EUROPEAN JOURNAL OF PURE AND APPLIED MATHEMATICS
Vol. 16, No. 2, 2023, 1059-1067
ISSN 1307-5543 – ejpam.com
Published by New York Business Global

A hybridization of the Hestenes-Stiefel and Dai-Yuan Conjugate Gradient Methods

Yoksal A. Laylani[1], Hisham M. Khudhur[2,*], Edrees M. Nori[2], Khalil K. Abbo[3]

[1] *Department of Mathematics, College of Sciences, University of Kirkuk, Kirkuk, Iraq*
[2] *Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Mosul, Iraq*
[3] *Department of Mathematics, College of Basic Education, University of Telafer, Tall'Afar, Iraq*

Abstract. The paper in discusses conjugate gradient methods, which are often used for unconstrained optimization and are the subject of this discussion. In the process of studying and implementing conjugate gradient algorithms, it is standard practice to assume that the descent condition is true. Despite the fact that this sort of approach very seldom results in search routes that slope in a downward direction, this assumption is made routinely. As a result of this research, we propose a revised method known as the improved hybrid conjugate gradient technique. This method is a convex combination of the Dai-Yuan and Hestenes-Stiefel methodologies. The descending property and global convergence are both exhibited by the Wolfe line search. The numerical data demonstrates that the strategy that was presented is an efficient one.

2020 Mathematics Subject Classifications: 90C30, 65K05, 49M37

Key Words and Phrases: Optimization, Conjugate gradient, Numerical results

---

1. Introduction

The following unconstrained optimization problems are dealt with in this paper:

$$minf(x), x \in R^n \tag{1}$$

Where $f : R^n \to R$ is a continuously differentiable function that is bounded from below. From an initial guess, a non linear conjugate gradient (CG) algorithm generates a sequence of points $\{x_k\}$, according to the formula for recurrence [6, 7]

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

---

*Corresponding author.
DOI: https://doi.org/10.29020/nybg.ejpam.v16i2.4746

Email addresses: yuksel@uokirkuk.edu.iq (Y. A. Laylani), hisham892020@uomosul.edu.iq (H. M. Khudhur), edreesnori@uomosul.edu.iq (E. M. Nori), dr.khalil-khuder@uotelafer.edu.iq (K. K. Abbo)

where $d_k \in R^n$ search direction and $\alpha_k \in R$ is a step length that is normally obtained using the Wolfe method. [16, 19] and, the standard Wofle condition (SWC)

$$f\left(x_k + \alpha_k d_k\right) - f\left(x_k\right) \leq \rho \alpha_k g_k^T d_k \tag{3}$$

$$g_{k+1}^T d_k \geq \sigma g_k^T d_k \tag{4}$$

with $0 < \rho < \sigma < 1$, or by strong Wolfe condition(STWC)

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \rho \alpha_k g_k^T d_k \tag{5}$$

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

and compute the search direction from the following equation

$$d_{k+1} = -g_{k+1} + \beta_k s_k, d_0 = -g_0 \tag{7}$$

Here β_k the conjugate gradient parameter which is a scalar known as $g_k = \nabla f(x_k)$ and $s_k = x_{k+1} - x_k$. Different conjugate gradient algorithms corresponding to various parameter choices β_k see [3, 11, 13, 17, 18, 20, 23]. As a consequence, every conjugate gradient algorithm formula definition would be the same is crucial β_k The formulas are well-known. β_k are as follows;

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

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

FR denotes Fletcher and Reeves [10] , HS denotes Hestenes and Steifel [13], PR denotes Polak and Ribiere [19] , DY denotes Dai and Yuan [8] and LS denotes Liu and Storey. It's worth noting that these formulas for β_k they are similar when the objective function is a strictly convex quadratic function, and α_k is an exact one-dimensional minimizer [9]. The following is the outline of the document. Section 2 introduces and demonstrates our proposed method (HKYE) for generating descent directions. Its convergence analysis is seen in section 3. In section 4, you'll find some numerical experiments.

