Investigating Integrative Non-Negative Matrix Factorization for Batch Effect Correction

Tianyi Liu

November 28, 2020

1 Frobenius Norm

$$\begin{aligned} \min_{W,H_k,V_k} \quad & \sum_{k=1}^{2} \|X_k - (W+V_k)H_k\|_F^2 + \lambda \sum_{k=1}^{2} \|V_k H_k\|_F^2 && (1) \\ \text{s.t.} \quad & W \geq 0, H_k \geq 0, V_k \geq 0, k \in [1,2] \\ & \|X\|_F^2 = \sum_{ij} X_{ij}^2 && (2) \end{aligned}$$

1.1 Update Rules

$$\begin{aligned} W &\leftarrow W \circ \frac{\sum_{k=1}^{2} X_k H_k^T}{\sum_{k=1}^{2} (W+V_k) H_k H_k^T} \\ V_k &\leftarrow V_k \circ \frac{X_k H_k^T}{(W+V_k) H_k H_k^T + \lambda V_k H_k H_k^T} \\ H_k &\leftarrow H_k \circ \frac{(W+V_k)^T X_k}{(W+V_k)^T (W+V_k) H_k + \lambda V_k^T V_k H_k} \end{aligned}$$

2 Frobenius Norm with Penalty

We consider a simplified scenario that there are only two batches in the dataset, i.e., $k = 1, 2$. The optimization problem is defined as follows:

$$\begin{aligned} \min_{W,H_k,V_k} \quad & \sum_{k=1}^{2} \|X_k - (W+V_k)H_k\|_F^2 + \lambda \sum_{k=1}^{2} \|V_k H_k\|_F^2 + \gamma \|H_1 - H_2\|_F^2 && (3) \\ \text{s.t.} \quad & W \geq 0, H_k \geq 0, V_k \geq 0, k \in [1,2] \end{aligned}$$

This is the conventional iNMF objective with an additional weighted Frobenius norm penalty.

2.1 Update Rules

$$
\begin{aligned}
W &\leftarrow W \circ \frac{\sum_{k=1}^{2} X_k H_k^T}{\sum_{k=1}^{2}(W+V_k)H_k H_k^T} \\
V_k &\leftarrow V_k \circ \frac{X_k H_k^T}{(W+V_k)H_k H_k^T + \lambda V_k H_k H_k^T} \\
H_1 &\leftarrow H_1 \circ \frac{(W+V_1)^T X_1}{(W+V_1)^T(W+V_1)H_1 + (\lambda V_1^T V_1 + \gamma)H_1 - \gamma H_2} \\
H_2 &\leftarrow H_2 \circ \frac{(W+V_2)^T X_2}{(W+V_2)^T(W+V_2)H_2 + (\lambda V_2^T V_2 + \gamma)H_2 - \gamma H_1}
\end{aligned}
$$

2.2 Gradient Derivation

2.2.1 Gradient for W

Now, we evaluate the gradient of (3) as follows:

$$
\begin{aligned}
f(X+\delta X) &\approx f(X) + \langle \nabla f(X), \delta X\rangle + o(\|\delta X\|) \\
f(W+\delta W) &= \sum_{k=1}^{2} \|X_k - (W+\delta W+V_k)H_k\|_F^2 \\
&= \sum_{k=1}^{2} \langle X_k - (W+\delta W+V_k)H_k, X_k - (W+\delta W+V_k)H_k\rangle \\
&= \sum_{k=1}^{2} \left[\|X_k - (W+V_k)H_k\|_F^2 + 2\langle -\delta W H_k, X_k - (W+V_k)H_k\rangle + \|\delta W H_k\|_F^2\right]
\end{aligned}
$$

