Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 3, pp. 461–472, 2020

SOME THREE-TERM CONJUGATE GRADIENT METHODS FOR SOLVING UNCONSTRAINED OPTIMIZATION PROBLEMS

L. Arman, Y. Xu, M. Rostami and F. Rahpeymaii

Abstract: In this paper, we investigate some three-term conjugate gradient methods, since these methods are an efficient class for unconstrained optimization problems. Three-term conjugate gradient methods are obtained to improve traditional conjugate gradient methods. Conjugate gradient methods have low memory requirements and strong local and global convergence properties. Here, we focus on the three-term conjugate gradient methods which generated directions satisfy the descent condition, the sufficient descent condition and the Dai-Liao conjugacy condition. We compare some three-term conjugate gradient methods on the unconstrained `CUTEst` test problems.

Key words: *unconstrained optimization, three-term conjugate gradient method, sufficient descent condition, inexact line search, numerical comparisons*

Mathematics Subject Classification: *90C06, 65K05, 90C26, 90C30*

1 Introduction

Conjugate gradient (CG) methods are usually effective for solving a large-scale unconstrained optimization problem

$$\min_{x\in\mathbb{R}^n} f(x), \tag{1.1}$$

where $f:\mathbb{R}^n\to\mathbb{R}$ is assumed to be a continuously differentiable function and bounded from below. Using an initial guess $x_0\in\mathbb{R}^n$, the conjugate gradient methods generate a new iterate by

$$x_{k+1}=x_k+\alpha_k d_k, \tag{1.2}$$

where the step-size $\alpha_k>0$ is obtained by an inexact line search and the direction d_k is computed by

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

Here, $g_k=\nabla f(x_k)$ and β_k is called the conjugate parameter. Some well-known formulas for β_k are as follows:

$$\beta_k^{FR}=\frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \text{Fletcher \& Reeves (FR) [16]} \tag{1.4}$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \text{Hestenes \& Stiefel (HS) [19]} \tag{1.5}$$

$$\beta_k^{PR} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \text{Polak \& Ribière (PR) [22]} \tag{1.6}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}}, \qquad \text{Dai \& Yuan (DY) [9]} \tag{1.7}$$

$$\beta_k^{HZ} = \beta_k^{HS} - 2\|y_{k-1}\|^2 \frac{d_{k-1}^T g_k}{(d_{k-1}^T y_{k-1})^2}, \qquad \text{Hager \& Zhang (HZ) [17]} \tag{1.8}$$

where $\|\cdot\|$ is the Euclidean norm and $y_{k-1} = g_k - g_{k-1}$. In the case that the objective function f is strictly convex quadratic and the exact line search is used, all the choices for β_k are equivalent, see [18].

In convergence analysis of the conjugate gradient methods, we say that the descent condition holds if for each search directions d_k

$$g_k^T d_k < 0, \quad \forall k \geq 1. \tag{1.9}$$

Also, the direction d_k satisfies the sufficient descent condition, if there exists a constant $c > 0$ such that

$$g_k^T d_k \leq -c\|g_k\|^2, \quad \forall k. \tag{1.10}$$

The conjugacy condition is utilized to accelerate the conjugate gradient methods. So, the order accuracy in the estimation of the curvature is improved. By modifying the HS method, Dai & Liao [10] proposed the conjugacy condition as follows

$$d_k^T y_{k-1} = -\xi g_k^T s_{k-1}, \tag{1.11}$$

where $\xi > 0$ is a constant and $s_{k-1} = x_k - x_{k-1}$. Recently, Babaie-Kafaki [4] and Fatemi [15] discussed how to find the optimal choice for the parameter ξ.

