International Journal of Mathematics and
Computer Science, **18**(2023), no. 2, 163–172

Enhanced spectral conjugate gradient methods for unconstrained optimization

Yoksal A. Laylani[1], Basim A. Hassan[2], Hisham M. Khudhur[2]

[1]Department of Mathematics
College of Sciences
University of Kirkuk
Kirkuk, Iraq

[2]Department of Mathematics
College of Computer Sciences and Mathematics
University of Mosul
Mosul, Iraq

yuksel@uokirkuk.edu.iq, basimah@uomosul.edu.iq,
hisham892020@uomosul.edu.iq

(Received November 1, 2022, Accepted December 6, 2022,
Published January 23, 2023)

Abstract

Spectral conjugate gradient techniques for unconstrained optimization are frequently utilized especially when the dimension is high. Based on the curvature information, updated spectral approaches for addressing unconstrained optimization problems are devised. The offered approaches satisfy the descending requirement. Moreover, the novel spectral approaches are shown to be globally convergent. When compared to the Fletcher-Reeves technique, the numerical results suggest that the proposed approaches are successful.

1 Introduction

Unconstrained optimization problems are widely used in a variety of fields including petroleum exploration, aircraft, and transportation. However, as

Key words and phrases: Spectral conjugate gradient, Global convergence, Numerical results.
AMS (MOS) Subject Classifications: 90C30, 65K05, 49M3.
ISSN 1814-0432, 2023, http://ijmcs.future-in-tech.net

the size becomes larger, the computation required climbs exponentially [5]. As a result, new approaches for solving optimization problems are necessary. For unconstrained optimization problems on a vast scale:

$$minf(x), x \in R^n, \tag{1.1}$$

where $f : R^n \to R$ is a smooth function. In the iteration of the method, the new point x_{k+1} is updated to:

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

where a line search is used to determine α_k (a step length) and d_k (a search direction) produced as:

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

where g represents the Gradient of the objective function f, and β_k is a scalar utilized to reduce a quadratic model that is strictly convex and to ensure that the directions d_{k+1} and d_k are conjugate with the objective Hessian's function. Different selections for the scalar parameter β_k correlate to different conjugate gradient techniques. The standard Wolfe (SW) criteria are frequently used in line search conjugate gradient algorithms:

$$f_{k+1} \leq f_k + \delta g_k^T s_k \tag{1.4}$$

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

where $s_k = x_{k+1} - x_k$, and $0 < \delta < \sigma < 1$. A well-known conjugate gradient approach is the Fletcher-Reeves (FR) method which specifies the parameter BB as follows:

$$\beta_k^{FR} = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}. \tag{1.6}$$

The reader is referred to [10] for further information. According to Zoutendijk [13], the FR technique with precise line search is globally convergent. Al-Baali [2] demonstrated the global convergence of the FR technique using a powerful Wolfe line search. Instructions are developed in our technique which is based on the concepts of Birgin and Martinez [3]

$$d_{k+1} = -\theta_k g_{k+1} + \beta_k s_k, \tag{1.7}$$

where $\theta_k = s_k^T s_k / y_k s_k^T$ is the spectral method, $\beta_k = (\theta_k y_k - s_k)^T g_{k+1} / s_k^T y_k$, where $y_k = g_{k+1} - g_k$. The presented numerical results suggest that the above strategy is efficient. Zhang et al. [12] suggested a modified FR approach (dubbed MFR) in which d_{k+1} is defined by

$$d_{k+1} = -\theta_k^{MFR} g_{k+1} + \beta_k^{FR} s_k, \tag{1.8}$$

where

$$\theta_k^{MFR} = \frac{y_k^T s_k}{g_k^T g_k}. \tag{1.9}$$

This is a separate descent direction from the line search. Many researchers pay great attention to improve the optimization methods [6, 7, 8]. In this paper, we develop new spectral approaches based on the curvature information. Every iteration of the technique results in a search direction that meets the descent requirement. We'll also use the Wolfe-line search to determine the proposed algorithm's global convergence.

2 Our spectral methods

In [4] new conjugate gradient formulas known as the BN1 and BN2 were straightforward and easy to apply:

$$\beta_k = \frac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k}. \tag{2.10}$$

When investigating the BN1 and BN2 methods, we respectively have:

$$\begin{array}{l} \rho_k^1 = \frac{1}{2}\frac{\alpha_k (g_k^T d_k)^2}{s_k^T y_k (s_k^T y_k + (f_{k+1} - f_k))} \\ \rho_k^2 = \frac{1}{2}\frac{\alpha_k (g_k^T d_k)^2}{s_k^T y_k (-s_k^T g_k + (f_{k+1} - f_k))}. \end{array} \tag{2.11}$$

Laylani et al. [4] produced several convergence conclusions applicable to any technique that can be stated as a ratio:

$$0 < \beta_k = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}. \tag{2.12}$$

