DOI: 10.13656/j.cnki.gxkx.2008.01.004
广西科学 Guangxi Sciences 2008, 15(1): 14~ 19

A Conjugate Gradient Formula Generated by PRP and HS Formulas*
基于 PRP和 HS公式产生的一个共轭梯度公式

CHEN Cui-ling[1], ZHAO Yan[2]
陈翠玲[1],赵 岩[2]

(1. College of Mathematics Science, Guangxi Normal University, Guilin, Guangxi, 541004, China; 2. School of Management, University of Shanghai for Science and Technology, Shanghai, 200093, China)
(1. 广西师范大学数学科学学院,广西桂林 541004; 2. 上海理工大学管理学院,上海 200093)

Abstract With the similarity of the form between PRP and HS formulas which have the same numerator and different denominators, and proper combining and composing, a new conjugate gradient formula is obtained. The present method based on this formula possesses a sufficient descent property with the strong Wolfe-Powell line search. Under some suitable assumptions and the weak Wolfe-Powell line search, the global convergence result is established. The preliminary numerical results show that the proposed method is efficient.

Key words: unconstrained optimization, conjugate gradient method, Wolfe-Powell conditions, global convergence

摘要: 根据 PRP和 HS公式具有相同分子只是分母不同的相似性,通过适当的结合和构造,给出一个新的共轭梯度公式.证明该公式的新方法在强 Wolfe-Powell线搜索下具有充分下降性,在适当的假设和弱 Wolfe-Powell线搜索下具有全局收敛性,并用数值试验证实新方法是有效的.

关键词: 无约束优化 共轭梯度法 Wolfe-Powell条件 全局收敛
中图法分类号: O241.7 文献标识码: A 文章编号: 1005-9164(2008)01-0014-06

Polak-Ribière-Polyak (PRP) nonlinear conjugate gradient method was reported separately by Polak, Ribière and Polyak in 1969. The form of this method is listed as follows

$$x_{k+1} = x_k + \alpha_k d_k, \qquad (0.1)$$

where α_k is a steplength which is computed by carrying out some line search, such as the weak Wolfe-Powell (WWP) line search:

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \delta\alpha_k g_k^T d_k, \qquad (0.2)$$

$$g(x_k + \alpha_k d_k)^T d_k \geqslant \sigma g_k^T d_k, \qquad (0.3)$$

or the strong Wolfe-Powell (SWP) line search: (0.2) and

$$|g(x_k + \alpha_k d_k)^T d_k| \leqslant -\sigma g_k^T d_k, \qquad (0.4)$$

where $\delta \in (0,1)$ and $\sigma \in (\delta, 1)$, d_k is the search direction defined by

$$d_k = \begin{cases} -g_k & \text{if } k = 1, \\ -g_k + \beta_k d_{k-1} & \text{if } k \geqslant 2, \end{cases} \qquad (0.5)$$

where the parameter β_k is computed by the following formula:

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T g_{k-1}}. \qquad (0.6)$$

In addition, the other formulae of β_k were also given by other authors:

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}}, \qquad (0.7)$$

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}, \qquad (0.8)$$

$$\beta_k^{CD} = -\frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}, \qquad (0.9)$$

收稿日期: 2006-12-12
修回日期: 2007-03-26
作者简介: 陈翠玲(1978-),女,助教,主要从事最优化理论与方法研究工作。
* Supported by the Natural Science Foundation of China (10161002), the Natural Science Foundation of Guangxi Province (0542043) and the Youth Foundation of Guangxi Normal University.

$$\beta_k^{LS} = -\frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T g_{k-1}}, \quad (0.10)$$

$$\beta_k^{DY} = \frac{g_k^T g_k}{(g_k - g_{k-1})^T d_{k-1}}, \quad (0.11)$$

which were successively called HS, FR, CD, LS and DY method.

