Conjugate Gradient Methods with Armijo-type Line Searches *

Y. H. Dai

State Key Laboratory of Scientific and Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering computing,
Chinese Academy of Sciences, P. O. Box 2719, Beijing 100080, China
Email-address: dyh@lsec.cc.ac.cn

Abstract

Two Armijo-type line searches are proposed in this paper for nonlinear conjugate gradient methods. Under these line searches, global convergence results are established for several famous conjugate gradient methods, including the Fletcher-Reeves method, the Polak-Ribiére-Polyak method, and the conjugate descent method.

Key words: unconstrained optimization, conjugate gradient method, line search, global convergence.

AMS classification: 65k, 90c.

1. Introduction

To minimize a smooth function of n variables,

$$\min\ f(x), \tag{1.1}$$

we consider the conjugate gradient method in the form

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

$$d_k = \begin{cases} -g_k, & \text{for } k=1; \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{1.3}$$

where $g_k = \nabla f(x_k)$, α_k is a stepsize obtained by a line search and β_k is a scalar. The scalar β_k is so chosen that (1.2)–(1.3) reduces to the linear conjugate gradient method when f is a strictly convex quadratic and when α_k is the exact one-dimensional minimizer. Several famous formulas for β_k are called the Fletcher-Reeves (FR), Polak-Ribiére-Polyak (PRP), Hestenes-Stiefel (HS), and conjugate descent (CD) formulas (see [9; 16, 17; 12; 8]), and are given by

$$\beta_k^{FR} = \|g_k\|^2 / \|g_{k-1}\|^2, \tag{1.4}$$

*Research partially supported by Chinese NSF grant 19801033.

$$\beta_k^{PRP} = g_k^T(g_k - g_{k-1})/\|g_{k-1}\|^2, \tag{1.5}$$

$$\beta_k^{HS} = g_k^T(g_k - g_{k-1})/d_{k-1}^T(g_k - g_{k-1}), \tag{1.6}$$

$$\beta_k^{CD} = -\|g_k\|^2/d_{k-1}^T g_{k-1}, \tag{1.7}$$

respectively, where $\|\cdot\|$ means the Euclidean norm. The conjugate gradient method is available for large-scale unconstrained optimization because its storage is relatively small. Numerical results in [15] showed that, if f is easy to be computed and if its dimension n is very large, the conjugate gradient method is still the best choice for solving (1.1).

In practical computations, it is generally believed that the conjugate gradient method prefers over relatively exact line searches. As a result, in the already-existing convergence analyses and implementations of the conjugate gradient method, the strong Wolfe conditions, namely,

$$f(x_k + \alpha_k d_k) - f(x_k) \le \delta \alpha_k g_k^T d_k, \tag{1.8}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k, \tag{1.9}$$

where $0 < \delta < \sigma < 1$, are often imposed on the line search. For example, see [1, 14, 6, 11, 21].

However, recent studies show that one can analyze the conjugate gradient method under several practical line searches other than the strong Wolfe line search, and good numerical results can be obtained. For example, the nonlinear conjugate gradient method in [4] converges globally provided that the stepsize satisfies the standard Wolfe conditions, namely, (1.8) and

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{1.10}$$

where also $0 < \delta < \sigma < 1$. Based on the new method, a hybrid and efficient algorithm was also presented in [4], which performs much better than the PRP method. Another example is that, [12] suggested an Armijo-type line search which can ensure the global convergence of the PRP method, and [16] noted that such a line search leads to a marginal improvement over the standard implementation of the PRP method. In addition, in [8] and [22], the FR method is shown to be globally convergent under the generalized Wolfe line search and the generalized Armijo line search, respectively.

In this paper, we will propose two Armijo-type searches, both of which are based on the standard Armijo line search (see [2]). The first is designed for any method (1.2)–(1.3) with β_k satisfying $\beta_k \ge 0$ and ensures a descent direction at every iteration. Under this line search, the FR method, the CD method, and the PRP method with nonnegative values of β_k are proved to converge globally. But such a line search can not ensure the global convergence of the original PRP method. For this, another Armijo-type line search is proposed for the original PRP method. The PRP method is proved to be globally convergent with this line search. Some discussion is also made in this paper.

