APPLICATIONES MATHEMATICAE
51,2 (2024), pp. 147–161

Abd Elhamid Mehamdia and Yacine Chaib (Souk Ahras)

# IMPROVED CONJUGATE GRADIENT METHODS AND APPLICATION TO NONPARAMETRIC ESTIMATION

*Abstract.* The conjugate gradient (CG) method is one of the most important ideas in scientific computing, applied to solving linear systems of equations and nonlinear optimization problems. In this paper, based on a variant of Dai–Yuan (DY) method and Fletcher–Reeves (FR) method, two modified CG methods (named IDY and IFR) 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 effectiveness of the methods. Finally, the methods are extended to solve the problem of conditional model regression function.

**1. Introduction.** Optimization methods are widely used to obtain numerical solutions of 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 focus on the CG method for the nonconvex unconstrained optimization problem

$$\min\{f(x) : x \in \mathbb{R}^n\}, \tag{1.1}$$

where $f$ is a continuously differentiable function. The iterative procedure of the nonlinear CG method is expressed as follows:

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

2020 *Mathematics Subject Classification*: Primary 90C30; Secondary 65K05, 62G05.
*Key words and phrases*: conjugate gradient method, inexact line search, descent condition, global convergence, numerical comparisons, conditional model regression, kernel estimator.
Received 16 December 2023; revised 13 March 2024.
Published online 26 August 2024.

DOI: 10.4064/am2512-6-2024

where $\alpha_k$ is the step length computed by a certain line search and $d_k \in \mathbb{R}^n$ is defined by

$$d_{k+1} = -g_{k+1} + \beta_k d_k, \qquad d_0 = -g_0, \tag{1.3}$$

where $g_k = \nabla f(x_k)$ and $\beta_k \in \mathbb{R}$. Some example formulas for $\beta_k$ include:

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

$$\beta_k^{\mathrm{HS}} = \frac{g_{k+1}^T y_k}{y_k^T d_k} \quad [14], \qquad \beta_k^{\mathrm{PRP}} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \quad [20–21], \qquad \beta_k^{\mathrm{LS}} = \frac{g_{k+1}^{\mathrm{T}} y_k}{-g_k^T d_k} \quad [17],$$

where $y_k = g_{k+1} - g_k$, and $\| \cdot \|$ represents the Euclidean norm. In the convergence analysis of conjugate gradient methods, it is often necessary for the line search to meet the Wolfe (WLS) conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{1.4}$$

and

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k. \tag{1.5}$$

Furthermore, the strong Wolfe (SWLS) conditions consist of (1.4) and

$$|g_{k+1}^T d_k| \leq -\sigma g_k^T d_k, \tag{1.6}$$

where $0 < \delta < \sigma < 1$. In practical computations, 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 probably poor unless a reboot along the negative gradient direction is executed [22]. Despite such a drawback, it has been shown that the FR method has strong convergence properties [6]. The numerical performances of the CD and DY methods are very similar to the FR method since the scalars $\beta_k$ in these three methods have the same numerator.

In the past few years, the PRP method has generally been regarded to be one of the most efficient CG methods in practical computations. A wonderful property of the PRP method is that it automatically performs a restart if a bad direction occurs [13]. The numerical performances of the HS and LS methods are very similar to the PRP method since the coefficients $\beta_k$ in those methods have the same numerator. However, the convergence properties of the PRP, HS and LS methods are not so good [23]. In recent years, based on the above six formulas and their hybridization, many works have been published putting effort into seeking new CG methods with not only good convergence properties but also excellent numerical effects.

Wei et al. [25] introduced a modified version of the PRP method, which we refer to as the WYL method:

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

Huang et al. [16] demonstrated that the WYL method adheres to the sufficient descent condition and achieves global convergence.

