Parallel Adaptive Stochastic Gradient Descent Algorithms for Latent Factor Analysis of High-Dimensional and Incomplete Industrial Data Supplementary File

Wen Qin, Xin Luo, *Senior Member*, *IEEE*, Shuai Li, *Senior Member*, *IEEE*, and MengChu Zhou, *Fellow*, *IEEE*

I. Introduction

This is the supplementary file for the paper entitled "Parallel Adaptive Stochastic Gradient Descent Algorithms for Latent Factor Analysis of High-Dimensional and Incomplete Industrial Data". We have put the proof of *Lemma* 1-2 and *Theorem* 1-2 in Section II, and some figures and tables of experimental results in Section III.

II. Proofs for Lemma 1-2, and Theorems 1-2

A. Proof of Lemma 1

Considering *Lemma* 1, we have the following inferences:

$$\begin{aligned}
\boldsymbol{x}_u^{\tau+1}+\boldsymbol{w}^{\tau+1} &= \boldsymbol{x}_u^{\tau+1}+\frac{\gamma}{1-\gamma}\left(\boldsymbol{x}_u^{\tau+1}-\boldsymbol{x}_u^{\tau}+\sigma\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right)=\frac{1}{1-\gamma}\boldsymbol{x}_u^{\tau+1}+\frac{\gamma}{1-\gamma}\left(\sigma\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\boldsymbol{x}_u^{\tau}\right)\\
&=\frac{1}{1-\gamma}\left(\boldsymbol{x}_u^{\tau}-\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)+\gamma\left[\left(\boldsymbol{x}_u^{\tau}-\sigma\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right)-\left(\boldsymbol{x}_u^{\tau-1}-\sigma\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau-1}\right)\right)\right]\right)+\frac{\gamma}{1-\gamma}\left(\sigma\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\boldsymbol{x}_u^{\tau}\right)\\
&=\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\frac{\eta}{1-\gamma}\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right).
\end{aligned}\tag{S1}$$

Thus, *Lemma* 1 stands.□

B. Proof of Lemma 2

Considering *Lemma* 2, we have the following inferences:

$$\begin{aligned}
\left\|\boldsymbol{x}_u^{\tau+1}+\boldsymbol{w}^{\tau+1}-\boldsymbol{x}_u\right\|^2 &= \left\|\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\frac{\eta}{1-\gamma}\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\boldsymbol{x}_u\right\|^2\\
&=\left\|\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\boldsymbol{x}_u\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\boldsymbol{x}_u,\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\|^2\\
&=\left\|\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\boldsymbol{x}_u\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{x}_u^{\tau}+\frac{\gamma}{1-\gamma}\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{\tau-1}+\sigma\eta\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau-1}\right)\right)-\boldsymbol{x}_u,\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\|^2\\
&=\left\|\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\boldsymbol{x}_u\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u,\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle-\frac{2\eta\gamma}{\left(1-\gamma\right)^2}\left\langle\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{\tau-1},\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle\\
&\quad-\frac{2\sigma\eta^2\gamma}{1-\gamma}\left\langle\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau-1}\right),\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\|^2.
\end{aligned}\tag{S2}$$

According to the properties of a convex function, the following inequality can be achieved:

$$\begin{aligned}
\left\|\boldsymbol{x}_u^{\tau+1}+\boldsymbol{w}^{\tau+1}-\boldsymbol{x}_u\right\|^2 &= \left\|\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\boldsymbol{x}_u\right\|^2-\frac{2\eta}{1-\gamma}\left\langle\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u,\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle-\frac{2\eta\gamma}{\left(1-\gamma\right)^2}\left\langle\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{\tau-1},\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle\\
&\quad-\frac{2\sigma\eta^2\gamma}{1-\gamma}\left\langle\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau-1}\right),\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\rangle+\left(\frac{\eta}{1-\gamma}\right)^2\left\|\nabla_{x_u}\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\right\|^2\\
&\le\left\|\boldsymbol{x}_u^{\tau}+\boldsymbol{w}^{\tau}-\boldsymbol{x}_u\right\|^2-\frac{2\eta}{1-\gamma}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u\right)\right)-\frac{2\eta\gamma}{\left(1-\gamma\right)^2}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau-1}\right)\right)+\left(\frac{\eta}{1-\gamma}\right)^2\left(2\sigma\gamma+1\right).
\end{aligned}\tag{S3}$$

Note that (S3) can yield the appearance of $\boldsymbol{x}_u^{-1}$ when τ = 0. Following previous research [57], by setting $\boldsymbol{x}_u^{-1}=\boldsymbol{x}_u^{0}$ the above inequality still holds. Hence, *Lemma* 2 stands.□

C.*Proof of Theorem* 1

Proof. We prove the correctness of *Theorem* 1 in the manuscript separately in the following.

