Convergence of a modified PRP conjugate gradient method with a new formula of step-size

Khelifa Bouaziz*
Department of Science and Technology
Larbi Tebessi University
Tebessa
Algeria
khalifa.bouaziz@univ-tebessa.dz

Tahar Bachawat
Mohamed cherif Messaadia University
Souk-Ahras
Algeria
t.bechouat@gmail.com

Abstract. We present in this paper the global convergence of a modified PRP (Polak-Ribière-Polyak) conjugate gradient method suggested by Min and Jing [11], by using a new formula of step-size that combination by Wu [14], and by Sun and colleagues [3, 12]. Some numerical results are also presented.
Keywords: conjugate gradient methods, global convergence, PRP method, step-size, line search.

1. Introduction

Let us consider the following unconstrained minimization problem: $f(x), x \in R^n$, where f is a differentiable objective function, has the following form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.1}$$

$$\text{where} \quad d_k = \begin{cases} -g_k, & \text{for } k = 1, \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{1.2}$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k.

Motivated by the ideas of Wei and al. [14] and Dai and Wen [5], which spured Min and Jing [11] construct two modified PRP methods, in which the parameter β_k is specified as follows:

$$\beta_k^{MPRP} = \frac{g_k^T y_{k-1}}{\mu |g_k^T d_{k-1}| + \|g_{k-1}\|^2}, \tag{1.3}$$

∗. Corresponding author

where $\|.\|$ means the Euclidean norm, $y_{k-1} = g_k - g_{k-1}$, and $\mu \geq 0$ is a constant. Let us remark that the descent direction d_k is defined by

$$g_k^T d_k = -c \|g_k\|^2, \tag{1.4}$$

where $0 < c < 1$.

The global convergence properties of conjugate gradient method have been studied by many researchers [2-9].

In the implementation of any conjugate gradient (CG) method, the step-size is often determined by certain line search conditions such as the Wolfe conditions [13]. These types of line search involve extensive computation of function values and gradients, which often becomes a significant burden for large-scale problems, which spured Sun [12], and Wu [14] to pursue the conjugate gradient method where they calculated the step-size instead of the line search. The new formula for step-size α_k in the form

$$\alpha_k = \frac{-\delta g_k^T d_k}{(\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2}, \tag{1.5}$$

where

$$\delta \in (0, (\kappa + \gamma)/\tau), \gamma \geq 0, \tag{1.6}$$

τ and κ confirm the Assumption 2.1 below, $\bar{g}_{k+1}$ denote $\bigtriangledown f(x_k + d_k)$.

In this paper, our goal is to employ the step-formula (1.5) to prove the convergence of a modified PRP conjugate gradient method.

This paper is organized as follows. Some preliminary results on the family of CG methods with the new-form step-size formula (1.5) are given in Section 2. Section 3 includes the main convergence properties of the modified PRP conjugate gradient method.

2. Properties of the new step-size

The present section gathers technical results concerning the step-size α_k generated by (1.5).

Assumption 2.1. *The function f is LC^1 and strongly convex in R^n, i.e, there exists constants $\tau > 0$ and $\kappa \geq 0$ such that*

$$\|\bigtriangledown f(u) - \bigtriangledown f(v)\| \leq \tau \|u - v\|, \forall u, v \in R^n, \tag{2.1}$$

and

$$[\bigtriangledown f(u) - \bigtriangledown f(v)]^T (u - v) \geq \kappa \|u - v\|^2, \forall u, v \in R^n. \tag{2.2}$$

Note that Assumption 2.1 implies that the level set $L = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded.

Lemma 2.2 *Suppose that Assumption 2.1 holds. Then the following inequalities*

$$\kappa \|s_k\|^2 \leq y_k^T s_k \leq \tau \|s_k\|^2, \tag{2.3}$$

where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$ *and*

$$(\kappa + \gamma) \|d_k\|^2 \leq (\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2 \leq (\tau + \gamma) \|d_k\|^2, \tag{2.4}$$

hold for all k.