2. Hybrid conjugate gradient method

The traditional numerator conjugate gradient methods $\|g_{k+1}\|^2$ in relation to the update parameter β_k (FR, DY, CD) Although they have good convergence properties, their functional performance is limited. While the methods with numerator $g_{k+1}^T y_k$such as (PR, HS, LS)often have better computation performance, but they may not generally be convergent. In this section we try to introduce a new hybrid conjugate gradient method as follows: Consider the following search direction defined by

$$d_{k+1} = -\, g_{k+1} + [(1-\theta_k)\beta_k^{HS} + \theta \beta_k^{DY}] s_k \tag{8}$$

By using the conjugacy condition (Dai-Liao)

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k$$

and let

$$t = \frac{(s_k^T g_{k+1})^2}{(y_k^T s_k)^2}$$

We get the equation (9)

$$\theta_k = \frac{(s_k^T g_{k+1})^3}{(y_k^T s_k)^2} \tag{9}$$

Note that $\theta_k < 1$, since $s_k^T g_{k+1} < y_k^T s_k$, but it may be $\theta_k < 0$, to avoid this situation we take $\theta = 0$. Therefor, equation (8) becomes

$$d_{k+1} = -\, g_{k+1} + [\beta_k^{HS}] s_k \tag{10}$$

We call the method defined by equation (8)(HKYE) method.

2.1. New Algorithms

Step 1. Given $x_1 \in R^n, \varepsilon > 0, \quad d_1 = -g_1$.

Step 2. If $\|g_k\|^2 < \varepsilon$ then stop. else go to Step3.

Step 3. Compute an $\alpha_k > 0$satisfying (5)and (6)

Step 4. Let $x_{k+1} = x_k + \alpha_k d_k$. If $\|g_{k+1}\|^2 < \varepsilon$ then stop, else go to Step5

Step 5 Compute $\theta$and β_k. If $\theta > 1$or $\theta < 0$then compute $d_{k+1} = -g_{k+1} + \beta_k^{HS} s_k$ else $d_{k+1} = -g_{k+1} + \left\{\left(1 - \theta\beta_k^{HS} + \theta\beta_k^{DY}\right\} s_k\right.$

2.2. Descent property and global convergence analysis

Next we will show that our CG method (8) satisfies the descent property and global converges.

Theorem 1. *Let $\{\theta_k\}$ and $\{d_{k+1}\}$ be the sequences generated by Eq.* (8) *and* (9) *with strong wolfe conditions and assume that $g_{k+1}^T g_k \geq 0$ and $\sigma \leq \frac{1}{2}$. Then the search directions d_{k+1} satisfies sufficient descent condition $g_{k+1}^T d_k \leq -c\|g_k\|$ Where c is constant.*

Proof. The prove is by in induction that is if $k = 0$ then $d_1 = -g_1$ and $g_1^T d_1 = -\|g_1\|$ Assume that $g_k^T d_k \leq -c\|g_k\|$ to prove for $k = k+1$ we have $y_k^T s_k = g_{k+1}^T s_k - g_k^T s_k \geq (\sigma - 1)\, g_k^T d_k > 0$ and $\theta_k = \frac{(s_k^T g_{k+1})^3}{(y_k^T s_k)^2}$<1, if $\theta_k < 0$. We set $\theta_k = 0$ therefor we assume $0 < \theta_k < 1$. Then

$$d_{k+1} = -g_{k+1} + \left[(1-\theta_k)\frac{y_k^T g_{k+1}}{y_k^T s_k} - \theta_k \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k}\right] s_k$$

Multiply both sides by g_{k+1}^T

$$g_{k+1}^T d_{k+1} = -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{y_k^T g_{k+1}}{y_k^T s_k} - \theta_k \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} \right] \ g_{k+1}^T s_k$$

$$\leq -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{y_k^T g_{k+1} g_{k+1}^T s_k}{(\sigma - 1)\, g_k^T d_k} - \frac{\left(g_{k+1}^T s_k\right)^4}{(y_k^T s_k * y_k^T s_k)} \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} \right]$$

Since the last term is positive

$$\therefore g_{k+1}^T d_{k+1} \leq -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{\left|y_k^T g_{k+1}\right| \left|g_{k+1}^T s_k\right|}{(\sigma - 1)\, g_k^T d_k} \right] \leq -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{\left|y_k^T g_{k+1}\right| (-\sigma g_k^T d_k)}{(\sigma - 1)\, g_k^T d_k} \right]$$

$$\therefore g_{k+1}^T d_{k+1} \leq -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{\left|y_k^T g_{k+1}\right| (-\sigma)}{(\sigma - 1)} \right] = -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{\acute{o} \left|y_k^T g_{k+1}\right|}{(1-\sigma)} \right]$$

Since $g_{k+1}^T g_k \geq 0$ and $\left|y_k^T g_{k+1}\right| = \left|g_{k+1}^T \left(|g_{k+1} - g_k|\right)\right| \leq ||g_{k+1}||^2$

$$\therefore g_{k+1}^T d_{k+1} \leq -g_{k+1}^T g_{k+1} + \left[(1-\theta_k) \frac{\sigma g_{k+1}^T g_{k+1}}{(1-\sigma)} \right] = \left(-1 + \left[\frac{\sigma (1-\theta_k)}{(1-\sigma)} \right] \right) g_{k+1}^T g_{k+1}$$

$d_{k+1}^T g_{k+1} \leq -c \left\|g_{k+1}\right\|^2$ where $c = (1 - \frac{\sigma(1-\theta)}{(1-\sigma)})$, the prove is complete.

