A Decoupled Stochastic Optimization Approach to Efficient Nonstandard Tensor Representation Learning Supplementary File

Qicong Hu, Hong Jia, *Member, IEEE*, Hao Wu, *Member, IEEE*, Xin Luo, *Fellow, IEEE*

This file provides a detailed convergence analysis of DecSGD, other experimental settings (including data preprocessing steps and hyperparameter settings), and additional experimental results (including numerical results and significance analysis).

I. Convergence Proof

A. Strong Convexity of Mode-Specific Subproblems (Lemma 1)

The goal is to show that each mode-specific subproblem is strongly convex, which will later help establish the convergence of the DecSGD algorithm.

Consider the mode-1 sub-objective $\mathcal{L}^{(1)}$, where the factor matrix $\mathbf{A}$ is optimized while $\mathbf{B}$ and $\mathbf{C}$ are held fixed. The loss for a sampled observed entry $(i,j,k) \in \Omega$ is defined as

$$\ell_{ijk}^{(1)}(\mathbf{a}_i) = \left(x_{ijk} - \mathbf{a}_i(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k)^\top\right)^2 + \lambda\|\mathbf{a}_i\|^2. \tag{1}$$

This is a quadratic loss function with respect to $\mathbf{a}_i$. For two factor vectors $\mathbf{a}_{i1}$ and $\mathbf{a}_{i2}$, we apply the second-order Taylor expansion to approximate the difference between $\ell_{ijk}^{(1)}(\mathbf{a}_{i1})$ and $\ell_{ijk}^{(1)}(\mathbf{a}_{i2})$:

$$\ell_{ijk}^{(1)}(\mathbf{a}_{i1}) - \ell_{ijk}^{(1)}(\mathbf{a}_{i2}) = \nabla\ell_{ijk}^{(1)}(\mathbf{a}_{i2})(\mathbf{a}_{i1} - \mathbf{a}_{i2})^\top + \frac{1}{2}(\mathbf{a}_{i1} - \mathbf{a}_{i2})\nabla^2\ell_{ijk}^{(1)}(\mathbf{a}_{i2})(\mathbf{a}_{i1} - \mathbf{a}_{i2})^\top. \tag{2}$$

If $\ell_{ijk}^{(1)}$ is strongly convex, its second-order Taylor expansion guarantees that

$$\ell_{ijk}^{(1)}(\mathbf{a}_{i1}) - \ell_{ijk}^{(1)}(\mathbf{a}_{i2}) \geq \nabla\ell_{ijk}^{(1)}(\mathbf{a}_{i2})(\mathbf{a}_{i1} - \mathbf{a}_{i2})^\top + \frac{\mu}{2}\|\mathbf{a}_{i1} - \mathbf{a}_{i2}\|^2. \tag{3}$$

Substituting (3) into (2) yields

$$(\mathbf{a}_{i1} - \mathbf{a}_{i2})\nabla^2\ell_{ijk}^{(1)}(\mathbf{a}_{i2})(\mathbf{a}_{i1} - \mathbf{a}_{i2})^\top \geq \mu\|\mathbf{a}_{i1} - \mathbf{a}_{i2}\|^2.$$

The Hessian matrix $\nabla^2\ell_{ijk}^{(1)}(\mathbf{a}_i)$ is

$$\nabla^2\ell_{ijk}^{(1)}(\mathbf{a}_i) = (\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k)^\top(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k) + \lambda I_R, \tag{4}$$

where I_R is the $R \times R$ identity matrix. Thus,

$$(\mathbf{a}_{i1} - \mathbf{a}_{i2})\left((\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k)^\top(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k) + \lambda I_R - \mu I_R\right)(\mathbf{a}_{i1} - \mathbf{a}_{i2})^\top \geq 0. \tag{5}$$

This inequality holds for all $\mathbf{a}_{i1}, \mathbf{a}_{i2} \in \mathbb{R}^R$ if μ is the smallest eigenvalue of the Hessian matrix $\nabla^2\ell_{ijk}^{(1)}$, bounded below by $\lambda > 0$. Similar arguments apply to the other mode-specific subproblems, $\mathcal{L}_B$ and $\mathcal{L}_C$, ensuring that all subproblems are strongly convex.

B. Lipschitz Continuity of Gradients (Lemma 2)

Next, we analyze the Lipschitz continuity of the gradients of the per-entry loss functions. For two factor vectors $\mathbf{a}_{i1}$ and $\mathbf{a}_{i2}$, the gradient difference is given by

$$\nabla\ell_{ijk}^{(1)}(\mathbf{a}_{i1}) - \nabla\ell_{ijk}^{(1)}(\mathbf{a}_{i2}) = (\mathbf{a}_{i1} - \mathbf{a}_{i2})\left((\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k)^\top(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k) + \lambda I_R\right). \tag{6}$$

Taking the norm of both yields

$$\|\nabla\ell_{ijk}^{(1)}(\mathbf{a}_{i1}) - \nabla\ell_{ijk}^{(1)}(\mathbf{a}_{i2})\|^2 \leq \|\mathbf{a}_{i1} - \mathbf{a}_{i2}\|^2 \left\|(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k)^\top(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k) + \lambda I_R\right\|^2. \tag{7}$$

Thus, the gradients are Lipschitz continuous with constant $L = \left\|(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k)^\top(\tilde{\mathbf{b}}_j \odot \tilde{\mathbf{c}}_k) + \lambda I_R\right\|^2$.

Q. Hu, H. Wu and X. Luo are with the College of Computer and Information Science, Southwest University, Chongqing 400715, China (E-mail: a1634688370@email.swu.edu.cn, haowuf@swu.edu.cn and luoxin@swu.edu.cn).

H. Jia is with the School of Computer Science, University of Auckland, Auckland 1023, New Zealand (E-mail: hong.jia@auckland.ac.nz)

C. Convergence of DecSGD (Theorem 1)

Consider the following update step for $\mathbf{a}_i$:

$$\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2 = \|\mathbf{a}_i^{(t-1)} - \eta \nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)}) - \mathbf{a}_i^*\|^2. \tag{8}$$

Expanding (8) yields

$$\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2 = \|\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*\|^2 - 2\eta \nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})(\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*)^\top + \eta^2 \|\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})\|^2. \tag{9}$$

Taking expectations over both sides:

$$\mathbb{E}[\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2] = \mathbb{E}[\|\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*\|^2] - 2\eta \mathbb{E}[\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})(\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*)^\top] + \eta^2 \mathbb{E}[\|\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})\|^2]. \tag{10}$$

From strong convexity (Lemma 1), we have:

$$\begin{cases} \ell_{ijk}^{(1)}(\mathbf{a}_i^*) \geq \ell_{ijk}^{(1)}(\mathbf{a}_i^{t-1}) + \nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{t-1})(\mathbf{a}_i^* - \mathbf{a}_i^{t-1})^\top + \frac{\mu}{2}\|\mathbf{a}_i^{t-1} - \mathbf{a}_i^*\|^2; \\ \ell_{ijk}^{(1)}(\mathbf{a}_i^{t-1}) \geq \ell_{ijk}^{(1)}(\mathbf{a}_i^*) + \nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^*)(\mathbf{a}_i^{t-1} - \mathbf{a}_i^*)^\top + \frac{\mu}{2}\|\mathbf{a}_i^{t-1} - \mathbf{a}_i^*\|^2. \end{cases} \tag{11}$$

