Filomat 30:11 (2016), 3083–3100
DOI 10.2298/FIL1611083D

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

New Hybrid Conjugate Gradient Method as a Convex Combination of FR and PRP Methods

Snežana S. Djordjević[a]

[a] *College for Textile, Leskovac, Serbia*

Abstract. We consider a new hybrid conjugate gradient algorithm, which is obtained from the algorithm of Fletcher-Reeves, and the algorithm of Polak-Ribière-Polyak. Numerical comparisons show that the present hybrid conjugate gradient algorithm often behaves better than some known algorithms.

1. Introduction

We consider the nonlinear unconstrained optimization problem

$$\min\{f(x) : x \in \mathbb{R}^n\}, \tag{1.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, bounded from below.

There are many different methods for solving the problem (1.1).

We are interested in conjugate gradient methods, which have low memory requirements and strong local and global convergence properties.

For solving the problem (1.1), we consider the conjugate gradient method, which starts from an initial point $x_0 \in \mathbb{R}^n$ and generates a sequence $\{x_k\} \subset \mathbb{R}^n$ as follows

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

where $t_k > 0$ is a step size, received from the line search, and the directions d_k are given by [2], [4]

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k s_k. \tag{1.3}$$

In the relation (1.3) β_k is the conjugate gradient parameter, $s_k = x_{k+1} - x_k$, $g_k = \nabla f(x_k)$.

Let the norm $\|\cdot\|$ be the Euclidean norm.

The standard Wolfe line search conditions are frequently used in the conjugate gradient methods; these conditions are given by [30], [31]

$$f(x_k + t_k d_k) - f(x_k) \le \delta t_k g_k^T d_k, \tag{1.4}$$

2010 *Mathematics Subject Classification*. 90C30.
Keywords. Hybrid conjugate gradient method; Convex combination; Conjugacy condition.
Received: 31 August 2014; Accepted: 16 September 2014
Communicated by Predrag Stanimirović
Email address: snezanadjordjevic1971@gmail.com (Snežana S. Djordjević)

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k, \tag{1.5}$$

where d_k is a descent direction and $0 < \delta \leq \sigma < 1$.

Strong Wolfe conditions consist of (1.4) and the next stronger version of (1.5)

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k. \tag{1.6}$$

In the generalized Wolfe conditions [12], the absolute value in (1.6) is replaced by a pair of inequalities:

$$\sigma_1 g_k^T d_k \leq g_{k+1}^T d_k \leq -\sigma_2 g_k^T d_k, \; 0 < \delta < \sigma_1 < 1, \; \sigma_2 \geq 0. \tag{1.7}$$

Now, let us denote

$$y_k = g_{k+1} - g_k. \tag{1.8}$$

There are many conjugate gradient methods; a great contribution in this sphere is given by Hager and Zhang [19]. Different conjugate gradient methods correspond to different values of the scalar parameter β_k.

Hybrid conjugate gradient methods combine different conjugate gradient methods to improve the behavior of these methods and to avoid the jamming phenomenon.

In order to choose the parameter β_k for the method in the present paper, we mention the following choices of β_k [2]:

$$\text{Fletcher and Reeves: [17] } \beta_k^{FR} = \frac{||g_{k+1}||^2}{||g_k||^2}; \tag{1.9}$$

$$\text{Dai and Yuan: [11] } \beta_k^{DY} = \frac{||g_{k+1}||^2}{y_k^T s_k}; \tag{1.10}$$

$$\text{Conjugate Descent, proposed by Fletcher: [16] } \beta_k^{CD} = \frac{||g_{k+1}||^2}{-g_k^T s_k}. \tag{1.11}$$

The conjugate gradient methods with the choice of β_k taken in (1.9), (1.10) and (1.11), have strong convergence properties and, in the same time, they may have modest practical performance, due to jamming [2], [3].

On the other hand, methods of Polak-Ribiére [24] and Polyak (PRP) [25], Hestenes and Stiefel (HS) [20] and also Liu and Storey [22] in general may not be convergent, but usually they have better computer performances [2], [3]. The choices of β_k in these methods are, respectively [2]:

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{||g_k||^2}, \tag{1.12}$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T s_k}, \tag{1.13}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-g_k^T s_k}. \tag{1.14}$$

We will use good convergence properties of the first group of methods and, in the same time, good computational performances of the second one; here we want to exploit choices of β_k in (1.9) and (1.12).

Touati-Ahmed and Storey [29] introduced one of the first hybrid conjugate gradient algorithms, calculating the parameter β_k as:

$$\beta_k^{TaS} = \begin{cases} \beta_k^{PRP} = \dfrac{g_{k+1}^T y_k}{\|g_k\|^2}, \text{ if } 0 \le \beta_k^{PRP} \le \beta_k^{FR}, \\ \beta_k^{FR} = \dfrac{\|g_{k+1}\|^2}{\|g_k\|^2}, \text{ otherwise.} \end{cases} \tag{1.15}$$

When the iterations jam, the method of Touati-Ahmed and Storey use PRP computational scheme, having in view that PRP method has a built-in restart feature that directly addresses to jamming. Namely, when s_k is small, then the factor y_k in the formula for β_k^{PRP} tends to zero, β_k^{PRP} becomes smaller and smaller, and the search direction d_{k+1} is very close to the steepest direction $-g_{k+1}$.

Hu and Storey [21] introduced another hybrid conjugate gradient method with the following choice of β_k:

$$\beta_k^{HuS} = \max\{0, \min\{\beta_k^{PRP}, \beta_k^{FR}\}\}. \tag{1.16}$$

Also, if the method of Hu and Storey jam, then PRP method is used.

The first global convergence result for FR method was given by Zoutendijk [33] in 1970. He proved that FR method converges globally when the line search is exact.

The first global convergence result of FR method for an inexact line search was given by Al-Baali [1] in 1985. Under the strong Wolfe conditions with $\sigma < \frac{1}{2}$, he proved that FR method generates sufficient descent directions. In fact, he proved that

$$\frac{1 - 2\sigma + \sigma^{k+1}}{1 - \sigma} \le \frac{-g_k^T d_k}{\|g_k\|^2} \le \frac{1 - \sigma^{k+1}}{1 - \sigma},$$

for all $k \ge 0$. As a consequence, global convergence is established. For $\sigma = \dfrac{1}{2}$, when d_k is a descent direction, this analysis didn't establish sufficient descent.