When $\tau\in\{0, \ldots, T\}$, the formula (21) is cumulatively summed to get:

$$\frac{2\eta}{1-\gamma}\sum_{\tau=0}^{T}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u\right)\right)\le\frac{2\eta\gamma}{\left(1-\gamma\right)^2}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{0}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u^{T}\right)\right)+\left\|\boldsymbol{x}_u^{0}-\boldsymbol{x}_u\right\|^2+\left(\frac{\eta}{1-\gamma}\right)^2\left(2\sigma\gamma+1\right)\left(T+1\right)G^2. \tag{S4}$$

Since $\boldsymbol{x}_u=\boldsymbol{x}_u^{*}$, and $\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)\ge\varepsilon_{u,i}\left(\boldsymbol{x}_u^{*}\right)$, we get:

$$\sum_{\tau=0}^{T}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{\tau}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u\right)\right)\le\frac{\gamma}{1-\gamma}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{0}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u^{*}\right)\right)+\frac{1-\gamma}{2\eta}\left\|\boldsymbol{x}_u^{0}-\boldsymbol{x}_u^{*}\right\|^2+\frac{\eta}{2\left(1-\gamma\right)}\left(2\sigma\gamma+1\right)\left(T+1\right)G^2. \tag{S5}$$

Let $\bar{\boldsymbol{x}}_u=\sum_{\tau=0}^{T}\boldsymbol{x}_u^{\tau}/\left(T+1\right)$, for a convex function $\varepsilon_{u,i}(x_u)$ have:

$$\varepsilon_{u,i}\left(\bar{\boldsymbol{x}}_u\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u^{*}\right)\le\frac{\gamma}{\left(1-\gamma\right)\left(T+1\right)}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{0}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u^{*}\right)\right)+\frac{1-\gamma}{2\eta\left(T+1\right)}\left\|\boldsymbol{x}_u^{0}-\boldsymbol{x}_u^{*}\right\|^2+\frac{\eta}{2\left(1-\gamma\right)}\left(2\sigma\gamma+1\right)G^2. \tag{S6}$$

The proof of **Theorem 1** can be done by plugging in the value of the learning rate η. According to the same principle, $\varepsilon_{u,i}(\boldsymbol{y}_i)$ converges by training $\boldsymbol{y}_i$ by fixing $\boldsymbol{x}_u$ as a constant. Therefore, *Theorem* 1 stands.□

D.Proof of Theorem 2

Proof. We prove the correctness of *Theorem* 2 in the manuscript separately in the following.

To facilitate subsequent proof, we transform the update rule (11) of A-ASGD into:

$$\begin{aligned}
&\boldsymbol{m}_u^{(0)}=0,\boldsymbol{z}_u^{(0)}=0,\\
&\text{for the } \tau\text{th instance } r_{u,i}\in\Lambda\left(u\right) \text{ as } \tau\in\left[1,2,\ldots,\left|\Lambda\left(u\right)\right|\right]:\\
&\boldsymbol{m}_u^{(\tau)}=\gamma_1\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_2\right)\boldsymbol{h}_u^{\tau},\quad \hat{\boldsymbol{m}}_u^{(\tau)}=\frac{\boldsymbol{m}_u^{(\tau)}}{\left(1-\gamma_1^{\tau}\right)},\\
&\boldsymbol{z}_u^{(\tau)}=\gamma_1\boldsymbol{z}_u^{(\tau-1)}+\left(1-\gamma_2\right)\left(\boldsymbol{h}_u^{\tau}\right)^2,\quad \hat{\boldsymbol{z}}_u^{(\tau)}=\frac{\boldsymbol{z}_u^{(\tau)}}{\left(1-\gamma_2^{\tau}\right)},\\
&\boldsymbol{x}_u^{(\tau)}\leftarrow\boldsymbol{x}_u^{(\tau-1)}-\frac{\eta\,\hat{\boldsymbol{m}}_u^{(\tau)}}{\sqrt{\hat{\boldsymbol{z}}_u^{(\tau)}}},
\end{aligned} \tag{S7}$$

where $\boldsymbol{h}_u^{\tau}=\nabla\varepsilon_{u,i}\left(x_u^{\tau}\right)$. According to the iteration formula of x in the (S7), we have:

