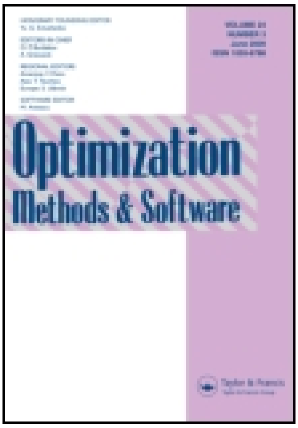

Optimization Methods and Software
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/goms20

Convergence of conjugate gradient methods with constant stepsizes

Yu-Hong Dai [a]
[a] State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, PO Box 2719, Beijing, 100190, People's Republic of China
Version of record first published: 09 Apr 2010.

To cite this article: Yu-Hong Dai (2011): Convergence of conjugate gradient methods with constant stepsizes, Optimization Methods and Software, 26:6, 895-909

To link to this article: http://dx.doi.org/10.1080/10556781003721042

Optimization Methods & Software
Vol. 26, No. 6, December 2011, 895–909

Taylor & Francis
Taylor & Francis Group

Convergence of conjugate gradient methods with constant stepsizes

Yu-Hong Dai*

State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, PO Box 2719, Beijing 100190, People's Republic of China

(*Received 17 November 2008; final version received 22 February 2010*)

We study the convergence properties of several conjugate gradient methods for nonlinear optimization under the assumptions that the objective function is bounded below and its gradient is Lipschitz continuous. Specifically, we strengthen the existing convergence result of the Polak–Ribière–Polyak method with constant stepsizes. For the method of shortest residuals, we establish global convergence of both the Fletcher–Reeves version and the Polak–Ribière–Polyak version using constant stepsizes. A numerical example is also presented.

Keywords: unconstrained optimization; nonconvex; conjugate gradient method; method of shortest residuals; descent property; global convergence

AMS Classifications: 49M37; 65K05; 90C30

1. Introduction

Consider the unconstrained optimization problem

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

where f is smooth and its gradient $\nabla f(x)$ is available. It is generally difficult to seek the global minimizer of f if the function has no particular properties, e.g. convexity. We are only interested in calculating a local minimizer of f, or even seeking an iterate $\{x_k\}$ satisfying the following property

$$\lim_{k\to\infty} \|g_k\| = 0, \tag{2}$$

where $g_k = \nabla f(x_k)$. The convergence relation (2) means that any cluster point of $\{x_k\}$ is a stationary point of f. Instead of (2), sometimes we use the slightly weaker relation

$$\liminf_{k\to\infty} \|g_k\| = 0, \tag{3}$$

which means that at least one cluster point of $\{x_k\}$ is a stationary point if $\{x_k\}$ is bounded.

*Email: dyh@lsec.cc.ac.cn

ISSN 1055-6788 print/ISSN 1029-4937 online

http://dx.doi.org/10.1080/10556781003721042
http://www.tandfonline.com

In this paper, we consider conjugate gradient methods for solving (1). They are suitable for large-scale problems since they do not need to store any matrices. The methods are of the form

$$x_{k+1} = x_k + \alpha_k d_k, \tag{4}$$

$$d_k = \begin{cases} -g_1, & \text{for } k = 1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{5}$$

in which α_k is a stepsize obtained by some method and β_k is the so-called conjugate gradient parameter. Two well-known formulae for β_k are Fletcher–Reeves (FR) [7] and Polak–Ribière–Polyak (PRP) [13,14], and are given by

$$\beta_k^{\rm FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{6}$$

and

$$\beta_k^{\rm PRP} = \frac{g_k^{\rm T} y_{k-1}}{\|g_{k-1}\|^2}, \tag{7}$$

respectively, where $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ is the Euclidean norm.

For general nonconvex functions, the FR method with exact or inexact line searches is shown to be globally convergent [1,3]. The PRP method performs much better than the FR method in practical computations. However, even with the exact line search, the PRP method needs not to converge, see Powell's [16] counter-examples. To change this unbalanced state, Gilbert and Nocedal [8] established the global convergence of the PRP method with the restriction $\beta_k^{\rm PRP} \geq 0$. Later, Grippo and Lucidi [9] designed an Armijo-type line search with which the original PRP method is shown to converge for general nonconvex functions. Sun and Zhang [19] established the convergence of the PRP method with the fixed stepsize

$$\alpha_k = -\frac{\delta g_k^{\rm T} d_k}{d_k^T Q_k d_k}. \tag{8}$$

They assumed that the positive-definite matrix Q_k in (8) satisfies $\nu_{\min}\|d\|^2 \leq d^{\rm T} Q_k d \leq \nu_{\max}\|d\|^2$ for all $d \in R^n$ and $k \geq 1$ and the constant $\delta \in (0, \nu_{\min}/L)$. Here and below, the word 'fixed stepsize' indicates 'without line search', i.e. 'without additional function evaluation'. The value L is the Lispchitz constant describing the gradient continuity (11). Some other results on the PRP method can be found in [6].

The Armijo-type line search designed for the PRP method in [9] seems somewhat complicated. By some further analysis, one can show that the Armijo-type line search in [9] is such that

$$\gamma \leq \alpha_k \leq \bar{\gamma}, \quad \text{for all } k \geq 1, \tag{9}$$

where $0 < \gamma < \bar{\gamma}$ are some positive constants. This nice property of the line search motivates us to consider the question: if constant stepsizes are used, namely,

$$\alpha_k \equiv \eta, \tag{10}$$

where $\eta > 0$ is some constant, does the PRP method converge? A part of the answer to this question is given in the Chinese monograph [5] by Dai and Yuan. If $\eta < 1/(4L)$, they show that the PRP method converges in the sense that (2) holds. In the next section, we will further show that if $\eta = 1/(4L)$, the PRP method provides the weak convergence relation (3). By considering two neighbouring iterations together (such a technique was once used in [4] to analyse the FR method), we can also prove the strong convergence relation (2) for the case that $\eta = 1/(4L)$. A counter-example is constructed showing that the PRP method may fail due to generating an uphill search direction provided that $\eta > 1/(4L)$. In Section 3, we will discuss the convergence properties of another class of conjugate gradient methods: the method of shortest residuals (SRs). The introduction of the SR method is presented in Section 3. Specifically, if $\eta < 2/L$, we are able to establish the convergence relation (3) for its FR version (FRSR), and (1.2) for its PRP version (PRPSR). Our analyses are made under the assumptions that the objective function is bounded below and its gradient is Lipschitz continuous. A numerical example is presented in Section 4 and some discussion is given in the last section.

