*ANZIAM J.* 61(2019), 195–203
doi:10.1017/S1446181119000063

# MATRIX ANALYSES ON THE DAI–LIAO CONJUGATE GRADIENT METHOD

Z. AMINIFARD[1] and S. BABAIE-KAFAKI✉[1]

(Received 17 April, 2018; accepted 8 March, 2019; first published online 9 May 2019)

**Abstract**

Some optimal choices for a parameter of the Dai–Liao conjugate gradient method are proposed by conducting matrix analyses of the method. More precisely, first the $\ell_1$ and $\ell_\infty$ norm condition numbers of the search direction matrix are minimized, yielding two adaptive choices for the Dai–Liao parameter. Then we show that a recent formula for computing this parameter which guarantees the descent property can be considered as a minimizer of the spectral condition number as well as the well-known measure function for a symmetrized version of the search direction matrix. Brief convergence analyses are also carried out. Finally, some numerical experiments on a set of test problems related to constrained and unconstrained testing environment, are conducted using a well-known performance profile.

2010 *Mathematics subject classification*: primary 90C53; secondary 49M37, 65K05.

*Keywords and phrases*: unconstrained optimization, conjugate gradient method, matrix norm, Dai–Liao parameter, condition number.

## 1. Introduction

We consider the unconstrained optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x}), \tag{1.1}$$

in which $f:\mathbb{R}^n\to\mathbb{R}$ is continuously differentiable and an analytic expression of its gradient is available. As a class of efficient techniques for solving the large-scale cases of problem (1.1), conjugate gradient (CG) methods are useful tools due to low memory requirements and strong global convergence properties [7, 19]. These methods are specified by iterative formulas in the form

$$\mathbf{x}_0\in\mathbb{R}^n,\quad \mathbf{x}_{k+1}=\mathbf{x}_k+\mathbf{s}_k,\quad \mathbf{s}_k=\alpha_k\mathbf{d}_k,\quad k=0,1,\ldots,$$

where $\alpha_k$ is a step length to be computed by a line search along the direction $\mathbf{d}_k$, defined by

$$\mathbf{d}_0=-\mathbf{g}_0,\quad \mathbf{d}_{k+1}=-\mathbf{g}_{k+1}+\beta_k\mathbf{d}_k,\quad k=0,1,\ldots,$$

in which $\mathbf{g}_k=\nabla f(\mathbf{x}_k)$ and $\beta_k$ is a scalar called the CG (update) parameter [15].

[1]Department of Mathematics, Faculty of Mathematics, Statistics and Computer Science, Semnan University, PO Box 35195-363, Semnan, Iran; e-mail: aminisor@semnan.ac.ir, sbk@semnan.ac.ir.

Employing the features of the quasi-Newton methods [21], Dai and Liao [9] (DL) proposed one of the well-known CG parameters

$$\beta_k^{DL} = \frac{\mathbf{g}_{k+1}^T \mathbf{y}_k}{\mathbf{d}_k^T \mathbf{y}_k} - t\frac{\mathbf{g}_{k+1}^T \mathbf{s}_k}{\mathbf{d}_k^T \mathbf{y}_k}, \tag{1.2}$$

where $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$ and $t$ is a nonnegative parameter, being an extension of the Hestenes–Stiefel parameter [16]. Note that if the line search fulfils the popular strong Wolfe conditions [19]

$$f(\mathbf{x}_k + \alpha_k \mathbf{d}_k) - f(\mathbf{x}_k) \leq \delta \alpha_k \mathbf{d}_k^T \mathbf{g}_k, \tag{1.3}$$

$$|\mathbf{d}_k^T \nabla f(\mathbf{x}_k + \alpha_k \mathbf{d}_k)| \leq -\sigma \mathbf{d}_k^T \mathbf{g}_k \tag{1.4}$$

with $0 < \delta < \sigma < 1$, then $\mathbf{d}_k^T \mathbf{y}_k > 0$ and so the parameter in (1.2) is well defined. We observe that the efficient CG methods proposed by Hager and Zhang [13] and Dai and Kou [8] can be regarded as adaptive versions of the DL method satisfying the sufficient descent property. However, numerical behaviour of the this method is very dependent on the parameter $t$, for which finding appropriate choices has attracted special attention, even considered as an open problem [1].

