SQU Journal for Science, 2021, 26(2), 141-151 DOI:10.53539/squjs.vol26iss2pp141-151
Sultan Qaboos University

Improved Fletcher–Reeves Methods Based on New Scaling Techniques†

Amal Al-Saidi[1,*] and Mehiddin Al-Baali[2]

*[1]Department of Mathematics, Sultan Qaboos University, Muscat, Sultanate of Oman. [2]Department of Mathematics, Sultan Qaboos University, Muscat, Sultanate of Oman; and GERAD-HEC, Montréal, Canada. *Email: Amal2.AlSaeedi@moe.om.*

ABSTRACT: This paper introduces a scaling parameter to the Fletcher-Reeves (FR) nonlinear conjugate gradient method. The main aim is to improve its theoretical and numerical properties when applied with inexact line searches to unconstrained optimization problems. We show that the sufficient descent and global convergence properties of Al-Baali for the FR method with a fairly accurate line search are maintained. We also consider the possibility of extending this result to less accurate line search for appropriate values of the scaling parameter. The reported numerical results show that several values for the proposed scaling parameter improve the performance of the FR method significantly.

Keywords: Unconstrained optimization; Large-scale optimization; Line search framework; Nonlinear conjugate gradient methods; Fletcher-Reeves method.

طرق فليتشر- ريفز المحسنة إستنادا إلى تقنيات موازنة جديدة

أمل السعيدي و محي الدين البعلي

الملخص: يقدم هذا البحث معامل موازنة لطريقة فليتشر- ريفز في طرق التدرج المترافقة غيرالخطية. الهدف الرئيسي هو تحسين الصفات النظرية والعددية لتلك الطريقة عند تطبيقها مع مستقيمات بحث تقريبية على مسائل الأمثليات غيرالمقيدة. نبين إمكانية الحفاظ على صفات البعلي في الإنحدارالكافي والتقارب الشامل لطريقة FR مع مستقيم بحث تقريبي بدقة معقولة. نأخذ في الإعتبارأيضًا إمكانية تعميم هذه النتيجة إلى مستقيم بحث تقريبي بإستخدام قيم مناسبة لمعامل الموازنة. تظهرالنتائج العددية وجود العديد من قيم معامل الموازنة لتحسين أداء طريقة FR بشكل ملحوظ.

الكلمات المفتاحية: الأمثليات غيرالمقيدة، مسائل ذات أبعاد عالية، مستقيم البحث ، طرق التدرج المترافقة غيرالخطية، طريقة فليتشر- ريفز.

1. Introduction

This paper is concerned with the line-search Fletcher-Reeves (FR) conjugate gradient (CG) method for solving the unconstrained optimization problem

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

where f: $R^n \rightarrow R$ is a smooth function and its gradient $g(x) = \nabla f(x)$ is available for any value of x. For a given starting point x_1, the method defines a sequence of points $\{x_k\}$ iteratively as follows. On each iteration, once the function value $f_k = f(x_k)$ and the gradient value $g_k = g(x_k)$ are calculated, a search direction d_k is defined such that the descent property $d_k^T g_k < 0$ holds, assuming $g_k \neq 0$ since $g_k = 0$ holds at a solution of problem (1). Then it is possible to find a positive steplength α_k and hence a new point

† Some preliminary numerical results were presented at 5th International Conference on Numerical Analysis and Optimization (NAO-V), January 2020, Muscat, Oman

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

such that the function reduction $f_k - f_{k+1}$ is sufficiently positive (see for example Fletcher [1]).

In the class of CG methods, the search direction d_k is defined initially by the steepest descent choice $d_1 = -g_1$ that satisfies the above descent property for $k = 1$. The other directions are defined by

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \tag{3}$$

where β_k is the CG parameter. In particular, Fletcher and Reeves [2] propose the positive value of

$$\beta_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \tag{4}$$

(referred to as β_k^{FR}), where $||\cdot||$ denotes the Euclidean norm. If the exact line search equation

$$d_k^T g_{k+1} = 0 \tag{5}$$

is satisfied, which is guaranteed if the steplength $\alpha_k = \alpha^*$ is obtained by solving the minimization subproblem $\alpha^* = \arg\min_\alpha f(x_k + \alpha d_k)$ exactly, then the above descent property with k replaced by $k+1$ is also satisfied. Since solving this subproblem exactly (referred to as exact line search) is impractical, inexact line search is usually used so that the descent property may not hold. In this case, choice (4) is replaced by $\beta_k = 0$ (i.e., direction (3) is reset to that of the steepest descent) so that the descent property holds for all k (for further detail, see for example Fletcher [1], Nocedal and Wright [3] and Pytlak [4]). Because this resetting may destroy the useful features of the CG method, in certain cases, many choices for replacing the CG parameter FR formula have been proposed (see for example Hager and Zhang [5] and the references therein). If exact line search is employed and $f(x)$ is quadratic, all the directions generated via (3) are mutually conjugate. Hence, the gradients of $f(x)$ at the different iterates are mutually orthogonal, i.e., we have

$$g_{k+1}^T g_k = 0, \tag{6}$$

and these formulae are reduced to the FR one (see e.g. [1, 3, 4]).

Our aim in this paper is to maintain the global convergence property of the FR method and improve its behavior by scaling formula (4) by ξ_k before using it in (3) (as in Al-Baali [6]). Therefore, the resulting scaled direction

$$d_{k+1} = -g_{k+1} + \xi_k \beta_k d_k \tag{7}$$

would be close to the steepest descent one if ξ_k is chosen sufficiently close to zero. Hence, by continuity, the corresponding class of scaled CG methods (referred to as ScFR when $\beta_k = \beta_k^{FR}$) would satisfy the descent and convergence properties that the steepest descent method has.

It is important to mention that Powell [7] and essentially Zoutendijk [8] show that the FR method with exact line searches converges globally, in the sense that $\lim_{k\to\infty} \inf \|g_k\| = 0$. This result has been extended by Al-Baali [9] to inexact line searches with a fairly accurate values of α_k, known as the first practical global convergence result for the FR method (see for example Nocedal [10]). Al-Baali has obtained this result based on showing that the sufficient descent condition

$$d_k^T g_k \le -c\|g_k\|^2 \tag{8}$$

holds for some positive constant c, assuming the steplength α_k satisfies the strong Wolfe conditions

$$f_{k+1} \le f_k + \rho \alpha_k g_k^T d_k, \quad |\, g_{k+1}^T d_k \,| \le -\sigma g_k^T d_k, \tag{9}$$

for $0 < \rho < \sigma < \frac{1}{2}$. It is worth noting that a value of $\alpha_k > 0$ which satisfies these conditions, for $0 < \rho < \frac{1}{2}$ and $\rho < \sigma < 1$, can be found in a finite number of operations whenever the above descent property holds (see for example Al-Baali and Fletcher [11] and Fletcher [1]).