Our concept comes from (2.12):

$$\begin{array}{l} \beta_k^{BN} g_k^T d_k = g_{k+1}^T d_{k+1} \\ g_{k+1}^T d_{k+1} = \frac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} g_k^T d_k. \end{array} \tag{2.13}$$

Since

$$y_k^T d_k = g_{k+1}^T d_k - g_k^T d_k, \tag{2.14}$$

we obtain the following from (2.12) and (2.13):

$$g_{k+1}^T d_{k+1} = -\frac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} \left[\frac{y_k^T d_k}{\rho_k s_k^T y_k} \rho_k s_k^T y_k - g_{k+1}^T d_k\right] \\ = -\frac{y_k^T d_k}{\rho_k s_k^T y_k} \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} g_{k+1}^T d_k. \tag{2.15}$$

Thus

$$g_{k+1}^T d_{k+1} = -\theta_k \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} g_{k+1}^T d_k, \tag{2.16}$$

where

$$\theta_k^{SBN} = \frac{y_k^T d_k}{\rho_k s_k^T y_k}. \tag{2.17}$$

Doing both sides of the inner product, we get

$$d_{k+1} = -\theta_k^{SBN} g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{\rho_k s_k^T y_k} d_k. \tag{2.18}$$

In this part, we provide the suggested scaled conjugate gradient method in its specific form and demonstrate its descent feature. Now we give an overview of the new algorithms that we have suggested.

2.1 New Algorithms

Step 0. Set up the $x_0 \in R^n$, $\varepsilon = 0.0001$.

Step 1. Check ; if $\|g_k\| \leq \varepsilon$, then stop.

Step 2. Calculate the β_k^{BN} and $\theta_k^{SBN1,2}$ respectively.

Step 3. Find the new point $x_{k+1} = x_k + \alpha_k d_k$.

Step 4. Determine d_{k+1} by the formulas $d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k$.

Step 5 Go back to **step 1.**

3 Global convergence

In this section, we examine global algorithm convergence. The evidence of global convergence requires the first of the following hypotheses.

Assumption 3.1. **i-** *The set of levels* $L = \{x \in R^n \,|f(x) \leq f(x_0)\}$ *is constrained.*

ii- *There is a constant $\mu_1 > 0$ such that in a region between U and L, f is continuously differentiable and its gradient is Lipschitz continuous:*

$$\|g(\chi) - g(\nu)\| \leq \mu_1 \|\chi - \nu\| \tag{3.19}$$

Further details appear in [12]. The descent property of Algorithm 2.1 is demonstrated by the next theorem.

Theorem 3.2. *Assume that the d_{k+1} is determined by* (1.7)*. Then*

$$g_{k+1}^T d_{k+1} \leq -c \|g_{k+1}\|^2 \tag{3.20}$$

Proof. The proof is by induction.
Suppose $k = 0$. We have $g_0^T d_0 = - \|g_0\|^2 < 0$ since $d_0 = -g_0$.
Assume that $g_k^T d_k < -c_1 \|g_k\|^2$ is true, for some positive constant c_1.
Multiplying (1.7) by g_{k+1}^T, we get:

$$g_{k+1}^T d_{k+1} = -\theta^{SBN} \|g_{k+1}\|^2 + \frac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} g_{k+1}^T d_k \tag{3.21}$$

$$\begin{aligned} g_{k+1}^T d_{k+1} &= \tfrac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} \left[g_{k+1}^T d_k - y_k^T d_k \right] \\ &= \tfrac{\|g_{k+1}\|^2}{\rho_k s_k^T y_k} g_k^T d_k \\ &= \tfrac{g_k^T d_k}{\rho_k s_k^T y_k} \|g_{k+1}\|^2 \end{aligned} \tag{3.22}$$

Since $g_k^T d_k < -c_1 \|g_k\|^2$, we have

$$g_{k+1}^T d_{k+1} < -c_1 \frac{\|g_k\|^2}{\rho_k s_k^T y_k} \|g_{k+1}\|^2 . \tag{3.23}$$

Put $c = c_1 \|g_k\|^2 / \rho_k s_k^T y_k$. Since c_1, ξ_{k+1} and $\|g_k\|^2$ are positive numbers, c is also a positive number with

$$g_{k+1}^T d_{k+1} \leq -c \|g_{k+1}\|^2 . \tag{3.24}$$

□

The conclusion in the following lemma, known as the Zoutendijk condition [13], is used to support the claim that the algorithms shown are globally convergent.

Lemma 3.3. *Under Assumption 1, any technique in* (1.1), (1.2) *and* (1.3) *or* d_{k+1} *that fulfills* (3.24) *satisfies*

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty. \tag{3.25}$$

The theorem demonstrates the offered approaches' global convergence.

