

Supplementary Material for Co-Regularized Deep Multi-Network Embedding

Jingchao Ni¹, Shiyu Chang², Xiao Liu³, Wei Cheng⁴, Haifeng Chen⁴, Dongkuan Xu¹, and Xiang Zhang¹

¹College of Information Sciences and Technology, Pennsylvania State University

²IBM T. J. Watson Research Center

³Department of Biomedical Engineering, Pennsylvania State University, ⁴NEC Laboratories America

¹{jzn47, dux19, xzhang}@ist.psu.edu, ²shiyu.chang@ibm.com

³xxl213@engr.psu.edu, ⁴{weicheng, haifeng}@nec-labs.com

Here we prove Theorem 1 and Theorem 2 using the Auxiliary Function approach [2]. First, we introduce the definition of an auxiliary function as below.

DEFINITION 1. [2] A function $Z(h, \tilde{h})$ is an auxiliary function for a given function $J(h)$ if the conditions $Z(h, \tilde{h}) \geq J(h)$ and $Z(h, h) = J(h)$ are satisfied.

1. PROOF OF THEOREM 1

In this section, we provide the detailed proof of Theorem 1. We first introduce a generic matrix inequality proposed in [1].

LEMMA 1. [1] For any matrices $\mathbf{A} \in \mathbb{R}_+^{k \times k}$, $\mathbf{B} \in \mathbb{R}_+^{k \times k}$, $\mathbf{V} \in \mathbb{R}_+^{n \times k}$, $\tilde{\mathbf{V}} \in \mathbb{R}_+^{n \times k}$, and \mathbf{A}, \mathbf{B} are symmetric, the following inequality holds

$$\text{Tr}(\mathbf{V}^T \mathbf{A} \mathbf{V} \mathbf{B}) \leq \sum_{xp} \frac{(\mathbf{A} \tilde{\mathbf{V}} \mathbf{B})_{xp} \mathbf{V}_{xp}^2}{\tilde{\mathbf{V}}_{xp}}$$

In the following, we use $\mathbf{U}^{(\tau)}$ ($1 \leq \tau \leq g$) to denote variable and $\tilde{\mathbf{U}}^{(\tau)}$ to denote current estimate (i.e., a constant).

PROOF. Using ED loss $\mathcal{L}_{ed}^{(ij)}$ in Eq. (12)¹, the sum of all terms in Eq. (12) that contains $\mathbf{U}^{(\tau)}$ is

$$\begin{aligned} \mathcal{L}_{ed}(\mathbf{U}^{(\tau)}) &= \alpha \sum_{(\tau, j) \in \mathcal{I}} \|\mathbf{O}^{(\tau j)} \mathbf{U}^{(\tau)} - \tilde{\mathbf{S}}^{(\tau j)} \mathbf{U}^{(j)}\|_F^2 \\ &\quad + \beta \|\mathbf{U}^{(\tau)} - \mathbf{H}^{(\tau)}\|_F^2 \end{aligned}$$

which is equivalent to (after removing some constants)

$$\begin{aligned} \mathcal{L}_{ed}(\mathbf{U}^{(\tau)}) &= \alpha \sum_{(\tau, j) \in \mathcal{I}} \text{Tr} \left((\mathbf{O}^{(\tau j)} \mathbf{U}^{(\tau)})^T \mathbf{O}^{(\tau j)} \mathbf{U}^{(\tau)} - 2(\tilde{\mathbf{S}}^{(\tau j)} \mathbf{U}^{(j)})^T \mathbf{O}^{(\tau j)} \mathbf{U}^{(\tau)} \right) \\ &\quad + \alpha \sum_{(j, \tau) \in \mathcal{I}} \text{Tr} \left((\tilde{\mathbf{S}}^{(j\tau)} \mathbf{U}^{(\tau)})^T \tilde{\mathbf{S}}^{(j\tau)} \mathbf{U}^{(\tau)} - 2(\mathbf{O}^{(j\tau)} \mathbf{U}^{(j)})^T \tilde{\mathbf{S}}^{(j\tau)} \mathbf{U}^{(\tau)} \right) \\ &\quad + \beta \text{Tr} \left((\mathbf{U}^{(\tau)})^T \mathbf{U}^{(\tau)} - 2(\mathbf{H}^{(\tau)})^T \mathbf{U}^{(\tau)} \right) \end{aligned} \quad (17)$$

where $\text{Tr}(\cdot)$ is the trace function.

