Conjugate gradient methods based on secant conditions that generate descent search directions for unconstrained optimization

Yasushi Narushima * and Hiroshi Yabe †

September 28, 2011

Abstract

Conjugate gradient methods have been paid attention to, because they can be directly applied to large-scale unconstrained optimization problems. In order to incorporate second order information of the objective function into conjugate gradient methods, Dai and Liao (2001) proposed a conjugate gradient method based on the secant condition. However, their method does not necessarily generate a descent search direction. On the other hand, Hager and Zhang (2005) proposed another conjugate gradient method which always generates a descent search direction.

In this paper, combining Dai-Liao's idea and Hager-Zhang's idea, we propose conjugate gradient methods based on secant conditions that generate descent search directions. In addition, we prove global convergence properties of the proposed methods. Finally, preliminary numerical results are given.

Keyword; Unconstrained optimization, conjugate gradient method, descent search direction, secant condition, global convergence.

KeyMathematics Subject Classification: 90C30, 90C06

1 Introduction

We deal with the following unconstrained optimization problems:

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

where $f : \boldsymbol{R}^n \to \boldsymbol{R}$ is continuously differentiable and its gradient $g \equiv \nabla f$ is available. For solving (1.1), the iterative method is widely used and its form is given by

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

where $x_k \in \boldsymbol{R}^n$ is the kth approximation to a solution of (1.1), $\alpha_k \in \boldsymbol{R}$ is a step size and $d_k \in \boldsymbol{R}^n$ is a search direction.

Recently, the conjugate gradient method is paid attention to as an effective numerical method for solving large-scale unconstrained optimization problems because it does not need the storage of any matrices. The search direction of the conjugate gradient method is defined by

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

*Corresponding author. Department of Communication and Information Science, Fukushima National College of Technology, 30 Nagao, Tairakamiarakawa-aza, Iwaki-shi, Fukushima 970-8034, Japan. e-mail: narushima(at)fukushima-nct.ac.jp

†Department of Mathematical Information Science, Tokyo University of Science, 1-3, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan. e-mail: yabe(at)rs.kagu.tus.ac.jp

where g_k denotes $g(x_k)$ and β_k is a parameter which characterizes the conjugate gradient method. Well known formulas for β_k are the Hestenes-Stiefel (HS) [15], Fletcher-Reeves (FR) [5], Polak-Ribière (PR) [19], Polak-Ribière Plus (PR+) [9], and Dai-Yuan (DY) [3] formulas, which are respectively given by

$$\begin{gathered}\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}, \qquad \beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \\ \beta_k^{PR} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2}, \qquad \beta_k^{PR+} = \max\left\{\frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2},\ 0\right\}, \qquad \beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}},\end{gathered} \tag{1.4}$$

where y_{k-1} is defined by

$$y_{k-1} = g_k - g_{k-1}$$

and $\|\cdot\|$ denotes the ℓ_2 norm. Furthermore, we define

$$s_{k-1} = x_k - x_{k-1},$$

which is used in the subsequent sections. Note that these formulas for β_k are equivalent if the objective function is a strictly convex quadratic function and α_k is the one dimensional minimizer. There are many researches on convergence properties of conjugate gradient methods with (1.4). The global convergence properties of these methods have been proved in the previous works (see [14, 19], for example).

In this decade, in order to incorporate the second-order information of the objective function into conjugate gradient methods, many researchers have proposed conjugate gradient methods based on secant conditions. Dai and Liao [2] proposed a conjugate gradient method based on the secant condition and proved its global convergence property. Later some researchers gave its variants based on other secant conditions, and they proved global convergence properties of their proposed methods [8, 22, 27]. Kobayashi et al. [16] proposed conjugate gradient methods based on structured secant conditions for solving nonlinear least squares problems. Although numerical experiments of the previous works show effectiveness of these methods for solving large-scale unconstrained optimization problems, they do not necessarily satisfy the descent condition ($g_k^T d_k < 0$ for all k), or the sufficient descent condition, namely, there exists a constant $\bar{c} > 0$ such that

$$g_k^T d_k \le -\bar{c}\|g_k\|^2 \qquad \text{for all } k. \tag{1.5}$$

In order to overcome this weakness, Sugiki et al. [21] proposed three-term conjugate gradient methods based on secant conditions which always satisfy the sufficient descent condition (1.5) with $\bar{c} = 1$, by combining the three-term conjugate gradient method by Narushima et al. [18] with parameters β_k given in [2, 8, 16, 22, 27].

On the other hand, Hager and Zhang [11] proposed a formula of β_k:

$$\beta_k^{\mathrm{HZ}} = \frac{1}{d_{k-1}^T y_{k-1}} g_k^T \left(y_{k-1} - 2d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}}\right) = \beta_k^{\mathrm{HS}} - 2\left(\frac{\|y_{k-1}\|}{d_{k-1}^T y_{k-1}}\right)^2 g_k^T d_{k-1}, \tag{1.6}$$

and prove that the conjugate gradient method with (1.6) satisfies the sufficient descent condition (1.5) with $\bar{c} = 7/8$, if $d_{k-1}^T y_{k-1} \neq 0$ holds for all k. Hager and Zhang [14] extended β_k^{HZ} and gave the following formula:

$$\beta_k^{\mathrm{MHZ}} = \frac{1}{d_{k-1}^T y_{k-1}} g_k^T \left(y_{k-1} - \lambda d_{k-1}\frac{\|y_{k-1}\|^2}{d_{k-1}^T y_{k-1}}\right) = \beta_k^{\mathrm{HS}} - \lambda\left(\frac{\|y_{k-1}\|}{d_{k-1}^T y_{k-1}}\right)^2 g_k^T d_{k-1}, \tag{1.7}$$

where $\lambda > 1/4$. Note that their method satisfies the sufficient descent condition with $\bar{c} = 1 - 1/(4\lambda)$. Following Hager-Zhang's idea, Yu, Guan and Li [23] proposed a modified Polak-Ribière method whose β_k is given by

$$\beta_k^{\mathrm{YGL}} = \frac{1}{\|g_{k-1}\|^2} g_k^T \left(y_{k-1} - \lambda d_{k-1} \frac{\|y_{k-1}\|^2}{\|g_{k-1}\|^2} \right) = \beta_k^{\mathrm{PR}} - \lambda \left(\frac{\|y_{k-1}\|}{\|g_{k-1}\|^2} \right)^2 g_k^T d_{k-1}, \tag{1.8}$$

where $\lambda > 1/4$. They showed that a conjugate gradient method with β_k^{YGL} also satisfies the sufficient descent condition with $\bar{c} = 1 - 1/(4\lambda)$. After that, Yuan [24] proposed some variants of the method of Yu et al.

Considering that β_k^{HZ} can be regarded as a modification of β_k^{HS}, we propose, in this paper, new conjugate gradient methods which are based on β_k in [2, 8, 22, 27] and satisfy the sufficient descent condition. The present paper is organized as follows. In Section 2, we propose the parameter β_k by making use of the technique of Hager and Zhang [14], and give its related algorithm. In Section 3, we show global convergence of our method given in Section 2. Finally, in Section 4, some numerical experiments are presented.

2 Conjugate gradient methods based on the secant conditions that generate descent search directions

In this section, we propose conjugate gradient methods based on the secant conditions that generate descent search directions. In Section 2.1, we review conjugate gradient methods based on secant conditions. In Section 2.2, making use of Hager and Zhang's idea, we give new formulas of β_k.

2.1 Conjugate gradient methods based on the secant conditions

The conjugacy condition of (nonlinear) conjugate gradient methods is given by

$$d_k^T y_{k-1} = 0. \tag{2.1}$$

In order to incorporate the second-order information into the conjugacy condition (2.1), Perry [20] extended the conjugacy condition (2.1) by using the secant condition of quasi-Newton methods:

$$B_k s_{k-1} = y_{k-1}, \tag{2.2}$$

and the search direction d_k of quasi-Newton methods:

$$B_k d_k = -g_k, \tag{2.3}$$

where B_k is a symmetric approximation matrix to the Hessian $\nabla^2 f(x_k)$. Specifically, based on the relations (2.2) and (2.3), Perry gave the following relation

$$d_k^T y_{k-1} = d_k^T (B_k s_{k-1}) = (B_k d_k)^T s_{k-1} = -g_k^T s_{k-1}.$$

Thus, Perry's conjugacy condition is defined by

$$d_k^T y_{k-1} = -g_k^T s_{k-1}. \tag{2.4}$$

After that, by incorporating nonnegative parameter t, Dai and Liao [2] proposed the following condition:

$$d_k^T y_{k-1} = -t g_k^T s_{k-1}. \tag{2.5}$$

Note that, if $t = 0$, then (2.5) reduces to the usual conjugacy condition (2.1), and if $t = 1$, (2.5) becomes Perry's condition (2.4). Moreover, if we use the exact line search, the condition (2.5) is equivalent to the conjugacy condition (2.1), independently of choices of t. By substituting (1.3) into condition (2.5), Dai and Liao proposed a parameter β_k as follows:

$$\beta_k^{DL} = \frac{g_k^T (y_{k-1} - t s_{k-1})}{d_{k-1}^T y_{k-1}}. \tag{2.6}$$

Note that $d_{k-1}^T y_{k-1} > 0$ holds for all k if the Wolfe conditions are used in the line search. They showed that the conjugate gradient method with β_k^{DL} converges globally for a uniformly convex objective function under the assumption that the method satisfies the descent condition. They also showed that the conjugate gradient method with $\beta_k^{DL+} = \max\{0, \beta_k^{DL}\}$ converges globally for a general objective function under the assumption that the method satisfies the sufficient descent condition (1.5).

Recently, following Dai and Liao, several conjugate gradient methods have been studied by using other secant conditions instead of the secant condition (2.2). We first introduce some secant conditions, and next review conjugate gradient methods based on these secant conditions.

