

Supplementary Material of VEST

MoonJeong Park¹, Jun-Gi Jang², and Lee Sael(☒)²

¹ Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea
moonjeong94@dgist.ac.kr

² Department of Computer Science and Engineering, Seoul National University, Seoul, Korea
{elnino4, saelllee}@snu.ac.kr

Abstract. In this supplementary material, we describe additional proofs, algorithms, analysis, and experimental results of VEST.

1 Proofs of Update Rules and Responsibilities

1.1 Update Rules with L_F Regularization

We provide proof of correctness for the entry-wise update rule with L_F regularization.

Lemma 1 (Update rule for factor matrices with L_F regularization).

$$a_{i_n j_n}^{(n)} \leftarrow \arg \min_{a_{i_n j_n}^{(n)}} \mathcal{L}_F(\mathcal{G}, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \frac{\left(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathbf{x}_\alpha \delta_\alpha^{(n)}(j_n) \right) - \left(\sum_{\forall t \neq j_n} \mathbf{v}_{i_n j_n}^{(n)}(t) \cdot a_{i_n t}^{(n)} \right)}{\mathbf{v}_{i_n j_n}^{(n)}(j_n) + \lambda}, \quad (1)$$

where $\mathbf{v}_{i_n j_n}^{(n)}$ is a length J_n vector whose j th element is

$$\mathbf{v}_{i_n j_n}^{(n)}(j) = \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \delta_\alpha^{(n)}(j) \delta_\alpha^{(n)}(j_n), \quad (2)$$

$\delta_\alpha^{(n)}$ is a length J_n vector whose j th element is

$$\delta_\alpha^{(n)}(j) = \sum_{\forall \beta_{j_n=j} \in \mathcal{G}} \mathcal{G}_{\beta_{j_n=j}} \prod_{k \neq n} a_{i_k j_k}^{(k)}, \quad (3)$$

$\Omega_{i_n}^{(n)}$ is the subset of Ω whose index of n th mode is i_n , and $\lambda > 0$ is a regularization parameter. \square

Proof. The value that makes partial derivative of L_F regularization loss with regard to the factor matrix entry $a_{i_n j_n}^{(n)}$ to zero is as follows:

$$\frac{\partial \left(\sum_{\forall \alpha \in \Omega} \left(\mathbf{x}_\alpha - \sum_{\forall \beta \in \mathcal{G}} \mathcal{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)} \right)^2 + \lambda \left(\sum_{\forall \beta \in \mathcal{G}} \mathcal{G}_\beta^2 + \sum_{n=1}^N \sum_{(i_n, j_n) \in \mathbf{A}^{(n)}} a_{i_n j_n}^{(n)2} \right) \right)}{\partial a_{i_n j_n}^{(n)}} = 0$$

$$\begin{aligned}
& \Leftrightarrow \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} ((\mathcal{X}_\alpha - \sum_{\forall \beta \in \mathcal{G}} \mathcal{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)}) \times (-\delta_\alpha^{(n)}(j_n))) + \lambda a_{i_n j_n}^{(n)} = 0 \\
& \Leftrightarrow - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n) + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\sum_{\forall \beta \in \mathcal{G}} \mathcal{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \delta_\alpha^{(n)}(j_n) + \lambda a_{i_n j_n}^{(n)} = 0 \\
& \Leftrightarrow - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n) + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\sum_{t=1}^{J_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)}) \cdot (\delta_\alpha^{(n)}(j_n)) + \lambda a_{i_n j_n}^{(n)} = 0 \\
& \Leftrightarrow - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n) + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\sum_{t \neq j_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)} + \delta_\alpha^{(n)}(j_n) a_{i_n j_n}^{(n)}) \cdot (\delta_\alpha^{(n)}(j_n)) + \lambda a_{i_n j_n}^{(n)} = 0 \\
& \Leftrightarrow (\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\delta_\alpha^{(n)}(j_n))^2 + \lambda) a_{i_n j_n}^{(n)} - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n) - \sum_{t \neq j_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)} \delta_\alpha^{(n)}(j_n)) = 0 \\
& \Leftrightarrow a_{i_n j_n}^{(n)} = \frac{\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n) - \sum_{t \neq j_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)} \delta_\alpha^{(n)}(j_n))}{\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\delta_\alpha^{(n)}(j_n))^2 + \lambda} \\
& \Leftrightarrow a_{i_n j_n}^{(n)} = \frac{(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n)) - (\sum_{\forall t \neq j_n} \mathbf{v}_{i_n j_n}^{(n)}(t) \cdot a_{i_n t}^{(n)})}{\mathbf{v}_{i_n j_n}^{(n)}(j_n) + \lambda},
\end{aligned}$$