The following assumptions are used to demonstrate the proposed algorithm's global convergence.

Assumption 1. *1. In the level set, the objective function $f(x)$ is bounded below*

$$\Omega = \{x \in R^n : f(x) \leq f(x_1)\};$$

2. In some parts of town N ofΩ, f There exists a constant that is constantly differentiable and has a Lipschitz continuous gradient $L > 0$ such that:

$$\|g(x) - g(y)\| \leq L \|x - y\|, \forall \mathrm{x}, \mathrm{y} \in \mathrm{N} \tag{11}$$

It's worth noting that these assumptions suggest the existence of a constant Γ, implying that $\|g_k\| \leq \Gamma$, for any x in the level set Ω, and $\|x\| \leq B, \forall x \in \Omega$

Lemma 1. *[1, 2, 12, 14, 15] Assume that $d_k$is a descent direction and g_k satisfies the Lipschitz condition$\|g(x) - g(y)\| \leq L\|x - y\|$for all $x$on the linking section of a line x_k and x_{k+1} L is the constant in this equation. If Wolfe's condition is met, the line quest is successful then*

$$\alpha_k \geq \frac{1-\sigma}{L} \frac{\left|g_k^T d_k\right|}{\|d_k\|^2}$$

Lemma 2. *[4, 5, 8, 15, 21, 22] Assume the Assumption is right. The descent condition is satisfied if the conjugate gradient method is used.*

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty \tag{12}$$

Theorem 2. *Assume the Assumption 2.3. is true. Let $\{g_k\}$and $\{d_k\}$ be the sequence generated by the algorithm (HKYE) when subjected to a strong Wolfe line search condition. Then* $\liminf_{k\to\infty} \|g_k\| = 0$

Proof. It follows from Assumption 2.3. that there is a positive constant Γ >0 Such that $\|g(x)\| \leq \Gamma$ for all x $\in \Omega$, and Theorem 2.1. $\alpha_k \geq \frac{(1-\sigma)}{L}\frac{(g_k^T d_k)^2}{\|d_k\|^2} = \lambda$ If the theorem is incorrect, then a positive constant must exist >0 such that is true for all k sufficiently large

$$\|g_k\| \geq \mu \tag{13}$$

By the Eq(6)and Theorem 2.1., it results that

$$y_k^T s_k = g_{k+1}^T s_k - g_k^T s_k \geq -(1-\sigma) g_k^T s_k \geq c(1-\sigma)\, g_k{}^2 \tag{14}$$

Lipschitz continuity of the gradient, gives

$$y_k = g_{i+1} - g_k \leq L\ x_{k+1} - x_k \leq LR \tag{15}$$

Where R =max$\{\|x-y\|\,; x, y \in \Omega\ \}$ is diameter of the level. Now

$$\begin{aligned} |\beta_k| &= \left|(1-\theta_k)\beta_k^{HS} + \theta\beta_k^{DY}\right| \\ &\leq \left|\beta_k^{HS}\right| + \left|\beta_k^{DY}\right| \\ &= \frac{\left|\alpha_k g_k^T y_k\right|}{\left|y_k^T s_k\right|} + \frac{\left|\alpha_k g_{k+1}^T g_{k+1}\right|}{\left|y_k^T s_k\right|} \\ &\leq \frac{\alpha_k \|g_{k+1}\|\,\|y_k\|}{c\,\|g_k\|} + \frac{\alpha_k \|g_{k+1}\|}{c(1-\sigma)\,\|g_k\|} \\ &\leq \frac{\alpha_k \Gamma L R}{c(1-\sigma)} + \frac{\alpha_k \Gamma^2}{c(1-\sigma)\mu^2} = V \end{aligned}$$

Therefor

$$\begin{aligned} \|d_{k+1}\| &= \|g_{k+1}\| + |\beta_k|\,\|x_{k+1} - x_k\| \\ &= \|g_{k+1}\| + \tfrac{|\beta_k|\|s_k\|}{\alpha_k} \leq \Gamma + \tfrac{VR}{\mu} \equiv Q \end{aligned} \tag{16}$$

Which gives

$$\sum_{k\geq 0} \frac{1}{\|d_k\|} = \infty \tag{17}$$

Theorem 2.1. and Eq(13) and the Zoutendijk condition, on the other hand, result from adequate descent

$$c^2\Gamma^4 \sum_{k\geq 0} \frac{1}{\|d_k\|} \leq \sum_{k\geq 0} \frac{c^2 \|g_k\|^4}{\|d_k\|^2} \leq \sum_{k\geq 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} = \infty$$

As a result of the inconsistency with (17), (13) does not hold, and as a result

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

The prove is complete.

