Numer. Math. (2006) 104:561–572
DOI 10.1007/s00211-006-0028-z

Numerische Mathematik

Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo-type line search

Li Zhang · Weijun Zhou · Donghui Li

Received: 1 July 2005 / Revised: 17 May 2006 /
Published online: 5 September 2006

Abstract In this paper, we are concerned with the conjugate gradient methods for solving unconstrained optimization problems. It is well-known that the direction generated by a conjugate gradient method may not be a descent direction of the objective function. In this paper, we take a little modification to the Fletcher–Reeves (FR) method such that the direction generated by the modified method provides a descent direction for the objective function. This property depends neither on the line search used, nor on the convexity of the objective function. Moreover, the modified method reduces to the standard FR method if line search is exact. Under mild conditions, we prove that the modified method with Armijo-type line search is globally convergent even if the objective function is nonconvex. We also present some numerical results to show the efficiency of the proposed method.

Mathematics Subject Classification (2000) 90C30 · 65K05

Supported by the 973 project (2004CB719402) and the NSF foundation (10471036) of China.

L. Zhang · W. Zhou (✉)
College of Mathematics and Computational Science,
Changsha University of Science and Technology, Changsha 410077, China
e-mail: zl606@tom.com

W. Zhou
e-mail: weijunzhou@126.com

D. Li
College of Mathematics and Econometrics, Hunan University, Changsha 410082, China
e-mail: dhli@hnu.cn

1 Introduction

Let $f : R^n \to R$ be continuously differentiable. Consider the unconstrained optimization problem:

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

We use $g(x)$ to denote the gradient of f at x. We are concerned with the conjugate gradient methods for solving (1.1). Let x_0 be the initial guess of the solution of problem (1.1). A conjugate gradient method generates a sequence of iterates by letting

$$x_{k+1} = x_k + \alpha_k d_k, \quad k = 0, 1, \ldots, \tag{1.2}$$

where the steplength α_k is obtained by carrying out some line search, and the direction d_k is defined by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if} \quad k > 0, \end{cases} \tag{1.3}$$

where β_k is a parameter such that when applied to minimize a strictly convex quadratic function, the directions d_k and d_{k-1} are conjugate with respective to the Hessian of the objective function. The Fletcher–Reeves (FR) method is a well-known conjugate gradient method. In the FR method, the parameter β_k is specified by

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

where g_k is the abbreviation of $g(x_k)$ and $\|\cdot\|$ stands for the Euclidean norm of vectors.

We see from (1.3) that for each $k \geq 1$, the directional derivative of f at x_k along direction d_k is given by

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

It is clear that if exact line search is used, then we have for any $k \geq 0$,

$$g_k^{\mathrm{T}} d_k = -\|g_k\|^2 < 0.$$

Consequently, vector d_k is a descent direction of f at x_k. Zoutendijk [33] proved that the FR method with exact line search is globally convergent. Another line search that ensures descent property of d_k is the strong Wolfe line search, that is, α_k satisfies the following two inequalities

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

and

$$|d_k^{\mathrm{T}} g(x_k + \alpha_k d_k)| \leq \sigma |d_k^{\mathrm{T}} g_k|,$$

where $0 < \delta < \sigma < \frac{1}{2}$. Al-Baali [1] proved the global convergence of FR method with the strong Wolfe line search. Liu et al. [20] and Dai and Yuan [7] extended this result to $\sigma = \frac{1}{2}$. Recently, Hager and Zhang [15] proposed a new conjugate gradient method which generates descent directions. If line search is exact, this method reduces to the Hestenes–Stiefel (HS) method [17], that is, the parameter

$$\beta_k = \beta_k^{HS} \triangleq \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \quad \text{with } y_{k-1} = g_k - g_{k-1}.$$

Hager and Zhang [15] established a global convergence theorem for their method with Wolfe line search. Other conjugate gradient methods and their global convergence can be found in [5–8,11–14,16–19,22–25,27,28,30,31] etc.

