Learning the Consistent Behavior of Common Users for Target Node Prediction across Social Networks: Supplementary Materials

Shan-Hung Wu SHWU@CS.NTHU.EDU.TW
Hao-Heng Chien HHCHIEN@NETDB.CS.NTHU.EDU.TW
Kuan-Hua Lin KHLIN@NETDB.CS.NTHU.EDU.TW
National Tsing Hua University, ROC

Philip S. Yu PSYU@CS.UIC.EDU
University of Illinois at Chicago, USA

1. Solving the Objective of CICF

In the following, we detail our steps in solving the objective of CICF.

1.1. Overview

The objective function of CICF, with the non-negative constraints on $\mathbf{V}^{(g)}$ and $\mathbf{E}^{(g)}$, is a special case of *Non-negative Matrix Factorization* (NMF) (Lee & Seung, 1999; Recht et al., 2012; Seung & Lee, 2001; Yang et al., 2012), and can be solved by a multiplicative update approach (Seung & Lee, 2001; Yang & Oja, 2010). However, this approach suffers from the fluctuation problem in convergence (Yang & Oja, 2011; Zhang et al., 2012), as shown in Fig. 1.

We first transform the original CICF objective into a *Projective Non-negative Matrix Factorization* (PNMF) (Yang & Oja, 2010) problem. The new PNMF objective, as compared to the original one, offers two advantages. First, it learns sparse latent factors which is helpful to our problem in identifying the key user behavior. Second, it can be solved with convergence guarantee if we are able to find an *auxiliary function* for the objective (Seung & Lee, 2001).

We then devise an auxiliary function and obtain an iterative update rule for $\mathbf{E}^{(t)}$, as shown by Eq. (5), and another for $\mathbf{E}^{(s)}$ similarly. We then update $\mathbf{E}^{(t)}$ and $\mathbf{E}^{(s)}$ alternately until convergence. Fig. 1 show the convergence rate of our algorithm. Empirically, 8 to 15 iterations suffice to reach convergence.

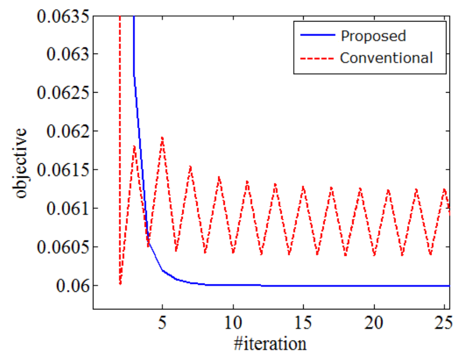

Figure 1. Comparison of convergence rates.

1.2. PNMF Formulation

Based on an intuition that a user's latent interests/character (modeled by the latent factors of nodes) can be summarized/averaged by what he/she has done (modeled by the latent factors of edges), we let $\mathbf{V}_{:,j} = \frac{1}{d(v_j)} \sum_{e_i \in I(v_j)} \mathbf{E}_{:,i}$, where $I(v_j)$ denotes the set of edges incident to v_j and $d(v_j)$ is the degree of v_j. Define a matrix $\mathbf{\Delta} \in \mathbb{R}^{m \times n}$ where $\mathbf{\Delta}_{i,j} = 1/d(v_j)$ if $e_i \in I(v_j)$ and 0 otherwise. We have $\mathbf{V} = \mathbf{E}\mathbf{\Delta}$, and the objective can be re-written as in Eq. (1).

Eq. (1) has the non-negative constraints on $\mathbf{E}^{(s)}$, and its first term is related to the *Projective Non-negative Matrix Factorization* (PNMF) (Yang & Oja, 2010). The PNMF, as compared to the traditional NMF (Lee & Seung, 1999; Recht et al., 2012; Seung & Lee, 2001; Yang et al., 2012), has an advantage in learning sparse latent factors and is helpful to our problem in identifying the key user behaviors. Typically, the PNMF is solved by a multiplicative update approach (Seung & Lee, 2001; Yang & Oja, 2010). However, this approach suffers from the fluctuation problem in convergence (Yang & Oja, 2011; Zhang et al., 2012).

