A New Three–Term Conjugate Gradient Method with Descent Direction for Unconstrained Optimization

XiaoLiang Dong[a], HongWei Liu[b], YuBo He[c], Saman Babaie-Kafaki[d] and Reza Ghanbari[e]

[a] *School of Mathematics and Information, Beifang University for Nationalities*
 Yinchuan, 750021, China

[b] *School of Mathematics and Statistics, Xidian University*
 No.2 South Taibai Road, Xian, 710071 Shannxi, China

[c] *Department of Mathematics and Applied Mathematics, Huaihua University*
 No.612 Yingfeng East Road, Huaihua, 418008 Hunan, China

[d] *Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, 35195-363 Semnan, Iran*

[e] *Department of Mathematical Science, Ferdowsi University of Mashhad, Mashhad, Iran*
 E-mail(*corresp.*): `dongxl@nun.edu.cn`

Abstract. In this paper, we propose a three–term PRP–type conjugate gradient method which always satisfies the sufficient descent condition independently of line searches employed. An important property of our method is that its direction is closest to the direction of the Newton method or satisfies conjugacy condition as the iterations evolve. In addition, under mild condition, we prove global convergence properties of the proposed method. Numerical comparison illustrates that our proposed method is efficient for solving the optimization problems.

Keywords: unconstrained optimization, conjugate gradient method, sufficient descent condition, global convergence.

AMS Subject Classification: 49M37; 65K05; 90C53.

1 Introduction

Consider the following unconstrained optimization problem:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function. We denote by $g(x)$ the gradient of f at x and abbreviate $g(x_k)$ and $f(x_k)$ by g_k and f_k, respectively. Also, we use $y_{k-1} = g_k - g_{k-1}$ and $|| \cdot ||$ to stand for the Euclidian norm.

The conjugate gradient (CG) methods are used for large–scale optimization problems, because their memory requirements are modest. Usually, the general procedure of the iterative computational scheme is presented by:

$$x_1 \in \mathbb{R}^n, \quad x_{k+1} = x_k + s_k, \quad s_k = \alpha_k d_k, k = 1, 2, \quad \cdots, \tag{1.1}$$

where $\alpha_k > 0$ is a steplength, and d_k is a search direction generated by

$$d_1 = -g_1, \quad d_k = -g_k + \beta_k d_{k-1}, \quad k = 2, 3, \cdots$$

Several famous formulas for β_k are called the Fletcher–Reeves [15], Hestenes–Stiefel [19], Polak–Ribière–Polyak [24] formulas. We refer to an excellent survey [18] for a review on recent advances in this area. The parameters of the PRP and HS methods are respectively given by

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \quad \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}. \tag{1.2}$$

Both above methods are usually recommended in actual computation due to its superior computational performance. This can be attributed to the property $(*)$, which was dated back to Gilbert and Nocedal [16]. That is, in the corresponding PRP$^+$ method, in which the parameters $\beta_k^{PRP+} = \max\{\beta_k^{PRP}, 0\}$, the current search direction will automatically adjust to the steepest descent direction when the step s_{k-1} is small, which prevents effectively jamming phenomenon from occurring.

Recently, various modifications of the HS method have received growing interests, in which conjugacy condition and sufficient descent condition are two key factors to be considered in designing new methods.

The HS method has the property that it can satisfy automatically the standard conjugacy condition independent of any line search used [19], i.e.,

$$d_k^T y_{k-1} = 0. \tag{1.3}$$

Later, the above "pure" conjugacy condition (1.3) was replaced by the so–called Dai–Liao conjugacy condition [9]:

$$d_k^T y_{k-1} = -\xi g_k^T s_{k-1},$$

where $\xi > 0$ is a constant. Also, the CG parameter proposed by Hager and Zhang [17], that is

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

can be viewed as adaptive version of (1.2) corresponding to $\xi = 2\frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}$.

Often, the sufficient descent condition has been used in the literature to analyze the global convergence of conjugate gradient methods with inexact line searches. That is, there exists a constant $c > 0$ such that

$$d_k^T g_k \le -c\|g_k\|^2, \quad \forall k \in N. \tag{1.4}$$

This property has been paid special attention by many authors, say, Dai and Kou [8] proposed a family of CGOPT methods. To achieve good computational performance, Babaie–Kafaki [3,5] focused on hybridizing different CG methods, in which the conjugacy condition (1.3) and (1.4) are always satisfied. Yu *et al.* [28],Andrei [1, 2],Babaie-Kafaki [4], Dong *et al.* [11, 12, 13, 14], Livieris and Pintelas [20, 21, 22] proposed several modified spectral CG methods.