In [23], the global convergence proof of Al-Baali is extended to the case $\sigma = \frac{1}{2}$.

Dai and Yuan analyzed this further, and showed that in consecutive FR iterations, at least one iteration satisfies the sufficient descent property.

In [26], it is observed that FR method with the exact line search may produce many small steps continuously. Precisely, if a small step is generated away from the solution, the subsequent steps may also be very short.

In that case, PRP method generates a search direction close to $-g_k$ to avoid the propensity of small steps.

In [26] it is proved that, for a nonlinear function, if

(a) s_k tends to zero,

(b) the line search is exact,

(c) Lipschitz condition holds,

then PRP method is globally convergent.

On the other hand, in [27] it is proved, using a threedimensional example, that PRP method with an exact line search, can diverge. So, the assumption that s_k tends to zero, is needed for convergence [19].

When a hybrid conjugate gradient method is built in which PRP method is presented, there is a possibility for some additional assumptions (see, for example, [2]).

In this paper we consider a convex combination of two methods. The first method is the method FR, which has strong convergence properties, and in the same time it may have modest practical performance, due to jamming. The second method is the method PRP, which in general may not be convergent, but it usually has good computer performances.

The paper is organized as follows. In Section 2 we construct a new hybrid method FRPRPCC (Fletcher-Reeves-Polak-Ribière-Polyak-conjugate-condition), using the convex combination of parameters from the

FR method and from the PRP method. In this section we also find the formula for computing the parameter $\theta_k \in [0, 1]$, which is relevant for our method. We also prove that under some assumptions the search direction of our method satisfies the sufficient descent condition. We use the inexact line search in order to determine the stepsize of the method. In Section 3 we present the algorithm FRPRPCC, which is developed in Section 2. In Section 4 we prove the global convergence theorem of the method FRPRPCC, assuming some conditions. Section 5 contains some numerical experiments.

2. A Convex Combination

In this paper we use another combination of PRP and FR methods. We use the following conjugate gradient parameter

$$\beta_k^{hyb} = (1 - \theta_k) \cdot \beta_k^{PRP} + \theta_k \cdot \beta_k^{FR}. \tag{2.1}$$

Hence, the direction d_k is given by:

$$d_0^{hyb} = -g_0, \;\; d_{k+1}^{hyb} = -g_{k+1} + \beta_k^{hyb} s_k. \text{ [2]} \tag{2.2}$$

The parameter θ_k is the scalar parameter to be determined later.

We can see that, if $\theta_k = 0$, then $\beta_k^{hyb} = \beta_k^{PRP}$, and if $\theta_k = 1$, then $\beta_k^{hyb} = \beta_k^{FR}$.

On the other hand, if $0 < \theta_k < 1$, then β_k^{hyb} is a proper convex combination of the parameters β_k^{PRP} and β_k^{FR}.

Theorem 2.1. *If the relations* (2.1) *and* (2.2) *hold, then*

$$d_{k+1}^{hyb} = \theta_k d_{k+1}^{FR} + (1 - \theta_k) d_{k+1}^{PRP}. \tag{2.3}$$

Proof. Having in view the relations (1.9) and (1.12), the relation (2.1) becomes:

$$\beta_k^{hyb} = (1 - \theta_k) \cdot \frac{g_{k+1}^T y_k}{\|g_k\|^2} + \theta_k \cdot \frac{\|g_{k+1}\|^2}{\|g_k\|^2}, \tag{2.4}$$

so, the relation (2.2) becomes

$$d_0^{hyb} = -g_0, \tag{2.5}$$

$$d_{k+1}^{hyb} = -g_{k+1} + (1 - \theta_k) \frac{g_{k+1}^T y_k}{\|g_k\|^2} \cdot s_k + \theta_k \cdot \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \cdot s_k. \tag{2.6}$$

In further consideration of the relation (2.6), we can get

$$d_{k+1}^{hyb} = -(\theta_k g_{k+1} + (1 - \theta_k) g_{k+1}) + \beta_k^{hyb} s_k,$$

$$d_{k+1}^{hyb} = -(\theta_k g_{k+1} + (1 - \theta_k) g_{k+1}) + (1 - \theta_k) \beta_k^{PRP} s_k + \theta_k \beta_k^{FR} s_k.$$

The last relation yields

$$d_{k+1}^{hyb} = \theta_k (-g_{k+1} + \beta_k^{FR} s_k) + (1 - \theta_k)(-g_{k+1} + \beta_k^{PRP} s_k). \tag{2.7}$$

From (2.7) we finally conclude

$$d_{k+1}^{hyb} = \theta_k d_{k+1}^{FR} + (1 - \theta_k) d_{k+1}^{PRP}. \tag{2.8}$$

□

Our way to find θ_k is to make that the conjugacy condition

$$y_k^T d_{k+1}^{hyb} = 0 \tag{2.9}$$

holds.

Multiplying (2.6) by y_k^T from the left and using (2.9), we get

$$-y_k^T g_{k+1} + (1-\theta_k)\frac{g_{k+1}^T y_k}{\|g_k\|^2}(y_k^T s_k) + \theta_k \frac{\|g_{k+1}\|^2}{\|g_k\|^2}(y_k^T s_k) = 0.$$

So,

$$\theta_k \left(\frac{\|g_{k+1}\|^2}{\|g_k\|^2}(y_k^T s_k) - \frac{g_{k+1}^T y_k}{\|g_k\|^2}(y_k^T s_k) \right) = y_k^T g_{k+1} - \frac{g_{k+1}^T y_k}{\|g_k\|^2}(y_k^T s_k),$$

i.e.

$$\theta_k \cdot \frac{g_{k+1}^T g_k}{\|g_k\|^2}(y_k^T s_k) = \frac{(\|g_k\|^2 - y_k^T s_k)(y_k^T g_{k+1})}{\|g_k\|^2}.$$

Finally,

$$\theta_k = \frac{(\|g_k\|^2 - y_k^T s_k)(y_k^T g_{k+1})}{(g_{k+1}^T g_k)(y_k^T s_k)}. \tag{2.10}$$

It is possible that θ_k, calculated as in (2.10), has the values outside the interval $[0,1]$.