Yao et al. [26] extended this concept to the HS method. This modification is referred to as the MHS approach and the parameter $\beta_k$ in this method is defined as follows:

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

The authors analyzed the sufficient descent property and global convergence when SWLS is employed [26]. Zhang [27] proposed two modified CG methods:

$$\beta_k^{\mathrm{NPRP}} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k|}{\|g_k\|^2}, \quad \beta_k^{\mathrm{NHS}} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k|}{d_k^T (g_{k+1} - g_k)}.$$

The NPRP and NHS methods have sufficient descent conditions and are globally convergent if the SWLS is utilized with $\sigma < 1/2$ [27]. Soon afterward, based on the DY CG method, Huang [15] proposed a new CG formula, where $\beta_k$ is given as

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

Huang [15] proved that the MDY method satisfies the sufficient descent condition and converges globally under the SWLS. Moreover, Du et al. [9] proposed two modified CG methods, denoted by NVHS$^*$ and NVPRP$^*$. The parameter $\beta_k$ in those methods is given by

$$\beta_k^{\mathrm{NVHS}^*} = \frac{\|g_{k+1}\|^2 - \frac{|g_{k+1}^T g_k|}{\|g_k\|^2} g_{k+1}^T g_k}{d_k^T (g_{k+1} - g_k)}, \quad \beta_k^{\mathrm{NVPRP}^*} = \frac{\|g_{k+1}\|^2 - \frac{|g_{k+1}^T g_k|}{\|g_k\|^2} g_{k+1}^T g_k}{\|g_k\|^2}.$$

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

**1.1. Motivation and some new formulas.** The two methods we presented are the result of monitoring the construction of CG parameters in the NHS and NPRP methods. Clearly, $\beta_k^{\mathrm{NHS}}$ and $\beta_k^{\mathrm{NPRP}}$ have the same expression in the numerator, namely $\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} |g_{k+1}^T g_k|$.

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

$$(1.7)\qquad \beta_k^{\mathrm{IDY}} = \frac{\|g_{k+1}\|^2 - \frac{\eta_1 (g_{k+1}^T d_k)^2 |g_{k+1}^T d_k|}{\|g_{k+1}\|\,\|d_k\|^3}}{d_k^T (g_{k+1} - g_k)}, \qquad \eta_1 \in [0,1].$$

That is, the term $\frac{\|g_{k+1}\|}{\|g_k\|}|g_{k+1}^T g_k|$ in $\beta_k^{\mathrm{NHS}}$ is replaced by $\frac{\eta_1 (g_{k+1}^T d_k)^2 |g_{k+1}^T d_k|}{\|g_{k+1}\|\,\|d_k\|^3}$ in $\beta_k^{\mathrm{IDY}}$.

Second, we define the parameters $\beta_k$ of the IFR method as follows:

$$(1.8)\qquad \beta_k^{\mathrm{IFR}} = \frac{\|g_{k+1}\|^2 - \frac{\eta_2 (g_{k+1}^T d_k)^2 |g_{k+1}^T d_k|}{\|g_{k+1}\|\,\|d_k\|^3}}{\|g_k\|^2}, \qquad \eta_2 \in [0,1].$$

The primary attributes of these methods are as follows:

- Two modified CG algorithms, based on the DY and FR method, for solving unconstrained optimization problems are developed.
- The search direction generated by these methods has descent at each iteration with a strong Wolfe line search. Furthermore, these modifications are proved to be globally convergent under usual assumptions.
- Numerical experiments demonstrate that our methods compare favorably with some existing methods for solving unconstrained optimization problems. Especially as regards the CPU time and the number of iterations, our methods are superior for the given test problems.
- The proposed methods can be successfully applied to solving the conditional model regression function problem at a lower computation cost.

**2. The sufficient descent direction and algorithm.** The following theorem shows that the search direction generated by our methods satisfies the sufficient descent conditions with the SWLS.

