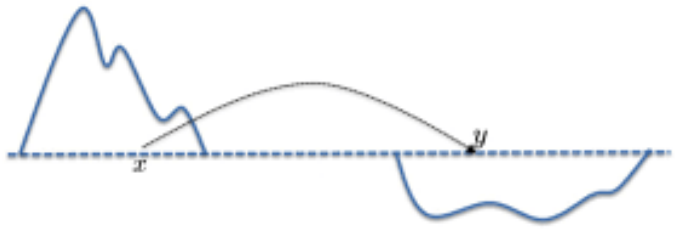
Graphical Optimal Transport and its Applications

Isabel Haasler[1] Ringh Axel[2] Jiaojiao Fan[3] Rahul Singh[3]
Qinsheng Zhang[3] Johan Karlsson[1] Yongxin Chen[3]

[1]KTH [2]Chalmers [3]Georgia Tech



Optimal transport



given: marginal distributions μ_1, μ_2 and cost function $c(x, y)$

Monge's formulation

$$\inf_{T_\sharp \mu_1 = \mu_2} \int c(x, T(x)) d\mu_1(x)$$

Kantorovich's formulation: linear programming

$$\inf_{\pi \in \Pi(\mu_1, \mu_2)} \int c(x, y) \pi(dxdy)$$

Wasserstein distance and dynamic formulations

Wasserstein distance:

$$W_p^p(\mu_1,\mu_2) = \inf_{\pi\in\Pi(\mu_1,\mu_2)} \int \|x-y\|^p \pi(dxdy)$$

Fluid dynamics:

$$\begin{aligned} W_p^p(\mu_1,\mu_2) &= \inf_{(\rho,u)} \int_{\mathbb{R}^m}\int_0^1 \|u(t,x)\|^p \rho(t,x)dtdx \\ & \frac{\partial\rho}{\partial t} + \nabla\cdot(u\rho) = 0 \\ & \rho(0,\cdot)=\mu_1, \quad \rho(1,\cdot)=\mu_2 \end{aligned}$$

Stochastic control:

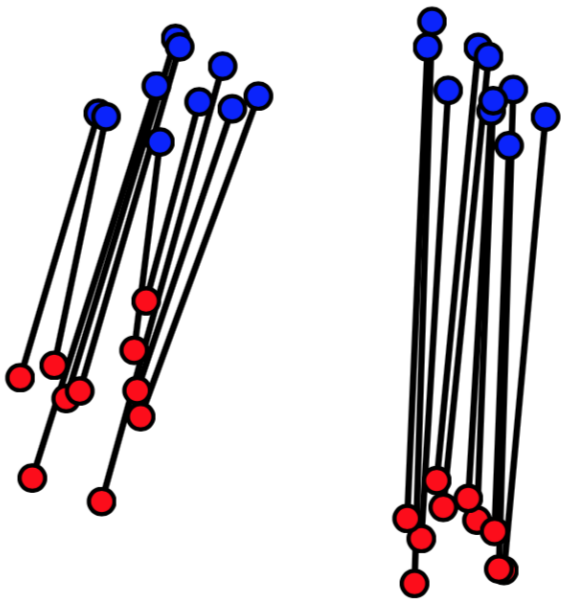
$$\begin{aligned} W_p^p(\mu_1,\mu_2) &= \inf_u \ \mathbb{E}\left\{\int_0^1 \|u(t)\|^p dt\right\} \\ & \dot{x}(t) = u(t,x) \\ & x(0)\sim\mu_1, \quad x(1)\sim\mu_2 \end{aligned}$$

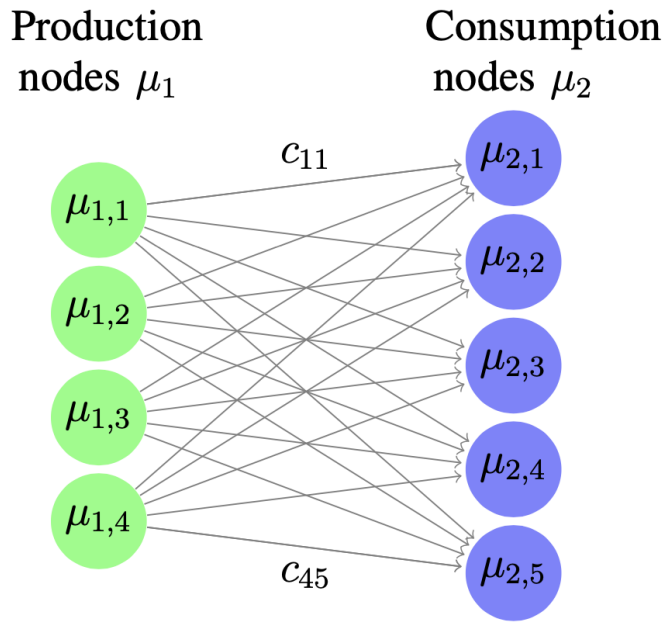
Discrete optimal transport

$$\min_{\mathbf{B}\in\mathbb{R}_+^{\mathtt{d}_1\times\mathtt{d}_2}} \quad \operatorname{trace}(\mathbf{C}^\mathsf{T}\mathbf{B})$$

$$\text{subject to } \mathbf{B}\mathbf{1} = \boldsymbol{\mu}_1$$

$$\mathbf{B}^\mathsf{T}\mathbf{1} = \boldsymbol{\mu}_2$$





Sinkhorn algorithm

Entropy regularization $\mathcal{H}(\mathbf{B}) = -\sum_{i_1,i_2} \mathrm{B}(i_1,i_2) \ln\ \mathrm{B}(i_1,i_2)$

$$\min_{\mathbf{B}\in\mathbb{R}_+^{d_1\times d_2}} \quad \mathrm{trace}(\mathbf{C}^\mathsf{T}\mathbf{B}) - \epsilon\mathcal{H}(\mathbf{B})$$

$$\text{subject to } \mathbf{B}\mathbf{1} = \boldsymbol{\mu}_1$$

$$\mathbf{B}^\mathsf{T}\mathbf{1} = \boldsymbol{\mu}_2$$

Sinkhorn iterations (kernel $\mathbf{K} = \exp(-\mathbf{C}/\epsilon)$)

$$\begin{aligned} \mathbf{u}_2 &= \boldsymbol{\mu}_2./(\mathbf{K}^\mathsf{T}\mathbf{u}_1) \\ \mathbf{u}_1 &= \boldsymbol{\mu}_1./(\mathbf{K}\mathbf{u}_2) \end{aligned}$$

