RESEARCH ARTICLE

ScienceAsia 39 (2013): **306–311**

doi: 10.2306/scienceasia1513-1874.2013.39.306

Global convergence of two spectral conjugate gradient methods

Mahdi Ghanbari, Tahir Ahmad*, **Norma Alias, Mohammadreza Askaripour**

Ibnu Sina Institute for Fundamental Science Studies, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

*Corresponding author, e-mail: tahir@ibnusina.utm.my

Received 19 Dec 2012
Accepted 8 May 2013

ABSTRACT: Two new nonlinear spectral conjugate gradient methods for solving unconstrained optimization problems are proposed. One is based on the Hestenes and Stiefel (HS) method and the spectral conjugate gradient method. The other is based on a mixed spectral HS-CD conjugate gradient method, which combines the advantages of the spectral conjugate gradient method, the HS method, and the CD method. The directions generated by the methods are descent directions for the objective function. Under mild conditions, we prove that the spectral conjugate gradient methods with an Armijo-type line search are globally convergent. Numerical results show the proposed methods are promising.

KEYWORDS: unconstrained optimization, inexact line search

INTRODUCTION

Unconstrained optimization problems have extensive applications, for example, in petroleum exploration, aerospace, and transportation[1–3]. The purpose of this paper is to study the global convergence properties and practical computational performance of two Hestenes and Stiefel (HS) spectral conjugate gradient methods for unconstrained optimization without restarts, and with suitable conditions.

Consider the following unconstrained optimization problem:

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

where $f : R^n \to R$ is continuously differentiable and its gradient is available. Iterative methods are widely used and the iterative formula is given by

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

where $x_k \in R^n$ is the kth approximation to the solution, α_k is a positive scalar and called the step-size which is determined by some line search and d_k is a search direction. There are many kinds of iterative methods that include Newton method[4,5], steepest descent method[6], and nonlinear conjugate gradient method[7,8]. The conjugate gradient methods are the most famous methods for solving (1), especially in the case of large scale optimization problems in scientific and engineering computation due to the simplicity of their iteration and low memory requirements. The search direction d_k is defined by

$$d_k = \begin{cases} -g_k, & k = 1, \\ -g_k + \beta_k d_{k-1}, & k \geqslant 2, \end{cases} \tag{3}$$

where $g_k = \nabla f(x_k)$ and β_k is a scalar which determines the different conjugate methods. The well-known formulae for β_k such as β_k^{HS} [9], β_k^{FR} [10], β_k^{PRP} [11], β_k^{CD} [12], β_k^{LS} [13], β_k^{DY} [14] can be found in many related papers. The convergence behaviour of conjugate gradient methods with these formulae under some different line search conditions has been widely studied by many authors (see Refs. 12–15). In the original HS method proposed by Hestenes and Stiefel[9] β_k is defined by

$$\beta_k^{\mathrm{HS}} = \frac{g_k^{\mathrm{T}} y_k}{y_k^{\mathrm{T}} d_{k-1}}, \tag{4}$$

where $y_k = g_k - g_{k-1}$. In practical computation, the HS method is generally believed to be one of the most efficient conjugate gradient methods. Recently, some modified HS formulae have been proposed[16–18]. In these methods, the search direction is constructed to possess the sufficient descent property, and the theory of global convergence is established with different line search strategies. Zhang[17] proposed a three-term HS conjugate gradient method (called TTHS), in which the direction d_k is given by

$$d_k = -g_k + \beta_k^{\mathrm{HS}} d_{k-1} + \theta_k y_{k-1}, \theta_k = \frac{g_k^{\mathrm{T}} d_{k-1}}{d_{k-1}^{\mathrm{T}} y_{k-1}}. \tag{5}$$

Fletcher[12] proposed the CD method, in which β_k is defined by

$$\beta_k^{\mathrm{CD}} = -\frac{\|g_k\|^2}{d_{k-1}^{\mathrm{T}} g_{k-1}}, \tag{6}$$

where $\|.\|$ denotes the Euclidean norm of vectors. An important property of the CD method is that the method will produce a descent direction under the strong Wolfe line search:

