Supplementary Material for Co-Regularized Deep Multi-Network Embedding

Jingchao Ni[1], Shiyu Chang[2], Xiao Liu[3], Wei Cheng[4], Haifeng Chen[4], Dongkuan Xu[1], and Xiang Zhang[1]
[1]College of Information Sciences and Technology, Pennsylvania State University
[2]IBM T. J. Watson Research Center
[3]Department of Biomedical Engineering, Pennsylvania State University, [4]NEC Laboratories America
[1]{jzn47, dux19, xzhang}@ist.psu.edu, [2]shiyu.chang@ibm.com
[3]xxl213@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*

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

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 $\mathscr{L}_{ed}^{(ij)}$ in Eq. (12)[1], the sum of all terms in Eq. (12) that contains $\mathbf{U}^{(\tau)}$ is

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

which is equivalent to (after removing some constants)

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

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

First, we derive a tight upper bound for $\mathscr{L}_{ed}(\mathbf{U}^{(\tau)})$. 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

$$\alpha\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T(\mathbf{O}^{(\tau j)})^T\mathbf{O}^{(\tau j)}\mathbf{U}^{(\tau)}\Big) \leq \sum_{xp}(\mathbf{\Phi}_{xp}^{(\tau j)})' \frac{(\mathbf{U}_{xp}^{(\tau)})^2}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \tag{18}$$

where

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

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

$$\alpha\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T(\tilde{\mathbf{S}}^{(j\tau)})^T\tilde{\mathbf{S}}^{(j\tau)}\mathbf{U}^{(\tau)}\Big) \leq \sum_{xp}(\mathbf{\Pi}_{xp}^{(j\tau)})' \frac{(\mathbf{U}_{xp}^{(i)})^2}{\tilde{\mathbf{U}}_{xp}^{(\tau)}} \tag{19}$$

where

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

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

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

Then we can derive an upper bound for the second term

$$\begin{aligned}&-2\alpha\mathrm{Tr}\Big((\tilde{\mathbf{S}}^{(\tau j)}\mathbf{U}^{(j)})^T\mathbf{O}^{(\tau j)}\mathbf{U}^{(\tau)}\Big) = -2\sum_{xp}(\mathbf{\Theta}_{xp}^{(\tau j)})'\mathbf{U}_{xp}^{(\tau)} \\ &\leq -2\sum_{xp}(\mathbf{\Theta}_{xp}^{(\tau j)})'\tilde{\mathbf{U}}_{xp}^{(\tau)}(1 + \log\frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}})\end{aligned} \tag{21}$$

where

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

Similarly, we derive an upper bound for the fourth term

$$\begin{aligned}&-2\alpha\mathrm{Tr}\Big((\mathbf{O}^{(j\tau)}\mathbf{U}^{(j)})^T\tilde{\mathbf{S}}^{(j\tau)}\mathbf{U}^{(\tau)}\Big) = -2\sum_{xp}(\mathbf{\Lambda}_{xp}^{(j\tau)})'\mathbf{U}_{xp}^{(\tau)} \\ &\leq -2\sum_{xp}(\mathbf{\Lambda}_{xp}^{(j\tau)})'\tilde{\mathbf{U}}_{xp}^{(\tau)}(1 + \log\frac{\mathbf{U}_{xp}^{(\tau)}}{\tilde{\mathbf{U}}_{xp}^{(\tau)}})\end{aligned} \tag{22}$$

where

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

and also for the last term, we have

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

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

The fifth term it is equivalent to

$$\beta\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T\mathbf{U}^{(\tau)}\Big) = \beta\sum_{xp}(\mathbf{U}^{(\tau)}_{xp})^2 \tag{24}$$

