Optimal Transport in Learning, Control, and Dynamical Systems

ICML Tutorial 2023

Charlotte Bunne and Marco Cuturi

ETH zürich

Disclaimers: We cannot cover everything OT is about!

- Optimal transport theory has made popular news' headlines for over a decade.

Four Are Awarded Medal in Mathematics

Dr. Villani, a Frenchman, looked at how order falls apart into disorder in systems like colliding gas molecules.

Fields Medals Awarded to 4 Mathematicians

The prize, bestowed every four years to mathematicians 40 years or younger, is often described as the subject's Nobel Prize.

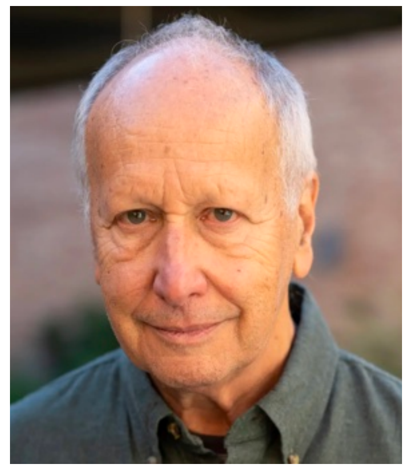
In a 2013 interview with the website Live Science, Dr. Figalli described his work with "optimal transport."

Abel Prize Goes to Mathematician Who Studied Equations That Describe Nature

The honor, like a Nobel Prize for mathematics, was given this year to Luis Caffarelli for his work on partial differential equations.

Villani

Figalli



Caffarelli

- Our focus today: foundations & computations relevant for ML (high-d, GPU, NN).
- We won't be able to cover everything, this is an invitation to read on & experiment with OT.
- This is WIP, we cite many refs but forgot many: please do reach out if we missed yours!

Outline of the Tutorial

Prelude **Warm-Up: Starting with Optimal Matchings**

- OT problems as extensions / generalizations of CS.101 optimal matchings

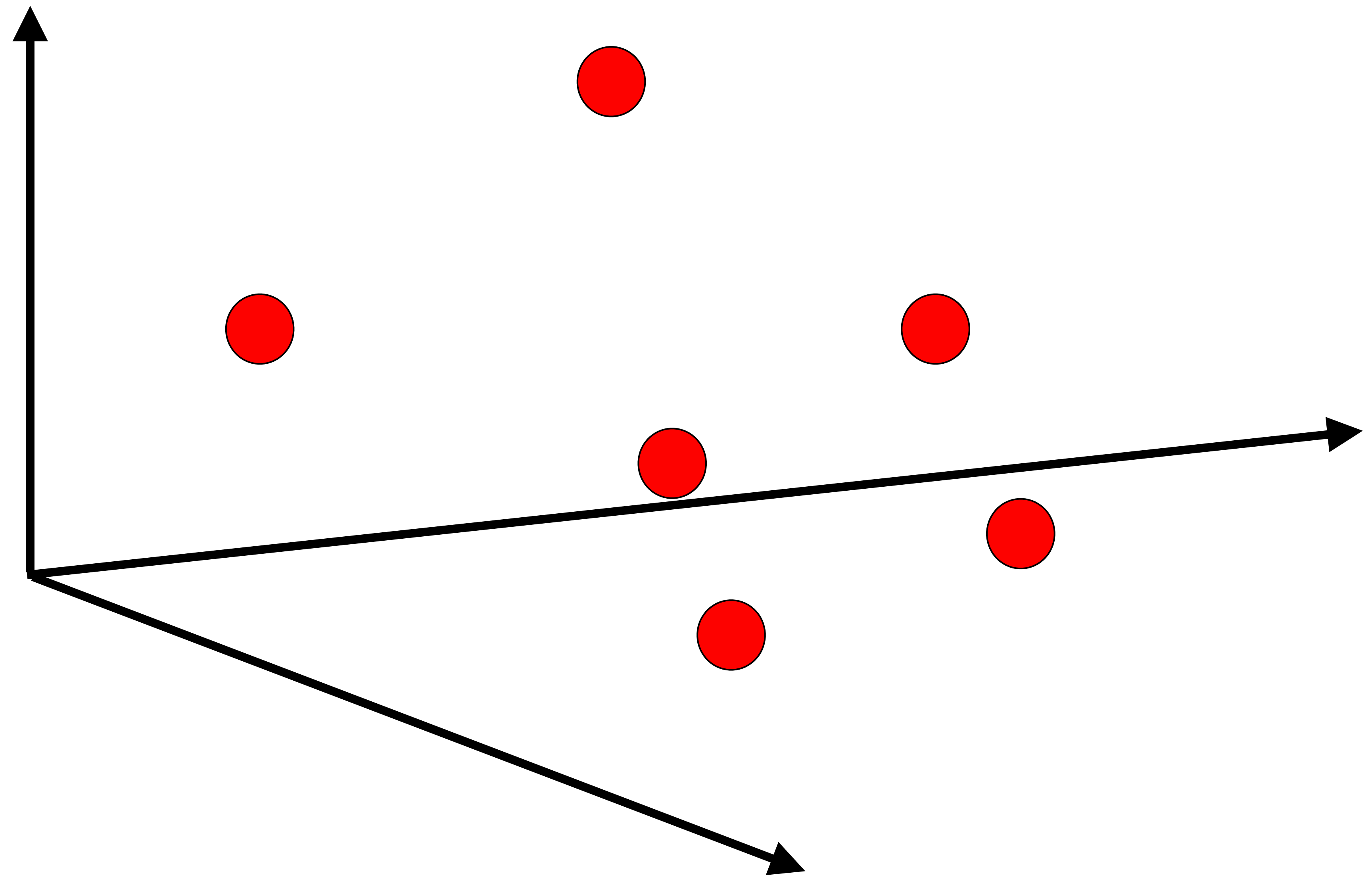
Part 1 Kantorovich Formulation of OT and Computations

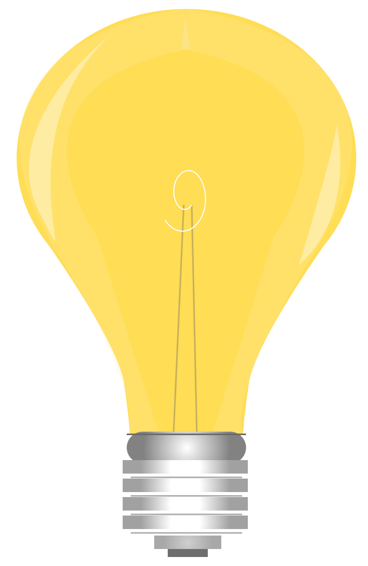
Part 2 Duality, Monge Formulations and Brenier Theorems

Part 3 Modeling Measure Dynamics with Optimal Transport

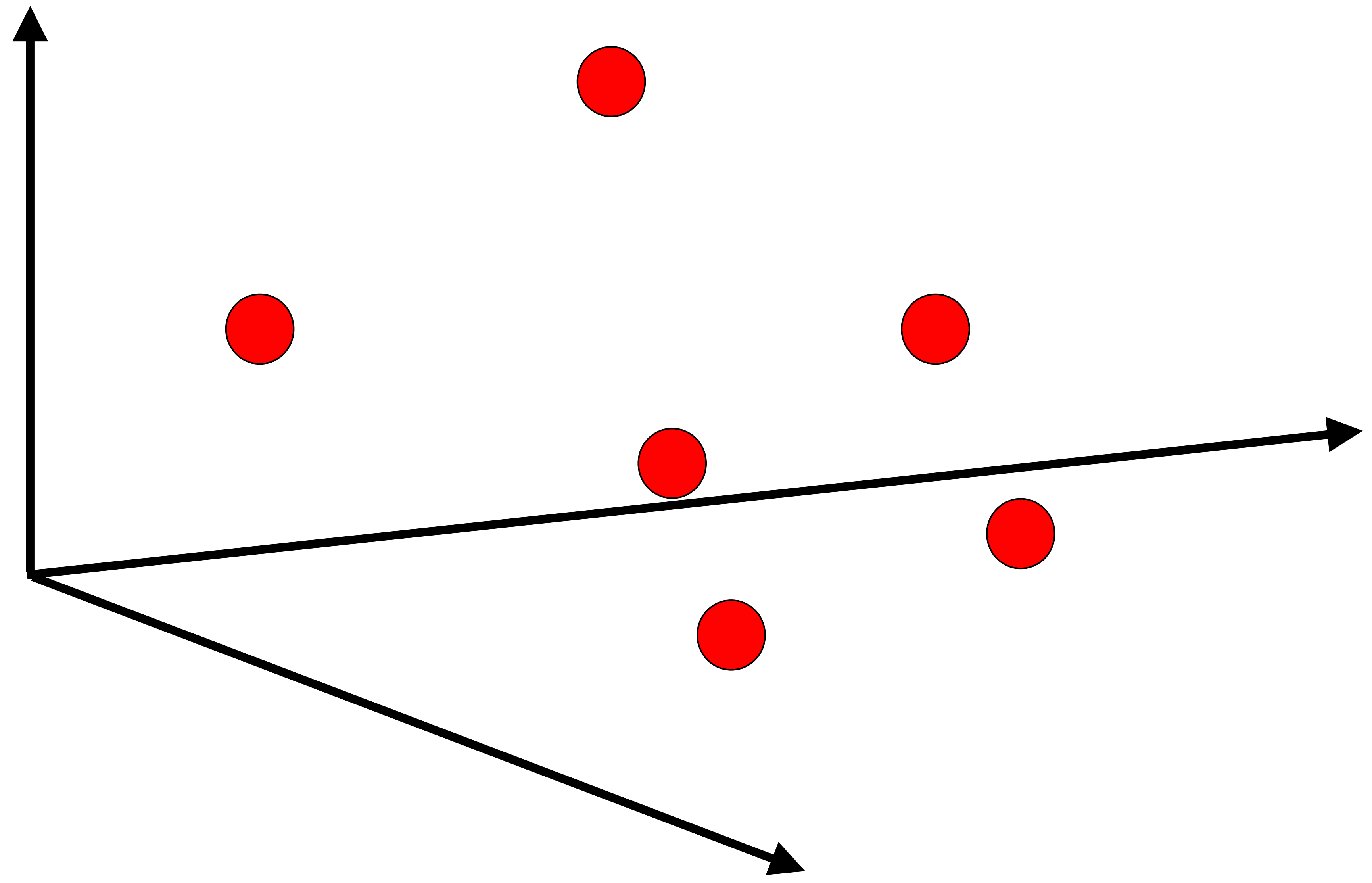
A Puzzle

A Puzzle



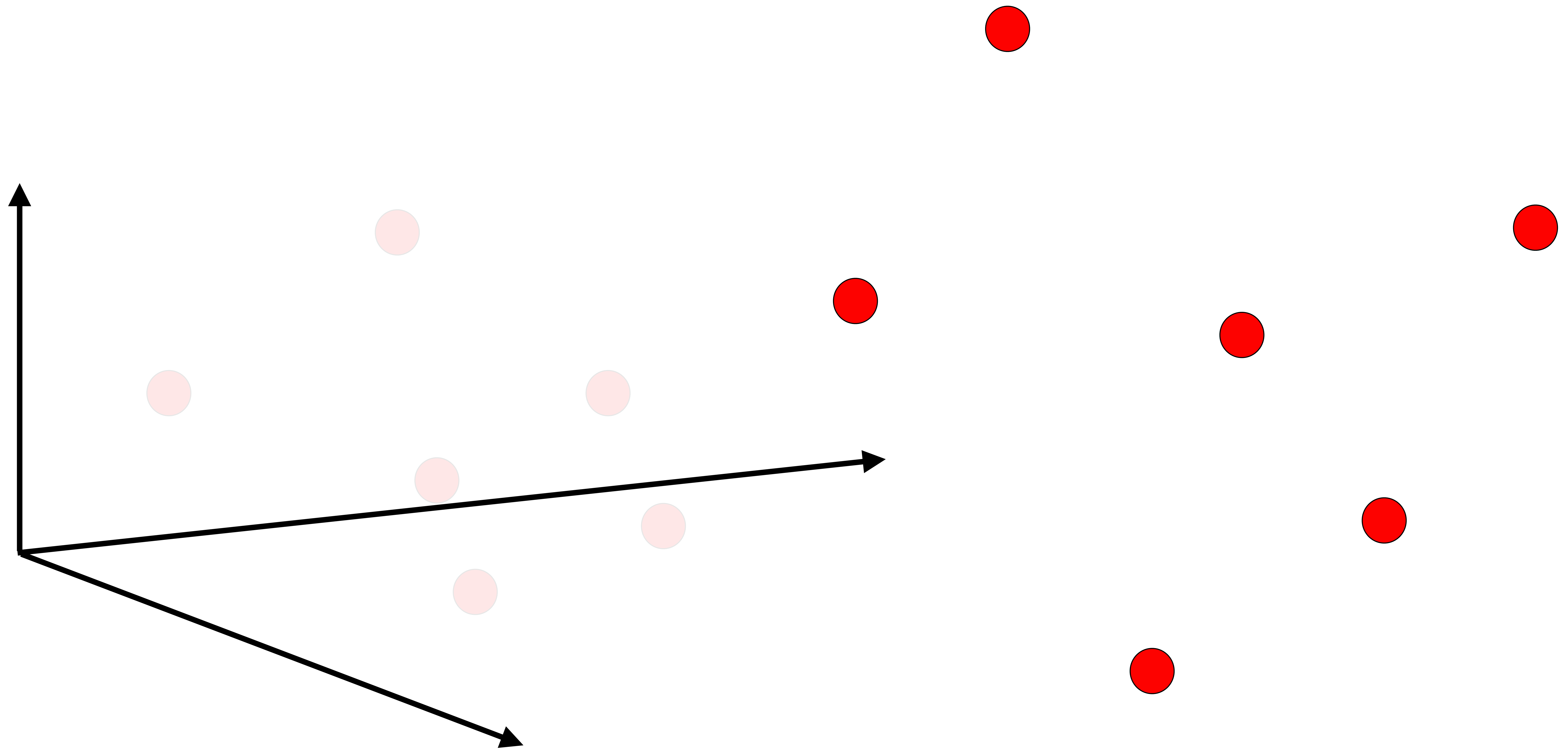


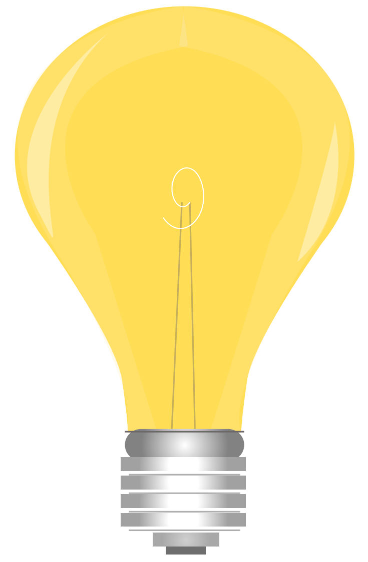
A Puzzle



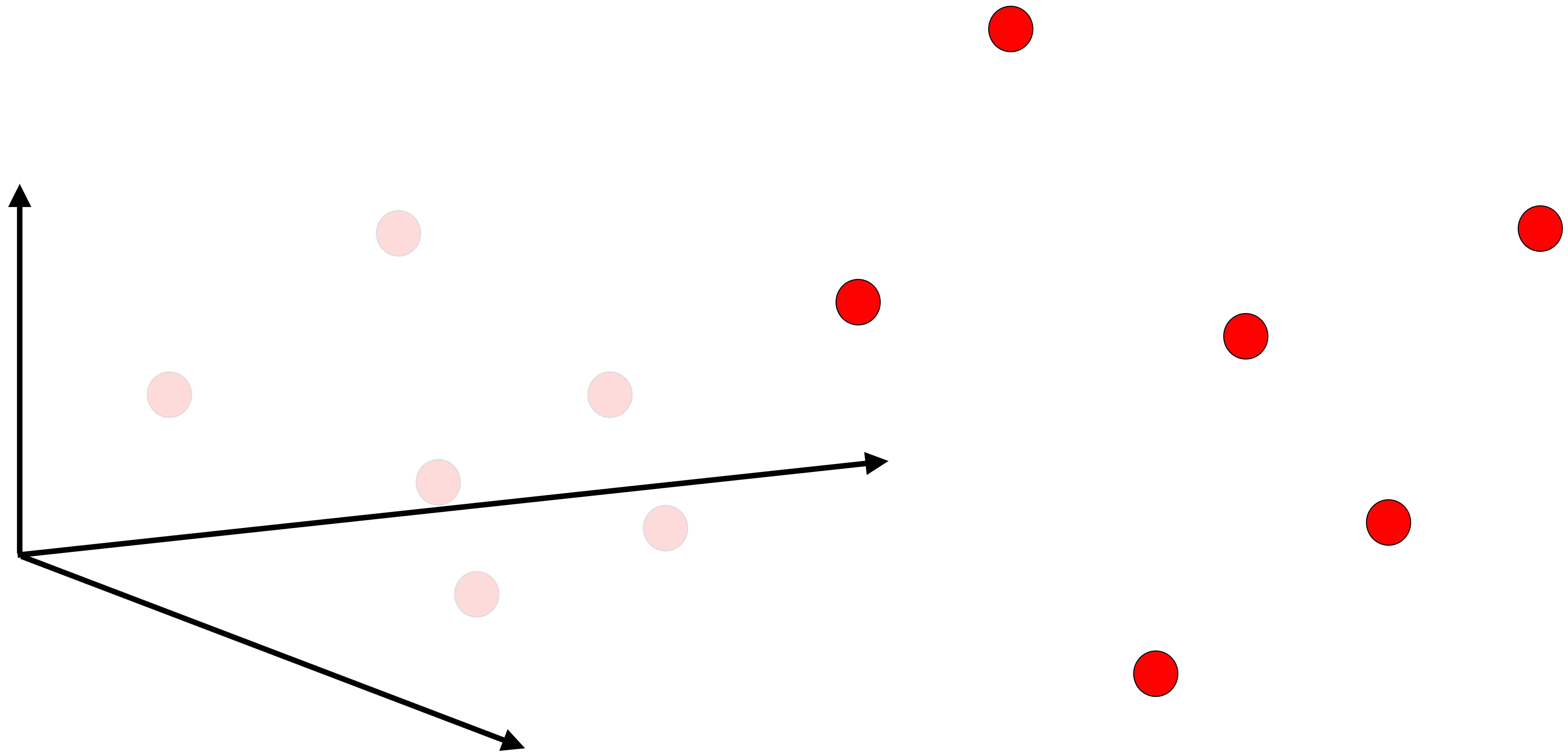
A Puzzle

A Puzzle

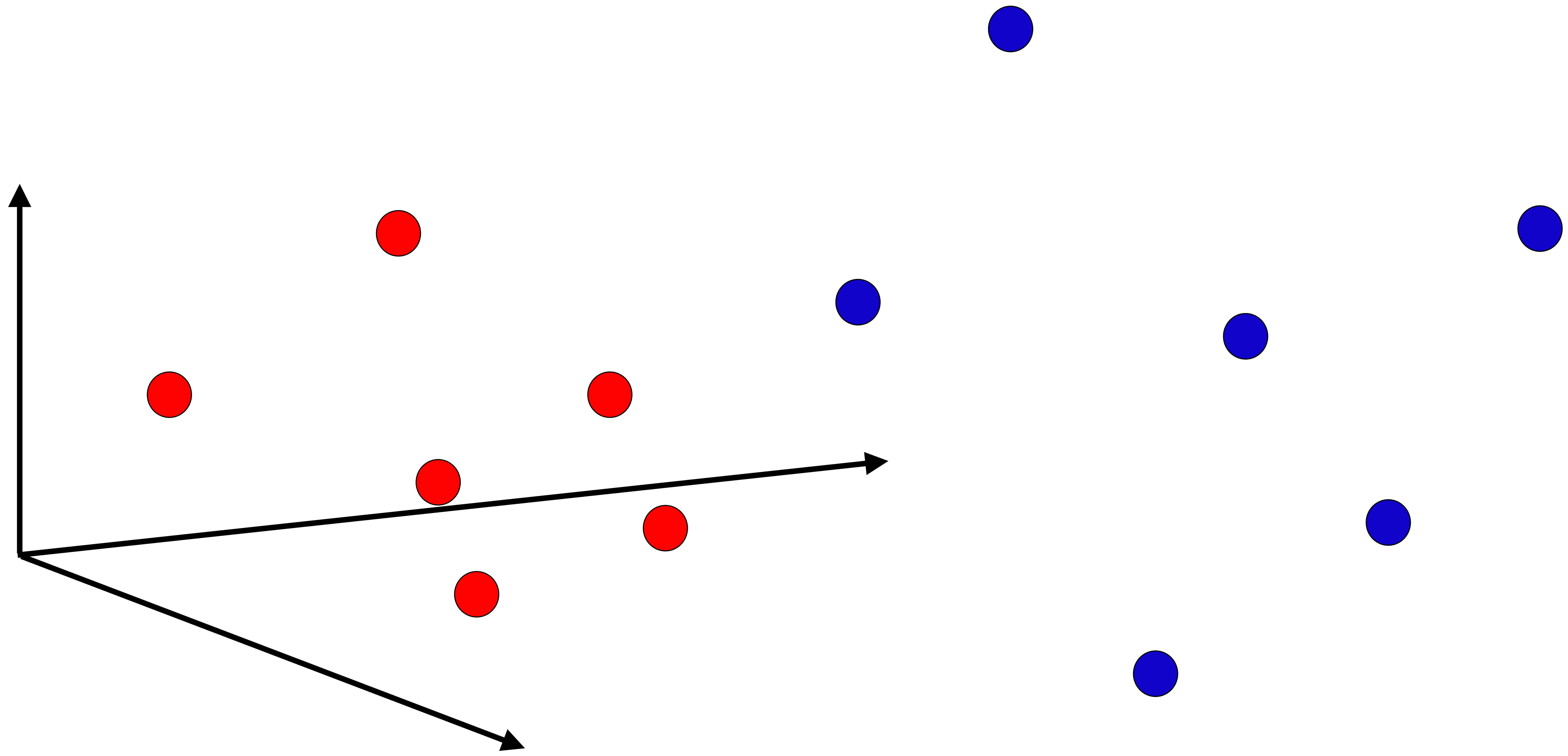




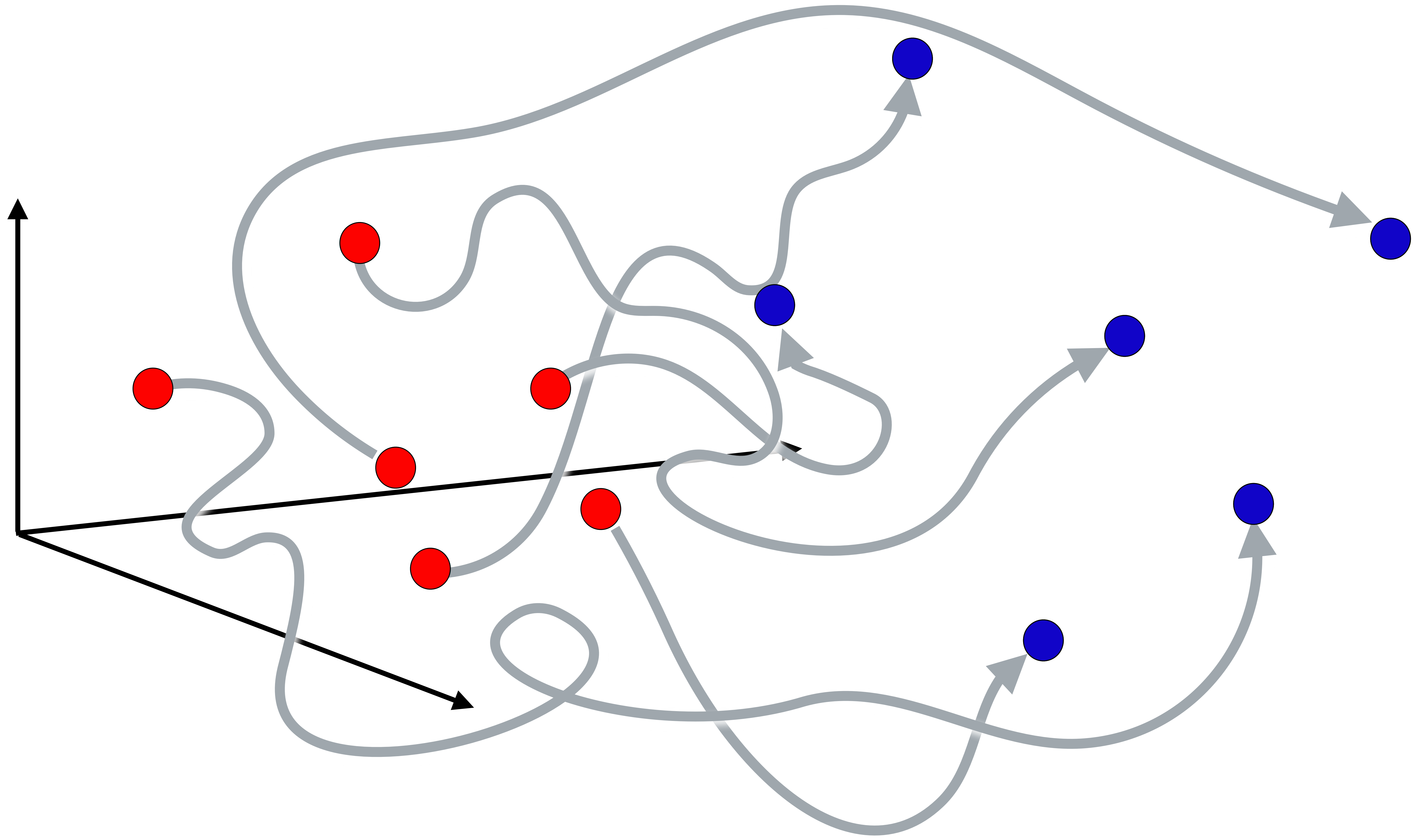
The Puzzle is: Who Went Were?



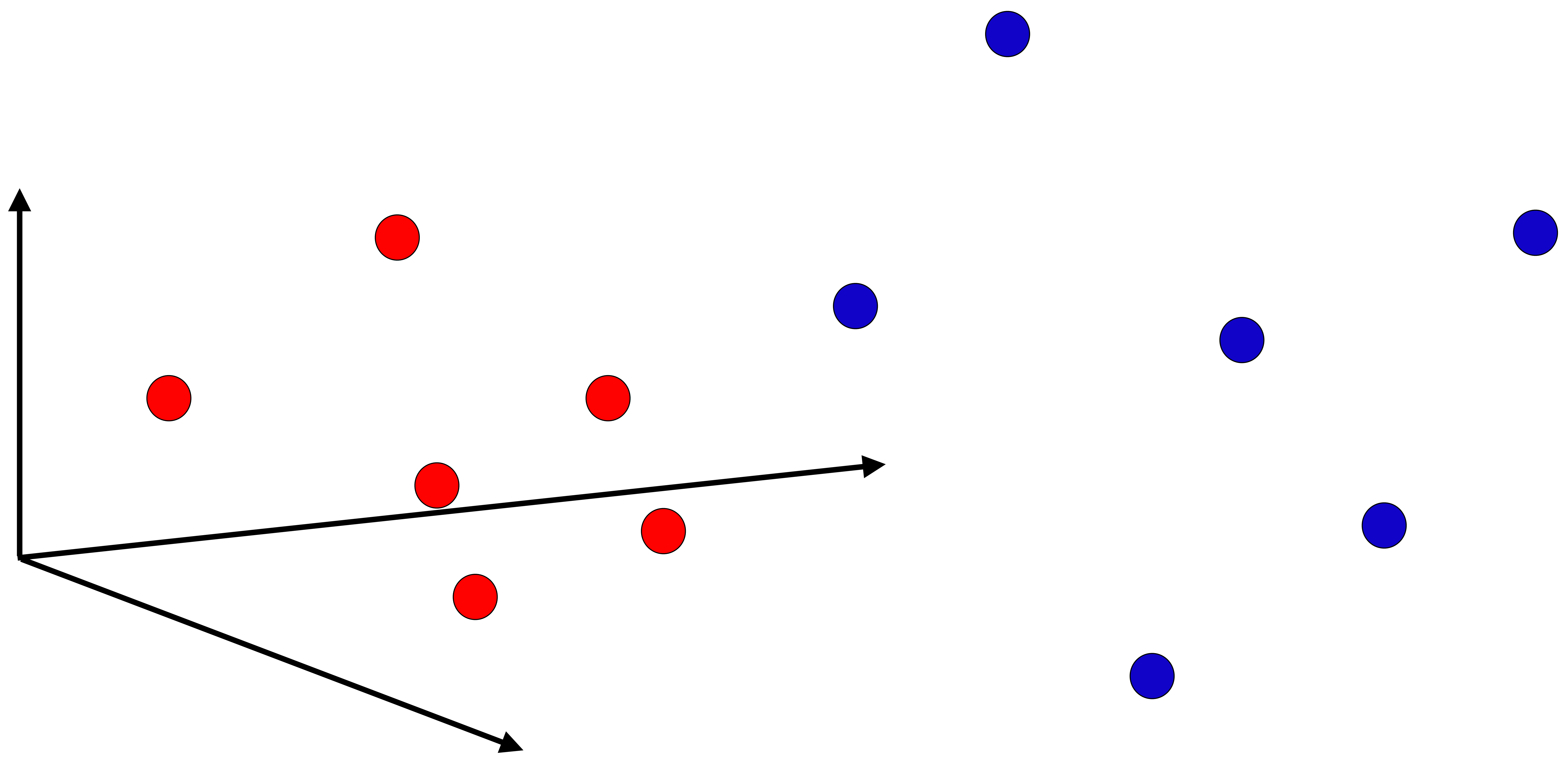
The Puzzle is: Who Went Were?



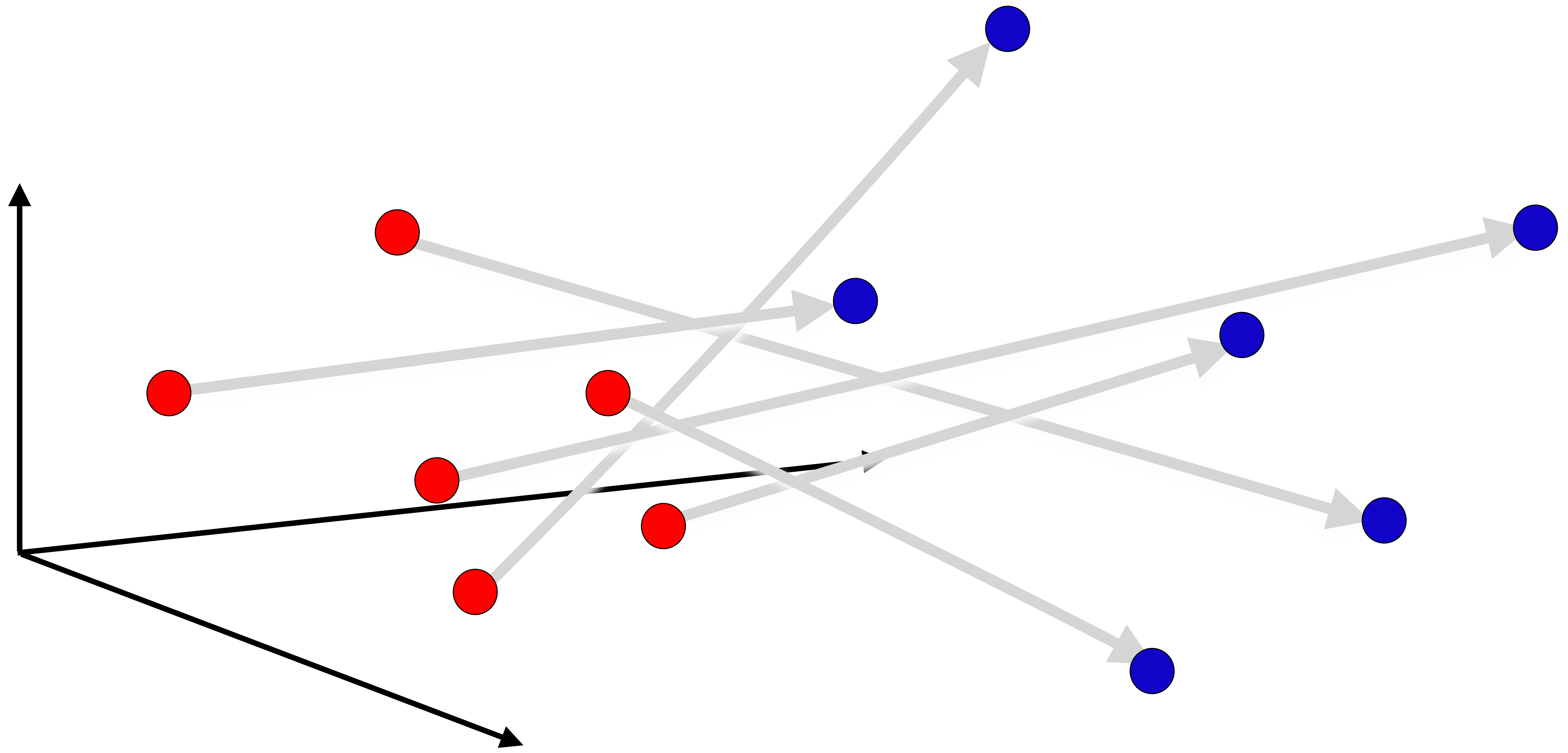
The Puzzle is: Who Went Were?



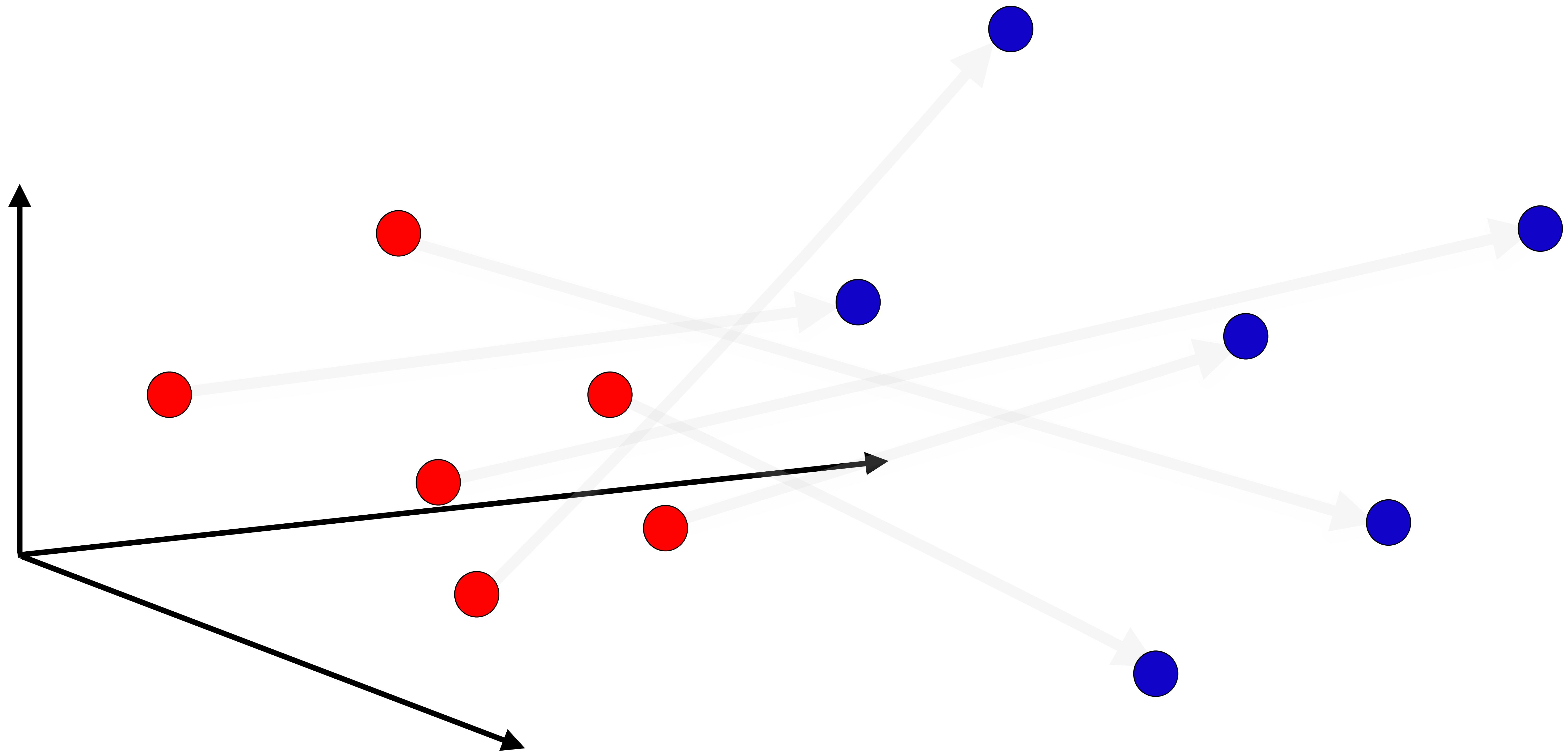
For Now, Assume Particles Moved Along Straight Lines



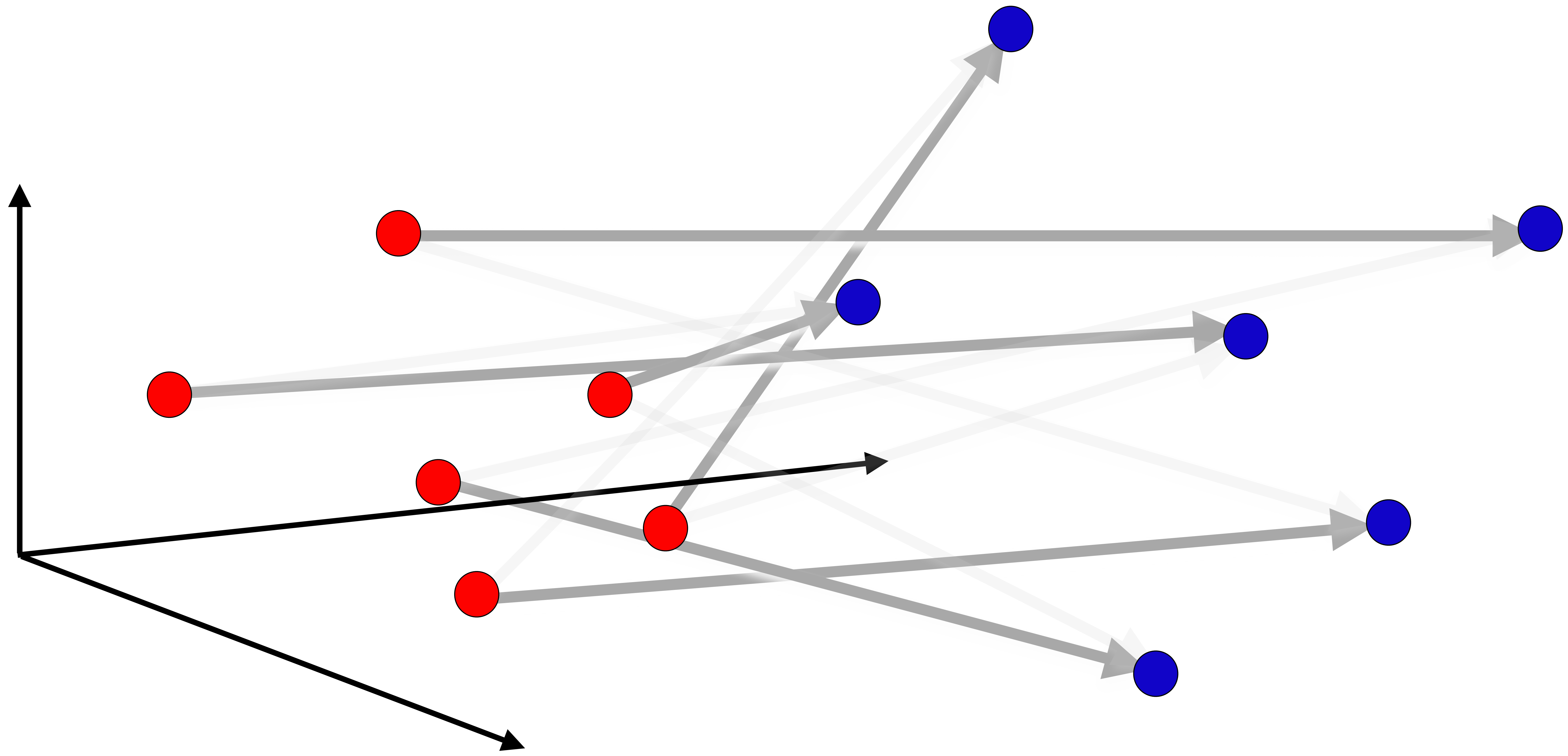
For Now, Assume Particles Moved Along Straight Lines



For Now, Assume Particles Moved Along Straight Lines

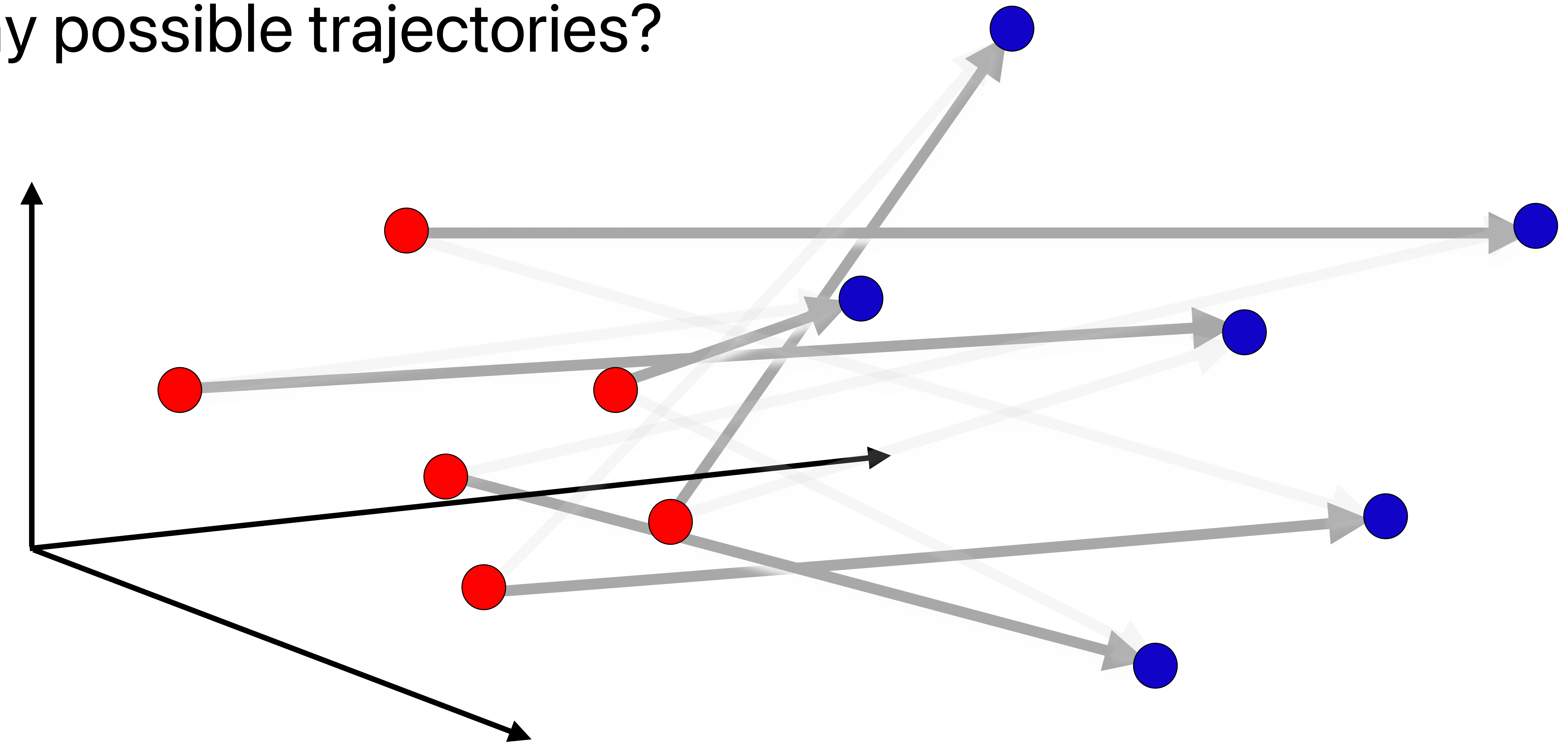


For Now, Assume Particles Moved Along Straight Lines



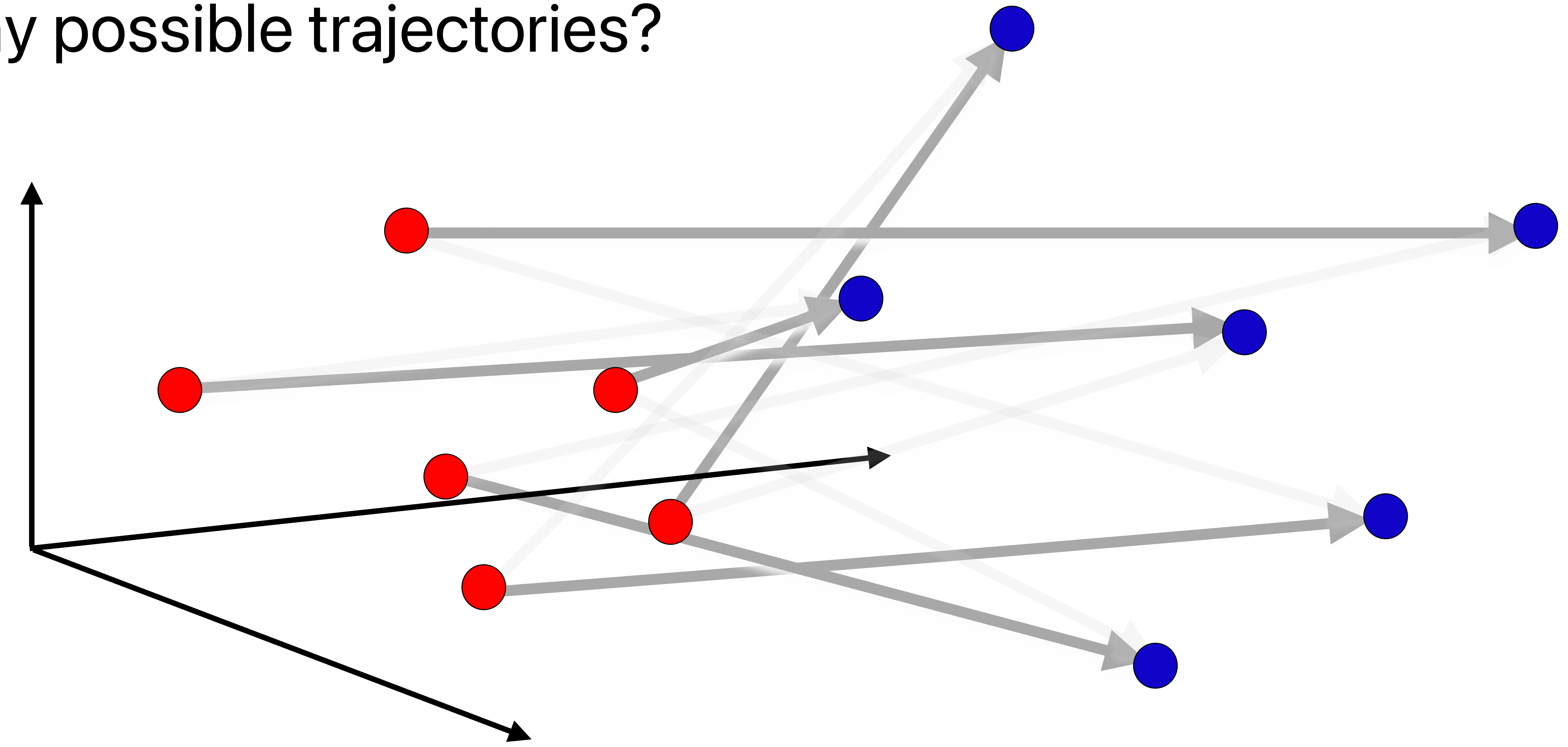
For Now, Assume Particles Moved Along Straight Lines

How many possible trajectories?



For Now, Assume Particles Moved Along Straight Lines

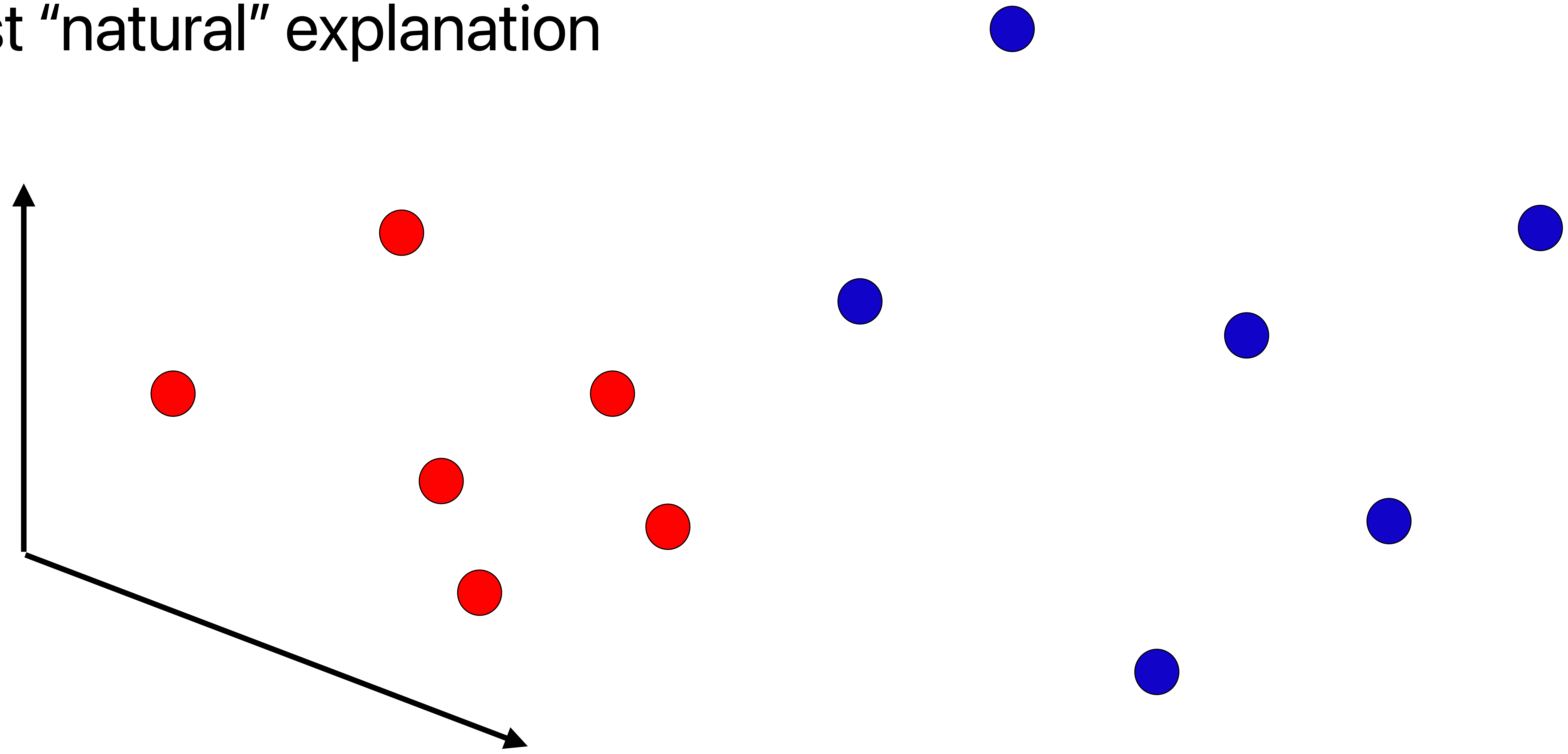
How many possible trajectories?



There are 6! = 6 x 5 x 4 x 3 x 2 = 720 permutations

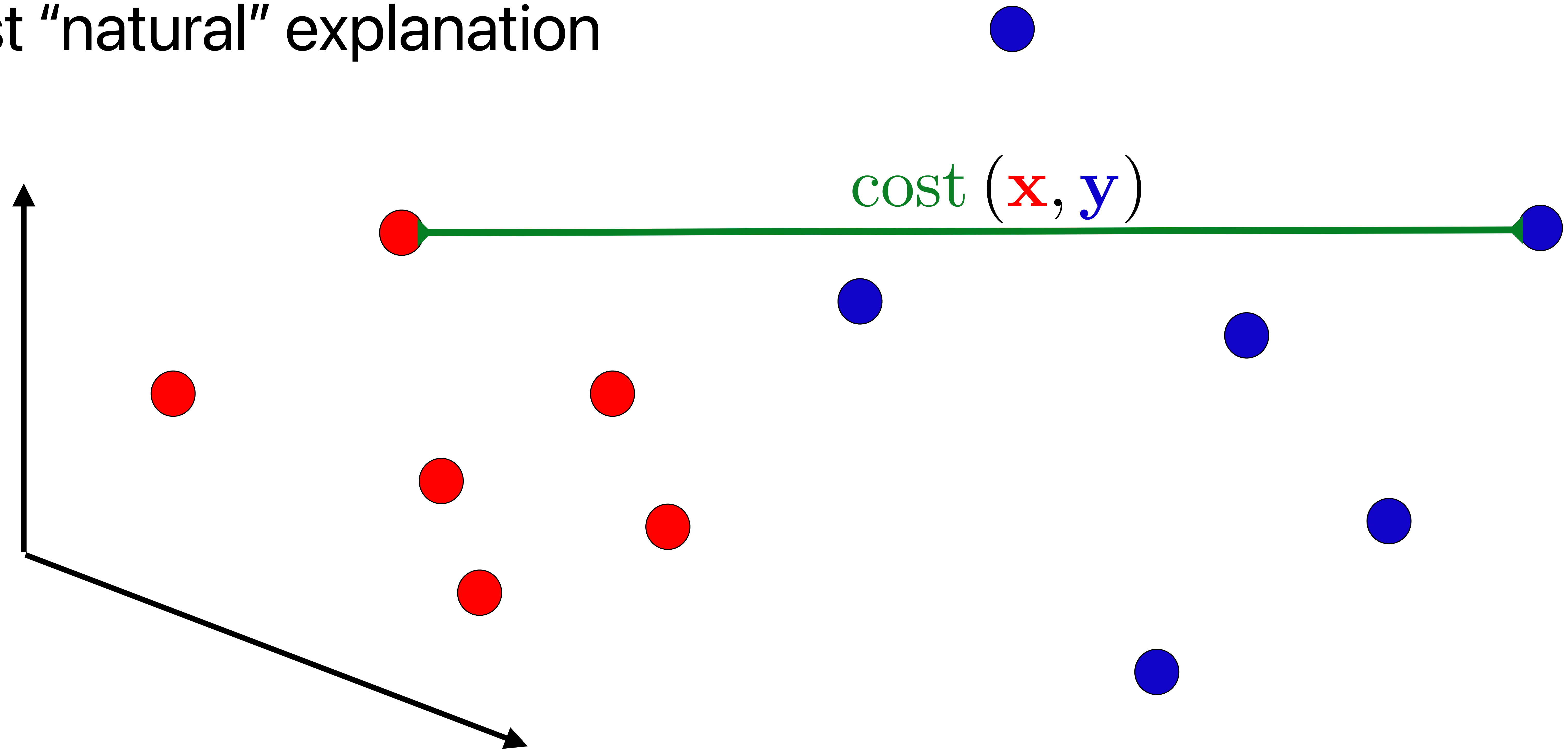
Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Goal: Find most “natural” explanation



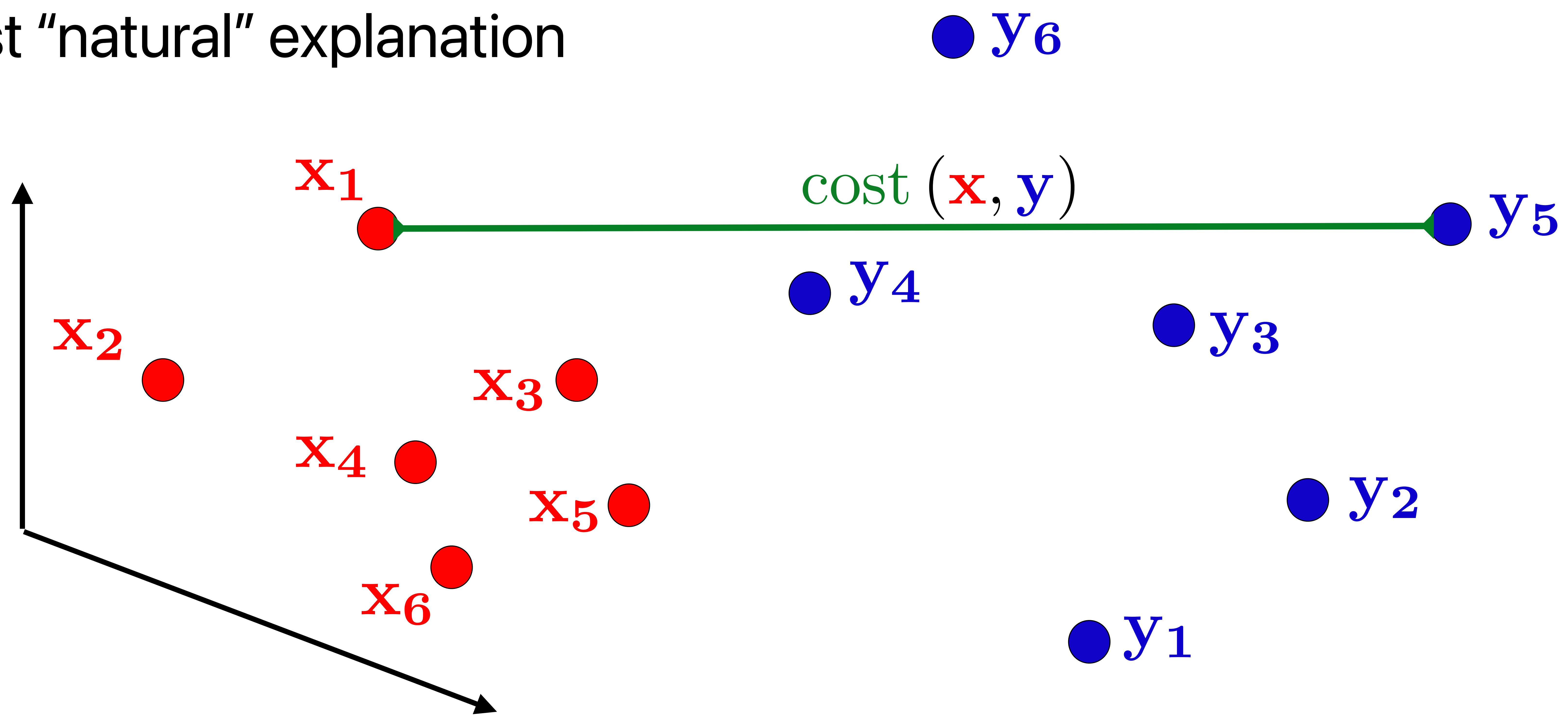
Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Goal: Find most “natural” explanation



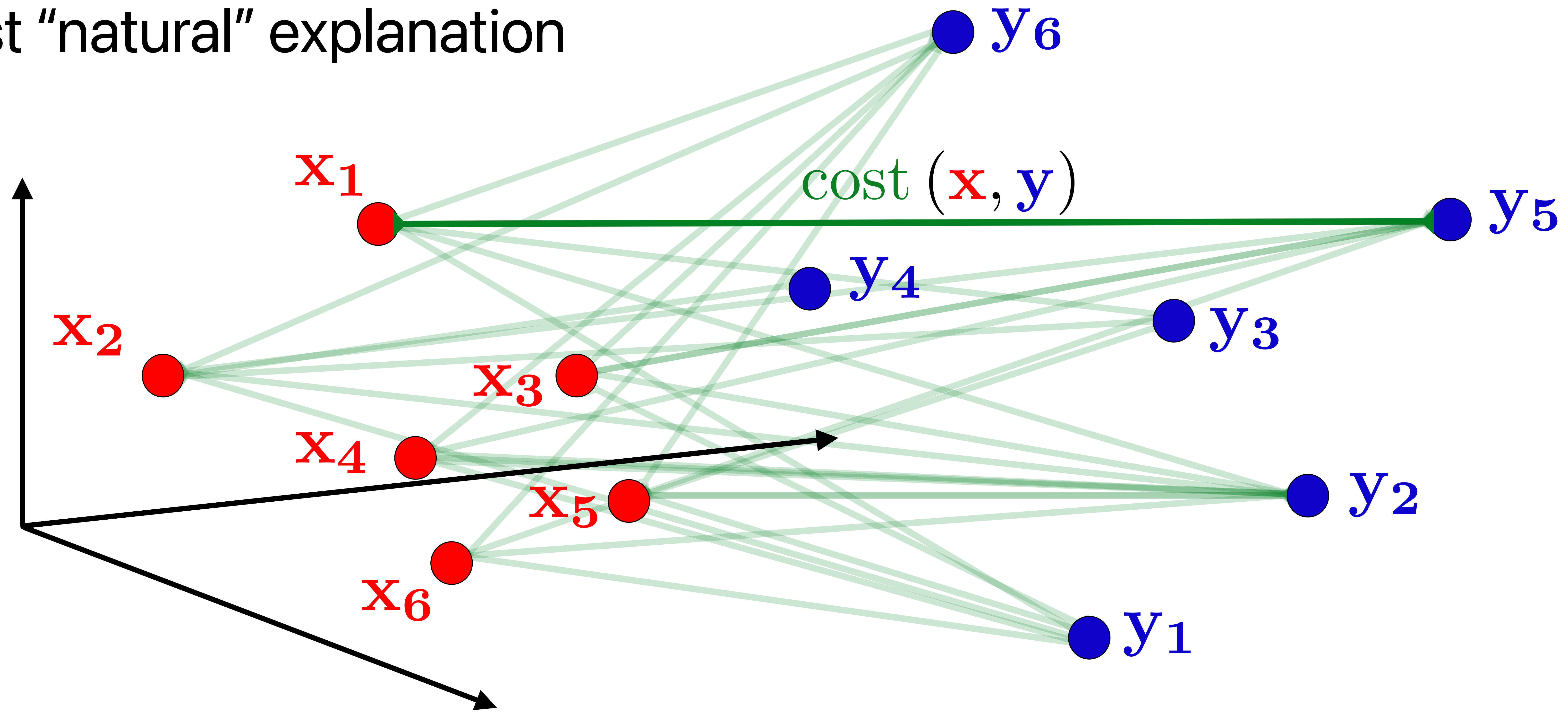
Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Goal: Find most "natural" explanation



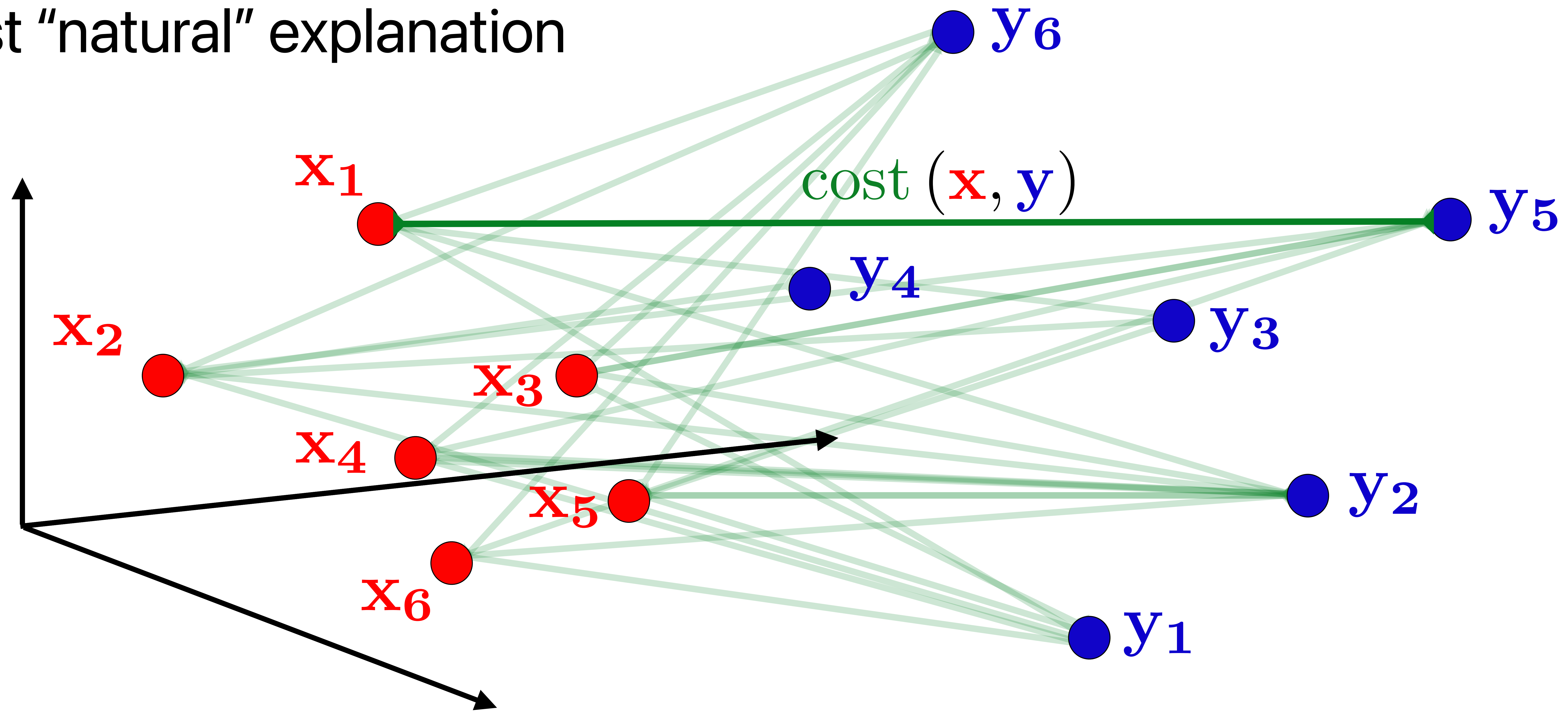
Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Goal: Find most "natural" explanation



Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Goal: Find most "natural" explanation

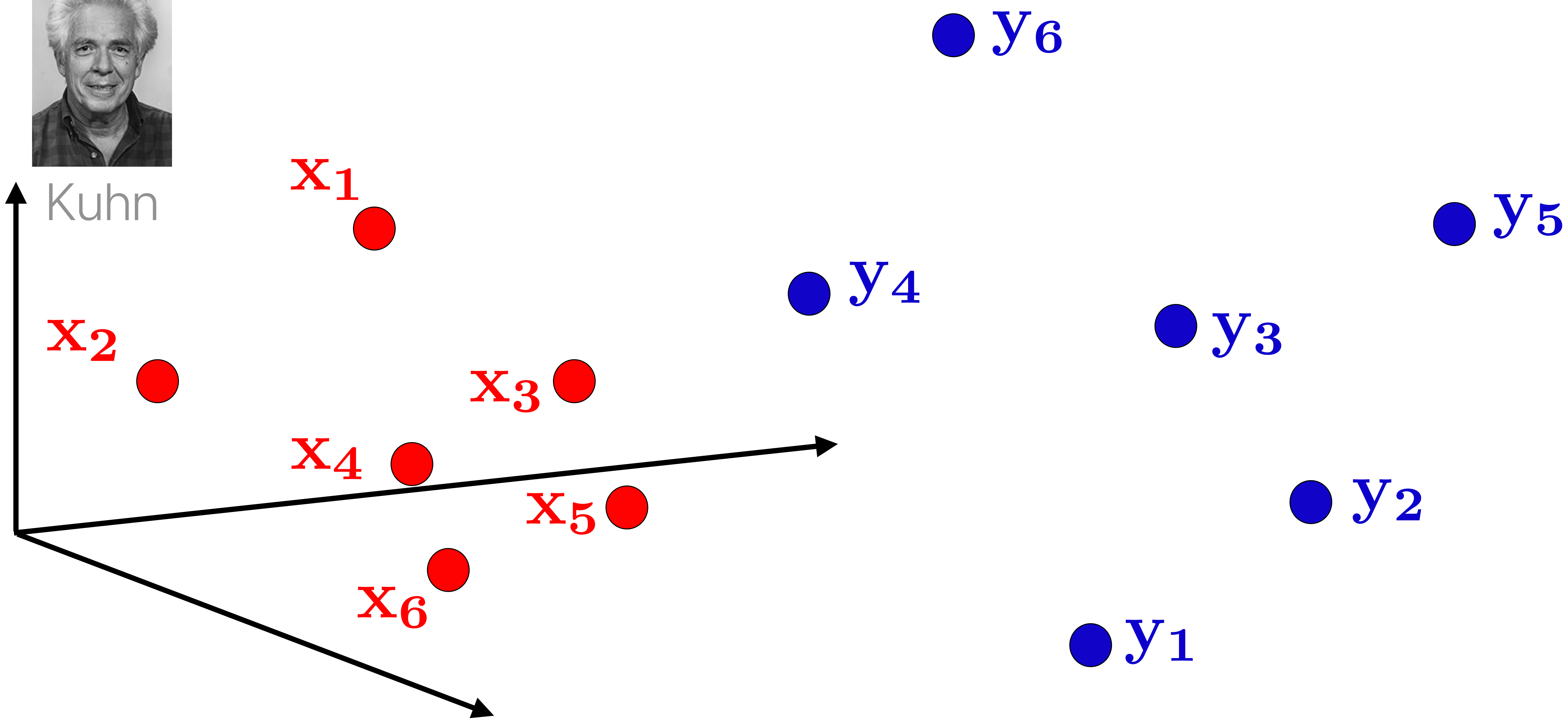


"natural" will be encoded with a cost between pairs of points, to find best permutation:

$$\min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} \boldsymbol{c}\left(\mathbf{x}_i, \mathbf{y}_{\sigma_i}\right)$$

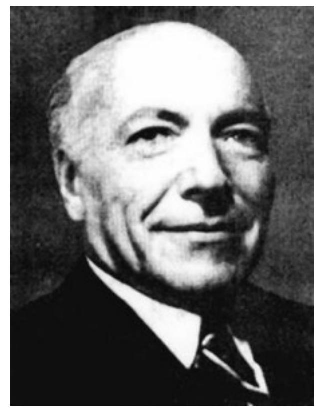
Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Egerváry König Kuhn

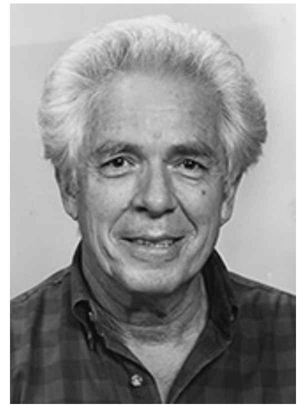


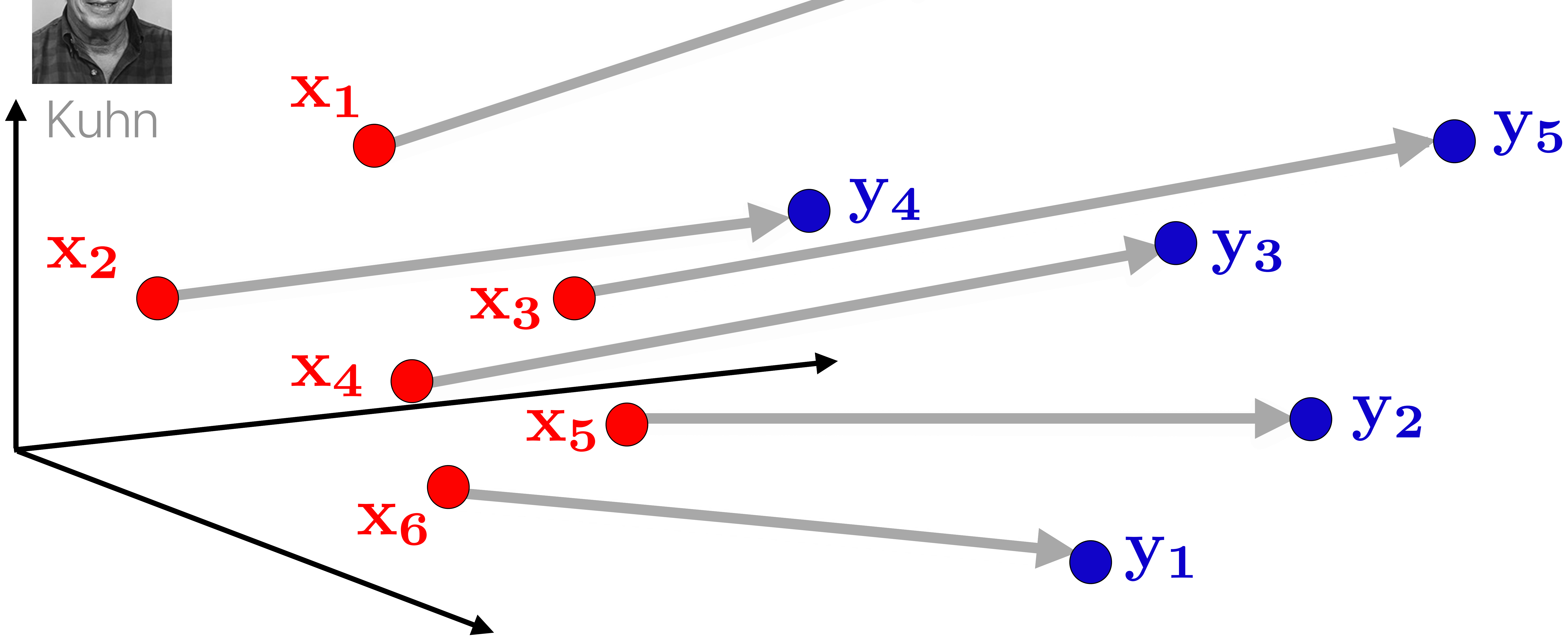
To get the best permutation $(6,4,5,3,2,1)$, run *Hungarian* algorithm (or variants)

Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem


Egerváry


König


Kuhn

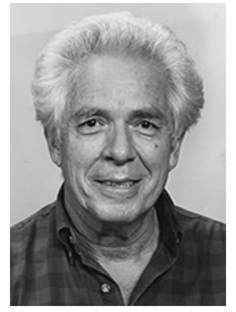


To get the best permutation $(6,4,5,3,2,1)$, run *Hungarian* algorithm (or variants)

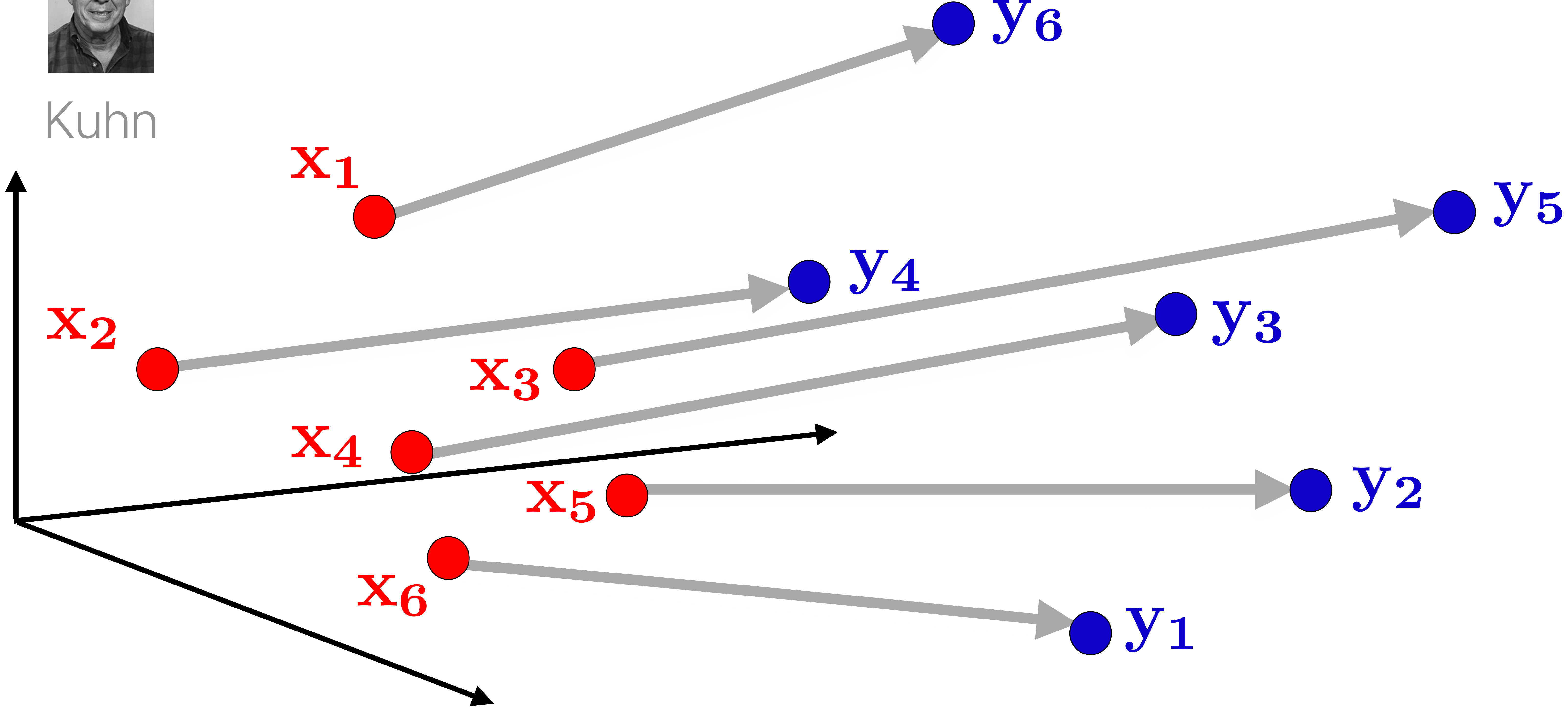
Optimal Transport v.0.0, *a.k.a.* Linear Assignment Problem

Egerváry

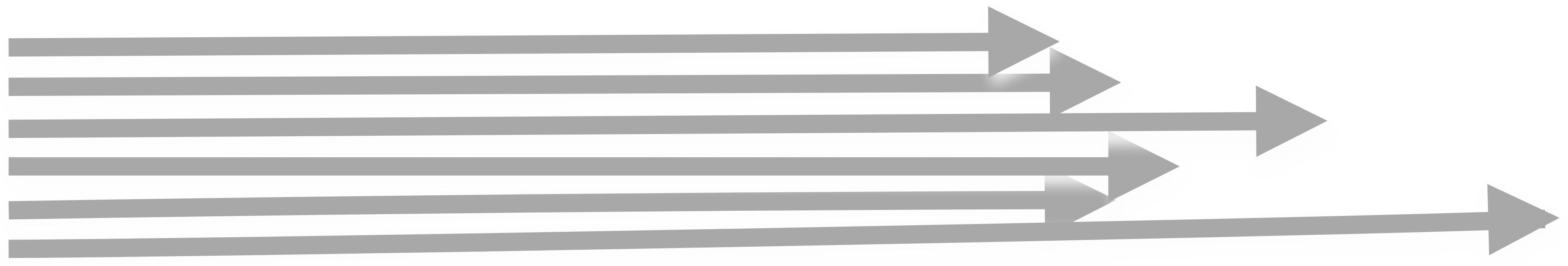
König



Kuhn



Sum of these lengths = {*Wasserstein, OT, Fréchet, Kantorovich, EMD*} distance between

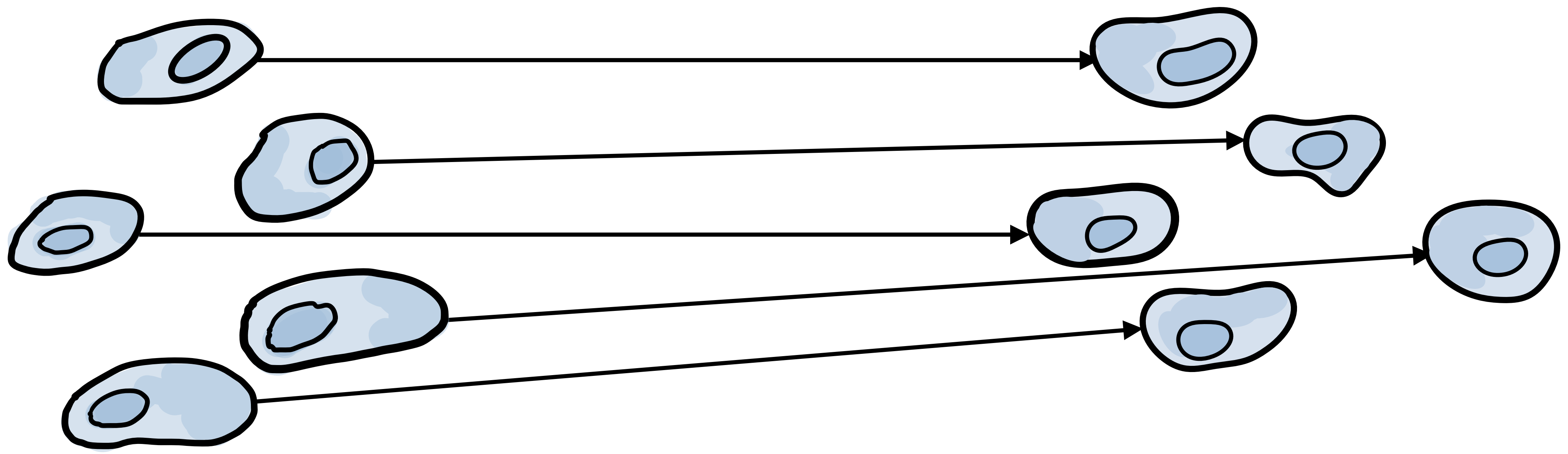


Linear Assignments Are Useful

Cell

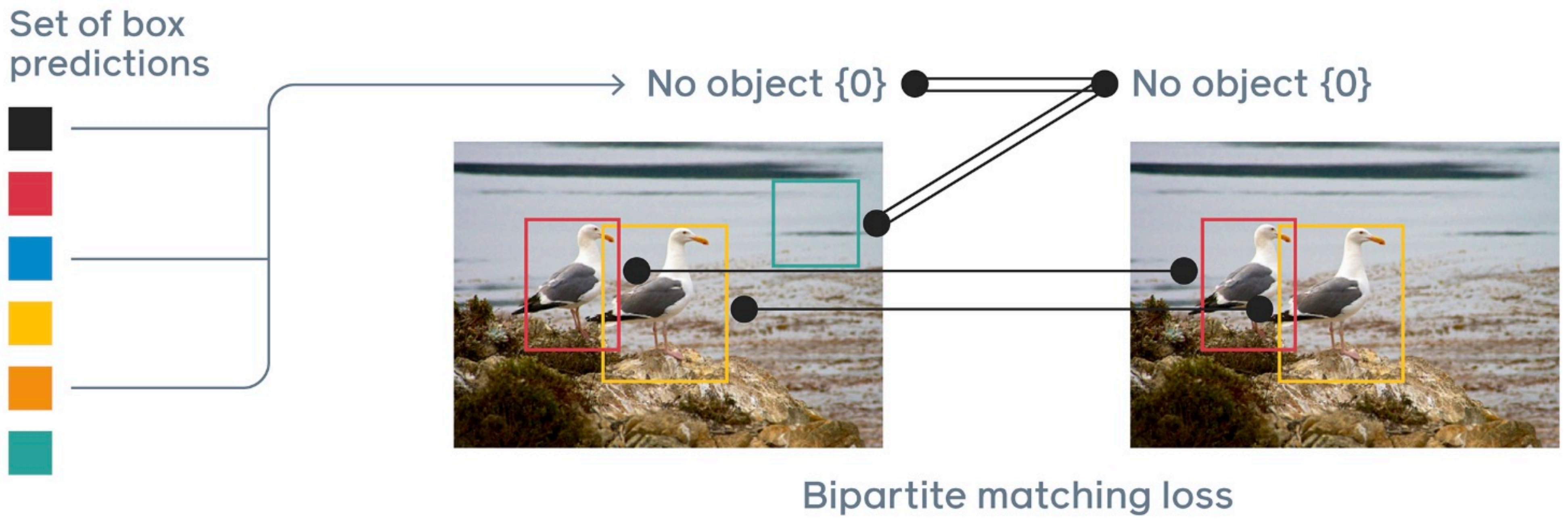
Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming

[Schiebinger+19]



End-to-end object detection with Transformers

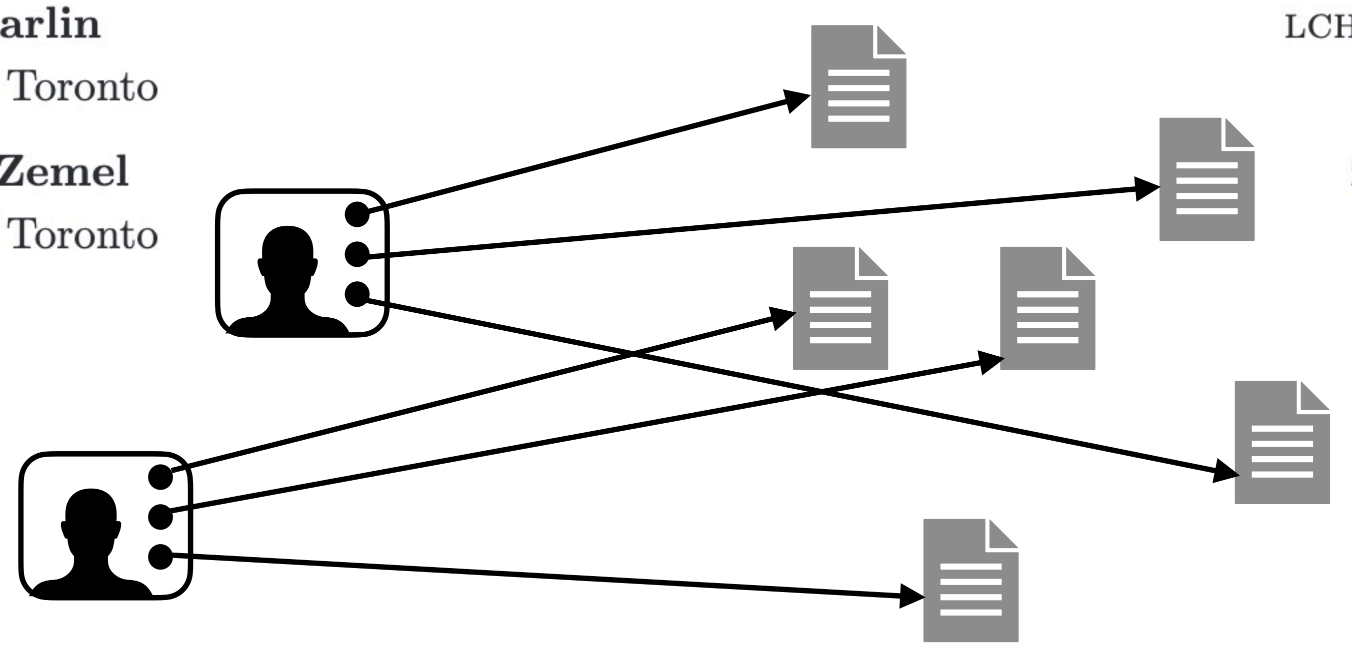
[Carion+20]



The Toronto Paper Matching System: An automated paper-reviewer assignment system

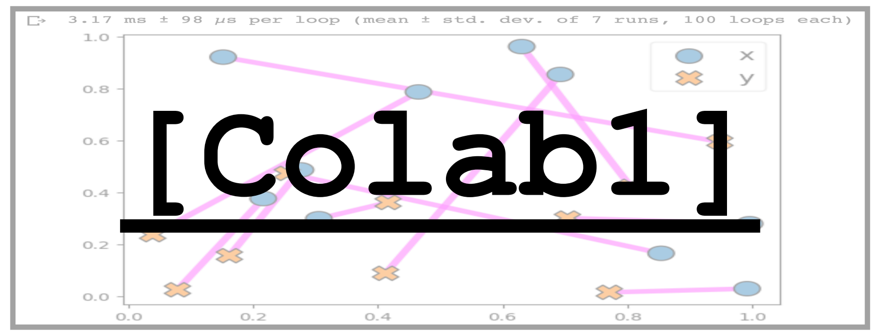
Laurent Charlin LCHARLIN@CS.TORONTO.EDU
University of Toronto

Richard S. Zemel ZEMEL@CS.TORONTO.EDU
University of Toronto

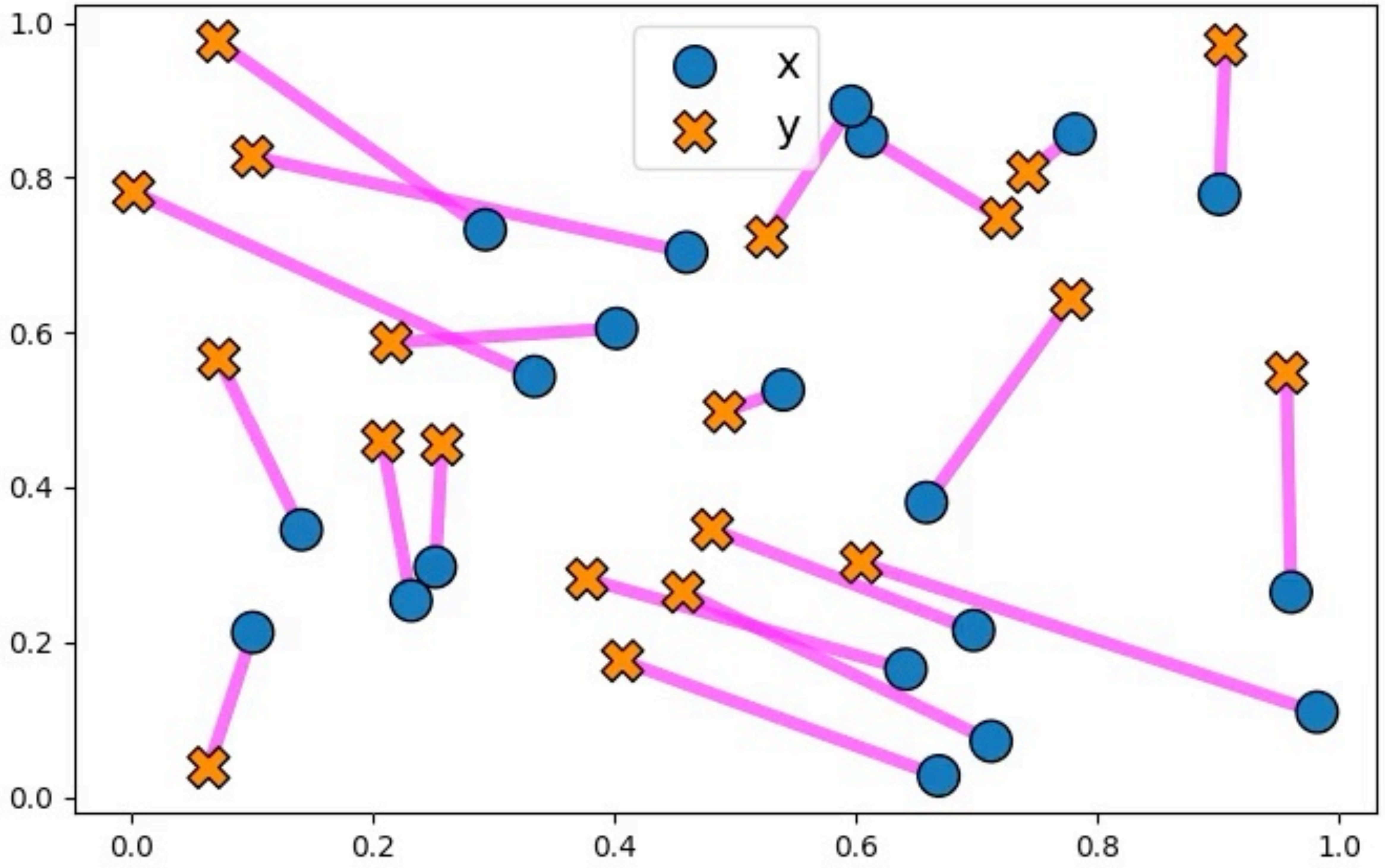


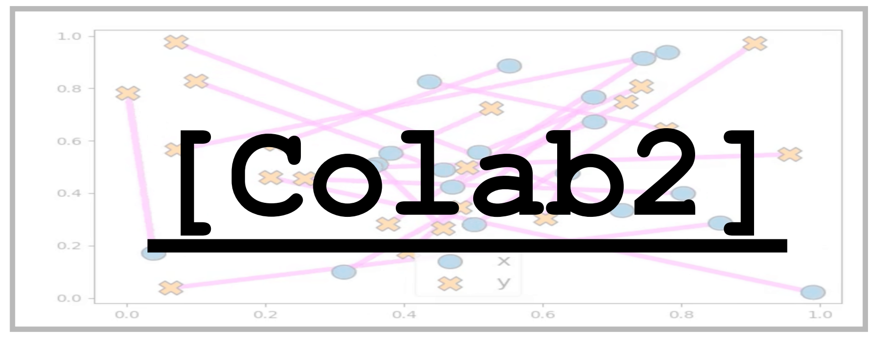
Unforuntately, optimal matchings (OM) are not convenient tools for ML:

Computing optimal matchings runs into multiple issues:



- Bad scaling, typically $O(n^3)$ complexity for n points
- Optimal matchings are unstable, and hard to parallelize efficiently on GPU

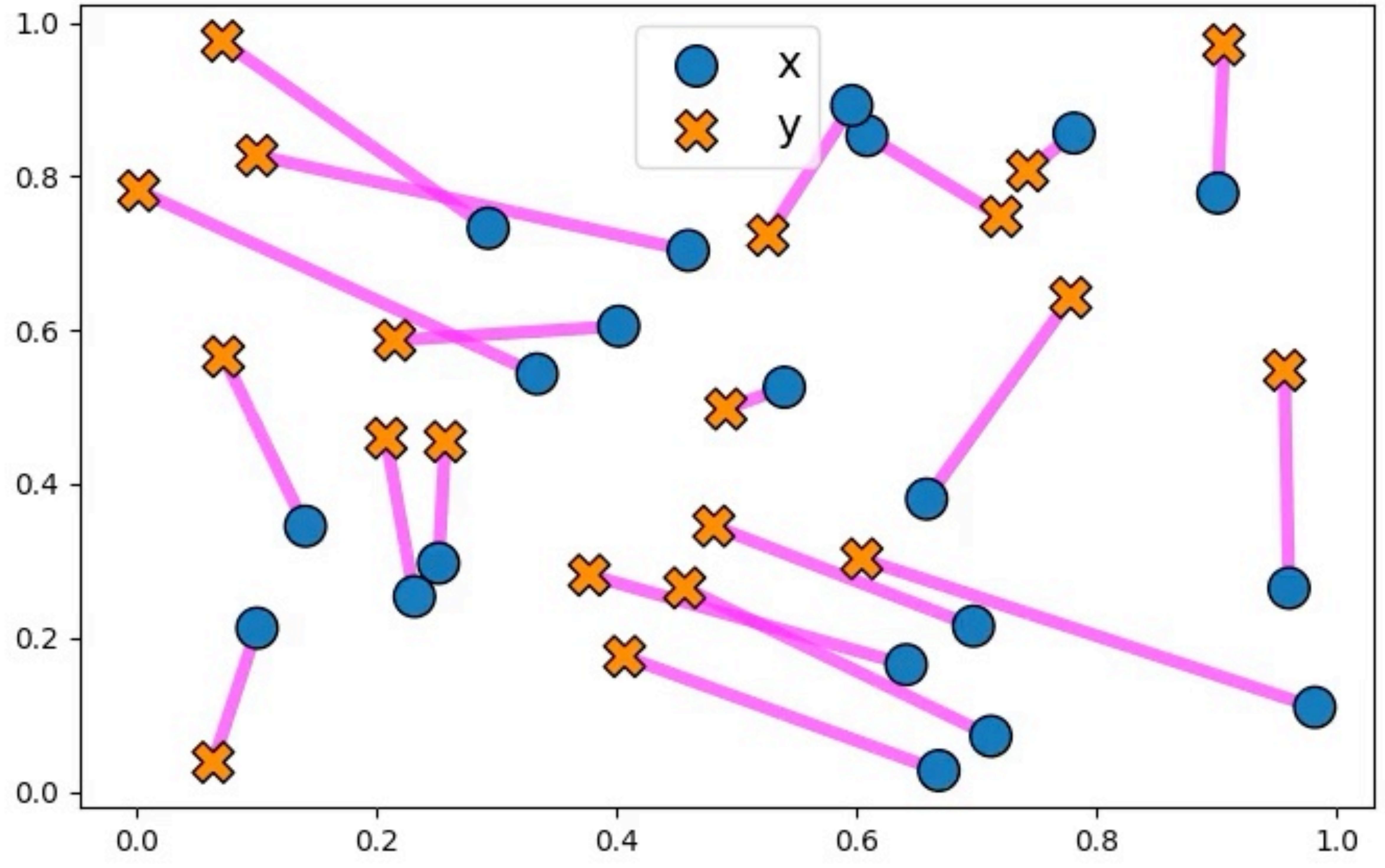


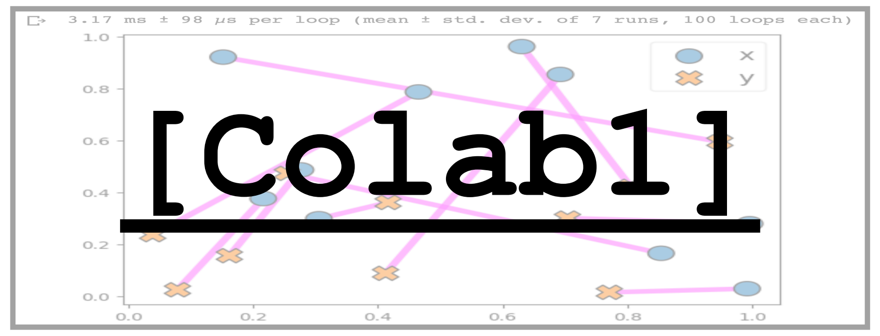


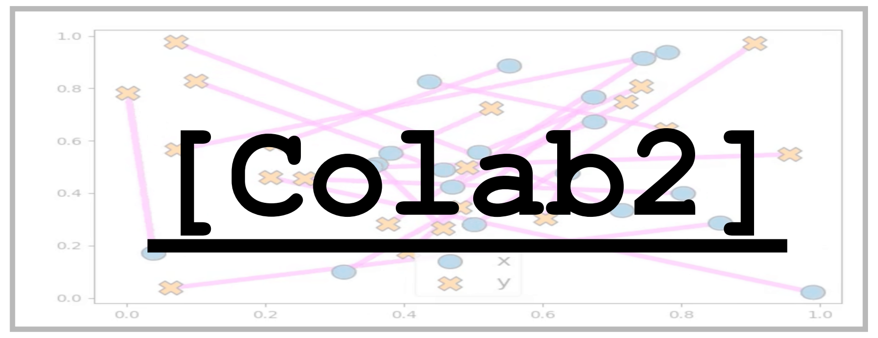
Unforuntately, optimal matchings (OM) are not convenient tools for ML:

Computing optimal matchings runs into multiple issues:

- Bad scaling, typically $O(n^3)$ complexity for n points
- Optimal matchings are unstable, and hard to parallelize efficiently on GPU



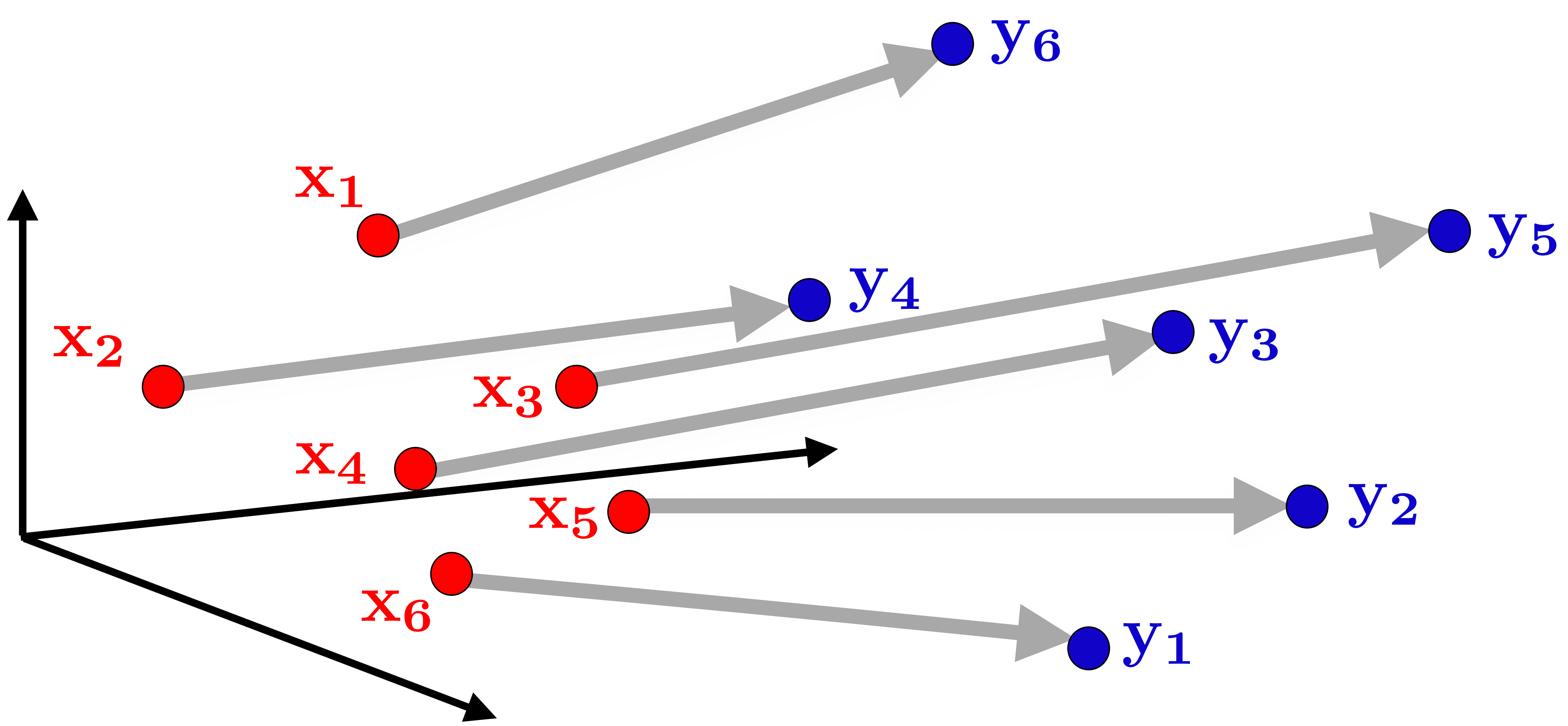




Optimal Matchings Raise More Questions than Answers

Even more problematically, OM are brittle, rigid, and lack generalization:

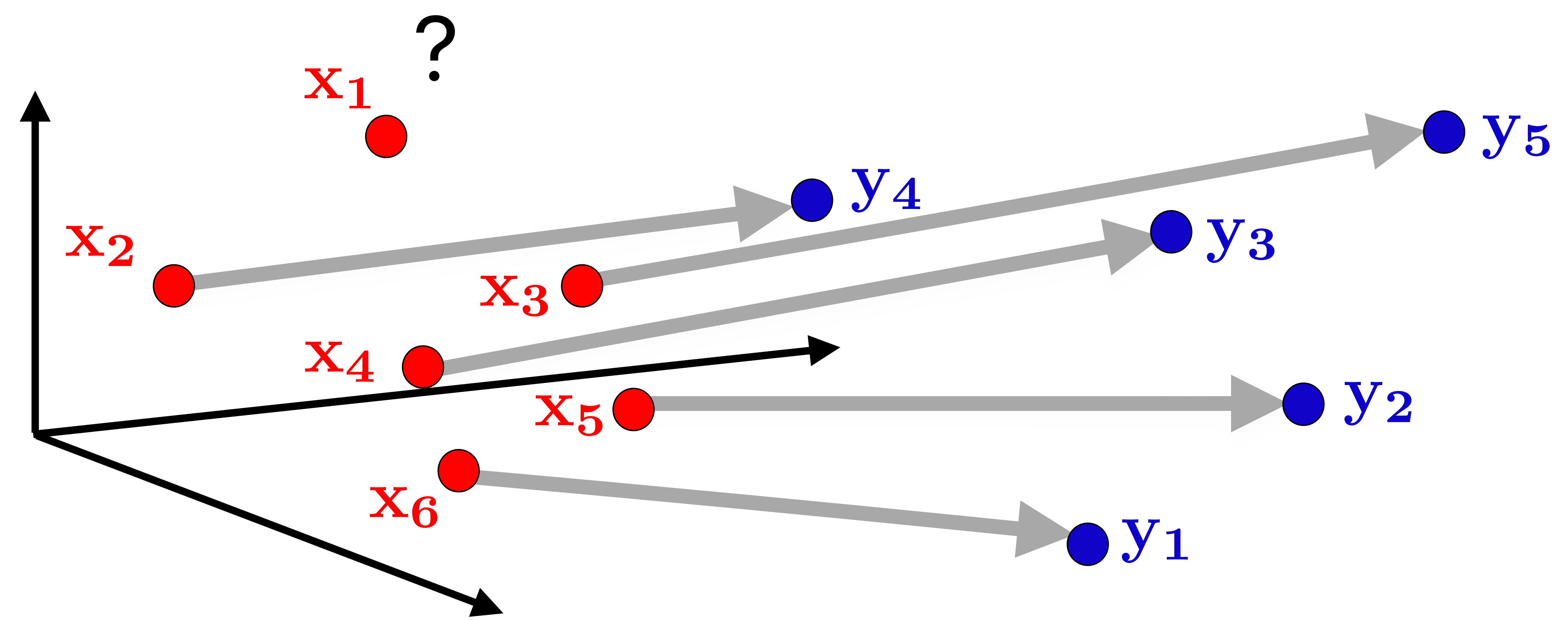
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new out-of-sample points?
- What if we want stochasticity? From straight-lines to random stochastic paths?



