A Fine-Grained Regularization Scheme for Nonnegative Latent Factorization of High-Dimensional and Incomplete Tensors Supplementary File

Hao Wu, *Member, IEEE*, and Xin Luo, *Senior Member*, *IEEE*

I. Introduction

This is the supplementary file for paper entitled *"A Fine-Grained Regularization Scheme for Nonnegative Latent Factorization of High-Dimensional and Incomplete Tensors"*. We have put the convergence proof of FRNL model in Section 3, and the supplementary table of empirical studies in Section 4.

II. Convergence Proof of FRNL

Considering the nonnegative constraints for latent feature matrices S, D, and T and linear bias vectors $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$, we have the Lagrangian function L for (6) as:

$$L=\varepsilon\left(\mathrm{S},\mathrm{D},\mathrm{T},\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}\right)-\sum_{i=1}^{|I|}\sum_{r=1}^{R}\tilde{s}_{ir}s_{ir}-\sum_{j=1}^{|J|}\sum_{r=1}^{R}\tilde{d}_{jr}d_{jr}-\sum_{k=1}^{|K|}\sum_{r=1}^{R}\tilde{t}_{kr}t_{kr}-\sum_{i=1}^{|I|}\tilde{a}_{i}a_{i}-\sum_{j=1}^{|J|}\tilde{b}_{j}b_{j}-\sum_{k=1}^{|K|}\tilde{c}_{k}c_{k}. \tag{S1}$$

where $\tilde{\mathrm{S}}$, $\tilde{\mathrm{D}}$, $\tilde{\mathrm{T}}$ $\tilde{\boldsymbol{a}}$, $\tilde{\boldsymbol{b}}$, and $\tilde{\boldsymbol{c}}$ denote Lagrangian multipliers for S, D, T, $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{c}$.

Considering the partial derivatives of L with latent feature and linear bias, they are highly similar for S, D, T and $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$. Hence, we consider the case of s_{ir} and a_i as follows:

$$\begin{cases}\dfrac{\partial L}{\partial s_{ir}}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta s_{ir}^{p}e^{-\beta s_{ir}}+s_{ir}^{p-1}+s_{ir}^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}}\right)^{p+1}-\tilde{s}_{ir}=0,\\ \dfrac{\partial L}{\partial a_{i}}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta a_{i}^{p}e^{-\beta a_{i}}+a_{i}^{p-1}+a_{i}^{p-1}e^{-\beta a_{i}}\right)\Big/\left(1+e^{-\beta a_{i}}\right)^{p+1}-\tilde{a}_{i}=0.\end{cases}$$
$$\Rightarrow\begin{cases}\tilde{s}_{ir}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta s_{ir}^{p}e^{-\beta s_{ir}}+s_{ir}^{p-1}+s_{ir}^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}}\right)^{p+1},\\ \tilde{a}_{i}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\cdot\left(\beta s_{ir}^{p}e^{-\beta s_{ir}}+s_{ir}^{p-1}+s_{ir}^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}}\right)^{p+1}.\end{cases} \tag{S2}$$

Then, considering the KKT conditions of (S1), i.e., $\forall s_{ir}$, $\tilde{s}_{ir}$: $s_{ir}\tilde{s}_{ir}=0$, and $\forall a_i$, $\tilde{a}_i$: $a_i\tilde{a}_i=0$, we have:

$$\begin{cases}s_{ir}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta s_{ir}^{p}e^{-\beta s_{ir}}+s_{ir}^{p-1}+s_{ir}^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}}\right)^{p+1}\right)=0,\\ a_{i}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta a_{i}^{p}e^{-\beta a_{i}}+a_{i}^{p-1}+a_{i}^{p-1}e^{-\beta a_{i}}\right)\Big/\left(1+e^{-\beta a_{i}}\right)^{p+1}\right)=0.\end{cases}$$
$$\Rightarrow\begin{cases}s_{ir}\sum\limits_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}=s_{ir}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta s_{ir}^{p}e^{-\beta s_{ir}}+s_{ir}^{p-1}+s_{ir}^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}}\right)^{p+1}\right),\\ a_{i}\sum\limits_{y_{ijk}\in\Lambda(i)}y_{ijk}=a_{i}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta a_{i}^{p}e^{-\beta a_{i}}+a_{i}^{p-1}+a_{i}^{p-1}e^{-\beta a_{i}}\right)\Big/\left(1+e^{-\beta a_{i}}\right)^{p+1}\right).\end{cases} \tag{S3}$$

With (S3), we conveniently achieve the iterative learning rules given in (16). Hence, an SLF-NMU-based learning scheme in an FRNL model is closely connected to the KKT conditions of its learning objective. From this point of view, we theoretically prove the convergence of FRNL in the following two steps:

Step 1: The objective function (6) is non-increasing and lower-bounded.

Step 2: FRNL is guaranteed to converge at a KKT stationary point of its learning objective with the SFR scheme and SLF-NMU-based learning rules.

To implement Step 1, we present the following *Lemma* 1.