2. Convergence of PRP method with constant stepsizes

We give the following basic assumption on the objective function.

ASSUMPTION 2.1 (1) $f(x)$ *is bounded below.* (2) *In a neighbourhood* $\mathcal{N}$ *of the level set* $\mathcal{L} = \{x \in \Re^n \colon f(x) \le f(x_1)\}$, *where* x_1 *is the starting point*, $f(x)$ *is differentiable and its gradient* $g(x)$ *is Lipschitz continuous*, *namely*, *there exists a constant* $L > 0$ *such that*

$$\|\nabla f(x) - \nabla f(\tilde{x})\| \le L\|x - \tilde{x}\|, \quad \text{for any } x, \quad \tilde{x} \in \mathcal{N}. \tag{11}$$

First, we provide the following lemmas for the PRP method. In [5], two important relations (3.5.5) and (3.5.6) are established to prove the convergence result with $\eta < 1/(4L)$. These relations are no longer true if $\eta = 1/(4L)$. However, when $\eta \in (0, 1/(4L)]$, we can establish the stronger relations (17) and (18). They enable us to prove the convergence for the case of $\eta = 1/(4L)$.

Here we should mention that the stepsize α_k is fixed to some constant value in $(0, 1/(4L)]$ in the following analysis. The following results can therefore be extended to the case when $\alpha_k \in [\tau, 1/(4L)]$ for some $\tau > 0$ and all large k. Further, as mentioned in Section 5, the PRP method gives the weak convergence relation (3) provided that $\alpha_k \in (0, 1/(4L)]$ is such that $\sum_{k\ge1} \alpha_k = +\infty$. This provides us with the possibility of designing some dynamic strategy for choosing α_k, for example, asking α_k to depend on the quantity d_k, g_k, etc.

LEMMA 2.2 *Assume that* η *is some constant in* $(0, 1/(4L)]$. *Define the sequence* $\{\xi_k\}$ *as follows*:

$$\xi_1 = 1; \quad \xi_{k+1} = 1 + L\eta\xi_k^2, \;\; k \ge 1. \tag{12}$$

Then we have that

$$1 \le \xi_k < c, \quad \text{for all } k \ge 1, \tag{13}$$

where c is the constant

$$c = 2\left(1 + \sqrt{1 - 4L\eta}\right)^{-1}, \tag{14}$$

that satisfies

$$1 < c \le 2. \tag{15}$$

Proof We show (13) by induction. Since $\xi_1 = 1$, we have that $1 \le \xi_1 < c$. Assume that $1 \le \xi_k < c$ for some k. It follows from (12)–(15) that

$$1 \le \xi_{k+1} = 1 + L\eta\xi_k^2 < 1 + L\eta c^2 = c. \tag{16}$$

Thus, by induction, (13) is true. ■

We see from the above proof that the condition $\eta \le 1/(4L)$ enables us to derive a uniform upper bound for the sequence ξ_k in (12). Otherwise, if η exceeds $1/(4L)$, the constant c in (14) fails its definition and we cannot derive an upper bound for ξ_k any more.

LEMMA 2.3 *Suppose that Assumption* 2.1 *holds. Consider the PRP method* (4)*,* (5) *and* (7) *with constant stepsizes* (10), *where* $\eta \in (0,\ 1/(4L)]$. *Then we have for all* $k \ge 1$,

$$(2-\xi_k)\|g_k\|^2 \le -g_k^{\mathrm{T}} d_k \le \xi_k \|g_k\|^2, \tag{17}$$

$$(2-\xi_k)\|g_k\| \le \|d_k\| \le \xi_k \|g_k\|, \tag{18}$$

where ξ_k *is the sequence defined in* (12)*. Further*, *each search direction* d_k *is downhill*, *namely*,

$$g_k^{\mathrm{T}} d_k < 0, \quad \text{for all } k \ge 1. \tag{19}$$

Proof Since $d_1 = -g_1$ and $\xi_1 = 1$, (17) and (18) clearly hold for $k = 1$. Assume that (17) and (18) hold for some k. Then by (5), (7), the Cauchy-Schwartz inequality, (11), (10) and the induction assumption, we can get that

$$\begin{aligned} \|d_{k+1} + g_{k+1}\| &= \|\beta_{k+1}^{\mathrm{PRP}} d_k\| \\ &\le \frac{\|g_{k+1} - g_k\|\|d_k\|}{\|g_k\|^2}\|g_{k+1}\| \\ &\le \frac{L\alpha_k\|d_k\|^2}{\|g_k\|^2}\|g_{k+1}\| \\ &\le L\eta\xi_k^2\|g_{k+1}\|. \end{aligned} \tag{20}$$

By the triangular inequality and relation (20), we obtain

$$\|d_{k+1}\| \le \|d_{k+1} + g_{k+1}\| + \|g_{k+1}\| \le (1 + L\eta\xi_k^2)\|g_{k+1}\| \tag{21}$$

and

$$\|d_{k+1}\| \ge \|g_{k+1}\| - \|d_{k+1} + g_{k+1}\| \ge (1 - L\eta\xi_k^2)\|g_{k+1}\|. \tag{22}$$

The above relations and the definition of $\{\xi_k\}$ in (12) indicate that (18) holds for $k+1$. By (20) and the Cauchy–Schwartz inequality, we have that

$$|g_{k+1}^{\mathrm{T}} d_{k+1} + \|g_{k+1}\|^2| \le \|g_{k+1}\|\|d_{k+1} + g_{k+1}\| \le L\eta\xi_k^2\|g_{k+1}\|^2. \tag{23}$$

Similarly, by (23), the triangular inequality and definition (12), we know that (17) holds for $k+1$. Therefore, by induction, (17) and (18) hold for all $k \ge 1$.

The descent property (19) then follows the first inequality in (13), (15) and (17). ■

Again, we see that the condition that $\eta \le 1/(4L)$ is essential. It enables us to derive a uniform upper bound for the sequence ξ_k, with which we can deduce the descent property of the search direction and obtain the bounds for the search direction. Further analysis implies that the sequence $\{-g_k^{\mathrm{T}} d_k\}$ is summable (Lemma 2.4). An illustrative example is given at the end of this section, showing that the method may generate an uphill search direction if $\eta > 1/(4L)$.