where $\mathbf{v}_{i_n j_n}^{(n)}$ is a length J_n vector whose j th entry is

$$\mathbf{v}_{i_n j_n}^{(n)}(j) = \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \delta_\alpha^{(n)}(j) \delta_\alpha^{(n)}(j_n),$$

$\delta_\alpha^{(n)}$ is a length J_n vector whose j th entry is

$$\delta_\alpha^{(n)}(j) = \sum_{\forall \beta_{j_n=j} \in \mathcal{G}} \mathcal{G}_{\beta_{j_n=j}} \prod_{k \neq n} a_{i_k j_k}^{(k)},$$

$\Omega_{i_n}^{(n)}$ is the subset of Ω whose index of n th mode is i_n , and $\lambda > 0$ is a regularization parameter. \square

Similar to the derivation of update rules for factor matrices, the entries of the core tensor is updated by making the partial derivative concerning \mathcal{G}_β to zero.

Lemma 2 (Update rule for core tensor with L_F regularization).

$$\mathcal{G}_\beta \leftarrow \frac{\sum_{\forall \alpha \in \Omega} (\mathcal{X}_\alpha - \sum_{\forall \gamma \neq \beta} \mathcal{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \prod_{n=1}^N a_{i_n j_n}^{(n)}}{\lambda + \sum_{\forall \alpha \in \Omega} (\prod_{n=1}^N a_{i_n j_n}^{(n)})^2} \quad (4)$$

\square

Proof.

$$\begin{aligned}
& \frac{\partial \left(\sum_{\forall \alpha \in \Omega} \left(\mathbf{x}_\alpha - \sum_{\forall \gamma \in \mathcal{G}} \mathcal{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)} \right)^2 + \lambda \left(\sum_{\forall \gamma \in \mathcal{G}} \mathcal{G}_\gamma^2 + \sum_{n=1}^N \sum_{(i_n, j_n) \in \mathbf{A}^{(n)}} a_{i_n j_n}^{(n)2} \right) \right)}{\partial \mathcal{G}_\beta} = 0 \\
& \Leftrightarrow \sum_{\forall \alpha \in \Omega} (\mathbf{x}_\alpha - \sum_{\forall \gamma \in \mathcal{G}} \mathcal{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot (- \prod_{n=1}^N a_{i_n j_n}^{(n)}) + \lambda \mathcal{G}_\beta = 0 \\
& \Leftrightarrow - \sum_{\forall \alpha \in \Omega} (\mathbf{x}_\alpha - \sum_{\forall \gamma \neq \beta} \mathcal{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)} - \mathcal{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \prod_{n=1}^N a_{i_n j_n}^{(n)} + \lambda \mathcal{G}_\beta = 0 \\
& \Leftrightarrow \left(\sum_{\forall \alpha \in \Omega} \left(\prod_{n=1}^N a_{i_n j_n}^{(n)} \right)^2 + \lambda \right) \mathcal{G}_\beta - \sum_{\forall \alpha \in \Omega} (\mathbf{x}_\alpha - \sum_{\forall \gamma \neq \beta} \mathcal{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \prod_{n=1}^N a_{i_n j_n}^{(n)} = 0 \\
& \Leftrightarrow \mathcal{G}_\beta = \frac{\sum_{\forall \alpha \in \Omega} (\mathbf{x}_\alpha - \sum_{\forall \gamma \neq \beta} \mathcal{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \prod_{n=1}^N a_{i_n j_n}^{(n)}}{\lambda + \sum_{\forall \alpha \in \Omega} \left(\prod_{n=1}^N a_{i_n j_n}^{(n)} \right)^2}
\end{aligned}$$

