Communications in Combinatorics and Optimization
Vol. 10, No. 2 (2025), pp. 405-421
https://doi.org/10.22049/cco.2023.28935.1784

CCO
Commun. Comb. Optim.

Research Article

A hybrid conjugate gradient method between MLS and FR in nonparametric statistics

Imane Guefassa[1,2,*], **Yacine Chaib**[1,2,†], **Tahar Bechouat**[2]

[1]Laboratory Informatics and Mathematics, Mohamed Cherif Messaadia University,
Souk Ahras, 41000, Algeria
*i.guefassa@univ-soukahras.dz
†y.chaib@univ-soukahras.dz

[2]Mohamed Cherif Messaadia University, Souk Ahras, 41000, Algeria
bachoite.t@gmail.com

Received: 19 August 2023; Accepted: 25 November 2023
Published Online: 3 December 2023

Abstract: This paper proposes a novel hybrid conjugate gradient method for nonparametric statistical inference.The proposed method is a convex combination of the modified linear search (MLS) and Fletcher-Reeves (FR) methods, and it inherits the advantages of both methods. The FR method is known for its fast convergence, while the MLS method is known for its robustness to noise. The proposed method combines these advantages to achieve both fast convergence and robustness to noise. Our method is evaluated on a variety of nonparametric statistical problems, including kernel density estimation, regression, and classification. The results show that the new method outperforms the MLS and FR methods in terms of both accuracy and efficiency.

Keywords: Hybrid conjugate gradient method, Strong Wolfe line search, Sufficient descent direction, Global convergence, Numerical comparisons, Mode function, Kernel estimator.

AMS Subject classification: 90C30, 65K05, 62G05

1. Introduction

The conjugate gradient method has played a special role in solving large-scale nonlinear optimization due to the simplicity of their iteration, very low memory requirements and good convergence analysis. For more references on advances in conjugate gradient

* *Corresponding Author*

© 2025 Azarbaijan Shahid Madani University

method see, Andrei [5, 6]. The nonlinear conjugate gradient method is designed to solve the following unconstrained optimization problem:

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

where $f : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function, bounded from below. The conjugate gradient method aims to solve the problem (1.1) starting from an initial point $x_0 \in \mathbb{R}^n$, it generates a sequence $\{x_k\}_{k\geq 0}$, such as:

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

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:

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

where $g_{k+1} = \nabla f(x_{k+1})$ is the gradient of f at x_{k+1} and the parameter β_k is known as the conjugate gradient coefficient.
The step length α_k is very important for global convergence of conjugate gradient methods. Usually, two major inexact line searches are the standard Wolfe line search, which

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

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

where $0 < \delta < \sigma < 1$. Also, the strong Wolfe conditions consist of (1.4) and

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

Now, we denote $y_k = g_{k+1} - g_k$ and $\|.\|$ the Euclidean norm. We will give some famous formulas of the parameter β_k as Polak-Ribière-Polyak (PRP) method [22, 23], Hestenes-Stiefel (HS) method [16] and Liu -Storey (LS) method [20]:

$$\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\|g_k\|^2},\ \beta_k^{HS} = \frac{g_{k+1}^T y_k}{y_k^T d_k},\ \beta_k^{LS} = \frac{g_{k+1}^T y_k}{-g_k^T d_k},$$

in general, they may not be convergent, but usually they have better numerical results. Moreover, although Fletcher-Reeves (FR) method [15], Dai-Yuan (DY) method [9] and Conjugate Decent (CD) proposed by Fletcher [14]:

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2},\ \beta_k^{DY} = \frac{\|g_{k+1}\|^2}{y_k^T d_k},\ \beta_k^{CD} = \frac{\|g_{k+1}\|^2}{-g_k^T d_k}.$$

These methods have strong convergent properties, but they may not perform well in practice due to jamming [2] and [3].
The first global convergence result for FR method was given by Zoutendijk [30] in 1970. He proved that FR method converges globally when the line search is exact. Al-Baali [1], is the first who proved the FR method converges globally when the line search is inexact. Under the strongWolfe conditions with $\sigma < \frac{1}{2}$, he proved that FR method generates sufficient descent directions. On the other hand, for the exact line search, the LS method is identical to the PRP method. Liu and Storey [20] studied this method, proving its global convergence. The techniques developed for the analysis of the PRP method may be applied to the LS method.
Throughout the years, many of the variety of the original conjugate gradient methods have been widely studied. For instance, Wei et al. [28] gave a variant of the PRP method, which we call WYL method, where the parameter β_k is obtained 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 Wei–Yao–Liu conjugate gradient method inherits the properties of PRP. Under the strong Wolfe line search with $\sigma \leq \frac{1}{4}$, Huang et al. [17] showed that the search direction of the WYL method satisfies the sufficient descent condition and the algorithm is globally convergent. Yao et al. [26] extended this idea to the LS method which we call the MLS method. The parameter β_k in the MLS method is given by:

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

If $\sigma \in \left(0, \frac{1}{2}\right)$ in the strong Wolfe line search, Yao et al. [26] proved that the MLS method also can produce sufficient descent direction and the method is globally convergent.
The hybrid conjugate gradient method is considered to combine the standard conjugate gradient methods in a two distinct ways. The first class is based on projection concept. Recently, Touati-Ahmed and Storey [27] introduced the first hybrid conjugate gradient algorithm, where the parameter β_k is computed as:

$$\beta_k^{TaS} = \min\left\{\beta_k^{FR},\ \beta_k^{PRP}\right\}.$$

The authors proved that β_k^{TaS} has good convergence properties and numerically outperforms both the β_k^{FR} and β_k^{PRP} algorithms. Also, Zhou et al. [29] combined LS method with CD method, proposing the following formula:

$$\beta_k^{H3} = \max\left\{0,\ \min\left\{\beta_k^{LS},\ \beta_k^{CD}\right\}\right\}.$$

Its global convergence under the strong Wolfe line search was proved by Zhou et al. [29].

