# Many-body Approximation for Non-negative Tensors

Kazu Ghalamkari

RIKEN AIP

Mahito Sugiyama

National Institute of Informatics
SOKENDAI

Yoshinobu Kawahara

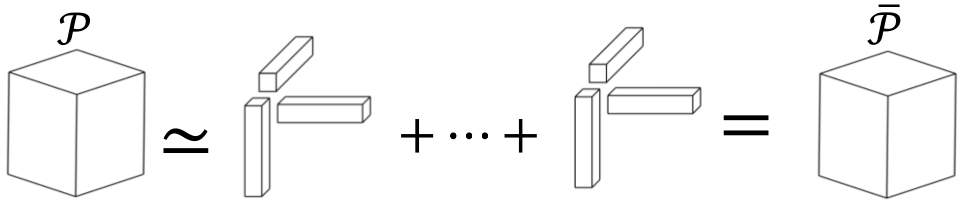
RIKEN AIP
Osaka University

The 37th Annual Conference on Neural Information Processing Systems (NeurIPS 2023), New Orleans in USA, December 11–14, 2023

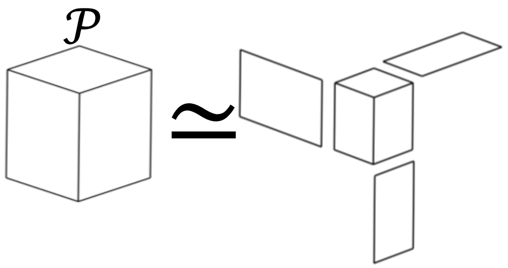
# Difficulties in tensor factorization

## ■ Model selection is not intuitive.

CP decomp.

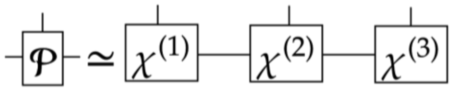



Tucker decomp.

