Chapter-2

Survey of Literature

The following chronological list gives us an idea about some choices for the conjugate gradient update parameter

In 1952, Hestenes and stiefel[45] proposed the formula as $\beta_k^{HS} = \frac{y_k^T g_{k+1}}{d_k^T y_k}$ in the original (linear) Conjugate Gradient paper.

Here we briefly discuss the Hestenes-Stiefel (HS) conjugate gradient method namely

$$x_{k+1} = x_k + \alpha_k d_k \tag{2.1}$$

Where α_k is the step size obtained by a line search and d_k is the search direction defined by

$$d_k = \begin{cases} -g_k, k = 1 \\ -g_k + \beta_k d_{k-1}, k \geq 2 \end{cases} \tag{2.2}$$

Where β_k is a parameter and g_k denotes $\nabla f(x_k)$ where β_k is calculated by

$$\beta_K^{HS} = -\frac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}} \tag{2.3}$$

Such a formula is first used by Hestenes and Stiefel in the proposition of the linear conjugate gradient method in 1952 .

A remarkable property of the HS method is that, no matter whether line search is exact or not by multiplying (2.2) with y_{k-1} and using (2.3) we always have

$$d_k^T y_{k-1} = 0 \tag{2.4}$$

In the quadratic case, y_{k-1} is parallel to Ad_{k-1}, where A is the Hessian of the function. Then (2.4) implies $d_k^T A d_{k-1} = 0$ namely d_k is conjugate to d_{k-1}.For this reason the relation (2.4) is often called conjugacy condition.

If the line search is exact, we have by

$$g_k^T d_k = -\| g_k \|^2 + \beta_k g_k^T d_{k-1} \tag{2.5}$$

That $g_k^T d_k = -\| g_k \|^2 \; and \; \beta_k^{HS} = \beta_k^{PRP}$ .Therefore the HS method is identical to the PRP method in case of exact line searches. One major observation is that, when $\| s_{k-1} \|$ is small both the nominator and denominator of β_k^{HS} become small so that β_k^{HS} might be unbounded. Another observation is that, for any one dimensional function, we always have

$$d_2 = -g_2 + \beta_2^{HS} d_1 = -g_2 + \frac{g_2.y_1}{d_1.y_1} d_1 = -g_2 + g_2 = 0 \qquad (2.6)$$

independent of the line search. Consequently there is some special to ensure the descent property of the HS method with exact line searches. Further , we can similarly modify the standard Wolfe line search to ensure the sufficient descent condition and global convergence for the HS^+ method. If the sufficient descent condition $g_k^T d_k \leq -c \| g_k \|^2$ is relaxed to the descent condition, Qi et al established the global convergence of the modified HS method, where β_k takes the form

$$\beta_k^{QHL} = \max\{o, \min\{\beta_k^{HS}, \frac{1}{\| g_k \|}\}\} \qquad (2.7)$$

Early 1977,Perry observed that the direction in the HS method can be written as

$$d_k = -P_k g_k \qquad (2.8)$$

Where $$P_k = I - \frac{d_{k-1} y_{k-1}^T}{d_{k-1}^T y_{k-1}}$$

Noting that $P_k^T y_{k-1} = 0, P_k$ is a affine transformation that transforms R^n into the null space of y_{k-1}.To ensure the descent property of d_k, however, we may wish the matrix P_k is positive definite. It is obvious that there is no positive definite matrix P_k such that the conjugacy condition (2.4) holds. In case of exact line searches, it is sufficient to require P_k to satisfy

$$P_k^T y_{k-1} = s_{k-1}, \qquad (2.9)$$

Which exactly the quasi-Newton equation.

First nonlinear Conjugate Gradient method ,proposed by Fletcher and Reeves in 1964 and proposed the formula as $\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}$.

