

Upscaling parameters for conjugate gradient method in unconstrained optimization

Basim A. Hassan *

Department of Mathematics

University of Mosul

College of Computer Science & Mathematics

Mosul

Iraq

Hawraz N. Jabbar [†]

Yoksal A. Laylani [§]

Department of Mathematics

University of Kirkuk

College of Sciences

Kirkuk

Iraq

Abstract

An effective iterative technique, conjugate gradient algorithm is based on the conjugate gradient parameters. By satisfying the conjugacy requirement, one may construct an effective parameters conjugate gradient using a second order Taylor series. After an analysis of the new algorithm's convergence characteristics, we provide some numerical findings that demonstrate the robustness and effectiveness of the improved methods.

Subject Classification: 90C30, 65K05, 49M37

Keywords: Optimization techniques, Conjugate-gradient, Numerical results.

1. Introduction

The majority of optimization techniques are iterative, [1]. We aim to resolve the following mathematical optimization issue for the smooth function

$$f : R^n \rightarrow R :$$

* E-mail: basimah@uomosul.edu.iq (Corresponding Author)

[†] E-mail: hawrazmath@uokirkuk.edu.iq

[§] E-mail: yuksel@uokirkuk.edu.iq

This research was conducted at University of Mosul.

$$\text{Min } f(x), \quad x \in \mathbb{R}^n \quad (1)$$

We shall examine many iterative techniques of the type:

$$x_{k+1} = x_k + \alpha_k d_k \quad (2)$$

where α_k is the step length and d_k is the search direction,[2]. Step length is determined by:

$$\alpha_k = \frac{-g_k^T d_k}{d_k^T Q d_k}. \quad (3)$$

Using the Wolfe conditions the step length α_k will be accepted:

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T d_k \quad (4)$$

$$d_k^T g(x_k + \alpha_k d_k) \geq \sigma d_k^T g_k \quad (5)$$

where $0 < \delta < \sigma < 1$, see [3,4]. Here, the following give the search direction:

$$d_{k+1} = -g_{k+1} + \beta_k d_k \quad (6)$$

where β_k is scalar that is calculated using a different formula depending on the conjugate gradient technique being employed. This approach was expanded by Fletcher and Reeves for use with broader functions[5]. The following is the formula for calculating β_k using Fletcher-Reeves:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}. \quad (7)$$

The stepsize and the CG parameter are the two main variables that have an impact on the numerical performance of the CG technique. Some famous classical choices of β_k can be found in [6-11]. Numerous CG approaches have been put forth recently that have both strong global convergence features and promising numerical performance. These methods are based on the conjugacy requirement of numerous CG parameters (see [12-15]). The conjugacy requirement was used in previous CG algorithms:

$$d_{k+1}^T Q d_k = 0. \quad (8)$$

that is based on the precise line search is crucial to the mathematical experiment and convergence analysis[16].

In this study, we provide novel conjugate gradient optimization techniques and analyze their convergence. Numerical tests reveal that our approaches can perform better than those already in use.

2. A New-Formulas For β_k

Now that we have derived the new formulae using the second order Taylor series, we get:

$$f_k = f_{k+1} - g_{k+1}^T s_k + \frac{1}{2} s_k^T Q s_k \quad (9)$$

Using exact line search in (9), we get:

$$s_k^T Q s_k = 2(f_k - f_{k+1}) \quad (10)$$

The conjugacy condition, we get:

$$d_{k+1}^T Q s_k = 0 \quad (11)$$

Applying (11) on (10) and after some algebra, we obtained:

$$\beta_k = \frac{\frac{2(f_k - f_{k+1})}{s_k^T y_k} g_{k+1}^T y_k}{d_k^T y_k} \quad (12)$$

