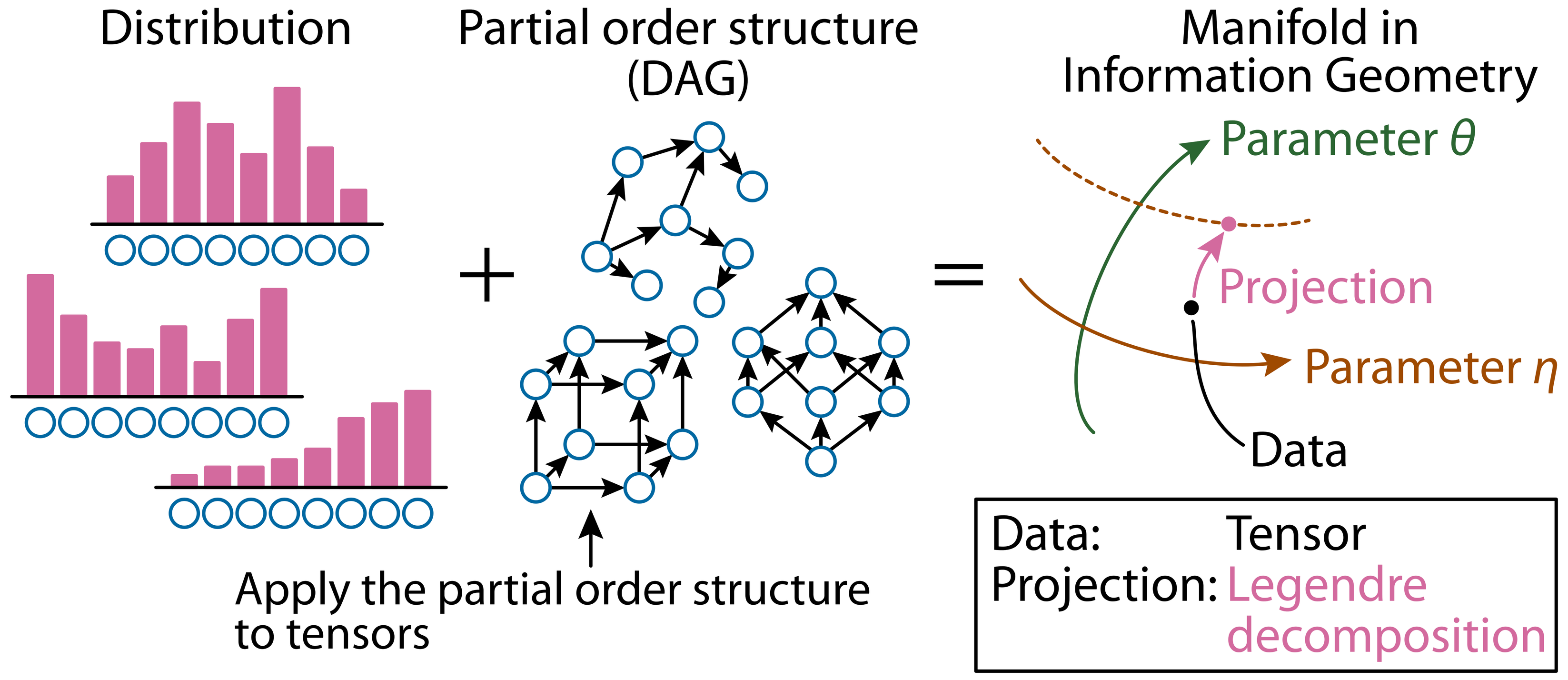
Legendre Decomposition for Tensors

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

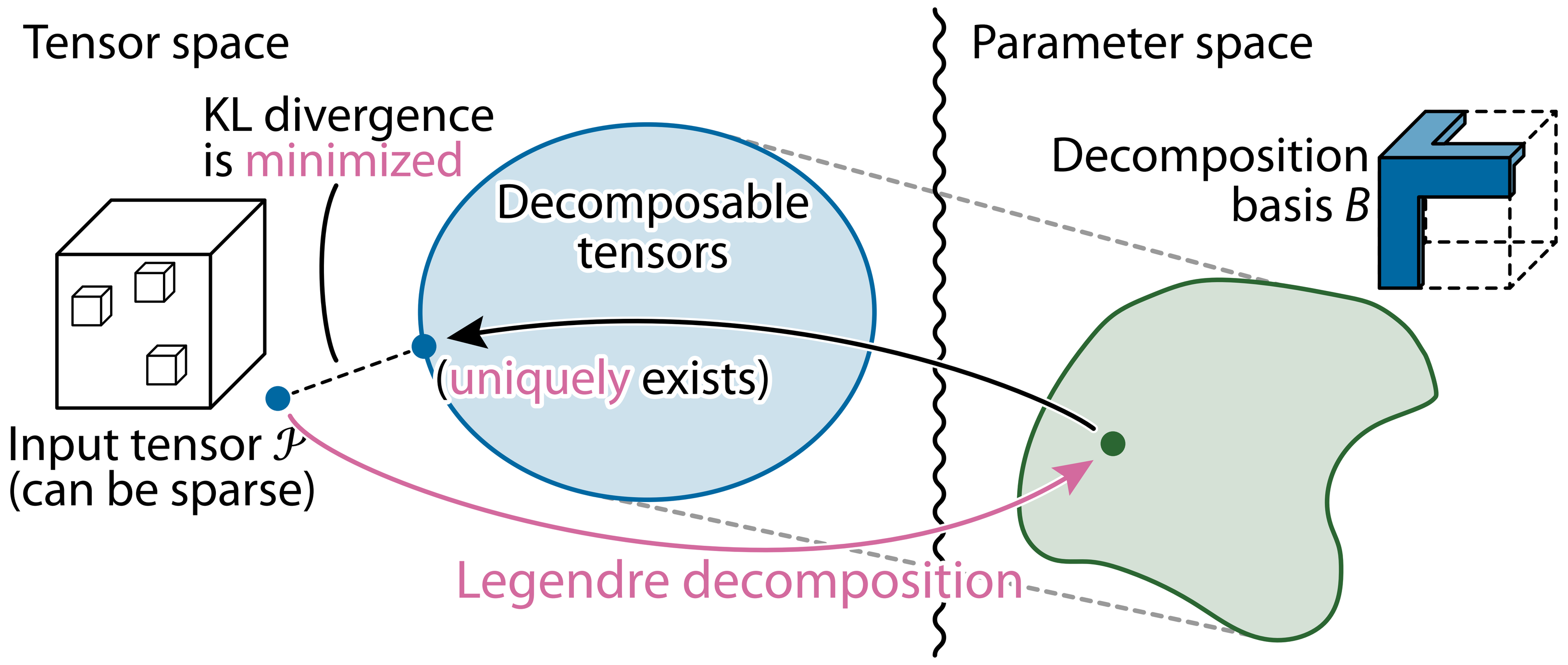
The 32nd Annual Conference on Neural Information Processing Systems (NeurIPS 2018), December 2–8, 2018

Our Approach



Summary

- We present Legendre decomposition for tensors
  - A new nonnegative decomposition method
  - A tensor is factorized into a multiplicative combination of parameters
- Our proposal is theoretically supported by information geometry
  - The reconstructed tensor is unique and always minimizes the KL divergence from an input tensor



Properties of Legendre Decomposition

- Given $\mathcal{P} \in \mathbb{R}_{\geq 0}^{I_1 \times I_2 \times \dots \times I_N}$, Legendre decomposition finds $\mathcal{Q}$, where
  (i) $\mathcal{Q}$ always exists, (ii) $\mathcal{Q}$ is unique, and
  (iii) $\mathcal{Q}$ is the best approximation in the sense of the KL divergence:

$$\mathcal{Q} = \mathrm{argmin}_{\mathcal{R} \in \boldsymbol{\mathcal{S}}_B} D_{\mathrm{KL}}(\mathcal{P}, \mathcal{R}),$$

$$\boldsymbol{\mathcal{S}}_B = \left\{ \mathcal{R} \in \mathbb{R}_{\geq 0}^{I_1 \times I_2 \times \dots \times I_N} \;\middle|\; \mathcal{R} \text{ is fully decomposable with } B \right\}$$

Legendre Decomposition

Input matrix:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{bmatrix}$$

Legendre decomposition →

Reconstructed matrix:

$$\begin{bmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \\ q_{41} & q_{42} & q_{43} & q_{44} \end{bmatrix}$$