In Section 2, we show that the ScFR class of methods with $\xi_k \in (0, 1]$ maintains the sufficient descent property that the FR method has for the strong Wolfe conditions (9) with $\sigma < \frac{1}{2}$. We will also enforce this property for any line search technique with sufficiently small values of the scaling parameter ξ_k. In addition, that section defines some choices for ξ_k. In Section 3, we introduce the quasi-Newton feature to ξ_k for the FR method as considered by Al-Saidi et al. [12] for the scaled CG methods. Section 4 shows that the proposed class of ScFR methods maintains the global convergence property that the FR method has when the strong Wolfe conditions are employed with $\sigma < \frac{1}{2}$. In Section 5, we study the behavior of the proposed ScFR methods by applying them to a set of standard test problems. It is shown that the proposed scaling technique improves the performance of the FR method significantly in many cases. Finally, Section 6 concludes the paper.

2. The ScFR Class of Methods

We now define the scaled FR (ScFR) class of methods as suggested by Al-Baali [6] for the CG class of methods. The author replaces the search direction (3) by (7), for $k > 1$, where $0 < \xi_k \le 1$ is the scaling parameter, and maintains $d_1 = -g_1$. Since the value of $\xi_k = 0$ reduces this direction to that of the steepest descent, it follows by continuity that the ScFR class of methods, with values of ξ_k sufficiently close to zero, has the sufficient descent and global convergence properties that the steepest descent method has.

We first define some values of ξ_k that enforce the sufficient descent condition (8) as follows (Al-Baali [6]). We note that this condition holds for $k = 1$ and $c = 1$. We now prove this condition for $k \geq 1$ as follows. On substituting (7) into (8) with k replaced by $k+1$, we have

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \xi_k \beta_k d_k^T g_{k+1} \leq -c\|g_{k+1}\|^2. \tag{10}$$

Hence,

$$\xi_k \beta_k (d_k^T g_{k+1}) \leq (1-c)\|g_{k+1}\|^2 \tag{11}$$

or equivalently by the FR formula (4),

$$\xi_k (d_k^T g_{k+1}) \leq (1-c)\|g_k\|^2, \tag{12}$$

where $0 < c \leq 1$. This upper bound on c is included because condition (11) cannot hold for $c > 1$ and $d_k^T g_{k+1} = 0$ (the exact line search option). We note that (11) holds if its left hand side is nonpositive or either ξ_k, $d_k^T g_{k+1}$ or β_k is sufficiently close to zero.

This paper assumes that the CG parameter β_k is given by the FR formula (4), unless otherwise stated, and defines ξ_k such that (11) holds whether d_k satisfies the sufficient descent condition (8) or not. For convenience, we start with the following result of Al-Saidi et al. [12] assuming β_k satisfies (8).

Theorem 1. *Consider the ScFR class of methods, defined by (2), (4) and (7) and assume that the search direction (7) with $\xi_k = 1$ satisfies the sufficient descent condition (8). Then, this condition remains satisfied for $0 \leq \xi_k \leq 1$.*

Proof. When $k = 1$, $d_k = -g_k$ and hence condition (8) holds with $c = 1$. For $k \geq 1$, it follows from (7) that

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 + \xi_k \beta_k (d_k^T g_{k+1}). \tag{13}$$

If $\beta_k (d_k^T g_{k+1}) < 0$, it follows that

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2,$$

which is (8) with $c = 1$ and k replaced by $k+1$, since $\xi_k \geq 0$. Otherwise, if $\beta_k (d_k^T g_{k+1}) \geq 0$, then (13) implies

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \beta_k (d_k^T g_{k+1}),$$

since $\xi_k \leq 1$. Hence, by the theorem assumption we obtain (8) with k replaced by $k+1$. □

We note that this result remains valid for any value of $c > 0$ if the choice $\xi_k = 1$ is used when $\beta_k (d_k^T g_{k+1}) \leq 0$ as we consider below for some choices of ξ_k. It is important to note that if the strong Wolfe conditions (9) are employed with $\sigma < 1/2$, then the sufficient descent condition (8) holds for $c = c_\sigma \equiv \frac{1-2\sigma}{1-\sigma}$ (see Al-Baali [9], for detail). Thus, for any choice of $c \leq c_\sigma$, $\xi_k = 1$ and the ScFR method is reduced to the standard FR method. However, using values of $c > c_\sigma$ require defining $\xi_k < 1$, for which we suggest the following way.

If condition (11) holds with $\xi_k = 1$, then we use this value so that the CG search direction is unchanged. Otherwise, we choose ξ_k such that condition (11) holds with equality. Thus, we choose

$$\xi_k = \begin{cases} \dfrac{(1-c)\|g_{k+1}\|^2}{\beta_k (d_k^T g_{k+1})} & \text{if } \beta_k (d_k^T g_{k+1}) > (1-c)\|g_{k+1}\|^2, \\ 1 & \text{otherwise,} \end{cases} \tag{14}$$

or equivalently for the ScFR method,

$$\xi_k^1 = \begin{cases} \dfrac{(1-c)\|g_k\|^2}{d_k^T g_{k+1}} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2, \\ 1 & \text{otherwise.} \end{cases} \tag{15}$$

In this case, it follows that

$$\xi_k^1 \beta_k = \begin{cases} \dfrac{(1-c)\|g_{k+1}\|^2}{d_k^T g_{k+1}} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2, \\ \beta_k^{FR} & \text{otherwise} \end{cases} \tag{16}$$

which shows that scaling is employed only when α_k is sufficiently remote away from the right hand side of a minimizer α^*. To be precise, $\xi_k = 1$ is used when $\beta_k (d_k^T g_{k+1}) \leq 0$, although in this case (11) holds for any value of $\xi_k > 0$. To obtain the least change in the CG parameter, c should be chosen close to zero (in practice, the value of $c = 0.001$ seems reasonable as observed by Al-Saidi [13]). We also note that $\xi_k^1 = 1$ usually is used for a fairly accurate line search ($\sigma \leq 0.1$). Thus, we consider the outline of the scaled FR (ScFR) method as in Algorithm 1 that is reduced to the standard FR method if $\xi_k = 1$ for all k. If in Step 2 exact line searches (or nearly so) are employed for all k, then Algorithm 1 is reduced to the standard FR method. In practice, we observed that the above scaling technique improves the performance of the FR method significantly in several cases (see Section 5, for details).

