Supplementary Material for Identifying Multiple Causal Anomalies by Modeling Local Propagations

Jingchao Ni [1], Wei Cheng [2], Kai Zhang [2], Dongjin Song [2], Tan Yan [2], Haifeng Chen [2], and Xiang Zhang [1]
[1]College of Information Sciences and Technology, Pennsylvania State University
[2]NEC Laboratories America
[1]{jzn47, xzhang}@ist.psu.edu,
[2]{weicheng, kzhang, dsong, yan, haifeng}@nec-labs.com

1. GRADIENT DEFINITIONS

First, from Sec. 3.3, $\mathcal{J}_{CR}(\mathbf{U})$ and $\mathcal{J}_{CR}(\mathbf{s})$ can be written as following

$$\mathcal{J}_{CR}(\mathbf{U}) = -\sum_{xy}\mathbf{A}_{xy}\log\hat{\mathbf{A}}_{xy} - (\alpha-1)\sum_{xi}\log\mathbf{U}_{xi} - \beta\sum_{xy}\mathbf{B}_{xy}\log(\sum_i \mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i) + \beta\sum_{xy}\mathbf{W}_{xy}\sum_i\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i$$

$$\mathcal{J}_{CR}(\mathbf{s}) = -\beta\sum_{xy}\mathbf{B}_{xy}\log(\sum_i\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i) + \beta\sum_{xyi}\mathbf{W}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i$$

where $\hat{\mathbf{A}}$ is defined in Eq. (2)[1].

Then the positive and non-positive parts of the gradients of $\mathcal{J}_{CR}(\mathbf{U})$ w.r.t. $\mathbf{U}$ and those of $\mathcal{J}_{CR}(\mathbf{s})$ w.r.t. $\mathbf{s}$ are given as following

$$(\nabla_U)^+_{xi} = 2(\mathbf{ZU})_{xi}\mathbf{h}_i^{-1} + \alpha\mathbf{U}_{xi}^{-1} + 2\beta(\mathbf{YU})_{xi}\mathbf{s}_i$$
$$(\nabla_U)^-_{xi} = (\mathbf{U}^T\mathbf{ZU})_{ii}\mathbf{h}_i^{-2} + \mathbf{U}_{xi}^{-1} + 2\beta(\mathbf{WU})_{xi}\mathbf{s}_u$$
$$(\nabla_s)^+_i = \beta(\mathbf{U}^T\mathbf{WU})_{ii} + \mathbf{s}_i^{-1}$$
$$(\nabla_s)^-_i = \beta(\mathbf{U}^T\mathbf{YU})_{ii} + \gamma\mathbf{s}_i^{-1}$$

where

$$\mathbf{Z}_{xy} = \mathbf{A}_{xy}(\sum_i\frac{\mathbf{U}_{xi}\mathbf{U}_{yi}}{\sum_z\mathbf{U}_{zi}})^{-1},\ \mathbf{h}_i = \sum_z\mathbf{U}_{zi}$$
$$\mathbf{Y}_{xy} = \mathbf{B}_{xy}(\sum_i\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i)^{-1}$$

2. PROOF OF THEOREM 3.1

Proof. First, the Lagrangian multipliers $\boldsymbol{\lambda}$ can be derived by using the equality constraints in Eq. (8). The gradient of $\mathcal{L}_U$ w.r.t. $\mathbf{U}_{xi}$ can be written as

$$\frac{\partial\mathcal{L}_U}{\partial\mathbf{U}_{xi}} = (\nabla_U)^+_{xi} - (\nabla_U)^-_{xi} + \lambda_x \tag{25}$$

The Karush-Kuhn-Tucker (KKT) completeness slackness condition [1] on the non-negative constraint of $\mathbf{U}$ suggests a fixed-point updating rule for $\mathbf{U}_{xi}$ as

$$\mathbf{U}^{\text{new}}_{xi} = \mathbf{U}_{xi}\frac{(\nabla_U)^-_{xi} - \lambda_x}{(\nabla_U)^+_{xi}} \tag{26}$$