First, we derive a tight upper bound for $\mathcal{L}_{ed}(\mathbf{U}^{(\tau)})$. Using the inequality introduced in Lemma 1, substituting \mathbf{A} with

$(\mathbf{O}^{(\tau j)})^T \mathbf{O}^{(\tau j)}$, \mathbf{B} with an identity matrix, and \mathbf{V} with $\mathbf{U}^{(\tau)}$, we can obtain an upper bound for the first term

$$\alpha \text{Tr} \left((\mathbf{U}^{(\tau)})^T (\mathbf{O}^{(\tau j)})^T \mathbf{O}^{(\tau j)} \mathbf{U}^{(\tau)} \right) \leq \sum_{xp} (\Phi_{xp}^{(\tau j)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^2}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \quad (18)$$

where

$$(\Phi^{(\tau j)})' = \alpha (\mathbf{O}^{(\tau j)})^T \mathbf{O}^{(\tau j)} \tilde{\mathbf{U}}^{(\tau)}$$

Similarly, substituting \mathbf{A} with $(\tilde{\mathbf{S}}^{(j\tau)})^T \tilde{\mathbf{S}}^{(j\tau)}$, \mathbf{B} with an identity matrix, and \mathbf{V} with $\mathbf{U}^{(\tau)}$, we obtain an upper bound for the third term

$$\alpha \text{Tr} \left((\mathbf{U}^{(\tau)})^T (\tilde{\mathbf{S}}^{(j\tau)})^T \tilde{\mathbf{S}}^{(j\tau)} \mathbf{U}^{(\tau)} \right) \leq \sum_{xp} (\Pi_{xp}^{(j\tau)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^2}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \quad (19)$$

where

$$(\Pi^{(j\tau)})' = \alpha (\tilde{\mathbf{S}}^{(j\tau)})^T \tilde{\mathbf{S}}^{(j\tau)} \tilde{\mathbf{U}}^{(\tau)}$$

Using the inequality $z > 1 + \log z$, which holds for $z > 0$, we have

$$\frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \geq 1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \quad (20)$$

Then we can derive an upper bound for the second term

$$\begin{aligned} -2\alpha \text{Tr} \left((\tilde{\mathbf{S}}^{(\tau j)} \mathbf{U}^{(j)})^T \mathbf{O}^{(\tau j)} \mathbf{U}^{(\tau)} \right) &= -2 \sum_{xp} (\Theta_{xp}^{(\tau j)})' \mathbf{U}_{xp}^{(\tau)} \\ &\leq -2 \sum_{xp} (\Theta_{xp}^{(\tau j)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} (1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}}) \end{aligned} \quad (21)$$

where

$$(\Theta^{(\tau j)})' = \alpha (\mathbf{O}^{(\tau j)})^T \tilde{\mathbf{S}}^{(\tau j)} \mathbf{U}^{(j)}$$

Similarly, we derive an upper bound for the fourth term

$$\begin{aligned} -2\alpha \text{Tr} \left((\mathbf{O}^{(j\tau)} \mathbf{U}^{(j)})^T \tilde{\mathbf{S}}^{(j\tau)} \mathbf{U}^{(\tau)} \right) &= -2 \sum_{xp} (\Lambda_{xp}^{(j\tau)})' \mathbf{U}_{xp}^{(\tau)} \\ &\leq -2 \sum_{xp} (\Lambda_{xp}^{(j\tau)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} (1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}}) \end{aligned} \quad (22)$$

where

$$(\Lambda^{(j\tau)})' = \alpha (\tilde{\mathbf{S}}^{(j\tau)})^T \mathbf{O}^{(j\tau)} \mathbf{U}^{(j)}$$

and also for the last term, we have

$$\begin{aligned} -2\beta \text{Tr} \left((\mathbf{H}^{(\tau)})^T \mathbf{U}^{(\tau)} \right) &= -2\beta \sum_{xp} \mathbf{H}_{xp}^{(\tau)} \mathbf{U}_{xp}^{(\tau)} \\ &\leq -2\beta \sum_{xp} \mathbf{H}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{xp}^{(\tau)} (1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}}) \end{aligned} \quad (23)$$

¹The equation numbers in this supplementary material are consistent with those in the original paper.