Optimal Matchings Raise More Questions than Answers

Even more problematically, OM are brittle, rigid, and lack generalization:

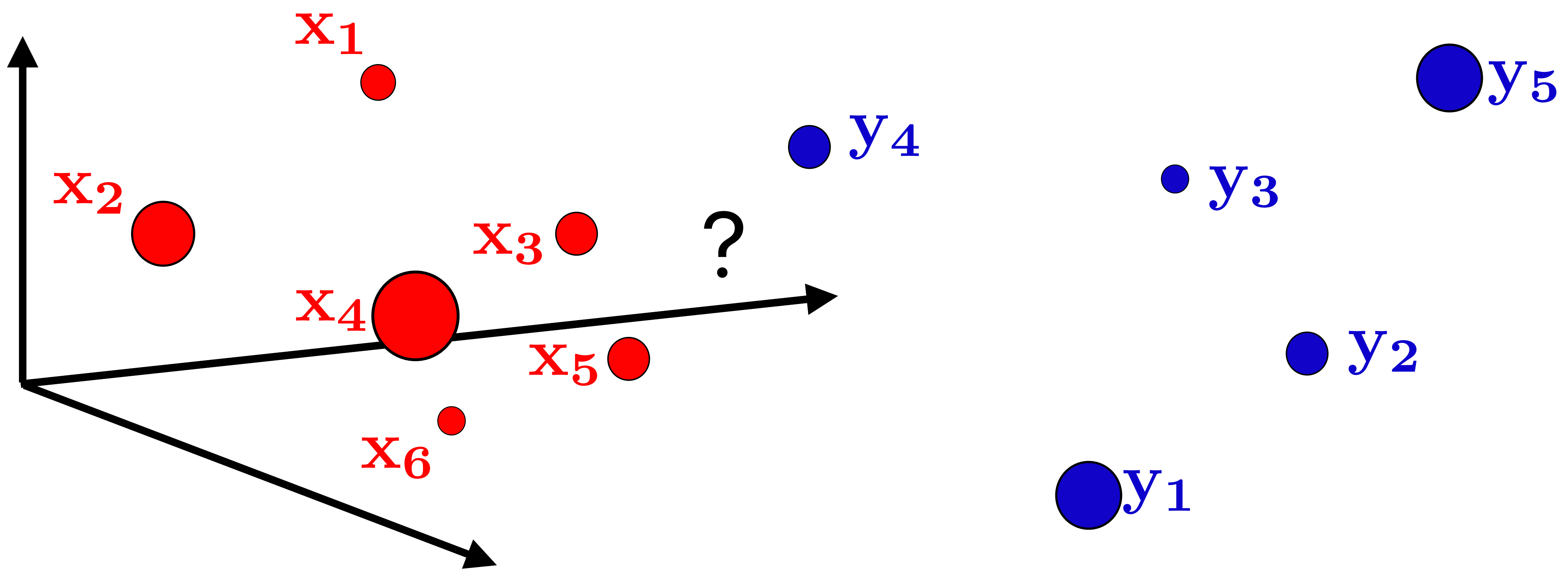
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new out-of-sample points?
- What if we want stochasticity? From straight-lines to random stochastic paths?



Optimal Matchings Raise More Questions than Answers

Even more problematically, OM are brittle, rigid, and lack generalization:

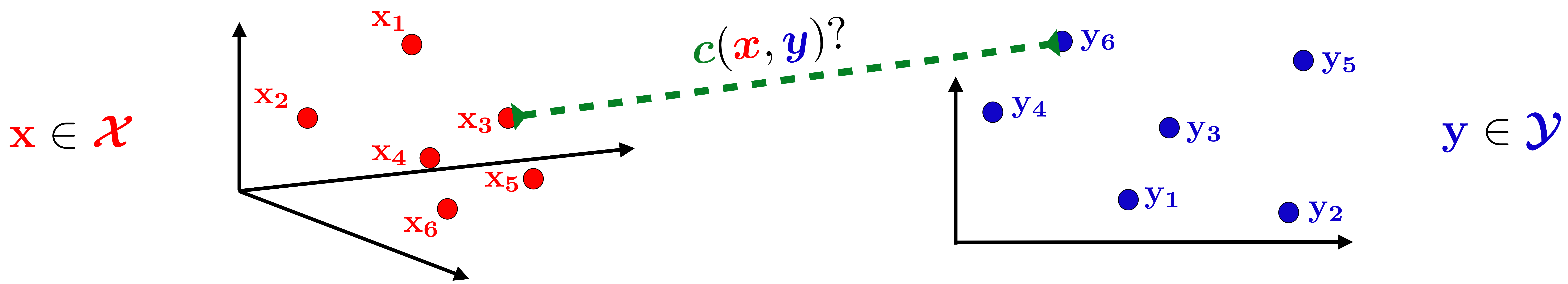
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new out-of-sample points?
- What if we want stochasticity? From straight-lines to random stochastic paths?



Optimal Matchings Raise More Questions than Answers

Even more problematically, they are brittle, rigid, and lack generalization:

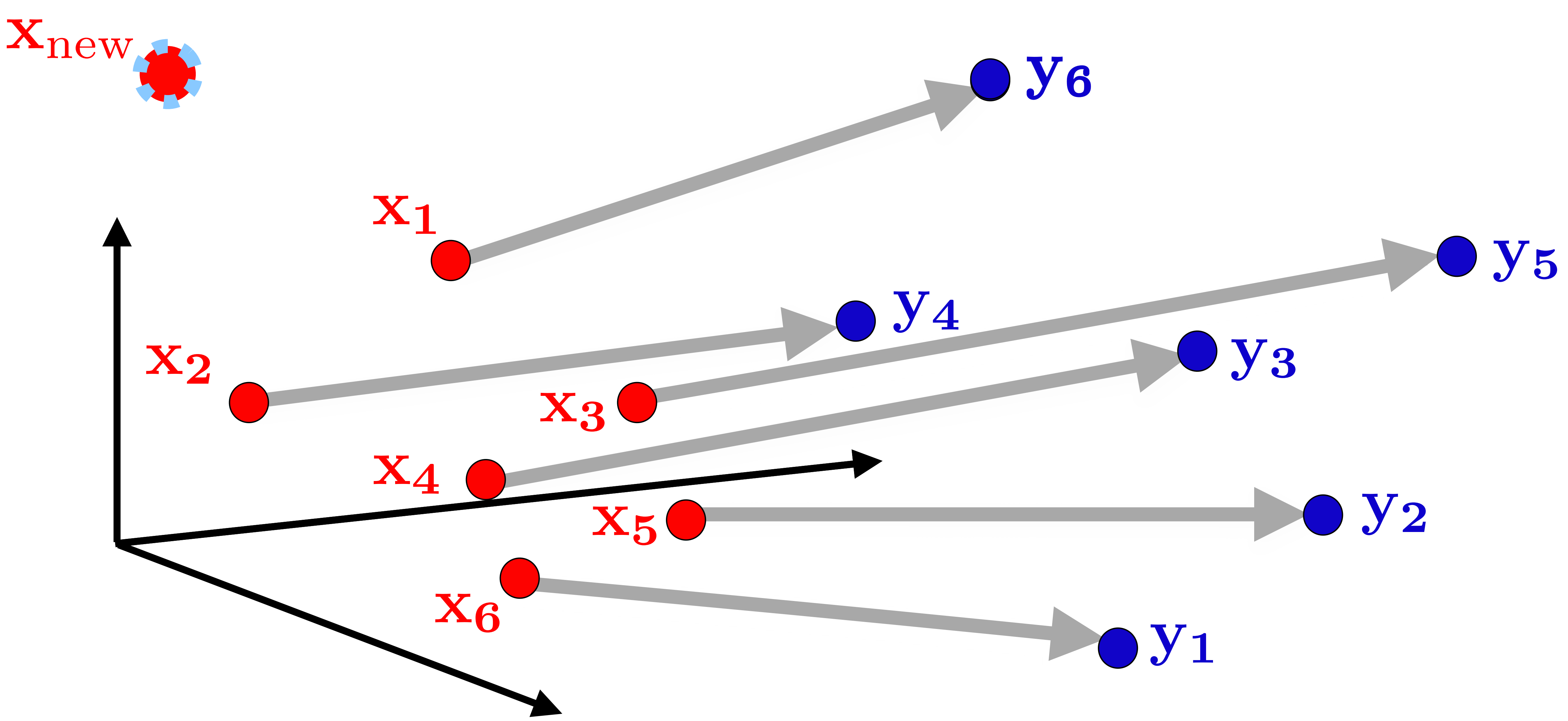
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is **no obvious** cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new out-of-sample points?
- What if we want stochasticity? From straight-lines to random stochastic paths?



Optimal Matchings Raise More Questions than Answers

Even more problematically, they are brittle, rigid, and lack generalization:

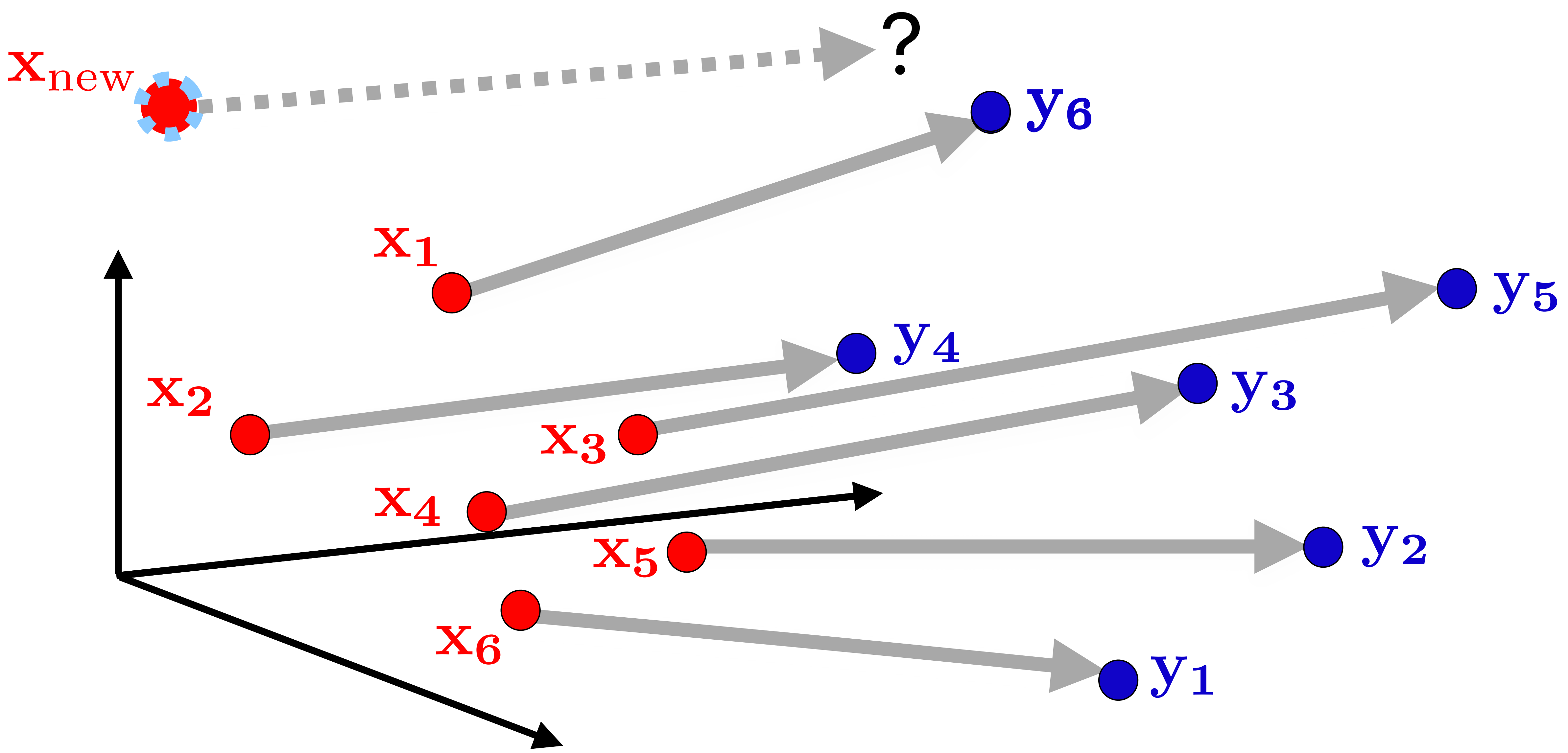
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new **out-of-sample** points?
- What if we want stochasticity? From straight-lines to random stochastic paths?



Optimal Matchings Raise More Questions than Answers

Even more problematically, they are brittle, rigid, and lack generalization:

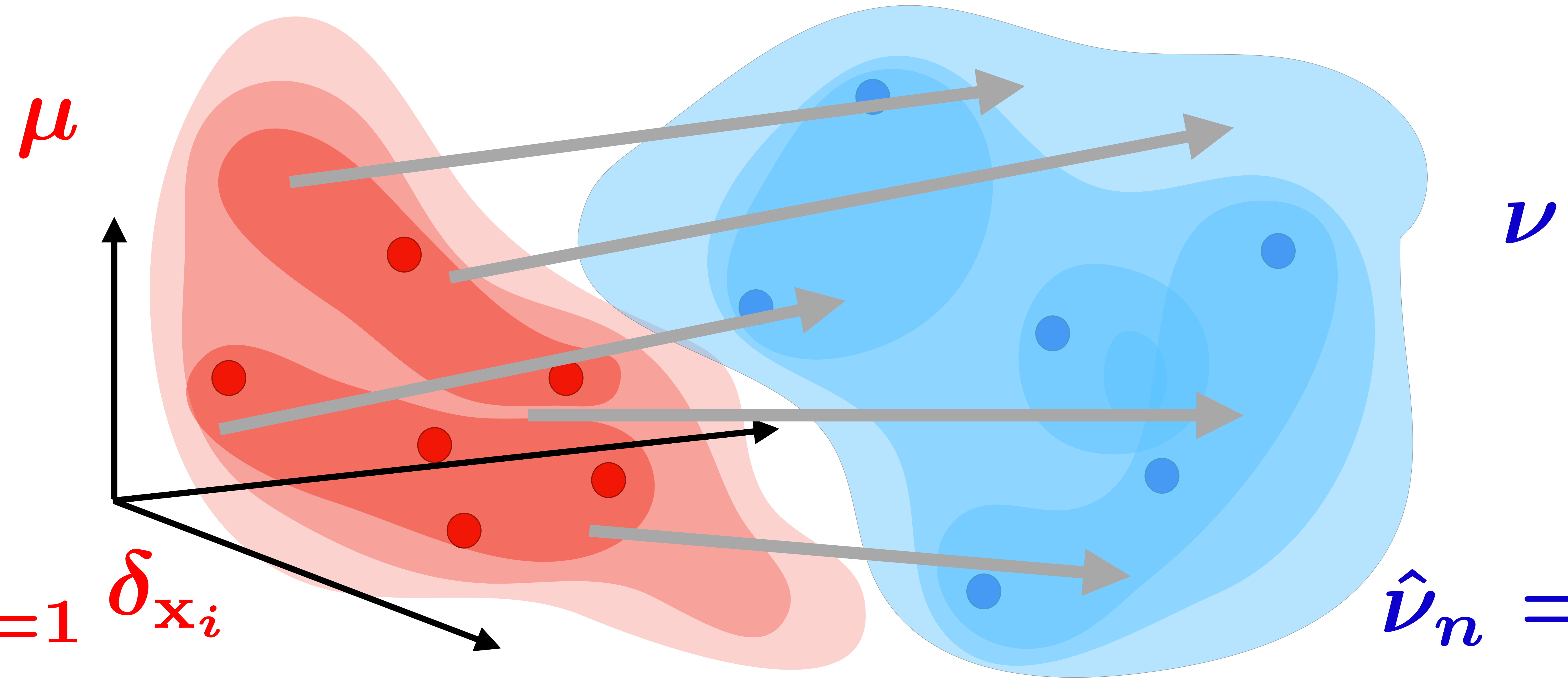
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new **out-of-sample** points?
- What if we want stochasticity? From straight-lines to random stochastic paths?



Optimal Matchings Raise More Questions than Answers

Even more problematically, they are brittle, rigid, and lack generalization:

- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new **out-of-sample** points?
- What if we want stochasticity? From straight-lines to random stochastic paths?

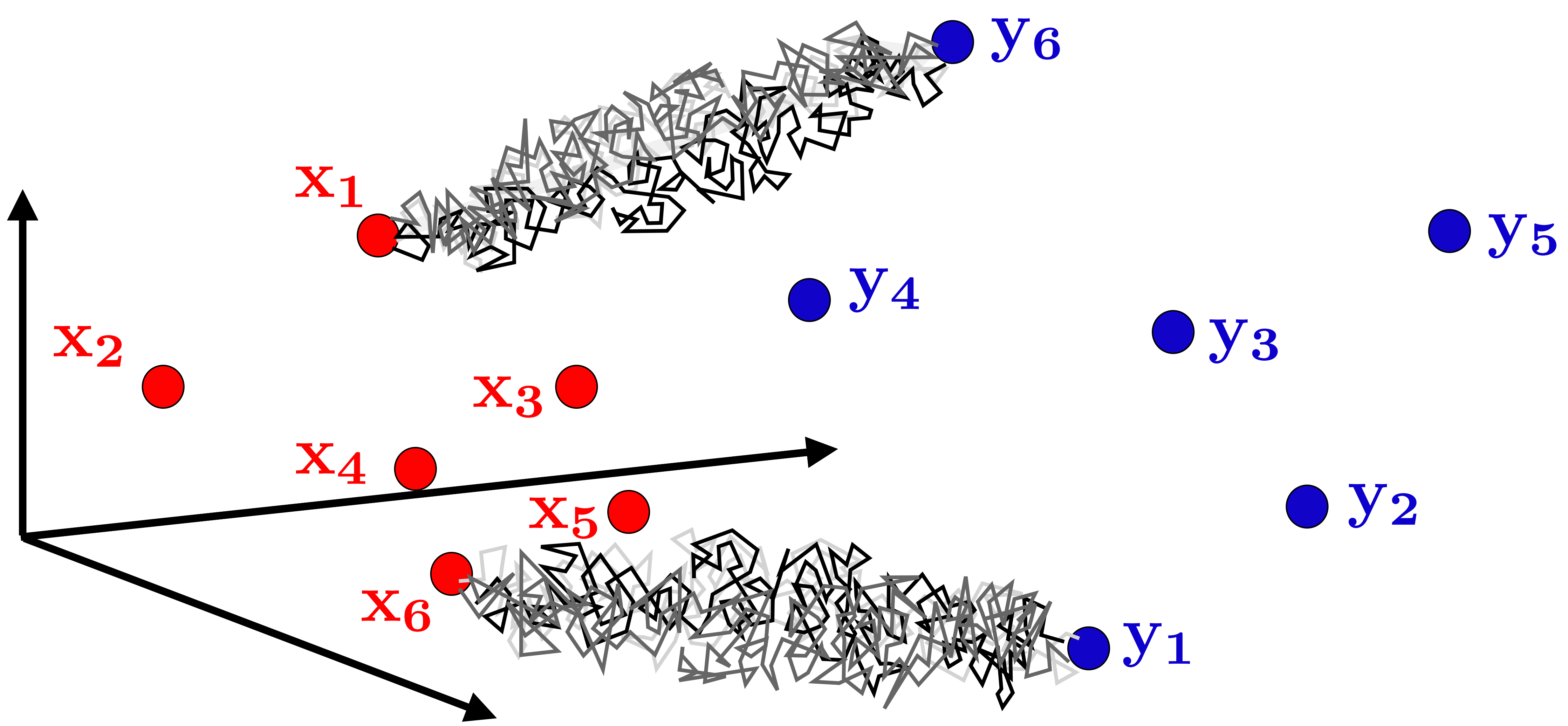


$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n}\delta_{\mathbf{x}_i}$ $\hat{\nu}_n = \frac{1}{n}\sum_{j=1}^{n}\delta_{\mathbf{y}_j}$

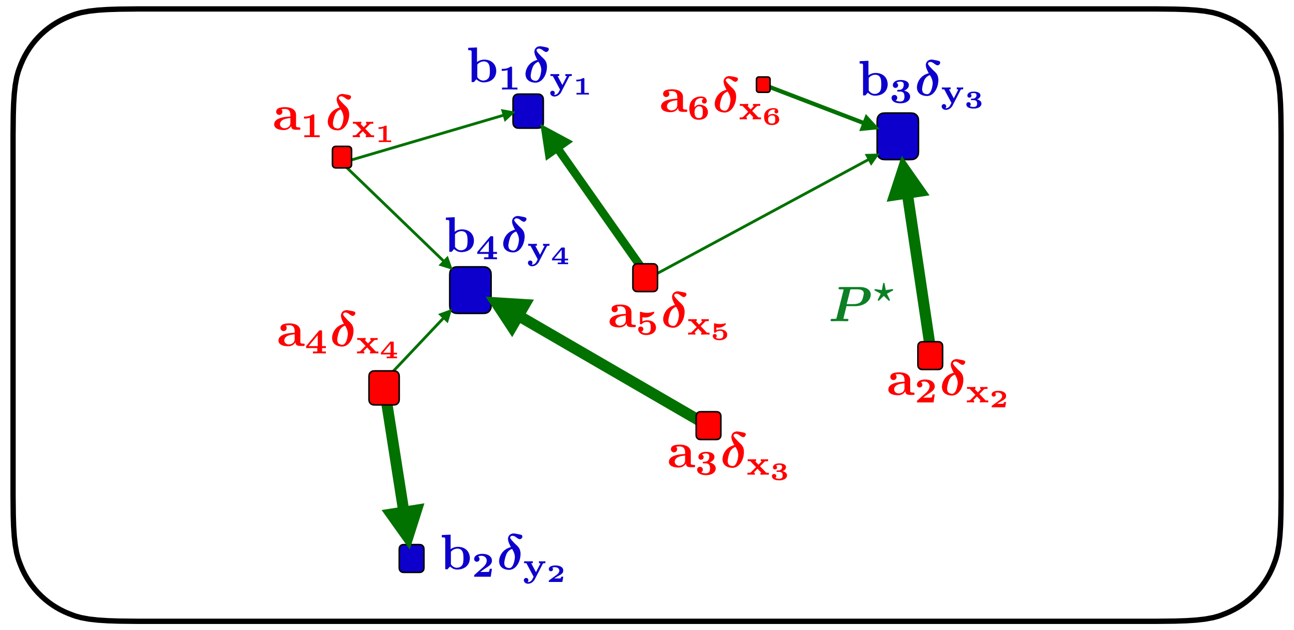
Optimal Matchings Raise More Questions than Answers

Even more problematically, they are brittle, rigid, and lack generalization:

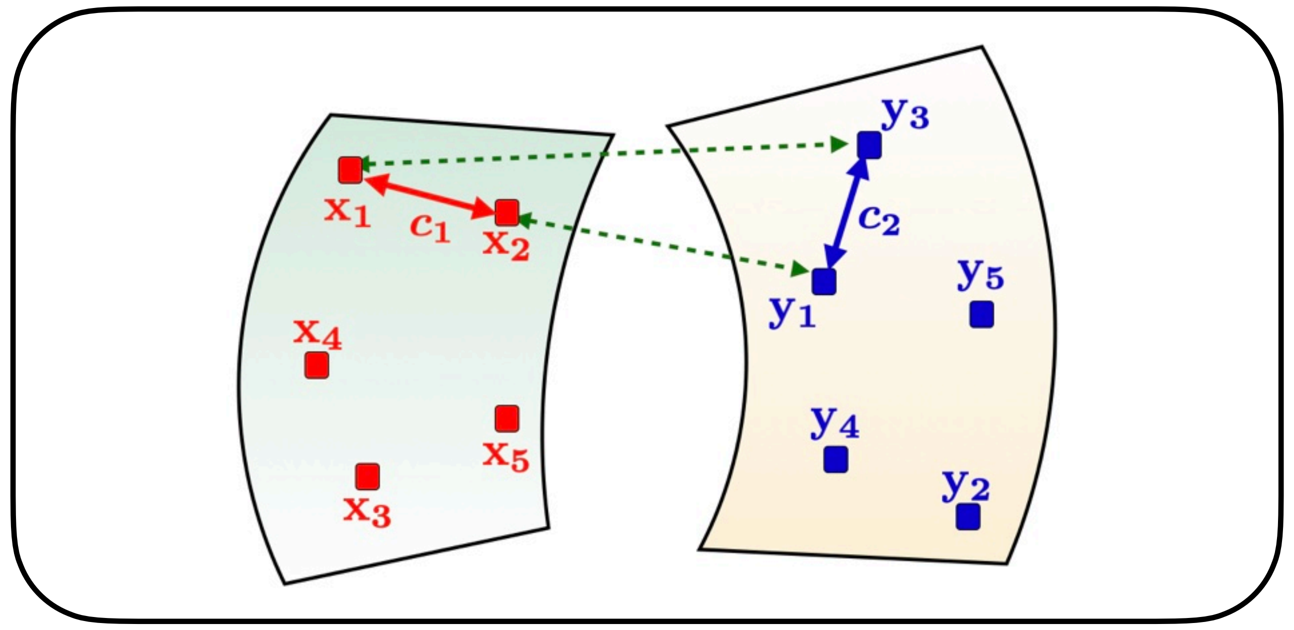
- What if the numbers of **x** and **y** points differ, $n \neq m$? weighted points?
- What if there is no obvious cost function? e.g. **x** and **y** points are in ≠ spaces?
- What if we need ways to generalize, and move new **out-of-sample** points?
- What if we need stochasticity? From straight-lines to stochastic processes?



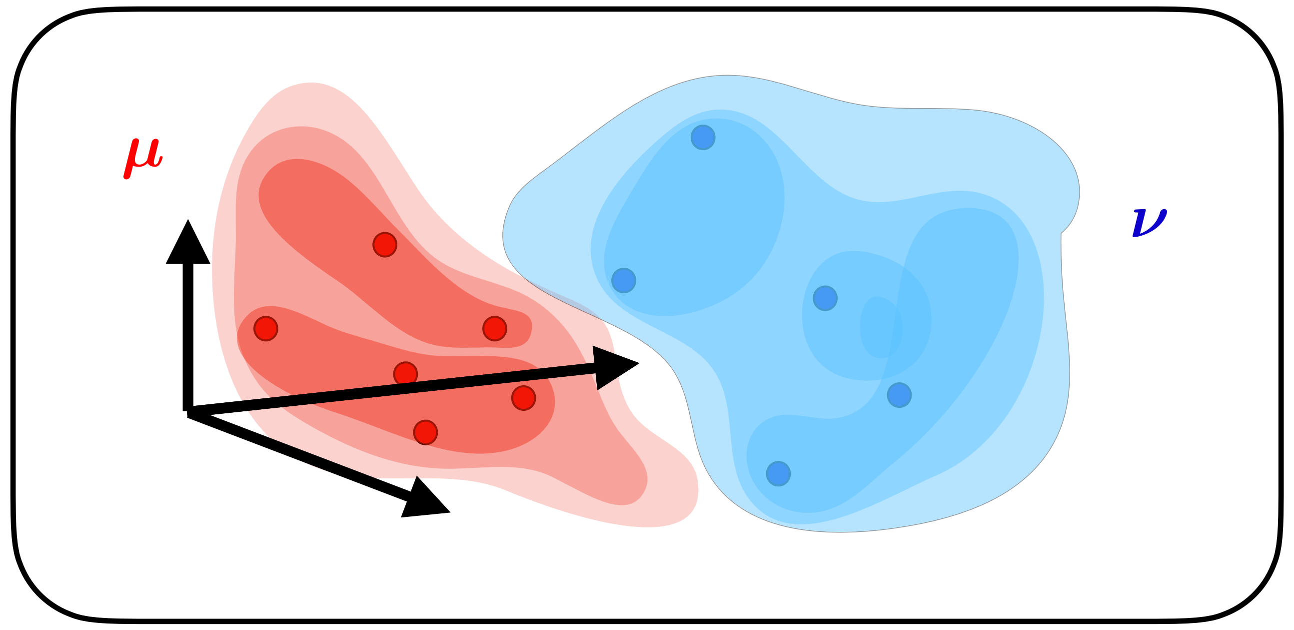
Landscape / Jargon of Optimal Transport Problems



- LAP, Kantorovich (Rubinstein), duality and potential.
- Optimal coupling, transport plans, EMD, Wasserstein



- QAP, Gromov-Wasserstein (GW) distance
- Fused GW, graph isomorphism problems



- Monge Problem, Brenier & Gangbo-McCann theorems
- Monge-Ampère PDE, Benamou-Brenier, JKO
- Schrödinger bridge, stochastic control

Outline of the Tutorial

Prelude Warm-Up: Starting with Optimal Matchings

Part 1 **Kantorovich Formulation of OT and Computations**

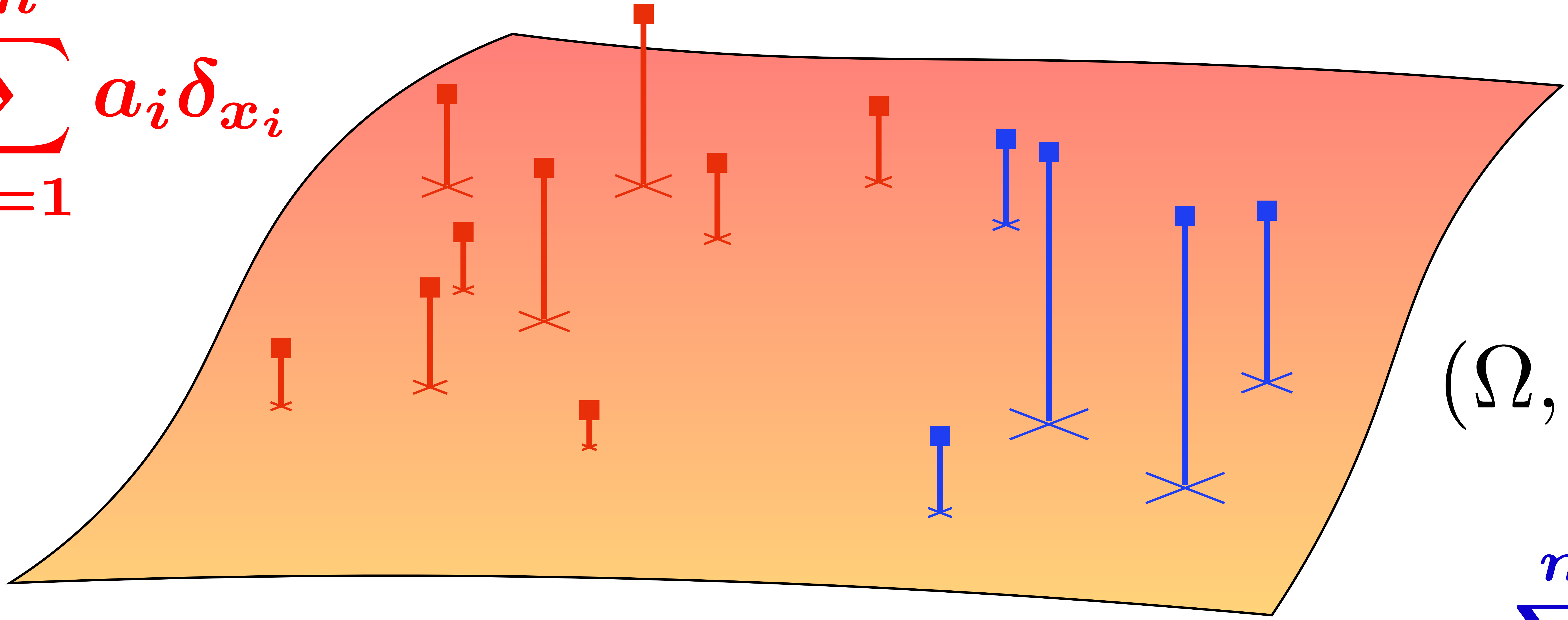
- Optimizing over space of coupling matrices, linear programs
- Regularized OT, entropic and rank constrained, unbalanced, quadratic
- Differentiability of OT plans