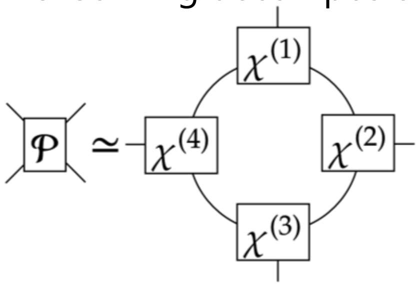



Decomp. with tensor networks

Tensor Train decomposion

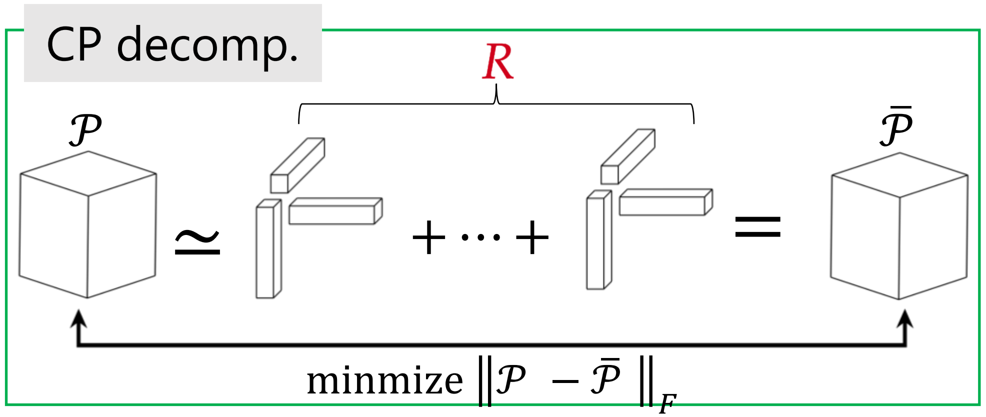


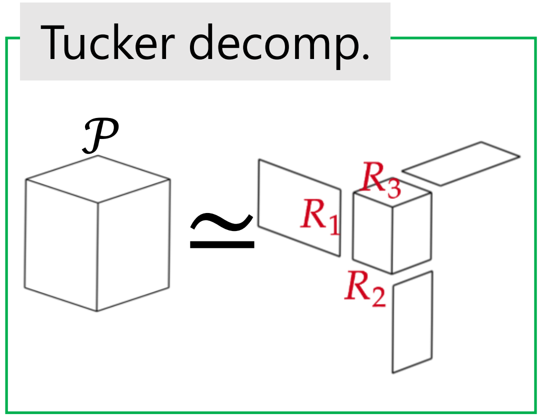

Tensor ring decomposion

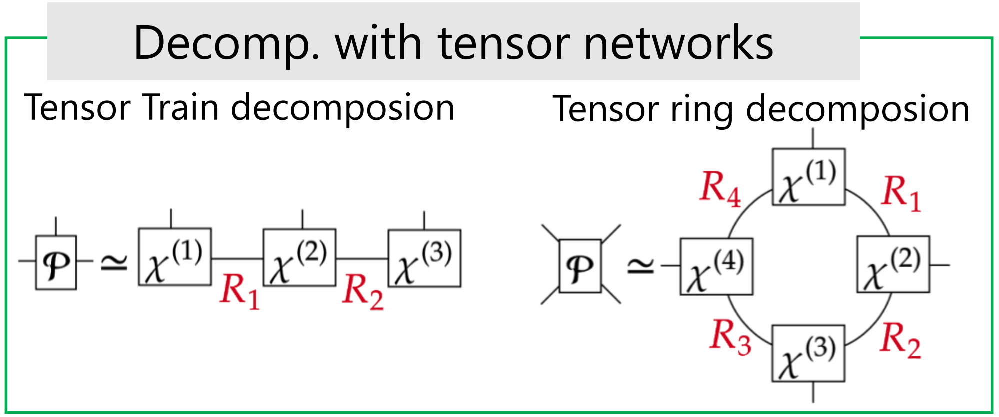


# Difficulties in tensor factorization

## ■ Model selection is not intuitive.



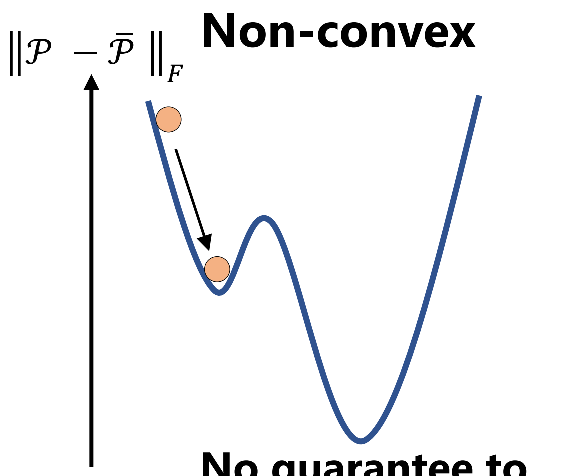




## ■ Optimization is difficult.

😟 **The objective function is typically non-convex.**
- Initial values dependency

😟 **Solution often might be indeterminate.**




**A convex, stable and intuitive tensor factorization is desired.**

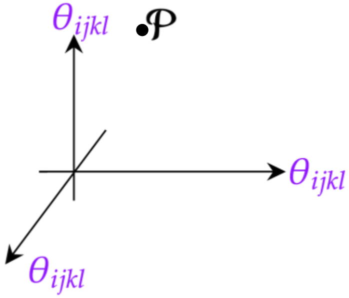
# Many-body approximation for non-negative tensors

**Energy function**

$$\mathcal{P}_{ijkl} = \frac{1}{Z}\exp[-E_{\theta}(i,j,k,l)]$$

Natural parameter
of exponential distribution family.

$$\sum_{ijkl} \mathcal{P}_{ijkl} = 1$$



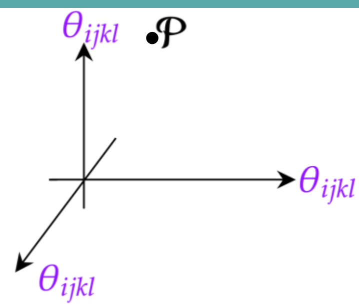
# Many-body approximation for non-negative tensors

Energy function

$$\mathcal{P}_{ijkl} = \frac{1}{Z}\exp[-\underbrace{E_{\theta}(i,j,k,l)}]$$

Natural parameter
of exponential distribution family.

$$\sum_{ijkl}\mathcal{P}_{ijkl} = 1$$




$$= \frac{1}{Z}\exp\left[E_i^{(1)} + \ldots + E_l^{(4)} + E_{ij}^{(12)} + \ldots + E_{kl}^{(34)} + E_{ijk}^{(123)} + \ldots + E_{jkl}^{(234)} + E_{ijkl}^{(1234)}\right]$$

