MANY-BODY APPROXIMATION FOR TENSORS

Anonymous authors
Paper under double-blind review

ABSTRACT

We propose a nonnegative tensor decomposition with focusing on the relationship between the modes of tensors. Traditional decomposition methods assume low-rankness in the representation, resulting in difficulties in global optimization and target rank selection. To address these problems, we present an alternative way to decompose tensors, a *many-body approximation* for tensors, based on an information geometric formulation. A tensor is treated via an energy-based model, where the tensor and its mode correspond to a probability distribution and a random variable, respectively, and many-body approximation is performed on it by taking the interaction between variables into account. Our model can be globally optimized in polynomial time in terms of the KL divergence minimization, which is empirically faster than low-rank approximations keeping comparable reconstruction error. Furthermore, we visualize interactions between modes as *tensor networks* and reveal a nontrivial relationship between many-body approximation and low-rank approximation.

1 INTRODUCTION

Tensors are generalization of vectors and matrices. Data in various fields such as neuroscience (Erol & Hunyadi, 2022), bioinformatics (Luo et al., 2017), signal processing (Cichocki et al., 2015), and computer vision (Panagakis et al., 2021) are often stored in the form of tensors, and features are extracted from them. *Tensor decomposition* is one of the most popular methods that extract features by approximating tensors by the sum of products of smaller size of tensors, often called *factors*. It usually tries to minimize the difference between the tensor reconstructed from obtained factors and an original tensor, called the reconstruction error.

In most of tensor decomposition approaches, a *low-rank structure* is typically assumed, where a given tensor is approximated by a linear combination of a small number of bases. Such decomposition requires the following two information. First, it requires the structure, which specifies the type of decomposition such as CP decomposition (Hitchcock, 1927) and Tucker decomposition (Tucker, 1966). In recent years, *tensor networks* (Cichocki et al., 2016) have been introduced, which can intuitively and flexibly design the structure including tensor train decomposition (Oseledets, 2011), tensor ring decomposition (Zhao et al., 2016), and tensor tree decomposition (Murg et al., 2010). Second, it requires the rank value, the number of bases used in the decomposition. Since larger ranks increase the capability of the model while increasing the computational cost, the user is required to find the appropriate rank in this tradeoff problem. Since the above tensor decomposition via minimization of the reconstruction error is non-convex, which causes initial value dependence (Kolda & Bader, 2009, Chapter 3), the problem of finding an appropriate setting of the low-rank structure is highly nontrivial in practice as it is hard to locate the cause if the decomposition does not perform well. As a result, to find proper structure and rank, the user often needs to perform decomposition multiple times with various settings, which is time and memory consuming.

Instead of the low-rank structure that has been the focus of attention in the past, in this paper, we propose a novel formulation of tensor decomposition, called *many-body approximation*, that focuses on the relationship among modes of tensors. We determine the structure of decomposition based on the existence of the interactions between modes. The proposed method requires only the decomposition structure naturally determined by the interactions between the modes and does not require the rank value, which traditional decomposition methods also require and often suffer to determine.

To describe interactions between modes, we follow the standard strategy in statistical mechanics that uses an energy function $\mathcal{H}(\cdot)$ to treat interactions and considers the corresponding distribution $\exp(\mathcal{H}(\cdot))$. This model is known to be an energy-based model in machine learning, which has been used in Legendre decomposition (Sugiyama et al., 2018; 2016) that decomposes tensors via convex optimization. Technically, it finds factors of a tensor by treating a tensor as a probability distribution and enforcing some of its natural parameters to be zero. We point out that interactions in the energy function $\mathcal{H}(\cdot)$ can be represented using natural parameters of distribution, and we can successfully formulate many-body approximation as a special case of Legendre decomposition by setting some of natural parameters to be zero. The advantage of this approach is that many-body approximation can be also achieved by convex optimization that minimizes the Kullback–Leibler (KL) divergence (Kullback & Leibler, 1951).

Furthermore, we introduce a way of representing tensor interactions, which visualizes the presence or absence of interactions between modes. We discuss the correspondence between our representation and the tensor network and point out that an operation called coarse-grained transformation (Levin & Nave, 2007), in which multiple tensors are viewed as a new tensor, reveals unexpected relationship between the proposed method and existing methods such as tensor ring and tensor tree decomposition.

We summarize our contribution as follows:

- By focusing on the interaction between modes of tensors, we introduce an alternative rank-free tensor decomposition, many-body approximation. This decomposition is realized by convex optimization.
- We present a way of describing tensor many-body approximation, interaction representation, a diagram that shows interactions within a tensor. This diagram can be transformed into tensor networks, which tells us the relationship between many-body approximation and existing low-rank approximation.
- We empirically show that many-body approximation is more efficient than low-rank approximation with competitive reconstruction errors.

2 Tensor many-body approximation

Our proposal, tensor many-body approximation, is based on the formulation of Legendre decomposition for tensors. We first review Legendre decomposition and its optimization in Section 2.1. We introduce interactions between modes and its visual representation to prepare for many-body approximation in Section 2.2. Using interactions between modes, we define many-body approximation in Section 2.3. Finally, we transform the interaction representation into a tensor network and point out the connection between many-body approximation and existing low-rank decomposition methods in Section 2.4.

In the following discussion, we consider D-order non-negative tensors whose size is $(I_1, \ldots, I_D)$. We assume the sum of all elements in $\mathcal{P}$ is 1 for simplicity, while this assumption can be eliminated using the general property of Kullback–Leibler (KL) divergence, $\lambda D_{KL}(\mathcal{P}, \mathcal{Q}) = D_{KL}(\lambda\mathcal{P}, \lambda\mathcal{Q})$, for any real number λ.

2.1 Reminder to Legendre Decomposition and its optimization

Legendre decomposition is a method to decompose a non-negative tensor by regarding the tensor as a discrete distribution and representing it with a limited number of parameters. We describe a non-negative tensor $\mathcal{P}$ using natural parameters $\boldsymbol{\theta} = (\theta_{1,\ldots,1}, \ldots, \theta_{I_1,\ldots,I_D})$ and its energy function $\mathcal{H}$ as

$$\mathcal{P}_{i_1,\ldots,i_D} = \exp(\mathcal{H}_{i_1,\ldots,i_D}), \quad \mathcal{H}_{i_1,\ldots,i_D} = \sum_{i'_1=1}^{i_1} \cdots \sum_{i'_D=1}^{i_D} \theta_{i'_1,\ldots,i'_D}, \tag{1}$$

where $\theta_{1,\ldots,1}$ has a role of normalization. Here it is clear that a tensor corresponds to a distribution whose sample space is its index set; that is, the value of each element is regarded as the probability of realizing the corresponding index (Sugiyama et al., 2017).

