Palestine Journal of Mathematics

Vol. 13(3)(2024) , 808–822

A new hybrid HS-DY conjugate gradient algorithm with application in mode function

L. Oumertem and Y. Chaib

Communicated by Harikrishnan

MSC 2010 Classifications: 90C30, 65K05, 62G05.

Keywords and phrases: Unconstrained optimization, Hybrid conjugate gradient method, Global convergence, Numerical comparisons, Mode function, Kernel estimator.

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

Abstract Conjugate gradient methods are an important class of methods for unconstrained optimization, especially for large-scale problems. Recently, they have been much studied. In this paper, a new hybrid conjugate gradient algorithm is proposed and analyzed. The proposed method inherits the features of the HS, DY and NHS conjugate gradient methods. The method can generate the descent direction at every iteration, moreover, this property doesn't depend on any line search. Under the strong Wolfe line search, the global convergence of the proposed method is established. The numerical results also show the feasibility and effectiveness of our algorithm. Furthermore, the proposed algorithm EHD was extended to solve problem of mode function.

1 Introduction

The optimization model is a needful mathematical problem since it has been connected to different fields such as economics, engineering and physics. Today there are many optimization algorithms, such as Newton, quasi-Newton and bundle algorithms. Note that these algorithms fail to solve large-scale optimization problems because they need to store and calculate relevant matrices. In contrast,Conjugate gradient (CG) method is one of iterative techniques prominently used in solving unconstrained optimization problems due to its simplicity, low memory storage, and good convergence analysis. In this work, we consider the unconstrained optimization problem

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

where f is continuously differentiable and bounded from below and its gradient $g_k = \nabla f(x_k)$ is available.

Conjugate gradient methods are very important methods for solving (1.1), especially when the dimension n is large. The iterative process of a conjugate gradient method for solving (1.1) is given by

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

where x_k is the current iterate point and d_k is the search direction generated by the following rule

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

where β_k is a parameter known as the conjugate gradient coefficient. The step-length α_k is very important for global convergence of conjugate gradient methods, one often requires the line search to satisfy the standard Wolfe conditions

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

and

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

Also, the strong Wolfe conditions consist of (1.4) and

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

where $0 < \delta < \sigma < 1$.

Now, we denote $y_k = g_{k+1} - g_k$, $\|.\|$ the Euclidean norm and $s_k = x_{k+1} - x_k$.

The scalar β_k is chosen so that the methods (1.2) and (1.3) reduces to the linear conjugate gradient method in the case when f is convex quadratic and exact line search, since the gradient are mutually orthogonal, and the parameters β_k in these methods are equal. For general nonlinear function, however, a different formula for scalar β_k result in distinct nonlinear conjugate gradient methods. Some of these methods as Polak- Ribière and Polyak (PRP) method [28, 29], Hestenes-Stiefel (HS) method [17] and Liu-Storey (LS) method [23]

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

in general may not be convergent, but they often have better computational performances.

Moreover, although Fletcher-Reeves (FR) method [13], Dai-Yuan (DY) method [8] 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 convergence properties, but they may not perform well in practice due to jamming [1] and [4].

Naturally, people try to devise some new methods, which have the advantages of these two kinds of methods. Touati-Ahmed and Storey [32] introduced one of the first hybrid conjugate gradient algorithms, 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. Soon afterwards, Hu and Storey [18], Gilbert and Nocedal [15] further studied other hybrid schemes about PRP and FR methods. Dai and Yuan [9] combined DY method with HS method, proposing the following two hybrid methods

$$\beta_k^{hDY} = \max\left\{-c\beta_k^{DY}, \; \min\left\{\beta_k^{HS}, \; \beta_k^{DY}\right\}\right\},$$

$$\beta_k^{hDYz} = \max\left\{0, \; \min\left\{\beta_k^{HS}, \; \beta_k^{DY}\right\}\right\},$$

where $c = \frac{1-\sigma}{1+\sigma}$. For the standard Wolfe conditions (1.4) and (1.5), under the Lipschitz continuity of the gradient, Dai and Yuan [9] established the global convergence of these hybrid computational schemes.

