A Proof of Propositions

A.1 Proof of Proposition 1

Proposition 1 (Bingo and Tucker rank). *If there are b_k bingos on mode-k, it holds that*

$$\mathrm{Rank}(\mathcal{P}^{(k)}) \leq I_k - b_k.$$

Proof. If there is a bingo on mode-k, the m-th row of the mode-k expansion of $\mathcal{P}$ is a constant multiple of the $(m-1)$-th row, where m is a number determined by the bingo position. Indeed,

$$
\begin{aligned}
\frac{\mathcal{P}^{(k)}_{m,j}}{\mathcal{P}^{(k)}_{m-1,j}} &= \frac{\mathcal{P}_{i_1,\ldots,i_{k-1},m,i_{k+1},\ldots,i_d}}{\mathcal{P}_{i_1,\ldots,i_{k-1},m-1,i_{k+1},\ldots,i_d}} \\
&= \frac{\exp\left(\sum_{i_1'=1}^{i_1}\cdots\sum_{i_{k-1}'=1}^{i_{k-1}}\sum_{i_k'=1}^{m}\sum_{i_{k+1}'=1}^{i_{k+1}}\cdots\sum_{i_d'=1}^{i_d}\theta_{i_1',\ldots,i_d'}\right)}{\exp\left(\sum_{i_1'=1}^{i_1}\cdots\sum_{i_{k-1}'=1}^{i_{k-1}}\sum_{i_k'=1}^{m-1}\sum_{i_{k+1}'=1}^{i_{k+1}}\cdots\sum_{i_d'=1}^{i_d}\theta_{i_1',\ldots,i_d'}\right)} \\
&= \exp\left(\sum_{i_1'=1}^{i_1}\cdots\sum_{i_{k-1}'=1}^{i_{k-1}}\sum_{i_{k+1}'=1}^{i_{k+1}}\cdots\sum_{i_d'=1}^{i_d}\theta_{i_1',\ldots,i_{k-1}',m,i_{k+1}',\ldots,i_d'}\right) \\
&= \exp\left(\theta_{1,\ldots,1,m,1,\ldots,1}\right)
\end{aligned}
$$

is just a constant that does not depend on j. When a row is a constant multiple of another row, the rank of the matrix is reduced by a maximum of one, which means $\mathrm{Rank}(\mathcal{P}^{(k)}) \leq I_k - 1$. In the same way, if there are b_k bingos, then b_k rows are constant multiple of the other rows, which means $\mathrm{Rank}(\mathcal{P}^{(k)}) \leq I_k - b_k$. $\qquad\square$

A.2 Proof of Proposition 2

Proposition 2 (rank-1 condition on θ). *For any positive tensor $\overline{\mathcal{P}}$, $\mathrm{rank}(\overline{\mathcal{P}}) = 1$ if and only if its all many-body $\overline{\theta}$ parameters are 0.*

Proof. First, we show that $\mathrm{rank}(\overline{\mathcal{P}}) = 1$ implies all many-body θ-parameters are 0. From the assumption of $\mathrm{rank}(\overline{\mathcal{P}}) = 1$, the m-th row of the mode-k expansion of $\overline{\mathcal{P}}$ have to be a constant multiple of the $(m-1)$-th row for all $m = \{2,\ldots,I_k\}$ and $k \in [d]$. That is,

$$
\frac{\overline{\mathcal{P}}^{(k)}_{m,j}}{\overline{\mathcal{P}}^{(k)}_{m-1,j}} = \frac{\overline{\mathcal{P}}_{i_1,\ldots,i_{k-1},m,i_{k+1},\ldots,i_d}}{\overline{\mathcal{P}}_{i_1,\ldots,i_{k-1},m-1,i_{k+1},\ldots,i_d}} = \exp\left(\sum_{i_1'=1}^{i_1}\cdots\sum_{i_{k-1}'=1}^{i_{k-1}}\sum_{i_{k+1}'=1}^{i_{k+1}}\cdots\sum_{i_d'=1}^{i_d}\theta_{i_1',\ldots,i_{k-1}',m,i_{k+1}',\ldots,i_d'}\right)
$$

can depend on only m. If any many-body parameter $\theta_{i_1',\ldots,m,\ldots i_d'}$ is not 0 for $i_1' + \cdots + i_{k-1}' + i_{k+1}' + \cdots + i_d' \neq d - 1$, the left side of the above equation depends on indices other than m. For example, if a many-body parameter $\theta_{2,1,\ldots,1,m,1,\ldots 1}$ is not 0, the equation depends on the value of i_1. Therefore, all many-body parameters of rank-1 tensor are 0.

