Menemui Matematik (Discovering Mathematics)
Vol. 40, No. x1: 20-30 (2018)

Preconditioned Conjugate Gradient Methods with Sufficient Descent Condition for Large-scale Unconstrained Optimization

M. M. Ling[1] and W. J. Leong[2]
[1]*Department of Mathematics, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor*
[1]lingmm4559@hotmail.com, [2] leongwj@upm.edu.my

ABSTRACT

In this paper, we propose preconditioned conjugate gradient method by applying preconditioning technique to the search direction. Besides, we also enforce the search direction to satisfy sufficient descent and boundedness condition. These two conditions are important to ensure the global convergence of the proposed method. Numerical results on a set of unconstrained optimization problems showed that the proposed method is efficient for different conjugate gradient methods.

Keywords: Unconstrained Optimization, Conjugate Gradient Method, Sufficient Descent

INTRODUCTION

We consider the unconstrained optimization problem

$$\min f(x) \quad \text{subject to } x \in R^n , \tag{1}$$

where $f: R^n \to R$ is a continuously differentiable and n is a dimension of x that assumed to be large. The iterative formula of conjugate gradient method is given by

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

where $\alpha_k > 0$ is step length obtained by some line searches and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k + \beta_k d_{k-1}, & k \geq 1 \\ -g_{k,} & k = 0 \end{cases} , \tag{3}$$

where $g_k = \nabla f(x_k)$ denotes the gradient vector of $f(x)$ at x_k and β_k is a parameter which determines different conjugate gradient methods. The well-known conjugate gradient methods are the Fletcher-Reeves (FR) , Polak-Ribière-Polyak (PRP) and Dai-Yuan (DY) methods, in which β_k is defined as follows:

$$\beta_k^{FR} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} , \tag{4}$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \quad , \tag{5}$$

$$\beta_k^{DY} = \frac{\|g_k\|^2}{d_{k-1}^T y_{k-1}} \quad , \tag{6}$$

where $y_{k-1} = g_k - g_{k-1}$ and $\|\cdot\|$ denotes the Euclidean norm.

The step length can be obtained by using exact line search

$$\alpha_k = \arg\min \{ f(x_k + \alpha d_k) \mid \alpha > 0 \} \quad . \tag{7}$$

But it is expensive or impossible to get the exact line search, so in this paper, the Armijo inexact line search are used to find the step length, such as

$$f(x_k + \alpha d_k) \le f(x_k) + \delta \alpha g_k^T d_k \quad , \tag{8}$$

for the constant $\delta \in (0,1)$.

To ensure the global convergence of the modified methods, we enforce the search direction to satisfy the sufficient descent and boundedness condition. The sufficient descent condition which appears to guaranteeing the descent property of conjugate gradient (CG) method has been first considered by Gilbert and Nocedal. The sufficient descent condition is defined as

$$g_k^T d_k \le -C_1 \|g_k\|^2 \quad . \tag{9}$$

It is used to establish the global convergence of different algorithms with inexact line searches by Hager and Zhang and Nakarima et al. followed to modify the technique of Hager and Zhang by considering a unified formula of parameters.

Another key condition that has been frequently used in the convergence analysis on CG method is the boundedness condition on d_k:

$$\|d_k\| \le C_2 \|g_k\| \quad . \tag{10}$$

Powell suggested that a projection matrix, P can be used to precondition the CG direction in the following way:

$$d_k = -P_k g_k + \beta_k d_{k-1} \quad , \tag{11}$$

where

$$P_k = I - \frac{g_k g_k^T}{\|g_k\|^2} \quad . \tag{12}$$

However, the projection matrix proposed by Powell does not satisfy the quasi-Newton equation. Motivated by the above observation, we propose a preconditioner in a similar form as suggested by Powell, but would satisfy the quasi-Newton equation.

The paper is organized as follows. Next section presents the derivation of the proposed preconditioner, search direction and the algorithm of the proposed preconditioned conjugate gradient method. Then we prove the global convergence of the proposed method and the performance of the methods in comparison with different conjugate gradient methods are presented in following section. Some conclusions are included in last section.

