A New Non-monotone Line Search Algorithm for Nonlinear Programming

Abstract. We study the application of a kind of non-monotone line search’s technique in conjugate gradient method. At present, most of the study of conjugate gradient methods are using Wolfe’s monotone line search, by constructing the condition of Zoutendijk, we can get the conclusion that it’s convergence by using reduction to absurdity. Here we study the global convergence of conjugate gradient methods with Armijo-type line search, the thought of proof wasn’t using the method above mentioned.

Keywords: Unconstrained optimization, Non-monotone line search, Global convergence.

Introduction

Unconstrained optimization problems

\( \min_{x \in \mathbb{R}^n} f(x) \)

where \( f: \mathbb{R}^n \to \mathbb{R} \) is continuously differentiable. The general form of the conjugate gradient method:

\[ x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, 2, \cdots \]

\[ d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \geq 1, \end{cases} \]

which \( \beta_k \) is a parameter, corresponding to different conjugate gradient method on the different emulated.

Since 1986 on Grippo, Lampariello and Lucidi[1] from first proposed a non-monotonic linear search technology non - line search provides a broader means. The early nineties of the last century, from Lucidi and Roma[2,3] only studied the form of non-monotonic strong Wolfe line search method the application of conjugate gradient algorithm to find the step size \( \alpha_0 \) to meet

\[ f(x_k + \alpha_k d_k) \leq \min_{\alpha \in \mathbb{R}} f(x_k + \alpha d_k) + \rho \alpha_k g_k^T d_k, \]

and

\[ |d_k^T g(x_k + \alpha_k d_k)| \leq -\delta_\alpha g_k^T d_k, \]

\[ 0 < \rho < \delta < 0.5, \quad M \text{ is a positive integer}, \quad \delta(0) = 0, \quad \delta(k) \leq \min(\delta(k-1)+1, M), \quad k \geq 1. \]


In this paper, a class of non-monotone line search (NLS), study the convergence properties of such an algorithm for general nonconvex function, and proved its global convergence. More conjugate gradient algorithm is used the Wolfe monotone line search, and by constructing Zoutendijk conditions, export contradictory to prove convergence. This article studies the global convergence of a class of non-monotone line search of Armijo-type conjugate gradient algorithm to prove that the idea comes from the on GLL search techniques. We study the condition of global convergence of four conjugate gradient methods with nonmonotone line searches. When the conditions is being increased, for nonconvex functions, we prove the global convergence of modified method.

Algorithm

Given \( \sigma > 0, \beta \in (0, 1), \delta \in (0, 1), M \) non-negative integer, and to make the initial test step

\[ r_k = -\frac{\nabla g_k^T d_k}{||d_k||^2}. \]

Take \( \alpha_k = \rho^m \delta^{(m)}, \quad m(k) = 0, 1, 2, \ldots, m(k) \) is to make the smallest non-negative integer set up by the following formula,

\[ f(x_k + \alpha_k d_k) \leq f(x_{k-1}) - \delta \alpha_k g_k^T d_k, \]

where, \( l(0) = 0, 0 \leq (k) \leq \min(\delta(k-1)+1, M), \quad k \geq 1. \) Obviously, if the descent direction \( d_k \) is met \( d_{k-1} \), then, when sufficiently large \( m(k) \), inequality (6) always holds, thus satisfying \( \alpha_k \) the conditions of existence. The above line search in each iteration, it is recommended that the initial test step \( r_k > 0 \) no longer remain the same, but can be adjusted automatically. On the value, change the initial test step approach can get better results, calculate a larger step size \( \alpha_0 \), thereby reducing the number of iterations.

Remarks

(1) In order to be able to use a non-monotonic linear search the NLS calculate the step length factor \( \alpha_k \) must be the search direction \( d_k \) is a descent direction. In the next section, we will prove that this study of the conjugate gradient algorithm to keep the search direction \( d_k \) is down.

