Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 16, No. 4, pp. 581–594, 2020

MODIFIED SPECTRAL CONJUGATE GRADIENT METHODS BASED ON THE QUASI-NEWTON ASPECTS*

Zohre Aminifard† and Saman Babaie–Kafaki

Abstract: A class of spectral conjugate gradient methods is developed for solving unconstrained optimization problems in which the conjugate gradient parameter is computed based on the quasi–Newton aspects and then, the spectral parameter is determined to improve condition number of the corresponding search direction matrix. The given conjugate gradient parameter can be considered as a two–parameter extension of the Dai–Liao conjugate gradient parameter. To achieve sufficient descent property, a three–term extension of the given class of conjugate gradient methods is developed. Global convergence analysis of the methods is conducted. Numerical experiments are done on a set of test problems of the CUTEr collection. They show practical efficiency of the given adaptive choices of the parameters in the sense of the Dolan–Moré performance profile.

Key words: *nonlinear programming, unconstrained optimization, spectral conjugate gradient method, condition number, sufficient descent property, global convergence*

Mathematics Subject Classification: *90C53, 65K05*

1 Introduction

Spectral conjugate gradient (CG) methods comprise a class of optimization tools for solving unconstrained optimization problem $\min_{x\in\mathbb{R}^n} f(x)$ where f is called the objective function, here assumed to be nonlinear and smooth. Initially proposed by Birgin and Martínez [9] as a hybridization of the CG method [22] and the spectral gradient technique [31], iterative formula of the method is given by

$$x_0 \in \mathbb{R}^n,\ x_{k+1} = x_k + s_k,\ s_k = \alpha_k d_k,\ k = 0, 1, ..., \tag{1.1}$$

where $\alpha_k > 0$ is a step length generated by line search [28] along the (descent) direction d_k defined by

$$d_0 = -g_0,\ d_{k+1} = -\theta_k g_{k+1} + \beta_k d_k,\ k = 0, 1, ..., \tag{1.2}$$

where $g_k = \nabla f(x_k)$, and $\theta_k > 0$ and β_k are respectively called the spectral parameter and the CG parameter.

Various spectral CG methods have been developed in the literature by investigating appropriate choices for θ_k and β_k in (1.2). For example, taking advantage of the strong

*This research was supported by Research Council of Semnan University.
†Corresponding author.

© 2020 Yokohama Publishers

local and global convergence properties of the quasi–Newton techniques, Andrei [3] introduced a scaled BFGS preconditioned CG method. Also, based on a penalization strategy, Fatemi [16] suggested effective choices for parameters of the search direction (1.2). Recently, employing quasi–Newton aspects as well as parameter determination scheme of [11], Jian et al. [23] introduced another approach for computing the spectral parameter which ensures global convergence for uniformly convex objective functions. Their method has been then improved by Faramarzi and Amini [15] using the modified secant condition proposed by Li and Fukushima [26], especially in the sense of achieving global convergence for nonconvex objective functions. By constructing a restrictive relationship, Liu et al. [27] proposed another spectral CG method which possesses the effective sufficient descent property, i.e.

$$d_k^T g_k \leq -c||g_k||^2, \ k \geq 0, \tag{1.3}$$

independent to the line search. More recently, Aminifard and Babaie–Kafaki [2] used a spectral parameter to modify the efficient PRP method [29, 30], achieving the sufficient descent property as well as the global convergence for general functions.

Here, based on the quasi–Newton aspects [28] as well as using the singular value analysis conducted in [6], we suggest adaptive choices for the parameters θ_k and β_k in (1.2). Our formula for β_k can be considered as an extension of the essential CG parameter proposed by Dai and Liao (DL) [12]; that is

$$\beta_k^{DL} = \frac{g_{k+1}^T y_k}{d_k^T y_k} - t\frac{g_{k+1}^T s_k}{d_k^T y_k}, \tag{1.4}$$

derived based on an extended conjugacy condition, where t is a nonnegative parameter and $y_k = g_{k+1} - g_k$. Note that if the line search fulfills the popular Wolfe conditions [28], i.e.

$$f(x_{k+1}) - f(x_k) \leq \delta\alpha_k d_k^T g_k, \tag{1.5}$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k, \tag{1.6}$$

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 the best formula has not been achieved yet [4]. Nevertheless, some choices for the DL parameter can be listed as follows:

$$t = t_{k_1} = 0.1, \text{ proposed in [12]}; \tag{1.7}$$

$$t = t_{k_2} = \frac{s_k^T y_k}{||s_k||^2} + \frac{||y_k||}{||s_k||}, \text{ proposed in [6]}; \tag{1.8}$$

$$t = t_{k_3} = \frac{s_k^T y_k}{||s_k||^2}, \text{ proposed in [5, 7]}; \tag{1.9}$$

$$t = t_{k_4} = \frac{||s_k||^2||y_k||^2}{(s_k^T y_k)^2}, \text{ proposed in [8]}; \tag{1.10}$$

$$t = t_{k_5} = \sqrt{\frac{||y_k||(s_k^T y_k)}{||s_k||^3}}, \text{ proposed in [7]}; \tag{1.11}$$

$$t = t_{k_6} = \sqrt{\frac{||y_k||_1}{||s_k||_1}\left(\frac{s_k^T y_k + ||s_k||_\infty||y_k||_1}{||s_k||^2 + ||s_k||_\infty||s_k||_1}\right)}, \text{ proposed in [1]}; \tag{1.12}$$

where here $||.||$ stands for the Euclidean norm.