PRECONDITIONED CONJUGATE GRADIENT METHOD

In this section, we present the proposed preconditioner which is in the similar form as Powell preconditioner. Initially we set our preconditioner P as:

$$P_k = I - \alpha \frac{u_k u_k^T}{\|u_k\|^2} \quad . \tag{13}$$

By the quasi-Newton relation, we have

$$H_{k+1} y_k = s_k \quad , \tag{14}$$

where $s_k = x_{k+1} - x_k$, $y_k = g_{k+1} - g_k$, and H_{k+1} is the inverse of the Hessian matrix.

To incorporate the quasi-Newton property into P, we let P satisfy the quasi-Newton equation, hence it gives namely

$$P_k y_k = s_k \quad , \tag{15}$$

and we have

$$\left(-\alpha \frac{u_k^T y_k}{\|u_k\|^2} \right) u_k = s_k - y_k \quad . \tag{16}$$

For simplicity, we set

$$-\alpha \frac{u_k^T y_k}{\|u_k\|^2} = 1 \quad , \tag{17}$$

and hence we obtain, from (16)

$$u_k = s_k - y_k \quad . \tag{18}$$

Substituting (18) into (17) gives

$$-\alpha = \frac{(s_k - y_k)^T (s_k - y_k)}{(s_k - y_k)^T y_k} \quad . \tag{19}$$

Then by substituting the (19) and (18) into (13), we have

$$P_k = I + \left[\frac{(s_k - y_k)^T (s_k - y_k)}{(s_k - y_k)^T y_k}\right]\left[\frac{(s_k - y_k)(s_k - y_k)^T}{(s_k - y_k)^T (s_k - y_k)}\right] \quad , \tag{20}$$

and lastly,

$$P_k = I + \frac{(s_k - y_k)(s_k - y_k)^T}{(s_k - y_k)^T y_k} \quad , \tag{21}$$

which is the memoryless SR1 update. From here, we showed that the proposed preconditioner is satisfy the quasi-Newton equation.

To ensure the satisfaction of sufficient descent condition and boundedness condition, our search direction is then defined as follow:

$$d_k = \begin{cases} -Pg_k + \beta_k^{FR} d_{k-1} & , \quad g_k^T d_k \le -C_1 \|g_k\|^2 \; and \; \|d_k\| \le C_2 \|g_k\| \\ -g_k & , \quad otherwise \end{cases} \quad . \tag{22}$$

The Preconditioned Conjugate Gradient Algorithm

The steps of the preconditioned algorithm are presented as follows:

Step 1: Choose an initial point $x_0 \in R^n$ and set $d_o = -g_0$. Given constants $\rho \in (0,1)$, let $k = 0$.

Step 2: Set $\varepsilon = 10^{-4}$, check $\|g_k\| \le \varepsilon$, if yes, stop. Else, proceed to step 3.

Step 3: Compute d_k by (22) where $C_1 = 0.01$ and $C_2 = 100$ are used.

Step 4: Find a step size $\alpha_k = \max\{\rho^{-j}, j = 0, 1, 2, \ldots\}$ ($\rho = \frac{1}{2}$ is chosen) satisfying

$$f(x_k + \alpha_k d_k) \le f(x_k) + \delta \alpha_k g_k^T d_k \quad . \tag{23}$$

Step 5: Let the next iteration be $x_{k+1} = x_k + \alpha_k d_k$.

Step 6: Let $k := k + 1$ then go to step 2.

CONVERGENCE ANALYSIS

In this paper, our focus is on the conjugate gradient method, which we enforce the search direction to satisfy the sufficient descent direction and the boundedness condition on search direction, d_k . Therefore, we shall look at the convergence properties of sufficient descent condition and the boundedness condition on direction, d_k . The convergence analysis is based on the convergence analysis given by Bryd et al. and Bryd and Nocedal.

In order to present convergence analysis, we make some assumptions about the objective function, f and the step size, α_k as follows:

Assumption 1

(1) The level set $\Omega = \{x \in R^n \mid f(x) \le f(x_1)\}$ is closed.

(2) The objective function, f is uniformly convex on Ω and there exist two positive constants m and M such that

$$m\|z\|^2 \le z^T G(x) z \le M\|z\|^2 \quad , \tag{24}$$

where $G(x) = \nabla^2 f(x)$ for all $x \in \Omega$ and all $z \in R^n$.

We use backtracking line search to find the step size, α . The algorithm of the backtracking line search is given as follows:

Algorithm

Step 1: Given the constants $\rho \in (0,1)$ and τ_1 , τ_2, with $0 < \tau_1 < \tau_2 < 1$.

Step 2: Set $\alpha = 1$.

