

A SURVEY ON THE DAI–LIAO FAMILY OF NONLINEAR CONJUGATE GRADIENT METHODS

SAMAN BABAIE–KAFKI* 

Abstract. At the beginning of this century, which is characterized by huge flows of emerging data, Dai and Liao proposed a pervasive conjugacy condition that triggered the interest of many optimization scholars. Recognized as a sophisticated conjugate gradient (CG) algorithm after about two decades, here we share our visions and thoughts on the method in the framework of a review study. In this regard, we first discuss the modified Dai–Liao methods based on the modified secant equations given in the literature, mostly with the aim of applying the objective function values in addition to the gradient information. Then, several adaptive, sort of optimal choices for the parameter of the method are studied. Especially, we devote a part of our study to the modified versions of the Hager–Zhang and Dai–Kou CG algorithms, being well-known members of the Dai–Liao class of CG methods. Extensions of the classical CG methods based on the Dai–Liao approach are also reviewed. Finally, we discuss the optimization models of practical disciplines that have been addressed by the Dai–Liao approach, including the nonlinear systems of equations, image restoration and compressed sensing.

Mathematics Subject Classification. 90C53, 49M37.

Received August 29, 2022. Accepted December 13, 2022.

1. INTRODUCTION

It is needless to say that in the contemporary world we are grappling with a level of huge data that makes the modeling a critical task. Large data sets that we are inundated and swarmed with need efficient memoryless approaches to be handled, in a way that we do not carry on their complexity.

In a wide range of practical disciplines such as machine learning and signal processing, large scale continuous optimization models emerge, often as the unconstrained optimization problem

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

where the objective function f is here assumed to be smooth. Scholar studies reflect the value of the CG techniques among the various continuous optimization algorithms [95]. Initially founded by Hestenes and Stiefel (HS) [59] in the midst of the previous century for solving positive definite systems of linear equations, and then adopted by Fletcher and Reeves [51] for the unconstrained optimization, CG algorithms benefit the low memory storage and simple iterations squarely as well as the second order information implicitly [57, 83].

Keywords. Unconstrained optimization, large scale optimization, conjugate gradient algorithm, Dai–Liao method, secant equation, quasi-Newton update.

Department of Mathematics, Semnan University, Semnan, Iran.

*Corresponding author: sbk@semnan.ac.ir

Successive approximations generated by the CG algorithms are in the form of

$$x_{k+1} = x_k + s_k, \quad k = 0, 1, \dots, \quad (1.2)$$

starting from a given point $x_0 \in \mathbb{R}^n$, where $s_k = \alpha_k d_k$, in which $\alpha_k > 0$ is a step length determined by a line search along the CG direction d_k , often defined by

$$d_0 = -g_0, \quad d_{k+1} = -g_{k+1} + \beta_k d_k, \quad k = 0, 1, \dots, \quad (1.3)$$

where β_k is the CG parameter and $g_k = \nabla f(x_k)$ [97]. There is a big bunch of formulas for β_k with completely different computational outcomes [8, 16, 57].

In the line search procedure of the CG algorithms, polynomial interpolation schemes are mostly used with the Wolfe conditions as the stopping criterion [86]. The Wolfe conditions simultaneously consist of the Armijo condition,

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k, \quad (1.4)$$

and the curvature condition,

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq -\sigma g_k^T d_k, \quad (1.5)$$

with $0 < \delta < \sigma < 1$, for a descent direction d_k . Meanwhile, consisting of (1.4) together with the following (to some extent) strict version of (1.5),

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

the strong Wolfe conditions are widely used to give more exact choices for α_k . Moreover, modified versions of the Wolfe conditions have been developed for the CG algorithms [114, 117]. It is also worth noting that sometimes backtracking line search schemes are also used in the CG algorithms, based upon the Armijo type conditions [121, 122].

As known, conjugacy conditions are the basis of the CG algorithms. The linear CG methods, *i.e.* the CG methods for minimizing a strictly convex quadratic function with the Hessian $G \in \mathbb{R}^{n \times n}$ by the exact line search, or equivalently, the CG methods for solving a system of linear equations with the positive definite coefficient matrix G , generate a sequence of the search directions $\{d_k\}_{k \geq 0}$ such that the following basic conjugacy condition holds [97]:

$$d_i^T G d_j = 0, \quad \forall i \neq j.$$

However, for general nonlinear functions, mean value theorem ensures that there exists some $\xi_k \in (0, 1)$ such that

$$d_{k+1}^T y_k = \alpha_k d_{k+1}^T \nabla^2 f(x_k + \xi_k \alpha_k d_k) d_k,$$

where $y_k = g_{k+1} - g_k$. Hence, the equation

$$d_{k+1}^T y_k = 0, \quad (1.6)$$

can be regarded as a conjugacy condition for general situations, which together with (1.3) yields the HS parameter:

$$\beta_k^{\text{HS}} = \frac{g_{k+1}^T y_k}{d_k^T y_k}.$$

Note that by the Wolfe line search conditions we have $d_k^T y_k > 0$, and so, β_k^{HS} is well-defined.

At the beginning of the current century, Dai and Liao (DL) [40] put forward a conjugacy condition which gave rise to a great one-parameter class of CG algorithms. Their approach has been established upon the quasi-Newton (QN) aspects. As known, QN iterations are in the form of (1.2) in which the search direction d_{k+1} is computed as a solution of the linear system $B_{k+1} d = -g_{k+1}$ where $B_{k+1} \in \mathbb{R}^{n \times n}$ is a symmetric (and often

positive definite) approximation of $\nabla^2 f(x_{k+1})$ [97]. The successive Hessian approximations $\{B_k\}_{k \geq 0}$ in the QN algorithms are classically updated based on the standard secant equation; that is

$$B_{k+1}s_k = y_k. \quad (1.7)$$

Now, from the algorithmic features of the QN methods, we can write

$$d_{k+1}^T y_k = d_{k+1}^T B_{k+1} s_k = -g_{k+1}^T s_k, \quad (1.8)$$

being another conjugacy condition which reduces to (1.6) under the exact line search. In an effective extension scheme, Dai and Liao [40] embedded a parameter $t \geq 0$ on (1.8) and suggested the following hybrid conjugacy condition:

$$d_{k+1}^T y_k = -t g_{k+1}^T s_k, \quad (1.9)$$

which together with (1.3) yields

$$\beta_k^{\text{DL}} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}. \quad (1.10)$$

It is worth mentioning that when $t = 0$ or the line search is performed exactly, then (1.9) reduces to (1.6). Alternatively, for $t = 1$, when the line search is approximately exact in the sense of $g_{k+1}^T d_k \approx 0$, the equation (1.9) can be regarded as a conjugacy condition that implicitly meets the QN features. After introducing (1.10), Dai and Liao [40] showed that an iterative method of the form (1.2)–(1.3) with β_k^{DL} as the CG parameter is globally convergent for the uniformly (strongly) convex functions in (1.1). Then, to establish convergence for general functions, based on the analysis of [54], they proposed the following modified version of β_k^{DL} :

$$\beta_k^{\text{DL}+} = \max \left\{ \frac{g_{k+1}^T y_k}{d_k^T y_k}, 0 \right\} - t \frac{g_{k+1}^T s_k}{d_k^T y_k}, \quad t \geq 0. \quad (1.11)$$

Many researchers volunteered their efforts to study different aspects of the DL method so that now it can be regarded as a sophisticated CG algorithm. Here, we plan to share our vision of the DL method in the framework of a review study. We classify our review into the following sections. Firstly, in Section 2, we discuss the studies in which by employing the modified secant equations [106] in (1.9), modified versions of the DL method have been proposed. Several optimal adaptive choices for the DL parameter t are discussed in Section 3. As well-known members of the DL class of CG algorithms, Section 4 is devoted to modified versions of the Hager–Zhang [55,56] and Dai–Kou [42] methods. Considering the similar transition from β_k^{HS} to β_k^{DL} , several DL type extended versions of the classical CG parameters are studied in Section 5. To discuss practical applications of the method, in Section 6 we review some researches that have addressed the nonlinear systems of equations, image restoration and compressed sensing by the DL method. Ultimately, the concluding remarks are given in Section 7.

