A PERRY DESCENT CONJUGATE GRADIENT METHOD WITH RESTRICTED SPECTRUM

DONGYI LIU AND GENQI XU

Abstract. A new nonlinear conjugate gradient method, based on Perry's idea, is presented. And it is shown that its sufficient descent property is independent of any line search and the eigenvalues of $P_{k+1}^{\mathrm{T}}P_{k+1}$ are bounded above, where P_{k+1} is the iteration matrix of the new method. Thus, the global convergence is proven by the spectral analysis for non-convex functions when the line search fulfills the Wolfe conditions. Preliminary numerical results for a set of 720 unconstrained optimization test problems verify the performance of the new method and show that it is competitive with the CG_DESCENT method.

1. Introduction and New Method

The nonlinear conjugate gradient method is a well-known method for large-scale unconstrained optimization problems

$$\min_{x\in R^n} f(x), \tag{1}$$

which can be tracked to the HS method ([14], Hestenes et al. 1952). Since then, many variants of the HS method have been developed, and some of them are widely used in practice, such as FR ([9], Fletcher et al.1964), PRP ([18], Polak et al., 1969, [19], Polyak, 1969), CG$^+$ ([10], Gilbert et al., 1992), DY ([7], Dai et al., 1999), SCG ([4], Birgin et al., 2001), CG_DESCENT (Hager et al, [11], 2005, [12], 2006), SCALCG ([2], Andrei, 2007) methods, and so on. On the conjugate gradient method, there are several survey articles, such as a recent one, [13] (Hager et al. 2006).

The classical conjugate gradient method with line searches is formulated by

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

and

$$d_1 = -g_1,\ d_{k+1} = -g_{k+1} + \beta_k d_k, \qquad \forall k \geq 1, \tag{3}$$

Key words and phrases. large scale optimization, conjugate gradient method, descent property, spectral analysis, global convergence.

The work of the authors research is supported by the Natural Science Foundation of China grant NSFC-60874034 and by the Seed Foundation of Tianjin University.

where $g_k = g(x_k) = \nabla f(x_k)$, the gradient at x_k.

In 1978, Perry ([17]) rewrote the line search direction of the HS conjugate gradient method as follows:

$$d_{k+1} = -D_{k+1}g_{k+1} \text{ with } D_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right), \tag{4}$$

where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$. Then, he added a rank one matrix to D_{k+1} such that

$$y_k^{\mathrm{T}} d_{k+1} = -s_k^{\mathrm{T}} g_{k+1}, \tag{5}$$

thus he deduced that

$$d_{k+1} = -P_{k+1}g_{k+1} = -g_{k+1} + \beta_k^P d_k,$$

where $\beta_k^P = \dfrac{(y_k - s_k)^{\mathrm{T}} g_{k+1}}{y_k^{\mathrm{T}} d_k}$ and $P_{k+1} = I - \dfrac{s_k (y_k - s_k)^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k}$.

In [16], the conjugacy condition (5) was replaced by

$$y_k^{\mathrm{T}} d_{k+1} = -\sigma s_k^{\mathrm{T}} g_{k+1}, \tag{6}$$

where the parameter σ may vary with the iterative computation, and the matrix P_{k+1} was reformulated by

$$P_{k+1} = D_{k+1} + u_k v_k^{\mathrm{T}}$$

where $u_k, v_k \in R^n$. Then, from the definition of D_{k+1} in (4), (6) and $d_{k+1} = -P_{k+1}g_{k+1}$, it followed that

$$(\sigma s_k - v_k y_k^{\mathrm{T}} u_k)^{\mathrm{T}} g_{k+1} = 0 \text{ and } v_k = \sigma s_k / (y_k^{\mathrm{T}} u_k).$$

Thus, the matrix P_{k+1} was written by

$$P_{k+1} = D_{k+1} + \sigma \frac{u_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} u_k} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \sigma \frac{u_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} u_k}\right), \tag{7}$$

where u_k is any vector in R^n such that $y_k^T u_k \neq 0$. So, a family of generalized Perry conjugate gradient method were presented in [16], whose line search directions were defined by

$$\begin{cases} d_1 = -g_1, \\ d_{k+1} = -P_{k+1}g_{k+1} = -g_{k+1} + \dfrac{y_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} y_k} s_k - \sigma \dfrac{s_k^{\mathrm{T}} g_{k+1}}{y_k^{\mathrm{T}} u_k} u_k, \forall k \geq 1. \end{cases} \tag{8}$$

The matrix P_{k+1} defined by (7) is called the *iteration matrix of the generalized Perry conjugate gradient method*, which satisfies that $P_{k+1}^{\mathrm{T}} y_k = \sigma s_k$.

In this paper, we let $u_k = s_k$, then (8) yields to

$$d_1 = -g_1, d_{k+1} = -g_{k+1} + \beta_k^{GP} d_k, k \geq 1 \tag{9}$$

with

$$\beta_k^{GP} = \frac{(y_k - \sigma s_k)^{\mathrm{T}} g_{k+1}}{d_k^{\mathrm{T}} y_k}. \tag{10}$$

In next section, it will be shown that the directions d_{k+1} calculated by (9) and (10) satisfy the sufficient descent property ([10], Gilbert et al., 1992)

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -c_0 \|g_{k+1}\|^2 \text{ with } c_0 > 0, \tag{11}$$

for any line search, under the following condition

$$\text{either } \sigma \geq \frac{a\|y_k\|^2}{2(s_k^{\mathrm{T}} y_k)} \text{ if } s_k^{\mathrm{T}} y_k > 0, \text{ or } \sigma \leq \frac{a\|y_k\|^2}{2(s_k^{\mathrm{T}} y_k)} \text{ if } s_k^{\mathrm{T}} y_k < 0, \tag{12}$$

where $a > 1/2$. So the scheme (2), (9) and (10) with (12) is called the *Perry descent conjugate gradient algorithm*, denoted by PDCGs, and the corresponding iteration matrix is formulated by

$$P_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \sigma \frac{s_k s_k^{\mathrm{T}}}{y_k^{\mathrm{T}} s_k} \right) = \left(I - \frac{s_k (y_k - \sigma s_k)^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} \right). \tag{13}$$

Hence, it is reasonable choice that the value of the parameter σ varies with the iterative process and satisfies the condition (12), that is, the conjugate condition (6) is an self-adaptive version.

A similar property of β_k^{GP} pointed out in [6] is that it is also the solution of the following one-parameter quadratic model on β:

$$\min_{\beta} g_{k+1}^{\mathrm{T}} d(\beta) + \frac{1}{2} d(\beta)^{\mathrm{T}} H_{k+1} d(\beta)$$

where $d(\beta) = -g_{k+1} + \beta d_k$, the symmetrical and positive definite matrix H_{k+1} is an approximation of the Hessen matrix $\nabla^2 f(x_{k+1})$ such that $H_{k+1} s_k = \sigma^{-1} y_k$ with $\sigma \neq 0$ (general quasi-Newton equation). That is to say, the solution of the symmetrical linear system

$$H_{k+1} d(\beta) = -g_{k+1}$$

can be formulated by $d(\beta) = -P_{k+1} g_{k+1}$, where P_{k+1} defined by (13) is not a symmetrical matrix. It is a remarkable phenomenon.