By imposing $\sum_{i=1}^k(\mathbf{U}_{xi})^{\text{new}} = 1$, we can derive

$$\lambda_x = (b_x - 1)/a_x \tag{27}$$

where

$$a_x = \sum_{i=1}^k\frac{\mathbf{U}_{xi}}{(\nabla_U)^+_{xi}},\quad b_x = \sum_{i=1}^k\mathbf{U}_{xi}\frac{(\nabla_U)^-_{xi}}{(\nabla_U)^+_{xi}}$$

Such inferred $\boldsymbol{\lambda}$ will adjust $\mathbf{U}$ toward the equality constraint. With the use of it, we prove Theorem 3.1. The proof follows a Majorization-Minimization approach [4], which is similar to the auxiliary function approach [2]. In the following, we use $\mathbf{U}$ to denote variable and $\tilde{\mathbf{U}}$ to denote current estimate (i.e., a constant). Let $\mathcal{L}_U(\mathbf{U})$ be the function in Eq. (14) w.r.t. $\mathbf{U}$ when fix $\boldsymbol{\lambda}$ as the constant in Eq. (27). Next, we derive a tight upper bound for $\mathcal{L}_U(\mathbf{U})$, which can be written by

$$\mathcal{L}_U(\mathbf{U}) = -\sum_{xy}\mathbf{A}_{xy}\log\hat{\mathbf{A}}_{xy} - (\alpha-1)\sum_{xi}\log\mathbf{U}_{xi} - \beta\sum_{xy}\mathbf{B}_{xy}\log(\sum_i\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i) + \beta\sum_{xy}\mathbf{W}_{xy}\sum_i\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i + \sum_x\lambda_x(\sum_i\mathbf{U}_{xi} - 1)$$

Let $\phi_{xyi} = \frac{\tilde{\mathbf{U}}_{xi}\tilde{\mathbf{U}}_{yi}}{\sum_z\tilde{\mathbf{U}}_{zi}}(\sum_j\frac{\tilde{\mathbf{U}}_{xj}\tilde{\mathbf{U}}_{yj}}{\sum_z\tilde{\mathbf{U}}_{zj}})^{-1}$, $\theta_{xyi} = \frac{\tilde{\mathbf{U}}_{xi}\tilde{\mathbf{U}}_{yi}\mathbf{s}_i}{\sum_j\tilde{\mathbf{U}}_{xj}\tilde{\mathbf{U}}_{yj}\mathbf{s}_j}$, using the convexity of $\log(\cdot)$, we have

$$\mathcal{L}_U(\mathbf{U}) \le -\sum_{xyi}\mathbf{A}_{xy}\phi_{xyi}(\log\mathbf{U}_{xi} + \log\mathbf{U}_{yi} - \log\sum_z\mathbf{U}_{zi}) - (\alpha-1)\sum_{xi}\log\mathbf{U}_{xi} - \beta\sum_{xyi}\mathbf{B}_{xy}\theta_{xyi}(\log\mathbf{U}_{xi} + \log\mathbf{U}_{yi}) + \beta\sum_{xyi}\mathbf{W}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i + \sum_{xi}\lambda_x\mathbf{U}_{xi} + C_1 \tag{28}$$

where C_1 is a constant. Using the fact that $d \ge 1 + \log d$ for any $d > 0$, we have

$$\log\sum_z\mathbf{U}_{zi} \le \frac{\sum_z\mathbf{U}_{zi}}{\sum_z\tilde{\mathbf{U}}_{zi}} - 1 + \log\sum_z\tilde{\mathbf{U}}_{zi}$$

We can further upper bound the first term in Eq. (28) by

$$-\sum_{xyi}\mathbf{A}_{xy}\phi_{xyi}(\log\mathbf{U}_{xi} + \log\mathbf{U}_{yi} - \frac{\sum_z\mathbf{U}_{zi}}{\sum_z\tilde{\mathbf{U}}_{zi}}) + C_2$$

where C_2 is another constant. Therefore, we define the fol-

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

lowing function as an upper bound of $\mathcal{L}_U(\mathbf{U})$.