Another hybrid conjugate gradient is a convex combination of the different conjugate gradient algorithms. Recently, Andrei [2] introduced a new hybrid conjugate gradient method based on HS and DY methods (denoted as HYBRID method) for solving unconstrained optimization problem (1.1), calculating the parameter β_k^c as a convex combination of β_k^{HS} and β_k^{DY} i.e.

$$\beta_k^c = (1-\theta_k)\,\beta_k^{HS} + \theta_k \beta_k^{DY},$$

where θ_k is a scalar parameter satisfying $0 \leq \theta_k \leq 1$. Convergence with the standard Wolfe condition was established. In 2009, this author [4] presented a new hybrid conjugate gradient algorithm between PRP and DY methods (denoted as CCOMB method) with the β_k is obtained by

$$\beta_k^c = (1-\theta_k)\,\beta_k^{PRP} + \theta_k \beta_k^{DY}.$$

Under the strong Wolfe line search, he proved the global convergence of this method. Recently, Liu and Li [22] proposed another hybrid conjugate gradient method as a convex combination of LS and DY method (denoted as HLSDY method) given by

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

The global convergence was established under strong Wolfe line search. Numerical result show that the method is efficient for the standard unconstrained problems in a CUTE library [3] .

In 2019, Mtagulwa and Kaelo [26] introduced another hybrid and three-term conjugate gradient method which computes β_k^{EPF} as

$$\beta_k^{EPF} = \begin{cases} \beta_k^{PRP}, & \text{if } \|g_{k+1}\|^2 > \left|g_{k+1}^T g_k\right| \\ (1-\theta_k)\,\beta_k^{NPRP} + \theta_k \beta_k^{FR}, & \text{otherwise} \end{cases},$$

where β_k^{NPRP} given in Zhang [34] by

$$\beta_k^{NPRP} = \frac{\|g_{k+1}\|^2 - \frac{\|g_{k+1}\|}{\|g_k\|}\left|g_{k+1}^T g_k\right|}{\|g_k\|^2},$$

and direction d_k defined as

$$d_0 = -g_0;\; d_{k+1} = -\left(1 + \beta_k^{EPF}\frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2}\right) g_{k+1} + \beta_k^{EPF} d_k.$$

The authors proved this method has global convergence under the strong Wolfe line search conditions.

This paper aims to propose new hybrid conjugate gradient algorithm. We establish, under a strong Wolfe line search, convergence properties of the proposed conjugate gradient method. Numerical results show that the EHD method is efficient and robust and outperforms as seven conjugate gradient methods famous. Finally, an application of our method in nonparametric mode estimator is also considered.

The rest of this paper is organized as follows. In section 2, we propose another hybrid conjugate gradient method, with combines the features of the DY method and HS method. In this section we also present the new algorithm and we prove the search direction of our method satisfies the sufficient descent condition. Section 3 includes the main convergence properties of the proposed method with strong Wolfe line search. The preliminary numerical results are presented in section 4. In section 5, we focus an application of the new method in statistics nonparametric. Finally, we make a summary of our paper.

2 Modified HS-DY hybrid conjugate gradient method

In this section, we construct a new hybrid conjugate gradient method relating to the HS and DY methods. Dai and Yuan [8] proved that the DY method always generate descent directions and converges globally with the Wolfe line conditions (1.4) and (1.5). On the other hand, the HS method is generally regarded to be one of the most efficient conjugate gradient methods, but their convergence property is not so good.

In the latest years, many works have devoted their time and effort to come up with new formulae in order to increase the efficiency and effectiveness of the DY and HS methods.

Yao et al. [33] gave a variant of the HS method which we call the MHS method. The parameter β_k in the MHS method is given by

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

If $\sigma < \frac{1}{3}$ in the strong Wolfe line search (1.6), Yao et al. [33] proved that the MHS method also can produce sufficient descent direction and global convergence. More recently, Zhang [34] took a little modification to the MHS method and constructed the NHS method as follows

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

Under the strong Wolfe line search (1.6) with the parameter σ is restricted in $\left(0, \frac{1}{2}\right)$, it has been shown that the NHS method can generate sufficient descent directions and converges globally.

Motivated by the ideas on the hybrid methods [2] and [26], this paper introduce a new hybrid choice for parameter β_k as follows

$$\beta_k^{EHD} = \begin{cases} \beta_k^{HS}, & \text{if } \|g_{k+1}\|^2 > \left|g_{k+1}^T g_k\right|, \\ (1-\theta_k)\,\beta_k^{NHS} + \theta_k \beta_k^{DY}, & \text{otherwise,} \end{cases} \tag{2.1}$$

where θ_k is a scalar parameter satisfying $0 \le \theta_k \le 1$ and the direction d_k defined as

$$d_0 = -g_0;\ d_{k+1} = -\left(1 + \beta_k^{EHD} \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2}\right) g_{k+1} + \beta_k^{EHD} d_k. \tag{2.2}$$

For convenience, we call this method as EHD method.

