A Proximal-ADMM-incorporated Nonnegative Latent-Factorization-of-Tensors Model for Representing Dynamic Cryptocurrency Transaction Network Supplementary File

Xin Liao, Hao Wu, *Member, IEEE,* Tiantian He, *Member, IEEE,* and Xin Luo, *Fellow, IEEE*

I. INTRODUCTION

This is the supplementary file for paper entitled *A Proximal-ADMM-incorporated Nonnegative Latent-Factorization-of-Tensors for Dynamic Cryptocurrency Transaction Network Embedding*. The convergence proof of PNL, supplementary procedure, and experimental results are put into this file.

II. CONVERGENCE PROOF OF PNL

Given $i\in I$, $j\in J$, and $k\in K$, the PNL model's convergence proof is presented as follows:

(a) Proof of Step 1: Note that we present the proof procedure for variable $\hat{u}_{ir}$, u_{ir}, and d_{ir}, and the similar variables also applies to the same conclusion.

***Lemma* 1**. With (11), $\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2$, $\left(h_{jr}^{t+1}-h_{jr}^{t}\right)^2$, and $\left(l_{kr}^{t+1}-l_{kr}^{t}\right)^2$ are bounded as:

$$\begin{aligned}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2 \le\ & 8\left(\left(\eta-1\right)^2\alpha_i^2+\rho^2\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2+8\alpha_i^2\left(\left(\eta-1\right)^2+1\right)\left(s_{ir}^{t+1}-s_{ir}^{t}\right)^2\\ &+8\rho^2\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)^2+8\alpha_i^2\left(u_{ir}^{t}-u_{ir}^{t-1}\right)^2+4\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)^2=\varphi_d\end{aligned} \tag{S1}$$

Where Δ_{ir}^{t+1} is defined as:

$$\Delta_{ir}^{t+1}=\sum_{y_{ijk}\in\Lambda}\left(y_{ijk}-\left(\sum_{f_1=1}^{r-1}\hat{u}_{if_1}^{t+1}\hat{v}_{jf_1}^{t+1}\hat{w}_{kf_1}^{t+1}+\hat{u}_{ir}^{t+1}\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}+\sum_{f_2=r+1}^{R}\hat{u}_{if_2}^{t}\hat{v}_{jf_2}^{t}\hat{w}_{if_2}^{t}+\hat{a}_i^{t}+\hat{c}_j^{t}+\hat{e}_k^{t}\right)\right)\left(-\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right). \tag{S2}$$

***Proof* 1**. Note that (7) is non-convex and its zero-gradient points, such as local/global optimum and saddle point, should be regarded as a feasible solution. Therefore, assuming that $\hat{u}_{ir}^{t+1}$ is the solution to $\hat{u}_{ir}$ by (11), the following condition is fulfilled:

$$\Delta_{ir}^{t+1}+\alpha_i\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t}+\frac{d_{ir}^{t}}{\alpha_i}\right)+\rho\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)=0. \tag{S3}$$

By substituting the update rule in (11) and (15) into (S3), the following equation is achieved:

$$d_{ir}^{t+1}=\left(\eta-1\right)\alpha_i\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t+1}\right)-\alpha_i\left(u_{ir}^{t+1}-u_{ir}^{t}\right)-\rho\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)-\Delta_{ir}^{t+1}. \tag{S4}$$

Further, the difference between d_{ir}^{t+1} and d_{ir}^{t} is given as:

$$\begin{aligned}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2=\ & (\left(\eta-1\right)\alpha_i\left(\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t+1}\right)-\left(\hat{u}_{ir}^{t}-u_{ir}^{t}\right)\right)-\alpha_i\left(\left(u_{ir}^{t+1}-u_{ir}^{t}\right)-\left(u_{ir}^{t}-u_{ir}^{t-1}\right)\right)\\ &-\rho\left(\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)-\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)\right)-\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right))^2.\end{aligned} \tag{S5}$$

With the inequality $\left(a-b-c-d\right)^2\le 4\left(a^2+b^2+c^2+d^2\right)$, we have:

$$\begin{aligned}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2\le\ & 4\left(\eta-1\right)^2\alpha_i^2\left(\left(\hat{u}_{ir}^{t+1}-u_{ir}^{t+1}\right)-\left(\hat{u}_{ir}^{t}-u_{ir}^{t}\right)\right)^2+4\alpha_i^2\left(\left(u_{ir}^{t+1}-u_{ir}^{t}\right)-\left(u_{ir}^{t}-u_{ir}^{t-1}\right)\right)^2\\ &+4\rho^2\left(\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)-\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)\right)^2+4\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)^2.\end{aligned} \tag{S6}$$

With (S6), we implement (S1) by using the inequality $\left(a-b\right)^2\le 2\left(a^2+b^2\right)$. Note that applying the same principle, we can get $\left(h_{jr}^{t+1}-h_{jr}^{t}\right)^2\le\varphi_h$ and $\left(l_{kr}^{t+1}-l_{kr}^{t}\right)^2\le\varphi_l$. Hence, ***Lemma* 1** stands.