In [11]–[13], Hager and Zhang chose $\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}} y_k)$ with $c > 1/4$, and presented the CG_DESCENT algorithm by restricting β_k and taking $c = 2$. So, we call the algorithm determined by the scheme (2) and

$$d_1 = -g_1, d_{k+1} = -g_{k+1} + \beta_k^{PHZ} d_k, k \geq 1 \tag{14}$$

where

$$\beta_k^{PHZ} = \frac{1}{d_k^{\mathrm{T}} y_k} \left(y_k - \frac{c\|y_k\|^2}{d_k^{\mathrm{T}} y_k} d_k \right)^{\mathrm{T}} g_{k+1}, \tag{15}$$

the *Perry-Hager-Zhang conjugate gradient algorithm*, denoted by PHZCG. Of course, there are also other choices for σ. If $\sigma = \dfrac{y_k^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} g_{k+1}}$, then $\beta_k^{GP} = 0$, from which the famous steepest method can be obtained. When $\sigma = 0$, the HS method can be deduced from (10). In [6], Dai and Liao suggested that σ is a bounded positive constant, especially, $\sigma = 0.1$ in the resulting method (the D-L method).

In [6] and [11], the authors proved the global convergence of their algorithms for general nonconvex functions using the similar approach taken in [10]. In this paper, we use a different method to establish the global convergence of the PDCGs algorithm for nonconvex functions, that is, we restrict the maximum eigenvalue of $P_{k+1}^{\mathrm{T}} P_{k+1}$. For this, we first restrict the value of $y_k^{\mathrm{T}} s_k$ as follows:

$$\eta_k^s = \begin{cases} y_k^{\mathrm{T}} s_k, & \text{if } \|g_k\|^2 \geq \eta \alpha_k \|d_k\|^2, \\ \|s_k\|^2, & \text{otherwise}, \end{cases} \tag{16}$$

where $\eta > 0$, then we let

$$\widetilde{\beta}_k^{GP} = \frac{1}{\eta_k^s} \left(y_k - \frac{c\|y_k\|^2}{\eta_k^s} s_k \right)^{\mathrm{T}} g_{k+1}, c > \frac{1}{4}. \tag{17}$$

Thus, the new line search directions d_k are generated by the rule

$$d_1 = -g_1, d_{k+1} = -g_{k+1} + \widetilde{\beta}_k^{GP} s_k, k \geq 1, \tag{18}$$

and the corresponding iteration matrix is

$$P_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{\eta_k^s} + \frac{c\|y_k\|^2 s_k s_k^{\mathrm{T}}}{(\eta_k^s)^2} \right). \tag{19}$$

In Section 3, Theorem 3.5 shows that (16) makes the maximum eigenvalue of $P_{k+1}^{\mathrm{T}} P_{k+1}$ bounded above. So the algorithm determined by (2), (17) and (18) is called the *Perry descent conjugate gradient algorithm with the restricted spectrum*, denoted by RSPDCGs. In this paper, we suggest $\eta = 0.001$ in (16) and $c = 1$ in (17).

The paper is organized as follows. In Section 2 we show that the directions generated by (17) and (18) satisfy the sufficient descent property, and analyze the spectra of $P_{k+1}^{\mathrm{T}} P_{k+1}$. In Section 3, by estimating the spectral upper bound of $P_{k+1}^{\mathrm{T}} P_{k+1}$, we prove the global convergence of these algorithms: PDCGs, PHZCG and RSPDCGs, under the Wolfe line searches. So, this proof method is called the spectral method. In Section 4 we compare the Dolan and Moré ([8], 2002) performance profile of the RSPDCGs algorithm with the profiles for the PHZCG algorithm and the CG_DESCENT algorithm. Finally, in Section 5 we conclude this paper.

2. Descent Property and Spectral Analysis

The descent property ([1], Al-Baali, 1985) and the sufficient descent property are important conditions for the global convergence, so we first consider the descent property.

Proposition 2.1. *Let $a > 1/2$ and $s_k^{\mathrm{T}} y_k \neq 0, \forall k \geq 1$. For the line search directions defined by*

$$d_1 = -g_1, d_{k+1} = -g_{k+1} + \frac{(y_k - \sigma s_k)^{\mathrm{T}} g_{k+1}}{\eta_k^s} s_k, k \geq 1, \tag{20}$$

where $\eta_k^s \neq 0$ and η_k^s is defined by (16) *with $\eta \geq 0$, if σ satisfies either*

$$\sigma \geq \frac{a\|y_k\|^2}{2(\eta_k^s)} \text{ when } \eta_k^s > 0, \text{ or } \sigma \leq \frac{a\|y_k\|^2}{2(\eta_k^s)} \text{ when } \eta_k^s < 0, \tag{21}$$

then

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -\left(1 - \frac{1}{2a}\right) \|g_{k+1}\|^2, k = 0, 1, 2, \dots . \tag{22}$$

Proof. The inequality (22) holds for $k = 0$, clearly. Now, we let $k \geq 1$. From the following inequality

$$u^{\mathrm{T}} v \leq \frac{1}{2}(\gamma\|u\|^2 + \gamma^{-1}\|v\|^2), \forall \gamma > 0 \text{ and } u, v \in R^n,$$

it can be derived that

$$((g_{k+1}^{\mathrm{T}} s_k) y_k)^{\mathrm{T}} (\eta_k^s g_{k+1}) \leq \frac{1}{2}\left(a(g_{k+1}^{\mathrm{T}} s_k)^2 \|y_k\|^2 + \frac{1}{a}(\eta_k^s)^2 \|g_{k+1}\|^2 \right). \tag{23}$$

So, it follows from (20) and (23) that

$$\begin{aligned} d_{k+1}^{\mathrm{T}} g_{k+1} &= -\|g_{k+1}\|^2 - \sigma \frac{\left(s_k^{\mathrm{T}} g_{k+1}\right)^2}{\eta_k^s} + g_{k+1}^{\mathrm{T}} s_k \frac{y_k^{\mathrm{T}} g_{k+1}}{(\eta_k^s)^2} \eta_k^s \\ &\leq -\|g_{k+1}\|^2 - \sigma \eta_k^s \frac{\left(s_k^{\mathrm{T}} g_{k+1}\right)^2}{(\eta_k^s)^2} + \frac{1}{2} \frac{(a(g_{k+1}^{\mathrm{T}} s_k)^2 \|y_k\|^2 + a^{-1}(\eta_k^s)^2 \|g_{k+1}\|^2)}{(\eta_k^s)^2} \\ &\leq -\left(1 - \frac{1}{2a}\right) \|g_{k+1}\|^2 + \left(\frac{a\|y_k\|^2}{2} - \sigma \eta_k^s \right) \frac{(s_k^{\mathrm{T}} g_{k+1})^2}{(\eta_k^s)^2}. \end{aligned}$$

Therefore, when $1 - \frac{1}{2a} > 0$ and $\frac{a\|y_k\|^2}{2} - \sigma \eta_k^s \leq 0$, that is, when (21) is fulfilled, the sufficient descent property (22) holds. □

