Temporal Relations-Aware Nonnegative Latent Factorization of Tensors for Dynamic Directed Graph Representation Supplementary File

Hao Wu, *Member, IEEE*, Xin Luo, *Senior Member, IEEE*, and Xinbo Gao, *Fellow, IEEE*

I. Introduction

This is the supplementary file for paper entitled "*Temporal Relations-Aware Nonnegative Latent Factorization of Tensors for Dynamic Directed Graph Representation*", which presents the convergence proof of the proposed TRNL model.

II. Convergence Proof of TRNL

In order to analysis the convergence of the proposed TRNL model, it is first essential to discover the relations between the KKT condition of the learning objective (7) and the SLF-NMUT algorithm [49, 50]. Hence, considering the nonnegative constraints for the LF tensor **X**, LF matrix A and linear biases vectors ***s***, ***d***, and ***t*** we have the Lagrangian function L for (7) as:

$$L=\varepsilon\left(\mathbf{X},\mathrm{A},\boldsymbol{s},\boldsymbol{d},\boldsymbol{t}\right)-\sum_{p=1}^{R}\sum_{q=1}^{R}x_{pqk}\tilde{x}_{pqk}-\sum_{i=1}^{|V|}\sum_{p=1}^{R}a_{ip}\tilde{a}_{ip}-\sum_{j=1}^{|V|}\sum_{q=1}^{R}a_{jq}\tilde{a}_{jq}-\sum_{i=1}^{|V|}s_i\tilde{s}_i-\sum_{j=1}^{|V|}d_j\tilde{d}_j-\sum_{k=1}^{|K|}t_k\tilde{t}_k. \tag{S1}$$

where $\tilde{x}_{pqk},\tilde{a}_{ip},\tilde{a}_{jq},\tilde{s}_i,\tilde{d}_j,\tilde{t}_k$ denote single element in $\tilde{\mathbf{X}}$, $\tilde{\mathrm{A}}$, $\tilde{\boldsymbol{s}}$, $\tilde{\boldsymbol{d}}$, and $\tilde{\boldsymbol{t}}$ respectively, which denote Lagrangian multipliers for **X**, A, ***s***, ***d***, and ***t***.

Considering the partial derivatives of L with **X**, A, ***s***, ***d***, and ***t***, they are highly similar for a_{ip}, a_{jp} *and* s_i, d_j, and t_k. Hence, we consider the case of x_{pqk}, a_{ip} and s_i as follows:

$$\begin{cases}\dfrac{\partial L}{\partial x_{pqk}}=\sum\limits_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right)+\lambda\left|\Lambda(k)\right|x_{pqk}-\tilde{x}_{pqk}=0,\\ \dfrac{\partial L}{\partial a_{ip}}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-\sum\limits_{q=1}^{R}x_{pqk}a_{jq}\right)\right)+\lambda\left|\Lambda(i)\right|a_{ip}-\tilde{a}_{ip}=0,\\ \dfrac{\partial L}{\partial s_i}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\lambda_b\left|\Lambda(i)\right|s_i-\tilde{s}_i=0.\end{cases}$$
$$\Rightarrow\begin{cases}\tilde{x}_{pqk}=\sum\limits_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right)+\lambda\left|\Lambda(k)\right|x_{pqk},\\ \tilde{a}_{ip}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-\sum\limits_{q=1}^{R}x_{pqk}a_{jq}\right)\right)+\lambda\left|\Lambda(i)\right|a_{ip},\\ \tilde{s}_i=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\lambda_b\left|\Lambda(i)\right|s_i.\end{cases} \tag{S2}$$

Then, considering the KKT conditions of (S1), i.e., $\forall x_{pqk}$, $\tilde{x}_{pqk}$: $x_{pqk}\tilde{x}_{pqk}=0$, $\forall a_{ir}$, $\tilde{a}_{ip}$: $a_{ip}\tilde{a}_{ip}=0$, and $\forall s_i$, $\tilde{s}_i$: $s_i\tilde{s}_i=0$, we have:

$$\begin{cases}x_{pqk}\left(\sum\limits_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right)+\lambda\left|\Lambda(k)\right|x_{pqk}\right)=0,\\ a_{ip}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-\sum\limits_{q=1}^{R}x_{pqk}a_{jq}\right)\right)+\lambda\left|\Lambda(i)\right|a_{ip}\right)=0,\\ s_i\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\lambda_b\left|\Lambda(i)\right|s_i\right)=0.\end{cases} \tag{S3}$$