Optimal solution

$$\mathbf{B} = \mathrm{diag}(\mathbf{u}_1)\,\mathbf{K}\,\mathrm{diag}(\mathbf{u}_2)$$

Multi-marginal optimal transport

$$\min_{\mathbf{B}\in\mathbb{R}_+^{d_1\times\cdots\times d_J}} \langle \mathbf{C}, \mathbf{B}\rangle$$
$$\text{subject to } P_j(\mathbf{B}) = \boldsymbol{\mu}_j, \text{ for } j \in \Gamma$$

Projection

$$P_j(\mathbf{B}) = \sum_{i_1,\ldots,i_{j-1},i_{j+1},i_J} B(i_1,\ldots,i_{j-1},i_j,i_{j+1},\ldots,i_J)$$

- Only some marginals are given ($\Gamma \subset \{1,2,\ldots,J\}$)
- C, B are J-dimensional tensors
- Difficulty to solve; number of variables d^J

Structured cost

Graphical structure: e.g., Wasserstein barycenter

$$C(\mathbf{x}) = \sum_{i \neq 1} c(x_i, x_1)$$

other structures

- Low rank structure: C is a low rank tensor
- Sparse cost: C is a sparse tensor

Graphical optimal transport

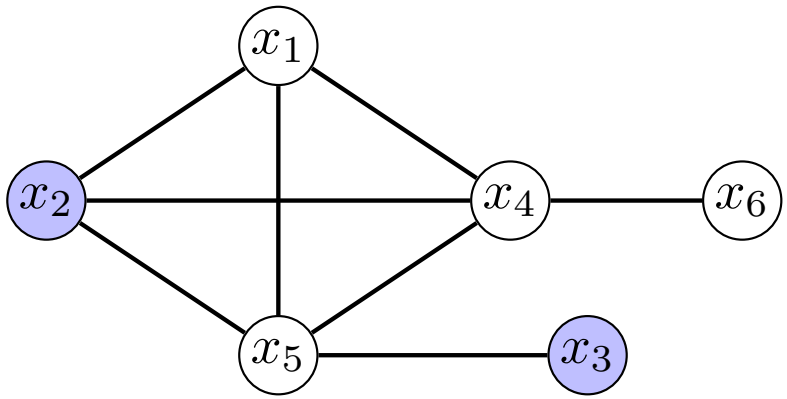
Multi-marginal optimal transport

$$\min_{\mathbf{B}\in\mathbb{R}_+^{d_1\times\cdots\times d_J}} \langle \mathbf{C}, \mathbf{B}\rangle$$
$$\text{subject to } P_j(\mathbf{B}) = \boldsymbol{\mu}_j, \text{ for } j \in \Gamma$$

with graphical cost (each α is a subset of $\{1, 2, \ldots, J\}$)

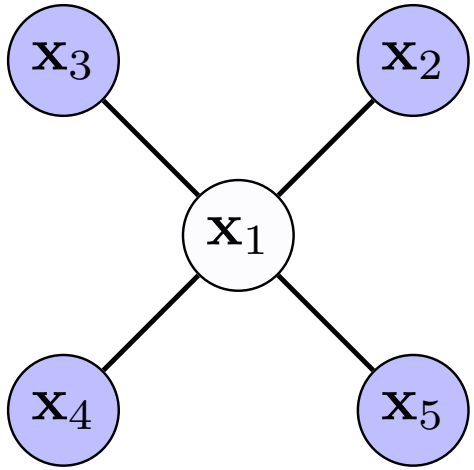
$$C(\mathbf{x}) = \sum_{\alpha\in F} C_\alpha(\mathbf{x}_\alpha)$$

$J = 6, \Gamma = \{2, 3\}, F = \{\{1, 2\}, \{1, 4\}, \{1, 5\}, \{2, 4\}, \{2, 5\}, \{4, 5\}, \{4, 6\}, \{3, 5\}\}$