For solving the unconstrained optimization problem

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

where *where* $f : R^n \to R$ is continuously differentiable .Dai and Yuan suggested the following nonlinear conjugate gradient algorithm

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

Where the step size α_k is positive and the direction d_k are computed by the rule

$$d_{k+1} = -g_{k+1} + \beta_k^{DY} s_k, d_0 = -g_0$$
$$where\ \beta_k^{DY} = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} \tag{2.10}$$

Using a standard Wolfe line search, the Dai and Yuan method always generates descent directions and under Lipschitz assumption it is globally convergent. Here we present a modification of the Dai and Yuan computational scheme as:

$$\begin{aligned} where\ \ & g_k = \nabla f(x_k) \\ & d_{k+1} = -g_{k+1} + \beta_k^A s_k \\ & d_0 = -g_0 \end{aligned} \tag{2.11}$$

$$where\ \beta_k^A = \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} - \delta_k \frac{(g_{k+1}^T s_k)(g_{k+1}^T g_{k+1})}{(y_k^T s_k)^2} \tag{2.12}$$

and δ_k is a positive parameter which follows to be determined. It is assume that $y_k^T s_k \neq 0$,so that β_k^A is well-defined.(Using a standard Wolfe line search, $y_k^T s_k > 0\ when\ g_k \neq 0$.)

The general form (2.12) which is used to deduce the β_k^A in (2.11) has also been proposed. The motivation of the proposed β_k^A in (2.12) were introduced here is that under suitable conditions it assures the descent character of the direction d_{k+1} in (2.11). The following theorem proves this.

Theorem: *If* $y_k^T s_k \neq 0\ and\ d_{k+1} = -g_{k+1} + \beta_k^A s_k (d_0 = -g_0)\ where\ \beta_k^A$ is given by (2.12),then

$$g_{k+1}^T d_{k+1} \leq -\left(1 - \frac{1}{4\delta}\right) \| g_{k+1} \|^2 \tag{2.13}$$

Proof:

Since

$$d_o = -g_0$$

therefore $g_0^T d_0 = -\| g_0 \|^2$ which satisfy (2.13)

Multiplying (2.11) by g_{k+1}^T

$$g_{k+1}^T d_{k+1} = -\| g_{k+1} \|^2 + \frac{(g_{k+1}^T g_{k+1})(g_{k+1}^T s_k)}{y_k^T s_k} - \delta_k \frac{\| g_{k+1} \|^2 (s_k^T g_{k+1})^T}{(y_k^T s_k)^2} \tag{2.14}$$

But

$$\frac{(g_{k+1}^T g_{k+1})(g_{k+1}^T s_k)}{y_k^T s_k} = \frac{\left[(y_k^T s_k) g_{k+1}/\sqrt{2\delta_k}\right]^T \left[\sqrt{2\delta_k}\,(g_{k+1}^T s_k) g_{k+1}\right]}{(y_k^T s_k)^2}$$

$$\le \frac{\frac{1}{2}\left[\frac{1}{2\delta_k}(y_k^T s_k)^2 \|g_{k+1}\|^2 + 2\delta_k (g_{k+1}^T s_k)^2 \|g_{k+1}\|^2\right]}{(y_k^T s_k)^2}$$

$$= \frac{1}{4\delta_k}\|g_{k+1}\|^2 + \delta_k \frac{(g_{k+1}^T s_k)^2 \|g_{k+1}\|^2}{(y_k^T s_k)^2} \qquad (2.15)$$

Using (2.15) in (2.14) gives (2.13)

To conclude the sufficient descent condition from(2.13),the quantity $1-\frac{1}{4\delta_k}$ is required to nonnegative. Supposing that $1-\frac{1}{4\delta_k}>0$,then the direction given by (2.11)and (2.12) is a descent direction. Dai and Yuan present conjugate gradient schemes with the property that $g_k^T d_k < 0$ when $y_k^T s_k > 0$.

If f is strongly convex or the line search satisfies the Wolfe conditions, then $y_k^T s_k > 0$ and the Dai and Yuan scheme yield descent.

It is also observed that ,if for all k , $\frac{1}{4\delta_k} \le 1$ and the line search satisfies the Wolfe conditions ,then for all the search direction (2.11) and (2.12) satisfy the sufficient descent condition.