(2) In order to be able to calculate a larger step size \( \alpha_k \), we can consider a mixed class of non-monotone line search, (6) can be rewritten as

\[ f(x_k + \alpha_k d_k) \leq \max_{\alpha \in \mathbb{R}} f(x_k) + \max_{\alpha \in \mathbb{R}} \{ -\delta_\alpha \alpha_k g_k^T d_k \}, \]

where \( \delta_\alpha, \delta_\beta \in (0, 1) \).

Convergence Analysis

First given the general assumptions of this section:

(A1) The level set

\[ L_0 = \{ x \mid f(x) \leq f(x_0), \quad x \in \mathbb{R}^n \} \]

is bounded.
A2. \( f(x) \) is differentiable in the level set \( L_0 \), and its gradient \( g(x) = V(f(x)) \) satisfies the Lipschitz condition, i.e., there is a constant \( L > 0 \), making
\[
\| g(x) - g(y) \| \leq L \| x - y \|, \forall x, y \in L_0.
\]

Lemma 1. Let the search direction \( d_t \) be a descent direction and step size factor \( \alpha_t \) be the non-monotone line search NLS, by (2) of iterates \( \{x_i\} \subseteq L_0 \).

Proof. By the non-monotone line search in the NLS (6) shows,
\[
\begin{align*}
    f(x_i) &\leq f(x_{i1}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    f(x_2) &\leq \max_{0 \leq t \leq 1} f(x_{i1}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    f(x_3) &\leq \max_{0 \leq t \leq 1} f(x_{i2}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    \vdots \\
    f(x_{k+1}) &\leq \max_{0 \leq t \leq 1} f(x_{i,k}) - \delta \| \alpha_t d_t \|^2 < f(x_0).
\end{align*}
\]

As a result, \( \{x_i\} \subseteq L_0 \).

Lemma 2. The step factor \( \alpha_t \) of non-monotone line search, by (2), amendment to the definition of method to meet
\[
g_t d_k = -\| g_k \|^2, \quad k = 0, 1, 2, \ldots
\]

Proof. Obviously
\[
g_0 d_0 = -\| g_0 \|^2.
\]

For \( k \geq 1 \),
\[
g_t^T d_k = \| g_k \|^2, k = 0, 1, 2, \ldots
\]

so
\[
g_t d_k = \| g_k \|^2, k = 0, 1, 2, \ldots
\]

Lemma 3. Under the conditions of the assumptions (A1), the sequence \( \{f(x_{i,k})\} \) satisfies the property of monotonicity.

Proof. Known \( f(x) \) was lower bound on the level set \( L_0 \), \( \{x_i\} \subseteq L_0 \) (see proof of Lemma 1) and sequence \( \{f(x_{i,k})\} \) decreases monotonically, then
\[
\lim_{k \to \infty} f(x_{i,k}) = 0.
\]

Proof. Known \( f(x) \) was lower bound on the level set \( L_0 \), \( \{x_i\} \subseteq L_0 \) (see proof of Lemma 1) and sequence \( \{f(x_{i,k})\} \) decreases monotonically, then
\[
\lim_{k \to \infty} f(x_{i,k}) = 0.
\]

Lemma 4. Assuming (A1) holds, then the limit
\[
\lim_{k \to \infty} f(x_{i,k}) = 0.
\]

Proof. Assume that
\[
\{h(k)\} \subset \{h(k)\}.
\]

Thus, by Lemma 3 shows that
\[
\lim_{k \to \infty} f(x_{i,k}) = 0.
\]

Namely \( h(k) \) is non-negative integer, and satisfy the following two formulas
\[
\begin{align*}
    f(x_i) &\leq f(x_{i1}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    f(x_2) &\leq \max_{0 \leq t \leq 1} f(x_{i1}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    f(x_3) &\leq \max_{0 \leq t \leq 1} f(x_{i2}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    \vdots \\
    f(x_{k+1}) &\leq \max_{0 \leq t \leq 1} f(x_{i,k}) - \delta \| \alpha_t d_t \|^2 < f(x_0).
\end{align*}
\]

As a result, \( \{x_i\} \subseteq L_0 \).