The second class of hybrid conjugate gradient methods is based on the convex combination of the standard methods. Recently, Liu and Li [19] introduced a hybrid conjugate gradient method based on LS and DY methods (denoted as HLSDY method) for solving unconstrained optimization problem (1.1), calculating the parameter β_k^{HLSDY} as a convex combination of β_k^{LS} and β_k^{DY}, i.e:

$$\beta_k^{HLSDY} = (1-\theta_k)\,\beta_k^{LS} + \theta_k \beta_k^{DY},$$

where θ_k is a scalar parameter satisfying $0 \leq \theta_k \leq 1$. Convergence with the strong Wolfe condition was established and numerical results show that this hybrid computational scheme outperforms some sophisticated conjugate gradient methods for many problems. Also, Djordjević in [10] proposed the hybridization of LS and CD by their convex combination, which call the HLSCD method, such as:

$$\beta_k^{HLSCD} = (1-\theta_k)\,\beta_k^{LS} + \theta_k \beta_k^{CD}.$$

The compilation of parameter θ_k in β_k^{HLSCD} is in such a way that the conjugacy condition is satisfied. The global convergence of this method is proved under the strong Wolfe line search without convexity assumption on the objective function. In 2019, this author also studied the global convergence of the HLSFR method [11] under the strong Wolfe line search, such that:

$$\beta_k^{HLSFR} = (1-\theta_k)\,\beta_k^{LS} + \theta_k \beta_k^{FR}.$$

Numerical results show that this method is efficient for the standard unconstrained problems in CUTE library [4].

The aim of this paper is to propose new hybrid conjugate gradient as a convex combination of MLS and FR conjugate gradient algorithms. Under a strong Wolfe line search, we establish the convergence properties of the proposed CG method. We prove that the modified method possesses sufficient descent property independent on any line search. Numerical results show that the new hybrid method is efficient and robust and outperform as five algorithms famous. Finally, an application of our method in nonparametric mode estimator is also considered. Now, we will organize our work as follows. In the next section, we compose a new hybrid method and determine the parameter θ_k. Also we present the specific algorithm and we prove the sufficient descent condition. In section 3, we prove the global convergence of the proposed method with strong Wolfe line search. The numerical results are interpreted in section 4. An application of the new method in nonparametric statistics is given in section 5. Finally, we end up this paper with a brief summary.

2. Convex combination method

In this section, we combine FR and MLS methods to get HMLSFR method in which the parameter β_k in the presented method, denoted as β_k^{HMLSFR}, is computed as a convex combination of β_k^{MLS} and β_k^{FR}, i.e:

$$\beta_k^{HMLSFR} = (1-\theta_k)\,\beta_k^{MLS} + \theta_k \beta_k^{FR}, \tag{2.1}$$

where θ_k is a scalar parameter satisfying $0 \le \theta_k \le 1$, which follows to be determined. The search direction d_k of our algorithm is computed by:

$$d_0 = -g_0, \; d_{k+1} = -g_{k+1} + \beta_k^{HMLSFR} d_k. \tag{2.2}$$

In the conjugate gradient method, the traditional conjugacy condition $d_{k+1}^T y_k = 0$, plays an important role in the convergence analyses and numerical calculation. To select the parameter θ_k, we consider the following lemma.

Lemma 1. *If the conjugacy condition $d_{k+1}^T y_k = 0$ is satisfied at every iteration, we get*

$$\theta_k = \begin{cases} 0 & if & \frac{\zeta+\vartheta}{\rho+\vartheta} \le 0, \\ 1 & if & \frac{\zeta+\vartheta}{\rho+\vartheta} \ge 1 \\ \frac{\zeta+\vartheta}{\rho+\vartheta} & & else \end{cases}, \tag{2.3}$$

where $\vartheta = \left(\|g_{k+1}\|^2 \|g_k\|^2 - \|g_{k+1}\| \|g_k\| g_{k+1}^T g_k\right) d_k^T y_k$, $\rho = d_k^T y_k d_k^T g_k \|g_{k+1}\|^2$ *a,d* $\zeta = \|g_k\|^2 g_{k+1}^T y_k d_k^T g_k$.