(b) Proof of Step 2: ***Lemma* 1** has been proved, then we perform Step 2. For simplicity, we first introduce seven functions and intermediate variables to express similar structures as follows:

$$F_1\left(\hat{u}_{ir},u_{ir},\alpha_i\right)=\left(\frac{8\left((\eta-1)^2\alpha_i^2+\rho^2\right)}{\eta\alpha_i}-\frac{1}{2}\left(\sum_{i\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^2+\alpha_i+\rho\right)\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2+\left(\frac{8\rho^2}{\eta\alpha_i}\right)\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)^2=F_1^{I}. \tag{S7}$$

With (S7), we define a function expression for $\{\hat{u}_{ir},u_{ir},\alpha_i\}$. Note that the above three variables are related to the I-dimension of the tensor, and we can get F_1^{J} and F_1^{K} for $\{\hat{v}_{jr},v_{jr},\beta_j\}$ in J-dimension and $\{\hat{w}_{kr},w_{kr},\delta_k\}$ in K-dimension by adopting the similar expression. Note that for the term $\sum_{i\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^{2}$ in I-dimension, its expression for the J-dimension and K-dimension are $\sum_{j\in\Lambda(j)}\left(\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}\right)^{2}$ and $\sum_{k\in\Lambda(k)}\left(\hat{u}_{ir}^{t}\hat{v}_{jr}^{t}\right)^{2}$, respectively. Hence, we define the first intermediate variable $A_1 = F_1^I + F_1^J + F_1^K$.

Next the second function expression is as follows:

$$F_2\left(\hat{u}_{ir},u_{ir},\alpha_i,\Delta_{ir}\right)=\left(\left(\frac{\alpha_i}{2}-\frac{8\alpha_i\left((\eta-1)^2+1\right)}{\eta}\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2-\frac{8\alpha_i\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t-1}\right)^2}{\eta}-\frac{4\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)^2}{\eta\alpha_i}\right)=F_2^I. \tag{S8}$$

Similarly, with (S8), we get F_2^I, F_2^J, F_2^K and thus the second intermediate variable $A_2 = F_2^I + F_2^J + F_2^K$.

The third function expression is given as follows:

$$F_3\left(\hat{u}_{ir},u_{ir},\alpha_i\right)=\sum_{i\in I}\alpha_i\left(\sum_{f_1=1}^{r-1}\left(\hat{u}_{if_1}^{t+1}-u_{if_1}^{t+1}\right)^2+\sum_{f_2=r}^{R}\left(\hat{u}_{if_2}^{t}-u_{if_2}^{t}\right)^2\right)=F_3^I. \tag{S9}$$

With (S9), the third intermediate variable is defined as $A_3 = F_3^I + F_3^J + F_3^K$. Similarly, the fourth function expression is:

$$F_4\left(\hat{u}_{ir}\right)=\frac{\rho}{2}\left(\sum_{i\in I}\left(\sum_{f_1=1}^{r-1}\left(\hat{u}_{if_1}^{t+1}-\hat{u}_{if_1}^{t}\right)^2+\sum_{f_2=r}^{R}\left(\hat{u}_{if_2}^{t}-\hat{u}_{if_2}^{t-1}\right)^2\right)\right)=F_4^I. \tag{S10}$$

Correspondingly, the fourth intermediate variable is $A_4 = F_4^I + F_4^J + F_4^K$. The next fifth function expression is given as follows:

$$\begin{aligned}F_5\left(\hat{u}_{ir},u_{ir},\alpha_i,\Delta_{ir}\right)=&\sum_{i\in I}\left(\sum_{f_1=2}^{r-1}\left(\alpha_i\left(u_{if_1}^{t+1}-u_{if_1}^{t}\right)+\Delta_{if_1}^{t+1}+\rho\left(\hat{u}_{if_1}^{t+1}-\hat{u}_{if_1}^{t}\right)\right)\left(u_{if_1}^{t+1}-\hat{u}_{if_1}^{t+1}\right)\right)\\&+\sum_{i\in I}\left(\sum_{f_2=r}^{R}\left(\alpha_i\left(u_{if_2}^{t}-u_{if_2}^{t-1}\right)+\Delta_{if_2}^{t}+\rho\left(\hat{u}_{if_2}^{t}-\hat{u}_{if_2}^{t-1}\right)\right)\left(u_{if_2}^{t}-u_{if_2}^{t}\right)\right)=F_5^I.\end{aligned} \tag{S11}$$

With (S11), the fifth intermediate variable is $A_5 = F_5^I + F_5^J + F_5^K$.

A functional expression is given as:

$$F_6\left(\hat{u}_{ir},u_{ir},d_{ir}\right)=\sum_{i\in I}\left(\sum_{f_1=1}^{r-1}d_{if_1}^{t+1}\left(\hat{u}_{if_1}^{t+1}-u_{if_1}^{t+1}\right)+\sum_{f_2=r}^{R}d_{if_2}^{t}\left(\hat{u}_{if_2}^{t}-u_{if_2}^{t}\right)\right)=F_6^I. \tag{S12}$$

With (S12), we can get the J-dimension and K-dimension versions, F_6^J and F_6^K. Hence, the sixth intermediate variable is $A_6 = F_6^I + F_6^J + F_6^K$. Finally, a functional expression for the bias is given as:

$$F_6\left(\hat{u}_{ir},u_{ir},d_{ir}\right)=\sum_{i\in I}\left(\sum_{f_1=1}^{r-1}d_{if_1}^{t+1}\left(\hat{u}_{if_1}^{t+1}-u_{if_1}^{t+1}\right)+\sum_{f_2=r}^{R}d_{if_2}^{t}\left(\hat{u}_{if_2}^{t}-u_{if_2}^{t}\right)\right)=F_6^I. \tag{S13}$$

With (S13), it defines a function expression for $\{\hat{a}_i,a_i,\sigma_i,o_i\}$ and is related to the I- dimension. We can obtain expressions about the J-dimension and K-dimension by similar expressions for $\{\hat{c}_j,c_j,\phi_j,s_j\}$ and $\{\hat{e}_k,e_k,\psi_k,z_k\}$, i.e., F_7^J and F_7^K. Therefore, we get the last intermediate variable as $A_7 = F_7^I + F_7^J + F_7^K$.