Combining (11) obtains

$$[\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{t-1}) - \nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^*)](\mathbf{a}_i^{t-1} - \mathbf{a}_i^*)^\top \geq \mu \|\mathbf{a}_i^{t-1} - \mathbf{a}_i^*\|^2. \tag{12}$$

Since $\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^*) = 0$, we get

$$\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{t-1})(\mathbf{a}_i^{t-1} - \mathbf{a}_i^*)^\top \geq \mu \|\mathbf{a}_i^{t-1} - \mathbf{a}_i^*\|^2. \tag{13}$$

Substituting this into expectation (10) yields

$$\mathbb{E}[\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2] \leq \mathbb{E}[\|\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*\|^2] - 2\eta\mu \mathbb{E}[\|\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*\|^2] + \eta^2 \mathbb{E}[\|\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})\|^2]. \tag{14}$$

Assuming bounded gradients, i.e., $\|\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})\|^2 \leq G^2$:

$$\mathbb{E}[\|\nabla \ell_{ijk}^{(1)}(\mathbf{a}_i^{(t-1)})\|^2] \leq G^2. \tag{15}$$

Thus, we get

$$\mathbb{E}[\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2] \leq (1 - 2\eta\mu)\mathbb{E}[\|\mathbf{a}_i^{(t-1)} - \mathbf{a}_i^*\|^2] + \eta^2 G^2. \tag{16}$$

By induction, we can get

$$\mathbb{E}[\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2] \leq (1 - 2\eta\mu)^{t-1}\mathbb{E}[\|\mathbf{a}_i^1 - \mathbf{a}_i^*\|^2] + \sum_{\tau=0}^{t-2}(1 - 2\eta\mu)^\tau (\eta G)^2. \tag{17}$$

For $0 < \eta < 1/(2\mu)$,

$$\sum_{\tau=0}^{t-2}(1 - 2\eta\mu)^\tau < \sum_{\tau=0}^{\infty}(1 - 2\eta\mu)^\tau = \frac{1}{2\eta\mu}. \tag{18}$$

Substituting (18) into (17), we get

$$\mathbb{E}[\|\mathbf{a}_i^{(t)} - \mathbf{a}_i^*\|^2] \leq (1 - 2\eta\mu)^{t-1}\mathbb{E}[\|\mathbf{a}_i^1 - \mathbf{a}_i^*\|^2] + \frac{\eta G^2}{2\mu}. \tag{19}$$

From the quadratic upper bound propert of L-smoothness, we have

$$\ell_{ijk}^{(1)}(\mathbf{a}_i^t) - \ell_{ijk}^{(1)}(\mathbf{a}_i^*) \leq \frac{L}{2}\|\mathbf{a}_i^t - \mathbf{a}_i^*\|^2. \tag{20}$$

Based on (19) and (20), we get

$$\mathbb{E}[\ell_{ijk}^{(1)}(\mathbf{a}_i^t) - \ell_{ijk}^{(1)}(\mathbf{a}_i^*)] \leq \frac{L}{2}\left[(1 - 2\eta\mu)^{t-1}\mathbb{E}[\|\mathbf{a}_i^1 - \mathbf{a}_i^*\|^2] + \frac{\eta G^2}{2\mu}\right]. \tag{21}$$

As $t \to \infty$,

$$\mathbb{E}[\ell_{ijk}^{(1)}(\mathbf{a}_i^t) - \ell_{ijk}^{(1)}(\mathbf{a}_i^*)] \leq \frac{L\eta G^2}{4\mu}. \tag{22}$$

Let $\Omega_i^{(1)}$ represent the set of known entries related to slice $\mathcal{X}(i,:,:)$, then (22) can be reformulated as:

$$\mathbb{E}\left[\sum_{(i,j,k)\in\Omega_i^{(1)}} \ell_{ijk}^{(1)}\left(\mathbf{a}_i^t\right) - \ell_{ijk}^{(1)}\left(\mathbf{a}_i^*\right)\right] \leq \frac{\left|\Omega_i^{(1)}\right| L\eta G^2}{4\mu}. \tag{23}$$

TABLE I
OPTIMAL HYPERPARAMETER CONFIGURATIONS

Algorithm	MovieLens 25M	MovieLens 32M	WS-DREAM RT	WS-DREAM TP
Ours	$\eta = 10^{-3}, \lambda = 10^{-5}$	$\eta = 10^{-3}, \lambda = 10^{-5}$	$\eta = 10^{-4}, \lambda = 0$	$\eta = 10^{-3}, \lambda = 0$
FP-CPD [1]	$\eta = 10^{-3}, \lambda = 10^{-3}$	$\eta = 10^{-3}, \lambda = 10^{-3}$	$\eta = 10^{-4}, \lambda = 10^{-4}$	$\eta = 10^{-3}, \lambda = 10^{-4}$
FTCDP [2]	$\eta = 10^{-3}, \lambda = 10^{-2}$	$\eta = 10^{-3}, \lambda = 10^{-2}$	$\eta = 10^{-4}, \lambda = 0$	$\eta = 10^{-4}, \lambda = 0$
DisMASTD [3]	$\lambda = 10^{-1}$	$\lambda = 10^{-1}$	$\lambda = 10^{-4}$	$\lambda = 10^{-4}$
MuLOT [4]	$\lambda = 10^{1}$	$\lambda = 10^{1}$	$\lambda = 10^{-3}$	$\lambda = 10^{-3}$
CSF-ALS [5]	$\lambda = 10^{-1}$	$\lambda = 10^{-1}$	$\lambda = 10^{-3}$	$\lambda = 10^{-1}$
CSF-SGD [5]	$\eta = 10^{-3}, \lambda = 10^{-2}$	$\eta = 10^{-3}, \lambda = 10^{-2}$	$\eta = 10^{-4}, \lambda = 10^{-4}$	$\eta = 10^{-3}, \lambda = 10^{-5}$
CDTF [6]	$\lambda = 10^{1}$	$\lambda = 10^{1}$	$\lambda = 10^{-1}$	$\lambda = 10^{-1}$
SALS [6]	$\lambda = 10^{1}$	$\lambda = 10^{1}$	$\lambda = 10^{1}$	$\lambda = 10^{-2}$
ASTEN [7]	$\eta = 10^{-3}, \lambda = 10^{-5}$	$\eta = 10^{-3}, \lambda = 10^{-5}$	$\eta = 10^{-4}, \lambda = 0$	$\eta = 10^{-2}, \lambda = 10^{-5}$
FlexiFact [8]	$\eta = 10^{-3}, \lambda = 10^{-2}$	$\eta = 10^{-3}, \lambda = 10^{-2}$	$\eta = 10^{-4}, \lambda = 10^{-5}$	$\eta = 10^{-3}, \lambda = 10^{-5}$

For all $i \in 1, \ldots, I$, (23) is reformulated as

$$\mathbb{E}\left[\sum_{i=1}^{I} \sum_{(i,j,k)\in\Omega_i^{(1)}} \ell_{ijk}^{(1)}\left(\mathbf{a}_i^t\right) - \ell_{ijk}^{(1)}\left(\mathbf{a}_i^*\right)\right] \leq \frac{|\Omega|\, L\eta G^2}{4\mu}. \tag{24}$$

Similarly for other subproblems, summing over all entries in Ω gives the bound for the global loss under the alternating scheme, as block-coordinate descent ensures monotonic decrease to a stationary point. This completes the proof.