***Lemma* 1**. With the following definitions,

$$\tau_1 = \begin{pmatrix} \left(s_{ir}^{n+1}-s_{ir}^{n}\right)^2 \left(\sum_{y_{ijk}\in\Lambda(i)} \left(d_{jr}^{n} t_{kr}^{n}\right)^2 + \gamma\left|\Lambda(i)\right|^{\alpha} p \frac{(p-1)\left(s_{ir}^{n+1}\right)^{p-2}\left(1+e^{-\beta s_{ir}^{n+1}}\right)^2 + \beta\left(s_{ir}^{n+1}\right)^{p-1} e^{-\beta s_{ir}^{n+1}}\left(p+2pe^{-\beta s_{ir}^{n+1}}+1\right) + p\beta^2\left(s_{ir}^{n+1}\right)^{p} e^{-2\beta s_{ir}^{n+1}}}{\left(1+e^{-\beta s_{ir}^{n+1}}\right)^{p+2}} \right) \\ +\left(d_{jr}^{n+1}-d_{jr}^{n}\right)^2 \left(\sum_{y_{ijk}\in\Lambda(j)} \left(s_{ir}^{n+1} t_{kr}^{n}\right)^2 + \gamma\left|\Lambda(j)\right|^{\alpha} p \frac{(p-1)\left(d_{jr}^{n+1}\right)^{p-2}\left(1+e^{-\beta d_{jr}^{n+1}}\right)^2 + \beta\left(d_{jr}^{n+1}\right)^{p-1} e^{-\beta d_{jr}^{n+1}}\left(p+2pe^{-\beta d_{jr}^{n+1}}+1\right) + p\beta^2\left(d_{jr}^{n+1}\right)^{p} e^{-2\beta d_{jr}^{n+1}}}{\left(1+e^{-\beta d_{jr}^{n+1}}\right)^{p+2}} \right) \\ +\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^2 \left(\sum_{y_{ijk}\in\Lambda(k)} \left(s_{ir}^{n+1} d_{jr}^{n+1}\right)^2 + \gamma\left|\Lambda(k)\right|^{\alpha} p \frac{(p-1)\left(t_{kr}^{n+1}\right)^{p-2}\left(1+e^{-\beta t_{kr}^{n+1}}\right)^2 + \beta\left(t_{kr}^{n+1}\right)^{p-1} e^{-\beta t_{kr}^{n+1}}\left(p+2pe^{-\beta t_{kr}^{n+1}}+1\right) + p\beta^2\left(t_{kr}^{n+1}\right)^{p} e^{-2\beta t_{kr}^{n+1}}}{\left(1+e^{-\beta t_{kr}^{n+1}}\right)^{p+2}} \right) \\ +\left(a_{i}^{n+1}-a_{i}^{n}\right)^2 \left(\left|\Lambda(i)\right| + \gamma\left|\Lambda(i)\right|^{\alpha} p \frac{(p-1)\left(a_{i}^{n+1}\right)^{p-2}\left(1+e^{-\beta a_{i}^{n+1}}\right)^2 + \beta\left(a_{i}^{n+1}\right)^{p-1} e^{-\beta a_{i}^{n+1}}\left(p+2pe^{-\beta a_{i}^{n+1}}+1\right) + p\beta^2\left(a_{i}^{n+1}\right)^{p} e^{-2\beta a_{i}^{n+1}}}{\left(1+e^{-\beta a_{i}^{n+1}}\right)^{p+2}} \right) \\ +\left(b_{j}^{n+1}-b_{j}^{n}\right)^2 \left(\left|\Lambda(j)\right| + \gamma\left|\Lambda(j)\right|^{\alpha} p \frac{(p-1)\left(b_{j}^{n+1}\right)^{p-2}\left(1+e^{-\beta b_{j}^{n+1}}\right)^2 + \beta\left(b_{j}^{n+1}\right)^{p-1} e^{-\beta b_{j}^{n+1}}\left(p+2pe^{-\beta b_{j}^{n+1}}+1\right) + p\beta^2\left(b_{j}^{n+1}\right)^{p} e^{-2\beta b_{j}^{n+1}}}{\left(1+e^{-\beta b_{j}^{n+1}}\right)^{p+2}} \right) \\ +\left(c_{k}^{n+1}-c_{k}^{n}\right)^2 \left(\left|\Lambda(k)\right| + \gamma\left|\Lambda(k)\right|^{\alpha} p \frac{(p-1)\left(c_{k}^{n+1}\right)^{p-2}\left(1+e^{-\beta c_{k}^{n+1}}\right)^2 + \beta\left(c_{k}^{n+1}\right)^{p-1} e^{-\beta c_{k}^{n+1}}\left(p+2pe^{-\beta c_{k}^{n+1}}+1\right) + p\beta^2\left(c_{k}^{n+1}\right)^{p} e^{-2\beta c_{k}^{n+1}}}{\left(1+e^{-\beta c_{k}^{n+1}}\right)^{p+2}} \right) \end{pmatrix},$$

and

$$\tau_2 = \begin{pmatrix} \left(s_{ir}^{n+1}-s_{ir}^{n}\right)^2 \left(\gamma\left|\Lambda(i)\right|^{\alpha} p\beta^2\left(s_{ir}^{n+1}\right)^{p} e^{-\beta s_{ir}^{n+1}}\right) \Big/ \left(1+e^{-\beta s_{ir}^{n+1}}\right)^{p+2} + \left(a_{i}^{n+1}-a_{i}^{n}\right)^2 \left(\gamma\left|\Lambda(i)\right|^{\alpha} p\beta^2\left(a_{i}^{n+1}\right)^{p} e^{-\beta a_{i}^{n+1}}\right) \Big/ \left(1+e^{-\beta a_{i}^{n+1}}\right)^{p+2} \\ +\left(d_{jr}^{n+1}-d_{jr}^{n}\right)^2 \left(\gamma\left|\Lambda(j)\right|^{\alpha} p\beta^2\left(d_{jr}^{n+1}\right)^{p} e^{-\beta d_{jr}^{n+1}}\right) \Big/ \left(1+e^{-\beta d_{jr}^{n+1}}\right)^{p+2} + \left(b_{j}^{n+1}-b_{j}^{n}\right)^2 \left(\gamma\left|\Lambda(j)\right|^{\alpha} p\beta^2\left(b_{j}^{n+1}\right)^{p} e^{-\beta b_{j}^{n+1}}\right) \Big/ \left(\left(1+e^{-\beta b_{j}^{n+1}}\right)^{p+2}\right) \\ +\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^2 \left(\gamma\left|\Lambda(k)\right|^{\alpha} p\beta^2\left(t_{kr}^{n+1}\right)^{p} e^{-\beta t_{kr}^{n+1}}\right) \Big/ \left(\left(1+e^{-\beta t_{kr}^{n+1}}\right)^{p+2}\right) + \left(c_{k}^{n+1}-c_{k}^{n}\right)^2 \left(\gamma\left|\Lambda(k)\right|^{\alpha} p\beta^2\left(c_{k}^{n+1}\right)^{p} e^{-\beta c_{k}^{n+1}}\right) \Big/ \left(\left(1+e^{-\beta c_{k}^{n+1}}\right)^{p+2}\right) \end{pmatrix},$$

if $\tau_1 \geq \tau_2$, then the following inequality holds:

$$\varepsilon\left(s_{ir}^{n+1}, d_{jr}^{n+1}, t_{kr}^{n+1}, a_{i}^{n+1}, b_{j}^{n+1}, c_{k}^{n+1}\right) - \varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right) \leq 0. \tag{S4}$$