□

1.2 Update Rules with L_1 Regularization

Lemma 3 (Update rule for factor matrix with L_1 regularization).

$$\arg \min_{a_{i_n j_n}^{(n)}} L_1(\mathcal{G}, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \begin{cases} (\lambda - g_{fm})/d_{fm} & \text{if } g_{fm} > \lambda \\ -(\lambda + g_{fm})/d_{fm} & \text{if } g_{fm} < -\lambda \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where

$$g_{fm} = 2 \left(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathbf{x}_\alpha \delta_\alpha^{(n)}(j_n) \right) - \left(\sum_{\forall t \neq j_n} \mathbf{v}_{i_n j_n}^{(n)}(t) \cdot a_{i_n t}^{(n)} \right), \quad (6)$$

$$d_{fm} = 2 \mathbf{v}_{i_n j_n}^{(n)}(j_n), \quad (7)$$

$\mathbf{v}_{i_n j_n}^{(n)}$ is a length J_n vector whose j th element is

$$\mathbf{v}_{i_n j_n}^{(n)}(j) = \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \delta_\alpha^{(n)}(j) \delta_\alpha^{(n)}(j_n), \quad (8)$$

$\delta_\alpha^{(n)}$ is a length J_n vector whose j th element is

$$\delta_\alpha^{(n)}(j) = \sum_{\forall \beta_{j_n=j} \in \mathcal{G}} \mathcal{G}_{\beta_{j_n=j}} \prod_{k \neq n} a_{i_k j_k}^{(k)}, \quad (9)$$

$\Omega_{i_n}^{(n)}$ is the subset of Ω whose index of n th mode is i_n , and $\lambda > 0$ is a regularization parameter.

□

Proof. The partial derivative of L_1 regularization loss function with regard to the factor matrix entry $a_{ijn}^{(n)}$ is

$$\begin{aligned}
& \frac{\partial L_1}{\partial a_{ijn}^{(n)}} \\
&= \left[2 \sum_{\forall \alpha \in \Omega_{in}^{(n)}} ((\mathcal{X}_\alpha - \sum_{\forall t \neq j_n} \delta_\alpha^{(n)}(t) a_{int}^{(n)}) \cdot (-\delta_\alpha^{(n)}(j_n))) \right] + \left[2 \sum_{\forall \alpha \in \Omega_{in}^{(n)}} (\delta_\alpha^{(n)}(j_n))^2 \cdot a_{ijn}^{(n)} \right] + \lambda \frac{\partial |a_{ijn}^{(n)}|}{\partial a_{ijn}^{(n)}} \\
&= g_{fm} + d_{fm} \cdot a_{ijn}^{(n)} + \lambda \frac{\partial |a_{ijn}^{(n)}|}{\partial a_{ijn}^{(n)}} \\
&= \begin{cases} g_{fm} + d_{fm} \cdot a_{ijn}^{(n)} + \lambda & \text{if } a_{ijn}^{(n)} > 0 \\ g_{fm} + d_{fm} \cdot a_{ijn}^{(n)} - \lambda & \text{if } a_{ijn}^{(n)} < 0 \end{cases}
\end{aligned} \tag{10}$$

Case 1. ($g_{fm} > \lambda (> 0)$) : if $a_{ijn}^{(n)} > 0$, then $\frac{\partial L_1}{\partial a_{ijn}^{(n)}} > 0$ since $g_{fm} + \lambda > 0$, and $d_{fm} > 0$. $\frac{\partial L_1}{\partial a_{ijn}^{(n)}} = 0$ when $a_{ijn}^{(n)} = (\lambda - g_{fm})/d_{fm} (> 0)$, so that the value of $\frac{\partial L_1}{\partial a_{ijn}^{(n)}}$ becomes negative when $a_{ijn}^{(n)} < (\lambda - g_{fm})/d_{fm}$. In sum, L_1 decreases if $a_{ijn}^{(n)} < (\lambda - g_{fm})/d_{fm}$, and increases if $a_{ijn}^{(n)} > (\lambda - g_{fm})/d_{fm}$. Thus, the loss becomes minimum when $a_{ijn}^{(n)} = (\lambda - g_{fm})/d_{fm}$.

