Iraqi Journal of Statistical Science (20) 2011
The Fourth Scientific Conference of the College of Computer Science & Mathematics
pp [246-258]

A Modified Conjugate Gradient Method using multi-step in Unconstrained Optimization

Dr. Huda Issam Ahmed[**] **Dr.Eman Tarik Hamed**

Abstract:

In this paper we proposed a new HS type conjugate gradient method by using two kind of modification, first derived a new non-quadratic model second using structure of the memoryless BFGS quasi-Newton method. The new proposed method always generates a descent condition. We give a sufficient condition for the global converges of the proposed general method. Finally, some numerical results are also reported.

تحسين طريقة التدرج المترافق باستخدام الخطوات المتعددة في الأمثليـة غير المقيدة

د. هدى عصام احمد **د. أيمان طارق حامد**

الملخص

في هذا البحث تم اقتراح طريقة جديدة لصيغة H/s في التدرج المترافق باستخدام نوعين من التحسينات ،أولا اشتقاق نموذج غير تربيعي جديد وثانيا تطبيق خاصية تقليل الخزن لصيغة BFGS.الطريقة الجديدة تم برهنتها أنها تحقق الشرط الضروري للانحدار وكذلك تحقق خاصية التقارب الشامل.

1. Introduction

We are concerned with the following unconstrained minimization problem:

$$\text{minimize } f(x) \qquad (1)$$

where $f: R^n \to R$ is smooth and its gradient $g(x) = \nabla f(x)$ is available. There are several kinds of numerical methods for solving (1), which include the steepest

[**]Assistant Prof. \ **College of Computers Sciences and Math.\ University of Mosul**.
[*]Assistant Prof. \ **College of Computers Sciences and Math.\ University of Mosul**.

Received:1/10 /2011 ____________________Accepted: 21 /12 / 2011

descent methods, the Newton method and quasi-Newton methods, for example. Among them the conjugate gradient method is one choice for solving large-scale problems, because it does not need any matrices. Conjugate gradient methods are iterative methods of the form

$$x_{k+1} = x_k + \lambda_k d_k \tag{2}$$

$$d_{k+1} = \begin{cases} -g_{k+1} & \text{for } k = 0 \\ -g_{k+1} + \beta_k d_k & \text{for } k \geq 1 \end{cases} \tag{3}$$

where $g_k = g(x_k)$ denotes $\nabla f(x_k)$,β_k is a positive scalar and α_k is a positive scalar which is determined by a line search step satisfying the sufficient descent condition:

$$g_{k+1}^T d_{k+1} \leq -c\|g_{k+1}\|^2 \tag{4}$$

where $\|.\|$ stands for Euclidean norm.

The HS, FR, PR and LS are four well-known conjugate gradient methods, they are specified by:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k} \text{ (Hestenes and Stiefel, 1952),} \tag{5}$$

$$\beta_k^{FR} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2} \text{ (Fletcher and Reeves, 1964),} \tag{6}$$

$$\beta_k^{PR} = \frac{g_{k+1}^T y_k}{\|g_k\|^2} \text{ (Polak, 1969)} \tag{7}$$

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k} \text{ (Liu and Story, 1991),} \tag{8}$$

where $y_k = g_{k+1} - g_k$.

Note that these formulas for β_k are equivalent for each other if the objective function is strictly convex quadratic function and λ_k is exact line search. There are many researches on convergence properties of these methods see for example (Hager and Zhang, 2006) and (Nocedal and Wright, 2006).

To establish the convergence results methods mentioned above, it is usually required that the step α_k should satisfy the following strong Wolfe conditions

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{9}$$

$$\left| g(x_k + \alpha_k d_k)^T d_k \right| \leq -\sigma\; g_k^T d_k \tag{10}$$

where $0 < \delta < \sigma < 1$. On the other hand, many numerical methods (e.g. the steepest descent method and quasi-Newton methods) for unconstrained optimization are proved to be convergent under the Wolfe conditions:

$$f(x_k + \alpha_k d_k) - f(x_k) \leq \delta \alpha_k g_k^T d_k \tag{11}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma \; g_k^T d_k \tag{12}$$

see (Wolfe,1969) and (Zoutendijk,1970).