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \delta \alpha_k g_k^{\mathrm{T}} d_k, \tag{7}$$

$$d_k^{\mathrm{T}} g(x_k + \alpha_k d_k) \geqslant \sigma d_k^{\mathrm{T}} g_k, \tag{8}$$

where $0 < \delta < \sigma < 1$. Some good results from the CD method have also been reported in recent years[19–21]. Another popular method of solving problem (1) is the spectral gradient method, which was developed originally by Barzilai and Borwein[22]. Raydan[23] further introduced the spectral gradient method for potentially large-scale unconstrained optimization problems. Recently, Birgin and Martinez[24] proposed a spectral conjugate gradient method by combining the conjugate gradient method and spectral gradient method. The direction d_k is given by

$$d_k = -\theta_k g_k + \beta_k s_{k-1} \tag{9}$$

where $s_k = x_k - x_{k-1}$ and

$$\beta_k^{\mathrm{SP}} = \frac{(\theta_k y_{k-1} - s_{k-1})^{\mathrm{T}} g_k}{s_{k-1}^{\mathrm{T}} y_{k-1}},$$
$$\beta_k^{\mathrm{SPR}} = \frac{\theta_k y_{k-1}^{\mathrm{T}} g_k}{\alpha_k \theta_{k-1} g_{k-1}^{\mathrm{T}} g_{k-1}},$$
$$\beta_k^{\mathrm{SFR}} = \frac{\theta_k g_k^{\mathrm{T}} g_k}{\alpha_k \theta_{k-1} g_{k-1}^{\mathrm{T}} g_{k-1}}, \tag{10}$$

and θ_k is taken to be the spectral gradient and is computed from

$$\theta_k = \frac{s_{k-1}^{\mathrm{T}} s_{k-1}}{s_{k-1}^{\mathrm{T}} y_{k-1}}. \tag{11}$$

The numerical results show that these methods are very effective. Unfortunately, the spectral conjugate gradient method[24] cannot guarantee to generate descent directions. Hence, based on the FR formula, Zhang et al[25] modified the FR method so that the direction generated is always a descent direction. The d_k is defined by the following

$$d_k = \begin{cases} -g_k, & k = 1, \\ -\theta_k g_k + \beta_k^{\mathrm{FR}} d_{k-1}, & k \geqslant 2, \end{cases} \tag{12}$$

where $\theta_k = (d_{k-1}^{\mathrm{T}} y_{k-1})/(\|g_{k-1}\|^2)$. They proved that this method can guarantee to generate descent directions and is globally convergent. In this paper, motivated by the success of the spectral gradient method, we first propose a new spectral conjugate gradient method by combining the HS method and the spectral gradient method. The direction is given by (3) and

$$\beta_k = \begin{cases} \beta_k^{\mathrm{HS}}, & g_k^{\mathrm{T}} d_{k-1} > 0, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

$$\theta_k = 1 - \frac{|g_k^{\mathrm{T}} d_{k-1}|}{g_{k-1}^{\mathrm{T}} d_{k-1}}. \tag{14}$$

Then, another new spectral conjugate gradient method obtained by combining the HS method and CD is proposed. The direction is given by (3) and

$$\beta_k = \begin{cases} \beta_k^{\mathrm{HS}}, & g_k^{\mathrm{T}} d_{k-1} > 0, \\ \beta_k^{\mathrm{CD}}, & g_k^{\mathrm{T}} d_{k-1} \leqslant 0. \end{cases} \tag{15}$$

$$\theta_k = 1 - \frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}}. \tag{16}$$

Under some mild conditions, we give the global convergence of the new spectral conjugate gradient methods with an Armijo-type line search[26]. The rest of this paper is organized as follows. First, we propose the corresponding algorithms and give some assumptions and lemmas, which are usually used in the proof of the global convergence properties of nonlinear conjugate gradient methods. Then, the global convergence of the new spectral conjugate methods will be proven. Some numerical experiments will be done to test the efficiency, especially in comparison with the modified FR[25] and the spectral PRP methods[26] in the last part of the paper.