II. EXPERIMENTAL SETUP

A. Data Preprocessing

To convert datasets into tensors, we apply the following preprocessing steps:

- **MovieLens Datasets:** The original timestamps are converted into calendar dates, and user–item interactions are aggregated on a weekly basis. Each week is treated as an independent temporal slice, forming a three-way user–item–time tensor.
- **WS-DREAM Datasets:** The QoS measurements are organized into a three-way user–service–time tensor. For WS-DREAM RT, which have a relatively small numeric range, values were linearly scaled to the interval [0, 5]. For the WS-DREAM TP, which exhibit large variance, a logarithmic (log) transformation is applied to stabilize variance and mitigate the influence of extreme outliers.

B. Hyperparameter Settings

To ensure the fairness of comparisons, hyperparameters are tuned via grid search. The learning rate η is selected from $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$ and the regularization coefficient λ from $\{10^{1}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 0\}$. The resulting optimal hyperparameter configurations for each algorithm are summarized in Table I.

III. NUMERIC RESULTS AND SIGNIFICANCE ANALYSIS

A. Numeric Results

Tables II and III report the training time and throughput results for each method under different ranks. Tables IV and V report the training time and throughput results for each method under different numbers of workers.

B. Significance Analysis

This section rigorously evaluates whether the observed performance gaps are statistically significant across methods. Two nonparametric procedures are adopted: (i) the Friedman rank test to detect overall performance differences among all methods on all datasets, and (ii) post-hoc Wilcoxon signed-rank tests to compare DecSGD against each baseline. To make metrics commensurate, negative throughput is used so that lower is better for both training time and the transformed throughput.

1) Friedman Rank Test: Let K denote the number of methods and N denote the number of paired evaluation blocks within a given case. Within each block $n \in \{1, \ldots, N\}$, algorithms are ranked by performance (lower is better), with ties receiving averaged ranks. Denote by $r_{k,n}$ the rank of algorithm $k \in \{1, \ldots, K\}$ in block n, and by

$$\bar{R}_k \;=\; \frac{1}{N}\sum_{n=1}^{N} r_{k,n} \tag{25}$$

its average rank over the N blocks.

TABLE II
TRAINING TIME COMPARISON VS. VARYING RANKS (SECONDS)

Method	Movielens 25M				Movielens 32M			
	$R=5$	$R=10$	$R=20$	$R=40$	$R=5$	$R=10$	$R=20$	$R=40$
DecSGD	$0.61_{\pm 0.00}$	$0.79_{\pm 0.01}$	$1.19_{\pm 0.00}$	$2.19_{\pm 0.00}$	$0.81_{\pm 0.01}$	$1.04_{\pm 0.00}$	$1.58_{\pm 0.01}$	$2.90_{\pm 0.01}$
FP-CPD	$14.37_{\pm 0.29}$	$22.97_{\pm 0.10}$	$42.38_{\pm 0.11}$	$78.00_{\pm 0.19}$	$20.15_{\pm 1.19}$	$31.77_{\pm 0.33}$	$57.18_{\pm 0.51}$	$100.30_{\pm 0.86}$
FTCDP	$4.91_{\pm 0.22}$	$7.93_{\pm 0.08}$	$12.59_{\pm 0.59}$	$18.83_{\pm 0.54}$	$8.32_{\pm 0.11}$	$12.40_{\pm 0.04}$	$17.37_{\pm 0.04}$	$25.51_{\pm 0.08}$
DisMASTD	$8.25_{\pm 0.19}$	$24.88_{\pm 0.41}$	$8.94_{\pm 0.66}$	$14.24_{\pm 0.04}$	$11.20_{\pm 0.37}$	$27.94_{\pm 0.98}$	$27.81_{\pm 4.39}$	$29.36_{\pm 0.54}$
MuLOT	$4.78_{\pm 0.29}$	$8.50_{\pm 0.08}$	$14.76_{\pm 0.13}$	$41.60_{\pm 1.23}$	$6.84_{\pm 0.18}$	$11.92_{\pm 0.13}$	$21.72_{\pm 0.34}$	$56.97_{\pm 0.45}$
CSF-ALS	$6.79_{\pm 0.02}$	$8.91_{\pm 0.02}$	$14.05_{\pm 0.03}$	$29.03_{\pm 0.07}$	$9.40_{\pm 0.34}$	$12.28_{\pm 0.20}$	$18.90_{\pm 0.07}$	$43.54_{\pm 0.99}$
CSF-SGD	$1.06_{\pm 0.02}$	$1.23_{\pm 0.00}$	$1.69_{\pm 0.05}$	$2.47_{\pm 0.03}$	$1.34_{\pm 0.05}$	$1.59_{\pm 0.02}$	$2.10_{\pm 0.08}$	$3.11_{\pm 0.08}$
CDTF	$47.38_{\pm 0.14}$	$91.47_{\pm 0.20}$	$188.33_{\pm 0.29}$	$366.42_{\pm 0.38}$	$72.48_{\pm 0.52}$	$143.17_{\pm 2.61}$	$293.11_{\pm 3.19}$	$561.33_{\pm 2.91}$
SALS	$15.06_{\pm 0.19}$	$28.64_{\pm 0.31}$	$56.42_{\pm 0.19}$	$111.23_{\pm 0.34}$	$22.43_{\pm 0.39}$	$42.68_{\pm 0.84}$	$92.63_{\pm 2.01}$	$167.54_{\pm 0.17}$
ASTEN	$6.01_{\pm 1.23}$	$12.02_{\pm 0.93}$	$27.85_{\pm 0.94}$	$80.37_{\pm 9.12}$	$7.31_{\pm 0.47}$	$16.06_{\pm 0.06}$	$36.81_{\pm 0.08}$	$111.86_{\pm 1.01}$
FlexiFaCT	$4.85_{\pm 0.24}$	$6.68_{\pm 0.15}$	$10.14_{\pm 0.04}$	$18.81_{\pm 0.03}$	$6.54_{\pm 0.31}$	$8.77_{\pm 0.18}$	$15.57_{\pm 0.31}$	$26.80_{\pm 0.49}$