Also the conjugate gradient methods always satisfy the sufficient descent condition

$$g_{k+1}^T d_{k+1} \leq -c_k \|g_{k+1}\|^2 \tag{13}$$

where c is a positive constant.

In this paper, we propose a new HS type conjugate gradient method by using structure memoryless BFGS quasi-Newton method (Nocedal,1980) and (Shanno, 1978). The present paper is organized as follows. In section 2 we derived a new non-quadratic model and use it in new multi-step of algorithm. In section 3 we define the memoryless BFGS quasi-Newton method. In section 4 we propose a specific new conjugate gradient method based on the new non-quadratic model and new multi-step quasi-Newton method and prove its global converges. Finally in section 5, some numerical experiments are presented.

2- The Non-Quadratic Models.

Most of the currently used optimization methods use a local quadratic representation of the objective function, but the use of the quadratic model may be inadequate to incorporate all the information (Fried, 1999) so that more general models than quadratic are proposed as a basic for CG algorithms, also (Al-Bayati, 1993) and (Tassopoulos and Story, 1984) have proposed further modifications of the conjugate gradient method whichs are based on some non-quadratic models. If $q(x)$ is a quadratic function defined by:

$$q(x) = \frac{1}{2} x^T G x + b^T x + c \tag{14}$$

where G is $n \times n$ symmetric and positive definite matrix and b is a constant vector in R^n and c is a constant. Then we say that f is defined as a nonlinear scaling of $q(x)$ if the following conditions hold (Boland et al., 1979):

$$f(x) = F(q(x)), \qquad q > 0 \quad \text{and} \frac{dF}{dq} = F' > 0$$

The following proportions are immediately derived from the above conditions:

1- Every contour line of $q(x)$ is a contour line of f.

2- If x^* is minimize of $q(x)$ then it's also a minimize of f.

In this area there are various published works.

(a) A CG methods which minimize the function:
$f(x) = (q(x))^p, \quad p > 0, \quad x \in R^n$, in at most n-step have been described by (Fried, 1991).

(b) The special polynomial case:
$F(q(x)) = \in_1 q(x) + \frac{1}{2} \in_2 q^2(x)$, where $\in_1, \in_2$ scalars, has been investigated by (Boland et al., 1979).

(c) A rational model has been developed by (Tassopoulos and Story, 1984) where: $F(q(x)) = \frac{\in_1 q(x) + 1}{\in_2 q(x)}, \quad \in_1 > 0, \ \in_2 > 0.$

(d) Another rational model was considered by (Al-Bayati, 1993) where:
$F(q(x)) = \frac{\in_1 q(x)}{1 - \in_2 q(x)}, \quad \in_1 > 0, \ \in_2 > 0.$

2.1 An Extended CG method

We consider a new non-quadratic model defined by:

$$F(q(x)) = e^{q(x)^2 + 2q(x) + 1} \tag{15}$$

Assume that $q > 0$ and $\frac{df}{dq} > 0$, the unknown quantities ρ_k were expressed in term of available quantities of the algorithm (i.e. function and gradient value of the objective function) using the expression for ρ_k

$$\rho_k = \frac{F'_k}{F'_{k+1}} \tag{16}$$

From the relations

$$g_{k+1} = F'_{k+1}\, G\left(x_{k+1} - x^*\right) \tag{17}$$

$$g_k = F'_k\, G\left(x_k - x^*\right) \tag{18}$$

Where G is the Hessian matrix, we have

$$\rho_k = \frac{F'_k}{F'_{k+1}} = \frac{g_k^T\left(x_{k+1} - x^*\right)}{g_{k+1}^T\left(x_k - x^*\right)}$$

furthermore

$$g_k^T\left(x_{k+1} - x^*\right) = g_k^T\left(x_k + \lambda_k d_k - x^*\right) = g_k^T\left(x_k - x^*\right) + \lambda_k g_k^T d_k$$

and

$$g_{k+1}^T\left(x_k - x^*\right) = g_{k+1}^T\left(x_{k+1} - \lambda_k d_k - x^*\right) = g_{k+1}^T\left(x_{k+1} - x^*\right)$$

since $g_{k+1}^T d_k = 0$. Therefore, we express ρ_k as follows:

$$\rho_k = \frac{g_k^T\left(x_k - x^*\right) + \lambda_k g_k^T d_k}{g_{k+1}^T\left(x_{k+1} - x^*\right)} \tag{19}$$

