5. QUASI-NEWTON METHODS

5.1 Introduction

Since computation of the Hessian matrix or its factors is impractical or costly, the idea underlying quasi-Newton methods is to approximate the Hessian matrix or its inverse using only the gradient and function values. In these methods, the search direction is computed as

\[ d^{(k)} = -A^{(k)} c^{(k)} \]  

(5.1.1)

where \( A^{(k)} \) is an \( n \times n \) approximation to the Hessian inverse. Different quasi-Newton methods correspond to different ways of updating the matrix \( A^{(k)} \) to include the new curvature information obtained during the \( k \)th iteration.

The search direction can also be calculated from the following linear system:

\[ H^{(k)} d^{(k)} = -c^{(k)} \]  

(5.1.2)

where \( H^{(k)} \) is approximation to the Hessian at the \( k \)th iteration. In quasi-Newton methods, approximation to the Hessian or its inverse is generated at each iteration using only the first order information (Gill, Murray and Wright 1981; Luenberger 1984). The approximate Hessian or its inverse is kept symmetric as well as positive definite.

Algorithm 5.1: General Quasi-Newton Method

Step 1. Initialize \( x^{(0)} \) and \( A^{(0)} \) (with any symmetric positive definite matrix). Calculate \( c^{(0)} \) and set \( k = 0 \).
Step 2. Stop, if convergence criteria are satisfied.
Step 3. Calculate \( d^{(k)} = -A^{(k)} c^{(k)} \). Or, solve \( H^{(k)} d^{(k)} = -c^{(k)} \) for \( d^{(k)} \).
Step 4. Calculate a step size \( \alpha_k \).
Step 5. \( x^{(k+1)} = x^{(k)} + \alpha_k d^{(k)} \).
Step 6. Calculate \( c^{(k+1)} \) and stop if convergence criteria are satisfied. Otherwise, calculate \( s^{(k)} = \alpha_k d^{(k)} \) and \( y^{(k)} = c^{(k+1)} - c^{(k)} \).
Step 7. Set \( k = k + 1 \), update \( A^{(k)} \) or \( H^{(k)} \) and go to Step 3.

5.2 Quasi-Newton Condition

To derive the quasi-Newton methods, let

\[ s^{(k)} = x^{(k+1)} - x^{(k)} = \alpha_k d^{(k)} \]  

(change in design)

\[ y^{(k)} = c^{(k+1)} - c^{(k)} \]  

(change in the gradient)
Let us expand the gradient vector at the point $x^{(k+1)}$ about the point $x^{(k)}$ as
\[
c^{(k+1)} = c^{(k)} + s^{(k)} = c^{(k)} + H^{(k)}s^{(k)} + ...
\]
The curvature of $f(x)$ along $s^{(k)}$ is given as $(s^{(k)}, H^{(k)}s^{(k)})$. Using the above expression, this curvature is approximated as
\[
(s^{(k)}, H^{(k)}s^{(k)}) \cong (y^{(k)}, s^{(k)}) \quad (5.2.1)
\]
In the Hessian updating procedure, the initial Hessian approximation $H^{(0)}$ is usually taken as the identity matrix if no other information is available. With this choice, the first iteration of the quasi-Newton method is the steepest descent iteration.

After $x^{(k+1)}$ has been calculated, a new Hessian approximation $H^{(k+1)}$ is obtained by modifying $H^{(k)}$ to account for the newly acquired curvature information. This update formula is of the form
\[
H^{(k+1)} = H^{(k)} + \Delta H^{(k)} \quad (5.2.2)
\]
The standard condition required of the updated Hessian is that it approximate the curvature of $f(x)$ along $s^{(k)}$; i.e., $(s^{(k)}, H^{(k+1)}s^{(k)}) = (s^{(k)}, H^{(k)}s^{(k)})$. Using Eq. (5.2.1) on the left hand side of this equation, it is seen that $H^{(k+1)}$ is required to satisfy the following equation:
\[
H^{(k+1)}s^{(k)} = y^{(k)} \quad (5.2.3)
\]
This is known as the quasi-Newton condition. If inverse Hessian is to be updated, then the quasi-Newton condition becomes
\[
A^{(k+1)}y^{(k)} = s^{(k)} \quad (5.2.4)
\]
These conditions are enforced while deriving the update formulas for the Hessian or its inverse.