Zhang, Deng and Chen [25] and Zhang and Xu [26] presented the following modified secant condition:

$$B_k s_{k-1} = z_{k-1}^{YT}, \quad z_{k-1}^{YT} = y_{k-1} + \phi_k \left(\frac{\theta_{k-1}}{s_{k-1}^T u_{k-1}} u_{k-1} \right), \tag{2.7}$$

where

$$\theta_{k-1} \quad = \quad 6(f_{k-1} - f_k) + 3(g_{k-1} + g_k)^T s_{k-1}, \tag{2.8}$$

and $\phi_k \geq 0$ is a scalar, f_k denotes $f(x_k)$ and $u_{k-1} \in \boldsymbol{R}^n$ is any vector such that $s_{k-1}^T u_{k-1} \neq 0$ holds. Li and Fukushima [17] gave the MBFGS secant condition:

$$B_k s_{k-1} = z_{k-1}^{ZZ}, \quad z_{k-1}^{ZZ} = y_{k-1} + \zeta \|g_k\|^q s_{k-1}, \tag{2.9}$$

where $\zeta > 0$ and $q > 0$ are constants. Ford and Moghrabi [6, 7] proposed the multi-step secant condition. Later on, Ford, Narushima and Yabe [8] introduced the following specific choices of the multi-step secant conditions:

$$B_k h_{k-1}^{F1} = z_{k-1}^{F1}, \quad h_{k-1}^{F1} = s_{k-1} - \xi_{k-1} s_{k-2}, \quad z_{k-1}^{F1} = y_{k-1} - \xi_{k-1} y_{k-2}, \tag{2.10}$$

and

$$B_k h_{k-1}^{F2} = z_{k-1}^{F2}, \quad h_{k-1}^{F2} = s_{k-1} - \xi_{k-1} s_{k-2}, \quad z_{k-1}^{F2} = y_{k-1} - t\xi_{k-1} y_{k-2}, \tag{2.11}$$

where

$$\xi_{k-1} = \frac{\delta_{k-1}^2}{1 + 2\delta_{k-1}}, \quad \delta_{k-1} = \eta_k \frac{\|s_{k-1}\|}{\|s_{k-2}\|}, \tag{2.12}$$

and $\eta_k \geq 0$ is a scaling factor.

Based on the modified secant condition (2.7), Yabe and Takano [22] proposed the following formula for β_k:

$$\beta_k^{YT} = \frac{g_k^T (z_{k-1}^{YT} - t s_{k-1})}{d_{k-1}^T z_{k-1}^{YT}}. \tag{2.13}$$

On the other hand, based on (2.9), Zhou and Zhang [27] proposed

$$\beta_k^{ZZ} = \frac{g_k^T(z_{k-1}^{ZZ} - ts_{k-1})}{d_{k-1}^T z_{k-1}^{ZZ}}. \tag{2.14}$$

In addition, based on (2.10) and (2.11), Ford et al. [8] proposed two types of formulas for β_k given by

$$\beta_k^{F1} = \frac{g_k^T(z_{k-1}^{F1} - th_{k-1}^{F1})}{d_{k-1}^T z_{k-1}^{F1}}, \tag{2.15}$$

$$\beta_k^{F2} = \frac{g_k^T(z_{k-1}^{F2} - th_{k-1}^{F2})}{d_{k-1}^T z_{k-1}^{F2}}. \tag{2.16}$$

We now treat a unified formula of β_k in (2.6) and (2.13)–(2.16). Secant conditions are generally represented by

$$B_k h_{k-1} = z_{k-1}. \tag{2.17}$$

In the case of $h_{k-1} = s_{k-1}$ and $z_{k-1} = y_{k-1}$, (2.17) reduces to the usual secant condition (2.2). Following Dai and Liao, we have a general form of conjugacy condition $d_k^T z_{k-1} = -tg_k^T h_{k-1}$. By substituting (1.3) into the above condition,

$$\beta_k d_{k-1}^T z_{k-1} = g_k^T z_{k-1} - tg_k^T h_{k-1} \tag{2.18}$$

is obtained. If $d_{k-1}^T z_{k-1} = 0$, then β_k satisfying (2.18) does not necessarily exist. Thus we set $\beta_k = 0$ when $d_{k-1}^T z_{k-1} = 0$. Taking into account the above arguments, we have the formula for β_k as follows:

$$\beta_k^{Secant} = g_k^T(z_{k-1} - th_{k-1})(d_{k-1}^T z_{k-1})^{\dagger}, \tag{2.19}$$

where † implies the following generalized inverse:

$$a^{\dagger} = \begin{cases} \dfrac{1}{a}, & a \neq 0, \\ 0, & a = 0. \end{cases}$$

In Table 1, we give z_{k-1} and h_{k-1} in (2.19) for the cases β_k^{DL}, β_k^{YT}, β_k^{ZZ}, β_k^{F1} and β_k^{F2}.

Table 1: z_{k-1} and h_{k-1} in (2.19)

Name	z_{k-1}	h_{k-1}
β_k^{DL}	y_{k-1}	s_{k-1}
β_k^{YT}	z_{k-1}^{YT} in (2.7)	s_{k-1}
β_k^{ZZ}	z_{k-1}^{ZZ} in (2.9)	s_{k-1}
β_k^{F1}	z_{k-1}^{F1} in (2.10)	h_{k-1}^{F1} in (2.10)
β_k^{F2}	z_{k-1}^{F2} in (2.11)	h_{k-1}^{F2} in (2.11)

The conjugate gradient method with (2.19) does not necessarily generate descent search directions. If we try to establish the global convergence of conjugate gradient method with (2.19), we need to assume that the search direction satisfies the (sufficient) descent condition.

In order to overcome this weakness, Sugiki et al. applied (2.19) to the three-term conjugate gradient method by Narushima et al. [18] and proposed three-term conjugate gradient methods satisfying the sufficient descent condition (1.5). On the other hand we propose, in this paper, a conjugate gradient method which satisfies the sufficient descent condition by modifying the unified formula (2.19).

2.2 Proposed method

In this section, we give conjugate gradient methods that are based on secant conditions and satisfy the sufficient descent condition. Taking into account β_k^{MHZ} in (1.7) or β_k^{YGL} in (1.8), we propose the following formula of β_k:

$$\begin{aligned}\beta_k^{DS} &= \beta_k^{Secant} - \lambda\|z_{k-1} - th_{k-1}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T z_{k-1})^2\}^\dagger \\ &= g_k^T(z_{k-1} - th_{k-1})(d_{k-1}^T z_{k-1})^\dagger - \lambda\|z_{k-1} - th_{k-1}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T z_{k-1})^2\}^\dagger, \quad (2.20)\end{aligned}$$

where λ is a parameter such that $\lambda > 1/4$, and "DS" denotes "Descent and Secant conditions". If $d_{k-1}^T z_{k-1} = 0$, then $\beta_k^{DS} = 0$ and $g_k^T d_k = -\|g_k\|^2$, otherwise, considering the fact that $u^T v \leq \frac{1}{2}(\|u\|^2 + \|v\|^2)$ holds for any vector u and v, we have

$$\begin{aligned}g_k^T d_k &= -\|g_k\|^2 + \beta_k^{DS} g_k^T d_{k-1} \\ &= -\|g_k\|^2 + \frac{1}{(d_{k-1}^T z_{k-1})^2} \\ &\quad \times\{d_{k-1}^T z_{k-1} g_k^T(z_{k-1} - th_{k-1}) g_k^T d_{k-1} - \lambda(g_k^T d_{k-1})^2\|z_{k-1} - th_{k-1}\|^2\} \\ &= -\|g_k\|^2 + \frac{1}{(d_{k-1}^T z_{k-1})^2} \\ &\quad \times\left[\frac{(d_{k-1}^T z_{k-1} g_k)^T}{\sqrt{2\lambda}}\{\sqrt{2\lambda} g_k^T d_{k-1}(z_{k-1} - th_{k-1})\} - \lambda(g_k^T d_{k-1})^2\|z_{k-1} - th_{k-1}\|^2\right] \\ &\leq -\|g_k\|^2 + \frac{1}{(d_{k-1}^T z_{k-1})^2} \\ &\quad \times\left\{\frac{\|d_{k-1}^T z_{k-1} g_k\|^2}{4\lambda} + \lambda\|g_k^T d_{k-1}(z_{k-1} - th_{k-1})\|^2 - \lambda(g_k^T d_{k-1})^2\|z_{k-1} - th_{k-1}\|^2\right\} \\ &= -\left(1 - \frac{1}{4\lambda}\right)\|g_k\|^2.\end{aligned}$$

Summarizing the above arguments, the following lemma is obtained.

Lemma 2.1. *Consider the conjugate gradient method* (1.2)–(1.3) *with* (2.20)*. Then the sufficient descent condition* (1.5) *holds with* $\bar{c} = 1 - \frac{1}{4\lambda}$.

To establish the global convergence of the methods for a general objective function, $\beta_k \geq 0$ is often needed. Then we replace β_k^{DS} by

$$\beta_k^{DS+} = \max\{0, \beta_k^{DS}\}. \quad (2.21)$$

Note that Lemma 2.1 still holds for the conjugate gradient method with (2.21). We now give an algorithm of conjugate gradient method with (2.20) or (2.21).

Algorithm 2.1.

Step 0 *Give an initial point* $x_0 \in \boldsymbol{R}^n$ *and positive parameters* $\lambda > 1/4$, $0 < \sigma_1 < \sigma_2 < 1$. *Set the initial search direction* $d_0 = -g_0$. *Let* $k = 0$ *and go to Step 2.*

Step 1 *Compute* d_k *by* (1.3) *with* (2.20) (*or* (2.21)).