Proof. We multiply both sides of the relation (2.2) by the vector y_k^T and compute β_k by (2.1) , we obtain

$$\theta_k = \frac{g_{k+1}^T y_k - \beta_k^{MLS} d_k^T y_k}{\left(\beta_k^{FR} - \beta_k^{MLS}\right) d_k^T y_k}.$$

From the formulas of β_k^{FR} and β_k^{MLS}, we get

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

Hence, Lemma 1 is proved. $\square$

2.1. Algorithm and the sufficient descent condition

The framework of the proposed HMLSFR algorithm is given as follows:
Step 1. Initialization.
Choose an initial point $x_0 \in \mathbb{R}^n$ and the parameters $0 < \delta < \sigma < \frac{1}{2}$. Compute $f(x_0)$ and g_0. Set $d_0 = -g_0$.
Step 2. Test for continuation of iterations.
If $\|g_k\|_\infty \leq 10^{-6}$, then stop. Otherwise, go to the next step.
Step 3. Line search.
Compute α_k by the strong Wolfe line searches (1.4), (1.6) and update the variables $x_{k+1} = x_k + \alpha_k d_k$.
Step 4. Compute θ_k.
If $\rho + \vartheta = 0$, then set $\theta_k = 0$, else set θ_k in (2.3)
Step 5. Compute β_k.
If $0 < \theta_k < 1$, then compute β_k by (2.1). If $\theta_k \geq 1$, then compute β_k by β_k^{FR}. If $\theta_k \leq 0$, then compute β_k by β_k^{MLS}.
Step 6. Compute the search direction.
Generate $d_{k+1} = -g_{k+1} + \beta_k^{HMLSFR} d_k$.
Step 7. Set $k = k + 1$ and go to Step 2.
Now, we prove that the search direction d_k obtained by the new hybrid conjugate gradient method satisfies in some condition the sufficient descent condition.

Theorem 1. *Let the sequences $\{d_k\}$ and $\{g_k\}$ be generated by the HMLSFR algorithm. Then the search direction d_k satisfies the sufficient descent direction*

$$g_k^T d_k \leq -c \|g_k\|^2, \ \forall\ k \geqslant 0. \tag{2.4}$$

Proof. The following proof is by induction. For $k = 0$, it holds $d_0 = -g_0$ then $g_0^T d_0 = -\|g_0\|^2$, we conclude that the sufficient descent condition holds for $k = 0$. Now, we assume (2.4) holds for k and prove that for $k + 1$.
From (2.1) and (2.2), we have

$$d_{k+1} = -g_{k+1} + ((1 - \theta_k)\beta_k^{MLS} + \theta_k \beta_k^{FR}) d_k.$$

Thus, we can obtain

$$d_{k+1} = \theta_k d_{k+1}^{FR} + (1 - \theta_k) d_{k+1}^{MLS}. \tag{2.5}$$

Multiplying (2.5) by g_{k+1}^T from the left, we get

$$g_{k+1}^T d_{k+1} = \theta_k g_{k+1}^T d_{k+1}^{FR} + (1 - \theta_k) g_{k+1}^T d_{k+1}^{MLS}. \tag{2.6}$$

Firstly, let $\theta_k = 0$. Then $d_{k+1} = d_{k+1}^{MLS}$. So,

$$g_{k+1}^T d_{k+1}^{MLS} = -\|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|} g_{k+1}^T g_k}{-d_k^T g_k} g_{k+1}^T d_k.$$

We have since the definition of β_k^{MLS} that

$$0 \leq \beta_k^{MLS} \leqslant 2\beta_k^{CD}. \tag{2.7}$$

Using Cauchy Schwarz inequality and the second strong Wolfe line search condition (1.6), we have

$$g_{k+1}^T d_{k+1}^{MLS} \leq -c_1 \|g_{k+1}\|^2. \tag{2.8}$$

where $c_1 = 1 - 2\sigma$.
Secondly, let $\theta_k = 1$. We get

$$d_{k+1} = d_{k+1}^{FR}.$$

The FR method satisfies the sufficient descent condition, see Theorem 1 in Al-Baali [1], so

$$g_{k+1}^T d_{k+1}^{FR} \leq -c_2 \|g_{k+1}\|^2, \ \forall k \geq 0, \tag{2.9}$$

where $c_2 = \frac{1-2\sigma}{1-\sigma}$, et $0 < \sigma < \frac{1}{2}$.
Finally, suppose that $0 < \theta_k < 1$, i.e. there exists two positive constants a_1 and a_2, such as

$$0 < a_1 \leqslant \theta_k \leqslant a_2 < 1.$$

From the relation (2.6), we conclude

$$g_{k+1}^T d_{k+1} \leqslant a_1 g_{k+1}^T d_{k+1}^{FR} + (1 - a_2)\, g_{k+1}^T d_{k+1}^{MLS}.$$

Using (2.8) and (2.9), we have

$$g_{k+1}^T d_{k+1} \leqslant -c \|g_{k+1}\|^2, \tag{2.10}$$

where $c = a_1 c_1 + (1 - a_2)\, c_2$. Thus, we complete the proof. $\square$

3. Global convergence of the proposed method

In order to establish the global convergence of our method, we need the following basic assumptions on the objective function.

Assumption A: The level set

$$S = \{x \in \mathbb{R}^n : f(x) \leq f(x_0)\},$$

is bounded.

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

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

From assumption B we can deduce that for all $x \in \mathcal{N}$ there exists a positive constant $\Gamma \geq 0$, such that

$$\| \nabla f(x) \| \leq \Gamma, \quad \forall x \in \mathcal{N}. \tag{3.2}$$

It follows from Dai et al. [8] it is proved that for any conjugate gradient method with strong Wolfe line search, it holds.

