Iranian Journal of Numerical Analysis and Optimization
Vol. 11, No. 1, (2021), pp 211–219
DOI:10.22067/ijnao.2021.67048.0

Research Article

An adaptive descent extension of the Polak–Rebière–Polyak conjugate gradient method based on the concept of maximum magnification†

Z. Aminifard* and S. Babaie-Kafaki

Abstract

Recently, a one-parameter extension of the Polak–Rebière–Polyak method has been suggested, having acceptable theoretical features and promising numerical behavior. Here, based on an eigenvalue analysis on the method with the aim of avoiding a search direction in the direction of the maximum magnification by a symmetric version of the search direction matrix, an adaptive formula for computing parameter of the method is proposed. Under standard assumptions, the given formula ensures the sufficient descent property and guarantees the global convergence of the method. Numerical experiments are done on a collection of CUTEr test problems. They show practical effectiveness of the suggested formula for the parameter of the method.

AMS subject classifications (2020): 90C53, 65K05.

Keywords: Unconstrained optimization; Conjugate gradient method; Maximum magnification; Line search.

*Corresponding author
Received 19 November 2020; revised 31 December 2020; accepted 31 January 2021
Zohre Aminifard
Department of Mathematics, Semnan University, P.O. Box: 35195–363, Semnan, Iran. E-mail: aminisor@semnan.ac.ir

Saman Babaie-Kafaki
Department of Mathematics, Semnan University, P.O. Box: 35195–363, Semnan, Iran. E-mail: sbk@semnan.ac.ir

† This article was suggested by the Scientific Committee of the "13^{th} International Conference of Iranian Operations Research Society" for publication in IJNAO, which was accepted after independent review.

1 Introduction

Conjugate gradient (CG) methods can be regarded as the most popular optimization techniques due to their wide applications in the practical fields [1, 11, 12, 17]. CG algorithms are advantageous because of affordable memory storage, the simple structure of the iterative formula, promising computational performance, and acceptable convergence properties [5, 9, 13].

General form of an unconstrained optimization problem can be given by

$$\min_{x \in \mathbb{R}^n} f(x),$$

where f is a smooth real-valued nonlinear function with the gradient $g(x)$. Starting from some point $x_0 \in \mathbb{R}^n$, iterations of the CG algorithms are in the form of $x_{k+1} = x_k + s_k$ and $s_k = \alpha_k d_k$, for all $k \geq 0$, where $\alpha_k > 0$ is a step length often determined by some inexact line search techniques along the direction d_k calculated by

$$d_0 = -g_0, \ d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k \geq 0, \tag{1}$$

in which $\beta_k \in \mathbb{R}$ is called the CG parameter and $g_k = g(x_k)$. Among the various classical CG techniques, the Polak–Rebière–Polyak (PRP) method with

$$\beta_k^{\mathrm{PRP}} = \frac{g_{k+1}^T y_k}{||g_k||^2},$$

in which $y_k = g_{k+1} - g_k$ and $|| \cdot ||$ denotes the ℓ_2 norm, is regarded as an efficiently popular classic method, mainly because of adaptive restarts when dealing with improper search directions [9].

Although being computationally advantageous, the PRP method fails to ensure the descent property [9]. So, significant attention have been paid to get descent modifications of the PRP method. For example, Zhang, Zhou, and Li [18] developed (ZZL) a three-term extension of the method by

$$d_0 = -g_0, \ d_{k+1}^{\mathrm{ZZL}} = -g_{k+1} + \beta_k^{\mathrm{PRP}} d_k - \frac{g_{k+1}^T d_k}{||g_k||^2} y_k, \ k \geq 0, \tag{2}$$

satisfying the sufficient descent condition, that is,

$$d_k^T g_k \leq -\tau ||g_k||^2, \quad k \geq 0, \tag{3}$$

where $\tau > 0$ is a constant. In another effort, Andrei [3] proposed a spectral PRP (SPRP) method with