Proof. It is straightforward from (2.1) and (2.2) that (2.3) holds. Now, we prove (2.4) is true

$$\begin{aligned}(\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2 &\leq \|\bar{g}_{k+1} - g_k\| \|d_k\| + \gamma \|d_k\|^2 \\ &\leq (\tau + \gamma) \|d_k\|^2.\end{aligned} \tag{2.5}$$

Then, by (2.2), we have

$$(\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2 \geq \kappa \|d_k\|^2 + \gamma \|d_k\|^2 \geq (\kappa + \gamma) \|d_k\|^2. \tag{2.6}$$

Hence, it follows from (2.5) and (2.6) that (2.4) hold for all k. □

Lemma 2.3. *Suppose that* x_k *is given by* (1.1), (1.2) *and* (1.5). *Then*

$$g_{k+1}^T d_k = \rho_k g_k^T d_k, \tag{2.7}$$

holds for all k, *where* $0 < \rho_k = 1 - \delta \Phi_k \|d_k\|^2 / [(\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2]$, *and*

$$\Phi_k = \begin{cases} 0, & \text{for } \alpha_k = 0, \\ (g_{k+1} - g_k)^T (x_{k+1} - x_k) / \|x_{k+1} - x_k\|^2, & \text{for } \alpha_k \neq 0. \end{cases} \tag{2.8}$$

Proof. If $\alpha_k = 0$, then $\rho_k = 1$ and $x_{k+1} = x_k$. Thus, (2.7) is true.

Now, we suppose that $\alpha_k \neq 0$. From (2.8) and (2.6), we have

$$\begin{aligned} g_{k+1}^T d_k &= g_k^T d_k + (g_{k+1} - g_k)^T d_k \\ &= g_k^T d_k + \alpha_k^{-1} (g_{k+1} - g_k)^T (x_{k+1} - x_k) \\ &= g_k^T d_k + \alpha_k^{-1} \Phi_k \|x_{k+1} - x_k\|^2 \\ &= g_k^T d_k + \alpha_k \Phi_k \|d_k\|^2 \\ &= g_k^T d_k - \{\delta g_k^T d_k / [(\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2]\} \Phi_k \|d_k\|^2 \\ &= \{1 - \delta \Phi_k \|d_k\|^2 / [(\bar{g}_{k+1} - g_k)^T d_k + \gamma \|d_k\|^2]\} g_k^T d_k \\ &= \rho_k g_k^T d_k. \end{aligned}$$

The proof is complete. □

Corollary 2.4. *Suppose that Assumption 2.1 holds. Then*

$$\frac{\delta\kappa}{\tau+\gamma} \leq 1-\rho_k \leq \frac{\delta\tau}{\kappa+\gamma}, \tag{2.9}$$

holds for all k.

Proof. It follows From (2.3) and (2.4), we obtain (2.9).

Lemma 2.5. *Suppose that Assumption 2.1 holds and $\{x_k\}$ is generated by* (1.1), (1.2) *and* (1.5). *Then*

$$\sum_{d_k\neq 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.10}$$

Proof. By the mean-value theorem, we have

$$f(x_{k+1}) - f(x_k) = \bar{g}^T(x_{k+1}-x_k), \tag{2.11}$$

where $\bar{g} = \triangledown f(\bar{x})$ for some $\bar{x} \in [x_{k,}x_{k+1}]$. Now, by the Cauchy-Schwartz inequality, (1.5), and Assumption 2.1 we obtain

$$\begin{aligned}
\bar{g}^T(x_{k+1}-x_k) &= g_k^T(x_{k+1}-x_k) + (\bar{g}-g_k)^T(x_{k+1}-x_k) \\
&\leq g_k^T(x_{k+1}-x_k) + \|\bar{g}-g_k\|\,\|x_{k+1}-x_k\| \\
&\leq g_k^T(x_{k+1}-x_k) + \tau\,\|x_{k+1}-x_k\|^2 \\
&= \alpha_k g_k^T d_k + \tau\alpha_k^2\,\|d_k\|^2 \\
&= \alpha_k g_k^T d_k - \tau\alpha_k\delta g_k^T d_k\,\|d_k\|^2 / [(\bar{g}_{k+1}-g_k)^T d_k + \gamma\,\|d_k\|^2] \\
&= \alpha_k g_k^T d_k\Big(1 - \frac{\tau\delta\,\|d_k\|^2}{(\bar{g}_{k+1}-g_k)^T d_k + \gamma\,\|d_k\|^2}\Big).
\end{aligned} \tag{2.12}$$