Moreover, if $\gamma \geq 0$, we constantly have:

$$\varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right) \geq 0. \tag{S5}$$

Proof of *Lemma* 1. Firstly, considering the difference between $\varepsilon\left(s_{ir}^{n+1}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right)$ and $\varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right)$, we have:

$$\begin{aligned} &\varepsilon\left(s_{ir}^{n+1}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right) - \varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right) \\ &\overset{\Delta}{=} \left(\sum_{y_{ijk}\in\Lambda(i)} \left(y_{ijk} - \left(\sum_{r=1}^{R} s_{ir}^{n+1} d_{jr}^{n} t_{kr}^{n} + a_{i}^{n} + b_{j}^{n} + c_{k}^{n} \right) \right) \left(-d_{jr}^{n} t_{kr}^{n}\right) + \gamma\left|\Lambda(i)\right|^{\alpha} p \frac{\left(s_{ir}^{n+1}\right)^{p-1} + \left(s_{ir}^{n+1}\right)^{p-1} e^{-\beta s_{ir}^{n+1}} + \beta\left(s_{ir}^{n+1}\right)^{p} e^{-\beta s_{ir}^{n+1}}}{\left(1+e^{-\beta s_{ir}}\right)^{p+1}} \right) \left(s_{ir}^{n+1}-s_{ir}^{n}\right) \\ &-\frac{1}{2} \left(\sum_{y_{ijk}\in\Lambda(i)} \left(d_{jr}^{n} t_{kr}^{n}\right)^2 + \gamma\left|\Lambda(i)\right|^{\alpha} p \frac{(p-1)\left(s_{ir}^{n+1}\right)^{p-2}\left(1+e^{-\beta s_{ir}^{n+1}}\right)^2 + \beta\left(s_{ir}^{n+1}\right)^{p-1} e^{-\beta s_{ir}^{n+1}}\left(p+2pe^{-\beta s_{ir}^{n+1}}+1\right) + \beta^2\left(s_{ir}^{n+1}\right)^{p} e^{-\beta s_{ir}^{n+1}}\left(pe^{-\beta s_{ir}^{n+1}}-1\right)}{\left(1+e^{-\beta s_{ir}^{n+1}}\right)^{p+2}} \right) \left(s_{ir}^{n+1}-s_{ir}^{n}\right)^2. \end{aligned} \tag{S6}$$

where $\overset{\Delta}{=}$ denotes the second-order approximation of a function. Based on SLF-NMU, considering s_{ir}'s optimal condition, (S6) is reformulated as:

$$\begin{aligned} &\varepsilon\left(s_{ir}^{n+1}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right) - \varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_{i}^{n}, b_{j}^{n}, c_{k}^{n}\right) \\ &= -\frac{1}{2} \left(\sum_{y_{ijk}\in\Lambda(i)} \left(d_{jr}^{n} t_{kr}^{n}\right)^2 + \gamma\left|\Lambda(i)\right|^{\alpha} p \frac{(p-1)\left(s_{ir}^{n+1}\right)^{p-2}\left(1+e^{-\beta s_{ir}^{n+1}}\right)^2 + \beta\left(s_{ir}^{n+1}\right)^{p-1} e^{-\beta s_{ir}^{n+1}}\left(p+2pe^{-\beta s_{ir}^{n+1}}+1\right) + \beta^2\left(s_{ir}^{n+1}\right)^{p} e^{-\beta s_{ir}^{n+1}}\left(pe^{-\beta s_{ir}^{n+1}}-1\right)}{\left(1+e^{-\beta s_{ir}^{n+1}}\right)^{p+2}} \right) \left(s_{ir}^{n+1}-s_{ir}^{n}\right)^2. \end{aligned} \tag{S7}$$

Similarly, we have:

$$\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)=$$
$$-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(j)}\left(s_{ir}^{n+1}t_{kr}^{n}\right)^{2}+\gamma\left|\Lambda(j)\right|^{\alpha}p\frac{(p-1)\left(d_{jr}^{n+1}\right)^{p-2}\left(1+e^{-\beta d_{jr}^{n+1}}\right)^{2}+\beta\left(d_{jr}^{n+1}\right)^{p-1}e^{-\beta d_{jr}^{n+1}}\left(p+2pe^{-\beta d_{jr}^{n+1}}+1\right)+\beta^{2}\left(d_{jr}^{n+1}\right)^{p}e^{-\beta d_{jr}^{n+1}}\left(pe^{-\beta d_{jr}^{n+1}}-1\right)}{\left(1+e^{-\beta d_{jr}^{n+1}}\right)^{p+2}}\right)\left(d_{jr}^{n+1}-d_{jr}^{n}\right)^{2}. \tag{S8}$$

$$\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)=$$
$$-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(k)}\left(s_{ir}^{n+1}d_{jr}^{n+1}\right)^{2}+\gamma\left|\Lambda(k)\right|^{\alpha}p\frac{(p-1)\left(t_{kr}^{n+1}\right)^{p-2}\left(1+e^{-\beta t_{kr}^{n+1}}\right)^{2}+\beta\left(t_{kr}^{n+1}\right)^{p-1}e^{-\beta t_{kr}^{n+1}}\left(p+2pe^{-\beta t_{kr}^{n+1}}+1\right)+\beta^{2}\left(t_{kr}^{n+1}\right)^{p}e^{-\beta t_{kr}^{n+1}}\left(pe^{-\beta t_{kr}^{n+1}}-1\right)}{\left(1+e^{-\beta t_{kr}^{n+1}}\right)^{p+2}}\right)\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^{2}. \tag{S9}$$

