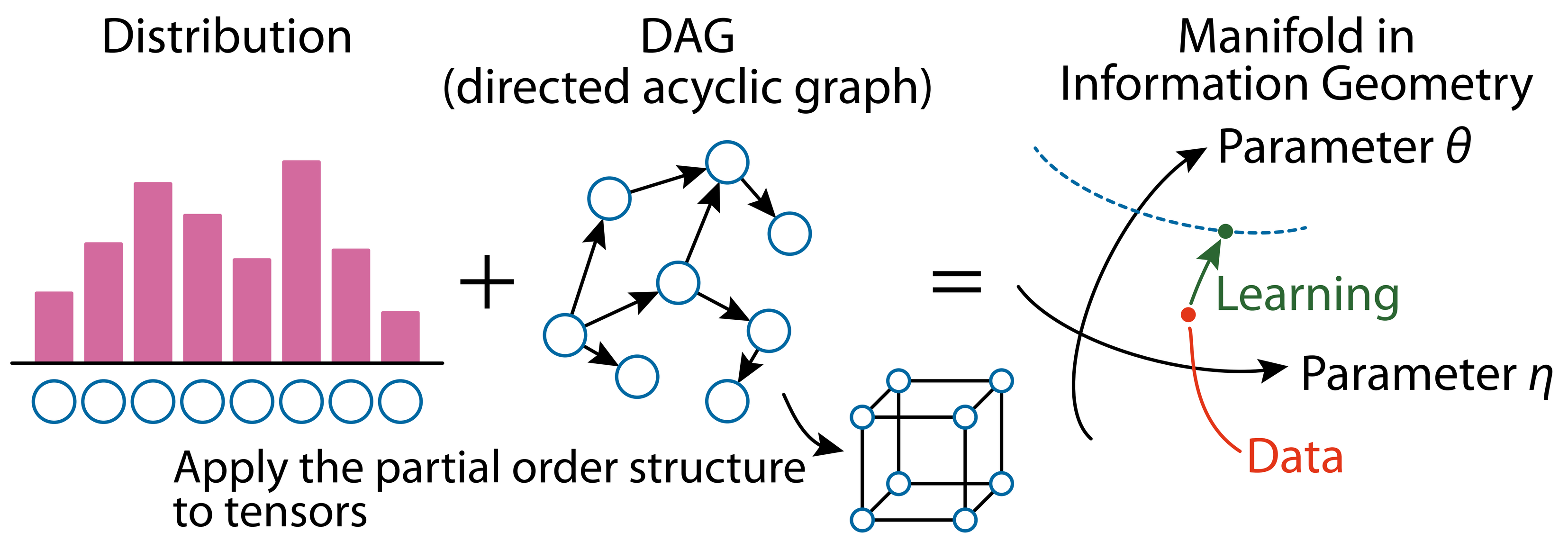
Legendre Decomposition for Tensors

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

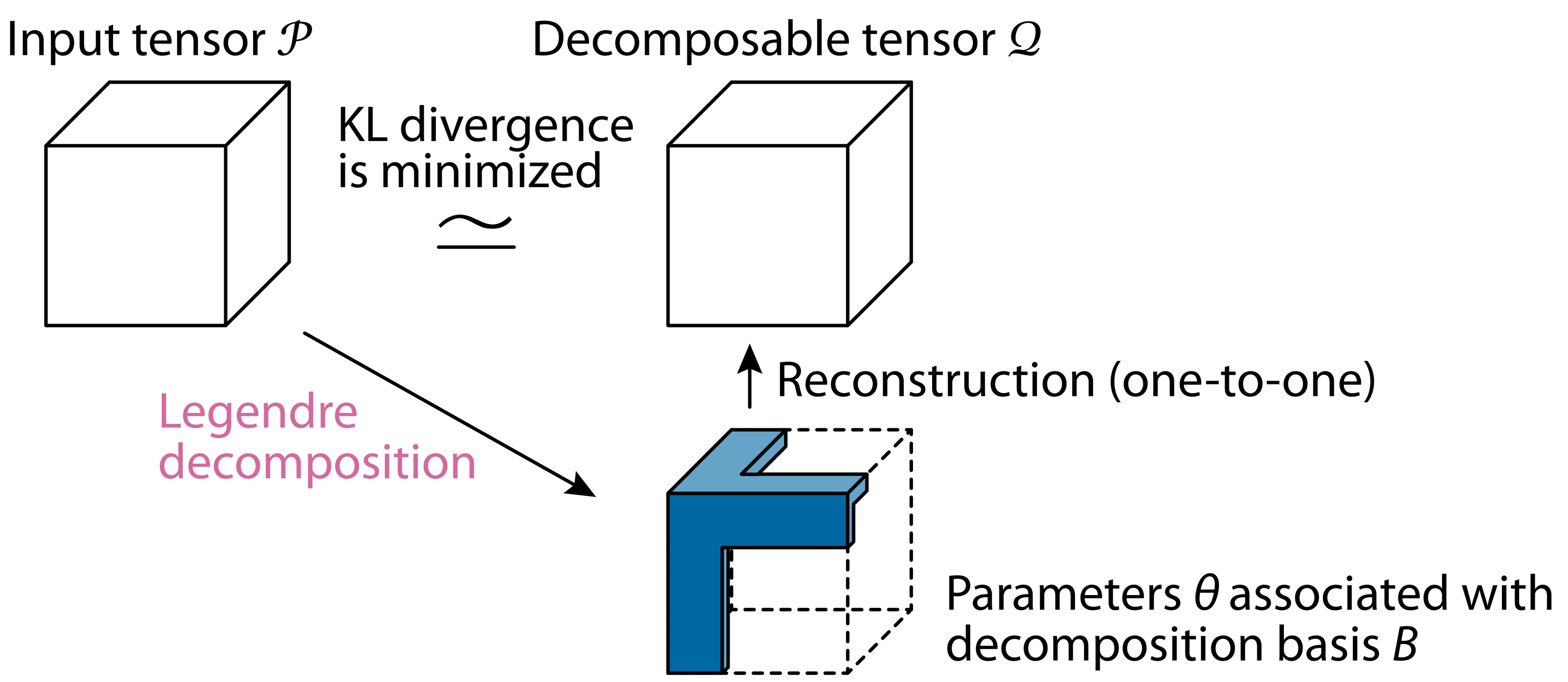
The 32nd Annual Conference on Neural Information Processing Systems, December 2–8, 2018