***Lemma* 2**. If the following inequality is satisfied:

$$A_1 \le A_2. \tag{S14}$$

Then the following inequalities holds:

$$L_p\left(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t+1}\right)-L_p\left(\mathcal{D}_1^{t}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}\right)\le 0. \tag{S15}$$

Note that we have:

$$L_p\left(\mathcal{D}_1^{t}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}\right)\ge 0. \tag{S16}$$

With the following condition:

$$\eta \ge \frac{1}{2}-\frac{A_4+A_5+A_7}{A_3}. \tag{S17}$$

***Proof* 2**. By expanding the second-order Taylor expansion of L_p at the point $\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}$, we can get the following equality:

$$\begin{aligned}L_p\left(D_1^{t+1}\cup D_2^{t}\cup D_3^{t}\right)-L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)\overset{(11)}{=}&-\frac{1}{2}\left(\sum_{y_{ik}\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^2+\alpha_i+\rho\right)\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)^2\\-\frac{1}{2}\left(\sum_{y_{ik}\in\Lambda(j)}\left(\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}\right)^2+\beta_j+\rho\right)\left(\hat{v}_{jr}^{t+1}-\hat{v}_{jr}^{t}\right)^2&-\frac{1}{2}\left(\sum_{y_{ik}\in\Lambda(k)}\left(\hat{u}_{ir}^{t}\hat{v}_{jr}^{t}\right)^2+\delta_k+\rho\right)\left(\hat{w}_{kr}^{t+1}-\hat{w}_{kr}^{t}\right)^2.\end{aligned} \tag{S18}$$

Note that the equality (S18) holds when the first-order terms in (9) are equal zero. Therefore, the difference between $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t})$ and $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t})$ is given as:

$$L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t}\right)-L_p\left(D_1^{t+1}\cup D_2^{t}\cup D_3^{t}\right)\overset{(9)}{\le}-\frac{\alpha_i}{2}\left(u_{ir}^{t}-u_{ir}^{t+1}\right)^2-\frac{\beta_j}{2}\left(v_{jr}^{t}-v_{jr}^{t+1}\right)^2-\frac{\delta_k}{2}\left(w_{kr}^{t}-w_{kr}^{t+1}\right)^2. \tag{S19}$$

Note that the inequality (S19) considers the optimal condition of (11) and the projection rule of (14), the first-order terms equal to or less than zero are omitted. Therefore, the difference between $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t+1})$ and $L_p(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t})$ is:

$$\begin{aligned}L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t+1}\right)-L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t}\right)&\overset{(11),(14)}{=}\frac{\left(d_{ir}^{t+1}-d_{ir}^{t}\right)^2}{\eta\alpha_i}+\frac{\left(h_{jr}^{t+1}-h_{jr}^{t}\right)^2}{\eta\beta_j}+\frac{\left(l_{kr}^{t+1}-l_{kr}^{t}\right)^2}{\eta\delta_k}\\&\overset{(S1)}{\le}\frac{\varphi_d}{\eta\alpha_i}+\frac{\varphi_h}{\eta\beta_j}+\frac{\varphi_l}{\eta\delta_k},\end{aligned} \tag{S20}$$

where the equality depends on (11) and (15), and the inequality follows ***Lemma* 1**. With (S18)-(S20), we have:

$$L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t+1}\right)-L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)\le A_1-A_2. \tag{S21}$$

Therefore, with (S14), the following inequality evidently holds:

$$L_p\left(D_1^{t+1}\cup D_2^{t+1}\cup D_3^{t+1}\right)-L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)\le 0. \tag{S22}$$

Hence, the proximal-incorporated augmented Lagrangian of (7) is non-increasing. Moreover, after the *t*-th iteration, (7) is formulated as:

$$L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)=\frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(y_{ijk}-\sum_{f_1=1}^{r-1}\hat{u}_{if_1}^{t+1}\hat{v}_{jf_1}^{t+1}\hat{w}_{kf_1}^{t+1}-\sum_{f_2=r}^{R}\hat{u}_{if_2}^{t}\hat{v}_{jf_2}^{t}\hat{w}_{kf_2}^{t}-\hat{a}_i^{t}-\hat{c}_j^{t}-\hat{e}_k^{t}\right)^2+\frac{A_3}{2}+A_4+A_6+A_7. \tag{S23}$$

By substituting (S4) into (S23), the following deduction is achieved:

$$L_p\left(D_1^{t}\cup D_2^{t}\cup D_3^{t}\right)=\frac{1}{2}\sum_{y_{ijk}\in\Lambda}\left(y_{ijk}-\sum_{f_1=1}^{r-1}\hat{u}_{if_1}^{t+1}\hat{v}_{jf_1}^{t+1}\hat{w}_{kf_1}^{t+1}-\sum_{f_2=r}^{R}\hat{u}_{if_2}^{t}\hat{v}_{jf_2}^{t}\hat{w}_{kf_2}^{t}-\hat{a}_i^{t}-\hat{c}_j^{t}-\hat{e}_k^{t}\right)^2+\frac{(2\eta-1)A_3}{2}+A_4+A_5+A_7. \tag{S24}$$

With (S17) and (S24), (S16) is fulfilled, i.e., (7) is lower-bounded. Therefore, ***Lemma* 2** holds.