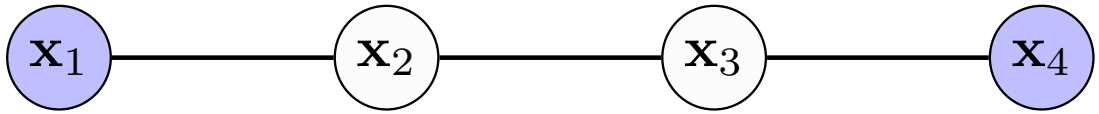


Examples

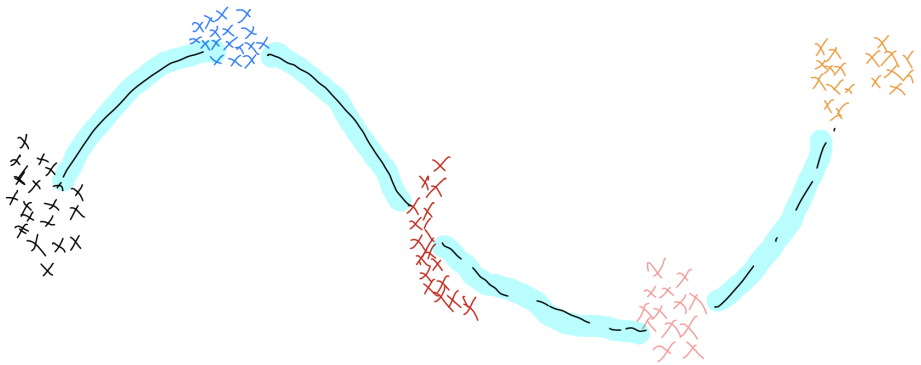
Wasserstein barycenter



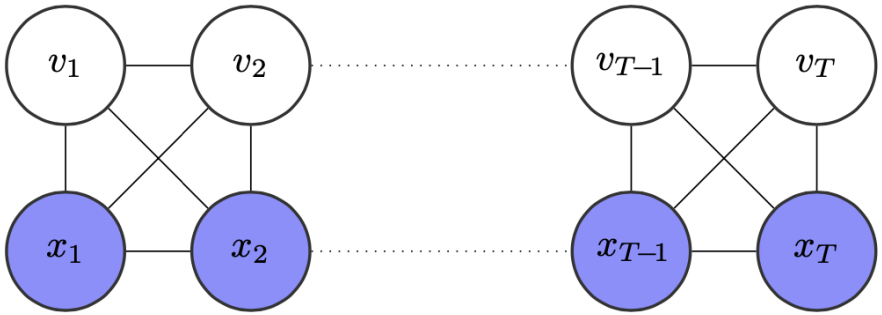
Wasserstein geodesic



Wasserstein spline



Cubic spline in Wasserstein space



How to utilize the graphical structure of the cost?

Entropy regularization

$$\text{Entropy } \mathcal{H}(\mathbf{B}) = -\sum_{i_1,\ldots,i_J} B(i_1,\ldots,i_J) \ln\ B(i_1,\ldots,i_J)$$

$$\min_{\mathbf{B}\in\mathbb{R}^{d_1\times\cdots\times d_J}} \langle \mathbf{C}, \mathbf{B}\rangle - \epsilon\mathcal{H}(\mathbf{B})$$
$$\text{subject to } P_j(\mathbf{B}) = \boldsymbol{\mu}_j, \text{ for } j \in \Gamma$$

Optimal solution $\mathbf{B} = \mathbf{K} \odot \mathbf{U}$

$$\begin{aligned}
\mathbf{K} &= \exp(-\mathbf{C}/\epsilon) \\
\mathbf{U} &= \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_J \\
\mathbf{u}_j &= \begin{cases} \exp\left(-\frac{1}{J} - \frac{\boldsymbol{\lambda}_j}{\epsilon}\right), & \text{if } j \in \Gamma \\ \exp\left(-\frac{1}{J}\right)\mathbf{1}, & \text{otherwise} \end{cases}
\end{aligned}$$

Benamou et al 15

Sinkhorn algorithm as dual block ascent

Dual problem

$$\max_{\{\boldsymbol{\lambda}_j, j\in\Gamma\}} -\epsilon\langle \mathbf{K}, \mathbf{U}\rangle - \sum_{j\in\Gamma} \boldsymbol{\lambda}_j^\top \boldsymbol{\mu}_j$$

Sinkhorn algorithm

1. Compute $\mathbf{U} = \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_J$
2. Update $\mathbf{u}_j$ as $\mathbf{u}_j \leftarrow \mathbf{u}_j \odot \boldsymbol{\mu}_j ./ P_j(\mathbf{K} \odot \mathbf{U})$

Advantages

- global linear convergent
- easily parallelizable

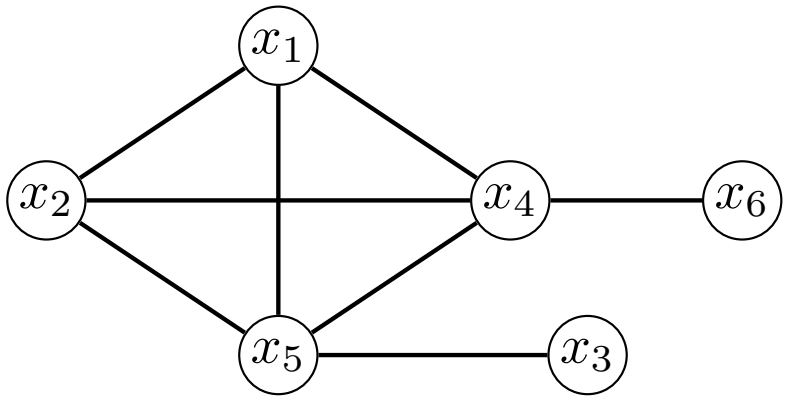
Luo & Tseng 92, Di Marino & Gerolin 20, Carlier 21

bottleneck: projection $P_j(\mathbf{K} \odot \mathbf{U})$ ($O(d^J)$ complexity)

Probabilistic graphical models

Markov random field (undirected probabilistic graphical models)

$$p(\mathbf{x}) := p(x_1, x_2, \ldots, x_J) = \frac{1}{Z} \prod_{j \in V} \phi_j(x_j) \prod_{\alpha \in F} \psi_\alpha(\mathbf{x}_\alpha)$$



$$p(x_1, \ldots, x_6) = \frac{1}{Z} \psi_{\alpha_1}(\mathbf{x}_{\alpha_1}) \psi_{\alpha_2}(\mathbf{x}_{\alpha_2}) \psi_{\alpha_3}(\mathbf{x}_{\alpha_3}) \prod_{j=1}^{6} \phi_j(x_j)$$

factors: $\mathbf{x}_{\alpha_1} = \{x_1, x_2, x_4, x_5\}$, $\mathbf{x}_{\alpha_2} = \{x_3, x_5\}$, $\mathbf{x}_{\alpha_3} = \{x_4, x_6\}$

Bayesian inference

Bayesian marginal inference

$$p(x_j) = \sum_{x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_J} p(\mathbf{x})$$

- Brutal force computation is expensive
- There are efficient ways to compute the marginals such as variable elimination and belief propagation

