An Elastic-Net regularized Approach to Nonnegative Latent Factorization of Tensors Supplementary File

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

I. Introduction

This is the supplementary file for paper entitled "*An Elastic-Net Regularized Approach to Nonnegative Latent Factorization of Tensors".* We have put the convergence proof of ER-NLFT model in Section II, and the supplementary tables and figures of empirical studies in Section III.

II. Convergence Proof of ER-NLFT

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

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

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

$$\begin{aligned}
&\begin{cases}
\dfrac{\partial L}{\partial s_{ir}}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}+s_{ir}\big/\sqrt{s_{ir}^2+\mu}\right)-\tilde{s}_{ir}=0,\\
\dfrac{\partial L}{\partial a_i}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i+a_i\big/\sqrt{a_i^2+\mu}\right)-\tilde{a}_i=0.
\end{cases}\\
\Rightarrow&\begin{cases}
\tilde{s}_{ir}=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}+s_{ir}\big/\sqrt{s_{ir}^2+\mu}\right),\\
\tilde{a}_i=\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i+a_i\big/\sqrt{a_i^2+\mu}\right).
\end{cases}
\end{aligned}\tag{S2}$$

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:

$$\begin{aligned}
&\begin{cases}
s_{ir}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}+s_{ir}\big/\sqrt{s_{ir}^2+\mu}\right)\right)=0,\\
a_i\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i+a_i\big/\sqrt{a_i^2+\mu}\right)\right)=0;
\end{cases}\\
\Rightarrow&\begin{cases}
s_{ir}\sum\limits_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}=s_{ir}\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}+s_{ir}\big/\sqrt{s_{ir}^2+\mu}\right)\right),\\
a_i\sum\limits_{y_{ijk}\in\Lambda(i)}y_{ijk}=a_i\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i+a_i\big/\sqrt{a_i^2+\mu}\right)\right).
\end{cases}
\end{aligned}\tag{S3}$$

With (S3), we conveniently achieve the iterative learning rules given in (17). Hence, an SLF-NMU-based learning scheme in an ER-NLFT model is closely connected to the KKT conditions of its learning objective. From this point of view, we theoretically prove the convergence of ER-NLFT in the following two steps:

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

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

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

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

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

and

$$\tau_2=\begin{pmatrix}\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(\left(s_{ir}^{n+1}-s_{ir}^{n}\right)^2\left(s_{ir}^{n+1}\right)^2\left(\left(s_{ir}^{n+1}\right)^2+\mu\right)^{-\frac{3}{2}}+\left(a_{i}^{n+1}-a_{i}^{n}\right)^2\left(a_{i}^{n+1}\right)^2\left(\left(a_{i}^{n+1}\right)^2+\mu\right)^{-\frac{3}{2}}\right)\\+\gamma\phi\left(\left|\Lambda(j)\right|\right)\left(\left(d_{jr}^{n+1}-d_{jr}^{n}\right)^2\left(d_{jr}^{n+1}\right)^2\left(\left(d_{jr}^{n+1}\right)^2+\mu\right)^{-\frac{3}{2}}+\left(b_{j}^{n+1}-b_{j}^{n}\right)^2\left(b_{j}^{n+1}\right)^2\left(\left(b_{j}^{n+1}\right)^2+\mu\right)^{-\frac{3}{2}}\right)\\+\gamma\phi\left(\left|\Lambda(k)\right|\right)\left(\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^2\left(t_{kr}^{n+1}\right)^2\left(\left(t_{kr}^{n+1}\right)^2+\mu\right)^{-\frac{3}{2}}+\left(c_{k}^{n+1}-c_{k}^{n}\right)^2\left(c_{k}^{n+1}\right)^2\left(\left(c_{k}^{n+1}\right)^2+\mu\right)^{-\frac{3}{2}}\right)\end{pmatrix},$$

