Tensor Balancing on Statistical Manifold

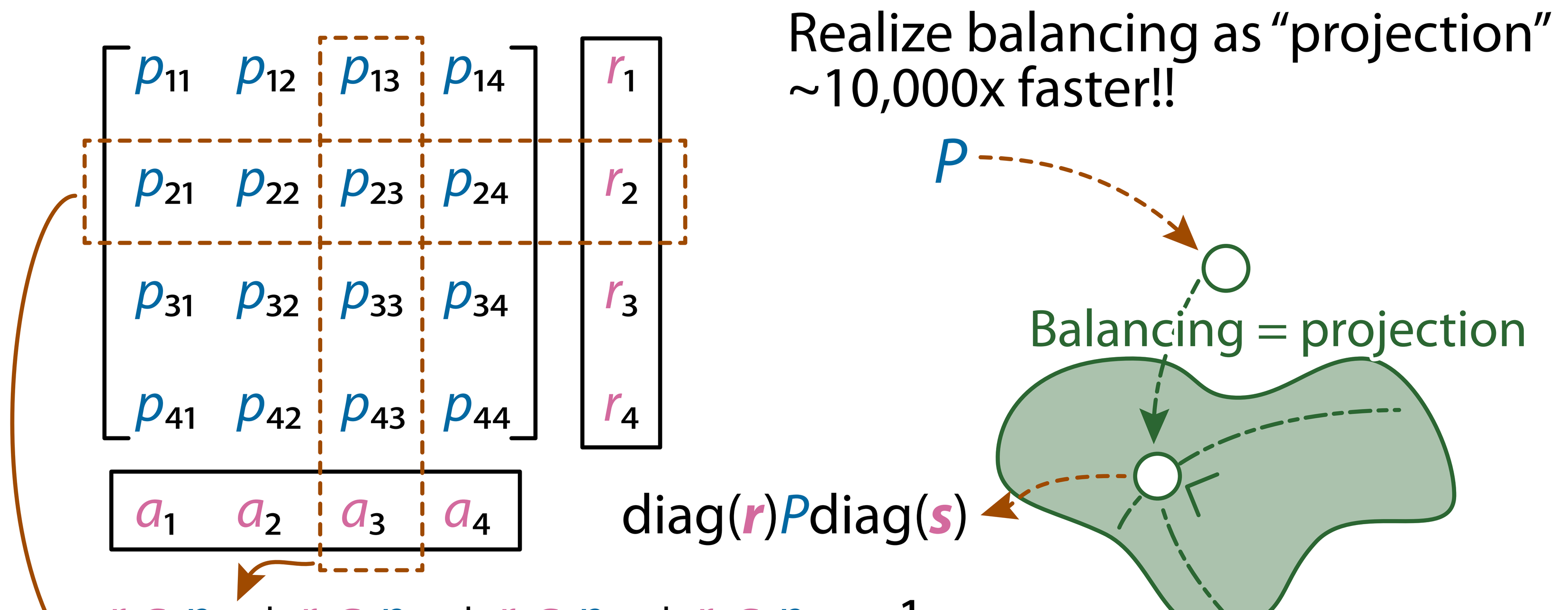
Mahito Sugiyama (NII, JST PRESTO), Hiroyuki Nakahara (RIKEN BSI), Koji Tsuda (UTokyo, RIKEN AIP, NIMS)

The 34th International Conference on Machine Learning (ICML 2017), August 6–11, 2017

Results

- Balancing of higher order (more than two) tensors is firstly (theoretically) achieved
- A fast balancing algorithm with quadratic convergence using Newton's method (an existing algorithm is linear convergence)
- **[Theory]** We provide dually flat Riemannian manifold of probability distributions with the structured outcome space