2. Main results

In this section, we always assume that $\|g_k\| \neq 0$ for all k, for otherwise a stationary point has been obtained. The following Armijo-type line search is designed for any method (1.2)–(1.3) with β_k satisfying $\beta_k \geq 0$.

Line Search (A). *Given constants $\lambda \in (0,1)$, $\delta > 0$ and $\sigma_1 \in (0,1]$. Determine the smallest integer $m \geq 0$ such that, if one defines*

$$\alpha_k = \lambda^m, \tag{2.1}$$

then the vectors x_{k+1} and d_{k+1} given by (1.2)–(1.3) satisfy

$$f_{k+1} - f_k \leq \delta \alpha_k g_k^T d_k \tag{2.2}$$

and

$$g_{k+1}^T d_{k+1} \leq -\sigma_1 \|g_{k+1}\|^2. \tag{2.3}$$

The condition (2.2) indicates that the objective function has certain reduction at the k-th iteration, whereas (2.3) means that the sufficient descent condition, namely, $-g_k^T d_k \geq c\|g_k\|^2$ holds for $k = k+1$ and some constant $c > 0$. To show that there must exist such stepsizes α_k for which the conditions (2.2) and (2.3) holds, we give a lemma under the following assumption on f.

Assumption 2.1 *(i) f is bounded below on the level set $\mathcal{L} = \{x \in \Re^n : f(x) \leq f(x_1)\}$; (ii) In some neighborhood $\mathcal{N}$ of $\mathcal{L}$, f is differentiable and its gradient g is Lipschitz continuous, namely, there exists a constant $L > 0$ such that*

$$\|g(x) - g(\tilde{x})\| \leq L\|x - \tilde{x}\|, \qquad \text{for any } x,\ \tilde{x} \in \mathcal{N}. \tag{2.4}$$

Lemma 2.2 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider any method (1.2)–(1.3), where $\beta_k \geq 0$ and where the stepsize α_k is computed by line search (A). Then for every k, line search (A) determines a positive stepsize $\alpha_k > 0$ and produces a descent direction d_{k+1}. Further, we have that*

$$\alpha_k \geq \min\{1, c_1 \frac{|g_k^T d_k|}{\|d_k\|^2}\}, \tag{2.5}$$

where c_1 is a positive constant.

Proof Since $d_1 = -g_1$, d_1 is a descent direction. Suppose that d_k is such that

$$g_k^T d_k < 0. \tag{2.6}$$

For any $\tilde{\alpha}_k$, define $\tilde{x}_{k+1} = x_k + \tilde{\alpha}_k d_k$ and $\tilde{d}_{k+1} = \tilde{g}_{k+1} + \beta_{k+1} d_k$. Because

$$\tilde{g}_{k+1}^T \tilde{d}_{k+1} = -\|\tilde{g}_{k+1}\|^2 + \beta_{k+1}[g_k^T d_k + (\tilde{g}_{k+1} - g_k)^T d_k], \tag{2.7}$$

for any $\sigma_1 \in (0,1]$, we have by (2.4) that

$$\tilde{g}_{k+1}^T \tilde{d}_{k+1} \leq -\sigma_1 \|\tilde{g}_{k+1}\|^2, \quad \text{for all } \tilde{\alpha} \in (0, \tfrac{|g_k^T d_k|}{L\|d_k\|^2}). \tag{2.8}$$

On the other hand, by the mean value theorem and (2.4),

$$\begin{aligned} f(\tilde{x}_{k+1}) - f_k &= \int_0^1 g(x_k + t\tilde{\alpha}_k d_k)^T(\tilde{\alpha}_k d_k)dt \\ &= \tilde{\alpha}_k g_k^T d_k + \int_0^1 [g(x_k + t\tilde{\alpha}_k d_k) - g_k]^T(\tilde{\alpha}_k d_k)dt \\ &\leq \tilde{\alpha}_k g_k^T d_k + \frac{1}{2}L\tilde{\alpha}_k^2\|d_k\|^2. \end{aligned} \tag{2.9}$$