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

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

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

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

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:

$$\begin{aligned}&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n},d_{jr}^{n},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)\\&\triangleq\left(\sum_{y_{ijk}\in\Lambda(i)}\left(y_{ijk}-\left(\sum_{r=1}^{R}s_{ir}^{n+1}d_{jr}^{n}t_{kr}^{n}+a_{i}^{n}+b_{j}^{n}+c_{k}^{n}\right)\right)\left(-d_{jr}^{n}t_{kr}^{n}\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^{n+1}+s_{ir}^{n+1}\Big/\sqrt{\left(s_{ir}^{n+1}\right)^2+\mu}\right)\right)\left(s_{ir}^{n+1}-s_{ir}^{n}\right)\\&-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(i)}\left(d_{jr}^{n}t_{kr}^{n}\right)^2+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(\sqrt{\left(s_{ir}^{n+1}\right)^2+\mu}-\left(s_{ir}^{n+1}\right)^2\left(\left(s_{ir}^{n+1}\right)^2+\mu\right)^{-\frac{1}{2}}\right)\Big/\left(\left(s_{ir}^{n+1}\right)^2+\mu\right)\right)\left(s_{ir}^{n+1}-s_{ir}^{n}\right)^2.\end{aligned} \tag{S6}$$

Based on SLF-NMU, considering s_{ir}'s optimal condition, (S6) is reformulated as:

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

Similarly, we have:

$$\begin{aligned}&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)\\&=-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(j)}\left(s_{ir}^{n+1}t_{kr}^{n}\right)^2+\gamma\phi\left(\left|\Lambda(j)\right|\right)\left(\sqrt{\left(d_{jr}^{n+1}\right)^2+\mu}-\left(d_{jr}^{n+1}\right)^2\Big/\sqrt{\left(d_{jr}^{n+1}\right)^2+\mu}\right)\Big/\left(\left(d_{jr}^{n+1}\right)^2+\mu\right)\right)\left(d_{jr}^{n+1}-d_{jr}^{n}\right)^2.\end{aligned} \tag{S8}$$

$$\begin{aligned}&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)\\&=-\frac{1}{2}\left(\sum_{y_{ijk}\in\Lambda(k)}\left(s_{ir}^{n+1}d_{jr}^{n+1}\right)^2+\gamma\phi\left(\left|\Lambda(k)\right|\right)\left(\sqrt{\left(t_{kr}^{n+1}\right)^2+\mu}-\left(t_{kr}^{n+1}\right)^2\Big/\sqrt{\left(t_{kr}^{n+1}\right)^2+\mu}\right)\Big/\left(\left(t_{kr}^{n+1}\right)^2+\mu\right)\right)\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^2.\end{aligned} \tag{S9}$$

$$\begin{aligned}&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n+1},b_{j}^{n},c_{k}^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_{i}^{n},b_{j}^{n},c_{k}^{n}\right)\\&=-\frac{1}{2}\left(\left|\Lambda(i)\right|+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(\sqrt{\left(a_{i}^{n+1}\right)^2+\mu}-\left(a_{i}^{n+1}\right)^2\Big/\sqrt{\left(a_{i}^{n+1}\right)^2+\mu}\right)\Big/\left(\left(a_{i}^{n+1}\right)^2+\mu\right)\right)\left(a_{i}^{n+1}-a_{i}^{n}\right)^2.\end{aligned} \tag{S10}$$

$$\begin{aligned}&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_i^{n+1},b_j^{n+1},c_k^{n}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_i^{n+1},b_j^{n},c_k^{n}\right)\\&=-\frac{1}{2}\left(\left|\Lambda(j)\right|+\gamma\phi\left(\left|\Lambda(j)\right|\right)\left(\sqrt{\left(b_j^{n+1}\right)^2+\mu}-\left(b_j^{n+1}\right)^2\Big/\sqrt{\left(b_j^{n+1}\right)^2+\mu}\right)\Big/\left(\left(b_j^{n+1}\right)^2+\mu\right)\right)\left(b_j^{n+1}-b_j^{n}\right)^2.\end{aligned}\tag{S11}$$