From (16), (17) and (19), we get:

$$\rho_k = \frac{F'_k\left(x_k - x^*\right) G\left(x_k - x^*\right) + \lambda_k g_k^T d_k}{F'_{k+1}\left(x_{k+1} - x^*\right) G\left(x_{k+1} - x^*\right)}$$

Therefore

$$\rho_i = \frac{2F'_k q_k + \lambda_k g_k^T d_k}{2F'_{k+1} q_{k+1}} \tag{20}$$

If we express F'_k and F'_{k+1} as follows, using the derivation the general exponential function,

$$F'_k = 2(q_k(x)+1)e^{(q_k(x)+1)^2} \tag{21}$$

$$F'_{k+1} = 2(q_{k+1}(x)+1)e^{(q_{k+1}(x)+1)^2} \tag{22}$$

Solving (15) for q(x) we have:

$$q(x) = (\ln f(x))^{\frac{1}{2}} - 1 \tag{23}$$

so that $F'_k = 2f_k(\ln(f_k))^{\frac{1}{2}}$ (24)

$$F'_{k+1} = 2f_{k+1}(\ln(f_{k+1}))^{\frac{1}{2}} \tag{25}$$

By substituting for the $f'_{k+1}q_{k+1}$ and $f'_k\, q_k$ in (20), we have

$$\rho_k = \frac{2f_k(\ln(f_k))^{\frac{1}{2}} + \hat{w}}{2f_{k+1}(\ln(f_{k+1}))^{\frac{1}{2}}} \tag{26}$$

Where $\hat{w} = \dfrac{\lambda_k g_k^T d_k}{2}$ (27)

3- The Memoryless BFGS Quasi-Newton Method

The direction d_{k+1} in the quasi-Newton BFGS method is given by:

$$d_{k+1} = -H_{k+1} g_{k+1} \tag{28}$$

where H_{k+1} is nxn symmetric and positive definite matrix and defined by:

$$H_{k+1} = H_k - [\frac{H_k y_k s_k^T + s_k y_k^T H_k}{s_k^T y_k}] + (1 + \frac{y_k^T H_k y_k}{s_k^T y_k}) \frac{s_k s_k^T}{s_k^T y_k} \tag{29}$$

If we use the memoryless BFGS (i.e. $H_k = I$, where I is the identity matrix) then the formula of H_{k+1} is defined by:

$$H_{k+1} = I_k - [\frac{y_k s_k^T + s_k y_k^T}{s_k^T y_k}] + (1 + \frac{y_k^T y_k}{s_k^T y_k}) \frac{s_k s_k^T}{s_k^T y_k} \tag{30}$$

Where $s_k = \alpha_k d_k = x_{k+1} - x_k$. In this case, d_{k+1} can be written as:

$$d_{k+1} = -g_{k+1} - [(1 + \frac{y_k^T y_k}{v_k^T y_k}) \frac{s_k^T g_{k+1}}{s_k^T y_k} - \frac{y_k^T g_{k+1}}{s_k^T y_k}] s_k + \frac{s_k^T g_{k+1}}{s_k^T y_k} y_k \tag{31}$$

4- New Multi-step Quasi Newton Method

In this section we drive a new multi-step quasi Newton method based on memoryless BFGS quasi-Newton method as a conjugate gradient method. This type of algorithms has been investigated for the first time by Ford and Maghrabi (Ford and Maghrabi, 1994). Now let us define a new multi-step by:

$$r_k = \rho_k (s_k - \mu_k s_{k-1}) \tag{32}$$

$$w_k = (y_k - \rho_k \mu_k y_{k-1}) \tag{33}$$

where $\mu_k = \dfrac{s_{k-1}^T s_k}{s_{k-1}^T s_{k-1}}$ (34)

and ρ_k is define in (26).

