

NeurIPS 2018  
Dec. 2–8, 2018



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



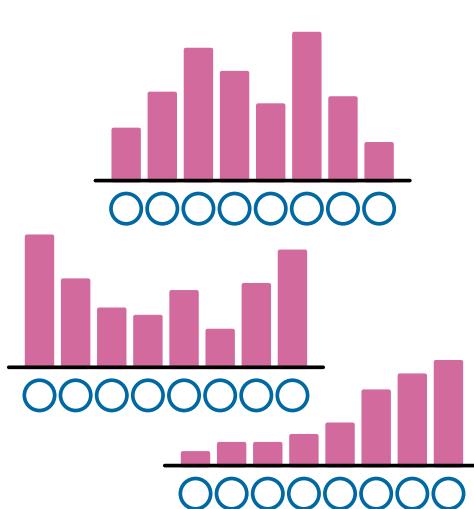
# Legendre Decomposition for Tensors

---

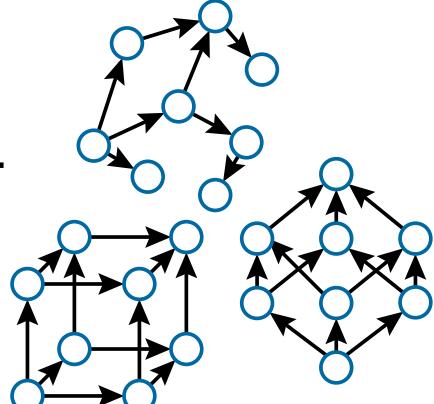
Mahito Sugiyama (National Institute of Informatics, JST PRESTO)  
Hiroyuki Nakahara (RIKEN CBS)  
Koji Tsuda (The University of Tokyo, NIMS, RIKEN AIP)

# Our Approach

Distribution



Partial order structure  
(DAG)



+

Apply the partial order structure  
to tensors

Manifold in  
Information Geometry

Parameter  $\theta$

Projection

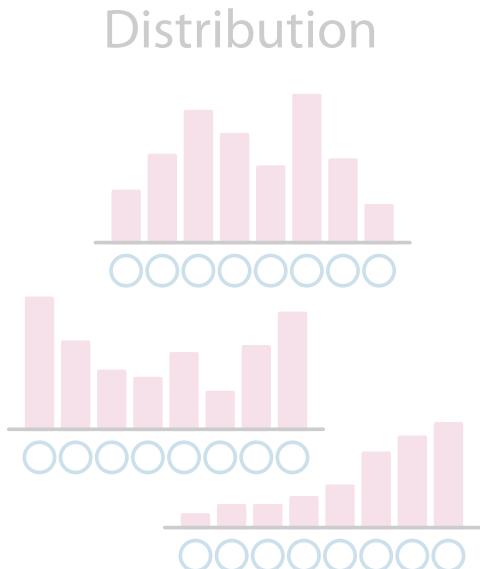
Parameter  $n$

Data

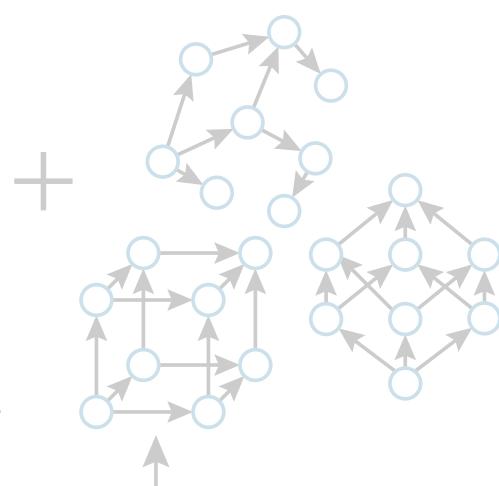
Data: Tensor

Projection: Legendre  
decomposition

# Our Approach



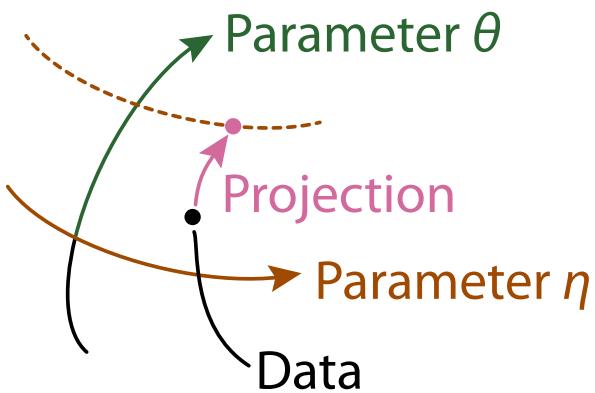
Partial order structure  
(DAG)



