Derive the Block Coordinate Descent Rules
for Non Negative Matrix Tri-Factorization (NMTF) with Regularization Terms

Objective

Given input matrix $X \in \mathbb{R}_{\geq 0}^{n \times m}$, find $U \in \mathbb{R}_{\geq 0}^{n \times k1}$, $S \in \mathbb{R}_{\geq > 0}^{k_1 \times k_2}$, and $V \in \mathbb{R}_{\geq 0}^{k2 \times m}$ that minimize the function

$$0 = \left\| X - USV^\top \right\|_\mathrm{F}^2 + \alpha_u \sum_{i=1}^{n} \left\| U_{i,.} \right\|_1 + \alpha_v \sum_{i=1}^{m} \left\| V_{i,.} \right\|_1 + \lambda_u \left\| U^\top U - I_n \right\|_1 + \lambda_v \left\| V^\top V - I_m \right\|_1 \tag{1}$$

subject to the norm of column vectors, u_i and v_i of U and V to be of unit norm.

Note on Normalization of U and V factors

We require that both u_i and v_i have unit norm. Consider the fallowing:

$$X \approx USV^\top \tag{2}$$

$$USV^\top = \sum_{i=1}^{k1} \sum_{j=1}^{k2} u_i s_{i,j} v_j^\top \tag{3}$$

$$USV^\top = \sum_{i=1}^{k1} \sum_{j=1}^{k2} \frac{u_i}{\left\| u_i \right\|_2} (\left\| u_i \right\|_2 s_{i,j} \left\| v_j \right\|_2) \frac{v_j}{\left\| v_j \right\|_2} \tag{4}$$

$$\tag{5}$$

In this way, $s_{i,j}$ can absorb the normalization factors of both u_i and v_j. This step needs to be performed between every iteration of the algorithm.

Optimize v_j

Consider the substitution $P = US \in \mathbb{R}_{\geq 0}^{nxk_2}$, then the objective function is equivalent to:

$$O = \left\| X - PV^\top \right\|_\mathrm{F}^2 + \alpha_u \sum_{i=1}^{n} \left\| U_{i,.} \right\|_1 + \alpha_v \sum_{i=1}^{m} \left\| V_{i,.} \right\|_1 + \lambda_u \left\| U^\top U - I_n \right\|_1 + \lambda_v \left\| V^\top V - I_m \right\|_1 \tag{6}$$

$$= \left\| X - \sum_{j=1}^{k_2} p_j v_j^\top \right\|_\mathrm{F}^2 + \alpha_v \sum_{j=1}^{k2} \left\| v_j \right\|_1 + \lambda_v \sum_{j=1}^{k2} \sum_{h \neq j} v_j^\top v_h + C \tag{7}$$

Where p_j is the jth column of P, v_j is the jth column of V, and C are terms that only contain U. Now 'pull out' terms correspond to the jth column:

$$O = \left\| X - p_j v_j^\top - \sum_{h \neq j} p_j v_j^\top \right\|_\mathrm{F}^2 + \alpha_v \left\| v_j \right\|_1 + \alpha_v \sum_{h \neq j} \left\| v_h \right\|_1 + \lambda_v \sum_{h \neq j} v_j^\top v_h + \lambda_v \sum_{k \neq j} \sum_{h \neq k} v_k^\top v_h + C \tag{8}$$

Substitute $R_j = X - \sum_{h \neq j} p_j v_j^\top$ and let C now represent all terms that do not include v_j.

$$O = \left\| R_j - p_j v_j^\top \right\|_\mathrm{F}^2 + \alpha_v \left\| v_j \right\|_1 + \lambda_v \sum_{h \neq j} v_j^\top v_h + C \tag{9}$$

$$\tag{10}$$

Now expand using the trace.