$$
\begin{aligned}
\sum_{k=1}^{2} 2\langle -\delta W H_k, X_k - (W+V_k)H_k\rangle &= 2\sum_{k=1}^{2} \mathbf{Tr}\left((-\delta W H_k)^T (X_k - (W+V_k)H_k)\right) \\
&= 2\sum_{k=1}^{2} \mathbf{Tr}\left((-\delta W)^T (X_k - (W+V_k)H_k)H_k^T\right) \\
&= -2\sum_{k=1}^{2} \langle X_k H_k^T - (W+V_k)H_k H_k^T, \delta W\rangle \\
&= -2\langle \sum_{k=1}^{2} X_k H_k^T - (W+V_k)H_k H_k^T, \delta W\rangle \\
\nabla f_W &= -2\sum_{k=1}^{2} X_k H_k^T - (W+V_k)H_k H_k^T \\
W &\leftarrow W + \alpha_W \left(2\sum_{k=1}^{2}\left(X_k H_k^T - (W+V_k)H_k H_k^T\right)\right)
\end{aligned}
$$

Set $\alpha_W = \frac{1}{2}\frac{W}{\sum_{k=1}^{2}(W+V_k)H_kH_k^T}$:

$$W \leftarrow W \circ \frac{\sum_{k=1}^{2} X_k H_k^T}{\sum_{k=1}^{2}(W+V_k)H_kH_k^T} \tag{4}$$

2.2.2 Gradient for V_k

$$\begin{aligned} f(V_k+\delta V_k) &= \|X_k-(W+V_k+\delta V_k)H_k\|_F^2 + \lambda\|(V_k+\delta V_k)H_k\|_F^2 \\ &= \langle X_k-(W+V_k+\delta V_k)H_k, X_k-(W+V_k+\delta V_k)H_k\rangle + \lambda\langle (V_k+\delta V_k)H_k, (V_k+\delta V_k)H_k\rangle \\ &= \|X_k-(W+V_k)H_k\|_F^2 + 2\langle -\delta V_kH_k, X_k-(W+V_k)H_k\rangle + \|\delta VH_k\|_F^2 \\ &\quad + \lambda\|V_kH_k\|_F^2 + 2\lambda\langle \delta V_kH_k, V_kH_k\rangle + \lambda\|\delta V_kH_k\|_F^2 \end{aligned}$$

$$\begin{aligned} 2\langle -\delta V_kH_k, X_k-(W+V_k)H_k\rangle & \\ +2\lambda\langle \delta V_kH_k, V_kH_k\rangle &= -2\mathbf{Tr}\left((\delta V_kH_k)^T(X_k-(W+V_k)H_k)\right) + 2\lambda\mathbf{Tr}\left((\delta V_kH_k)^TV_kH_k\right) \\ &= -2\mathbf{Tr}\left(\delta V_k^T(X_k-(W+V_k)H_k)H_k^T\right) + 2\lambda\mathbf{Tr}\left((\delta V_k)^TV_kH_kH_k^T\right) \\ &= -2\langle (X_k-(W+V_k)H_k)H_k^T, \delta V_k\rangle + 2\lambda\langle V_kH_kH_K^T, \delta V_k\rangle \\ \nabla f_{V_k} &= -2(X_kH_k^T-(W+V_k)H_kH_k^T) + 2\lambda V_kH_kH_k^T \\ V_k &\leftarrow V_k - \alpha_V\left(-2(X_kH_k^T-(W+V_k)H_kH_k^T) + 2\lambda V_kH_kH_k^T\right) \end{aligned}$$

Set $\alpha_V = \frac{V_k}{2(W+V_k)H_kH_k^T+2\lambda V_kH_kH_k^T}$:

$$V_k \leftarrow V_k \circ \frac{X_kH_k^T}{(W+V_k)H_kH_k^T+\lambda V_kH_kH_k^T} \tag{5}$$