2. IMPROVED DAI-LIAO METHODS BASED ON THE MODIFIED SECANT EQUATIONS

As the results of conscious efforts to enhance validity and reliability of the QN approximations of the Hessian, modified versions of the standard secant equation (1.7) have been suggested in the literature. A holistic model of several essential modified secant equations has been given in [13] as follows:

$$B_{k+1}s_k = z_k, \quad z_k = y_k + \xi \frac{\theta_k}{s_k^T u_k} u_k + C \|g_k\|^r s_k, \quad (2.1)$$

in which ξ , C and r are nonnegative constants, $u_k \in \mathbb{R}^n$ is a vector parameter satisfying $s_k^T u_k \neq 0$, $\|\cdot\|$ represents the Euclidean norm, and

$$\theta_k = 2(f_k - f_{k+1}) + s_k^T (g_k + g_{k+1}). \quad (2.2)$$

A review of the literature reveals that the vector parameter u_k is often set as $u_k = s_k$, or, under the Wolfe line search conditions, $u_k = y_k$. Now, to see how the classical (modified) secant equations are special cases of (2.1), firstly note that if $\xi = C = 0$, then (2.1) reduces to (1.7). Alternatively, when $C = 0$, the equation (2.1) yields the modified secant equations proposed by Wei *et al.* [104] for $\xi = 1$, Biglari *et al.* [32] for $\xi = 2$, and Zhang *et al.* [119, 120] for $\xi = 3$. All the equations given in [32, 104, 119, 120] have been obtained based on the Taylor expansion with the aims of enhancing the accuracy of the Hessian approximation as well as benefiting the available function values in addition to the gradient information. It can be (loosely) stated that order of the accuracy of the mentioned modified secant equations gradually increases by the growth of ξ . For $\xi = 0$, the equation (2.1) reduces to the modified secant equation proposed by Li and Fukushima [67, 68] which is capable to guarantee global convergence of the QN methods needless to the convexity assumptions. Meantime, multi-step secant equations have been developed by Ford and Moghrabi [52], utilizing the information available from more than one previous iteration.

Now, taking (1.8) into consideration, it can be observed that all the mentioned modified secant equations can be attached to the DL approach, as a measure to benefit the mentioned merits of the modified secant equations. In this context, Li *et al.* [70] used a version of the modified secant equation proposed by Wei *et al.* [104] in the sense of imposing a nonnegative restriction on θ_k given by (2.2). More exactly, they let

$$\theta_k \leftarrow \theta_k^+ = \max\{\theta_k, 0\}, \quad (2.3)$$

simply ensuring positiveness of $s_k^T z_k$ under the Wolfe conditions which is crucial both in the theoretical and numerical viewpoints. Yabe and Takano [107] employed the modified secant equation suggested by Zhang and Xu [119] to get another modified DL method. Babaie-Kafaki *et al.* [30] dealt with an extended version of the modified secant equation of [119] together with using a nonnegative restriction similar to (2.3), and proposed a modified DL algorithm with larger parametric convergence interval rather than that of [107]. Peyghami *et al.* [89] studied improved versions of the DL type methods of [30] by tuning ξ adaptively and setting u_k as a convex combination of s_k and y_k in (2.1) with $C = 0$. That is, in [89] an adaptive version of the modified secant equation of [119] has been used. Dehghani and Bidabadi [44] put forward another DL type algorithm using the modified secant equation proposed by Yuan [112]. Inherited from the corresponding modified secant equations, all the modified DL methods suggested in [30, 44, 70, 89, 119] benefit the objective function values in addition to the gradient information. Also, they have been shown to be globally convergent for uniformly convex functions with the modified versions of (1.10), while, for which global convergence regardless of the convexity has been established with the modified versions of (1.11). To get global convergence for general functions without any restriction on the CG parameter (1.10), Zhou and Zhang [127] applied the modified secant equation of [67]. In a generalization scheme, to simultaneously take advantage of the objective function values as in [30, 44, 70, 89, 119], and to get global convergence without convexity supposition or the mentioned nonnegativity restriction, Arazm *et al.* [13] proposed a modified DL method using the extended secant equation (2.1). Another study with similar aim has been carried out by Dehghani *et al.* [45] with improving the method of [30] based on the modified secant equation of [67]. Using modified structured secant equations, Kobayashi *et al.* [64] addressed the nonlinear least squares problems by the CG algorithms of the DL framework as well.

In most of the DL type methods discussed above, a version of the extended secant equation (2.1) was used with θ_k or θ_k^+ respectively defined by (2.2) or (2.3). However, Ford *et al.* [53] utilized the multi-step secant equations [52] in the DL approach to develop a class of multi-step nonlinear CG algorithms, employing information available from more than one previous iteration. Also, based on the concept of the spectral scaling secant equation, Liu *et al.* [76] developed a spectral DL technique with an adaptive choice for t .

3. ADAPTIVE CHOICES FOR THE DAI-LIAO PARAMETER

At the beginning of the previous decade, Andrei [10] classified several problems in the CG algorithms that have remained open. The given list then served as the origin of significant studies in the field. Especially, the second item of the list is related to the optimal choices of the DL parameter t , which mainly affects the theoretical

aspects as well as computational behavior of the method. Here, we review several studies that addressed the adaptive, sort of optimal choices of t . As will be discussed, several efforts targeted the modified secant equations to get appropriate choices for the DL parameter. It should be noted that in [30, 40, 53, 70, 89, 127] promising computational outputs have been reported with constant settings of t determined by simple trial and error schemes. Especially, in [14], based on the holistic conjugacy condition (1.9), it has been numerically shown that for small values of $g_{k+1}^T s_k$, it is better to set $t = 1$ to benefit the QN aspects, while, otherwise, the setting $t = 0$ is more reasonable.

To conduct convergence analysis of the CG algorithms, it is often of great necessity for the search directions to satisfy the descent condition [43],

$$g_k^T d_k < 0, \quad \forall k \geq 0. \quad (3.1)$$

Besides, sufficient descent condition may be pivotal to establish convergence of the methods [40, 54]; that is

$$g_k^T d_k \leq -C \|g_k\|^2, \quad \forall k \geq 0, \quad (3.2)$$

where $C > 0$ is a constant. Also, (3.2) has been classically considered as a superiority of the CG methods [38, 57]. So, taking these facts into consideration and inspired by the Shanno's matrix viewpoint on the CG algorithms [93], Babaie-Kafaki and Ghanbari [20] noted that the DL search directions can be written as $d_{k+1} = -\mathcal{P}_{k+1} g_{k+1}$, for all $k \geq 0$, where

$$\mathcal{P}_{k+1} = I - \frac{s_k y_k^T}{s_k^T y_k} + t \frac{s_k s_k^T}{s_k^T y_k}, \quad (3.3)$$

being nonsingular when $t > 0$ and $s_k^T y_k \neq 0$ [19]. Then, conducting an eigenvalue analysis on a symmetrized version of \mathcal{P}_{k+1} given by

$$\mathcal{A}_{k+1} = \frac{\mathcal{P}_{k+1} + \mathcal{P}_{k+1}^T}{2},$$

they obtained the following two-parameter choices for t :

$$t_k^{p,q} = p \frac{\|y_k\|^2}{s_k^T y_k} - q \frac{s_k^T y_k}{\|s_k\|^2}, \quad (3.4)$$

which ensures the descent condition (3.1) with $p > \frac{1}{4}$ and $q < \frac{1}{4}$. It is notable that the choices $(p, q) = (2, 0)$ and $(p, q) = (1, 0)$ respectively yield the CG parameters proposed by Hager and Zhang [55], and Dai and Kou [42]. Hence, the CG algorithms of [42, 55] lie within the DL family of CG methods. Moreover, as established in [15, 57], by setting

$$(p, q) = \left(\tau_k \frac{\|y_k\|^2}{s_k^T y_k}, 0 \right), \quad \text{or equivalently, } t = t_k^\tau = \tau_k \frac{\|y_k\|^2}{s_k^T y_k},$$

in which $\tau_k \geq \bar{\tau}$ with the constant $\bar{\tau} > \frac{1}{4}$, the sufficient descent condition (3.2) is achieved with $C = 1 - \frac{1}{4\bar{\tau}}$. It is also worth mentioning that, as established in [2], $(p, q) = \left(\frac{1}{4}, -\frac{3}{4} \right)$ yields the minimizer of the Byrd-Nocedal measure function [34] of the matrix \mathcal{A}_{k+1} .