Recently, based on a matrix point of view, Babaie-Kafaki and Ghanbari [2–5] proposed several adaptive choices for the parameter $t$ in (1.2) (see also the article by Fatemi [11]). More precisely, they noted that search directions of the DL method can be written as

$$\mathbf{d}_{k+1} = -\mathbf{Q}_{k+1}\mathbf{g}_{k+1}, \quad k = 0, 1, \ldots,$$

in which

$$\mathbf{Q}_{k+1} = \mathbf{I} - \frac{\mathbf{s}_k \mathbf{y}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} + t\frac{\mathbf{s}_k \mathbf{s}_k^T}{\mathbf{s}_k^T \mathbf{y}_k} \tag{1.5}$$

is called the search direction matrix. Then they suggested several choices for $t$ by minimizing some upper bounds for the condition number of $\mathbf{Q}_{k+1}$ [2, 4] as well as approaching $\mathbf{Q}_{k+1}$ to some well-known matrices (such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) updating formula) as much as possible [4, 5]. Moreover, symmetrizing the matrix $\mathbf{Q}_{k+1}$, Babaie-Kafaki and Ghanbari showed that the following family of choices for $t$ guarantees the descent property

$$t_k^{p,q} = p\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T \mathbf{y}_k} - q\frac{\mathbf{s}_k^T \mathbf{y}_k}{\|\mathbf{s}_k\|^2}, \tag{1.6}$$

where $p \geq 1/4$ and $q \leq 0$ are real parameters and $\|\cdot\|$ denotes the Euclidean norm [3].

Here, conducting matrix analyses on the search direction matrix of the DL method, we propose several choices for the parameter $t$ in (1.2). More precisely, in Section 2 we suggest two adaptive choices for the DL parameter by minimizing the $\ell_1$ and $\ell_\infty$ norm condition numbers of $\mathbf{Q}_{k+1}$. In Section 3 we discuss optimality of the formula (1.6) with appropriate choices for the parameters $p$ and $q$ in the sense of minimizing the spectral condition number as well as the measure function of Byrd and Nocedal [6] for

https://doi.org/10.1017/S1446181119000063 Published online by Cambridge University Press

a symmetrized version of $\mathbf{Q}_{k+1}$. In Section 4 we make some comparative numerical experiments and report the results. Finally, we present the concluding remarks in Section 5. Hereafter, we assume that the strong Wolfe line search conditions (1.3) and (1.4) are satisfied.

## 2. Two adaptive choices for the Dai–Liao parameter based on the $\ell_1$ and $\ell_\infty$ norm condition numbers

First, we briefly present some properties of the two nonsmooth matrix norms $\ell_1$ and $\ell_\infty$. Then we obtain an optimal choice for the DL parameter by minimizing the $\ell_1$ norm condition number of the search direction matrix $\mathbf{Q}_{k+1}$ given by (1.5). In a similar scheme, an optimal choice for $t$ is given by minimizing the $\ell_\infty$ norm condition number of $\mathbf{Q}_{k+1}$.

As is well known, for a matrix $\mathbf{A} = [a_{ij}] \in \mathbb{R}^{n\times n}$,

$$\|\mathbf{A}\|_1 = \max_{j=1,2,\ldots,n} \Sigma_{i=1}^{n} |a_{ij}|, \qquad \|\mathbf{A}\|_\infty = \max_{i=1,2,\ldots,n} \Sigma_{j=1}^{n} |a_{ij}|.$$

Hence, if $\mathbf{A} = \mathbf{x}\mathbf{y}^T$ with the arbitrary vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, then

$$\|\mathbf{A}\|_1 = \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty, \qquad \|\mathbf{A}\|_\infty = \|\mathbf{x}\|_\infty\|\mathbf{y}\|_1.$$

