We use the inverse of the approximated Hessian to get next iteration point with the secant equation as

$$H_k y_k = s_k \tag{0.1}$$

Then using the following optimization setting to get BFGS

$$\begin{aligned} &\underset{H}{\text{minimize}} && ||H - H_k||_W \\ &\text{subject to} && H = H^T \quad Hy_k = s_k \quad Ws_k = y_k \end{aligned}$$

Here $||.||_W$ means Frobenius norm with weights put on each entry. For example, $||H - H_k||_W$ under this weighted norm is $\sum_{i,j}(W_{ij}(H_{ij} - (H_k)_{ij})(W_{ij}(H_{ij} - (H_k)_{ij})$. W is also a positive symmetric definite matrix. Using Lagrange multiplier method, we can turn the optimization problem into finding local minimal of the following equation

$$\begin{aligned} \mathcal{L}(H, \lambda_1, \lambda_2) &= \sum_{i,j}(W_{ij}(H_{ij} - (H_k)_{ij})(W_{ij}(H_{ij} - (H_k)_{ij}) \\ &\quad - \sum_{i<j}(\lambda_1)_{ij}(H_{ij} - (H^T)_{ij}) - \lambda_2^T(Hy_k - s_k) \\ &= \sum_{i,j}(W_{ij}(H_{ij} - (H_k)_{ij})(W_{ij}(H_{ij} - (H_k)_{ij}) \\ &\quad - \sum_{i<j}(\lambda_1)_{ij}(H_{ij} - H_{ji}) - \lambda_2^T(Hy_k - s_k) \end{aligned}$$

here λ_1 is a matrix and λ_2 is a vector.

We can further simplify the above equation by setting $(\lambda_1)_{ij} = -(\lambda_1)_{ji}$, then we get

$$\mathcal{L}(H, \lambda_1, \lambda_2) = \sum_{i,j}(W_{ij}(H_{ij} - (H_k)_{ij})(W_{ij}(H_{ij} - (H_k)_{ij}) - \sum_{ij}(\lambda_1)_{ij}H_{ij} - \lambda_2^T(Hy_k - s_k) \tag{0.2}$$

Now take derivatives on this equation with respect to H, we have

$$\frac{\partial \mathcal{L}(H, \lambda_1, \lambda_2)}{\partial H} = 2W(H - H_k)W^T - \lambda_1 - \lambda_2 y_k^T \tag{0.3}$$

Because $(\lambda_1)_{ij} = -(\lambda_1)_{ji}$ therefore $\lambda_1^T = -\lambda_1$. Also, notice that H_k, H should be symmetric, and if we take a transpose on above equation we will have

$$\frac{\partial \mathcal{L}(H, \lambda_1, \lambda_2)}{\partial H} = 2W(H - H_k)W^T + \lambda_1 - y_k\lambda_2^T \tag{0.4}$$

Set equation (1.19) and equation (1.20) both equal to zero. Then add those two equations we can get

$$0 = 4W(H - H_k)W^T - \lambda_2 y_k^T - y_k\lambda_2^T \tag{0.5}$$

Based on this equation and W is symmetric, we know that

$$H = \frac{1}{4}W^{-1}(\lambda_2 y_k^T + y_k\lambda_2^T)W^{-1} + H_k \tag{0.6}$$

therefore, once we find $\frac{1}{4}W^{-1}(\lambda_2 y_k^T + y_k\lambda_2^T)W^{-1}$ we find H. Since W is symmetric, $Ws_k = y_k$, and $Hy_k = s_k$, we multiply s_k on both side of equation (1.22) we get

$$\begin{aligned}
0 &= 4W(H - H_k)Ws_k - \lambda_2 y_k^T s_k - y_k\lambda_2^T s_k \\
&= 4W(H - H_k)y_k - \lambda_2 y_k^T s_k - y_k\lambda_2^T s_k \\
&= 4WHy_k - 4WH_k y_k - \lambda_2 y_k^T s_k - y_k\lambda_2^T s_k \\
&= 4Ws_k - 4WH_k y_k - \lambda_2 y_k^T s_k - y_k\lambda_2^T s_k \\
&= 4y_k - 4WH_k y_k - \lambda_2 y_k^T s_k - y_k\lambda_2^T s_k
\end{aligned}$$

Notice in the above equation, 0 is a vector. Now we multiply both side by s_k^T (dot product), and we get

$$0 = 4s_k^T y_k - 4s_k^T WH_k y_k - s_k^T\lambda_2 y_k^T s_k - s_k^T y_k\lambda_2^T s_k$$

Since in this equation all terms are just numbers, we can rearrange the order of the product without changing the product values and we can get

$$\begin{aligned}
0 &= 4y_k^T s_k - 4y_k^T H_k y_k - s_k^T\lambda_2 y_k^T s_k - s_k^T\lambda_2 y_k^T s_k \\
&= 4y_k^T s_k - 4y_k^T H_k y_k - 2s_k^T\lambda_2 y_k^T s_k
\end{aligned}$$

Therefore, we get

$$s_k^T\lambda_2 = \lambda_2^T s_k = \frac{2}{y_k^T s_k} y_k^T(s_k - H_k y_k) \tag{0.7}$$

Substitute $\lambda_2^T s_k$ in the equation below equation (1.23), we have

$$0 = 4y_k - 4WH_k y_k - \lambda_2 y_k^T s_k - y_k\frac{2}{y_k^T s_k} y_k^T(s_k - H_k y_k) \tag{0.8}$$

which gives us

$$\lambda_2 = \frac{4}{y_k^T s_k}(y_k - WH_k y_k) - \frac{2}{(y_k^T s_k)^2} y_k y_k^T(s_k - H_k y_k) \tag{0.9}$$