2.1 The conjugate condition

In 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 2.1. *If the conjugacy condition $d_{k+1}^T y_k = 0$ is satisfied at every iteration, we get*

$$\theta_k = \frac{\eta - \zeta \beta_k^{NHS}}{\zeta \mu}, \tag{2.3}$$

where $\eta = y_k^T g_{k+1}$, $\zeta = y_k^T d_k - \eta \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2}$ and $\mu = \frac{\frac{\|g_{k+1}\|}{\|g_k\|}\left|g_{k+1}^T g_k\right|}{y_k^T d_k}$.

Proof. : If $\|g_{k+1}\|^2 \le \left|g_{k+1}^T g_k\right|$, we have $\beta_k^{EHD} = \beta_k^{NHS} + \theta_k\left(\beta_k^{DY} - \beta_k^{NHS}\right)$, then from (2.2) we get

$$d_{k+1} = -g_{k+1} + \left[\beta_k^{NHS} + \theta_k\left(\beta_k^{DY} - \beta_k^{NHS}\right)\right]\left[d_k - \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} g_{k+1}\right]. \tag{2.4}$$

We multiply both sides of the relation (2.4) by the vector y_k^T, we obtain

$$\theta_k = \frac{y_k^T g_{k+1} - \beta_k^{NHS}\left[y_k^T d_k - \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} y_k^T g_{k+1}\right]}{(\beta_k^{DY} - \beta_k^{NHS})\left[y_k^T d_k - \frac{d_k^T g_{k+1}}{\|g_{k+1}\|^2} y_k^T g_{k+1}\right]}.$$

From the above equality of β_k^{DY} and β_k^{NHS}, after some algebra, we get the result. □

Remark 2.2. Having in view the relation (2.3) , we define

$$\theta_k = \begin{cases} 0 & \text{if} \quad \frac{\eta - \zeta\beta_k^{NHS}}{\zeta\mu} \leq 0 \text{ or } \zeta\mu = 0, \\ \frac{\eta - \zeta\beta_k^{NHS}}{\zeta\mu} & \text{if} \quad 0 < \frac{\eta - \zeta\beta_k^{NHS}}{\zeta\mu} < 1, \\ 1 & \text{if} \quad \frac{\eta - \zeta\beta_k^{NHS}}{\zeta\mu} \geq 1. \end{cases} \tag{2.5}$$

2.2 EHD Algorithm and the sufficient descent condition

The framework of the proposed EHD algorithm is given as follows

Step 1: Initialization.

Choose an initial point $x_0 \in \mathbb{R}^n$ and the parameters $0 < \delta < \sigma < 1$. 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 using (2.5).

Step 5: Compute β_k^{EHD} using (2.1).

Step 6: Compute the search direction. If the restart criterion of Powell condition

$$\left|g_{k+1}^T g_k\right| > 0.2 \left\|g_{k+1}\right\|^2, \tag{2.6}$$

is satisfied, then set $d_{k+1} = -g_{k+1}$, otherwise generate d_{k+1} by (2.2) .

Step 7: Set $k = k + 1$ and go to Step 2.

Now, we prove that it generates search direction d_k obtained by new hybrid conjugate gradient method satisfying in some condition the sufficient descent conditions.

Theorem 2.3. *Let the sequences $\{d_k\}_{k\geq 0}$ and $\{g_k\}_{k\geq 0}$ be generated by EHD method. Then the search direction d_k satisfies the sufficient descent for all k*

$$g_k^T d_k = -\left\|g_k\right\|^2. \tag{2.7}$$

Proof. Multiplying (2.2) by g_{k+1}^T from the left, we get

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

So, we can get

$$g_{k+1}^T d_{k+1} = -\left\|g_{k+1}\right\|^2.$$

Hence true for $k \geq 1$. The proof is completed. □

3 Global convergence

To analyze the global convergence property of our hybrid method, the following Assumptions are required. These assumptions have been used extensively in the literature for the global convergence analysis of conjugate gradient methods.

Assumption A. The level set

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

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

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

These assumptions imply that there exists a positive constant $\Gamma \ge 0$ such that

$$\| \nabla f(x) \| \le \Gamma, \text{for all } x \in N. \tag{3.2}$$

The following result was essentially proved by Dai et al. [7].

Lemma 3.1. *Let Assumptions **A** and **B** hold. Let the sequence* $\{x_k\}_{k\ge 0}$ *be generated by* (1.2) *and search direction* d_k *is a descent direction, and* α_k *is received from the strong Wolfe line search. If*

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

then

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

The following Lemma gives some interesting properties of the EHD method.

