Australian Journal of Basic and Applied Sciences, 4(12): 6069-6085, 2010
ISSN 1991-8178

An Extended CG-Method with a New Non-Monotone Line Search Procedure for Constrained Optimization

[1]Runak M. Abdullah [2]Abbas Y. AL-Bayati

[1]Lecturer of Numerical Optimization , Department of Mathematics,
College of Science, University of Sulaimani, Iraq.
[2]Professor of Numerical Optimization, Department of Mathematics,
College of Computer Sciences and Mathematics, Mosul University, Iraq.

Abstract: In this paper, a new type of non-quadratic models is proposed which based on non-quadratic conic model. The new model is based upon a non-linear scaling of a quadratic form and it employs a new line search technique which is suitable for non-convex optimization problems. A restarting procedure due to Powell, and based upon some earlier work of Beale, is implemented in this work. The new proposed method is derived and tested on several standard constraint test functions with promising numerical results.

Key words: Conjugate Gradient Method, Non-Quadratic Conic Model, Restarting Criterion, Non-Monotone Line Search, Constrained Optimization.

INTRODUCTION

The general problem we deal with is:

$$\min_{x\in R^n} imize\ f(X) \tag{1a}$$

Subject to, h(x) = 0 (1b)

where $f: R^n \to R$, $C: R^n \to R^m$ and $h: R^n \to R^m$ are continuously differentiable functions in n variables

x = $(x_1, .. ,x_n)^T$, 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 Descent (SD); Newton and Quasi-Newton (QN) methods, for example. Among them, the Conjugate Gradient (CG) method is one choice for solving large scale problems, because it doesn't need any matrices. CG methods are iterative methods of the form, Bonnans and Gilbert (2006):

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

where $\alpha_k > 0$ is a step-size and d_k is a search direction. Search direction are usually defined by:

$$d_k = \begin{cases} -g_1, & for\ \ k = 1, \\ -g_k + \beta_k d_{k-1}, & for\ \ k \ge 2, \end{cases} \tag{3}$$

where g_k denotes $\nabla f(x_k)$ and β_k is a scalar. If β_k is a strictly convex quadratic function

Corresponding Author: Runak M. Abdullah, Lecturer of Numerical Optimization , Department of Mathematics, College of Science, University of Sulaimani, Iraq.

$f(x)=\frac{1}{2}x^T Ax-b^T x,$ where $A\in R^{nxn}$ is a symmetric positive definite matrix, and if α_k is the exact one-dimensional minimizer, then the method (2)-(3) is called linear CG method. The linear CG-method was originally proposed by Hestenes and Stiefel, (HS), (1952) for solving linear system of equations *Ax=b* and several formulae of β_k were considered, which are equivalent each other for the strictly convex quadratic objective function when exact line searches (ELS) are used. Within the framework of linear gradient methods, the conjugacy condition is defined by:

$$d_i^T A d_j = 0, \qquad i \neq j \tag{4}$$

For search directions, and this condition guarantees the finite termination of the linear CG-methods. On the other hand, the method (2)-(3) is called nonlinear CG-method for general unconstrained optimization problem. Nonlinear CG-methods were first proposed by Fletcher and Reeves, (FR), (1964). Within the framework of nonlinear CG-methods, the conjugacy condition is replaced by:

$$d_k^T y_{k-1} = 0 \tag{5a}$$

where $y_k = g_{k+1} - g_k$, and using the mean value theorem yields:

$$d_k^T y_{k-1} = \alpha_{k-1} d_k^T \nabla^2 f(x_{k-1} + \xi\alpha_{k-1} d_{k-1}) d_{k-1} \tag{5b}$$

for some $\xi \in (0,1)$. The above condition means that the search directions d_k and d_{k-1} are mutually conjugate with respect to the Hessian matrix $\nabla^2 f(x)$ at some point. Well-known formulae for β_k are the (FR); Polak-Ribière-Polyak (PRP), (1969) and (HS) formulae and their details are given in Hagar and Zhang (2006):

$$\beta_k^{FR} = \frac{\|g_k\|^2}{\|g_{k-1}\|^2} \tag{6a}$$

$$\beta_k^{PRP} = \frac{g_k^T y_{k-1}}{\|g_{k-1}\|^2} \tag{6b}$$

$$\beta_k^{HS} = \frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \tag{6c}$$

where $\|.\|$ denotes the Euclidean norm.

There are some merits and demerits in each method. In the FRCG-method, if a bad direction and a tiny step from x_{k-1} to x_k are generated, the next direction d_k and the next step $S_k = a_k d_k$ are also likely to be poor unless a restart along the gradient direction is performed. In spite of such a defect, Zoutendijk (1970) proved that the FRCG-method with exact line search is globally convergent for general functions. Al-Baali (1985) extended this result to inexact line searches (ILS). On the other hand, the PRP and the HS CG-methods perform similarly in terms of theoretical property. Both methods are preferred to the FRCG-method in its numerical performance, because the methods essentially perform a restart after it encounters a bad direction. Nevertheless, Powell (1984) showed that the PRP and the HS CG-methods can cycle infinitely without approaching a solution, which implies that they do not have globally convergence. The global convergence properties of the FR, PRP and HS CG-methods without regular restarts have been studied by many researchers, including Al-Baali (1985) and Gilbert and Nocedal (1992). Restarting CG-methods were discussed by Beale (1989) and Powell (1977). New kinds of nonlinear CG-methods were developed by Dai and Liao (2001); Dai and Yuan (1999) and Wei et al (2008). However, scaled CG-methods were investigated widely by Andrei (2007) and (2009) and Al-Bayati Mohammed (2010).

