Fast Tucker Rank Reduction for Non-Negative Tensors Using Mean-Field Approximation

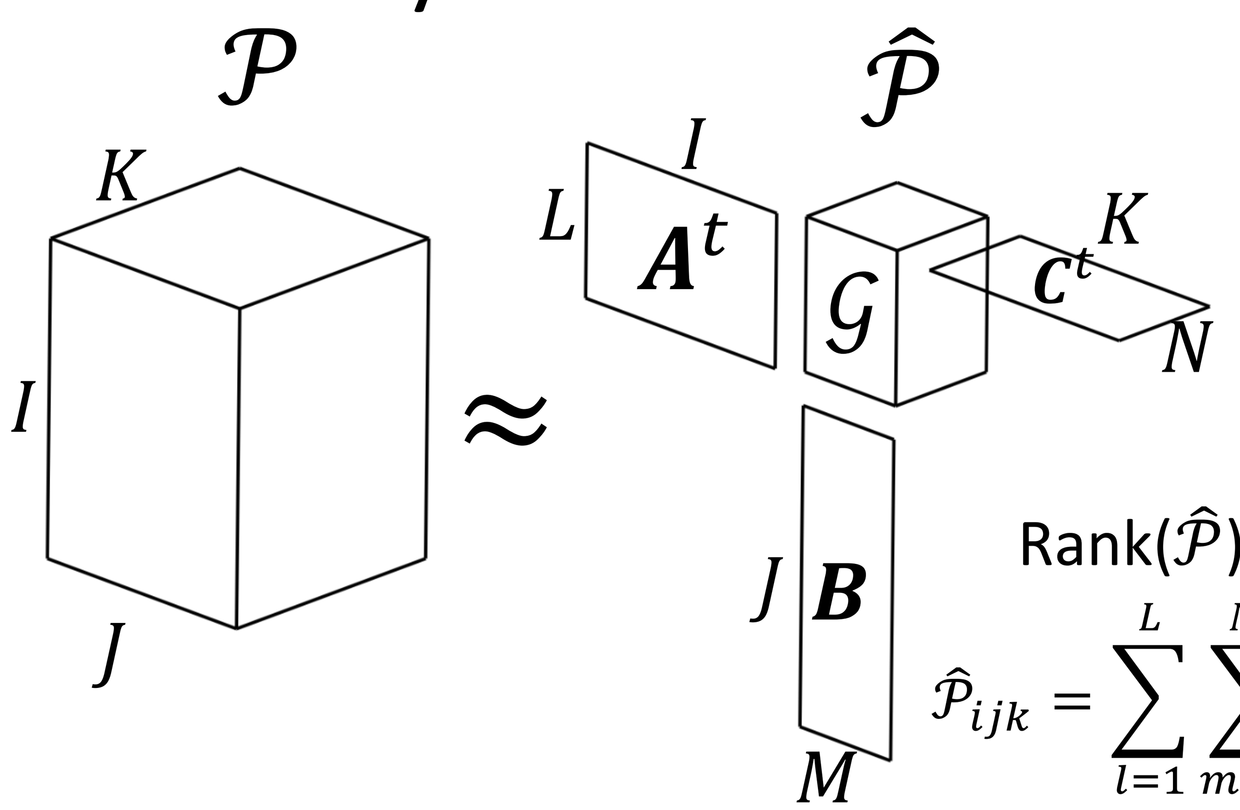
Kazu Ghalamkari[1,2], **Mahito Sugiyama**[1,2]

[1]National Institute of Infomatics @ Tokyo, Japan [2]The Graduate University for Advanced Studies, SOKENDAI

Inter-University Research Institute Corporation / Research Organization of Information and Systems
National Institute of Informatics

The 34th Annual Conference on Neural Information Processing Systems (NeurIPS 2021), December 6–14, 2021

Summary



Rank($\hat{\mathcal{P}}$)= (L, M, N)