The quasi-Newton condition is exact for quadratic functions, but in general, it is only approximate. Consider the quadratic function:
\[
q(x) = \frac{1}{2} (x, Qx) - (b, x) \quad (5.2.5)
\]
where $Q$ is a constant $n \times n$ matrix and $b$ is an $n \times 1$ vector. The quasi-Newton condition can be derived exactly as
\[
c^{(k+1)} = Q x^{(k+1)} - b; \quad c^{(k)} = Q x^{(k)} - b
\]
\[
y^{(k)} = c^{(k+1)} - c^{(k)} = Q x^{(k+1)} - b - Q x^{(k)} + b
\]
\[
= Q(x^{(k+1)} - x^{(k)}) = Q s^{(k)}
\]
Thus for the quadratic function the quasi-Newton condition becomes

\[ y^{(k)} = Q s^{(k)} \]  \hspace{1cm} (5.2.6)

that is satisfied at every iteration. This is an important condition. It is shown later that if the approximate Hessian satisfies the condition for all \( i \leq k \), then the approximation converges to the matrix \( Q \) at the \( n \)th iteration.

### 5.3 Direct Hessian Updating

Several formulas can be derived for updating of the Hessian. In this section, we will derive rank one and rank two updates.

#### 5.3.1 Rank One Update

This procedure adds a correction matrix of rank one to the current approximation of the Hessian \( H^{(k)} \). To preserve symmetry, let the update be given as

\[ H^{(k+1)} = H^{(k)} + a_k u^{(k)} u^{(k)T} \]  \hspace{1cm} (5.3.1)

where scalar \( a_k \) and vector \( u^{(k)} \) have to be determined using the quasi-Newton condition of Eq. (5.2.3) (the matrix \( u^{(k)} u^{(k)T} \) is of rank one since its columns and rows are scalar multiples of the vector \( u^{(k)} \)):

\[ y^{(k)} = H^{(k)} s^{(k)} \]  \hspace{1cm} (5.3.2)

Or,

\[ a_k u^{(k)} (u^{(k)}, s^{(k)}) = (y^{(k)} - H^{(k)} s^{(k)}) \equiv e^{(k)} \]  \hspace{1cm} (5.3.3)

which shows that \( u^{(k)} \) is in the direction of \( e^{(k)} \). Substituting \( u^{(k)} \) from Eq. (5.3.3) into Eq. (5.3.1), we get

\[ H^{(k+1)} = H^{(k)} + e^{(k)} e^{(k)T} a_k (u^{(k)}, s^{(k)}) \]  \hspace{1cm} (5.3.4)

Taking scalar product of both sides of Eq. (5.3.3) with \( s^{(k)} \),

\[ a_k (u^{(k)}, s^{(k)})^2 = (s^{(k)}, e^{(k)}) \]  \hspace{1cm} (5.3.5)

Substituting Eq. (5.3.5) into Eq. (5.3.4), we get
This is a symmetric rank one update formula. Note that there is no other symmetric rank one update formula.

The update in Eq. (5.3.6) has been derived to satisfy the quasi-Newton condition in Eq. (5.2.3). The following theorem shows that in case the Hessian is a constant matrix $Q$, as for problem (5.2.5), then the condition is satisfied for any $i < k$ and at the $n$th iteration $H^{(n)} = Q$.