Step 2 *Determine a step size* α_k *satisfying the Wolfe conditions*:

$$f(x_k+\alpha_k d_k)-f(x_k) \leq \sigma_1\alpha_k g_k^T d_k, \tag{2.22}$$
$$g(x_k+\alpha_k d_k)^T d_k \geq \sigma_2 g_k^T d_k. \tag{2.23}$$

Step 3 *Update* x_{k+1} *by* (1.2).

Step 4 *If the stopping criterion is satisfied, then stop. Otherwise go to Step 5.*

Step 5 *Let* $k := k+1$ *and go to Step 1.*

Note that the Wolfe conditions and (1.5) yield

$$d_{k-1}^T y_{k-1} \geq (\sigma_2-1)g_{k-1}^T d_{k-1} \geq \bar{c}(1-\sigma_2)\|g_{k-1}\|^2 (>0). \tag{2.24}$$

Now we introduce the concrete choices of β_k^{DS} and β_k^{DS+} by using the same arguments in Section 2.1. Considering Table 1 and (2.20), concrete choices of β_k^{DS} and β_k^{DS+} are respectively given by the following:

$$\begin{aligned}
\beta_k^{DSDL} &= g_k^T(y_{k-1}-ts_{k-1})(d_{k-1}^T y_{k-1})^\dagger - \lambda\|y_{k-1}-ts_{k-1}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T y_{k-1})^2\}^\dagger,\\
\beta_k^{DSYT} &= g_k^T(z_{k-1}^{YT}-ts_{k-1})(d_{k-1}^T z_{k-1}^{YT})^\dagger - \lambda\|z_{k-1}^{YT}-ts_{k-1}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T z_{k-1}^{YT})^2\}^\dagger,\\
\beta_k^{DSZZ} &= g_k^T(z_{k-1}^{ZZ}-ts_{k-1})(d_{k-1}^T z_{k-1}^{ZZ})^\dagger - \lambda\|z_{k-1}^{ZZ}-ts_{k-1}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T z_{k-1}^{ZZ})^2\}^\dagger,\\
\beta_k^{DSF1} &= g_k^T(z_{k-1}^{F1}-th_{k-1}^{F1})(d_{k-1}^T z_{k-1}^{F1})^\dagger - \lambda\|z_{k-1}^{F1}-th_{k-1}^{F1}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T z_{k-1}^{F1})^2\}^\dagger,\\
\beta_k^{DSF2} &= g_k^T(z_{k-1}^{F2}-th_{k-1}^{F2})(d_{k-1}^T z_{k-1}^{F2})^\dagger - \lambda\|z_{k-1}^{F2}-th_{k-1}^{F2}\|^2 g_k^T d_{k-1}\{(d_{k-1}^T z_{k-1}^{F2})^2\}^\dagger,
\end{aligned}$$

and

$$\beta_k^{DSDL+} = \max\{0, \beta_k^{DSDL}\}, \tag{2.25}$$
$$\beta_k^{DSYT+} = \max\{0, \tilde{\beta}_k^{DSYT}\}, \tag{2.26}$$
$$\beta_k^{DSF1+} = \max\{0, \beta_k^{DSF1}\}, \tag{2.27}$$
$$\beta_k^{DSF2+} = \max\{0, \beta_k^{DSF2}\}, \tag{2.28}$$

where $\tilde{\beta}_k^{DSYT}$ is β_k^{DS} with $h_{k-1}=s_{k-1}$ and

$$z_{k-1} = z_{k-1}^{YT+} = y_{k-1} + \phi_k\left(\frac{\max\{0,\theta_{k-1}\}}{s_{k-1}^T u_{k-1}}u_{k-1}\right). \tag{2.29}$$

3 Global convergence of the proposed methods

In this section, we investigate the global convergence property of Algorithm 2.1. For this purpose, we make the following assumptions for the objective function.

Assumption 3.1.

A1. *The level set* $\mathcal{L}=\{x|f(x)\leq f(x_0)\}$ *at* x_0 *is bounded, namely, there exists a constant* $\widehat{a}>0$ *such that*

$$\|x\| \leq \widehat{a} \qquad \textit{for all} \quad x\in\mathcal{L}.$$

A2 *In some open convex neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is continuously differentiable, and its gradient g is Lipschitz continuous, namely, there exists a positive constant L such that*

$$\|g(x) - g(\bar{x})\| \le L\|x - \bar{x}\| \qquad \textit{for all} \quad x, \bar{x} \in \mathcal{N}.$$

Note that Assumption 3.1 means that there exists a positive constant γ such that

$$\|g(x)\| \le \gamma \qquad \text{for all} \quad x \in \mathcal{L}. \tag{3.1}$$

We also assume $g_k \neq 0$ for all k, otherwise a stationary point has been found.

To establish the global convergence of the methods, we give a lemma for general iterative methods. The lemma can be easily shown by using the Zoutendijk condition [28], and hence we omit the proof (see [18], for example).

Lemma 3.1. *Suppose that Assumption* 3.1 *holds. Consider any iterative method of the form* (1.2)*, where d_k and α_k satisfy the sufficient descent condition* (1.5) *and the Wolfe conditions* (2.22) *and* (2.23)*, respectively. If*

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

then $\liminf_{k\to\infty} \|g_k\| = 0$ *holds.*

By using Lemma 3.1, we have the following theorem.

Theorem 3.1. *Suppose that Assumption* 3.1 *holds. Let the sequence $\{x_k\}$ be generated by Algorithm* 2.1*. If there exist positive constants c_1 and c_2 such that z_{k-1} and h_{k-1} satisfy*

$$\begin{aligned} \|z_{k-1} - th_{k-1}\| &\le c_1\|s_{k-1}\|, & (3.2)\\ \alpha_{k-1}\|d_{k-1}\|^2\, |d_{k-1}^T z_{k-1}|^{\dagger} &\le c_2 & (3.3) \end{aligned}$$

for all k, then the method converges globally in the sense that $\liminf\limits_{k\to\infty} \|g_k\| = 0$.

Proof. By (2.22) and Lemma 2.1, the sequence $\{x_k\}$ is contained in the level set $\mathcal{L}$. It follows from (2.20), (3.1), (3.2) and (3.3) that

$$\begin{aligned} |\beta_k^{DS}|\ \|d_{k-1}\| &\le \left| g_k^T(z_{k-1} - th_{k-1})(d_{k-1}^T z_{k-1})^{\dagger} \right| \|d_{k-1}\| \\ &\qquad + \lambda \left| g_k^T d_{k-1} \right| \ \|z_{k-1} - th_{k-1}\|^2 \{(d_{k-1}^T z_{k-1})^2\}^{\dagger}\, \|d_{k-1}\| \\ &\le c_1\alpha_{k-1}\|g_k\|\|d_{k-1}\|^2|d_{k-1}^T z_{k-1}|^{\dagger} + \lambda c_1^2\alpha_{k-1}^2\|g_k\|\|d_{k-1}\|^4(|d_{k-1}^T z_{k-1}|^{\dagger})^2 \\ &\le (c_1c_2 + \lambda c_1^2 c_2^2)\gamma, \end{aligned}$$

and hence we have from (1.3) that

$$\|d_k\| \ \le \ \|g_k\| + |\beta_k^{DS}|\ \|d_{k-1}\| \le (1 + c_1c_2 + \lambda c_1^2 c_2^2)\gamma,$$

which implies that $\sum_{k=0}^{\infty} 1/\|d_k\|^2 = \infty$ holds. Therefore from Lemma 3.1, the proof is complete. □

By using Theorem 3.1, we can show global convergence properties of Algorithm 2.1 with β_k^{DSDL}, β_k^{DSYT}, β_k^{DSF1}, and β_k^{DSF2} for a uniformly convex objective function, and Algorithm 2.1 with β_k^{DSZZ} for a general objective function.

First we give the definition of uniformly convex function. The function f is said to be uniformly convex (on $\boldsymbol{R}^n$) with modulus μ if $f((1-\lambda)x + \lambda y) \le (1-\lambda)f(x) + \lambda f(y) - \frac{1}{2}\mu(1 -$

$\lambda)\lambda\|x-y\|^2$ holds for any $x,\ y \in \boldsymbol{R}^n$ and $\lambda \in (0,\ 1)$. Note that, if f is a continuously differentiable uniformly convex function, the following holds:

$$(g(x) - g(\bar{x}))^T(x - \bar{x}) \geq \mu\|x - \bar{x}\|^2 \qquad \text{for all } x, \bar{x} \in \boldsymbol{R}^n. \tag{3.4}$$

We also note that, if f is a uniformly convex function, then the level set $\mathcal{L}$ is bounded for any x_0, and hence Assumption A1 is satisfied.

The proof of the following theorem is similar to that of [21, Theorem 2], but we do not omit the proof for readability.

Theorem 3.2. *Suppose that Assumption* 3.1 *holds and that f is a uniformly convex function. Let x^* be a unique optimal solution to* (1.1)*.*

(i) *Algorithm* 2.1 *with β_k^{DSDL} converges globally, i.e.* $\lim_{k\to\infty} x_k = x^*$.

(ii) *Assume that ϕ_k and u_k satisfy $0 \leq \phi_k \leq \bar{\phi}$ and*

$$|s_{k-1}^T u_{k-1}| \geq \bar{m}\|s_{k-1}\|\|u_{k-1}\|, \tag{3.5}$$

where $\bar{\phi}$ is a positive constant such that $\bar{\phi} < \mu/(3L)$, and $\bar{m}$ is some positive constant. Then Algorithm 2.1 *with β_k^{DSYT} converges globally, i.e.* $\lim_{k\to\infty} x_k = x^*$.

(iii) *If η_k satisfies $0 \leq \eta_k \leq \bar{\eta}$ for some positive constant $\bar{\eta}$ such that $2\mu - \bar{\eta}L > 0$ holds, then Algorithm* 2.1 *with β_k^{DSF1} converges globally, i.e.* $\lim_{k\to\infty} x_k = x^*$.