$$d_0 = -g_0, \ d_{k+1}^{\mathrm{SPRP}} = -\frac{s_k^T y_k}{||g_k||^2} g_{k+1} + \beta_k^{\mathrm{PRP}} s_k - \frac{g_{k+1}^T s_k}{||g_k||^2} y_k, \ k \geq 0, \tag{4}$$

which in addition to (3), fulfills the effective Dai–Liao conjugacy condition [6]. Also, Babaie-Kafaki and Ghanbari [4] developed a class of one-parameter extension of β_k^{PRP} (EPRP) based on the Dai–Liao approach [6]; that is,

$$\beta_k^{\mathrm{EPRP}} = \beta_k^{\mathrm{PRP}} - t\frac{g_{k+1}^T d_k}{||g_k||^2}, \tag{5}$$

where t is a positive parameter. Then, to find an optimal choice for t, they noted that from (1) and (5) search directions of EPRP can be written as

$$d_{k+1} = -H_{k+1} g_{k+1},$$

where

$$H_{k+1} = I - \frac{d_k y_k^T}{||g_k||^2} + t\frac{d_k d_k^T}{||g_k||^2}.$$

Symmetrizing H_{k+1} by

$$P_{k+1} = \frac{H_{k+1} + H_{k+1}^T}{2} = I - \frac{1}{2}\frac{d_k y_k^T + y_k d_k^T}{||g_k||^2} + t\frac{d_k d_k^T}{||g_k||^2}, \tag{6}$$

in light of an eigenvalue analysis, the following family of two-parameter choices for t was suggested in [4]:

$$t_k^{p,q} = p\frac{||y_k||^2}{||g_k||^2} + q\left(\frac{1}{2}\frac{d_k^T y_k}{||d_k||\ ||g_k||} - \frac{||g_k||}{||d_k||}\right)^2, \tag{7}$$

with $p > \frac{1}{4}$ and $q \geq -1$, guaranteeing the descent condition.

Following such studies, here we deal with another choice for parameter of the EPRP method based on the concept of the maximum magnification by a matrix. Organization of our study is summarized as follows. In Section 2, after analyzing eigenvalues of P_{k+1}, we introduce our new formula for the parameter t of the EPRP method. Also, we conduct a brief global convergence analysis. In Section 3, we make some competitive computational experiments on a collection of CUTEr problems, using the Dolan–Moré performance profile. Finally, concluding remarks are given in Section 4.

2 An adaptive choice for parameter of the extended Polak–Ribièr–Polyak method

Here, firstly we conduct a concise eigenvalue analysis on P_{k+1} in order to explain our adaptive way of computing t in (5). Hereafter, we assume that $d_k^T y_k > 0$, as ensured by the strong Wolfe line search conditions [14], that is,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \tag{8}$$

$$|\nabla f(x_k + \alpha_k d_k)^T d_k| \leq -\sigma g_k^T d_k, \tag{9}$$

with $0 < \delta < \sigma < 1$. The following basic definition is the kernel of our analysis.

Definition 1. [15] For an arbitrary matrix $A \in \mathbb{R}^{n\times m}$, the scalar

$$maxmag(A) = \max_{x\neq 0} \frac{||Ax||}{||x||},$$

is called the maximum magnification by A. Hence, $maxmag(A) = ||A||$, and also, the vector $x \neq 0$ for which $||Ax|| = ||A||\ ||x||$, is in the direction of the maximum magnification by the matrix A.