$$\begin{aligned}\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}}) &= -\sum_{xyi}\mathbf{A}_{xy}\phi_{xyi}(\log\mathbf{U}_{xi}+\log\mathbf{U}_{yi}-\frac{\sum_z\mathbf{U}_{zi}}{\sum_z\tilde{\mathbf{U}}_{zi}})\\ &-(\alpha-1)\sum_{xi}\log\mathbf{U}_{xi}-\beta\sum_{xyi}\mathbf{B}_{xy}\theta_{xyi}(\log\mathbf{U}_{xi}+\log\mathbf{U}_{yi})\\ &+\beta\sum_{xyi}\mathbf{W}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i+\sum_{xi}\lambda_x\mathbf{U}_{xi}\\ &+\sum_{xi}(\frac{1}{a_x}+\frac{1}{\tilde{\mathbf{U}}_{xi}})\tilde{\mathbf{U}}_{xi}(\frac{\mathbf{U}_{xi}}{\tilde{\mathbf{U}}_{xi}}-\log\frac{\mathbf{U}_{xi}}{\tilde{\mathbf{U}}_{xi}}-1)+C_3\end{aligned}$$

where $C_3 = C_1 + C_2$ is a constant. The term before C_3 is non-negative and is added to guarantee the non-negativity of the multiplicative updating rule. It is known as the "moving term" technique in existing work [3].

Note $\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})$ is a tight upper bound of $\mathcal{L}_U(\mathbf{U})$. That is, $\mathcal{Z}_U(\mathbf{U},\mathbf{U}) = \mathcal{L}_U(\mathbf{U})$ (Some terms are hidden in C_3). According to Definition 3.5, $\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})$ is also an *auxiliary function* of $\mathcal{L}_U(\mathbf{U})$ [2].

Next, we derive the minimal solution to $\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})$. The gradient of $\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})$ w.r.t. $\mathbf{U}_{xi}$ is

$$\begin{aligned}\frac{\partial\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})}{\partial\mathbf{U}_{xi}} &= -2\sum_y\mathbf{A}_{xy}\phi_{xyi}\frac{1}{\mathbf{U}_{xi}}+\sum_y\mathbf{A}_{xy}\phi_{xyi}\frac{1}{\sum_z\tilde{\mathbf{U}}_{zi}}\\ &-(\alpha-1)\frac{1}{\mathbf{U}_{xi}}-2\beta\sum_y\mathbf{B}_{xy}\theta_{xyi}\frac{1}{\mathbf{U}_{xi}}\\ &+2\beta\sum_y\mathbf{W}_{xy}\mathbf{U}_{yi}\mathbf{s}_i+\lambda_x+(\frac{1}{a_x}+\frac{1}{\tilde{\mathbf{U}}_{xi}})\tilde{\mathbf{U}}_{xi}(\frac{1}{\tilde{\mathbf{U}}_{xi}}-\frac{1}{\mathbf{U}_{xi}})\end{aligned}$$

Equivalently,

$$\begin{aligned}\frac{\partial\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})}{\partial\mathbf{U}_{xi}} &= ((\nabla_U)^+_{xi}-\frac{1}{\tilde{\mathbf{U}}_{xi}})-\frac{\tilde{\mathbf{U}}_{xi}}{\mathbf{U}_{xi}}((\nabla_U)^-_{xi}-\frac{1}{\tilde{\mathbf{U}}_{xi}})\\ &+\frac{b_x-1}{a_x}+\frac{1}{a_x}+\frac{1}{\tilde{\mathbf{U}}_{xi}}-\frac{\tilde{\mathbf{U}}_{xi}}{a_x\mathbf{U}_{xi}}-\frac{1}{\mathbf{U}_{xi}}\\ &= -\frac{\tilde{\mathbf{U}}_{xi}}{\mathbf{U}_{xi}}((\nabla_U)^-_{xi}+\frac{1}{a_x})+((\nabla_U)^+_{xi}+\frac{b_x}{a_x})\end{aligned}$$