This work is organized as follows. In Section 2, based on the quasi–Newton aspects as well as conducting a singular value analysis, we propose a modified family of spectral CG methods. Then, using the approach of Zhang et al. [35], we suggest a descent three–term extension of the given family of CG methods and discuss the global convergence as well. We provide a test bed to shed light on the merits of our modification schemes 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 Wolfe conditions (1.5) and (1.6).

2 A Class of Spectral Conjugate Gradient Methods

In this section, we deal with computing parameters of the spectral CG method with the search directions (1.2). Inspired by the features of quasi–Newton methods, here we let

$$d_{k+1} = -B_{k+1}^{-1} g_{k+1},$$

where B_{k+1} is a symmetric approximation of the Hessian $\nabla^2 f(x_{k+1})$ satisfying the secant condition $B_{k+1} s_k = y_k$ [28]. Now, we can write

$$-B_{k+1}^{-1} g_{k+1} = -\theta_k g_{k+1} + \beta_k d_k,$$

and so,

$$-s_k^T g_{k+1} = -\theta_k s_k^T B_{k+1} g_{k+1} + \beta_k s_k^T B_{k+1} d_k,$$

which under the secant condition yields

$$\beta_k = \theta_k \frac{y_k^T g_{k+1}}{d_k^T y_k} - \frac{s_k^T g_{k+1}}{d_k^T y_k}. \tag{2.1}$$

Now, based on the approach of [12], we suggest the following extended version of (2.1):

$$\beta_k^{\theta,t} = \theta_k \frac{y_k^T g_{k+1}}{d_k^T y_k} - t \frac{s_k^T g_{k+1}}{d_k^T y_k}, \tag{2.2}$$

where t is a nonnegative parameter, being also a two–parameter extension of (1.4). In what follows, based on the singular value analysis carried out in [6], we deal with finding an appropriate value for the parameter θ_k in terms of t.

Firstly, note that from (1.2) and (2.2), search directions of the spectral CG method can be written as

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

in which the matrix P_{k+1}, called the search direction matrix, is defined by

$$P_{k+1} = \theta_k I - \theta_k \frac{s_k y_k^T}{s_k^T y_k} + t \frac{s_k s_k^T}{s_k^T y_k}.$$

Since P_{k+1} presents a rank–two update, its determinant can be computed by

$$\det(P_{k+1}) = \theta_k^{n-1} t \frac{||s_k||^2}{s_k^T y_k}. \tag{2.3}$$

So, if $t > 0$, then P_{k+1} is nonsingular. In light of the similar discussion presented in [6], P_{k+1} has $n-2$ singular values equal to θ_k. Next, we find the two remaining singular values of the matrix P_{k+1}, namely σ_k^+ and σ_k^-.

Since $||P_{k+1}||_F^2$ is equal to $\text{tr}(P_{k+1}^T P_{k+1})$ as well as to sum of squares of all the singular values of P_{k+1} [34], we get

$$\sigma_k^{+2} + \sigma_k^{-2} = t^2 \frac{||s_k||^4}{(s_k^T y_k)^2} + \theta_k^2 \frac{||s_k||^2 ||y_k||^2}{(s_k^T y_k)^2}, \tag{2.4}$$

where $||.||_F$ stands for the Frobenius norm. Also, from (2.3) and since $|\det(P_{k+1})|$ is equal to product of all the singular values of P_{k+1} [33], we have

$$\sigma_k^+ \sigma_k^- = t\theta_k \frac{||s_k||^2}{s_k^T y_k}. \tag{2.5}$$

Now, from (2.4) and (2.5), after some algebraic manipulations we get

$$\begin{aligned} \sigma_k^{\pm} &= \frac{\theta_k}{2} \frac{\sqrt{(\frac{t}{\theta_k}||s_k||^2 + s_k^T y_k)^2 + ||s_k||^2||y_k||^2 - (s_k^T y_k)^2}}{s_k^T y_k} \\ &\quad \pm \frac{\theta_k}{2} \frac{\sqrt{(\frac{t}{\theta_k}||s_k||^2 - s_k^T y_k)^2 + ||s_k||^2||y_k||^2 - (s_k^T y_k)^2}}{s_k^T y_k}. \end{aligned}$$

Next, to make P_{k+1} be a well–conditioned matrix which is advantageous in matrix computations [34], we aim to minimize an upper bound of $\kappa(P_{k+1})$ where $\kappa(.)$ stands for the spectral condition number.