It is observed that if f is a quadratic function and f is selected to achieve the exact minimum of in the direction d_k ,then $s_k^T g_{k+1} = 0$ and the formula (2.12) for β_k^A reduces to the Dai and Yuan computational scheme. However here they consider general nonlinear functions and inexact line search.

The numerical experiments with algorithm (2.11) and (2.12) shows that for different choices of δ_k .its performance is quite different. Therefore in order to get an efficient algorithm in the following present a procedure for δ_k computation. Mainly this is based on the conjugacy condition. Dai and Liao introduced a new conjugacy condition as $y_k^T d_{k+1} = -t s_k^T g_{K+1}$,where $t \ge 0$ is a scalar. This is reasonable since in the general inexact line search is used in real computation. However this condition is very dependent by the nonnegative parameter t. It is observed that using (2.12) in (2.11) the following direction can be obtained:

$$d_{k+1} = -g_{k+1} + \frac{g_{k+1}^T g_{k+1}}{y_k^T s_k} s_k - \delta_k \frac{\|g_{k+1}\|^2}{(y_k^T s_k)^2}(g_{k+1}^T s_k) s_k \tag{2.16}$$

Which can be written as

$$d_{k+1} = -Q_{K+1} g_{k+1} \tag{2.17}$$

Where the matrix Q_{K+1} is

$$Q_{K+1} = I - \frac{s_k g_{k+1}^T}{y_k^T s_k} + \delta_k \frac{\|g_{k+1}\|^2}{(y_k^T s_k)^2}(s_k s_k^T) \tag{2.18}$$

Now by symmetrisation Q_{k+1} as:

$$\overline{Q}_{k+1} = I - \frac{s_k g_{k+1}^T + g_{k+1} s_k^T}{y_k^T s_k} + \delta_k \frac{\|g_{k+1}\|^2}{(y_k^T s_k)^2}(s_k s_k^T) \tag{2.19}$$

Which can consider the direction

$$d_{k+1} = -\overline{Q_{k+1}} g_{k+1} \tag{2.20}$$

The reason of symmetrisation the $Q_{K+1}\, as\, \overline{Q_{k+1}}$ in (2.19) is that the direction d_{k+1} computed as in (2.20) resembles the quasi-Newton methods. However it uses only the symmetry and do not modify further $\overline{Q_{k+1}}$ in order to satisfy the quasi-Newton equation.

Now from the conjugacy condition $y_k^T d_{k+1} = 0$

$$\text{i.e.} \quad y_k^T \overline{Q_{k+1}} g_{k+1} = 0 \tag{2.21}$$

after some computation it follows that

$$\delta_k = \frac{y_k^T s_k}{g_{k+1}^T s_k} + \frac{g_{k+1}^T y_k}{\|g_{k+1}\|^2} - \frac{(g_{k+1}^T y_k)(y_k^T s_k)}{\|g_{k+1}\|^2 (g_{k+1}^T s_k)} \tag{2.22}$$

Therefore using (2.22) in (2.12) gives the expression for β_k^A as follows

$$\beta_k^A = \frac{1}{y_k^T s_k}\left(y_k - \frac{g_{k+1}^T y_k}{y_k^T s_k} s_k \right)^T g_{k+1} \tag{2.23}$$

Observe that β_k^A from (2.23)can be written as:

$$\beta_k^A = \frac{y_k^T g_{k+1}}{y_k^T s_k}\left[1 - \frac{s_k^T g_{k+1}}{y_k^T s_k} \right] = \frac{y_k^T g_{k+1}}{y_k^T s_k}\left(- \frac{s_k^T g_{k+1}}{y_k^T s_k} \right) \tag{2.24}$$

Assuming that d_k is a descent direction i.e $g_k^T s_k \le 0$ and the step length α_k is determined by the Wolfe line search conditions, it is observed that β_k^A is positive multiplicative modification of $\beta_k^{HS} = y_k^T g_{k+1} / y_k^T s_k$.

If the line search is exact, then in this case $s_k^T g_{k+1} = 0$ and therefore $\beta_k^A = \beta_k^{HS}$,for which the conjugacy condition holds.

