Supplementary Material: Automated Medical Diagnosis by Ranking Clusters Across the Symptom-Disease Network

Jingchao Ni [1], Hongliang Fei [2], Wei Fan [2], and Xiang Zhang [1]
[1]College of Information Sciences and Technology, Pennsylvania State University, [2]Baidu Research Big Data Lab
[1]{jzn47, xzhang}@ist.psu.edu, [2]{hongliangfei, fanwei03}@baidu.com

1. GRADIENT DEFINITIONS

Let $\mathcal{J}(\mathbf{H}^{(i)})$ and $\mathcal{J}(\mathbf{S}^{(ij)})$ be the objective function in Eq. (6) w.r.t. $\mathbf{H}^{(i)}$ and $\mathbf{S}^{(ij)}$ when fixing other variables as constants, respectively. Then $\mathcal{J}(\mathbf{H}^{(i)})$ and $\mathcal{J}(\mathbf{S}^{(ij)})$ can be written as following

$$\begin{aligned}
\mathcal{J}(\mathbf{H}^{(i)}) &= -\sum_{xy}\mathbf{A}_{xy}^{(i)}\log\hat{\mathbf{A}}_{xy}^{(i)} - (\alpha-1)\sum_{xu}\log\mathbf{H}_{xu}^{(i)} \\
&\quad -\beta\sum_{j,j\neq i}\sum_{xv}(\tilde{\mathbf{B}}^{(ij)}\tilde{\mathbf{H}}^{(j)})_{xv}\log(\mathbf{H}^{(i)}\mathbf{S}^{(ij)})_{xv} \\
&\quad -\beta\sum_{j,j\neq i}\sum_{yu}(\tilde{\mathbf{B}}^{(ji)}\mathbf{H}^{(i)})_{yu}\log(\tilde{\mathbf{H}}^{(j)}\mathbf{S}^{(ji)})_{yu} \\
\mathcal{J}(\mathbf{S}^{(ij)}) &= -\beta\sum_{xv}(\tilde{\mathbf{B}}^{(ij)}\mathbf{H}^{(j)})_{xv}\log(\mathbf{H}^{(i)}\mathbf{S}^{(ij)})_{xv}
\end{aligned}$$

where $\hat{\mathbf{A}}_{xy}^{(i)}$ is defined in Eq. (1)[1].

Then the positive and non-positive parts of the gradients of $\mathcal{J}(\mathbf{H}^{(i)})$ w.r.t. $\mathbf{H}^{(i)}$ and those of $\mathcal{J}(\mathbf{S}^{(ij)})$ w.r.t. $\mathbf{S}^{(ij)}$ are given as following

$$\begin{aligned}
(\nabla_H^{(i)})_{xu}^{+} &= ((\mathbf{H}^{(i)})^T\mathbf{W}^{(i)}\mathbf{H}^{(i)})_{uu}(s_u^{(i)})^{-2} + (\mathbf{H}_{xu}^{(i)})^{-1} \\
&\quad -\beta\sum_{j,\ j\neq i}((\tilde{\mathbf{B}}^{(ji)})^T\mathbf{Q}^{(ji)})_{xu} \\
(\nabla_H^{(i)})_{xu}^{-} &= 2(\mathbf{W}^{(i)}\mathbf{H}^{(i)})_{xu}(s_r^{(i)})^{-1} + \alpha(\mathbf{H}_{xu}^{(i)})^{-1} \\
&\quad +\beta\sum_{j,\ j\neq i}(\mathbf{Y}^{(ij)}(\mathbf{S}^{(ij)})^T)_{xu} \\
(\nabla_S^{(ij)})_{uv}^{+} &= (\mathbf{S}_{uv}^{(ij)})^{-1} \\
(\nabla_S^{(ij)})_{uv}^{-} &= \beta((\mathbf{H}^{(i)})^T\mathbf{Y}^{(ij)})_{uv} + (\mathbf{S}_{uv}^{(ij)})^{-1}
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{W}_{xy}^{(i)} &= \mathbf{A}_{xy}^{(i)}/\hat{\mathbf{A}}_{xy}^{(i)}, \quad s_u^{(i)} = \sum_{x=1}^{n_i}\mathbf{H}_{xu}^{(i)} \\
\mathbf{Q}_{yu}^{(ji)} &= \log(\mathbf{H}^{(j)}\mathbf{S}^{(ji)})_{yu}, \quad \mathbf{Y}_{xv}^{(ij)} = \frac{(\tilde{\mathbf{B}}^{(ij)}\mathbf{H}^{(j)})_{xv}}{(\mathbf{H}^{(i)}\mathbf{S}^{(ij)})_{xv}}
\end{aligned}$$