Remark 2.2. *When $\eta = 0$ in* (16)*, $\eta_k^s \equiv s_k^{\mathrm{T}} y_k$ for all $k \geq 1$, so, it is proven that the PDCGs algorithm satisfies* (22)*. Let $a = 2c$ and $c > 1/4$, thus for the PHZCG algorithm with $\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}} y_k)$, the following inequality*

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -\left(1 - \frac{1}{4c}\right)\|g_{k+1}\|^2 \tag{24}$$

holds. In fact, for the PHZCG algorithm, $\eta_k^s \equiv s_k^{\mathrm{T}} y_k$ for all $k \geq 1$ in (21)*, thus, when $s_k^{\mathrm{T}} y_k \neq 0$, $\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}} y_k)$ satisfies the condition* (21)*. If $\sigma = c\|y_k\|^2/\eta_k^s$ in* (20)*, then the directions defined by* (20) *is the restricted forms* (17)–(18)*, so the RSPDCGs algorithm satisfies the sufficient descent property* (24) *with $c > 1/4$.*

In what follows, we analyze the property of the iteration matrix of the PDCGs algorithm, P_{k+1}.

Proposition 2.3. *Let P_{k+1} be defined by (13) and $\sigma s_k^{\mathrm{T}} y_k \neq 0$. Then, the maximum and minimum eigenvalues of $P_{k+1}^{\mathrm{T}} P_{k+1}$, $\mu_k^{\max}$ and $\mu_k^{\min}$, are formulated by*

$$\mu_k^{\max} = \frac{1}{2}\left[\omega_k + a_k + \sqrt{(\omega_k + a_k)^2 - 4a_k}\right] \tag{25}$$

and

$$\mu_k^{\min} = \frac{1}{2}\left[\omega_k + a_k - \sqrt{(\omega_k + a_k)^2 - 4a_k}\right], \tag{26}$$

respectively, where $\omega_k = \dfrac{y_k^{\mathrm{T}} y_k s_k^{\mathrm{T}} s_k}{(s_k^{\mathrm{T}} y_k)^2} \geq 1$ and $a_k = \left(\sigma \dfrac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\right)^2$. Moreover, the 2-norm condition number of P_{k+1}, $\kappa_2(P_{k+1}) = \sqrt{\mu_k^{\max}/\mu_k^{\min}}$, reaches the minimum $\sqrt{\omega_k} + \sqrt{\omega_k - 1}$ when $\sigma = \pm\|y_k\|/\|s_k\|$.

Proof. Firstly, according to the fundamental algebra formula

$$\det(I + xy^{\mathrm{T}}) = 1 + y^{\mathrm{T}} x, \tag{27}$$

where $x, y \in R^n$ and I is the $n \times n$ identity matrix, it can be derived that

$$\det(P_{k+1}) = \det\left(I - \frac{s_k(y_k - \sigma s_k)^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) = \sigma \frac{s_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k}. \tag{28}$$

Next, it follows from (13) that

$$\begin{aligned} P_{k+1}^{\mathrm{T}} P_{k+1} &= I - \left(1 + \sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\right) \frac{y_k s_k^{\mathrm{T}} + s_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} \\ &\quad + \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\left(\frac{y_k y_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} + \frac{\sigma^2 s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}\right) + 2\sigma \frac{s_k s_k^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k}. \end{aligned}$$

So, the real number 1 is an eigenvalue of $P_{k+1}^{\mathrm{T}}P_{k+1}$ with the geometric multiplicity $n-2$, because $P_{k+1}^{\mathrm{T}}P_{k+1}\xi = \xi, \forall \xi \in \operatorname{span}\{y_k, s_k\}^{\perp}$. The other two eigenvalues are denoted by μ_1, μ_2. In addition, the trace of matrix $P_{k+1}^{\mathrm{T}}P_{k+1}$

$$\begin{aligned} Tr(P_{k+1}^{\mathrm{T}}P_{k+1}) &= n - 2\left(1 + \sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right) + \frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\left(\frac{y_k^{\mathrm{T}}y_k}{s_k^{\mathrm{T}}y_k} + \frac{\sigma^2 s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k}\right) + 2\sigma\frac{s_k^{\mathrm{T}}s_k}{s_k^{\mathrm{T}}y_k} \\ &= n - 2 + \omega_k + a_k. \end{aligned}$$

Thus, by the relationship between the trace and the eigenvalues of matrix and (28), it can be deduced that

$$\mu_1 + \mu_2 = a_k + \omega_k \text{ and } \mu_1\mu_2 = a_k. \tag{29}$$

Therefore, μ_1 and μ_2 are two solutions of the following equation

$$\lambda^2 - (a_k + \omega_k)\lambda + a_k = 0. \tag{30}$$

Obviously, (30) claim that (25) and (26) hold, so

$$\kappa_2(P_{k+1}) = \frac{\left[\omega_k + a_k + \sqrt{(\omega_k + a_k)^2 - 4a_k}\right]}{2\sqrt{a_k}}. \tag{31}$$

Finally, let $t = (\omega_k + a_k)/\sqrt{4a_k}$, then $t \geq \sqrt{\omega_k}$ and

$$\kappa_2(P_{k+1}) = \psi(t) = t + \sqrt{t^2 - 1},$$

where $\psi(t)$ is a strictly increasing function on $[1, +\infty)$, so $\psi(t) \geq \psi(\sqrt{\omega_k})$ when $t \geq \sqrt{\omega_k}$. Thus, $\psi(t)$ arrives at the minimum, $\sqrt{\omega_k} + \sqrt{\omega_k - 1}$, when $t = \sqrt{\omega_k}$. Note that $t = \sqrt{\omega_k}$ if and only if $\omega_k = a_k$. Hence, $\kappa_2(P_{k+1})$ attains the minimum, $\sqrt{\omega_k} + \sqrt{\omega_k - 1}$ when $\omega_k = a_k$, i.e., $\sigma = \pm\|y_k\|/\|s_k\|$. □

According to Proposition 2.1 and 2.3, when $\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}}y_k)$ and $c = 1/\sqrt{\omega_k}$, $\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}}y_k) = \pm\|y_k\|/\|s_k\|$, the condition number of P_{k+1} defined by (13) arrives at the minimum. So, for the PDCGs algorithm, a suitable choice of parameter σ is

$$\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}}y_k) \text{ and } c = \max\{c_\beta, 1/\sqrt{\omega_k}\}, c_\beta > 1/4. \tag{32}$$

This choice makes the condition number of P_{k+1} defined by (13) approach its minimum and the sufficient descent property (24) true.

3. Global Convergence

In this section, we analyze the global convergence of these algorithms: PDCGs, PHZCG and RSPDCGs by the spectral method. For this, we first introduction the following assumptions about the objective function $f(x)$, which are often used in convergent analysis.

H1: f is bounded below in R^n and continuously differentiable in a neighborhood $\mathcal{N}$ of the level set $\mathcal{L} = \{x : f(x) \leq f(x_1)\}$, where x_1 is the starting point of the iteration.