Part 2 Duality, Monge Formulations and Brenier Theorems

Part 3 Modeling Measure Dynamics with Optimal Transport

Optimal Transport between Two Empirical Measures

$$\mu = \sum_{i=1}^{n} a_i \delta_{x_i}$$



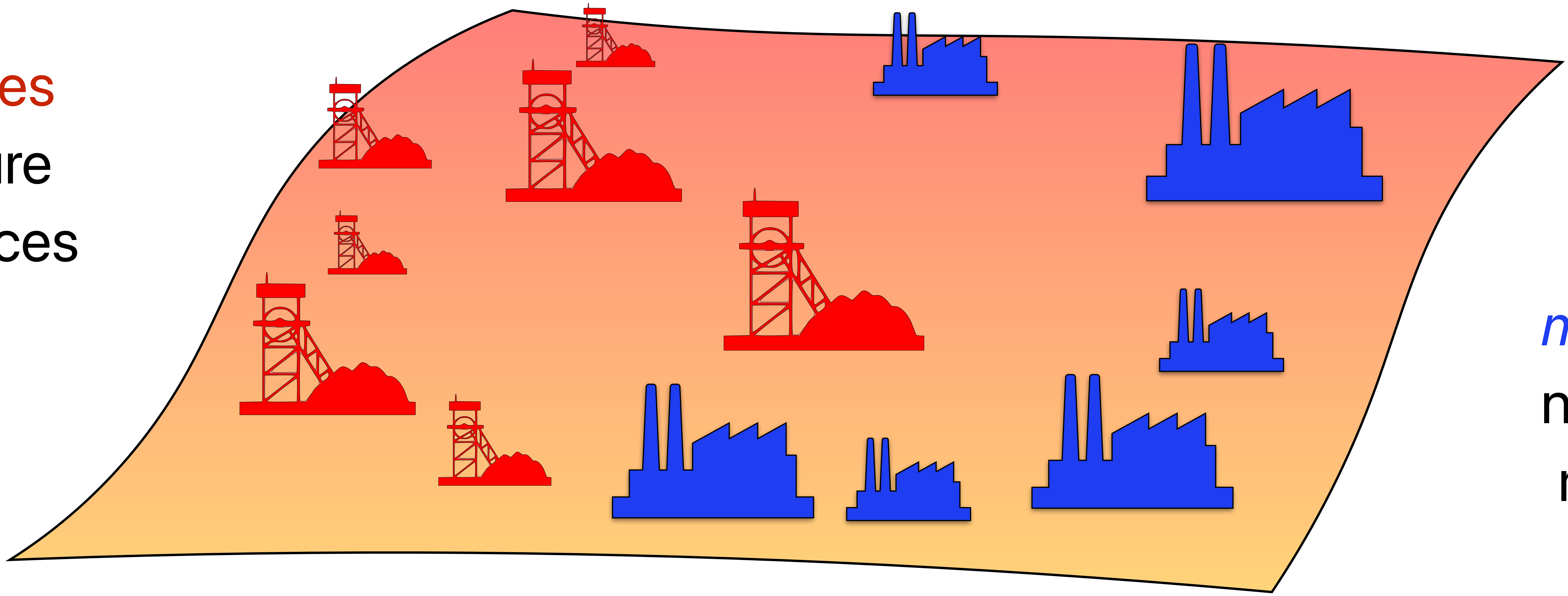
$$(\Omega, c)$$

$$\nu = \sum_{j=1}^{m} b_j \delta_{y_j}$$

Kantorovich Formulation of Optimal Transport

Hitchcock Kantorovich Dantzig Koopmans

n mines
capture
resources

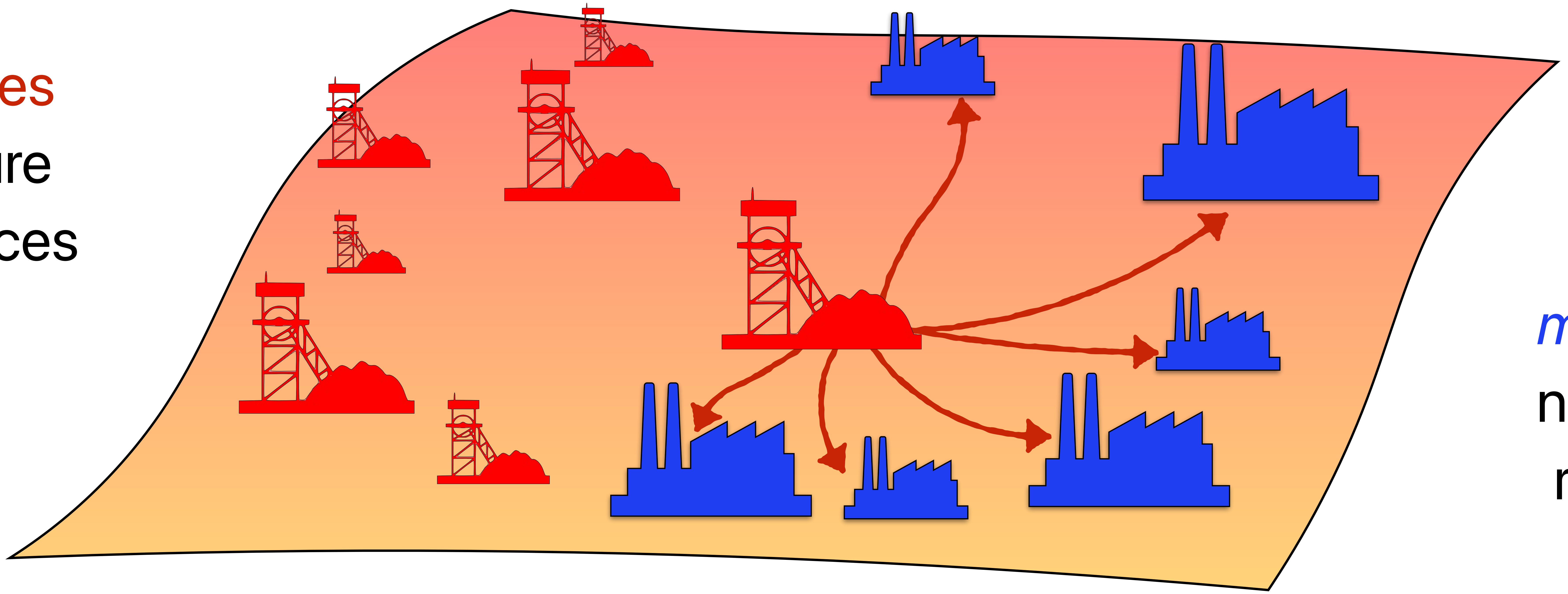


m factories
need those
resources

Kantorovich Formulation of Optimal Transport

Each of the *n* mines must dispatch its production

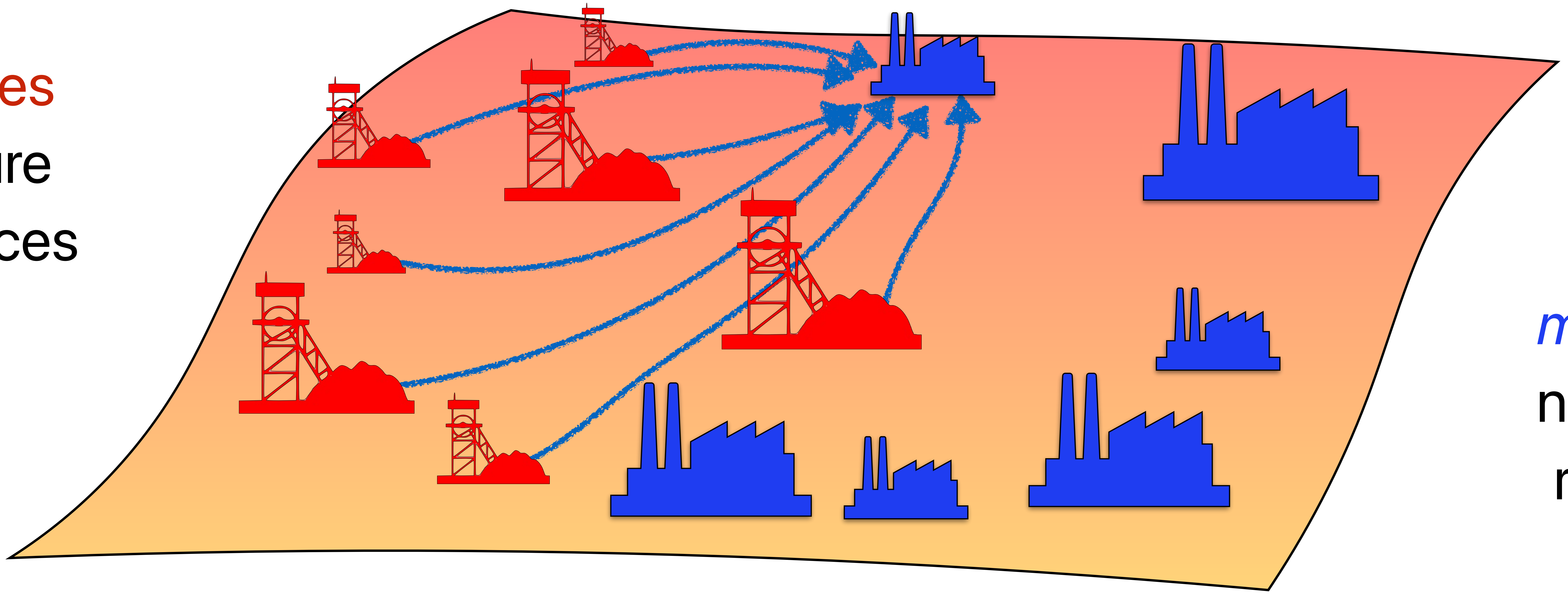
n mines capture resources



m factories need those resources

Kantorovich Formulation of Optimal Transport

n mines
capture
resources

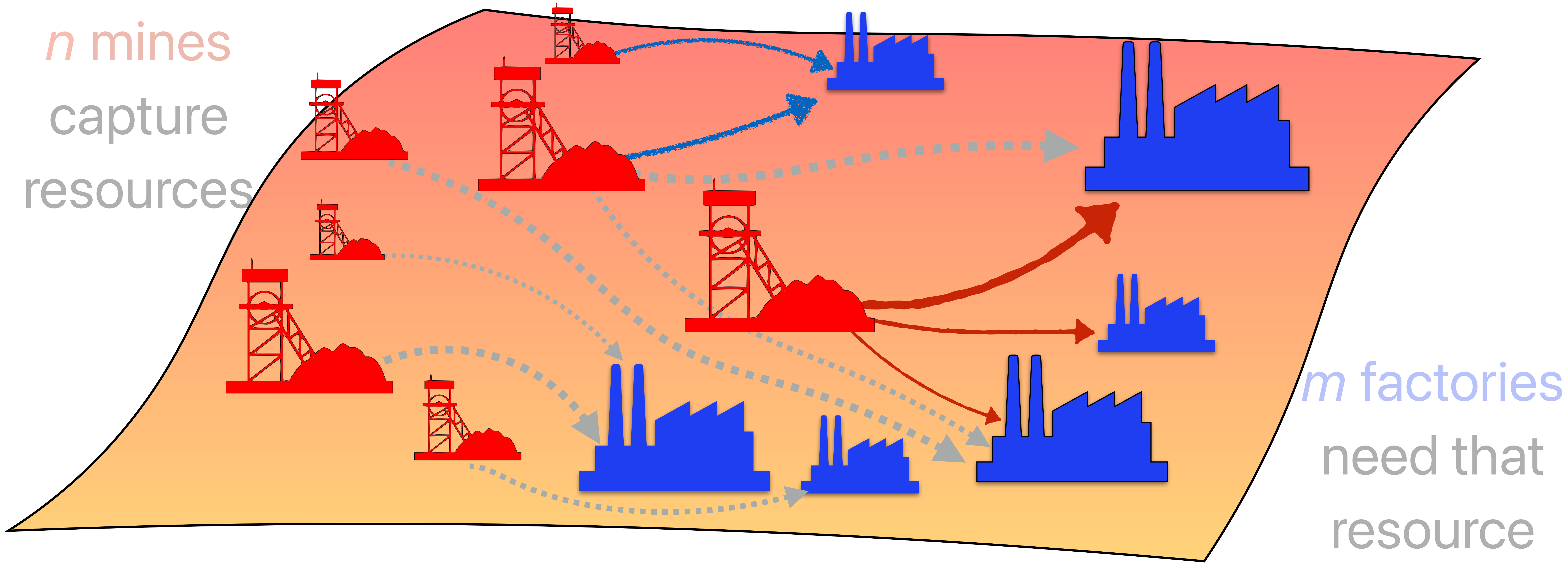


m factories
need those
resources

Each of the *m* factories must get resource it needs

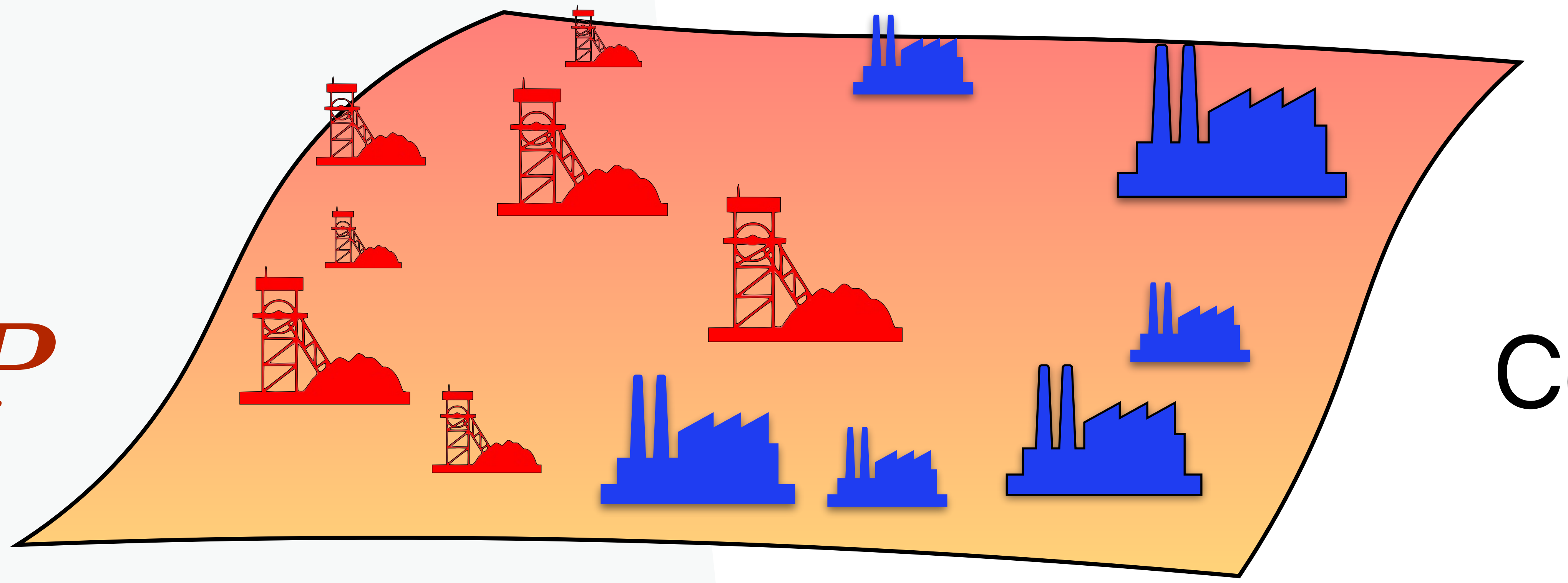
OT on Two Empirical Measures

Optimal Transport computes the **least-costly** transport matrix

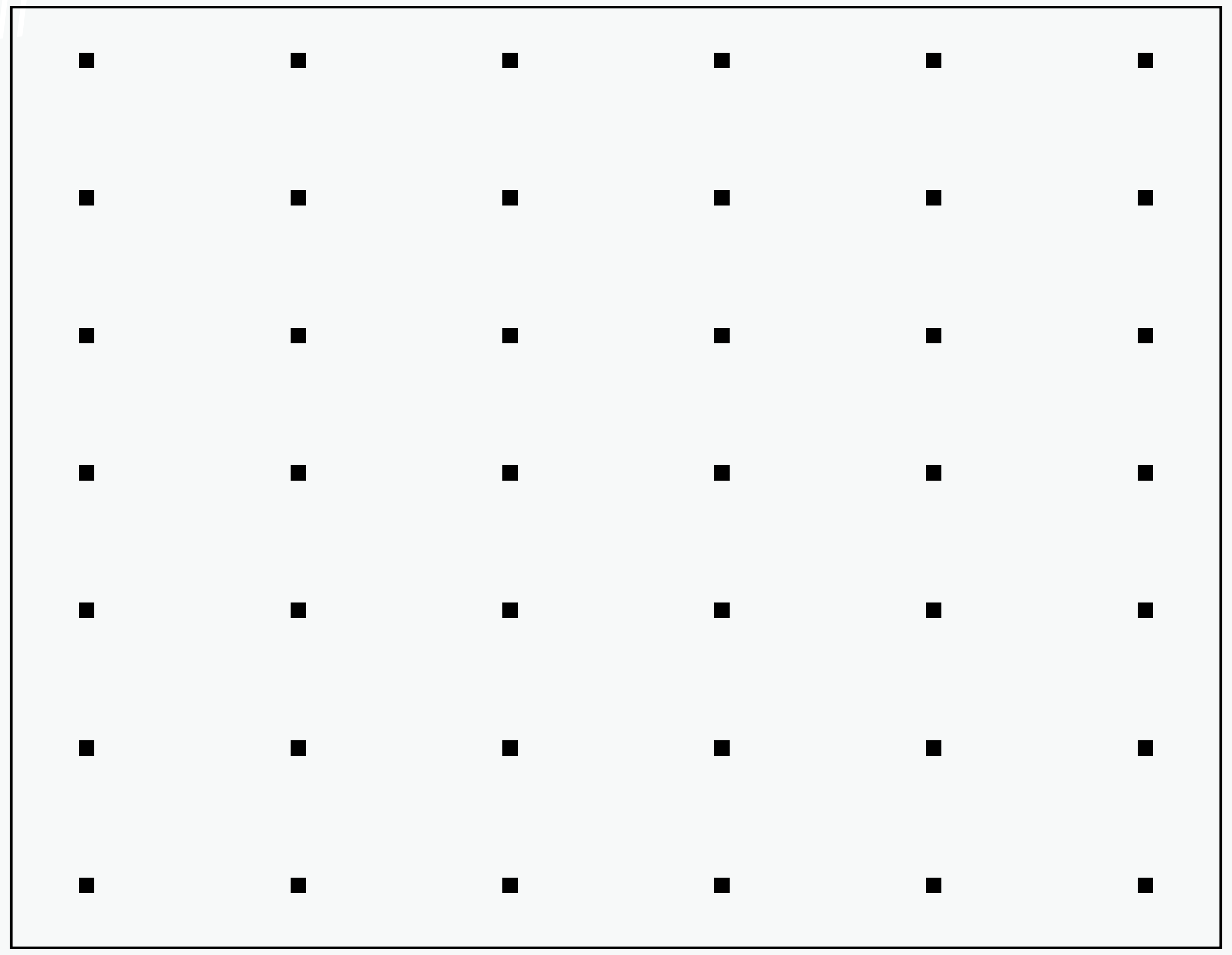


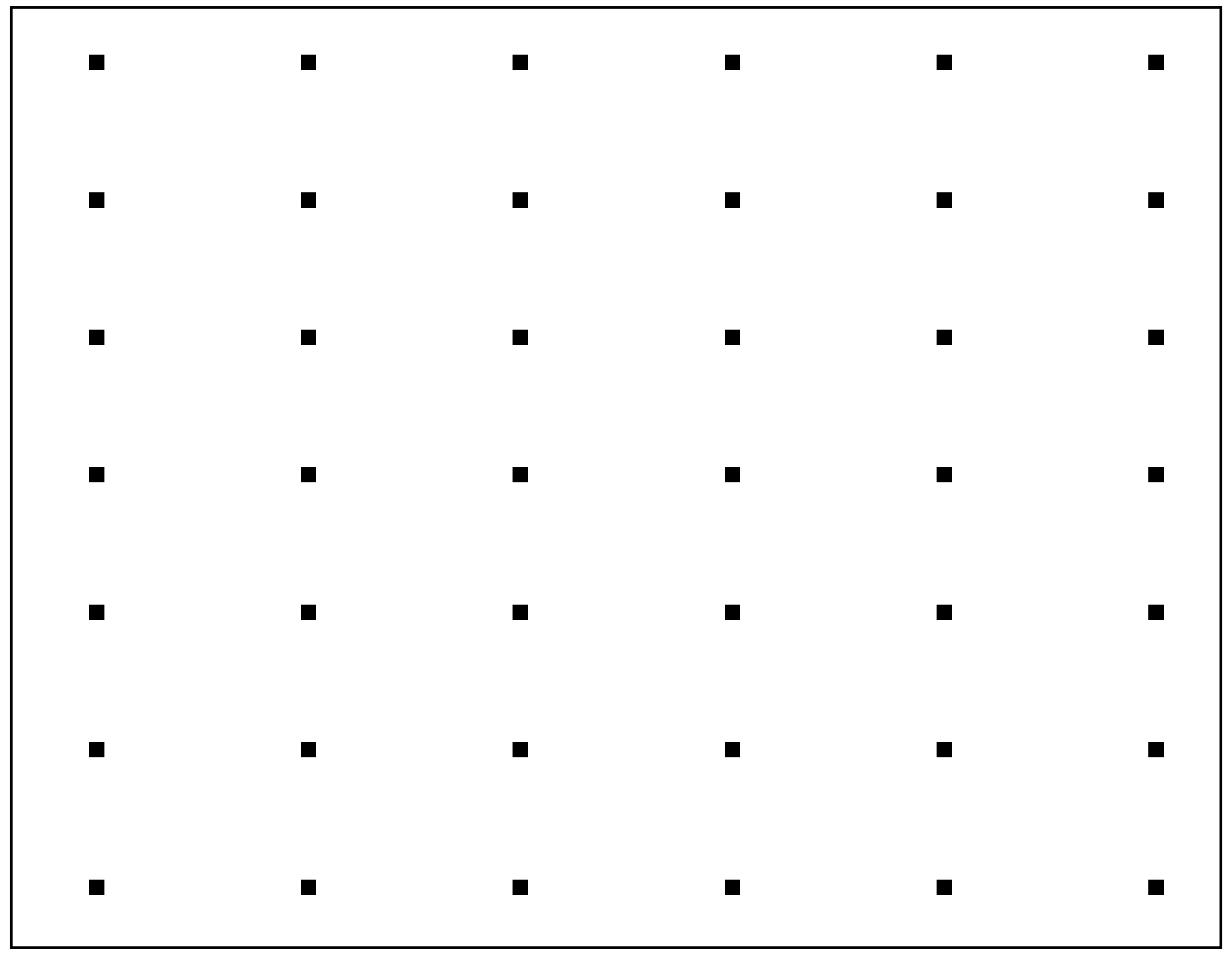
OT on Two Empirical Measures: Matrices

Couplings $\mathbf{P}$



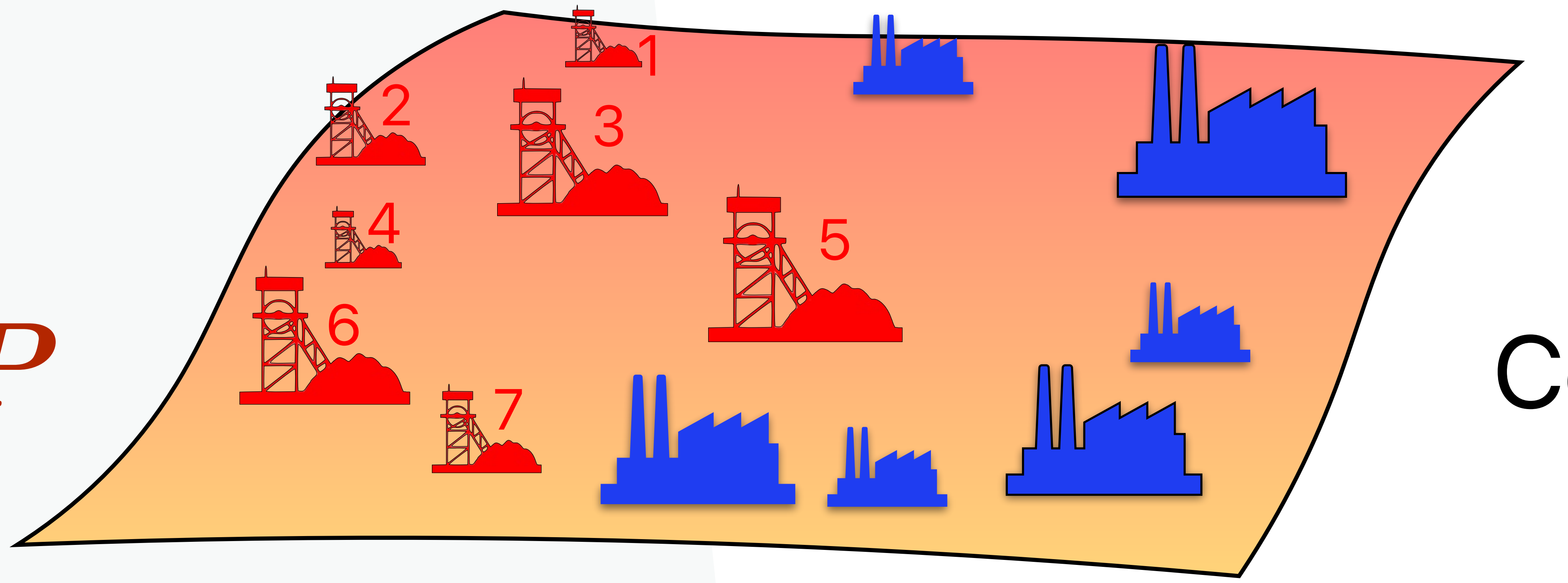
Cost matrix $M_{\mathbf{X}\mathbf{Y}}$



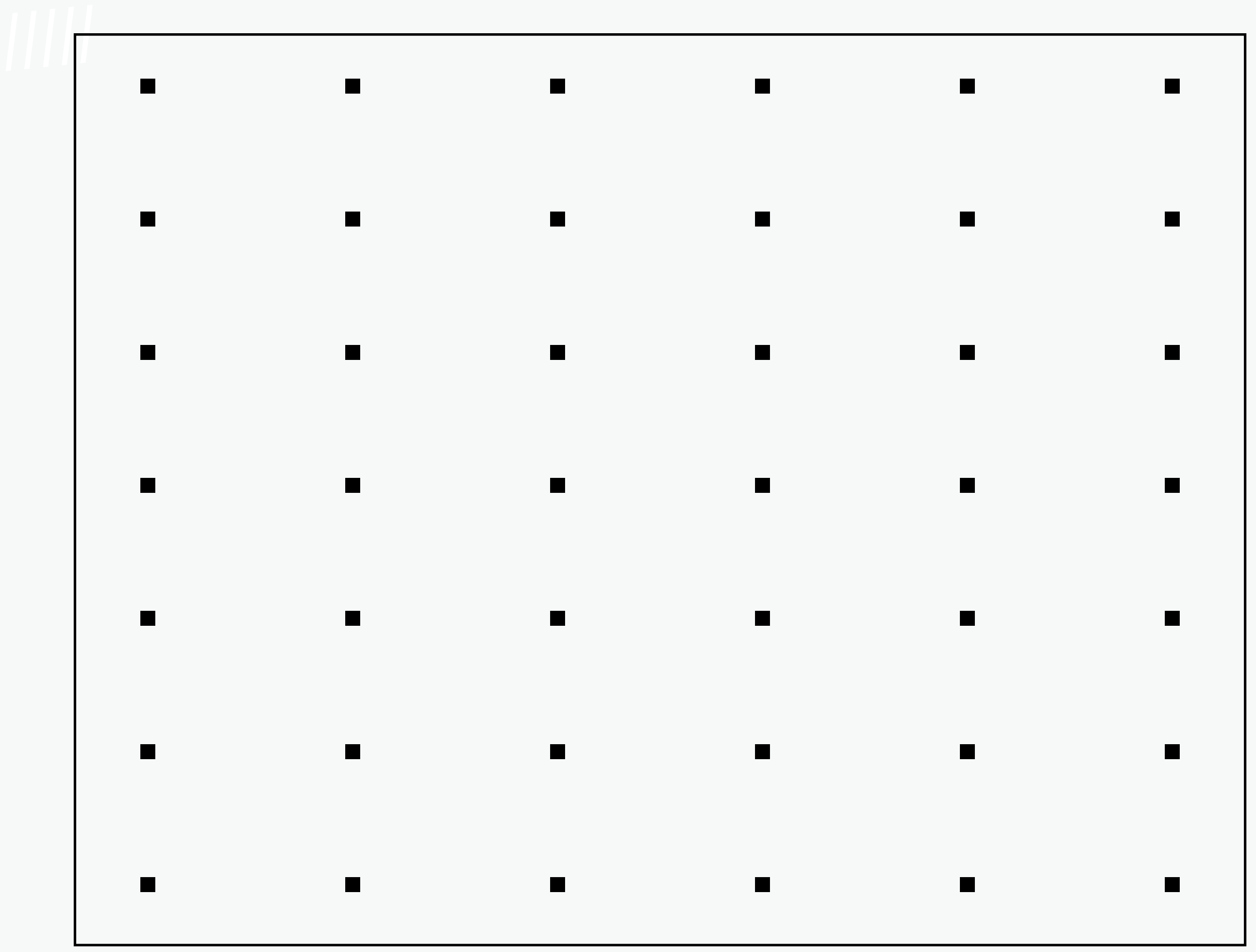


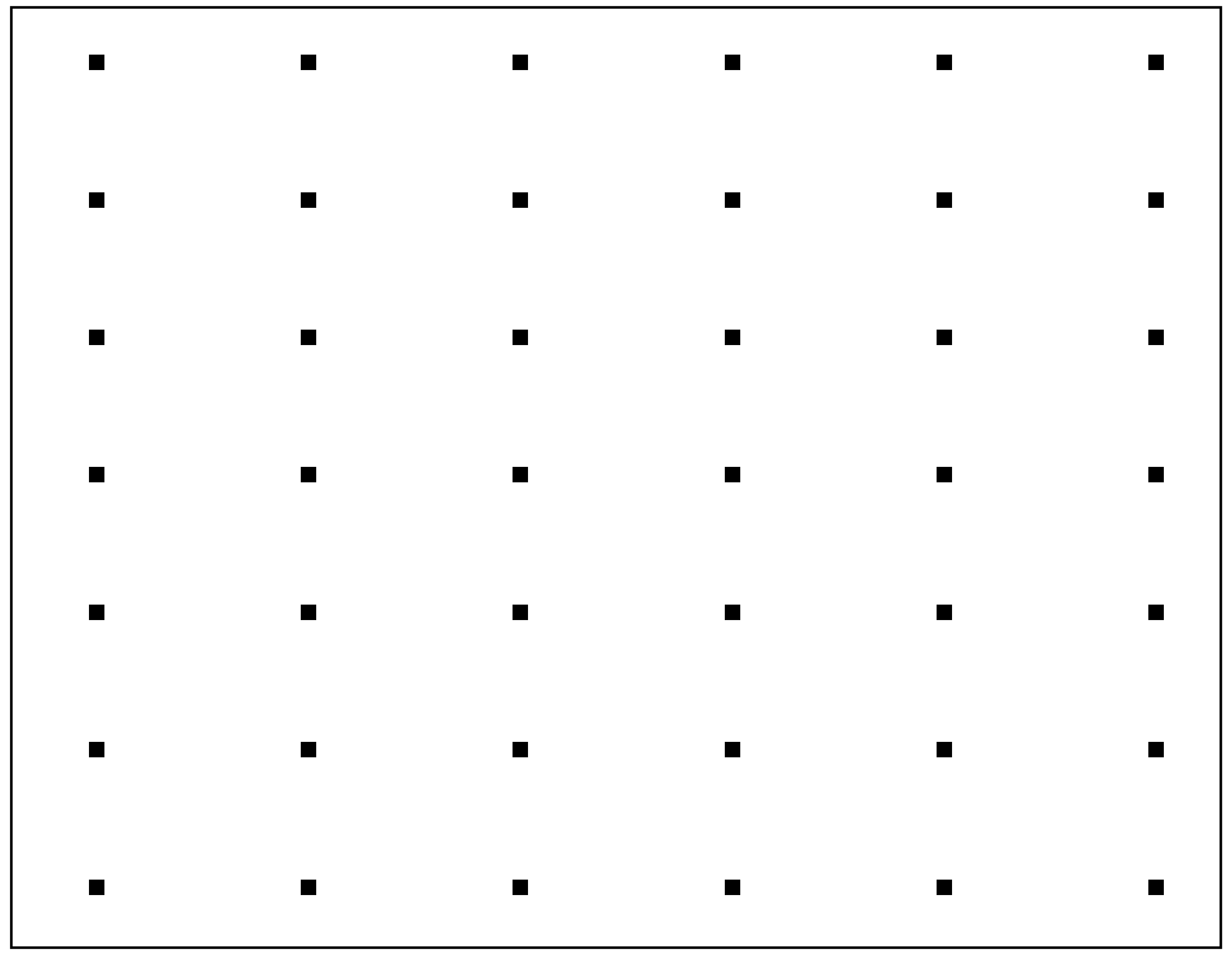
OT on Two Empirical Measures: Matrices

Couplings P



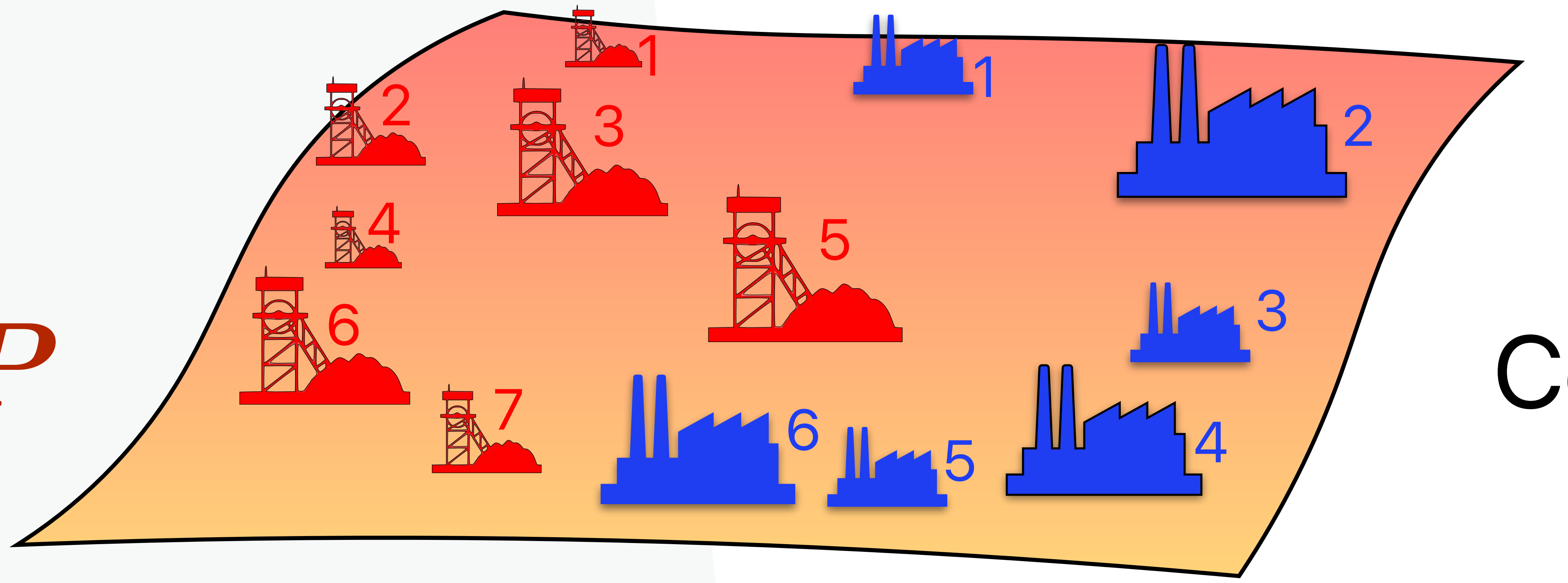
Cost matrix $M_{\mathbf{X}\mathbf{Y}}$



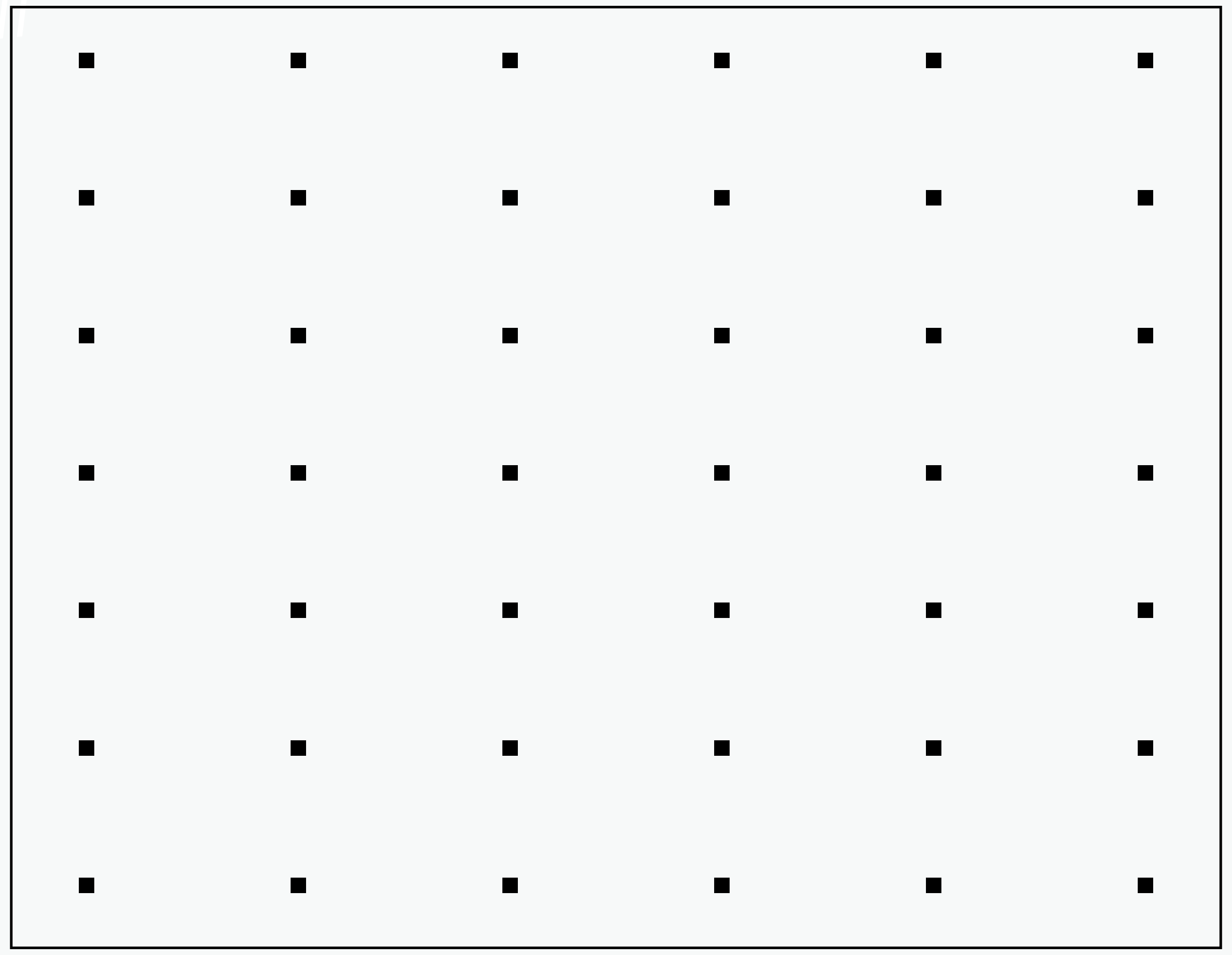


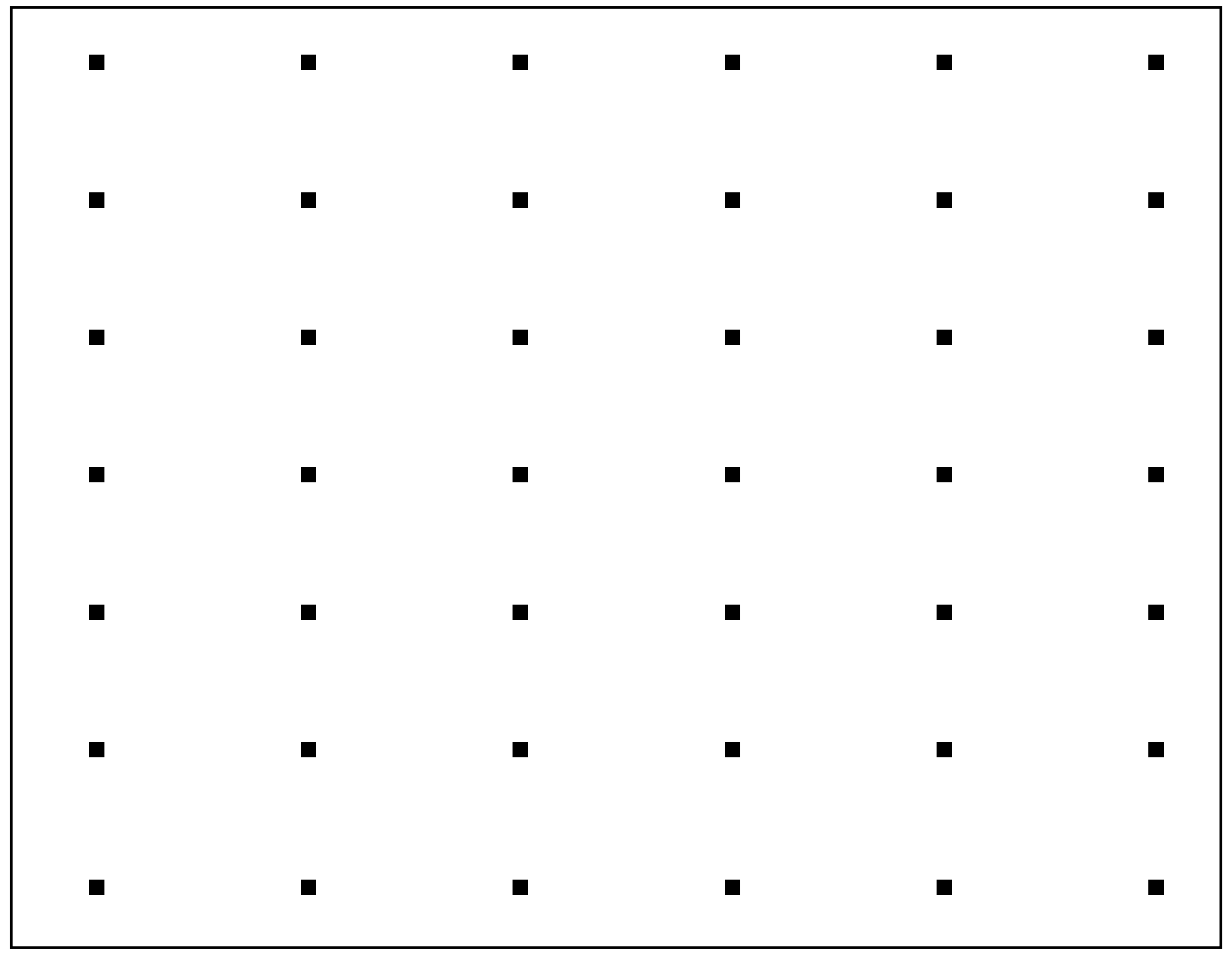
OT on Two Empirical Measures: Matrices

Couplings $\mathbf{P}$

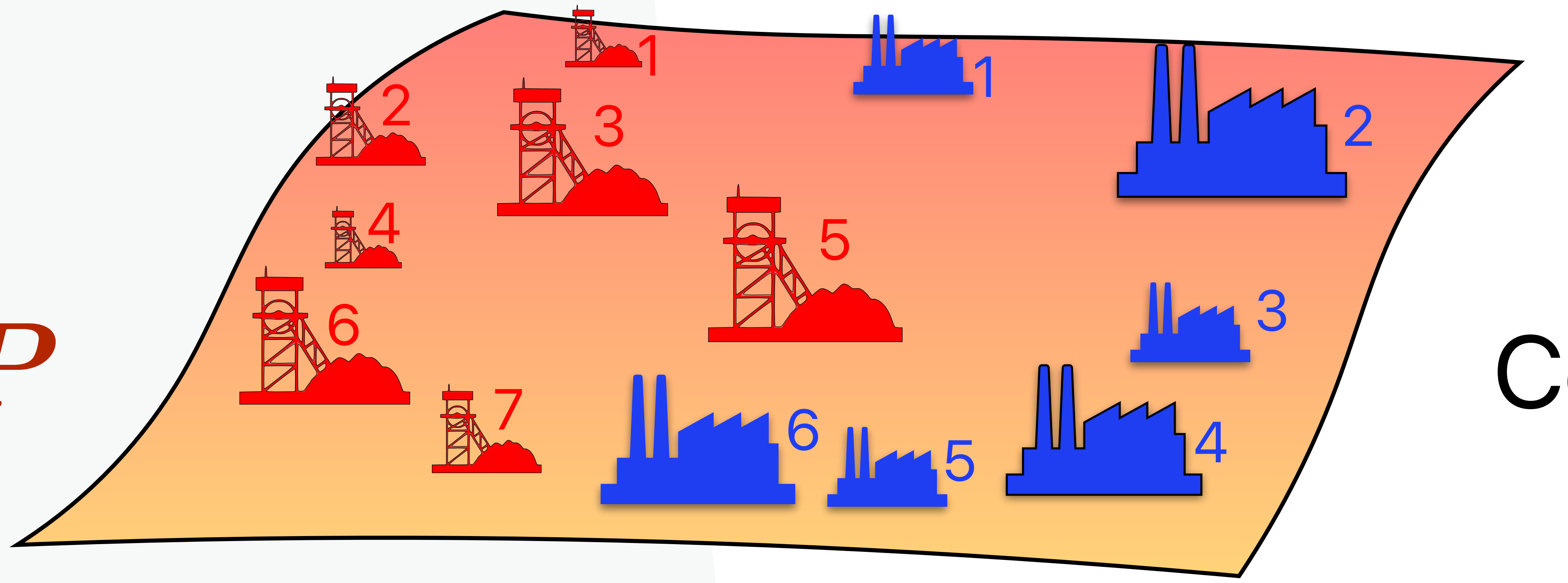


Cost matrix $M_{\mathbf{XY}}$

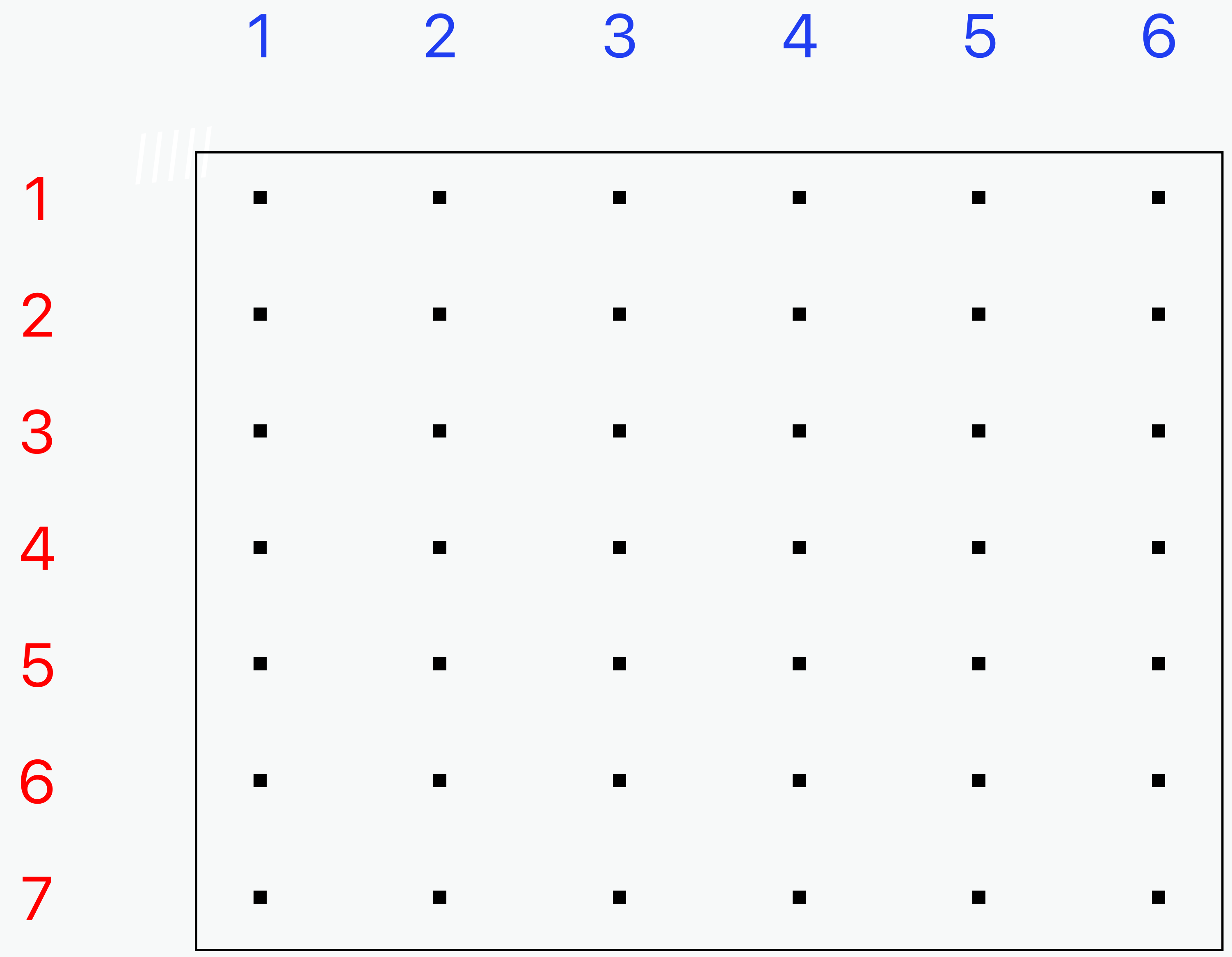




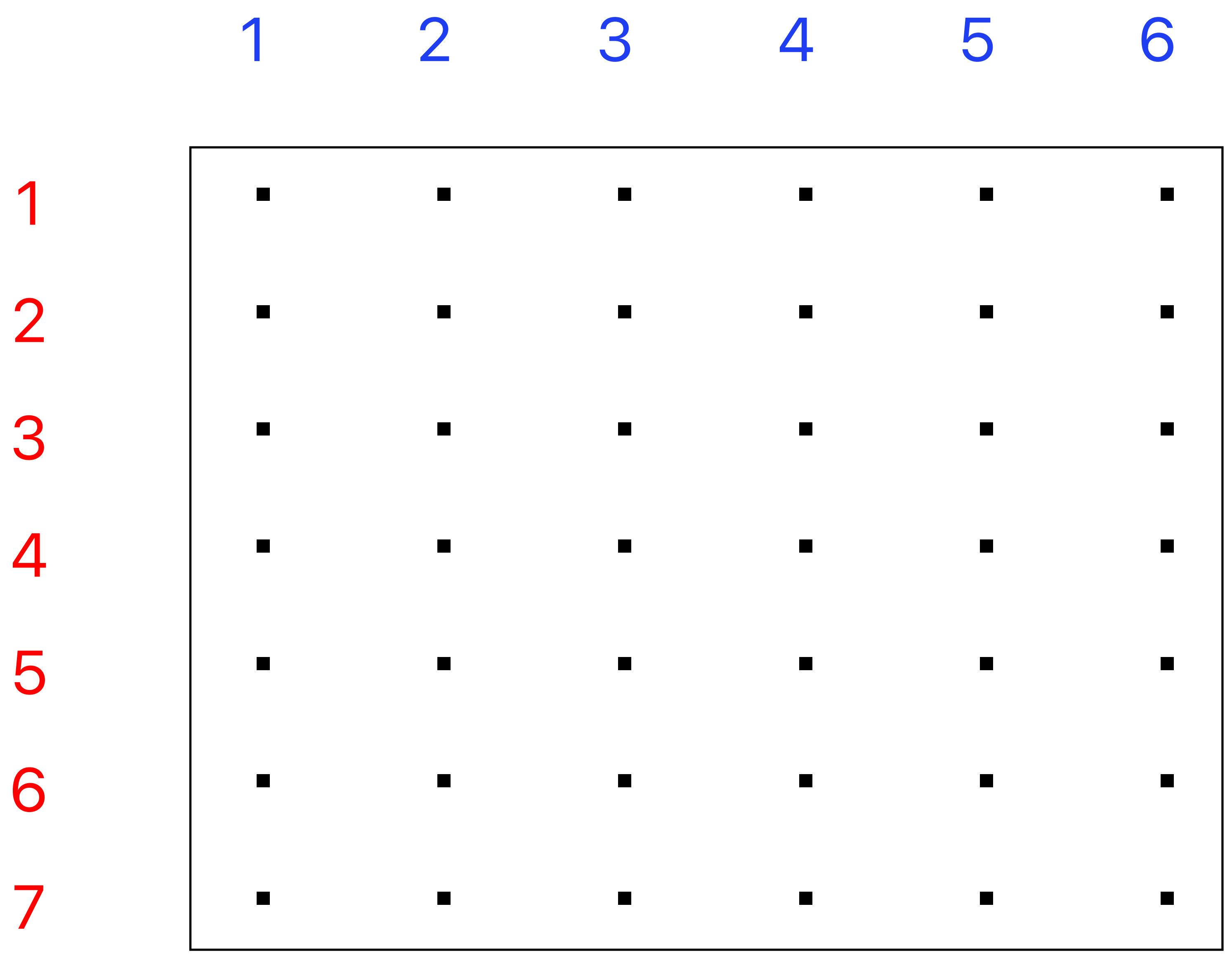
OT on Two Empirical Measures: Matrices



Couplings $\boldsymbol{P}$

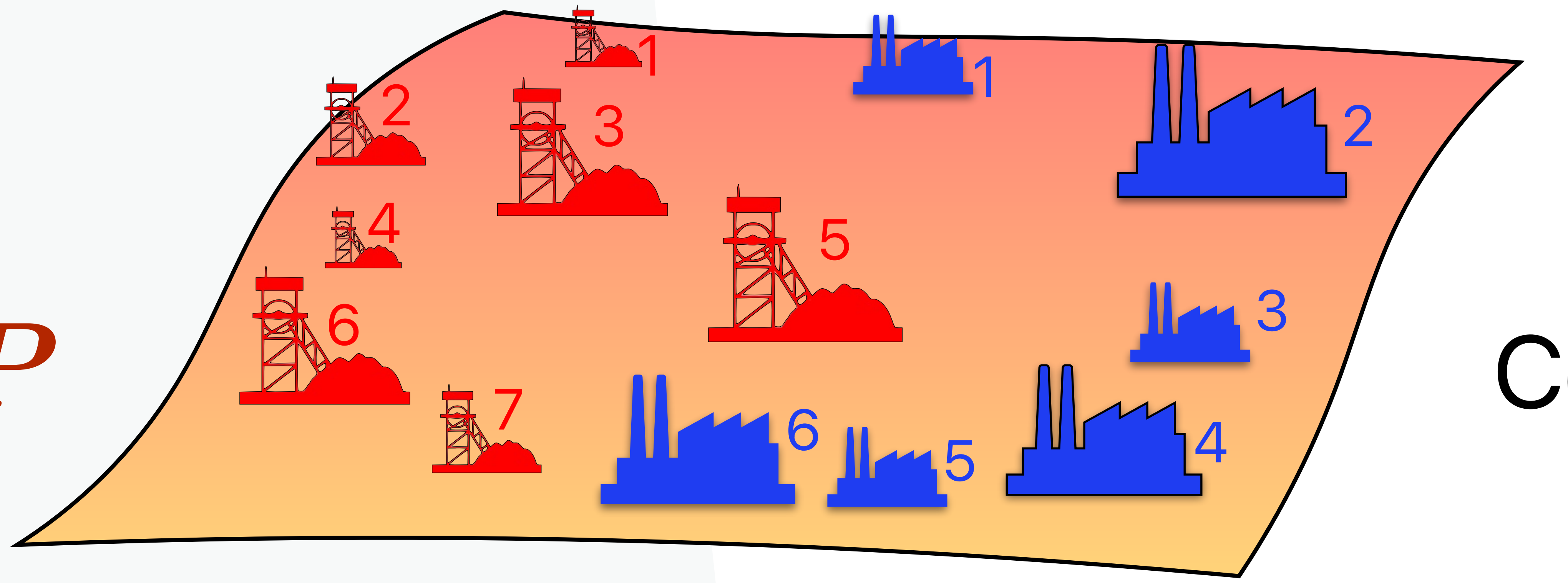


Cost matrix $M_{\boldsymbol{X}\boldsymbol{Y}}$



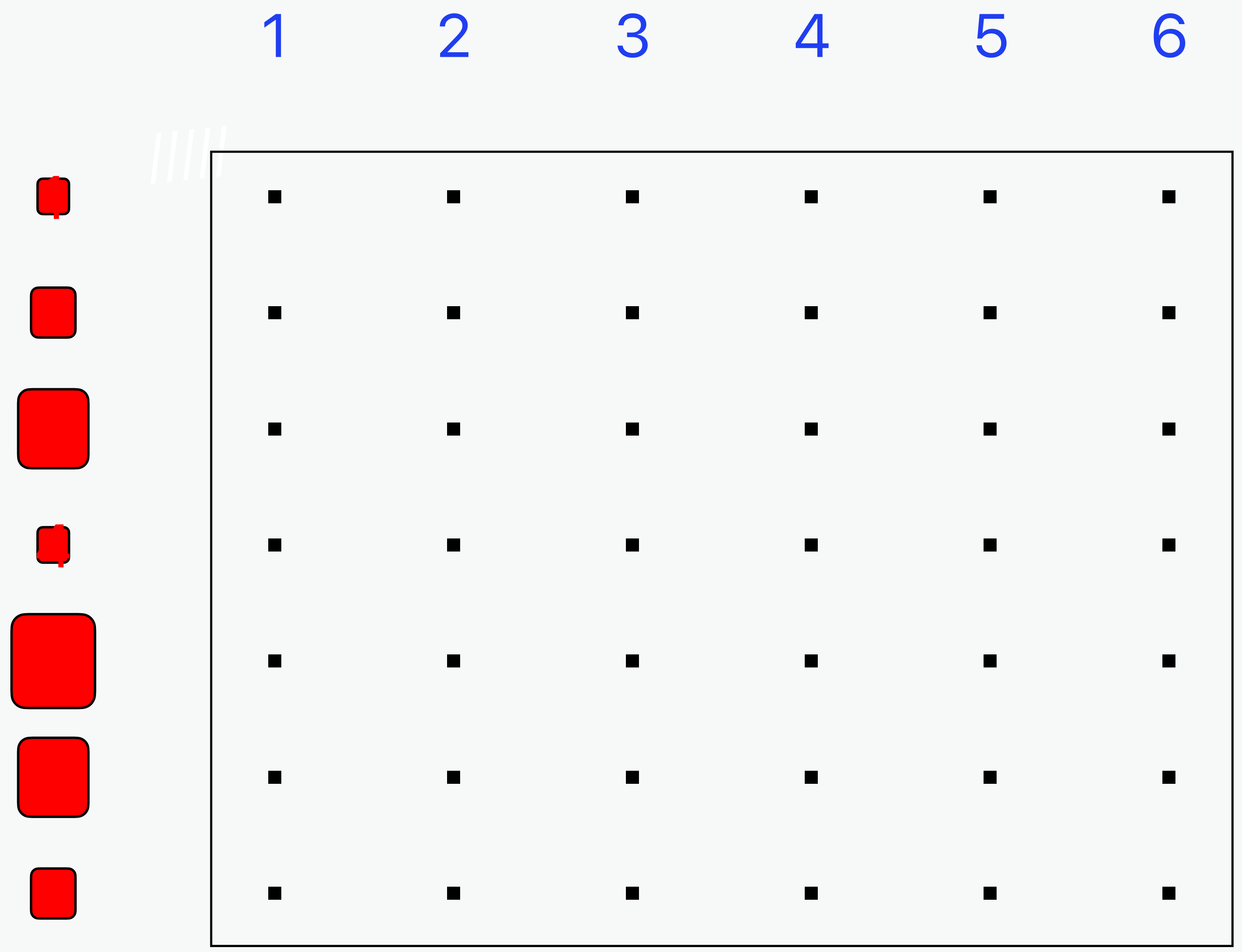
OT on Two Empirical Measures: Matrices

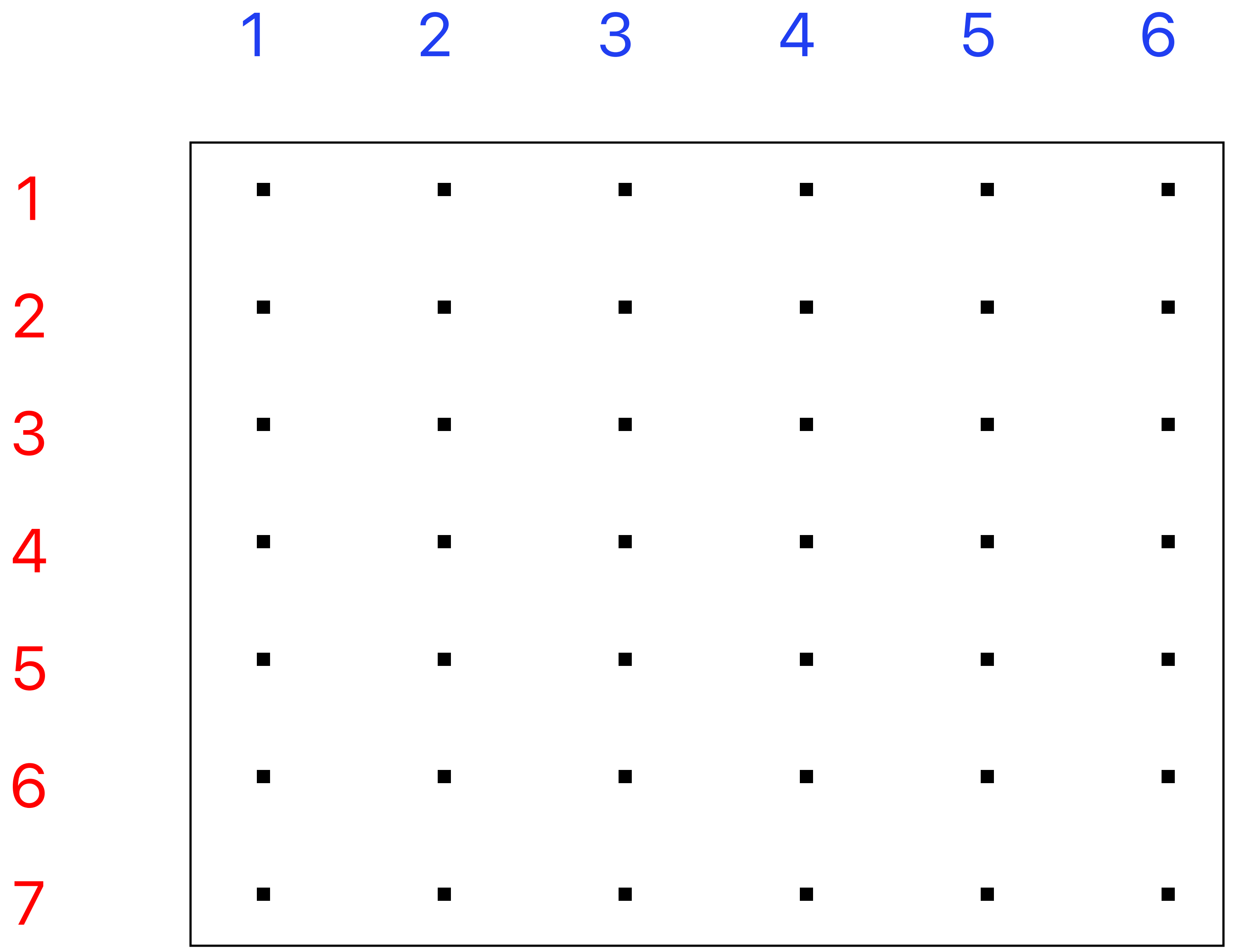
Couplings P



Cost matrix M_{XY}

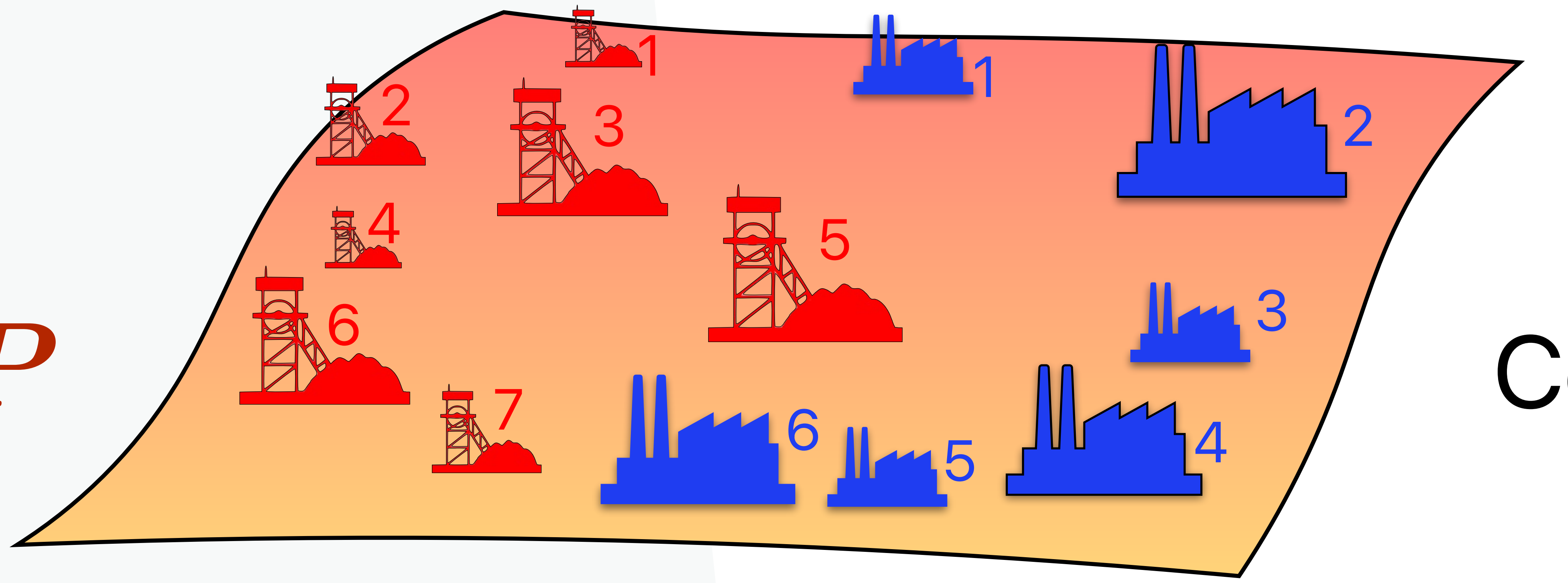
a



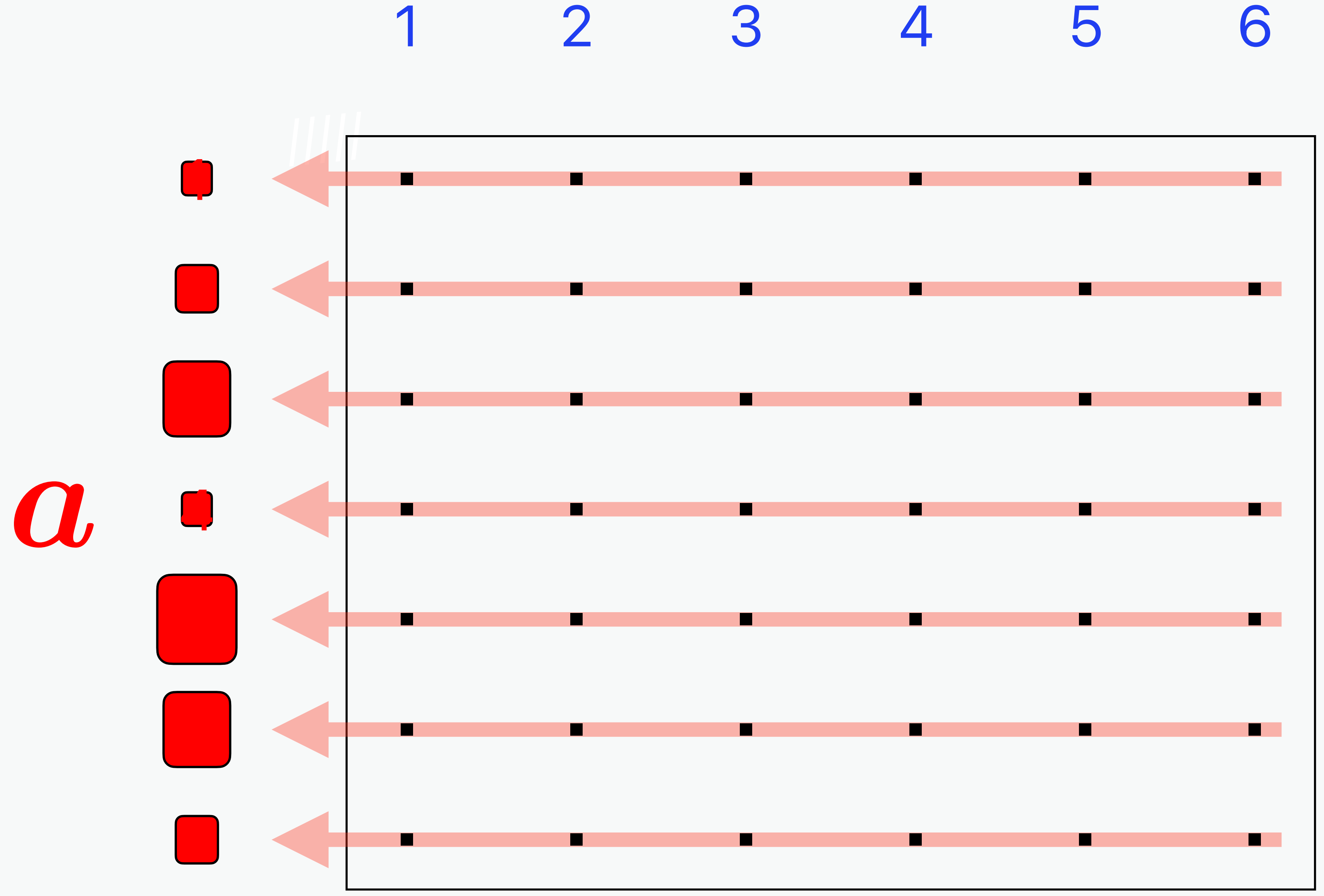


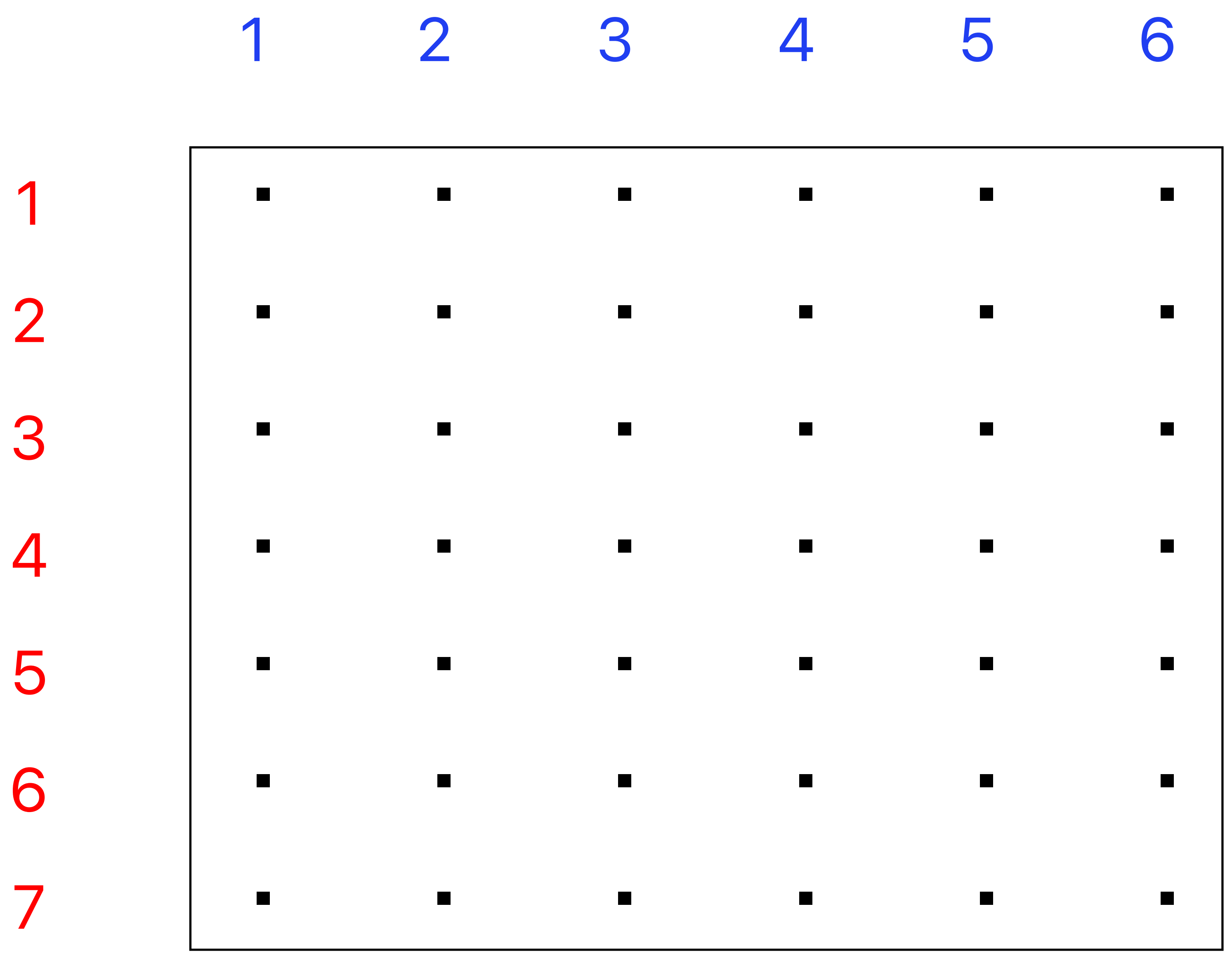
OT on Two Empirical Measures: Matrices

Couplings P

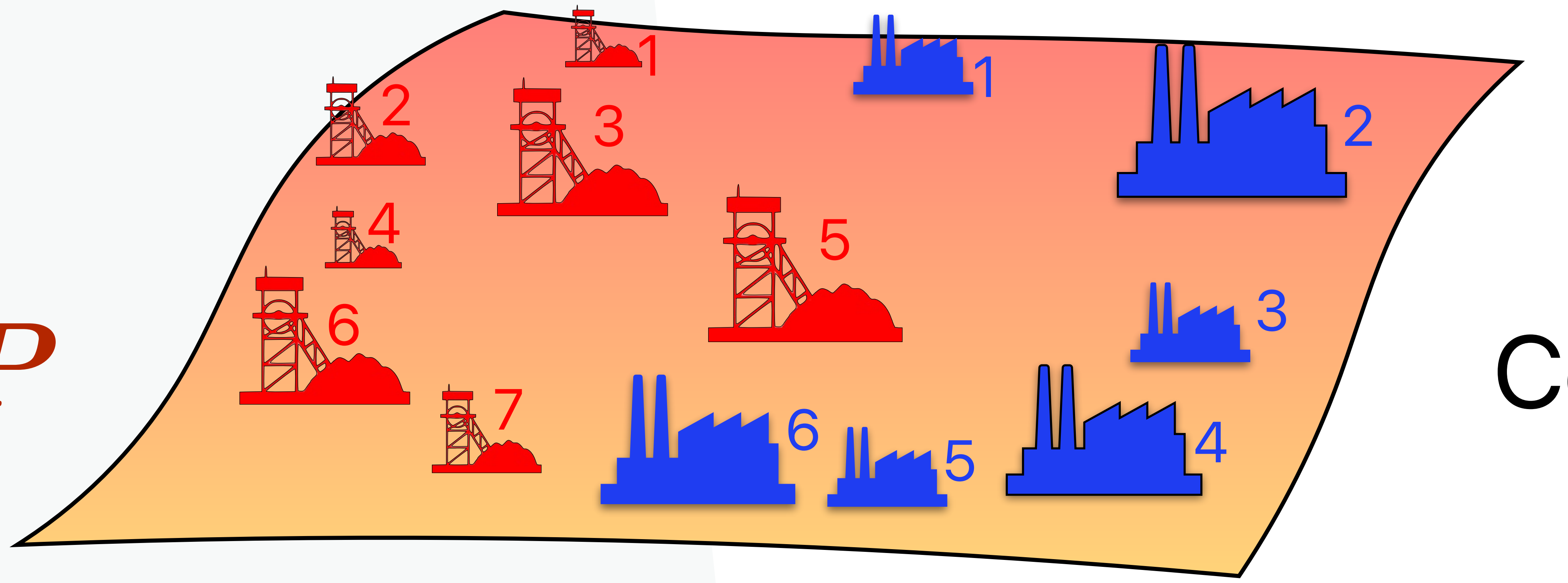


Cost matrix M_{XY}

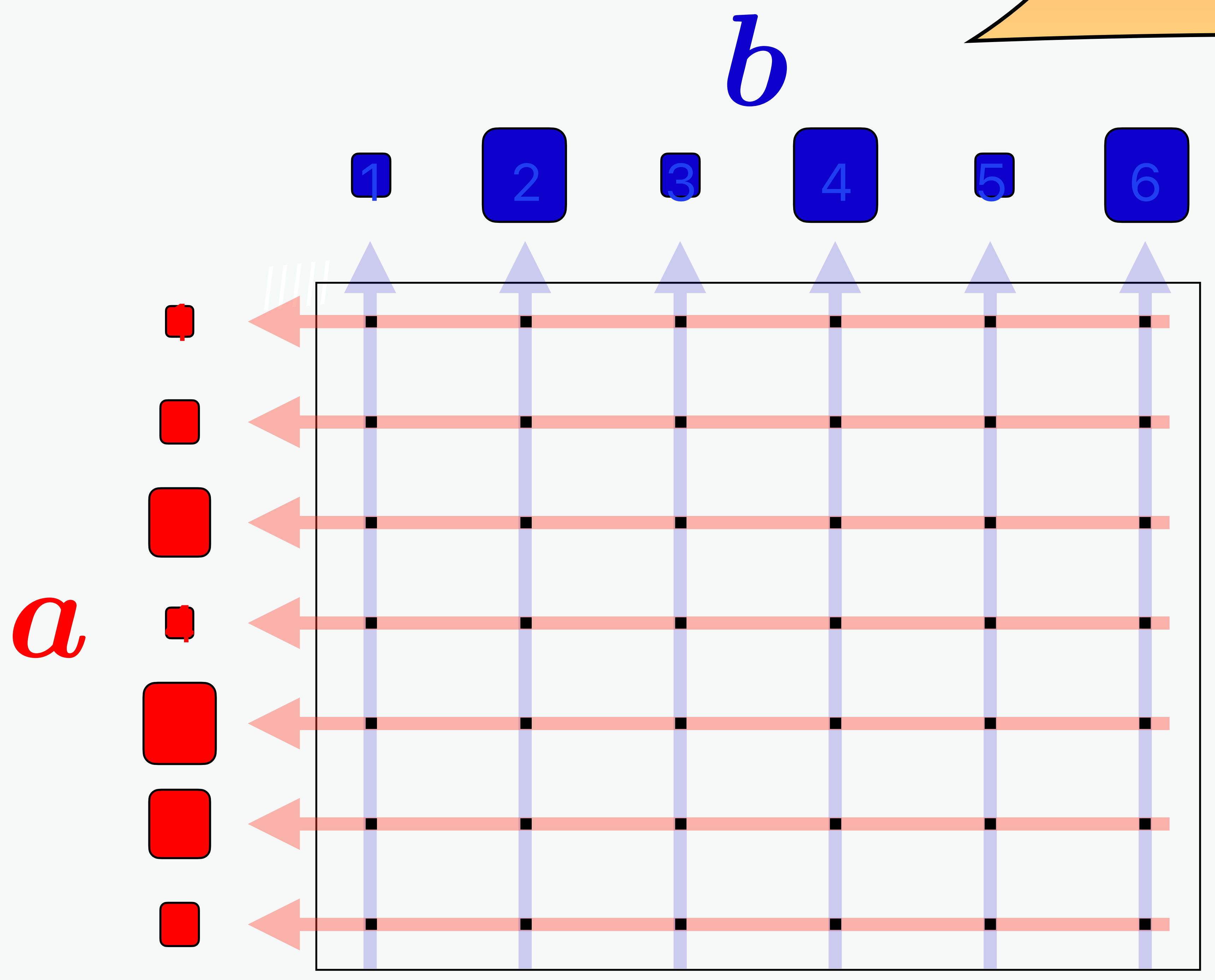




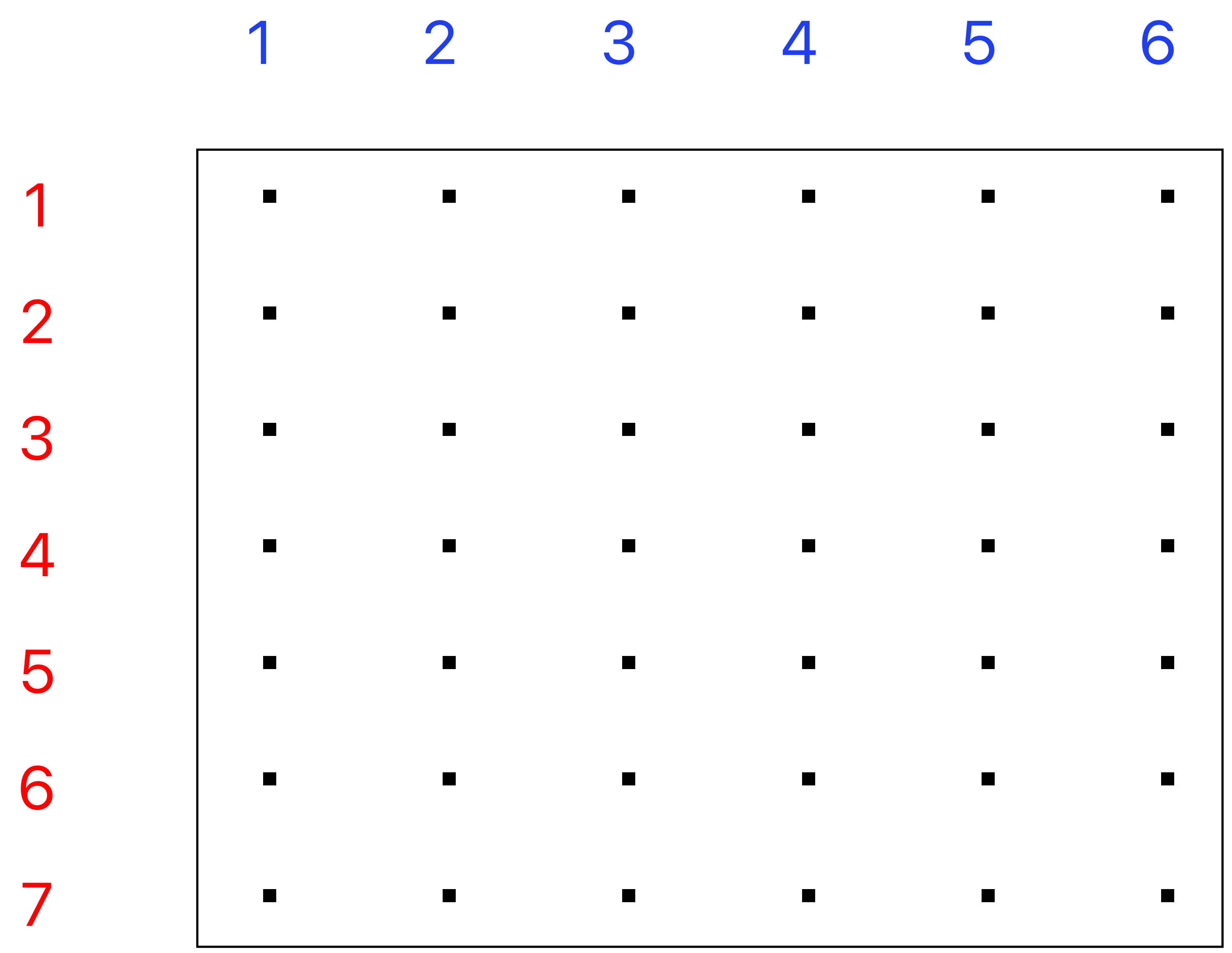
OT on Two Empirical Measures: Matrices



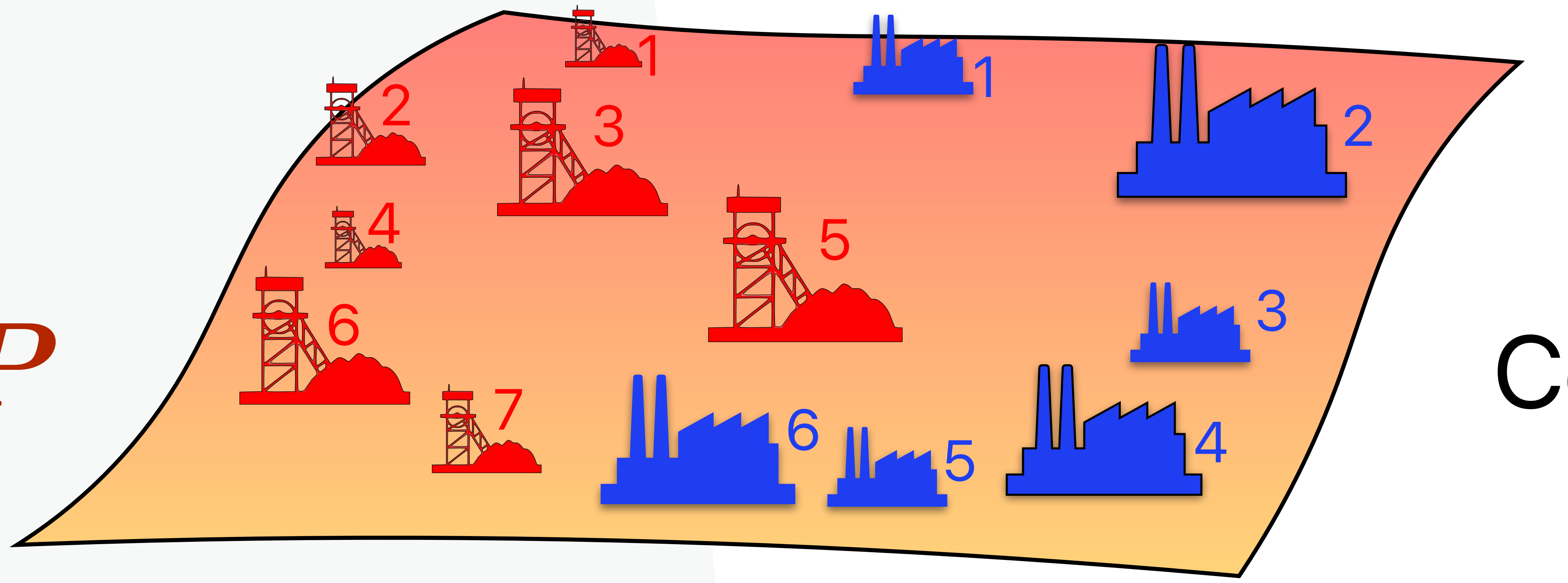
Couplings $\boldsymbol{P}$



Cost matrix $M_{\boldsymbol{XY}}$

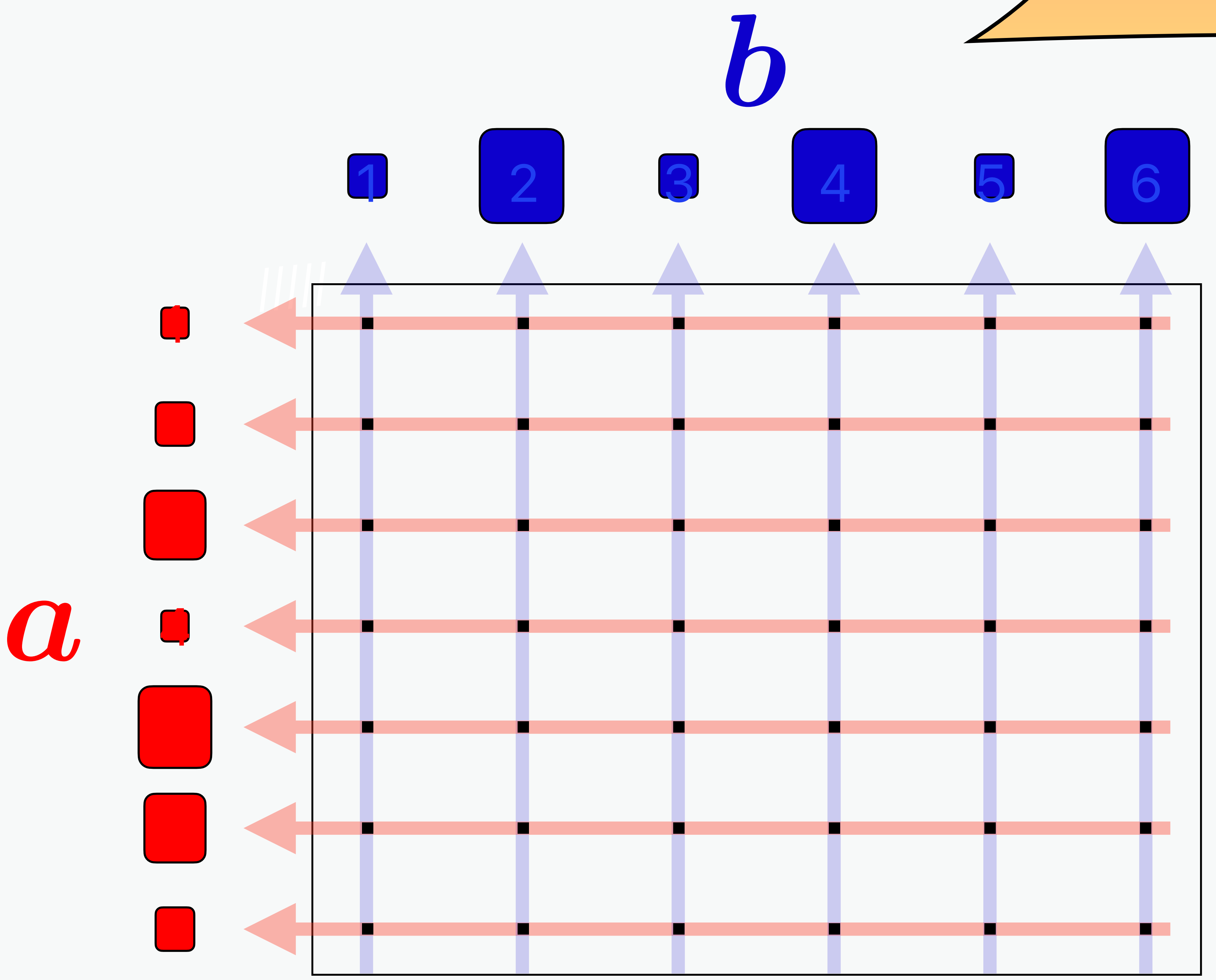


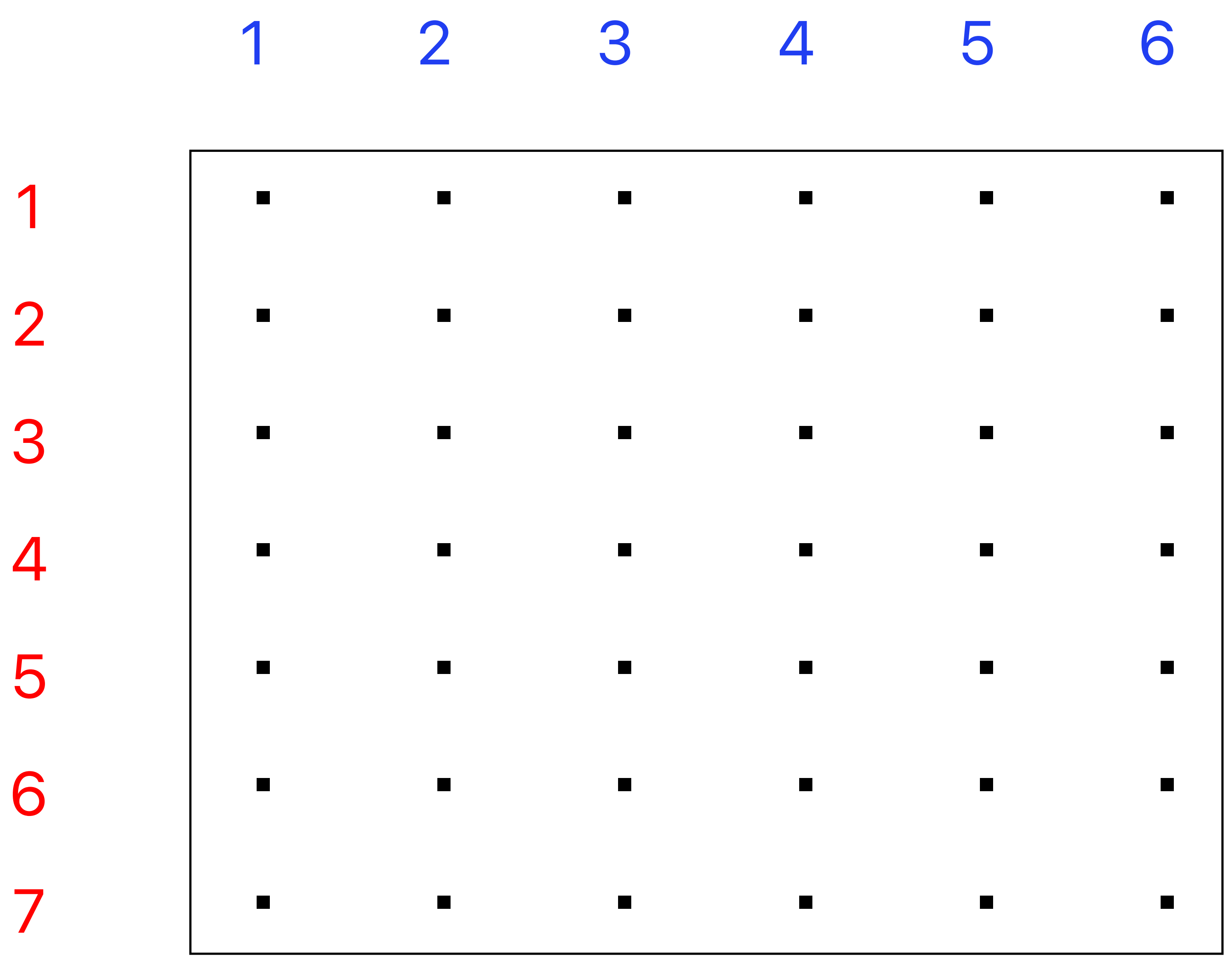
OT on Two Empirical Measures: Matrices



Couplings $\boldsymbol{P}$

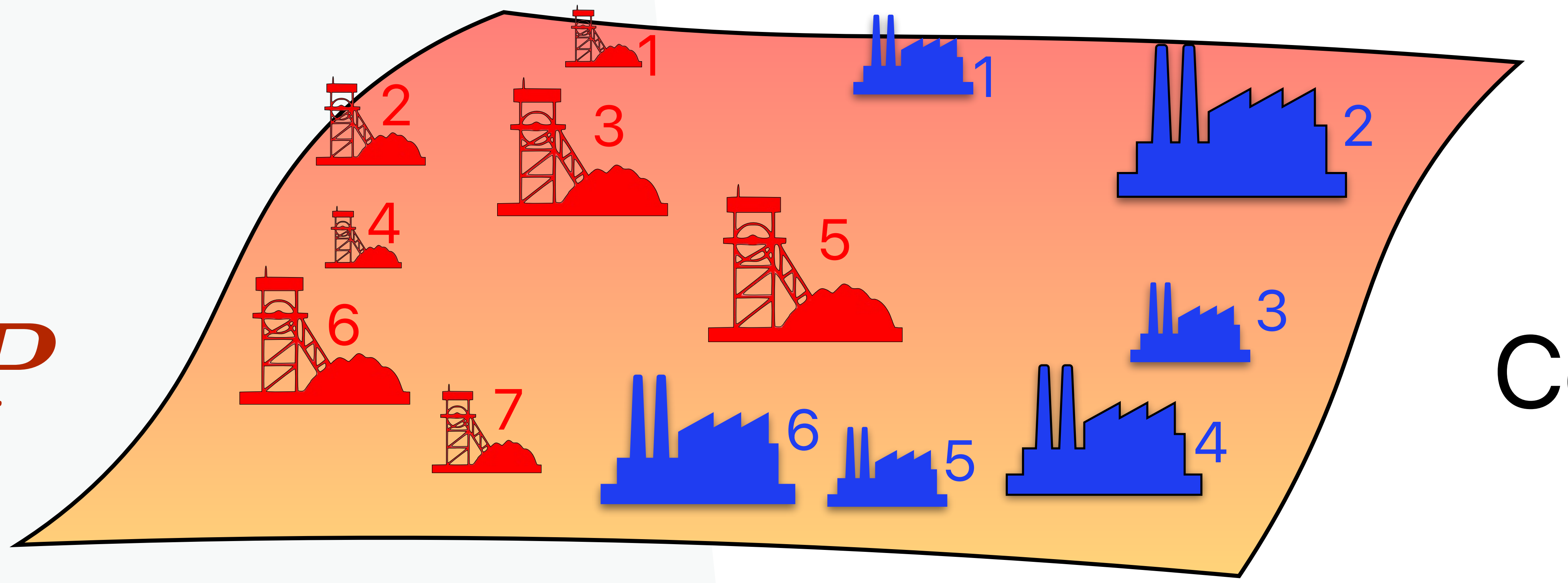
Cost matrix $M_{\boldsymbol{XY}}$





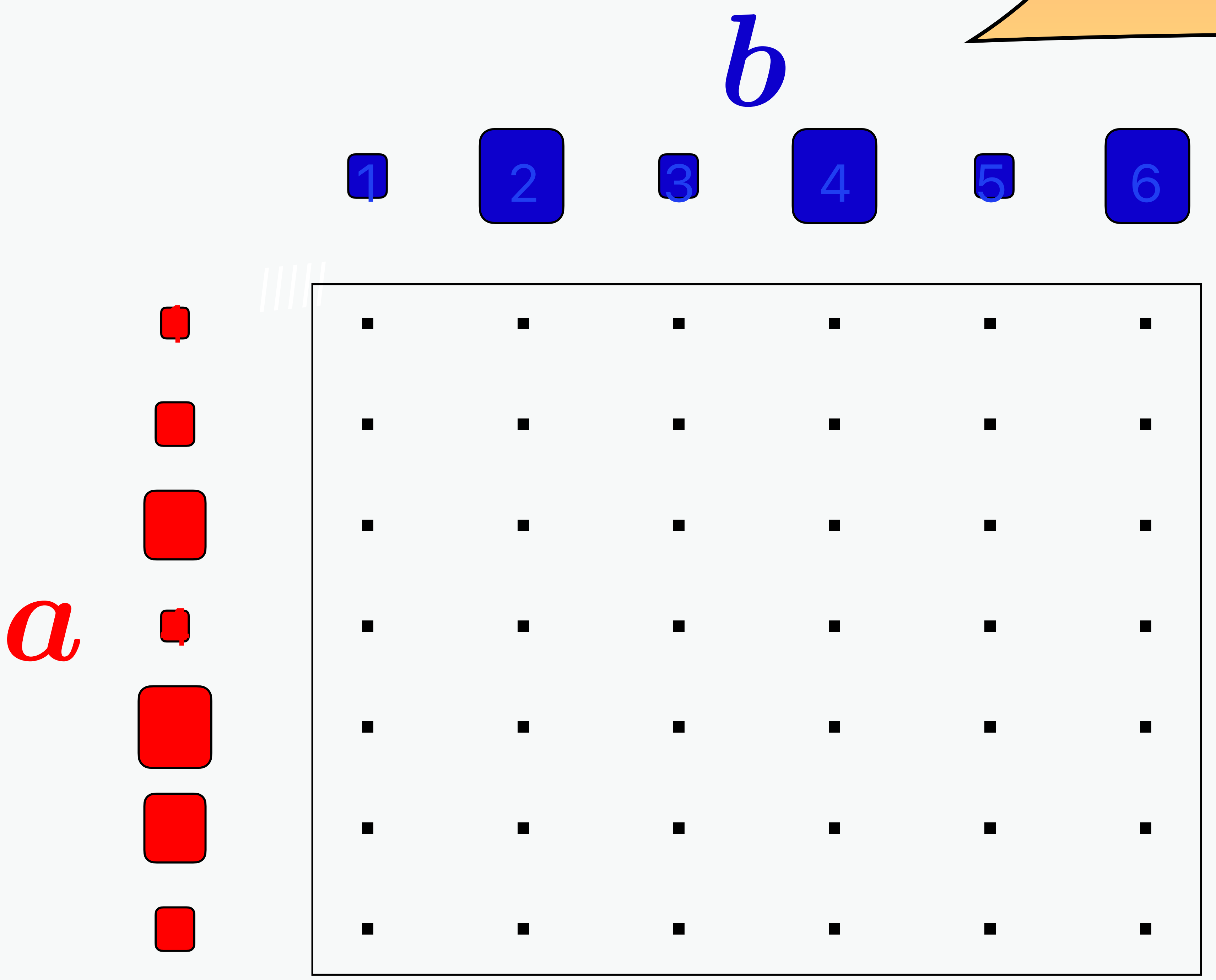
$$U(\boldsymbol{a},\boldsymbol{b}) \stackrel{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}_+^{n\times m} | \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_n = \boldsymbol{b}\}$$

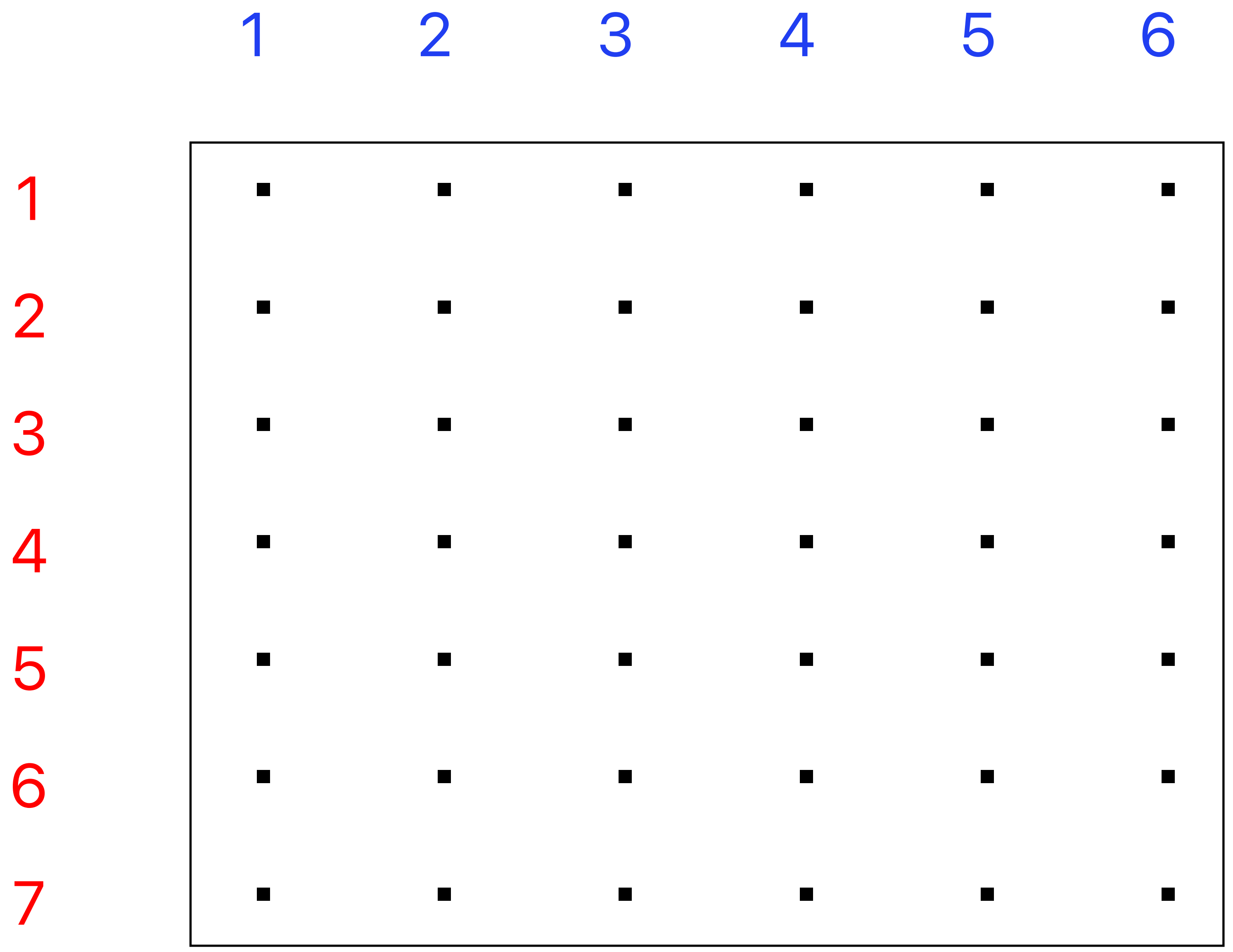
OT on Two Empirical Measures: Matrices



Couplings $\boldsymbol{P}$

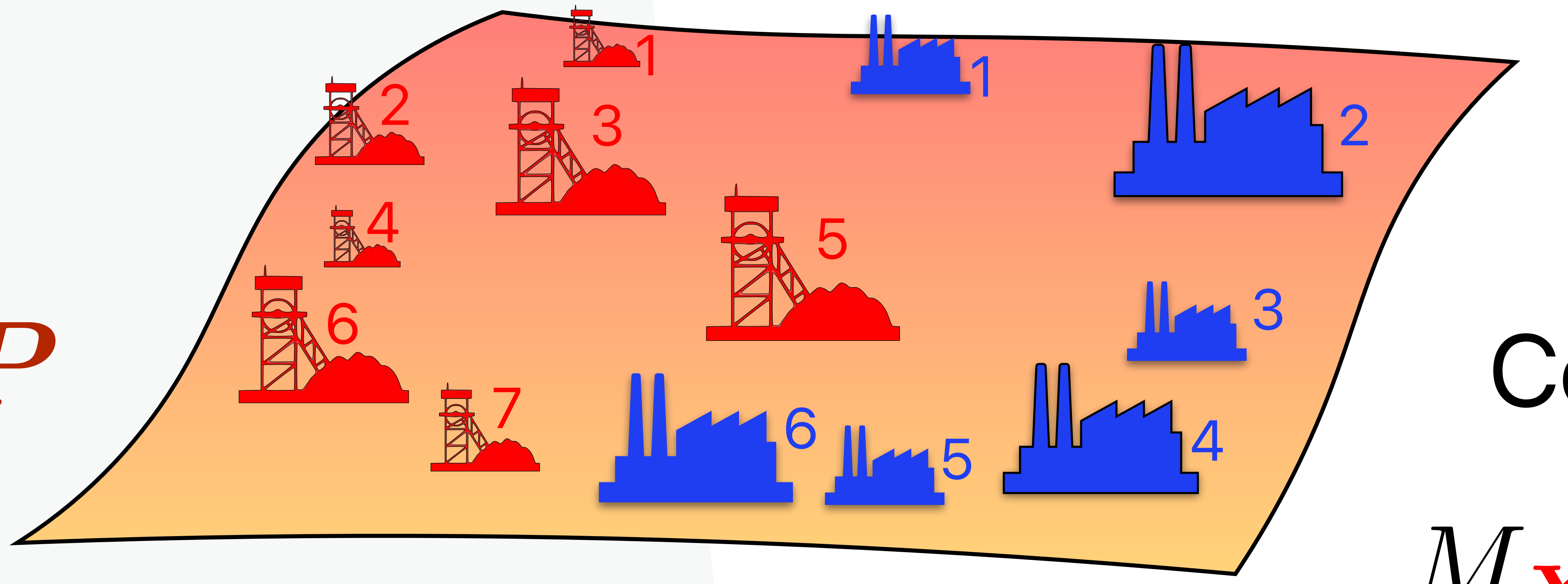
Cost matrix $M_{\boldsymbol{XY}}$



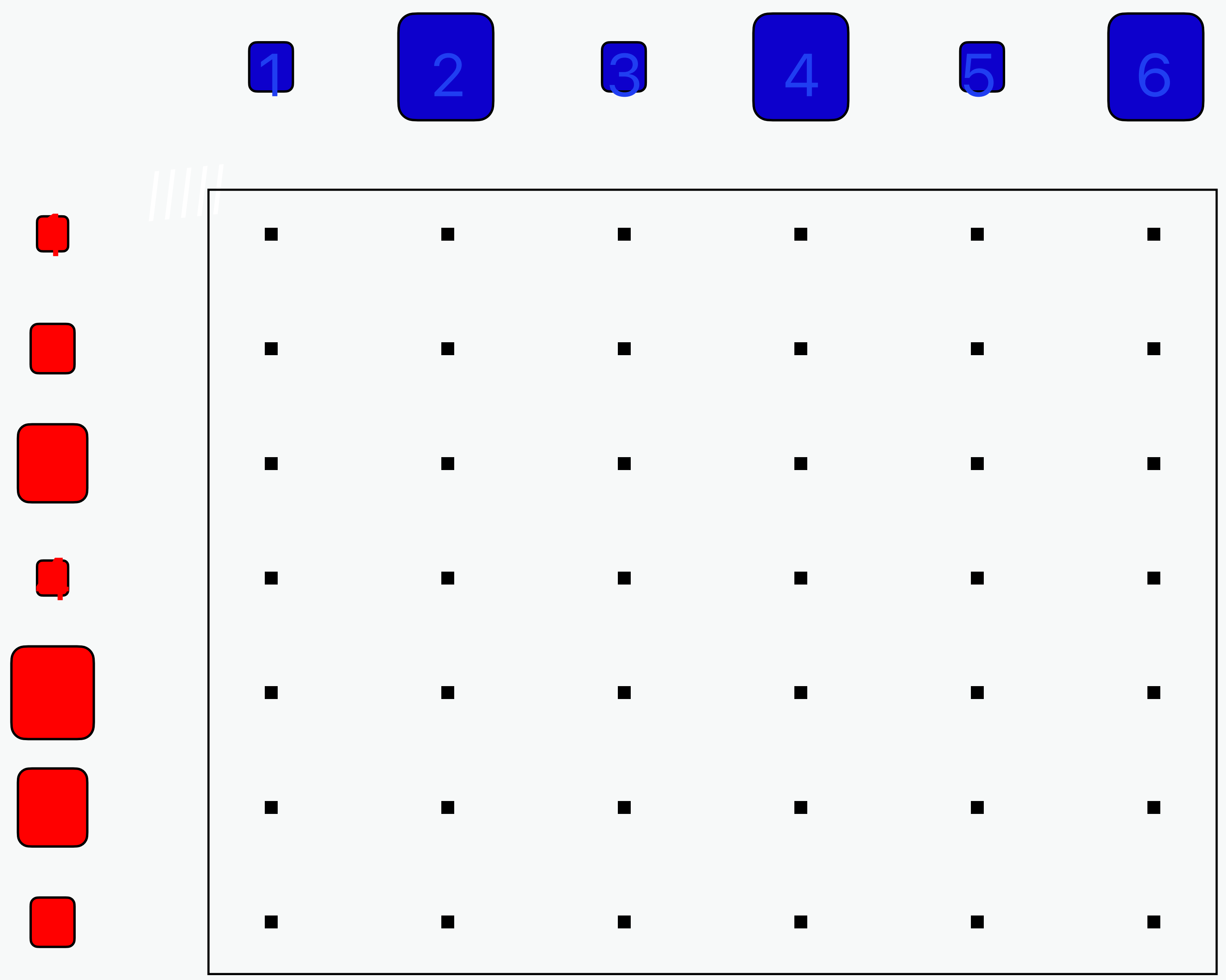


$$U(\boldsymbol{a},\boldsymbol{b}) \stackrel{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}_+^{n\times m} | \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_n = \boldsymbol{b}\}$$

OT on Two Empirical Measures: Matrices



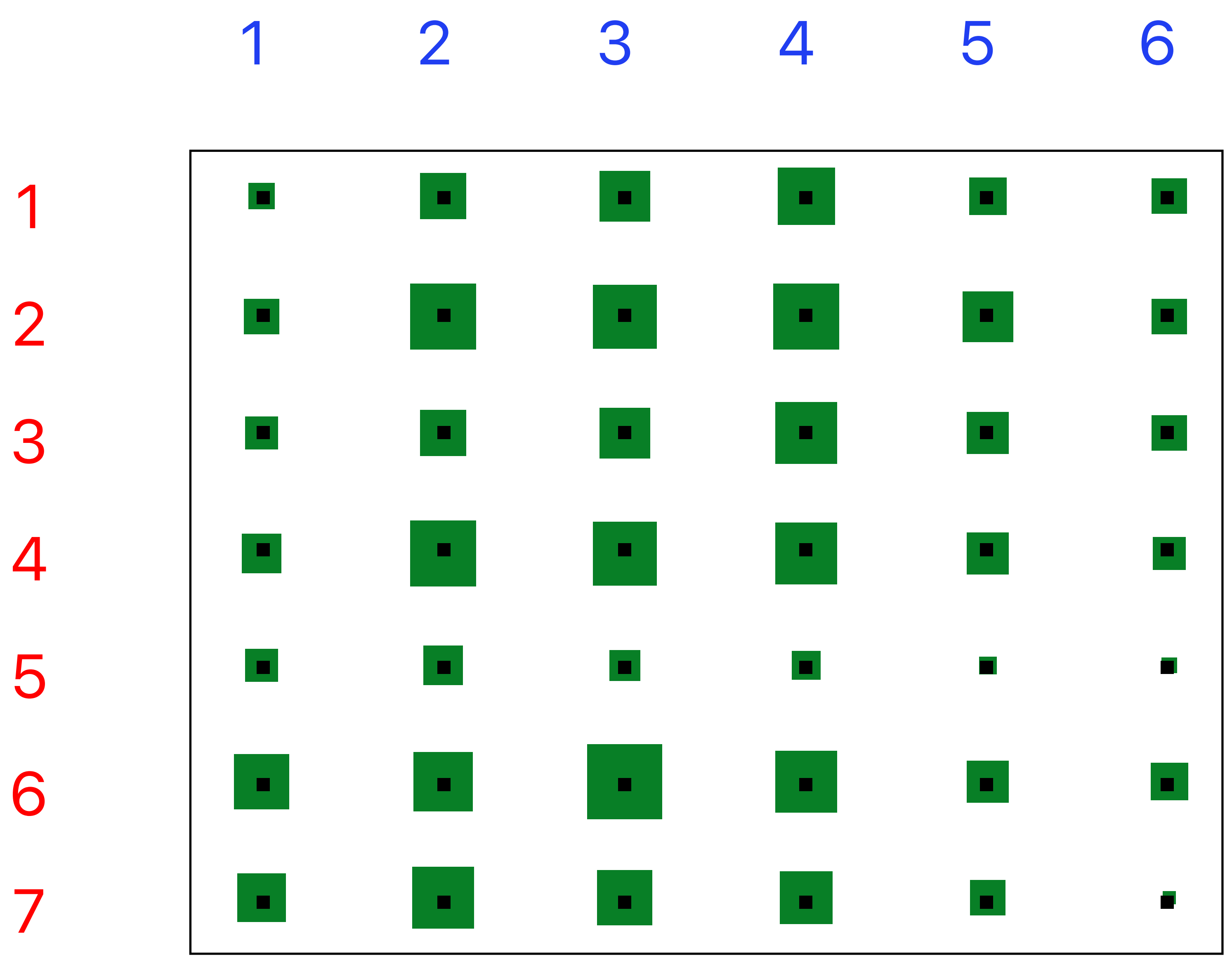
Couplings $\boldsymbol{P}$



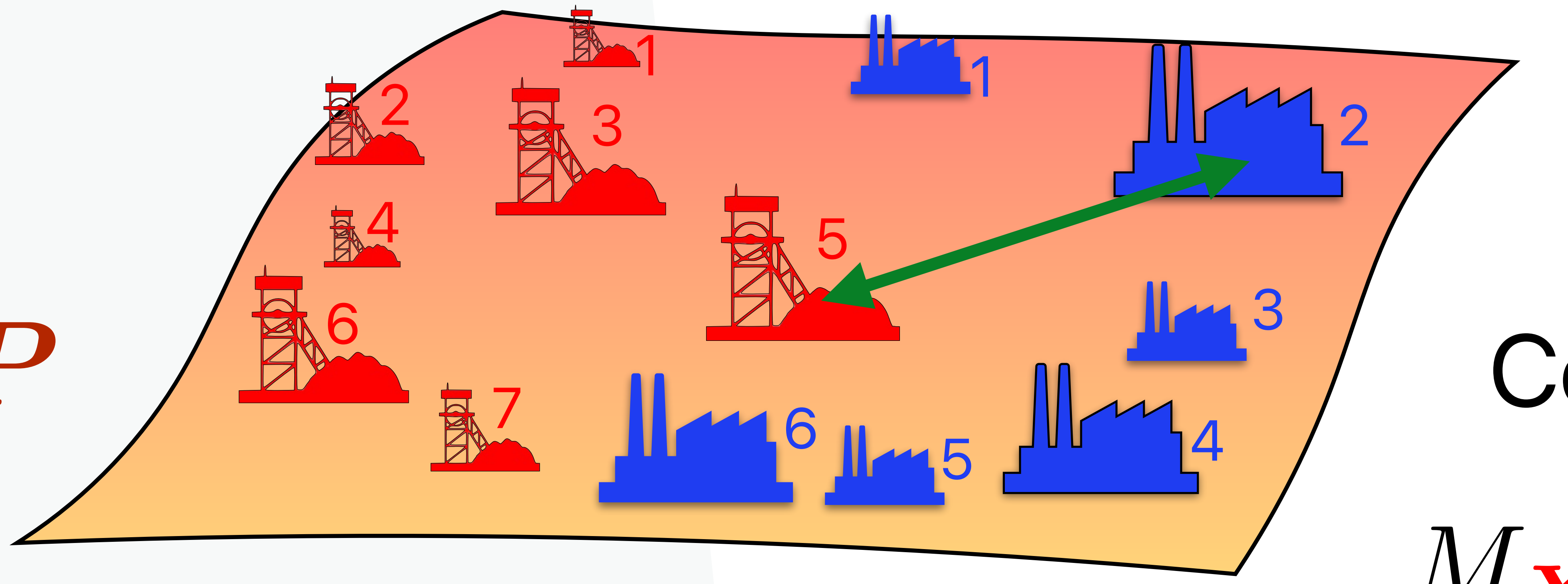
$$U(\boldsymbol{a},\boldsymbol{b}) \stackrel{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}_+^{n\times m} | \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_n = \boldsymbol{b}\}$$

Cost matrix $M_{\boldsymbol{XY}}$

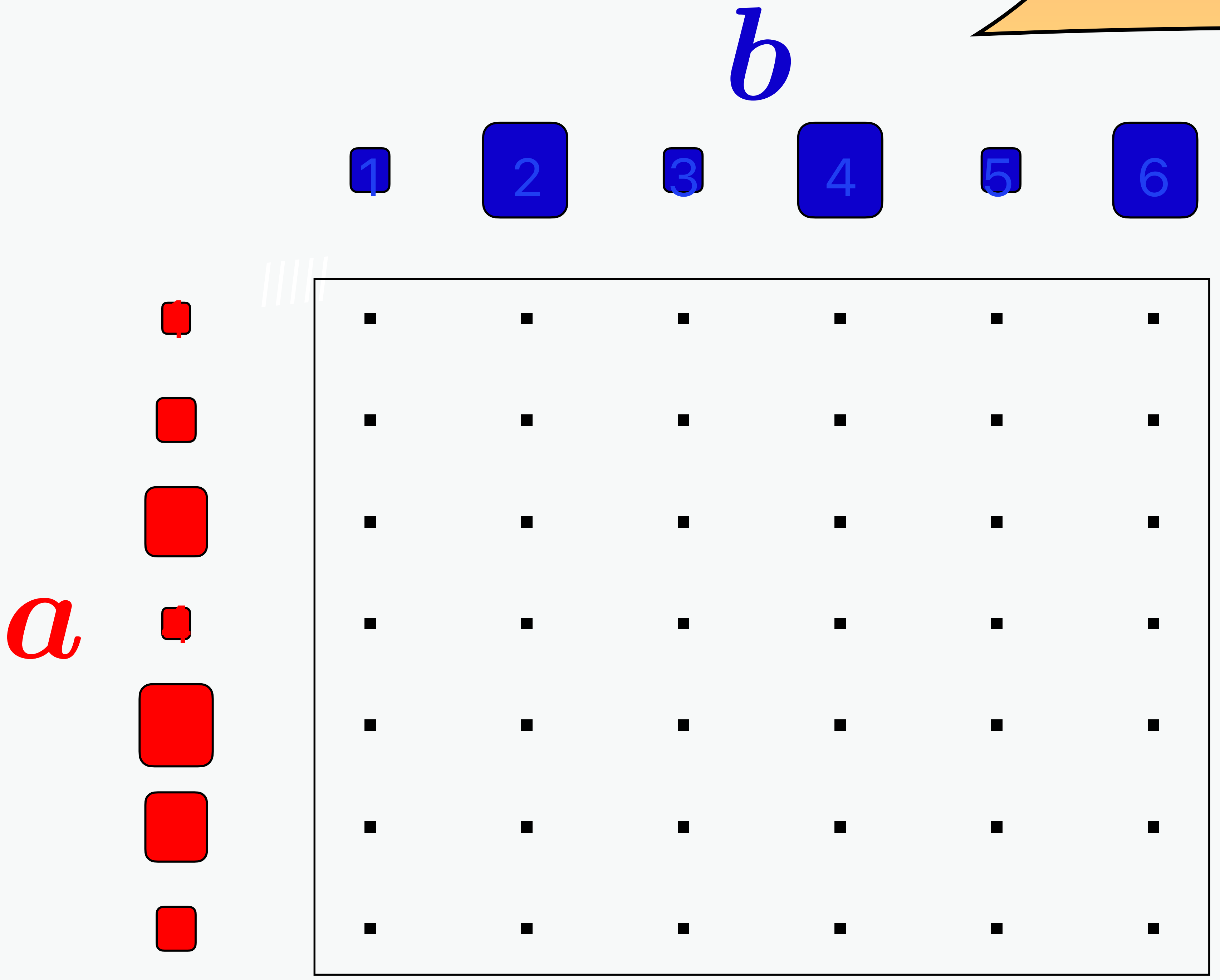
$$M_{\boldsymbol{XY}} \stackrel{\text{def}}{=} [\boldsymbol{c}(\boldsymbol{x}_i, \boldsymbol{y}_j)]_{ij}$$



OT on Two Empirical Measures: Matrices

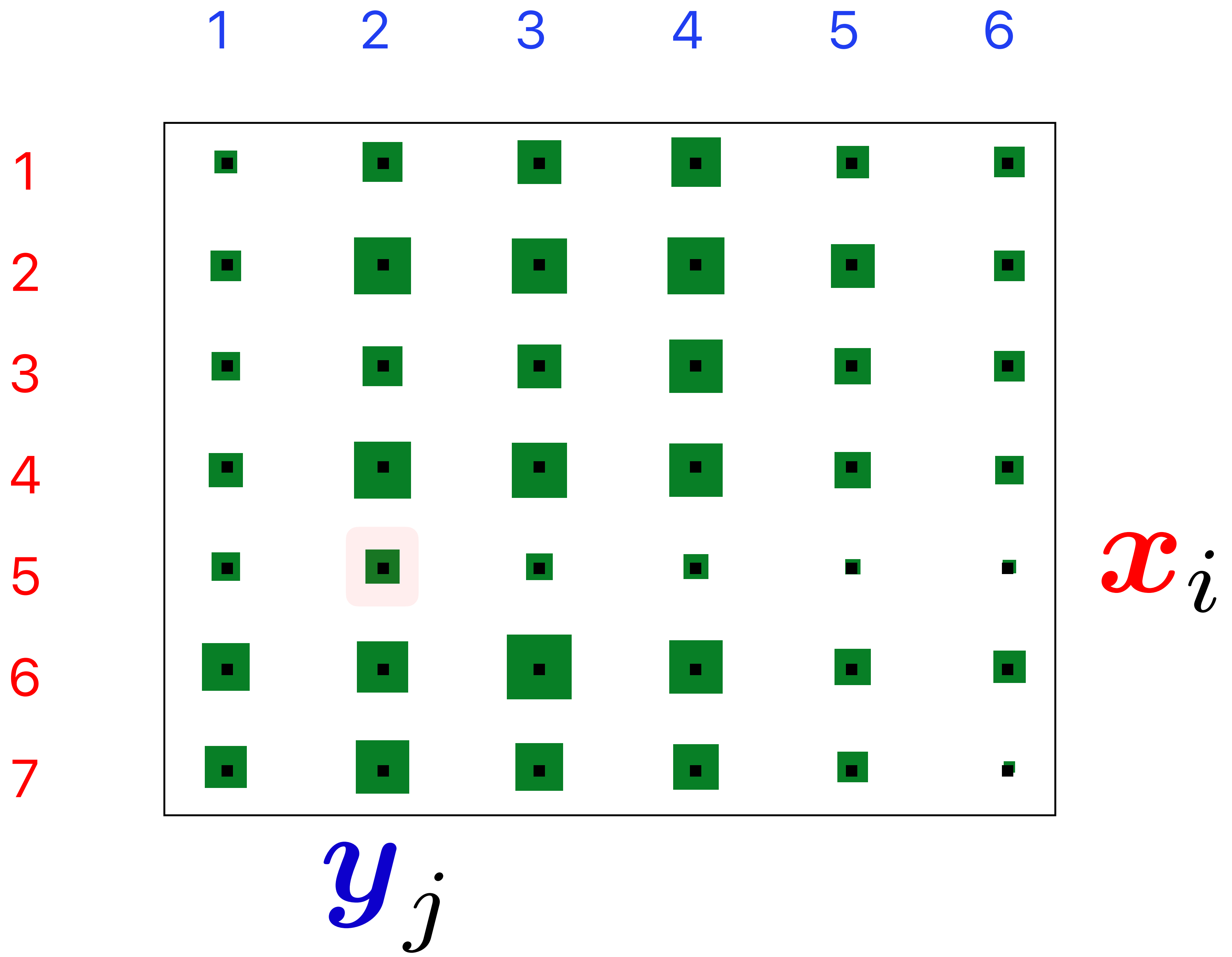


Couplings $\boldsymbol{P}$



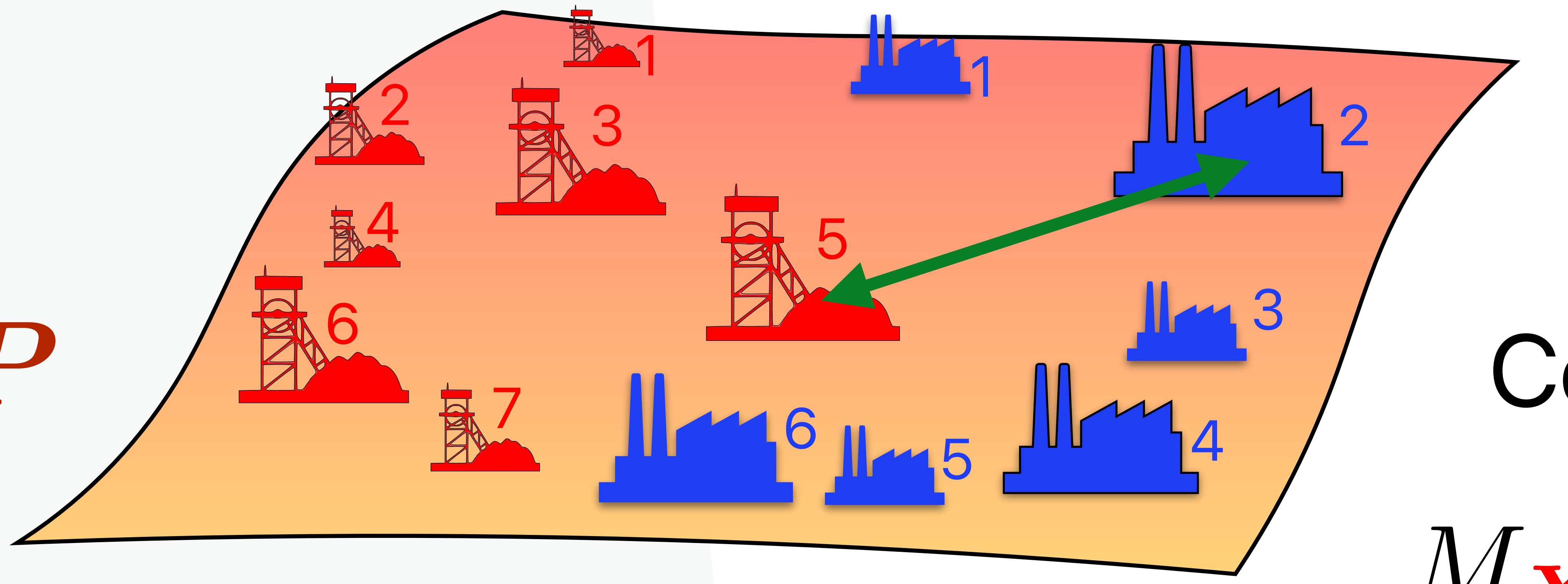
Cost matrix $M_{\boldsymbol{X}\boldsymbol{Y}}$

$$M_{\boldsymbol{X}\boldsymbol{Y}} \overset{\text{def}}{=} [\boldsymbol{c}(\boldsymbol{x}_i, \boldsymbol{y}_j)]_{ij}$$



$$U(\boldsymbol{a}, \boldsymbol{b}) \overset{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} | \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T \mathbf{1}_n = \boldsymbol{b}\}$$

OT on Two Empirical Measures: Matrices

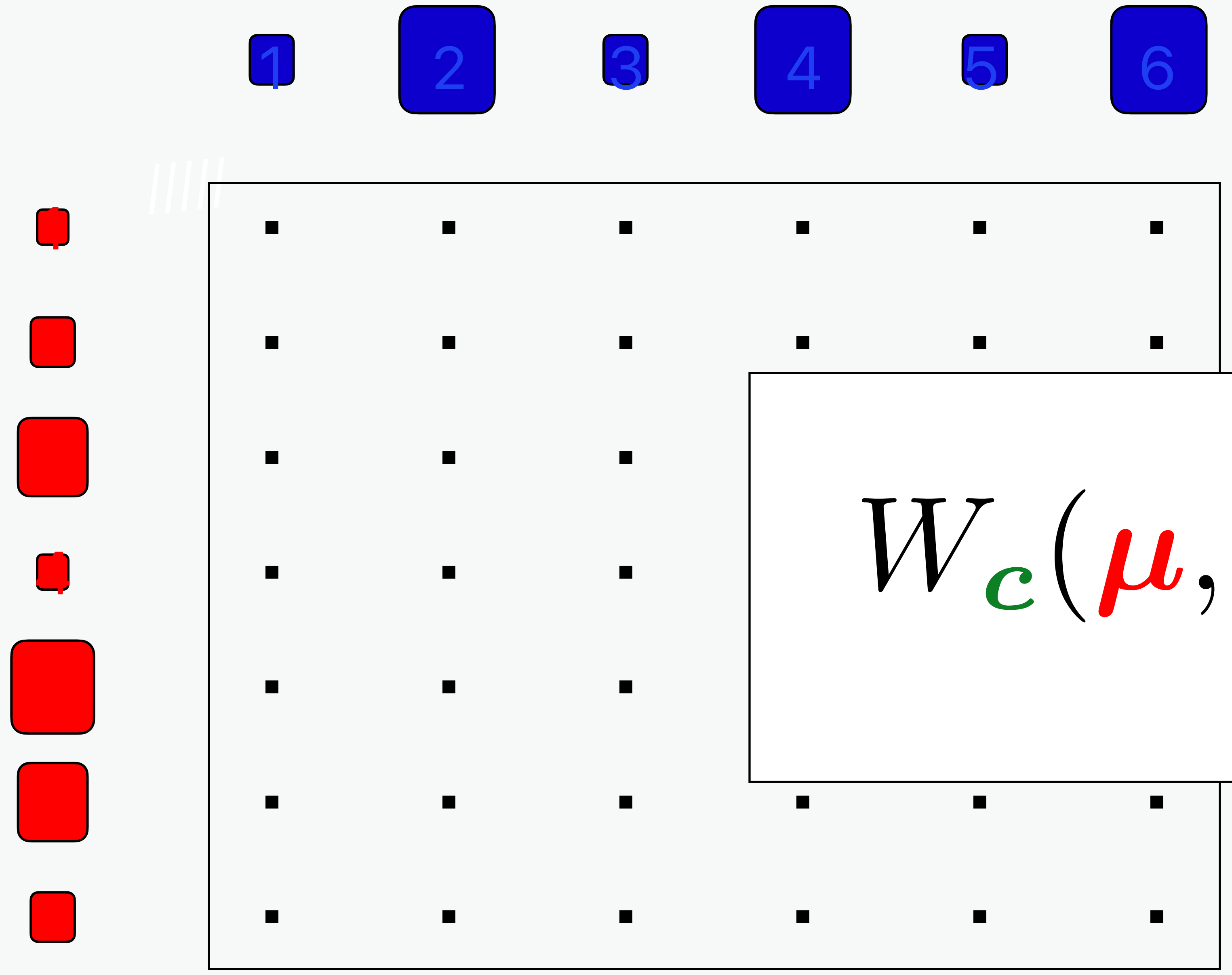


Couplings $\boldsymbol{P}$

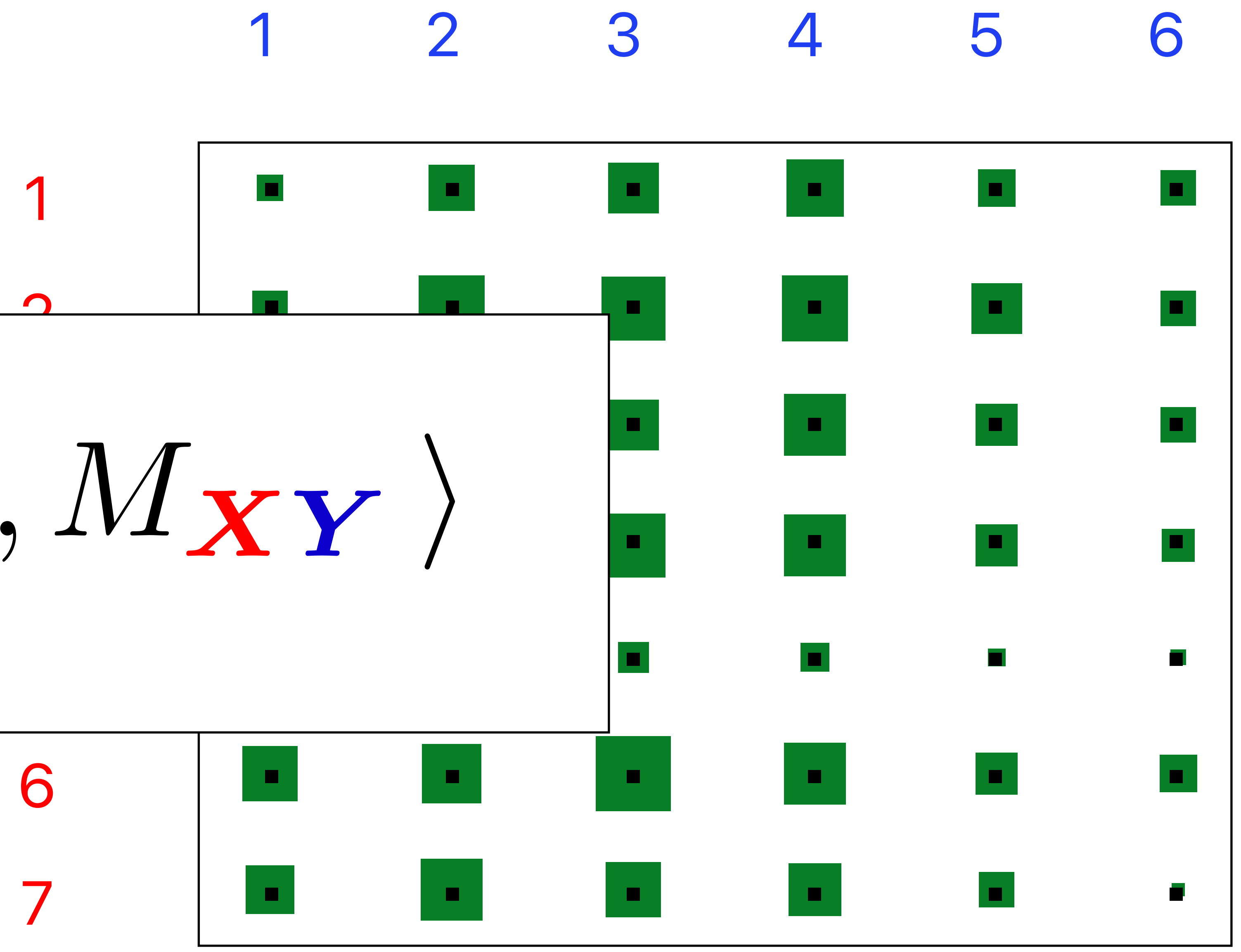
Cost matrix $M_{\mathbf{XY}}$

$$M_{\mathbf{XY}} \stackrel{\text{def}}{=} [\boldsymbol{c}(\boldsymbol{x}_i, \boldsymbol{y}_j)]_{ij}$$

$\boldsymbol{b}$



$\boldsymbol{a}$



$\boldsymbol{x}_i$

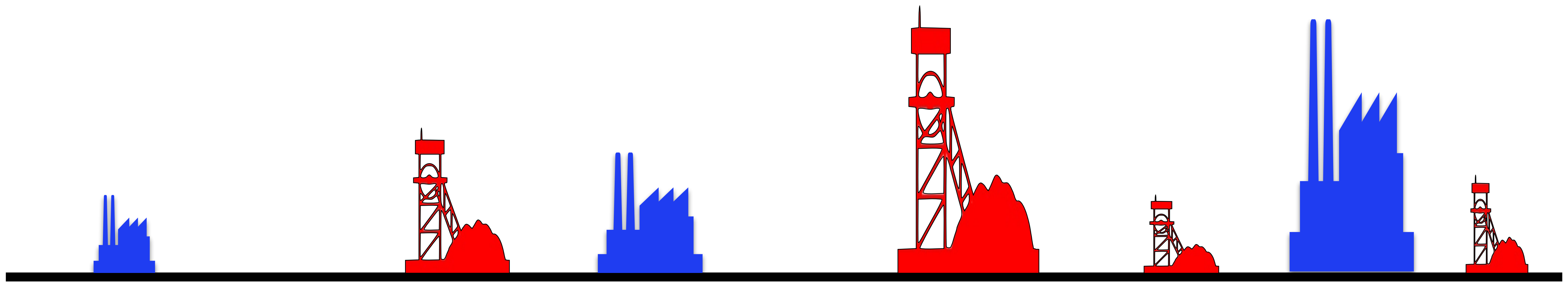
$\boldsymbol{y}_j$

$$W_{\boldsymbol{c}}(\boldsymbol{\mu}, \boldsymbol{\nu}) = \min_{\boldsymbol{P} \in U(\boldsymbol{a}, \boldsymbol{b})} \langle \boldsymbol{P}, M_{\mathbf{XY}} \rangle$$

$$U(\boldsymbol{a}, \boldsymbol{b}) \stackrel{\text{def}}{=} \{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} | \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_n = \boldsymbol{b}\}$$

Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

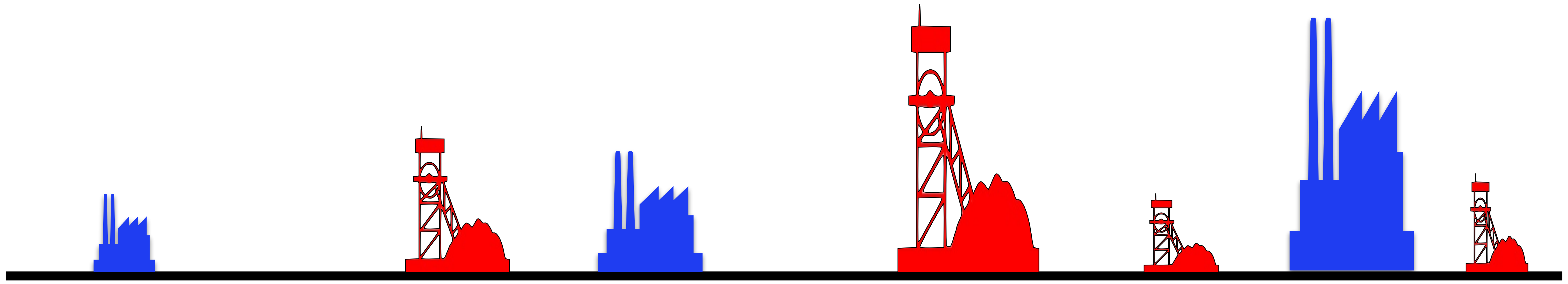


Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
“Repeat until termination: the leftmost mine transfers as much as it can to the leftmost factory still in need of resource.”

Computation: $O(n \log n + m \log m)$

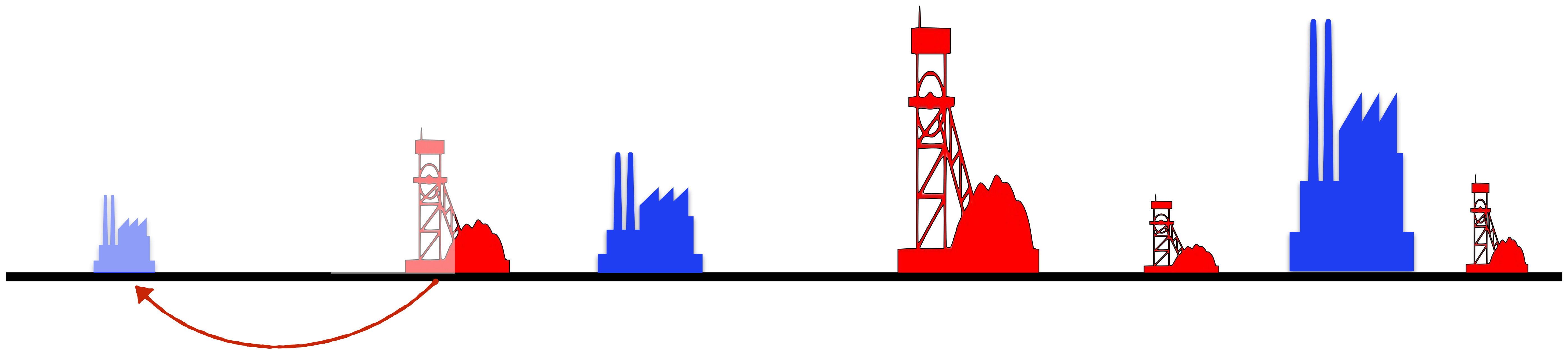


Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
"Repeat until termination: the leftmost mine transfers
as much as it can to the leftmost factory still in need of resource."

Computation: $O(n \log n + m \log m)$

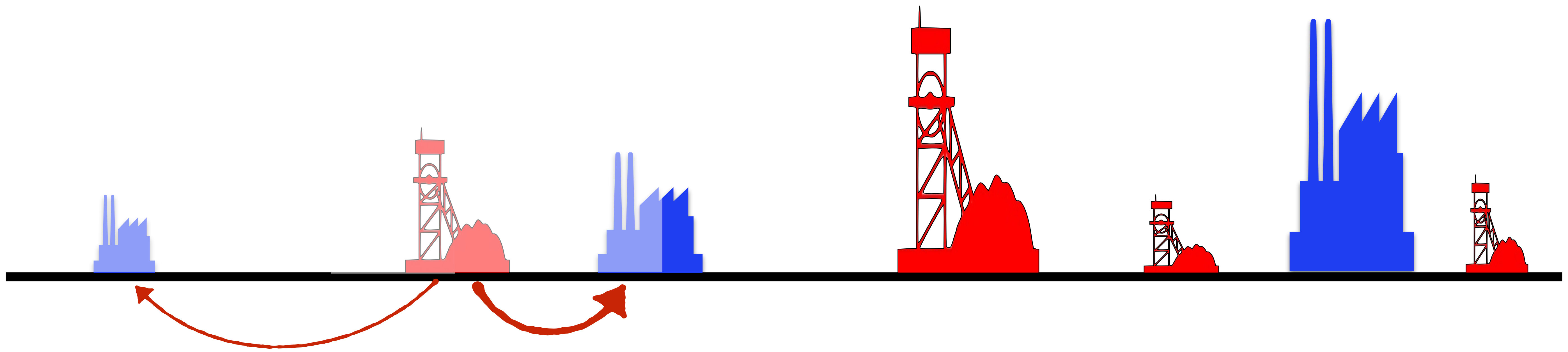


Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
"Repeat until termination: the leftmost mine transfers
as much as it can to the leftmost factory still in need of resource."

Computation: $O(n \log n + m \log m)$

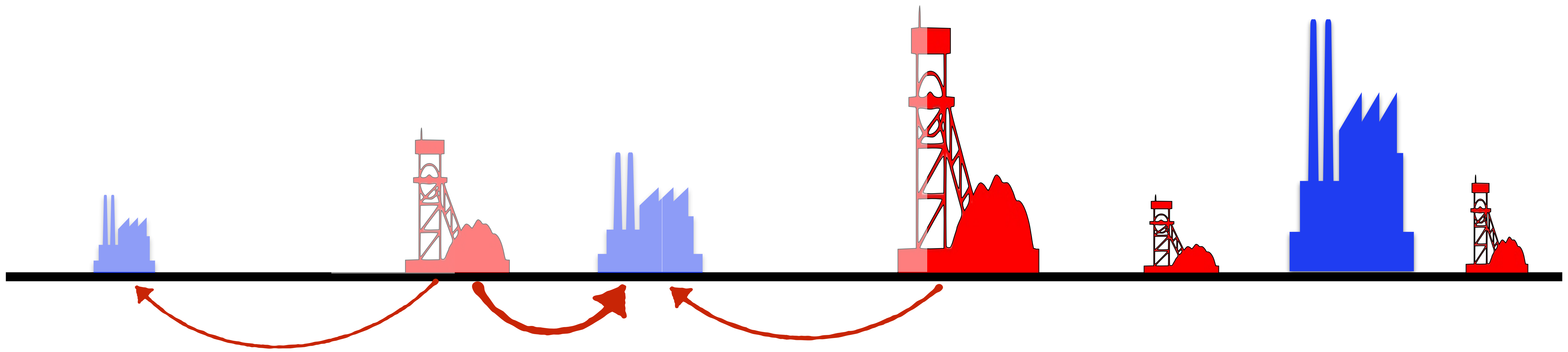


Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
"Repeat until termination: the leftmost mine transfers
as much as it can to the leftmost factory still in need of resource."

Computation: $O(n \log n + m \log m)$

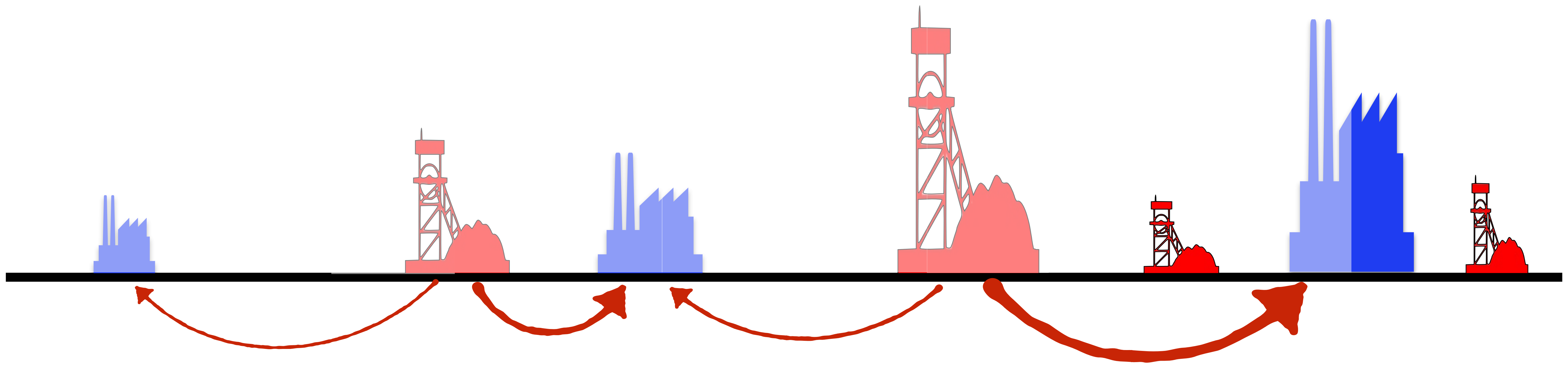


Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
"Repeat until termination: the leftmost mine transfers
as much as it can to the leftmost factory still in need of resource."

Computation: $O(n \log n + m \log m)$

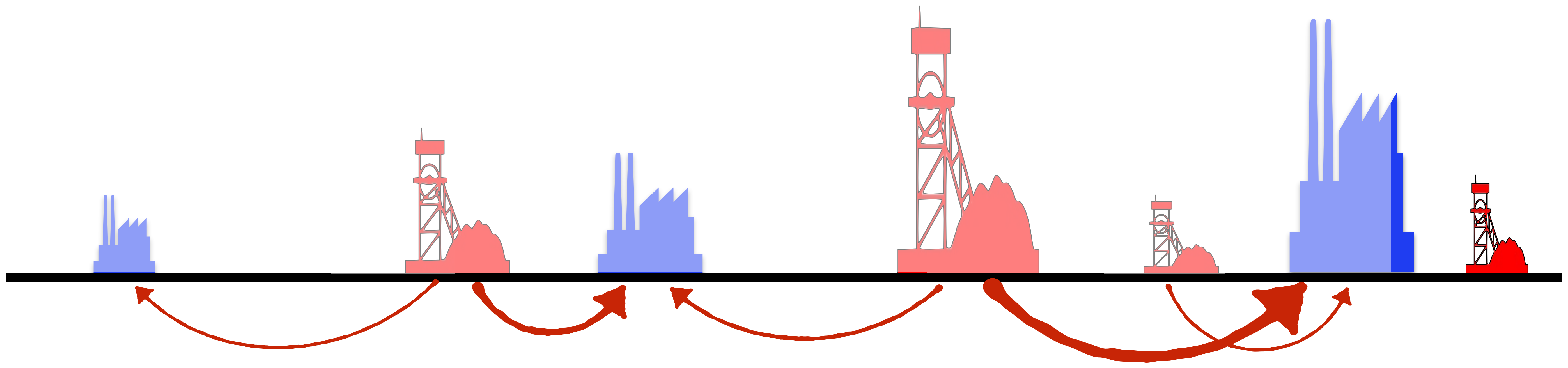


Easy Case: 1D Point Clouds

Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
"Repeat until termination: the leftmost mine transfers as much as it can to the leftmost factory still in need of resource."

Computation: $O(n \log n + m \log m)$

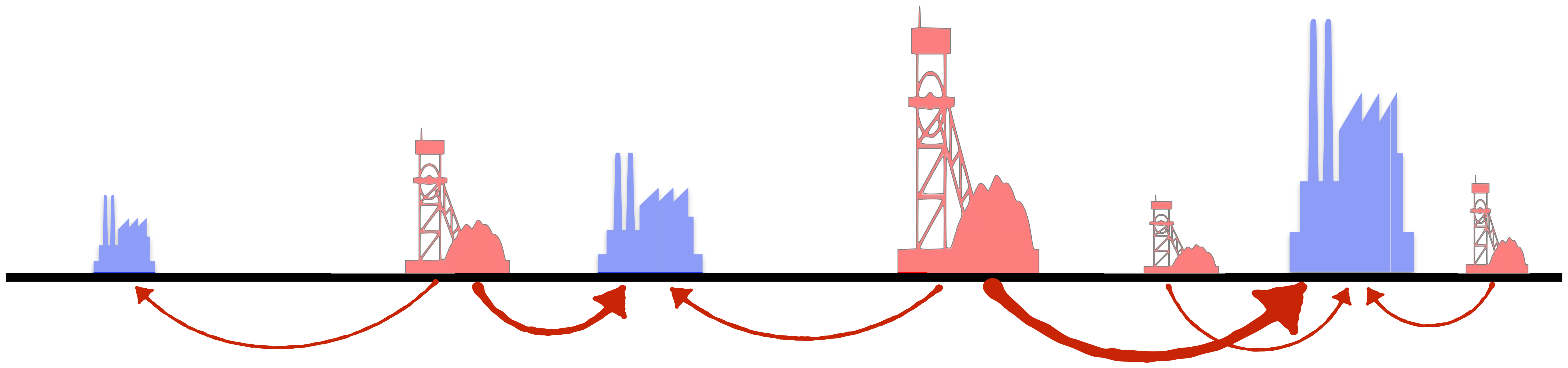


Easy Case: 1D Point Clouds

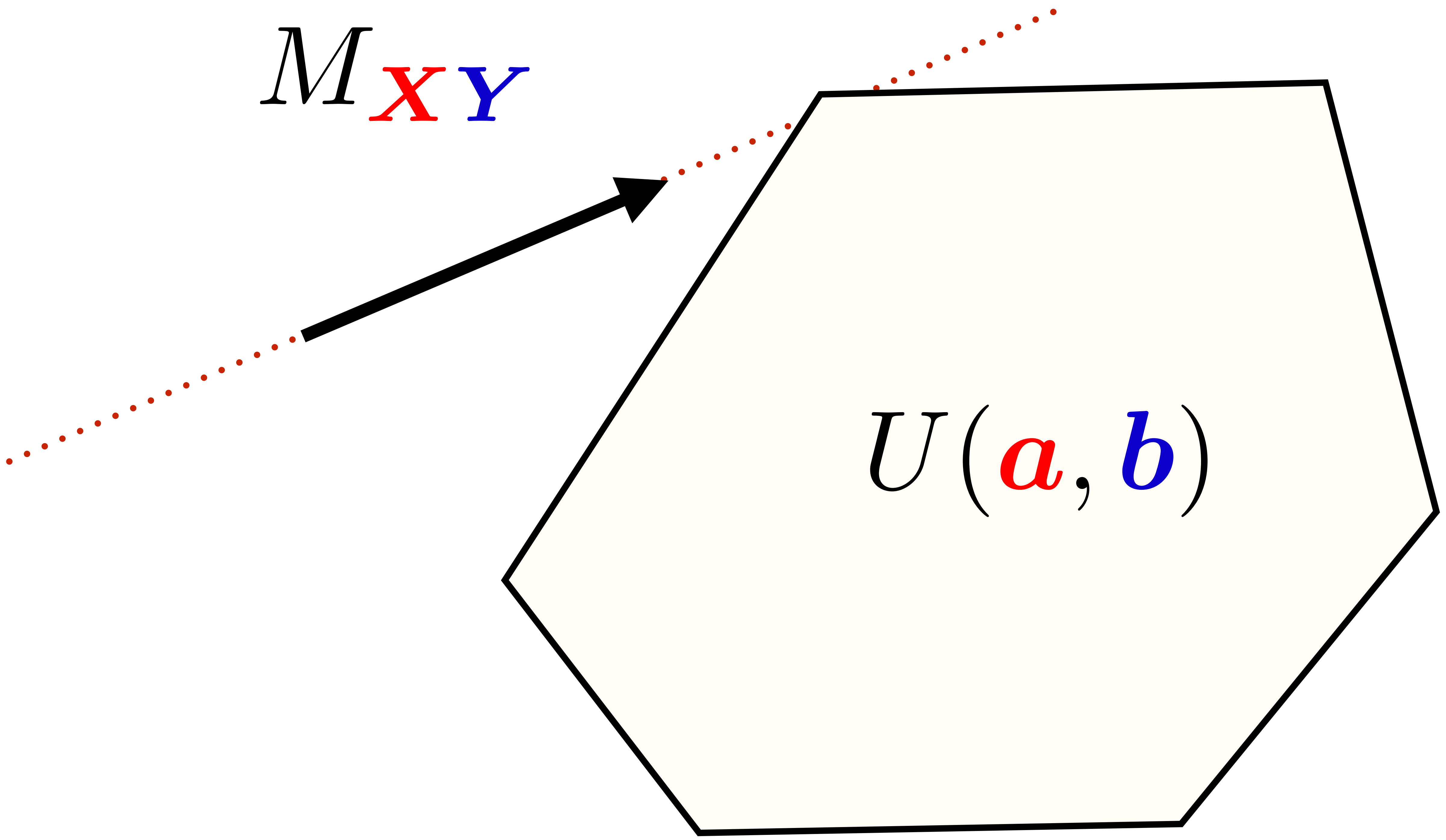
Suppose the mines and factories live in a 1D world.

Optimal Transport is **solved in 1D** using **sorting:**
"Repeat until termination: the leftmost mine transfers
as much as it can to the leftmost factory still in need of resource."

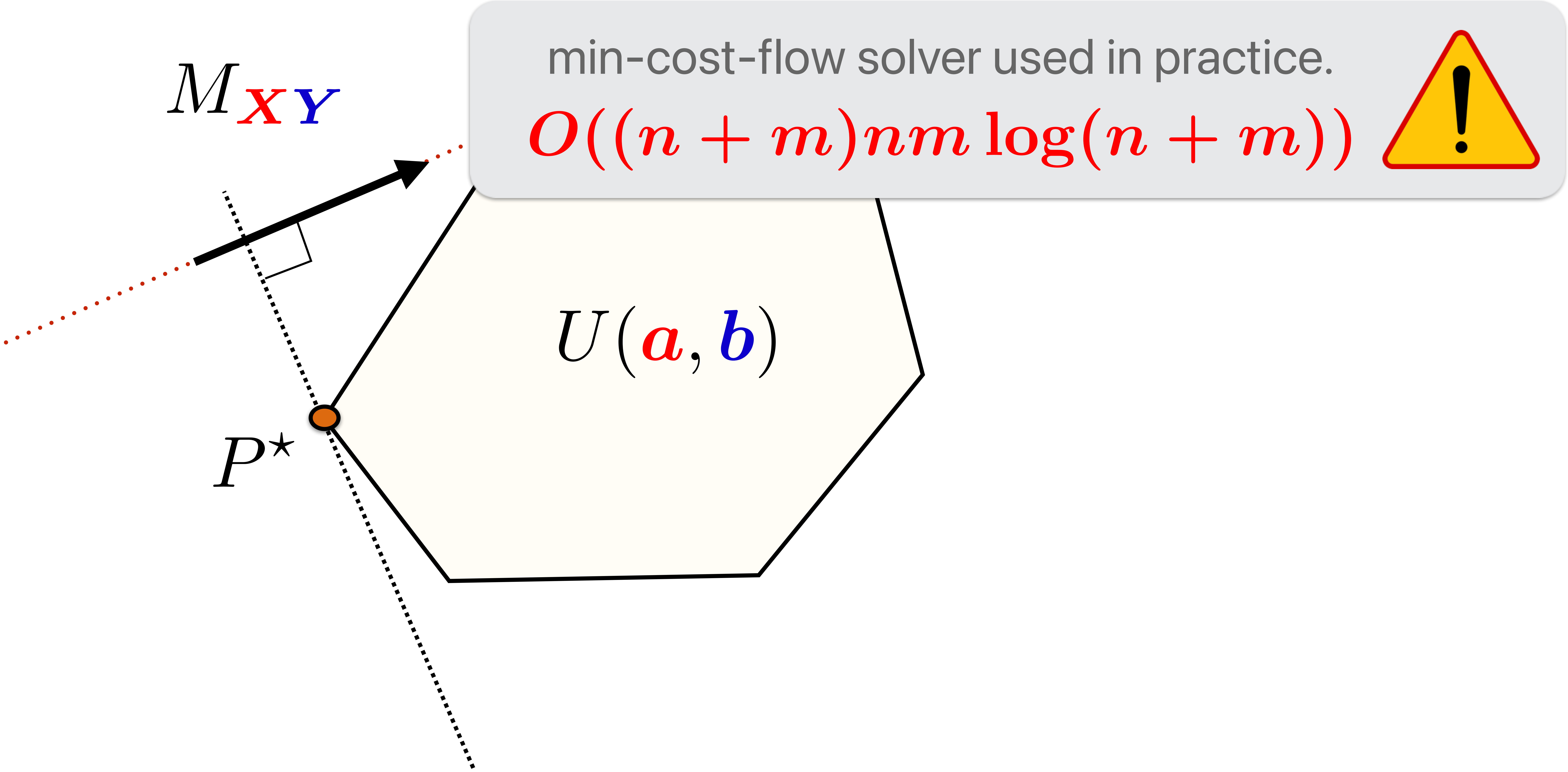
Computation: $O(n \log n + m \log m)$



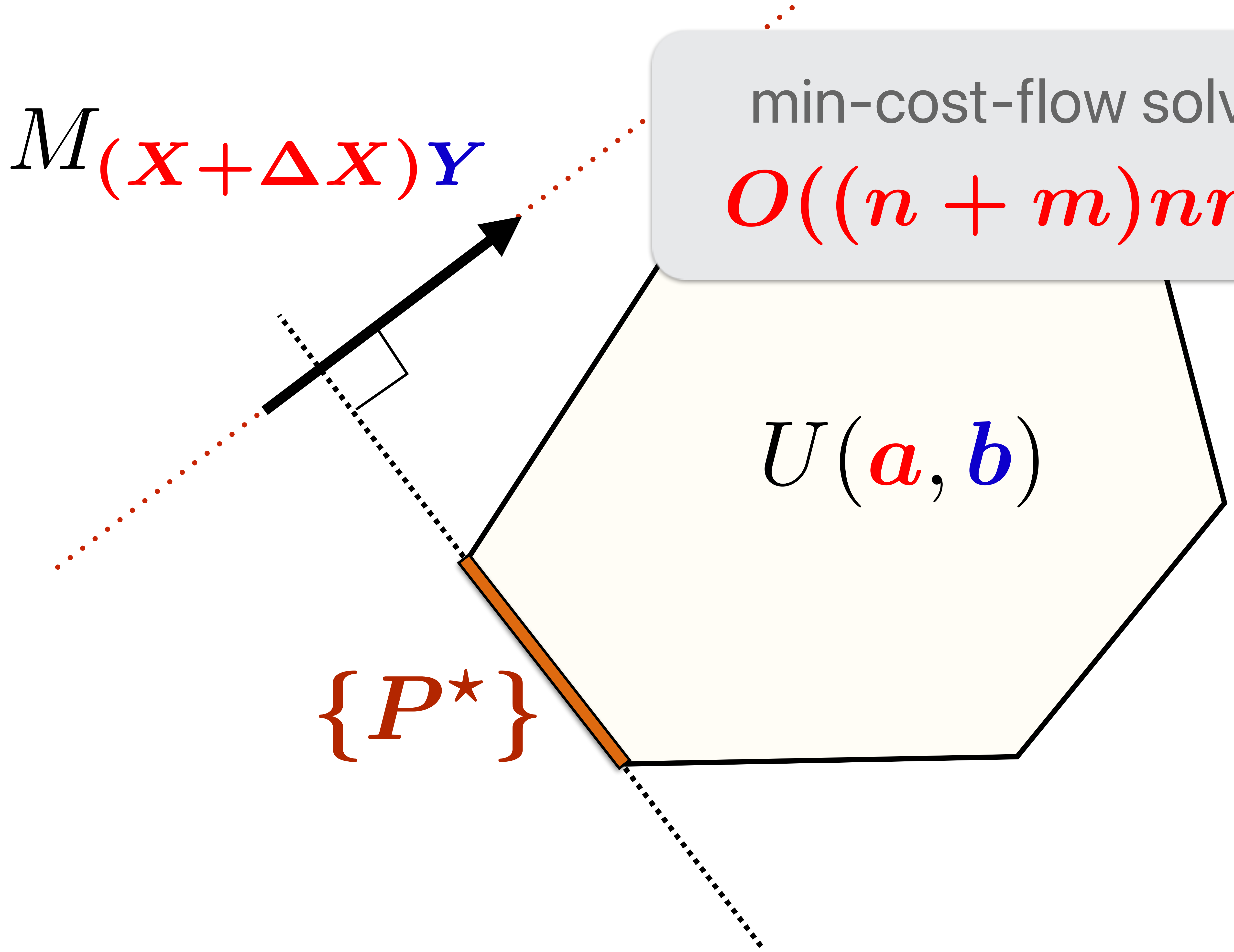
Solving the Discrete OT Problem in Full Generality



Solving the Discrete OT Problem in Full Generality



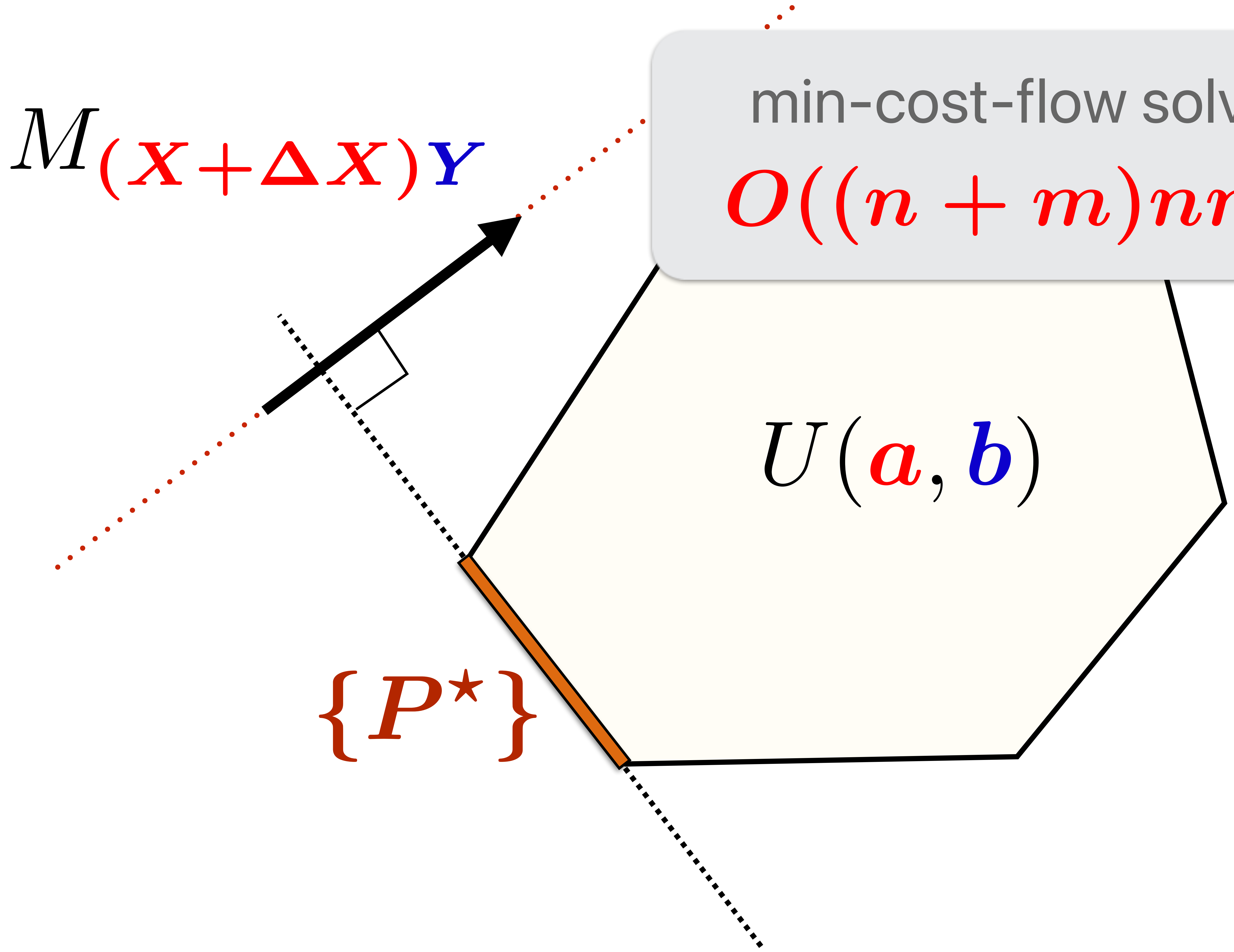
Solving the Discrete OT Problem in Full Generality



min-cost-flow solver used in practice.

$$O((n+m)nm\log(n+m))$$

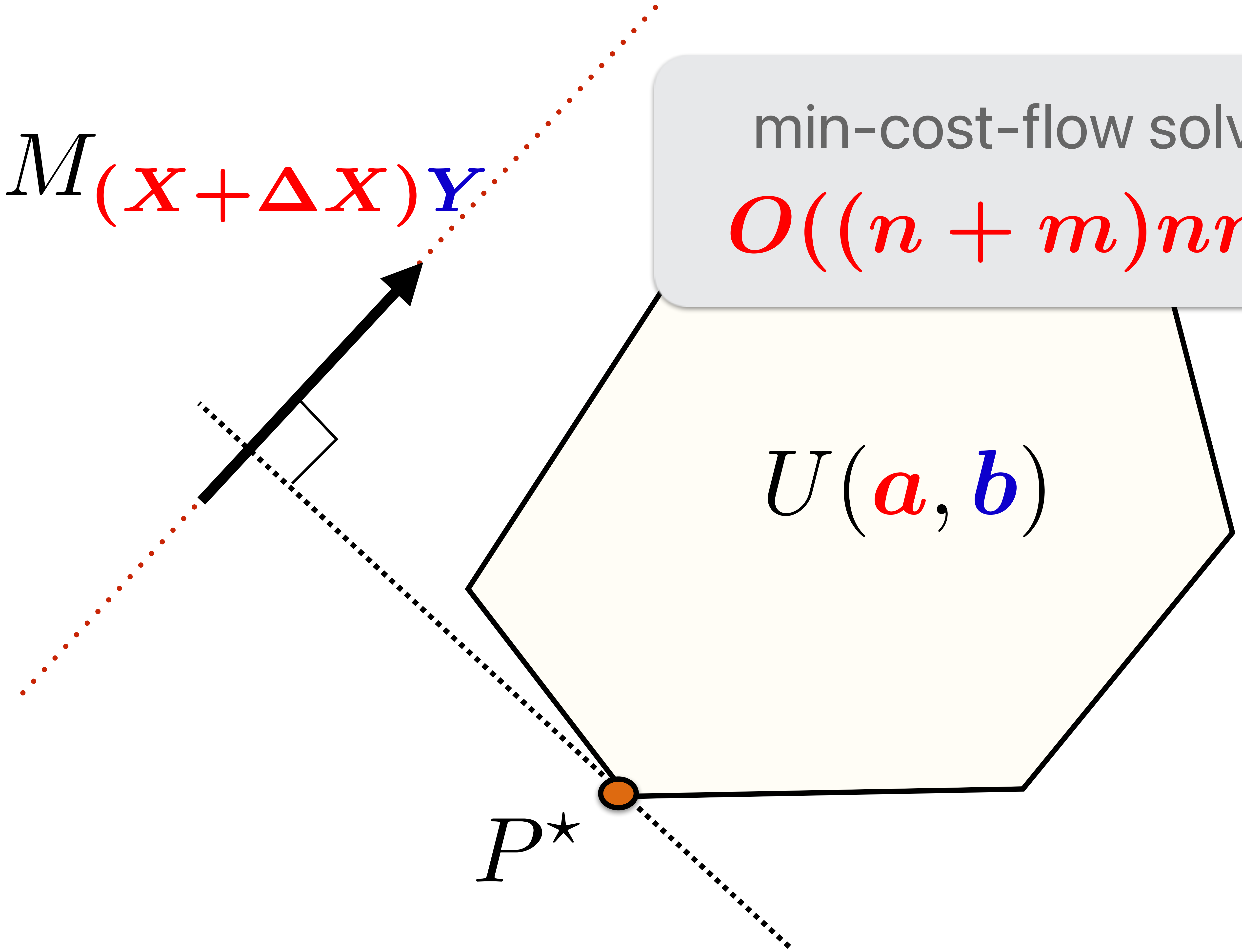
Solving the Discrete OT Problem in Full Generality



min-cost-flow solver used in practice.

$O((n+m)nm\log(n+m))$

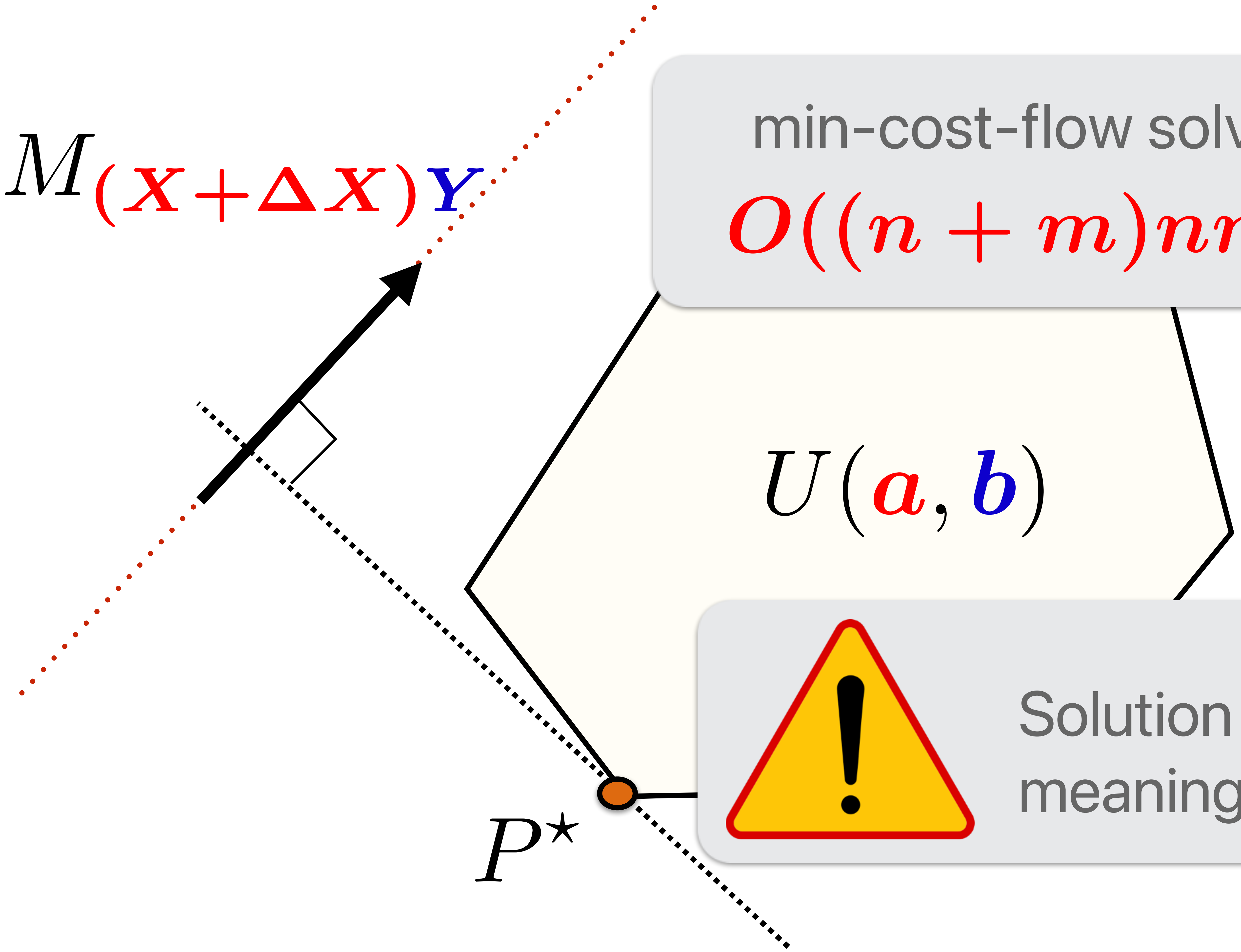
Solving the Discrete OT Problem in Full Generality



min-cost-flow solver used in practice.

$O((n+m)nm\log(n+m))$

Solving the Discrete OT Problem in Full Generality

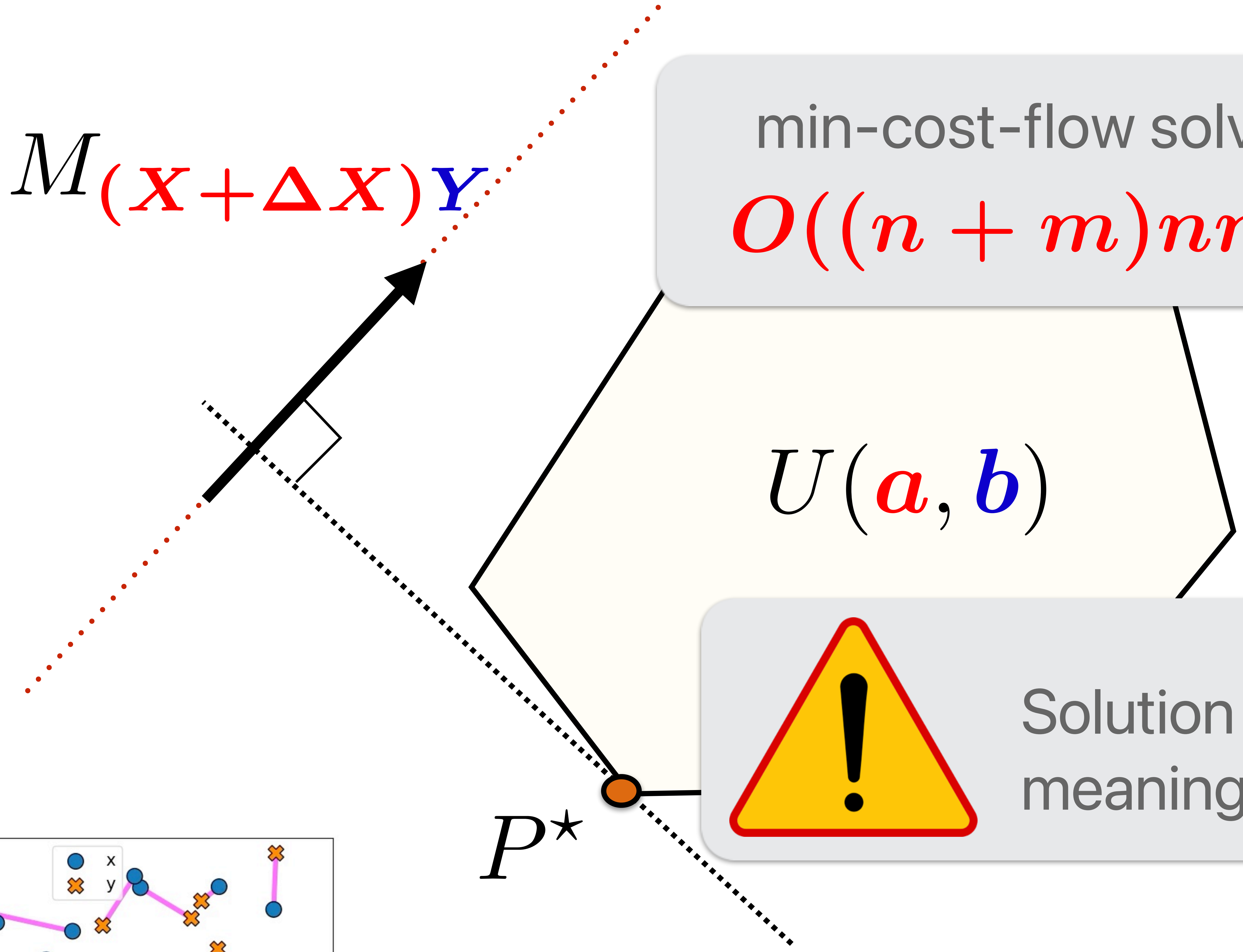


min-cost-flow solver used in practice.

$$O((n+m)nm\log(n+m))$$

Solution $P^\star$ is not unique, and has no meaningful Jacobian w.r.t input X or Y.