$$\begin{aligned}
\boldsymbol{m}_u^{\tau}&=\gamma_1\boldsymbol{m}_u^{\tau-1}+\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau}=\gamma_1^{2}\boldsymbol{m}_u^{\tau-2}+\gamma_1\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau-1}+\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau}\\
&=\gamma_1^{\tau}\boldsymbol{m}_u^{0}+\gamma_1^{\tau-1}\left(1-\gamma_1\right)\boldsymbol{h}_u^{1}+\ldots+\gamma_1^{2}\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau-2}+\gamma_1\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau-1}+\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau}\\
&\overset{\left(\boldsymbol{m}^{(0)}=0\right)}{=}\gamma_1^{\tau-1}\left(1-\gamma_1\right)\boldsymbol{h}_u^{1}+\ldots+\gamma_1^{2}\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau-2}+\gamma_1\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau-1}+\left(1-\gamma_1\right)\boldsymbol{h}_u^{\tau}\\
&=\sum_{p=1}^{\tau}\left(1-\gamma_1\right)\gamma_1^{\tau-p}\boldsymbol{h}_u^{p}=\left(1-\gamma_1\right)\sum_{p=1}^{\tau}\gamma_1^{\tau-p}\boldsymbol{h}_u^{p}.
\end{aligned} \tag{S8}$$

Note that $\boldsymbol{z}_u^{\tau}=\sum_{p=1}^{\tau}\left(1-\gamma_2\right)\gamma_2^{\tau-p}\left(\boldsymbol{h}_u^{p}\right)^2=\left(1-\gamma_2\right)\sum_{p=1}^{\tau}\gamma_2^{\tau-p}\left(\boldsymbol{h}_u^{p}\right)^2$. So we observe $\mathrm{E}\left[\boldsymbol{m}_u^{\tau}\right],\mathrm{E}\left[\boldsymbol{z}_u^{\tau}\right]$:

$$\mathrm{E}\left[\boldsymbol{m}_u^{\tau}\right]=\left(1-\gamma_1\right)\sum_{p=1}^{\tau}\gamma_1^{\tau-p}\mathrm{E}\left[\boldsymbol{h}_u^{p}\right]=\mathrm{E}\left[\boldsymbol{h}_u^{p}\right]\left(1-\gamma_1^{\tau}\right),\quad \mathrm{E}\left[\boldsymbol{z}_u^{\tau}\right]=\left(1-\gamma_2\right)\sum_{p=1}^{\tau}\gamma_2^{\tau-p}\mathrm{E}\left[\left(\boldsymbol{h}_u^{p}\right)^2\right]=\mathrm{E}\left[\left(\boldsymbol{h}_u^{p}\right)^2\right]\left(1-\gamma_2^{\tau}\right). \tag{S9}$$

To make $\mathrm{E}\left[\boldsymbol{m}_u^{\tau}\right]=\mathrm{E}\left[\boldsymbol{h}_u^{p}\right]$, $\mathrm{E}\left[\boldsymbol{z}_u^{\tau}\right]=\mathrm{E}\left[\left(\boldsymbol{h}_u^{p}\right)^2\right]$, we have $\boldsymbol{m}_u^{\tau}\to\hat{\boldsymbol{m}}_u^{\tau}=\frac{\boldsymbol{m}_u^{\tau}}{\left(1-\gamma_1^{\tau}\right)}$, $\boldsymbol{z}_u^{\tau}\to\hat{\boldsymbol{z}}_u^{\tau}=\frac{\boldsymbol{z}_u^{\tau}}{\left(1-\gamma_2^{\tau}\right)}$. The weight assigned by A-ASGD decreases as τ increases. By making $\forall p$: $\left\|\left(\boldsymbol{h}_u^{p}\right)^2\right\|_2\le I$ as I be a positive consant, then we ahve $\left\|\boldsymbol{z}_u^{\tau}\right\|_2=\sum_{p=1}^{\tau}\left(1-\gamma_2\right)\gamma_2^{\tau-p}\left\|\left(\boldsymbol{h}_u^{p}\right)^2\right\|_2\le I$.

Since $\varepsilon_{u,i}(\boldsymbol{x}_u)$ is convex when y_i is fixed [55], we have $R(T)=\sum_{\tau=1}^{T}\left(\varepsilon_{u,i}\left(\boldsymbol{x}_u^{(\tau)}\right)-\varepsilon_{u,i}\left(\boldsymbol{x}_u^{*}\right)\right)\le\sum_{\tau=1}^{T}\boldsymbol{h}_u^{\tau}\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)$. According to the update scheme of A-ASGD, we have:

$$\boldsymbol{x}_u^{(\tau+1)}=\boldsymbol{x}_u^{(\tau)}-\eta\frac{\widehat{\boldsymbol{m}}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}=\boldsymbol{x}_u^{(\tau)}-\eta\frac{1}{\left(1-\gamma_1^{\tau}\right)}\frac{\boldsymbol{m}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}=\boldsymbol{x}_u^{(\tau)}-\eta\frac{1}{\left(1-\gamma_1^{\tau}\right)}\frac{\gamma_1\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_1\right)\boldsymbol{h}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}. \tag{S10}$$