In each conjugate gradient iterate, the step-size α_k is chosen as an approximate minimum of one dimensional optimization problem

$$\min_{\alpha \geq 0} f(x_k + \alpha d_k). \tag{1.12}$$

The termination conditions for the line search are often based on some versions of the Wolfe conditions. The standard Wolfe conditions [21] are

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k, \tag{1.13}$$

$$g_{k+1}^T d_k \geq \sigma_2 g_k^T d_k, \tag{1.14}$$

where $0 < \sigma_1 < \sigma_2 < 1$. Furthermore, the strong Wolfe conditions are presented by

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \sigma_1 \alpha_k g_k^T d_k, \tag{1.15}$$

$$|g_{k+1}^T d_k| \leq -\sigma_2 g_k^T d_k. \tag{1.16}$$

Assumption 1.1 The gradient g is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \qquad \forall x, y \in \mathbb{R}^n.$$

Assumption 2.1 The function $f : \mathbb{R}^n \to \mathbb{R}$ is an uniformly convex function, i.e., there exists a constant $\mu > 0$ such that

$$\big(g(x) - g(y)\big)^T (x - y) \geq \mu \|x - y\|^2, \qquad \forall x, y \in \mathbb{R}^n.$$

Birgin & Martínez [7] proposed a spectral conjugate gradient method by combining the spectral gradient method with the conjugate gradient ideas. In the spectral conjugate gradient methods, the direction d_k is computed by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -\theta_k g_k + \beta_k d_{k-1}, & k \geq 1, \end{cases} \tag{1.17}$$

where β_k is the conjugate parameter and θ_k is a scalar. Recently, the three-term conjugate gradient (TTCG) methods have been proposed to solve the unconstrained optimization problem (1.1). In the first, these methods proposed by Beale [6] as

$$d_k = -g_k + \beta_k d_k + \gamma_k d_t, \tag{1.18}$$

where $\beta_k = \beta_k^{FR}, \beta_k^{HS}, \beta_k^{DY}$. Also, d_t is a restart direction and

$$\gamma_k = \begin{cases} 0, & k = t + 1, \\ \dfrac{g_k^T y_t}{d_t^T y_t}, & k > t + 1. \end{cases} \tag{1.19}$$

In recent years, many researchers obtain further results the three-term conjugate gradient algorithms and established their global convergence properties. Furthermore, they showed that the three-term conjugate gradient algorithms are robust and efficient, especially for large scale problems. [2, 26, 27].

2 TTCG Methods with Sufficient Descent Condition

All proposed three-term conjugate gradient methods by Al-Baali et al. [1] (TTM1), Babaie-Kafaki [3] (TTM2) and Yuan et al. [23] (TTM3) satisfy the sufficient descent condition. In TTM1, the descent direction is given by

$$d_k = \begin{cases} -g_k, & k = 0 \text{ or } |g_k^T y_{k-1}| \leq \theta \|g_k\| \|y_{k-1}\|, \\ -g_k + \beta_k d_{k-1} + \eta_k y_{k-1}, & k \geq 1, \end{cases} \tag{2.1}$$

in which $\beta_k = \beta_k^{HS}$, β_k^{PR}, β_k^{HZ}, $0 < \theta < 1$ is a constant and

$$\eta_k = -\frac{(\gamma_k - 1)\|g_k\|^2 + \beta_k g_k^T d_{k-1}}{g_k^T y_{k-1}}. \tag{2.2}$$

The condition $|g_k^T y_{k-1}| \leq \theta \|g_k\| \|y_{k-1}\|$ is as a restarting criterion, since it holds for sufficiently large values of θ. In numerical experiments, we use $\theta = 10^{-5}$. Also, there are 16 different choices for the parameter γ_k, one of which is robust than others, proposed by Barzilai & Borwein with the following form [5]:

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

The global convergence property for TTM1 is established under Lipschitz continuous for the gradient g while the step-size α_k satisfies the standard Wolfe conditions.