So, we fix it:

$$\theta_k = \begin{cases} 0, & \text{if } \frac{(\|g_k\|^2 - y_k^T s_k)(y_k^T g_{k+1})}{(g_{k+1}^T g_k)(y_k^T s_k)} \le 0, \\ \frac{(\|g_k\|^2 - y_k^T s_k)(y_k^T g_{k+1})}{(g_{k+1}^T g_k)(y_k^T s_k)}, & \text{if } 0 < \frac{(\|g_k\|^2 - y_k^T s_k)(y_k^T g_{k+1})}{(g_{k+1}^T g_k)(y_k^T s_k)} < 1, \\ 1, & \text{if } \frac{(\|g_k\|^2 - y_k^T s_k)(y_k^T g_{k+1})}{(g_{k+1}^T g_k)(y_k^T s_k)} \ge 1. \end{cases} \tag{2.11}$$

Either of the following assumptions is often utilized in convergence analysis for CG algorithms.

Boundedness Assumption: The level set $\mathcal{S} = \{x \in \mathbb{R}^n | f(x) \le f(x_0)\}$ is bounded, i.e. there exists a constant $B > 0$, such that

$$\|x\| \le B, \ \text{ for all } x \in \mathcal{S}. \tag{2.12}$$

Lipschitz Assumption: In a neighborhood $\mathcal{N}$ of $\mathcal{S}$ the function f is continuously differentiable and its gradient $\nabla f(x)$ is Lipschitz continuous, i.e. there exists a constant $0 < L < \infty$ such that

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \text{ for all } x,\ y \in \mathcal{N}. \tag{2.13}$$

Under these assumptions, there exists a constant $\Gamma \ge 0$, such that

$$\|\nabla f(x)\| \le \Gamma \tag{2.14}$$

for all $x \in S$ [2].

Theorem 2.2. *Assume that* (2.12) *and* (2.13) *hold and let strong Wolfe conditions* (1.4)-(1.6) *hold with* $\sigma < \frac{1}{2}$. *Also, let* $\{\|s_k\|\}$ *tend to zero, and let there exist some nonnegative constants* η_1, η_2 *such that*

$$\|g_k\|^2 \geq \eta_1 \|s_k\|^2, \tag{2.15}$$

$$\|g_{k+1}\|^2 \leq \eta_2 \|s_k\|. \tag{2.16}$$

Then d_k^{hyb} *satisfies the sufficient descent condition for all k.*

Proof. It holds $d_0 = -g_0$. So, for $k = 0$, it holds $g_0^T d_0 = -\|g_0\|^2$.

Multiplying (2.8) by g_{k+1}^T from the left, we get

$$g_{k+1}^T d_{k+1}^{hyb} = \theta_k g_{k+1}^T d_{k+1}^{FR} + (1-\theta_k) g_{k+1}^T d_{k+1}^{PRP}. \tag{2.17}$$

If $\theta_k = 0$, the relation (2.17) becomes

$$g_{k+1}^T d_{k+1}^{hyb} = g_{k+1}^T d_{k+1}^{PRP}.$$

So, if $\theta_k = 0$, the sufficient descent holds for the hybrid method, if it holds for PRP method. We can prove the sufficient descent for PRP method under the conditions of Theorem 2.2. It holds

$$d_{k+1}^{PRP} = -g_{k+1} + \beta_k^{PRP} s_k. \tag{2.18}$$

Multiplying (2.18) by g_{k+1}^T from the left, we get

$$g_{k+1}^T d_{k+1}^{PRP} = -\|g_{k+1}\|^2 + \beta_k^{PRP} g_{k+1}^T s_k.$$

Using (1.12), we get

$$g_{k+1}^T d_{k+1}^{PRP} = -\|g_{k+1}\|^2 + \frac{(g_{k+1}^T y_k)(g_{k+1}^T s_k)}{\|g_k\|^2}. \tag{2.19}$$

From (2.19), we get

$$g_{k+1}^T d_{k+1}^{PRP} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 \|y_k\| \|s_k\|}{\|g_k\|^2}. \tag{2.20}$$

From Lipschitz condition we have $\|y_k\| \leq L\|s_k\|$, so

$$g_{k+1}^T d_{k+1}^{PRP} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 L \|s_k\|^2}{\|g_k\|^2}. \tag{2.21}$$

But, using (2.16), we get

$$g_{k+1}^T d_{k+1}^{PRP} \leq -\|g_{k+1}\|^2 + \frac{L\eta_2 \|s_k\|^3}{\|g_k\|^2}. \tag{2.22}$$

Using (2.15), we get

$$g_{k+1}^T d_{k+1}^{PRP} \leq -\|g_{k+1}\|^2 + \frac{1}{\eta_1} L\eta_2 \|s_k\|. \tag{2.23}$$

But, because of the assumption $\|s_k\| \to 0$, the second summand in (2.23) tends to zero, so there exists a number $0 < \delta \ll 1$, such that

$$\frac{1}{\eta_1} L\eta_2 \|s_k\| \leq \delta \|g_{k+1}\|^2. \tag{2.24}$$

Now, (2.23) becomes

$$g_{k+1}^T d_{k+1}^{PRP} \le -\|g_{k+1}\|^2 + \delta\|g_{k+1}\|^2, \tag{2.25}$$

i.e.

$$g_{k+1}^T d_{k+1}^{PRP} \le -(1-\delta)\|g_{k+1}\|^2. \tag{2.26}$$

On the other hand, for $\theta_k = 1$, the relation (2.17) becomes

$$g_{k+1}^T d_{k+1}^{hyb} = g_{k+1}^T d_{k+1}^{FR}.$$

But, under the strong Wolfe line search, FR method satisfies the sufficient descent condition [19].
Now, let $0 < \theta_k < 1$. So, we can write $0 < \mu_1 \le \theta_k \le \mu_2 < 1$.
Now, from (2.17), we get

$$g_{k+1}^T d_{k+1}^{hyb} \le \mu_1 g_{k+1}^T d_{k+1}^{FR} + (1-\mu_2) g_{k+1}^T d_{k+1}^{PRP}.$$

We obviously can conclude now that there exists a number $K > 0$, such that

$$g_{k+1}^T d_{k+1}^{hyb} \le -K\|g_{k+1}\|^2. \tag{2.27}$$

□

Recall that in [26], it is shown that for general functions PRP method is globally convergent if $s_k = x_{k+1} - x_k$ tend to zero, i.e. $\|s_k\| \le \|s_{k-1}\|$ is a sufficient condition for convergence.