$$\begin{aligned}&\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_i^{n+1},b_j^{n+1},c_k^{n+1}\right)-\varepsilon\left(s_{ir}^{n+1},d_{jr}^{n+1},t_{kr}^{n+1},a_i^{n+1},b_j^{n+1},c_k^{n}\right)\\&=-\frac{1}{2}\left(\left|\Lambda(k)\right|+\gamma\phi\left(\left|\Lambda(k)\right|\right)\left(\sqrt{\left(c_k^{n+1}\right)^2+\mu}-\left(c_k^{n+1}\right)^2\Big/\sqrt{\left(c_k^{n+1}\right)^2+\mu}\right)\Big/\left(\left(c_k^{n+1}\right)^2+\mu\right)\right)\left(c_k^{n+1}-c_k^{n}\right)^2.\end{aligned}\tag{S12}$$

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\phi\left(\left|\Lambda(i)\right|\right)\left(\sqrt{\left(s_{ir}^{n+1}\right)^2+\mu}-\left(s_{ir}^{n+1}\right)^2\Big/\sqrt{\left(s_{ir}^{n+1}\right)^2+\mu}\right)\Big/\left(\left(s_{ir}^{n+1}\right)^2+\mu\right)\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\phi\left(\left|\Lambda(j)\right|\right)\left(\sqrt{\left(d_{jr}^{n+1}\right)^2+\mu}-\left(d_{jr}^{n+1}\right)^2\Big/\sqrt{\left(d_{jr}^{n+1}\right)^2+\mu}\right)\Big/\left(\left(d_{jr}^{n+1}\right)^2+\mu\right)\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\phi\left(\left|\Lambda(k)\right|\right)\left(\sqrt{\left(t_{kr}^{n+1}\right)^2+\mu}-\left(t_{kr}^{n+1}\right)^2\Big/\sqrt{\left(t_{kr}^{n+1}\right)^2+\mu}\right)\Big/\left(\left(t_{kr}^{n+1}\right)^2+\mu\right)\right)\left(t_{kr}^{n+1}-t_{kr}^{n}\right)^2\\&-\frac{1}{2}\left(\left|\Lambda(i)\right|+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(\sqrt{\left(a_i^{n+1}\right)^2+\mu}-\left(a_i^{n+1}\right)^2\Big/\sqrt{\left(a_i^{n+1}\right)^2+\mu}\right)\Big/\left(\left(a_i^{n+1}\right)^2+\mu\right)\right)\left(a_i^{n+1}-a_i^{n}\right)^2\\&-\frac{1}{2}\left(\left|\Lambda(j)\right|+\gamma\phi\left(\left|\Lambda(j)\right|\right)\left(\sqrt{\left(b_j^{n+1}\right)^2+\mu}-\left(b_j^{n+1}\right)^2\Big/\sqrt{\left(b_j^{n+1}\right)^2+\mu}\right)\Big/\left(\left(b_j^{n+1}\right)^2+\mu\right)\right)\left(b_j^{n+1}-b_j^{n}\right)^2\\&-\frac{1}{2}\left(\left|\Lambda(k)\right|+\gamma\phi\left(\left|\Lambda(k)\right|\right)\left(\sqrt{\left(c_k^{n+1}\right)^2+\mu}-\left(c_k^{n+1}\right)^2\Big/\sqrt{\left(c_k^{n+1}\right)^2+\mu}\right)\Big/\left(\left(c_k^{n+1}\right)^2+\mu\right)\right)\left(c_k^{n+1}-c_k^{n}\right)^2.\end{aligned}\tag{S13}$$

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

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

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

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

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

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

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

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]:

$$\lim_{n\to+\infty}\left(s_{ir}^{n+1}-s_{ir}^{n}\right)=0, \lim_{n\to+\infty}\left(a_{i}^{n+1}-a_{i}^{n}\right)=0;$$
$$\lim_{n\to+\infty}\left(d_{jr}^{n+1}-d_{jr}^{n}\right)=0, \lim_{n\to+\infty}\left(b_{j}^{n+1}-b_{j}^{n}\right)=0; \tag{S17}$$
$$\lim_{n\to+\infty}\left(t_{kr}^{n+1}-t_{kr}^{n}\right)=0, \lim_{n\to+\infty}\left(c_{k}^{n+1}-c_{k}^{n}\right)=0.$$