Matrix (Tensor) Balancing

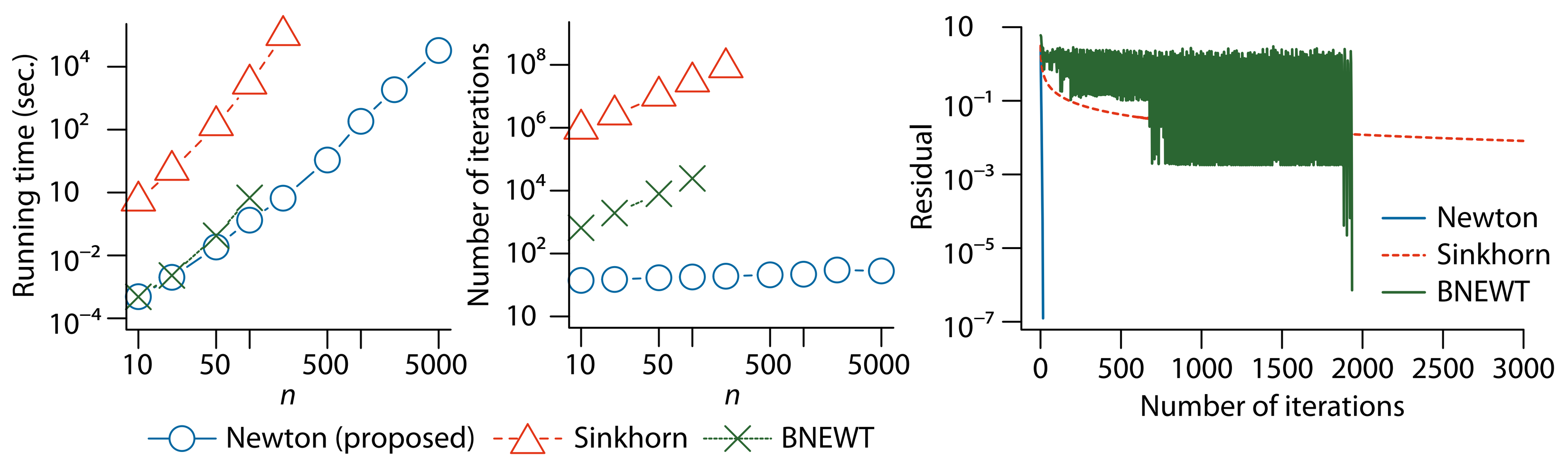


$r_1a_3p_{13} + r_2a_3p_{23} + r_3a_3p_{33} + r_4a_3p_{43} = 1$

$r_2a_1p_{21} + r_2a_2p_{22} + r_2a_3p_{23} + r_2a_4p_{24} = 1$

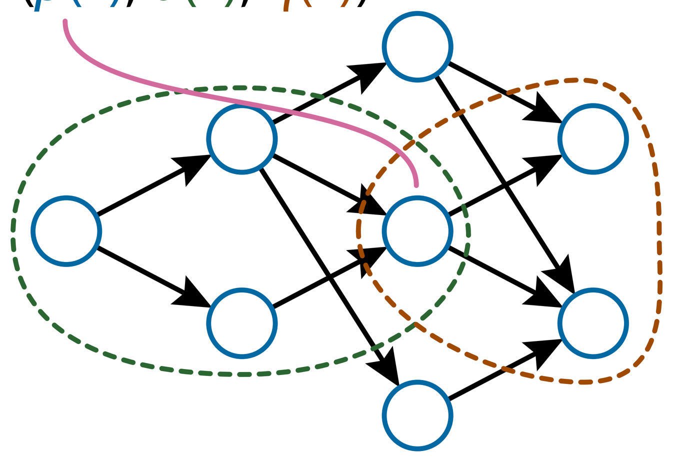
Use geometric structure in Information Geometry

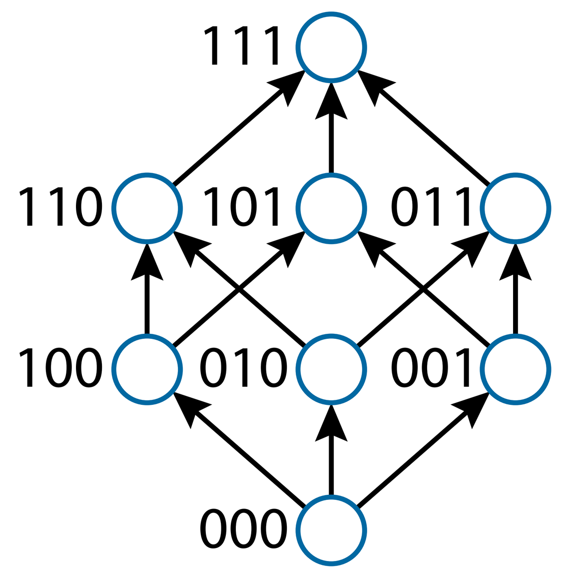
- **Matrix balancing**: Given a nonnegative matrix $P = (p_{ij}) \in \mathbb{R}_+^{n \times n}$, find $\boldsymbol{r}, \boldsymbol{s} \in \mathbb{R}^n$ s.t.
  $(RPS)\mathbf{1} = \mathbf{1}$ and $(RPS)^T\mathbf{1} = \mathbf{1}$
  - R = diag($\boldsymbol{r}$), S = diag($\boldsymbol{s}$), each entry is given as $p'_{ij} = p_{ij}r_is_j$
  - Applications: input-output analysis, Hi-C data analysis, the Sudoku puzzle, and Wasserstein metric approximation
  - Standard balancing algorithm: Sinkhorn–Knopp algorithm