Method	WS-DREAM RT				WS-DREAM TP			
	$R=5$	$R=10$	$R=20$	$R=40$	$R=5$	$R=10$	$R=20$	$R=40$
DecSGD	$1.47_{\pm 0.44}$	$1.84_{\pm 0.13}$	$1.99_{\pm 0.10}$	$3.13_{\pm 0.09}$	$1.77_{\pm 0.24}$	$1.65_{\pm 0.15}$	$2.06_{\pm 0.02}$	$3.37_{\pm 0.19}$
FP-CPD	$112.12_{\pm 9.44}$	$187.50_{\pm 0.97}$	$257.05_{\pm 3.90}$	$345.57_{\pm 3.96}$	$107.12_{\pm 7.36}$	$183.30_{\pm 5.40}$	$264.32_{\pm 5.90}$	$345.45_{\pm 1.62}$
FTCDP	$4.69_{\pm 0.42}$	$5.15_{\pm 0.26}$	$5.72_{\pm 0.19}$	$6.68_{\pm 0.19}$	$5.82_{\pm 0.75}$	$5.49_{\pm 0.09}$	$5.66_{\pm 0.12}$	$6.87_{\pm 0.06}$
DisMASTD	$4.01_{\pm 0.10}$	$12.49_{\pm 2.53}$	$11.31_{\pm 0.52}$	$24.06_{\pm 1.05}$	$2.87_{\pm 0.22}$	$16.41_{\pm 2.13}$	$12.67_{\pm 0.55}$	$23.04_{\pm 0.68}$
MuLOT	$6.49_{\pm 0.17}$	$22.47_{\pm 0.57}$	$15.24_{\pm 0.45}$	$28.85_{\pm 0.84}$	$9.73_{\pm 1.62}$	$19.30_{\pm 1.53}$	$15.85_{\pm 0.41}$	$27.35_{\pm 1.83}$
CSF-ALS	$3.10_{\pm 0.14}$	$4.10_{\pm 0.47}$	$5.79_{\pm 0.11}$	$10.23_{\pm 0.08}$	$2.92_{\pm 0.39}$	$3.97_{\pm 0.18}$	$4.43_{\pm 0.37}$	$9.01_{\pm 0.66}$
CSF-SGD	$5.47_{\pm 0.31}$	$7.18_{\pm 0.66}$	$6.84_{\pm 0.47}$	$8.51_{\pm 0.18}$	$6.02_{\pm 0.44}$	$7.66_{\pm 0.51}$	$7.25_{\pm 0.09}$	$8.51_{\pm 0.15}$
CDTF	$15.66_{\pm 0.58}$	$30.39_{\pm 0.22}$	$62.88_{\pm 0.26}$	$121.01_{\pm 0.31}$	$16.87_{\pm 1.30}$	$31.24_{\pm 1.12}$	$63.26_{\pm 2.25}$	$121.88_{\pm 0.35}$
SALS	$4.83_{\pm 0.38}$	$9.26_{\pm 0.84}$	$17.98_{\pm 0.51}$	$34.53_{\pm 1.79}$	$4.94_{\pm 0.05}$	$10.02_{\pm 0.91}$	$19.45_{\pm 1.67}$	$34.45_{\pm 1.39}$
ASTEN	$32.31_{\pm 1.93}$	$48.66_{\pm 6.19}$	$76.85_{\pm 3.86}$	$149.16_{\pm 5.74}$	$32.48_{\pm 1.00}$	$50.24_{\pm 1.74}$	$71.22_{\pm 12.56}$	$133.77_{\pm 15.83}$
FlexiFaCT	$11.43_{\pm 2.41}$	$14.24_{\pm 0.27}$	$17.09_{\pm 0.59}$	$25.38_{\pm 0.44}$	$11.58_{\pm 0.73}$	$14.20_{\pm 0.36}$	$17.88_{\pm 0.45}$	$25.04_{\pm 0.21}$

Note: The values in the table represent the average training time (in seconds) plus/minus the standard deviation (Mean ± Std). The Red marking indicates the minimum training time (best performance) result for a given rank R.

The Friedman statistic is then computed as

$$\chi_F^2 = \frac{12N}{K(K+1)} \left(\sum_{k=1}^{K} \bar{R}_k^2 - \frac{K(K+1)^2}{4} \right), \tag{26}$$

which, under the null hypothesis H_0 that all methods perform equivalently, approximately follows a chi-square distribution with $k-1$ degrees of freedom.

We report the Friedman test results for $K = 11$ methods (DecSGD and ten baselines) over $N = 4$ (datasets)$\times 2$ (metrics) $= 8$ blocks, where the metrics are training time and throughput. Table VI reports the Friedman test results for training time and throughput under different ranks R (Tables II and III). Using a significance level of $\alpha = 0.05$, we observe that all p-values in Table VI are far below this threshold (e.g., for $R = 40$, $\chi_F^2 = 60.00$ with $p = 3.62 \times 10^{-9}$). Therefore, H_0 is rejected for all ranks, indicating that the performance differences among the methods are statistically significant. In all cases, DecSGD achieves an average rank of 1.00, meaning that it consistently ranks first across all evaluation blocks. Competing methods exhibit substantially larger average ranks, which confirms that DecSGD is globally superior in training time and throughput at different ranks. Table VII reports the Friedman test results for training time and throughput at different numbers of workers W (Table IV and V). Using a significance level of $\alpha = 0.05$, all reported p-values are far below this threshold (e.g., for $W = 32$, $\chi_F^2 = 69.66$ with $p = 5.16 \times 10^{-11}$). Thus, H_0 is rejected for each worker configuration, indicating that the observed differences are statistically significant. Across all worker counts, DecSGD again achieves an average rank of 1.00, and other methods exhibit substantially larger average ranks. These results reaffirm that DecSGD delivers statistically significant advantages in training efficiency across all workers.

2) Wilcoxon Signed-Rank Test: To identify the source of significance, pairwise Wilcoxon signed-rank tests compare DecSGD against each baseline on the same N paired blocks used above. For a given case, let $\{d_n\}_{n=1}^{N}$ denote the paired differences (baseline minus DecSGD) on a "lower-is-better" scale. Under the null hypothesis H_0 that the two methods perform equally, these differences are assumed to be symmetrically distributed around zero. After ranking the absolute differences, the sums of

TABLE III
THROUGHPUT COMPARISON VS. VARYING RANKS (UPDATES/SEC)