H2: The gradient of f is Lipschitz continuous in $\mathcal{N}$, that is, there exists a constant $L > 0$ such that

$$\|\nabla f(\bar{x}) - \nabla f(x)\| \leq L\|\bar{x} - x\|, \quad \forall \bar{x}, x \in \mathcal{N}.$$

H3: f is a uniformly convex function on $\mathcal{N}$, i.e., there exists a constant $m > 0$ such that

$$(\nabla f(\bar{x}) - \nabla f(x))^{\mathrm{T}}(\bar{x} - x) \geq m\|\bar{x} - x\|^2 \quad \forall \bar{x}, x \in \mathcal{N}.$$

Second, we assume that the line search strategy satisfies the Wolfe conditions:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + c_1 \alpha_k d_k^{\mathrm{T}} g_k \tag{33}$$

and

$$d_k^{\mathrm{T}} g(x_k + \alpha_k d_k) \geq c_2 d_k^{\mathrm{T}} g_k, \tag{34}$$

where $0 < c_1 < c_2 < 1$.

Next, we introduce the spectral condition theorem of the global convergence for objective functions satisfying H1 and H2, which generalizes Theorem 4.1 in [15] (Liu et al., 2009).

Theorem 3.1. *Assume that H1 and H2 are fulfilled. For a nonlinear conjugate gradient method, its iterative sequence is generated by (2) and its line search directions are calculated by*

$$d_1 = -g_1, \; d_k = -M_k g_k, \quad \forall k > 1, \tag{35}$$

where M_k is the conjugate gradient iteration matrix and the maximum eigenvalue of $M_k^{\mathrm{T}} M_k$ is denoted by Λ_k. If the following conditions hold:

(a): *the sufficient descent property (11) holds,*

(b): *the Wolfe line searches (33) and (34) are implemented, and*

(c): $\sum_{k=1}^{\infty} (\Lambda_k)^{-1} = +\infty$ *(the spectral condition),*

then either $g_k = 0$ for some $k > 1$, or

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{36}$$

Moreover, if the condition (c) is replaced by the following condition

(c)': $\Lambda_k \leq \widetilde{\Lambda}$ *for large enough k, where $\widetilde{\Lambda}$ is a positive constant,*

then

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{37}$$

Proof. Assume that $g_k \neq 0, \forall k \geq 1$ and $\liminf_{k\to\infty} \|g_k\| \neq 0$, then there exists $\gamma > 0$ such that $\|g_k\| > \gamma, \forall k \geq 1$, and the sufficient descent condition (11) implies that $d_k \neq 0, \forall k \geq 1$. From (35) and the fact that $M_k^{\mathrm{T}} M_k$ is symmetric and positive semi-definite, it follows that

$$\|d_k\|^2 = g_k^{\mathrm{T}} M_k^{\mathrm{T}} M_k g_k \leq \Lambda_k \|g_k\|^2. \tag{38}$$

Thus, from (11) and the above inequality, it can be deduced that

$$\cos^2\theta_k = \frac{(-d_k^{\mathrm{T}} g_k)^2}{\|d_k\|^2\|g_k\|^2} \geq c_0^2 \frac{\|g_k\|^2}{\|d_k\|^2} \geq \frac{c_0^2}{\Lambda_k}, \tag{39}$$

where θ_k is the angle between d_k and $-g_k$. Thus,

$$\sum_{k\geq 0} \|g_k\|^2 \cos^2\theta_k \geq \gamma^2 \sum_{k=0}^{\infty} \frac{c_0^2}{\Lambda_k} = \infty,$$

which contradicts to the Zoutendijk's condition ([20], Zoutendijk, 1970),

$$\sum_{k\geq 0} \|g_k\|^2 \cos^2\theta_k = \sum_{k=0}^{\infty} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} < \infty. \tag{40}$$

Therefore, (c) implies that either $g_k = 0$ for some $k > 1$, or (36) holds.

If $\Lambda_k \leq \widetilde{\Lambda}$ for large enough k, then $\cos^2\theta_k \geq c_0^2/\widetilde{\Lambda} > 0$, which, together with (40), implies (37). □

Now, we discuss the global convergence of these algorithms.

3.1. **The convergence for uniformly convex functions.**

Theorem 3.2. *Assume that H1, H2 and H3 hold, and that σ satisfies (12) and $|\sigma| \leq \sigma_0$, where σ_0 is a positive constant. If the PDCGs algorithm implements the Wolfe line searches (33) and (34), then either $g_k = 0$ for some $k > 1$, or $\lim_{k\to\infty} \|g_k\| = 0$.*

Proof. For the PDCGs algorithm, the direction d_{k+1} is calculated by (9) and (10), and σ satisfies (12). It follows from Remark 2.2 of Proposition 2.1 that the algorithm PDCGs satisfies the descent property (11) with $c_0 = (2a-1)/(2a)$ and $a > 1/2$.

By Proposition 2.3, $\mu_k^{\max}$, the maximum eigenvalue of $P_{k+1}^{\mathrm{T}} P_{k+1}$, satisfies

$$\mu_k^{\max} \leq \frac{y_k^{\mathrm{T}} y_k s_k^{\mathrm{T}} s_k}{(s_k^{\mathrm{T}} y_k)^2} + \left(\sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\right)^2.$$

So, it can be deduced from assumptions of the theorem that

$$\mu_k^{\max} \leq \frac{L^2 s_k^{\mathrm{T}} s_k s_k^{\mathrm{T}} s_k}{(m s_k^{\mathrm{T}} s_k)^2} + \left(\sigma \frac{s_k^{\mathrm{T}} s_k}{m s_k^{\mathrm{T}} s_k}\right)^2 \leq \frac{L^2 + \sigma_0^2}{m^2}. \tag{41}$$

Thus, Theorem 3.1 claims that either $g_k = 0$ for some $k > 1$, or $\lim_{k\to\infty} \|g_k\| = 0$. □

Remark 3.3. *For the PHZCG algorithm, the direction d_{k+1} is defined by (9) and* (15)*, thus, the inequality (24) in Remark 2.2 shows that the descent property (11) holds, and $\sigma = c\|y_k\|^2/s_k^{\mathrm{T}} y_k$ implies that* (41) *holds with $\sigma_0 = m^{-1}cL^2$. So, the Perry-Hager-Zhang conjugate gradient algorithm is global convergence for uniformly convex functions.*

3.2. **The convergence for nonlinear nonconvex functions.** To prove the global convergence of the RSPDCGs algorithm, we need the following lemma.