With (S3), we can achieve the following parameters equations:

$$\begin{cases} x_{pqk}\sum_{y_{ijk}\in\Lambda(k)} y_{ijk}a_{ip}a_{jq} = x_{pqk}\left(\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}\right), \\ a_{ip}\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}\sum_{q=1}^{R}x_{pqk}a_{jq} = a_{ip}\left(\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}\sum_{q=1}^{R}x_{pqk}a_{jq}+\lambda\left|\Lambda(i)\right|a_{ip}\right), \\ s_i\sum_{y_{ijk}\in\Lambda(i)} y_{ijk} = s_i\left(\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}+\lambda_b\left|\Lambda(i)\right|s_i\right). \end{cases}$$

$$\Rightarrow\begin{cases} x_{pqk} = x_{pqk}\dfrac{\sum_{y_{ijk}\in\Lambda(k)} y_{ijk}a_{ip}a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}}, \\ a_{ip} = a_{ip}\dfrac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}\sum_{q=1}^{R}x_{pqk}a_{jq}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}\sum_{q=1}^{R}x_{pqk}a_{jq}+\lambda\left|\Lambda(i)\right|a_{ip}}, \\ s_i = s_i\dfrac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}+\lambda_b\left|\Lambda(i)\right|s_i}. \end{cases} \tag{S4}$$

In particular, with (S4), we can conveniently achieve the parameters update rule given in (8). Therefore, the SLF-NMUT-based learning scheme in the TRNL 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 (7) is non-increasing.

Step 2: Sequences { x_{pqk}^n , a_{ip}^n , a_{jq}^n , s_i^n , d_j^n , t_k^n } converge to an equilibrium point (x_{pqk}^* , a_{ip}^* , a_{jq}^* , s_i^* , d_j^* , t_k^*).

In the Step 1, we aim to prove that objective function (7) is nonincreasing with the SLF-NMUT-based learning scheme (15). To do so, we have:

Theorem 1: (7) is nonincreasing with (15).

In particular, an auxiliary function is essential and vital to prove Theorem 1 [50]. Hence, the following function is defined:

Definition 2: $H(x, x')$ is an auxiliary function of $F(x)$ if

$$H(x, x') \geq F(x),\ H(x, x')=F(x). \tag{S5}$$

Accordingly, we further recall the following lemma [50, 51],

Lemma 1: $F(x)$ keeps nonincreasing with the following rule:

$$x^{n+1}=\mathrm{argmin}_x H(x, x'). \tag{S6}$$

Proof of Lemma 1: With Definition 2, we deduce that

$$F(x^n) = H(x^n, x^n) \geq H(x^{n+1}, x^n) \geq F(x^{n+1}). \tag{S7}$$

Note that we have $F(x^{n+1}) = F(x^n)$ when x^n guarantees a local minimum of $H(x, x^n)$. Hence, $\nabla F(x^n)=0$ holds if $F(x^n)$ is differentiable around x^n. Thus, (S7) can be extended into the following converging sequence to $x_{\min}= \mathrm{argmin}_x F(x)$:

$$F(x_{\min}) \leq \cdots \leq F(x^{n+1}) \leq F(x^n) \leq \cdots \leq F(x_1) \leq F(x_0). \tag{S8}$$

Next, we aim to achieve that (7) for is exactly consistent with that in (S6) with a specifically designed H. Considering $x_{pqk}\in\mathbf{X}$, let Fx_{pqk} be the partial loss from (7) $\varepsilon(\mathbf{X}, \mathrm{A}, \boldsymbol{s}, \boldsymbol{d}, \boldsymbol{t})$ related to x_{pqk} only,

$$F_{x_{pqk}} = \frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)^2+\lambda\left(\sum_{p=1}^{R}\sum_{q=1}^{R}x_{pqk}^2+\sum_{p=1}^{R}a_{ip}^2+\sum_{q=1}^{R}a_{jp}^2\right)+\lambda_b\left(s_i^2+d_j^2+t_k^2\right)\right). \tag{S9}$$

