Computational experience with modified conjugate gradient methods for unconstrained optimization

Lukšan, Ladislav
2008

Institute of Computer Science
Academy of Sciences of the Czech Republic

Computational experience with modified conjugate gradient methods for unconstrained optimization

L.Lukšan, C.Matonoha, J.Vlček

Technical report No. 1038

December 2008

Pod Vodárenskou věží 2, 182 07 Prague 8 phone: +420 2 688 42 44, fax: +420 2 858 57 89,
e-mail:e-mail:ics@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Republic

Computational experience with modified conjugate gradient methods for unconstrained optimization

L.Lukšan, C.Matonoha, J.Vlček [1]

Technical report No. 1038

December 2008

Abstract:

In this report, several modifications of the nonlinear conjugate gradient method are described and investigated. Theoretical properties of these modifications are proved and their practical performance is demonstrated using extensive numerical experiments.

Keywords:
Numerical optimization, conjugate direction methods, conjugate gradient methods, global convergence, numerical experiments.

[1]This work was supported by the Grant Agency of the Czech Academy of Sciences, project No. IAA1030405, the Grant Agency of the Czech Republic, project No. 201/06/P397, and the institutional research plan No. AV0Z10300504. L. Lukšan is also from Technical University of Liberec, Hálkova 6, 461 17 Liberec.

1 Introduction

Conjugate gradient methods are widely studied and used for unconstrained minimization of function $F : R^n \to R$, see [1]–[75], [77]–[117], [120]–[133]. These methods are descent direction methods. It means that after introducing a starting approximation $x_1 \in R^n$ they generate a sequence of points $\{x_i\} \subset R^n$ by the rule

$$x_{i+1} = x_i + \alpha_i s_i, \quad i \in N, \tag{1}$$

where $s_i \in R^n$ is a direction vector satisfying a descent condition $s_i^T g(x_i) < 0$ ($g(x_i)$ is a gradient of the function F at a point x_i), and $\alpha_i > 0$ is a step-length chosen in such a way that the generalized Wolfe conditions

$$F(x_{i+1}) - F(x_i) \le \varepsilon_1 \alpha_i s_i^T g(x_i), \tag{2}$$

$$\varepsilon_2 s_i^T g(x_i) \le s_i^T g(x_{i+1}) \le \varepsilon_3 |s_i^T g(x_i)|, \tag{3}$$

with $0 < \varepsilon_1 < \varepsilon_2 < 1$ and $\varepsilon_3 \ge 0$, are satisfied, see [118]–[119]. If $\varepsilon_3 = \varepsilon_2$, we speak about the strong Wolfe conditions and if $\varepsilon_3 = \infty$, we speak about the weak Wolfe conditions. In case that $s_i^T g(x_{i+1}) = 0$, the step length is exact, otherwise is inexact.

We will use a shortened notation $F_i = F(x_i)$, $g_i = g(x_i)$, $G_i = G(x_i)$, $i \in N$, where $G(x_i)$ is a Hessian matrix of the function F at a point x_i. We will assume that function F is twice continuously differentiable and satisfies the conditions

$$F(x_i) \ge \underline{F} \qquad \forall x_i \in R^n, \tag{4}$$

$$\|G(x_i)\| \le \overline{G} \qquad \forall x_i \in R^n, \tag{5}$$

where $\underline{F}$ and $\overline{G}$ are suitable constants. We will investigate methods where vectors $s_i \in R^n$, $i \in N$, satisfy the condition

$$-s_i^T g_i \ge \varepsilon_0 \|g_i\|^2, \tag{6}$$

with $\varepsilon_0 > 0$. It can be shown (see e.g. [24]) that this condition implies the inequality

$$\sum_{i=1}^{\infty} \frac{\|g_i\|^4}{\|s_i\|^2} < \infty. \tag{7}$$

Definition 1 *We say that a descent direction method is a conjugate gradient method if*

$$s_1 = -g_1 \quad and \quad s_{i+1} = -g_{i+1} + \beta_i s_i \quad for \quad i \in N, \tag{8}$$

where a parameter β_i is chosen so that the direction vectors s_i, $1 \le i \le n$, were mutually G–orthogonal, i.e. $s_j^T G s_i = 0$, $1 \le j < i \le n$, if we apply this method to a strictly convex quadratic function

$$Q(x) = \frac{1}{2}(x - x^*)^T G (x - x^*)$$

with an exact choice of a step-length.

If we denote $d_i = x_{i+1} - x_i = \alpha_i s_i$ and $y_i = g_{i+1} - g_i$, then for the quadratic function Q we have $y_i = G d_i$ and the G–orthogonality condition of vectors s_i, s_{i+1} can be written as $\alpha_i s_i^T G s_{i+1} = y_i^T s_{i+1} = 0$ (we assume that $\alpha_i \neq 0$). This together with (8) lead to the equation $\beta_i y_i^T s_i - y_i^T g_{i+1} = 0$ or

$$\beta_i = \frac{y_i^T g_{i+1}}{y_i^T s_i}. \tag{9}$$

It can be shown (see e.g. [107]) that this value already assures mutual orthogonality of direction vectors and gradients and finding a minimum of a strictly convex quadratic function after a finite number of steps if a step-length is exact. We call this property a quadratic termination.

If the step-length is exact, we can write by (8)

$$y_i^T s_i = g_{i+1}^T s_i - g_i^T s_i = -g_i^T s_i = g_i^T g_i - \beta_{i-1} g_i^T s_{i-1} = g_i^T g_i.$$

Moreover, if the minimized function is quadratic, then its gradients are mutually orthogonal, and so

$$y_i^T g_{i+1} = g_{i+1}^T g_{i+1} - g_i^T g_{i+1} = g_{i+1}^T g_{i+1}.$$

This fact implies that we can use three different denominators and two different numerators in expression (9) without violation a quadratic termination property. Thus we obtain six basic conjugate gradient methods:

$$\beta_i^{HS} = \frac{y_i^T g_{i+1}}{y_i^T s_i}, \quad \beta_i^{PR} = \frac{y_i^T g_{i+1}}{g_i^T g_i}, \quad \beta_i^{LS} = \frac{y_i^T g_{i+1}}{|g_i^T s_i|} \tag{10}$$

(HS – Hestenes and Stiefel [62], PR – Polak and Ribiére [94], LS – Liu and Storey [74]),

$$\beta_i^{DY} = \frac{g_{i+1}^T g_{i+1}}{y_i^T s_i}, \quad \beta_i^{FR} = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i}, \quad \beta_i^{CD} = \frac{g_{i+1}^T g_{i+1}}{|g_i^T s_i|} \tag{11}$$

(DY – Dai and Yuan [32], FR – Fletcher and Reeves [50], CD – conjugate descent [49]).

