A Descent Modified HS Conjugate Gradient Method with an Optimal Property. *

Fenfei Chen † Jiahao Zhang ‡ Jianhao Ye § Zixin Chen ¶

Abstract

In this paper, by minimizing the distance between the CG direction and the direction of the improved Perry conjugate gradient method [?], we propose a descent modified HS conjugate gradient method. A remarkable property of the modified HS method is that it can produce sufficient descent property, which is independent of the line search used. Under suitable conditions, we prove that the modified HS method with the standard Armijo line search is globally convergent for uniformly convex functions and the modified HS+ method with standard Wolfe line search is globally convergent for general nonlinear functions. Extensive numerical experiments show that the proposed method is efficient.

Keywords: HS method, unconstrained optimization, global convergence

AMS subject classification: 90C06, 90C25, 65Y20, 94A08

1. Introduction

Consider the unconstrained optimization problem

$$\min f(x), \quad x \in \mathcal{R}, \tag{1.1}$$

where f is smooth and its gradient g is available. Conjugate gradient methods are welcome methods for solving (**??**), especially for large scale problems due to their simplicity and low storage. A typical conjugate gradient method has the form:

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \cdots,$$

*Supported by the Chinese NSF Grant (nos. 12271187, 11961011 and 11971106), by the Ministry of Education, Humanities and Social Sciences project (no.17YJAZH011) and by special projects in key fields of Guangdong Universities(2021ZDZX1054) and by the Natural Science Foundation of Guangdong province (2022A1515010567).

†School of Economics and Management, Beijing Jiaotong University, Beijing 523000, China. E-mail: ffchen@bjtu.edu.cn

‡College of Computer and Science Technology, Dongguan University of Technology, Dongguan 523000 China. E-mail: 1095199414@qq.com

§College of Computer and Science Technology, Dongguan University of Technology, Dongguan 523000 China. E-mail: 458523944@qq.com

¶Corresponding author, College of Computer and Science Technology, Dongguan University of Technology, Dongguan 523000 China. E-mail: 275191090@qq.com

where α_k is the stepsize determined by some line search and d_k is the search direction determined by

$$d_k = \begin{cases} -g_k, & \text{if } \ k = 0, \\ -g_k + \beta_{k-1} d_{k-1} & \text{if } \ k > 0, \end{cases} \tag{1.2}$$

where β_{k-1} is the conjugate gradient parameter. Different choices of β_k lead to different conjugate gradients. Well known formulate for β_k are the Fletcher-Reeves(FR) [?], Hestenes-Stiefel [?], Polak-Ribiere-Polyak [?, ?]and Dai-Yuan [?] which are given by

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \quad \beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k},$$

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}, \quad \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k},$$

where $y_k = g_{k+1} - g_k$. It is well known that the HS method is generally believed to be one of the most efficient conjugate gradient methods. However, the HS method lacks the descent property. Much efforts has been made to find some modified HS methods which not only satisfy some conjugate condition but also produce sufficient descent direction. Dai and Liao [?] considered the conjugacy condition

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k \tag{1.3}$$

and derived a new formula for β_k

$$\beta_k^{DL}(t) = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{1.4}$$

where t is some parameter and $s_k = x_{k+1} - x_k$. Dai and Liao [?] proved that the conjugate gradient method with

$$\beta_k^{DL+}(t) = \max\{\frac{g_{k+1}^T y_k}{d_k^T y_k}, 0\} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}$$

is globally convergent for general functions. Combining with self-scaling memoryless BFGS method, Hager and Zhang [?] proposed the formula

$$\beta_k^N = \frac{g_{k+1}^T y_k}{d_k^T y_k} - 2 \frac{\|y_k\|^2}{d_k^T y_k} \frac{g_{k+1}^T s_k}{d_k^T y_k}$$

which can be regarded as (??) with $t = 2\frac{\|y_k\|^2}{s_k^T y_k}$. A good property of their method is that the direction d_k with β_k^N satisfies sufficient descent property $g_{k+1}^T d_{k+1} \le -\frac{7}{8}\|g_{k+1}\|^2$ which is independent of the line search used. To establish the global convergence of general functions, they proposed the truncated form:

$$\bar{\beta}_k^N = \max\{\beta_k^N, \frac{-1}{\|d_k\| \min\{\eta, \|g_k\|\}}\}.$$