LEMMA 2.4 *Suppose that Assumption* 2.1 *holds. Consider the PRP method* (4)*,* (5) *and* (7) *with constant stepsizes* (10) *where* $\eta \in (0,\ 1/(4L))$. *Then we have that*

$$\sum_{k=1}^{\infty} g_k^{\mathrm{T}} d_k > -\infty. \tag{24}$$

Proof By the mean value theorem, (4) and (11), we have that

$$\begin{aligned} f(x_{k+1}) - f(x_k) &= \int_0^1 g(x_k + t\alpha_k d_k)^{\mathrm{T}}(\alpha_k d_k)\mathrm{d}t \\ &= \alpha_k g_k^{\mathrm{T}} d_k + \int_0^1 [g(x_k + t\alpha_k d_k) - g_k]^{\mathrm{T}}(\alpha_k d_k)\mathrm{d}t \\ &\le \alpha_k g_k^{\mathrm{T}} d_k + \int_0^1 Lt\alpha_k^2 \|d_k\|^2 \mathrm{d}t \\ &\le \alpha_k g_k^{\mathrm{T}} d_k + \frac{1}{2} L\alpha_k^2 \|d_k\|^2. \end{aligned} \tag{25}$$

Now we estimate an upper bound of the quantity $\|d_k\|^2$ for the method. Relation (5) indicates that

$$d_k + g_k = \beta_k d_{k-1}. \tag{26}$$

By (12), (13), (15), (20) with $k+1$ replaced by k and (26), we have that

$$\beta_k^2 \|d_{k-1}\|^2 \le (L\eta\xi_{k-1}^2)^2 \|g_k\|^2 = (\xi_k - 1)^2 \|g_k\|^2 \le \|g_k\|^2. \tag{27}$$

By squaring the norm of (26), we can also obtain

$$\|d_k\|^2 = -\|g_k\|^2 - 2g_k^{\mathrm{T}} d_k + \beta_k^2 \|d_{k-1}\|^2. \tag{28}$$

Then it follows by (27) and (28) that

$$\|d_k\|^2 \le -2g_k^{\mathrm{T}} d_k. \tag{29}$$

Now, by (10), (25) and (29), we get that

$$f(x_{k+1}) - f(x_k) \le (\eta - L\eta^2) g_k^{\mathrm{T}} d_k. \tag{30}$$

Summing (30) over k and noting that $\eta - L\eta^2 > 0$, we obtain

$$\sum_{i=1}^{k} g_i^{\mathrm{T}} d_i \ge (\eta - L\eta^2)^{-1} (f(x_{k+1}) - f(x_1)). \tag{31}$$

Since by (30), $f(x_k)$ is monotonically decreasing, we have that $\{x_k\} \subset \mathcal{L}$. Further, we know by Assumption 2.1 that $\{f(x_k)\}$ is bounded below. Thus, (24) follows from (31). ■

We are now ready to establish the convergence of the PRP method using constant stepsizes. Part (i) of Theorem 2.5 has been obtained in [6].

THEOREM 2.5 *Suppose that Assumption* 2.1 *holds. Consider the PRP method* (4), (5) *and* (7) *with constant stepsizes* (10)*, where* $\eta \in (0, 1/(4L)]$. *Then*

(i) *if* $0 < \eta < 1/(4L)$, *relation (2) holds. Thus every cluster point of* $\{x_k\}$ *is a stationary point of* f;
(ii) *if* $\eta = 1/(4L)$, *relation* (3) *holds. Thus, at least one of the cluster points of* $\{x_k\}$ *is a stationary point of* f.

Proof (i) Since $0 < \eta < 1/(4L)$, we have by (13), (15) and (17) that

$$-g_k^{\mathrm{T}} d_k \geq (2 - c)\|g_k\|^2. \tag{32}$$

Then we get from (15), (32) and Lemma 2.4 that

$$\sum_{k=1}^{\infty} \|g_k\|^2 < +\infty. \tag{33}$$

Therefore, $\lim_{k\to\infty} g_k = 0$, which implies that every cluster point of $\{x_k\}$ is a stationary point of f.

(ii) Define $\tau_k = 2 - \xi_k$. Then it follows from (12) and $\eta = 1/(4L)$ that

$$\tau_{k+1} = \tau_k \left(1 - \frac{1}{4}\tau_k\right). \tag{34}$$

Since $\tau_1 = 1$, we have that

$$0 < \tau_k \leq 1, \quad \text{for all } k \geq 1. \tag{35}$$

We now prove that

$$\tau_k \geq k^{-1}, \quad \text{for all } k \geq 1. \tag{36}$$

In fact, since $\xi_1 = 1$, it is obvious that $\tau_1 = 2 - \xi_1 \geq 1$. Assume that $\tau_k \geq k^{-1}$. Noting that $\tau_k(1 - (1/4)\tau_k)$ is monotonically increasing with τ_k in [0, 1], we have from (34), (35) and the induction assumption that

$$\tau_{k+1} \geq k^{-1}\left[1 - \frac{1}{4}k^{-1}\right] \geq (k+1)^{-1}. \tag{37}$$

Thus, by induction, (36) holds. We now proceed by contradiction, assuming that (3) does not hold. Then there exists some constant $\delta > 0$ such that

$$\|g_k\| \geq \delta, \quad \text{for all } k \geq 1. \tag{38}$$

By (38), the first inequality in (17) and (36), we can get that

$$\sum_{k=1}^{\infty} g_k^{\mathrm{T}} d_k \leq -\delta^2 \sum_{k=1}^{\infty} k^{-1} = -\infty, \tag{39}$$

which contradicts Lemma 2.4. The contradiction shows that $\liminf_{k\to\infty} \|g_k\| = 0$, which completes our proof. ■

From relation (32) in the above proof, we see that if $\eta < 1/(4L)$, the PRP method with constant stepsizes (10) is such that the sufficient descent condition

$$g_k^{\mathrm{T}} d_k \leq -\bar{c}\,\|g_k\|^2, \quad \text{for all } k \geq 1 \tag{40}$$

holds with $\bar{c} = 2 - c > 0$. At the same time, we have from (13) and (18) that

$$\bar{c}\,\|g_k\| \leq \|d_k\| \leq c\,\|g_k\|. \tag{41}$$

The above relation implies that the size of $\|d_k\|$ relative to $\|g_k\|$ is both upper and lower bounded. Further, letting θ_k be the angle between d_k and $-g_k$, we have from the definition of θ_k, (13), (18) and (40) that

$$\cos\theta_k = -\frac{g_k^{\mathrm{T}} d_k}{(\|g_k\|\|d_k\|)} \geq \frac{\bar{c}}{c}. \tag{42}$$

Thus, in this case, the angle θ_k is not greater than some angle $\theta < \pi/2$. If in addition the objective function is uniformly convex, it is not difficult to establish the linear convergence rate of the method by (30) and (42).