**Theorem.** Let $Q$ be a fixed symmetric matrix and suppose that $s^{(i)}$, $i = 0$ to $k$ are given vectors. Define the vectors $y^{(i)} = Qs^{(i)}$, $i = 0$ to $k$. Starting with any initial symmetric matrix $H^{(0)}$, let $H^{(k+1)}$ be updated as in Eq. (5.3.6). Then

$$H^{(k+1)}s^{(i)} = y^{(i)}, \text{ for all } i \leq k \ (5.3.7)$$

**Proof.** The proof is by induction. Suppose that Eq. (5.3.7) holds for $H^{(k)}$ and $i \leq (k - 1)$. The relation is true for $H^{(k+1)}$ and $i = k$, as enforced in Eq. (5.3.2). For $i < k$, use of Eq. (5.3.6) gives

$$H^{(k+1)}s^{(i)} = H^{(k)}s^{(i)} + (e^{(k)}, s^{(i)})p^{(k)} \quad \text{with} \quad p^{(k)} = \frac{e^{(k)}}{(s^{(k)}, e^{(k)})} \quad (5.3.8)$$

By the induction hypothesis and substitution for $e^{(k)}$ from Eq. (5.3.3), Eq. (5.3.8) becomes

$$H^{(k+1)}s^{(i)} = y^{(i)} + [(y^{(k)}, s^{(i)}) - (s^{(k)}, H^{(k)}s^{(i)})]p^{(k)}$$

$$= y^{(i)} + [(y^{(k)}, s^{(i)}) - (s^{(k)}, y^{(i)})]p^{(k)} \quad (5.3.9)$$

But

$$(s^{(k)}, y^{(i)}) = (s^{(k)}, Qs^{(i)}) = (s^{(i)}, y^{(k)})$$

Thus, Eq. (5.3.9) reduces to

$$H^{(k+1)}s^{(i)} = y^{(i)} \quad \text{for all } i < k$$

and Eq. (5.3.7) holds for all $i \leq k$.

If Eq. (5.3.7) holds for $n$ linearly independent vectors $s^{(i)}$, $i = 0$ to $(n-1)$, a quasi-Newton method will terminate in finite number of iterations for a quadratic function with constant Hessian $Q$. Let $S$ be matrix whose ith column is the vector $s^{(i)}$. Then Eq. (5.3.7) becomes

$$H^{(n)}S = QS$$
since $y^{(i)} = Q s^{(i)}$. Therefore, since $S$ is nonsingular ($s^{(i)}$ are linearly independent vectors), $H^{(n)} = Q$.

There can be certain numerical difficulties with the rank one update formula. First, the formula preserves positive definiteness only if $(s^{(k)}, e^{(k)}) > 0$ which cannot be guaranteed. Also, even if it is positive, it may have a small value which leads to numerical difficulties. Thus, although it is an excellent example of how the information generated during the iterative process can be used to approximate the Hessian, the rank one method can be numerically unstable. Thus, it is rarely used in practice.

### 5.3.2 Rank Two Updates

Note that if a vector $w^{(k)}$ is orthogonal to $s^{(k)}$, any rank one matrix $z w^T$ annihilates $s^{(k)}$. The quasi-Newton condition will continue to hold if we add further rank one modifications of the form $z w^T$ to $H^{(k+1)}$ in Eq. (5.3.1). This suggests that there can be many rank two updates for the Hessian or its inverse. The most popular update has been derived by Broyden-Fletcher-Goldfarb-Shanno, abbreviated as BFGS.

