Iranian Journal of Numerical Analysis and Optimization
Vol. ??, No. ??, ??, pp ??
https://doi.org/??
https://ijnao.um.ac.ir/

Research Article

Global convergence of modified conjugate gradient methods with application in conditional model regression function

C. Yacine and M. Abd Elhamid*

Abstract

The conjugate gradient method is one of the most important ideas in scientific computing. It is applied to solving linear systems of equations and nonlinear optimization problems. In this paper, based on a variant of the Hestenes–Stiefel (HS) method and Polak–Ribière–Polyak (PRP) method, two modified conjugate gradient methods (named MHS* and MPRP*) are presented and analyzed. The search direction of the presented methods fulfills the sufficient descent condition at each iteration. We establish the global convergence of the proposed algorithms under normal assumptions and strong Wolfe line search. Preliminary elementary numerical experiment results are presented, demonstrating the promise and the effective-

*Corresponding author

Received ??? ; revised ??? ; accepted ???

Chaib Yacine
Laboratory Informatics and Mathematics (LIM), Mohamed Cherif Messaadia University, Souk Ahras, 41000, Algeria. e-mail: y.chaib@univ-soukahras.dz

Mehamdia Abd Elhamid
Laboratory Informatics and Mathematics (LIM), Mohamed Cherif Messaadia University, Souk Ahras, 41000, Algeria. e-mail: a.mehamdia@univ-soukahras.dz

How to cite this article
Yacine, C. and Abd Elhamid, M., Global convergence of modified conjugate gradient methods with application in conditional model regression function. *Iran. J. Numer. Anal. Optim.*, ??; ??(??): ??-??. ??

ness of the proposed methods. Finally, the proposed methods are further extended to solve the problem of conditional model regression function.

AMS subject classifications (2020): Primary 90C30; Secondary 65K05, 62G05.

Keywords: Conjugate gradient method; Inexact line search; Descent condition; Global convergence; Numerical comparisons; Conditional model regression; Kernel estimator.

1 Introduction

Optimization methods are widely used to obtain the numerical solution of the optimal control problems arising in scientific and engineering computation, especially for solving large-scale problems. The nonlinear conjugate gradient (CG) method is welcomed for its simple iteration and little storage. In this work, we consider the unconstrained optimization problem

$$\min \left\{ f(x) : x \in \mathbb{R}^n \right\}, \tag{1}$$

where f is a continuously differentiable function. The nonlinear CG method is one of the convincing methods for solving problem (1). Its iterative procedure is expressed as follows:

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

where x_k is the current iteration point, the stepsize α_k is a positive scalar determined by some line search, and d_k is the search direction defined by the following formula:

$$d_{k+1} = -g_{k+1} + \beta_k d_k; \quad d_0 = -g_0, \tag{3}$$

where $g_k = \nabla f(x_k)$ is the gradient of f at x_k and β_k is known as the CG coefficient. There are some established formulas for β_k, which are provided as follows:

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k} \text{ [6]}, \quad \beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \text{[11]}, \quad \beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T d_k} \text{ [10]},$$

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k} \text{ [13]}, \quad \beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \text{ [19]} - \text{[20]}, \quad \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-g_k^T d_k} \text{ [16]},$$

where y_k is defined as the difference between g_{k+1} and g_k, and $\|\cdot\|$ represents the Euclidean norm. The step length α_k is very important for the global convergence of CG methods. One often requires the line search to satisfy the Wolfe line search (WLS) conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{4}$$

and

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k. \tag{5}$$

Also, the strong Wolfe line search (SWLS) conditions consist of (4) and

$$\left| g_{k+1}^T d_k \right| \leq -\sigma g_k^T d_k, \tag{6}$$

where $0 < \delta < \sigma < 1$. From a practical computations point of view, if the FR method produces a bad direction and a little step from x_k to x_{k+1}, the next direction and the next step are also probable to be poor unless a reboot along the negative gradient direction is executed [21]. Although there is such a drawback, it has been shown that the FR method has strong convergent properties [7]. The numerical performances of the CD and DY methods are very similar to the FR method since the scalar β_k in these three methods has the same numerator.

In the past few years, the Polak–Ribière–Polyak (PRP) method has generally been regarded to be one of the most efficient CG methods in practical computation. A wonderful property of the PRP method is that it automatically performs a restart if a bad direction occurs [12]. The numerical performances of the Hestenes–Stiefel (HS) and LS methods are very similar to the PRP method since the coefficient β_k in these methods has the same numerator. However, the convergence properties of the PRP, HS, and LS methods are not so good [22]. In recent years, based on the above six formulas and their hybridization, many works putting effort into seeking new CG methods with only good convergence properties and also excellent numerical effects were published.

In (2006) ,Wei, Yao, and Liu [24] gave a variant of the PRP method called the WYL method, where the parameter β_k is yielded by

$$\beta_k^{WYL} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{\|g_k\|^2}.$$

The WYL method inherits the properties of PRP. Under the SWLS with $\sigma < \frac{1}{4}$, Huang, Wei, and Yao [15] demonstrated that the WYL method adheres to the sufficient descent condition and achieves global convergence.

Yao, Wei, and Huang [25] expanded upon this concept to the HS method. This modification is referred to as the MHS approach, and the parameter β_k within this method is defined as follows:

$$\beta_k^{MHS} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{d_k^T (g_{k+1} - g_k)}.$$

The authors analyzed the sufficient descent property and global convergence when SWLS is employed [25]. In 2009, Zhang [26] gave two modified CG methods, proposing the following formula:

$$\beta_k^{NPRP} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \left|g_{k+1}^T g_k\right|}{\|g_k\|^2} \text{ and } \beta_k^{NHS} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} \left|g_{k+1}^T g_k\right|}{d_k^T (g_{k+1} - g_k)}.$$

The NPRP and NHS methods have sufficient descent conditions and are globally convergent if the SWLS is utilized with the parameter $\sigma < \frac{1}{2}$ [26]. Soon afterward, based on the CG method DY , Huang [14] proposed a new CG formula, where β_k is written as

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

Huang [14] proved that the MDY method satisfies the sufficient descent condition and converges globally under the SWLS. Moreover, Du, Zhang, and Ma [9] proposed two modified CG methods, denoted by NVHS * and NVPRP*. The parameter β_k in the NVHS* and NVPRP* methods are given by

$$\beta_k^{\text{NVHS}^*} = \frac{\|g_{k+1}\|^2 - \frac{\left|g_{k+1}^T g_k\right|}{\|g_k\|^2} g_{k+1}^T g_k}{d_k^T (g_{k+1} - g_k)} \text{ and } \beta_k^{\text{NVPRP}^*} = \frac{\|g_{k+1}\|^2 - \frac{\left|g_{k+1}^T g_k\right|}{\|g_k\|^2} g_{k+1}^T g_k}{\|g_k\|^2}.$$

The convergence of the two methods with the SWLS is established, and numerical results show that these computational schemes are efficient [9].

Continuing previous results, we propose two efficient CG methods for solving unconstrained optimization problems. Under the SWLS, we establish the convergence properties of the MHS* and MPRP* CG methods. Numerical results show that the two modifications are efficient, robust and each of these modifications outperforms the four CG methods famous. Finally, an application of our methods in nonparametric mode conditional estimator is also considered.

