Control and Optimization in Applied Mathematics (COAM)
Vol. 3, No. 1, Spring-Summer 2018(27-43), ©2016 Payame Noor University, Iran

A New Hybrid Conjugate Gradient Method Based on Eigenvalue Analysis for Unconstrained Optimization Problems

F. Rahpeymaii [1,*], M. Rostami[2]

[1] Department of Mathematics, Payame Noor University,
P.O. BOX. 19395-3697, Tehran, Iran
[2] Young Researchers and Elite Club, Hamedan Branch,
Islamic Azad University, Hamedan, Iran

Received: February 18, 2019; **Accepted:** September 14, 2019.

Abstract. In this paper, two extended three-term conjugate gradient methods based on the Liu-Storey (LS) conjugate gradient method are presented to solve unconstrained optimization problems. A remarkable property of the proposed methods is that the search direction always satisfies the sufficient descent condition independent of line search method, based on eigenvalue analysis. The global convergence of proposed algorithms is established under suitable conditions. Preliminary numerical results show that the proposed methods are efficient and robust to solve the unconstrained optimization problems..

Keywords. Unconstrained optimization, Conjugate gradient methods, Eigenvalue analysis, Global convergence, Numerical comparisons.

MSC. 65K05; 90C26; 15A18; 90C30.

* Corresponding author
rahpeyma_83@yahoo.com, majid403rostami@yahoo.com
http://mathco.journals.pnu.ac.ir

1 Introduction

Consider the unconstrained optimization problem

$$\min \ f(x), \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function and bounded from below, whose gradient is available. Let $g_k = g(x_k) := \nabla f(x_k)$ be the gradient of f at x_k and $\|\cdot\|$ be as the Euclidian norm. Some iterative methods such as the Newton method [24], the quasi-Newton methods [13, 14], the trust-region methods [9, 18, 29, 38] and the conjugate gradient methods [5, 12, 19, 22] have been used to solve (1). In an iterative method, by starting from initial point $x_0 \in \mathbb{R}^n$, the sequence $\{x_k\}_{k\geq 0}$ is generated using the formula

$$x_{k+1} := x_k + \alpha_k d_k, \tag{2}$$

in which the step-size α_k is obtained by an inexact monotone or a nonmonotone line search [15, 24, 34]. In addition, d_k is the search direction, satisfying the descent condition $g_k^T d_k < 0$, or the sufficient descent condition

$$g_k^T d_k < -c\|g_k\|^2, \tag{3}$$

where $c > 0$ is the constant.

Conjugate gradient (CG) methods are a well-known class of iterative methods to solve the unconstrained optimization problems. Since CG methods have low memory requirement and simple computational scheme, they are suitable especially for large-scale problems. CG methods have strong local and global convergence properties which are investigated in many papers, see for example [16, 25]. In the CG methods, the search direction d_k is

$$d_k := \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \geq 1, \end{cases} \tag{4}$$

where β_k is called the conjugate gradient parameter. We can obtain the various CG methods with distinct choices of parameter β_k. Some prominent CG methods are called Fletcher-Reeves (FR) [12], Dai-Yuan (DY) [5], Liu-Storey (LS) [22], the conjugate descent method of Fletcher (CD) [11], Hestenes and Stiefel (HS) [17] and Polak and Ribière [25] and Polyak [26] (PRP), respectively, which are as follows:

$$\beta_k^{FR} := \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \beta_k^{DY} := \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{LS} := -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}, \tag{5}$$

$$\beta_k^{CD} := -\frac{\|g_k\|^2}{g_k^T d_{k-1}}, \qquad \beta_k^{HS} := \frac{y_{k-1}^T g_k}{g_{k-1}^T s_{k-1}}, \qquad \beta_k^{PRP} := \frac{y_{k-1}^T g_k}{\|g_{k-1}\|^2}, \tag{6}$$

where $y_{k-1} := g_k - g_{k-1}$ and $s_{k-1} := x_k - x_{k-1}$. The exact step-size α_k can be obtained by solving the one-dimensional minimization problem

$$\min_{\alpha>0} \ f(x_k + \alpha d_k).$$

Remark 1. The exact line search to obtain the step-size α_k implies that $d_k^T g_{k+1} = 0$. Hence,

(i) $-g_{k-1}^T d_{k-1} = -g_{k-1}^T(-g_{k-1} + \beta_{k-1} d_{k-2}) = \|g_{k-1}\|^2$.

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

(iii) Since the gradients are mutually orthogonal in `CG` methods, we get

$$g_k^T y_{k-1} = g_k^T g_k - g_k^T g_{k-1} = \|g_k\|^2.$$

Based on Remark 1, for strictly convex quadratic function with the exact line search, the `CG` parameters in (5) and (6) are equivalent [24]. In this paper, we use the nonmonotone Armijo-type line search [1] to obtain the inexact step-size α_k, which sufficiently employs the current value of the objective function $f(x)$ as

$$f(x_k + \alpha_k d_k) \leq R_k + \rho\, \alpha_k g_k^T d_k, \tag{7}$$

in which $0 < \rho < 1$ and

$$R_k := \eta_k f_{l(k)} + (1 - \eta_k) f_k, \tag{8}$$

where $f_k := f(x_k)$, $\eta_k \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0, 1)$, $\eta_{\max} \in [\eta_{\min}, 1]$ and

$$f_{l(k)} := \max_{0 \leq j \leq n(k)} \{f_{k-j}\}, \quad k \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}, \tag{9}$$

in which

$$n(k) := \begin{cases} 0, & k = 0, \\ \min\{n(k-1) + 1, N\}, & k \geq 1, \end{cases}$$

with $N \geq 0$.

