Hindawi Publishing Corporation
Journal of Applied Mathematics
Volume 2012, Article ID 569795, 10 pages
doi:10.1155/2012/569795

Research Article

A Mixed Spectral CD-DY Conjugate Gradient Method

Liu Jinkui,[1] Du Xianglin,[1] and Wang Kairong[2]

[1] *School of Mathematics and Statistics, Chongqing Three Gorges University, Wanzhou 404100, China*
[2] *College of Mathematics and Physics, Chongqing University, Chongqing 401331, China*

Correspondence should be addressed to Liu Jinkui, liujinkui2006@126.com

Received 14 November 2011; Revised 22 January 2012; Accepted 23 January 2012

Academic Editor: Shan Zhao

A mixed spectral CD-DY conjugate descent method for solving unconstrained optimization problems is proposed, which combines the advantages of the spectral conjugate gradient method, the CD method, and the DY method. Under the Wolfe line search, the proposed method can generate a descent direction in each iteration, and the global convergence property can be also guaranteed. Numerical results show that the new method is efficient and stationary compared to the CD (Fletcher 1987) method, the DY (Dai and Yuan 1999) method, and the SFR (Du and Chen 2008) method; so it can be widely used in scientific computation.

1. Introduction

The purpose of this paper is to study the global convergence properties and practical computational performance of a mixed spectral CD-DY conjugate gradient method for unconstrained optimization without restarts, and with appropriate conditions.

Consider the following unconstrained optimization problem:

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

where $f : R^n \to R$ is continuously differentiable and its gradient $g(x) = \nabla f(x)$ is available. Generally, we use the iterative method to solve (1.1), and the iterative formula is given by

$$x_{k+1} = x_k + \alpha_k d_k, \tag{1.2}$$

where x_k is the current iteration, α_k is a positive scalar and called the step-size, which is determined by some line search, d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, & \text{for } k = 1, \\ -g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{1.3}$$

where $g_k = \nabla f(x_k)$ and β_k is a scalar which determines the different conjugate gradient methods [1, 2].

There are many kinds of iterative method that include the steepest descent method, Newton method, and conjugate gradient method. The conjugate direction method is a commonly used and effective method in optimization, and it only needs to use the information of the first derivative. However, it overcomes the shortcoming of the steepest descent method in the slow convergence and avoids the defects of Newton method in storage and computing the second derivative.

The original CD method was proposed by Fletcher [3], in which β_k is defined by

$$\beta_k^{\text{CD}} = -\frac{\|g_k\|^2}{d_{k-1}^T g_{k-1}}, \tag{1.4}$$

where $\|\cdot\|$ denotes the Euclidean norm of vectors. An important property of the CD method is that the method in each iteration will produce a descent direction under the strong Wolfe line search:

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

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma g_k^T d_k, \tag{1.6}$$

where $0 < \delta < \sigma < 1$. Dai and Yuan [4] first proposed the DY method, in which β_k is defined by

$$\beta_k = \frac{\|g_k\|^2}{d_{k-1}^T (g_k - g_{k-1})}. \tag{1.7}$$

Dai and Yuan [4] also strictly proved that the DY method in each iteration produces a descent direction under the Wolfe line search (1.5) and

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k. \tag{1.8}$$

Some good results about the CD method and DY method have also been reported in recent years [5–11].

Quite recently, Birgin and Martinez [12] proposed a spectral conjugate gradient method by combining conjugate gradient method and spectral gradient method. Unfortunately, the spectral conjugate gradient method [12] cannot guarantee to generate descent directions. So, based on the FR formula, Zhang et al. [13] modified the FR method such that the direction generated is always a descent direction. Based on the modified FR conjugate gradient method [13], Du and Chen [14] proposed a new spectral conjugate gradient method:

$$d_k = \begin{cases} -g_k, & \text{for } k = 1, \\ -\theta_k g_k + \beta_k^{\mathrm{FR}} d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{1.9}$$

where

$$\beta_k^{\mathrm{FR}} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2}, \qquad \theta_k = \frac{d_k^T (g_k - g_{k-1})}{\|g_{k-1}\|^2}. \tag{1.10}$$

And they proved the global convergence of the modified spectral FR method (In this paper, we call it SFR method) with the mild conditions.