In this paper, we are interested in the three–term CG method. It is well–known that a good direction to follow is the Newton direction $-\nabla^2 f(x_k)^{-1} g_k$, provided that the current point x_k is close enough to a local minimizer x^*. In the need of solving large–scale problems, computationally, though, it is difficult or even impossible to evaluate or store the Hessian matrix of the function exactly at each iteration. For this reason, various modified secant equations and quasi–Newton algorithms are employed in the sense that they have the attractive property of achieving a high–order accuracy in approximating the second–order curvature information of the objective function.

Based on the analysis above, the extension and application of some properties of quasi–Newton methods to nonlinear CG methods successfully revolutionized the field of nonlinear optimization. For example, Zhang, Zhou and Li proposed some three–term CG methods, which satisfied $d_k^T g_k = -\|g_k\|^2$, $\forall k \in N$. Furthermore, Narushima et al. [23] extended this method to a general three–term conjugate gradient method. Based on a combination of the scaled memoryless BFGS method and the preconditioning technique, Andrei [1, 2] proposed the THREECG and TTCG method.

Inspired by the above papers, we construct another way to generate the search direction and proposed another three–term CG method, which is always sufficiently descent as well as being close to the quasi–Newton direction.

2　Motivation and Properties

In this section, we begin to describe our motivation, and then present a detailed description of our method.

Recently, Zhang, Zhou and Li [29] proposed a three–term MPRP method and establish its global convergence with the Armijo line search, in which

$$d_k^{MPRP} = -g_k + \beta_k^{PRP} d_{k-1} - \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} y_{k-1}.$$

By using the Gram–Schmidt orthogonalization on to d_{k-1} and g_k, Cheng [7] proposed a TMPRP method with the strong Wolfe line search, in which

$$d_k^{TMPRP} = -g_k + \beta_k^{PRP} d_{k-1} - \beta_k^{PRP} \frac{g_k^T d_{k-1}}{\|g_k\|^2} g_k. \qquad (2.1)$$

The structures of the search direction d_k^{MPRP} and d_k^{TMPRP} are to some extent alike in the sense that both of them can be regarded as the variants of the PRP method. Also, a nice property is that, at each iteration, the generated direction in the above two methods satisfied sufficient descent condition

independent of any line search used, that is,

$$d_k^T g_k = -\|g_k\|^2. \tag{2.2}$$

However, the above methods do not satisfy the conjugacy condition and the probably effective combination between them have been ignored, which motivate this paper.

In this paper, we propose another three–term PRP–type method by noting that our chief concern is to take advantage of the property of the Newton method or the conjugacy condition in constructing the CG method.

We first simply merge the directions above into a single direction:

$$d_k^{NPRP} = (1 - \xi_k)d_k^{MPRP} + \xi_k d_k^{TMPRP},$$

where ξ_k is a hybridization parameter.

After some simple algebraic manipulations, we have

$$d_k^{NPRP} = d_k^{MPRP} + \xi_k \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \left(y_{k-1} - \frac{g_k^T y_{k-1}}{\|g_k\|^2} g_k \right), \tag{2.3}$$

or equivalently,

$$d_k^{NPRP} = d_k^{MPRP} + \xi_k \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} T_k y_{k-1},$$

where the orthogonal projection matrix

$$T_k = I - \frac{g_k g_k^T}{\|g_k\|^2}$$

is a symmeric non–negative definite matrix with $\|T_k\| = 1$ and I denotes the identity matrix.

Next, we explain in more detail how to obtain the desired value of ξ_k.

Recall that, in the QN method, an approximation matrix B_{k-1} of the Hessian matrix $\nabla^2 f(x_{k-1})$ is updated such that the new matrix B_k satisfies the quasi–Newton equation

$$B_k s_{k-1} = y_{k-1}, \tag{2.4}$$

which leads to the the quasi–Newton search direction, calculated by

$$d_k^{QN} = -B_k^{-1} g_k.$$

The hybrid CG parameter ξ_k is justified by the following condition:

$$-B_k^{-1} g_k = d_k^{MPRP} - \xi_k \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \left(\frac{g_k^T y_{k-1}}{\|g_k\|^2} g_k - y_{k-1} \right),$$

or equivalently,

$$-g_k = B_k d_k^{MPRP} - \xi_k B_k \frac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \left(\frac{g_k^T y_{k-1}}{\|g_k\|^2} g_k - y_{k-1} \right).$$

Note that $||g_k||^2 = -d_k^T g_k$. Pre-multiplying both sides of s_{k-1} and using the effective standard secant equation (2.4) give the following relationship, that is

$$-g_k^T s_{k-1} = - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T y_{k-1} + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} ||y_{k-1}||^2$$
$$+\xi_k \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \left(\frac{(g_k^T y_{k-1})^2}{||g_k||^2} - ||y_{k-1}||^2 \right). \qquad (2.5)$$

Obviously, Equation (2.5) is satisfied trivially for $g_k^T d_{k-1} = 0$. We next consider the case where $g_k^T d_{k-1} \neq 0$.