In the case that $\eta = 1/(4L)$, we have by (14) that $c = 2$. Thus, we cannot deduce the sufficient descent condition from (17). Nevertheless, we have from (36) that

$$g_k^{\mathrm{T}} d_k \leq -k^{-1}\|g_k\|^2, \quad \text{for all } k \geq 1, \tag{43}$$

with which a weaker convergence relation can be established. Further, we still have from (18) and $c = 2$ that $\|d_k\| \leq 2\|g_k\|$. By this and (43), we can get the following relation for the angle θ_k:

$$\cos\theta_k \geq \frac{1}{2k}. \tag{44}$$

The previous analysis was to estimate the quantity

$$\gamma_k = -\frac{g_k^{\mathrm{T}} d_k}{\|g_k\|^2} \tag{45}$$

for every iteration. For the case that $\eta = 1/(4L)$, if we consider two neighbouring steps together (such technique was once used in [4] to analyse the FR method), we can establish the relation

$$\max(\gamma_k, \gamma_{k+1}) \geq \frac{2}{3}, \quad \text{for any } k \geq 1. \tag{46}$$

This relation means that the sufficient descent condition $-g_k^{\mathrm{T}} d_k \geq 2/3\|g_k\|^2$ holds for at least one of any neighbouring iterations. Along this line, we can prove the strong convergence relation (2) for the case that $\eta = 1/(4L)$. Further, by (41) and (46), we could establish the two-step linear convergence result in this case.

THEOREM 2.6 *Suppose that Assumption* 2.1 *holds. Consider the PRP method (*4*), (*5*) and (*7*) with constant stepsizes (*10*) where $\eta \equiv 1/(4L)$. Then the relation (*46*) holds. Further, the method gives the convergence relation (*2*), which means that every cluster point of $\{x_k\}$ is a stationary point of f.*

Proof By the definition of γ_k and relation (29), we have that

$$\|d_k\|^2 \leq 2\,\gamma_k\,\|g_k\|^2. \tag{47}$$

From the proofs of relations (20) and (23), we can see that

$$|g_{k+1}^{\mathrm{T}} d_{k+1} + \|g_{k+1}\|^2| \leq \|g_{k+1}\|\|d_{k+1} + g_{k+1}\| \leq \frac{L\alpha_k\|d_k\|^2}{\|g_k\|^2}\|g_{k+1}\|^2. \tag{48}$$

Therefore, using (47) and $\eta = 1/(4L)$, it follows from the above relation that

$$-g_{k+1}^{\mathrm{T}} d_{k+1} \geq \left(1 - \frac{\gamma_k}{2}\right)\|g_{k+1}\|^2, \tag{49}$$

or, equivalently,

$$\gamma_{k+1} \geq 1 - \frac{\gamma_k}{2}. \tag{50}$$

If $\gamma_k \leq 2/3$, we must have by (50) that $\gamma_{k+1} \geq 2/3$. Thus, relation (46) always holds.

To show (2), we note from $\gamma_k > 0$ and (50) that

$$\gamma_k + \gamma_{k+1} \geq \frac{1}{2}(\gamma_k + 2\gamma_{k+1}) \geq 1. \tag{51}$$

By Lemma 2.4, (29) and the definition of γ_k, we get that

$$\sum_{k=1}^{\infty}(\gamma_k \|g_k\|^2 + \gamma_{k+1}\|g_{k+1}\|^2) \leq 2\sum_{k=1}^{\infty}(-g_k^{\mathrm{T}} d_k) < +\infty. \tag{52}$$

On the other hand, by Lipschitz condition (11), relation (18), $\xi_k \leq 2$ and $\eta = 1/(4L)$, we can get that

$$\|g_{k+1}\| \leq \|g_k\| + \alpha_k L\|d_k\| \leq \left(1 + \frac{1}{4}\xi_k\right)\|g_k\| \leq \frac{3}{2}\|g_k\|. \tag{53}$$

Thus, we have that

$$\gamma_k\|g_k\|^2 + \gamma_{k+1}\|g_{k+1}\|^2 \geq \left(\frac{4}{9}\gamma_k + \gamma_{k+1}\right)\|g_{k+1}\|^2 \geq \frac{4}{9}(\gamma_k + \gamma_{k+1})\|g_{k+1}\|^2 \geq \frac{4}{9}\|g_{k+1}\|^2. \tag{54}$$

Therefore, by (52) and (54), we know that (2) is true. ■

If the parameter η is greater than $1/(4L)$, we can construct a one-dimensional counter-example showing that the PRP method may fail due to generating an uphill search direction. Suppose that $L = 1$ and η is any constant greater than $1/4$. Consider the sequence $\{\xi_k\}$ defined in (12). Since in this case $L\eta > 1/4$, it is easy to show that ξ_k tends to $+\infty$ with k. Then we can pick a positive integer $K \geq 1$ such that

$$\xi_K > \frac{\max\{3, 1 + 2\eta\}}{\eta}. \tag{55}$$

The desired counter-example is based on the one-dimensional function

$$f(x) = -\frac{1}{2}x^2, \quad x \in [0, x_K], \tag{56}$$

where x_K is the Kth iteration generated by the PRP method with the starting point $x_1 = 1$ and constant stepsizes $\alpha_k \equiv \eta$. Since $x_{k+1} = x_k + \eta d_k$ and $g_k = -x_k$, we have that $g_{k+1} = g_k - \eta d_k$. Noting this and the dimension $n = 1$, we can see that

$$d_{k+1} = -g_{k+1} - \eta\left(\frac{d_k^2}{g_k^2}\right)g_{k+1}. \tag{57}$$

Define $\xi_k = -d_k/g_k$. The above relation and $d_1 = -g_1$ implies that $\{\xi_k\}$ has the same recursive relation as (12). Consequently, we have that $x_{k+1} = x_k - \eta\xi_k g_k = (1 + \eta\xi_k)x_k$, which with $x_1 = 1$ yields

$$x_K = \prod_{i=1}^{K-1}(1 + \eta\xi_i). \tag{58}$$

On other hand, by the choice of integer K and the definition of ξ_k,

$$-\frac{d_K}{g_K} > \frac{\max\{3,\ 1+2\eta\}}{\eta}. \tag{59}$$

Now we define f in the interval $[x_K, x_K + \eta d_K]$ so that

$$f(x_K) = -\frac{1}{2}x_K^2; \quad \nabla f(x_K) = g_K; \quad \nabla^2 f(x_K + \eta t d_K) = -1 + 2\sqrt{t}, \quad \text{for } t \in (0, 1]. \tag{60}$$