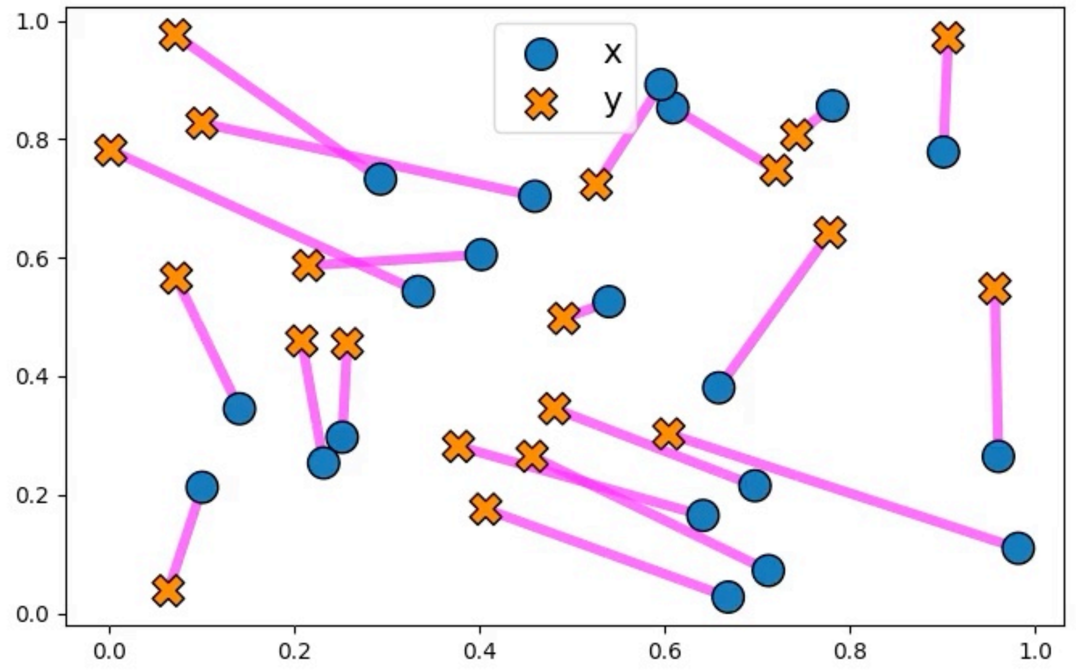
Solving the Discrete OT Problem in Full Generality



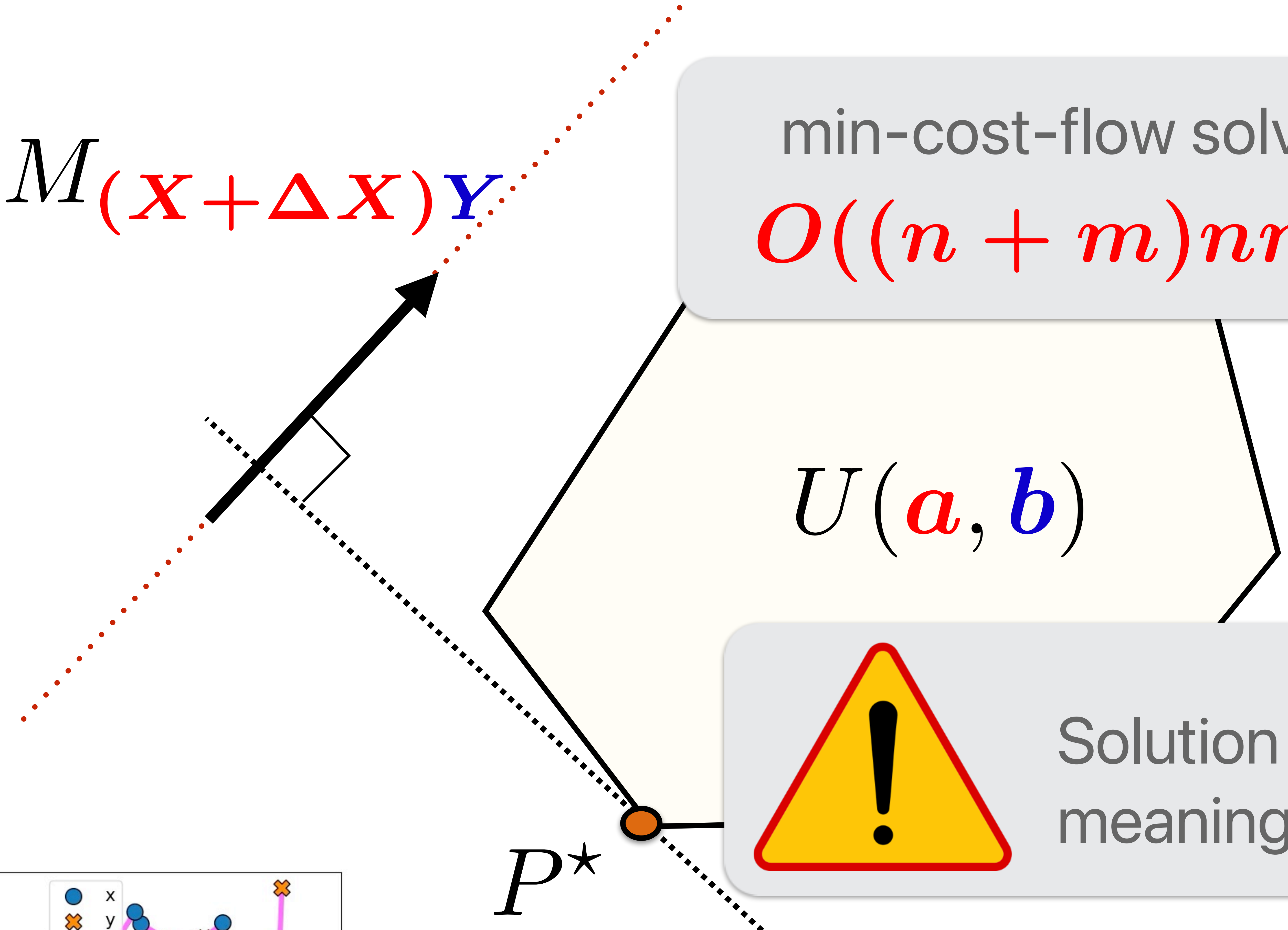
min-cost-flow solver used in practice.

$$O((n+m)nm\log(n+m))$$

Solution $P^\star$ is not unique, and has no meaningful Jacobian w.r.t input X or Y.



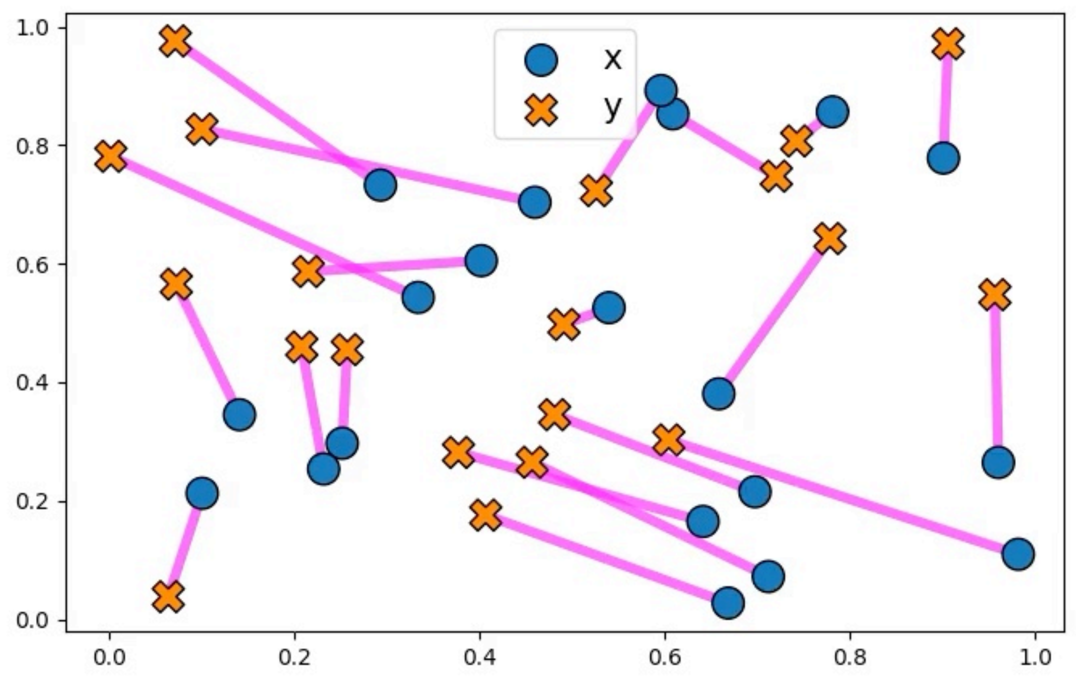
Solving the Discrete OT Problem in Full Generality



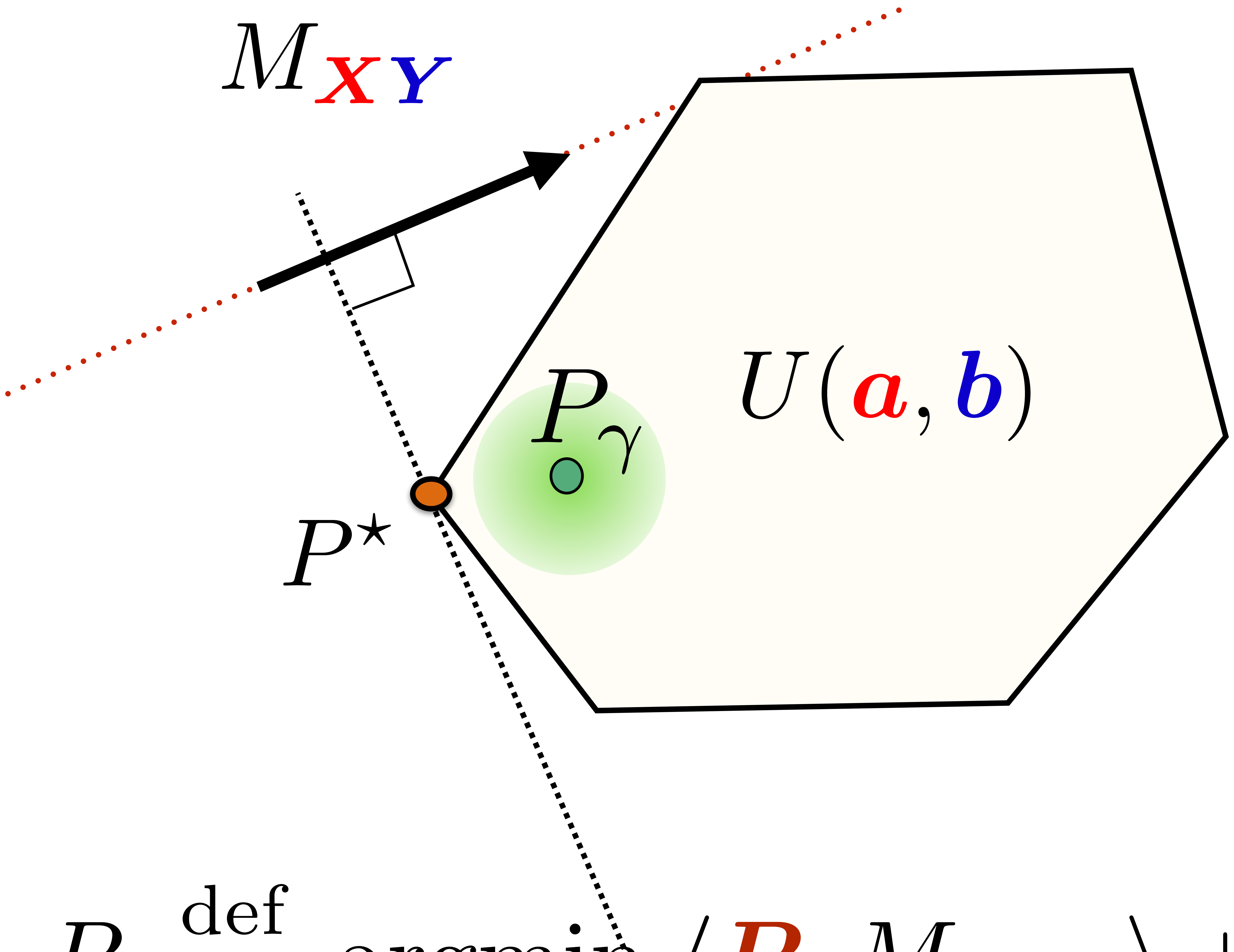
min-cost-flow solver used in practice.
$O((n+m)nm\log(n+m))$

Solution $P^\star$ is not unique, and has no meaningful Jacobian w.r.t input X or Y.

$W_c(\boldsymbol{\mu}, \boldsymbol{\nu})$ not differentiable



Solution: Regularize Linear Program

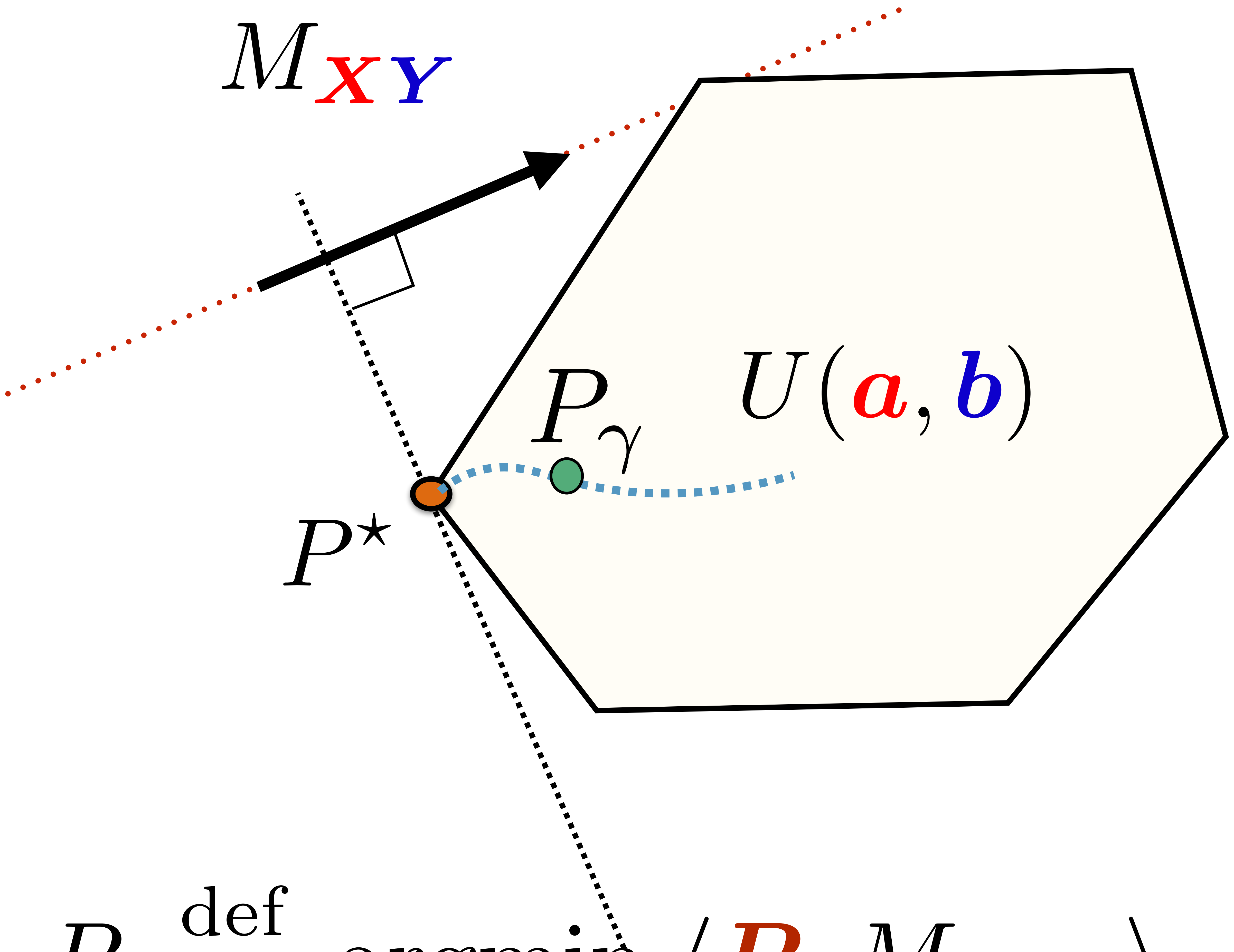


Why regularize?

- Computational efficiency
- Statistical regularity
- Robustness
- Differentiability
- Structure

$$P_\gamma \stackrel{\text{def}}{=} \underset{P \in U(a,b)}{\operatorname{argmin}} \langle P, M_{XY} \rangle + \gamma \text{Reg}(P)$$

Solution: Regularize Linear Program



Why regularize?

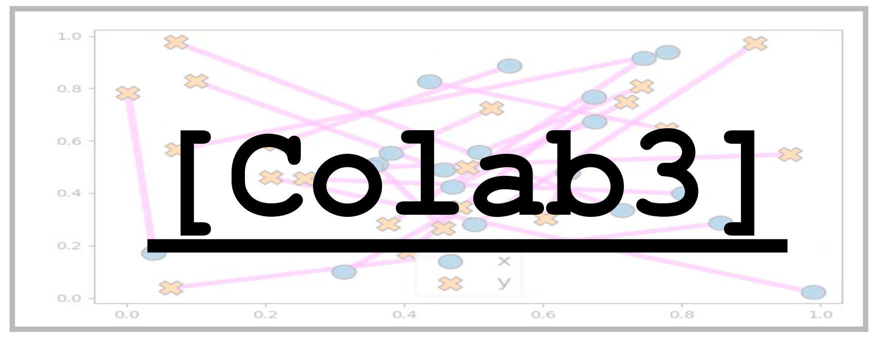
- Computational efficiency
- Statistical regularity
- Robustness
- Differentiability
- Structure

[Dessein+16] [Blondel+18]

$$P_\gamma \stackrel{\text{def}}{=} \underset{\boldsymbol{P}\in U(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle - \gamma E(\boldsymbol{P})$$

Solving Entropy Regularized OT

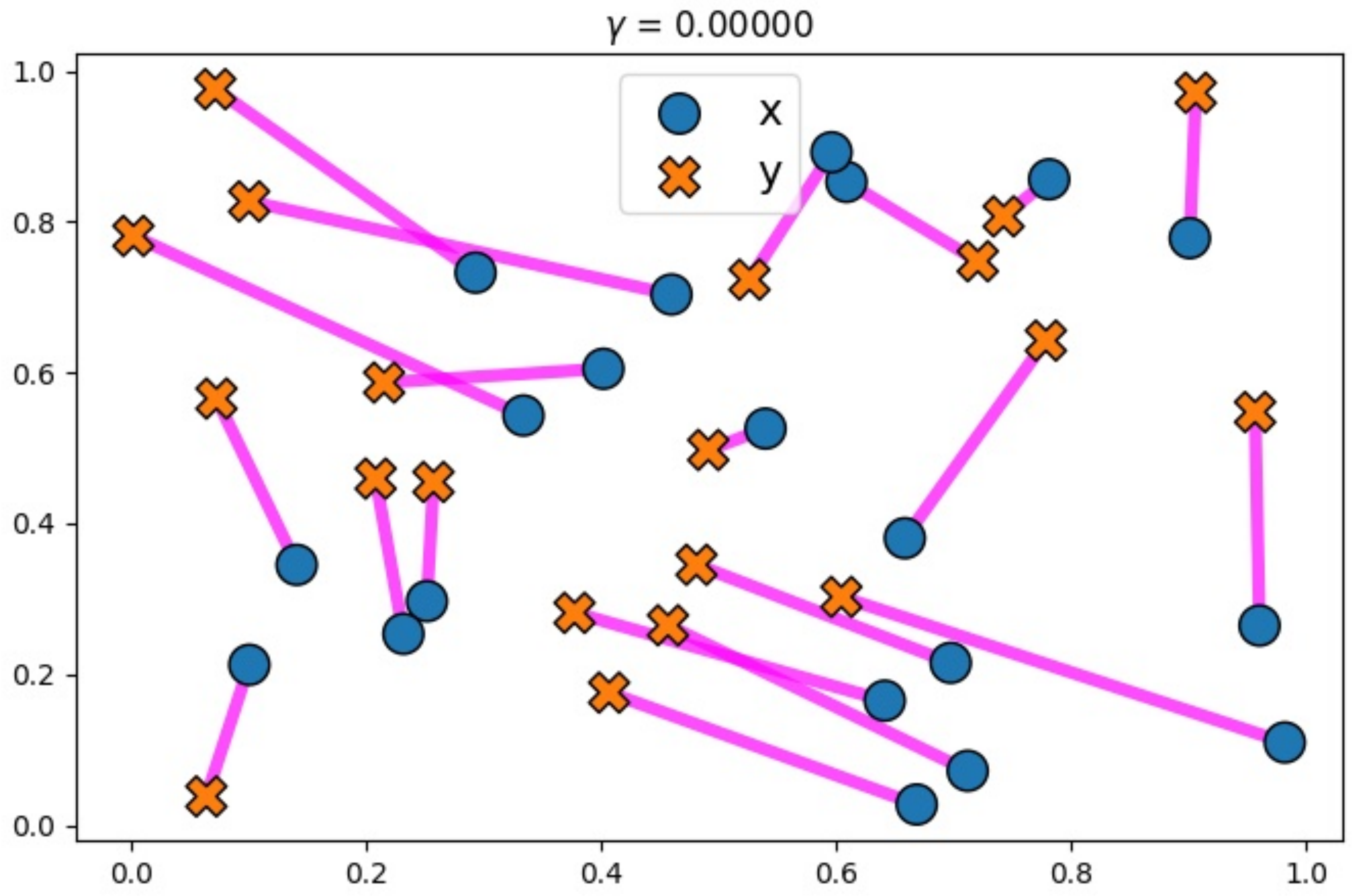
$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle - \gamma E(\boldsymbol{P})$$

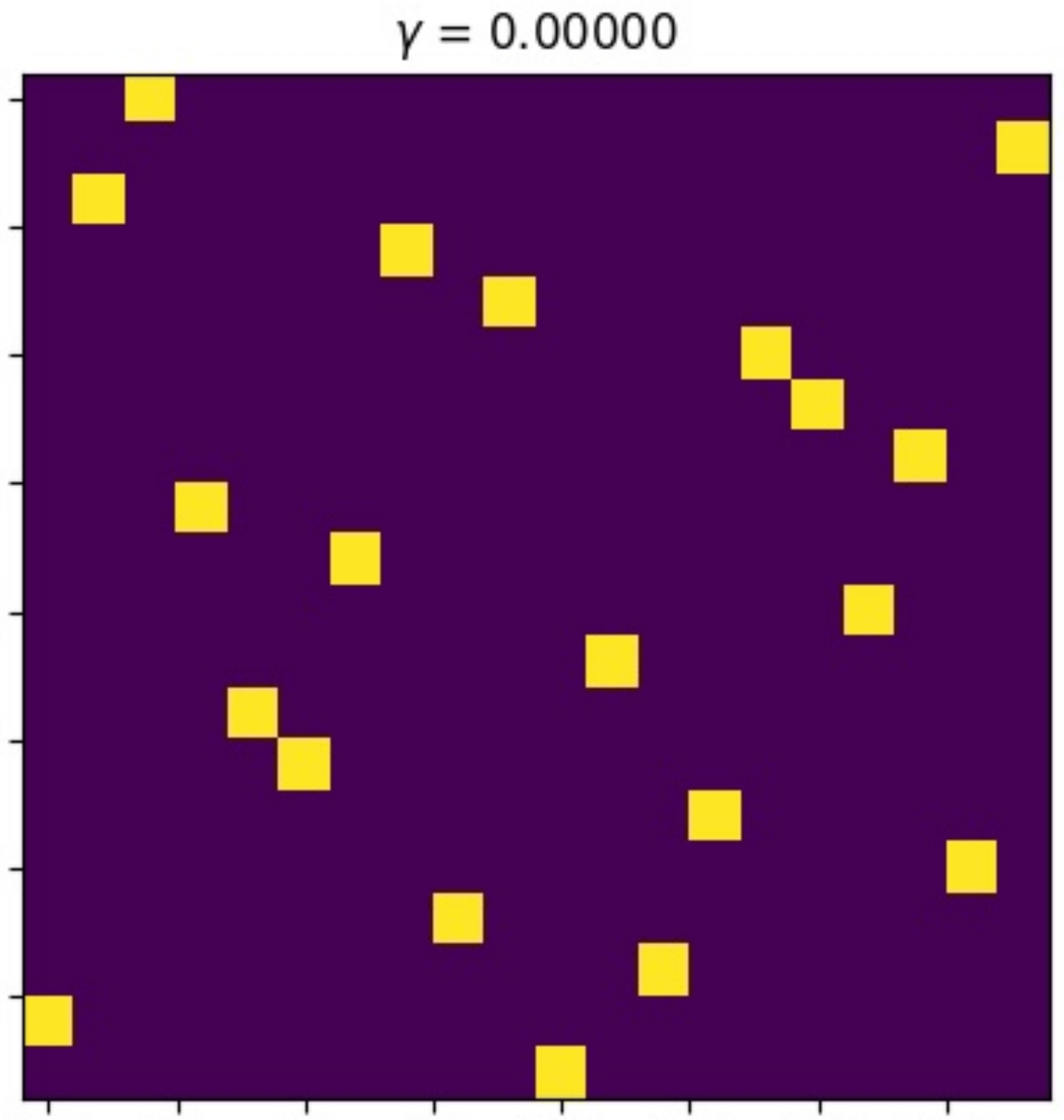


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle - \gamma E(\boldsymbol{P})$$

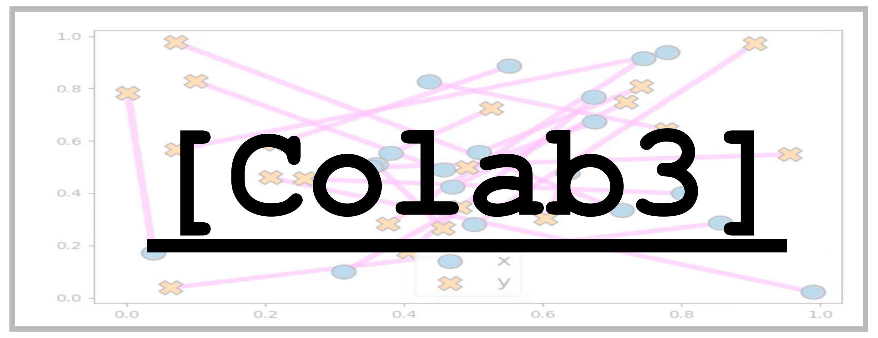
[Colab3]

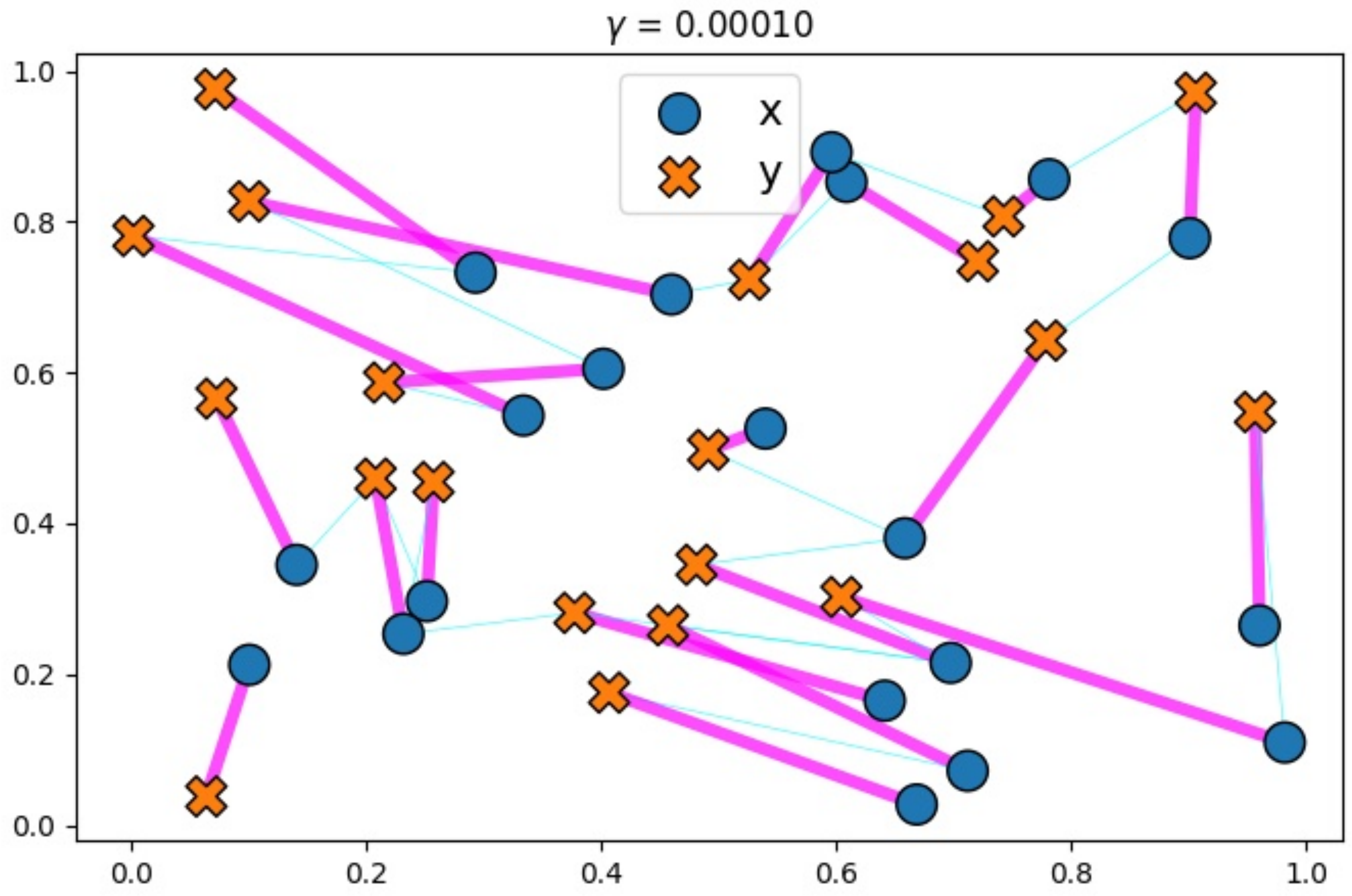


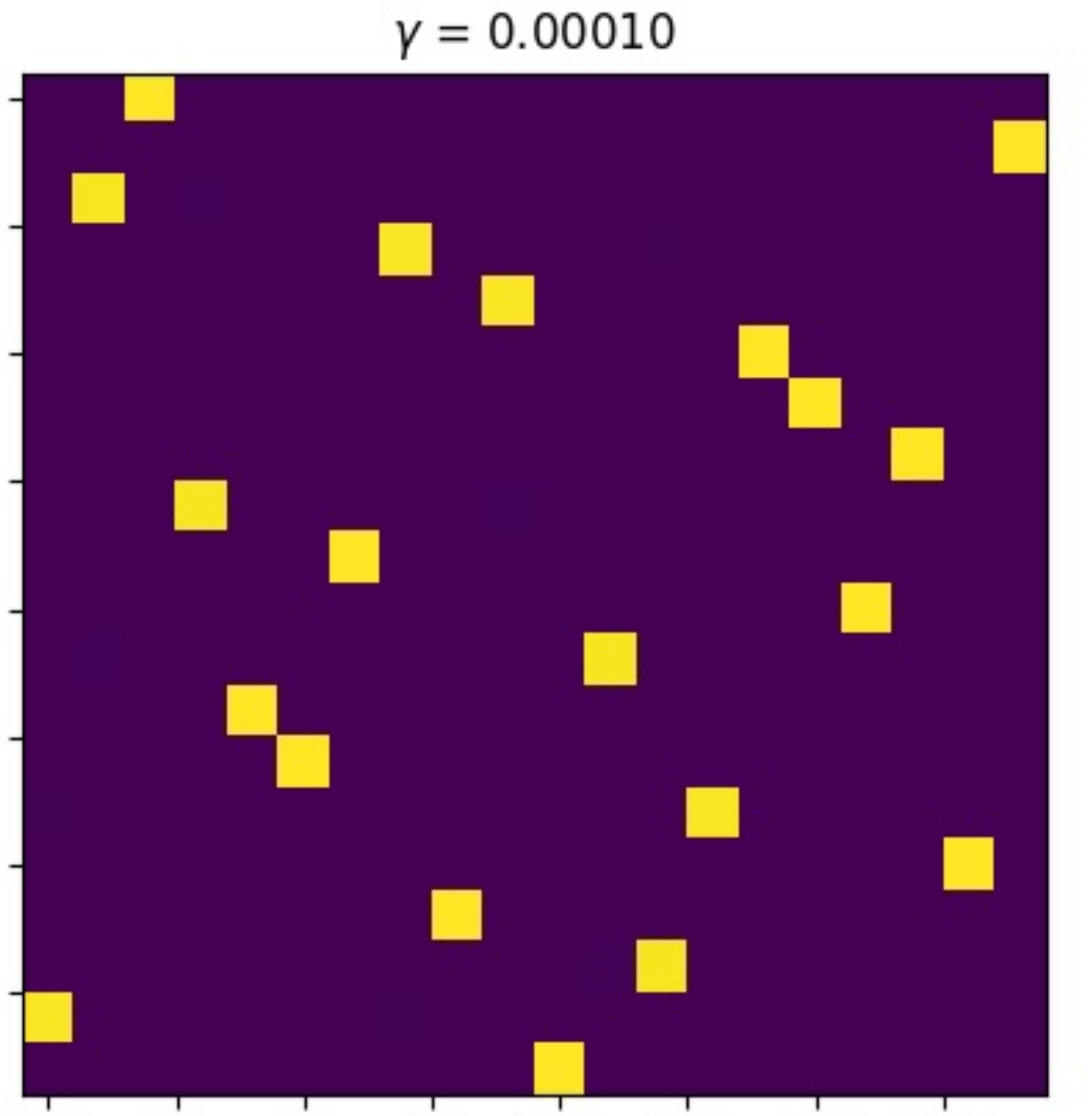


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle - \gamma E(\boldsymbol{P})$$

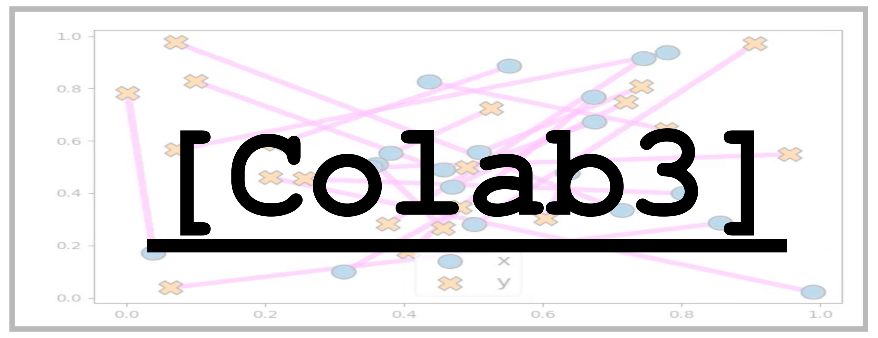


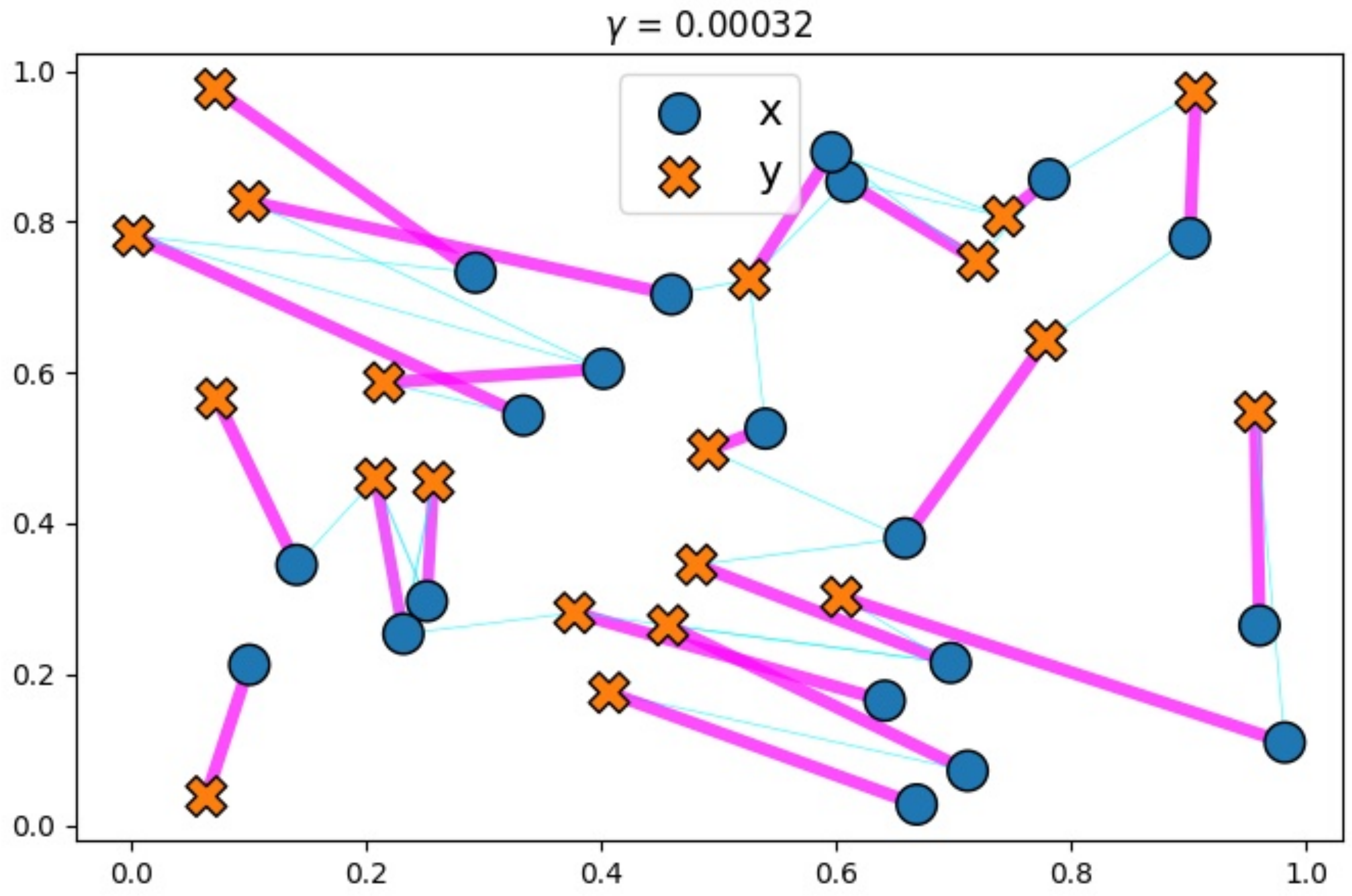


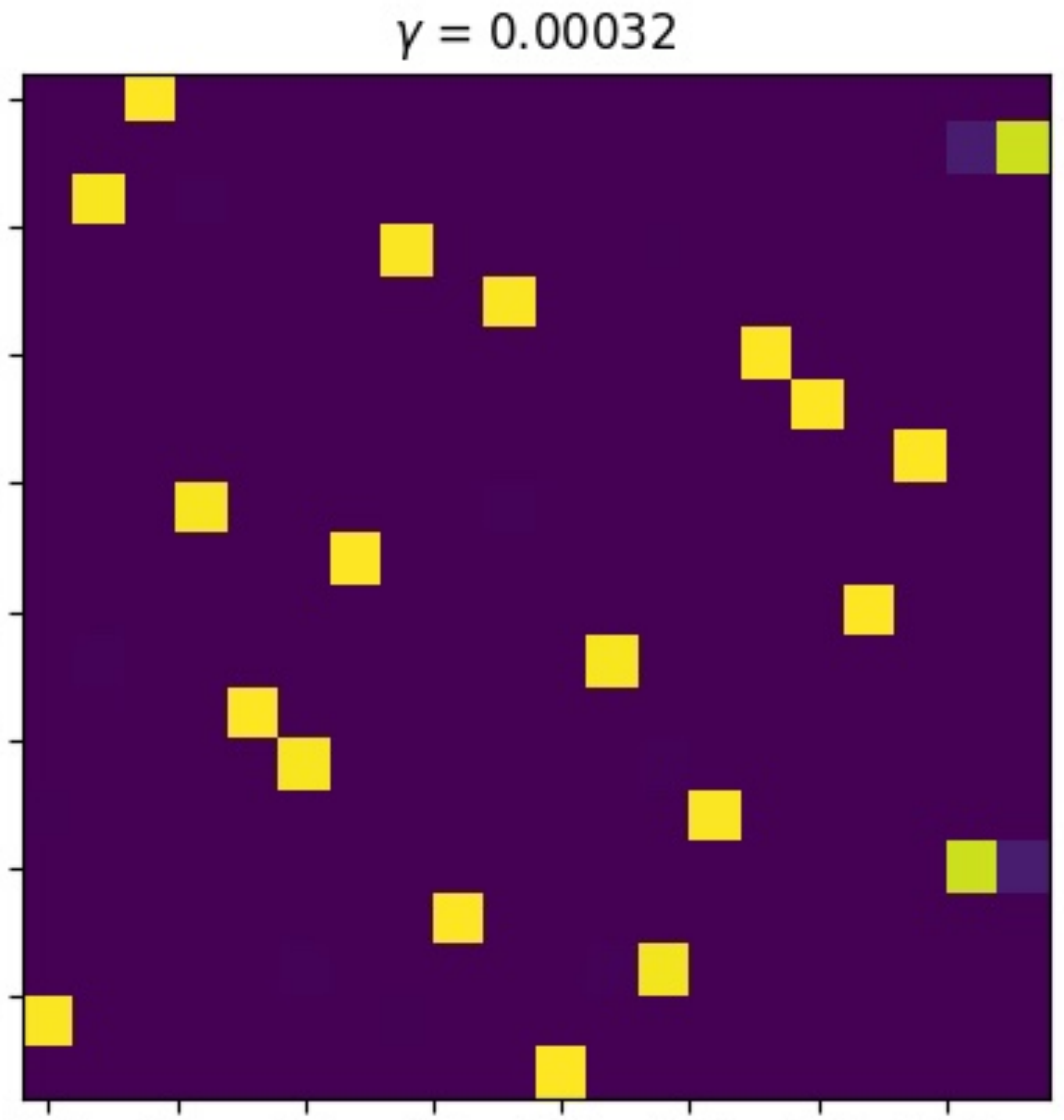


Solving Entropy Regularized OT

$$\min_{P \in U(a,b)} \langle P, M_{XY} \rangle - \gamma E(P)$$

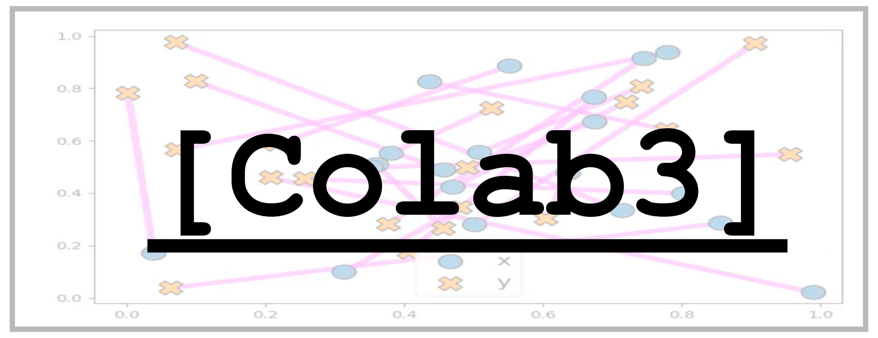


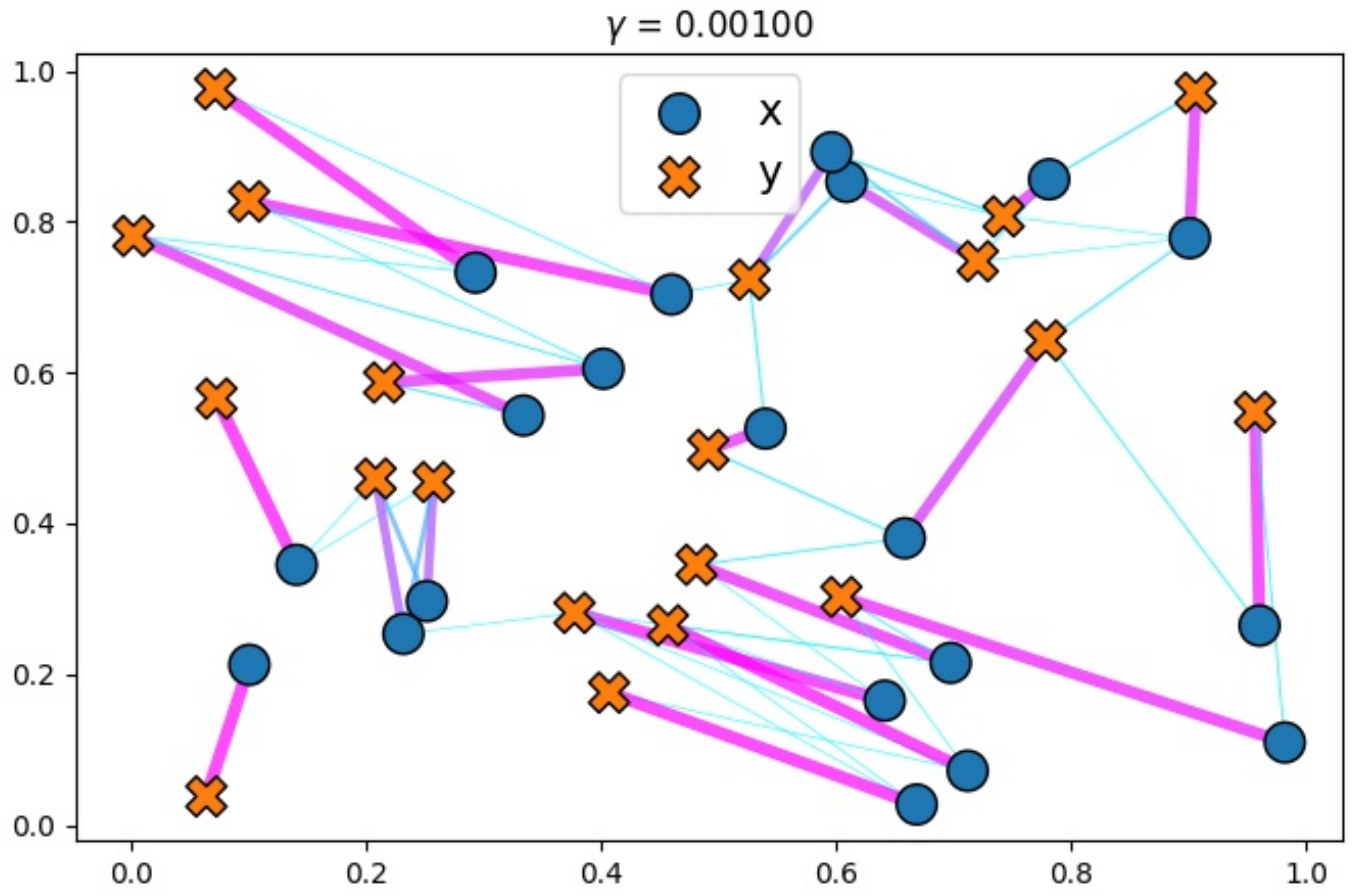


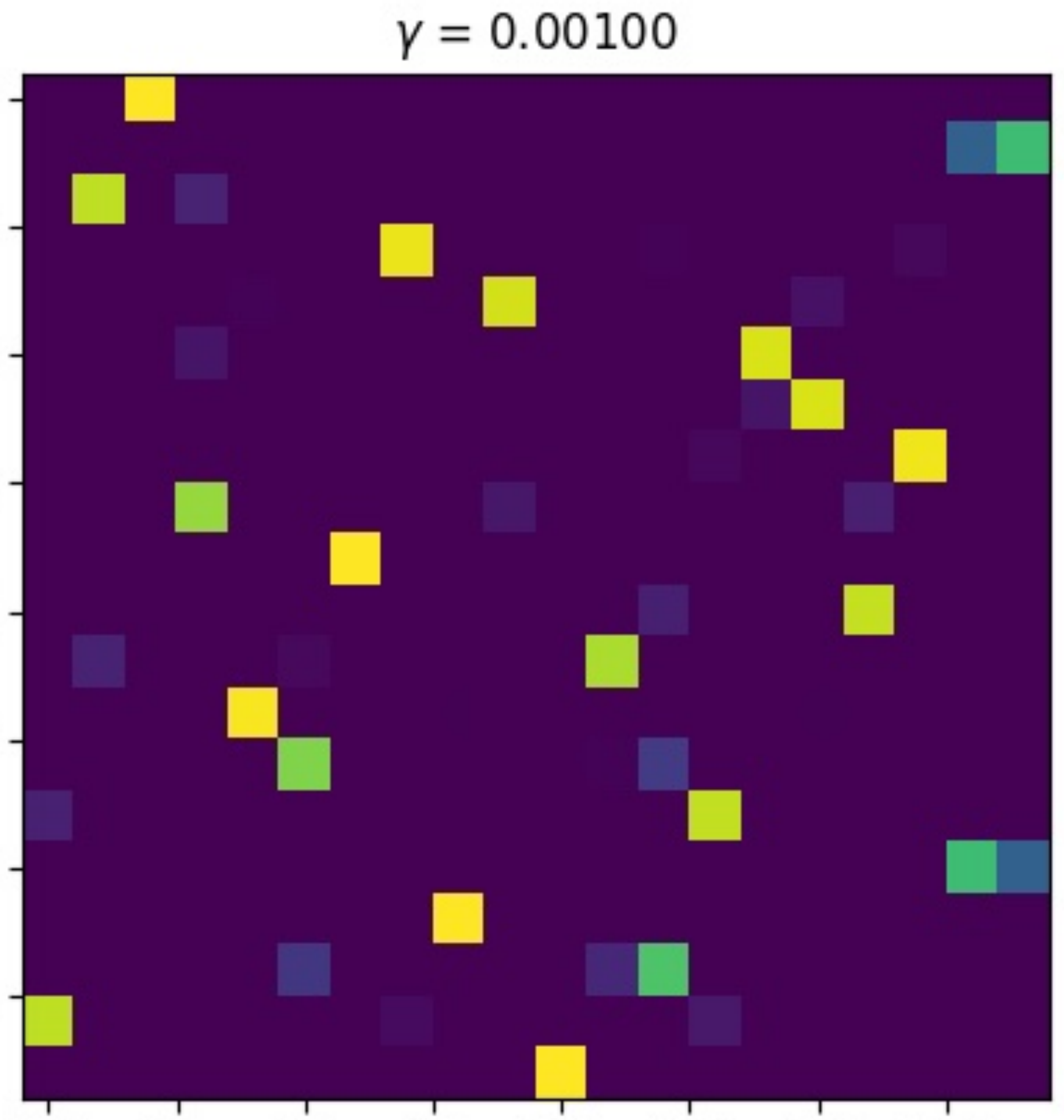


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle - \gamma E(\boldsymbol{P})$$

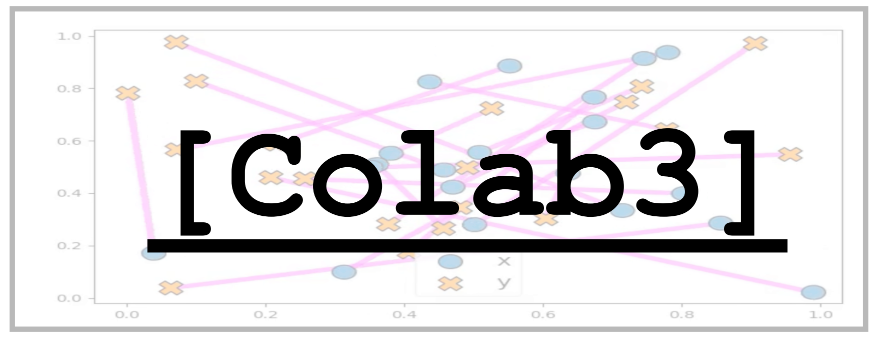


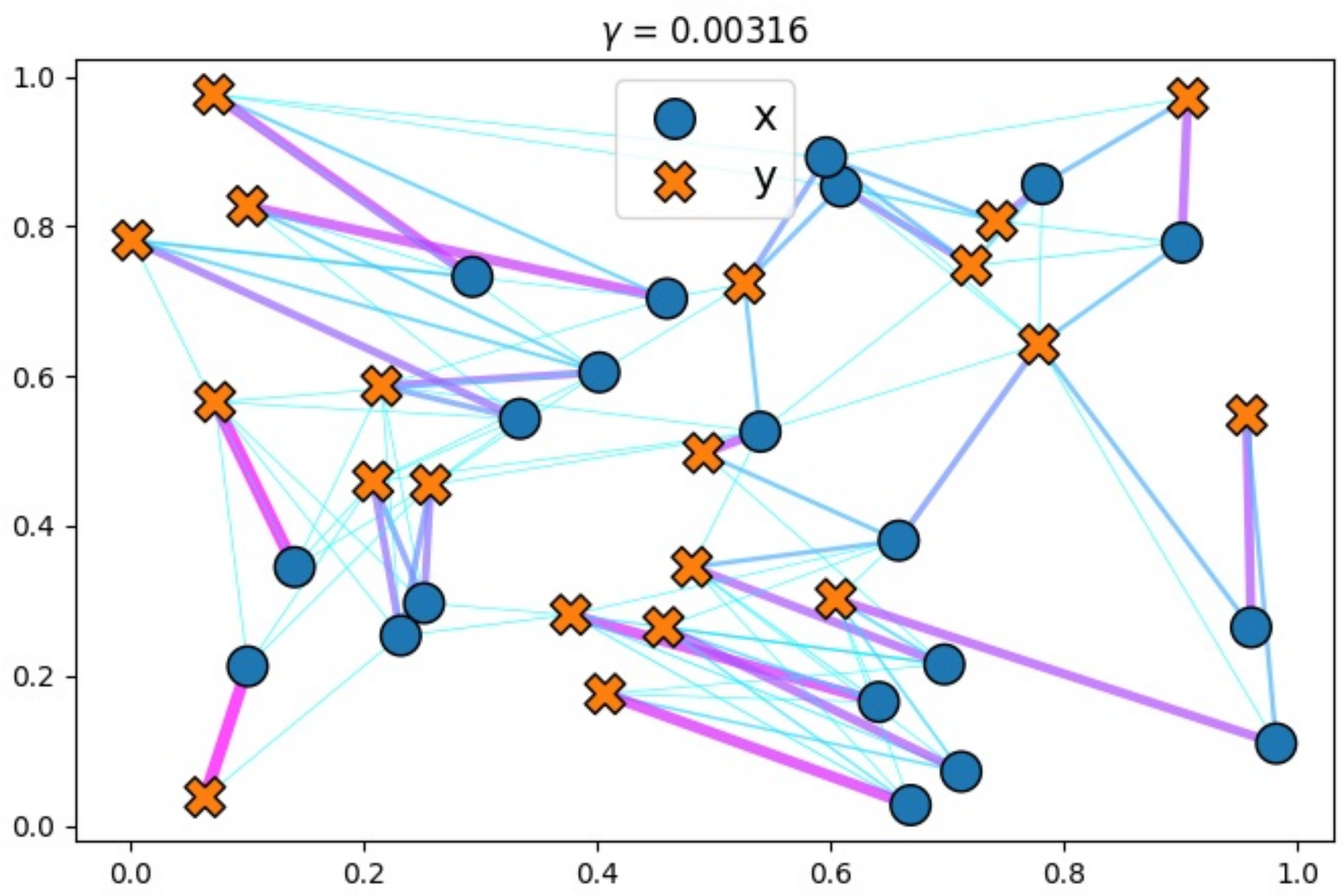


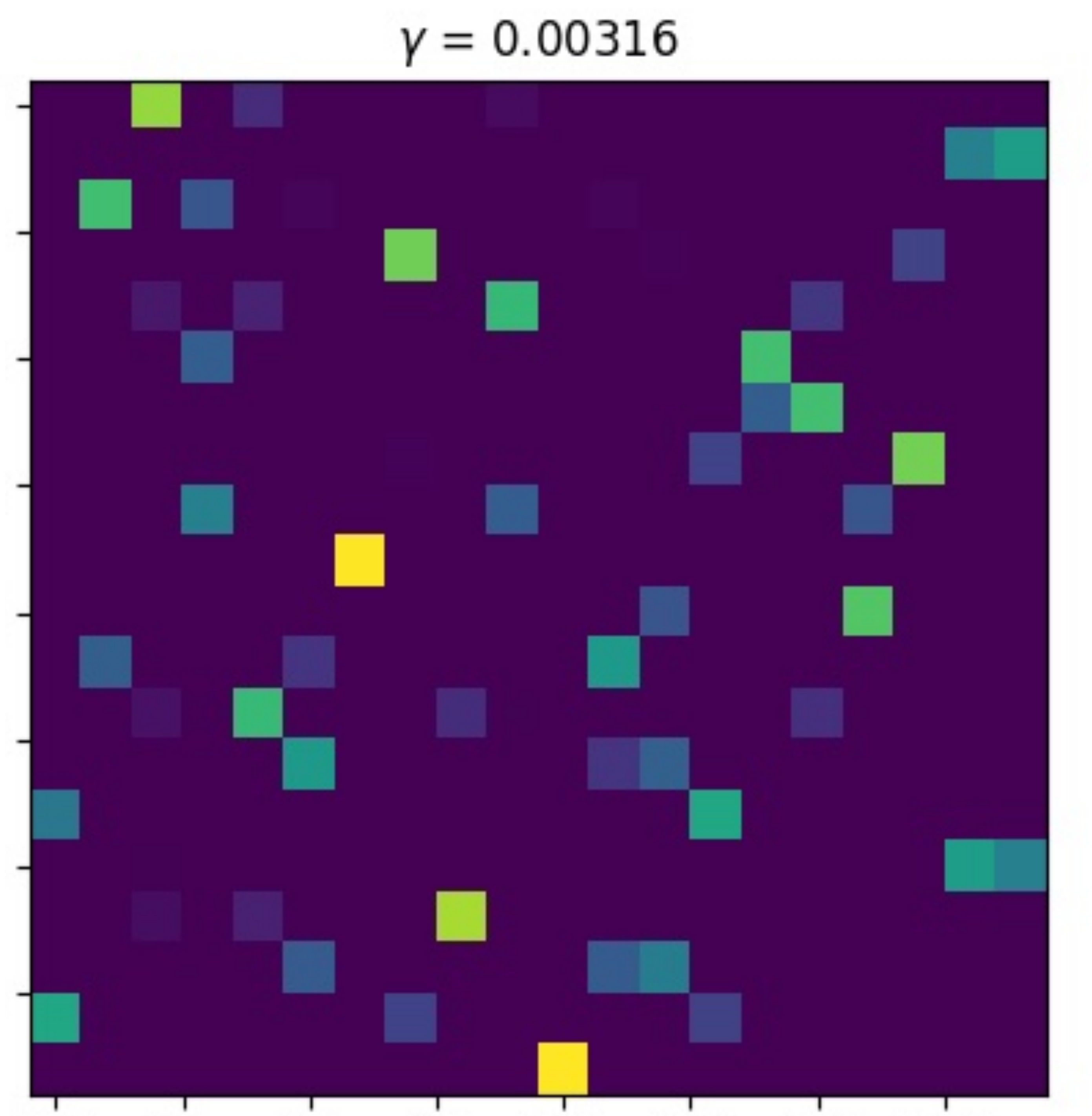


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P}\in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle - \gamma E(\boldsymbol{P})$$

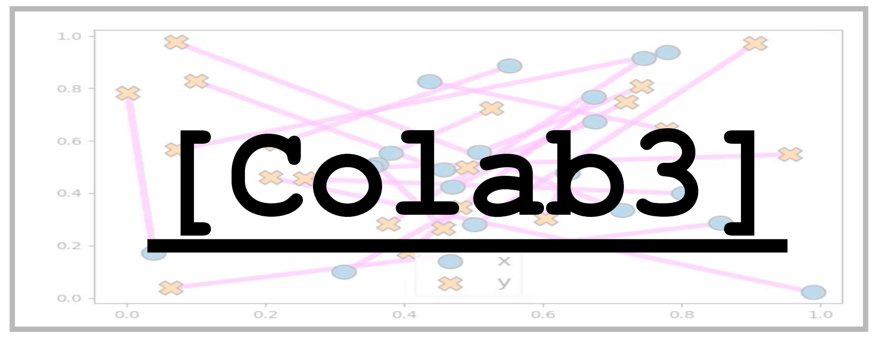


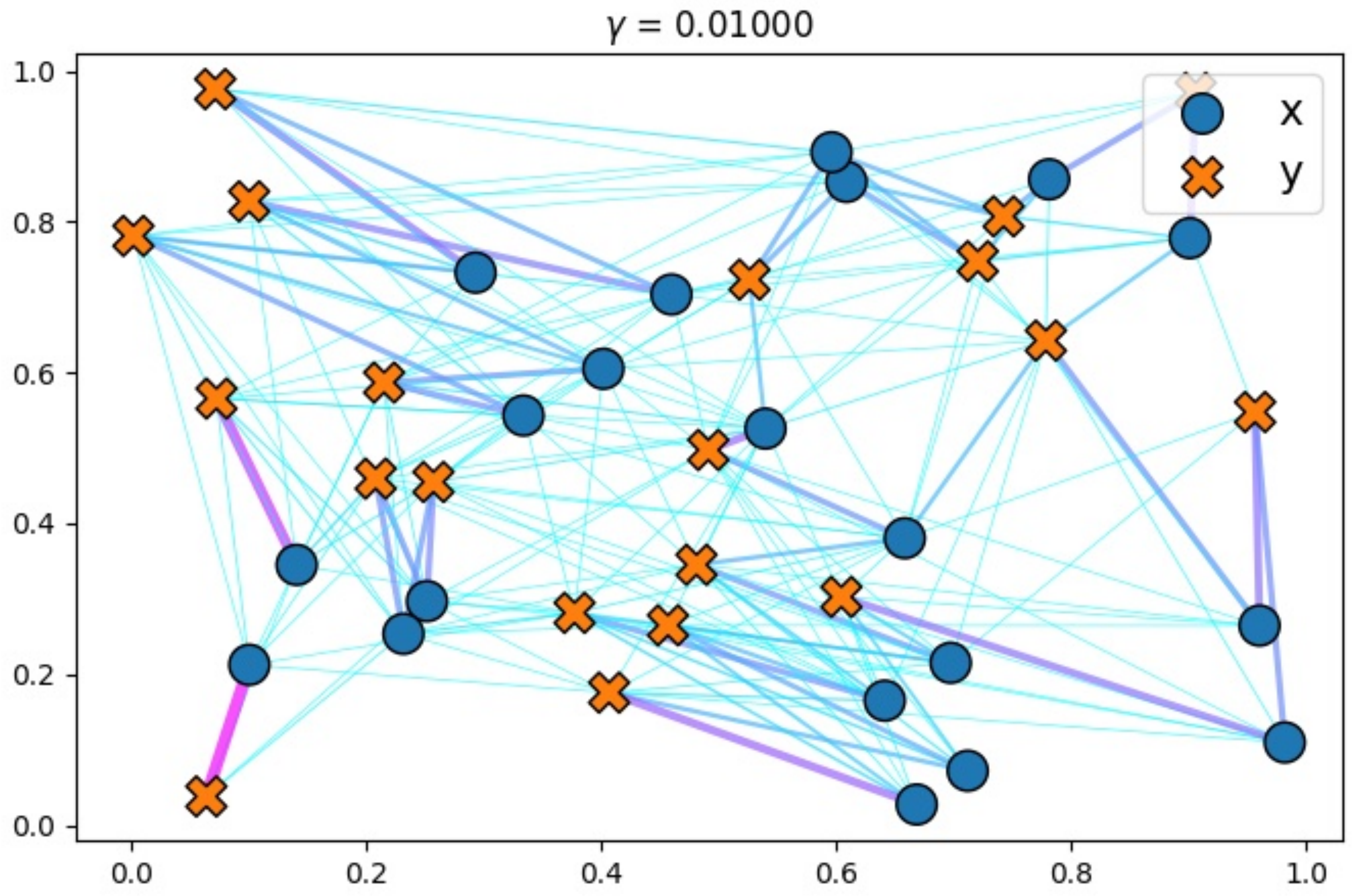


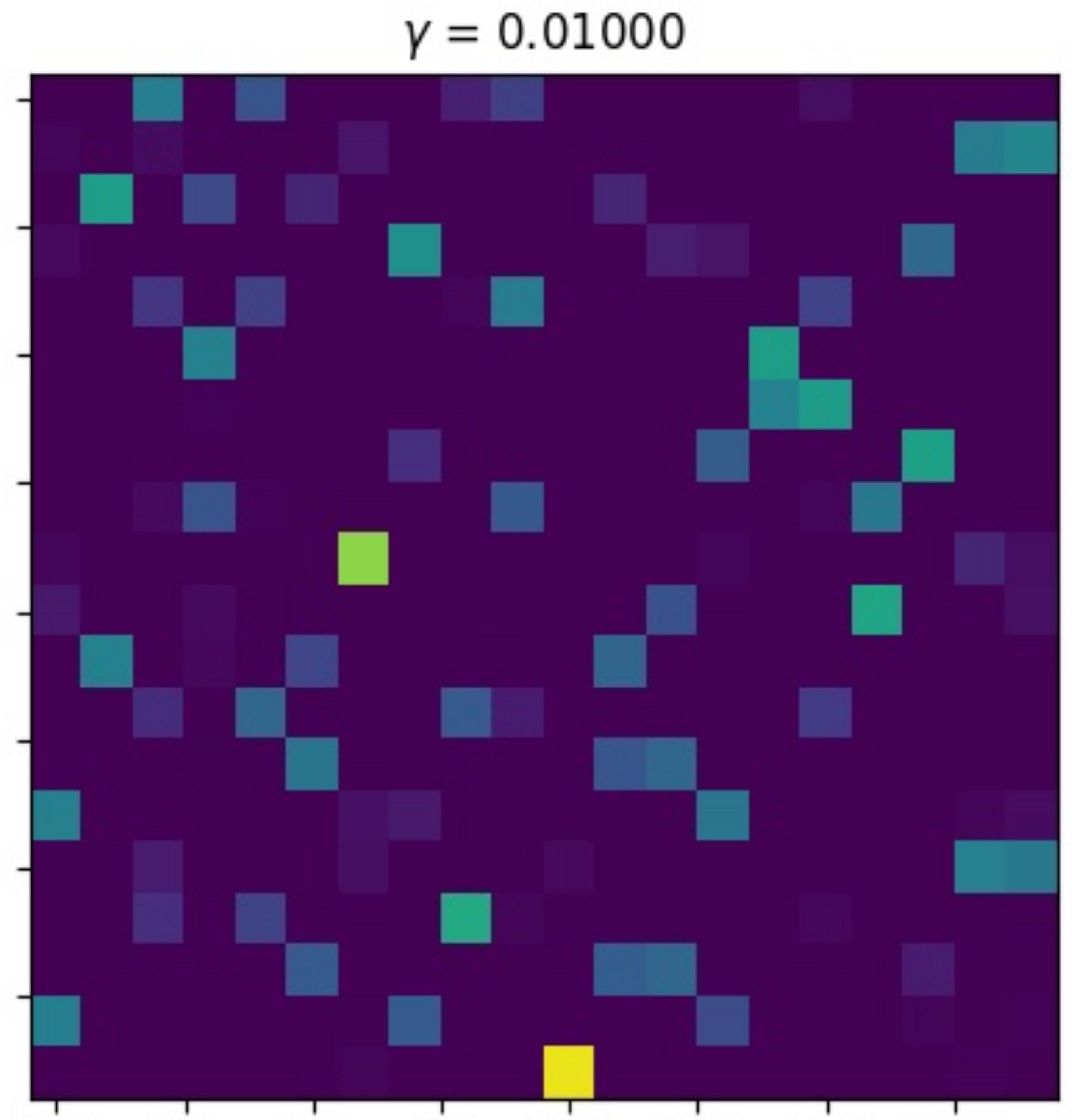


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle - \gamma E(\boldsymbol{P})$$

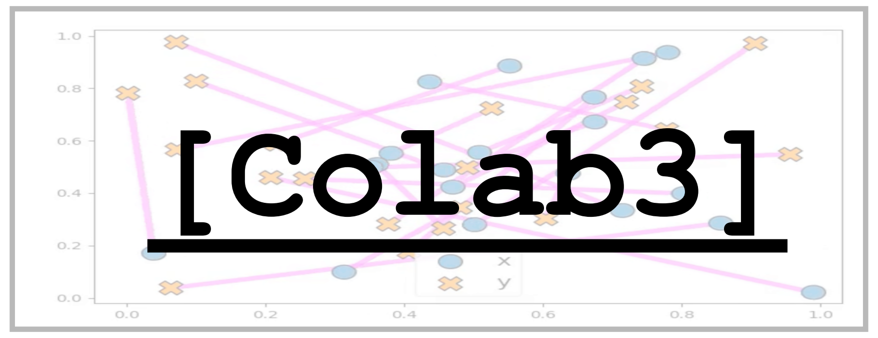


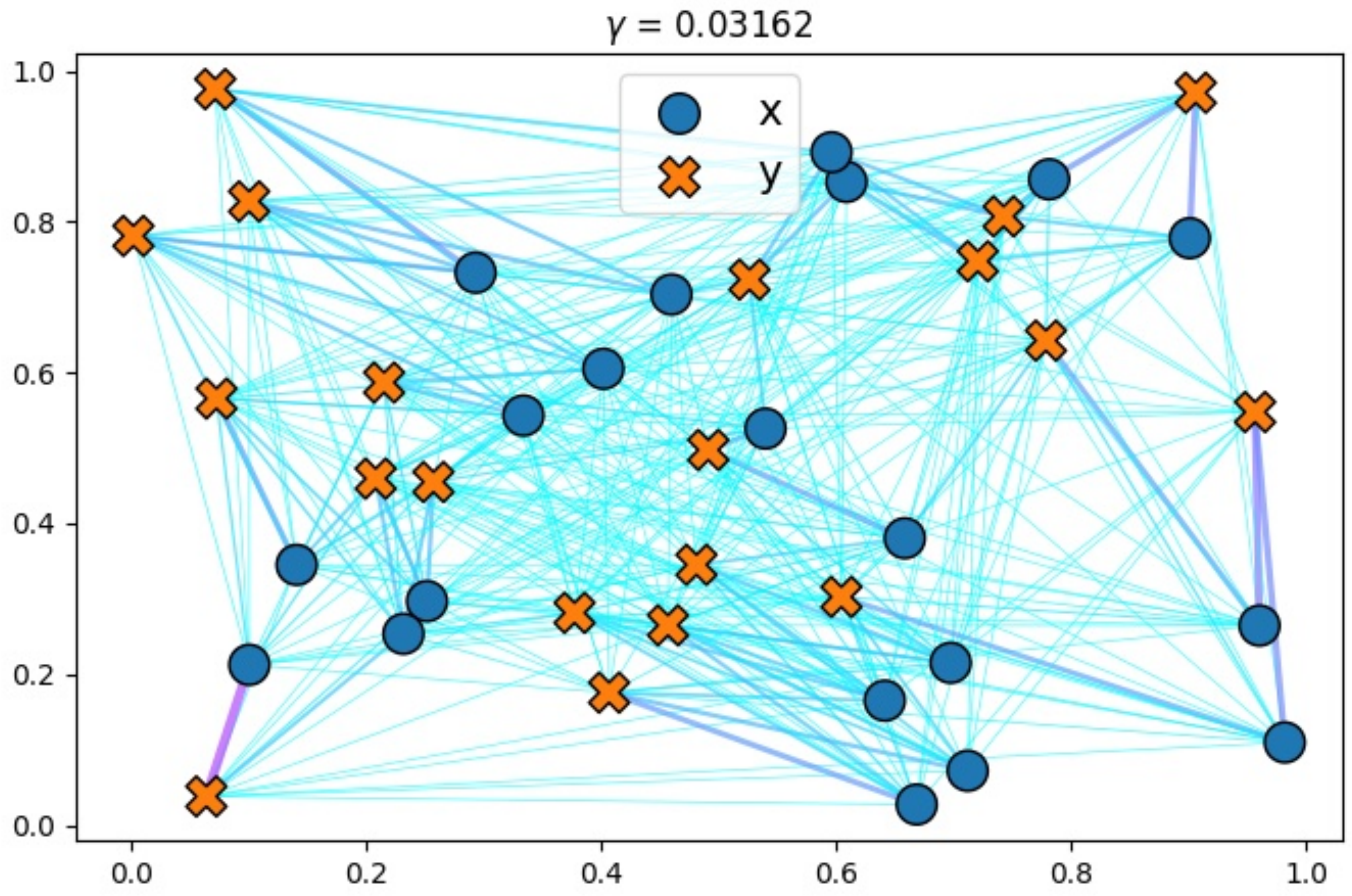


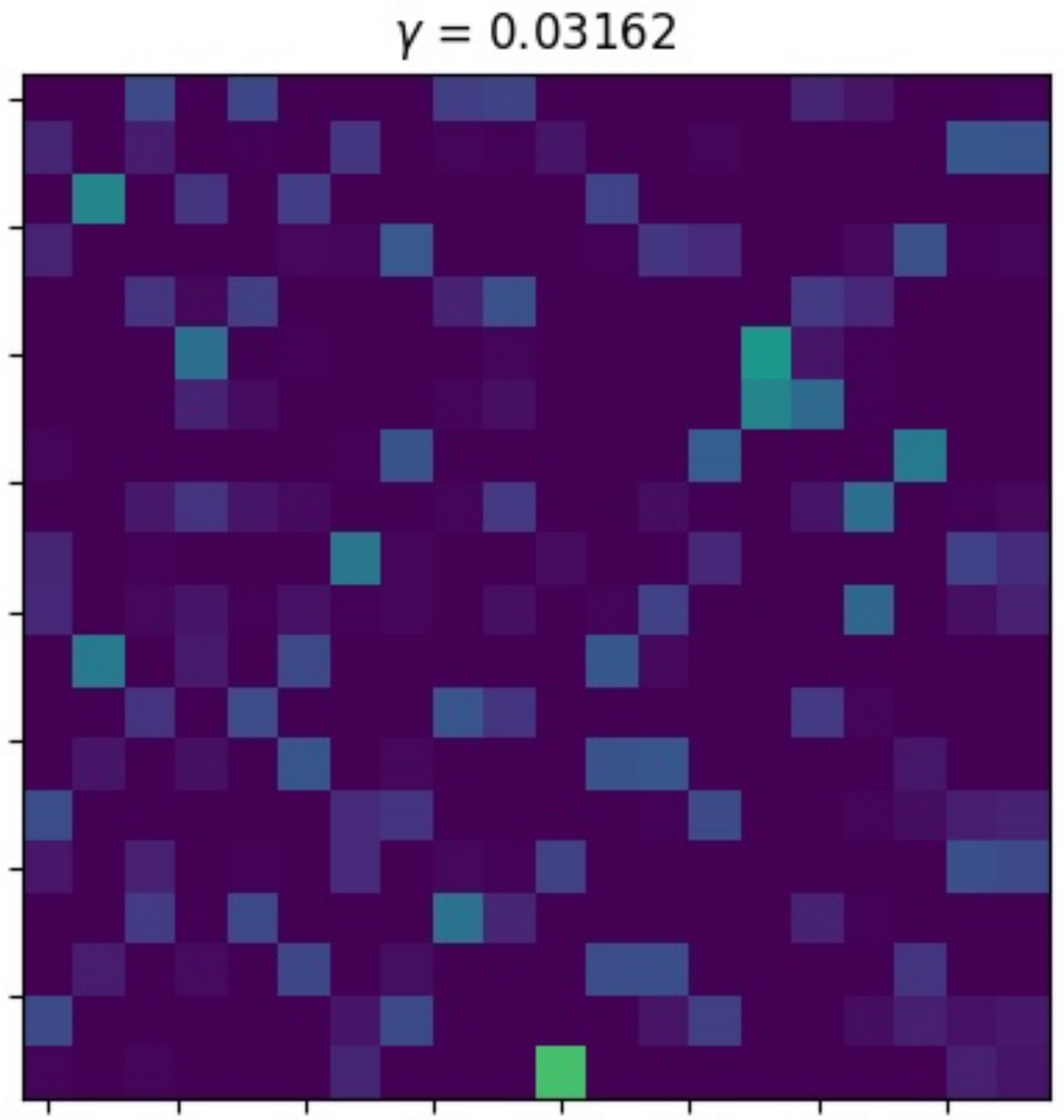


Solving Entropy Regularized OT

$$\min_{P \in U(a,b)} \langle P, M_{XY} \rangle - \gamma E(P)$$

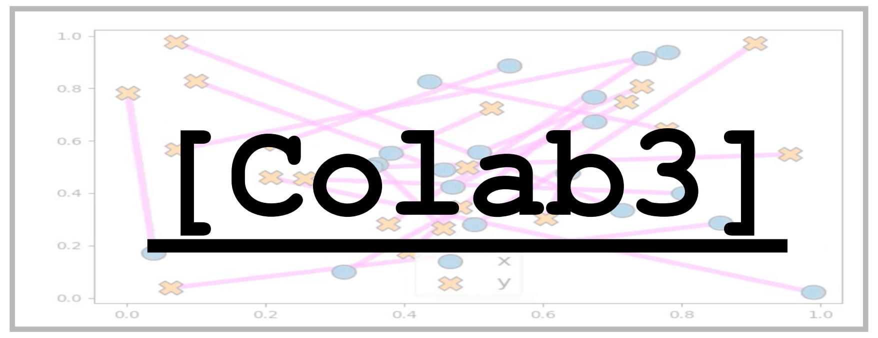


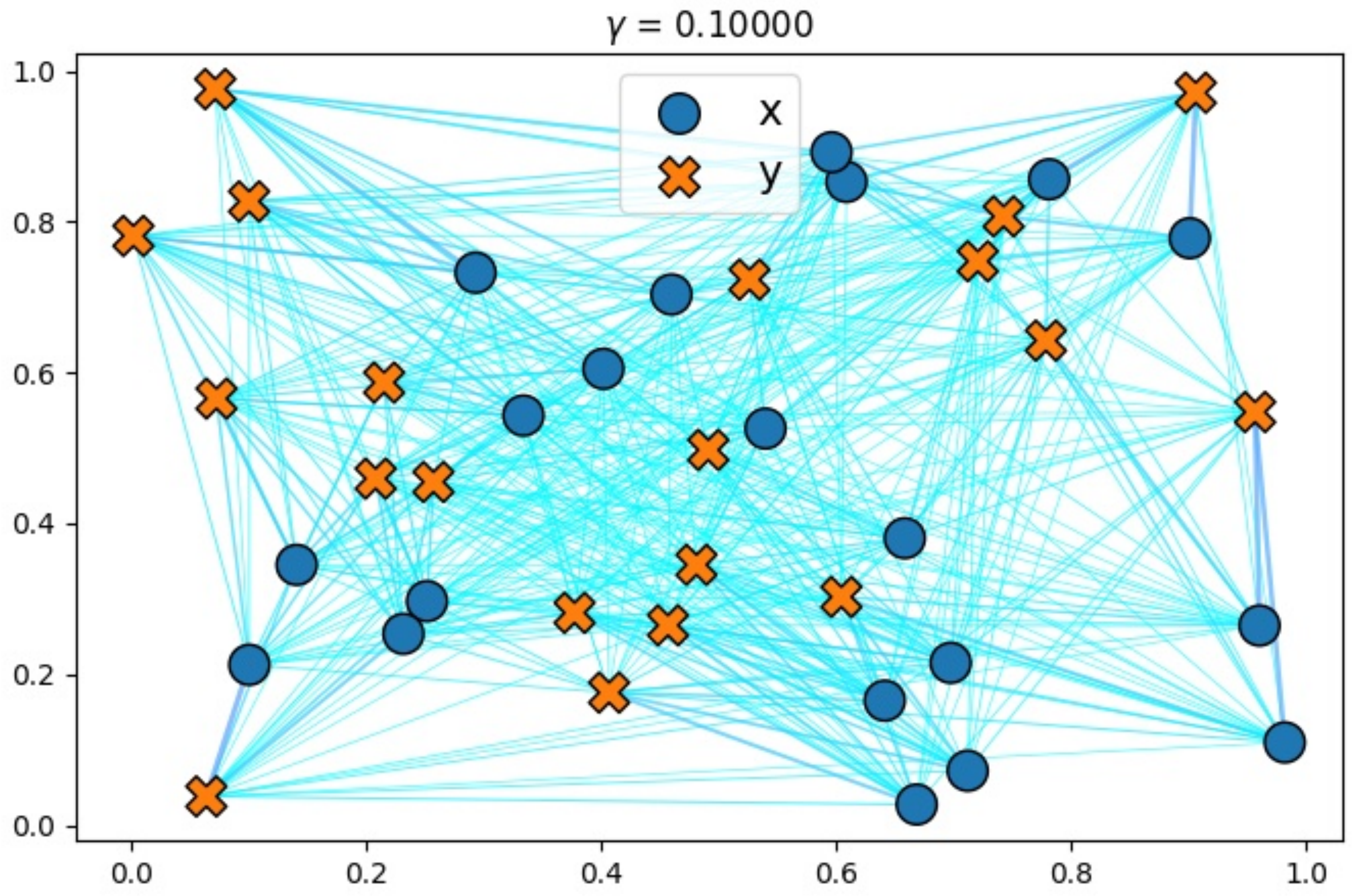


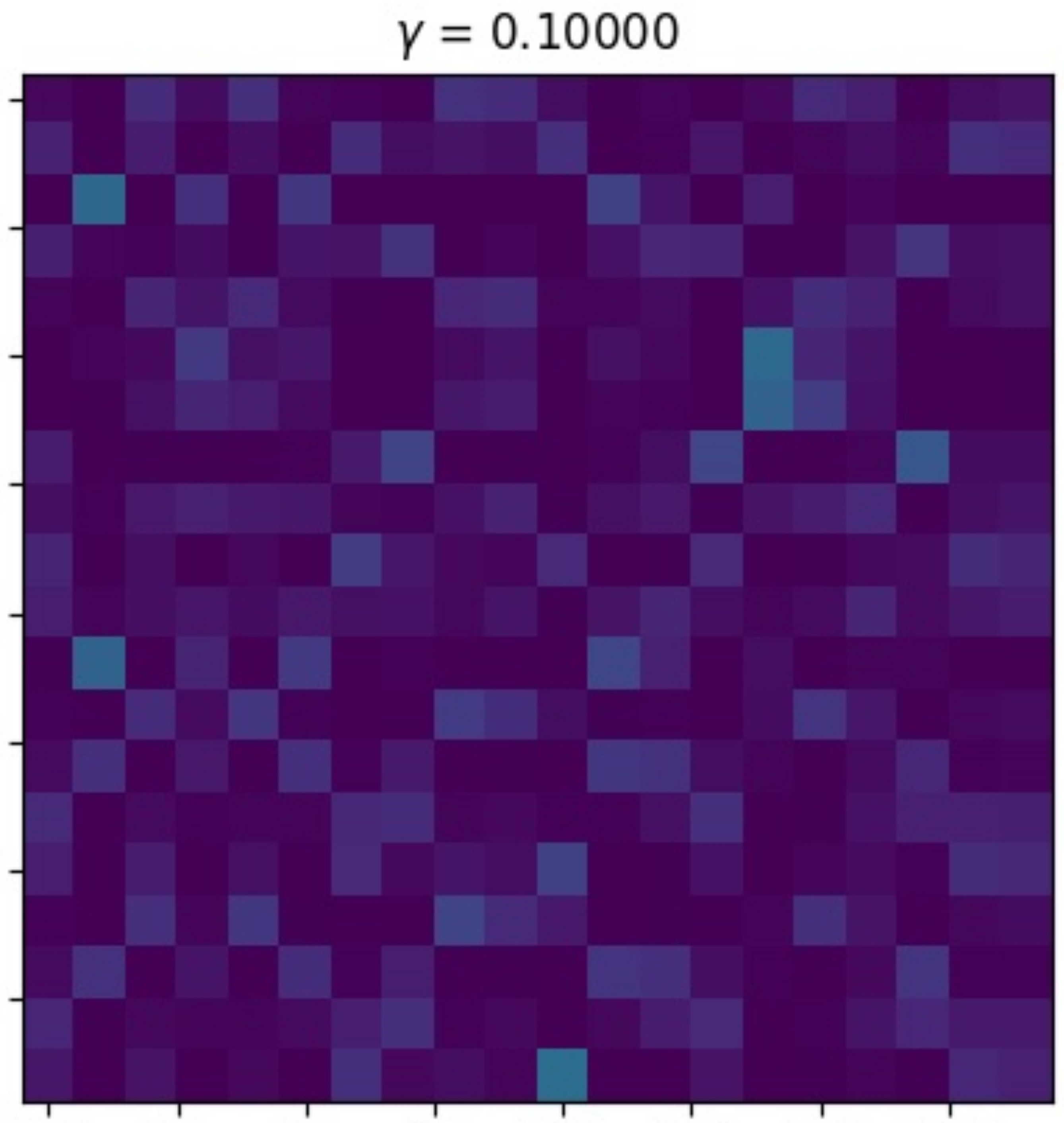


Solving Entropy Regularized OT

$$\min_{P \in U(a,b)} \langle P, M_{XY} \rangle - \gamma E(P)$$

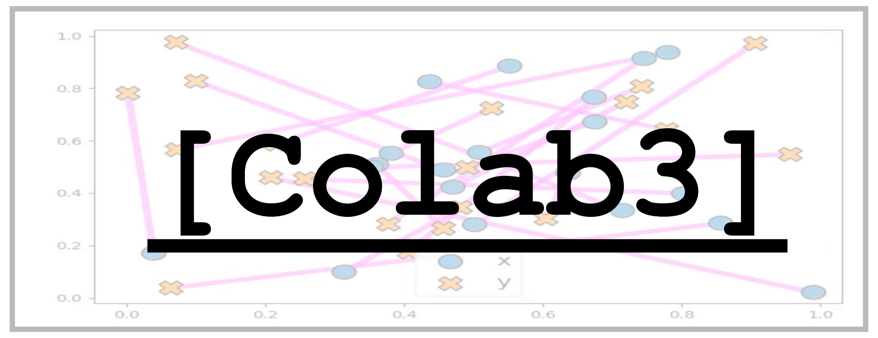


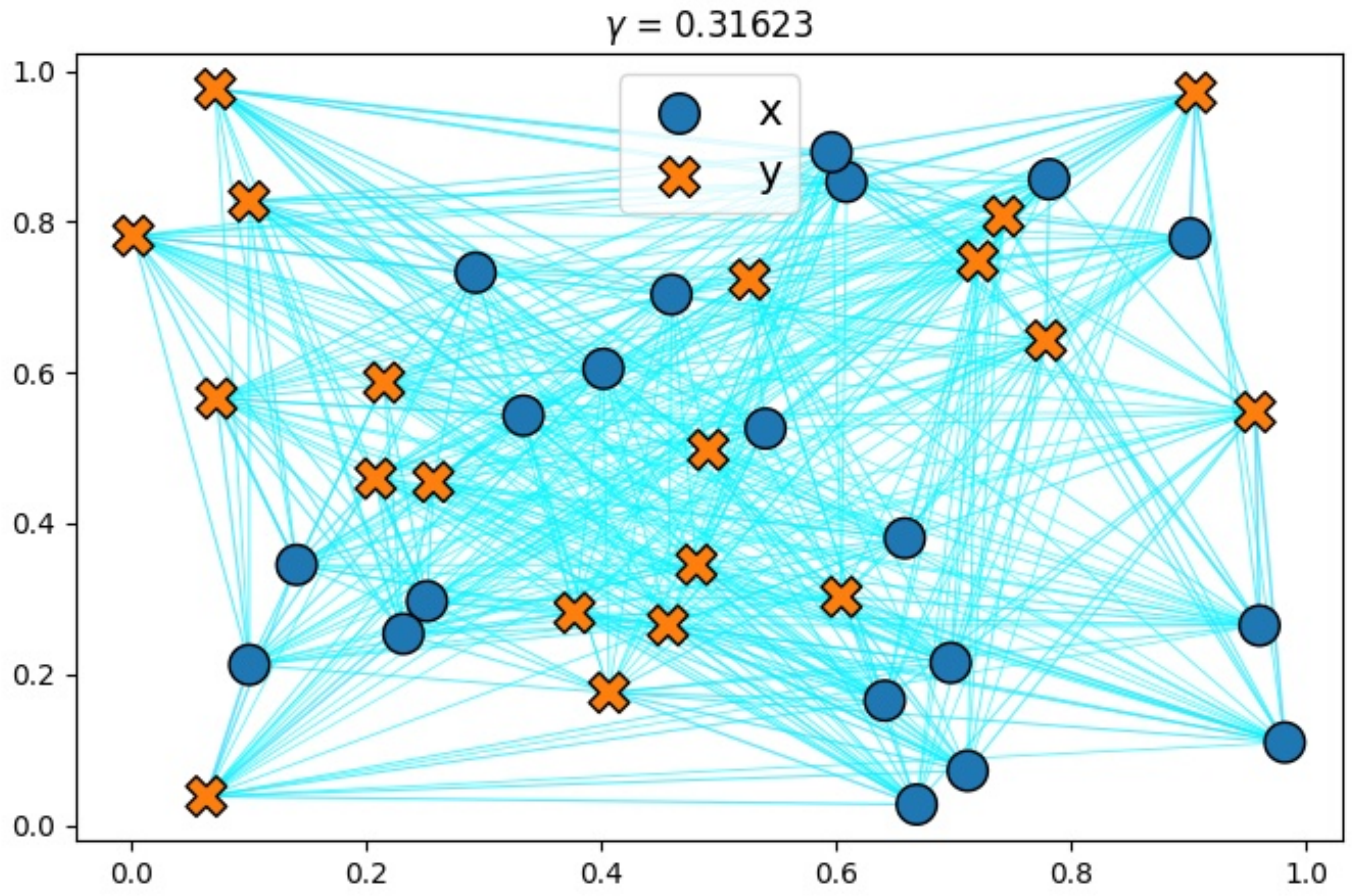


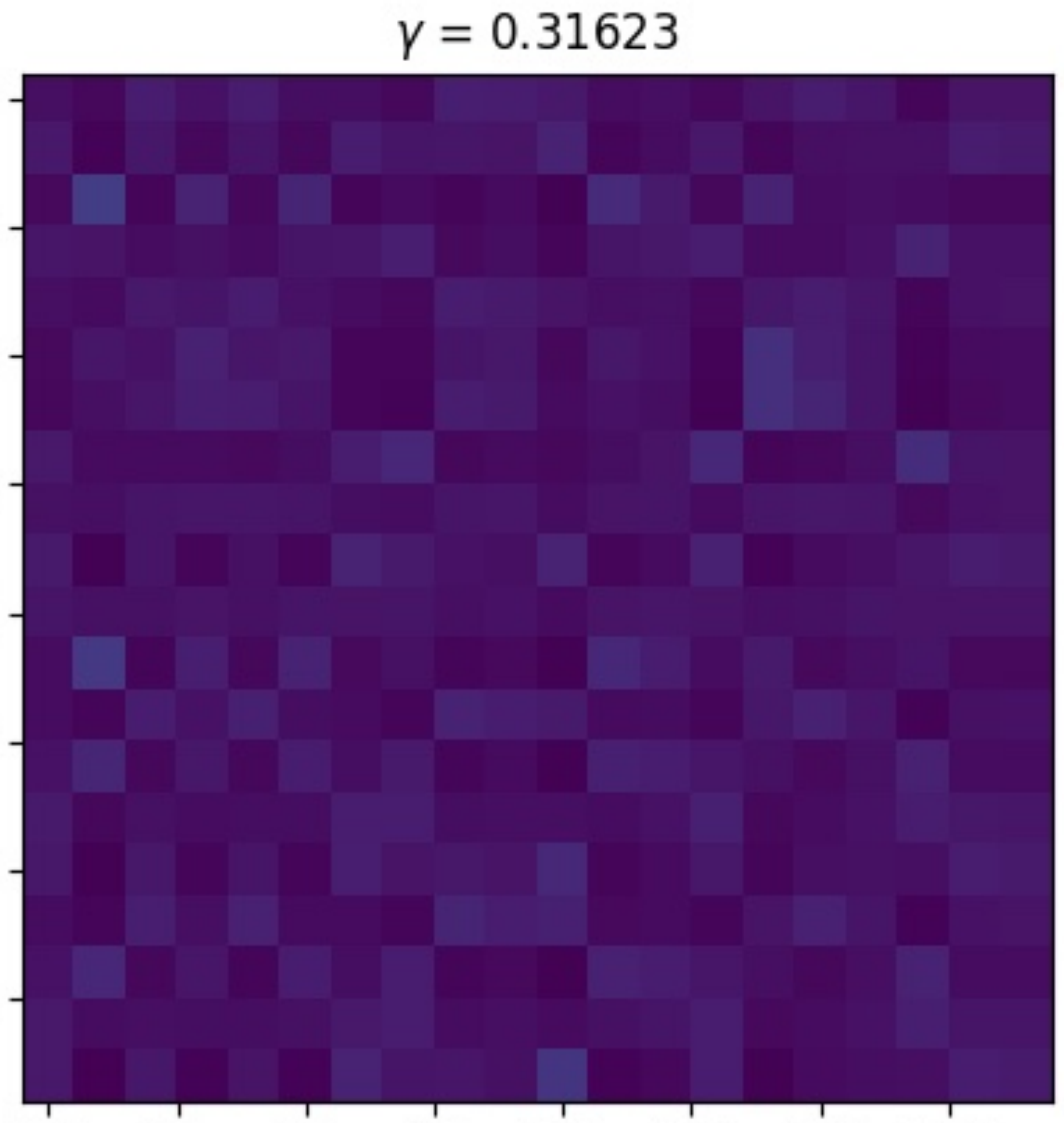


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle - \gamma E(\boldsymbol{P})$$

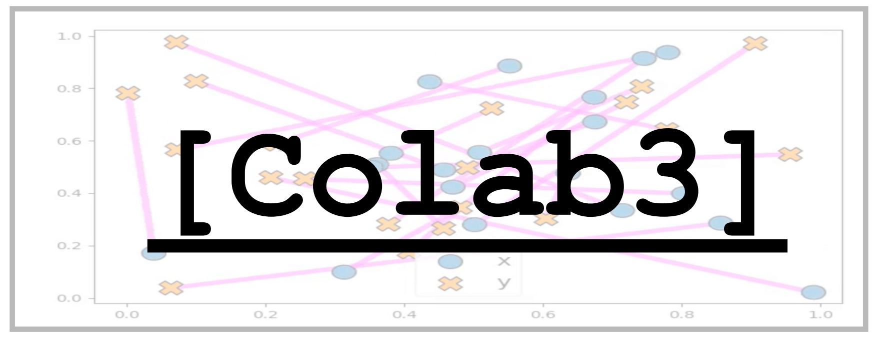


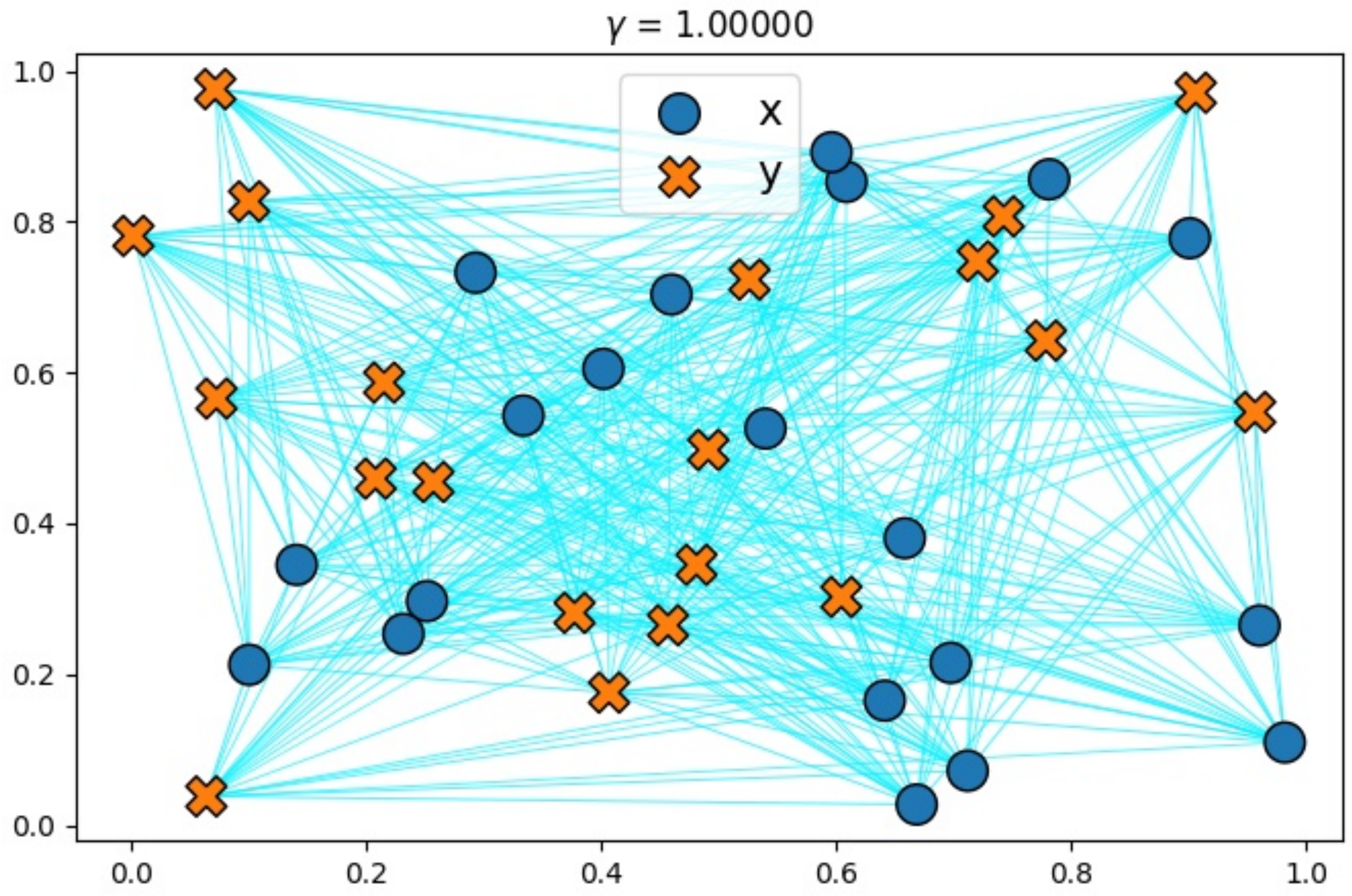


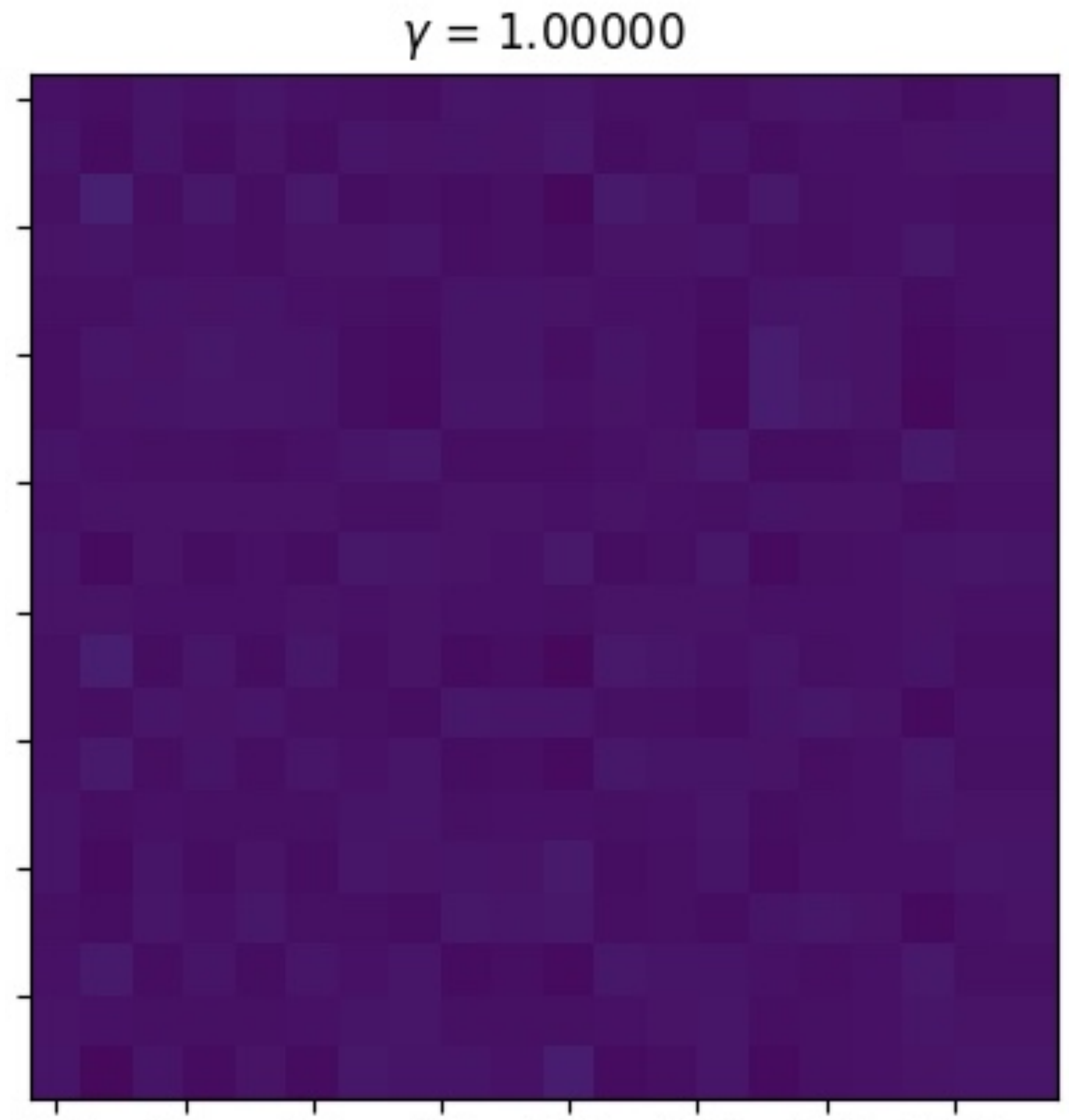


Solving Entropy Regularized OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle - \gamma E(\boldsymbol{P})$$







Solving Entropy Regularized OT

$$\textbf{Prop.}\ \text{If } P_\gamma \stackrel{\text{def}}{=} \underset{\boldsymbol{P}\in U(\boldsymbol{a},\boldsymbol{b})}{\text{argmin}} \langle \boldsymbol{P}, M_{\mathbf{XY}} \rangle - \gamma E(\boldsymbol{P})$$

$$\text{then } \exists! \boldsymbol{u} \in \mathbb{R}^n_+, \boldsymbol{v} \in \mathbb{R}^m_+, \text{ such that}$$

$$P_\gamma = \textbf{diag}(\boldsymbol{u}) K \textbf{diag}(\boldsymbol{v}), \quad K \stackrel{\text{def}}{=} e^{-M_{\mathbf{XY}}/\gamma}$$

Sinkhorn's Algorithm (a.k.a. IPFP, RAS): Initialize $\boldsymbol{v}$, then repeat

$$\boldsymbol{u} \leftarrow \frac{\boldsymbol{a}}{K\boldsymbol{v}} \qquad \boldsymbol{v} \leftarrow \frac{\boldsymbol{b}}{K^T\boldsymbol{u}}$$

[Deming+40, Wilson62, Sinkhorn64, Erlander92, Galichon+10, Cuturi13]

Solving Entropy Regularized OT

Prop. If $P_\gamma \stackrel{\text{def}}{=} \underset{\boldsymbol{P}\in U(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle - \gamma E(\boldsymbol{P})$

then $\exists! \boldsymbol{u} \in \mathbb{R}^n_+, \boldsymbol{v} \in \mathbb{R}^m_+$, such that

$$P_\gamma = \mathbf{diag}(\boldsymbol{u}) K \mathbf{diag}(\boldsymbol{v}), \quad K \stackrel{\text{def}}{=} e^{-M_{\boldsymbol{XY}}/\gamma}$$

