A Family of Conjugate Gradient Methods[1]

Chunyan Hu

College of Electronic Engineering and Automation,

Guilin University of Electronic Technology, Guilin, 541004, P. R. China

Benxin Zhang, Weijuan Shi

School of Mathematics and Computing Science,

Guilin University of Electronic Technology, Guilin, 541004, P. R. China

Abstract. Conjugate gradient method is a method for solving unconstrained optimization problems, especially for large-scale problems. In this paper, a new parameter is given and we propose a new family of conjugate gradient methods. In particular, some famous conjugate gradient methods are special cases. The global convergence is proved with an inexact line search.

Key words. Unconstrained optimization; Conjugate gradient method; New Armijo-type line search; Global convergence

1 Introduction

Consider the unconstrained optimization problem

$$\min \quad f(x),\ x \in R^n, \tag{1.1}$$

where $f(x)$ is continuously differentiable function, and $g(x)$ is the gradient function of $f(x)$, i.e., $g(x) = \nabla f(x)$. The iterative method has the form as follows:

$$x_{k+1} = x_k + \alpha_k d_k, k = 0, 1, 2, \cdots, n. \tag{1.2}$$

where x_k is the current iterate point, α_k the step size, d_k the search direction. The conjugate gradient direction has the following formula

$$d_k = \begin{cases} -g_0,\ k = 0, \\ -g_k + \beta_k d_{k-1},\ k \geq 1, \end{cases} \tag{1.3}$$

where $g_k = g(x_k)$. the β_k has the well-know types of FR,DY,PRP,LS as follows:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2},\ \beta_k^{DY} = \frac{g_k^T g_k}{d_{k-1}^T y_{k-1}}, \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2},\ \beta_k^{LS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \tag{1.4}$$

where $\|\cdot\|$ is the Euclidean norm, $y_{k-1} = g_k - g_{k-1}$, $s_{k-1} = x_k - x_{k-1}$.

In the line search method, in order to guarantee d_k to be a descent direction, d_k is required to satisfy

$$g_k^T d_k < 0. \tag{1.5}$$

[1]This work was supported in part by NNSF(No.11361018) of China and Guangxi Fund for Distinguished Young Scholars (2012GXSFFA060003).

Corresponding author: E-mail: huchyel@hotmail.com

AMO -Advanced Modeling and Optimization. ISSN: 1841-4311

Sometimes d_k is required to satisfy $g_k^T d_k \leq -c\|g_k\|^2$ which guarantees the global convergence.

Some literatures and related papers made a lot of work about the construction of β_k and the convergence of the corresponding algorithm. In [1], Zhang proposed a three-term PRP conjugate gradient method, where d_k was given by

$$d_{k+1} = -g_{k+1} + \beta_k^{PRP} d_k - \theta_k y_k,\ \theta_k = \frac{g_{k+1}^T d_k}{\|g_k\|^2}. \tag{1.6}$$

A good property of the three-term PRP conjugate gradient method is that the direction generated by the method satisfies $g_k^T d_k = -\|g_k\|^2$. Dai [4] gave the new parameter

$$\beta_k^{VPRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2 + \tau s_{k-1}^T d_{k-1}},$$

then proposed the corresponding three-term PRP conjugate gradient method and proved its global convergence with the standard Armijo line search

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k.$$

In [2], the author gave an Armijo-type line search. Shi [5] proposed a new Armijo-type line search and proved the convergence of LS method. Shi [6] proved the convergence of PRP and other conjugate gradient methods by using the Armijo-type line search [2]. Motivated by Dai [4], we present a new β_k as follows:

$$\beta_k = \frac{g_k^T y_{k-1}}{(1-u)\|g_{k-1}\|^2 - u g_{k-1}^T d_{k-1}}, \tag{1.7}$$

where $u \in [0,1]$. For $u = 0$, β_k is the β_k^{PRP}, and for $u = 1$, β_k is the β_k^{LS}.