The rest of the paper is organized as follows. In section 2, we introduce the two modified methods and algorithms. In section 3 presents the sufficient descent condition and the global convergence proof of the two proposed methods. The numerical results and discussions are contained in section 4. In section 5, we focus on applying the new methods in nonparametric statistics. Conclusions and discussions are made in the last section.

2 Modified formulas and algorithms

In this section, we propose modified CG methods to solve unconstrained optimization problems (1) The sufficient descent condition of our methods is analyzed and established.

2.1 Main contributions

The two methods presented are the result of monitoring the construction of CG parameters in the NHS and NPRP methods. Clearly, β_k^{NHS} and β_k^{NPRP} have the same mathematical expression for the numerator, that is, $\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}\left|g_{k+1}^T g_k\right|$.

By considering the numerators of the previous two methods, we see that the parameter β_k can also be chosen as

$$\beta_k^{MHS^*} = \frac{\|g_{k+1}\|^2 - \eta_1 \frac{\left|g_{k+1}^T d_k\right|\left|g_{k+1}^T g_k\right|}{\|d_k\|\|g_k\|}}{d_k^T\left(g_{k+1} - g_k\right) + \xi_1 \|d_k\| \|g_{k+1}\|}, \quad \eta_1 \in [0,1] \text{ and } \xi_1 > 1. \tag{7}$$

That is, we replace the term $\frac{\|g_{k+1}\|}{\|g_k\|}\left|g_{k+1}^T g_k\right|$ in β_k^{NHS} by $\eta_1 \frac{\left|g_{k+1}^T d_k\right|\left|g_{k+1}^T g_k\right|}{\|d_k\|\|g_k\|}$ in $\beta_k^{MHS^*}$. Also, add $\xi_1 \|g_{k+1}\| \|d_k\|$ to the denominator.

Second, we define the parameters $\beta_k^{MPRP^*}$ of the MPRP* method as follows:

$$\beta_k^{MPRP^*} = \frac{\|g_{k+1}\|^2 - \eta_2 \frac{\left|g_{k+1}^T d_k\right|\left|g_{k+1}^T g_k\right|}{\|d_k\|\|g_k\|}}{\|g_k\|^2 + \xi_2 \|d_k\| \|g_{k+1}\|}, \quad \eta_2 \in [0,1] \text{ and } \xi_2 > 0. \tag{2.2}$$

The primary attributes of these methods are as follows:

- A modified CG methods are introduced for solving nonparametric estimators of the conditional mode function.
- All search directions satisfy the sufficient descent condition.
- The proposed methods provide global convergence.
- Evaluation of performance based on the tool of Dolan and Mor [8] showed that the proposed methods are more efficient and effective than conventional methods.

2.2 Algorithms

In this subsection, we present the MHS* and MPRP* algorithms with the SWLS.

2.2.1 MHS* algorithm

2.2.2 MPRP* algorithm

The MPRP* algorithm shares similarities with the MHS* algorithm, with the key difference being that in Step 4, we substitute (7) with (2.2).

Algorithm 1:

Step 1: Initialization.

Select $x_0 \in \mathbb{R}^n$, and choose parameters δ and σ such that $0 < \delta < \sigma < 1$. Calculate $f(x_0)$ and g_0. Let $d_0 = -g_0$.

Step 2: Test for a continuation of iterations.

If the value of $\|g_k\|_\infty$ is less than or equal to 10^{-6}, then terminate the procedure. Otherwise, continue to the next step.

Step 3: Line search.

Determine the value of α_k that satisfies (4) and (6), and update the variables with the following equation $x_{k+1} = x_k + \alpha_k d_k$.

Step 4: Calculate β_k using (7).

Step 5: Use (3) to determine the search direction.

Step 6: Go to Step 2 after setting $k = k + 1$.

2.3 The sufficient descent direction

If $g_k^T d_k \leq -c\|g_k\|^2$ with $c \geq 0$, this indicates that the search direction d_k possesses the sufficient descent conditions, which is an important property for the global convergence.

The following Theorem shows that the MHS* method generates satisfies sufficient descent conditions directions with the strong WLS.

Theorem 1. Let the sequences $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ be generated by MHS* algorithm, then for positive constant c,

$$g_k^T d_k \leq -c\|g_k\|^2, \quad \text{for all} k \geqslant 0. \tag{8}$$

Proof. The following proof is by induction. For $k = 0$, $g_0^T d_0 = -\|g_0\|^2$, we conclude that the sufficient descent condition holds for $k = 0$. Now, we assume that (8) holds for k and prove that for $k + 1$.

From (6) and (8), we obtain

$$d_k^T (g_{k+1} - g_k) \geq (1-\sigma)\left(-d_k^T g_k\right) \geq 0. \tag{9}$$

It follows from (9) and Cauchy–Schwarz inequality, that

$$\beta_k^{MHS^*} \geq \frac{\|g_{k+1}\|^2 - \eta_1 \frac{\|g_{k+1}\|\|d_k\|\|g_{k+1}\|\|g_k\|}{\|d_k\|\|g_k\|}}{d_k^T(g_{k+1}-g_k) + \xi_1 \|g_{k+1}\|\|d_k\|}$$
$$= \frac{\|g_{k+1}\|^2(1-\eta_1)}{d_k^T(g_{k+1}-g_k) + \xi_1 \|g_{k+1}\|\|d_k\|} \geq 0. \quad (10)$$

Using the definition of $\beta_k^{MHS^*}$ and (9), we have

$$\beta_k^{MHS^*} = \frac{\|g_{k+1}\|^2 - \frac{\eta_1 |g_{k+1}^T d_k||g_{k+1}^T g_k|}{\|d_k\|\|g_k\|}}{d_k^T(g_{k+1}-g_k) + \xi_1 \|g_{k+1}\|\|d_k\|} \leq \frac{\|g_{k+1}\|^2}{\xi_1 \|g_{k+1}\|\|d_k\|}. \quad (11)$$

From (3), (10), (11) and the Cauchy–Schwarz inequality, it is clear that

$$g_{k+1}^T d_{k+1} \leq -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\xi_1 \|g_{k+1}\|\|d_k\|}\|g_{k+1}\|\|d_k\| = -c\|g_{k+1}\|^2,$$

where $c = 1 - \frac{1}{\xi_1}$, so there is a constant $c > 0$ with $\xi_1 > 1$. □

We give a theorem, which shows that the MPRP* method possesses the sufficient descent property if the step size α_k is determined by the SWLS with $0 < \sigma < \frac{1}{2}$.

Theorem 2. Let the direction d_k be yielded by the MPRP* method. If $\sigma < \frac{1}{2}$, then the relations

$$-\frac{1}{1-\sigma} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -\frac{1-2\sigma}{1-\sigma}, \quad (12)$$

hold. So, the search direction d_k generated by the MPRP* method is sufficient descent.