These methods can be divided into two groups by the numerator used. Methods of the first group (HS, PR, LS) are more suitable for practical computations but they are globally convergent only with necessary modifications. Methods of the second group (DY, FR, CD) are globally convergent under certain assumptions (put on a choice of a step-length) but the direction vectors stay worse conjugate if a step-length is inexact and the minimized function is not quadratic.

The properties of methods (10) can be improved by eliminating negative values, so

$$\beta_i^{HS+} = \max(0, \beta_i^{HS}), \quad \beta_i^{PR+} = \max(0, \beta_i^{PR}), \quad \beta_i^{LS+} = \max(0, \beta_i^{LS}). \tag{12}$$

We use nonnegative values in order to prevent possible cycling [96]. Methods (10) can also be combined with methods (11). Such combined methods use relations

$$
\begin{aligned}
\beta_i^{HSC} &= \max(0, \min(\beta_i^{HS}, \beta_i^{DY})), \\
\beta_i^{PRC} &= \max(0, \min(\beta_i^{PR}, \beta_i^{FR})), \\
\beta_i^{LSC} &= \max(0, \min(\beta_i^{LS}, \beta_i^{CD})).
\end{aligned} \tag{13}
$$

Combined methods HSC, PRC, LSC are globally convergent under the same conditions as methods DY, FR, CD. Furthermore, they are efficient for practical computations because values β_i^{HS}, β_i^{PR}, β_i^{LS} are used sufficiently often.

2 Modification of conjugate gradient methods

Relation (8) can be variously modified in order to improve effectiveness of conjugate gradient methods. It is usually performed so that we add terms proportional to $s_i^T g_{i+1}$, which vanish in case of the exact choice of a step-length and the quadratic termination property stays unchanged. One possibility, used in [130], is to replace (8) with

$$
s_1 = -g_1 \quad and \quad s_{i+1} = -\left(1 + \beta_i \frac{g_{i+1}^T s_i}{g_{i+1}^T g_{i+1}}\right) g_{i+1} + \beta_i s_i \quad for \quad i \in N, \tag{14}
$$

where β_i is one of the values in (10) or (11).

Theorem 1 *Modified conjugate gradient method (14) has the quadratic termination property. Moreover, for $i \in N$ we have*

$$
-g_{i+1}^T s_{i+1} = g_{i+1}^T g_{i+1}. \tag{15}
$$

Proof If a step-length is exact, one has $g_{i+1}^T s_i = 0$, so (14) will change into (8) and the quadratic termination property will stay unchanged. Multiplying (14) by a vector g_{i+1}, we obtain equality (15). □

If we substitute the value β_i^{CD} into (14), we will get $s_{i+1} = -\vartheta_i^{CD} g_{i+1} + \beta_i^{CD} s_i$, where $\vartheta_i^{CD} = -y_i^T s_i / g_i^T s_i$. Method FR can be modified in a similar way, see [131]. These modifications allow to weaken substantially conditions for global convergence.

Theorem 2 *Consider modified methods DY, FR, CD given by the rule*

$$
s_1 = -g_1 \quad and \quad s_{i+1} = -\vartheta_i g_{i+1} + \beta_i s_i \quad for \quad i \in N, \tag{16}
$$

where the values β_i^{DY}, β_i^{FR}, β_i^{CD} are determined by (11) and

$$
\vartheta_i^{DY} = \frac{y_i^T s_i}{y_i^T s_i} = 1, \quad \vartheta_i^{FR} = \frac{y_i^T s_i}{g_i^T g_i}, \quad \vartheta_i^{CD} = \frac{y_i^T s_i}{|g_i^T s_i|}. \tag{17}
$$

These methods have the quadratic termination property. If a function $F : R^n \to R$ satisfies conditions (4)–(5) and if we use generalized Wolfe conditions (2)–(3) with $0 < \varepsilon_1 < \varepsilon_2 < 1$ and $0 \le \varepsilon_3 < \infty$ during a choice of a step-length, then these methods are globally convergent.

Proof If a step-length is exact, then $y_i^T s_i = -g_i^T s_i = g_i^T g_i$, or $\vartheta_i^{DY} = \vartheta_i^{FR} = \vartheta_i^{CD} = 1$, so (16) changes into (8) and the quadratic termination property stays unchanged. Now we will prove global convergence.

(a) Since $\vartheta_i^{DY} = 1$, method DY is unchanged using (16), so global convergence follows from the theorem proved in [32].

(b) For modified method FR we have

$$g_{i+1}^T s_{i+1} = -y_i^T s_i \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i} + \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i} g_{i+1}^T s_i = \frac{g_{i+1}^T g_{i+1}}{g_i^T g_i} g_i^T s_i < 0.$$

Since $g_1^T s_1 = -g_1^T g_1$, we obtain equality (15) with sequential substituting into the previous relation (by induction). Thus modified method FR is identical to modified method CD and equality (15) is fulfilled for both methods.