Denoting $d_{k+1}\, and\, d_{k+1}^{HS}$ to be the search directions given by β_k^A and β_k^{HS} respectively, namely $d_{k+1} = -g_{k+1} + \beta_k^A s_k$ and $d_{k+1}^{HS} = -g_{k+1} + \beta_k^{HS} s_k$

Then it was found that $$g_{k+1}^T d_{k+1} = g_{k+1}^T d_{k+1}^{HS} - \beta_k^{HS} \frac{(s_k^T g_{k+1})^2}{y_k^T s_k} \qquad (2.24)$$

Again assuming that $g_{k+1}^T d_{k+1}^{HS} < 0$ and $\beta_k^{HS} \ge 0$ (2.24) gives the condition $g_{k+1}^T d_{k+1} < 0$.

Thus if the direction generated by the HS method is descent, $\beta_k^{HS} \ge 0$ and if the line search is given by the Wolfe conditions, then the direction generated by β_k^A must also be a descent direction.

Here we can observe that the direction

$$d_{k+1} = -g_{k+1} - \frac{(y_k^T g_{k+1})(s_k^T g_k)}{(y_k^T s_k)^2} s_k \qquad (2.25)$$

Is not a descent direction at every iteration. However, since $-s_k^T g_k / y_k^T s_k \ge 0$,when d_k is a descent direction, then if

$$\frac{y_k^T s_k}{s_k^T g_k} \| g_{k+1} \|^2 + \frac{(y_k^T g_{k+1})(s_k^T g_{k+1})}{y_k^T s_k} \le 0$$

It follows that $g_{k+1}^T d_{k+1} \le 0$.

On the other hand, we observe that

$$\frac{y_k^T s_k}{s_k^T g_k} \| g_k \|^2 \le 0$$

and tends to zero.

Therefore if

$$\frac{(y_k^T g_{k+1})(s_k^T g_{k+1})}{y_k^T s_k} \le \frac{y_k^T s_k}{\left| s_k^T g_k \right|} \| g_{k+1} \|^2 \qquad (2.26)$$

then $g_{k+1}^T d_{k+1} \le 0$.

In 1967 Denial proposed a new method which requires evaluation of the Hessian $\nabla^2 f(x)$,and the formula used by him is $\beta_k^D = \dfrac{g_{k+1}^T \nabla^2 f(x) d_k}{d_k^T \nabla^2 f(x) d_k}$

In 1969, Polak and Riebiere proposed the formula for β as $\beta_k^{PRP} = \max\{0, \dfrac{g_{k+1}^T y_k}{g_k^T g_k}\}$

In PRP method they considered the following unconstrained optimization problem

To find arg $\min\{f(x) : x \in R^n\}$ (2.27)

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

The Conjugate Gradient method to solve the general nonlinear problem defined by (2.27) is of the form

$$x_{k+1} = x_k + \alpha_k d_k \qquad (2.28)$$

Where α_k is a step size obtained by a line search and d_k is the search direction obtained by

$$d_k = \begin{cases} -g_k, k = 1 \\ -g_k + \beta_k d_{k-1}, k \geq 2 \end{cases} \qquad (2.29)$$

Where β_k is a parameter and g_k denotes $\nabla f(x_k)$ where the gradient $\nabla f(x_k)$ of f at x_k is a row vector and g_k is a column vector .Different C.G methods correspond to different choices for the scalar β_k .Plenty of C.G methods are known and excellent survey of these methods, with special attention on their global convergence properties .

In the PRP method parameter β_k is computed from

$$\beta_k = \frac{g_k y_{k-1}}{\| g_{k-1} \|^2} \qquad (2.30)$$

where $y_{k-1} = g_k - g_{k-1}$ and $\|.\|$ stands for the Euclidean norm. When the step-size α_k is small and the direction d_k is very close to the steepest descent direction $-g_k$ .Thus this PRP method has a built-in restart feature that directly addresses the jamming problem. This feature means that the PRP method has been one of the most efficient Conjugate methods in practical computation for many years Dai[61].Constructed an example to indicate that the PRP method may generate an upward direction resulting in the iterative scheme falling even if the objective function is uniformly convex under the strong Wolfe line search. So, far the convergence of the PRP method has not been completely proved under the Wolfe line line search.

