MATHEMATICAL MODELING AND COMPUTING, Vol. 9, No. 1, pp. 111–118 (2022)

A new modified conjugate gradient method under the strong Wolfe line search for solving unconstrained optimization problems

Ishak M. I., Marjugi S. M., June L. W.

Department of Mathematics, Universiti Putra Malaysia,
43400 Serdang, Selangor, Malaysia

(Received 7 July 2021; Revised 19 December 2021; Accepted 20 December 2021)

Conjugate gradient (CG) method is well-known due to efficiency to solve the problems of unconstrained optimization because of its convergence properties and low computation cost. Nowadays, the method is widely developed to compete with existing methods in term of their efficiency. In this paper, a modification of CG method will be proposed under strong Wolfe line search. A new CG coefficient is presented based on the idea of make use some parts of the previous existing CG methods to retain the advantages. The proposed method guarantees that the sufficient descent condition holds and globally convergent under inexact line search. Numerical testing provides strong indication that the proposed method has better capability when solving unconstrained optimization compared to the other methods under inexact line search specifically strong Wolfe–Powell line search.

Keywords: *conjugate gradient, global convergence, inexact line search, strong Wolfe–Powell line search, unconstrained optimization.*

2010 MSC: 65K10, 90C06, 90C30 **DOI:** 10.23939/mmc2022.01.111

1. Introduction

The conjugate gradient (CG) methods are capable to find the optimum solution for the nonlinear unconstrained optimization problems. It is widely used due to relatively little memory required for large-scale problems and no numerical linear algebra required, so each step is quite fast. For the unconstrained optimization problem as

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

where $f\colon \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable function, the CG method would construct iteratively according to

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

where $k = 0, 1, 2, \dots$ and d_k is the search direction, described by

$$d_k = \begin{cases} -g_k, & \text{if} \quad k = 0, \\ -g_k + \beta_k d_{k-1}, & \text{if} \quad k \geqslant 1, \end{cases} \tag{3}$$

where the current iterate point assigned as x_k, α_k is regarded as positive stepsize, g_k denotes the gradient coefficient and β_k is a scalar which is the CG coefficient. Many conjugate gradient methods have already been established including the Hestenes Stiefel (HS) method [1], the Fletcher–Reeves (FR) method [2], the Polak–Ribiere–Polyak (PRP) method [3] and [4], the conjugate descent (CD) method [5] and a lot of other recent methods where the corresponding coefficient beta β_k of the

This work was supported by Putra Grant no.vot 9544900 of Universiti Putra Malaysia.

mentioned methods are shown below

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

$$\beta_k^{\mathrm{RR}} = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}},$$

$$\beta_k^{\mathrm{PRP}} = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}},$$

$$\beta_k^{\mathrm{CD}} = \frac{g_k^T g_k}{d_{k-1}^T g_{k-1}}.$$

The important part in line search algorithm is a condition which is often called the sufficient descent property, defined in the form:

$$g_k^T d_k \leqslant -c \left\| g_k \right\|^2, \tag{4}$$

where the constant c plays a significant role which guarantees the global convergence of the nonlinear conjugate gradient method.

The properties of conjugate gradient methods that have been explored greatly through its global convergence properties. The global convergence of the FR method under the exact line has been proved by Zoutendijk [6]. It was later countered in an example by Powell [7]. Then in 1986, Powell [8] identified that the method is not able to possess global convergence properties. Before that in 1977, Powell [9] also proved that other methods are not preferable methods compared to FR method. Further research on FR method related to the global convergence was conducted by some researchers including Al-Baali [10], Touati–Ahmed and Storey [11] and Gilbert and Nocedal [12]. The researchers ran the FR method along with the inexact line search under a strong Wolfe condition. Strong Wolfe line search also guaranteed that other methods turned to be a globally convergent ones. Nowadays, a lot of attempts have been placed on proposing and structuring a modified formula from the existing CG methods through improvising the performances of numerical which possessed global convergent properties. In recent year, some researchers successfully did some modification on classical methods to improve the performance. X. Jiang and J. Jian succeed in improving the performance of FR and DY method [13] and P. Mtagulwa and P. Kaelo introduced modified PRP-FR hybrid conjugate gradient which is more efficient [14]. All the past and recent research give us some motivation to improve CG by modifying the β_k. Thus, this research focuses on the modification of CG through the exploration of new β_k performance under line search and section 2 will give further explanation on its algorithm. In the next section, we establish the sufficient descent condition with the global convergence proof of the new method. For the section 4, we highlight the numerical experiment results and also the discussion. Finally, we summed up and concluded them in section 5.