The fifth term it is equivalent to

$$\beta \text{Tr} \left((\mathbf{U}^{(\tau)})^T \mathbf{U}^{(\tau)} \right) = \beta \sum_{xp} (\mathbf{U}_{xp}^{(\tau)})^2 \quad (24)$$

Therefore, with the inequalities Eq. (18), (19), (21), (22), (23) and the equation Eq. (24), we can formulate an upper bound for $\mathcal{L}_{ed}(\mathbf{U}^{(\tau)})$ in Eq. (17) as below.

$$\begin{aligned} Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)}) &= \sum_{(\tau,j) \in \mathcal{I}} \left(\sum_{xp} [(\Phi_{xp}^{(\tau,j)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^2}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} - 2(\Theta_{xp}^{(\tau,j)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} (1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}})] \right) \\ &+ \sum_{(j,\tau) \in \mathcal{I}} \left(\sum_{xp} [(\Pi_{xp}^{(j,\tau)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^2}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} - 2(\Lambda_{xp}^{(j,\tau)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} (1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}})] \right) \\ &+ \left(\sum_{xp} [\beta (\mathbf{U}_{xp}^{(\tau)})^2 - 2\beta \mathbf{H}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{xp}^{(\tau)} (1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}})] \right) \end{aligned}$$

We can verify that $Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ is a tight upper bound of $\mathcal{L}_{ed}(\mathbf{U}^{(\tau)})$. That is, it satisfies

$$Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)}) \geq \mathcal{L}_{ed}(\mathbf{U}^{(\tau)}), \quad Z_{ed}(\mathbf{U}^{(\tau)}, \mathbf{U}^{(\tau)}) = \mathcal{L}_{ed}(\mathbf{U}^{(\tau)}) \quad (25)$$

Thus, according to Definition 1, $Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ is an auxiliary function of $\mathcal{L}_{ed}(\mathbf{U}^{(\tau)})$.

Next, we derive the minimal solution to $Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$. The gradient of $Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ w.r.t. $\mathbf{U}^{(\tau)}$ is

$$\begin{aligned} \frac{\partial Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})}{\partial \mathbf{U}_{xp}^{(\tau)}} &= \sum_{(\tau,j) \in \mathcal{I}} \left(2(\Phi_{xp}^{(\tau,j)})' \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} - 2(\Theta_{xp}^{(\tau,j)})' \frac{\tilde{\mathbf{U}}_{xp}^{(\tau)}}{\mathbf{U}_{xp}^{(\tau)}} \right) \\ &+ \sum_{(j,\tau) \in \mathcal{I}} \left(2(\Pi_{xp}^{(j,\tau)})' \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} - 2(\Lambda_{xp}^{(j,\tau)})' \frac{\tilde{\mathbf{U}}_{xp}^{(\tau)}}{\mathbf{U}_{xp}^{(\tau)}} \right) \\ &+ \left(2\beta \mathbf{U}_{xp}^{(\tau)} - 2\beta \mathbf{H}_{xp}^{(\tau)} \frac{\tilde{\mathbf{U}}_{xp}^{(\tau)}}{\mathbf{U}_{xp}^{(\tau)}} \right) \end{aligned}$$

Then the global minimum of $Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ can be obtained by solving $\frac{\partial Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})}{\partial \mathbf{U}_{xp}^{(\tau)}} = 0$. This gives

$$\mathbf{U}_{xp}^{(\tau)} = \mathbf{U}_{xp}^{(\tau)} \left(\frac{\sum_{(\tau,j) \in \mathcal{I}} (\Phi_{xp}^{(\tau,j)})' + \sum_{(j,\tau) \in \mathcal{I}} (\Lambda_{xp}^{(j,\tau)})' + \beta \mathbf{H}_{xp}^{(\tau)}}{\sum_{(\tau,j) \in \mathcal{I}} (\Phi_{xp}^{(\tau,j)})' + \sum_{(j,\tau) \in \mathcal{I}} (\Pi_{xp}^{(j,\tau)})' + \beta \mathbf{U}_{xp}^{(\tau)}} \right)^{\frac{1}{2}}$$

which is consistent with the updating rule in Eq. (13).