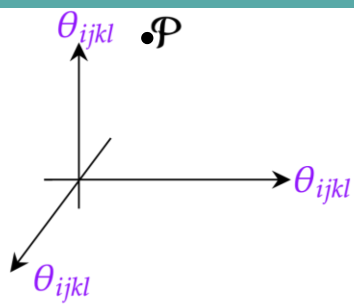
# Many-body approximation for non-negative tensors

**Energy function**

$$\mathcal{P}_{ijkl} = \frac{1}{Z}\exp[-E_{\theta}(i,j,k,l)]$$

Natural parameter
of exponential distribution family.

$$\sum_{ijkl} \mathcal{P}_{ijkl} = 1$$




$$= \frac{1}{Z}\exp\left[E^{(1)}_{i} + \ldots + E^{(4)}_{l} + E^{(12)}_{ij} + \ldots + \boxed{E^{(34)}_{kl}} + E^{(123)}_{ijk} + \ldots + \boxed{E^{(234)}_{jkl}} + E^{(1234)}_{ijkl}\right]$$

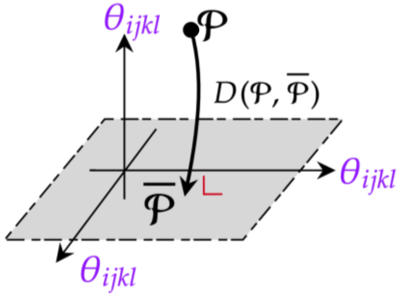
Control relation between
mode-*k* and mode-*l*.

Control relation among
mode-*j*, -*k* and -*l*.

# Many-body approximation for non-negative tensors

$$\mathcal{P}_{ijkl} = \frac{1}{Z}\exp[-E_{\theta}(i,j,k,l)]$$

$$= \frac{1}{Z}\exp\left[E_i^{(1)} + \ldots + E_l^{(4)} + E_{ij}^{(12)} + \ldots + E_{kl}^{(34)} + E_{ijk}^{(123)} + \ldots + E_{jkl}^{(234)} + E_{ijkl}^{(1234)}\right]$$




**One-body approx.**

$$\overline{\mathcal{P}}_{ijkl} = p_i^{(1)} p_j^{(2)} p_k^{(3)} p_l^{(4)}$$

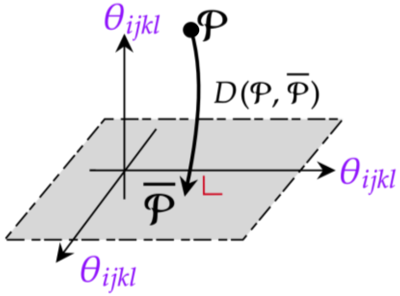
**Rank-1 approximation**
**（mean-field approximation）**

［**NeurIPS 2021** Ghalamkari, K., Sugiyama, M. ］

# Many-body approximation for non-negative tensors

$$\mathcal{P}_{ijkl} = \frac{1}{Z}\exp[-E_{\theta}(i,j,k,l)]$$


$$= \frac{1}{Z}\exp\left[E_i^{(1)} + \ldots + E_l^{(4)} + E_{ij}^{(12)} + \ldots + E_{kl}^{(34)} + E_{ijk}^{(123)} + \ldots + E_{jkl}^{(234)} + E_{ijkl}^{(1234)}\right]$$




Control relation between mode-*k* and mode-*l*.

**One-body approx.**

$$\overline{\mathcal{P}}_{ijkl} = p_i^{(1)} p_j^{(2)} p_k^{(3)} p_l^{(4)}$$

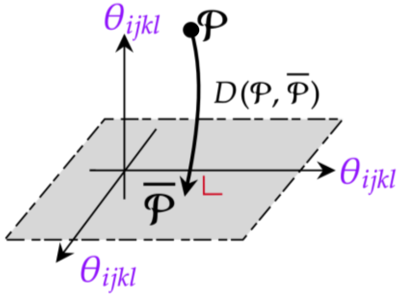
**Two-body approx.**

$$\overline{\mathcal{P}}_{ijkl} = X_{ij}^{(12)} X_{ik}^{(13)} X_{il}^{(14)} X_{jk}^{(23)} X_{jl}^{(24)} X_{kl}^{(34)}$$

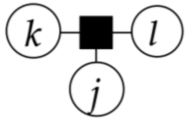
**Rank-1 approximation**
**(mean-field approximation)**
[NeurIPS 2021 Ghalamkari, K., Sugiyama, M. ]

Larger Capability

# Many-body approximation for non-negative tensors




$$\mathcal{P}_{ijkl} = \frac{1}{Z}\exp[-E_{\theta}(i,j,k,l)]$$

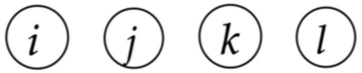
$$= \frac{1}{Z}\exp\left[E_i^{(1)} + \ldots + E_l^{(4)} + E_{ij}^{(12)} + \ldots + E_{kl}^{(34)} + E_{ijk}^{(123)} + \ldots + E_{jkl}^{(234)} + E_{ijkl}^{(1234)}\right]$$

Control relation between mode-*k* and mode-*l*.