Lemma 3.2. *Let Assumptions **A** and **B** hold. If* d_k *is a descent direction and* α_k *satisfies the standard Wolfe condition* (1.5)*. Then*

$$\alpha_k \ge \frac{(1-\sigma) \| g_k \|^2}{L \| d_k \|^2}. \tag{3.3}$$

Proof. See the proof of Lemma 3.2 in Liu and Li [22]. □

Remark 3.3. From (1.6) and (2.7), the step-size α_k obtained in the EHD algorithm satisfies (3.3). This indicates, the step size α_k obtained in EHD method is not equal to zero, i.e., there exists a constant $\lambda > 0$, such that

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

The following Theorem establishes the global convergence of EHD method with the strong Wolfe line search.

Theorem 3.4. *Suppose that Assumptions **A** and **B** hold. Consider the sequences* $\{g_k\}_{k\ge 0}$ *and* $\{d_k\}_{k\ge 0}$ *generated by EHD algorithm. Then this method converges in the sense that*

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

Proof. For the sake of contradiction, assume that (3.5) doesn't hold. Then there exists a positive constant γ such that

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

We have for the definition of β_k^{NHS} and Cauchy Schwarz inequality, that

$$0 \le \beta_k^{NHS} \le \beta_k^{DY}. \tag{3.7}$$

From (2.1) and (3.7), we have

$$\left|\beta_k^{EHD}\right| \leq \left|\beta_k^{HS}\right| + \beta_k^{DY}.$$

For all k sufficiently large. By using (1.6) and from the sufficient descent condition we obtain

$$d_k^T y_k = d_k^T (g_{k+1} - g_k) \geq (1-\sigma) \left\|g_k\right\|^2. \tag{3.8}$$

So, using (3.6) we get

$$d_k^T y_k \geq (1-\sigma)\gamma^2. \tag{3.9}$$

On the other hand, using the Cauchy Schwarz inequality, (3.1) and (3.2), we obtain

$$\left|g_{k+1}^T y_k\right| \leq \|g_{k+1}\| \|y_k\| \leq \Gamma L D,$$

where D is a diameter of the level set $\mathcal{N}$.
Now we use (3.9), we have

$$\left|\beta_k^{HS}\right| \leq \frac{\Gamma L D}{(1-\sigma)\gamma^2}. \tag{3.10}$$

On the other side,

$$\beta_k^{DY} = \frac{\left\|g_{k+1}\right\|^2}{d_k^T y_k} \leq \frac{\Gamma^2}{(1-\sigma)\gamma^2}. \tag{3.11}$$

From (3.10) and (3.11) , we have

$$\left|\beta_k^{EHD}\right| \leq \frac{\Gamma}{(1-\sigma)\gamma^2} (LD + \Gamma) = E. \tag{3.12}$$

Thus, it follows from (2.2) that

$$\|d_{k+1}\| \leq \| g_{k+1} \| + \left|\beta_k^{EHD}\right| \left(\frac{\left|d_k^T g_{k+1}\right|}{\|g_{k+1}\|} + \frac{\| s_k \|}{\alpha_k} \right).$$

Cauchy Schwarz inequality, (3.4) and (3.12) yields

$$\|d_{k+1}\| \leq M,$$

where $M = \Gamma + 2E\frac{D}{\lambda}$.
By take the summation $k \geq 0$, we get

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

So, applying Lemma 3.1, we conclude that which impels 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). The test problems have been taken to the CUTE library [3] and [6] . All the algorithms have been coded in MATLAB 2013 and compiler settings on the PC machine (2.5 GHz, 3.8 GB RAM memory) with windows XP operating system. We compare the computational results of our method (EHD method) against the NHS [34], DY [8], hDYz [9], CCOMB [4], HYBRID [2], HLSDY [22] and CG_DESCENT [16] methods. 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 500 s. We show the performance difference clearly between our method EHD and seven conjugate gradient algorithms. We choose the performance profile introduced by Dolan and Morè [10] to compare the performance according to the number of iteration and CPU time with 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 of iteration or CPU time required to solve problems $p \in P$ by solver $s \in S$. Then comparsion 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 problem and solvers chosen, and $r_M = r_{p,s}$ if and only if solver s does not solve problem p. The overall evaluation of performance of the solvers is then given by the performance profile function given by

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

where $t \geq 1$ and $size\{p:\ 1 \leq p \leq n_p,\ r_{p,s} \leq t\}$ is the number of elements in the set $\{p:\ 1 \leq p \leq n_p,\ r_{p,s}$ 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.