Following carefully the analyses conducted in [6], it can be seen that $0 \leq \sigma_k^- \leq \theta_k \leq \sigma_k^+$. Thus, $||P_{k+1}|| = \sigma_k^+$ and $||P_{k+1}^{-1}|| = {\sigma_k^-}^{-1}$. Hence,

$$\kappa(P_{k+1}) = \frac{\sigma_k^+}{\sigma_k^-} = \frac{\sigma_k^{+2}}{\sigma_k^+ \sigma_k^-} \leq \frac{\sigma_k^{+2} + \sigma_k^{-2}}{\sigma_k^+ \sigma_k^-}, \tag{2.6}$$

and so, from (2.4) and (2.5) we can write

$$\kappa(P_{k+1}) \leq \frac{t}{\theta_k} \frac{||s_k||^2}{s_k^T y_k} + \frac{\theta_k}{t} \frac{||y_k||^2}{s_k^T y_k}.$$

Now, to compute an optimal value for the spectral parameter θ_k in our method, we obtain minimizer of the upper bound of $\kappa(P_{k+1})$ given by (2.6); that is

$$\theta_k^* = t\frac{||s_k||}{||y_k||} = \arg\min_{\theta_k > 0} \left\{ \frac{t}{\theta_k} \frac{||s_k||^2}{s_k^T y_k} + \frac{\theta_k}{t} \frac{||y_k||^2}{s_k^T y_k} \right\}. \tag{2.7}$$

Here, a spectral CG method of the form (1.1)–(1.2) with the parameter $\beta_k^{\theta,t}$ given by (2.2) in which θ_k is computed by (2.7) is called the SDL method. That is,

$$d_{k+1}^{SDL} = -\theta_k^* g_{k+1} + \left(\theta_k^* \frac{g_{k+1}^T y_k}{d_k^T y_k} - t \frac{g_{k+1}^T s_k}{d_k^T y_k} \right) d_k, \ \forall k \geq 0. \tag{2.8}$$

As known, descent condition may be crucial in convergence analysis of the CG methods, also being considered as a desirable effective feature of the methods [10, 20]. Based on this fact and using the approach of [35], here we suggest a descent three–term extension of the given family of CG methods, namely DSDL, in which

$$d_{k+1}^{DSDL} = d_{k+1}^{SDL} - \tau_k \theta_k^* \frac{g_{k+1}^T d_k}{d_k^T y_k} y_k, \ \forall k \geq 0, \tag{2.9}$$

in which $\tau_k = 0$ if $g_{k+1}^T d_{k+1}^{SDL} < 0$, and $\tau_k = 1$, otherwise. So, when $g_{k+1}^T d_{k+1}^{SDL} < 0$, since $\tau_k = 0$ we have $g_{k+1}^T d_{k+1}^{DSDL} = g_{k+1}^T d_{k+1}^{SDL} < 0$. Also, when $g_{k+1}^T d_{k+1}^{SDL} \geq 0$, since $\tau_k = 1$ we have

$$\begin{aligned} g_{k+1}^T d_{k+1}^{DSDL} &= -\theta_k^* ||g_{k+1}||^2 + \theta_k^* \frac{g_{k+1}^T y_k}{d_k^T y_k}(g_{k+1}^T d_k) - t\frac{g_{k+1}^T s_k}{d_k^T y_k}(g_{k+1}^T d_k) \\ &\quad -\theta_k^* \frac{g_{k+1}^T d_k}{d_k^T y_k}(g_{k+1}^T y_k) \leq -\theta_k^* ||g_{k+1}||^2 - t\alpha_k \frac{(g_{k+1}^T d_k)^2}{d_k^T y_k} < 0, \end{aligned}$$

ensuring descent property of the DSDL method. Next, we discuss global convergence of the DSDL method. In this context, the following standard assumptions are needed [32].

Assumption 2.1. (i) The level set $\mathcal{L} = \{x : f(x) \leq f(x_0)\}$ at x_0 is bounded, namely, there exists a constant $\hat{a}$ such that

$$||x|| \leq \hat{a}, \ \forall x \in \mathcal{L}.$$

(ii) In some open convex neighborhood $\mathcal{N}$ of $\mathcal{L}$, the objective function f is continuously differentiable and its gradient is Lipschitz continuous; that is, there exists a positive constant L such that

$$||\nabla f(x) - \nabla f(\tilde{x})|| \leq L||x - \tilde{x}||, \ \forall x, \tilde{x} \in \mathcal{N}.$$

Lemma 2.1 ([32])**.** *Suppose that Assumptions 2.1 hold. Consider any iterative method in the form (1.1), where d_k and α_k satisfy the sufficient descent condition (1.3) and the Wolfe conditions (1.5) and (1.6), respectively. If*

$$\sum_{k=0}^{\infty} \frac{1}{||d_k||^2} = \infty,$$

then the method converges globally in the sense that

$$\liminf_{k\to\infty} ||g_k|| = 0.$$

Theorem 2.2. *Suppose that Assumptions 2.1 hold. Consider the DSDL method and for which assume that θ_k^* and t are bounded above by a positive constant M. If the objective function f is uniformly convex on $\mathcal{N}$, the search directions satisfy the sufficient descent condition (1.3) and the step lengths are determined to fulfill the Wolfe conditions (1.5) and (1.6), then the method converges in the sense that $\lim_{k\to\infty} ||g_k|| = 0$.*

