We have already met two symmetric rank-two updating strategies for the approximate inverse Hessian \( B_k \) in the quasi-Newton algorithm:

1. The DFP method:  
\[
B_{k+1} = B_k + \frac{\delta_k \delta_k^T}{\delta_k^T \gamma_k} - \frac{B_k \gamma_k \gamma_k^T B_k}{\gamma_k^T B_k \gamma_k} 
\]

2. The BFGS method:  
\[
B_{k+1} = B_k - \frac{B_k \gamma_k \gamma_k^T B_k}{\gamma_k^T B_k \gamma_k} + \left( 1 + \frac{\gamma_k^T B_k \gamma_k}{\gamma_k^T B_k \gamma_k} \right) \frac{\delta_k \delta_k^T}{\gamma_k^T \delta_k} 
\]

Recall that we arrived at the DFP method by making two rank one updates with the vectors:  
\( \delta_k \) and  
\( B_k \gamma_k \). Let us examine now how much freedom we have in constructing symmetric rank-two updating methods, in terms of the vectors  
\( \delta_k \) and  
\( B_k \gamma_k \), that satisfy the secant condition.

The most general symmetric rank-two correction formula in terms of the given vectors can be written:

\[
B_{k+1} = B_k + [\delta_k \ B_k \gamma_k] \begin{pmatrix} a & b & c \\ b & c & \gamma_k^T B_k \end{pmatrix} \begin{pmatrix} \delta_k \\ \gamma_k^T B_k \end{pmatrix} 
= B_k + a \delta_k^T \delta_k + b (B_k \gamma_k \delta_k^T + \delta_k \gamma_k^T B_k) + c B_k \gamma_k \gamma_k^T B_k 
\]

Now, if we impose the secant condition:

\[
B_{k+1} \gamma_k = \delta_k 
\]

we obtain (with some rearranging):

\[
B_k \gamma_k + (a \delta_k^T \delta_k + b \gamma_k^T B_k \gamma_k) \gamma_k + (b \delta_k^T \gamma_k + c \gamma_k^T B_k \gamma_k) B_k \gamma_k = \delta_k. 
\]

Assuming that  
\( \gamma_k \) and  
\( B_k \gamma_k \) are linearly independent then we need  
\( a, b, c \) to satisfy:

\[
\begin{align*}
  a \delta_k^T \delta_k + b \gamma_k^T B_k \gamma_k &= 1 \\
b \delta_k^T \gamma_k + c \gamma_k^T B_k \gamma_k &= -1
\end{align*} 
\]

The key point is that this is only two equations for determining three unknowns. This leaves one degree of freedom that we are free to choose and this leads to a whole family of possible methods.

Now, after some more rearranging (messy) equation (1) can be rewritten as:

\[
B_{k+1} = B_k + \left( \frac{a \delta_k^T \delta_k + b \gamma_k^T B_k \gamma_k}{\delta_k^T \gamma_k} \right) \delta_k^T \delta_k + \left( \frac{b \delta_k^T \gamma_k + c \gamma_k^T B_k \gamma_k}{\gamma_k^T B_k \gamma_k} \right) B_k \gamma_k \gamma_k^T B_k 

- b \delta_k^T \gamma_k \gamma_k^T B_k \gamma_k \left( \frac{\delta_k}{\delta_k^T \gamma_k} - \frac{B_k \gamma_k}{\gamma_k^T B_k \gamma_k} \right) \left( \frac{\delta_k}{\delta_k^T \gamma_k} - \frac{B_k \gamma_k}{\gamma_k^T B_k \gamma_k} \right)^T. 
\]

Substituting in conditions (3) and (4) gives:

\[
B_{k+1} = B_k + \frac{\delta_k \delta_k^T}{\delta_k^T \gamma_k} - \frac{B_k \gamma_k \gamma_k^T B_k}{\gamma_k^T B_k \gamma_k} + \phi_k v_k v_k^T 
\]

where  
\( \phi_k \) is a scalar parameter of the form:

\[
\phi_k = b \delta_k^T \gamma_k 
\]

and  
\( v_k \) is the vector:

\[
v_k = \left( \gamma_k^T B_k \gamma_k \right)^{\frac{1}{2}} \left( \frac{\delta_k}{\delta_k^T \gamma_k} - \frac{B_k \gamma_k}{\gamma_k^T B_k \gamma_k} \right). 
\]