Algorithm 1 (ScFR)
Step 0: Given $\varepsilon > 0$, $0 < c \leq 1$ and x_1, let $d_1 = -g_1$ and set $k = 1$
Step 1: Stop if $\|\|g_1\|\| \leq \varepsilon$
Step 2: Compute a positive steplength α_k which satisfies certain standard conditions
Step 3: Compute a new point by (2); $x_{k+1} = x_k + \alpha_k d_k$
Step 4: Define the FR parameter β_k by (4)
Step 5: Define ξ_k (e.g., by (14))
Step 6: Compute a new search direction by (7); $d_{k+1} = -g_{k+1} + \xi_k \beta_k d_k$
Step 7: Set $k := k + 1$ and go to Step 1.

It is important to realize that the choice of ξ_k^1 in (15) is independent of the line search technique. However, if the strong Wolfe conditions (9) are employed, the first case of (15) could be replaced by a smaller value to obtain

$$\xi_k^2 = \begin{cases} \dfrac{(1-c)\|g_k\|^2}{\sigma |d_k^T g_k|} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2, \\ 1 & \text{otherwise.} \end{cases} \tag{17}$$

This choice satisfies the sufficient descent condition (11) with $\xi_k = \xi_k^2$, because $\xi_k^2 \leq \xi_k^1$ and we have $d_k^T g_{k+1} > 0$ and $\xi_k^2 (d_k^T g_{k+1}) \leq \xi_1 (d_k^T g_{k+1}) = (1 - c) \|g_k\|^2$. Indeed, replacing ξ_k^2 by a smaller value, which we consider below, maintains the sufficient descent condition (12). In practice, for c = 0.001, we observed that choice (17) works better than (15).

We now consider the rearrangement of Al-Baali [6] for the FR parameter as

$$\beta_k = \eta_k^T g_{k+1}, \qquad \eta_k = \frac{g_{k+1}}{\|g_k\|^2}, \tag{18}$$

which is possible for most CG methods (e.g., see Hager and Zhang [5] and Narushima and Yabe [14] and the references therein). Using this form and observing that

$$d_k^T g_{k+1} \leq \|d_k\| \|g_{k+1}\|, \tag{19}$$

the first case of (15) can be reduced to obtain

$$\xi_k^3 = \begin{cases} \frac{(1-c)\|g_k\|^2}{\|d_k\|\|g_{k+1}\|} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2, \\ 1 & \text{otherwise.} \end{cases} \tag{20}$$

To increase the interval for the scaling parameter, we use (19) again in the condition of (20) so that we obtain the following choice of Al-Baali [6]:

$$\xi_k^4 = \begin{cases} \frac{(1-c)\|g_k\|^2}{\|d_k\|\|g_{k+1}\|} & \text{if } \|d_k\|\|g_{k+1}\| > (1-c)\|g_k\|^2, \\ 1 & \text{otherwise .} \end{cases} \tag{21}$$

We note that the interval for using $\xi_k < 1$ in (21) is larger than or equal to that used by (20) so that sufficiently large values of c might be used. Indeed, in practice, we observed that choice (17) is preferable to (21). Thus, we will not provide further details about (21).

3. A Quasi-Newton Feature

One feature for choices (20) and (21) is given by Al-Baali [6] who shows that the values of $\xi_k \leq \frac{(1-c)}{\|\eta_k\|\|d_k\|}$, where η for the FR formula is given in (18), enforce some positive eigenvalues of the ScCG matrix

$$H_{k+1} = I - \xi_k d_k \eta_k^T, \tag{22}$$

noting that direction (7) (similar to that of Perry [15]) can be rearranged as follows

$$d_{k+1} = -H_{k+1} g_{k+1}. \tag{23}$$

Thus, we introduce another feature to the ScFR method like that of Al-Saidi et al. [12] for the CG methods by modifying the scaling parameter ξ_k such that ideally the quasi-Newton condition

$$H_{k+1} \gamma_k = \delta_k, \tag{24}$$

Where

$$\delta_k = x_{k+1} - x_k, \quad \gamma_k = g_{k+1} - g_k, \tag{25}$$

holds, assuming H_{k+1} approximates the inverse of Hessian of $f(x)$ at x_k. Because, in general, we cannot fulfill condition (24), we let

$$\xi_k = \arg\min_{\xi} \|H_{k+1}\gamma_k - \delta_k\| = \arg\min_{\xi} \|\gamma_k - \delta_k - \xi(\eta_k^T\gamma_k)d_k\|. \tag{26}$$

Solving this minimization problem (see for example Al-Saidi *et al*. [12]), it follows that

$$\xi_k^q = \frac{(\gamma_k-\delta_k)^T d_k}{\eta_k^T\gamma_k\|d_k\|^2}. \tag{27}$$

On substituting η_k for the FR formula (18), (27) is reduced to

$$\xi_k^q = \frac{(\gamma_k-\delta_k)^T d_k\|g_k\|^2}{\gamma_k^T g_{k+1}\|d_k\|^2}. \tag{28}$$

To maintain the search direction differs from that of the steepest descent, the values of ξ_k should be bounded away from zero. Since, in addition, the first case of (15) enforces the sufficient descent condition (12) with equality, and that replacing ξ_k by a smaller value keeps inequality (12) holds, we may enforce $\xi_k \in [\hat{c}, 1]$, where $\hat{c}$ is a small positive number. Thus, we modify (15), which defines ξ_k^1, to

$$\xi_k^{q1} = \begin{cases} \dfrac{(1-c)\|g_k\|^2}{d_k^T g_{k+1}} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2 \text{ and } \xi_k^q \notin [\hat{c}, 1], \\ \min\left\{\dfrac{(1-c)\|g_k\|^2}{d_k^T g_{k+1}}, \xi_k^q\right\} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2 \text{ and } \xi_k^q \in [\hat{c}, 1], \\ 1 & \text{otherwise.} \end{cases} \tag{29}$$

and modify (17), which defines ξ_k^2, to

$$\xi_k^{q2} = \begin{cases} \dfrac{(1-c)\|g_k\|^2}{\sigma|d_k^T g_k|} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2 \text{ and } \xi_k^q \notin [\hat{c}, 1], \\ \min\left\{\dfrac{(1-c)\|g_k\|^2}{\sigma|d_k^T g_k|}, \xi_k^q\right\} & \text{if } d_k^T g_{k+1} > (1-c)\|g_k\|^2 \text{ and } \xi_k^q \in [\hat{c}, 1], \\ 1 & \text{otherwise.} \end{cases} \tag{30}$$

Since choices (29) and (30) can be rewritten as $\xi_k^{qi} = \min(\max(\xi_k^q, \hat{c}), \xi_k^i)$, for $i = 1,2$ respectively, we observe that $\xi_k^{qi} \le \xi_k^i$. Similarly, we modify any ξ_k^i, for i = 1, 2, 3, 4, to $\xi_k^{qi} = \min(\max(\xi_k^q, \hat{c}), \xi_k^i)$. In practice, we observed that these modified choices with $\sigma = 0.1$ and $\hat{c} = 0.001$ improves the performance of ξ_k^2 (further possible choices and details can be seen in Al-Saidi [13]). Therefore, the above choices for ξ_k are chosen to enforce the sufficient descent condition as shown in the following result which is an extension to Theorem 1.