Case 2. ($g_{fm} < -\lambda (< 0)$) : likewise, if $a_{ijn}^{(n)} < 0$, then $\frac{\partial L_1}{\partial a_{ijn}^{(n)}} < 0$ since $g_{fm} - \lambda < 0$, and $d_{fm} > 0$. $\frac{\partial L_1}{\partial a_{ijn}^{(n)}} = 0$ when $a_{ijn}^{(n)} = -(\lambda + g_{fm})/d_{fm} (> 0)$, so that the value of $\frac{\partial L_1}{\partial a_{ijn}^{(n)}}$ becomes positive when $a_{ijn}^{(n)} > -(\lambda + g_{fm})/d_{fm}$. In sum, the loss decreases if $a_{ijn}^{(n)} < -(\lambda + g_{fm})/d_{fm}$, and increases if $a_{ijn}^{(n)} > -(\lambda + g_{fm})/d_{fm}$. Thus, the loss becomes minimum when $a_{ijn}^{(n)} = -(\lambda + g_{fm})/d_{fm}$ respect to $a_{ijn}^{(n)}$.

Case 3. ($-g_{fm} < \lambda$, and $g_{fm} < \lambda$) : If $a_{ijn}^{(n)} > 0$, then $\frac{\partial L}{\partial a_{ijn}^{(n)}} > 0$; if $a_{ijn}^{(n)} < 0$, then $\frac{\partial L}{\partial a_{ijn}^{(n)}} < 0$. In sum, the loss decreases if $a_{ijn}^{(n)} < 0$, and increases if $a_{ijn}^{(n)} > 0$. Thus, the loss becomes minimum when $a_{ijn}^{(n)} = 0$. □

Lemma 4 (Update rule for core tensor with L_1 regularization).

$$\arg \min_{\mathcal{G}_\beta} L_1(\mathcal{G}, \mathbf{A}^{(1)}, \dots, \mathbf{A}^{(N)}) = \begin{cases} (\lambda - g_c)/d_c & \text{if } g_c > \lambda \\ -(\lambda + g_c)/d_c & \text{if } g_c < -\lambda \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

where

$$g_c = -2 \sum_{\forall \alpha \in \Omega} (\mathcal{X}_\alpha - \sum_{\forall \gamma \neq \beta} \mathcal{G}_\gamma \prod_{n=1}^N a_{ijn}^{(n)}) \cdot \prod_{n=1}^N a_{ijn}^{(n)}, \tag{12}$$

and

$$d_c = 2 \sum_{\forall \alpha \in \Omega} \left(\prod_{n=1}^N a_{i_n j_n}^{(n)} \right)^2. \quad (13)$$

□

Proof. The partial derivative of L_1 regularization loss function with regard to core tensor entry \mathfrak{G}_β is

$$\begin{aligned} & \frac{\partial L_1}{\partial \mathfrak{G}_\beta} \\ &= 2 \sum_{\forall \alpha \in \Omega} (\mathfrak{X}_\alpha - \sum_{\forall \gamma \neq \beta} \mathfrak{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)} - \mathfrak{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \left(- \prod_{n=1}^N a_{i_n j_n}^{(n)} \right) + \lambda \frac{\partial |\mathfrak{G}_\beta|}{\partial \mathfrak{G}_\beta} \\ &= \left[-2 \sum_{\forall \alpha \in \Omega} (\mathfrak{X}_\alpha - \sum_{\forall \gamma \neq \beta} \mathfrak{G}_\gamma \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \left(\prod_{n=1}^N a_{i_n j_n}^{(n)} \right) \right] + \left[2 \sum_{\forall \alpha \in \Omega} \left(\prod_{n=1}^N a_{i_n j_n}^{(n)} \right)^2 \cdot \mathfrak{G}_\beta \right] + \lambda \frac{\partial |\mathfrak{G}_\beta|}{\partial \mathfrak{G}_\beta} \\ &= g_c + d_c \cdot \mathfrak{G}_\beta + \lambda \frac{\partial |\mathfrak{G}_\beta|}{\partial \mathfrak{G}_\beta} \\ &= \begin{cases} g_c + d_c \cdot \mathfrak{G}_\beta + \lambda & \text{if } \mathfrak{G}_\beta > 0 \\ g_c + d_c \cdot \mathfrak{G}_\beta - \lambda & \text{if } \mathfrak{G}_\beta < 0 \end{cases} \end{aligned} \quad (14)$$

