Homework 7
Shengchao Liu

1. First we have

$$\hat{H}_{k+1} = I - \frac{s \cdot y^T}{y^T \cdot s}$$

$$= \frac{1}{y^T \cdot s} \begin{bmatrix} s_1 y_1 - \sum_i s_i y_i & s_1 y_2 & \cdots & s_1 y_n \\ s_2 y_1 & s_2 y_2 - \sum_i s_i y_i & \cdots & s_2 y_n \\ \vdots & \vdots & \ddots & \vdots \\ s_n y_1 & s_n y_2 & \cdots & s_n y_n - \sum s_i y_i \end{bmatrix}$$

Suppose c_i is the i-th column of $\hat{H}_{k+1}$, so we can get

$$c_1 s_1 + c_2 s_2 + \ldots + c_n s_n = \begin{bmatrix} s_1 s_1 y_1 - s_1 \sum s_i y_i + s_1 s_2 y_2 + \ldots + s_1 s_n y_n \\ s_1 s_2 y_1 + s_2 s_2 y_2 - s_2 \sum s_i y_i + \ldots + s_2 s_n y_n \\ \vdots \\ s_1 s_n y_1 + s_2 s_n y_2 + \ldots + s_n s_n y_n - s_n \sum s_i y_i \end{bmatrix}$$

$$= \begin{bmatrix} s_1 (s_1 y_1 + s_2 y_2 + \ldots + s_n y_n - \sum s_i y_i) \\ s_2 (s_1 y_1 + s_2 y_2 + \ldots + s_n y_n - \sum s_i y_i) \\ \vdots \\ s_n (s_1 y_1 + s_2 y_2 + \ldots + s_n y_n - \sum s_i y_i) \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

So $\hat{H}_{k+1}$ is linearly dependent, so it is singular.

2. $f(x_{k+1}) = f(x_k + \alpha_k p_k)$

   According to exact line search, we get $p_k^T \bigtriangledown f(x_k + \alpha_k p_k) = 0$

   That is $p_k^T \bigtriangledown f(x_{k+1}) = 0$.

   $p_k = -(I - \frac{s_k y_k^t}{y_k^T s_k} - \frac{y_k s_k^T}{y_k^T s_k} + \frac{(y_k^T y_k) s_k s_k^T}{(y_k^T s_k)^2} + \frac{s_k s_k^T}{y_k^T s_k}) \cdot \bigtriangledown f_{k+1}$

   And $s_k = \alpha_k p_k$, so $s_k^T \bigtriangledown f_{k+1} = \alpha_k p_k^T \bigtriangledown f_{k+1} = 0$.

   Plug this into original equation, we get

   $p_k = -(I - \frac{p_k y_k^T}{y_k^T p_k}) \cdot \bigtriangledown f_{k+1} = -\bigtriangledown f_{k+1} + \frac{p_k y_k^T \bigtriangledown f_{k+1}}{y_k^T p_k} = -\bigtriangledown f_{k+1} + \frac{\bigtriangledown f_{k+1}^T y_k}{y_k^T p_k} p_k$

3. Using inductive method, first we assume that after k steps, we have Q_k and $\Sigma_k = D_k + L_k + L_k^T - S_k^T B_0 S_k$.

Then given a new pair s_k, y_k:

$$Q_k = [Y_k, y_k] - B_0[S_k, s_k] = [Q_k, y_k - B_0 s_k]$$

And according to definition in formula 7.25 - 7.27, we get:

$$\Sigma_k = \begin{bmatrix} s_0^T y_0 - s_0^T B_0 s_0 & s_1^T y_0 - s_0^T B_0 s_1 & s_2^T y_0 - s_0^T B_0 s_2 & \cdots & s_{k-1}^T y_0 - s_0^T B_0 s_{k-1} \\ s_1^T y_0 - s_1^T B_0 s_0 & s_1^T y_1 - s_1^T B_0 s_1 & s_2^T y_1 - s_1^T B_0 s_2 & \cdots & s_{k-1}^T y_1 - s_1^T B_0 s_{k-1} \\ s_2^T y_0 - s_2^T B_0 s_0 & s_2^T y_1 - s_2^T B_0 s_1 & s_2^T y_2 - s_2^T B_0 s_2 & \cdots & s_{k-1}^T y_2 - s_2^T B_0 s_{k-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{k-1}^T y_0 - s_{k-1}^T B_0 s_0 & s_{k-1}^T y_1 - s_{k-1} B_0 s_1 & s_{k-1}^T y_2 - s_{k-1} B_0 S_2 & \cdots & s_{k-1}^T y_{k-1} - s_{k-1}^T B_0 s_{k-1} \end{bmatrix}$$