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

$$\begin{aligned}
&Z_{ed}(\mathbf{U}^{(\tau)}, \tilde{\mathbf{U}}^{(\tau)}) \\
&= \sum_{(\tau,j)\in\mathcal{I}}\Big(\sum_{xp}[(\mathbf{\Phi}^{(\tau j)}_{xp})'\frac{(\mathbf{U}^{(\tau)}_{xp})^2}{\tilde{\mathbf{U}}^{(\tau)}_{xp}} - 2(\mathbf{\Theta}^{(\tau j)}_{xp})'\tilde{\mathbf{U}}^{(\tau)}_{xp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}})]\Big) \\
&+ \sum_{(j,\tau)\in\mathcal{I}}\Big(\sum_{xp}[(\mathbf{\Pi}^{(j\tau)}_{xp})'\frac{(\mathbf{U}^{(\tau)}_{xp})^2}{\tilde{\mathbf{U}}^{(\tau)}_{xp}} - 2(\mathbf{\Lambda}^{(j\tau)}_{xp})'\tilde{\mathbf{U}}^{(\tau)}_{xp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}})]\Big) \\
&+ \Big(\sum_{xp}[\beta(\mathbf{U}^{(\tau)}_{xp})^2 - 2\beta\mathbf{H}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{xp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}})]\Big)
\end{aligned}$$

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

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

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

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

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

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

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

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

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

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*

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

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

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

which is equivalent to (after removing some constants)

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

Next, we derive a tight upper bound for $\mathscr{L}_{pd}(\mathbf{U}^{(\tau)})$. 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

$$\begin{aligned}
&\alpha\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T(\mathbf{O}^{(\tau j)})^T\mathbf{O}^{(\tau j)}\mathbf{U}^{(\tau)}(\mathbf{U}^{(\tau)})^T(\mathbf{O}^{(\tau j)})^T\mathbf{O}^{(\tau j)}\mathbf{U}^{(\tau)}\Big) \\
&\leq \sum_{xp}(\hat{\mathbf{\Phi}}^{(\tau j)}_{xp})'\frac{(\mathbf{U}^{(\tau)}_{xp})^4}{(\tilde{\mathbf{U}}^{(\tau)}_{xp})^3}
\end{aligned} \tag{27}$$

where

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

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

$$\begin{aligned}
&\alpha\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T(\tilde{\mathbf{S}}^{(j\tau)})^T\tilde{\mathbf{S}}^{(j\tau)}\mathbf{U}^{(\tau)}(\mathbf{U}^{(\tau)})^T(\tilde{\mathbf{S}}^{(j\tau)})^T\tilde{\mathbf{S}}^{(j\tau)}\mathbf{U}^{(\tau)}\Big) \\
&\leq \sum_{xp}(\hat{\mathbf{\Pi}}^{(j\tau)}_{xp})'\frac{(\mathbf{U}^{(\tau)}_{xp})^4}{(\tilde{\mathbf{U}}^{(\tau)}_{xp})^3}
\end{aligned} \tag{28}$$

where

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

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

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

Then we can derive an upper bound for the second term

$$\begin{aligned}
&-2\alpha\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T(\mathbf{O}^{(\tau j)})^T\tilde{\mathbf{S}}^{(\tau j)}\mathbf{U}^{(j)}(\tilde{\mathbf{S}}^{(\tau j)}\mathbf{U}^{(j)})^T\mathbf{O}^{(\tau j)}\mathbf{U}^{(\tau)}\Big) \\
&\leq -2\sum_{xyp}(\hat{\mathbf{\Gamma}}^{(\tau j)}_{xy})'\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}\Big(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}\mathbf{U}^{(\tau)}_{yp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}}\Big)
\end{aligned} \tag{29}$$

where

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

Similarly, we derive an upper bound for the fourth term

$$-2\alpha\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T(\tilde{\mathbf{S}}^{(j\tau)})^T\mathbf{O}^{(j\tau)}\mathbf{U}^{(j)}(\mathbf{O}^{(j\tau)}\mathbf{U}^{(j)})^T\tilde{\mathbf{S}}^{(j\tau)}\mathbf{U}^{(\tau)}\Big)$$
$$\le -2\sum_{xyp}(\hat{\mathbf{\Omega}}^{(j\tau)}_{xy})'\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}\Big(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}\mathbf{U}^{(\tau)}_{yp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}}\Big) \tag{30}$$