By seeking the conjugate gradient direction that is closest to the direction of the scaled memory BFGS method, Dai and Kou [?] proposed the following formula

$$\beta_k(\tau_k) = \frac{g_{k+1}^T y_k}{d_k^T y_k} - (\tau_k + \frac{\|y_k\|^2}{d_k^T y_k} - \frac{s_k^T y_k}{\|s_k\|^2}) \frac{g_{k+1}^T s_k}{d_k^T y_k} \tag{1.5}$$

where τ_k is a parameter. The parameter $\beta_k(\tau_k)$ can also be regarded as (**??**) with $t = \tau_k + \frac{\|y_k\|^2}{s_k^T y_k} - \frac{s_k^T y_k}{\|s_k\|^2}$. To establish the global convergence of general functions, they considered the following truncated form

$$\beta_k^+(\tau_k) = \max\{\beta_k(\tau_k), \eta \frac{g_{k+1}^T d_k}{\|d_k\|^2}\}.$$

Numerical results [**?**] shows that their method with $\tau_k = \frac{s_k^T y_k}{\|s_k\|^2}$ performs best among four different choices of τ_k. We refer to a recent review papers [**?**, **?**, **?**, **?**, **?**, **?**] for details about the progress of conjugate gradient methods.

Recently, Yao et al. [**?**], proposed an improved Perry conjugate gradient method in which the direction can be written as a quasi-Newton direction. They showed that the method is not only globally convergence but also produces sufficient descent direction independent of the line search. For a large collection of test problems from CUTE library [?], they showed that this method works well in practice. Motivated by theoretical and numerical features of the method [**?**], by the use of the idea of Dai and Kou [**?**] we proposed a descent modified HS method. Specifically, we obtain the new parameter β_k by minimizing the distance between the CG direction and the direction of the improved Perry conjugate gradient method [**?**]. The proposed method satisfies Dai-Liao conjugate condition and the new β_k can be seen as a special case of $\beta_k^{DL}(t)$. A common property of the modified HS method is that it can produce sufficient descent property, which is independent of the line search used. Moreover, if the exact line search is used, the method reduces to the standard HS method. Under suitable conditions, we prove that the modified HS method with the standard Armijo line search is globally convergent for uniformly convex functions and the modified HS+ method with standard Wolfe line search is globally convergent for general nonlinear functions. Extensive numerical experiments show that the proposed method is efficient.

The remainder of the paper is organized as follows. In Section 2, a modified HS method is proposed which satisfied the sufficient descent property independent of line search. In Section 3, we show global convergence of the proposed method. Some numerical results are given in Section 4. Finally, some concluding remarks are listed in Section 5.

2. A descent modified HS conjugate gradient method

In this section, by minimizing the distance between the CG direction and the direction of the improved Perry conjugate gradient method [**?**], we propose a descent modified HS method. The descent modified HS method can produce sufficient descent property, which is independent of the line search used.

Yao et al. [**?**] proposed an improved Perry conjugate gradient method as follows

$$d_{k+1}^p = -g_{k+1} + \beta_k d_k + \theta_k y_k, \text{ and } d_0 = -g_0, \tag{2.6}$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k and

$$\beta_k = \frac{g_{k+1}^T y_k}{d_k^T y_k} - 2\frac{\|y_k\|^2}{s_k^T y_k}\frac{g_{k+1}^T s_k}{d_k^T y_k}, \quad \theta_k = \frac{g_{k+1}^T d_k}{d_k^T y_k}.$$

Note that the direction d_{k+1}^p can be written a quasi-Newton direction

$$d_{k+1}^p = -Q_{k+1}g_{k+1} = -(I - \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + 2\frac{\|y_k\|^2}{s_k^T y_k}\frac{s_k s_k^T}{s_k^T y_k})g_{k+1}.$$