PRP method is one of the methods with the best numerical behavior among the above methods till now. When the algorithm generates a small steplength, the search direction d_k generated by PRP method approaches automatically to a negative gradient direction, and then avoids efficiently the drawback of FR method, which generates continuously small steplengths. Powell[1] proved the global convergence of PRP method when the steplength $s_k = x_{k+1} - x_k$ approaches to zero. But for a general nonconvex function, Powell[2] gave a counterexample which showed PRP method was not convergent. Dai[3] also illustrated with examples, even if $f(x)$ is uniform convex, and the parameter $\sigma \in (0,1)$ is sufficiently small, PRP method is very likely to generate an ascent direction. For a general nonconvex function, Powell[4] suggested to restrict β_k^{PRP} to be nonnegative

$$\beta_k = \max\{0, \beta^{PRP}\}. \quad (0.12)$$

Gilbert and Nocedal[5] considered the above suggestion of Powell's and established the global convergence of the above varied PRP methods for a general nonconvex function under the proper line search. However, Gilbert and Nocedal[5] also showed by examples that even if the object function is uniform convex, β_k^{PRP} is also likely to be negative. Over all the above results, the global convergence of PRP method is not optimistic. Therefore, in order to find a conjugate gradient method which has good numerical result and converges globally, it is very significant to do further study of PRP formula.

In this paper, we investigate a conjugate gradient method generated by a new formula. In section 1, we represent the new algorithm of this method and its properties. The global convergence result is given in section 2. The preliminary numerical results are contained in section 3. Finally, we have a conclusion section.

1 Algorithm

Since the conjugate gradient methods belong to the descent methods for solving unconstrained optimization problems, the new β_k should be chosen such that $g_k^T d_k \leqslant 0$ if a line search is used. Furthermore, due to the sufficient descent condition

$$g_k^T d_k \leqslant -c\|g_k\|^2 \quad (1.1)$$

is a very nice and important property for conjugate gradient methods, we hope that the new formula β_k satisfies (1.1). In the following, we will find β_k such that d_k satisfies (1.1).

Noting the similarity of the form between PRP and HS formulae which have the same numerator and different dinominators, through combining and composing properly, we get a conjugate gradient formula as follows:

$$\beta_k^{PH}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \frac{\lambda_1\|g_k\|^2 - \lambda_4|g_k^T g_{k-1}|}{\lambda_2|(g_k - g_{k-1})^T d_{k-1}| + \lambda_3\|g_{k-1}\|^2}, \quad (1.2)$$

where $\lambda_1 \in (0,+\infty)$, $\lambda_2 \in (\lambda_1\frac{\sigma}{1-\sigma},+\infty)$, $\lambda_3 \in (0,+\infty)$, $\lambda_4 \in (0,+\infty)$. In order to ensure the nonnegative of β_k, we define

$$\beta_k^{PH+}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \max\{0, \beta_k^{PH}(\lambda_1,\lambda_2,\lambda_3,\lambda_4)\}. \quad (1.3)$$

In the following, we investigate whether the method generated by (1.2) and (1.3) satisfies the sufficient descent condition (1.1).

A sequence β_k is called a descent sequence (or sufficient descent sequence) for conjugate gradient methods if there exists a constant $\xi \in [0,1]$ (or $\xi \in [0,1)$) such that for all $k \geqslant 2$,

$$\beta_k g_k^T d_{k-1} \leqslant \xi\|g_k\|^2. \quad (1.4)$$

If we use the SWP conditions (0.2) and (0.4) to choose α_k, then it is easy to check that $\beta_k^{PH+}(\lambda_1,\lambda_2,\lambda_3,\lambda_4)$ is a descent sequence for conjugate gradient methods if $g_{k-1}^T d_{k-1} \leqslant 0$. In fact, if $\beta_k^{PH+}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) = 0$, then it is obvious that $\beta_k^{PH+}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) g_k^T d_{k-1} = 0 \leqslant \xi\|g_k\|^2$, it is true with any line search; and if $\beta_k^{PH+}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) = \beta_k^{PH}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) > 0$, and $g_{k-1}^T d_{k-1} \leqslant 0$, then from the SWP conditions, we have

$$|(g_k - g_{k-1})^T d_{k-1}| \geqslant |g_{k-1}^T d_{k-1}| - |g_k^T d_{k-1}| \geqslant \frac{1}{\sigma}|g_k^T d_{k-1}| - |g_k^T d_{k-1}| = (\frac{1}{\sigma} - 1)|g_k^T d_{k-1}|. \quad (1.5)$$