A descent property is very important for the global convergence of an iterative method, especially if it is the conjugate gradient method [1].

From (2.15) and (2.16), we see that in the aim to realize a sufficient descent property, the gradient g_k must be bounded, i.e., it holds:

$$\eta_1\|s_k\|^2 \le \|g_k\|^2 \le \eta_2\|s_{k-1}\|.$$

If the Powell condition is satisfied, i.e. $\|s_k\|$ tends to zero, then $\|s_k\|^2 \ll \|s_{k-1}\|$ and therefore the norm of gradient can satisfy (2.15) and (2.16) [2].

The conditions (2.15) and (2.16) are used, for example, in [2].

Now we give the corresponding algorithm.

Algorithm 2.1. *(Algorithm FRPRPCC)*

Step 1. Initialization. Select the initial point
$x_0 \in dom(f)$, $\epsilon > 0$ *and select the parameters*
$0 < \rho \le \sigma < 1$. *Set* $k = 0$.
Compute $f(x_k)$, $g_k = \nabla f(x_k)$. *Set* $d_k = -g_k$.
Set the initial guess $t_k = \dfrac{1}{\|g_k\|}$.

Step 2. Test for continuation of iterations. If $\|g_k\| \le \epsilon$ *then STOP.*

Step 3. Line search. Compute $t_k > 0$ *satisfying the strong Wolfe line search conditions (1.4)-(1.6) and update the variables*
$x_{k+1} = x_k + t_k d_k$.
Compute $f(x_{k+1})$, $g_{k+1} = \nabla f(x_{k+1})$, $s_k = x_{k+1} - x_k$,
$y_k = g_{k+1} - g_k$.

Step 4. θ_k *parameter computation. If* $(g_{k+1}^T g_k)(y_k^T s_k) = 0$, *then set* $\theta_k = 0$, *else set* θ_k *as in (2.11).*

Step 5. Compute β_k as in (2.1).

Step 6. Compute $d = -g_{k+1} + \beta_k s_k$.

Step 7. If the restart criterion of Powell

$$|g_{k+1}^T g_k| \geq 0.2\|g_{k+1}\|^2 \tag{2.28}$$

is satisfied, then set $d_{k+1} = -g_{k+1}$, else let $d_{k+1} = d$.

Step 8. Calculate the initial guess

$$t_k = t_{k-1}\|d_{k-1}/\|d_k\|. \tag{2.29}$$

Step 9. Set $k = k + 1$ and go to Step 2.

In every iteration $k \geq 1$ of Algorithm FRPRPCC, the starting value of the step size t_k is calculated as in (2.29); the selection (2.29) is considered for the first time in CONMIN [28] by Shanno and Phua.

Also, this selection is used by Birgin and Martinez in [9], in the package SCG, and in SCALCG by Andrei, [5], [6], [7].

In our algorithm, when the Powell restart condition is satisfied, we restart the algorithm, with the negative gradient $-g_{k+1}$.

3. Convergence Analysis

For the purpose of this section we remind to the next theorem.

Theorem 3.1. *[19] Consider any iterative method of the form (1.2)-(1.3), where d_k satisfies a descent condition $g_k^T d_k < 0$ and t_k satisfies strong Wolfe conditions (1.4)-(1.6). If the Lipschitz condition holds, then either*

$$\liminf_{k\to\infty} \|g_k\| = 0, \tag{3.1}$$

or

$$\sum_{k=1}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.2}$$

Now we give the next theorem.

Theorem 3.2. *Consider the iterative method of the form (1.2), (2.1), (2.2), (2.11). Let all conditions of Theorem 2.2 hold. Then either $g_k = 0$ for some k, or*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.3}$$

Proof. Let $g_k \neq 0$ for all k. Then, we are going to prove (3.3).

Suppose, on the contrary, that there exists a number $c > 0$, such that

$$\|g_k\| \geq c, \text{ for all } k. \tag{3.4}$$

From (2.8), we get

$$\|d_{k+1}^{hyb}\| \leq \|d_{k+1}^{FR}\| + \|d_{k+1}^{PRP}\|. \tag{3.5}$$

Next, it holds

$$\|d_{k+1}^{FR}\| \leq \|g_k\| + |\beta_k^{FR}|\|s_k\|.$$

Further,

$$|\beta_k^{FR}| = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \leq \frac{\Gamma^2}{c^2},$$

where Γ is given by (2.14). Now,

$$\|d_{k+1}^{FR}\| \leq \Gamma + \frac{\Gamma^2}{c^2} \cdot D, \tag{3.6}$$

where D is a diameter of the level set $\mathcal{S}$.

Also,

$$|\beta_k^{PRP}| = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \leq \frac{1}{c^2}\|g_{k+1}\|\|y_k\| \leq \frac{2\Gamma^2}{c^2},$$

because of

$$\|y_k\| = \|g_{k+1} - g_k\| \leq \|g_{k+1}\| + \|g_k\| \leq 2\Gamma.$$

Now,

$$\|d_{k+1}^{PRP}\| \leq \Gamma + \frac{2\Gamma^2}{c^2} \cdot D. \tag{3.7}$$

So, using (3.5), (3.6) and (3.7), we get

$$\|d_{k+1}^{hyb}\| \leq \Gamma + \frac{\Gamma^2}{c^2} \cdot D + \Gamma + \frac{2\Gamma^2}{c^2} \cdot D,$$

i.e.

$$\|d_{k+1}^{hyb}\| \leq 2\Gamma + \frac{3\Gamma^2}{c^2} \cdot D. \tag{3.8}$$

But, now we can get

$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \frac{c^4}{\left(2\Gamma + \frac{3\Gamma^2}{c^2} \cdot D\right)^2},$$

wherefrom

$$\sum_{k=0}^{\infty} \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} = \infty.$$

