Jing ZHANG
Beijing Union University

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.*

***Streszczenie.** Przeprowadzono studia nad zastosowaniem niemonotonicznego badania prostej w sprzężonej metodzie gradientowej. Obecnie najczęściej wykorzystuje się metodę Wolfa ale nasze badania wykazały że lepsze wyniki uzyskuje się w metodach globalnej zbieżności sprzężonej metody gradientowej. (**Nowy niemonotoniczny algorytm poszukiwania prostej w programowaniu nieliniowym**)*

Keywords: Unconstrained optimization, Non-monotone line search, Global convergence.
Słowa kluczowe: zbieżność globalna, metody gradientowe, programowanie nieliniowe.

Introduction

Unconstrained optimization problems

(1) $$\min_{x\in\mathbf{R}^n} f(x)$$

where $f:\mathbf{R}^n\rightarrow\mathbf{R}$ is continuously differentiable. The general form of the conjugate gradient method:

(2) $$x_{k+1}=x_k+\alpha_k d_k\ ,\ k=0\,,1\,,2\,,\cdots$$

(3) $$d_k=\begin{cases}-g_k\ , & k=0\ ,\\ -g_k+\beta_k d_{k-1}\ , & k\geq 1\ ,\end{cases}$$

which β_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 α_k to meet

(4) $$f(x_k+\alpha_k d_k)\leq \max_{0\leq j\leq l(k)} f(x_{k-j})+\rho\alpha_k g_k^T d_k\ ,$$

and

(5) $$|\,d_k^T g(x_k+\alpha_k d_k)\,|\leq -\delta g_k^T d_k\ ,$$

which $0<\rho<\delta<0.5$, M is a positive integer, $l(0)=0$, $0\leq l(k)\leq\min\{l(k\text{-}1)+1, M\}$, $k\geq1$.

In 1999, Guang-Hui Liu, Jing, LX Han, and D. Han[4] in the above-mentioned nonmonotonic strong Wolfe line search method to prove that two types of classical conjugate gradient algorithm of the PRP method and HS method for convex objective function in these non-monotonic Wolfe line search , global convergence. In 2002, Dai Yu Hong[5] the initial non-monotone line search method on Grippo-Lucidi-Lampariello search and hybrid conjugate DY-CD method combined with the study. In 2006, Chinese scholars Zhang Li, Wei-Jun Zhou and Li Dong Hui[7] on this basis, given the amendments.

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 searchs. 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

$$r_k=-\frac{\sigma g_k^T d_k}{\|\,d_k\,\|^2}\ .$$

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

(6) $$f(x_k+\alpha_k d_k)\leq \max_{0\leq j\leq l(k)} f(x_{k-j})-\delta\,\|\,\alpha_k d_k\,\|^2\ ,$$

where, $l(0)=0$, $0\leq l(k)\leq\min\{l(k\text{-}1)+1, M\}$, $k\geq1$. Obviously, if the descent direction d_k is met $d_k^T g_k<0$, then, when sufficiently large $m(k)$, inequality (6) always holds, thus satisfying α_k the conditions of existence. The above line search in each iteration, it is recommended that the initial test step r_k 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 α_k, 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 α_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 α_k, we can consider a mixed class of non-monotone line search, (6) can be rewritten as

$$f(x_k+\alpha_k d_k)\leq \max_{0\leq j\leq l(k)} f(x_{k-j})+ \max\{-\delta_1\,\|\,\alpha_k d_k\,\|^2\ ,\ \delta_2\alpha_k g_k^T d_k\}\ ,$$

where δ_1, $\delta_2\in(0,\ 1)$.

Convergence Analysis

First given the general assumptions of this section:

(A1) The level set

$$L_0=\{x\,|\,f(x)\leq f(x_0)\ ,\ x\in\mathbf{R}^n\}$$

is bounded.

(A2) $f(x)$ is differentiable in the level set L_0, and its gradient $g(x)=\nabla f(x)$ satisfies the Lipschitz condition, ie, there is a constant $L>0$, making

(7) $$\|g(x)-g(y)\| \le L\,\|x-y\|\,,\forall x\,,\, y\in L_0\,.$$

Lemma 1. Let the search direction d_k is a descent direction and step size factor α_k by the non-monotone line search NLS, by (2) of iterates $\{x_k\}\subset L_0$.