So

$$\beta_k^{PH+}(\lambda_1,\lambda_2,\lambda_3,\lambda_4) g_k^T d_{k-1} = \frac{\lambda_1\|g_k\|^2 - \lambda_4|g_k^T g_{k-1}|}{\lambda_2|(g_k - g_{k-1})^T d_{k-1}| + \lambda_3\|g_{k-1}\|^2} g_k^T d_{k-} \leqslant$$

$$\frac{\mu_1\|g_k\|^2-\mu_4|g_k^Tg_{k-1}|}{\mu_2|(g_k-g_{k-1})^Td_{k-1}|+\mu_3\|g_{k-1}\|^2}|g_k^Td_{k-1}|\leqslant$$
$$\frac{\mu_1\|g_k\|^2}{\mu_2|(g_k-g_{k-1})^Td_{k-1}|}|g_k^Td_{k-1}|\leqslant$$
$$\frac{\mu_1\|g_k\|^2}{\mu_2(\frac{1}{\sigma}-1)|g_k^Td_{k-1}|}|g_k^Td_{k-1}|\leqslant$$
$$\frac{\mu_1\|g_k\|^2}{\mu_2(\frac{1}{\sigma}-1)|g_k^Td_{k-1}|}|g_k^Td_{k-1}|=$$
$$\frac{\mu_1}{\mu_2}\frac{\sigma}{1-\sigma}\|g_k\|^2. \qquad (1.6)$$

From the above inequalities, whether $\beta_k^{PH*}=0$ or $\beta_k^{PH*}>0$, we can deduce that β_k^{PH*} satisfies (1.4) with $c=\frac{\mu_1}{\mu_2}\frac{\sigma}{1-\sigma}$. Therefore, we have the following results.

Theorem 1.1 Suppose $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)$ is defined by (1.2) and (1.3). Then with the SWP line search, for all $k\geqslant 1$, we have

$$g_k^Td_k\leqslant -(1-\frac{\mu_1}{\mu_2}\frac{\sigma}{1-\sigma})\|g_k\|^2. \qquad (1.7)$$

Proof For any $k\geqslant 1$, suppose that $g_k^Td_k\leqslant 0$. If $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)=0$, then $d_{k+1}=-g_{k+1}$. So we have $g_{k+1}^Td_{k+1}=-\|g_{k+1}\|^2\leqslant -(1-\frac{\mu_1}{\mu_2}\frac{\sigma}{1-\sigma})\|g_{k+1}\|^2$, for some positive scalar $\frac{\mu_1}{\mu_2}\frac{\sigma}{1-\sigma}$. Otherwise, from the definition of $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)$, we can obtain

$$g_{k+1}^Td_{k+1}=-\|g_{k+1}\|^2+$$
$$\frac{\mu_1\|g_{k+1}\|^2-\mu_4|g_{k+1}^Tg_k|}{\mu_2|(g_{k+1}-g_k)^Td_k|+\mu_3\|g_k\|^2}g_{k+1}^Td_k\leqslant$$
$$-\|g_{k+1}\|^2+$$
$$\frac{\mu_1\|g_{k+1}\|^2-\mu_4|g_{k+1}^Tg_k|}{\mu_2|(g_{k+1}-g_k)^Td_k|+\mu_3\|g_k\|^2}|g_{k+1}^Td_k|\leqslant$$
$$-\|g_{k+1}\|^2+\frac{\mu_1\|g_{k+1}\|^2}{\mu_2|(g_{k+1}-g_k)^Td_k|}|g_{k+1}^Td_k|\leqslant$$
$$-\|g_{k+1}\|^2+\frac{\mu_1\|g_{k+1}\|^2}{\mu_2(\frac{1}{\sigma}-1)|g_{k+1}^Td_k|}|g_{k+1}^Td_k|=$$
$$-(1-\frac{\mu_1}{\mu_2}\frac{\sigma}{1-\sigma})\|g_{k+1}\|^2.$$