Proof. The proof is by induction. The result clearly holds for $k = 0$, since the middle term equals -1 and $0 \leq \sigma < 1$. Assume that (12) holds for some $k \geq 0$. This implies that $g_k^T d_k < 0$, since

$$-\frac{1-2\sigma}{1-\sigma} < 0.$$

From the Cauchy–Schwarz inequality, $\eta_2 \in [0,1]$, and $\xi_2 > 0$, we have

$$\beta_k^{MPRP^*} \geq \frac{\|g_{k+1}\|^2 - \eta_2 \frac{\|g_{k+1}\|\|d_k\|\|g_{k+1}\|\|g_k\|}{\|d_k\|\|g_k\|}}{\|g_k\|^2 + \xi_2 \|g_{k+1}\|\|d_k\|}$$

$$= \frac{\|g_{k+1}\|^2 (1-\eta_2)}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|} \geq 0.$$

On the other hand,

$$\beta_k^{MPRP^*} = \frac{\|g_{k+1}\|^2 - \frac{\eta_2 |g_{k+1}^T d_k| |g_{k+1}^T g_k|}{\|d_k\| \|g_k\|}}{\|g_k\|^2 + \xi_2 \|g_{k+1}\| \|d_k\|} \leq \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_k^{FR}.$$

We concluded

$$0 \leq \beta_k^{MPRP^*} \leq \beta_k^{FR}. \tag{13}$$

From (3), we have

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} = -1 + \beta_k^{MPRP^*} \frac{g_{k+1}^T d_k}{\|g_{k+1}\|^2} = -1 + \frac{\beta_k^{MPRP^*}}{\beta_k^{FR}} \frac{g_{k+1}^T d_k}{\|g_k\|^2}. \tag{14}$$

Using (6) and (13), we have

$$\beta_k^{MPRP^*} |g_{k+1}^T d_k| \leq -\sigma \beta_k^{MPRP^*} g_k^T d_k,$$

which, together with (14), gives

$$-1 + \sigma \frac{\beta_k^{MPRP^*}}{\beta_k^{FR}} \frac{g_k^T d_k}{\|g_k\|^2} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 - \sigma \frac{\beta_k^{MPRP^*}}{\beta_k^{FR}} \frac{g_k^T d_k}{\|g_k\|^2}.$$

From the left-hand side of the induction hypothesis (12), we obtain

$$-1 - \frac{\sigma}{1-\sigma} \frac{\beta_k^{MPRP^*}}{\beta_k^{FR}} \leq \frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2} \leq -1 + \frac{\sigma}{1-\sigma} \frac{\beta_k^{MPRP^*}}{\beta_k^{FR}}.$$

Using the bound (13), we conclude that (12) holds for $k+1$. □

3 Convergence analysis

To establish the global convergence of the proposed methods, we need the following basic assumptions on the objective function.

Assumption 1. Given an initial point x_0, the level set $S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\}$, is bounded.

Assumption 2. In a neighborhood $\mathcal{N}$ of S, the objective function f is continuously differentiable and its gradient is Lipschitz continuous, namely, there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|, \quad \text{for all} \quad x, y \in \mathcal{N}. \tag{15}$$

Assumption 2 implies that there exists a positive constant $\Gamma \geq 0$, such that

$$\| \nabla f(x) \| \leq \Gamma, \qquad \text{for all } x \in \mathcal{N}. \tag{16}$$

Under Assumptions 1–2, the following theorem, due to Zoutendijk [27], is essential in proving the global convergence results of the unconstrained optimization algorithms.

Theorem 3. We assume that Assumptions 1 and 2 hold. Let the sequence $\{x_k\}_{k\geq 0}$ be generated by (2), if the direction satisfies (8), and α_k satisfies the SWLS. Then the Zoutendijk condition

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty, \tag{17}$$

holds, by using (12), we conclude that the condition (17) can also be expressed as

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{18}$$

Proof. From (5) it follows that

$$(g_{k+1} - g_k)^T d_k \geq (\sigma - 1)\, g_k^T d_k.$$

On the other hand, the Lipschitz continuity (15) results in

$$(g_{k+1} - g_k)^T d_k \leq L\alpha_k \|d_k\|^2.$$

Therefore, the combination of these two relations gives

$$\alpha_k \geq \frac{\sigma - 1}{L} \frac{g_k^T d_k}{\|d_k\|^2}. \tag{19}$$

Now, using (4), (12), and (19), it results that

$$f(x_{k+1}) \leq f(x_k) + \delta \frac{(\sigma - 1)}{L} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2}. \tag{20}$$

Summing (20) for $k \geq 0$, we have

$$\delta\frac{(\sigma-1)}{L}\sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} \le (f(x_0)-f(x_1))+(f(x_1)-f(x_2))+\ldots \le f(x_0).$$

Having in view that f is bounded below, (17) is obtained.

Now, inequality (12) implies that

$$(g_k^T d_k)^2 \ge c_1^2 \|g_k\|^4, \tag{21}$$

where $c_1 = \frac{1-2\sigma}{1-\sigma}$.

Dividing by $\|d_k\|^2$ of (21), yields

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} \le \sum_{k=0}^{\infty}\frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

The theorem below demonstrates the global convergence of the MHS* method if the SWLS is used. □

Theorem 4. Assume that assumptions 1 and 2 hold. Let the iterative sequence $\{d_k\}_{k\ge 0}$ and $\{g_k\}_{k\ge 0}$ be generated by the MHS* method. Then

$$\lim_{k\to\infty}\inf\|g_k\| = 0. \tag{22}$$

Proof. If condition (22) is not satisfied, then there is a positive constant $\gamma_1 > 0$, such that

$$\|g_k\| \ge \gamma_1, \quad \text{for all } k \ge 0. \tag{23}$$

From (9), it is clear that

$$\beta_k^{MHS^*} \le \frac{\|g_{k+1}\|^2}{d_k^T(g_{k+1}-g_k)} = \beta_k^{DY}.$$

In addition, we have

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k^{DY} g_{k+1}^T d_k = \frac{\|g_{k+1}\|^2}{d_k^T(g_{k+1}-g_k)} g_k^T d_k \\ &= \beta_k^{DY} g_k^T d_k, \end{aligned}$$

which implies that

$$\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T(g_{k+1}-g_k)} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{24}$$

By used (24), we have

$$\beta_k^{MHS^*} \leq \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{25}$$

Hence, by using (3)

$$d_{k+1} + g_{k+1} = \beta_k^{MHS^*} d_k.$$

So,

$$\|d_{k+1}\|^2 = \left(\beta_k^{MHS^*}\right)^2 \|d_k\|^2 - \|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1}. \tag{26}$$

Substituting (25) into (26), we obtain

$$\|d_{k+1}\|^2 \leq \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}\right)^2 \|d_k\|^2 - \|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1}. \tag{27}$$

Divided the both sides of (27) by $\left(g_{k+1}^T d_{k+1}\right)^2$, we obtain

$$\begin{aligned}\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} &\leq \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} - \frac{\|g_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} - \frac{2}{g_{k+1}^T d_{k+1}} \\ &= \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}}\right)^2 + \frac{1}{\|g_{k+1}\|^2}. \end{aligned} \tag{28}$$

Combining with $\frac{\|d_0\|^2}{\left(g_0^T d_0\right)^2} = \frac{1}{\|g_0\|^2}$, by using (20) and a recurrence of relation (28), we have

$$\frac{\|d_{k+1}\|^2}{\left(g_{k+1}^T d_{k+1}\right)^2} \leq \frac{\|d_k\|^2}{\left(g_k^T d_k\right)^2} + \frac{1}{\|g_{k+1}\|^2} \leq \sum_{i=0}^{k+1} \frac{1}{\|g_i\|^2} \leq \frac{k+2}{\gamma_1^2}.$$

Then,

$$\sum_{k\geq 0} \frac{\left(g_k^T d_k\right)^2}{\|d_k\|^2} \geq \gamma_1^2 \sum_{k\geq 0} \frac{1}{k+1} = \infty.$$

This contradicts the Zoutendjik condition (17), concluding the proof. □

Now, we can give the global convergence result of the MPRP* method.