The observation of the above formula motivates us to construct a new formula; which combines the advantage of the spectral gradient method, CD method, and DY method as follows:

$$d_k = \begin{cases} -g_k, & \text{for } k = 1, \\ -\theta_k g_k + \beta_k d_{k-1}, & \text{for } k \geq 2, \end{cases} \tag{1.11}$$

where β_k is specified by

$$\beta_k = \beta_k^{\mathrm{CD}} + \min\left\{0, \varphi_k \cdot \beta_k^{\mathrm{CD}}\right\}, \tag{1.12}$$

$$\theta_k = 1 - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}, \tag{1.13}$$

$$\varphi_k = -\frac{g_k^T d_{k-1}}{d_{k-1}^T (g_k - g_{k-1})}. \tag{1.14}$$

And under some mild conditions, we give the global convergence of the mixed spectral CD-DY conjugate gradient method with the Wolfe line search.

This paper is organized as follows. In Section 2, we propose the corresponding algorithm and give some assumptions and lemmas, which are usually used in the proof of the global convergence properties of nonlinear conjugate gradient methods. In Section 3, global convergence analysis is provided with suitable conditions. Preliminary numerical results are presented in Section 4.

2. Algorithm and Lemmas

In order to establish the global convergence of the proposed method, we need the following assumption on objective function, which have been used often in the literature to analyze the global convergence of nonlinear conjugate gradient methods with inexact line search.

Assumption 2.1. (i) The level set $\Omega = \{x \mid f(x) \le f(x_1)\}$ is bounded, where x_1 is the starting point.

(ii) In some neighborhood N of Ω, the objective function is continuously differentiable, and its gradient is Lipchitz continuous, that is, there exists a constant $L > 0$ such that

$$\|g(x) - g(y)\| \le L\|x - y\|, \quad \forall x, y \in N. \tag{2.1}$$

Now we give the mixed spectral CD-DY conjugate gradient method as follows.

Algorithm 2.2.

Step 1. Data $x_1 \in R^n$, $\varepsilon \ge 0$. Set $d_1 = -g_1$; if $\|g_1\| \le \varepsilon$, then stop.

Step 2. Compute α_k by the strong Wolfe line search (1.5) and (1.8).

Step 3. Let $x_{k+1} = x_k + \alpha_k d_k, g_{k+1} = g(x_{k+1})$; if $\|g_{k+1}\| \le \varepsilon$, then stop.

Step 4. Compute β_{k+1} by (1.12), and generate d_{k+1} by (1.11).

Step 5. Set $k = k + 1$; go to Step 2.

The following lemma shows that Algorithm 2.2 produces a descent direction in each iteration with the Wolfe line search.

Lemma 2.3. *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.2, and let the step-size α_k be determined by the Wolfe line search* (1.5) *and* (1.8)*, then*

$$g_k^T d_k < 0. \tag{2.2}$$

Proof. The conclusion can be proved by induction. Since $g_1^T d_1 = -\|g_1\|^2$, the conclusion holds for $k = 1$. Now we assume that the conclusion is true for $k - 1$, $k \ge 2$. Then from (1.8), we have

$$d_{k-1}^T (g_k - g_{k-1}) \ge (\sigma - 1) g_{k-1}^T d_{k-1} > 0. \tag{2.3}$$

If $g_k^T d_{k-1} \le 0$, then from (1.14) and (2.3), we have $\beta_k = \beta_k^{\mathrm{CD}}$.

From (1.4), (1.11), and (1.13), we have

$$g_k^T d_k = -\left(1 - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}\right) \cdot \|g_k\|^2 - \frac{\|g_k\|^2}{g_{k-1}^T d_{k-1}} \cdot g_k^T d_{k-1} = -\|g_k\|^2 < 0. \tag{2.4}$$

If $g_k^T d_{k-1} > 0$, then from (1.14) and (2.3), we have $\beta_k = \beta_k^{\mathrm{CD}} + \varphi_k \cdot \beta_k^{\mathrm{CD}} = \beta_k^{\mathrm{DY}}$.