Liu and Story [22] introduced the parameter β_k^{LS} for which the global convergence of the `LS` method with Grippo-Lucidi line search stablished in [20]. Some researchers have studied for variants of the `LS` methods [21, 31, 33]. In [33], Zhang proposed a `MLS+` method and proved that `MLS+`, independent of the line search, can always generate the descent directions satisfying the following sufficient descent condition

$$g_k^T d_k \leq -\left(1 - \frac{1}{t}\right) \|g_k\|^2,$$

where $t > 1$. Li and Feng [21] improved the `MLS+` method to obtain well-defined modified `MLS+` method to generate the sufficient descent directions independent of the line search. They proved that the modified `MLS+` method is globally convergent with the strong Wolfe line search. A hybridization of known `LS-CD` conjugate gradient algorithms presented to solve unconstrained optimization problem in [31]. In addition, the Wolfe-type line search can guarantee the global convergence of the `LS-CD` conjugate gradient method.

For the first time, Beale [3] proposed a three-term conjugate gradient method whose the search direction d_k has the form

$$d_k := -g_k + \beta_k d_{k-1} + \gamma_k d_t, \tag{10}$$

where γ_k and d_t are the scaler and restart direction, respectively. In general, the three-term conjugate gradient algorithms are numerically strong, efficient, reliable, and robust compared with two-term conjugate gradient algorithms [2, 36]. Recently, some researchers have focused on the three-term conjugate gradient methods, generating a descent search direction [2, ?]. Moreover, the general class of three-term conjugate gradient methods are presented in [23], satisfying the sufficient descent condition.

Some researchers used the CG methods for solving the eigenvalue problems of a symmetric matrix [8, 32]. They introduced the unconstrained optimization problems and obtained some variational characterizations for the minimum and maximum eigenvalues. Hence, we introduce two modified three-term conjugate gradient algorithms based on the LS conjugate gradient method to solve unconstrained optimization problems. We will obtain the search directions which satisfy the sufficient descent condition using both quasi-Newton method and eigenvalues analysis. The global convergence of the new algorithms will be investigated. We give some numerical examples to show the efficiency our algorithms in comparison with several CG algorithms based on LS method.

The rest of this paper is organized as follows. In Section 2, we introduce two new three-term conjugate gradient algorithms based on LS conjugate gradient method. In the next section, the global convergence of proposed algorithms will be established under mild assumptions. In Section 4, numerical results are reported. Finally, some conclusions are given in Section 5.

Remark 2. Let $A \in \mathbb{R}^{n\times n}$ be a square matrix. $\det(A)$ and $\mathrm{tr}(A)$ stand for the determinant and trace of A [28].

(a) $\|A\|_F^2 = \mathrm{tr}(A^TA)$ in which $\|\cdot\|_F$ denotes Frobenius norm, i.e.,

$$\|A\|_F := \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}.$$

(b) If $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of matrix A, then

$$\mathrm{tr}(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n,$$

and

$$\mathrm{tr}(A^2) = \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_n^2.$$

2 Motivation and Properties

In this section, we present a new three-term conjugate gradient method to solve unconstrained optimization problems based on LS conjugate gradient method to obtain a descent search direction. Then, we will modify the new three-term conjugate gradient method using the eigenvalues analysis to improve the efficiency of numerical results. Liu and Storey [22] proposed the LS conjugate gradient method whose direction d_k can be obtained by

$$d_k := \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k^{LS} d_{k-1}, & k \geq 1, \end{cases} \tag{11}$$

where

$$\beta_k^{LS} := -\frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{12}$$

If the exact line search is used, the `LS` method is equivalent to `PRP` method [16], which has been regarded as one of the most efficient `CG` methods in practical computation. We now combine the `LS` conjugate gradient direction with y_{k-1} to obtain a new three-term conjugate gradient method (`N3TCG`) as follows:

$$d_k := \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k^{LS} d_{k-1} + \theta_k y_{k-1}, & k \geq 1, \end{cases} \tag{13}$$

in which

$$\theta_k := \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{14}$$

We will show that (13) satisfies the sufficient descent condition, $g_k^T d_k = -\|g_k\|^2 < 0$, independent of the line search and the objective function convexity. Furthermore, `N3TCG` is reduced to `LS` using the exact line search. To augment the efficiency of `N3TCG`, we now consider a modification on `N3TCG` to get `MN3TCG` using the eigenvalue analysis. In `MN3TG`, the search directions are generated by

$$d_k := \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k^{LS} d_{k-1} + t_k \theta_k y_{k-1}, & k \geq 1, \end{cases} \tag{15}$$

where $t_k \in \mathbb{R}$ is a parameter. `MN3TCG` has the following properties:

- For exact line search or $t_k = 0$, $\forall k \geq 0$, `MN3TCG` reduces to the `LS` method.
- If $t_k = 1$, $\forall k \geq 0$, `MN3TCG` reduces to the `N3TCG`.