2. PROOF OF THEOREM 3.1

We use an auxiliary function approach [1] to prove Theorem 3.1. First, we introduce the definition of an auxiliary function as below.

Definition 1. *[1] A function $Z(u,\tilde{u})$ is an* auxiliary function *for a given function $\mathcal{L}(u)$ if the conditions $Z(u,\tilde{u}) \geq \mathcal{L}(u)$ and $Z(u,u) = \mathcal{L}(u)$ are satisfied.*

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

Next, we provide the proof for Theorem 3.1.

Proof. In the following, we use $\mathbf{H}^{(i)}$ to denote variable and $\tilde{\mathbf{H}}^{(i)}$ to denote current estimate (i.e., a constant).

Let $\mathcal{L}_H(\mathbf{H}^{(i)})$ be the function in Eq. (8) when fixing $\boldsymbol{\lambda}^{(i)}$ as constant by Eq. (11), which can be written by

$$\begin{aligned}
\mathcal{L}_H(\mathbf{H}^{(i)}) &= -\sum_{xy}\mathbf{A}_{xy}^{(i)}\log\hat{\mathbf{A}}_{xy}^{(i)} - (\alpha-1)\sum_{xu}\log\mathbf{H}_{xu}^{(i)} \\
&\quad -\beta\sum_{j,j\neq i}\sum_{xv}(\tilde{\mathbf{B}}^{(ij)}\tilde{\mathbf{H}}^{(j)})_{xv}\log(\mathbf{H}^{(i)}\mathbf{S}^{(ij)})_{xv} \\
&\quad -\beta\sum_{j,j\neq i}\sum_{yu}(\tilde{\mathbf{B}}^{(ji)}\mathbf{H}^{(i)})_{yu}\log(\tilde{\mathbf{H}}^{(j)}\mathbf{S}^{(ji)})_{yu} + \sum_{xu}\lambda_x^{(i)}(\mathbf{H}_{xu}^{(i)}-1)
\end{aligned}$$

First, we derive a tight upper bound for $\mathcal{L}_H(\mathbf{H}^{(i)})$. Let $\pi_{xyu}^{(i)} = \frac{\tilde{\mathbf{H}}_{pr}^{(i)}\tilde{\mathbf{H}}_{qr}^{(i)}}{\hat{\mathbf{A}}_{xy}^{(i)}\sum_{z=1}^{n_i}\tilde{\mathbf{H}}_{zr}^{(i)}}$ and $\rho_{xuv}^{(ij)} = \frac{\tilde{\mathbf{H}}_{xu}^{(i)}\mathbf{S}_{uv}^{(ij)}}{\sum_{r=1}^{k_i}\tilde{\mathbf{H}}_{ur}^{(i)}\mathbf{S}_{rv}^{(ij)}}$, using the convexity of $\log(\cdot)$, we have