When $\mathbf{A}$ is nonsingular, we denote its $\ell_1$ and $\ell_\infty$ norm condition numbers [20], respectively, by

$$\kappa_1(\mathbf{A}) = \|\mathbf{A}\|_1\|\mathbf{A}^{-1}\|_1 \quad \text{and} \quad \kappa_\infty(\mathbf{A}) = \|\mathbf{A}\|_\infty\|\mathbf{A}^{-1}\|_\infty.$$

These are effective tools to measure sensitivity of the solution to data perturbations in matrix-based numerical problems [20]. A matrix with a large condition number is called ill-conditioned, since instability may occur in the corresponding matrix computations. Considering this fact, here we determine the parameter $t$ in (1.2) to minimize an upper bound on the $\ell_1$ and $\ell_\infty$ norm condition numbers of the matrix $\mathbf{Q}_{k+1}$, making the matrix well-conditioned.

Based on the above discussion,

$$\|\mathbf{Q}_{k+1}\|_1 \le 1 + \frac{\|\mathbf{s}_k\|_1\|\mathbf{y}_k\|_\infty}{\mathbf{s}_k^T\mathbf{y}_k} + t\frac{\|\mathbf{s}_k\|_1\|\mathbf{s}_k\|_\infty}{\mathbf{s}_k^T\mathbf{y}_k}.$$

Also, from the Sherman–Morrison formula [19],

$$\mathbf{Q}_{k+1}^{-1} = \mathbf{I} + \frac{1}{t}\frac{\mathbf{s}_k\mathbf{y}_k^T}{\|\mathbf{s}_k\|^2} - \frac{\mathbf{s}_k\mathbf{s}_k^T}{\|\mathbf{s}_k\|^2}$$

and, therefore,

$$\|\mathbf{Q}_{k+1}^{-1}\|_1 \le 1 + \frac{1}{t}\frac{\|\mathbf{s}_k\|_1\|\mathbf{y}_k\|_\infty}{\|\mathbf{s}_k\|^2} + \frac{\|\mathbf{s}_k\|_1\|\mathbf{s}_k\|_\infty}{\|\mathbf{s}_k\|^2}.$$

Hence,

$$\kappa_1(\mathbf{Q}_{k+1}) \le h(t) + \zeta, \tag{2.1}$$

https://doi.org/10.1017/S1446181119000063 Published online by Cambridge University Press

where $\zeta$ is a positive constant and

$$h(t) = \frac{1}{t}\frac{\|\mathbf{s}_k\|_1\|\mathbf{y}_k\|_\infty}{\|\mathbf{s}_k\|^2}\left(1 + \frac{\|\mathbf{s}_k\|_1\|\mathbf{y}_k\|_\infty}{\mathbf{s}_k^T\mathbf{y}_k}\right) + t\frac{\|\mathbf{s}_k\|_1\|\mathbf{s}_k\|_\infty}{\mathbf{s}_k^T\mathbf{y}_k}\left(1 + \frac{\|\mathbf{s}_k\|_1\|\mathbf{s}_k\|_\infty}{\|\mathbf{s}_k\|^2}\right), \quad t > 0.$$

Then, after some algebraic manipulations,

$$t_{k_1}^* = \arg\min_t h(t) = \sqrt{\frac{\|\mathbf{y}_k\|_\infty}{\|\mathbf{s}_k\|_\infty}\left(\frac{\mathbf{s}_k^T\mathbf{y}_k + \|\mathbf{s}_k\|_1\|\mathbf{y}_k\|_\infty}{\|\mathbf{s}_k\|^2 + \|\mathbf{s}_k\|_1\|\mathbf{s}_k\|_\infty}\right)} \tag{2.2}$$

is a minimizer of the upper bound of the $\ell_1$ norm condition number given by (2.1). Also, if we conduct a similar analysis using the $\ell_\infty$ norm, then

$$t_{k_2}^* = \sqrt{\frac{\|\mathbf{y}_k\|_1}{\|\mathbf{s}_k\|_1}\left(\frac{\mathbf{s}_k^T\mathbf{y}_k + \|\mathbf{s}_k\|_\infty\|\mathbf{y}_k\|_1}{\|\mathbf{s}_k\|^2 + \|\mathbf{s}_k\|_\infty\|\mathbf{s}_k\|_1}\right)} \tag{2.3}$$

as a minimizer of an upper bound of $\kappa_\infty(\mathbf{Q}_{k+1})$.

