Journal of Mathematical Modeling
Vol. 8, No. 1, 2020, pp. 1-20

JMM

A modified conjugate gradient method based on a modified secant equation

Parvaneh Faramarzi and Keyvan Amini*

Department of Mathematics, Faculty of Science, Razi University, Kermanshah, Iran
Emails: faramarzi.parvaneh@razi.ac.ir, kamini@razi.ac.ir

Abstract. Quasi-Newton methods are one of the popular iterative schemes to solve unconstrained optimization problems. The high convergence rate and excellent precision are two prominent characteristics of the quasi-Newton methods. In this paper, according to the preferable properties of a modified secant condition, a modified conjugate gradient method is introduced. The new algorithm satisfies the sufficient descent property independent of the line search. The convergence properties of the proposed algorithm are investigated both for uniformly convex and general functions. Numerical experiments show the superiority of the proposed method.

Keywords: Conjugate gradient methods, Modified secant condition, Sufficient descent condition, Global convergence.
AMS Subject Classification: 90C30, 90C06, 65K05, 90C53.

1 Introduction

The conjugate gradient (CG) methods are one of the useful iterative methods to solve the unconstrained optimization problem as follows:

$$\min f(x), \qquad x \in \mathbb{R}^n, \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, and its gradient $g(x) = \nabla f(x)$ is available. In each iteration, the CG methods only need to evaluate the objective function and its gradient. For this reason, this

*Corresponding author.
Received: 22 October 2019 / Revised: 30 November 2019 / Accepted: 8 December 2019.
DOI: 10.22124/jmm.2019.14807.1343

method is very desirable to solve large-scale unconstrained optimization problems. The iterative formula of these methods is as follows:

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

where α_k is a step length. The search direction d_k is calculated by

$$d_k = \begin{cases} -g_k, & k = 0, \\ -g_k + \beta_k d_{k-1}, & k \geq 1, \end{cases} \tag{3}$$

where β_k is a scalar describing the characteristics of the CG methods and $g_k = g(x_k)$. Various CG methods are deduced by the different choices of the conjugate parameter β_k. Some of the most famous conjugate parameters are listed as follows:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \ , \ \beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \ , \ \beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \ , \ \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},$$

which are called the Fletcher-Reeves (FR) [9], Hestenes-Stiefel (HS) [13], Polak-Ribire-Polyak (PRP) [22, 23], and Dai-Yuan (DY) [6], respectively. Note that $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ denotes the L_2 norm.

The convergence properties of these methods have been widely studied [8, 19]. Despite the appropriate numerical results of the HS and PRP methods, their convergence properties are not efficient. On the other hand, the DY and FR methods have some strong convergence characteristics, but their numerical results are weaker than the other methods [19].

Quasi-Newton methods are another prominent class of iterative techniques which have been marked due to the high convergence rate and their good accuracy. The secant equation plays a basic role in constructing the Hessian approximation, and quasi-Newton algorithms are built based on it. Despite the good properties of this equation, quasi-Newton schemes only use gradient information while their efficiency may be increased by using the function information. In this regard, many researchers modified the secant equation to improve convergence conditions of quasi-Newton schemes. They modified the secant equation to using both the objective function information and the gradient information. In the following, some of these methods are mentioned.

In 2001, Zhang and Xu [27] introduced the following modified secant equations:

$$B_k s_{k-1} = z_{k-1}; \qquad z_{k-1} = y_{k-1} + \frac{\theta_{k-1}}{s_{k-1}^T u_{k-1}} u_{k-1}, \tag{4}$$

where

$$\theta_{k-1} = 6\left(f_{k-1} - f_k\right) + 3\left(g_{k-1} + g_k\right)^T s_{k-1}, \tag{5}$$

$f_k = f(x_k)$, $s_{k-1} = x_k - x_{k-1}$, and $u_{k-1} \in \mathbb{R}^n$ is an arbitrary vector satisfying $s_{k-1}^T u_{k-1} \neq 0$. Zhang and Xu [27] established the local and super-linear convergence of the modified method (4), under some suitable conditions. The numerical results presented by Zhang and Xu [27] showed that quasi-Newton methods based on the equation (4) are comparable to the standard quasi-Newton methods. If $u_{k-1} = s_{k-1}$, then the modified secant equation (4) is an extension of the modified secant equation by Zhang et al. [26]. Yabe and Takano [25] incorporated a non-negative parameter ψ_k to the modified secant equation suggested by Zhang and Xu [27] and proposed the following equations:

$$B_k s_{k-1} = z_{k-1}, \qquad z_{k-1} = y_{k-1} + \psi_k \frac{\theta_{k-1}}{s_{k-1}^T u_{k-1}} u_{k-1}. \tag{6}$$

It is clear that the equation (6) maintains the desired properties of the equation (4).

The convergence properties of the BFGS method have been extensively investigated by some scholars for convex functions [2, 3, 16]. To improve the convergence conditions and the extension of the BFGS method for non-convex functions, Li and Fukushima [14] suggested a modified secant equation as follows

$$B_k s_{k-1} = y_{k-1}^*; \qquad y_{k-1}^* = y_{k-1} + t_{k-1} \|g_{k-1}\| s_{k-1}, \tag{7}$$

where

$$t_{k-1} = 1 + \max\left\{ \frac{-y_{k-1}^T s_{k-1}}{\|s_{k-1}\|^2}, 0 \right\}.$$

They established the local and super-linear convergence of the BFGS method for non-convex functions. In 2006, based on (7), Zhou and Zhang [28] introduced the following modified secant equation

$$B_k s_{k-1} = z_{k-1}; \qquad z_{k-1} = y_{k-1} + h_{k-1} \|g_{k-1}\|^r s_{k-1}, \tag{8}$$

where

$$h_{k-1} = C + \max\left\{ \frac{-y_{k-1}^T s_{k-1}}{\|s_{k-1}\|^2}, 0 \right\} \|g_{k-1}\|^{-r}.$$

The modified secant equation (8) improves the numerical efficiency and convergence conditions of the quasi-Newton methods.

