Journal of Mathematical Modeling
Vol. -, No. -, 2022, pp. -

JMM

Improving the Dai–Liao parameter choices using a fixed point equation

Saman Babaie-Kafaki*, Zohre Aminifard

Department of Mathematics, Semnan University, P.O. Box: 35195–363, Semnan, Iran
Email(s): sbk@semnan.ac.ir, aminisor@semnan.ac.ir

Abstract. Recently, based on a singular value analysis on the Dai–Liao conjugate gradient method, Babaie-Kafaki and Aminifard suggested a fixed point equation. The prominent feature of the proposed equation is that its solutions may increase numerical stability of the method while improving the global convergence. Here, the same fixed point equation is employed to upgrade previously proposed choices of the Dai–Liao parameter based on the well-known functional iteration method. Global convergence analysis is conducted and numerical experiments are done to support our discussions.

Keywords: Unconstrained optimization, conjugate gradient method, maximum magnification, fixed point equation, functional iteration.
AMS Subject Classification 2010: 90C53, 65K05, 37C25.

1 Introduction

As of late, the one-parameter class of conjugate gradient (CG) methods proposed by Dai and Liao (DL) [9] has attracted special attention in developing efficient tools for solving the unconstrained optimization problem $\min_{x \in \mathbb{R}^n} f(x)$ where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth nonlinear function. Iterative formula of the method is

$$x_0 \in \mathbb{R}^n,\ x_{k+1} = x_k + s_k,\ s_k = \alpha_k d_k,\ k \geq 0, \tag{1}$$

where $\alpha_k > 0$ is a step size generated by the line search discussed in [17] along the direction

$$d_0 = -g_0,\ d_{k+1} = -g_{k+1} + \beta_k d_k,\ k \geq 0, \tag{2}$$

with the CG parameter

$$g_k := g(x_k) := \nabla f(x_k),\ \beta_k := \beta_k^{DL} := \frac{g_{k+1}^T y_k}{d_k^T y_k} - t\frac{g_{k+1}^T s_k}{d_k^T y_k},\ y_k = g_{k+1} - g_k. \tag{3}$$

*Corresponding author.
Received: 23 June 2020/ Revised: 25 February 2021/ Accepted: 5 March 2021
DOI: 10.22124/jmm.2021.16900.1466

Here t is a nonnegative parameter. Note that if the line search fulfills the strong Wolfe conditions [17],

$$f(x_k+\alpha_k d_k)-f(x_k) \leq \delta\alpha_k g_k^T d_k, \tag{4}$$
$$|d_k^T g_{k+1}| \leq -\sigma d_k^T g_k, \tag{5}$$

with $0<\delta<\sigma<1$, then we have $d_k^T y_k>0$, and so β_k^{DL} is well-defined.

It is worth noting that performance of the DL method is closely dependent on the choice of t for which an optimal formula has not been achieved yet [1]. Nevertheless, some recent adaptive choices for the DL parameter can be listed as follows:

$$t=t_k^{(1)}=2\frac{||y_k||^2}{s_k^T y_k}, \text{ by Hager and Zhang [14];} \tag{6}$$
$$t=t_k^{(2)}=\frac{||y_k||^2}{s_k^T y_k}, \text{ by Dai and Kou [8];} \tag{7}$$
$$t=t_k^{(3)}=\frac{s_k^T y_k}{||s_k||^2}+\frac{||y_k||}{||s_k||}, \text{ by Babaie-Kafaki and Ghanbari [5];} \tag{8}$$
$$t=t_k^{(4)}=\frac{||y_k||}{||s_k||}, \text{ by Babaie-Kafaki and Ghanbari [5];} \tag{9}$$
$$t=t_k^{(5)}=\frac{s_k^T y_k}{||s_k||^2}, \text{ by Andrei [2];} \tag{10}$$
$$t=t_k^{(6)}=\frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}, \text{ by Babaie-Kafaki and Ghanbari [7];} \tag{11}$$
$$t=t_k^{(7)}=\sqrt{\frac{||y_k||(s_k^T y_k)}{||s_k||^3}}, \text{ by Babaie-Kafaki and Ghanbari [6];} \tag{12}$$

where $||.||$ stands for the Euclidean norm. Especially, the DL method with

$$t=\vartheta\frac{||y_k||^2}{s_k^T y_k},$$

where $\vartheta>\frac{1}{4}$ is a constant, has been established to satisfy the sufficient descent condition

$$d_k^T g_k \leq -c||g_k||^2,\ k\geq 0, \tag{13}$$

where c is a positive constant, independent of the objective function convexity and the line search conditions (see also [3]).

