# *Chapter-5*

# *A modified form of conjugate direction for general nonlinear function and its convergence*

In this chapter we have derived a new expression of the Conjugate Gradient parameter $\beta$ for the non linear Conjugate Gradient method. The convergence of the CG method with the proposed $\beta$ has been discussed .Also, the proposed $\beta$ has been tested for a good number of nonlinear test functions.

Conjugate Gradient(C.G) method has a very important role in the field of large-scale unconstrained optimization problems. In this paper we have obtained a new conjugate direction for general unconstrained nonlinear function with its convergence analysis. The result is then verified with different standards test functions and compared with other existing results

$$\min\{f(x): x \in R^n\} \tag{5.1}$$

Where $f: R^n \to R$ is a continuously differentiable function especially if the dimension $n$ is large.

The Conjugate Gradient method to solve the general nonlinear problem defined by (5.1) of the form

$$x_{k+1} = x_k + \alpha_k d_k \tag{5.2}$$

Where $\alpha_k$ is a step size obtained by a line search and $d_k$ is the search direction defined by

$$d_k = \begin{cases} -g_k, k = 0 \\ -g_k + \beta_k d_{k-1}, k \geq 1 \end{cases} \tag{5.3}$$

Where $\beta_k$ is a parameter and $g_k$ denotes $\nabla f(x_k)$ where the gradient of $f$ at $x_k$ is row vector and $g_k$ is a column vector .

Different C.G methods corresponds to different choices for the scalar $\beta_k$ are as follows

$$\beta_k^{HS} = \frac{g_{k+1}^T}{d_k^T y_k}$$

was proposed in 1952 in the original (linear) CG paper of Hestenes and Stiefel, in1964 Fletcher and Reeves proposed the first nonlinear Conjugate Gradient method

$$\beta_k^{FR} = \frac{\| g_{k+1}^T \|^2}{\| g_k^T \|^2}$$

in 1967 Daniel proposed

$$\beta_k^{D} = \frac{g_{k+1}^T \nabla^2 f(x_k) d_k}{d_k^T \nabla^2 f(x_k) d_k}$$

Which requires evaluation of the Hessian, $\nabla^2 f(x_k)$. Polak Ribiere and by Polyak in 1969 proposed $\beta_k^{PRP} = \frac{g_{k+1}^T y_k}{\| g_k^T \|^2}$

In the year 1987

$$\beta_k^{CD} = \frac{\| g_{k+1}^T \|^2}{-d_k^T g_k}$$

was proposed by Fletcher , CD stands for "Conjugate Descent", ž a new beta

$$\beta_k^{LS} = \frac{g_{k+1}^T y_k}{-d_k^T g_k}$$

was proposed by Liu and Storeyin ,1991, Dai and Yuan proposed

$$\beta_k^{DY} = \frac{\| g_{k+1}^T \|^2}{d_k^T y_k}$$

In 1991,

$$\beta_k^{N} = \frac{1}{d_k^T y_k}\left( y_k - \frac{2 d_k \| y_k \|^2}{d_k^T y_k} \right) g_{k+1}$$

was proposed by Hager and Zhang in2005.

It is known from (5.2) and (5.3) that .only the step size $\alpha_k$ and the parameter $\beta_k$ remain to be determined in the definition of Conjugate Gradient. In this case that if $f$ is a convex quadratic, the choice of $\beta_k$ should be such that the method (5.2)-(5.3) reduces to the linear Conjugate Gradient method if the line search is exact namely

$$\alpha_k = \arg\min\{ f(x_k + \alpha d_k); \alpha > 0\} \qquad (5.4)$$

For nonlinear functions, different formulae for the parameter $\beta_k$ result in different Conjugate Gradient methods and their properties can be significantly different. To differentiate the linear Conjugate Gradient method ,sometimes we cal the Conjugate Gradient method for unconstrained optimization by nonlinear Conjugate Gradient method. Meanwhile the parameter $\beta_k$ is called Conjugate Gradient parameter . In the general nonlinear Conjugate Gradient method , without restarts is only linearly convergent (Crowder and Wolfe[54] while n-step quadratic convergence rate can be established if the method is restarted along the along the negative gradient every n-step(see Cohen [55] and McCormick and Ritter [56]) .

In 1964 the method has been extended to nonlinear problems by Fletcher and Reeves[44],which is usually considered as the first nonlinear Conjugate Gradient algorithm. Since there is a large number of variations of Conjugate Gradient algorithms have been suggested. A survey on their definition including 40 nonlinear Conjugate Gradient algorithms for unconstrained optimization is given by Andrei[57]. Since the exact line search is usually expensive and impractical, the strong Wolfe line search is often consider the implementation of the nonlinear Conjugate Gradient methods. Its aims to find a step size satisfying the strong Wolfe conditions

$$
\begin{aligned}
f(x_k + \alpha_k d_k) - f(x_k) &\le \rho \alpha_k g_k^T d_k \\
\left| g(x_k + \alpha_k d_k)^T \right| &\le -\sigma g_k^T d_k \\
\textit{where } 0 < \rho < \sigma < 1.&
\end{aligned} \tag{5.5}
$$