In the case where Armijo-type line search or Wolfe-type line search is used, the descent property of d_k determined by (1.3) is in general not guaranteed. In order to ensure descent property, Dixon [9] and Al-Baali [1] suggested to use the steepest descent direction $-g_k$ instead of d_k determined by (1.3) in the case where d_k is not a descent direction. By the use of this hybrid technique, Dixon [9] and Al-Baali [1] obtained the global convergence of the conjugate gradient methods with some inexact line searches. Quite recently, Zhang et al. [32] proposed a descent modified Polak–Ribière–Polyak (PRP) method [24,26], i.e.,

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

and proved its global convergence with some kind of Armijo-type line search.

Birgin and Martínez [3] proposed a spectral conjugate gradient method by combining conjugate gradient method and spectral gradient method [29] in the following way:

$$d_k = -\theta_k g_k + \beta_k d_{k-1},$$

where θ_k is a parameter and

$$\beta_k = \frac{(\theta_k y_{k-1} - s_{k-1})^{\mathrm{T}} g_k}{d_{k-1}^{\mathrm{T}} y_{k-1}}.$$

The reported numerical results show that the above method performs very well if θ_k is taken to be the spectral gradient:

$$\theta_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}, \tag{1.5}$$

where $s_{k-1} = x_k - x_{k-1}$. Unfortunately, the spectral conjugate gradient method [3] cannot guarantee to generate descent directions. So in [2], based on the quasi-Newton BFGS update formula, Andrei proposed scaled conjugate gradient algorithms that are descent method if Wolfe line search is used. However, it is not known whether the method with Armijo-type line search can generate descent directions.

In this paper, we take a little modification to the FR method such that the direction generated by the modified FR method is always a descent direction of the objective function. This property is independent of the line search used. Under mild conditions, we prove that this modified FR method with Armijo-type line search is globally convergent.

In the next section, we present the modified FR method. In Sect. 3, we prove the global convergence of the modified FR method. In Sect. 4, we report numerical results to test the proposed method. We also compare the performance of the modified FR method with the standard FR method, the steepest descent method, the CG_DESCENT method [15] and the PRP$^+$ [13] method by using the test problems in the CUTE [4] library.

2 Algorithm

In this section, we describe the modified FR method whose form is similar to that of [3] but with different parameters θ_k and β_k. Let x_k be the current iterate. Let d_k be defined by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -\theta_k g_k + \beta_k^{FR} d_{k-1}, & \text{if} \quad k > 0, \end{cases} \tag{2.1}$$

where β_k^{FR} is specified by (1.4) and

$$\theta_k = \frac{d_{k-1}^{\mathrm{T}} y_{k-1}}{\|g_{k-1}\|^2}, \tag{2.2}$$

where $y_k = g_{k+1} - g_k$. It is easy to see from (1.4), (2.1) and (2.2) that for any $k \geq 0$,

$$d_k^{\mathrm{T}} g_k = -\|g_k\|^2. \tag{2.3}$$

This implies that d_k provides a descent direction of f at x_k. It is also clear that if exact line search is used, then $g_k^{\mathrm{T}} d_{k-1} = 0$. In this case, we have

$$\theta_k = \frac{g_k^{\mathrm{T}} d_{k-1} - d_{k-1}^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} = -\frac{d_{k-1}^{\mathrm{T}} g_{k-1}}{\|g_{k-1}\|^2} = 1.$$

Consequently, the modified FR method reduces to the standard FR method.

Based on the above discussion, we propose a modified FR method which we call MFR method as follows.