From (1.11), (1.7), and (1.13), we have

$$\begin{aligned} g_k^T d_k &= -\theta_k \cdot \|g_k\|^2 + \beta_k^{\mathrm{DY}} \cdot g_k^T d_{k-1} \\ &= \beta_k^{\mathrm{DY}} \cdot \left[-\theta_k \cdot d_{k-1}^T (g_k - g_{k-1}) + g_k^T d_{k-1}\right] \\ &= \beta_k^{\mathrm{DY}} \cdot \left[g_{k-1}^T d_{k-1} + \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}} \cdot d_{k-1}^T (g_k - g_{k-1})\right] \\ &\le \beta_k^{\mathrm{DY}} \cdot g_{k-1}^T d_{k-1} < 0. \end{aligned} \tag{2.5}$$

From the above inequality (2.4) and (2.5), we obtain that the conclusion holds for k. □

The conclusion of the following lemma, often called the Zoutendijk condition, is used to prove the global convergence properties of nonlinear conjugate gradient methods. It was originally given by Zoutendijk [15].

Lemma 2.4 (see [15]). *Suppose that Assumption 2.1 holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.2, and let the step-size α_k be determined by the Wolfe line search* (1.5) *and* (1.8), *and Lemma 2.3 holds. Then*

$$\sum_{k \ge 1} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty. \tag{2.6}$$

Lemma 2.5. *Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.2, and let the step-size α_k be determined by the Wolfe line search* (1.5) *and* (1.8), *and Lemma 2.3 holds. Then*

$$\beta_k \le \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}. \tag{2.7}$$

Proof. If $g_k^T d_{k-1} \le 0$, then from Lemma 2.3, we have $\beta_k = \beta_k^{\mathrm{CD}}$. From (1.11), (1.4), and (1.13), we have

$$\begin{aligned} g_k^T d_k &= -\left(1 - \frac{g_k^T d_{k-1}}{g_{k-1}^T d_{k-1}}\right) \cdot \|g_k\|^2 + \beta_k^{\mathrm{CD}} \cdot g_k^T d_{k-1} \\ &= \beta_k^{\mathrm{CD}} \cdot \left(g_{k-1}^T d_{k-1} - g_k^T d_{k-1}\right) + \beta_k^{\mathrm{CD}} \cdot g_k^T d_{k-1} \\ &= \beta_k^{\mathrm{CD}} \cdot g_{k-1}^T d_{k-1}. \end{aligned} \tag{2.8}$$

From the above equation, we have

$$\beta_k^{CD} = \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}. \tag{2.9}$$

If $g_k^T d_{k-1} > 0$, from (2.5), we have

$$\beta_k^{DY} \le \frac{g_k^T d_k}{g_{k-1}^T d_{k-1}}. \tag{2.10}$$

From (1.12), (2.9), and (2.10), we obtain that the conclusion (2.7) holds. □

3. Global Convergence Property

The following theorem proves the global convergence of the mixed spectral CD-DY conjugate gradient method with the Wolfe line search.

Theorem 3.1. *Suppose that Assumption 2.1 holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be generated by Algorithm 2.2, and let the step-size α_k be determined by the Wolfe line search* (1.5) *and* (1.8), *and Lemma 2.3 holds. Then*

$$\liminf_{k \to +\infty} \|g_k\| = 0. \tag{3.1}$$

Proof. Suppose by contradiction that there exists a positive constant $\rho > 0$, such that

$$\|g_k\| \ge \rho \tag{3.2}$$

holds for all $k \ge 1$. □

From (1.11), we have $d_k + \theta_k g_k = \beta_k d_{k-1}$, and by squaring it, we get

$$\|d_k\|^2 = \beta_k^2 \|d_{k-1}\|^2 - 2\theta_k g_k^T d_k - \theta_k^2 \|g_k\|^2. \tag{3.3}$$