Although the PR method is numerically efficient, but the descent condition is not established even for uniformly convex objective functions, in many cases. TTM2 is a hybridization of PR method and a three-term conjugate gradient method with the sufficient descent condition, i.e., $g_k^T d_k = -\|g_k\|^2$, $\forall k \geq 0$, based on an eigenvalue analysis for convex objective functions, independent of the line search. In this method, the search direction is computed by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k^{PR} d_{k-1} - t_k \dfrac{g_k^T d_{k-1}}{\|g_{k-1}\|^2} y_{k-1}, & k \geq 1, \end{cases} \tag{2.3}$$

in which $t_k \in [0, 1]$. Note that if $t_k = 0$ or the exact line search is used, TTM2 reduces to the PR method. The hybridization parameter t_k is computed such that TTM2 satisfies the sufficient descent condition. Hence, TTM2 is considered as a quasi-Newton method, in which the inverse Hessian is approximated by the non-symmetric matrix. The parameter t_k is given by

$$t_k = \begin{cases} 1, & d_{k-1}^T y_{k-1} + \|d_{k-1}\|\|y_{k-1}\| = 0 \text{ or } \bar{t}_k < 0, \\ \bar{t}_k, & \bar{t}_k \geq 0, \end{cases} \tag{2.4}$$

in which

$$\bar{t}_k = 1 + 2(\zeta - 1)\frac{\|g_{k-1}\|^2}{d_{k-1}^T y_{k-1} + \|d_{k-1}\|\|y_{k-1}\|},$$

with $\zeta = 0.80$. The global convergence of TTM2 has been established under the Armijo-type line search and the Lipschitz continuous of the gradient g.

A derivative-free approach for the large-scale nonlinear equations is proposed in TTM3. It is clear that solving the nonlinear equation $F(x) = 0$ ($F : \mathbb{R}^n \to \mathbb{R}^m$) is equivalent with the optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2}\|F(x)\|^2.$$

To solve the above optimization problem, Yuan et al. [23] defined the search direction d_k by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \delta_k d_{k-1} - \eta_k y_{k-1}, & k \geq 1, \end{cases} \tag{2.5}$$

where

$$\delta_k = \frac{g_k^T y_{k-1}}{\max\{\mu\|d_{k-1}\|\|y_{k-1}\|, \|g_{k-1}\|^2\}}, \tag{2.6}$$

$$\eta_k = \frac{g_k^T d_{k-1}}{\max\{\mu\|d_{k-1}\|\|y_{k-1}\|, \|g_{k-1}\|^2\}}, \tag{2.7}$$

with $\mu = 0.01$. Now, for $\|g(z_k)\| < 10^{-4}$, $z_k = x_k + \alpha_k d_k$ is a new iterate. In other words, if the search direction belongs to a trust-region without special conditions, then TTM3 has the trust-region technique. Also, the scaling term $\max\{\mu\|d_{k-1}\|\|y_{k-1}\|, \|g_{k-1}\|^2\}$ in the denominator guarantees that all search directions automatically will stay in a trust-region because

$$\|g_k\| \leq \|d_k\| \leq \Big(1 + \frac{2}{\mu}\Big)\|g_k\|.$$

As well as, for $\|g(z_k)\| \geq 10^{-4}$ the new iterate is computed by

$$x_{k+1} = x_k - \frac{g(z_k)^T(x_k - z_k)}{\|g(z_k)\|^2} g(z_k).$$

TTM3 is stopped in the finite number of iterates and its global convergence results are established with a new line search technique as follows [20]:

$$-F(x_k + \alpha_k d_k)^T d_k \geq \sigma \alpha_k \|F(x_k + \alpha_k d_k)\| \|d_k\|^2.$$

3 TTCG Methods with Descent and Conjugacy Conditions

Recently, many three-term conjugate gradient algorithms have introduced satisfying both the descent condition and the conjugacy condition. The presented three-term conjugate gradient method by ANDREI [2] (TTM4) generates the search direction by minimizing an one dimensional quadratic model of the objective function. TTM4 uses the quadratic Taylor approximate of the objective function at x_{k+1} such that the symmetric matrix B_k is as an approximation of the Hessian matrix satisfying $B_k s_{k-1} = \omega^{-1} y_{k-1}$ with $\omega \neq 0$. After a simple algebra, the search direction in TTM4 has the following form