The updating scheme (\( \ast \)) is known as the Broyden family of Quasi-Newton methods. It is a one-parameter family of methods, where we are free to choose  \( \phi_k \).
Note the following points:

- Both the DFP and BFGS methods are just special cases of the Broyden family. \( \phi_k = 0 \Rightarrow \text{DFP method} \) and \( \phi_k = 1 \Rightarrow \text{BFGS method} \).
- The Broyden update is equivalent to the DFP update plus the rank-one correction \( \phi_k v_k v_k^T \).
- Since the DFP method satisfies the secant condition, it is easy to check that the Broyden methods (**) do indeed satisfy \( B_{k+1} \gamma_k = \delta_k \).
- Since the DFP update produces a positive definite matrix (see proof in lecture notes) under the assumption that \( \gamma_k^T \delta_k > 0 \), it is also easy to check that the Broyden update also produces a positive definite matrix provided that \( 0 \leq \phi_k \leq 1 \). (exercise)

**Convergence results**

Since the DFP method and the BFGS method are part of the larger Broyden family of methods, we can analyse their convergence properties collectively. First, we consider their performance in minimising quadratic objective functions (for which performing exact line searches is trivial).

### Theorem 1 (Minimising quadratic functions)

Consider minimising the quadratic objective function:

\[
Q(x) = \frac{1}{2} x^T G x + b^T x + c,
\]

where \( G \in \mathbb{R}^{n \times n} \) is a given symmetric and positive definite matrix, \( b \in \mathbb{R}^n \) and \( c \in \mathbb{R} \), via the Quasi-Newton method with exact line searches. If the approximate inverse Hessian is updated at each iteration by an arbitrary member of the Broyden family with \( \phi_k \geq 0 \) then, at the start of the \( i \)th quasi-Newton step,

\[
B_i \gamma_j = \delta_j, \quad \text{for all } j \text{ satisfying } j < i \quad (\dagger)
\]

\[
\gamma_j^T \delta_j = 0 \quad \text{for all } k, j, \text{ satisfying } k < j < i \quad (\ddagger)
\]

**Proof:** Exercise. See question 4 on problem sheet 3 and follow the hints.

Now, the trick is to see that condition (\dagger) implies that the search directions obtained in the Quasi-Newton iteration are conjugate with respect to \( G \). Since we are using exact line searches, we know from a previous Theorem that the resulting method must have quadratic termination (ie converge in at most \( n \) steps). If the method had not terminated after \( n \) iterations, then at the \( (n+1) \)st iteration, with \( i = n + 1 \) in (\dagger), we obtain,

\[
B_{n+1} \gamma_j = \delta_j, \quad \text{for all } j = 1 : n
\]

For any \( j \) a simple calculation gives that:

\[
\gamma_j = G \delta_j.
\]

so that

\[
B_{n+1} \gamma_j = G^{-1} \gamma_j, \quad \text{for all } j = 1 : n.
\]

Since the set of \( n \) vectors \( \delta_j, j = 1 : n \), are conjugate, they are linearly independent and so \( B_{n+1} = G^{-1} \). That is, the \( (n+1) \)st approximation to the inverse Hessian is exact. (\( G \) is the exact Hessian of the given quadratic objective function).

### Theorem 2 (Minimising quadratic functions)

Under the conditions of Theorem 1, and starting from the same initial guess \( x_0 \), the same sequence of iterates \( \{x_i\} \) is generated, independent of the choice of \( \phi_k \geq 0 \).

**Proof:** Omitted. Similar to proof of Theorem 3 below. The idea is to determine that the search vectors \( \{p_i\} \) differ only in magnitude and not in direction, for each of the different updating strategies from the Broyden class.
Next, we consider the performance of the Broyden quasi-Newton methods in minimising general nonlinear objective functions (more interesting). If exact line searches are possible then we have the same result as in the quadratic case.

