MATEMATIKA, MJIAM, Volume 41, Number 1, 135–150

An Improved Hestenes-Stiefel Conjugate Gradient Method and Its Application

[1,2]Sijun Tao, [3]Mohd Rivaie* and [1]Nur Hamizah Abdul Ghani

[1]College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia.

[2]College of Mathematics and Computer, Xinyu University, 338004 Xinyu Jiangxi, P.R. China.

[3]Department of Computer Science and Mathematics, Universiti Teknologi MARA (UITM), Campus Kuala Terengganu, 21080 Kuala Terengganu, Malaysia.

*Corresponding author: rivaie75@uitm.edu.my

Article history
Received: 17 July 2024
Received in revised form: 26 December 2024
Accepted: 6 March 2025
Published online: 30 April 2025

Abstract The conjugate gradient method is an important method in numerical optimization, which uses the gradient information to construct the conjugate search direction, ensures fast convergence, and aims to achieve fast minimization of the objective function. In this paper, the author proposes a new conjugate gradient method (referred to as the NMHS method) with a simple form and easy to implement and proves the sufficient descent and global convergence of the algorithm under strong Wolfe-Powell line search. To verify the numerical performance of the NMHS method, numerical experiments were conducted on the algorithm, and the experimental results show that the NMHS method is significantly better than the existing method in terms of convergence of the method, the number of iterations required for convergence, and the CPU time consumed. Finally, the NMHS method was applied in practice, and a regression analysis model was established to predict the admission rate of Chinese master's degree students. Through calculations, it can be found that the NMHS method is superior to the Least Squares method and Trend Line method in practical applications, and its prediction is more accurate than the Least Squares method and Trend Line method.

Keywords Conjugate gradient method; global convergence; numerical experiments; sufficient descent; strong Wolfe-Powell line search; regression analysis.

Mathematics Subject Classification 62P20; 62P25; 65Z05; 68T99.

1 Introduction

The conjugate gradient (CG) method, as an effective algorithm for solving unconstrained optimization problems, is favoured for its simple iteration format, small storage capacity, and excellent numerical performance [1]. Unconstrained optimization problems are usually of the following form:

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

where the function $f(x) : R^n \to R$ is a real-valued function, which is continuous and differentiable. For convenience, the gradient of $f(x)$ was denoted by $g(x)$, defined $y_k = g_k - g_{k-1}$. The core of the CG method is to gradually approach the optimal solution of the problem through the update strategy of the following iterative formula:

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

Where x_k is kth current iterate point and $\alpha_k > 0$ is the step size, which is determined by a certain type of line search. In this paper, the author mainly uses the Strong Wolfe-Powell (SWP) inexact line search, which aims to find α_k that satisfies these conditions.

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k, \tag{3}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \leq \sigma |g_k^T d_k|, \tag{4}$$

where $g_k = \nabla f(x_k), \delta, \sigma$ are constant and $0 < \delta < \sigma < 1$.

In Equation (2),d_k is the th-step search direction, also known as the descent direction, which is defined as follows

$$d_k = \begin{cases} -g_k & \text{for} \quad k = 0 \\ -g_k + \beta_k d_{k-1} & \text{for} \quad k \geq 1 \end{cases} \tag{5}$$

where β_k is the coefficient of CG. Different CG methods correspond to different choices of CG coefficients β_k. The classical CG formulas include the Hestenes-Stiefel (HS) method [2], Fletcher-Reeves (FR) method [3], Polak-Ribiere-Polyak (PRP) method [4,5], Fletcher-Reevess (CD) method [6], Liu-Storey (LS) method [7] and Dai-Yuan (DY) method [8]. The specific forms are as follows:

$$\begin{aligned} &\beta_k^{HS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T(g_k - g_{k-1})}, \beta_k^{PRP} = \frac{g_k^T(g_k - g_{k-1})}{||g_{k-1||^2}}, \beta_k^{LS} = \frac{g_k^T(g_k - g_{k-1})}{d_{k-1}^T g_{k-1}} \\ &\beta_k^{FR} = \frac{||g_k||^2}{||g_{k-1}||^2}, \beta_k^{DY} = \frac{||g_k||^2}{d_k^T g_{k-1}}, \beta_k^{CD} = -\frac{||g_k||^2}{d_k^T g_{k-1}}. \end{aligned} \tag{6}$$

where $|| \bullet ||$ stands for the vectors' Euclidean norm.

Among the six methods above, the HS, PRP, and LS methods have the advantages of fast convergence speed and automatic correction to avoid jamming; however, their convergence is poor, and they may not converge even under exact line search [9]. The advantage of the FR, CD, and DY methods is strong convergence, but the convergence speed is slow and easily jammed [10]. Powell [9] provided a counterexample to demonstrate that, even in the case of an exact line search, a non-convex function might exist and that PRP, HS, and LS do not satisfy the convergence properties [11]. Powell recommended using non-negative values for the PRP formulas to guarantee global convergence. Gilbert and Nocedal [12] proposed another formula that shows that the PRP method converges globally under exact and inexact line search, named the PRP+ method, as shown below:

$$\beta_k^{PRP^+} = \max \quad \{\beta_k^{PRP}, 0\} \tag{7}$$

Following this, researchers made various improvements to the above CG method. Wei *et al.* [13] proposed the WYL method. The formula is:

$$\beta_k^{WYL} = \frac{g_k^T(g_k - \frac{||g_k||}{||g_{k-1}||}g_{k-1})}{||g_{k-1}||^2}. \tag{8}$$