Sinkhorn's Algorithm, in log domain:

$$(\boldsymbol{u}, \boldsymbol{v}) \stackrel{\text{def}}{=} (e^{\boldsymbol{\alpha}/\gamma}, e^{\boldsymbol{\beta}/\gamma})$$

$$\boldsymbol{\alpha} \leftarrow \gamma \left(\log \boldsymbol{a} - \log \; K(e^{\boldsymbol{\beta}/\gamma}) \right) \qquad \boldsymbol{\beta} \leftarrow \gamma \left(\log \boldsymbol{b} - \log \; K^T(e^{\boldsymbol{\alpha}/\gamma}) \right)$$

Unbalanced Generalizations for OT

$$\min_{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle$$

[Frogner+15, Chizat+16]

$$\min_{\boldsymbol{P} \in \mathbb{R}_+^{\boldsymbol{n} \times \boldsymbol{m}}} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle + \rho_{\boldsymbol{a}} \mathrm{KL}(\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}} \| \boldsymbol{a}) + \rho_{\boldsymbol{b}} \mathrm{KL}(\boldsymbol{P}^T \mathbf{1}_{\boldsymbol{n}} \| \boldsymbol{b})$$

Unbalanced Generalizations for OT

$$\min_{P \in U(a,b)} \langle P, M_{XY} \rangle$$

[Frogner+15, Chizat+16]

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$

Still a Sinkhorn-esque Algorithm!

$$u = (a / Kv)^{\frac{\rho_a}{\rho_a + \gamma}} \qquad v = (b / K^T u)^{\frac{\rho_b}{\rho_b + \gamma}}$$

Unbalanced Generalizations for OT

$$\min_{P \in U(a,b)} \langle P, M_{XY} \rangle$$

[Frogner+15, Chizat+16]

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$

Still a Sinkhorn-esque Algorithm!

$$u = (a / Kv)^{\frac{\rho_a}{\rho_a + \gamma}} \qquad v = (b / K^T u)^{\frac{\rho_b}{\rho_b + \gamma}}$$

$$0 \leq \tau_a \leq 1 \qquad 0 \leq \tau_b \leq 1$$

Semi-Balanced Sinkhorn

$$\min_{\boldsymbol{P} \in \mathbb{R}_+^{\boldsymbol{n} \times \boldsymbol{m}}} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle + \rho_{\boldsymbol{a}} \mathrm{KL}(\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}} \| \boldsymbol{a}) + \rho_{\boldsymbol{b}} \mathrm{KL}(\boldsymbol{P}^T \mathbf{1}_{\boldsymbol{n}} \| \boldsymbol{b}) - \gamma E(\boldsymbol{P})$$

Semi-Balanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$

$$\rho_b = \infty$$

Semi-Balanced Sinkhorn

$$\min_{\substack{\boldsymbol{P}\in\mathbb{R}_+^{\boldsymbol{n}\times\boldsymbol{m}} \\ \boldsymbol{P}^T\mathbf{1}_n=\boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle + \rho_{\boldsymbol{a}}\mathrm{KL}(\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}}\|\boldsymbol{a}) - \gamma E(\boldsymbol{P})$$

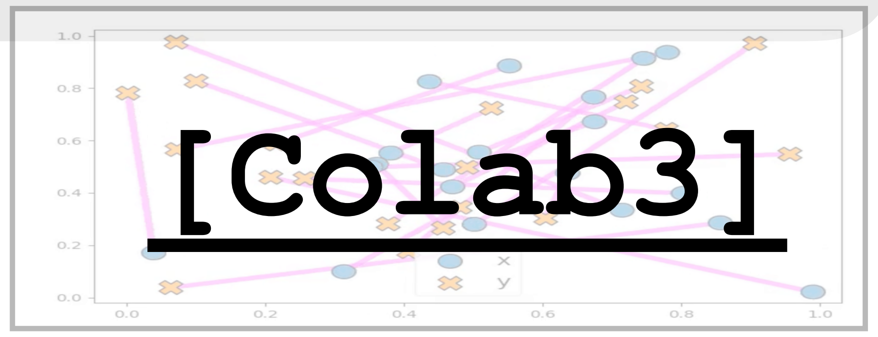
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$

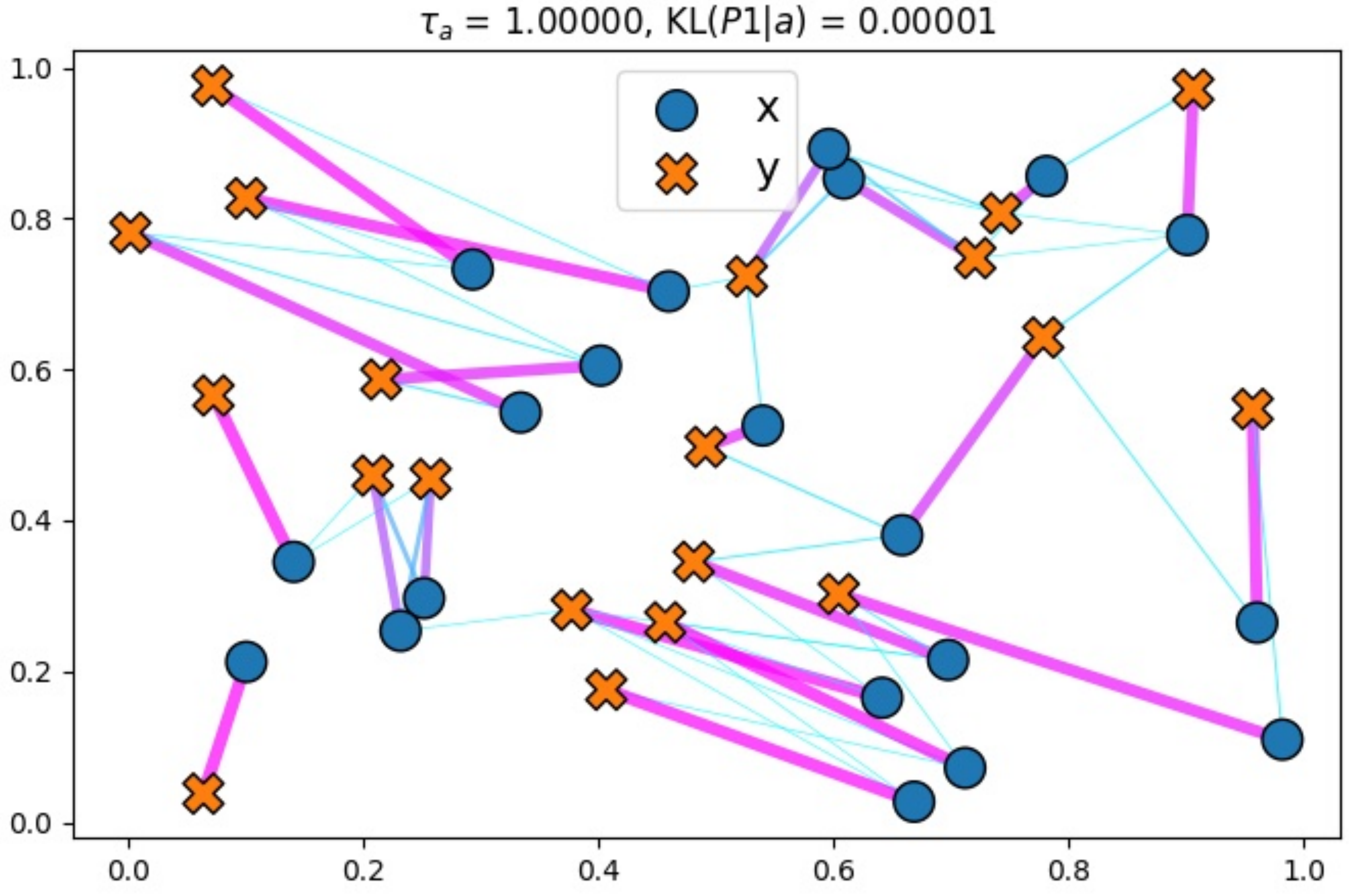
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$

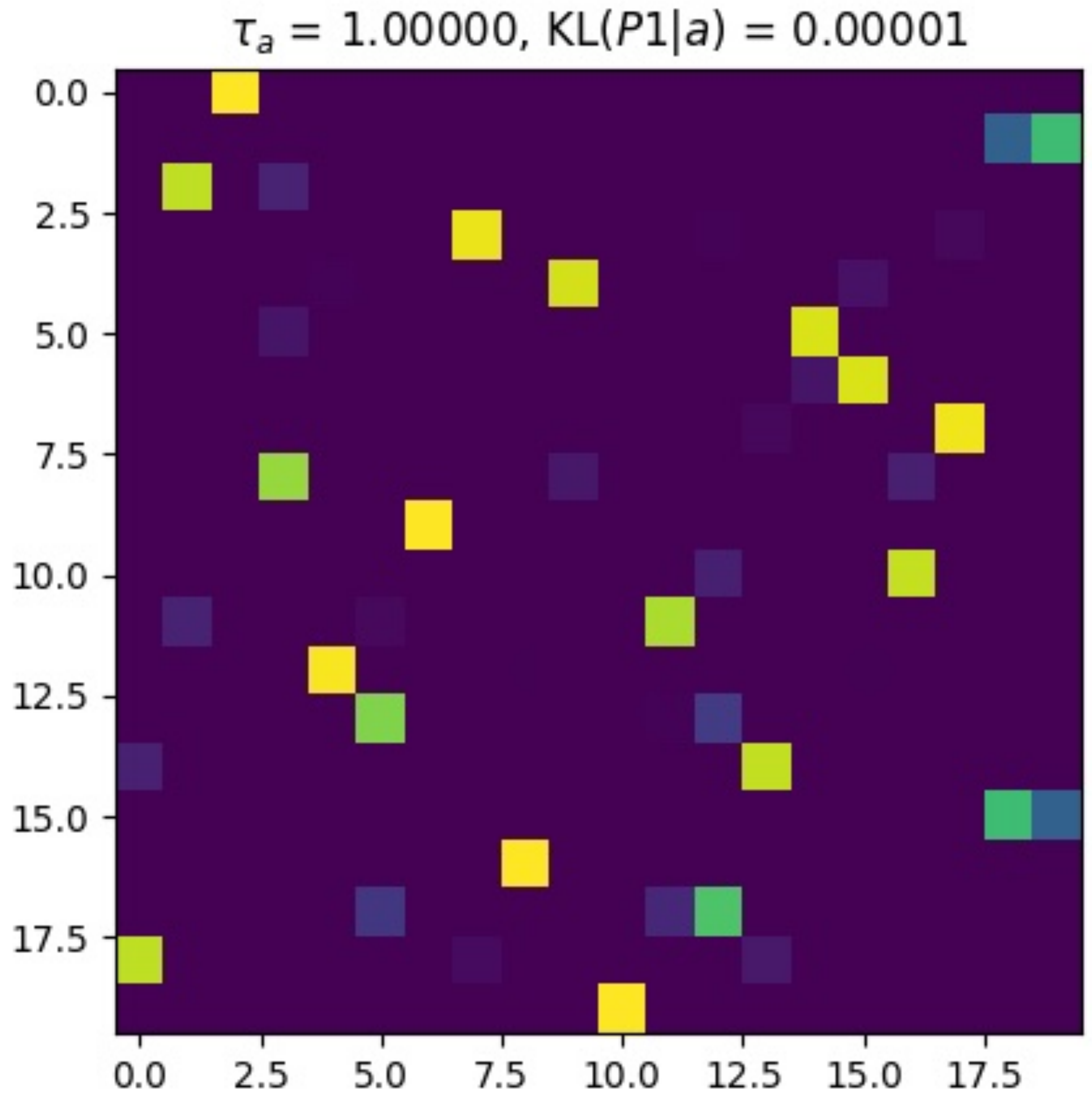
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



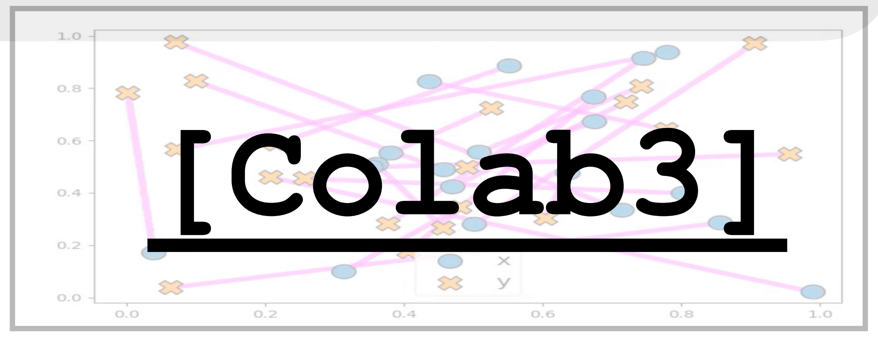
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



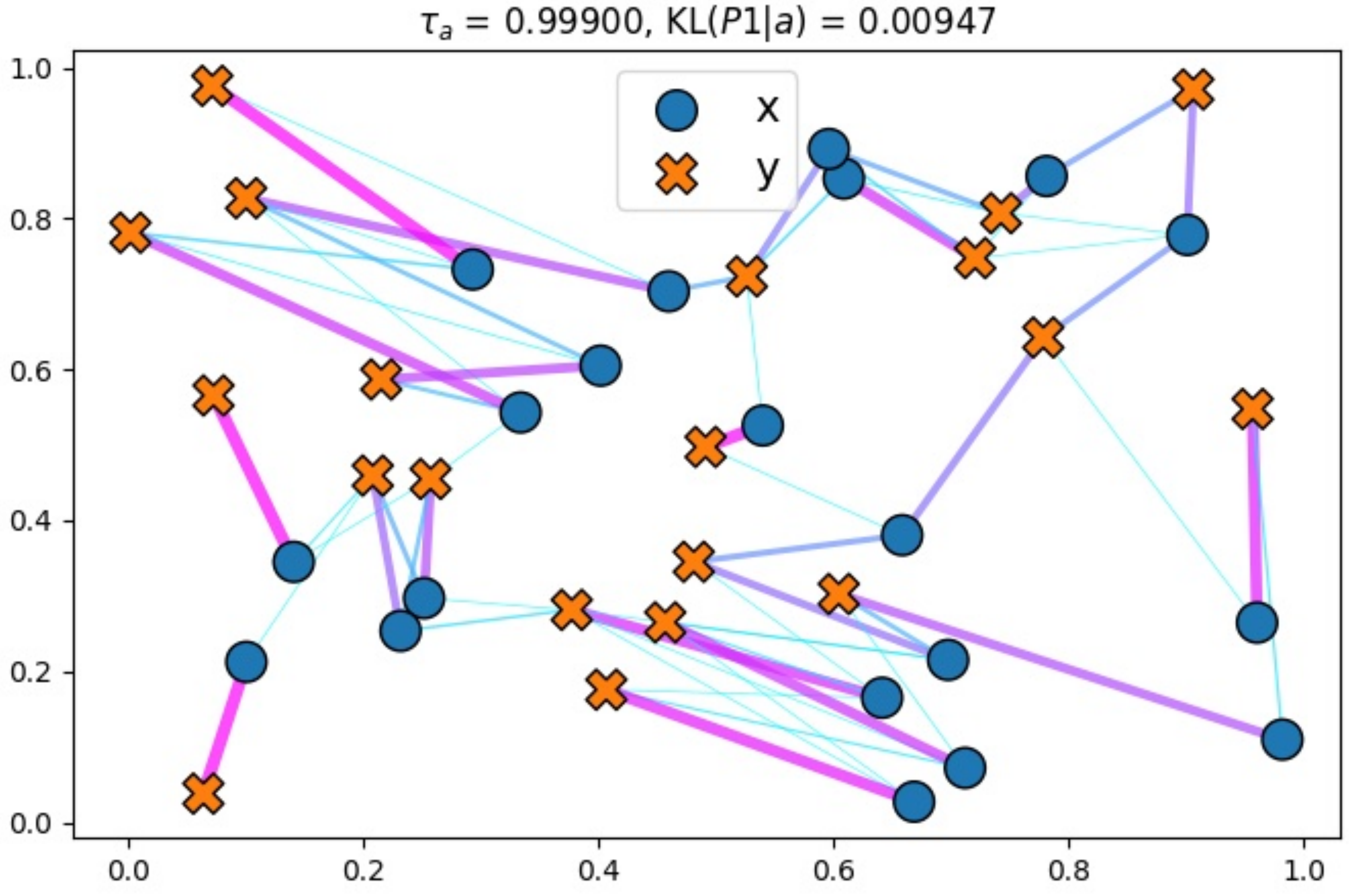


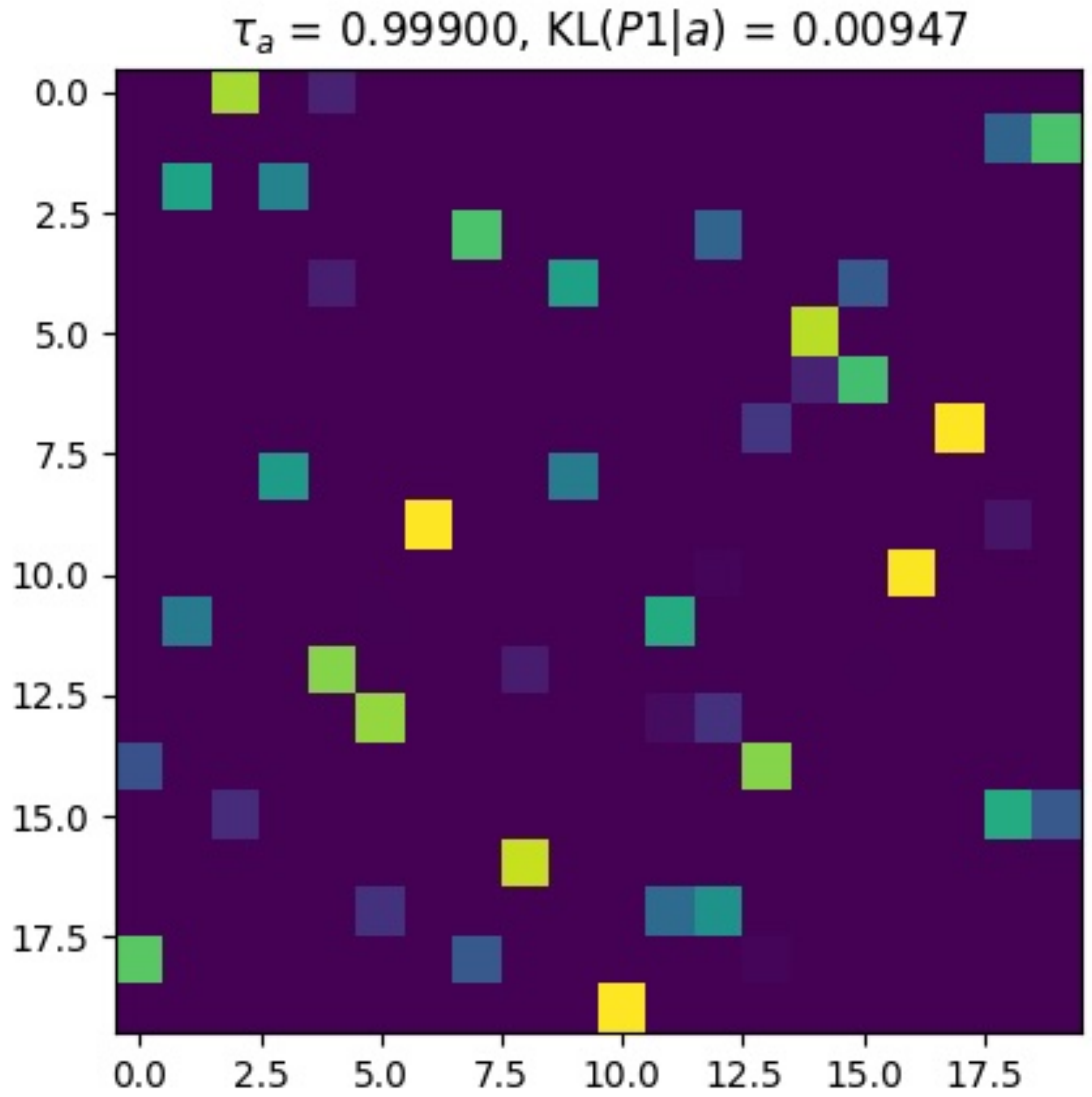
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



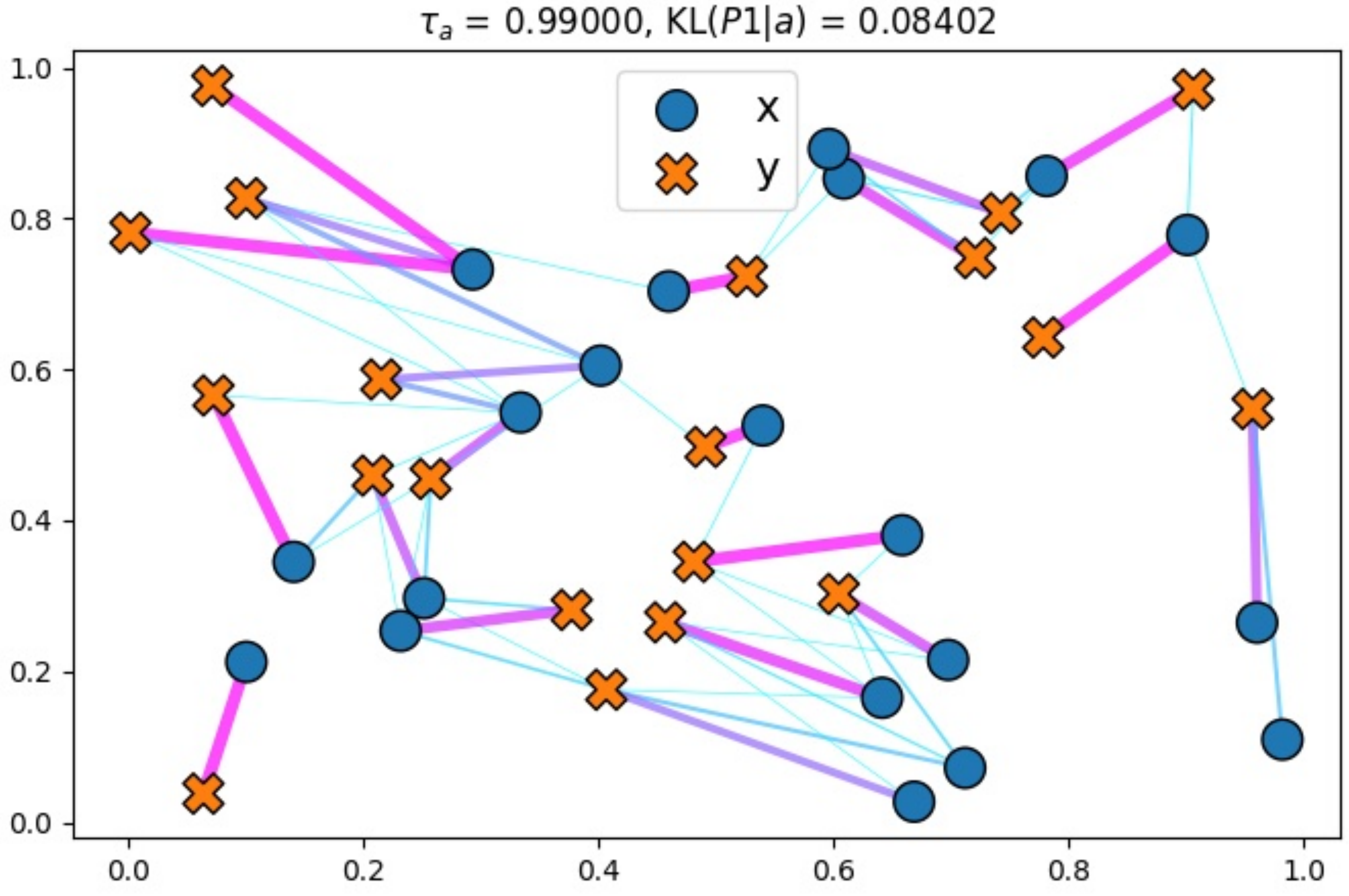


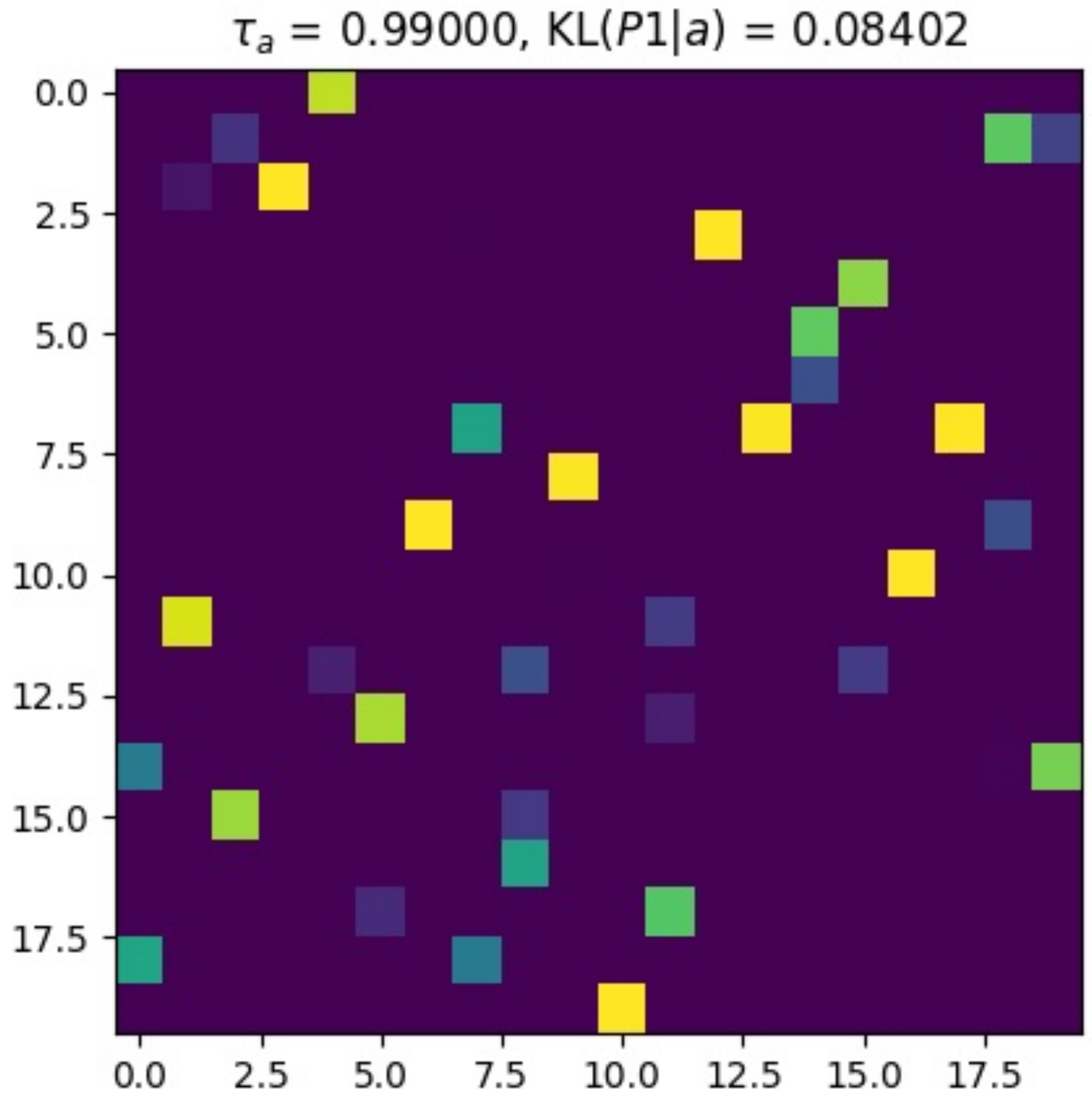
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$

[Colab3]

$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$

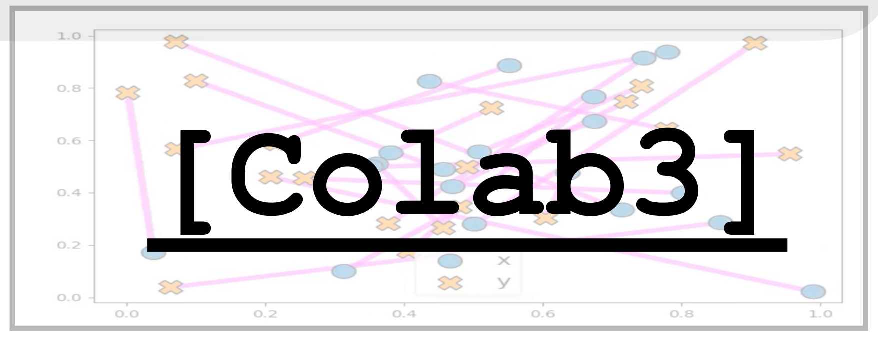


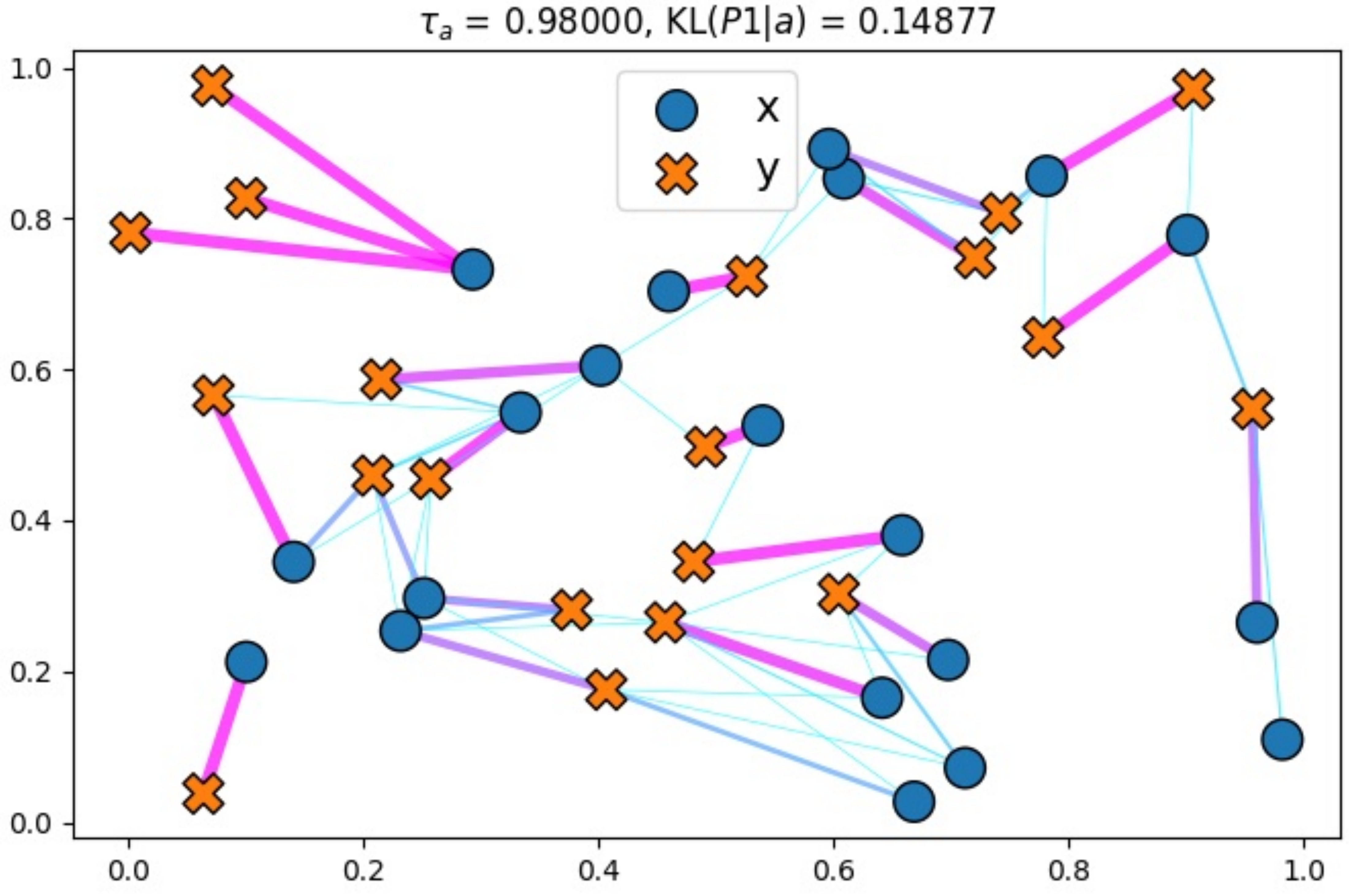


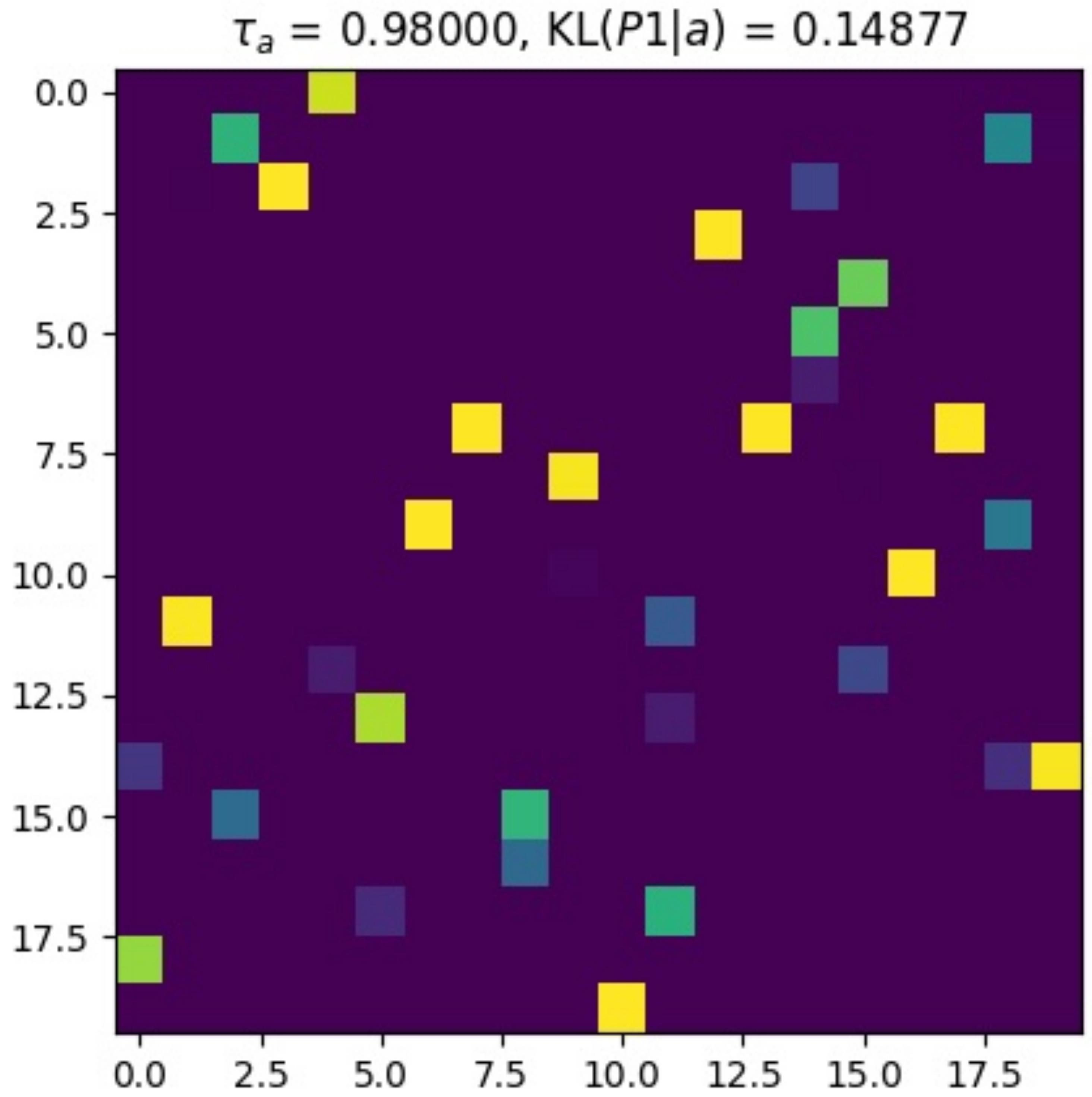
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$

$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



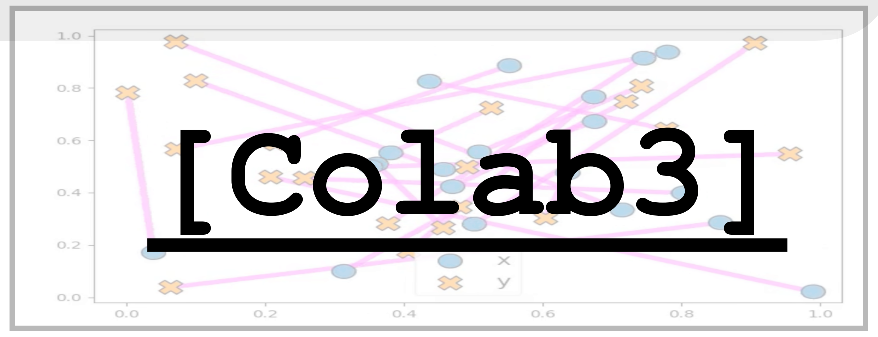


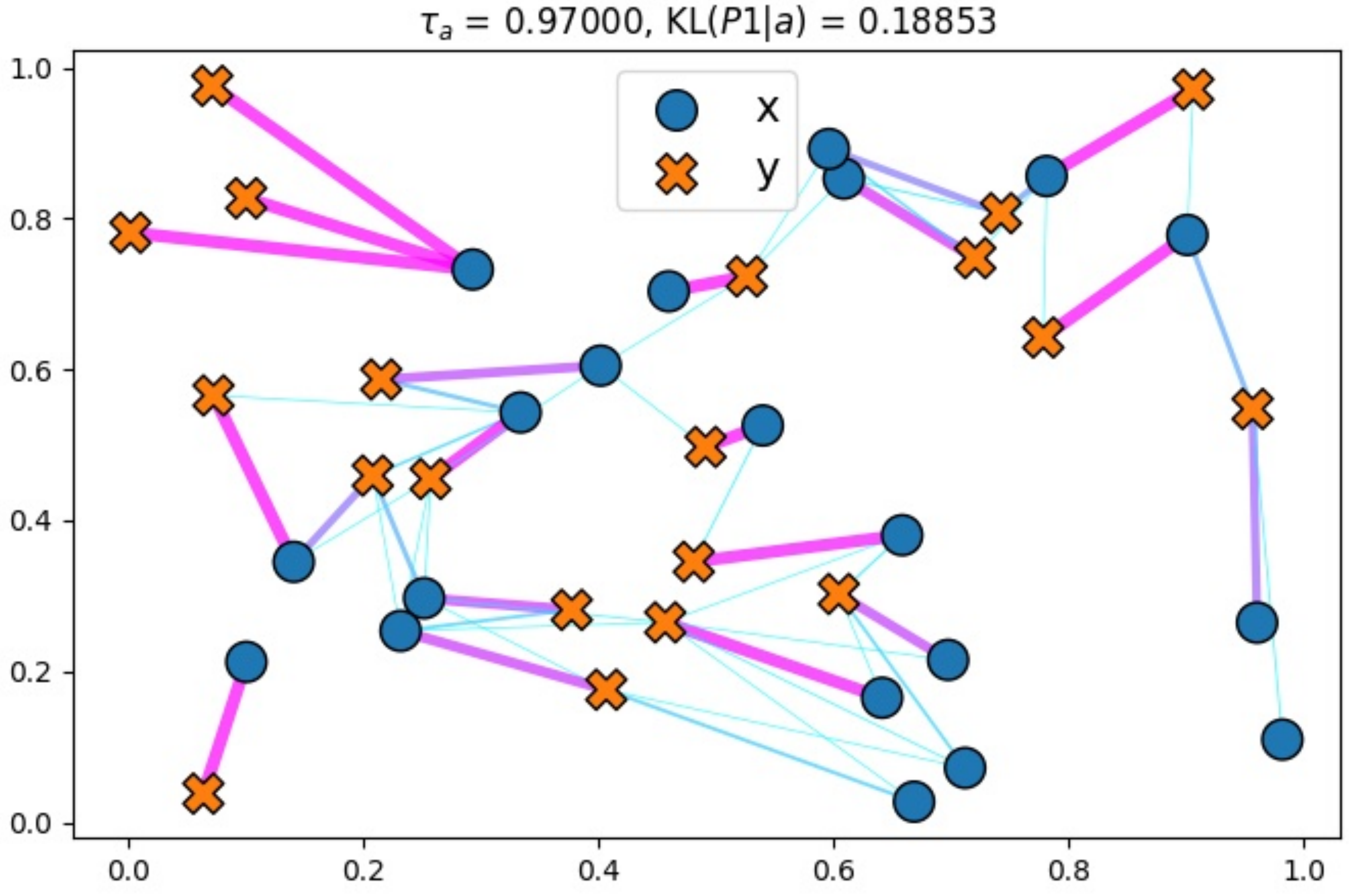


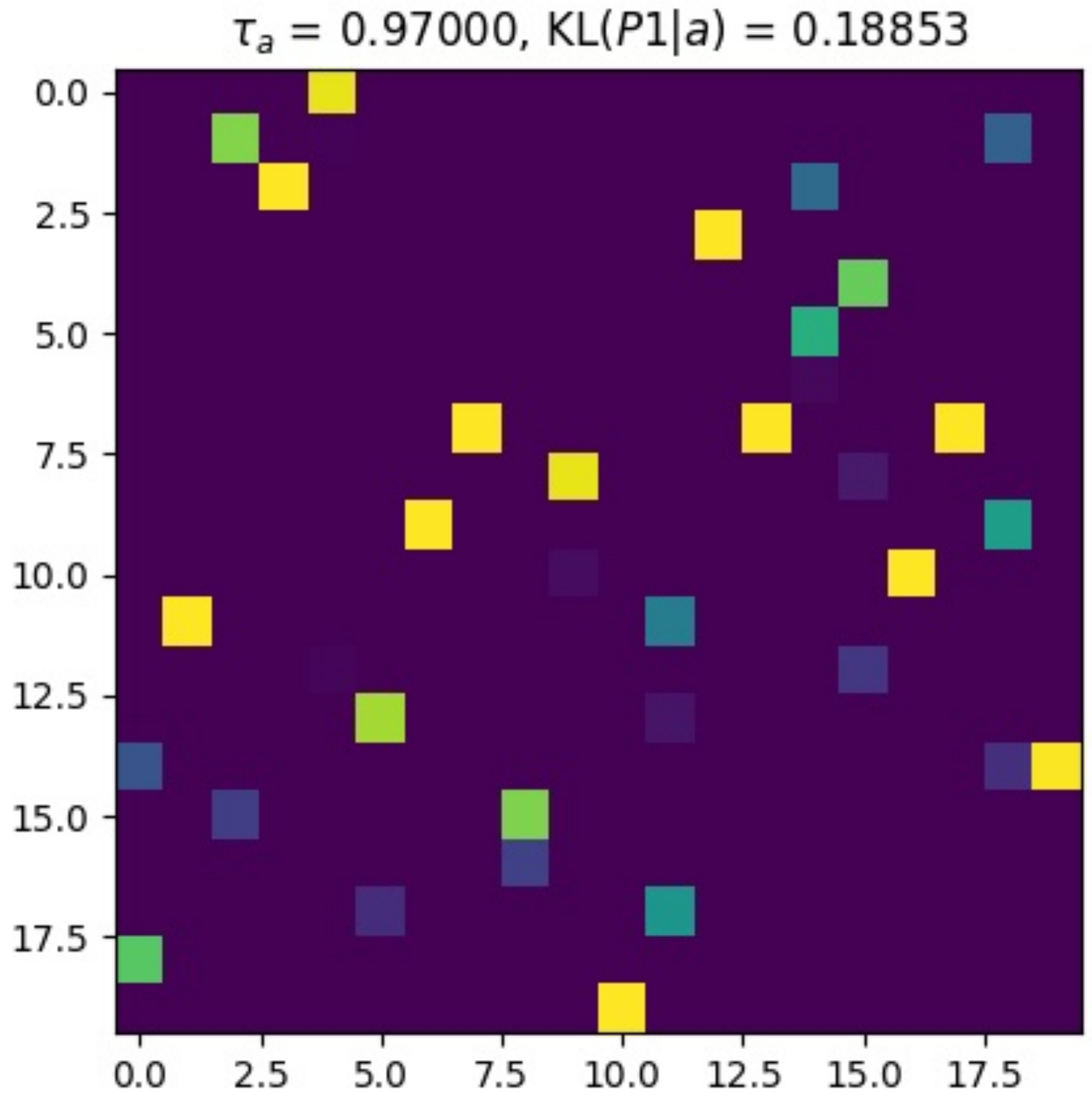
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$

$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$





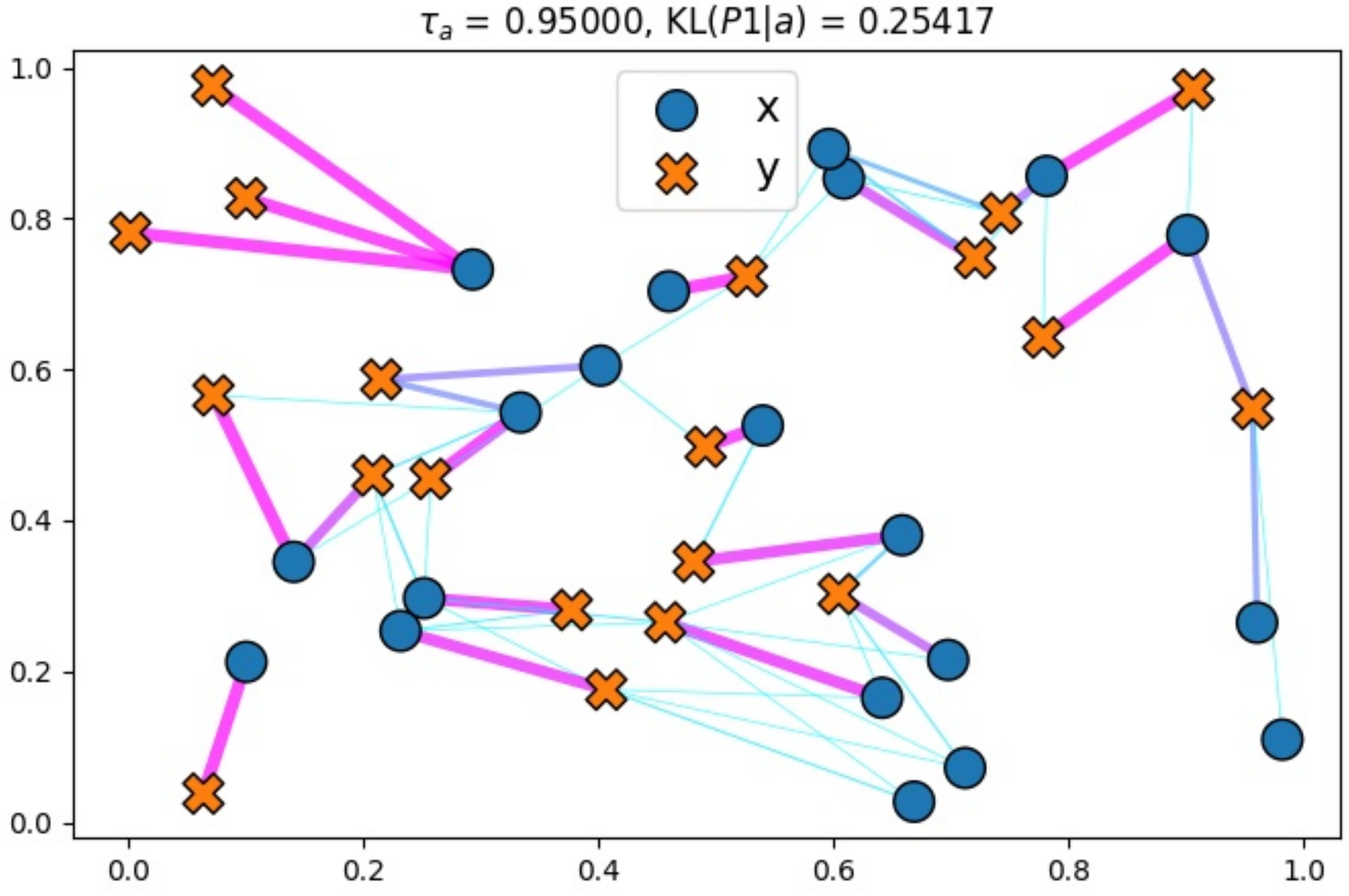


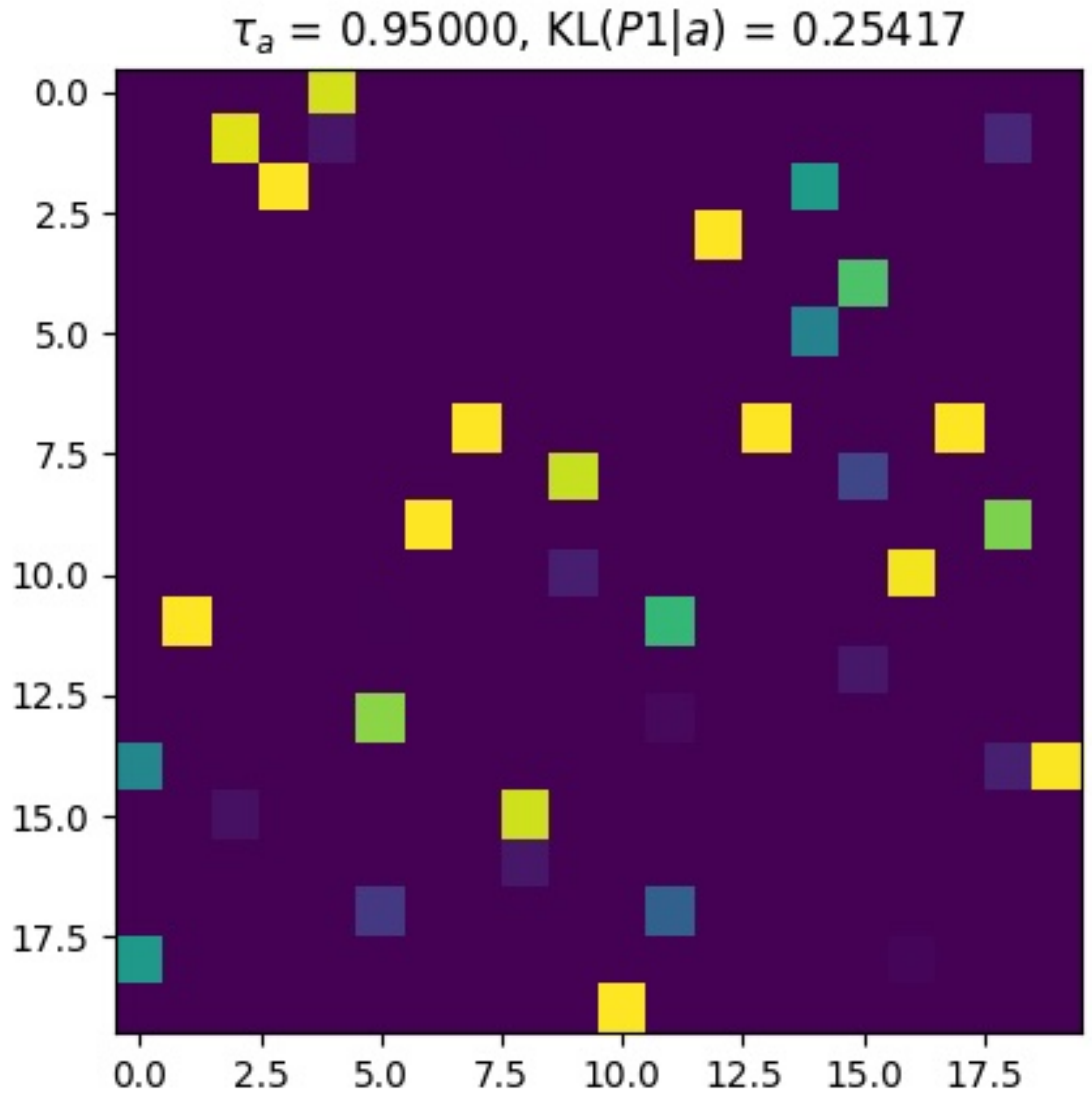
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$

[Colab3]

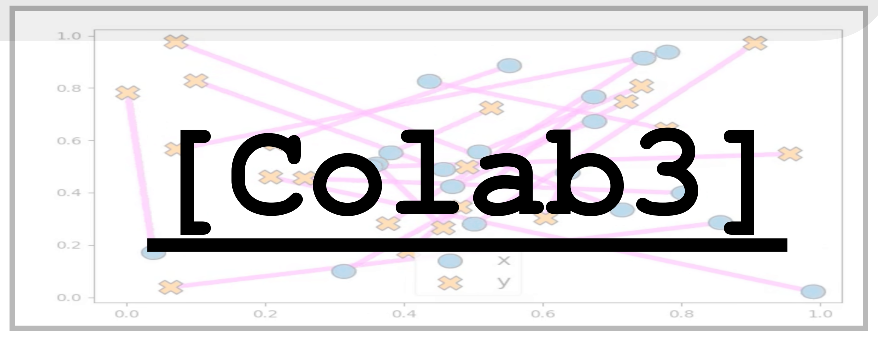
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



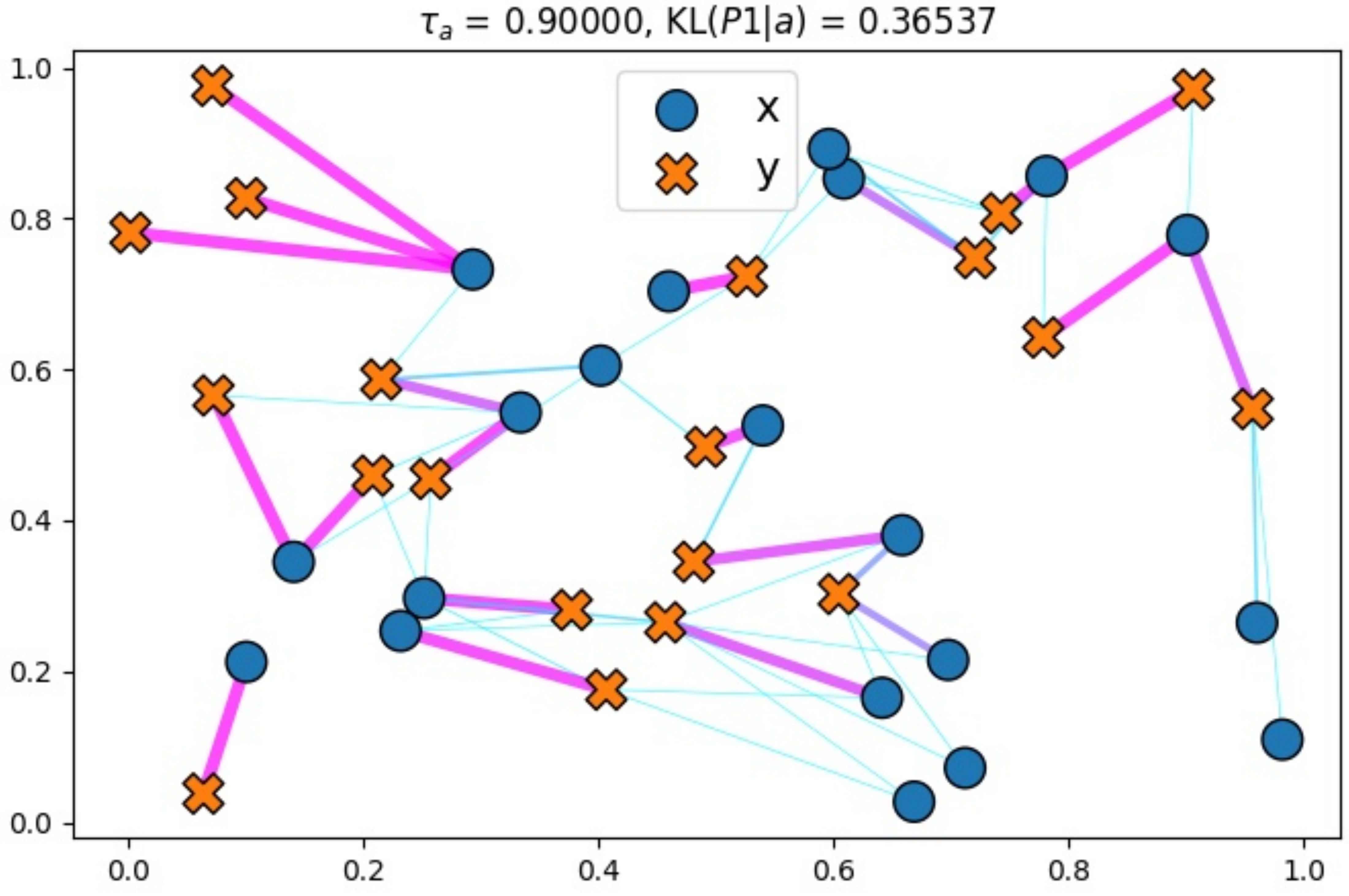


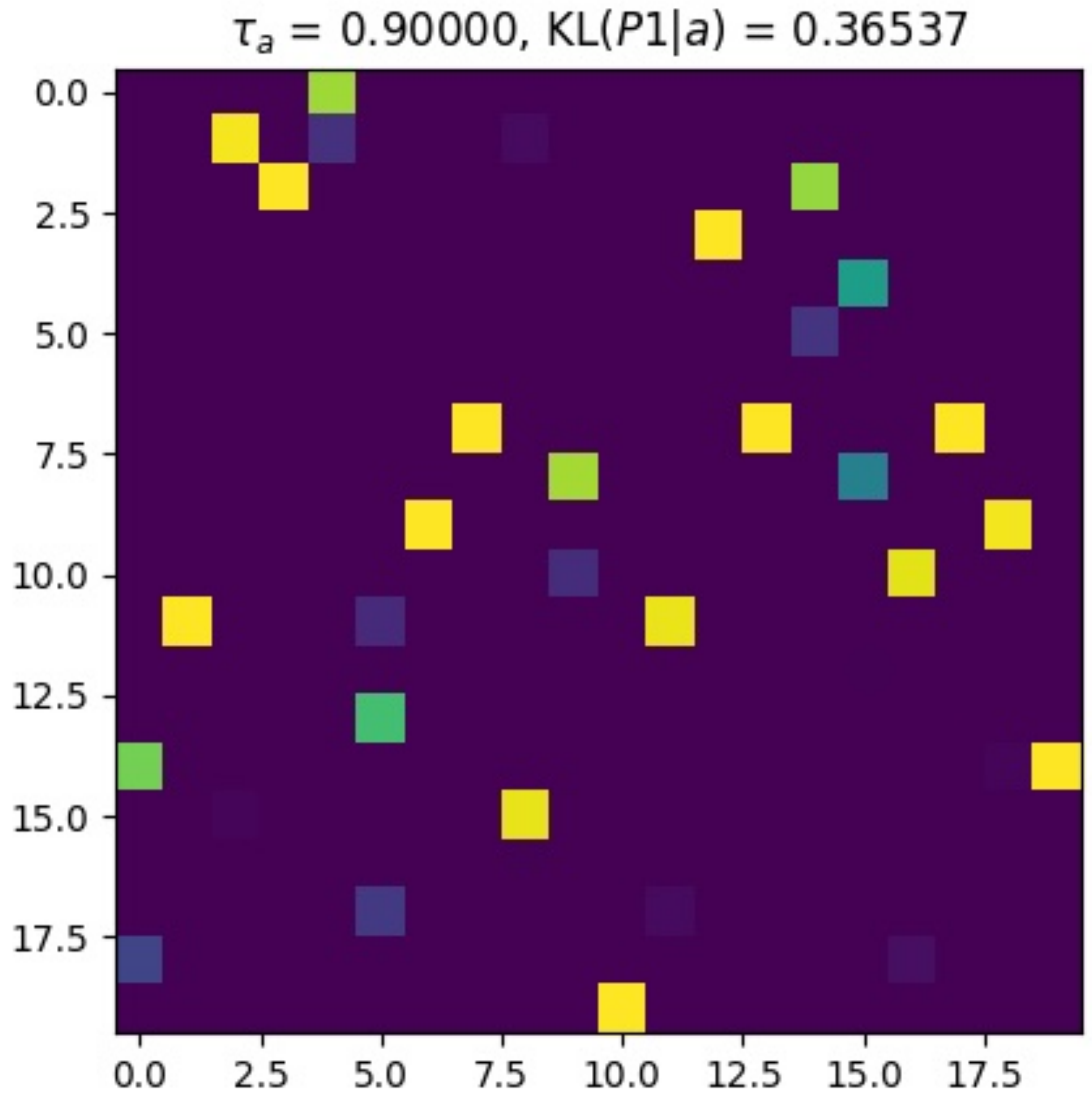
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



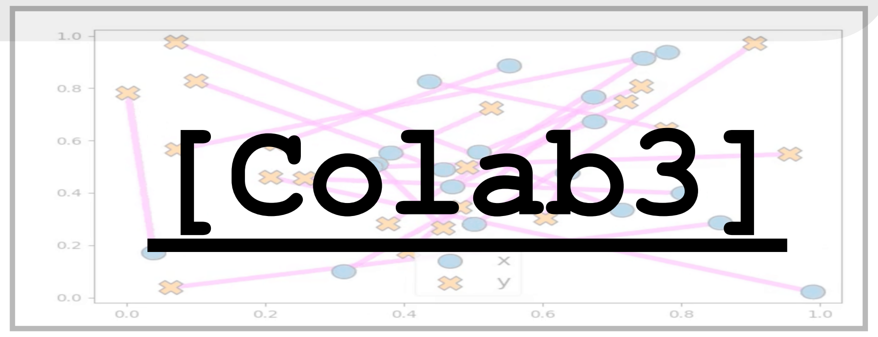
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



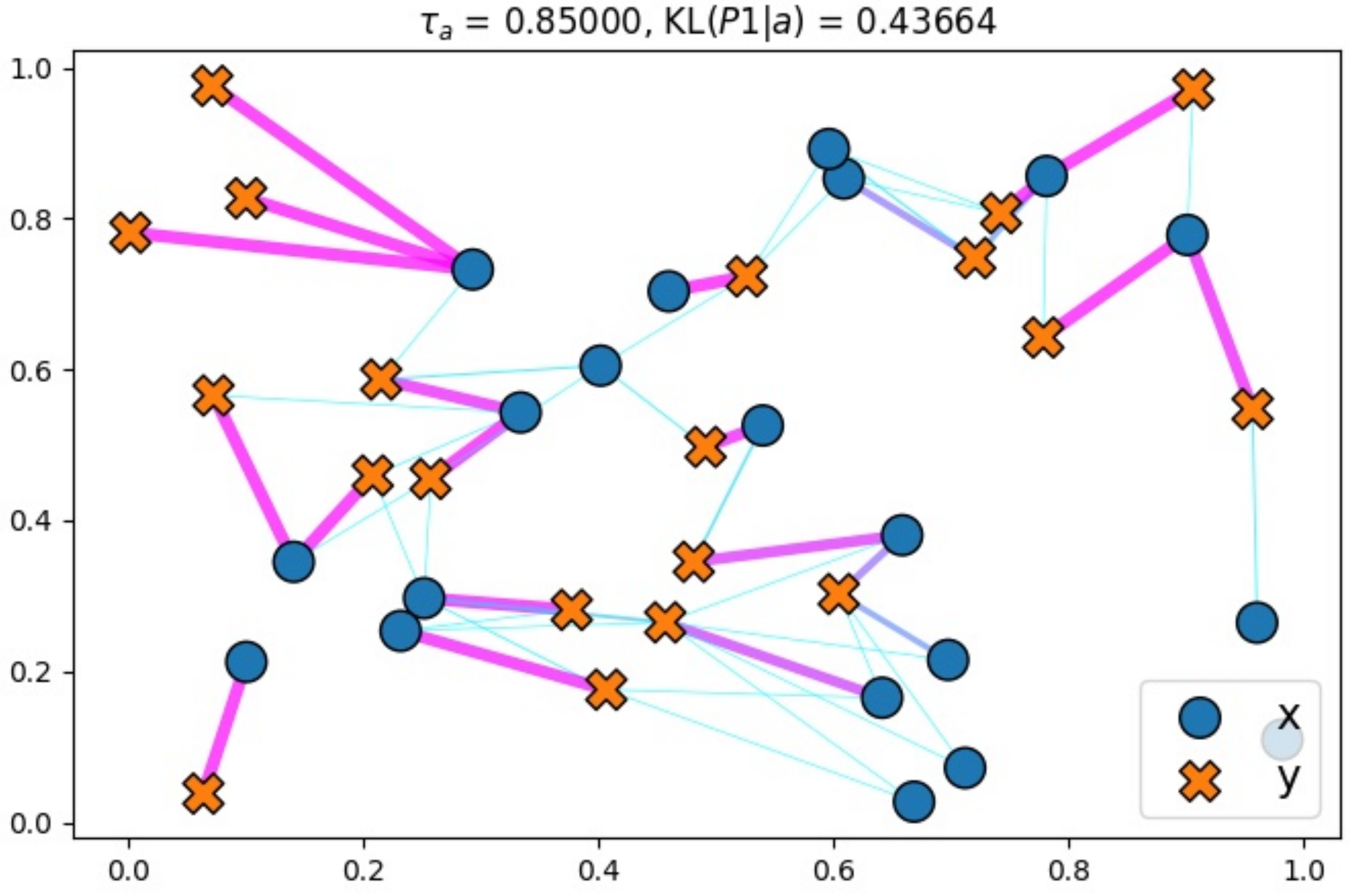


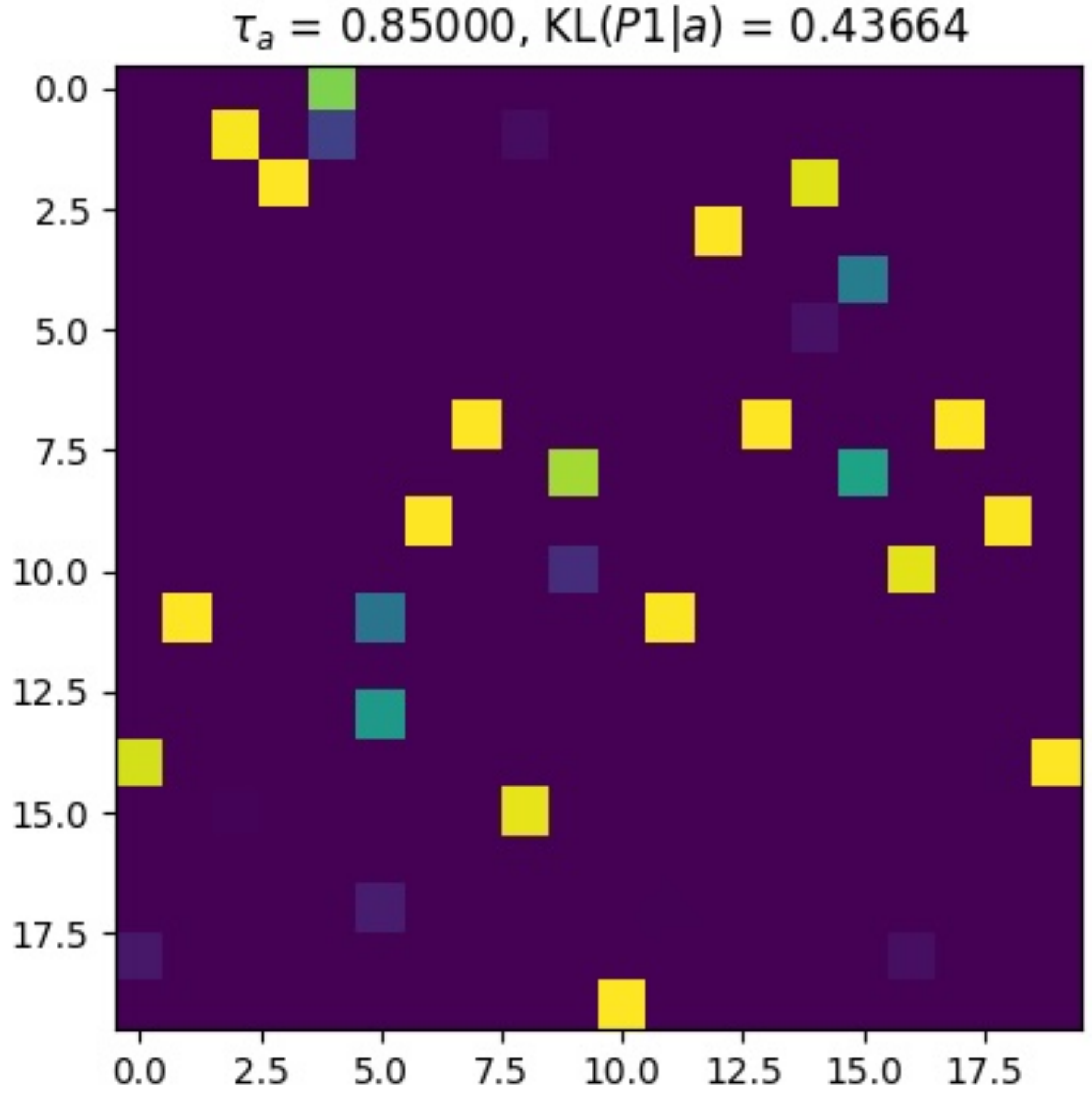
Semi-Balanced Sinkhorn

$$\min_{\substack{P\in\mathbb{R}_+^{n\times m}\\ P^T\mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



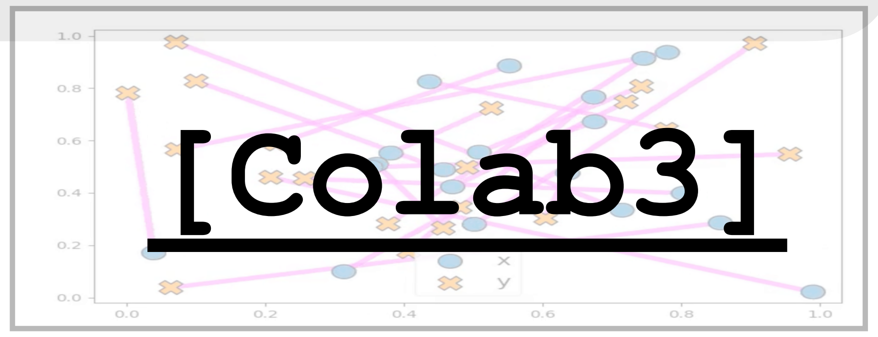
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



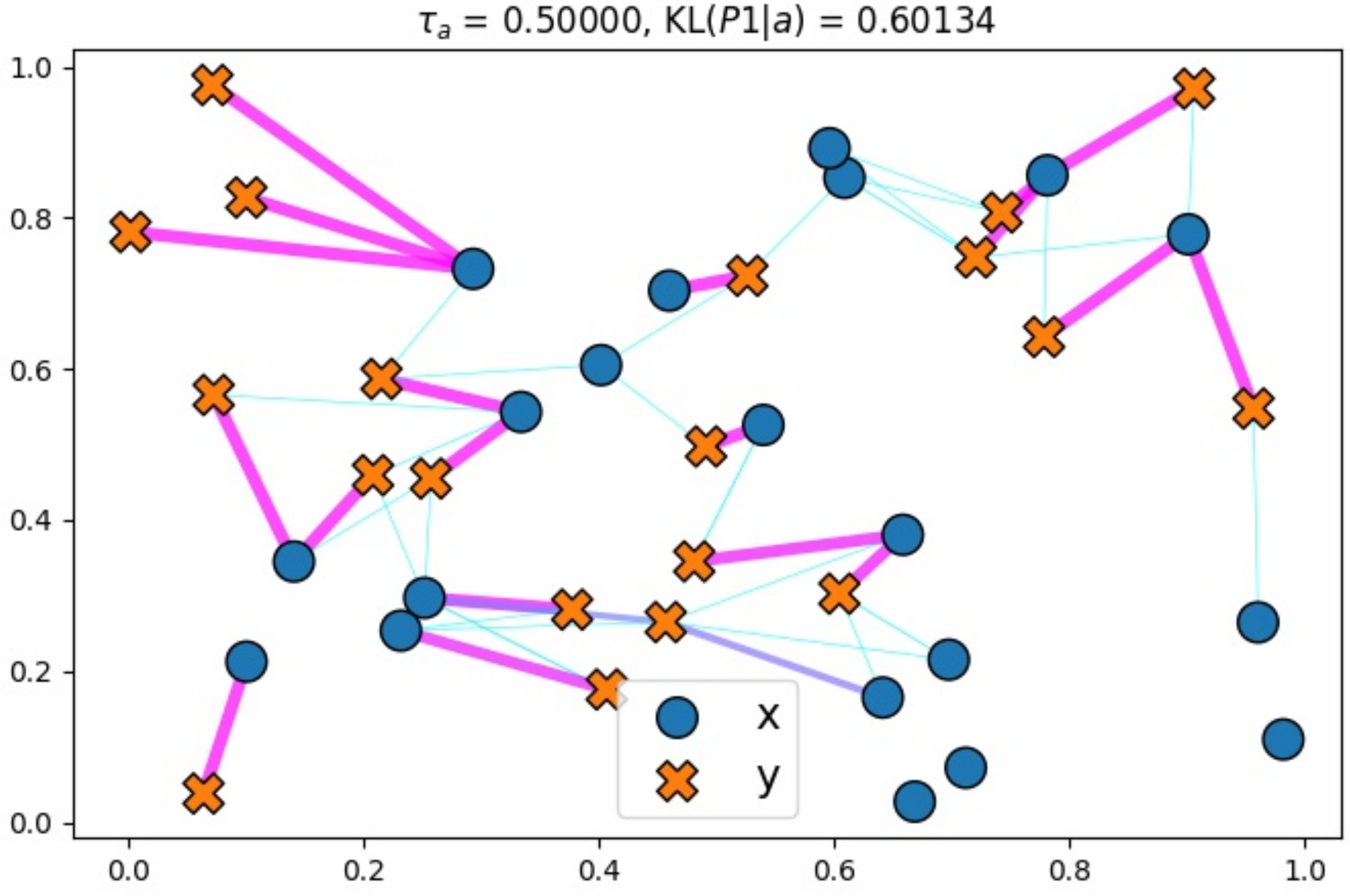


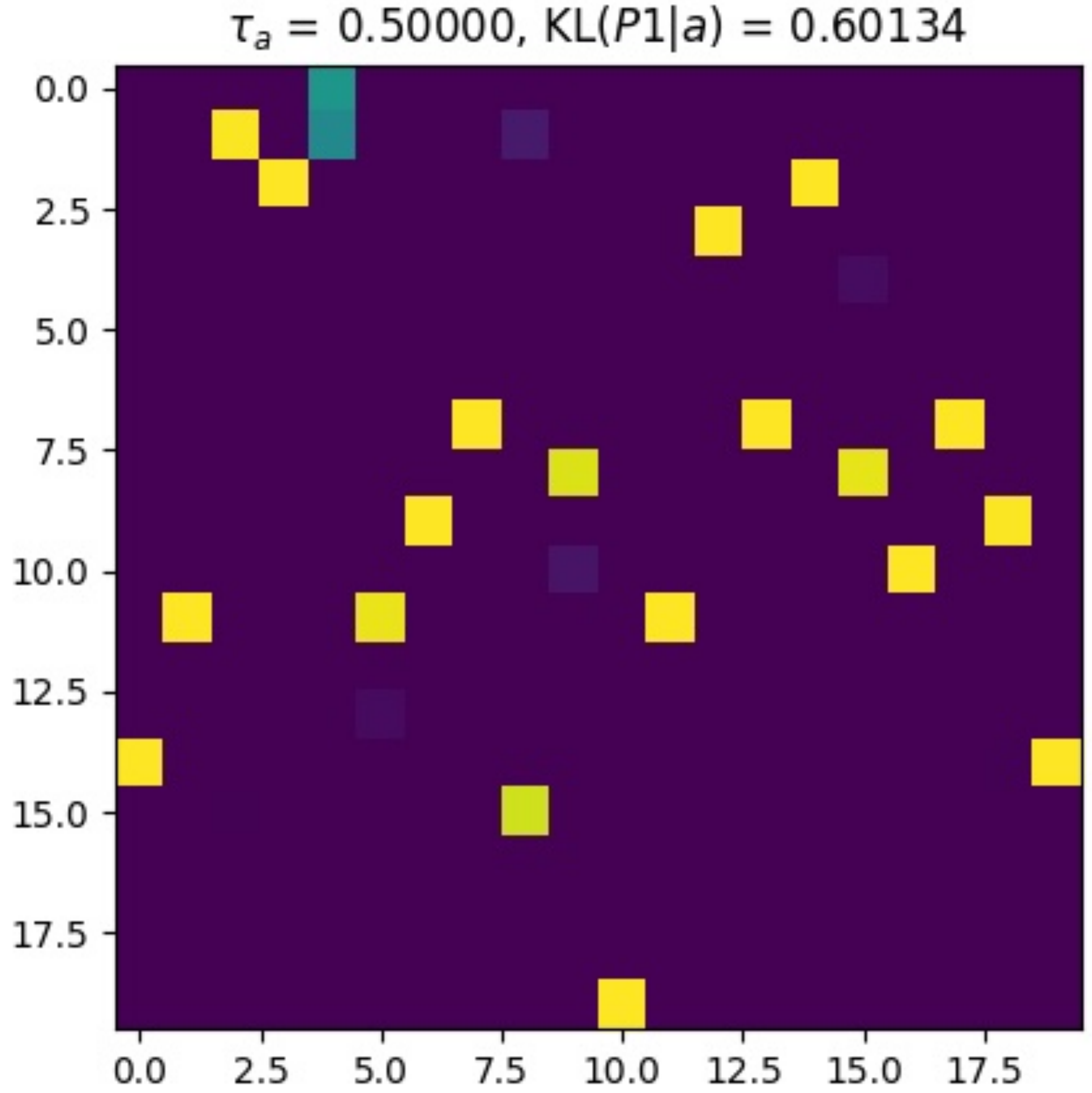
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



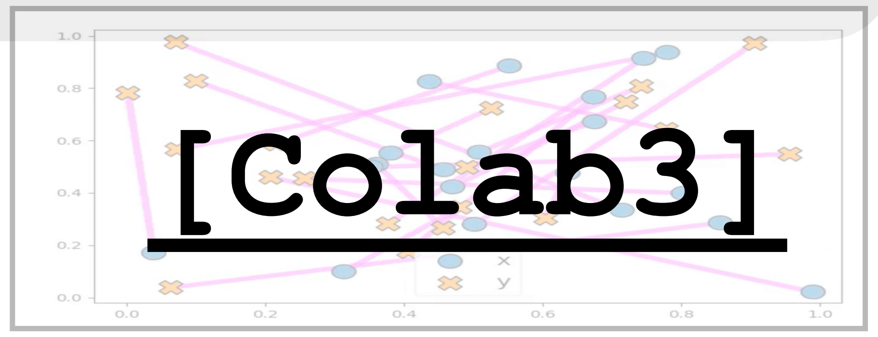
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$



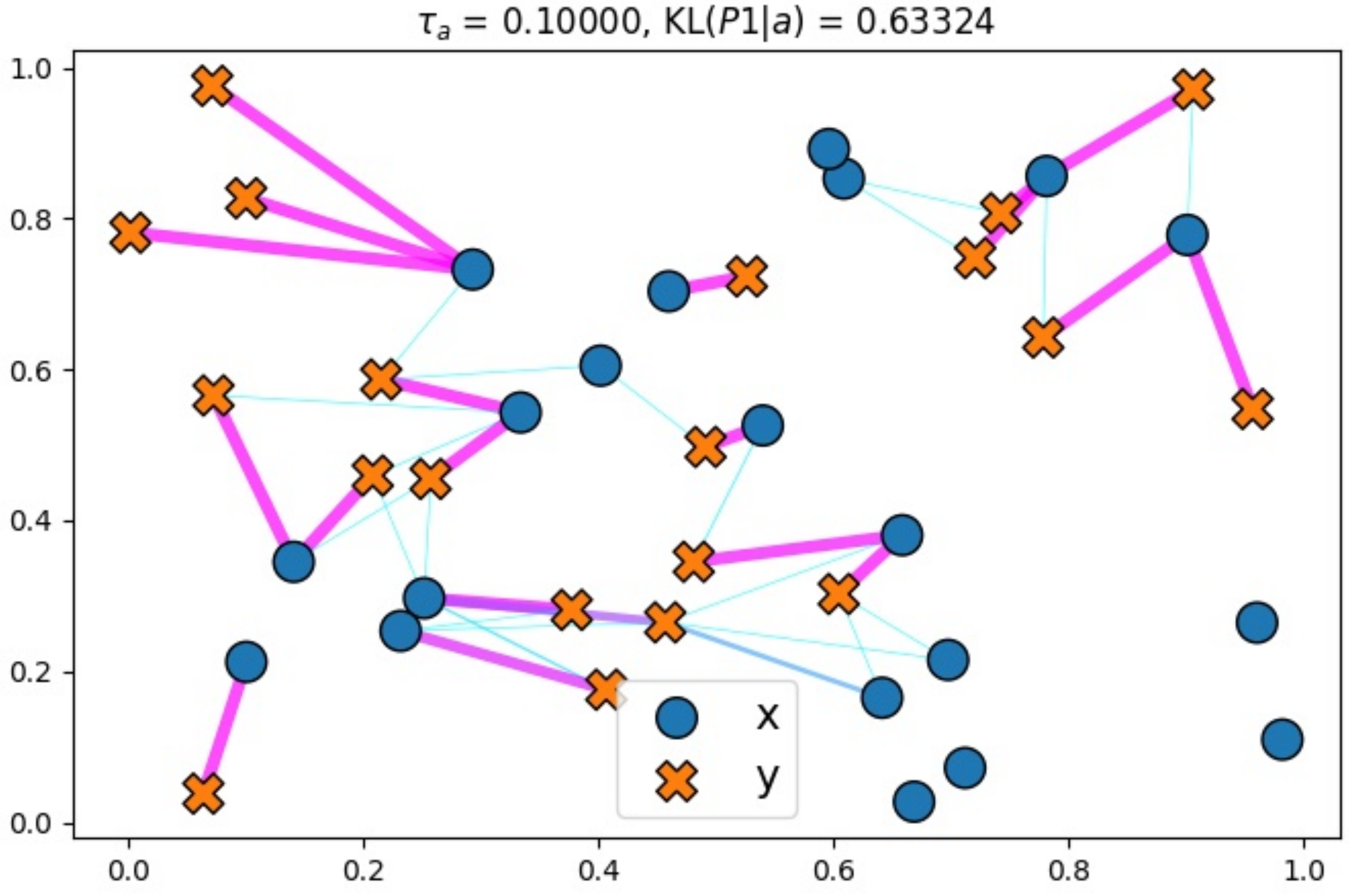


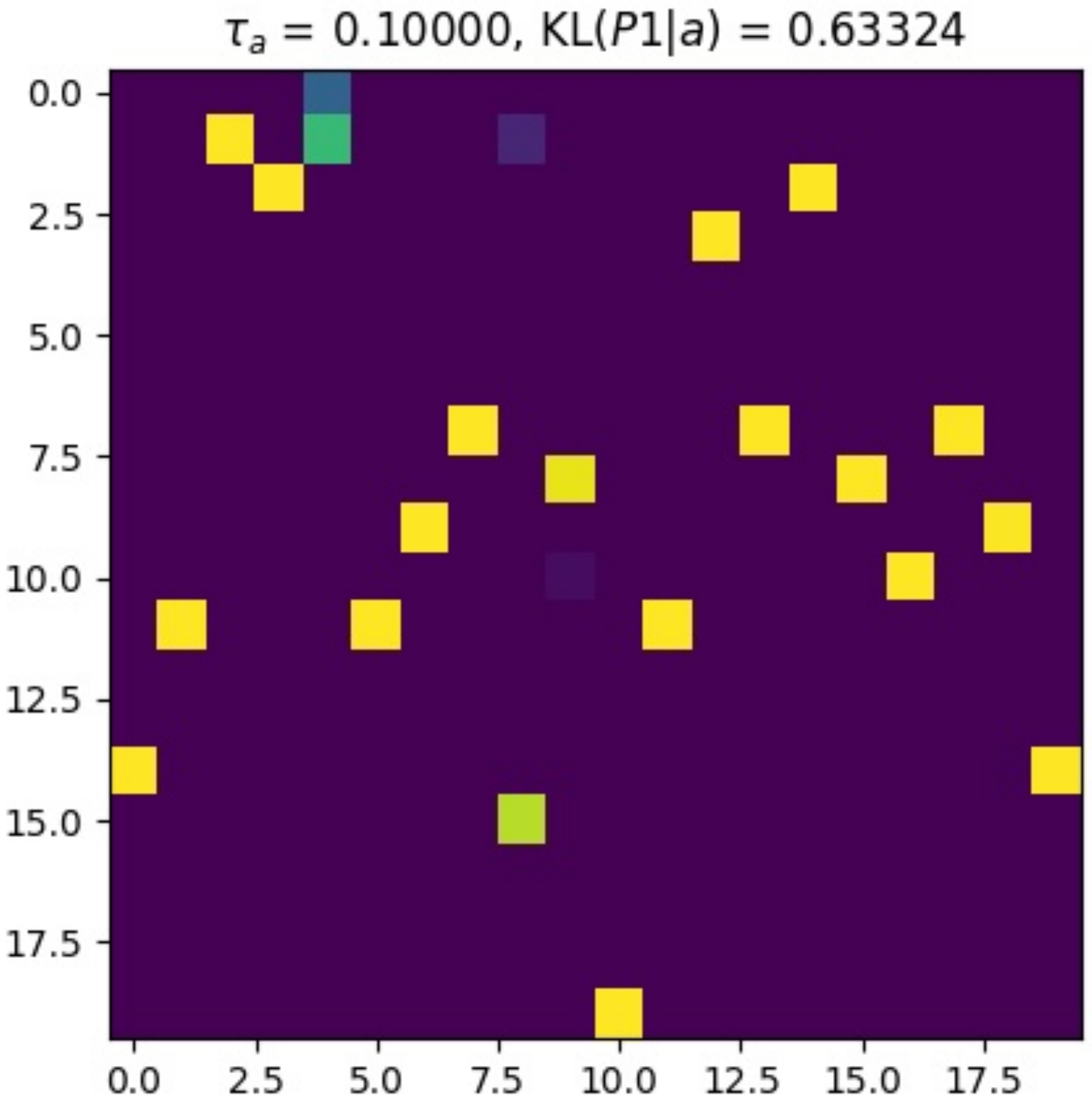
Semi-Balanced Sinkhorn

$$\min_{\substack{P\in\mathbb{R}_+^{n\times m}\\ P^T\mathbf{1}_n=b}} \langle P, M_{XY}\rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



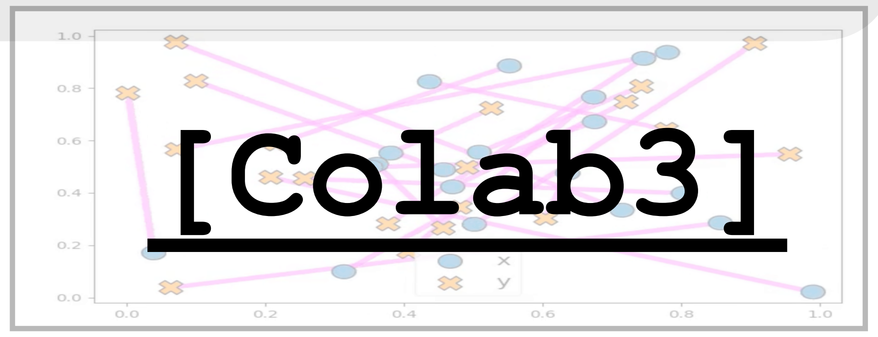
$$\tau_a := \frac{\rho_a}{\rho_a+\gamma}$$



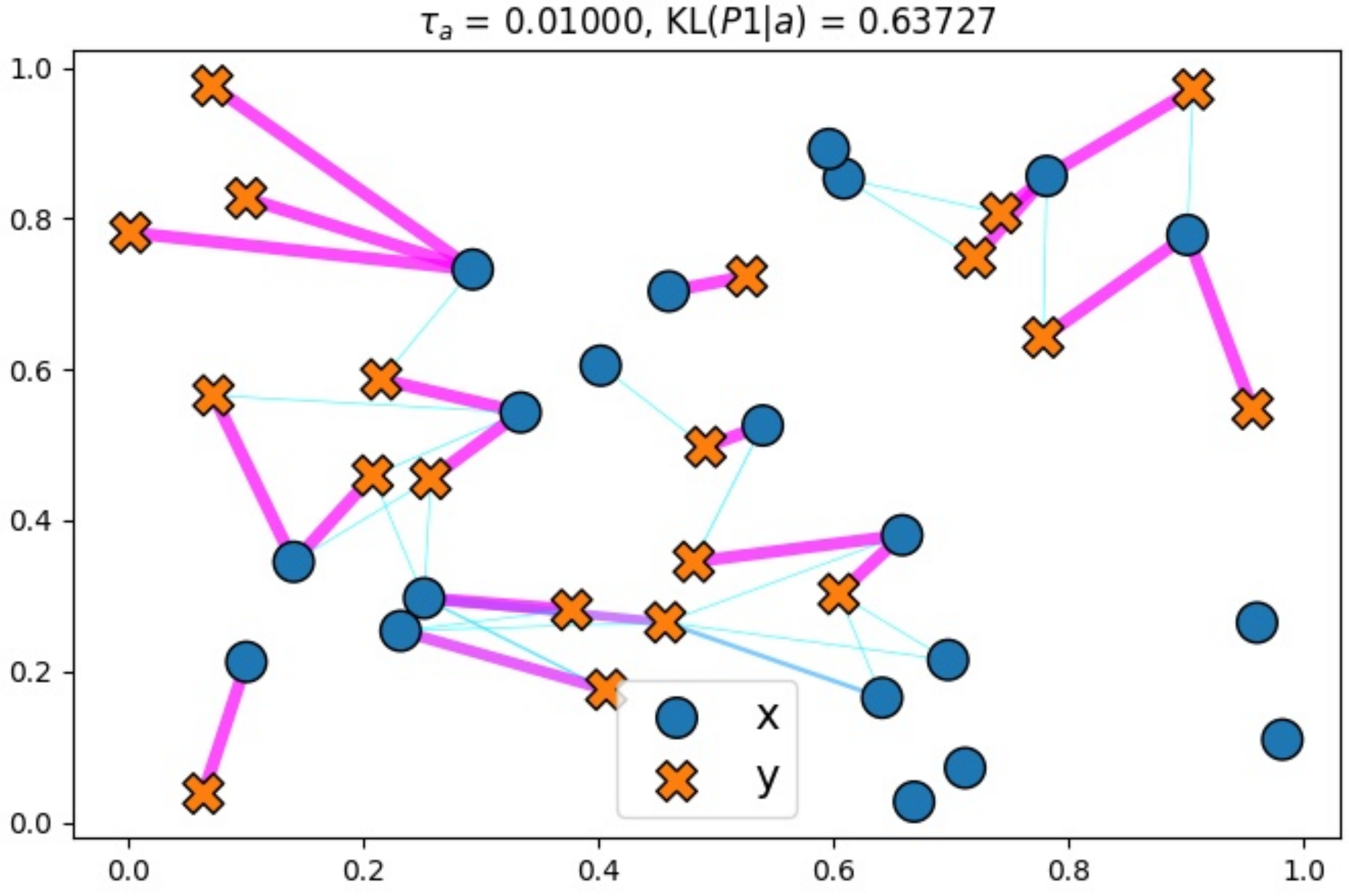


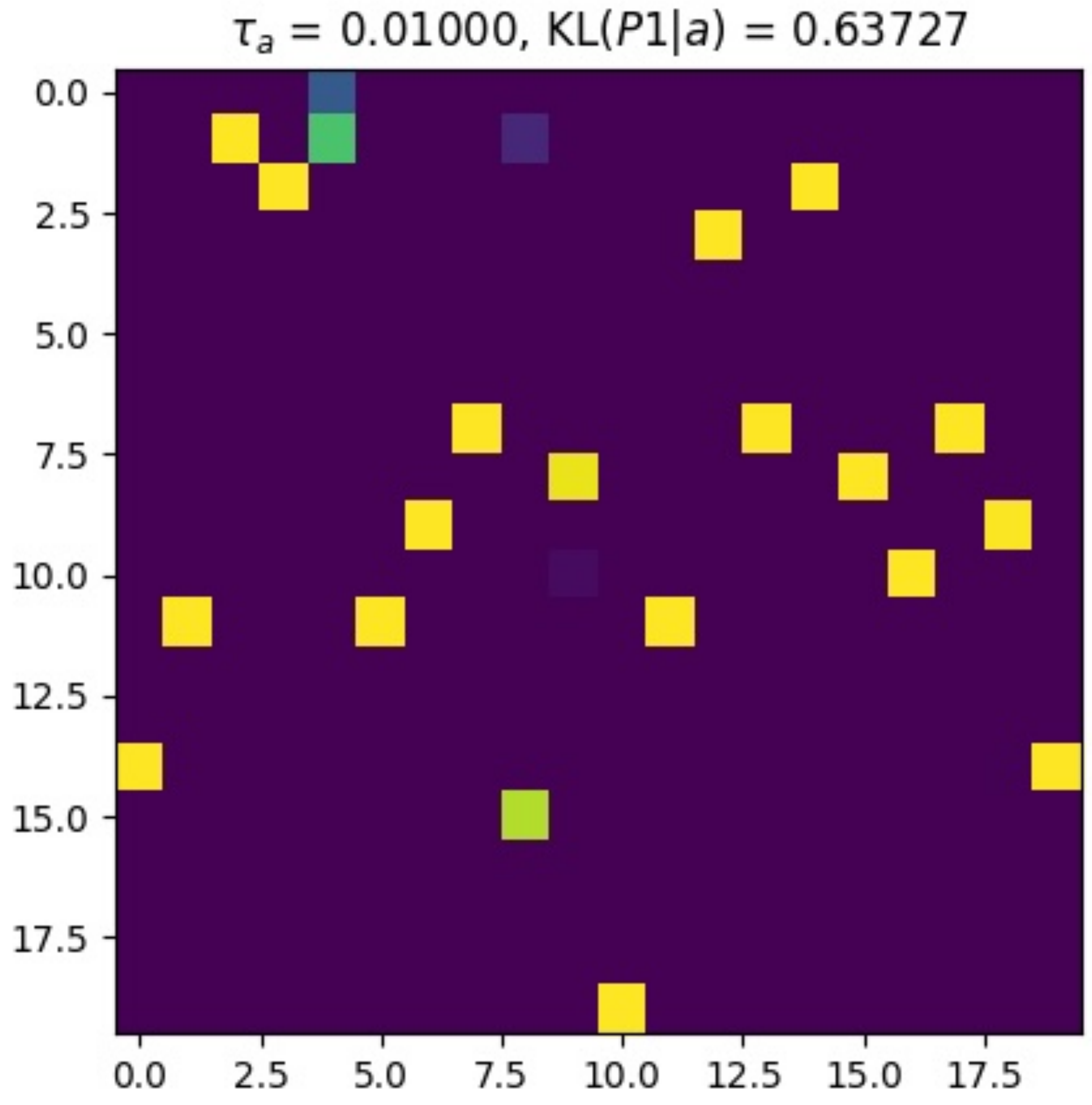
Semi-Balanced Sinkhorn

$$\min_{\substack{P \in \mathbb{R}_+^{n \times m} \\ P^T \mathbf{1}_n = b}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) - \gamma E(P)$$



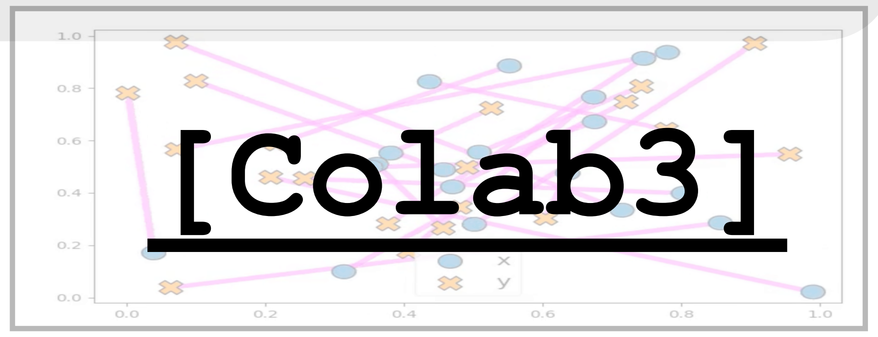
$$\tau_a := \frac{\rho_a}{\rho_a + \gamma}$$





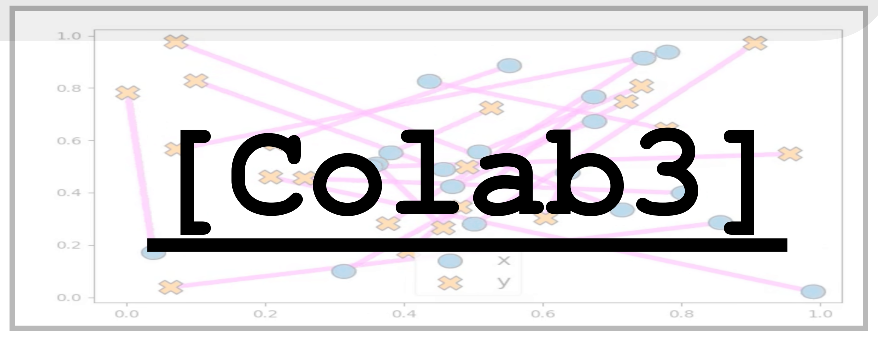
Unbalanced Sinkhorn

$$\min_{\boldsymbol{P}\in\mathbb{R}_+^{\boldsymbol{n}\times\boldsymbol{m}}} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle + \rho_{\boldsymbol{a}}\mathrm{KL}(\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}} \| \boldsymbol{a}) + \rho_{\boldsymbol{b}}\mathrm{KL}(\boldsymbol{P}^T\mathbf{1}_{\boldsymbol{n}} \| \boldsymbol{b}) - \gamma E(\boldsymbol{P})$$

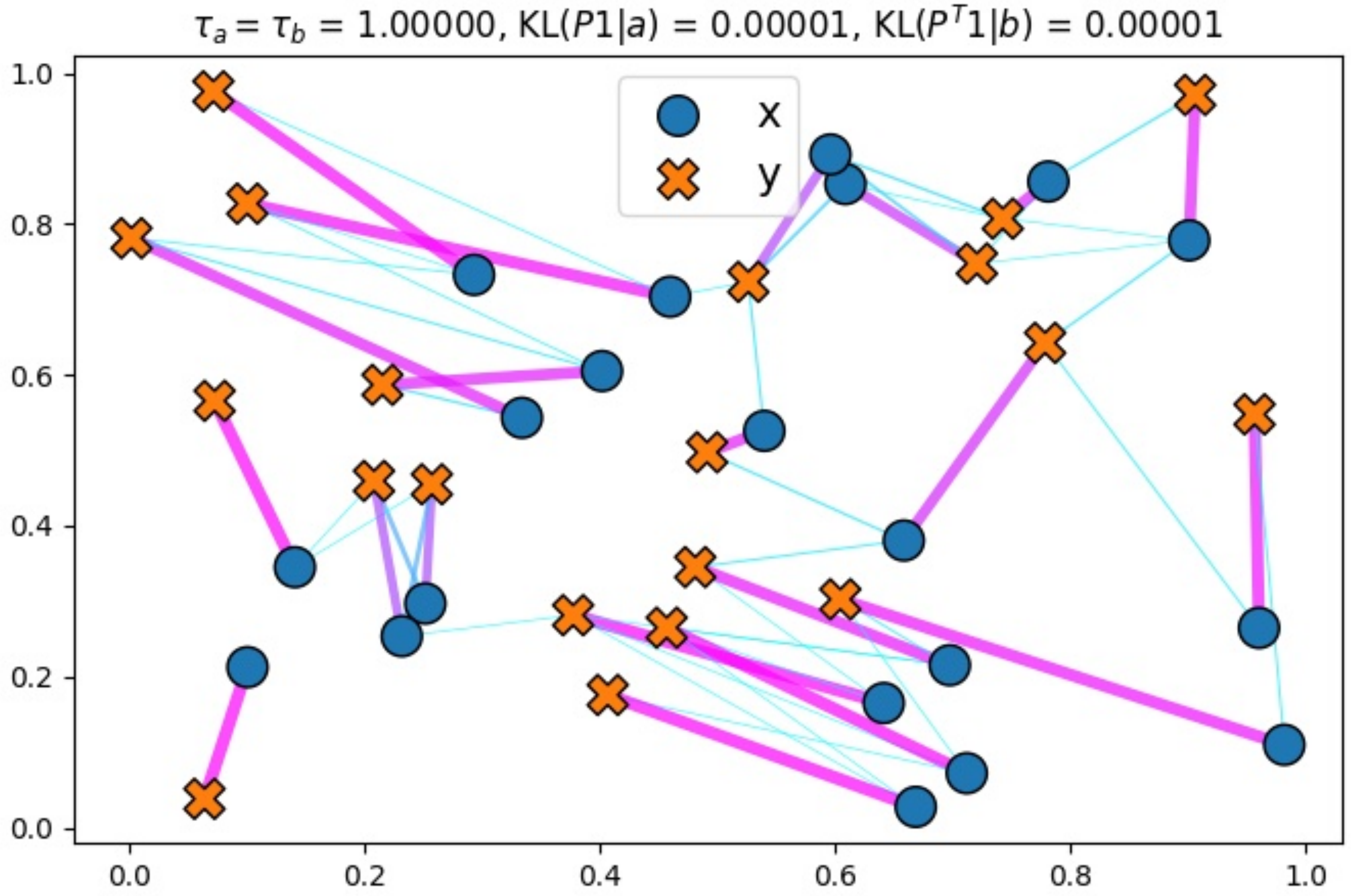


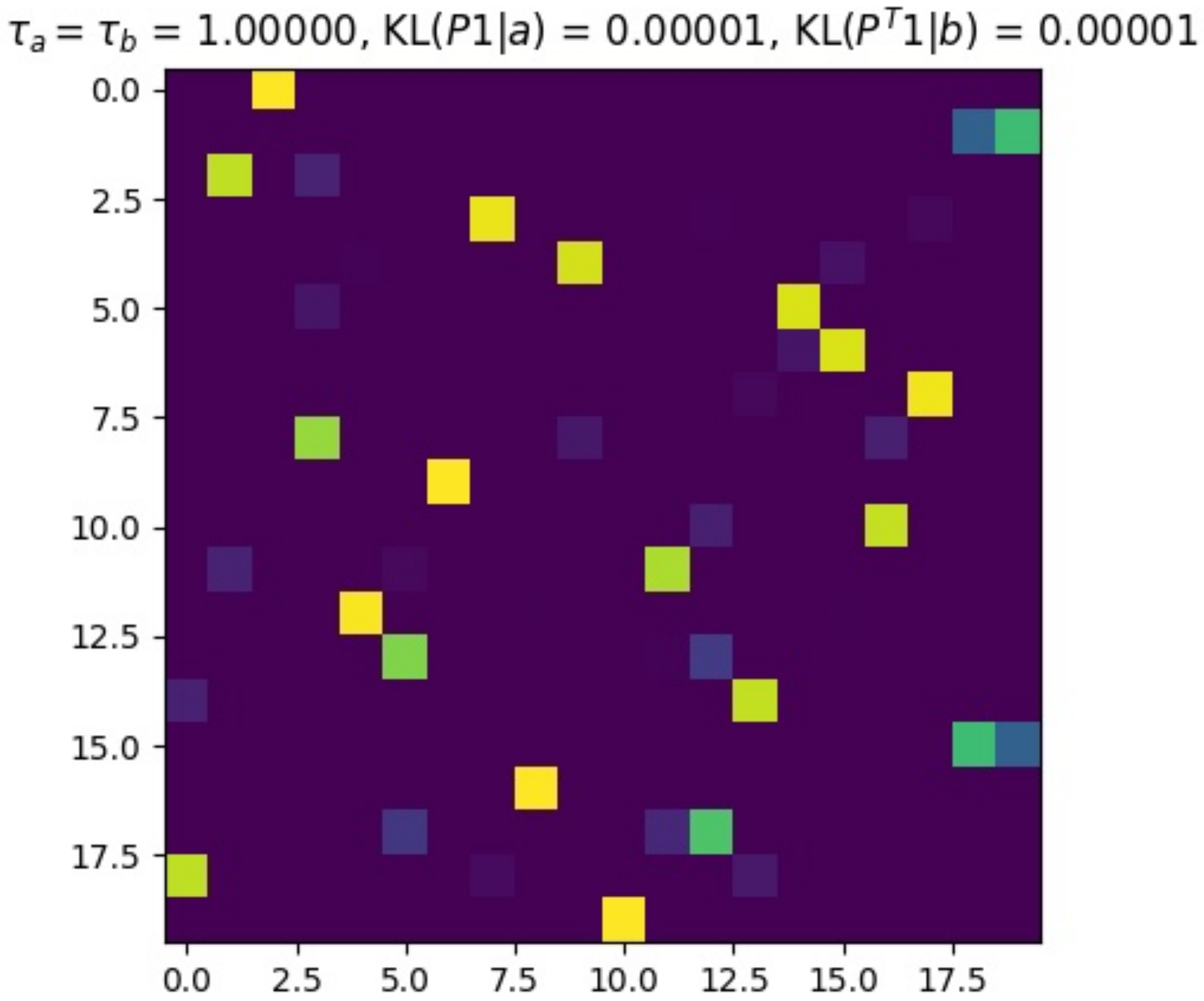
Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$