2.2.3 Gradient for H_k

$$f_1(H_k) = \sum_{k=1}^{2}\|X_k-(W+V_k)H_k\|_F^2 + \lambda\sum_{k=1}^{2}\|V_kH_k\|_F^2$$

$$\begin{aligned} f_2(H_k) &= \|H_1-H_2\|_F^2 \\ f(H_k) &= f_1(H_k)+f_2(H_k) \\ \nabla f_{H_k} &= \nabla f_{1H_k} + \nabla f_{2H_k} \\ f_1(H_k+\delta H_k) &= \|X_k-(W+V_k)(H_k+\delta H_k)\|_F^2 + \lambda\|V_k(H_k+\delta H_k)\|_F^2 \\ &= \langle X_k-(W+V_k)(H_k+\delta H_k), X_k-(W+V_k)(H_k+\delta H_k)\rangle \\ &\quad + \lambda\langle V_k(H_k+\delta H_k), V_k(H_k+\delta H_k)\rangle \\ &= \|X_k-(W+V_k)H_k\|_F^2 - 2\langle (W+V_k)\delta H_k, X_k-(W+V_k)H_k\rangle + \|(W+V_k)\delta H_k\|_F^2 \\ &\quad + \lambda\|V_kH_k\|_F^2 + 2\lambda\langle V_k\delta H_k, V_kH_k\rangle + \lambda\|V_k\delta H_k\|_F^2 \end{aligned}$$

$$
\begin{aligned}
-2\langle (W+V_k)\delta H_k, X_k-(W+V_k)H_k\rangle &+ 2\lambda\langle V_k\delta H_k, V_kH_k\rangle \\
&= -2\mathbf{Tr}\left(((W+V_k)\delta H_k)^T(X_k-(W+V_k)H_k)\right) + 2\lambda\mathbf{Tr}\left((V_k\delta H_k)^T V_kH_k\right) \\
&= -2\mathbf{Tr}\left((\delta H_k)^T(W+V_k)^T(X_k-(W+V_k)H_k)\right) + 2\lambda\mathbf{Tr}\left((\delta H_k)^T V_k^T V_kH_k\right) \\
&= -2\langle (W+V_k)^T(X_k-(W+V_k)H_k), \delta H_k\rangle + 2\lambda\langle V_k^T V_kH_k, \delta H_k\rangle \\
\nabla f_{1H_k} &= -2(W+V_k)^T(X_k-(W+V_k)H_k) + 2\lambda V_k^T V_k H_k \\
f_{2H_1} &= \gamma\|H_1+\delta H_1-H_2\|_F^2 \\
&= \gamma\langle H_1+\delta H_1-H_2, H_1+\delta H_1-H_2\rangle \\
&= \gamma\|H_1-H_2\|_F^2 + \|\delta H_1\|_F^2 + 2\langle \delta H_1, H_1-H_2\rangle \\
\nabla f_{2H_1} &= 2\gamma(H_1-H_2) \\
f_{2H_2} &= \gamma\|H_1-(H_2+\delta H_2)\|_F^2 \\
&= \gamma\langle H_1-(H_2+\delta H_2), H_1-(H_2+\delta H_2)\rangle \\
&= \gamma\|H_1-H_2\|_F^2 + \|\delta H_2\|_F^2 - 2\langle \delta H_2, H_1-H_2\rangle \\
\nabla f_{2H_2} &= 2\gamma(H_2-H_1) \\
\nabla f_{H_1} &= -2(W+V_1)^T X_1 + 2(W+V_1)^T(W+V_1)H_1 + 2\lambda V_1^T V_1 H_1 + 2\gamma(H_1-H_2) \\
&= -2(W+V_1)^T X_1 + 2(W+V_1)^T(W+V_1)H_1 + 2(\lambda V_1^T V_1+\gamma)H_1 - 2\gamma H_2 \\
\nabla f_{H_2} &= -2(W+V_2)^T X_2 + 2(W+V_2)^T(W+V_2)H_2 + 2\lambda V_2^T V_2 H_2 - 2\gamma(H_1-H_2) \\
&= -2(W+V_2)^T X_2 + 2(W+V_2)^T(W+V_2)H_2 + 2(\lambda V_2^T V_2+\gamma)H_2 - 2\gamma H_1 \\
H_1 &\leftarrow H_1 + \alpha_{H_1}\left[2(W+V_1)^T X_1 - 2(W+V_1)^T(W+V_1)H_1 - 2(\lambda V_1^T V_1+\gamma)H_1 + 2\gamma H_2\right] \\
H_2 &\leftarrow H_2 + \alpha_{H_2}\left[2(W+V_2)^T X_2 - 2(W+V_2)^T(W+V_2)H_2 - 2(\lambda V_2^T V_2+\gamma)H_2 + 2\gamma H_1\right]
\end{aligned}
$$