Proof. Under Assumptions 2.1, there exists a positive constant $\bar{\gamma}$ such that

$$||g_k|| \leq \bar{\gamma}, \ \forall x \in \mathcal{L}. \tag{2.10}$$

Uniform convexity of the differentiable function f ensures that there exists a positive constant μ such that

$$d_k^T y_k \geq \mu\alpha_k ||d_k||^2, \ \forall k \geq 0.$$

(See Theorem 1.3.16 of [33].) By (1.2), (2.2) and (2.10) we have

$$\begin{aligned} ||d_{k+1}|| &\leq |\theta_k^*| \ ||g_{k+1}|| + \frac{(2L|\theta_k^*| + t)||g_{k+1}|| \ ||s_k||}{\mu\alpha_k ||d_k||^2} ||d_k|| \\ &\leq \mu^{-1}\bar{\gamma}(\mu|\theta_k^*| + 2L|\theta_k^*| + t) \leq \mu^{-1}\bar{\gamma}M(\mu + 2L + 1). \end{aligned}$$

Now, since $\{||d_k||\}_{k\geq 0}$ is bounded above, from Lemma 2.1 the proof is complete. □

To make θ_k^* and t to be bounded above, we can let $t \leftarrow \min\{t, M\}$ and $\theta_k^* \leftarrow \min\{\theta_k^*, M\}$ where M is an enough large positive constant.

3 Numerical Experiments

In this section, we investigate computational efficiency of the SDL and DSDL methods respectively with the search directions (2.8) and (2.9), with the different six adaptive choices (1.7)–(1.12) for t. Here, for $i = 1, 2, ..., 6$, the method with $\theta_k = 1$ in (2.8) which is called DLi is compared by its spectral versions with $\theta^*_{k_i} = t_{k_i} \frac{||s_k||}{||y_k||}$ which is called SDLi. Similarly, DSDLi is compared by the descent version of the corresponding DL method (the method with $\theta_k = 1$ in (2.9)), called DDLi, $i = 1, 2, ..., 6$.

The runs were performed on a set of 81 unconstrained optimization test problems of the CUTEr collection [18] with the minimum dimension being equal to 50, as given in Table 1. In hardware point of view, we used a computer Intel(R) Core(TM)2 Duo CPU 2.3 GHz with 8 GB of RAM. Also, in software point of view we applied MATLAB 7.7.0.471 (R2008b) on a Centos 6.2 server Linux operation system.

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	10000	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	10000
COSINE	10000	EIGENALS	2550	QUARTC	5000
CRAGGLVY	5000	EIGENBLS	2550	SCHMVETT	5000
CURLY10	10000	EIGENCLS	2652	SENSORS	100
CURLY20	10000	ENGVAL1	5000	SINQUAD	5000
CURLY30	10000	ERRINROS	50	SPARSQUR	10000
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

In the line search procedure, the strong Wolfe conditions have been employed, consisting of (1.5) and the following strengthened version of (1.6):

$$|d_k^T g_{k+1}| \leq -\sigma d_k^T g_k,$$

using Algorithm 3.5 of [28] with $\delta = 0.0001$ and $\sigma = 0.99$. The algorithms were stopped by reaching a maximum of 10000 iterations or achieving a solution with $||g_k|| < 10^{-5}(1 + |f(x_k)|)$. Moreover, efficiency comparisons were drawn using the Dolan–Moré performance profile [14] on the running time and the total number of function and gradient evaluations being equal to $N_f + 3N_g$ [21], where N_f and N_g respectively denote the number of function and gradient evaluations. 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 ω of the best. Figures 1–4 illustrate the results of comparisons (based on the considerations of [19]). As seen, generally the given spectral CG methods as well as their descent versions are practically promising (except the case where DDL6 slightly outperforms DSDL6 in part (f) of Figures 3 and 4). Also, the experiments showed that DSDL2 outperforms the other methods.

In the second part of our numerical experiments, we compared performance of DSDL2 with some recent efficient CG methods suggested by Kou and Dai (KD) [25] as well as the methods IFR (improved Fletcher–Reeves (FR) method [17]), IDY (improved Dai–Yuan method [13]) and IFD (a hybridization of FR and IDY methods) proposed by Jiang and Jian [24]. Results of comparisons are illustrated by Figure 5. As seen, although IFR and IDY outperform DSDL2, performance of DSDL2 is preferable to IFD and KD. So, our descent spectral CG method can be regarded as being somewhat intermediate between the methods of IFR, IDY, IFD and KD.

4 Conclusions

In this study, firstly features of the quasi–Newton method has been employed to achieve a general structure for the spectral conjugate gradient method (as an extension of the Dai–Liao method [12]) and then, following singular value analysis conducted in [6], an optimal value for parameter of the method has been suggested. In order to achieve descent property, an extended version of the method has been proposed based on the approach of Zhang et al. [35]. A brief convergence analysis has been conducted when the line search fulfills the Wolfe conditions. 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 class of spectral conjugate gradient methods can be considered as an appropriate practical tool for solving unconstrained optimization problems.

Acknowledgements

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.

