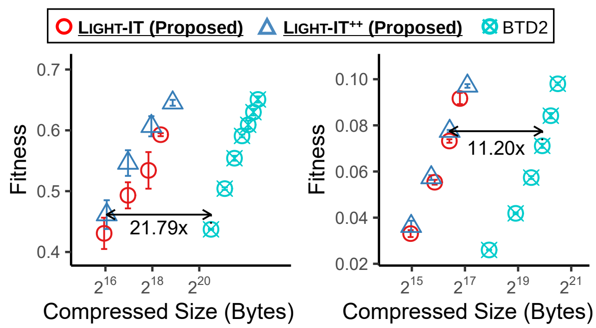
Compact Decomposition of Irregular Tensors for Data Compression: From Sparse to Dense to High-Order Tensors (Supplementary Material)

1 PERFORMANCE COMPARISON WITH BTD2

We add the compression performances of BTD2 [1] on 4-order tensors in Figure 1. Still, our algorithms perform the best in terms of accuracy and compression ability. Note that BTD2 is tailored for dense 4-order irregular tensors. We reduced the size of the 4-order tensors by sampling the second-order and third-order indices (i.e from `Enron` to `Enron_small` and `Delicious` to `Delicious_small`) because O.O.M occurred when running BTD2 on the original tensors. The sizes of `Enron_small` and `Delicious_small` are in Table 1.



(a) `Enron_small` (b) `Delicious_small`

Figure 1: Light-IT and Light-IT^{++} concisely and accurately compress 4-order irregular tensors. Notably, the output size of Light-IT^{++} is up to 21.79× smaller than that of BTD2, showing a similar fitenss.

Table 1: Statistics of real-world datasets.

Name	N_{max}	N_{avg}	Size (except the 1st mode)	Order	Density
`Enron_small`	408	59.4	$50 \times 50 \times 759$	4	5.07×10^{-3}
`Delicious_small`	39	13.9	$50 \times 50 \times 1,001$	4	1.16×10^{-3}

2 THEORETICAL ANALYSIS

We assume that all equations in the main paper are written in the version that can handle a tensor of any order.

Theorem 1 (Compressed size of Light-IT). *The compressed size of Light-IT is $O(R(D + \sum_{i=1}^{d-2} M_i + K) + \log D \sum_{k=1}^{K} N_k)$.*

Proof. The matrix $\mathbf{P}$ requires $O(DR)$, $\{\mathbf{V}_i\}_{i=1}^{d-2}$ require $O(\sum_{i=2}^{d-2} M_i R)$, and S requires $O(KR)$. The mappings $\{\pi_k\}_{k=1}^{K}$ require $O(\log D)$ bits per index of the first mode when the integers follow a uniform distribution, which is the case that the compression result of Huffman encoding is the largest. Thus, $\{\pi_k\}_{k=1}^{K}$ requires $O(\log N_{max} \sum_{k=1}^{K} N_k)$ bits. Hence, the compressed size is $O(R(D + \sum_{i=1}^{d-2} M_i + K) + \log D \sum_{k=1}^{K} N_k)$. □

Theorem 2 (Compressed size of Light-IT^{++}). *The compressed size of Light-IT^{++} is $O(R(D + \sum_{i=1}^{d-2} M_i + K) + \log D \sum_{k=1}^{K} N_k + R^d)$.*

Proof. By Theorem 1, the sum of sizes of $\mathbf{P}$, $\{\mathbf{V}_i\}_{i=1}^{d-2}$, S, $\{\pi_k\}_{k=1}^{K}$ is $O(R(D + \sum_{i=1}^{d-2} M_i + K) + \log D \sum_{k=1}^{K} N_k)$. The core tensor $\mathcal{G}$ takes $O(R^d)$ parameters. Therefore, the total number of parameters is $O(R(D + \sum_{i=1}^{d-2} M_i + K) + \log D \sum_{k=1}^{K} N_k + R^d)$. □

Theorem 3 (Compression time of Light-IT). *The compression time of Light-IT is $O\big((\prod_{i=1}^{d-2} M_i)(\sum_{k=1}^{K} N_k)R + (\sum_{k=1}^{K} N_k)DR\big)$.*