The remaining steps are the same as those of update rule for factor matrices with L_1 regularization (Lemma 3). □

1.3 Derivation of Responsibility Values of Factor Matrix Entries

Proof. From its definition,

$$\begin{aligned} (RE)^2 \|\mathfrak{X}\|_F^2 &= \sum_{\forall \alpha \in \Omega} (\mathfrak{X}_\alpha - B(\alpha))^2 \\ &= \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathfrak{X}_\alpha - B(\alpha))^2 + \sum_{\forall \alpha \notin \Omega_{i_n}^{(n)}} (\mathfrak{X}_\alpha - B(\alpha))^2 \\ &= \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathfrak{X}_\alpha - B_{j_n=j}(\alpha) - B_{j_n \neq j}(\alpha))^2 + \sum_{\forall \alpha \notin \Omega_{i_n}^{(n)}} (\mathfrak{X}_\alpha - B(\alpha))^2 \end{aligned} \quad (15)$$

Note that

$$\sum_{\forall \alpha \notin \Omega_{i_n}^{(n)}} (\mathfrak{X}_\alpha - B(\alpha))^2 = (RE)^2 \|\mathfrak{X}\|_F^2 - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathfrak{X}_\alpha - B_{j_n=j}(\alpha) - B_{j_n \neq j}(\alpha))^2 \quad (16)$$

Thus,

$$\begin{aligned}
& (RE(a_{ij}^{(n)}))^2 \|\mathcal{X}\|_F^2 \\
&= \left(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathcal{X}_\alpha - B_{j_n \neq j}(\alpha))^2 \right) + \sum_{\forall \alpha \notin \Omega_{i_n}^{(n)}} (\mathcal{X}_\alpha - B(\alpha))^2 \\
&= \left(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathcal{X}_\alpha - B_{j_n \neq j}(\alpha))^2 \right) + (RE)^2 \|\mathcal{X}\|_F^2 - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\mathcal{X}_\alpha - B_{j_n=j}(\alpha) - B_{j_n \neq j}(\alpha))^2 \\
&= (RE)^2 \|\mathcal{X}\|_F^2 + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (2\mathcal{X}_\alpha - 2B_{j_n \neq j}(\alpha) - B_{j_n=j}(\alpha)) \cdot B_{j_n=j}(\alpha) \\
&= (RE)^2 \|\mathcal{X}\|_F^2 + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (2 \cdot (\mathcal{X}_\alpha - B(\alpha)) + B_{j_n=j}(\alpha)) \cdot B_{j_n=j}(\alpha)
\end{aligned} \tag{17}$$

Dividing both sides of Eq. (17) with $\|\mathcal{X}\|_F^2$, we get

$$(RE(a_{ij}^{(n)}))^2 = RE^2 + \frac{\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (2 \cdot (\mathcal{X}_\alpha - B(\alpha)) + B_{j_n=j}(\alpha)) \cdot (B_{j_n=j}(\alpha))}{\|\mathcal{X}\|_F^2} \tag{18}$$

□

2 Core Tensor Update Algorithm

Element-wise update of the core tensor \mathcal{G} using either the L_F or L_1 regularization is detailed in Algorithm 1. Elements of a core tensor are highly dependent on each other and thus updating them cannot be made parallel, although a part of required computations can be made parallel (line 1 of Algorithm 1). However, considering that typical size $|\mathcal{G}|$ of a core tensor is small, the core tensor updates are a minor burden in the computational process.

3 Theoretical Analysis

3.1 Convergence Analysis.

We theoretically prove the convergence of VEST update rules.