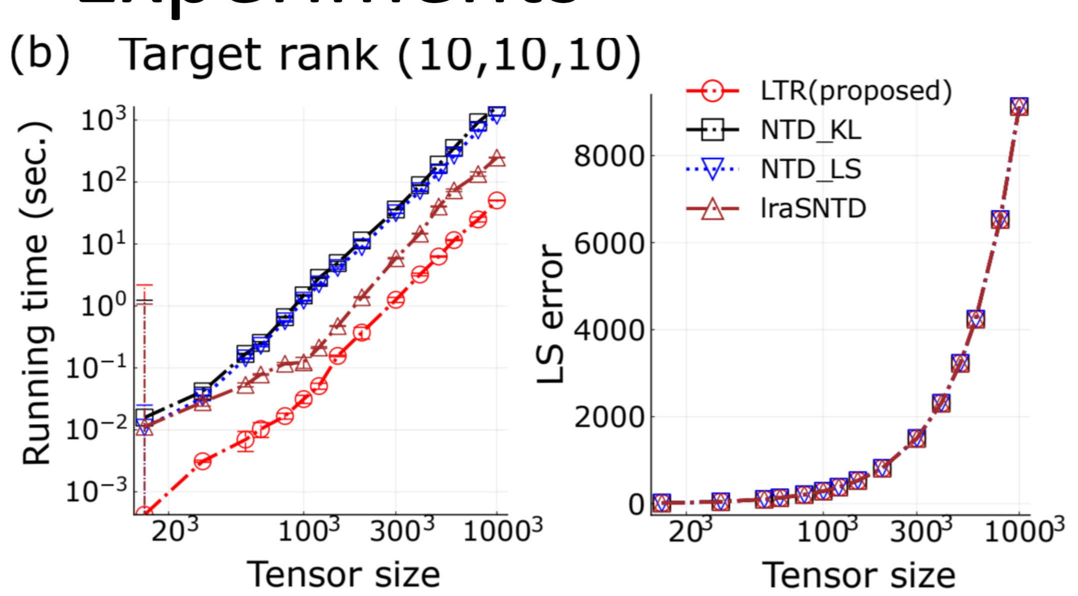
$$\hat{\mathcal{P}}_{ijk} = \sum_{l=1}^{L}\sum_{m=1}^{M}\sum_{n=1}^{N} \mathcal{G}_{lmn} A_{il} B_{jm} C_{kn}$$

- Low-rank tensors reduce memory requirements.
- Many non-negative low-rank approximation methods are based on a **gradient method**.

  ⇒ Initial values, stopping criterion, learning rate…

We developed fast low-rank approximation **without gradient method**, called **Legendre Rank Reduction** (LTR).

Experiments



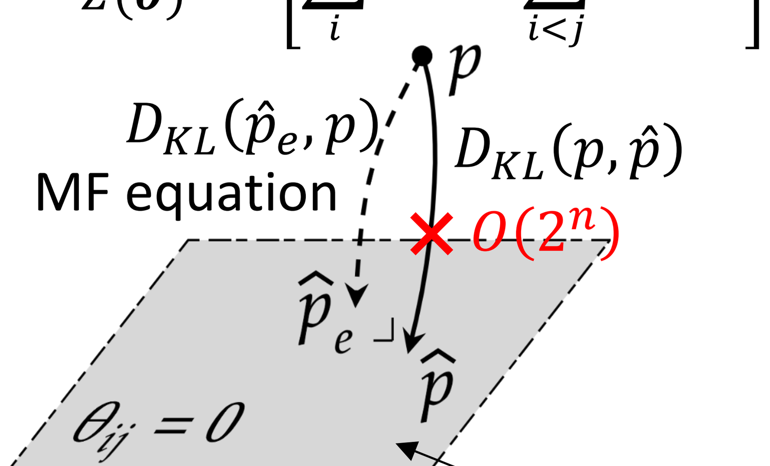
LTR is Faster.

LTR has Competitive error.

Rank-1 approximation as a Mean field approximation

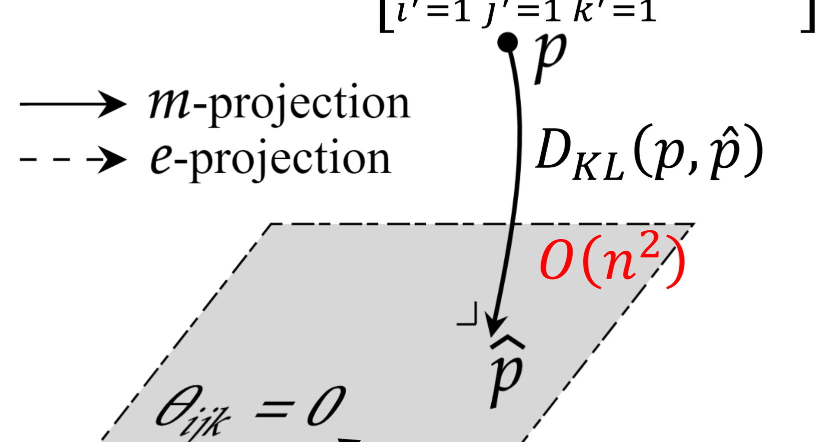
Mean field approximation of BM

$$p(\boldsymbol{x}) = \frac{1}{Z(\boldsymbol{\theta})}\exp\left[\sum_i \theta_i x_i + \sum_{i<j}\theta_{ij}x_i x_j\right]$$