From the above equation and Lemma 2.5, we have

$$\|d_k\|^2 \le \left(\frac{g_k^T d_k}{g_{k-1}^T d_{k-1}} \right)^2 \cdot \|d_{k-1}\|^2 - 2\theta_k g_k^T d_k - \theta_k^2 \|g_k\|^2. \tag{3.4}$$

Dividing the above inequality by $(g_k^T d_k)^2$, we have

$$\begin{aligned}
\frac{\|d_k\|^2}{(g_k^T d_k)^2} &\le \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} - \frac{2\theta_k}{g_k^T d_k} - \theta_k^2 \cdot \frac{\|g_k\|^2}{(g_k^T d_k)^2} \\
&= \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} - \left(\theta_k \cdot \frac{\|g_k\|}{g_k^T d_k} + \frac{1}{\|g_k\|}\right)^2 + \frac{1}{\|g_k\|^2} \\
&\le \frac{\|d_{k-1}\|^2}{(g_{k-1}^T d_{k-1})^2} + \frac{1}{\|g_k\|^2}.
\end{aligned} \tag{3.5}$$

Using (3.5) recursively and noting that $\|d_1\|^2 = -g_1^T d_1 = \|g_1\|^2$, we get

$$\frac{\|d_k\|^2}{(g_k^T d_k)^2} \le \sum_{i=1}^{k} \frac{1}{\|g_i\|^2}. \tag{3.6}$$

Then from (3.2) and (3.6), we have

$$\frac{(g_k^T d_k)^2}{\|d_k\|^2} \ge \frac{\rho^2}{k}, \tag{3.7}$$

which indicates

$$\sum_{k\ge1} \frac{1}{\|g_k\|^2} = +\infty. \tag{3.8}$$

The above equation contradicts the conclusion of Lemma 2.4. Therefore, the conclusion (3.1) holds.

4. Numerical Experiments

In this section, we report some numerical results. We used MATLAB 7.0 to test some problems that are from [16] and compare the performance of the mixed spectral CD-DY method (Algorithm 2.2) with the CD method, DY method, and SFR method. The global convergence of the CD method has not still been proved under the Wolfe line search, so our line search subroutine computes α_k such that the strong Wolfe line search conditions hold with $\delta = 0.01$ and $\sigma = 0.1$. We also use the condition $\|g_k\| \le 10^{-6}$ or It-max > 9999 as the stopping criterion. It-max denotes the maximal number of iterations.

The numerical results of our tests are reported in the following table. The first column "Problem" represents the name of the tested problem in [16]. "Dim" denotes the dimension of the tested problem. The detailed numerical results are listed in the form NI/NF/NG, where NI, NF, and NG denote the number of iterations, function evaluations, and gradient evaluations respectively. If the limit of 9999 function evaluations was exceeded, the run was stopped; this is indicated by "—".

In order to rank the average performance of all above methods, one can compute the total number of function and gradient evaluation by the formula

$$N_{\text{total}} = \text{NF} + l * \text{NG}, \tag{4.1}$$

where l is some integer. According to the results on automatic differentiation [17, 18], the value of l can be set to 5

$$N_{\text{total}} = \text{NF} + 5 * \text{NG}. \tag{4.2}$$

That is to say, one gradient evaluation is equivalent to five function evaluations if automatic differentiation is used.

By making use of (4.2), we compare the mixed spectral CD-DY method as follows: for the ith problem, compute the total number of function evaluations and gradient evaluations required by the CD method, the DY method, the SFR method, and the mixed spectral CD-DY method by formula (4.2), and denote them by $N_{\text{total},i}(\text{CD})$, $N_{\text{total},i}(\text{DY})$, $N_{\text{total},i}(\text{SFR})$, and $N_{\text{total},i}(\text{CD-DY})$; then calculate the ratios:

$$\begin{aligned} \gamma_i(\text{CD}) &= \frac{N_{\text{total},i}(\text{CD})}{N_{\text{total},i}(\text{CD-DY})}, \\ \gamma_i(\text{DY}) &= \frac{N_{\text{total},i}(\text{DY})}{N_{\text{total},i}(\text{CD-DY})}, \\ \gamma_i(\text{SFR}) &= \frac{N_{\text{total},i}(\text{SFR})}{N_{\text{total},i}(\text{CD-DY})}. \end{aligned} \tag{4.3}$$

From Table 1, we know that some problems are not run by some methods. So, if the i_0th problem is not run by the given method, we use a constant $\tau = \max\{\gamma_i(\text{the given method}) \mid i \in S_1\}$ instead of $\gamma_{i_0}(\text{the given method})$, where S_1 denotes the set of test problems, which can be run by the given method.