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:

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

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

Legendre Decomposition

Input matrix:

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

Reconstructed matrix:

Decomposition basis (arbitrary subset of indices)

↕ θ: parameter

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

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

↕ η: constraint

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

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

Input matrix:

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

Reconstructed matrix:

Find $\boldsymbol{\theta}$ s.t. $\hat{\boldsymbol{\eta}} = \boldsymbol{\eta}$

Information Geometry

- Consider statistical manifold: $\boldsymbol{\mathcal{S}} = \{P \mid 0 < p_v < 1 \text{ and } \sum p_v = 1\}$
  - Each distribution P corresponds to a tensor $\mathcal{P}$; p_v is a probability
- $\boldsymbol{\mathcal{S}}$ is dually flat as (θ, η) is connected via Legendre transformation:
  $\theta = \nabla\varphi(\eta)$, $\quad \eta = \nabla\psi(\theta)$
  with two convex functions
  $\psi(\theta) = -\theta_\perp = -\log p_\perp$, $\quad \varphi(\eta) = \sum_{v \in \Omega} p_v \log p_v$
  - $\perp$ is the least element and corresponds to $(1, 1, \dots, 1)$
- Riemannian metric g, Fisher information (gradient), is given as
  $g(\theta) = \nabla\nabla\psi(\theta)$, $\quad g(\eta) = \nabla\nabla\varphi(\eta)$,

$$g_{uv}(\theta) = \frac{\partial \eta_u}{\partial \theta_v} = \mathbf{E}\left[\frac{\partial \log p_w}{\partial \theta_u}\frac{\partial \log p_w}{\partial \theta_v}\right] = \sum_{w \in \Omega} \zeta(u, w)\zeta(v, w)p_w - \eta_u\eta_v,$$

$$g_{uv}(\eta) = \frac{\partial \theta_u}{\partial \eta_v} = \mathbf{E}\left[\frac{\partial \log p_w}{\partial \eta_u}\frac{\partial \log p_u}{\partial \eta_v}\right] = \sum_{w \in \Omega} \mu(w, u)\mu(w, v)p_w^{-1}$$

  - We use zeta function ζ and Möbius function μ

Legendre Decomposition as Projection

- The e-flat and m-flat submanifolds:

$$\boldsymbol{\mathcal{S}}_B = \{\, \mathcal{Q} \in \boldsymbol{\mathcal{S}} \mid \theta_v = 0 \text{ for all } v \in \Omega \setminus B \,\}$$

$$\boldsymbol{\mathcal{S}}'_B = \{\, \mathcal{Q} \in \boldsymbol{\mathcal{S}} \mid \eta_v = \hat{\eta}_v \text{ for all } v \in B \,\}$$

- Legendre decomposition finds the intersection $\boldsymbol{\mathcal{S}}_B \cap \boldsymbol{\mathcal{S}}'_B$
  - From some $P \in \boldsymbol{\mathcal{S}}_B$ to $\boldsymbol{\mathcal{S}}_B \cap \boldsymbol{\mathcal{S}}'_B$ is called e-projection

Experiments on MNIST