2. Methodology

Recently, Rivaie et al. [15] constructed a new coefficient β_k that is useful for handling the problem of non-convergence in which it is stated as the following:

$$\beta_k^{\mathrm{RMIL}} = \frac{g_k^T (g_k - g_{k-1})}{d_{k-1}^T (d_{k-1} - g_k)}. \tag{5}$$

The proposed β_k maintained the numerator as in the PRP, LS, and HS formula to deliver it restart properties as observed by Pytlak [16]. This formula is special because of its simplicity and capable to ensure that this method retained four important requirements of CG formula, global convergence properties, the sufficient descent conditions, angle conditions and linear convergence rate.

On the other hand, recently modification based on PRP method proposed recently by Hamoda et al. [17], namely, HRM. The formula are shown below

$$\beta_k^{\mathrm{HRM}} = \frac{g_k^T\left(g_k - \frac{\|g_k\|}{\|g_{k-1}\|}g_{k-1}\right)}{\mu\left\|g_{k-1}\right\|^2 + (1-\mu)\left\|d_{k-1}\right\|^2}. \tag{6}$$

The value of parameter μ is in the range within $0 < \mu < 1$. In this research scope, arbitrary value will be set as $\mu = 0.4$, as applied by the HRM method to test our new method.

Based on the properties of RMIL and HRM methods, a new modified β which is known as β_k^{ISL}, where ISL denotes Izwan, Siti Mahani, and Leong was proposed as follows

$$\beta_k^{ISL} = \frac{g_k^T\left(g_k - g_{k-1}\right)}{\mu\left\|g_{k-1}\right\|^2 + (1-\mu)\|d_{k-1}\|^2}. \tag{7}$$

By incorporating the proposed β_k (7), the new modified CG method algorithm is executed as:
Step 1: (Initialization). Given x_0, choose $k = 0$.
Step 2: Compute β_k, according to formula (7).
Step 3: Compute d_k by formula (3).
Step 4: Compute α_k by exact or inexact line search.
Step 5: Updating new point based on iterative formula (2).
Step 6: Convergent analysis test and stopping criteria: $\|g_k\| \leqslant \in$ then stop.
Or else go to Step 1 with $k = k + 1$.

3. Convergence analysis

In this section, we will show that our new modified CG guarantees the sufficient descent conditions hold and globally convergent for both exact and inexact line searches. However, we use the inexact line search or strong Wolfe–Powell line search to obtain the numerical results due to the fact that inexact line search is more practical compared to the exact line search [18].

Sufficient descent condition of ISL method with exact line search. The exact line search requires a condition as follows:

$$\min_{\infty > 0} f(x_k + \alpha d_k). \tag{8}$$

The next theorem asserts that the ISL direction with exact line search satisfies the sufficient descent condition.

Theorem 1. *Consider the sequences β_k and d_k established by the formula* (3) *and* (7)*, and the step length α_k is generated by the exact line search* (8) *then the sufficient descent condition* (4) *holds true for all $k \geqslant 0$.*

Proof. Theorem 1 shall be proved through induction; if $k = 0$ as $g_0^T d_0 = -C\left\|g_0\right\|^2$. Then, consider that condition (4) also holds true for some $k \geqslant 0$. For

$$g_{k+1}^T d_{k+1} = -\left\|g_{k+1}\right\|^2 + \beta_{k+1}^{\mathrm{ISL}} g_{k+1}^T d_k. \tag{9}$$

Since the exact line search implies that $g_{k+1}^T d_k = 0$, $g_{k+1}^T d_{k+1} = -\left\|g_{k+1}\right\|^2$. Thus, the sufficient descent condition holds, for all $k \geqslant 0$. ■

Sufficient descent condition of ISL method with inexact line search. Based on the previous research, the applicable and reliable inexact line search that commonly used is strong Wolfe–Powell line search (SWP) which contains two conditions

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \rho\,\alpha_k g_k^T d_k, \tag{10}$$

$$\left|g(x_k+\alpha_k d_k)^T d_k\right| \leqslant \sigma \left|g_k^T d_k\right|, \tag{11}$$

where $0<\rho<1$.