We now use the eigenvalues analysis to compute the parameter t_k in `MN3CG`, satisfying the sufficient descent condition. From (15), the search directions of `MN3TCG` can be written as

$$d_k = -Q_k g_k, \qquad \forall k \geq 0, \tag{16}$$

where

$$Q_k := I + \frac{d_{k-1} y_{k-1}^T}{g_{k-1}^T d_{k-1}} - t_k \frac{y_{k-1} d_{k-1}^T}{g_{k-1}^T d_{k-1}}. \tag{17}$$

Hence, `MN3TCG` can be considered as a quasi-Newton method, in which the non-symmetric matrix Q_k is an approximation for the inverse Hessian matrix. Now, (16) implies that

$$d_k^T g_k = -g_k^T Q_k^T g_k = -g_k^T A_k g_k, \tag{18}$$

in which

$$A_k := \frac{Q_k^T + Q_k}{2}.$$

The symmetric matrix A_k can be rewritten as follows:

$$A_k = I - \frac{t_k - 1}{2}\frac{y_{k-1}d_{k-1}^T}{g_{k-1}^T d_{k-1}} - \frac{t_k - 1}{2}\frac{d_{k-1}y_{k-1}^T}{g_{k-1}^T d_{k-1}}. \tag{19}$$

If $d_{k-1} = 0$ or $y_{k-1} = 0$, then $A_k = I$; consequently all the eigenvalues of A_k are 1. Otherwise, there exists a set of orthonormal vectors $\{z_k^i\}_{i=1}^{n-2}$ such that

$$d_{k-1}^T z_k^i = y_{k-1}^T z_k^i = 0, \qquad i = 1, \cdots, n-2. \tag{20}$$

Therefore, from (19) and (20), we get

$$A_k z_k^i = z_k^i, \qquad i = 1, \cdots, n-2.$$

Hence, the matrix A_k has the eigenvalues 1 corresponding to the eigenvectors z_k^i for $i = 1, \cdots, n-2$. Finally, we must to find two remaining eigenvalues λ_k and μ_k of the matrix A_k. Remark 2 gives us

$$\mathrm{tr}(A_k) = n - (t_k - 1)\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}} = \underbrace{1 + \cdots + 1}_{n-2 \text{ times}} + \lambda_k + \mu_k,$$

so that

$$\lambda_k + \mu_k = 2 - (t_k - 1)\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}}. \tag{21}$$

The symmetry of the matrix A_k along with Remark 2 implies that

$$\begin{aligned}
\|A_k\|_F^2 &= \mathrm{tr}(A_k^T A_k) = \mathrm{tr}(A_k^2) \\
&= n - 2 + 2\Big(1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}}\Big)^2 \\
&+ \frac{1}{2}(t_k - 1)^2 \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^2} \\
&= \underbrace{1 + \cdots + 1}_{n-2} + \lambda_k^2 + \mu_k^2,
\end{aligned}$$

leading to

$$\lambda_k^2 + \mu_k^2 = 2\Big(1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}}\Big)^2 + \frac{1}{2}(t_k - 1)^2 \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^2}. \tag{22}$$

From (21) and (22), we have

$$\begin{aligned}
\lambda_k \mu_k &= \frac{1}{2}\Big[(\lambda_k + \mu_k)^2 - (\lambda_k^2 + \mu_k^2)\Big] \\
&= \frac{1}{2}\Big[4\Big(1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}}\Big)^2 - 2\Big(1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}}\Big)^2 \\
&- \frac{1}{2}(t_k - 1)^2 \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^2}\Big] \\
&= \Big(1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}}\Big)^2 - \frac{1}{4}(t_k - 1)^2 \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^2}.
\end{aligned} \tag{23}$$

Finally, the eigenvalues λ_k and μ_k using (21) and (23) are computed

$$\lambda_k = 1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}} + \frac{t_k - 1}{2}\frac{\|y_{k-1}\|}{\|g_{k-1}\|},$$
$$\mu_k = 1 - \frac{t_k - 1}{2}\frac{d_{k-1}^T y_{k-1}}{g_{k-1}^T d_{k-1}} - \frac{t_k - 1}{2}\frac{\|y_{k-1}\|}{\|g_{k-1}\|}.$$

Let $\lambda_k = \xi \in (0, 1)$. Then, to guarantee the sufficient descent condition for directions generated by (15), we compute the well-defined parameter t_k as follows:

$$t_k := \begin{cases} 1, & if\ \Gamma_k = 0, \\ \min\{\tau_1, \max\{1, \tilde{t}_k\}\}, & if\ \Gamma_k \neq 0\ \&\ (g_k^T d_{k-1})(g_k^T y_{k-1}) \geq 0, \\ \min\{\tau_2, \min\{1, \tilde{t}_k\}\}, & if\ \Gamma_k \neq 0\ \&\ (g_k^T d_{k-1})(g_k^T y_{k-1}) < 0, \end{cases} \tag{24}$$

in which $\tau_2 \leq 1 \leq \tau_1$, $\Gamma_k := \|y_{k-1}\| - d_{k-1}^T y_{k-1}$ and

$$\tilde{t}_k := 1 + 2(\xi - 1)\frac{g_{k-1}^T d_{k-1}}{\|y_{k-1}\| - d_{k-1}^T y_{k-1}}.$$

We now describe the proposed three-term conjugate gradient algorithms to solve unconstrained optimization problems:

Algorithm 1. New three-term conjugate gradient method (`N3TCG`)
Input: Choose $x_0 \in \mathbb{R}^n$, $\rho \in (0, 1)$, $\eta_{\min} \in (0, 1)$, $\eta_{\max} \in [\eta_{\min}, 1]$, $N > 0$ and $\epsilon > 0$.

begin
  set $k = 0$ and $d_k = -g_k$;
  while $\|g_k\| > \epsilon$
    determine α_k by (7)-(9);
    set $x_{k+1} = x_k + \alpha_k d_k$;
    compute β_{k+1}^{LS} by (12) and obtain θ_k by (14);
    compute d_{k+1} by (13);
    $k \leftarrow k + 1$;
  end
end