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \delta_k s_{k-1} - \eta_k y_{k-1}, & k \geq 1, \end{cases} \tag{3.1}$$

where the parameters δ_k and η_k are

$$\delta_k = \frac{y_{k-1}^T g_k - \omega s_{k-1}^T g_k}{y_{k-1}^T s_{k-1}}, \tag{3.2}$$

$$\eta_k = \frac{s_{k-1}^T g_k}{y_{k-1}^T s_{k-1}}. \tag{3.3}$$

There are obtained many choices for the parameter ω in (3.2). For example, if $\omega = 0$ (3.1) reduces to the HS direction. For $\omega = 0.1$, TTM4 reduces to the DAI & LIAO method [10]. However, for $\omega > 0$ the search direction (3.1) satisfies the descent condition and the Dai–Liao conjugacy condition (1.11) with $\xi = \omega + \|y_{k-1}\|^2 / y_{k-1}^T s_{k-1}$. Therefore, a suitable choice for the parameter ω is

$$\omega = \frac{2}{\|s_{k-1}\|^2} \sqrt{\|s_{k-1}\|^2 \|y_{k-1}\|^2 - (y_{k-1}^T s_{k-1})^2}, \tag{3.4}$$

which is well-defined whenever $\|s_{k-1}\| \neq 0$. The global convergence of TTM4 is established by using the strong Wolfe line search (1.15)- and (1.16) and Lipschitz continuous of the gradient for the uniformly convex objective function.

Another three-term conjugate gradient algorithm with descent and conjugacy conditions to solve large-scale unconstrained optimization problem is developed in [11], called TTM5. In this method, the search direction is close to the Newton direction which does not need to compute or store any Hessian matrix of the objective function, but the search direction satisfies an approximate secant equation such that the numerical performance is greatly improved. In TTM5, the search direction is computed by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k - \delta_k s_{k-1} - \eta_k y_{k-1}, & k \geq 1, \end{cases} \tag{3.5}$$

where

$$\delta_k = \Big(1 - \min\Big\{1, \frac{\|y_{k-1}\|^2}{y_{k-1}^T s_{k-1}}\Big\}\Big) \frac{s_{k-1}^T g_k}{y_{k-1}^T s_{k-1}} - \frac{y_{k-1}^T g_k}{y_{k-1}^T s_{k-1}}, \tag{3.6}$$

$$\eta_k = \frac{s_{k-1}^T g_k}{y_{k-1}^T s_{k-1}}. \tag{3.7}$$

It is established that (3.5) is the descent direction. Furthermore, the conjugacy condition (1.11) is established with

$$\xi = 1 - \min\Big\{1, \frac{\|y_{k-1}\|^2}{y_{k-1}^T s_{k-1}}\Big\} + \frac{\|y_{k-1}\|^2}{y_{k-1}^T s_{k-1}} > 0.$$

It is clear that, if $\|y_{k-1}\| < y_{k-1}^T s_{k-1}$ then $\xi = 1$. Similar to TTM4, the global convergence is obtained under the strong Wolfe line search using Lipschitz continuous of the gradient for the uniformly convex objective function in TTM5.

4 TTCG Methods with Sufficient Descent and Conjugacy Conditions

Two new versions of the three-term conjugate gradient methods with sufficient descent and Dai–Liao conjugacy conditions are introduced by Dong et al. [13] (TTM6) and [14] (TTM7). In TTM6, the search direction d_k satisfied the Dai–Liao conjugacy condition and the sufficient descent condition, simultaneously. They started from the descent steepest direction $d_k = -g_k$ and then the search direction is presented by

$$d_k = \begin{cases} -g_k, & \|y_{k-1}\|\|d_{k-1}\| \geq \mu\|g_k\|, \\ -g_k + \beta_k^{MHZ+} d_{k-1} - \theta_k y_{k-1}, & \|y_{k-1}\|\|d_{k-1}\| < \mu\|g_k\|, \end{cases} \tag{4.1}$$

where $\mu > 0$ is a large enough constant. Furthermore, they used $\theta_k = g_k^T d_{k-1}/d_{k-1}^T y_{k-1}$ and

$$\beta_k^{MHZ+} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - t_k \frac{\|y_{k-1}\|^2}{(d_{k-1}^T y_{k-1})^2} \max\{0, g_k^T d_{k-1}\}, \tag{4.2}$$