Lemma 2. *Let Assumptions A and B hold. Consider the method (1.2) and (1.3), where d_k is a descent direction, and α_k is obtained by the strong Wolfe line search. If*

$$\sum_{k \geq 0} \frac{1}{\|d_k\|^2} = \infty, \tag{3.3}$$

then

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

Now, we need also this Lemma to prove the convergence of our method.

Lemma 3. *Let Assumptions A and B hold and the sequence $\{x_k\}$ is obtained by the HMLSFR method, α_k satisfies the strong Wolfe conditions (1.4) and (1.6). Then*

$$\alpha_k \geq \frac{(1-\sigma) \mid g_k^T d_k \mid}{L \| d_k \|^2}.$$

Proof. From (1.6), we have $(g_{k+1}^T d_k - g_k^T d_k) \geq (\sigma - 1) g_k^T d_k$. Using the Cauchy Schwarz inequality and (3.1), it holds that $(\sigma - 1) g_k^T d_k \leq (g_{k+1} - g_k)^T d_k \leq L\alpha_k \| d_k \|^2$. By combining these two inequalities, the result can be achieved. □

This indicates that α_k obtained by the HMLSFR method is not equal to zero, hence there exists a constant $\lambda > 0$, such as

$$\alpha_k \geq \lambda, \quad \text{for all } k \geq 0. \tag{3.4}$$

The following Theorem is established to the global convergence of the HMLSFR method with the strong Wolfe line search.

Theorem 2. *Consider the iterative method in the form (1.2) and(1.3), with the conjugate gradient parameter β_k^{HMLSFR} defined by (2.1), suppose that Assumptions A and B hold. Then either $g_k = 0$, for some k, or*

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

Proof. Suppose by contradiction that (3.5) does not hold, then there exists a positive constant γ, such that

$$\|g_k\| \geq \gamma, \text{ for all } k. \tag{3.6}$$

Since $0 \leq \theta_k \leq 1$, from (2.1) we have

$$\beta_k^{HMLSFR} \leq \beta_k^{MLS} + \beta_k^{FR}. \tag{3.7}$$

The sufficient descent condition holds for the CD method too see [12], then

$$\beta_k^{CD} \leq \frac{\Gamma^2}{c_1\gamma^2}. \tag{3.8}$$

On the other side,

$$\beta_k^{FR} \leq \frac{\Gamma^2}{\gamma^2}. \tag{3.9}$$

From (2.7), (3.8) and (3.9), we have

$$\left|\beta_k^{HMLSFR}\right| \leq \frac{\Gamma^2}{\gamma^2}\left(\frac{2}{c_1} + 1\right) = E. \tag{3.10}$$

Thus, it follows from (2.2) ,(3.4) and (3.10) that

$$\|d_{k+1}\| \leq \| g_{k+1} \| + \left|\beta_k^{HMLSFR}\right| \frac{\| x_{k+1} - x_k \|}{\alpha_k} \leq M, \tag{3.11}$$

where $M = \Gamma + E\frac{D}{\lambda}$ and D is a diameter of the level set $\mathcal{N}$.
By taking the summation $k \geq 0$, $\sum_{k\geq 0} \frac{1}{\|d_k\|^2} = \infty$.
So, applying Lemma 2, we conclude that (3.5) is true. This is a contradiction with (3.6), so we have proved (3.5). □

4. Numerical Experiments

In this section, we present some numerical experiments obtained with the new proposed conjugate gradient method with the hybridization parameter β_k given by (2.1). In this numerical study, we selected 35 unconstrained optimization test problems in Table 1 have been taken to the CUTE library [4] and [7] collection. Dimensions of the test problems vary from 2 to 10000.
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. We compare the computational results of our method (HMLSFR method) against the FR [15], MLS [26], HLSCD [10], HLSFR [11], H3 [29]. In this numerical result, all algorithms implement the strong Wolfe line search condition with $\delta = 10^{-4}$ and $\sigma = 10^{-3}$. The iteration is terminated if one of the following conditions is satisfied (i) $\|g_k\|_\infty < 10^{-6}$, where $\|.\|_\infty$ is the maximum absolute component of a vector, (ii) The number of iterations exceeded 2000, (iii) The computing time is more than 500s. We show the performance difference clearly between the our method HLMLSFR and five conjugate gradient algorithms.
We use the performance profile introduced by Dolan and Moré [12] to compare the performance according to number iteration and CPU time to rule as follows. Let S is the set of methods and P is the set of the test problems with n_p , n_s are 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 $\tau_{p,s}$ be the computing time (the number of iterations or CPU time) required to solve problems $p \in P$ by solver $s \in S$. Then comparison between different solvers based on the performance ratio is given by

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

Suppose that a parameter $r_M \geq r_{p,s}$ for all problems and solvers chosen, and $r_M = r_{p,s}$ 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 Figure 1 and Figure 2 give a performance comparison of the HMLSFR method with those methods for the number of iterations and the CPU time. It very well may be seen from these Figures, it is shown that the method of HMLSFR is superior when compared to FR [15], MLS [26], HLSCD [10], HLSFR [11] , H3 [29] with the least duration of CPU time corresponds to those in iteration number and vice versa. The highest percentage of successful comparison is with HMLSFR at 98.82%, followed by MLS which is 95.29%. However, the successful rate comparison for HLSCD, H3, FR and HLSFR are low at 94.82%, 84.94% , 80% and 78.35% respectively. Hence, our method (HMLSFR) successfully solved the test problems, and it is competitive with the well-known conjugate gradient methods for unconstrained optimization. From Table 2, it is clear that the average performance of the HMLSFR, FR, MLS, HLSCD, HLSFR and H3 CG methods is very similar to the results obtained from Figs 1-2. From these Figures and Table 2 of the simulation results we can see that the new method HMLSFR performs better than the FR [15], MLS [26], HLSCD [10], HLSFR

[11], H3 [29] methods for the given test problems. These obtained preliminary results are indeed encouraging,

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

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.