As a result, the first-order and second-order derivatives of Fx_{pqk} with respect to x_{pqk} can be obtained as:

$$\begin{aligned} F'_{x_{pqk}} &= \frac{\partial\varepsilon}{\partial x_{pqk}} = \sum_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right)+\lambda\left|\Lambda(k)\right|x_{pqk}, \\ F''_{x_{pqk}} &= \frac{\partial^2\varepsilon}{\partial\left(x_{pqk}\right)^2} = \sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^2+\lambda\left|\Lambda(k)\right|. \end{aligned} \tag{S10}$$

According to (S8)-(S10), we obtain the following proposition:

Proposition 1: The auxiliary function of Fx_{pqk} is given as:

$$H\left(x, x_{pqk}^n\right) = F_{x_{pqk}}\left(x_{pqk}^n\right)+F'_{x_{pqk}}\left(x_{pqk}^n\right)\left(x-x_{pqk}^n\right)+\frac{1}{2}\left(\left(\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^n\right)\Big/x_{pqk}^n\right)\left(x-x_{pqk}^n\right)^2. \tag{S11}$$

With (S11), $H(x, x)= Fx_{pqk}(x)$ holds.

Next, we prove $H\left(x,x_{pqk}^{n}\right)\ge F_{x_{pqk}}\left(x\right)$. To do so, the quadratic approximation to Fx_{pqk} at x_{pqk}^{n} needs to be first obtained as:

$$F_{x_{pqk}}\left(x\right)=F_{x_{pqk}}\left(x_{pqk}^{n}\right)+F'_{x_{pqk}}\left(x_{pqk}^{n}\right)\left(x-x_{pqk}^{n}\right)+\frac{1}{2}F''_{x_{pqk}}\left(x_{pqk}^{n}\right)\left(x-x_{pqk}^{n}\right)^{2}. \tag{S12}$$

By combine (S10)-(S12), we can see that $H\left(x,x_{pqk}^{n}\right)$ is an auxiliary function of Fx_{pqk} if the following inequality holds:

$$\left(\frac{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{n}}{x_{pqk}^{n}}\right)\ge\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^{2}+\lambda\left|\Lambda(k)\right|. \tag{S13}$$

Note that $\hat{y}_{ijk}$, a_{ip}, and a_{jq} are nonnegative, thus, (S13) is equal to

$$\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}\ge x_{pqk}^{n}\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^{2}. \tag{S14}$$

Then, we reformulate the left term of (S14) as follows:

$$\begin{aligned}\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}&=\sum_{y_{ijk}\in\Lambda(k)}a_{ip}a_{jq}\left(x_{pqk}^{n}a_{ip}a_{jq}+\sum_{u=1,u\ne p}^{R}\sum_{v=1,v\ne q}^{R}x_{uvk}a_{iu}a_{jv}+s_{i}+d_{j}+t_{k}\right)\\&=x_{pqk}^{n}\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^{2}+\sum_{y_{ijk}\in\Lambda(k)}a_{ip}a_{jq}\left(\sum_{u=1,u\ne p}^{R}\sum_{v=1,v\ne q}^{R}x_{uvk}a_{iu}a_{jv}+s_{i}+d_{j}+t_{k}\right)\\&\ge x_{pqk}^{n}\sum_{y_{ijk}\in\Lambda(k)}\left(a_{ip}a_{jq}\right)^{2}.\end{aligned} \tag{S15}$$

Note that (S13) holds with (S15), making $H\left(x,x_{pqk}^{n}\right)$ is an auxiliary function of Fx_{pqk}.

Based on Proposition 1, we achieve the following proof.
Proof of theorem 1: Based on (S6), (S10) and (S11), we have:

$$\begin{aligned}&x_{pqk}^{n+1}=\arg\min_{x}H\left(x,x_{pqk}^{n}\right)\\&\Rightarrow F'_{x_{pqk}}\left(x_{pqk}^{n}\right)+\frac{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{n}}{x_{pqk}^{n}}\left(x-x_{pqk}^{n}\right)^{2}=0\\&\Rightarrow x_{pqk}^{n+1}\leftarrow x_{pqk}^{n}\frac{\sum_{y_{ijk}\in\Lambda(k)}y_{ijk}a_{ip}a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{n}}.\end{aligned} \tag{S16}$$

Based on (S8), it is clearly shown that Fx_{pqk} is nonincreasing with (15). Naturally, $\forall i, j\in V$, $k\in K$, p, $q\in\{1, \ldots, R\}$, (S16) holds. Therefore, Theorem 1 holds.