Theorem 1. $\text{VEST}_{L_1}^*$ and $\text{VEST}_{L_F}^*$ converges.

Proof. Both of the loss functions (Eq.(1) and (2) of the main paper) are bounded by 0. Since the proposed element-wise update rule minimizes the loss function at each update sessions, the loss function decreases in every update process and never increases. Thus, VEST converges. □

Algorithm 1: Parallel element-wise core tensor update

Input : Tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_N}$,
factor matrices $A^{(n)} \in \mathbb{R}^{I_n \times J_n} (n = 1, \dots, N)$, and
core tensor $\mathcal{G} \in J_1 \times \dots \times J_N$.
Output: Updated core tensor $\mathcal{G} \in J_1 \times \dots \times J_N$

```

1 for  $\alpha = \forall(i_1, \dots, i_N) \in \Omega$  do ▷ in parallel
2   calculate  $B(\alpha) = \sum_{\forall \beta = (j_1, \dots, j_N) \in \mathcal{G}} \mathcal{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)}$ 
3 for  $\beta = \forall(j_1, \dots, j_N) \in \mathcal{G}$  do
4   calculate  $\sum_{\forall \alpha \in \Omega} (\mathcal{X}_\alpha - B(\alpha) + \mathcal{G}_\beta \prod_{n=1}^N a_{i_n j_n}^{(n)}) \cdot \prod_{n=1}^N a_{i_n j_n}^{(n)}$ 
5   calculate  $\sum_{\forall \alpha \in \Omega} (\prod_{n=1}^N a_{i_n j_n}^{(n)})^2$ 
6   update  $\mathcal{G}_\beta$  using Eq. (4) for  $L_F$  (use Eq. (11) for  $L_1$ ).

```

3.2 Complexity Analysis

We analyze the time and memory complexities of VEST. Assuming equal mode dimensions for a input tensor and a core tensor, i.e., $I_1 = I_2 = \dots = I_N = I$ and $J_1 = J_2 = \dots = J_N = J$, the time complexity per iteration (lines 3-7 in Algorithm 2 of the main paper) of VEST is $O(N^2 J |\mathcal{G}| |\Omega| / T)$, and the memory complexity is $O(TJ + J^N + NIJ)$. Note that we focus on memory complexities of intermediate data, not for all variables.

Theorem 2. *The time complexity per iteration of VEST is*

$$O(N^2 J |\mathcal{G}| |\Omega| / T + NIJ \log(IJ) + |\mathcal{G}| \log |\mathcal{G}|)$$

Proof. Given an $a_{ij}^{(n)}$ that is not pruned (lines 1-5 in Algorithm 4 of the main paper), computing $\delta_\alpha^{(n)}$ (line 8) for an α takes $O(N |\mathcal{G}|)$. Calculating $\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n)$ and $v_{i_n j_n}^{(n)}$ (line 9) takes $O(N |\Omega_{i_n}^{(n)}| |\mathcal{G}|)$. Therefore, the time complexity for updating an element $a_{ij}^{(n)}$ (line 8) is $O(N |\Omega_{i_n}^{(n)}| |\mathcal{G}|)$. Updating an i -th row $a_{i:}^{(n)}$ takes $O(NJ |\Omega_{i_n}^{(n)}| |\mathcal{G}|)$. Updating all elements of $A^{(n)}$ takes $O(NJ |\mathcal{G}| |\Omega| / T)$, using T threads. Thus, updating all factor matrices takes $O(N^2 J |\mathcal{G}| |\Omega| / T)$. According to Algorithm 1, updating core tensor takes $O(N |\mathcal{G}| |\Omega|)$. Calculating $Resp$ of factor matrices and core tensor, and sorting them take $O(N^2 |\mathcal{G}| |\Omega| / T + NIJ \log(IJ))$ and $O(N |\mathcal{G}| |\Omega| / T + |\mathcal{G}| \log |\mathcal{G}|)$ respectively. Therefore, the time complexity per iteration of VEST is $O(N^2 J |\mathcal{G}| |\Omega| / T + NIJ \log(IJ) + |\mathcal{G}| \log |\mathcal{G}|)$. \square

Theorem 3. *The memory complexity of VEST is*

$$O(TJ + J^N + NIJ),$$

where T is the number of threads, I is the dimensionality of a mode of \mathcal{X} , J is the dimensionality of a mode of \mathcal{G} , and N is the order of \mathcal{X} .