The subsequent theorem displays that the formula ISL with SWP line search will leads to the sufficient descent condition.

Theorem 2. *Suppose that the sequences $\{x_k\}$ and $\{d_k\}$ are generated by (2), (3) and (7), and the step length α_k is determined through the SWP line search (10) and (11). If $g_k \neq 0$, then the sequence $\{d_k\}$ satisfies the sufficient descent condition (4).*

Proof. Firstly, note that

$$\begin{aligned}
\beta_k^{\mathrm{ISL}} &= \frac{g_k^T(g_k-g_{k-1})}{\mu\left\|g_{k-1}\right\|^2+(1-\mu)\|d_{k-1}\|^2} \\
&= \frac{\left\|g_k\right\|^2-g_k^T g_{k-1}}{\mu\left\|g_{k-1}\right\|^2+(1-\mu)\|d_{k-1}\|^2} \\
&\geqslant \frac{\left\|g_k\right\|^2-\left\|g_k^T g_{k-1}\right\|}{\mu\left\|g_{k-1}\right\|^2+(1-\mu)\|d_{k-1}\|^2} \\
&\geqslant \frac{\left\|g_k\right\|^2-\left\|g_k^T\right\|\left\|g_{k-1}\right\|}{\mu\|g_{k-1}\|^2+(1-\mu)\left\|d_{k-1}\right\|^2} \geqslant 0.
\end{aligned}$$

On the other hand,

$$\frac{\left\|g_k\right\|^2+\left\|g_k^T g_{k-1}\right\|}{\mu\|g_{k-1}\|^2+(1-\mu)\left\|d_{k-1}\right\|^2} \leqslant \frac{\left\|g_k\right\|^2+\left\|g_k^T\right\|\left\|g_{k-1}\right\|}{\mu\|g_{k-1}\|^2+(1-\mu)\left\|d_{k-1}\right\|^2} \leqslant \frac{\left\|g_k\right\|^2}{\mu\left\|g_{k-1}\right\|^2}.$$

*Restart if $\left|g_k^T g_{k-1}\right| > c\|g_k\|^2$, $c\in[0,1)$.
*Not restart if $\left|g_k^T g_{k-1}\right| \leqslant c\|g_k\|^2$, $c\in(0,1)$.

Thus, by setting $\mu=0.4$ we get

$$0 \leqslant \beta_k^{\mathrm{ISL}} \leqslant \frac{2.5\|g_k\|^2}{\|g_{k-1}\|^2}. \tag{12}$$

Using (7) and (11),

$$\left|\beta_{k+1}^{\mathrm{lSL}} g_{k+1}^T d_k\right| \leqslant \frac{2.5\|g_{k+1}\|^2}{\|g_k\|^2}\sigma\left|g_k^T d_k\right|. \tag{13}$$

By (3), $d_{k+1}=-g_{k+1}+\beta_{k+1}d_k$

$$\frac{g_{k+1}^T d_{k+1}}{\left\|g_{k+1}\right\|^2} = -1+\beta_{k+1}\frac{g_{k+1}^T d_k}{\left\|g_{k+1}\right\|^2}. \tag{14}$$

We prove the descent property of d_k by induction. Since $g_0^T d_0 = -\|g_0\|^2<0$, if $g_0\neq 0$, now assume that d_i, $l=1,2,\ldots,k$, are all descent directions, that is $g_i^\tau d_i<0$.

By (13),

$$\left|\beta_{k+1}^{\mathrm{ISL}} g_{k+1}^T d_k\right| \leqslant \frac{2.5\|g_{k+1}\|^2}{\|g_k\|^2}\sigma\left(-g_k^T d_k\right). \tag{15}$$

That is,

$$\beta_{k+1}^{lSL} g_{k+1}^T d_k \leqslant -\frac{\left\|g_{k+1}\right\|^2}{\left\|g_k\right\|^2} - \frac{\left\|g_{k+1}\right\|^2}{\left\|g_k\right\|^2} 2.5\sigma g_k^T d_k \tag{16}$$

(12) and (14) deduce

$$-1+\frac{5\sigma g_k^T d_k}{2\left\|g_k\right\|^2} \leqslant \frac{g_{k+1}^T d_{k+1}}{2\left\|g_{k+1}\right\|^2} \leqslant -1-\frac{5\sigma g_k^T d_k}{2\left\|g_k\right\|^2}.$$