Theorem 5. Consider that assumptions 1 and 2 hold. Let the sequences $\{g_k\}_{k\geq 0}$ and $\{d_k\}_{k\geq 0}$ be generated by the MPRP* algorithm. Then

$$\lim_{k\to\infty} \inf \|g_k\| = 0. \tag{29}$$

Proof. Suppose that (29) does not hold. Then there exists a constant $\gamma_2 > 0$, such that

$$\|g_k\| \geq \gamma_2, \quad \text{for all } k \geqslant 0. \tag{30}$$

Using the definition of d_k $(k \geq 1)$,

$$d_{k+1} = -g_{k+1} + \beta_k^{MPRP^*} d_k.$$

Then,

$$\|d_{k+1}\|^2 = \left(\beta_k^{MPRP^*}\right)^2 \|d_k\|^2 - 2\beta_k^{MPRP^*} g_{k+1}^T d_k + \|g_{k+1}\|^2. \tag{31}$$

Also, by (6), (12) and (13),

$$-2\beta_k^{MPRP^*} g_{k+1}^T d_k \leq 2\beta_k^{MPRP^*} \left|g_{k+1}^T d_k\right| \leq \frac{-2\|g_{k+1}\|^2 \sigma g_k^T d_k}{\|g_k\|^2} \leq \frac{2\sigma \|g_{k+1}\|^2}{1-\sigma}. \tag{32}$$

Substituting (13) and (32) into (31), we obtain

$$\|d_{k+1}\|^2 \leq \frac{\|g_{k+1}\|^4}{\|g_k\|^4} \|d_k\|^2 + \left(\frac{\sigma+1}{1-\sigma}\right) \|g_{k+1}\|^2. \tag{33}$$

Divided (33) by $\|g_{k+1}\|^4$, we obtain

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{\|d_k\|^2}{\|g_k\|^4} + \left(\frac{\sigma+1}{1-\sigma}\right) \frac{1}{\|g_{k+1}\|^2}. \tag{34}$$

Noting that $\frac{\|d_0\|^2}{\left(g_0^T d_0\right)^2} = \frac{1}{\|g_0\|^2}$ and using (34) recursively yields

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \left(\frac{\sigma+1}{1-\sigma}\right) \sum_{i=0}^{k+1} \frac{1}{\|g_i\|^2} \leq \left(\frac{\sigma+1}{1-\sigma}\right) \frac{k+2}{\gamma_2^2}.$$

This implies that

$$\sum_{k\geq 0} \frac{\|g_k\|^4}{\|d_k\|^2} \geq \gamma_2^2 \left(\frac{1-\sigma}{1+\sigma}\right) \sum_{k\geq 0} \frac{1}{k+1} = \infty.$$

This contradicts the Zoutendjik condition (18), concluding the proof. □

4 Numerical experiments

To demonstrate more clearly the efficiency of the MHS* and MPRP* algorithms to some other CG algorithms famous, we have run two groups of preliminary numerical experiments for the MHS* and MPRP* CG methods, respectively.

- In Group A, we compare the MHS* with the NHS [26], NVHS* [9], MHS [25], and MDY [14] CG methods.
- In Group B, the MPRP* is compared with the NPRP [26], NVPRP* [9], PRP [19, **?**], and WYL [24] CG methods.

To evaluate the CG algorithm's performance, we utilized a collection of well-established benchmark functions commonly used to test optimization algorithms, which have been taken from the CUTE library [1, **?**] collections. All the algorithms have been coded in MATLAB 2013 and compiler settings on the PC machine (2.5 GHz, 3.8 GB RAM) with Windows XP operating system. In this numerical results, all algorithms implement the strong WLS condition with $\delta = 10^{-3}$, $\sigma = 10^{-1}$. A different parameters are mentioned in this work by $\eta_1 = 0.8$, $\eta_2 = 0.7$, $\xi_1 = 1.5$, and $\xi_2 = 1.3$.

The iteration is terminated if one of the following conditions is satisfied

(i) $\|g_k\|_\infty < 10^{-6}$,

(ii) reaching a maximum of 2000 iterations,

(iii) the computing time is more than 500 s. We also employ a performance profiling tool suggested by Dolan and Morè [8]. To test effectiveness, we measure the number of iterations and CPU time, applying the following criteria. Let S represent the set of methods and let P represent the set of test problems. Referring to n_p and n_s, these represent the number of test problems and methods, respectively. For every problem p in the set P and every solver s in the set S, let $\tau_{p,s}$ represent the count of iterations or the CPU time needed to solve problems p by solver s. The assessment of various solvers in terms of performance ratio can be expressed as follows:

$$r_{p,s} = \frac{\tau_{p,s}}{\min\{\tau_{p,i},\ 1 \le i \le n_s\}}.$$

Assume a parameter r_M such that $r_M \ge r_{p,s}$ for all selected problems and solvers and $r_{p,s} = r_M$ if and only if s is unable to solve p. The comprehensive assessment of solver performance is determined by the performance profile function as provided by

$$F_s(t) = \frac{size\{p:\ 1 \le p \le n_p,\ r_{p,s} \le t\}}{n_p},$$

where $t \ge 1$ and size$\{p:\ 1 \le p \le n_p,\ r_{p,s} \le t\}$ is the number of elements in the set $\{p:\ 1 \le p \le n_p,\ r_{p,s} \le t\}$. This function $F_s : [1, \infty[\to [0, 1]$ is the distribution function for the performance ratio. The value of $F_s(1)$ is the probability that the solver will win the rest of the solvers.

Figure 1 displays the performance profile in relation to CPU time. In terms of this metric, MHS* demonstrates the highest performance, followed by the NHS method, the MDY method, the MHS method, and the NVHS* method sequentially.

As observed in Figure 2, the MHS* curve consistently positions itself above the NHS, NVHS*, MHS, and MDY CG curves. This suggests that the MHS* algorithm surpasses the NHS, NVHS*, MHS, and MDY methods in terms of the number of iterations.

From previous numerical experiments, the MHS* algorithm can successfully solve 95.03% of the test problems 85.32%, 85.18%, 83.19%, and 83.09% of the tested problems successfully solved by MHS, NHS, NVHS *, and MDY algorithms, respectively. So, the MHS* algorithm proposed in this paper can solve more tested problems than other algorithms.

From Table 1, we can see that the average performance of the MHS*, NHS, NVHS*, MHS, and MDY methods are very similar to the results obtained from Figures 1 and 2.

In Figure 3, a performance evaluation of the MPRP* method is presented in comparison to NPRP, NVPRP*, PRP, and WYL methods. The results demonstrate the superiority of the new algorithm over all other methods in terms of CPU time, underscoring the effectiveness of the MPRP* method.

Notably, the NPRP method exhibits a behavior closely resembling that of the PRP method.