THEOREM 2.1. *Let the direction $d_k$ be obtained by the* IDY *or* IFR *method. If $\sigma < 1/2$, then*

$$(2.1)\qquad -\frac{1}{1-\sigma} \le \frac{g_k^T d_k}{\|g_k\|^2} \le -\frac{1-2\sigma}{1-\sigma}.$$

*So, the direction $d_k$ has sufficient descent.*

*Proof.* CASE (i): $\beta_k = \beta_k^{\mathrm{IDY}}$. From (1.7) and the Cauchy–Schwarz inequality,

$$(2.2)\qquad \beta_k^{\mathrm{IDY}} \ge \frac{\|g_{k+1}\|^2 - \frac{\eta_1 \|g_{k+1}\|^2 \|d_k\|^2 \|g_{k+1}\|\,\|d_k\|}{\|g_{k+1}\|\,\|d_k\|^3}}{d_k^T (g_{k+1} - g_k)} = \frac{\|g_{k+1}\|^2 (1-\eta_1)}{d_k^T (g_{k+1} - g_k)} \ge 0.$$

Also,

$$(2.3)\qquad \beta_k^{\mathrm{IDY}} = \frac{\|g_{k+1}\|^2 - \frac{\eta_1 (g_{k+1}^T d_k)^2 |g_{k+1}^T d_k|}{\|g_{k+1}\|\,\|d_k\|^3}}{d_k^T (g_{k+1} - g_k)} \le \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)}.$$

From (1.3), (1.6), (1.7) and (2.2), it is clear that

$$\begin{aligned} g_{k+1}^T d_{k+1} &\le -\|g_{k+1}\|^2 + \beta_k^{\mathrm{IDY}} |g_{k+1}^T d_k| \\ &\le -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{d_k^T (g_{k+1} - g_k)} |g_{k+1}^T d_k| \\ &\le -\frac{1 - 2\sigma}{1 - \sigma} \|g_{k+1}\|^2. \end{aligned}$$

On the other hand,

$$g_{k+1}^T d_{k+1} \ge -\|g_{k+1}\|^2 - \beta_k^{\mathrm{IDY}} |g_{k+1}^T d_k| \ge \frac{-1}{1 - \sigma}.$$

So, the IDY method satisfies (2.1).

Case (ii): $\beta_k = \beta_k^{\mathrm{IFR}}$. Applying the Cauchy–Schwarz inequality, we obtain

$$\beta_k^{\mathrm{IFR}} \ge \frac{\|g_{k+1}\|^2 - \eta_2 \frac{\|g_{k+1}\|^2 \|d_k\|^2 \|g_{k+1}\|\,\|d_k\|}{\|g_{k+1}\|\,\|d_k\|^3}}{\|g_{k+1}\|^2} = \frac{\|g_{k+1}\|^2 (1 - \eta_2)}{\|g_{k+1}\|^2} \ge 0.$$

On the other hand,

$$\beta_k^{\mathrm{IFR}} = \frac{\|g_{k+1}\|^2 - \frac{\eta_2 (g_{k+1}^T d_k)^2 |g_{k+1}^T d_k|}{\|g_{k+1}\|\,\|d_k\|^3}}{\|g_k\|^2} \le \frac{\|g_{k+1}\|^2}{\|g_k\|^2} = \beta_k^{\mathrm{FR}}.$$

Thus

$$(2.4)\qquad 0 \le \beta_k^{\mathrm{IFR}} \le \beta_k^{\mathrm{FR}}.$$

By (2.4) and [12, Lemma 3.1], we immediately see that the IFR method satisfies the sufficient descent condition (2.1). Therefore, the proof is complete. ∎

Now, based on the proposed conjugate parameters (1.7) and (1.8), the outlines of the corresponding algorithms are stated as follows.

IDY Algorithm

Step 1: Initializing. Select positive constants $0 < \delta < \sigma < 1/2$, choose any initial point $x_0 \in \mathbb{R}^n$, let $d_0 = -g_0$.

Step 2: Testing the iterations. If $\|g_k\|_\infty \le 10^{-6}$ is satisfied, then stop. Otherwise, go to next step.

Step 3: Line search. Find the step length $\alpha_k > 0$ satisfying the strong Wolfe line search condition (1.6), and compute $x_{k+1} = x_k + \alpha_k d_k$.

Step 4: Calculate $\beta_k$ by (1.7).

Step 5: Compute the search direction $d_k$ by (1.3).
Step 6: Let $k = k + 1$ and go to Step 2.