In another research line, well-conditioning of the DL search direction matrix \mathcal{P}_{k+1} defined by (3.3) has been considered, to enhance stability of the method. As known, condition number is a crucial factor in matrix computations that should be preferably small to the possible extent [99]. Initially, upon a singular value analysis on \mathcal{P}_{k+1} in the Euclidean matrix norm in the sense of minimizing the spectral condition number, Babaie-Kafaki and Ghanbari [19] suggested the following two adaptive choices for the DL parameter:

$$t_{k_1}^E = \frac{s_k^T y_k}{\|s_k\|^2} + \frac{\|y_k\|}{\|s_k\|}, \quad \text{and } t_{k_2}^E = \frac{\|y_k\|}{\|s_k\|},$$

which the first one is shown to be computationally outstanding. Then, carrying out simultaneous singular value and eigenvalue analyses, Zhang *et al.* [124] showed that the choice $(p, q) = \left(1, \frac{1}{4}\right)$ in (3.4) minimizes another upper bound of the spectral condition number of \mathcal{P}_{k+1} , also ensuring the sufficient descent condition. Moreover, Babaie-Kafaki and Ghanbari [28] considered the DL parameter in the forms of

$$t = 1 + \vartheta \frac{\|y_k\|^2}{s_k^T y_k}, \quad (3.5)$$

and

$$t = \vartheta + \frac{\|y_k\|^2}{s_k^T y_k}, \quad (3.6)$$

with the parameter $\vartheta > 0$, making \mathcal{P}_{k+1} similar to the scaled memoryless BFGS (Broyden-Fletcher-Goldfarb-Shanno) updating formula for the inverse Hessian [97], *i.e.*

$$H_{k+1}^\vartheta = \vartheta_k I - \vartheta_k \frac{s_k y_k^T + y_k s_k^T}{s_k^T y_k} + \left(1 + \vartheta_k \frac{y_k^T y_k}{s_k^T y_k}\right) \frac{s_k s_k^T}{s_k^T y_k} \approx \nabla^2 f(x_{k+1})^{-1}, \quad (3.7)$$

in which $\vartheta_k > 0$ is called the scaling parameter. Through this initiative, the DL method may benefit the second order information more explicitly. Then, studying the spectral condition number of \mathcal{P}_{k+1} with the settings (3.5) and (3.6), they obtained optimal values of ϑ , yielding

$$t_{k_1}^{\text{BFGS}} = 1 + \sqrt{1 + \frac{\|y_k\|^2}{\|s_k\|^2}},$$

corresponding to (3.5), and

$$t_{k_2}^{\text{BFGS}} = \sqrt{\frac{\|y_k\|^4}{(s_k^T y_k)^2} + \frac{\|y_k\|^2}{\|s_k\|^2}} + \frac{\|y_k\|^2}{s_k^T y_k},$$

corresponding to (3.6). Also, in an attempt to minimize the condition number of \mathcal{P}_{k+1} in the Frobenius norm, Babaie-Kafaki and Ghanbari [22] achieved another choice for t as

$$t_k^{\text{F}} = \sqrt{\frac{\|y_k\| (s_k^T y_k)}{\|s_k\|^3}}. \quad (3.8)$$

In another study, Aminifard and Babaie-Kafaki [2] showed that

$$t_k^{\ell_1} = \sqrt{\frac{\|y_k\|_\infty}{\|s_k\|_\infty} \left(\frac{s_k^T y_k + \|s_k\|_1 \|y_k\|_\infty}{\|s_k\|^2 + \|s_k\|_1 \|s_k\|_\infty} \right)},$$

is a minimizer of an upper bound of the ℓ_1 -norm condition number of \mathcal{P}_{k+1} , while

$$t_k^{\ell_\infty} = \sqrt{\frac{\|y_k\|_1}{\|s_k\|_1} \left(\frac{s_k^T y_k + \|s_k\|_\infty \|y_k\|_1}{\|s_k\|^2 + \|s_k\|_\infty \|s_k\|_1} \right)},$$

is the minimizer of an upper bound of the ℓ_∞ -norm condition number of \mathcal{P}_{k+1} .

Several other studies focused on the least squares models to get appropriate choices for the DL parameter, inspired by the approach of Dai and Kou [42]. In this regard, to take advantages of the merits of the three-term CG algorithm proposed by Zhang *et al.* (ZZL) [123] with the search directions

$$d_0 = -g_0, \quad d_{k+1}^{\text{ZZL}} = -g_{k+1} + \beta_k^{\text{HS}} d_k - \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k, \quad \forall k \geq 0,$$

especially satisfying an equality form of (3.2) with $\mathcal{C} = 1$ regardless of the line search and the objective function convexity, Babaie-Kafaki and Ghanbari [22] calculated t as the solution of the least squares problem $\min_{t \geq 0} \|d_{k+1}^{\text{DL}} - d_{k+1}^{\text{ZZL}}\|^2$, achieving

$$t_k^{\text{ZZL}} = \frac{s_k^T y_k}{\|s_k\|^2}, \quad (3.9)$$

which can be also obtained as the minimizer of the distance between the maximum and the minimum singular values of \mathcal{P}_{k+1} [17]. As put forward in [17], t_k^{ZZL} is shown to be the minimizer of an upper bound of the spectral condition number of \mathcal{P}_{k+1} given by Piazza and Politi [90]. Also, Andrei [12] acquired t_k^{ZZL} as a result of clustering the eigenvalues of \mathcal{P}_{k+1} . In a similar scheme, Li *et al.* [72] used Andrei's adaptive three-term CG algorithm [11] and obtained (3.4) with an adaptive choice for p besides the fixed choice $q = 1$. In another attempt with the aim of taking advantage of the second order information provided by the BFGS update, by solving $\min_{t \geq 0} \|\mathcal{P}_{k+1} - H_{k+1}^\vartheta\|_{\text{F}}^2$, where $\|\cdot\|_{\text{F}}$ stands for the Frobenius matrix norm and H_{k+1}^ϑ is defined by (3.7), Babaie-Kafaki and Ghanbari [25] obtained the following one-parameter choices for t :

$$t_k^\vartheta = 1 + \vartheta_k \frac{\|y_k\|^2}{s_k^T y_k} - \vartheta_k \frac{s_k^T y_k}{\|s_k\|^2}.$$

By considering the DL parameter in the two-parameter framework of (3.4), in a similar approach Li *et al.* [73] proposed an adaptive setting for (p, q) . Moreover, Babaie-Kafaki [17] noted that it is reasonable to compute the DL parameter in a way that \mathcal{P}_{k+1} tends to an orthonormal matrix [97], being perfectly conditioned with respect to the Euclidean norm in the sense of having unit spectral condition number. Thus, by solving $\min_{t \geq 0} \|\mathcal{P}_{k+1}^T - \mathcal{P}_{k+1}^{-1}\|_{\text{F}}^2$ where \mathcal{P}_{k+1}^{-1} is obtained by the Sherman-Morrison formula [97], t_k^{F} given by (3.8) has been regained in [17].

Aminifard and Babaie-Kafaki [3] noted that when the gradient approaches the direction of the maximum magnification [99] by the search direction matrix \mathcal{P}_{k+1} , some computational difficulties arose alongside undesirable convergence behavior of the DL method. To resolve the issues, they determined the parameter t in a way to make the gradient orthogonal to the direction of the maximum magnification by \mathcal{P}_{k+1} . In the analysis of [3] a fixed point equation plays a pillar role. Thus, as a popular technique for solving fixed point equations, in an omnipresent scheme Babaie-Kafaki and Aminifard [29] used the functional iteration method to improve effectiveness of the adaptive choices of the DL parameter proposed in the literature. Based on the concept of the maximum magnification, Aminifard and Babaie-Kafaki [4] proposed a restart strategy for the DL method which is capable to advance the computational performance.