On the other hand, the efficiency of the CG methods can be improved by using the second-order information of the objective function. Dai and

Liao [5] was a pioneer this idea and presented the CG method with the following parameter based on the standard secant equation:

$$\beta_k^{DL} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}},$$

where $t > 0$ is a parameter. They proposed the following modification of β_k^{DL} to ensure global convergence for general functions:

$$\beta_k^{DL+} = \max\left\{\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, 0\right\} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}.$$

They showed that DL+ method has suitable convergence properties and generates appropriate results in comparison to PRP and HS methods.
According to the modified secant equation (8), Similar to the idea of Dai and Liao [5], Zhou and Zhang [28] presented the CG with the following parameter as follows:

$$\beta_k^{ZZ} = \frac{g_k^T z_{k-1}}{d_{k-1}^T z_{k-1}} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T z_{k-1}}.$$

Many researchers have been presented the CG methods satisfying the sufficient descent property, see for example [1, 4, 12, 17, 18]. Although sufficient descent condition is an extremely important property in proving the global convergence of the CG methods, some CG methods lack it. The search direction d_k fulfills the sufficient descent property whenever there exists a constant $c_1 > 0$ such that

$$g_k^T d_k \leq -c_1\|g_k\|^2, \quad \text{for all} \quad k \geq 0.$$

Hager and Zhang [12] modified the well-known conjugate parameter β_k^{HS} as follows:

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

which called CG-DESCENT method and is one of the best-known CG methods both in terms of theoretical features and practical performances. They established the global convergence for general functions by considering the conjugate parameter as follows:

$$\beta_k^{HZ+} = \max\left\{\beta_k^{HZ}, \eta_k\right\}; \qquad \eta_k = \frac{-1}{\|d_k\|\min\{\eta, \|g_k\|\}}, \tag{9}$$

where $\eta > 0$ is a constant.

To deduce an effective family of the CG methods, Dai and Kou [4] paid attention to the self-scaling memoryless BFGS method presented by Perry [21] and Shanno [24], in which the search direction is defined by

$$d_{k+1} = -H_{k+1} g_{k+1}, \tag{10}$$

where

$$H_{k+1} = \frac{1}{\tau_k}\left(I - \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k}\right) + \left(1 + \frac{1}{\tau_k}\frac{\|y_k\|^2}{s_k^T y_k}\right)\frac{s_k s_k^T}{s_k^T y_k}, \tag{11}$$

and τ_k is a scaling parameter. Setting (11) in (10), the search direction with a different multiplier is computed as follows:

$$d_{k+1}^{PS} = -g_{k+1} + \left[\frac{g_{k+1}^T y_k}{s_k^T y_k} - \left(\tau_k + \frac{\|y_k\|^2}{s_k^T y_k}\right)\frac{g_{k+1}^T s_k}{s_k^T y_k}\right] s_k + \frac{g_{k+1}^T s_k}{s_k^T y_k} y_k. \tag{12}$$

The idea of Dai and Kou [4] was to seek the closest two-term direction to the search direction (12), and they obtained the following conjugate parameter:

$$\beta_k^{DK}(\tau_k) = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \left(\tau_k + \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}} - \frac{s_{k-1}^T y_{k-1}}{\|s_{k-1}\|^2}\right)\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}.$$

Dai and Kou [4] used the following scaling parameter by Oren and Luenberger [20]

$$\tau_k = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}.$$

By this, the following conjugate parameter results

$$\beta_k^{DK} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} - \frac{\|y_{k-1}\|^2}{s_{k-1}^T y_{k-1}}\frac{g_k^T s_{k-1}}{d_{k-1}^T y_{k-1}}. \tag{13}$$

Based on the following conjugate parameter, they proved the global convergence of this method

$$\beta_k^{DK+} = \max\left\{\beta_k^{DK}(\tau_k), \eta\frac{g_k^T d_{k-1}}{\|d_{k-1}\|^2}\right\}, \tag{14}$$

where $\eta \in [0,1)$ is a parameter. Numerical experiments showed that this algorithm is one of the most efficient CG methods.

In this regard, similar to the idea of Dai and Kou [4], Amini et al. [1] sought the closest two-term direction to three-term direction of Narushima et al. [18] and introduced the following parameter:

$$\beta_k^{MHS} = \beta_k^{HS}\left(1 - \frac{(g_k^T d_{k-1})^2}{\|g_k\|^2\|d_{k-1}\|^2}\right).$$

They proved the global convergence of this method for general functions by considering the following truncated form:

$$\bar{\beta_k}^{MHS+} = \max\left\{\bar{\beta_k}^{MHS}, 0\right\},$$

with

$$\bar{\beta_k}^{MHS} = \beta_k^{MHS} - \lambda \left(\frac{\|y_{k-1}\|\theta_k}{d_{k-1}^T y_{k-1}}\right)^2 g_k^T d_{k-1}; \quad \theta_k = 1 - \frac{(g_k^T d_{k-1})^2}{\|g_k\|^2 \|d_{k-1}\|^2},$$

where $\lambda > \frac{1}{4}$ is a parameter. The numerical experiments showed this algorithm is effective in comparison with other algorithms.

In this study, using a modified secant equation, we improve the CG method by Dai and Kou [4] and present a modified CG method. The new search direction satisfies the sufficient descent property independent of the line search and the convexity assumption on the objective function. The new algorithm has the global convergence for general functions, under the standard assumptions. Numerical experiments indicated that the new algorithm is preferable, dealing with unconstrained optimization problems.

