Using a new coefficient conjugate gradient method for solving unconstrained optimization problems

Ranen M. Sulaiman, Basim A. Hassan
Department of Mathematics, College of Computers Sciences and Mathematics, University of Mosul, Mosul, Iraq

ABSTRACT

The conjugate gradient technique is a numerical solution strategy for finding minimization in mathematics. We present a simple, straightforward, efficient, and resilient conjugate gradient technique in this study. To address the convergence difficulty and descent property, the new technique is built on the quadratic model. Under some assumptions, the new improved approach meets the convergence characteristics and the adequate descent criterion. The suggested unique strategy is substantially more efficient than the classic FR method, according to our numerical analysis. The number of function evaluations, iterations and restarts are all included in the numerical results. The computational efficiency of the proposed approach is proved by comparative results.

Keywords:
Conjugate gradient
Descent property
Global convergence
Numerical results
Optimization

1. INTRODUCTION

Use the following formula to get the minimum of a continuously differentiable function:

\[ \text{Min } f(x), x \in \mathbb{R}^n \]  

(1)

the iterative methods we use are iterative approaches of the following form:

\[ x_{k+1} = x_k + \lambda_k d_k \]

(2)

in (2) shows that various stepsizes \( \lambda_k \) and directions \( d_k \) result in distinct approaches, as shown in [1]. For example, in the quadratic case, \( \lambda_k \) is an accurate step size as (3):

\[ \lambda_k = -\frac{g_k^T d_k}{d_k^T Q d_k} \]

(3)

for further information, see [2]. After that, the step length \( \lambda_k \) is selected to meet the Wolfe conditions, which are as (4) and (5):

\[ f_{k+1} \leq f_k + \delta \lambda_k g_k^T d_k \]

(4)
\[ d_{k+1}^{\sigma} \geq \sigma d_{k}^{\sigma} g_k \]  

where \( 0 < \delta < \sigma < 1 \), see [3]. The search directions in conjugate gradient algorithms can be specified recursively:

\[ d_{k+1} = -g_{k+1} + \beta_k s_k \]  

where \( \beta_k \) is selected in a way that \( d_k \) and \( d_{k+1} \) must satisfy the conjugacy property. To compute the scalar \( \beta_k \), a number of formulae have been presented. Fletcher and Reeves (FR) [4] and Dai and Yuan (DY) [5] are two well-known formulae. They’re provided by:

\[ \beta_{FK} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}, \beta_{DR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} \]  

The other nonlinear conjugate gradient techniques, for example, are the subject of a lot of study in this area (e.g. see [6]-[9]). Many writers have researched the nonlinear conjugate gradient technique in recent years, particularly from the perspective of global convergence. The nonlinear conjugate gradient technique was often studied independently because its characteristics might vary greatly depending on \( \beta_k \) (see Powell [10]).

We’d want to close the new search direction to the quasi-Newton direction later because of the theoretical usefulness of quasi-Newton approaches. In this situation, we’re looking for a parameter that will allow us to:

\[ -Q_{k+1}^{-1} g_{k+1} = -g_{k+1} + \beta_k s_k \]  

where \( Q_{k+1} \) is the Hessian matrix. (See [11], [12]) for a useful resource for research describing the most recent CG coefficients with notable results and numerous \( \beta_k \) adjustments. The approaches are efficient in reality, according to numerical findings, and the methods’ convergence guarantees are comparable to the classical variations. Our key contribution is a novel coefficient derivation based on the second-order Taylor’s series, which we utilized to build an inverse Hessian matrix for computing the search direction and ensuring global convergence.

### 2. OUR NEW COEFFICIENT CONJUGATE

A second order Taylor series is used to derive the new coefficient conjugate. Let we clarify:

\[ f(x) = F(x_{k+1}) + g_{k+1}^T (x - x_{k+1}) + \frac{1}{2} (x - x_{k+1})^T Q(x_{k+1})(x - x_{k+1}) \]  

the \( f(x) \) has the following gradient:

\[ g_{k+1} = g_k + Q(x_{k+1})s_k \]  

putting (3) in (9) and using ELS search, we get:

\[ s_k^T Q(x_{k+1})s_k = (F_{k+1} - F_k) + 3/2 \lambda s_k^T g_k \]  

the yielded matrix \( Q(x_{k+1}) \) can be as:

\[ Q(x_{k+1}) = \frac{(F_{k+1} - F_k) + 3/2 \lambda s_k^T g_k}{s_k^T s_k} I_{n \times n} \]  

putting \( Q(x_{k+1}) \) in (8) we get:

\[ \beta_k = \left( 1 - \frac{s_k^T s_k}{(F_{k+1} - F_k) + 3/2 \lambda s_k^T g_k} \right) \frac{g_{k+1}^T y_k}{y_k^T y_k} \]  

