International Journal of Engineering & Technology, 7 (3.28) (2018) 312-315

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Secant Condition Free of a Spectral Hestenses-Stiefel (SHS) Conjugate Gradient Method and its Sufficient Descent Properties

Usman Abbas Yakubu[1,2], Mustafa Mamat[2]*, Mohamad Afendee Mohamed[2], Puspa Liza Ghazali[3], Mohd Rivaie[4]

[1]Department of Mathematics, Northwest University, Kano. Nigeria
[2]Faculty of Informatics & Computing, Universiti Sultan Zainal Abidin (UniSZA), Kuala Terengganu, Malaysia
[3]Faculty of Economic and Management Sciences, Universiti Sultan Zainal Abidin (UniSZA), Kuala Terengganu, Malaysia
[4]Department of Computer Science and Mathematics, Universiti Teknologi MARA (UiTM), Kuala Terengganu, Malaysia
**Corresponding author E-mail: must@unisza.edu.my*

Abstract

The conjugate gradient method have been used widely to solve unconstrained minimization problems as a result of less storage locations and less computational expensive in dealing with the large-scale problems. In this work, we suggested a spectral HS conjugate gradient method without employing the secant condition and use some unconstrained problems with many variables to prove its sufficient descent as well as global convergence, the results is certified by apply exact line search procedure.

Keywords: *Global convergence; exact line search; spectral CG; secant condition; sufficient descent property.*

1. Introduction

In the paper [1] Raydan have introduced a spectral gradient method (SGM) for a large-scale unconstrained optimization problems and recently, paper [16] have presented spectral WYL CG method without using the secant condition and also Birgin and Martinez [3] purposed a spectral CG method and computed their spectral parameter using standard secant equation as initially acquainted by Barzilai and Borwein [11]. A spectral CG method normally combines the advantages of a spectral parameter and classical CG direction by creating a new search direction, see Du et al. [12], Zull et al. [13], Usman et al. [17], Abashar et al. [21], Abashar et al. [22] and Raydan et al. [1] for more details.
In this research work, we derived the spectral parameter θ_k by using a classical CG direction and a scalar parameter β_k of Hestenses-Stiefel (HS) method with no means of secant equations. A method is open to solve large-scale unconstrained minimization problems and compare its performance with classical HS, PRP and FR conjugate gradient methods. Therefore, we consider the general unconstrained optimization problem:

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

where $f: R^n \to R$ is continuous as well as differentiable, g_k is a gradient vector of a function f and initial point $x_0 \in R^n$ is usually solved iteratively according to the recurrence expression

$$x_{k+1} = x_k + \gamma_k d_k\,, \qquad k = 0,1,2,3,4,\ldots \tag{2}$$

where x_k stand for a current iteration, $\gamma_k > 0$ represent a step size obtained by line search method called exact line search, given as

$$\gamma_k = arg \min_{\gamma>0} f(x_k + \gamma d_k) \tag{3}$$

also d_k is a search direction defined as

$$d_k = \begin{cases} -g_k, & if\ k = 0 \\ -g_k + \beta_k d_{k-1}, & if\ k \geq 1 \end{cases} \tag{4}$$

and $g_k = \nabla f(x)$, parameter $\beta_k \in R$ are the gradient vector and CG coefficient respectively. A classical β_k of FR, PRP and HS are given below:

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{5}$$

$$\beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{\|g_{k-1}\|^2} \tag{6}$$

$$\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})} \tag{7}$$

where g_k and g_{k-1} in the above equations are gradient vectors of function f at points x_k, x_{k-1} respectively, and $\|.\|$ stand for a Euclidian norm. Consideration has been made on a global convergence of a CG methods above. In the paper [5] G. Zoutendijk proved FR method on a line search (3) converge globally. Later in paper [9-10] Powell disproved this assertion and noted clearly that the method have demonstrated a poor practical behaviour and its convergence is not global. In general PRP is the best CG method amongst others but their convergence analysis for a nonlinear function is uncertain [14]. Nonetheless, in paper [6] J. C. Gilbert and Nocedal established a global convergence result of the PRP method by limiting a scalar β_k to be non-negative that is $\beta_k{}^{PRP} = max\{0, \beta_k{}^{PRP}\}$. However, spectral CG method is worthy in terms of a numerical performances see also Raydan et al.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

[1], Yakubu et al. [17-18], Du et al. [12], Zull et al. [13], Hu [7], Wu [4] Abidin et al. [19], Mamat et al. [20] and Huang et al. [8]. In this research, we present a new spectral HS CG method without using the secant condition and tested its performance with classical formula HS, PRP and FR methods. The details of the method is presented in section 2, we demonstrate a sufficient descent condition and global convergence in section 3. Numerical results are given in section 4 based on CPU time and number of iterations. Conclusion follows immediately in section 5.