By some simple algebraic manipulations, we have from (2.5) that

$$\xi_k = \frac{||y_{k-1}||^2 + s_{k-1}^T g_{k-1} - g_k^T y_{k-1}}{||y_{k-1}||^2 - \left(\frac{g_k^T y_{k-1}}{||g_k||} \right)^2}. \qquad (2.6)$$

Notice first that ξ_k given by (2.6) is well–defined when $||y_{k-1}|| \neq \frac{|g_k^T y_{k-1}|}{||g_k||}$. We expect to take advantage of the fast local convergence of the Newton method in at least the final iterations, but the PRP method will have to be modified in order to converge globally. Having the view that the PRP$^+$ method has the property $\beta_k^{PRP} > 0$ ($\beta_k = 0$ can be regard as its truncation form), which is equivalent to the inequality $g_k^T y_{k-1} > 0$. To eliminate the probable effect of unboundedness of ξ_k and establish the global convergence of our proposed method, we proposed the following strategy in constructing the search direction:

$$d_k = \begin{cases} -g_k, & k \notin K, \\ d_k^{NPRP}, & k \in K. \end{cases} \qquad (2.7)$$

In (2.7), the index set K is defined by

$$K = \{k | k \in N, \ 0 < \zeta_k < 1 - \eta\}, \qquad (2.8)$$

where $\eta \in (0, 1)$ is a given constant and

$$\zeta_k = \frac{g_k^T y_{k-1}}{||g_k^T|| ||y_{k-1}||}.$$

Remark 1. The condition (2.5) was originated form Perry [25], that is, $d_k^T y_{k-1} = d_k^T (B_k s_{k-1}) = (B_k d_k)^T s_{k-1} = -g_k^T s_{k-1}$. Without the loss of generality, instead of using the above condition, we choose the general form of the conjugacy condition, that is, $d_k^T y_{k-1} = -t g_k^T s_{k-1}$. Obviously, if $t = 0$, then the above condition reduces to the usual conjugacy condition, otherwise, it becomes the DL conjugacy condition.

With the change of conjugacy condition comes a change of the choice of the parameter ξ_k in (2.6), given by

$$\widehat{\xi_k} = \frac{||y_{k-1}||^2 + ts_{k-1}^T g_{k-1} - g_k^T y_{k-1}}{||y_{k-1}||^2 - \left(\frac{g_k^T y_{k-1}}{||g_k||}\right)^2}, \tag{2.9}$$

where $t \geq 0$ is a constant.

For convenience, we call our method (2.3) and (2.9) as NPRP method in later part of this paper, and state the steps of this method as follows.

Algorithm 2.1   (NPRP method)

Step 0. Chosen positive constants $\rho < \sigma < 1$, t, ε and η . Select $x_1 \in R^n$, set $k = 1$.

Step 1. Test a criterion for stopping the iterations. If $||g_k|| < \varepsilon$, then stop, otherwise calculate d_k by (2.7) with the parameters $\widehat{\xi_k}$ defined by (2.9).

Step 2. Determine the steplength α_k by the strong Wolfe conditions, i.e.,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k g_k^T d_k, \quad |g\left(x_k + \alpha_k d_k\right)^T d_k| \leq -\sigma g_k^T d_k. \tag{2.10}$$

Step 3. Set $x_{k+1} = x_k + \alpha_k d_k$ and $k = k + 1$, return to Step 1.

3   Convergence Analysis

In this section, we prove the global convergence of the proposed method. For general nonlinear functions, similar to [16] and [17], we can obtain a weaker global convergence in the sense that $\liminf_{k \to \infty} ||g_k|| = 0$. To this end, we proceed by contradiction that, there exists a constant $\epsilon > 0$ such that

$$||g_k|| \geq \epsilon, \quad for \quad all \quad k \in N; \tag{3.1}$$

otherwise a stationary point has been found. We also assume that the objective function always satisfies the following assumptions.

Assumption 3.1

–Boundedness Assumption: The level set, defined by $\Omega = \{x \in R^n | f(x) \leq f(x_1)\}$ is bounded; that is, there exists a constant $B > 0$ such that

$$||x|| \leq B, \forall x \in \Omega. \tag{3.2}$$

–Lipschitz Assumption: In some neighborhood Ω_0 of Ω, f is continuously differentiable, and g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$||g(x) - g(y)|| \leq L||x - y||, \forall x, y \in \Omega_0.$$

Clearly, there exists a constant $\gamma > 0$ such that $||g(x)|| \leq \gamma$, for all $x \in \Omega$.

The following Zoutendijk condition is often used to prove convergence of CG method.