Then we have that

$$f(\tilde{x}_{k+1}) - f_k \leq -\delta\tilde{\alpha}_k g_k^T d_k, \quad \text{for all } \tilde{\alpha}_k \in (0, \tfrac{2(1-\delta)}{L}\tfrac{|g_k^T d_k|}{\|d_k\|^2}). \tag{2.10}$$

Therefore due to (2.8) and (2.10), line search (A) determines a positive stepsize $\alpha_k > 0$ and further, (2.5) holds with constant

$$c_1 = \min\{\frac{\lambda}{L}, \frac{2\lambda(1-\delta)}{L}\}. \tag{2.11}$$

Meanwhile, (2.3) shows that (2.6) holds for $k+1$. By induction, we know that this lemma is true. □

Note that (2.3) is the so-called sufficient descent condition (for example, see [11]). In the case that the sufficient condition holds, it is obvious that

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|g_k\|^2} = \infty. \tag{2.12}$$

By the above relation and Lemma 2.2, we can have the following theorem. The proof of the theorem is quite similar to the proof of Theorem 3.2 in [3] and so is omitted.

Theorem 2.3 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the FR method (1.2), (1.3) and (1.4), where the stepsize α_k is obtained by line search (A). Then each d_k is a descent search direction. Further, we have that*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{2.13}$$

We also formulate the following assumption on the objective function.

Assumption 2.4 *The level set $\mathcal{L} = \{x \in R^n : f(x) \leq f(x_1)\}$ is bounded.*

If f satisfies Assumptions 2.1 and 2.4, there exists a constant $\overline{\gamma} > 0$ such that

$$\|g(x)\| \leq \overline{\gamma}, \quad \text{for all } x \in \mathcal{L}. \tag{2.14}$$

Consider the FR method using line search (A), where (2.3) is replaced by

$$g_{k+1}^T d_{k+1} < 0. \tag{2.15}$$

Since (2.15) is weaker than (2.3), it is easy to see that Lemma 2.2 also holds for such a modified line search. In this case, it follows from (2.14) that

$$\sum_{k\geq 1} \frac{1}{\|g_k\|^2} = \infty. \tag{2.16}$$

Therefore similarly to Theorem 3.4 in [3], we can have the following result.

Theorem 2.5 *Suppose that x_1 is a starting point for which Assumptions 2.1 and 2.4 hold. Consider the FR method (1.2), (1.3) and (1.4), where the stepsize α_k is obtained by line search (A) with (2.3) replaced by (2.15). Then each d_k is a descent direction. Further, we have that* $\liminf \|g_k\| = 0$.

Now we turn to the CD method. In the case that the parameter σ_1 in (2.3) is equal to one, it follows from (1.4), (1.7) and (2.3) with k replaced by $k+1$ that

$$0 \le \beta_k^{CD} \le \beta_k^{FR}. \tag{2.17}$$

Since Theorem 3.2 in [3] is also true for a family of methods (1.2)-(1.3) with $0 \le \beta_k \le \beta_k^{FR}$ (see [21] for the introduction of the family), we clearly have the following convergence result for the CD method.

Theorem 2.6 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the CD method (1.2), (1.3) and (1.7), where the stepsize α_k is obtained by line search (A) with $\sigma_1 = 1$. Then each d_k is a descent search direction. Further, we have that* $\liminf \|g_k\| = 0$.

For the PRP method using line search (A), similarly to the results in [11], it can be shown that a suggestion in [20] of setting

$$\beta_k = \max\{0, \beta_k^{PRP}\} \tag{2.18}$$

results in global convergence.

Theorem 2.7 *Suppose that x_1 is a starting point for which Assumptions 2.1 and 2.4 hold. Consider the modified PRP method (1.2), (1.3) and (2.18), where the stepsize α_k is obtained by line search (A). Then each d_k is a descent search direction. Further, we have that* $\liminf \|g_k\| = 0$.