Fatemi [48] obtained another adaptive choice for t based on the following penalty model:

$$\min_{\beta} (g_{k+1}^T d_{k+1} + M ((g_{k+2}^T s_k)^2 + (d_{k+1}^T y_k)^2)), \quad (3.10)$$

in which $M > 0$ is the penalty parameter and d_{k+1} is defined by (1.3). The model has been designed with the aim of achieving the sufficient descent condition (3.2) as well as the conjugacy condition (1.6) and the orthogonality of the gradient to the previous search directions as in the linear CG algorithms [97]. He also claimed that it is reasonable to set the DL parameter in the interval $\left(0, \frac{1}{2}\right]$. In a similar framework, Fatemi [49] dealt with the following variant of (3.10) to get another adaptive choice for the DL parameter as well:

$$\min_{\beta} \left(g_{k+1}^T d_{k+1} + M \sum_{i=0}^m (g_{k+2}^T s_{k-i})^2 \right).$$

In another initiative to employ higher order information of the objective function, Momeni and Peyghami [81] determined another adaptive formula for the DL parameter as a function of the step length obtained by the quadratic and/or cubic local models of the objective function.

Modified secant equations have been also employed to achieve proper choices for the DL parameter, benefiting their advantages presented in Section 2. For example, Zheng [125] used the modified secant equation of Yabe and Takano [107], and proposed an adaptive choice for t to be used in (1.11). Also, using the Newton direction in the sense of setting $d_{k+1}^{\text{DL}} = d_{k+1}^{\text{Newton}}$, Lotfi and Hosseini [78], and Lu *et al.* [79, 80] dealt with the following equation:

$$-\nabla^2 f(x_{k+1})^{-1} g_{k+1} = -g_{k+1} + \frac{g_{k+1}^T y_k}{d_k^T y_k} d_k - t \frac{g_{k+1}^T s_k}{d_k^T y_k} d_k.$$

Performing inner product on both sides of the above equation by $s_k^T \nabla^2 f(x_{k+1})$, they obtained

$$t_k^{\text{Secant}} = \frac{s_k^T g_{k+1} - s_k^T \nabla^2 f(x_{k+1}) g_{k+1} + \frac{g_{k+1}^T y_k}{s_k^T y_k} s_k^T \nabla^2 f(x_{k+1}) s_k}{\frac{g_{k+1}^T s_k}{s_k^T y_k} s_k^T \nabla^2 f(x_{k+1}) s_k}.$$

Then, Lotfi and Hosseini [78] simplified the above formula using the modified secant equation of [67], while, Lu *et al.* [79, 80] applied the modified secant equations of [67, 70, 104] in the framework of (2.1).

4. MODIFIED VERSIONS OF THE HAGER–ZHANG AND DAI–KOU METHODS

As a deep study at the first years of the 21st century, Hager and Zhang (HZ) [56] founded a great algorithm, called CG_DESCENT, which today is entitled as a quite helpful tool to handle the large scale continuous optimization models. In particular, CG_DESCENT can be regarded as a DL type algorithm [75] with a judicious choice for t , *i.e.*

$$t_k^{\text{HZ}} = 2 \frac{\|y_k\|^2}{s_k^T y_k}.$$

Although satisfying the sufficient descent condition [15], the DL method with $t = t_k^{\text{HZ}}$ fails to guarantee the global converge without convexity supposition. To resolve the weak spot, Hager and Zhang [55, 56] proposed the following CG parameter:

$$\bar{\beta}_k^{\text{HZ}} = \max \{ \beta_k^{\text{HZ}}, \eta_k \}, \quad \eta_k = \frac{-1}{\|d_k\| \min \{ \eta, \|g_k\| \}},$$

where $\eta > 0$ is a constant and β_k^{HZ} is β_k^{DL} with $t = t_k^{\text{HZ}}$. Afterwards, Dai and Kou (DK) [42] suggested another choice for the DL parameter as well, with

$$t_k^{\text{DK}} = \vartheta_k + \frac{\|y_k\|^2}{s_k^T y_k} - \frac{s_k^T y_k}{\|s_k\|^2}, \quad (4.1)$$

where, loosely speaking, ϑ_k represents the scaling parameter of the scaled memoryless BFGS updating formula (3.7). They found t_k^{ZZL} given by (3.9) as the best choice for ϑ_k . To get the global convergence for general functions, they introduced the following restricted CG parameter:

$$\bar{\beta}_k^{\text{DK}} = \max \left\{ \beta_k^{\text{DK}}, \eta \frac{g_{k+1}^T d_k}{\|d_k\|^2} \right\},$$

where $\eta \in [0, 1)$ is a parameter and β_k^{DK} is β_k^{DL} with $t = t_k^{\text{DK}}$.

Validity and reliability of the CG_DESCENT algorithm triggered the interest of several scholars. For example, Li and Huang [69] proposed a modified CG_DESCENT algorithm based on the Yabe–Takano modified secant equation [107]. Thus, the method of [69] fulfills the sufficient descent condition while using the objective function

values. Then, based on a singular value analysis, Babaie–Kafaki and Ghanbari [23] noted that large (absolute) values of the second term of

$$\beta_k^{\text{HZ}} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t_k^{\text{HZ}} \frac{g_{k+1}^T s_k}{d_k^T y_k},$$

may lead to an ill-conditioned search direction matrix and so, in such situations it is better to ignore the term in the sense of using β_k^{HS} instead of β_k^{HZ} . As a result, in [23] a discrete hybridization of the HS and HZ methods has been suggested.

Improving performance of the DK method has also attracted significant attentions. As examples, Kou [65], and Faramarzi and Amini [47] developed improved versions of the DK method using the modified secant equations given in [30, 107]. Hence, the methods of [47, 65] possess the mentioned properties of the method of [69], *i.e.* benefiting the function values while satisfying (3.2). To improve orthogonality of the gradient vectors generated by the DK method as an advantageous feature of the linear CG methods, Liu *et al.* [77] developed a special scaled version of the DK method using a matrix obtained based on a QN update. Huang and Liu [60] set ϑ_k in (4.1) as a convex combination of the Oren–Luenberger [87] and Oren–Spedicato [88] scaling parameters and developed modified DK algorithms based on some new line search conditions.

5. EXTENSIONS OF THE CLASSICAL CONJUGATE GRADIENT PARAMETERS BASED ON THE DAI-LIAO APPROACH

As a work in progress, scholars waged significant studies to expand DL approach on the classical CG parameters, achieving one-parameter extensions of the parameters. Such attempts have been devised to get the descent property or to enhance the efficiency. Amidst suchlike studies, Babaie–Kafaki and Ghanbari [18] proposed an extension of the Polak–Ribière–Polyak (PRP) [91, 92] parameter as follows:

$$\beta_k^{\text{EPRP}} = \beta_k^{\text{PRP}} - t \frac{g_{k+1}^T d_k}{\|g_k\|^2}, \quad (5.1)$$

where $\beta_k^{\text{PRP}} = \frac{g_{k+1}^T y_k}{\|g_k\|^2}$, and t is a nonnegative parameter. Then, in light of the eigenvalue analysis carried out in [20], they acquired a two-parameter choice for t which ensures the descent property. It is worth mentioning that the DPRP method suggested by Yuan [113] is a member of the EPRP class of CG algorithms with $t = \mu \frac{\|y_k\|^2}{\|g_k\|^2}$,

ensuring the sufficient descent condition (3.2) for a constant setting of μ in $\left(\frac{1}{4}, +\infty\right)$. Global convergence of DPRP has been analyzed by Yu *et al.* [110] under some modified line searches. To employ the objective function values in the PRP framework, Yuan *et al.* [115] made a hybrid modification on the DPRP parameter using a version of the modified secant equation proposed by Li *et al.* [70]. Also, Babaie–Kafaki and Ghanbari [27] conducted a singular value analysis on a rank-two perturbation of the identity matrix, being a generalization of the DL and EPRP search direction matrices, to get an optimal choice for the EPRP parameter t as (sort of) minimizer of the spectral condition number of the updating formula. Making the EPRP search direction to bend to the direction of the efficient three-term CG algorithm proposed by Zhang *et al.* [121] in a least square model, another adaptive choice for the EPRP parameter has been given in [27] as well. Performing an eigenvalue analysis in light of the concept of the maximum magnification by a symmetrized version of the EPRP search direction matrix, Aminifard and Babaie–Kafaki [5] proposed another formula for t in (5.1). Their formula is capable to simultaneously improve the convergence and the numerical behavior of EPRP. Moreover, Babaie–Kafaki *et al.* [31] applied the modified PRP parameter proposed by Sun and Liu [98] to get another CG parameter in the DL framework. Andrei [9] studied a DL generalization of the Dai–Yuan CG parameter [39] with a mix of acceleration, taking into account the optimal choice of t given in [42]. A similar extension of the modified Liu–Storey (LS) [74] CG parameter proposed by Yao *et al.* [94] has been analyzed by Cheng *et al.*

[37]. Also, Yao *et al.* [108] studied a DL generalization of the modified HS parameter given in [94]. Aminifard and Babaie–Kafaki [6] put forward such extension on the effective hybrid CG parameter proposed by Jian *et al.* [63]. In addition, Zheng and Zheng [126] targeted the modified CG parameters given by Dai and Wen [41], being improved versions of the HS and LS parameters, to suggest other CG parameters with the DL structure. Nakamura *et al.* [82] developed extended versions of several classical CG parameters based on the DL scheme, taking into account the choices of t in the HZ framework [57].