Let $\gamma_1=\gamma_{1,\tau}$, then γ_1 varies with itearions and will not increase monotonically, i.e., we have $\gamma_{1,1}\geqslant\gamma_{1,2}\geqslant\ldots\geqslant\gamma_{1,\tau}\ldots$. Hene, the momentum will gradually disappear. By making $\boldsymbol{m}_u\rightarrow\boldsymbol{h}_u$ [59], we have:

$$\boldsymbol{x}_u^{(\tau+1)}=\boldsymbol{x}_u^{(\tau)}-\eta\frac{1}{1-\prod_{p=1}^{\tau}\gamma_{1,p}}\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau+1)}}}. \tag{S11}$$

Let $\alpha_\tau=\eta\frac{1}{1-\prod_{p=1}^{\tau}\gamma_{1,p}}$, then $\boldsymbol{x}_u^{(\tau+1)}=\boldsymbol{x}_u^{(\tau)}-\alpha_\tau\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}$. So we separate $\boldsymbol{h}_u^{\tau}\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)$ as:

$$\begin{aligned}&\boldsymbol{x}_u^{(\tau+1)}=\boldsymbol{x}_u^{(\tau)}-\alpha_\tau\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}\\&\Rightarrow\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2=\left[\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)-\alpha_\tau\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}\right]^2\\&\Rightarrow2\alpha_\tau\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}}{\sqrt{\widehat{\boldsymbol{z}}_u^{(t)}}}\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)=\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2+\alpha_\tau^2\frac{\left[\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}\right]^2}{\widehat{\boldsymbol{z}}_u^{(\tau)}}.\end{aligned} \tag{S12}$$

Since $\boldsymbol{m}_u^{(\tau)}=\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}$,

$$\boldsymbol{h}_u^{\tau}\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)=\underbrace{\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}\left[\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\alpha_\tau\left(1-\gamma_{1,\tau}\right)}}_{(1)}-\underbrace{\frac{\gamma_{1,t}\boldsymbol{m}_u^{(\tau-1)}}{\left(1-\gamma_{1,\tau}\right)}\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)}_{(2)}+\underbrace{\frac{\alpha_\tau}{2\left(1-\gamma_{1,\tau}\right)}\frac{\left(\boldsymbol{m}_u^{\tau}\right)^2}{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}}}_{(3)}. \tag{S13}$$

Summing and scaling each item of the upper bound of $R(T)$, we have:

$$\begin{aligned}\sum_{\tau=1}^{\boldsymbol{T}}\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}\left[\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\alpha_\tau\left(1-\gamma_{1,\tau}\right)}&=\sum_{\tau=1}^{\boldsymbol{T}}\frac{\sqrt{\widehat{\boldsymbol{z}}_{u,i}^{\tau}}\left[\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\eta\frac{1}{1-\prod_{p=1}^{\tau}\beta_{1,p}}\left(1-\gamma_{1,\tau}\right)}\\&=\sum_{\tau=1}^{\boldsymbol{T}}\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}\left[\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]\left(1-\prod_{p=1}^{\tau}\gamma_{1,p}\right)}{2\eta\left(1-\gamma_{1,\tau}\right)}\\&\le\sum_{\tau=1}^{\boldsymbol{T}}\frac{\sqrt{\widehat{\boldsymbol{z}}_{\boldsymbol{u}}^{\tau}}\left[\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\eta\left(1-\gamma_{1,1}\right)}.\end{aligned} \tag{S14}$$

And the dislocation recombination summation is given as:

$$\begin{aligned}\sum_{\tau=1}^{T}\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}\left[\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\alpha_\tau\left(1-\gamma_{1,\tau}\right)}&\le\sum_{\tau=1}^{T}\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2}{2\eta\left(1-\gamma_{1,1}\right)}\\&=\frac{\sqrt{\widehat{\boldsymbol{z}}_{\boldsymbol{u}}^{1}}\left(\boldsymbol{x}_u^{1}-\boldsymbol{x}_u^{*}\right)^2}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{T}}\left(\boldsymbol{x}_u^{(T+1)}-\boldsymbol{x}_u^{*}\right)^2}{2\eta\left(1-\gamma_{1,1}\right)}+\sum_{\tau=2}^{T}\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2\cdot\left[\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{\tau}}}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau-1)}}}{2\eta\left(1-\gamma_{1,1}\right)}\right].\end{aligned} \tag{S15}$$