Table 1. The list of test problems.

Number	Function	Number	Function
01	Zakharov	19	Griewank
02	Sumsquares	20	Exponential
03	Styblinski	21	Dixon
04	Sphere	22	Diagonal 2
05	Schwefel 223	23	Diagonal 1
06	Schwefel	24	Chang
07	Rosenbrock	25	Alpine 1
08	Ridge	26	Matyas
09	Raydan 2	27	Leon
10	Raydan 1	28	Branin
11	Rastrigin	29	Booth
12	Quartic	30	Beale
13	Quadratic	31	Nondia
14	Qing	32	Diag
15	Perquadratic	33	Linear Perturbed
16	Penalty	34	Cube
17	Himmelblau	35	Liarwhd
18	Hager		

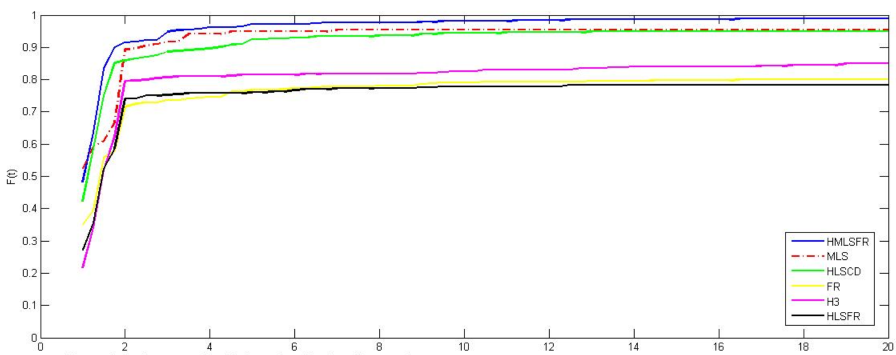

Figure 1. **Performance profile based on the iteration number**

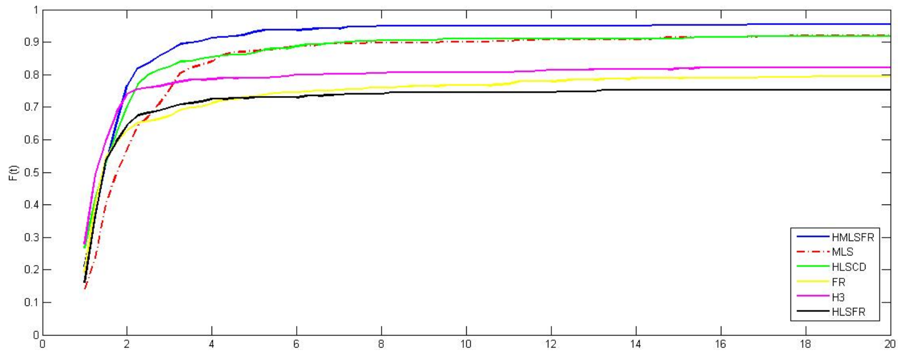

Figure 2. Performance profile based on the CPU time

Table 2: The simulation results.