Therefore, according to the properties of auxiliary function, i.e., Eq. (25), at any iteration $\kappa \geq 1$ when updating $\mathbf{U}^{(\tau)}$ using Eq. (13), we have

$$\begin{aligned} \mathcal{L}_{ed}((\mathbf{U}^{(\tau)})^{(\kappa)}) &= Z_{ed}((\mathbf{U}^{(\tau)})^{(\kappa)}, (\mathbf{U}^{(\tau)})^{(\kappa)}) \\ &\geq Z_{ed}((\mathbf{U}^{(\tau)})^{(\kappa+1)}, (\mathbf{U}^{(\tau)})^{(\kappa)}) \geq \mathcal{L}_{ed}((\mathbf{U}^{(\tau)})^{(\kappa+1)}) \end{aligned}$$

where $(\mathbf{U}^{(\tau)})^{(\kappa)}$ denotes the updated $\mathbf{U}^{(\tau)}$ at the κ -th iteration. Therefore, using ED loss, the objective value in Eq. (12) monotonically decreases. This completes the proof of Theorem 1. \square

2. PROOF OF THEOREM 2

In this section, we provide the detailed proof of Theorem 2. The idea is similar to the proof of Theorem 1. We first introduce a generic matrix inequality proposed in [3].

LEMMA 2. [3] *For any matrices $\mathbf{A} \in \mathbb{R}_+^{k \times k}$, $\mathbf{B} \in \mathbb{R}_+^{k \times k}$, $\mathbf{V} \in \mathbb{R}_+^{n \times k}$, $\tilde{\mathbf{V}} \in \mathbb{R}_+^{n \times k}$, and \mathbf{A}, \mathbf{B} are symmetric, the following inequality holds*

$$\text{Tr}(\mathbf{V} \mathbf{A} \mathbf{V}^T \mathbf{V} \mathbf{B} \mathbf{V}^T) \leq \sum_{xp} \left(\frac{\tilde{\mathbf{V}} \mathbf{A} \tilde{\mathbf{V}}^T \tilde{\mathbf{V}} \mathbf{B} + \tilde{\mathbf{V}} \mathbf{B} \tilde{\mathbf{V}}^T \tilde{\mathbf{V}} \mathbf{A}}{2} \right) \frac{\mathbf{V}_{xp}^4}{\tilde{\mathbf{V}}_{xp}^3}$$

In the following, we use $\mathbf{U}^{(\tau)}$ ($1 \leq \tau \leq g$) to denote variable and $\tilde{\mathbf{U}}^{(\tau)}$ to denote current estimate (i.e., a constant).

PROOF. Using PD loss $\mathcal{L}_{pd}^{(ij)}$ in Eq. (12), the sum of all terms in Eq. (12) that contains $\mathbf{U}^{(\tau)}$ is

$$\begin{aligned} \mathcal{L}_{pd}(\mathbf{U}^{(\tau)}) &= \alpha \sum_{(\tau,j) \in \mathcal{I}} \|(\mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)})(\mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)})^T - (\tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)})(\tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)})^T\|_F^2 \\ &+ \beta \|\mathbf{U}^{(\tau)} - \mathbf{H}^{(\tau)}\|_F^2 \end{aligned}$$

which is equivalent to (after removing some constants)

$$\begin{aligned} \mathcal{J}_{pd}(\mathbf{U}^{(\tau)}) &= \alpha \sum_{(\tau,j) \in \mathcal{I}} \text{Tr} \left(\mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)} (\mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)})^T \mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)} (\mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)})^T \right. \\ &\quad \left. - 2\tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)} (\tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)})^T \mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)} (\mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)})^T \right) \\ &+ \alpha \sum_{(j,\tau) \in \mathcal{I}} \text{Tr} \left(\tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)} (\tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)})^T \tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)} (\tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)})^T \right. \\ &\quad \left. - 2\mathbf{O}^{(j,\tau)} \mathbf{U}^{(j)} (\mathbf{O}^{(j,\tau)} \mathbf{U}^{(j)})^T \tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)} (\tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)})^T \right) \\ &+ \beta \text{Tr} \left((\mathbf{U}^{(\tau)})^T \mathbf{U}^{(\tau)} - 2(\mathbf{H}^{(\tau)})^T \mathbf{U}^{(\tau)} \right) \end{aligned} \quad (26)$$