Firstly, note that the matrix P_{k+1} given by (6) can be regarded as a symmetric approximation of the search direction matrix H_{k+1}. Based on the analysis of [4], eigenvalues of P_{k+1} are 1 with multiplicity $n-2$, and λ_k^+ and λ_k^- are given by

$$\lambda_k^\pm = 1 + \frac{1}{2||g_k||^2}\left(t||d_k||^2 - d_k^T y_k\right) \pm \frac{1}{2||g_k||^2}\sqrt{(t||d_k||^2 - d_k^T y_k)^2 + ||d_k||^2||y_k||^2 - (d_k^T y_k)^2}.$$

It can be seen that with the choice (7), we have $\lambda_k^+ \geq 1 \geq \lambda_k^- > 0$, and consequently, $||P_{k+1}|| = \lambda_k^+$. Also, in light of similar analysis carried out in [2], the eigenvector of P_{k+1} corresponding to λ_k^+, here called v_1^k, can be written as $v_1^k = \gamma d_k + \vartheta y_k$ in which

$$\gamma = \frac{2(1-\lambda_k^+)||g_k||^2 - d_k^T y_k}{||d_k||^2}\vartheta.$$

So, v_1^k as a vector in the direction of the maximum magnification by P_{k+1} is specified.

As explained in [2], when the gradient is approximately parallel with the direction of the maximum magnification by H_{k+1}, then EPRP may face with some numerical errors and also, it may converge hardly. Based on this fact and since P_{k+1} is a symmetric approximation of H_{k+1}, it can be stated that if g_{k+1} is as far away as possible from the direction of the maximum magnification by P_{k+1}, then the mentioned possible errors may be diminished and the convergence may be improved. Hence, we obtain a formula for the EPRP parameter by making v_1^k to be orthogonal to g_{k+1} in the sense of solving the equation $g_{k+1}^T v_1^k = 0$; that is,

$$\bar{t}_k = \frac{||y_k||^2(g_{k+1}^T d_k)^2 - ||d_k||^2(g_{k+1}^T y_k)^2}{2(g_{k+1}^T d_k)\left((d_k^T y_k)(g_{k+1}^T d_k) - ||d_k||^2(g_{k+1}^T y_k)\right)}. \tag{10}$$

Now, for the sake of positiveness of the EPRP parameter and to achieve the sufficient descent property, we suggest the following modified version of (10):

$$t_k^* = \begin{cases} \max\left\{\bar{t}_k, \vartheta \dfrac{||y_k||^2}{||g_k||^2}\right\}, & \text{if denominator of } \bar{t}_k \text{ is nonzero,} \\ \vartheta \dfrac{||y_k||^2}{||g_k||^2}, & \text{otherwise,} \end{cases} \tag{11}$$

with $\vartheta > \dfrac{1}{4}$.

Now, similar to the analysis conducted in the proof of [16, Theorem 3.2], the following convergence result can be established for the EPRP method. The proof is omitted to avoid repetition.

Theorem 1. Suppose that the level set $\Omega = \{x \in \mathbb{R}^n \mid f(x) \leq f(x_0)\}$ is bounded and in some neighborhood $\mathcal{N}$ of Ω, that the objective function f is smooth and also, that ∇f is Lipschitz continuous. For the EPRP method with the parameter (5), assume that t is equal to t_k^* defined by (11) and the line search fulfills the strong Wolfe conditions (8) and (9). If there exists a positive lower bound α^* for the step lengths α_k (for all $k \geq 0$), then $\lim_{k\to\infty} ||g_k|| = 0$.

3 Computational experiments

In this section, we examine the numerical efficiency of the EPRP method in which t is computed by (11) with $\vartheta = 0.26$ and (7) with $(p, q) = (1, 0)$; here the corresponding methods are, respectively, called EPRP1 and EPRP2. The methods are compared by the two modified PRP methods of ZZL and SPRP, respectively, with the search directions (2) and (4) [3, 18]. We have implemented all the algorithms on a set of 45 test functions of the CUTEr library [8] with $n \geq 50$, as given in Table 1, in MATLAB software environment. Hardware and software detailed specifications have been clarified in [2], together with the strong Wolfe line search features and the stopping criteria. Detailed outputs have been provided in Table 1.