Control relation among mode-*j*, -*k* and -*l*.

**One-body approx.**

$$\overline{\mathcal{P}}_{ijkl} = p_i^{(1)} p_j^{(2)} p_k^{(3)} p_l^{(4)}$$

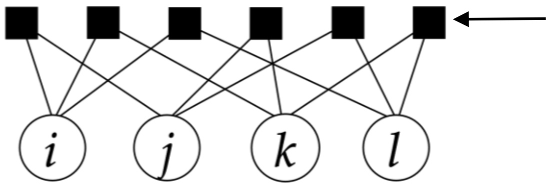


**Rank-1 approximation**
**(mean-field approximation)**

[NeurIPS 2021 Ghalamkari, K., Sugiyama, M. ]

**Two-body approx.**

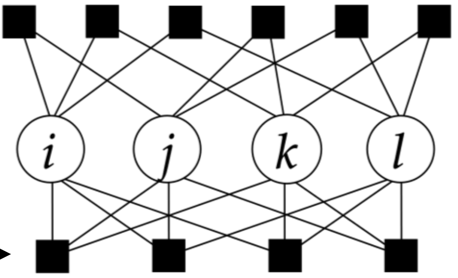
$$\overline{\mathcal{P}}_{ijkl} = X_{ij}^{(12)} X_{ik}^{(13)} X_{il}^{(14)} X_{jk}^{(23)} X_{jl}^{(24)} X_{kl}^{(34)}$$



Two-body Interaction

**Three-body approx.**

$$\overline{\mathcal{P}}_{ijkl} = \chi_{ijk}^{(123)} \chi_{ijl}^{(124)} \chi_{ikl}^{(134)} \chi_{jkl}^{(234)}$$



Three-body Interaction

**Intuitive modeling focusing on interactions between modes**

Larger Capability

**The global optimal solution $\overline{\mathcal{P}}$ minimizing KL divergence from $\mathcal{P}$ can be obtained by a convex optimization.**