2. Detail of the New Method

Generally, the spectral CG method was proposed by Barzilai and Borwein [11], the direction d_k is generated as $d_k = -\theta g_k + \beta_k s_{k-1}$ where $s_{k-1} = \gamma_{k-1} d_{k-1}$ and θ_k is a spectral parameter derived using standard secant equation. The direction of our new method is defined as

$$d_k = \begin{cases} -g_k, & if\ k = 0 \\ -\dfrac{1}{\theta_k} g_k + \beta_k^{HS} d_{k-1}, & if\ k \geq 1 \end{cases} \quad (8)$$

From equation (8), we have $d_k = -\frac{1}{\theta_k} g_k + \beta_k^{HS} d_{k-1}$ $\rightarrow$ $d_k - \beta_k^{HS} d_{k-1} = -\frac{1}{\theta_k} g_k$

using the fact that $d_k = -g_k$ from (8) and substituting (7), we have

$$\frac{1}{\theta_k} = 1 + \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (g_k - g_{k-1}) g_k} d_{k-1} = 1 + \frac{g_k^T g_k - g_k^T g_{k-1}}{(d_{k-1}^T g_k - d_{k-1}^T g_{k-1}) g_k} d_{k-1}$$

Note that $g_k^T g_{k-1} = 0$ and $d_{k-1}^T g_k = 0$ implies that

$$\theta_k = \left(1 - \frac{g_k^T d_{k-1}}{d_{k-1}^T g_{k-1}}\right)^{-1} \quad (9)$$

where θ_k is a new spectral parameter computed without employing secant condition by using exact line search procedure.

Algorithm 2.1 (SHS Method)
Step 1: Given a starting point $x_0 \in R^n$ set $k = 0$
Step 2: Calculate β_k as given in formula (7) above
Step 3: Calculate d_k as given in (8). If $\|g_k\| = 0$, then stop.
Step 4: Calculate γ_k as given in (3).
Step 5: Update the new point as given in the recurrence expression(2).
Step 6: If $f(x_{k+1}) < f(x_k)$ and $\|g_k\| < \varepsilon$ then stop, otherwise go to step 1 with $k = k + 1$

3. Global Convergence Analysis

3.1. Sufficient descent condition

For a sufficient descent condition to hold, we consider

$$g_k^T d_k \leq -C\|g_k\|^2 \quad for\ k \geq 0\ and\ C \in (0,1] \quad (10)$$

Theorem 3.1: *Suppose a CG method with search direction* (8) *and* β_k^{HS} *given by* (7), *the condition* (10) *holds* $\forall\, k \geq 0$.

Proof: To prove the above theorem, we proceed by induction, since $g_0^T d_0 = -\|g_0\|^2$, the condition (10) satisfied as $k = 0$. Now we assume it is true for $k \geq 0$. Condition (10) as well hold, from equation (8), multiplying both sides by g_{k+1}^T

$$g_{k+1}^T d_{k+1} = -\frac{1}{\theta_k} g_{k+1}^T g_{k+1} + \beta_k^{HS} g_{k+1}^T d_k$$

$$g_{k+1}^T d_{k+1} = -\left(1 - \frac{g_k^T d_{k-1}}{d_{k-1}^T g_{k-1}}\right)\|g_{k+1}\|^2 + \beta_k^{HS} g_{k+1}^T d_k$$

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

It is known that $g_{k+1}^T d_k = 0$. Hence, (10) is true for $k + 1$. ■

3.2. Global convergence properties

The following assumptions applied on the objective function to psychoanalyze the global convergence of a general CG method.

Assumptions 3.1
(i) A level set $\Omega = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded, the function f is continuously differentiable in a neighborhood N of the level set Ω and x_0 is a starting point.
(ii) $g(x)$ is globally Lipschitz continuous in N that is $\exists$ a constant $L > 0$ such that $\|g(x) - g(y)\| \leq L\|x - y\|$ for any $x, y \in N$.

Lemma 3.1: *Suppose the assumptions 3.1 holds and consider any recurrence expression* (2) *and direction*(8), γ_k *is found in equation*(3). *Then, Zoutendijk condition below holds.*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty \quad (11)$$

Proof of this lemma can be found in [5].