Next, we show that $\mathrm{rank}(\overline{\mathcal{P}}) = 1$ if all many-body θ-parameters are 0. If all many-body θ-parameters are 0, we have

$$
\overline{\mathcal{P}}_{i_1,\ldots,i_d} = \exp\left(\theta_{1,1,\ldots,1}\right)\prod_{k=1}^{d}\exp\left(\sum_{i_k'=2}^{i_k}\theta^{(k)}_{i_k'}\right).
$$

Then we can represent the tensor $\overline{\mathcal{P}}$ as the outer products of d vectors $\boldsymbol{s}^{(1)} \in \mathbb{R}^{I_1}, \boldsymbol{s}^{(2)} \in \mathbb{R}^{I_2},\ldots, \boldsymbol{s}^{(d)} \in \mathbb{R}^{I_d}$, whose elements are described as

$$
s^{(k)}_{i_k} = \exp\left(\frac{\theta_{1,\ldots,1}}{d}\right)\exp\left(\sum_{i_k'=2}^{i_k}\theta^{(k)}_{i_k'}\right)
$$

for each $k \in [d]$. Thus, $\mathrm{rank}(\overline{\mathcal{P}}) = 1$ followed by the definition of the tensor rank.

$\qquad\square$

A.3 Proof of Proposition 3

The following proposition is related with the second paragraph in Section 3.4. We have succeeded in describing the rank-1 condition using η-parameter as well as on θ-parameter.

Proposition 3 (rank-1 condition as η form). *For any positive dth-order tensor $\overline{\mathcal{P}} \in \mathbb{R}_{>0}^{I_1 \times \cdots \times I_d}$, $\mathrm{rank}(\overline{\mathcal{P}}) = 1$ if and only if its all many-body η parameters are factorizable as*

$$\overline{\eta}_{i_1,\ldots,i_d} = \prod_{k=1}^{d} \overline{\eta}_{i_k}^{(k)}. \tag{14}$$

Proof. First, we show that all many-body η-parameters are factorizable if $\mathrm{rank}(\overline{\mathcal{P}}) = 1$. Since we can decompose a rank-1 tensor as a product of normalized independent distributions $s^{(k)} \in \mathbb{R}^{I_k}$ as shown in Section 3.4, we can decompose many-body η parameters of $\overline{\mathcal{P}}$ as follows:

$$
\begin{aligned}
\overline{\eta}_{i_1,\ldots,i_d} &= \sum_{i_1'=i_1}^{I_1} \cdots \sum_{i_d'=i_d}^{I_d} \overline{\mathcal{P}}_{i_1',\ldots,i_d'} \\
&= \sum_{i_1'=i_1}^{I_1} \cdots \sum_{i_d'=i_d}^{I_d} \left(s_{i_1'}^{(1)} s_{i_2'}^{(2)} \cdots s_{i_d'}^{(d)} \right) \\
&= \left(\sum_{i_1'=i_1}^{I_1} s_{i_1'}^{(1)} \right) \left(\sum_{i_2'=i_2}^{I_2} s_{i_2'}^{(2)} \right) \cdots \left(\sum_{i_d'=i_d}^{I_d} s_{i_d'}^{(d)} \right) \\
&= \prod_{k=1}^{d} \left(\sum_{i_k'=i_k}^{I_k} s_{i_k'}^{(k)} \right) \\
&= \prod_{k=1}^{d} \left(\sum_{i_k'=i_k}^{I_k} s_{i_k'}^{(k)} \sum_{i_1'=1}^{I_1} s_{i_1'}^{(1)} \sum_{i_2'=1}^{I_2} s_{i_2'}^{(2)} \cdots \sum_{i_d'=1}^{I_d} s_{i_d'}^{(d)} \right) \\
&= \prod_{k=1}^{d} \left(\overline{\eta}_{i_k}^{(k)} \sum_{i_k'=1}^{I_k} s_{i_k'}^{(k)} \right) \\
&= \prod_{k=1}^{d} \overline{\eta}_{i_k}^{(k)},
\end{aligned}
$$

where we use the normalization condition

$$\sum_{i_k'=1}^{I_k} s_{i_k'}^{(k)} = 1$$

for each $k \in [d]$.