By from (2.4) and (2.12), we obtain

$$\begin{aligned}
\alpha_k g_k^T d_k &= -\frac{\delta}{(\bar{g}_{k+1}-g_k)^T d_k + \gamma\,\|d_k\|^2}(g_k^T d_k)^2 \\
&\leq -\frac{\delta}{(\tau+\gamma)}\frac{(g_k^T d_k)^2}{\|d_k\|^2},
\end{aligned} \tag{2.13}$$

by (2.12) and (1.6), we have

$$1 - \frac{\tau\delta\,\|d_k\|^2}{(\bar{g}_{k+1}-g_k)^T d_k + \gamma\,\|d_k\|^2} \geq 1 - \frac{\tau\delta}{\kappa+\gamma} > 0. \tag{2.14}$$

From (2.13) and (2.14), it follows that

$$\Omega = \frac{\delta}{\tau+\gamma}\Big(1 - \frac{\tau\delta}{\kappa+\gamma}\Big) > 0. \tag{2.15}$$

From (2.11) we have,

$$f(x_{k+1}) - f(x_k) \leq -\Omega \frac{(g_k^T d_k)^2}{\|d_k\|^2} \leq 0, \tag{2.16}$$

which implies $f(x_{k+1}) \leq f(x_k)$. Hence, it follows from (2.16) that (2.10) is true. The proof is complete. □

Lemma 2.6. *Suppose that Assumption* 2.1 *holds, then we have*

$$\sum_k \alpha_k^2 \|d_k\|^2 < \infty. \tag{2.17}$$

Proof. By (1.5) and (2.4) we have

$$\begin{aligned} \sum_k \alpha_k^2 \|d_k\|^2 &= \sum_k \frac{(\delta g_k^T d_k)^2}{[(\bar{g}_{k+1} - g_k)^T d_k + \|d_k\|^2]^2} \|d_k\|^2 \\ &\leq (\frac{\delta}{\kappa + \gamma})^2 \sum_{d_k \neq 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \end{aligned} \tag{2.18}$$

The proof is complete. □

3. Global convergence of the modified PRP method

In this section, we discuss the convergence properties of a modified PRP method conjugate gradient method, in which β_k^{MPRP} is given by (1.3).

We give the following algorithm firstly.

Algorithm 3.1
Step 0: Given $x_1 \in R^n$, set $d_1 = -g_1$, $k = 1$.
Step 1: If $\|g_k\| = 0$ then stop else go to Step 2.
Step 2: Set $x_{k+1} = x_k + \alpha_k d_k$ where d_k is defined by (1.2), and α_k is defined by (1.5).
Step 3: Compute β_{k+1}^{MPRP} using formula (1.3).
Step 4: Set $k := k + 1$, go to Step 1.

In 1992, Gilbert and Nocedal introduced the property (*) which plays an important role in the studies of CG methods. This property means that the next research direction approaches the steepest direction automatically when a small step-size is generated, and the step-sizes are not produced successively [15].

Property (*). Consider a CG method of the form (1.1) and (1.2). Suppose that, for all k,

$$0 < r \leq \|g_k\| \leq \bar{r}, \tag{3.1}$$

where r and $\bar{r}$ are two constants. If there exist $b > 1$ and $\lambda > 0$ such that for all k,

$$|\beta_k^{MPRP}| \leq b, \tag{3.2}$$

and

$$\|s_k\| \leq \lambda \Longrightarrow |\beta_k^{MPRP}| \leq \frac{1}{2b}, \tag{3.3}$$

where $s_{k-1} = \alpha_{k-1} d_{k-1}$.

The following Lemma shows that the MPRP method has Property (*).

Lemma 3.2. *Consider the method of form* (1.1) *and* (1.2). *Suppose that Assumption 2.1 hold, then, the method β_k^{MPRP} has Property* (*).

Proof. Consider any constant r and $\bar{r}$ which satisfy (3.1).