Theorem 3.2: *Suppose the assumptions 3.1 holds, for any* $\{x_k\}$, $\{d_k\}$ *be given as spectral SH CG method,* γ_k *are determined by* (3) *and* β_k *can be found as in* (5). *Then*

$$\lim_{k\to\infty} \|g_k\| = 0 \quad (12)$$

Proof: From equation (8), we have $d_{k+1} + \frac{1}{\theta_k} g_{k+1} = \beta_k^{HS} d_k$ and square both sides of the equation

$$\left(d_{k+1} + \frac{1}{\theta_k} g_{k+1}\right)^2 = \left(\beta_k^{HS} d_k\right)^2$$

$$\|d_{k+1}\|^2 = \left(\beta_k^{HS}\right)^2 \|d_k\|^2 - \frac{2}{\theta_k} g_{k+1}^T d_{k+1} - \frac{1}{\theta_k^2}\|g_{k+1}\|^2 \quad (13)$$

Substituting (3) into(13), we obtain

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

$$\|d_{k+1}\|^2 = \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_k\|^2 - \frac{2}{\theta_k} g_{k+1}^T d_{k+1} - \frac{1}{\theta_k^2}\|g_{k+1}\|^2 \quad (14)$$

Also substituting $g_{k+1}^T d_{k+1} = -C\|g_{k+1}\|^2$ in (14), where $C = 1$ for this particular method

$$\|d_{k+1}\|^2 = \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_k\|^2 + \frac{2}{\theta_k}\|g_{k+1}\|^2 - \frac{1}{\theta_k^2}\|g_{k+1}\|^2$$

$$\|d_{k+1}\|^2 = \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_k\|^2 - \|g_{k+1}\|^2\left(\frac{1}{\theta_k^2} - \frac{2}{\theta_k}\right) \quad (15)$$

Multiply both side of equation (15) by $\frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}$, we have

$$\|d_{k+1}\|^2 \frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2} = \frac{\|g_{k+1}\|^2}{\|d_{k+1}\|^2}\left(\frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_k\|^2 - \|g_{k+1}\|^2\left(\frac{1}{\theta_k^2} - \frac{2}{\theta_k}\right)\right)$$

$$\frac{\|d_{k+1}\|^2\|g_{k+1}\|^2}{\|d_{k+1}\|^2} = \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}\left(\left(\frac{2}{\theta_k} - \frac{1}{\theta_k^2}\right) + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_k\|^2\right) \quad (16)$$

Substituting (9) in (16) and note that for exact line search $g_{k+1}^T d_k = 0$, we have

$$\frac{\|d_{k+1}\|^2\|g_{k+1}\|^2}{\|d_{k+1}\|^2} = \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}\left((2-1) + \frac{\|g_k\|^4}{\|g_{k-1}\|^4}\|d_k\|^2\right)$$

$$\frac{\|d_{k+1}\|^2\|g_{k+1}\|^2}{\|d_{k+1}\|^2} \le \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \qquad (17)$$

From the Lemma 3.1, $\lim_{k\to\infty}\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} < 0$. It implies that Theorem 3.2 does not hold true, then $\lim_{k\to\infty}\frac{(g_{k+1}^T d_{k+1})^2}{\|d_{k+1}\|^2} = \infty$ and from equation (17) this is true $\infty \le \frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2}$. So, Theorem 3.2 is true for a sufficient large k. ■

4. Results and Discussion

We have tested the Algorithm 2.1 (SHS Method) and compared its performance with a classical HS, FR and PRP methods. The comparisons are based on CPU time as well as number of iterations. The step size γ_k is found by exact line search, the stopping criteria used for both methods is $\varepsilon = 10^{-6}$ as suggested by Hillstrom and $\|g_k\| < \varepsilon$. A test problems in paper [15] N. Andrei were used, and we established a problems in Table 1 with four different initial values by MatlabR2015a subroutine programming using Intel® Core™ i5-3317U (1.7GHz with 4 GB (RAM)). In this research, we represent failure due to (i) memory requirement (ii) number of iterations exceed 1000. The results shown in Figure 1 and 2 are performance profile acquainted by paper [2] Dolan and More. Performances are established on CPU time and number of iterations respectively. Therefore, the highest value of a percentage of probability $P_s(t)$ will be regarded as the best performing method and also the method that reached the top foremost is considered most sophisticated upon all the methods.