(iv) *If η_k satisfies $0 \leq \eta_k \leq \bar{\eta}$ for some positive constant $\bar{\eta}$ such that $2\mu - t\bar{\eta}L > 0$ holds, then Algorithm* 2.1 *with β_k^{DSF2} converges globally, i.e.* $\lim_{k\to\infty} x_k = x^*$.

Proof. **(i)** By Table 1, we have

$$\|z_{k-1} - th_{k-1}\| = \|y_{k-1} - ts_{k-1}\| \leq (L + t)\|s_{k-1}\|, \tag{3.6}$$

which implies (3.2) with $c_1 = L + t$. Since from (3.4), $y_{k-1}^T s_{k-1} \geq \mu\|s_{k-1}\|^2$ holds, we have

$$|d_{k-1}^T z_{k-1}| = |d_{k-1}^T y_{k-1}| \geq \mu\alpha_{k-1}\|d_{k-1}\|^2.$$

Thus (3.3) is satisfied with $c_2 = 1/\mu$. It follows from Theorem 3.1 that $\liminf_{k\to\infty} \|g_k\| = 0$ holds.

(ii) By the mean value theorem, the following holds:

$$f_{k-1} - f_k = -g(\tau' x_{k-1} + (1 - \tau')x_k)^T s_{k-1}$$

for some $\tau' \in (0, 1)$. Then it follows from (2.8) and Assumption A2 that

$$\begin{aligned}
|\theta_{k-1}| &= |6(f_{k-1} - f_k) + 3(g_{k-1} + g_k)^T s_{k-1}| \\
&= |-6g(\tau' x_{k-1} + (1 - \tau')x_k)^T s_{k-1} + 3(g_{k-1} + g_k)^T s_{k-1}| \\
&\leq 3\{\|g_{k-1} - g(\tau' x_{k-1} + (1 - \tau')x_k)\| + \|g_k - g(\tau' x_{k-1} + (1 - \tau')x_k)\|\}\|s_{k-1}\| \\
&\leq 3L\{\|x_{k-1} - (\tau' x_{k-1} + (1 - \tau')x_k)\| + \|x_k - (\tau' x_{k-1} + (1 - \tau')x_k)\|\}\|s_{k-1}\| \\
&= 3L\|s_{k-1}\|^2.
\end{aligned} \tag{3.7}$$

We have from Table 1, (3.5), (3.7) and Assumption A2 that

$$\begin{aligned}
\|z_{k-1} - th_{k-1}\| &= \left\| y_{k-1} + \phi_k \left(\frac{\theta_{k-1}}{s_{k-1}^T u_{k-1}} u_{k-1} \right) - ts_{k-1} \right\| \\
&\leq L\|s_{k-1}\| + \phi_k \frac{3L\|s_{k-1}\|^2}{\bar{m}\|s_{k-1}\|\|u_{k-1}\|}\|u_{k-1}\| + t\|s_{k-1}\| \\
&\leq (L + \bar{\phi}\frac{3L}{\bar{m}} + t)\|s_{k-1}\|.
\end{aligned} \tag{3.8}$$

The relations (3.7) and (3.4) yield

$$
\begin{aligned}
|d_{k-1}^T z_{k-1}| &= \left| d_{k-1}^T y_{k-1} + \phi_k \frac{\theta_{k-1}}{s_{k-1}^T u_{k-1}} d_{k-1}^T u_{k-1} \right| \\
&\geq |d_{k-1}^T y_{k-1}| - \frac{\phi_k}{\alpha_{k-1}} 3L \|s_{k-1}\|^2 \\
&\geq (\mu - 3\bar{\phi}L)\alpha_{k-1}\|d_{k-1}\|^2. \qquad (3.9)
\end{aligned}
$$

We note that $\mu - 3\bar{\phi}L > 0$. The relations (3.8) and (3.9) imply (3.2) and (3.3) with $c_1 = (L+3\bar{\phi}L/\bar{m}+t)$ and $c_2 = 1/(\mu-3\bar{\phi}L)$, and hence we get, from Theorem 3.1, that $\liminf_{k\to\infty} \|g_k\| = 0$.

(iii) The relation (2.12) yields

$$
\xi_{k-1} \leq \frac{\delta_{k-1}}{2} = \frac{\eta_k}{2} \frac{\|s_{k-1}\|}{\|s_{k-2}\|}, \qquad (3.10)
$$

and hence we have from Table 1 and (3.10) that

$$
\begin{aligned}
\|z_{k-1} - th_{k-1}\| &= \|y_{k-1} - \xi_{k-1} y_{k-2} - t(s_{k-1} - \xi_{k-1}s_{k-2})\| \\
&\leq \|y_{k-1}\| + \xi_{k-1}\|y_{k-2}\| + t\|s_{k-1}\| + t\xi_{k-1}\|s_{k-2}\| \\
&\leq (L+t)\|s_{k-1}\| + \frac{\eta_k}{2}\frac{\|s_{k-1}\|}{\|s_{k-2}\|}(L+t)\|s_{k-2}\| \\
&\leq \left(1 + \frac{\bar{\eta}}{2}\right)(L+t)\|s_{k-1}\|, \qquad (3.11)
\end{aligned}
$$

which implies (3.2) with $c_1 = (1+\bar{\eta}/2)(L+t)$. We have from (3.4), (3.10), and Assumption A2

$$
\begin{aligned}
|d_{k-1}^T z_{k-1}| &= |d_{k-1}^T y_{k-1} - \xi_{k-1} d_{k-1}^T y_{k-2}| \\
&\geq |d_{k-1}^T y_{k-1}| - \xi_{k-1}|d_{k-1}^T y_{k-2}| \\
&\geq \mu\alpha_{k-1}\|d_{k-1}\|^2 - \frac{\eta_k}{2}\frac{\|s_{k-1}\|}{\|s_{k-2}\|}\|d_{k-1}\|\|y_{k-2}\| \\
&\geq \mu\alpha_{k-1}\|d_{k-1}\|^2 - \frac{\bar{\eta}}{2}L\alpha_{k-1}\|d_{k-1}\|^2 \\
&= \left(\mu - \frac{\bar{\eta}}{2}L\right)\alpha_{k-1}\|d_{k-1}\|^2.
\end{aligned}
$$

We note that $\mu-\bar{\eta}L/2 > 0$, and hence (3.3) holds with $c_2 = 1/(\mu-\bar{\eta}L/2)$. Thus by Theorem 3.1, we get $\liminf_{k\to\infty} \|g_k\| = 0$.

(iv) By Table 1 and (3.10), we have

$$
\begin{aligned}
\|z_{k-1} - th_{k-1}\| &= \|y_{k-1} - t\xi_{k-1} y_{k-2} - t(s_{k-1} - \xi_{k-1}s_{k-2})\| \\
&\leq (L+t)\|s_{k-1}\| + t\frac{\eta_k}{2}(1+L)\|s_{k-1}\| \\
&\leq \left\{L + t + \frac{\bar{\eta}}{2}t\,(1+L)\right\}\|s_{k-1}\|. \qquad (3.12)
\end{aligned}
$$

It follows from (3.4) and (3.10) that

$$
\begin{aligned}
|d_{k-1}^T z_{k-1}| &= |d_{k-1}^T y_{k-1} - t\xi_{k-1} d_{k-1}^T y_{k-2}| \\
&\geq \mu\alpha_{k-1}\|d_{k-1}\|^2 - tL\frac{\eta_k}{2}\|s_{k-1}\|\|d_{k-1}\| \\
&\geq \left(\mu - \frac{\bar{\eta}}{2}Lt\right)\alpha_{k-1}\|d_{k-1}\|^2. \qquad (3.13)
\end{aligned}
$$

We note $\mu - \bar{\eta}Lt/2 > 0$. Therefore, we have from (3.12) and (3.13) that (3.2) and (3.3) hold with $c_1 = L + t + \bar{\eta}t(1 + L)/2$ and $c_2 = 1/(\mu - \bar{\eta}Lt/2)$. It follows from Theorem 3.1 that we have $\liminf_{k\to\infty} \|g_k\| = 0$.

Summarizing (i)–(iv), we have $\liminf_{k\to\infty} \|g_k\| = 0$ for each case. Since f is uniformly convex, we obtain the desired result. Therefore the theorem is proved. □

Although condition (3.5) for u_{k-1} is assumed in (ii) of Theorem 3.2, this condition is reasonable. For example, if $u_{k-1} = s_{k-1}$, (3.5) is satisfied with $\bar{m} = 1$. If $u_{k-1} = y_{k-1}$, it follows from (3.4) that

$$|s_{k-1}^T u_{k-1}| = |s_{k-1}^T y_{k-1}| \geq \mu \|s_{k-1}\|^2 \geq \frac{\mu}{L} \|s_{k-1}\| \, \|y_{k-1}\|,$$

which implies that (3.5) is satisfied with $\bar{m} = \mu/L$.

Next we show global convergence of Algorithm 2.1 with β_k^{DSZZ} for a general function.

Theorem 3.3. *Suppose that Assumptions* 3.1 *holds. Consider Algorithm* 2.1 *with* β_k^{DSZZ}. *Then the method converges globally in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

Proof. To prove this theorem by contradiction, we suppose that there exists a positive constant ε such that

$$\|g_k\| > \varepsilon \qquad \text{for all } k. \tag{3.14}$$

It follows from Table 1 and (3.1) that

$$\|z_{k-1} - th_{k-1}\| = \|y_{k-1} + \zeta\|g_k\|^q s_{k-1} - ts_{k-1}\| \leq (L + \zeta\gamma^q + t)\|s_{k-1}\|. \tag{3.15}$$

We have from (2.24) that $d_{k-1}^T y_{k-1} > 0$, and hence it follows from (3.14) that

$$|d_{k-1}^T z_{k-1}| = d_{k-1}^T y_{k-1} + \zeta\|g_{k-1}\|^q d_{k-1}^T s_{k-1} > \zeta\varepsilon^q \alpha_{k-1}\|d_{k-1}\|^2. \tag{3.16}$$

Therefore, it follows from (3.15) and (3.16) that (3.2) and (3.3) hold with $c_1 = L + \zeta\gamma^q + t$ and $c_2 = 1/(\zeta\varepsilon^q)$. Although, by Theorem 3.1, we have $\liminf_{k\to\infty} \|g_k\| = 0$, this contradicts (3.14). Therefore the proof is complete. □

Although we proved, in Theorem 3.2, global convergence properties of the method with β_k^{DSDL}, β_k^{DSYT}, β_k^{DSF1} and β_k^{DSF2} for uniformly convex objective functions, we have not shown their global convergence properties for general objective functions. Accordingly, in the rest of this section, we consider the global convergence properties of the methods for such functions.