When using exact line search, β_k is such that:

$$\beta_k^{BK1} = \frac{\frac{2(f_k - f_{k+1})}{s_k^T y_k} \|g_{k+1}\|^2}{d_k^T y_k} \quad (13)$$

and

$$\beta_k^{BK2} = \frac{\frac{2(f_k - f_{k+1})}{s_k^T g_k} \|g_{k+1}\|^2}{d_k^T g_k} \quad (14)$$

and

$$\beta_k^{BK3} = \frac{\frac{2(f_k - f_{k+1})}{\alpha_k g_k^T g_k} \|g_{k+1}\|^2}{g_k^T g_k}. \quad (15)$$

The ways that the aforementioned parameters produced are known as the BK1, BK2, BK3, respectively.

The suggested method's algorithm is then stated as follows:

- I. Given $x_0 \in R^n$. Let $d_0 = -g_0, k=0$.
- II. If $\|g_{k+1}\| \leq 10^{-6}$ yes, then stop; if not, continue.
- III. Compute α_k by using (4)-(5).
- IV. Let $x_{k+1} = x_k + \alpha_k d_k$, and compute β_k by (13-15).
- V. Compute the $d_{k+1} = -g_{k+1} + \beta_k d_k$.
- VI. Set $k = k + 1$ and continue with step 1.

3. Convergence analysis:

We assume that $f(x)$ meets the following conditions in what follows:

- i. $f(x)$ is bounded on the set $\Psi = \{x \in R^n : f(x) \leq f(x_0)\}$.
- ii. g is Lipschitz continuous, i.e. there exists a constant $L > 0$. That is:

$$\|g(z) - g(u)\| \leq L \|z - u\|, \quad \forall z, u \in R^n. \quad (17)$$

With a constant $\forall \geq 0$ such that $\|\nabla f(x)\| \leq \forall$, see [2, 13].

Lemma 1.1: *Assuming that the assumptions are correct, think about any recurrence expression (2) with search direction (3). The Zoutendijk criterion (16) is satisfied in that case [17].*

Theorem 3.1: *Let d_{k+1} be a sequence generated by new Algorithms. Then, it holds that:*

$$d_{k+1}^T g_{k+1} < 0 \text{ and } d_{k+1}^T g_{k+1} = \beta_k d_k^T g_k. \quad (18)$$

Proof: Using (3), we get:

$$d_{k+1}^T g_{k+1} = -g_{k+1}^T g_{k+1} + \beta_k d_k^T g_{k+1} \quad (19)$$

Equation (14) indicates that,

$$\|g_{k+1}\|^2 = \frac{\beta_k^{BK1} d_k^T y_k}{\frac{2(f_k - f_{k+1})}{s_k^T y_k}}. \quad (20)$$

Adding (20) to (19), we got:

$$d_{k+1}^T g_{k+1} = \beta_k^{BK1} [d_k^T g_{k+1} - d_k^T y_k] = \beta_k^{BK1} d_k^T g_k < 0 \quad (21)$$

The proof has been completed.

Theorem 3.2: *Assume that the direction d_{k+1} was produced using new algorithms. Then:*

$$\liminf_{k \rightarrow \infty} \|g_k\| = 0. \quad (22)$$

Proof: The proof is simple, similar to the Theorem 2 proof in [12][6].

4. Numerical results

This section includes the outcomes of novel techniques on a number of test problems. Calculations are done with double precision and the programs are written in Fortran. All exams are taken on computers. The 50 nonlinear conditions that a transducer is able to produce serve as the foundation for our experiments. These issues are covered in Andrei[18] and are tested on in

the CUTE exam. In all algorithms, the standard stop, s.t. $\|g_{k+1}\| \leq 10^{-6}$, is used. For assessing the performance of the algorithms, both the total number of function evaluations (NOF) and the total number of iterations are considered (NOI). The findings are shown in Tables 1 and 2. There are important papers in the field of optimization as [19-25].

Table 1
The numerical findings using the FR and New techniques with n=100.