The last inequality follows (1.5). Combining with $\mu_2\geqslant\mu_1\frac{\sigma}{1-\sigma}$, we have (1.7) provided that $g_k^Td_k\leqslant 0$. Hence, from $g_1^Td_1=-\|g_1\|^2\leqslant 0$, we can deduce that (1.7) holds for all $k\geqslant 1$.

The following theorem tells us that the range of $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)$ can be defined when its parameters satisfy some conditions.

Theorem 1.2 Suppose that $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)$ is defined by (1.2) and (1.3).

(a) If $\mu_1\leqslant\mu_2$ and WWP line search is used, then for all

$$k\geqslant 1, 0\leqslant\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)\leqslant\beta_k^{DY} \qquad (1.8)$$

hold;

(b) If $\mu_1\leqslant\mu_3$, then with any line search, for all $k\geqslant 1$,

$$0\leqslant\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)\leqslant\beta_k^{FR} \qquad (1.9)$$

hold.

Proof It is clear that the inequalities (1.8) and (1.9) hold under their own conditions when $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)=0$, where $\beta_k^{DY}>0$ is used with WWP line search. We consider the case where $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)>0$. So we have

(a) $\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)=$
$$\frac{\mu_1\|g_k\|^2-\mu_4|g_k^Tg_{k-1}|}{\mu_2|(g_k-g_{k-1})^Td_{k-1}|+\mu_3\|g_{k-1}\|^2}\leqslant$$
$$\frac{\mu_1\|g_k\|^2}{\mu_2|(g_k-g_{k-1})^Td_{k-1}|+\mu_3\|g_{k-1}\|^2}\leqslant$$
$$\frac{\mu_1\|g_k\|^2}{\mu_2|(g_k-g_{k-1})^Td_{k-1}|}\leqslant\beta_k^{DY},$$

where the last inequality uses that $\beta_k^{DY}>0$ with WWP line search.

$\beta_k^{PH*}(\mu_1,\mu_2,\mu_3,\mu_4)=$
$$\frac{\mu_1\|g_k\|^2-\mu_4|g_k^Tg_{k-1}|}{\mu_2|(g_k-g_{k-1})^Td_{k-1}|+\mu_3\|g_{k-1}\|^2}\leqslant$$
$$\frac{\mu_1}{\mu_3}\frac{\|g_k\|^2}{\|g_{k-1}\|^2}\leqslant\beta_k^{FR}.$$

Now we give the corresponding algorithm.

Algorithm 1.1 Step 0 Give $x_1\in R^n$, set $d_1=-g_1$, $k=1$. If $g_1=0$, then stop.

Step 1 Find a $\alpha_k>0$ satisfying the SWP conditions (0.2) and (0.4).

Step 2 Let $x_{k+1}=x_k+\alpha_kd_k$ and $g_{k+1}=g(x_{k+1})$. If $g_{k+1}=0$, then stop.

Step 3 Compute $\beta_{k+1}^{PH*}(\mu_{1(k+1)},\mu_{2(k+1)},\mu_{3(k+1)},\mu_{4(k+1)})$ by the formulae (1.2) and (1.3), then generate d_{k+1} by (0.5).

Step 4 Set $k:=k+1$, go to Step 1.