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

$$\left.\frac{\partial L}{\partial s_{ir}}\right|_{s_{ir}=s_{ir}^*} = \sum_{y_{ijk}\in\Lambda(i)}\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^*+s_{ir}^*\Big/\sqrt{\left(s_{ir}^*\right)^2+\mu}\right)-\tilde{s}_{ir}^*=0, \tag{S18a}$$

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

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

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

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

$$\tilde{s}_{ir}^* = \sum_{y_{ijk}\in\Lambda(i)}\left(y_{ijk}-\hat{y}_{ijk}\right)\left(-d_{jr}t_{kr}\right)+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^*+s_{ir}^*\Big/\sqrt{\left(s_{ir}^*\right)^2+\mu}\right). \tag{S19}$$

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

$$\xi_{ir}^n=\frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}d_{jr}t_{kr}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^n+s_{ir}^n\Big/\sqrt{\left(s_{ir}^n\right)^2+\mu}\right)}. \tag{S20}$$

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

$$0\le\xi_{ir}^*=\lim_{n\to+\infty}\xi_{ir}^n=\frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}d_{jr}t_{kr}}{\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^*+s_{ir}^*\Big/\sqrt{\left(s_{ir}^*\right)^2+\mu}\right)}. \tag{S21}$$

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:

$$\lim_{n\to+\infty} s_{ir}^*\xi_{ir}^*-s_{ir}^*=0, s_{ir}^*>0\Rightarrow\xi_{ir}^*=1\Rightarrow\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^*+s_{ir}^*\Big/\sqrt{\left(s_{ir}^*\right)^2+\mu}\right)-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}=0. \tag{S24}$$

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

$$\tilde{s}_{ir}^*=\sum_{y_{ijk}\in\Lambda(i)}\hat{y}_{ijk}d_{jr}t_{kr}+\gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^*+s_{ir}^*\Big/\sqrt{\left(s_{ir}^*\right)^2+\mu}\right)-\sum_{y_{ijk}\in\Lambda(i)}y_{ijk}d_{jr}t_{kr}=0\Rightarrow\tilde{s}_{ir}^*\cdot s_{ir}^*=0. \tag{S25}$$

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:

$$s_{ir}^0>0, s_{ir}^0 \lim_{n\to+\infty}\prod_{h=1}^{n}\xi_{ir}^h = s_{ir}^* = 0 \Rightarrow \lim_{n\to+\infty}\prod_{h=1}^{n}\xi_{ir}^h = 0$$
$$\Rightarrow \lim_{n\to+\infty}\xi_{ir}^n = \xi_{ir}^* = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk} d_{jr} t_{kr}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} d_{jr} t_{kr} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^* + s_{ir}^* \Big/ \sqrt{\left(s_{ir}^*\right)^2+\mu}\right)} \le 1 \tag{S27}$$
$$\Rightarrow \tilde{s}_{ir}^* = \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} d_{jr} t_{kr} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(s_{ir}^* + s_{ir}^* \Big/ \sqrt{\left(s_{ir}^*\right)^2+\mu}\right) - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} d_{jr} t_{kr} \ge 0.$$

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

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

Let a_i^* denotes the converging state of sequence $\{a_i^n\}$, 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:

$$\left.\frac{\partial L}{\partial a_i}\right|_{a_i=a_i^*} = \sum_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right) + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^* + a_i^* \Big/ \sqrt{\left(a_i^*\right)^2+\mu}\right) - \tilde{a}_i^* = 0, \tag{S28a}$$
$$\tilde{a}_i^*\cdot a_i^* = 0. \tag{S28b}$$
$$a_i^* \ge 0, \tag{S28c}$$
$$\tilde{a}_i^* \ge 0. \tag{S28d}$$

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

$$\tilde{a}_i^* = \sum_{y_{ijk}\in\Lambda(i)}\left(\left(y_{ijk}-\hat{y}_{ijk}\right)(-1)\right) + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^* + a_i^* \Big/ \sqrt{\left(a_i^*\right)^2+\mu}\right). \tag{S29}$$

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

$$\varsigma_i^n = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^n + a_i^n \Big/ \sqrt{\left(a_i^n\right)^2+\mu}\right)}. \tag{S30}$$

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

$$0 \le \varsigma_i^* = \lim_{n\to+\infty}\varsigma_i^n = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^n + a_i^n \Big/ \sqrt{\left(a_i^n\right)^2+\mu}\right)}. \tag{S31}$$

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), $a_i^* \ge 0$ with a non-negatively initial hypothesis. Hence, we have the following inferences.