$$\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)=$$
$$-\frac{1}{2}\left(\left|\Lambda(i)\right|+\gamma\left|\Lambda(i)\right|^{\alpha}p\frac{(p-1)\left(a_{i}^{n+1}\right)^{p-2}\left(1+e^{-\beta a_{i}^{n+1}}\right)^{2}+\beta\left(a_{i}^{n+1}\right)^{p-1}e^{-\beta a_{i}^{n+1}}\left(p+2pe^{-\beta a_{i}^{n+1}}+1\right)+\beta^{2}\left(a_{i}^{n+1}\right)^{p}e^{-\beta a_{i}^{n+1}}\left(pe^{-\beta a_{i}^{n+1}}-1\right)}{\left(1+e^{-\beta a_{i}^{n+1}}\right)^{p+2}}\right)\left(a_{i}^{n+1}-a_{i}^{n}\right)^{2}. \tag{S10}$$

$$\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n+1},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n},c_{k}^{n}\right)=$$
$$-\frac{1}{2}\left(\left|\Lambda(j)\right|+\gamma\left|\Lambda(j)\right|^{\alpha}p\frac{(p-1)\left(b_{j}^{n+1}\right)^{p-2}\left(1+e^{-\beta b_{j}^{n+1}}\right)^{2}+\beta\left(b_{j}^{n+1}\right)^{p-1}e^{-\beta b_{j}^{n+1}}\left(p+2pe^{-\beta b_{j}^{n+1}}+1\right)+\beta^{2}\left(b_{j}^{n+1}\right)^{p}e^{-\beta b_{j}^{n+1}}\left(pe^{-\beta b_{j}^{n+1}}-1\right)}{\left(1+e^{-\beta b_{j}^{n+1}}\right)^{p+2}}\right)\left(b_{j}^{n+1}-b_{j}^{n}\right)^{2}. \tag{S11}$$

$$\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n+1},c_{k}^{n+1}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n+1},c_{k}^{n}\right)=$$
$$-\frac{1}{2}\left(\left|\Lambda(k)\right|+\gamma\left|\Lambda(k)\right|^{\alpha}p\frac{(p-1)\left(c_{k}^{n+1}\right)^{p-2}\left(1+e^{-\beta c_{k}^{n+1}}\right)^{2}+\beta\left(c_{k}^{n+1}\right)^{p-1}e^{-\beta c_{k}^{n+1}}\left(p+2pe^{-\beta c_{k}^{n+1}}+1\right)+\beta^{2}\left(c_{k}^{n+1}\right)^{p}e^{-\beta c_{k}^{n+1}}\left(pe^{-\beta c_{k}^{n+1}}-1\right)}{\left(1+e^{-\beta c_{k}^{n+1}}\right)^{p+2}}\right)\left(c_{k}^{n+1}-c_{k}^{n}\right)^{2}. \tag{S12}$$

With (S7)-(S12), we have:

$$\begin{aligned}
&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n+1},c_{k}^{n+1}\right)-\varepsilon\left(s_{ir}^{n},d_{jr}^{n},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)=\\
&-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(i)}\left(d_{jr}^{n}t_{kr}^{n}\right)^{2}+\gamma\left|\Lambda(i)\right|^{\alpha}p\frac{(p-1)\left(s_{ir}^{n+1}\right)^{p-2}\left(1+e^{-\beta s_{ir}^{n+1}}\right)^{2}+\beta\left(s_{ir}^{n+1}\right)^{p-1}e^{-\beta s_{ir}^{n+1}}\left(p+2pe^{-\beta s_{ir}^{n+1}}+1\right)+\beta^{2}\left(s_{ir}^{n+1}\right)^{p}e^{-\beta s_{ir}^{n+1}}\left(pe^{-\beta s_{ir}^{n+1}}-1\right)}{\left(1+e^{-\beta s_{ir}^{n+1}}\right)^{p+2}}\right)\left(s_{ir}^{n+1}-s_{ir}^{n}\right)^{2}\\
&-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(j)}\left(s_{ir}^{n+1}t_{kr}^{n}\right)^{2}+\gamma\left|\Lambda(j)\right|^{\alpha}p\frac{(p-1)\left(d_{jr}^{n+1}\right)^{p-2}\left(1+e^{-\beta d_{jr}^{n+1}}\right)^{2}+\beta\left(d_{jr}^{n+1}\right)^{p-1}e^{-\beta d_{jr}^{n+1}}\left(p+2pe^{-\beta d_{jr}^{n+1}}+1\right)+\beta^{2}\left(d_{jr}^{n+1}\right)^{p}e^{-\beta d_{jr}^{n+1}}\left(pe^{-\beta d_{jr}^{n+1}}-1\right)}{\left(1+e^{-\beta d_{jr}^{n+1}}\right)^{p+2}}\right)\left(d_{jr}^{n+1}-d_{jr}^{n}\right)^{2}\\
&-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(k)}\left(s_{ir}^{n+1}d_{jr}^{n+1}\right)^{2}+\gamma\left|\Lambda(k)\right|^{\alpha}p\frac{(p-1)\left(t_{kr}^{n+1}\right)^{p-2}\left(1+e^{-\beta t_{kr}^{n+1}}\right)^{2}+\beta\left(t_{kr}^{n+1}\right)^{p-1}e^{-\beta t_{kr}^{n+1}}\left(p+2pe^{-\beta t_{kr}^{n+1}}+1\right)+\beta^{2}\left(t_{kr}^{n+1}\right)^{p}e^{-\beta t_{kr}^{n+1}}\left(pe^{-\beta t_{kr}^{n+1}}-1\right)}{\left(1+e^{-\beta t_{kr}^{n+1}}\right)^{p+2}}\right)\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^{2}\\
&-\frac{1}{2}\left(\left|\Lambda(i)\right|+\gamma\left|\Lambda(i)\right|^{\alpha}p\frac{(p-1)\left(a_{i}^{n+1}\right)^{p-2}\left(1+e^{-\beta a_{i}^{n+1}}\right)^{2}+\beta\left(a_{i}^{n+1}\right)^{p-1}e^{-\beta a_{i}^{n+1}}\left(p+2pe^{-\beta a_{i}^{n+1}}+1\right)+\beta^{2}\left(a_{i}^{n+1}\right)^{p}e^{-\beta a_{i}^{n+1}}\left(pe^{-\beta a_{i}^{n+1}}-1\right)}{\left(1+e^{-\beta a_{i}^{n+1}}\right)^{p+2}}\right)\left(a_{i}^{n+1}-a_{i}^{n}\right)^{2}\\
&-\frac{1}{2}\left(\left|\Lambda(j)\right|+\gamma\left|\Lambda(j)\right|^{\alpha}p\frac{(p-1)\left(b_{j}^{n+1}\right)^{p-2}\left(1+e^{-\beta b_{j}^{n+1}}\right)^{2}+\beta\left(b_{j}^{n+1}\right)^{p-1}e^{-\beta b_{j}^{n+1}}\left(p+2pe^{-\beta b_{j}^{n+1}}+1\right)+\beta^{2}\left(b_{j}^{n+1}\right)^{p}e^{-\beta b_{j}^{n+1}}\left(pe^{-\beta b_{j}^{n+1}}-1\right)}{\left(1+e^{-\beta b_{j}^{n+1}}\right)^{p+2}}\right)\left(b_{j}^{n+1}-b_{j}^{n}\right)^{2}\\
&-\frac{1}{2}\left(\left|\Lambda(k)\right|+\gamma\left|\Lambda(k)\right|^{\alpha}p\frac{(p-1)\left(c_{k}^{n+1}\right)^{p-2}\left(1+e^{-\beta c_{k}^{n+1}}\right)^{2}+\beta\left(c_{k}^{n+1}\right)^{p-1}e^{-\beta c_{k}^{n+1}}\left(p+2pe^{-\beta c_{k}^{n+1}}+1\right)+\beta^{2}\left(c_{k}^{n+1}\right)^{p}e^{-\beta c_{k}^{n+1}}\left(pe^{-\beta c_{k}^{n+1}}-1\right)}{\left(1+e^{-\beta c_{k}^{n+1}}\right)^{p+2}}\right)\left(c_{k}^{n+1}-c_{k}^{n}\right)^{2}.
\end{aligned} \tag{S13}$$

Hence, if $\tau_1 \geq \tau_2$, the following inequality evidently holds:

$$\varepsilon\left(s_{ir}^{n+1}, d_{jr}^{n+1}, t_{kr}^{n+1}, a_i^{n+1}, b_j^{n+1}, c_k^{n+1}\right) - \varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_i^{n}, b_j^{n}, c_k^{n}\right) \leq 0. \tag{S14}$$

Thus, the objective function (6) is non-increasing.

Moreover, after the n-th iteration, (6) is formulated as:

$$\begin{aligned}
&\varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_i^{n}, b_j^{n}, c_k^{n}\right) \\
&= \frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(y_{ijk} - \left(\sum_{r=1}^{R} s_{ir}^{n} d_{jr}^{n} t_{kr}^{n} + a_i^n + b_j^n + c_k^n\right)\right)^2 \\
&+\gamma\sum_{i=1}^{|I|}\left|\Lambda(i)\right|^{\alpha}\left(\sum_{r=1}^{R}\left(\frac{s_{ir}^n}{1+e^{-\beta s_{ir}}}\right)^p + \left(\frac{a_i^n}{1+e^{-\beta a_i}}\right)^p\right) \\
&+\gamma\sum_{j=1}^{|J|}\left|\Lambda(j)\right|^{\alpha}\left(\sum_{r=1}^{R}\left(\frac{d_{jr}^n}{1+e^{-\beta d_{jr}}}\right)^p + \left(\frac{b_j^n}{1+e^{-\beta b_j}}\right)^p\right) \\
&+\gamma\sum_{k=1}^{|K|}\left|\Lambda(k)\right|^{\alpha}\left(\sum_{r=1}^{R}\left(\frac{t_{kr}^n}{1+e^{-\beta t_{kr}}}\right)^p + \left(\frac{c_k^n}{1+e^{-\beta c_k}}\right)^p\right).
\end{aligned} \tag{S15}$$

From (S15), we see that if $\gamma \geq 0$ is satisfied, the following inequality must be true:

$$\varepsilon\left(s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_i^{n}, b_j^{n}, c_k^{n}\right) \geq 0. \tag{S16}$$

Hence, the objective function (6) is low-bounded. According to the above inferences, *Lemma* 1 holds. Then to implement Step 2, we present the following *Theorem* 1.

***Theorem* 1**. Sequences $\left\{s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}, a_i^{n}, b_j^{n}, c_k^{n}\right\}$ learnt from update rules in (17) converge to a stationary point $\left\{s_{ir}^{*}, d_{jr}^{*}, t_{kr}^{*}, a_i^{*}, b_j^{*}, c_k^{*}\right\}$ of $\varepsilon\left(s_{ir}, d_{jr}, t_{kr}, a_i, b_j, c_k\right)$ in (6).

Note that the proof process of $\left\{s_{ir}^{n}, d_{jr}^{n}, t_{kr}^{n}\right\}$ is similar and $\left\{a_i^{n}, b_j^{n}, c_k^{n}\right\}$ is also similar, hence, for conciseness, we only present the proof with $\left\{s_{ir}^{n}\right\}$ and $\left\{a_i^{n}\right\}$.