Rank-1 approximation

$$p_\theta(i,j,k) = \exp\left[\sum_{i'=1}^{i}\sum_{j'=1}^{j}\sum_{k'=1}^{k}\theta_{i'j'k'}\right]$$



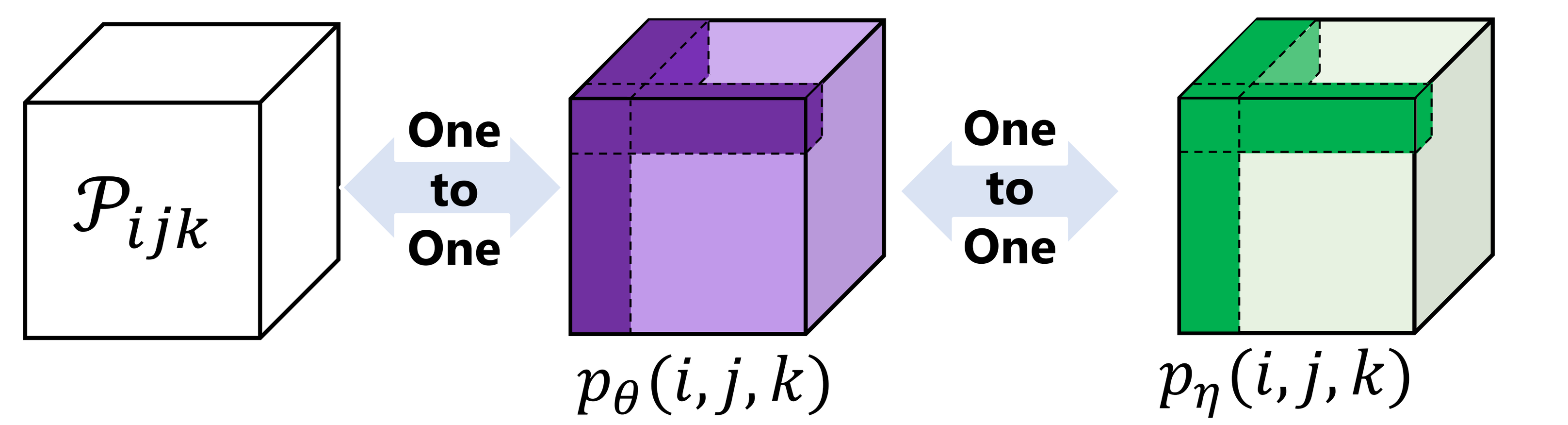
Set of product of **independent distributions**

Theory

■ Tensor as distribution

$$p_\theta(i,j,k) = \exp\left[\sum_{i'=1}^{i}\sum_{j'=1}^{j}\sum_{k'=1}^{k}\theta_{i'j'k'}\right] \qquad \eta_{ijk} = \sum_{i'=i}^{I}\sum_{j'=j}^{J}\sum_{k'=k}^{K}\mathcal{P}_{i'j'k'}$$

Index set is a sample space



one-body parameters: $\theta_{i11}, \theta_{1j1}, \theta_{11k}, \eta_{i11}, \eta_{1j1}, \eta_{11k}$. Only one index is 1.

many-body parameters: Other than a one-body parameter

■ Describe rank-*1* condition using $(\boldsymbol{\theta}, \boldsymbol{\eta})$

Rank-1 condition ($\boldsymbol{\eta}$-representation)

$$\mathrm{rank}(\mathcal{P}) = 1 \Leftrightarrow \eta_{ijk} = \eta_{i11}\eta_{1j1}\eta_{11k}$$

Rank-1 condition ($\boldsymbol{\theta}$-representation)

$\mathrm{rank}(\mathcal{P}) = 1 \Leftrightarrow$ its all many–body θ parameters are 0

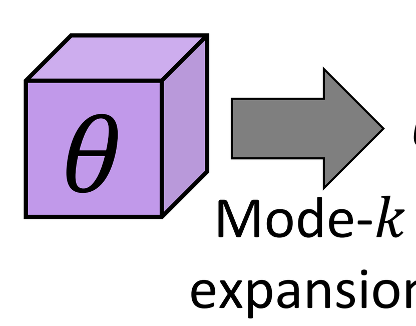
Best rank-1 tensor formula for minimizing KL divergence