# Theoretical idea behind proposal

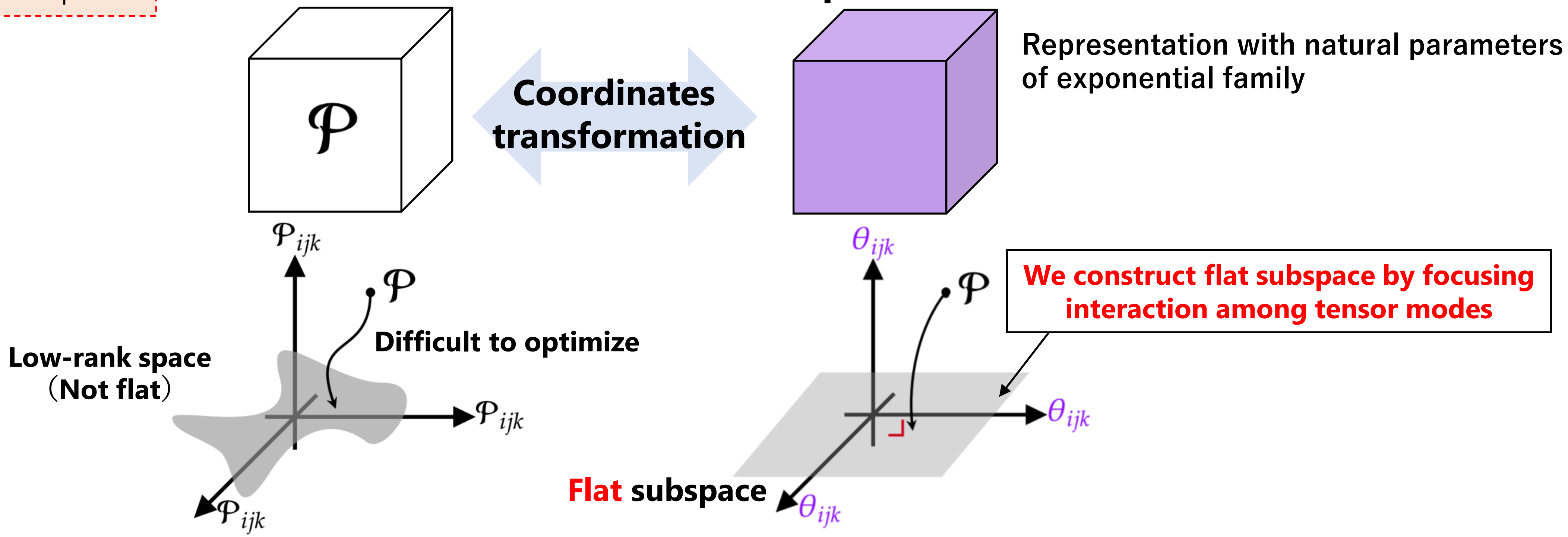
Index is discrete random variable

$$\sum_{ijk} \mathcal{P}_{ijk} = 1, \quad (i, j, k) \in \Omega = \{(1,1,1), \ldots, (I, J, K)\}$$

- **We regard a normalized tensor $\mathcal{P}$ as a discrete joint probability distribution whose sample space is an index set**
- **We use information geometry to formulate factorization as convex problem**

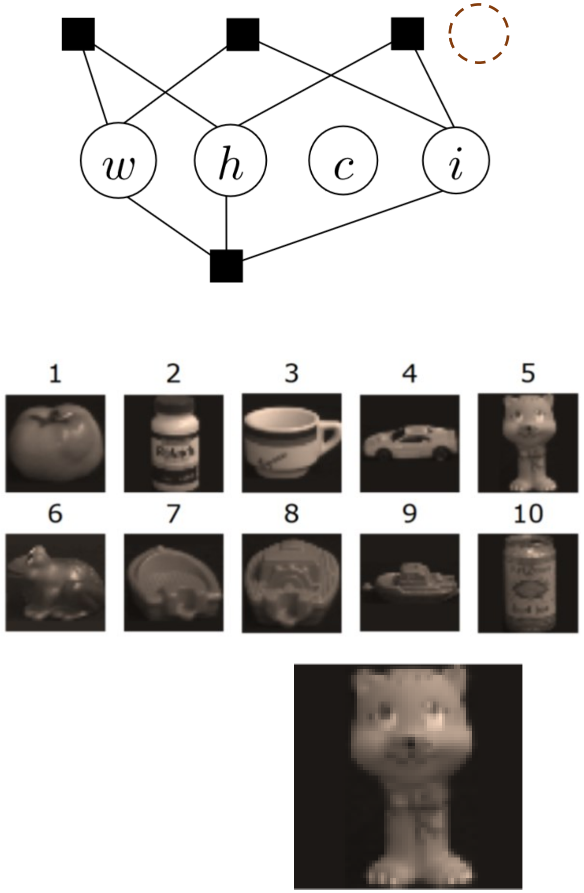
Geometry of $\boldsymbol{\theta}$-space

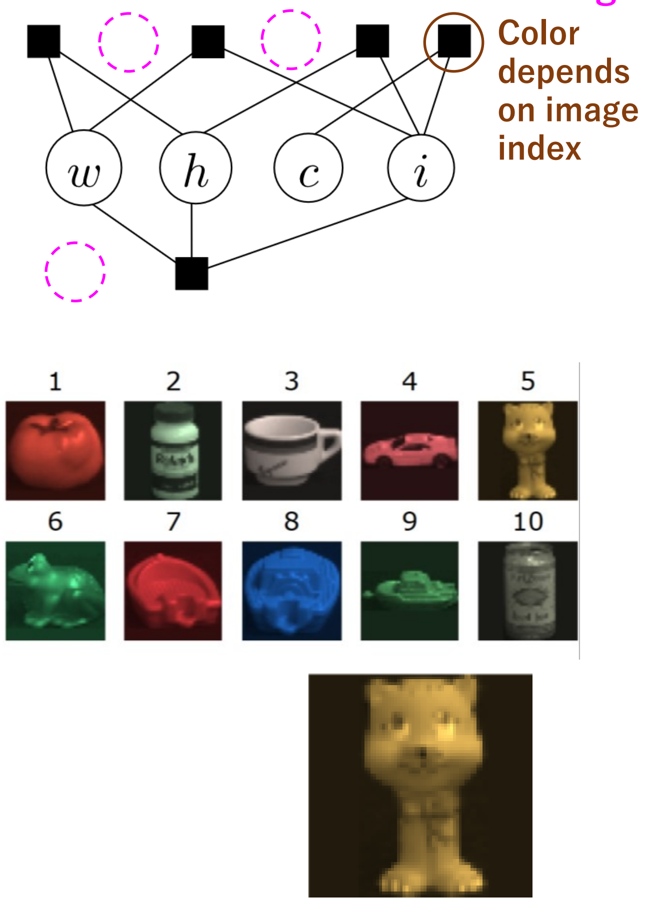
## $\boldsymbol{\theta}$- Representation