$x_b := x_k$; $f_b := f_k$;
Output: x_b, f_b

Algorithm 2. Modified three-term conjugate gradient method (`MN3TCG`)
Input: Choose $x_0 \in \mathbb{R}^n$, $\rho \in (0, 1)$, $\eta_{\min} \in (0, 1)$, $\eta_{\max} \in [\eta_{\min}, 1]$, $\tau_1 \leq 1 \leq \tau_1$, $N > 0$ and $\epsilon > 0$.

begin
  set $k = 0$ and $d_k = -g_k$;
  while $\|g_k\| > \epsilon$
    determine α_k by (7)-(9);
    set $x_{k+1} = x_k + \alpha_k d_k$;
    compute β_{k+1}^{LS} by (12) and obtain θ_k by (14);
    obtain t_{k+1} by (24);
    compute d_{k+1} by (15);

$k \leftarrow k+1$;
end
end
$x_b := x_k;\ \ f_b := f_k$;
Output: x_b, f_b

The following example is used to explain the eigenvalue analysis in `MN3TCG`.

Example 1. Consider the vectors

$$g_{k-1} = \begin{bmatrix}1\\2\\3\end{bmatrix}, \qquad d_{k-1} = \begin{bmatrix}0\\1\\-2\end{bmatrix}, \qquad y_{k-1} = \begin{bmatrix}4\\2\\-3\end{bmatrix}.$$

The relation (19) gives the symmetric matrix A_k as follows

$$A_k = \begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix} - \frac{t_k-1}{2}\frac{\begin{bmatrix}0&4&-8\\0&2&-4\\0&-3&6\end{bmatrix}}{-4} - \frac{t_k-1}{2}\frac{\begin{bmatrix}0&0&0\\4&2&-3\\-8&-4&0\end{bmatrix}}{-4}$$
$$= \begin{bmatrix}1 & \frac{t_k-1}{2} & 1-t_k\\ \frac{t_k-1}{2} & \frac{t_k+1}{2} & \frac{7-7t_k}{8}\\ 1-t_k & \frac{7-7t_k}{8} & \frac{3t_k+1}{4}\end{bmatrix}.$$

Now, from (21)-(23), we get

$$\begin{aligned}\lambda_k+\mu_k &= \frac{t_k+3}{2},\\ \lambda_k^2+\mu_k^2 &= \frac{65}{56}t_k^2 - \frac{37}{28}t_k + \frac{121}{56},\\ \lambda_k\mu_k &= -\frac{51}{112}t_k^2 - \frac{37}{56}t_k + \frac{121}{112}.\end{aligned} \tag{25}$$

By solving (25), we obtain

$$\lambda_k = \frac{9t_k-2}{7}, \qquad \mu_k = \frac{-11t_k+25}{14}.$$

Let $\xi = 0.15$ and $\tau_1 = 5$. Then, $\tilde{t}_k = -1.6$, $(g_k^T d_{k-1})(g_k^T y_{k-1}) = 4$ and

$$t_k = \min\{5, \max\{1, -1.6\}\} = 1.$$

3 Convergence Analysis

In this section, we investigate the global convergence results of the proposed methods. For these, the following assumptions are needed.

Assumption 3.1 The level set $L(x_0) := \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ is bounded, i.e., there exists a positive constant $B > 0$ such that $\|x\| \leq B$ for all $x \in L(x_0)$.

Assumption 3.2 In the neighborhood Ω of $L(x_0)$, the gradient $g(x)$ is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \Omega. \tag{26}$$

Assumptions 3.1 and 3.2 imply that there exists a constant $\zeta > 0$ such that

$$\|g_k\| \leq \zeta, \qquad \forall k \geq 0. \tag{27}$$

We now show that the generated directions in both algorithms satisfy the sufficient descent condition (3) independent of line search type.

Lemma 1. Suppose that d_k is generated by (13) or (15). Then, d_k is the sufficient descent direction, i.e., $g_k^T d_k \leq -\|g_k\|^2$.

Proof. From (12)-(14), we get

$$g_k^T d_k = -\|g_k\|^2 - \frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T y_{k-1} = -\|g_k\|^2 < 0.$$

Suppose that d_k is generated by `MN3TCG`. Then, (12), (14) and (15) imply that

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 - \frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} + t_k \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T y_{k-1} \\ &= -\|g_k\|^2 + (t_k - 1) \frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1}. \end{aligned} \tag{28}$$

If $\Gamma_k = 0$, then $t_k = 1$. Therefore, `MN3TCG` is reduced to `N3TCG` satisfying the sufficient descent condition. We continue the proof with the induction over k. Using the induction hypothesis, we have $g_{k-1}^T d_{k-1} \leq -\|g_{k-1}\|^2 < 0$. Now, we have two cases:
Case (i): If $\Gamma_k \neq 0$ and $(g_k^T d_{k-1})(g_k^T y_{k-1}) \geq 0$, then

$$t_k = \min\{\tau_1, \max\{1, \tilde{t}_k\}\} \geq 1.$$

This inequality, along with (28), results in

$$g_k^T d_k = -\|g_k\|^2 + (t_k - 1) \frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \leq -\|g_k\|^2.$$

Case (ii): If $\Gamma_k \neq 0$ and $(g_k^T d_{k-1})(g_k^T y_{k-1}) < 0$, then

$$t_k = \min\{\tau_2, \min\{1, \tilde{t}_k\}\} \leq 1.$$

Similar to case (i), we have

$$g_k^T d_k = -\|g_k\|^2 + (t_k - 1) \frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}} g_k^T d_{k-1} \leq -\|g_k\|^2.$$