IFR ALGORITHM. The IFR algorithm is similar to the IDY algorithm, with the difference being that in Step 4, we replace (1.7) with (1.8).

**3. Convergence analysis.** In order to prove the global convergence of the new methods, the following assumptions are required.

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, that is, there exists a constant $L > 0$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\| \quad \text{for all } x, y \in \mathcal{N}. \tag{3.1}$$

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

$$\|\nabla f(x)\| \leq \varGamma \quad \text{for all } x \in \mathcal{N}. \tag{3.2}$$

Next, we state the famous Zoutendijk condition [28], which is essential for the global convergence of CG methods.

LEMMA 3.1. *Assume that Assumptions* 1 *and* 2 *hold. Let the sequence* $\{x_k\}_{k\geq 0}$ *be generated by* (1.2) *where* $\alpha_k$ *satisfies the SWLS conditions. Then the Zoutendijk condition holds:*

$$\sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{3.3}$$

By using (2.1), we conclude that (3.3) can also be expressed as

$$\sum_{k=0}^{\infty} \frac{\|g_k\|^4}{\|d_k\|^2} < \infty. \tag{3.4}$$

The following theorem establishes the global convergence of the IDY method with the SWLS.

THEOREM 3.1. *Suppose Assumptions* 1 *and* 2 *hold. Let* $\{g_k\}_{k\geq 0}$ *and* $\{d_k\}_{k\geq 0}$ *be produced by the IDY Algorithm. Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.5}$$

*Proof.* Suppose that (3.5) does not hold. Then there exists a constant $\gamma_1 > 0$ such that

$$\|g_k\| \geq \gamma_1 \quad \text{for all } k \geq 0. \tag{3.6}$$

According to Dai and Yuan [7], we have

$$\beta_k^{\mathrm{DY}} = \frac{\|g_{k+1}\|^2}{d_k^T(g_{k+1}-g_k)} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{3.7}$$

From (2.3) and (3.7), it is clear that

$$\beta_k^{\mathrm{IDY}} \le \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{3.8}$$

Hence, by using (1.3),

$$d_{k+1} + g_{k+1} = \beta_k^{\mathrm{IDY}} d_k,$$

so

$$\|d_{k+1}\|^2 = (\beta_k^{\mathrm{IDY}})^2\|d_k\|^2 - \|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1}. \tag{3.9}$$

Substituting (3.8) into (3.9), we obtain

$$\|d_{k+1}\|^2 \le \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}\right)^2 \|d_k\|^2 - \|g_{k+1}\|^2 - 2g_{k+1}^T d_{k+1}. \tag{3.10}$$

Dividing both sides of (3.10) by $(g_{k+1}^T d_{k+1})^2$ gives

$$\begin{aligned} \frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} &\le \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \frac{\|g_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} - \frac{2}{g_{k+1}^T d_{k+1}} \\ &= \frac{\|d_k\|^2}{(g_k^T d_k)^2} - \left(\frac{1}{\|g_{k+1}\|} + \frac{\|g_{k+1}\|}{g_{k+1}^T d_{k+1}}\right)^2 + \frac{1}{\|g_{k+1}\|^2}. \end{aligned} \tag{3.11}$$

Combining this with $\frac{\|d_0\|^2}{(g_0^T d_0)^2} = \frac{1}{\|g_0\|^2}$, by using (3.6) and the recurrence relation (3.11), we have

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

so

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

This contradicts the Zoutendjik condition (3.3), concluding the proof. ∎

Now, we can give the global convergence result for the IFR method.