↕ θ: parameter

Decomposition basis (arbitrary subset of indices)

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & 0 \\ \theta_{31} & \theta_{32} & 0 \end{bmatrix} \qquad \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} \\ \theta_{21} & \theta_{22} & 0 & 0 \\ \theta_{31} & \theta_{32} & 0 & 0 \\ \theta_{41} & \theta_{42} & 0 & 0 \end{bmatrix}$$

$$q_{ij} = \prod_{k \leq i, l \leq j} \exp(\theta_{kl})$$

↕ η: constraint

$$\begin{bmatrix} q_{22} & q_{23} & q_{24} \\ q_{32} & q_{33} & q_{34} \\ q_{42} & q_{43} & q_{44} \end{bmatrix} \qquad \begin{bmatrix} \eta_{11} & \eta_{12} & \eta_{13} & \eta_{14} \\ \eta_{21} & \eta_{22} & & \\ \eta_{31} & \eta_{32} & & \\ \eta_{41} & \eta_{42} & & \end{bmatrix}$$

$$\eta_{ij} = \sum_{k \geq i, l \geq j} q_{kl}$$

Input matrix:

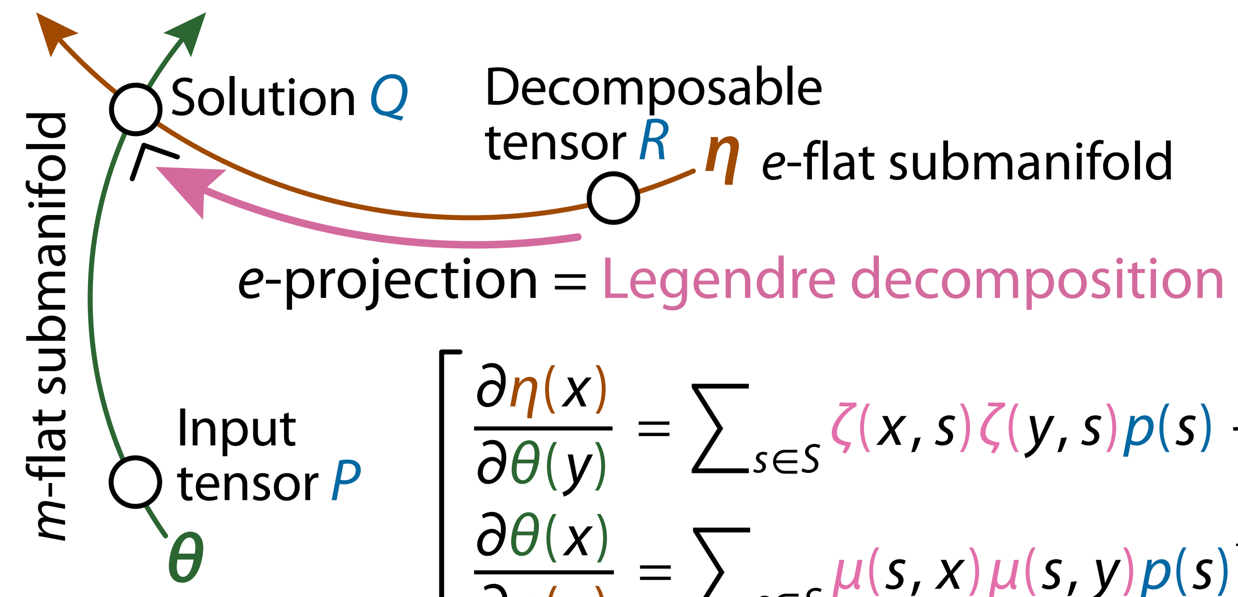
$$\begin{bmatrix} \hat{\eta}_{11} & \hat{\eta}_{12} & \hat{\eta}_{13} & \hat{\eta}_{14} \\ \hat{\eta}_{21} & \hat{\eta}_{22} & & \\ \hat{\eta}_{31} & \hat{\eta}_{32} & & \\ \hat{\eta}_{41} & \hat{\eta}_{42} & & \end{bmatrix}$$