Theorem 1. *[27] Suppose that Assumptions 3.1 hold. Consider any iterative method of the form* (1.1) *where d_k is a descent direction and α_k satisfies the Wolfe conditions. Consider the general conjugate gradient method, we have $\sum_{k \geq 1} (g_k^T d_k)^2 / \|d_k\|^2 < +\infty$.*

Subsequently, analogous to [16], we state some properties for the direction d_k, formula β_k and step s_k.

Theorem 2. *Suppose that Assumptions 3.1 and* (3.1) *hold. Let $\{x_k\}$ be generated by Algorithm 2.1. Then, there are two positive constants C_1 and M such that*

$$|\beta_k^{PRP}| \leq C_1 \|s_{k-1}\| \tag{3.3}$$

$$\|R_k\| \leq M, \tag{3.4}$$

where

$$R_k = \begin{cases} -g_k, & k \notin K, \\ -g_k - \dfrac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} y_{k-1} + \widehat{\xi}_k \dfrac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} \left(y_{k-1} - \dfrac{g_k^T y_{k-1}}{\|g_k\|^2} g_k \right), & k \in K. \end{cases} \tag{3.5}$$

Proof. Clearly, it suffices to show the case where $0 < \zeta_k \leq 1 - \eta$ is satisfied. We first estimate the upper bound for $\left| g_k^T d_{k-1} / d_{k-1}^T y_{k-1} \right|$. Observe that

$$g_k^T d_{k-1} = d_{k-1}^T y_{k-1} + g_{k-1}^T d_{k-1} < d_{k-1}^T y_{k-1}.$$

On the other hand, we have from the Wolfe condition (2.10) that

$$g_k^T d_{k-1} > \sigma g_{k-1}^T d_{k-1} = \sigma d_{k-1}^T y_{k-1} + \sigma g_k^T d_{k-1}$$

and therefore $g_k^T d_{k-1} > -\frac{\sigma}{1-\sigma} d_{k-1}^T y_{k-1}$ holds. Combining the above upper and lower bound for $g_k^T d_{k-1}$ yields

$$|g_k^T d_{k-1}| \leq c_\sigma d_{k-1}^T y_{k-1}, \tag{3.6}$$

where $c_\sigma = \max \{1, \sigma/(1-\sigma)\}$. The above inequations (3.6) gives

$$|g_{k-1}^T s_{k-1}| \leq y_{k-1}^T s_{k-1} + |g_k^T s_{k-1}| \leq (1 + c_\sigma) s_{k-1}^T y_{k-1} \leq 2B(1 + c_\sigma)\|y_{k-1}\|$$

and

$$|\widehat{\xi}_k| \cdot \|y_{k-1}\| \leq \frac{\|y_{k-1}\|^2 + t|s_{k-1}^T g_{k-1}| + |g_k^T y_{k-1}|}{\|y_{k-1}\|^2 - \left(g_k^T y_{k-1}/\|g_k\|\right)^2} \|y_{k-1}\|$$

$$\leq \frac{\|y_{k-1}\| + 2tB(1 + c_\sigma) + \|g_k\|}{1 - \left(g_k^T y_{k-1}/(\|g_k\| \, \|y_{k-1}\|)\right)^2} \leq \frac{2tB(1 + c_\sigma) + 3\gamma}{1 - (1-\eta)^2}.$$

Setting $c_\lambda = \dfrac{2tB(1 + c_\sigma) + 3\gamma}{1 - (1-\eta)^2}$, we have $|\widehat{\xi}_k| \cdot \|y_{k-1}\| \leq c_\lambda$.

The above inequality together with the strong Wolfe condition (2.10) and Equations (2.2) gives that

$$
\begin{aligned}
||R_k|| &\leq ||g_k|| + \frac{|g_k^T d_{k-1}|}{||g_{k-1}||^2}||y_{k-1}|| + |\widehat{\xi_k}|\frac{g_k^T d_{k-1}}{||g_{k-1}||^2}\left(y_{k-1} - \frac{g_k^T y_{k-1}}{||g_k||^2}g_k\right) \\
&= ||g_k|| + \sigma||y_{k-1}|| + \sigma|\widehat{\xi_k}| \cdot ||T_k|| \cdot ||y_{k-1}|| \\
&\leq ||g_k|| + \sigma L||s_{k-1}|| + \sigma c_\lambda L||s_{k-1}||.
\end{aligned}
$$

Setting $M = \gamma + 2(1 + c_\lambda)\sigma c_\lambda LB$, we get $||R_k|| \leq M$.