THEOREM 3.2. *Suppose Assumptions* 1 *and* 2 *hold. Let the sequences* $\{g_k\}_{k\ge 0}$ *and* $\{d_k\}_{k\ge 0}$ *be generated by the IFR Algorithm. Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{3.12}$$

*Proof.* Suppose that (3.12) does not hold. Then there exists a constant $\gamma_2 > 0$ such that

$$\|g_k\| \geq \gamma_2 \quad \text{for all } k \geq 0. \tag{3.13}$$

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

$$d_{k+1} = -g_{k+1} + \beta_k^{\mathrm{IFR}} d_k,$$

so

$$\|d_{k+1}\|^2 = (\beta_k^{\mathrm{IFR}})^2 \|d_k\|^2 - 2\beta_k^{\mathrm{IFR}} g_{k+1}^T d_k + \|g_{k+1}\|^2. \tag{3.14}$$

Also, by (1.6), (2.1) and (2.4),

$$-2\beta_k^{\mathrm{IFR}} g_{k+1}^T d_k \leq 2\beta_k^{\mathrm{IFR}} |g_{k+1}^T d_k|,$$

and

$$2\beta_k^{\mathrm{IFR}} |g_{k+1}^T d_k| \leq \frac{-2\|g_{k+1}\|^2 \sigma g_k^T d_k}{\|g_k\|^2} \leq \frac{2\sigma \|g_{k+1}\|^2}{1-\sigma}. \tag{3.15}$$

Substituting (2.4) and (3.15) into (3.14), we obtain

$$\|d_{k+1}\|^2 \leq \frac{\|g_{k+1}\|^4}{\|g_k\|^4} \|d_k\|^2 + \frac{\sigma+1}{1-\sigma} \|g_{k+1}\|^2. \tag{3.16}$$

Dividing (3.16) by $\|g_{k+1}\|^4$, we obtain

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{\|d_k\|^2}{\|g_k\|^4} + \frac{\sigma+1}{1-\sigma} \frac{1}{\|g_{k+1}\|^2}. \tag{3.17}$$

Noting that $\frac{\|d_0\|^2}{(g_0^T d_0)^2} = \frac{1}{\|g_0\|^2}$, $\|g_k\| \geq \gamma_2$ and using (3.17) recursively yields

$$\frac{\|d_{k+1}\|^2}{\|g_{k+1}\|^4} \leq \frac{\sigma+1}{1-\sigma} \sum_{i=0}^{k+1} \frac{1}{\|g_i\|^2} \leq \frac{\sigma+1}{1-\sigma} \frac{k+2}{\gamma_2^2},$$

so

$$\frac{\|g_{k+1}\|^4}{\|d_{k+1}\|^2} \geq \frac{1-\sigma}{1+\sigma} \frac{\gamma_2^2}{2+k}.$$

Consequently,

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

This contradicts the Zoutendjik condition (3.4), concluding the proof. ∎

**4. Numerical experiments.** In this section, we present some numerical experiments obtained with the proposed conjugate gradient method. The test problems have been taken from the CUTE library [1, 4]. All the algorithms have been coded in MATLAB 2013 and compiler settings on a PC machine (2.5 GHz, 3.8 GB RAM) with Windows XP operating system. We compare the computational results for the IDY method against NHS [27], NVHS* [9], MHS [26] and MDY [15]. On the other hand, we compare the computational results of the IFR method against NPRP [27], NVPRP* [9], PRP [20, 21] and WYL [25]. All algorithms implement the SWLS condition with $\delta = 10^{-3}$ and $\sigma = 10^{-1}$. The iteration is terminated if one of the following conditions is satisfied: (i) $\|g_k\|_\infty < 10^{-6}$, where $\|\cdot\|_\infty$ is the maximum absolute component of a vector, (ii) the number of iterations exceeds 2000, (iii) the computing time is more than 500 s. We show the difference in performance between our methods IDY, IFR and four conjugate gradient algorithms. We choose the performance profile introduced by Dolan and Morè [8] to compare the performance according to the number of iterations and the CPU time, as follows. Let $S$ be a set of methods and $P$ a set of test problems with $n_p$, $n_s$ being the number of the test problems and the number of the methods, respectively. For each problem $p \in P$ and solver $s \in S$, denote by $\tau_{p,s}$ the number of iterations or CPU time required to solve problem $p$ by solver $s$. Then a comparison between different solvers based on the performance ratio is given by

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

Suppose there exists a parameter $r_M$, with $r_M \ge r_{p,s}$ for all problems and solvers, such that $r_M = r_{p,s}$ if and only if solver $s$ does not solve problem $p$. The overall evaluation of the performance of the solvers is then given by the performance profile function given by

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

where $t \ge 1$ and "size" is the number of elements. The 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 against the other solvers.