Therefore, the global minimum of $\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})$ can be obtained by solving $\frac{\partial\mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}})}{\partial\mathbf{U}_{xi}} = 0$. This gives the solution

$$\mathbf{U}_{xi} = \tilde{\mathbf{U}}_{xi}\frac{a_x(\nabla_U)^-_{xi}+1}{a_x(\nabla_U)^+_{xi}+b_x}$$

which is consistent with the updating rule in Eq. (16). According to the following tight upper bound properties

$$(1)\ \mathcal{Z}_U(\mathbf{U},\tilde{\mathbf{U}}) \geq \mathcal{L}_U(\mathbf{U}),\ (2)\ \mathcal{Z}_U(\mathbf{U},\mathbf{U}) = \mathcal{L}_U(\mathbf{U})$$

At any iteration $\kappa \geq 1$ during updating $\mathbf{U}$ by Eq. (16), we have

$$\begin{aligned}\mathcal{L}_U(\mathbf{U}^{(\kappa)},\boldsymbol{\lambda}) &= \mathcal{Z}_U(\mathbf{U}^{(\kappa)},\mathbf{U}^{(\kappa)})\\ &\geq \mathcal{Z}_U(\mathbf{U}^{(\kappa+1)},\mathbf{U}^{(\kappa)}) \geq \mathcal{L}_U(\mathbf{U}^{(\kappa+1)},\boldsymbol{\lambda})\end{aligned}$$

which completes the proof of Theorem 3.1. □

3. PROOF OF THEOREM 3.2

PROOF. In the following, we use $\mathbf{s}$ to denote variable and $\tilde{\mathbf{s}}$ to denote current estimate (i.e., a constant). First, let $\mathcal{J}_{CR}(\mathbf{s})$ be the objective function in Eq. (8) w.r.t. $\mathbf{s}$, which can be written by

$$\mathcal{J}_{CR}(\mathbf{s}) = -\beta\sum_{xy}\mathbf{B}_{xy}\log(\sum_i\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i)+\beta\sum_{xyi}\mathbf{W}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i$$

Let $\pi_{xyi} = \frac{\mathbf{U}_{xi}\mathbf{U}_{yi}\tilde{\mathbf{s}}_i}{\sum_j\mathbf{U}_{xj}\mathbf{U}_{yj}\tilde{\mathbf{s}}_j}$, using the convexity of $\log(\cdot)$, we have

$$\begin{aligned}\mathcal{J}_{CR}(\mathbf{s}) &\leq -\beta\sum_{xyi}\mathbf{B}_{xy}\pi_{xyi}\log\mathbf{s}_i+\beta\sum_{xyi}\mathbf{W}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\mathbf{s}_i\\ &+\sum_i(\frac{\mathbf{s}_i}{\tilde{\mathbf{s}}_i}-\log\frac{\mathbf{s}_i}{\tilde{\mathbf{s}}_i}-1)+C_4\end{aligned}$$

where C_4 is a constant, the term before C_4 is non-negative and is added to guarantee the non-negativity of the multiplicative updating rule.

Let function $\mathcal{Z}_s(\mathbf{s},\tilde{\mathbf{s}})$ be the above upper bound, it is a tight upper bound of $\mathcal{J}_{CR}(\mathbf{s})$, i.e., we have $\mathcal{Z}_s(\mathbf{s},\mathbf{s}) = \mathcal{J}_{CR}(\mathbf{s})$ (some terms are hidden in C_4).

Next, we enforce the inequality constraint $\mathbf{s}_i \leq 1$ of Eq. (8) and solve the following optimization problem, which will give us a solution that decreases the objective value in Eq. (8) and satisfies the constraints in Eq. (8).

$$\min_{\mathbf{s}}\ \mathcal{Z}_s(\mathbf{s},\tilde{\mathbf{s}}),\ \text{ s.t. } 0 \leq \mathbf{s}_i \leq 1,\ \forall 1 \leq i \leq k \tag{29}$$

Let $\mathcal{L}_s$ be the Lagrangian function of the above optimization problem, which can be represented by

$$\mathcal{L}_s(\mathbf{s}) = \mathcal{Z}_s(\mathbf{s},\tilde{\mathbf{s}})+\sum_{i=1}^k\mu_i(\mathbf{s}_i-1)$$

where μ_i $(1 \leq i \leq k)$ is the Lagrangian multiplier.