Let $b = \dfrac{2\bar{r}^2}{r^2} > 1$, $\lambda = \dfrac{r^4}{4\tau\bar{r}^3}$. By (1.3) we have

$$|\beta_k^{MPRP}| \leq |\frac{g_k^T y_{k-1}}{\mu|g_k^T d_{k-1}| + \|g_{k-1}\|^2}| \leq \frac{\|g_k\|^2 + \|g_k\| \, \|g_{k-1}\|}{\|g_{k-1}\|^2} \leq \frac{2\bar{r}^2}{r^2} = b. \tag{3.4}$$

From (2.1), holds. If then

$$|\beta_k^{MPRP}| \leq \frac{\|g_k\| \, \|g_k - g_{k-1}\|}{\|g_{k-1}\|^2} \leq \frac{\tau \, \|s_{k-1}\| \, \|g_k\|}{\|g_{k-1}\|} \leq \frac{\tau\lambda\bar{r}}{r^2} = \frac{1}{2b}. \tag{3.5}$$

The proof is finished. □

Theorem 3.3. *Under Assumption 2.1, the method defined by* (1.1), (1.2), (1.5) *and* (1.3) *will generate a sequence $\{x_k\}$ such that* $\lim_{k\longrightarrow\infty} \inf \|g_k\| = 0$.

Proof. Suppose on the contrary that $\|g_k\| \geq \psi$, for all k.

Since L is bounded, both $\{x_k\}$ and $\{g_k\}$ are bounded. By using

$$\|d_k\| \leq \|g_k\| + |\beta_k^{MPRP}| \, \|d_{k-1}\| \,, \tag{3.6}$$

one can show that $\{\|d_k\|\}$ is uniformly bounded. Definition (1.2) implies the following relation

$$|g_k^T d_k| = |g_k^T(-g_k + \beta_k^{MPRP} d_{k-1})| \tag{3.7}$$

$$\geq \|g_k\|^2 - |\beta_k^{MPRP}| \, \|g_k\| \, \|d_{k-1}\| \,. \tag{3.8}$$

From (1.3) and using the Cauchy-Schwarz inequality, we have

$$|\beta_k^{MPRP}| = |\frac{g_k^T(g_k - g_{k-1})}{\mu|g_k^T d_{k-1}| + \|g_{k-1}\|^2}|. \tag{3.9}$$

From (2.1) and (2.18) we have

$$\begin{aligned} \|g_k - g_{k-1}\| &\le \tau\alpha_{k-1}\|d_{k-1}\| \\ &\le (\frac{\tau\delta}{\kappa+\gamma})\frac{|g_{k-1}^T d_{k-1}|}{\|d_{k-1}\|} \le \frac{|g_{k-1}^T d_{k-1}|}{\|d_{k-1}\|}. \end{aligned} \tag{3.10}$$

From (1.4), (2.7) we have

$$\mu|g_k^T d_{k-1}| + \|g_{k-1}\|^2 = (\mu\rho_{k-1} + \frac{1}{c})|g_{k-1}^T d_{k-1}| = m|g_{k-1}^T d_{k-1}|, (m > 1). \tag{3.11}$$

By (3.9), (3.10), and (3.11) we have

$$|\beta_k^{MPRP}|\,\|d_{k-1}\| \le \frac{\|g_k\|}{m}. \tag{3.12}$$

Hence by substituting (3.12) in (3.8), we have

$$|g_k^T d_k| \ge A\|g_k\|^2, A = \frac{m-1}{m}, \tag{3.13}$$

for large k. Thus we have

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2\|g_k\|^2} \ge A^2\frac{\|g_k\|^2}{\|d_k\|^2}. \tag{3.14}$$

Since $\|g_k\| \ge \psi$ and $\|d_k\|$ is bounded above, we conclude that there is $\varepsilon > 0$ such that $\frac{(g_k^T d_k)^2}{\|d_k\|^2\|g_k\|^2} \ge \varepsilon$, which implies $\sum_{d_k \ne 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty$.

This is a contradiction to Lemma 2.5. □

4. Numerical experiments and discussions

In this part, we present the numerical experiments of the new formula (1.5) and apply it using (1.3), computer
(Processor: Intel(R)core(TM)i3-3110M cpu@2.40GHZ, Ram 4.00 GB) through the Matlab programme.
10 testing problems have been taken from [1].