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

$$a_i^*\varsigma_i^* - a_i^* = 0, a_i^* \ge 0 \Rightarrow \varsigma_i^* = 1 \Rightarrow \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^* + a_i^* \Big/ \sqrt{\left(a_i^*\right)^2+\mu}\right) - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} = 0. \tag{S34}$$

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

$$\tilde{a}_i^* = \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^* + a_i^* \Big/ \sqrt{\left(a_i^*\right)^2+\mu}\right) - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} = 0 \Rightarrow \tilde{a}_i^*\cdot a_i^* = 0. \tag{S35}$$

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:

$$\begin{aligned} & a_i^0 > 0, a_i^0 \lim_{n\to+\infty} \prod_{h=1}^{n} \varsigma_i^h = a_i^* = 0 \Rightarrow \lim_{n\to+\infty} \prod_{h=1}^{n} \varsigma_i^h = 0 \\ & \Rightarrow \lim_{n\to+\infty} \varsigma_i^n = \varsigma_i^* = \frac{\sum_{y_{ijk}\in\Lambda(i)} y_{ijk}}{\sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^* + a_i^* \Big/ \sqrt{\left(a_i^*\right)^2 + \mu}\right)} \le 1 \\ & \Rightarrow \tilde{a}_i^* = \sum_{y_{ijk}\in\Lambda(i)} \hat{y}_{ijk} + \gamma\phi\left(\left|\Lambda(i)\right|\right)\left(a_i^* + a_i^* \Big/ \sqrt{\left(a_i^*\right)^2 + \mu}\right) - \sum_{y_{ijk}\in\Lambda(i)} y_{ijk} \ge 0. \end{aligned} \tag{S37}$$

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 ER-NLFT model's convergence on a nonnegative HDI tensor is guaranteed.

III. Supplementary Tables and Figures of Empirical Studies

TABLE S.I
Hyper-parameter settings of M1-M6.

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

TABLE S.II
RMSE, MAE, Win/Loss Counts and Friedman test of M1-6 on D1-6.

Dataset	Metric	M1	M2	M3	M4	M5	M6
D1	RMSE	**0.2826±0.0005**	0.3091±0.0006●	0.3082±0.0041●	0.3189±0.0006●	Intractable	Intractable
	MAE	**0.2088±0.0004**	0.2171±0.0004●	0.2207±0.0010●	0.2375±0.0005●	Intractable	Intractable
D2	RMSE	**0.2989±0.0004**	0.3221±0.0003●	0.3223±0.0025●	0.3342±0.0007●	Intractable	Intractable
	MAE	**0.2228±0.0003**	0.2287±0.0002●	0.2340±0.0010●	0.2499±0.0005●	Intractable	Intractable
D3	RMSE	**0.2597±0.0003**	0.2723±0.0005●	0.2920±0.0054●	0.2837±0.0005●	Intractable	Intractable
	MAE	**0.1817±0.0003**	0.1823±0.0003●	0.1834±0.0014●	0.1938±0.0003●	Intractable	Intractable
D4	RMSE	**0.2629±0.0005**	0.2738±0.0004●	0.2935±0.0043●	0.2895±0.0005●	Intractable	Intractable
	MAE	**0.1862±0.0004**	0.1868±0.0004●	0.1871±0.0005●	0.1997±0.0006●	Intractable	Intractable
D5	RMSE	**0.3213±0.0020**	0.3311±0.0026●	0.3315±0.0024●	0.3337±0.0019●	0.3346±0.0022●	0.3529±0.0025●
	MAE	**0.2125±0.0009**	0.2136±0.0009●	0.2143±0.0010●	0.2229±0.0006●	0.2426±0.0012●	0.2570±0.0019●
D6	RMSE	**0.3255±0.0018**	0.3382±0.0018●	0.3358±0.0053●	0.3457±0.0042●	0.3337±0.0027●	0.3506±0.0015●
	MAE	**0.2221±0.0019**	0.2258±0.0011●	0.2249±0.0021●	0.2355±0.0021●	0.2420±0.0023●	0.2569±0.0015●
Statistical analysis	Win/Loss	--	12/0	12/0	12/0	4/0	4/0
	F-rank	1.00	2.33	3.00	3.92	5.08	5.67