Proof Assume that there exists an infinite subsequence $\{k_i\}$ such that

$$\alpha_{k_i} = 1, \quad \text{for all } i \ge 1. \tag{2.19}$$

In this case, since the limit of f_k exists, it follows from (2.2) and (2.19) that

$$\lim_{i\to\infty} g_{k_i}^T d_{k_i} = 0. \tag{2.20}$$

Thus by (2.20) and (2.3), we get that

$$\lim_{i\to\infty} \|g_{k_i}\| = 0. \tag{2.21}$$

Otherwise, by (2.5), we can assume that

$$\alpha_k \ge c_1 \frac{|g_k^T d_k|}{\|d_k\|^2} \tag{2.22}$$

holds for all large k. Summing (2.2) and noting that f is bounded below, we get that the Zoutendijk condition holds:

$$\sum_{k\geq 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{2.23}$$

Thus similarly to [11], we can also prove (2.13). This completes our proof. $\Box$

Since the values of β_k^{PRP} may be negative even if f is strongly convex (see [11]), one would be interested in the convergence properties of the original PRP method. Note that by (1.5) and (2.4), we have that

$$|\beta_k^{PRP}| \leq \frac{|g_k^T(g_{k+1} - g_k)|}{\|g_k\|^2} + \frac{\|g_{k+1} - g_k\|^2}{\|g_k\|^2} \leq \frac{L\alpha_k\|d_k\|}{\|g_k\|} + \frac{L^2\alpha_k^2\|d_k\|^2}{\|g_k\|^2}. \tag{2.24}$$

This relation enables us to prove that line search (A) is adaptable for the original PRP method provided that the parameter σ_1 in (2.3) is strictly less than one. However, we can test that (2.2) and (2.3) are always satisfied in the $n = 2, m = 8$ example in [19], and hence such an example can also be used to show that the PRP method using line search (A) needs not converge globally. One reason for which Powell's counter-example exists is that the quantity $\|d_k\|$ generated by the PRP method may increases exponentially. If $\|d_k\|^2$ increases at most linearly, namely, $\|d_k\|^2 \leq \gamma_1 + \gamma_2 k$, [5] proved that any descent method (1.2)–(1.3) with the strong Wolfe conditions (1.8)–(1.9) gives the convergence relation (2.13).

For the above reason, instead of (2.3) we use condition

$$0 \neq g_{k+1}^T d_{k+1} \leq -\sigma_2\|d_{k+1}\|^2, \tag{2.25}$$

where $0 < \sigma_2 < 1$, and consider the following Armijo-type line search for the PRP method.

Line Search (B). *Given $\lambda \in (0,1)$, $\delta > 0$ and $\sigma_2 \in (0,1)$. Determine the smallest integer $m \geq 0$ such that, if α_k is given in (2.1), then the vectors x_{k+1} and d_{k+1} given by (1.2), (1.3) and (1.5) satisfy (2.2) and (2.25).*

The condition (2.25) gives a restriction to the quantity $\|d_k\|^2$ and hence makes it possible for us to consider the global convergence of the PRP method under Assumption 2.1 on f.

Lemma 2.8 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the PRP method (1.2), (1.3) and (1.5), where the stepsize α_k is obtained by line search (B). Then for every k, line search (B) determines a positive stepsize $\alpha_k > 0$. Further, we have that*

$$\alpha_k \geq c_2, \tag{2.26}$$

where c_2 is a positive constant.

Proof Since $d_1 = -g_1$, we have that

$$g_k^T d_k \le -\sigma_2 \|d_k\|^2 \tag{2.27}$$

for $k = 1$. Suppose that (2.27) holds for some k. Using the fact that $a^T b \le \|a\| \, \|b\|$ for any vector a, $b \in R^n$, we can get from (2.27) that

$$\|d_k\| \le \sigma_2^{-1} \|g_k\|. \tag{2.28}$$

For any $\tilde{\alpha}_k$, define $\tilde{x}_{k+1} = x_k + \tilde{\alpha}_k d_k$ and $\tilde{d}_{k+1} = -\tilde{g}_{k+1} + \beta_{k+1} d_k$. Noting that (2.10) is still available, it follows from this and (2.27) that

$$f(\tilde{x}_{k+1}) - f_k \le \delta \tilde{\alpha}_k g_k^T d_k, \quad \text{for all } \tilde{\alpha}_k \in (0, c_3), \tag{2.29}$$

where $c_3 = 2\sigma_2(1-\delta)/L$. By the definition of $\tilde{d}_{k+1}$ and (2.4), we have