As we can see in equation 1, we can uniquely identify tensors from natural parameters $\boldsymbol{\theta}$. We can compute the natural parameter $\boldsymbol{\theta}$ from a given tensor as

$$\theta_{i_1,\dots,i_D} = \sum_{i'_1=1}^{I_1} \cdots \sum_{i'_D=1}^{I_D} \mu_{i_1,\dots,i_D}^{i'_1,\dots i'_D} \log \mathcal{P}_{i'_1,\dots,i'_D} \tag{2}$$

using the Möbius function defined inductively as follows:

$$\mu_{i_1,\dots,i_D}^{i'_1,\dots,i'_D} = \begin{cases} 1 & \text{if } i_d = i'_d \text{ for all } d \in \{1,\dots,D\}, \\ -\prod_{d=1}^{D} \sum_{j_d=i_d}^{i'_d-1} \mu_{i_1,\dots,i_D}^{j_1,\dots j_D} & \text{if } i_d < i'_d \text{ for all } d \in \{1,\dots,D\}, \\ 0 & \text{otherwise.} \end{cases}$$

The above modelling for non-negative tensors is an instance of the log-linear model on posets (Sugiyama et al., 2017).

Since distribution described by equation 1 belongs to the exponential family, we can also identify each tensor by expectation parameters $\boldsymbol{\eta} = (\eta_{1,\dots,1}, \dots, \eta_{I_1,\dots,I_D})$ using the Möbius inversion formula as

$$\eta_{i_1,\dots,i_D} = \sum_{i'_1=1}^{i_1} \cdots \sum_{i'_D=1}^{i_D} \mathcal{P}_{i'_1,\dots,i'_D}, \quad \mathcal{P}_{i_1,\dots,i_d} = \sum_{i'_1=1}^{I_1} \cdots \sum_{i'_D=1}^{I_D} \mu_{i_1,\dots,i_d}^{i'_1,\dots,i'_d} \eta_{i'_1,\dots,i'_d}. \tag{3}$$

Since distribution is determined by specifying either θ-parameters or η-parameters, they form two coordinate systems called the θ-coordinate system and the η-coordinate system, respectively. By using the dual flatness and orthogonality of these coordinate systems, Legendre decomposition achieves convex optimization as shown in the following.

2.1.1 Optimization

Legendre decomposition approximates a tensor by setting some θ values to be zero, which corresponds to dropping some parameters for regularization. Let B be the set of indices of θ parameters that are imposed to be 0 . Then Legendre decomposition coincides with a projection of a given nonnegative tensor $\mathcal{P}$ onto the subspace $\mathcal{B} = \{\boldsymbol{\theta} \mid \theta_{i_1,\dots,i_D} = 0 \text{ if } (i_1,\dots,i_D) \in B\}$.

Let us consider projection of a given tensor $\mathcal{P}$ onto $\mathcal{B}$. The space of probability distributions is not a Euclidean space. Therefore, it is necessary to consider geometry of probability distributions, which is studied in information geometry. It is known that a subspace with linear constraints on natural parameters θ is flat, called e-flat (Amari, 2016, Chapter 2). The subspace $\mathcal{B}$ is e-flat, meaning that the logarithmic combination, or called e-geodesic, $\mathcal{R} \in \{(1-t)\log \mathcal{Q}_1 + t \log \mathcal{Q}_2 - \phi(t) \mid 0 < t < 1\}$ of any two points $\mathcal{Q}_1, \mathcal{Q}_2 \in \mathcal{B}$ is included in the subspace $\mathcal{B}$, where $\phi(t)$ is a normalizer. There is always a unique point $\overline{\mathcal{P}}$ on the e-flat subspace that minimizes the KL divergence from any point $\mathcal{P}$.

$$\overline{\mathcal{P}} = \operatorname*{arg\,min}_{\mathcal{Q};\mathcal{Q}\in\mathcal{B}} D_{KL}(\mathcal{P}, \mathcal{Q}) \tag{4}$$

This projection is called the m-projection. The m-projection onto a e-flat subspace is a convex optimization. We define two vectors $\boldsymbol{\theta}^B = \{\theta_b \mid b \in B\}$ and $\boldsymbol{\eta}^B = \{\eta_b \mid b \in B\}$. We write as $|\mathcal{B}|$ the number of elements in these vectors since it is equal to the dimension of $\mathcal{B}$. The derivative of the KL divergence and the Hessian matrix $G \in \mathbb{R}^{|\mathcal{B}|\times|\mathcal{B}|}$ are given as

$$\frac{\partial}{\partial \boldsymbol{\theta}^B} D_{KL}(\mathcal{P}, \mathcal{Q}) = \boldsymbol{\eta}^B - \hat{\boldsymbol{\eta}}^B, \quad G_{u,v} = \eta_{\max(i_1,j_1),\dots,\max(i_D,j_D)} - \eta_{i_1,\dots,i_D}\eta_{j_1,\dots,j_D} \tag{5}$$

where $\boldsymbol{\eta}^B$ and $\hat{\boldsymbol{\eta}}^B$ are the expectation parameter of $\mathcal{Q}$ and $\mathcal{P}$, respectively, and $u = (i_1,\dots,i_D), v = (j_1,\dots,j_D) \in B$. This matrix G is also known as the negative Fisher information matrix. Using natural gradient, we can update $\boldsymbol{\theta}^B$ in each iteration t as

$$\boldsymbol{\theta}_{t+1}^B = \boldsymbol{\theta}_t^B - G^{-1}(\boldsymbol{\eta}_t^B - \hat{\boldsymbol{\eta}}^B) \tag{6}$$

The distribution $\mathcal{Q}_{t+1}$ is calculated from the updated natural parameters $\boldsymbol{\theta}_{t+1}$. This step finds a point $\mathcal{Q}_{t+1} \in \mathcal{B}$ that is closer to the destination $\overline{\mathcal{P}}$ along with the e-geodesic from $\mathcal{Q}_t$ to $\overline{\mathcal{P}}$. We can also calculate the expected value parameters $\boldsymbol{\eta}_{t+1}$ from the distribution. By repeating this process until convergence, we can always find the globally optimal solution satisfying equation 4. This procedure is illustrated in Figure. 1(a).