Hence, d_k satisfies the sufficient descent condition in both algorithms. □

Lemma 2. Suppose that d_k is a sufficient descent direction and Assumptions 3.1 and 3.2 hold. Then

$$\sum_{l(k)-1} \frac{\left(g_{l(k)-1}^T d_{l(k)-1}\right)^2}{\|d_{l(k)-1}\|^2} < \infty.$$

Proof. See Lemma 3.4 in [10]. □

Lemma 3. Let d_k be the direction generated by `N3TCG` or `MN3TCG` and the step-size α_k be obtained by the nonmonotone Armijo-type line search (7)-(9). Also, Assumptions 3.1 and 3.2 hold. If

$$\sum_{l(k)-1} \frac{1}{\|d_{l(k)-1}\|^2} = +\infty, \tag{29}$$

then

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

Proof. By contradiction, we suppose that $\lim_{k\to\infty} \inf \|g_k\| \neq 0$. Hence, there exists a constant $\gamma > 0$ such that

$$\|g_k\| \geq \gamma, \qquad \forall k. \tag{30}$$

Define

$$\Pi_k := \frac{g_k^T d_k}{\|g_k\|\|d_k\|}. \tag{31}$$

Using the Lemma 1, we have $g_k^T d_k \leq -\|g_k\|^2$; hence

$$\Pi_k \leq -\frac{\|g_k\|}{\|d_k\|},$$

so that

$$\Pi_k^2 \geq \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{32}$$

Therefore, (30)-(32) lead to

$$\frac{\gamma^2}{\|d_k\|^2} \leq \frac{\|g_k\|^2}{\|d_k\|^2} \leq \Pi_k^2 = \frac{(g_k^T d_k)^2}{\|g_k\|^2\|d_k\|^2} \leq \frac{(g_k^T d_k)^2}{\gamma^2\|d_k\|^2}. \tag{33}$$

Without any loss of generality, we can take $k := l(k) - 1$. From Lemma 2, we obtain

$$\sum_{l(k)-1} \frac{1}{\|d_{l(k)-1}\|^2} \leq \sum_{l(k)-1} \frac{\left(g_{l(k)-1}^T d_{l(k)-1}\right)^2}{\|d_{l(k)-1}\|^2} < \infty,$$

which contradicts with (29). Hence, the proof of the desired result is completed. □

Theorem 1. Let d_k be the direction generated by `N3TCG` or `MN3TCG` and the step-size α_k be obtained by the nonmonotone Armijo-type line search (7)-(9). If Assumptions 3.1 and 3.2 hold, then

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

Proof. In the conjugate gradient algorithms, it is clear that iterations can fail when $\|g_k\| > \varepsilon$ for all $k \geq 0$ [27]. By using Lemma 1 and (27), we get

$$|\theta_k| = \Big|\frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}\Big| = \frac{|g_k^T d_{k-1}|}{|g_{k-1}^T d_{k-1}|} \leq \frac{\|g_k\|\|d_{k-1}\|}{\|g_{k-1}\|^2} \leq \frac{\zeta}{\varepsilon^2}\|d_{k-1}\|. \tag{34}$$

Assumption 3.2 results in

$$\|y_{k-1}\| \leq L\alpha_{k-1}\|d_{k-1}\|. \tag{35}$$

Now, Cauchy–Schwarz inequality along with Lemma 1, (12) and (35) gives us

$$|\beta_k^{LS}| = \Big| - \frac{g_k^T y_{k-1}}{g_{k-1}^T d_{k-1}}\Big| = \frac{|g_k^T y_{k-1}|}{|g_{k-1}^T d_{k-1}|} \leq \frac{\|g_k\|\|y_{k-1}\|}{\|g_{k-1}\|^2} \leq \frac{\zeta}{\varepsilon^2}L\alpha_{k-1}\|d_{k-1}\|. \tag{36}$$

From (15) and (27), we get

$$\begin{aligned}\|d_k\| &\leq \|g_k\| + |\beta_k^{LS}|\|d_{k-1}\| + |t_k||\theta_k|\|y_{k-1}\| \\ &\leq \zeta + \frac{\zeta}{\varepsilon^2}L\alpha_{k-1}\|d_{k-1}\|^2 + |t_k|\frac{\zeta}{\varepsilon^2}\|d_{k-1}\|L\alpha_{k-1}\|d_{k-1}\| \\ &= \zeta + \frac{\zeta}{\varepsilon^2}L\alpha_{k-1}\|d_{k-1}\|^2 + |t_k|\frac{\zeta}{\varepsilon^2}L\alpha_{k-1}\|d_{k-1}\|^2. \end{aligned} \tag{37}$$

Finally, we have three cases:
Case (i): In `N3TCG`, we have $t_k = 1$, then

$$\|d_k\| \leq \zeta + \frac{2L\zeta}{\varepsilon^2}\alpha_{k-1}\|d_{k-1}\|^2.$$

Case (ii): If $\Gamma_k \neq 0$ and $(g_k^T d_{k-1})(g_k^T y_{k-1}) \geq 0$, then $|t_k| \leq \tau_1$

$$\|d_k\| \leq \zeta + \frac{(1+\tau_1)L\zeta}{\varepsilon^2}\alpha_{k-1}\|d_{k-1}\|^2.$$

Case (iii): If $\Gamma_k \neq 0$ and $(g_k^T d_{k-1})(g_k^T y_{k-1}) < 0$, then $|t_k| \leq \tau_2$

$$\|d_k\| \leq \zeta + \frac{(1+\tau_2)L\zeta}{\varepsilon^2}\alpha_{k-1}\|d_{k-1}\|^2.$$