This paper is organized as follows. In the next section, after scheming the new idea, some basic properties of the proposed method are proven. In the third section, global convergence of the new algorithm is investigated both for general functions and uniformly convex functions. In Section 4, numerical experiments of the new method are evaluated. Some conclusions are given in the final section.

2 New algorithm

In this section, according to a modified secant equation, a modified CG method is proposed.

Quasi-Newton methods are one of the famous classes to solve the problem (1). We recall that both the modified secant equations (4) and (6) approximate the Hessian matrix with the high accuracy in comparison with the standard secant equation. Inspired by the idea of Dai and Liao [5], Yabe and Takano [25] proposed the following conjugate parameter:

$$\beta_k^{YT} = \frac{g_k^T z_{k-1}}{d_{k-1}^T z_{k-1}} - t\frac{g_k^T s_{k-1}}{d_{k-1}^T z_{k-1}},$$

where z_{k-1} is defined by (6). They showed that the numerical results of this method are improved by using the modified secant equation (6).

Motivated by the idea of Yabe and Takano [25], strong theory features of the equation (6), and considering high efficiency of the DK method [4], we propose the new conjugate parameter β_k^{MDK} as follows:

$$\beta_k^{MDK} = \frac{g_k^T y_{k-1}}{d_{k-1}^T z_{k-1}} - \frac{\|y_{k-1}\|^2}{d_{k-1}^T z_{k-1}} \frac{g_k^T d_{k-1}}{d_{k-1}^T z_{k-1}}, \tag{15}$$

where z_{k-1} is defined by (6). This formula is an extension of DK CG parameter [4]. The relation (15) is resulted by replacing vector y_{k-1} with z_{k-1} in the denominator of the relation (13).
Because $d_{k-1}^T z_{k-1}$ may turn out to zero for a general function f, so the formula (15) may not well-defined. To overcome this drawback, similar to Li et al. [15]; we consider a modified version of (6) as follows:

$$B_k s_{k-1} = z_{k-1}^+; \qquad z_{k-1}^+ = y_{k-1} + \psi_k \frac{\max\{0, \theta_{k-1}\}}{s_{k-1}^T u_{k-1}} u_{k-1}, \tag{16}$$

with θ_{k-1} defined by (5).

The following lemma shows that the new direction satisfies the sufficient descent property independent of the line search.

Lemma 1. *Consider the CG method (2), (3) with (15) . If $d_{k-1}^T z_{k-1} \neq 0$, then we have*

$$g_k^T d_k \leq -\frac{3}{4}\|g_k\|^2. \tag{17}$$

Proof. Since $d_0 = -g_0$, we have

$$g_0^T d_0 = -\|g_0\|^2 \leq -\frac{3}{4}\|g_0\|^2,$$

which implies that (17) holds for $k = 0$. For $k \geq 1$, multiplying (3) by g_k^T, we have

$$\begin{aligned} g_k^T d_k &= -\|g_k\|^2 + \left[\frac{g_k^T y_{k-1}}{d_{k-1}^T z_{k-1}} - \frac{\|y_{k-1}\|^2}{d_{k-1}^T z_{k-1}} \frac{g_k^T d_{k-1}}{d_{k-1}^T z_{k-1}}\right] g_k^T d_{k-1} \\ &= \frac{\left(g_k^T y_{k-1}\right)\left(d_{k-1}^T z_{k-1}\right)\left(g_k^T d_{k-1}\right) - \|g_k\|^2 \left(d_{k-1}^T z_{k-1}\right)^2 - \|y_{k-1}\|^2 \left(g_k^T d_{k-1}\right)^2}{\left(d_{k-1}^T z_{k-1}\right)^2}. \end{aligned} \tag{18}$$

Applying the inequality $u_k^T v_k \leq \frac{1}{2}\left(\|u_k\|^2 + \|v_k\|^2\right)$ to the first term (18) with

$$u_k = \frac{1}{\sqrt{2}}(d_{k-1}^T z_{k-1}) g_k, \qquad v_k = \sqrt{2}(g_k^T d_{k-1}) y_{k-1},$$

we obtain (17). $\square$

The new method with $\theta_{k-1} = 0$ reduces to the HS method. It is expected that the convergence properties of the proposed parameter (15) be similar to the HS method. Therefore, according to the idea of Gilbert and Nocedal [10] to prove the global convergence of the HS method for general objective functions, we consider the following truncating

$$\beta_k^{MDK+} = \max\left\{0, \beta_k^{MDK}\right\}. \tag{19}$$

Using this truncating, the sufficient descent condition still holds for the new search direction. Based on the above, the new algorithm is presented for the proposed method.

Algorithm 1 (MDK algorithm)

Input. Given constant $\epsilon > 0$. Chosen $x_0 \in \mathbb{R}^n$ and set $k = 0$, $d_0 = -g_0$.
Step 1. If $\|g_k\|_\infty \leq \epsilon$, stop.
Step 2. Compute d_k by (3) and (19).
Step 3. Determine the step length α_k by an appropriate line search.
Step 4. Set $x_{k+1} = x_k + \alpha_k d_k$.
Step 5. Set $k = k + 1$ and go to Step 1.

---

3 Global convergence properties

In this part, the global convergence of the MDK algorithm is analyzed. Therefore, the following assumptions will be needed through this study. Without loss of generality, suppose $g_k \neq 0$ for all $k \geq 0$; otherwise, we assume that a stationary point can be found.