as the conjugate parameters. Since the numerical stability is a significant property of numerical algorithms, an optimal parameter for t_k in (4.2) is obtained based on an eigenvalue study and a singular value analysis in [13]. Hence, they discussed on eigenvalue and singular value of iterate matrix

$$Q_k = I - \frac{s_{k-1}}{s_{k-1}^T y_{k-1}} \Big(y_{k-1} - t\frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} s_{k-1}\Big)^T + \frac{y_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}},$$

and obtained the optimal parameter t^* as follows:

$$t^* = \frac{2}{\gamma_k}\sqrt{\gamma_k - 1},$$

in which

$$\gamma_k = \Big(\frac{\|s_{k-1}\|\|y_{k-1}\|}{s_{k-1}^T y_{k-1}}\Big)^2. \tag{4.3}$$

However, this choice for t^* is suggested in the sense that the condition number of the iterate matrix could arrive at its minimum, which can be regarded as the inheritance and development of the spectral scaling quasi-Newton equation.

TTM7 is the modified HS conjugate gradient method for solving unconstrained optimization problems which generates the sufficient descent direction at each iterate as well as being close to the Newton direction. Based on some three-term conjugate gradient methods and Gram-Schmidt orthogonalization into d_k and g_k, the search direction is obtained as follows

$$d_k = \begin{cases} -g_k, & g_k^T y_{k-1} \le 0, \\ d_k^{NHS}, & \text{otherwise} \end{cases} \tag{4.4}$$

where d_k^{NHS} is an affine combination as follows

$$\begin{aligned} d_k^{NHS} &= (1-\lambda_k) d_k^{HS3} + \lambda_k \overbrace{\Big(-g_k + \beta_k^{HS} d_{k-1} - \beta_k^{HS} \frac{g_k^T d_{k-1}}{\|g_k\|^2} g_k \Big)}^{\beta_k^{HS2}} \\ &= \underbrace{\Big(-g_k + \beta_k^{HS} d_{k-1} - \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} y_{k-1} \Big)}_{d_k^{HS3}} + \lambda_k \frac{g_k^T d_{k-1}}{d_{k-1}^T y_{k-1}} \Big(y_{k-1} - \frac{g_k^T y_{k-1}}{\|g_k\|^2} g_k \Big). \end{aligned}$$

Note that the parameter λ_k is chosen such that the quasi-Newton condition $B_k s_{k-1} = y_{k-1}$ holds. Thus, the optimal parameter λ_k^* is given by

$$\lambda_k^* = \begin{cases} \|g_k\|^2 \dfrac{\|y_{k-1}\|^2 - t s_{k-1}^T y_{k-1}}{\|g_k\|^2 \|y_{k-1}\|^2 - (g_k^T y_{k-1})^2}, & k \in K, \\ 0, & k \in \mathbb{N} \backslash K, \end{cases} \tag{4.5}$$

where the index set K is

$$K = \Big\{ k \mid k \in \mathbb{N},\ 0 < \frac{g_k^T y_{k-1}}{\|g_k\| \|y_{k-1}\|} \le 1 - \eta \Big\}.$$

In the numerical experiments, the parameter η is chosen 10^{-6} and the parameter t has different values $t \in \{0.1k\}_{k=1}^{10}$. The optimal choice for this parameter is $t = 0.5$, which performs slightly better than others.

5 Numerical Experiments

In this section, we report the numerical results on a set of 142 nonlinear unconstrained optimization test problems from the CUTEst collection [8], which are given in Table 1. The initial points are standard ones proposed in CUTEst. All algorithms are implemented in Matlab 2015 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. The stop criteria for all algorithms is $\|g_k\|_\infty \le 10^{-6}$ or the total number of iterates exceeds 10000.

We use the performance profiles proposed by Dolan & Moré [12] to display the performance of each algorithm where $\mathtt{N_i}$, $\mathtt{N_f}$ and $\mathtt{C_t}$ are the total number of iterates, the total number of function evaluations and the times in second, respectively. In Dolan & Moré performance profile, the top curve corresponds to the method that solved most test problems in a time that was within a given factor of the best time. The percentage of the test problems for which a method is the fastest is given on the left axis of the plot. The right side of the plot gives the percentage of the test problems that were successfully solved by these algorithms, which is a measure of the robustness of an algorithm.