Now assume that the level set $\mathcal{L} = \{\mathbf{x} \mid f(\mathbf{x}) \le f(\mathbf{x}_0)\}$ is bounded and also, in a neighbourhood $\mathcal{N}$ of $\mathcal{L}$, $f$ is continuously differentiable and its gradient is Lipschitz continuous. Consider a CG method with the parameter $\beta_k^{DL}$ defined by (1.2) in which for all $k \ge 0$, $t = t_{k_1}^*$ or $t = t_{k_2}^*$ as defined by (2.2) or (2.3), respectively. It can be shown that there exists a positive constant $M$ as an upper bound of $t_{k_1}^*$ and $t_{k_2}^*$. Hence, if the search directions are descent directions and the step lengths are determined to satisfy the strong Wolfe conditions (1.3) and (1.4), then a result of Dai and Liao [9, Theorem 3.3] ensures the global convergence of the method for uniformly convex objective functions. Furthermore, if the CG parameter (1.2) is truncated as

$$\beta_k^{DL+} = \max\left\{\frac{\mathbf{g}_{k+1}^T\mathbf{y}_k}{\mathbf{d}_k^T\mathbf{y}_k}, 0\right\} - t\frac{\mathbf{g}_{k+1}^T\mathbf{s}_k}{\mathbf{d}_k^T\mathbf{y}_k}$$

with $t = t_{k_1}^*$ or $t = t_{k_2}^*$, and the search directions satisfy the sufficient descent condition, then another result of Dai and Liao [9, Theorem 3.6] ensures the global convergence of the method for general objective functions.

## 3. On optimality of the formula (1.6)

In this section we show that appropriate choices for the parameters $p$ and $q$ in (1.6) may enhance numerical stability of the DL method in the sense of decreasing the condition number of a symmetrized version of $\mathbf{Q}_{k+1}$, that is,

$$\mathbf{A}_{k+1} = \frac{\mathbf{Q}_{k+1} + \mathbf{Q}_{k+1}^T}{2} = \mathbf{I} - \frac{1}{2}\frac{\mathbf{s}_k\mathbf{y}_k^T + \mathbf{y}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k} + t\frac{\mathbf{s}_k\mathbf{s}_k^T}{\mathbf{s}_k^T\mathbf{y}_k},$$

which is shown to be positive definite when

$$t > \frac{1}{4}\left(\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k} - \frac{\mathbf{s}_k^T\mathbf{y}_k}{\|\mathbf{s}_k\|^2}\right), \tag{3.1}$$

guaranteeing the descent property [3].

https://doi.org/10.1017/S1446181119000063 Published online by Cambridge University Press

As shown by Babaie-Kafaki and Ghanbar [3], the matrix $\mathbf{A}_{k+1}$ has $n-2$ eigenvalues being equal to 1 and, also,

$$\operatorname{tr}(\mathbf{A}_{k+1}) = n - 1 + t\frac{\|\mathbf{s}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}, \tag{3.2}$$

$$\det(\mathbf{A}_{k+1}) = \frac{1}{4} + t\frac{\|\mathbf{s}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k} - \frac{1}{4}\frac{\|\mathbf{s}_k\|^2\|\mathbf{y}_k\|^2}{(\mathbf{s}_k^T\mathbf{y}_k)^2}, \tag{3.3}$$