Three-term CG algorithms weighed in acquiring generalized versions of the DL method as well, yielding the sufficient descent condition (3.2) by a simple but meaningful plan. Especially, Sugiki *et al.* [96] used the class of three-term CG algorithms established by Narushima *et al.* [84] as a role model to get a three-term generalization of the Yabe–Takano [107] CG parameter which is a member of the DL family of CG parameters. Then, founded upon the three-term CG method of [123], Babaie–Kafaki and Ghanbari [21] developed another three-term extension of the DL method in which the parameter t is determined based on the standard secant equation. In light of a matrix point of view, Yao *et al.* [109] put forward another three-term extension of the DL method in which t is determined based on the conjugacy condition (1.9). Babaie–Kafaki and Fatemi [50] developed a three-term version of the DL method based on a penalty model similar to (3.10). Besides, a four-term generalization of the DL method has been studied by Babaie–Kafaki and Ghanbari [26], together with an adaptive formula for t given by approaching the search direction matrix of the method to the scaled memoryless BFGS updating formula (3.7) in the Frobenius norm. In another guide line, Babaie–Kafaki and Ghanbari [24] proposed a special symmetrized version of the DL search direction matrix \mathcal{P}_{k+1} given by (3.3) that contains the memoryless BFGS updating formula as a special case. Then, performing an eigenvalue analysis, they gained two adaptive formulas for the parameter t , leading to two other generalized DL algorithms.

6. APPLICATIONS OF THE DAI–LIAO CLASS OF CONJUGATE GRADIENT METHODS IN PRACTICAL DISCIPLINES

Recently, as the well-known, well-studied and well-developed memoryless algorithms, DL type methods put to the test for solving several practical optimization problems. As a result, they are now technically recognized as an efficient tool to address real world optimization models, even being capable to reach their full potential in an era defined by broad-based, inclusive growth in the size of the data sets. Here, we briefly review several attempts around highlighting practical aspects of the DL algorithms.

6.1. Nonlinear systems of equations

Consider a nonlinear system of equations as $F(x) = 0$, where $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is continuous. As known, F is called monotone [97] when

$$(F(x) - F(y))^T(x - y) \geq 0, \quad \forall x, y \in \mathbb{R}^n.$$

Nonlinear systems of equations often straightly emerge in engineering applications [58, 103, 114, 116]. Especially, monotone cases of the model appear as the subproblems of the generalized proximal algorithms with Bregman distances [61]. In the following, several studies for developing the DL algorithms (by derivative free schemes) to solve the nonlinear systems of monotone equations are reviewed.

As an initial study on the issue, Abubakar and Kumam [1] applied the DL method given in [20] to solve the problem. Then, based on eigenvalue analyses, Waziri *et al.* [100, 103] developed several DL methods with the parameter choices in the framework of HZ [55, 57] for solving the nonlinear systems of monotone equations. Especially, in [103] they assessed efficiency of their algorithms in compressed sensing. Waziri *et al.* [101, 102] also used the DL algorithms given in [13, 30, 70, 107] to address the model. Based on a scalar approximation of the matrix $F' = \nabla F$ yielding an adaptive formula for t , Halilu *et al.* [58] proposed another CG algorithm of the DL family to solve the problem and evaluated its performance in robotics. Recently, in a similar scheme, an accelerated DL projection algorithm for solving the systems of monotone equations has been developed by Branislav *et al.* [62].

6.2. Image restoration

As a frequently reported issue, images may be corrupted *via* impulse noise, one of the fashionable noise models in which just a part of the pixels is contaminated. Generally, image restoration (reconstruction) approaches take on meaningful noise suppression schemes to recover the original image. A well-known image restoration model consists of minimizing a composite function [35], *i.e.* sum of a smooth function and a convex continuous but (often) nonsmooth function. Nonsmooth structure of the model makes the noise removal procedure challenging. So, smooth relaxations of the image restoration model have been devised in the literature that can be addressed by large scale optimization algorithms [111]. As already mentioned, Babaie-Kafaki *et al.* [6,31], Lu *et al.* [79,80], Waziri *et al.* [103] and Branislav *et al.* [62] tackled the impulse noise removal of the images by the CG algorithms of the DL family.

6.3. Compressed sensing

As known, compressed sensing is a signal processing strategy for effectively acquiring and reconstructing signals in a sparse structure which makes it possible to get memoryless compact storage of the signals [33]. Importance of the compressed sensing in real world applications such as machine learning, compressive imaging and radar, wireless sensors network, medical imaging, astrophysical signals and video coding has been referred in [7].

Also known as sparse recovery, compressed sensing principally deals with sparse solutions of an extremely underdetermined system $Ax = b$ with $A \in \mathbb{R}^{m \times n}$ ($m \ll n$) and $b \in \mathbb{R}^m$, for which it is mostly fashionable to address the following (composite) unconstrained optimization model:

$$\min_{x \in \mathbb{R}^n} f(x) = \frac{1}{2} \|Ax - b\|_2^2 + \mu \|x\|_1, \quad (6.1)$$

where $\mu > 0$ is the penalty parameter, embedded to balance the sparsity and reconstruction quality of the solution [46]. The model (6.1) for the compressed sensing is called the basis pursuit denoising (BPD) problem which has been significantly analyzed in the literature. Technically, presence of the nonsmooth ℓ_1 penalty term in (6.1) makes the problem to some extent challenging. So, recognizing Nesterov's underpinning strategy [85] as a role model, Zhu *et al.* [128] proposed a relaxation of BPD which can be effectively solved by classic optimization tools. As already mentioned, Waziri *et al.* [103] developed two modified HZ methods for monotone nonlinear systems of equations and then, as a case study, investigated capacity of the methods for compressed sensing. Also, Aminifard and Babaie-Kafaki [6] investigated efficiency of some DL extensions of the hybrid CG parameter of [63] for solving the compressed sensing problem.

7. CONCLUSIONS

Since the turn of the century, one of the most outstanding conjugacy conditions has been proposed by Dai and Liao [40], being wellspring of broad-ranging deep studies. Actually, they laid foundation of a class of meaningful one-parameter conjugate gradient algorithms which contains the efficient CG_DESCENT algorithm. Keeping on evolving to expose their potentials, the Dai-Liao methods are principally enriched by the quasi-Newton aspects. Although computational evidences are generally convincing, Dai-Liao algorithms technically face some major challenges such as generating uphill search directions and being vulnerable to improper settings of their parameter. To advance the algorithms both in theoretical and practical fields, researchers set out to improving the Dai-Liao method in several baselines. Here, we have classified such attempts to make them crystal clear and unveil their borders in a way that the readers can get an instant evaluation of the progress delivered on the Dai-Liao algorithms in different ramifications.

At the first step, because of the close merger between the Dai-Liao algorithms and the quasi-Newton aspects, we revealed how the modified secant equations have been attached to the Dai-Liao algorithms. However, the main part of our study has been devoted to review origins of the optimal values of the Dai-Liao parameter

given in the literature. We also depicted how the scope of the Dai–Liao scheme has been widened to the other classical conjugate gradient parameters, yielding one-parameter extensions of the traditional parameters. Finally, we emphasized practical efficiency of the Dai–Liao algorithms when dealing with the practical issues such as signal and image processing models as the real world case studies.