(c) For modified method CD, equality (15) is fulfilled. Therefore, direction vectors s_i, $i \in N$, are descent and (6) holds with $\varepsilon_0 = 1$, which implies inequality (7). Because generalized Wolfe conditions (2)–(3) are used during a choice of a step-length, we have

$$y_i^T s_i = g_{i+1}^T s_i - g_i^T s_i \le \varepsilon_3 |g_i^T s_i| - g_i^T s_i = (1+\varepsilon_3)|g_i^T s_i|,$$

or $\vartheta_i \le 1+\varepsilon_3$. If we use this estimate together with relations (15)–(17), we can write

$$\begin{aligned} \|s_{i+1}\|^2 &= \left(-\vartheta_i g_{i+1} + \frac{\|g_{i+1}\|^2}{|g_i^T s_i|} s_i\right)^T \left(-\vartheta_i g_{i+1} + \frac{\|g_{i+1}\|^2}{|g_i^T s_i|} s_i\right) \\ &= \vartheta_i^2 \|g_{i+1}\|^2 - 2\vartheta_i \frac{\|g_{i+1}\|^2}{|g_i^T s_i|} g_{i+1}^T s_i + \frac{\|g_{i+1}\|^4}{|g_i^T s_i|^2} \|s_i\|^2 \\ &\le (1+\varepsilon_3)^2 \|g_{i+1}\|^2 + 2\varepsilon_2(1+\varepsilon_3)\|g_{i+1}\|^2 + \frac{\|g_{i+1}\|^4}{|g_i^T s_i|^2}\|s_i\|^2, \end{aligned}$$

or

$$\frac{\|s_{i+1}\|^2}{\|g_{i+1}\|^4} \le \frac{(1+\varepsilon_3)(1+2\varepsilon_2+\varepsilon_3)}{\|g_{i+1}\|^2} + \frac{\|s_i\|^2}{\|g_i\|^4}.$$

Now suppose that

$$\liminf_{i\to\infty} \|g_i\| = 0$$

does not hold. Then there exists a constant $\underline{\varepsilon} > 0$ such that $\|g_i\| \ge \underline{\varepsilon}\ \forall i \in N$, so from the previous inequality it follows that

$$\frac{\|s_{i+1}\|^2}{\|g_{i+1}\|^4} \le \frac{(1+\varepsilon_3)(1+2\varepsilon_2+\varepsilon_3)}{\underline{\varepsilon}^2} + \frac{\|s_i\|^2}{\|g_i\|^4} \le \frac{(1+\varepsilon_3)(1+2\varepsilon_2+\varepsilon_3)}{\underline{\varepsilon}^2}(i+1)$$

(we assume without loss of generality that $\underline{\varepsilon}^2\|s_1\|^2/\|g_1\|^4 \le (1+\varepsilon_3)(1+2\varepsilon_2+\varepsilon_3)$). Thus

$$\sum_{i=1}^{\infty} \frac{\|g_i\|^4}{\|s_i\|^2} \ge \frac{\underline{\varepsilon}^2}{(1+\varepsilon_3)(1+2\varepsilon_2+\varepsilon_3)} \sum_{i=1}^{\infty} \frac{1}{i} = \infty,$$

which is in contradiction with inequality (7). $\square$

It follows from Theorem 2 that modification (16) allows to weaken conditions for global convergence of methods FR [1] and CD [30]. It suffices to choose a step-length by generalized Wolfe conditions (2) and (3), where $\varepsilon_3 \geq 0$ is arbitrarily large but finite number. This condition does not differ much from the weak Wolfe conditions, where $\varepsilon_3 = \infty$.

Relation (16) can also be used to improve conjugation of direction vectors in methods PR and LS.

Theorem 3 *Consider modifications of methods HS, PR, LS given by the rule*

$$s_1 = -g_1 \quad \textit{and} \quad s_{i+1} = -\vartheta_i g_{i+1} + \beta_i s_i \quad \textit{for} \quad i \in N,$$

where the values β_i^{HS}, β_i^{PR}, β_i^{LS} are determined by (10) and

$$\vartheta_i^{HS} = \frac{y_i^T s_i}{y_i^T s_i} = 1, \quad \vartheta_i^{PR} = \frac{y_i^T s_i}{g_i^T g_i}, \quad \vartheta_i^{LS} = \frac{y_i^T s_i}{|g_i^T s_i|}. \tag{18}$$

Then the quadratic termination property stays unchanged and moreover,

$$y_i^T s_{i+1} = 0 \quad \textit{for} \quad i \in N. \tag{19}$$

Proof As in the proof of Theorem 2 we have $\vartheta_i^{HS} = \vartheta_i^{PR} = \vartheta_i^{LS} = 1$, if a step-length is exact. So (16) changes into (8) and the quadratic termination property stays unchanged. Method HS, for which (19) holds, is unchanged. In case of methods PR and LS we obtain

$$y_i^T s_{i+1} = -\frac{y_i^T s_i}{g_i^T g_i} y_i^T g_{i+1} + \frac{y_i^T g_{i+1}}{g_i^T g_i} y_i^T s_i = 0$$

and

$$y_i^T s_{i+1} = -\frac{y_i^T s_i}{|g_i^T s_i|} y_i^T g_{i+1} + \frac{y_i^T g_{i+1}}{|g_i^T s_i|} y_i^T s_i = 0.$$

$\square$

Formula (16) does not assure a descent of direction vectors of methods HS, PR, LS. This requirement is guaranteed by relation (14) or by formula

$$s_1 = -g_1 \quad \text{and} \quad s_{i+1} = -g_{i+1} + \beta_i s_i - \gamma_i y_i \quad \text{for} \quad i \in N, \tag{20}$$

where the values β_i^{HS}, β_i^{PR}, β_i^{LS} are determined by (10) and

$$\gamma_i^{HS} = \frac{g_{i+1}^T s_i}{y_i^T s_i}, \quad \gamma_i^{PR} = \frac{g_{i+1}^T s_i}{g_i^T g_i}, \quad \gamma_i^{LS} = \frac{g_{i+1}^T s_i}{|g_i^T s_i|}, \tag{21}$$

see [132]. Multiplying (20) by a vector g_{i+1} we can easily check a validity of (15). From the practical point of view, formula (20) is less efficient than (14).

Basic conjugate gradient methods can also be combined so that we choose

$$\beta_i = \frac{\lambda_i^1 g_{i+1}^T y_i + \lambda_i^2 g_{i+1}^T g_{i+1}}{\mu_i^1 y_i^T s_i + \mu_i^2 g_i^T g_i - \mu_i^3 g_i^T s_i} = \frac{g_{i+1}^T (g_{i+1} - \lambda_i^1 g_i)}{\mu_i^1 y_i^T s_i + \mu_i^2 g_i^T g_i - \mu_i^3 g_i^T s_i}, \tag{22}$$

where λ_i^1, λ_i^2, μ_i^1, μ_i^2, μ_i^3 are nonnegative numbers such that $\lambda_i^1 + \lambda_i^2 = 1$ and $\mu_i^1 + \mu_i^2 + \mu_i^3 = 1$. One possibility is the choice $\lambda_i^1 = \min(1, \|g_{i+1}\|/\|g_i\|)$, see [117], [122], which leads to modifications

$$\beta_i^{HSM} = \frac{g_{i+1}^T \tilde{y}_i}{y_i^T s_i}, \quad \beta_i^{PRM} = \frac{g_{i+1}^T \tilde{y}_i}{g_i^T g_i}, \quad \beta_i^{LSM} = \frac{g_{i+1}^T \tilde{y}_i}{|g_i^T s_i|}, \tag{23}$$

where

$$\tilde{y}_i = g_{i+1} - \min\left(1, \frac{\|g_{i+1}\|}{\|g_i\|}\right) g_i. \tag{24}$$