Assumption 1
(A1) The level set $L_0 = \{x | f(x) \leq f(x_0)\}$ is bounded, namely, there exists a constant $B > 0$ such that

$$\|x\| \leq B, \quad \text{for all} \ \ x \in L_0. \tag{20}$$

(A2) In some open neighborhood $N \in L_0$, f is continuously differentiable and its gradient g is Lipschitz continuous, namely, there exists a positive constant L such that

$$\|g(x) - g(y)\| \leq L\,\|x - y\|, \quad \text{for all} \ \ x, y \in N. \tag{21}$$

Assumption 1 conclude that there is a constant $\gamma > 0$ such that

$$\|g(x)\| \leq \gamma, \quad \text{for all} \ \ x \in L_0. \tag{22}$$

To guarantee the global convergence of a nonlinear CG method, we need to impose a line search on it. Suppose that the step length α_k fulfills the Wolfe conditions, i.e.,

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{23}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{24}$$

where $0 < \delta \leq \sigma < 1$. Note that the Wolfe conditions and (17) yield

$$d_{k-1}^T z_{k-1}^+ \geq d_{k-1}^T y_{k-1} \geq (\sigma - 1)\, g_{k-1}^T d_{k-1} \geq \frac{3}{4}\,(1-\sigma)\,\|g_{k-1}\|^2 > 0,$$

which implies that (15) and (19) are well defined.

The next lemma, called Zoutendijk condition [29], plays an important role to establish the MDK algorithm.

Lemma 2. *[29] Suppose that Assumption 1 holds. Consider any CG method in the forms (2)-(3), where step length α_k satisfies the Wolfe line search and the search direction d_k is decreasing. If*

$$\sum_{k=0}^{\infty} \frac{1}{\|d_k\|^2} = \infty, \tag{25}$$

then,

$$\liminf_{n \longrightarrow \infty} \|g_k\| = 0. \tag{26}$$

Lemma 3. *[25] Suppose that Assumption 1 holds. For θ_{k-1} given by (5), we have*

$$|\theta_{k-1}| \leq 3L\|s_{k-1}\|^2. \tag{27}$$

Proof. This relation immediately is resulted from the relation (5.12) from [25]. □

Subsequently, we establish the global convergence of the MDK algorithm for uniformly convex functions.

Theorem 1. *Suppose that f is uniformly convex, i.e., there exists a constant $\mu > 0$, such that*

$$(g(x) - g(y))^T (x - y) \geq \mu\|x - y\|^2, \qquad \forall x, y \in \mathbb{R}^n. \tag{28}$$

Also, Assumption 1 holds and ψ_k satisfies $0 \leq \psi_k \leq \bar{\rho}$ where $\bar{\rho}$ is a positive constant such that $\bar{\rho} < \dfrac{\mu}{3L}$. Let $\{x_k\}$ be generated by Algorithm 1, then either $\|g_k\| = 0$ for some k or

$$\liminf_{n \longrightarrow \infty} \|g_k\| = 0.$$

Proof. Due to the sufficient descent condition (17), we get $d_k \neq 0$. By considering Lemma 2, it suffices to show that $\|d_k\|$ is bounded above. The relations (6), (27), and (28) conclude that

$$\begin{aligned} |d_{k-1}^T z_{k-1}| =& |d_{k-1}^T y_{k-1} + \psi_k \frac{\theta_{k-1}}{s_{k-1}^T u_{k-1}} d_{k-1}^T u_{k-1}| \\ \geq & |d_{k-1}^T y_{k-1}| - \frac{3\psi_k L \|s_{k-1}\|^2}{\alpha_{k-1}} \\ \geq & \left(\mu - 3\bar{\rho} L\right) \alpha_{k-1} \|d_{k-1}\|^2. \end{aligned} \tag{29}$$

Because $\mu - 3\bar{\rho}L > 0$, we obtain by this, (15),(21), and (29) that

$$\begin{aligned} |\beta_k^{MDK}| =& |\frac{g_k^T y_{k-1}}{d_{k-1}^T z_{k-1}} - \frac{\|y_{k-1}\|^2}{d_{k-1}^T z_{k-1}} \frac{g_k^T d_{k-1}}{d_{k-1}^T z_{k-1}}| \\ \leq & \frac{L\|g_k\| \|s_{k-1}\|}{\left(\mu - 3\bar{\rho}L\right) \alpha_{k-1} \|d_{k-1}\|^2} + \frac{L^2 \|s_{k-1}\|^2 \|g_k\| \|d_{k-1}\|}{\left(\mu - 3\bar{\rho}L\right)^2 \alpha_{k-1}^2 \|d_{k-1}\|^4} \\ =& A \frac{\|g_k\|}{\|d_{k-1}\|}, \end{aligned}$$

where

$$A = \frac{L}{\mu - 3\bar{\rho}L} + \frac{L^2}{\left(\mu - 3\bar{\rho}L\right)^2}.$$

This relation together with (3) and (22) yield

$$\|d_k\| \leq \|g_k\| + |\beta_k^{MDK}| \|d_{k-1}\| \leq (1 + A)\, \gamma < \infty.$$

This completes the proof. □

Property (*) plays an important role to prove the global convergence of the CG methods. In the following, this property is expressed.

Property (*). Consider the CG method (2)-(3) and suppose that there exist positive constants γ and $\bar{\gamma}$ such that $0 < \bar{\gamma} \leq \|g_k\| \leq \gamma$, for all $k \geq 1$. Then, the method has property (*) if there exist constants $b > 1$ and $\nu > 0$ such that for all k

$$|\beta_k| \leq b,$$

and

$$\|s_{k-1}\| \leq \nu \Longrightarrow |\beta_k| \leq \frac{1}{2b}.$$

Property (*) has been first formally stated by Gilbert and Nocedal [10].