Proof. VEST generates intermediate data of size $O(J)$ to update an element in a factor matrix. Using T threads, VEST requires $O(TJ)$ space while updating factor matrices. VEST uses a marking table to indicate pruned and un-pruned elements, which requires $O(J^N + NIJ)$ memory. While pruning, VEST save *Resp* values which require $O(J^N + NIJ)$. In sum, the memory complexity of VEST is $O(TJ + J^N + NIJ)$. \square

4 Hyperparameter Selection

4.1 Selecting Core Sizes

In the case of $I_n \leq 200$, J_n is assigned to $I_n/10$. When $I_n > 200$, we observed the change of reconstruction error value to determine J_n . Reconstruction Error was measured by $VEST_*^{man}$ while increasing all undetermined J_n from 1 to 10. Then the point where the error value becomes minimum is determined as the rank.

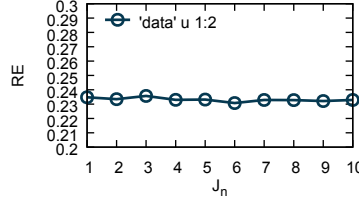


Fig. 1: Reconstruction error value of MovieLens data varying input order.

Figure 1 shows the change in the reconstruction error value of MovieLens data. The x-axis represents J_n and y-axis represent Reconstruction error. The data of the figure is the average value of the three times run. We note that the change value of RE is substantially small, but has a minimum value when $J_n = 6$. In this way, Rank of MovieLens data is determined as $6 \times 6 \times 2 \times 2$. The Rank for the remaining Real data was determined in the same way. However, Rank $5 \times 5 \times 5$ is used for the sampled-data used for comparison with the competition code. This is because the competing code is implemented only for the $J_1 = J_2 = \dots = J_n$ case.

5 Additional Experimental Results

5.1 Comparative Studies

Figure 1 of the main paper is plotted based on the datasets used.

5.2 Sparsity

Figure 3 shows normalized reconstruction error values varying lambda values of $VEST_*^{man}$ on test set (test RE). Results of $VEST_{L_1}^{auto}$ and $VEST_{L_F}^{auto}$ are plotted to show the auto mode find reasonable sparsity at the elbow point of the curve.

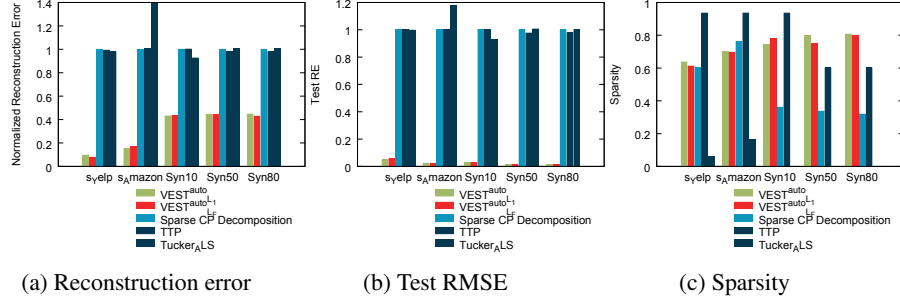


Fig. 2: Accuracy of VEST compared to standard tensor factorization methods on three real world data. Measurement values are averages of five randomly initialized independent runs.

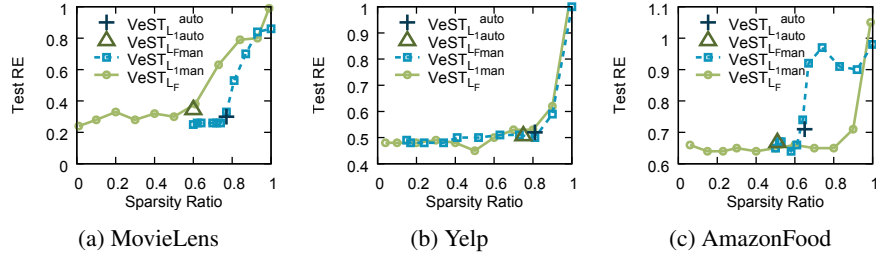


Fig. 3: Sparsity against test RE of $VEST_{L_F}^{man}$ and $VEST_{L_1}^{man}$ with varying target sparsity s from 0 to 0.99. The values are averaged from five randomly initialized independent runs.