Lemma 3.4. *Let P_{k+1} be defined by*

$$P_{k+1} = \left(I - \frac{s_k y_k^{\mathrm{T}}}{\eta_k^s} + \sigma \frac{s_k s_k^{\mathrm{T}}}{\eta_k^s}\right) = \left(I - \frac{s_k (y_k - \sigma s_k)^{\mathrm{T}}}{\eta_k^s}\right), \tag{42}$$

where η_k^s is defined by (16) *and $\eta > 0$. Denote the maximum and the minimum eigenvalues of $P_{k+1}^{\mathrm{T}} P_{k+1}$ by $\mu_k^{\max}$ and $\mu_k^{\min}$, respectively. If $\|y_k\| \le L\|s_k\|$ and $|\sigma| \le \sigma_0$, where L and σ_0 are two positive constants, then*

$$0 \le \mu_k^{\min} \le \mu_k^{\max} \le \frac{L^2 + \sigma_0^2}{\widetilde{\eta}^2}, \text{ when } |y_k^{\mathrm{T}} s_k| \ge \widetilde{\eta}\|s_k\|^2 \text{ and } \eta_k^s = s_k^{\mathrm{T}} y_k, \tag{43}$$

where $\widetilde{\eta} > 0$, and

$$0 \le \mu_k^{\min} \le \mu_k^{\max} \le 1 + (1 + L + \sigma_0)^2, \text{ when } \eta_k^s = \|s_k\|^2. \tag{44}$$

Proof. When $|y_k^{\mathrm{T}} s_k| \ge \widetilde{\eta}\|s_k\|^2$ and $\eta_k^s = s_k^{\mathrm{T}} y_k$, then P_{k+1} defined by (42) becomes the matrix P_{k+1} defined by (13). (25) in Proposition 2.3 implies that

$$\mu_k^{\max} \le \omega_k + \left(\sigma \frac{s_k^{\mathrm{T}} s_k}{s_k^{\mathrm{T}} y_k}\right)^2 \le \frac{y_k^{\mathrm{T}} y_k s_k^{\mathrm{T}} s_k}{\widetilde{\eta}^2 \|s_k\|^4} + \left(\sigma \frac{s_k^{\mathrm{T}} s_k}{\widetilde{\eta}\|s_k\|^2}\right)^2 \le \frac{L^2 + \sigma_0^2}{\widetilde{\eta}^2}.$$

Thus, Proposition 2.3 shows that these inequalities in (43) hold.

Next, we prove the inequalities in (44). It follows from (42) that

$$\begin{aligned} P_{k+1}^{\mathrm{T}} P_{k+1} &= I - \left(1 + \sigma \frac{s_k^{\mathrm{T}} s_k}{\eta_k^s}\right) \frac{y_k s_k^{\mathrm{T}} + s_k y_k^{\mathrm{T}}}{\eta_k^s} \\ &\quad + \frac{s_k^{\mathrm{T}} s_k y_k y_k^{\mathrm{T}}}{(\eta_k^s)^2} + \left(2\sigma + \frac{\sigma^2 s_k s_k^{\mathrm{T}}}{\eta_k^s}\right) \frac{s_k s_k^{\mathrm{T}}}{\eta_k^s}. \end{aligned}$$

Since $P_{k+1}^{\mathrm{T}} P_{k+1} \xi = \xi, \forall \xi \in \mathrm{span}\{y_k, s_k\}^{\perp}$, the real number 1 is an eigenvalue of $P_{k+1}^{\mathrm{T}} P_{k+1}$ with the geometric multiplicity $n-2$. The other two eigenvalues are denoted by μ_1, μ_2, and assume that $\mu_1 \le \mu_2$. It from (27) and (42) follows that

$$\det(P_{k+1}) = \det\left(I - \frac{s_k(y_k - \sigma s_k)^{\mathrm{T}}}{\eta_k^s}\right) = 1 - \frac{s_k^{\mathrm{T}} y_k}{\eta_k^s} + \sigma \frac{s_k^{\mathrm{T}} s_k}{\eta_k^s},$$

so, similar to (29) in Proposition 2.3, it can be derived that

$$\begin{cases} \mu_1 + \mu_2 = 2 - \left(1 + \sigma \dfrac{s_k^{\mathrm{T}} s_k}{\eta_k^s}\right) \dfrac{2 s_k^{\mathrm{T}} y_k}{\eta_k^s} + \dfrac{s_k^{\mathrm{T}} s_k}{\eta_k^s} \left(\dfrac{y_k^{\mathrm{T}} y_k}{\eta_k^s} + \dfrac{\sigma^2 s_k^{\mathrm{T}} s_k}{\eta_k^s}\right) + 2\sigma \dfrac{s_k^{\mathrm{T}} s_k}{\eta_k^s}, \\ \mu_1 \mu_2 = \left(1 - \dfrac{s_k^{\mathrm{T}} y_k}{\eta_k^s} + \sigma \dfrac{s_k^{\mathrm{T}} s_k}{\eta_k^s}\right)^2 \ge 0. \end{cases}$$

Thus, when $\eta_k^s = \|s_k\|^2$, it can be deduced that

$$\begin{aligned} \mu_1 + \mu_2 &= 1 - (1 + \sigma) \frac{2 s_k^{\mathrm{T}} y_k}{\|s_k\|^2} + \frac{y_k^{\mathrm{T}} y_k}{\|s_k\|^2} + (1 + \sigma)^2 \\ &\le 1 + 2L|1 + \sigma| + L^2 + (1 + \sigma)^2 \le 1 + (1 + L + \sigma_0)^2, \end{aligned}$$

which, together with the fact that μ_1 and μ_2 are nonnegative, $\mu_k^{\min} = \min\{1, \mu_1\}$ and $\mu_k^{\max} = \max\{1, \mu_2\}$, claims that these inequalities in (44) hold. □

Now, we prove the convergence of the RSPDCGs algorithm.

Theorem 3.5. *Assume that H1 and H2 hold. For the RSPDCGs algorithm with $c > 1/4$, if the line searches satisfy the Wolfe conditions* (33) *and* (34)*, then either $g_k = 0$ for some $k > 1$, or* $\lim_{k\to\infty} \|g_k\| = 0$.

Proof. We assume that $g_k \ne 0$ for all $k > 1$, and show that $\lim_{k\to\infty} \|g_k\| = 0$.

Denote the maximum eigenvalue of $P_{k+1}^{\mathrm{T}} P_{k+1}$ by $\mu_k^{\max}$, where P_{k+1} is the iteration matrix of the RSPDCGs algorithm defined by (16) and (19), also determined by (42) with $\sigma = \dfrac{c\|y_k\|^2}{\eta_k^s}$.

When $\|g_k\|^2 \ge \eta \alpha_k \|d_k\|^2$, then $\eta_k^s = y_k^{\mathrm{T}} s_k$, and it follows from (11) and (34) that

$$y_k^{\mathrm{T}} s_k \ge -\alpha_k (1 - c_2) d_k^{\mathrm{T}} g_k \ge \alpha_k c_0 (1 - c_2) \|g_k\|^2 \ge \eta c_0 (1 - c_2) \|s_k\|^2,$$

thus, from (16) and H2, it can be deduced that

$$|\sigma| = \frac{c\|y_k\|^2}{|\eta_k^s|} \le \frac{c\|y_k\|^2}{\eta c_0 (1 - c_2) \|s_k\|^2} \le \frac{cL^2}{\eta c_0 (1 - c_2)}.$$

And it is derived from (42) and (43) in Lemma 3.4 that

$$0 \le \mu_k^{\max} \le \frac{L^2 + \left(\dfrac{cL^2}{\eta c_0 (1 - c_2)}\right)^2}{\eta^2 c_0^2 (1 - c_2)^2}.$$