[14–16] respectively proved the global convergence of the WYL method under exact and inexact line searches. Rivaie *et al.* [17] proposed the RMIL method with the formula:

$$\beta_k^{RML} = \frac{g_k^T(g_k - g_{k-1})}{||d_{k-1}||^2} \tag{9}$$

Ghani *et al.* [18] proposed the HPRP method and validated the numerical experiments of the HPRP method; the formula is defined as follows:

$$\beta^{HPRP} = \begin{cases} \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_k - g_{k-1}\|}\left|g_k^T g_{k-1}\right|}{\|g_{k-1}\|^2}, \|g_k\|^2 > \frac{\|g_k\|}{\|g_k - g_{k-1}\|}\left|g_k^T g_{k-1}\right|. \\ 0, \text{otherwise} \end{cases} \tag{10}$$

Later, some researchers proposed formulas with parameters, such as the ISL and the VRMIL methods proposed by Ishak *et al.* [19] and Wu *et al.* [20], respectively. Numerous new CG formulas have emerged in recent years [20–22] to continuously enhance the running speed and convergence of the CG method. However, these formulas are usually complex and challenging to apply to real-world fields. Therefore, it is necessary to design a new CG method that is simple, easy to understand, and has good running speed and convergence.

The structure of this paper is as follows. The proposed new algorithm, Algorithm 2.1, will be discussed in the second section. The third section will demonstrate the global convergence under SWP line search and the sufficient descent property of Algorithm 2.1. Section 4 contains the algorithm's numerical experiments and comparisons with other methods. In section 5, the NMHS method will be applied to build a regression analysis model to predict the admission rate of Chinese master's degree students. The algorithm is finally summarized in Section 6.

2 New conjugate gradient method

As analysed in Section 1, the HS method has a fast convergence speed when solving partially smooth objective functions. Still, it is prone to generate non-descent directions in non-smooth or complex problems, which leads to poor convergence. The HPRP method improves the stability of the direction generation through conditional screening and performs better in terms of convergence, but its convergence speed is relatively slow. To combine the advantages of the two methods and make the method applicable to more complex objective functions to ensure convergence speed and convergence, a new and improved conjugate gradient method, called the NMHS method, is proposed in this study where 'N' stands for 'New' and 'M' stands for 'Modified', i.e., the new modified HS method. The formula is:

$$\beta_k^{NMHS} = \begin{cases} \frac{\|g_k\|^2 - \frac{\|g_k\|}{\|g_k - g_{k-1}\|}\left|g_k^T g_{k-1}\right|}{d_{k-1}^T(g_k - g_{k-1})}, \text{if } \|g_k\|^2 > \frac{\|g_k\|}{\|g_k - g_{k-1}\|}\left|g_k^T g_{k-1}\right|. \\ 0, \text{otherwise} \end{cases} \quad (11)$$

The denominator of the NMHS method is the same as the denominator of the HS method, and the convergence speed is accelerated by regulating the new conjugate direction using $d_{k-1}^T(g_k - g_{k-1})$. The numerator of the NMHS method is the same as the numerator of the HPRP method, using the conditional filtering condition $\|g_k\|^2 > \frac{\|g_k\|}{\|g_k - g_{k-1}\|}\left|g_k^T g_{k-1}\right|$ to ensure that the generated direction satisfies the descent condition, avoiding the non-descent direction due to the gradient change magnitude being too small, which enhances the descent property of the algorithm. This design can enhance the algorithm's direction generation problem in nonlinear problems. The NMHS method theoretically integrates the advantages of the HS and HPRP methods, avoids the main defects, achieves a balance between dependability and convergence speed, and provides a reliable and efficient solution for optimizing complex objective functions. The following is the algorithm:

Algorithm 1 (NMHS method)

Step 1. Provide a starting point $x_0 \in R^2, \epsilon > 0$, and $0 < \delta < 1/2, \delta < \sigma < 1$. Set $d_0 = -g_0$, let $k = 0$.

Step 2. If satisfied $||g_k|| \leq \epsilon$,then stop. If not, proceed to Step 3.

Step 3. Computing β_k by Equation (11) and computing the search direction d_k by Equation (5).

Step 4. Computing α_k using Equation (3) and Equation (4).

Step 5. Update new point x_{k+1} using Equation (2).

Step 6. Let $k = k + 1$ then return to Step 2.

3 Global Convergence Analysis of the NMHS method

To demonstrate the global convergence of the proposed method, the following assumptions need to be made:

Assumption 1

A: The set $K = x \in R^n | f(x) \leq f(x_0)$ also known as the level set is bounded.

B: Within a certain neighbourhood D of the set K , the function $f(x)$ is continuously differentiable, and its derivative satisfies the Lipschitz condition; that is, there is a constant $L > 0$ make

$$||g(x) - g(y)|| \leq L||x - y||, \text{for any} \quad x, y \in D \quad (12)$$

In the proof of global convergence of the CG method, it is required that satisfies sufficient descent, which takes the general form:

$$g_k^T d_k \leq -c||g_k||^2, \forall k \in N \quad (13)$$

Lemma 1. Let $\{x_k\}$ be created by Algorithm 1, β_k is given by Equation (11), if α_k is determined by SWP search denoted by Equation (3) and Equation (4) with $\sigma < 1/3$, then Equation (13) holds.

Proof. When $\|g_k\|^2 \leq \frac{\|g_k\|}{\|g_k - g_{k-1}\|} \left|g_k^T g_{k-1}\right|, \beta_k^{NMHS} = 0$, then $g_k^T d_k = -||g_k||^2$, thus Equation (13) holds.