(c) Proof of Step 3: Considering $\mathrm{U}\ge 0$, $\mathrm{V}\ge 0$, and $\mathrm{W}\ge 0$, the proximal augmented Lagrangian of (7) is extended as:

$$L_p^{\#}=L_p-tr(\mathrm{GU})-tr(\mathrm{MV})-tr(\mathrm{PW})=L_p-\sum_{i\in I}\sum_{r=1}^{R}g_{ir}u_{ir}-\sum_{j\in J}\sum_{r=1}^{R}m_{jr}v_{jr}-\sum_{k\in K}\sum_{r=1}^{R}p_{kr}w_{kr}, \tag{S25}$$

where $tr(\cdot)$ calculates the trace of an involved matrix. G, M, and P denote Lagrangian multiplier for PNL's nonnegative constraint. Then, ***Theorem* 1** is presented:

***Theorem* 1**. If the following conditions hold:

$$\begin{cases}\dfrac{8\left((\eta-1)^2(\alpha_i)^2+\rho^2\right)}{\eta\alpha_i}\ne-\dfrac{1}{2}\left(\displaystyle\sum_{y_{ijk}\in\Lambda(i)}\left(\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}\right)^2+\alpha_i+\rho\right)\\ \dfrac{8\left((\eta-1)^2(\beta_j)^2+\rho^2\right)}{\eta\beta_j}\ne-\dfrac{1}{2}\left(\displaystyle\sum_{y_{ijk}\in\Lambda(j)}\left(\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}\right)^2+\beta_j+\rho\right)\\ \dfrac{8\left((\eta-1)^2(\delta_k)^2+\rho^2\right)}{\eta\delta_k}\ne-\dfrac{1}{2}\left(\displaystyle\sum_{y_{ijk}\in\Lambda(k)}\left(\hat{u}_{ir}^{t}\hat{v}_{jr}^{t}\right)^2+\delta_k+\rho\right)\\ \dfrac{\alpha_i}{2}\ne\dfrac{8\alpha_i\left((\eta-1)^2+1\right)}{\eta},\dfrac{\beta_j}{2}\ne\dfrac{8\beta_j\left((\eta-1)^2+1\right)}{\eta},\dfrac{\delta_k}{2}\ne\dfrac{8\delta_k\left((\eta-1)^2+1\right)}{\eta}\\ \rho\ne 0,\eta\ne 0,\alpha_i\ne 0,\beta_j\ne 0,\delta_k\ne 0.\end{cases} \tag{S26}$$

With (11), the equilibrium point $\mathcal{D}_1^{*}\cup\mathcal{D}_2^{*}\cup\mathcal{D}_3^{*}$ of $\mathcal{D}_1\cup\mathcal{D}_2\cup\mathcal{D}_3$ is a KKT stationary point, and the following KKT conditions holds:

$$\hat{u}_{ir}^{*}-u_{ir}^{*}=0,\hat{v}_{jr}^{*}-v_{jr}^{*}=0,\hat{w}_{kr}^{*}-w_{kr}^{*}=0, \tag{S27a}$$

$$\left.\frac{\partial L_p^{\#}}{\partial\hat{u}_{ir}}\right|_{\hat{u}_{ir}=\hat{u}_{ir}^{*}}=\Delta_{ir}^{*}+d_{ir}^{*}=0;\left.\frac{\partial L_p^{\#}}{\partial\hat{v}_{jr}}\right|_{\hat{v}_{jr}=\hat{v}_{jr}^{*}}=\Delta_{jr}^{*}+h_{jr}^{*}=0;\left.\frac{\partial L_p^{\#}}{\partial\hat{w}_{kr}}\right|_{\hat{w}_{kr}=\hat{w}_{kr}^{*}}=\Delta_{kr}^{*}+l_{kr}^{*}=0, \tag{S27b}$$

$$\left.\frac{\partial L_p^{\#}}{\partial u_{ir}}\right|_{u_{ir}=u_{ir}^*}=-\alpha_i\left(\hat{u}_{ir}^*-u_{ir}^*+\frac{d_{ir}^*}{\alpha_i}\right)-g_{ir}^*=0;\left.\frac{\partial L_p^{\#}}{\partial v_{jr}}\right|_{v_{jr}=v_{jr}^*}=-\beta_j\left(\hat{v}_{jr}^*-v_{jr}^*+\frac{h_{jr}^*}{\beta_j}\right)-m_{jr}^*=0;\left.\frac{\partial L_p^{\#}}{\partial w_{kr}}\right|_{w_{kr}=w_{kr}^*}=-\delta_k\left(\hat{w}_{kr}^*-w_{kr}^*+\frac{l_{kr}^*}{\delta_k}\right)-p_{kr}^*=0, \tag{S27c}$$

$$g_{ir}^*u_{ir}^*=0, m_{jr}^*v_{jr}^*=0, p_{kr}^*w_{kr}^*=0, \tag{S27d}$$

$$u_{ir}^*\geq 0, v_{jr}^*\geq, w_{kr}^*\geq 0, \tag{S27e}$$

$$g_{ir}^*\geq 0, m_{jr}^*\geq 0, p_{kr}^*\geq 0. \tag{S27f}$$

***Proof* 3**. With ***Lemma* 2**, the following inequality holds since $t\to\infty$:

$$L_p\left(\mathcal{D}_1^{t+1}\cup\mathcal{D}_2^{t+1}\cup\mathcal{D}_3^{t+1}\right)-L_p\left(\mathcal{D}_1^{t}\cup\mathcal{D}_2^{t}\cup\mathcal{D}_3^{t}\right)\leq A_1-A_2\to 0. \tag{S28}$$

Based on (S24) and (S28), we can get:

$$\begin{cases}\lim\limits_{t\to\infty}\left(\hat{u}_{ir}^{t+1}-\hat{u}_{ir}^{t}\right)\to 0,\lim\limits_{t\to\infty}\left(u_{ir}^{t+1}-u_{ir}^{t}\right)\to 0,\lim\limits_{t\to\infty}\left(\Delta_{ir}^{t+1}-\Delta_{ir}^{t}\right)\to 0\\ \lim\limits_{t\to\infty}\left(\hat{v}_{jr}^{t+1}-\hat{v}_{jr}^{t}\right)\to 0,\lim\limits_{t\to\infty}\left(v_{jr}^{t+1}-v_{jr}^{t}\right)\to 0,\lim\limits_{t\to\infty}\left(\Delta_{jr}^{t+1}-\Delta_{jr}^{t}\right)\to 0\\ \lim\limits_{t\to\infty}\left(\hat{w}_{kr}^{t+1}-\hat{w}_{kr}^{t}\right)\to 0,\lim\limits_{t\to\infty}\left(w_{kr}^{t+1}-w_{kr}^{t}\right)\to 0,\lim\limits_{t\to\infty}\left(\Delta_{kr}^{t+1}-\Delta_{kr}^{t}\right)\to 0.\end{cases} \tag{S29}$$