This review is meaningfully capable to plant a seed for devising new modifications of the Dai–Liao algorithms. Future studies on the Dai–Liao methods may target global convergence under the nonmonotone line searches, or promoting the methods for the nonsmooth optimization problems. As a final note, we encourage the researchers to put a practical spin on their studies around the Dai–Liao algorithms in the sense of targeting state-of-the-art real world issues. Alongside the mentioned practical models, as examples, the nonnegative matrix factorization [71] and the Muskingum model [118] can be also addressed by the Dai–Liao algorithms. Moreover, the algorithms are capable to be used for designing the kernel methods [105] as well as the adaptive filtering techniques [36, 66] which are significantly employed in signal processing and machine learning disciplines, such as support vector machine, support vector regression and extreme learning machine.

Acknowledgements. Since the main part of the author’s Ph.D. thesis is devoted to modifying the Dai–Liao method, he owes a great debt of gratitude to his thesis supervisor, Professor Nezam Mahdavi–Amiri from Faculty of Mathematical Sciences of Sharif University of Technology. He also is grateful to Professor Marcos Raydan from Center of Mathematics and Applications (CMA) of Nova University of Lisbon for his useful hints and suggestions helped to enhance quality of this work. The author appreciates helpful comments of the anonymous reviewers as well.

Conflict of interest. There are no conflicts of interest.

REFERENCES

- [1] A.B. Abubakar and P. Kumam, A descent Dai–Liao conjugate gradient method for nonlinear equations. *Numer. Algorithms* **81** (2019) 197–210.
- [2] Z. Aminifard and S. Babaie–Kafaki, Matrix analysis on the Dai–Liao conjugate gradient method. *ANZIAM J.* **61** (2019) 195–203.
- [3] Z. Aminifard and S. Babaie–Kafaki, 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.
- [4] Z. Aminifard and S. Babaie–Kafaki, A restart scheme for the Dai–Liao conjugate gradient method by ignoring a direction of maximum magnification by the search direction matrix. *RAIRO:RO* **54** (2020) 981–991.
- [5] Z. Aminifard and S. Babaie–Kafaki, An adaptive descent extension of the Polak–Rebière–Polyak conjugate gradient method based on the concept of maximum magnification. *Iran. J. Numer. Anal. Optim.* **11** (2021) 211–219.
- [6] Z. Aminifard and S. Babaie–Kafaki, Dai–Liao extensions of a descent hybrid nonlinear conjugate gradient method with application in signal processing. *Numer. Algorithms* **89** (2022) 1369–1387.
- [7] Z. Aminifard, A. Hosseini and S. Babaie–Kafaki, Modified conjugate gradient method for solving sparse recovery problem with nonconvex penalty. *Signal Process.* **193** (2022) 108424.
- [8] N. Andrei, Numerical comparison of conjugate gradient algorithms for unconstrained optimization. *Stud. Inform. Control* **16** (2007) 333–352.
- [9] N. Andrei, New accelerated conjugate gradient algorithms as a modification of Dai–Yuan’s computational scheme for unconstrained optimization. *J. Comput. Appl. Math.* **234** (2010) 3397–3410.
- [10] N. Andrei, Open problems in nonlinear conjugate gradient algorithms for unconstrained optimization. *Bull. Malays. Math. Sci. Soc.* **34** (2011) 319–330.
- [11] N. Andrei, An adaptive conjugate gradient algorithm for large scale unconstrained optimization. *J. Comput. Appl. Math.* **292** (2016) 83–91.
- [12] N. Andrei, A Dai–Liao conjugate gradient algorithm with clustering of eigenvalues. *Numer. Algorithms* **77** (2018) 1273–1282.
- [13] M.R. Arazm, S. Babaie–Kafaki and R. Ghanbari, An extended Dai–Liao conjugate gradient method with global convergence for nonconvex functions. *Glas. Mat.* **52** (2017) 361–375.
- [14] S. Babaie–Kafaki, An adaptive conjugacy condition and related nonlinear conjugate gradient methods. *Int. J. Comput. Methods* **11** (2014) 1350092.
- [15] S. Babaie–Kafaki, On the sufficient descent condition of the Hager–Zhang conjugate gradient methods. *4OR* **12** (2014) 285–292.
- [16] S. Babaie–Kafaki, Computational approaches in large scale unconstrained optimization, in *Big Data Optimization: Recent Developments and Challenges*. Springer (2016) 391–417.
- [17] S. Babaie–Kafaki, On optimality of two adaptive choices for the parameter of Dai–Liao method. *Optim. Lett.* **10** (2016) 1789–1797.

- [18] S. Babaie-Kafaki and R. Ghanbari, A descent extension of the Polak-Ribière-Polyak conjugate gradient method. *Comput. Math. Appl.* **68** (2014) 2005–2011.
- [19] S. Babaie-Kafaki and R. Ghanbari, The Dai-Liao nonlinear conjugate gradient method with optimal parameter choices. *Eur. J. Oper. Res.* **234** (2014) 625–630.
- [20] S. Babaie-Kafaki and R. Ghanbari, A descent family of Dai-Liao conjugate gradient methods. *Optim. Methods Softw.* **29** (2014) 583–591.
- [21] S. Babaie-Kafaki and R. Ghanbari, Two modified three-term conjugate gradient methods with sufficient descent property. *Optim. Lett.* **8** (2014) 2285–2297.
- [22] S. Babaie-Kafaki and R. Ghanbari, Two optimal Dai-Liao conjugate gradient methods. *Optimization* **64** (2015) 2277–2287.
- [23] S. Babaie-Kafaki and R. Ghanbari, An adaptive Hager-Zhang conjugate gradient method. *Filomat* **30** (2016) 3715–3723.
- [24] S. Babaie-Kafaki and R. Ghanbari, Descent symmetrization of the Dai-Liao conjugate gradient method. *Asia-Pac. J. Oper. Res.* **33** (2016) 1650008.
- [25] S. Babaie-Kafaki and R. Ghanbari, A class of adaptive Dai-Liao conjugate gradient methods based on the scaled memoryless BFGS update. *JOR* **15** (2017) 85–92.
- [26] S. Babaie-Kafaki and R. Ghanbari, A class of descent four-term extension of the Dai-Liao conjugate gradient method based on the scaled memoryless BFGS update. *J. Ind. Manag. Optim.* **13** (2017) 649.
- [27] S. Babaie-Kafaki and R. Ghanbari, An optimal extension of the Polak-Ribière-Polyak conjugate gradient method. *Numer. Func. Anal. Optim.* **38** (2017) 1115–1124.
- [28] S. Babaie-Kafaki and R. Ghanbari, Two adaptive Dai-Liao nonlinear conjugate gradient methods. *Iran. J. Sci. Technol. – Trans. A: Sci.* **42** (2018) 1505–1509.
- [29] S. Babaie-Kafaki and Z. Aminifard, Improving the Dai-Liao parameter choices using a fixed point equation. *J. Math. Model.* **10** (2022) 11–20.
- [30] S. Babaie-Kafaki, R. Ghanbari and Nezam Mahdavi-Amiri, Two new conjugate gradient methods based on modified secant equations. *J. Comput. Appl. Math.* **234** (2010) 1374–1386.
- [31] S. Babaie-Kafaki, N. Mirhoseini and Z. Aminifard, A descent extension of a modified Polak-Ribière-Polyak method with application in image restoration problem. *Optim. Lett.* (2022).
- [32] F. Biglari, M.A. Hassan and W.J. Leong, New quasi-Newton methods via higher order tensor models. *J. Comput. Appl. Math.* **235** (2011) 2412–2422.
- [33] A.M. Bruckstein, D.L. Donoho and M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* **51** (2009) 34–81.
- [34] 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.
- [35] R.H. Chan, C.-Wa. Ho and M. Nikolova, Salt-and-pepper noise removal by median type noise detectors and detail-preserving regularization. *IEEE Trans. Image Process.* **14** (2005) 1479–1485.
- [36] P.S. Chang and A.N. Willson, Analysis of conjugate gradient algorithms for adaptive filtering. *IEEE Trans. Signal Process.* **48** (2000) 409–418.
- [37] Y. Cheng, Q. Mou, X. Pan and S. Yao, A sufficient descent conjugate gradient method and its global convergence. *Optim. Methods Softw.* **31** (2016) 577–590.
- [38] Y.-H. Dai, New properties of a nonlinear conjugate gradient method. *Numer. Math.* **89** (2001) 83–98.
- [39] Y.-H. Dai and Y.-X. Yuan, A nonlinear conjugate gradient method with a strong global convergence property. *SIAM J. Optim.* **10** (1999) 177–182.
- [40] Y.-H. Dai and L.-Z. Liao, New conjugacy conditions and related nonlinear conjugate gradient methods. *Appl. Math. Optim.* **43** (2001) 87–101.
- [41] Z. Dai and F. Wen, Another improved Wei-Yao-Liu nonlinear conjugate gradient method with sufficient descent property. *Appl. Math. Comput.* **218** (2012) 7421–7430.
- [42] 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.
- [43] Y. Dai, J. Han, G. Liu, D. Sun, H. Yin and Y.-X. Yuan, Convergence properties of nonlinear conjugate gradient methods. *SIAM J. Optim.* **10** (2000) 345–358.
- [44] R. Dehghani and N. Bidabadi, Two-step conjugate gradient method for unconstrained optimization. *Comput. Appl. Math.* **39** (2020) 1–15.
- [45] R. Dehghani, N. Bidabadi, H. Fahs and M.M. Hosseini, A conjugate gradient method based on a modified secant relation for unconstrained optimization. *Numer. Funct. Anal. Optim.* **41** (2020) 621–634.
- [46] H. Esmaili, S. Shabani and M. Kimiaei, A new generalized shrinkage conjugate gradient method for sparse recovery. *Calcolo* **56** (2019) 1–38.
- [47] P. Faramarzi and K. Amini, A modified conjugate gradient method based on a modified secant equation. *Appl. Math. Model.* **8** (2020) 1–20.
- [48] M. Fatemi, A new efficient conjugate gradient method for unconstrained optimization. *J. Comput. Appl. Math.* **300** (2016) 207–216.
- [49] M. Fatemi, An optimal parameter for Dai-Liao family of conjugate gradient methods. *J. Optim. Theory Appl.* **169** (2016) 587–605.