Lemma 4. *Suppose that Assumption 1 holds. Also, ψ_k satisfies $0 \le \psi_k \le \bar{\rho}$ where $\bar{\rho}$ is a fixed positive constant. Then, the new algorithm with (19) satisfies property (*).*

Proof. We assume that there exists a positive constant $\bar{\gamma}$ such that

$$\bar{\gamma} \le \|g_k\|, \tag{30}$$

holds for any $k \ge 0$. Because $\{x_k\} \subset L_0$ and the level set L_0 is bounded, from (20), we have

$$\|s_{k-1}\| < 2B. \tag{31}$$

From (17), (24), and (30), we conclude that

$$d_{k-1}^T y_{k-1} \ge \frac{3}{4}(1-\sigma)\|g_{k-1}\|^2 \ge \frac{3}{4}(1-\sigma)\bar{\gamma}^2. \tag{32}$$

Using relation $g_{k-1}^T d_{k-1} < 0$ and (24), result

$$g_k^T d_{k-1} \le g_k^T d_{k-1} - g_{k-1}^T d_{k-1} = d_{k-1}^T y_{k-1},$$

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

Since $\sigma \in (0,1)$ and $d_{k-1}^T y_{k-1} > 0$, we get

$$|g_k^T d_{k-1}| \le \max\left\{1, \frac{\sigma}{1-\sigma}\right\} d_{k-1}^T y_{k-1} = c_1 d_{k-1}^T y_{k-1}, \tag{33}$$

where $c_1 = \max\left\{1, \frac{\sigma}{1-\sigma}\right\}$. Using (16), (32) and $\psi_k \ge 0$, we conclude that

$$\begin{aligned} |d_{k-1}^T z_{k-1}^+| &= |d_{k-1}^T y_{k-1} + \frac{\psi_k}{\alpha_{k-1}} \max\{0, \theta_{k-1}\}| \\ &\ge d_{k-1}^T y_{k-1} \ge \frac{3}{4}(1-\sigma)\bar{\gamma}^2. \end{aligned} \tag{34}$$

From (33) and (34), we have

$$|g_k^T d_{k-1}| \le c_1 |d_{k-1}^T y_{k-1}| \le c_1 |d_{k-1}^T z_{k-1}^+|. \tag{35}$$

Therefore, from (31), (34) and (35), we obtain

$$\begin{aligned} |\beta_k^{MDK+}| &\le \frac{\|g_k\|\|y_{k-1}\|}{|d_{k-1}^T z_{k-1}^+|} + \frac{\|y_{k-1}\|^2 |g_k^T d_{k-1}|}{\left(d_{k-1}^T z_{k-1}^+\right)^2} \\ &\le \frac{\gamma L \|s_{k-1}\|}{\frac{3}{4}(1-\sigma)\bar{\gamma}^2} + \frac{c_1 L^2 \|s_{k-1}\|^2}{\frac{3}{4}(1-\sigma)\bar{\gamma}^2} \\ &\le \frac{\gamma L \|s_{k-1}\|}{\frac{3}{4}(1-\sigma)\bar{\gamma}^2} + \frac{2Bc_1 L^2}{\frac{3}{4}(1-\sigma)\bar{\gamma}^2}\|s_{k-1}\| \\ &= c_2 \|s_{k-1}\|, \end{aligned}$$

where

$$c_2 = \frac{4\gamma L}{3\,(1-\sigma)\,\bar{\gamma}^2} + \frac{8BL^2 c_1}{3\,(1-\sigma)\,\bar{\gamma}^2}.$$

This relation implies that $|\beta_k^{MDK+}| \leq c_2 \|s_{k-1}\|$, for all k. By putting $\nu = \frac{1}{2bc_2}$, we obtain the desired result. $\square$

The following theorem corresponds to Theorem 4.3 in [10], and we mention it for readability.

Theorem 2. *[10] Suppose that Assumption 1 holds. Consider the method (2)-(3) which satisfies the following conditions:*
(C_1) $\beta_k \geq 0$ for all $k \geq 0$;
(C_2) the Zoutendijk condition;
(C_3) the sufficient descent condition;
(C_4) property (*);
then $\liminf_{k \longrightarrow \infty} \|g_k\| = 0$.

Now, we can provide the global convergence result for MDK algorithm.

Theorem 3. *Suppose that all conditions of Lemma 4 hold. Then, MDK algorithm with (19) is globally convergent, i.e.,*

$$\liminf_{n \longrightarrow \infty} \|g_k\| = 0.$$

Proof. It is sufficient to show that conditions of Theorem 2 are held. From (19), clearly, $\beta_k^{MDK+} \geq 0$ and the MDK algorithm with β_k^{MDK+} satisfies the sufficient descent condition. The relation (21) and Wolfe conditions get the Zoutendijk condition. Hence, conditions (C_1), (C_2), and (C_3) of the Theorem 2 are satisfied. The previous lemma provides property (*) for the MDK algorithm; hence (C_4) is satisfied. This completes the proof. $\square$

4 Numerical experiments

In this section, we report the numerical performance of the new algorithm. To investigate the numerical efficiency of the MDK algorithm, we compare this algorithm with the HZ+ method, developed by Hager and Zhang [12], and the DK+ method, proposed by Dai and Kou [4], on a collection of the 119 test problems of the CUTEst library [11], with dimensions between 2 to 20000. The codes were written in MATLAB 8.1 and were implemented on a Vaio Laptop with 2.30 GHz of CPU, 4 GB of RAM. The line search in all the algorithms is the strong Wolfe line search proposed by Nocedal

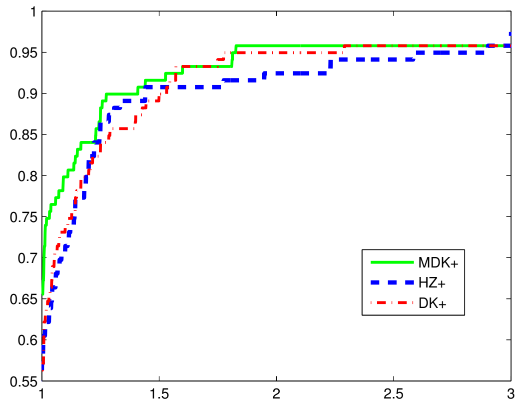

Figure 1: Performance profile for the number of iterations.