On the other hand, we obtain from (2.10) that

$$
y_{k-1}^T d_{k-1} > -(1 - \sigma)g_{k-1}^T d_{k-1} = (1 - \sigma)||g_{k-1}||^2 > (1 - \sigma)\epsilon^2.
$$

We now insert the lower bound for $y_{k-1}^T d_{k-1}$ and the Lipschitz estimate (2.10) for y_{k-1} into β_k^{PRP} to get:

$$
|\beta_k^{PRP}| \leq \frac{||g_k||||y_{k-1}||}{||g_{k-1}||^2} \leq \frac{L||g_k||||s_{k-1}||}{\epsilon^2}.
$$

The conclusion (3.3) is seen to hold by letting $C_1 = L\gamma/\epsilon^2$.   □

The next theorem will show that, asymptotically, the search direction changes slowly. Similar to [16], it suffices to establish a bound for the change $u_k - u_{k-1}$ in the normalized direction $u_k = d_k/||d_k||$.

Theorem 3. *Suppose that Assumptions 3.1 and (3.1) hold. Let $\{x_k\}$ be generated by Algorithm 2.1. Then, we have $d_k \neq 0$ and $\sum\limits_{k=1}^{+\infty} ||u_k - u_{k-1}||^2 < \infty$.*

Proof.   It's clear that $||d_k|| \geq ||g_k|| \neq 0$ and therefore u_k is well defined. For convenience, set $r_k = R_k/||d_k||$ and

$$
\delta_k = \begin{cases} 0, & k \notin K, \\ \beta_k^{PRP}||d_{k-1}||/||d_k||, & k \in K. \end{cases} \tag{3.7}
$$

By the definition R_k in (3.5) and δ_k in (3.7), we get $u_k = r_k + \delta_k u_{k-1}$. Using this relation with the identity $||u_k|| = ||u_{k-1}|| = 1$, we have that

$$
||r_k|| = ||u_k - \delta_k u_{k-1}|| = ||u_{k-1} - \delta_k u_k||.
$$

The above equality, together with $\delta_k \geq 0$ implies that

$$
\begin{aligned}
||u_k - u_{k-1}|| &\leq (1 + \delta_k)||u_k - u_{k-1}|| \\
&\leq ||u_k - \delta_k u_{k-1}|| + ||u_{k-1} - \delta_k u_k|| = 2||r_k||. \tag{3.8}
\end{aligned}
$$

We have from (3.4) and the Zoutendijk condition that

$$
\begin{aligned}
\sum_{k\geq 1} ||r_k||^2 &= \sum_{k\geq 1}\frac{||R_k||^2}{||d_k||^2} \leq \sum_{k\geq 1}\frac{M^2}{||d_k||^2} \\
&= \sum_{k\geq 1}\frac{M^2}{||g_k||^4}\frac{||g_k||^4}{||d_k||^2} \leq \frac{M^2}{\epsilon^4}\sum_{k\geq 1}\frac{(g_k^T d_k)^2}{||d_k||^2} < \infty. \tag{3.9}
\end{aligned}
$$

Thus, the conclusion follows from (3.8) and (3.9). □

Note that the parameter β_k^{PRP} have the property $(*)$, firstly proposed by Gilbert and Nocedal in [16].

Theorem 4. *Suppose that Assumptions 4.1 hold. Let $\{x_k\}$ be generated by Algorithm 2.1. Then, there exist constants $b > 1$ and $\lambda > 0$ such that*

$$|\beta_k^{PRP}| \leq b, \quad and \quad ||s_{k-1}|| \leq \lambda \Longrightarrow |\beta_k^{PRP}| \leq 1/b.$$

Subsequently, using the Theorems 2, 3 and 4, we establish the global convergence of Algorithm 2.1 for the general functions.

Theorem 5. *Suppose that Assumptions 3.1 hold. Let $\{x_k\}$ be generated by Algorithm 2.1. Then the method converges in the sense that $\liminf\limits_{k\to\infty} ||g_k|| = 0$ holds.*

Proof. The proof is divided into the following two steps.

Step 1: A bound on the steps s_k. Let $\triangle \in N$, chosen large enough such that $\triangle \geq 4BC_1$, where B and C_1 are defined in (3.2) and (3.3), respectively. For any $l > k \geq k_0$ with $l - k \leq \triangle$, proceeding the same proof as the case two of Theorem 3.2 in [16] we get $\sum\limits_{j=k}^{l-1} ||s_k|| \leq 2B$.

Step 2: A bound on the directions d_k. From the (3.3), we obtain that

$$\|d_k^{NPRP}\|^2 \leq \left(||R_k|| + |\beta_k^{PRP}| \, ||d_{k-1}|| \right)^2$$
$$\leq (M + C_1||s_{k-1}|| \, ||d_{k-1}||||)^2 \leq 2M^2 + 2 \left(C_1||s_{k-1}|| \, ||d_{k-1}|| \right)^2 .$$

Now, the remaining argument is similar to the case (iii) in Theorem 2 in [17], thus we omit it. This completes the proof. □

4 Numerical Experiments

In this section, we report numerical results on a set of 145 nonlinear unconstrained optimization test problems on the CUTEr collection [6], with default dimensions as presented in Hager's homepage:
"http://www.math.ufl.edu/hager/", and also listed in the Table 1.