Another popular method to solve the problem(2.27) is the spectral gradient method proposed originally by Barzilai and Borwein.
The direction d_k is generated by

$$d_k = -\theta_k g_k + \beta_k s_{k-1} \tag{2.31}$$

where $s_{k-1} = \alpha_{k-1} d_{k-1}$ and θ_k is the spectral gradient parameter. In [62] Raydan introduce the spectral gradient method for large scale unconstrained optimization problems. An attractive property of this method is that it only needs gradient directions at each line search whereas a non monotone strategy guarantees the global convergence.

Out performs the sophisticated C.G method in many problems. Begin and Martine [67] proposed a spectral gradient method in which is computed from (2.31).One parameter θ_k is generated by

$$\beta_k = \frac{\theta_k g_k^T y_{k-1}}{\alpha_{k-1}\theta_{k-1} \| g_{k-1} \|^2} \tag{2.32}$$

If $\theta_k = \theta_{k-1} = 1$, this is the classical parameter (4). Motivated by the success of spectral gradient method ,they also compute θ_k using

$$\theta_k = \frac{s_{k-1}^T s_{k-1}}{s_{k-1}^T y_{k-1}} \tag{2.33}$$

Under the standard Wolfe line search, they show that the scaled PRP method,(2.31)-(2.33) is very effective. However, the scaled PRP method cannot guarantee the descent direction at each iteration which lead may lead to failure of the iterative scheme.

Because of the advantages of the PRP method and the scaled PRP method, a new method spectral PRP(SPRP) Conjugate Gradient method was developed. The SPRP method not only processes the sufficient descent property and global convergence property, but also satisfies the famous conjugacy condition.

The SPRP method and its descent property : In this method the problem of the type (2.27) was solved using a new iterative method , in which the iterative point is generated by (2.28) and the direction d_k is obtained by

$$d_k = \begin{cases} -g_k, k = 1 \\ -\theta_k g_k + \beta_k d_{k-1}, k \geq 2 \end{cases} \tag{2.34}$$

Where θ_k is the spectral gradient parameter ,and $\beta_k = \beta_k^{PRP}$, Obviously if $\theta_k = 1$,it reduces to the PRP method. In this method, the parameter θ_k is selected in such a way that at each iteration the conjugacy condition is satisfied independent of the line search. Multiplying (2.34) by y_{k-1}^T, we have

$$d_{k-1}^T y_{k-1} = -\theta_k g_k^T y_{k-1} + \beta_k d_{k-1}^T y_{k-1}$$

Hence from the conjugacy condition $d_k^T y_{k-1} = 0$, we obtain

$$\theta_k = \frac{d_{k-1}^T y_{k-1}}{\| g_{k-1} \|^2} \tag{2.35}$$

so the method constructed by (2.34) and (2.35) always satisfies the conjugacy condition, and has structure feature of the spectral gradient method.

In the following algorithm, the specific iterative algorithm is given, and refer to it as the SPRP method.

Algorithm1:

Step 1: Data : $x_1 \in \mathbb{R}^n, \in \geq 0$ .Set $d_1 = -g_0 \; if \| g_1 \| < \in$, then stop.

Step 2: Compute $\alpha_k > 0$ using the standard Wolfe line search.

$$f(x_k + \alpha_k d_k) \leq f(x_k) + \delta \alpha_k g_k^T \tag{2.36}$$

$$g(x_k + \alpha_k d_k)^T d_k \geq \sigma g_k^T d_k \tag{2.37}$$

Where $0 < \rho < \sigma < 1$

Step 3: Let $x_{k+1} = x_k + \alpha_k d_k ; g_{k+1} = g(x_{k+1}),$ if $\| g_{k+1} \| < \in$, then stop`

Step 4: Compute β_{k+1} using (2.30) ; generate θ_{k+1} using (3.35)