Methods		HMLSFR		FR		MLS		H3		HLSCD		HLSFR	
F	Dim	TIME	ITR	TIME	ITR	TIME	ITR	Time	ITR	TIME	ITR	TIME	ITR
1	100	0.051	3	0.046	3	0.042	3	0.041	4	0.041	4	0.036	3
2	100	0.696	121	0.126	106	0.126	7	0.677	116	0.636	112	0.615	107
	500	7.146	288	5.980	238	0.435	7	6.394	245	6.240	253	6.340	252
3	100	0.052	6	0.051	6	0.100	6	0.075	6	0.086	6	0.055	6
	5000	2.111	6	2.152	6	3.340	6	2.691	6	2.801	6	2.194	6
4	1000	0.077	3	0.371	10	0.340	7	0.093	3	0.086	3	3.003	85
	2000	0.086	2	0.720	10	0.522	7	0.152	3	0.152	3	Inf	Inf
	5000	0.205	2	1.996	12	1.415	7	0.379	3	0.360	3	0.515	4
5	200	0.047	4	0.025	3	0.033	3	0.023	3	0.041	4	0.024	3
6	25	0.005	3	0.005	2	0 .109	9	0.004	2	0.012	4	0.005	2
7	2000	0.437	5	0.312	4	0.390	3	0.265	4	0.390	5	0.593	5
	5000	0.998	5	0.733	4	0.842	3	0.686	4	0.921	5	1.482	5
	10000	1.825	5	1.466	4	1.419	3	1.373	4	1.125	5	2.932	5
8	30	0.031	27	Inf	Inf	0.063	26	1.108	707	Inf	Inf	Inf	Inf
9	1000	0.031	1	0.046	2	0.062	2	0.031	2	0.031	1	0.047	2
	1500	0.047	1	0.047	2	0.063	2	0.031	2	0.047	1	0.046	2
10	1000	0.031	1	0.047	2	0.047	2	0.047	2	0.063	1	0.047	2
	5000	0.203	1	0.171	2	0.249	2	0.265	2	0.250	1	0.203	2
	10000	0.374	1	0.343	2	0.452	2	0.546	2	0.421	1	0.375	2
11	60	0.031	5	3.307	816	0.109	8	Inf	Inf	0.140	21	0.172	30
12	500	0.281	5	0.203	5	0.281	4	0.231	6	0.296	5	0.156	5
13	500	0,031	15	0.015	15	0.047	12	0.015	15	0.032	15	0.032	15
14	800	0.437	6	0.593	5	0.406	6	0.827	7	0.265	6	0.655	5
	900	0.374	5	0.593	5	0.390	6	0.765	6	0.281	5	0.546	5
15	600	0.062	2	0.031	2	0.344	9	0.031	2	0.031	2	0.047	3
	2000	0.266	4	0.187	3	0.889	9	0.141	3	0.156	3	0.187	4
16	4000	0.125	5	0.109	4	0.249	4	0.078	5	0.109	5	0.093	4
	5000	0.250	5	0.203	4	0.406	4	0.156	4	0.203	5	0.203	4
	10000	1.622	5	0.765	4	1.264	4	0.531	4	0.765	5	0.639	4
17	500	0.125	4	0.094	4	0.156	6	0.125	5	0.219	4	0.094	4
	1000	0.219	4	0.171	4	0.296	6	0.218	5	0.421	4	0.172	4
18	1000	0.047	1	0.031	2	0.047	2	0.047	2	0.032	1	0.047	2
	5000	0.187	1	0.249	2	0.265	2	0.219	2	0.188	1	0.219	2

Table 2: The simulation results. (Continued)

Methods		HMLSFR		FR		MLS		H3		HLSCD		HLSFR	
F	Dim	TIME	ITR	TIME	ITR	TIME	ITR	Time	ITR	TIME	ITR	TIME	ITR
19	2000	0.093	1	1.591	35	0.125	1	8.810	197	0.109	1	8.610	203
	5000	0.234	1	2.667	44	0.297	1	31.20	288	0.328	1	41.90	294
	10000	0.422	1	8.845	44	0.281	1	66.78	347	0.515	1	40.37	294
20	1600	0.234	3	2.168	26	0.140	2	0.125	2	0.171	3	0.156	3
	2000	0.281	3	1.669	21	0.140	2	0.140	2	0.219	3	0.156	3
21	400	0.141	5	0.093	4	0.109	3	0.110	5	0.125	5	0.093	5
22	3000	0.009	1	0.010	2	0.010	2	0.008	2	0.007	1	0.006	2
	5000	0.018	1	0.020	2	0.020	2	0.019	2	0.019	1	0.015	2
	8000	0.044	1	0.037	2	0.047	2	0.038	2	0.035	1	0.030	2
23	3000	0.002	1	0.223	76	0.003	2	1.843	606	0.002	1	1.923	603
	5000	0.005	1	0.458	93	0.005	2	3.718	664	0.005	1	3.812	663
	8000	0.009	1	1.482	110	0.010	2	9.759	793	0.009	1	10.311	811
24	50	0.067	4	0.040	3	0.074	5	0.042	4	0.036	5	0.034	3
25	88	2.626	6	2.661	6	1.320	2	1.666	4	1.668	4	4.567	10
26	2	0.011	2	0.015	2	0.033	2	0.100	2	1.119	67	0.025	3
27	2	0.066	4	0.090	4	0.078	5	0.064	5	0.073	5	0.070	4
28	2	0.032	6	0.033	6	Inf	Inf	0.034	6	0.038	6	0.036	6
29	2	0.021	3	0.027	3	0.024	2	0.015	2	0.056	5	0.066	5
30	2	0.061	4	0.076	4	0.086	4	0.073	4	0.226	18	0.072	4
31	300	0.097	3	0.087	3	0.139	3	0.085	3	0.074	3	0.064	3
	1000	0.264	3	0.313	3	0.350	3	0.389	3	0.303	3	0.182	3
32	1000	0.195	4	0.154	4	0.292	3	0.184	4	0.439	5	0.183	3
	1500	0.262	4	0.225	4	0.319	3	0.263	4	0.641	5	0.265	3
33	60	0.058	48	0.061	51	Inf	Inf	0.063	50	0.064	52	0.068	51
34	150	0.091	3	0.086	3	0.082	2	0.073	3	0.072	3	0.082	3
35	1000	0.184	4	0.167	4	0.238	3	0.174	4	0.390	5	0.181	2
	1500	0.263	4	0.212	4	0.316	3	0.259	4	0.563	5	0.257	2

5. Application in mode function

Nonparametric estimation has received a great deal of attention in both theoretical and applied statistics literature. For the historical and mathematical survey, we refer the reader to Sager [24]. In statistics, it is always interesting to study the central tendency of the data, that is usually quantified using the location parameters (mean, mode, median). The problem of estimating the mode function of a probability density function (p.d.f.) has taken considerable attention in the past for both independent and dependent data, and a number of distinguished papers deal with this topic. For example, Parzen [21] and Eddy [13] for estimation of the unconditional mode in the independent and identically distributed (i.i.d.) case.