The following property is originally given by [9], and this property shows that β_k will be small when the step s_{k-1} is small.

Property A. *Consider the conjugate gradient method* (1.2)–(1.3) *and suppose that there exist positive constants* ε *and* γ *such that* $\varepsilon \leq \|g_k\| \leq \gamma$ *for all* k. *If there exist* $b > 1$ *and* $\nu > 0$ *such that* $|\beta_k| \leq b$ *and*

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

then we say that the method has Property A.

The general result under Property A is the following (for example, see [14]).

Theorem 3.4. *Suppose that Assumption* 3.1 *holds. Let* $\{x_k\}$ *be the sequence generated by the conjugate gradient method* (1.2)–(1.3) *which satisfies the following conditions*:

(C1) $\beta_k \geq 0$ *for all* k,

(C2) *the sufficient descent condition,*

(C3) *the Zoutendijk condition,*

(C4) *Property A.*

Then the sequence $\{x_k\}$ *converges globally in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

In order to adopt Theorem 3.4 to Algorithm 2.1, β_k must be nonnegative. As mentioned around (2.21), we consider Algorithm 2.1 with β_k^{DS+} instead of β_k^{DS}. We note that, if $\beta_k^{DS} < 0$ and $\beta_k^{DS+} = 0$, the search direction becomes the steepest descent direction (i.e. $d_k = -g_k$), and hence Algorithm 2.1 with β_k^{DS+} still satisfies the sufficient descent condition (1.5).

Now we give the following global convergence property for general objective functions.

Theorem 3.5. *Suppose that Assumption* 3.1 *holds.*

(i) *Algorithm* 2.1 *with* β_k^{DSDL+} *in* (2.25) *converges globally in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

(ii) *Assume that* u_k *and* ϕ_k *satisfy* (3.5) *and* $0 \leq \phi_k \leq \bar{\phi}$ *where* $\bar{\phi}$ *is any fixed positive constant. Then Algorithm* 2.1 *with* β_k^{DSYT+} *in* (2.26) *converges globally in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

(iii) *Assume that there exists a positive constant* φ_1 *such that, for all* k,

$$\max\{|g_{k-1}^T d_{k-1}|,\ |g_k^T d_{k-1}|\}\, |d_{k-1}^T z_{k-1}^{F1}|^{\dagger} \leq \varphi_1 \tag{3.17}$$

holds. If η_k *satisfies* $0 \leq \eta_k \leq \bar{\eta}$ *for any fixed positive constant* $\bar{\eta}$, *then Algorithm* 2.1 *with* β_k^{DSF1+} *in* (2.27) *converges globally in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

(iv) *Assume that there exists a positive constant* φ_2 *such that, for all* k,

$$\max\{|g_{k-1}^T d_{k-1}|,\ |g_k^T d_{k-1}|\}\, |d_{k-1}^T z_{k-1}^{F2}|^{\dagger} \leq \varphi_2 \tag{3.18}$$

holds. If η_k *satisfies* $0 \leq \eta_k \leq \bar{\eta}$ *for any fixed positive constant* $\bar{\eta}$, *then Algorithm* 2.1 *with* β_k^{DSF2+} *in* (2.28) *converges globally in the sense that* $\liminf_{k\to\infty} \|g_k\| = 0$.

Proof. By (2.21), $\beta_k^{DS+} \geq 0$ holds, and hence condition (C1) is satisfied. As mentioned above, the method with β_k^{DS+} satisfies the sufficient descent condition, which implies condition (C2). Assumption A2 and the Wolfe conditions yield the Zoutendijk condition, and hence condition (C3) is also satisfied. Therefore, we need to prove Property A only, and hence we assume, in the rest of the proof, that there exists a positive constant ε such that

$$\varepsilon \leq \|g_k\| \tag{3.19}$$

holds for any k. Note that there exists a positive constant $\widehat{a}$ such that

$$\|s_{k-1}\| < \widehat{a} \tag{3.20}$$

because the level set $\mathcal{L}$ is bounded and $\{x_k\} \subset \mathcal{L}$. Under the condition (3.19), if there exists a positive constant c_3 satisfying

$$|\beta_k^{DS+}| \leq c_3 \|s_{k-1}\| \tag{3.21}$$

for all k, then we have, by putting $\nu = 1/(2bc_3)$, that $|\beta_k^{DS+}| \le \max\{1, \widehat{a}c_3\} \equiv b$ and

$$\|s_{k-1}\| \le \nu \quad \Longrightarrow \quad |\beta_k^{DS+}| \le \frac{1}{2b},$$

which implies that Property A is satisfied. Thus it suffices to prove that (3.21) holds for cases (i)–(iv).

Here, we give some facts. It follows from (2.24) and (3.19) that

$$d_{k-1}^T y_{k-1} \ge \bar{c}(1-\sigma_2)\|g_{k-1}\|^2 \ge \bar{c}(1-\sigma_2)\varepsilon^2. \tag{3.22}$$

The relation $g_{k-1}^T d_{k-1} < 0$ and (2.23) yield

$$\begin{aligned} g_k^T d_{k-1} &\le g_k^T d_{k-1} - g_{k-1}^T d_{k-1} = d_{k-1}^T y_{k-1}, \\ g_k^T d_{k-1} &\ge \sigma_2 g_{k-1}^T d_{k-1} = -\sigma_2 d_{k-1}^T y_{k-1} + \sigma_2 g_k^T d_{k-1}, \end{aligned}$$

and hence we have from $\sigma_2 \in (0,1)$ and $d_{k-1}^T y_{k-1} > 0$ that

$$|g_k^T d_{k-1}| \le \max\left\{1,\ \frac{\sigma_2}{1-\sigma_2}\right\} d_{k-1}^T y_{k-1} = c_4 d_{k-1}^T y_{k-1}, \tag{3.23}$$

where $c_4 = \max\{1, \sigma_2/(1-\sigma_2)\}$.

Now we prove (3.21) for each case.

(i) By taking into account $z_{k-1}^{DL} = y_{k-1}$ and $h_{k-1}^{DL} = s_{k-1}$, it follows from (3.6), (3.22), (3.23), (3.20) and the Lipschitz continuity of g that

$$\begin{aligned} |\beta_k^{DSDL+}| &\le |g_k^T(z_{k-1}^{DL} - th_{k-1}^{DL})|\ |d_{k-1}^T z_{k-1}^{DL}|^\dagger + \lambda\|z_{k-1}^{DL} - th_{k-1}^{DL}\|^2 |g_k^T d_{k-1}| \{(d_{k-1}^T z_{k-1}^{DL})^2\}^\dagger \\ &\le \frac{\gamma(L+t)}{\bar{c}(1-\sigma_2)\varepsilon^2}\|s_{k-1}\| + \lambda\frac{c_4(L+t)^2}{\bar{c}(1-\sigma_2)\varepsilon^2}\|s_{k-1}\|^2 \\ &\le \left(\frac{\gamma(L+t)}{\bar{c}(1-\sigma_2)\varepsilon^2} + \lambda\frac{c_4(L+t)^2\widehat{a}}{\bar{c}(1-\sigma_2)\varepsilon^2}\right)\|s_{k-1}\|, \end{aligned}$$

which implies that (3.21) holds.

(ii) By (2.29), (3.22) and $\phi_k \ge 0$, we have

$$|d_{k-1}^T z_{k-1}^{YT+}| = \left| d_{k-1}^T y_{k-1} + \frac{\phi_k}{\alpha_{k-1}}\max\{0, \theta_{k-1}\}\right| \ge d_{k-1}^T y_{k-1} \ge \bar{c}(1-\sigma_2)\varepsilon^2. \tag{3.24}$$

It follows from (3.23) and (3.24) (namely, $|d_{k-1}^T z_{k-1}^{YT+}| \ge |d_{k-1}^T y_{k-1}|$) that

$$|g_k^T d_{k-1}| \le c_4 |d_{k-1}^T y_{k-1}| \le c_4 |d_{k-1}^T z_{k-1}^{YT+}|. \tag{3.25}$$

Similar to (3.8), we have from (2.29) that

$$\|z_{k-1}^{YT+} - th^{YT}\| = \left\| y_{k-1} + \phi_k \frac{\max\{0, \theta_{k-1}\}}{s_{k-1}^T u_{k-1}} u_{k-1} - ts_{k-1} \right\| \le c_5 \|s_{k-1}\|,$$

where $c_5 = t + (1 + 3\bar{\phi}/\bar{m})L$. Therefore, it follows from (3.24), (3.25) and (3.20) that

$$\begin{aligned} |\beta_k^{DSTY+}| &\le |g_k^T(z_{k-1}^{YT+} - th_{k-1}^{YT})|\ |d_{k-1}^T z_{k-1}^{YT+}|^\dagger + \lambda\|z_{k-1}^{YT+} - th_{k-1}^{YT}\|^2 |g_k^T d_{k-1}| \{(d_{k-1}^T z_{k-1}^{YT+})^2\}^\dagger \\ &\le \frac{\gamma c_5}{\bar{c}(1-\sigma_2)\varepsilon^2}\|s_{k-1}\| + \lambda\frac{c_4 c_5^2}{\bar{c}(1-\sigma_2)\varepsilon^2}\|s_{k-1}\|^2 \\ &\le \left(\frac{\gamma c_5}{\bar{c}(1-\sigma_2)\varepsilon^2} + \lambda\frac{c_4 c_5^2 \widehat{a}}{\bar{c}(1-\sigma_2)\varepsilon^2}\right)\|s_{k-1}\|, \end{aligned}$$

which implies that (3.21) holds.