From Theorem 1.1, we have the following conclusions which indicate that Algorithm 1.1 is well-defined and has a nice property, i.e. the sufficient descent property (1.1) provided that we choose μ_{1k}

and $\_{2k}$ such that

$$u \equiv \inf\{u_k\} > 0, \qquad (1.10)$$

where $u_k = 1 - \frac{\_{1k}}{\_{2k}} \frac{e}{1-e}$.

Corollary 1.1 Suppose that (1.10) holds. Then Algorithm 1.1 is well-defined, either there is a k_0 such that $g_{k_0} = 0$ or generates a sequence $\{x_k\}$ such that for all k, the property (1.1) holds.

Proof If $g_1 = 0$, then we have finished our proof. Suppose $g_1 \neq 0$, Then $d_1 = -g_1$ and $g_1^T d_1 = -\|g_1\|^2 \neq 0$. So

$$g_1^T d_1 \leqslant -u\|g_1\|^2.$$

Then, we can generate (t_1, x_2, g_2). If $g_2 \neq 0$, we can have U_2 and d_2. Using Theorem 1.1, we have

$$g_2^T d_2 \leqslant -u_k\|g_2\|^2 \leqslant -u\|g_2\|^2.$$

Repeating the above discussions, and noting that $u_k > 0$, we can deduce our conclusion by induction. The following conclusion is a direct result of Theorem 1.1.

Corollary 1.2 For all $k \geqslant 1$,

$$g_k^T d_k \leqslant -u\|g_k\|^2. \qquad (1.11)$$

Remark 1.1 The nonnegative of the parameter U_k is important for some formulaes. For example, Powell[4] suggested that in PRP method, the parameter in (0.6) is not allowed to be negative, i.e. (0.12). By using a complicated line search, Gilbert and Nocedal[5] were able to establish the global convergence result of PRP and HS methods by restricting the scalar U_k to be nonnegative.

Remark 1.2 From Corollary 1.2, Algorithm 1.1 always generates a descent direction d_k in every step, and furthermore, the sufficient descent property (1.1) holds provided that $u > 0$. But it is not always the same for some formulaes. For example, even if f is a uniform convex function, PRP method with SWP line search may be fail to generate a descent direction[6]. The sufficient descent property (1.1) is very important for the global convergence of the conjugate gradient methods[5]. So we hope to keep (1.1) for the conjugate gradient methods. From Theorem 1.1, formula U_k^{PH} ($\_1, \_2, \_3, \_4$) possesses the sufficient descent property with SWP line search.

Remark 1.3 In Algorithm 1.1, formula U_k can be chosen with $\_{1k} = \_{2k} + \_{3k}$ or $\_{1k} \neq \_{2k} + \_{3k}$. If $\_{1k} \neq \_{2k} + \_{3k}$, it is not any nonlinear conjugate gradient method, but it has a close relationship with the conjugate gradient methods.

2 Global convergence results

Assumption A The level set $K = \{x \in R^n \mid f(x) \leqslant f(x_1)\}$ is bounded.

Assumption B There exists a constant L such that for any $x, y \in K$,

$$\|g(x) - g(y)\| \leqslant L\|x - y\|. \qquad (2.1)$$

Theorem 2.1[7] Suppose that x_1 is a starting point for which Assumptions A, B hold. Consider the methods (0.1) and (0.5), where t_k is computed by WWP line search, and U_k is

$$r_k \in [-c, 1], \qquad (2.2)$$

where $r_k = \frac{U_k}{U_k^{DY}}$. Then if $g_k \neq 0$ for all $k \geqslant 1$, we have

$$g_k^T d_k \leqslant 0 \qquad (2.3)$$

for all $k \geqslant 1$.

Further, the method converges in the sense that

$$\liminf_{k\to\infty} \|g_k\| = 0. \qquad (2.4)$$

Theorem 2.2 Suppose Assumptions A, B hold, and $\{x_k\}$ is generated by Algorithm 1.1. If $g_k \neq 0$, for all $k \geqslant 1$, (2.3) holds. Further, so does (2.4).

The above theorem 2.2 is the direct result of Theorem 2.1 and (1.7).