They showed that Q_{k+1} is symmetric and positive definite. Correspondingly, the direction d_{k+1}^p produces sufficient descent property independent of line search. Since the direction of a typical conjugate gradient method is a combination of $-g_{k+1}$ and d_k, this and the good properties of this direction (**??**) motivate us to find a new two-term direction d_{k+1} close to d_{k+1}^p. Similar to the idea of Dai and Kou [**?**], we find the parameter β_k by solving the following optimization problem

$$\min_{\beta} \| - g_{k+1} + \beta d_k - d_{k+1}^P\|^2.$$

After easily computation, we obtain a unique solution as follows

$$\beta_k = \frac{g_{k+1}^T y_k}{d_k^T y_k} - (2\frac{\|y_k\|^2}{s_k^T y_k} + \frac{s_k^T y_k}{\|s_k\|^2})\frac{g_{k+1}^T s_k}{d_k^T y_k}. \tag{2.7}$$

If the exact line search is used, then the β_k reduces to the standard β_k^{HS}. Moreover, it is worth noticing that the new parameter β_k corresponds to the Dai-Liao formula (**??**) with $\bar{t} = 2\frac{\|y_k\|^2}{s_k^T y_k} + \frac{s_k^T y_k}{\|s_k\|^2}$. This implies that the direction d_{k+1} with the parameter (**??**) satisfies Dai-Liao conjugate condition (**??**) with $t = \bar{t}$. Also it corresponds to Dai-Kou formula (**??**) with $\tau_k = \frac{\|y_k\|^2}{s_k^T y_k} + 2\frac{s_k^T y_k}{\|s_k\|^2}$.

The following lemma shows that the direction (**??**) with the parameter (**??**) produces sufficient descent method indepentdent of the line search and the convexity of the objection function.

Lemma 2.1. *Consider the CG method with the parameter (**??**). If $d_k^T y_k \neq 0$, then*

$$g_{k+1}^T d_{k+1} \leq -\frac{7}{8}\|g_{k+1}\|^2. \tag{2.8}$$

Proof. Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2$, which satisfies (**??**). By the definition of d_{k+1}, we have

$$\begin{aligned}
d_{k+1}^T g_{k+1} &= -\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k \\
&= -\|g_{k+1}\|^2 + \frac{g_{k+1}^T y_k}{d_k^T y_k} g_{k+1}^T d_k - 2\frac{\|y_k\|^2}{s_k^T y_k}\frac{g_{k+1}^T s_k}{d_k^T y_k} g_{k+1}^T d_k - \frac{s_k^T y_k}{\|s_k\|^2}\frac{g_{k+1}^T s_k}{d_k^T y_k} g_{k+1}^T d_k \\
&= -\|g_{k+1}\|^2 + \frac{g_{k+1}^T}{2}\frac{2y_k g_{k+1}^T d_k}{d_k^T y_k} - 2\frac{\|y_k\|^2}{(d_k^T y_k)^2}(g_{k+1}^T d_k)^2 - \frac{(g_{k+1}^T d_k)^2}{\|d_k\|^2} \\
&\leq -\|g_{k+1}\|^2 + \frac{1}{2}(\frac{1}{4}\|g_{k+1}\|^2 + 4\frac{\|y_k\|^2 (g_{k+1}^T d_k)^2}{(d_k^T y_k)^2}) - 2\frac{\|y_k\|^2}{(d_k^T y_k)^2}(g_{k+1}^T d_k)^2 \\
&= -\frac{7}{8}\|g_{k+1}\|^2.
\end{aligned}$$

The proof is completed. □

3. Convergence analysis

In this section, under suitable conditions, we prove that the modified HS method with the standard Armijo line search is globally convergent for uniformly convex functions and the modified HS+ method with Wolfe line search is globally convergent for general nonlinear functions. The standard Armijo line search is to find steplength $\alpha_k = \max\{\rho^j\}_{j=0,1,\cdots}$ satisfying

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g(x_k)^T d_k, \tag{3.1}$$

where $\delta \in (0,1)$ and $\rho \in (0,1)$ are constants. And the Wolfe line search is to find steplength α_k satisfying

$$\begin{cases} f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g(x_k)^T d_k, \\ g(x_k + \alpha_k d_k)^T d_k \geq -\sigma g(x_k)^T d_k, \end{cases} \tag{3.2}$$

where $0 < \delta < \sigma < 1$ are constants. We make the following assumptions on the objective function.