$$O = Tr[(R_j - p_j v_j^\top)^\top (R_j - p_j v_j^\top)] + \alpha_v \left\| v_j \right\|_1 + \lambda_v \sum_{h \neq j} v_j^\top v_h + C \tag{11}$$

$$= Tr(R_j^\top R_j) - 2Tr(R_j^\top p_j v_j^\top) + Tr(v_j p_j^\top p_j v_j^\top) + \alpha_v \left\| v_j \right\|_1 + \lambda_v \sum_{h \neq j} v_j^\top v_h + C \tag{12}$$

$$= Tr(R_j^\top R_j) - 2(R_j^\top p_j)^\top v_j + p_j^\top p_j v_j^\top v_j + \alpha_v \left\| v_j \right\|_1 + \lambda_v \sum_{h \neq j} v_j^\top v_h + C \tag{13}$$

We can expand the 1 norm on v_j with the inner product $I_m^\top * v_j$ due to the non-negativity of V.

$$= Tr(R_j^\top R_j) - 2(R_j^\top p_j)^\top v_j + p_j^\top p_j v_j^\top v_j + \alpha_v I_m^\top * v_j + \lambda_v \sum_{h \neq j} v_j^\top v_h + C \tag{14}$$

Take the derivative and minimize with respect to v_j

$$\frac{\partial O}{\partial v_j} = 0 - 2R_j^\top p_j + 2p_i^\top p_j v_j + \alpha_v I_m + \lambda_v \sum_{h \neq j} v_j \tag{15}$$

$$0 = -R_j^\top p_j + \|p_i\|_2^2 v_j + \frac{\alpha_v}{2} I_m + \frac{\lambda_v}{2} \sum_{h \neq j} v_j \tag{16}$$

$$v_j = \frac{R_j^\top p_j - \frac{\alpha_v}{2} I_m - \frac{\lambda_v}{2} \sum_{h \neq j} v_j}{\|p_j\|_2^2} \tag{17}$$

Finally we need to constrain v_j to the non-negative constraint

$$v_j = \frac{[R_j^\top p_j - \frac{\alpha_v}{2} I_m - \frac{\lambda_v}{2} \sum_{h \neq j} v_j]_+}{\|p_j\|_2^2} \tag{18}$$

Optimize u_i

The optimization of u_i is similar. Consider the substitution $Q = (SV^\top)^\top = VS^\top \in \mathbb{R}_{\geq 0}^{m*k_1}$, then the objective function is equivalent to:

$$O = \left\|X - UQ^\top\right\|_{\mathrm{F}}^2 + \alpha_u \sum_{i=1}^{n} \|U_{i,.}\|_1 + \alpha_v \sum_{i=1}^{m} \|V_{i,.}\|_1 + \lambda_u \left\|U^\top U - I_n\right\|_1 + \lambda_v \left\|V^\top V - I_m\right\|_1 \tag{19}$$

$$= \left\| X - \sum_{i=1}^{k_1} u_i q_i^\top \right\|_{\mathrm{F}}^2 + \alpha_v \sum_{j=1}^{k1} \|u_j\|_1 + \lambda_v \sum_{j=1}^{k1} \sum_{h \neq j} u_j^\top u_h + C \tag{20}$$

Where u_i is the ith column of U, q_i is the ith column of Q, and C are terms that only contain V. Now 'pull out' terms correspond to the ith column:

$$O = \left\| X - u_i q_i^\top - \sum_{h \neq i} u_h q_i^\top \right\|_{\mathrm{F}}^2 + \alpha_u \|u_i\|_1 + \alpha_u \sum_{j \neq i} \|u_j\|_1 + \lambda_u \sum_{j \neq i} u_i^\top u_j + \lambda_u \sum_{h \neq i} \sum_{j \neq h} u_h^\top u_j + C \tag{21}$$

Substitute $R_i = X - \sum_{h \neq i} u_h q_i^\top$ and let C represent all terms that do not contain u_i:

$$O = \left\|R - u_i q_i^\top\right\|_{\mathrm{F}}^2 + \alpha_u \|u_i\|_1 + \lambda_u \sum_{j \neq i} u_i^\top u_j + C \tag{22}$$