Note that the ahove inferences adopt the condition of $\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}{\eta}\ge\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau-1)}}}{\eta}$, which holds according to (8). Thus, we further have:

$$\sum_{\tau=2}^{T}\left(\boldsymbol{x}_u^{\tau}-\boldsymbol{x}_u^{*}\right)^2\cdot\left[\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau-1)}}}{2\eta\left(1-\gamma_{1,1}\right)}\right]\leq\sum_{\tau=1}^{T}D^2\cdot\left[\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau-1)}}}{2\eta\left(1-\gamma_{1,1}\right)}\right]=D^2\cdot\left[\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(T)}}}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(1)}}}{2\eta\left(1-\gamma_{1,1}\right)}\right], \tag{S16}$$

where the final scaling is organized as：

$$\sum_{\tau=1}^{T}\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}\left[\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\alpha_\tau\left(1-\gamma_{1,\tau}\right)}\leq\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(1)}}D^2}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(T)}}\left(\boldsymbol{x}_{u,i}^{(T+1)}-\boldsymbol{x}_{u,i}^{*}\right)^2}{2\eta\left(1-\gamma_{1,1}\right)}+\left[\frac{D^2\sqrt{\widehat{\boldsymbol{z}}_u^{(\boldsymbol{T})}}}{2\eta\left(1-\gamma_{1,1}\right)}-\frac{D^2\sqrt{\widehat{\boldsymbol{z}}_u^{(1)}}}{2\eta\left(1-\gamma_{1,1}\right)}\right]\leq\frac{D^2\sqrt{\widehat{\boldsymbol{z}}_u^{(T)}}}{2\eta\left(1-\gamma_{1,1}\right)}. \tag{S17}$$

By scaling $\widehat{z}_u^{(T)}$ with the previously mentioned expected approximation and bounded gradient, we finally scaling the term of (S13)-(1) as:

$$\sum_{\tau=1}^{T}\frac{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}\left[\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)^2-\left(\boldsymbol{x}_u^{(\tau+1)}-\boldsymbol{x}_u^{*}\right)^2\right]}{2\alpha_\tau\left(1-\gamma_{1,\tau}\right)}\leq\frac{D^2\sqrt{\widehat{\boldsymbol{z}}_u^{(T)}}}{2\eta\left(1-\gamma_{1,1}\right)}\leq\frac{D^2G}{2\eta\left(1-\gamma_{1,1}\right)}. \tag{S18}$$

Considering the term of (S13)-(2), we have:

$$\sum_{\tau=1}^{T}-\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}}{\left(1-\gamma_{1,\tau}\right)}\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)=\sum_{\tau=1}^{T}\frac{\gamma_{1,\tau}}{\left(1-\gamma_{1,\tau}\right)}\boldsymbol{m}_u^{(\tau-1)}\left[-\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)\right]\leq\sum_{\tau=1}^{T}\frac{\gamma_{1,\tau}}{\left(1-\gamma_{1,\tau}\right)}\left|\boldsymbol{m}_u^{(\tau-1)}\right|D^2. \tag{S19}$$

According to the following inference,

$$\begin{aligned}\boldsymbol{m}_u^{(\tau)}&=\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}=\gamma_{1,\tau}\gamma_{1,\tau-1}\boldsymbol{m}_u^{(\tau-2)}+\gamma_{1,\tau}\left(1-\gamma_{1,\tau-1}\right)\boldsymbol{h}_u^{(\tau-1)}+\left(1-\gamma_{1,\tau}\right)\boldsymbol{h}_u^{(\tau)}\\&=\sum_{p=1}^{\tau}\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)\boldsymbol{h}_u^{(p)},\\\left|\boldsymbol{m}_u^{(\tau)}\right|&\leq\sum_{p=1}^{\tau}\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)\left|\boldsymbol{h}_u^{(p)}\right|\leq\sum_{p=1}^{\tau}\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)G^2\leq G^2.\end{aligned} \tag{S20}$$

we formulate the term (S13)-(2) as:

$$\sum_{\tau=1}^{T}-\frac{\gamma_{1,\tau}\boldsymbol{m}_u^{(\tau-1)}}{\left(1-\gamma_{1,\tau}\right)}\left(\boldsymbol{x}_u^{(\tau)}-\boldsymbol{x}_u^{*}\right)\leq\sum_{\tau=1}^{T}\frac{\gamma_{1,\tau}}{\left(1-\gamma_{1,\tau}\right)}G^2D^2=G^2D^2\sum_{\tau=1}^{T}\frac{\gamma_{1,\tau}}{\left(1-\gamma_{1,\tau}\right)}. \tag{S21}$$

Regarding (S13)-(3), we focus on the following formula:

$$\frac{\left(\boldsymbol{m}_u^{(\tau)}\right)^2}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}=\left(1-\gamma_2^{\tau}\right)\frac{\left(\boldsymbol{m}_u^{(\tau)}\right)^2}{\sqrt{\boldsymbol{z}_u^{(\tau)}}}\leq\frac{\left(\boldsymbol{m}_u^{(\tau)}\right)^2}{\sqrt{\boldsymbol{z}_u^{(\tau)}}}. \tag{S22}$$