Efficiency of the algorithms was compared by applying the performance profile proposed by Dolan and Moré [7] on the norm of gradient, the CPU time (CPUT), and the total number of function and gradient evaluations (TNFGE), following the notation of [10]. Results are shown by Figures 1–3. As seen, EPRP is preferable to the other methods. Particularly, the results show that the choice (11) for the EPRP parameter is practically effective. Our experiments showed that averagely in 62.63% of the iterations of EPRP1, we had $t_k^* = \bar{t}_k$.

Table 1: Outputs

Function	n	EPRP1 TNFGE	EPRP1 CPUT	EPRP2 TNFGE	EPRP2 CPUT	SPRP TNFGE	SPRP CPUT	ZZL TNFGE	ZZL CPUT
ARGLINA	200	12	2.13E-01	12	9.40E-02	12	7.97E-02	12	8.93E-02
ARWHEAD	5000	103	2.17E-01	20095	4.25E+00	40053	8.25E+00	23730	4.89E+00
BDEXP	5000	12	9.73E-02	12	7.68E-02	12	6.97E-02	12	6.44E-02
BDQRTIC	5000	26021	1.62E+01	26069	1.34E+01	40023	1.13E+01	26073	6.95E+00
BQPGABIM	50	299	7.67E-02	238	4.76E-02	2131	1.02E-01	685	5.10E-02
BQPGASIM	50	299	3.74E-02	238	3.18E-02	2131	9.70E-02	685	5.05E-02
BROWNAL	200	4641	6.26E-01	38227	4.25E+00	40084	3.93E+00	40084	4.07E+00
BRYBND	5000	659	4.07E-01	1853	6.93E-01	12334	4.62E+00	493	2.71E-01
CHENHARK	5000	816	2.57E-01	633	1.79E-01	3196	5.23E-01	6832	1.46E+00
COSINE	10000	27	1.29E-01	37	1.15E-01	46	1.33E-01	38	1.18E-01
CRAGGLVY	5000	431	6.99E-01	447	5.75E-01	3204	2.90E+00	1127	1.38E+00
CURLY10	10000	457	3.66E-01	479	3.60E-01	2597	1.27E+00	1273	8.53E-01
DIXMAANA	3000	36	6.02E-02	31	4.79E-02	31	4.95E-02	31	4.92E-02
DIXMAANB	3000	32	6.26E-02	27	5.28E-02	32	5.96E-02	32	4.76E-02
DIXMAANC	3000	32	6.10E-02	32	4.90E-02	32	5.10E-02	32	5.40E-02
DIXMAAND	3000	37	5.74E-02	37	5.18E-02	37	5.10E-02	37	5.09E-02
DQDRTIC	5000	1408	5.81E-01	1652	5.66E-01	11183	2.71E+00	4661	1.62E+00
DQRTIC	5000	4	4.89E-02	4	3.88E-02	4	4.15E-02	4	4.44E-02
DRCAV1LQ	4489	4	9.58E-02	4	8.56E-02	4	8.18E-02	4	7.50E-02
DRCAV2LQ	4489	4	8.63E-02	4	1.03E-01	4	8.63E-02	4	1.03E-01
DRCAV3LQ	4489	4	8.66E-02	4	8.95E-02	4	9.29E-02	4	8.79E-02
EDENSCH	2000	76	8.14E-02	89	7.10E-02	94	6.73E-02	89	7.07E-02
EG2	1000	23	3.52E-02	23	4.09E-02	23	3.23E-02	23	2.21E-02
ENGVAL1	5000	58	1.14E-01	58	9.18E-02	58	9.11E-02	62	9.48E-02
FLETCBV2	5000	4	1.07E-01	4	8.07E-02	4	8.45E-02	4	7.25E-02
FLETCBV3	5000	164	1.89E-01	164	1.62E-01	160	1.88E-01	164	1.61E-01
FLETCHBV	5000	156	1.50E-01	156	1.44E-01	148	1.45E-01	156	1.48E-01
FREUROTH	5000	275	2.21E-01	167	1.73E-01	2109	9.82E-01	987	5.74E-01
GENHUMPS	5000	4	6.25E-02	4	6.28E-02	4	5.86E-02	4	6.33E-02
MANCINO	100	93	3.20E-01	93	2.49E-01	98	2.56E-01	98	2.55E-01
MOREBV	5000	95	9.49E-02	115	9.50E-02	971	2.78E-01	380	1.70E-01
NCB20	5010	372	6.34E-01	905	1.26E+00	1505	1.85E+00	545	7.49E-01
NONCVXU2	5000	4	5.49E-02	4	4.40E-02	4	4.72E-02	4	3.32E-02
PENALTY2	200	4	2.88E-02	4	1.95E-02	4	8.67E-03	4	9.30E-03
QUARTC	5000	4	5.72E-02	4	3.82E-02	4	4.24E-02	4	3.13E-02
SCHMVETT	5000	55	1.94E-01	54	1.63E-01	58	1.67E-01	54	1.54E-01
SENSORS	100	88	4.51E-01	87	3.83E-01	110	5.27E-01	98	4.10E-01
SINQUAD	5000	123	3.00E-01	106	2.88E-01	191	3.37E-01	117	2.82E-01
SPARSQUR	10000	274	4.10E-01	187	2.81E-01	226	3.12E-01	235	3.36E-01
TOINTGOR	50	655	7.83E-02	725	5.58E-02	4425	1.96E-01	1812	1.26E-01
TOINTGSS	5000	83	1.81E-01	64	1.07E-01	1072	1.02E+00	321	3.57E-01
TOINTQOR	50	123	2.57E-02	145	2.39E-02	166	2.21E-02	244	2.67E-02
VARDIM	200	8	4.23E-02	8	1.94E-02	8	3.49E-03	8	8.71E-03
VAREIGVL	50	113	3.45E-02	122	3.85E-02	132	3.05E-02	133	8.70E-03
WOODS	4000	164	1.22E-01	180	9.18E-02	746	1.84E-01	159	8.35E-02