$\|g_k\|^2 \leq \frac{\|g_k\|}{\|g_k - g_{k-1}\|} \left|g_k^T g_{k-1}\right|$ can be shown by induction. If $k = 0$, as $g_1^T d_1 = -||g_1||^2$, then Lemma 1 holds for $k = 0$.

For $k \geq 1$, Assume Equation (13) holds for $k-1$,

$$0 \leq \beta_k^{NMHS} = \frac{||g_k||^2 - \frac{||g_k||}{||g_k - g_{k-1}|g_k^T g_{k-1}}}{d_{k-1}^T (g_k - g_{k-1})} \leq \frac{||g_k||^2}{d_{k-1}^T (g_k - g_{k-1})}. \tag{14}$$

According to the SWP condition, Equation (4) can be rewritten as

$$\sigma g_{k-1}^T d_{k-1} \leq g_k^T d_{k-1} \leq -\sigma g_{k-1}^T d_{k-1}. \tag{15}$$

Thus, it can derive the following inequality

$$-(1-\sigma) g_{k-1}^T d_{k-1} \leq d_{k-1}^T (g_k - g_{k-1}) \leq -(1+\sigma) g_{k-1}^T d_{k-1}. \tag{16}$$

From Equation (5), multiplying both sides by g_k^T, obtain

$$g_k^T d_k = -||g_k||^2 + \beta_k^{NMHS} g_k^T d_{k-1} \leq -||g_k||^2 + \frac{||g_k||^2 g_k^T d_{k-1}}{d_{k-1}^T (g_k - g_{k-1})}. \tag{17}$$

Divide both sides of Equation (17) by $||g_k||^2$, then combine with Equation (15) and Equation (16), yields

$$\frac{g_k^T d_k}{||g_k||^2} \leq -1 + \frac{-\sigma g_{k-1}^T d_{k-1}}{-(1-\sigma) g_{k-1}^T d_{k-1}} = -1 + \frac{\sigma}{(1-\sigma)} < 0, \text{when} \quad \sigma < \frac{1}{3} \tag{18}$$

Let $c = 1 - \frac{\sigma}{1-\sigma}$, Equation (18) be written as $g_k^T d_k \leq -c||g_k||^2$, Indicating that this conclusion holds for k. The proof is finished.

Lemma 2. If the step size α_k satisfies the SWP line search Equation (3) and Equation (4) with $\sigma < 1/3$, then the conclusion can be obtained as follows.

$$0 \leq \beta_k^{NMHS} \leq \frac{1}{1-2\sigma} \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}} \tag{19}$$

Proof. According to Equation (14) and Equation (16), obtain

$$0 \leq \beta_k^{NMHS} \leq \frac{||g_k||^2}{d_{k-1}^T (g_k - g_{k-1})} \leq \frac{||g_k||^2}{(\sigma - 1) g_{k-1}^T d_{k-1}}$$

Because $c = 1 - \frac{\sigma}{1-2\sigma} = \frac{1-2\sigma}{1-\sigma}$, Equation (18) can be written as

$$||g_k||^2 \leq -\frac{g_k^T d_k}{c} = \frac{\sigma - 1}{1 - 2\sigma} g_k^T d_k.$$

Thus, it can be obtained

$$0 \leq \beta_k^{NMHS} \leq \frac{||g_k||^2}{(\sigma - 1) g_{k-1}^T d_{k-1}} \leq \frac{1}{1-2\sigma} \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}$$

So, the proof is finished.

The following lemma, also known as the Zoutendijk condition, takes the following form.

Lemma 3. If Assumption 1 holds, take into consideration any CG method that takes the form of Equation (2) and Equation (5), where d_k satisfy $g_k^T < 0$, and let α_k satisfy the SWP line search Equation (3) and Equation (4), then the following condition holds

$$\sum_{k=1}^{\infty} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} < \infty \tag{20}$$

This Lemma 3 has been proved by Dai & Yuan [8], and the author will not prove it again.

Theorem 1. Suppose that Assumption 1 holds. Be produced by Algorithm 1, which is given by Equation (11); if α_k satisfy the SWP line search Equation (3) and Equation (4), $\mu > 1, 0 < \delta < \sigma < 1/3$, then

$$\lim_{k \to +\infty} \|g_k\| = 0$$

Proof. This Theorem 1 can be proved by contradiction. In other words, if the result of Theorem 1 is false, then there exists a constant $\epsilon > 0$, such that.

$$||g_k|| \geq \epsilon, \forall k \geq 1.$$

From Equation (5), by squaring both sides, obtain

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

Dividing both sides by $(g_k^T d_k)^2$ and combining Equation (19) to obtain

$$\begin{aligned}
\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} &\leq \frac{\|d_{k-1}\|^2}{(1-2\sigma)^2 \left(g_{k-1}^T d_{k-1}\right)^2} - \frac{2}{g_k^T d_k} - \frac{\|g_k\|^2}{\left(g_k^T d_k\right)^2} \\
&= \frac{\|d_{k-1}\|^2}{(1-2\sigma)^2 \left(g_{k-1}^T d_{k-1}\right)^2} - \left(\frac{1}{\|g_k\|} + \frac{\|g_k\|}{g_k^T d_k}\right)^2 + \frac{1}{\|g_k\|^2} \leq \frac{\|d_{k-1}\|^2}{(1-2\sigma)^2 \left(g_{k-1}^T d_{k-1}\right)^2} + \frac{1}{\|g_k\|^2}.
\end{aligned}$$