In this section, we consider the problem of estimating the mode of a multivariate unimodal probability density f with support in $\mathbb{R}^n$ from i.i.d. standard normal random variables $X_1, \ldots, X_n$ with common probability density function f. This problem has been investigated in numerous papers. To quote a few of them, Konakov [18] and Samanta [25]. We assume that density f has a unique mode denoted by θ and defined by

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

A kernel estimator of the mode θ is defined as the random variable $\hat{\theta}$ which maximizer the kernel estimator $f_n(x)$ of $f(x)$, that is

$$f_n\left(\hat{\theta}\right) = \max_{x \in \mathbb{R}^n} f_n(x), \tag{5.2}$$

where

$$f_n(x) = \frac{1}{nh_n^n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right). \tag{5.3}$$

The bandwidth (h_n) is a sequence of positive real numbers which goes to zero as n goes to infinity and the kernel K is a p.d.f. on $\mathbb{R}^n$.
In this simulation, we choose standard Gaussian kernel defined by

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

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 context, we employ our proposed method to solve the

Table 3: The simulation result of **HMLSFR, MLS, FR, HLSFR and H3 methods** for solving problem (5.2).

Kernel	Initial points	Dim	HMLSFR		MLS		FR		HLSFR		H3	
			TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR	TIME	ITR
Gaussien	(-1.5,……,-1.5)	20	1.038	9	0.108	2	1.733	36	0.962	19	1.531	30
		50	0.301	1	0.584	2	3.480	12	9.905	33	0.295	1
		60	0.440	1	INF	INF	11.44	27	INF	INF	5.585	13
		80	6.159	8	INF	INF	24.34	33	22.98	30	61.70	81
	(5.25,……….,5.25)	20	0.523	6	0.115	2	0.213	4	26.17	335	INF	INF
		60	0.646	1	0.853	2	2.076	3	0.868	2	1.058	2
		80	0.751	1	1.541	2	3.887	5	4.255	3	2.279	3
		100	1.163	1	2.374	2	1.166	1	5.129	3	2.364	3
	(4.25,……..,4.25)	90	0.957	1	3.724	2	3.809	4	3.819	4	3.813	4
		100	3.507	3	4.820	2	3.543	3	6.931	6	7.051	6
		150	5.327	2	9.844	2	10.58	4	5.210	2	18.48	7
		200	7.554	1	15.32	2	9.465	2	49.35	9	14.10	9
	(-3.078,……..,-3.078)	60	2.736	3	INF	INF	0.437	1	0.879	2	INF	INF
		70	5.067	6	INF	INF	INF	INF	INF	INF	0.700	1
		80	3.118	3	INF	INF	INF	INF	INF	INF	2.522	3
		90	1.267	1	1.997	2	3.934	4	0.984	1	INF	INF
		100	2.608	2	INF	INF	INF	INF	INF	INF	INF	INF
		200	14.48	3	9.870	2	196.5	40	INF	INF	45.27	9

problem (5.2) under strong Wolfe line search technique and compare our method with HLSFR [11], H3 [29], MLS [26] and FR [15] methods. We choose some initial points and we obtain the result as in the Table 3. According to this Table, it is clear that the HMLSFR method is more efficient than HLSFR [11], H3 [29], MLS [26] and FR [15] methods based on the number of iterations and CPU time for solving the problem (5.2).

6. Conclusion

In the realm of optimization, this paper introduces a novel hybrid conjugate gradient method named HMLSFR, which artfully blends the strengths of the MLS and FR conjugate gradient algorithms. Our primary objective is to enhance the convergence and overall efficiency of conventional conjugate gradient algorithms in resolving optimization problems. Utilizing the strong Wolfe line search technique, we meticulously establish the global convergence characteristics and the sufficient descent property of our proposed method. Additionally, extensive numerical results convincingly demonstrate the remarkable robustness and effectiveness of our approach. Furthermore, we delve into the practical applicability of our method in the context of nonparametric estimation of the mode function. This application showcases the versatility and potential of our method beyond traditional optimization.

Conflict of Interest: The authors declare that they have no conflict of interest.

Data Availability: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

[1] M. Al-Baali, *Descent property and global convergence of the Fletcher-Reeves method with inexact line search*, IMA J. Numer. Anal. **5** (1985), no. 1, 121–124.
https://doi.org/10.1093/imanum/5.1.121.

[2] N. Andrei, *A hybrid conjugate gradient algorithm for unconstrained optimization as a convex combination of hestenes-stiefel and dai-yuan*, Studies in Informatics and Control **17** (2008), no. 1, 55–70.

[3] ______, *Hybrid conjugate gradient algorithm for unconstrained optimization*, J. Optim. Theory Appl. **141** (2008), no. 2, 249–264.
https://doi.org/10.1007/s10957-008-9505-0.

[4] ______, *An unconstrained optimization test functions collection*, Adv. Model. Optim **10** (2008), no. 1, 147–161.