For any given positive tensor $\mathcal{P}$, its best rank-1 approximation is

$$\hat{\mathcal{P}}_{ijk} = \left(\sum_{j'=1}^{J}\sum_{k'=1}^{K}\mathcal{P}_{ij'k'}\right)\left(\sum_{k'=1}^{K}\sum_{i'=1}^{I}\mathcal{P}_{i'jk'}\right)\left(\sum_{i'=1}^{I}\sum_{j'=1}^{J}\mathcal{P}_{i'j'k}\right).$$

K.Huang, et al. "Kullback-Leibler principal component for tensors is not NP-hard." *ACSSC* 2017

■ Describe rank-*r* condition using $(\boldsymbol{\theta}, \boldsymbol{\eta})$



Scaling Factor

$$\theta^{(1)} = \begin{bmatrix} \theta_{111} & \theta_{121} & \theta_{131} & \theta_{112} & 0 & 0 & \theta_{113} & 0 & 0 \\ \theta_{211} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{311} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Two bingos

$$\theta^{(2)} = \begin{bmatrix} \theta_{111} & \theta_{211} & \theta_{311} & \theta_{112} & 0 & 0 & \theta_{311} & 0 & 0 \\ \theta_{121} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_{131} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Two bingos

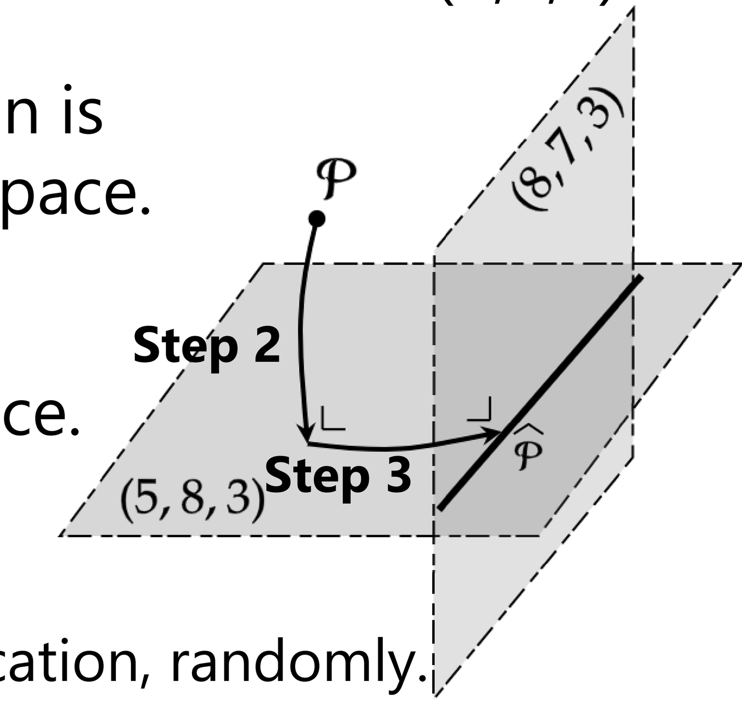
Bingo Rule

The mode-k expansion $\theta^{(k)}$ of the natural parameter has b_k bingos
⇒ $\mathrm{rank}(\mathcal{P}) \leq (I - b_1, J - b_2, K - b_3)$