Then the gradient of $\mathcal{L}_s(\mathbf{s})$ w.r.t. $\mathbf{s}_i$ is

$$\frac{\partial\mathcal{L}_s(\mathbf{s})}{\mathbf{s}_i} = -\beta\sum_{xy}\mathbf{B}_{xy}\pi_{xyi}\frac{1}{\mathbf{s}_i}+\beta\sum_{xy}\mathbf{W}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}+\frac{1}{\tilde{\mathbf{s}}_i}-\frac{1}{\mathbf{s}_i}+\mu_i$$

Equivalently,

$$\begin{aligned}\frac{\partial\mathcal{L}_s(\mathbf{s})}{\mathbf{s}_i} &= ((\nabla_s)^+_i-\frac{1}{\tilde{\mathbf{s}}_i})-\frac{\tilde{\mathbf{s}}_i}{\mathbf{s}_i}((\nabla_s)^-_i-\frac{1}{\tilde{\mathbf{s}}_i})+\frac{1}{\tilde{\mathbf{s}}_i}-\frac{1}{\mathbf{s}_i}+\mu_i\\ &= -\frac{\tilde{\mathbf{s}}_i}{\mathbf{s}_i}(\nabla_s)^-_i+(\nabla_s)^+_i+\mu_i\end{aligned}$$

According to the Karush-Kuhn-Tucker (KKT) conditions [1], for $1 \leq i \leq k$, we have

- Gradient: $\mu_i = \frac{\tilde{\mathbf{s}}_i}{\mathbf{s}_i}(\nabla_s)^-_i-(\nabla_s)^+_i$
- Non-negativity: $\mu_i \geq 0$
- Feasibility: $\mathbf{s}_i \leq 1$
- Complementary slackness: $\mu_i(\mathbf{s}_i-1) = 0$

The gradient and complementary slackness conditions suggest the optimal solution $\mathbf{s}^*_i$ can be either $\tilde{\mathbf{s}}_i[(\nabla_s)^-_i/(\nabla_s)^+_i]$ or 1. The non-negativity and feasibility conditions suggest $\mathbf{s}^*_i \leq \tilde{\mathbf{s}}_i[(\nabla_s)^-_i/(\nabla_s)^+_i]$ and $\mathbf{s}^*_i \leq 1$.

Therefore, we obtain the global minimal $\mathbf{s}^*_i$ to Eq. (29) as

$$\mathbf{s}^*_i = \min(\tilde{\mathbf{s}}_i[(\nabla_s)^-_i/(\nabla_s)^+_i],1)$$

which is consistent with the updating rule in Eq. (17). According to the following tight upper bound properties

$$(1)\ \mathcal{Z}_s(\mathbf{s},\tilde{\mathbf{s}}) \geq \mathcal{J}_{CR}(\mathbf{s}),\ (2)\ \mathcal{Z}_s(\mathbf{s},\mathbf{s}) = \mathcal{J}_{CR}(\mathbf{s})$$

At any iteration $\kappa \geq 1$ when we update $\mathbf{s}$ by Eq. (17), we have

$$\mathcal{J}_{CR}(\mathbf{s}^{(\kappa)}) = \mathcal{Z}_s(\mathbf{s}^{(\kappa)},\mathbf{s}^{(\kappa)}) \geq \mathcal{Z}_s(\mathbf{s}^{(\kappa+1)},\mathbf{s}^{(\kappa)}) \geq \mathcal{J}_{CR}(\mathbf{s}^{(\kappa+1)})$$

which completes the proof of Theorem 3.2. □

4. PROOF OF THEOREM 3.4

PROOF. First, we can rewrite the objective function in Eq. (13) as (after removing some constants)

$$\mathcal{J}_H(\mathbf{E}) = -2tr(\mathbf{Q}(\mathbf{U}\circ\mathbf{E})(\mathbf{U}\circ\mathbf{E})) + tr(\mathbf{R}\mathbf{R}) + \tau\sum_{xi}\mathbf{E}_{xi}$$

where

$$\mathbf{Q} = \mathbf{H}^T(\mathbf{B}\circ\mathbf{C})^T\mathbf{H}$$
$$\mathbf{R} = \mathbf{C}\circ[\mathbf{H}(\mathbf{U}\circ\mathbf{E})(\mathbf{U}\circ\mathbf{E})^T\mathbf{H}^T]$$