and Wright [19] with $\delta = 0.01$ and $\sigma = 0.1$, in which the initial step length selected as follows:

$$\alpha_k^{(0)} = \begin{cases} 1, & k = 1, \\ \alpha_{k-1}\frac{g_{k-1}^T d_{k-1}}{g_k^T d_k}, & k > 1. \end{cases}$$

All algorithms are terminated when

$$\|g_k\|_\infty \leq 10^{-6},$$

or the number of iterations exceeds 10000, and show it with the symbol "Failed" in Table 1. In the figures, the curves have the following meaning:
MDK+ : The CG method with β_k^{MDK+} given by (19), $\psi_k = 0.6$ and $u_{k-1} = y_{k-1}$.
HZ+ : The CG method proposed by Hager and Zhang [12] with β_k^{HZ+} given by (9), $\eta = 0.01$.
DK+ : The CG method proposed by Dai and Kou [4] with β_k^{DK+} given by (14), $\eta = 0.5$ and

$$\tau_k = \frac{s_{k-1}^T y_{k-1}}{s_{k-1}^T s_{k-1}}.$$

The numerical results of the mentioned algorithms are presented in Table 1. This table includes the name of the test problems (Prob), their dimensions (DIM), the total number of iterations (k), the total number of function evaluations (Nf), and the total number of gradient evaluations (Ng). Based on the performance profile proposed by Dolan and Moré [7],

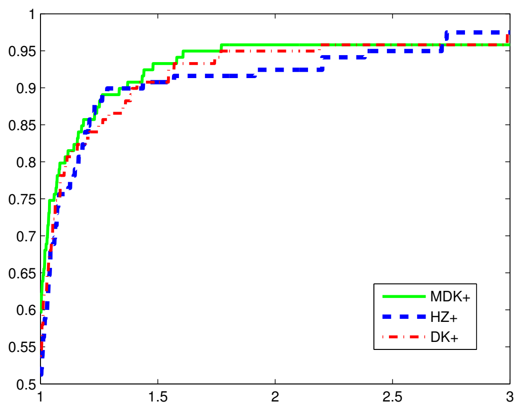

Figure 2: Performance profile based on the number of function evaluations.

the numerical results of the algorithms are compared, while including the number of iterations, the number of function evaluations, and the number of gradient evaluations, respectively.

The preliminary experiments demonstrate the efficiency of the new algorithm. In details, Figure 1 shows that "MDK+" has the best performance among all algorithms, which solves about 65% of the test problems with the least number of iterations, while "HZ+" and "DK+" solve almost 56% of the test problems. Figure 2 shows that "MDK+" solves 59% of the test problems with the least number of function evaluations, where "HZ+" and "DK+" solve almost 51% and 53% of the test problems, respectively. Figure 3 is used to report the number of gradient evaluations, where the "MDK+" solves 59% of the test problems with the least number of gradient evaluations, while "HZ+" and "DK+" can solve about 54% and 52%, respectively. Hence, the numerical results indicate that the "MDK+" performs better than the other algorithms (HZ+, DK+).

Table 1. Numerical results.

Prob	Dim	MDK+ k/Nf/Ng	HZ+ k/Nf/Ng	DK+ k/Nf/Ng
AIRCRFTB	8	43/201/168	45/203/165	35/171/138
ALLINITU	4	9/41/28	11/50/36	13/57/40
ARGLINA	200	1/4/3	1/4/3	1/4/3
ARGLINB	10	2/7/5	2/7/5	2/7/5
ARGLINC	10	2/7/5	2/7/5	2/7/5
ARWHEAD	500	9/81/58	9/84/61	9/84/61
BARD	3	24/97/81	62/231/198	28/110/95
BDEXP	5000	2/6/5	2/6/5	2/6/5
BEALE	2	12/71/53	12/60/45	12/62/47
BIGGS3	6	165/673/600	368/1481/1349	510/2011/1836
BIGGS5	6	165/673/600	368/1481/1349	510/2011/1836

Table 1. Numerical results. (*continued*)