(iii) The relations (1.5) and (3.17) yield

$$|d_{k-1}^T z_{k-1}^{F1}|^\dagger \le \frac{\varphi_1}{\bar{c}\varepsilon^2}, \qquad |g_k^T d_{k-1}|\ |d_{k-1}^T z_{k-1}^{F1}|^\dagger \le \varphi_1$$

and therefore, it follows from (3.11) and (3.20) that

$$\begin{aligned} |\beta_k^{DSF1+}| &\le |g_k^T(z_{k-1}^{F1} - th_{k-1}^{F1})|\ |d_{k-1}^T z_{k-1}^{F1}|^\dagger + \lambda \|z_{k-1}^{F1} - th_{k-1}^{F1}\|^2 |g_k^T d_{k-1}| \{(d_{k-1}^T z_{k-1}^{F1})^2\}^\dagger \\ &\le \frac{\varphi_1 \gamma c_6}{\bar{c}\varepsilon^2}\|s_{k-1}\| + \lambda \frac{\varphi_1^2 c_6^2}{\bar{c}\varepsilon^2}\|s_{k-1}\|^2 \\ &\le \left(\frac{\varphi_1 \gamma c_6}{\bar{c}\varepsilon^2} + \lambda \frac{\varphi_1^2 c_6^2 \widehat{a}}{\bar{c}\varepsilon^2}\right)\|s_{k-1}\|, \end{aligned}$$

where $c_6 = (1+\bar{\eta}/2)(L+t)$, which implies that (3.21) holds.

(iv) The relations (1.5) and (3.18) yield

$$|d_{k-1}^T z_{k-1}^{F2}|^\dagger \le \frac{\varphi_2}{\bar{c}\varepsilon^2}, \qquad |g_k^T d_{k-1}|\ |d_{k-1}^T z_{k-1}^{F2}|^\dagger \le \varphi_2$$

and therefore, it follows from (3.12) and (3.20) that

$$\begin{aligned} |\beta_k^{DSF2+}| &\le |g_k^T(z_{k-1}^{F2} - th_{k-1}^{F2})|\ |d_{k-1}^T z_{k-1}^{F2}|^\dagger + \lambda \|z_{k-1}^{F2} - th_{k-1}^{F2}\|^2 |g_k^T d_{k-1}| \{(d_{k-1}^T z_{k-1}^{F2})^2\}^\dagger \\ &\le \frac{\varphi_2 \gamma c_7}{\bar{c}\varepsilon^2}\|s_{k-1}\| + \lambda \frac{\varphi_2^2 c_7^2}{\bar{c}\varepsilon^2}\|s_{k-1}\|^2 \\ &\le \left(\frac{\varphi_2 \gamma c_7}{\bar{c}\varepsilon^2} + \lambda \frac{\varphi_2^2 c_7^2 \widehat{a}}{\bar{c}\varepsilon^2}\right)\|s_{k-1}\|, \end{aligned}$$

where $c_7 = L + t + \bar{\eta}t(L+1)/2$, which implies that (3.21) holds.

Summarizing (i)–(iv), the proof is complete. □

Although (3.17) and (3.18) look like strong assumptions, these are reasonable if we use (2.22) and the condition

$$-\sigma_3 g_k^T d_k \ge g(x_k + \alpha_k d_k)^T d_k \ge \sigma_2 g_k^T d_k, \tag{3.26}$$

as the line search rules, where $0 < \sigma_1 < \sigma_2 < 1$ and $0 < \sigma_3 < 1$. Note that the conditions (2.22) and (3.26) are the Wolfe conditions with the additional condition $-\sigma_3 g_k^T d_k \ge g(x_k + \alpha_k d_k)^T d_k$. We also note that, if $\sigma_3 = \sigma_2$, then the conditions (2.22) and (3.26) become the strong Wolfe condition: (2.22) and

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma_2 g_k^T d_k,$$

and in addition, if $\sigma_3 = 1 - 2\sigma_1$ and $\sigma_1 < 1/2$, then (2.22) and (3.26) become the approximate Wolfe condition: (2.22) and

$$-(1-2\sigma_1) g_k^T d_k \ge g(x_k + \alpha_k d_k)^T d_k \ge \sigma_2 g_k^T d_k. \tag{3.27}$$

We now demonstrate why (3.17) is reasonable under the conditions (2.22) and (3.26). If $d_{k-1}^T z_{k-1}^{F1} = 0$, then (3.17) is automatically satisfied. So that, we consider the case $d_{k-1}^T z_{k-1}^{F1} \neq 0$. Then we need to justify

$$|g_{k-1}^T d_{k-1}| \le \varphi_1 |d_{k-1}^T z_{k-1}^{F1}|, \tag{3.28}$$

$$|g_k^T d_{k-1}| \le \varphi_1 |d_{k-1}^T z_{k-1}^{F1}|. \tag{3.29}$$

It follows from (3.26) that

$$|g_k^T d_{k-1}| \le \max\{\sigma_2, \sigma_3\}|g_{k-1}^T d_{k-1}| \le |g_{k-1}^T d_{k-1}|,$$

which implies that (3.29) holds if (3.28) is satisfied. Thus it suffices to consider (3.28). From the definition of z_{k-1}^{F1} in (2.10), we have

$$d_{k-1}^T z_{k-1}^{F1} = d_{k-1}^T y_{k-1} - \xi_{k-1} d_{k-1}^T y_{k-2}. \tag{3.30}$$

If $s_{k-1}^T y_{k-2} \le 0$, then we have, from (3.30), $\xi_{k-1} > 0$ and (3.26), that

$$d_{k-1}^T z_{k-1}^{F1} \ge d_{k-1}^T y_{k-1} \ge -(1-\sigma_2) g_{k-1}^T d_{k-1} \quad (>0),$$

which implies (3.28) with $\varphi_1 = 1/(1-\sigma_2)$. If $s_{k-1}^T y_{k-2} > 0$, we can control the magnitude of the last term in (3.30) by using the parameter η_k. For example, if we choose $\eta_k = 0$, then $-\xi_{k-1} d_{k-1}^T y_{k-2} = 0$, which implies (3.28) holds with $\varphi_1 = 1/(1-\sigma_2)$. Thus (3.28) is justified. Therefore, (3.17) is reasonable. We can also justify (3.18) in a similar way for (3.17).

4 Numerical results

In this section, we give numerical results of Algorithm 2.1 to compare our methods with other conjugate gradient methods. We investigated numerical performance of the proposed algorithms on 70 problems from the CUTEr library [1, 10]. Note that we tested the proposed algorithms on 145 problems, but we omit the numerical results on small-scale problems. Table 2 shows problem names and dimensions of 70 problems.

Table 2: Test problems (problem name & dimension)

ARWHEAD	5000	DIXMAAND	9000	FLETCHCR	10000	PENALTY1	10000
BDEXP	5000	DIXMAANE	9000	FMINSRF2	5625	POWELLSG	20000
BDQRTIC	5000	DIXMAANF	9000	FMINSURF	5625	POWER	20000
BIGGSB1	5000	DIXMAANG	9000	FREUROTH	5000	QUARTC	10000
BOX	7500	DIXMAANH	9000	GENHUMPS	5000	SCHMVETT	5000
BROYDN7D	5000	DIXMAANI	9000	GENROSE	5000	SINQUAD	10000
BROYDN7D	10000	DIXMAANJ	9000	GENROSE	10000	SPARSINE	5000
BRYBND	10000	DIXMAANK	3000	LIARWHD	10000	SPARSQUR	10000
CHAINWOO	4000	DIXMAANL	9000	MODBEALE	10000	SROSENBR	10000
CHAINWOO	10000	DIXON3DQ	10000	MOREBV	5000	TESTQUAD	5000
COSINE	10000	DQDRTIC	5000	MOREBV	10000	TOINTGSS	10000
CRAGGLVY	5000	DQRTIC	5000	NONCVXU2	5000	TQUARTIC	10000
CURLY10	10000	EDENSCH	10000	NONDIA	10000	TRIDIA	10000
CURLY20	10000	EG2	1000	NONDQUAR	5000	VAREIGVL	5000
CURLY30	5000	ENGVAL1	10000	NONDQUAR	10000	WOODS	4000
DIXMAANA	9000	EXTROSNB	1000	NONSCOMP	5000	WOODS	10000
DIXMAANB	9000	EXTROSNB	10000	OSCIPATH	10000		
DIXMAANC	9000	FLETCHCR	1000	PENALTY1	1000		

We tested the following methods:

CG-DESCENT	:	Software by Hager and Zhang [11–13],
DSDL+	:	Algorithm 2.1 with β_k^{DSDL+} and $(\lambda, t) = (2, 0.3)$,
DSYT+	:	Algorithm 2.1 with β_k^{DSYT+}, $(\lambda, t, \phi_k) = (2, 0.3, 0.3)$ and $u_k = y_k$,
DSZZ+	:	Algorithm 2.1 with $\beta_k^{DSZZ+} \equiv \max\{0,\ \beta_k^{DSZZ}\}$ and $(\lambda, t) = (2, 0.3)$,
DSF1+	:	Algorithm 2.1 with β_k^{DSF1+} and $(\lambda, t, \eta_k) = (2, 0.3, 0.3)$,
DSF2+	:	Algorithm 2.1 with β_k^{DSF2+} and $(\lambda, t, \eta_k) = (2, 0.3, 0.3)$.