Recently, conducting a singular value analysis [18], Babaie-Kafaki and Aminifard [4] computed direction of the maximum magnification by search direction matrix of the DL method [5],

$$Q_{k+1}=I-\frac{s_k y_k^T}{s_k^T y_k}+t\frac{s_k s_k^T}{s_k^T y_k}, \tag{14}$$

where I is an $n \times n$ identity matrix. In light of the analysis carried out in [5], the largest singular value of Q_{k+1}, namely σ_k^+, can be computed as follows:

$$\sigma_k^+ = \frac{1}{2}\frac{\sqrt{a_k^2 + b_k^2 - q_k^2}}{q_k} + \frac{1}{2}\frac{\sqrt{p_k^2 + b_k^2 - q_k^2}}{q_k},$$

with $q_k = s_k^T y_k$, $a_k = t||s_k||^2 + q_k$, $b_k = ||s_k||\,||y_k||$, and $p_k = t||s_k||^2 - q_k$.

Now, considering (14), the DL search directions can be generated by:

$$d_{k+1} = -Q_{k+1} g_{k+1},\ k \geq 0. \tag{15}$$

As discussed in [4], when the gradient and the direction of the maximum magnification by the search direction matrix are approximately parallel, from (15) we may have $||g_{k+1}|| \ll ||d_{k+1}||$. Therefore, numerical errors (such as data swamping) may corrupt the solution and slow down the convergence speed (see Theorem 2.1 of [11]). Hence, if the gradient g_{k+1} is far from the direction of the maximum magnification by Q_{k+1} as much as possible (kind of being orthogonal), impact of the mentioned troubles may be decreased. Based on this fact, in [4] it has been noted that the fixed point equation $t = \varphi(t)$ with

$$\varphi(t) = \frac{(\sigma_k^+)^2}{1-(\sigma_k^+)^2}\left(\frac{s_k^T y_k}{||s_k||^2} - \frac{g_{k+1}^T y_k}{g_{k+1}^T s_k}\right), \tag{16}$$

should be solved to achieve an appropriate formula for computing the DL parameter. As a result, they suggested the following adaptive choice:

$$t = t_k^{(8)} = \begin{cases} \max\left\{\dfrac{e_k \pm \sqrt{e_k^2 + 4h_k^2 b_k^2}}{2h_k ||s_k||^2}, \vartheta\dfrac{||y_k||^2}{s_k^T y_k}\right\}, & g_{k+1}^T s_k \neq 0 \text{ and } h_k \neq 0, \\ \vartheta\dfrac{||y_k||^2}{s_k^T y_k}, & \text{otherwise}, \end{cases} \tag{17}$$

in which

$$h_k = \frac{s_k^T y_k}{||s_k||^2} - \frac{g_{k+1}^T y_k}{g_{k+1}^T s_k}, \quad e_k = ||y_k||^2 - h_k^2 ||s_k||^2 - \frac{(s_k^T y_k)^2}{||s_k||^2},$$

and

$$\vartheta > \frac{1}{4},$$

is a constant.

Here, using the fixed point equation (16) obtained based on the maximum magnification by the matrix Q_{k+1}, we propose a general modification approach for the adaptive choices of the DL parameter. This is the subject of Section 2, together with a brief convergence analysis. We provide a test bed to shed light on the merits of our modification scheme in Section 3. Finally, in Section 4 we come out with concluding remarks. Hereafter, we assume that $d_k^T y_k > 0$, as guaranteed by the strong Wolfe conditions (4) and (5).