then the matrix in eq(29) is defined by:

$$H_{k+1}^{new} = H_k - [\frac{H_k w_k r_k^T + r_k w_k^T H_k}{r_k^T w_k}] + (1 + \frac{w_k^T H_k w_k}{r_k^T w_k}) \frac{r_k r_k^T}{r_k^T w_k} \tag{35}$$

If we use the memoryless BFGS then the new direction d_{k+1} is defined by: $d_{k+1}^{new} = -g_{k+1} - [(1 + \dfrac{w_k^T w_k}{r_k^T w_k}) \dfrac{r_k^T g_{k+1}}{r_k^T w_k} - \dfrac{w_k^T g_{k+1}}{r_k^T w_k}] r_k + \dfrac{r_k^T g_{k+1}}{r_k^T w_k} w_k$ (36)

since $r_k^T g_{k+1} = 0$، then equation (36) becomes:

$$d_{k+1}^{new} = -g_{k+1} + \frac{w_k^T g_{k+1}}{r_k^T w_k} r_k \tag{37}$$

Or equivalent to $d_{k+1}^{new} = -g_{k+1} + \dfrac{g_{k+1}^T w_k}{r_k^T w_k} r_k$ (38)

This search direction can be rewriten as the form:

$$d_{k+1}^{new} = -g_{k+1} + \beta_k^{MHS} r_k \tag{39}$$

where $\beta_k^{MHS} = \dfrac{g_{k+1}^T w_k}{r_k^T w_k}$, and ρ_k, w_k, μ_k, r_k are defined in (26, 32-34) respectively, the property of this multi-step is satisfying the QN-condition:

$$H_{k+1}y_k = s_k \tag{40}$$

we can rewrite (40) by equivalent new form:

$$H_{k+1}\ w_k = r_k \tag{41}$$

since w_k, r_k are defined in (32-33), so the relation (41) must satisfy a modified of the form:

$$H_{k+1}(y_k - \rho_k\mu_k\ y_{k-1}) = (s_k - \rho_k\mu_k\ s_{k-1}) \tag{42}$$

where ρ_k, μ_k are positive scalar and defined in (26,43) respectively .

now from (42) we obtain:

$$H_{k+1}(y_k - \rho_k\mu_k\ y_{k-1}) = \rho_k s_k - \rho_k\mu_k s_{k-1}$$
$$H_{k+1}y_k - \rho_k\mu_k H_{k+1}\ y_{k-1} = \rho_k s_k - \rho_k\mu_k s_{k-1}$$
$$H_{k+1}y_k = \rho_k s_k - \rho_k\mu_k\ s_{k-1} + \rho_k\mu_k H_{k+1}\ y_{k-1}$$
$$= \rho_k s_k - \rho_k\mu_k\ [s_{k-1} - H_{k+1}\ y_{k-1}]$$

since $(H_{k+1}\ y_{k-1} = s_{k-1})$

$\therefore\ H_{k+1}y_k = \rho_k s_k$, which is equivalent to (standard quasi-Newton condition).

4.1- Outline of New Algorithm

Step1: Set x_0 , $\in$ (initial point, scalar termination).

Step2: Set $k = 0$, $d_k = -g_k$.

Step3: Set $x_{k+1} = x_k + \lambda_k d_k$, $k \ge 0$ where λ_k is obtained from the line search procedure.

Step4: check for convergence, i.e. if $\|g_{k+1}\| < \in$, stop; otherwise continue.

Step5: Compute $\rho_k = \dfrac{2f_k(\ln(f_k))^{\frac{1}{2}} + \hat{w}}{2f_{k+1}(\ln(f_{k+1}))^{\frac{1}{2}}}$, where $\hat{w} = \dfrac{\lambda_k g_k^T d_k}{2}$.

Step6: Compute w_k, r_k , μ_k which are defined in (32-34).

Step7: Compute the new search direction defined by:

$d_{k+1} = -g_{k+1} + \beta_K^{MHS} d_k$, $k \ge 1$, where β_K^{MHS} is computed by the following formula $\beta_k^{MHS} = \dfrac{g_{k+1}^T w_k}{r_k^T w_k}$.

