Projective Non-Negative Matrix Factorization for unsupervised graph clustering
Supplementary Material

1 Appendix 1

For any two matrices $\mathbf{X}$, $\mathbf{H}$ we have that:

$$\left\|\mathbf{X} - \mathbf{X}\mathbf{H}\mathbf{H}^\top\right\|^2 = \mathrm{Tr}\Big\{(\mathbf{X} - \mathbf{X}\mathbf{H}\mathbf{H}^\top)^\top(\mathbf{X} - \mathbf{X}\mathbf{H}\mathbf{H}^\top)\Big\} \tag{1}$$

where Tr{.} denotes the trace of the corresponding matrix. In order to solve our minimization one needs to turn the constrained optimization problem into an unconstrained one. So, we introduce the respective Lagrange multipliers ϕ_{ij} for the non-negativity of H_{ij} by defining a $n \times c$ matrix $\Phi = [\phi_{ij}]$ and then seek to minimize an unconstrained version of $J(\mathbf{H})$, i.e:

$$\min_{\mathbf{H}} \hat{J}(\mathbf{H}), \text{ with } \hat{J}(\mathbf{H}) = J(\mathbf{H}) + \mathrm{Tr}(\mathbf{\Phi}\mathbf{H}^\top) \tag{2}$$

For the purposes of this section, we set $\mathbf{X} \leftarrow \mathbf{X}^\top$. Then, the Projective NMF minimization is re-formulated as:

$$\begin{aligned}\min_{\mathbf{H}} \hat{J}(\mathbf{H}) = \min_{\mathbf{H}} \Big\{ &Tr(\mathbf{X}^\top\mathbf{X} - 2\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X} \\ &+ \mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X} + \lambda\mathbf{H}^\top\mathbf{L}\mathbf{H} + \mathbf{\Phi}\mathbf{H}^\top)\Big\}\end{aligned} \tag{3}$$

where we also add the free parameter $\lambda \in R$ that reflects the tradeoff between spatial regularity and the data in $\mathbf{X}$. Taking the gradient with respect to $\mathbf{H}$ yields:

$$\begin{aligned}\frac{\partial \hat{J}}{\partial \mathbf{H}} &= -2(\mathbf{X} - \mathbf{H}\mathbf{H}^\top\mathbf{X})\mathbf{X}^\top\mathbf{H} - 2\mathbf{X}(\mathbf{X}^\top - \mathbf{X}^\top\mathbf{H}\mathbf{H}^\top)\mathbf{H} \\ &+ \lambda(\mathbf{L} + \mathbf{L}^\top)\mathbf{H} + \mathbf{\Phi} \\ &= -4\mathbf{X}\mathbf{X}^\top\mathbf{H} + 2\mathbf{H}\mathbf{H}^\top\mathbf{X}\mathbf{X}^\top\mathbf{H} + 2\mathbf{X}\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H} \\ &+ 2\lambda\mathbf{L}\mathbf{H} + \mathbf{\Phi}\end{aligned}$$

where we also use that $\mathbf{L}$ is symmetric. By replacing $\mathbf{L} = \mathbf{D} - \mathbf{S}$ and setting the gradient to 0 we get:

$$\begin{aligned}4\mathbf{X}\mathbf{X}^\top\mathbf{H} + 2\lambda\mathbf{S}\mathbf{H} = &\,2\mathbf{H}\mathbf{H}^\top\mathbf{X}\mathbf{X}^\top\mathbf{H} \\ &+ 2\mathbf{X}\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H} + 2\lambda\mathbf{D}\mathbf{H} + \mathbf{\Phi}\end{aligned} \tag{4}$$

Then, we multiply (4) element wise by H_{ij}^4:

$$\begin{aligned}4[\mathbf{X}\mathbf{X}^\top\mathbf{H}]_{ij}H_{ij}^4 + 2\lambda[\mathbf{S}\mathbf{H}]_{ij}H_{ij}^4 = &\,2[\mathbf{H}\mathbf{H}^\top\mathbf{X}\mathbf{X}^\top\mathbf{H}]_{ij}H_{ij}^4 \\ &+ 2[\mathbf{X}\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H}]_{ij}H_{ij}^4 + 2\lambda[\mathbf{D}\mathbf{H}]_{ij}H_{ij}^4 + \Phi_{ij}H_{ij}^4\end{aligned}$$

Finally, using the KKT conditions $\Phi_{ij} H_{ij} = 0$ and re-arranging yields the update rule.