We then derive upper bounds for the above three terms. Let $\tilde{\mathbf{E}}$ be current estimate (i.e., a constant). Using the general inequality $d \geq 1 + \log d$ for any $d > 0$, we have

$$\frac{\mathbf{E}_{xi}\mathbf{E}_{yi}}{\tilde{\mathbf{E}}_{xi}\tilde{\mathbf{E}}_{yi}} \geq 1 + \log\frac{\mathbf{E}_{xi}\mathbf{E}_{yi}}{\tilde{\mathbf{E}}_{xi}\tilde{\mathbf{E}}_{yi}}$$

Then we can write an upper bound for the first term

$$\begin{aligned}
&-tr(\mathbf{Q}(\mathbf{U}\circ\mathbf{E})(\mathbf{U}\circ\mathbf{E})) = -\sum_{xyi}\mathbf{Q}_{xy}\mathbf{U}_{xi}\mathbf{E}_{xi}\mathbf{U}_{yi}\mathbf{E}_{yi}\\
&\leq -\sum_{xyi}\mathbf{Q}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\tilde{\mathbf{E}}_{xi}\tilde{\mathbf{E}}_{yi}(1+\log\frac{\mathbf{E}_{xi}\mathbf{E}_{yi}}{\tilde{\mathbf{E}}_{xi}\tilde{\mathbf{E}}_{yi}})
\end{aligned} \tag{30}$$

For the second term, using the following property of Hadamard product (i.e., entry-wise product) [5]

$$tr((\mathbf{P}\circ\mathbf{V})^T(\mathbf{P}\circ\mathbf{M})) = tr((\mathbf{P}\circ\mathbf{V})^T\mathbf{M})$$
$$\forall\ \mathbf{P}\in\{0,1\}^{n\times n},\ \mathbf{V}\in\mathbb{R}_+^{n\times n},\ \mathbf{M}\in\mathbb{R}_+^{n\times n}$$

We can write

$$\begin{aligned}
&tr(\mathbf{R}\mathbf{R}) = tr\big(\mathbf{R}[\mathbf{H}(\mathbf{U}\circ\mathbf{E})(\mathbf{U}\circ\mathbf{E})^T\mathbf{H}^T]\big)\\
&= \sum_{xyzirpjq}\mathbf{C}_{xy}\mathbf{H}_{xz}\mathbf{U}_{zi}\mathbf{E}_{zi}\mathbf{U}_{ri}\mathbf{E}_{ri}\mathbf{H}_{yr}\mathbf{H}_{xp}\mathbf{U}_{pj}\mathbf{E}_{pj}\mathbf{U}_{qj}\mathbf{E}_{qj}\mathbf{H}_{yq}
\end{aligned} \tag{31}$$

Let $\mathcal{F}(\mathbf{E}_{zi}, \mathbf{E}_{ri}, \mathbf{E}_{pj}, \mathbf{E}_{qj})$ be a function w.r.t. $\mathbf{E}_{zi}$, $\mathbf{E}_{ri}$, $\mathbf{E}_{pj}$ and $\mathbf{E}_{qj}$ as defined by

$$\begin{aligned}
&\mathcal{F}(\mathbf{E}_{zi}, \mathbf{E}_{ri}, \mathbf{E}_{pj}, \mathbf{E}_{qj})\\
&= \mathbf{C}_{xy}\mathbf{H}_{xz}\mathbf{U}_{zi}\mathbf{E}_{zi}\mathbf{U}_{ri}\mathbf{E}_{ri}\mathbf{H}_{yr}\mathbf{H}_{xp}\mathbf{U}_{pj}\mathbf{E}_{pj}\mathbf{U}_{qj}\mathbf{E}_{qj}\mathbf{H}_{yq}
\end{aligned}$$