Projection as Bayesian inference

Bottleneck: Projection $P_j(\mathbf{K} \odot \mathbf{U})$

$$\mathbf{K} = \exp(-\mathbf{C}/\epsilon), \quad \mathbf{U} = \mathbf{u}_1 \otimes \mathbf{u}_2 \otimes \cdots \otimes \mathbf{u}_J$$

Probabilistic graphical model

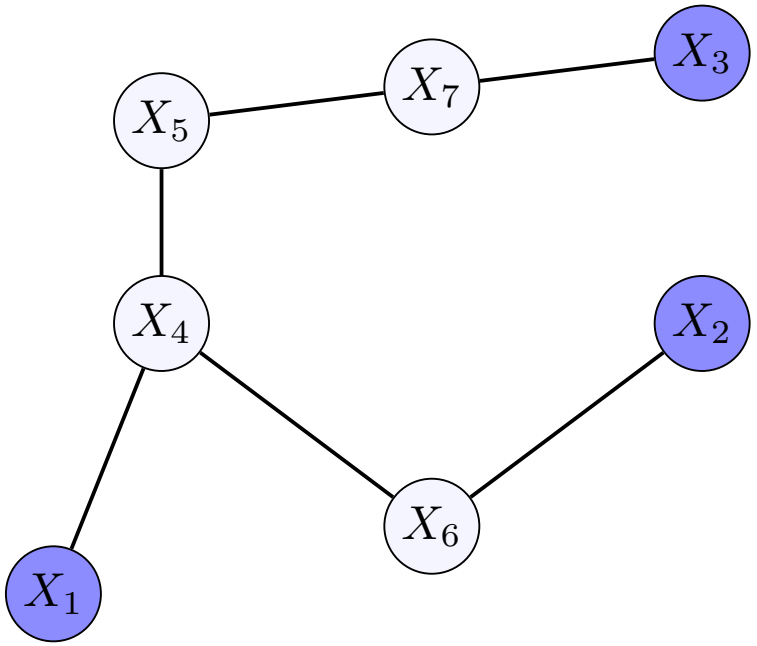
$$\begin{aligned} p(\mathbf{x}) &= \prod_{\alpha \in F} \exp(-C_\alpha/\epsilon)(\mathbf{x}_\alpha) \prod_{j \in V} u_j(\mathbf{x}_j) \\ &= \frac{1}{Z} \prod_{\alpha \in F} \psi_\alpha(\mathbf{x}_\alpha) \prod_{j \in V} \phi_j(x_j) \end{aligned}$$

$$\boxed{P_j(\mathbf{K} \odot \mathbf{U}) = p(x_j)}$$

Tree-structured OT

$$\min_{\mathbf{B}\in\mathbb{R}_+^{d_1\times\cdots\times d_J}} \langle \mathbf{C}, \mathbf{B}\rangle$$
$$\text{subject to } P_j(\mathbf{B}) = \boldsymbol{\mu}_j, \text{ for } j \in \Gamma$$

- Tree-structured cost $C(\mathbf{x}) = \sum_{\alpha\in F} C_\alpha(\mathbf{x}_\alpha)$
- Marginal constraints on leaf nodes



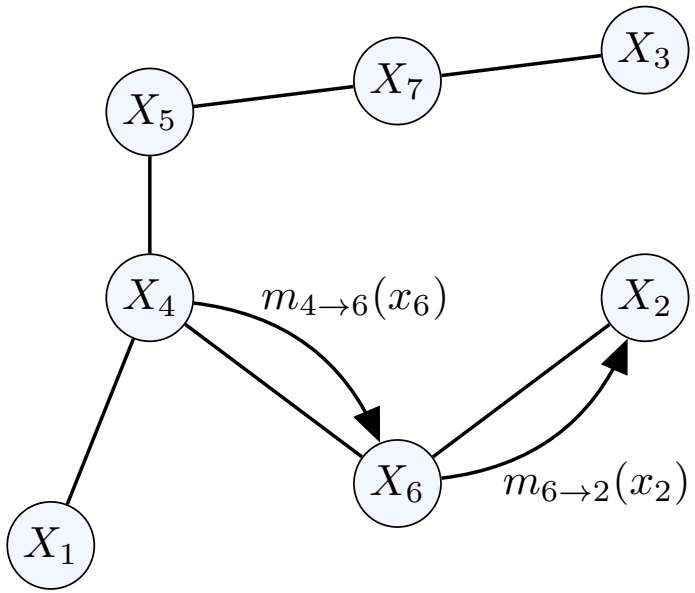
Belief propagation

message passing for $p(\mathbf{x}) = \frac{1}{Z} \prod_{j \in V} \phi_j(x_j) \prod_{\alpha \in F} \psi_\alpha(\mathbf{x}_\alpha)$

$$m_{i \to j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \phi_i(x_i) \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i)$$

marginal distribution

$$b_i(x_i) \propto \phi_i(x_i) \prod_{k \in N(i)} m_{k \to i}(x_i)$$



Sinkhorn belief propagation

Apply belief propagation to compute $\mathrm{P_j}(\mathbf{K} \odot \mathbf{U}) = \mathrm{p}(x_j)$

Sinkhorn belief propagation

1. update

$$m_{i \to j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \frac{\mu_i(x_i)}{m_{j \to i}(x_i)}$$

2. update all the messages on the path from i to i_{next} with

$$m_{i \to j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \prod_{k \in N(i) \setminus j} m_{k \to i}(x_i)$$