Using Theorem 3.1, we conclude that this is a contradiction. So, we finish the proof. □

4. Numerical Experiments

In this section we present the computational performance of a Mathematica implementation of FRPRPCC algorithm on a set of unconstrained optimization test problems from [8]. Each problem is tested for a number of variables: $n = 50$, $n = 60$, $n = 70$, $n = 90$, $n = 100$, $n = 120$. We present comparisons with CCOMB from [2], HYBRID from [3], denoted on the pictures by HSDY, then with the algorithm of Touatti-Ahmed and Storey (ToAhS) from [29], the algorithm of Hu and Storey (HuS) from [21] and the algorithm (GN) of Gilbert and Nocedal from [18], using the performance profiles of Dolan and Moré [15]. The criterion of comparation is CPU time. The stopping criterion of all algorithms is $\|g_k\| < 10^{-6}$. From the figures below, we can conclude that FRPRPCC algorithm behaves similar to or better than CCOMB, HYBRID, ToAhS, HuS and GN.

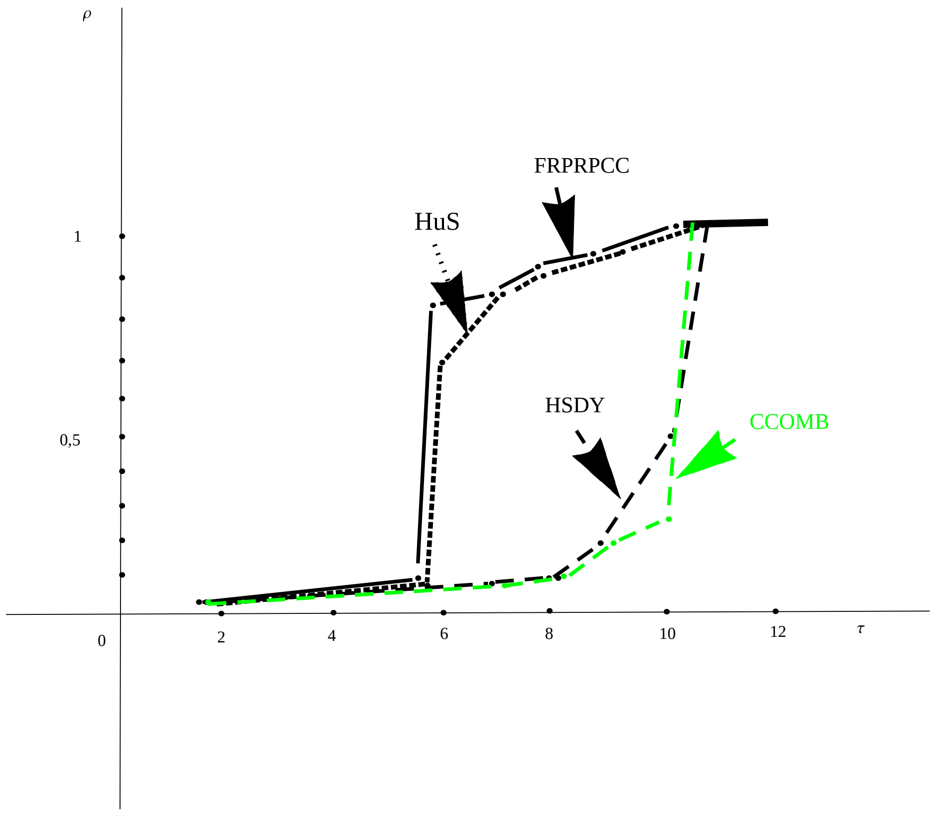

Figure 1. (n=50)

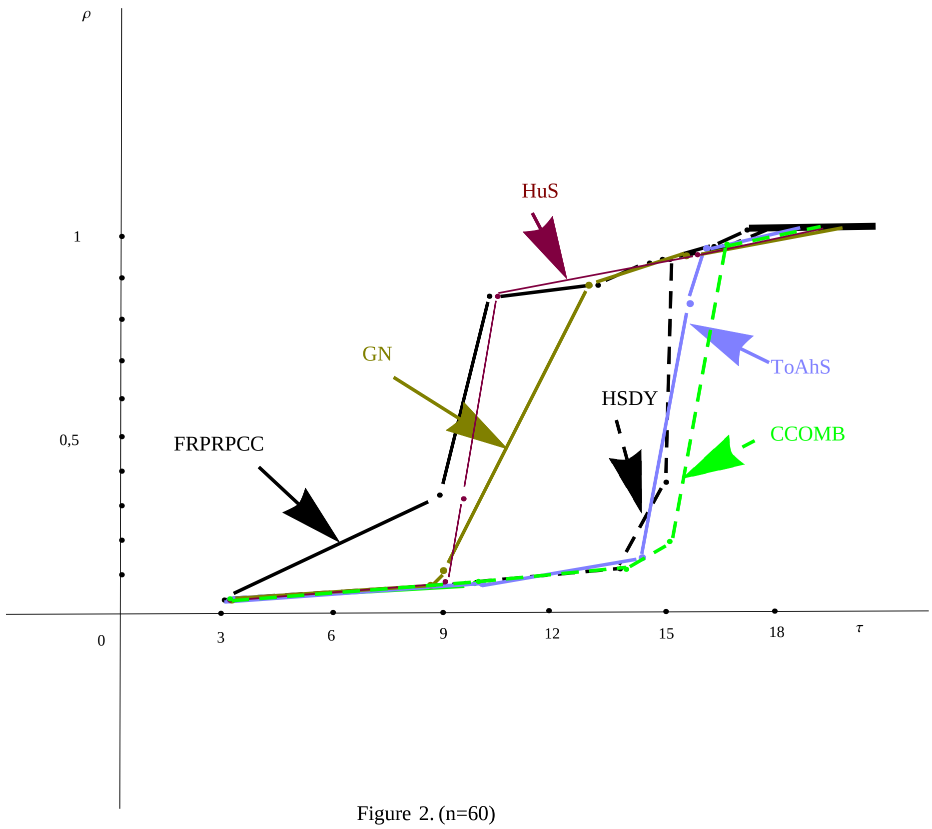

Figure 2. (n=60)