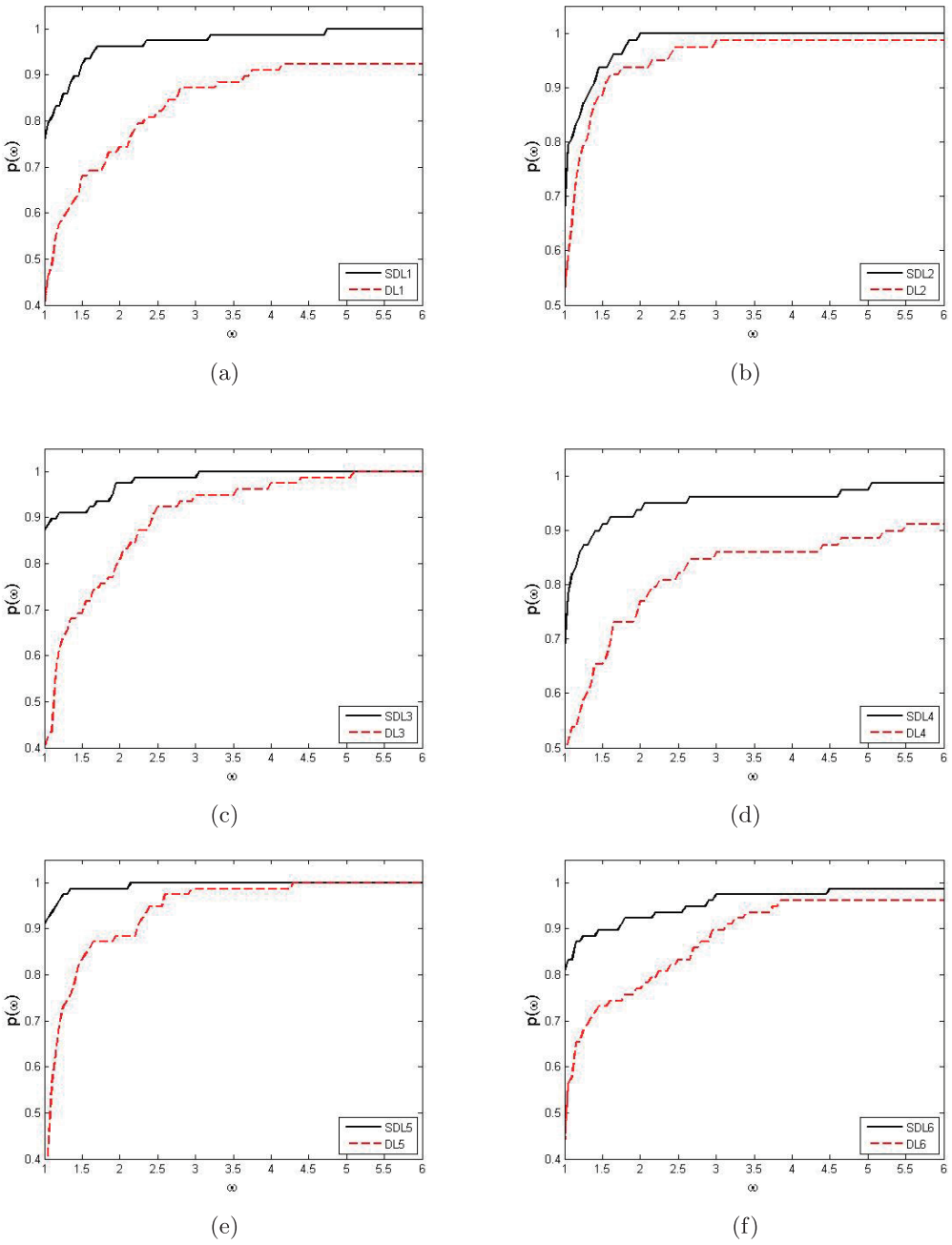

Figure 1: Total number of function and gradient evaluations performance profiles for the SDL and DL methods