BIGGS6	6	133/590/533	142/614/556	150/640/574
BIGGSB1	5000	Failed	Failed	Failed
BOX	100	9/41/35	9/40/34	9/40/34
BOX2	3	14/55/43	42/149/132	22/85/69
BOX3	3	14 /55/43	42/149/132	22/85/69
BRKMCC	2	7/22/15	6/19/13	7/22/15
BROWNBS	2	16/63/48	17/72/54	13/57/42
BROYDN7D	500	207/637/604	204/630/594	206/647/608
BRYBND	5000	64/247/205	40/185/136	40/185/136
CHAINWOO	1000	410/1644/1257	326/1250/1043	343/1199/1092
CHNROSNB	50	298/976/934	273/902/854	291/959/916
COSINE	1000	9/40/25	9/35/24	9/35/24
CRAGGLVY	4	38/165/138	37/165/138	33/142/120
CUBE	2	21/171/133	24/153/116	27/185/145
DECONVU	63	1807/5602/5550	1178/3679/3616	998/3164/3098
DENSCHNA	2	9/50/36	9/50/36	9/50/36
DENSCHNB	2	7/39/27	7/39/27	7/39/27
DENSCHND	3	20/479/235	24/298/155	35/527/285
DENSCHNE	3	12/116/85	13/149/114	15/158/121
DENSCHNF	2	10/66/45	9/64/44	10/67/46
DIXMAANA	3000	6/41/28	6/41/28	6/41/28
DIXMAANB	3000	5/54/37	5/54/37	5/54/37
DIXMAANC	3000	6/67/45	6/67/45	6/65/44
DIXMAAND	3000	7/73/49	7/73/49	7/71/48
DIXMAANE	3000	264/695/682	264/695/682	264/695/682
DIXMAANF	3000	205/571/551	205/571/551	204/567/547
DIXMAANG	3000	200/571/549	200/571/549	202/572/549
DIXMAANH	3000	200/583/555	200/583/555	189/545/518
DIXMAANI	3000	4260/10511/10499	4732/11758/11746	4410/11067/11055
DIXMAANJ	3000	360/902/883	360/902/883	369/912/893
DIXMAANK	3000	313/806/784	313/806/784	326/828/804
DIXMAANL	3000	290/766/741	290/766/741	286/752/726
DIXMAANM	15	57/192/166	57/192/166	57/192/166
DIXMAANN	15	73/240/211	72/239/207	72/239/207
DIXMAANO	15	54/192/161	55/195/163	55/195/163
DIXMAANP	3000	1704/4238/4219	3320/8111/8092	2432/5980/5961
DIXON3DQ	1000	6757/16526/16522	7158/17513/17509	7122/17385/17381
DQDRTIC	5000	5/22/19	5/22/19	5/22/19
DQRTIC	5000	16/281/188	16/281/188	16/281/188
EDENSCH	36	23/117/82	24/121/86	24/121/86
EG2	1000	2/20/7	2/20/7	2/20/7
ENGVAL1	100	25/122/82	25/122/82	22/113/80
ENGVAL2	3	45/265/226	2319/9078/8634	431/1801/1659
ERRINROS	10	Failed	593/2404 /2158	494/1989/1801
EXTROSNB	1000	Failed	9459/31456/30249	8304/27547/26544
FLETCHCR	1000	4454/11659/11650	4412/11489/11476	4429/11567/11555
FMINSRF2	5625	362/909/909	365/915/915	365/915/915
FMINSURF	5625	557/1383/1383	554/1358/1358	554/1358/1358
FREUROTH	2	12/90/70	13/98/76	11/87/67
GENHUMPS	5000	9358/28365/28096	9407/28628/28400	9424/28706/28382
GENROSE	500	1091/3382/3325	1092/3454/3290	1099/3402/3333
GROWTHLS	3	1/21/3	1/21/3	1/21/3
GULF	3	201/816/753	1031/4116/3840	308/1422/1275
HATFLDD	3	52/272/235	65/312/269	52/262/227
HATFLDFL	3	31/147/128	103/493/442	40/186/167
HEART6LS	6	Failed	Failed	Failed
HEART8LS	8	5030/15605/15467	811/2940/2731	1136/3667/3543

Table 1. Numerical results. (*continued*)

HELIX	3	31/137/114	55/215/180	55/215/180
HILBERTA	2	2/7/6	2/7/6	2/7/6
HILBERTB	10	4/13/9	4/13/9	4/13/9
HIMMELBG	2	7/36/26	8/35/26	8/34/25
HIMMELBH	2	8/29/20	7/28/19	8/31/21
HUMPS	2	29/213/155	16/144/99	24/186/122
KOWOSB	4	65/269/244	148/587/530	51/215/196
LIARWHD	5000	16/129/99	19/138/107	20/140/108
LOGHAIRY	2	53/420/308	29/314/210	35/293/221
LMINSURF	5625	350/878/878	347/872/872	347/872/872
MANCINO	100	11/34/23	11/34/23	11/34/23
MODBEALE	20000	554/2317/2025	798/2734/2584	Failed
MOREBV	5000	167/392/390	167/392/390	167/392/390
MSQRTALS	1024	4611/11430/11419	4588/11279/11268	4565/11231/11220
MSQRTBLS	1024	2994/7335/7324	2970/7309/7298	2995/7359/7348
NLMSURF	5625	392 985 983	404/1027/1024	404/1027/1024
NONCVXU2	5000	Failed	Failed	Failed
NONDIA	5000	10/96/72	11/102/78	11/102/78
NONDQUAR	5000	7191/27269/25533	2724/10760/9838	1890/7493/6845
NONSCOMP	5000	43/177/125	43/179/126	43/179/126
OSCIPATH	10	Failed	Failed	Failed
OSBORNEB	11	399/1384/1311	277/961/886	297/1042/950
PALMER5C	6	6/19/14	6/19/14	6/19/14
PENALTY1	100	27/287/212	29/278/204	26/269/198
PENALTY2	50	1608/6119/5679	384/1649/1485	879/3609/3184
POWELLSG	5000	139/572/508	91/362/313	308/1317/1161
POWER	100	36/203/149	36/203/149	36/203/149
QUARTC	5000	16/281/188	16/281/188	16/281/188
ROSENBR	2	27/158/127	33/186/150	31/171/135
S308	2	7/69/46	7/69/46	7/69/46
SCHMVETT	100	43/135/109	43/135/109	43/135/109
SENSORS	10	20/76/56	20/79/60	20/80/60
SINEVAL	2	51/246/212	52/282/235	50/271/225
SINQUAD	5	31/115/95	22/91/74	34/126/104
SISSER	2	5/76/51	5/76/51	5/76/51
SNAIL	2	5/27/16	6/28/18	6/26/16
SPARSINE	5000	Failed	Failed	Failed
SPARSQUR	10000	15/199/136	15/199/136	15/199/136
SPMSRTLS	4999	224/609/601	234/647/638	234/647/638
SROSENBR	5000	11/71/52	11/71/52	11/73/53
TESTQUAD	5000	6794/18025/18018	7889/20705/20698	6336/16829/16822
TOINTGSS	5000	5/16/11	5/16/11	4/13/9
TOINTQOR	50	29/88/64	30/89/64	29/88/64
TQUARTIC	5000	15/88/72	16/106/90	12/98/84
TRIDIA	5000	2398/5741/5736	2381/5717/5712	2415/5785/5780
VARDIM	200	16/586/325	16/586/325	16/586/325
VAREIGVL	50	23/71/48	23/71/48	23/71/48
WATSON	12	1475/5684/5212	1429/5786/5261	1739/6782/6203
WOODS	4000	226/783/747	243/824/783	243/824/783
YFITU	3	1176/4416/4057	946/3560/3249	1322/4811/4514
ZANGWIL2	2	1/4/4	1/4/4	1/4/4