By repeating this procedure and the theory where $g_0^T d_0 = -\|g_0\|^2$, we have,

$$-\sum_{j=0}^{k}(2.5\sigma)^j \leqslant \frac{g_{k+1}^T d_{k+1}}{2\left\|g_{k+1}\right\|^2} \leqslant -2 + \sum_{j=0}^{k}(2.5\sigma)^j. \tag{17}$$

Since

$$\sum_{j=0}^{k}(2.5\sigma)^j < \sum_{j=0}(2.5\sigma)^j = \frac{1}{1-2.5\sigma}$$

(12) can be written as

$$-\frac{1}{1-2.5\sigma} \leqslant \frac{g_{k+1}^T d_{k+1}}{2\left\|g_{k+1}\right\|^2} - 2 + \frac{1}{1-2.5\sigma}. \tag{18}$$

By setting the restriction $\sigma \in (0, 0.1)$, we get $g_{k+1}^T d_{k+1} < 0$. Thus, by induction, $g_k^T d_k < 0$ holds for all $k \geqslant 0$.

Denote $c = 2 - \frac{1}{1-2.5\sigma}$ then $0 \leqslant c < 1$, and (18) can be expressed as

$$(c-2)\left\|g_k\right\|^2 \leqslant g_k^T d_k \leqslant -\left\|g_k\right\|^2. \tag{19}$$

Then, it is indicate that the sufficient descent condition holds. Thus, the theorem is proved. ■

4. Results and discussion

This section reveals the numerical results involving 20 different functions with different variable and initial point [19] coded on MATHLAB program version R2015b. We plotted some comparison graphs of the new modified conjugate gradient method with other classical and modified methods under strong Wolfe–Powell line search. We selected $\varepsilon = 10^{-6}$ and established the gradient value to act as stopping criteria which is $\|g_k\| < \varepsilon$ suggested by Hillsterm [20]. Thus, the termination of all computational experiments is active where $\|g_k\| \leqslant 10^{-6}$. Table 1 showed the problem functions used in the computational experiment. The experiment was performed on a PC with CPU processor specifically Intel (R) Core (TM) i5 2450M (2.50 GHz) under RAM capacity, 8 GB. For some problems, under consideration, numerical results are not available due to breakdown of the line search to compute the positive step size, then the process assumed as a failure. Numerical results are exploited in term of the CPU time and number of iteration. The result of the comparative experimental are shown in Figure 1 and Figure 2 respectively, applying a performance profile established by Dolan and More [21]. The performance profile is a bench-marking to analyze the performance of optimization methods.

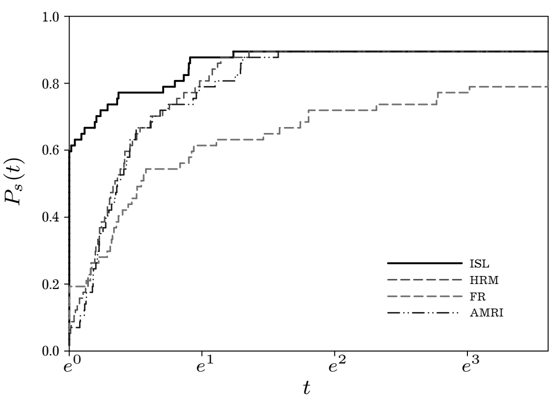

Fig. 1. Performance profile relative to the CPU time.

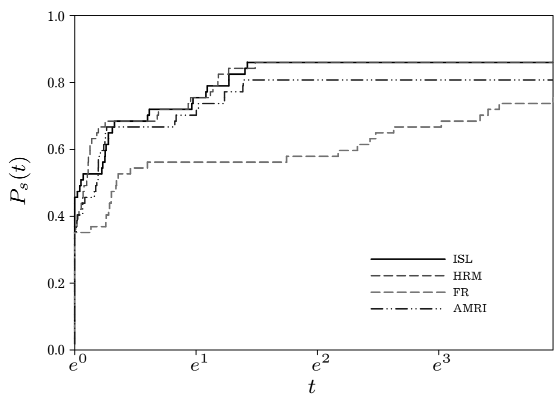

Fig. 2. Performance profile relative to the number of iterations.

Table 1. A list of problem functions.