1.3. Auxiliary Function

In this paper, we propose an iterative update algorithm based on the *auxiliary function* (Seung & Lee, 2001) to solve the objective of CICF while guaranteeing the convergence.

$$\arg\min_{\{\mathbf{E}^{(g)}>\mathbf{O}\}_g} \sum_g \|\mathbf{E}^{(g)\top}\mathbf{E}^{(g)}\boldsymbol{\Delta}^{(g)} - \mathbf{G}^{(g)}\|_{\mathcal{F}}^2 + \alpha\|\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\mathbf{P}^{(t)} - \mathbf{E}^{(s)}\boldsymbol{\Delta}^{(s)}\mathbf{P}^{(s)}\|_{\mathcal{F}}^2 + \beta\sum_g \mathrm{tr}(\mathbf{E}^{(g)}\mathbf{L}(\mathbf{K}(E^{(g)}))\mathbf{E}^{(g)\top}). \quad (1)$$

$$\begin{aligned} g_1(\widetilde{\mathbf{E}}^{(t)}, \mathbf{E}^{(t)}) = \quad & \textstyle\sum_{i,j} \frac{\widetilde{\mathbf{E}}_{i,j}^{(t)4}}{2\mathbf{E}_{i,j}^{(t)3}} (\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)} + \mathbf{E}^{(t)}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top})_{i,j} \\ & + \textstyle\sum_{i,j} \frac{\widetilde{\mathbf{E}}_{i,j}^{(t)2}}{\mathbf{E}_{i,j}^{(t)}} (\alpha\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\mathbf{P}^{(t)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top} + \beta\mathbf{E}^{(t)}\mathbf{L}_{+}^{(t)})_{i,j} \\ & -2\textstyle\sum_{i,j} \widetilde{\mathbf{E}}_{i,j}^{(t)}(\mathbf{E}^{(t)}(2\mathbf{G}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \beta\mathbf{L}_{-}^{(t)}))_{i,j} \\ & -2\textstyle\sum_{i,j}(1 + \log\frac{\widetilde{\mathbf{E}}_{i,j}^{(t)}}{\mathbf{E}_{i,j}^{(t)}})\mathbf{E}_{i,j}^{(t)}(\alpha\mathbf{E}^{(s)}\boldsymbol{\Delta}^{(s)}\mathbf{P}^{(s)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top})_{i,j} \end{aligned} \quad (2)$$

$$\begin{aligned} g_2(\widetilde{\mathbf{E}}^{(t)}, \mathbf{E}^{(t)}) = \quad & \textstyle\sum_{i,j} \frac{\widetilde{\mathbf{E}}_{i,j}^{(t)4}}{2\mathbf{E}_{i,j}^{(t)3}} (\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)} + \mathbf{E}^{(t)}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \alpha\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\mathbf{P}^{(t)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top} \\ & + \beta\mathbf{E}^{(t)}\mathbf{L}_{+}^{(t)})_{i,j} - 2\textstyle\sum_{i,j}(1 + \log\frac{\widetilde{\mathbf{E}}_{i,j}^{(t)}}{\mathbf{E}_{i,j}^{(t)}})\mathbf{E}_{i,j}^{(t)}(\mathbf{E}^{(t)}(2\mathbf{G}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \beta\mathbf{L}_{-}^{(t)}) + \alpha\mathbf{E}^{(s)}\boldsymbol{\Delta}^{(s)}\mathbf{P}^{(s)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top})_{i,j} \end{aligned} \quad (3)$$

$$\begin{aligned} 0 = \frac{\partial g_2(\widetilde{\mathbf{E}}^{(t)}, \mathbf{E}^{(t)})}{\partial \widetilde{\mathbf{E}}_{i,j}^{(t)}} = \quad & \frac{2\widetilde{\mathbf{E}}_{i,j}^{(t)3}}{\mathbf{E}_{i,j}^{(t)3}} (\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)} + \mathbf{E}^{(t)}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \alpha\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\mathbf{P}^{(t)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top} \\ & + \beta\mathbf{E}^{(t)}\mathbf{L}_{+}^{(t)})_{i,j} - 2\frac{\mathbf{E}_{i,j}^{(t)}}{\widetilde{\mathbf{E}}_{i,j}^{(t)}}(\mathbf{E}^{(t)}(2\mathbf{G}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \beta\mathbf{L}_{-}^{(t)}) + \alpha\mathbf{E}^{(s)}\boldsymbol{\Delta}^{(s)}\mathbf{P}^{(s)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top})_{i,j} \end{aligned} \quad (4)$$