When $||g_k||^2 < \eta\alpha_k||d_k||^2$, then $\eta_k^s = ||s_k||^2$ and

$$|\sigma| = \frac{c||y_k||^2}{|\eta_k^s|} = \frac{c||y_k||^2}{||s_k||^2} \le cL^2.$$

So, (44) in Lemma 3.4 implies that

$$0 \le \mu_k^{\max} \le 1 + \left(1 + L + cL^2\right)^2.$$

So, we conclude from above two cases that the maximum eigenvalue of $P_{k+1}^{\mathrm{T}} P_{k+1}$ is bounded above. Thus, it follows from Theorem 3.1 that $\lim_{k\to\infty} ||g_k|| = 0$. □

4. Numerical Experiments

In this section, we demonstrate the algorithm: RSPDCGs and PHZCG. We use the notation PHZCG(c_β) to indicate the dependence of the PHZCG algorithm on the scalar c_β (see (32)). For example, PHZCG(0.5) shows that the parameter $c_\beta = 0.5$ in (32). The test functions are the 73 unconstrained functions but the 71-st, coded by N. Andrei, referring to website:

http://camo.ici.ro/forum/SCALCG/evalfg.for.

These functions come from the CUTE library [5] and other optimization functions, see [3] for the details. For every test function we have considered 10 numerical experiments with the number of variables $n = 1000, 2000, \ldots,$ 10000, thus there are 720 unconstrained optimization test problems. The starting points used are those given in the code, evalfg.for.

For comparison with the CG_DESCENT algorithm ([13], Hager et al.,2006), all codes are written in Fortran 77 according to CG_DESCENT codes, which can be obtained from Hager's web page at

http://www.math.ulf.edu/hager/papers/CG.

These tests use the approximate Wolfe line search and the default parameters in CG_DESCENT. The termination criterion of all algorithms is that $||g||_\infty < 10^{-6}$, where $||\cdot||_\infty$ is the infinity norm of a vector. The tests are performed on PC, Intel(R) Core(TM) 2 Duo, E4600 2.40GHz 2.39GHz, RAM 2.00 GB, using f77 compiler. The detailed numerical results are placed on the web site

http://cid-9553f8a2fd68847d.office.live.com/browse.aspx/cgm.

Figure 1 and Figure 2 present the Dolan and Moré performance profiles of the PHZCG algorithm relative to the number of iteration (Nite) and the metric $NF + 3NG$, respectively, where NF is the number of function evaluations and NG is the number of gradient evaluations. The metric $NF+3NG$ is suggested by Hager et al. ([13], 2006). E. D. Dolan and J. J. Moré use the term *performance profile* for the cumulative distribution function for

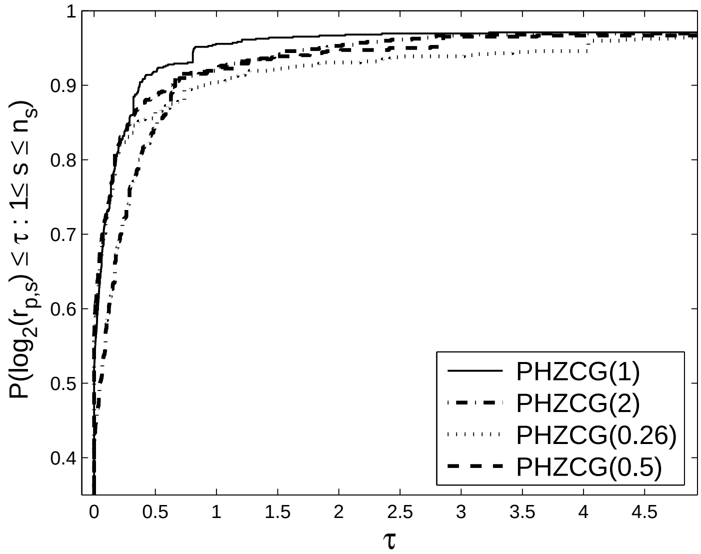

Figure 1. Performance based on Nite of the PHZCG algorithm for different c_β.

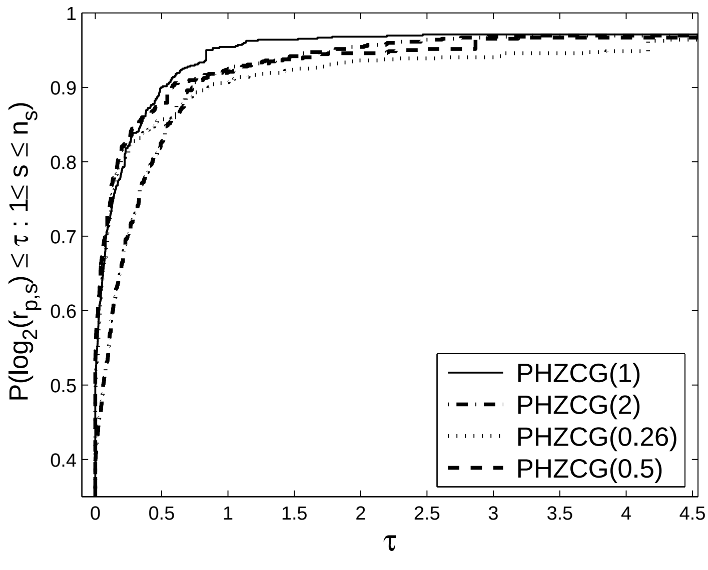

Figure 2. Performance based on $NF + 3NG$ of the PHZCG algorithm for different c_β.

the performance ratio that is within a factor $\tau \in R$ of the best possible ratio, where we use the logarithm to base 2 for the scale, see [8] for the details.

For the PHZCG algorithm, we also test other values of c_β for the set of 720 unconstained optimization test problems, and find the suitable choice of c in β_k^{PHZ} is that $c = \max\{c_\beta, 1/\sqrt{\omega_k}\}$ with $c_\beta \in [0.8, 1]$. Therefore, for the RSPDCGs algorithm, we suggest $c = 1$ in (17). We also test the value $c = 2$ (denoted by RSPDCGs(2)) to compare with the CG_DESCENT algorithm, thus, in what follows, we denote the RSPDCGs algorithm with $c = 1$ by RSPDCGs(1) to distinguish the one with $c = 2$. Figure 3 and Figure 4 present the performance profiles of the RSPDCGs and CG_DESCENT algorithms relative to Nite and $NF + 3NG$, respectively.

Figure 5 and Figure 6 present the performance profiles of the algorithms: PHZCG(1), RSPDCGs(1) and CG_DESCENT(1), relative to Nite and $NF + 3NG$, respectively. The scalar in parentheses of CG_DESCENT(1) means that the parameter $c = 1$ in β_k^{PHZ} defined by (15).

From the numerical results, we find that none of the three algorithms can solve all test problems, for example, the CG_DESCENT algorithm only solves the 59-th test function (DIXMAANH function) with the number of variable $n = 1000$, the RSPDCGs algorithm with $c = 1$ only solves it with the number of variables $n = 1000, 2000, \ldots, 6000$, 9000, 10000. Table 1 lists the numbers (Nfail) that these algorithms can not solve problems with a fixed the number of variables n, where Alg is the abbreviation for ‘algorithm’.