**BFGS Direct Update: First Form**

The symmetric rank two update is assumed as

$$H^{(k+1)} = H^{(k)} + \beta y^{(k)} y^{(k)T} + \gamma w^{(k)} w^{(k)T}$$  \hspace{1cm} (5.3.7)

where $w^{(k)} = H^{(k)} s^{(k)}$, and constants $\beta$ and $\gamma$ need to be determined. To enforce the quasi-Newton condition of Eq. (5.2.3), we proceed as follows:

$$H^{(k+1)} s^{(k)} = H^{(k)} s^{(k)} + \beta y^{(k)} (y^{(k)}, s^{(k)}) + \gamma w^{(k)} (w^{(k)}, s^{(k)})$$

$$= w^{(k)} + \beta y^{(k)} a_k + \gamma w^{(k)} b_k$$  \hspace{1cm} (5.3.8)

where

$$a_k = (y^{(k)}, s^{(k)}), \quad b_k = (w^{(k)}, s^{(k)})$$  \hspace{1cm} (5.3.9)

If we choose $\beta = 1/a_k$ and $\gamma = -1/b_k$, then the quasi-Newton condition will be satisfied in Eq. (5.3.8). Therefore, the BFGS update formula for updating the Hessian is given from Eq. (5.3.7) as

$$H^{(k+1)} = H^{(k)} + \frac{y^{(k)} y^{(k)T}}{(y^{(k)}, s^{(k)})} - H^{(k)} s^{(k)} s^{(k)T} H^{(k)}$$  \hspace{1cm} (5.3.10)

**BFGS Direct Update: Second Form**

The BFGS formula can be transcribed into several equivalent forms. For example, it can be written as
\[
H^{(k+1)} = H^{(k)} + \frac{y^{(k)} y^{(k)T}}{(y^{(k)}, s^{(k)})} + \frac{c^{(k)} c^{(k)T}}{(d^{(k)}, c^{(k)})}
\]

(5.3.11)

To show this, we proceed as follows:

\[
w^{(k)} = H^{(k)} s^{(k)} = H^{(k)} (\alpha_k d^{(k)}) = -\alpha_k c^{(k)}, \text{ since } H^{(k)} d^{(k)} = -c^{(k)}
\]

\[
(s^{(k)}, H^{(k)} s^{(k)}) = \alpha_k^2 (d^{(k)}, H^{(k)} d^{(k)}) = -\alpha_k^2 (d^{(k)}, c^{(k)})
\]

Therefore,

\[
\frac{H^{(k)} s^{(k)} s^{(k)T}}{(s^{(k)}, H^{(k)} s^{(k)})} = -\frac{c^{(k)} c^{(k)T}}{(d^{(k)}, c^{(k)})}
\]

(5.3.11)

### 5.3.3 Positive Definiteness of Rank Two Updates

If \(H^{(k)}\) is positive definite, then BFGS update maintains positive definiteness of \(H^{(k+1)}\).

For any \(x \in \mathbb{R}^n\), Eq. (5.3.10) gives (assuming a precise step size, i.e., \((s^{(k)}, c^{(k+1)}) = 0\))

\[
(x, H^{(k+1)} x) = (x, H^{(k)} x) + (x, y^{(k)})^2 / a_k - (x, H^{(k)} s^{(k)})^2 / b_k
\]

\[
a_k = (s^{(k)}, y^{(k)}) = (s^{(k)}, (c^{(k+1)} - c^{(k)})) = -(s^{(k)}, c^{(k)})
\]

\[
= (s^{(k)}, H^{(k)} d^{(k)}) = \frac{1}{\alpha_k} (s^{(k)}, H^{(k)} s^{(k)}) = b_k / \alpha_k > 0
\]

(5.3.12)

Let \(H^{(k)} = H^{(k)1/2} H^{(k)1/2}, p = H^{(k)1/2} x, q = H^{(k)1/2} s\)

Therefore,

\[
(x, H^{(k+1)} x) = (p, p) + \alpha_k (x, y^{(k)})^2 / b_k - \frac{(p, q)^2}{(q, q)}
\]

\[
= \frac{(p, p)(q, q) - (p, q)^2}{(q, q)} + \frac{\alpha_k (x, y^{(k)})^2}{b_k}
\]

(5.3.13)