$$\widetilde{\mathbf{E}}_{i,j}^{(t)} = \mathbf{E}_{i,j}^{(t)}\left(\frac{(\mathbf{E}^{(t)}(2\mathbf{G}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \beta\mathbf{L}_{-}^{(t)}) + \alpha\mathbf{E}^{(s)}\boldsymbol{\Delta}^{(s)}\mathbf{P}^{(s)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top})_{i,j}}{(\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)} + \mathbf{E}^{(t)}\mathbf{E}^{(t)\top}\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\boldsymbol{\Delta}^{(t)\top} + \alpha\mathbf{E}^{(t)}\boldsymbol{\Delta}^{(t)}\mathbf{P}^{(t)}\mathbf{P}^{(t)\top}\boldsymbol{\Delta}^{(t)\top} + \beta\mathbf{E}^{(t)}\mathbf{L}_{+}^{(t)})_{i,j}}\right)^{\frac{1}{4}} \quad (5)$$

Definition 1 (Auxiliary Function)**.** Given a function $h: \mathbb{R}^{m\times n} \to \mathbb{R}$. A function $g: (\mathbb{R}^{m\times n})^2 \to \mathbb{R}$ is called an *auxiliary function* of h iff $g(\widetilde{\mathbf{Z}}, \mathbf{Z}) \geq h(\widetilde{\mathbf{Z}})$ for all $\widetilde{\mathbf{Z}}$ and $\mathbf{Z}$ and the equation holds when $\widetilde{\mathbf{Z}} = \mathbf{Z}$.

If we can find an auxiliary function g of h, then the minimizer $\mathbf{Z}^*$ of h can be found using the following iterative update rule:

$$\mathbf{Z}^{(i+1)} = \arg\min_{\mathbf{Z}} g(\mathbf{Z}, \mathbf{Z}^{(i)}). \quad (6)$$

Notice that $h(\mathbf{Z}^{(i+1)})$ is monotonically decreasing during the iterations, as $h(\mathbf{Z}^{(i+1)}) \leq g(\mathbf{Z}^{(i+1)}, \mathbf{Z}^{(i)}) \leq g(\mathbf{Z}^{(i)}, \mathbf{Z}^{(i)}) = h(\mathbf{Z}^{(i)})$.

Eq. (1) is a function of both $\mathbf{E}^{(t)}$ and $\mathbf{E}^{(s)}$. We solve our objective using an iterative alternate approach. In one iteration, some $\mathbf{E}^{(s)}$, $s = a$ or p, is fixed and we look for another by solving Eq. (6), then in the next iteration we alternate s and solve the opposite. This process is repeated until the objective converges.

Without loss of generality, we first assume $\mathbf{E}^{(s)}$ is fixed and denote Eq. (1) as $h(\mathbf{E}^{(t)})$. Our target is to derive an auxiliary function for $h(\mathbf{E}^{(t)})$ so that we can employ Eq. (6) to find the $\mathbf{E}^{(t)}$ that will be fixed in the next iteration. We rewrite the Laplacian matrix $\mathbf{L}(\mathbf{K}(E^{(t)}))$ as $\mathbf{L}_+^{(t)} - \mathbf{L}_-^{(t)}$, where $\mathbf{L}_-^{(t)}$ is the similarity matrix $\mathbf{K}(E^{(t)})$ and $\mathbf{L}_+^{(t)}$ is a diagonal matrix with the ith entry on the diagonal being the sum of the ith column (or row) in $\mathbf{L}_-^{(t)}$. Note that all entries in $\mathbf{L}_+^{(s)}$ and $\mathbf{L}_-^{(s)}$ are non-negative.

Theorem 2. *Given a positive semi-definite $\mathbf{L}_-^{(t)}$. Ignoring the constant terms, the function g_1 defined in Eq. (2) is an auxiliary function of $h(\mathbf{E}^{(t)})$.*

Proof. See Appendix. □

Based on the above theorem, the similarity matrix $\mathbf{K}(E^{(t)})$ is required to be positive semi-definite. To ensure this, we need to use a positive definite kernel function, such as the Gaussian RBF kernel, to set each $\mathbf{K}(E^{(t)})_{i,j}$.

1.4. Further Simplification

In fact, the objective in Eq. (6) based on g_1 is still difficult to solve because when we set the derivative of g_1 to zero to find the minimum, we cannot derive the analytical form solution directly from an equation containing more than two monomials (Yang & Oja, 2011). However, we can combine 1) the first and second terms and 2) the third and fourth terms respectively to derive another auxiliary function g_2, which contains only two monomials and is easy to solve.