Theorem 3.4. *Under Assumption 3.1, we allow Algorithm 2.1 to construct sequences* $\{g_{k+1}\}$ *and* $\{d_{k+1}\}$*. Then the Wolfe line search* (1.4) *and* (1.5) *to identify sequence* α_k *yields*

$$\liminf_{k \Rightarrow \infty} \|g_k\| = 0. \tag{3.26}$$

Proof. Lemma 4.1 holds under the specified conditions. We'll get the conclusion (3.26) by contradiction in the next section. Assume that there exists constant $\varepsilon_1 > 0$ such that

$$\|g_{k+1}\| > \varepsilon_1 \tag{3.27}$$

Rewriting (2.18) as follows:

$$d_{k+1} + \theta_k^{SBN} g_{k+1} = \beta_k^{BN} d_k \tag{3.28}$$

and squaring both sides, we get

$$\|d_{k+1}\|^2 + (\theta_k^{SBN})^2 \|g_{k+1}\|^2 + 2\theta_k^{SBN} d_{k+1}^T g_{k+1} = (\beta_k^{BN})^2 \|d_k\|^2 . \tag{3.29}$$

We obtain (3.30) from (3.29):

$$\|d_{k+1}\|^2 = (\beta_k^{BN})^2 \|d_k\|^2 - 2\theta_k^{SBN} d_{k+1}^T g_{k+1} - (\theta_k^{SBN})^2 \|g_{k+1}\|^2 \tag{3.30}$$

The previous equation and (2.12) allow us to infer the following:

$$\|d_{k+1}\|^2 \le \left(\frac{g_{k+1}^T d_{k+1}}{g_k^T d_k} \right)^2 \|d_k\|^2 - 2\theta_k^{SBN} d_{k+1}^T g_{k+1} - (\theta_k^{SBN})^2 \|g_{k+1}\|^2 \tag{3.31}$$

If we divide both inequalities by $(g_{k+1}^T d_{k+1})^2$, then

$$\begin{aligned} &\tfrac{\|d_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} \le \tfrac{\|d_k\|^2}{(g_k^T d_k)^2} - (\theta_k^{SBN})^2 \tfrac{\|g_{k+1}\|^2}{(d_{k+1}^T g_{k+1})^2} - 2\theta_k^{SBN} \tfrac{1}{d_{k+1}^T g_{k+1}} \\ &\le \tfrac{\|d_k\|^2}{(g_k^T d_k)^2} - \left(\theta_k^{SBN} \tfrac{\|g_{k+1}\|}{c\|g_{k+1}\|^2} + \tfrac{1}{\|g_{k+1}\|} \right)^2 + \tfrac{1}{\|g_{k+1}\|^2} \\ &\le \tfrac{\|d_k\|^2}{(g_k^T d_k)^2} + \tfrac{1}{\|g_{k+1}\|^2} \end{aligned} \tag{3.32}$$

Using (3.32) recursively and observing that $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$, we get:

$$\frac{\|d_{k+1}\|^2}{(g_{k+1}^T d_{k+1})^2} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2}. \tag{3.33}$$

From (3.33) and (3.27), we deduce

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\varepsilon_1^2}{k}. \tag{3.34}$$

From (3.34), we get

$$\sum_{k=1}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \sum_{k=1}^{\infty} \frac{\varepsilon_1^2}{k} = \infty. \tag{3.35}$$

This is in direct contradiction to the Zoutendijk condition (3.25). As a result, conclusion (3.26) is correct. □

4 Numerical results

In this section, several well-known unconstrained optimization test functions are used to evaluate the effectiveness of FORTRAN implementations of the FR technique. Numerical tests with n = 100 and 1000 variables for each function are also included [1]. We compare the algorithm FR's acne performance to the well-known FR-algorithm, defined in (1.6). In general, the most natural stopping criterion is $\|g_{k+1}\| \le 10^{-6}$. The following are included in the comparison

NOI : number of iterations.

NF : number of evaluation functions. Table 1 shows that Algorithms 2.1 outperforms the FR approach for the most often examined issues using solely Wolfe line search. As a result, the proposed approach appears to be viable and equivalent to the FR method. Problems numbers indicant for : 1. is the Trigonometric, 2. is the Extended Rosenbrock, 3. is the Hager, 4. is the Extended Tridiagonal 1, 5. is the Generalized Tridiagonal 2, 6. is the Extended PSC1, 7. is the Extended Tridiagonal 2, 8. is EDENSCH (CUTE), 9. is the STAIRCASE S1, 10. is the DENSCHNA (CUTE), 11. is the DENSCHNC (CUTE), 12. is the Extended White & Holst, 13. is the Extended Block-Diagonal BD2, 14. is the Generalized quartic GQ2, 15. is the Extended Beale.