Both the terms in Eq. (5.3.13) are nonnegative - the first by Cauchy-Schwarz inequality. We need to show that both do not vanish simultaneously. The first term vanishes only if \(p\) and \(q\) are proportional. This implies that \(x\) and \(s^{(k)}\) are proportional, e.g., \(x = \theta s^{(k)}\). In this case,

\[
(x, y^{(k)}) = \theta (s^{(k)}, y^{(k)})
\]

which is strictly positive as shown in Eq. (5.3.12). Thus \((x, H^{(k+1)} x) > 0\) for all \(x \neq 0\).
Note that in the above proof, it has been assumed that an exact line search is performed, so that \((s^{(k)}, c^{(k+1)}) = 0\) in Eq. (5.3.12). Actually, we need not make this assumption; any \(\alpha_k\) can be used as long as \(a_k = (s^{(k)}, y^{(k)}) > 0\).

### 5.3.4 Cholesky Factors Updating

It turns out that Cholesky factors of the Hessian can be updated directly. Let

\[ H^{(k)} = L^{(k)} D^{(k)} L^{(k)T} \]

\[ H^{(k+1)} = L^{(k+1)} D^{(k+1)} L^{(k+1)T} \]

Knowing \(L^{(k)}, D^{(k)}, y^{(k)}\) and \(s^{(k)}\), \(L^{(k+1)}\) and \(D^{(k+1)}\) can be evaluated (Gill, Murray and Wright 1981). There are several computational advantages of updating Cholesky factors:

1. Condition number of \(H^{(k+1)}\) can be estimated as

\[ \text{Cond}(H^{(k+1)}) \geq \frac{d_{i\text{max}}}{d_{i\text{min}}} \]

where \(d_{ii}\) are the diagonal elements of the matrix \(D^{(k+1)}\)

2. Updating can be numerically guaranteed to remain positive definite. \(D^{(k+1)}\) and \(L^{(k+1)}\) can be modified so that all \(d_{ii}\) are positive.

3. When \(H^{(k)}\) is ill-conditioned, the direction \(d^{(k)}\) is likely to have no correct digits. In such a situation, the procedure for computing and updating Cholesky factors can be modified so that the condition number of the Hessian approximation does not exceed a fixed upper bound.

4. When the minimization is completed, an estimate of the condition number can be used to give an indication of whether or not the algorithm has converged successfully.

### 5.4 Inverse Hessian Updating

Many times, it is useful to update the inverse Hessian matrix rather than the Hessian itself. If updated inverse \(A^{(k)}\) is available, then the search direction can be obtained simply as \(d^{(k)} = -A^{(k)} c^{(k)}\). The procedures for updating the inverse matrix are the same as for the Hessian itself, except that the inverse form of the quasi-Newton condition given in Eq. (5.2.4) needs to be used. The update formulas for inverse Hessian are completely complementary to the formulas for the Hessian.

#### 5.4.1 Rank One Update

The rank one update for the inverse can be obtained directly from Eq. (5.3.6) by replacing
COMPUTATIONAL METHODS FOR UNCONSTRAINED MINIMIZATION

\( \mathbf{y}^{(k)} \) with \( \mathbf{s}^{(k)} \) and \( \mathbf{s}^{(k)} \) with \( \mathbf{y}^{(k)} \) as

\[
\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \xi^{(k)} \xi^{(k)\mathsf{T}} / (\mathbf{y}^{(k)}, \xi^{(k)}) ; \quad \xi^{(k)} = (\mathbf{s}^{(k)} - \mathbf{A}^{(k)} \mathbf{y}^{(k)})
\] (5.4.1)

### 5.4.2 Rank Two Update

Replacing \( \mathbf{y}^{(k)} \) by \( \mathbf{s}^{(k)} \) and \( \mathbf{s}^{(k)} \) by \( \mathbf{y}^{(k)} \) in Eq. (5.3.10) we get rank-two update formula for inverse Hessian as

\[
\mathbf{A}^{(k+1)} = \mathbf{A}^{(k)} + \frac{\mathbf{s}^{(k)} \mathbf{s}^{(k)\mathsf{T}} / (\mathbf{y}^{(k)}, \mathbf{s}^{(k)}) - \mathbf{A}^{(k)} \mathbf{y}^{(k)} \mathbf{y}^{(k)\mathsf{T}} / (\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)})}{(\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)})}
\] (5.4.2)

This formula was derived by Davidon, and developed further by Fletcher and Powell. It is known as DFP update.