With (S1), we infer that:

$$\lim_{t\to\infty}\left(d_{ir}^{t+1}-d_{ir}^{t}\right)\to 0;\lim_{t\to\infty}\left(h_{jr}^{t+1}-h_{jr}^{t}\right)\to 0;\lim_{t\to\infty}\left(l_{kr}^{t+1}-l_{kr}^{t}\right)\to 0. \tag{S30}$$

According to (15) and (S30), (S27a) holds. (11) can be reconstructed as:

$$\begin{cases}\left(\hat{u}_{ir}^{t}-\hat{u}_{ir}^{t+1}\right)\left(\sum\limits_{y_{ijk}\in\Lambda(i)}\hat{v}_{jr}^{t}\hat{w}_{kr}^{t}+\alpha_i+\rho\right)=\Delta_{ir}^{t+1}+\alpha_i\left(\hat{u}_{ir}^{t}-u_{ir}^{t}\right)+d_{ir}^{t}\\ \left(\hat{v}_{jr}^{t}-\hat{v}_{jr}^{t+1}\right)\left(\sum\limits_{y_{ijk}\in\Lambda(j)}\hat{u}_{ir}^{t}\hat{w}_{kr}^{t}+\beta_j+\rho\right)=\Delta_{jr}^{t+1}+\beta_j\left(\hat{v}_{jr}^{t}-v_{jr}^{t}\right)+h_{jr}^{t}\\ \left(\hat{w}_{kr}^{t}-\hat{w}_{kr}^{t+1}\right)\left(\sum\limits_{y_{ijk}\in\Lambda(k)}\hat{u}_{ir}^{t}\tilde{v}_{jr}^{t}+\delta_k+\rho\right)=\Delta_{kr}^{t+1}+\delta_k\left(\hat{w}_{kr}^{t}-w_{kr}^{t}\right)+l_{kr}^{t}.\end{cases} \tag{S31}$$

Note that (S27b) holds via (S27a), (S29), and (S31). Further, the following inference holds by applying the partial derivation of $L_p^{\#}$ to u_{ir}, v_{jr}, and w_{kr}:

$$\begin{cases}\frac{\partial L_p^{\#}}{\partial u_{ir}}=-\alpha_i\left(\hat{u}_{ir}-u_{ir}+\frac{d_{ir}}{\alpha_i}\right)-g_{ir}=0\\ \frac{\partial L_p^{\#}}{\partial v_{jr}}=-\beta_j\left(\hat{v}_{jr}-v_{jr}+\frac{h_{jr}}{\beta_j}\right)-m_{jr}=0\\ \frac{\partial L_p^{\#}}{\partial w_{kr}}=-\delta_k\left(\hat{t}_{kr}-w_{kr}+\frac{l_{kr}}{\delta_k}\right)-p_{kr}=0\end{cases}\Rightarrow\begin{cases}g_{ir}=-\alpha_i\left(\hat{u}_{ir}-u_{ir}+\frac{d_{ir}}{\alpha_i}\right)\\ m_{jr}=-\beta_j\left(\hat{v}_{jr}-v_{jr}+\frac{h_{jr}}{\beta_j}\right)\\ p_{kr}=-\gamma_k\left(\hat{w}_{kr}-w_{kr}+\frac{l_{kr}}{\delta_k}\right).\end{cases} \tag{S32}$$

With (S32), g_{ir}, m_{jr}, and p_{kr} are implicitly updated and generate their limits g_{ir}^*, m_{jr}^*, p_{kr}^*. Considering (S26)'s KKT conditions that $\forall g_{ir},u_{ir}:g_{ir}u_{ir}=0$, $\forall m_{jr},v_{jr}:m_{jr}v_{jr}=0$, and $\forall p_{kr},w_{kr}:p_{kr}w_{kr}=0$, we have:

$$\begin{cases}-\alpha_i u_{ir}\left(\hat{u}_{ir}-u_{ir}+\frac{d_{ir}}{\alpha_i}\right)=0\\ -\beta_j v_{jr}\left(\hat{v}_{jr}-v_{jr}+\frac{h_{jr}}{\beta_j}\right)=0\\ -\delta_k w_{kr}\left(\hat{w}_{kr}-w_{kr}+\frac{l_{kr}}{\delta_k}\right)=0\end{cases}\Rightarrow\begin{cases}u_{ir}=\hat{u}_{ir}+\frac{d_{ir}}{\alpha_i}\\ v_{jr}=\hat{v}_{jr}+\frac{h_{jr}}{\beta_j}\\ w_{kr}=\hat{w}_{kr}+\frac{l_{kr}}{\delta_k}.\end{cases} \tag{S33}$$

Hence, we can update u_{ir}, v_{jr}, and w_{kr} by (S33). Additionally, the nonnegative truncation is applied to u_{ir}, v_{jr}, and w_{kr} to ensure its nonnegativity and (S26)'s KKT conditions, it's given as:

$$u_{ir}=\max\left(0,\hat{u}_{ir}+\frac{d_{ir}}{\alpha_i}\right);v_{jr}=\max\left(0,\hat{v}_{jr}+\frac{h_{jr}}{\beta_j}\right);w_{kr}=\max\left(0,\hat{w}_{kr}+\frac{l_{kr}}{\delta_k}\right). \tag{S34}$$

Note that the update rules of (14) and (S34) are equivalent. When $t\to\infty$, (S27c)-(S27e) are hold via (14) and (S32)-(S34).