Step 5: If $d_{k+1} g_{k+1} > -10^{-3} \| d_{k+1} \| . \| g_{k+1} \|$ is satisfied, we set $d_{k+1} = -\theta_{k+1} g_{k+1}$, otherwise we compute d_{k+1} by

$$d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_{k+1} d_k$$

Step 6 : Set $k = k+1$ go to step 2

Lemma 1: Let the sequences $\{g_k\}$ and $\{d_k\}$ obtained by the SPRP method in which α_k satisfies any line search. Then ,

$$g_k^T d_k < -(w \frac{\| d_{k=1} \|}{\| g_{k-1} \|} \| g_{k-1} \|^2) \qquad (2.38)$$

Where $w > 0$

Lemma 2: Suppose assumptions H holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be obtained by the SPRP method. Then

$$\frac{\sum_{k \geq 1} (g_k^T d_k)^2}{\| d_k \|^2} < +\infty$$

Lemma 3: Suppose assumptions H holds. Let the sequences $\{g_k\}$ and $\{d_k\}$ be obtained by the SPRP method. If there exist a constant $r > 0$ such that

$$\| g_k \|, \forall k \geq 1$$

then , $$\sum_{k \geq 1} \frac{\| d_{k-1} \|^2}{\| d_k \|^2} < +\infty$$

Powell proposed a formula and analysed by Gilbert and Nocedal for β as

$$\beta_k^{PRP} = \max\{0, \frac{g_{k+1}^T y_k}{g_k^T g_k}\}$$

Powell's restart criterion and descent property: As mentioned in the first section, if

$$\sigma < 1$$

any of the FR, PRP and HS method with the strong Wolfe line search may produce ascent search directions even if the objective function is quadratic. Thus special attentions must be given to the problem how to keep the descent property of the Conjugate Gradient methods .In this section we will prove that if the Powell [65]'s restart criterion is applied , the three parameter family of methods with the strong Wolfe line search can guarantee the descent property of each search direction.

When dealing with Beale's three term Conjugate Gradient method, Powell [65] suggested a restart with

$$d_k = -g_k$$

if the following condition is satisfied

$$\left| g_k^T g_{k-1} \right| \leq \in \| g_k \|^2$$

where $\in > 0$ is some positive constant. As Powell [65] observed ,such a restart criterion can avoid that Beale's recurrence to a non-stationary point (a strict convergence result was given in [66] for Beale's method with Powell's restart criterion), and improve the numerical behaviours of Beale's method.

In fact the standard Conjugate Gradient methods, if the function is convex quadratic and the line search is exact, then the relation $g_k^T g_{k-1} = 0$ implies that no restart would take place and finite termination would occur. Thus the quantity $\frac{g_k^T g_{k-1}}{\| g_k \|^2}$ would indicate strong local non quadratic behaviour and hence would be indicative of a need for restarting. In the implementations of Conjugate Gradient methods, Powell restart criterion has been used by many authors, for example Buckley and Lenir [64] and Koda et al [63].

To show the importance of Powell's restart criterion in keeping the descent property of Conjugate Gradient methods , Koda et al [63] first take the HS method as an illustrative example. For this purpose, they define

$$\tau_k = \frac{g_k^T d_k}{\| g_k \|^2}$$

It is obvious that d_k is a descent direction if and if $\tau_k > 0$. For the HS method (2.28),(2.29) and $\beta_k^{HS} = \dfrac{g_k^T y_{k-1}}{d_{k-1}^T y_{k-1}}$, direct calculations yield

$$\tau_k = \frac{g_{k-1}^T d_{k-1}}{d_{k-1}^T y_{k-1}}\left[1 - \frac{g_k^T g_{k-1} g_k^T d_{k-1}}{\| g_k \|^2 \, g_{k-1}^T d_{k-1}}\right]$$

Fletcher proposed a formula for β as $\beta_K^{CD} = \dfrac{g_{k+1}^T g_{k+1}}{-d_k^T g_k}$

Liu and Store proposed a formula for $\beta_K^{LS} = \dfrac{g_{k+1}^T y_k}{-d_k^T g_k}$