Algorithm 1 (MFR method with Armijo-type line search)

Step 0 Given constants $\delta_1, \rho \in (0,1)$, $\delta_2 > 0$. Choose an initial point $x_0 \in R^n$. Let $k := 0$.
Step 1 Compute d_k by (2.1).
Step 2 Determine a stepsize $\alpha_k = \max\{\rho^{-j}, j = 0, 1, 2, \ldots\}$ satisfying

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta_1 \alpha_k g_k^{\mathrm{T}} d_k - \delta_2 \alpha_k^2 \|d_k\|^2. \tag{2.4}$$

Step 3 Let the next iterate be $x_{k+1} = x_k + \alpha_k d_k$.
Step 4 Let $k := k + 1$ and go to Step 1.

Since d_k is a descent direction of f at x_k, Algorithm 1 is well defined.

3 Global convergence

In this section, we prove the global convergence of Algorithm 1 under the following assumption.

Assumption A
(1) The level set $\Omega = \{x \in R^n | f(x) \leq f(x_0)\}$ is bounded.
(2) In some neighborhood N of Ω, f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leq L\|x - y\|, \quad \forall x, y \in N. \tag{3.1}$$

Since $\{f(x_k)\}$ is decreasing, it is clear that the sequence $\{x_k\}$ generated by Algorithm 1 is contained in Ω. In addition, we can get from Assumption A that there is a constant $\gamma_1 > 0$, such that

$$\|g(x)\| \leq \gamma_1, \quad \forall x \in \Omega. \tag{3.2}$$

In the latter part of the paper, we always suppose that the conditions in Assumption A hold. Without specification, we let $\{x_k\}$ and $\{d_k\}$ be generated by Algorithm 1.

Lemma 3.1 *There exists a constant $c_1 > 0$ such that the following inequality holds for all k sufficiently large,*

$$\alpha_k \geq c_1 \frac{\|g_k\|^2}{\|d_k\|^2}. \tag{3.3}$$

Proof We have from (2.4) and Assumption A that

$$\sum_{k \geq 0} (-\delta_1 \alpha_k g_k^{\mathrm{T}} d_k + \delta_2 \alpha_k^2 \|d_k\|^2) < \infty.$$

This together with equality (2.3) yields

$$\sum_{k\geq 0}\alpha_k^2\|d_k\|^2 < \infty \tag{3.4}$$

and

$$\sum_{k\geq 0}\alpha_k\|g_k\|^2 = -\sum_{k\geq 0}\alpha_k g_k^{\mathrm{T}} d_k < \infty. \tag{3.5}$$

In particular, we have

$$\lim_{k\to\infty}\alpha_k\|d_k\| = 0, \quad \text{and} \quad \lim_{k\to\infty}\alpha_k\|g_k\|^2 = 0. \tag{3.6}$$

We now prove (3.3) by considering the following two cases.

Case (i): $\alpha_k = 1$. We get from (2.3) $\|g_k\| \leq \|d_k\|$. In this case, inequality (3.3) is satisfied with $c_1 = 1$.

Case (ii): $\alpha_k < 1$. By the line search step, i.e., Step 2 of Algorithm 1, $\rho^{-1}\alpha_k$ does not satisfy inequality (2.4). This means

$$f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k) > \delta_1\alpha_k\rho^{-1}g_k^{\mathrm{T}}d_k - \delta_2\rho^{-2}\alpha_k^2\|d_k\|^2. \tag{3.7}$$

By the mean-value theorem and inequality (3.2), there is a $t_k \in (0,1)$ such that $x_k + t_k\rho^{-1}\alpha_k d_k \in N$ and

$$\begin{aligned}
&f(x_k + \rho^{-1}\alpha_k d_k) - f(x_k)\\
&\quad = \rho^{-1}\alpha_k g(x_k + t_k\rho^{-1}\alpha_k d_k)^{\mathrm{T}} d_k\\
&\quad = \rho^{-1}\alpha_k g_k^{\mathrm{T}} d_k + \rho^{-1}\alpha_k(g(x_k + t_k\rho^{-1}\alpha_k d_k) - g_k)^{\mathrm{T}} d_k\\
&\quad \leq \rho^{-1}\alpha_k g_k^{\mathrm{T}} d_k + L\rho^{-2}\alpha_k^2\|d_k\|^2.
\end{aligned}$$

Substituting the last inequality into (3.7), we get

$$\alpha_k > \frac{(1-\delta_1)\rho\|g_k\|^2}{(L+\delta_2)\|d_k\|^2}.$$

Letting $c_1 = \min\left\{1, \frac{(1-\delta_1)\rho}{L+\delta_2}\right\}$, we get (3.3).