Since $\tau_2 \leq 1 \leq \tau_1$, for all cases, we have

$$\|d_k\| \leq \zeta + \frac{(1+\tau_1)L\zeta}{\varepsilon^2}\alpha_{k-1}\|d_{k-1}\|^2. \tag{38}$$

Similar to the proof of Lemma 3.1 in [35], there exists a positive constant M such that

$$\|d_k\| \leq M, \qquad \forall k \geq 0.$$

Moreover, for $k := l(k) - 1$, it is clear that

$$\sum_{l(k)-1} \frac{1}{\|d_{l(k)-1}\|^2} \geq \sum_{l(k)-1} \frac{1}{M^2} = +\infty.$$

Hence, Lemma 3 implies that

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

□

4 Numerical Experiments

In this section, we compare N3TCG and MN3TCG with the LS conjugate gradient method [22], the hybrid conjugate gradient method (HLSFR) [6] and the sufficient descent LS conjugate gradient method (MMLS) [21] to solve the unconstrained optimization problems. In our experiments, all codes are written in Matlab 2017a programming environment on a 2.3Hz Intel core i3 processor laptop and 4GB of RAM with the double precision data type in Linux operations system. We selected a number of 150 test functions of nonlinear unconstrained optimization problems from the CUTEst [4] library. The parameters are chosen as $\xi := 0.15$, $\rho := 0.01$, $\tau_1 := 5$, $\tau_2 := 0.99$ and $N := 10$. The parameter η_k is updated by

$$\eta_k := \begin{cases} \eta_0/2, & \text{if } k = 1, \\ (\eta_{k-1} + \eta_{k-2})/2, & \text{if } k > 1, \end{cases}$$

in which $\eta_0 := 0.15$. All algorithms are stopped when $\|g_k\| \leq 10^{-6}$ or the total number of iterates exceeds 10000.

Figures 1-3 show the performance of LS, HLSFR, MMLS, N3TCG and MN3TCG to solve the unconstrained optimization problems, which are evaluated using the profiles of Dolan and Moré [7]. In these figures, N_i, N_f, N_g and C_t indicate the total number of iterations, the total number of function evaluations, the total number of gradient evaluations and CPU times, respectively. Also, we will plot the fraction $P(\tau)$ of optimization minimization problems for which the algorithm is within a factor τ of the best time. Figure 1 shows that MN3TCG has the best performance about the number of iterations since it can solve about 37% of the test problems with the smallest number of iterations. It is easy to see from Figure 2 MN3TCG is more competitive than LS, HLSFR, MMLS and N3TCG since this algorithm can solve 38% of test problems better than others in terms of $N_f + 3N_g$. Finally, we can obtain from Figure 3 that N3TCG method is better than LS, HLSFR, MMLS and MN3TCG methods about 43% of the most wins in terms of C_t. Thus, the modified three-term conjugate gradient method turns out to be practically efficient.

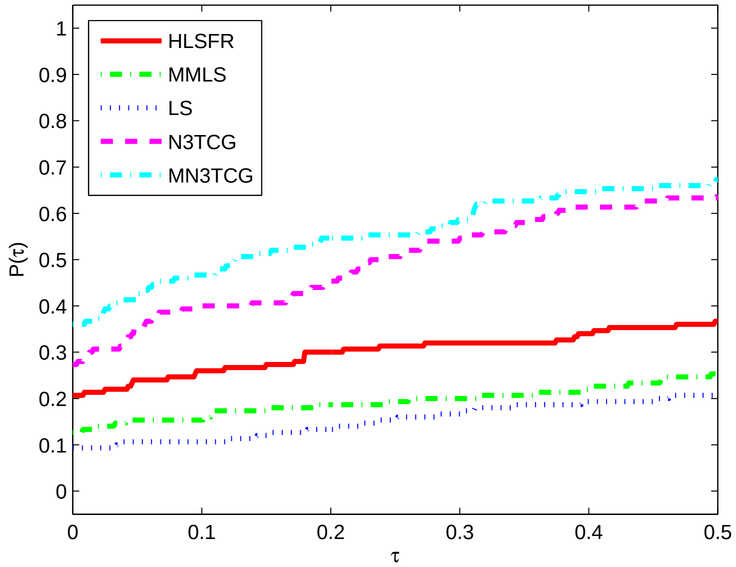

Figure 1: The performance profile in terms of $\mathtt{N_i}$ for `HLSFR`, `MMLS`, `LS`, `N3TCG` and `MN3TCG`

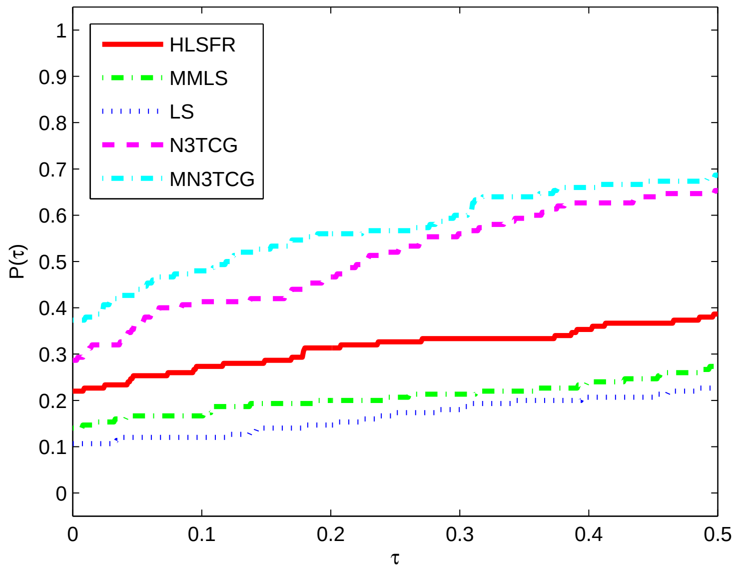

Figure 2: The performance profile in terms of $\mathtt{N_f} + 3\mathtt{N_g}$ for `HLSFR`, `MMLS`, `LS`, `N3TCG` and `MN3TCG`