Repeat the above process and combine it with $||d_1||^2/(g_1^T d_1)^2 = 1/||g_1||^2$, the yields

$$\begin{aligned}
\frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} &\leq \frac{\|d_{k-1}\|^2}{(1-2\sigma)^2 \left(g_{k-1}^T d_{k-1}\right)^2} + \frac{1}{||g_k||^2} \leq \frac{1}{(1-2\sigma)^4} \frac{||d_{k-2}||^2}{(g_{k-2}^T d_{k-2})^2} + \frac{1}{(1-2\sigma)^2} \frac{1}{||g_{k-1}||^2} + \frac{1}{||g_k||^2} \\
&\leq \ldots \leq \frac{1}{(1-2\sigma)^{2k-2}} \frac{1}{||g_1||^2} + \frac{1}{(1-2\sigma)^{2k-4}} \frac{1}{||g_2||^2} + \ldots + \frac{1}{(1-2\sigma)^2} \frac{1}{||g_k||^2} + \frac{1}{||g_k||^2}.
\end{aligned}$$

Because of $||g_k||^2 > \epsilon^2$, obtain

$$\frac{||d_k||^2}{(g_k^T d_k)^2} \leq \frac{1}{\epsilon^2} \sum_{i=0}^{k-1} \frac{1}{(1-2\sigma)^{2i}} = \left(1 - \frac{1}{(1-2\sigma)^{2k}}\right) / \epsilon^2 \left(1 - \frac{1}{(1-2\sigma)^2}\right).$$

Thus,

$$\frac{(g_k^T d_k)^2}{||d_k||^2} \geq \epsilon^2 \left(1 - \frac{1}{(1-2\sigma)^2}\right) / \left(1 - \frac{1}{(1-2\sigma)^{2k}}\right)$$

Therefore,

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{||d_k||^2} = \infty.$$

This result contradicts the formula Equation (20). Therefore, the proof is finished.

4 Numerical experiments

In this section, numerical experiments were conducted using the proposed NMHS method. To comprehensively evaluate the algorithm's performance, 20 unconstrained optimization test functions have been selected for numerical experiments. These test functions are all derived from Andrei [23], covering different types and difficulty problems. Each test function selects four different test points. According to the characteristics of the test function, select different test dimensions, from the lowest 2 dimensions to the highest 10,000 dimensions. Test functions and related information are displayed in Table 1.

Table 1: Functions and Related Information.

No	Test Functions	Test Dimensions	Initial Points
1	Booth	2	2,6,15,25
2	Six Hump Camel	2	-2,7,15,25
3	Three Hump Camel	2	-5,2,10,41
4	Treccani	2	-1,5,10,20
5	NONSCOMP function	2	6,16,26,36
6	Zettl	2	3,12,20,30
7	Extended Wood	4	-1,5,10,20
8	Extended quadratic penalty QP1	2,4,10	3,10,20,40
9	Raydan 1 function	4,20,100	4,14,30,40
10	Extended Freudenstein & Roth	2,500,1000	3,10,20,40
11	Hager	2,4,10,100	3,8,20,30
12	Extended Tridiagonal 1	4,20,100,1000,10000	6,15,30,60
13	Fletcher	2,10,100,500, 1000	12,22,32,62
14	Diagonal 4	2, 500, 1000, 5000, 10000	4,10,40,90
15	Extended Beale	2, 500, 1000, 5000, 10000	-4,1,2,4
16	Shallow	2, 500, 2000, 6000, 10000	3,6,30,60
17	Extended White & Holst	2, 500, 2000, 6000, 10000	-2,4,12,16
18	Extended DENSCHNB	2, 500, 2000, 6000, 10000	2,8,16,20
19	Extended Himmelblau	4, 500, 2000, 6000, 10000	2,10,25,30
20	Extended Rosenbrock	4, 500, 2000, 6000, 10000	3,8,15,20

For the experimental environment, a computer equipped with an Intel i5-1155G7 + 16GB RAM was used, ensuring the algorithm had enough computing power and storage space. In terms of software platform, and chose MatlabR2023b programming, which has powerful numerical computing capabilities and provides excellent convenience for implementing and testing algorithms. To evaluate the NMHS method's operational efficacy, the NMHS method with the HS, HPRP, PRP, and FR methods were compared. To test accurately, set $\epsilon = 10^{-6}$. When $||g_k|| \leq \epsilon$, the iteration is stopped; at the same time, it is stipulated that the iteration ends when the total number of iterations reaches 10,000. Using SWP line search and applying the performance curves of Dolan and More [24] to compare the numerical performance of the four methods, NMHS, HS, TPRP, and FR. That is, compare the number of iterations and CPU running time of the 20 test functions in Table 1 under four different methods. Figure 1 and Figure 2 represent the iteration times and CPU time performance profiles of these four methods, respectively.