Following Zhou and Zhang [27], we chose, in DSZZ+, $\zeta = 0.001$, and $q = 1.0$ if $\|g_k\| \geq 1.0$, otherwise $q = 3.0$. Although Algorithm 2.1 with β_k^{DSZZ} converges globally for a general objective function, we used β_k^{DSZZ+}, instead of β_k^{DSZZ}, because the methods with β_k^{DSZZ+} performed a little better than the methods with $\beta_k = \beta_k^{DSZZ}$ did. We should note that the global convergence property of the method with β_k^{DSZZ+} is still established. To choose values of parameters λ, t, ϕ_k and η_k, we had preliminarily performed the algorithm by using two or three kinds of values for each parameter. In this numerical experiment, we chose values of parameters which relatively performed better in these preliminary numerical results. CG-DESCENT is a software package of conjugate gradient method with (1.6) and an efficient line search which computes the step size α_k satisfying the approximate Wolfe conditions (2.22) and (3.27). We coded DSDL+, DSYT+, DSZZ+, DSF1+ and DSF2+ by using CG-DESCENT [11–13], in which parameters were set as $\sigma_1 = 10^{-4}$ and $\sigma_2 = 0.1$. The stopping condition was

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

We also stopped the algorithm if CPU time exceeded 500(sec). We note from the numerical results in [21] that the three-term conjugate gradient method given by Sugiki et al. [21] is almost comparable with CG-DESCENT. Thus we omit numerical comparisons our methods with the three-term conjugate gradient method by Sugiki et al.

We adopt the performance profiles by Dolan and Moré [4] to compare the performance among the tested methods. For n_s solvers and n_p problems, the performance profile $P : \boldsymbol{R} \to [0, 1]$ is defined as follows:

Let $\mathcal{P}$ and $\mathcal{S}$ be the set of problems and the set of solvers, respectively. For each problem $p \in \mathcal{P}$ and for each solver $s \in \mathcal{S}$, we define $t_{p,s} :=$ (computing time (or number of iterations, etc.) required to solve problem p by solver s). The performance ratio is given by $r_{p,s} := t_{p,s}/ \min_{s\in\mathcal{S}} t_{p,s}$. Then, the performance profile is defined by $P(\tau) := \frac{1}{n_p}\text{size}\{p \in \mathcal{P} | r_{p,s} \leq \tau\}$ for all $\tau \in \boldsymbol{R}$, where size$\{p \in \mathcal{P} | r_{p,s} \leq \tau\}$ stands for the number of elements of the set $\{p \in \mathcal{P} | r_{p,s} \leq \tau\}$. Note that if the performance profile of a method is over the performance profiles of the other methods, then this method performed better than the other methods.

Figures 1–4 are the performance profiles measured by CPU time, the number of iterations, the number of function evaluations and the number of gradient evaluations, respectively. From the viewpoint of CPU time, we see from Figure 1 that CG-DESCENT performed well in the interval $1 \leq \tau \leq 2$, and DSF1+ and DSF2+ were at least comparable with CG-DESCENT in the interval $2 \leq \tau \leq 5$. On the other hand, DSDL+ and DSYT+ were outperformed by CG-DESCENT. From the viewpoint of the number of iterations, the number of function evaluations and the number of gradient evaluations, Figures 2–4 show that DSF1+ and DSF2+ were superior to CG-DESCENT, and that DSDL+, DSYT+ and DSZZ+ were almost comparable with CG-DESCENT. CG-DESCENT is coded to reduce the computational costs of inner-products that is needed in Hager-Zhang's method [11]. On the other hand, DSDL+, DSYT+, DSZZ+, DSF1+ and DSF2+ are not tuned to reduce the computational costs of inner-products, and hence, as mentioned above, DSDL+, DSYT+, DSZZ+, DSF1+ and DSF2+ need more computational costs for inner-products than CG-DESCENT does. This is a reason why CG-DESCENT is superior to the other methods in the interval $1 \leq \tau \leq 2$ of Figular 1. We may reduce computational costs of DSDL+, DSYT+, DSZZ+, DSF1+ and DSF2+ by effectively tuning the code. However, it is beyond the scope of this paper.

Summarizing the above observations, we can conclude that DSF1+ and DSF2+ are efficient in this numerical experiments, and DSDL+, DSYT+ and DSZZ+ are almost comparable with CG-DESCENT. On the other hand, since the methods have some parameters, a suitable choice of parameters in the methods is our further study.