Lemma 3. *Given $a, b, x \in \mathbb{R}$, the inequality $x^a \leq \frac{a}{b}x^b + 1 - \frac{a}{b}$ holds if $b > a > 1$, and the equality holds iff $x = 1$.*

Proof. Let $f(t) = x^a$. Because $f(t)$ is convex, we have $f(t) = f((\frac{a}{b})b + (1-\frac{a}{b})0) \leq \frac{a}{b}f(b) + (1-\frac{a}{b})f(0) = \frac{a}{b}x^b + 1 - \frac{a}{b}$. □

Thus, for any $\mathbf{S}, \mathbf{S}' \in (\mathbb{R}^+)^{k\times m}$ and a symmetric $\mathbf{A} \in (\mathbb{R}^+)^{m\times m}$ we have

$$\begin{aligned}\frac{(\mathbf{S}'\mathbf{A})_{i,j}\,\mathbf{S}_{i,j}^2}{\mathbf{S}'_{i,j}} &= (\mathbf{S}'\mathbf{A})_{i,j}\,\mathbf{S}'_{i,j}(\frac{\mathbf{S}_{ip}}{\mathbf{S}'_{ip}})^2 \\ &\leq \frac{2}{4}(\mathbf{S}'\mathbf{A})_{i,j}\,\mathbf{S}'_{i,j}(\frac{\mathbf{S}_{ip}}{\mathbf{S}'_{ip}})^4 + \text{constant} \\ &= \frac{(\mathbf{S}'\mathbf{A})_{i,j}\,\mathbf{S}_{i,j}^4}{2\mathbf{S}'^3_{i,j}} + \text{constant}.\end{aligned}$$

Regarding respectively $\mathbf{S} = \widetilde{\mathbf{E}}^{(t)}$, $\mathbf{S}' = \mathbf{E}^{(t)}$, and $\mathbf{A} = \alpha\mathbf{\Delta}^{(t)}\mathbf{P}^{(t)}\mathbf{P}^{(t)\top}\mathbf{\Delta}^{(t)\top} + \beta\mathbf{L}_+^{(t)}$, we can raise the power of $\widetilde{\mathbf{E}}_{i,j}^{(t)}$ in the second term and merge this raised term into the first term. To merge the third and fourth terms, we use the inequality $x \geq 1 + \log(x)$ for any positive $x \in \mathbb{R}$. Regarding $x = \widetilde{\mathbf{E}}_{i,j}^{(t)}/\mathbf{E}_{i,j}^{(t)}$, we can see that the third term is upper-bounded by

$$-2\sum_{i,j}(1+\log(\frac{\widetilde{\mathbf{E}}_{i,j}^{(t)}}{\mathbf{E}_{i,j}^{(t)}}))\mathbf{E}_{i,j}^{(t)}(\mathbf{E}^{(t)}(2\mathbf{G}^{(t)}\mathbf{\Delta}^{(t)\top} + \beta\mathbf{L}_-^{(t)}))_{i,j},$$

which can be merged into the fourth term. After the merging, we can write g_2 as in Eq. (3). Finally, we solve Eq. (6) based on g_2 by setting the derivative of g_2 to 0, as shown in Eq (4), and obtain the update rule for $\mathbf{E}^{(t)}$, as listed in Eq. (5).

The update rule for $\mathbf{E}^{(s)}$ can be obtained similarly and is omitted.

Appendix: Proof of Theorem 2

Our goal is to prove that g_1 is an auxiliary function of h. We first rewrite each term in h using Eqs. (7), (8), and (9) and obtain Eq. (10). It is easy to verify that the equality $g_1(\widetilde{\mathbf{E}}^{(a)}, \mathbf{E}^{(a)}) + \text{constant} = h(\widetilde{\mathbf{E}}^{(a)})$ holds when $\widetilde{\mathbf{E}}^{(a)} = \mathbf{E}^{(a)}$. We prove that $g_1(\widetilde{\mathbf{E}}^{(a)}, \mathbf{E}^{(a)}) + \text{constant} \geq h(\widetilde{\mathbf{E}}^{(a)})$ for any $\widetilde{\mathbf{E}}^{(a)}$ by comparing the g_1 and h term-by-term.