Theorem 2.3 Suppose $f(x)$ is continuous, Assumptions A, B hold, and $\{x_k\}$ is generated by the Algorithm 1.1. If $\_{1k} \leqslant \_{3k}$, $u \equiv \sup\{\frac{\_{1k}}{\_{2k}} \frac{e}{1-e} \mid k \geqslant 1\} < 1$, then (2.4) holds.

Proof We consider the case where $U_{k+1}^{PH} = U_{k+1}^{PH} > 0$. From the definition of d_k we have

$$\|d_k\|^2 = \|g_k\|^2 - 2U_k^{PH} g_k^T d_{k-1} + (U_k^{PH})^2\|d_{k-1}\|^2 \leqslant \|g_k\|^2 + 2|U_k^{PH}|\,|g_k^T d_{k-1}| + (U_k^{FR})^2\|d_{k-1}\|^2 \leqslant \|g_k\|^2 + 2\frac{\_{1k}}{\_{2k}}\frac{e}{1-e}\|g_k\|^2 + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_{k-1}\|^2 \leqslant 3\|g_k\|^2 + \|g_k\|^4 \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}$$

for all $k \geqslant 2$

Hence

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leqslant \frac{3}{\|g_k\|^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}. \qquad (2.5)$$

In (1.3), if $U_k^{PH} = 0$, (2.5) holds. Suppose that (2.4) does not hold, then we have a positive constant M, such that for all $k \geqslant 1$,

$$\|g_k\| \geqslant M \qquad (2.6)$$

So follows (2.5), we obtain

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leqslant \frac{3}{M^2} + \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4}.$$

Using the facts that $\frac{\|d_1\|}{\|g_1\|} = 1$, we get

$$\frac{\|d_k\|^2}{\|g_k\|^4} \leqslant \sum_{k-1}^{i=1} \frac{3}{M^2} + \frac{1}{\|g_1\|^2} \leqslant \sum_{i=1}^{k} \frac{3}{M^2} \text{ i.e.}$$

$$\frac{\|g_k\|^4}{\|d_k\|^2} \geqslant \frac{M^2}{3k}.$$

According to the sufficient descent condition, we have $(g_k^T d_k)^2 \geqslant (1 - \frac{\_1k}{\_2k}\frac{e}{1-e})\|g_k\|^4$ from which we can get

$$\|g_k\|^4 \leqslant \frac{1}{1 - \frac{\_1k}{\_2k}\frac{e}{1-e}} (g_k^T d_k)^2 \leqslant$$

$$\frac{1}{1 - \underline{u}} (g_k^T d_k)^2.$$

Hence

$$\frac{1}{1-\underline{u}} \sum_{k=1}^{+\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \geqslant \sum_{k=1}^{+\infty} \frac{\|g_k\|^4}{\|d_k\|^2} \geqslant \sum_{k=1}^{+\infty} \frac{M^2}{3k}$$

which implies that

$$\sum_{k=1}^{+\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = +\infty.$$

It contradicts with the Zoutendijk condition. The proof of the theorem is completed.

Theorems 2.2 and 2.3 show that the present method has global convergence under the relatively weak conditions, which satisfies our hope for the formula of U_k stated in the former section of this article.

3 Numerical experiment

In this section, we will test the following three methods.

PRPSWP: PRP formula with SWP line search, where $W = 0.01$, $e = 0.1$.

PRP^+ SWP: PRP^+ formula with SWP line search, where $W = 0.01$, $e = 0.1$.

PH^+ SWP: PH^+ formula with SWP line search, where $W = 0.01$, $e = 0.1$, $\_{}_1 = 3$, $\_{}_2 = 2$, $\_{}_3 = 1$, $\_{}_4 = 1$.

The problems tested are from Reference [8]. For each tested problem, the termination condition is $\|g(x_k)\| \leqslant 10^{-5}$.

The comparison results of the performance of the three methods are showed in Table 1.