Method	Movielens 25M				Movielens 32M			
	$R=5$	$R=10$	$R=20$	$R=40$	$R=5$	$R=10$	$R=20$	$R=40$
DecSGD	$285.57_{\pm 2.14}$	$221.57_{\pm 2.00}$	$147.31_{\pm 0.43}$	$80.04_{\pm 0.18}$	$277.99_{\pm 2.55}$	$214.05_{\pm 1.21}$	$142.29_{\pm 0.30}$	$77.15_{\pm 0.36}$
FP-CPD	$12.18_{\pm 0.17}$	$7.62_{\pm 0.14}$	$4.12_{\pm 0.00}$	$2.24_{\pm 0.00}$	$11.16_{\pm 0.63}$	$7.05_{\pm 0.07}$	$3.94_{\pm 0.01}$	$2.23_{\pm 0.02}$
FTCDP	$35.73_{\pm 1.67}$	$22.08_{\pm 0.42}$	$13.87_{\pm 0.03}$	$9.29_{\pm 0.01}$	$26.93_{\pm 0.34}$	$18.06_{\pm 0.06}$	$12.88_{\pm 0.00}$	$8.78_{\pm 0.03}$
DisMASTD	$21.21_{\pm 0.15}$	$7.08_{\pm 0.57}$	$20.05_{\pm 1.35}$	$12.36_{\pm 0.39}$	$20.01_{\pm 0.65}$	$8.03_{\pm 0.28}$	$8.35_{\pm 0.96}$	$7.63_{\pm 0.14}$
MuLOT	$36.65_{\pm 1.50}$	$20.61_{\pm 0.78}$	$11.60_{\pm 0.12}$	$4.21_{\pm 0.03}$	$32.77_{\pm 0.86}$	$18.79_{\pm 0.21}$	$10.27_{\pm 0.04}$	$3.93_{\pm 0.03}$
CSF-ALS	$25.60_{\pm 0.54}$	$19.66_{\pm 0.44}$	$12.50_{\pm 0.23}$	$6.03_{\pm 0.08}$	$23.86_{\pm 0.89}$	$18.24_{\pm 0.28}$	$11.62_{\pm 0.08}$	$5.15_{\pm 0.12}$
CSF-SGD	$165.18_{\pm 3.15}$	$141.86_{\pm 0.41}$	$105.24_{\pm 0.88}$	$71.03_{\pm 0.79}$	$167.02_{\pm 5.50}$	$140.81_{\pm 1.25}$	$108.41_{\pm 1.43}$	$72.12_{\pm 1.85}$
CDTF	$3.70_{\pm 0.10}$	$1.91_{\pm 0.02}$	$0.93_{\pm 0.00}$	$0.48_{\pm 0.01}$	$3.09_{\pm 0.02}$	$1.57_{\pm 0.03}$	$0.76_{\pm 0.00}$	$0.40_{\pm 0.00}$
SALS	$11.62_{\pm 0.17}$	$6.04_{\pm 0.11}$	$3.06_{\pm 0.01}$	$1.57_{\pm 0.01}$	$9.99_{\pm 0.17}$	$5.25_{\pm 0.10}$	$2.44_{\pm 0.02}$	$1.34_{\pm 0.00}$
ASTEN	$29.16_{\pm 1.39}$	$14.55_{\pm 0.09}$	$6.17_{\pm 0.02}$	$2.18_{\pm 0.03}$	$30.77_{\pm 1.91}$	$13.95_{\pm 0.05}$	$6.06_{\pm 0.05}$	$2.00_{\pm 0.02}$
FlexiFaCT	$36.23_{\pm 2.07}$	$26.22_{\pm 0.39}$	$17.56_{\pm 0.03}$	$9.30_{\pm 0.10}$	$34.34_{\pm 1.61}$	$25.56_{\pm 0.63}$	$14.66_{\pm 0.05}$	$8.36_{\pm 0.15}$
Method	**WS-DREAM RT**				**WS-DREAM TP**			
	$R=5$	$R=10$	$R=20$	$R=40$	$R=5$	$R=10$	$R=20$	$R=40$
DecSGD	$155.66_{\pm 1.91}$	$115.69_{\pm 1.13}$	$105.71_{\pm 0.82}$	$67.84_{\pm 0.61}$	$121.79_{\pm 1.99}$	$129.66_{\pm 1.16}$	$101.00_{\pm 0.80}$	$63.11_{\pm 0.62}$
FP-CPD	$45.58_{\pm 3.90}$	$41.32_{\pm 2.15}$	$37.21_{\pm 0.93}$	$31.76_{\pm 0.90}$	$37.13_{\pm 5.24}$	$38.61_{\pm 0.64}$	$37.75_{\pm 1.27}$	$30.89_{\pm 0.27}$
FTCDP	$52.98_{\pm 1.33}$	$17.69_{\pm 3.59}$	$18.72_{\pm 0.27}$	$8.63_{\pm 0.39}$	$74.36_{\pm 5.48}$	$13.14_{\pm 1.67}$	$17.64_{\pm 0.21}$	$9.27_{\pm 0.21}$
DisMASTD	$32.68_{\pm 0.68}$	$9.44_{\pm 0.25}$	$14.03_{\pm 0.21}$	$7.35_{\pm 0.22}$	$22.36_{\pm 3.41}$	$11.06_{\pm 0.93}$	$13.27_{\pm 0.58}$	$7.79_{\pm 0.55}$
MuLOT	$68.53_{\pm 3.07}$	$52.49_{\pm 6.35}$	$39.46_{\pm 0.61}$	$20.72_{\pm 0.16}$	$73.91_{\pm 10.22}$	$53.50_{\pm 2.38}$	$41.77_{\pm 1.70}$	$23.65_{\pm 1.73}$
CSF-ALS	$57.53_{\pm 2.61}$	$51.46_{\pm 3.51}$	$44.33_{\pm 2.90}$	$35.37_{\pm 2.94}$	$58.37_{\pm 4.81}$	$49.94_{\pm 1.89}$	$33.68_{\pm 1.63}$	$33.11_{\pm 0.91}$
CSF-SGD	$38.86_{\pm 2.22}$	$29.79_{\pm 2.55}$	$30.91_{\pm 1.08}$	$24.91_{\pm 0.53}$	$35.44_{\pm 2.71}$	$27.82_{\pm 1.92}$	$28.90_{\pm 0.24}$	$24.93_{\pm 0.46}$
CDTF	$44.22_{\pm 3.63}$	$23.07_{\pm 1.97}$	$11.96_{\pm 0.96}$	$6.16_{\pm 0.31}$	$42.93_{\pm 4.48}$	$21.34_{\pm 1.90}$	$11.43_{\pm 0.57}$	$6.16_{\pm 0.24}$
SALS	$6.59_{\pm 0.40}$	$4.43_{\pm 0.60}$	$2.93_{\pm 0.05}$	$1.42_{\pm 0.05}$	$6.53_{\pm 0.21}$	$4.23_{\pm 0.15}$	$2.99_{\pm 0.28}$	$1.61_{\pm 0.17}$
ASTEN	$18.55_{\pm 4.72}$	$14.99_{\pm 0.29}$	$12.29_{\pm 0.15}$	$8.35_{\pm 0.18}$	$18.39_{\pm 1.21}$	$14.94_{\pm 0.37}$	$12.06_{\pm 0.45}$	$8.47_{\pm 0.19}$
FlexiFaCT	$38.40_{\pm 0.16}$	$8.08_{\pm 0.82}$	$1.80_{\pm 0.01}$	$1.93_{\pm 0.06}$	$16.71_{\pm 0.94}$	$11.33_{\pm 0.80}$	$3.77_{\pm 0.01}$	$3.59_{\pm 0.62}$

Note: The values in the table represent the average throughput (in operations per second) plus/minus the standard deviation (Mean $\pm$ Std). The Red marking indicates the maximum throughput (best performance) result for a given rank R.

positive and negative ranks are defined as

$$R^+ = \sum_{d_n>0} \text{rank}(|d_n|), \qquad R^- = \sum_{d_n<0} \text{rank}(|d_n|). \tag{27}$$

The Wilcoxon statistic is

$$T = \min(R^+, R^-),$$

from which the p-value is computed under H_0.

Table VIII presents the Wilcoxon test results for all ranks. Since there are $N = 8$ blocks, the total sum of ranks is 36. For each pairwise comparison, we observe

$$R^+ = 36.0, \quad R^- = 0.0, \quad p = 1.2 \times 10^{-2}. \tag{28}$$

Using a significance level of $\alpha = 0.05$, these p-values lead to rejection of H_0 for all pairwise comparisons, indicating DecSGD outperforms each baseline for all ranks. Table IX presents the Wilcoxon test results for all worker counts. The results mirror those across ranks:

$$R^+ = 36.0, \quad R^- = 0.0, \quad p = 1.2 \times 10^{-2}. \tag{29}$$

for each pair-wise comparison. Consequently, H_0 is rejected in all worker settings, indicating that DecSGD outperforms each baseline for all numbers of workers.

3) Summary of Significance Findings: Across all cases, the Friedman tests consistently detect overall differences among algorithms, and the Wilcoxon tests show that DecSGD outperforms each baseline on every paired comparison. The improvements of DecSGD are therefore practically and statistically consistent.