More general concept of conjugacy has been suggested for optimization problems by considering conjugate direction methods based on quadratic objective functions by minimizing sequentially along lines in conjugate directions. The concept of conjugacy used in these methods is invariant under affine transforms for the non-quadratic objective functions using the following transformation $\delta(w) = x_0 + Jw$.

Definition (1.1): Coxeter (1964)

A mapping δ: W à X between sets W and X $\subseteq R^n$ is a collineartion if and only if:

$\delta(w) = x_0 + Jw / (1+p^T w)$, for some invertible J $R^{n\times n}$

Definition (1.2): Coxeter (1964)

A function f: X à R is conic if and only if its composition fs: W à R with some collineartion s: W à X is quadratic. It is normal if this quadratic composition has a positive definite Hessian.

Definition (1.3): Coxeter (1964)

Two lines in the domain X of a normal conic function f: Xà R are conjugate if and only if they images of orthogonal ones in W under any collineartion s: W à X which makes fs: W à R quadratic and unit Hessian.

Various authors in the past have introduced non-quadratic methods for function minimization. These methods which are based on non-quadratic models are called conic model methods. Byrd et al. (1987) have analyzed and tested methods using conic models for one dimensional optimization and found them generally superior to these using cubic models. Dennis and Schnabel (1983) has run some tests on an n-dimensional method of this type which has been encouraging though preliminary. Davidon (1980), introduced conic models for unconstrained optimization. Nocedal and Liu (1989), are presented an method for minimizing conic function of a CG method. Among the most popular recent non-quadratic models are Al-Bayati (1993, 2001, 2007); Al-Bayati and Al-Naemi (1995) and Al-Bayati and Jabbar (2010).

2. Conjugate Gradient Methods Based on Standard Conic Models:

Given an initial point $x_0 \in R^n$, we will write any $x \in R^n$ as $x = x_0 + s$. With respect to this reference point, we define the conic function as:

$$f(x) = f(x_0 + s) = f_0 + \frac{g_0^T s}{1 - a_0^T s} + \frac{1}{2}\frac{s^T G s}{\left(1 - a_0^T s\right)^2} \tag{7}$$

where $g_0 \in R^n$, $a_0 \in R^n$ and G is an n by n positive definite and symmetric matrix. We call a_0 the horizon vector and denote the domain of f by $\hat{D}$ i.e. $\hat{D} = \left\{x : 1 - a_0^T s \neq 0\right\}$, Since the term $\frac{s}{1 - a_0^T s}$ appear once in the second term on the right-hand side of (7), and twice in the third term, it is clear that by letting

$$w = \frac{s}{1 - a_0^T s} \tag{8}$$

the conic function becomes a quadratic in the variable w,

$$f(x) = f(x_0 + s) = f_0 + g_0^T w + \frac{1}{2} w^T G w = \hat{h}(w) \tag{9}$$

We can express s in term of w:

$$s = \frac{w}{1 + a_0^T w} \tag{10}$$

To simplify the formulas, we define

$$\gamma(x) = 1 - a_0^T s = \frac{1}{1 + a_0^T w} \tag{11}$$

Where $x = x_0 + s$, so that $w = s/\gamma$. We call $E = \{x : 1 - a_0^T s = 0\}$, the singular hyper plane, and note that if $\gamma(x)\gamma(y) < 0$, then x and y lie on opposite sides of E. We will need to relate the derivative of *f* to that of $\hat{h}$, since

$$f(x) = f(x_0 + s) = f\left(x_0 + \frac{w}{1 + a_0^T w}\right) = \hat{h}(w) \tag{12}$$

It follows from the chain rule that:

$$\hat{h}'(w) = \gamma(x)\left(I - a_0^T s\right) g(x) \tag{13}$$

where g denotes the gradient of *f*. The conic function (7) is useful as a model function for minimization only if it has a unique minimize. This is ensured by the conditions

$$G > 0 \text{ and } a_0^T G^{-1} g_0 \neq 1 \tag{14}$$

If equation (14) holds, *f* will be called a normal conic function, see Davidon (1980).

The horizon vector a can be computed using function and gradient values at any three collinear points.

In particular, consider the iterates x_k and x_{k+1} let x_t be given by $x_t = x_k + \lambda_t d_k$, where $\lambda_t \neq \lambda_k$. Then

$$a_0 = \frac{(\lambda_t - \lambda_k)\gamma_k^2 g_k + \lambda_k \gamma_t^2 g_t - \lambda_t \gamma_{k+1}^2 g_{k+1}}{(\lambda_t - \lambda_k)\lambda_k^2 s_k^T g_k + \lambda_k \gamma_k^2 s_t^T g_k - \lambda_t \gamma_{k+1}^2 s_{k+1}} \tag{15}$$

where as before $s_i = x_i - x_0, i = k, t, k+1$; and where η_t is defined by

$$\eta_{k+1} = \frac{-g_k^T (x_{k+1} - x_k)}{f_k - f_{k+1} + \varsigma_{k,k+1}} \eta_k, \tag{16}$$

With $\eta_0 = 1$, and

$$\varsigma_{k,k+1} = +\left[(f_k - f_{k+1})^2 - g_{k+1}^T (x_{k+1} - x_k) g_k^T (x_{k+1} - x_k)\right]^{1/2} \tag{17}$$

For more details see Nocedal and Liu (1989).

2.1: Outlines of the Standard Conic Model Method:

Step(1): Set x_0, ε, μ_0 (initial point, scalar termination, initial parameter) and let $\beta_0 = 0, \delta_{-1} = 0$ and

$w_0 = 0$. For i =0, 1, …., n

Step(2): $d_0 = -\hat{h}_0$

Step(3): Set $x_{i+1} = x_i + \lambda_i \delta_i, i \geq 0,$ where is obtained from the line search procedure.