The strong Wolfe line search is often regarded as a suitable extension of the exact line search since it reduces to the latter. If $\sigma$ is equal to zero, in practical computation a typical choice for $\sigma$ that controls the inexactness of the line search is $\sigma$=0.1.On the other hand general non linear function ,one may be satisfy with a step size satisfying the standard wolf conditions ,namely (5.5) and

$$
g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k \quad \textit{where } 0 < \rho < \sigma < 1. \tag{5.6}
$$

As is well known the standard Wolf line search is normally used in the implementation of Quasi-Newton methods, another important class of methods for unconstrained optimization. The work of Dai and Yuan indicates that the use of standard Wolfe line search is possible in the nonlinear Conjugate Gradient field. A requirement for an optimization method to use the above line searches is that, the search direction $d_k$ must have descent property namely

$$g_k^T d_k < 0 \tag{5.7}$$

For Conjugate Gradient method, by multiplying (5.3) with $g_k^T$ , we have

$$g_k^T d_k = -\| g_k \|^2 + \beta_k g_k^T d_{k-1}$$

Thus if the line search is exact, we have $g_k^T d_k = -\| g_k \|^2$ since $g_k^T d_{k-1} = 0$ .Consequently $d_k$ is descent provided $g_k \neq 0$.In this paper we say that a Conjugate Gradient method is descent if (5.7) holds for all k and is sufficient descent if the sufficient descent condition

$$g_k^T d_k \leq -c \| g_k \|^2$$

Holds for all k and some constant $c > 0$ . However we have to point out that the borderlines between these Conjugate Gradient methods are not strict.

If $s_k = x_{k+1} - x_k$ and in the following $y_k = g_{k+1} - g_k$ .Different Conjugate Gradient algorithms corresponds to different choices for the parameter $\beta_k$. Either of the following assumptions are often utilised in convergence analysis for C.G algorithms

**1.1** Lipschitz Assumption **:** In some neighbourhood $N$ Of the level set $\Omega = \{x \in R^n : f(x) \leq f(x_0)\}$ .The gradient $\nabla f(x)$ is Lipschitz continuous. That is there exist a constant $L < \infty$ such that $\| \nabla f(x) - \nabla f(y) \| \leq L \| x - y \| \forall x, y \in N$

**1.2** Boundedness Assumption **:** The level set $\Omega$ is bounded. That is there exist a constant $B < \infty$ such that $\| x \| \le B \forall x \in \Omega$

The conclusion of the following theorem is often called the Zoutendijk condition, is used to prove the global convergence of the non linear CG methods ; it was originally given by Zoutendijk [58] and Wolfe [59,60]

The objective of this present work is to investigate the conjugacy condition of the general nonlinear functions taking an arbitrary initial search direction although the convergence for general linear as well as nonlinear functions with an initial search direction already exist.

In this chapter we obtain a new conjugacy condition of the general nonlinear functions taking an arbitrary initial search directions as discussed by Sumit Saha and Biprajit Nath in Chapter 4

Here we have taken the initial direction as

$$d_0 = -g_0 + \gamma g_0 \text{ instead of } d_0 = -g_0 \text{ where } \lambda \in (0,1) \tag{5.8}$$

So the general formula for $d_k$

$$d_k = \begin{cases} -g_k + \gamma g_k, k = 0 \\ (-g_k + \gamma g_k) + \beta_k d_{k-1}, k \ge 1 \end{cases} \tag{5.9}$$

Again from the findings of Dai,Liao[7]

$$d_k^T g_{k-1} = -t g_k^T s_{k-1} \tag{5.10}$$

$$\Rightarrow g_k^T g_{k-1} + \beta_k d_{k-1}^T g_{k-1} = -t g_k^T s_{k-1}$$

$$\Rightarrow d_{k-1}^T g_{k-1} = \frac{-t g_k^T s_{k-1} - (\gamma - 1) g_k^T g_{k-1}}{\beta_k}$$

$$\Rightarrow \beta_k = \frac{-t g_k^T s_{k-1} - (\gamma - 1) g_k^T g_{k-1}}{d_{k-1}^T g_{k-1}} \tag{5.11}$$

Where $t \ge 0$ is a scalar

From assumptions 1.1 and 1.2 we have $|g_k| \le BL < \infty$ .Also from sufficient descent condition $g_k^T d_k \le c \| g_k \|^2$, holds for all k and some constant $c > 0$

From (5.10) we have

$$\beta_k d_k^T g_{k-1} + (\gamma - 1) g_k^T g_{k-1} = -t g_k^T s_{k-1}$$

$$\Rightarrow s_{k-1} = \left[ \frac{\beta_k d_k^T g_{k-1}^T + (\gamma - 1) g_k^T g_{k-1}}{-t \| g_k \|^2} \right] g_k$$

$$\Rightarrow |s_{k-1}| < \beta_k cBL + \frac{|1-\gamma|}{t}$$

$$\Rightarrow |s_{k-1}| < \frac{|1-\gamma| BL}{t} \tag{5.12}$$

Therefore the C.G method converges for $t > |1-\gamma|$