Effectiveness of conjugate gradient methods can be improved by suitable restarts. This is performed so that we test fulfilling a prescribed condition after computation of a direction vector. If this condition is not satisfied, then the computed direction vector is replaced with a negative gradient (which corresponds to a choice $\beta_i = 0$). It is very convenient to test a uniform descent condition $-g_{i+1}^T s_{i+1} \geq \varepsilon_0 \|g_{i+1}\| \|s_{i+1}\|$, where $\varepsilon_0 > 0$ is a small number (e.g. $\varepsilon_0 = 10^{-8}$). Such a modified conjugate gradient method is globally convergent without occurring restarts too often. If methods (11) are used, then it is suitable to test a conjugation of direction vectors. In this case, we interrupt the iteration process if the condition

$$y_i^T s_{i+1} \leq \eta_1 \|s_{i+1}\| \|y_i\| \tag{25}$$

does not hold, where the value η_1 depends on the Wolfe conditions chosen. It is also possible to test orthogonality of gradients. The iteration process is restarted if

$$g_i^T g_{i+1} \leq \eta_2 \|g_{i+1}\| \|g_i\| \tag{26}$$

does not hold, where the value η_2 again depends on the Wolfe conditions chosen. If the number of variables is sufficiently large, then it is worth interrupting the iteration process after every n steps counted from the last restart.

3 Numerical experiments

We present a numerical comparison of conjugate gradient methods for minimization of 60 test functions taken from [76] with 1000 variables (NIT is a total number of iterations, NFV is a total number of function evaluations, F is a total number of failures, and T is a total computational time in seconds). The first table contains the results for methods using strong Wolfe conditions (2)–(3) with $\varepsilon_1 = 10^{-4}$, $\varepsilon_2 = 10^{-1}$ and $\varepsilon_3 = 10^{-1}$; the value $\eta_1 = 0.05$ is used in condition (25). The second table contains the results for methods using weak Wolfe conditions (2)–(3) with $\varepsilon_1 = 10^{-4}$, $\varepsilon_2 = 0.9$ and $\varepsilon_3 = \infty$; the value $\eta_1 = 0.2$ is used in condition (25). The third table contains the results for methods using a special line search described in [57].

From the data stated in Tables 1-3 we can deduce several conclusions:

Realization	Method HS NIT	NFV	F	T	Method PR NIT	NFV	F	T	Method LS NIT	NFV	F	T
(10)	139774	212632	-	49.26	174261	283226	2	58.61	168932	265329	2	53.94
(13)	119210	180182	-	40.63	161040	249186	2	55.91	147376	226064	2	46.45
(14)	124991	192584	-	45.04	135756	205897	-	45.66	135960	206329	-	46.86
(16)	139774	212632	-	49.17	141557	225649	-	50.20	142055	221960	-	47.20
(20)	138044	225193	-	50.52	146081	234783	-	52.39	145463	234396	-	52.22
(23)	135025	201969	-	44.91	170704	276052	2	55.48	172540	272644	2	58.91
Realization	Method DY NIT	NFV	F	T	Method FR NIT	NFV	F	T	Method CD NIT	NFV	F	T
(11)	205447	267994	4	52.13	227498	321939	5	64.92	257187	346818	5	73.49
(14)	269027	367475	6	68.78	211859	277978	4	53.12	215700	282033	5	56.02
(16)	205447	267994	4	52.17	211783	278629	4	55.93	210948	272413	5	54.56
(11) + (25)	136535	218387	1	46.86	150147	234202	2	50.67	140509	222035	1	44.91
(14) + (25)	142833	228778	1	47.40	147208	221234	2	45.91	141765	221279	2	51.25
(16) + (25)	136535	218387	1	46.75	139604	223049	-	48.05	141318	217956	1	48.50

Table 1: The strong Wolfe conditions.

Realization	Method HS NIT	NFV	F	T	Method PR NIT	NFV	F	T	Method LS NIT	NFV	F	T
(10)	278645	350503	2	65.84	239625	315005	1	49.40	254874	338751	1	56.36
(13)	309881	386999	3	77.77	275419	354945	2	60.77	318272	404547	4	73.03
(14)	298873	371150	3	74.45	249197	309460	1	59.55	267303	332440	1	60.00
(16)	278645	350503	2	66.50	99198	229502	-	47.72	300630	374527	3	71.94
(20)	419046	619145	6	116.39	303858	406525	3	75.53	303835	405806	3	71.49
(23)	313271	362428	4	77.94	264157	348745	1	63.61	285764	375206	3	66.36
Realization	Method DY NIT	NFV	F	T	Method FR NIT	NFV	F	T	Method CD NIT	NFV	F	T
(11)	267710	272646	4	56.38	371666	451181	5	89.36	450286	509099	10	106.05
(14)	513154	588460	9	123.89	275391	280049	4	58.13	286283	291727	5	63.76
(16)	267710	272646	4	56.31	270054	274863	4	57.53	276589	281664	4	60.75
(11) + (25)	192988	206408	1	51.00	249538	302095	2	61.09	255854	301273	1	59.34
(14) + (25)	368022	439971	6	75.36	231142	245748	2	55.53	195636	209153	1	50.39
(16) + (25)	192988	206408	1	50.84	186556	199588	1	47.20	196423	210158	1	51.24

Table 2: The weak Wolfe conditions.