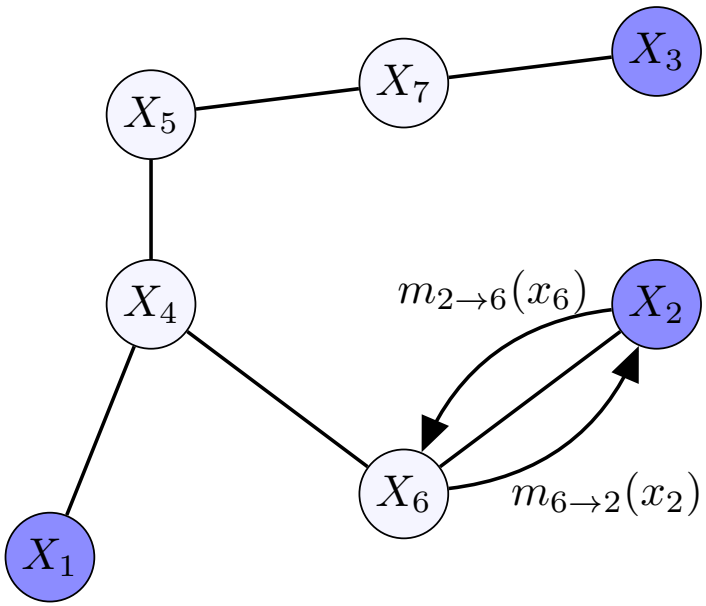
3. $i = i_{\text{next}} \in \Gamma$

Teh & Welling 01

Sinkhorn belief propagation

Sinkhorn belief propagation

1. bouncing messages
   $m_{i\to j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \frac{\mu_i(x_i)}{m_{j\to i}(x_i)}$
2. traditional messages
   $m_{i\to j}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j) \prod_{k\in N(i)\setminus j} m_{k\to i}(x_i)$



Complexity

Sinkhorn belief propagation with random order

Tree structure (graph diameter D)

$$\tilde{\mathcal{O}}\left(\frac{D J d^2}{\delta^2}\right)$$

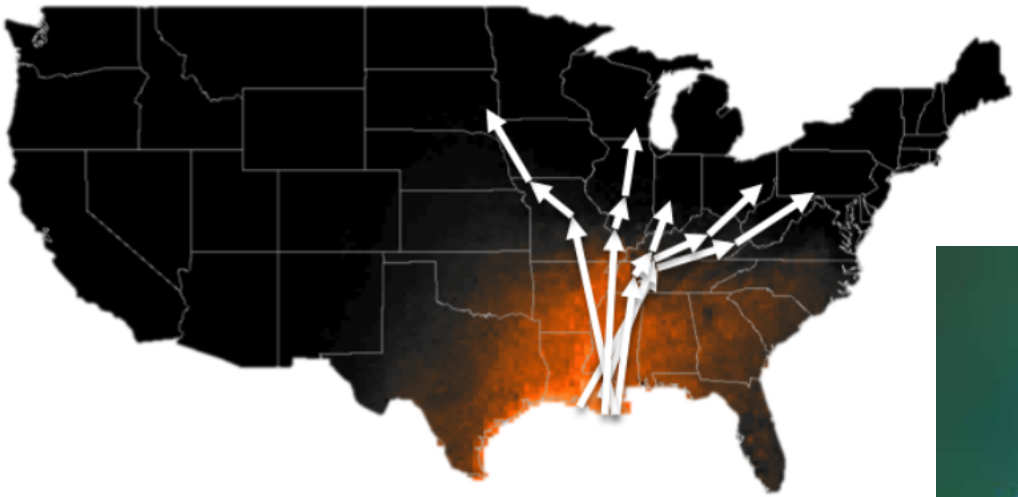
General graph (tree-width $w(G)$)

$$\tilde{\mathcal{O}}\left(\frac{D J d^{w(G)+1}}{\delta^2}\right)$$

Fan et al 21

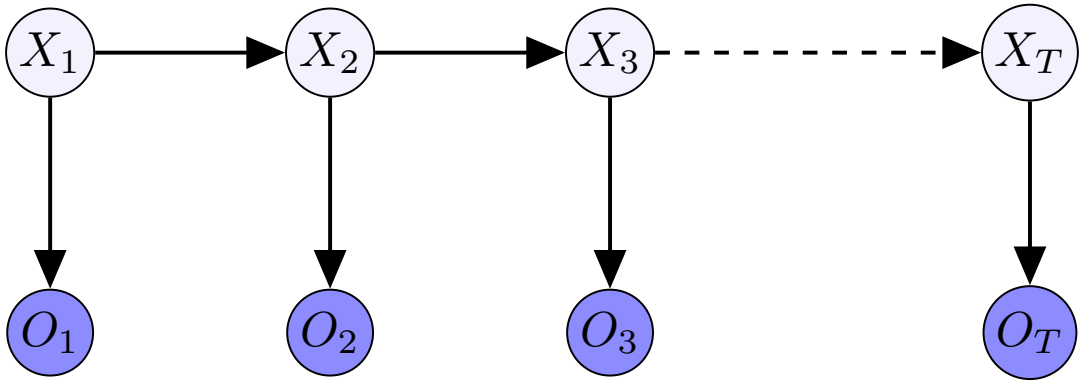
Inference with aggregate observations

- Privacy preservation, cost of data collection, measurement fidelity
- Examples: bird migration, disease spread, human mobility





Inference for aggregate hidden Markov models



$$\begin{aligned} X_{t+1} &\sim p(X_{t+1}|X_t), \quad X_1 \sim p(X_1) \\ O_t &\sim p(O_t|X_t) \end{aligned}$$

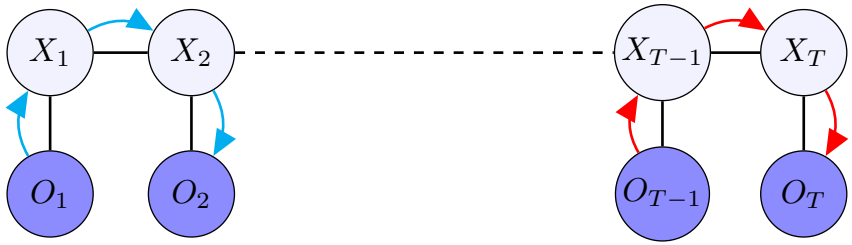
Given: measurements are the empirical distributions of M individuals

$$O_1 = \frac{1}{M}\sum_{i=1}^{M} \delta_{o_1^i}, O_2 = \frac{1}{M}\sum_{i=1}^{M} \delta_{o_2^i}, \ldots, O_T = \frac{1}{M}\sum_{i=1}^{M} \delta_{o_T^i}$$