TABLE 1. The numbers of failure for different algorithms.

Alg	PHZCG(0.26)	PHZCG(0.5)	PHZCG(1)	PHZCG(2)
Nfail	26	24	21	22
Alg	CG_DESCENT(1)	CG_DESCENT	RSPDCGs(1)	RSPDCGs(2)
Nfail	22	22	13	19

The preliminary computational results show that the RSPDCGs algorithm outperforms the CG_DESCENT algorithm, which means that the restriction for β_k in (17) not only perfects the global convergence of the PHZCG algorithm for nonconvex functions, but also improves its performance.

The preliminary results also show that the choice of β_k, as follows:

$$\beta_k = \frac{1}{d_k^{\mathrm{T}} y_k}\left(y_k - \frac{\|y_k\|^2}{d_k^{\mathrm{T}} y_k} d_k\right)^{\mathrm{T}} g_{k+1} \tag{45}$$

is slightly better than

$$\beta_k = \frac{1}{d_k^{\mathrm{T}} y_k}\left(y_k - \frac{2\|y_k\|^2}{d_k^{\mathrm{T}} y_k} d_k\right)^{\mathrm{T}} g_{k+1} \tag{46}$$

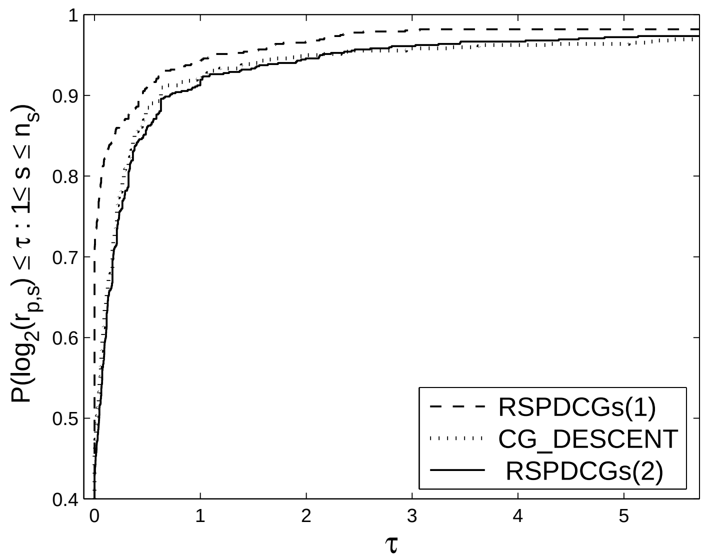

FIGURE 3. Performance based on Nite for RSPDCGs and CG_DESCENT algorithms.

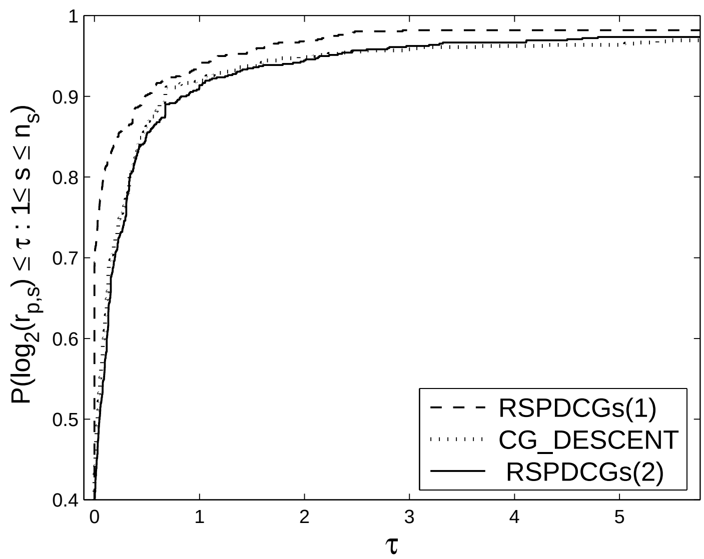

FIGURE 4. Performance based on $NF + 3NG$ for RSPDCGs and CG_DESCENT algorithms.

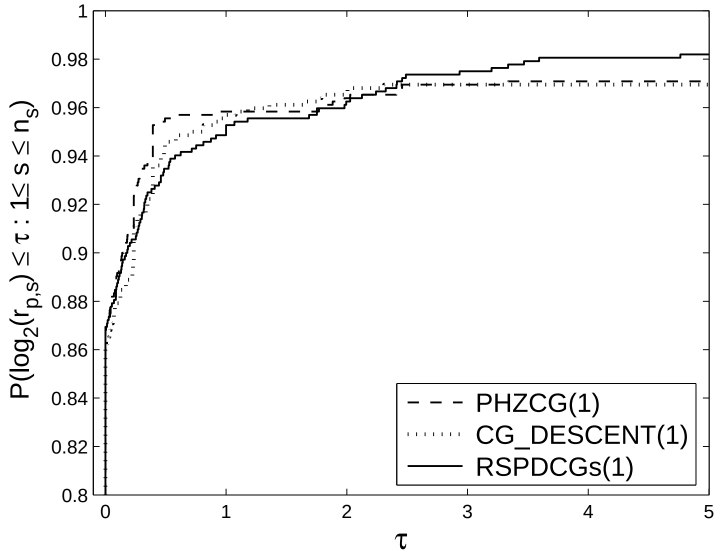

Figure 5. Performance based on Nite for PHZCG(1), RSPDCGs(1) and CG_DESCENT(1).

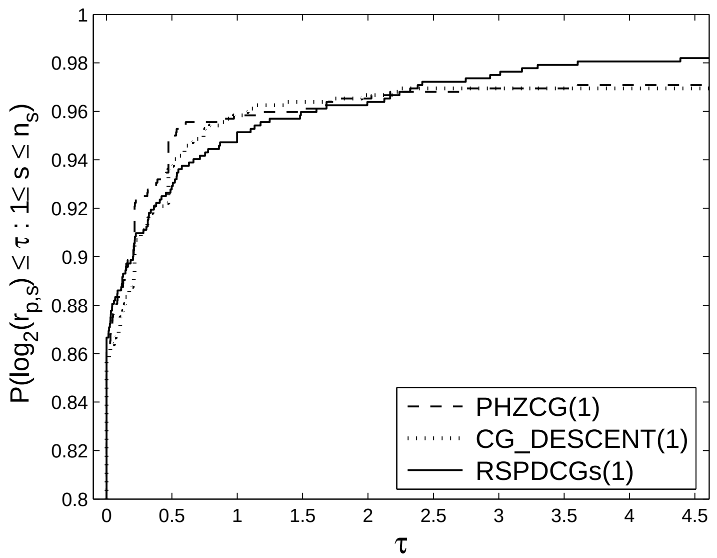

Figure 6. Performance based on $NF + 3NG$ for PHZCG(1), RSPDCGs(1) and CG_DESCENT(1).