Realization	Method HS NIT	NFV	F	T	Method PR NIT	NFV	F	T	Method LS NIT	NFV	F	T
(10)	100585	300068	-	72.78	104260	308173	-	74.28	102661	307106	-	69.94
(13)	89728	268513	-	62.41	89847	268416	-	58.94	96601	289303	-	69.89
(14)	93631	282202	-	60.84	98031	293426	-	71.72	105638	315271	1	83.44
(16)	100614	300229	-	72.70	103395	308503	-	61.99	92335	276284	-	62.92
(20)	101023	300800	1	73.56	94165	280980	-	69.16	101486	301235	-	79.20
(23)	93373	277795	-	68.81	96783	288230	-	63.99	101383	303307	-	71.19

Realization	Method DY NIT	NFV	F	T	Method FR NIT	NFV	F	T	Method CD NIT	NFV	F	T
(11)	163046	489494	5	96.78	165868	496121	5	98.31	177997	530560	4	107.88
(14)	169538	502837	2	122.53	162684	486854	4	94.77	161931	484030	4	92.84
(16)	163046	489494	5	96.88	162315	485926	5	92.76	165392	494846	4	99.43
(11) + (25)	108705	328674	1	81.20	96632	292572	1	58.92	135110	404208	1	83.34
(14) + (25)	119501	356686	2	88.36	105748	320766	1	74.63	108234	328114	1	83.95
(16) + (25)	109326	328889	1	81.29	109797	332741	1	81.97	108563	329042	1	78.84

Table 3: The special Hager–Zhang line search.

- It is advantageous to use the strong Wolfe conditions with $\varepsilon_2 = 10^{-1}$ (this value was obtained experimentally) at a realization of conjugate gradient methods, particularly methods HS, PR, LS, and their modifications.
- In case that we use the strong Wolfe conditions, method HS gives the best results. Combination (13) or modifications (14) and (16) (particularly (14)) improve effectiveness of methods HS, PR, LS. Modification (20) improves effectiveness of methods PR and LS. Modification (23) slightly improves effectiveness of method HS.
- In case that we use the strong Wolfe conditions, methods DY, FR, CD give worse results than methods HS, PR, LS. The properties of methods DY, FR, CD are considerably improved if they are restarted each time condition (25) is not fulfilled. The choice of a value η_1 in (25) depends on the Wolfe conditions used (a suitable value must be determine experimentally).
- In general, modifications (14) and (16) considerably improve effectiveness of methods FR and CD. This observation is independent of a choice of the Wolfe conditions which confirms a significance of Theorem 2. If we supplement the stated modifications with conjugation test (25), then the resulting methods are competitive with the best modifications of methods HS and PR. Modification (14) is unsuitable for method DY.
- In case that we use the weak Wolfe conditions, method PR gives better results than methods HS (particularly if we use modification (16)). Modifications (20) and (23) are unsuitable.
- A special choice of a step-length described in [57] allows to find a more accurate solution. Properties of individual methods and their modifications are in this case in accord with the previous conclusions.

References

[1] M.Al-Baali: Descent property and global convergence of the Fletcher-Reeves method with inexact line search. IMA J. Numerical Analysis 5 (1985) 121-124.

[2] N.Andrei: Scaled conjugate gradient algorithms for unconstrained optimization. Computational Optimization and Applications 38 (2007) 401-416.

[3] N.Andrei: Another hybrid conjugate gradient algorithm for unconstrained optimization. Numerical Algorithms 47 (2008) 143-156.

[4] N.Andrei: A Dai-Yuan conjugate gradient algorithm with sufficient descent and conjugacy conditions for unconstrained optimization. Applied Mathematics Letters 21 (2008) 165-171.

[5] P.Armand: Modification of the Wolfe line search rule to satisfy the descent conditions in the Polak-Ribiere-Polyak conjugate gradient method. J. Optimization Theory and Applications 132 (2007) 287-305.

[6] P.Baptist, J. Stoer: On the relation between quadratic termination and convergence properties of minimization algorithms. Part II: Applications. Numerische Mathematik 28 (1977) 367-392.

[7] E.M.L.Beale: A derivative of conjugate gradients. In: Numerical Methods for Nonlinear Optimization (F.A.Lootsma, ed.), Academic Press, London, 1972 39-43.

[8] M.Bertocchi, E.Spedicato: Computational experience with conjugate gradient algorithms. Estratto da Calcolo 16 (1979) 255-269.

[9] M.C.Biggs: Minimization algorithms making use of non-quadratic properties of the objective function. J. Inst. Math. Appl. 8 (1971) 315-327.

[10] E.G.Birgin, J.M.Martinez: A spectral conjugate gradient method for unconstrained optimization. Applied Mathematics and Optimization 43 (2001) 117-128.

[11] A.G.Buckley: Extending the relationship between the conjugate gradient and BFGS algorithms. Mathematical Programming 15 (1978) 343-348.

[12] A.G.Buckley: A combined conjugate-gradient quasi-Newton minimization algorithm, Mathematical Programming 15 (1978) 200-210.

[13] A.G.Buckley: Conjugate gradient methods. In: Nonlinear Optimization 1981 (M.J.D.Powell, ed.), Academic Press, London, 1982 17-22.

[14] X.Chen, J.Sun: Global convergence of a two-parameter family of conjugate gradient methods without line-search. J. of Computational and Applied Mathematics 146 (2002) 37-45.

[15] W.Cheng, Y.Xiao, Q.Hu: A family of derivative-free conjugate gradient methods for large scale nonlinear systems of equations. J. of Computational and Applied Mathematics 224 (2009) 11-19.

[16] A.Cohen: Rate of convergence of several conjugate gradient algorithms. SIAM J. Numerical Analysis 9 (1972) 248-259.

[17] H.P.Crowder and P.Wolfe: Linear convergence of the conjugate gradient method. IBM J. Res. Dev. 16 (1969) 431-433.

[18] Y.Dai: Analysis of conjugate gradient methods. Ph.D. thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1997.

[19] Y.Dai: Convergence of nonlinear conjugate gradient methods. Journal of Computational Mathematics 19 (5), 539-548.

[20] Y.Dai: Convergence of Polak-Ribiere-Polyak conjugate gradient method with constant stepsizes. Research report AMSS-2001-040, Academy of Mathematics and Systems Sciences, Chinese Academy of Sciences 2001.

[21] Y.Dai: Further insight into the convergence of the Fletcher-Reeves method. Sci. China Ser. A 42 (1999) 905-916.

[22] Y.Dai: New properties of a nonlinear conjugate gradient method. Numerische Mathematik 89 (2001) 83-98.