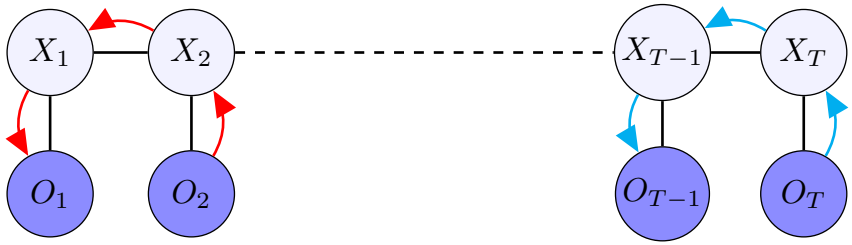
Goal: the posterior marginal distribution of X_t

Collective forward-backward algorithm

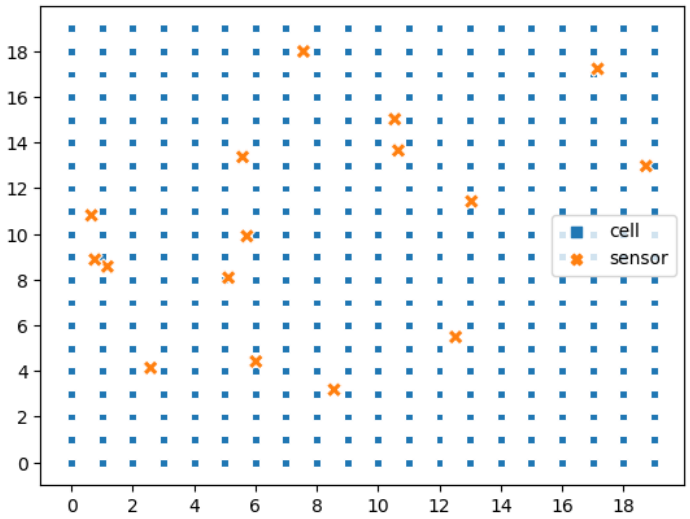
Forward pass

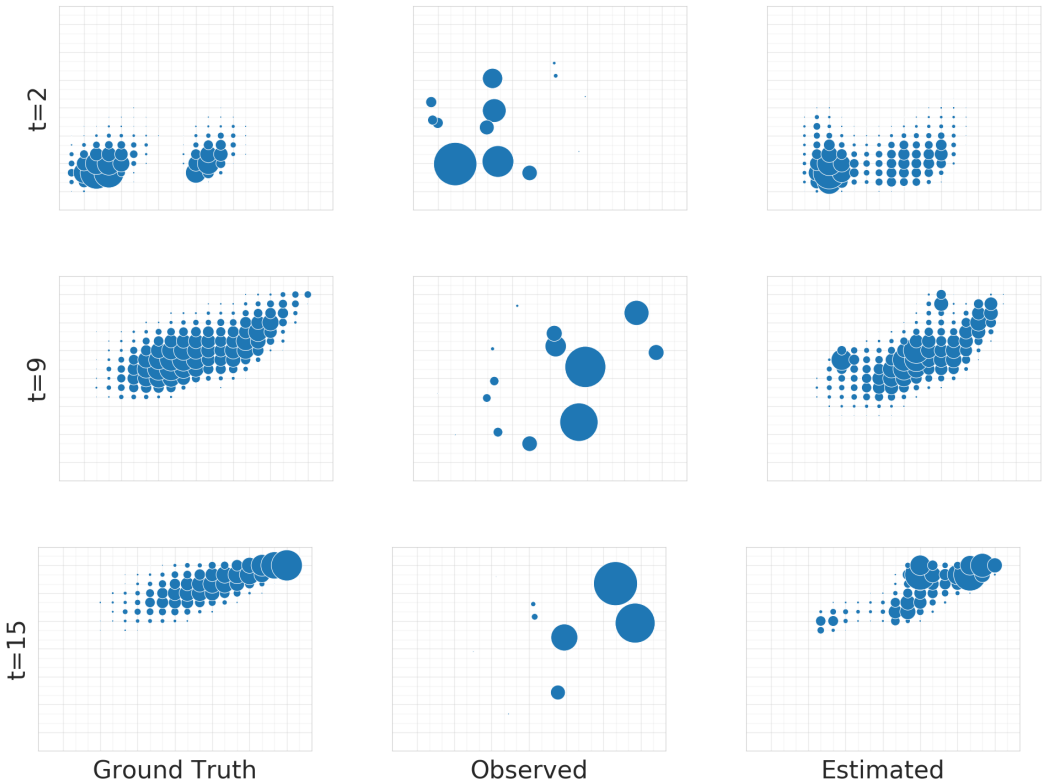


Backward pass

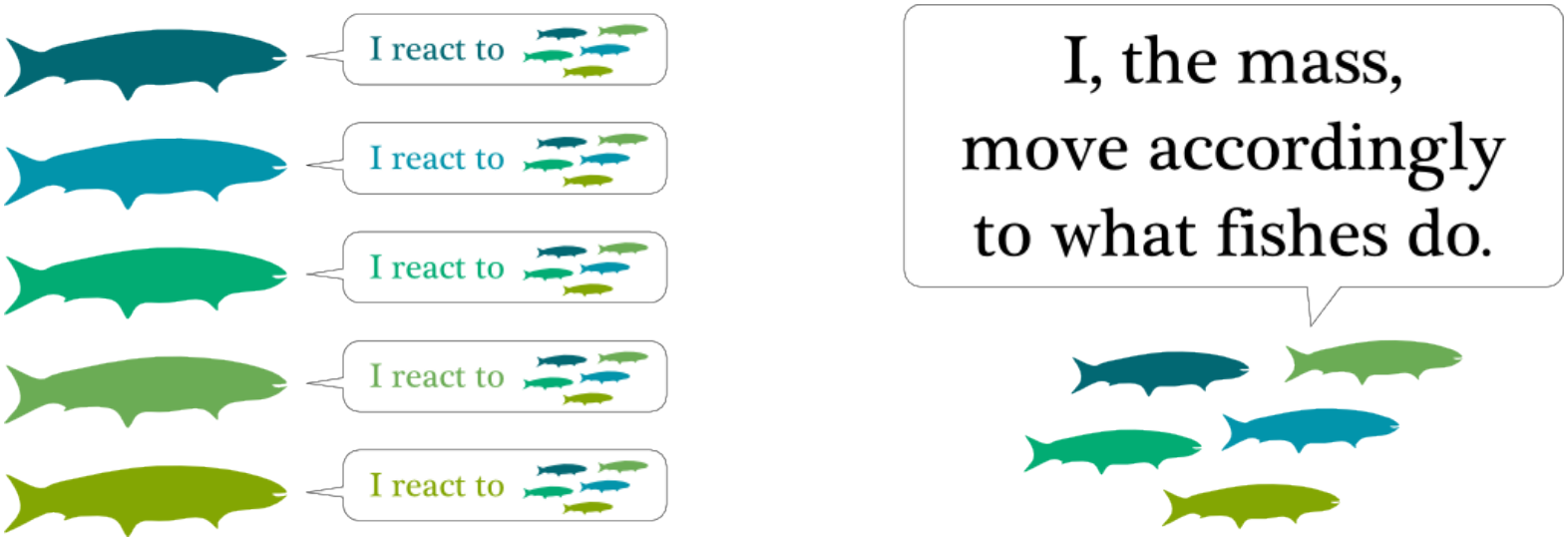


Inference for aggregate hidden Markov models





Mean field control/games



Use distributions to approximate group behavior

Mean field control/games

individual dynamics

$$dx(t) = f(x(t))dt + B(x(t))\big(v(x(t),t)dt + \sqrt{\epsilon}dw\big)$$

density control formulation

$$\min_{\rho,v} \int_0^1 \int \frac{1}{2}\|v\|^2 \rho dx dt + \int_0^1 F_t(\rho(t,\cdot))dt + G(\rho(1,\cdot))$$

$$\text{s.t. } \frac{\partial \rho}{\partial t} + \nabla \cdot ((f + Bv)\rho) - \frac{\epsilon}{2}\sum_{i,k=1}^{n} \frac{\partial^2(\sigma_{ik}\rho)}{\partial x_i \partial x_k} = 0$$

$$\rho(0,\cdot) = \rho_0$$

Mean field control/games

Girsanov theorem

$$\mathrm{KL}(\mathcal{P}^v \| \mathcal{P}^0) = \frac{1}{2\epsilon}\mathbb{E}_{\mathcal{P}^v}\{\int_0^1 \|v\|^2 dt\} = \frac{1}{2\epsilon}\int\int_0^1 \|v\|^2 \rho dt dx$$