Step(4): Compute $a_0 = \dfrac{(\lambda_t - \lambda_k)\gamma_k^2 g_k + \lambda_k \gamma_t^2 g_t - \lambda_t \gamma_{k+1}^2 g_{k+1}}{(\lambda_t - \lambda_k)\lambda_k^2 s_k^T g_k + \lambda_k \gamma_k^2 s_t^T g_k - \lambda_t \gamma_{k+1}^2 s_{k+1}}$

Step(5): The new search direction compute $d_i = \hat{h}_i' + \beta_i \delta_{i-1}$ with

$$\beta_i = \frac{\hat{h}_i'^T y_{i-1}}{y_i^T \delta_{i-1}}, \tag{18}$$

where β_k is the conjugacy coefficient of conic model and $y_i = \hat{h}_{i+1} - \hat{h}_i$.

Step(6): Check for convergence if $\frac{1}{\mu}\sum_{i=1}^{l} h_i(x) < \varepsilon$, then stop. Otherwise set $\mu_{i+1} = \mu_i \times 10$, and take $x = x^*$ as a new starting point.

Step(7): Check for restarting criterion if $w_{i+1}^T g_{i+1} \geq -0.8 |g_{i+1}^T g_{i+1}|$ is satisfied, Powell (1982), then go to **step(2)**. Otherwise, go to **step(3)**.

2.2 Standard Non-monotone Line Search Approach:

Let σ_1 , σ_2 be two forcing functions. Let *m* be a positive integer. Given parameters $0 < \sigma_1 < \beta < 1$ and $M \geq 0$, $\sigma_2 > 0$. At the iteration *k*, the step length α_k , satisfies that:

$$f(x_k + \alpha_k d_k) \leq C_k - \sigma_1 \min\{\sigma_1(\mu_k), \sigma_2(\zeta_k)\} \tag{19}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \beta_k g_k^T d_k \tag{20}$$

$$\mu_k = -\frac{g_k^T d_k}{\|d_k\|} \zeta_k = -\alpha_k g_k^T d_k \tag{21}$$

$C_0 = f_0$, $Q_0 = 1$ and $C_{k+1} = \dfrac{\gamma_k Q_k C_k + f_{k+1}}{Q_{k+1}}$ (22)

with $\gamma_k \in [0,1]$ and

$$Q_{k+1} = \gamma_k Q_{k+1} \tag{23}$$

Clearly from (22) and (23), C_k is a convex combination of the function values $f_0, f_1, \ldots, f_k$. The choice of γ_k controls the degree of the non-monotonicity. If $\gamma_k = 0$ and $\min\{\sigma_1, \sigma_2\} = \zeta_k$ for each *k* then (19) reduces to monotone line search. If $\gamma_k = 1$ for each *k* then C_k is the average of $f_0, f_1, \ldots, f_k$ and (19) reduces to the monotone non-increasing function. For more details see Yin and Du (2007).

3. A New Non-Monotone Line Search Approach:

Given parameters $0<\delta<\beta<1$, and a nonnegative integer *M*, let σ_1 , σ_2 be two forcing functions. At the iteration *k*, the step length α_k , satisfies that:

$$f(x_k+\alpha_k d_k) \le \max_{0\le j\le M} f(x_{k-j}) - \delta \min\{\sigma_1(\mu_k), \sigma_2(v_k)\} - \alpha_k^2 \|d_k\|^2 \tag{24}$$

$$g(x_k+\alpha_k d_k)^T d_k \ge \beta g_k^T d_k \tag{25}$$

where d_k is the search direction $\mu_k = -\dfrac{g_k^T d_k}{\|d_k\|}$, and $v_k = -\alpha_k g_k^T d_k$ (26)

This new line search procedure will make descent property more rapid, and hence increase its rate of convergence.

Lemma (3.1):

Consider any iteration of the form $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction and α_k is a step length Satisfying (24). Then there exist a subsequence of $\{\alpha_k\}$, denoted as $\{x_{p(i)}\}$, for which

$$\sum \phi_{p(i)-1} < +\infty \tag{27}$$

where

$$\phi_{p(i)-1} = \min\{\sigma_1(\mu_{p(i)-1}), \sigma_2(v_{p(i)-1})\} - \alpha_{p(i)-1}^2 \|d_{p(i)-1}\|^2 \tag{28}$$

New Theorem (3.2):

Consider $x_{k+1} = x_k + \alpha_k d_k$, where d_k is a descent direction and α_k is a step length satisfying (24)-(26). If the sequence $\{f_k\}$ is infinite and bounded from below and satisfying lemma (3.1), then

$$\sum_{k=0}^{\infty} \frac{\min\{\sigma_1(\mu_k), \sigma_2(\zeta_k)\} + \alpha_k^2 \|d_k\|^2}{Q_{k+1}} < \infty \tag{29}$$

and hence $\liminf_{k\to\infty} \|g_k\| = 0$ (30)

Proof:

From (24) it is clear that $\alpha_k^2 \|d_k\|^2 \ge 0$ and since d_k is a descent direction, $\mu_k \ge 0$ and $\zeta_k \ge 0$

$\Rightarrow \quad \min\{\sigma_1, \sigma_2\} \ge 0$

$\Rightarrow \quad f_{k+1} \le \max f(x_{k-j})$ from (24) (31)

And from (22)

$$C_{k+1} = \frac{\gamma_k Q_k C_k + f_{k+1}}{Q_{k+1}} \quad \text{, where} \quad C_k = \max_{0\le j\le M} f(x_{k-j}) \tag{32}$$

$$\leq= \frac{\gamma_k Q_k C_k + \{C_k - \min\{(\sigma_1, \sigma_2\} - \alpha_k^2 \|d_k\|^2\}}{Q_{k+1}} \tag{33}$$

$$= C_k - \frac{\min\{(\sigma_1, \sigma_2\} + \alpha_k^2 \|d_k\|^2}{Q_{k+1}} \tag{34}$$

$\therefore$ $\{C_k\}$ is monotone non-increasing since $\{f_k\}$ is bounded.