[23] Y.Dai: A nonmonotone conjugate gradient algorithm for unconstrained optimization. J. Syst. Sci. Complex. 15 (2002) 139-145.

[24] Y.Dai: J.Han, G.Liu, D.Sun, H.Yin, Y.Yuan: Convergence properties of nonlinear conjugate gradient methods. SIAM J. Optimization 10 (1999) 345-358.

[25] Y.Dai, L.Liao, D.Li: New conjugacy conditions and related nonlinear conjugate gradient methods. Applied Mathematics and Optimization 43 (2001) 87-101.

[26] Y.Dai, L.Liao, D.Li: On restart procedures for the conjugate gradient method. Numerical Algorithms 35 (2004) 249-260.

[27] Y.Dai, J.M.Martinez, Y.Yuan : An increasing-angle property of the conjugate gradient method and the implementation of large-scale minimization algorithms with line searches. Numerical Linear Algebra with Applications 10 (2003) 323-334.

[28] Y.Dai, Q.Ni: Testing different conjugate gradient methods for large-scale unconstrained optimization. J. of Computational Mathematics 21 (2003) 311-320.

[29] Y.Dai, Y.Yuan, Convergence properties of the Fletcher-Reeves method. IMA J. of Numerical Analysis 16 (1996) 155-164.

[30] Y.Dai, Y.Yuan: Convergence properties of the conjugate descent method. Adv. Math. (China) 26 (1996) 552-562.

[31] Y.Dai, Y.Yuan: Further studies on the Polak-Ribiere-Polyak method. Research report ICM-95-040, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1995.

[32] Y.Dai, Y.Yuan: A nonlinear conjugate gradient method with a strong global convergence property. SIAM J. Optimization 10 (1999) 177-182.

[33] Y.Dai, Y.Yuan: Global convergence of the method of shortest residuals. Numerische Mathematik 83 (1999) 581-598.

[34] Y.Dai, Y.Yuan: Convergence of the Fletcher-Reeves method under a generalized Wolfe search. J. of Computational Mathematics 2 (1996) 142-148.

[35] Y.Dai, Y.Yuan: Convergence properties of the Beale-Powell restart algorithm. Sci. China Ser. A 41 (1998) 1142-1150.

[36] Y.Dai, Y.Yuan: Some properties of a new conjugate gradient method. In: Advances in Nonlinear Programming (Y. Yuan ed.), Kluwer Publications, Boston, (1998) 251-262.

[37] Y.Dai, Y.Yuan: A note on the nonlinear conjugate gradient method. J. of Computational Mathematics 20 (2002) 575-582.

[38] Y.Dai, 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, 1998.

[39] Y.Dai, Y.Yuan:, Extension of a class of nonlinear conjugate gradient methods. Research report ICM-98-049, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 1998.

[40] Y.Dai, Y.Yuan:, A three-parameter family of hybrid conjugate gradient method. Mathematics of Computation 70 (2001) 1155-1167.

[41] Y.Dai, Y.Yuan:, An efficient hybrid conjugate gradient method for unconstrained optimization. Ann. Oper. Res. 103 (2001) 33-47.

[42] Y.Dai, Y.Yuan:, A class of globally convergent conjugate gradient methods. Sci. China Ser. A, 46 (2003) 251-261.

[43] J.W.Daniel: The conjugate gradient method for linear and nonlinear operator equations. SIAM J. Numerical Analysis, 4 (1967) 10-26.

[44] J.W.Daniel: A correction concerning the convergence rate for the conjugate gradient method. SIAM J. Numerical Analysis 7 (1970) 277-280.

[45] N.Deng, Z.Li: Global convergence of three terms conjugate gradient methods. Optimization Methods and Software 4 (1995) 273-282.

[46] L.C.W.Dixon, P.G.Ducksbury, P.Singh: A new three term conjugate gradient method. Technical Report No. 130, Numerical Optimization Centre, The Hatfield Polytechnic, 1985.

[47] S.Du, Y.Chen: Global convergence of a modified spectral FR conjugate gradient method. Applied Mathematics and Computation 202 (2008) 766-770.

[48] R.Fletcher: A FORTRAN subroutine for minimization by the method of conjugate gradients. Atomic Energy Research Establishment, Harwell, Oxfordhire, England, Report No. R-7073, 1972.

[49] R.Fletcher: Practical Methods of Optimization (second edition). John Wiley & Sons, New York, 1987.

[50] R.Fletcher, C.Reeves: Function minimization by conjugate gradients. Computer J. 7 (1964) 149-154.

[51] J.A.Ford, Y.Narushima, H.Yabe: Multi-step nonlinear conjugate gradient methods for unconstrained minimization. conjugate gradients. Computational Optimization and Applications 40 (2007) 191-216.

[52] J.C.Gilbert, J.Nocedal: Global convergence properties of conjugate gradient methods for optimization. SIAM J. Optimization 2 (1992) 21-42.

[53] P.E.Gill, W.Murray: Conjugate-gradient methods for large-scale nonlinear optimization. Systems Optimization Laboratory, Department of Operations Research, Stanford Univarsity, Report No. SOL-79-15, 1979.

[54] G.H.Golub, D.P.OLeary: Some history of the conjugate gradient methods and Lanczos algorithms: 1948 -1976. SIAM Review 31 (1989) 50-100.

bibitemgrl1 L.Grippo, S.Lucidi: A globally convergent version of the Polak-Ribiere conjugate gradient method. Mathematical Programming 78 (1997) 375-391.

[55] L.Grippo, S.Lucidi: Convergence conditions, line search algorithms and trust region implementations for the Polak-Ribiere conjugate gradient method. Optimization Methods and Software 20 (2005) 71-98.

[56] W.W.Hager, H.Zhang: A survey of nonlinear conjugate gradient methods. Pacific Journal of Optimization 2 (2006) 35-58.

[57] W.W.Hager, H.Zhang: A New conjugate gradient method with guaranteed descent and an efficient line search. SIAM J. Optimization 16 (2005) 170-192.

[58] W.W.Hager, H.Zhang: Algorithm 851: CG_DESCENT, a conjugate gradient method with guaranted descent. ACM Transactions on Mathematical Software 32 (2006) 113-137.

[59] J.Han, G.Liu, D.Sun, H.Yin: Two fundamental convergence theorems for nonlinear conjugate gradient methods and their applications. Acta Math. Appl. Sinica, 17 (2001) 38-46.