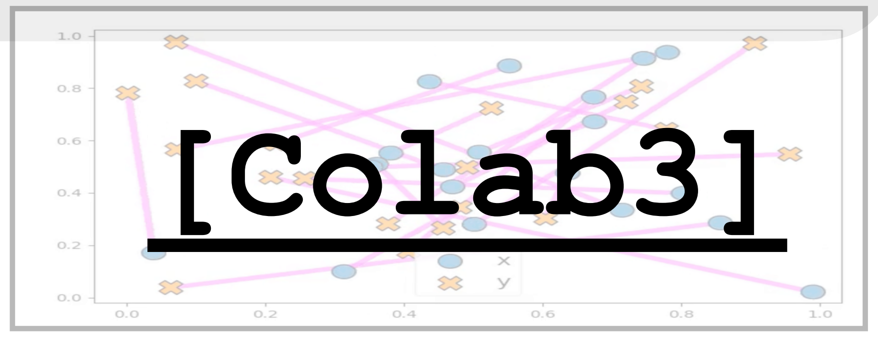
[Colab3]

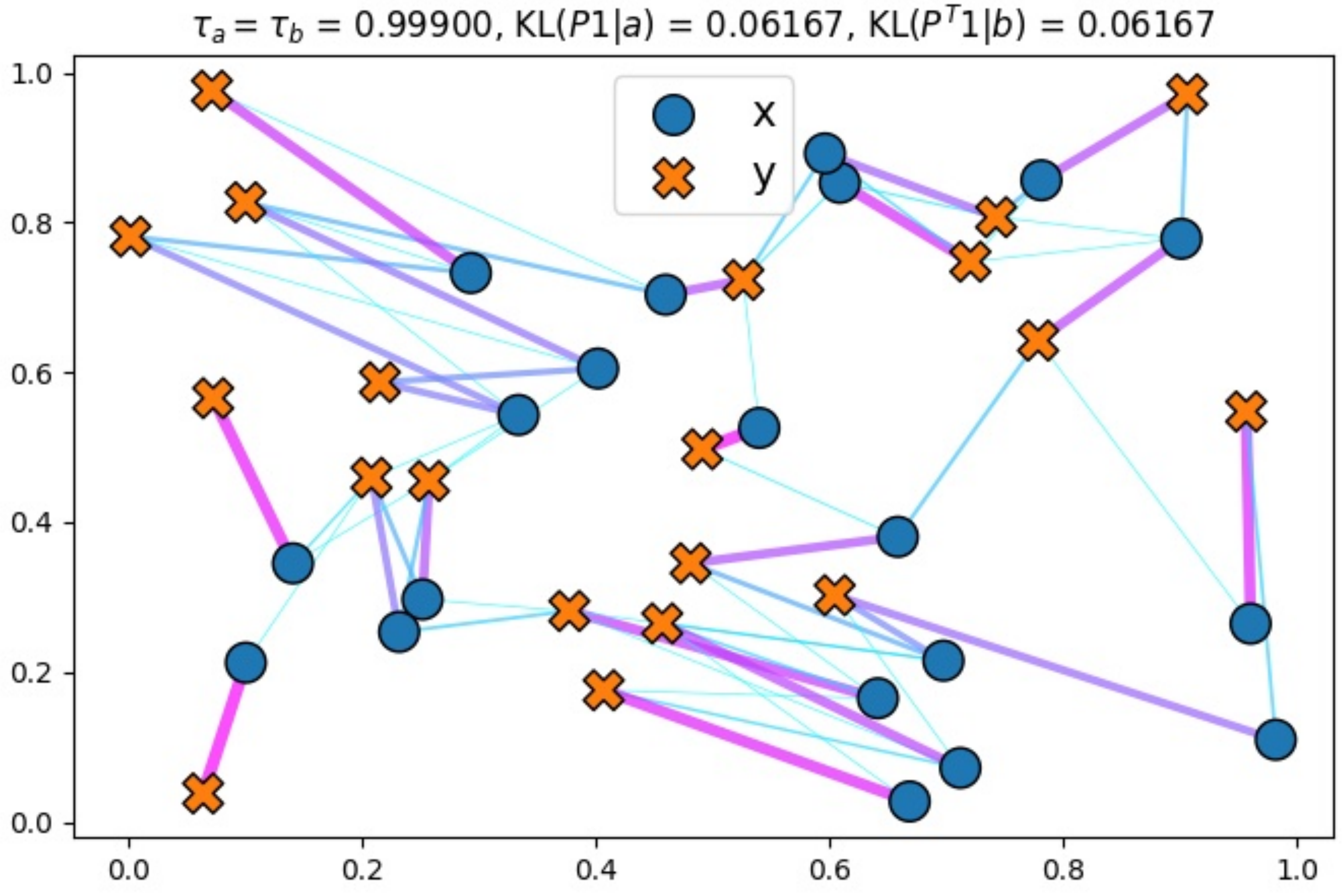


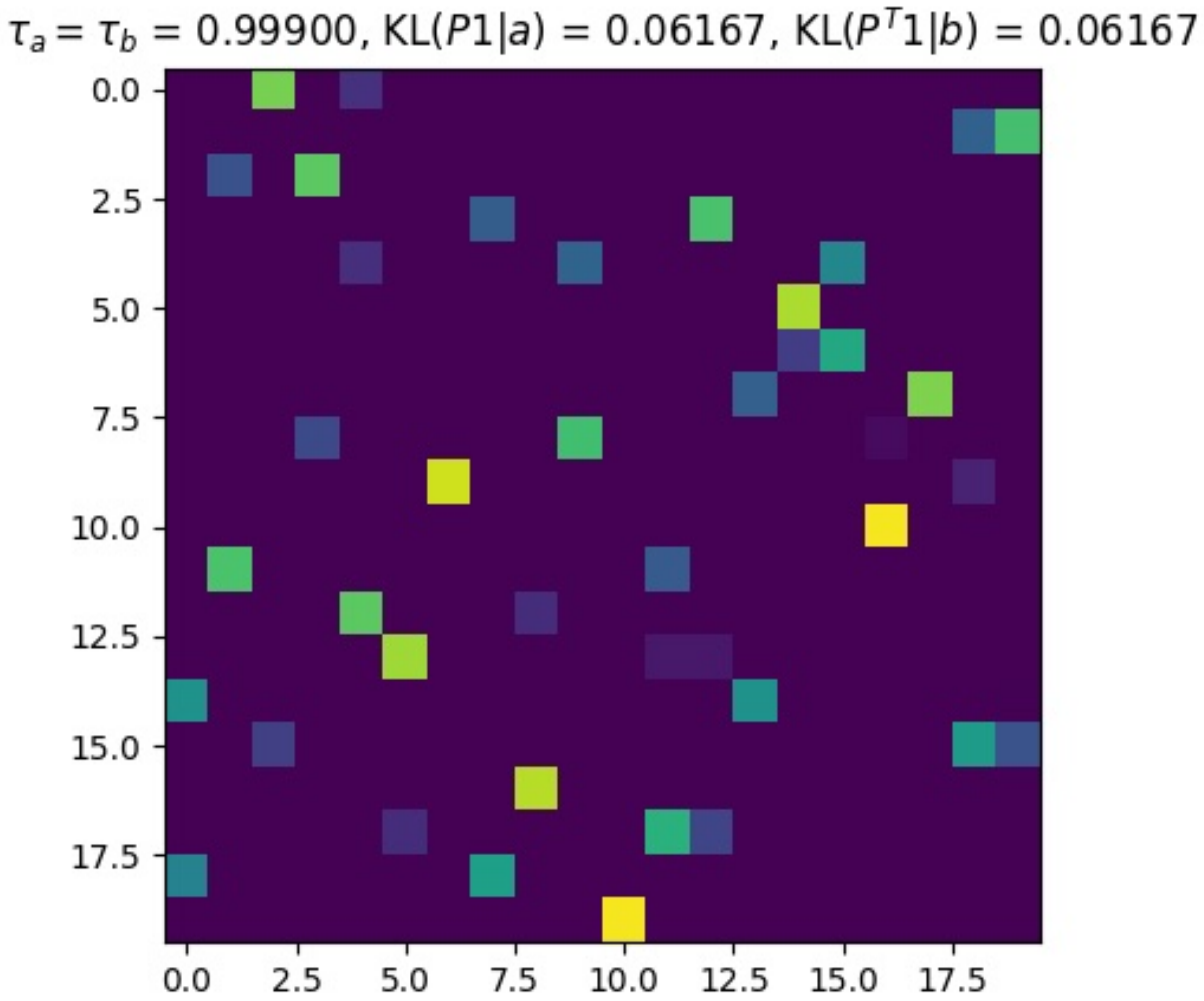


Unbalanced Sinkhorn

$$\min_{\boldsymbol{P}\in\mathbb{R}_+^{\boldsymbol{n}\times \boldsymbol{m}}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}}\rangle + \rho_{\boldsymbol{a}} \mathrm{KL}(\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}}\|\boldsymbol{a}) + \rho_{\boldsymbol{b}} \mathrm{KL}(\boldsymbol{P}^T\mathbf{1}_{\boldsymbol{n}}\|\boldsymbol{b}) - \gamma E(\boldsymbol{P})$$

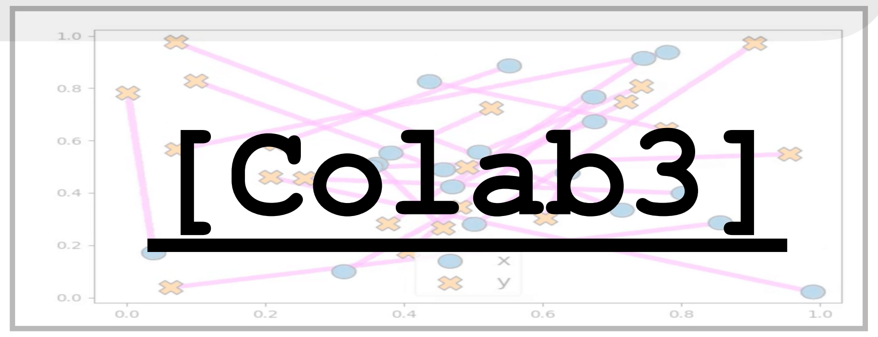


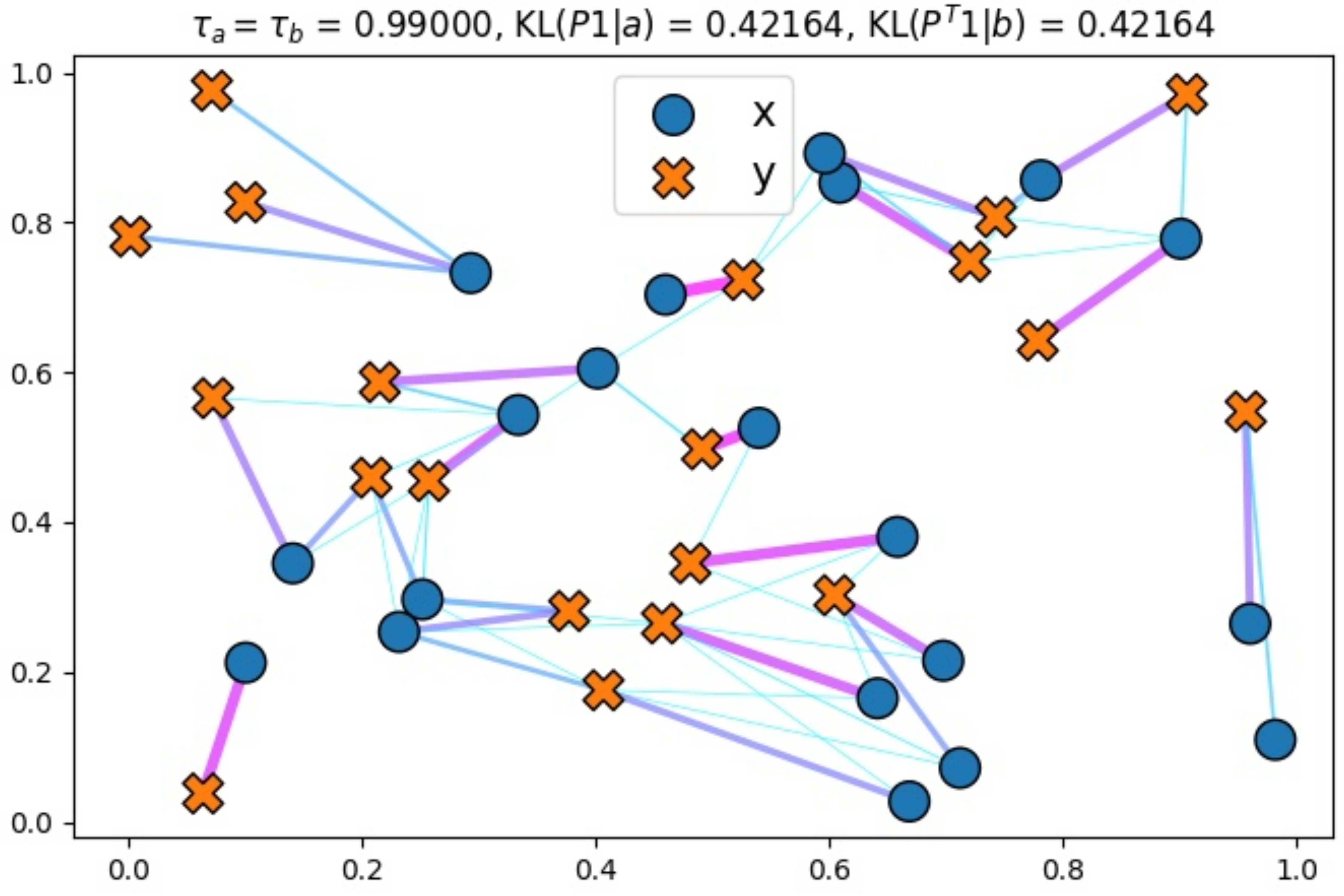


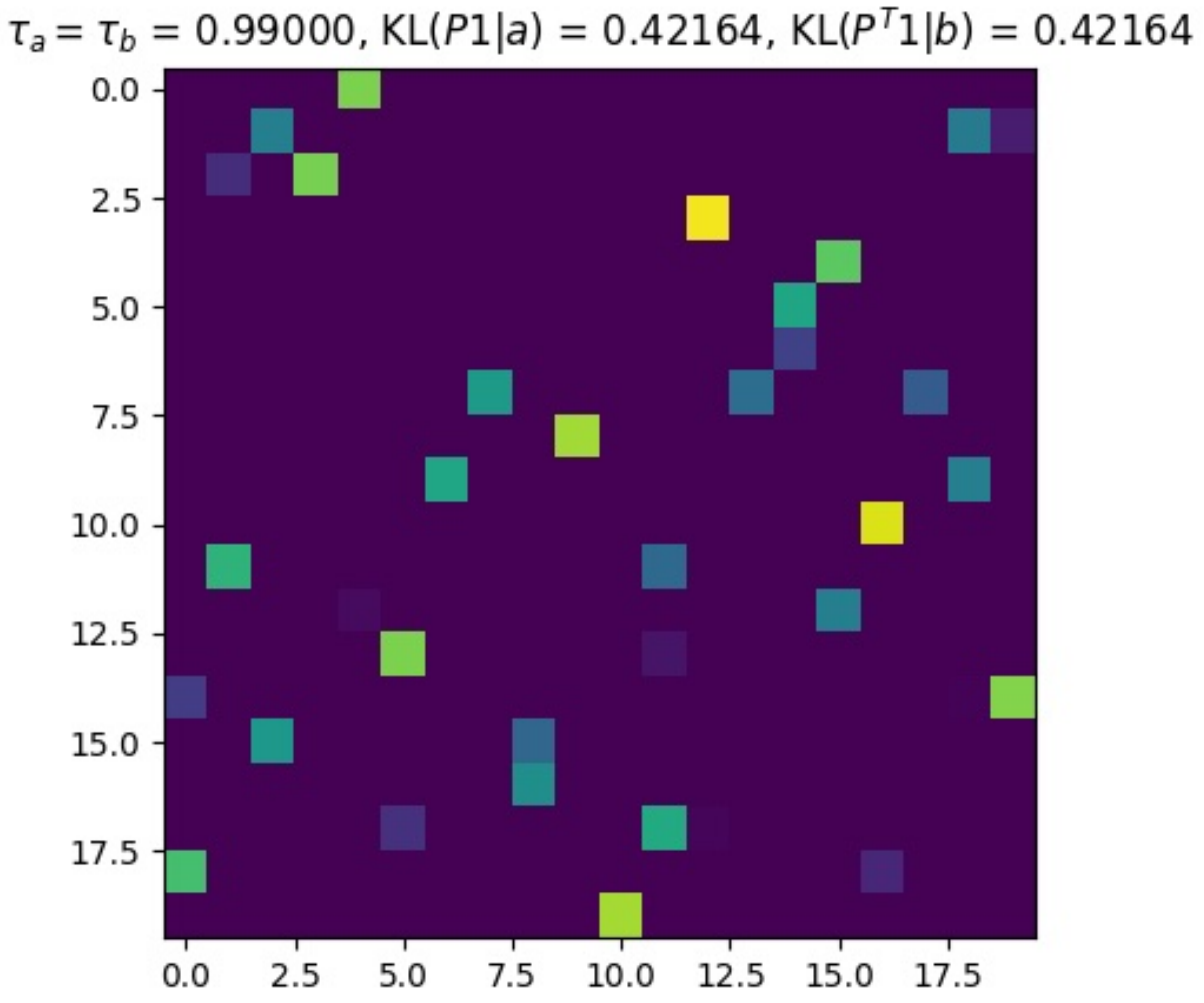


Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T\mathbf{1}_n \| b) - \gamma E(P)$$

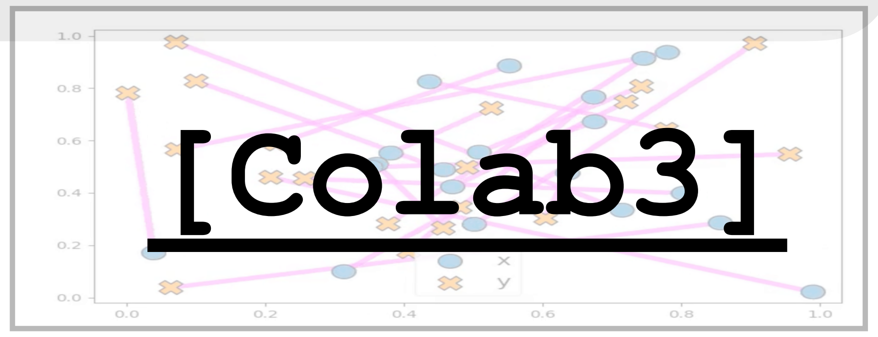


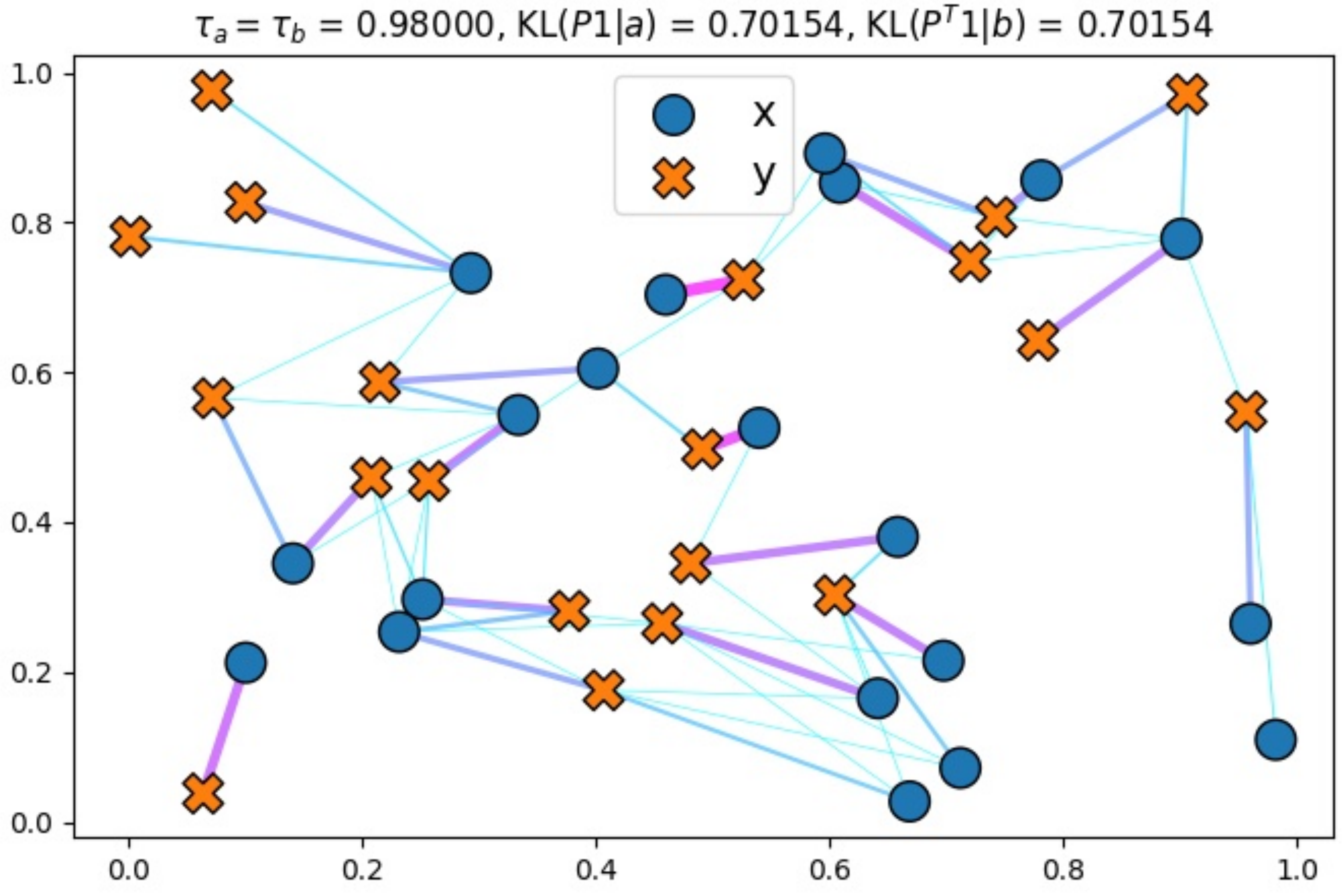


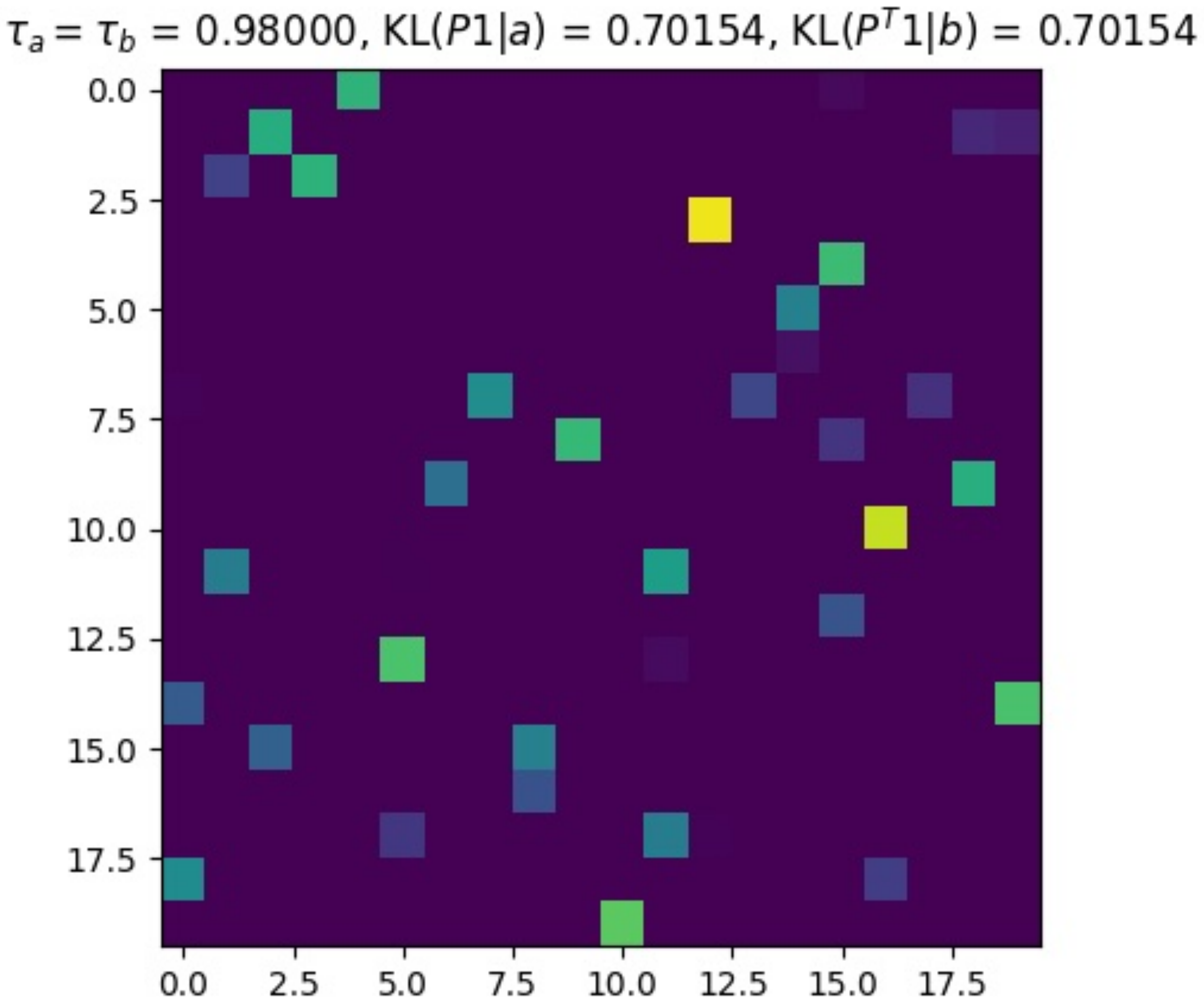


Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$

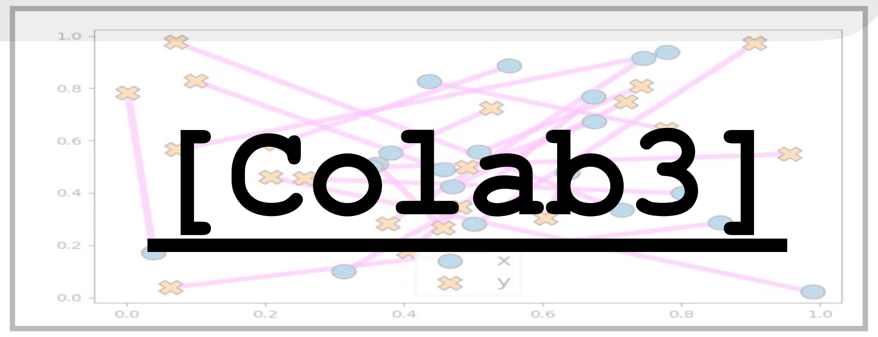




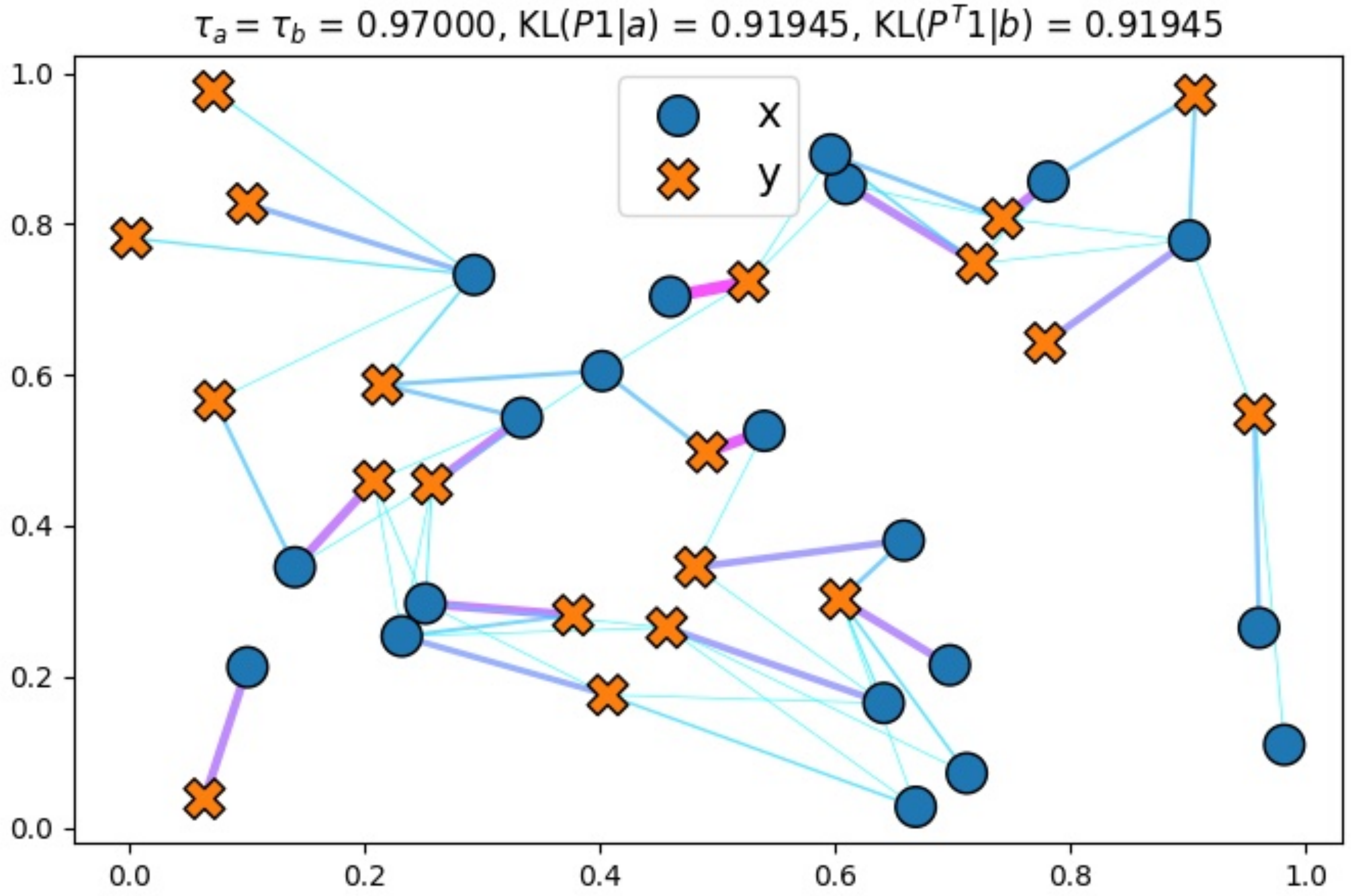


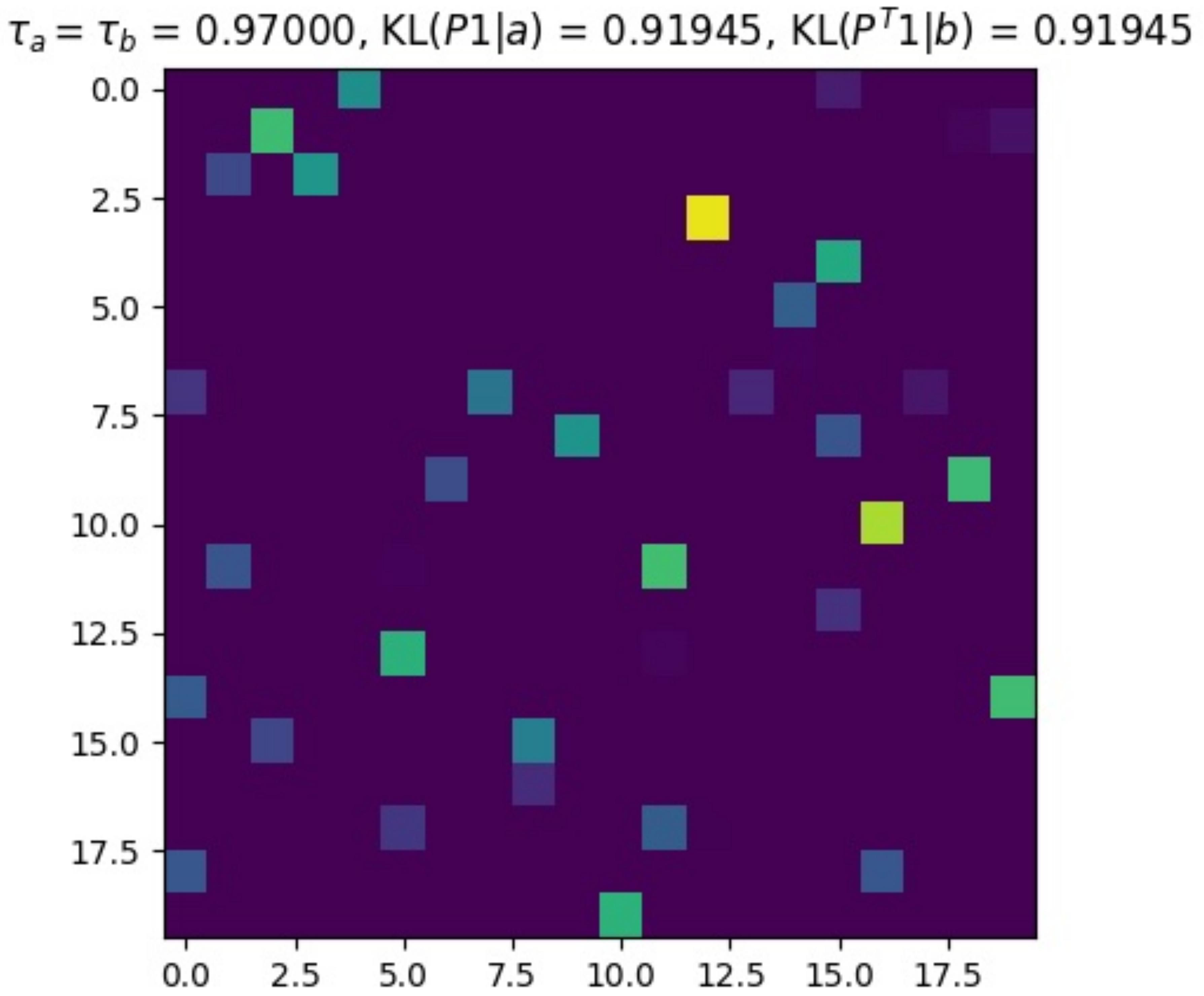
Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T\mathbf{1}_n \| b) - \gamma E(P)$$



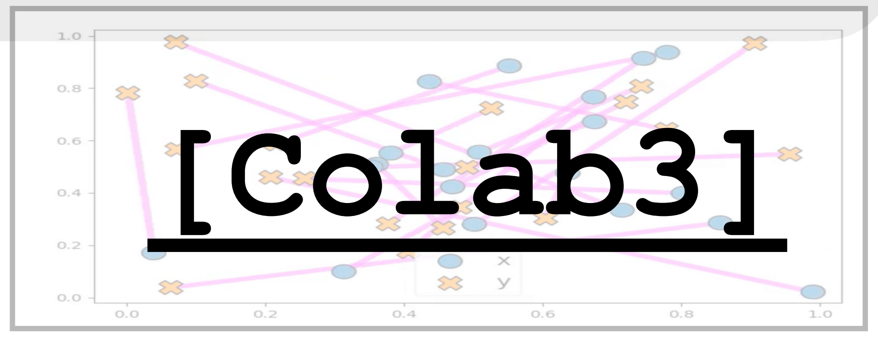
[Colab3]

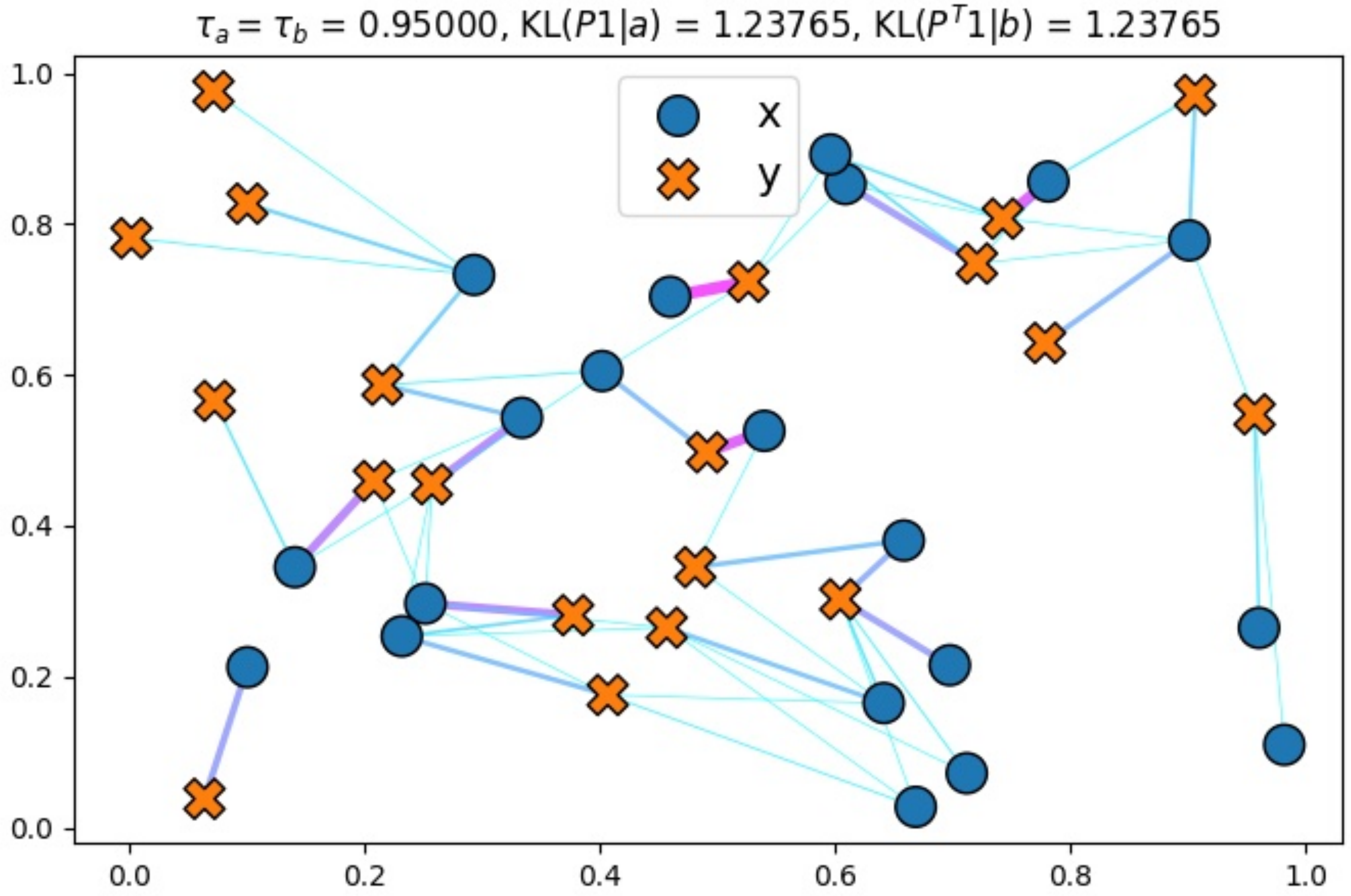


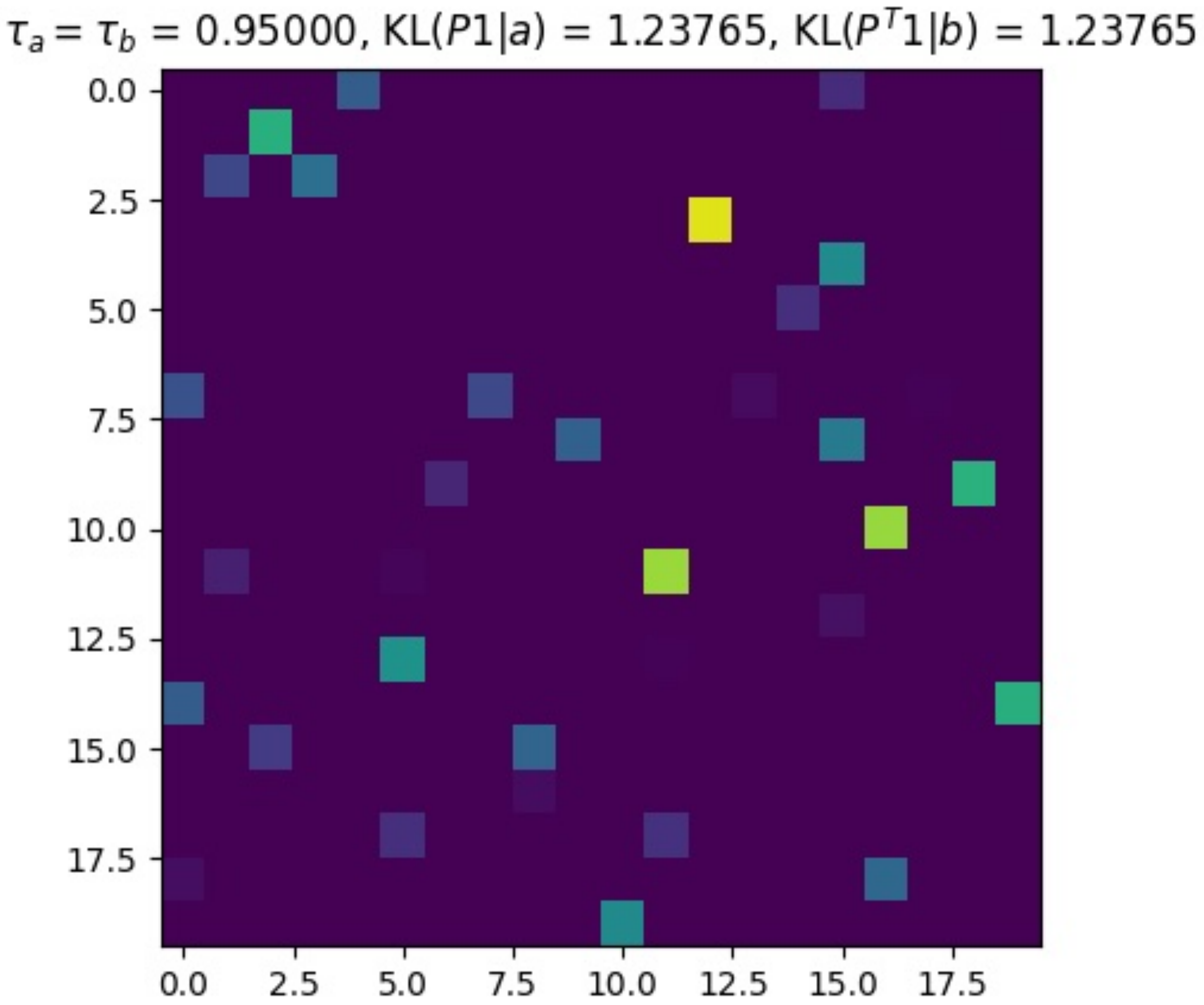


Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$

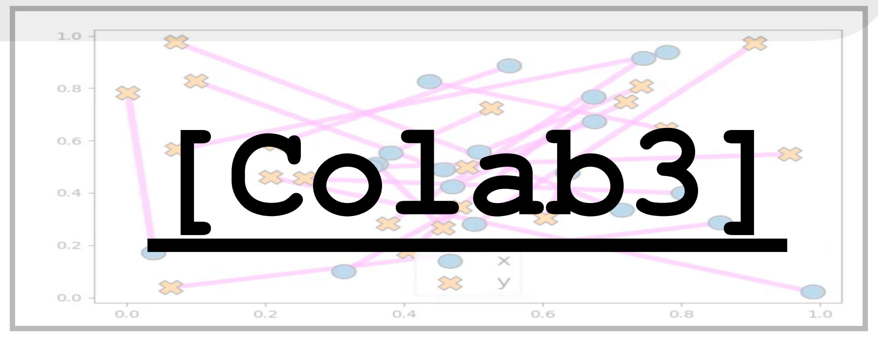


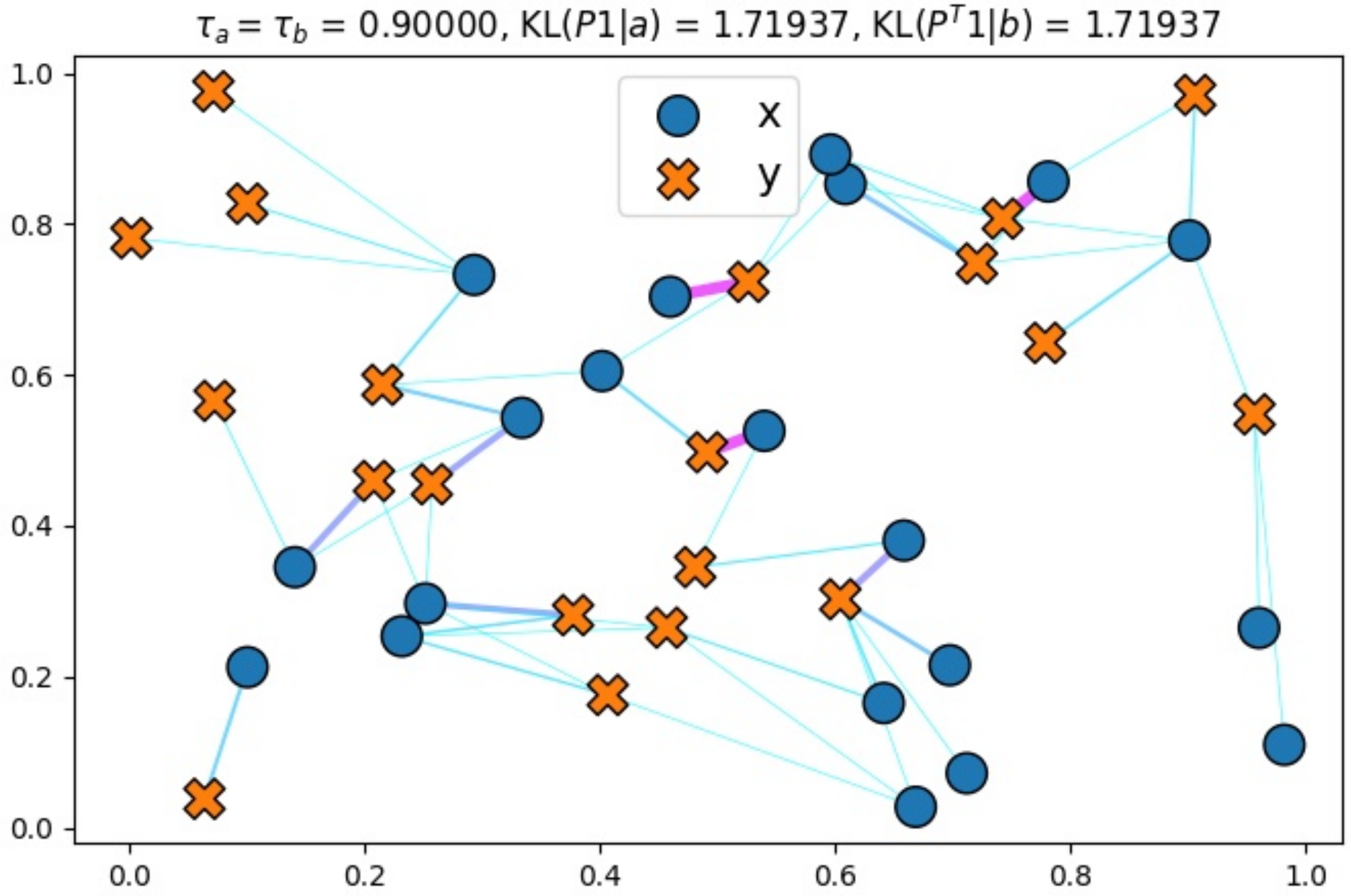


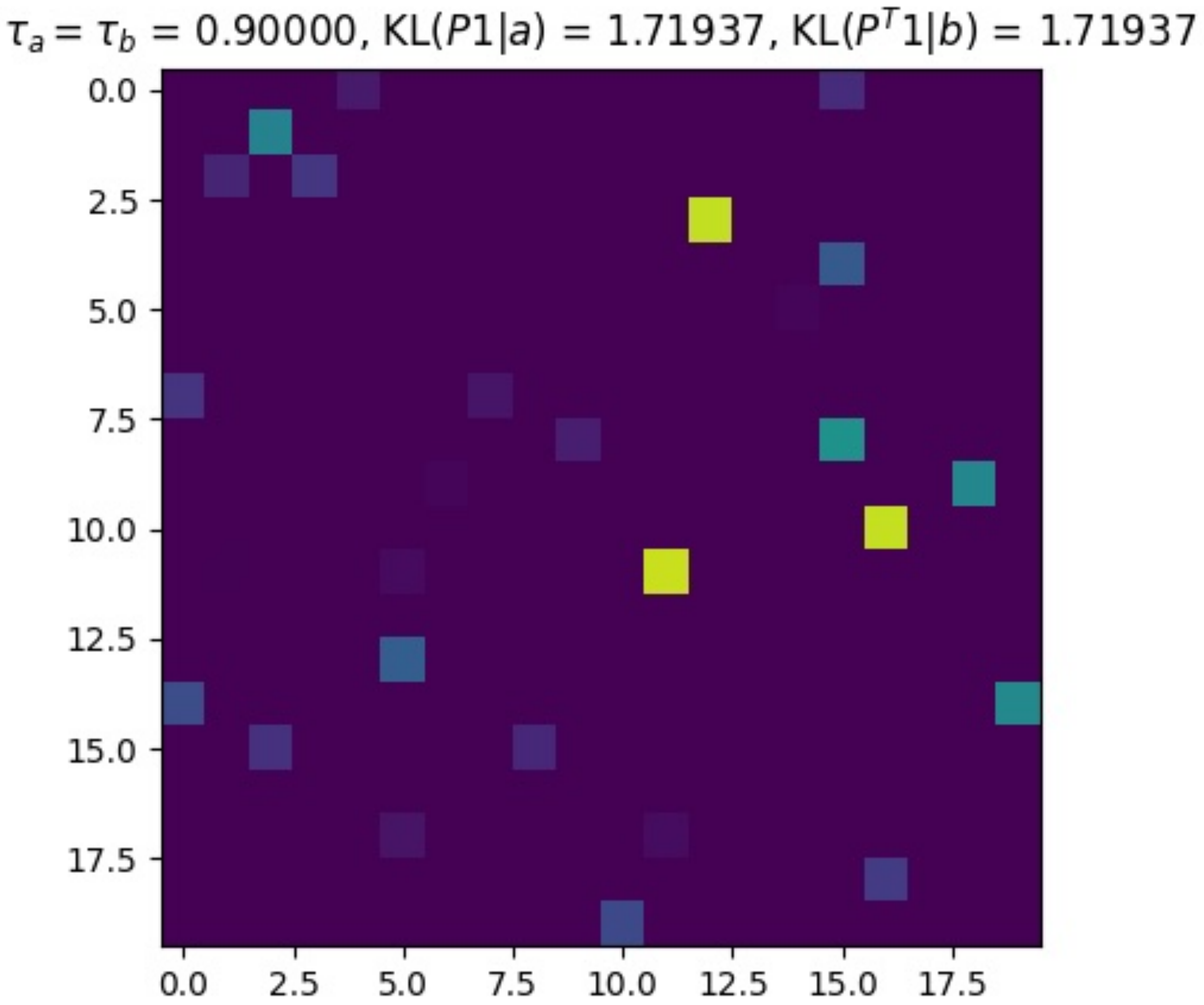


Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$

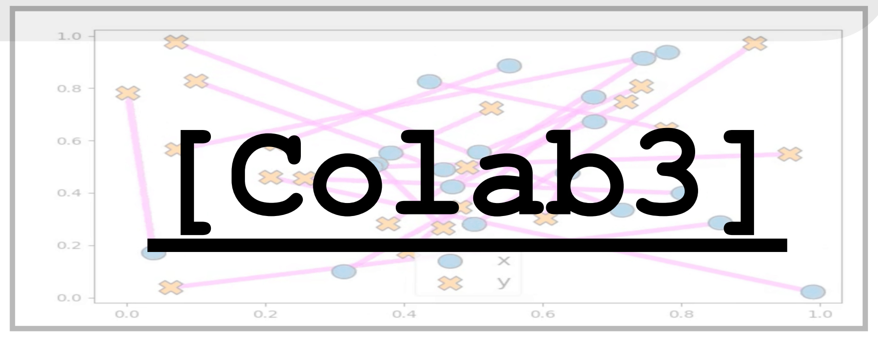


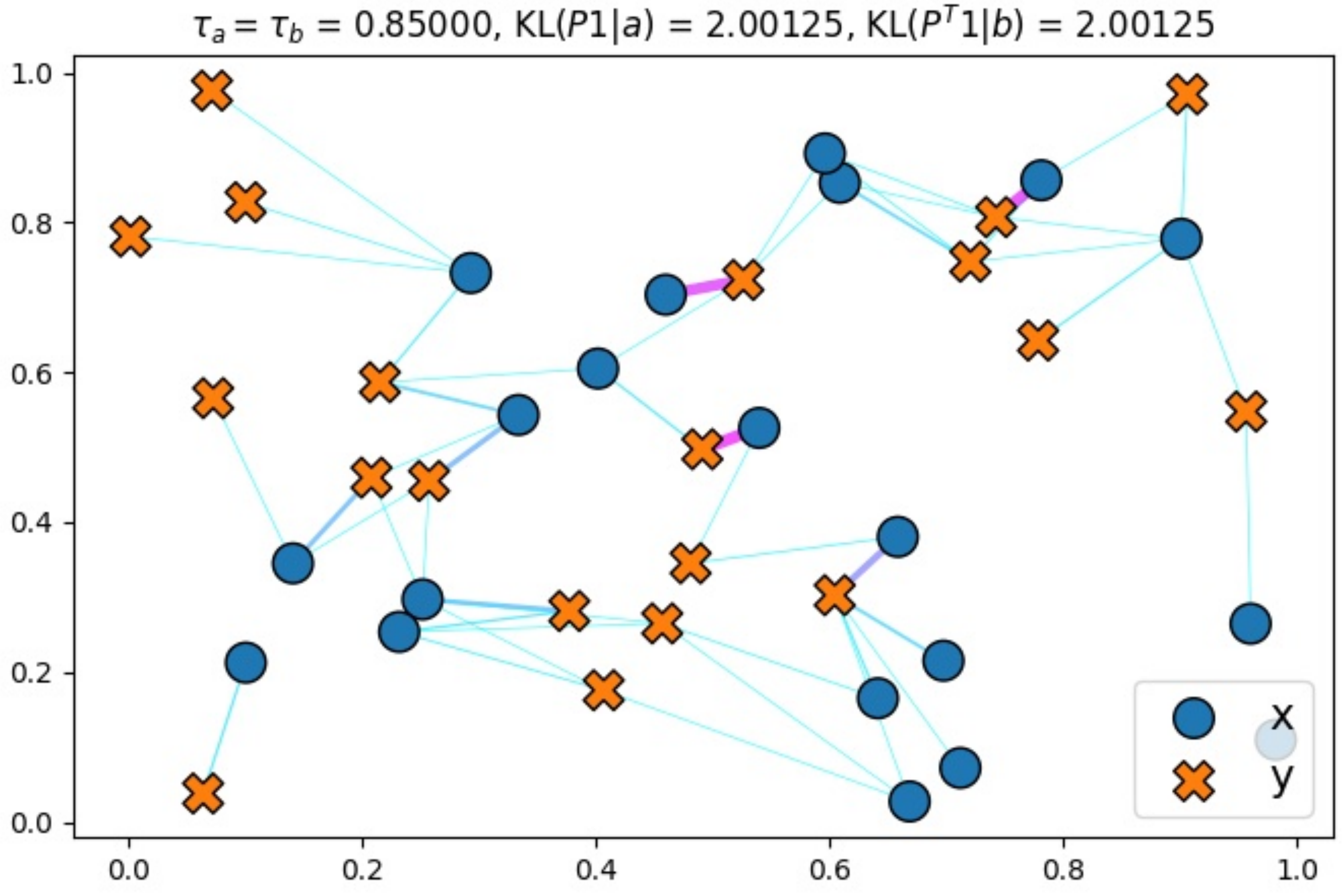


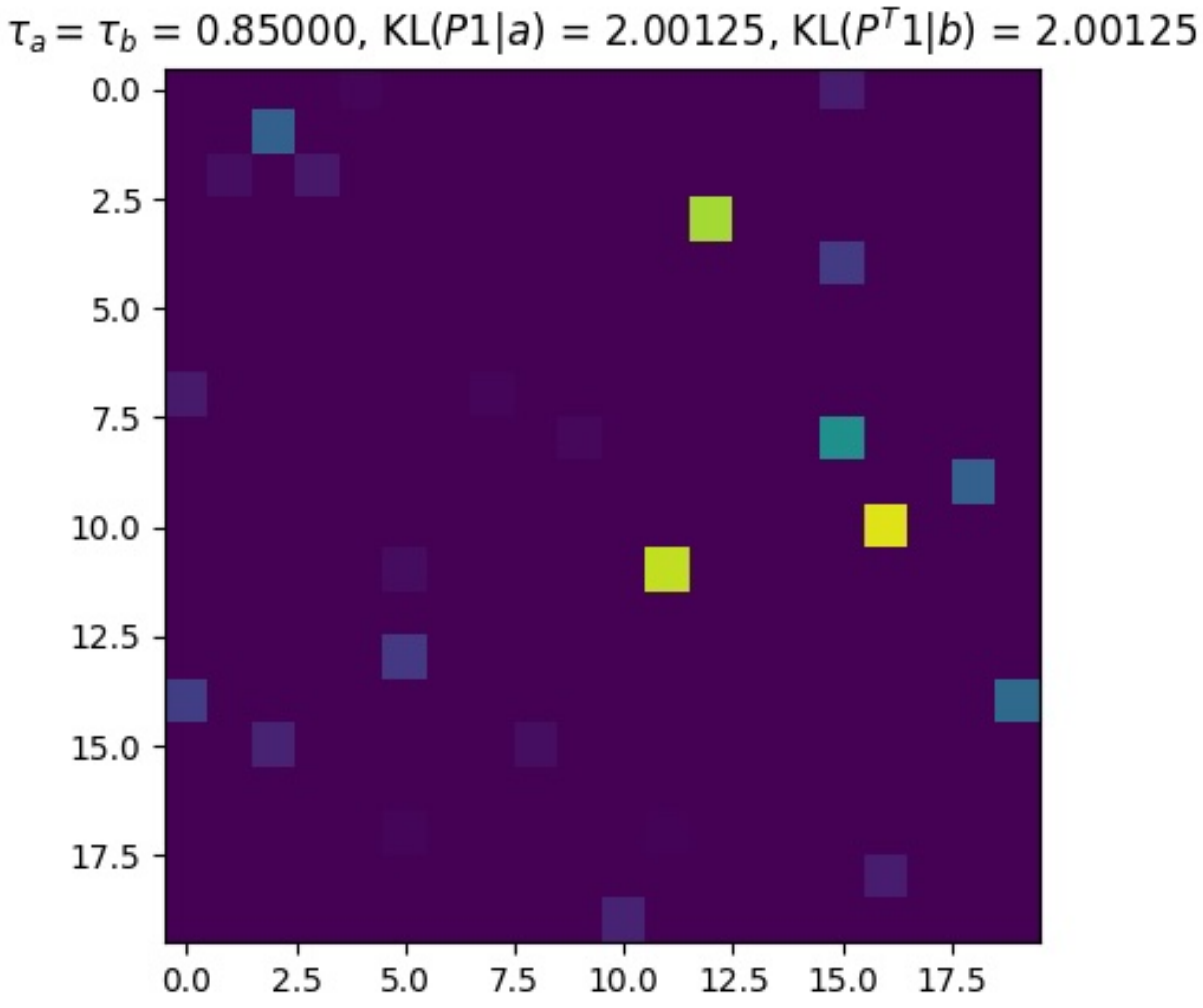


Unbalanced Sinkhorn

$$\min_{P\in\mathbb{R}_+^{n\times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T\mathbf{1}_n \| b) - \gamma E(P)$$

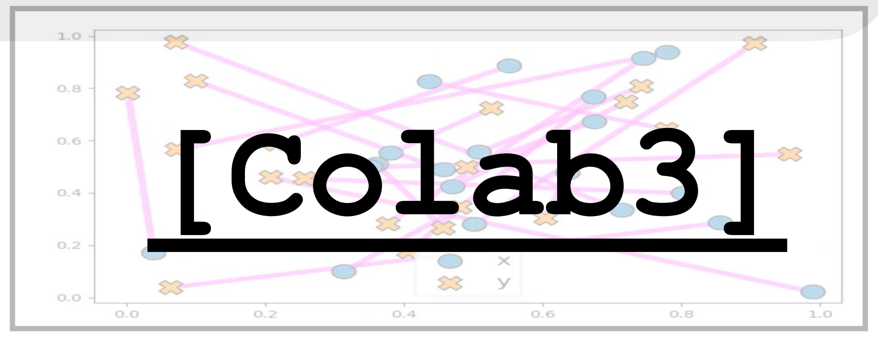


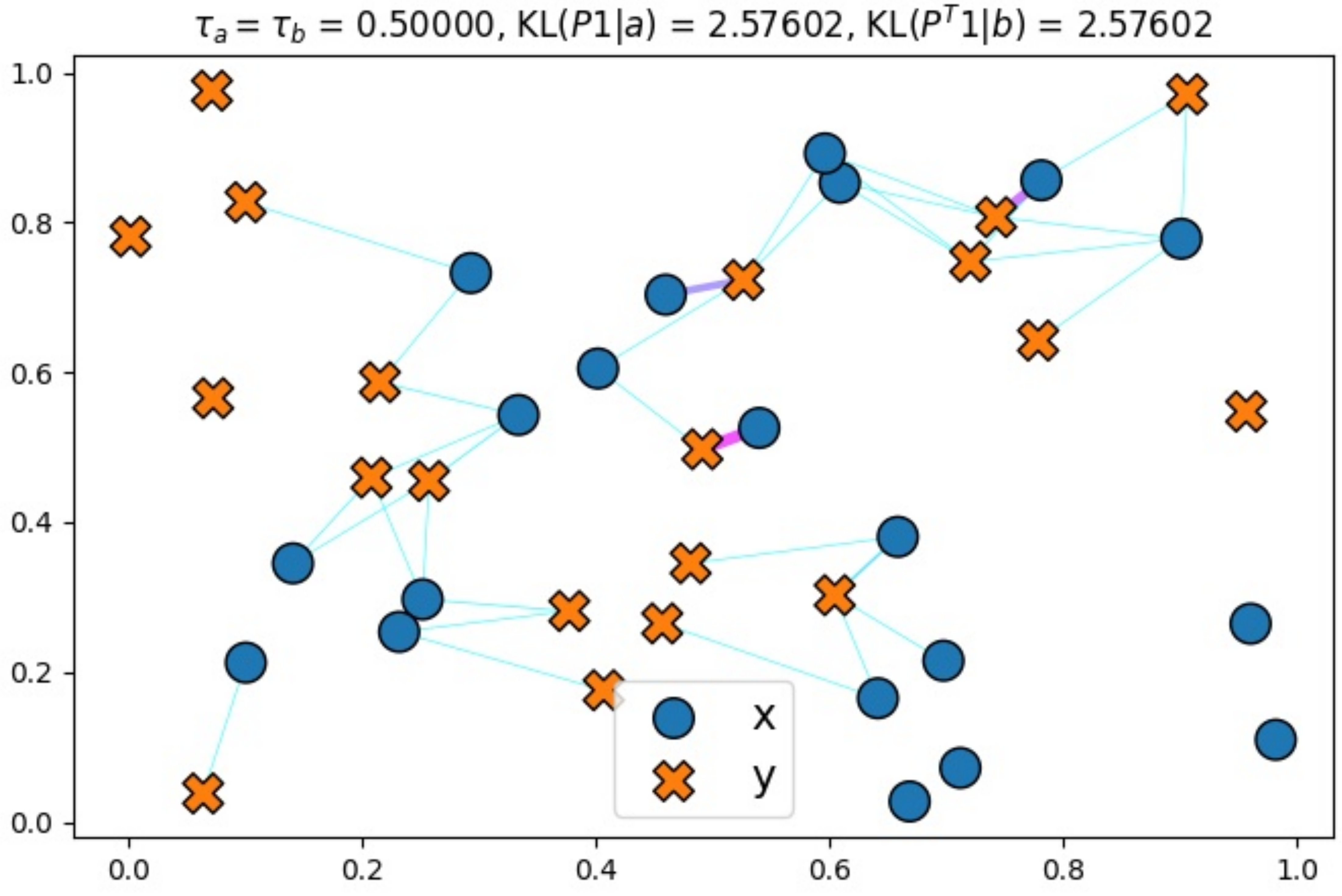


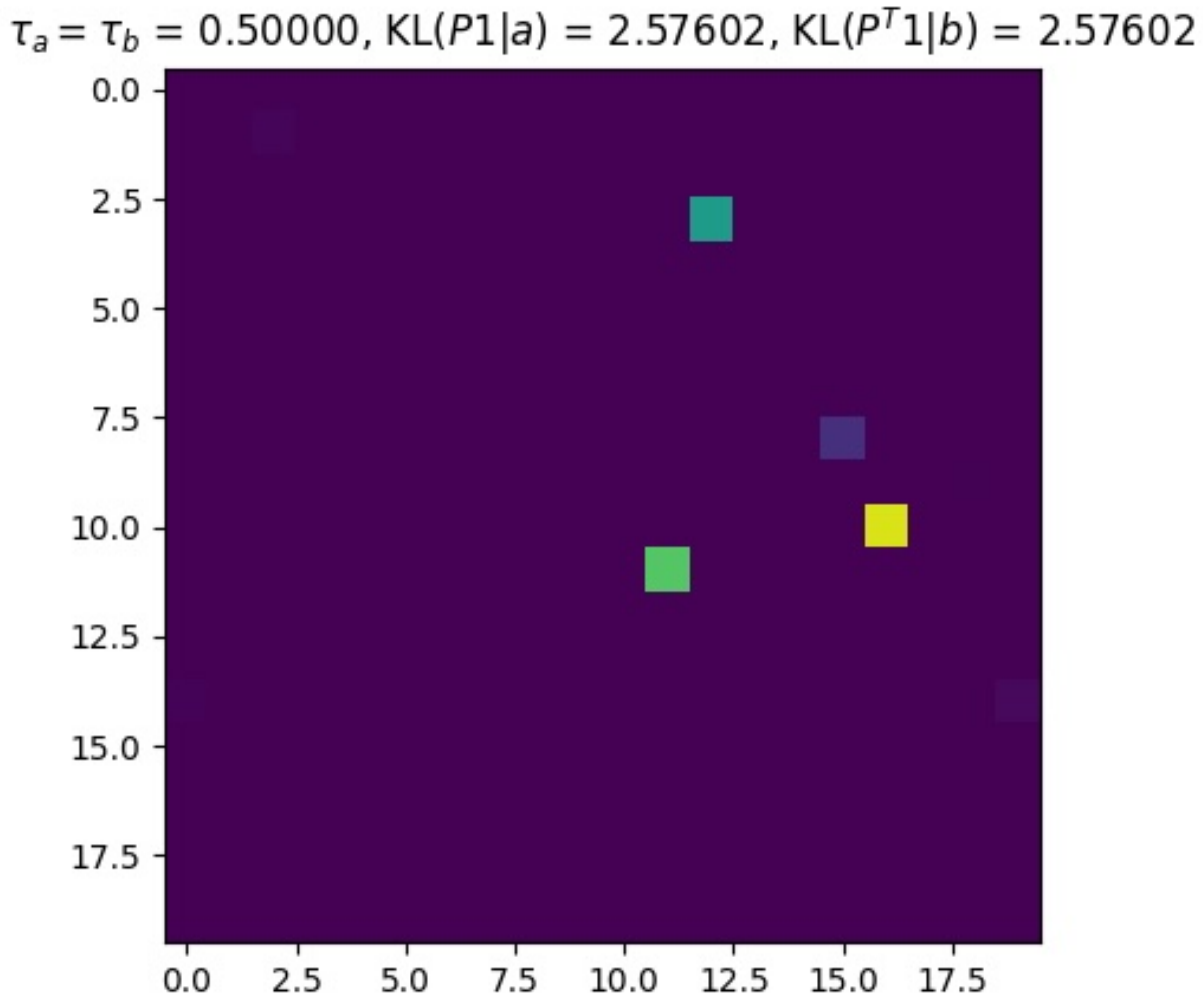


Unbalanced Sinkhorn

$$\min_{P\in\mathbb{R}_+^{n\times m}} \langle P, M_{XY}\rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T\mathbf{1}_n \| b) - \gamma E(P)$$

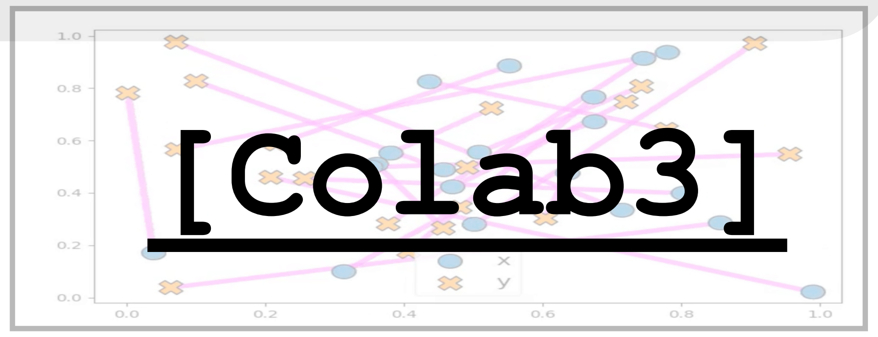


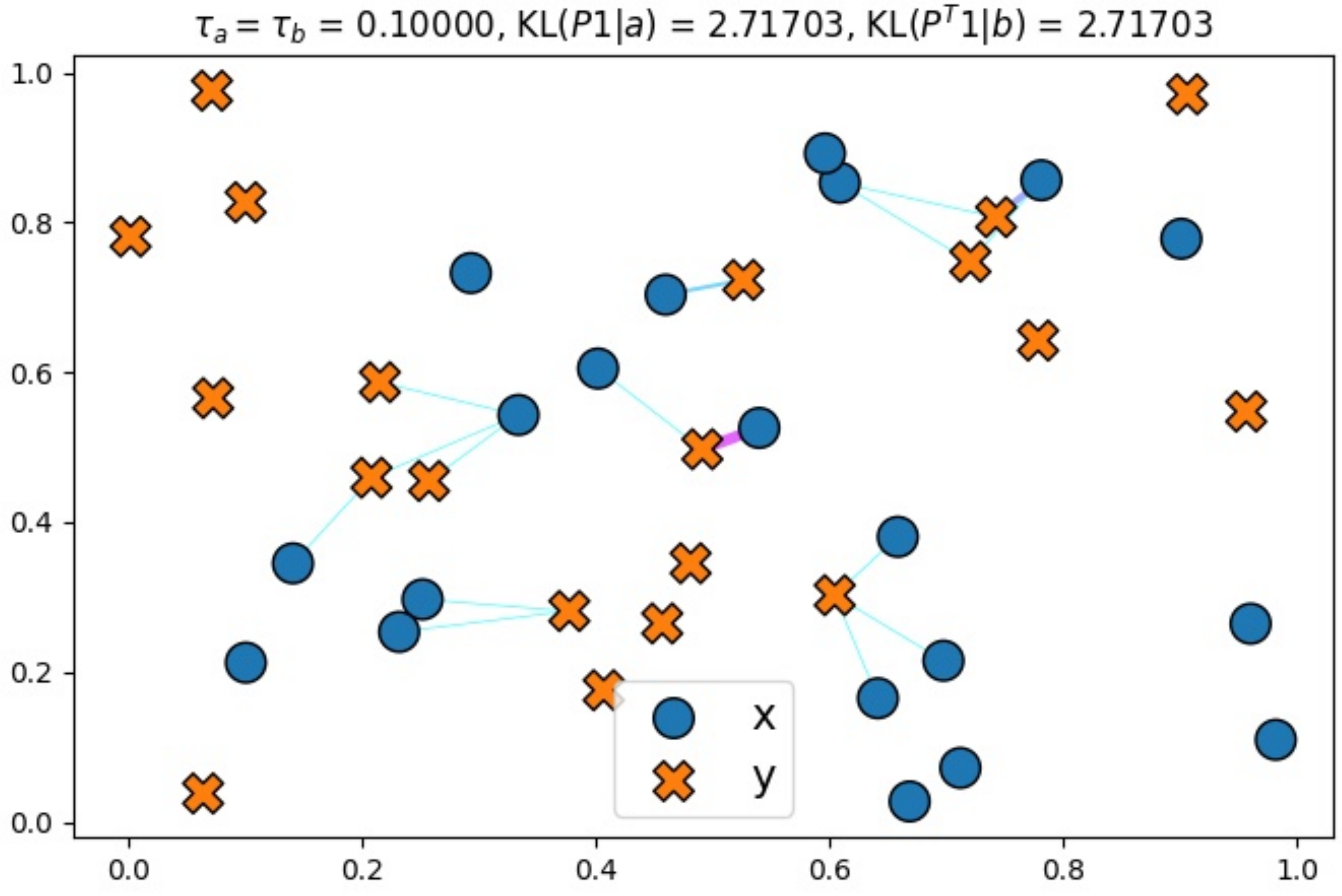


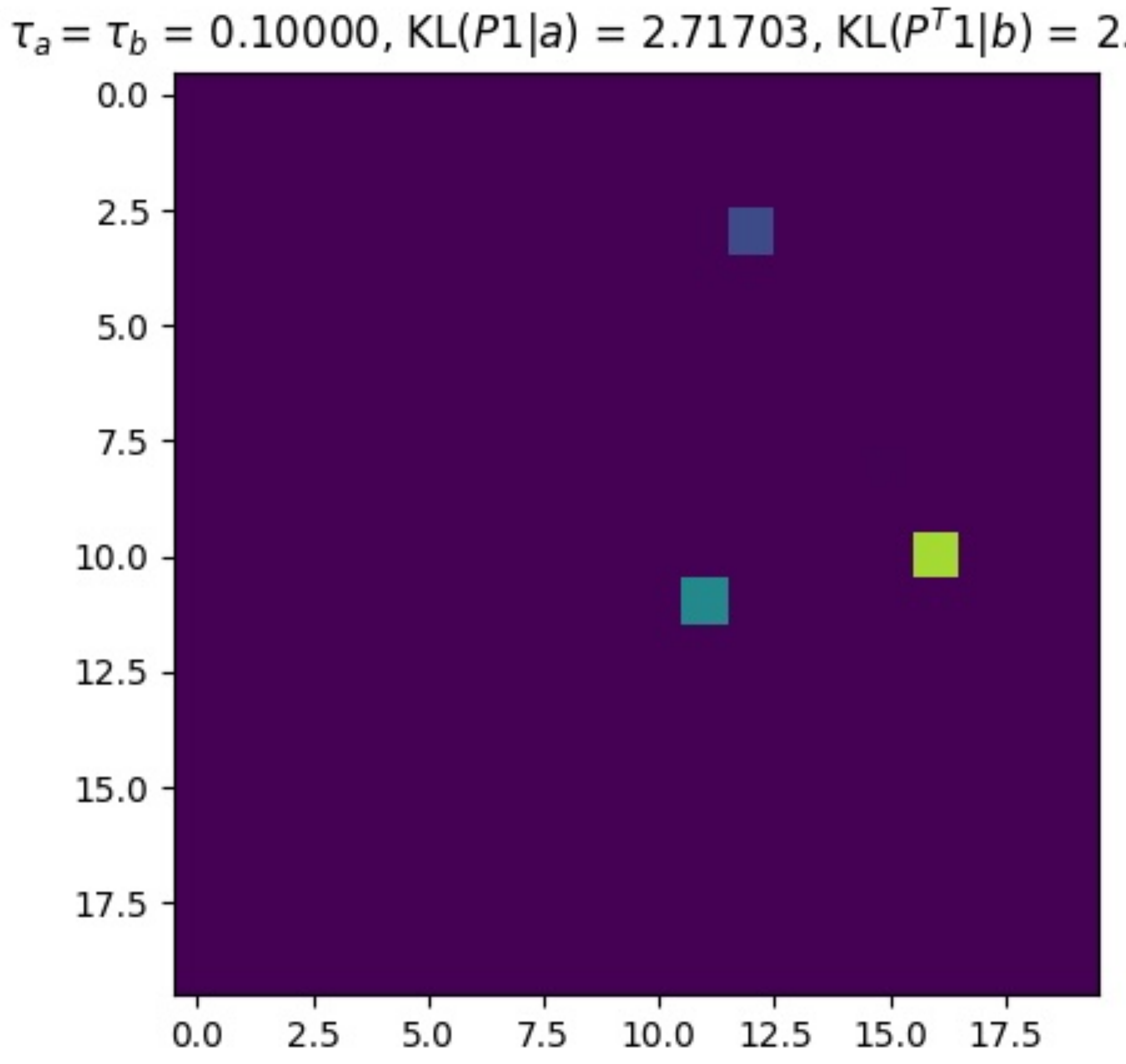


Unbalanced Sinkhorn

$$\min_{\boldsymbol{P}\in\mathbb{R}_+^{\boldsymbol{n}\times\boldsymbol{m}}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle + \rho_{\boldsymbol{a}}\mathrm{KL}(\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}}\|\boldsymbol{a}) + \rho_{\boldsymbol{b}}\mathrm{KL}(\boldsymbol{P}^T\mathbf{1}_{\boldsymbol{n}}\|\boldsymbol{b}) - \gamma E(\boldsymbol{P})$$

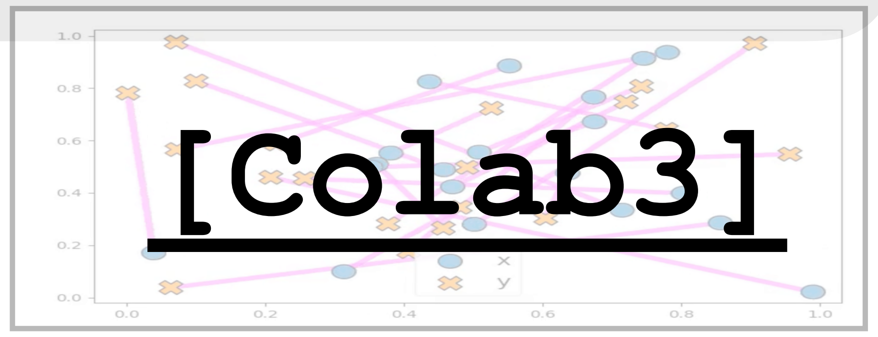




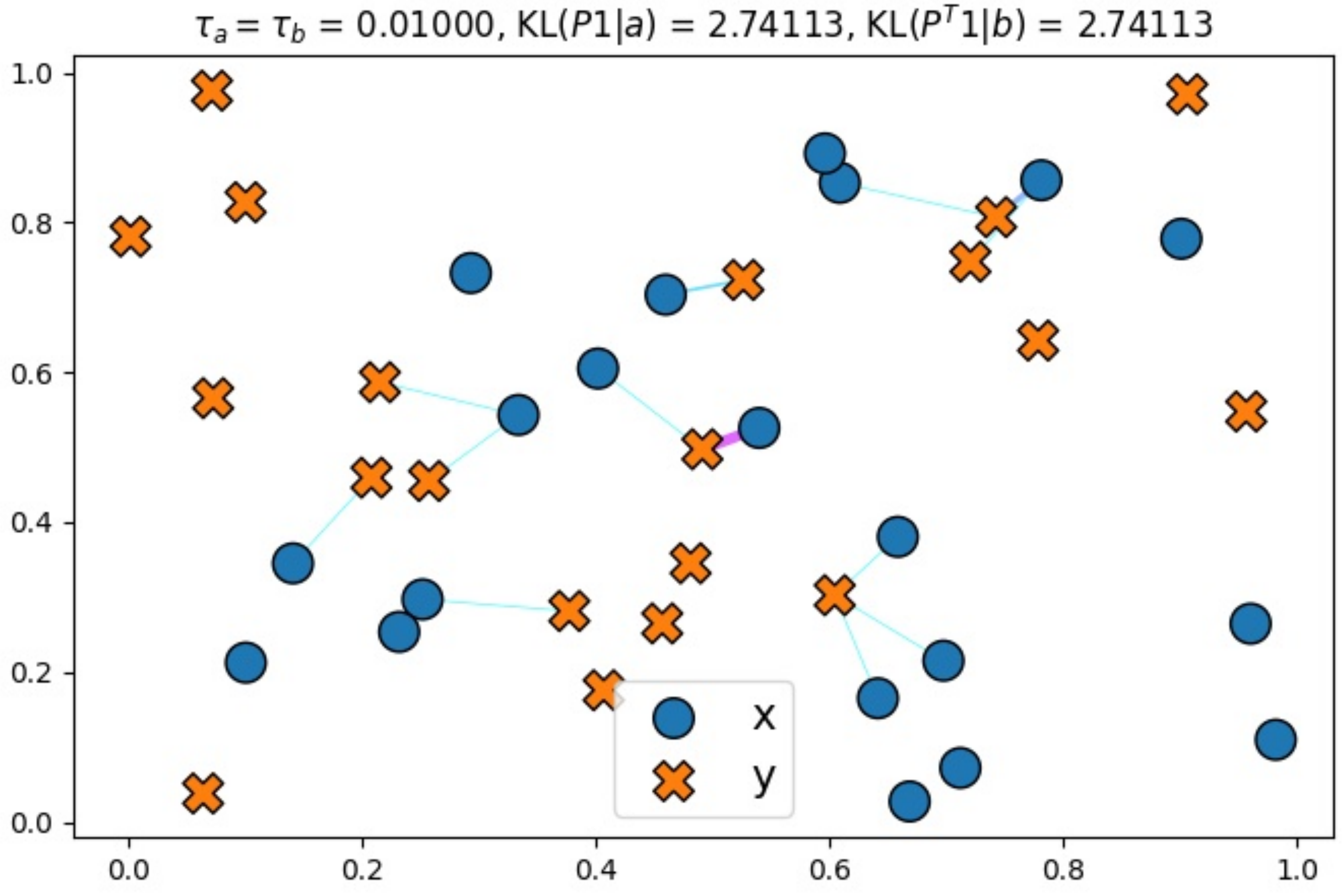


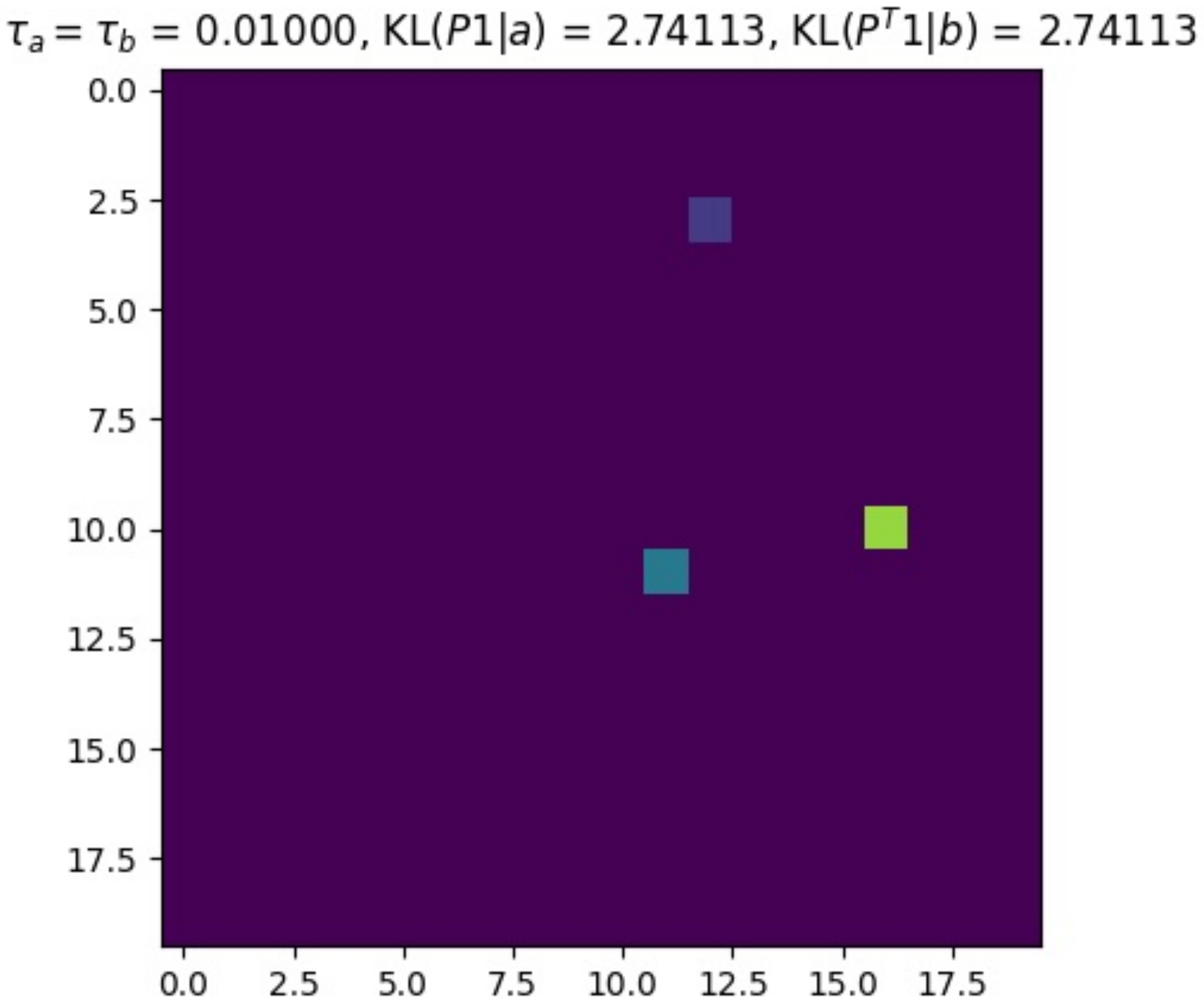
Unbalanced Sinkhorn

$$\min_{P \in \mathbb{R}_+^{n \times m}} \langle P, M_{XY} \rangle + \rho_a \mathrm{KL}(P\mathbf{1}_m \| a) + \rho_b \mathrm{KL}(P^T \mathbf{1}_n \| b) - \gamma E(P)$$



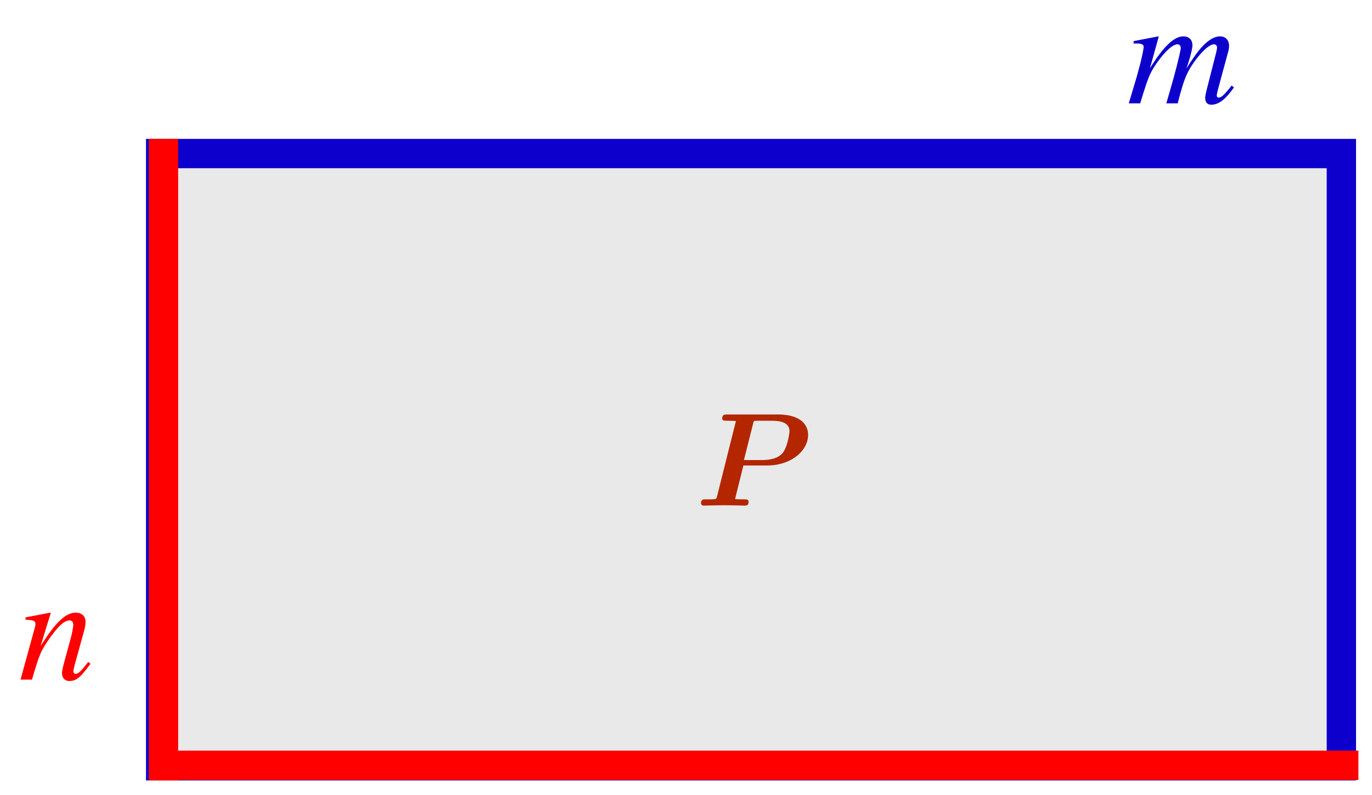
[Colab3]

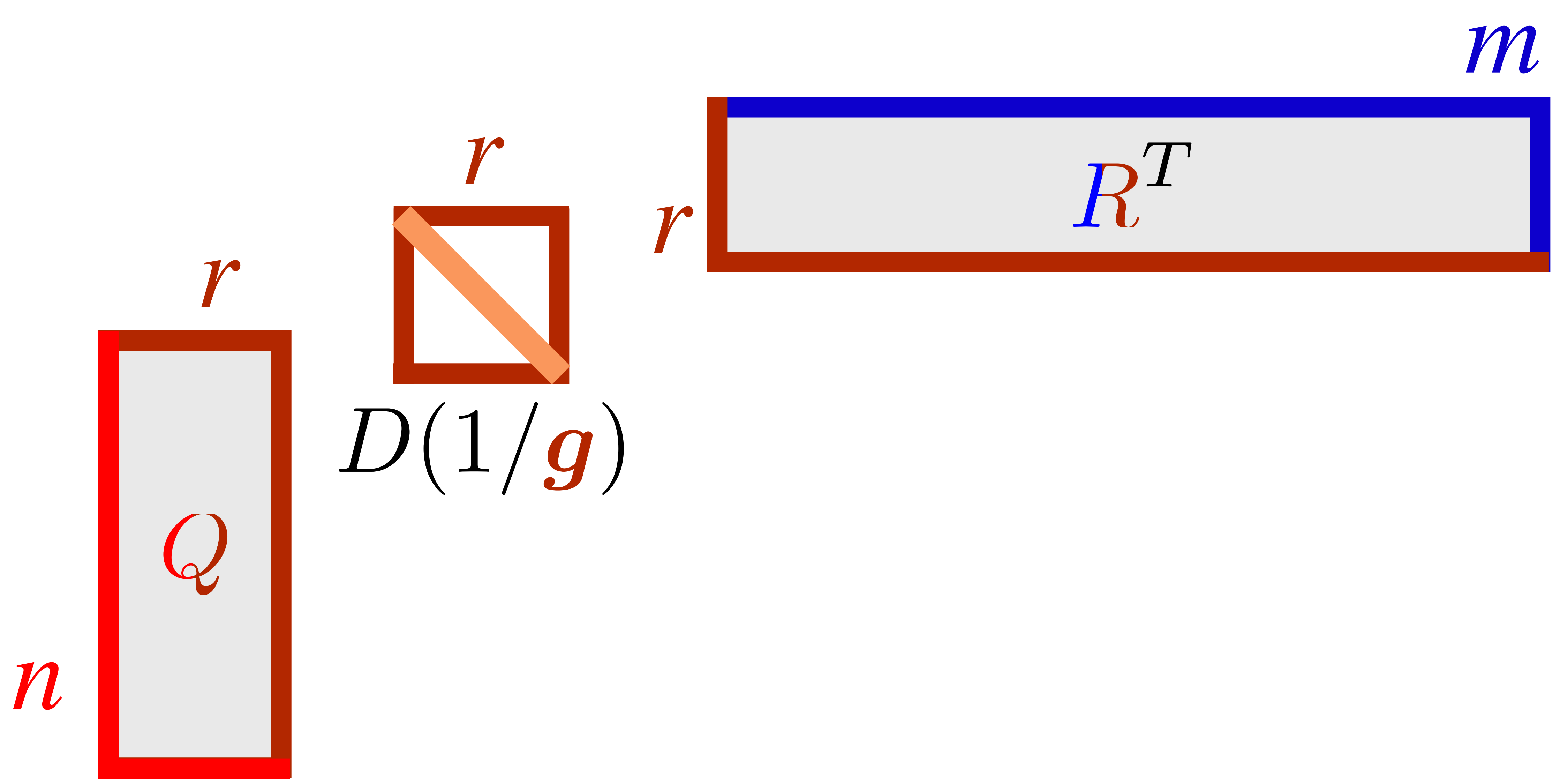




Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\boldsymbol{P}\in U(\boldsymbol{a},\boldsymbol{b})} \langle \boldsymbol{P}, M_{\mathbf{XY}} \rangle$$





$$Q \in U(\boldsymbol{a},\boldsymbol{g}),\ R \in U(\boldsymbol{b},\boldsymbol{g}), \boldsymbol{g} \in \Sigma_r \Rightarrow Q D(1/\boldsymbol{g})\ R^T \in U(\boldsymbol{a},\boldsymbol{b})$$

Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle.$$

[Forrow+19, Scetbon+21]

Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle.$$

[Forrow+19, Scetbon+21]

$$\min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle$$

$$\min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle$$

$$\min_{\boldsymbol{g} \in \Sigma_r} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle$$

Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle.$$

[Forrow+19, Scetbon+21]

$$\min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle \qquad \min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q, M_{\boldsymbol{XY}} R D(1/\boldsymbol{g}) \rangle$$

$$\min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle \longrightarrow \min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle R, M_{\boldsymbol{XY}}^T Q D(1/\boldsymbol{g}) \rangle$$

$$\min_{\boldsymbol{g} \in \Sigma_r} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle \qquad \min_{\boldsymbol{g} \in \Sigma_r} \langle D(1/\boldsymbol{g}), Q^T M_{\boldsymbol{XY}} R \rangle$$

Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}} \rangle.$$

[Forrow+19, Scetbon+21]

$$\min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}} \rangle \longrightarrow \min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q, \underbrace{M_{\mathbf{XY}} R D(1/\boldsymbol{g})}_{C_Q} \rangle$$

$$\min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}} \rangle \longrightarrow \min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle R, \underbrace{M_{\mathbf{XY}}^T Q D(1/\boldsymbol{g})}_{C_R} \rangle$$

$$\min_{\boldsymbol{g} \in \Sigma_r} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}} \rangle \longrightarrow \min_{\boldsymbol{g} \in \Sigma_r} \langle D(1/\boldsymbol{g}), \underbrace{Q^T M_{\mathbf{XY}} R}_{c_{\boldsymbol{g}}} \rangle$$

Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{X}\mathbf{Y}} \rangle.$$

[Forrow+19, Scetbon+21]

$$\min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{X}\mathbf{Y}} \rangle \quad\longrightarrow\quad \min_{Q \in U(\boldsymbol{a},\boldsymbol{g})} \langle Q, \underbrace{M_{\mathbf{X}\mathbf{Y}} R D(1/\boldsymbol{g})}_{C_Q} \rangle \quad -\gamma E(Q, \boldsymbol{g}, R)$$

$$\min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{X}\mathbf{Y}} \rangle \quad\longrightarrow\quad \min_{R \in U(\boldsymbol{b},\boldsymbol{g})} \langle R, \underbrace{M_{\mathbf{X}\mathbf{Y}}^T Q D(1/\boldsymbol{g})}_{C_R} \rangle \quad -\gamma E(Q, \boldsymbol{g}, R)$$

$$\min_{\boldsymbol{g} \in \Sigma_r} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{X}\mathbf{Y}} \rangle \quad\longrightarrow\quad \min_{\boldsymbol{g} \in \Sigma_r} \langle D(1/\boldsymbol{g}), \underbrace{Q^T M_{\mathbf{X}\mathbf{Y}}\, R}_{c_{\boldsymbol{g}}} \rangle \quad -\gamma E(Q, \boldsymbol{g}, R)$$

Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}),\\ R\in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g}\in\Sigma_r}} \langle QD(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}}\rangle.$$

[Forrow+19, Scetbon+21]

$$O(\boldsymbol{n}+\boldsymbol{m}) \text{ if } M_{\mathbf{XY}} = \boldsymbol{AB}^T$$

$$(\boldsymbol{n},\boldsymbol{d})\cdot(\boldsymbol{d},\boldsymbol{m})\cdot(\boldsymbol{m},\boldsymbol{r})\cdot\boldsymbol{r} = O\left((\boldsymbol{n}+\boldsymbol{m})(\boldsymbol{d}+\boldsymbol{r})\right)$$

$$\min_{Q\in U(\boldsymbol{a},\boldsymbol{g})} \langle QD(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}}\rangle \qquad \min_{Q\in U(\boldsymbol{a},\boldsymbol{g})} \langle Q, \underbrace{\boldsymbol{AB}^T RD(1/\boldsymbol{g})}_{C_Q}\rangle \quad -\gamma E(Q,\boldsymbol{g},R)$$

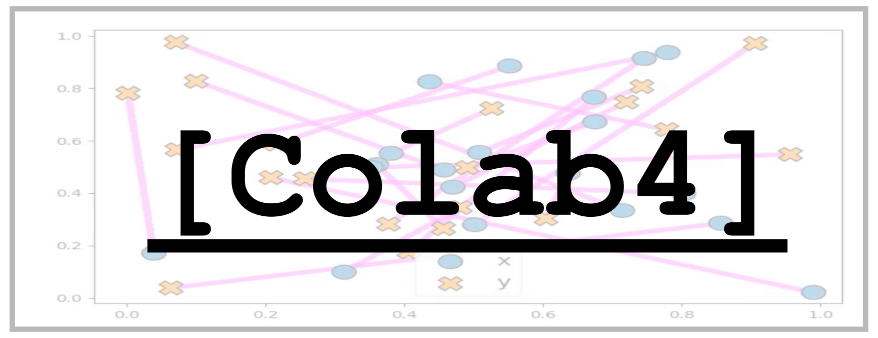
$$\min_{R\in U(\boldsymbol{b},\boldsymbol{g})} \langle QD(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}}\rangle \longrightarrow \min_{R\in U(\boldsymbol{b},\boldsymbol{g})} \langle R, \underbrace{M_{\mathbf{XY}}^T\, QD(1/\boldsymbol{g})}_{C_R}\rangle \quad -\gamma E(Q,\boldsymbol{g},R)$$

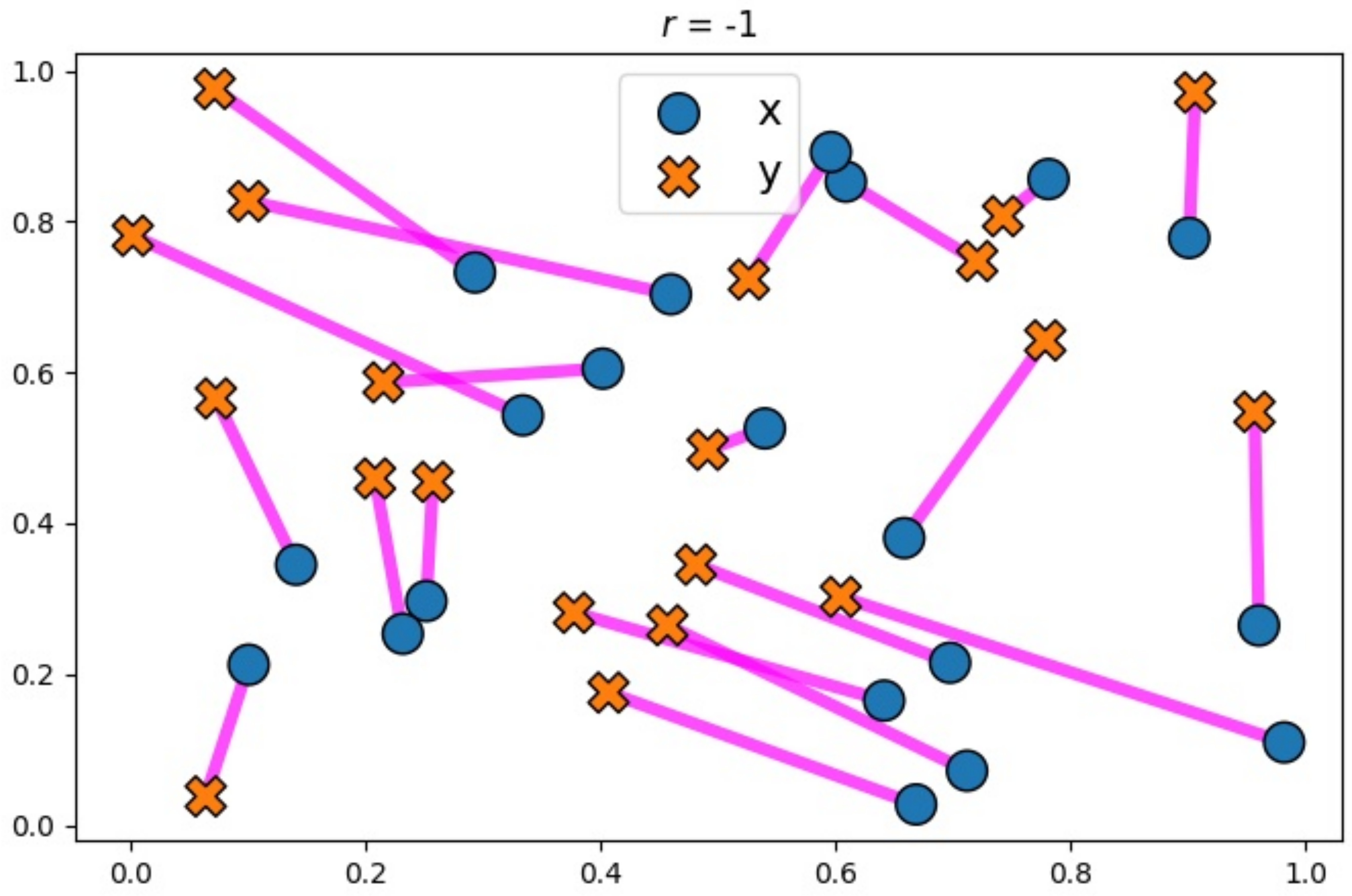
$$\min_{\boldsymbol{g}\in\Sigma_r} \langle QD(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}}\rangle \qquad \min_{\boldsymbol{g}\in\Sigma_r} \langle D(1/\boldsymbol{g}), \underbrace{Q^T M_{\mathbf{XY}}\, R}_{c_{\boldsymbol{g}}}\rangle \quad -\gamma E(Q,\boldsymbol{g},R)$$

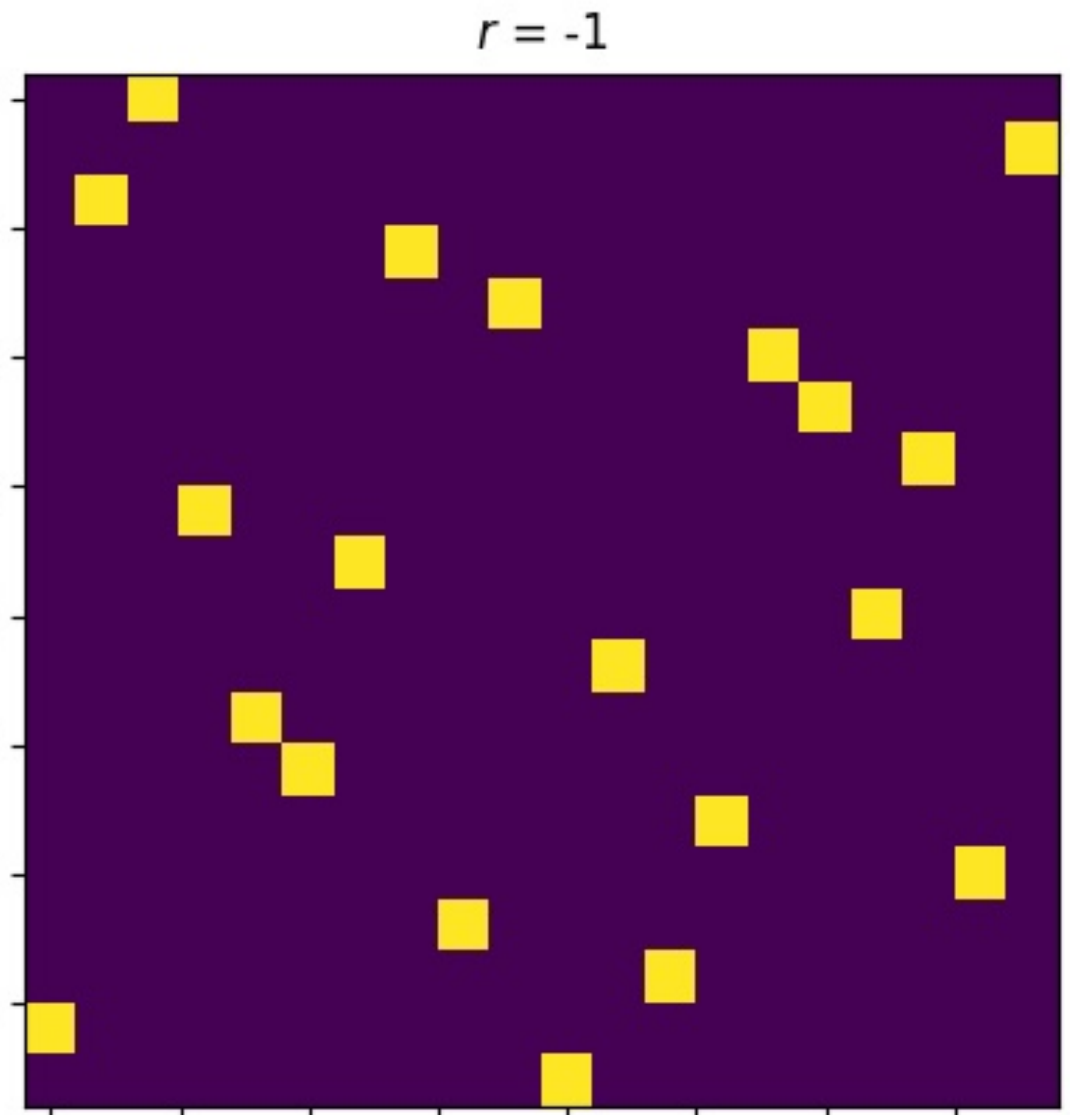
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}),\\ R\in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g}\in\Sigma_r}} \langle QD(1/\boldsymbol{g})\,R^T, M_{\mathbf{XY}}\rangle.$$