2 New Armijo-type line search

We firstly make the following assumption.

Assumption A:

(a) The level set

$$\Omega = \{x \in R^n | f(x) \leq f(x_0)\} \tag{2.1}$$

is bound.

(b) The function f is continuously differentiable with *Lipschitz* continuous gradient on an open ball D containing Ω, i.e., there is a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|. \tag{2.2}$$

$L > 0$ is a *Lipschitz* constant and we do not know it at priori. So, it needs to be estimated. In [7, 8], some methods of estimating L were proposed as follows:

$$L = \frac{\|y_{k-1}\|}{\|s_{k-1}\|}, \tag{2.3}$$

$$L = \frac{\|y_{k-1}\|^2}{|s_{k-1}^T y_{k-1}|}, \tag{2.4}$$

$$L = \frac{|s_{k-1}^T y_{k-1}|}{\|s_{k-1}\|^2}. \tag{2.5}$$

By the Armijo-type line search, we should choose *Lipschitz* constants as small as possible in practical computation. In the *kth* iteration we take the estimated *Lipschitz* constants as:

$$L_k = max\{L_0, \frac{\|y_{k-1}\|}{\|s_{k-1}\|}\}, \tag{2.6}$$

$$L_k = max\{L_0, min\{\frac{\|y_{k-1}\|^2}{|s_{k-1}^T y_{k-1}|}, M_0'\}\}, \tag{2.7}$$

$$L_k = max\{L_0, \frac{|s_{k-1}^T y_{k-1}|}{\|s_{k-1}\|^2}\}, \tag{2.8}$$

where L_0 and M_0' being a large positive number. Motivated by [5], we present a different new Armijo-type line search.

Moreover, we present the new Armijo-type line search. Given $\delta \in (0, \frac{1}{2})$, $\rho \in (0, 1)$, $c \in (\frac{1}{2}, 1)$, and $u \in [0, 1]$, set $\eta_k = \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2}$ and α_k is the largest in $\{\eta_k, \eta_k\rho, \eta_k\rho^2, \cdots, \}$, which satisfies

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta\alpha_k g_k^T d_k. \tag{2.9}$$

where L_k is estimated by (2.6),(2.7),(2.8) respectively.

Algorithm:

Step0: Given an initial point $x_0 \in R^n$, $d_0 = -g_0$, $k := 0$.

Step1: If $\|g_k\| = 0$, then stop, else go to Step 2.

Step2: Set $x_{k+1} = x_k + \alpha_k d_k$. Compute the direction d_k by (1.3), where

$$\beta_k = \frac{g_k^T y_{k-1}}{(1-u)\|g_{k-1}\|^2 - u g_{k-1}^T d_{k-1}}.$$

α_k is defined by the above new Armijo-type line search.

Step3: Set $k := k + 1$, go to Step 1.

Lemma 2.1. *Let Assumption A hold, and the algorithm with the new Armijo-type line search generates an infinite sequence $\{x_k\}$. Then, there exist $m_0 > 0$ and $M_0 > 0$ such that*

$$m_0 \le L_k \le M_0. \tag{2.10}$$

Proof. By (2.6), (2.7) or (2.8), we can easy to obtain $L_k \ge L_0$. We take $m_0 = L_0$. For (2.3) and (2.6), we have

$$L_k = max\{L_0, \frac{\|y_{k-1}\|}{\|s_{k-1}\|}\} \le max\{L_0, L\}.$$

For (2.4) and (2.7), we have

$$L = max\{L_0, min\{\frac{\|y_{k-1}\|^2}{|s_{k-1}^T y_{k-1}|}, M_0'\}\} \le max\{L_0, M_0'\}.$$

For (2.5) and (2.8), we have

$$L = max\{L_0, \frac{|s_{k-1}^T y_{k-1}|}{\|s_{k-1}\|^2}\} \le \{L_0, L\}.$$

So we can take $M_0 = max\{L_0, L, M_0'\}$. The proof is finished. ∎

Lemma 2.2. *Let AssumptionA hold, under the new line search, for all $k > 1$, it holds that*

$$\alpha_k \leq \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2}. \tag{2.11}$$

Then the following inequality

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2 \tag{2.12}$$

holds.

