Supplementary Material of VeST

MoonJeong Park[1], Jun-Gi Jang[2], 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,saellee}@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).

$$a_{i_n j_n}^{(n)} \longleftarrow \underset{a_{i_n j_n}^{(n)}}{\arg\min}\ \mathcal{L}_F(\mathcal{G}, \mathbf{A}^{(1)}, ..., \mathbf{A}^{(N)}) = \frac{\big(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathcal{X}_\alpha \delta_\alpha^{(n)}(j_n)\big) - \big(\sum_{\forall t \neq j_n} \boldsymbol{v}_{i_n j_n}^{(n)}(t) \cdot a_{i_n t}^{(n)}\big)}{\boldsymbol{v}_{i_n j_n}^{(n)}(j_n) + \lambda}, \tag{1}$$

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

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

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

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

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

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:

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

$$\Leftrightarrow \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \big((\mathfrak{X}_\alpha - \sum_{\forall \beta \in \mathfrak{G}} \mathfrak{G}_\beta \prod_{n=1}^{N} a_{i_n j_n}^{(n)}) \times (-\delta_\alpha^{(n)}(j_n)) \big) + \lambda a_{i_n j_n}^{(n)} = 0$$

$$\Leftrightarrow - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathfrak{X}_\alpha \delta_\alpha^{(n)}(j_n) + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\sum_{\forall \beta \in \mathfrak{G}} \mathfrak{G}_\beta \prod_{n=1}^{N} a_{i_n j_n}^{(n)}) \cdot \delta_\alpha^{(n)}(j_n) + \lambda a_{i_n j_n}^{(n)} = 0$$

$$\Leftrightarrow - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathfrak{X}_\alpha \delta_\alpha^{(n)}(j_n) + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\sum_{t=1}^{J_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)}) \cdot (\delta_\alpha^{(n)}(j_n)) + \lambda a_{i_n j_n}^{(n)} = 0$$

$$\Leftrightarrow - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathfrak{X}_\alpha \delta_\alpha^{(n)}(j_n) + \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\sum_{t \neq j_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)} + \delta_\alpha^{(n)}(j_n) a_{i_n j_n}^{(n)}) \cdot (\delta_\alpha^{(n)}(j_n)) + \lambda a_{i_n j_n}^{(n)} = 0$$

$$\Leftrightarrow \big(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\delta_\alpha^{(n)}(j_n))^2 + \lambda \big) a_{i_n j_n}^{(n)} - \sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \big(\mathfrak{X}_\alpha \delta_\alpha^{(n)}(j_n) - \sum_{t \neq j_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)} \delta_\alpha^{(n)}(j_n) \big) = 0$$

$$\Leftrightarrow a_{i_n j_n}^{(n)} = \frac{\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \big(\mathfrak{X}_\alpha \delta_\alpha^{(n)}(j_n) - \sum_{t \neq j_n} \delta_\alpha^{(n)}(t) a_{i_n t}^{(n)} \delta_\alpha^{(n)}(j_n) \big)}{\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} (\delta_\alpha^{(n)}(j_n))^2 + \lambda}$$

$$\Leftrightarrow a_{i_n j_n}^{(n)} = \frac{\big(\sum_{\forall \alpha \in \Omega_{i_n}^{(n)}} \mathfrak{X}_\alpha \delta_\alpha^{(n)}(j_n) \big) - \big(\sum_{\forall t \neq j_n} \boldsymbol{v}_{i_n j_n}^{(n)}(t) \cdot a_{i_n t}^{(n)} \big)}{\boldsymbol{v}_{i_n j_n}^{(n)}(j_n) + \lambda},$$

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

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

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

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

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

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

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

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

□

Proof.

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

$$\Leftrightarrow \sum_{\forall\alpha\in\Omega}(\mathcal{X}_\alpha - \sum_{\forall\gamma\in\mathcal{G}}\mathcal{G}_\gamma\prod_{n=1}^{N}a_{i_nj_n}^{(n)})\cdot(-\prod_{n=1}^{N}a_{i_nj_n}^{(n)}) + \lambda\mathcal{G}_\beta = 0$$

$$\Leftrightarrow -\sum_{\forall\alpha\in\Omega}(\mathcal{X}_\alpha - \sum_{\forall\gamma\neq\beta}\mathcal{G}_\gamma\prod_{n=1}^{N}a_{i_nj_n}^{(n)} - \mathcal{G}_\beta\prod_{n=1}^{N}a_{i_nj_n}^{(n)})\cdot\prod_{n=1}^{N}a_{i_nj_n}^{(n)} + \lambda\mathcal{G}_\beta = 0$$

$$\Leftrightarrow (\sum_{\forall\alpha\in\Omega}(\prod_{n=1}^{N}a_{i_nj_n}^{(n)})^2 + \lambda)\mathcal{G}_\beta - \sum_{\forall\alpha\in\Omega}(\mathcal{X}_\alpha - \sum_{\forall\gamma\neq\beta}\mathcal{G}_\gamma\prod_{n=1}^{N}a_{i_nj_n}^{(n)})\cdot\prod_{n=1}^{N}a_{i_nj_n}^{(n)} = 0$$

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

□

1.2 Update Rules with L_1 Regularization

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

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

where

$$g_{fm} = 2(\sum_{\forall\alpha\in\Omega_{i_n}^{(n)}}\mathcal{X}_\alpha\delta_\alpha^{(n)}(j_n)) - (\sum_{\forall t\neq j_n}\boldsymbol{v}_{i_nj_n}^{(n)}(t)\cdot a_{i_nt}^{(n)}), \tag{6}$$

$$d_{fm} = 2\boldsymbol{v}_{i_nj_n}^{(n)}(j_n), \tag{7}$$

$\boldsymbol{v}_{i_nj_n}^{(n)}$ is a length J_n vector whose jth element is

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

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

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

$\Omega_{i_n}^{(n)}$ is the subset of Ω whose index of nth 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 $a^{(n)}_{i_n j_n}$ is