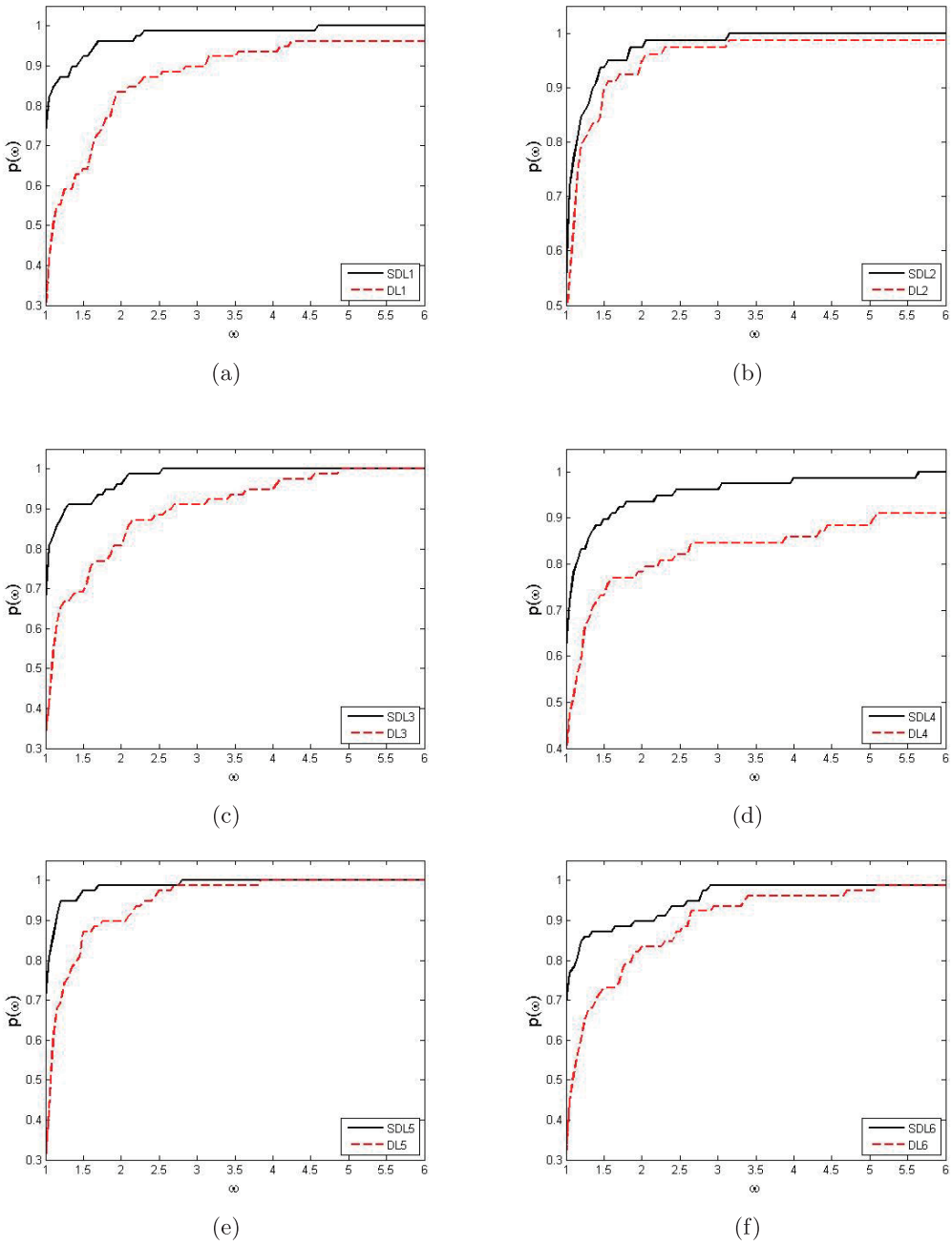

Figure 2: CPU time performance profiles for the SDL and DL methods

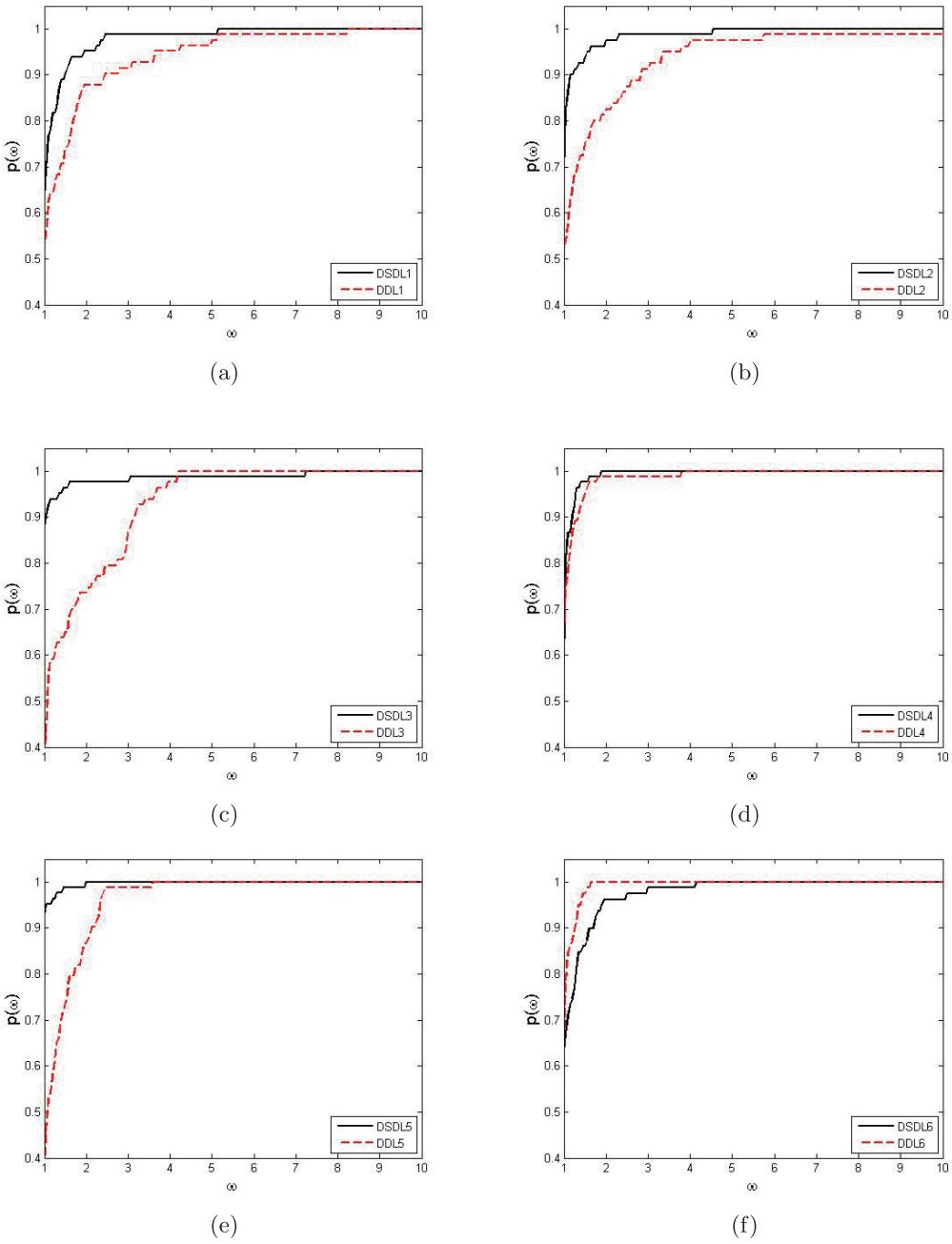

Figure 3: Total number of function and gradient evaluations performance profiles for the DSDL and DDL methods