Proof. By the non-monotone line search in the NLS (6) shows,

$$f(x_1)\le f(x_0)-\delta\,\|\alpha_0 d_0\|^2 < f(x_0)\,,$$

$$f(x_2)\le \max_{0\le j\le l(1)} f(x_{1-j})-\delta\,\|\alpha_1 d_1\|^2 < f(x_0)\,,$$

$$f(x_3)\le \max_{0\le j\le l(2)} f(x_{2-j})-\delta\,\|\alpha_2 d_2\|^2 < f(x_0)\,,$$

$$\cdots\cdots$$

$$f(x_{k+1})\le \max_{0\le j\le l(k)} f(x_{k-j})-\delta\,\|\alpha_k d_k\|^2 < f(x_0)\,.$$

$$\cdots\cdots$$

As a result, $\{x_k\}\subset L_0$.

Lemma 2[7]. The step factor α_k NLS1 of non-monotone line search, by (2), amendment to the definition of method to meet

$$g_k^T d_k = -\|g_k\|^2\,,\ k=0\,,1\,,2\,,\cdots$$

Proof. Obviously

$$g_0^T d_0 = -\|g_0\|^2\,.$$

For $k\ge 1$,

$$\begin{aligned}
g_k^T d_k &= g_k^T(-\theta_k g_k + \beta_k^{FR} d_{k-1})\\
&= g_k^T\Big(-\frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}\,g_k + \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,d_{k-1}\Big)\\
&= -\frac{d_{k-1}^T y_{k-1}}{\|g_{k-1}\|^2}\,\|g_k\|^2 + \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,g_k^T d_{k-1}\\
&= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,(-d_{k-1}^T y_{k-1} + g_k^T d_{k-1})\\
&= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,[-d_{k-1}^T(g_k-g_{k-1}) + g_k^T d_{k-1}]\\
&= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,d_{k-1}^T g_{k-1}\,,
\end{aligned}$$

Thus, we find a recurrence relation

$$g_k^T d_k = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,g_{k-1}^T d_{k-1}\,,$$

so

$$\begin{aligned}
g_k^T d_k &= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,g_{k-1}^T d_{k-1}\\
&= \frac{\|g_k\|^2}{\|g_{k-1}\|^2}\,\frac{\|g_{k-1}\|^2}{\|g_{k-2}\|^2}\,g_{k-2}^T d_{k-2}\\
&= \frac{\|g_k\|^2}{\|g_{k-2}\|^2}\,g_{k-2}^T d_{k-2}\\
&= \cdots\\
&= \frac{\|g_k\|^2}{\|g_0\|^2}\,g_0^T d_0\\
&= -\|g_k\|^2\,.
\end{aligned}$$

For simplicity, we introduce the notation:

$$h(k)=\max\{i \mid 0\le k-i\le l(k)\,,\quad f(x_i)=\max_{0\le j\le l(k)} f(x_{k-j})\}\,.$$

Namely $h(k)$ is non-negative integer, and satisfy the following two formulas

(8) $$k-l(k)\le h(k)\le k\,,$$

(9) $$f(x_{h(k)})=\max_{0\le j\le l(k)} f(x_{k-j})\,.$$

So non-monotone line search NLS1 (4) can be rewritten as

(10) $$f(x_{k+1})\le f(x_{h(k)})-\delta\,\|\alpha_k d_k\|^2\,.$$

Lemma 3. Under the conditions of the assumptions (A1), the sequence $\{f(x_{h(k)})\}$ is decreases monotonically.