Hence, considering (S27e)'s conditions that $g_{ir}^* \geq 0$, $m_{jr}^* \geq 0$, and $p_{kr}^* \geq 0$, there are two cases as:

- If $u_{ir}^* = 0$, $v_{jr}^* = 0$, and $w_{kr}^* = 0$, the following inequality holds according to (14):

$$\hat{u}_{ir}^* + \frac{d_{ir}^*}{\alpha_i} \leq 0; \hat{v}_{jr}^* + \frac{h_{jr}^*}{\beta_j} \leq 0; \hat{w}_{kr}^* + \frac{l_{kr}^*}{\delta_k} \leq 0, \tag{S35}$$

which indicates that $g_{ir}^* \geq 0$, $m_{jr}^* \geq 0$, and $p_{kr}^* \geq 0$ by collectively analyzing (S32);

- If $u_{ir}^* > 0$, $v_{jr}^* > 0$, and $w_{kr}^* > 0$, the following equality is inferred according to (14):

$$u_{ir}^* = \hat{u}_{ir}^* + \frac{d_{ir}^*}{\alpha_i}; v_{jr}^* = \hat{v}_{jr}^* + \frac{h_{jr}^*}{\beta_j}; w_{kr}^* = \hat{w}_{kr}^* + \frac{l_{kr}^*}{\delta_k}. \tag{S36}$$

With (S32) and (S36), $g_{ir}^* = 0$, $m_{jr}^* = 0$, and $p_{kr}^* = 0$. Therefore, (S27f) holds and ***Theorem* 1** holds. Based on the above inferences, the implemented steps 1-3 demonstrate that the PNL's convergence is theoretically guaranteed.

III. ADDITIONAL PROCEDURES

Procedure: Parallel_update_$\hat{\mathrm{W}}$

Input: $\hat{\boldsymbol{u}}_r, \hat{\boldsymbol{v}}_r, \hat{\boldsymbol{w}}_r, \boldsymbol{u}_r, \boldsymbol{d}_r, \boldsymbol{\tau}_{(p)}, \Lambda, K, Q$

Output: Updated $\hat{\mathrm{W}}$, W, L

Operation	Cost
1. Init $\mathrm{W\_U}^{\lvert K\rvert\times R}, \mathrm{W\_D}^{\lvert K\rvert\times R} = 0$	$\Theta(2\times N\times R)$
2. for each $q\in Q$ **Parallelization**	$\times Q$
3. for each r=1 **to** R **do**	$\times R$
4. for each $y_{ijk}\in\Lambda_q$	$\times\lvert\Lambda_q\rvert$
5. $err = y_{ijk} - \tilde{y}_{ijk}$	$\Theta(1)$
6. $\mathrm{W\_U}_{kr} += \hat{u}_{ir}\hat{v}_{jr}(err + \hat{u}_{ir}\hat{v}_{jr}\hat{w}_{kr})$	$\Theta(1)$
7. $\mathrm{W\_D}_{kr} += (\hat{u}_{ir}\hat{v}_{jr})^2$	$\Theta(1)$
8. for each $k\in K_q$	$\times\lvert K_q\rvert$
9. $\hat{w}_{kr} = \dfrac{\mathrm{W\_U}_{kr} + \lambda_{(p)}\lvert\Lambda_{(k)}\rvert w_{kr} - l_{kr} + \rho_{(p)}\hat{w}_{kr}^{t}}{\mathrm{W\_D}_{kr} + \lambda_{(p)}\lvert\Lambda_{(k)}\rvert + \rho_{(p)}}$	$\Theta(1)$
10. $w_{kr} = \max\left(0, \hat{w}_{kr} + \dfrac{l_{kr}}{\lambda_{(p)}\lvert\Lambda_{(k)}\rvert}\right)$	$\Theta(1)$
11. $l_{kr} += \eta_{(p)}\lambda_{(p)}\lvert\Lambda_{(k)}\rvert(\hat{w}_{kr} - w_{kr})$	$\Theta(1)$

Procedure: Parallel_update_$\hat{\boldsymbol{e}}$

Input: $\hat{\boldsymbol{e}}, \boldsymbol{e}, \boldsymbol{z}, \boldsymbol{\tau}_{(p)}, \Lambda, K, Q$

Output: Updated $\hat{\boldsymbol{e}}$, $\boldsymbol{e}$, $\boldsymbol{z}$

Operation	Cost
1. Init $\mathrm{E\_U}^{\lvert N\rvert} = 0$	$\Theta(N)$
2. for each $q\in Q$ **Parallelization**	$\times Q$
3. for each $y_{ijk}\in\Lambda_q$	$\times\lvert\Lambda_q\rvert$
4. $err = y_{ijk} - \tilde{y}_{ijk}$	$\Theta(1)$
5. $\mathrm{E\_U}_k += err + \hat{e}_k$	$\Theta(1)$
6. for each $k\in K_q$	$\times\lvert K_q\rvert$
7. $\hat{e}_k = \dfrac{\mathrm{E\_U}_k + \lambda_{(p)}\lvert\Lambda_{(k)}\rvert e_k - z_k + \rho_{(p)}\hat{e}_k^{t}}{\lvert\Lambda_{(k)}\rvert + \lambda_{(p)}\lvert\Lambda_{(k)}\rvert + \rho_{(p)}}$	$\Theta(1)$
9. $e_k = \max(0, \hat{e}_k + \dfrac{z_k}{\lambda_{(p)}\lvert\Lambda_{(k)}\rvert})$	$\Theta(1)$
10. $z_k += \eta_{(p)}\lambda_{(p)}\lvert\Lambda_{(k)}\rvert(\hat{e}_k - e_k)$	$\Theta(1)$