Next, we show the opposite direction. If all many-body η-parameters are factorizable, it follows that

$$
\begin{aligned}
\overline{\mathcal{P}}_{i_1,\ldots,i_d} &= \sum_{(i_1',\ldots,i_d') \in \Omega_d} \left(\mu_{i_1,\ldots,i_d}^{i_1',\ldots,i_d'} \prod_{k=1}^{d} \overline{\eta}_{i_k'}^{(k)} \right) \\
&= \sum_{(i_1'\ldots i_d') \in \Omega_d} \left(\prod_{k=1}^{d} \mu_{i_k}^{i_k'} \overline{\eta}_{i_k'}^{(k)} \right) \\
&= \prod_{k=1}^{d} \left(\overline{\eta}_{i_k}^{(k)} - \overline{\eta}_{i_k+1}^{(k)} \right) \\
&\equiv \prod_{k=1}^{d} s_{j_k}^{(k)}.
\end{aligned}
$$

2

Thus, $\mathrm{rank}(\overline{\mathcal{P}}) = 1$ holds by the definition of the tensor rank. □

A.4 Proof of Proposition 4

The following proposition is related to the second paragraph in Section 3.4. The factorizability of η-parameter of rank-1 tensor and bingo rule reproduces the closed formula of the best rank-1 approximation minimizing KL divergence [3].

Proposition 4 (m-projection onto factorizable subspace). *For any positive tensor $\mathcal{P} \in \mathbb{R}_{>0}^{I_1 \times \cdots \times I_d}$, its m-projection onto the rank-1 space is given as*

$$\overline{\mathcal{P}}_{i_1,\ldots,i_d} = \prod_{k=1}^{d} \left(\sum_{i_1'=1}^{I_1} \cdots \sum_{i_{k-1}'=1}^{I_{k-1}} \sum_{i_{k+1}'=1}^{I_{k+1}} \cdots \sum_{i_d'=1}^{I_d} \mathcal{P}_{i_1',\ldots,i_{k-1}',i_k,i_{k+1}',\ldots,i_d} \right). \tag{15}$$

Proof.

$$\overline{\mathcal{P}}_{i_1,\ldots,i_d} = \sum_{(i_1'\ldots i_d') \in \Omega_d} \mu_{i_1\ldots i_d}^{i_1',\ldots,i_d'} \overline{\eta}_{i_1',\ldots,i_d'}$$

$$\overset{(14)}{=} \sum_{(i_1',\ldots,i_d') \in \Omega_d} \left(\mu_{i_1,\ldots,i_d}^{i_1',\ldots,i_d'} \prod_{k=1}^{d} \overline{\eta}_{i_k'}^{(k)} \right)$$

$$\overset{(6)}{=} \sum_{(i_1'\ldots i_d') \in \Omega_d} \left(\mu_{i_1,\ldots,i_d}^{i_1',\ldots,i_d'} \prod_{k=1}^{d} \eta_{i_k'}^{(k)} \right)$$

$$= \sum_{(i_1',\ldots,i_d') \in \Omega_d} \left(\prod_{k=1}^{d} \mu_{i_k}^{i_k'} \eta_{i_k'}^{(k)} \right)$$

$$= \prod_{k=1}^{d} \left(\eta_{i_k}^{(k)} - \eta_{i_k+1}^{(k)} \right)$$

$$\overset{(4)}{=} \prod_{k=1}^{d} \left(\sum_{i_1'=1}^{I_1} \cdots \sum_{i_{k-1}'=1}^{I_{k-1}} \sum_{i_{k+1}'=1}^{I_{k+1}} \cdots \sum_{i_d'=1}^{I_d} \mathcal{P}_{i_1',\ldots,i_{k-1}',i_k,i_{k+1}',\ldots,i_d} \right).$$

