Used lecture 16 notes as ref.

(3) $$F(W,H) = \frac{1}{2}\|WH - X\|_F^2 + \lambda\|H\|_1$$

subj: $W_{ij} \geq 0 \;\forall i,j$ & $H_{ij} \geq 0 \;\forall i,j$, $\|H\|_1 = \sum_i \sum_j |H_{i,j}|$

Grad w.r. to H:

$$\frac{\partial F(W,H)}{\partial H} = W^T(WH - X) + \lambda = \nabla_H F(W,H) \quad \text{— (1)}$$

The induction hypothesis at iteration k is $W_k, H_k > 0$ we will show that $W_{k+1}, H_{k+1} > 0$

Step size: $$\frac{[\bar{H}_k]}{[W_k^T W_k H_k] + \delta 11^T} > 0 \quad \text{— (2)}$$

We will look at 2 cases:

(1) $\nabla_H F(W,H) < 0$

(2) $\nabla_H F(W,H) \geq 0$

(1) From (2) and gradient update rule, we have:

$$[H_{k+1}]_{ij} = [H_k]_{ij} - \underbrace{\frac{[\bar{H}_k]}{[W_k^T W_k H_k] + \delta 11^T} \cdot \nabla_H F(W,H)}_{<0}$$

Since its an addition of 2 positive non zero terms, $[H_{k+1}]_{ij}$ is positive $\Rightarrow$

$$[H_{k+1}]_{ij} \geq [H_k]_{ij} > 0$$

② We know that $[\bar{H}_k]_{ij} = \begin{cases} [H_k]_{ij} & \text{if } \nabla_H F(W,H) \geq 0 \\ \max([H_k]_{ij}, \sigma) & \text{otherwise} \end{cases}$

Since $\nabla_H F(W,H) \geq 0$, we have:

$$[\bar{H}_k]_{ij} = [H_k]_{ij} \quad \text{and} \quad [W_k^T W_k H_k]_{ij} \geq [W_k^T W_k H_k]_{ij}$$

Using the gradient update for H_k, we have:

$$[H_{k+1}]_{ij} = [H_k]_{ij} - \frac{[\bar{H}_k]_{ij}}{[W_k^T W_k \bar{H}_k]_{ij} + \delta 11^T}[\nabla_H F(W,H)]_{ij}$$

$$\geq [H_k]_{ij} - \frac{[H_k]_{ij}}{[W_k^T W_k H_k]_{ij} + \delta 11^T}[\nabla_H F(W,H)]_{ij}$$

$$= [H_k]_{ij}\left(1 - \frac{[\nabla_H F(W,H)]_{ij}}{[W_k^T W_k H_k]_{ij} + \delta}\right)$$

$$= [H_k]_{ij}\left(1 - \frac{[W_k^T(W_k H - X)]_{ij} + \lambda}{[W_k^T W_k H_k]_{ij} + \delta}\right)$$

Let $d = [W_k^T W_k H_k]_{ij} + \delta$ be the denominator

$$= [H_k]_{ij}\left(1 - \frac{[W_k^T W_k H_k]_{ij}}{d} + \frac{[W_k^T X]_{ij}}{d} - \frac{\lambda}{d}\right)$$

$$[H_{k+1}]_{ij} = [H_k]_{ij}\left(\frac{[W_k^T X]_{ij} - \lambda + \delta}{d}\right)$$

If we set $\delta \geq \lambda$, we will gurantee non-negative updates