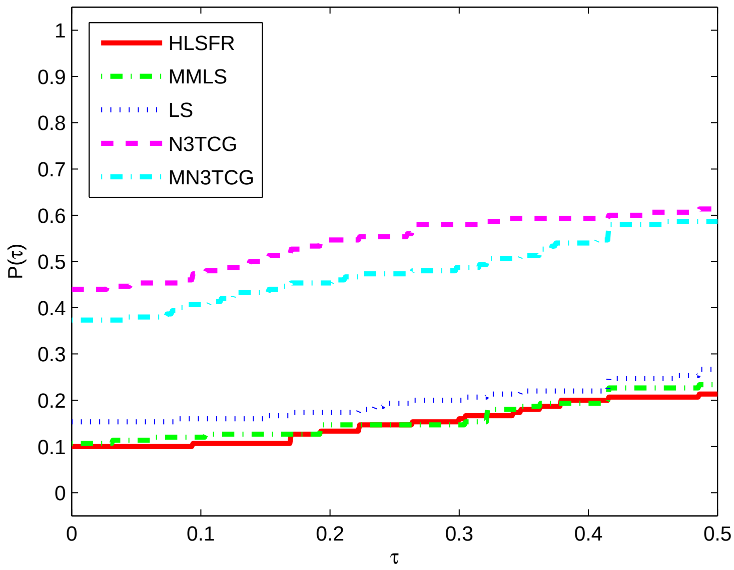

Figure 3: The performance profile in terms of $\mathtt{C_t}$ for `HLSFR`, `MMLS`, `LS`, `N3TCG` and `MN3TCG`

5 Conclusions

We have presented two modifications of the `LS` conjugate gradient algorithm to solve unconstrained optimization problems. Also, we have shown that the generated directions by the new three-term conjugate gradient methods satisfy the sufficient descent condition. Furthermore, the conjugate gradient parameter is obtained using the eigenvalue analysis for an approximation of the inverse Hessian matrix. We have established the global convergence of our methods under mild conditions. The numerical results have indicated that the proposed methods are efficient and robust to solve the unconstrained optimization problems.

References

[1] Ahookhosh M., Amini K., Peyghami M.R. (2012). “A nonmonotone trust–region line search method for large-scale unconstrained optimization”, Applied Mathematical Modelling, 36(1), 478–487.

[2] Andrei N. (2015). “A new three–term conjugate gradient algorithm for unconstrained optimization”, Applied Mathematical Modelling, 68, 305–321.

[3] Beale E.M.L. (1972). "A derivative of conjugate gradients, in: Lootsma F A, eds. Numerical methods for nonlinear optimization", London, Academic Press, 39–43.

[4] Conn A. R., Gould N. I. M., Toint Ph. L. (1995). "CUTE: constrained and unconstrained testing environment", ACM T. Math. Software, 21, 123–160.

[5] Dai Y., Yuan Y. (1999). "A nonlinear conjugate gradient method with a strong global convergence property", SIAM Journal on Optimization, 10(1), 177–182.

[6] Djordjevic S.S. (2019). "New hybrid conjugate gradient method as a convex combination of LS and FR methods", Acta Mathematica Scientia, 39B(1), 214–228.

[7] Dolan E.D., Moré J. (2012). "Benchmarking optimization software with performance profiles", Mathematical Programming Ser A, 91(2), 201–213.

[8] Edelman A., Mith T.S. (1996). "On conjugate gradient–like methods for eigen–like problems", BIT Numerical Mathematics, 36(1), 1–16.

[9] Esmaeili H., Kimiaei M. (2014). "An improved adaptive trust–region method for unconstrained optimization", Mathematical Modelling and Analysis, 19(4), 469–490.

[10] Esmaeili H., Rostami M., Kimiaei M. (2018). "Extended Dai–Yuan conjugate gradient strategy for large–scale unconstrained optimization with applications to compressive sensing", Filomat, 32(6), 323–337.

[11] Fletcher R. (1987). "*Practical methods of optimization, in: Unconstrained Optimization*", John Wiley & Sons, New York.

[12] Fletcher R., Reeves C. (1964). "Function minimization by conjugate gradients", Computer Journal, 72(2), 149–154.

[13] Gill P., Leonard M.W. (2001). "Reduced–Hessian Quasi–Newton methods for unconstrained optimization", SIAM Journal on Optimization, 12(1), 209–237.

[14] Gill P., Murry W. (1972). "Quasi–Newton Methods for Unconstrained Optimization", IMA Journal of Applied Mathematics, 9(1), 91–108.

[15] Grippo L., Lamparillo F., Lucidi S. (1986). "A nonmonotone line search technique for Newtons method", SIAM Journal on Numerical Analysis, 23, 707–716.

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

[17] Hestenes M.R., Stiefel E.L. (1952). "Methods of conjugate gradients for solving linear systems", Journal of Research of the National Bureau of Standards, 49, 409–436.