Assumption 3.1:

(1) The level set $\Theta := \{x \in \mathcal{R}^n : f(x) \leq f(x^0)\}$ is bounded.

(2) In some neighbourhood $\mathcal{N}$ of Θ, f is continuous differentiable and the gradient of f is Lipschitz continuous with constant $L > 0$, i.e.,

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathcal{N}. \tag{3.3}$$

Under Assumption 3.1, there exist positive constants B and γ such that

$$\|x - y\| \leq B, \quad \forall x, y \in \mathcal{N} \tag{3.4}$$

and

$$\|g(x)\| \leq \gamma, \quad \forall x \in \mathcal{N}. \tag{3.5}$$

The following lemma plays an important role in establishing global convergence of the proposed method for uniformly convex function and general functions, respectively.

Lemma 3.1. *Suppose that Assumption 3.1 holds and $f(x)$ is bounded below. Let $\{x_k\}$ be generated by the modified HS method, where α_k is obtained by Armijo line search or Wolfe line search. Then we have*

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.6}$$

Proof. Since $f(x)$ is bounded below and is monotonically nonincreasing, there exists a constant $\bar{f}$ such that

$$\lim_{k\to\infty} f(x_k) = \bar{f}. \tag{3.7}$$

We shall give a low bounded for α_k. Now, we consider the following two cases. Case 1. Suppose that α_k satisfies the Wolfe conditions. By the second inequality of (**??**), we have

$$(g(x_{k+1}) - g(x_k))^T d_k \geq (\sigma - 1) g(x_k)^T d_k.$$

By (**??**) and Lemma **??**, we get

$$\alpha_k \geq \frac{(\sigma-1)g(x^k)^T d_k}{L\|d_k\|^2} \geq \frac{7(1-\sigma)\|g_k\|^2}{8L\|d_k\|^2}. \tag{3.8}$$

Case 2. Suppose that α_k satisfies the Armijo condition. By the use of Cauchy-Schwarz inequality to (**??**), we get

$$\|g_k\| \leq \frac{8}{7}\|d_k\|.$$

If $\alpha_k = 1$, there exists a positive constant c_1 such that

$$\alpha_k = 1 > c_1 \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{3.9}$$

If $\alpha_k \neq 1$, by the line search rule (**??**), we get

$$f(x_k + \frac{\alpha_k}{\rho} d_k) > f(x_k) + \frac{\alpha_k}{\rho}\delta g_k^T d_k.$$

By (**??**) we get

$$f(x_k + \frac{\alpha_k}{\rho} d_k) - f(x_k) \leq \frac{\alpha^k}{\rho} g_k^T d_k + \frac{L}{2}(\frac{\alpha_k}{\rho})^2\|d_k\|^2.$$

The last two inequalities yield

$$\alpha_k \geq \frac{(1-\delta)7\rho\|g_k\|^2}{4L\|d_k\|^2}. \tag{3.10}$$

(**??**), (**??**) and (**??**) imply that there exists a positive constant $c > 0$ such that

$$\alpha_k \geq c\frac{\|g_k\|^2}{\|d_k\|^2} \tag{3.11}$$

for all k. By (**??**), (**??**) and $\|g_k\| \geq \epsilon$, we get

$$\frac{\|g_k\|^4}{\|d_k\|^2} c\delta \leq f(x_k) - f(x_{k+1}).$$

Summing these inequalities and using (**??**), we get the conclusion. □

3.1 Global convergence for uniformly convex functions

In this subsection, we prove that the modified HS method with the standard Armijo line search is globally convergent for uniformly convex functions. By the uniform convexity of $f(x)$, there exists a constant $\mu > 0$ such that

$$(g(x) - g(y))^T (x - y) \geq \mu\|x - y\|^2, \quad \forall x, y \in \mathcal{R}^n. \tag{3.12}$$

The following theorem established the global convergence of the modified HS method with the standard Armijo line search for uniformly convex functions.