Step 3: Test the step size, α by using

$$f(x_k + \alpha d_k) \le f(x_k) + \rho \alpha g_k^T d_k \quad . \tag{25}$$

Step 4: If (25) is satisfied, set $\alpha_k = \alpha$ and $x_{k+1} = x_k + \alpha_k d_k$. If not, find a new α in $[\tau_1\alpha, \tau_2\alpha]$ then go to step 3.

In our cases, we set $\tau_1 = \tau_2 = \frac{1}{2}$. From the backtracking line search, we state the following lemma which is due to Bryd and Nocedal.

Lemma 1 From the assumption 1, there exist positive constants c_1 and c_2 such that, for any x_k and d_k with $g_k^T d_k < 0$, the step size, α_k produced by the Algorithm will satisfy either

$$f(x_k + \alpha_k d_k) - f(x_k) \leq -c_1 \frac{(g_k^T d_k)^2}{\|d_k\|^2} \quad , \tag{26}$$

or

$$f(x_k + \alpha_k d_k) - f(x_k) \leq c_2 g_k^T d_k \quad . \tag{27}$$

Proof: See Bryd and Nocedal.

Since our methods will enforce the search direction to satisfy the sufficient descent and the boundedness condition, hence the global convergence of the proposed methods can be proven by proving that these two conditions can ensure the global convergence. The following definition is made by refer to Dussault.

Definition 1 A direction d_k is consider as sufficient descent if d_k satisfies the following conditions:

(i) Sufficient descent condition

$$g_k^T d_k \leq -c_3 \|g_k\|^2 \quad , \tag{28}$$

and

(ii) Boundedness condition on d_k

$$\|d_k\| \leq c_4 \|g_k\| \tag{29}$$

where c_3 and c_4 are positive constants.

Theorem 1 Let x_1 be the initial point which f satisfies the Assumption 1, where α_k is chosen from Algorithm 1. If (28) – (29) hold, then $\{x_k\} \to x^*$. Moreover, there is a constant $0 \leq r < 1$ such that

$$f(x_k) - f(x^*) \leq r^k [f(x_1) - f(x^*)] \tag{30}$$

holds for all k and

$$\sum_{k=1}^{\infty} \|x_k - x^*\| < \infty \quad . \tag{31}$$

Proof:

From Lemma 1, if (26) is satisfied, then by (28) - (29):

$$f(x_k + \alpha_k d_k) - f(x_k) \le -c_1 \frac{c_3^2 \|g_k\|^4}{c_4^2 \|g_k\|^2} = -\xi \|g_k\|^2 \quad , \tag{32}$$

where $\xi = \dfrac{c_1 c_3^2}{c_4^2}$.

On the other hand, if (27) is satisfied, we have

$$f(x_k + \alpha_k d_k) - f(x_k) \le -c_2 c_3 \|g_k\|^2 = -\xi \|g_k\|^2 \quad , \tag{33}$$

where $\xi = c_2 c_3$.

Thus, in either case, we can obtain

$$f(x_k + \alpha_k d_k) - f(x_k) \le -\xi \|g_k\|^2 \quad . \tag{34}$$

Then from Assumption 1, for all $k \ge 1$, it can be seen that

$$\frac{1}{2} m \|x_k - x^*\|^2 \le f(x_k) - f(x^*) \le \frac{1}{m} \|g_k\|^2 \quad . \tag{35}$$

From (34) and (35), we have

$$f(x_{k+1}) - f(x^*) \le r[f(x_k) - f(x^*)] \le \ldots \le r^{k+1}[f(x_1) - f(x^*)] \quad , \tag{36}$$

where $r = 1 - \xi m$. Since Algorithm 1 ensures that $f(x_k)$ is decreasing, then $0 \le r < 1$. Together with (35),

$$\begin{aligned} \sum_{k=1}^{\infty} \|x_k - x^*\| &\le \left(\frac{2}{m}\right)^{1/2} \sum_{k=1}^{\infty} [f(x_k) - f(x^*)]^{1/2} \\ &\le \left[\frac{2(f(x_1) - f(x^*)}{m}\right]^{1/2} \sum_{k=1}^{\infty} (r^{1/2})^k \\ &< \infty \quad . \end{aligned}$$

NUMERICAL RESULT AND DISCUSSION

In this section, the performance of the preconditioned conjugate gradient on a set of test functions is presented. The unconstrained test functions used in testing the unconstrained optimization are given by Andrei. The test functions we used are listed in below:

Table 1: Test functions

Raydan 1	Raydan
Extended Beale	Diagonal 1
Diagonal 2	Diagonal 3
Hager	Generalized Tridiagonal 1
Generalized Tridiagonal 2	Extended Tridiagonal 1
Diagonal 4	Diagonal 5
Extended Himmelblau	Generalized PSC1
Extended PSC1	Extended Powell
Extended BD1	Extended Maratos
Extended Quadratic Exponential EP1	Extended Tridiagonal 2
ENGVAL1	EDENSCH
NONSCOMP	QUARTC
Extended DENSCHNB	Extended DENSCHNF
Generalized Quartic	Full Hessian FH3
Diagonal 7	Diagonal 8
HIMMELBG	HIMMELH

Since we are interested in large-scale of unconstrained optimization problems, so the number of variable (n) are set at 1000, 5000 and 10000. The stopping criterion is set as

$$\|g_k\| \leq 10^{-4} \quad , \tag{37}$$

which means that the runs were terminated if the norm of the final gradient is below 10^{-4}. Besides that, the iteration will also force to stop if the number of iterations exceed 1000 or the numbers of function evaluations exceed 10000.

To validate the performance of the proposed method, comparison is made between Powell conjugate gradient method with the modified SR1 conjugate gradient method in term of number of iterations, number of function evaluations and the CPU time in seconds. For both Powell and SR1 method, we test for three different conjugate gradient method, which is FR method, PRP method and also DY method. The numerical results are simplified using the performance profile by Dolan and Moré. We evaluate the performance with that of

Powell-FR : Powell preconditioner in FR conjugate gradient method,
SR1-FR : SR1 preconditioner in FR conjugate gradient method,
Powell-PRP : Powell preconditioner in PRP conjugate gradient method,
SR1-PRP : SR1 preconditioner in PRP conjugate gradient method,
Powell-DY : Powell preconditioner in DY conjugate gradient method,
SR1-DY : SR1 preconditioner in DY conjugate gradient method.

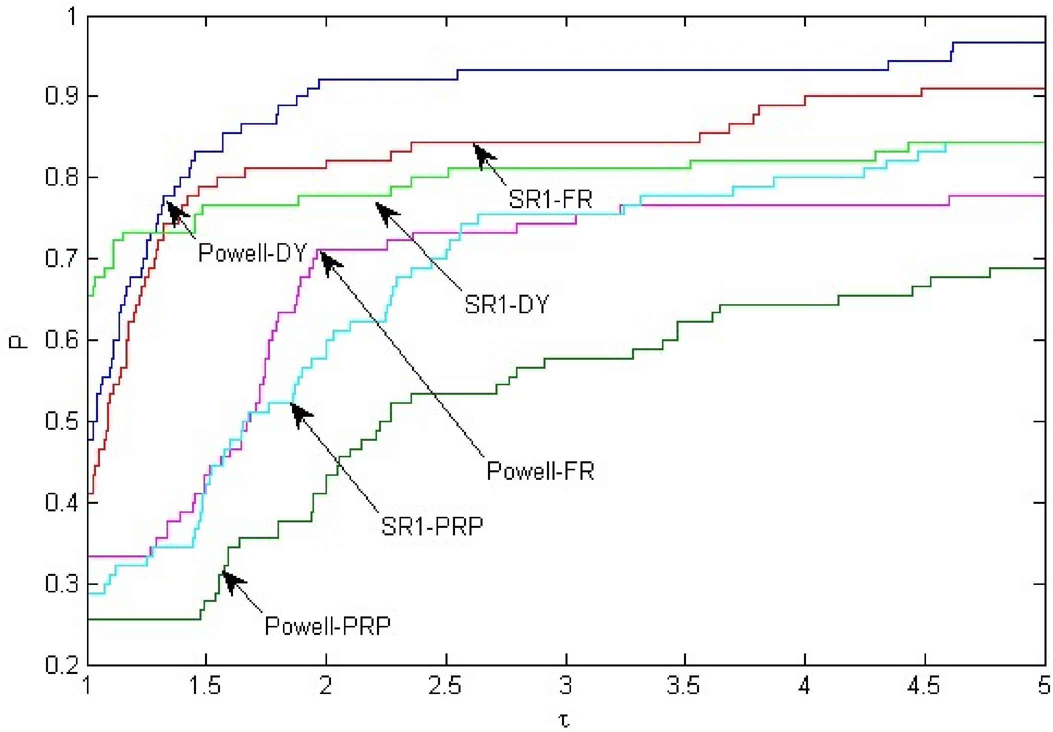

Figure 1: Performance profile based on number of iterations.