From (21), $\{f_k\}$ is non-increasing. This implies

$$\sum_{k=0}^{\infty} \frac{\min\{\sigma_1(\mu_k), \sigma_2(\zeta_k)\} + \alpha_k^2 \|d_k\|^2}{Q_{k+1}} < \infty \tag{35}$$

4. A New Extended Conic Model for the CG-Method:

The general problem we deal with is defined in (1a) and (1b). The constraint problem defined by:

$$h(x) = 0 \tag{36}$$

can be transformed into an unconstrained problem, which is:

$$\text{Minimize } \phi(x) = f(x) + \frac{1}{\mu} \sum_{i=1}^{l} \left(h_i(x)\right)^2 \tag{37}$$

where $\mu > 0$.
Standard methods for unconstrained optimization are based on a quadratic model of the form

$$q(x) = \frac{1}{2}\left(x - x^*\right)^T G\left(x - x^*\right) + q\left(x^*\right) \tag{38}$$

where x^* is the minimizer of $q(x)$ of and G is the Hessian matrix.

A conic function has the form:

$$c(x) = \frac{1}{2} \frac{x^T G x}{(1 + a_0^T x)^2} + \frac{b^T x}{(1 + a_0 x)} + a \tag{39}$$

where $a_0 \in R^n$ is the vector defining the horizon of the conic function. If c(x) is a conic function, then a function f is defined as nonlinear scaling of c(x) if the following condition holds:

$$f = F(c(x)), \quad df/dc = f' > 0 \text{ and } c(x) > 0 \tag{40}$$

where x^* is the minimizer of c(x) with respect to x, see Spedicato (1976).

The following properties are immediately derived from the above condition, Boland & Kowalik, (1979): every contour line of c(x) is a contour line of *f*.

if x^* is a minimizer of c(x), then it is a minimizer of *f* .

that x^* is a global minimum of c(x) dose not necessarily mean that it is a global minimum of *f* .

Many authors have suggested special models to determine ρ_i , where ρ_i is defined as follows:

$$\rho_i = \frac{f'_{i-1}}{f'_i} \tag{41}$$

where f is defined as follows, Fried (1971).

$$d f/dc = f' > 0 \text{ and } c > 0. \tag{42}$$

and f is an increasing monotonic function, which may better represent the objective function and c is a conic function. These authors employ non-quadratic models to extend the CG-method. But in this paper we will employ the conic model to extend the CG-method.

Now, a new specific model is investigated which extend the standard CG-method. This new model is as follows:

$$F(c(x)) = \frac{\varepsilon c(x)}{1-\varepsilon c(x)}, \quad 0< \varepsilon < 1 \tag{43}$$

where c(x) is a conic function and ε is a positive scalar.

4.1 Derivation of the New Parameter ρ_i^{NEW} in the New Proposed CG-Method:

It is first assumed that ε is not zero in equation (43) for c(x).

$$\therefore \quad f = \frac{\varepsilon c}{1-\varepsilon c} \tag{44}$$

Then $$c = \frac{f}{\varepsilon(f+1)} \tag{45}$$

Now, the derivation of equation (44) is:

$$f' = \frac{(1-\varepsilon c)(\varepsilon) - \varepsilon c(-\varepsilon)}{(1-\varepsilon c)^2} \tag{46}$$

Substituting (45) in (46), yields:

$$f' = \frac{\varepsilon}{\left(1-\varepsilon\dfrac{f}{\varepsilon(f+1)}\right)^2}$$

or

$$f' = \varepsilon(f+1)^2 \tag{50}$$

and using the expression for ρ_i^{NEW}

$$\rho_i^{NEW} = \frac{f_i'}{f_{i+1}'} = \left(\frac{f_i+1}{f_{i+1}+1}\right)^2 \tag{51}$$

where $f_i' = \varepsilon(f_i+1)^2$ and $f_{i+1}' = \varepsilon(f_{i+1}+1)^2$ (52)

By using the Chain Rule we have:

$$\frac{dF(c(x))}{dx} = \frac{dF(c(x))}{d(c(x))}.\frac{d(c(x))}{dx}$$

i.e.

$$g^T(x) = f'(x)\nabla c(x) \tag{53}$$

The conic function is

$$c(x) = \frac{1}{2}\frac{x^T G x}{(1+a_0^T x)^2} + \frac{b^T x}{(1+a_0 x)} + a$$

Then $\nabla c(x) = \left(\frac{Gx}{1+a_0^T x}\right)\frac{\left(1+a_0^T x\right)(1) - x\left(a_0^T\right)}{(1+a_0^T x)^2} + b^T\frac{\left(1+a_0^T x\right)(1) - x\left(a_0^T\right)}{(1+a_0^T x)^2}$

i.e.

$$\nabla c(x) = \frac{Gx}{(1+a_0^T x)^3} + \frac{b^T}{(1+a_0^T x)^2}$$

or

$$\nabla c(x) = \frac{1}{(1+a_0^T x)^2}\left[\frac{Gx}{(1+a_0^T x)} + b^T\right] \tag{54}$$