2 Appendix 2 - Convergence of the update rule

Definition: $F(\mathbf{H}, \mathbf{H}')$ is an auxiliary function of $J(\mathbf{H}')$ if the following two conditions hold:

1. $F(\mathbf{H}, \mathbf{H}') \geq J(\mathbf{H}')$
2. $F(\mathbf{H}', \mathbf{H}') = J(\mathbf{H}')$

Lemma: if F is an auxiliary function then $J(\mathbf{H})$ is non-increasing under the update:

$$\mathbf{H}^{t+1} = \arg\min_{\mathbf{H}} F(\mathbf{H}^{t+1}, \mathbf{H}^{t}) \tag{5}$$

Proof: Set $\mathbf{H} \leftarrow \mathbf{H}^{t+1}$ and $\mathbf{H}' \leftarrow \mathbf{H}^{t}$. Then, $J(\mathbf{H}^t) = F(\mathbf{H}^t, \mathbf{H}^t) \geq F(\mathbf{H}^{t+1}, \mathbf{H}^t) \geq J(\mathbf{H}^{t+1})$.
In this work, we have $J(\mathbf{H}) = J_{data}(\mathbf{H}) = Tr(\mathbf{X}^\top\mathbf{X} - 2\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X} + \mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X})$.
Now consider the following function [1], [2]:

$$\begin{aligned} F(\mathbf{H}^{t+1}, \mathbf{H}^t) = \|\mathbf{X}\|^2 - 2\sum_{ijk}(H^t_{ji}(\mathbf{X}\mathbf{X}^\top)_{jk}H^t_{ki}(1+ \\ \log\frac{H^{t+1}_{ji}H^{t+1}_{ki}}{H^t_{ji}H^t_{ki}}) + \frac{1}{2}\sum_{ji}(\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{X}\mathbf{X}^\top\mathbf{H}^t + \\ \mathbf{X}\mathbf{X}^\top\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{H}^t)_{ji}\frac{H^{4,t+1}_{ji}}{H^{3,t}_{ji}} \end{aligned} \tag{6}$$

where $j, k = 1 \dots n$ and $i, l = 1 \dots c$, where $H^{x,t}_{ji}$ denotes the element (j,i) of the matrix $\mathbf{H}$ raised to the power of x at the t iteration. We will prove that $F(\mathbf{H}^{t+1}, \mathbf{H}^t)$ is an auxiliary function for J_{data} by proving properties 1 and 2. Firstly, note that $F(\mathbf{H}^t, \mathbf{H}^t) = J_{data}(\mathbf{H}^t)$ since the following two properties hold:

$$\begin{aligned} Tr(\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X}) &= Tr(\mathbf{H}^\top\mathbf{X}\mathbf{X}^\top\mathbf{H}) \\ &= \sum_{ijk} H_{ji}(\mathbf{X}\mathbf{X}^\top)_{jk}H_{ki} \ \forall\ \mathbf{H} \end{aligned} \tag{7}$$

$$\begin{aligned} \frac{1}{2}\sum_{ji}(\mathbf{H}\mathbf{H}^\top\mathbf{X}\mathbf{X}^\top\mathbf{H} + \mathbf{X}\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H})_{ji}H_{ji} \\ = Tr(\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X}) \ \forall\ \mathbf{H} \end{aligned} \tag{8}$$

The next step is to prove that the second term of $F(\mathbf{H}^{t+1}, \mathbf{H}^t)$ is larger than the one in $J_{data}(\mathbf{H}^{t+1})$, i.e. :

$$\begin{aligned} -2\sum_{ijk}(H^t_{ji}(\mathbf{X}\mathbf{X}^\top)_{jk}H^t_{ki}(1+\log\frac{H^{t+1}_{ji}H^{t+1}_{ki}}{H^t_{ji}H^t_{ki}}) \\ \geq Tr(-2\mathbf{X}^\top\mathbf{H}^{t+1}\mathbf{H}^{t+1\top}\mathbf{X}) \end{aligned} \tag{9}$$

where we set $z = \dfrac{H^{t+1}_{ji}H^{t+1}_{ki}}{H^t_{ji}H^t_{ki}}$ and use the inequality $z \geq 1 + \log(z)$ for $z > 0$. Then:

$$\begin{aligned} (\mathbf{X}\mathbf{X}^\top)_{jk}\frac{H^{t+1}_{ji}H^{t+1}_{ki}}{H^t_{ji}H^t_{ki}} \geq (\mathbf{X}\mathbf{X}^\top)_{jk}(1+\log\frac{H^{t+1}_{ji}H^{t+1}_{ki}}{H^t_{ji}H^t_{ki}}) \\ \Rightarrow (\mathbf{X}\mathbf{X}^\top)_{jk}H^{t+1}_{ji}H^{t+1}_{ki} \geq (\mathbf{X}\mathbf{X}^\top)_{jk}H^t_{ji}H^t_{ki}(1+\log\frac{H^{t+1}_{ji}H^{t+1}_{ki}}{H^t_{ji}H^t_{ki}}) \end{aligned}$$

and (9) is proved. For the third term, we have to show that:

$$\begin{aligned} \frac{1}{2}\sum_{ik}(\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{X}\mathbf{X}^\top\mathbf{H}^t + \mathbf{X}\mathbf{X}^\top\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{H}^t)_{ik}\frac{\mathbf{H}^{4,t+1}_{ik}}{\mathbf{H}^{3,t}_{ik}} \\ \geq Tr(\mathbf{X}^\top\mathbf{H}^{t+1}\mathbf{H}^{t+1\top}\mathbf{H}^{t+1}\mathbf{H}^{t+1\top}\mathbf{X}) \end{aligned} \tag{10}$$

where $i = 1\ldots n$ and $k = 1\ldots c$ by an appropriate change of indexes. For this part of the proof we will prove an extended version of the Lemma 4 in [1] by introducing a $n \times n$ matrix $\mathbf{Q}$. Consider the following expression:

$$\sum_{ik}\Big[\mathbf{H}'\mathbf{A}(\mathbf{Q}\mathbf{H}')^\top\mathbf{H}'\mathbf{B} + \mathbf{Q}\mathbf{H}'\mathbf{B}\mathbf{H}'^\top\mathbf{H}'\mathbf{A}\Big]_{ik}\frac{H^4_{ik}}{H'^3_{ik}} \tag{11}$$

where $\mathbf{H} = [H_{ik}]$ and $\mathbf{H}' = [\mathbf{H}'_{ik}]$ are $n \times c$ matrices. Now, let $\mathbf{U} = [u_{ik}] \in R^{n\times c}$ a matrix such that $H_{ik} = H'_{ik}u_{ik}$. Regarding the first term:

$$\begin{aligned} \sum_{ik}\Big[\mathbf{H}'\mathbf{A}(\mathbf{Q}\mathbf{H}')^\top\mathbf{H}'\mathbf{B}\Big]_{ik}\frac{H^4_{ik}}{H'^3_{ik}} \\ = \sum_{ijkrpq}H'_{jr}A_{rk}(\mathbf{Q}\mathbf{H}')_{ik}H'_{ip}B_{pq}H'_{jq}u^4_{jq} \end{aligned} \tag{12}$$

By setting $i \leftrightarrow j$, $k \leftrightarrow r$, $p \leftrightarrow q$:

$$\begin{aligned} \sum_{ik}\Big[\mathbf{H}'\mathbf{A}(\mathbf{Q}\mathbf{H}')^\top\mathbf{H}'\mathbf{B}\Big]_{ik}\frac{H^4_{ik}}{H'^3_{ik}} \\ = \sum_{ijkrpq}H'_{ik}A_{kr}(\mathbf{Q}\mathbf{H}')_{jr}H'_{jq}B_{qp}H'_{ip}u^4_{ip} \end{aligned} \tag{13}$$

Regarding the second term:

$$\sum_{ik}\Big[\mathbf{QH'BH'^{\top}H'A}\Big]_{ik}\frac{H_{ik}^4}{H_{ik}'^3} \\ = \sum_{ijkrpq}(\mathbf{QH'})_{ip}B_{pq}H'_{jq}H'_{jr}A_{rk}H'_{ik}u_{ik}^4 \tag{14}$$