Set $\alpha_{H_1} = \frac{H_1}{2(W+V_1)^T(W+V_1)H_1+2(\lambda V_1^T V_1+\gamma)H_1-2\gamma H_2}, \alpha_{H_2} = \frac{H_2}{2(W+V_2)^T(W+V_2)H_2+2(\lambda V_2^T V_2+\gamma)H_2-2\gamma H_1}$
:

$$
\begin{aligned}
H_1 &\leftarrow H_1 \circ \frac{(W+V_1)^T X_1}{(W+V_1)^T(W+V_1)H_1 + (\lambda V_1^T V_1+\gamma)H_1 - \gamma H_2} \\
H_2 &\leftarrow H_2 \circ \frac{(W+V_2)^T X_2}{(W+V_2)^T(W+V_2)H_2 + (\lambda V_2^T V_2+\gamma)H_2 - \gamma H_1}
\end{aligned}
$$

3 KL Divergence

$$
\begin{aligned}
\min_{W,H_k,V_k} \quad & \sum_{k=1}^{2} D_{KL}(X_k \parallel (W+V_k)H_k) + \lambda\sum_{k=1}^{2}\|H_kV_k\|_F^2 \qquad (6) \\
\text{s.t.} \quad & W \geq 0, H_k \geq 0, V_k \geq 0, k \in [1,2]
\end{aligned}
$$

$$
D_{KL}(A \parallel B) = \sum_{ij}\left(A_{ij}\log\frac{A_{ij}}{B_{ij}} - A_{ij} + B_{ij}\right) \qquad (7)
$$

3.1 Update Rules

$$
\begin{aligned}
W_{ij} &\leftarrow W_{ij} \frac{\sum_{k=1}^{2} \sum_m \frac{X_{k,im} H_{k,jm}}{((W+V_k)H_k)_{im}}}{\sum_{k=1}^{2} \sum_m H_{k,jm}} \\
V_{k,ij} &\leftarrow V_{k,ij} \frac{\sum_m \frac{X_{k,im} H_{k,jm}}{((W+V_k)H_k)_{im}}}{\sum_m H_{k,jm} + 2\lambda (V_k H_k H_k^T)_{ij}} \\
H_{k,ij} &\leftarrow H_{k,ij} \frac{\sum_m \frac{(W+V_k)_{mi} X_{k,mj}}{((W+V_k)H_k)_{mj}}}{\sum_m (W+V_k)_{mi} + 2\lambda (V_k^T V_k H_k)_{ij}}
\end{aligned}
$$

4 KL Divergence with Penalty

$$
\begin{aligned}
\min_{W,H_k,V_k} \quad & \sum_{k=1}^{2} D_{KL}(X_k \parallel (W+V_k)H_k) + \lambda \sum_{k=1}^{2} \|V_k H_k\|_F^2 + \gamma D_{KL}(H_1 \parallel H_2) \\
\text{s.t.} \quad & W \geq 0, H_k \geq 0, V_k \geq 0, k \in [1,2]
\end{aligned}
\tag{8}
$$