Then considering the closed solution formula of numerator and denominator, we have

$$\begin{aligned}&\boldsymbol{m}_u^{(\tau)}=\sum_{p=1}^{\tau}\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)h_u^{(p)},\ \boldsymbol{z}_u^{(\tau)}=\left(1-\gamma_2\right)\sum_{p=1}^{\tau}\gamma_2^{\tau-p}\left(h_u^{(p)}\right)^2,\\&\left(\boldsymbol{m}_u^{(\tau)}\right)^2=\left(\sum_{p=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)}{\sqrt{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}}\sqrt{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}\boldsymbol{h}_u^{(p)}\right)^2.\end{aligned} \tag{S23}$$

With the Cauchy's inequality, (S23) yields:

$$\begin{aligned}\left(\boldsymbol{m}_u^{(\tau)}\right)^2&=\left(\sum_{p=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)}{\sqrt{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}}\sqrt{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}\boldsymbol{h}_u^{(p)}\right)^2\leq\sum_{p=1}^{t}\left(\frac{\left(1-\gamma_{1,p}\right)\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)}{\sqrt{\left(1-\gamma_2\right)\gamma_{2,s}^{\tau-p}}}\right)^2\sum_{p=1}^{\tau}\left(\sqrt{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}\boldsymbol{h}_u^{(p)}\right)^2\\&=\sum_{p=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)^2\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)^2}{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}\underbrace{\sum_{p=1}^{\tau}\left(\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}\left(\boldsymbol{h}_u^{(p)}\right)^2\right)}_{z_u^{(\tau)}}.\end{aligned} \tag{S24}$$

Then the term of (S13)-(3) is reformulated as:

$$\sum_{\tau=1}^{T}\frac{\alpha_\tau}{2\left(1-\gamma_{1,\tau}\right)}\frac{\left(\boldsymbol{m}_u^{(\tau)}\right)^2}{\sqrt{\widehat{\boldsymbol{z}}_u^{(\tau)}}}\le\sum_{t=1}^{T}\frac{\alpha_\tau}{2\left(1-\beta_{1,t}\right)}\sum_{p=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)^2\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)^2}{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}\frac{\sum_{p=1}^{\tau}\left(\left(1-\gamma_2\right)\gamma_{2,p}^{t-p}\left(\boldsymbol{h}_u^{(p)}\right)^2\right)}{\sqrt{\sum_{p=1}^{\tau}\left(\left(1-\gamma_2\right)\gamma_{2,p}^{t-p}\left(\boldsymbol{h}_u^{(p)}\right)^2\right)}}$$

$$=\sum_{\tau=1}^{T}\frac{\alpha_\tau}{2\left(1-\gamma_{1,\tau}\right)}\sum_{s=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)^2\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)^2}{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}\sqrt{\boldsymbol{z}_i^{(t)}}\le\sum_{\tau=1}^{T}\frac{\alpha_\tau}{2\left(1-\gamma_{1,t}\right)}\sum_{p=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)^2\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)^2}{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}G^2. \tag{S25}$$

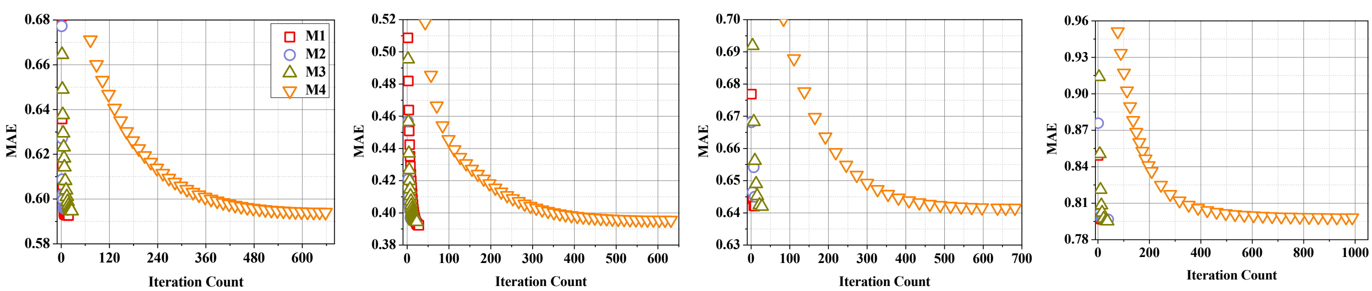
By integrating the scaling of (S13)-(1) and (S13)-(3), we have:

$$R\left(T\right)\le\frac{D^2G}{2\eta\left(1-\gamma_{1,1}\right)}+GD\sum_{\tau=1}^{T}\frac{\gamma_{1,\tau}}{\left(1-\gamma_{1,\tau}\right)}+G\sum_{\tau=1}^{T}\frac{\alpha_\tau}{2\left(1-\gamma_{1,\tau}\right)}\sum_{p=1}^{\tau}\frac{\left(1-\gamma_{1,p}\right)^2\left(\prod_{k=p+1}^{\tau}\gamma_{1,k}\right)^2}{\left(1-\gamma_2\right)\gamma_{2,p}^{\tau-p}}. \tag{S26}$$