IV. Experimental Results

TABLE S1
P^2SO PARAMETERS SETTINGS

Parameters	Settings
P	5
ω	0.724
c_1	2
c_2	2
r_1	random number ∈ [0, 1]
r_2	random number ∈ [0, 1]
μ	1
$[\breve{\rho},\hat{\rho}]$	[10, 100]
$[\breve{\lambda},\hat{\lambda}]$	[0.1, 1]
$[\breve{\eta},\hat{\eta}]$	[0.1, 1]
$[\breve{\upsilon}_\rho,\hat{\upsilon}_\rho]$	$[-0.2\times(\hat{\rho}-\breve{\rho}),0.2\times(\hat{\rho}-\breve{\rho})]$
$[\breve{\upsilon}_\lambda,\hat{\upsilon}_\lambda]$	$[-0.2\times(\hat{\lambda}-\breve{\lambda}),0.2\times(\hat{\lambda}-\breve{\lambda})]$
$[\breve{\upsilon}_\eta,\hat{\upsilon}_\eta]$	$[-0.2\times(\hat{\eta}-\breve{\eta}),0.2\times(\hat{\eta}-\breve{\eta})]$

TABLE S2
THE OPTIMAL HYPER-PARAMETERS BY MANUAL TUNING

Datasets	Optimal Hyper-parameters		
	ρ	λ	η
D1	90	0.3	0.2
D2	100	0.3	0.1
D3	90	0.9	0.1
D4	100	0.1	0.3
D5	80	1.0	0.1
D6	100	0.1	0.3
D7	80	0.4	0.2
D8	90	0.2	0.3

TABLE S3
HYPER-PARAMETERS SETTING OF M1-8

Datasets	M1	M2	M3	M4	M5	M6	M7	M8
D1	P^2SO	η=5×10^{-2}	η=10^{-5}	λ=10^{-2}	η=5×10^{-3}	α=5×10^{-3}, β_1=0.7	η=10^{-2}	K=3, η=1, λ=0.5
	Q=16	λ=10^{-2}	γ=40	λ_b=10^{-2}	λ=5×10^{-3}	β_2=0.999	λ=10^{-2}	mini-batch size=2048
D2	P^2SO	η=5×10^{-2}	η=10^{-6}	λ=10^{-2}	η=5×10^{-3}	α=5×10^{-3} β_1=0.7	η=10^{-1}	K=3, η=1, λ=1
	Q=16	λ=10^{-2}	γ=80	λ_b=10^{-2}	λ=5×10^{-3}	β_2=0.999	λ=10^{-1}	mini-batch size=2048
D3	P^2SO	η=10^{-2}	η=10^{-6}	λ=0.5	η=5×10^{-3}	α=5×10^{-3}, β_1=0.7	η=10^{-1}	K=3, η=1, λ=10^{-1}
	Q=16	λ=10^{-1}	γ=80	λ_b=10^{-1}	λ=10^{-2}	β_2=0.999	λ=10^{-1}	mini-batch size=2048
D4	P^2SO	η=10^{-2}	η=10^{-5}	λ=0.5	η=5×10^{-3}	α=5×10^{-2}, β_1=0.7	η=10^{-1}	K=3, η=1, λ=1
	Q=16	λ=10^{-1}	γ=20	λ_b=10^{-1}	λ=10^{-2}	β_2=0.999	λ=5×10^{-2}	mini-batch size=2048
D5	P^2SO	η=10^{-2}	η=10^{-5}	λ=0.5	η=5×10^{-3}	α=5×10^{-2}, β_1=0.5	η=10^{-1}	K=3, η=5×10^{-2}, λ=5×10^{-2}
	Q=16	λ=10^{-1}	γ=40	λ_b=10^{-1}	λ=5×10^{-3}	β_2=0.999	λ=5×10^{-2}	mini-batch size=2048
D6	P^2SO	η=10^{-2}	η=10^{-5}	λ=0.5	η=5×10^{-3}	α=5×10^{-2}, β_1=0.5	η=10^{-1}	K=3, η=1, λ=1
	Q=16	λ=0.5	γ=40	λ_b=10^{-1}	λ=5×10^{-3}	β_2=0.999	λ=5×10^{-3}	mini-batch size=2048
D7	P^2SO	η=10^{-2}	η=10^{-5}	λ=10^{-1}	η=10^{-3}	α=5×10^{-3}, β_1=0.7	η=10^{-1}	K=3, η=10^{-1}, λ=10^{-1}
	Q=16	λ=10^{-1}	γ=40	λ_b=5×10^{-2}	λ=10^{-3}	β_2=0.999	λ=5×10^{-3}	mini-batch size=2048
D8	P^2SO	η=5×10^{-3}	η=10^{-5}	λ=0.5	η=10^{-3}	α=5×10^{-3}, β_1=0.9	η=10^{-1}	K=3, η=10^{-1}, λ=10^{-1}
	Q=16	λ=10^{-3}	γ=40	λ_b=0.5	λ=10^{-3}	β_2=0.999	λ=5×10^{-3}	mini-batch size=2048