P.No.	FR		BK1		BK2		BK3	
	NI	NF	NI	NF	NI	NF	NI	NF
Trigonometric	19	35	19	36	16	34	18	34
Extended-rosenbrok	47	93	44	90	43	88	34	72
Extended-White & .Holst	43	88	42	85	41	85	38	89
Extended-Beale	32	52	16	30	13	26	17	32
Extended-ridiagonal	32	64	9	19	10	21	15	31
Extended Three Expo Terms	15	25	10	15	14	20	17	27
Generalized Tridiagonal 2	37	67	37	59	43	68	45	64
Extended Himmelblau	12	25	10	19	10	19	10	19
Extended Powell	189	313	87	163	63	118	87	166
Quad. Diagonal Perturbed	124	231	47	85	45	81	36	101
Extended Wood	71	110	29	54	25	47	33	60
ARWHEAD	9	18	8	16	8	16	10	19
NONDIA	13	25	13	26	11	21	11	22
DIXMAANE	121	218	82	129	81	124	75	122
Partial Perturbed Quadratic	74	123	77	120	74	113	77	126
LIARWHD	23	45	19	34	20	38	20	36
ENGVAL1	34	57	24	46	22	42	24	47
DENSCHNA	20	33	11	20	10	19	13	23
DENSCHNC	49	80	19	33	13	24	18	31
Extended Block-Diagonal	122	156	11	21	11	20	12	23
Generalized quartic GQ1	11	24	7	16	7	16	8	18

Generalized quartic GQ2	112	147	34	56	35	53	39	61
STAIRCASE S1	671	1066	503	816	504	793	462	747
TRIDIA	398	605	292	468	322	496	345	548
Extended PSC1	15	31	8	17	8	17	8	17
Total	2293	3731	1005	2473	1449	2399	1472	2535

Table 2

The numerical findings using the FR and New techniques with n=1000.

	FR		BK1		BK2		BK3	
P.No.	NI	NF	NI	NF	NI	NF	NI	NF
Trigonometric	38	65	37	67	36	66	40	69
Extended-rosenbrok	78	171	37	83	40	89	42	92
Extended-White & .Holst	46	92	41	90	41	84	44	94
Extended-Beale	22	42	12	24	12	23	15	29
Extended-ridiagonal	77	129	13	27	13	26	18	36
Extended Three Expo Terms	fail	Fail	61	1447	19	32	10	17
Generalized Tridiagonal 2	73	115	52	84	49	79	56	89
Extended Himmelblau	14	29	11	21	10	19	10	19
Extended Powell	Fail	Fail	97	185	70	133	84	156
Quad. Diagonal Perturbed	445	711	162	288	149	260	186	340
Extended Wood	47	84	26	52	25	49	26	50
ARWHEAD	12	82	7	34	8	55	8	16
NONDIA	15	29	13	25	14	29	16	33
DIXMAANE	345	634	237	376	234	357	256	403
Partial Perturbed Quadratic	370	616	253	428	274	457	272	463
LIARWHD	27	55	21	44	22	48	18	39
ENGVAL1	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail
DENSCHNA	19	35	9	18	18	30	9	18
DENSCHNC	129	166	13	29	13	26	12	24
Extended Block-Diagonal	130	166	13	25	12	23	10	20
Generalized quartic GQ1	9	22	7	18	7	18	7	18

Generalized quartic GQ2	110	145	35	54	31	56	34	57
STAIRCASE S1	Fail	Fail	Fail	Fail	Fail	Fail	Fail	Fail
TRIDIA	1255	1991	Fail	fail	1279	1996	1026	1657
Extended PSC1	8	17	7	15	7	15	7	15
Total	2014	3405	792	1802	1015	1809	855	1924

According to this data, the BK1, BK2 and BK3 aroaches have a clear advantage over the FR technique since they save around (38-42)% in NOF but only (42-59)% in NOI. (See Table 3)