[Forrow+19, Scetbon+21]





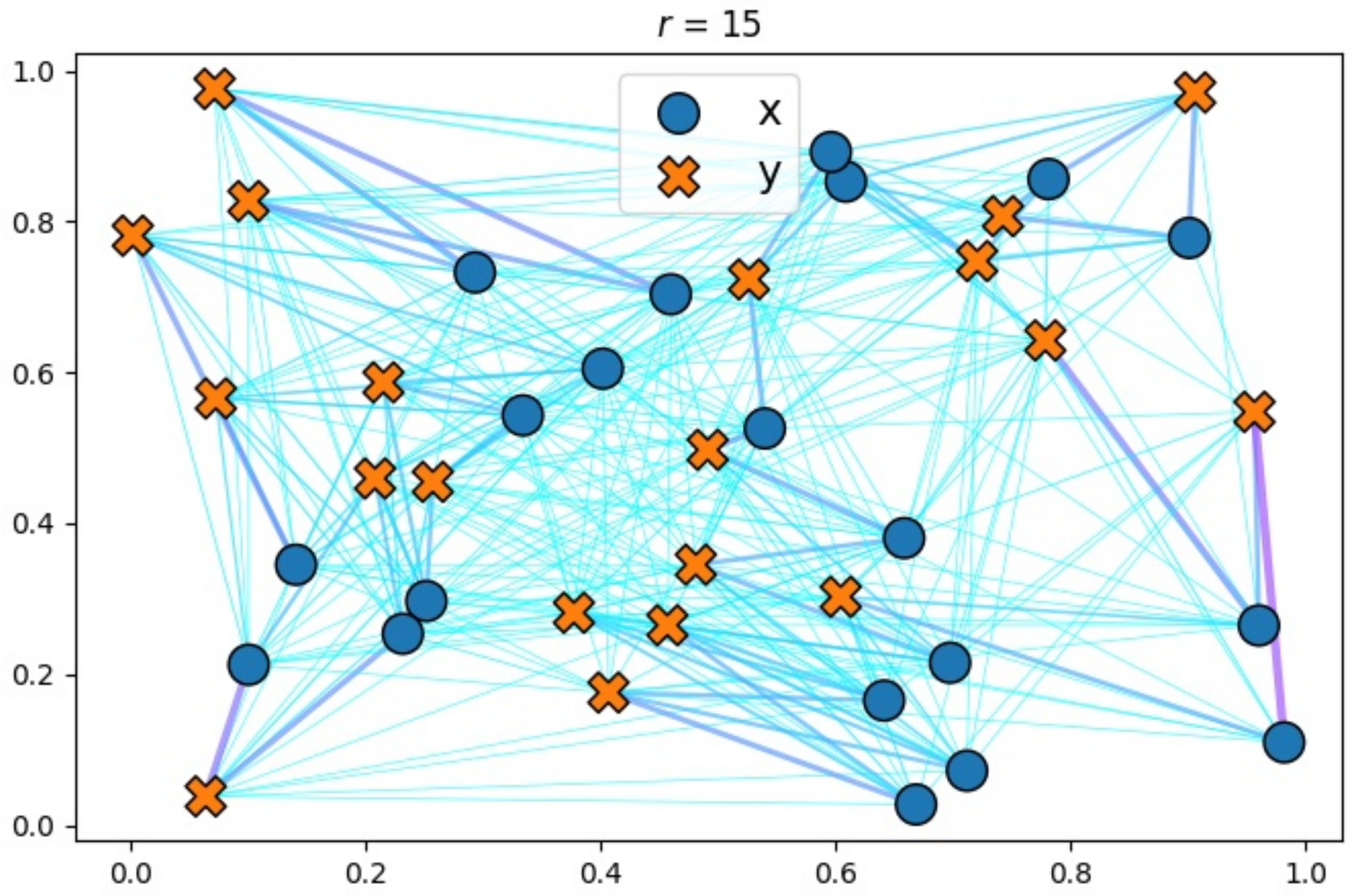


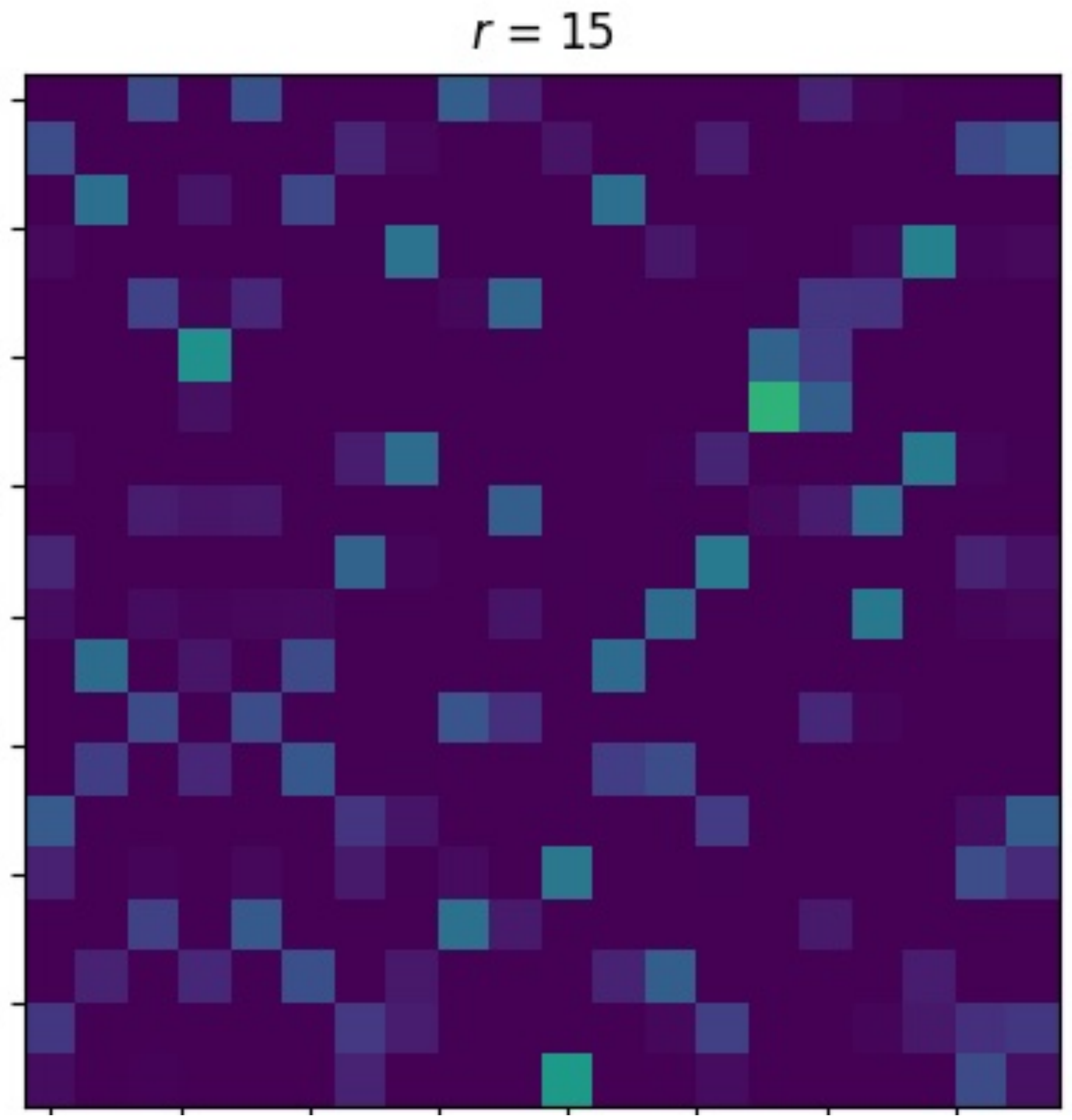
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a}, \boldsymbol{g}), \\ R \in U(\boldsymbol{b}, \boldsymbol{g}) \\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle.$$

[Forrow+19, Scetbon+21]

[Colab4]

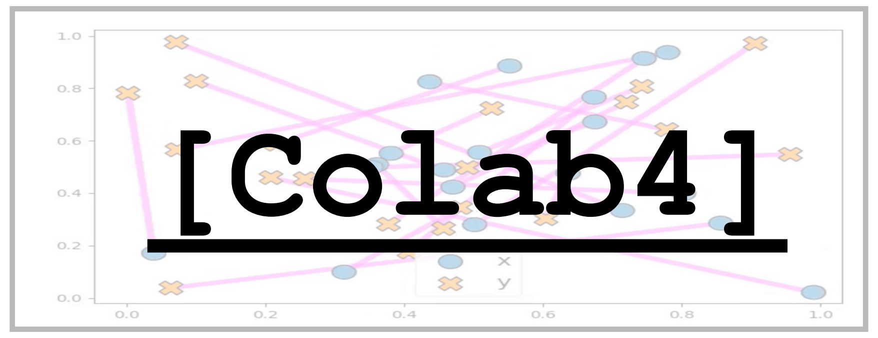


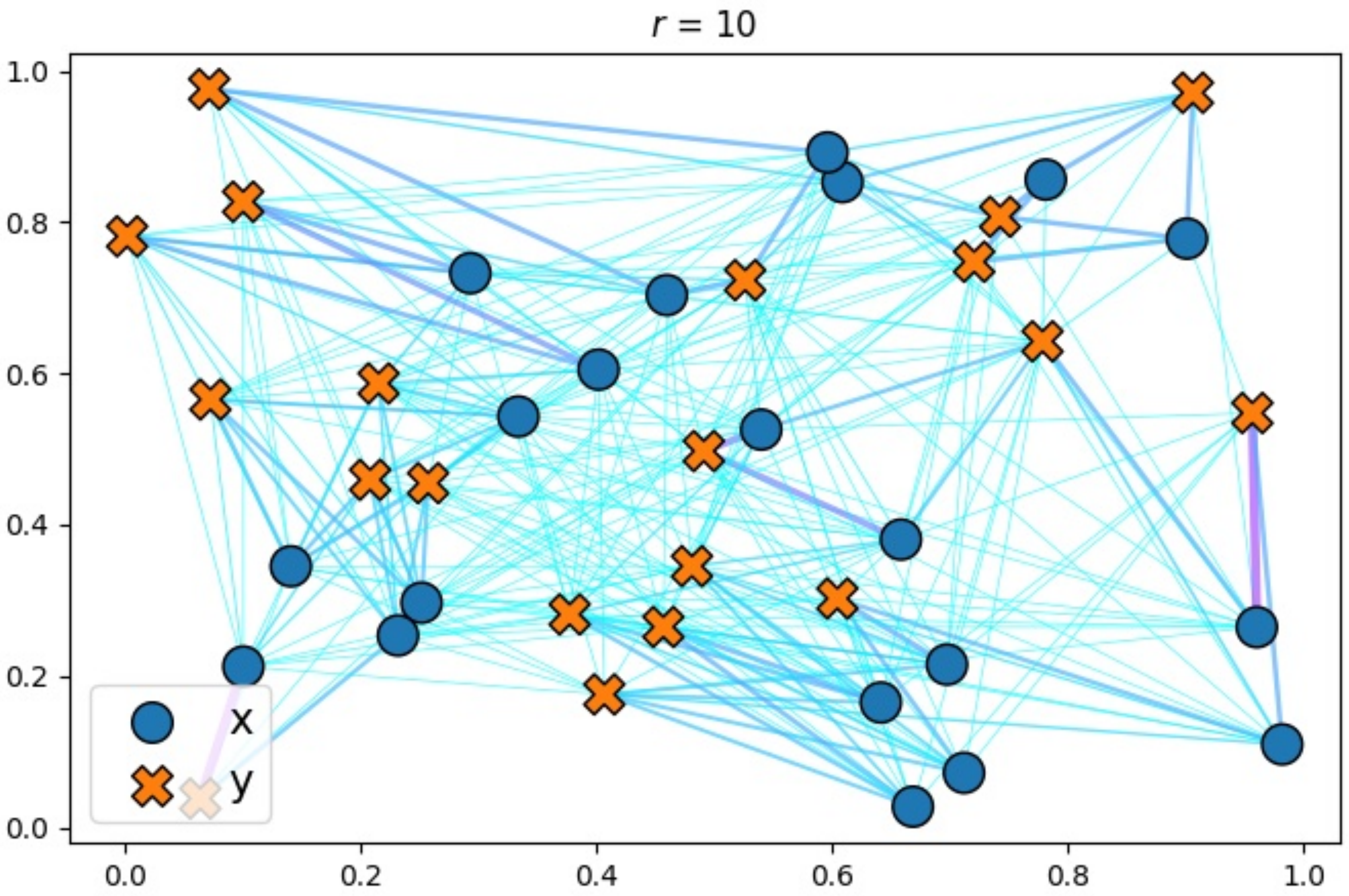


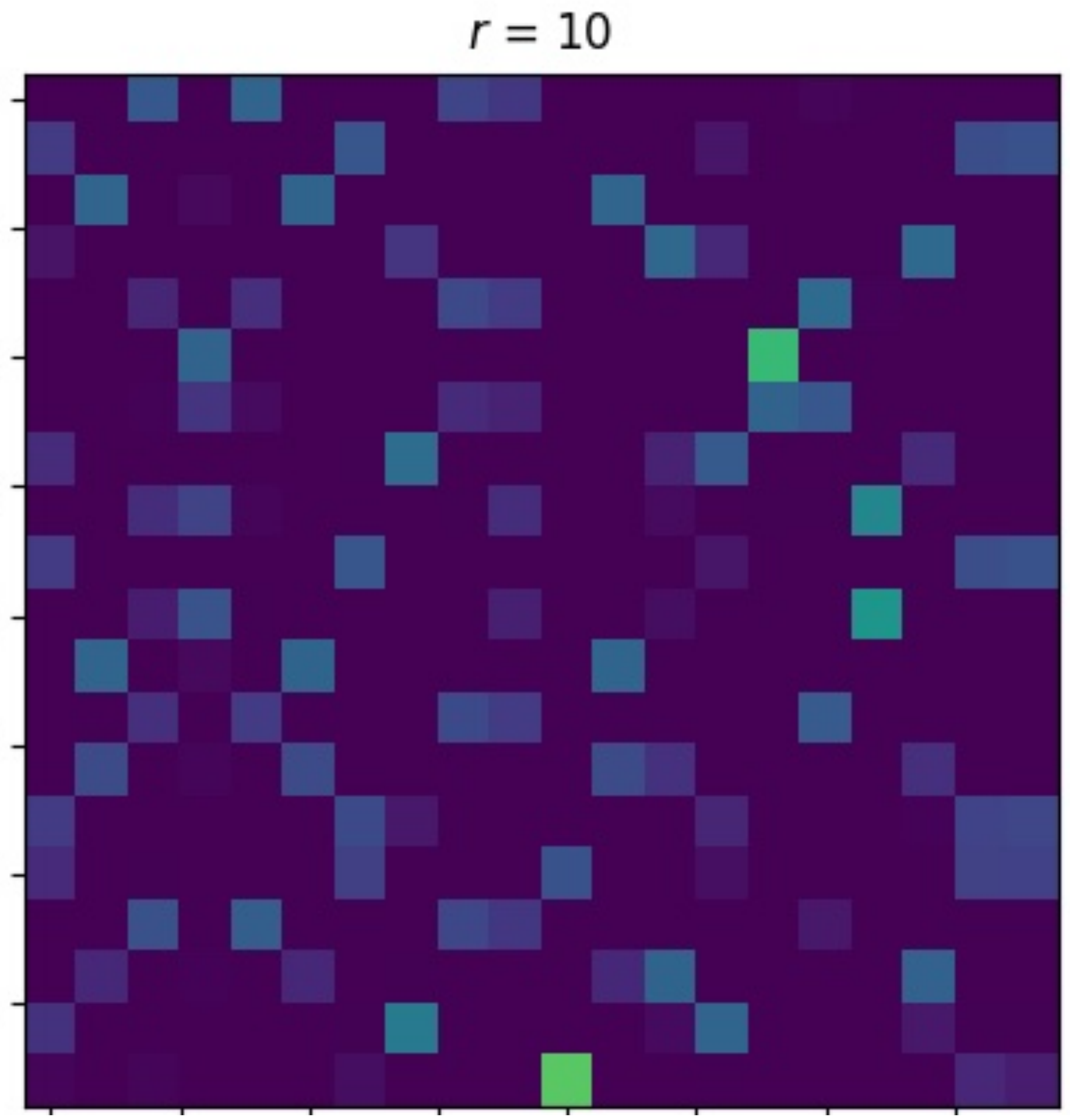
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\mathbf{XY}} \rangle.$$

[Forrow+19, Scetbon+21]



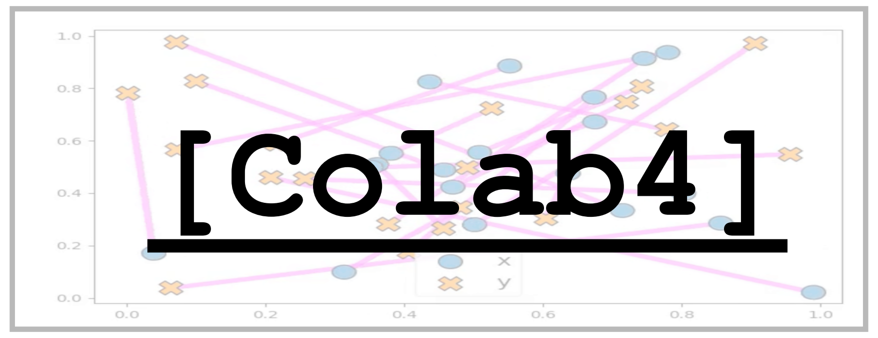


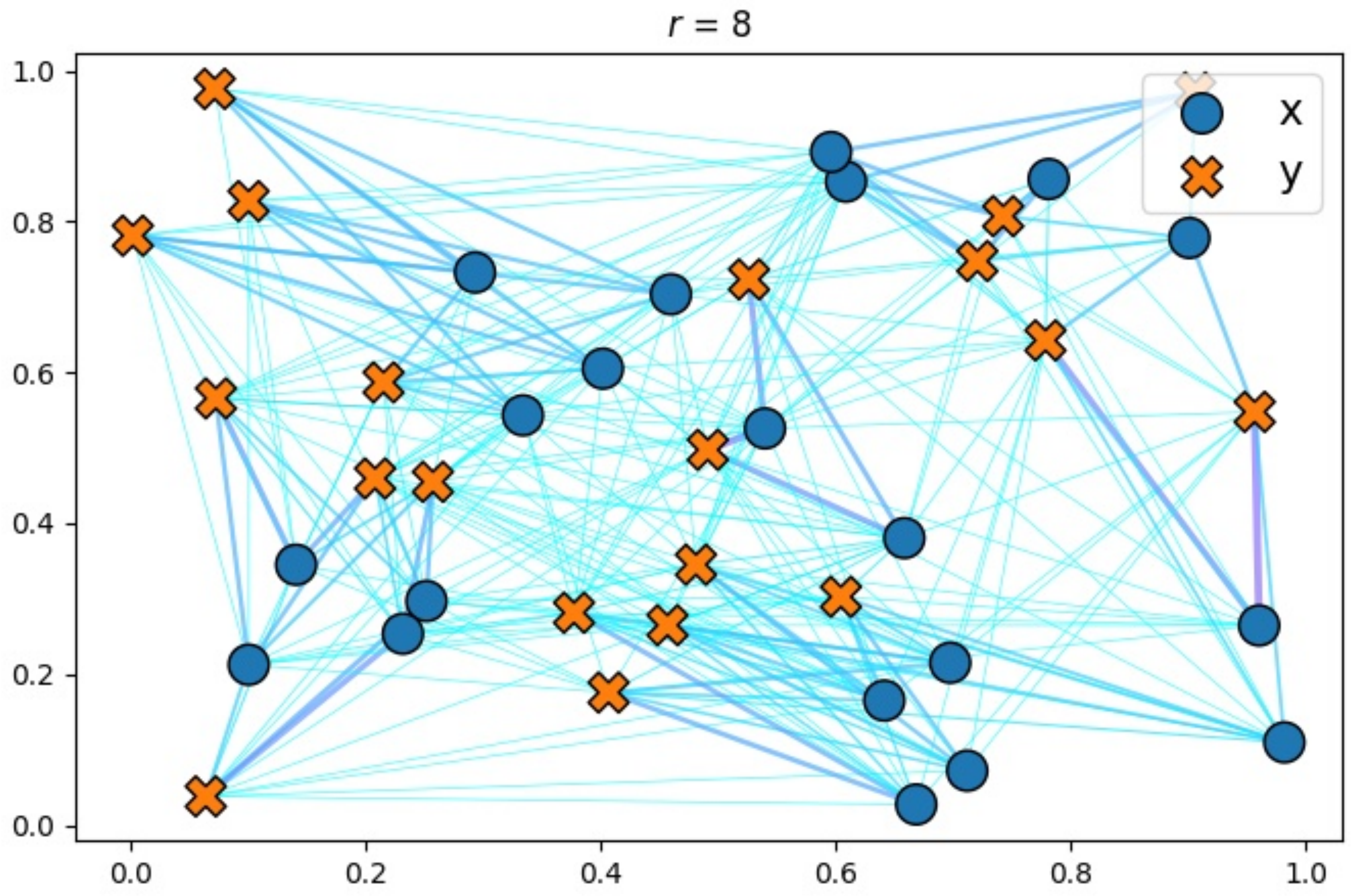


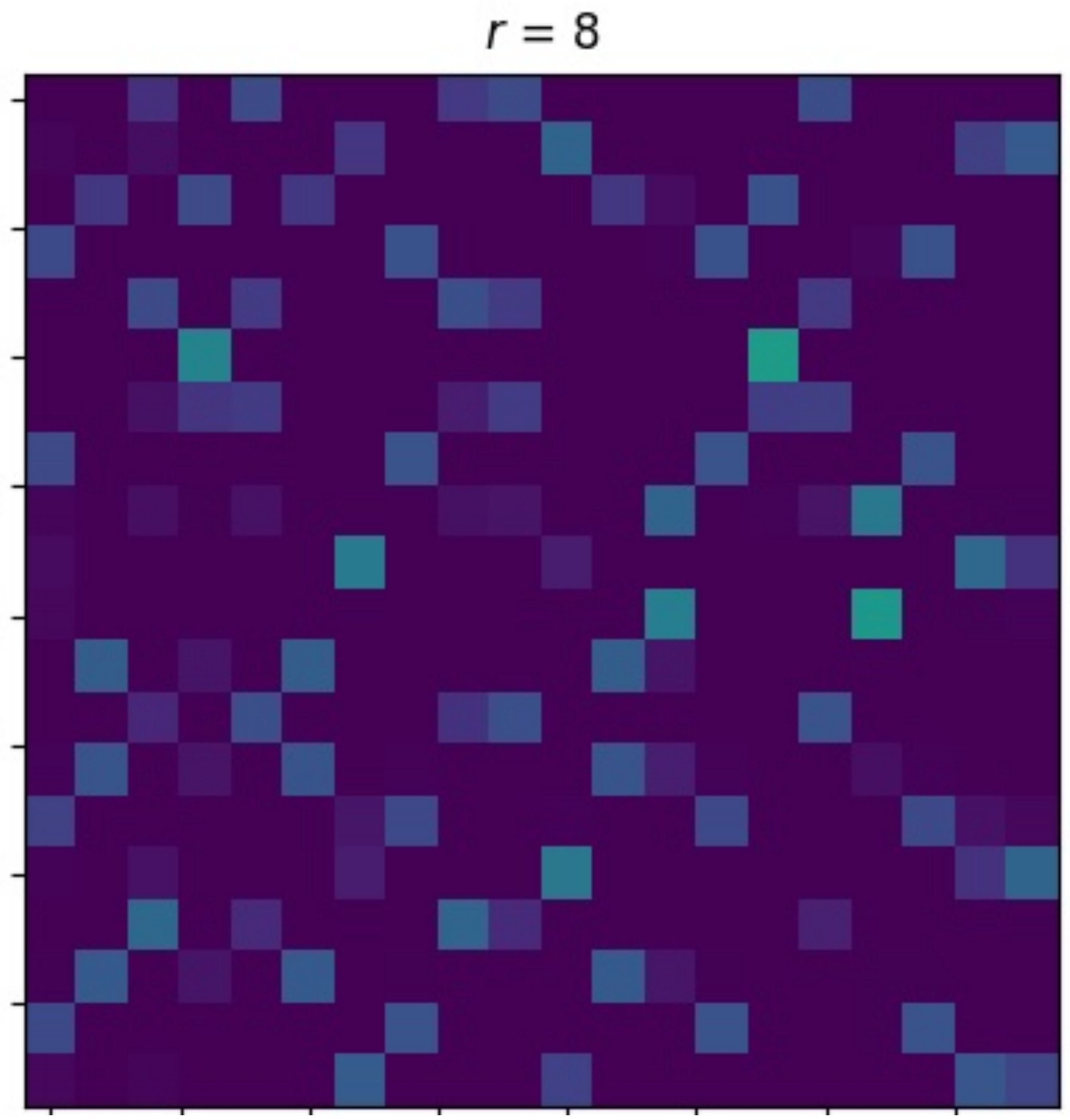
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{XY}} \rangle.$$

[Forrow+19, Scetbon+21]





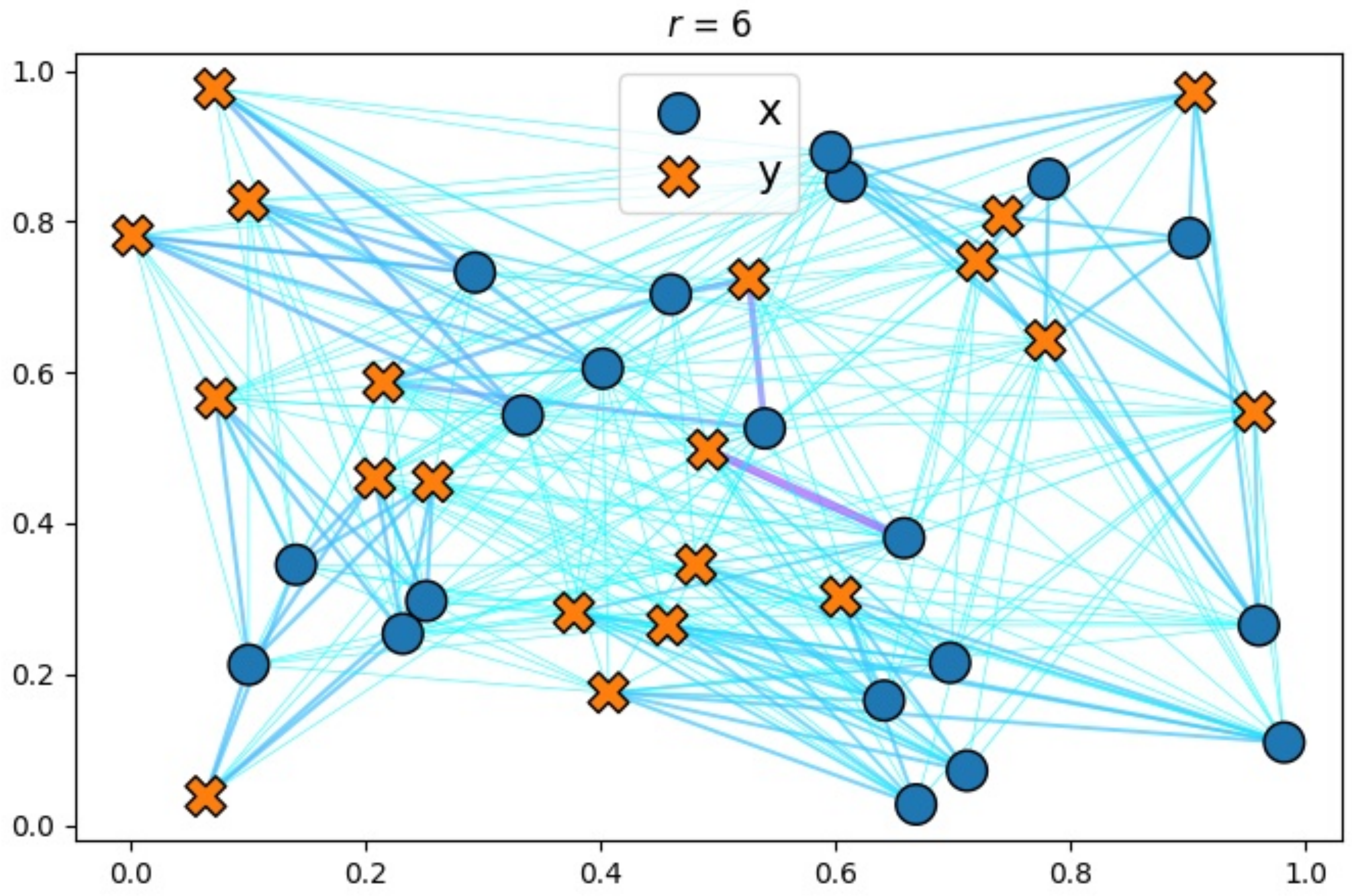


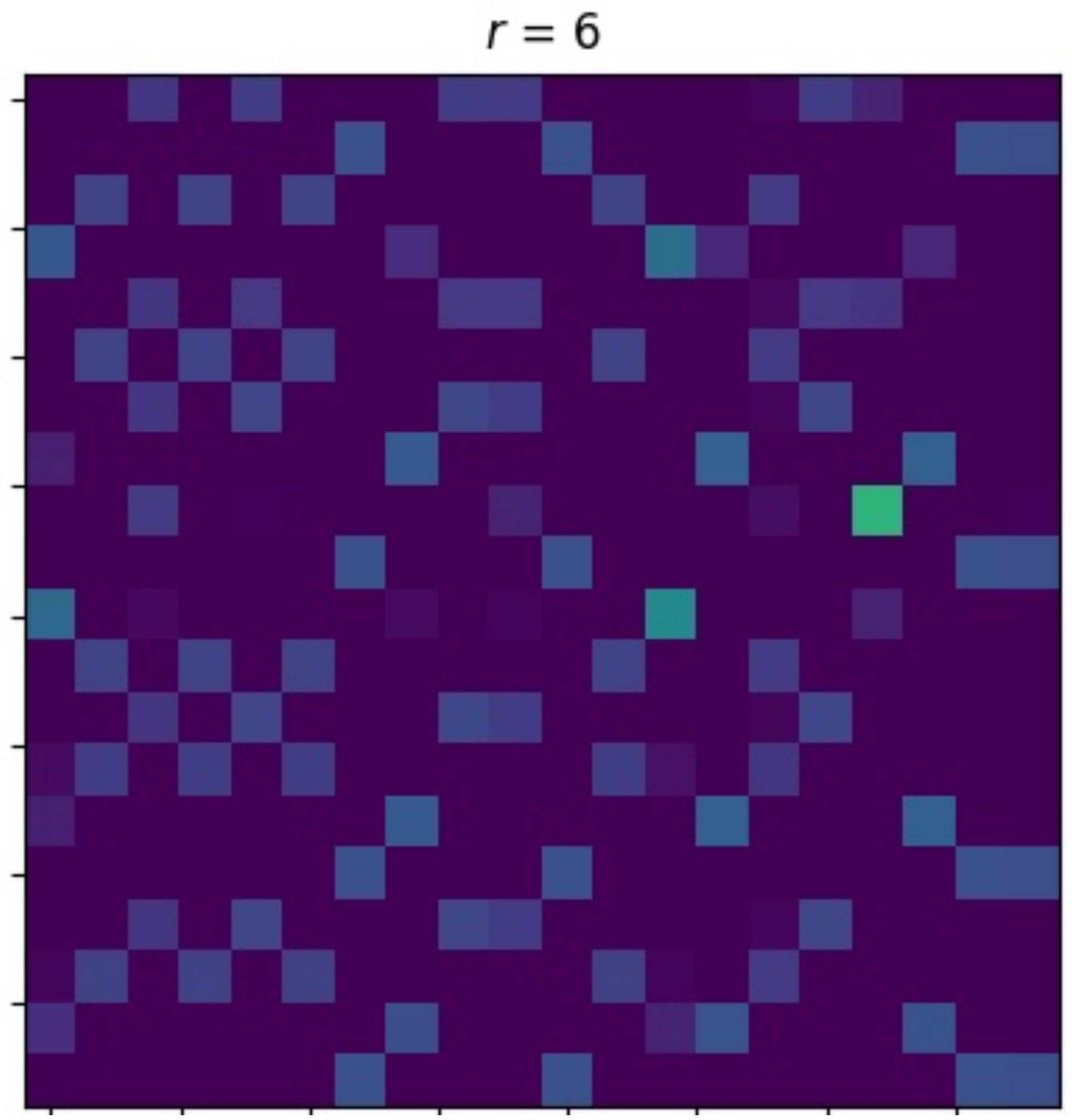
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a}, \boldsymbol{g}),\\ R \in U(\boldsymbol{b}, \boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle.$$

[Forrow+19, Scetbon+21]

[Colab4]

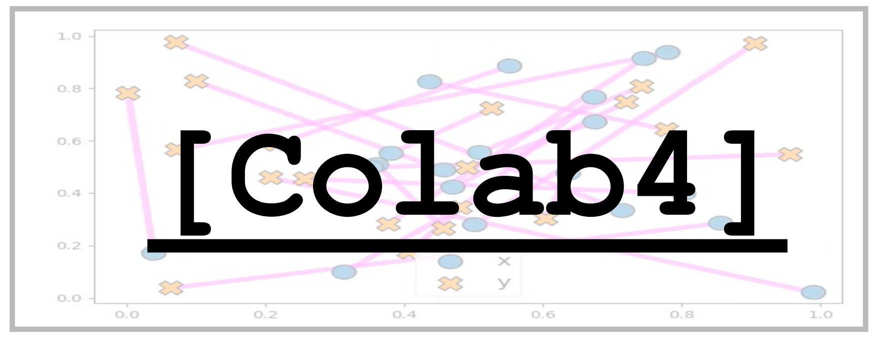


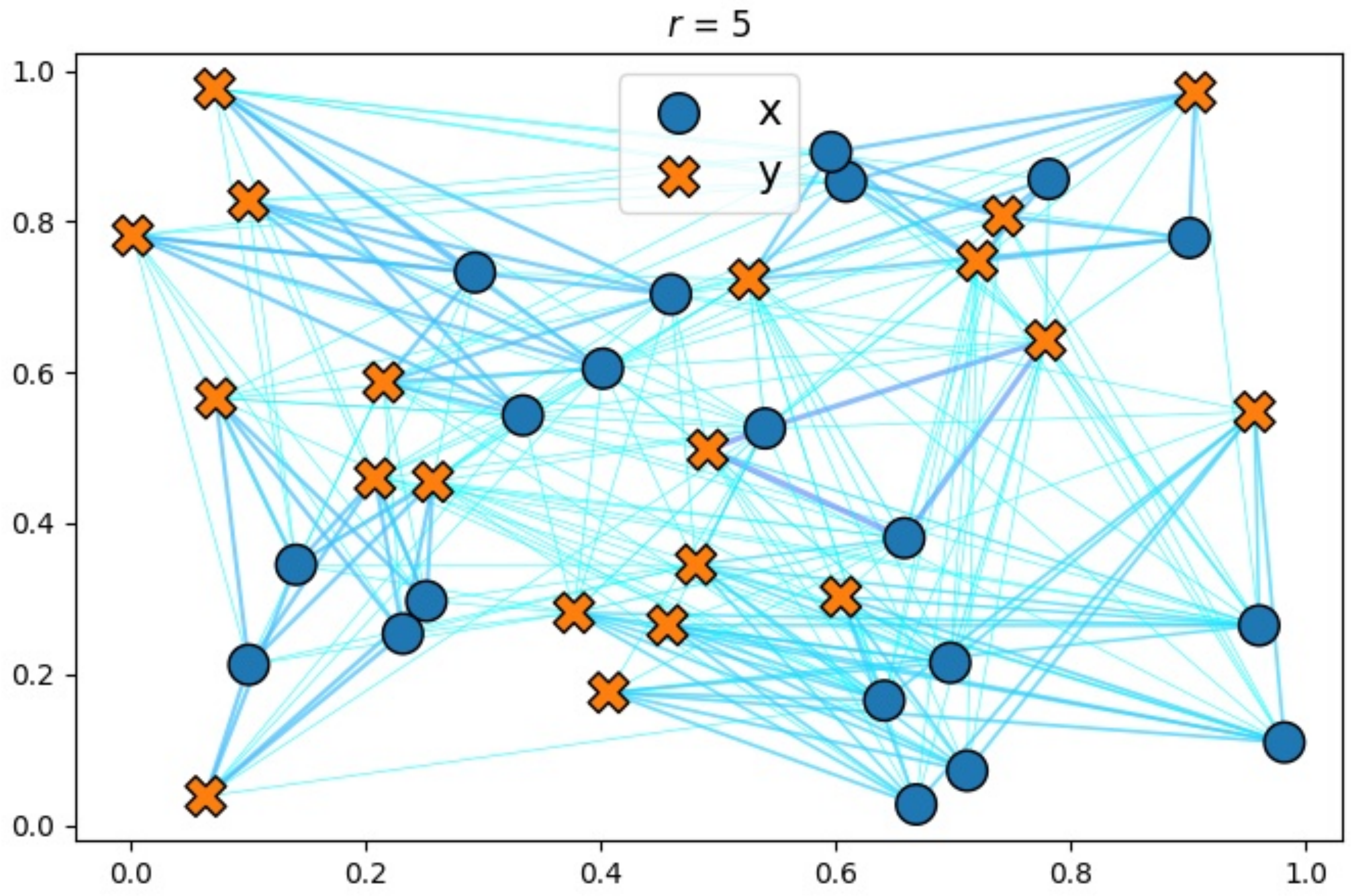


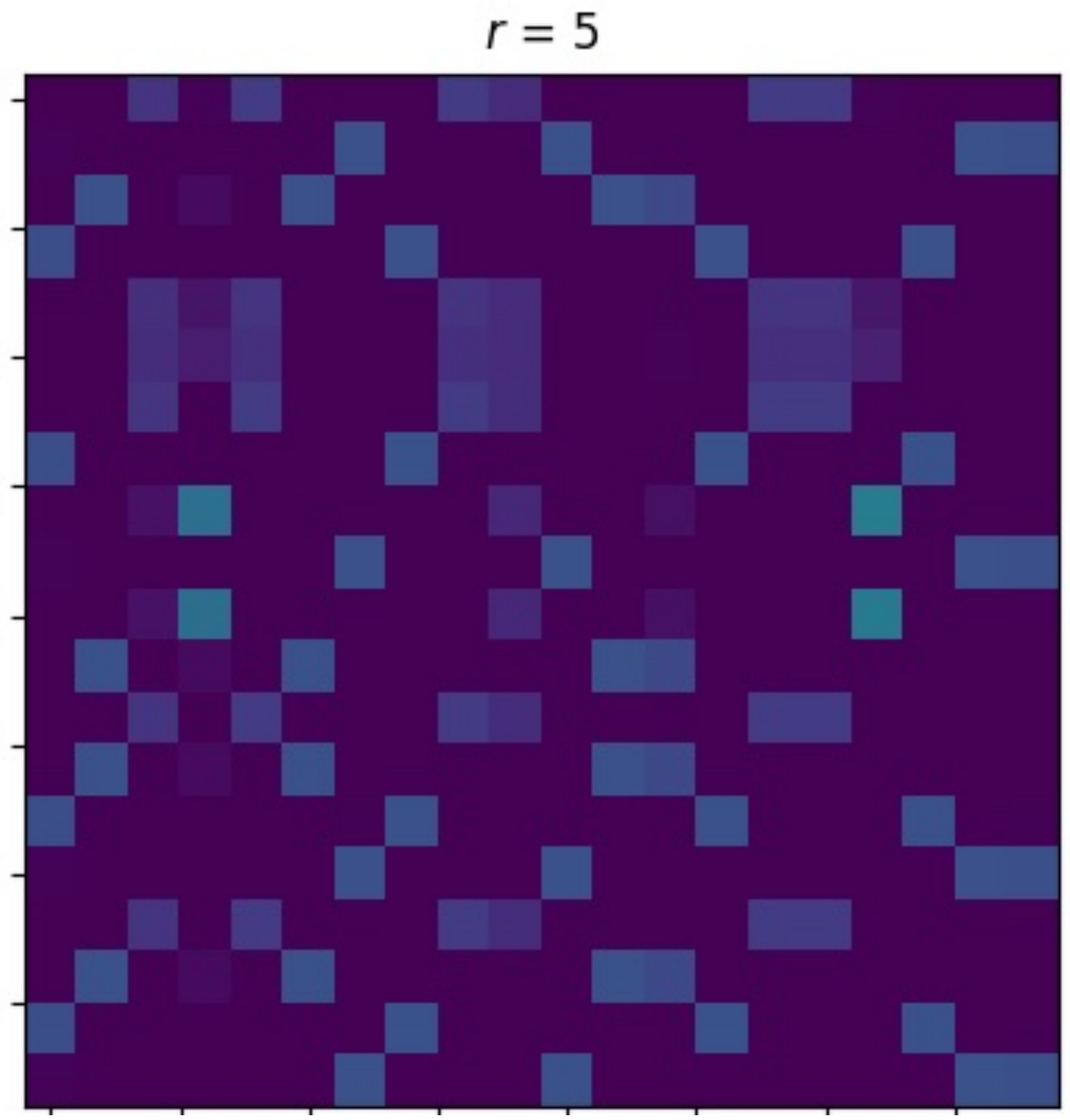
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}),\\ R\in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g}\in\Sigma_r}} \langle\, Q D(1/\boldsymbol{g})\, R^T, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle.$$

[Forrow+19, Scetbon+21]



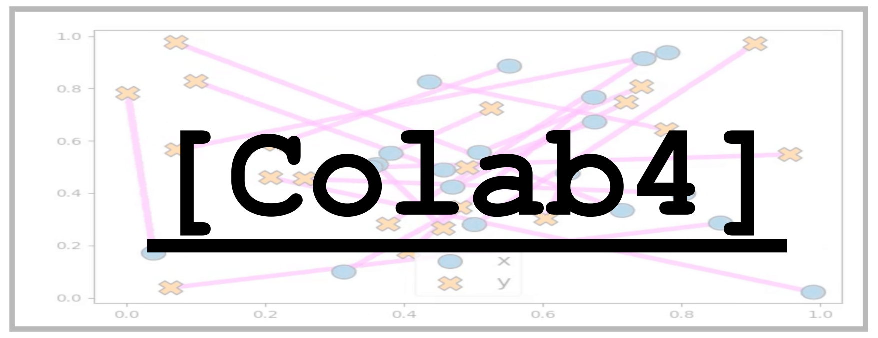


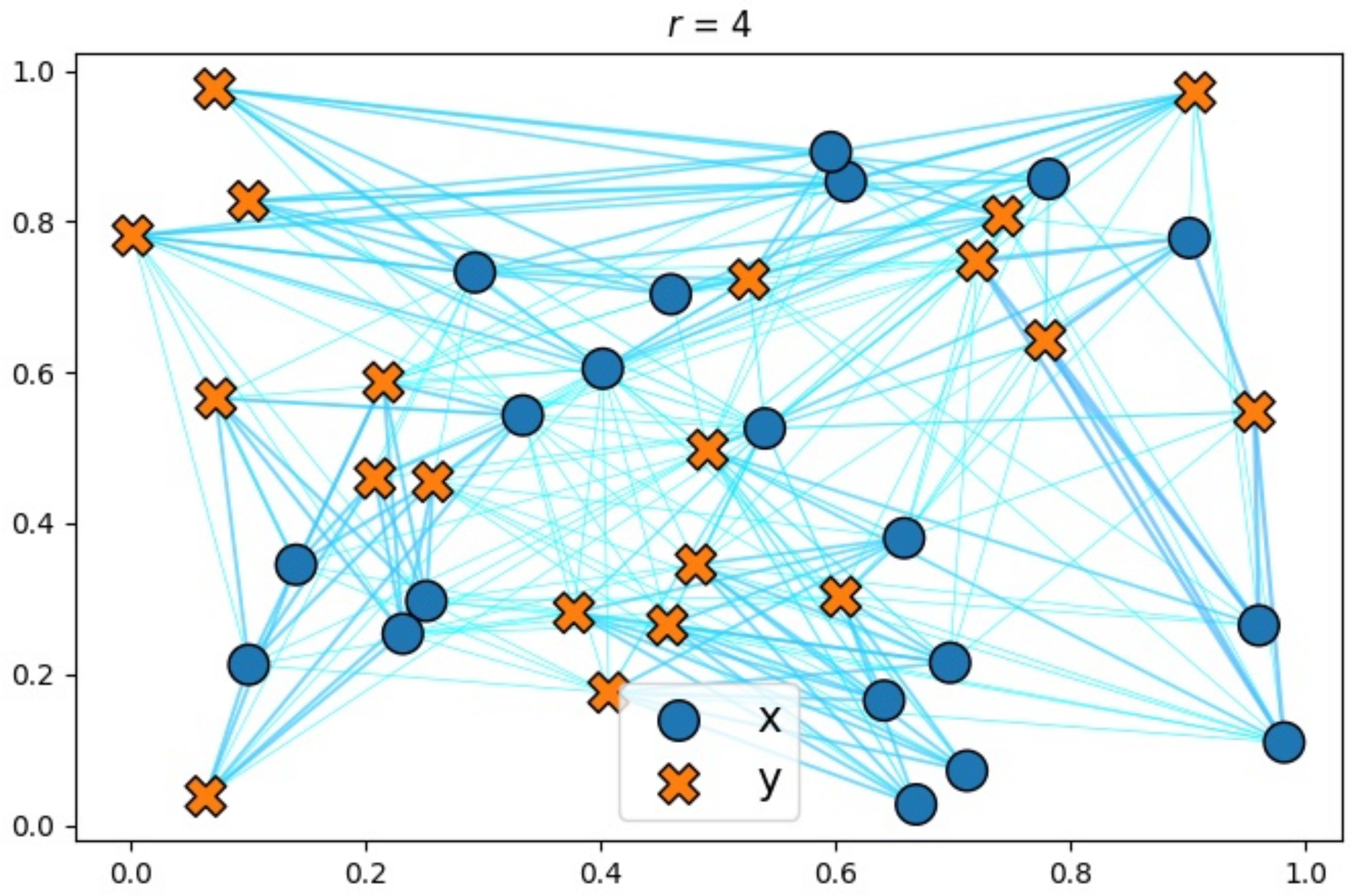


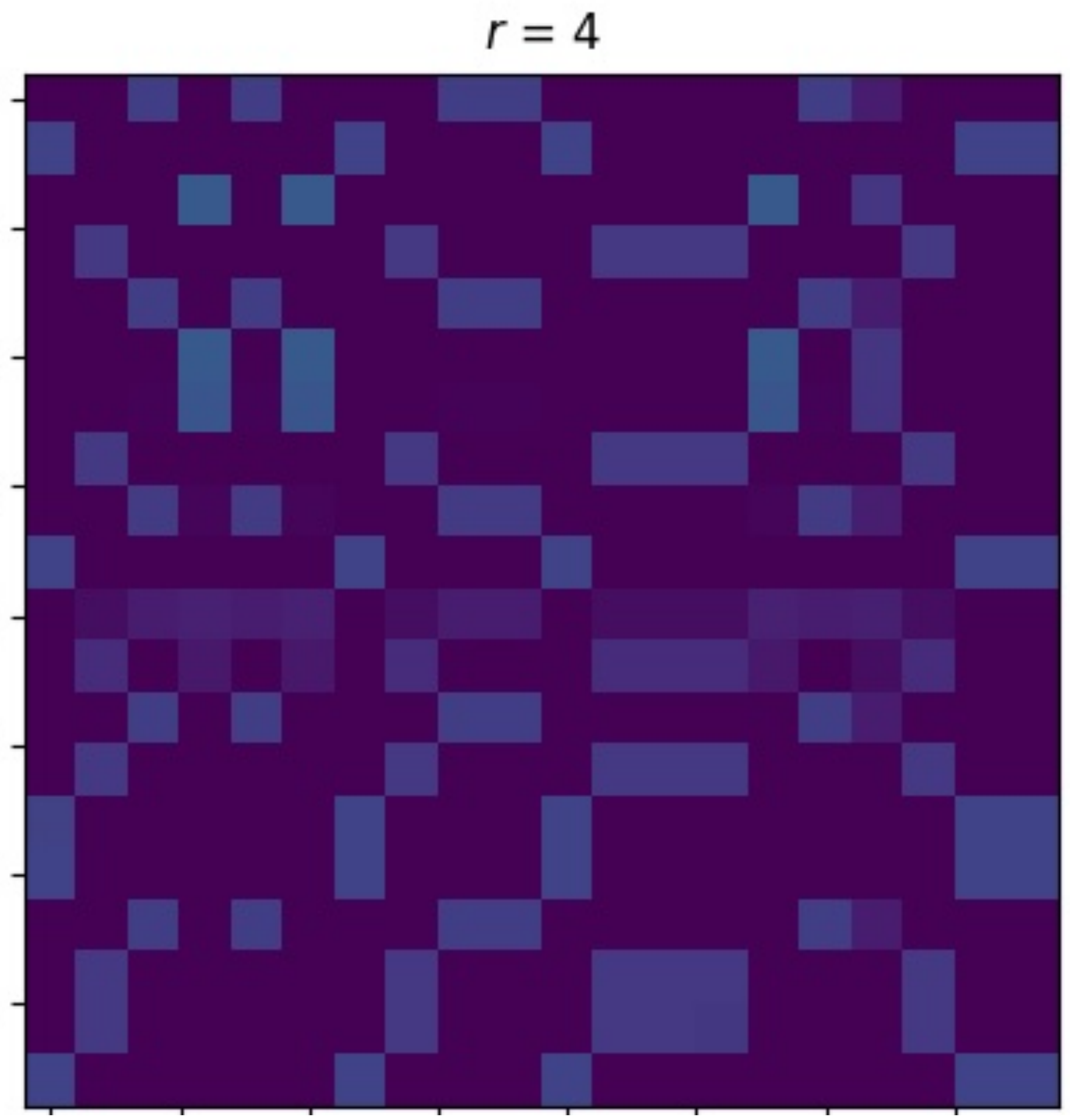
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}),\\ R\in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g}\in\Sigma_r}} \langle QD(1/\boldsymbol{g})\, R^T, M_{\boldsymbol{XY}}\rangle.$$

[Forrow+19, Scetbon+21]



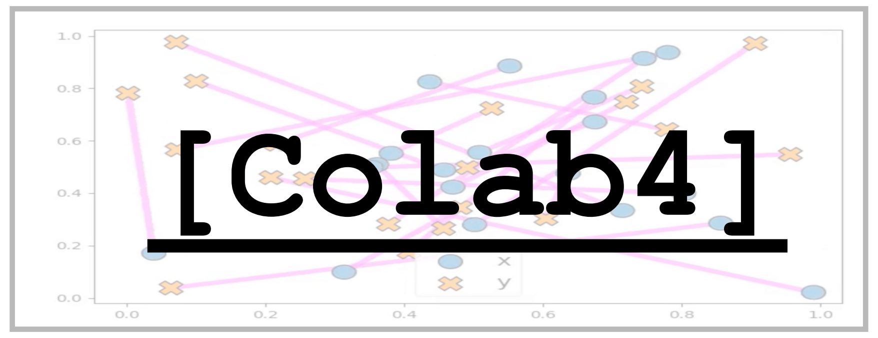


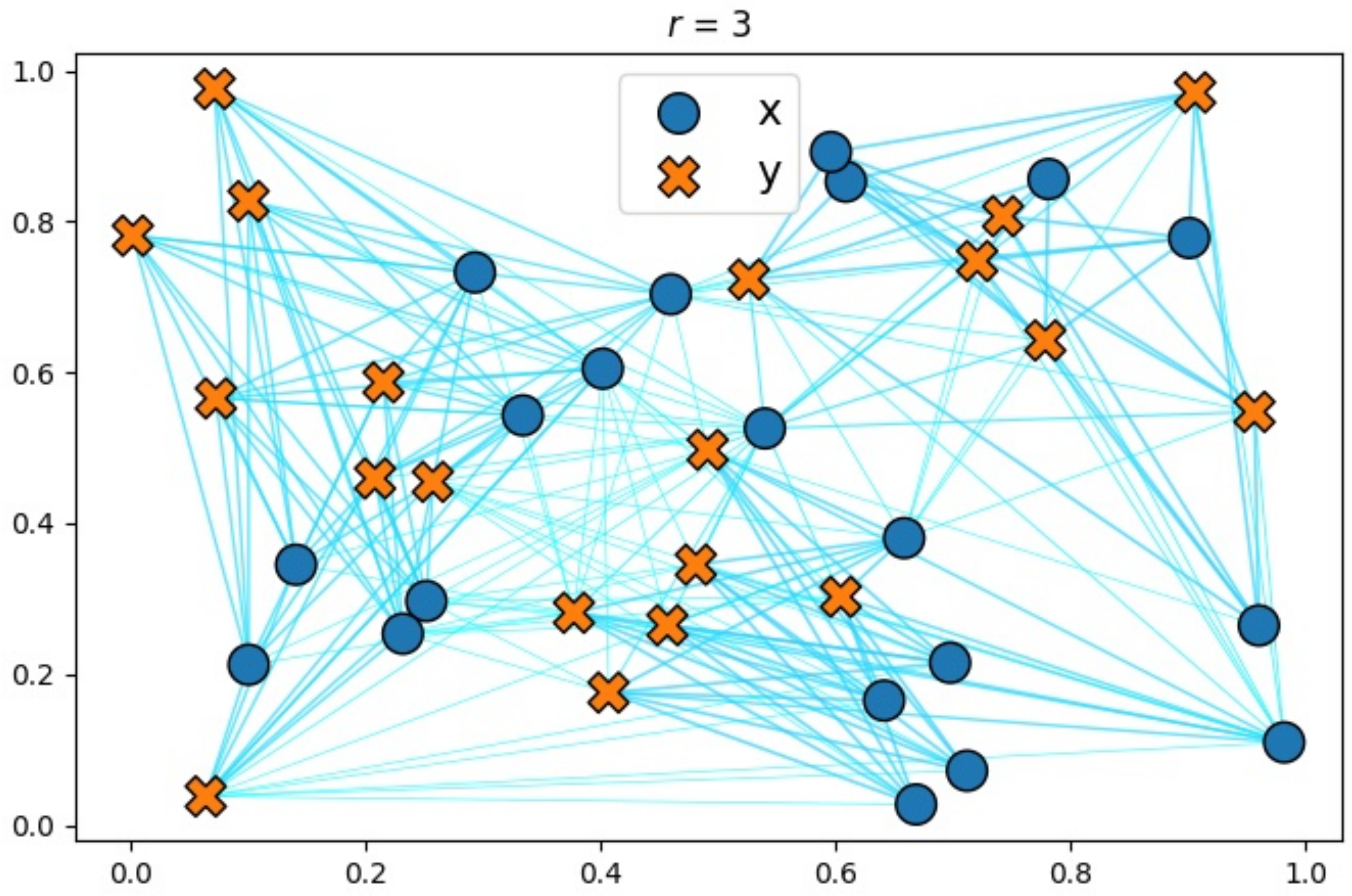


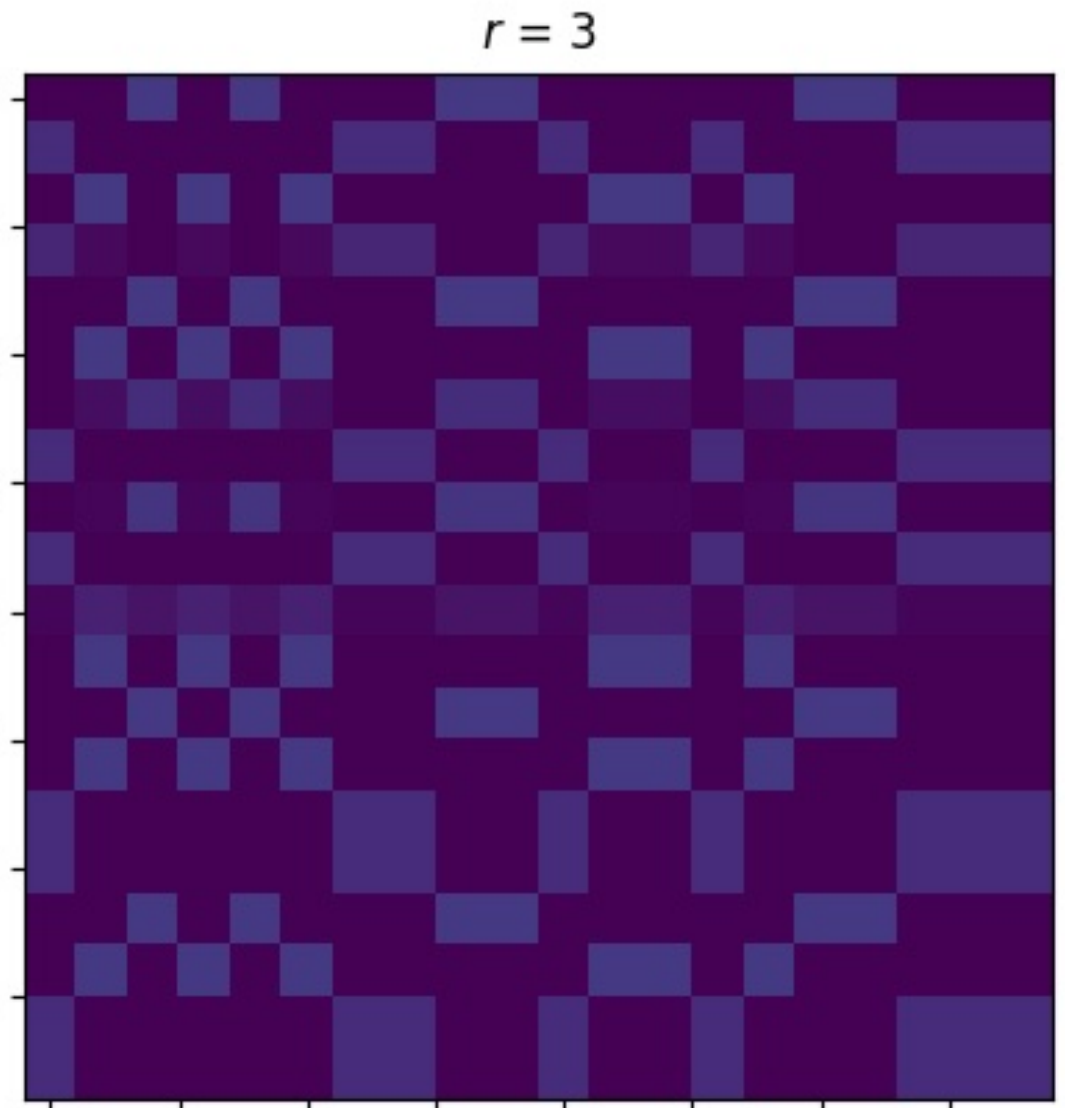
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}),\\ R\in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g}\in\Sigma_r}} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\boldsymbol{X}\boldsymbol{Y}}\rangle.$$

[Forrow+19, Scetbon+21]



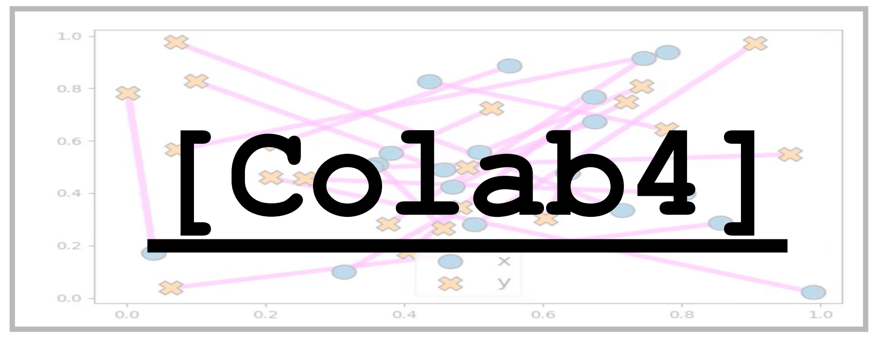


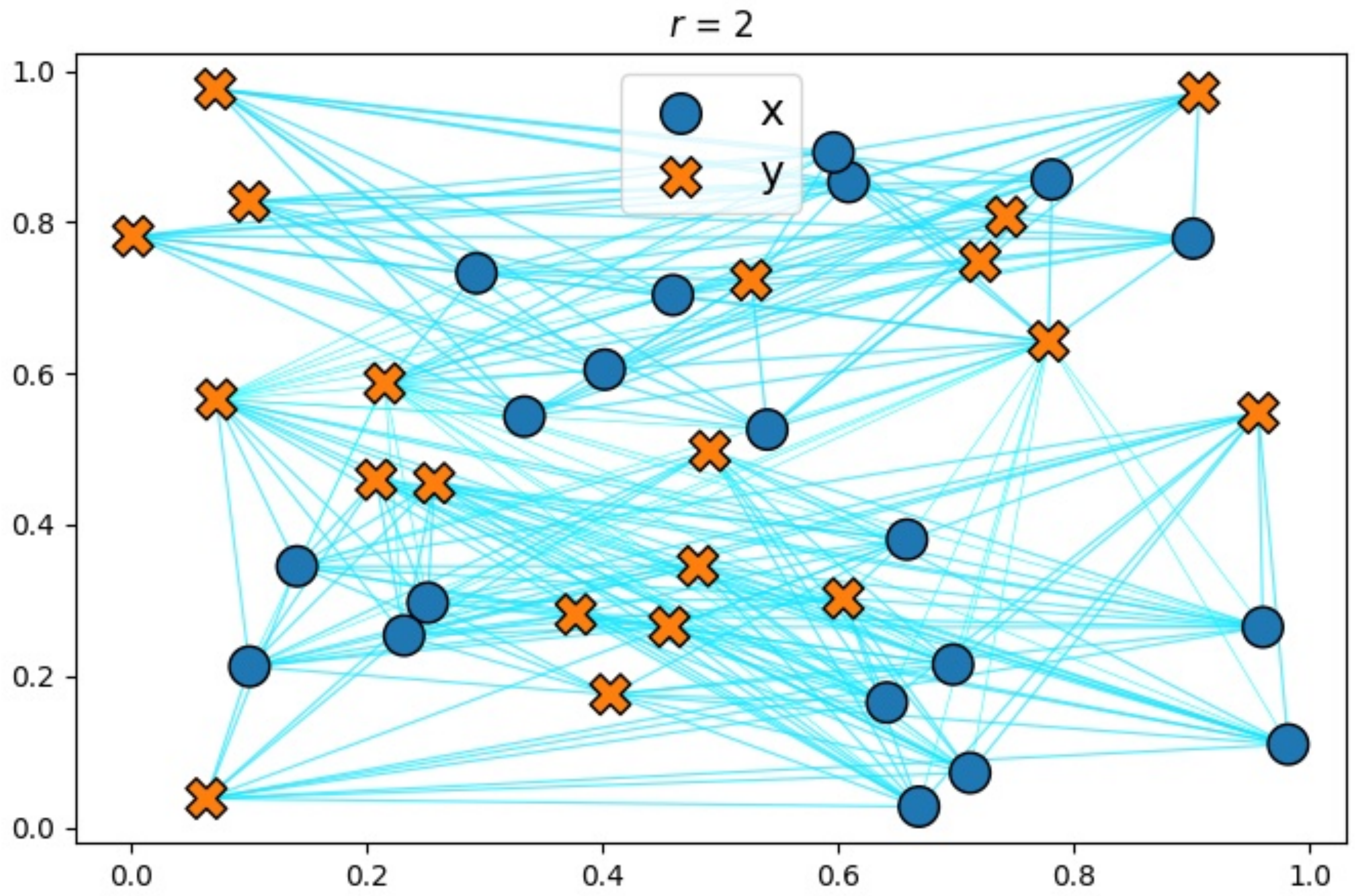


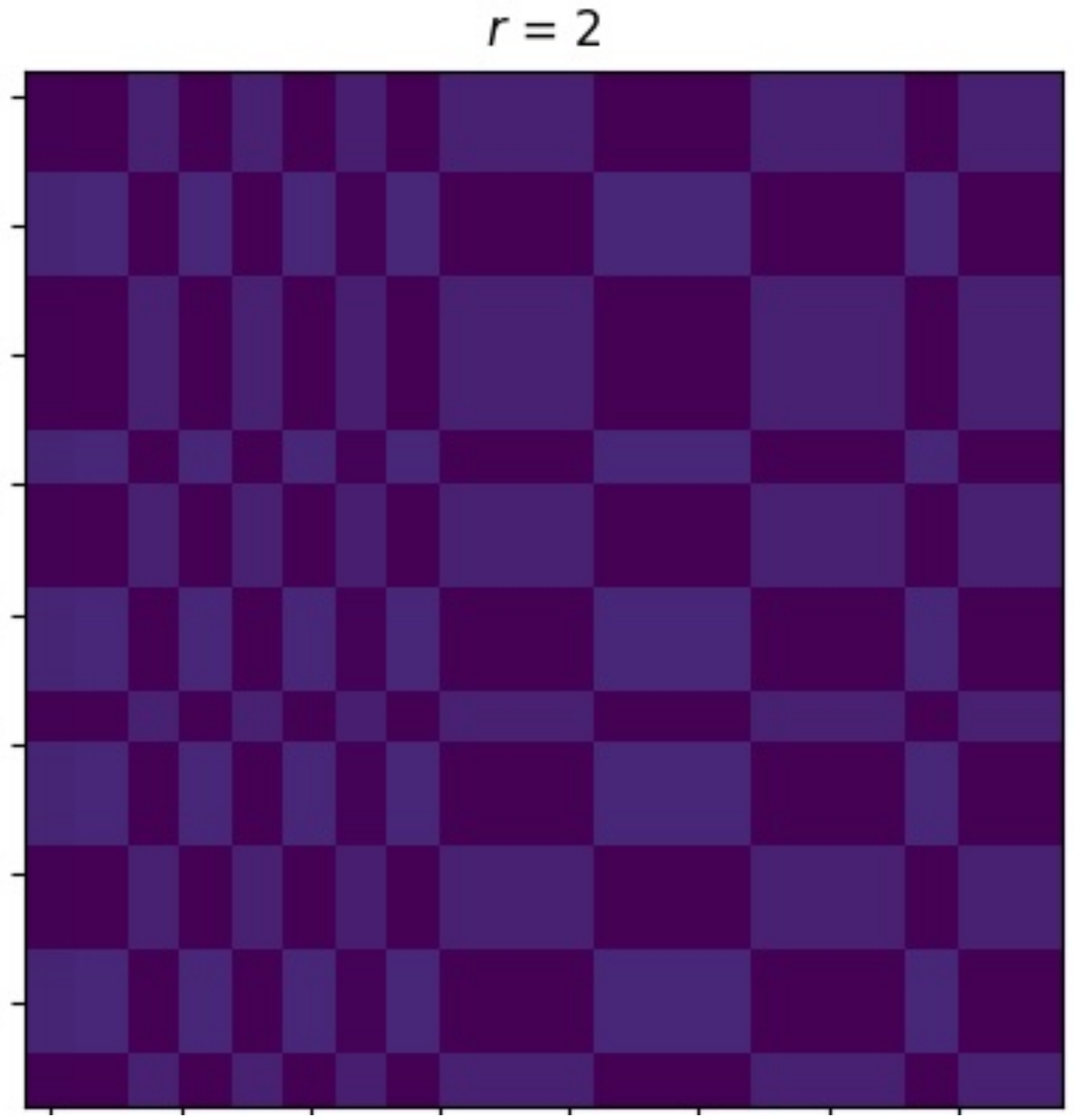
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}),\\ R\in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g}\in\Sigma_r}} \langle Q D(1/\boldsymbol{g})\, R^T, M_{\boldsymbol{X}\boldsymbol{Y}}\rangle.$$

[Forrow+19, Scetbon+21]



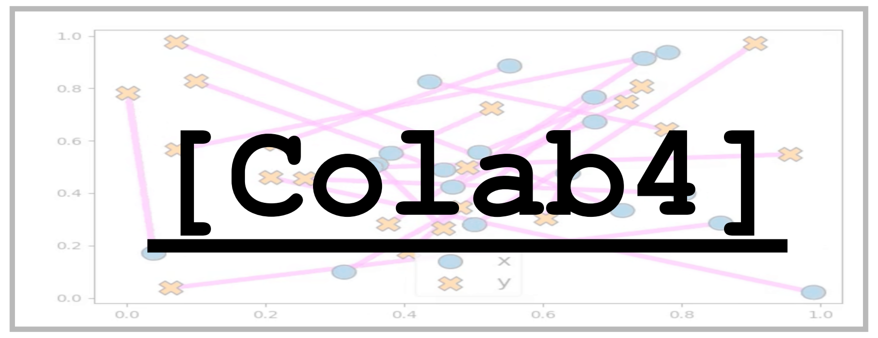


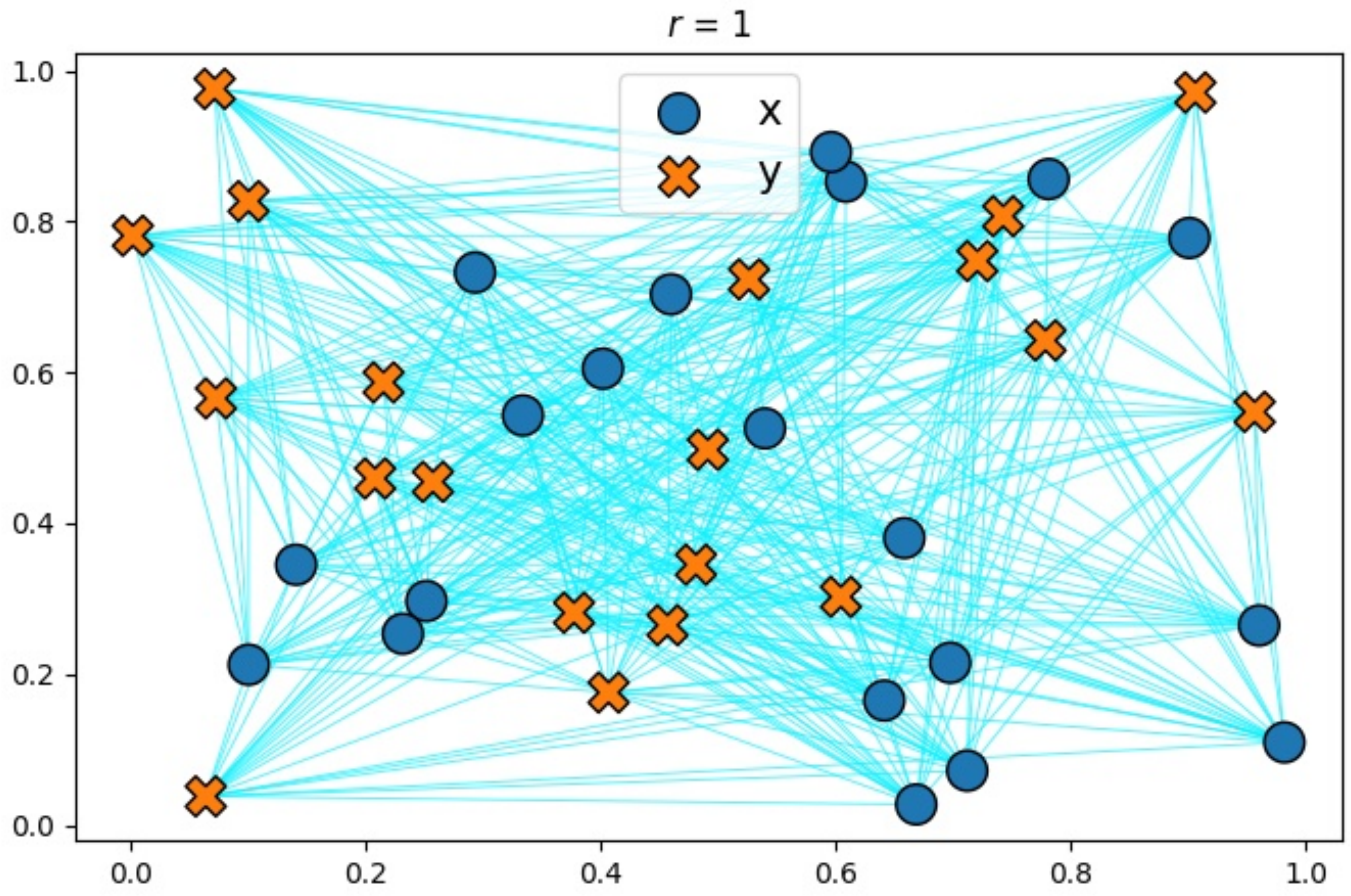


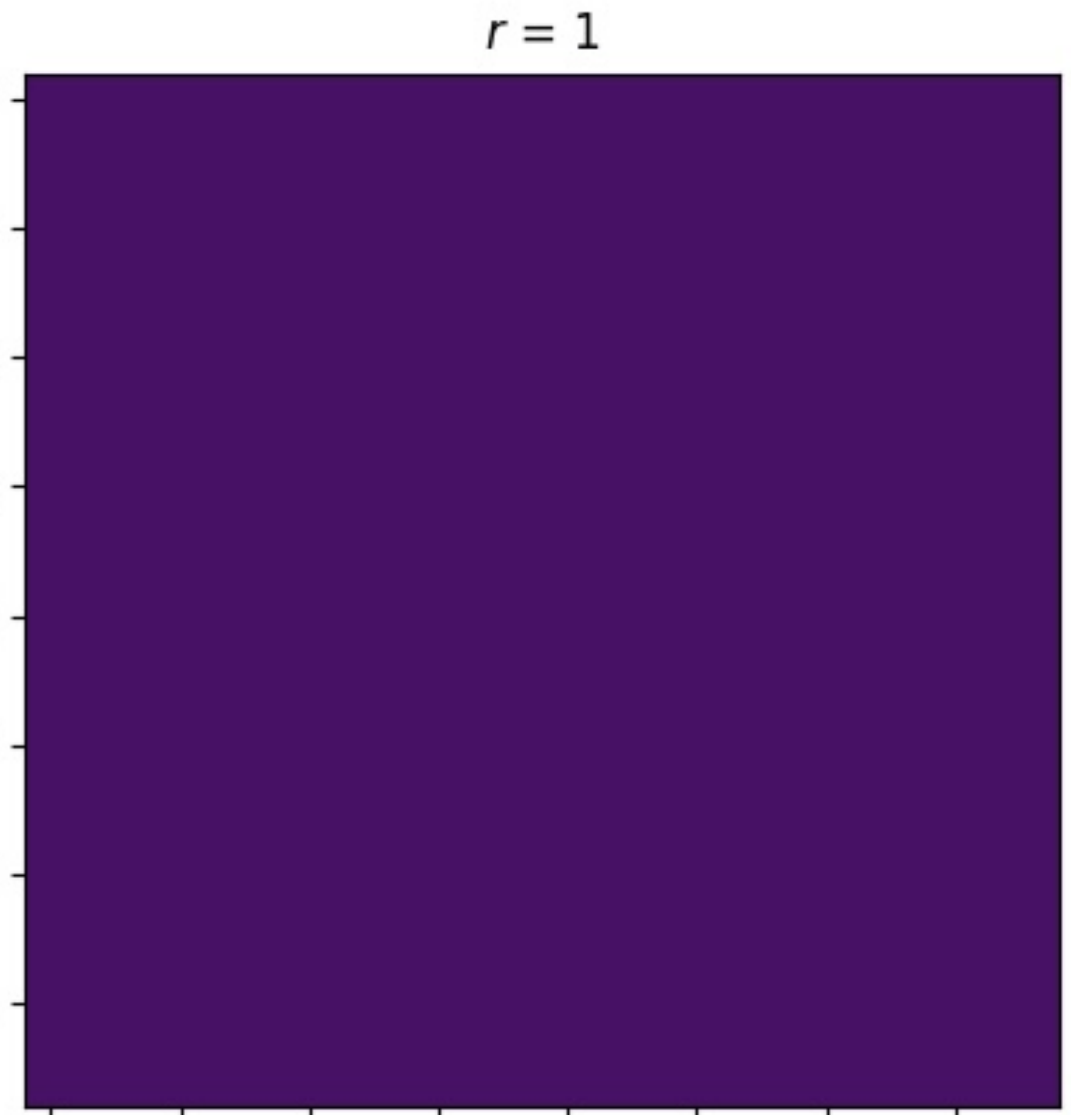
Low-Rank Constrained Discrete Optimal Transport Problem

$$\min_{\substack{Q \in U(\boldsymbol{a},\boldsymbol{g}),\\ R \in U(\boldsymbol{b},\boldsymbol{g})\\ \boldsymbol{g} \in \Sigma_r}} \langle Q D(1/\boldsymbol{g}) R^T, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle.$$

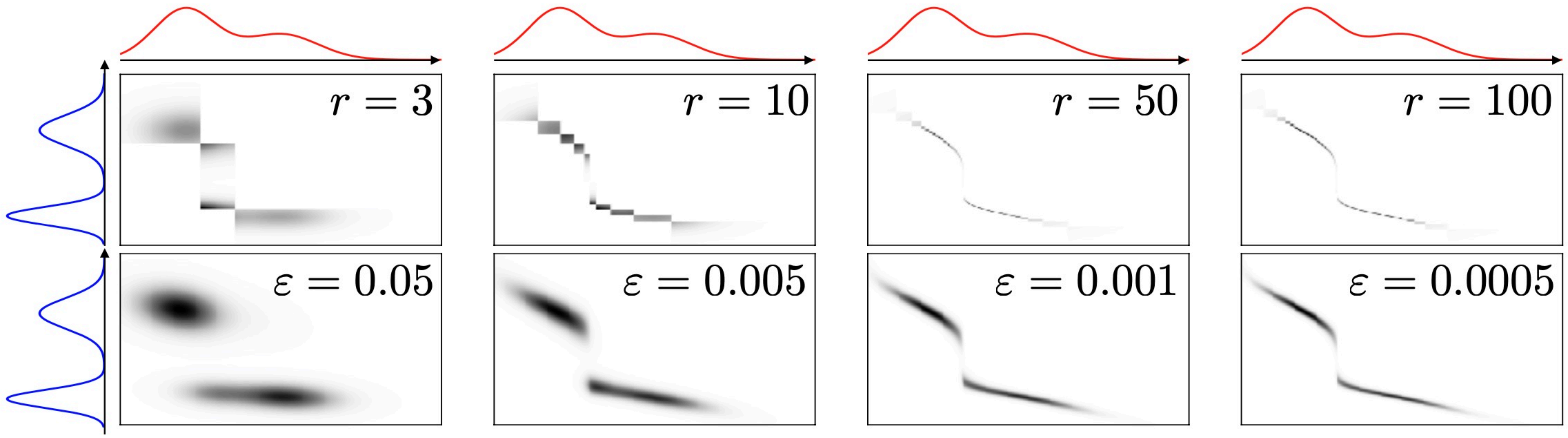
[Forrow+19, Scetbon+21]





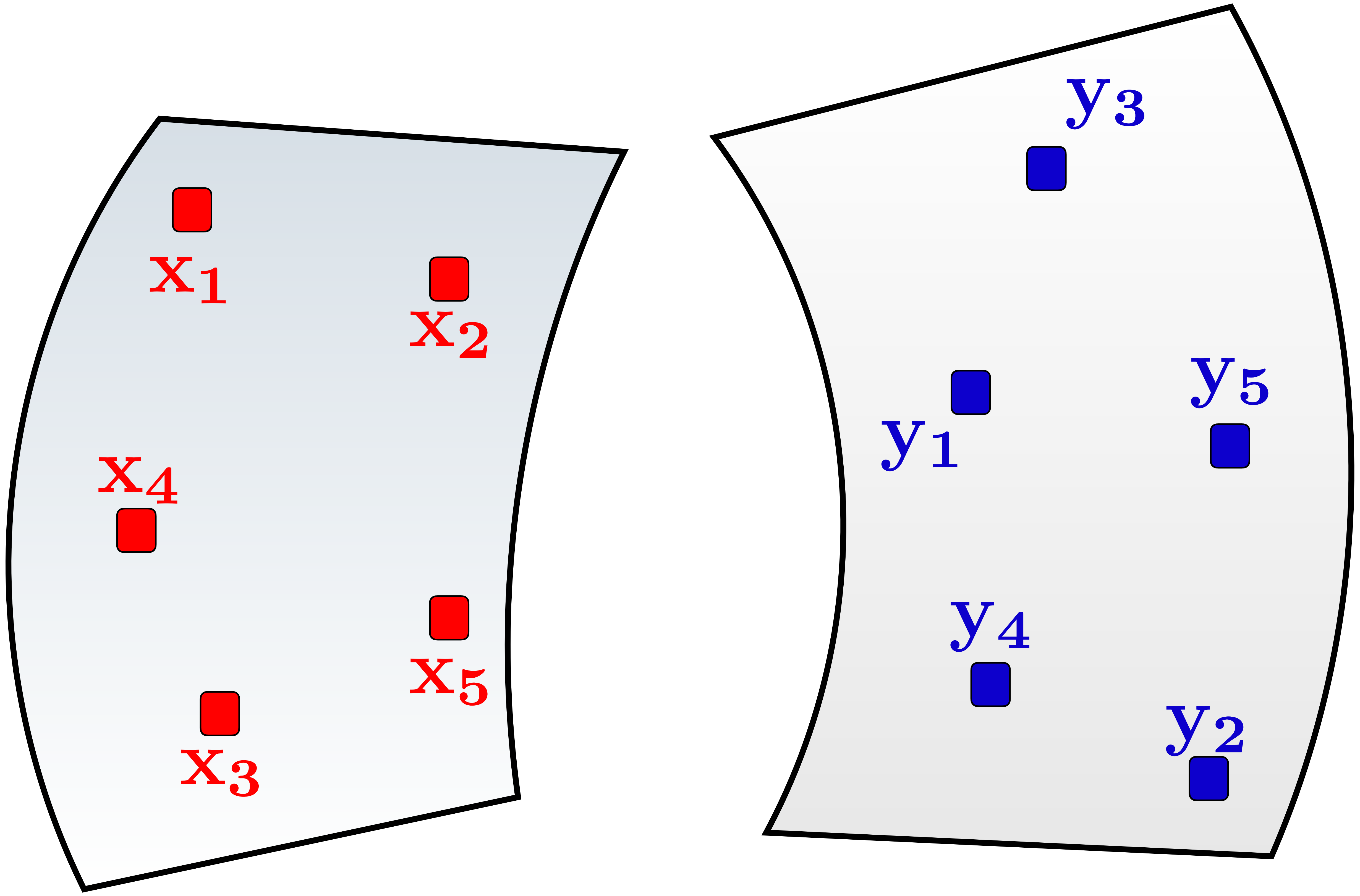


On a Simpler 1D Problem



The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2$$



Beckman

Koopmans

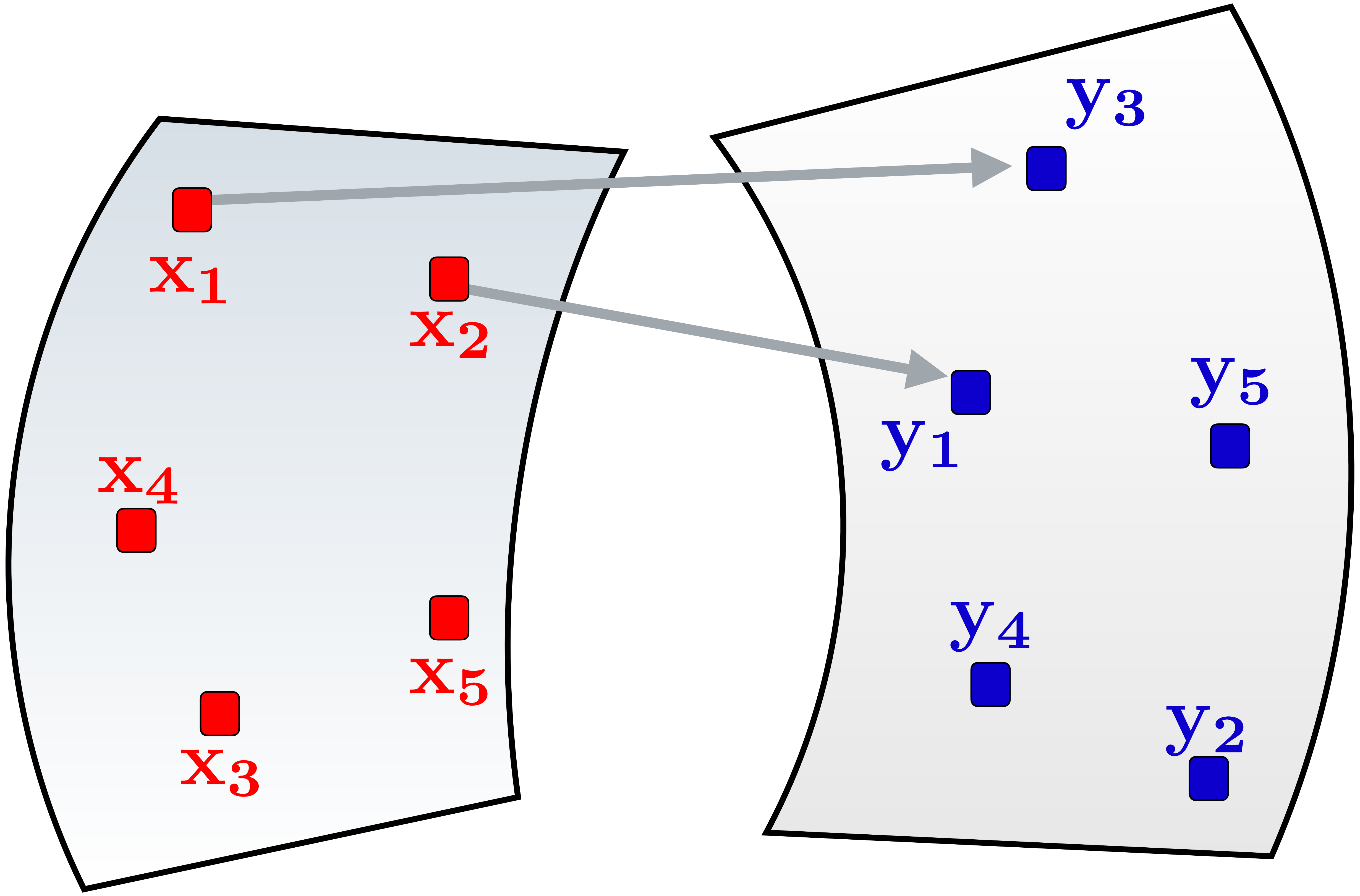
Memoli

Sturm

$$P^\star = \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\text{argmin}} \ \mathcal{E}(P)$$

The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathrm{x}_i, \mathrm{x}_{i'}) - c_2(\mathrm{y}_j, \mathrm{y}_{j'}))^2$$



$$P^\star = \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\operatorname{argmin}} \mathcal{E}(P)$$


Beckman

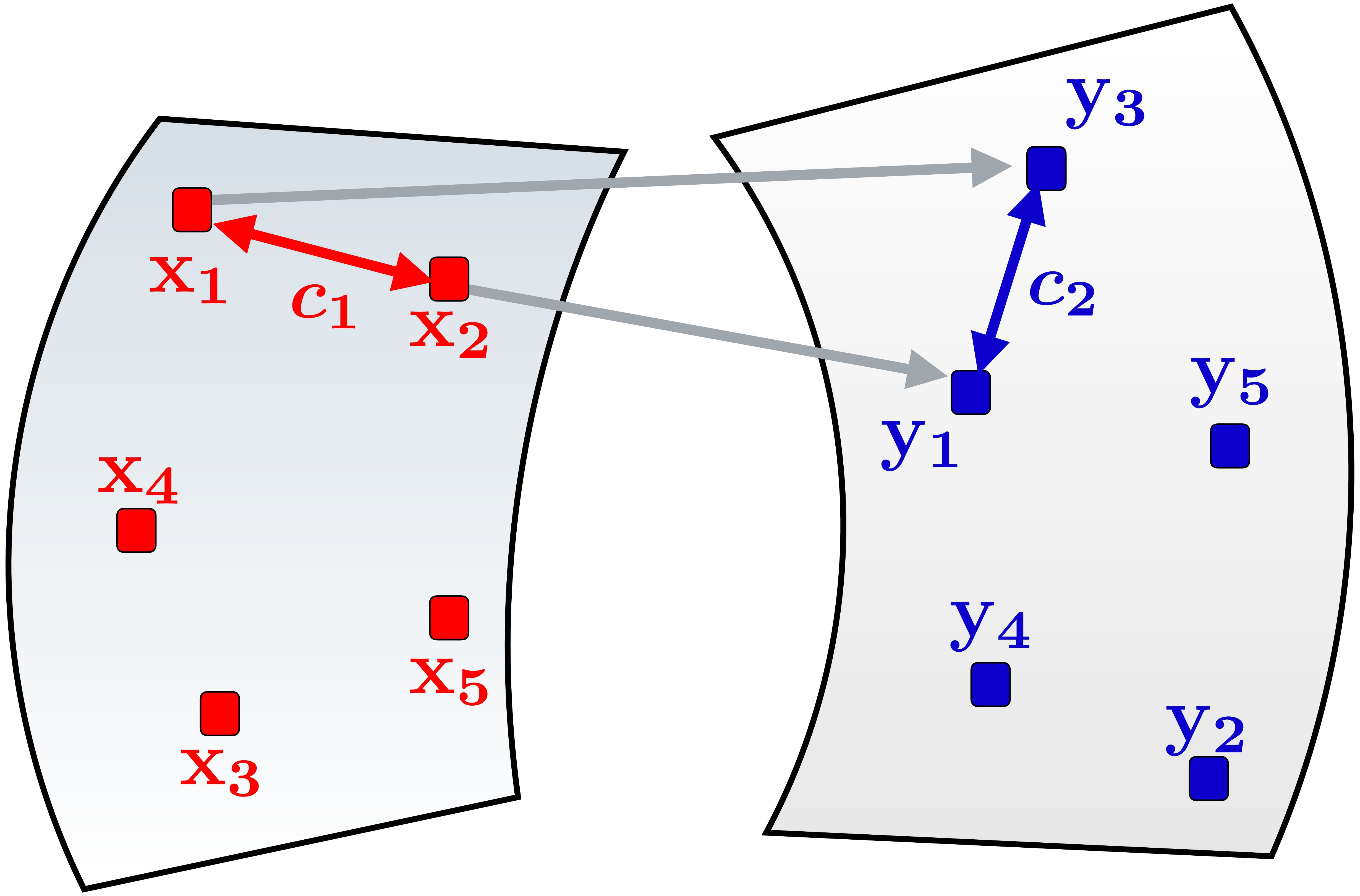

Koopmans

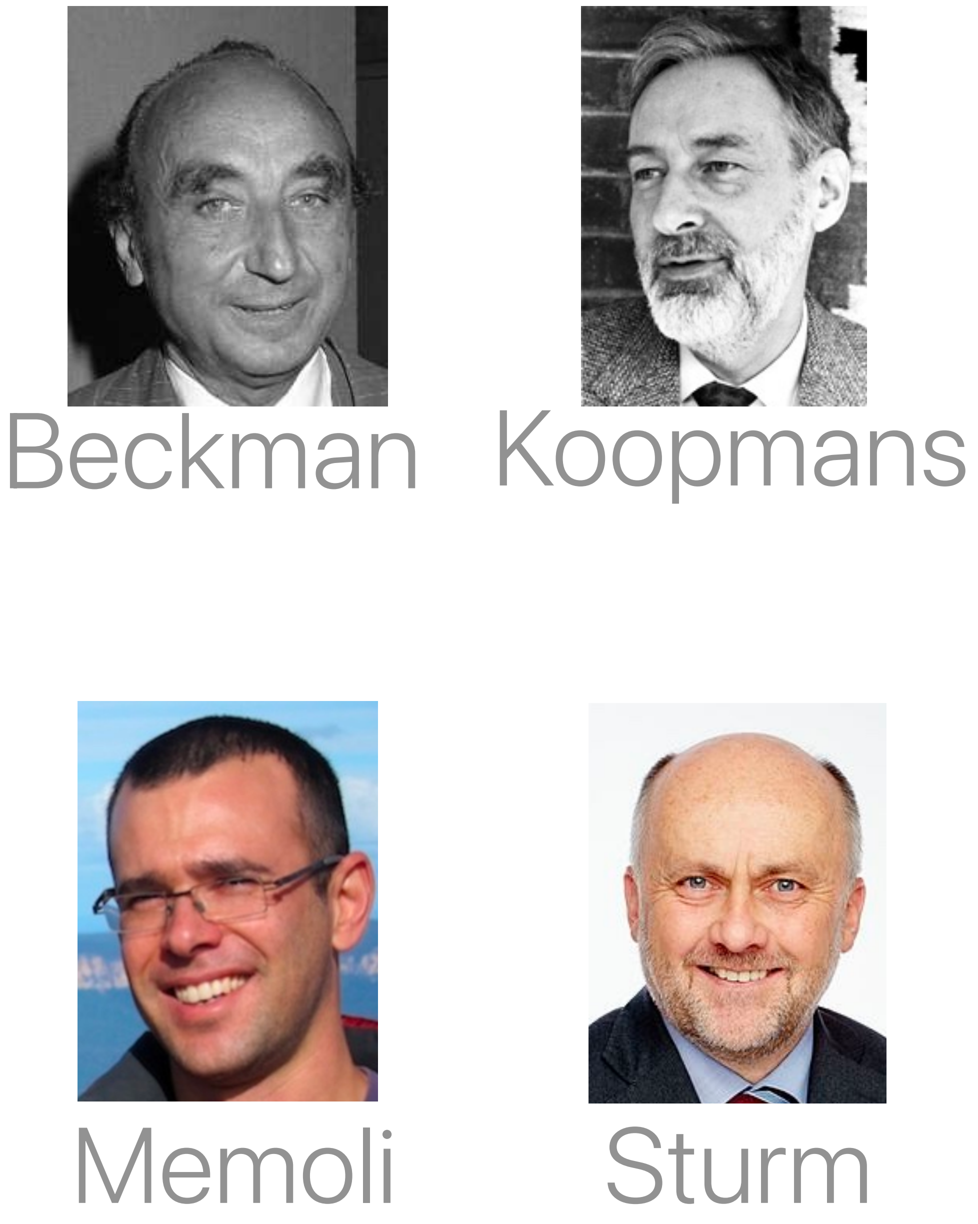

Memoli


Sturm

The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2$$



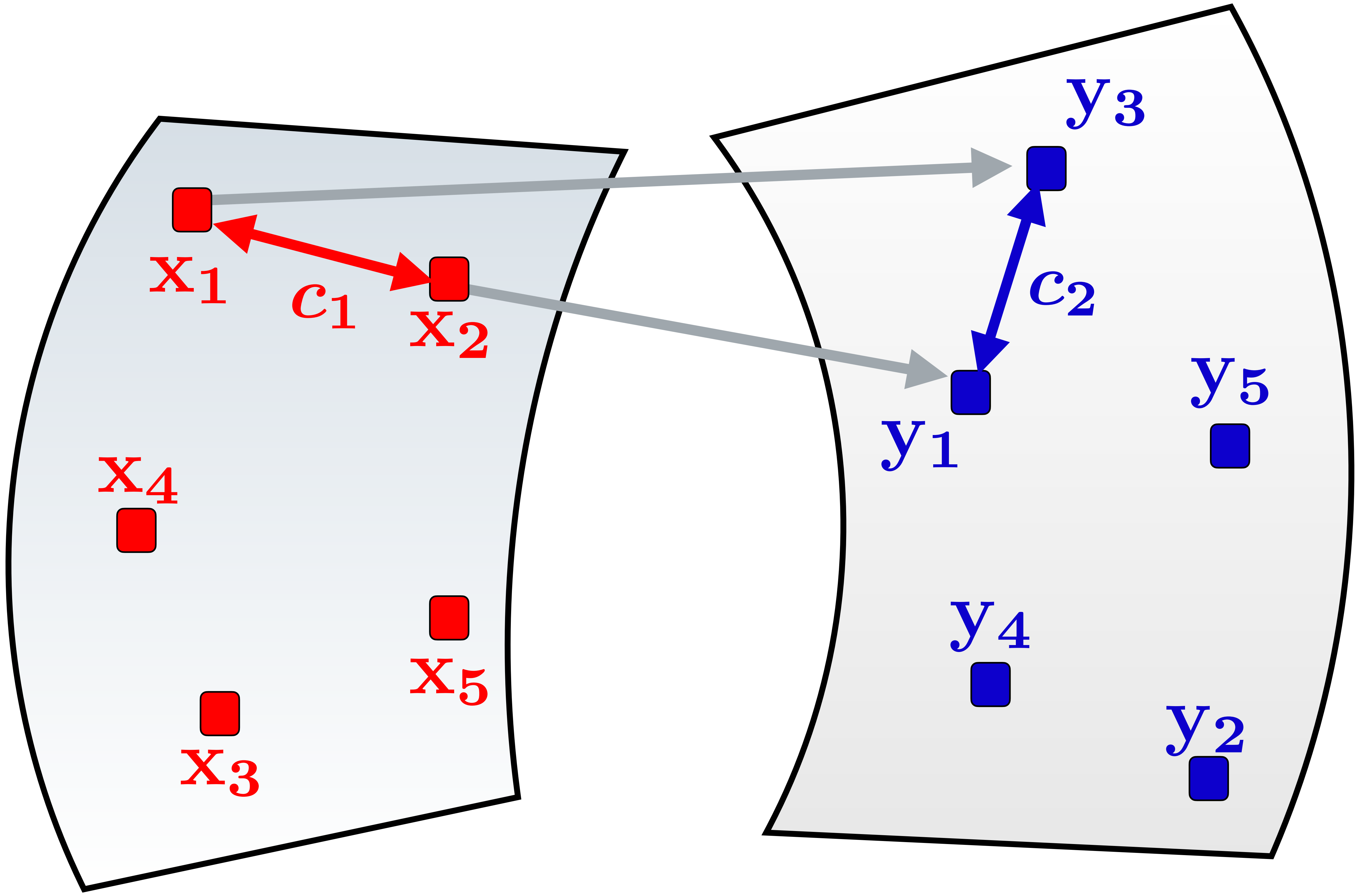


$$P^\star = \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\operatorname{argmin}} \; \mathcal{E}(P)$$

The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(\boldsymbol{P}) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2$$

$$\nabla \mathcal{E}(\boldsymbol{P}) \propto -\boldsymbol{C}_1 \boldsymbol{P} \boldsymbol{C}_2$$




Beckman


Koopmans


Memoli

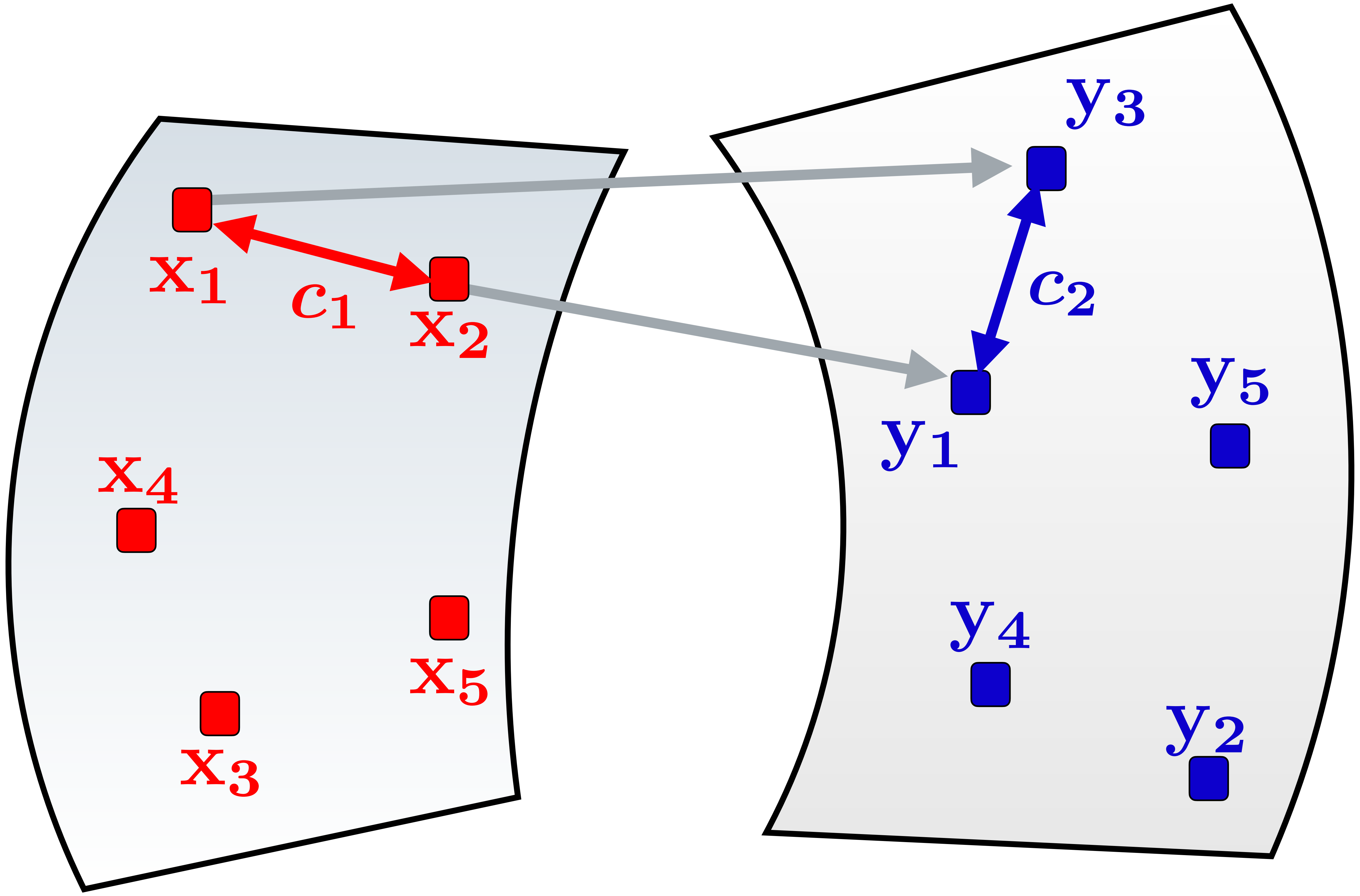

Sturm

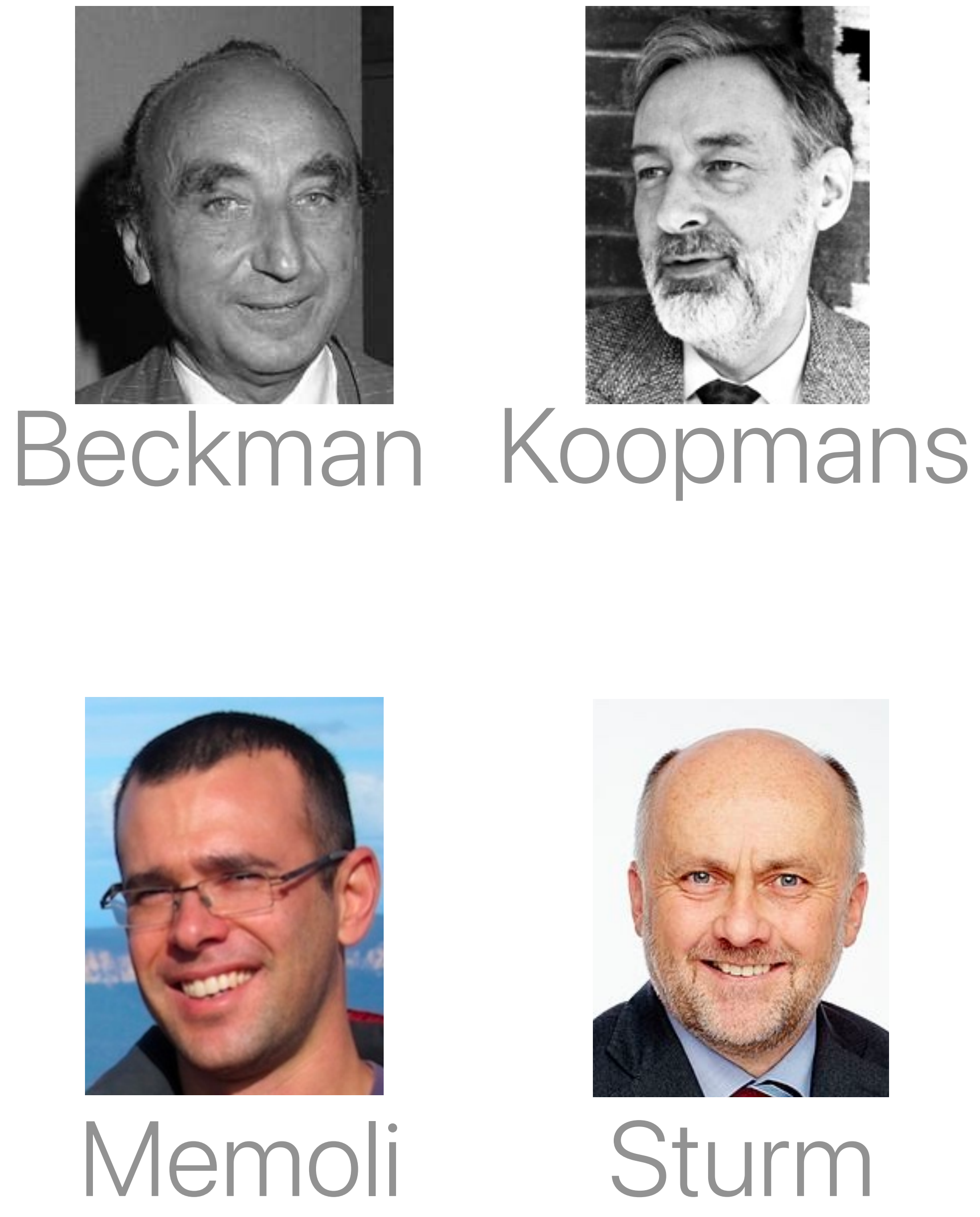
$$\boldsymbol{P}_{t+1} \leftarrow \underset{\boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T \mathbf{1}_n = \boldsymbol{b}}{\operatorname{argmin}} \langle \nabla \mathcal{E}(\boldsymbol{P}_t), \boldsymbol{P} \rangle$$

The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(x_i, x_{i'}) - c_2(y_j, y_{j'}))^2$$

$$\nabla \mathcal{E}(P) \propto -C_1 P C_2$$





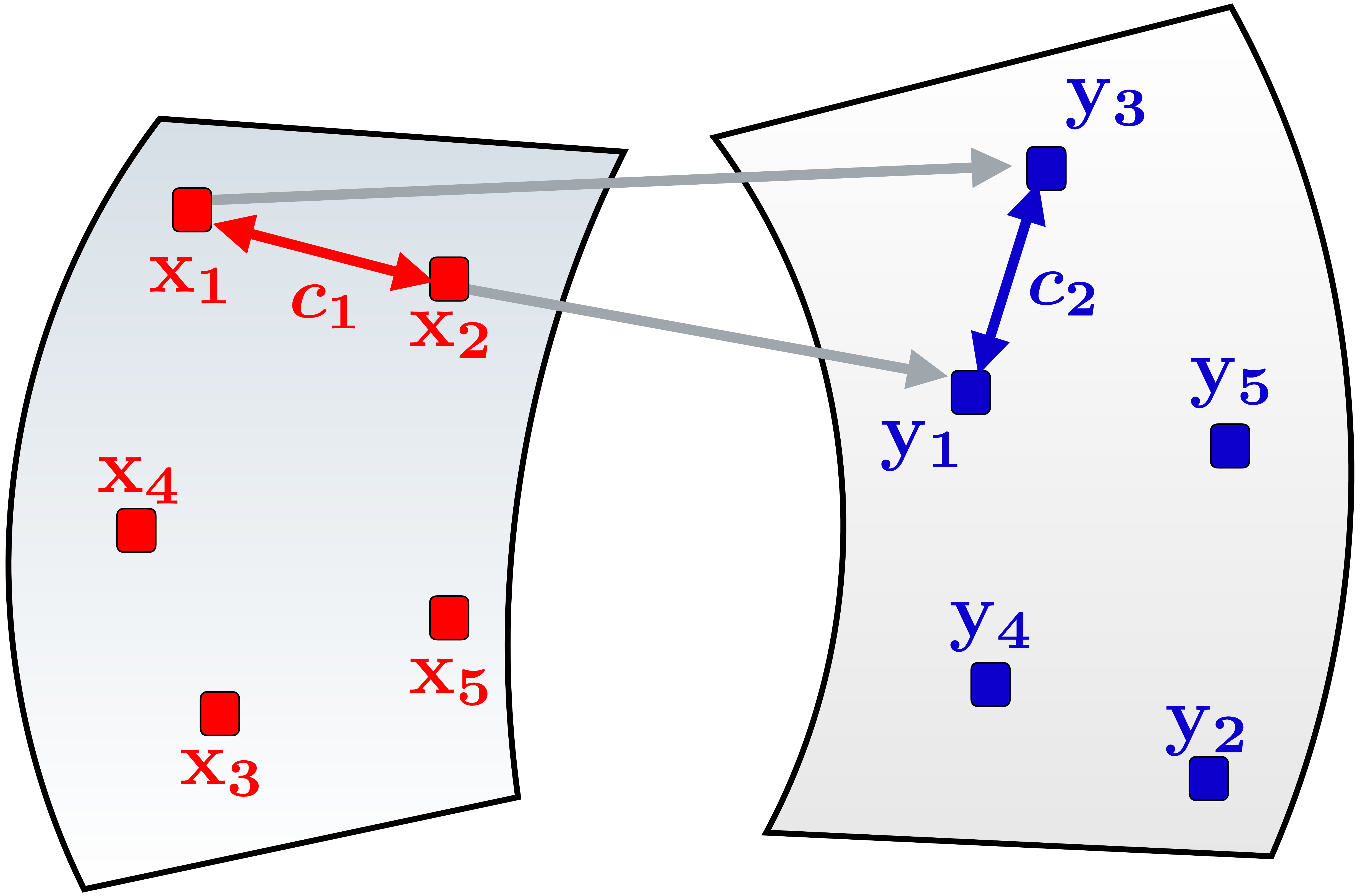
$$P_{t+1} \leftarrow \underset{P\mathbb{1}_m = a, P^T\mathbb{1}_n = b}{\text{argmin}} \langle \nabla \mathcal{E}(P_t), P \rangle$$

Any linear solver can be used

The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathrm{x}_i, \mathrm{x}_{i'}) - c_2(\mathrm{y}_j, \mathrm{y}_{j'}))^2$$

$$\nabla \mathcal{E}(P) \propto -C_1 P C_2$$




Beckman


Koopmans


Memoli


Sturm

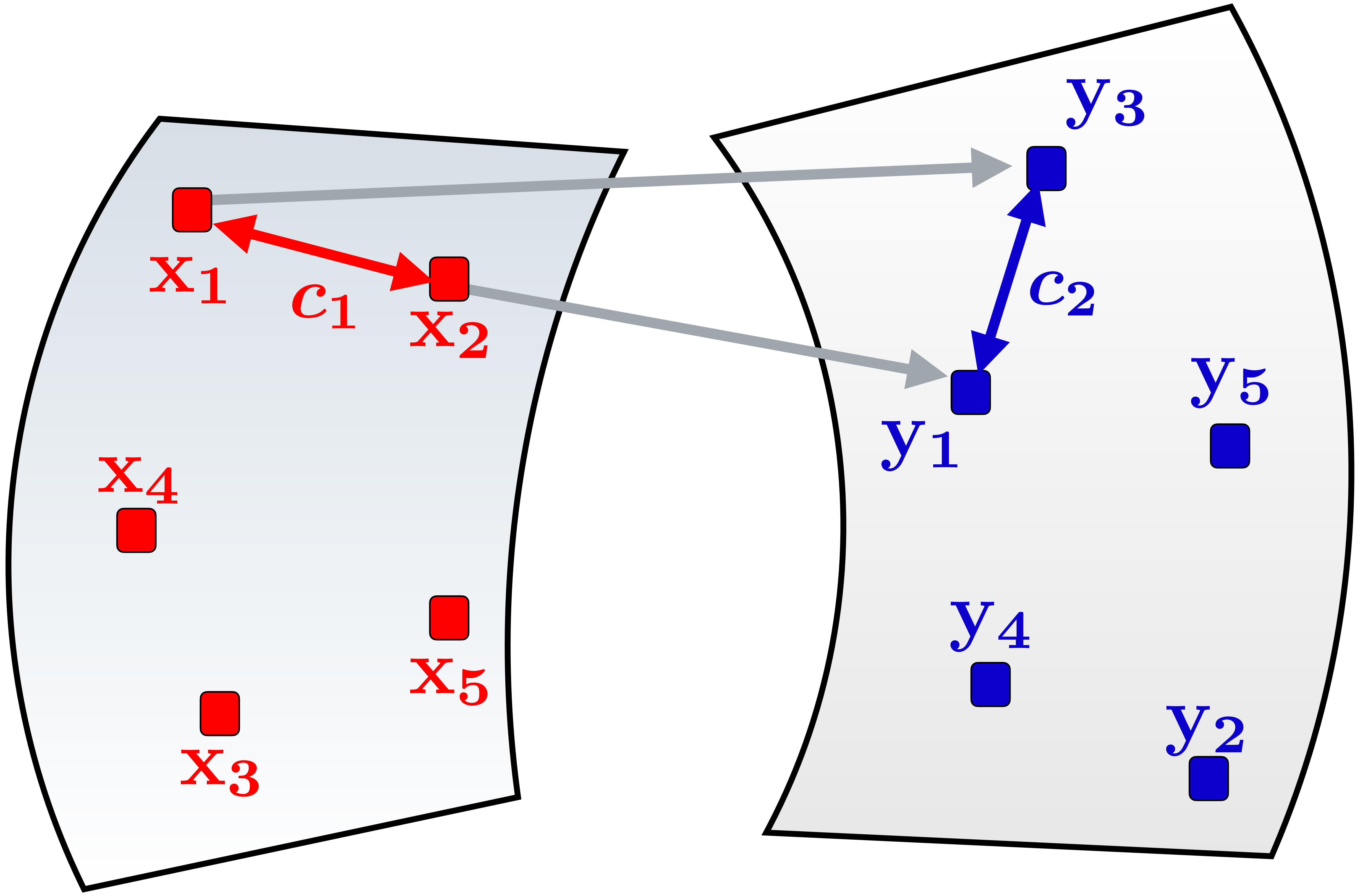
$$P_{t+1} \leftarrow \underset{P \mathbf{1}_m = a, P^T \mathbf{1}_n = b}{\mathrm{argmin}} \; \langle \nabla \mathcal{E}(P_t), P \rangle - \gamma E(P)$$

Any linear solver can be used

The Quadratic (Gromov-Wasserstein) Optimal Tranport Problem

$$\mathcal{E}(\boldsymbol{P}) := \sum_{i,i',j,j'} \boldsymbol{P}_{ij}\boldsymbol{P}_{i'j'}(\boldsymbol{c}_1(\mathbf{x}_i,\mathbf{x}_{i'}) - \boldsymbol{c}_2(\mathbf{y}_j,\mathbf{y}_{j'}))^2$$

$$\nabla\mathcal{E}(\boldsymbol{P}) \propto -\boldsymbol{C}_1\boldsymbol{P}\boldsymbol{C}_2$$




Beckman


Koopmans


Memoli


Sturm

$$\boldsymbol{P}_{t+1} \leftarrow \underset{\substack{\boldsymbol{P}\mathbf{1}_m=\boldsymbol{a},\boldsymbol{P}^T\mathbf{1}_n=\boldsymbol{b} \\ \mathrm{rank}(\boldsymbol{P})=r}}{\mathrm{argmin}} \langle \nabla\mathcal{E}(\boldsymbol{P}_t), \boldsymbol{P}\rangle$$

Any linear solver can be used

GW

Low-Rank and GW are Compatible, yielding Linear Time Algorithm

[Scetbon+22]

$$\mathcal{E}(\boldsymbol{P}) := \sum_{\boldsymbol{i},\boldsymbol{i'},\boldsymbol{j},\boldsymbol{j'}} \boldsymbol{P}_{\boldsymbol{ij}} \boldsymbol{P}_{\boldsymbol{i'}\boldsymbol{j'}} (\boldsymbol{c_1}(\mathbf{x}_{\boldsymbol{i}}, \mathbf{x}_{\boldsymbol{i'}}) - \boldsymbol{c_2}(\mathbf{y}_{\boldsymbol{j}}, \mathbf{y}_{\boldsymbol{j'}}))^2$$

$$\nabla \mathcal{E}(\boldsymbol{P}) \propto -\boldsymbol{C_1} \boldsymbol{P} \boldsymbol{C_2} \qquad \boldsymbol{C_1} = \boldsymbol{A}\boldsymbol{A}^T \qquad \boldsymbol{C_2} = \boldsymbol{B}\boldsymbol{B}^T$$

$$\boldsymbol{P_{t+1}} \leftarrow \underset{\substack{\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}} = \boldsymbol{a}, \boldsymbol{P}^T \mathbf{1}_{\boldsymbol{n}} = \boldsymbol{b} \\ \mathrm{rank}(\boldsymbol{P}) = r}}{\mathrm{argmin}} \langle \nabla \mathcal{E}(\boldsymbol{P_t}), \boldsymbol{P} \rangle$$

Low-Rank and GW are Compatible, yielding Linear Time Algorithm

$$\mathcal{E}(\boldsymbol{P}) := \sum_{\boldsymbol{i},\boldsymbol{i'},\boldsymbol{j},\boldsymbol{j'}} \boldsymbol{P}_{\boldsymbol{ij}} \boldsymbol{P}_{\boldsymbol{i'}\boldsymbol{j'}} (\boldsymbol{c_1}(\mathbf{x}_{\boldsymbol{i}}, \mathbf{x}_{\boldsymbol{i'}}) - \boldsymbol{c_2}(\mathbf{y}_{\boldsymbol{j}}, \mathbf{y}_{\boldsymbol{j'}}))^2$$

[Scetbon+22]

$$\nabla\mathcal{E}(\boldsymbol{P}) \propto -\boldsymbol{C_1}\boldsymbol{P}\boldsymbol{C_2} \qquad \boldsymbol{C_1} = \boldsymbol{A}\boldsymbol{A}^T \qquad \boldsymbol{C_2} = \boldsymbol{B}\boldsymbol{B}^T$$

$$\boldsymbol{P_{t+1}} \leftarrow \underset{\substack{\boldsymbol{P}\mathbf{1}_{\boldsymbol{m}}=\boldsymbol{a}, \boldsymbol{P}^T\mathbf{1}_{\boldsymbol{n}}=\boldsymbol{b} \\ \mathrm{rank}(\boldsymbol{P})=r}}{\mathrm{argmin}} \langle \nabla\mathcal{E}(\boldsymbol{P_t}), \boldsymbol{P} \rangle$$

$$Q_{t+1}, \boldsymbol{g}_{t+1}, R_{t+1} \leftarrow \arg \min_{\substack{Q\in U(\boldsymbol{a},\boldsymbol{g}), \\ R\in U(\boldsymbol{b},\boldsymbol{g}), \\ \boldsymbol{g}\in\Sigma_r}} -\langle \boldsymbol{C_1} \overbrace{Q_t D(\boldsymbol{g}_t)^{-1} R_t^T}^{\boldsymbol{P_t}} \boldsymbol{C_2}, \overbrace{Q D(1/\boldsymbol{g}) R^T}^{\boldsymbol{P}} \rangle.$$

Low-Rank and GW are Compatible, yielding Linear Time Algorithm

[Scetbon+22]

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2$$

$$\nabla\mathcal{E}(P) \propto -C_1 P C_2 \qquad C_1 = AA^T \qquad C_2 = BB^T$$

$$P_{t+1} \leftarrow \underset{\substack{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b \\ \mathrm{rank}(P) = r}}{\operatorname{argmin}} \langle \nabla\mathcal{E}(P_t), P \rangle$$

$$Q_{t+1}, g_{t+1}, R_{t+1} \leftarrow \arg \min_{\substack{Q \in U(a,g), \\ R \in U(b,g), \\ g \in \Sigma_r}} -\langle \underbrace{C_1 \overbrace{Q_t D(g_t)^{-1}}}_{} \underbrace{\overbrace{R_t^T} C_2}_{}, \overbrace{Q D(1/g) R^T} \rangle.$$

with the overbraces labelled P_t (over $Q_t D(g_t)^{-1} R_t^T$) and P (over $Q D(1/g) R^T$).