where $\operatorname{tr}(\cdot)$ and $\det(\cdot)$ stand for the trace and determinant of a matrix, respectively. So, after some algebraic manipulations, the two other eigenvalues $\lambda_k^+$ and $\lambda_k^-$ of $\mathbf{A}_{k+1}$ are computed as

$$\lambda_k^{\pm} = \frac{1}{2}\left(1 + t\frac{\|\mathbf{s}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k}\right) \pm \frac{1}{2}\sqrt{\left(t\frac{\|\mathbf{s}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k} - 1\right)^2 + \frac{\|\mathbf{s}_k\|^2\|\mathbf{y}_k\|^2}{(\mathbf{s}_k^T\mathbf{y}_k)^2} - 1},$$

for which, when inequality (3.1) holds, it follows that $0 < \lambda_k^- \le 1 \le \lambda_k^+$. Thus,

$$\kappa(\mathbf{A}_{k+1}) = \frac{\lambda_k^+}{\lambda_k^-},$$

where $\kappa(\cdot)$ stands for the spectral condition number. Now, minimizing $\kappa(\mathbf{A}_{k+1})$,

$$t_{k_3}^* = \arg\min_t \kappa(\mathbf{A}_{k+1}) = \frac{1}{2}\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k} + \frac{1}{2}\frac{\mathbf{s}_k^T\mathbf{y}_k}{\|\mathbf{s}_k\|^2}, \tag{3.4}$$

ensuring optimality of the choice $(p, q) = (1/2, -1/2)$ in equation (1.6). In what follows, based on a different strategy, we achieve another optimal choice for $(p, q)$.

In an analysis on the convergence of the quasi-Newton methods, Byrd and Nocedal [6] introduced the following function on the set of positive-definite matrices:

$$\psi(\mathbf{A}) = \operatorname{tr}(\mathbf{A}) - \ln(\det(\mathbf{A})),$$

where $\mathbf{A} \in \mathbb{R}^{n\times n}$ is an arbitrary symmetric matrix with the eigenvalues $\lambda_1 \ge \cdots \ge \lambda_n > 0$. The definition of $\psi(\cdot)$ originates from the work of Powell [18], in which the trace and determinant of the Hessian approximation given by the quasi-Newton methods have been combined using the arithmetic–geometric mean inequality. Note that $\psi(\mathbf{A}) > 0$, because

$$\psi(\mathbf{A}) = \sum_{i=1}^{n}(\lambda_i - \ln\lambda_i).$$

Note that for all $z > 0$, the function $\omega(z) = z - \ln z$ is a strictly convex function with the minimum value 1 at $z = 1$. So, $\psi(\mathbf{A}) \ge n$ and, as pointed out by Byrd and Nocedal [6], $\psi(\mathbf{A})$ can be considered as a measure of closeness of the matrix $\mathbf{A}$ to the identity matrix for which $\psi(\mathbf{I}) = n$. Furthermore, $\omega(z) > \ln z$; thus,

$$\psi(\mathbf{A}) \ge \ln\lambda_1 - \ln\lambda_n = \ln\frac{\lambda_1}{\lambda_n} = \ln\kappa(\mathbf{A}),$$

which shows that $\psi(\mathbf{A})$ is large when $\mathbf{A}$ is an ill-conditioned matrix. Hence, small values of $\psi(\mathbf{A})$ are favourable in actual computations. Based on this fact and

https://doi.org/10.1017/S1446181119000063 Published online by Cambridge University Press

Table 1. Test problems data.