Following Theorem 1, with a positive initialization, i.e., $x_{pqk}^{0}\ge 0,a_{ip}^{0}\ge 0,a_{iq}^{0}\ge 0,s_{i}^{0}\ge 0,s_{j}^{0}\ge 0,t_{k}^{0}\ge 0$, we have the following recursion:

$$F\left(x_{pqk}^{0}\right)\ge F\left(x_{pqk}^{n}\right)\ge H\left(x_{pqk}^{n+1},x_{pqk}^{n}\right)\ge F\left(x_{pqk}^{n+1}\right)\ge 0, \tag{S17}$$

which demonstrates that a sequence $\left\{F\left(x_{pqk}^{n}\right)\right\}$ is monotonically nonincreasing and bounded. Therefore, we have:

$$\lim_{n\to+\infty}\left(F\left(x_{pqk}^{n+1}\right)-F\left(x_{pqk}^{n}\right)\right)=0. \tag{S18}$$

With (S18), we further have the following inference:

$$\lim_{n\to+\infty}\left|x_{pqk}^{n+1}-x_{pqk}^{n}\right|=0. \tag{S19}$$

Hence, the sequence $\left\{x_{pqk}^{n}\right\}$ is bounded and convergent. Similarly, the sequence $\left\{a_{ip}^{n}\right\}$, $\left\{a_{jq}^{n}\right\}$, $\left\{s_{i}^{n}\right\}$, $\left\{d_{j}^{n}\right\}$, $\left\{t_{k}^{n}\right\}$ are also bounded and convergent.

In the Step 2, we aim to prove that sequences { x_{pqk}^{n}, a_{ip}^{n}, a_{jq}^{n}, s_{i}^{n}, d_{j}^{n}, t_{k}^{n} } obtained by SLF-NMUT-based learning scheme converge to a KKT equilibrium point (x_{pqk}^{*}, a_{ip}^{*}, a_{jq}^{*}, s_{i}^{*}, d_{j}^{*}, t_{k}^{*}) of (7). To prove it, we have:

Theorem 2: Sequences { x_{pqk}^{n}, a_{ip}^{n}, a_{jq}^{n}, s_{i}^{n}, d_{j}^{n}, t_{k}^{n} } by (15) converge to an equilibrium point (x_{pqk}^{*}, a_{ip}^{*}, a_{jq}^{*}, s_{i}^{*}, d_{j}^{*}, t_{k}^{*}) of ε (**X**, A, ***s***, ***d***, ***t***) in (7).

Proof of Theorem 2: From (S19) we see that a sequence $\left\{x_{pqk}^{n}\right\}$ converges with the update rule (14). Let $\left\{x_{pqk}^{*}\right\}$ denotes the converging state of $\left\{x_{pqk}^{n}\right\}$, i.e., $0 \le x_{pqk}^{*} = \lim_{n\to+\infty} x_{pqk}^{n} < +\infty$. Then for the learning objective (7), the following KKT conditions related to $\left\{x_{pqk}^{n}\right\}$ should be fulfilled if $\left\{x_{pqk}^{*}\right\}$ is one of its stationary point.

$$\left.\frac{\partial L}{\partial x_{pqt}}\right|_{x_{pqt}=x_{pqt}^{*}} = \sum_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right)+\lambda\left|\Lambda(k)\right|x_{pqt}^{*}-\tilde{x}_{pqt}^{*}=0, \tag{S20a}$$

$$x_{pqt}^{*}\cdot\tilde{x}_{pqt}^{*}=0, \tag{S20b}$$

$$x_{pqt}^{*}\ge 0, \tag{S20c}$$

$$\tilde{x}_{pqt}^{*}\ge 0. \tag{S20d}$$

Note that following (S1)-(S3), condition (S20a) is evidently fulfilled with parameter update rule (15), thus, we have the following equation holds:

$$\tilde{x}_{pqt}^{*} = \sum_{y_{ijk}\in\Lambda(k)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-a_{ip}a_{jq}\right)\right)+\lambda\left|\Lambda(k)\right|x_{pqt}^{*}. \tag{S21}$$

Next, we mainly analyze condition (S20c) and (S20d). We first construct β_{pqk}^{n} as:

$$\beta_{pqk}^{n} = \frac{\sum_{y_{ijk}\in\Lambda(k)} y_{ijk}a_{ip}a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{n}}. \tag{S22}$$