The geometric mean of these ratios for the CD method, the DY method and SFR method, over all the test problems is defined by

$$\begin{aligned} \gamma(\text{CD}) &= \left(\prod_{i \in S} \gamma_i(\text{CD})\right)^{1/|S|}, \\ \gamma(\text{DY}) &= \left(\prod_{i \in S} \gamma_i(\text{DY})\right)^{1/|S|}, \\ \gamma(\text{SFR}) &= \left(\prod_{i \in S} \gamma_i(\text{SFR})\right)^{1/|S|}, \end{aligned} \tag{4.4}$$

where S denotes the set of the test problems and $|S|$ denotes the number of elements in S. One advantage of the above rule is that, the comparison is relative and hence is not be dominated a few problems for which the method requires a great deal of function evaluations and gradient functions.

Table 1: The performance of the CD method, DY method, CD-DY method, and SFR method.

Problem	Dim	CD	DY	CD-DY	SFR
ROSE	2	88/250/223	64/190/172	60/188/168	64/190/172
FROTH	2	—	42/168/138	38/151/123	—
BADSCP	2	682/1885/1637	—	2704/6550/6438	—
BADSCB	2	272/941/800	—	726/2051/1786	—
BEALE	2	73/177/155	75/186/164	68/175/145	75/186/164
JENSAM	2 ($m = 6$)	17/61/43	10/49/26	26/80/57	15/48/32
HELIX	3	56/157/132	37/118/98	50/145/120	37/118/98
BRAD	3	75/224/189	66/208/177	37/120/98	66/208/177
SING	4	454/1074/1009	2286/4555/4545	850/1894/1863	1476/2901/2891
WOOD	4	184/438/399	100/291/240	139/396/337	100/291/240
KOWOSB	4	173/516/449	536/1449/1271	144/421/365	504/1386/1211
BD	4	43/169/132	39/158/121	28/144/113	37/152/116
WATSON	5	89/279/239	127/348/299	158/438/373	128/352/304
BIGGS	6	200/579/509	294/824/712	236/680/599	288/812/703
OSB2	11	3243/5413/5398	7006/11059/11048	584/1262/1226	7013/11102/11091
VAEDIM	5	6/57/38	6/57/38	6/57/38	6/57/38
	10	7/81/52	7/81/52	7/81/52	7/81/52
PEN1	50	2209/2565/2536	1727/2117/2043	116/221/190	1727/2117/2043
	100	62/223/182	31/157/121	31/167/131	31/157/121
TRIG	100	—	305/399/398	88/145/144	305/399/398
	500	—	343/424/423	109/189/188	344/427/425
ROSEX	500	92/267/238	68/207/186	65/205/182	68/207/186
	1000	98/287/255	68/207/186	65/205/182	68/207/186
SINGX	100	682/1593/1517	1488/3139/3073	1159/2366/2614	2411/5326/5084
	1000	511/1245/1135	2092/4451/4321	1374/3104/3064	5213/10042/10032
BV	500	1950/2543/2542	4796/6823/6822	1311/2131/2130	4784/6793/6792
	1000	632/833/832	414/449/448	414/449/448	414/449/448
IE	500	7/15/8	7/15/8	7/15/8	7/15/8
	1000	7/15/8	7/15/8	7/15/8	7/15/8
TRID	500	52/112/107	49/106/101	43/94/89	49/106/101
	1000	70/149/145	64/137/133	55/119/115	64/137/133

Table 2: Relative efficiency of the CD, DY, SFR, and the mixed spectral CD-DY methods.