| Function | $n$ | Function | $n$ | Function | $n$ |
|---|---|---|---|---|---|
| ARGLINA | 200 | DIXMAANK | 3000 | MANCINO | 100 |
| BDEXP | 5000 | DIXMAANL | 3000 | MOREBV | 5000 |
| BIGGSB1 | 5000 | DIXON3DQ | 10 000 | MSQRTALS | 1024 |
| BQPGABIM | 50 | DMN15103 | 99 | MSQRTBLS | 1024 |
| BQPGASIM | 50 | DQDRTIC | 5000 | NCB20 | 5010 |
| BROYDN7D | 5000 | DQRTIC | 5000 | NCB20B | 5000 |
| BRYBND | 5000 | DRCAV1LQ | 4489 | NONCVXU2 | 5000 |
| CHAINWOO | 4000 | DRCAV2LQ | 4489 | NONDQUAR | 5000 |
| CHENHARK | 5000 | DRCAV3LQ | 4489 | PENALTY2 | 200 |
| CHNROSNB | 50 | EDENSCH | 2000 | POWELLSG | 5000 |
| CLPLATEB | 5041 | EG2 | 1000 | POWER | 10 000 |
| COSINE | 10 000 | EIGENALS | 2550 | QUARTC | 5000 |
| CRAGGLVY | 5000 | EIGENBLS | 2550 | SCHMVETT | 5000 |
| CURLY10 | 10 000 | EIGENCLS | 2652 | SENSORS | 100 |
| CURLY20 | 10 000 | ENGVAL1 | 5000 | SINQUAD | 5000 |
| CURLY30 | 10 000 | ERRINROS | 50 | SPARSQUR | 10 000 |
| DECONVU | 63 | EXTROSNB | 1000 | SPMSRTLS | 4999 |
| DIXMAANA | 3000 | FLETCBV2 | 5000 | SROSENBR | 5000 |
| DIXMAANB | 3000 | FLETCBV3 | 5000 | TESTQUAD | 5000 |
| DIXMAANC | 3000 | FLETCHBV | 5000 | TOINTGOR | 50 |
| DIXMAAND | 3000 | FLETCHCR | 1000 | TOINTGSS | 5000 |
| DIXMAANE | 3000 | FMINSRF2 | 5625 | TOINTPSP | 50 |
| DIXMAANF | 3000 | FMINSURF | 5625 | TOINTQOR | 50 |
| DIXMAANG | 3000 | FREUROTH | 5000 | TRIDIA | 5000 |
| DIXMAANH | 3000 | GENHUMPS | 5000 | VARDIM | 200 |
| DIXMAANI | 3000 | GENROSE | 500 | VAREIGVL | 50 |
| DIXMAANJ | 3000 | LIARWHD | 5000 | WOODS | 4000 |

considering equations (3.2) and (3.3), after some algebraic manipulations,

$$t_{k_4}^* = \arg\min_t \psi(\mathbf{A}_{k+1}) = \frac{1}{4}\frac{\|\mathbf{y}_k\|^2}{\mathbf{s}_k^T\mathbf{y}_k} + \frac{3}{4}\frac{\mathbf{s}_k^T\mathbf{y}_k}{\|\mathbf{s}_k\|^2}. \tag{3.5}$$

Thus, the choice $(p, q) = (1/4, -3/4)$ in (1.6) is optimal in the sense of minimizing the measure function of Byrd and Nocedal [6].

## 4. Numerical experiments

Here, we computationally compare performance of the DL method with the adaptive choices $t_{k_1}^*$, $t_{k_2}^*$, $t_{k_3}^*$ and $t_{k_4}^*$ for the parameter $t$ in (1.2), given by (2.2), (2.3),

https://doi.org/10.1017/S1446181119000063 Published online by Cambridge University Press

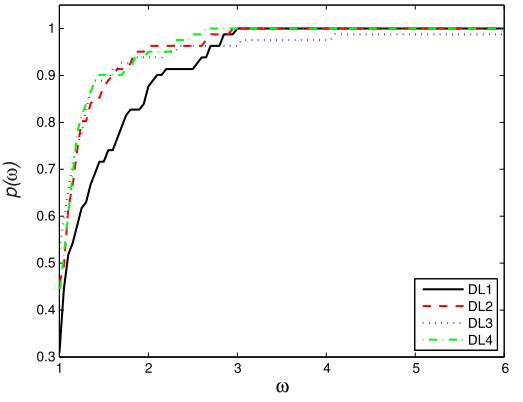


FIGURE 1. Total number of function and gradient evaluations performance profiles.