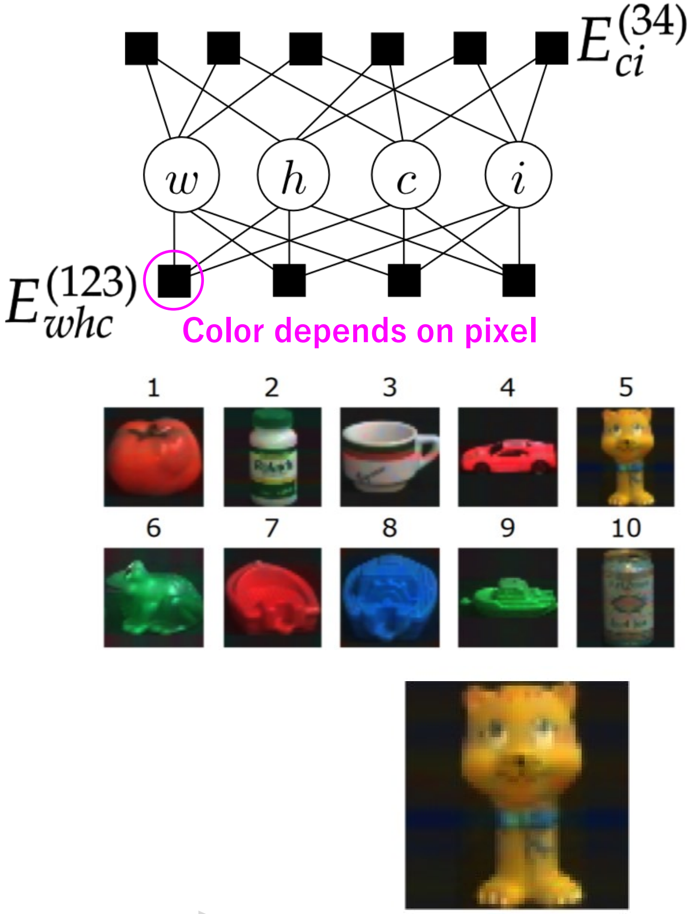

**Describing tensor factorization in $\boldsymbol{\theta}$-coordinate system makes it convex problem**

# Reconstruction for 40 × 40 × 3 × 10 tensor
(width, height, colors, # images)




**Three-body Approx.**




**Larger**
**ity**

**Intuitive model design that captures the relationship between modes**

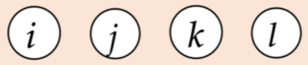
# Rank-free convex nonnegative tensor factorization

## Many-body Approximation

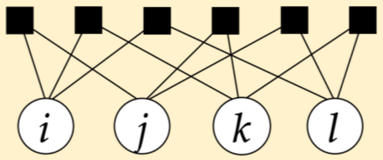
$$\mathcal{P}_{ijkl} = \exp\left[\sum_{i'=1}^{i}\sum_{j'=1}^{j}\sum_{k'=1}^{k}\sum_{l'=1}^{l}\theta_{i'j'k'l'}\right]$$

$$= \frac{1}{Z}\exp\left[E_i^{(1)}+\ldots+E_l^{(4)}+E_{ij}^{(12)}+\ldots+E_{kl}^{(34)}+E_{ijk}^{(123)}+\ldots+E_{jkl}^{(234)}+E_{ijkl}^{(1234)}\right]$$

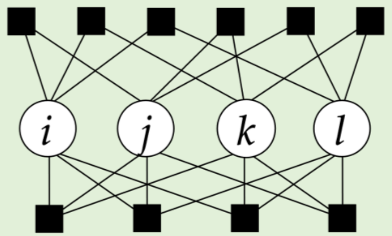
One-body Approx.




Two-body Approx.




Three-body Approx.




- **Convex optimization always provide unique solution**
- **More intuitive design than rank tuning**