From inequalities (3.3) and (3.5), we can easily obtain the following Zoutendijk condition.

Lemma 3.2 *We have*

$$\sum_{k\geq 0}\frac{\|g_k\|^4}{\|d_k\|^2} < +\infty. \tag{3.8}$$

We now establish the global convergence theorem for Algorithm 1.

Theorem 3.3 *We have*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.9}$$

Proof For the sake of contradiction, we suppose that the conclusion is not true. Then there exists a constant $\varepsilon > 0$ such that

$$\|g_k\| \geq \varepsilon, \quad \forall k \geq 0. \tag{3.10}$$

We get from (2.1) that

$$\|d_k\|^2 = (\beta_k^{FR})^2 \|d_{k-1}\|^2 - 2\theta_k d_k^T g_k - \theta_k^2 \|g_k\|^2.$$

Dividing both sides of this equality by $(g_k^T d_k)^2$, we get from (2.3), (3.10) and (1.4) that

$$\begin{aligned}
\frac{\|d_k\|^2}{\|g_k\|^4} = \frac{\|d_k\|^2}{(g_k^T d_k)^2} &= (\beta_k^{FR})^2 \frac{\|d_{k-1}\|^2}{(g_k^T d_k)^2} - \frac{2\theta_k}{d_k^T g_k} - \frac{\theta_k^2 \|g_k\|^2}{(g_k^T d_k)^2} \\
&= \left(\frac{\|g_k\|^2}{\|g_{k-1}\|^2}\right)^2 \frac{\|d_{k-1}\|^2}{\|g_k\|^4} + \frac{2\theta_k}{\|g_k\|^2} - \frac{\theta_k^2}{\|g_k\|^2} \\
&= \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} - \frac{1}{\|g_k\|^2}(\theta_k^2 - 2\theta_k + 1 - 1) \\
&= \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} - \frac{(\theta_k - 1)^2}{\|g_k\|^2} + \frac{1}{\|g_k\|^2} \\
&\leq \frac{\|d_{k-1}\|^2}{\|g_{k-1}\|^4} + \frac{1}{\|g_k\|^2} \\
&\leq \sum_{i=0}^{k-1} \frac{1}{\|g_i\|^2} \leq \frac{k}{\varepsilon^2}.
\end{aligned}$$

The last inequalities implies

$$\sum_{k\geq 1} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \varepsilon^2 \sum_{k\geq 1} \frac{1}{k} = \infty,$$

which contradicts Zountendijk condition (3.8). The proof is then complete.

4 Numerical experiments

In this section, we report some numerical experiments. We test Algorithm 1 on problems in the CUTE [4] library and compare its performance with that of the steepest descent method, the standard FR method, the CG_DESCENT method [15] and the PRP$^+$ [13] method. We test the performance of these methods with different Armijo-type line search and Wolfe line search, respectively. Note that since the direction d_k generated by the standard FR method with Armijo-type line search may not be descent, in the case where an ascent direction is generated, we restart the FR method by setting $d_k = -g_k$. The stop criterions are given below: for the methods with Armijo-type line search, we stop the iteration if the iteration number exceeds 2×10^4 or the function evaluation number exceeds 4×10^5 or the inequality $\|g(x_k)\|_\infty \leq 10^{-6}$ is satisfied; for the methods with Wolfe-type line search, we stop the iteration if the inequality $\|g(x_k)\|_\infty \leq 10^{-6}$ is satisfied. The detailed numerical results are listed on the web site

http://blog.sina.com.cn/u/4928efd5010003cz.