Now, we can show that:

$$\nabla c(x) = \frac{Gx}{(1+a_0^T x)^3} + \frac{b^T}{(1+a_0^T x)^2}$$

or

$$\nabla c(x^*) = \frac{Gx^*}{(1+a_0^T x^*)^3} + \frac{b^T}{(1+a_0^T x^*)^2}$$

$\because$ x^* is the minimum point, this implies that $\nabla c\left(x^*\right) = 0$ (55)

$$\Rightarrow \quad \frac{Gx^*}{(1+a_0^T x^*)^3} + \frac{b^T}{(1+a_0^T x^*)^2} = 0$$

$$\Rightarrow \quad \frac{Gx^*}{(1+a_0^T x^*)^3} = -\frac{b^T}{(1+a_0^T x^*)^2}$$

$$\Rightarrow \quad \frac{Gx^*}{(1+a_0^T x^*)} = -b^T$$

$$\Rightarrow \quad b^T = -\frac{Gx^*}{(1+a_0^T x^*)} \tag{56}$$

$$\nabla c(x) - \nabla c(x^*) = \frac{Gx}{(1+a_0^T x)^3} + \frac{b^T}{(1+a_0^T x)^2} - \frac{Gx^*}{(1+a_0^T x^*)^3} - \frac{b^T}{(1+a_0^T x^*)^2} \tag{57}$$

Substituting (55) and (56) in (57), we have:

$$\nabla c(x) = \frac{Gx}{(1+a_0^T x)^3} - \frac{Gx^*}{(1+a_0^T x^*)(1+a_0^T x)^2}$$

$$\nabla c(x) = \frac{1}{(1+a_0^T x)^2}\left[\frac{Gx}{(1+a_0^T x)} - \frac{Gx^*}{(1+a_0^T x^*)}\right]$$

$$\nabla c(x) = \frac{1}{(1+a_0^T x)^2}\left[\frac{Gx(1+a_0^T x^*)}{(1+a_0^T x)} - \frac{Gx^*(1+a_0^T x)}{(1+a_0^T x^*)}\right]$$

$$\nabla c(x) = \frac{1}{(1+a_0^T x)^2}\left[\frac{Gx + a_0^T Gxx^* - Gx^* - a_0^T Gx^* x}{(1+a_0^T x)(1+a_0^T x^*)}\right]$$

$$\nabla c(x) = \frac{1}{(1+a_0^T x)^2}\left[\frac{Gx - Gx^*}{(1+a_0^T x)(1+a_0^T x^*)}\right]$$

$$\nabla c(x) = \frac{1}{(1+a_0^T x)^2}\left[\frac{G\left(x - x^*\right)}{(1+a_0^T x)(1+a_0^T x^*)}\right]$$

$$\nabla c(x) = \frac{G\left(x - x^*\right)}{(1+a_0^T x)^3(1+a_0^T x^*)} \tag{58}$$

Substituting (58) in (53), we have:

$$g' = f'G\frac{(x-x^*)}{(1+a_0^T x)^3(1+a_0^T x^*)} \tag{59}$$

Then

$$g_i^T = f_i'G\frac{(x_i-x^*)}{(1+a_0^T x_i)^3(1+a_0^T x^*)} \tag{60}$$

and

$$g_{i+1}^T = f_{i+1}'G\frac{(x_{i+1}-x^*)}{(1+a_0^T x_{i+1})^3(1+a_0^T x^*)} \tag{61}$$

by definition of ρ_i^{NEW} we have:

$$\rho_i^{NEW} = \frac{f_i'}{f_{i+1}'} = \frac{g_i^T(x_{i+1}-x^*)(1+a_0^T x_i)^3}{g_{i+1}^T(x_i-x^*)(1+a_0^T x_{i+1})^3}$$

$$= \frac{g_i^T(x_{i+1}-x^*)(1+a_0^T x_i)^3}{g_{i+1}^T(x_i-x^*)(1+a_0^T x_{i+1})^3} \tag{62}$$

Furthermore

$$\begin{aligned} g_i^T\left(x_{i+1}-x^*\right) &= g_i^T\left(x_i+\lambda_i d_i - x^*\right) \\ &= g_i^T\left(x_i-x^*\right)+\lambda_i g_i^T d_i \end{aligned} \tag{63}$$

and

$$g_{i+1}^T\left(x_i-x^*\right)$$

$$= g_{i+1}^T\left(x_{i+1}-d_i-x^*\right)$$

$$= g_{i+1}^T\left(x_{i+1}-x^*\right)-g_{i+1}^T d_i \tag{64}$$

$$= g_{i+1}^T\left(x_{i+1}-x^*\right)$$

since $g_{i+1}^T d_i = 0$, for exact line search

therefore we can express ρ_i^{NEW} as follows:

$$\rho_i^{NEW} = \frac{\left(g_i^T(x_i - x^*) + \lambda_i g_i^T d_i\right)(1 + a_0^T x_i)^3}{g_{i+1}^T(x_{i+1} - x^*)(1 + a_0^T x_{i+1})^3} \tag{65}$$

from (60) and (61), we get

$$\rho_i^{NEW} = \frac{f_i' \dfrac{(x_i - x^*)^T G(x_i - x^*)}{(1 + a_0^T x_i)^3 (1 + a_0^T x^*)}(1 + a_0^T x_i)^3 + \lambda_i g_i^T d_i (1 + a_0^T x_i)^3}{f_{i+1}' \dfrac{(x_{i+1} - x^*)^T G(x_{i+1} - x^*)}{(1 + a_0^T x_i)^3 (1 + a_0^T x^*)}(1 + a_0^T x_{i+1})^3}$$