The following CG methods are test, the involved parameters are listed:

- 1. The HZ method [17]: The CG method with the truncation of
  $$\beta_k^{HZ+} = \max\left\{ \beta_k^{HZ}, \frac{-1}{||d_{k-1}|| \min\{\eta, ||g_{k-1}||\}} \right\}, \text{ where } \eta = 0.1.$$

- 2. The TMPRP method [7]: The CG method with the search direction defined by (2.1).

- 3. The NPRP method (Algorithm 2.1): The CG method with the direction defined by (2.7), where $\eta = 10^{-5}$ in (2.8).

Table 1. Specifications of the test functions

| Name	Dim	Name	Dim	Name	Dim	Name	Dim
AKIVA	2	ALLINITU	4	ARGLINA	200	ARGLINB	200
ARWHEAD	5000	BARD	3	BDQRTIC	5000	BEALE	2
BIGGS6	6	BOX3	3	BOX	10000	BRKMCC	2
BROWNAL	200	BROWNBS	2	BROWNDEN	4	BROYDN7D	5000
BRYBND	5000	CHAINWOO	4000	CHNROSNB	50	CLIFF	2
COSINE	10000	CRAGGLVY	5000	CUBE	2	CURLY10	10000
CURLY20	10000	CURLY30	10000	DECONVU	63	DENSCHNA	2
DENSCHNB	2	DENSCHNC	2	DENSCHND	3	DENSCHNE	3
DENSCHNF	2	DIXMAANA	3000	DIXMAANB	3000	DIXMAANC	3000
DIXMAAND	3000	DIXMAANE	3000	DIXMAANF	3000	DIXMAANG	3000
DIXMAANH	3000	DIXMAANI	3000	DIXMAANJ	3000	DIXMAANK	3000
DIXMAANL	3000	DIXON3DQ	10000	DJTL	2	DQDRTIC	5000
DQRTIC	5000	EDENSCH	2000	EG2	1000	EIGENALS	2550
EIGENBLS	2550	EIGENCLS	2652	ENGVAL1	5000	ENGVAL2	3
ERRINROS	50	EXPFIT	2	EXTROSNB	1000	FLETCBV2	5000
FLETCHCR	1000	FMINSRF2	5625	FMINSURF	5625	FREUROTH	5000
GENHUMPS	5000	GENROSE	500	GROWTHLS	3	GULF	3
HAIRY	2	HATFLDD	3	HATFLDE	3	HATFLDFL	3
HEART6LS	6	HEART8LS	8	HELIX	3	HIELOW	3
HILBERTA	2	HILBERTB	10	HIMMELBB	2	HIMMELBF	4
HIMMELBG	2	HIMMELBH	2	HUMPS	2	JENSMP	2
JIMACK	3549	KOWOSB	4	LIARWHD	5000	LOGHAIRY	2
MANCINO	100	MARATOSB	2	MEXHAT	2	MOREBV	5000
MSQRTALS	1024	MSQRTBLS	1024	NCB20B	5000	NCB20	5010
NONCVXU2	5000	NONDIA	5000	NONDQUAR	5000	OSBORNEA	5
OSBORNEB	11	OSCIPATH	10	PALMER1C	8	PALMER1D	7
PALMER2C	8	PALMER3C	8	PALMER4C	8	PALMER5C	6
PALMER6C	8	PALMER7C	8	PALMER8C	8	PARKCH	15
PENALTY1	1000	PENALTY2	200	PENALTY3	200	POWELLSG	5000
POWER	10000	QUARTC	5000	ROSENBR	2	S308	2
SCHMVETT	5000	SENSORS	100	SINEVAL	2	SINQUAD	5000
SISSER	2	SNAIL	2	SPARSINE	5000	SPARSQUR	10000
SPMSRTLS	4999	SROSENBR	5000	STRATEC	10	TESTQUAD	5000
TOINTGOR	50	TOINTGSS	5000	TOINTPSP	50	TOINTQOR	50
TQUARTIC	5000	TRIDIA	5000	VARDIM	200	VAREIGVL	50
VIBRBEAM	8	WATSON	12	WOODS	4000	YFITU	3
ZANGWIL2	2						

For the methods of TMPRP and NPRP, similar to the CG_Descent algorithm, we used the approximate Wolfe conditions proposed by Hager and Zhang (2005) in the line search procedure, with the same parameter values as considered in CG_Descent version 5.3. We stop the iterations if the inequality $||g_k||_\infty \leq 10^{-6}$ is satisfied. The implementations are run on a mobile computer with 3.2 GHz of CPU, 1 GB of RAM and Centos 6.2 server Linux operation system.