Figures 1, 2 and 3 show the performance of these methods relative to CPU time, which were evaluated using the profiles of Dolan and Moré [10].

All codes were written in Fortran and run on PC with 2.66 GHz CPU processor and 1 GB RAM memory and Linux operation system. The PRP$^+$ code is coauthored by Liu, Nocedal and Waltz and the CG_DESCENT code is coauthored by Hager and Zhang. The PRP$^+$ [13] codes were obtained from Nocedal's web page at

http://www.ece.northwestern.edu/˜ nocedal/software.html,

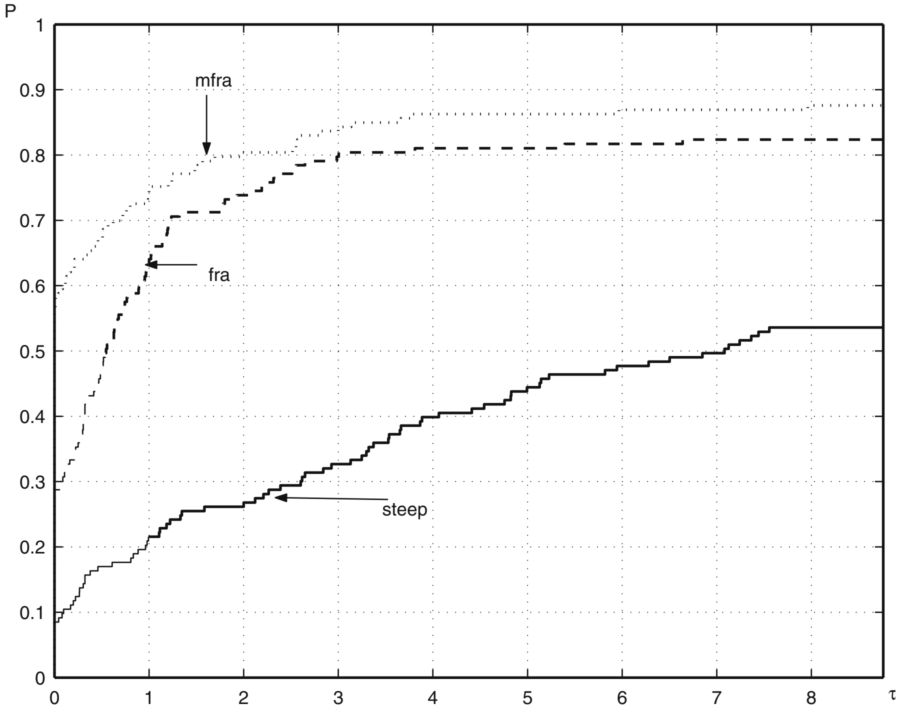

Fig. 1 Performance profiles

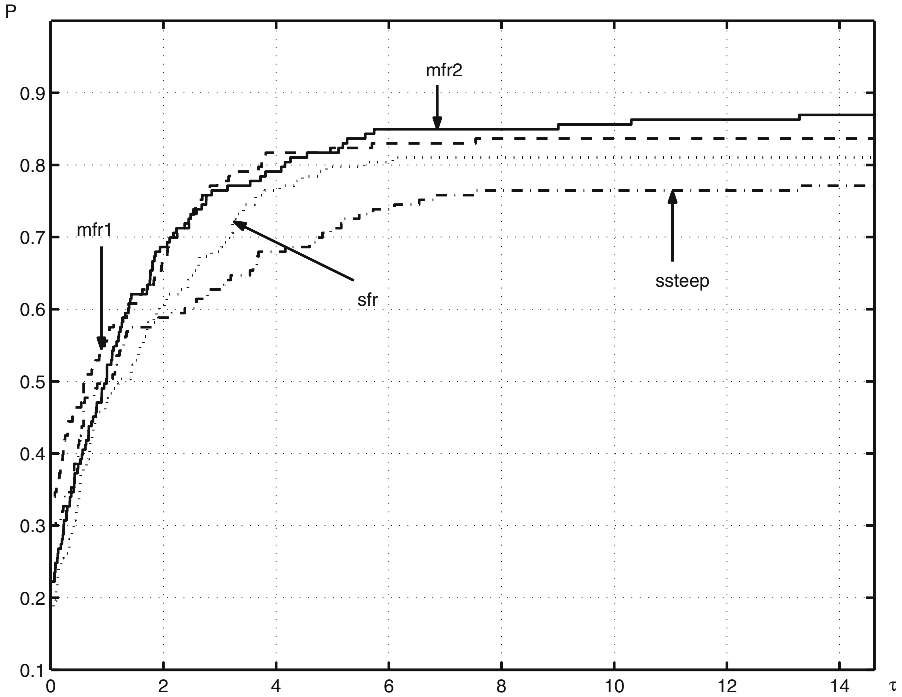

Fig. 2 Performance profiles

and the CG_DESCENT [15] code can be get from Hager's web page at

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

In Fig. 1, we compare the performance of the MFR method with the steepest descent method and the standard FR method. The line search used is the standard Armijo line search: determine a stepsize $\alpha_k = \max\{\rho^{-j}, j = 0, 1, 2, \ldots\}$ satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^{\mathrm{T}} d_k, \tag{4.1}$$

with $\delta = 10^{-3}$ and $\rho = 0.5$. In Fig. 1, "steep", "mfra" and "fra" stand for the steepest descent method, the MFR method and the standard FR method, respectively. Figure 1 shows that "mfra" outperforms "steep" and "fra" about 57% (87 out of 153) test problems. Moreover, "mfra" solves about 87% of the test problems successfully. The method "steep" performs worst since it can't solve many problems.

In Fig. 2,

- "steep" stands for the steepest descent method with the following modified Armijo line search. If $s_{k-1}^T y_{k-1} > 0$, we set $a_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}$, otherwise $a_k = 1$; determine a stepsize $\alpha_k = \max\{a_k \rho^{-j}, j = 0, 1, 2, \ldots\}$ satisfying