2 A modification scheme for the Dai–Liao parameter choices

As known, functional iteration method is a popular technique for solving fixed point equations [16]. Especially, near the solution computational behavior of the method is acceptable. So, in order to take advantage of the fixed point equation (16), if we have the appropriate choice $t = t_k$ for the DL parameter, then we may suggest its modified version as $\bar{t}_k = \varphi(t_k)$. However, the choice $t = \bar{t}_k$ may be nonpositive. In addition, generally it may not guarantee the descent property for the DL method. Hence, based on the eigenvalue analysis conducted in [3], we suggest the following one-parameter class of choices for the DL parameter:

$$t = \tau(t_k) = \begin{cases} \max\left\{\varphi(t_k), \vartheta \dfrac{||y_k||^2}{s_k^T y_k}\right\}, & g_{k+1}^T s_k \neq 0 \text{ and } \sigma_k^+ > 1, \\ \vartheta \dfrac{||y_k||^2}{s_k^T y_k}, & \text{otherwise}, \end{cases} \tag{18}$$

with the constant $\vartheta > \dfrac{1}{4}$. As can be seen, different choices for t_k yield different formulas for $\tau(t_k)$. Now, similar to the proof of Lemma 3.1 of [4], the following result can be established for the DL parameter $t = \tau(t_k)$. The proof is ignored to avoid verbosity.

Lemma 1. *For the DL method with $t = \tau(t_k)$ given by (18), if the line search guarantees that $d_k^T y_k > 0$, for all $k \geq 0$, then the sufficient descent condition (13) holds.*

In what follows, we discuss the global convergence of the DL method with the parameter $t = \tau(t_k)$ given by (18). In this context, the following preliminaries are needed [11].

Assumption 1. *For arbitrary $x_0 \in \mathbb{R}^n$, $\mathscr{L} = \{x : f(x) \leq f(x_0)\}$ is a bounded set and in some neighborhood $\mathscr{U}$ of $\mathscr{L}$, $\nabla f(x)$ is Lipschitz continuous; that is, there exists a constant $L > 0$ such that*

$$||\nabla f(x) - \nabla f(\check{x})|| \leq L||x - \check{x}||, \ \forall x, \check{x} \in \mathscr{U}. \tag{19}$$

The following theorem ensures the global convergence of the method for uniformly convex functions [17]. The proof is similar to the proof of Theorem 3.3 of [9] and here is omitted.

Theorem 1. *Suppose that Assumption 1 holds and the CG method using (1)–(2) and (18) is applied. If t is uniformly bounded, the objective function f is uniformly convex on $\mathscr{U}$ and the step size α_k is computed to fulfill the strong Wolfe conditions (4) and (5), then the method converges in the sense that $\lim_{k\to\infty} ||g_k|| = 0$.*

To ensure boundedness of t (as assumed in Theorem 1), we can set $t \leftarrow \min\{t, M\}$, where M is a large positive constant. Also, in order to achieve global convergence for general functions, we can employ the following restricted version of the CG parameter (3) [9]:

$$\beta_k^{DL+} = \max\left\{\frac{g_{k+1}^T y_k}{d_k^T y_k}, 0\right\} - t\frac{g_{k+1}^T s_k}{d_k^T y_k}. \tag{20}$$

3 Numerical experiments

In this section, we investigate computational effect of the choice $t = \tau(t_k)$ given by (18) on the DL+ method in which the CG parameter is computed by (20), with different eight adaptive choices (6)–(12) and (17) for t_k. Here, for $i = 1,2,\ldots,8$, the method with $t = t_k^{(i)}$ which is called DL*i*+ is compared by its modified version with $t = \tau(t_k^{(i)})$ which is called MDL*i*+. The runs were performed on a set of 153 unconstrained optimization test problems of the CUTEr collection [12] as given in Table 1. The hardware and software specifications have been clarified in [4]. In the line search procedure, the strong Wolfe conditions (4) and (5) have been employed using Algorithm 3.5 of [17] with $\delta = 0.0001$ and $\sigma = 0.99$. The steepest descent direction employed when an uphill search direction was generated [9]. The efficiency of all algorithms was evaluated by applying the performance profile of Dolan–Moré [10]. To do so, the total number of function and gradient evaluations (TNFGE) being equal to $N_f + 2N_g$, where N_f and N_g respectively denote the number of function and gradient evaluations, and the times in second (TSEC) were used as the cost measures. Moreover, the algorithms were stopped when TNFGE$>$ $20000+50n$ (where n denotes the dimension of the problem) or achieving a solution with $||g_k|| < 10^{-6}$.