Theorem 3 (Shanno-Kettler Theorem)

Consider minimising a nonlinear objective function \( f(x) \) via the Quasi-Newton method with exact line searches. If the approximate inverse Hessian is updated at each iteration by an arbitrary member of the Broyden family with \( \phi_k \geq 0 \) then the search direction \( p_{k+1} \) is proportional to \( v_k \), assuming that a certain degenerate value of \( \phi_k \) is not used.

Proof: In the quasi-Newton method, at the \((k+1)\)st step, the search direction is calculated via

\[
p_{k+1} = -B_k g_{k+1} - \frac{\delta_k \gamma_k g_k}{\delta_k} + \frac{B_k \gamma_k g_k}{\gamma_k} \gamma_k g_{k+1} \]

With exact line-searches, we obtain \( \delta_k g_{k+1} = 0 \) (see exercise sheet one) and we also have:

\[
B_k g_{k+1} = B_k (g_k + g_k - g_k) = B_k g_k + B_k \gamma_k.
\]

Therefore, (*) becomes:

\[
p_{k+1} = -B_k g_k - B_k \gamma_k + \frac{B_k \gamma_k \gamma_k}{\gamma_k} (B_k g_k + B_k \gamma_k) + \phi_k v_k v_k^T g_{k+1}
\]

Now, we can use: \( p_k = -B_k g_k \) and \( \delta_k = x_{k+1} - x_k = \alpha_k p_k \) to give:

\[
p_{k+1} = \frac{1}{\alpha_k} \delta_k - B_k \gamma_k - \frac{B_k \gamma_k \gamma_k}{\alpha_k \gamma_k^2} B_k \gamma_k \gamma_k + \phi_k v_k v_k^T g_{k+1} = \frac{1}{\alpha_k} \left( \delta_k - \frac{B_k \gamma_k \gamma_k}{\gamma_k} \delta_k \right) + \phi_k v_k v_k^T g_{k+1}
\]

and applying the definition of \( v_k \) gives:

\[
p_{k+1} = \frac{\delta_k \gamma_k v_k}{\alpha_k (\gamma_k^T B_k \gamma_k)^{\frac{3}{2}}} + \phi_k v_k v_k^T g_{k+1}
\]

Using, once again, the relation: \( g_{k+1} = g_k + \gamma_k \) gives:

\[
v_k^T g_{k+1} = v_k^T g_k + v_k^T \gamma_k = v_k^T g_k + \left( \gamma_k^T B_k \gamma_k \right)^{\frac{1}{2}} \left( \frac{\delta_k \gamma_k - \gamma_k^T B_k \gamma_k}{\gamma_k \gamma_k} \right) = v_k^T g_k
\]

Substituting this into (**) gives:

\[
p_{k+1} = v_k \left( \frac{\delta_k \gamma_k}{\alpha_k (\gamma_k^T B_k \gamma_k)^{\frac{3}{2}}} + \phi_k v_k^T g_k \right).
\]

Hence, the search direction \( p_{k+1} \) is just \( v_k \) multiplied by a constant. The constant depends on the choice of \( \phi_k \) but the vector \( v_k \) does not. Hence the theorem holds, provided that \( \phi_k \) satisfies:

\[
\left( \frac{\delta_k \gamma_k}{\alpha_k (\gamma_k^T B_k \gamma_k)^{\frac{3}{2}}} + \phi_k v_k^T g_k \right) \neq 0.
\]

Hence, if exact line searches are used, the direction of the search vectors is independent of the choice of \( \phi_k \). This leads immediately to the following result.

Theorem 4 (Minimising general nonlinear functions)

Under the conditions of Theorem 3, and starting from the same initial guess \( x_0 \), all the quasi-Newton methods from the Broyden family produce the same sequence of iterates \( \{x_i\} \).

Proof: The result follows immediately from Theorem 3.