TABLE IV
TRAINING TIME COMPARISON VS. NUMBER OF WORKERS (SECONDS)

Method	Movielens 25M					Movielens 32M				
	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$
DecSGD	$12.65_{\pm0.04}$	$6.49_{\pm0.01}$	$3.33_{\pm0.09}$	$1.87_{\pm0.14}$	$1.19_{\pm0.00}$	$17.47_{\pm0.05}$	$8.77_{\pm0.01}$	$5.03_{\pm0.90}$	$2.48_{\pm0.19}$	$1.58_{\pm0.01}$
FP-CPD	$33.70_{\pm0.14}$	$33.87_{\pm0.28}$	$41.50_{\pm0.54}$	$46.96_{\pm0.57}$	$42.38_{\pm0.66}$	$46.04_{\pm1.13}$	$42.85_{\pm0.22}$	$51.08_{\pm0.73}$	$57.01_{\pm0.83}$	$57.18_{\pm0.51}$
FTCDP	$14.43_{\pm0.12}$	$13.14_{\pm0.11}$	$16.33_{\pm0.31}$	$15.48_{\pm0.13}$	$12.59_{\pm0.04}$	$20.03_{\pm0.08}$	$17.81_{\pm0.07}$	$22.92_{\pm0.21}$	$20.77_{\pm0.15}$	$17.37_{\pm0.04}$
DisMASTD	$47.88_{\pm3.08}$	$33.86_{\pm9.18}$	$29.08_{\pm1.50}$	$17.74_{\pm3.08}$	$8.94_{\pm0.75}$	$86.16_{\pm6.81}$	$55.44_{\pm14.31}$	$38.77_{\pm3.19}$	$19.31_{\pm0.52}$	$27.81_{\pm4.39}$
MuLOT	$60.78_{\pm2.68}$	$43.66_{\pm1.13}$	$30.86_{\pm0.44}$	$22.19_{\pm0.17}$	$14.76_{\pm0.19}$	$76.72_{\pm0.87}$	$57.60_{\pm0.90}$	$43.72_{\pm0.88}$	$31.09_{\pm0.33}$	$21.72_{\pm0.34}$
CSF-ALS	$55.92_{\pm0.63}$	$41.73_{\pm0.20}$	$28.71_{\pm0.08}$	$20.58_{\pm0.26}$	$14.05_{\pm0.29}$	$71.78_{\pm1.13}$	$53.58_{\pm0.24}$	$39.58_{\pm0.89}$	$27.33_{\pm0.42}$	$18.90_{\pm0.07}$
CSF-SGD	$13.37_{\pm1.97}$	$8.52_{\pm0.12}$	$4.70_{\pm0.14}$	$2.78_{\pm0.03}$	$1.69_{\pm0.05}$	$17.69_{\pm0.87}$	$11.26_{\pm0.26}$	$6.41_{\pm0.19}$	$3.63_{\pm0.07}$	$2.10_{\pm0.08}$
CDTF	$779.52_{\pm5.58}$	$602.65_{\pm14.63}$	$427.03_{\pm3.30}$	$294.45_{\pm2.23}$	$188.33_{\pm0.94}$	$1084.01_{\pm11.29}$	$848.54_{\pm17.37}$	$622.22_{\pm9.04}$	$432.28_{\pm2.34}$	$293.11_{\pm3.19}$
SALS	$252.17_{\pm2.14}$	$196.11_{\pm4.22}$	$137.60_{\pm3.49}$	$95.33_{\pm0.56}$	$56.42_{\pm0.59}$	$338.00_{\pm5.29}$	$254.98_{\pm5.07}$	$188.56_{\pm4.04}$	$131.44_{\pm1.12}$	$92.63_{\pm2.01}$
ASTEN	$248.75_{\pm1.32}$	$191.30_{\pm2.14}$	$95.82_{\pm0.37}$	$49.62_{\pm0.54}$	$27.85_{\pm0.13}$	$328.29_{\pm1.16}$	$248.16_{\pm6.26}$	$128.02_{\pm2.04}$	$66.34_{\pm0.13}$	$36.81_{\pm0.08}$
FlexiFaCT	$15.41_{\pm0.29}$	$15.23_{\pm0.53}$	$18.14_{\pm0.13}$	$15.97_{\pm0.10}$	$10.14_{\pm0.11}$	$19.93_{\pm0.04}$	$19.84_{\pm0.21}$	$23.99_{\pm0.74}$	$20.72_{\pm0.63}$	$15.57_{\pm0.31}$
Method	**WS-DREAM RT**					**WS-DREAM TP**				
	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$
DecSGD	$10.77_{\pm0.05}$	$6.03_{\pm0.09}$	$3.48_{\pm0.13}$	$2.64_{\pm0.13}$	$1.99_{\pm0.10}$	$11.41_{\pm0.37}$	$6.27_{\pm0.09}$	$3.34_{\pm0.16}$	$2.36_{\pm0.01}$	$2.06_{\pm0.02}$
FP-CPD	$35.08_{\pm3.22}$	$48.94_{\pm2.28}$	$92.19_{\pm2.11}$	$157.91_{\pm5.49}$	$257.05_{\pm3.90}$	$27.31_{\pm0.24}$	$49.78_{\pm1.85}$	$91.43_{\pm1.15}$	$156.66_{\pm4.80}$	$264.32_{\pm5.90}$
FTCDP	$13.27_{\pm0.17}$	$11.22_{\pm0.42}$	$7.94_{\pm0.15}$	$6.22_{\pm0.34}$	$5.72_{\pm0.19}$	$13.61_{\pm0.12}$	$11.11_{\pm1.11}$	$8.30_{\pm0.19}$	$6.20_{\pm0.16}$	$5.66_{\pm0.12}$
DisMASTD	$38.65_{\pm0.47}$	$21.75_{\pm0.30}$	$22.50_{\pm3.28}$	$12.87_{\pm0.13}$	$11.31_{\pm0.52}$	$41.42_{\pm1.34}$	$20.63_{\pm0.40}$	$11.01_{\pm0.25}$	$7.82_{\pm0.35}$	$12.67_{\pm0.55}$
MuLOT	$42.31_{\pm3.26}$	$27.17_{\pm4.04}$	$21.11_{\pm1.98}$	$18.41_{\pm3.52}$	$15.24_{\pm0.45}$	$50.85_{\pm7.34}$	$29.76_{\pm6.32}$	$27.72_{\pm3.63}$	$19.51_{\pm2.06}$	$15.85_{\pm0.41}$
CSF-ALS	$43.44_{\pm0.61}$	$22.90_{\pm0.16}$	$12.48_{\pm0.12}$	$6.88_{\pm0.05}$	$5.79_{\pm0.11}$	$42.73_{\pm0.68}$	$22.72_{\pm0.22}$	$11.86_{\pm0.04}$	$6.88_{\pm0.10}$	$4.43_{\pm0.37}$
CSF-SGD	$12.60_{\pm1.95}$	$11.36_{\pm0.41}$	$8.82_{\pm0.35}$	$8.00_{\pm0.17}$	$6.84_{\pm0.47}$	$12.87_{\pm0.27}$	$11.23_{\pm0.65}$	$9.27_{\pm0.34}$	$7.85_{\pm0.29}$	$7.25_{\pm0.09}$
CDTF	$633.25_{\pm2.89}$	$397.61_{\pm3.09}$	$201.35_{\pm2.36}$	$104.76_{\pm0.70}$	$62.88_{\pm0.26}$	$616.00_{\pm11.08}$	$400.27_{\pm1.48}$	$203.98_{\pm0.75}$	$104.56_{\pm0.91}$	$63.26_{\pm2.25}$
SALS	$191.25_{\pm3.16}$	$114.11_{\pm2.12}$	$58.79_{\pm0.51}$	$29.85_{\pm0.18}$	$17.98_{\pm0.51}$	$185.74_{\pm0.82}$	$112.62_{\pm1.12}$	$57.73_{\pm0.41}$	$29.71_{\pm0.29}$	$19.45_{\pm1.67}$
ASTEN	$230.12_{\pm1.94}$	$277.16_{\pm3.82}$	$207.52_{\pm4.26}$	$114.47_{\pm5.33}$	$76.85_{\pm3.86}$	$239.02_{\pm12.60}$	$284.32_{\pm6.65}$	$175.28_{\pm10.01}$	$117.48_{\pm5.41}$	$71.22_{\pm12.56}$
FlexiFaCT	$12.26_{\pm0.50}$	$11.23_{\pm0.19}$	$12.54_{\pm0.12}$	$14.83_{\pm0.42}$	$17.09_{\pm0.59}$	$11.77_{\pm1.22}$	$10.60_{\pm0.30}$	$12.38_{\pm0.58}$	$14.46_{\pm0.03}$	$17.88_{\pm0.45}$

Note: The values in the table represent the average training time (in seconds) plus/minus the standard deviation (Mean $\pm$ Std). The Red marking indicates the minimum training time (best performance) result for a given worker count W.