3. Comparisons and numerical results

The computational output of a Fortran implementation of the latest suggested (HKYE) algorithm on a set of 400 unconstrained optimization test problems is presented in this section. The unconstrained optimization problems are the test problems in [6]. We chose 40 test problems that were either extended or generalized. Each problem is evaluated ten times, with the number of variables increasingly increasing n=100,200,. . . ,1000. The code for Fortran from We compare our results to those of the HS and DY algorithms, as well as the Dolan and More performance profiles [23]. The strong Wolfe line search conditions are used in all algorithms with$\rho = 0.0001$and$\sigma = 0.4$. The same criteria for stopping $\|g_k\|_2 < 10^{-6}$is used. Algorithm comparisons are presented in the following context. Let $f_i^{ALG1}$and f_i^{AIG2} to be the best value found by ALG1 andALG2 for problem i=1,..,400, respectively. If the performance of ALG1 was better than the performance of ALG2 in the specific problem I we can assume that the performance of ALG1 was better than the performance of ALG2.

$$\left|f_i^{ALG1} - f_i^{AIG2}\right| < 10^{-3}$$

The number of iterations (#ite), function-gradient evaluations (#fg), or CPU time corresponding to ALG1 was less than the number of (#ite), (#fg), or CPU time corresponding to ALG2, respectively. To compare the performance of (HKYE) to that of HS and DY, we consider the number of iterations (#ite) in fig.1. Figure 2 shows how these algorithms perform when evaluating the number of function gradients (#fg), while Figure 3 shows how they perform when evaluating CPU time. We can see that (HKYE) is the best performer. These codes vary in their search path since they use the same strong Wolfe line search and stopping criteria. As a result, it appears that (HKYE) generates the best search direction among these methods.