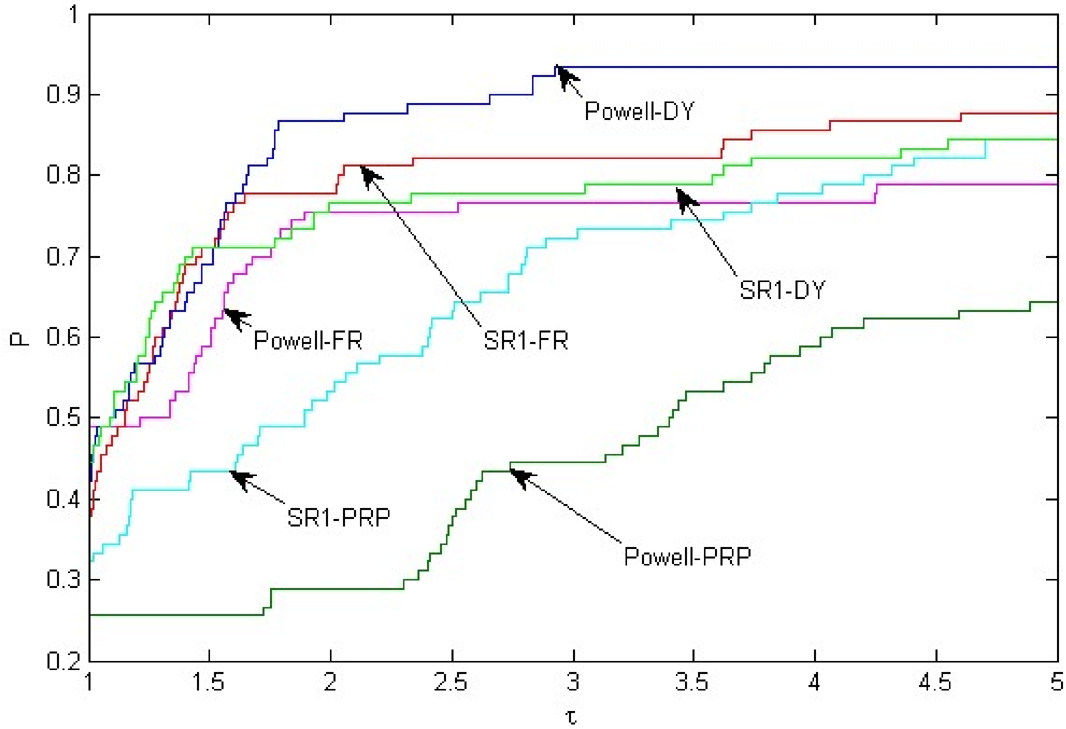

Figure 2: Performance profile based on number of function evaluations.