Table 1: Comparing several CG-algorithms with various test functions and dimensions

		FR algorithm		BN1-algorithm		BN2- algorithm	
P. No.	n	NI	NF	NI	NF	NI	NF
1	100	19	35	21	39	22	40
	1000	38	65	36	64	36	68
2	100	47	93	47	98	44	91
	1000	78	131	97	169	73	127
3	100	61	1024	37	66	24	41
	1000	Fail	Fail	Fail	Fail	Fail	Fail
4	100	32	64	14	28	20	37
	1000	77	129	36	64	23	43
5	100	37	67	48	81	38	67
	1000	73	115	71	121	56	98
6	100	15	31	19	39	19	39
	1000	8	17	9	19	9	19
7	100	40	65	42	72	42	75
	1000	43	68	46	78	46	76
8	100	69	1202	Fail	Fail	70	1345
	1000	98	1967	76	842	31	67
9	100	671	1066	549	771	526	871
	1000	Fail	Fail	Fail	Fail	Fail	Fail
10	100	20	33	45	75	13	26
	1000	19	35	43	71	31	47
11	100	49	80	33	53	23	39
	1000	129	166	42	72	22	40
12	100	43	88	56	114	42	99
	1000	46	92	53	122	40	83
13	100	122	156	25	42	25	42
	1000	130	166	39	63	21	36
14	100	112	147	Fail	Fail	40	70
	1000	110	145	52	87	38	65
15	100	32	52	23	43	26	50
	1000	22	42	23	42	28	54
Total		2059	5032	1582	3335	1318	2340

Table 2: Ratio of algorithm new cost to FR cost

	FR algorithm	BN1- algorithm	BN2- algorithm
NI	100 %	76.83 %	64.01 %
NF	100 %	66.27 %	46.50 %

5 Conclusion

We have given (2.12) a novel spectral descent-type approach for estimating spectral conjugate gradient based on this described relationship. Our approach is straightforward and improves the speed of gradient-type algorithms while requiring minimum storage. The greatest outcomes come from the BN1 and BN2 approaches. The findings of the BN1 approach are less than those of the BN2 method. The novel spectral approaches save (24-36) percent NI and (34-54) percent NF, according to this table, notably for our selected test issues. Table 2 shows the percentage performance of Table 1 improvements (the paper's discussion of the relative efficacy of the various methods).

References

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

[2] Mehiddin Al-Baali, Descent property and global convergence of the Fletcher-Reeves method with inexact line search, IMA Journal of Numerical Analysis, **5,** (1985), 121–124.

[3] Ernesto G. Birgin, José Mario Martinez, A spectral conjugate gradient method for unconstrained optimization, Applied Mathematics and optimization, **43,** (2001), 117–128.

[4] Yoksal A. Laylani, Basim A. Hassan, Hisham M. Khudhur, A New Class of Optimization Methods Based on Coefficient Conjugate Gradient, European Journal of Pure and Applied Mathematics, **15,** no. 4, (2022), 1908–1916.

[5] Basim A. Hassan, M. S. Younis, M. W. Taha, A. H. Ibrahim, A new type of step sizes for unconstrained optimization, Journal of Physics: Conference Series, no. 1, (2008), 147–161.

[6] Basim A. Hassan, Mohammed Abdulrazaq Kahya, A new class of quasi-Newton updating formulas for unconstrained optimization, Journal of Interdisciplinary Mathematics, **24**, no. 8, (2008), 2355–2366.

[7] Basim A. Hassan, Ranen M. Sulaiman, Using a New Type Quasi-Newton Equation for Unconstrained Optimization, 7th International Engineering Conference, "Research and Innovation amid Global Pandemic", (IEC), IEEE, (2021), 118–122.

[8] Basim A Hassan, Ranen M. Sulaiman, A new class of self-scaling for quasi-Newton method based on the quadratic model, Indonesian Journal of Electrical Engineering and Computer Science, **21,** no. 3, (2021), 1830–1836.

[9] Elizabeth D. Dolan, Jorge J. Moré, Benchmarking optimization software with performance profiles, Mathematical programming, **91,** (2002), 201–213.

[10] Reeves Fletcher, Colin M. Reeves, Function minimization by conjugate gradients, The computer Journal, **7,** no. 2, (1964), 149–154.

[11] Hawraz N. Jabbar, Basim A. Hassan, Two-versions of descent conjugate gradient methods for large-scale unconstrained optimization, Indonesian Journal of Electrical Engineering and Computer Science, **22,** no. 3, (2021), 1643–1649.

[12] Li Zhang, Weijun Zhou, Donghui Li, Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search, Numerische Mathematik, **104,** no. 4, (2006) , 561–572.

[13] G. Zoutendijk, Nonlinear programming computational methods, In: J. Abadie, Ed., Integer and Nonlinear Programming, North Holland, Amsterdam, 1970, 37–86.