Step8: if k=n or if $\|g_{k+1}^T g_k\| > 0.2\|g_{k+1}\|$ is satisfied go to step (2), else set k=k+1and go to step then stop. Otherwise go to step 3.

4.2 global convergence

In order to establish the descent condition and the global convergence of the new proposed method, we make the following additional assumption.

4.2.1 Assumption

1) The level set $\Omega = \{x / f(x) \le f(x_0)\}$ at x_0 is bounded.

2) In some neighborhood N of Ω, f is continuously differentiable and its gradient is Lipschitz continuous with Lipschitz constant $L > 0$, i.e.

$$\|g(x) - g(y)\| \le L\|x - y\| \text{ , for all } x, y \in N . \tag{43}$$

The above Assumption implies that there exists a positive constant γ such that :

$$\|g(x)\| \le \gamma \text{ for all } x \in \Omega . \tag{44}$$

4.2.2Theorem

The direction d_{k+1} given in (38) satisfies the descent condition

$$g_{k+1}^T d_{k+1} = -\|g_{k+1}\|^2 \tag{45}$$

Proof:

Since $d_0 = -g_0$, we have $g_0^T d_0 = -\|g_0\|^2$, which satisfies (45).

Now from (8), we have:

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \beta_k^{MHS} r_k g_{k+1} \text{, where } \beta_k^{MHS} = \frac{g_{k+1}^T w_k}{r_k^T w_k}$$

$$d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 + \frac{g_{k+1}^T w_k}{r_k^T w_k} r_k^T g_{k+1}$$

Since $r_k^T g_{k+1} = 0$

$$\therefore \; d_{k+1}^T g_{k+1} = -\|g_{k+1}\|^2 \; .$$

We note that this method always satisfies $g_k^T d_k = -\|g_k\|^2 < 0$ for all k, which implies the sufficient descent condition (13) with c=1.

Property (*): Consider the method (2) and (3) **,** assume that there exists a positive constant $\bar{\gamma}$ such that $\|g_{k+1}\| \ge \bar{\gamma}$ holds for all k. then we say that the method has Property (*) if there exists constants $b > 1$ and $\xi > 0$ such that for all k:

$$|\beta_k| \le b \tag{46}$$

$$\text{and } \|s_k\| \le \xi \Rightarrow |\beta_k| \le \frac{1}{b} \tag{47}$$

Also we need the following additional Assumption (Ford et al., 2009) to prove the global convergence of the proposed method.

4.2.3 Assumption

1) Assume that there exists appositive constant τ_1 such that, for all k

$$\left|g_k^T r_k\right| > \tau_1 \left|g_k^T d_k\right| \tag{48}$$

2) Assume that there exists appositive constant τ_2 such that satisfies $0 \le \tau_1 \le 1$ and

$$\rho_k \left|\mu_k\right| \le \tau_2 \min\left\{ \frac{\left|g_{k+1}^T y_k\right|}{\left|g_{k+1}^T y_{k-1}\right|}, \frac{\left|r_k^T y_k\right|}{\left|r_k^T y_{k-1}\right|} \right\} \text{ for all k} \tag{49}$$

4.2.4 Theorem: Consider the method (2) and (3) that satisfies the following conditions:

(1) $\beta_k \ge 0$ for all k.

(2) Property (*) holds.

Assume that λ_k satisfies the strong Wolfe condition (9) and (10). If Assumption 4.2.1 holds, then the method converges in the sense that $\lim_{k\to\infty} \left\|g_{k+1}\right\| = 0$ (Ford et al., 2009).

Now using the above theorem, we obtain the following global convergence property.

4.2.5 New Theorem:

Suppose that Assumption 4.2.1 and 4.2.3 are satisfied. Consider the method (2)-(3) with (38). Assume that λ_k satisfies the strong Wolfe condition (9) and (10) then the new method converges in the sense that $\lim_{k\to\infty} \left\|g_{k+1}\right\| = 0$.