Figures 1 and 2 illustrate the results of comparisons (based on the considerations of [13]) in which $p(\omega)$ stands for the Dolan–Moré performance profile in the level ω [10]. As seen, MDL*i*+ for $i = 3,\ldots,8$ are superior to DL*i*, with respect to TNFGE and TSEC. Also, from Figures 1 and 2 it can be concluded that MDL1+ and MDL2+ respectively are preferable to DL1+ and DL2+ with respect to TNFGE while in viewpoint of TSEC, MDL1+ and DL1+ as well as MDL2+ and DL2+ are competitive. Hence, employing the given fixed point equation can enhance effectiveness of the previously proposed adaptive choices of the DL parameter. As a final remark, Table 2 showed that averagely at almost $\frac{1}{3}$ of the iterations of the MDL*i*+ methods ($i = 1,2,\ldots,8$) our modification scheme were employed in the sense that we had $t = \varphi(t_k^{(i)})$ in (18).

4 Conclusions and future works

Finding optimal choices for the parameter of the Dai–Liao method is a classical open problem in the conjugate gradient methods. In this regard, by orthogonalizing the direction of the maximum magnification by the search direction matrix of the method to the gradient vector, a fixed point equation has been obtained. As known, functional iteration method is a popular technique for solving fixed point equations, being computationally promising near the solution. Based on this fact, a modification scheme has been proposed for the classical reasonable choices of the Dai–Liao parameter, yielding a one-parameter class of adaptive choices. To investigate practical effect of our approach, several pairwise comparisons have been done on a set of CUTEr test problems, using the Dolan–Moré performance profile. They showed the proposed modified methods outperform their classical versions. Especially, the experiments demonstrated that averagely in 32.55% of the iterations of the modified methods our fixed point scheme were employed.

As a final note, the functional iteration method for the fixed point equation $t = \varphi(t)$ can have good convergence behavior when $\varphi(t)$ given by (16) is a contractive function on a compact interval containing the solution as an inner point [16]. It is worth noting that in our experiments we have observed that $\varphi'(t_k) < 1$, for all $k \geq 0$. So, as a future work, one can establish the inequality $\varphi'(t) < 1$ which ensures that $\varphi(t)$ is contractive. Also, employing such fixed point approaches on the other iterative methods of

Table 1: Test problems data.

Function	n	Function	n	Function	n
AKIVA	2	DIXMAANL	3000	KOWOSB	4
ALLINIT	4	DIXMAANM	15	LIARWHD	5000
ALLINITU	4	DIXMAANN	15	LOGHAIRY	2
ARGLINA	200	DIXON3DQ	10000	MANCINO	100
ARGTRIGL	10	DJTL	2	MARATOSB	2
ARWHEAD	5000	DJTL	2	MEXHAT	2
BARD	3	DMN15102	66	MOREBV	5000
BDEXP	5000	DMN15103	99	MSQRTALS	1024
BDQRTIC	5000	DMN37142	66	MSQRTBLS	1024
BEALE	2	DMN37143	99	NONCVXU2	5000
BENNETT5	3	DQDRTIC	5000	NONDIA	5000
BIGGS6	6	DQRTIC	5000	NONDQUAR	5000
BIGGSB1	5000	DRCAV1LQ	4489	OSBORNEA	5
BOX3	3	DRCAV2LQ	4489	OSBORNEB	11
BQP1VAR	1	DRCAV3LQ	4489	OSCIPATH	10
BQPGABIM	50	ECKERLE4	3	PALMER1D	7
BQPGASIM	50	EDENSCH	2000	PALMER2C	8
BRKMCC	2	EG2	1000	PALMER3C	8
BROWNAL	200	EIGENALS	2550	PALMER4C	8
BROWNBS	2	EIGENBLS	2550	PALMER5C	6
BROWNBS	2	EIGENCLS	2652	PALMER6C	8
BROWNDEN	4	ENGVAL2	3	PALMER7C	8
BROYDN7D	5000	ERRINROS	50	PALMER8C	8
BRYBND	5000	EXPFIT	2	PARKCH	15
CHAINWOO	4000	EXTROSNB	1000	PENALTY1	1000
CHENHARK	5000	FLETCBV2	5000	PENALTY2	200
CHNROSNB	50	FLETCBV3	5000	POWELLSG	5000
CHWIRUT2	3	FLETCHCR	1000	POWER	10000
CLIFF	2	FMINSRF2	5625	QUARTC	5000
CLPLATEB	5041	FMINSURF	5625	ROSENBR	2
CUBE	2	GAUSS1LS	8	S308	2
DECONVU	63	GAUSS3LS	8	SINEVAL	2
DENSCHNA	2	GENHUMPS	5000	SISSER	2
DENSCHNB	2	GENROSE	500	SNAIL	2
DENSCHNC	2	GROWTHLS	3	SPARSINE	5000
DENSCHNC	2	GULF	3	SPARSQUR	10000
DENSCHND	3	HAIRY	2	SPMSRTLS	4999
DENSCHND	3	HATFLDD	3	SROSENBR	5000
DENSCHNE	3	HATFLDE	3	TESTQUAD	5000
DENSCHNF	2	HATFLDFL	3	TOINTGOR	50
DIXMAANA	3000	HEART6LS	6	TOINTGSS	5000
DIXMAANB	3000	HEART8LS	8	TOINTPSP	50
DIXMAANC	3000	HELIX	3	TOINTQOR	50
DIXMAAND	3000	HILBERTA	2	TQUARTIC	5000
DIXMAANE	3000	HILBERTB	10	TRIDIA	5000
DIXMAANF	3000	HIMMELBB	2	VARDIM	200
DIXMAANG	3000	HIMMELBF	4	VAREIGVL	50
DIXMAANH	3000	HIMMELBG	2	VIBRBEAM	8
DIXMAANI	3000	HIMMELBH	2	WATSON	12
DIXMAANJ	3000	HUMPS	2	WOODS	4000
DIXMAANK	3000	JENSMP	2	YFITU	3