Log-Linear Model on Poset

Each $x \in S$ has a triple: $(p(x), \theta(x), \eta(x))$





- Let $(S, \leq)$ be a poset (= DAG)
- A probability vector $p{:}S \to (0,1)$ s.t. $\sum_{x \in S} p(x) = 1$
  - (Normalized) weight for each node
- We introduce $\theta{:}S \to \mathbb{R}$ and $\eta{:}S \to \mathbb{R}$ as
  $\log p(x) = \sum_{s \leq x} \theta(s)$, $\eta(x) = \sum_{s \geq x} p(s)$
- Our model is generalization of the log-linear model on binary vectors with $\boldsymbol{x} \in \{0,1\}^n = S$:

$$\log p(\boldsymbol{x}) = \sum_i \theta^i x^i + \sum_{i<j} \theta^{ij} x^i x^j + \dots + \theta^{1 \dots n} x^1 x^2 \dots x^n - \psi,$$

$$\eta^i = \mathbf{E}[x^i] = \Pr(x^i = 1),$$

$$\eta^{ij} = \mathbf{E}[x^i x^j] = \Pr(x^i = x^j = 1), \dots$$

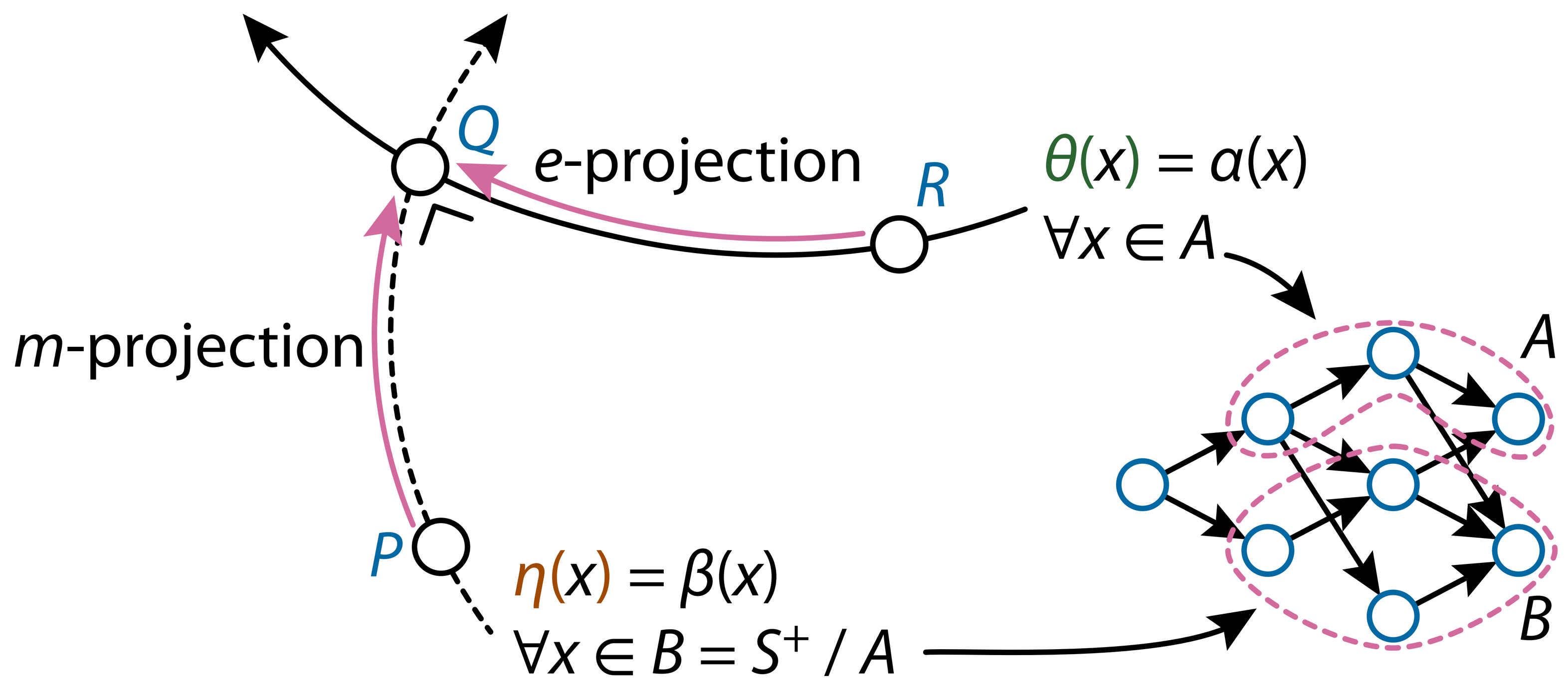
Dually Flat Structure

- θ and η form a dual coordinate system: $\nabla\psi(\theta) = \eta$, $\nabla\varphi(\eta) = \theta$
  - $\psi(\theta) = -\theta(\bot) = -\log p(\bot)$, $\varphi(\eta) = \sum_{x \in S} p(x) \log p(x)$
  - $\psi(\theta)$ and $\varphi(\eta)$ are connected via the Legendre transformation:
    $\varphi(\eta) = \max_{\theta'} \left(\theta'\eta - \psi(\theta')\right)$, $\theta'\eta = \sum_{x \in S \setminus \{\bot\}} \theta'(x)\eta(x)$
- The gradients: $g(\theta) = \nabla\nabla\psi(\theta) = \nabla\eta$, $g(\eta) = \nabla\nabla\varphi(\eta) = \nabla\theta$