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

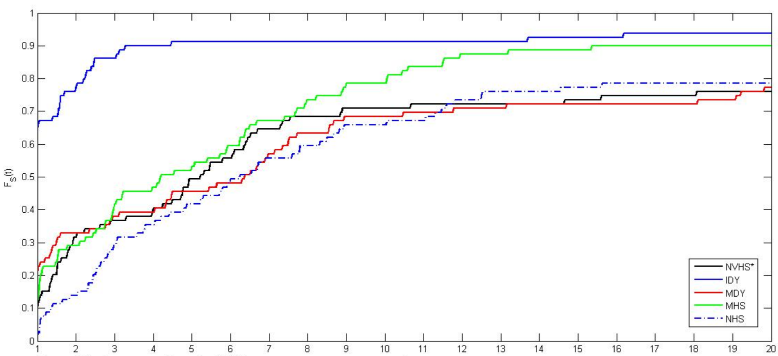


Fig. 1. Performance profile with respect to the CPU time

Figure 1 shows the performance profile for the CPU time. Relative to this metric, IDY achieves the top performance, followed by MHS, NHS, NVHS* and MDY methods.

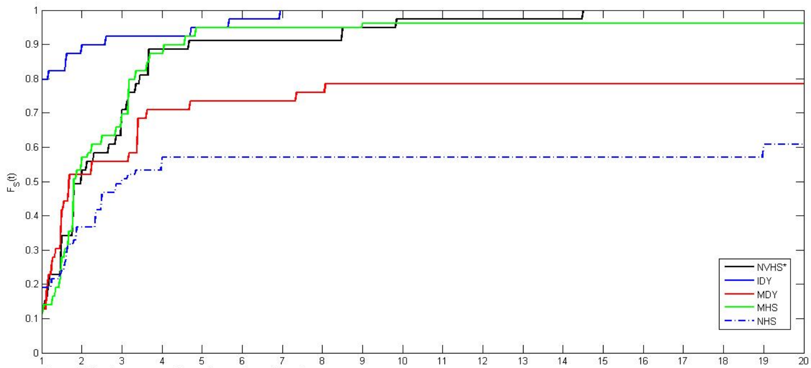


Fig. 2. Performance profile with respect to the number of iterations

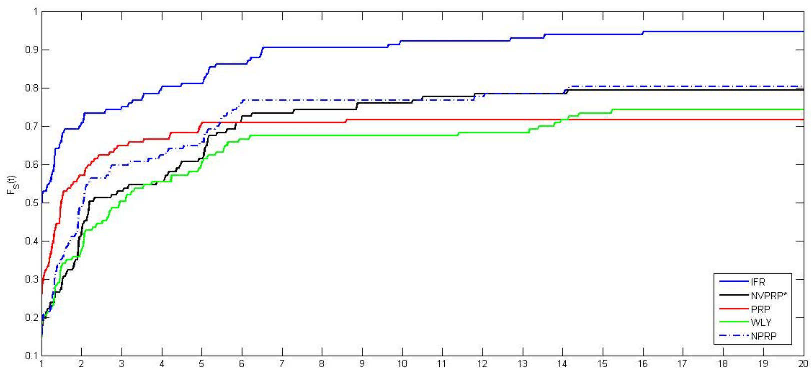


Fig. 3. Performance profile with respect to the CPU time

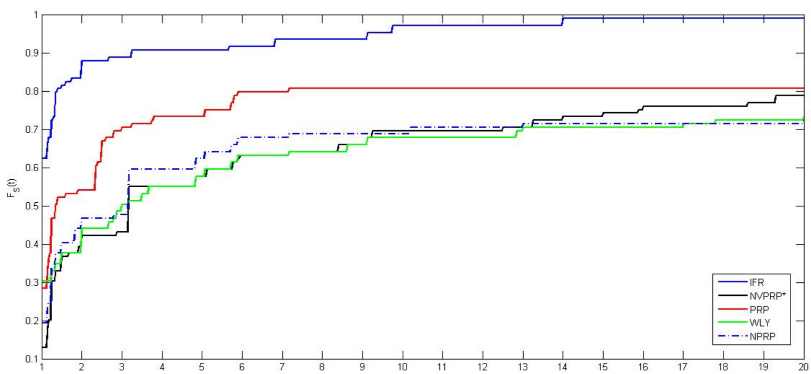


Fig. 4. Performance profile with respect to the number of iterations

It can be seen from Figure 2 that the IDY curve is mostly above the NHS, NVHS*, MHS and MDY CG curves, indicating that the IDY algorithm outperforms the NHS, NVHS*, MHS and MDY methods based on the number of iterations.

Figure 3 gives a performance comparison of the IFR method versus NPRP, NVPRP*, PRP and WYL methods. As the figure indicates, the new algorithm prevails over all other methods with respect to CPU time; this clearly confirms the effectiveness of the IFR method. The NVPRP* method behaves almost like the NPRP method.

On the other hand, Figure 4 shows the performance profile of all methods based on the number of iterations. From this figure, it can be seen that the IFR method performs better than the NPRP, NVPRP*, PRP and WYL methods, from this viewpoint.

**5. Application in conditional mode regression.** The conjugate gradient method has played an important role in solving large scale unconstrained optimization problems that may arise in nonparametric statistics [19], portfolio selection [2] and image restoration problems [18].

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 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

$$f(\cdot \mid x) = \frac{f(x,\cdot)}{l(x)}$$

the conditional probability density function (p.d.f) of $Y$ given $X = x$, where $f(\cdot,\cdot)$ is the joint p.d.f. of $(X, Y)$ and $l(\cdot)$ is the marginal density of $X$. Assuming that $f(\cdot \mid x)$ has a unique mode $\theta(x)$, it is given by