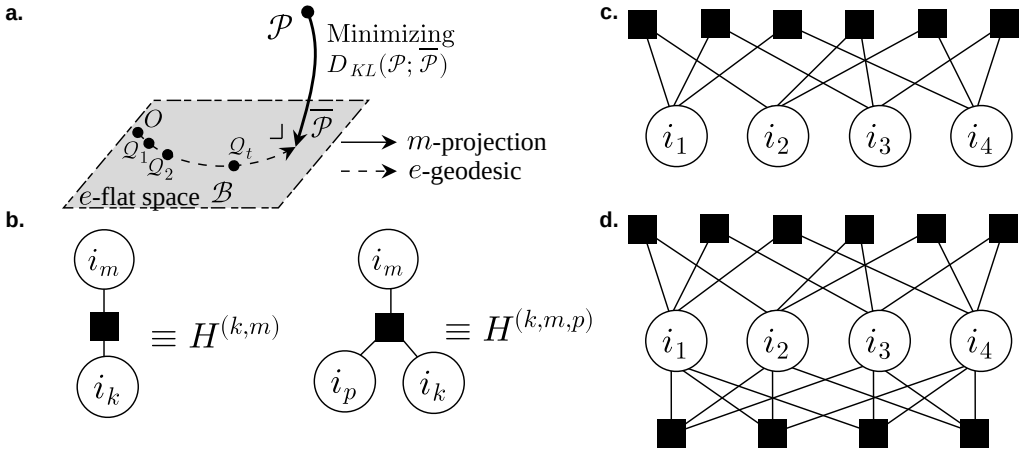

Figure 1: (a) An illustration of optimization of Legendre decomposition. Interaction representations corresponding to (c) equation 10 and (d) equation 11.

2.2 Interaction and its representation of tensors

In this subsection, we introduce interactions between modes and its visual representation to prepare for many-body approximation. The following discussion enables us to intuitively describe relationships between modes and formulate our novel rank-free tensor decomposition.

First we introduce n-body parameters, which is a generalized concept of one-body and two-body parameters in (Ghalamkari & Sugiyama, 2022). Let n of a n-body parameter be the number of non-one indices; for example, $\theta_{1,2,1,1}$ is a one-body parameter, $\theta_{4,3,1,1}$ is a two-body parameter and $\theta_{1,2,4,3}$ is a three-body parameter. We also use the following notation for n-body parameters:

$$\theta^{(k)}_{i_k} = \theta_{1,\dots,1,i_k,1,\dots,1}, \quad \theta^{(k,m)}_{i_k,i_m} = \theta_{1,\dots,1,i_k,1,\dots,1,i_m,1,\dots,1}, \quad \theta^{(k,m,p)}_{i_k,i_m,i_p} = \theta_{1,\dots,i_k,\dots,i_m,\dots,i_p,\dots,1}. \tag{7}$$

We write the energy function $\mathcal{H}$ with n-body parameters as

$$\mathcal{H}_{i_1,\cdots,i_D} = H_0 + \sum_{k=1}^{D} H^{(k)}_{i_k} + \sum_{m=1}^{k-1}\sum_{k=1}^{D} H^{(k,m)}_{i_k,i_m} + \sum_{p=1}^{m-1}\sum_{m=1}^{k-1}\sum_{k=1}^{D} H^{(k,m,p)}_{i_k,i_m,i_p} + \cdots + H^{(1,\dots,D)}_{i_1,\dots,i_D} \tag{8}$$

where the n-th order energy is introduced as

$$H^{(l_1,\dots,l_n)}_{i_{l_1},\dots,i_{l_n}} = \sum_{i'_{l_1}=2}^{i_{l_1}} \cdots \sum_{i'_{l_n}=2}^{i_{l_n}} \theta^{(l_1,\dots,l_n)}_{i'_{l_1},\dots,i'_{l_n}}. \tag{9}$$

For simplicity, we suppose that $1 \leq l_1 < l_2 < \cdots < l_n \leq D$ holds. We set $H_0 = \theta_{1,\dots,1}$. We say that an n-body interaction exists between modes $(l_1, \dots, l_n)$ when $H^{(l_1,\dots,l_n)}_{i_{l_1},\dots,i_{l_n}} \neq 0$.

The first term H_0 in equation 8 is called the normalized factor or the partition function. The terms $H^{(k)}$ are called bias in machine learning and magnetic field or self-energy in statistical physics. The terms $H^{(k,m)}$ are called the weight of the Boltzmann machine in machine learning and two-body interaction or electron-electron interaction in physics.

To visualize the existence of interactions within a tensor, we newly introduce a diagram called *interaction representation*, which is inspired by factor graphs in graphical modelling (Bishop & Nasrabadi, 2006, Chapter 8). The graphical representation of the product of tensors is widely known as tensor networks. However, displaying the relations between the modes of a tensor as a factor graph is our novel approach. We represent the n-body interaction as a black square, ■, connected with n modes. We describe examples of the two-body interaction between modes (k, m) and the three-body interaction among modes (k, m, p) in Figure 1(b). Combining these interactions, the energy function

including all two-body interactions is shown in Figure 1(c), and the energy function including all two-body and three-body interactions is shown in Figure 1(d) for $D = 4$.

This visualization allows us to intuitively understand the relationship between modes of tensors. For simplicity, we abbreviate one-body interactions in the diagrams, while we always assume them. Once interaction representation is given, we can determine the corresponding decomposition of tensors.

2.3 MANY-BODY APPROXIMATION

Our proposed method, tensor many-body approximation, approximate a given tensor with assuming the existence of dominant interactions between the modes of the tensor and ignoring other interactions. Since this operation can be understood as setting some natural parameters of the distribution to be zero, it can be achieved by convex optimization through the theory of Legendre decomposition. Moreover, we can represent any approximated tensor as an element-wise product of small order tensors.