- [50] M. Fatemi and S. Babaie-Kafaki, Two extensions of the Dai-Liao method with sufficient descent property based on a penalization scheme. *Bull. Comput. Appl. Math.* **4** (2016) 7–19.
- [51] R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients. *Comput. J.* **7** (1964) 149–154.
- [52] J.A. Ford and I. Moghrabi, Multi-step quasi-Newton methods for optimization. *J. Comput. Appl. Math.* **50** (1994) 305–323.
- [53] J.A. Ford, Y. Narushima and H. Yabe, Multi-step nonlinear conjugate gradient methods for unconstrained minimization. *Comput. Optim. Appl.* **40** (2008) 191–216.
- [54] J.C. Gilbert and J. Nocedal, Global convergence properties of conjugate gradient methods for optimization. *SIAM J. Optim.* **2** (1992) 21–42.
- [55] 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.
- [56] W.W. Hager and H. Zhang, Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranteed descent. *ACM Trans. Math. Softw.* **32** (2006) 113–137.
- [57] W.W. Hager and H. Zhang, A survey of nonlinear conjugate gradient methods. *Pacific J. Optim.* **2** (2006) 35–58.
- [58] A.S. Halilu, A. Majumder, M.Y. Waziri, K. Ahmed and A. Muhammed Awwal, Motion control of the two joint planar robotic manipulators through accelerated Dai-Liao method for solving system of nonlinear equations. *Eng. Comput.* **39** (2022) 1802–1840.
- [59] M.R. Hestenes and E. Stiefel, Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* **49** (1952) 409.
- [60] Y. Huang and C. Liu, Dai-Kou type conjugate gradient methods with a line search only using gradient. *J. Inequal. Appl.* **2017** (2017) 1–17.
- [61] N.A. Iusem and V.M. Solodov, Newton-type methods with generalized distances for constrained optimization. *Optimization* **41** (1997) 257–278.
- [62] B. Ivanov, G.V. Milovanović and P.S. Stanimirović, Accelerated Dai-Liao projection method for solving systems of monotone nonlinear equations with application to image deblurring. *J. Glob. Optim.* (2022).
- [63] J. Jian, L. Han and X. Jiang, A hybrid conjugate gradient method with descent property for unconstrained optimization. *Appl. Math. Model.* **39** (2015) 1281–1290.
- [64] M. Kobayashi, Y. Narushima and H. Yabe, Nonlinear conjugate gradient methods with structured secant condition for nonlinear least squares problems. *J. Comput. Appl. Math.* **234** (2010) 375–397.
- [65] C. Kou, An improved nonlinear conjugate gradient method with an optimal property. *Sci. China Math.* **57** (2014) 635–648.
- [66] C.-H. Lee, B.D. Rao and H. Garudadri, A sparse conjugate gradient adaptive filter. *IEEE Signal Process. Lett.* **27** (2020) 1000–1004.
- [67] D.-H. Li and M. Fukushima, A modified BFGS method and its global convergence in nonconvex minimization. *J. Comput. Appl. Math.* **129** (2001) 15–35.
- [68] D.-H. Li and M. Fukushima, On the global convergence of the BFGS method for nonconvex unconstrained optimization problems. *SIAM J. Optim.* **11** (2001) 1054–1064.
- [69] S. Li and Z. Huang, Guaranteed descent conjugate gradient methods with modified secant condition. *J. Ind. Manag. Optim.* **4** (2008) 739.
- [70] G. Li, C. Tang and Z. Wei, New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. *J. Comput. Appl. Math.* **202** (2007) 523–539.
- [71] X. Li, W. Zhang and X. Dong, A class of modified FR conjugate gradient method and applications to nonnegative matrix factorization. *Comput. Math. Appl.* **73** (2017) 270–276.
- [72] M. Li, H. Liu and Z. Liu, A new family of conjugate gradient methods for unconstrained optimization. *J. Appl. Math. Comput.* **58** (2018) 219–234.
- [73] X. Li, W. Zhao and X. Dong, A new CG algorithm based on a scaled memoryless BFGS update with adaptive search strategy, and its application to large scale unconstrained optimization problems. *J. Comput. Appl. Math.* **398** (2021) 113670.
- [74] Y. Liu and C. Storey, Efficient generalized conjugate gradient algorithms, part 1: theory. *J. Optim. Theory Appl.* **69** (1991) 129–137.
- [75] H. Liu, H. Wang and Q. Ni, On Hager and Zhangs conjugate gradient method with guaranteed descent. *Appl. Math. Comput.* **236** (2014) 400–407.
- [76] H. Liu, Y. Yao, X. Qian and H. Wang, Some nonlinear conjugate gradient methods based on spectral scaling secant equations. *Comput. Appl. Math.* **35** (2016) 639–651.
- [77] Z. Liu, H. Liu and Y.-H. Dai, An improved Dai-Kou conjugate gradient algorithm for unconstrained optimization. *Comput. Optim. Appl.* **75** (2020) 145–167.
- [78] M. Lotfi and S.M. Hosseini, An efficient Dai-Liao type conjugate gradient method by reformulating the CG parameter in the search direction equation. *J. Comput. Appl. Math.* **371** (2020) 112708.
- [79] J. Lu, Y. Li and H. Pham, A modified Dai-Liao conjugate gradient method with a new parameter for solving image restoration problems. *Math. Probl. Eng.* **2020** (2020) 6279543.
- [80] J. Lu, G. Yuan and Z. Wang, A modified Dai-Liao conjugate gradient method for solving unconstrained optimization and image restoration problems. *J. Appl. Math. Comput.* **68** (2022) 681–703.
- [81] M. Momeni and M.R. Peyghami, A new conjugate gradient algorithm with cubic Barzilai-Borwein stepsize for unconstrained optimization. *Optim. Methods Softw.* **34** (2019) 650–664.