Proof. By the inequality (2.11) and the $Cauchy - Schwarz$ inequality, we have

$$\begin{aligned}
(1-c)[(1-u)\|g_k\|^2 - ug_k^T d_k] &\geq \alpha_k L\|d_k\|^2 \\
&= \frac{\alpha_k L\|g_{k+1}\|\|d_k\|}{\|g_{k+1}\|^2}\|g_{k+1}\|\|d_k\| \\
&\geq \frac{\|g_{k+1}\|\|g_{k+1} - g_k\|}{\|g_{k+1}\|^2}|g_{k+1}^T d_k| \\
&\geq \frac{g_{k+1}^T(g_{k+1} - g_k)}{(1-u)\|g_k\|^2 - ug_k^T d_k}\frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|g_{k+1}\|^2}g_{k+1}^T d_k \\
&\geq \beta_{k+1}\frac{(1-u)\|g_k\|^2 - uy_k^T d_k}{\|g_{k+1}\|^2}g_{k+1}^T d_k.
\end{aligned}$$

Therefore

$$(1-c)\|g_{k+1}\|^2 \geq \beta_{k+1}g_{k+1}^T d_k.$$

And thus

$$-c\|g_{k+1}\|^2 \geq -\|g_{k+1}\|^2 + \beta_{k+1}g_{k+1}^T d_k = g_{k+1}^T d_{k+1}.$$

The proof is finished. ∎

Lemma 2.3. *Let AssumptionA hold, if the stepsize α_k is generated by the new Armijo line search, there exists a constant $c_1 > 0$ such that the following inequality holds for all k sufficiently large,*

$$\alpha_k \geq c_1\frac{\|g_k\|^2}{\|d_k\|^2}. \tag{2.13}$$

Proof. We have from (2.9) and Assumption A that

$$\sum_{i=1}^{\infty} -\delta\alpha_k g_k^T d_k < +\infty. \tag{2.14}$$

This together with (2.12) yields

$$\sum_{i=1}^{\infty} c\alpha_k\|g_k\|^2 \leq -\sum_{i=1}^{\infty}\delta\alpha_k g_k^T d_k < +\infty. \tag{2.15}$$

We prove (2.13) by considering the following two cases.

Case (i). $\alpha_k = \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2}$, we can obtain

$$\alpha_k = \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - ug_k^T d_k}{\|d_k\|^2} \geq \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2}{\|d_k\|^2}. \tag{2.16}$$

In this case the inequality (2.13) is satisfied with $c_1 = \frac{(1-c)(1-u)}{L}$.

Case (ii). $\alpha_k < \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2}$. By the line search condition, $\alpha = \rho^{-1}\alpha_k$ does not satisfy (2.9), this means

$$f(x_k) - f(x_k + \alpha d_k) < -\delta \alpha_k g_k^T d_k. \tag{2.17}$$

Using the mean value theorem on the left-hand side of the above inequality, there exists some $t_k \in (0,1)$ such that

$$-\alpha g(x_k + t_k \alpha d_k)^T d_k < -\delta \alpha_k g_k^T d_k,$$

i.e.

$$g(x_k + t_k \alpha d_k)^T d_k > \delta \alpha_k g_k^T d_k.$$

By $Assumption A(b)$, the $Cauchy - Schwarz$ inequality, the above inequality, and Lemma 2.1, we have

$$\begin{aligned} L\alpha\|d_k\|^2 &\geq \|g(x_k + t_k \alpha d_k)^T - g_k\|\|d_k\| \\ &\geq (g(x_k + t_k \alpha d_k)^T - g_k)^T d_k \geq -(1-\delta) g_k^T d_k \\ &\geq c(1-\delta)\|g_k\|^2. \end{aligned}$$

We have

$$\alpha_k \geq \frac{c\rho(1-\delta)}{L} \frac{\|g_k\|^2}{\|d_k\|^2}.$$

Letting $c_1 = min\{\frac{(1-c)(1-u)}{L}, \frac{c\rho(1-\delta)}{L}\}$, we can get $\alpha_k \geq c_1 \frac{\|g_k\|^2}{\|d_k\|^2}$. The proof is finished. ∎

From inequalities (2.13) and (2.15), we can easily obtain the following Zoutendijk condition.