Next, we derive a tight upper bound for $\mathcal{L}_{pd}(\mathbf{U}^{(\tau)})$. Using the inequality introduced in Lemma 2, substituting both \mathbf{A} and \mathbf{B} with $(\mathbf{O}^{(\tau,j)})^T \mathbf{O}^{(\tau,j)}$, and \mathbf{V} with $(\mathbf{U}^{(\tau)})^T$, we can obtain an upper bound for the first term

$$\begin{aligned} \alpha \text{Tr} \left((\mathbf{U}^{(\tau)})^T (\mathbf{O}^{(\tau,j)})^T \mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)} (\mathbf{U}^{(\tau)})^T (\mathbf{O}^{(\tau,j)})^T \mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)} \right) \\ \leq \sum_{xp} (\Phi_{xp}^{(\tau,j)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^4}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^3} \end{aligned} \quad (27)$$

where

$$(\Phi^{(\tau,j)})' = \alpha (\mathbf{O}^{(\tau,j)})^T \mathbf{O}^{(\tau,j)} \tilde{\mathbf{U}}^{(\tau)} (\tilde{\mathbf{U}}^{(\tau)})^T (\mathbf{O}^{(\tau,j)})^T \mathbf{O}^{(\tau,j)} \tilde{\mathbf{U}}^{(\tau)}$$

Similarly, substituting both \mathbf{A} and \mathbf{B} with $(\tilde{\mathbf{S}}^{(j,\tau)})^T \tilde{\mathbf{S}}^{(j,\tau)}$, and \mathbf{V} with $(\mathbf{U}^{(\tau)})^T$, we can obtain an upper bound for the third term

$$\begin{aligned} \alpha \text{Tr} \left((\mathbf{U}^{(\tau)})^T (\tilde{\mathbf{S}}^{(j,\tau)})^T \tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)} (\mathbf{U}^{(\tau)})^T (\tilde{\mathbf{S}}^{(j,\tau)})^T \tilde{\mathbf{S}}^{(j,\tau)} \mathbf{U}^{(\tau)} \right) \\ \leq \sum_{xp} (\tilde{\Pi}_{xp}^{(j,\tau)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^4}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^3} \end{aligned} \quad (28)$$

where

$$(\tilde{\Pi}^{(j,\tau)})' = \alpha (\tilde{\mathbf{S}}^{(j,\tau)})^T \tilde{\mathbf{S}}^{(j,\tau)} \tilde{\mathbf{U}}^{(\tau)} (\tilde{\mathbf{U}}^{(\tau)})^T (\tilde{\mathbf{S}}^{(j,\tau)})^T \tilde{\mathbf{S}}^{(j,\tau)} \tilde{\mathbf{U}}^{(\tau)}$$

Using the inequality $z > 1 + \log z$, which holds for $z > 0$, we have

$$\frac{\mathbf{U}_{xp}^{(\tau)} \mathbf{U}_{yp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)}} \geq 1 + \log \frac{\mathbf{U}_{xp}^{(\tau)} \mathbf{U}_{yp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)}}$$

Then we can derive an upper bound for the second term

$$\begin{aligned} -2\alpha \text{Tr} \left((\mathbf{U}^{(\tau)})^T (\mathbf{O}^{(\tau,j)})^T \tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)} (\tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)})^T \mathbf{O}^{(\tau,j)} \mathbf{U}^{(\tau)} \right) \\ \leq -2 \sum_{xyp} (\hat{\mathbf{r}}_{xy}^{(\tau,j)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)} \left(1 + \log \frac{\mathbf{U}_{xp}^{(\tau)} \mathbf{U}_{yp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)}} \right) \end{aligned} \quad (29)$$

where

$$(\hat{\mathbf{r}}^{(\tau,j)})' = \alpha (\mathbf{O}^{(\tau,j)})^T \tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)} (\tilde{\mathbf{S}}^{(\tau,j)} \mathbf{U}^{(j)})^T \mathbf{O}^{(\tau,j)}$$

Similarly, we derive an upper bound for the fourth term

$$\begin{aligned} & -2\alpha \text{Tr} \left((\mathbf{U}^{(\tau)})^T (\tilde{\mathbf{S}}^{(j\tau)})^T \mathbf{O}^{(j\tau)} \mathbf{U}^{(j)} (\mathbf{O}^{(j\tau)} \mathbf{U}^{(j)})^T \tilde{\mathbf{S}}^{(j\tau)} \mathbf{U}^{(\tau)} \right) \\ & \leq -2 \sum_{xy} (\hat{\mathbf{\Omega}}_{xy}^{(j\tau)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)} \left(1 + \log \frac{\mathbf{U}_{xp}^{(\tau)} \mathbf{U}_{yp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)}} \right) \end{aligned} \quad (30)$$

where

$$(\hat{\mathbf{\Omega}}^{(j\tau)})' = \alpha (\tilde{\mathbf{S}}^{(j\tau)})^T \mathbf{O}^{(j\tau)} \mathbf{U}^{(j)} (\mathbf{O}^{(j\tau)} \mathbf{U}^{(j)})^T \tilde{\mathbf{S}}^{(j\tau)}$$