Proof. Computing $\pi_k(i)$ according to Eq. (3) for all k in $\{1, \cdots, K\}$ and i in $[N_k]$ takes $O((\sum_{k=1}^{K} N_k)DR)$ time. Given $\mathbf{P}'_k$ and $((\odot_{i=1}^{d-2}\mathbf{V}_i) \odot \mathrm{S}(k,:))$, computing $\mathbf{P}'_k((\odot_{i=1}^{d-2}\mathbf{V}_i) \odot \mathrm{S}(k,:))^\intercal$ in Eq. (25) for all $k \in \{1, \ldots, K\}$ is the most dominant part during a single epoch. Since the size of $\mathbf{P}'_k$ is $N_k \times R$ and the size of $((\odot_{i=1}^{d-2}\mathbf{V}_i) \odot \mathrm{S}(k,:))$ is $(\prod_{i=1}^{d-2} M_i) \times R$, computing $\mathbf{P}'_k((\odot_{i=1}^{d-2}\mathbf{V}_i) \odot \mathrm{S}(k,:))^\intercal$ takes $O(N_k (\prod_{i=1}^{d-2} M_i)R)$ time, and computing it for all k takes $O((\sum_{k=1}^{K} N_k)(\prod_{i=1}^{d-2} M_i)R)$. Hence, an epoch of Light-IT requires $O((\prod_{i=1}^{d-2} M_i)(\sum_{k=1}^{K} N_k)R + (\sum_{k=1}^{K} N_k)DR)$ time. □

Theorem 4 (Compression time of Light-IT for sparse irregular tensors). *The compression time of the spares version of Light-IT is $O\big((\sum_{k=1}^{K} N_k + \sum_{i=1}^{d-2} M_i)R^2 + \sum_{k=1}^{K} nnz(\mathcal{X}_k)dR + (\sum_{k=1}^{K} N_k)DR\big)$.*

Proof. In the proof, we only mention the most dominant parts of computations. First, Computing $\pi_k(i)$ according to Eq. (3) for all k in $\{1, \cdots, K\}$ and i in $[N_k]$ takes $O((\sum_{k=1}^{K} N_k)DR)$ time. let's consider Eq. (26). Computing $(\prod_{i=1}^{d-2} \mathbf{V}_i^\intercal \mathbf{V}_i)$ consumes $O(\sum_{i=1}^{d-2} M_i R^2)$ time as the size of $\mathbf{V}_i$ is $M_i \times R$. Computing $(\sum_{k=1}^{K} \mathrm{S}(k, r_1)\ \mathrm{S}(k, r_2)\ \mathbf{P}'_k(:, r_1)^\intercal \mathbf{P}'_k(:, r_2))$ for all r_1, r_2 in $\{1, \ldots, R\}$ and k in $\{1, \ldots, K\}$ consumes $O(\sum_{k=1}^{K} N_k R^2)$ time as the sizes of $\mathbf{P}'_k(:, r_1)$ and $\mathbf{P}'_k(:, r_2)$ are both $1 \times N_k$. Approximation of a single entry by Light-IT requires $O(dR)$ time, and approximating whole non-zero entries requires $\sum_{k=1}^{K} O(nnz(\mathcal{X}_k)dR)$ time. It is equivalent to the time complexity for computing the second term in the final formula of Eq. (7). To sum up, the running time for an epoch of Light-IT when compressing a sparse irregular tensor is $O((\sum_{k=1}^{K} N_k + \sum_{i=1}^{d-2} M_i)R^2 + \sum_{k=1}^{K} nnz(\mathcal{X}_k)dR + (\sum_{k=1}^{K} N_k)DR)$. □

Theorem 5 (Compression time of Light-IT^{++}). *The compression time of Light-IT^{++} is $O\big(R^{3(d-2)} + KR^4 + R^6 + R^{2d-1}D + d(\sum_{k=1}^{K} N_k)\prod_{i=1}^{d-2} M_i)R^{d-1} + KR^{2d-1}\big)$.*