$$\begin{aligned}
&\mathcal{L}_H(\mathbf{H}^{(i)}) \\
&\leq -\sum_{xyu}\mathbf{A}_{xy}^{(i)}\pi_{xyu}^{(i)}(\log\mathbf{H}_{xu}^{(i)} + \log\mathbf{H}_{yu}^{(i)} - \log\sum_{z}\mathbf{H}_{zu}^{(i)}) \\
&\quad -(\alpha-1)\sum_{xu}\log\mathbf{H}_{xu}^{(i)} - \beta\sum_{jxuv}(\tilde{\mathbf{B}}^{(ij)}\tilde{\mathbf{H}}^{(j)})_{xv}\rho_{xuv}^{(ij)}\log\mathbf{H}_{xu}^{(i)} \\
&\quad -\beta\sum_{jyu}(\tilde{\mathbf{B}}^{(ji)}\mathbf{H}^{(i)})_{yu}\log(\tilde{\mathbf{H}}^{(j)}\mathbf{S}^{(ji)})_{yu} + \sum_{xu}\lambda_x^{(i)}\mathbf{H}_{xu}^{(i)} + C_1
\end{aligned}$$

where C_1 is a constant. Using the fact that $q \geq 1 + \log q$ for any $q > 0$, we can further upper bound the first term in the above equation by

$$-\sum_{xyu}\mathbf{A}_{xy}^{(i)}\pi_{xyu}^{(i)}(\log\mathbf{H}_{xu}^{(i)} + \log\mathbf{H}_{yu}^{(i)} - \frac{\sum_z\mathbf{H}_{zu}^{(i)}}{\sum_z\tilde{\mathbf{H}}_{zu}^{(i)}}) + C_2$$

where C_2 is a constant. Therefore, we obtain the following function as an upper bound of $\mathcal{L}_H(\mathbf{H}^{(i)})$.

$$\begin{aligned}
&\mathcal{Z}_H(\mathbf{H}^{(i)},\tilde{\mathbf{H}}^{(i)}) \\
&= -\sum_{xyu}\mathbf{A}_{xy}^{(i)}\pi_{xyu}^{(i)}(\log\mathbf{H}_{xu}^{(i)} + \log\mathbf{H}_{yu}^{(i)} - \frac{\sum_z\mathbf{H}_{zu}^{(i)}}{\sum_z\tilde{\mathbf{H}}_{zu}^{(i)}}) \\
&\quad -(\alpha-1)\sum_{xu}\log\mathbf{H}_{xu}^{(i)} - \beta\sum_{jxuv}(\tilde{\mathbf{B}}^{(ij)}\tilde{\mathbf{H}}^{(j)})_{xv}\rho_{xuv}^{(ij)}\log\mathbf{H}_{xu}^{(i)} \\
&\quad -\beta\sum_{jyu}(\tilde{\mathbf{B}}^{(ji)}\mathbf{H}^{(i)})_{yu}\log(\tilde{\mathbf{H}}^{(j)}\mathbf{S}^{(ji)})_{yu} + \sum_{xu}\lambda_x^{(i)}\mathbf{H}_{xu}^{(i)} \\
&\quad +\sum_{xu}(\frac{1}{a_x^{(i)}} + \frac{1}{\tilde{\mathbf{H}}_{xu}^{(i)}})\tilde{\mathbf{H}}_{xu}^{(i)}(\frac{\mathbf{H}_{xu}^{(i)}}{\tilde{\mathbf{H}}_{xu}^{(i)}} - \log\frac{\mathbf{H}_{xu}^{(i)}}{\tilde{\mathbf{H}}_{xu}^{(i)}} - 1) + C_3
\end{aligned}$$

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 in the multiplicative updating rule. It is known as the "moving term" technique in existing work [2, 3].

Note $\mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})$ is a tight upper bound of $\mathcal{L}_H(\mathbf{H}^{(i)})$. That is, it satisfies (note some terms are hidden in C_3).

$$\mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)}) \geq \mathcal{L}_H(\mathbf{H}^{(i)}),\ \mathcal{Z}_H(\mathbf{H}^{(i)}, \mathbf{H}^{(i)}) = \mathcal{L}_H(\mathbf{H}^{(i)}) \quad (18)$$

Thus, according to Definition 1, $\mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})$ is an auxiliary function of $\mathcal{L}_H(\mathbf{H}^{(i)})$.