Then we will have that

$$\nabla f(x_K + \eta t d_K) = g_K + \int_0^t \nabla^2 f(x_K + \eta\tau d_K)\eta d_K \,\mathrm{d}\tau = g_K + \left(-t + \frac{4}{3}t^{3/2}\right)\eta d_K \tag{61}$$

and

$$\begin{aligned} f(x_K + \eta t d_K) &= f(x_K) + \int_0^t \nabla f(x_K + \eta\tau d_K)\eta d_K \,\mathrm{d}\tau \\ &= f(x_K) + \eta g_K d_K t - \frac{1}{2}\eta^2 d_K^2 t^2 + \frac{8}{15}\eta^2 d_K^2 t^{5/2}. \end{aligned} \tag{62}$$

Letting $t = 1$ in (61) and noticing (59) and $g_K = -x_K < 0$ that

$$g_{K+1} = g_K + \frac{\eta}{3} d_K = g_K \left(1 + \frac{\eta}{3}\frac{d_K}{g_K}\right) > 0. \tag{63}$$

At the same time, we know by

$$d_{K+1} = -g_{K+1} + \frac{1}{3}\eta\left(\frac{d_K^2}{g_K^2}\right) g_{K+1}, \tag{64}$$

the relation (59) and $\eta > 1/4$ that $d_{K+1}^{\mathrm{T}} g_{K+1} > 0$ and hence d_{K+1} is an uphill search direction.

To complete the structure of f, we define f for $x \ge x_{K+1}$ as follows:

$$f(x) = \frac{1}{2}x^2 + (g_{K+1} - x_{K+1})x + \left(f_{K+1} - g_{K+1}x_{K+1} + \frac{1}{2}x_{K+1}^2\right) \tag{65}$$

so that $\nabla^2 f(x) = 1$ for $x \ge x_{K+1}$ and the whole function f is twice-differentiable at the point x_{K+1}. The convexity of f over the interval $[x_{K+1}, +\infty)$ and $g_{K+1} > 0$ implies that the minimizer of f over $[0, +\infty)$ lies in the interval (x_K, x_{K+1}). However, some direct calculations show that

$$\frac{d_{k+1}}{g_{k+1}} = -1 + \eta\frac{d_k^2}{g_k^2}, \quad k \ge K + 1. \tag{66}$$

By (59), (64) and considering the two cases $\eta \in (1/4, 1]$ and $\eta > 1$, we can obtain

$$\frac{d_{K+1}}{g_{K+1}} > -1 + \frac{\eta}{3}\frac{[\max\{3, 1+2\eta\}]^2}{\eta^2} > \frac{1+\sqrt{1+4\eta}}{2\eta}. \tag{67}$$

It follows from (66) and (67) that d_{k+1}/g_{k+1} tends to $+\infty$ monotonically. Meanwhile, noting that $\nabla^2 f(x) = 1$ now, we have that $x_{k+1} = x_k + \eta d_k$ for $k \ge K + 1$. Therefore, d_k is always an uphill search direction for $k \ge K + 1$ and x_{k+1} goes to $+\infty$.

It is not easy to see that the above-defined function satisfies Assumption 2.1 with $L = 1$. However, the PRP method with the starting point $x_1 = 1$ and constant stepsize $\eta > 1/4$ fails to provide or converge to any stationary point. This counter-example shows that the bound $\eta = 1/(4L)$ is strict.

3. Convergence of PRPSR and FRSR with constant stepsizes

In this section, we discuss the convergence properties of another class of conjugate gradient methods: the method of SRs. The SR method was presented by Hestenes in his monograph [10] on conjugate direction methods and it can be viewed as a special case of conjugate subgradient method developed in Wolfe [20] and Lemaréchal [12] for minimizing a convex function. Pytlak [17] generalized the SR method and considered the following family of methods:

$$d_k = -\mathrm{Nr}\{g_k, -\beta_k d_{k-1}\}, \tag{68}$$

where $\mathrm{Nr}\{a,\ b\}$ is defined as the point from a line segment spanned by the vectors a and b which has the smallest norm, namely,

$$\|\mathrm{Nr}\{a, b\}\| = \min\{\|\lambda a + (1-\lambda)b\| : 0 \le \lambda \le 1\}. \tag{69}$$

In (68), we still use the symbol β_k to differentiate variants of the SR method. If $g_k^{\mathrm{T}} d_{k-1} = 0$, the above family with

$$\beta_k \equiv 1 \tag{70}$$

reduces to the SR method described by Hestenes [10]. Further, if the function is quadratic, the vector d_k can be proved to be the SR in the $(k-1)$-simplex whose vertices are $-g_1, -g_2, \ldots, -g_k$. Another choice for the scalar β_k is that

$$\beta_k = \frac{\|g_k\|^2}{g_k^{\mathrm{T}} y_{k-1}}, \tag{71}$$

where $y_{k-1} = g_k - g_{k-1}$ as before. In case of exact line searches, the family with the choices (70) and (71) produce the same search directions for general objective functions as the FR method and the PRP method, respectively. In this paper, we describe the family of methods (4) and (68) as method of SRs, and the corresponding methods with (70) and (71) as FRSR and PRPSR. In [4,17], the choice $\beta_k = \|g_k\|^2/|g_k^{\mathrm{T}} y_{k-1}|$ is also considered for the PRPSR method. It is easy to see that the following analysis of PRPSR applies to this choice.