Lemma 2.4. *Suppose Assumption A holds.* x_k *is generated by the Algorithm and* α_k *is generated by the new Armijo line search, then we have*

$$\sum_{k \geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{2.18}$$

3 Global Convergence

Theorem 3.1. *Let Assumption A hold, the algorithm generates an infinite sequence* $\{x_k\}$*. Then we have*

$$\lim_{k \to \infty} \inf \|g_k\| = 0. \tag{3.1}$$

Proof. For the sake of contradiction, we suppose that the conclusion is not true. Then there exists a constant $\epsilon > 0$ such that

$$\|g_k\| > \epsilon, \quad \forall k \geq 0. \tag{3.2}$$

Since $g_k \neq 0$ and with (2.12), it follows that $d_k \neq 0$.

$$\alpha_k \leq \frac{(1-c)}{L_k} \cdot \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2} \leq \frac{(1-c)}{m_0} \cdot \frac{(1-u)\|g_k\|^2 - u g_k^T d_k}{\|d_k\|^2}.$$

By the β_k formula and the $Cauchy - Schwarz$ inequality, we have

$$\begin{aligned}\|d_{k+1}\| &= \|-g_{k+1}+\beta_{k+1}d_k\| \\ &\leq \|g_{k+1}\| + \frac{|g_{k+1}(g_{k+1}-g_k)|}{(1-u)\|g_k\|^2 - ug_k^T d_k}\|d_k\| \\ &\leq \|g_{k+1}\|(1+\frac{\alpha_k L\|d_k\|^2}{(1-u)\|g_k\|^2 - ug_k^T d_k}) \\ &\leq \|g_{k+1}\|(1+\frac{L(1-c)}{m_0}).\end{aligned}$$

Let$\sqrt{A} = (1+\frac{L(1-c)}{m_0})$,then $\|d_k\|^2 \leq A\|g_k\|^2$. So

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} \geq \sum_{k=0}^{\infty}\frac{\epsilon^2}{A} = \infty.$$

Which contradicts with (2.18). Hence, $\lim_{k\to\infty}\inf\|g_k\| = 0$.

4 Numerical experiments

In this section, we carry out some numerical experiments. Our algorithm has been tested on some problems as follows, where x_0 is the initial point, and x_k is the final point.

Example 1 $f(x) = 4*(x_1-5)^2 + (x_2-6)^2$, $x_0 = (8,9)$, $x_k = (5.0000, 6.0000)$.

Example 2 $f(x) = (x_2 - x_1^2)^2 + 100*(1-x_1)^2$, $x_0 = (-1.2, 1)$, $x_k = (1.0000, 1.0000)$.

Example 3 $f(x) = (x_1 - x_2 + x_3)^2 + (-x_1 + x_2 + x_3)^2 + (x_1 + x_2 - x_3)^2$, $x_0 = (1,2,3)$, $x_k = (0,0,0)$.

Example 4 $f(x) = (x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2$, $x_0 = (1,1)$, $x_k = (3.0000, 2.0000)$.

Example 5 $f(x) = (x_1-2)^2 + (x_2-1)^2 + 0.04/(-x_1^2/4 - x_2^2 + 1) + (x_1 - 2x_2 + 1)^2/0.2$, $x_0 = (2,2)$, $x_k = (1.7954, 1.3779)$.