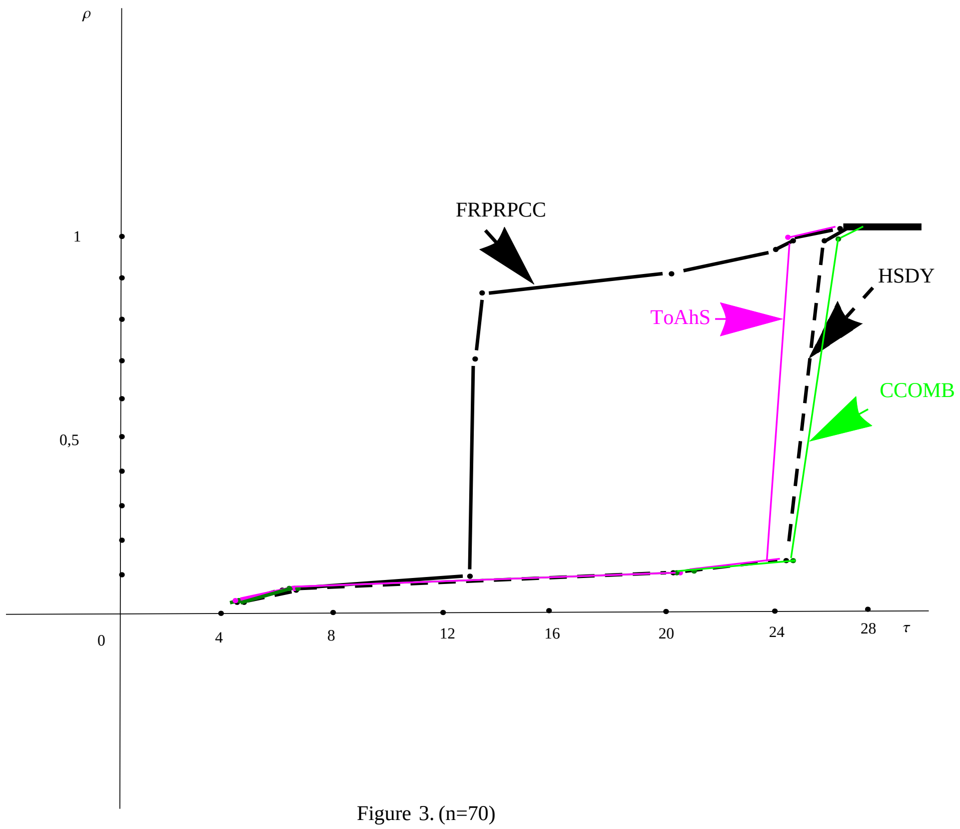

Figure 3. (n=70)

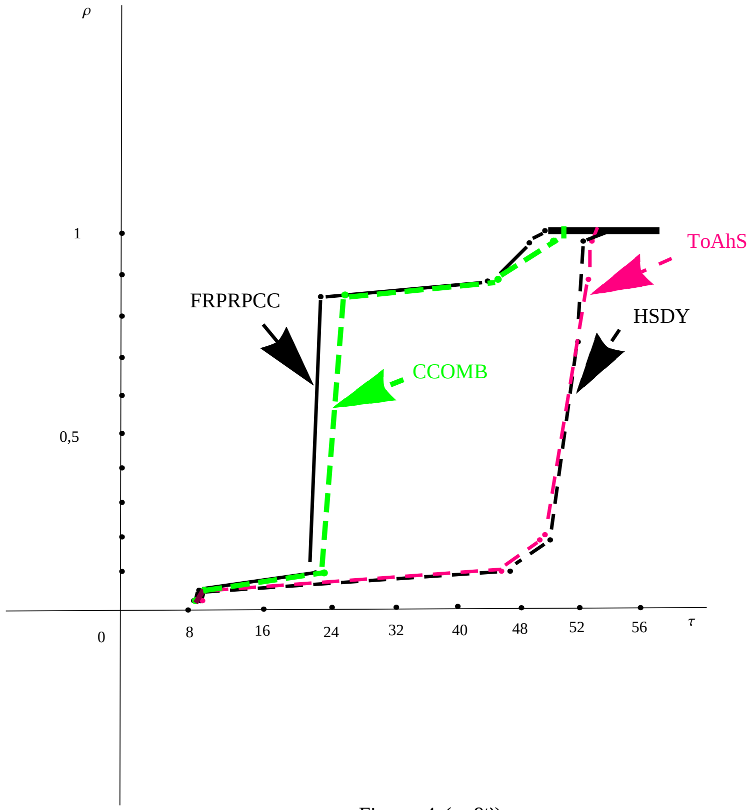

Figure 4. (n=90)

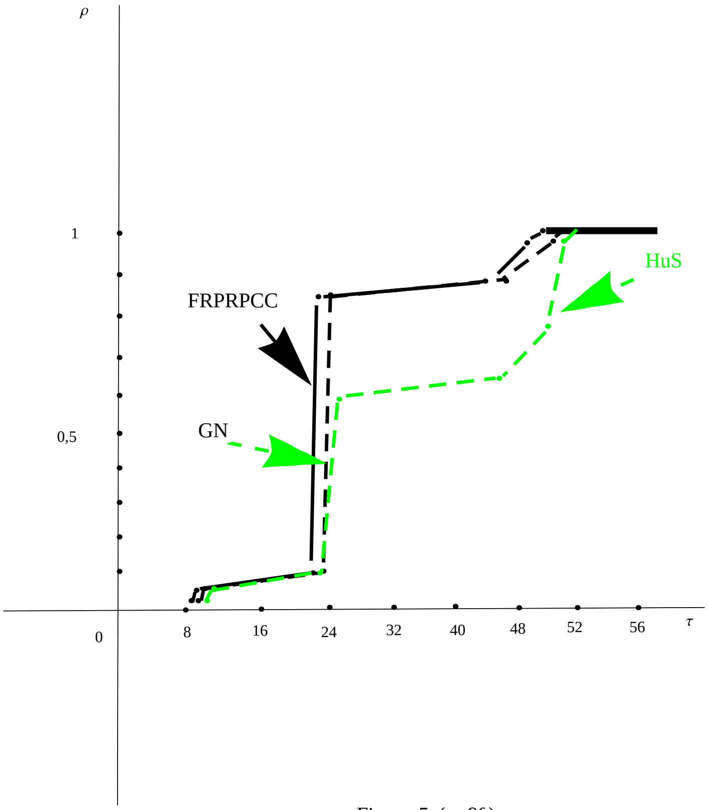

Figure 5. (n=90)