□

Since the m-projection minimizes the KL divergence, it is guaranteed that $\overline{\mathcal{P}}$ obtained by Equation (15) minimizes the KL divergence from $\mathcal{P}$ within the set of rank-1 tensors. If a given tensor is not normalized, we need to divide the right-hand side of Equation (15) by the $d-1$-th power sum of all entries of the tensor in order to match the scales of the input and the output tensors. To summarize, the output of LTR $\overline{\mathcal{P}}$ is the best rank-1 approximation that always minimizes KL divergence,

$$\overline{\mathcal{P}} = \underset{\hat{\mathcal{P}}, \mathrm{rank}(\hat{\mathcal{P}})=1}{\mathrm{argmin}} D_{\mathrm{KL}}(\mathcal{P}; \hat{\mathcal{P}}),$$

where the generalized KL divergence is defined as

$$D_{\mathrm{KL}}(\mathcal{P}; \overline{\mathcal{P}}) = \sum_{i_1=1}^{I_1} \cdots \sum_{i_d=1}^{I_d} \left\{ \mathcal{P}_{i_1\ldots i_d} \log \frac{\mathcal{P}_{i_1,\ldots,i_d}}{\overline{\mathcal{P}}_{i_1,\ldots,i_d}} - \mathcal{P}_{i_1,\ldots,i_d} + \overline{\mathcal{P}}_{i_1,\ldots,i_d} \right\}.$$

The generalized KL divergence for positive tensors is an extension of generalized KL divergence for non-negative matrices in [5], which enables us to treat non-normalized tensors.

A.5 Proof of Proposition 5

The following discussion is related to the third paragraph in Section 3.4.

In the typical Boltzmann machine, which is defined as $p(\boldsymbol{x}) = \exp(\sum_i b_i x_i + \sum_{ij} w_{ij} x_i x_j)$ for a bias parameter $\boldsymbol{b} = (b_i) \in \mathbb{R}^n$, an interaction parameter $\boldsymbol{W} = (w_{ij}) \in \mathbb{R}^{n \times n}$, and a binary random

variable vector $x \in \{0,1\}^n$, mean-field approximation is a projection onto the special manifold where $b_i = \log \frac{\eta_i}{1-\eta_i}$ holds for $\eta_i = \mathbb{E}_p [x_i]$.

In the rank-1 space, we show that θ-parameters can be easily computed from η-parameters, as discussed in the Boltzmann machines in the following proposition; this supports our claim that rank-1 approximation can be regarded as mean-field approximation.

Proposition 5. *For any positive dth-order rank-1 tensor $\overline{\mathcal{P}} \in \mathbb{R}_{>0}^{I_1 \times \cdots \times I_d}$, its one-body η-parameters and one-body θ-parameters satisfy the following equations*

$$\overline{\theta}_j^{(k)} = \log \left(\frac{\overline{\eta}_j^{(k)} - \overline{\eta}_{j+1}^{(k)}}{\overline{\eta}_{j-1}^{(k)} - \overline{\eta}_j^{(k)}} \right), \qquad \overline{\eta}_j^{(k)} = \frac{\sum_{i_k=j}^{I_k} \exp \sum_{i_k'=2}^{i_k} \overline{\theta}_{i_k'}^{(k)}}{1 + \sum_{i_k=2}^{I_k} \exp \left(\sum_{i_k'=2}^{i_k} \overline{\theta}_{i_k'}^{(k)} \right)},$$

where we assume $\overline{\eta}_0^{(k)} = \overline{\eta}_{I_k+1}^{(k)} = 0$.

Proof. As shown in Theorem 2 in Sugiyama et al. [9], the relation between θ and η is obtained by the differentiation of Helmholtz's free energy $\psi(\theta)$, which is defined as the sign inverse normalization factor. For the rank-1 tensor $\overline{\mathcal{P}}$, Helmholtz's free energy $\psi(\theta)$ is given as

$$\psi(\overline{\theta}) = \log \prod_{k=1}^{d} \left(1 + \sum_{i_k=2}^{I_k} \exp \left(\sum_{i_k'=2}^{i_k} \overline{\theta}_{i_k'}^{(k)} \right) \right).$$