We show that the first term of g_1 is larger than or equal to that of h using the Jensen's inequality

$$\varphi(\frac{\sum_{s,t}\lambda_{s,t}\mathbf{X}_{s,t}}{\sum_{s,t}\lambda_{s,t}}) \leq \frac{\sum_{s,t}\lambda_{s,t}\varphi(\mathbf{X}_{s,t})}{\sum_{s,t}\lambda_{s,t}},$$

where φ is a real convex function and $\lambda_{s,t}$ are non-negative weights with positive sum. Let $\varphi(x) = x^2$ and $\lambda_{i,j,s,t} = \mathbf{E}_{t,i}^{(a)}\mathbf{E}_{t,s}^{(a)}\mathbf{\Delta}_{s,j}^{(a)}/(\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)})_{i,j}$ ($\sum_{s,t}\lambda_{i,j,s,t} = 1$), we have

$$\begin{aligned}\|\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}-\mathbf{G}^{(a)}\|_{\mathcal{F}}^2 &= \mathrm{tr}((\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}-\mathbf{G}^{(a)})^\top(\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}-\mathbf{G}^{(a)}))\\ &= -2\,\mathrm{tr}(\mathbf{E}^{(a)}\mathbf{G}^{(a)}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top})+\textstyle\sum_{i,j}(\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)})_{i,j}^2+\text{constant} \qquad (7)\end{aligned}$$

$$\mathrm{tr}(\mathbf{E}^{(a)}\mathbf{L}(\mathbf{K}(E^{(a)}))\mathbf{E}^{(a)\top})=\mathrm{tr}(\mathbf{E}^{(a)}(\mathbf{L}_+^{(a)}-\mathbf{L}_-^{(a)})\mathbf{E}^{(a)\top})=\mathrm{tr}(\mathbf{E}^{(a)}\mathbf{L}_+^{(a)}\mathbf{E}^{(a)\top}-\mathbf{E}^{(a)}\mathbf{L}_-^{(a)}\mathbf{E}^{(a)\top}) \qquad (8)$$

$$\begin{aligned}\|\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{P}^{(a)}-\mathbf{E}^{(p)}\mathbf{\Delta}^{(p)}\mathbf{P}^{(p)}\|_{\mathcal{F}}^2 &= \mathrm{tr}((\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{P}^{(a)}-\mathbf{E}^{(p)}\mathbf{\Delta}^{(p)}\mathbf{P}^{(p)})^\top(\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{P}^{(a)}-\mathbf{E}^{(p)}\mathbf{\Delta}^{(p)}\mathbf{P}^{(p)})) \qquad (9)\\ &= \mathrm{tr}(\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{P}^{(a)}\mathbf{P}^{(a)\top}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top}-2\mathbf{E}^{(p)}\mathbf{\Delta}^{(p)}\mathbf{P}^{(p)}\mathbf{P}^{(a)\top}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top})+\text{constant}\end{aligned}$$

$$\begin{aligned}h(\mathbf{E}^{(a)}) &= \textstyle\sum_{i,j}(\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)})_{i,j}^2+\mathrm{tr}(\beta\mathbf{E}^{(a)}\mathbf{L}_+^{(a)}\mathbf{E}^{(a)\top}+\alpha\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{P}^{(a)}\mathbf{P}^{(a)\top}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top})\\ &\quad -\mathrm{tr}(\mathbf{E}^{(a)}(2\mathbf{G}^{(a)}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top}+\beta\mathbf{L}_-^{(a)})\mathbf{E}^{(a)\top})-\mathrm{tr}(2\alpha\mathbf{E}^{(p)}\mathbf{\Delta}^{(p)}\mathbf{P}^{(p)}\mathbf{P}^{(a)\top}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top})+\text{constant} \qquad (10)\end{aligned}$$

$$\begin{aligned}&\textstyle\sum_{i,j}(\widetilde{\mathbf{E}}^{(a)\top}\widetilde{\mathbf{E}}^{(a)}\mathbf{\Delta}^{(a)})_{i,j}^2\\ &\quad=\textstyle\sum_{i,j}(\sum_{s,t}\lambda_{i,j,s,t}(\widetilde{\mathbf{E}}^{(a)\top}\widetilde{\mathbf{E}}^{(a)}\mathbf{\Delta}^{(a)})_{i,j})^2\\ &\quad\le\textstyle\sum_{i,t}\frac{\widetilde{\mathbf{E}}_{t,i}^{(a)4}}{2\mathbf{E}_{t,i}^{(a)3}}(\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{\Delta}^{(a)\top}\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\\ &\qquad\quad+\mathbf{E}^{(a)}\mathbf{E}^{(a)\top}\mathbf{E}^{(a)}\mathbf{\Delta}^{(a)}\mathbf{\Delta}^{(a)\top})_{t,i}.\end{aligned}$$

We omit the detailed derivation here.

