Consider the problem,

$$\min_x f(x) \quad \text{where} \quad f : \mathbb{R}^n \to \mathbb{R}$$

To solve above problem in Newton method f should be C^2 contionous

1 Newton Method :

The quadratic approximation of $f(x)$ about x_k is

$$m_k(x) \approx f(x)\Big|_{x_k} = f(x_k) + {g_k}^T(x - x_k) + \frac{1}{2}(x - x_k)^T H_k(x - x_k)$$

where g_k and H_k are gradient and hessian at x_k respectively. Newton descent direction is : $p_k = -{H_k}^{-1}g^k$

- The operation count for calculating inverse of matrix is $O(n^3)$ which is expensive.
- we can implement newton's method only if there exists positive definite Hessian.

To overcome these quasi newton methods are introduced.

2 Quasi Newton Methods :

The basic idea of quasi newton methods is to approximate Hessian(H_k) or inverse of Hessian(${H_k}^{-1}$) by some symmetric positive definite matrix, which brings down our requirement of f to C^1 continous.

Lets B_k be approximaton of Hessian,then the quadratic approximation of f about x_k is

$$m_k(x) \approx f(x)\Big|_{x_k} = f(x_k) + g_k^T(x - x_k) + \frac{1}{2}(x - x_k)^T B_k(x - x_k)$$

where g_k and B_k are gradient and approximate Hessian at x_k respectively.
Quasi Newton descent direction is : $p_k = -{B_k}^{-1}g^k$ and

$$x_{k+1} = x_k + \alpha_k p_k \qquad \text{such that } \alpha_k \text{ should satisfy wolfe conditions.}$$

But the problem here is, how to update approximate Hessian(B_{k+1}) preserving its symmetric positive definiteness property for next iteration ?

Based on how we approximate the Hessian(B_{k+1}) we have different methods. The basic idea in approximating Hessian(B_{k+1}) is to take into account the curvature measured during the most recent step and previous step.

$$m_{k+1}(x) \approx f(x)\Big|_{x_{k+1}} = f(x_{k+1}) + g_{k+1}^T(x - x_{k+1}) + \frac{1}{2}(x - x_{k+1})^T B_{k+1}(x - x_{k+1})$$

$$\nabla m_{k+1}(x) = g_{k+1} + B_{k+1}(x - x_{k+1})$$

from above,

$$\begin{aligned}\nabla m_{k+1}(x_k) &= g_{k+1} + B_{k+1}(x_k - x_{k+1}) \\ &= \nabla f(x_{k+1}) + B_{k+1}(x_k - x_{k+1})\end{aligned} \tag{1}$$

we need $\nabla m_{k+1}(x_k) = \nabla f(x_k)$, on substituting in equation 1

$$\begin{aligned}\nabla f(x_k) &= \nabla f(x_{k+1}) + B_{k+1}(x_k - x_{k+1}) \\ \nabla f(x_k) - \nabla f(x_{k+1}) &= B_{k+1}(x_k - x_{k+1})\end{aligned} \tag{2}$$

Let, $\nabla f(x_k) - \nabla f(x_{k+1}) = y_k$ and $x_k - x_{k+1} = s_k$ then equation 2 becomes

$$B_{k+1}s_k = y_k$$

This is known as Secant Equation.

As we need B_{k+1} to be positive definite,

$$s_k^T B_{k+1} s_k = s_k^T y_k > 0$$

So we need to enforce this condition for a general non linear problem.

How to update B_{k+1} ?

- we have n equalities form secant equation.
- As B_{k+1} to be positive definite, all principal minors should be positive. so we have n inequalities.

As B_{k+1} should be symmetric we have $\frac{n(n+1)}{2}$ unknowns. So, this is under-constrained problem.So we can get multiple solutions for B_{k+1}.

2.1 SR1 Method:

This is symmetric rank 1 update

$$B_{k+1} = B_k + \sigma v v^T \tag{3}$$

Where σ and v are chosen such that B_{k+1} satisfies the secant equation.

$$\begin{aligned}B_{k+1}s_k &= y_k \\ (B_k + \sigma v v^T)s_k &= y_k \\ \sigma v v^T s_k &= y_k - B_k s_k \\ (\sigma v^T s_k)v &= y_k - B_k s_k\end{aligned}$$

$$\text{so,} \qquad \sigma v^T s_k = 1 \qquad \text{and} \qquad v = y_k - B_k s_k$$

$$\sigma = \frac{1}{(y_k - B_k s_k)^T s_k}$$

Substituting in equation 3, we get

$$B_k + 1 = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{(y_k - B_k s_k)^T s_k}$$

This is known as SR1 update formula.