As an example, we consider approximations of a nonnegative tensor $\mathcal{P}$ by tensors represented in Figure 1(c) and Figure 1(d). Substituting 0 for energies greater than 2nd and 3rd order in equation 8, $\mathcal{P}$ is approximated as follows:

$$\mathcal{P}_{i_1,i_2,i_3,i_4} \simeq \mathcal{P}^{\leq 2}_{i_1,i_2,i_3,i_4} = X^{(1,2)}_{i_1,i_2} X^{(1,3)}_{i_1,i_3} X^{(1,4)}_{i_1,i_4} X^{(2,3)}_{i_2,i_3} X^{(2,4)}_{i_2,i_4} X^{(3,4)}_{i_3,i_4}, \tag{10}$$

$$\mathcal{P}_{i_1,i_2,i_3,i_4} \simeq \mathcal{P}^{\leq 3}_{i_1,i_2,i_3,i_4} = \chi^{(1,2,3)}_{i_1,i_2,i_3} \chi^{(1,2,4)}_{i_1,i_2,i_4} \chi^{(1,3,4)}_{i_1,i_3,i_4} \chi^{(2,3,4)}_{i_2,i_3,i_4}, \tag{11}$$

where each factor on the right-hand side is represented as

$$X^{(k,m)}_{i_k,i_m} = \frac{1}{\sqrt[6]{Z}} \exp\left(\frac{1}{3} H^{(k)}_{i_k} + H^{(k,m)}_{i_k,i_m} + \frac{1}{3} H^{(m)}_{i_m}\right),$$

$$\chi^{(k,m,p)}_{i_k,i_m,i_p} = \frac{1}{\sqrt[4]{Z}} \exp\left(\frac{H^{(k)}_{i_k} + H^{(m)}_{i_m} + H^{(p)}_{i_p}}{3} + \frac{1}{2} H^{(k,m)}_{i_k,i_m} + \frac{1}{2} H^{(m,p)}_{i_m,i_p} + \frac{1}{2} H^{(k,p)}_{i_k,i_p} + H^{(k,m,p)}_{i_k,i_m,i_p}\right).$$

The partition function is given as $Z = \exp(-\theta_{1,\ldots,1}/6)$ and $Z = \exp(-\theta_{1,\ldots,1}/4)$, respectively, which do not depend on indices (i_1, i_2, i_3, i_4). In these examples, $\mathcal{P}$ is approximated by the element-wise product of an n-order tensor for $n = 2$ and $n = 3$.

In the above discussion, we consider many-body approximation with all n-body parameters, while our formulation allows us to use only a part of n-body interactions as shown in the following. We consider the situation where only one-body interaction and two-body interaction between $(k, k+1)$ exist for all $k \in \{1, \ldots, D\}$ ($D + 1$ implies 1 for simplicity). Figure 2(a) shows the interaction representation of the approximated tensor. As we can confirm by substituting 0 for $H^{(k,l)}_{i_k,i_l}$ if $l \neq k+1$, we can describe the approximated tensor as the element-wise cyclic product of matrices,

$$\mathcal{P}_{i_1,\ldots,i_D} \simeq \mathcal{P}^{\text{cyc}}_{i_1,\ldots,i_D} = X^{(1)}_{i_1,i_2} X^{(2)}_{i_2,i_3} \ldots X^{(D)}_{i_D,i_1} \tag{12}$$

where

$$X^{(k)}_{i_k,i_{k+1}} = \frac{1}{\sqrt[D]{Z}} \exp\left(\frac{1}{2} H^{(k)}_{i_k} + H^{(k,k+1)}_{i_k,i_{k+1}} + \frac{1}{2} H^{(k+1)}_{i_{k+1}}\right). \tag{13}$$

The partition function is given as $Z = \exp(-\theta_{1,\ldots,1}/D)$, which does not depend on indices $(i_1, \ldots, i_k)$. We call this approximation cyclic two-body approximation.

We show the connection between cyclic two-body approximation and existing tensor ring decomposition in the following subsection.

2.4 CONNECTION TO TENSOR NETWORK

Tensor interaction representation is a diagram that focuses on the relationship between modes. Tensor networks, which are well known as diagrams that focus on factors after decomposition, represent a tensor as an undirected graph, whose nodes correspond to matrices or tensors and edges are the modes of summation in tensor products (Cichocki et al., 2016).

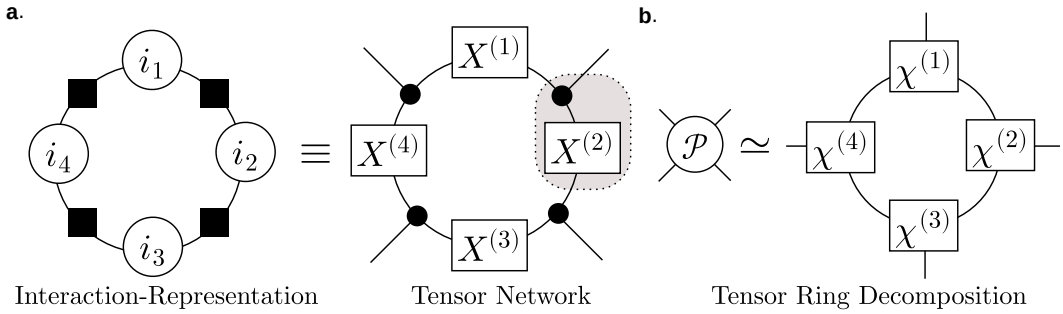

Figure 2: (a) Interaction representation of an example of cyclic two-body approximation and its transformed tensor network for $D = 4$. (b) Tensor network of tensor ring decomposition.