Proof:

It's clearly $\beta_k \ge 0$. So we only prove that the proposed method satisfies condition (2) of Theorem (**4.2.4).** To this end, we assume that there exists a constant $\bar{\gamma}$ such that $\left\|g_{k+1}\right\| \ge \bar{\gamma}$ for all k.

It follows from (33) and (49) that:

$$\begin{aligned} \left|g_{k+1}^T w_k\right| &= \left|g_{k+1}^T y_k\right| + \rho_k \left|\mu_k g_{k+1}^T y_{k-1}\right| \\ &\le \left|g_{k+1}^T y_k\right| + \tau_2 \left|g_{k+1}^T y_k\right| \\ &\le (1+\tau_2) \left|g_{k+1}^T y_k\right| \\ &\le (1+\tau_2) L \left\|g_{k+1}\right\| \left\|s_k\right\|. \end{aligned} \tag{50}$$

and also from (33), (49), and the fact $g_{k+1}^T r_k = 0$ we have

$$\left|r_k^T w_k\right| \geq \left|r_k^T y_k\right| - \rho_k \left|\mu_k r_k^T y_{k-1}\right|$$
$$\geq (1-\tau_2)\left|r_k^T y_k\right|$$
$$= (1-\tau_2)\left|g_k^T r_k\right| \tag{51}$$

It follow from (48) and (45) that:

$$\left|g_k^T r_k\right| \geq \tau_1 \left|g_k^T d_k\right| = \tau_1 \left\|g_k\right\|^2 \tag{52}$$

Therefore, from (51) yield

$$\left|r_k^T w_k\right| \geq \tau_1(1-\tau_2)\left\|g_k\right\|^2 \tag{53}$$

By definition of β_k^{MHS}, (50) and (53) and since $\left\|s_k\right\| < A$,we have

$$\left|\beta_k^{MHS}\right| \leq \frac{\left|g_{k+1}^T w_k\right|}{\left|r_k^T w_k\right|} \leq \frac{(1+\tau_2)L\left\|g_{k+1}\right\|\left\|s_k\right\|}{\tau_1(1-\tau_2)\left\|g_k\right\|} \leq \frac{(1+\tau_2)L\gamma A}{\tau_1(1-\tau_2)\bar{\gamma}^2} = b \tag{54}$$

Now let $\xi = \dfrac{(1+\tau_2)\,\bar{\gamma}^2}{\tau_1(1-\tau_2)L\,\gamma\,b}$

Then, if $\left\|s_k\right\| \leq \xi$, we have

$$\left|\beta_k^{MHS}\right| \leq \frac{(1+\tau_2)L\gamma A}{\tau_1(1-\tau_2)\bar{\gamma}^2} \leq \frac{1}{b} \tag{55}$$

Therefore, Property (*) holds. Thus from Theorem (**4.2.4**), the Theorem is true (i.e. $\lim\limits_{k\to\infty}\left\|g_{k+1}\right\| = 0$).

5 Numerical results

A new non-quadratic model is derived and a new implicit multi-step quasi Newton method have also been derived, using these two derivations in memoryless BFGS method and obtain the new modification of HS (modified Hestenes-Stiefel) method.

We compare MHS with standard HS method. The parameters in the strong Wolfe line search were chosen to be $\delta = 0.001$ and $\sigma = 0.9$. For each test problem, the termination criterion is $\left\|g_{k+1}\right\| < 10^{-5}$, also the value of μ_k where $\mu_k = \dfrac{s_{k-1}^T s_k}{s_{k-1}^T s_{k-1}}$ is between the (0,1). The numerical results of our experiment are reported from Table (1). Each problem was tested with various value of n changing from n=100, 1000, 10000. The numerical results are given in the form of NOI (Number of Iterations), NOE (Number of Evaluations) and using a program written in FORTRAN language.