ALGORITHMS AND LEMMAS

In this section, we will give the following assumption on objective function, which have often been used in the literature to analyse the global convergence of nonlinear conjugate gradient method and the spectral conjugate gradient method with inexact line searches.

Assumption 1: the level set $\Omega = \{x | f(x) \leqslant f(x_1)\}$ is bounded, where x_1 is the starting point.

Assumption 2: in some neighbourhood N of Ω, the objective function is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \leqslant L \|x - y\| \ \text{ for } x, y \in N. \tag{17}$$

Now we present the new spectral conjugate gradient method.

Algorithm 1 (Spectral HS Conjugate Gradient)
Step 1: Given constant δ_1, $\rho \in (0,1)$, $\delta_2 > 0$, $\epsilon > 0$. Choose an initial point $x_1 \in R^n$, let $k = 1$.
Step 2: If $\|g_k\| \leqslant \epsilon$, then the algorithm stops. Otherwise, compute d_k by (3), (14) and β_k by (13).
Step 3: Compute step-size $\alpha_k = \max\{\rho^j, j = 0, 1, 2, \dots\}$ such that

$$f(x_k+\alpha_k d_k) \leqslant f(x_k)+\delta_1\alpha_k g_k^{\mathrm{T}} d_k-\delta_2\alpha_k^2 \|d_k\|^2 \tag{18}$$

Step 4: Set $x_{k+1} = x_k+\alpha_k d_k$, and $k = k+1$. Return to Step 2.

Algorithm 2 (Spectral HS-CD Conjugate Gradient)
Whole steps of the spectral HS-CD algorithm are defined as Algorithm 1 except d_k and β_k in Step 2 that are computed by (3), (16) and (15), respectively.

The following theorem shows that both Algorithms 1 and 2 possess a descent direction in each iteration.

Theorem 1 *Let the sequences g_k and d_k be generated by Algorithms 1 and 2 and let the step-size α_k be determined by any line search, then*

$$g_k^{\mathrm{T}} d_k < 0. \tag{19}$$

Proof: We can prove the conclusion by induction. From $\|g_1\|^2 = -g_1^{\mathrm{T}} d_1$, the conclusion (19) holds for $k = 1$. Now we assume that the conclusion is true for $k-1$ and $g_k \neq 0$, that is $g_{k-1}^{\mathrm{T}} d_{k-1} < 0$. In the following, we need to prove that the conclusion holds for k. If $g_k^{\mathrm{T}} d_{k-1} > 0$, then, from both Algorithms 1 and 2 $\beta_k = \beta_k^{\mathrm{HS}}$. From (4), (3), (14), (16) and our assumption $g_{k-1}^{\mathrm{T}} d_{k-1} < 0$ we have

$$\begin{aligned} g_k^{\mathrm{T}} d_k &= -\left(1-\frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}}\right)\|g_k\|^2 \\ &\quad + \frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{(g_k-g_{k-1})^{\mathrm{T}} d_{k-1}} g_k^{\mathrm{T}} d_{k-1} \\ &= -\|g_k\|^2 + \frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}} g_k^{\mathrm{T}} g_k \\ &\quad + \frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{(g_k-g_{k-1})^{\mathrm{T}} d_{k-1}} g_k^{\mathrm{T}} d_{k-1} \\ &= -\frac{d_{k-1}^{\mathrm{T}}(g_k-g_{k-1})}{(g_k-g_{k-1})^{\mathrm{T}} d_{k-1}}\|g_k\|^2 \\ &\quad + \frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}} g_k^{\mathrm{T}} g_k \\ &\quad + \frac{g_k^{\mathrm{T}}(g_k-g_{k-1})}{(g_k-g_{k-1})^{\mathrm{T}} d_{k-1}} g_k^{\mathrm{T}} d_{k-1} \\ &= \frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}} g_k^{\mathrm{T}} g_k < 0. \end{aligned} \tag{20}$$