Using the inequality $2ab \leq a^2 + b^2$, we can derive an upper bound for the fifth term

$$\beta \text{Tr} \left((\mathbf{U}^{(\tau)})^T \mathbf{U}^{(\tau)} \right) = \beta \sum_{xp} (\mathbf{U}_{xp}^{(\tau)})^2 \leq \beta \sum_{xp} \frac{(\mathbf{U}_{xp}^{(\tau)})^4 + (\tilde{\mathbf{U}}_{xp}^{(\tau)})^4}{2(\tilde{\mathbf{U}}_{xp}^{(\tau)})^2} \quad (31)$$

Using the inequality in Eq. (20), we obtain an upper bound for the last term

$$-2\beta \text{Tr} \left((\mathbf{H}^{(\tau)})^T \mathbf{U}^{(\tau)} \right) \leq -2\beta \sum_{xp} \mathbf{H}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{xp}^{(\tau)} \left(1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \right) \quad (32)$$

Therefore, with the inequalities Eq. (27), (28), (29), (30), (31) and (32), we can formulate an upper bound for $\mathcal{L}_{pd}(\mathbf{U}^{(\tau)})$ in Eq. (26) as below.

$$\begin{aligned} Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)}) &= \sum_{(\tau,j) \in \mathcal{I}} \left(\sum_{xp} (\hat{\mathbf{\Phi}}_{xp}^{(\tau j)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^4}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^3} \right. \\ & - 2 \sum_{xy} (\hat{\mathbf{F}}_{xy}^{(\tau j)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)} \left(1 + \log \frac{\mathbf{U}_{xp}^{(\tau)} \mathbf{U}_{yp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)}} \right) \\ & + \sum_{(j,\tau) \in \mathcal{I}} \left(\sum_{xp} (\hat{\mathbf{\Pi}}_{xp}^{(j\tau)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^4}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^3} \right. \\ & - 2 \sum_{xy} (\hat{\mathbf{\Omega}}_{xy}^{(j\tau)})' \tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)} \left(1 + \log \frac{\mathbf{U}_{xp}^{(\tau)} \mathbf{U}_{yp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{yp}^{(\tau)}} \right) \\ & \left. + \sum_{xp} \left(\beta \frac{(\mathbf{U}_{xp}^{(\tau)})^4 + (\tilde{\mathbf{U}}_{xp}^{(\tau)})^4}{2(\tilde{\mathbf{U}}_{xp}^{(\tau)})^2} - 2\beta \mathbf{H}_{xp}^{(\tau)} \tilde{\mathbf{U}}_{xp}^{(\tau)} \left(1 + \log \frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \right) \right) \right) \end{aligned}$$

We can verify that $Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ is a tight upper bound of $\mathcal{L}_{pd}(\mathbf{U}^{(\tau)})$. That is, it satisfies

$$Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)}) \geq \mathcal{L}_{pd}(\mathbf{U}^{(\tau)}), \quad Z_{pd}(\mathbf{U}^{(\tau)}, \mathbf{U}^{(\tau)}) = \mathcal{L}_{pd}(\mathbf{U}^{(\tau)}) \quad (33)$$

Thus, according to Definition 1, $Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ is an auxiliary function of $\mathcal{L}_{pd}(\mathbf{U}^{(\tau)})$.