Figures 1–3 show the numerical performance of the three-term conjugate gradient algorithms. As shown by Figure 1, TTM5 outperforms respect to other methods in terms of the total number of iterates. Figure 2 shows that TTM2, TTM5 and TTM7 are robust than others in terms of the total number of function evaluations. Finally, from Figure 3, we see that TTM5 has acceptable numerical results for the times in second about 31%. As a result, TTM5 has excellent performance in comparison with the other algorithms.

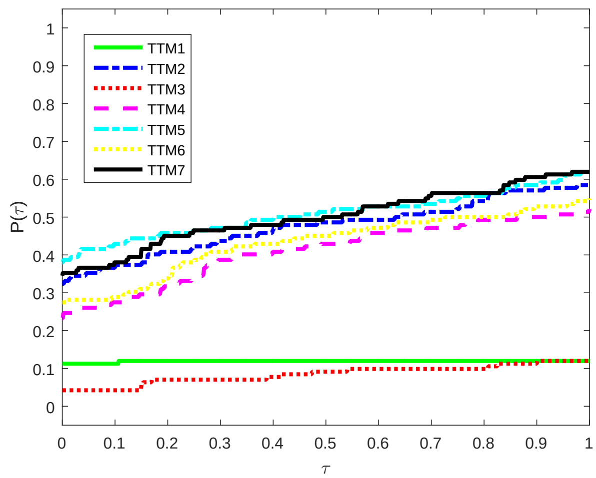

Figure 1: A comparison among TTM1, TTM2, TTM3, TTM4, TTM5, TTM6 and TTM7 by performance profiles for number of iterates (N_i).

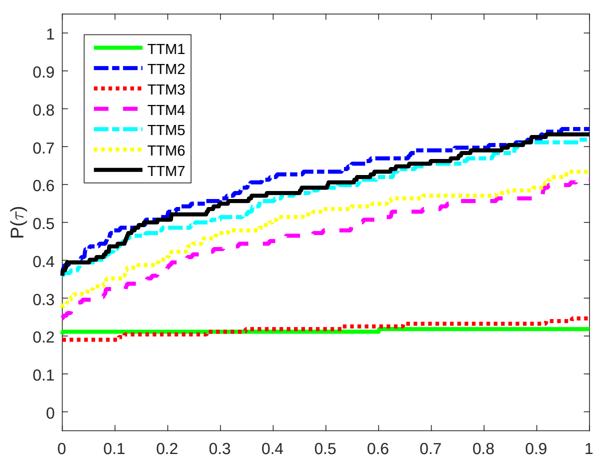

Figure 2: A comparison among TTM1, TTM2, TTM3, TTM4, TTM5, TTM6 and TTM7 by performance profiles for number of function evaluations (N_f).