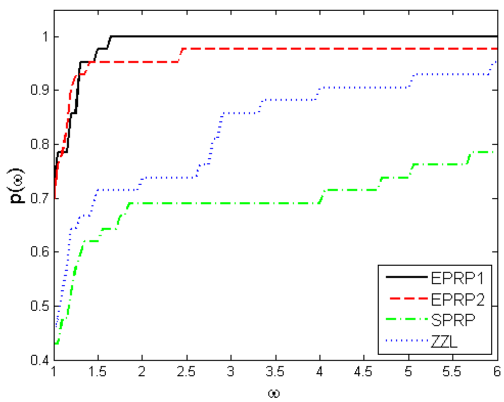

Figure 1: TNFGE performance profiles

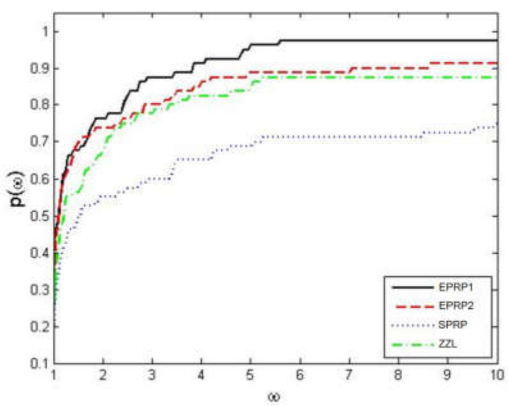

Figure 2: CPUT performance profiles

4 Conclusion

Based on the concept of maximum magnification, we have conducted an eigenvalue analysis to suggest an optimal choice for the parameter of the recently proposed EPRP method. The suggested formula guarantees the sufficient descent property as well as the global convergence of the method. Effect of the proposed formula has been numerically investigated in contrast to several other modifications made on the classical PRP method. Results showed the effectiveness of the suggested choice for the EPRP parameter.