Table 3
The Performance-Percentage for the BK1,BK2 and BK3 algorithms-compared with FR method

	FR algorithm	BK1 (13)	BK2 (14)	BK3 (15)
NI	100 %	41.72%	58.02%	54.02%
NF	100 %	59.90%	58.96%	62.48%

5. Conclusions

In addition to the innovative modified conjugate gradient formula, we also introduced the conjugate gradient techniques BK1, BK2, and BK3. Wolfe line search settings allowed us to identify its worldwide convergence. Simulations have demonstrated that BK1, BK2, BK3 may decrease function evaluations and iterations. Compared to the traditional secant relation, which just uses gradient values, the modified conjugate gradient approach uses-both gradient and function values.

References

- [1] W. W.Hager and H. Zhang, A Survey of Nonlinear Conjugate Gradient Methods, *Pacific J. Optim.*, 2(1), 35-58 (2006), <http://www.ybook.co.jp/online2/ojo/vol2/p35>.
- [2] J. C. Gilbert and J. Nocedal, Global Convergence Properties of Conjugate Gradient Methods for Optimization, *SIAM J. Optim.*, 2(1), 21-42 (1992), <http://doi:10.1137/0802003>.

- [3] P. Wolfe, Convergence Conditions for Ascent Methods. II: Some Corrections, *SIAM Rev.*, 13(2), 185-188 (1971), <http://doi:10.1137/1013035>.
- [4] P. Wolfe, Convergence Conditions for Ascent Methods, *SIAM Rev.*, 11(2), 226-235 (1969) <http://doi:10.1137/1011036>.
- [5] R. Fletcher, Function minimization by conjugate gradients, *Comput. J.*, 7(2), 149-154 (1964), <http://doi:10.1093/comjnl/7.2.149>.
- [6] Y.H. Dai, Y. Yuan, A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property, *SIAM J. Optim.*, 10(1), 177-182 (1999), <http://doi:10.1137/S1052623497318992>.
- [7] J. W. Daniel, The Conjugate Gradient Method for Linear and Nonlinear Operator Equations, *SIAM J. Numer. Anal.*, 4(1), 10-26 (1967), <http://doi:10.1137/0704002>.
- [8] M. R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bur. Stand.* (1934), 49(6), 409-501 (1952), <http://doi:10.6028/jres.049.044>.
- [9] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, part 1: Theory, *J. Optim. Theory Al.*, 69(1), 129-137 (1991), <http://doi:10.1007/BF00940464>.
- [10] E. Polak and G. Ribiere, Note sur la convergence de méthodes de directions conjuguées, *Rev. française d'informatique Rech. opérationnelle. Série rouge*, 3(16), 35-43 (1969), <http://doi:10.1051/m2an/196903R100351>.
- [11] B. T. Polyak, The conjugate gradient method in extremal problems, *USSR Comput. Math. Math. Phys.*, 9(4), 94-112 (1969), [http://doi:10.1016/0041-5553\(69\)90035-4](http://doi:10.1016/0041-5553(69)90035-4).
- [12] B.A. Hassan, H. O. Dahawi, A.S. Younus, A new kind of parameter conjugate gradient for unconstrained optimization, *Indones. J. Electr. Eng. Comput. Sci.*, 17(1), 404-503 (2020), <http://doi:10.11591/ijeecs.v17.i1.404-411>.
- [13] B.A. Hassan, Z.M. Abdullah, S.A. Hussein, A new family of conjugate gradient methods to solve unconstrained optimization problems, *J. Inf. Optim. Sci.*, 43(4), 811-820 (2022), <http://doi:10.1080/02522667.2022.2103299>.
- [14] B.A. Hassan, K. Muangchoo, F. Alfarag, A. H. Ibrahim, A. Bala Abubakar, An improved quasi-Newton equation on the quasi-Newton methods for unconstrained optimizations, *Indones. J. Electr.*