$$f(\theta(x) \mid x) = \max_{y\in\mathbb{R}^n} f(y \mid x). \tag{5.1}$$

The estimation of the conditional mode has a long history and has been studied by many authors. The nonparametric estimator of conditional mode was first considered in the case of complete data, for independent and identically distributed (i.i.d.) random variables by Samanta and Thavaneswaran [24], and by Collomb et al. [5] in the dependent case.

For complete data, the kernel estimator of the conditional mode $\theta(x)$ is defined as the random variable $\hat{\theta}_n(x)$ which maximizes the kernel estimator $\hat{f}_n(y \mid x)$ of $f(y \mid x)$, that is,

$$\hat{f}_n(\hat{\theta}_n(x) \mid x) = \max_{y\in\mathbb{R}^n} \hat{f}_n(y \mid x), \tag{5.2}$$

where

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

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 is $\frac{0}{0} = 0$. The functions $K$ and $H$ are p.d.f. (called kernels) defined on $\mathbb{R}^n$ and $(h_n)$ is a sequence of positive real numbers (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, distributed identically to $(X, Y)$ which is a random pair having values 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 a nonlinear model (parabolic case) such that

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

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

In practice, some tuning parameters have to be fixed: the kernel $K$ is chosen as

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

and the kernel $H$ is defined by

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

where $1_A$ is the indicator of a set $A$. The selection of the bandwidth $h$ is an important and basic problem in kernel smoothing techniques. In this simulation, we chose the optimal bandwidth by the cross-validation method.

In this context, we employ the IDY and IFR algorithms to solve the problem (5.2) under strong Wolfe line search technique. According to Table 1, it is clear that the IDY and IFR are efficient for solving the problem (5.2) based on number of iterations and CPU time.