● indicates M1 has a lower RMSE/MAE than the comparison models.

TABLE S.III
Total Time Cost in RMSE/MAE(Seconds), Win/Loss Counts and Friedman test of M1-6 on D1-6.

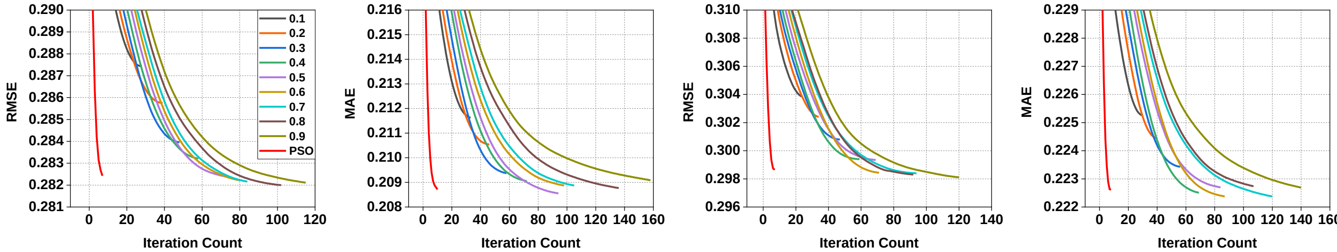
Dataset	Metric	M1	M2	M3	M4	M5	M6
D1	RMSE	**7338±1380**	495716±18683●	303003±43492●	147879±9603●	Intractable	Intractable
	MAE	**9942±1923**	642177±123801●	312970±60759●	156782±16735●	Intractable	Intractable
D2	RMSE	**5274±590**	510699±63729●	182743±43469●	87766±5855●	Intractable	Intractable
	MAE	**6592±590**	722891±71839●	188198±41723●	99132±7937●	Intractable	Intractable
D3	RMSE	**442±52**	104987±4051●	35755±3367●	11888±737●	Intractable	Intractable
	MAE	**814±52**	144982±6662●	32238±7620●	20850±963●	Intractable	Intractable
D4	RMSE	**380±77**	91194±2012●	19394±4173●	7931±428●	Intractable	Intractable
	MAE	**605±71**	120565±11957●	20741±3418●	13078±734●	Intractable	Intractable
D5	RMSE	**11.95±3.09**	6583.24±0.00●	327.95±79.17●	175.14±7.95●	3603082±0●	1157406±106958●
	MAE	**57.76±8.44**	6583.24±0.00●	327.95±82.29●	158.93±4.59●	3603082±0●	1150391±128675●
D6	RMSE	**2.40±0.70**	4102.06±0.00●	118.56±30.25●	74.13±10.43●	517117±0●	93264±5589●
	MAE	**12.57±1.11**	4102.06±0.00●	114.00±39.23●	63.61±5.03●	517117±0●	93753±5694●
Statistical analysis	Win/Loss	--	12/0	12/0	12/0	4/0	4/0
	F-rank	1.00	4.00	3.00	2.00	5.67	5.33

● indicates M1 has a less total time cost than the comparison models in RMSE/MAE.