By solving (69) without the restriction $0 \le \lambda \le 1$, Dai and Yuan [4] obtain the following direction

$$d_k = -(1-\lambda_k) g_k + \lambda_k \beta_k d_{k-1}, \tag{72}$$

where

$$\lambda_k = \frac{\|g_k\|^2 + \beta_k g_k^{\mathrm{T}} d_{k-1}}{\|g_k + \beta_k d_{k-1}\|^2}. \tag{73}$$

By direct calculations, we can obtain the following two important relations for (72) and (73) [4]:

$$-g_k^{\mathrm{T}} d_k = \|d_k\|^2 \tag{74}$$

and

$$\|d_k\|^2 = \frac{\beta_k^2 \left(\|g_k\|^2 \|d_{k-1}\|^2 - (g_k^{\mathrm{T}} d_{k-1})^2\right)}{\|g_k + \beta_k d_{k-1}\|^2}. \tag{75}$$

Relation (74) shows that d_k is a descent direction unless $d_k = 0$. However, it is possible that $d_k = 0$ if g_k and d_{k-1} are collinear. For convenience, we assume in this paper that $d_k \neq 0$ and the denominator of (71) $g_k^{\mathrm{T}} y_{k-1} \neq 0$ so that the method is well defined.

To present our convergence results for FRSR and PRPSR using the constant stepsize (10) with $\eta \in (0, 2/L)$, we give the following lemma for the SR method at first.

LEMMA 3.1 *Suppose that Assumption 2.1 holds. Consider the SR method (4) and (72), that takes $d_1 = -g_1$ and constant stepsizes (10) with $0 < \eta < 2/L$. Then we have that*

$$|g_k^{\mathrm{T}} d_{k-1}| \leq c_1 \|d_{k-1}\|^2, \tag{76}$$

where $c_1 = 1 + \eta L$ is constant, and

$$\sum_{k\geq 1} \|d_k\|^2 < +\infty. \tag{77}$$

Proof By the triangle inequality, (10), (11) and (74), we have that

$$\begin{aligned}
|g_k^{\mathrm{T}} d_{k-1}| &= |(g_k - g_{k-1})^{\mathrm{T}} d_{k-1} + g_{k-1}^{\mathrm{T}} d_{k-1}| \\
&\leq \|g_k - g_{k-1}\| \|d_{k-1}\| + |g_{k-1}^{\mathrm{T}} d_{k-1}| \\
&\leq \alpha_{k-1} L \|d_{k-1}\|^2 + \|d_{k-1}\|^2 \\
&\leq (1 + \eta L) \|d_{k-1}\|^2.
\end{aligned} \tag{78}$$

Thus, (76) holds with $c_1 = 1 + \eta L$. In addition, note that the relation (25) still holds. By this, (10) and (74), we get that

$$f(x_{k+1}) - f(x_k) \leq \eta \left(-1 + \frac{L}{2}\eta \right) \|d_k\|^2. \tag{79}$$

Summing (79) over k and noting that $\eta(1 - (L/2)\eta) > 0$, we obtain

$$\sum_{i=1}^{k} \|d_i\|^2 \leq \left[\eta (1 - \frac{L}{2}\eta) \right]^{-1} (f(x_1) - f(x_{k+1})). \tag{80}$$

Since by (79), $f(x_k)$ is monotonically decreasing, we have that $\{x_k\} \subset \mathcal{L}$. Further, we know by Assumption 2.1 that $\{f(x_k)\}$ is bounded below. Thus, (77) follows (80). ∎

Next, we prove the following theorem for the FRSR method.

THEOREM 3.2 *Suppose that Assumption 2.1 holds. Consider the FRSR method (4) and (72) where β_k is given by (70). If constant stepsizes (10) with $0 < \eta < 2/L$ are used, the method converges in the sense that (3) holds.*

Proof From (70) and (75), we can get that

$$\frac{1}{\|d_k\|^2} = \frac{1}{\|d_{k-1}\|^2} (1 + r_k), \tag{81}$$

where

$$r_k = \frac{\|d_{k-1}\|^2 + 2 g_k^{\mathrm{T}} d_{k-1} + (g_k^{\mathrm{T}} d_{k-1})^2 / \|d_{k-1}\|^2}{\|g_k\|^2 - (g_k^{\mathrm{T}} d_{k-1})^2 / \|d_{k-1}\|^2}. \tag{82}$$

The recursive use of (81) yields

$$\frac{1}{\|d_k\|^2} = \frac{1}{\|d_1\|^2} \prod_{i=2}^{k} (1 + r_i). \tag{83}$$

By (77) in Lemma 3.1, we know that

$$\lim_{k\to\infty} \|d_k\| = 0. \tag{84}$$

If the statement of this theorem is not true, there must exist some constant $\delta > 0$ such that

$$\|g_k\| \geq \delta, \quad \text{for all } k \geq 1. \tag{85}$$

By (76), there exists an integer k_1 such that $\|d_k\| \leq \delta/2c_1$ for all $k \geq k_1$. It then follows from (82), (84) and (85) that

$$r_k \leq c_2\|d_{k-1}\|^2, \quad \text{for all } k \geq k_1 + 1, \tag{86}$$

where $c_2 = 4(1 + c_1)^2/3\delta^2$ is constant. The above relation and (77) imply that $\sum_{k\geq 1} r_k < +\infty$ and hence the product

$$\prod_{k\geq 2}(1 + r_k) = \exp\left(\sum_{k\geq 2}\log(1 + r_k)\right) \leq \exp\left(\sum_{k\geq 2} r_k\right) < +\infty. \tag{87}$$

The above relation and (83) indicate that $\|d_k\|$ is bounded away from zero, a contradiction to (84). Therefore, this theorem is true. ■

Finally, we establish the following convergence result for the PRPSR method.