Another rank two update formula can be obtained by writing inverse of the BFGS formula given in Eq. (5.3.10). The following identity, known as Sherman-Morrison formula, can be used for this purpose:

\[
(A + \alpha \mathbf{a} \mathbf{b}^\mathsf{T})^{-1} = A^{-1} - \frac{\alpha A^{-1} \mathbf{a} \mathbf{b} A^{-1}}{1 + \alpha (\mathbf{b}, A^{-1} \mathbf{a})}
\] (5.4.3)

Applying this identity to Eq. (5.3.10), we get

\[
\mathbf{A}_{\text{BFGS}}^{(k+1)} = \mathbf{A}^{(k)} + \left[ 1 + \frac{(\mathbf{y}^{(k)} \mathbf{A}^{(k)\mathsf{T}} / (\mathbf{y}^{(k)}, \mathbf{A}^{(k)})) \mathbf{s}^{(k)} \mathbf{s}^{(k)\mathsf{T}} / (\mathbf{y}^{(k)}, \mathbf{s}^{(k)})}{(\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)})} \right]
- \frac{\mathbf{s}^{(k)} \mathbf{s}^{(k)\mathsf{T}} / (\mathbf{y}^{(k)}, \mathbf{s}^{(k)})}{(\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)})}
\] (5.4.4)

This formula can be used exactly like \( \mathbf{A}_{\text{DFP}}^{(k+1)} \) given in Eq. (5.4.2). Numerical experiments show this formula to have superior performance.

The BFGS update formula for \( \mathbf{A}^{(k+1)} \) can also be written in the product form (Dennis and Schnabel 1983; Nocedal 1980):

\[
\mathbf{A}^{(k+1)} = \mathbf{V}^{(k)} \mathbf{A}^{(k)} \mathbf{V}^{(k)\mathsf{T}} + \rho_k \mathbf{s}^{(k)} \mathbf{s}^{(k)\mathsf{T}}
\] (5.4.5)

where
\[ \rho_k = \frac{1}{a_k} = \frac{1}{(y^{(k)} , s^{(k)})} ; \quad V^{(k)} = I - \rho_k y^{(k)} s^{(k)^T} \] (5.4.6)

It can be shown that if \( A^{(0)} \) is positive definite, the foregoing updates maintain positive definiteness of \( A^{(k)} \) throughout the process.

### 5.5 The Broyden Family

It is noted that both the DFP and the BFGS inverse updates have symmetric rank two correction that are constructed using vectors \( s^{(k)} \) and \( A^{(k)} y^{(k)} \). Weighted combinations of these formulas will also be of this type. This observation leads to a whole collection of updates, known as the Broyden family, defined as

\[ A_\phi = (1 - \phi)A_{\text{DFP}} + \phi A_{\text{BFGS}} \] (5.5.1)

where \( \phi \) may take any real value. \( \phi = 0 \) gives DFP update and \( \phi = 1 \) gives BFGS update. Rank one update is also included in Eq. (5.5.1). Computational and theoretical evidence strongly supports the BFGS update as the best procedure amongst the Broyden class (Fletcher 1980; Shanno and Phua 1978).