therefore

$$\rho_i^{NEW} = \frac{2 f_i' c_i (1 + a_0^T x_i)^2 (1 + a_0^T x^*) + \lambda_i g_i^T d_i (1 + a_0^T x_i)^3}{2 f_{i+1}' c_{i+1} (1 + a_0^T x_{i+1})^2 (1 + a_0^T x^*)}$$

or

$$\rho_i^{NEW} = \rho_i \left(\frac{c_i}{c_{i+1}}\right)\left(\frac{1 + a_0^T x_i}{1 + a_0^T x_{i+1}}\right)^2 + \frac{n(1 + a_0^T x_i)^3}{(1 + a_0^T x_{i+1})^2 (1 + a_0^T x^*) f_{i+1}' c_{i+1}} \tag{66}$$

the quantities $\left(\dfrac{c_i}{c_{i+1}}\right)$ and $f_{i+1}' c_{i+1}$ can be rewritten as:

$$\frac{c_i}{c_{i+1}} = \left(\frac{f_i}{\varepsilon\left(f_i + 1\right)}\right)\left(\frac{\varepsilon\left(f_{i+1} + 1\right)}{f_{i+1}}\right)$$

$$= \left(\frac{1}{\sqrt{\rho_i}}\right)\left(\frac{f_i}{f_{i+1}}\right) \tag{67}$$

$$f_{i+1}' c_{i+1} = f_{i+1}(f_{i+1} + 1) \tag{68}$$

Substituting (67) and (68) in (66), we have:

$$\rho_i^{NEW} = \sqrt{\rho_i^{NEW}} \left(\frac{f_i}{f_{i+1}}\right)\left(\frac{1 + a_0^T x_i}{1 + a_0^T x_{i+1}}\right)^2 + \frac{n\left(1 + a_0^T x_i\right)^3}{\left(1 + a_0^T x_{i+1}\right)^2 \left(1 + a_0^T x^*\right) f_{i+1}\left(f_{i+1} + 1\right)} \tag{69}$$

$$= \left(\frac{f_i}{f_{i+1}}\right)\left(\frac{f_i + 1}{f_{i+1} + 1}\right)\left(\frac{1 + a_0^T x_i}{1 + a_0^T x_{i+1}}\right)^2 + \frac{n\left(1 + a_0^T x_i\right)^3}{\left(1 + a_0^T x_{i+1}\right)^2 \left(1 + a_0^T x^*\right) f_{i+1}\left(f_{i+1} + 1\right)}$$

$$=\left(\frac{1+a_0^T x_i}{1+a_0^T x_{i+1}}\right)^2\left[\left(\frac{f_i}{f_{i+1}}\right)\left(\frac{f_i+1}{f_{i+1}+1}\right)+\left(\frac{n}{f_{i+1}\left(f_{i+1}+1\right)}\right)\left(\frac{1+a_0^T x_i}{1+a_0^T x^*}\right)\right] \tag{70}$$

$$\Rightarrow \rho_i^{NEW}=\left(\frac{1+a_0^T x_i}{1+a_0^T x_{i+1}}\right)^2\left[\left(\frac{f_i}{f_{i+1}}\right)\left(\frac{f_i+1}{f_{i+1}+1}\right)+\left(\frac{n}{f_{i+1}\left(f_{i+1}+1\right)}\right)\left(\frac{1+a_0^T x_i}{1+a_0^T x^*}\right)\right]$$

$$\rho_i^{NEWC}=\vartheta^2\left[\left(\frac{f_i}{f_{i+1}}\right)\left(\frac{f_i+1}{f_{i+1}+1}\right)+\left(\frac{n}{f_{i+1}\left(f_{i+1}+1\right)}\right)\xi\right] \tag{71}$$

where

$$\vartheta=\frac{1+a_0^T x_i}{1+a_0^T x_{i+1}} \quad , \quad \vartheta=\frac{g_{i+1}^T \Delta x}{\Delta f+k_i} \tag{72}$$

where $\Delta f=f_{i+1}-f_i$ and $\Delta x=x_{i+1}-x_i$ and $k_i=\left[\Delta f^2-g_{i+1}^T \Delta x g_i^T \Delta x\right]^{1/2}$

and $\xi=\frac{1+a_0^T x_i}{1+a_0^T x^*}$, $\xi=\frac{\Delta \hat{f}+\hat{k}_i}{g_i^T \Delta \hat{x}}$ (73)

where

$$\Delta \hat{f}=f_i-f^* \quad , \quad \Delta \hat{x}=x_i-x^* \quad , \quad \hat{k}_i=\left[\Delta \hat{f}^2-g_i^T \Delta \hat{x}\, g^{*T} \Delta \hat{x}\right]^{1/2}$$

since x^* is the minimum, then $g^*=g\left(x^*\right)= \nabla f\left(x^*\right)=0 \rightarrow \quad \bar{k}_i=\Delta f$
then

$$\xi=\frac{2 \Delta \hat{f}}{g_i^T \Delta \hat{x}} \tag{74}$$

5. Outlines of The New Extended Conic Model CG-Method:

Step (1): Set x_0 the initial point, ε the scalar termination, μ_0 the initial parameter, and let β_0 = 0, $d_{-1}=0$ and w_0=0 for i = 0, 1, . . . , n.