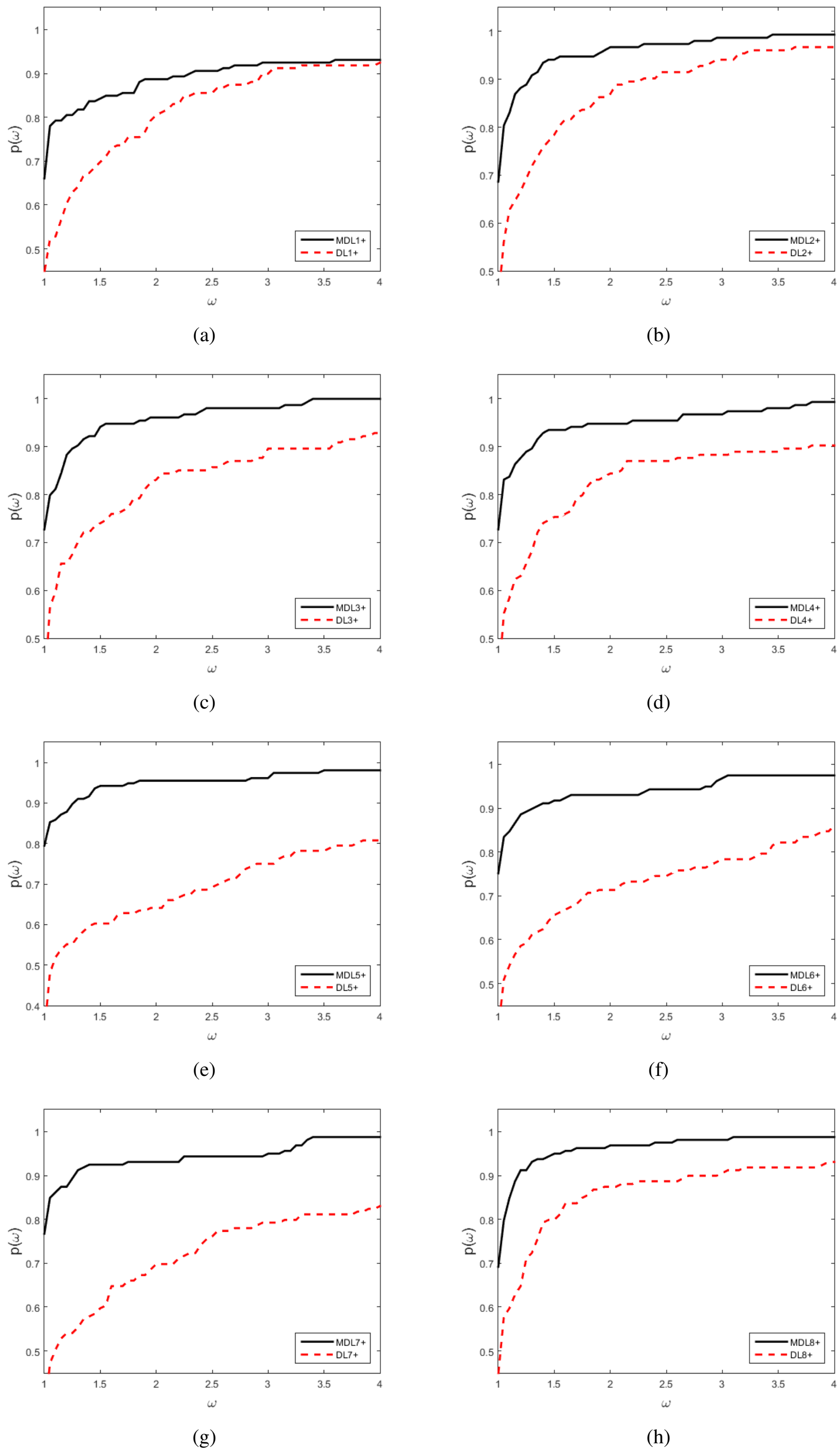