In this numerical study, Table 1 lists the names of the test functions and Table 2 shows the performance of the eight methods which gives the number of the test problems (N°), the dimension of functions (Dim), the total number of iterations (NI), the CPU time in seconds (CPU) and 'INF' indicates that the algorithm failed to solve the problem. Table2, Figure 1 and Figure 2 give a performance comparison of the EHD method with those for the number of iterations and the CPU time. From these Figures and Table 2, we can see that the new method EHD performs better than NHS [34], DY [8], hDYz [9], CCOMB [4], HYBRID [2], CG_DESCENT [16] and HLSDY [22] methods, for the given test problems. These obtained preliminary results are indeed encouraging.

Table 1: The test functions.

Number	function	Number	function
1	Beale	21	Himmelbleau
2	Booth	22	Liarwhd
3	Branin	23	Penalty
4	Lion	24	Perquadratic
5	Matyas	25	Power
6	Almost Perturbed Quadratic	26	Qing
7	Almost Perturbed Quartic	27	Quadratic
8	Alpine 1	28	Quartic
9	Chung	29	Rastring
10	DIAG	30	Raydan 1
11	Diag-aup 1	31	Raydan 2
12	Diagonal 1	32	Ridge
13	Diagonal 2	33	Rosenbrock
14	Diagonal 4	34	Schwefel
15	Dixon	35	Schwefel 220
16	Engval 1	36	Schwefel 221
17	Exponential	37	Schwefel 223
18	Extended Hiebert	38	Styblinski
19	Greinwak	39	Sumsquares
20	Hager	40	Zakharov

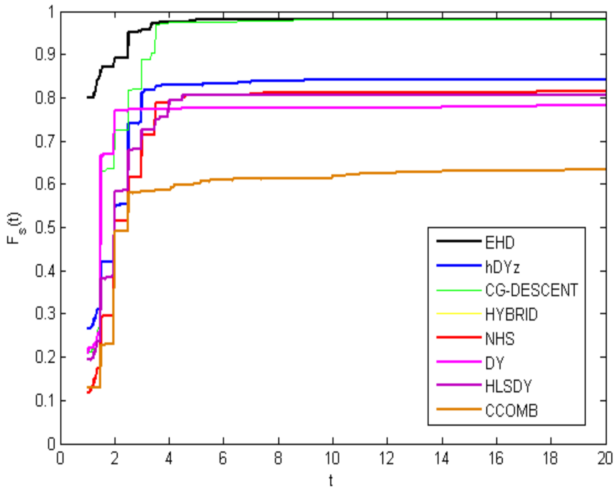

Figure 1: Performance profile on the number of iterations.

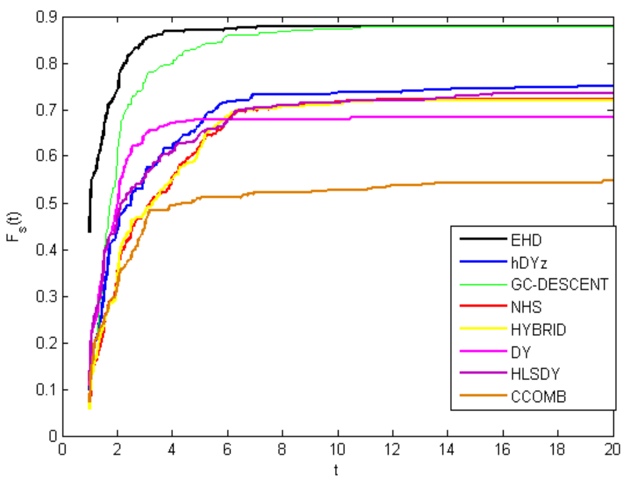

Figure 2: Performance profile on the CPU time.

Table2 : Numerical results of the eight methods.