THEOREM 3.3 *Suppose that Assumption* 2.1 *holds. Consider the PRPSR method* (4) *and* (72) *where* β_k *is given by* (71). *If constant stepsizes* (10) *with* $0 < \eta < 2/L$ *are used, the method converges in the sense that* (2) *holds.*

Proof By the choice (10), (11), and (71), we have that

$$\|\beta_k d_{k-1}\| = \frac{\|g_k\|^2\|d_{k-1}\|}{|g_k^{\mathrm{T}}(g_k - g_{k-1})|} \geq \frac{\|g_k\|^2\|d_{k-1}\|}{L\|g_k\|\alpha_{k-1}\|d_{k-1}\|} \geq (L\eta)^{-1}\|g_k\|. \tag{88}$$

By (75) and (88), for k sufficiently large, we have that

$$\begin{aligned}\frac{1}{\|d_k\|^2} &= \left(\frac{\|g_k + \beta_k d_{k-1}\|^2}{\beta_k^2\|g_k\|^2\|d_{k-1}\|^2}\right)\left(1 - \frac{(g_k^{\mathrm{T}}d_{k-1})^2}{\|g_k\|^2\|d_{k-1}\|^2}\right)^{-1}\\ &\leq \left(\frac{1}{\|\beta_k d_{k-1}\|} + \frac{1}{\|g_k\|^2}\right)^2\left(1 - \frac{(g_k^{\mathrm{T}}d_{k-1})^2}{\|g_k\|^2\|d_{k-1}\|^2}\right)^{-1}\\ &\leq \frac{(1 + L\eta)^2}{\|g_k\|^2}\left(1 - \frac{(g_k^{\mathrm{T}}d_{k-1})^2}{\|g_k\|^2\|d_{k-1}\|^2}\right)^{-1}.\end{aligned} \tag{89}$$

Assuming $\|g_k\|^2 \geq 2c_1\|d_{k-1}\|^2$, we can get by (89) and (76) that

$$\|g_k\|^2 \leq (1 + L\eta)^2\|d_k\|^2\left(1 - \frac{c_1\|d_{k-1}\|^2}{\|g_k\|^2}\right)^{-1} \leq 2(1 + L\eta)^2\|d_k\|^2. \tag{90}$$

Therefore we always have

$$\|g_k\|^2 \leq \max\{2c_1\|d_{k-1}\|^2,\ 2(1 + L\eta)^2\|d_k\|^2\}. \tag{91}$$

The above relation with (77) indicates the truth of this theorem. ■

4. A numerical example

The following objective function is used for our preliminary numerical tests,

$$f(x) = \frac{1}{2}x^{\mathrm{T}}Hx, \quad x \in R^5, \tag{92}$$

where the Hessian H is the Hilbert matrix with entries $H_{i,j} = 1/(i + j - 1), i, j = 1, \ldots, 5$. In this case, the condition number of H is $4.7661e + 5$. The tight Lipshitz constant for the objective function is the maximal eightvalue of H, that is $L = 1.5671$ in four digits decimal precision. In all tests, we use the initial point $x_1 = \sqrt{5}/5(1, -1, 1, -1, 1)^{\mathrm{T}}$ and the stopping condition is

$$\|g_k\| \leq 10^{-4}\|g_1\|. \tag{93}$$

We tested the following seven methods: (1) steepest descent (SD) method; (2) FR method; (3) PRP method; (4) FRSR, defined by (70), (72), and (73); (5) PRPSR, defined by (74)–(76); (6) SDFR, where d_k is the SD direction and the FR direction alternatively; (7) SDPRP, where d_k is the SD direction and the PRP direction alternatively. For these methods, we used the constant stepsize of the form

$$\alpha_k \equiv \frac{\mu}{L}, \tag{94}$$

where $\mu \in (0, 2)$ is some fixed parameter. In our tests, nine values are used (Table 1). The iteration number for each method to achieve (93) is listed in Table 1.

From Table 1, we see that the FRSR and PRPSR methods with different values of μ in (0, 2) successfully solved the given example, as predicted by Theorems 3.2 and 3.3. This theoretical property can also be extended to the SD method. In fact, if $d_k = -g_k$ and α_k is given by (94) with $\mu \in (0,\ 2)$, and if Assumption 2.1 holds, we can establish some relations similar to (20) and (79) and finally

$$f(x_{k+1}) - f(x_k) \leq -\frac{\mu(2-\mu)}{L}\ \|g_k\|^2. \tag{95}$$

By summing (95)) over k, we can easily see that $\sum_{k\geq 1} \|g_k\|^2 < +\infty$, which gives the convergence relation $\lim_{k\to\infty} \|g_k\| = 0$. This extends the result in Armijo [2], which shows the convergence of the SD method when $\mu \in (0, 1/2)$.

Although Theorems 2.5 and 2.6 only guarantee the global convergence of the PRP method when $\mu \in (0, 1/4]$, we see from Table 1 that the PRP method with large values of μ requires fewer iterations to achieve the same stopping condition. Nevertheless, it was found that the PRP directions in the SDPRP method with $\mu = 1.75$ or 1.90 are always uphill. This leads to the failure

Table 1. Numerical results for different methods with (94).

	Method						
μ	SD	FR	PRP	FRSR	PRPSR	SDFR	SDPRP
0.10	8739	390	8748	9351	17466	5829	8744
0.25	3495	244	3503	4313	6980	2333	3500
0.50	1747	170	1755	2558	3484	1167	1751
0.75	1165	135	1172	1192	2320	779	1168
1.00	873	116	880	3424	1739	586	877
1.25	699	106	703	730	1596	500	700
1.50	582	101	584	649	931	456	561
1.75	499	92	492	476	715	470	Failed
1.90	459	88	412	462	673	488	Failed

Table 2. Numerical results for different methods with $\alpha_k = 1/L_k$.

Method	SD	FR	PRP	FRSR	PRPSR	SDFR	SDPRP
# Iter	870	99	876	902	1729	584	873

of the SDPRP method due to numerical overflows. Similarly, although Table 1 shows that the FR and SDFR methods perform much better than the SD method for the given problem, we have not presented any convergence results for the FR and the SDFR methods. Considerable efforts are still required to establish efficient numerical algorithms based on the methods in both deterministic optimization and stochastic approximation.

We can also see from Table 1 that for each method the choice of μ influences its numerical performance significantly. Particularly, when μ is relatively small, they always require quite many iterations to achieve the stopping condition. In practice, to ensure that the SD method works, a small value for the stepsize is preferred, since it is difficult to know the value of the Lipschitz constant L. In the deterministic case, one may estimate the value of L, for example, in the following way

$$L_1 = 0.01; \quad L_k = \max\left\{\frac{\|y_{k-i}\|}{\|s_{k-i}\|};\ i = 1, \ldots, k-1\right\}, \quad \text{for } k \geq 2. \tag{96}$$

In (96), $y_{k-1} = g_k - g_{k-1}$ as before and $s_{k-1} = x_k - x_{k-1}$. With this estimation, we tested the seven methods again and listed the iteration numbers required for achieving (93) in Table 2.