Table 1: Test Problems functions

Functions	Dimensions	Initial Points
Trecanni	2	(2,2), (4,4), (5,5), (15,15)
Booth	2	(12,12), (-12,-12), (40,40), (-40,-40)
Zettl	2	(11,11), (-11,-11), (110,110), (-110,-110)
Leon	2	(3,3), (5,5), (7,7), (-7,-7)
Quartic	4	(22,22), (-22,-22), (50,50), (60,60)
Colville	4	(2,2), (13,13), (-13,-13), (19,19)
Wood	4	(5,5), (35,35),(50,50),(70,70)
Gen. Tridiagonal 1	10	(3,.,3), (-3,.,-3), (5,.,5), (-5,.,-5)
Gen. Tridiagonal 2	10	(3,.,3), (-3,.,-3), (11,.,11), (21,.,21)
Fletcher	10	(2,.,2), (-2,.,-2), (21,.,21),), (-21,.,-21)
Hager	2,4,10,100	(12,.,12), (-12,.,-12), (17,.,17), (-19,.,-19)
Quadratic Penalty QP1	2,4,10,100	(2,.,2),(-2,.,-2),(23,.,23),(-23,.,-23)
Raydan 1	2,4,10,100	(4,.,4), (-5,.,-5), (-7,.,-7), (-23,.,-23)
Ext. Tridiagonal 1	2,4,10,100,1000	(-5,.,-5), (21,.,21),(21,.,21), (25,.,25)
Extended Penalty	2,4,10,100,1000	(2,.,2), (8,.,8), (14,.,14), (25,.,-25)
Freud. and Roth	2,4	(4,.,-4), (8,.,8), (9,.,9), (10,.,10)
Himmelblau	2,4,10,100,1000, 10000,100000	(8,.,8), (-8,.,-8), (11,.,11), (-11,.,-11)
White and Holst	2,4,10,100,1000, 10000,100000	(5,.,5), (7,.,7), (-7,.,-7), (10,.,10)
Shallow	2,4,10,100,1000, 10000,100000	(3,.,3)(-3,.,-3), (-11,.,-11), (20,.,20)
Rosenbrock	2,4,10,100,1000, 10000,100000	(15,..,15),(-15,..,-15),(17,..,17), (-17,..,-17)

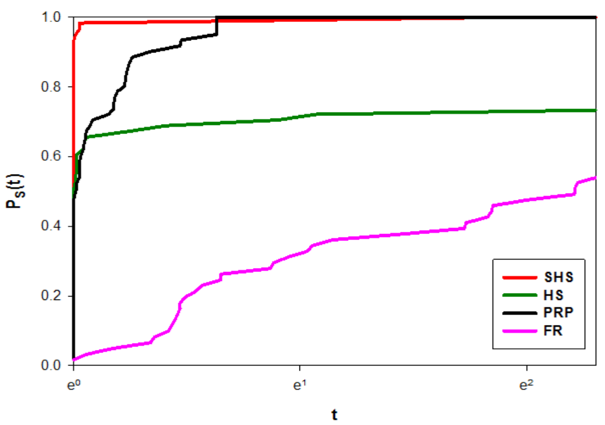

Fig. 1: Performance profile based on number of iterations

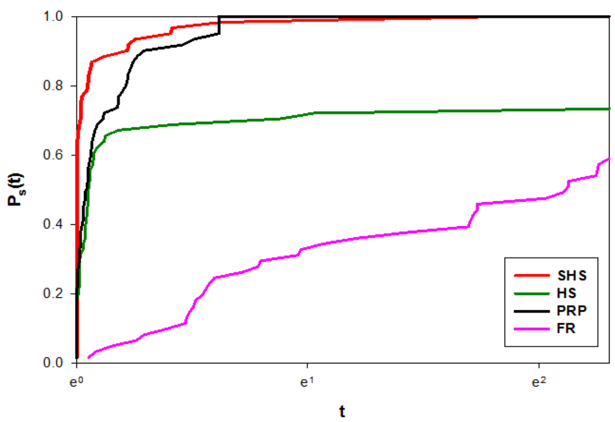

Fig. 2: Performance profile based on CPU time

Lastly, the new method has potentials when compared with the classical HS, PRP and FR methods based on that comparison in the figures shown, it shows that spectral HS method and PRP method successfully reached the solution points. Certainly, the efficiency of the spectral HS method is highly exhilarated.

5. Conclusion

A proposed spectral HS conjugate gradient method satisfies a sufficient descent conditions and converges globally. The method overcomes the weaknesses of HS, FR and PRP methods and presented numerical results apparently shows that the method performed wonderfully in terms of CPU time and number of iterations respectively.