Table 1 Test results for the PRPSWP / PRP^+ SWP / PH^+ SWP methods

Problem	Dim	NI/NF/NG		
		PRPSWP	PRP^+ SWP	PH^+ SWP
ROSE	2	29/502/65	22/394/60	28/405/60
FROTH	2	12/30/20	10/28/20	12/31/23
BADSCP	2	–	41/509/100	39/444/84
BADSCB	2	13/80/22	11/123/22	13/76/22
BEALE	2	9/126/21	9/173/20	16/89/30
JENSAM	2	–	–	13/34/23
HELIX	3	49/255/83	32/265/55	33/129/56
BARD	3	23/98/37	27/152/43	16/86/25
GAUSS	3	4/57/6	4/57/6	4/9/5
MEYER	3	–	–	–
GULF	3	1/2/2	1/2/2	1/2/2
BOX	3	–	–	–
SING	4	199/611/338	49/155/79	72/251/121
WOOD	4	169/1103/302	101/549/195	104/842/178
KOWOSB	4	55/300/94	51/249/79	66/282/108
BD	4	–	–	–
OSB1	5	–	–	–
BIGGS	6	264/875/423	–	68/338/117
OSB2	11	254/1061/418	250/1011/412	1832/7912/3115
WATSON	20	2795/7733/4425	2143/5780/3396	1472/3633/2284
ROSEX	8	23/402/59	25/371/62	29/315/59
	50	31/533/77	24/492/60	25/268/61
	100	28/337/74	35/514/101	30/276/63
SINGX	4	199/611/338	49/155/79	72/251/121
PEN1	2	5/18/12	6/20/14	5/18/12
PEN2	4	12/134/28	12/136/27	11/175/25
	50	613/2795/1063	136/898/282	128/797/245
VARDIM	2	3/9/7	3/9/7	3/9/7
	50	10/52/36	10/52/36	10/52/36
TRIG	3	12/81/24	14/131/25	18/138/31
	50	41/279/72	41/230/72	39/225/74
	100	46/342/87	46/341/85	53/446/101
BV	3	12/25/16	12/25/16	11/21/14
	10	75/241/117	75/241/117	45/90/69
IE	3	5/12/7	5/12/7	5/12/7
	50	6/13/7	5/11/6	5/11/6
	100	6/13/8	6/13/8	5/11/7
	200	6/13/8	6/13/8	5/11/7
	500	6/13/8	6/13/8	6/13/8
TRID	3	10/75/16	13/33/19	15/37/22
	50	26/55/31	26/55/31	28/107/32
	100	30/67/36	30/67/36	29/65/35
	200	30/66/36	30/66/36	31/68/37
BAND	3	9/68/13	10/23/17	7/64/12
	50	18/183/24	16/331/25	19/670/26
	100	18/183/24	16/373/26	19/715/29
	200	19/283/27	17/340/27	19/679/27
LIN	2	1/3/3	1/3/3	1/3/3
	50	1/3/3	1/3/3	1/3/3
	500	1/3/3	1/3/3	1/3/3
	1000	1/3/3	1/3/3	1/3/3
LIN 1	2	1/51/2	1/51/2	1/51/2
	10	1/3/3	1/3/3	1/3/3

Table 1 shows the computation results, where the columns have the following meanings: Problem is the name of the test problem in MATLAB; Dim: the dimension of the problem, NI is the number of iterations, NF is the number of function evaluations, NG is the number of gradient evaluations.

In order to rank the iterative numerical methods, we compute the total numbers of function and gradient evaluations by the following formula

$$N_{total} = NF + m * NG, \qquad (3.1)$$

where m is an integer. According to the results on automatic differentiation[8,9], the value of m can be set to $m = 5$.