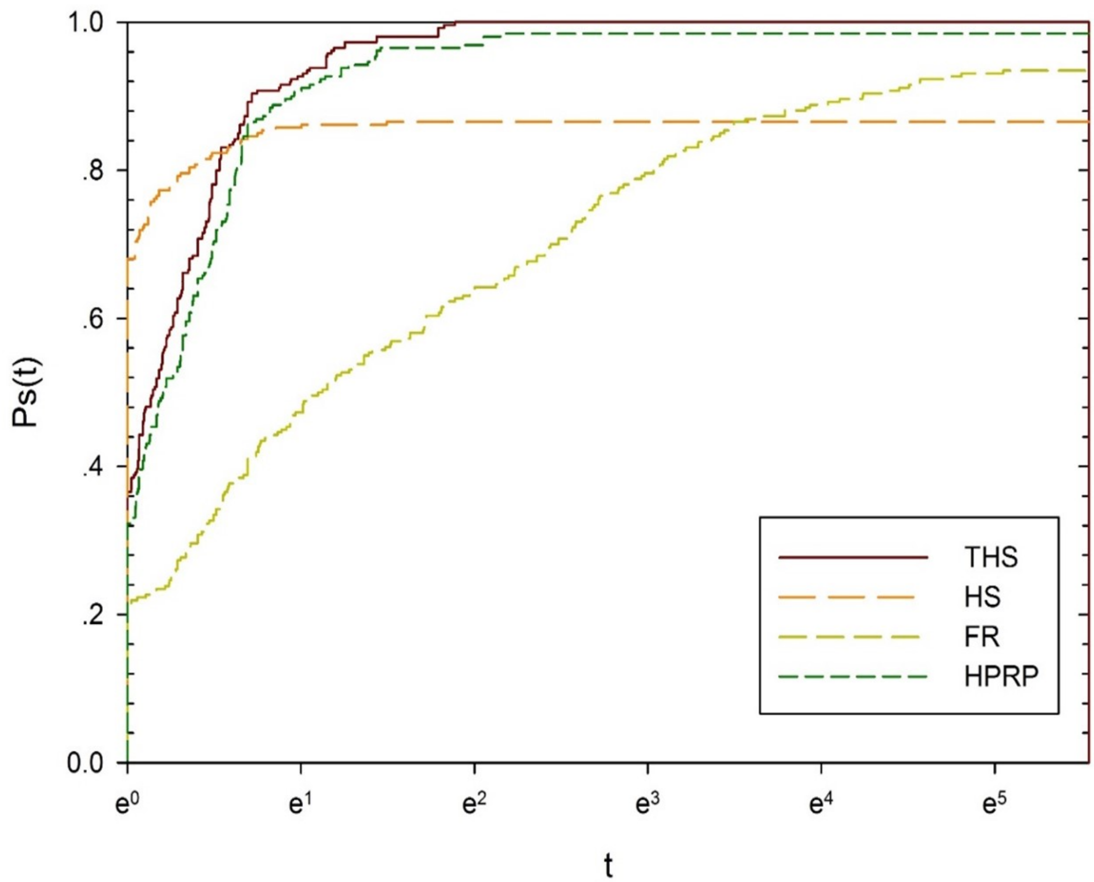

Figure 1: Performance Profile for the Number of Iterations

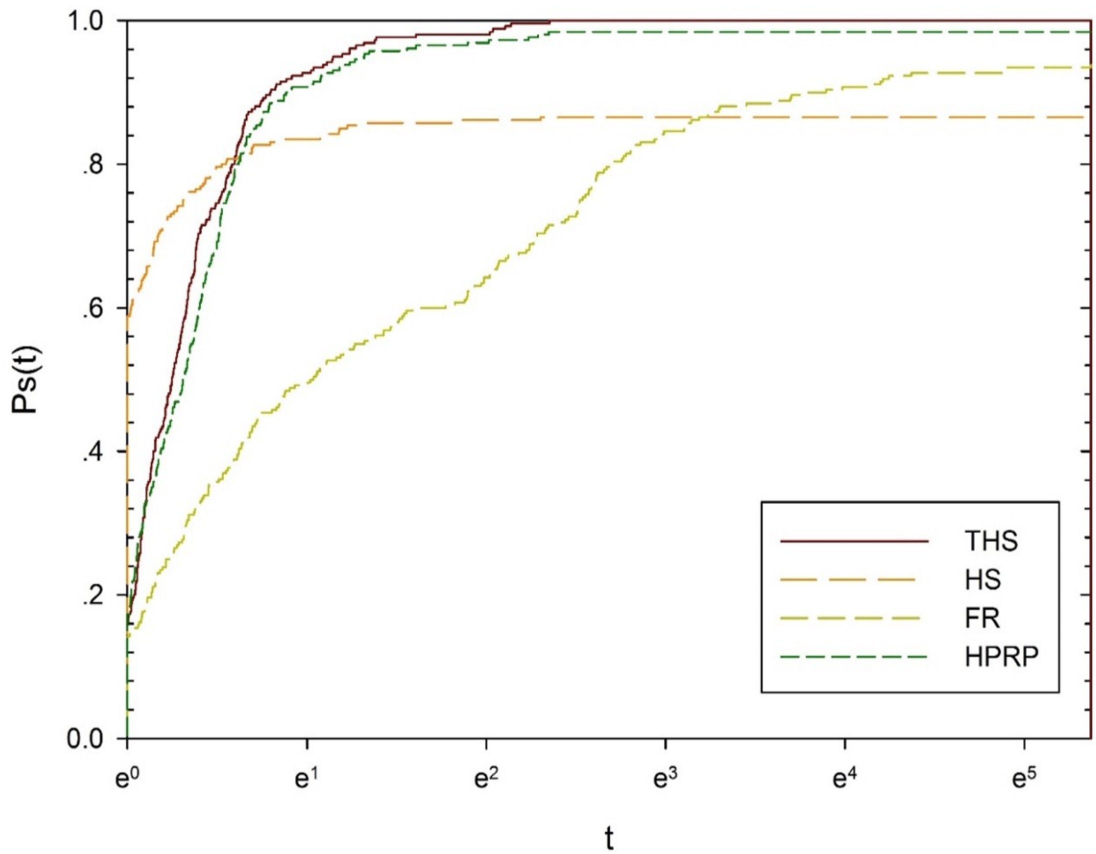

Figure 2: Performance Profile for the CPU Times

As is well known, the left side of the performance profile figure gives the percentage of test problems solved using the fastest method, and the right side of the figure shows the percentage of test problems successfully solved by each method. From the curves on the left side of Figure 1 and Figure 2, it can be seen that the HS method has the fastest test speed, and the curve is at the highest position, followed by the NMHS method and the HPRP method, and the slowest method is the FR method. From the right curve, it can be found that the NMHS method performs the best and may solve every test problem. The HPRP method comes in second and can resolve 98% of test problems. The next is the FR method, which can solve 94% of the test problems, and the HS method has the poorest convergence and can only solve 86% of the test problems. Figure 1 and Figure 2 show that in terms of overall test performance, the NMHS method can solve all test problems and has a faster test speed. Therefore, the NMHS method is superior to the other methods.