No	Function	Dimension	Initial points
1	Booth	2	(10),(25),(50),(100)
2	Matyas Function	2	(1),(5),(10),(15)
3	Extended Freudenstem and Roth	2	(1),(2),(3),(7)
4	Extended Powell	2	(2),(10),(15),(50)
5	Extended Quadratic Penalty QP1 (0, 0,...)	2	(5),(7),(8),(30)
6	Extended Quadratic Penalty QP1 (1, 1,...)	2	(5),(7),(8),(30)
7	Power Function	2	(1),(10),(30),(50)
8	Sphere	2	(1),(10),(30),(50)
9	Ex-Penalty	2	(10),(30),(50),(80)
10	Hager	2	(1),(5),(7),(10)
11	Quadratic QF1	2, 4, 10	(3),(5),(8),(10)
12	Diagonal 4	2, 4, 10, 100, 500,1000	(1),(3),(6),(12)
13	Ex-Tridiagonal 1	2,4,10,100,500,1000,10000	(6),(12),(17),(20)
14	Perturbed Quadratic	2, 4, 10, 100, 500,1000	(1),(3),(5),(10)
15	Extended Denschnb	2,4,10,100,500,1000,10000	(8),(13),(30),(50)
16	Generalized Quartic	2,4,10,100,500,1000,10000	(1),(2),(5),(7)
17	Quadrtic QF2	2,4,10,100,500,1000	(5),(20),(50),(100)
18	Diagonal 2	2,4,10,100,500,1000	(1),(5),(10),(15)
19	Sum Squares	2,4,10,100,500,1000	(1),(3),(7),(10)
20	Generalized Tridiagonal 1	2,4,10,100	(7),(10),(13),(21)

The performance profile function is defined as,

$$\rho_s(t) = \frac{1}{n_p} \operatorname{size} \big(p\colon 1 \leqslant s \leqslant n_p, \log(r_{s,p} \leqslant t)\big), \tag{20}$$

where

$$r_{s,p} = \frac{f_{s,p}}{\min\left(f_{s,p}\colon 1 \leqslant s \leqslant n_s\right)}, \tag{21}$$

$\rho_s(t)$ is the probability for solver s which the performance ratio $r_{s,p}$ is inside the range of factor t. When $t = 1$, the probability of the solver to be superior to the other solvers is very high. For the better comparison between the results, the value of $p_s(1)$ needs to be considered by the experimentalist.

Figure 1 and Figure 2 exposed that the performances of all experimental methods were constructed according to the number of iterations and CPU time. For both figures, we generated each graph from the overall problem used to compare ISL, HRM, FR and AMRI method to find out the best solver method. Figure 1 is analyzed based on CPU times in seconds. The analysis was carried out to estimate the duration needed to generate search direction with specific end goal to execute line search and convergence test. Figure 1 showed that the performance of ISL utilizes a shortest time to converge compared to HRM, FR and AMRI. Meanwhile, Figure 2 showed that our proposed method have a better performance in term of number of iteration where it just consumed less number of iteration loops as compared to HRM, FR and AMRI.

Based on the analysis, ISL is obviously better than the other experimental existing methods under strong Wolfe–Powell line search in terms of number of iterations and CPU time. Overall, ISL method shows that it performed well under inexact line search compared to the other existing methods.

5. Conclusion

This paper explores a new coefficient on conjugate gradient method solely to solve unconstrained optimization problems with the assists from inexact line search. Based on this paper, we executed the new method along with inexact line searches and compared with other methods. This new method possesses the global convergence condition under the line search used. Numerical experiment results show that our method is better under inexact line search, namely strong Wolfe–Powell line search compared to the other method. For further study, we apply spectral on the ISL method under the strong Wolfe–Powell line search and form spectral conjugate gradient method. This method takes on idea combining two methods which are, conjugate gradient method and spectral gradient method in a certain way.

[1] Hestenes M. R., Stiefel E. Methods of conjugate gradients for solving linear systems. Journal of Research of the National Bureau of Standards. **49** (6), 409–436 (1952).

[2] Fletcher R., Reeves C. M. Function minimization by conjugate gradients. The Computer Journal. **7** (2), 149–154 (1964).

[3] Polak E., Ribiere G. Note sur la convergence de méthodes de directions conjuguées. Revue Franзaise d'informatique et de Recherche Opérationnelle. Série Rouge. **3** (16), 35–43 (1969).

[4] Polyak B. T. The conjugate gradient method in extremal problems. USSR Computational Mathematics and Mathematical Physics. **9** (4), 94-–112 (1969).