If $g_k^{\mathrm{T}} d_{k-1} \leqslant 0$ then from Algorithm 2 $\beta_k = \beta_k^{\mathrm{CD}}$. From (6), (3), and (16), we have

$$\begin{aligned} g_k^{\mathrm{T}} d_k &= -\theta_k \|g_k\|^2 + \beta_k^{\mathrm{CD}} g_k^{\mathrm{T}} d_{k-1} \\ &= -\left[1-\frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}}\right]\|g_k\|^2 - \frac{\|g_k\|^2 g_k^{\mathrm{T}} d_{k-1}}{d_{k-1}^{\mathrm{T}} g_{k-1}} \\ &= -\|g_k\|^2 < 0. \end{aligned} \tag{21}$$

If $g_k^{\mathrm{T}} d_{k-1} \leqslant 0$ then from Algorithm 1 $\beta_k = 0$. From (3) and (14), we have

$$\begin{aligned} g_k^{\mathrm{T}} d_k &= -\theta_k \|g_k\|^2 = -\left(1+\frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}}\right)\|g_k\|^2 \\ &= -\|g_k\|^2 - \frac{g_k^{\mathrm{T}} d_{k-1}}{g_{k-1}^{\mathrm{T}} d_{k-1}}\|g_k\|^2 \\ &\leqslant -\|g_k\|^2 < 0. \end{aligned} \tag{22}$$

From (20), (21) and (22), we know that the conclusion (19) holds for all k. □

Proposition 1 *(See Ref. 26). Let $f : R^n \to R$ be a continuously differentiable function. Suppose that d is a descent direction of f at x. Then there exists J_0 such that*

$$f(x+\alpha d) \leqslant f(x) + \delta_1 \alpha g^{\mathrm{T}} d - \delta_2 \alpha^2 \|d\|^2, \tag{23}$$

where $\alpha = \rho^{J_0}$, g is the gradient vector of f at x, δ_1, $\rho \in (0,1)$, $\delta_2 > 0$ are given constant scalars.

Remark 1 From Proposition 1, it is known that both Algorithms 1 and 2 are well defined. In addition, it is easy to see that a larger descent magnitude can be obtained at each step by the modified Armijo-type line search (18). The following conclusion is given in Lemma 3.3 of Ref. 26.

Lemma 1 *With Assumptions 1 and 2 there exists a constant $m > 0$ such that*

$$\alpha_k > m\frac{|g_k^{\mathrm{T}} d_k|}{\|d_k\|^2}, \tag{24}$$

holds for all sufficiently large k.

Lemma 2 *Under Assumptions 1 and 2,*

$$\sum_{k\geqslant 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|g_k\|^2} < \infty, \tag{25}$$

$$\lim_{k\to\infty} \alpha_k^2 \|d_k\|^2 = 0. \tag{26}$$

Proof: From the line search rule (18) and Assumption 1, there exists a constant M such that

$$\sum_{k=0}^{n-1} -\delta_1 \alpha_k g_k^{\mathrm{T}} d_k + \delta_2 \alpha_k^2 \|d_k\|^2 \leqslant \sum_{k=0}^{n-1} (f(x_k) - f(x_{k+1})) < 2M. \quad (27)$$

Then from Lemma 1 we have

$$\begin{aligned} 2M &\geqslant \sum_{k=0}^{n-1} (-\delta_1 \alpha_k g_k^{\mathrm{T}} d_k + \delta_2 \alpha_k^2 \|d_k\|^2) \\ &\geqslant \sum_{k=0}^{n-1} \left(\delta_1 m \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} + \delta_2 m^2 \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^4} \|d_k\|^2 \right) \\ &\geqslant \sum_{k=0}^{n-1} (\delta_1 m + \delta_2 m^2) \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2}. \end{aligned} \quad (28)$$

Hence the first conclusion is proven. Since

$$\begin{aligned} 2M &\geqslant \sum_{k=0}^{n-1} (-\delta_1 \alpha_k g_k^{\mathrm{T}} d_k + \delta_2 \alpha_k^2 \|d_k\|^2) \\ &\geqslant \delta_2 \sum_{k=0}^{n-1} \alpha_k^2 \|d_k\|^2. \end{aligned} \quad (29)$$