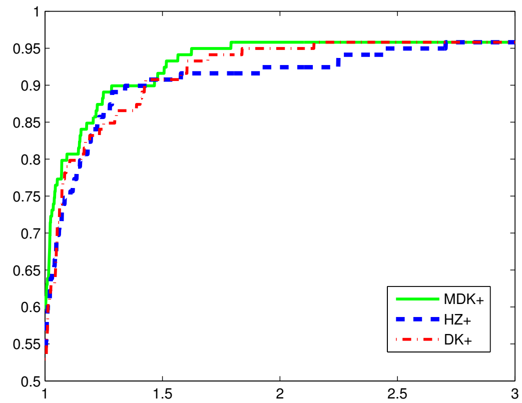

Figure 3: Performance profile based on the number of gradient evaluations.

5 Conclusion

The modified secant equations by Zhang et al. [26] and Zhang and Xu [27] possess strong theoretical properties. The modified secant equation approximates the Hessian matrix with high accuracy. In this study, regarding the strong theory features of a modified secant equation, we decided to modify the DK method [4] by using it, then a new CG method is presented. The new search direction always satisfies the sufficient descent condition, and the proposed method has the global convergence for general functions. Numerical experience on the CUTEst collection shows the MDK algorithm is capable.

References

[1] K. Amini, P. Faramarzi and N. Pirfalah, *A modified Hestenes-Stiefel conjugate gradient method with an optimal property*, Optim. Methods. Softw. **34**(4) (2019), 770-782.

[2] R. Byrd and J. Nocedal, *A tool for the analysis of quasi-Newton methods with application to unconstrained minimization*, SIAM J. Numer. Anal. **26** (1989) 727–739.

[3] Y.H. Dai, *Convergence properties of the BFGS algorithm*, SIAM J. Optim. **13** (2002) 693–701.

[4] Y.H. Dai and C.X. Kou, *A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search*, SIAM J. Optim. **23** (2013) 296-320.

[5] Y.H. Dai and L.Z. Liao, *New conjugacy conditions and related conjugate gradient methods*, Appl. Math. Optim. **43** (2001) 87-101.

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

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

[8] R. Fletcher, *Practical Methods of Optimization*, John Wiley and Sons, New york, 1987.

[9] R. Fletcher, C.M. Reeves, *Function minimization by conjugate gradient methods*, J. Comput. **7** (1964) 149-154.

[10] J.C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. **2** (1992) 21-42.

[11] N.I.M. Gould, D. Orban, P.L. Toint, *CUTEr and SifDec: A constrained and unconstrained testing environment*, revisited. ACM Trans. Math. Softw. **29** (2003) 373-394.

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

[13] M.R. Hestenes and E. Stiefel, *Method of conjugate gradient for solving linear systems*, J. Res. Nat. Bur. Stand. **49** (1952) 409-436.

[14] D.H. Li and M. Fukushima, *A modified BFGS method and its global convergence in non-convex minimization*, J. Comput. Appl. Math. **129** (2001) 15–35.

[15] G.Y. Li, C.M. Tang and Z.X. Wei, *New conjugacy condition and related new conjugate gradient methods for unconstrained optimization*, J. Comput. Appl. Math. **202** (2007) 523-539.

[16] W.F. Mascarenhas, *The BFGS method with exact line searches fails for non-convex objective functions*, Math. Program. **99** (2004) 49-61.

[17] Y. Narushima and H. Yabe, *Conjugate gradient methods on secant conditions that generate descent search directions for unconstrained optimization*, J. Comput. Appl. Math. **236** (2012) 4303–4317.

[18] Y. Narushima, H. Yabe and J.A. Ford, *A three-term conjugate gradient method with sufficient descent property for unconstrained optimization*, SIAM J. Optim. **21** (2011) 212–230.

[19] J. Nocedal, S.J. Wright, *Numerical Optimization, Springer Series in Operations Research*, Springer, New York, 1999,.

[20] S.S. Oren, D.G. Luenberger, *Self-scaling variable metric (SSVM) algorithms, part I: Criteria and sufficient conditions for scaling a class of algorithms*, Management Science. **20**(5) (1974) 845–862.

[21] J. M. Perry, *A class of conjugate gradient algorithms with a two-step variable-metric memory*, Discussion Paper 269, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University (Evanston, Illinois), 1977.

[22] E. Polak, G. Ribiére, *Note sur la convergence de méthods de directions conjugées*, Rev. Franccaise Informat. Rech. Opiérationnelle **3** (1969) 35–43.

[23] B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comput. Math. Math. Phys. **9** (1969) 94–112 .

[24] D.F. Shanno, *On the convergence of a new conjugate gradient algorithm*, SIAM J. Numer. Anal. **15** (1978) 1248–1257.

[25] H. Yabe and M. Takano, *Global convergence properties of nonlinear conjugate gradient methods with modified secant condition*, Comput. Optim. Appl. **28** (2004) 203–225.

[26] J.Z. Zhang, N.Y. Deng and L.H. Chen, *New quasi-Newton equation and related methods for unconstrained optimization*, J. Optim. Theory Appl. **102** (1999) 147–167.

[27] J.Z. Zhang and C.X. Xu, *Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations*, J. Comput. Appl. Math. **137** (2001) 269–278.

[28] W.J. Zhou and L. Zhang, *A nonlinear conjugate gradient method based on the MBFGS secant equation*, Optim. Meth. Softw. **21** (2006) 707–714.

[29] G. Zoutendijk, *Nonlinear programming. In: Abadie, J. (ed.) Integer and Nonlinear Programming*, North-Holland, Amsterdam, (1970) 37–86.