Next, we derive the minimal solution to $\mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})$. The gradient of $\mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})$ w.r.t. $\mathbf{H}^{(i)}$ is

$$\begin{aligned}
&\frac{\partial \mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})}{\partial \mathbf{H}^{(i)}_{xu}} \\
&= -2\sum_y \mathbf{A}^{(i)}_{xy}\pi^{(i)}_{xyu}\frac{1}{\mathbf{H}^{(i)}_{xu}} + 2\sum_{xy}\mathbf{A}^{(i)}_{xy}\pi^{(i)}_{xyu}\frac{1}{\sum_z \tilde{\mathbf{H}}^{(i)}_{zr}} - (\alpha-1)\frac{1}{\mathbf{H}^{(i)}_{xu}} \\
&- \beta\sum_{jv}(\tilde{\mathbf{B}}^{(ij)}\tilde{\mathbf{H}}^{(j)})_{xv}\rho^{(ij)}_{xuv}\frac{1}{\mathbf{H}^{(i)}_{xu}} - \beta\sum_{jy}\tilde{\mathbf{B}}^{(ji)}_{yx}\log(\tilde{\mathbf{H}}^{(j)}\mathbf{S}^{(ji)})_{yu} \\
&+ \lambda^{(i)}_x + (\frac{1}{a^{(i)}_x} + \frac{1}{\tilde{\mathbf{H}}^{(i)}_{xu}})\tilde{\mathbf{H}}^{(i)}_{xu}(\frac{1}{\tilde{\mathbf{H}}^{(i)}_{xu}} - \frac{1}{\mathbf{H}^{(i)}_{xu}})
\end{aligned}$$

Equivalently,

$$\begin{aligned}
\frac{\partial \mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})}{\partial \mathbf{H}^{(i)}_{xu}} &= (\nabla^{(i)}_H)^+_{xu} - \frac{1}{\tilde{\mathbf{H}}^{(i)}_{xu}} - \frac{\tilde{\mathbf{H}}^{(i)}_{xu}}{\mathbf{H}^{(i)}_{xu}}((\nabla^{(i)}_H)^-_{xu} - \frac{1}{\tilde{\mathbf{H}}^{(i)}_{xu}}) \\
&+ \lambda^{(i)}_x + (\frac{1}{a^{(i)}_x} + \frac{1}{\tilde{\mathbf{H}}^{(i)}_{xu}})\tilde{\mathbf{H}}^{(i)}_{xu}(\frac{1}{\tilde{\mathbf{H}}^{(i)}_{xu}} - \frac{1}{\mathbf{H}^{(i)}_{xu}}) \\
&= -\frac{\tilde{\mathbf{H}}^{(i)}_{xu}}{\mathbf{H}^{(i)}_{xu}}((\nabla^{(i)}_H)^-_{xu} + \frac{1}{a^{(i)}_x}) + ((\nabla^{(i)}_H)^+_{xu} + \frac{b^{(i)}_x}{a^{(i)}_x})
\end{aligned}$$

Therefore, the global minimum of $\mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})$ can be obtained by solving $\frac{\partial \mathcal{Z}_H(\mathbf{H}^{(i)}, \tilde{\mathbf{H}}^{(i)})}{\partial \mathbf{H}^{(i)}_{xu}} = 0$. This gives

$$\mathbf{H}^{(i)}_{xu} = \tilde{\mathbf{H}}^{(i)}_{xu}\frac{a^{(i)}_x(\nabla^{(i)}_H)^-_{xu} + 1}{a^{(i)}_x(\nabla^{(i)}_H)^+_{xu} + b^{(i)}_x}$$