Theorem 3.1. *Suppose that Assumption 3.1 holds, $f(x)$ is uniformly convex and bounded below. Let $\{x_k\}$ be generated by the modified HS method, where β_k is determined by (**??**) and α_k is obtained by Armijo line search or Wolfe line search. Then we have*

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

Proof. It follows that (**??**) and (**??**) that

$$\|y_k\| \le L\|s_k\| \quad \text{and} \quad d_k^T y_k \ge \mu\|d_k\|\|s_k\|.$$

Thus,

$$\frac{\|y_k\|^2}{s_k^T y_k} \le \frac{L^2}{\mu} \quad \text{and} \quad \frac{s_k^T y_k}{\|s_k\|^2} \le L.$$

By (**??**), we get

$$\begin{aligned} \|d_{k+1}\| &\le \|g_{k+1}\| + |\frac{g_{k+1}^T y_k}{d_k^T y_k}|\|d_k\| + (L + 2\frac{L^2}{\mu})|\frac{g_{k+1}^T s_k}{d_k^T y_k}|\|d_k\| \\ &\le \gamma + \frac{\gamma L\|s_k\|}{\mu\|d_k\|\|s_k\|}\|d_k\| + (L + 2\frac{L^2}{\mu})\frac{\gamma\|s_k\|}{\mu\|d_k\|\|s_k\|}\|d_k\| \\ &\le \gamma(1 + 2\frac{L}{\mu} + 2\frac{L^2}{\mu^2}). \end{aligned}$$

By (**??**), we get the conclusion. □

3.2 Global convergence for general functions

If the exact line search is used, then the new β_k reduces to the standard β_k^{HS}. The example constructed by Powell [**?**] shows that the proposed method can not guarantee global convergence for general functions. Similar to the idea of Dai and Kou [**?**], we replace (**??**) by

$$\beta_k^+ = \max\{\beta_k, \eta\frac{g_{k+1}^T d_k}{\|d_k\|^2}\}. \tag{3.13}$$

where $0 \le \eta < 1$. If $\beta_k^+ = \eta\frac{g_{k+1}^T d_k}{\|d_k\|^2}$, by Gauchy-Schwarz inequality, we have

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \eta\frac{(g_{k+1}^T d_k)^2}{\|d_k\|^2} \le -(1-\eta)\|g_{k+1}\|^2.$$

This together with Lemma **??** implies that there exists a positive constant b such that

$$d_{k+1}^T g_{k+1} \le -b\|g_{k+1}\|^2. \tag{3.14}$$

Note that Lemma **??** still holds for the modified HS method with (**??**).

The following lemma is similar to Lemma 3.4 in [**?**].

Lemma 3.2. *Suppose that Assumption 3.1 holds and $f(x)$ is bounded below. Let $\{x_k\}$ be generated by the modified HS method with (**??**) and α_k be obtained by Wolfe line search. If $\|g_k\| \geq \epsilon$ for all $k \geq 1$, then $d_k \neq 0$ and*

$$\sum_{k=1}^{\infty} \|u_k - u_{k-1}\|^2 < \infty,$$

where $u_k = \frac{d_k}{\|d_k\|}$.