5 Application of the NMHS method to regression analysis

This study applies the NMHS method in regression analysis to solve real world problem in data modeling. Table 2 shows the number and proportion of Chinese university students enrolled and admitted to master's degrees in the last 16 years (https://www.dxsbb.com/news/133463.html).

Table 2: China Master's Degree Application and Admission Information.

No	Year	Number of applicants (in 10,000)	Number of admissions (in 10,000)	Admission rate
1	2008	120	38.67	32.23%
2	2009	124.6	44.9	36.04%
3	2010	140.6	47.44	33.74%
4	2011	151.1	49.46	32.73%
5	2012	165.6	52.13	31.48%
6	2013	176	54.09	30.73%
7	2014	172	54.87	31.90%
8	2015	164.9	57.06	34.60%
9	2016	177	58.98	33.32%
10	2017	201	72.22	35.93%
11	2018	238	76.25	32.04%
12	2019	290	81.13	27.98%
13	2020	341	99.05	29.05%
14	2021	377	106.2	28.17%
15	2022	457	110.35	24.15%
16	2023	474	124.2	26.20%

As shown in Table 2, despite the fluctuations in the number of students enrolled and admitted to master's degrees from year to year, there is a correlation between the two from a statistical point of view. Our main goal is to establish the relationship between the proportion of master's degree students admitted and the year, i.e., to find a regression equation that describes the variation of the proportion of admissions with the years. From the data in Table 2, it can be observed that there may be some linear or non-linear functional relationship between the proportion of master's degree students admitted in different years and the year. For this reason, this paper conducted regression analyses on the data in Table 2 using the Trend Line method, the Least Squares method, and the NMHS method to predict the admission rate of Chinese master's degree students in future years. The calculation excluded the data from 2023, which was used to calculate the relative error between the predicted and actual values to test the accuracy of the algorithms.

5.1 Trend Line method

The year and admission rate data presented in Table 2 are analyzed using Microsoft Excel. The trend line method is employed to construct both linear and quadratic regression models, resulting in the following equations, respectively

$$y = -0.0047403571x + 9.8678796429, \tag{21}$$

$$y = -0.0008019958x^2 + 3.2273027101x - 3246.40054, \tag{22}$$

where x represents the year and y is the admission rate. Figure 3 presents a graphical representation of the models, showing the trend in the admission ratio for Master's programs in China from 2008 to 2022.

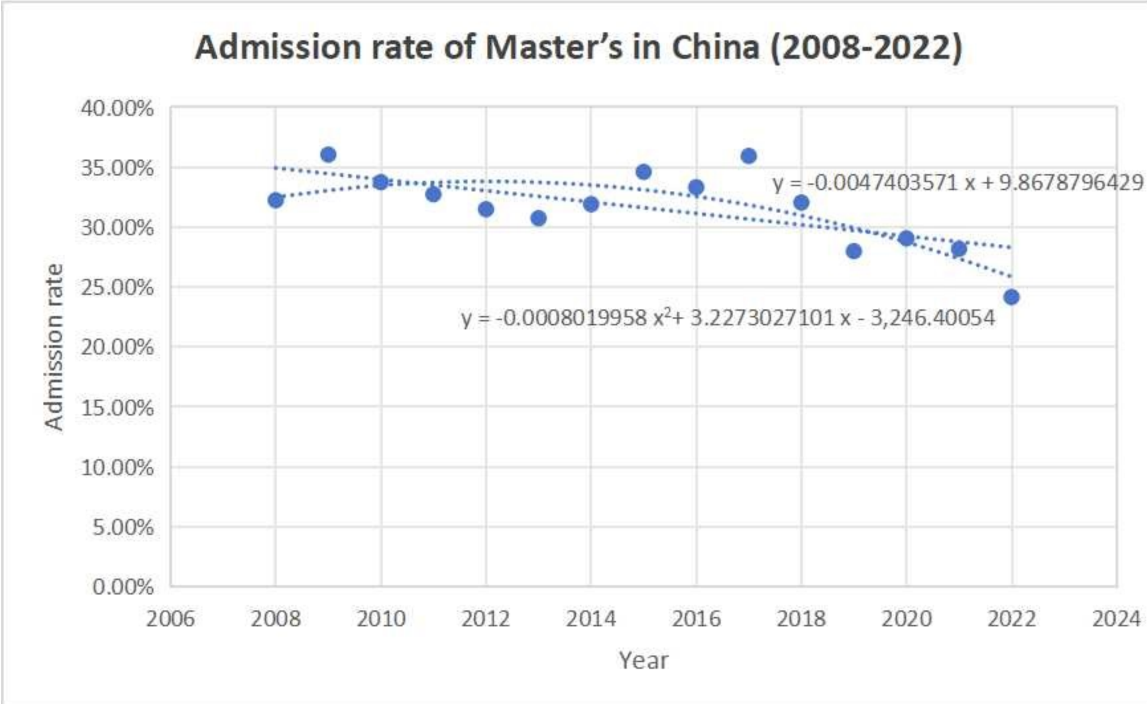

Figure 3: Admission Ratio of Master's in China (2008-2022).