The detailed numerical results, including the CPU time in seconds and the number of iterations, the number of function evaluations, and gradient evaluations implementation for each of the tested method, can be found in [26].

It should be pointed out that inspired by Livieris and Pintelas [20, 21, 22], we consider to use the different values $t \in \{0.1k\}_{k=0}^{10}$ in actual computation. We obtained from the data that Algorithm 2.1 with the optimal choice for the parameter $t = 0.8$, performed slightly better than others and are correspondingly named NPRP method on the website above. Also, the item NFG is denoted by $NFG = NF + 3NG$, where NF and NG denote the the number of function evaluations and gradient evaluations, respectively.

Consequently, efficiency comparisons are made, using the performance profiles of Dolan and Moré [10], which will present a wealth of information including efficiency and robustness. More analytically, the left side of the figure presents the percentage of test problem for which a method performs fastest, the right side gives the percentage of the test problems that are successfully solved. The top curve is the method that solved the most problems in a time

that was within a factor ω of the best time.
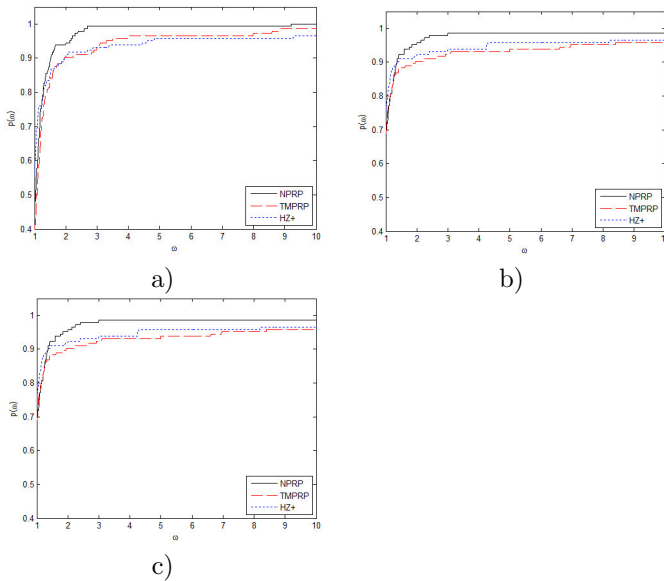
Figure 1. Performance profiles with respect to: a) the number of iteration, b) the total CPU-time, c) the total number of function and gradient evaluations

As shown by the figures,the NPRP method slightly outperforms the TM-PRP and HZ+ methods in the perspectives of iteration numbers, the total number of function and gradient evaluations, and the CPU time. Since the Figures 1 graphically illustrate that the curve of "NPRP" is always the top performer for almost all values of ω, it seems to conclude that the NPRP method are effective for solving these test problem.

5 Conclusions

In this paper, we propose a three–term conjugate gradient method which always satisfy the sufficient descent condition independently of line searches employed. The corresponding search direction is closest to the direction of the Newton method or satisfies conjugacy condition as the iterations evolve. Global convergence of the methods has been briefly discussed. Numerical results show that the proposed method is efficient for solving the optimization problems.

Acknowledgements

References

[1] N. Andrei. On three-term conjugate gradient algorithms for unconstrained optimization. *Applied Mathematics and Computation*, **219**(11):6316–6327, 2013. http://dx.doi.org/10.1016/j.amc.2012.11.097.

[2] N. Andrei. A simple three-term conjugate gradient algorithm for unconstrained optimization. *Journal of Computational and Applied Mathematics*, **241**:19–29, 2013. http://dx.doi.org/10.1016/j.cam.2012.10.002.

[3] S. Babaie-Kafaki and R. Ghanbari. A descent family of Dai-Liao conjugate gradient methods. *Optimization Methods and Software*, **29**(3):583–591, 2014. http://dx.doi.org/10.1080/10556788.2013.833199.

[4] S. Babaie-Kafaki and R. Ghanbari. A hybridization of the Polak-Ribiẽre-Polyak and Fletcher-Reeves conjugate gradient methods. *Numerical Algorithms*, **68**(3):481–495, 2014. http://dx.doi.org/10.1007/s11075-014-9856-6.

[5] S. Babaie-Kafaki and N. Mahdavi-Amiri. Two modified hybrid conjugate gradient methods based on a hybrid secant equation. *Mathematical Modelling and Analysis*, **18**(1):32–52, 2013. http://dx.doi.org/10.3846/13926292.2013.756832.

[6] I. Bongartz, A.R. Conn, N. Gould and Ph.L. Toint. CUTE: Constrained and unconstrained testing environment. *ACM Transactions on Mathematical Software*, **21**(1):123–160, 1995. http://dx.doi.org/10.1145/200979.201043.

[7] W. Cheng. A two-term PRP-based descent method. *Numerical Functional Analysis and Optimization*, **28**(11-12):1217–1230, 2007. http://dx.doi.org/10.1080/01630560701749524.