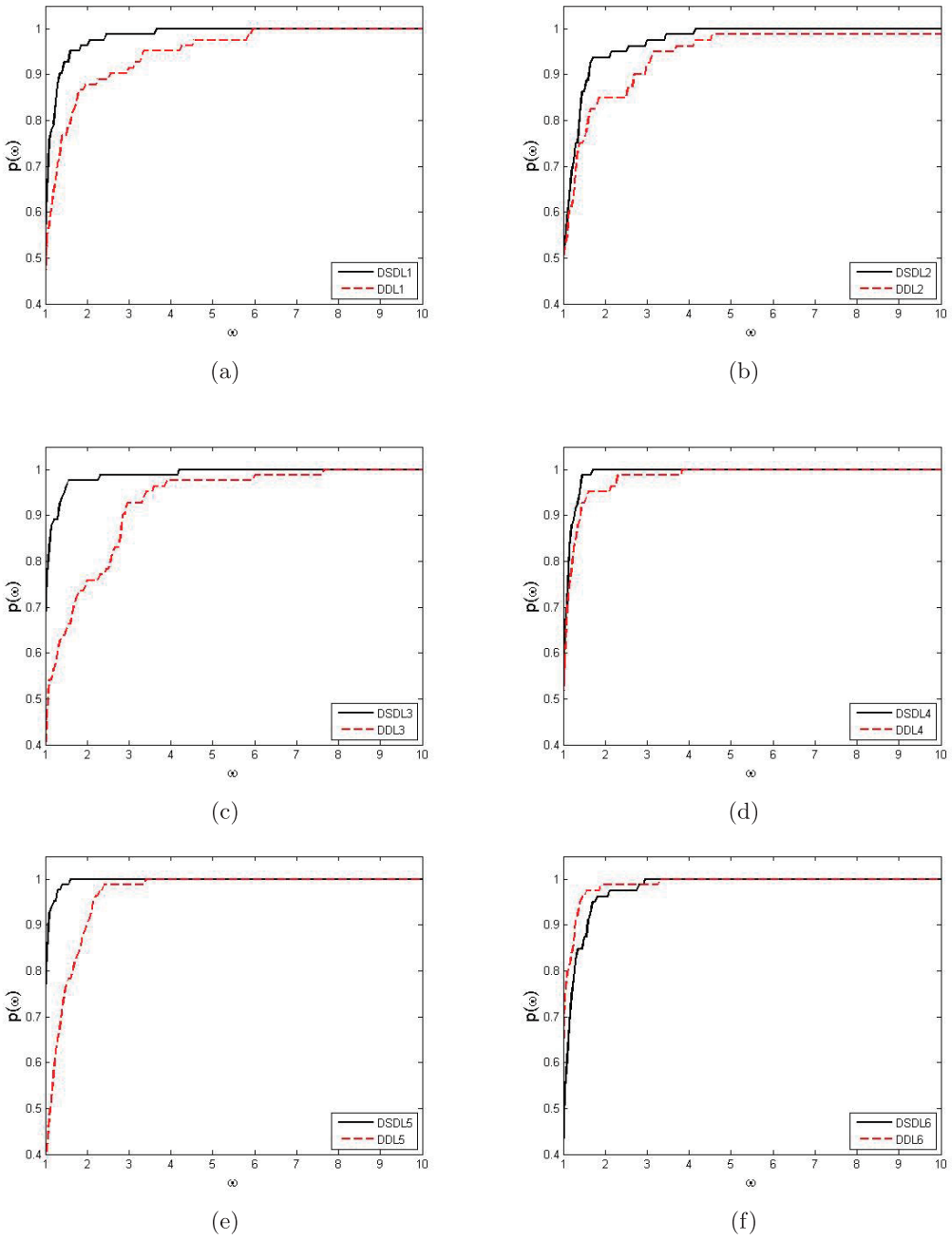

Figure 4: CPU time performance profiles for the DSDL and DDL methods

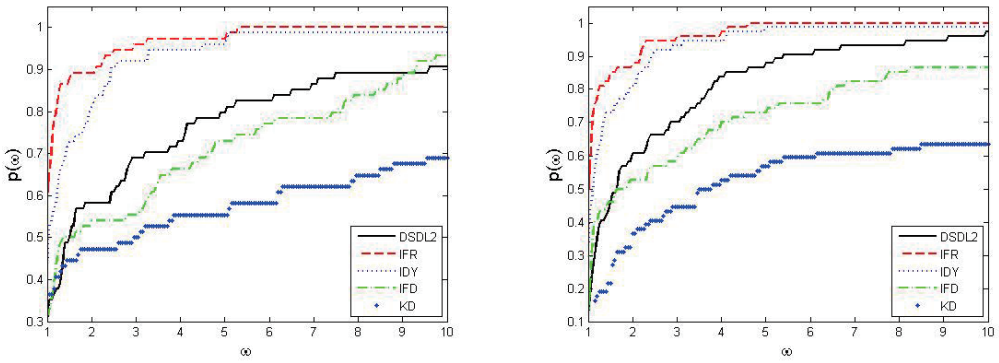

Figure 5: Performance profiles of the total number of function and gradient evaluations (a) and the CPU time (b) for the DSDL2, IFR, IDY, IFD and KD methods

References

[1] Z. Aminifard and S. Babaie–Kafaki, Matrix analyses on the Dai–Liao conjugate gradient method, *ANZIAM J.* 61(2019)195–203.

[2] Z. Aminifard and S. Babaie–Kafaki, A modified descent Polak–Ribiére–Polyak conjugate gradient method with global convergence property for nonconvex functions, *Calcolo* 56 (2019):16.

[3] N. Andrei, A scaled BFGS preconditioned conjugate gradient algorithm for unconstrained optimization, *Appl. Math. Lett.* 20 (2007) 645–650.

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

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

[6] 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.

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

[8] 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.

[9] E. Birgin and J.M. Martínez, A spectral conjugate gradient method for unconstrained optimization, *Appl. Math. Optim.* 43 (2001), 117–128.

[10] Y.H. Dai, New properties of a nonlinear conjugate gradient method, *Numer. Math.* 89 (2001), 83–98.

[11] 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.

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

[13] 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.

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

[15] P. Faramarzi and K. Amini, A modified spectral conjugate gradient method with global convergence, *J. Optim. Theory Appl.* 182 (2019) 667–690.

[16] M. Fatemi, A scaled conjugate gradient method for nonlinear unconstrained optimization, *Optim. Methods Softw.* 32 (2017) 1095–1112.

[17] R. Fletcher and C.M. Reeves, Function minimization by conjugate gradients, *Comput. J.* 7 (1964) 149–154.

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

[19] N.I.M. Gould and J. Scott., A note on performance profiles for benchmarking software, *ACM Trans. Math. Software* 43(2016):Art. 15.