We will do some algebra manipulations on (14), in order to achieve an ideal direction of descent:
\[ \beta_k = \frac{1}{\|y_k\|^2} \left( y_k - \omega \frac{\|y_k\|^2}{\|y_k\|^2} g_k \right)^T g_{k+1} \]  

(14)

where:

\[ \omega = \frac{\frac{y_k^T y_k}{\|y_k\|^2} \frac{y_k^T y_k}{\|y_k\|^2}}{\|y_k\|^2} \frac{y_k^T y_k}{\|y_k\|^2} + \frac{\lambda_k}{2} \frac{y_k^T y_k}{\|y_k\|^2} \]  

(15)

This is the formula that will be utilized to do the convergence analysis. The application of optimization theory and methods to new formulations is a vast field of applied mathematics. The method described by is denoted by New (13). As follows, we suggest a novel conjugate gradient method.

**New Algorithm:**

1) Give \( x_1 \in \mathbb{R}^n \). Set \( k = 1 \) and \( d_1 = -g_1 \). If \( \|g_1\| \leq 10^{-6} \), then stop.
2) Evaluate \( \lambda_k > 0 \) satisfying a (4-5).
3) Let \( x_{k+1} = x_k + \lambda_k d_k \). If \( \|g_{k+1}\| \leq 10^{-6} \), then come to a halt.
4) Evaluate \( \beta_k \) by the formulae (14) and \( d_{k+1} \) by (6).
5) Let \( k = k + 1 \) and continue with step 2.

**Theorem (2.2):** Consider the CG technique (2), (4), (5), and the descent direction \( d_{k+1} \) provided by (6) with (14) is adequate.

**Proof:** Since \( d_0 = -g_0 \), we get \( g_0^T d_0 = -\|g_0\|^2 \leq 0 \). Suppose that \( g_k^T d_k < 0 \) for all \( k \in n \). To finish the proof, we must prove that the theorem holds for all \( k + 1 \). Because \( Q(x_{k+1}) \) is the quadratic model's search direction matrix, we may write it like this:

\[ Q(x_{k+1}) = \frac{(f_{k+1} - f_k) + 3/2 \lambda_k g_k^T g_{k+1}}{y_k^T y_k} l_{nxn} = \omega l_{nxn} \]  

(16)

Now, we have to prove that \( \omega > 0 \). Using Wolfe's condition for determining the value of \( \omega \), we have:

\[ \omega = \frac{\lambda_k \delta_k d_k^T g_k + \lambda_k \delta_k g_k^T g_k}{\eta_k} \]  

(17)

Since \( d_k^T g_k < -c g_k^T g_k \), then, using the equation above, we obtain:

\[ \omega = \frac{3/2 \lambda_k g_k^T g_k + \lambda_k \delta_k g_k^T g_k}{\eta_k} \]  

(18)

Because the first half of (18) is greater than the second, we get:

\[ \omega > 0 \]  

(19)

We may describe the search directions of the new approach as follows using (6) and (14) and some algebraic manipulations:

\[ d_{k+1} = -Q_{k+1}^{-1} d_{k+1} = -\frac{g_k^T g_k}{(f_{k+1} - f_k) + 3/2 \lambda_k g_k^T g_k} g_{k+1} \]  

(20)

Multiplying (20) by \( g_{k+1} \), we have:

\[ d_{k+1}^T g_{k+1} = -\frac{g_k^T g_k}{(f_{k+1} - f_k) + 3/2 \lambda_k g_k^T g_k} \| g_{k+1} \|^2 = -\omega \| g_{k+1} \|^2 \]  

(21)

Since \( \omega > 0 \), from (17) we obtained:

\[ d_{k+1}^T g_{k+1} = -\omega^{-1} \| g_{k+1} \|^2 \leq -c \| g_{k+1} \|^2 \]  

(22)
3. CONVERGENCE ANALYSIS

To do this, establish the “global convergence” of new Algorithm is one of the most property of numerical algorithms, the assumptions must be made: The $\Omega = \{x \in R^n / f(x) \leq f(x_0)\}$ is a confined level set. The gradient of function $g$ is Lipschitz continuous in some neighborhood $\Lambda$ of $\Omega$, i.e., there exists a constant, $L > 0$ such that:

$$\|g(o) - g(\tau)\| \leq L\|o - \tau\|, \forall o, \tau \in \Lambda$$

for more details see [8]. We show why the Dai et al. [13] theorem is crucial for determining global convergence.