CD	DY	SFR	CD-DY
1.3956	1.6092	1.6580	1

According to the above rule, it is clear that $\gamma(\text{CD-DY}) = 1$. From Table 2, we can see that average performance of the mixed spectral CD-DY conjugate gradient method (Algorithm 2.2) works the best. So, the mixed spectral CD-DY conjugate gradient method has some practical values.

Acknowledgments

The authors wish to express their heartfelt thanks to the referees and the editor for their detailed and helpful suggestions for revising the paper. This work was supported by The Nature Science Foundation of Chongqing Education Committee (KJ091104).

References

[1] G. H. Yu, *Nonlinear self-scaling conjugate gradient methods for large-scale optimization problems*, Doctorial thesis, Sun Yat-Sen University, Guangzhou, China, 2007.

[2] G. Yuan and Z. Wei, "New line search methods for unconstrained optimization," *Journal of the Korean Statistical Society*, vol. 38, no. 1, pp. 29–39, 2009.

[3] R. Fletcher, *Practical Methods of Optimization. Volume 1: Unconstrained Optimization*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1987.

[4] Y. H. Dai and Y. Yuan, "A nonlinear conjugate gradient method with a strong global convergence property," *SIAM Journal on Optimization*, vol. 10, no. 1, pp. 177–182, 1999.

[5] B. Qu, G. F. Hu, and X. C. Zhang, "A global convergence result for the conjugate descent method," *Journal of Dalian University of Technology*, vol. 42, no. 1, pp. 13–16, 2002.

[6] X. W. Du, L. Q. Ye, and C. X. Xu, "Global convergence of a class of unconstrained optimal methods include the conjugate descent method," *Journal of Engineering Mathematics*, vol. 18, no. 2, pp. 120–122, 2001.

[7] C. Y. Pan and L. P. Chen, "A class of efficient new descent methods," *Acta Mathematicae Applicatae Sinica*, vol. 30, no. 1, pp. 88–98, 2007.

[8] Y. Dai and Y. Yuan, "Convergence properties of the conjugate descent method," *Advances in Mathematics*, vol. 25, no. 6, pp. 552–562, 1996.

[9] Y. H. Dai, "New properties of a nonlinear conjugate gradient method," *Numerische Mathematik*, vol. 89, no. 1, pp. 83–98, 2001.

[10] Y. H. Dai and Y. Yuan, "A class of globally convergent conjugate gradient methods," Research Report ICM-98-030, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, Bejing, China, 1998.

[11] Y.-H. Dai, "Convergence of nonlinear conjugate gradient methods," *Journal of Computational Mathematics*, vol. 19, no. 5, pp. 539–548, 2001.

[12] E. G. Birgin and J. M. Martinez, "A spectral conjugate gradient method for unconstrained optimization," *Applied Mathematics and Computation*, vol. 180, pp. 46–52, 2006.

[13] L. Zhang, W. Zhou, and D. Li, "Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search," *Numerische Mathematik*, vol. 104, no. 4, pp. 561–572, 2006.

[14] S.-Q. Du and Y.-Y. Chen, "Global convergence of a modified spectral FR conjugate gradient method," *Applied Mathematics and Computation*, vol. 202, no. 2, pp. 766–770, 2008.

[15] G. Zoutendijk, "Nonlinear programming, computational methods," in *Integer and Nonlinear Programming*, J. Abadie, Ed., pp. 37–86, North-Holland, 1970.

[16] J. J. More, B. S. Garbow, and K. E. Hillstrom, "Testing unconstrained optimization software," *ACM Transactions on Mathematical Software*, vol. 7, no. 1, pp. 17–41, 1981.

[17] Y.-H. Dai and Q. Ni, "Testing different conjugate gradient methods for large-scale unconstrained optimization," *Journal of Computational Mathematics*, vol. 21, no. 3, pp. 311–320, 2003.

[18] A. Griewank, "On automatic differentiation," in *Mathematical Programming: Recent Developments and Applications*, M. Iri and K. Tannabe, Eds., pp. 84–108, Kluwer Academic Publishers, 1989.