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

Therefore, according to the properties of auxiliary function, i.e., Eq. (18), at any iteration $t \geq 1$, we have

$$\begin{aligned}
\mathcal{L}_H((\mathbf{H}^{(i)})^{(t)}) &= \mathcal{Z}_H((\mathbf{H}^{(i)})^{(t)}, (\mathbf{H}^{(i)})^{(t)}) \\
&\geq \mathcal{Z}_H((\mathbf{H}^{(i)})^{(t+1)}, (\mathbf{H}^{(i)})^{(t)}) \geq \mathcal{L}_H((\mathbf{H}^{(i)})^{(t+1)})
\end{aligned}$$

which completes the proof of Theorem 3.1. □

3. PROOF OF THEOREM 3.2

PROOF. The proof of Theorem 3.2 is similar to the proof of Theorem 3.1. In the following, we use $\mathbf{S}^{(ij)}$ to denote variable and $\tilde{\mathbf{S}}^{(ij)}$ to denote current estimate (i.e., a constant). First, the Lagrangian function $\mathcal{L}_S(\mathbf{S}^{(ij)}, \boldsymbol{\eta}^{(ij)})$ is

$$\mathcal{L}_S(\mathbf{S}^{(ij)}, \boldsymbol{\eta}^{(ij)}) = \mathcal{J}(\mathbf{S}^{(ij)}) + \sum_{u=1}^{k_i}\eta^{(ij)}_u(\sum_{v=1}^{k_j}\mathbf{S}^{(ij)}_{uv} - 1) \quad (19)$$

where $\boldsymbol{\eta}^{(ij)} = (\eta^{(ij)}_1, ..., \eta^{(ij)}_{k_i})^T$ are Lagrangian multipliers.

Let $\mathcal{L}_S(\mathbf{S}^{(ij)})$ be the function in Eq. (19) w.r.t. $\mathbf{S}^{(ij)}$ when fixing $\boldsymbol{\eta}^{(ij)}$ as constant by Eq. (14), which is

$$\begin{aligned}
\mathcal{L}_S(\mathbf{S}^{(ij)}) = &-\beta\sum_{xv}(\tilde{\mathbf{B}}^{(ij)}\mathbf{H}^{(j)})_{xv}\log(\mathbf{H}^{(i)}\mathbf{S}^{(ij)})_{xv} \\
&+ \sum_{uv}\eta^{(ij)}_u(\mathbf{S}^{(ij)}_{uv} - 1)
\end{aligned}$$

Then we drive a tight upper bound for $\mathcal{L}_S(\mathbf{S}^{(ij)})$. Let $\omega^{(ij)}_{xuv} = \frac{\mathbf{H}^{(i)}_{xu}\tilde{\mathbf{S}}^{(ij)}_{uv}}{\sum_u \mathbf{H}^{(i)}_{xu}\tilde{\mathbf{S}}^{(ij)}_{uv}}$, using the convexity of $\log(\cdot)$, we have

$$\begin{aligned}
&\mathcal{L}_S(\mathbf{S}^{(ij)}) \leq -\beta\sum_{xuv}(\tilde{\mathbf{B}}^{(ij)}\mathbf{H}^{(j)})_{xv}\omega^{(ij)}_{xuv}\log(\mathbf{H}^{(i)}_{xu}\mathbf{S}^{(ij)}_{uv}) \\
&+ \sum_{uv}\eta^{(ij)}_u\mathbf{S}^{(ij)}_{uv} + \sum_{uv}(\frac{1}{c^{(ij)}_u} + \frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}})\tilde{\mathbf{S}}^{(ij)}_{uv}(\frac{\mathbf{S}^{(ij)}_{uv}}{\tilde{\mathbf{S}}^{(ij)}_{uv}} - \log\frac{\mathbf{S}^{(ij)}_{uv}}{\tilde{\mathbf{S}}^{(ij)}_{uv}} - 1) \\
&+ C_4
\end{aligned}$$

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

Let $\mathcal{Z}_H(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})$ be the above upper bound, it is a tight upper bound of $\mathcal{L}_S(\mathbf{S}^{(ij)})$. That is, it satisfies (note some terms are hidden in C_4).

$$\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)}) \geq \mathcal{L}_S(\mathbf{S}^{(ij)}),\ \mathcal{Z}_S(\mathbf{S}^{(ij)}, \mathbf{S}^{(ij)}) = \mathcal{L}_S(\mathbf{S}^{(ij)}) \quad (20)$$

Thus, according to Definition 1, $\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})$ is an auxiliary function of $\mathcal{L}_S(\mathbf{S}^{(ij)})$.