Lemma 4. Assuming (A1) holds, then the limit

$$\lim_{k\to\infty} f(x_{h(k)})$$

exists, and

$$\lim_{k\to\infty}\alpha_{h(k)-1}\,\|d_{h(k)-1}\|=0\,.$$

Proof. Known $f(x)$ was lower bound on the level set L_0, $\{x_k\}\subset L_0$ (see proof of Lemma 1) and sequence $\{f(x_{h(k)})\}$ decreases monotonically, then

$$\lim_{k\to\infty} f(x_{h(k)})$$

exists, by (10),

$$f(x_{h(k)})\le f(x_{h(h(k)-1)})-\delta\,\|\alpha_{h(k)-1}d_{h(k)-1}\|^2\,,$$

On both sides of order $k\to\infty$, and noted $\delta>0$ that it. So

$$\lim_{k\to\infty}\|\alpha_{h(k)-1}d_{h(k)-1}\|^2=0\,.$$

Lemma 5. Assuming (A1) holds, then

$$\lim_{k\to\infty}\alpha_k\,\|d_k\|=0\,.$$

Proof. Remember

$$\hat h(k)=h(k+M+2)\,,$$

first proved by mathematical induction, for any $i\ge 1$, two of the following holds:

(11) $$\lim_{k\to\infty}\alpha_{\hat h(k)-i}\,\|d_{\hat h(k)-i}\|=0\ ,$$

(12) $$\lim_{k\to\infty} f(x_{\hat h(k)-i})=\lim_{k\to\infty} f(x_{h(k)})\ .$$

That $i=1$, by $\hat h$ definition, apparently

$$\{\hat h(k)\}\subset\{h(k)\}\,.$$

Thus, by Lemma 3 shows that

$$\lim_{k\to\infty} f(x_{\hat h(k)})$$

is existence and

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

Know (11) was established. But

$$x_{\hat h(k)}-x_{\hat h(k)-1}=s_{\hat h(k)}=\alpha_{\hat h(k)-1}d_{\hat h(k)-1}\,,$$

this shows that

$$\|x_{\hat h(k)}-x_{\hat h(k)-1}\|\to 0\ \ (k\to\infty\,),$$

and $f(x)$ then by in L_0 the uniformly continuous, it is

$$\lim_{k\to\infty} f(x_{\hat h(k)-1})=\lim_{k\to\infty} f(x_{\hat h(k)})=\lim_{k\to\infty} f(x_{h(k)})\,.$$

i=1, (12) was established.

It is assumed that the establishment of a given i, (11) and (12). By (10),

$$f(x_{\hat{h}(k)-i}) \le f(x_{h(\hat{h}(k)-i-1)}) - \delta \| \alpha_{\hat{h}(k)-i-1} d_{\hat{h}(k)-i-1} \|^2,$$

On the order of the above equation on both sides $k\to\infty$, by (12), and

$$\lim_{k\to\infty} f(x_{h(\hat{h}(k)-i-1)}) = \lim_{k\to\infty} f(x_{h(k)}),$$

and noted δ>0, so

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

This indicates that, on the arbitrary $i\ge1$, (11) was established.

The (14) also implies

$$\| x_{\hat{h}(k)-i} - x_{\hat{h}(k)-(i+1)} \| \to 0 \ (k\to\infty),$$

$f(x)$ was uniformly continuous on the level set L_0, which

$$\begin{aligned} &\lim_{k\to\infty} f(x_{\hat{h}(k)-(i+1)}) \\ &= \lim_{k\to\infty} f(x_{\hat{h}(k)-i}) \\ &= \lim_{k\to\infty} f(x_{\hat{h}(k)}) \\ &= \lim_{k\to\infty} f(x_{h(k)}) \end{aligned}$$

This shows that the arbitrary $i\ge1$, (12) have also set up.

By $\hat{h}$ definition and (8) are available

$$\hat{h}(k) = h(k+M+2) \le k+M+2,$$

namely

(15) $$\hat{h}(k) - k - 1 \le M + 1.$$

Thus, for any k, do deformed

$$\begin{aligned} & x_{k+1} \\ &= x_{\hat{h}(k)} - \sum_{i=1}^{\hat{h}(k)-k-1} (x_{\hat{h}(k)-i+1} - x_{\hat{h}(k)-i}) \\ &= x_{\hat{h}(k)} - \sum_{i=1}^{\hat{h}(k)-k-1} \alpha_{\hat{h}(k)-i} d_{\hat{h}(k)-i}. \end{aligned}$$

where $x_{\hat{h}(k)}$ transposition, and noting (15), was

(16) $$\begin{aligned} \| x_{k+1} - x_{\hat{h}(k)} \| &= \| - \sum_{i=1}^{\hat{h}(k)-k-1} \alpha_{\hat{h}(k)-i} d_{\hat{h}(k)-i} \| \\ &\le \sum_{i=1}^{M+1} \| \alpha_{\hat{h}(k)-i} d_{\hat{h}(k)-i} \|. \end{aligned}$$