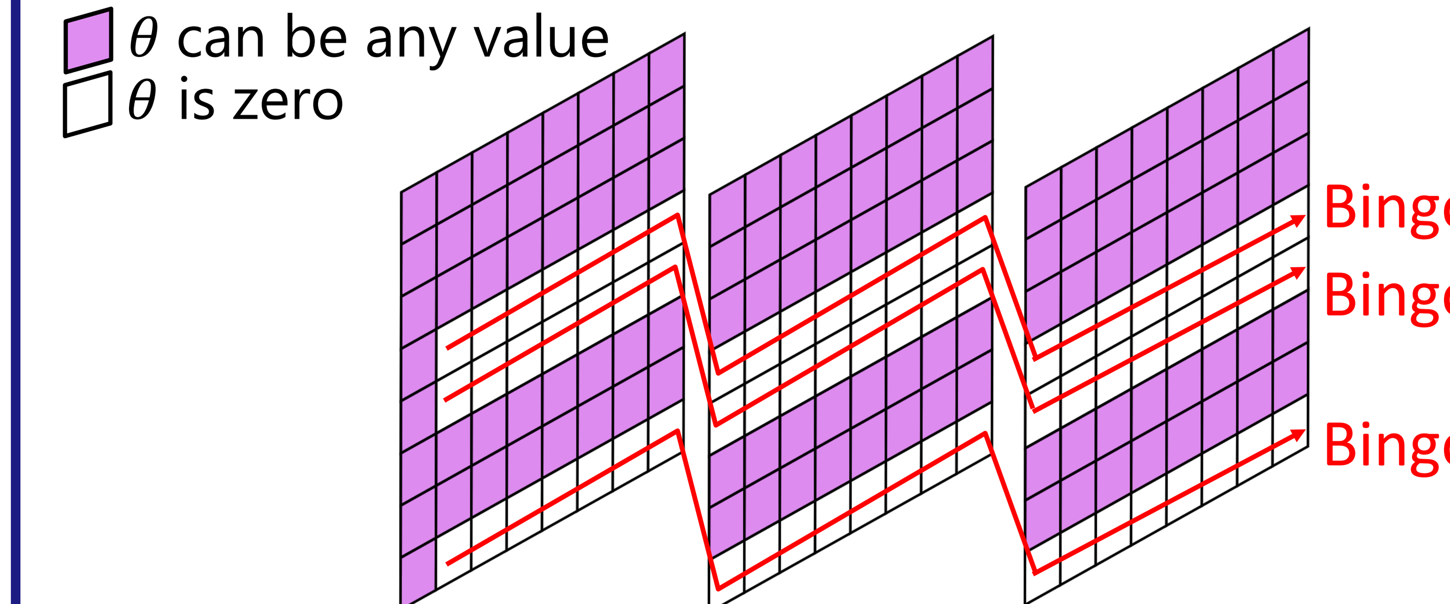
Legendre Rank Reduction

Let us reduce a (8,8,3)-tensors rank to (**5**,**7**,3).

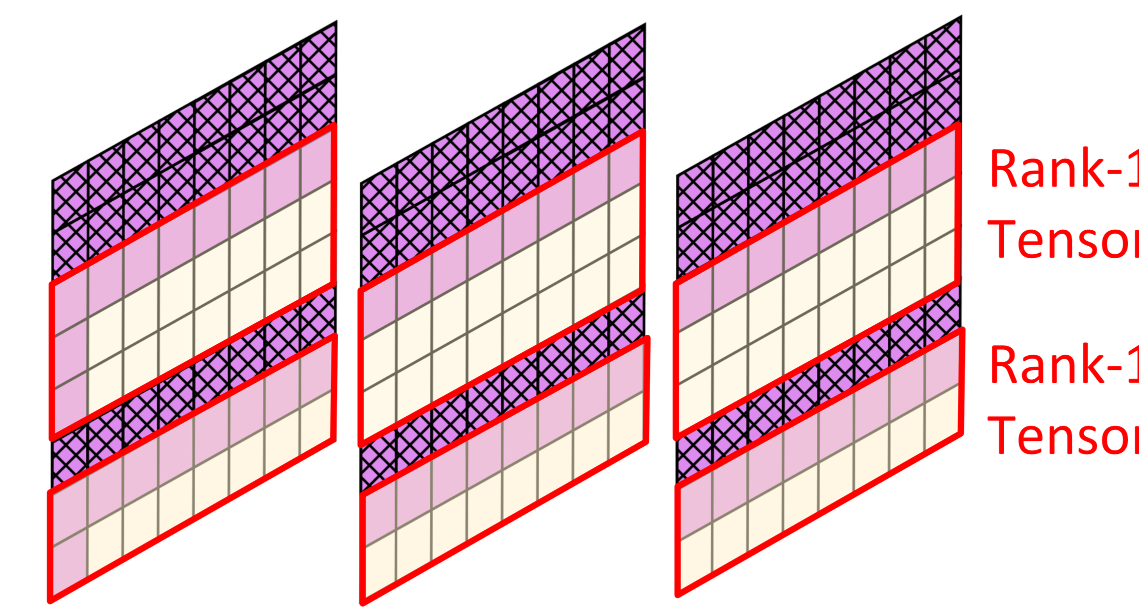
- Low-rank approximation is a projection onto Bingo space.
- $\hat{\mathcal{P}}$ is the best tensor in the specified bingo space.



Step 1 : Choose a bingo location, randomly.



Step 2 : Replace the bingo part with the rank-1 tensor.



The shaded areas do not change their values in the projection.

Step 3 : Do Step1 and Step2 in other axes.