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

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

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

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

□

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

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

where

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

and

$$d_c = 2 \sum_{\forall \alpha \in \Omega} (\prod_{n=1}^{N} a^{(n)}_{i_n j_n})^2. \tag{13}$$

□

Proof. The partial derivative of L_1 regularization loss function with regard to core tensor entry $\mathcal{G}_\beta$ is

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

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,

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

Note that

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

Thus,

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

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

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

□

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. VEST$^*_{L_1}$ *and* 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 𝒳 ∈ ℝ^{I_1×···×I_N},
        factor matrices A^{(n)} ∈ ℝ^{I_n×J_n} (n = 1, ··· , N), and
        core tensor 𝒢 ∈ J_1 × ...J_N.
Output: Updated core tensor 𝒢 ∈ J_1 × ...J_N
for α = ∀(i_1, ..., i_N) ∈ Ω do                                   ▷ in parallel
    calculate B(α) = Σ_{∀β=(j_1,··· ,j_N)∈𝒢} 𝒢_β Π_{n=1}^{N} a^{(n)}_{i_n j_n}
for β = ∀(j_1, ..., j_N) ∈ 𝒢 do
    calculate Σ_{∀α∈Ω} (𝒳_α − B(α) + 𝒢_β Π_{n=1}^{N} a^{(n)}_{i_n j_n}) · Π_{n=1}^{N} a^{(n)}_{i_n j_n}
    calculate Σ_{∀α∈Ω} (Π_{n=1}^{N} a^{(n)}_{i_n j_n})^2
    update 𝒢_β 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., $I_1 = I_2 = \cdots = I_N = I$ and $J_1 = J_2 = \cdots = J_N = J$, 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 $a^{(n)}_{ij}$ that is not pruned (lines 1-5 in Algorithm 4 of the main paper), computing $\delta^{(n)}_{\alpha}$ (line 8) for an α takes $O(N|\mathcal{G}|)$. Calculating $\sum_{\forall\alpha\in\Omega^{(n)}_{i_n}} \mathcal{X}_\alpha \delta^{(n)}_\alpha(j_n)$ and $\boldsymbol{v}^{(n)}_{i_n j_n}$ (line 9) takes $O(N|\Omega^{(n)}_{i_n}||\mathcal{G}|)$. Therefore, the time complexity for updating an element $a^{(n)}_{ij}$ (line 8) is $O(N|\Omega^{(n)}_{i_n}||\mathcal{G}|)$. Updating an i-th row $a^{(n)}_{i:}$ takes $O(NJ|\Omega^{(n)}_{i_n}||\mathcal{G}|)$. 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 $O(N^2 J|\mathcal{G}||\Omega|/T + NIJ\log(IJ) + |\mathcal{G}|\log|\mathcal{G}|)$. □

Theorem 3. *The memory complexity of* VEST *is*

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

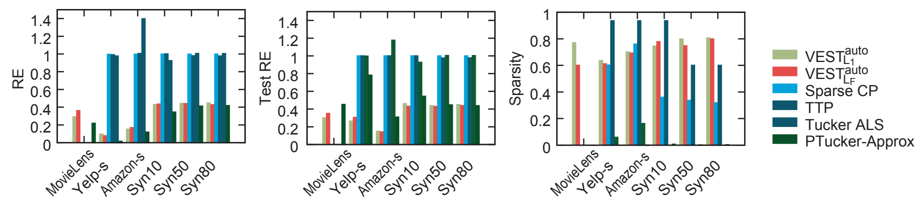
where T is the number of threads, I is the dimensionality of a mode of $\mathcal{X}$, J is the dimensionality of a mode of $\mathcal{G}$, and N is the order of $\mathcal{X}$.

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 $O(TJ + J^N + NIJ)$. □

4 Additional Experimental Results

4.1 Comparative Studies

In Figure 1, we decompose the results of Figure 1 of the main paper in terms of datasets.

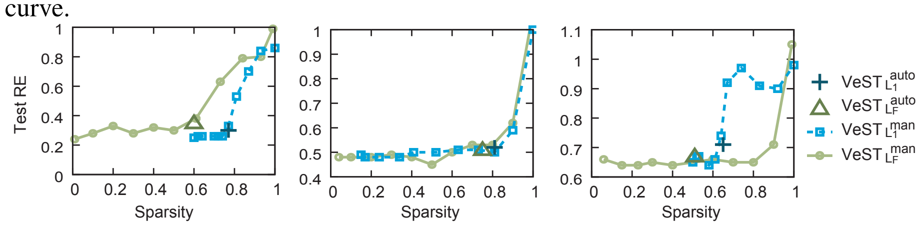


(a) RE (b) Test RE (c) Sparsity

Fig. 1: Accuracy of VEST compared to standard tensor factorization methods on three real world datasets and three synthetic dataset with densities 0.1, 0.5 and 0.8. Measurement values are averages of five randomly initialized independent runs.

4.2 Sparsity

Figure 2 shows test reconstruction error (Test RE) values measured by varying lambda values of VEST$_*^{man}$ on *test set*. Similar to the conclusion from Figure 3 of the main paper, VEST successfully removes redundant information in the decomposition without hurting the accuracy, and automatically finds the sparsity at the elbow point of the RE curve.



(a) MovieLens (b) Yelp (c) AmazonFood

Fig. 2: 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.