5.2 Application of the Least Squares method and the NMHS method

Assuming the linear and quadratic approximation functions under the Least Squares method are $y = \alpha_0+\alpha_1 x$ and $y = \alpha_0+\alpha_1 x+\alpha_2 x^2$, respectively, the Least Squares method is then applied to the data in Table 2, resulting in the following linear and quadratic regression functions.

$$y = 35.3982857143 - 0.4740357143x, \tag{23}$$

$$y = 31.7625714286 + 0.809157563x - 0.0801995798x^2. \tag{24}$$

Transforming the above two regression functions into an unconstrained minimization problem gives

$$\min f(a) = \sum_{i=1}^{n}[y_i - (a_0 + a_1 x_i)]^2, \tag{25}$$

$$\min f(a) = \sum_{i=1}^{n}[y_i - (a_0 + a_1 x_i + a_2 x_i^2)]^2, \tag{26}$$

where x_i denotes the serial number, which takes the value of $x_i = 1, 2, ..., 15$, and y_i denotes the master's admission rate for the corresponding year. To calculate the relative error between the predicted and actual values, we excluded the 2023 data from the calculation. Once the data points are selected, these data are substituted into Equations (25) and (26), respectively, and solved using Matlab2023b software to obtain the mathematical models for the two unconstrained minimization problems:

$$f(a_0, a_1) = 15a_0^2 + 240a_0a_1 - 948.18a_0 + 1240a_1^2 - 7319.98a_1 + 15128.6915 \tag{27}$$

$$\begin{aligned} f(a_0, a_1, a_2) = 15a_0^2 + 240a_0a_1 + 240a_0a_2 - 948.18a_0 + 1240a_1^2 + 28800a_1a_2 \\ - 7319.98a_1 + 178312a_2^2 - 73473.82a_2 + 15128.6915 \end{aligned} \tag{28}$$

For the minimum value functions Equations (27) and (28) of unconstrained optimization, four different initial points are arbitrarily selected and solved using the NMHS method under SWP line search. The solution results are obtained as shown in Table 3.

Table 3: NMHS method for solving linear and quadratic regression models.

Regression Model	Initial Point	Iterations	CPU	Value of Variable
Linear	(-8,-8)	2	0.00015	[35.39828571426852 -0.474035714446821]
Linear	(3,3)	2	0.00014	[35.39828571428290 -0.474035714268453]
Linear	(18,18)	3	0.00018	[35.39828571428362 -0.474035714285593]
Linear	(26,26)	3	0.00015	[35.39828571428462 -0.474035714290037]
Quadratic	(-6,-6,-6)	6	0.00027	[31.76257140821985, 0.809157568349284, -0.080199580120367]
Quadratic	(3,3,3)	6	0.00044	[31.76257136625478, 0.809157579328706, -0.080199580715205]
Quadratic	(18,18,18)	6	0.00024	[31.76257137289269, 0.809157577592752, -0.080199580621181]
Quadratic	(24,24,24)	6	0.00021	[31.76257142487385, 0.809157563988443, -0.080199579884083]

The linear and quadratic regression functions under the NMHS method are obtained by taking the average of the solution results for each initial point in Table 3, respectively. The two functions obtained are:

$$y = 35.3982857143 - 0.4740357143x, \tag{29}$$

$$y = 31.7625713931 + 0.8091575723x - 0.0801995803x^2 \tag{30}$$

5.3 Error Analysis

Error analysis is used to evaluate the accuracy of the Trend Line, Least Squares and NMHS methods by predicting the admission rate for Masters's student in 2023by using each model. The relative error is then calculated by comparing the predicted values with the actual admission rate with the formula as follows.

$$\text{relative error} = \frac{|\text{exact value - predictive value}|}{|\text{exact value}|} \tag{31}$$

According to Equations (21), (22), (23), (28), and (29), the predicted values and relative errors of 2023 acceptance rates under linear and quadratic line searches for each of the three methods can be calculated as shown in Table 4 below.

Table 4: Prediction values and relative errors of three methods

Method	Regression Model	Exact Value (%)	Predictive value (%)	Relative Error
NMHS	Linear	26.20	27.8137	0.0615916031
	Quadratic	26.20	24.1781	0.0771717557
Least Squares	Linear	26.20	27.8137	0.0615916031
	Quadratic	26.20	24.1780	0.0771755725
Trend Line	Linear	26.20	27.8137	0.0615916031
	Quadratic	26.20	24.1773	0.0772022901

From Table 4 above, it can be seen that all three methods have linear models with more minor relative errors than the quadratic model; the linear models of the three methods are the same, and the quadratic model of the NMHS method is slightly better than the Least Squares method, and the Trend Line method has a more significant relative error than the NMHS method and the Least Squares method. Overall, the NMHS method's relative error is lower than the other two methods' relative error, and the method has more accurate predicted values.

6 Conclusion

Inspired by the HS and HPRP methods, this paper proposes a NMHS method that is simple in form and easy to implement. The paper proves the sufficient descent of the method and the global convergence under SWP inexact line search. The method's effectiveness is verified by numerical experiments, which show that the new method has better convergence and faster convergence than other methods, and the overall performance is better than other methods. Finally, the algorithm is applied in practice to establish a regression analysis model to predict the admission ratio of Chinese master's degree students. Through computational comparison, it can be found that the NMHS method is better than the Least Squares and Trend Line methods in practical application.

Acknowledgments

The authors are grateful to College of Computing, Informatics, and Mathematics, Universiti Teknologi Mara (UiTM) for providing the facilities support on this research. This work is also supported by the Department of Education of Jiangxi Province, China, as a scientific and technological research project entitled 'A study on the nature of the solutions of complex (partial) differential-difference equations' (Project No. GJJ212305).