On the flip side, Figure 4 illustrates the performance profile of all methods. The conclusion drawn from this Figure is that the MPRP* method outperforms the NPRP, NVPRP*, PRP, and WYL methods in terms of the number of iterations.

From previous numerical experiments, the MPRP* algorithm can successfully solve 95.28% of the test problems 92.43%, 92.27%, 81.36%, and 78.84% of the tested problems successfully solved by PRP, NVPRP*, NPRP, and WYL algorithms, respectively. Therefore, the MPRP* algorithm is more effective in solving the tested problems than other algorithms.

From Table 2, we can see that the average performance of the MPRP*, NPRP, NVPRP*, PRP, and WYL methods are very similar to the results obtained from Figures 3 and 4.

The final conclusion is that the proposed methods are more efficient than some existing methods.

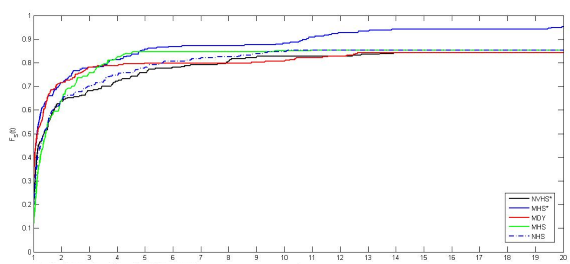

Figure 1: Performance profile on the CPU time.

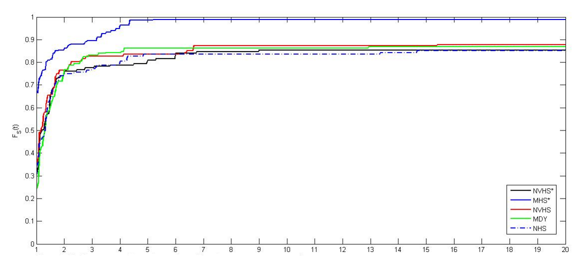

Figure 2: Performance profile on the number of iterations.

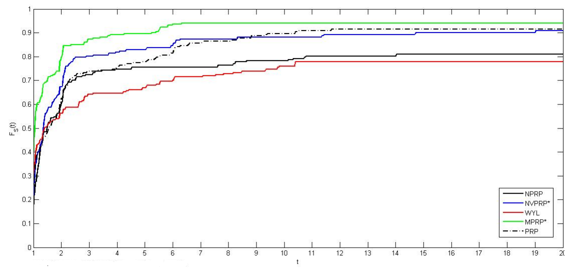

Figure 3: Performance profile the CPU time.

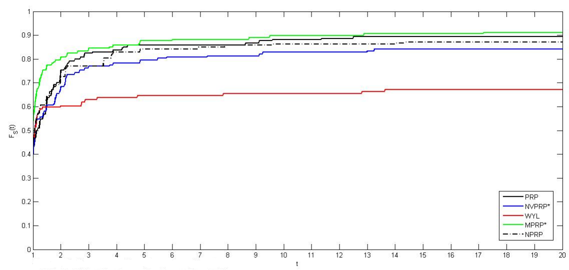

Figure 4: Performance profile on the number of iterations.

5 Application in conditional mode regression

The CG method has played an important role in solving large scale unconstrained optimization problems that may arise in statistics nonparametric [18], portfolio selection [2], and image restoration problems [17].

The regression function estimation is the most important tool for addressing nonparametric prediction problems. The study of the relationship between a variable of interest Y and a covariate X is one of the most important problems in statistics. Recent years have witnessed a renewal of interest in regression modal estimation, we refer the reader to Boente and Fraiman [3]. For any x denote by $f\left(\cdot|x\right)=\frac{f(x,\cdot)}{l(x)}$ the conditional probability density function (p.d.f) of Y given $X=x$, where $f\left(\cdot,\cdot\right)$ is the joint p.d.f. of (X,Y) and $l\left(\cdot\right)$ is the marginal density of X. Assuming that $f\left(\cdot|x\right)$ has a unique mode $\theta\left(x\right)$, the latter is given by

$$f\left(\theta\left(x\right)|x\right)=\max_{y\in\mathbb{R}} f\left(y|x\right). \tag{35}$$

Table 1: The simulation results of MHS*, NVHS*, NHS, MHS and MDY methods.