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

$$\beta\mathrm{Tr}\Big((\mathbf{U}^{(\tau)})^T\mathbf{U}^{(\tau)}\Big) = \beta\sum_{xp}(\mathbf{U}^{(\tau)}_{xp})^2 \le \beta\sum_{xp}\frac{(\mathbf{U}^{(\tau)}_{xp})^4+(\tilde{\mathbf{U}}^{(\tau)}_{xp})^4}{2(\tilde{\mathbf{U}}^{(\tau)}_{xp})^2} \tag{31}$$

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

$$-2\beta\mathrm{Tr}\Big((\mathbf{H}^{(\tau)})^T\mathbf{U}^{(\tau)}\Big) \le -2\beta\sum_{xp}\mathbf{H}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{xp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}}) \tag{32}$$

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

$$Z_{pd}(\mathbf{U}^{(\tau)},\tilde{\mathbf{U}}^{(\tau)}) = \sum_{(\tau,j)\in\mathcal{I}}\Big(\sum_{xp}(\hat{\mathbf{\Phi}}^{(\tau j)}_{xp})'\frac{(\mathbf{U}^{(\tau)}_{xp})^4}{(\tilde{\mathbf{U}}^{(\tau)}_{xp})^3}$$
$$-2\sum_{xyp}(\hat{\mathbf{\Gamma}}^{(\tau j)}_{xy})'\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}\mathbf{U}^{(\tau)}_{yp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}})\Big)$$
$$+\sum_{(j,\tau)\in\mathcal{I}}\Big(\sum_{xp}(\hat{\mathbf{\Pi}}^{(j\tau)}_{xp})'\frac{(\mathbf{U}^{(\tau)}_{xp})^4}{(\tilde{\mathbf{U}}^{(\tau)}_{xp})^3}$$
$$-2\sum_{xyp}(\hat{\mathbf{\Omega}}^{(j\tau)}_{xy})'\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}\mathbf{U}^{(\tau)}_{yp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{yp}})\Big)$$
$$+\sum_{xp}\Big(\beta\frac{(\mathbf{U}^{(\tau)}_{xp})^4+(\tilde{\mathbf{U}}^{(\tau)}_{xp})^4}{2(\tilde{\mathbf{U}}^{(\tau)}_{xp})^2}-2\beta\mathbf{H}^{(\tau)}_{xp}\tilde{\mathbf{U}}^{(\tau)}_{xp}(1+\log\frac{\mathbf{U}^{(\tau)}_{xp}}{\tilde{\mathbf{U}}^{(\tau)}_{xp}})\Big)$$

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

$$Z_{pd}(\mathbf{U}^{(\tau)},\tilde{\mathbf{U}}^{(\tau)}) \ge \mathscr{L}_{pd}(\mathbf{U}^{(\tau)}),\quad Z_{pd}(\mathbf{U}^{(\tau)},\mathbf{U}^{(\tau)}) = \mathscr{L}_{pd}(\mathbf{U}^{(\tau)}) \tag{33}$$

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

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

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

where

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

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

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

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 $\kappa \ge 1$ when updating $\mathbf{U}^{(\tau)}$ using Eq. (14), we have

$$\mathscr{L}_{pd}((\mathbf{U}^{(\tau)})^{(\kappa)}) = Z_{pd}((\mathbf{U}^{(\tau)})^{(\kappa)},(\mathbf{U}^{(\tau)})^{(\kappa)})$$
$$\ge Z_{pd}((\mathbf{U}^{(\tau)})^{(\kappa+1)},(\mathbf{U}^{(\tau)})^{(\kappa)}) \ge \mathscr{L}_{pd}((\mathbf{U}^{(\tau)})^{(\kappa+1)})$$

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

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.