Theorem 2. *Consider the ScFR class of methods, defined by (2), (4) and (7), and let* $\xi_k = \xi_k^i$ *or* $\xi_k = \xi_k^{qi}$, *for* $i = 1, 2, 3, 4$. *Then, the sufficient descent condition (8) holds for all* $k \ge 1$.

Proof. It is obvious for $k = 1$. For $k \ge 1$, $\xi_k = \xi_k^i$, $1 \le i \le 4$, are chosen such that inequality (12) holds which is equivalent to the sufficient descent condition (10). Since $\xi_k^{qi} \le \xi_k^i$, $1 \le i \le 4$, inequalities (11) and (12) remain satisfied so that condition (10) holds. Hence, condition (10) implies (8). □

Introducing other features to the scaling parameter ξ_k are also possible. In particular, we could introduce the useful conjugacy condition $\gamma_k^T d_{k+1} = 0$. This equation with (7) holds if $\xi_k = \xi_k^c \equiv \frac{\gamma_k^T g_{k+1}}{\beta_k \gamma_k^T d_k}$ which exists if the Wolfe condition holds. Similarly if we consider the modified conjugacy condition of Dai and Liao [16], $\gamma_k^T d_{k+1} = t d_k^T g_{k+1}$ where $t \ge 0$ is a scalar, we obtain $\xi_k = \frac{(\gamma_k - t d_k)^T g_{k+1}}{\beta_k \gamma_k^T d_k}$ which reduces direction (7), for any β_k to the class of Dai and Liao.

It is worth noting that Al-Baali [6] defines a bound on ξ_k to ensure that matrix (22) is positive definite, but we do not consider it here because it may increase the value of ξ_k so that the sufficient descent condition (11) may not hold.

4. Global Convergence Property

We now study the possibility of imposing the global convergence property

$$\liminf_{k\to\infty} \|g_k\| = 0 \tag{31}$$

to the ScFR methods which we proposed in the previous sections such that the sufficient descent condition (8) holds for all $k \ge 1$. Therefore, we first state the following standard assumptions on the objective function.

Assumption 1.

1. The level set $\mathcal{L} = \{x: f(x) \le f(x_0)\}$ *is bounded,*

2. The objective function f is continuously differentiable in some neighborhood $\mathcal{N}$ *of* $\mathcal{L}$,

3. The gradient $g(x)$ *is Lipschitz continuous, that is there exists a positive constant L such that*

$$\| g(x) - g(\tilde{x}) \| \le L \| x - \tilde{x} \| \quad \text{for all } x, \tilde{x} \in \mathcal{N}. \tag{32}$$

Supposing Assumption 1 holds, Powell [7] and Zoutendijk [8] prove the ideal result that the FR method, defined by (2), (3) and (4), satisfies the global convergence property (31) if exact line searches are used. This means that impractical choice of $\sigma = 0$ in (9) is considered. Al-Baali [9], however, considers the practical strong Wolfe conditions with $\sigma < \frac{1}{2}$ to show that the sufficient descent condition (8) holds and extends the above global convergence result to the practical FR method. This result has been extended by Touati-Ahmed and Storey [17] to the interval $0 \leq \beta_k \leq \beta_k^{FR}$, while Gilbert and Nocedal [18] have extended it to

$$-\beta_k^{FR} \leq \beta_k \leq \beta_k^{FR}. \tag{33}$$

Thus, we obtain the following convergence result for the ScFR methods.

Theorem 3. *Suppose Assumption 1 holds and let the ScFR class of methods be defined by (2), (4), (7), and $|\xi_k| \leq 1$. If the strong Wolfe conditions (9) are employed with $\sigma < \frac{1}{2}$, then the ScFR methods converge globally in the sense that limit (31) holds.*

Proof. The ScFR methods are defined in the previous sections like the FR method, except that the FR formula $\beta_k^{FR} > 0$, defined by (4), is replaced by $\beta_k^{ScFR} = \xi_k \beta_k^{FR}$. Because $|\xi_k| \leq 1$, conditions (33) hold and hence the result follows from that of Gilbert and Nocedal [18]. □

5. Numerical Results

To show the efficiency of the proposed scaling technique for the Fletcher-Reeves method, we describe some numerical results. The results are obtained by applying the FR method with several scaled choices to a set of 46 different type of unconstrained optimization problems; their function names are given in Table 1. They belong to the CUTE [19] and Moré, Garbow and Hillstrom [20] collections of test problems. We also include the simple quadratic function of Al-Baali (see for example Al-Saidi [13]). For each type of problem, we let the dimension n varies as 2, 10, 100, 1000, 10000 so that the total number of test problems is 230. All the ScFR methods that we consider are defined by Algorithm 1 and differ only in Step 5 for defining the scaling parameter ξ_k unless otherwise stated. For the purposes of an accurate comparison, we use in Step 2 for all methods the Matlab line search program routine of Al-Baali (which essentially written in Fortran by Fletcher) which computes a value of the steplength α_k that satisfies the strong Wolfe conditions (9) as described in Al-Baali and Fletcher [11] and Fletcher [1]. The run was stopped as in Step 1 when either $||g_k|| \leq 10^{-6}$ or the number of iterations reaches 10^5. We consider the following algorithms:

- FR; the standard FR method, defined by Algorithm 1 with $\rho = 10^{-4}$ and $\sigma = 0.1$ in Step 2 and $\xi_k = 1$ in Step 5.

Table 1. The test functions.