TABLE V
THROUGHPUT COMPARISON VS. NUMBER OF WORKERS (UPDATES/SEC)

Method	Movielens 25M					Movielens 32M				
	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$
DecSGD	$13.83_{\pm0.04}$	$26.98_{\pm0.05}$	$52.64_{\pm1.49}$	$94.14_{\pm6.53}$	$147.31_{\pm0.43}$	$12.82_{\pm0.04}$	$25.55_{\pm0.03}$	$45.78_{\pm7.23}$	$90.79_{\pm6.50}$	$142.29_{\pm0.30}$
FP-CPD	$5.19_{\pm0.02}$	$5.17_{\pm0.04}$	$4.22_{\pm0.06}$	$3.73_{\pm0.05}$	$4.12_{\pm0.00}$	$4.87_{\pm0.12}$	$5.23_{\pm0.02}$	$4.38_{\pm0.06}$	$3.93_{\pm0.06}$	$3.94_{\pm0.01}$
FTCDP	$12.13_{\pm0.10}$	$13.32_{\pm0.12}$	$10.72_{\pm0.20}$	$11.31_{\pm0.09}$	$13.87_{\pm0.03}$	$11.18_{\pm0.05}$	$12.58_{\pm0.05}$	$9.77_{\pm0.09}$	$10.79_{\pm0.08}$	$12.88_{\pm0.00}$
DisMASTD	$3.67_{\pm0.23}$	$5.63_{\pm1.73}$	$6.04_{\pm0.31}$	$10.14_{\pm1.59}$	$20.05_{\pm1.35}$	$2.62_{\pm0.20}$	$4.29_{\pm0.99}$	$5.82_{\pm0.49}$	$11.61_{\pm0.31}$	$8.35_{\pm0.96}$
MuLOT	$2.88_{\pm0.13}$	$4.01_{\pm0.10}$	$5.67_{\pm0.08}$	$7.89_{\pm0.06}$	$11.60_{\pm0.12}$	$2.92_{\pm0.03}$	$3.89_{\pm0.06}$	$5.13_{\pm0.10}$	$7.20_{\pm0.08}$	$10.27_{\pm0.04}$
CSF-ALS	$3.13_{\pm0.03}$	$4.19_{\pm0.02}$	$6.10_{\pm0.02}$	$8.51_{\pm0.11}$	$12.50_{\pm0.23}$	$3.12_{\pm0.05}$	$4.18_{\pm0.02}$	$5.66_{\pm0.13}$	$8.20_{\pm0.13}$	$11.82_{\pm0.08}$
CSF-SGD	$13.37_{\pm1.90}$	$20.54_{\pm0.29}$	$37.23_{\pm1.10}$	$62.94_{\pm0.64}$	$105.24_{\pm0.88}$	$12.69_{\pm0.61}$	$19.91_{\pm0.45}$	$34.99_{\pm1.05}$	$61.81_{\pm1.21}$	$108.41_{\pm1.43}$
CDTF	$0.22_{\pm0.00}$	$0.29_{\pm0.01}$	$0.41_{\pm0.00}$	$0.59_{\pm0.00}$	$0.93_{\pm0.00}$	$0.21_{\pm0.00}$	$0.27_{\pm0.00}$	$0.36_{\pm0.00}$	$0.52_{\pm0.00}$	$0.76_{\pm0.00}$
SALS	$0.70_{\pm0.00}$	$0.89_{\pm0.02}$	$1.27_{\pm0.03}$	$1.84_{\pm0.01}$	$3.06_{\pm0.01}$	$0.66_{\pm0.01}$	$0.88_{\pm0.02}$	$1.19_{\pm0.03}$	$1.71_{\pm0.01}$	$2.44_{\pm0.02}$
ASTEN	$0.70_{\pm0.00}$	$0.91_{\pm0.01}$	$1.83_{\pm0.00}$	$3.53_{\pm0.04}$	$6.17_{\pm0.02}$	$0.68_{\pm0.00}$	$0.90_{\pm0.02}$	$1.75_{\pm0.03}$	$3.38_{\pm0.00}$	$6.06_{\pm0.05}$
FlexiFaCT	$11.36_{\pm0.22}$	$11.50_{\pm0.41}$	$9.64_{\pm0.07}$	$10.96_{\pm0.07}$	$17.56_{\pm0.03}$	$11.24_{\pm0.02}$	$11.30_{\pm0.12}$	$9.35_{\pm0.29}$	$10.82_{\pm0.32}$	$14.66_{\pm0.05}$
Method	**WS-DREAM RT**					**WS-DREAM TP**				
	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$	$W=2$	$W=4$	$W=8$	$W=16$	$W=32$
DecSGD	$19.69_{\pm0.09}$	$35.18_{\pm0.54}$	$61.07_{\pm2.13}$	$80.48_{\pm3.90}$	$105.71_{\pm1.80}$	$18.60_{\pm0.61}$	$33.82_{\pm0.52}$	$63.57_{\pm3.00}$	$89.78_{\pm0.35}$	$101.00_{\pm3.77}$
FP-CPD	$6.09_{\pm0.55}$	$4.34_{\pm0.20}$	$2.30_{\pm0.05}$	$1.34_{\pm0.05}$	$0.82_{\pm0.01}$	$7.76_{\pm0.07}$	$4.27_{\pm0.16}$	$2.32_{\pm0.03}$	$1.36_{\pm0.04}$	$0.80_{\pm0.01}$
FTCDP	$15.98_{\pm0.20}$	$18.92_{\pm0.71}$	$26.72_{\pm0.50}$	$34.22_{\pm1.98}$	$37.21_{\pm0.93}$	$15.58_{\pm0.14}$	$19.29_{\pm2.08}$	$25.57_{\pm0.59}$	$34.21_{\pm0.91}$	$37.75_{\pm1.21}$
DisMASTD	$5.49_{\pm0.06}$	$9.75_{\pm0.14}$	$9.61_{\pm1.28}$	$16.48_{\pm0.17}$	$18.72_{\pm0.27}$	$5.12_{\pm0.17}$	$10.22_{\pm0.21}$	$19.36_{\pm0.40}$	$26.18_{\pm2.01}$	$16.84_{\pm0.17}$
MuLOT	$5.04_{\pm0.38}$	$7.99_{\pm1.26}$	$10.14_{\pm0.96}$	$11.90_{\pm2.02}$	$14.03_{\pm0.21}$	$4.27_{\pm0.69}$	$7.51_{\pm1.83}$	$7.77_{\pm0.95}$	$11.00_{\pm1.25}$	$13.27_{\pm0.58}$
CSF-ALS	$4.88_{\pm0.07}$	$9.26_{\pm0.07}$	$17.00_{\pm0.17}$	$30.82_{\pm0.24}$	$39.46_{\pm0.61}$	$4.97_{\pm0.08}$	$9.34_{\pm0.09}$	$17.88_{\pm0.06}$	$30.84_{\pm0.44}$	$41.77_{\pm1.73}$
CSF-SGD	$17.19_{\pm2.40}$	$18.68_{\pm0.67}$	$24.07_{\pm0.94}$	$26.49_{\pm0.55}$	$30.91_{\pm1.08}$	$16.48_{\pm0.35}$	$18.95_{\pm1.07}$	$22.91_{\pm0.83}$	$27.04_{\pm1.00}$	$28.90_{\pm0.24}$
CDTF	$0.34_{\pm0.00}$	$0.53_{\pm0.00}$	$1.05_{\pm0.01}$	$2.03_{\pm0.01}$	$3.35_{\pm0.02}$	$0.34_{\pm0.00}$	$0.53_{\pm0.00}$	$1.04_{\pm0.00}$	$2.03_{\pm0.02}$	$3.38_{\pm0.01}$
SALS	$1.11_{\pm0.02}$	$1.86_{\pm0.04}$	$3.61_{\pm0.03}$	$7.10_{\pm0.04}$	$11.36_{\pm0.06}$	$1.14_{\pm0.00}$	$1.88_{\pm0.02}$	$3.67_{\pm0.02}$	$7.14_{\pm0.07}$	$11.43_{\pm0.57}$
ASTEN	$0.92_{\pm0.00}$	$0.77_{\pm0.01}$	$1.02_{\pm0.02}$	$1.86_{\pm0.09}$	$2.83_{\pm0.05}$	$0.89_{\pm0.05}$	$0.75_{\pm0.02}$	$1.21_{\pm0.07}$	$1.81_{\pm0.09}$	$2.99_{\pm0.28}$
FlexiFaCT	$17.32_{\pm0.71}$	$18.88_{\pm0.33}$	$16.91_{\pm0.17}$	$14.31_{\pm0.41}$	$12.29_{\pm0.22}$	$18.22_{\pm1.99}$	$20.02_{\pm0.58}$	$17.16_{\pm0.82}$	$14.66_{\pm0.03}$	$12.06_{\pm0.45}$