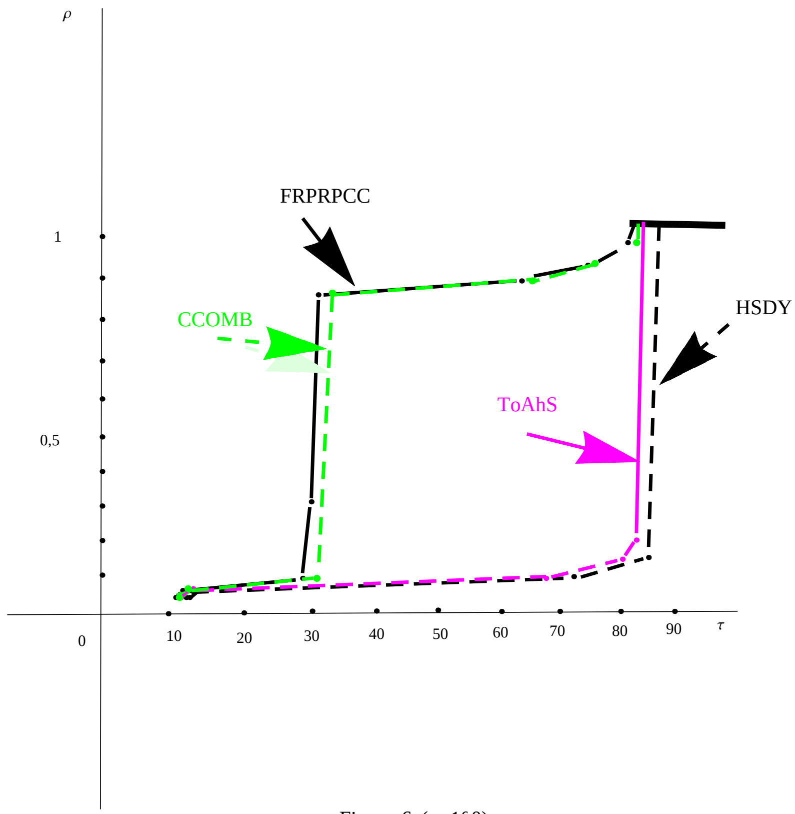

Figure 6. (n=1C0)

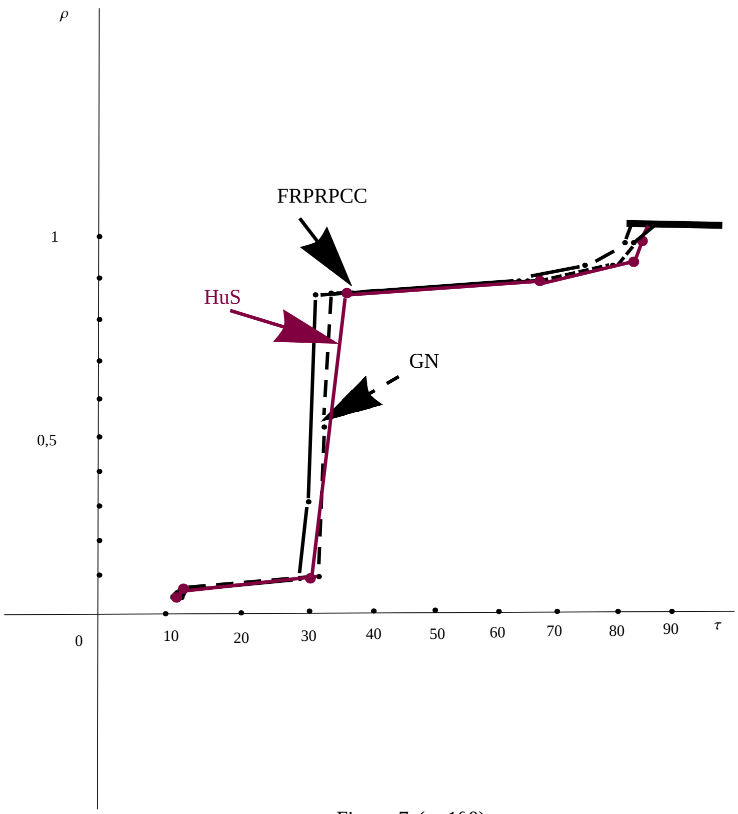

Figure 7. (n=100)