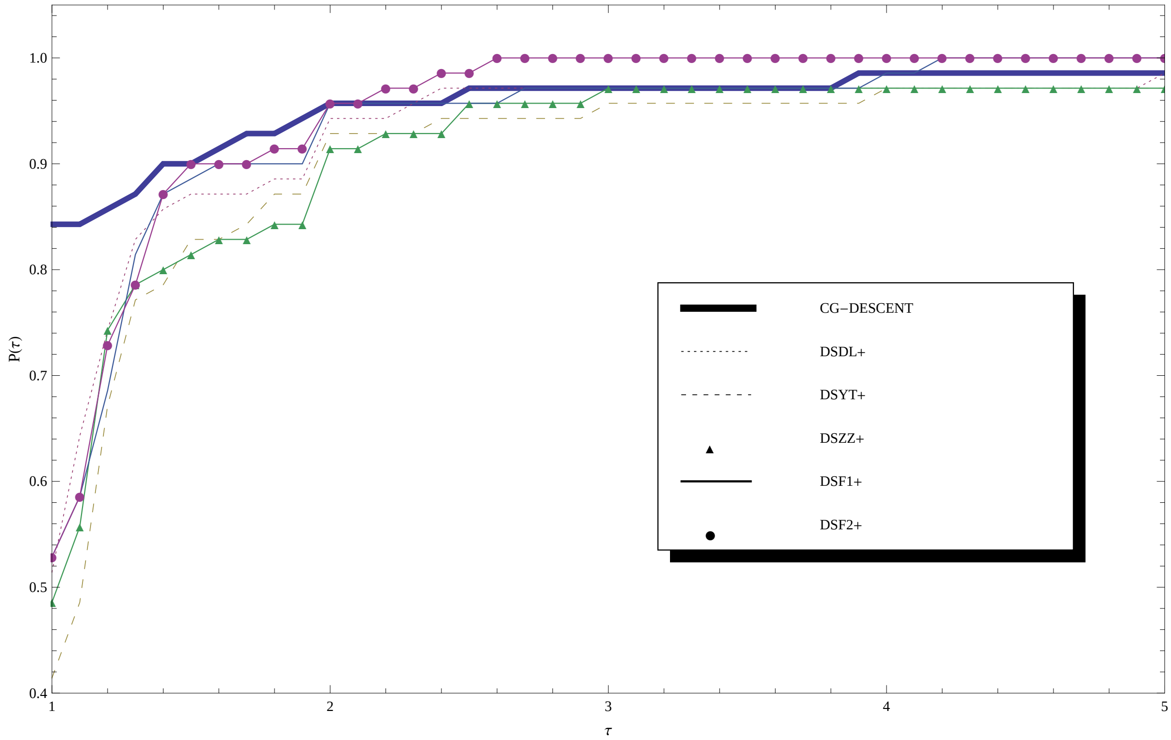

Figure 1: Performance based on CPU time

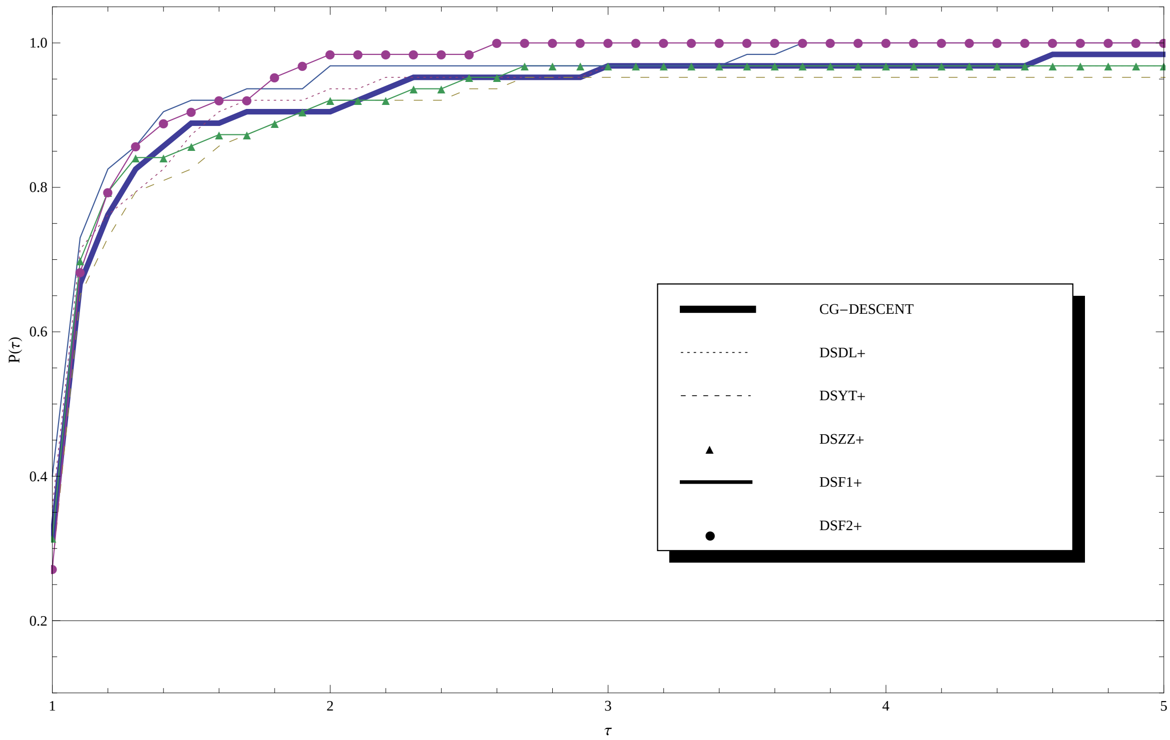

Figure 2: Performance based on the number of iterations

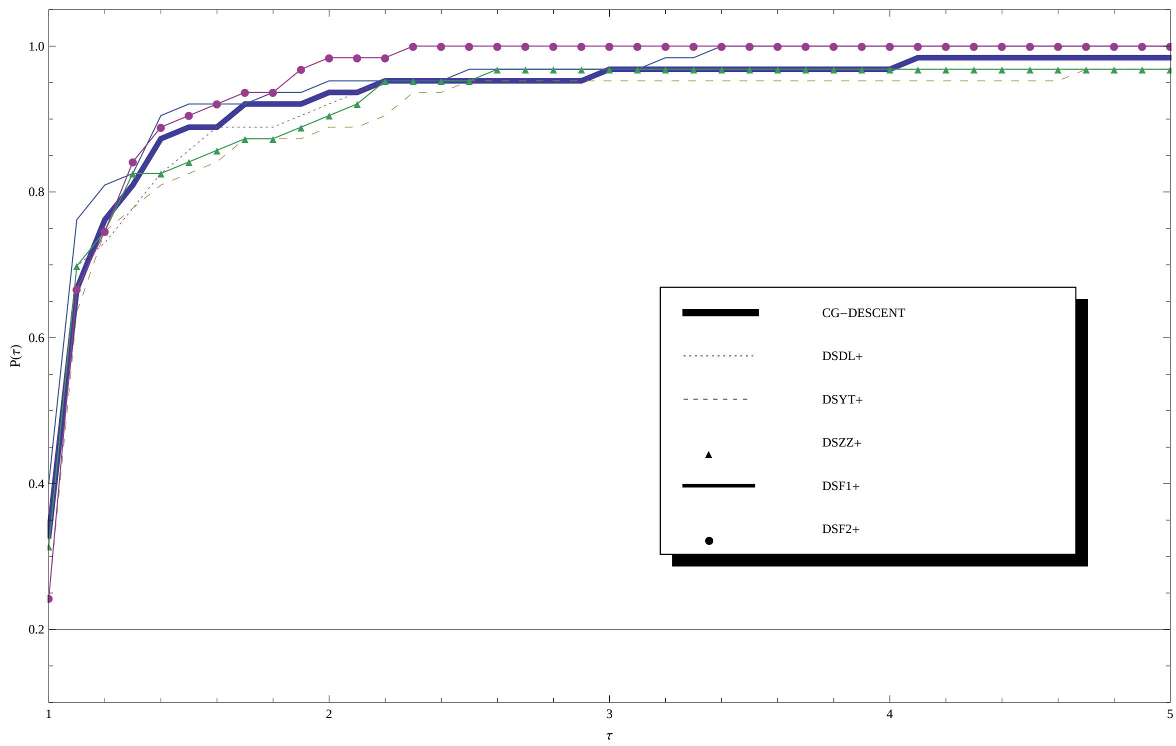

Figure 3: Performance based on the number of function evaluations

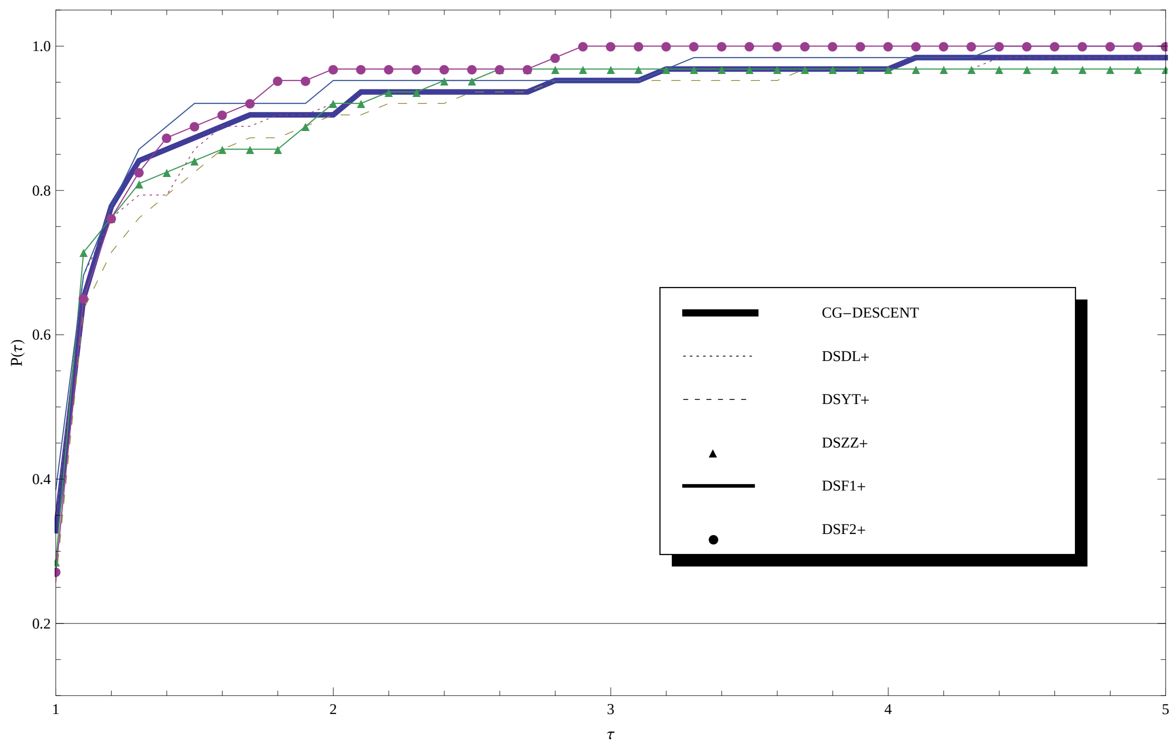

Figure 4: Performance based on the number of gradient evaluations

5 Conclusion

In this paper, we have proposed conjugate gradient methods based on secant conditions that generate descent search directions. Under suitable assumptions, our methods have been shown to converge globally. In numerical experiments, we have confirmed the effectiveness of the proposed methods by using performance profiles.

Acknowledgements

The authors are supported in part by the Grant-in-Aid for Scientific Research (C) 21510164 of Japan Society for the Promotion of Science.

References

[1] I. Bongartz, A.R. Conn, N.I.M. Gould and P.L. Toint, CUTE: constrained and unconstrained testing environment, *ACM Transactions on Mathematical Software*, **21** (1995), 123–160.

[2] Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, *Applied Mathematics and Optimization*, **43** (2001), 87–101.

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

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

[5] R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients, *Computer Journal*, **7** (1964), 149–154.

[6] J.A. Ford and I.A. Moghrabi, Alternative parameter choices for multi-step quasi-Newton methods, *Optimization Methods and Software*, **2** (1993), 357–370.

[7] J.A. Ford and I.A. Moghrabi, Multi-step quasi-Newton methods for optimization, *Journal of Computational and Applied Mathematics*, **50** (1994), 305–323.

[8] J.A. Ford, Y. Narushima and H. Yabe, Multi-step nonlinear conjugate gradient methods for unconstrained minimization, *Computational Optimization and Applications*, **40** (2008), 191–216.

[9] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, *SIAM Journal on Optimization*, **2** (1992), 21–42.

[10] N.I.M. Gould, D. Orban and P.L. Toint, CUTEr web site, *http://www.cuter.rl.ac.uk/*.

[11] W.W. Hager and H. Zhang, A new conjugate gradient method with guaranteed descent and an efficient line search, *SIAM Journal on Optimization*, **16** (2005), 170–192.

[12] W.W. Hager and H. Zhang, CG_DESCENT Version 1.4 User' Guide, University of Florida, November 2005, *http://www.math.ufl.edu/~hager/*.

[13] W.W. Hager and H. Zhang, Algorithm 851: CG_DESCENT, A conjugate gradient method with guaranteed descent, *ACM Transactions on Mathematical Software*, **32** (2006), 113–137.

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

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

[16] M. Kobayashi, Y. Narushima and H. Yabe, Nonlinear conjugate gradient methods with structured secant condition for nonlinear least squares problems, *Journal of Computational and Applied Mathematics*, **234** (2010), 375–397.

[17] D.H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization, *Journal of Computational and Applied Mathematics*, **129** (2001), 15–35.

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

[19] J. Nocedal and S.J. Wright, *Numerical Optimization: Second Edition*, Springer Series in Operations Research, Springer-Verlag, New York, 2006.

[20] A. Perry, A modified conjugate gradient algorithm, *Operations Research*, **26** (1978), 1073–1078.

[21] K. Sugiki, Y. Narushima and H. Yabe, Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization, The Institute of Statistical Mathematics Cooperative Research Report, Optimization -Modeling and Algorithms-, **23** (2010),165–183.

[22] H. Yabe and M. Takano, Global convergence properties of nonlinear conjugate gradient methods with modified secant condition, *Computational Optimization and Applications*, **28** (2004), 203–225.

[23] G. Yu, L. Guan and G. Li, Global convergence of modified Polak-Ribiére-Polyak conjugate gradient methods with sufficient descent property, *Journal of Industrial and Management Optimization*, **4** (2008), 565–579

[24] G. Yuan, Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems, *Optimization Letters*, **3** (2009), 11–21.

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

[26] J.Z. Zhang and C.X. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations, *Journal of Computational and Applied Mathematics*, **137** (2001), 269–278.

[27] W. Zhou and L. Zhang, A nonlinear conjugate gradient method based on the MBFGS secant condition, *Optimization Methods and Software*, **21** (2006), 707–714.

[28] G. Zoutendijk, Nonlinear programming, computational methods, in Integer and Nonlinear Programming, J.Abadie, ed., North-Holland, Amsterdam, 37-86, 1970.