for the test set consisting of 720 problems. In fact, when $\sigma = c\|y_k\|^2/(s_k^{\mathrm{T}} y_k)$, it can be obtained from (31) that

$$\kappa_2(P_{k+1}(c)) = \psi_{k+1}(c) = t(c) + \sqrt{(t(c))^2 - 1}$$

where

$$t(c) = \left(\omega_k + \left(\sigma \frac{s_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k}\right)^2\right) \Bigg/ \left|2\sigma \frac{s_k^{\mathrm{T}} s_k}{y_k^{\mathrm{T}} s_k}\right| = \frac{1 + c^2 \omega_k}{2c} \geq \sqrt{\omega_k}.$$

It can be verified easily that $t(c)$ and $\psi_{k+1}(c)$ are strictly increasing functions with respect to c on $[\sqrt{\omega_k}^{-1}, \infty)$, so, when $c_\beta \leq 1$ in (32),

$$\sqrt{\omega_k} = t\left(1/\sqrt{\omega_k}\right) \leq t\left(\max\{c_\beta, 1/\sqrt{\omega_k}\}\right) \leq t(1) < t(2)$$

and

$$\psi_{k+1}(\max\{c_\beta, 1/\sqrt{\omega_k}\}) \leq \psi_{k+1}(1) < \psi_{k+1}(2).$$

Therefore, the spectral condition number of the iteration matrix of the PHZCG algorithm as $c = 1$ is less than the one as $c = 2$, thus, the numerical stability of the PHZCG algorithm for (45) is stronger than that for (46). It may be the partial reason that the performance of (45) is slightly better than that of (46). Based on the same reason, the algorithm performance should be better as c_β is smaller, since $c = \max\{c_\beta, 1/\sqrt{\omega_k}\}$, but this is not the case (see Figure 1 and Figure 2). How the parameter c in (32) (or (15)) is chosen such that the algorithm performance is best? It is still an open problem.

5. Conclusions and Remarks

In this paper, we have presented a new Perry conjugate gradient method with the restricted spectrum and have proven its global convergence by the spectral method. The formula (8) also suggests new ideas and prototypes for constructing more effective nonlinear conjugate gradient algorithms, hence it is worthy of studying further, especially the choices of u_k and σ, such as $u_k = \nu_1 s_k + \nu_2 y_k$, the linear combination of s_k and y_k. In [16] we discuss the case that $u_k = y_k$.

The restriction of $y_k^{\mathrm{T}} s_k$ in (16), in fact, gives a new scaling strategy for the iteration matrix of the PDCGs algorithm, namely, when $\eta_k^s = s_k^{\mathrm{T}} s_k$, the matrix P_{k+1} defined by (13) is scaled by

$$P_{k+1} = I - \frac{s_k^{\mathrm{T}} y_k}{s_k^{\mathrm{T}} s_k} \frac{s_k (y_k - \sigma s_k)^{\mathrm{T}}}{s_k^{\mathrm{T}} y_k} = I - \frac{s_k (y_k - \sigma s_k)^{\mathrm{T}}}{s_k^{\mathrm{T}} s_k},$$

from which the following direction is deduced

$$d_{k+1} = -P_{k+1} g_{k+1} = -g_{k+1} + \frac{(y_k - \sigma s_k)^{\mathrm{T}} g_{k+1}}{s_k^{\mathrm{T}} s_k} s_k,$$

such that

$$d_{k+1}^{\mathrm{T}} g_{k+1} \leq -(1 - \frac{1}{4c})\|g_{k+1}\|^2,$$

when $\sigma = \frac{c\|y_k\|^2}{\|s_k\|^2}$ and $c > 1/4$, according to Proposition 2.1. So, it is also an interesting method.

Acknowledgements

The authors thank N. Andrei for the Fortran code, evalfg.for, W.W. Hager and H. Zhang for CG_DESCENT code and J. J. Moré for Matlab code, perf.m.

References

1. Al-Baali, M. (1985). Descent Property and Global Convergence of the Fletcher-Reeves Method with Inexact Line-Search, IMA Journal of Numerical Analysis, 5, 121–124.
2. Andrei, N. (2007). Scaled conjugate gradient algorithms for unconstrained optimization, Comput. Optim. Appl. 38, 401–416.
3. Andrei, N. (2008). An unconstrained optimization test functions collection, Advanced Modeling and Optimization, 10,147–161.
4. Birgin, E. G., & Martínez, J. M. (2001). A Spectral Conjugate Gradient Method for unconstrained optimization, Applied Mathematics and Optimization, 43, 117–128.
5. Bongartz I., Conn A. R., Gould N. I. M., & Toint P.L., CUTE: constrained and unconstrained testing environments. ACM Trans. Math. Softw. 21 (1995) 123–160.
6. Dai, Y. H., & Liao, L. Z. (2001). New conjugate conditions and related nonlinear conjugate gradient methods, Appl. Math. Optim. 43, 87–101.
7. Dai, Y. H., & Yuan, Y. X. (1999). A nonlinear conjugate gradient with a strong global convergence properties, SIAM J. Optim. 10, 177–182.
8. Dolan, E. D., & Moré, J. J. (2002). Benchmarking optimization software with performance profiles, Mathematical Programming Ser. A, 91, 201–213.
9. Fletcher, R. M., & Reeves, C. M. (1964). Function minimization by conjugate gradients, Comput. J. 7, 149–154.
10. Gilbert, J. C., & Nocedal, J. (1992). Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim. 2(1), 21–42.
11. Hager, W. W., & Zhang, H. (2005). A new conjugate gradient method with guaranteed descent and an efficient line search, SIAM J. Optim. 16, 170–192.
12. Hager, W. W., & Zhang, H. (2006a). Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent, ACM Transaction on Mathematical Software, 32(1), 113–137.
13. Hager, W. W., & Zhang, H. (2006b). A survey of nonlinear conjugate gradient methods, Pac. J. Optim. 2, 35–58.
14. Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems, J. Res. Nat. Bur. Standards, 49(6), 409–439.
15. Liu, D., & Xu, G. (2009). Applying Powells symmetrical technique to conjugate gradient methods, Computational Optimization and Applications, DOI 10.1007/s10589-009-9302-1.

16. Liu, D. & Shang Y. (2010). A New Perry Conjugate Gradient Method with the Generalized Conjugacy Condition, In: Computational Intelligence and Software Engineering (CiSE), 2010 International Conference on Issue Date: 10-12 Dec. 2010, DOI: 10.1109/CISE.2010.5677114
17. Perry, A. (1978). A Modified Conjugate Gradient Algorithm, Operations Research, Technical Notes, 26(6), 1073–1078.
18. Polak, E., & Ribière, G. (1969). Note sur la convergence de méthodes de directions conjuguées, Revue Françise d' Informatique et de Recherche Opérationnelle, 3(16), 35–43.
19. Polyak, B. T. (1969). The conjugate gradient method in extreme problems, USSR Comput. Math. Math. Phys. 9, 94–112.
20. Zoutendijk, G. (1970). Nonlinear Programming, Computational Methods. In J. Abadie (Eds.), Integer and Nonlinear Programming (pp. 37–86). Amsterdam: North-Holland.

Department of Mathematics, University of Tianjin, Tianjin, 300072, P R China
E-mail address: `dyliu@tju.edu.cn, corresponding author.`

Department of Mathematics, University of Tianjin, Tianjin, 300072, P R China
E-mail address: `gqxu@tju.edu.cn`