Proof. Here, the complexities for the most dominant parts are listed. In Eq. (34), given $(\otimes_{i=1}^{d-2}\mathbf{V}_i^\intercal \mathbf{V}_i)$, computing its pseudoinverse requires $O(R^{3(d-2)})$. Computing $\mathbf{P}_k^\intercal \mathbf{P}_k \otimes \mathrm{S}(k,:)^\intercal \mathrm{S}(k,:)$ from $\mathbf{P}_k^\intercal \mathbf{P}_k$ and $\mathrm{S}(k,:)^\intercal \mathrm{S}(k,:)$ for all k in $\{1, \ldots, K\}$ needs $O(KR^4)$ time. Computing its pseudoinverse takes $O(R^6)$ time. In Eq. (36), earning $\mathbf{G}^{(1)}\Big((\otimes_{i=1}^{d-2}\mathbf{V}_i^\intercal \mathbf{V}_i) \otimes \big(\sum_{(j,k)\in T_i} \mathrm{S}(k,:)^\intercal \mathrm{S}(k,:)\big)\Big)\mathbf{G}^{(1)\,\intercal}$ from $\mathcal{G}$ and

$(\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i)\otimes\sum_{(j,k)\in T_i}\mathbf{S}(k,:)^\top\mathbf{S}(k,:)$ spends $O(R^{2d-1}D)$ time due to their sizes. In Eq. (22), computing $\mathbf{X}_k^{(i)}(\mathbf{P}_k\otimes\mathbf{S}(k,:)\otimes(\otimes_{j\neq i}\mathbf{V}_j))$ from $\mathcal{X}_k$ and $\mathbf{P}_k\otimes\mathbf{S}(k,:)\otimes(\otimes_{j\neq i}\mathbf{V}_j)$ for all k requires $O((\sum_{k=1}^K N_k)$ $(\prod_{i=1}^{d-2}M_i)\ R^{d-1})$ time due to their sizes. It takes $O(d(\sum_{k=1}^K N_k)$ $(\prod_{i=1}^{d-2}M_i)\ R^{d-1})$ in total since Eq. (22) is conducted for $d-2$ times. In Eq. (38), computing $(\mathbf{P}_k^\top\mathbf{P}_k\otimes(\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i))$ from $\mathbf{P}_k^\top\mathbf{P}_k$ and $\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i$ for all k requires $O(KR^{2d-1})$ time. In conclusion, total running time for an epoch is $O(R^{3(d-2)}+KR^4+R^6+R^{2d-1}D+d(\sum_{k=1}^K N_k)\prod_{i=1}^{d-2}M_i)R^{d-1}+KR^{2d-1})$. □

Theorem 6 (Compression time of Light-IT^{++} for spare irregular tensors). *The compression time of the sparse version of Light-IT^{++} is* $O\big(\prod_{i=1}^{d-2}M_i\,R^4+\sum_{k=1}^K nnz(\mathcal{X}_k)\,R^d d+R^{3(d-2)}+(\prod_{i=1}^{d-2}M_i)R^d+KR^4+R^6+R^{2d-1}D+\sum_{i=1}^{d-2}M_iKR^d+d(\sum_{k=1}^K N_k)R^2+KR^{2d-1}\big)$.