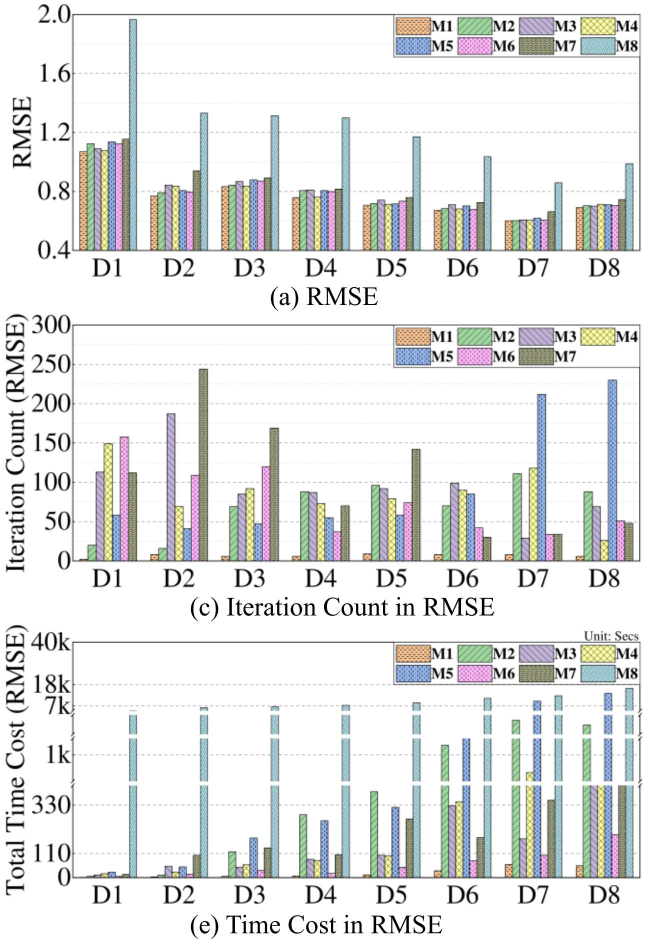
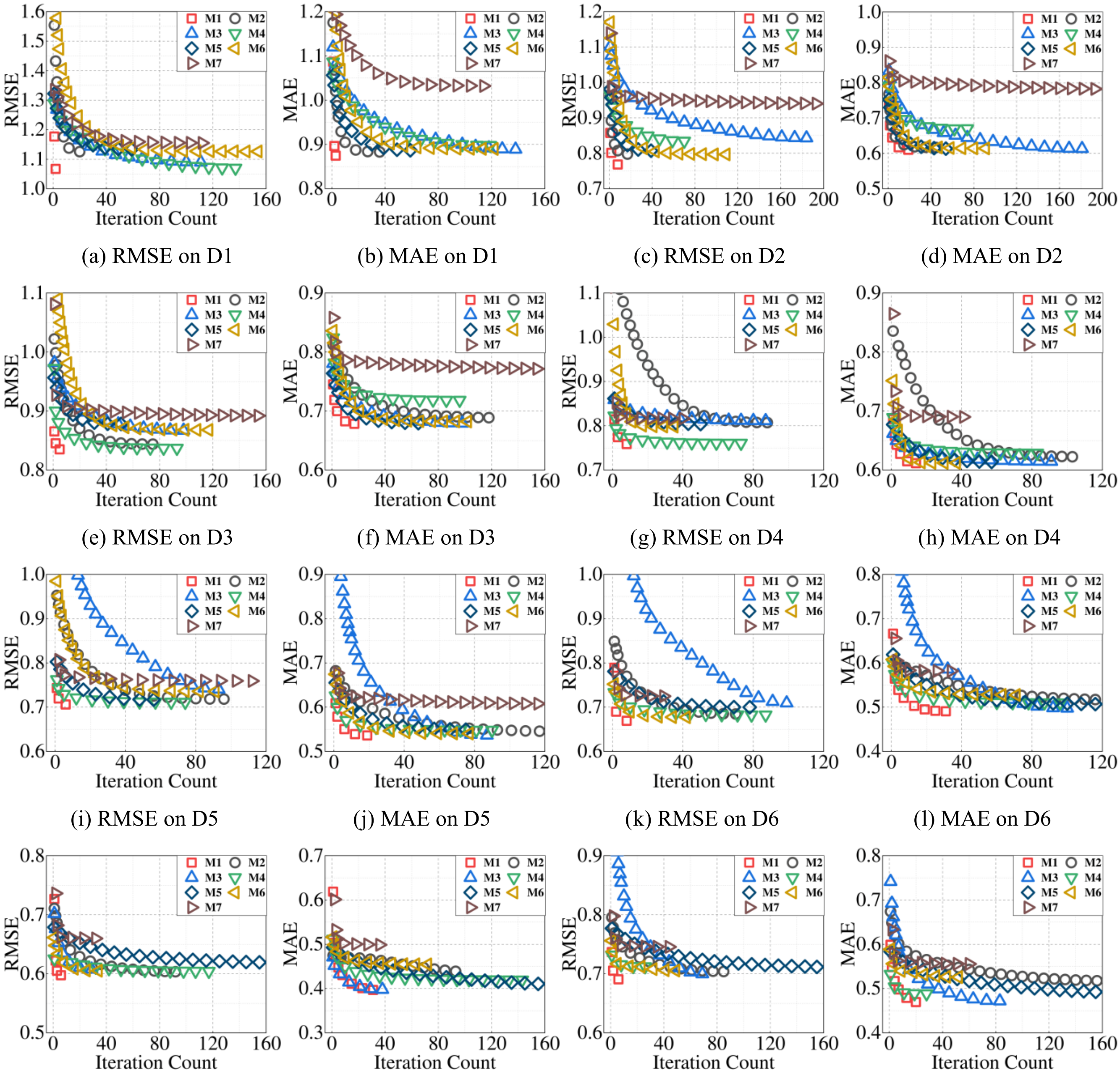


Fig. **S1**. Performance comparison of M1-8 on D1-8.



(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

(m) RMSE on D7 (n) MAE on D7 (o) RMSE on D8 (p) MAE on D8

Fig. S2. Training curves of M1-7 in RMSE and MAE on D1-8.