We can now expand using trace:

$$O = Tr[(R_i - u_i q_i^\top)^\top (R_i - u_i q_i^\top)] + \alpha_u \|u_i\|_1 + \lambda_u \sum_{j \neq i} u_i^\top u_j + C \tag{23}$$

$$= Tr(R_i^\top R_i) - 2Tr(R_i^\top u_i q_i^\top) + Tr(u_i q_i^\top u_i q_i^\top) + \alpha_u \|u_i\|_1 + \lambda_u \sum_{j \neq i} u_i^\top u_j + C \tag{24}$$

$$= Tr(R_i^\top R_i) - 2u_i^\top R_i q_i + (u_i^\top u_i)(q_i \top q_i) + \alpha_u \|u_i\|_1 + \lambda_u \sum_{j \neq i} u_i^\top u_j + C \tag{25}$$

We can expand the 1 norm on u_i with the inner product $I_n^\top * u_i$ due to the non-negativity of U:

$$O = Tr(R_i^\top R_i) - 2u_i^\top R_i q_i + (u_i^\top u_i)(q_i \top q_i) + \alpha_u I_n^\top * u_i + \lambda_u \sum_{j \neq i} u_i^\top u_j + C \tag{26}$$

Take the derivative and minimize with respect to u_i

$$\frac{\partial O}{\partial u_i} = 0 - 2R_j q_i + 2q_i^\top q_i u_i + \alpha_u I_n + \lambda_u \sum_{j \neq i} u_j \tag{27}$$

$$0 = -R_i q_i + \|q_i\|_2^2 u_i + \frac{\alpha_u}{2} I_n + \frac{\lambda_u}{2} \sum_{j \neq i} u_j \tag{28}$$

$$u_i = \frac{R_i q_i - \frac{\alpha_u}{2} I_n - \frac{\lambda_u}{2} \sum_{j \neq i} u_j}{\|q_i\|_2^2} \tag{29}$$

Finally we need to constrain u_i to the non-negative constraint

$$u_i = \frac{[R_i q_i - \frac{\alpha_u}{2} I_n - \frac{\lambda_u}{2} \sum_{j \neq i} u_j]_+}{\|q_i\|_2^2} \tag{30}$$

Optimize $s_{i,j}$

The optimization of $s_{i,j}$ is unaffected by the additional regularization terms. For this reason I will just ignore those terms and show the derivation for $s_{i,j}$. For the final variable we must expand out in terms of both i and j.

$$O = \left\|X - USV^\top\right\|_\mathrm{F}^2 \tag{31}$$

$$O = \left\|X - \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} u_i s_{i,j} v_j^\top \right\|_\mathrm{F}^2 \tag{32}$$

Now 'pull out' the term that corresponds to a particular i and j.

$$O = \left\|X - u_i s_{i,j} v_j^\top - \sum_{c \neq i} \sum_{d \neq j} u_c s_{c,d} v_d^\top \right\|_\mathrm{F}^2 \tag{33}$$

Let $R_{i,j} = X - \sum_{c \neq i} \sum_{d \neq j} u_c s_{c,d} v_d^\top$, the expand using the trace:

$$O = Tr[(R_{i,j} - u_i s_{i,j} v_j^\top)^\top (R_{i,j} - u_i s_{i,j} v_j^\top)] \tag{34}$$

$$= Tr(R_{i,j}^\top R_{i,j}) - 2Tr(R_{i,j}^\top u_i s_{i,j} v_j \top) + Tr(s_{i,j}^2 v_j u_i^\top u_i v j^\top) \tag{35}$$

$$= Tr(R_{i,j}^\top R_{i,j}) - 2s_{i,j}(R_{i,j} \top u_i)^\top v_j + s_{i,j}^2 \|u_i\|_2^2 \|v_j\|_2^2 \tag{36}$$

Finally take the derivative with respect to $s_{i,j}$, solve, and apply the positivity constraint.

$$\frac{\partial O}{\partial s_{i,j}} = -2u_i^\top R_{i,j} v_j + 2s_{i,j} \|u_i\|_2^2 \|v_j\|_2^2 \tag{37}$$

$$s_{i,j} = \frac{u_i^\top R_{i,j} v_j}{\|u_i\|_2^2 \|v_j\|_2^2} \tag{38}$$

$$s_{i,j} = \frac{[u_i^\top R_{i,j} v_j]_+}{\|u_i\|_2^2 \|v_j\|_2^2} \tag{39}$$