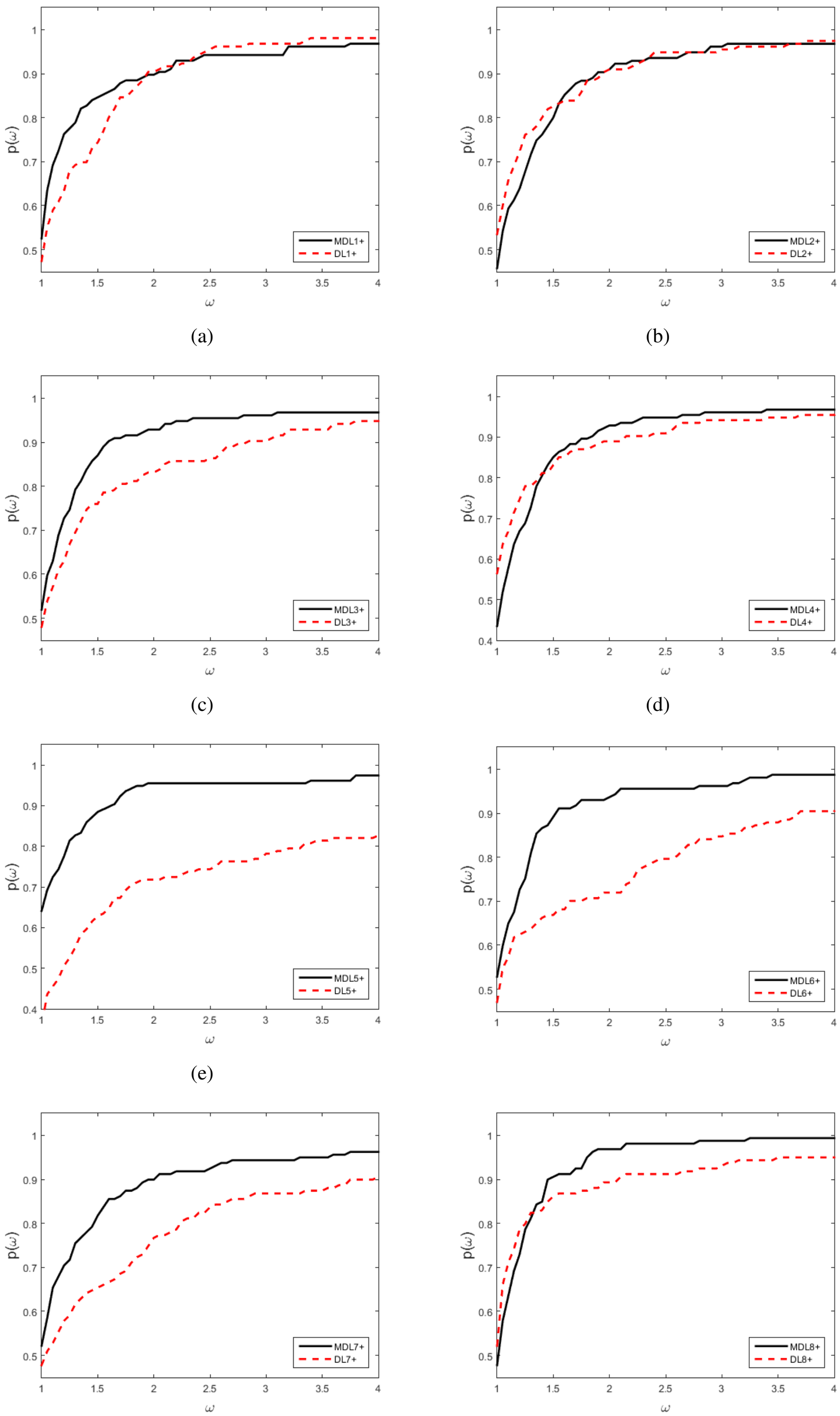