Number	Function's name	Number	Function's name
1	A Simple Quadratic function	24	Trecanni
2	Extended White and Holst	25	Zettl
3	Extended Rosenbrock	26	Shallow
4	Extended Freudenstein and Roth	27	Generalized Quartic
5	Extended Beale	28	Axis Parallel hyper-ellipsoid
6	Extended Wood	29	Leon
7	Raydan 1	30	Generalized Tridiagonal 1
8	Generalized Tridiagonal 1	31	Generalized Tridiagonal 2
9	Diagonal 4	32	POWER
10	Extended Himmelblau	33	Quadratic QF1
11	Extended Hiebert	34	Extended quadratic penalty QP2
12	FLETCHCR	35	Dixon and Price
13	Extended Powell	36	Quartic
14	NONSCOMP	37	Diagonal4
15	Extended DENSCHNB	38	Colville
16	Extended quadratic penalty QP1	39	Schumer Steilitz
17	Extended Penalty I	40	Sphere
18	De Jong's	41	Sum Squares
19	Hager	42	Powell Singular
20	Extended Maratos	43	Qing
21	Cube	44	Generalized PSC1
22	Three hump function	45	Perturbed Quadratic
23	Booth	46	Diagonal 2

- ScFRi; same as FR except that in Step 5, $\xi_k = \xi_k^i$, for $i = 1, 2, 3, 4$, defined by (15), (17), (20), and (21), respectively.
- ScFRqi, $i \leq 4$; same as ScFRi except that in Step 5, $\xi_k = \xi_k^{qi} = \min\left(\max\left(\xi_k^q, \hat{c}\right), \xi_k^i\right)$ with $\hat{c} = 0.001$, where ξ_k^q is given by (28).

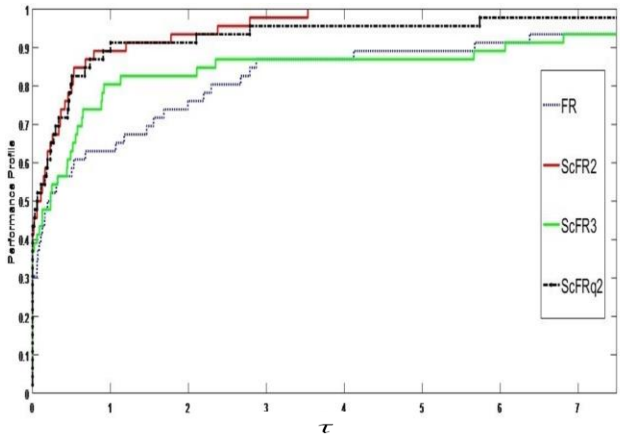
We observed that ScFR1, ScFR4 and ScFRqi, for $i = 1, 3, 4$, are less efficient than the others. Therefore, we will not give further details about them. Although it is possible to study the behaviour of other methods that we considered here, we do not consider them in our comparisons because they were less efficient than the above algorithms.

For a useful comparison, we used the performance profiles tool of Dolan and Moré [21], which compares some solvers on the above set of problems in terms of the number of line searches, the number of function evaluations, the number of gradient evaluations and CPU time, required to solve the problems. In general the performance profile $P_M(\tau)$, $\tau \geq 0$, is defined by the formula

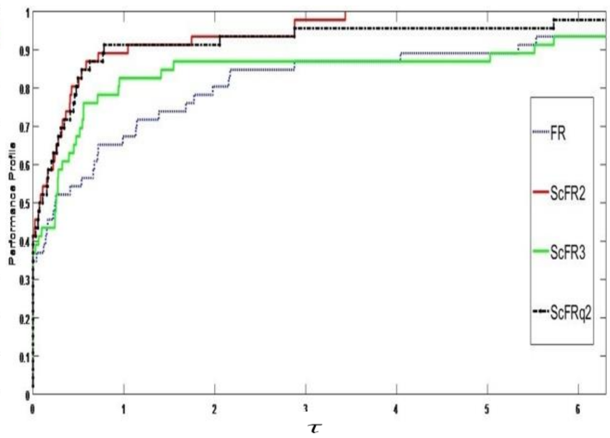
$$P_M(\tau) = \frac{\text{number of problems where } \log_2(\tau_{P,M}) \leq \tau}{\text{total number of problems}}, \tag{34}$$

where $\tau_{P,M}$ is the performance ratio of the number of line searches (similarly for the other measures) required to solve problem p by a method M to the lowest number of line searches required to solve problem p. The ratio $\tau_{P,M}$ is set to ∞ (or some large number) if the method M fails to solve problem p. The value of $P_M(\tau)$ at $\tau = 0$ gives the percentage of test problems for which the M method is the best. The value for τ large enough is the percentage of test problems that M method can solve. The relative efficiency and reliability of each method can be directly seen from the performance profiles: the higher the particular curve, the better is the corresponding method.

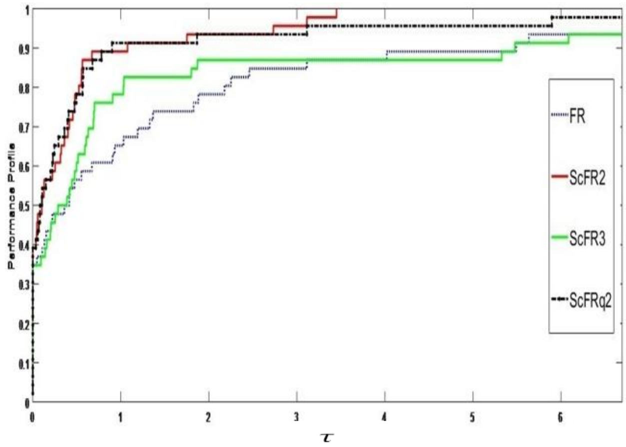
The performance profiles for the FR, ScFR2, ScFR3 and ScFRq2 algorithms are given in Figures 1. We observe that all scaling techniques improve the performance of the FR method substantially. We also observe that ScFR2 is slightly better than ScFRq2 and both of them perform better than ScFR1. Thus, ScFR2 seems to be the best of the ScFR algorithms.