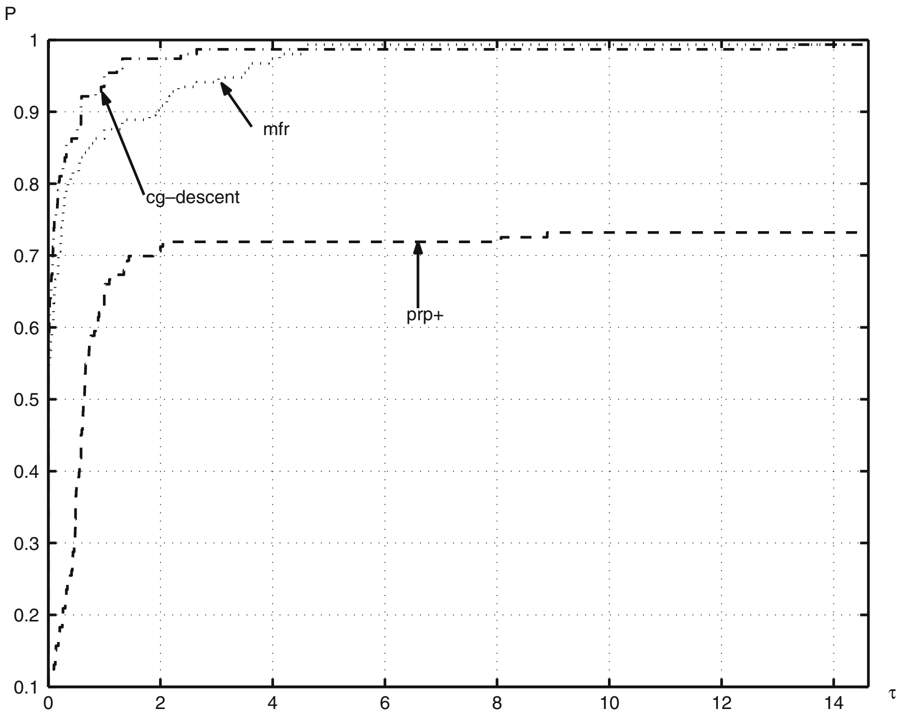

Fig. 3 Performance profiles

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^{\mathrm{T}} d_k, \tag{4.2}$$

where $\delta = 10^{-3}$ and $\rho = 0.5$. This initial steplength is called the spectral gradient [29].

- "sfr" stands for the standard FR method with the same line search as "steep".
- "mfr1" is the MFR method with the following line search rule. If $s_{k-1}^{\mathrm{T}} y_{k-1} > 0$, we set $a_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}$. Otherwise we set $a_k = 1$. Determine a stepsize $\alpha_k = \max\{a_k \rho^{-j}, j = 0, 1, 2, \ldots\}$ satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta_1 \alpha_k g_k^{\mathrm{T}} d_k - \delta_2 \alpha_k^2 \|d_k\|^2, \tag{4.3}$$

where $\delta_1 = 10^{-3}$, $\delta_2 = 10^{-8}$ and $\rho = 0.5$.

- "mfr2" is the MFR method with line search (2.4), where we set $\delta_1 = 10^{-3}$, $\delta_2 = 10^{-8}$ and $\rho = 0.5$.

We see from Fig. 2 that "mfr1" has the best performance since it solves about 33% (51 out of 153) of the test problems with the least CPU time. The method "ssteep" has the second best performance in which it solves about 30% (46 out of 153) of the test problems in the same situation. However, it fails to terminate successfully for many problems. The method “mfr2” can solve about 85% test problems successfully.

In Fig. 3, “cg-descent” stands for the CG_DESCENT method with the approximate Wolfe line search proposed by Hager and Zhang [15]; “mfr” is the MFR method with the same line search as “cg-descent”; “PRP+” means the PRP$^+$ method with the strong Wolfe line search proposed by Moré and Thuente [21]. We see from Fig. 3 that “cg-descent" performs slightly better than “mfr” does, “PRP+” performs worst.

Acknowledgements The authors would like to thank two anonymous referees for their valuable suggestions and comments, which improve this paper greatly. We are very grateful to Professors W. W. Hager and J. Nocedal for their codes