Table 1: Test problems (name & dimensions); Collected by CUTEst

name	dimension	name	dimension	name	dimension
3PK	30	ALLINIT	4	ALLINITU	4
ARGLINA	500	ARGLINB	200	ARWHEAD	5000
BARD	3	BDQRTIC	100	BEALE	2
BIGGS6	6	BOX2	3	BOX3	3
BRKMCC	2	BROWNDEN	4	BROYDN3D	5000
BROYDN7D	500	BROYDNBD	5000	BRYBND	500
CHAINWOO	1000	CHNROSNB	50	CLIFF	2
COSINE	1000	CRAGGLVY	1000	CUBE	2
CUBENE	2	DALLASM	196	DALLASS	46
DECONVU	63	DENSCHNA	2	DENSCHNB	2
DENSCHNC	2	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	1000
DJTL	2	DQDRTIC	10000	DQRTIC	5000
EDENSCH	100	EG2	1000	EG3	10001
EIGENA	2550	ENGVAL1	100	ENGVAL2	3
ERRINROS	50	EXPFIT	2	EXTROSNB	1000
FLETCBV2	5000	FLETCHCR	500	FMINSRF2	5625
FMINSURF	5625	FREUROTH	2	GENHUMPS	5000
GENROSE	500	GROWTHLS	3	GULF	3
HAIRY	2	HATFLDD	3	HATFLDF	3
HATFLDFL	3	HEART6LS	6	HEART8LS	8
HELIX	3	HILBERTA	10	HILBERTB	10
HIMMELBA	2	HIMMELBC	2	HIMMELBF	4
HIMMELBG	2	HIMMELBH	2	HUMPS	2
JENSMP	2	JIMACK	3549	KOWOSB	4
LIARWHD	5000	LOGHAIRY	2	MANCINO	100
MATRIX2	6	METHANOL	12005	MOREBV	5000
MSQRTALS	1024	MSQRTBLS	1024	NINE5D	10733
NONCVXU2	1000	NONDIA	5000	NONDQUAR	5000
OSCIPANE	5000	OSCIPATH	10	OSLBQP	8
PALMER1C	8	PALMER1D	7	PALMER2C	8
PALMER3C	8	PALMER4C	8	PALMER5C	6
PALMER6C	8	PALMER7C	8	PALMER8A	6
PALMER8C	8	PENALTY1	200	PENALTY2	50
POWELLBC	1000	POWELLSG	5000	QR3DLS	610
QUARTC	5000	ROSENBR	2	S308	2
SCHMVETT	1000	SENSORS	100	SINEVAL	2
SINVALE	2	SISSER	2	SNAIL	2
SPARSINE	5000	SPARSQUR	10000	SPMSRTLS	4999
SROSENBR	10000	TAME	2	TESTQUAD	5000
TOINTGOR	50	TOINTGSS	5000	TOINTPSP	50
TOINTQOR	50	TQUARTIC	5000	TRIDIA	5000
VAREIGVL	1000	VIBRBEAM	8	WATSON	12
WEEDS	3	WOODS	1000	YFITU	3
ZANGWIL2	2				

6 Conclusion

In this paper, we compare several versions of three-term conjugate gradient methods to solve the unconstrained optimization problems. The selected three-term conjugate gradient methods satisfy the descent condition, the sufficient descent condition and the Dai–Liao con-

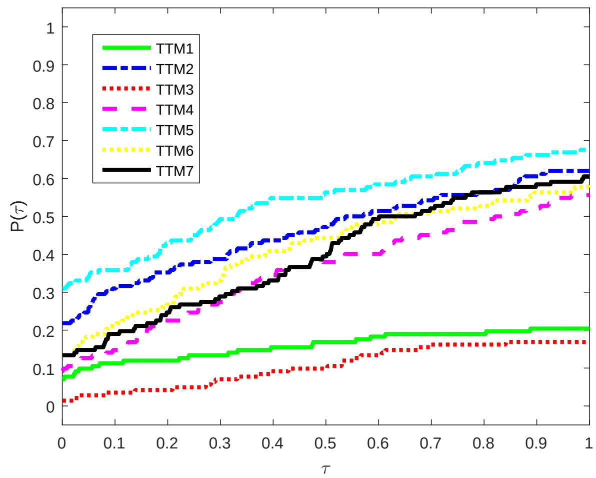

Figure 3: A comparison among `TTM1`, `TTM2`, `TTM3`, `TTM4`, `TTM5`, `TTM6` and `TTM7` by performance profiles of CPU times ($\mathtt{C_t}$).

jugacy condition. To review these methods, we use the nonlinear unconstrained optimization test problems on the `CUTEst` collection.

References

[1] M. Al-Baali, Y. Narushima and H. Yabe, A family of three-term conjugate gradient methods with sufficient descent property for unconstrained optimization, *Comput. Optim. Appl.* 60 (2015) 89–110.

[2] N. Andrei, A new three-term conjugate gradient algorithm for unconstrained optimization, *Numerical Algorithm* 68 (2015) 305–321.

[3] S. Babaie–Kafaki, A modified three-term conjugate gradient method with sufficient descent property, *Appl. Math. J. Chinese Univ.* 30(3) (2015) 263–272.

[4] S. Babaie-Kafaki, On optimality of two adaptive choices for the parameter of Dai-Liao method, *Optim. Lett.* 10(8) (2016) 1789–1797.