Also, let $\mathbf{E}_{zi} = \tilde{\mathbf{E}}_{zi}v_{zi}$, $\mathbf{E}_{ri} = \tilde{\mathbf{E}}_{ri}v_{ri}$, $\mathbf{E}_{pj} = \tilde{\mathbf{E}}_{pj}v_{pj}$ and $\mathbf{E}_{qj} = \tilde{\mathbf{E}}_{qj}v_{qj}$ for non-negative values v_{zi}, v_{ri}, v_{pj} and v_{qj}, we can rewrite Eq. (31) by

$$\begin{aligned}
&tr(\mathbf{R}\mathbf{R}) = \sum_{xyzirpjq}\mathcal{F}(\tilde{\mathbf{E}}_{zi}, \tilde{\mathbf{E}}_{ri}, \tilde{\mathbf{E}}_{pj}, \tilde{\mathbf{E}}_{qj})v_{zi}v_{ri}v_{pj}v_{qj}\\
&\leq \sum_{xyzirpjq}\mathcal{F}(\tilde{\mathbf{E}}_{zi}, \tilde{\mathbf{E}}_{ri}, \tilde{\mathbf{E}}_{pj}, \tilde{\mathbf{E}}_{qj})\frac{v_{zi}^4+v_{ri}^4+v_{pj}^4+v_{qj}^4}{4}\\
&= \frac{1}{4}\Big(\sum_{zi}(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{zi}\frac{\mathbf{E}_{zi}^4}{\tilde{\mathbf{E}}_{zi}^3} + \sum_{ri}(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{ri}\frac{\mathbf{E}_{ri}^4}{\tilde{\mathbf{E}}_{ri}^3}\\
&+ \sum_{pj}(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{pj}\frac{\mathbf{E}_{pj}^4}{\tilde{\mathbf{E}}_{pj}^3} + \sum_{qj}(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{qj}\frac{\mathbf{E}_{qj}^4}{\tilde{\mathbf{E}}_{qj}^3}\Big)\\
&= \sum_{xi}(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{xi}\frac{\mathbf{E}_{xi}^4}{\tilde{\mathbf{E}}_{xi}^3}
\end{aligned} \tag{32}$$

where $\tilde{\mathbf{\Theta}} = \mathbf{H}^T\{\mathbf{C}\circ[\mathbf{H}(\mathbf{U}\circ\tilde{\mathbf{E}})(\mathbf{U}\circ\tilde{\mathbf{E}})^T\mathbf{H}^T]\}\mathbf{H}(\mathbf{U}\circ\tilde{\mathbf{E}})$. The last equation is obtained by switching indexes.

For the third term, using the fact that $2ab \leq a^2 + b^2$, we have

$$\tau\sum_{xi}\mathbf{E}_{xi} \leq \frac{\tau}{2}\sum_{xi}\frac{\mathbf{E}_{xi}^2+\tilde{\mathbf{E}}_{xi}^2}{\tilde{\mathbf{E}}_{xi}} \leq \frac{\tau}{4}\sum_{xi}\frac{\mathbf{E}_{xi}^4+3\tilde{\mathbf{E}}_{xi}^4}{\tilde{\mathbf{E}}_{xi}^3} \tag{33}$$

Therefore, by collecting Eq. (30), Eq. (32) and Eq. (33), we obtain an auxiliary function for $\mathcal{J}_H(\mathbf{E})$ as