Reformulation over path distribution $\mathcal{P}^v$

$$\begin{aligned} \min_{\mathcal{P}^v} \quad & \epsilon\, \mathrm{KL}(\mathcal{P}^v \| \mathcal{P}^0) + \int_0^1 F_t(\mathcal{P}_t^v) dt + G(\mathcal{P}_1^v) \\ \text{s.t.} \quad & \mathcal{P}_0^v = \rho_0 \end{aligned}$$

Graphical optimal transport reformulation

discretization over space-time

$$\begin{aligned} \min_{\substack{\mathbf{M}\in\mathbb{R}_+^{N^{\mathcal{T}+1}},\\ \mu_1,\dots,\mu_{\mathcal{T}}\in\mathbb{R}_+^N}} \quad & \langle \mathbf{C}, \mathbf{M}\rangle - \epsilon\mathcal{H}(\mathbf{M}) + \Delta t \sum_{j=1}^{\mathcal{T}-1} F_j(\mu_j) + G(\mu_{\mathcal{T}}) \\ \text{s.t.} \quad & P_j(\mathbf{M}) = \mu_j,\ j = 1, 2, \dots, \mathcal{T} \\ & P_0(\mathbf{M}) = \rho_0 \end{aligned}$$

graphical cost

$$\mathbf{C}_{i_0,\dots,i_{\mathcal{T}}} = \sum_{j=0}^{\mathcal{T}-1} C_{i_j,i_{j+1}}$$

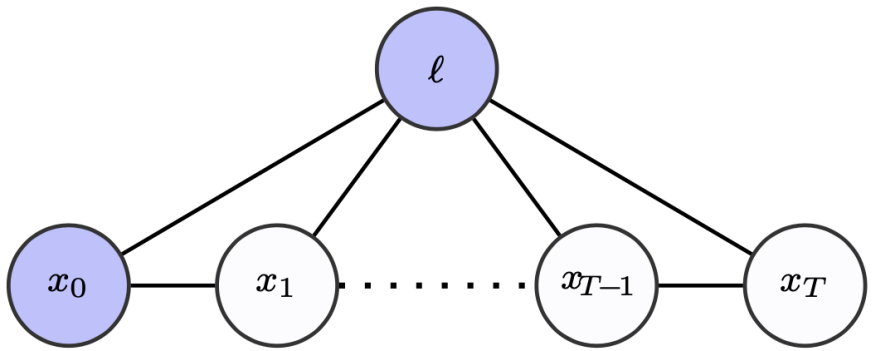
Multi-species Mean field control

$$\min_{\rho,\rho_\ell,v_\ell} \int_0^1 \int \sum_{\ell=1}^{L} \frac{1}{2}\|v_\ell\|^2 \rho_\ell \, dx dt + \int_0^1 F_t(\rho(t,\cdot))dt + G(\rho(1,\cdot))$$