- Eng. Comput. Sci.*, 22(2), 997 (2021), <http://doi:10.11591/ijeecs.v22.i2.997-1005>.
- [15] B.A. Hassan, Z.M. Abdullah, H.N. Jabbar, A descent extension of the Dai -Yuan conjugate gradient technique, *Indones. J. Electr. Eng. Comput. Sci.*, 16(2), 661-700 (2019), <http://doi:10.11591/ijeecs.v16.i2.661-668>.
- [16] J. Nocedal S.J. Wright. Quadratic Programming, in Numerical Optimization, Springer New York, 448-492. http://doi:10.1007/978-0-387-40065-5_16.
- [17] R.Madani . Computational Methods for Nonlinear Programming, in Building and Solving Mathematical Programming Models in Engineering and Science, Hoboken, NJ, USA : John Wiley & Sons, Inc., 2011, 235-282. <http://doi:10.1002/9780471225294.ch9>.
- [18] N. Andrei, Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization, *Bull. Malaysian Math. Sci. Soc.*, 34(2), 319-330 (2011).
- [19] W. Gao , M. R. Farahani, The hyper-zagreb index for an infinite family of nanostar dendrimer, *Journal of Discrete Mathematical Sciences and Cryptography*, 20(2), 515-523 (2017), <http://doi:10.1080/09720529.2016.1220088>.
- [20] W. Gao, L. Shi, M. R. Farahani, Szeged Related Indices of TUAC6[p,q], *Journal of Discrete Mathematical Sciences and Cryptography*. 20(2), 553-563 (2017), <http://doi:10.1080/09720529.2016.1228312>.
- [21] W. Gao , M. R. Farahani, The Zagreb topological indices for a type of Benzenoid systems jagged-rectangle, *Journal of Interdisciplinary Mathematics*, 20(5), 1341-1348 (2017), <http://doi:10.1080/09720502.2016.1232037>.
- [22] W. Sajjad, M. S. Sardar, M. Cancan, S. Ediz, A. Q. Baig, On modified eccentric connectivity index of nanotube. *Journal of Information and Optimization Sciences* 41(4), 959-972 (May 2020), <http://doi:10.1080/02522667.2020.1745380>.
- [23] S. Ediz, İdris Çiftçi, M. Cancan, M. R. Farahani, On k-total distance degrees and k-total Wiener polarity index. *Journal of Information and Optimization Sciences*. 42(7), 1469-1477 (2021). <https://doi.org/10.1080/02522667.2021.1896652>.
- [24] F. Afzal, Shah Rukh, D. Afzal, M. R. Farahani, M. Cancan, S. Ediz. A study of non-commutative Von-Neumann regular rings. *Journal of*

- Information and Optimization Sciences*. 42(7), 1425-1436 (2021) <https://doi.org/10.1080/02522667.2021.1896650>.
- [25] M. Shoaib Sardar, M. Alaeiyan, M. R. Farahani, M. Cancan, S. Ediz. Resistance distance in some classes of rooted product graphs obtained by Laplacian generalized inverse method. *Journal of Information and Optimization Sciences* 42(7), 1447-1467 (2021). <https://doi.org/10.1080/02522667.2021.1899210>.
- [26] M. Nadeem, S. Ahmad, M. K. Siddiqui, M. A. Ali, M. R. Farahani, A. J. M. Khalaf. On some applications related with algebraic structures through different well known graphs. *Journal of Discrete Mathematical Sciences and Cryptography*. 24(2), 451-471 (2021). <https://doi.org/10.1080/09720529.2021.1885806>.
- [27] A. Alsinai, H.M.U. Rehman, Y. Manzoor, M. Cancan, Z. Taş, M.R. Farahani, Sharp upper bounds on forgotten and SK indices of cactus graph, *Journal of Discrete Mathematical Sciences and Cryptography*, 1-22 (2022). <https://doi.org/10.1080/09720529.2022.2027605>.