Our tensor interaction representation has a tight connection to tensor networks, and we can convert a tensor interaction representation to a tensor network. For the conversion, we use a hyper-diagonal tensor Ω, that is defined as $\Omega_{ijk} = \delta_{ij}\delta_{jk}\delta_{ki}$, where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise. The tensor Ω is often represented by $\bullet$ in tensor networks. In the community of tensor network, the tensor Ω appears in the CNOT gate and a special case of Z spider (Nielsen & Chuang, 2010). The tensor network in Figure 2(a) represents the following formula

$$\prod_{d=1}^{D} \left(\sum_{j_d} \sum_{l_d} X^{(d)}_{l_d, j_{d+1}} \Omega_{j_{d+1}, i_{d+1}, l_{d+1}} \right), \tag{14}$$

where $j_{D+1} = j_1, i_{D+1} = i_1, l_{D+1} = l_1$. Substituting the definition of Ω in equation 14, we realize that the tensor network corresponds to equation 12.

We point out that the tensor network representation of cyclic two-body approximation is similar to the tensor network of the tensor ring decomposition. The tensor ring decomposition is an extension of the tensor train decomposition, and its representation is shown in Figure 2(b) using a tensor network. In fact, if we consider the region enclosed by the dotted line in the tensor network as a new tensor, the tensor network of the cyclic two-body approximation coincides with the tensor network of the tensor ring decomposition. This operation, in which multiple tensors are regarded as a new tensor in a tensor network, is called renormalization or coarse-graining transformation (Evenbly & Vidal, 2015).

Comparing the number of parameters The number of elements of an input tensor is $I_1 \times I_2 \times \cdots \times I_D$. After the cyclic two-body approximation, the number of parameters is given as

$$|\mathcal{B}| = 1 + \sum_{k=1}^{d} (I_k - 1) + \sum_{k=1}^{d} (I_k - 1)(I_{k+1} - 1). \tag{15}$$

The first term is for a normalizer, the second is the number of one-body parameters, and the final term is the number of two-body parameters. In contrast, in the tensor ring decomposition with the target rank $(R_1, \ldots, R_D)$, the number of parameters is given as $|\mathcal{R}| = \sum_{k=1}^{d} R_k I_k R_{k+1}$. The ratio of the number of parameters of these two methods $|\mathcal{B}|/|\mathcal{R}|$ is proportional to dI/R^2 if we assume $R_k = R$ and $I_k = I$ for all $k \in \{1, \ldots, D\}$ for simplicity. Therefore, when the target rank is small and the size of the input tensor is large, the proposed method has more parameters than the tensor ring decomposition.

2.4.1 Other example of many-body approximation and its tensor network

In the same way, we can find a correspondence between another example of many-body approximation and the existing low-rank approximation. For $D = 9$, we consider three-body and two-body interactions among (i_1, i_2, i_3), (i_4, i_5, i_6), and (i_7, i_8, i_9) and three-body approximation among (i_3, i_6, i_9). We provide the interaction representation of the target energy function in Figure 3(a). In this approximation, the decomposed tensor can be described as

$$\mathcal{P}_{i_1, \ldots, i_9} = \mathcal{A}_{i_1, i_2, i_3} \mathcal{B}_{i_4, i_5, i_6} \mathcal{C}_{i_7, i_8, i_9} \mathcal{G}_{i_3, i_6, i_9}. \tag{16}$$

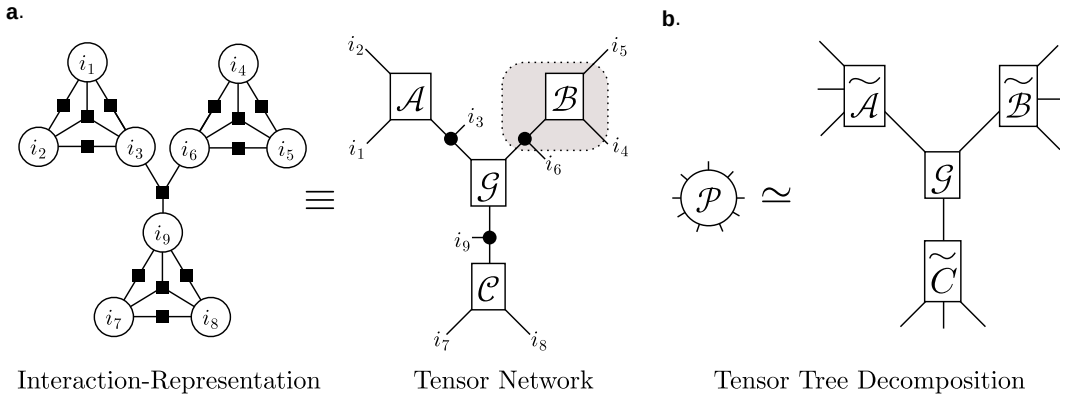

Figure 3: (a) Interaction representation corresponding to equation 16 and its transformed tensor network for $D = 9$. (b) Tensor network of a variant of tensor tree decomposition.

In the same way in the case of the cyclic two-body approximation, we can convert the interaction representation to a tensor network, as described in Figure 3(a). A tensor network of tensor tree decomposition in Figure 3(b) emerges when the region enclosed by the dotted line is replaced with a new tensor (shown with tilde) in Figure 3(a). Such tensor tree decomposition is used in generative modeling (Cheng et al., 2019), computational chemistry (Murg et al., 2015) and quantum many-body physics (Shi et al., 2006).

As we have seen above, by transforming tensor interaction representation to tensor networks and applying coarse-graining, we can reveal the relationship between tensor many-body approximations and low-rank approximations.

2.5 Many-body approximation as generalization of mean-field approximation

It has been already pointed out that any tensor $\mathcal{P}$ can be represented by vectors $\boldsymbol{x}^{(d)} \in \mathbb{R}^{I_d}$ for $d \in \{1, \ldots, D\}$ as

$$\mathcal{P}_{i_1,\ldots,i_D} = x^{(1)}_{i_1} x^{(2)}_{i_2} \ldots x^{(D)}_{i_D} \tag{17}$$

if and only if all $n(\geq 2)$-body θ-parameters are 0 (Ghalamkari & Sugiyama, 2021). The right-hand side is equal to the Kronecker product of D vectors $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(D)}$, and therefore this approximation is equivalent to the rank-1 approximation since the rank of the tensor that can be represented by the Kronecker product is always 1, which is also known to correspond to mean-field approximation. In this study, we propose many-body approximation by relaxing the condition for the mean-field approximation that ignores $n(\geq 2)$-body interactions. Therefore many-body approximation is generalization of rank-1 approximation and mean-field approximation.