**Lemma (3.1):** Let $x_k$ be produced by (2), $d_k$ satisfy descent property and $\alpha_k$ be satisfy (4-5). If:

$$\sum_{k=0}^{\infty} \frac{1}{\|d_{k+1}\|^2} = \infty$$

then:

$$\lim \inf_{k \to \infty} \|g_{k+1}\| = 0$$

we began by stating our paper’s key theorem.

**Theorem (3.2):** Assume that $f(x)$ meets Assumptions 1 and 2. Let $\{x_k\}$ be the sequence that (6) generates (14). If Wolfe criteria (4) and (5) are satisfied by step size $\lambda_k$, then:

$$\lim \inf_{k \to \infty} \|g_k\| = 0$$

**Proof:** Using (14) as an example of $\beta_k$ in (6), we get:

$$\|d_{k+1}\| = \left\| -g_{k+1} + \beta_k^{\text{new}}d_k \right\| \leq \|g_{k+1}\| + \left\| \left( y_k - \omega \frac{\|y_k\|^2}{\|y_k\|^2} y_k \right) \right\| \|y_k\| \|d_k\|$$

and combining $\delta_k = \lambda_k d_k$ we get:

$$\|d_{k+1}\| \leq \|g_{k+1}\| + \left\| y_k \|g_{k+1}\| + \omega \|g_{k+1}\| \|y_k\| \right\| \|d_k\| \leq \left[ \frac{\lambda_k + \omega}{\lambda_k} \right] \|g_{k+1}\|$$

which results in:

$$\sum_{k=1}^{\infty} \frac{1}{\|d_k\|^2} \leq \left[ \frac{\lambda_k}{\lambda_k + 1} \right] \frac{1}{\|g_{k+1}\|} = \infty$$

we may deduce from Lemma 1 that $\lim \inf_{k \to \infty} \|g_k\| = 0$ is identical to $\lim \inf_{k \to \infty} \|g_k\| = 0$ for a uniformly convex function.

4. NUMERICAL RESULTS

On a series of unconstrained optimization test problems, this section shows the computing efficiency of a Fortran implementation of the novel CG technique and the FR-Algorithm. Readers who are interested can access the papers and references listed below ([14], [15]). The unconstrained concerns in [16] are the test problems. We investigated numerical experiments with 100 and 1000 variables for each test function for 15 large scale unconstrained optimization problems in extended or generalized form. Many papers have proposed this method for optimization problems [17]-[20]. As for the papers, it is concerned with the convergence feature [21]-[24]. As a termination condition, we employ the inequality $\|g_{k+1}\| \leq 10^{-6}$. The $\delta = 0.001$ and $\sigma = 0.9$ were used to evaluate both methods. The numerical findings are reported in Table 1. "The following are the definitions for each column: NI: the total number of iterations, NR: the total number of restart, NF: the total number of evaluation functions". Table 1 shows how many issues these algorithms have solved in terms of iterations (NI), restart (NR) and function evaluations (NF).
Using a new coefficient conjugate gradient method for solving unconstrained optimization problems

Table 1. The FR and new methods' numerical results

| P. No.                | FR algorithm | New algorithm |
|-----------------------|--------------|---------------|
|                       | n  | NI | NR | NF  | n  | NI | NR | NF  |
| Extended Rosenbrock   | 100| 47 | 18 | 93  | 43 | 23 | 89 |
| Extended Beale        | 100| 32 | 15 | 52  | 18 | 11 | 34 |
| Penalty               | 100| 22 | 10 | 42  | 19 | 11 | 35 |
| Extended PSC1         | 100| 15 | 9  | 31  | 11 | 7  | 23 |
| Extended Maratos      | 100| 8  | 6  | 17  | 8  | 6  | 17 |
| Extended Q. Penalty   | 100| 32 | 12 | 65  | 24 | 15 | 56 |
| Quadratic QF2         | 100| 130| 49 | 196 | 117| 36 | 184|
| ARWHEAD (CUTE)        | 100| 9  | 4  | 18  | 10 | 6  | 18 |
| NONDIA (CUTE)         | 100| 13 | 7  | 25  | 12 | 7  | 23 |
| Partial Quadratic     | 100| 15 | 7  | 29  | 12 | 6  | 25 |
| Broyden Triagonal     | 100| 30 | 10 | 49  | 28 | 6  | 49 |
| EDENSCH (CUTE)        | 100| 34 | 10 | 63  | 38 | 11 | 67 |
| EDENSCHNC (CUTE)      | 100| 19 | 11 | 35  | 19 | 11 | 32 |
| LIARWHD (CUTE)        | 100| 23 | 11 | 45  | 24 | 11 | 29 |
| DENSCHNA (CUTE)       | 100| 20 | 11 | 33  | 18 | 11 | 30 |
| DENSCHNC (CUTE)       | 100| 19 | 11 | 35  | 19 | 11 | 32 |
| Total                 |      | 2002| 818| 6527| 1569| 633| 3189|