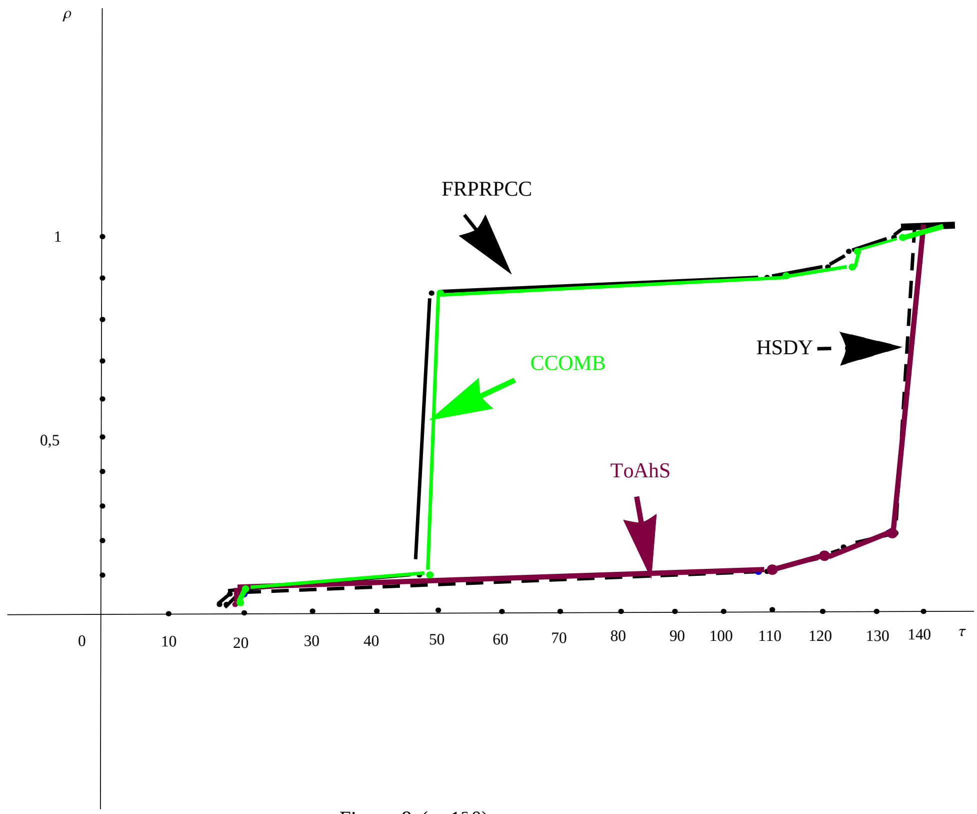

Figure 8. (n=120)

References

[1] X.-M. An, D.-H. Li, Y. Xiao, *Sufficient descent directions in unconstrained optimization*, Computational Optimization and Applications, 48, 3 (2011) 515–532.

[2] N. Andrei, *New hybrid conjugate gradient algorithms for unconstrained optimization*, Encyclopedia of Optimization (2009) 2560–2571.

[3] N. Andrei, *A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of Hestenes-Stiefel and Dai-Yuan*, Studies in Informatics and Control, 17, 1 (2008) 55–70.

[4] N. Andrei, *Another hybrid conjugate gradient algorithm for unconstrained optimization*, Numerical Algorithms, 47, 2 (2008) 143–156.

[5] N. Andrei, *Scaled conjugate gradient algorithms for unconstrained optimization*, Computational Optimization and Applications, 38, 3 (2007) 401–416.

[6] N. Andrei, *Scaled memoryless BFGS preconditioned conjugate gradient algorithm for unconstrained optimization*, Optimization Methods and Software, 22, 4 (2007) 561–571.

[7] N. Andrei, *A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization*, Appl. Math. Letters, 20, 6, (2007) 645–650.
[8] N. Andrei, *Test functions for unconstrained optimization*, http://www.ici.ro/ camo/neculai/HYBRID/evalfg.for.
[9] E. Birgin, J. M. Martinez, *A spectral conjugate gradient method for unconstrained optimization*, Appl. Math. Optim. 43 (2001) 117–128.
[10] Y.H. Dai, *A family of hybrid conjugate gradient methods for unconstrained optimization*, Mathematics of Computation, 72, 243 (2003) 1317–1328.
[11] Y.H. Dai, Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim., 10 (1999), 177-182.
[12] Y.H. Dai, Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal., 16 (1996) 155–164.
[13] Y.H. Dai, J.Y. Han, G.H. Liu, D.F. Sun, X. Yin, Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM Journal on Optimization, 10 (1999) 348–358.
[14] Y.H. Dai, L.Z. Liao, Duan Li, *On restart procedures for the conjugate gradient method*, Numerical Algorithms, 35 (2004) 249–260.
[15] E.D. Dolan, J. J. Moré, *Benchmarking optimization software with performance profiles*, Math. Programming, 91 (2002) 201–213.
[16] R. Fletcher, *Practical Methods of Optimization*, vol. 1: Unconstrained Optimization, John Wiley and Sons, New York, 1987.
[17] R. Fletcher, C. Reeves, *Function minimization by conjugate gradients*, Comput. J., 7 (1964) 149–154.
[18] J.C. Gilbert, J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. 2 (1992) 21–42.
[19] W.W. Hager, H. Zhang, *A survey of nonlinear conjugate gradient methods*, Pacific journal of Optimization, 2 (2006) 35–58.
[20] M.R. Hestenes, E.L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952) 409–436.
[21] Y.F. Hu, C. Storey, *Global convergence result for conjugate gradient methods*, J. Optim. Theory Appl., 71 (1991) 399–405.
[22] Y. Liu, C. Storey, *Efficient generalized conjugate gradient algorithms, Part 1: Theory*, JOTA, 69 (1991) 129–137.
[23] G.H. Liu, J.Y. Han, H.X. Yin, *Global convergence of the Fletcher-Reeves algorithm with an inexact line search*, Appl. Math. J. Chinese Univ. Ser. B, 10 (1995) 75–82.
[24] E. Polak, G. Ribiére, *Note sur la convergence de directions conjuguée*, Rev. Francaise Informat Recherche Operationelle, 3e Année 16 (1969), 35-43.
[25] B.T. Polyak, *The conjugate gradient method in extreme problems*, USSR Comp. Math. Math. Phys., 9 (1969) 94–112.
[26] M.J.D. Powell, *Restart procedures of the conjugate gradient method*, Mathematical Programming, 2 (1977) 241–254.
[27] M.J.D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, Numerical Analysis (Dundee, 1983), Lecture Notes in Mathematics, Vol. 1066, Springer-Verlag, Berlin, (1984) 122–141.
[28] D. F. Shanno, K. H. Phua, *Algorithm 500, Minimization of uncon- strained multivariate functions*, ACM Trans. on Math. Soft. 2 (1976), 87-94.
[29] D. Touati-Ahmed, C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optim. Theory Appl., 64 (1990) 379–397.
[30] P. Wolfe, *Convergence conditions for ascent methods*, SIAM Review, 11 (1969) 226–235.
[31] P. Wolfe, *Convergence conditions for ascent methods. II: Some corrections*, SIAM Review, 11 (1969) 226–235.
[32] X. Li, X. Zhao, *A hybrid conjugate gradient method for optimization problems*, Natural Science, Vol. 3, No. 1, 85-90 (2011), openly accessible at `http://www.scirp.org/journal/NS/`.
[33] G. Zoutendijk, *Nonlinear Programming, Computational Methods*, in Integer and Nonlinear Programming, J. Abadie, ed., North-Holland, Amsterdam, (1970) 37–86.