$$\begin{cases} g_{xy}(\theta) = \dfrac{\partial \eta(x)}{\partial \theta(y)} = \displaystyle\sum_{s \in S} \zeta(x,s)\zeta(y,s)p(s) - \eta(x)\eta(y) \\ g_{xy}(\eta) = \dfrac{\partial \theta(x)}{\partial \eta(y)} = \displaystyle\sum_{s \in S} \mu(s,x)\mu(s,y)p(s)^{-1} \end{cases}$$

  - Zeta function: $\zeta(s,x) = 1$ if $s \leq x$, $\zeta(s,x) = 0$ otherwise
  - Möbius function: $\mu{:}S \times S \to \mathbb{Z}$ satisfying $\zeta\mu = I$
  - The manifold $(\boldsymbol{\mathcal{S}}, g(\xi))$ is a Riemannian manifold with the set $\boldsymbol{\mathcal{S}}$ of probability vectors and the Riemannian metric $g(\xi)$

e-Projection and *m*-Projection



- *e*-projection and *m*-projection can be computed by Newton's Method
- Each step of Newton's method in *e*-projection:

$$\begin{bmatrix} \vdots \\ \eta^{(t)}_{P_\beta}(x) - \beta(x) \\ \vdots \end{bmatrix} + J \begin{bmatrix} \vdots \\ \theta^{(t+1)}_{P_\beta}(y) - \theta^{(t)}_{P_\beta}(y) \\ \vdots \end{bmatrix} = \mathbf{0},$$

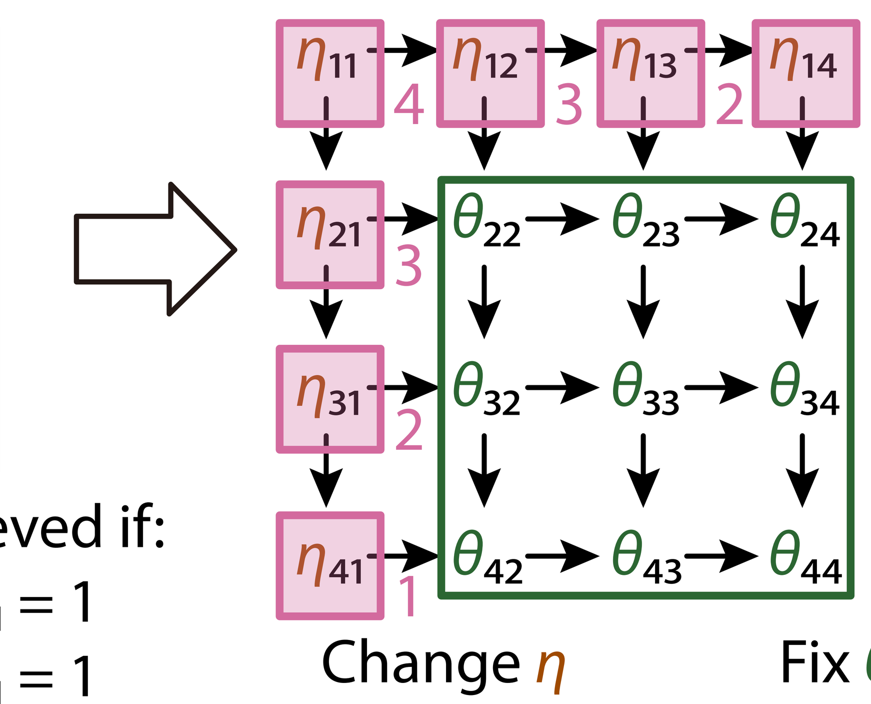
  - J is the $|B| \times |B|$ Jacobian matrix given as

$$J_{xy} = \frac{\partial \eta^{(t)}_{P_\beta}(x)}{\partial \theta^{(t)}_{P_\beta}(y)} = \sum_{s \in S} \zeta(x,s)\zeta(y,s)p_\beta^{(t)}(s) - \eta^{(t)}_{P_\beta}(x)\eta^{(t)}_{P_\beta}(y)$$

for each $x, y \in B$

Balancing = *e*-Projection

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



Matrix balancing is achieved if:

$\eta_{11} = 4, \eta_{21} = 3, \eta_{31} = 2, \eta_{41} = 1$

$\eta_{11} = 4, \eta_{12} = 3, \eta_{13} = 2, \eta_{41} = 1$