2.6 Computational Complexity

We analyze the computational complexity of many-body approximation. The complexity of computing the inverse of an $n \times n$ matrix is $O(n^3)$. Therefore, the computational complexity of many-body approximation is $\mathcal{O}(\gamma|\mathcal{B}|^3)$, where γ is the number of iterations.

This complexity can be reduced if we reshape tensors so that the size of each mode becomes small. For example, let us consider a 3-order tensor whose size is (J^2, J^2, J^2) and its cyclic two-body approximation. In this case, the time complexity is $\mathcal{O}(\gamma J^{12})$ since it holds that $|\mathcal{B}| \propto J^4$ (See equation 15). In contrast, if we reshape the input tensor to a 6-order tensor whose size is (J, J, J, J, J, J), the time complexity becomes $\mathcal{O}(\gamma J^6)$ since it holds that $|\mathcal{B}| \propto J^2$.

This technique of reshaping a tensor into a larger-order tensor is used practically not only in the proposed method but also in various methods based on tensor networks, such as tensor ring decomposition (Malik & Becker, 2021).

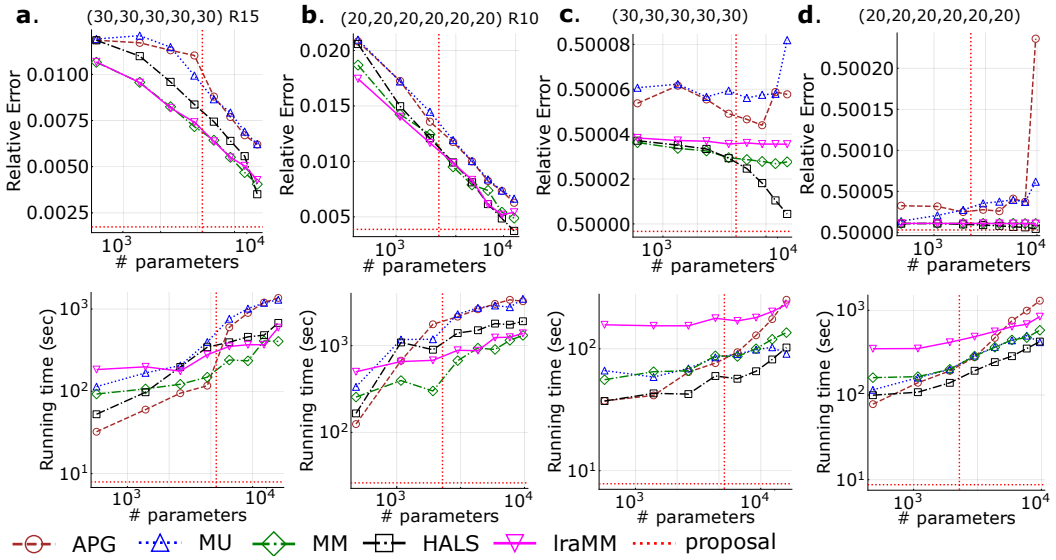

Figure 4: (a)(b) Results for low ring rank tensor. (c)(d) Results for tensors sampled from uniform distribution.

3 Experiments

As seen in Section 2.4, many-body decomposition has a close connection to low-rank approximation. For example, in a tensor ring decomposition, if we impose that decomposed factors can be represented as products with hyper-diagonal tensors Ω, this decomposition is equivalent to a cyclic two-body approximation (see Figure 2). Therefore, to examine our conjecture that cyclic two-body approximation is as capable of approximating as tensor ring decomposition, we empirically examine the efficiency and effectiveness of cyclic two-body approximation compared with tensor ring decomposition. As baselines, we use five existing methods of non-negative tensor ring decomposition, NTR-APG, NTR-HALS, NTR-MU, NTR-MM and NTR-lraMM (Yu et al., 2021; 2022). These methods minimize the reconstruction error defined with the Frobenius norm by the gradient method.

Our method is implemented in Julia 1.8. We implemented these methods by translating MATLAB code provided by the authors into Julia code for fair comparison. Experiments were conducted on Ubuntu 20.04.1 with a single core of 2.1GHz Intel Xeon CPU Gold 5218 and 128GB of memory.

We evaluate the approximation performance by the relative error

$$\frac{\|\mathcal{T} - \overline{\mathcal{T}}\|_F}{\|\mathcal{T}\|_F} \tag{18}$$

for an input tensor $\mathcal{T}$ and a reconstracted tensor $\mathcal{T}$ with the Frobenius norms $\|\cdot\|_F$. Since all the existing methods are based on nonconvex optimization, we plot the best score (minimum relative error) among 5 restarts with random initialization. In contrast, the score of our method is obtained by a single run as it is convex optimization and such restarts are fundamentally unnecessary. We compare the total running time of them.

3.1 Synthetic Data

We performed experiments on four synthetic datasets. The first two are synthetic data with low tensor ring rank. This setting is often used in evaluation of tensor ring decomposition. We create D core tensors of size $R \times I \times R$ by sampling from uniform distribution. Then a tensor with the size I^D and the tensor ring rank $(R, \ldots, R)$ is obtained by the product of these D tensors. Results for $R = 15, D = 5, I = 30$ are shown in Figure 4(a)., and those for $R = 10, D = 6, I = 20$ in Figure 4(b). Relative error and computation time are plotted with gradually increasing the target rank of the tensor ring decomposition, which is compared to the score of our method, plotted as the cross point of horizontal and vertical red dotted lines. Please note that our method does not have the