Proof. Again, we focus on the most dominant parts of the computations. In Eq. (32), computing $(\otimes_{i=1}^{d-2}\mathbf{V}_i)\mathbf{G}^{(2,\cdots,d-1)}\sum_{k=1}^K(\mathbf{P}_k^\top\mathbf{P}_k\otimes\mathbf{S}(k,:)^\top\mathbf{S}(k,:))$ from $(\otimes_{i=1}^{d-2}\mathbf{V}_i)$ and $\mathbf{G}^{(2,\cdots,d-1)}\sum_{k=1}^K(\mathbf{P}_k^\top\mathbf{P}_k\otimes\mathbf{S}(k,:)^\top\mathbf{S}(k,:))$ requires $O(\prod_{i=1}^{d-2}M_iR^4)$. Computing the second term in Eq. (28) using Eq. (20) requires $O(\sum_{k=1}^K nnz(\mathcal{X}_k)\ R^d d)$ as approximating a single entry takes $O(R^d d)$. In Eq. (36), earning $\mathbf{G}^{(1)}\Big((\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i)\otimes\big(\sum_{(j,k)\in T_i}\mathbf{S}(k,:)^\top\mathbf{S}(k,:)\big)\Big)\mathbf{G}^{(1)\top}$ from $\mathcal{G}$ and $(\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i)\otimes\sum_{(j,k)\in T_i}\mathbf{S}(k,:)^\top\mathbf{S}(k,:)$ spends $O(R^{2d-1}D)$ time due to their sizes. In Eq. (22), computing $\mathbf{X}_k^{(i)}(\mathbf{P}_k\otimes(\otimes_{j\neq i}\mathbf{V}_j)\otimes\mathbf{S}(k,:))\mathbf{G}^{(i)\top}$ from $\mathbf{X}_k^{(i)}(\mathbf{P}_k\otimes(\otimes_{j\neq i}\mathbf{V}_j)\otimes\mathbf{S}(k,:))$ and $\mathcal{G}$ for all k and i takes $O(\sum_{i=1}^{d-2}M_iKR^d)$ due to their sizes. In Eq. (36), computing $\mathbf{P}_k^\top\mathbf{P}_k$ for all k requires $O(\sum_{k=1}^k N_kR^2)$ due to the size of $\mathbf{P}_k$. Since Eq. (36) is repeated for $d-2$ times, this part requires $O(d\sum_{k=1}^k N_kR^2)$ in total. In Eq. (38), computing $(\mathbf{P}_k^\top\mathbf{P}_k\otimes(\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i))$ from $\mathbf{P}_k^\top\mathbf{P}_k$ and $\otimes_{i=1}^{d-2}\mathbf{V}_i^\top\mathbf{V}_i$ for all k requires $O(KR^{2d-1})$ time. To sum up, total running time of Light-IT^{++} for an epoch when it is executed on a sparse irregular tensor is $O(\prod_{i=1}^{d-2}M_iR^4+\sum_{k=1}^K nnz(\mathcal{X}_k)\ R^d d+R^{3(d-2)}+(\prod_{i=1}^{d-2}M_i)R^d+KR^4+R^6+R^{2d-1}D+\sum_{i=1}^{d-2}M_iKR^d+d(\sum_{k=1}^K N_k)R^2+KR^{2d-1})$. □

3 VOCABULARY SIZE

We investigate how the accuracies of our models change depending on the size of the vocabulary, and the results are depicted in Fig. 2. The ranks of the modes are fixed to 5 in the CMS and MIMIC-III datasets, and 4 in the Korea-stock and US-stock datasets. The accuracies of the models are nearly saturated when the vocabulary sizes are near the maximum mode length of the first mode. Thus, the settings used in Section 6.2 are empirically validated.

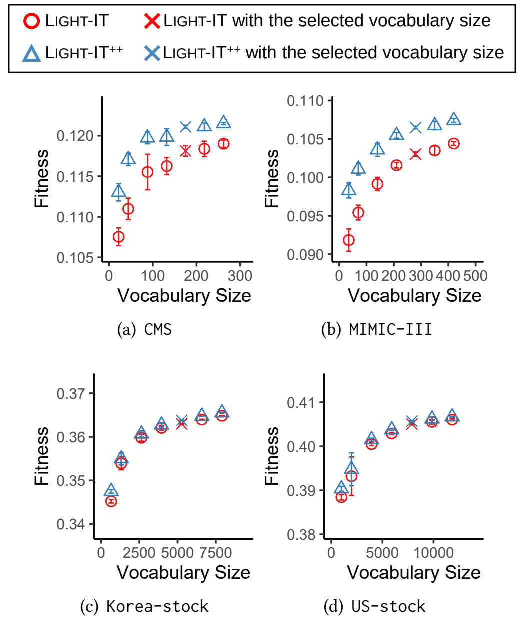


Figure 2: Setting the vocabulary size to the maximum mode length of the first mode is sufficient.

REFERENCES

[1] Christos Chatzichristos, Eleftherios Kofidis, and Sergios Theodoridis. 2017. PARAFAC2 and its block term decomposition analog for blind fMRI source unmixing. In *EUSIPCO*.