5.3 Effect of λ in $VEST_{*}^{auto}$ on Performance.

Fig. 4, shows the effect of λ values on normalized reconstruction error on the test set (Test RE) for $VEST_{L_F}^{auto}$, $VEST_{L_1}^{auto}$, and L_1 , i.e., VEST optimized on L_1 regularizer without pruning. In all three datasets, $VEST_{L_1}^{auto}$ was able to generate sparse results ($s \geq 0.4$) without a significant difference in the test RE values. To balance the error term and λ term in the loss functions, Eq.(1) and (2) in the main paper, we plotted against an adjusted λ' such that $\lambda = |\Omega|/(|\mathcal{S} + \sum_{n=1}^N |A^{(n)}|) \times \lambda'$.

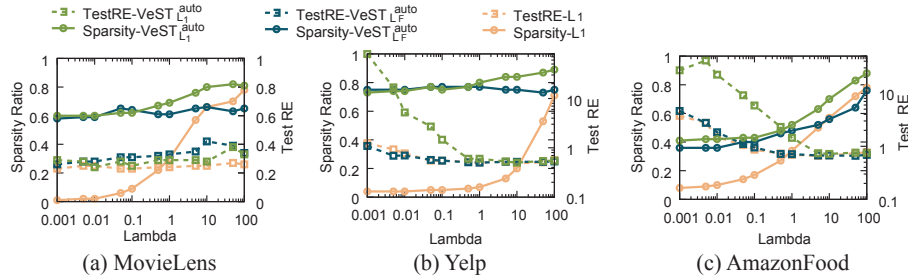


Fig. 4: Sparsity and test RE of $VEST_{L_1}^{man}$ and $VEST_{L_1}^{auto}$ with varying λ values. Measurement values are averaged from five runs.

For $VEST_{L_F}^{auto}$, sufficiently large λ' values were required to generalize on the test data set (higher test RE values at $\lambda' = 0.001$), while the training set RE values remained relatively consistent over the λ' values (Figure 5). However, sparsity and RE values after

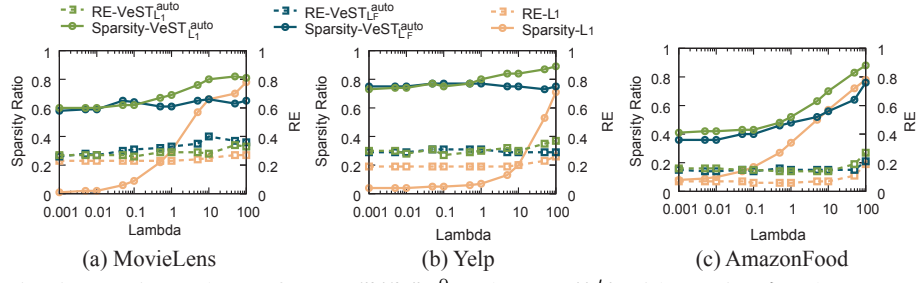


Fig. 5: Sparsity and RE of $\text{VEST}_{L_1}^{\text{man } p=0}$ and $\text{VEST}_{L_1}^{\text{auto}}$ with varying λ values. Measurement values are averaged from five runs.

$\lambda' \geq 5$ remained relatively consistent. For both $\text{VEST}_{L_F}^{\text{auto}}$ and $\text{VEST}_{L_1}^{\text{auto}}$, sparsity and RE values were more consistent compared to that of L_1 , where the sparsity showed a high dependency on λ' . Also, as expected, if the λ' is too small, factorization did not generalize well on the test set as shown by the higher test RE values at $\lambda' = 0.001$.