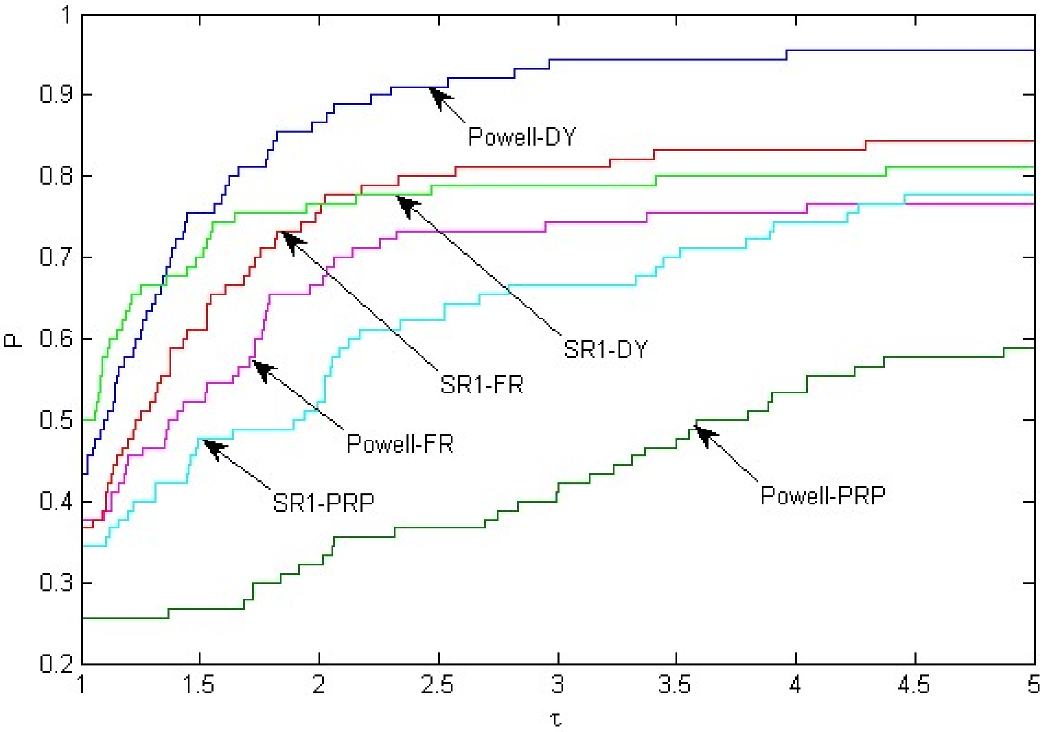

Figure 3: Performance profile based on CPU time

.

Figure 1-3 shows the performance profile in terms of number of iterations, number of function evaluations and CPU time. Based on Figure 1, the proposed SR1 preconditioner show better results in three different conjugate gradient methods compare to Powell preconditioner. SR1-FR solves 41% of the test functions compared to 33% by Powell-FR. SR1-DY improve almost 40% than Powell-DY. From Figure 2, Powell-FR, SR1-FR, Powell-DY and SR1-DY solve around 40% to 50% of the test functions.

We can clearly see that the SR1-PRP performs better in terms of number of iterations, number of function evaluations and CPU time compared to Powell-PRP. In overall, the performance has improved around 30%.

CONCLUSION

In this paper, we propose a preconditioner that satisfy the quasi-Newton equation. Besides, we also ensure the global convergence of the modified method by forcing the search direction to satisfy sufficient descent and boundedness condition. Numerical results showed that the proposed methods are promising and effective.

REFERENCES

Andrei, N. (2008), An Unconstrained Optimization Test Functions Collection, *Adv. Model. Optim.*, **10(2009)**: 147 – 161.

Byrd, R. H. and Nocedal, J. (1989), A Tool for the Analysis of Quasi-Newton Methods with Application to Unconstrained Minimization, *SIAM J. Numer. Anal.*, **26**: 727 – 739.

Byrd, R. H., Nocedal, J. and Yuan, Y. X. (1987), Global Convergence of a Class of Quasi-Newton Methods on Convex Problems, *SIAM J. Numer. Anal.*, **24**: 1171 – 1190.

Dai, Y. H. and Yuan, Y. (2000), A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property, *SIAM J. Optim.*, **10**: 177 – 182.

Dolan, E. D. and Moré, J. J. (2002), Benchmarking Optimization Software with Performance Profiles, *Mathematical Programming.*, **91**: 201 – 213.

Dussault, J. P. (2000), Convergence of Implementation Descent Algorithms for Unconstrained Optimization, *Journal of Optimization Theory and Applications.*, **104**: 739 – 745.

Fletcher, R. and Reeves, C. M. (1964), Function Minimization by Conjugate Gradients, *Comput. J.*, **7**: 149 – 154.

Gilbert, J. C. and Nocedal, J. (1992), Global Convergence Properties of Conjugate Gradient Methods for Optimization, *SIAM J. Optim.*, **2**: 21 – 42.

Hager, W. W. and Zhang, H. (2005), A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search, *SIAM J. Optim.*, **16**: 170 – 192.

Nakamura, W., Narushima, Y. and Yabe, H. (2013), Nonlinear Conjugate Gradient Methods with Sufficient Descent Properties for Unconstrained Optimization, *Journal of Industrial and Management Optimization.*, **9**: 595 – 619.

Polak, E. and Ribière, G. (1969), Note sur la convergence de directions conjuguèe, *Rev. Francaise Inf. Rech. Opèr. 3e Ann.,* **16**: 35 – 43.

Powell, M. J. (1977), Restart Procedures for the Conjugate Gradient Method, *Mathematical Programming.*, **12**: 241 – 254.