N°	Dim	EHD		hDYz		CG-DESCENT		NHS		HYBRID		DY		HLSDY		CCOMB	
		NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU
1	2	15	0.0780	29	0.1410	10	0.0780	11	0.0630	11	0.0620	10	0.0470	80	0.3600	105	0.5310
2	2	6	0.0310	6	0.0320	8	0.0470	7	0.0320	7	0.0320	6	0.0320	8	0.0470	60	0.3440
3	2	8	0.0470	14	0 .0780	16	0.0780	10	0.0480	10	0.0620	8	0.0620	15	0.0940	88	0.5620
4	2	2	0.0150	2	0.0150	3	0.0160	3	0.0160	3	0.0160	3	0.0160	3	0.0160	5	0.0310
5	2	2	0.0150	2	0.0160	2	0.0160	3	0.0160	3	0.0320	3	0.0160	2	0.0160	2	0.0160
6	70 200 300	10 14 22	0.0310 0.0480 0.0620	840 1999 1999	2.4370 20.9230 30.6070	10 13 25	0.0310 0.0460 0.0670	84 109 194	0.2500 0.8920 2.2540	84 100 190	0.2510 0.8300 2.2020	86 140 190	0.2650 1.1410 2.2120	271 333 1999	0.8130 2.7850 32.9200	647 1999 1999	2.0680 20.7530 33.7580
7	800	2	0.0470	3	0.0940	4	0.0780	3	0.1100	3	0.0930	3	0.0940	3	0.01090	3	0.1120
8	70 300	6 11	0.1590 0.2750	42 I NF	7.4280 INF	5 11	0.1400 0.2780	1999 20	195.5290 1.8090	1999 24	198.0890 1.8200	109 INF	9.8520 INF	8 15	0.2000 0.6440	1999 INF	329.8640 INF
9	100	5	0.0310	5	0.0320	5	0.0310	9	0.0400	9	0.0350	7	0.0870	INF	INF	6	0.0330
10	200	2	0.0150	4	0.0310	5	0.0310	4	0.0160	4	0.0310	3	0.0320	4	0.0160	4	0.0160
11	600	4	0.0470	5	0.1100	6	0.0940	6	0.0930	6	0.0940	4	0.0610	4	0.0620	5	0.0680
12	3000	1	0.0180	2	0.0320	1	0.0220	2	0.0300	2	0.0210	233	5.4850	4	0.0210	1999	20.3680
13	500	2	0.0110	3	0.0160	3	0.0200	4	0.0250	4	0.0250	4	0.0280	2	0.1390	2	0.1050
14	5000	6	0.1410	2	0.0990	4	0.1110	6	0.3760	5	0.3670	3	0.1970	4	0.2030	3	0.4100
15	2000	4	0.0160	4	0.0820	4	0.0510	4	0.0205	4	0.2080	3	0.0320	5	0.0160	5	0.0310
16	50	2	0.0150	5	0.0770	5	0.0360	5	0.0160	5	0.0160	3	0.0160	7	0.0980	1999	28.2610
17	3000	3	0.1250	1999	28.3400	5	0.1400	3	0.1270	3	0.1260	4	0.1880	5	0.1410	4	0.1460
18	120	3	0.0150	5	0.0310	6	0.0320	5	0.0310	5	0.0310	4	0.0160	4	0.0460	80	1.2810
19	1000	1	0.0160	1	0.0630	1	0.0470	1	0.0460	1	0.0470	1	0.0460	2	0.0310	2	0.0160
20	2000	2	0.0150	4	0.0310	4	0.0160	3	0.0160	3	0.0310	3	0.0310	4	0.0320	5	0.0620

Table 2: (Continued).

N°	Dim	EHD		hDYz		CG-DESCENT		NHS		HYBRID		DY		HLSDY		CCOMB	
		NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU	NI	CPU
21	200	3	0.0160	6	0.0780	4	0.0470	6	0.0630	7	0.0940	4	0.0310	4	0.0310	6	0.0320
22	80	2	0.0150	5	0.0310	7	0.0310	8	0.0320	7	0.0320	3	0.0160	7	0.0310	4	0.0160
23	20 2000	2 2	0.0630 0.0620	5 5	0.1470 0.1230	2 2	0. 0680 0.1090	7 7	0.1560 0.3130	7 6	0.1600 0.2800	12 11	0.2810 0.6730	INF INF	INF INF	5 5	0.0940 0.1420
24	200 400	25 18	0.0320 0.0470	INF INF	INF INF	20 15	0.0310 0.0430	273 428	1.2960 3.7650	276 425	1.3280 3.8270	755 INF	4.2650 INF	261 390	1.2650 3.4990	1999 1999	17.5680 31.4430
25	10 500	23 4	0.0160 0.0150	8 1999	0.0150 42.3770	24 5	0.0310 0.0310	8 1999	0.0150 41.9810	8 1999	0.0150 41.8280	11 INF	0.0160 INF	INF 10	INF 0.0310	59 1999	0.9690 41.8080
26	150	3	0.0150	6	0.0310	4	0.0160	8	0.0320	8	0.0310	4	0.0160	9	0.0470	5	0.0310
27	200 1000	15 20	0.0480 0.0630	INF INF	INF INF	12 18	0.0470 0.0620	305 389	4.6080 8.2010	319 367	4.6600 8.2480	323 390	4.9550 8.3900	296 381	4.7960 8.0190	1431 1999	31.3520 66.7380
28	800	2	0.0310	4	0.0470	2	0.5620	5	0.0470	5	0.0460	3	0.0320	5	0.0470	4	0.0320
29	200	9	0.0780	215	2.0620	6	0.2030	6	0.0460	6	0.0480	9	0.0630	INF	INF	35	0.3280
30	20 1000	5 9	0.0160 0.0310	8 30	0.0310 0.0470	9 8	0.0310 0.0320	35 1999	0.5270 720.4940	35 INF	0.3510 INF	35 INF	0.5310 INF	56 INF	0.5930 INF	84 1999	0.9690 631.3020
31	5000	5	0.1100	8	0.0940	4	0.0780	8	0.0940	9	0.0970	31	4.1930	7	0.0720	653	109.3600
32	400	2	0.0620	2	0.1090	2	0.0630	2	0.0940	2	0.0930	1030	41.6090	146	0.4220	1999	22.3730
33	10	7	0.0160	6	0.0160	6	0.0930	7	0.0310	7	0.0320	82	0.0780	86	0.0940	7	0.0160
34	40	11	0.0470	3	0.0160	10	0.0160	3	0.0160	3	0.0160	4	0.0420	INF	INF	1999	71.3760
35	2000	2	0.0310	2	0.0470	3	0.0940	3	0.1400	3	0.0780	3	0.0620	3	0.0670	1999	199.9170
36	2000	3	0.0310	3	0.1860	2	0.0180	4	0.0470	4	0.0470	3	0.0940	3	0.0320	1999	42.4450
37	500	2	0.0150	2	0.0160	2	0.0320	3	0.0310	3	0.0160	3	0.0310	3	0.0160	3	0.0160
38	5000	4	0.2829	14	3.4680	5	0.2890	15	2.4210	15	2.4840	62	18.8240	37	10.0450	5	0.3680
39	200	226	1.5460	553	5.1890	224	1.4550	158	1.4210	159	1.6710	230	1.5840	810	9.2390	INF	INF
40	30	6	0.0150	6	0.0160	6	4.7490	8	0.0320	8	0.0310	13	0.0460	7	0.0320	6	0.0160