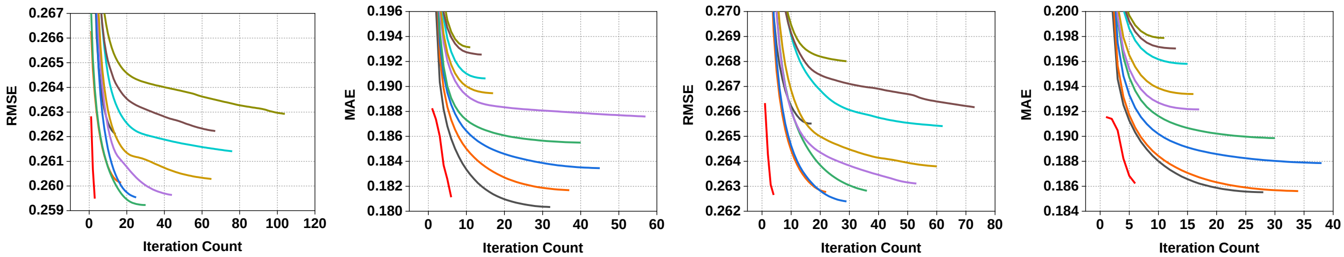
TABLE S.IV
Iteration Count in RMSE/MAE, Win/Loss Counts and Friedman test of M1-6 on D1-6.

Dataset	Metric	M1	M2	M3	M4	M5	M6
D1	RMSE	**6±1**	29±2●	25±4●	50±3●	Intractable	Intractable
	MAE	**9±2**	38±7●	26±5●	53±3●	Intractable	Intractable
D2	RMSE	**7±1**	47±6●	22±5●	46±3●	Intractable	Intractable
	MAE	**8±1**	67±7●	23±5●	52±4●	Intractable	Intractable
D3	RMSE	**3±1**	67±4●	31±3●	43±3●	Intractable	Intractable
	MAE	**6±1**	92±5●	28±6●	76±4●	Intractable	Intractable
D4	RMSE	**4±1**	74±3●	24±5●	44±3●	Intractable	Intractable
	MAE	**6±1**	98±9●	26±4●	72±4●	Intractable	Intractable
D5	RMSE	**5±1**	200±0●	7±2●	36±2●	200±0●	66±6●
	MAE	24±3	200±0●	**7±3**	33±2●	200±0●	66±7●
D6	RMSE	**4±1**	200±0●	9±3●	49±7●	200±0●	48±3●
	MAE	23±2	200±0●	**8±3**	42±4●	200±0●	48±3●
Statistical analysis	Win/Loss	--	12/0	10/2	12/0	4/0	4/0
	F-rank	1.17	4.33	1.83	3.25	5.50	4.92

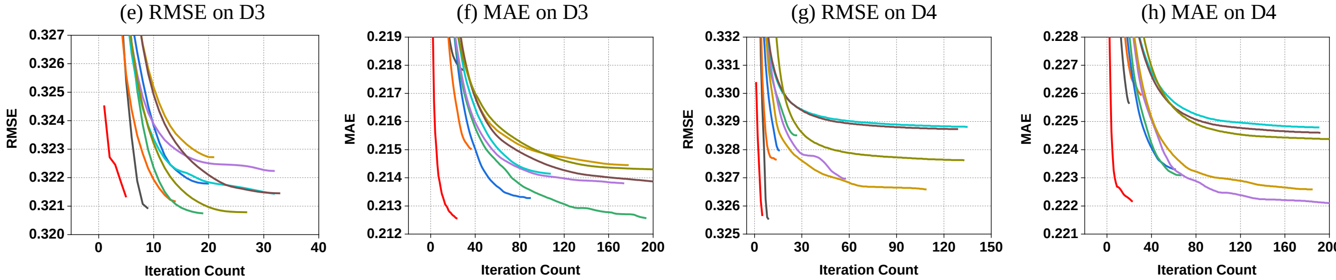
● indicates M1 has a less converging iteration count than the comparison models in RMSE/MAE.