[5] J. Barzilai and J.M. Borwein, Two point step size gradient method, *IMA J. Numer. Anal.* 8 (1988) 141–148.

[6] E.M.L. Beale, A derivative of conjugate gradients in: Numerical Methods for Nonlinear Optimization, Lootsma, F.A (ed.) , Academic, London, 1972. pp. 39–43.

[7] E.G. Birgin and J.M. Martínez, A spectral conjugate gradient method for unconstrained optimization, *Applied Mathematics and Optimization* 43 (2001) 117–128.

[8] I. Bongartz, A.R. Conn and N.I.M. Gould, Cute: constrained and unconstrained testing environment, *ACM Trans Math Softw., Ser. A* 21 (1995) 123–160.

[9] Y.H. Dai and Y. Yuan, A nonlinear conjugate gradient method with a strong global convergence property, *IEEE SIAM J. Optim.* 10 (1999) 177–182.

[10] Y. Dai and L. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Appl. Math. Optim.* 43 (2001) 87–101.

[11] S. Deng and Z. Wan, A three-term conjugate gradient algorithm for large-scale unconstrained optimization problems, *Applied Numerical Mathematics* 92 (2015) 70–81.

[12] E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles, *Math. Program.* 91 (2002) 201–213.

[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 (2015) 606–617.

[14] X.L. Dong, D.R. Han, R. Ghanbari, X.L. Li and Z.F. Dai, Some new three-term HestenesStiefel conjugate gradient methods with affine combination, *Optimization* 66 (2017) 759–776.

[15] M. Fatemi, An optimal parameter for Dai-Liao family of conjugate gradient methods, *Optim Theory Appl.* 169(2) (2016) 1–19.

[16] R. Fletcher and C. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964) 149–154.

[17] W.W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an effcient line search, *SIAM J. Optim.* 16 (2005) 170–192.

[18] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods, This material is based upon work supported by the National Science Foundation under Grant No. 0203270 (2005).

[19] M.R. Hestenes and E.L. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Research Nat. Bur. Standards* 49 (1952) 409–436.

[20] Q. Li and D. Li, A class of derivative-free methods for large-scale nonlinear monotone equations, *IMA J. Numer. Anal.* 31 (2011) 1625–1635.

[21] J. Nocedal and S.J. Wright, *Numerical Optimization*, Springer, NewYork, 2006.

[22] E. Polyak and G. Ribière, Note surla convergence de directions conjugées, *Francaise Informat Recherche Opertionelle, 3e Année.* 16 (1969) 35–43.

[23] G. Yuan and M. Zhang, A three-terms Polak–Ribière–Polyak conjugate gradient algorithm for large-scale nonlinear equations, *Journal of Computational and Applied Mathematics* 286 (2015) 186–195.

[24] 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 (2006) 629–640.

[25] L. Zhang, A derivative-free conjugate residual method using secant condition for general large-scale nonlinear equations, *Numerical Algorithms* https://doi.org/10.1007/s11075-019-00725-7.

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

[27] L. Zhang and W. Zhou, A note on the convergence properties of the original three-term Hestenes-Stiefel method, *Adv. Model Optim.* 14 (2012) 159–163.

[28] W. Zhou and D. Li, On the Q-linear convergence rate of a class of methods for monotone nonlinear equations, *Pacific Journal of Optimization* 14 (2018) 723–737.

Manuscript received 14 May 2019
revised 19 September 2019
accepted for publication 17 December 2019

L. Arman School of Aeronautic Science and Engineering
Beihang University, Beijing, China
E-mail address: ladan.arman@gmail.com

Y. Xu
School of Aeronautic Science and Engineering
Beihang University, Beijing, China
E-mail address: Xuyng @sina.com

M. Rostami
Young Researchers and Elite Club, Hamedan Branch
Islamic Azad University, Hamedan, Iran
E-mail address: majid403rostami@yahoo.com

F. Rahpeymaii
Department of Mathematics, Payame Noor University
PO BOX 19395-3697, Tehran, Iran
E-mail address: rahpeyma_83@yahoo.com