By setting $i \leftrightarrow j$, $k \leftrightarrow q$, $p \leftrightarrow r$ and $i \leftrightarrow j$, $k \leftrightarrow r$, $p \leftrightarrow q$ we get respectively:

$$\sum_{ik}\Big[\mathbf{QH'BH'^{\top}H'A}\Big]_{ik}\frac{H_{ik}^4}{H_{ik}'^3} \\ = \sum_{ijkrpq}(QH')_{jr}B_{rk}H'_{ik}H'_{ip}A_{pq}H'_{jq}u_{jq}^4 \tag{15}$$

$$\sum_{ik}\Big[\mathbf{QH'BH'^{\top}H'A}\Big]_{ik}\frac{H_{ik}^4}{H_{ik}'^3} \\ = \sum_{ijkrpq}(\mathbf{QH'})_{jq}B_{qp}H'_{ip}H'_{ik}A_{kr}H'_{jr}u_{jr}^4 \tag{16}$$

We set $\mathbf{A} = \mathbf{B} = \mathbf{B}^{\top} = \mathbf{A}^{\top} = \mathbf{I} \in R^{c\times c}$ which implies that $A_{kr} = B_{kr} = B_{rk}$ and $B_{qp} = A_{qp} = A_{pq}$. Note that except for the u^4 terms we have:

1. $RHS_1 = RHS_2$
2. $RHS_3 = RHS_4 = RHS_5$
3. $RHS_2 = RHS_5$

By adding up (12), (13), (14) and (16) we get that:

$$\frac{1}{2}\sum_{ik}\Big[\mathbf{H'A(QH')^{\top}H'B + QH'BH'^{\top}H'A}\Big]_{ik}\frac{H_{ik}^4}{H_{ik}'^3} \\ = \sum_{ijkrpq}H'_{ip}B_{pq}(QH')_{jq}H'_{jr}A_{rk}H'_{ik}\frac{u_{ik}^4 + u_{jr}^4 + u_{jq}^4 + u_{ip}^4}{4}$$

We also write the trace of the third term in $J_{data}(\mathbf{H})$ as:

$$Tr(\mathbf{HH^{\top}QHH^{\top}}) = Tr(\mathbf{HAH^{\top}QHBH^{\top}}) \\ = \sum_{ijkrpq}H'_{ip}B_{pq}(\mathbf{QH'})_{jq}H'_{jr}A_{rk}H'_{ik}u_{ik}u_{jr}u_{jq}u_{ip} \tag{17}$$

Then it is straightforward to notice that:

$$\sum_{ijkrpq}H'_{ip}B_{pq}(\mathbf{QH'})_{jq}H'_{jr}A_{rk}H'_{ik}\frac{u_{ik}^4 + u_{jr}^4 + u_{jq}^4 + u_{ip}^4}{4} \\ \geq Tr(\mathbf{HAH^{\top}QHBH^{\top}}) \tag{18}$$

using the fact that $a^4+b^4+c^4+d^4 \geq 2(a^2b^2+c^2d^2) \geq 4abcd\ \forall a,b,c,d$. This concludes the proof for (10) when one sets $\mathbf{Q} \leftarrow \mathbf{X}\mathbf{X}^\top$, $\mathbf{H} \leftarrow \mathbf{H}^{t+1}$ and $\mathbf{H}' \leftarrow \mathbf{H}^t$. Therefore, $F(\mathbf{H}^{t+1}, \mathbf{H}^t)$ is an auxiliary function. Since we update $\mathbf{H}$ using (5), the minimum value is obtained by setting

$$\frac{\partial F(\mathbf{H}^{t+1}, \mathbf{H}^t)}{\partial H_{jk}^{t+1}} = 0 \tag{19}$$

For the derivative with respect to the second and the third term notice that:

$$\begin{aligned} &\frac{\partial}{\partial H_{jk}^{t+1}}\left[-2\sum_{ijk} H_{ji}^t(\mathbf{X}\mathbf{X}^\top)_{jk}H_{ki}^t\left(1+\log\frac{H_{ji}^{t+1}H_{ki}^{t+1}}{H_{ji}^tH_{ki}^t}\right)\right] \\ &\quad = -4(\mathbf{X}\mathbf{X}^\top\mathbf{H}^t)_{jk}\frac{H_{jk}^t}{H_{jk}^{t+1}} \end{aligned} \tag{20}$$

$$\begin{aligned} &\frac{\partial}{\partial H_{jk}^{t+1}}\left[\sum_{jk}(\mathbf{H}^t\mathbf{H}^{t\top}XX^\top\mathbf{H}^t + XX^\top\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{H}^t)_{jk}\frac{H_{jk}^{4,t+1}}{H_{jk}^{3,t}}\right] \\ &\quad = 4(\mathbf{H}^t\mathbf{H}^{t\top}XX^\top\mathbf{H}^t + XX^\top\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{H}^t)_{jk}\frac{H_{jk}^{3,t+1}}{H_{jk}^{3,t}} \end{aligned} \tag{21}$$