SO when we expand Σ_k to Σ_{k+1}, we get:

$$\Sigma_{k+1} = \begin{bmatrix} s_0^T y_0 - s_0^T B_0 s_0 & s_1^T y_0 - s_0^T B_0 s_1 & \cdots & s_{k-1}^T y_0 - s_0^T B_0 s_{k-1} & s_k^T y_0 - s_0^T B_0 s_k \\ s_1^T y_0 - s_1^T B_0 s_0 & s_1^T y_1 - s_1^T B_0 s_1 & \cdots & s_{k-1}^T y_1 - s_1^T B_0 s_{k-1} & s_k^T y_1 - s_1^T B_0 s_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{k-1}^T y_0 - s_{k-1}^T B_0 s_0 & s_{k-1}^T y_1 - s_{k-1} B_0 s_1 & \cdots & s_{k-1}^T y_{k-1} - s_{k-1}^T B_0 s_{k-1} & s_k^T y_{k-1} - s_{k-1}^T B_0 s_k \\ s_k^T y_0 - s_k B_0 s_0 & s_k y_1 - s_k B_0 s_1 & \cdots & s_k^T y_{k-1} - s_k^T B_0 s_k & s_k^T y_k - s_k^T B_0 s_k \end{bmatrix}$$

$$= \begin{bmatrix} s_0^T y_0 - s_0^T B_0 s_0 & s_1^T y_0 - s_0^T B_0 s_1 & \cdots & s_{k-1}^T y_0 - s_0^T B_0 s_{k-1} & y_0^T s_k - s_0^T B_0 s_k \\ s_1^T y_0 - s_1^T B_0 s_0 & s_1^T y_1 - s_1^T B_0 s_1 & \cdots & s_{k-1}^T y_1 - s_1^T B_0 s_{k-1} & y_1^T s_k - s_1^T B_0 s_k \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ s_{k-1}^T y_0 - s_{k-1}^T B_0 s_0 & s_{k-1}^T y_1 - s_{k-1} B_0 s_1 & \cdots & s_{k-1}^T y_{k-1} - s_{k-1}^T B_0 s_{k-1} & y_{k-1}^T s_k - s_{k-1}^T B_0 s_k \\ s_k^T y_0 - s_k B_0 s_0 & s_k y_1 - s_k B_0 s_1 & \cdots & s_k^T y_{k-1} - s_k^T B_0 s_k & s_k^T y_k - s_k^T B_0 s_k \end{bmatrix}$$

$$= \begin{bmatrix} \Sigma_k & Q_k^T s_k \\ s_k^T Q_k & s_k^T y_k - s_k^T B_0 s_k \end{bmatrix}$$

Therefore prove that we can represent Q_{k+1} and Σ_{k+1}. And these two matrix has dimension which requires less storage.

4. *Multiplication forward sweep*:

$x_k = x_i \cdot x_j, \frac{\partial x_k}{\partial x_i} = x_j, \frac{\partial x_k}{\partial x_j} = x_i$

one multiplication.

Multiplication reverse sweep:

$\bar{x}_k = 1, \bar{x}_i+ = \bar{x}_k \cdot x_j = x_j, \bar{x}_j+ = \bar{x}_k \cdot x_i = x_i$

two addictions, and two multiplications.

Cosine forward sweep:

$x_k = \cos(x_i), \frac{\partial x_k}{x_i} = -\sin(x_i)$

one cosine function, one sine function, and one multiplication.

Cosine reverse sweep:

$\bar{x}_k = 1$

$\bar{x}_i += = \bar{x}_k \cdot -\sin(x_i)$

one addiction, one multiplication.

5. First we can say that this computational graph is directed acyclic graph, and only one ending point, with out-degree as 0 at first. And we can remove the node once it gets finalized.

Suppose at one reverse sweep step, no nodes can get finalized. That is to say, all nodes have out-degree greater than or equal to 1. But this can't be true for directed acyclic graph, so the assumption doesn't hold.

Claim is, at each step, we can find at least one node with out-degree as 0, and such 0 out-degree nodes can get finalized.