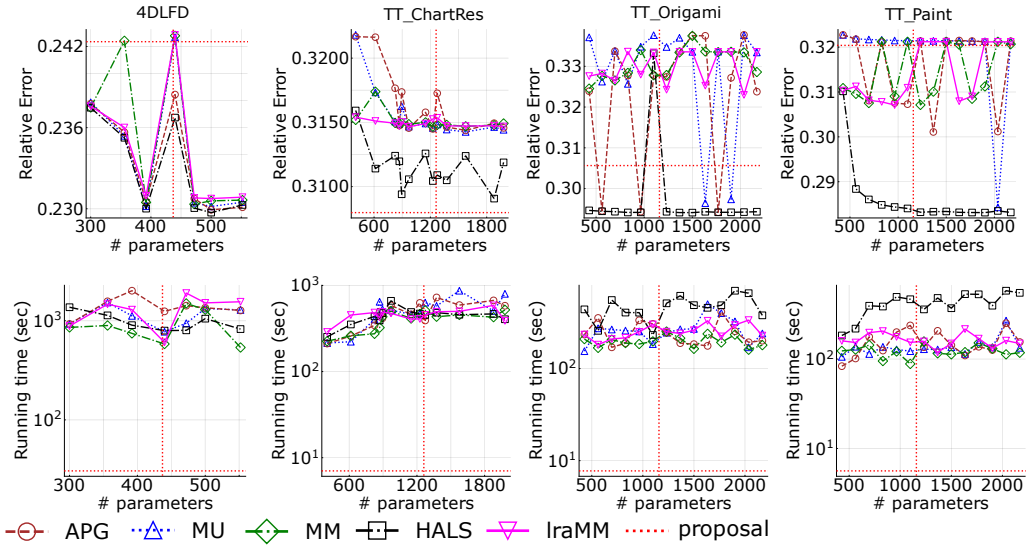

Figure 5: Experimental results for real datasets.

concept of the rank, thus the score of our method is invariant to changes of the target rank unlike other methods. If the cross point of red dotted lines is lower than other lines, the proposed method is better than other methods.

In addition to the above case in which we assumed the low-rankness, we also generated synthetic datasets without such an assumption. We created a tensor of size 30^5 and a tensor of size 20^5 by sampling from uniform distribution and performed the same experiment. Results are shown in Figure 4(c) and Figure 4(d). In all experiments, the proposed method is superior to comparison partners in both efficiency and effectiveness. It should be noted that the relative error of the proposed method is smaller even when the target rank of the tensor ring decomposition is large and the number of parameters is several times larger than the proposed method.

3.2 Real Data

Next, we evaluate our method on real data. `4DLFD` is a 9-order tensor, which is produced from 4D Light Field Dataset described in (Honauer et al., 2016). `TT_ChartRes`, `TT_Origami` and `TT_Paint` are 7-order tensors, which is produced from TokyoTech Hyperspectral Image Dataset (Monno et al., 2015; 2017). Each tensor has been reshaped to reduce the computational complexity. See the dataset details in the Supplemental Materials. The proposed method is always faster than baselines with keeping the competitive relative errors.

4 Conclusion

We propose *many-body approximation* for tensors, which decomposes tensors with focusing on the relationship between modes represented by an energy-based model. It approximates tensors by ignoring the energy corresponding to some interactions, which can be viewed as generalization of mean-field approximation that considers only one-body interactions. Our novel formulation enables us to achieve convex optimization of the model, while the existing approaches based on the low-rank structure are non-convex. Furthermore, we introduce a way of visualize interactions between modes, called interaction representation, to see activated interactions between modes. We have established transformation between our representation and tensor networks, which reveals the nontrivial connection between many-body approximation and the classical tensor low-rank tensor decomposition.

REPRODUCIBILITY STATEMENT

We provide implementation and dataset details in Supplemental Material. We provide both of code for the proposed method and code for comparison methods in Supplemental Material.

ETHICS STATEMENT

This study is theoretical analysis of tensors and we believe that our theoretical discussion would not have any negative societal impacts.

REFERENCES

S. Amari. *Information Geometry and Its Applications*. Springer, 2016.

Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

Song Cheng, Lei Wang, Tao Xiang, and Pan Zhang. Tree tensor networks for generative modeling. *Physical Review B*, 99(15):155131, 2019.

Andrzej Cichocki, Danilo Mandic, Lieven De Lathauwer, Guoxu Zhou, Qibin Zhao, Cesar Caiafa, and Huy Anh Phan. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE signal processing magazine*, 32(2):145–163, 2015.

Andrzej Cichocki, Namgil Lee, Ivan Oseledets, Anh-Huy Phan, Qibin Zhao, Danilo P Mandic, et al. Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions. *Foundations and Trends® in Machine Learning*, 9(4-5):249–429, 2016.

Aybüke Erol and Borbála Hunyadi. Tensors for neuroimaging: A review on applications of tensors to unravel the mysteries of the brain. *Tensors for Data Processing*, pp. 427–482, 2022.

Glen Evenbly and Guifre Vidal. Tensor network renormalization. *Physical review letters*, 115(18): 180405, 2015.

K. Ghalamkari and M. Sugiyama. Fast tucker rank reduction for non-negative tensors using mean-field approximation. In *Advances in Neural Information Processing Systems*, volume 34, pp. 443–454, Virtual Event, December 2021.

K. Ghalamkari and M. Sugiyama. Fast rank-1 NMF for missing data with KL divergence. In *Proceedings of the 25th International Conference on Artificial Intelligence and Statistics*, pp. 2927–2940, Virtual Event, March 2022.

Frank L Hitchcock. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, 6(1-4):164–189, 1927.

Katrin Honauer, Ole Johannsen, Daniel Kondermann, and Bastian Goldluecke. A dataset and evaluation methodology for depth estimation on 4d light fields. In *Asian Conference on Computer Vision*, pp. 19–34. Springer, 2016.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM review*, 51(3): 455–500, 2009.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Michael Levin and Cody P Nave. Tensor renormalization group approach to two-dimensional classical lattice models. *Physical review letters*, 99(12):120601, 2007.

Yuan Luo, Fei Wang, and Peter Szolovits. Tensor factorization toward precision medicine. *Briefings in bioinformatics*, 18(3):511–514, 2017.

Osman Asif Malik and Stephen Becker. A sampling-based method for tensor ring decomposition. In *International Conference on Machine Learning*, pp. 7400–7411. PMLR, 2021.

Yusuke Monno, Daisuke Kiku, Masayuki Tanaka, and Masatoshi Okutomi. Adaptive residual interpolation for color and multispectral image demosaicking. *Sensors*, 17(12):2787, 2017.

Yusukex Monno, Sunao Kikuchi, Masayuki Tanaka, and Masatoshi Okutomi. A practical one-shot multispectral imaging system using a single image sensor. *IEEE Transactions on Image Processing*, 24(10):3048–3059, 2015.

Valentin Murg, Frank Verstraete, Örs Legeza, and Reinhard M Noack. Simulating strongly correlated quantum systems with tree tensor networks. *Physical Review B*, 82(20):205105, 2010.