Proof. It is clear that $d_k \neq 0$. Otherwise the condition (**??**) would imply $g_k = 0$. We divided formula (**??**) into the following two parts:

$$\beta_k^1 = \max\{\frac{g_{k+1}^T y_k}{d_k^T y_k} - 2\frac{\|y_k\|^2 g_{k+1}^T d_k}{(d_k^T y_k)^2} + (1-\eta)\frac{g_{k+1}^T d_k}{\|d_k\|^2}, 0\}$$

and

$$\beta_k^2 = \eta \frac{g_{k+1}^T d_k}{\|d_k\|^2}.$$

Define

$$w_k = \frac{-g_k + \beta_{k-1}^2 d_{k-1}}{\|d_k\|}, \quad \text{and} \quad \delta_{k-1} = \frac{\beta_{k-1}^1 \|d_{k-1}\|}{\|d_k\|}.$$

By $d_k = -g_k + \beta_{k-1} d_{k-1}$, we have for $k \geq 1$

$$u_k = w_k + \delta_{k-1} u_{k-1}.$$

Since $\|u_k\| = \|u_{k-1}\| = 1$ and $\delta_{k-1} > 0$, we get

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

and

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

Note that

$$\|w_k\| \leq \frac{\|g_k\| + \eta \frac{|g_k^T d_{k-1}|}{\|d_{k-1}\|^2}\|d_{k-1}\|}{\|d_k\|} \leq \frac{(1+\eta)\|g_k\|}{\|d_k\|}.$$

Since $\|g_k\| \geq \epsilon$, by the last two inequalities and (**??**) we get

$$\sum_{k \geq 1} \|u_k - u_{k-1}\|^2 \leq \sum_{k \geq 1} 4(1+\eta)^2 \frac{\|g_k\|^2}{\|d_k\|^2} \leq \sum_{k \geq 1} 4\frac{(1+\eta)^2}{\epsilon^2} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty.$$

□

The following theorem established the global convergence of the modified HS method with (**??**) for general functions.

Theorem 3.2. *Suppose that Assumption 3.1 holds and $f(x)$ is bounded below. Let $\{x_k\}$ be generated by the modified HS method, where β_k is determined by (**??**) and α_k is obtained by Wolfe line search. Then we have*

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

Proof. Assume that the conclusion does not hold. Then there exists a positive constant ϵ such that

$$\|g_k\| \geq \epsilon \tag{3.15}$$

for all k. Letting $\gamma_1 = B + \gamma$, by (**??**) and (**??**), we get

$$\|x_k\| \leq \gamma_1 \ \text{ and } \ \|g_k\| \leq \gamma_1, \ \ \forall k \geq 1. \tag{3.16}$$

If $\beta_k = \eta \frac{g_{k+1}^T d_k}{\|d_k\|^2}$, then we have

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \eta \frac{|g_{k+1}^T d_k|}{\|d_k\|^2}\|d_k\| \leq (1+\eta)\|g_{k+1}\| \leq (1+\eta)\gamma_1.$$

By (**??**), we get

$$\|d_k\| \to \infty,$$

which implies that $\beta_k^+ = \beta_k$ for all sufficiently large k. The second inequality of (**??**) indicates

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k. \tag{3.17}$$

It follows from (**??**), (**??**) and (**??**) that

$$d_k^T y_k \geq (\sigma - 1) g_k^T d_k \geq b(1-\sigma)\epsilon^2 := d. \tag{3.18}$$

By (**??**), (**??**) and the last inequality, we get

$$\frac{\sigma}{\sigma - 1} \leq \frac{d_k^T g_{k+1}}{d_k^T y_k} \leq 1.$$

By (**??**), (**??**), (**??**) and the last inequality, we have

$$\begin{aligned}
|\beta_k| &\leq |\frac{g_{k+1}^T y_k}{d_k^T y_k}| + (2\frac{\|y_k\|^2}{s_k^T y_k} + \frac{s_k^T y_k}{\|s_k\|^2})|\frac{g_{k+1}^T s_k}{d_k^T y_k}| \\
&\leq \frac{\gamma_1 L}{d}\|s_k\| + \frac{L\|s_k\|}{\|s_k\|^2}\frac{\|g_{k+1}\|\|s_k\|^2}{d} + 2\frac{\|y_k\|^2}{d_k^T y_k}|\frac{g_{k+1}^T d_k}{d_k^T y_k}| \\
&\leq 2\frac{\gamma_1 L}{d}\|s_k\| + 2\frac{L^2\|s_k\|^2}{d}\max\{\frac{\sigma}{1-\sigma}, 1\} \\
&\leq 2\frac{\gamma_1 L}{d}\|s_k\| + \frac{4\gamma_1 L^2\|s_k\|}{d}\max\{\frac{\sigma}{1-\sigma}, 1\} \\
&= C\|s_k\|
\end{aligned}$$