5. Discussion

One advantage of the PRP method over the FR method in practical computations is that [15], if a small step is generated far away from the solution point, the direction in the PRP method will tend to the negative gradient direction. In this paper, we use this property and establish the global convergence of the PRP method with constant stepsizes. Hence, the result of this paper provides some further insights into the convergence theory of the PRP method and a better understanding of the convergence result in [9].

Sun and Zhang [19] established the global convergence of the FR method with fixed stepsize $\alpha_k = c_3(|g_k^T d_k|/\|d_k\|^2)$, where c_3 is some positive constant. Nevertheless, it is difficult to analyse that the FR method with constant stepsizes may produce an uphill direction since in this case $\|d_k\|$ may increase linearly with k. Comparing the result in [19] for the FR method and Theorem 3.2, which shows the global convergence of the FRSR method using constant stepsizes, we can see that the method of SRs has better convergence properties than the standard conjugate gradient method in the form (4) and (5).

In a recent paper, Pytlak and Tarnawski [18] strengthened the convergence relation (3) by Dai and Yuan [4] to (2) for the PRPSR method with strong Wolfe line searches (or Wolfe line searches and $\alpha_k \leq M$). Since (76)) is true in these cases, the analysis of this paper can be also used to establish the strong convergence result for the PRPSR method. More results of the SR method with the restriction $0 \leq \lambda \leq 1$ can be found in [17,18].

Under Assumption 2.1, we can show that the PRP method gives the weak convergence relation (3) provided that $\alpha_k \in (0, 1/(4L))$ is such that

$$\sum_{k\geq 1} \alpha_k = +\infty. \tag{97}$$

Here we provide a sketch of the proof by (25) and the related results in Section 2. Assume that there is an infinite sequence $\{k_i\}$ such that $\alpha_{k_i} \in [\tau, 1/(4L)]$ for some $\tau > 0$. Then in at least one of the k_ith and $(k_i + 1)$th iterations the objective function can achieve a descent, which is proportional to $\|g_{\hat{i}}\|^2$ where $\hat{i} = k_i$ or $k_i + 1$. As the summation of the achieved descents is finite, we know that the subsequence $\|g_{\hat{i}}\|$ tends to zero. Otherwise, we have that $\lim_{k\to\infty} \alpha_k = 0$. In this case, we will eventually have $g_k^T d_k \approx -\|g_k\|^2$ and $\|d_k\|^2 \approx \|g_k\|^2$ with k. Thus, the second term in the right-hand side of (25) is an infinitesimal of higher order comparing with the first term and the objective function can achieve a descent proportional to $\alpha_k\|g_k\|^2$. Therefore, we have that $\sum_{k\geq 1} \alpha_k\|g_k\|^2 < +\infty$. By this, the condition $\sum_{k\geq 1} \alpha_k = +\infty$ and the contradiction principle, we know that $\liminf_{k\to\infty} \|g_k\| = 0$. Thus in either cases, (3) is true.

As is well known, the steepest descent method is widely used in stochastic approximation [11], where the stepsize is usually chosen to be a tiny constant or satisfy

$$\alpha_k > 0, \quad \sum_{k=1}^{\infty} \alpha_k = +\infty, \quad \sum_{k=1}^{\infty} \alpha_k^2 < +\infty. \tag{98}$$

The results of this paper enable the possible use of the conjugate gradient method in this field. We wonder whether there exist more efficient stochastic approximation algorithms based on the conjugate gradient method.

Acknowledgements

The author thanks the area editor and the anonymous referees very much for their many useful comments about an early version of this paper. This research was supported by the Chinese NSF grants 10571171 and 10831006, the CAS grant kjcx-yw-s7-03, and the Alexander von Humboldt Foundation (grant no. CHN/1112740 STP).

References

[1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact line search,* IMA J. Numer. Anal. 5 (1985), pp. 121–124.
[2] L. Armijo, *Minimization of functions having Lipschitz continuous first partial derivatives*, Pacific J. Math. 16(1) (1966), pp. 1–3.
[3] Y.H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal. 16(2) (1996), pp. 155–164.
[4] Y.H. Dai and Y. Yuan, *Global convergence of the method of shortest residuals*, Numer. Math. 82 (1999), pp. 581–598.
[5] Y.H. Dai and Y. Yuan, *Nonlinear Conjugate Gradient Methods*, Shang Hai Science and Technology Publisher, Beijing, 2000 (in Chinese).
[6] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, H.X. Yin, and Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim. 10(2) (1999), pp. 345–358.
[7] R. Fletcher and C. Reeves, *Function minimization by conjugate gradients*, Comput. J. 7 (1964), pp. 149–154.
[8] J.C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. 2(1) (1992), pp. 21–42.
[9] L. Grippo and S. Lucidi, *A globally convergent version of the Polak–Ribière conjugate gradient method,* Math. Program. 78 (1997), pp. 375–391.
[10] M.R. Hestenes, *Conjugate Direction Methods in Optimization*, Springer-Verlag, New York, Heidelberg, Berlin, 1980.
[11] H.J. Kushner and G.G. Yin, *Stochastic Approximation Algorithms and Applications*, Springer, New York, 1997.
[12] C. Lemaréchal, *An extension of Davidon methods to nondifferentiable problems*, Math. Program. Study 3 (1975), pp. 95–109.
[13] E. Polak and G. Ribière, *Note sur la convergence de directions conjugées*, Rev. Francaise Informat Recherche Opertionelle, 3e Année 16 (1969), pp. 35–43.
[14] B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comp. Math. Math. Phys. 9 (1969), pp. 94–112.
[15] M.J.D. Powell, *Restart procedures for the conjugate gradient method*, Math. Program. 2 (1977), pp. 241–254.
[16] M.J.D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, Lecture Notes in Mathematics 1066, Springer-Verlag, Berlin, 1984, pp. 122–141.
[17] R. Pytlak, *On the convergence of conjugate gradient algorithms*, IMA J. Numer. Anal. 14 (1994), pp. 443–460.
[18] R. Pytlak and T. Tarnawski, *On the method of shortest residuals*, J. Optim. Theory Appl. 133 (2007), pp. 99–110.
[19] J. Sun and J. Zhang, *Conjugate gradient methods without line search*, Ann. Oper. Res. 103 (2001), pp. 161–173.
[20] P. Wolfe, *A method of conjugate subgradients for minimizing nondifferentiable functions*, Math. Program. Study 3 (1975), pp. 145–173.