[8] Y.H. Dai and C.X. Kou. A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search. *SIAM Journal on Optimization*, **23**(1):296–320, 2013. http://dx.doi.org/10.1137/100813026.

[9] Y.H. Dai and L.Z. Liao. New conjugacy conditions and related nonlinear conjugate gradient methods. *Applied Mathematics and Optimization*, **43**(1):87–101, 2001. http://dx.doi.org/10.1007/s002450010019.

[10] D.E. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, **91**(2):201–213, 2002. http://dx.doi.org/10.1007/s101070100263.

[11] X.L. Dong, H. Liu, Y.L. Xu and X.M. Yang. Some nonlinear conjugate gradient methods with sufficient descent condition and global convergence. *Optimization Letters*, **9**(7):1421–1432, 2014. http://dx.doi.org/10.1007/s11590-014-0836-5.

[12] X.L. Dong, H.W. Liu and Y.B. He. A self-adjusting conjugate gradient method with sufficient descent condition and conjugacy condition. *Journal of Optimization Theory and Applications*, **165**(1):225–241, 2014. http://dx.doi.org/10.1007/s10957-014-0601-z.

[13] X.L. Dong, H.W. Liu and Y.B. He. New version of the three-term conjugate gradient method based on spectral scaling conjugacy condition that generates descent search direction. *Applied Mathematics and Computation*, **269**:606–617, 2015. http://dx.doi.org/10.1016/j.amc.2015.07.067.

[14] X.L. Dong, H.W. Liu, Y.B. He and X.M. Yang. A modified Hestenes-Stiefel conjugate gradient method with sufficient descent condition and conjugacy condition. *Journal of Computational and Applied Mathematics*, **281**:239–249, 2015. http://dx.doi.org/10.1016/j.cam.2014.11.058.

[15] R. Fletcher and C.M. Reeves. Function minimization by conjugate gradients. *The Computer Journal*, **7**(2):149–154, 1964. http://dx.doi.org/10.1093/comjnl/7.2.149.

[16] J.Ch. Gilbert and J. Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*, **2**(1):21–42, 1992. http://dx.doi.org/10.1137/0802003.

[17] W.W. Hager and H. Zhang. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization*, **16**(1):170–192, 2005. http://dx.doi.org/10.1137/030601880.

[18] W.W. Hager and H. Zhang. A survey of nonlinear conjugate gradient methods. *Pacific journal of Optimization*, **2**(1):35–58, 2006.

[19] M.R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, **49**(6):409–436, 1952.

[20] I.E. Livieris and P. Pintelas. A descent Dai-Liao conjugate gradient method based on a modified secant equation and its global convergence. *ISRN Computational Mathematics*, **2012**, 2012. http://dx.doi.org/10.5402/2012/435495.

[21] I.E. Livieris and P. Pintelas. A new class of spectral conjugate gradient methods based on a modified secant equation for unconstrained optimization. *Journal of Computational and Applied Mathematics*, **239**:396–405, 2013. http://dx.doi.org/10.1016/j.cam.2012.09.007.

[22] I.E. Livieris and P. Pintelas. A modified Perry conjugate gradient method and its global convergence. *Optimization Letters*, **9**(5):999–1015, 2015. http://dx.doi.org/10.1007/s11590-014-0820-0.

[23] Y. Narushima, H. Yabe and J.A. Ford. A three-term conjugate gradient method with sufficient descent property for unconstrained optimization. *SIAM Journal on Optimization*, **21**(1):212–230, 2011. http://dx.doi.org/10.1137/080743573.

[24] B.T. Polyak. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, **9**(4):94–112, 1969. http://dx.doi.org/10.1016/0041-5553(69)90035-4.

[25] D.F. Shanno. On the convergence of a new conjugate gradient algorithm. *SIAM Journal on Numerical Analysis*, **15**(6):1247–1257, 1978. http://dx.doi.org/10.1137/0715085.

[26] Sohu.com. Numerical results on NPRP/TMPRP/HZ. Available from Internet: http://shuxueyou.blog.sohu.com/310274007.html.

[27] Ph. Wolfe. Convergence conditions for ascent methods. *SIAM review*, **11**(2):226–235, 1969. http://dx.doi.org/10.1137/1011036.

[28] G. Yu, L. Guan and W. Chen. Spectral conjugate gradient methods with sufficient descent property for large-scale unconstrained optimization. *Optimization Methods and Software*, **23**(2):275–293, 2008. http://dx.doi.org/10.1080/10556780701661344.

[29] L. Zhang, W. Zhou and D.H. Li. A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence. *IMA Journal of Numerical Analysis*, **26**(4):629–640, 2006. http://dx.doi.org/10.1093/imanum/drl016.