+

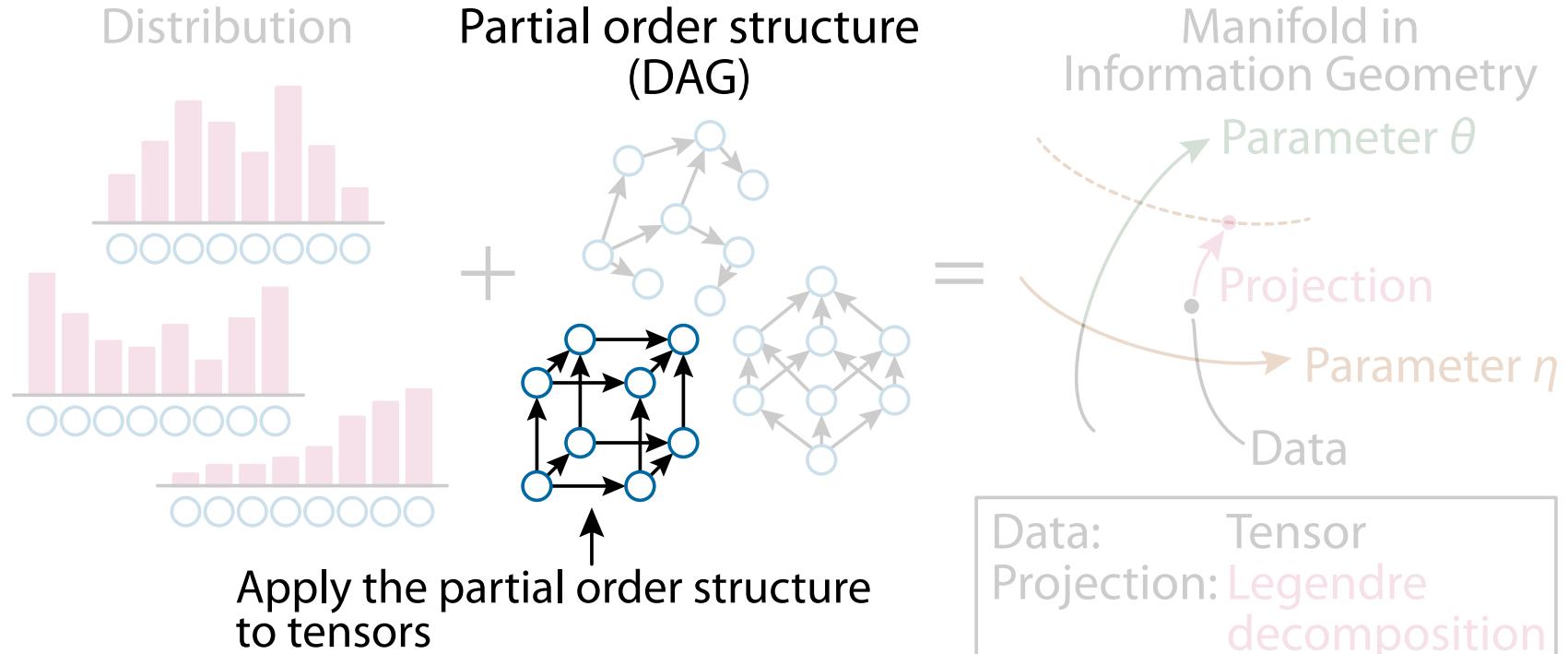
=

Manifold in  
Information Geometry



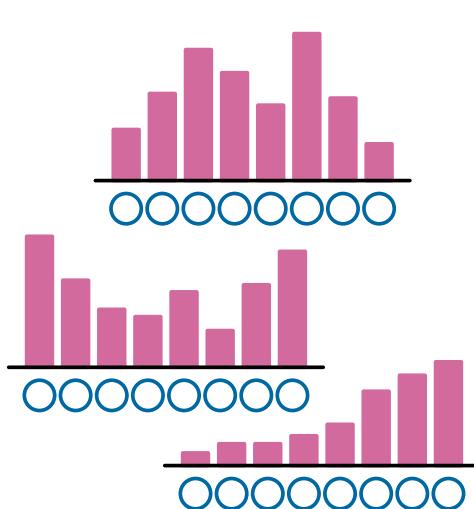
Data: Tensor  
Projection: Legendre decomposition