[5] ______, *Nonlinear Conjugate Gradient Methods for Unconstrained Optimization*, Springer Cham, 2020.

[6] ______, *Modern Numerical Nonlinear Optimization*, vol. 195, Springer Cham, 2022.

[7] I. Bongartz, A.R. Conn, N. Gould, and P.L. Toint, *Cute: Constrained and unconstrained testing environment*, ACM Trans. Math. Software **21** (1995), no. 1, 123–160.
https://doi.org/10.1145/200979.201043.

[8] Y.H. Dai, J. Han, G.H. Liu, D. Sun, H. Yin, and Y. Yuan, *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim. **10** (2000), no. 2, 345–358.
https://doi.org/10.1137/S1052623494268443.

[9] Y.H. Dai and Y. Yuan, *A nonlinear conjugate gradient method with a strong global convergence property*, SIAM J. Optim. **10** (1999), no. 1, 177–182.
https://doi.org/10.1137/S1052623497318992.

[10] S.S. Djordjević, *New hybrid conjugate gradient method as a convex combination of LS and CD methods*, Filomat **31** (2017), no. 6, 1813–1825.
http://www.jstor.org/stable/24902273.

[11] ______, *New hybrid conjugate gradient method as a convex combination of LS and FR methods*, Acta Math. Sin. **39** (2019), 214–228.
https://doi.org/10.1007/s10473-019-0117-6.

[12] E.D. Dolan and J.J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002), 201–213.
https://doi.org/10.1007/s101070100263.

[13] W.F. Eddy, *Optimum kernel estimators of the mode*, Ann. Statist. **8** (1980), no. 4, 870 – 882.
https://doi.org/10.1214/aos/1176345080.

[14] R. Fletcher, *Practical Methods of Optimization, Second ed.*, John Wiley & Sons, New York, 2000.

[15] R. Fletcher and C.M. Reeves, *Function minimization by conjugate gradients*, Comput. J. **7** (1964), no. 2, 149–154.
https://doi.org/10.1093/comjnl/7.2.149.

[16] M.R. Hestenes and E.L. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Inst. Stan. **49** (1952), no. 6, 409–436.

[17] H. Huang, Z. Wei, and Y. Shengwei, *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), no. 2, 1241–1245.
https://doi.org/10.1016/j.amc.2006.12.006.

[18] V.D. Konakov, *On the asymptotic normality of the mode of multidimensional distributions*, Theory Probab. Appl. **18** (1974), no. 4, 794–799.
https://doi.org/10.1137/1118104.

[19] J.K. Liu and S.J. Li, *New hybrid conjugate gradient method for unconstrained optimization*, Appl. Math. Comput. **245** (2014), 36–43.
https://doi.org/10.1016/j.amc.2014.07.096.

[20] Y. Liu and C. Storey, *Efficient generalized conjugate gradient algorithms, part 1: theory*, J. Optim. Theory Appl. **69** (1991), 129–137.

https://doi.org/10.1007/BF00940464.

[21] E. Parzen, *On estimation of a probability density function and mode*, Ann. Math. Statist. **33** (1962), no. 3, 1065–1076.
http://www.jstor.org/stable/2237880.

[22] E. Polak and G. Ribiere, *Notesurla convergence de directions conjuguées, rev*, Francaise Informat Recherche Operationelle. 3e Anne **16** (1969), no. 3, 35–43.
https://doi.org/10.1051/m2an/196903R100351.

[23] B.T. Polyak, *The conjugate gradient method in extremal problems*, USSR Comp. Math. Math. Phys. **9** (1969), no. 4, 94–112.
https://doi.org/10.1016/0041-5553(69)90035-4.

[24] T.W. Sager, *An iterative method for estimating a multivariate mode and isopleth*, J. Amer. Statist. Assoc. **74** (1979), no. 366a, 329–339.
https://doi.org/10.1080/01621459.1979.10482514.

[25] M. Samanta, *Nonparametric estimation of the mode of a multivariate density*, South African Statist. J. **7** (1973), no. 2, 109–117.
https://hdl.handle.net/10520/AJA0038271X_165.

[26] Y. Shengwei, Z. Wei, and H. Huang, *A note about WYL's conjugate gradient method and its applications*, Appl. Math. Comput. **191** (2007), no. 2, 381–388.
https://doi.org/10.1016/j.amc.2007.02.094.

[27] D. Touati-Ahmed and C. Storey, *Efficient hybrid conjugate gradient techniques*, J. Optim. Theory Appl. **64** (1990), 379–397.
https://doi.org/10.1007/BF00939455.

[28] Z. Wei, S. Yao, and L. Liu, *The convergence properties of some new conjugate gradient methods*, Appl. Math. Comput. **183** (2006), no. 2, 1341–1350.
https://doi.org/10.1016/j.amc.2006.05.150.

[29] A. Zhou, Z. Zhu, H. Fan, and Q. Qing, *Three new hybrid conjugate gradient methods for optimization*, Appl. Math. **2** (2011), no. 3, 303–308.
http://dx.doi.org/10.4236/am.2011.23035.

[30] G. Zoutendijk, *Nonlinear programming, computational methods*, Integer and Nonlinear Programming (J. Abadie, ed.), North-Holland, Amsterdam, 1970, pp. 37–86.