Figure 1: Performance profile in terms of TNFGE.



(a) (b)

(c) (d)

(e)

(f) (g)

Figure 2: Performance profile in terms of TSEC.

Table 2: Average percentage of the iterations with $t = \varphi(t_k^{(i)})$ for the methods MDLi+, $i = 1, 2, \ldots, 8$.

MDL1+	MDL2+	MDL3+	MDL4+	MDL5+	MDL6+	MDL7+	MDL8+
29.89%	31.81%	31.74%	33.21%	32.72%	37.49%	31.55%	39.34%

unconstrained optimization can be investigated.

Acknowledgements

This research was supported by Research Council of Semnan University. The authors thank the anonymous reviewers for their valuable comments and suggestions helped to improve the quality of this work. They are also grateful to Professor Michael Navon for providing the line search code.

References

[1] N. Andrei, *Open problems in conjugate gradient algorithms for unconstrained optimization*, B. Malays. Math. Sci. So. **34** (2011) 319–330.

[2] N. Andrei, *A Dai–Liao conjugate gradient algorithm with clustering of eigenvalues*, Numer. Algorithms **77** (2018) 1273–1282.

[3] S. Babaie-Kafaki, *On the sufficient descent condition of the Hager–Zhang conjugate gradient methods*, 4OR **12** (2014) 285–292.

[4] S. Babaie-Kafaki, Z. Aminifard, *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** (2019) 317–330.

[5] S. Babaie-Kafaki, R. Ghanbari, *The Dai–Liao nonlinear conjugate gradient method with optimal parameter choices*, European J. Oper. Res. **234** (2014) 625–630.

[6] S. Babaie-Kafaki, R. Ghanbari, *Two optimal Dai–Liao conjugate gradient methods*, Optimization **64** (2015) 2277–2287.

[7] S. Babaie-Kafaki, R. Ghanbari, *A class of adaptive Dai–Liao conjugate gradient methods based on the scaled memoryless BFGS update*, 4OR **15** (2017) 85–92.

[8] Y.H. Dai, C.X. Kou, *A nonlinear conjugate gradient algorithm with an optimal property and an improved Wolfe line search*, SIAM J. Optim. **23** (2013) 296–320.

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

[10] E.D. Dolan, J.J. Moré, *Benchmarking optimization software with performance profiles*, Math. Program. **91** (2002) 201–213.

[11] J.C. Gilbert, J. Nocedal, *Global convergence properties of conjugate gradient methods for optimization*, SIAM J. Optim. **2** (1992) 21–42.

[12] N.I.M. Gould, D. Orban, Ph.L. Toint, *CUTEr: a constrained and unconstrained testing environment, revisited*, ACM Trans. Math. Software **29** (2003) 373–394.

[13] N.I.M. Gould, J. Scott, *A note on performance profiles for benchmarking software*, ACM Trans. Math. Software **43** (2016) 1–5.

[14] W.W. Hager, H. Zhang, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim. **16** (2005) 170–192.

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

[16] D. Kincaid, W. Cheney, *Numerical Analysis*, Brooks/Cole Publishing Company, California, 1991.

[17] J. Nocedal, S.J. Wright. *Numerical Optimization*, Springer, New York, 2006.

[18] D.S. Watkins, *Fundamentals of Matrix Computations*, John Wiley and Sons, New York, 2002.