5 Application in mode function

The conjugate gradient method has played an important role in solving large scale unconstrained optimization problems that may arise in statistics nonparametric [19, 25], portfolio selection [20, 5] and image restoration problems [12, 24].
Estimation nonparametric 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 [30]. 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 [27] and Eddy [11] 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 uni-modal probability density f with support in $\mathbb{R}^n$ from i.i.d. standard normal random variables $X_1, ..., X_n$ with common probability density function f. This problem has been investigated in numerous paper. To quote a few of them, Konakov [21] and Samanta [31] . We assume that density f has an 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 between two different types of kernel: while standard Gaussian kernel defined by

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

and Epanechnikov kernel obtained by

$$K(x) = \left(\frac{3}{4}\right)^n \prod_{j=1}^{n} \left(1 - x_j^2\right).$$

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 problem (5.2) under strong Wolfe line search technique and compare the computational results of the EHD method against the CG_DESCENT method [16]. We choose some initial points and we obtain the result as in the Table 3. According to these results, it is clear that the EHD method more efficient than CG_DESCENT method based on the number of iterations and CPU time for solving the problem (5.2).

Table 3: The simulation result of EHD and GC-DESCENT methods for solving problem (5.2).

Kernel	Point initial	Dim	EHD		GC-DESCENT	
			NI	CPU	NI	CPU
Gaussian	(0.001,...,0.001)	90	6	4.7730	6	2.8770
		120	8	10.5020	27	18.8140
		200	2	7.1660	3	8.9070
	(0.025,...,0.025)	40	3	0.4610	2	0.1400
		50	17	3.7430	78	8.6610
		250	3	16.4670	1	14.4060
		400	2	16.5270	3	23.0380
	(-1.01,...,-1.01)	110	11	11.2970	38	20.5470
		130	2	2.8750	6	4.5460
		270	33	114.0290	34	214.3990
Epanechnikov	(0.25,...,0.25)	50	6	1.3740	2	0.4370
		180	5	17.1650	3	9.7560
		350	2	25.1610	3	32.9640
	(-0.45,...,-0.45)	45	4	0.6800	7	1.2180
		120	2	2.5300	4	4.9830
		220	9	38.6710	5	22.0620
	(-0.75,...,-0.75)	30	5	0.3910	7	0.5470
		70	16	6.6420	2	0.8280
		100	10	8.5350	7	5.9990
		120	3	3.9400	19	23.8080
		300	2	14.9650	4	32.0633
	(0.005,...,0.005)	15	29	0.6580	35	0.8620
		40	5	0.9720	6	1.0430
		220	4	17.2210	5	21.4220

6 Conclusion

We have presented a new hybrid conjugate gradient algorithm. The proposed method possesses a good descent search direction at each iteration and this is independent of the line search. The global convergence properties of the proposed method have been established under strong Wolfe line search conditions. We present the computational evidence that the performance of our method EHD is better than to some well-known conjugate gradient methods. The practical applicability of our method is also explored in nonparametric estimation of the mode function.