Proof of Theorm 1. Firstly, based on (S4) and (S5), $\forall i\in I$, $j\in J$, $k\in K$, we have the following references [58]:

$$\begin{aligned}
&\lim_{n\to+\infty}\left(s_{ir}^{n+1} - s_{ir}^{n}\right) = 0, \lim_{n\to+\infty}\left(a_i^{n+1} - a_i^{n}\right) = 0, \\
&\lim_{n\to+\infty}\left(d_{jr}^{n+1} - d_{jr}^{n}\right) = 0, \lim_{n\to+\infty}\left(b_j^{n+1} - b_j^{n}\right) = 0, \\
&\lim_{n\to+\infty}\left(t_{kr}^{n+1} - t_{kr}^{n}\right) = 0, \lim_{n\to+\infty}\left(c_k^{n+1} - c_k^{n}\right) = 0.
\end{aligned} \tag{S17}$$

From (S14) we see that a sequence $\left\{s_{ir}^{n}\right\}$ converges with the update rule (17). Let $\left\{s_{ir}^{*}\right\}$ denotes the converging state of $\left\{s_{ir}^{n}\right\}$, i.e., $0 \leq s_{ir}^{*} = \lim_{n\to+\infty} s_{ir}^{n} < +\infty$. Then for the objective (6), the following KKT conditions related to $\left\{s_{ir}^{n}\right\}$ should be fulfilled if $\left\{s_{ir}^{*}\right\}$ is one of its stationary point.

$$\left.\frac{\partial L}{\partial s_{ir}}\right|_{s_{ir}=s_{ir}^{*}} = \sum_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk} - \hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right) + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(\tilde{s}_{ir}^{*}\right)^{p} e^{-\beta\tilde{s}_{ir}^{*}} + \left(\tilde{s}_{ir}^{*}\right)^{p-1} + \left(\tilde{s}_{ir}^{*}\right)^{p-1} e^{-\beta\tilde{s}_{ir}^{*}}\right)\Big/\left(1+e^{-\beta\tilde{s}_{ir}^{*}}\right)^{p+1} - \tilde{s}_{ir}^{*} = 0, \tag{S18a}$$

$$\tilde{s}_{ir}^{*}\cdot s_{ir}^{*} = 0, \tag{S18b}$$

$$s_{ir}^{*} \geq 0, \tag{S18c}$$

$$\tilde{s}_{ir}^{*} \geq 0. \tag{S18d}$$

Note that following (S1)-(S3), condition (S18a) is evidently fulfilled with parameter update rule (17), making the following equation holds:

$$\tilde{s}_{ir}^{*} = \sum_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk} - \hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right) + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(\tilde{s}_{ir}^{*}\right)^{p} e^{-\beta\tilde{s}_{ir}^{*}} + \left(\tilde{s}_{ir}^{*}\right)^{p-1} + \left(\tilde{s}_{ir}^{*}\right)^{p-1} e^{-\beta\tilde{s}_{ir}^{*}}\right)\Big/\left(1+e^{-\beta\tilde{s}_{ir}^{*}}\right)^{p+1}, \tag{S19}$$

Hence, we focus on analyzing condition (S18c) and (S18d). We first construct ξ_{ir}^{n} as:

$$\xi_{ir}^{n}=\frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}d_{jr}t_{kr}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta s_{ir}^{p}e^{-\beta s_{ir}}+s_{ir}^{p-1}+s_{ir}^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}}\right)^{p+1}}. \tag{S20}$$

Naturally, (S20) is bounded by non-negative s_{ir}:

$$0\le\xi_{ir}^{*}=\lim_{n\to+\infty}\xi_{ir}^{n}=\frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}d_{jr}t_{kr}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta\left(s_{ir}^{*}\right)^{p}e^{-\beta s_{ir}}+\left(s_{ir}^{*}\right)^{p-1}+\left(s_{ir}^{*}\right)^{p-1}e^{-\beta s_{ir}^{*}}\right)\Big/\left(1+e^{-\beta s_{ir}^{*}}\right)^{p+1}}. \tag{S21}$$

Thus, we write the update rule of s_{ir} with SLF-NMU as:

$$s_{ir}^{n+1}=s_{ir}^{n}\xi_{ir}^{n}. \tag{S22}$$

By combining (S17) and (S22), we have:

$$\lim_{n\to+\infty}\left(s_{ir}^{n+1}-s_{ir}^{n}\right)=0\Rightarrow s_{ir}^{*}\xi_{ir}^{*}-s_{ir}^{*}=0. \tag{S23}$$

Note that following the update rule (17), $s_{ir}^{*}\ge 0$ with a non-negatively initial hypothesis. Hence, we have the following inferences.

a) **When** s_{ir}^{*} **>0.** Based on (S20) and (S23), we have:

$$\lim_{n\to+\infty}s_{ir}^{*}\xi_{ir}^{*}-s_{ir}^{*}=0,s_{ir}^{*}>0\Rightarrow\xi_{ir}^{*}=1\Rightarrow\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta\left(s_{ir}^{*}\right)^{p}e^{-\beta s_{ir}}+\left(s_{ir}^{*}\right)^{p-1}+\left(s_{ir}^{*}\right)^{p-1}e^{-\beta s_{ir}^{*}}\right)\Big/\left(1+e^{-\beta s_{ir}^{*}}\right)^{p+1}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}=0. \tag{S24}$$

By combing (S19) and (S24), we achieve condition (S18b):

$$\tilde{s}_{ir}^{*}=\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta\left(s_{ir}^{*}\right)^{p}e^{-\beta s_{ir}}+\left(s_{ir}^{*}\right)^{p-1}+\left(s_{ir}^{*}\right)^{p-1}e^{-\beta s_{ir}^{*}}\right)\Big/\left(1+e^{-\beta s_{ir}^{*}}\right)^{p+1}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}=0\Rightarrow\tilde{s}_{ir}^{*}\cdot s_{ir}^{*}=0. \tag{S25}$$