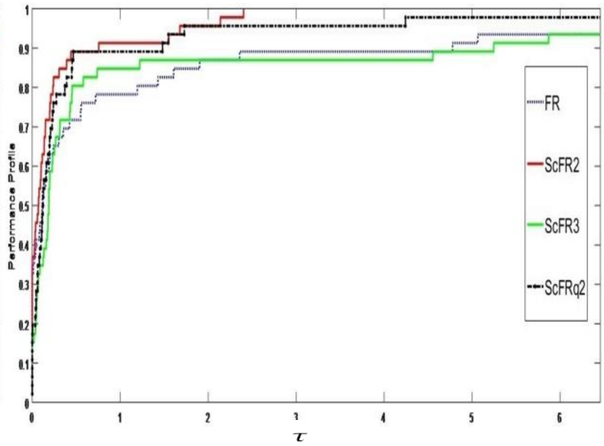
(a) # Line searches



(b) # Function Evaluation



(c) # Gradient Evaluation



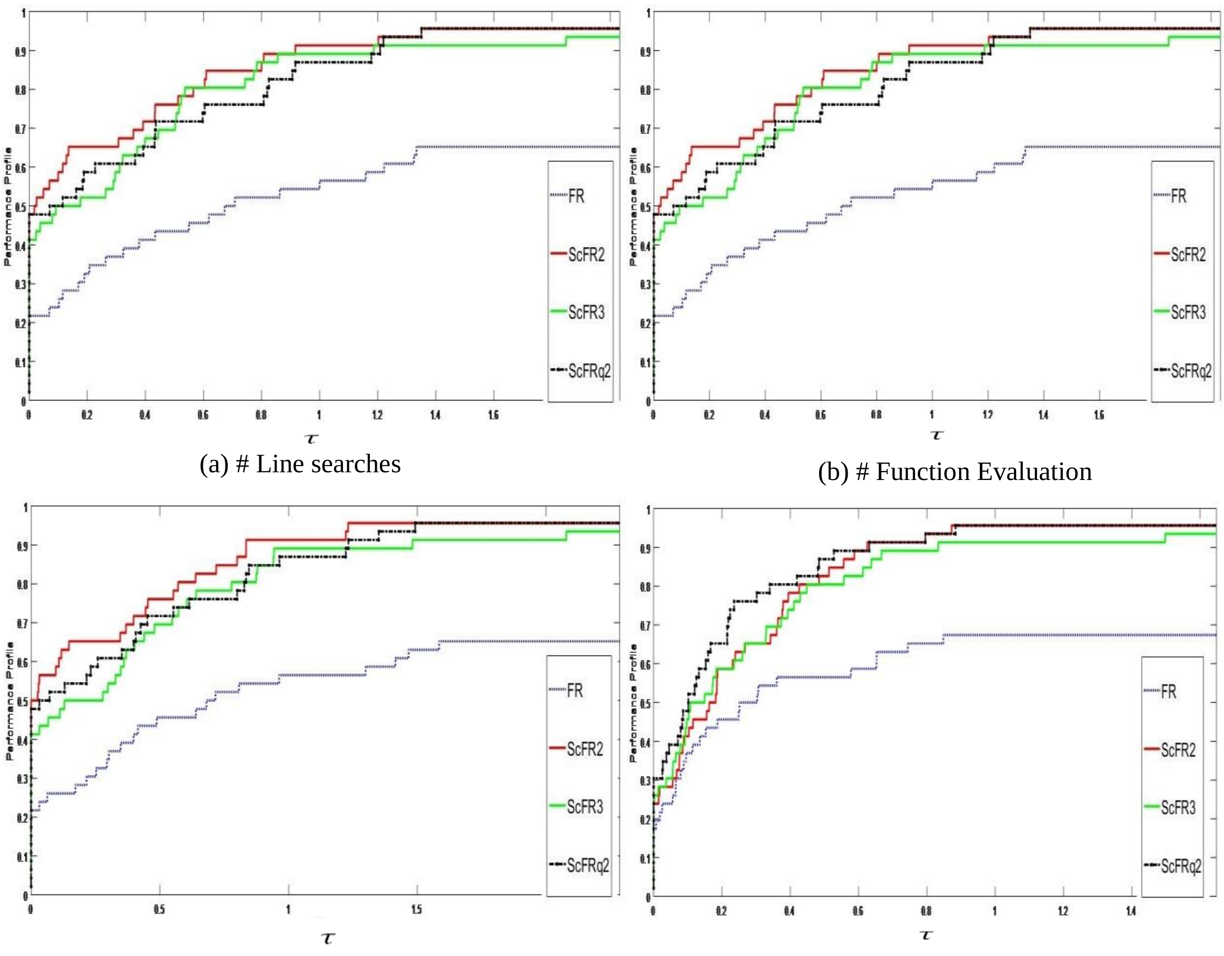
(d) CPU time

Figure 1. Comparison among FR, ScFR2, ScFR3 and ScFRq2 for $\sigma = 0.1$.

To give further idea about the behaviour of these methods for a large value of σ, we repeated the run for the previous four methods but with $\sigma = 0.9$ instead of $\sigma = 0.1$ in Step 2. The comparisons are given in Figures 2. Although these figures show that the FR method failed to solve about 40% of the test problems which is expected, the three ScFR methods solved almost all problems successfully. Thus, the ScFR methods are superior to the FR method.

To choose the best choice for σ, we repeated the run for the best ScFR2 method, using the values of $\sigma = 0.1, 0.4, 0.7, 0.9$ and 0.9999 (for the first value, the method is referred to as ScFR2). The comparisons are given in Figures 3. We observe that the performance of the ScFR method as improves as σ decreases in terms of the number of line searches, function evaluations and gradient evaluations. However, the choice $\sigma = 0.1$ is significantly better than the others in terms of CPU time. Thus, $\sigma = 0.1$ is the best value for all the measurements. Another useful observation is that the scaled methods solved all problems for not only $\sigma < 1/2$ but also for $\sigma \geq 1/2$. Thus, the scaled technique improves not only the good behaviour of the FR method when σ is sufficiently small but also solved problems that cannot be solved by the FR method.

We now compare the ScFR2 method with some useful CG methods such as the Polak-Ribière-Polyak [22, 23], Dai-Yuan [24] and Hager-Zhang [25] methods. These methods (referred to as PRP, DY and HZ) are defined by Algorithm 1 with $\xi_k = 1$ in Step 5 and β_k in Step 4 is given respectively by the following formulae:



(a) # Line searches

(b) # Function Evaluation

(c) # Gradient Evaluation

(d) CPU time

Figure 2. Comparison among FR, ScFR2, ScFR3 and ScFRq2 for $\sigma = 0.9$.

$$\beta_k^{PRP} = \frac{\gamma_k^T g_{k+1}}{\|g_k\|^2}, \tag{34}$$

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T \gamma_k} \tag{35}$$

and

$$\beta_k^{HZ} = \left(\gamma_k - 2\frac{\|\gamma_k\|^2}{\gamma_k^T d_k} d_k\right)^T \frac{g_{k+1.}}{\gamma_k^T d_k}. \tag{36}$$

Using $\sigma = 0.1$, the comparisons of the methods are given in Figures 4. We see that ScFR2 performs a little better than DY and a little worse than PRP and HZ. Thus, the ScFR2 method is efficient and there is a room for improving it. The above numerical results show indeed that all the ScFR methods perform well and the ScFR2 method is the best of them.