References

1. Al-Baali, M.: Descent property and global convergence of the Fletcher–Reeves method with inexact line search. IMA J. Numer. Anal. **5**, 121–124 (1985)
2. Andrei, N.: Scaled conjugate gradient algorithms for unconstrained optimization. Comput. Optim. Appl. (to apper)
3. Birgin, E., Martínez, J.M.: A spectral conjugate gradient method for unconstrained optimization. Appl. Math. Optim. **43**, 117–128 (2001)
4. Bongartz, K.E., Conn, A.R., Gould, N.I.M., Toint, P.L.: CUTE: constrained and unconstrained testing environments. ACM Trans. Math. Softw. **21**, 123–160 (1995)
5. Dai, Y.H., Yuan, Y.: Convergence properties of the conjugate descent method. Adv. Math. **25**, 552–562 (1996)
6. Dai, Y.H., Yuan, Y.: Convergence properties of the Fletcher–Reeves method. IMA J. Numer. Anal. **16**, 155–164 (1996)
7. Dai, Y.H., Yuan, Y.: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optim. **10**, 177–182 (2000)
8. Dai, Y.H.: Some new properties of a nonlinear conjugate gradient method. Research report ICM-98-010, Insititute of Computational Mathematics and Scientific/Engineering Computing. Chinese Academy of Sciences, Beijing (1998)
9. Dixon, L.C.W.: Nonlinear optimization: a survey of the state of the art. In: Evans, D.J. (eds.) Software for Numerical Mathematics pp. 193–216. Academic, New York (1970)
10. Dolan E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**, 201–213 (2002)
11. Fletcher, R.: Practical Methods of Optimization, Unconstrained Optimization. vol. I. Wiley, New York (1987)
12. Fletcher, R., Reeves, C.: Function minimization by conjugate gradients. J. Comput. **7**, 149–154 (1964)
13. Gilbert, J.C., Nocedal, J.: Global convergence properties of conjugate gradient methods for optimization. SIAM. J. Optim. **2**, 21–42 (1992)
14. Grippo, L., Lucidi, S.: A globally convergent version of the Polak–Ribiére gradient method. Math. Program. **78**, 375–391 (1997)
15. Hager, W. W., Zhang, H.: A new conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optim. **16**, 170–192 (2005)
16. Hager, W.W., Zhang, H.: A survey of nonlinear conjugate gradient methods. Pacific J. Optim. **2**, 35–58 (2006)
17. Hestenes, M.R., Stiefel, E.L.: Methods of conjugate gradients for solving linear systems. J. Res. Nat. Bur. Stand. Sect. B. **49**, 409–432 (1952)
18. Hu, Y., Storey, C.: Global convergence results for conjugate gradient methods. J. Optim. Theory Appl. **71**, 399–405 (1991)
19. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization methods. Math. Program. **45**, 503–528 (1989)

20. Liu, G., Han, J., Yin, H.: Global convergence of the Fletcher–Reeves algorithm with inexact line search. Appl. Math. J. Chin. Univ. Ser. B. **10**, 75–82 (1995)
21. Moré, J.J., Garbow, B.S., Hillstrome, K.E.: Testing unconstrained optimization software. ACM Trans. Math. Softw. **7**, 17–41 (1981)
22. Moré, J.J., Thuente, D.J.: Line search algorithms with guaranteed sufficient decrease. ACM Trans. Math. Softw. **20**, 286–307 (1994)
23. Nocedal, J.: Conjugate gradient methods and nonlinear optimization. In: Adams, L., Nazareth, J.L. (eds.) Linear and Nonlinear Conjugate Gradient-Related Methods. pp. 9–23. SIAM, Philadelphia (1996)
24. Polak, E.: Optimization: Algorithms and Consistent Approximations. Springer, Berlin Heidelberg New York (1997)
25. Polak, B., Ribiere, G.: Note surla convergence des méthodes de directions conjuguées. Rev. Francaise Imformat Recherche Opertionelle **16**, 35–43 (1969)
26. Polyak, B.T.: The conjugate gradient method in extreme problems. USSR Comp. Math. Math. Phys. **9**, 94–112 (1969)
27. Powell, M.J.D.: Nonconvex minimization calculations and the conjugate gradient method. In: Lecture Notes in Mathematics **1066**, 121–141 (1984)
28. Powell, M.J.D.: Convergence properties of algorithms for nonlinear optimization. SIAM Rev. **28**, 487–500 (1986)
29. Raydan, M.: The Barzilain and Borwein gradient method for the large unconstrained minimization problem. SIAM J. Optim. **7**, 26–33 (1997)
30. Sun, J., Zhang, J.:Convergence of conjugate gradient methods without line search. Ann. Oper. Res. **103**, 161–173 (2001)
31. Touati-Ahmed, D., Storey, C.: Efficient hybrid conjugate gradient techniques. J. Optim. Theory Appl. **64**, 379–397 (1990)
32. Zhang, L., Zhou, W.J., Li, D.H.: A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. IMA J. Numer. Anal. (to appear)
33. Zoutendijk, G. : Nonlinear programming, computational methods. In: Abadie, J. (eds.) Integer and Nonlinear Programming, pp. 37–86. North-Holland, Amsterdam (1970)