Method		MHS*		NVHS*		NHS		MHS		MDY	
Function	Dim	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR
Schwefel 223	1800	5.8930	154	12.5850	287	6.6420	6.6420	5.9890	155	6.5780	157
	1900	6.2970	156	15.8020	351	16.4540	16.4540	7.1470	169	6.3280	156
	2700	13.1950	190	29.0620	376	13.5800	13.5800	14.2060	192	13.5800	195
ZAKHAROV	800	0.2030	5	16.0780	554	23.1090	23.1090	8.6250	302	8.8050	304
	2000	0.5050	5	169.6590	1887	74.1530	74.1530	46.1030	542	33.1080	484
	3000	0.8520	5	266.6500	1634	146.812	146.812	294.4740	1658	87.2810	613
Extended Rosenbrock	200	0.0620	10	0.0780	14	0.1870	0.1870	0.1100	15	0.1090	16
	800	0.3750	20	0.5620	26	0.6090	0.6090	0.8600	36	1.9220	84
	2600	0.6090	18	0.7660	20	2.4690	2.4690	3.2660	65	1.6250	36
Quartic	1000	0.0780	4	0.1400	5	0.1400	0.1400	0.1250	5	0.1250	5
	3000	0.2010	4	0.3280	5	0.3280	0.3280	0.3280	5	0.3430	5
	3500	0.2500	4	0.3910	5	0.3910	0.3910	0.3750	5	0.3750	5
Raydan 2	2800	0.8090	17	0.8280	17	0.8750	0.8750	Inf	Inf	1.0620	22
	3000	0.9840	17	0.9530	18	1.1250	1.1250	Inf	Inf	1.2180	20
	4000	1.1090	17	1.2200	18	1.5060	1.5060	Inf	Inf	1.3900	20
Raydan 1	80	0.0780	56	0.1350	102	0.0940	0.0940	0.1400	104	0.0780	57
	120	0.1410	59	0.1800	80	0.5780	0.5780	0.1720	71	0.1560	60
	140	0.2030	81	0.2750	123	Inf	Inf	Inf	Inf	Inf	Inf
Styblinski	600	0.9690	243	Inf	Inf	5.4840	5.4840	Inf	Inf	Inf	Inf
	1000	2.5870	338	Inf	Inf	5.7970	5.7970	Inf	Inf	Inf	Inf
	2000	2.9710	244	Inf	Inf	16.0440	16.0440	Inf	Inf	Inf	Inf
Sphere	5000	0.7030	11	0.6250	11	0.4220	0.4220	0.5620	11	0.5620	11
	6000	0.5560	10	0.6560	11	0.5630	0.5630	0.7180	11	0.6720	11
	12000	1.1870	10	1.2970	11	1.0620	1.0620	1.2970	11	1.3360	11
Rastrigin	200	0.1560	32	0.1880	33	0.2190	0.2190	0.5000	66	0.2350	36
	700	0.9690	62	23.1660	738	3.0000	3.0000	23.6610	711	1.2970	69
	1600	13.8900	259	Inf	Inf	10.7030	10.7030	Inf	Inf	28.9070	483
Quadratic	1400	0.5000	8	2.7190	70	2.4380	2.4380	2.3750	66	2.8980	66
	1500	0.5320	8	4.2650	103	2.7030	2.7030	3.0630	73	2.9380	72
	1700	4.0220	8	3.4380	79	3.4840	3.4840	3.6720	78	3.5000	73
Qing	1000	0.1090	3	0.1250	3	0.1560	0.1560	0.1090	3	0.1250	4
	2800	0.2500	3	0.2660	3	0.4220	0.4220	0.2970	3	0.3750	4
	6000	0.6130	3	0.7340	3	0.7810	0.7810	0.8060	3	0.7340	4
	10000	1.0000	3	1.0310	3	1.4220	1.4220	0.9220	3	1.3910	4
Power	2400	7.1840	99	16.0780	226	1.4370	1.4370	46.0710	548	7.9230	101
	2600	0.7650	9	2.6090	33	26.5080	26.5080	19.5310	119	1.6560	30
	3000	1.4480	16	8.2970	92	3.3120	3.3120	8.1720	92	21.5130	207
	3200	3.5780	35	0.7660	12	10.7340	10.7340	3.2660	38	14.7180	145
Perquadritic	2000	0.3590	8	3.2030	46	3.1560	3.1560	3.1410	46	3.1880	46
	3200	0.8280	6	6.3120	45	6.3280	6.3280	6.3440	45	6.2810	45
	5000	0.4690	3	4.9850	21	5.0000	5.0000	4.9690	21	4.9690	21
Penalty	2000	0.8280	29	1.7650	61	1.7770	1.7770	1.2340	43	1.3060	45
	2400	1.0630	28	1.5620	52	1.9530	1.9530	1.4530	36	1.1010	32
	2800	1.5000	29	2.9540	65	2.4690	2.4690	2.4380	37	2.1880	45
	4600	2.1410	30	4.5000	65	3.0000	3.0000	2.3900	39	3.4530	48
Extended Himmelblau	1000	0.0470	3	0.0470	3	0.0470	0.0470	0.0470	3	0.0780	4
	1600	0.2030	5	0.0620	4	0.0780	0.0780	0.0780	4	0.1090	4
	3400	0.02810	3	1.7650	17	0.4620	0.4620	0.4060	5	12.2190	104
	5000	0.7820	4	3.3000	20	Int	Int	2.0940	421	16.0630	90
Hager	2000	0.6570	15	2.7970	50	6.5940	6.5940	2.7970	50	Inf	Inf
	2300	0.5790	12	1.9840	34	4.9220	4.9220	5.4060	84	2.7340	43
	2500	1.1100	19	16.5290	226	1.4690	1.4690	46.2500	548	7.6400	101
	2600	1.0820	15	11.4530	119	7.6210	7.6210	11.6270	119	2.4580	30
Griewank	3000	1.2340	22	1.6090	30	1.3750	1.3750	1.1560	23	1.7500	34
	4600	2.6870	34	3.7650	39	2.8280	2.8280	2.8900	38	2.6090	37
	5000	2.9530	35	4.1490	45	5.0790	5.0790	4.6880	50	4.7650	53
Dixon	1000	0.8400	18	0.8280	19	0.8590	0.8590	0.8590	19	0.9080	20
	1400	1.3910	14	14370	15	1.4060	1.4060	1.4380	15	1.4540	16
	5000	8.7980	21	9.6190	21	9.0540	9.0540	8.9110	21	9.3650	21
Ridge	3000	0.1690	4	0.3280	13	0.3190	0.3190	0.2410	12	0.4220	23
	1000	0.3040	13	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf
	1100	0.2080	4	Int	Int	Int	Int	Int	Int	Int	Int
	1200	0.2500	4	Int	Int	Int	Int	Int	Int	Int	Int
Sumsquares	1800	4.8280	159	9.5150	287	10.7050	10.7050	18.3190	272	5.6280	166
	3000	2.2880	36	3.0370	48	2.5720	2.5720	3.2830	47	2.4060	44
	4000	3.3400	39	4.0010	48	3.5660	3.5660	4.2710	48	3.4580	42

Table 2: The simulation results of MPRP*, NVPRP*, WYL, PRP and NPRP methods.

Method		MPRP*		NVPRP*		WYL		PRP		NPRP	
Function	Dim	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR
Schwefel 223	1000	0.0470	4	0.0630	4	0.0470	4	0.0780	4	0.0630	4
	3400	7.7040	67	10.9100	90	10.9250	89	7.9820	68	9.2580	78
	8000	6.0270	25	Inf	Inf	334.1070	965	12.3040	48	45.2360	159
Sumsquares	1000	5.6600	205	8.0410	1073	5.2070	205	5.0290	194	5.5110	211
Extended Rosenbrock	900	1.6560	106	4.0000	243	1.6590	107	1.6570	107	1.6880	107
	2000	0.7190	22	0.8130	24	0.7730	23	0.7500	23	0.7350	23
	3900	18.7120	226	38.7920	430	75.8970	711	21.1710	250	19.0480	227
	5000	30.4250	260	104.9940	782	Inf	Inf	30.4920	282	10.5610	286
Ridge	800	Inf	Inf	Inf	Inf	Inf	Inf	1.3370	234	Inf	Inf
	1700	Inf	Inf	Inf	Inf	Inf	Inf	2.6310	256	Inf	Inf
	1900	Inf	Inf	Inf	Inf	Inf	Inf	3.0630	277	Inf	Inf
Raydan 2	3000	0.1720	3	0.1720	3	0.2500	5	0.1870	4	0.1880	4
	4000	0.2040	3	0.2340	3	0.3440	3	0.3360	4	0.2190	3
Raydan 1	3200	0.3590	4	5.8910	45	6.0190	45	7.8030	45	6.8320	45
	3400	0.2190	4	8.0800	64	7.7660	64	9.6520	64	8.6580	64
	5000	0.4690	5	4.7100	21	4.6250	21	4.6530	21	5.4530	21
Styblinski	1800	0.3430	7	0.3440	7	0.3440	7	0.3490	7	0.3590	7
	7000	1.1250	6	1.4060	7	1.3440	7	1.4220	7	1.3520	7
	8000	1.4220	6	1.7810	7	1.7970	7	1.8590	7	1.8590	7
Sphere	2600	22.7710	197	52.1760	466	232.257	1508	Inf	Inf	75.5450	610
	2700	60.0100	455	249.939	1653	75.8830	603	Inf	Inf	178.269	1163
	3000	19.6130	168	273.112	1645	263.094	1560	167.216	1026	50.3130	327
	4000	85.5000	456	388.803	1758	247.534	1215	374.112	1599	262.803	1256
Rastrigin	750	0.3440	16	0.9190	39	0.9220	39	2.8910	108	3.0390	108
	1300	1.1570	31	1.0470	31	10470	31	7.3120	179	12.9600	299
	1800	0.2650	9	1.4220	31	0.5470	14	0.6560	15	0.9060	20
Quadratic	800	1.7930	126	2.2940	214	2.3100	216	2.3120	215	2.3610	214
	1400	3.9150	151	5.1660	246	5.0470	238	5.1720	246	4.9740	238
	1600	3.0780	133	4.4530	222	4.6650	229	4.4960	223	4.5360	225
Qing	1200	0.3430	7	0.5320	11	0.5150	11	0.5150	11	0.3750	8
	1600	0.4970	7	0.5150	9	0.5160	9	0.5310	9	0.3880	7
	2800	0.5620	6	0.8600	8	0.9060	8	0.8900	8	0.6390	6
Power	1200	0.6870	21	1.4610	38	2.7230	66	1.8960	49	8.3080	189
	2000	1.5220	28	2.3210	41	2.0740	38	5.5060	78	7.0570	109
	3400	0.7340	7	0.8590	8	5.5470	53	0.7190	7	63.4470	497
Perquadratic	3000	1.9020	34	2.2140	43	2.0650	39	2.3250	38	2.2270	39
	4300	2.8610	29	2.0230	28	1.9550	26	1.9740	26	2.8720	33
	5000	4.2000	44	4.3980	46	5.6870	52	5.6520	61	4.5270	49
Penalty	900	7.4370	271	20.7990	674	7.8850	274	8.8600	314	7.7280	278
	1400	13.9670	337	38.8820	827	15.0460	358	16.6210	394	14.6970	361
	1800	21.3640	393	40.7450	705	21.6250	395	25.9340	362	21.8620	400
Extended Himmelblau	1600	0.4310	15	0.6240	20	0.5790	16	0.6070	20	0.5080	17
	2400	0.5270	12	0.7640	19	1.1160	29	0.8540	19	1.3820	33
	2600	0.7040	12	1.0480	16	1.0600	21	1.0280	16	1.0000	20
Hager	4000	0.4150	10	0.5640	12	1.0400	12	0.5370	12	0.5530	12
	6000	0.6570	11	1.3760	13	0.9450	13	0.9380	13	0.7660	13
	20000	2.0940	12	2.2650	13	2.2070	13	2.3130	13	2.3670	13
Griewank	1000	0.0160	4	0.0780	44	0.0780	40	0.0310	5	0.0150	4
	1200	0.1570	4	0.0780	38	0.0940	39	0.0160	5	0.0310	5
	1500	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf
	2000	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf	Inf
Dixon	800	0.7190	17	1.3880	30	1.3440	30	0.9280	18	1.3750	31
	1960	1.0150	21	2.7960	56	1.3130	30	1.4200	25	3.3210	66
Quadratic	2000	1.0310	35	1.8130	69	1.7650	61	1.1720	43	1.3440	45
	2200	1.0780	32	2.6100	76	2.6250	68	2.0160	57	1.5940	49
	2700	1.9840	39	3.0000	61	3.3910	69	1.5940	32	2.1410	43
	3500	2.3910	38	3.3280	60	3.5940	52	3.0620	45	2.8120	42
ZAKHAROV	600	5.6480	243	14.7470	608	15.0020	619	16.9050	608	5.6750	259
	1000	12.1720	335	37.2960	940	25.2410	655	37.7970	940	12.2500	338
	2000	38.8590	494	118.938	1395	117.014	1378	118.009	1395	35.1350	498