We obtain the expectation parameters η by the differentiation of Helmholtz's free energy $\psi(\theta)$ by θ, given as

$$\overline{\eta}_j^{(k)} = \frac{\partial}{\partial \overline{\theta}_j^{(k)}} \psi(\overline{\theta}) = \frac{\sum_{i_k=j}^{I_k} \exp \sum_{i_k'=2}^{i_k} \overline{\theta}_{i_k'}^{(k)}}{1 + \sum_{i_k=2}^{I_k} \exp \left(\sum_{i_k'=2}^{i_k} \overline{\theta}_{i_k'}^{(k)} \right)}.$$

By solving the above equation inverse, we obtain

$$\overline{\theta}_j^{(k)} = \log \left(\frac{\overline{\eta}_j^{(k)} - \overline{\eta}_{j+1}^{(k)}}{\overline{\eta}_{j-1}^{(k)} - \overline{\eta}_j^{(k)}} \right).$$

$\square$

A.6 Proof of Proposition 6

The following proposition is related to the third paragraph in Section 3.5.

Proposition 6. *Let θ denote canonical parameters of given tensor $\mathcal{P} \in \mathbb{R}^{I_1 \times \cdots \times I_d}$ and $\overline{\theta}$ denote canonical parameters of $\overline{\mathcal{P}}$ which is the best rank-1 approximation that minimizes KL divergence from $\mathcal{P}$. If some one-body canonical parameter $\theta_{i_j}^{(j)} = 0$ for some $i_j \in [I_j]$, its values after the best rank-1 approximation $\overline{\theta}_{i_j}^{(j)}$ remain 0.*

Proof. When $\theta_{i_j}^{(j)} = 0$, it holds that

$$\frac{\mathcal{P}_{1,\ldots,1,i_j,1,\ldots,1}}{\mathcal{P}_{1,\ldots,1,i_j-1,1,\ldots,1}} = \exp \left(\theta_{i_j}^{(j)} \right) = 1.$$

By using the closed formula of the best rank-1 approximation (15), we obtain

$$\overline{\mathcal{P}}_{1,\ldots,1,i_j,1,\ldots,1} = \prod_{k=1}^{d} \left(\sum_{i_1'=1}^{I_1} \cdots \sum_{i_{k-1}'=1}^{I_{k-1}} \sum_{i_{k+1}'=1}^{I_{k+1}} \cdots \sum_{i_d'=1}^{I_d} \mathcal{P}_{i_1',\ldots,i_{k-1}',1,i_{k+1}',\ldots,i_j,\ldots,i_d'} \right)$$

$$= \prod_{k=1}^{d} \left(\sum_{i_1'=1}^{I_1} \cdots \sum_{i_{k-1}'=1}^{I_{k-1}} \sum_{i_{k+1}'=1}^{I_{k+1}} \cdots \sum_{i_d'=1}^{I_d} \mathcal{P}_{i_1',\ldots,i_{k-1}',1,i_{k+1}',\ldots,i_j-1,\ldots,i_d'} \right)$$

$$= \overline{\mathcal{P}}_{1,\ldots,1,i_j-1,1,\ldots,1}.$$

4

It follows that

$$\frac{\overline{\mathcal{P}}_{1,\ldots,1,i_j,1,\ldots,1}}{\overline{\mathcal{P}}_{1,\ldots,1,i_j-1,1,\ldots,1}} = \exp\left(\overline{\theta}_{i_j}^{(j)}\right) = 1.$$

Finally, we obtain $\overline{\theta}_{i_j}^{(j)} = 0$. $\qquad\square$

A.7 Proof of Theorem 1

Theorem 1 (LTR). *For a positive tensor $\mathcal{P} \in \mathbb{R}_{>0}^{I_1 \times \cdots \times I_d}$, the m-projection destination onto the bingo space $\mathcal{B} = \mathcal{B}^{(1)} \cap \cdots \cap \mathcal{B}^{(d)}$ is given as iterative application of m-projection d times, starting from $\mathcal{P}$ onto subspace $\mathcal{B}^{(1)}$, and then from there onto subspace $\mathcal{B}^{(2)}$, ..., and finally onto subspace $\mathcal{B}^{(d)}$.*