$$\|\tilde{d}_{k+1}\| \le \|\tilde{g}_{k+1}\| \left[1 + \frac{L \tilde{\alpha}_k \|d_k\|^2}{\|g_k\|^2} \right] \tag{2.30}$$

and

$$\begin{aligned} -\tilde{g}_{k+1}^T \tilde{d}_{k+1} &= \|\tilde{g}_{k+1}\|^2 \left[1 - \frac{\tilde{g}_{k+1}^T (\tilde{g}_{k+1} - g_k)}{\|g_k\|^2} \frac{d_k^T \tilde{g}_{k+1}}{\|\tilde{g}_{k+1}\|^2} \right] \\ &\ge \|\tilde{g}_{k+1}\|^2 \left[1 - \frac{L \tilde{\alpha}_k \|d_k\|^2}{\|g_k\|^2} \right]. \end{aligned} \tag{2.31}$$

Thus if we choose a positive constant $c_4 > 0$ such that

$$1 - c_4 \ge \sigma_2 (1 + c_4)^2, \tag{2.32}$$

due to (2.30) and (2.31), we have that

$$\tilde{g}_{k+1}^T \tilde{d}_{k+1} \le -\sigma_2 \|\tilde{d}_{k+1}\|^2, \quad \text{for all } \tilde{\alpha}_k \in (0, \tfrac{c_4 \|g_k\|^2}{L \|d_k\|^2}). \tag{2.33}$$

From (2.29), (2.33) and (2.28), we know that line search (B) determines a positive stepsize $\alpha_k > 0$ and further, (2.26) holds with

$$c_2 = \min\{1, \lambda c_3, \frac{\lambda c_4 \sigma_2^2}{L}\}. \tag{2.34}$$

(2.25) implies that (2.27) holds for $k+1$. By induction, we conclude the truth of the lemma. □

By the above lemma, we can prove the following strong convergence result for the PRP method. Here it should be noted that [12] has given some strong convergence result for the PRP method under another Armijo-type line search, which is based on the relation

$$f_{k+1} \le f_k - \gamma \alpha_k^2 \|d_k\|^2 \quad (\gamma < 1), \tag{2.35}$$

whereas the line search used in Theorem 2.9 is based on (2.2). In addition, our proof to Theorem 2.9 is a direct one, and hence is much simpler than that in [12].

Theorem 2.9 *Suppose that x_1 is a starting point for which Assumption 2.1 holds. Consider the PRP method (1.2), (1.3) and (1.5), where the stepsize α_k is obtained by line search (B). Then we have that*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{2.36}$$

Proof Since the limit of f_k exists, it follows from (2.2) and (2.26) that

$$\lim_{k\to\infty} g_k^T d_k = 0, \tag{2.37}$$

which with (2.25) indicates that

$$\lim_{k\to\infty} \|d_k\| = 0. \tag{2.38}$$

Using relations (1.3), (1.5), (2.4) and (2.28) and the fact that $\alpha_k \le 1$, we have for all $k \ge 1$,

$$\begin{aligned} \|g_{k+1}\| &\le \|d_{k+1}\| + \frac{|g_{k+1}^T(g_{k+1}-g_k)|}{\|g_k\|^2}\|d_k\| \\ &\le \|d_{k+1}\| + \frac{\alpha_k\|d_k\|^2}{\|g_k\|^2}\|g_{k+1}\| \\ &\le \|d_{k+1}\| + \sigma_2^2\|g_{k+1}\|. \end{aligned} \tag{2.39}$$

Therefore

$$\|g_{k+1}\| \le (1-\sigma_2^2)^{-1}\|d_{k+1}\|, \tag{2.40}$$

which and (2.38) implies that (2.36) holds. □

3. Some Discussions

We have proposed two Armijo-type line searches based on the standard Armijo line search and established convergence results for several famous conjugate gradient methods, including the FR, PRP and CD methods. We do not know whether there are similar results for other conjugate gradient methods, especially for the HS method. To the best of my knowledge, there is no any kind of line search yet which can ensure the convergence of the original HS method for generally unconstrained optimization problem.