[18] Kimiaei M., Ghaderi S. (2017). "A new restarting adaptive trust–region method for unconstrained optimization", Journal of the Operations Research Society of China, 5(4), 487–507.

[19] Kimiaei M., Rostami M. (2016). "Impulse noise removal based on new hybrid conjugate gradient approach", Kybernetika, 52(5), 791–823.

[20] Li Z.F., Chen J., Deng N.Y. (1997). "A New Conjugate Gradient Method and Its Global Convergence Properties", Mathematical Programming, 78, 375–391.

[21] Li M., Feng H. (2011). "A sufficient descent LS conjugate gradient method for unconstrained optimization problems", Applied Mathematics and Computation, 218, 1577–1586.

[22] Liu Y.L., Storey C. (1991). "Efficient generalized conjugate gradient algorithms", part1: Journal of Optimization Theory and Applications, 69(1), 129–137.

[23] Narushima Y., Yabe H., Ford J.A. (2011). "A three–term conjugate gradient method with sufficient descent property for unconstrained optimization", SIAM Journal on Optimization, 21, 212–230.

[24] Nocedal J, Wright S. (2006). "*Numerical Optimization*", Springer, New York.

[25] Polak E., Ribière G. (1969). "Note sur la convergence de directions conjuguée", Rev. Francaise Informat Recherche Operationelle, 3e Année, 16, 35–43.

[26] Polyak B.T. (1969). "The conjugate gradient method in extreme problems", USSR Computational Mathematics and Mathematical Physics, 9, 94–112.

[27] Powell M.J.D. (1984). "Nonconvex minimization calculations and the conjugate gradient method", in: Numerical Analysis (Dundee, 1983), Lecture Notes in Mathematics. 1066, Springer, Berlin, 122–141.

[28] Strang G. (2016). " *Introduction to Linear Algebra*", Springer, Fifth Edition, New York.

[29] Wang F., Zhang K., Wang C., Wang L. (2008). "A variant of trust–region methods for unconstrained optimization", Applied Mathematics and Computation, 203, 297–307.

[30] Yuan G., Hu W. (2018). "A conjugate gradient algorithm for large–scale unconstrained optimization problems and nonlinear equations", Journal of Inequalities and Applications, 113, https://doi.org/10.1186/s13660–018–1703–1.

[31] Yang X., Luo Z., Dai X. (2013). "A global convergence of LS–CD hybrid conjugate gradient method", Advances in Numerical Analysis, 517–452.

[32] Yang X., Sarkar T. P., Arvas E. (1989). "A survey of conjugate gradient algorithms for solution of extreme eigenproblems of a symmetric matrix", IEEE Trans. Acoust. Speech Signal Processing, 37, 1550–1556.

[33] Zhang L. (2009). "A new Liu–Storey type nonlinear conjugate gradient method for unconstrained optimization problems", J. Comput. Appl. Math, 225, 146–157.

[34] Zhang H., Hager W.W. (2014). "A nonmonotone line search technique and its application to nuconstrained optimization", SIAM Journal on Optimization, 14(4), 1043–1056.

[35] Zhang L., Zhou W., Li D.H. (2006). "A descent modified Polak-Ribière-Polyak conjugate gradient method and its global convergence", IMA Journal of Numerical Analysis, 26(4), 629–640.

[36] Zhang L., Zhou W., Li D. (2007). "Some descent three–term conjugate gradient methods and their global convergence", Optim Methods Software, 22(4), 697–711.

[37] Zheng X., Shi J. (2018). "A Modified Sufficient Descent Polak–Ribiére–Polyak Type Conjugate Gradient Method for Unconstrained Optimization Problems", Algorithms, 11(9), https://doi.org/10.3390/a11090133.

[38] Zhou Q., Zhou F., Cao F. (2013). "A nonmonotone trust–region method based on simple conic models for unconstrained optimization", Applied Mathematics and Computation, 225, 295–305.

یک روش گرادیان مزدوج ترکیبی جدید بر پایه آنالیز مقادیر ویژه برای مسائل بهینه‌سازی نامقید

راهپیمایی، ف.
استادیار، گروه ریاضی کاربردی- نویسنده مسئول
ایران، تهران، دانشگاه پیام نور، گروه ریاضی، صندوق پستی ۱۹۳۹۵-۳۶۹۷
rahpeyma_83@yahoo.com

رستمی، م.
استادیار، گروه ریاضی کاربردی،
ایران، همدان، دانشگاه آزاد اسلامی واحد همدان، باشگاه پژوهشگران جوان و نخبگان
majid403rostami@yahoo.com

تاریخ دریافت: ۲۹ بهمن ۱۳۹۷ تاریخ پذیرش: ۲۳ شهریور ۱۳۹۸

چکیده

در این مقاله، دو روش گرادیان مزدوج سه‌جمله‌ای تعمیم‌یافته بر پایه روش گرادیان مزدوج لیو-استوری برای حل مسائل بهینه‌سازی نامقید ارائه شده است. ویژگی مهم و اصلی روش‌های جدید، تولید جهت‌های جستجوی کاهشی بر پایه آنالیز مقادیر ویژه و مستقل از نوع جستجوی خطی است. همگرایی سراسری الگوریتم‌های جدید ارائه شده تحت برخی فرض‌های مناسب اثبات شده است. نتایج عددی نشان می‌دهند که روش‌های پیشنهادی برای حل مسائل بهینه‌سازی نامقید کارا و قوی هستند.

کلمات کلیدی

بهینه‌سازی نامقید، روش‌های گرادیان مزدوج، آنالیز مقادیر ویژه، همگرایی سراسری، مقایسه‌های عددی.