Meanwhile, when $\tilde{s}_{ir}^{*}$=0 and s_{ir}^{*}>0, condition (S18c) and (S18d) are naturally fulfilled. Hence, when s_{ir}^{*}>0, KKT conditions in (S18) are all satisfied.

b) **When** s_{ir}^{*} **=0.** The conditions (S18b) and (S18c) naturally holds. Hence, we only need to justify that whether condition (S18d) is fulfilled or not. To do so, we reformulate s_{ir}^{*} as follows:

$$s_{ir}^{*}=s_{ir}^{0}\lim_{n\to+\infty}\prod_{h=1}^{n}\xi_{ir}^{h}. \tag{S26}$$

Based on (S26) we further have the following deduction:

$$\begin{aligned}&s_{ir}^{0}>0,s_{ir}^{0}\lim_{n\to+\infty}\prod_{h=1}^{n}\xi_{ir}^{h}=s_{ir}^{*}=0\Rightarrow\lim_{n\to+\infty}\prod_{h=1}^{n}\xi_{ir}^{h}=0\\&\Rightarrow\lim_{n\to+\infty}\xi_{ir}^{n}=\xi_{ir}^{*}=\frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}d_{jr}t_{kr}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta\left(s_{ir}^{*}\right)^{p}e^{-\beta s_{ir}}+\left(s_{ir}^{*}\right)^{p-1}+\left(s_{ir}^{*}\right)^{p-1}e^{-\beta s_{ir}^{*}}\right)\Big/\left(1+e^{-\beta s_{ir}^{*}}\right)^{p+1}}\le 1\\&\Rightarrow\tilde{s}_{ir}^{*}=\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta\left(s_{ir}^{*}\right)^{p}e^{-\beta s_{ir}}+\left(s_{ir}^{*}\right)^{p-1}+\left(s_{ir}^{*}\right)^{p-1}e^{-\beta s_{ir}}\right)\Big/\left(1+e^{-\beta s_{ir}^{*}}\right)^{p+1}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}\ge 0.\end{aligned} \tag{S27}$$

Hence, the condition (S18d) holds. Therefore, when s_{ir}^{*}=0, KKT conditions in (S18) are all satisfied.

By analogy, we can prove that sequences $\left\{d_{jr}^{n}\right\}$ and $\left\{t_{kr}^{n}\right\}$ converge to a stationary point of (6), too. Next, we prove the convergence of sequence $\left\{a_{i}^{n}\right\}$.

Let a_{i}^{*} denotes the converging state of sequence $\left\{a_{i}^{n}\right\}$, i.e., $\forall i\in I:0\le a_{i}^{*}=\lim_{n\to+\infty}a_{i}^{n}\le+\infty$. If a_{i}^{*} is one of a_{i}^{n}'s stationary point, the following KKT conditions should be fulfilled:

$$\left.\frac{\partial L}{\partial a_{i}}\right|_{a_{i}=a_{i}^{*}}=\sum_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\left|\Lambda(i)\right|^{\alpha}p\left(\beta\left(a_{i}^{*}\right)^{p}e^{-\beta a_{i}^{*}}+\left(a_{i}^{*}\right)^{p-1}+\left(a_{i}^{*}\right)^{p-1}e^{-\beta a_{i}^{*}}\right)\Big/\left(1+e^{-\beta a_{i}^{*}}\right)^{p+1}-\tilde{a}_{i}^{*}=0, \tag{S28a}$$

$$\tilde{a}_{i}^{*}\cdot a_{i}^{*}=0. \tag{S28b}$$

$$a_{i}^{*}\ge 0, \tag{S28c}$$

$$\tilde{a}_i^* \geq 0. \tag{S28d}$$

Following (S1)-(S3), we see that condition (S28a) naturally holds. Hence, we have:

$$\tilde{a}_i^* = \sum_{y_{ijk}\in\Lambda(i)} \left(\left(y_{ijk} - \hat{y}_{ijk}\right)(-1)\right) + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(a_i^*\right)^p e^{-\beta a_i^*} + \left(a_i^*\right)^{p-1} + \left(a_i^*\right)^{p-1} e^{-\beta a_i^*}\right) \Big/ \left(1 + e^{-\beta a_i^*}\right)^{p+1}. \tag{S29}$$

Thus, we focus on condition (S28c) and (S28d), we first construct ς_i^n as follows:

$$\varsigma_i^n = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta a_i^p e^{-\beta a_i} + a_i^{p-1} + a_i^{p-1} e^{-\beta a_i}\right) \Big/ \left(1 + e^{-\beta a_i}\right)^{p+1}}. \tag{S30}$$

Obviously, (S30) is bounded by non-negative a_i^n, hence, we have:

$$0 \leq \varsigma_i^* = \lim_{n\to+\infty} \varsigma_i^n = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(a_i^n\right)^p e^{-\beta a_i^n} + \left(a_i^n\right)^{p-1} + \left(a_i^n\right)^{p-1} e^{-\beta a_i^n}\right) \Big/ \left(1 + e^{-\beta a_i^n}\right)^{p+1}}. \tag{S31}$$

Accordingly, the update rule of a_i^n can be rewrite as:

$$a_i^{n+1} = a_i^n \varsigma_i^n \tag{S32}$$

By combining (S17) and (S32), we have:

$$\lim_{n\to+\infty}\left(a_i^{n+1} - a_i^n\right) = 0 \Rightarrow a_i^* \varsigma_i^* - a_i^* = 0. \tag{S33}$$

Note that following the update rule (17), $a_i^* \geq 0$ with a non-negatively initial hypothesis. Hence, we have the following inferences.