# Our Approach



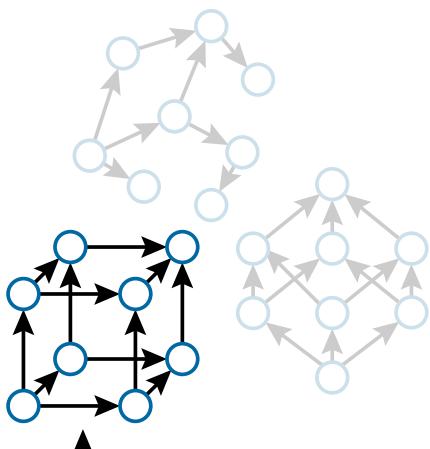
# Our Approach

Distribution



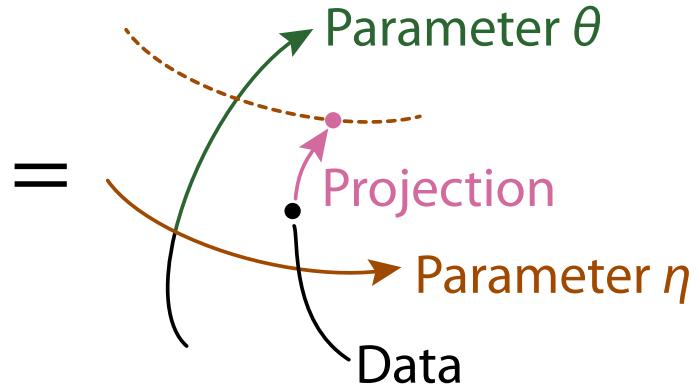
Partial order structure  
(DAG)

+



Apply the partial order structure  
to tensors

Manifold in  
Information Geometry



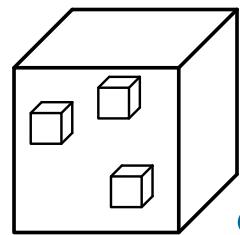
Data: Tensor

Projection: Legendre  
decomposition

# Legendre Decomposition

---

Tensor space

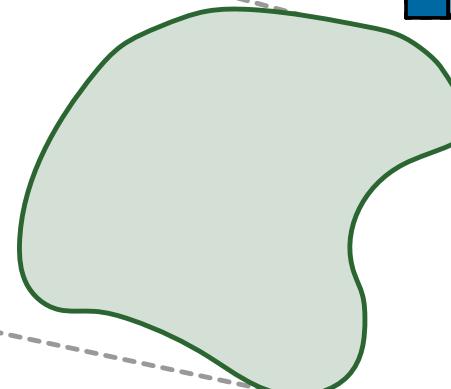


Input tensor  $\mathcal{P}$   
(can be sparse)

Decomposable  
tensors

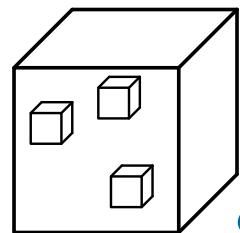
Parameter space

Decomposition  
basis  $B$



# Legendre Decomposition

Tensor space

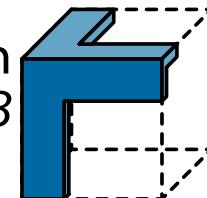


Input tensor  $\mathcal{P}$   
(can be sparse)

Decomposable  
tensors

Parameter space

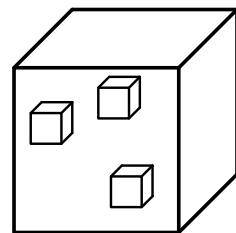
Decomposition  
basis  $B$



Legendre decomposition

# Legendre Decomposition

Tensor space



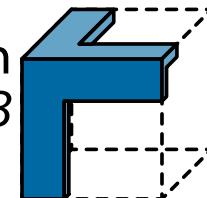
Input tensor  $\mathcal{P}$   
(can be sparse)

Decomposable  
tensors

(uniquely exists)