[20] 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.

[21] 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.

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

[23] J. Jian, Q. Chen, X. Jiang, Y. Zeng, and J. Yin, A new spectral conjugate gradient method for large–scale unconstrained optimization, *Optim. Methods Softw.* 32 (2017), 503–515.

[24] X. Jiang and J. Jian, Improved Fletcher–Reeves and Dai–Yuan conjugate gradient methods with the strong Wolfe line search, *J. Comput. Appl. Math.* 348 (2019) 525–534.

[25] C.X. Kou and Y.H. Dai, A modified self–scaling memoryless Broyden–Fletcher–Goldfarb–Shanno method for unconstrained optimization, *J. Optim. Theory Appl.* 165 (2015) 209–224.

[26] 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.

[27] J.K. Liu, Y.M. Feng, and L.M. Zou, A spectral conjugate gradient method for solving large–scale unconstrained optimization, *Comput. Math. Appl.* 77 (2019) 731–739.

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

[29] E. Polak and G. Ribiére, Note sur la convergence de méthodes de directions conjuguées. Rev. Française Informat, *Recherche Opérationnelle* 3 (1969), 35–43.

[30] B.T. Polyak, The conjugate gradient method in extreme problems, *USSR Comp. Math. Math. Phys.* 9 (1969) 94–112.

[31] M. Raydan, The Barzilai and Borwein gradient method for the large–scale unconstrained minimization problem. *SIAM J. Optim.* 7 (1997) 26–33.

[32] 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.

[33] W. Sun and Y.X. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, Springer, New York, 2006.

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

[35] L. Zhang, W. Zhou, and D.H. Li, Some descent three–term conjugate gradient methods and their global convergence, *Optim. Methods Softw.* 22 (2007) 697–711.

Manuscript received 10 August 2019
revised 9 February 2020
accepted for publication 9 February 2020

Zohre Aminifard
Department of Mathematics, Faculty of Mathematics
Statistics and Computer Science, Semnan University
P.O. Box: 35195–363, Semnan, Iran
E-mail address: aminisor@semnan.ac.ir

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