The series $\sum_{k=0}^{n-1} \alpha_k^2 \|d_k\|^2$ is convergent. Thus $\lim_{k\to 0} \alpha_k^2 \|d_k\|^2 = 0$. The second conclusion is obtained. □

Lemma 3 *Suppose that Assumptions 1 and 2 hold. Consider both Algorithms 1 and 2, where α_k is obtain by the modified Armijo line search* (18)*. If there exists a constant $\epsilon > 0$ such that for all $k > 0$, $\|g_k\| > \epsilon$. Then there exists a sufficiently large number k_0 such that for $k \geqslant k_0$, the scalars β_k in Algorithms 1 and 2 satisfy*

$$|\beta_k| \leqslant \left| \frac{g_k^{\mathrm{T}} d_k}{g_{k-1}^{\mathrm{T}} d_{k-1}} \right|. \quad (30)$$

Proof: If $g_k^{\mathrm{T}} d_{k-1} > 0$, from (2), (3), (17) and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned} |\beta_k| &= \left| \frac{g_k^{\mathrm{T}} (g_k - g_{k-1})}{(g_k - g_{k-1})^{\mathrm{T}} d_{k-1}} \right| \\ &\leqslant \frac{|g_k^{\mathrm{T}} (g_k - g_{k-1})|}{|-g_{k-1}^{\mathrm{T}} d_{k-1}|} \\ &\leqslant \frac{\|g_k\| \, \|g_k - g_{k-1}\|}{|g_{k-1}^{\mathrm{T}} d_{k-1}|} \\ &\leqslant \frac{\|g_k\| \, L\alpha_{k-1} \|d_{k-1}\|}{|g_{k-1}^{\mathrm{T}} d_{k-1}|.} \end{aligned} \quad (31)$$

From (26) in Lemma 2, it follows that

$$\lim_{k\to\infty} \alpha_{k-1}^2 \|d_{k-1}\|^2 = 0. \quad (32)$$

In particular, we have

$$\lim_{k\to\infty} \alpha_{k-1} \|d_{k-1}\| = 0. \quad (33)$$

Thus there exists a sufficient large number k_0 such that for $k \geqslant k_0$,

$$0 \leqslant \alpha_{k-1} \|d_{k-1}\| < \frac{|g_k^{\mathrm{T}} d_k|}{\|g_k\| \, L}. \quad (34)$$

Hence by substituting (34) in (31), we have (30).

If $g_k^{\mathrm{T}} d_{k-1} \leqslant 0$, then from (6), (15), (13) and (21), we have (30). □

GLOBAL CONVERGENCE PROPERTY

Theorem 2 *Under Assumptions 1 and 2,*

$$\liminf_{k\to\infty} \|g_k\| = 0. \quad (35)$$

Proof: Suppose that there exists a positive constant $\epsilon > 0$ such that $\|g_k\| > \epsilon$ for all k. From (3), it follows that

$$\begin{aligned} \|g_k\|^2 &= (-\theta_k g_k + \beta_k d_{k-1})^{\mathrm{T}} (-\theta_k g_k + \beta_k d_{k-1}) \\ &= \theta_k^2 \|g_k\|^2 - 2\theta_k \beta_k d_{k-1}^{\mathrm{T}} g_k + \beta_k^2 \|d_{k-1}\|^2 \\ &= \theta_k^2 \|g_k\|^2 - 2\theta_k (d_k^{\mathrm{T}} + \theta_k g_k^{\mathrm{T}}) g_k + \beta_k^2 \|d_{k-1}\|^2 \\ &= \theta_k^2 \|g_k\|^2 - 2\theta_k d_k^{\mathrm{T}} g_k - 2\theta_k^2 \|g_k\|^2 + \beta_k^2 \|d_{k-1}\|^2 \\ &= \beta_k^2 \|d_{k-1}\|^2 - 2\theta_k d_k^{\mathrm{T}} g_k - \theta_k^2 \|g_k\|^2. \end{aligned} \quad (36)$$