Proof. Let η denote η-parameters of $\mathcal{P}$. Let $\Omega_{\mathcal{B}^{(k)}}$ be the set of bingo indices for $\mathcal{B}^{(k)}$:

$$\mathcal{B}^{(k)} = \left\{ \mathcal{P} \mid \theta_{i_1,\ldots,i_d} = 0 \text{ for } (i_1,\ldots,i_d) \in \Omega_{\mathcal{B}^{(k)}} \right\}.$$

In the first m-projection onto $\mathcal{B}^{(1)}$ from input $\mathcal{P}$, given us $\mathcal{P}_{(1)}$ whose parameters satisfy

$$\begin{aligned} \tilde{\theta}_{i_1,\ldots,i_d} &= 0 && \text{if } (i_1,\ldots,i_d) \in \Omega_{\mathcal{B}^{(1)}}, \\ \tilde{\eta}_{i_1,\ldots,i_d} &= \eta_{i_1,\ldots,i_d} && \text{otherwise.} \end{aligned}$$

The θ-condition comes from the definition of the bingo space $\mathcal{B}^{(1)}$ and the η-condition comes from the conservation low of η-parameters in Equation (6). The second m-projection onto $\mathcal{B}^{(2)}$ from $\mathcal{P}_{(1)}$, we get $\mathcal{P}_{(1,2)}$ whose parameters satisfy

$$\begin{aligned} \tilde{\theta}_{i_1,\ldots,i_d} &= 0 && \text{if } (i_1,\ldots,i_d) \in \Omega_{\mathcal{B}^{(1)}} \cup \Omega_{\mathcal{B}^{(2)}}, \\ \tilde{\eta}_{i_1,\ldots,i_d} &= \eta_{i_1,\ldots,i_d} && \text{otherwise.} \end{aligned}$$

Proposition 6 ensures the above θ condition. The conservation low of η-parameters ensures the above η condition. Similarly, in the final m-projection onto $\mathcal{B}^{(d)}$ from input $\mathcal{P}_{(1,\ldots,d-1)}$, we get $\mathcal{P}_{(1,\ldots,d-1,d)}$ whose parameters satisfy

$$\begin{aligned} \tilde{\theta}_{i_1,\ldots,i_d} &= 0 && \text{if } (i_1,\ldots,i_d) \in \Omega_{\mathcal{B}^{(1)}} \cup \Omega_{\mathcal{B}^{(2)}} \cup \cdots \cup \Omega_{\mathcal{B}^{(d)}}, \\ \tilde{\eta}_{i_1,\ldots,i_d} &= \eta_{i_1,\ldots,i_d} && \text{otherwise.} \end{aligned}$$

The distribution satisfying these two conditions is the m-projection from $\mathcal{P}$ to $\mathcal{B}$. $\qquad\square$

B Theoretical Remarks

Invariance of the summation in each axial direction   The definition of η in Equation (4) suggests that one-body η-parameters are related to the summation of elements of a tensor in each axial direction. The i_k-th summation in the k-th axis is given by

$$\sum_{i_1=1}^{I_1} \cdots \sum_{i_{k-1}=1}^{I_{k-1}} \sum_{i_{k+1}=1}^{I_{k+1}} \cdots \sum_{i_d=1}^{I_d} \mathcal{P}_{i_1,\ldots,i_d} = \eta_{i_k}^{(k)} - \eta_{i_{k+1}}^{(k)}.$$

Since the one-body η-parameters do not change by the m-projection, it can be immediately proved that the best rank-1 approximation of a positive tensor in the sense of the KL divergence does not change the sum in each axial direction of the input tensor. Our information geometric insight leads to the fact that the conservation law of sums essentially comes from constant one-body η-parameters during m-projection. This property is a natural extension of the property, such that row sums and column sums are preserved in NMF, which minimizes the KL divergence [1] to tensors. Since the rank-1 reduction preserves the sum in each axial direction of the input tensor, LTR for general Tucker rank reduction also preserves it.

C Experiment Setup

All evaluation code are attached in the supplementary material.