(3.4) and (3.5), respectively; the corresponding methods are respectively called DL$i$, $i = 1, 2, 3, 4$. The runs were performed on a set of 81 unconstrained optimization test problems of the CUTEr collection [12] with the minimum dimension being equal to 50, as given in Table 1. For our experiments, we used a computer with a 2.3 GHz Intel (R) CORE (TM) 2 Duo CPU, and 8 GB of RAM. We implemented the method in MATLAB 7.7.0.471 (R2008b) running on a Centos 6.2 server Linux operating system.

In the line search procedure, the strong Wolfe conditions (1.3) and (1.4) have been employed using an algorithm of Nocedal and Wright [17, Algorithm 3.5] with $\delta = 0.0001$ and $\sigma = 0.9$. Since DL1 and DL2 are not necessarily descent methods, in these cases, when an uphill search direction was encountered, the negative gradient was used [9].

The algorithms were stopped by reaching a maximum of 10 000 iterations or achieving a solution with $\|\mathbf{g}_k\| < 10^{-6}(1 + |f(\mathbf{x}_k)|)$. Moreover, efficiency comparisons were drawn using the Dolan–Moré performance profile [10] on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$, where $N_f$ and $N_g$ respectively denote the numbers of function and gradient evaluations [14]. The performance profile gives, for every $\omega \geq 1$, the proportion $p(\omega)$ of the test problems that each considered algorithmic variant has a performance within a factor of $\omega$ of the best. Figures 1 and 2 demonstrate the results of comparisons.

As the figures show, the methods DL2, DL3 and DL4 are more likely competitive with each other and, also, they outperform the method DL1. Hence, formula (2.2) fails to introduce an appropriate choice for the DL parameter in contrast to the others. Also, note that DL3 and DL4 are slightly preferable to DL2 with respect to the CPU time. Thus, proper choices for the parameters $p$ and $q$ in (1.6) (previously proposed in [3]) may enhance efficiency of the DL method.

https://doi.org/10.1017/S1446181119000063 Published online by Cambridge University Press

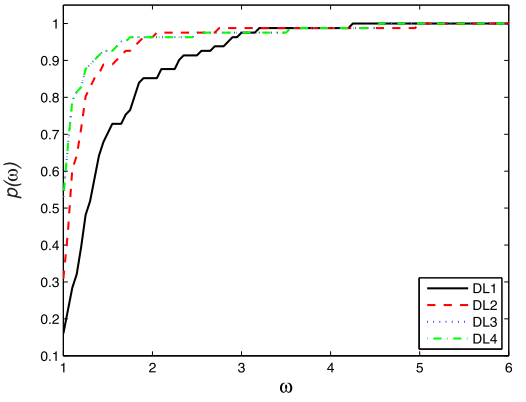


FIGURE 2. CPU time performance profiles.

## 5. Conclusions

As an open problem in nonlinear conjugate gradient methods, finding proper choices for the Dai–Liao parameter has attracted special attention. Here, based on matrix analyses, we have suggested several adaptive choices for this parameter. More precisely, we have focused on improving the condition number in different matrix norms as well as minimizing a well-known measure function.

Numerical experiments have been conducted to compare the effectiveness of the proposed choices of the Dai–Liao parameter.

## Acknowledgements

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

## References

[1] N. Andrei, "Open problems in conjugate gradient algorithms for unconstrained optimization", *Bull. Malays. Math. Sci. Soc.* **34** (2011) 319–330; http://emis.ams.org/journals/BMMSS/pdf/v34n2/v34n2p11.pdf.

[2] S. Babaie-Kafaki and R. Ghanbari, "The Dai–Liao nonlinear conjugate gradient method with optimal parameter choices", *European J. Oper. Res.* **234** (2014) 625–630; doi:10.1016/j.ejor.2013.11.012.

[3] S. Babaie-Kafaki and R. Ghanbari, "A descent family of Dai–Liao conjugate gradient methods", *Optim. Methods Softw.* **29** (2014) 583–591; doi:10.1080/10556788.2013.833199.