Example 6 $f(x) = \sum\limits_{i=1}^{n}(16-i)(x(i)-1)^2 + 10*\sum\limits_{i=1}^{n}(16-i)((x(i)-1)^2)^2$, $x_0 = (0,\cdots,0)$, $x_k = (1,\cdots,1)$.

Example 7 $f(x) = \sum\limits_{i=1}^{n-1} x_i^2 + (x_{i+1}+x_i^2)^2$, $x_0 = (1,\cdots,1)$, $x_k = (0,\cdots,0)$.

Example 8 $f(x) = \sum\limits_{i=1}^{n} ix_i^2 + (\sum\limits_{i=1}^{n} x_i^2)^2$, $x_0 = (1,\cdots,1)$, $x_k = (0,\cdots,0)$.

Example 9 $f(x) = \sum\limits_{i=1}^{n}(e^{x_i} - x_i)$, $x_0 = (1,\cdots,1)$, $x_k = (0,\cdots,0)$.

We set the parameters $\delta = 0.25, \rho = 0.5, c = 0.75$ and $L = 1$ in the numerical experiment. For the parameters $u = 0, 0.5, 1$, we named LS method (corresponding to $u = 0$), L+P method (corresponding to $u = 0.5$) and PRP method (corresponding to $u = 1$). The results are summarized in Table 1. For the test problem, n is the dimension, $\|g_k\|$ is the norm maximum of the LS, L+P and PRP methods, and k is the number of iteration for the problem. The stop criterion is

$$\|g_k\| \leq 10^{-6},$$

and the numerical results are given in Table 1.

Table 1 The detail information of numerical experiments for our algorithm

NO.	n	$\|g_k\|$	k(u=0)	k(u=0.5)	k(u=1)
1	2	1.9439e-007	23	18	21
2	2	2.6062e-007	28	34	21
3	3	7.0589e-007	34	37	40
4	2	2.8989e-007	18	24	18
5	2	6.2970e-007	31	23	29
6	10	5.9188e-007	45	45	23
7	50	6.7852e-007	18	24	31
8	100	5.7160e-007	67	67	34
9	1000	9.3892e-007	30	37	46
	10000	9.3807e-007	31	38	46

Table 1 shows the performance of the our algorithm about relative to the iteration. It is easy to see that, for all problems, the algorithm is very efficient. The results for each problem are accurate, and with less number of times of iteration.

5 Conclusions

In this paper, we propose a new family conjugate gradient formula for computing unconstrained optimization problems. PRP and LS conjugate gradient methods are special cases. Motivated by [5], we present a different new Armijo-type line search. The global convergence of this method is established under the new Armijo line search.

References

[1] L. Zhang, W.J. Zhou and D.H. Li. *A descent modified Polak-Ribire-Polyak conjugate gradient method and its global convergence*. IMA J. Numer. Anal, (2006),26.629-640.

[2] L. Grippo, S. Lucidi, *A globally convergent version of the Polak-Ribie'‘re conjugate gradient method*, Mathematical Programming, (1997), 78.375-391.

[3] Y. Yuan and W. Sun. *Theory and Methods of Optimization*, Science Press of China, Beijing, (1999).

[4] Z.F Dai. *Two modified Polak-Ribire-Polyak-type nonlinear conjugate methods with sufficient descent property.*NFNO.J. (2010),31(8).892-906.

[5] Z.J Shi, J. She. *Convergence of Liu-Storey conjugate gradient method*, European Journal of Operational Research, (2007),182.552-560.

[6] Z.J Shi,,J. She. *Convergence of Polak-Ribire-Polyak conjugate gradient method*, Nonlinear Anlysis, (2007),66.1428-1441.

[7] Z.J. Shi, J. Shen. *Convergence of descent method without line search*, Applied Mathematics and Computation,(2005),167.94-107.

[8] Z.J. Shi, J. Shen. *Step-size estimation for unconstrained optimization methods*,Computational and Applied Mathematics, (2005),24(3).399-416.