Lemma 2. The step factor \( \alpha_t \) of non-monotone line search, by (2), amendment to the definition of method to meet
\[
g_t d_k = -\| g_k \|^2, \quad k = 0, 1, 2, \ldots
\]

Proof. Obviously
\[
g_0 d_0 = -\| g_0 \|^2.
\]

For \( k \geq 1 \),
\[
g_t d_k = g_t^T (-\varphi_t g_k + \beta_{2R} d_{k-1})
\]

so
\[
g_t d_k = \| g_k \|^2, k = 0, 1, 2, \ldots
\]

Lemma 3. Under the conditions of the assumptions (A1), the sequence \( \{f(x_{i,k})\} \) satisfies the property of monotonicity.

Proof. Assume that
\[
\{h(k)\} \subset \{h(k)\}.
\]

Thus, by Lemma 3 shows that
\[
\lim_{k \to \infty} f(x_{i,k}) = 0.
\]

Namely \( h(k) \) is non-negative integer, and satisfy the following two formulas
\[
\begin{align*}
    f(x_i) &\leq f(x_{i1}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    f(x_2) &\leq \max_{0 \leq t \leq 1} f(x_{i1}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    f(x_3) &\leq \max_{0 \leq t \leq 1} f(x_{i2}) - \delta \| \alpha_t d_t \|^2 < f(x_0), \\
    \vdots \\
    f(x_{k+1}) &\leq \max_{0 \leq t \leq 1} f(x_{i,k}) - \delta \| \alpha_t d_t \|^2 < f(x_0).
\end{align*}
\]

As a result, \( \{x_i\} \subseteq L_0 \).

Lemma 2. The step factor \( \alpha_t \) of non-monotone line search, by (2), amendment to the definition of method to meet
\[
g_t d_k = -\| g_k \|^2, \quad k = 0, 1, 2, \ldots
\]

Proof. Obviously
\[
g_0 d_0 = -\| g_0 \|^2.
\]

For \( k \geq 1 \),
\[
g_t^T d_k = g_t^T (-\varphi_t g_k + \beta_{2R} d_{k-1})
\]

so
\[
g_t^T d_k = \| g_k \|^2, k = 0, 1, 2, \ldots
\]
on both sides of order \( i \). \( f(x) \) was uniformly continuous on the level set \( L_0 \), which
\[
\lim_{k \to \infty} f(x_{h(k)-(i+1)}) = \lim_{k \to \infty} f(x_{h(k)}),
\]
and noted \( \delta > 0 \), so
\[
\lim_{k \to \infty} \alpha_{h(k)-(i+1)} \| d_{h(k)-(i+1)} \| = 0.
\]
Thus, this indicates that, on the arbitrary \( i \geq 1 \), (11) was established.

The (14) also implies
\[
\| x_{h(k)-i} - x_{h(k)-(i+1)} \| \to 0 \quad (k \to \infty).
\]

This shows that the arbitrary \( i \geq 1 \), (12) have also set up.

By \( \hat{h} \) definition and (8) are available
\[
\hat{h}(k) = h(M + 2) \leq k + M + 2,
\]
namely
\[
\hat{h}(k) - k - 1 \leq M + 1.
\]

Thus, for any \( k \), do deformed
\[
x_{k+1} = x_{\hat{h}(k)} - \sum_{i=1}^{M} (x_{\hat{h}(k)-i} - x_{\hat{h}(k)-i}).
\]
where \( x_{\hat{h}(k)} \) transposition, and noting (15), was
\[
\| x_{k+1} - x_{\hat{h}(k)} \| = \sum_{i=1}^{M} (x_{\hat{h}(k)-i} - d_{\hat{h}(k)-i}) \leq \sum_{i=1}^{M} \alpha_{\hat{h}(k)-i} \| d_{\hat{h}(k)-i} \|.
\]