Notice that $y_k^T(s_k - H_k y_k)$ is just a number we can rewrite the above equation as

$$\lambda_2 = \frac{4}{y_k^T s_k}(y_k - WH_k y_k) - \frac{2}{(y_k^T s_k)^2} y_k^T(s_k - H_k y_k)y_k \tag{0.10}$$

Now let's multiply y_k^T to the above equation, and we have

$$\lambda_2 y_k^T = \frac{4}{y_k^T s_k}(y_k - WH_k y_k)y_k^T - \frac{2}{(y_k^T s_k)^2} y_k^T(s_k - H_k y_k)y_k y_k^T \tag{0.11}$$

Apply transpose to the above equation, and we have

$$y_k\lambda_2^T = \frac{4}{y_k^T s_k} y_k(y_k^T - y_k^T H_k W) - \frac{2}{(y_k^T s_k)^2} y_k^T(s_k - H_k y_k)y_k y_k^T \tag{0.12}$$

Thus, $\lambda_2 y_k^T + y_k\lambda_2^T$ is

$$\lambda_2 y_k^T + y_k\lambda_2^T = \frac{4}{y_k^T s_k}(2y_k y_k^T - WH_k y_k y_k^T - y_k y_k^T H_k W) - \frac{4}{(y_k^T s_k)^2} y_k^T(s_k - H_k y_k)y_k y_k^T \tag{0.13}$$

Since we have $Ws_k = y_k$ and W is symmetric, we have $s_k = W^{-1}y_k$ and $s_k^T = y_k^T W^{-1}$. Therefore, $\frac{1}{4}W^{-1}(\lambda_2 y_k^T + y_k\lambda_2^T)W^{-1}$ is

$$\begin{aligned}
\frac{1}{4}W^{-1}(\lambda_2 y_k^T + y_k\lambda_2^T)W^{-1} =& \frac{1}{y_k^T s_k}W^{-1}(2y_k y_k^T - WH_k y_k y_k^T - y_k y_k^T H_k W)W^{-1} \\
&- \frac{1}{(y_k^T s_k)^2}W^{-1}y_k^T(s_k - H_k y_k)y_k y_k^T W^{-1} \\
=& \frac{1}{y_k^T s_k}(2W^{-1}y_k y_k^T W^{-1} - W^{-1}WH_k y_k y_k^T W^{-1} - W^{-1}y_k y_k^T H_k W W^{-1}) \\
&- \frac{1}{(y_k^T s_k)^2}y_k^T(s_k - H_k y_k)W^{-1}y_k y_k^T W^{-1} \\
=& \frac{1}{y_k^T s_k}(2s_k s_k^T - H_k y_k s_k^T - s_k y_k^T H_k) \\
&- \frac{1}{(y_k^T s_k)^2}y_k^T(s_k - H_k y_k)s_k s_k^T \\
=& \frac{2s_k s_k^T}{y_k^T s_k} - \frac{H_k y_k s_k^T}{y_k^T s_k} - \frac{s_k y_k^T H_k}{y_k^T s_k} - \frac{y_k^T s_k s_k s_k^T}{(y_k^T s_k)^2} + \frac{y_k^T H_k y_k s_k s_k^T}{(y_k^T s_k)^2} \\
=& \frac{2s_k s_k^T}{y_k^T s_k} - \frac{H_k y_k s_k^T}{y_k^T s_k} - \frac{s_k y_k^T H_k}{y_k^T s_k} - \frac{s_k s_k^T}{y_k^T s_k} + \frac{y_k^T H_k y_k s_k s_k^T}{(y_k^T s_k)^2} \\
=& \frac{s_k s_k^T}{y_k^T s_k} - \frac{H_k y_k s_k^T}{y_k^T s_k} - \frac{s_k y_k^T H_k}{y_k^T s_k} + \frac{y_k^T H_k y_k s_k s_k^T}{(y_k^T s_k)^2}
\end{aligned}$$

Now we based on equation (1.23) we have

$$\begin{aligned}
H =& \frac{1}{4}W^{-1}(\lambda_2 y_k^T + y_k\lambda_2^T)W^{-1} + H_k \\
=& \frac{s_k s_k^T}{y_k^T s_k} - \frac{H_k y_k s_k^T}{y_k^T s_k} - \frac{s_k y_k^T H_k}{y_k^T s_k} + \frac{y_k^T H_k y_k s_k s_k^T}{(y_k^T s_k)^2} + H_k \\
=& H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k} - \frac{s_k y_k^T H_k}{y_k^T s_k} + \frac{s_k y_k^T H_k y_k s_k^T}{(y_k^T s_k)^2} \\
=& H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) - \frac{s_k y_k^T}{y_k^T s_k}H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k} \\
=& (I - \frac{s_k y_k^T}{y_k^T s_k})H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k}
\end{aligned}$$

Therefore, the solution is given by

$$(BFGS) \quad H_{k+1} = (I - \frac{s_k y_k^T}{y_k^T s_k})H_k(I - \frac{y_k s_k^T}{y_k^T s_k}) + \frac{s_k s_k^T}{y_k^T s_k} \tag{0.14}$$

We can also get the B_k update from this solution by applying the Sherman Morrison Woodbury formula.

$$(BFGS) \quad B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k} \tag{0.15}$$

References

[1] Jorge Nocedal and Stephen J. Wright. Numerical Optimization. Springer. 2006