Legendre decomposition →
Find θ s.t. $\hat{\eta} = \eta$

Reconstructed matrix:

$$\begin{bmatrix} \eta_{11} & \eta_{12} & \eta_{13} & \eta_{14} \\ \eta_{21} & \eta_{22} & & \\ \eta_{31} & \eta_{32} & & \\ \eta_{41} & \eta_{42} & & \end{bmatrix}$$

Information Geometry

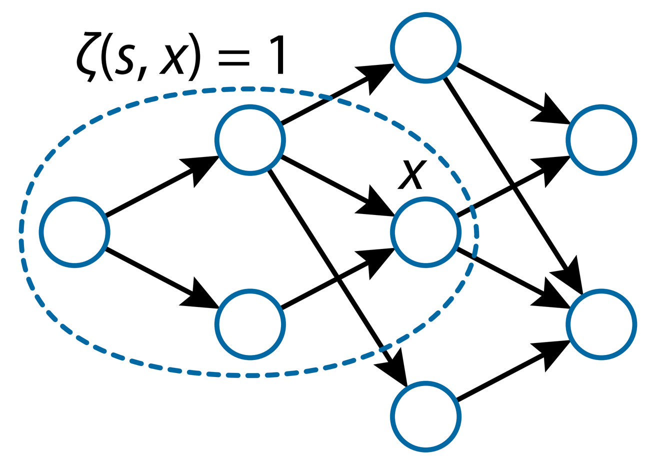


e-projection = Legendre decomposition (= MLE)

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

$$\frac{\partial \theta(x)}{\partial \eta(y)} = \sum_{s \in S} \mu(s, x)\mu(s, y)p(s)^{-1}$$

Capture partial order structure



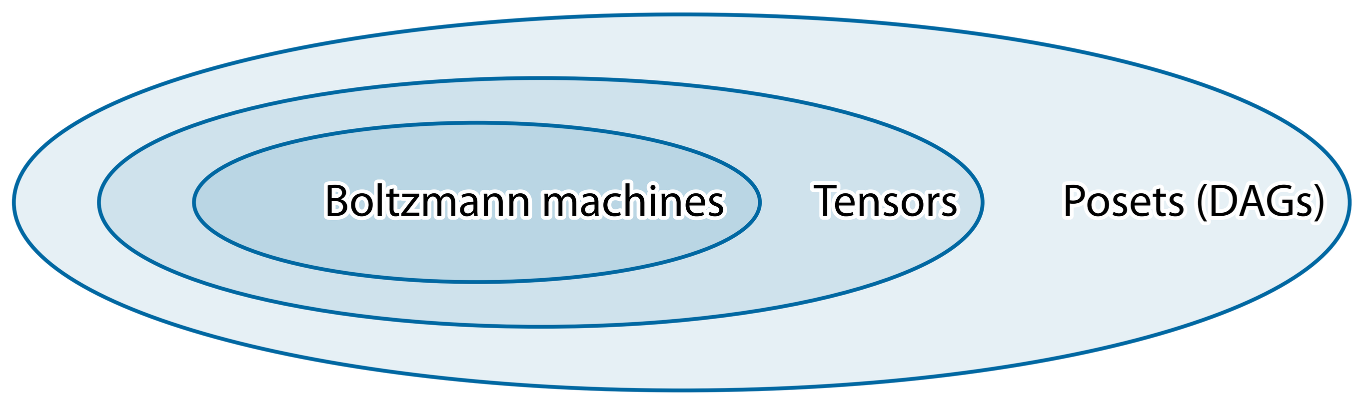
- Zeta function $\zeta : S \times S \to \{0, 1\}$

$$\zeta(s, x) = \begin{cases} 1 & \text{if } s \leq x, \\ 0 & \text{otherwise.} \end{cases}$$

- Möbius function $\mu : S \times S \to \mathbb{Z}$

$$\mu(x, y) = \begin{cases} 1 & \text{if } x = y, \\ -\sum_{x \leq s < y} \mu(x, s) & \text{if } x < y, \\ 0 & \text{otherwise.} \end{cases}$$

  - We have $\zeta\mu = I$, that is; $\sum_{s \in S} \zeta(s, y)\mu(x, s) = \sum_{x \leq s \leq y} \mu(x, s) = \delta_{xy}$



Experiments on MNIST