where $C = 2\frac{\gamma_1 L}{d} + \frac{4\gamma_1 L^2}{d}\max\{\frac{\sigma}{1-\sigma}, 1\}$. Define $b = 2C\gamma_1$ and $\upsilon = \frac{1}{2C^2\gamma_1}$. It follows from (**??**) and the last inequality that

$$|\beta_k| \leq b$$

and

$$\|s_k\| \leq v \Rightarrow |\beta_k| \leq \frac{1}{b}.$$

The last two inequalities imply that β_k has Property (*) in [?]. Proceeding the similar analysis of Theorem 4.4 [?], we get the conclusion. □

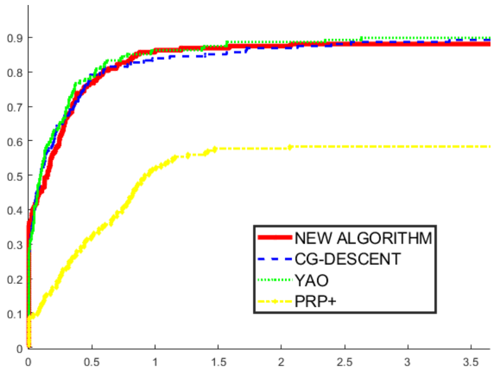
4. Numerical Experiments

In this section, we present some numerical results of the new algorithm. All tests are performed under the configuration of Windows10 operating system (64-bit), Intel(R) Core(TM) i3-8145U CPU @2.10GHz 2.30GHz, 4.00GB. We compare the new algorithm (??) with the following algorithm: CG_DESCENT method [?], Yao method [?] and PRP+ [?]. We set $\eta = 0.7$ in (??). We use the CG_DESCENT code (version 5.3) for testing, which can be obtained from Hager's homepage: https://people.clas.ufl.edu/hager/software-archive/. The PRP+ code was obtained from Jorge Nocedal's homepage page: http://www.ece.northwestern.edu/ nocedal/software.html. All parameters of CG_DESCENT and PRP+ are default. The new algorithm (??) and Yao method [?] is performed with the same line search as that of CG_DESCENT method. We tested 84 problems with dimensions 10000 and 150000 in [?]. We stopped all algorithms when the maximum number of iterations exceeds 20000 or $\|\nabla f(x_k)\|_\infty \leq 10^{-6}$.

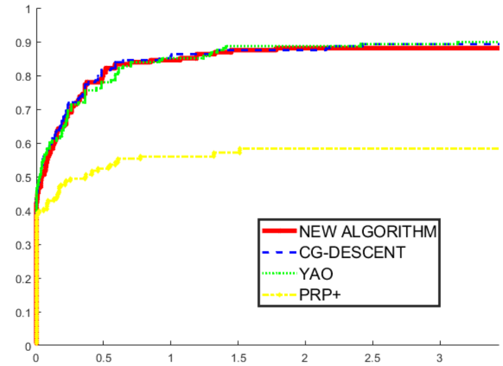
We use the performance profile by Dolan and Moré [?] to evaluate the numerical results. The related data can be downloaded from the website: https://github.com/feizaine/data. Figure 1 shows the comparison of the CPU time, the number of iterations, the number of function evaluations and the number of gradient evaluations of the four algorithms, respectively. The proposed algorithm solves 89% of the problem and it is significantly better than the PRP+ method and is very competitive with the CG_DESCENT method and Yao method, which shows that the proposed algorithm is processing robust to high-dimensional problems.