[5] Fletcher R. Practical Methods of Optimization. Wiley-Interscience, New York, NY, USA (1987).

[6] Zoutendijk G. Some algorithms based on the principle of feasible directions. Nonlinear programming. Proceedings of a Symposium Conducted by the Mathematics Research Center, the University of Wisconsin–Madison, May 4–6, 1970. 93–121 (1970).

[7] Powell M. J. D. Nonconvex minimization calculations and the conjugate gradient method. In: Griffiths D. F. (eds) Numerical Analysis. Lecture Notes in Mathematics, vol. 1066. Springer, Berlin, Heidelberg. 122–141 (1984).

[8] Powell M. J. D. Convergence Properties of Algorithms for Nonlinear Optimization. SIAM Review. **28** (4), 487–500 (1986).

[9] Powell M. J. D. Restart procedures for the conjugate gradient method. Mathematical programming. **12** (1), 241–254 (1977).

[10] Al-Baali M. Descent property and global convergence of the Fletcher–Reeves method with inexact line search. IMA Journal of Numerical Analysis. **5** (1), 121–124 (1985).

[11] Touati-Ahmed D., Storey C. Efficient hybrid conjugate gradient techniques. Journal of optimization theory and applications. **64** (2), 379–397 (1990).

[12] Gilbert J. C., Nocedal J. Global Convergence Properties of Conjugate Gradient Methods for Optimization. SIAM Journal on Optimization. **2** (1), 21–42 (1992).

[13] Jiang X., Jian J. Improved Fletcher–Reeves and Dai–Yuan conjugate gradient methods with the strong Wolfe line search. Journal of Computational and Applied Mathematics. **348**, 525–534 (2019).

[14] Mtagulwa P., Kaelo P. An efficient modified PRP-FR hybrid conjugate gradient method for solving unconstrained optimization problems. Applied Numerical Mathematics. **145**, 111–120 (2019).

[15] Rivaie M., Mamat M., June L. W., Mohd I. A new class of nonlinear conjugate gradient coefficients with global convergence properties. Applied Mathematics and Computation. **218** (22), 11323–11332 (2012).

[16] Pytlak R. Conjugate gradient algorithms in nonconvex optimization. Springer Science & Business Media. Vol. 89 (2008).

[17] Hamoda M., Mamat M., Rivaie M., Salleh Z. A conjugate gradient method with Strong Wolfe–Powell line search for unconstrained optimization. Applied Mathematical Sciences. **10** (15), 721–734 (2016).

[18] Rivaie M., Mamat M., Abashar A. A new class of nonlinear conjugate gradient coefficients with exact and inexact line searches. Applied Mathematics and Computation. **268**, 1152–1163 (2015).

[19] Andrei N. An Unconstrained Optimization Test Functions Collection. Advanced Modelling and Optimization. **10** (1), 147–161 (2008).

[20] Hillstrom K. E. A Simulation Test Approach to the Evaluation of Nonlinear Optimization Algorithms. ACM Transactions on Mathematical Software. **3** (4), 305–315 (1977).

[21] Dolan E. D., Moré J. J. Benchmarking optimization software with performance profiles. Mathematical Programming. Series B. **91** (2), 201–213 (2012).

Новий модифікований метод спряженого градієнта при сильному лінійному пошуку Вульфа для розв'язання проблеми необмеженої оптимізації

Ішак М. І., Маржуги С. М., Джун Л. В.

Кафедра математики, Університет Путра Малайзія,
43400 Серданг, Селангор, Малайзія

Метод спряженого градієнта (СГ) добре відомий своєю ефективністю для вирішення проблем необмеженої оптимізації через його збіжні властивості та низьку вартість обчислень. На сьогоднішній день цей метод широко розроблений, щоб конкурувати з існуючими методами за їх ефективностями. У цій статті пропонується модифікація методу СГ при сильному лінійному пошуку Вульфа. Новий коефіцієнт СГ подано на підставі ідеї використання деяких частин попередніх існуючих методів СГ, щоб зберегти їхні переваги. Чисельне тестування однозначно вказує на те, що запропонований метод має кращу можливість для розв'язання необмеженої оптимізації у порівнянні з іншими методами при неточному сильному лінійному пошуку Вулфа–Пауелла.

Ключові слова: *спряжений градієнт, глобальна збіжність, неточний лінійний пощук, сильний лінійний пошук Вульфа–Пауелла, необмежена оптимізація.*