References

[1] 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) 55-70.

[2] N. Andrei, Another hybrid conjugate gradient algorithm for unconstrained optimization, Numer. Algorithms, 47 (2008) 143–156.

[3] N. Andrei, An unconstrained optimization test functions, Advanced Modeling and Optimization, 10 (2008) 147-161.

[4] N. Andrei, New hybrid conjugate gradient algorithms for unconstrained optimization, Encyclopedia of Optimization, (2009) 2560–2571.

[5] A. M. Awwal, I. M. Sulaiman, M. Malik, M. Mamat, P. Kumam, 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.

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

[7] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, X. Yin and Y. Yuan, Convergence properties of nonlinear conjugate gradient methods, SIAM J. Optim., 10 (1999) 348–358.

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

[9] Y. H. Dai and Y. Yuan, An efficient hybrid conjugate gradient method for unconstrained optimization, Ann. Oper. Res., 103 (2001) 33–47.

[10] E. D. Dolan and J. J. Morè, Benchmarking optimization software with performance profiles, Math. Programming, 91 (2002) 201-213.

[11] W. F. Eddy, Optimum kernel estimators of the mode, Ann. Statist., 8 (1980) 870-882.

[12] A. Fita, W. Geremew and L. Lemecha, A DC optimization approach for constrained clustering with l_1 norm, Palestine Journal of Mathematics, 11 (2022) 337–347.

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

[14] R. Fletcher, Practical Methods of Optimization, vol. 1: Unconstrained Optimization, John Wiley and Sons, New York, 1987.

[15] J. C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization, SIAM J. Optim., 2 (1992) 21-42.

[16] W. W. Hager and H. Zhang, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent., ACM Trans. Math. Software, 32 (2006) 113–137.

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

[18] Y. F. Hu and C. Storey, Global convergence result for conjugate gradient methods, J. Optim. Theory Appl., 71 (1991) 399-405.

[19] M. B. Huban and M. Gürdal, On deferred generalzed statistical convrergence of complex unertain triple sequences, Palestine Journal of Mathematics, 11 (2022) 255–266.

[20] B. Jouilik, J. Daoudi, C. Tajani and J. Abouchabaka, Optimization approach based on genetic algorithm for a robin coefficient problem, Palestine Journal of Mathematics,11 (2022) 708–718.

[21] V. D. Konakov, On the asymptotic normality of the mode of multidimensional distributions, Theory Probab. Appl., 19 (1974) 794 799.

[22] J. K. Liu and S. J. Li, New hybrid conjugate gradient method for unconstrained optimization, Appl. Math. Comput., 245 (2014) 36-43.

[23] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, Theory JOTA, 69 (1991) 129-137.

[24] G. Ma, H. Lin, W. Jin, D. Han, Two modified conjugate gradient methods for unconstrained optimization with applications in image restoration problems, Journal of Applied Mathematics and Computing, (2022) 1-26.

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

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

[27] E. Parzen, On estimating probability density function and mode, Ann. Math. Statist., 33 (1962) 1065-1076.

[28] E. Polak and G. Ribière, Note sur la convergence de directions conjuguée, Rev. Francaise Informat Recherche Operationelle, 3e Année 16 (1969) 35-43.

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

[30] T. W. Sager, An iterative method for estimating a multivariate mode and isopleth, J. Amer. Statist. Assoc., 74 (1975) 329-339.

[31] M. Samanta, Nonparametric estimation of the mode of a multivariate density, South Africain Statist. J., 7 (1973) 109-117.

[32] D. Touati-Ahmed and C. Storey, Efficient hybrid conjugate gradient techniques, J. Optim. Theory Appl., 64 (1990) 379-397.

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

[34] L. Zhang, An improved Wei–Yao–Liu nonlinear conjugate gradient method for optimization computation, Applied Mathematics and Computation, 6 (2009) 2269-2274.

Author information

L. Oumertem, Laboratory Informatics and Mathematics (LIM)
Department of Mathematics, Faculty of Science and Technology. University of Mohamed Cherif Messaadia, Souk Ahras, 41000, Algeria.
E-mail: l.oumertem@univ-soukahras.dz

Y. Chaib, Laboratory Informatics and Mathematics (LIM)
Department of Mathematics, Faculty of Science and Technology. University of Mohamed Cherif Messaadia, Souk Ahras, 41000, Algeria.
E-mail: y.chaib@univ-soukahras.dz

Received: 2023-04-25.
Accepted: 2023-12-27.