4.1 Update Rules

$$
\begin{aligned}
W_{ij} &\leftarrow W_{ij} \frac{\sum_{k=1}^{2} \sum_m \frac{X_{k,im} H_{k,jm}}{((W+V_k)H_k)_{im}}}{\sum_{k=1}^{2} \sum_m H_{k,jm}} \\
V_{k,ij} &\leftarrow V_{k,ij} \frac{\sum_m \frac{X_{k,im} H_{k,jm}}{((W+V_k)H_k)_{im}}}{\sum_m H_{k,jm} + 2\lambda (V_k H_k H_k^T)_{ij}} \\
H_{1,ij} &\leftarrow H_{1,ij} \frac{\sum_m \frac{(W+V_1)_{mi} X_{k,mj}}{((W+V_1)H_1)_{mj}}}{\sum_m (W+V_1)_{mi} + \gamma \log \frac{H_{1,ij}}{H_{2,ij}} + 2\lambda (V_1^T V_1 H_1)_{ij}} \\
H_{2,ij} &\leftarrow H_{2,ij} \frac{\sum_m \frac{(W+V_2)_{mi} X_{k,mj}}{((W+V_2)H_2)_{mj}}}{\sum_m (W+V_2)_{mi} + \gamma \left(1 - \frac{H_{1,ij}}{H_{2,ij}}\right) + 2\lambda (V_2^T V_2 H_2)_{ij}}
\end{aligned}
$$

4.2 Gradient Derivation

4.2.1 Gradient for W

$$\begin{aligned}
\frac{\partial D_{KL}(X_k \parallel (W+V_k)H_k)}{\partial W_{ij}} &= \frac{\partial}{\partial W_{ij}} \sum_{ij} \left(X_{ij} \log \frac{X_{ij}}{((W+V_k)H_k)_{ij}} - X_{ij} + ((W+V_k)H_k)_{ij} \right) \\
&= -\sum_m \frac{X_{k,im}H_{k,jm}}{((W+V_k)H_k)_{im}} + \sum_m H_{k,jm} \\
W_{ij} \leftarrow W_{ij} &+ \alpha_W \left(\sum_{k=1}^{2} \sum_m \frac{X_{k,im}H_{k,jm}}{((W+V_k)H_k)_{im}} - \sum_{k=1}^{2} \sum_m H_{k,jm} \right)
\end{aligned}$$

Set $\alpha_W = \frac{W_{ij}}{\sum_{k=1}^{2} \sum_m H_{k,jm}}$:

$$W_{ij} \leftarrow W_{ij} \frac{\sum_{k=1}^{2} \sum_m \frac{X_{k,im}H_{k,jm}}{((W+V_k)H_k)_{im}}}{\sum_{k=1}^{2} \sum_m H_{k,jm}}$$

4.2.2 Gradient for V_k

$$\begin{aligned}
\frac{\partial D_{KL}(X_k \parallel (W+V_k)H_k)}{\partial V_{k,ij}} &= \frac{\partial}{\partial V_{k,ij}} \sum_{ij} \left(X_{ij} \log \frac{X_{ij}}{((W+V_k)H_k)_{ij}} - X_{ij} + ((W+V_k)H_k)_{ij} \right) \\
&= -\sum_m \frac{X_{k,im}H_{jm}}{((W+V_k)H_k)_{im}} + \sum_m H_{k,jm} \\
\frac{\partial \lambda \|V_k H_k\|_F^2}{\partial V_k} &= 2\lambda V_k H_k H_k^T
\end{aligned}$$

$$V_{k,ij} \leftarrow V_{k,ij} + \alpha_V \left(\sum_m \frac{X_{k,im}H_{k,jm}}{((W+V_k)H_k)_{im}} - \sum_m H_{k,jm} - 2\lambda (V_k H_k H_k^T)_{ij} \right)$$

Set $\alpha_V = \frac{V_{k,ij}}{\sum_m H_{k,jm} + 2\lambda(V_k H_k H_k^T)_{ij}}$:

$$V_{k,ij} \leftarrow V_{k,ij} \frac{\sum_m \frac{X_{k,im}H_{k,jm}}{((W+V_k)H_k)_{im}}}{\sum_m H_{k,jm} + 2\lambda(V_k H_k H_k^T)_{ij}}$$