Next, we derive the minimal solution to $Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$. The gradient of $Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ w.r.t. $\mathbf{U}^{(\tau)}$ is

$$\begin{aligned} & \frac{\partial Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})}{\partial \mathbf{U}_{xp}^{(\tau)}} \\ &= \sum_{(\tau,j) \in \mathcal{I}} \left(4(\hat{\mathbf{\Phi}}_{xp}^{(\tau j)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^3}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^3} - 4(\hat{\mathbf{\Theta}}_{xp}^{(\tau j)})' \frac{\tilde{\mathbf{U}}_{xp}^{(\tau)}}{\mathbf{U}_{xp}^{(\tau)}} \right) \\ &+ \sum_{(j,\tau) \in \mathcal{I}} \left(4(\hat{\mathbf{\Pi}}_{xp}^{(j\tau)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^3}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^3} - 4(\hat{\mathbf{\Lambda}}_{xp}^{(j\tau)})' \frac{\tilde{\mathbf{U}}_{xp}^{(\tau)}}{\mathbf{U}_{xp}^{(\tau)}} \right) \\ &+ (2\beta \frac{(\mathbf{U}_{xp}^{(\tau)})^3}{(\tilde{\mathbf{U}}_{xp}^{(\tau)})^2} - 2\beta \mathbf{H}_{xp}^{(\tau)} \frac{\tilde{\mathbf{U}}_{xp}^{(\tau)}}{\mathbf{U}_{xp}^{(\tau)}}) \end{aligned}$$

where

$$(\hat{\mathbf{\Theta}}^{(\tau j)})' = (\hat{\mathbf{F}}^{(\tau j)})' \tilde{\mathbf{U}}^{(\tau)}, \quad (\hat{\mathbf{\Lambda}}^{(j\tau)})' = (\hat{\mathbf{\Omega}}^{(j\tau)})' \tilde{\mathbf{U}}^{(\tau)}$$

Then the global minimum of $Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})$ can be ob-

tained by solving $\frac{\partial Z_{pd}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)})}{\partial \mathbf{U}_{xp}^{(\tau)}} = 0$. This gives

$$\mathbf{U}_{xp}^{(\tau)} = \mathbf{U}_{xp}^{(\tau)} \left(\frac{\sum_{(\tau,j) \in \mathcal{I}} 2(\hat{\mathbf{\Theta}}_{xp}^{(\tau j)})' + \sum_{(j,\tau) \in \mathcal{I}} 2(\hat{\mathbf{\Lambda}}_{xp}^{(j\tau)})' + \beta \mathbf{H}_{xp}^{(\tau)}}{\sum_{(\tau,j) \in \mathcal{I}} 2(\hat{\mathbf{\Phi}}_{xp}^{(\tau j)})' + \sum_{(j,\tau) \in \mathcal{I}} 2(\hat{\mathbf{\Pi}}_{xp}^{(j\tau)})' + \beta \tilde{\mathbf{U}}_{xp}^{(\tau)}} \right)^{\frac{1}{4}}$$

which is consistent with the updating rule in Eq. (14). Note the factor 2 before $(\hat{\mathbf{\Theta}}^{(\tau j)})'$, $(\hat{\mathbf{\Phi}}^{(\tau j)})'$, $(\hat{\mathbf{\Lambda}}^{(j\tau)})'$, $(\hat{\mathbf{\Pi}}^{(j\tau)})'$ in the above equation has been incorporated into $\hat{\mathbf{\Theta}}^{(\tau j)}$, $\hat{\mathbf{\Phi}}^{(\tau j)}$, $\hat{\mathbf{\Lambda}}^{(j\tau)}$, $\hat{\mathbf{\Pi}}^{(j\tau)}$ in Eq. (14).

Therefore, according to the properties of auxiliary function, i.e., Eq. (33), at any iteration $\kappa \geq 1$ when updating $\mathbf{U}^{(\tau)}$ using Eq. (14), we have

$$\begin{aligned} \mathcal{L}_{pd}((\mathbf{U}^{(\tau)})^{(\kappa)}) &= Z_{pd}((\mathbf{U}^{(\tau)})^{(\kappa)}, (\mathbf{U}^{(\tau)})^{(\kappa)}) \\ &\geq Z_{pd}((\mathbf{U}^{(\tau)})^{(\kappa+1)}, (\mathbf{U}^{(\tau)})^{(\kappa)}) \geq \mathcal{L}_{pd}((\mathbf{U}^{(\tau)})^{(\kappa+1)}) \end{aligned}$$

which indicates that, using PD loss, the objective value in Eq. (12) monotonically decreases. This completes the proof of Theorem 2. \square

3. REFERENCES

- [1] C. H. Ding, T. Li, and M. I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, 2010.
- [2] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.
- [3] H. Wang, H. Huang, and C. Ding. Simultaneous clustering of multi-type relational data via symmetric nonnegative matrix tri-factorization. In *CIKM*, 2011.