Step (2): $d_0 = -\hat{h}_0'$

Step (3): Set $x_{i+1} = x_i + \lambda_i d_i$, $i \geq 0$
where λ_i is obtained from the new non-monotone line search procedure defined in section 3
Step (4): Compute

$$a_0 = \frac{(\lambda_t - \lambda_i)\gamma_i^2 g_i + \lambda_i \gamma_t^2 g_t - \lambda_t \gamma_{i+1}^2 g_{i+1}}{(\lambda_t - \lambda_i)\gamma_i^2 d^T g_i + \lambda_i \gamma_i^2 d_t^T g_t - \lambda_t \gamma_{i+1}^2 d_{i+1}}$$

Step (5): If $|\Delta f| \le \varepsilon$ or $|f_{i+1}| \le \varepsilon$, then set $\rho_i = 1.0$ and go to step (7).

Else go to step (6), where ε is a small number (i.e. 0.1E-5).

Step (6): Compute ρ_i which is given in (71).

Step (7): Calculate the new direction:

$$d_i = -\hat{h}_i' + \beta_i d_{i-1}$$

where β_i is the conjugacy coefficient of conic model, and β_i is expressed as follows:

$$\beta_i = \rho_i \frac{\| \hat{h}_{i+1}' \|^2}{\| \hat{h}_i' \|^2} \text{ , (modified F/R),}$$

Step (8): Check for convergence if $\frac{1}{\mu} \sum_{i=1} h_i(x) < \in$ then stop. Otherwise set $\mu_{i+1} = \mu_i * 10$ and takes $x = x^*$ as a new starting point.

Step (9): Check for restarting criterion if $d_i^T g_{i+1} \ge -0.8 | g_{i+1}^T g_{i+1} |$, is satisfied, then go to step (2), otherwise go to step (3).

6. Numerical Results:

In order to test the effectiveness of the new constraint optimization Method which had used a non-quadratic conic model, a number of non-linear constraint test functions have been chosen and solved numerically by utilizing the new and standard established methods. So this new model has been compared with the standard conic model for CG-methods. It is found that the new Method is better than the previous published model shown in Table (6.1). All the results are obtained using (Pentium 4 computer). All programs were written in Fortran language and for all cases the stopping criterion taken to be, Bunday (1984).

$$\|g_{k+1}\| \le 1 * 10^{-5} \quad \text{and} \quad \frac{1}{\mu} \sum_{i=1}^{l} h_i(x_k) \qquad (75)$$

The comparative performance for all of these methods are evaluated by considering NOF, NOI, and NOC where NOF is the number of function evaluations, NOI is the number of iterations and NOC is the number of constraint evaluations.

Table 6.1: Comparison between Standard and New conic model methods

Test Functions	Standard Model Method			New Conic Model Method		
	NOI	NOF	NOC	NOI	NOF	NOC
1	1759	3525	1761	1756	3521	1751
2	16	47	21	14	43	19
3	73	158	103	64	135	99
4	27	56	31	24	56	26
5	3	33	883	2	29	802
6	19	43	22	19	43	22
7	50	242	56	43	140	45
8	15	33	23	13	30	20
9	55	125	236	55	125	236
Total	2017	4262	3136	1990	4122	3020

It is clear from the above table that taking NOI, NOF, NOC of the standard conic model we have an improvements for all NOI, NOF and NOC. The new proposed method which is based on the non-quadratic conic model and for all the nine cases of NOI, NOF and NOC is the best or equal. i.e. there are an improvement of 100% in NOI, NOF and NOC. However, overall total, taking the standard conic model method as 100% (NOI, NOF and NOC) separately, then the new proposed method requires about 98%(NOI), 96% (NOF) and 95% (NOC). This indicates that the new method is comparably effective against the standard method, for our selected set of test problems.

7. Conclusions:

In this paper we have been, first changed Wolfe's conditions defined by:

$$f(x_k) - f(x_k + \alpha_k d_k) \geq -\delta \alpha_k g_k^T d_k \tag{76}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \tag{77}$$

With $0 < \delta < \sigma < 1$, into a new non-monotone line search defined in (24) and (25). The new extended non-quadratic conic model defined in the new proposed Method satisfy the global convergence property with the new line search procedure defined in (24) and (25). The new proposed extended CG-Method with it's non-monotone line search defined in section (3) satisfies the descent condition and hence it has super linear convergence.

REFERENCES

Al-Baali, M., 1985. Descent Property and Global Convergence of the Fletcher-Reeves Method with Inexact Line Search, IMA Journal of Numerical Analysis, 5(1): 121-124.

Al-Bayati, A.., 1993. A New Non-Quadratic Model for Unconstrained Non-Linear Optimization. Mu'tah J. Res. Stud. Nat. Applied Science. Series., 8: 131-155.

Al-Bayati, A.Y., 2001. New Generalized CG-Methods for the Non-Quadratic Model in Unconstrained Optimizations. J. Al-Yarmouk, 10: 9-25.

Al-Bayati, A.Y., 2007. New Extended CG-methods for Combined Non-Quadratic Models in Non-Linear Optimization. J. Al-Yarmouk, 16: 347-366.

Al-Bayati, A.Y. and G.M. Al-Naemi, 1995. New Extended CG-Methods for Non-Linear Optimization, Mu'tah J. Res. Stud. Nat. Applied Science. Series., 10: 69-87.

Al-Bayati, A.Y. and H.N. Jabbar, 2010. A New Separable Logarithmic Algorithm for Solving Nonlinear Optimization, The American Journal of Applied Science, Science Publication, New York, 6(5): 533-540.

Al-Bayati, A.Y. and R.S. Mohammed, 2010. New Scaled Sufficient Descent Conjugate Gradient Algorithm for Unconstraint Optimization, The American Journal of Computer Sciences, Science Publication, New York, 6(5): 546-553.

Andrei, N., 2007. Scaled Conjugate Gradient Algorithms for Unconstrained Optimization, Computational Optimization and Applications, 38(3): 401-416.