Implementation Details All methods were implemented in `Julia 1.6` with `TensorToolbox`[1] library [7], hence runtime comparison is fair. We implemented lraSNTD referring to the original papers [10]. We used the `TensorLy` implementation [4] for NTDs. Experiments were conducted on CentOS 6.10 with a single core of 2.2 GHz Intel Xeon CPU E7-8880 v4 and 3TB of memory. We use default values of hyper parameters of `tensorly` [4] for NTD. We used default values of the hyper parameters of `sklearn` [6] for the NMF module in lraSNTD.

Dataset Details We describe the details of each dataset in the following. 4DLFD is a $(9, 9, 512, 512, 3)$ tensor, which is produced by 4D Light Field Dataset described in [2]. Its license is Creative Commons Attribution-Non-Commercial-ShareAlike 4.0 International License. We used dino images and their depth and disparity map in training scenes and concatenated them to produce a tensor. AttFace is a $(92, 112, 400)$ tensor that is produced by the entire data in The Database of Faces (AT&T) [8], which includes 400 grey-scale face photos. The size of each image is $(92, 112)$. AttFace is public on Kaggle but the license is not specified.

D   Tensor Operations

D.1   mode-k expansion

The mode-k expansion of a tensor $\mathcal{P} \in \mathbb{R}^{I_1 \times \cdots \times I_d}$ is an operation that focuses on the kth axis of the tensor $\mathcal{P}$ and converts $\mathcal{P} \in \mathbb{R}^{I_1 \times \cdots \times I_d}$ into a matrix $\mathcal{P}^{(k)} \in \mathbb{R}^{I_k \times \prod_{m=1 (m \neq k)}^{d} I_m}$. The relation between tensor $\mathcal{P}$ and its mode-k expansion $\mathcal{P}^{(k)}$ is given as,

$$\left(\mathcal{P}^{(k)} \right)_{i_k, j} = \mathcal{P}_{i_1, \dots, i_d},$$

$$j = 1 + \sum_{l=1, (l \neq k)}^{d} (i_l - 1) J_l,$$

$$J_l = \prod_{m=1, (m \neq k)}^{l-1} I_m.$$

D.2   Kronecker product for vectors

Given d vectors $\boldsymbol{a}^{(1)} \in \mathbb{R}^{I_1}$, $\boldsymbol{a}^{(2)} \in \mathbb{R}^{I_2}$, ..., $\boldsymbol{a}^{(d)} \in \mathbb{R}^{I_d}$, the Kronecker product $\mathcal{P}$ of these d vectors, written as

$$\mathcal{P} = \boldsymbol{a}^{(1)} \otimes \boldsymbol{a}^{(2)} \otimes \cdots \otimes \boldsymbol{a}^{(d)},$$

is a tensor in $\mathbb{R}^{I_1 \times \cdots \times I_d}$, where each element of $\mathcal{P}$ is given as

$$\mathcal{P}_{i_1, \dots, i_d} = a_{i_1}^{(1)} a_{i_2}^{(2)} \dots a_{i_d}^{(d)},$$

where $a_i^{(k)}$ is an ith element of a vector $\boldsymbol{a}^{(k)}$.

E   Experimental results with KL divergence

The cost function of LTR is the KL divergence from the input tensor to the low-rank tensor. Our experimental results in Figure 4 show that LTR also has better or competitive accuracy of approximation in terms of the KL divergence.