(a) RMSE on D1 (b) MAE on D1 (c) RMSE on D2 (d) MAE on D2

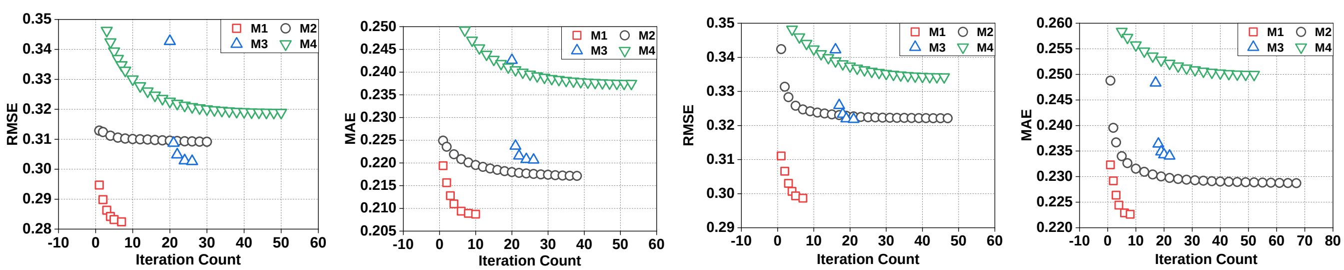


(e) RMSE on D3 (f) MAE on D3 (g) RMSE on D4 (h) MAE on D4

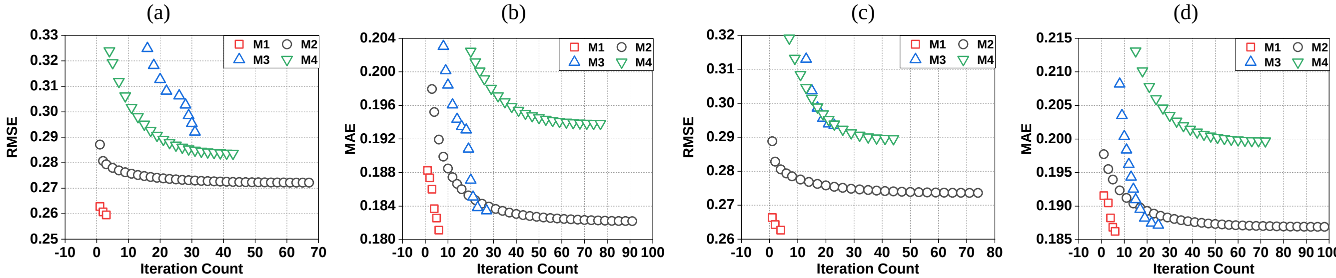


(i) RMSE on D5 (j) MAE on D5 (k) RMSE on D6 (l) MAE on D6

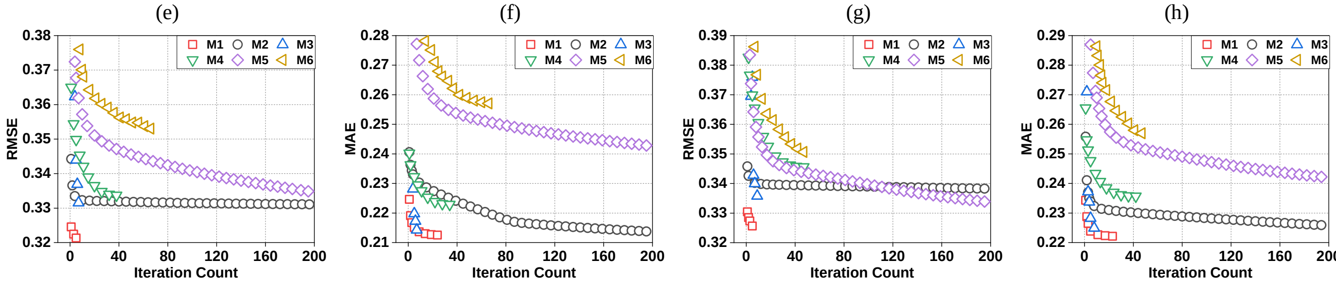
Fig. S1. Performance of ER-NLFT as γ varies on D1-6.



(a) (b) (c) (d)



(e) (f) (g) (h)



(i) (j) (k) (l)

Fig. S2. Training curves of M1-6 in RMSE and MAE on D1-6.