Parameter space

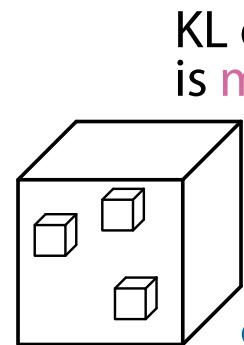
Decomposition  
basis  $B$



Legendre decomposition

# Legendre Decomposition

Tensor space



Input tensor  $\mathcal{P}$   
(can be sparse)

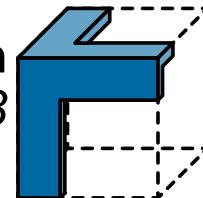
KL divergence  
is minimized

Decomposable  
tensors

(uniquely exists)

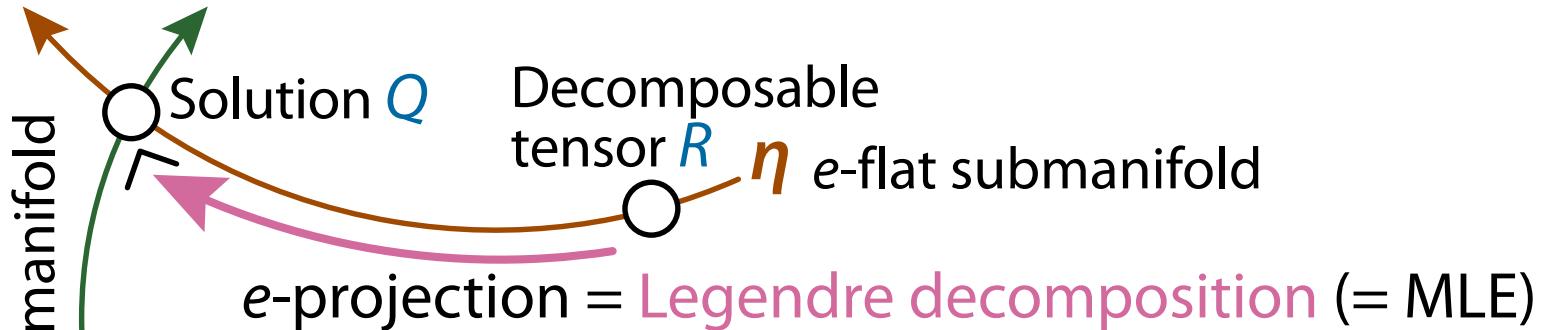
Parameter space

Decomposition  
basis  $B$



Legendre decomposition

# Information Geometry



$$\begin{bmatrix} \frac{\partial \eta(x)}{\partial \theta(y)} \\ \frac{\partial \theta(x)}{\partial \eta(y)} \end{bmatrix} = \sum_{s \in S} \underbrace{\zeta(x, s) \zeta(y, s) p(s)} - \eta(x) \eta(y)$$

Capture partial order structure

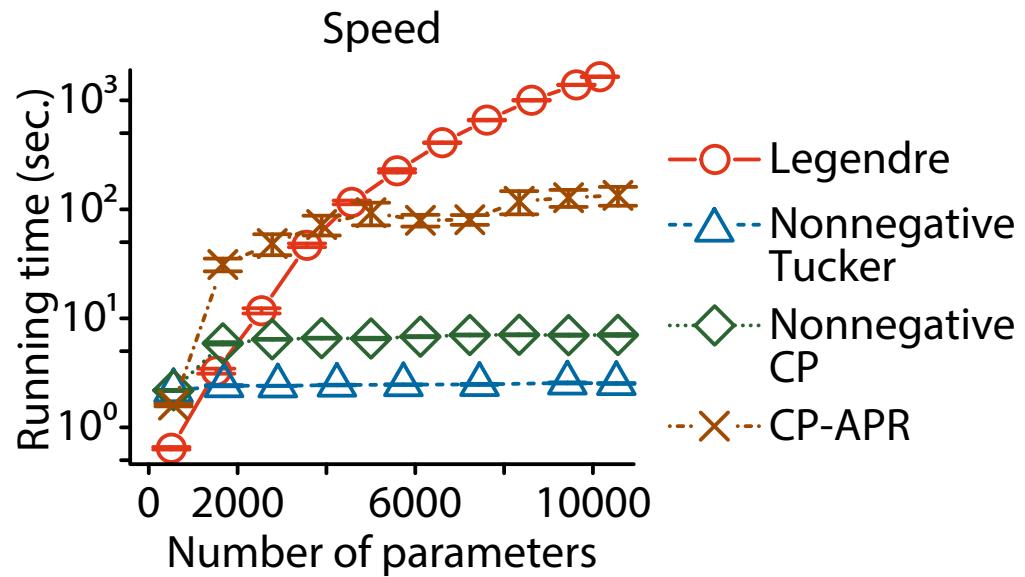
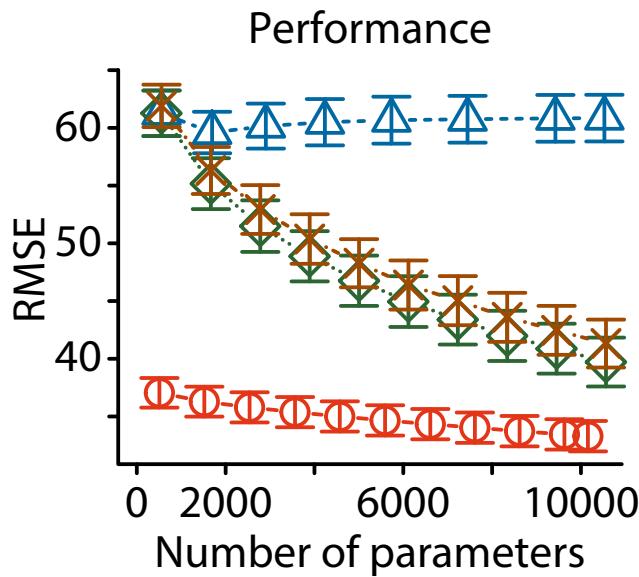
# Information Geometry