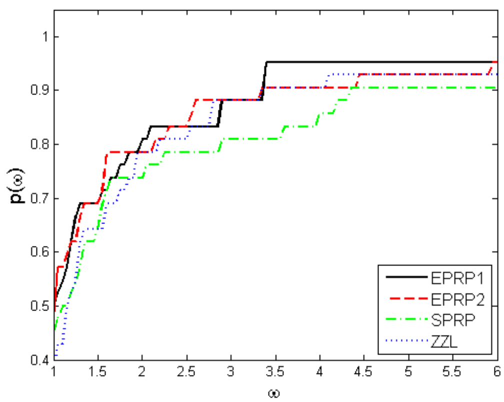

Figure 3: Norm of gradient performance profiles

References

1. Abubakar, A.B., Kumam, P. and Awwal, A.M. *Global convergence via descent modified three-term conjugate gradient projection algorithm with applications to signal recovery*, Result. Appl. Math. 4 (2019), 100069.

2. Aminifard, Z. and Babaie-Kafaki, S. *An optimal parameter choice for the Dai–Liao family of conjugate gradient methods by avoiding a direction of the maximum magnification by the search direction matrix*, 4OR, 17(3) (2019), 317–330.

3. Andrei, N. *A modified Polak–Ribière–Polyak conjugate gradient algorithm for unconstrained optimization*, Optimization, 60(12) (2011), 1457–1471.

4. Babaie-Kafaki, S. and Ghanbari, R. *A descent extension of the Polak–Ribière–Polyak conjugate gradient method*, Comput. Math. Appl. 68(12) (2014), 2005–2011.

5. Dai, Y.H., Han, J.Y., Liu, G.H., Sun, D.F., Yin, H.X. and Yuan, Y.X. *Convergence properties of nonlinear conjugate gradient methods*, SIAM J. Optim. 10(2) (1999), 348–358.

6. Dai, Y. H. and Liao, L. Z. *New conjugacy conditions and related nonlinear conjugate gradient methods*, Appl. Math. Optim. 43(1) (2001), 87–101.

7. Dolan, E.D. and Moré, J.J. *Benchmarking optimization software with performance profiles*, Math. Programming (Ser. A), 91(2) (2002), 201–213.

8. Gould, N.I.M., Orban, D. and Toint, Ph.L. *CUTEr: A constrained and unconstrained testing environment*, ACM Trans. Math. Software, 29(4) (2003), 373–394.

9. Hager, W.W. and Zhang, H. *A survey of nonlinear conjugate gradient methods*, Pac. J. Optim. 2(1) (2006), 35–58.

10. Hager, W. W. and Zhang, H. *Algorithm 851: CG-Descent, a conjugate gradient method with guaranteed descent*, ACM Trans. Math. Software, 32(1) (2006), 113–137.

11. Heravi, A.R. and Hodtani, G.A. *A new correntropy-based conjugate gradient backpropagation algorithm for improving training in neural networks*, IEEE Trans. Neural Netw. Learn. Syst. 29(12) (2018), 6252–6263.

12. Lin, J. and Jiang, C., *An improved conjugate gradient parametric detection based on space-time scan*, Signal Process. 169 (2020), 107412.

13. Nocedal, J. and Wright S.J. *Numerical optimization*, Springer, New York, 2006.

14. Sun, W. and Yuan, Y.X. *Optimization theory and methods: Nonlinear programming*, Springer, New York, 2006.

15. Watkins, D.S. *Fundamentals of matrix computations*, John Wiley and Sons, New York, 2002.

16. Yuan, G., *Modified nonlinear conjugate gradient methods with sufficient descent property for large-scale optimization problems*, Optim. Lett. 3(1) (2009), 11–21.

17. Yuan, G., Lu, J. and Wang, Z. *The PRP conjugate gradient algorithm with a modified WWP line search and its application in the image restoration problems*, Appl. Numer. Math. 152 (2020), 1–11.

18. Zhang, L., Zhou, W. and Li, D.H. *A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence*, IMA J. Numer. Anal. 26(4) (2006), 629–640.