---

[1]MIT Expat License

(a) Random (30,30,30,30,30) tensor

(b) Target rank (10,10,10)
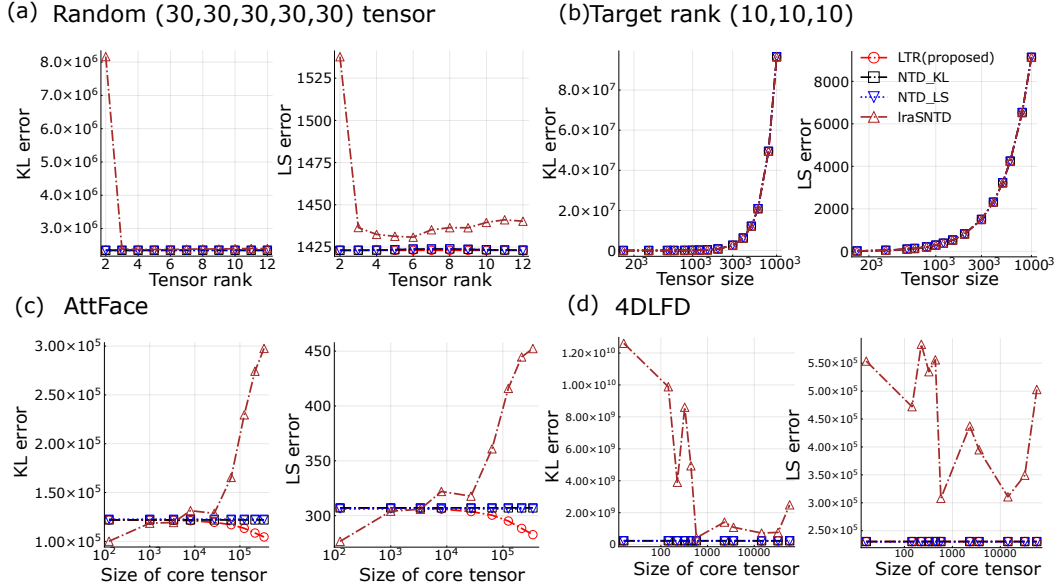
(c) AttFace

(d) 4DLFD

Figure 4: Experimental results for synthetic (a, b) and real-world (c, d) datasets. The left-hand panels are KL reconstruction error and the right-hand panels are LS reconstruction error. (a) The horizontal axis is r for target tensor rank (r, r, r, r, r). (b) The horizontal axis is n^3 for input (n, n, n) tensor.

Checklist

1. For all authors...

   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] Scope and contributions are clearly described in Section 1.

   (b) Did you describe the limitations of your work? [Yes] The limitations are clearly described in Section 1.

   (c) Did you discuss any potential negative societal impacts of your work? [Yes]

   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]

2. If you are including theoretical results...

   (a) Did you state the full set of assumptions of all theoretical results? [Yes] All assumptions are described in Section 3.

   (b) Did you include complete proofs of all theoretical results? [Yes] We have included them in Supplement.

3. If you ran experiments...

   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] All the data and code necessary for the reproduction experiment are attached as Supplement.

   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [N/A] Our method is not based on training.

   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes] Error bar is plotted in Section 4.

   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] We have stated in Supplement.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...

   (a) If your work uses existing assets, did you cite the creators? [Yes]

   (b) Did you mention the license of the assets? [Yes]

(c) Did you include any new assets either in the supplemental material or as a URL? [N/A]

(d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]

(e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]

5. If you used crowdsourcing or conducted research with human subjects...

(a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]

(b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]

(c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]

References

[1] Ho, N.-D. and Van Dooren, P. (2008). Non-negative matrix factorization with fixed row and column sums. *Linear Algebra and its Applications*, 429(5–6):1020–1025.

[2] Honauer, K., Johannsen, O., Kondermann, D., and Goldluecke, B. (2016). A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*, pages 19–34. Springer.

[3] Huang, K. and Sidiropoulos, N. D. (2017). Kullback-leibler principal component for tensors is not np-hard. In *2017 51st Asilomar Conference on Signals, Systems, and Computers*, pages 693–697. IEEE.

[4] Kossaifi, J., Panagakis, Y., Anandkumar, A., and Pantic, M. (2019). Tensorly: Tensor learning in python. *Journal of Machine Learning Research*, 20(26):1–6.

[5] Lee, D. and Seung, H. S. (2001). Algorithms for non-negative matrix factorization. In Leen, T., Dietterich, T., and Tresp, V., editors, *Advances in Neural Information Processing Systems*, volume 13, pages 556–562. MIT Press.

[6] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[7] Periša, L. and Arslan, A. (2019). lanaperisa/tensortoolbox.jl v1.0.2.

[8] Samaria, F. S. and Harter, A. C. (1994). Parameterisation of a stochastic model for human face identification. In *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, pages 138–142. IEEE.

[9] Sugiyama, M., Nakahara, H., and Tsuda, K. (2017). Tensor balancing on statistical manifold. *arXiv preprint arXiv:1702.08142*.

[10] Zhou, G., Cichocki, A., and Xie, S. (2012). Fast nonnegative matrix/tensor factorization based on low-rank approximation. *IEEE Transactions on Signal Processing*, 60(6):2928–2940.