[60] J.Han, G.Liu, H.Yin: Convergence properties of conjugate gradient methods with strong Wolfe linesearch. Systems Sci. Math. Sci. 11 (1998) 112-116.

[61] M.R.Hestenes, Conjugate direction methods in optimization. Springer-Verlag, New York, 1980.

[62] M.R.Hestenes, E.L.Stiefel: Methods of conjugate gradients for solving linear systems. J. Research Nat. Bur. Standards 49 (1952) 409-436.

[63] H.Hirst: N-step quadratic convergence in the conjugate gradient method. PhD Dissertation, Department of Mathematics, Pennsylvania State University, State College, PA, 1989.

[64] Y.Hu, C.Storey: Efficient generalized conjugate gradient algorithms. Part 2: Implementation. J. Optimization Theory and Applications 69 (1991) 139-152.

[65] Y.Hu, C.Storey: Global convergence result for conjugate gradient methods. J. Optimization Theory and Applications 71 (1991) 399-405.

[66] Y.Hu, C.Storey: Motivating quasi-Newton updates by preconditioned conjugate gradient methods. Report No. A150, Dept. of Math. Sci., Loughborough Univ. of Technology, Loughborough 1991.

[67] H.Huang, Z.Wei, Y.Shengwei: The proof of the sufficient descent condition of the Wei-Yao-Liu conjugate gradient method under the strong Wolfe-Powell line search. Applied Mathematics and Computation 189 (2007) 1241-1245.

[68] K.M.Khoda, Y.Liu, C.Storey: Generalized Polak-Ribiere algorithm. J. Optimization Theory and Applications 75 (1992) 345-354.

[69] R.Klessig, E.Polak: Efficient implementation of the Polak-Ribiere conjugate gradient algorithm. SIAM J. Control 10 (1972) 524-549.

[70] G.Li, C.Tang, Z.Wei: New conjugacy condition and related new conjugate gradient methods for unconstrained optimization. J. of Computational and Applied Mathematics 202 (2007) 523-539.

[71] G.Liu, J.Han, H.Yin: Global convergence of the Fletcher-Reeves algorithm with an inexact line search. Appl. Math. J. Chinese Univ. Ser. B 10 (1995) 75-82.

[72] G.Liu, L.Jing, L.Han, D.Han: A class of nonmonotone conjugate gradient methods for unconstrained optimization. J. Optimization Theory and Applications 101 (1999) 127-140.

[73] D.Liu, J.Nocedal: On the limited memory BFGS method for large scale optimization. Mathematical Programming 45 (1989) 503-528.

[74] Y.Liu, C.Storey: Efficient generalized conjugate gradient algorithms. Part 1: Theory. J. Optimization Theory and Applications 69 (1991) 129-137.

[75] L.Lukšan: Computational experience with improved conjugate gradient methods for unconstrained minimization. Kybernetika 28 (1992) 249-262.

[76] L.Lukšan, J.Vlček: Sparse and partially separable test problems for unconstrained and equality constrained optimization. Report V-767. Prague, ICS AS CR, 1998.

[77] A.Miele, J.W.Cantrell: Study on a memory gradient method for the minimization of functions. J. Optimization Theory and Applications 3 (1969) 459-185.

[78] G.P.McCormick, K.Ritter: Alternative Proofs of the convergence properties of the conjugate-gradient method. J. Optimization Theory and Applications 13 (1975) 497-518.

[79] M.F.McGuire, P.Wolfe: Evaluating a restart procedure for conjugate gradients. Report RC-4382, IBM Research Center, Yorktown Heights, 1973.

[80] H.Mukai: Readily implementable conjugate gradient methods. Mathematical Programming 17 (1979) 298-319.

[81] Y.Narushima, H.Yabe: Global convergence of a memory gradient method for unconstrained optimization. Computational Optimization and Applications 35 (2006) 325-346.

[82] J.L.Nazareth: A conjugate direction algorithm without line searches. J. Optimization Theory and Applications 23 (1977) 373-387.

[83] J.L.Nazareth: A relationship between the BFGS and conjugate gradient algorithms and its implications for the new algorithms. SIAM J. Numerical Analysis 16 (1979) 794-800.

[84] J.L.Nazareth: Conjugate gradient methods less dependent on conjugacy. SIAM Review 28 (1986) 501-511.

[85] J.L.Nazareth: A view of conjugate gradient-related algorithms for nonlinear optimization. In: Proceedings of the AMS-IMS-SIAM Summer Research Conference on Linear and Nonlinear Conjugate Gradient-Related Methods, University of Washington, Seattle, WA (July 9-13, 1995).

[86] J.L.Nazareth: Conjugate-gradient methods. IN: Encyclopedia of Optimization (C.Floudas, P.Pardalos, eds.) Kluwer Academic Publishers, Boston, 1999.

[87] J.L.Nazareth, J.Nocedal: Properties of conjugate gradient methods with inexact line searches. Systems Optimization Laboratory, Department of Operations Research, Stanford Univarsity, Report No. SOL-78-1, 1978.

[88] A.Neumaier: On convergence and restart conditions for a nonlinear conjugate gradient method. Preprint 1997.

[89] J.Nocedal: Updating quasi-Newton matrices with limited storage. Mathematics of Computation 35 (1980) 773-782.

[90] J.Nocedal: Theory of Algorithm for Unconstrained Optimization. Acta Numerica, Cambridge University Press (1991) 199-242.

[91] J.Nocedal: Conjugate Gradient Methods and Nonlinear Optimization. In: Proceedings of the AMS-IMS-SIAM Summer Research Conference on Linear and Nonlinear Conjugate gradient-Related Methods, University of Washington, Seattle, WA (July 9-13, 1995).

[92] J.Nocedal: Large scale unconstrained optimization. In: State of the Art in Numerical Analysis (A. Watson, I. Du., eds.) Oxford University Press, (1997) 311-338.

[93] J.M.Perry: A class of conjugate gradient algorithms with a two-step variable-metric memory. Discussion Paper 269, Center for Mathematical Studies in Economics and Management Sciences, Northwestern University, Evanston, Illinois, 1977.

[94] E.Polak, G.Ribière: Note sur la convergence de directions conjugeés. Rev. Francaise Informat. Recherche Opertionelle, 3e Annee 16 (1969) 35-43.