Table 2. The Performance percentage for the new algorithm compared with FR method

|         | NI   | NR   | NF   |
|---------|------|------|------|
| FR      | 100 %| 100 %| 100 %|
| New     | 78.37%| 77.38%| 48.85%|

Figure 1. Performance measure based on the NI
5. CONCLUSION

On the basis of the quadratic model, we proposed a unique conjugate gradient technique. The proposed technique satisfies both the descent and convergence requirements. According to the numerical data, the new strategy outperformed the FR-method in terms of the number of iterations, restart, and evaluation functions. In addition, the new method surpasses the previous FR.

REFERENCES

[1] Y. H Dai and L. Z Liao, “New conjugacy conditions and related nonlinear conjugate gradient methods,” Applied Mathematics and Optimization, vol. 43, no. 1, pp. 87–101, 2001, doi: 10.1007/s002450010019.
[2] J. Nocedal and S. J. Wright, “Sequential quadratic programming,” in Numerical optimization. New York: Springer, 1999.
[3] Z. Dai and F. Wen, “Some improved sparse and stable portfolio optimization problems,” Finance Research Letters, vol. 27, pp. 46–52, 2018, doi: 10.1016/j.frl.2018.02.026.
[4] R. Fletcher and C. M. Reeves, “Function minimization by conjugate gradients,” The Computer Journal, vol. 7, no. 2, pp. 149–154, 1964, doi: 10.1093/comjnl/7.2.149.
[5] Y.-H. Dai and Y. Yuan, “A nonlinear conjugate gradient method with a strong global convergence property,” SIAM Journal on Optimization, vol. 10, no. 1, pp. 177–182, 1999, doi: 10.1137/S1052623497318992.
[6] R. F. RF, Practical Method of Optimization, Unconstrained Optimization, New York: Wiley, 1987.
[7] M. R. Hestenes and E. Stiefel, “Methods of Conjugate Gradients for Solving,” Journal of Research of the National Bureau of Standards, vol. 49, no. 6, pp. 409–436, 1952.
[8] M. R. Hajmohammadi, “Design and analysis of multi-scale annular fins attached to a pin fin,” International Journal of Refrigeration, vol. 88, pp. 16–23, 2018, doi: 10.1016/j.ijrefrig.2017.11.032.
[9] E. Polak and C. Ribiere, “Note sur la convergence de directions conjuguées,” ESAIM: Mathematical Modelling and Numerical Analysis - Modélisation Mathématique et Analyse, vol. 3, no. 1, pp. 35–43, 1969.
BIOGRAPHIES OF AUTHORS

Ranen M. Sulaiman is currently a Professor in Department of Mathematics, College of Computer Science and Mathematics, University of Mosul. He obtained his M.Sc degrees in Mathematics from the University of Mosul, in 2020 and 2021, respectively with specialization in optimization. To date, he has published more than 4 research paper in various international journals and conferences. He currently works on iterative methods. His research interest in applied mathematics, with a field of concentration of optimization include conjugate gradient, steepest descent methods, Broyden’s family and quasi-Newton methods. He can be contacted at email: ranen.sulyman@uomosul.edu.iq.

Basim A. Hassan is currently a Professor in Department of Mathematics, College of Computer Science and Mathematics, University of Mosul. He obtained his M.Sc and Ph.D degrees in Mathematics from the University of Mosul, in 2000 and 2010, respectively with specialization in optimization. To date, he has published more than 80 research paper in international journal and conferences. His research interest in applied mathematics, with a field of concentration of optimization include conjugate gradient, steepest descent methods, Broyden’s family and quasi-Newton methods with application in signal recovery and image restoration. He can be contacted at email: basimah@uomosul.edu.iq.