On both sides of order \( k \to \infty \), by (15),
\[
\lim_{k \to \infty} \| x_{k+1} - x_{\hat{h}(k)} \| = 0.
\]

Thus, by the uniform continuity of \( f(x) \),
\[
\lim_{k \to \infty} f(x_{k+1}) = \lim_{k \to \infty} f(x_{\hat{h}(k)}).
\]

Then from (16), we can see
\[
\lim_{k \to \infty} f(x_{k+1}) = \lim_{k \to \infty} f(x_{\hat{h}(k)}).
\]

(10) on both sides so that \( k \to \infty \), by (17), and noted \( \delta > 0 \) that Lemma 5 holds.

**Remarks**

If (4) change the formation of line search is a monotone class, \( f(x) \) is clearly seen to be monotone decreasing, if \( f(x) \) there is a lower bound, easy to get
\[
\sum_{k=1}^{\infty} \alpha_k^2 \| d_k \|^2 < +\infty,
\]
in particular, have Lemma 5. Where the weak non-monotonic search, to prove Lemma 5 spent a lot of twists and turns.

Lemma 6. Under the assumptions (A1), (A2) conditions, the search direction \( d_k \) is a descent direction, the step size factor \( \alpha_k \) by the nonmonotone line search NLS, then there exists a constant \( C > 0 \) such that for any \( k \geq 0 \),
\[
\alpha_k \geq C \| g_k \|^2 \| d_k \|^2.
\]

Lemma 7. Under the assumptions (A1), (A2) conditions for Algorithm 1, if there exists a constant \( C > 0 \), \( \varepsilon > 0 \) such that for any \( k \geq 0 \),
\[
g_k^T d_k \leq - C \| g_k \|^2,
\]
and
\[
\| g_k \| \geq \varepsilon.
\]

Then
\[
\lim_{k \to \infty} \| d_k \| = +\infty.
\]

Theorem. Assumptions (A1), (A2) conditions for the algorithm where the search direction using the modified method. There is a constant \( C > 0 \), making sufficiently large \( k \) to meet
\[
\| d_k \| \leq C \| g_k \|^2 \| g_k \| \leq C \| g_k \|^2.
\]
Either there is some \( k \), making the
\[
\| g_k \| = 0.
\]

Either have
\[
\liminf_{k \to \infty} \| g_k \| = 0.
\]

**Conclusions**

The program prepared by the Matlab6.5 in general on a PC. Test functions from [11], indicated in brackets after the function name is the number of variables. NLS is a line search method proposed in this paper, by GLL on Grippo-Lampariello-Lucidi from non-monotonic line search. \( n_i \) represents the number of iterations, \( n_r \) the number of times that the function value, gradient calculation is the number \( n_i+1 \). We were calculated for different values of \( M \), when \( M=0 \), ie, monotone line search. To the pros and cons of the algorithm, the parameters are uniform taken as, \( \alpha = 1, \beta = 0.2 \), \( \delta = 0.9, \varepsilon = 10^{-9} \). Our conjugate gradient method is divided into two kinds of numerical experiment for a class of initial testing step according to this formula to the case of correction, the other is the case of an initial test step length fixed for a. From the results of the comparison, the proposed line search termination criterion has the following advantages:

1. Monotone line search \( (M = 0) \), or non-monotone line search \( (M > 0) \), the number of iterations of the NLS method, the function value calculation times are reduced.
2. Usually better than the initial test step fixed the case when the initial testing step according to this formula be amended.
3. Non-monotone strategy is effective for most of the functions, especially high-dimensional, or initial testing step fixed the situation.
Acknowledgments
The work is supported by Funding Project for Academic Human Resources Development in Institutions of Higher Learning under the Jurisdiction of Beijing Municipality (PHR (IHLB)) (PHR201108407). Thanks for the help.

REFERENCES

Authors: Jing Zhang, associate professor, basic courses department, Beijing union university, Beijing 100101, China. E-mail: zhang1jing4@sohu.com.