Dividing (36) by $(g_k^{\mathrm{T}} d_k)^2$ then from Lemma 3, there exists a sufficient large k_0 such that for $k \geqslant k_0$, we obtain

$$\begin{aligned} \frac{\|d_k\|^2}{(g_k^{\mathrm{T}} d_k)^2} &= \frac{\beta_k^2 \|d_{k-1}\|^2 - 2\theta_k d_k^{\mathrm{T}} g_k - \theta_k^2 \|g_k\|^2}{(g_k^{\mathrm{T}} d_k)^2} \\ &\leqslant \frac{(g_k^{\mathrm{T}} d_k)^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} \frac{\|d_{k-1}\|^2}{(g_k^{\mathrm{T}} d_k)^2} - \frac{2\theta_k}{g_k^{\mathrm{T}} d_k} - \theta_k^2 \frac{\|g_k\|^2}{(g_k^{\mathrm{T}} d_k)^2} \\ &= \frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} - \left(\theta_k \frac{\|g_k\|}{g_k^{\mathrm{T}} d_k} + \frac{1}{\|g_k\|} \right)^2 + \frac{1}{\|g_k\|^2} \\ &\leqslant \frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} + \frac{1}{\|g_k\|^2}. \end{aligned} \quad (37)$$

Table 1 The numerical results of the MFR (M), SHS-CD (C), SHS (S), and SPRP (P) methods.

problems	Dim	GV/10^{-6}				number of iterations				number of function evaluations			
		M	C	S	P	M	C	S	P	M	C	S	P
Brown almost-linear	100	6.54	7.60	7.60	6.54	24	29	29	24	2161	2126	2126	2161
Trigonometric function	100	9.99	9.87	9.64	7.24	589	818	239	146	18 959	26 211	1097	17 151
Linear function full rank	100	9.78	9.78	8.28	9.78	169	169	89	169	70	70	90	70
Linear function - rank 1	40	8.20	2.16	3.89	5.22	117	117	129	1499	23 383	23 305	33 553	469 714
Brown almost-linear	20	9.89	7.22	9.16	7.28	1254	833	1898	1680	63 366	33 982	93 890	89 935
Discrete boundary	20	9.57	9.84	9.66	8.61	1714	1769	7899	2699	38 461	47 241	166 750	50 001
Linear function	20	2.54	9.88	8.58	9.29	228	168	177	198	34 273	25 177	26 394	29 728
Penalty function II	20	5.91	6.97	5.65	9.23	250	899	699	4499	6948	51 848	22 771	140 659
Linear function full rank	12	3.66	1.47	1.64	1.68	8	11	11	6	97	73	99	25
Linear function - rank 1	10	7.40	1.29	2.64	9.48	42	44	35	74	4789	4843	3777	3026
Broyden tridiagonal	9	4.26	1.69	6.82	6.87	109	59	54	32	2125	2319	2121	6380
Variably dimensioned	8	1.33	1.47	2.32	4.25	45	21	21	21	1163	802	774	506
Extended power singular	8	9.76	9.74	3.84	2.57	846	526	9999	7899	10 199	5643	362 113	267 924
Biggs EXP6 function	6	6.30	3.98	6.08	9.31	20 999	19 999	31 999	31 899	975 041	796 701	277 479	238 730
Penalty function I	5	4.45	4.45	9.56	10.00	11 999	11 999	8999	6486	333 012	333 012	22 596	6530
Broyden tridiagonal	4	9.93	1.24	9.16	7.14	119	42	37	43	2864	783	694	816
Extended power singular	4	9.23	9.98	9.31	6.93	863	406	5989	3899	10 199	6030	216 981	13 206
Brown almost-linear	4	7.06	9.77	8.53	4.70	122	121	168	120	249	283	1984	1379
Discrete boundary value	4	9.42	4.21	7.23	7.70	35	27	39	32	724	561	771	843
Powell singular	4	9.60	9.18	5.49	7.79	667	408	679	2369	10 824	6030	24 736	80 201
Brown and Dennis	4	9.85	9.45	2.56	9.74	239	262	262	338	791	519	882	281
Wood function	4	9.71	8.05	7.49	3.41	292	198	254	209	18 669	11 208	14 688	270 532
Kowalik and Osboren function	4	6.92	7.80	8.18	9.78	9995	9899	2999	3044	416 535	415 016	28 014	24 340
Gaussian function	3	7.61	3.11	5.74	7.44	8	6	6	8	101	51	59	107
Bard function	3	9.78	9.81	6.33	-	715	614	1834	-	1102	590	36 939	-
Box three -dimensional	3	9.16	9.26	2.72	9.13	597	579	599	677	1190	1100	4681	1124
Helical valley	3	9.63	8.39	5.94	9.29	19	31	30	19	879	1493	1400	2613
Jennrich and Sampson	2	7.95	5.74	5.74	4.28	55	20	20	21	2183	1129	1129	882
Freudenstein and Roth	2	1.13	2.92	3.27	2.65	137	37	56	68	4739	1010	1461	704
Rosenbrock	2	4.98	1.68	9.45	3.55	238	66	175	106	2598	1824	1639	1073
Beal	2	7.56	5.51	2.93	6.93	74	39	45	45	1920	1165	1063	1263
Brown badly scan function	2	-	6.68	6.68	9.44	-	59	59	88	-	16 581	16 581	24 285