Valentin Murg, Frank Verstraete, Reinhold Schneider, Peter R Nagy, and O Legeza. Tree tensor network state with variable tensor order: An efficient multireference method for strongly correlated systems. *Journal of Chemical Theory and Computation*, 11(3):1027–1036, 2015.

Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. doi: 10.1017/CBO9780511976667.

Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5): 2295–2317, 2011.

Yannis Panagakis, Jean Kossaifi, Grigorios G Chrysos, James Oldfield, Mihalis A Nicolaou, Anima Anandkumar, and Stefanos Zafeiriou. Tensor methods in computer vision and deep learning. *Proceedings of the IEEE*, 109(5):863–890, 2021.

Y-Y Shi, L-M Duan, and Guifre Vidal. Classical simulation of quantum many-body systems with a tree tensor network. *Physical review a*, 74(2):022320, 2006.

M. Sugiyama, H. Nakahara, and K. Tsuda. Information decomposition on structured space. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 575–579, Barcelona, Spain, July 2016.

M. Sugiyama, H. Nakahara, and K. Tsuda. Tensor balancing on statistical manifold. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3270–3279, Sydney, Australia, August 2017.

M. Sugiyama, H. Nakahara, and K. Tsuda. Legendre decomposition for tensors. In *Advances in Neural Information Processing Systems 31*, pp. 8825–8835, Montréal, Canada, December 2018.

Ledyard R Tucker. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 31(3): 279–311, 1966.

YuYuan Yu, Kan Xie, Jinshi Yu, Qi Jiang, and Shengli Xie. Fast nonnegative tensor ring decomposition based on the modulus method and low-rank approximation. *Science China Technological Sciences*, 64(9):1843–1853, 2021.

Yuyuan Yu, Guoxu Zhou, Ning Zheng, Yuning Qiu, Shengli Xie, and Qibin Zhao. Graph-regularized non-negative tensor-ring decomposition for multiway representation learning. *IEEE Transactions on Cybernetics*, 2022.

Qibin Zhao, Guoxu Zhou, Shengli Xie, Liqing Zhang, and Andrzej Cichocki. Tensor ring decomposition. *arXiv preprint arXiv:1606.05535*, 2016.

A Implementation detail

We describe the implementation details of methods in the following.

Baseline methods As we can see from their original papers, NTR-APG, NTR-HALS, NTR-MU, NTR-MM and NTR-lraMM have an inner and outer loop to find a local solution. We repeat the inner loop 100 times. We stop the outer loop when the difference between the relative error of the previous and the current iteration is less than 10e-4. NTR-MM and NTR-lraMM require diagonal parameters matrix Ξ. We define $\Xi = \omega I$ where I is an identical matrix and $\omega = 0.1$. The NTR-lraMM method performs low-rank approximation to the matrix obtained by mode expansion of an input tensor. The target rank is set to be 20. This setting is the default setting in the provided code. The initial positions of baseline methods were sampled from uniform distribution on $(0, 1)$.

Proposed method We follow Algorithm 2 in Sugiyama et al. (2018). For non-normalized tensors, we conduct the following procedure. First, we compute the total sum of elements of an input tensor. Then, we normalize the tensor. After that, we conduct Legendre decomposition for the normalized tensor. Finally, we get the product of the result of the previous step and the total sum we compute initially.

B DATASET DETAIL

We describe the details of each dataset in the following.

Synthetic Datasets For all experiments on synthetic datasets, we change the target ring-rank as $(r, \ldots, r)$ for $r = 2, 3, \ldots, 9$.

Real Datasets `4DLFD` is originally a $(9, 9, 512, 512, 3)$ tensor, which is produced from 4D Light Field Dataset described in Honauer et al. (2016). Its license is Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. We use `dino` images and their depth and disparity map in training scenes. We concatenate them to produce a tensor. We reshaped the tensor as $(6, 8, 6, 8, 6, 8, 6, 8, 12)$. We chose the target ring-rank as $(2, 3, 2, 2, 2, 2, 2, 2, 2)$, $(2, 3, 2, 2, 3, 2, 2, 3, 2)$, $(2, 2, 2, 2, 2, 2, 2, 2, 5)$, $(2, 5, 2, 2, 5, 2, 2, 2, 2)$, $(2, 2, 2, 2, 2, 2, 2, 2, 7)$, $(2, 2, 2, 2, 3, 2, 2, 2, 7)$, $(2, 2, 2, 2, 2, 2, 2, 2, 9)$. `TT_ChartRes` is originally a $(736, 736, 31)$ tensor, which is produced from TokyoTech 31-band Hyperspectral Image Dataset. We use `ChartRes.mat`. We reshaped the tensor as $(23, 8, 4, 23, 8, 4, 31)$. We chose the target ring-rank as $(2, 2, 2, 2, 2, 2, 2)$ $(2, 2, 2, 2, 2, 2, 5)$, $(2, 2, 2, 2, 2, 2, 8)$, $(3, 2, 2, 3, 2, 2, 5)$, $(2, 2, 2, 2, 2, 2, 9)$, $(3, 2, 2, 3, 2, 2, 6)$, $(4, 2, 2, 2, 2, 2, 6)$, $(3, 2, 2, 4, 2, 2, 8)$, $(3, 2, 2, 3, 2, 2, 9)$, $(3, 2, 2, 3, 2, 2, 10)$, $(3, 2, 2, 3, 2, 2, 12)$, $(3, 2, 2, 3, 2, 2, 15)$, $(3, 2, 2, 3, 2, 2, 16)$. `TT_Origami` and `TT_Paint` are originally $(512, 512, 59)$ tensors, which are produced from TokyoTech 59-band Hyperspectral Image Dataset. We use `Origami.mat` and `Paint.mat`. In `TT_Origami`, 0.0016% of elements were negative, hence we preprocessed all elements of `TT_Origami` by subtracting -0.000764, the smallest value in `TT_Origami`, to make all elements non-negative. We reshaped the tensor as $(8, 8, 8, 8, 8, 8, 59)$. We chose the target ring-rank as $(2, 2, 2, 2, 2, 2, r)$ for $r = 2, 3, \ldots, 15$.