[95] B.T.Polyak: The conjugate gradient method in extreme problems. USSR Comp. Math. Math. Phys. 9 (1969) 94-112.

[96] M.J.D.Powell: Some convergence properties of the conjugate gradient method. Mathematical Programming 11 (1976) 42-49.

[97] M.J.D.Powell: Restart procedures of the conjugate gradient method. Mathematical Programming 2 (1977) 241-254.

[98] M.J.D.Powell: Nonconvex minimization calculations and the conjugate gradient method. In: Numerical Analysis (Dundee, 1983), Lecture Notes in Mathematics, Vol. 1066, Springer-Verlag, Berlin, (1984) 122-141.

[99] M.J.D.Powell: Convergence properties of algorithms for nonlinear optimization. SIAM Review 28 (1986) 487-500.

[100] R.Pytlak: On the convergence of conjugate gradient algorithm. IMA J. of Numerical Analysis 14 (1989) 443-460.

[101] R.Pytlak: Global convergence of the method of shortest residuals by Y.Dai and Y.Yuan. Numerische Mathematik 91 (2002) 319-321.

[102] R.Pytlak, T.Tarnawski: Preconditioned conjugate gradient algorithms for nonconvex problems. Pacific Journal of Optimization 2 (2006) 81-104.

[103] S.Sanmayias, E.Vercher: A generalized conjugate gradient algorithm. J. Optimization Theory and Applications 98 (1998) 489-502.

[104] D.F.Shanno: On the convergence of a new conjugate gradient algorithm. SIAM J. Numerical Analysis 15 (1978) 1247-1257.

[105] D.F.Shanno: Conjugate gradient methods with inexact searches. Math. Oper. Res. 3 (1978) 244-256.

[106] D.F.Shanno: Globally convergent conjugate gradient algorithms. Mathematical Programming 33 (1985) 61-67.

[107] J.R.Shewchuk: An introduction to the conjugate gradient method without the agonizing pain. See http://www.cs.cmu.edu/~jrs/jrspapers.html, 1994.

[108] Z.Shi, J.Guo: A new family of conjugate gradient methods. J. of Computational and Applied Mathematics 224 (2009) 444-457.

[109] Z.Shi, J.Shen: Convergence of Liu-Storey conjugate gradient method. European J. of Operational Research 182 (2007) 552-560.

[110] J.Stoer: On the relation between quadratic termination and convergence properties of minimization algorithms. Numerische Mathematik 28 (1977) 343-366.

[111] J.Sun, J.Zhang: Global convergence of conjugate gradient methods without line search. Ann. Oper. Res., 163 (2001) 161-173.

[112] C.Tang, Z Wei, G.Li: A new version of the Liu-Storey conjugate gradient method. Applied Mathematics and Computation 189 (2007) 302-313.

[113] D.Touati-Ahmed and C. Storey: Efficient hybrid conjugate gradient techniques. J. Optimization Theory and Applications 64 (1990) 379-397.

[114] C.Wang, S.Lian: Global convergence properties of two new dependent Fletcher-Reeves conjugate gradient methods. Applied Mathematics and Computation 181 (2006) 920-931.

[115] C.Wang, Y.Zhang: Global convergence properties of s-related conjugate gradient methods. Chinese Science Bulletin, 43 (1998) 1959-1965.

[116] Z.Wei, G.Li, L.Qi: New nonlinear conjugate gradient formulas for large-scale unconstrained optimization problems Applied Mathematics and Computation 179 (2006) 407-430.

[117] Z.Wei, S.Yao, L.Liu: The convergence properties of some new conjugate gradient methods. Applied Mathematics and Computation 183 (2006) 1341-1350.

[118] P.Wolfe: Convergence conditions for ascent methods. SIAM Review 11 (1969) 226-235.

[119] P.Wolfe: Convergence conditions for ascent methods II: Some corrections. SIAM Review, 13 (1971) 185-188.

[120] P.Wolfe: A method of conjugate subgradients for minimizing nondifferentiable functions. Mathematical Programming Study 3 (1975) 145-173.

[121] H.Yabe, M.Takano: Global convergence properties of nonlinear conjugate gradient methods with modified secant conditions. Computational Optimization and Applications 28 (2004) 203-225.

[122] S.Yao, Z.Wei, H.Huang: A note about WYL's conjugate gradient method and its applications. Applied Mathematics and Computation 191 (2007) 381-388.

[123] G.Yu, Y.Zhao, Z.Wei: A descent nonlinear conjugate gradient method for large-scale unconstrained optimization. Applied Mathematics and Computation 187 (2007) 636-643.

[124] Y.Yuan: Analysis on the conjugate gradient method. Optimization Methods and Software 2 (1993) 19-29.

[125] Y.Yuan: On the truncated conjugate gradient method. Mathematical Programming A-87 (2000) 561-573.

[126] Y.Yuan, J.Stoer: A subspace study on conjugate algorithms. Z. Angew. Math. Mech. 75 (1995) 69-77.

[127] L.Zhang: Two modified Dai-Yuan nonlinear conjugate gradient methods. Numerical Algorithms (2009). To appear.

[128] J.Zhang, N.Deng, L.Chen: New quasi-Newton equation and related methods for unconstrained optimization. J. Optimization Theory and Applications 102 (1999) 147-167.

[129] J.Zhang, C.Xu: Properties and numerical performance of quasi-Newton methods with modified quasi-Newton equations. J. of Computational and Applied Mathematics 137 (2001) 269-278.

[130] L.Zhang, W.Zhou: Two descent hybrid conjugate gradient methods for optimization. J. of Computational and Applied Mathematics 216 (2008) 251-264.

[131] L.Zhang, W.Zhou, D.Li: Global convergence of a modified Fletcher-Reeves conjugate gradient method with Armijo-type line search. Numerische Mathematik 104 (2006) 561-572.

[132] L.Zhang, W.Zhou, D.Li: A descent modified Polak-Ribiere-Polyak conjugate gradient method and its global convergence. IMA J. of Numerical Analysis 26 (2006) 629-640.

[133] G.Zoutendijk: Nonlinear Programming. Computational Methods. In: Integer and Nonlinear Programming (J. Abadie, ed.), North-Holland, Amsterdam, 1970 37-86.