6. Conclusion

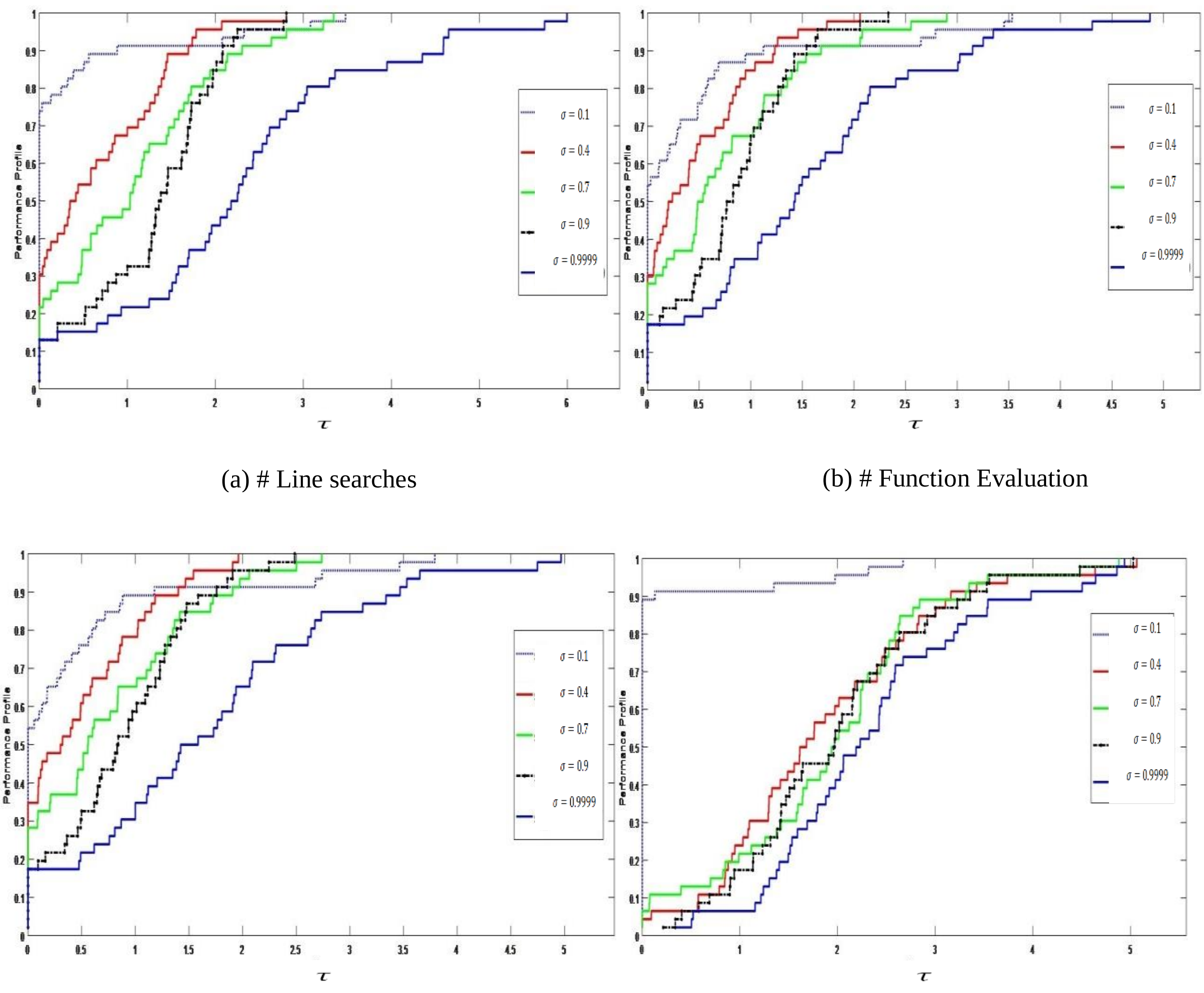
We show that introducing a simple scaling technique to the well-known Fletcher-Reeves method maintains its global convergence property, improves its performance significantly and solved problems that cannot be solved by the FR method. Since the ScFR methods work well when the strong Wolfe conditions are used with $\sigma \geq 1/2$, it is expected to extend their global convergence to $\sigma < 1$. It is also worth testing further possible choices for the FR scaled parameter (e.g., similar to that applied to the DY method by Esmaeili *et al.* [26]).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgment

The authors would like to thank the editor and two anonymous referees for their critical and constructive comments