Since (20) and (21) hold $\forall j = 1 \ldots n$ and $k = 1 \ldots c$ then by combining (19), (20) and (21), we get the update equation:

$$H_{ik}^{t+1} = H_{ik}^t \sqrt[4]{\frac{2(\mathbf{X}\mathbf{X}^\top\mathbf{H}^t)_{ik}}{(\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{X}\mathbf{X}^\top\mathbf{H}^t + \mathbf{X}\mathbf{X}^\top\mathbf{H}^t\mathbf{H}^{t\top}\mathbf{H}^t)_{ik}}} \tag{22}$$

$\forall i = 1 \ldots n, k = 1 \ldots c$, which is the same as the update rule when no spatial term is used, i.e. $\lambda = 0$. When the spatial term is added, we can use the exact same procedure by taking into account $J(\mathbf{H}) = Tr(\mathbf{X}^\top\mathbf{X} - 2\mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X} + \mathbf{X}^\top\mathbf{H}\mathbf{H}^\top\mathbf{H}\mathbf{H}^\top\mathbf{X} + \lambda\mathbf{H}^\top\mathbf{L}\mathbf{H})$. Then, we can use the exact same auxiliary function $F(\mathbf{H}^{t+1}, \mathbf{H}^t)$ by adding the following term:

$$\lambda\sum_{i=1}^n\sum_{k=1}^c\frac{(\mathbf{L}\mathbf{H}^t)_{ik}H_{ik}^{2,t+1}}{H_{ik}^t} \tag{23}$$

Obviously, when $\mathbf{H}^{t+1} = \mathbf{H}^t$ this term is equal to:

$$\sum_{i=1}^n\sum_{k=1}^c\frac{(\mathbf{L}\mathbf{H}^t)_{ik}H_{ik}^{2,t}}{H_{ik}^t} = Tr(\mathbf{H}^{t\top}\mathbf{L}\mathbf{H}^t) \tag{24}$$

which is the corresponding term for $J(\mathbf{H}^t)$. And now we have to show that

$$\sum_{i=1}^n\sum_{k=1}^c\frac{(\mathbf{L}\mathbf{H}^t)_{ik}H_{ik}^{2,t+1}}{H_{ik}^t} \geq Tr(\mathbf{H}^{t+1\top}\mathbf{L}\mathbf{H}^{t+1}) \tag{25}$$

We use Proposition 5 from [3] where we set $\mathbf{A} \leftarrow \mathbf{L}$, $\mathbf{S}' \leftarrow \mathbf{H}^t$, $\mathbf{B} \leftarrow \mathbf{I} \in R^{c \times c}$ and $\mathbf{S} \leftarrow \mathbf{H}^{t+1}$ which proves (25). Also, note that:

$$\frac{\partial}{\partial H_{ik}^{t+1}} \left[\lambda \sum_{i=1}^{n} \sum_{k=1}^{c} \frac{(\mathbf{L}\mathbf{H}^t)_{ik} H_{ik}^{2,t+1}}{H_{ik}^t} \right] = 2\lambda \frac{(LH^t)_{ik} H_{ik}^{t+1}}{H_{ik}^t} \tag{26}$$

Finally, the update rule is derived as before and the correctness and the convergence of the update rule is proved.

References

[1] H. Wang, H. Huang, and C. Ding, "Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization," in *Proc. of the 20th ACM Int'l Conf. on Inform. and Know. Management*, ser. CIKM '11, New York, NY, USA, 2011, pp. 279–284. [Online]. Available: http://doi.acm.org/10.1145/2063576.2063621

[2] Y. Chen, M. Rege, M. Dong, and J. Hua, "Non-negative matrix factorization for semi-supervised data clustering," *Knowledge and Information Systems*, vol. 17, no. 3, pp. 355–379, 2008.

[3] C. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 1, pp. 45–55, Jan. 2010.