Andrei, N., 2009. Another Non-Linear Conjugate Gradient Algorithm for Unconstrained Optimization, Optimization Methods & Software, 24(1): 89-104.

Beale, E.M.L., 1972. A Derivation of Conjugate Gradient, In: Numerical Methods For Nonlinear Optimization. Edited by Lootsma, F.A., Academic Press, pp: 39-43.

Boland, W. and J. Kowalik, 1979. Extended CG-Method with Restarts. JOTA, 28: 1-9.

Bonnans, F.J. and C.J. Gilbert, 2006. Numerical Optimization: Theoretical and Practical Aspects, Springer-Verlag Belin Heidelberg.

Byrd, R., J. Nocedal and Y. Yuan, 1987. Global Convergence of a Class of Quasi-Newton Method on Convex Problems. SIAM, J. Number. Analysis, 24: 1171-1190.

Bunday, B., 1984. A Basic Optimization Methods, Edward Arnold, Bedford Square, London, LTD.

Coxeter, H., 1964. Projective Geometry, Blasidell Publishing Company, New-York.

Dai, Y.H. and L.Z. Liao, 2001. New Conjugacy Conditions and Related Nonlinear Conjugate Gradient Methods, Applied mathematics an optimization, 43: 87-101.

Dai, Y.H. and Y. Yuan, 1999. A Nonlinear Conjugate Gradient Method with a Strong Global Convergence Property, SIAM Journal on optimization, 10: 177-182.

Davidon, W., 1980. Conic Approximation and Collinear Scaling for Optimizers, SIAM J. Numerical Analysis, 17(2): 268-281.

Dennis, J. and R. Schnabel, 1983. Numerical Methods for Unconstrained Optimization and Non-Linear Equations, Prentice-Hall, New York.

Fletcher, R. and C.M. Reeves, 1964. Function Minimization by Conjugate Gradients, Computer. J., 7: 149-154.

Fried, I., 1971. N-Step Conjugate Gradient Minimization Scheme for Non-Quadratic Function, AIAA Journal, 19: 2286-2287.

Gilbert, J. and J. Nocedal, 1992. Global Convergence Properties of Conjugate Gradient Methods for Optimization, SIAM Journal on Optimization, 2(1): 21-42.

Hager, W.W. and H. Zhang, 2006. A New Conjugate Gradient Method with Guaranteed Descent and an Efficient Line Search, SIAMJ. Optimization, 16: 170-192.

Hestenes, M.R. and E.L. Stiefel, 1952. Methods of Conjugate Gradient for Solving Linear Systems, Journal of Research of the National Bureau of Standards, 49: 409-436.

Nocedal, J. and D. Liu, 1989. Algorithms with Conic Termination for Non-Linear Optimization, SIAM, J. Sc. Computer, 10(1): 1-7.

Polak, E. and G. Ribière, 1969. Note Sur la Convergence De Directions Conjug´ees, Revue Francaised 'Informatique et de Recherche Operationnelle, 16: 35-43.

Powell, M.J.D., 1977. Restart Procedures for the Conjugate Gradient Method, Mathematical Programming, 12: 241-254.

Powell, M., 1982. Non-Linear Optimization, Academic Press. Inc., London, LTD.

Powell, M., 1984. Non-Convex Minimization Calculations and the Conjugate Gradient Method, in Numerical Analysis (Dundee, 1983), 106: 122-141, Springer, Berlin, Germany.

Spedicato, E., 1976. A Variable Metric Method for Function Minimization Derived From Invariancy to Non-Linear Scaling, JOTA, 20: 315-329.

Wei, Z., G. Li and L. Qi, 2008. Global Convergence of the Polak-Ribi`ere-Polyak Conjugate Gradient Method with an Armijo-type Inexact Line Search for Non-Convex Unconstrained Optimization Problems, Mathematics of Computation, 77(264): 2173-2193.

Yin, H.X. and D.L. Du, 2007. The Global Convergence of Self–Scaling BFGS Algorithm with Non-Monotone Line Search for Non-Convex Optimization Problems, Acta Mathematica, 23(7): 1233-1240.

Zoutendijk, G., 1970. Nonlinear Programming Computational Methods in Integer and Nonlinear Programming, Abadie. J. ed., North-Holland, Amsterdam, pp: 37-86.

Appendix:

All the test functions used in this paper are from general literature:

1. $\min f(x) = (x_1 - 2)^4 + (x_1 - 2x_2)^2$

s.t: $x_1^2 - x_2 = 0$

2. $\min f(x) = x_1 x_2^2 + 2$

s.t.: $x_1^2 - x_2^2 = -2$

3. $\min f(x) = (x_1 - 2)^2 + (x_1 - 2x_2)^2$

s.t.: $x_1^2 - x_2^2 = -4$

4. $\min f(x) = x_1^2 + x_2^2$

s.t.: $x_1^2 - x_2^2 = -1$

5. $\min f(x) = 0.5x_1^2 + 2.5x_2^2$

s.t.: $x_1 - x_2 - 1 = 0$

6. $\min f(x) = x_1^2 + x_2^2$

s.t.: $(x_1 - 1)^2 + x_2^2 = 0$

7. $\min f(x) = (x_1 - 1)^2 + (x_2 - 2)^2 + (x_3 + 2)^2$

s.t.: $x_1^2 + 4x_2^2 + x_3^2 + 36 = 0$

8. $\min f(x) = x_1^2 + x_2^2$

s.t.: $1 - x_1^2 - x_2^2$

9. $\min f(x) = x_1^2 + x_2^2$

s.t.: $(x_1 - 1)^3 - x_2^2 + 4 = 0$