Next, we derive the minimal solution to $\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})$. The gradient of $\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})$ w.r.t. $\mathbf{S}^{(ij)}$ is

$$\begin{aligned}
\frac{\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})}{\partial \mathbf{S}^{(ij)}_{uv}} &= -\beta\sum_x(\tilde{\mathbf{B}}^{(ij)}\mathbf{H}^{(j)})_{xv}\omega^{(ij)}_{xuv}\frac{1}{\mathbf{S}^{(ij)}_{uv}} \\
&+ \eta^{(ij)}_u + (\frac{1}{c^{(ij)}_u} + \frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}})\tilde{\mathbf{S}}^{(ij)}_{uv}(\frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}} - \frac{1}{\mathbf{S}^{(ij)}_{uv}})
\end{aligned}$$

Equivalently,

$$\begin{aligned}
&\frac{\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})}{\partial \mathbf{S}^{(ij)}_{uv}} = (\nabla^{(ij)}_S)^+_{uv} - \frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}} - \frac{\tilde{\mathbf{S}}^{(ij)}_{uv}}{\mathbf{S}^{(ij)}_{uv}}((\nabla^{(ij)}_S)^-_{uv} - \frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}}) \\
&+ \eta^{(ij)}_u + (\frac{1}{c^{(ij)}_u} + \frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}})\tilde{\mathbf{S}}^{(ij)}_{uv}(\frac{1}{\tilde{\mathbf{S}}^{(ij)}_{uv}} - \frac{1}{\mathbf{S}^{(ij)}_{uv}}) \\
&= -\frac{\tilde{\mathbf{S}}^{(ij)}_{uv}}{\mathbf{S}^{(ij)}_{uv}}((\nabla^{(ij)}_S)^-_{uv} + \frac{1}{c^{(ij)}_u}) + ((\nabla^{(ij)}_S)^+_{uv} + \frac{d^{(ij)}_u}{c^{(ij)}_u})
\end{aligned}$$

Therefore, the global minimum of $\mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})$ can be obtained by solving $\frac{\partial \mathcal{Z}_S(\mathbf{S}^{(ij)}, \tilde{\mathbf{S}}^{(ij)})}{\partial \mathbf{S}^{(ij)}_{uv}} = 0$. This gives

$$\mathbf{S}^{(ij)}_{uv} = \tilde{\mathbf{S}}^{(ij)}_{uv}\frac{c^{(ij)}_u(\nabla^{(ij)}_S)^-_{uv} + 1}{c^{(ij)}_u(\nabla^{(ij)}_S)^+_{uv} + d^{(ij)}_u}$$

which is consistent with the updating rule in Eq. (16).

Therefore, according to the properties of auxiliary function, i.e., Eq. (20), at any iteration $t \geq 1$, we have

$$\begin{aligned}
\mathcal{L}_S((\mathbf{S}^{(ij)})^{(t)}) &= \mathcal{Z}_S((\mathbf{S}^{(ij)})^{(t)}, (\mathbf{S}^{(ij)})^{(t)}) \\
&\geq \mathcal{Z}_S((\mathbf{S}^{(ij)})^{(t+1)}, (\mathbf{S}^{(ij)})^{(t)}) \geq= \mathcal{L}_S((\mathbf{S}^{(ij)})^{(t+1)})
\end{aligned}$$

which completes the proof of Theorem 3.2. □

4. REFERENCES

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

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

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