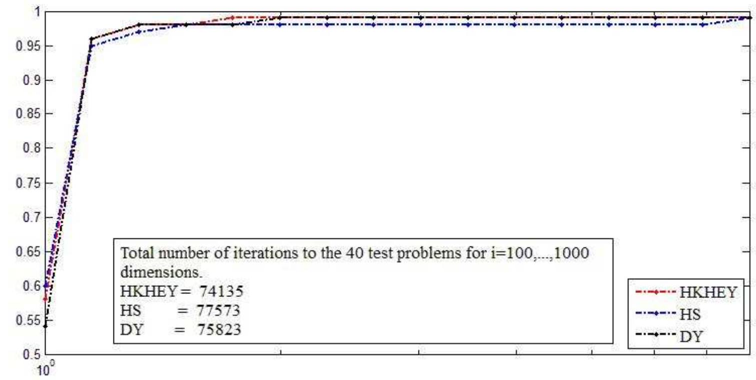

Figure 1: Iteration-based efficiency

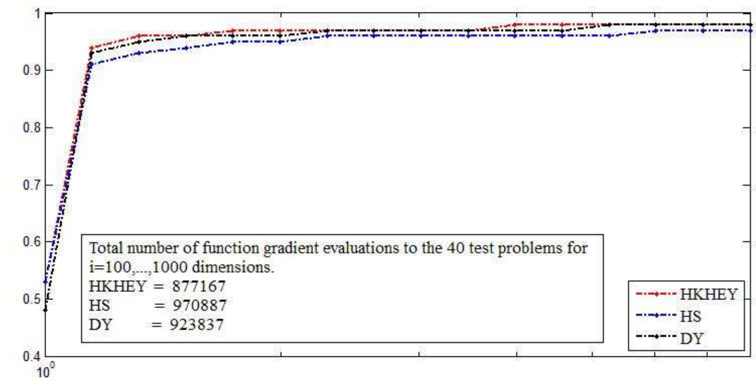

Figure 2: Performance by function

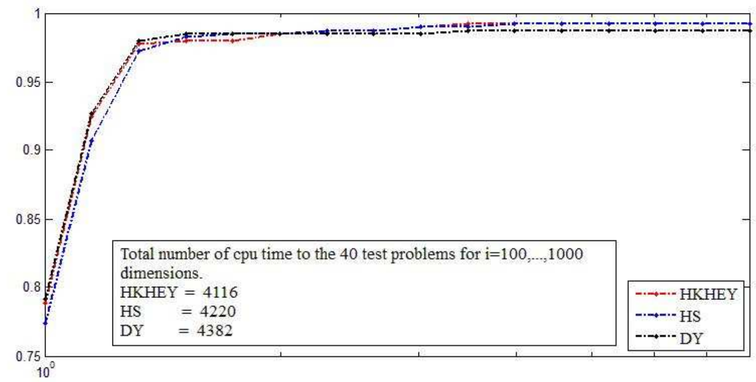

Figure 3: Time-Based Performance

4. Conclusion

We presented in this research a new type of conjugate gradient technique for solving unconstrained optimization problems, and the proposed algorithm has shown high efficiency in solving these problems with the least number of iterations and higher accuracy in reaching the approximate solution of the function.

References

[1] K K Abbo and H M Khudhur. New a hybrid conjugate gradient fletcher-reeves and polak-ribiere algorithm for unconstrained optimization. *Tikrit Journal of Pure Science*, 21:124–129, 2016.

[2] K K Abbo and H M Khudhur. New a hybrid hestenes-stiefel and dai-yuan conjugate gradient algorithms for unconstrained optimization. *Tikrit Journal of Pure Science*, 21:118–123, 2016.

[3] K K Abbo, Y A Laylani, and H M Khudhur. Proposed new scaled conjugate gradient algorithm for unconstrained optimization. *Int. J. Enhanc. Res. Sci. Technol. Eng*, 5:269–276, 2016.

[4] K K Abbo, Y A Laylani, and H M Khudhur. A new spectral conjugate gradient algorithm for unconstrained optimization. *Int. J. Math. Comput. Appl. Res*, 8:1–9, 2018.

[5] Z M Abdullah, M Hameed, H M Khudhur, and M A Khaleel. Modified new conjugate gradient method for unconstrained optimization. *Tikrit Journal of Pure Science*, 24:86–90, 2019.

[6] N Andrei. An unconstrained optimization test functions collection. *Adv. Model. Optim*, 10:147–161, 2008.

[7] Y Dai and L Liao. New conjugacy conditions and related non-linear cg methods' appl. *Math optim.,(43). Spring-verlag, New York*, 2001.

[8] Y H Dai and Y Yuan. A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on optimization*, 10:177–182, 1999.

[9] R Fletcher. *Practical methods of optimization.* John Wiley & Sons, 2013.

[10] R Fletcher and C M Reeves. Function minimization by conjugate gradients. *The computer journal*, 7:149–154, 1964.

[11] J C Gilbert and J Nocedal. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on optimization*, 2:21–42, 1992.

[12] W Hager and H Zhang. A survey of nonlinear conjugate gradient methods. *Pacific journal of Optimization*, 2:35–58, 2006.

[13] M R Hestenes and E Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of research of the National Bureau of Standards*, 49:409–436, 1952.

[14] H M Khudhur and K K Abbo. A new conjugate gradient method for learning fuzzy neural networks. *Journal of Multidisciplinary Modeling and Optimization*, 3:57–69, 2020.

[15] H M Khudhur and K K Abbo. A new type of conjugate gradient technique for solving fuzzy nonlinear algebraic equations. In *Journal of Physics: Conference Series*, volume 1879, page 022111, 2021.

[16] Y A Laylani, B A Hassan, and H M Khudhur. A new class of optimization methods based on coefficient conjugate gradient. *European Journal of Pure and Applied Mathematics*, 15:1908–1916, 2022.

[17] J Liu and S Wang. Modified nonlinear conjugate gradient method with sufficient descent condition for unconstrained optimization. *Journal of Inequalities and Applications*, 2011:1–12, 2011.

[18] J Nocedal and S J Wright. *Numerical optimization*. 2006.

[19] E Polak and G Ribiere. Note sur la convergence de méthodes de directions conjuguées. *Revue française d'informatique et de recherche opérationnelle. Série rouge*, 3:35–43, 1969.

[20] M Powell. A survey of numerical methods for unconstrained optimization. *SIAM review*, 12:79–97, 1970.

[21] A Sahiner, I A M Abdulhamid, and S A Ibrahem. A new filled function method with two parameters in a directional search. *Journal of Multidisciplinary Modeling and Optimization*, 2:34–42, 2019.

[22] A Sahiner and S A Ibrahem. A new global optimization technique by auxiliary function method in a directional search. *Optimization Letters*, 13:309–323, 2019.

[23] P Wolfe. Convergence conditions for ascent methods. ii: Some corrections. *SIAM review*, 13:185–188, 1971.