5. Conclusion

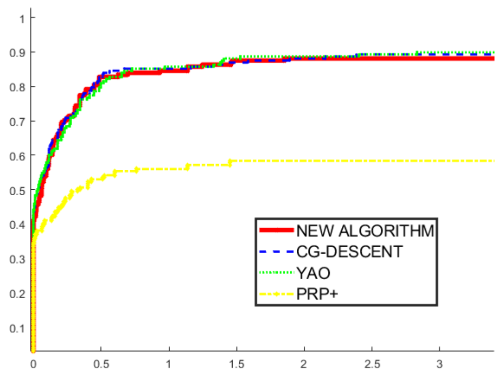
In this paper, we have proposed a modified HS method by minimizing the distance between the CG direction and the direction of the improved Perry conjugate gradient method [?]. A remarkble property of the new method is that it can always produce the descent direction which is independent of the line search. Under suitable conditions, we prove that the modified HS method is globally convergent with the Armijo line search for uniformly convex functions, and for the general functions, the convergence can also be guaranteed with the standard Wolfe line search. Extensive numerical experiments show that the proposed method is efficient.



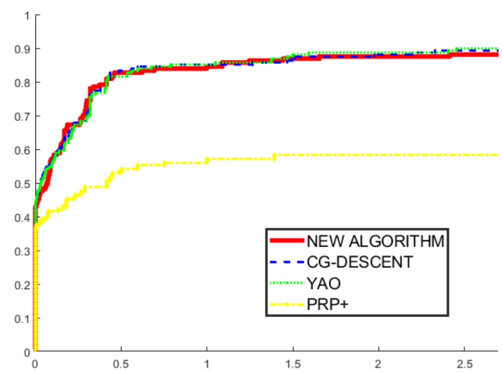
(a) cpu time



(b) iterations



(c) function evaluations



(d) gradient evaluations

Figure 1: *The performance profile on the CPU time (a), the number of iterations(b), the number of function evaluations(c) and the number of gradient evaluations(d).*

References

1. N. Andrei, An unconstrained optimization test functions collection, Adv. Model. Optim, 10(2008), pp. 147-161.

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

3. Y.H. Dai and L.Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim., 43(2001), pp. 87-101.

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), pp. 296-230.

5. E.D. Dolan and J.J. Moré, Benchmarking optimization software with performance profiles, Math. Program., 91 (2002), pp. 201-213.

6. R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients, Comput. J., 7 (1964), pp. 149-154.

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

8. N.I.M. Gould, D. Orban, and Ph.L. Toint, CUTEr: A constrained and unconstrained testing environment, ACM Trans. Math. Softw., 29(4) (2003), pp. 373-394.

9. M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Stand, 49 (1952), pp. 409-436.

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

11. W.W. Hager, H.C. Zhang, A survey of nonlinear conjugate gradient methods. Pacific J. Optim., 2(2006), pp. 35-58.

12. Z.X. Liu, H.W. Liu and Y.H. Dai, An improved Dai-Kou conjugate gradient algorithm for unconstrained optimization, Comput. Optim. Appl., 75 (2020), pp. 145-167.

13. M.J.D. Powell, Non convex minimization calculations and the conjugate gradient method, in D.F. Griffiths, ed.,Numerical Analysis, Lecture Notes in Mathematics 1066, Springer, Berlin, 1984, pp. 122-141.

14. E. Polak and G. Ribiére, Note sur la convergence de methods de directions conjugees, Rev. Franc Inform Recher Opiération, 16 (1969), pp. 35-43.

15. B.T. Polyak, The conjugate gradient method in extremal problems, U.S.S.R Comput. Math. Math. Phys., 9 (1969), pp. 94-112.

16. S.W. Yao, D.L. He and L.H. Shi, An improved Perry conjugate gradient method with adaptive parameter choice, Numer Algor., 78(2018), pp. 1255-1269.

17. G.L. Yuan, X.L. Wang and Z. Sheng, Family weak conjugate gradient algorithms and their convergence analysis for nonconvex functions, Numer. Algorithms., 84 (2020), pp. 935-956.

18. G.L. Yuan, J.G Lu and Z. Wang, The modified PRP conjugate gradient algorithm under a non-descent line search and its application in the Muskingum model and image restoration problems, Soft. Compu., 25 (2021), pp. 5867-5879.