$$\begin{aligned}
&\mathcal{Z}_E(\mathbf{E}, \tilde{\mathbf{E}}) = -2\sum_{xyi}\mathbf{Q}_{xy}\mathbf{U}_{xi}\mathbf{U}_{yi}\tilde{\mathbf{E}}_{xi}\tilde{\mathbf{E}}_{yi}(1+\log\frac{\mathbf{E}_{xi}\mathbf{E}_{yi}}{\tilde{\mathbf{E}}_{xi}\tilde{\mathbf{E}}_{yi}})\\
&+ \sum_{xi}(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{xi}\frac{\mathbf{E}_{xi}^4}{\tilde{\mathbf{E}}_{xi}^3} + \frac{\tau}{4}\sum_{xi}\frac{\mathbf{E}_{xi}^4+3\tilde{\mathbf{E}}_{xi}^4}{\tilde{\mathbf{E}}_{xi}^3}
\end{aligned}$$

which is a tight upper bound satisfies $\mathcal{Z}_E(\mathbf{E}, \tilde{\mathbf{E}}) = \mathcal{J}_H(\mathbf{E})$. Then the derivative

$$\begin{aligned}
\frac{\partial\mathcal{Z}_E(\mathbf{E}_{xi}, \tilde{\mathbf{E}}_{xi})}{\partial\mathbf{E}_{xi}} &= -4\{[\mathbf{Q}(\mathbf{U}\circ\tilde{\mathbf{E}})]\circ\mathbf{U}\}_{xi}\frac{\tilde{\mathbf{E}}_{xi}}{\mathbf{E}_{xi}}\\
&+ 4(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{xi}\frac{\mathbf{E}_{xi}^3}{\tilde{\mathbf{E}}_{xi}^3} + \tau\frac{\mathbf{E}_{xi}^3}{\tilde{\mathbf{E}}_{xi}^3}
\end{aligned}$$

By setting $\frac{\partial\mathcal{Z}_E(\mathbf{E}_{xi}, \tilde{\mathbf{E}}_{xi})}{\partial\mathbf{E}_{xi}} = 0$, we obtain the optimal solution

$$\mathbf{E}_{xi} = \tilde{\mathbf{E}}_{xi}\left(\frac{4(\tilde{\mathbf{\Phi}}\circ\mathbf{U})_{xi}}{4(\tilde{\mathbf{\Theta}}\circ\mathbf{U})_{xi}+\tau}\right)^{\frac{1}{4}} \tag{34}$$

where $\tilde{\mathbf{\Theta}} = \mathbf{H}^T\{\mathbf{C}\circ[\mathbf{H}(\mathbf{U}\circ\tilde{\mathbf{E}})(\mathbf{U}\circ\tilde{\mathbf{E}})^T\mathbf{H}^T]\}\mathbf{H}(\mathbf{U}\circ\tilde{\mathbf{E}})$ and $\tilde{\mathbf{\Phi}} = \mathbf{H}^T(\mathbf{B}\circ\mathbf{C})\mathbf{H}(\mathbf{U}\circ\tilde{\mathbf{E}})$.

Note Eq. (34) is consistent with the updating rule of Eq. (19) in Theorem 3.4. Because of the properties of the auxiliary function [2], i.e.,

$$(1)\ \mathcal{Z}_E(\mathbf{E}, \tilde{\mathbf{E}}) \geq \mathcal{J}_H(\mathbf{E}),\ (2)\ \mathcal{Z}_E(\mathbf{E}, \mathbf{E}) = \mathcal{J}_H(\mathbf{E})$$

We have

$$\mathcal{J}_H(\mathbf{E}^{(t)}) = \mathcal{Z}_E(\mathbf{E}^{(t)}, \mathbf{E}^{(t)}) \geq \mathcal{Z}_E(\mathbf{E}^{(t+1)}, \mathbf{E}^{(t)}) \geq \mathcal{J}_H(\mathbf{E}^{(t+1)})$$

which completes the proof of Theorem 3.4. □

5. REFERENCES

[1] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.

[2] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.

[3] Z. Yang and E. Oja. Linear and nonlinear projective nonnegative matrix factorization. *IEEE Trans. Neural Netw.*, 21(5):734–749, 2010.

[4] Z. Yang and E. Oja. Clustering by low-rank doubly stochastic matrix decomposition. In *ICML*, 2012.

[5] Y. Zhang and D.-Y. Yeung. Overlapping community detection via bounded nonnegative matrix tri-factorization. In *KDD*, 2012.