Table (1)

Comparison the new modified HS method with the standard HS method

Test fn.	Dim.	New modified HS method		HS method		Test fn.	Dim.	New modified HS method		HS method	
		NOF	NOI	NOF	NOI			NOF	NOI	NOF	NOI
Powell	100	93	40	109	41	**Penalty**	100	8	2	8	2
	1000	188	70	109	41		1000	8	2	8	2
	10000	177	73	163	57		10000	10	3	10	3
Cantrel	100	198	30	272	36	**Cubic**	100	49	19	44	16
	1000	214	46	328	40		1000	48	19	44	16
	10000	269	37	370	43		10000	49	19	44	16
Miele	100	186	61	110	34	**Sum**	100	65	12	65	12
	1000	163	60	172	47		1000	87	18	82	18
	10000	289	94	429	157		10000	173	38	185	36
Wolfe	100	89	44	99	49	**Extended**	100	27	5	27	5
	1000	103	51	141	70	**cilff**	1000	27	5	27	5
	10000	280	137	399	164		10000	29	6	29	6
Shallow	100	24	10	25	10	**Denschnb**	100	21	9	15	6
	1000	27	11	25	10	(CUTE)	1000	25	11	18	7
	10000	27	11	24	9		10000	24	10	18	7
Rosen	100	60	24	54	22	**Dixmaana**	100	14	6	12	5
	1000	72	29	88	32	(CUTE)	1000	13	5	12	5
	10000	67	28	54	22		10000	17	7	12	5
Recip	100	18	6	16	5	**Dixmaanb**	100	13	5	13	5
	1000	18	6	16	5	(CUTE)	1000	13	5	13	5
	10000	25	8	18	6		10000	14	5	14	5

References

Al-Bayati, A., (1993),”A new non-quadratic model for unconstrained non-linear optimization”, Mutah. J. For Research and Studies, 8.

Boland, W., Kamgnia, E. and Kowallik, J., (1979),” A conjugate gradient optimization method invariant to non-linear scaling “, Journal of optimization theory and Applications, 2.

Dai, Y. H. and Yuan, Y., (1999),”A nonlinear conjugate gradient method with a strong global convergence property”, SIAM J. Optim., 10, pp.177-182.

Ford, A. J. and Moghrabi, A. I., (1994),” Multi-step quasi-Newton methods for optimization”, Journal of Computational and Applied Mathematics, 50, pp.305-323.

Ford, A. J., Narushima, Y. and Yabe, H. (2009),” A three-term conjugate gradient method with sufficient descent property for unconstrained optimization”, Computational Optimization and Applications, 40, pp.191-216.

Fletcher, R. and Reeves, C., (1964),” Function minimization by conjugate gradients”, Computer Journal, 7,pp.149-154.

Hager, W. W. and Zhang, H., (2005),”A new conjugate gradient method with guaranteed descent and an efficient line search “, SIAM. J. Optim. ,16,pp.170-192.

Hager,W. W. and Zhang, H.,(2006),”A survey of nonlinear conjugate gradient methods”, Pacific Journal of Optimization, 2, pp.35-58.

Hestenes, M. R. and Stiefel, E. L., (1952)” Methods of conjugate gradients for solving linear systems”,Journal of Research of the National Bureau of Standards, Section B,49, pp.409-432.

Liu, Y. L. and Storey, S. C., (1991),”Efficient generalized conjugate gradient algorithms”, Part 1: Theory, J. Optima. Theory Appl.,69, pp.129-137.

Nocedal, J., (1980)," Updating quasi-Newton matrices with limited storage", Math. Comput., 35, pp. 773-782.

Nocedal, J. and Wright, S.J., (2006),"Numerical Optimization (Second Edition)", Springer Series in Operation Research, Springer Verlag, New York.

Polyak, B. T., (1969), "The conjugate gradient method in extreme problems", USSR Comp. Math. Phys., Vol. 9, pp. 94-112.

Shanno, D. F., (1978)," Conjugate gradient methods with inexact searches", Math. Oper. Res. 3, pp. 244-256.

Tassopoulus, A. and Story, C., (1984),"A variable metric method using non-quadratic model", Journal of optimization theory and Applications, 43, pp.200-217.

Wolfe, P. (1969)," Convergence conditions for ascent methods", SIAM. Rev. 11, pp. 226-235.

Zoutendijk, G., (1970)," Nonlinear Programming ", Computational Methods, in: J. Abadie (Ed.), Integer and Nonlinear Programming, North-Holland.