References

[1] Andrei, N. Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. *Bulletin of the Malaysian Mathematical Sciences Society*. Second Series. 2011. 34(2): 319-330.

[2] Hestenes, M. and Stiefel, E. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards.* Vol. 6. 1952. 409-436. `https://doi.org/10.6028/jres.049.044`.

[3] Fletcher, R. and Reeves, C. Function minimization by conjugate gradient. *Computer Journal*. 1964. 7(2): 149-154. `https://doi.org/10.1093/comjnl/7.2.149`.

[4] Polyak, B.T. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*. 1969. 9(4): 94-112. `https://doi.org/10.1016/0041-5553(69)90035-4`.

[5] Polak, E. and Ribiere, G. Note sur la convergence de méthodes de directions conjuguées. *Revue franaise d'informatique et de recherche opérationnelle. Srie rouge*. 1969. 3(16): 35-43. `https://doi.org/10.1051/m2an/196903R100351`.

[6] Fletcher, R. *Practical methods of optimization*. Vol. 1. 1987, New York: John Wiley & Sons Press. 10-30. `https://doi.org/10.1007/BF00940464`.

[7] Liu, Y. and Storey, C. Efficient generalized conjugate gradient algorithms, part 1: theory. *Journal of optimization theory and applications*. 1991. 69: 129-137.

[8] Dai, Y.H. and Yuan, Y. A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property. *SIAM Journal on Optimization*. 1999. 10(1): 177-182. `https://doi.org/10.1137/S1052623497318992`.

[9] Powell, M.J.D. Nonconvex minimization calculations and the conjugate gradient method. *in Numerical Analysis: Proceedings of the 10th Biennial Conference held at Dundee, Scotland*. 1984. 1066: 122-141. `https://doi.org/10.1007/BFb0099521` .

[10] Salleh, Z., Alhamzi, G., Masmali, I. and Alhawarat, A. A Modified Liu and Storey Conjugate Gradient Method for Large Scale Unconstrained Optimization Problems. *Algorithms*. 2021. 14(8). `https://doi.org/10.3390/a14080227`.

[11] Alhawarat, A., Salleh, Z., Mamat, M. and Rivaie, M. An efficient modified Polak-Ribière-Polyak conjugate gradient method with global convergence properties. *Optimization Methods and Software*. 2016. 32(6): 1299-1312. https://doi.org/10.1080/10556788.2016.1266354.

[12] Gilbert, J.C. and Nocedal, J. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on Optimization*. 1992. 2(1): 21-42. https://doi.org/10.1137/0802003.

[13] Zengxin, W., Shengwei, Y. and Liying, L. The convergence properties of some new conjugate gradient methods. *Applied Mathematics and Computation*. 2006. 183(2): 1341-1350. https://doi.org/10.1016/j.amc.2006.05.150

[14] Huang, H., Zengxin, W. and Shengwei, Y. The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search. *Applied Mathematics and Computation*. 2007. 189. https://doi.org/10.1016/j.amc.2006.12.006.

[15] Shengwei, Y., Zengxin, W. and Huang, H. A note about WYL's conjugate gradient method and its applications. *Applied Mathematics and Computation*. 2007. 191. https://doi.org/10.1016/j.amc.2007.02.094.

[16] Sha, L., Zengxin, W. and Liliu, M. Some global convergence properties of the Wei-Yao-Liu conjugate gradient method with inexact line search. *Applied Mathematics and Computation*. 2011. 217(17): 7132-7137. https://doi.org/10.1016/j.amc.2011.01.097.

[17] Rivaie, M., Mamat, M., June, L.W. and Mohd, I. A new class of nonlinear conjugate gradient coefficients with global convergence properties. *Applied Mathematics and Computation*. 2012. 218(22): 11323-11332. https://doi.org/10.1016/j.amc.2012.05.030.

[18] Abdul Ghani, N.H., Mamat, M. and Rivaie, M. A new family of Polak-Ribière-Polyak conjugate gradient method with the strong-Wolfe line search, in *American Institute of Physics*. 2017. https://doi.org/10.1063/1.4995892.

[19] Ishak, M. I., Marjugi, S.M. and June, L.W. A new modified conjugate gradient method under the strong Wolfe line search for solving unconstrained optimization problems. *Mathematical Modeling and Computing*. 2022. 9(1): 111-118. https://doi.org/10.23939/mmc2022.01.111.

[20] Xiaoyu, W., Shao, H., Pengjie, L., Zhang, Y. and Zhuo, Y. An efficient conjugate gradient-based algorithm for unconstrained optimization and its projection extension to large-scale constrained nonlinear equations with applications in signal recovery and image denoising problems. *Journal of Computational and Applied Mathematics*. 2023. 422. https://doi.org/10.1016/j.cam.2022.114879.

[21] Yahaya, J. and Kumam, P. Efficient hybrid conjugate gradient techniques for vector optimization. *Results in Control and Optimization*. 2024. 14. https://doi.org/10.1016/j.rico.2023.100348.

[22] Yousif, O.O.O. and Saleh, M.A. Another modified version of RMIL conjugate gradient method. *Applied Numerical Mathematics*. 2024. 202(120-126).

[23] Andrei, N. An unconstrained optimization test functions collection. *Advanced Modeling and Optimization*. 2008. 10(1): 147-161.

[24] Dolan, E.D. and Moré, J.J. Benchmarking optimization software with performance profiles. *Mathematical Programming*. 2002. 91(201-213). https://doi.org/10.1007/s101070100263.