Since PRPSWP method is one of the commonly efficient conjugate gradientmethods, we compare PRP+ SWP, PH+ SWP methods with PRPSWP method as follows: for each tested example i, compute the total numbers of function evaluations and gradient evaluations required by the evaluated method j (EM(j)) and PRPSWP method by the formula (3.1), and denote them by $N_{total,i}$(EM(j)) and $N_{total,i}$(PRPSWP); then calculate the ratio

$$r_i(\mathrm{EM}(j)) = \frac{N_{total,i}(\mathrm{EM}(j))}{N_{total,i}(\mathrm{PRPSWP})}. \qquad (3.2)$$

If EM (j_0) does not work for example i_0, we replace r_{i_0}(EM(j_0)) by a positive constant f which is defined as follows

$$f = \max\{r_i(\mathrm{EM}(j)): (i,j) \notin S_1\},$$

where

$S_1 = \{(i,j)$: method j does not work for example $i\}$.

The geometric meaning of these ratios for method EM (j) is

$$r(\mathrm{EM}(j)) = \left(\prod_{i \in S} r_i(\mathrm{EM}(j))\right)^{1/|S|}, \qquad (3.3)$$

where S denotes the set of the test problems and $|S|$ is the number of elements in S. One advantage of the above rule is that, the comparison is relative and hence is not to be dominated by a few problems for which the method requires a great deal of function evaluations and gradient functions.

According to the above rule, it is clear that r (PRPSWP) = 1. The values of r (PRP+ SWP) and r (PH+ SWP) are 0.9049 and 0.7704.

Remark 3.1 It is very clear that when the parameters in (1.2) satisfy _1 = _2 + _3, (1.2) is a conjugate gradient formula in case of exact line search. So we choose _1 = 3, _2 = 2, _3 = 1, _4 = 1 when doing computational experiment. From the computational results in r (PRPSWP) = 1, r (PRP+ SWP) = 0.9049 and r (PH+ SWP) = 0.7704, we can see the superiority of PH+ method over other two methods.

4 Conclusions and future work

In this paper, we discussed a new conjugate gradient formula, which generates by PRP and HS formulas. The new method based on this formula satisfies the sufficient descent condition, and under particular assumptions and with WWP line search, it converges globally.

From our preliminary numerical results, PH+ method not only possesses global convergence, but also performs much better than any other conjugate gradient methods given in the related literatures, such as PRPSWP method and PRP+ SWP method. However, since the numerical results given in this paper are dependent on the choosing of the parameters, it is very necessary to do further study for choosing more suitable parameters to have more numerical results to this algorithm.

References

[1] Powell M J D. Restart procedures of the conjugate gradient method [J]. Math Program, 1977(2): 241-254.

[2] Powell M J D. Nonconvex minimization calculations and the conjugate gradient method [M]//Lecture Notes in Mathematics. Berlin: Springer-Verlag, 1984: 122-141.

[3] Dai Y H. Analyses of conjugate gradient methods [Ph. D. thesis]. Institute of Computational Mathematics and Scientific /Engineering Computing, Chinese Academy of Sciences, 1997 (in Chinese).

[4] Powell M J D. Convergence properties of algorithms for nonlinear optimization [J]. SIAM Review, 1986, 28: 487-500.

[5] Gibert J C, Nocedal J. Global convergence properties of conjugate gradient methods for optimization [J]. SIAM Journal on Optimization, 1992(2): 21-42.

[6] Dai Y H, Yuan Y Y. Nonlinear conjugate gradient methods [M]. Shanghai: Science Press of Shanghai, 2000.

[7] Dai Y H, Yuan Y Y. An efficient hybrid conjugate gradient method for unconstrained optimization [J]. Annals of Operations Research, 2001, 103: 33-47.

[8] Moré J J, Garbow B S, Hillstrome K E. Testing unconstrained optimization software [J]. ACM Trans Math Software, 1981, 7: 17-41.

[9] Griewank A. On automatic differentiation [M]//Iri M, Tanabe K. Mathematical Programming: Recent Developments and Applications. Dordrecht: Kluwer Academic Publishers, 1989: 84-108.

(责任编辑: 邓大玉)