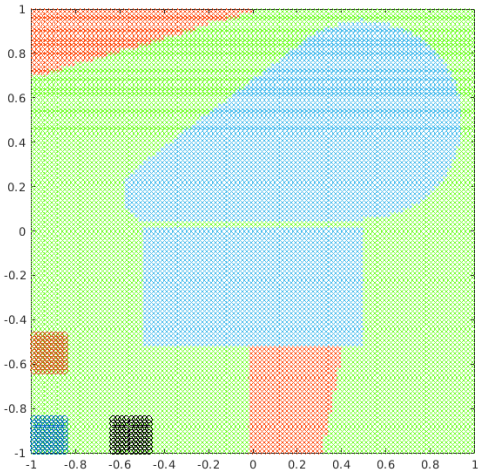
$$+ \int_X \sum_{\ell=1}^{L} \left(\int_0^1 \mathcal{F}_\ell(t,x)\rho_\ell(t,x)dt + \mathcal{G}_\ell(x)\rho_\ell(1,x) \right) dx$$

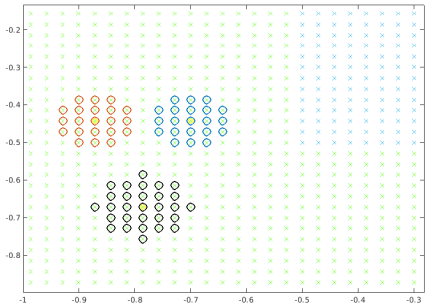
$$\text{s.t.} \quad \frac{\partial \rho_\ell}{\partial t} + \nabla \cdot ((f(x) + B(x)v_\ell)\rho_\ell) - \frac{\epsilon}{2} \sum_{i,k=1}^{n} \frac{\partial^2(\sigma_{ik}\rho_\ell)}{\partial x_i \partial x_k} = 0$$

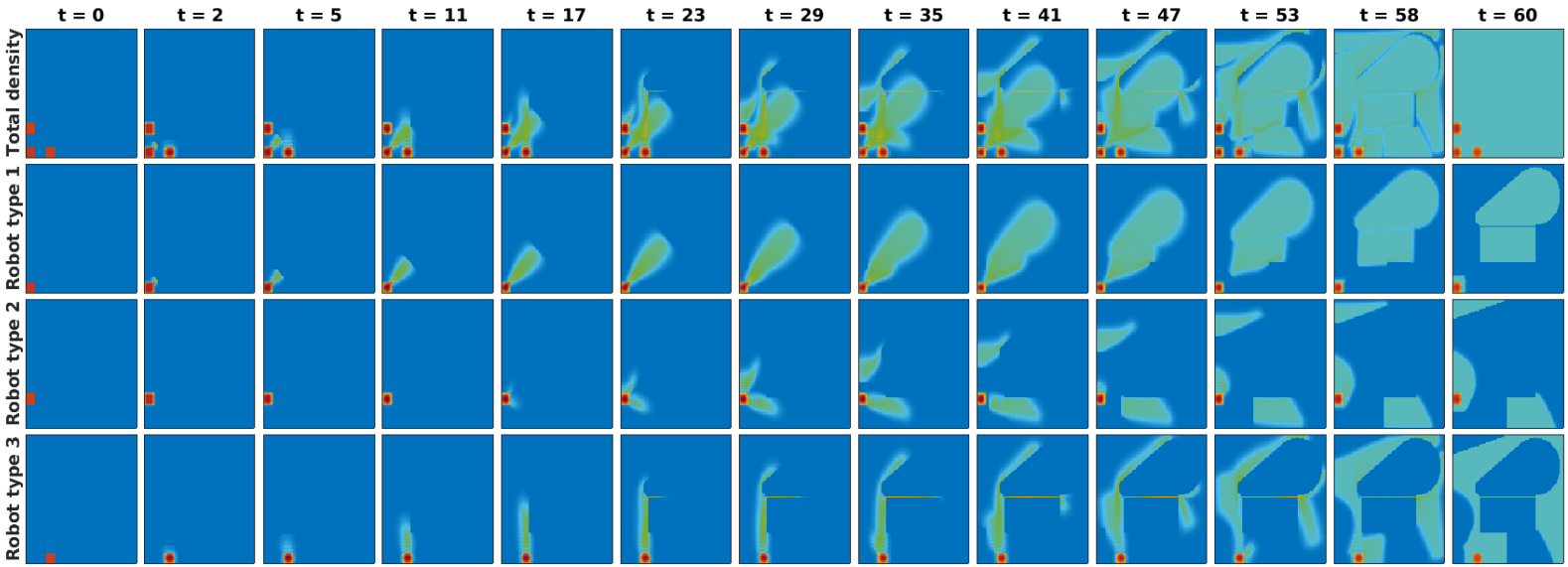
$$\rho_\ell(0,\cdot) = \rho_{0,\ell}, \quad \rho(t,x) = \sum_{\ell=1}^{L} \rho_\ell(t,x)$$



Robot coordination







Takeaway

- Graphical structure of cost in MOT is great
- Fundamental connections between MOT and PGM
- Plenty of applications: estimation/inference, mean-field control, operation research

Papers

1. Multi-marginal Optimal Transport with a Tree-structured Cost and the Schrödinger Bridges Problem
2. Multi-marginal Optimal Transport and Probabilistic Graphical Models
3. Inference with Aggregate Data in Probabilistic Graphical Models: An Optimal Transport Approach
4. Efficient Computations of Multi-species Mean Field Games via Graph-structured Optimal Transport

Acknowledgment








National Science Foundation

Thank you for your attention!