**Table 1.** The simulationresult of IDY and IFR methods for solving problem (5.2)

| Model | Initial points | Dim | IDY | | IFR | |
|---|---|---|---|---|---|---|
| | | | ITR | TIME | ITR | TIME |
| | | 70 | 38 | 0.7960 | 73 | 0.2500 |
| | | 75 | 49 | 0.4220 | 127 | 1.7500 |
| | | 80 | 89 | 0.4220 | 78 | 0.7660 |
| Linear | $(0.2, \ldots, 0.2)$ | 90 | 26 | 2.6530 | 60 | 3.6560 |
| | | 95 | 24 | 2.8590 | 20 | 2.6560 |
| | | 100 | 8 | 7.8590 | 11 | 10.578 |
| | | 110 | 4 | 17.141 | 5 | 18.171 |
| | | 40 | 38 | 14.9030 | 125 | 44.0020 |
| | | 70 | 70 | 45.6850 | 49 | 34.3210 |
| | | 100 | 51 | 52.7280 | 38 | 29.4030 |
| | | 110 | 33 | 144.478 | 27 | 53.1510 |
| | | 120 | 46 | 123.983 | 50 | 130.565 |
| Nonlinear | $(0.1, \ldots, 0.1)$ | 130 | 99 | 358.897 | 43 | 141.241 |
| | | 150 | 80 | 298.807 | 27 | 103.661 |
| | | 190 | Inf | Inf | 72 | 356.302 |
| | | 220 | 10 | 48.6080 | Inf | Inf |
| | | 250 | 34 | 199.015 | 5 | 16.3270 |
| | | 260 | Inf | Inf | Inf | Inf |

**Conclusion.** This paper presented two modified conjugate gradient methods for unconstrained optimization models, that is, IDY and IFR methods. Under basic assumptions, we proved that the two methods satisfy the descent condition with the strong Wolfe line search and exhibit 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 was also explored in the nonparametric estimation of the regression function.

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

## References

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

[2] A. M. Awwal, I. M. Sulaiman, M. Malik, M. Mamat, P. Kumam and 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] G. Boente and R. Fraiman, *Robust nonparametric regression estimation*, J. Multivariate Anal. 29 (1989), 180–198.

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

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

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

[7] Y. H. Dai and Y. Yuan, *Nonlinear Conjugate Gradient Methods*, Shanghai Scientific and Technical Publ., Shanghai, 2000.

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

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

[10] R. Fletcher, *Practical Methods of Optimization*, 2nd ed., Wiley, New York, 1987.

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

[12] J. Ch. Gilbert and J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. 2 (1992), 21–42.

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

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

[15] H. Huang, *A new conjugate gradient method for nonlinear unconstrained optimization problems*, J. Henan Univ. Natural Sci. Ed. 44 (2014), 141–145.

[16] H. Huang, Z. Wei and S. Yao, *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.

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

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

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

[20] E. Polak et G. Ribière, *Note sur la convergence de méthodes de directions conjuguées*, Rev. Française Informat. Recherche Opérationnelle 3 (1969), no. 16, 35–43.

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

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

[23] M. J. D. Powell, *Nonconvex minimization calculations and the conjugate gradient method*, in: Numerical Analysis (Dundee, 1983), Lecture Notes in Math. 1066, Springer, 1984, 122–141.

[24] M. Samanta and A. Thavaneswaran, *Nonparametric estimation of the conditional mode*, Comm. Statist. Theory Methods 19 (1990), 4515–4524.
[25] Z. Wei, S. Yao and L. Liu, *The convergence properties of some new conjugate gradient methods*, Appl. Math. Comput. 183 (2006), 1341–1350.
[26] S. Yao, Z. Wei and H. Huang, *A notes about WYL's conjugate gradient method and its applications*, Appl. Math. Comput. 191 (2007), 381–388.
[27] L. Zhang, *An improved Wei–Yao–Liu nonlinear conjugate gradient method for optimization computation*, Appl. Math. Comput. 6 (2009), 2269–2274.
[28] G. Zoutendijk, *Nonlinear programming, computational methods*, in: Integer and Nonlinear Programming, North-Holland, 1970, 37–86.

Abd Elhamid Mehamdia, Yacine Chaib
Laboratory Informatics and Mathematics
Mohamed Cherif Messaadia University
Souk Ahras, 41000, Algeria
E-mail: a.mehamdia@univ-soukahras.dz
y.chaib@univ-soukahras.dz