Explicit expression for the Broyden family is obtained after considerable algebra as

\[ A^{(k+1)}_\phi = A^{(k)} + \frac{s^{(k)} s^{(k)^T}}{(s^{(k)} , y^{(k)})} - \frac{A^{(k)} y^{(k)^T} A^{(k)}}{(y^{(k)} , A^{(k)} y^{(k)})} + \phi v^{(k)} v^{(k)^T} \]

However, it is difficult to adjust \( \phi \) due to lack of any rational basis. So, usually one \( \phi \) is selected and used throughout. Such a method is called the pure Broyden method. Also for \( \phi < 0 \), it is not possible to ensure positive definiteness of the updates. Therefore, usually \( \phi \geq 0 \) is chosen. If \( \phi \in [0,1] \), the updates defined by the Broyden family are called convex class of formulas.

**Theorem.** If \( f \) is quadratic with positive definite Hessian \( Q \), then for a Broyden method

\[ (s^{(i)} , Q s^{(j)}) = 0, \quad \text{for } 0 \leq i < j \leq k \] (5.5.4)

\[ A^{(k+1)} Q s^{(i)} = s^{(i)}, \quad \text{for } 0 \leq i \leq k \] (5.5.5)

Note that Eq. (5.5.5) is an eigenvalue problem whose first \( k \) eigenvalues are one with \( s^{(i)} \), i
= 1 to k as eigenvectors. This result will be used in Section 9 to precondition the truncated
Newton methods.

In the foregoing, the Broyden family has been defined for inverse Hessian updating. It
can also be defined for the direct Hessian updating as

\[ \mathbf{H}_\phi = (1 - \phi) \mathbf{H}_{\text{DFP}} + \phi \mathbf{H}_{\text{BFGS}} \]  

(5.5.6)

Oren (1974), Oren and Luenberger (1976), and Oren and Spedicato (1976) have
developed a modification to the Broyden class of quasi-Newton methods (Eq. 5.5.2) as

\[ \mathbf{A}^{(k+1)} = \left( \mathbf{A}^{(k)} - \frac{\mathbf{A}^{(k)} \mathbf{y}^{(k)} \mathbf{y}^{(k)T}}{(\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)})} + \phi_k (\mathbf{v}^{(k)} \mathbf{v}^{(k)T}) \right) \gamma + \frac{\mathbf{s}^{(k)} \mathbf{s}^{(k)T}}{(\mathbf{y}^{(k)}, \mathbf{s}^{(k)})} \]  

(5.5.7)

where the new scalar parameter \( \gamma \) is added to make the sequence invariant under multiplication of
the cost function by a constant. Oren and Spedicato then minimize the condition number of
\( \mathbf{A}^{-1} \mathbf{A}^{(k+1)} \), and derive the relationship

\[ \phi_k = \frac{a_k (r_k - a_k \gamma)}{\gamma (c_k r_k - a_k^2)} \]  

(5.5.8)

where \( a_k = (\mathbf{y}^{(k)}, \mathbf{s}^{(k)}) \), \( c_k = (\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)}) \), and \( r_k = (\mathbf{s}^{(k)} \mathbf{s}^{(k)T}) \).

Shanno and Phua (1978) explored computational efficiency of using \( \gamma \). They found for
nonlinear functions, using \( \gamma \) at each step was harmful, as it introduced truncation and
approximation errors in the estimate of the inverse Hessian. However, use of \( \gamma \) at the initial step
was critical, especially for larger problems, to eliminate the truncation errors that resulted from
using the identity matrix as an initial approximation to the inverse Hessian. Substituting \( \phi = 1 \) in
Eq. (5.5.8) yields

\[ \gamma = \frac{(\mathbf{y}^{(k)}, \mathbf{s}^{(k)})}{(\mathbf{y}^{(k)}, \mathbf{A}^{(k)} \mathbf{y}^{(k)})} \]  

(5.5.9)

It was observed that initial scaling of \( \mathbf{A}^{(0)} = \mathbf{I} \) by this factor greatly increased computational
stability. In the next section, Eq. (6.3.18) is a scaled conjugate gradient iterations proposed by
Shanno (1978) that uses \( \gamma \) given in Eq. (5.5.9). Also, in Section 8, it will be seen how \( \gamma \) is
considered to scale the limited memory quasi-Newton methods.