This will lead us to test for the global convergence properties of our method. Stopping criteria is set to $\|g_k\| \le \varepsilon$ where $\varepsilon = 10^{-6}$. Taking into consideration the following parameters: $\gamma = 1.5$ and $\mu = 0.5$.

Table 1 list numerical results. The meaning of each column is as follows: “Problem ”the name of the test problem, “δ”, “Xzero”, “k ”the number of iterations, “Time”, “Xoptimal”.

The following results showed the effectiveness of the proposed method.

Table 1

	Problem	δ	Xzero	k	Time	Xopimal
1	Booth	1	(1 1)	46	0.118	(1.0 3.0)
2	Branin	1.5	(1 1)	54	0.113	(3.1416 2.275)
3	Sphere	1	(-1 1)	64	0.015	(-0.230 -0.230)
4	Exponential	1	(-1 1)	59	0.082	(-0.6406 -0.6406)
5	Himmelblau	2	(1 1)	258	0.084	(0.6403 -0.6403)
6	Matyas	1	(-1 1)	34	0.047	(0.6403 -0.6403)
7	McCormick	1	(-1 1)	36	0.048	(-0.5472 -1.5472)
8	Rosenbrock	0.4	(1 1)	4999	0.735	(0.4198 1.9116)
9	SIX-HUMP CAMEL	2	(1 1)	15	0.031	(-0.0898 0.7127)
10	THREE-HUMP CAMEL	1.5	(1 1)	46	0.1180	(0.2665 -0.2935)

Acknowledgements

The authors would like to thank the referees and the editors for their careful reading and some useful comments on improving the presentation of this paper.

References

[1] N. Andrei, *An unconstrained optimization test functions collection*, Adv. Model. Optim., 10 (2008), 147-161.

[2] Kh. Bouaziz, Y. Laskri, *Convergence of a two-parameter family of conjugate gradient methods with a fixed formula of stepsize*, Boletim da Sociedade Paranaense de Matematica, 38 (2020), 127-140.

[3] X. Chen, J. Sun, *Global convergence of a two-parameter family of conjugate gradient methods without line search*, Journal of Computational and Applied Mathematics, 146 (2002), 37-45.

[4] Y. Dai, Y. Yuan, *A three-parameter family of nonlinear conjugate gradient methods*, Mathematics of Computation, 70 (2001), 1155-1167.

[5] Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM Journal on Optimization, 10 (1999), 177-182.

[6] R. Fletcher, *Practical methods of optimization*, John Wiley & Sons, 1987.

[7] R. Fletcher, C. M. Reeves, *Function minimization by conjugate gradients*, The Computer Journal, 7 (1964), 149-154.

[8] M. R. Hestenes, E. Stiefel, *Methods of conjugate gradients for solving linear systems*, NBS, 49 (1952).

[9] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms*, part 1: Theory, Journal of Optimization Theory and Applications, 69 (1991), 129-137.

[10] B. T. Polyak, *The conjugate gradient method in extremal problems*, USSR Computational Mathematics and Mathematical Physics, 9 (1969), 94-112.

[11] S. Min, L. Jing, *Three modified Polak-Ribière-Polyak conjugate gradient methods with sufficient descent property*, Journal of Inequalities and Applications, 1 (2015), 125.

[12] J. Sun, J. Zhang, *Global convergence of conjugate gradient methods without line search*, Annals of Operations Research, 103 (2001), 161-173.

[13] P. Wolfe, *Convergence conditions for ascent methods*, II: Some corrections, SIAM review, 13 (1971), 185-188.

[14] Q.-j. Wu, *A nonlinear conjugate gradient method without line search and its global convergence*, In Computational and Information Sciences (ICCIS), 2011 International Conference on IEEE, 2001, 1148-1152.

[15] Z. Yueqin, Z. Hao and Z. Chuanlin, *Global convergence of a modified PRP conjugate gradient method*, Procedia Engineering, Elsevier, 31 (2012), 986-995.

[16] W. Zengxin, Y. Shengwei, L, Liying, *The convergence properties of some new conjugate gradient methods*, Applied Mathematics and Computation, Elsevier, 183 (2006), 1341-1350.

[17] D. Zhifeng, W. Fenghua, *Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property*, Applied Mathematics and Computation, Elsevier, 2018 (2012), 7421-7430.

Accepted: November 23, 2020