The estimation of the conditional mode has a long history and has been studied by many authors in the statistics literature. The nonparametric estimator of conditional mode has first been considered in the case of complete data. For independent and identically distributed (i.i.d) random variables, see Samanta and Thavaneswaran [23], while Collomb, Hardle, and Hassani [5] in dependent case.

For the complete data presented, it is well known that the kernel estimator of the conditional mode $\theta(x)$ is defined as the random variable $\hat{\theta}_n\ (x)$, which maximizer the kernel estimator $\hat{f}_n(y|x)$ of $f(y|x)$, that is,

$$\hat{f}_n\left(\hat{\theta}_n(x)\,|x\right) = \max_{y\in\mathbb{R}} \hat{f}_n(y|x)\,, \tag{36}$$

where

$$\hat{f}_n(y|x) = \frac{\hat{f}_n(x,y)}{l_n(x)},$$

with

$$\hat{f}_n(x,y) = \frac{1}{nh_n^{2n}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right) H\left(\frac{y-Y_i}{h_n}\right),$$

and

$$l_n(x) = \frac{1}{nh_n^{n}} \sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right).$$

Here, the convention $\frac{0}{0} = 0$. The function K and H are a p.d.f. (so-called kernel) defined on $\mathbb{R}^n$, and (h_n) is a sequence of positive real numbers (so-called bandwidth) which goes to zero as n goes to infinity.

Simulation study

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be n independent pairs, identically distributed as (X, Y), which is a random pair valued in $\mathbb{R}^n \times \mathbb{R}^n$.

We first consider the classical linear model with normal errors

$$Y_i = X_i + \upsilon\epsilon_i.$$

Second, we consider nonlinear model (parabolic case) such that

$$Y_i = X_i^2 + \upsilon\epsilon_i,$$

where $(X_i)_{1\leq i\leq n}$ and $(\epsilon_i)_{1\leq i\leq n}$ are two i.i.d. sequences distributed as $N(0,1)$, and υ is an appropriately chosen constant (here we take $\upsilon = 0.2$).

In practice, some tuning parameters have to fixed: The kernel K is chosen by

$$K(x) = \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n} x_j^2\right),$$

and the kernel H is defined by

$$H(y) = \left(\frac{3}{4}\right)^n \prod_{j=1}^{n} \left(1 - y_j^2\right).$$

The selection of the bandwidth h is an important and basic problem in kernel smoothing techniques. In this simulation, we choose the optimal bandwidth by the cross-validation method.

In this numerical study, "Dim" denotes the dimension of the problem, "ITR" denotes the number of iterations, "TIME" denotes the "CPU" time, and "Inf" denotes the algorithm failed to yield a solution for the problem.

In this context, we employ the MHS* and MPRP* algorithms to solve the problem (36) under the SWLS technique. According to Tables 3–4, it is clear that the MHS* and MPRP* are efficient for solving the problem (36) based on the number of iterations and CPU time.

6 Conclusion

This paper presented two modified CG methods for unconstrained optimization models, that is, MHS* and MPRP* methods. Under basic assumptions, we proved that the two improved CG methods satisfy the descent condition with the SWLS and produce good convergence properties for unconstrained optimization problems.

Preliminary numerical results show that these improved methods are very robust and effective for given test problems. The practical applicability of our methods are also explored in the nonparametric estimation of the conditional mode function.

Table 3: The simulation result of MHS*, MDY, and NHS methods for solving problem (36).

Model	Initial Points	Dim	MHS*		MDY		NHS	
			ITR	TIME	ITR	ITR	TIME	ITR
Linear	$(0.2, \dots, 0.2)$	8	10	0.0620	85	10	0.0620	85
		10	12	0.1830	62	12	0.1830	62
		12	19	0.0930	41	19	0.0930	41
		14	7	0.0630	22	7	0.0630	22
		16	20	0.2820	8	20	0.2820	8
		18	9	0.1560	8	9	0.1560	8
		80	4	1.1410	5	4	1.1410	5
		100	3	4.3600	3	3	4.3600	3
	$(-0.5, \dots, -0.5)$	50	74	12.6221	462	74	12.6221	462
		54	41	8.8380	380	41	8.8380	380
		56	64	19.485	290	64	19.485	290
		62	82	31.320	50	82	31.320	50
		66	120	51.312	468	120	51.312	468
		68	10	4.9220	27	10	4.9220	27
		70	42	20.089	170	42	20.089	170
		76	11	5.4790	11	11	5.4790	11
Nonlinear	$(1, \dots, 1)$	20	189	7.5940	315	189	7.5940	315
		52	25	6.8020	44	25	6.8020	44
		56	67	20.3370	175	67	20.3370	175
		58	195	64.7330	28	195	64.7330	28
		68	77	27.4450	158	77	27.4450	158
		68	13	5.6560	28	13	5.6560	28
		70	12	5.0780	30	12	5.0780	30
		72	6	3.0190	33	6	3.0190	33
		80	10	5.2500	81	10	5.2500	81
		110	15	15.8280	Inf	15	15.8280	Inf
		120	5	4.6410	71	5	4.6410	71

Acknowledgments

The authors are very grateful to the anonymous referees for their valuable comments and useful suggestions, which improved the quality of this paper.