### 5.6 Summary of Quasi-Newton Methods

#### 5.6.1 Quasi-Newton Condition

Define

\[ \mathbf{s}^{(k)} = \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} = \alpha_k \mathbf{d}^{(k)} \]  

\[ \mathbf{y}^{(k)} = \mathbf{c}^{(k+1)} - \mathbf{c}^{(k)} \]  

(1)  

(2)
Quasi-Newton condition is \[ \begin{align*}
H^{(k+1)} s^{(k)} &= y^{(k)} \quad \text{(direct update)} \\
A^{(k+1)} y^{(k)} &= s^{(k)} \quad \text{(inverse update)}
\end{align*} \]

5.6.2 Rank One Update

**Hessian:** \[ H^{(k+1)} = H^{(k)} + \frac{1}{(e^{(k)}, s^{(k)})} e^{(k)} e^{(k)^T} \]

\[ e^{(k)} = y^{(k)} - w^{(k)}; \quad w^{(k)} = H^{(k)} s^{(k)} \]

Guaranteed to be positive definite if \((e^{(k)}, s^{(k)}) > 0\).

**Inverse:** \[ A^{(k+1)} = A^{(k)} + \frac{1}{(y^{(k)}, \xi^{(k)})} \xi^{(k)} \xi^{(k)^T} \]

\[ \xi^{(k)} = s^{(k)} - z^{(k)}; \quad z^{(k)} = A^{(k)} y^{(k)} \]

5.6.3 Rank Two Updates

**DFP:** developed for the inverse Hessian

\[ A_{DFP}^{(k+1)} = A^{(k)} + \frac{s^{(k)} s^{(k)^T}}{a_k} - \frac{z^{(k)} z^{(k)^T}}{c_k} \]

\[ a_k = (s^{(k)}, y^{(k)}) \quad \text{and} \quad c_k = (y^{(k)}, A^{(k)} y^{(k)}) \]

**BFGS:** developed for the Hessian (complementary to DFP in Eq. 8)

\[ H_{BFGS}^{(k+1)} = H^{(k)} + \frac{y^{(k)} y^{(k)^T}}{a_k} - \frac{w^{(k)} w^{(k)^T}}{b_k} \]

\[ b_k = (s^{(k)}, H^{(k)} s^{(k)}) \]

Guaranteed to be positive definite if \((s^{(k)}, y^{(k)}) > 0\)

**BFGS:** inverse Hessian updated obtained by using the Sherman-Morrison formula:

\[ (A + \alpha a b^T)^{-1} = A^{-1} - \frac{\alpha A^{-1} a b A^{-1}}{1 + \alpha (b, A^{-1} a)} \]
\[
A_{\text{BFGS}}^{(k+1)} = A^{(k)} + \left(1 + \frac{c_k}{a_k}\right) \frac{s^{(k)} T s^{(k)}}{a_k} - \frac{s^{(k)} T z + z^{(k)} T s^{(k)}}{a_k}
\]

**DFP:** direct Hessian update (complementary to BFGS in Eq. 13)

\[
H_{\text{DFP}}^{(k+1)} = H^{(k)} + \left(1 + \frac{b_k}{a_k}\right) \frac{y^{(k)} y^{(k) T}}{a_k} - \frac{y^{(k)} w + w^{(k)} y^{(k) T}}{a_k}
\]

**BFGS:** product form of the inverse Hessian update

\[
A^{(k+1)} = V^{(k)} T A^{(k)} V^{(k)} + \rho_k s^{(k)} s^{(k) T}
\]

\[
\rho_k = \frac{1}{a_k} = \frac{1}{(y^{(k)}, s^{(k)})} ; \quad V^{(k)} = I - \rho_k y^{(k)} (k) T
\]