- [82] W. Nakamura, Y. Narushima and H. Yabe, Nonlinear conjugate gradient methods with sufficient descent properties for unconstrained optimization. *J. Ind. Manag. Optim.* **9** (2013) 595–619.
- [83] Y. Narushima and H. Yabe, A survey of sufficient descent conjugate gradient methods for unconstrained optimization. *SUT J. Math.* **50** (2014) 167–203.
- [84] Y. Narushima, H. Yabe and J.A. Ford, A three-term conjugate gradient method with sufficient descent property for unconstrained optimization. *SIAM J. Optim.* **21** (2011) 212–230.
- [85] Y. Nesterov, Smooth minimization of nonsmooth functions. *Math. Program.* **103** (2005) 127–152.
- [86] J. Nocedal and S. Wright, Numerical Optimization. Springer, New York (2006).
- [87] S.S. Oren and D.G. Luenberger, Self-scaling variable metric (SSVM) algorithms: Part I: Criteria and sufficient conditions for scaling a class of algorithms. *Manag. Sci.* **20** (1974) 845–862.
- [88] S.S. Oren and E. Spedicato, Optimal conditioning of self-scaling variable metric algorithms. *Math. Program.* **10** (1976) 70–90.
- [89] M.R. Peyghami, H. Ahmadzadeh and A. Fazli, A new class of efficient and globally convergent conjugate gradient methods in the Dai–Liao family. *Optim. Methods Softw.* **30** (2015) 843–863.
- [90] G. Piazza and T. Politi, An upper bound for the condition number of a matrix in spectral norm. *J. Comput. Appl. Math.* **143** (2002) 141–144.
- [91] B.T. Polyak, The conjugate gradient method in extreme problems. *USSR Comput. Math. Math. Phys.* **9** (1969) 94–112.
- [92] E. Polak and G. Ribière, Note sur la convergence de méthodes de directions conjuguées. *ESAIM: Math. Model. Numer. Anal.* **3** (1969) 35–43.
- [93] D.F. Shanno, Conjugate gradient methods with inexact searches. *Math. Oper. Res.* **3** (1978) 244–256.
- [94] Y. Shengwei, Z. Wei and H. Huang, A note about WYL’s conjugate gradient method and its applications. *Appl. Math. Comput.* **191** (2007) 381–388.
- [95] P.S. Stanimirović, B. Ivanov, H. Ma and D. Mosić, A survey of gradient methods for solving nonlinear optimization. *Electron. Res. Arch.* **28** (2020) 1573.
- [96] K. Sugiki, Y. Narushima and H. Yabe, Globally convergent three-term conjugate gradient methods that use secant conditions and generate descent search directions for unconstrained optimization. *J. Optim. Theory Appl.* **153** (2012) 733–757.
- [97] W. Sun and Y.-X. Yuan, Optimization Theory and Methods: Nonlinear Programming. Springer, New York (2006).
- [98] M. Sun and J. Liu, Three modified Polak–Ribière–Polyak conjugate gradient methods with sufficient descent property. *J. Inequal. Appl.* **2015** (2015) 1–14.
- [99] D.S. Watkins, Fundamentals of Matrix Computations. John Wiley & Sons (2004).
- [100] M.Y. Waziri, K. Ahmed and J. Sabi’u, A family of Hager–Zhang conjugate gradient methods for system of monotone nonlinear equations. *Appl. Math. Comput.* **361** (2019) 645–660.
- [101] M.Y. Waziri, K. Ahmed and J. Sabi’u, A Dai–Liao conjugate gradient method via modified secant equation for system of nonlinear equations. *Arab. J. Math.* **9** (2020) 443–457.
- [102] M.Y. Waziri, K. Ahmed, J. Sabi’u and A.S. Halilu, Enhanced Dai–Liao conjugate gradient methods for systems of monotone nonlinear equations. *SEMA J.* **78** (2021) 15–51.
- [103] M.Y. Waziri, K. Ahmed, A.S. Halilu and J. Sabi’u, Two new Hager–Zhang iterative schemes with improved parameter choices for monotone nonlinear systems and their applications in compressed sensing. *RAIRO:RO* **56** (2022) 239–273.
- [104] Z. Wei, G. Li and L. Qi, New quasi-Newton methods for unconstrained optimization problems. *Appl. Math. Comput.* **175** (2006) 1156–1188.
- [105] K. Xiong, H.H.-C. Iu and S. Wang, Kernel correntropy conjugate gradient algorithms based on half-quadratic optimization. *IEEE Trans. Cybern.* **51** (2020) 5497–5510.
- [106] C. Xu and J. Zhang, A survey of quasi-Newton equations and quasi-Newton methods for optimization. *Ann. Oper. Res.* **103** (2001) 213–234.
- [107] H. Yabe and M. Takano, Global convergence properties of nonlinear conjugate gradient methods with modified secant condition. *Comput. Optim. Appl.* **28** (2004) 203–225.
- [108] S. Yao, X. Lu and Z. Wei, A conjugate gradient method with global convergence for large scale unconstrained optimization problems. *J. Appl. Math.* **2013** (2013) 730454.
- [109] S. Yao, Q. Feng, L. Li and J. Xu, A class of globally convergent three-term Dai–Liao conjugate gradient methods. *Appl. Numer. Math.* **151** (2020) 354–366.
- [110] G. Yu, L. Guan and G. Li, Global convergence of modified Polak–Ribière–Polyak conjugate gradient methods with sufficient descent property. *J. Ind. Manag. Optim.* **4** (2008) 565–579.
- [111] G. Yu, J. Huang and Y. Zhou, A descent spectral conjugate gradient method for impulse noise removal. *Appl. Math. Lett.* **23** (2010) 555–560.
- [112] Y.-X. Yuan, A modified BFGS algorithm for unconstrained optimization. *IMA J. Numer. Anal.* **11** (1991) 325–332.
- [113] G. Yuan, Modified nonlinear conjugate gradient methods with sufficient descent property for large scale optimization problems. *Optim. Lett.* **3** (2009) 11–21.
- [114] G. Yuan and M. Zhang, A three-terms Polak–Ribière–Polyak conjugate gradient algorithm for large scale nonlinear equations. *J. Comput. Appl. Math.* **286** (2015) 186–195.
- [115] G. Yuan, Z. Wei and Q. Zhao, A modified Polak–Ribière–Polyak conjugate gradient algorithm for large scale optimization problems. *IIE Trans.* **46** (2014) 397–413.
- [116] G. Yuan, T. Li and W. Hu, A conjugate gradient algorithm for large scale nonlinear equations and image restoration problems. *Appl. Numer. Math.* **147** (2020) 129–141.

- [117] G. Yuan, J. Lu and Z. Wang, 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.
- [118] G. Yuan, J. Lu and Z. Wang, The modified PRP conjugate gradient algorithm under a nondescent line search and its application in the Muskingum model and image restoration problems. *Soft Comput.* **25** (2021) 5867–5879.
- [119] J. Zhang and C. Xu, Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations. *J. Comput. Appl. Math.* **137** (2001) 269–278.
- [120] J. Zhang, N. Deng and L. Chen, New quasi-Newton equation and related methods for unconstrained optimization. *J. Optim. Theory Appl.* **102** (1999) 147–167.
- [121] L. Zhang, W. Zhou and D.-H. Li, A descent modified Polak–Ribière–Polyak conjugate gradient method and its global convergence. *IMA J. Numer. Anal.* **26** (2006) 629–640.
- [122] L. Zhang, W. Zhou and D. Li, Global convergence of a modified Fletcher–Reeves conjugate gradient method with Armijo type line search. *Numer. Math.* **104** (2006) 561–572.
- [123] L. Zhang, W. Zhou and D. Li, Some descent three-term conjugate gradient methods and their global convergence. *Optim. Methods Softw.* **22** (2007) 697–711.
- [124] K. Zhang, H. Liu and Z. Liu, A new Dai–Liao conjugate gradient method with optimal parameter choice. *Numer. Funct. Anal. Optim.* **40** (2019) 194–215.
- [125] Y. Zheng, A new family of Dai–Liao conjugate gradient methods with modified secant equation for unconstrained optimization. *RAIRO:RO* **55** (2021) 3281–3291.
- [126] Y. Zheng and B. Zheng, Two new Dai–Liao type conjugate gradient methods for unconstrained optimization problems. *J. Optim. Theory Appl.* **175** (2017) 502–509.
- [127] W. Zhou and L. Zhang, A nonlinear conjugate gradient method based on the MBFGS secant condition. *Optim. Methods Softw.* **21** (2006) 707–714.
- [128] H. Zhu, Y. Xiao and S.-Y. Wu, Large sparse signal recovery by conjugate gradient algorithm based on smoothing technique. *Comput. Math. Appl.* **66** (2013) 24–32.



Please help to maintain this journal in open access!

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting subscribers@edpsciences.org.

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.