Dim: the dimension of the objective function; GV: the gradient value of the objective function when the algorithm stops; MFR: the modified FR conjugate gradient method in Ref. 25; SHS-CD: the new spectral HS-CD method presented in this paper; SHS: the new spectral HS method presented in this paper; SPRP: the spectral PRP conjugate gradient method in Ref. 26.

Therefore

$$\begin{aligned}\frac{\|d_k\|^2}{(g_k^{\mathrm{T}} d_k)^2} &\leqslant \frac{\|d_{k-1}\|^2}{(g_{k-1}^{\mathrm{T}} d_{k-1})^2} + \frac{1}{\|g_k\|^2} \\ &\leqslant \dots \leqslant \frac{\|d_{k_0}\|^2}{(g_{k_0}^{\mathrm{T}} d_{k_0})} + \sum_{i=k_0+1}^{k} \frac{1}{\|g_k\|^2} \\ &\leqslant \frac{c_0}{\epsilon^2} + \sum_{i=k_0+1}^{k} \frac{1}{\epsilon^2} = \frac{c_0 + k - k_0}{\epsilon^2},\end{aligned}$$

where $c_0 = \epsilon^2 \|d_{k_0}\|^2 / (g_{k_0}^{\mathrm{T}} d_{k_0})^2$ is a nonnegative constant. The last inequality implies

$$\sum_{k\geqslant 1} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} \geqslant \sum_{k\geqslant k_0} \frac{(g_k^{\mathrm{T}} d_k)^2}{\|d_k\|^2} > \epsilon^2 \sum_{k\geqslant k_0} \frac{1}{c_0 + k + k_0}.$$

The right-hand side of this is infinite which contradicts the result of Lemma 2. Hence the conclusion (35) holds. □

NUMERICAL EXPERIMENTS

In this section, we report some numerical results. Under the modified Armijo line search (18), we compare the performances of the iteration number and the function evaluation number of the new methods with that of modified FR[25], spectral PRP[26] on the given test problems which come from[27].

All codes were written in MATLAB 7.0.1 and were implemented on a PC with 2.0 GHz CPU, 1 GB RAM, and Windows 7. The parameters used were $\epsilon = 10^{-5}$, $\rho = 0.9$, $\delta_1 = 0.25$, $\delta_2 = 0.45$.

Comparison of the results in Table 1 shows that the proposed algorithms in this paper are promising.

CONCLUSIONS

In this paper, two new spectral HS conjugate gradient algorithms have been developed for solving unconstrained minimization problems. Under some mild conditions, the global convergence has been proven with an Armijo-type line search rule. Compared with the other similar algorithms, the numerical performances of the developed algorithms are promising.