Hence, where $T\to\infty$, and the average value $R(T)/T\to 0$ of $R(T)$, the algorithm converges. According to the same principle, $\varepsilon_{u,i}(\boldsymbol{y}_i)$ converges by training $\boldsymbol{y}_i$ by fixing $\boldsymbol{x}_u$ as a constant. Therefore, *Thereom* 2 stands.□

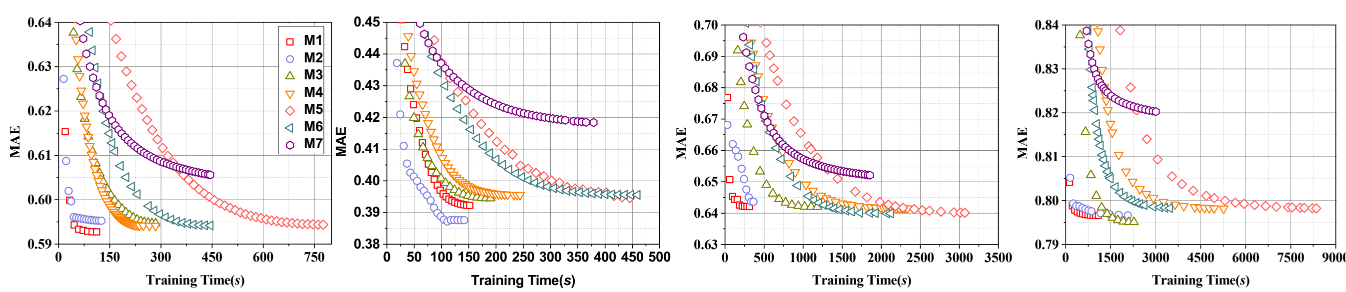
III. Some Figures and Tables of Experimental Results

Some experimental results are shown in this section.



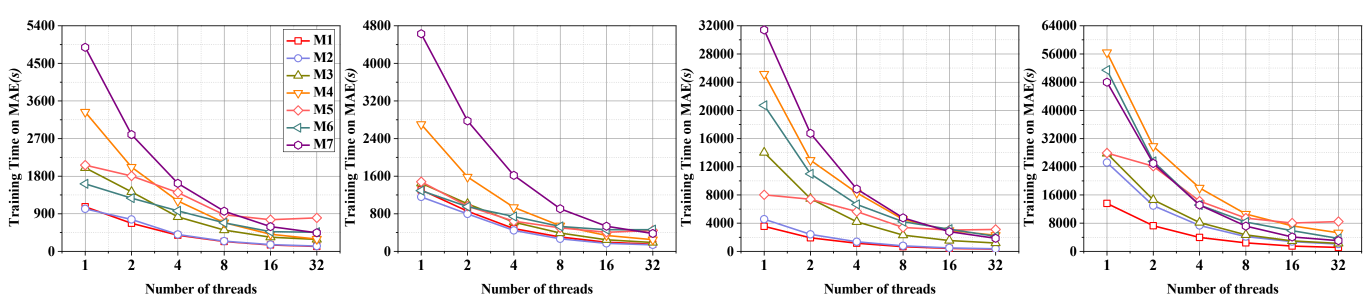
(a) D1 (b) D2 (c) D3 (d) D4

Fig. S.1. M1-4's training curves with iteration count in MAE. All panels share the legend in panel (a).



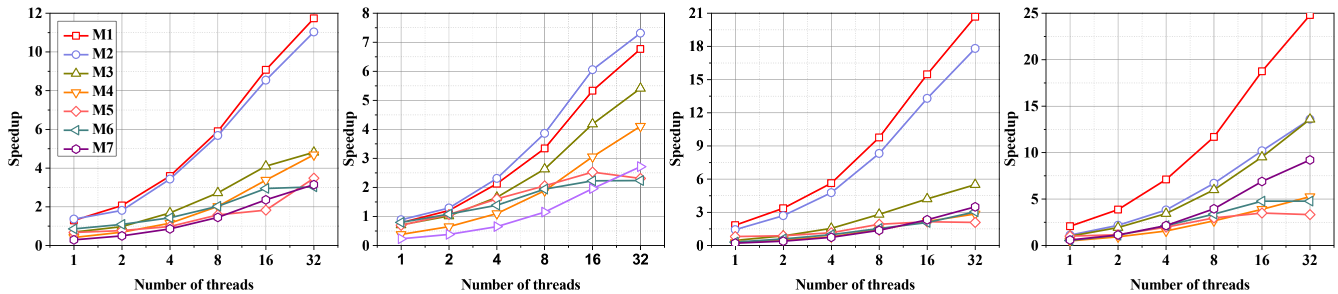
(a) D1 (b) D2 (c) D3 (d) D4

Fig. S.2. M1-7's training curves with time in MAE. All panels share the legend in panel (a).