Hybrid Liu and Storey-conjugate descent

$$\beta_k^{Hu-Storey} = \max\{0, \min\{\beta_k^{PRP}, \beta_K^{FR}\}\}$$

Hu and Storey proposed a formula for

$$\beta_k^{TA-S} = \beta_k^{PRP} if\, 0 \le \beta_k^{PRP} \le \beta_k^{FR}$$
$$= \beta_k^{FR} otherwise$$

Dai and Liao proposed a formula for $\beta_k^{LD} = \dfrac{g_{k+1}^T (y_k - tS_k)}{d_k^T y_k}$,t>o

In 1987,Fletcher ,CD stands for "Conjugate Descent" $\beta_k^{CD} = \dfrac{\|g_{k+1}\|^2}{-d_k^T g_k}$

In 1991,Liu and Storey proposed a formula for $\beta_k^{LS} = \dfrac{g_{k+1}^T y_k}{-d_k^T g_k}$

In 1999,Dai and Yuan proposed a formula for $\beta_k^{DY} = \dfrac{\|g_{k+1}\|^2}{d_k^T y_k}$

or as $d_{k+1} = -\theta_{k+1} g_{k+1} + \beta_k S_K$ where the parameter θ_{k+1} is a scalar approximation of the inverse Hessian of the function $\boldsymbol{f}$ and β_k is selected as suggested by Bringin and Martinez

$\beta_k^{BM} = \dfrac{g_{k+1}^T (\theta_k y_k - s_k)}{y_k^T s_k}$ scaled form of Perry

$\beta_k^{BM+} = \max\left\{0, \dfrac{\theta_k g_{k+1}^T y_k}{y_k^T y_k}\right\} - \dfrac{g_{k+1}^T s_k}{y_k^T s_k}$ scaled form of Perry

$\beta_k^{SPRP} = \dfrac{\theta_k g_{k+1}^T y_k}{\alpha_k \theta_{k-1} g_k^T g_k}$ scaled form of Polak-Ribiere- Polyak

$$\beta_k^{SFR} = \frac{\theta_k g_{k+1}^T g_{k+1}}{\alpha_k \theta_{k-1} g_k^T g_k}$$ scaled form of Fletcher-Reeves

Recently in the year 2009 some modifications have been done by Abbas Y.AL-Bayati, A.J.Salim and Khalel K.Abbo and some new algorithms proposed by them are

Algorithm1:

$d = -g_{k+1} + \beta_k^{V1} d_k \qquad \sum_{k=0}^{\infty} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < +\infty$ where β_k^{V1} is computed from

$\beta_k^{V1} = \left(1 - \frac{s_k^T y_k}{y_k^T y_k}\right) \frac{g_{k+1}^T y_k}{s_k^T y_k}$ and the initial value $\alpha_k = \alpha_{k-1} \frac{\|d_k\|}{\|d_{k-1}\|}$ where $x_{k+1} = x_k + \alpha_k d_k$

Algorithm 2:

$d = -g_{k+1} + \beta_k^{V2} d_k$ where β_k^{V2} is computed from

$$\beta_k^{V2} = \left(1 - \frac{s_k^T y_k}{y_k^T y_k}\right) \frac{g_{k+1}^T y_k}{d_k^T y_k} + \frac{s_k^T g_{k+1}}{d_k^T y_k}$$

Also in the year 2010, some modifications have been done by Zeng Zin WEI, Hai Dong Huang, Yan Rong Tao and Some new algorithms proposed by them are

Algorithm 3:

$$\beta_k^{HS} = \frac{g_k^T \left(g_k - \frac{g_k^T g_{k-1}}{\|g_{k-1}\|^2} \right) g_{k-1}}{d_{k-1}^T \left(g_k - g_{k-1} \right)} \qquad (2.39)$$

where β_k is a scalar and important feature of this new choice of β_k is that when the line search is adopted, the value β_k is greater than zero.