Acknowledgement

The authors thank the referees for their contributions and also grateful to the Malaysian government and Universiti Sultan Zainal Abidin for funding this research under the Fundamental Research Grant Scheme (FRGS/1/2017/STG06/Unisza/01/1).

References

[1] M. Raydan (1997). The Barzilai and J.M. Borwein gradient methods for the large scale unconstrained minimization in extreme problems. SIAM. J. Optim., 7 (1), 26-33.
[2] E. Dolan and J. J. More (2002). Benchmarking optimization software with performance profile. Math. Prog., 91, 201-213.
[3] E.G. Birgin and J. M. Martinez (2011). A spectral conjugate gradient method for unconstrained optimization. Appl. Math. Optim., 43 (2), 117-128.
[4] X. Wu (2015). A new spectral Polak- Ribière -Polak conjugate gradient method. ScienceAsia, 41, 345-349.
[5] G. Zoutendijk (1970). Nonlinear programming, computational methods. Integer and Nonlinear Programming, 1970, 37-86.
[6] J. C. Gilbert and J. Nocedal (1992). Global convergence properties of conjugate gradient methods for optimization. SIAM. J. Optim., 2, 21-42.
[7] C. Hu and Z. Wan (2013). An extended spectral conjugate gradient method for unconstrained optimization problems. British Journal of Math. and Computer Science, 3, 86-98.
[8] H. Huang, Z. Wei and Y. Shengwei (2007). The proof of the sufficient decent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search. Applied Mathematics and Computation, 189, 1241–1245.
[9] M. J. D. Powell (1984). Non-convex minimization calculations and the conjugate gradient method. Lecture Notes in Mathematics, 1066, 122-241.
[10] M. J. D. Powell (1977). Restart procedures for the conjugate gradient method. Math. Program., 12, 241-254.
[11] J. Barzilai and J. M. Borwein (1988). Two-point step size gradient methods, IMA J Numer Anal, 8, 141-148.
[12] X. Du and J. Liu (2011). Global convergence of a spectral HS conjugate gradient method. Procedia Engineering, 15, 1487-1492.
[13] N. Zull, M. Rivaie, M. Mamat, Z Salleh and Z. Amani (2015). Global convergence of a spectral conjugate gradient by using strong Wolfe line search. Appl. Math. Sci., 63, 3105-3117.
[14] W. W. Hager and H. Zhang (2006). A survey of nonlinear conjugate gradient methods. Pacific Journal of Optimization, 2(1), 35-58.
[15] N. Andrei (2008). An unconstrained optimization test functions collection, Adv. Modell. Optim., 10, 147-161.
[16] A. Y. Usman, M. Mamat, M. Rivaie, A. M. Mohamad and B. Y. Rabi'u (2018). Secant free condition of a spectral WYL and its global convergence properties. Far East Journal of Mathematical Science, 103(12), 1889-1902.
[17] A. Y. Usman, M. Mamat, M. Rivaie, A. M. Mohamad and J. Sabi'u (2018). A recent modification on Dai-Liao conjugate gradient method for solving symmetric nonlinear equations. Far East Journal of Mathematical Science, 103(12), 1961-1974.
[18] K. U. Kamfa, M. Mamat, A. Abashar, M. Rivaie, P. L. B. Ghazali and Z. Salleh (2015). Another modified conjugate gradient coefficient with global convergence properties. Applied Mathematical Sciences, 9, 1833-1844.
[19] N. Z. Abidin, M. Mamat, B. Dangerfield, J. H. Zulkepli, M. A. Baten and A. Wibowo (2014). Combating obesity through healthy eating behavior: A call for system dynamics optimization. Plos One, 9(12), 1-17.
[20] M. Mamat, Y. Rokhayati, N. M. M. Noor and I. Mohd (2011). Optimizing human diet problem with puzzy price using fuzzy linear programming approach. Pakistan Journal of Nutrition, 10(6), 594-598.
[21] A. Abashar, M. Mamat, M. Rivaie and I. Mohd (2014). Global convergence properties of a new class of conjugate gradient method for unconstrained optimization. Applied Mathematical Sciences Issue, 65-68, 3307-3319.
[22] A. Abashar, M. Mamat, M. Rivaie, I. Mohd and O. Omer (2014). The proof of sufficient descent condition for a new type of conjugate gradient methods. AIP Conference Proceedings, 1602, 296-303.