Acknowledgements: The authors would like to thank the school of postgraduate studies of Universiti Teknologi Malaysia for funding this research under the SPS Fellowship grant.

REFERENCES

1. Hu ZZ, Hu XY, He ZX (2006) Pseudo-three-dimensional magnetotelluric inversion with nonlinear conjugate gradients. *Chin J Geophys* **49**, 1226–34.
2. Keane A, Nair P (2005) *Computational Approaches for Aerospace Design: The Pursuit of Excellence*, Wiley.
3. Xie C, Travis Waller S (2012) Stochastic traffic assignment, Lagrangian dual, and unconstrained convex optimization. *Transport Res B* **46**, 1023–42.
4. Frontini M, Sormani E (2003) Some variants of Newton's method with third-order convergence. *Appl Math Comput* **140**, 419–26.
5. Zban AYO (2004) Some new variants of Newton's method. *Appl Math Lett* **17**, 677–82.
6. Yuan Y (2004) *A New Stepsize for the Steepest Descent Method*, Research report, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences.
7. Dai YH, Yuan YX (2000) *Nonlinear Conjugate Gradient Methods*, Shanghai Scientific and Technical Publishers, Shanghai.
8. Nocedal J, Wright SJ (1999) *Numerical Optimization*, Springer-Verlag, New York.
9. Hestenes MR, Stiefel ER (1964) Methods of conjugate gradients for solving linear systems. *J Res Natl Bur Stand* **49B**, 409–32.
10. Fletcher R, Reeves C (1964) Function minimization by conjugate gradients. *Comput J* **7**, 149–54.
11. Polak B, Ribière G (1969) Note sur la convergence des méthodes de directions conjuguées. *Rev Française Inf Rech Opér Ser 3e Ann* **16**, 35–43.
12. Fletcher R (1987) Practical methods of optimization. In: *Unconstrained Optimization* vol I, Wiley, New York.
13. Liu YL, Storey CS (1991) Efficient generalized conjugate gradient algorithms. part 1: theory. *J Optim Theor Appl* **69**, 129–37.
14. Dai YH, Yuan Y (2000) A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J Optim* **10**, 177–82.
15. Polyak BT (1969) The conjugate gradient method in extreme problems. *USSR Comp Math Math Phys* **9**, 94–112.
16. Hager WW, Zhang H (2005) A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM J Optim* **16**, 170–92.
17. Zhang L, Zhou W, Li D (2007) Some descent three-term conjugate gradient methods and their global convergence. *Optim Meth Software* **22**, 697–711.
18. Dai Z-F (2011) Two modified HS type conjugate gradient methods for unconstrained optimization problems. *Nonlin Anal* **74**, 927–36.
19. Du X-W, Ye L-Q, Xu C-X (2001) Global convergence of a class of unconstrained optimal methods that include the conjugate descent method. *J Appl Math* **18**, 120–2.
20. Dai Y, Yuan Y (1996) Convergence properties of the conjugate descent method. *Adv Math* **25**, 552–62.
21. Pan C-Y, Chen L-P (2007) A class of efficient new descent methods. *Acta Math Applicatae Sin* **30**, 88–98.
22. Barzilai J, Borwein JM (1988) Two-point step size gradient methods. *IMA J Numer Anal* **8**, 141–8.
23. Raydan M (1997) The Barzilai and Borwein gradient method for the large scale unconstrained minimization problem. *SIAM J Optim* **7**, 26–33.
24. Birgin EG, Martínez JM (2001) A spectral conjugate gradient method for unconstrained optimization. *Appl Math Optim* **43**, 117–28.
25. Zhang L, Zhou W, Li D (2006) Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. *Numer Math* **104**, 561–72.
26. Jiang H, Deng S, Zheng X, Wan Zh (2012) Global convergence of a modified spectral conjugate gradient method. *J Appl Math* **2012**, 641276.
27. More JJ, Garbow BS, Hillstrom KE (1981) Testing unconstrained optimization software. *ACM Trans Math Software* **7**, 17–41.