$$\begin{bmatrix} \frac{\partial \eta(x)}{\partial \theta(y)} \\ \frac{\partial \theta(x)}{\partial \eta(y)} \end{bmatrix} = \sum_{s \in S} \zeta(x, s) \zeta(y, s) p(s) - \eta(x) \eta(y)$$

Capture partial order structure

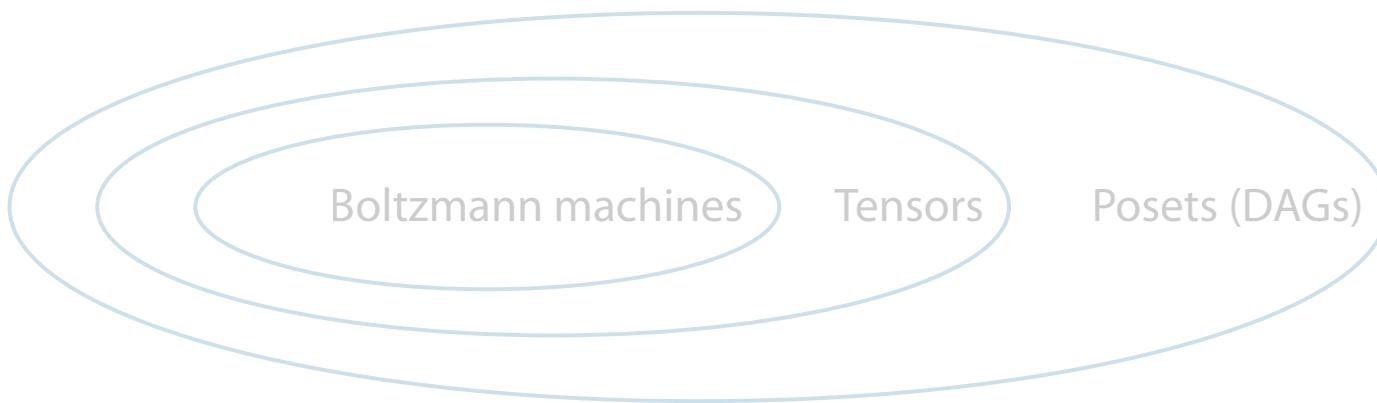
# Experimental Results on MNIST



# Summary

---

- We present **Legendre decomposition** for tensors
  - The solution always uniquely exists and minimizes the KL divergence
  - Dually flat manifold in information geometry is used
    - Parameters  $\theta$  and constraints  $\eta$  are connected via **Legendre transformation**



# Summary

---

- We present **Legendre decomposition** for tensors
  - The solution always uniquely exists and minimizes the KL divergence
  - Dually flat manifold in information geometry is used
    - Parameters  $\theta$  and constraints  $\eta$  are connected via **Legendre transformation**