As compared with (2.2), condition (2.35) enforces a greater reduction of f, for large values of $\alpha_k\|d_k\|$, while it becomes more tolerant when $\alpha_k\|d_k\|$ is small. In practical computations, to make the stepsize α_k easily accepted, it may be very useful to use the following *ad hoc* condition

$$f(x_k+\alpha_k d_k) - f_k \le \max\{\delta\alpha_k g_k^T d_k, -\gamma\alpha_k^2\|d_k\|^2\}, \tag{3.1}$$

where δ and $\gamma < 1$ are constants in $(0,1)$. We wonder whether similar convergence results could be established for (3.1).

Finally, we would like to make a numerical remark as follows. Currently, the studies on conjugate gradient methods with Armijo-type line searches still remain in theory, for example see [12] and [22]. We do not know yet whether the use of Armijo-type line searches in practical conjugate gradient algorithms is worthwhile. This topic still remains to be examined.

Acknowledgements. The author would like to thank the referees for their valuable suggestions on the early version of this paper.

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 Journal of Mathematics, 16:1 (1966), pp. 1-3

[3] Y. H. Dai, *Further Insight Into the Convergence of the Fletcher-Reeves Method*, Sciences in China (Series A), 42:9 (1999) pp. 905-916

[4] Y. H. Dai and Y. Yuan, *A Nonlinear Conjugate Gradient Method with A Strong Global Convergence Property*, SIAM Journal on Optimization, 10:1 (1999) pp. 177-182

[5] 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 Journal on Optimization, 10:2 (1999) pp. 345-358

[6] Y. H. Dai and Y. Yuan, *Convergence properties of the Fletcher-Reeves method*, IMA J. Numer. Anal., 16:2 (1996), pp. 155-164

[7] Y. H. Dai and Y. Yuan, *Convergence Properties of the Conjugate Descent Method*, Mathematical Advances, 6 (1996), pp. 552-562

[8] Y. H. Dai and Y. Yuan, *Convergence of the Fletcher-Reeves Method under A Generalized Wolfe Search*, Journal Computational Mathematics of Chinese Universities, 2 (1996), pp. 142-148

[9] R. Fletcher, *Practical Methods of Optimization vol. 1: Unconstrained Optimization*, John Wiley & Sons (New York), 1987

[10] R. Fletcher and C. Reeves, *Function minimization by conjugate gradients*, Comput. J., 7 (1964), pp. 149-154

[11] J. C. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM Journal on Optimization, 2 : 1 (1992), pp. 21-42

[12] L. Grippo and S. Lucidi, *A globally convergent version of the Polak-Ribière conjugate gradient method,* Math. Prog., 78 (1997), pp. 375-391

[13] M. R. Hestenes and E. L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards Sect., 5:49 (1952), 409-436

[14] G. H. Liu, J. Y. Han and H. X. Yin, *Global convergence of the Fletcher-Reeves algorithm with an inexact line search*, Appl. Math. J. Chinese Univ. Ser. B, 10 (1995), pp. 75-82

[15] D. C. Liu and J. Nocedal, *On the limited memory method for large scale optimization*, Math. Prog., 45 (1989), pp. 503-528

[16] J. Nocedal, *Large Scale Unconstrained Optimization*, Research Report, Northwestern University, 1996.

[17] E. Polak and G. Ribière, *Note sur la convergence de directions conjugées*, Rev. Francaise Informat Recherche Opertionelle, 3e Année 16 (1969), pp. 35-43

[18] B. T. Polyak, *The conjugate gradient method in extremem problems*, USSR Comp. Math. and Math. Phys., 9 (1969), pp. 94-112

[19] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in: D. F. Griffths ed., Numerical Analysis, Lecture Notes in Mathematics 1066, Springer-Verlag (Berlin) (1984), pp. 122-141

[20] M. J. D. Powell, *Convergence Properties of Algorithms for Nonlinear Optimization*, SIAM Review, 28 (1986), pp. 487-500

[21] D. Touati-Ahmed and C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optimization Theory Appl., 64 (1990), pp. 379-397

[22] C. Y. Wang and Y. Z. Zhang, *Global convergence properties of s-related conjugate gradient methods*, Chinese Science Bulletin 43:23 (1998), pp. 1959-1965