(a) D1 (b) D2 (c) D3 (d) D4

Fig. S.3. M1-7's training time in MAE as thread count increases. Both panels share the legend in panel (a).



(a) D1 (b) D2 (c) D3 (d) D4

Fig.S.4. M1-7's speedup with MAE as thread count increases. Both panels share the legend in panel (a).

TABLE S.I. M1-4's converging iteration count in MAE.

No.	Iteration count			
	D1	**D2**	**D3**	**D4**
M1	**18**$_{\pm3.2}$	**28**$_{\pm4.8}$	**10**$_{\pm3}$	**26**$_{\pm1.5}$
M2	**17**$_{\pm1.2}$	**18**$_{\pm3.5}$	**12**$_{\pm1.2}$	**41**$_{\pm7}$
M3	**25**$_{\pm1.7}$	**23**$_{\pm1.5}$	**29**$_{\pm5.2}$	**38**$_{\pm4.1}$
M4	659$_{\pm7.9}$	634$_{\pm12}$	669$_{\pm1.4}$	990$_{\pm3.7}$

TABLE S.II. Comparison results on Speedup with 32 threads, including win/loss counts and Friedman test (●, ○ and ✪ indicates M1-3 has a higher computational efficiency than its peers, respectively).

NO.	**CASE**	**M1●**	**M2○**	**M3✪**	**M4**	**M5**	**M6**	**M7**
D1	**RMSE-Speedup**	**11.54**	11.53	5.34	4.62 ●○✪	1.64 ●○✪	3.04 ●○✪	2.88 ●○✪
	MAE-Speedup	**11.73**	11.03	4.82	4.69 ●○✪	3.48 ●○✪	3.03 ●○✪	1.56 ●○✪
D2	**RMSE-Speedup**	5.06	**5.20**	3.30	3.14 ●○✪	1.64 ●○✪	1.78 ●○✪	1.46 ●○✪
	MAE-Speedup	6.77	**7.32**	5.41	4.11 ●○✪	2.31 ●○✪	2.24 ●○✪	2.71 ●○✪
D3	**RMSE-Speedup**	**12.48**	9.50	5.42	2.85 ●○✪	1.77 ●○✪	2.98 ●○✪	2.88 ●○✪
	MAE-Speedup	**20.68**	17.81	5.50	2.85 ●○✪	2.09 ●○✪	3.05 ●○✪	3.50
D4	**RMSE-Speedup**	**24.81**	16.43	10.14	5.39 ●○✪	3.38 ●○✪	7.90 ●○✪	8.14 ●○✪
	MAE-Speedup	**24.79**	13.58	13.58	5.26 ●○✪	3.33 ●○✪	4.78 ●○✪	9.20 ●○✪
	●Win/Loss	--	--	--	8/0	8/0	8/0	8/0
	○Win/Loss	--	--	--	8/0	8/0	8/0	8/0
	✪Win/Loss	--	--	--	8/0	8/0	8/0	8/0
Friedman Rank		**1.22**	1.83	2.94	4.78	6.56	5.44	5.22

* A lower F-rank value indicates a higher computational efficiency

TABLE S.III. Results of the Wilcoxon signed-ranks test on Speedup with the different thread of Figs. 7 and S.4.

Vs.	M4			M5			M6			M7		
	R+	R-	*p*-value*	R+	R-	*p*-value*	R+	R-	*p*-value*	R+	R-	*p*-value*
M1	**1176**	**0**	**8.42E-10**	**1162**	**14**	**2.03E-09**	**1173**	**3**	**1.02E-09**	**1176**	**0**	**8.42E-10**
M2	**1176**	**0**	**8.42E-10**	**1156**	**0**	**2.93E-09**	**1174**	**2**	**9.55E-10**	**1176**	**0**	**8.42E-10**
M3	**1176**	**10**	**8.42E-10**	**1069**	**107**	**4.15E-07**	**1117**	**59**	**2.97E-08**	**1176**	**0**	**8.42E-10**

* The accepted hypotheses with a significance level of 0.1 are highlighted.