Naturally, (S22) is bounded by non-negative y_{ijk}, a_{ip} and a_{iq}:

$$0\le\beta_{pqk}^{*} = \lim_{n\to+\infty}\beta_{pqk}^{n} = \frac{\sum_{y_{ijk}\in\Lambda(k)} y_{ijk}a_{ip}a_{jq}}{\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{*}}. \tag{S23}$$

Hence, the update rule of x_{pqk} with SLF-NMU can be written as:

$$x_{pqk}^{n+1} = x_{pqk}^{n}\beta_{pqk}^{n}. \tag{S23}$$

By combining (S20) and (S23), we have:

$$\lim_{n\to+\infty}\left(x_{pqk}^{n+1}-x_{pqk}^{n}\right)=0 \Rightarrow x_{pqk}^{*}\beta_{pqk}^{*}-x_{pqk}^{*}=0. \tag{S24}$$

According to the update rule (15), $x_{pqk}^{*}\ge 0$ with a non-negatively initial hypothesis. Hence, we have the following inferences.

a) **When x_{pqk}^{*} >0.** Based on (S21) and (S24), we have:

$$\lim_{n\to+\infty} x_{pqk}^{*}\beta_{pqk}^{*}-x_{pqk}^{*}=0, x_{pqk}^{*}>0 \Rightarrow \beta_{pqk}^{*}=1 \Rightarrow \sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqt}^{*}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}a_{ip}a_{jq}=0. \tag{S25}$$

By combing (S20) and (S25), we achieve condition (S20b):

$$\tilde{x}_{pqt}^{*} = \sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqt}^{*}-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}a_{ip}a_{jq} \Rightarrow \tilde{x}_{pqt}^{*}\cdot x_{pqt}^{*}\text{=0}. \tag{S26}$$

Meanwhile, when $\tilde{x}_{pqk}^{*}$ =0 and x_{pqk}^{*} >0, condition (S20c) and (S20d) are naturally fulfilled. Hence, when x_{pqk}^{*} >0, KKT conditions in (S20) are all satisfied.

b) **When x_{pqk}^{*} =0.** The conditions (S20b) and (S20c) naturally holds. Hence, we only need to justify that whether condition (S20d) is fulfilled or not. To do so, we reformulate x_{pqk}^{*} as follows:

$$x_{pqk}^{*} = x_{pqk}^{0}\lim_{n\to+\infty}\prod_{h=1}^{n}\beta_{pqk}^{h}. \tag{S27}$$

Based on (S27) we further have the following deduction:

$$\begin{aligned}
& x_{pqk}^{0}>0, x_{pqk}^{0}\lim_{n\to+\infty}\prod_{h=1}^{n}\beta_{pqk}^{h}=x_{pqk}^{*}=0 \Rightarrow \lim_{n\to+\infty}\prod_{h=1}^{n}\beta_{pqk}^{h}=0 \\
& \Rightarrow \lim_{n\to+\infty}\beta_{pqk}^{n}=\beta_{pqk}^{*}=\frac{\sum\limits_{y_{ijk}\in\Lambda(k)}y_{ijk}a_{ip}a_{jq}}{\sum\limits_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{*}}\le 1 \\
& \Rightarrow \tilde{x}_{pqk}^{*}=\sum_{y_{ijk}\in\Lambda(k)}\hat{y}_{ijk}a_{ip}a_{jq}+\lambda\left|\Lambda(k)\right|x_{pqk}^{*}-\sum_{y_{ijk}\in\Lambda(k)}y_{ijk}a_{ip}a_{jq}\ge 0.
\end{aligned} \tag{S28}$$

There, the condition (S20d) holds. Hence, when x_{pqk}^{*} =0, KKT conditions in (S20) are all satisfied. By analogy, the sequences $\left\{a_{ip}^{n}\right\}$, $\left\{a_{jq}^{n}\right\}$, $\left\{s_{i}^{n}\right\}$, $\left\{d_{j}^{n}\right\}$ and $\left\{t_{k}^{n}\right\}$ can also be proven to converge to a stationary point of (7).

As a result, according to Theorems 1-2, it can be proven that TRNL is guaranteed to converge at a KKT stationary point of its learning objective. Therefore, the convergence of TRNL model on a nonnegative HDI tensor is guaranteed.