[4] S. Babaie-Kafaki and R. Ghanbari, "Two optimal Dai–Liao conjugate gradient methods", *Optimization* **64** (2015) 2277–2287; doi:10.1080/02331934.2014.938072.

[5] S. Babaie-Kafaki and R. Ghanbari, "A class of adaptive Dai–Liao conjugate gradient methods based on the scaled memoryless BFGS update", *4OR* **15** (2017) 85–92; doi:10.1007/s10288-016-0323-1.

[6] R. H. Byrd and J. Nocedal, "A tool for the analysis of quasi-Newton methods with application to unconstrained minimization", *SIAM J. Numer. Anal.* **26** (1989) 727–739; doi:10.1137/0726042.
[7] Y. H. Dai, J. Y. Han, G. H. Liu, D. F. Sun, H. X. Yin and Y. X. Yuan, "Convergence properties of nonlinear conjugate gradient methods", *SIAM J. Optim.* **10** (1999) 348–358; doi:10.1137/S1052623494268443.
[8] Y. H. Dai and 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; doi:10.1137/100813026.
[9] Y. H. Dai and L. Z. Liao, "New conjugacy conditions and related nonlinear conjugate gradient methods", *Appl. Math. Optim.* **43** (2001) 87–101; doi:10.1007/s002450010019.
[10] E. D. Dolan and J. J. Moré, "Benchmarking optimization software with performance profiles", *Math. Program. Ser. A* **91**(2) (2002) 201–213; doi:10.1007/s101070100263.
[11] M. Fatemi, "An optimal parameter for Dai–Liao family of conjugate gradient methods", *J. Optim. Theory Appl.* **169** (2016) 587–605; doi:10.1007/s10957-015-0786-9.
[12] N. I. M. Gould, D. Orban and P. L. Toint, "CUTEr: a constrained and unconstrained testing environment, revisited", *ACM Trans. Math. Software* **29** (2003) 373–394; doi:10.1145/962437.962439.
[13] W. W. Hager and H. Zhang, "A new conjugate gradient method with guaranteed descent and an efficient line search", *SIAM J. Optim.* **16** (2005) 170–192; doi:10.1137/030601880.
[14] W. W. Hager and H. Zhang, "Algorithm 851: CG-descent, a conjugate gradient method with guaranteed descent", *ACM Trans. Math. Software* **32** (2006) 113–137; doi:10.1145/1132973.1132979.
[15] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods", *Pac. J. Optim.* **2**(1) (2006) 35–58; https://www.caam.rice.edu/~yzhang/caam554/pdf/cgsurvey.pdf.
[16] M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems", *J. Res. Nat. Bur. Stand.* **49** (1952) 409–436; doi:10.6028/jres.049.044.
[17] J. Nocedal and S. J. Wright, *Numerical optimization*, *Springer Ser. Oper. Res. Financ. Eng.* (Springer, New York, 2006); doi:10.1007/978-0-387-40065-5.
[18] M. J. D. Powell, "Some global convergence properties of a variable metric algorithm for minimization without exact line searches", in: *Nonlinear programming*, Volume 9 of *SIAM–AMS Proc.* (eds R. W. Cottle and C. E. Lemke), (American Mathematical Society, Providence, RI, 1976) 53–72.
[19] W. Sun and Y. X. Yuan, *Optimization theory and methods: nonlinear programming* (Springer, New York, 2006); https://www.springer.com/gp/book/9780387249759.
[20] D. S. Watkins, *Fundamentals of matrix computations* (John Wiley, New York, 2002); https://www.wiley.com/enus/Fundamentals+of+Matrix+Computations%2C+3rd+Edition-p-9780470528334.
[21] C. Xu and J. Z. Zhang, "A survey of quasi-Newton equations and quasi-Newton methods for optimization", *Ann. Oper. Res.* **103** (2001) 213–234; doi:10.1023/A:1012959223138.