We can easily verify that the second term of g_1 is no less than that of h using the lemma below:

Lemma 4. *For any* $\mathbf{S}\in(\mathbb{R}^+)^{k\times m}$, $\mathbf{S}'\in(\mathbb{R}^+)^{k\times m}$, *and symmetric* $\mathbf{A}\in(\mathbb{R}^+)^{m\times m}$, *the following inequality holds*

$$\sum_{i=1}^{k}\sum_{j=1}^{m}\frac{(\mathbf{S}'\mathbf{A})_{i,j}\mathbf{S}_{i,j}^2}{\mathbf{S}'_{i,j}}\ge\mathrm{tr}(\mathbf{S}\mathbf{A}\mathbf{S}^\top).$$

Proof. Let $u_{i,j}=\mathbf{S}_{i,j}/\mathbf{S}'_{i,j}$. Subtracting the right from left equals

$$\begin{aligned}&\textstyle\sum_{i,j}\sum_{p=1}^{m}\mathbf{S}'_{i,p}\mathbf{A}_{p,j}\mathbf{S}'_{i,j}(u_{i,j}^2-u_{i,p}u_{i,j})\\ &\quad=\textstyle\sum_{i,j,p}\mathbf{S}'_{i,p}\mathbf{A}_{p,j}\mathbf{S}'_{i,j}(\frac{u_{i,j}^2+u_{i,p}^2}{2}-u_{i,p}u_{i,j})\\ &\quad=\textstyle\frac{1}{2}\sum_{i,j,p}\mathbf{S}'_{i,p}\mathbf{A}_{p,j}\mathbf{S}'_{i,j}(u_{i,j}-u_{i,p})^2,\end{aligned}$$

which is larger then or equal to 0. □

The third term of h is upper-bounded by its first order Taylor expansion at $\mathbf{E}^{(a)}$:

$$-\mathrm{tr}(\widetilde{\mathbf{E}}^{(a)}\mathbf{X}\widetilde{\mathbf{E}}^{(a)\top})\le-2\sum_{i,j}\widetilde{\mathbf{E}}_{i,j}^{(a)}(\mathbf{E}^{(a)}\mathbf{X})_{i,j}+\text{constant},$$

where $\mathbf{X}=2\mathbf{G}^{(a)}\mathbf{\Delta}^{(a)\top}+\beta\mathbf{L}_-^{(a)}$, since each term in $\mathbf{X}$ is positive semi-definite and the left hand side is concave with respect to $\widetilde{\mathbf{E}}^{(a)}$.

We compare the fourth terms of g_1 and h using the inequality $x\ge1+\log(x)$ for any positive $x\in\mathbb{R}$. Regarding $x=\widetilde{\mathbf{E}}_{i,j}^{(a)}/\mathbf{E}_{i,j}^{(a)}$, we can see that the fourth term of h is always smaller.

With the above results, we obtain the proof. Note that the Hessian matrix of g_1 with respect to $\widetilde{\mathbf{E}}^{(a)}$ is positive semi-definite (i.e., $\nabla\nabla_{\widetilde{\mathbf{E}}^{(a)}}g_1(\widetilde{\mathbf{E}}^{(a)},\mathbf{E}^{(a)})\succeq0$), so g_1 is convex and the solution to Eq. (6) exists.

References

Lee, Daniel D and Seung, H Sebastian. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

Recht, Ben, Re, Christopher, Tropp, Joel, and Bittorf, Victor. Factoring nonnegative matrices with linear programs. In *Advances in Neural Information Processing Systems*, pp. 1223–1231, 2012.

Seung, D. and Lee, L. Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems*, 13:556–562, 2001.

Yang, Z. and Oja, E. Unified development of multiplicative algorithms for linear and quadratic nonnegative matrix factorization. *IEEE Trans. on Neural Networks*, 22(12): 1878–1891, 2011.

Yang, Zhirong and Oja, Erkki. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Trans. on Neural Networks*, 21(5):734–749, 2010.

Yang, Zhirong, Hao, Tele, Dikmen, Onur, Chen, Xi, and Oja, Erkki. Clustering by nonnegative matrix factorization using graph random walk. In *Advances in Neural Information Processing Systems*, pp. 1088–1096, 2012.

Zhang, H., Yang, Z., and Oja, E. Adaptive multiplicative updates for projective nonnegative matrix factorization. In *Proc. of the Neural Information Processing*, pp. 277–284, 2012.