From $g(x_k+t_kd_k)^Td_k \ge \sigma g_k^Td_k$ and $g_k^Td_k<0$ it implies that

$d_{k-1}^T(g_k-g_{k-1}) \ge (\sigma-1)g_{k-1}^Td_{k-1} \ge 0$

Which along with the equation (2.39) gives

$$\beta_k^{HS*} = \frac{g_k^Tg_k - \dfrac{g_k^Tg_{k-1}}{\|g_{k-1}\|^2}g_k^Tg_{k-1}}{d_{k-1}^T\left(g_k-g_{k-1}\right)}$$

$$= \frac{\dfrac{\|g_k^T\|^2\|g_{k-1}^T\|^2 - g_k^Tg_{k-1}g_k^Tg_{k-1}}{\|g_{k-1}\|^2}}{d_{k-1}^T\left(g_k-g_{k-1}\right)}$$

$$= \frac{\dfrac{\|g_k^T\|^2\|g_{k-1}^T\|^2 - \|g_k^T\|^2\|g_{k-1}^T\|^2\cos^2\alpha_k}{\|g_{k-1}\|^2}}{d_{k-1}^T\left(g_k-g_{k-1}\right)}$$

$$= \frac{\dfrac{\|g_k^T\|^2\|g_{k-1}^T\|^2\left(1-\cos^2\alpha_k\right)}{\|g_{k-1}\|^2}}{d_{k-1}^T\left(g_k-g_{k-1}\right)} \ge 0$$

where α_k is the angle between g_k and g_{k-1}

Huang, Yan Rong Tao made the following assumptions to develop new models:

Assumption A: The level set $\Omega=\{x\in R^n : f(x)\le f(x_1)\}$ is bounded.

Assumption B: In some neighbourhood N of Ω, $f(x)$ is differentiable. The gradient $g(x)=\Delta f(x)$ is Lipschitz continuous,i.e.

$$\exists L>O \text{ s.t.} \| g(x)-g(y) \| \le L \| x-y \|, \forall x,y\in N \qquad (2.40)$$

In the CG method, the step-length α_k is obtained by exact or inexact line search. In practical computation, exact line search is sometimes difficult and the workload is very large, so usually the following inexact line search is used by many researchers. The standard Wolfe linear search is to find the step-length α_k. Satisfying

$$f(x_k+\alpha_kd_k)-f(x_k)\le\delta\alpha_kg_k^Td_k$$
$$g(x_k+\alpha_kd_k)^Td_k>\sigma g_k^Td_k$$
$$where\, 0<\delta<\sigma<1$$

Dai and Yuan put forward a new method $\beta_k^{DY} = \frac{\|g_{k+1}\|^2}{d_k^T y_k}$,such that the line search satisfying the standard Wolfe condition is global convergence.

An equivalent $\beta_k^{DY} = \frac{g_{k+1}^T d_{k+1}}{g_k^T d_k}$ was also proposed by them.

In their work they put forward a new formula of β_k as

$$\beta_k = \frac{\lambda g_{k+1}^T d_{k+1} - \mu \|g_{k+1}\|^2}{g_k^T d_k} \tag{2.41}$$

where λ together with μ is 1.

In the case $g_{k+1}^T d_k \le 0$,we have $0 \le \lambda \le 1$,otherwise, $\lambda > 1$.

For (2.41),in the case λ is 0, $\beta_k = \beta_k^{CD}$. In the case μ is 0, $\beta_k = \beta_k^{DY}$.

Due to $d_{k+1} = -g_{k+1} + \beta_k d_k$ and (2.1),we get

$$\begin{aligned} g_{k+1}^T d_{k+1} &= -\|g_{k+1}\|^2 + \beta_k g_{k+1}^T d_k \\ &= -\|g_{k+1}\|^2 + \frac{\lambda g_{k+1}^T d_{k+1} - \mu \|g_{k+1}\|^2}{g_k^T d_k} g_{k+1}^T d_k \end{aligned} \tag{2.42}$$

The relation simplifies as $g_{k+1}^T d_{k+1} = -\frac{g_k^T d_k + \mu g_{k+1}^T d_k}{g_k^T d_k - \lambda g_{k+1}^T d_k} \|g_{k+1}\|^2$ (2.43)

Thus by the first equality in (2.41) and (2.42),we deduce an equivalent form of β_k i.e.

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

The above form for β_k can be used for practical computations.