(a) # Line searches

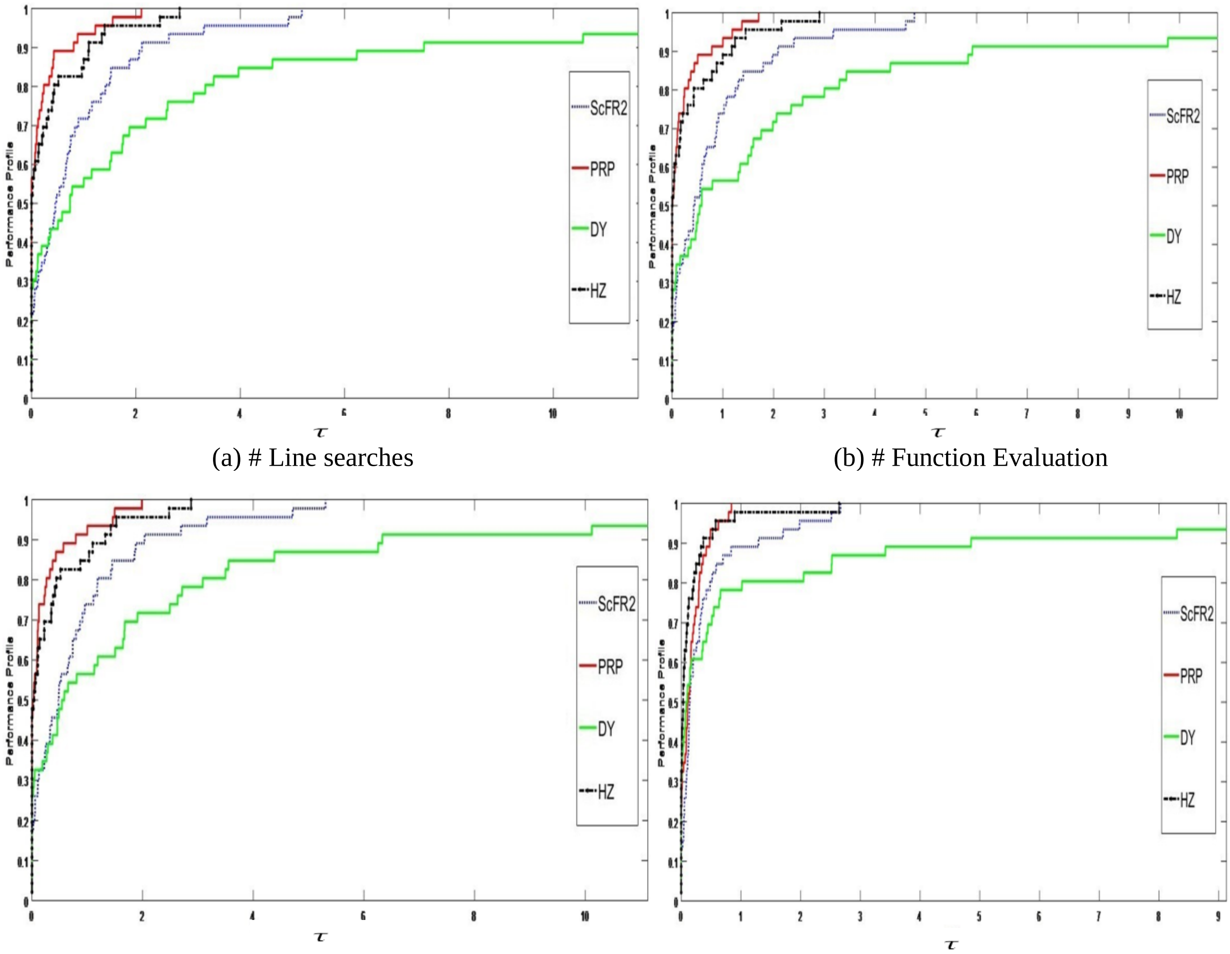
(b) # Function Evaluation

(c) # Gradient Evaluation

(d) CPU time

Figure 3. ScFR2 Comparison among the values of σ = 0.1, 0.4, 0.7, 0.9 and 0.9999.

which improve the quality of the paper. We would also like to thank Lucio Grandinetti, Calabria University, Italy, and Ekkehard Sachs, Trier University, Germany, for the useful discussions on the preparation of the paper. The first author would like to thank the Oman Ministry of Education for the award of a scholarship.



(a) # Line searches

(b) # Function Evaluation

(c) # Gradient Evaluation

(d) CPU time

Figure 4. Comparison among ScFR2, PRP, DY and HZ for σ = 0.1.

References

1. Fletcher, R. *Practical Methods of Optimization*, 2nd edition, 2013, Wiley, Chichester, England.
2. Fletcher, R., and Reeves, C. Function minimization by conjugate gradients, *Computational Journal*, 1964, **7**, 149-154.
3. Nocedal, J., and Wright, S. J. *Numerical Optimization*, 2nd Edition, 2006, Springer, London.
4. Pytlak, R. *Conjugate Gradient Algorithms in Nonconvex Optimization*, 2009, Springer, Berlin.
5. Hager W., and Zhang, H. A survey of nonlinear conjugate gradient methods, *Pacific Journal of Optimization*, 2006, **2**, 35-58.
6. Al-Baali, M. A general class of conjugate gradient methods, *22nd International Symposium on Mathematical Programming (ISMP)*, Pittsburg, USA, 2015.
7. Powell, M.J.D. Nonconvex minimization calculations and the conjugate gradient method. In Numerical Analysis, Lecture Notes in Mathematics (Ed. Griffiths D.F.), *Springer*, 1984, **1066**, 122-141.
8. Zoutendijk, G. Nonlinear Programming, Computational Methods, in Integer and Nonlinear Programming (Ed. J. Abadie), North-Holland, Amsterdam, 1970, 37-86.
9. Al-Baali, M. Descent property and global convergence of the Fletcher-Reeves method with inexact line search, *IMA Journal of Numerical Analysis*, 1985, **5**, 121-124.
10. Nocedal, J. Theory of algorithms for unconstrained optimization, *Acta Numerica*, 1992, **1**, 199-242.
11. Al-Baali, M., and Fletcher, R. An efficient line search for nonlinear least squares, *Journal of Optimization Theory and Applications*, 1986, **48**, 359-378.
12. Al-Saidi, A., Al-Baali, M., Fasano, G., and Roma, M. New scaled and shifted conjugate gradient methods for

large-scale unconstrained optimization, Technical Report DM215, Department of Mathematics, Sultan Qaboos University, Muscat, Oman, 2021.
13. Al-Saidi, A. Improved conjugate gradient methods for large-scale unconstrained optimization, PhD Thesis, Sultan Qaboos University, Muscat, Oman, 2021.
14. Narushima, Y., and Yabe, H. A survey of sufficient descent conjugate gradient methods for unconstrained optimization, *SUT Journal of Mathematics*, 2014, **50**, 167-203.
15. Perry, J. M. A class of conjugate gradient algorithms with a two-step variable-metric memory, Discussion Paper 269, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University, Evanston, Illinois, USA, 1977.
16. Dai, Y. H., and Liao, L. Z. New conjugacy conditions and related nonlinear conjugate gradient methods, *Applied Mathematics and Optimization*, 2001, **43**, 87-101.
17. Touati-Ahmed, D., and Storey, C. Globally convergent hybrid conjugate gradient methods, *Journal of Optimization Theory and Applications*, 1990, **64**, 379-397.
18. Gilbert, J. C., and Nocedal, J. Global convergence properties of conjugate gradient methods for optimization, *SIAM Journal on Optimization*, 1992, **2**, 21-42.
19. Bongartz, I., Conn A. R., Gould, N., and Toint, Ph. L. CUTE: Constrained and unconstrained testing environment, *ACM Transactions on Mathematical Software*, 1995, **21**, 123-160.
20. Moré, J. J., Garbow, B. S., Hillstrom, K. E. Testing unconstrained optimization software, *ACM Transactions on Mathematical Software*, 1981, **7**, 17-41.
21. Dolan, E. D., Moré, J. J. Benchmarking optimization software with performance profiles, *Mathematical Programming*, 2002, **91**, 201-213.
22. Polak, E., and Ribière, G. Note sur la convergence de directions conjugées, *Rev. Francaise Informat Recherche Opertionelle*, 1969, 3e Annee **16**, 35-43.
23. Polyak, B. The conjugate gradient method in extreme problems, *USSR Computational Mathematics and Mathematical Physics*, 1969, **9**, 94-112.
24. Dai, Y. H., and Yuan, Y. X. A nonlinear conjugate gradient method with a strong global convergence property, *SIAM Journal on Optimization*, 1999, **10**, 177-182.
25. Hager, W., and Zhang, H. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on Optimization*, 2005, **16**, 170-192.
26. Esmaeili. H., Rostami, M., and Kimiaei. M. Extended Dai-Yuan conjugate gradient strategy for large-scale unconstrained optimization with applications to compressive sensing, *Filomat,* 2018, **32 (6)**, 2173-2191.

Received 30 June 2021
Accepted 6 September 2021