a) **When** $a_i^* > 0$. Based on (S30) and (S33), we have:

$$a_i^* \varsigma_i^* - a_i^* = 0, a_i^* \geq 0 \Rightarrow \varsigma_i^* = 1 \Rightarrow \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(a_i^*\right)^p e^{-\beta a_i^*} + \left(a_i^*\right)^{p-1} + \left(a_i^*\right)^{p-1} e^{-\beta a_i^*}\right) \Big/ \left(1 + e^{-\beta a_i^*}\right)^{p+1} - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} = 0. \tag{S34}$$

By combing (S29) and (S34), we achieve condition (S28b):

$$\tilde{a}_i^* = \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(a_i^*\right)^p e^{-\beta a_i^*} + \left(a_i^*\right)^{p-1} + \left(a_i^*\right)^{p-1} e^{-\beta a_i^*}\right) \Big/ \left(1 + e^{-\beta a_i^*}\right)^{p+1} - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} = 0 \Rightarrow \tilde{a}_i^* \cdot a_i^* = 0. \tag{S35}$$

Meanwhile, when $a_i^* > 0$ and $\tilde{a}_i^* = 0$, conditions (S28c) and (S28d) naturally hold. Therefore, when $a_i^* > 0$, KKT conditions in (S28) are all satisfied.

b) **When** $a_i^* = 0$. Under such circumstance, conditions (S28b) and (S28c) are naturally fulfilled. Thus, we need to justify that whether condition (S28d) is fulfilled or not. To this end, we formulated $\tilde{a}_i^*$ as follows:

$$a_i^* = a_i^0 \lim_{n\to+\infty} \prod_{h=1}^{n} \varsigma_i^h. \tag{S36}$$

Following (S36), we have:

$$\begin{aligned}
& a_i^0 > 0, a_i^0 \lim_{n\to+\infty} \prod_{h=1}^{n} \varsigma_i^h = a_i^* = 0 \Rightarrow \lim_{n\to+\infty} \prod_{h=1}^{n} \varsigma_i^h = 0 \\
& \Rightarrow \lim_{n\to+\infty} \varsigma_i^n = \varsigma_i^* = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(a_i^*\right)^p e^{-\beta a_i^*} + \left(a_i^*\right)^{p-1} + \left(a_i^*\right)^{p-1} e^{-\beta a_i^*}\right) \Big/ \left(1 + e^{-\beta a_i^*}\right)^{p+1}} \leq 1 \\
& \Rightarrow \tilde{a}_i^* = \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\left|\Lambda(i)\right|^{\alpha} p\left(\beta\left(a_i^*\right)^p e^{-\beta a_i^*} + \left(a_i^*\right)^{p-1} + \left(a_i^*\right)^{p-1} e^{-\beta a_i^*}\right) \Big/ \left(1 + e^{-\beta a_i^*}\right)^{p+1} - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} \geq 0.
\end{aligned} \tag{S37}$$

Hence, when $a_i^* = 0$, KKT conditions in (S28) are also satisfied. Therefore, based on the above inference, *Theorem* 1 stands. According to Theorem 1, Step 2 is implemented. By combining steps 1-2, we conclude that with the SLF-NMU-based learning scheme (17), an FRNL model's convergence on a nonnegative HDI tensor is guaranteed.

III. Supplementary Table of Empirical Studies

TABLE S.1 Hyper-parameter settings of M1-M7 on D1-8.

Dataset	**Hyper-parameter Setting**				
D1	M1: Self-adaptation	M2: λ=0.001, η=0.001	M3: λ=0.01, η=0.0005	M4: λ_a=0.05, λ_b=0.05, λ_c=0.0001	M5: λ=0.05, η=0.01
D2	M1: Self-adaptation	M2: λ=0.001, η=0.001	M3: λ=0.01, η=0.0005	M4: λ_a=0.05, λ_b=0.05, λ_c=0.0001	M5: λ=0.03, η=0.01
D3	M1: Self-adaptation	M2: λ=0.01, η=0.004	M3:λ=0.001,η=0.0005	M4: λ_a=0.01, λ_b=0.01, λ_c=0.0005	M5: λ=0.01, η=0.02
D4	M1: Self-adaptation	M2: λ=0.01, η=0.005	M3:λ=0.001,η=0.0005	M4: λ_a=0.005, λ_b=0.01,λ_c=0.0005	M5: λ=0.01, η=0.02
D5	M1: Self-adaptation	M2: λ=0.0001, η=0.005	M3: λ=0.01, η=0.0001	M4: λ_a=0.01, λ_b=0.001, λ_c=0.05	M5: λ=0.001, η=0.04
	M6: $\rho_{A\bar{n}}$=1, $\rho_{\tilde{A}n}$=1, $\rho_{d\tilde{n}}$=1,μ=0.0001		M7: α=0.0001, β=0.1		
D6	M1: Self-adaptation	M2: λ=0.05, η=0.0005	M3: λ=0.01, η=0.0002	M4: λ_a=0.01, λ_b=0.001, λ_c=0.001	M5: λ=0.05, η=0.02
	M6: $\rho_{A\bar{n}}$=0.1, $\rho_{\tilde{A}n}$=0.1, $\rho_{d\tilde{n}}$=0.1,μ=0.0001		M7: α=0.001, β=0.1		
D7	M1: Self-adaptation	M2: λ=0.00001, η=0.005	M3: λ=0.1, η=0.0005	M4: λ_a=0.01, λ_b=0.001, λ_c=0.01	M5: λ=0.1, η=0.01
	M6: $\rho_{A\bar{n}}$=1, $\rho_{\tilde{A}n}$=1, $\rho_{d\tilde{n}}$=1,μ=0.0001		M7: α=0.0001, β=0.1		
D8	M1: Self-adaptation	M2: λ=0.05, η=0.0001	M3: λ=0.1, η=0.0003	M3: λ_a=0.01, λ_b=0.001, λ_c=0.001	M4: λ=0.1, η=0.02
	M6: $\rho_{A\bar{n}}$=0.1, $\rho_{\tilde{A}n}$=0.1, $\rho_{d\tilde{n}}$=0.1,μ=0.00001		M7: α=0.001, β=0.1		