4.2.3 Gradient for H_k

$$\begin{aligned}
\frac{\partial D_{KL}(X_k \parallel (W+V_k)H_k)}{\partial H_{k,ij}} &= \frac{\partial}{\partial H_{k,ij}} \sum_{ij} \left(X_{ij} \log \frac{X_{k,ij}}{((W+V_k)H_k)_{ij}} - X_{k,ij} + ((W+V_k)H_k)_{ij} \right) \\
&= -\sum_m \frac{(W+V_k)_{mi} X_{mj}}{((W+V_k)H_k)_{mj}} + \sum_m (W+V_k)_{mi} \\
\frac{\partial D_{KL}(H_1 \parallel H_2)}{\partial H_{1,ij}} &= \frac{\partial}{\partial H_{1,ij}} \sum_{ij} \left(H_{1,ij} \log \frac{H_{1,ij}}{H_{2,ij}} - H_{1,ij} + H_{2,ij} \right) \\
&= \log \frac{H_{1,ij}}{H_{2,ij}} \\
\frac{\partial D_{KL}(H_1 \parallel H_2)}{\partial H_{2,ij}} &= \frac{\partial}{\partial H_{2,ij}} \sum_{ij} \left(H_{1,ij} \log \frac{H_{1,ij}}{H_{2,ij}} - H_{1,ij} + H_{2,ij} \right) \\
&= 1 - \frac{H_{1,ij}}{H_{2,ij}} \\
\frac{\partial \lambda \sum_{k=1}^{2} \|V_k H_k\|_F^2}{\partial H_k} &= 2\lambda V_k^T V_k H_k
\end{aligned}$$

$$\begin{aligned}
H_{1,ij} &\leftarrow H_{1,ij} - \alpha_H \left(-\sum_m \frac{(W+V_1)_{mi} X_{k,mj}}{((W+V_1)H_1)_{mj}} + \sum_m (W+V_1)_{mi} + \gamma \log \frac{H_{1,ij}}{H_{2,ij}} + 2\lambda (V_1^T V_1 H_1)_{ij} \right) \\
H_{2,ij} &\leftarrow H_{2,ij} - \alpha_H \left(-\sum_m \frac{(W+V_2)_{mi} X_{k,mj}}{((W+V_2)H_2)_{mj}} + \sum_m (W+V_2)_{mi} + \gamma \left(1 - \frac{H_{1,ij}}{H_{2,ij}} \right) + 2\lambda (V_2^T V_2 H_2)_{ij} \right)
\end{aligned}$$

Set $\alpha_{H1} = \frac{H_{1,ij}}{\sum_m (W+V_1)_{mi} + \gamma \log \frac{H_{1,ij}}{H_{2,ij}} + 2\lambda (V_1^T V_1 H_1)_{ij}}, \alpha_{H2} = \frac{H_{2,ij}}{\sum_m (W+V_2)_{mi} + \gamma \left(1 - \frac{H_{1,ij}}{H_{2,ij}}\right) + 2\lambda (V_2^T V_2 H_2)_{ij}}$:

$$\begin{aligned}
H_{1,ij} &\leftarrow H_{1,ij} \frac{\sum_m \frac{(W+V_1)_{mi} X_{k,mj}}{((W+V_1)H_1)_{mj}}}{\sum_m (W+V_1)_{mi} + \gamma \log \frac{H_{1,ij}}{H_{2,ij}} + 2\lambda (V_1^T V_1 H_1)_{ij}} \\
H_{2,ij} &\leftarrow H_{2,ij} \frac{\sum_m \frac{(W+V_2)_{mi} X_{k,mj}}{((W+V_2)H_2)_{mj}}}{\sum_m (W+V_2)_{mi} + \gamma \left(1 - \frac{H_{1,ij}}{H_{2,ij}} \right) + 2\lambda (V_2^T V_2 H_2)_{ij}}
\end{aligned}$$