Table 4: The simulation result of MPRP*, NVPRP*, and PRP methods for solving problem (36).

Model	Initial Points	Dim	MPRP*		NVPRP*		PRP	
			ITR	TIME	ITR	ITR	TIME	ITR
Linear	$(0.2, \ldots, 0.2)$	8	18	0.1100	130	18	0.1100	130
		10	84	0.4690	105	84	0.4690	105
		12	7	0.0500	9	7	0.0500	9
		14	33	0.8310	20	33	0.8310	20
		16	8	0.1800	14	8	0.1800	14
		18	9	0.1560	8	9	0.1560	8
		80	3	2.7360	3	3	2.7360	3
		100	3	5.6220	3	3	5.6220	3
Nonlinear	$(-0.5, \ldots, -0.5)$	50	790	75.0800	Inf	790	75.0800	Inf
		54	268	34.2490	598	268	34.2490	598
		56	340	47.7880	Inf	340	47.7880	Inf
		62	291	59.0760	1343	291	59.0760	1343
		66	78	27.2570	54	78	27.2570	54
		68	67	26.2340	43	67	26.2340	43
		70	208	121.769	Inf	208	121.769	Inf
		76	172	94.8020	Inf	172	94.8020	Inf
		80	394	182.109	520	394	182.109	520
		84	41	31.0860	Inf	41	31.0860	Inf
		86	89	112.903	Inf	89	112.903	Inf
	$(1, \ldots, 1)$	20	48	8.49400	167	48	8.49400	167
		52	236	114.427	509	236	114.427	509
		56	69	81.4080	58	69	81.4080	58
		58	23	18.0570	43	23	18.0570	43
		68	45	88.5130	134	45	88.5130	134
		68	19	42.2830	148	19	42.2830	148
		70	24	55.9300	16	24	55.9300	16
		72	79	220.753	inf	79	220.753	Inf
		80	31	93.9150	35	31	93.9150	35
		110	27	91.1610	9	27	91.1610	9
		120	28	60.1150	31	28	60.1150	31

References

[1] Andrei, N. *An unconstrained optimization test functions*, Adv. Model. Optim. 10 (2008), 147–161.

[2] Awwal, A.M., Sulaiman, I.M., Malik, M., Mamat, M. and Kumam, P.K. *Sitthithakerngkiet, A spectral RMIL conjugate gradient method for unconstrained optimization with applications in portfolio selection and motion control*, IEEE Access, 9 (2021), 75398–75414.

[3] Boente, G. and Fraiman, R. *Nonparametric regression estimation*, J. Multivariate Anal. 29 (1989), 180–198.

[4] Bongartz, I., Conn, A.R., Gould, N. and Toint, P.L. *Constrained and unconstrained testing environment*, ACM Trans. Math. Softw. 21 (1995), 123–160.

[5] Collomb, G., Hardle, W. and Hassani, S. *A note on prediction via estimation of the conditional mode function*, J. Stat. Plan. Inference, 15 (1987), 227–236.

[6] Dai, Y.H. and Yuan, Y. *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J, Optim. 10 (1999), 177–182.

[7] Dai, Y.H. and Yuan, Y. *Nonlinear conjugate gradient methods*, Shanghai Scientific and Technical Publishers, 2000.

[8] Dolan, E.D. and Morè, J. J. *Benchmarking optimization software with performance profiles*, Math. Program. 91 (2002), 201–213.

[9] Du, X.W., Zhang, P. and Ma, W. *Some modified conjugate gradient methods for unconstrained optimization*, J. Comput. Appl. Math. 305 (2016), 92–114.

[10] Fletcher, R. *Practical methods of optimization*, Wiley, New York, 1987.

[11] Fletcher, R. and Reeves, C. *Function minimization by conjugate gradients*, Comput. J. 7 (1964), 149–154.

[12] Hager, W.W. and Zhang, H. *A survey of nonlinear conjugate gradient methods*, Pacific J. Optim. 2 (2006), 35–58.

[13] Hestenes, M.R. and Stiefel, E.L. *Methods of conjugate gradients for solving linear systems*, J. Res. Nation. Bur. Stand. 49 (1952), 409–436.

[14] Huang, H. *A new conjugate gradient method for nonlinear unconstrained optimization problems*, J. Hunan Univ. 44 (2014), 141–145.

[15] Huang, H., Wei, Z. and Yao, S. *The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search*, Appl. Math. Comput. 189 (2007), 1241–1245.

[16] Liu, Y. and Storey, C. *Efficient generalized conjugate gradient algorithms*, J. Optim. Theory Appl. 69 (1991), 129–137.

[17] Ma, G., Lin, H., Jin, W. and Han, D. *Two modified conjugate gradient methods for unconstrained optimization with applications in image restoration problems*, J.Appl. Math. Comput. (2022), 1–26.

[18] Mehamdia, A.E., Chaib, Y. and Bechouat, T. *Two modified conjugate gradient methods for unconstrained optimization and applications*, RAIRO - Oper. Res. 57 (2023), 333–350.

[19] Polak, E. and Ribière, G. *Note sur la convergence de directions conjuguée*, Revue Francaise d'Informatique et de Recherche Opérationnelle, 16 (1969), 35–43.

[20] Polyak, B.T. *The conjugate gradient method in extreme problems*, Comput. Math. Math. Phys. 9 (1969), 94–112.

[21] Powell, M.J.D. *Restart procedures for the conjugate gradient method*, Math. Program.12 (1977), 241–254.

[22] Powell, M.J.D. *Nonconvex minimization calculations and the conjugate gradient method*, Lect. Notes Math. 1066 (1984), 122–141.

[23] Samanta, M. and Thavaneswaran, A. *Nonparametric estimation of the conditional mode*, Commun. Stat. - Theory Methods, 16 (1990), 4515–4524.

[24] Wei, Z., Yao, S. and Liu, L. *The convergence properties of some new conjugate gradient methods*, Appl. Math.Comput. 183 (2006), 1341–1350.

[25] Yao, S., Wei, Z. and Huang, H. *A notes about WYL's conjugate gradient method and its applications*, Appl. Math. Comput. 191 (2007), 381–388.

[26] Zhang, L. *An improved Wei-Yao-Liu nonlinear conjugate gradient method for optimization computation*, Appl. Math. Comput. 6 (2009), 2269–2274.

[27] Zoutendijk, G. *Nonlinear programming computational methods*, Integer and Nonlinear Programming, (1970), 37–86.