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 $f(x)$,

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

Then from (16), we can see

(17) $$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(x_{h(k)}).$$

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

Remarks

If (4) change the formation of line search is a monotone class, $f(x_k)$ is clearly seen to be monotone decreasing, if $f(x)$ there is a lower bound, easy to get

$$\sum_{k=0}^{\infty} \alpha_k^2 \| d_k \|^2 < +\infty,$$

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 α_k by the nonmonotone line search NLS, then there exists a constant C_2>0 such that for any $k\ge0$,

(18) $$\alpha_k \ge C_2 \frac{| g_k^T d_k |}{\| d_k \|^2}.$$

Lemma 7. Under the assumptions (A1), (A2) conditions for Algorithm 1, if there exists a constant C_3>0, ε>0 such that for any $k\ge0$,

(19) $$g_k^T d_k \le -C_3 \|g_k\|^2,$$

and

(20) $$\|g_k\| \ge \varepsilon.$$

Then

(21) $$\lim_{k\to\infty} \| d_k \| = +\infty.$$

Theorem. Assumptions (A1), (A2) conditions for the algorithm where the search direction using the modified method. There is a constant C_4>0, making sufficiently large k to meet

(22) $$\| d_k \| \le C_4 \frac{\| g_k \|^2}{\| g_{k+1} \|^2}.$$

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_f 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, σ=1, β=0.2, δ=0.9, $\varepsilon=10^{-6}$. 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

[1] GRIPPO L, LAMPARIELLO F, LUCIDI S. A nonmonotone line search technique for Newton's method. SIAM Journal on Numerical Analysis, 1986, 23 (4): 707-716.

[2] LUCIDI S, ROMA M. Nonmonotone conjugate gradient methods for optimization. Technical Report IASI-CNR, R. 342, 1992.

[3] LUCIDI S, ROMA M. Nonmonotone conjugate gradient methods for optimization. HENRY Jacques, YVON Jean-Pierre. System Modeling and Optimization, Proceedings of the 16th IFIP-TC7 Conference, Compiègne, France — July 5–9, 1993. Berlin: Springer Verlag, 1994. 204-214.

[4] LIU G H, JING L L, HAN L X, HAN D. A class of nonmonotone conjugate gradient methods for unconstrained optimization. Journal of Optimization Theory and Applications, 1999, 101 (1): 127-140.

[5] DAI Yuhong. A nonmonotone conjugate gradient algorithm for unconstrained optimization. Journal of Systems Science and Complexity. 2002, 15 (2): 139-145.

[6] FLETCHER R, REEVES C M. Function minimization by conjugate gradients. The Computer Journal, 1964, 7 (2): 149–154.

[7] ZHANG Li, ZHOU Weijun, LI Donghui. Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search. Numerische Mathematik, 2006, 104 (4): 561-572.

[8] SUN Wenyu, HAN Jiye, SUN Jie. Global convergence of nonmonotone descent methods for unconstrained optimization problems. Journal of Computational and Applied Mathematics, 2002, 146 (1): 89-98.

[9] SHI Zhen-Jun, SHEN Jie. Convergence of nonmonotone line search method. Journal of Computational and Applied Mathematics. 2006, 193 (2): 397–412.

[10] SUN Wenyu, YUAN Ya-xiang. Optimization theory and methods: Nonlinear programming. New York: Springer, 2006. 47-48.

[11] MORÉ Jorge J, GARBOW Burton S, HILLSTROM Kenneth E. Testing unconstrained optimization software. ACM Transactions on Mathematical Software, 1981, 7 (1): 17-41.

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