Note: The values in the table represent the average throughput (in operations per second) plus/minus the standard deviation (Mean $\pm$ Std). The Red marking indicates the maximum throughput (best performance) result for a given worker count W.

TABLE VI
FRIEDMAN RANK TEST FOR TRAINING TIME AND THROUGHPUT UNDER DIFFERENT R

R	χ_F^2	p-value	DecSGD	FP-CPD	FTCDP	DisMASTD	MuLOT	CSF-ALS	CSF-SGD	CDTF	SALS	ASTEN	FlexiFaCT
5	45.52	1.75×10^{-6}	1.00	8.63	4.63	6.75	4.13	5.00	4.25	9.13	8.88	7.63	6.00
10	51.84	1.22×10^{-7}	1.00	8.00	5.13	8.25	4.75	4.00	3.25	9.25	9.00	7.88	5.50
20	57.66	1.00×10^{-8}	1.00	8.13	4.38	5.50	5.38	4.13	3.25	10.00	9.63	8.50	6.13
40	60.00	3.62×10^{-9}	1.00	7.50	4.00	5.25	6.50	4.50	2.75	10.00	9.75	8.75	6.00

TABLE VII
FRIEDMAN RANK TEST FOR TRAINING TIME AND THROUGHPUT UNDER DIFFERENT W

W	χ_F^2	p-value	DecSGD	FP-CPD	FTCDP	DisMASTD	MuLOT	CSF-ALS	CSF-SGD	CDTF	SALS	ASTEN	FlexiFaCT
2	77.78	1.37×10^{-12}	1.00	5.00	3.75	6.50	7.50	7.00	2.50	11.00	9.44	9.56	2.75
4	76.34	2.61×10^{-12}	1.00	6.75	2.75	5.38	7.50	6.38	3.00	11.00	9.50	9.50	3.25
8	76.18	2.80×10^{-12}	1.00	8.50	2.50	5.50	6.75	5.00	2.50	10.75	9.00	9.75	4.75
16	71.75	2.03×10^{-11}	1.00	9.50	3.00	4.38	7.00	4.50	3.13	10.00	9.00	9.50	5.00
32	69.66	5.16×10^{-11}	1.00	10.00	3.63	5.00	6.25	3.88	3.00	10.00	9.00	9.00	5.25

TABLE VIII
WILCOXON SIGNED-RANK TEST RESULTS FOR DECSGD VS. BASELINES ON TRAINING TIME AND THROUGHPUT UNDER DIFFERENT R

DecSGD vs.	$R=5$			$R=10$			$R=20$			$R=40$		
	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value
CSF-SGD	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
FTCDP	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
CSF-ALS	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
MuLOT	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
FlexiFaCT	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
DisMASTD	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
FP-CPD	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
ASTEN	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
SALS	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
CDTF	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}

Note: R^+ is the sum of ranks where DecSGD performs better (Positive difference). R^- is the sum of ranks where the comparison method performs better (Negative difference).

TABLE IX
WILCOXON SIGNED-RANK TEST RESULTS FOR DECSGD VS. BASELINES ON TRAINING TIME AND THROUGHPUT UNDER DIFFERENT W

DecSGD vs.	$W=2$			$W=4$			$W=8$			$W=16$			$W=32$		
	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value	R^+	R^-	p-value
CSF-SGD	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
FTCDP	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
CSF-ALS	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
FlexiFaCT	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
MuLOT	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
DisMASTD	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
FP-CPD	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
ASTEN	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
SALS	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}
CDTF	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}	36.0	0.0	1.2×10^{-2}

Note: R^+ is the sum of ranks where DecSGD performs better (Positive difference). R^- is the sum of ranks where the comparison method performs better (Negative difference).

REFERENCES

[1] A. Anaissi, B. Suleiman, W. Alyassine, and S. M. Zandavi, "A fast parallel tensor decomposition with optimal stochastic gradient descent: an application in structural damage identification," *International Journal of Data Science and Analytics*, vol. 17, no. 4, pp. 359–371, 2024.

[2] Z. Yang, B. Xiong, K. Chen, L. T. Yang, X. Deng, C. Zhu, and Y. He, "Differentially private federated tensor completion for cloud–edge collaborative aiot data prediction," *IEEE Internet of Things Journal*, vol. 11, no. 1, pp. 256–267, 2024.

[3] K. Yang, Y. Gao, Y. Shen, B. Zheng, and L. Chen, "Dismastd: An efficient distributed multi-aspect streaming tensor decomposition," in *Proceedings of the ACM Turing Award Celebration Conference-China 2023*, 2023, pp. 127–128.

[4] A. Gillet, É. Leclercq, and N. Cullot, "Multi-level optimization of the canonical polyadic tensor decomposition at large-scale: Application to the stratification of social networks through deflation," *Information Systems*, vol. 112, p. 102142, 2023.

[5] S. Smith, J. Park, and G. Karypis, "Hpc formulations of optimization algorithms for tensor completion," *Parallel Computing*, vol. 74, pp. 99–117, 2018.

[6] K. Shin, L. Sael, and U. Kang, "Fully scalable methods for distributed tensor factorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 1, pp. 100–113, 2017.

[7] Q. Do and W. Liu, "Asten: An accurate and scalable approach to coupled tensor factorization," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 99–106.

[8] A. Beutel, P. P. Talukdar, A. Kumar, C. Faloutsos, E. E. Papalexakis, and E. P. Xing, "Flexifact: Scalable flexible factorization of coupled tensors on hadoop," in *Proceedings of the 2014 SIAM international conference on data mining*. SIAM, 2014, pp. 109–117.