If we want to approximate inverse of hessian which is nothing but B_{k+1}^{-1}, we can similarly derive or using sherman morrison formula, we get

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k)(s_k - B_k^{-1} y_k)^T}{(s_k - B_k^{-1} y_k)^T y_k}$$

- Rank 1 updates tend to produce better approximations of hessain
- numerical difficulties arise if $(y_k - B_k s_k)^T s_k = 0$, then SR1 update is not possible which led to SR2 updates.

2.2 SR2 Method:

This is symmetric rank 2 update.

2.2.1 BFGS Method:

In this method we update Hessain B_{k+1}

$$B_{k+1} = B_k + \sigma_1 u u^T + \sigma_2 v v^T \tag{4}$$

Where σ_1, σ_2, u and v are chosen such that B_{k+1} satisfies the secant equation.

$$\begin{aligned} B_{k+1} s_k &= y_k \\ (B_k + \sigma_1 u u^T + \sigma_2 v v^T) s_k &= y_k \\ \sigma_1 u u^T s_k + \sigma_2 v v^T s_k &= y_k - B_k s_k \\ (\sigma_1 u^T s_k) u + (\sigma_2 v^T s_k) v &= y_k - B_k s_k \end{aligned}$$

So,

$$\begin{aligned} \sigma_1 u^T s_k = 1 \qquad & u = y_k \to \sigma_1 = \frac{1}{y_k^T s_k} \\ \sigma_2 v^T s_k = -1 \qquad & v = B_k s_k \to \sigma_2 = \frac{-1}{s_k^T B_k s_k} \end{aligned}$$

On substituing in equation 4, we get

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{5}$$

This is known as **BFGS update formula**.

We can update inverse of hessain directly instead of hessain,

Let $H_k = B_k^{-1}$, then H_{k+1} is

$$H_{k+1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}$$

This is known as **BFGS update formula for inverse of hessain**.

2.2.2 DFP Method:

If we want to approximate inverse of hessian which is nothing but B_{k+1}^{-1} we can derive similarly.

$$B_{k+1} s_k = y_k$$
$$s_k = B_{k+1}^{-1} y_k$$

Let $H_{k+1} = B_{k+1}^{-1}$

$$s_k = H_{k+1} y_k \tag{6}$$

The symmetric rank 2 update of H_{k+1} is

$$H_{k+1} = H_k + \sigma_1 u u^T + \sigma_2 v v^T \tag{7}$$

Where σ_1, σ_2, u and v are chosen such that H_{k+1} satisfies equation 6.

$$\begin{aligned} s_k &= H_{k+1} y_k \\ s_k &= (H_k + \sigma_1 u u^T + \sigma_2 v v^T) y_k \\ s_k - H_k y_k &= \sigma_1 u u^T y_k + \sigma_2 v v^T y_k \\ s_k - H_k y_k &= (\sigma_1 u^T y_k) u + (\sigma_2 v^T y_k) v \end{aligned}$$

So,

$$\sigma_1 u^T y_k = 1 \qquad u = s_k \rightarrow \sigma_1 = \frac{1}{s_k^T y_k}$$
$$\sigma_2 v^T y_k = -1 \qquad v = H_k y_k \rightarrow \sigma_2 = \frac{-1}{y_k^T H_k y_k}$$

On substituing in equation 7, we get

$$H_{k+1} = H_k + \frac{s_k s_k^T}{s_k^T y_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k} \tag{8}$$

This is known as **DFP update formula.**

We can update hessain direclty instead of inverse of hessain,

$$B_{k+1} = \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) B_k \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) + \frac{y_k y_k^T}{y_k^T s_k}$$

This is known as **DFP update formula for hessain.**

2.2.3 Comparison between SR1 and SR2 methods:

- Operation cost has been reduced.Now each iteration can be performed at $O(n^2)$, where in newton method it is $O(n^3)$.
- BFGS Method has self adjusting properties i.e even if at some iteration B_k becomes a poor approximation to true hessain then it will tend to correct on itself within few iterations which is not possible in SR1.
- Till today, nothing better than BFGS is available.

3 Broyden Class/Family:

In Broyden class, a family of updates are specificed by general formula

$$B_{k+1} = (1 - \phi_k)B_{k+1}^{BFGS} + \phi_k B_{k+1}^{DFP}$$

The BFGS and DFP methods are members of the Broyden class-we recover BFGS by setting $\phi_k = 0$ and DFP by setting $\phi_k = 1$.