Low-Rank and GW are Compatible, yielding Linear Time Algorithm

[Scetbon+22]

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathrm{x}_i, \mathrm{x}_{i'}) - c_2(\mathrm{y}_j, \mathrm{y}_{j'}))^2$$

$$\nabla \mathcal{E}(P) \propto -C_1 P C_2 \qquad C_1 = AA^T \qquad C_2 = BB^T$$

$$P_{t+1} \leftarrow \underset{\substack{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b \\ \mathrm{rank}(P) = r}}{\mathrm{argmin}} \langle \nabla \mathcal{E}(P_t), P \rangle$$

$$Q_{t+1}, g_{t+1}, R_{t+1} \leftarrow \arg \min_{\substack{Q \in U(a,g), \\ R \in U(b,g), \\ g \in \Sigma_r}} -\langle C_1 \overbrace{Q_t D(g_t)^{-1} R_t^T}^{P_t} C_2, \overbrace{Q D(1/g) R^T}^{P} \rangle.$$

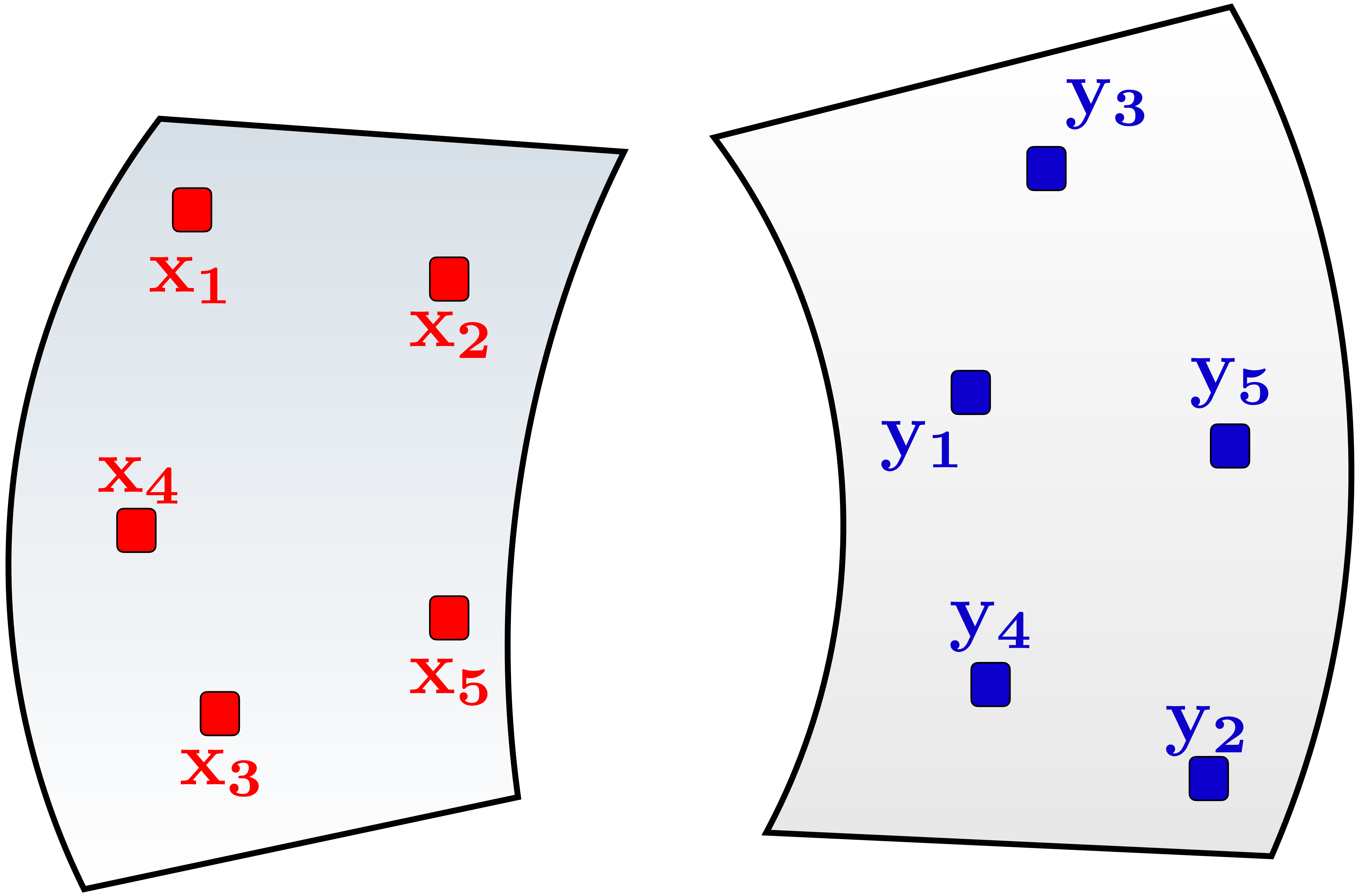
$$Q_{t+1}, g_{t+1}, R_{t+1} \leftarrow \arg \min_{\substack{Q \in U(a,g), \\ R \in U(b,g), \\ g \in \Sigma_r}} -\langle \underbrace{AA^T Q_t D(g_t^{-1/2})}_{A_t} \underbrace{D(g_t^{-1/2}) R_t^T B^T B}_{B_t}, Q D(1/g) R^T \rangle.$$

LRGW

Fused Gromov-Wasserstein

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2$$

[Vayer+'18]

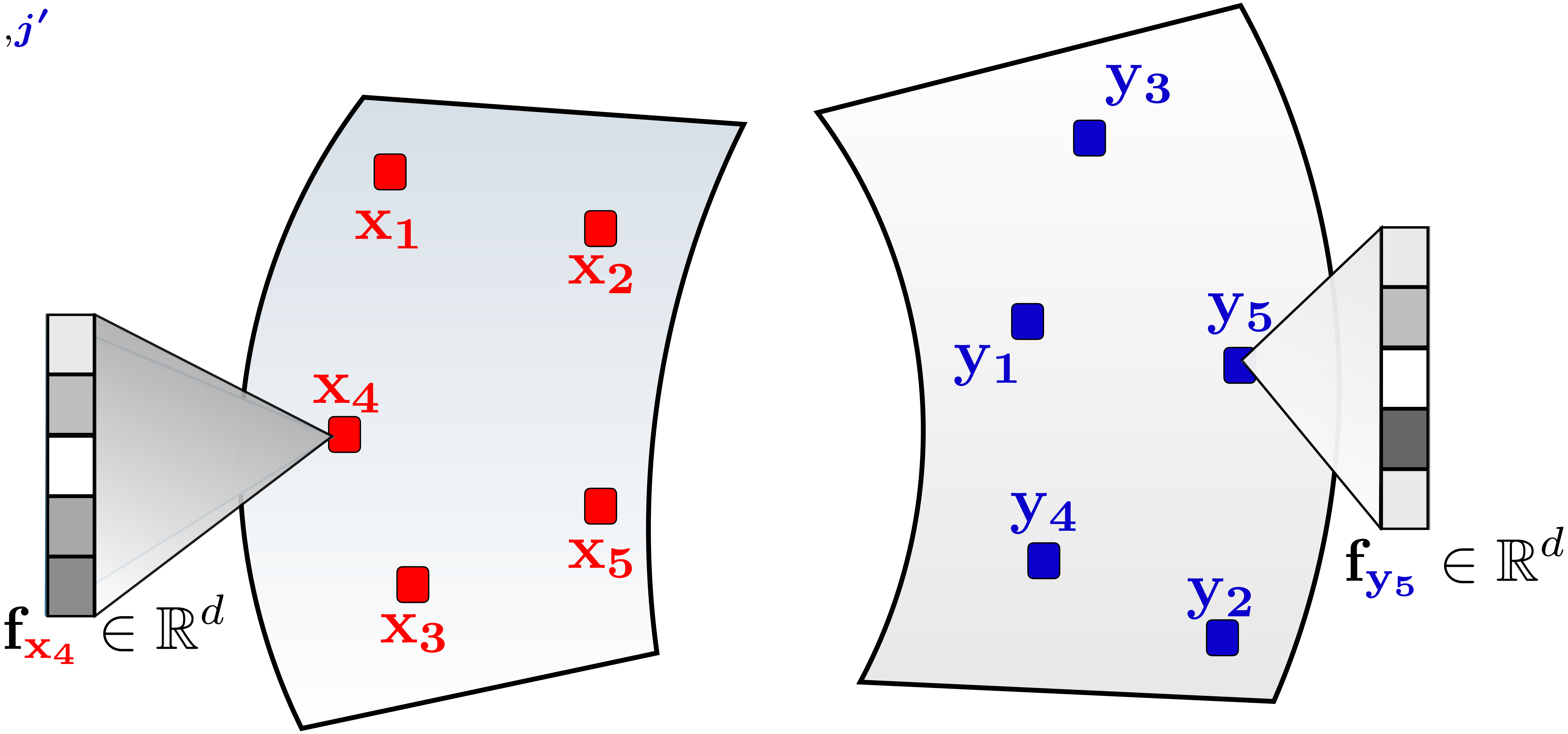


$$P_{t+1} \leftarrow \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\text{argmin}} \; \langle -C_1 P_t C_2 + \alpha \; C_3, P \rangle$$

Fused Gromov-Wasserstein

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2$$

[Vayer+'18]

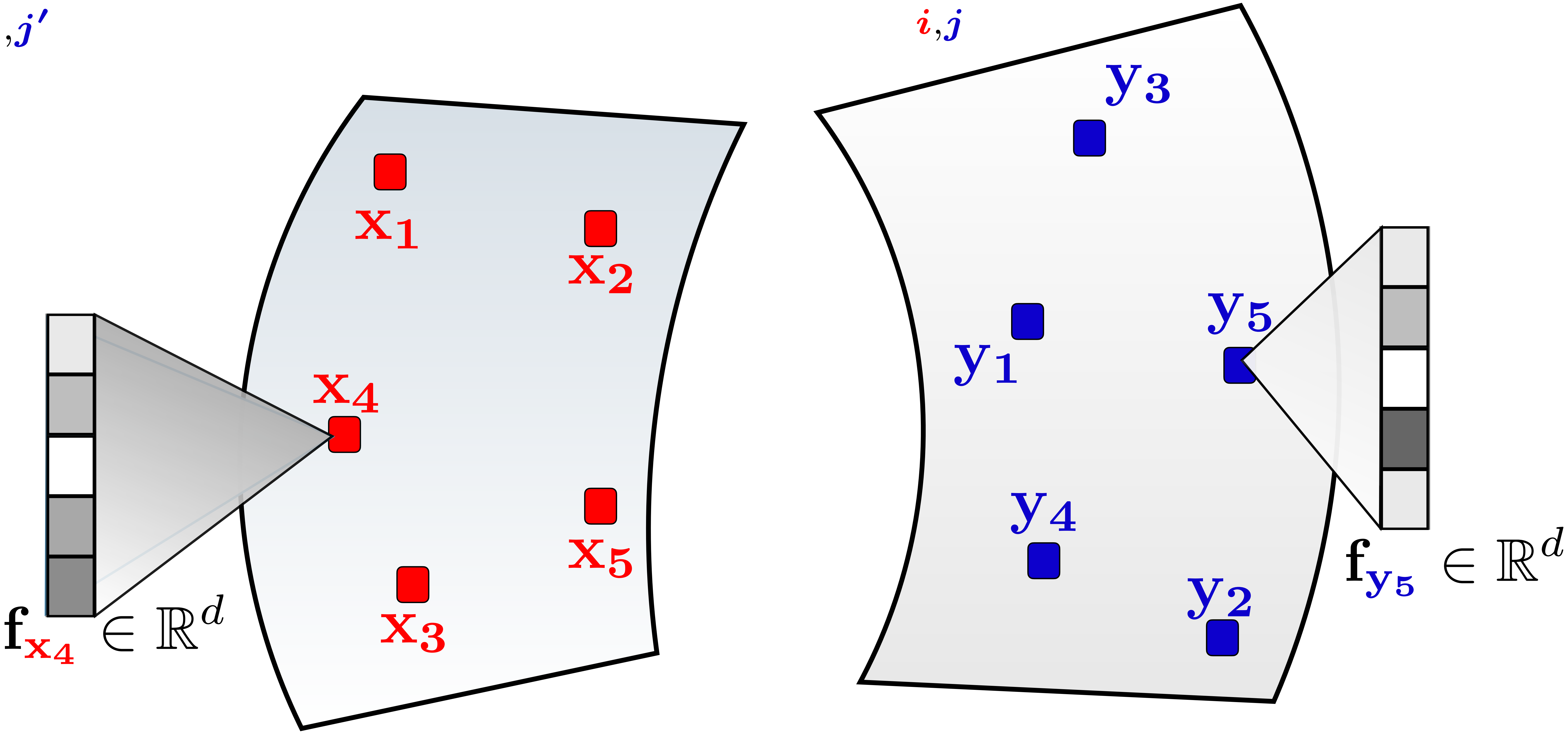


$$P_{t+1} \leftarrow \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\operatorname{argmin}} \langle -C_1 P_t C_2 + \alpha\, C_3, P \rangle$$

Fused Gromov-Wasserstein

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2 + \alpha \sum_{i,j} P_{ij}\, c_3(\mathbf{f}_{\mathbf{x}_i}, \mathbf{f}_{\mathbf{y}_j})$$

[Vayer+'18]

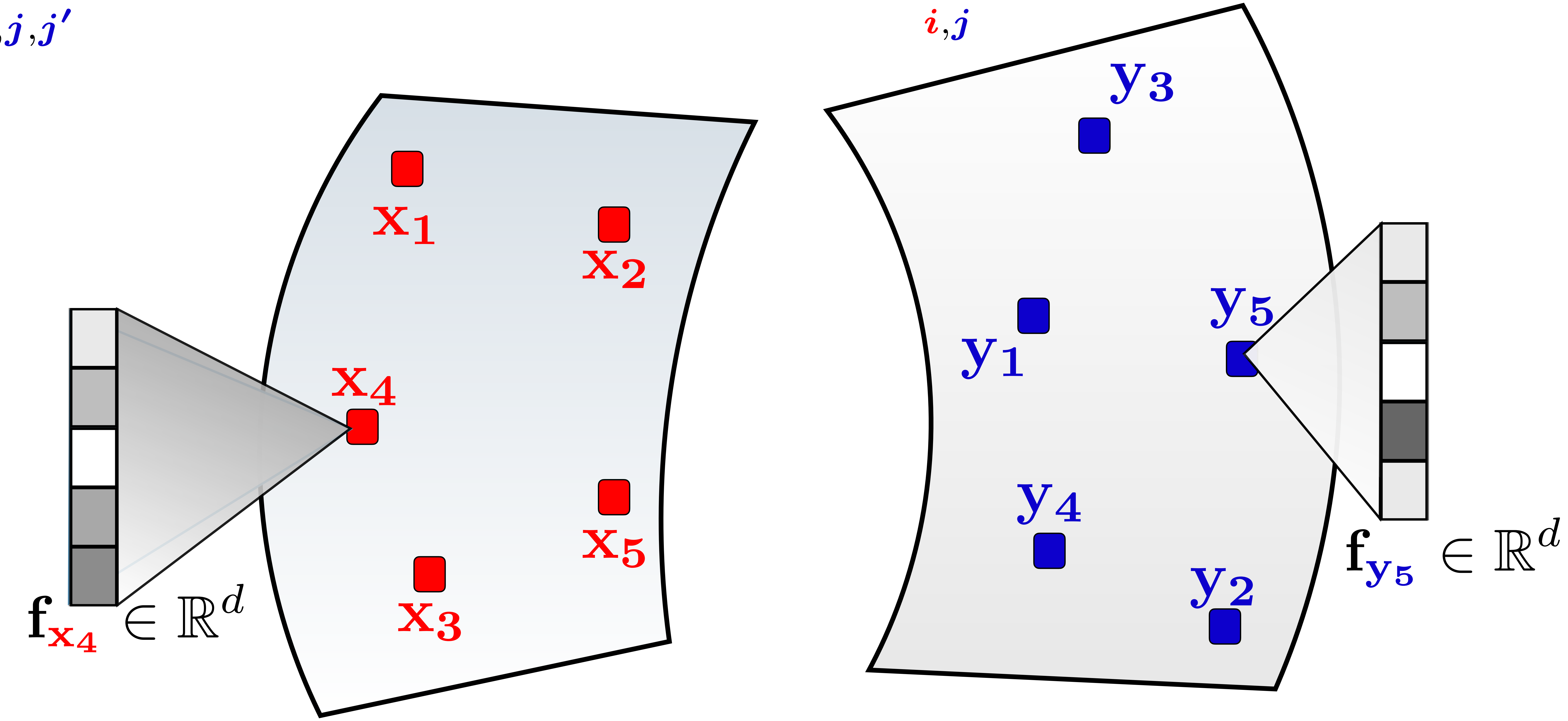


$$P_{t+1} \leftarrow \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\text{argmin}} \ \langle -C_1 P_t C_2 + \alpha\, C_3, P \rangle$$

Fused Gromov-Wasserstein

$$\mathcal{E}(P) := \sum_{i,i',j,j'} P_{ij} P_{i'j'} (c_1(\mathbf{x}_i, \mathbf{x}_{i'}) - c_2(\mathbf{y}_j, \mathbf{y}_{j'}))^2 + \alpha \sum_{i,j} P_{ij}\, c_3(\mathbf{f}_{\mathbf{x}_i}, \mathbf{f}_{\mathbf{y}_j})$$

[Vayer+'18]



$$P_{t+1} \leftarrow \underset{P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}{\text{argmin}} \langle -C_1 P_t C_2 + \alpha\, C_3, P \rangle$$

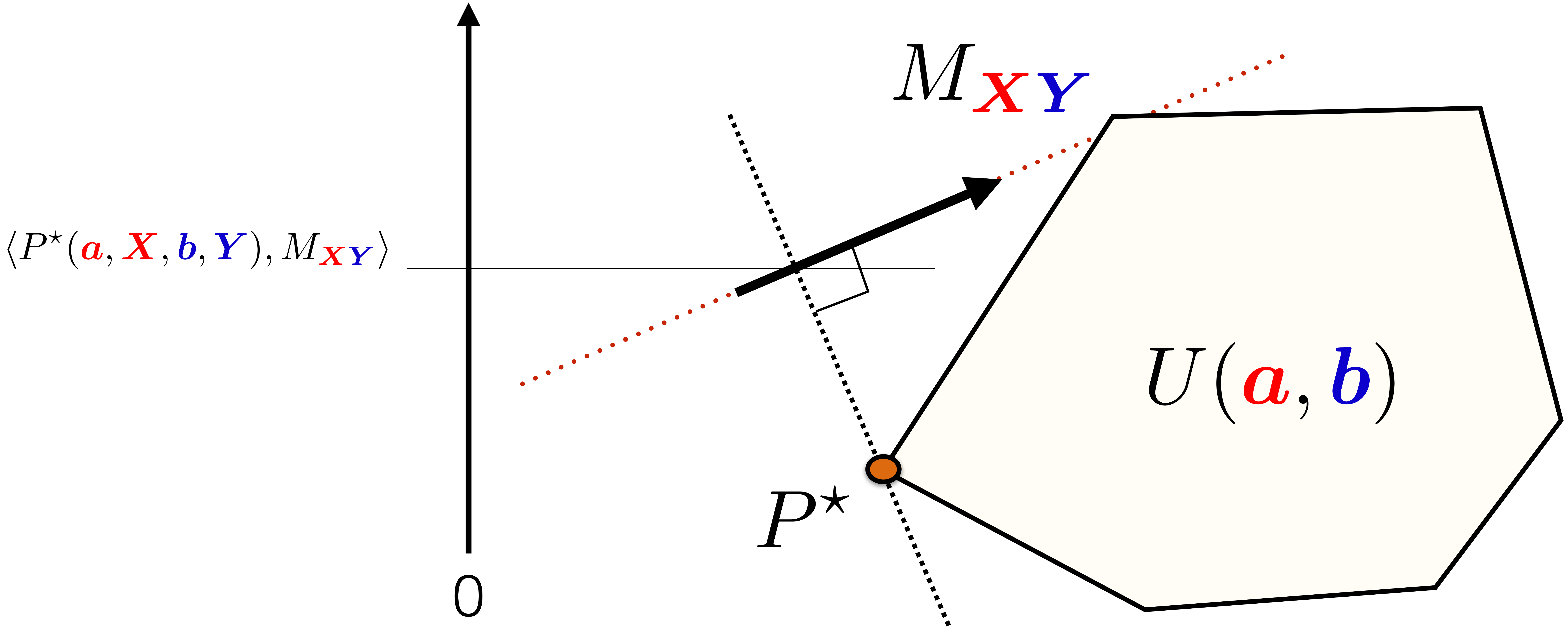
... also compatible with LR formulation

MOSCOT FGW

Rough Picture of What is Doable and What is Not [REDO with REFS]

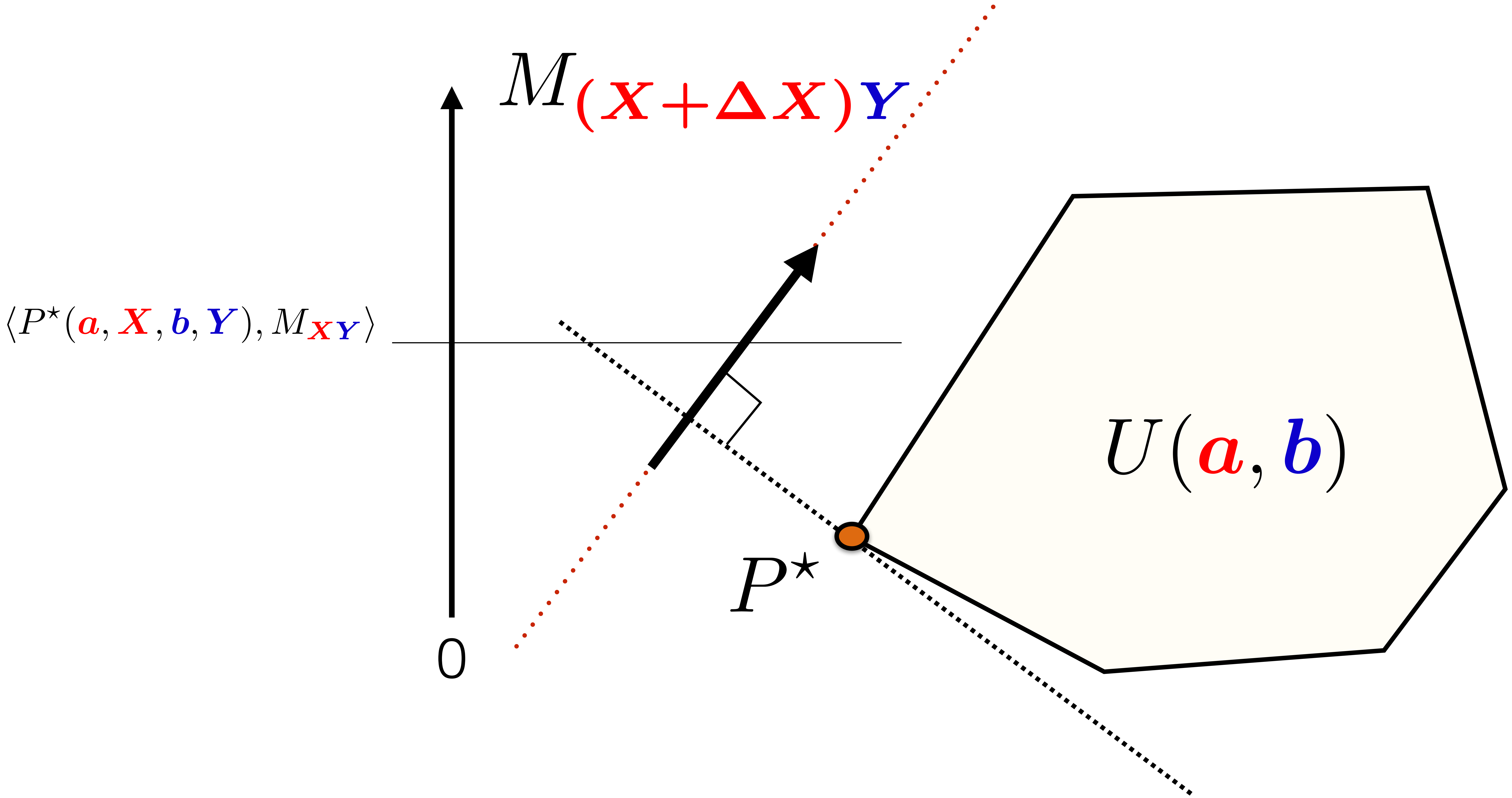
Problem \ Solver	Linear Program	Entropic Regularization $-\varepsilon E(\boldsymbol{P})$	Low-Rank Constraint $\text{rank}(\boldsymbol{P}) = r$
Linear OT Problem	**N < 10k, unstable, no GPU**	**linear / quadratic (N~100k) Convex (Sinkhorn), GPU**	**linear (N ~ 500k) not cvx, GPU friendly 2021~**
Fused-GW Problem	**unstable**	**quadratic / cubic (N~20k) not cvx, GPU. 2016~**	**linear (N ~ 500k) not cvx, GPU 2022**
Unbalanced Linear OT	**Partial OT formulations**	**linear / quadratic (N~100k) cvx (Sinkhorn), GPU 2015~**	**linear(N ~ 500k) not cvx, GPU 2023**
Unbalanced Fused-GW	**NA**	**quadratic / cubic (N < 10k) not cvx, GPU 2022~**	

Differentiating the OT Objective Function



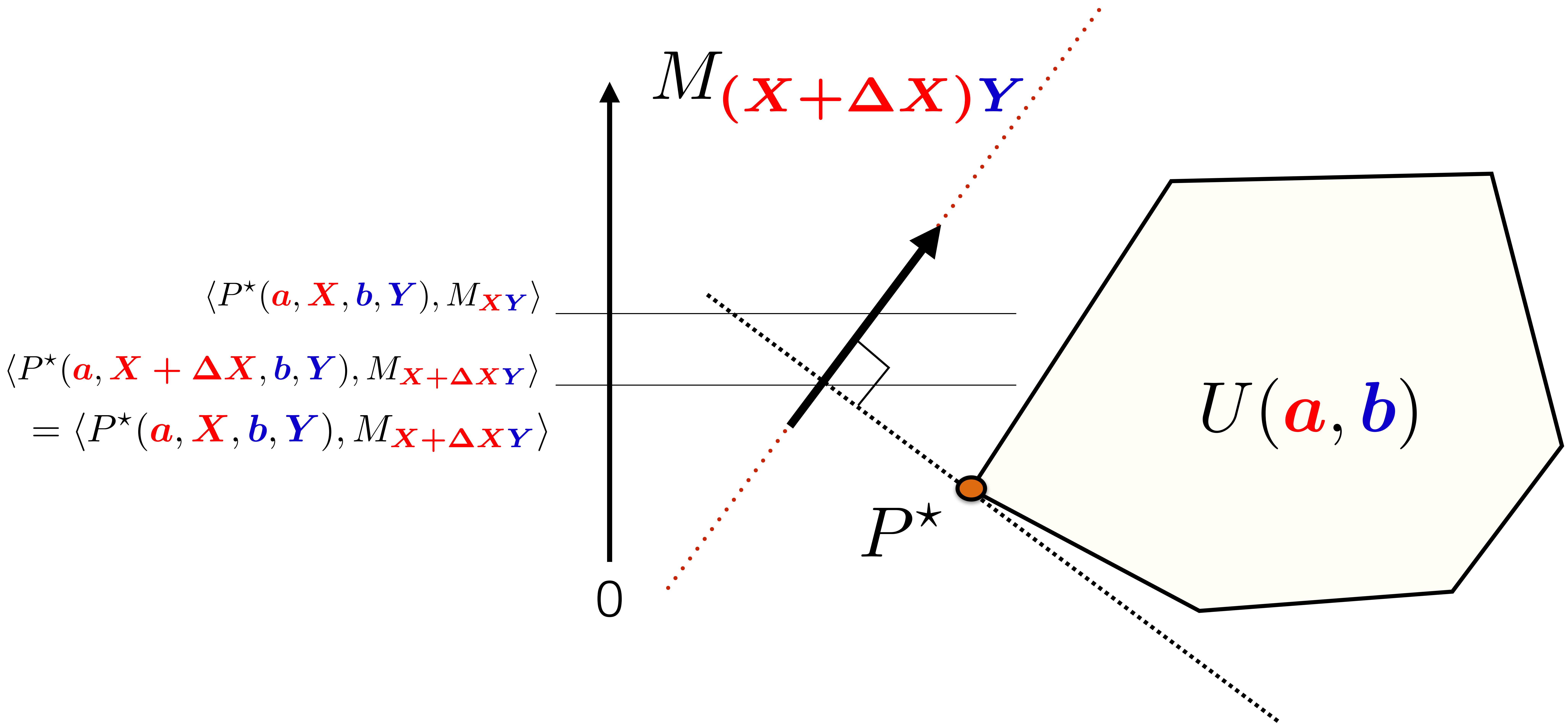
Computing $\nabla_{\blacksquare} W$ can be usually handled with the Envelope (Danskin) theorem, where $\blacksquare$ can stand for any parameter defining the problem.

Differentiating the OT Objective Function



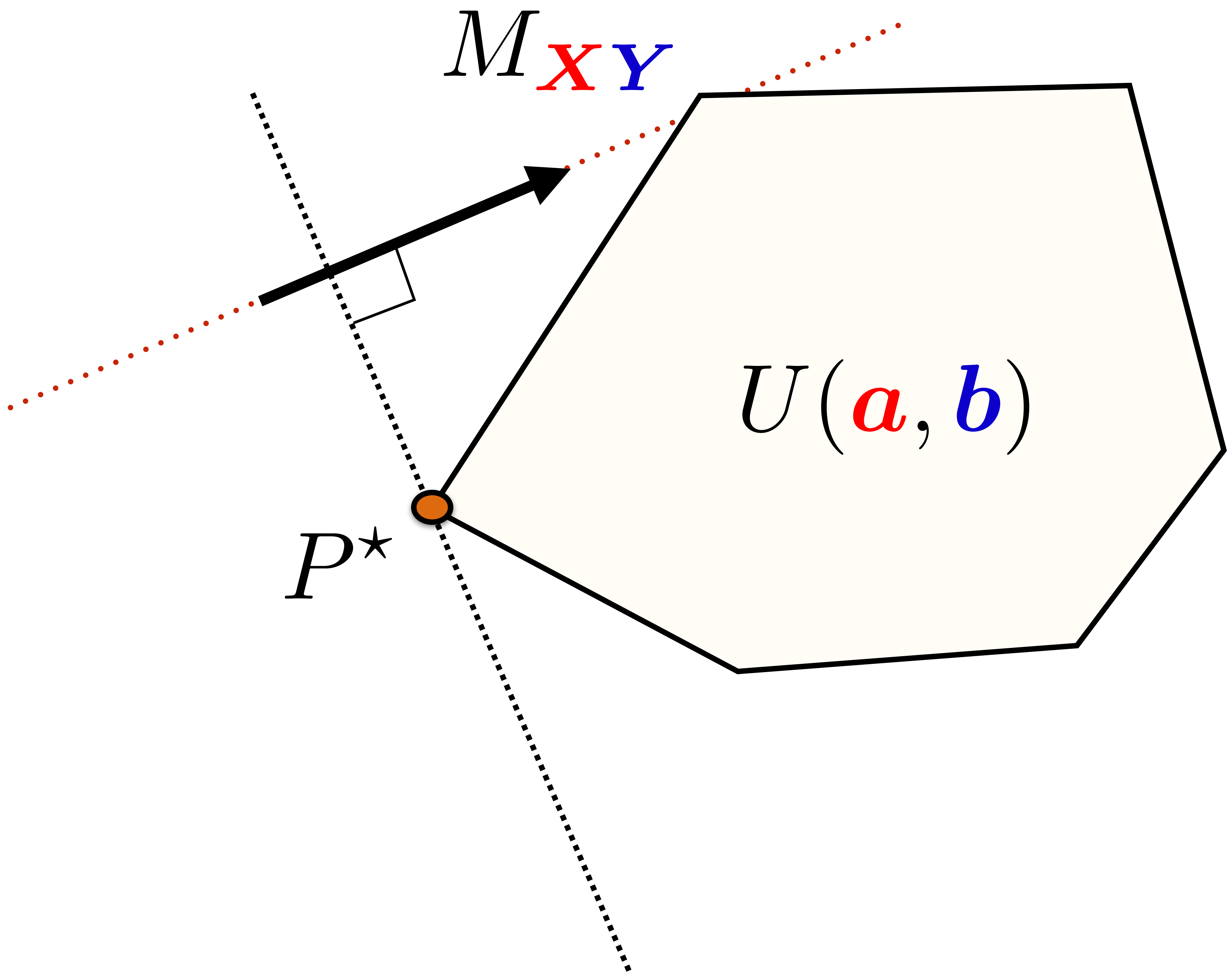
Computing $\nabla_{\blacksquare} W$ can be usually handled with the Envelope (Danskin) theorem, where ■ can stand for any parameter defining the problem.

Differentiating the OT Objective Function

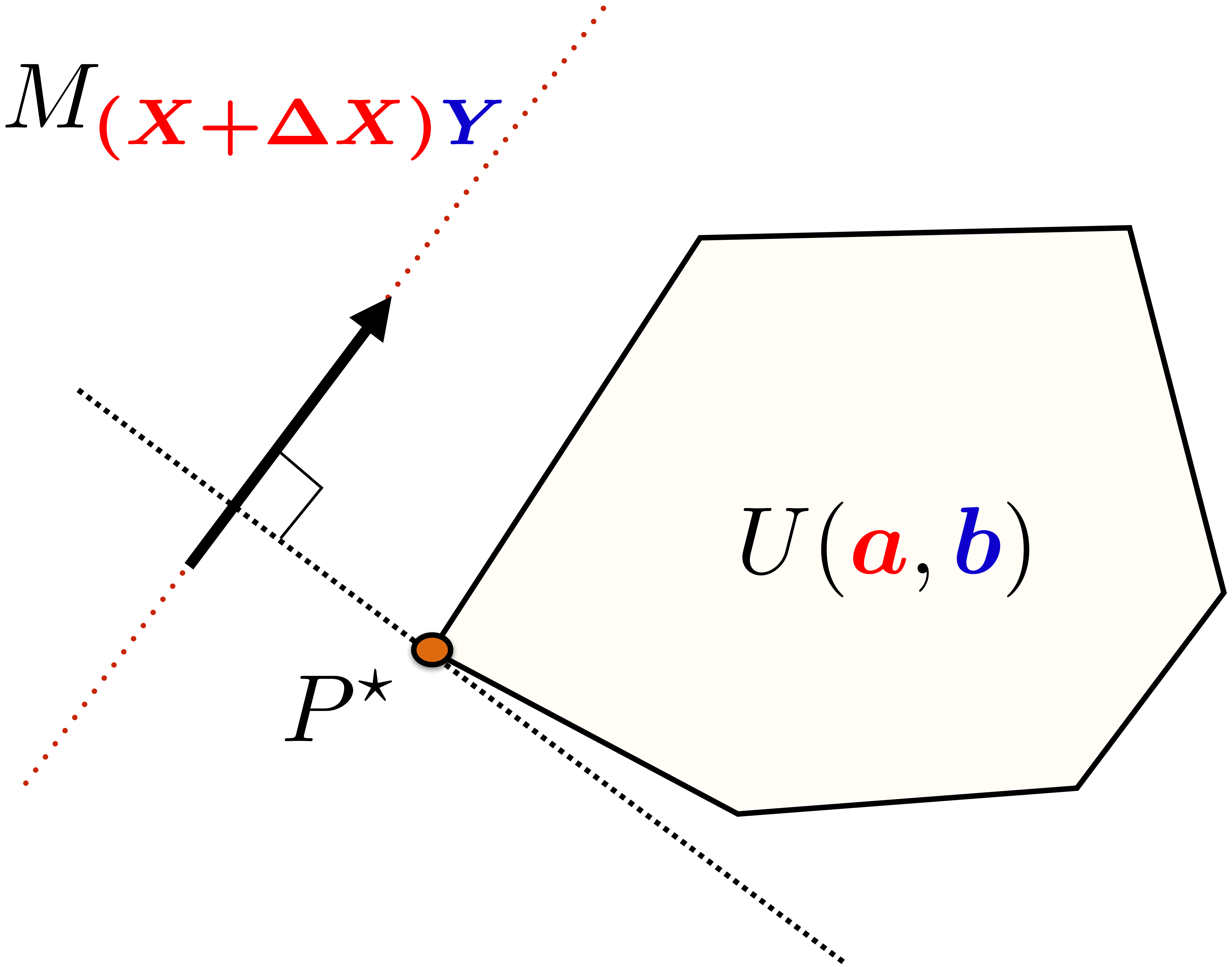


Computing $\nabla_{\blacksquare} W$ can be usually handled with the Envelope (Danskin) theorem, where $\blacksquare$ can stand for any parameter defining the problem.

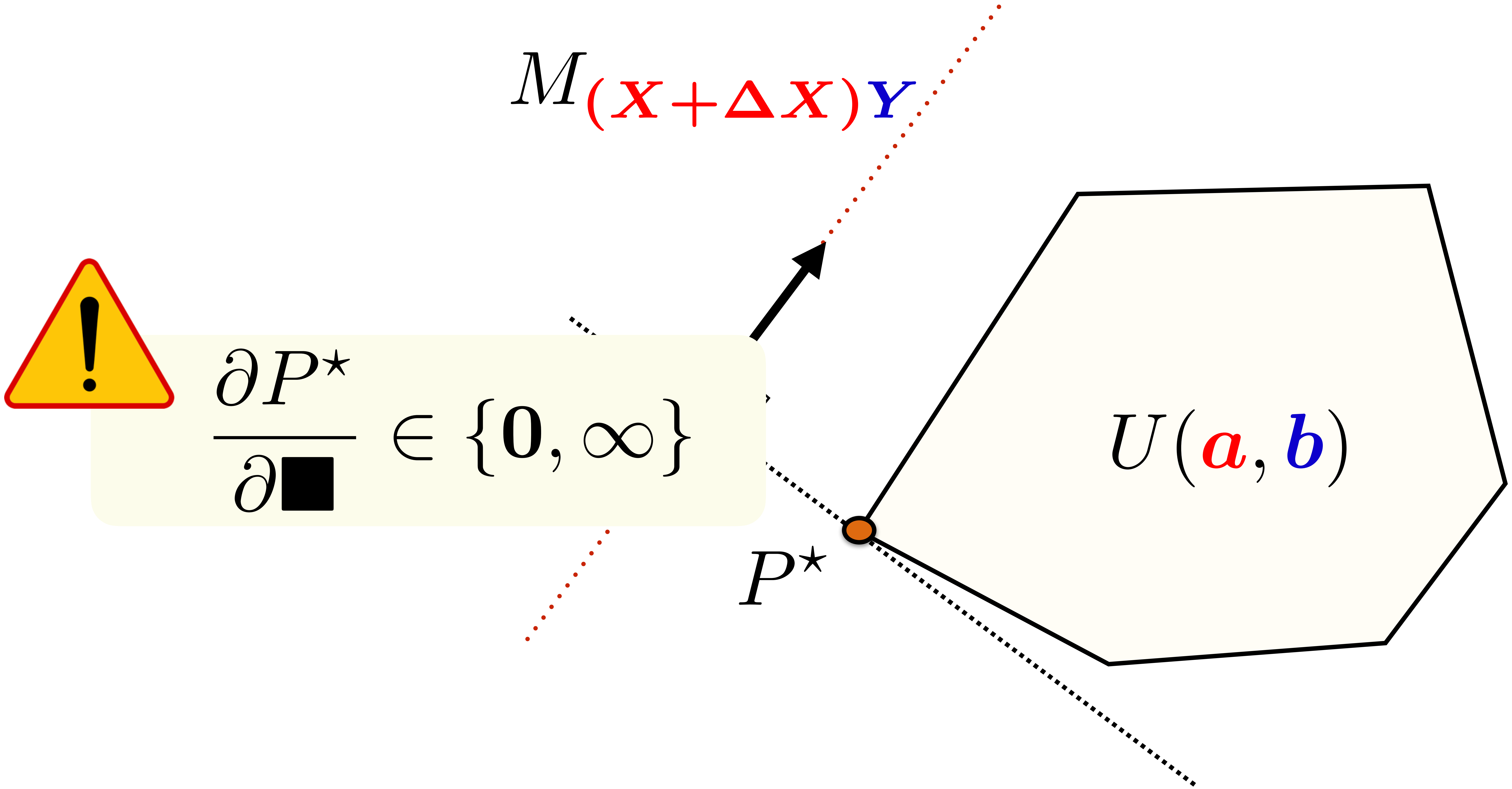
Differentiating the OT Solution



Differentiating the OT Solution

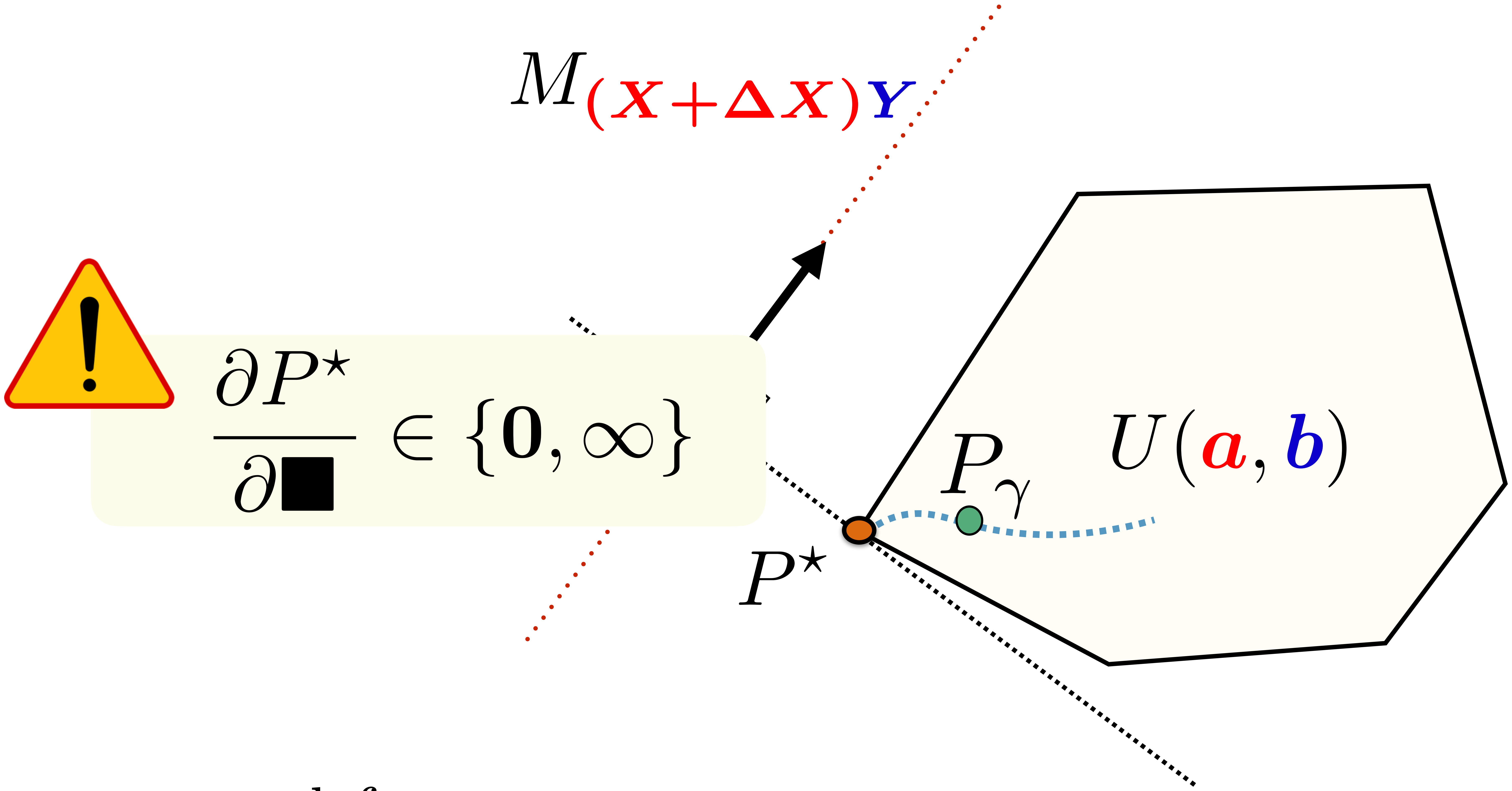


Differentiating the OT Solution



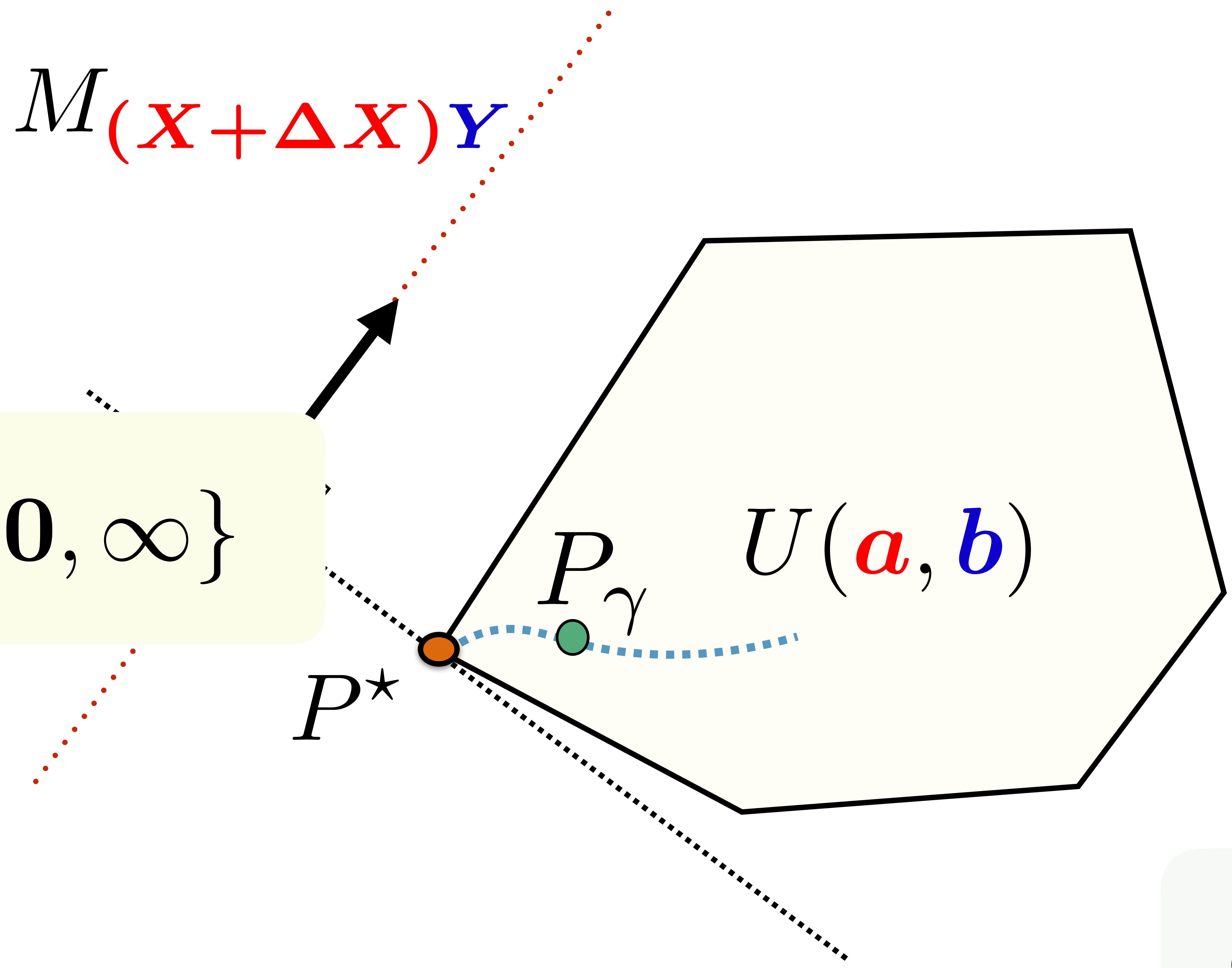
$$\frac{\partial P^\star}{\partial \blacksquare} \in \{\mathbf{0}, \infty\}$$

Differentiating the OT Solution



$$P_\gamma \stackrel{\text{def}}{=} \underset{\boldsymbol{P} \in U(\boldsymbol{a},\boldsymbol{b})}{\operatorname{argmin}} \langle \boldsymbol{P}, M_{\mathbf{X}\mathbf{Y}} \rangle - \gamma E(\boldsymbol{P})$$

Differentiating the OT Solution

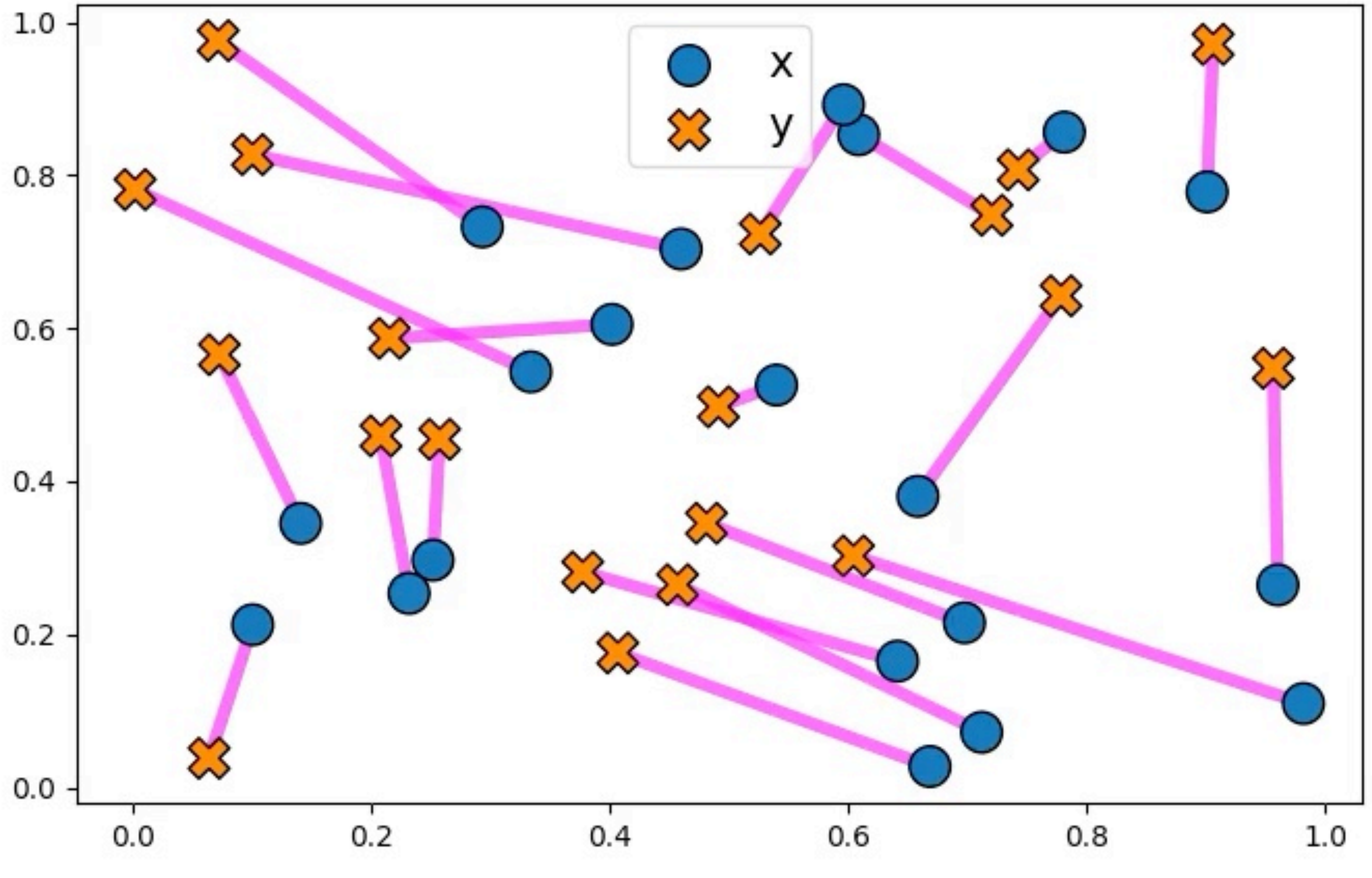


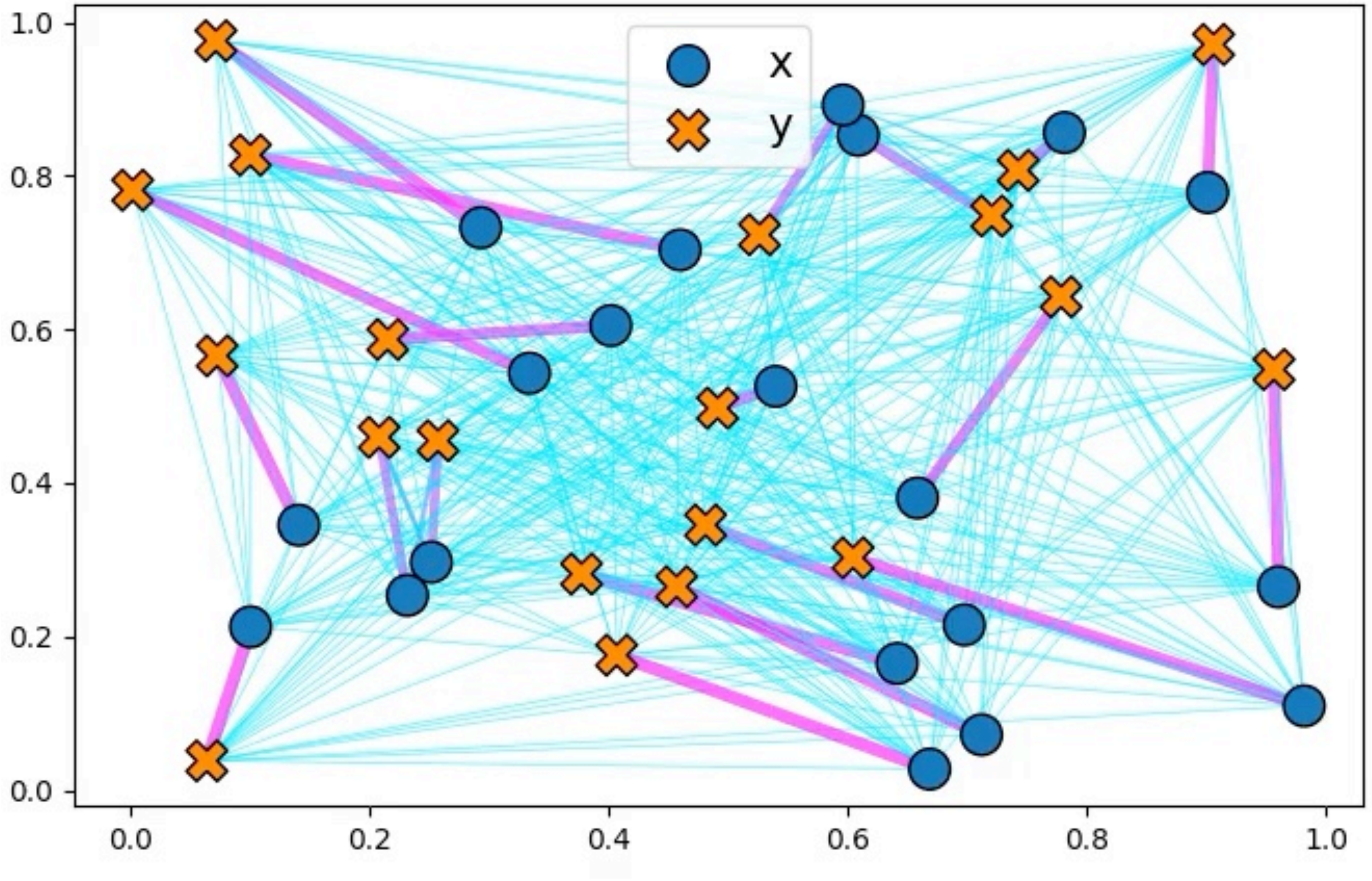
$$\frac{\partial P^\star}{\partial \blacksquare} \in \{\mathbf{0}, \infty\}$$

$$P_\gamma \stackrel{\text{def}}{=} \underset{\boldsymbol{P} \in U(\boldsymbol{a}, \boldsymbol{b})}{\operatorname{argmin}} \langle \boldsymbol{P}, M_{\mathbf{X}\mathbf{Y}} \rangle - \gamma E(\boldsymbol{P})$$

$$\frac{\partial P_\gamma}{\partial \blacksquare}?$$

Differentiating the OT Solution

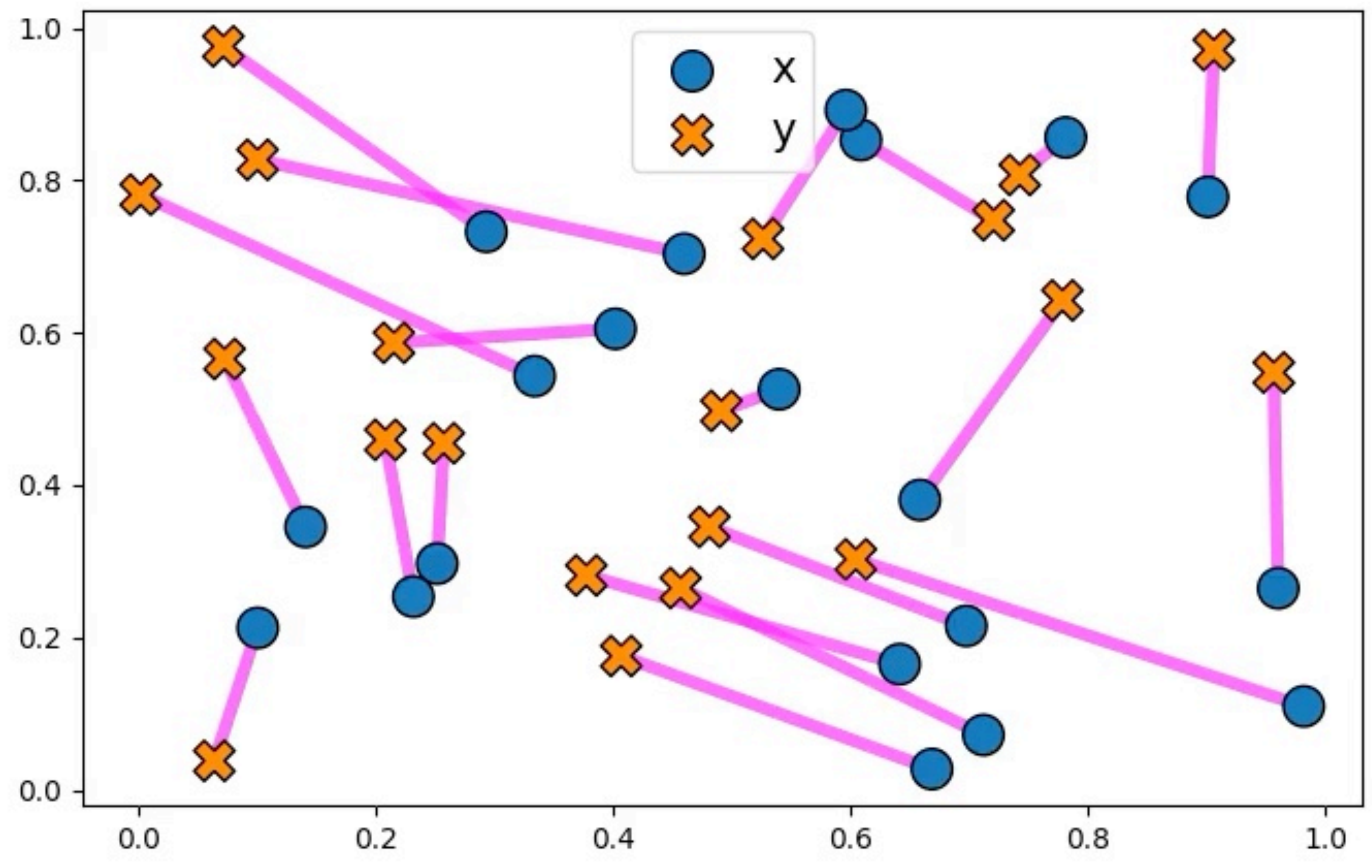


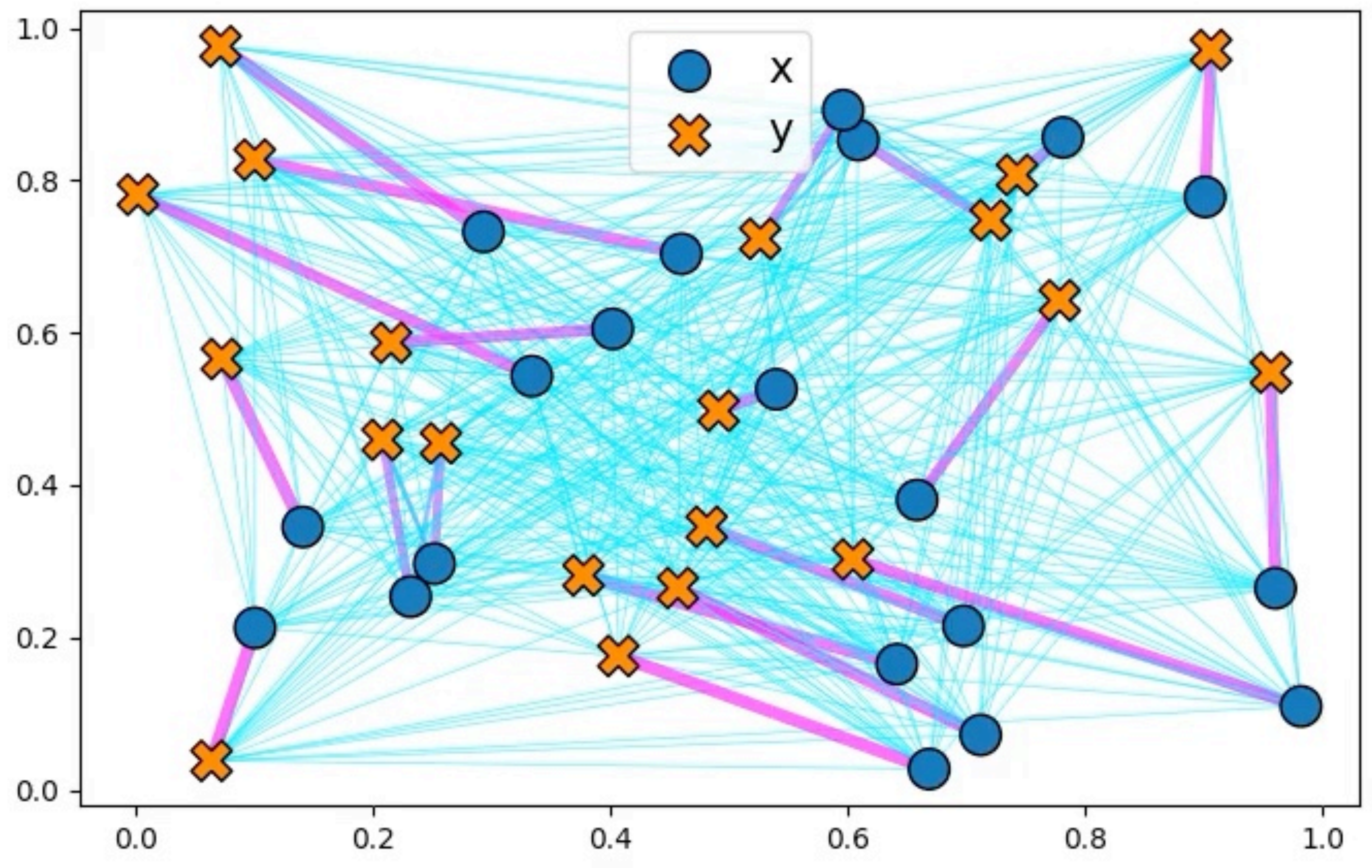


$\left(\frac{\partial P_\gamma}{\partial \blacksquare}\right)^T$ can be evaluated using either unrolling or the implicit function theorem.

Sinkhorn Hessians

Differentiating the OT Solution



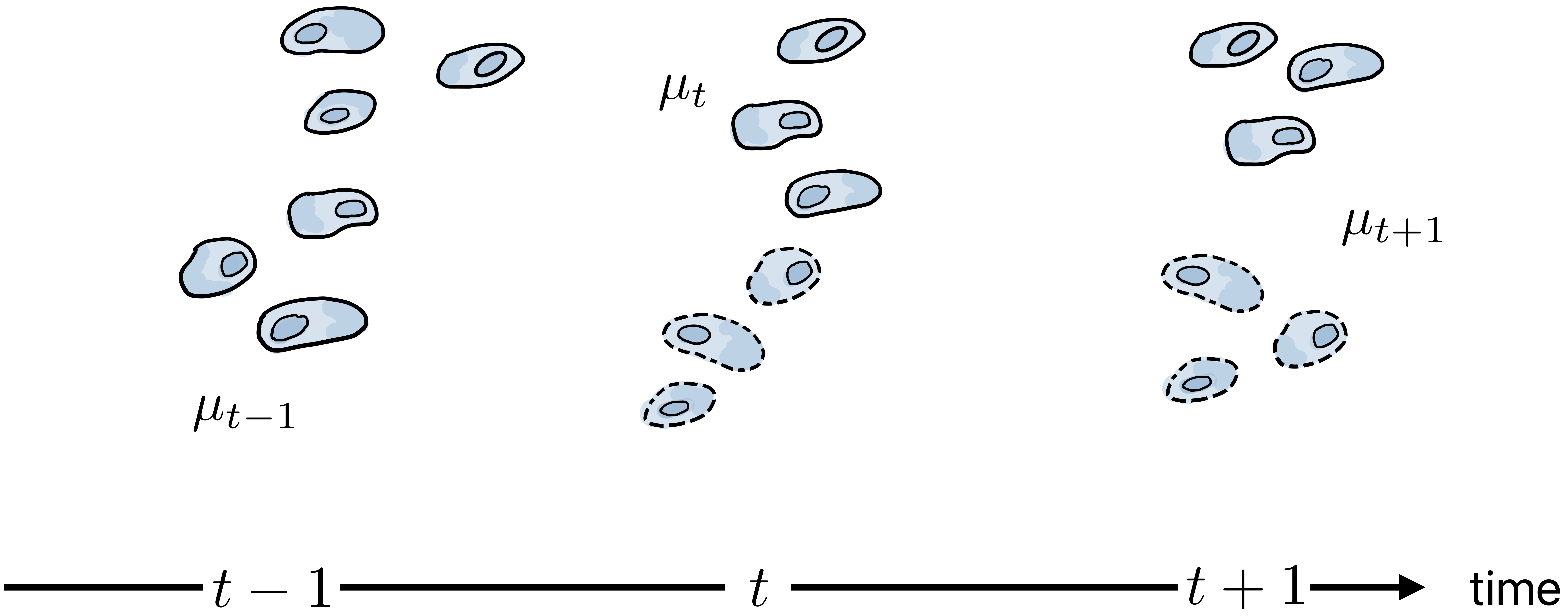


$\left(\frac{\partial P_\gamma}{\partial \blacksquare}\right)^T$ can be evaluated using either unrolling or the implicit function theorem.

Sinkhorn Hessians

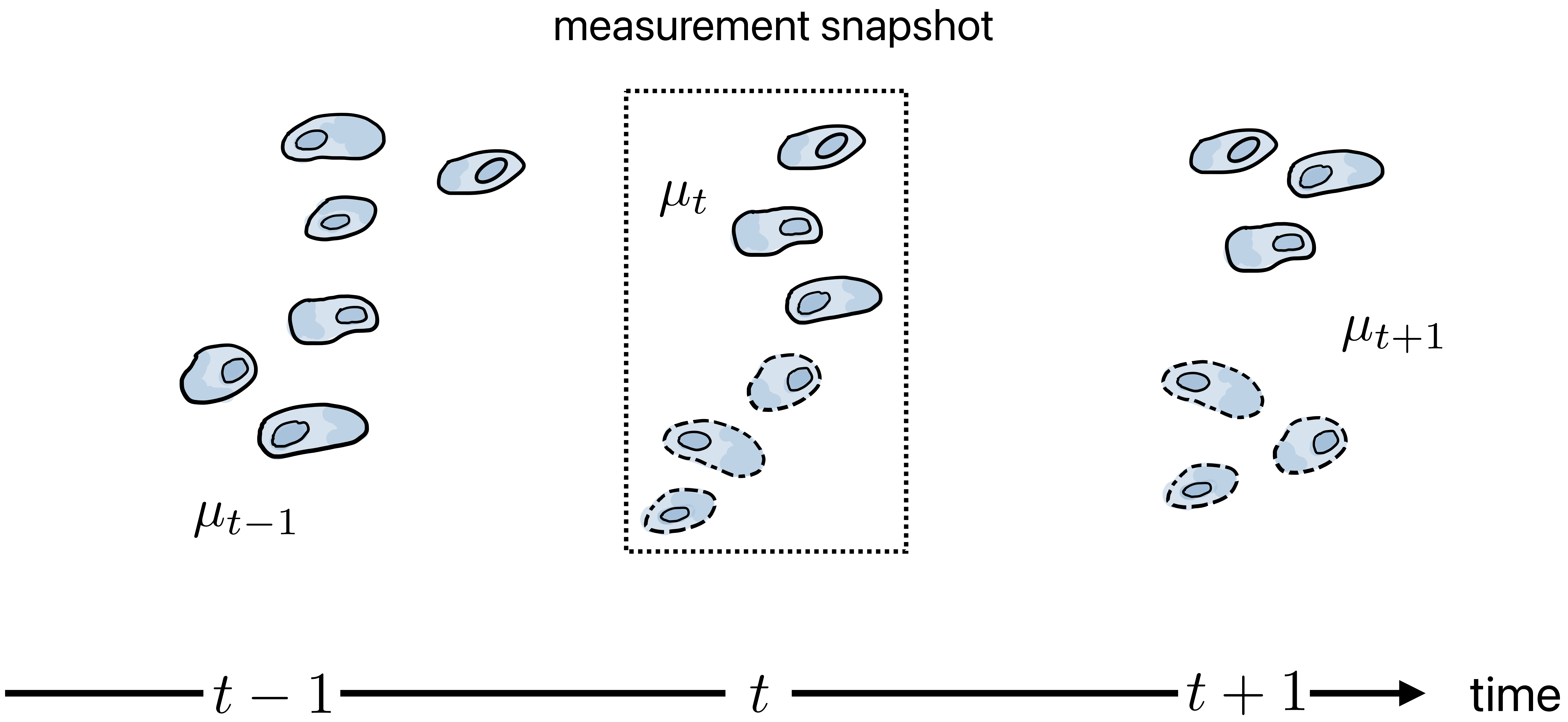
Motivating Example: **Cell Differentiation**

Optimal transport plan for recovering developmental trajectories



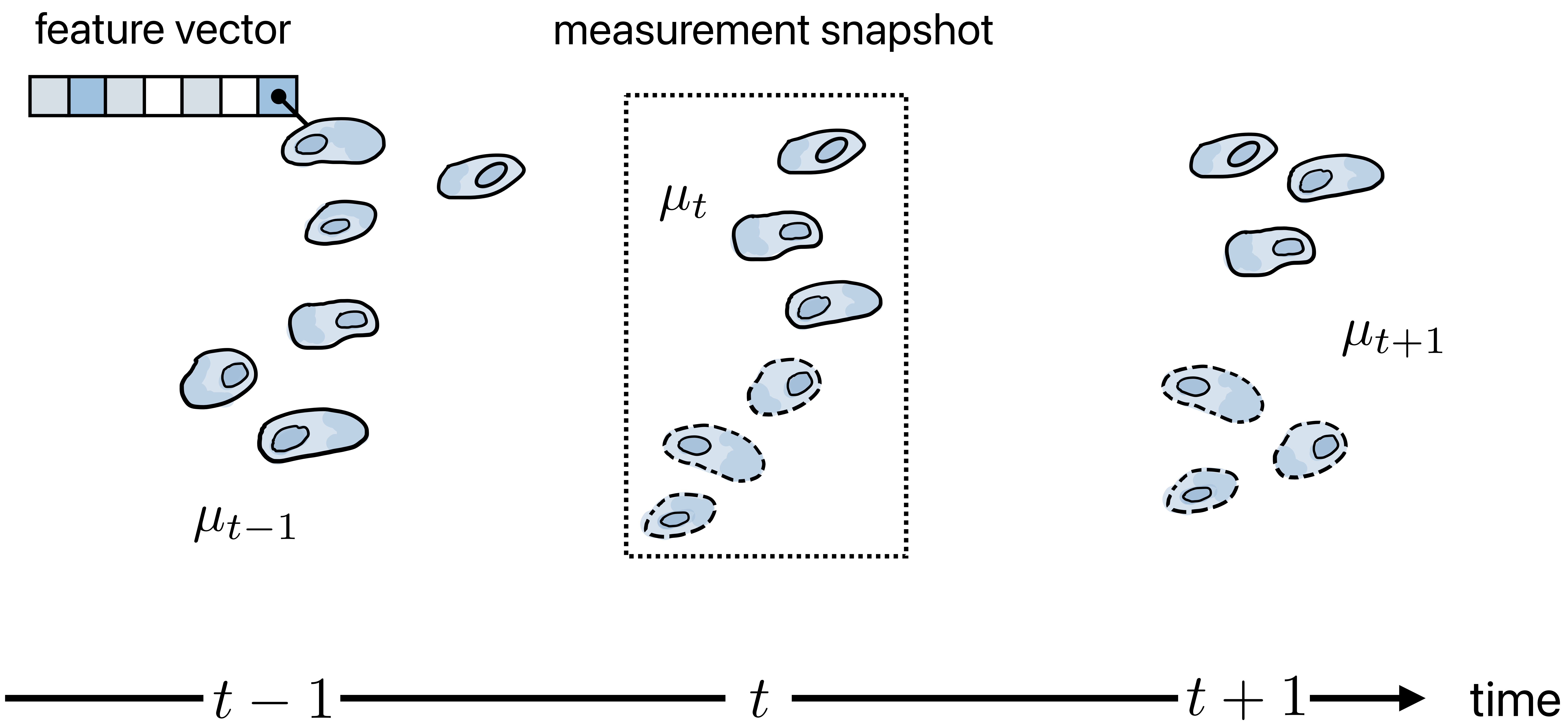
Motivating Example: **Cell Differentiation**

Optimal transport plan for recovering developmental trajectories



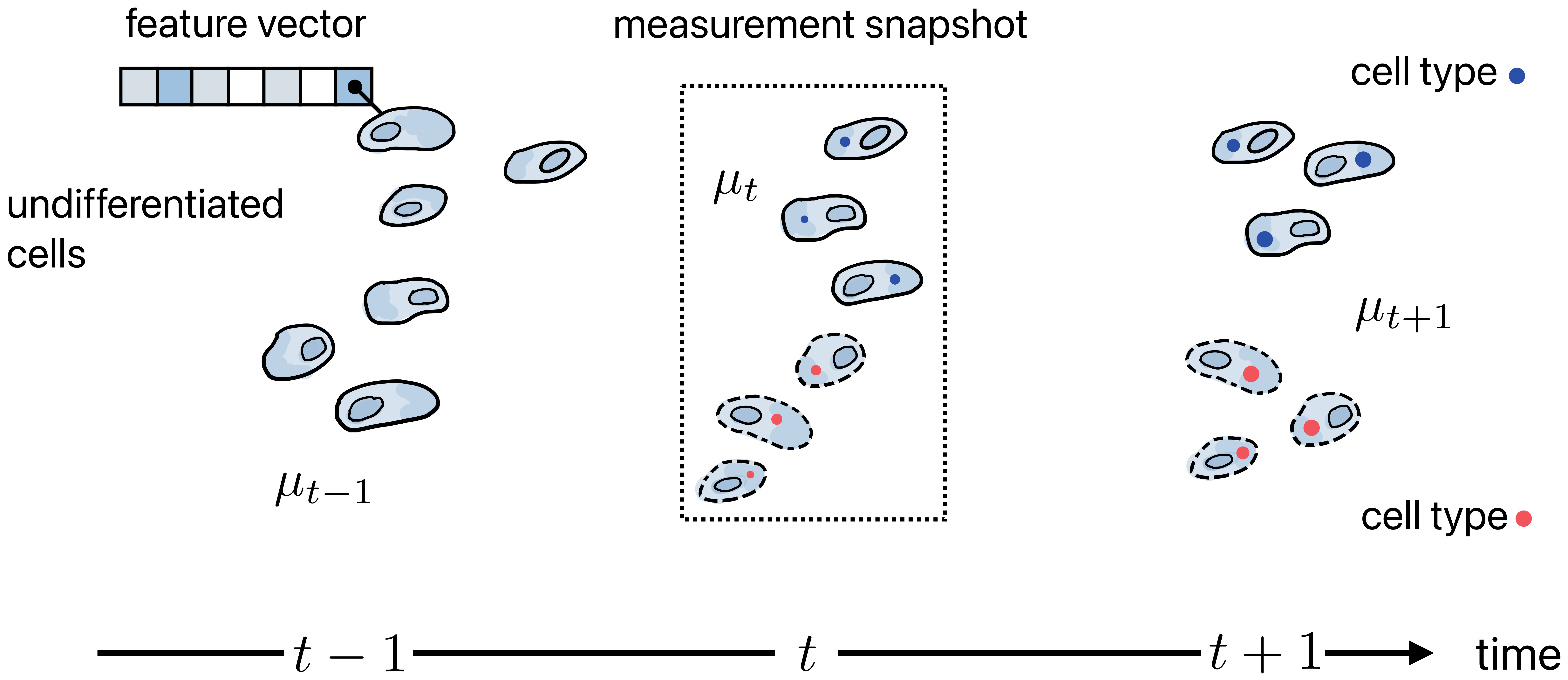
Motivating Example: **Cell Differentiation**

Optimal transport plan for recovering developmental trajectories



Motivating Example: **Cell Differentiation**

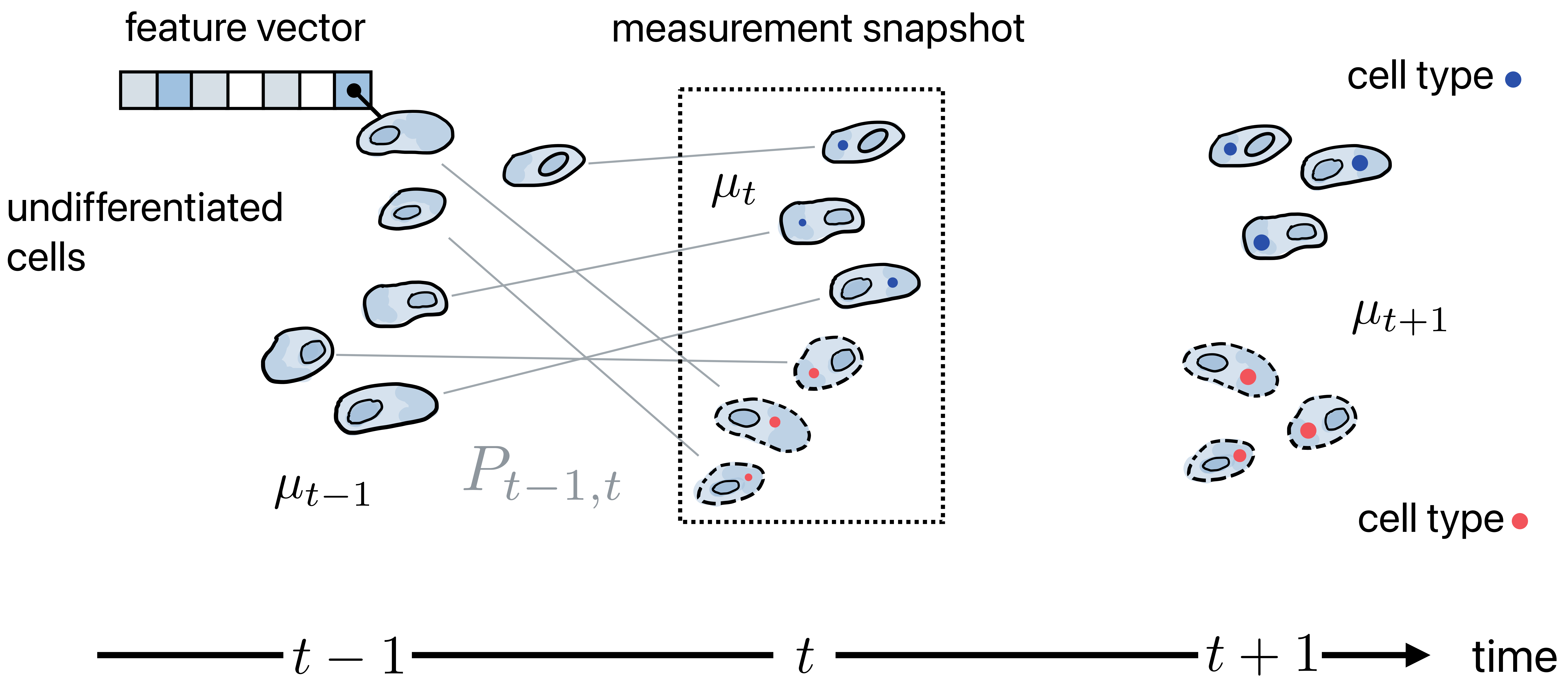
Optimal transport plan for recovering developmental trajectories



Motivating Example: **Cell Differentiation**

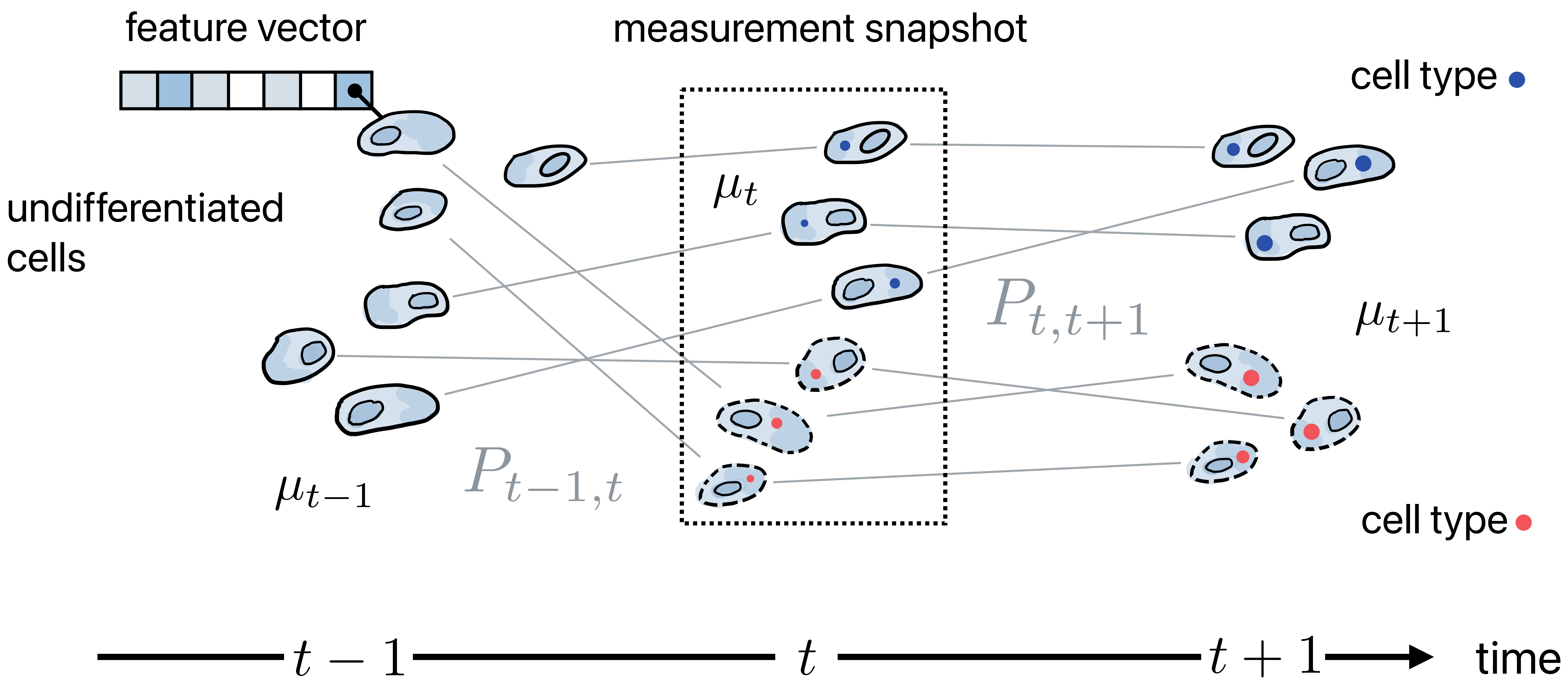
Part 2

Optimal transport plan for recovering developmental trajectories



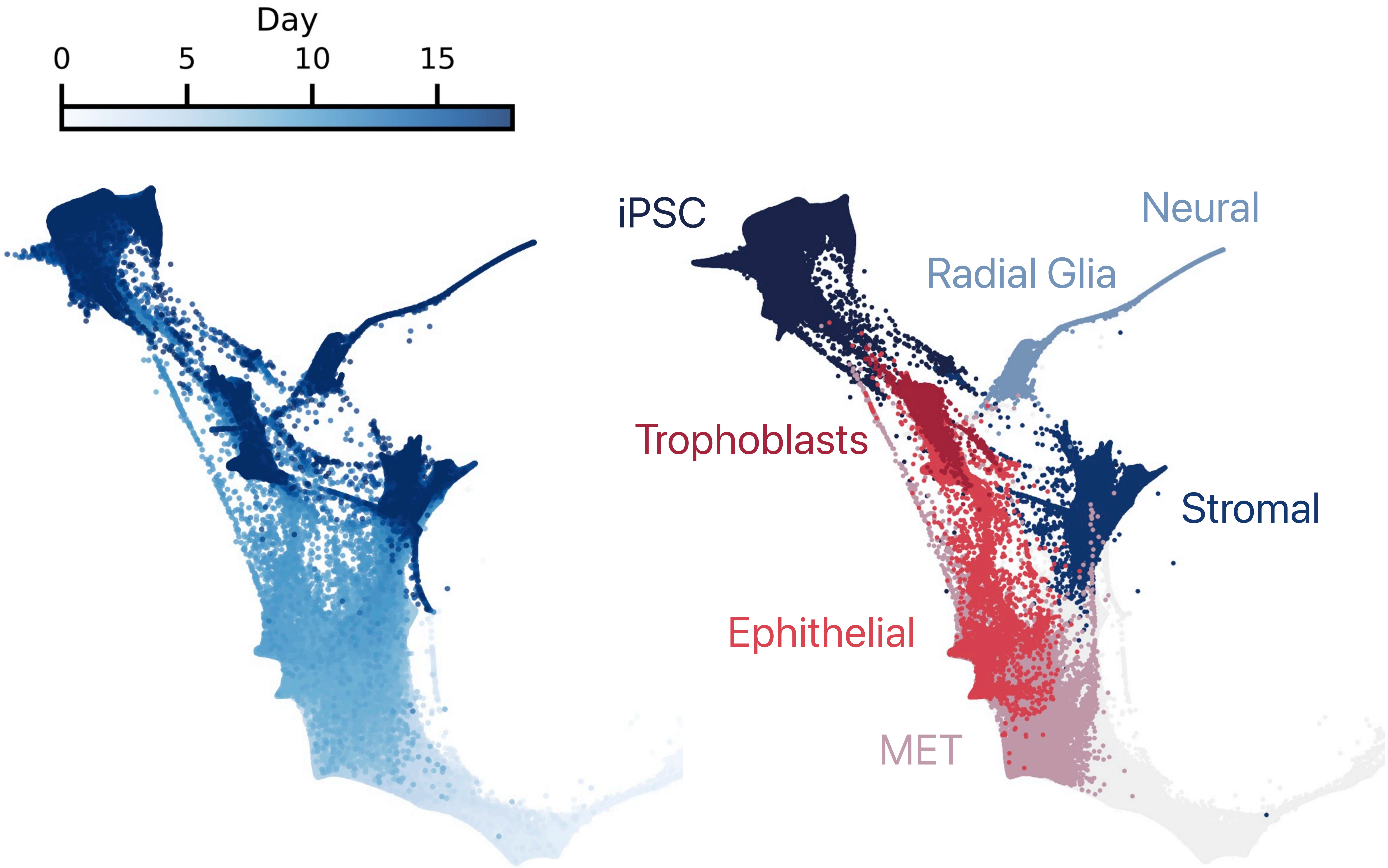
Motivating Example: **Cell Differentiation**

Optimal transport plan for recovering developmental trajectories



Motivating Example: **Cell Differentiation**

Time course of induced pluripotent stem celll development



$P^\star$

fate transition table

	iPSC	Stromal	Neural	Trophoblast	Epithelial	Other
iPSC	0.11		0.01			0.01
Stromal		0.30	0.01		0.01	0.01
Neural			0.04			
Trophoblast				0.06	0.01	0.01
Epithelial			0.05	0.01	0.07	0.07
Other		0.02	0.11		0.03	0.03

Application: **Reconstruct Cell Differentiation Processes**

Part 1

Search

- Installation
- User API
- Developer API
- Contributing
- Tutorials
  - Mapping lineage-traced cells across time points with moslin
  - Analyzing HSPCs with the `TemporalProblem`
  - Alignment of spatial transcriptomics data
  - Mapping gene expression in space
  - Spatiotemporal trajectory of mouse organogenesis
  - Translating multiomics single-cell data
- Examples
- References

Moscot - Multiomics Single-cell Optimal Transport

`moscot` is a framework for optimal transport applications in single cell genomics.

If you find a model useful for your research, please consider citing the `moscot` manuscript (in preparation) as well as the publication introducing the model, which can be found in the corresponding documentation.

Installation

Learn how to install `moscot`.

User API

Find a detailed documentation of `moscot`.

Contributing

Add a functionality or report a bug.

Examples

Find brief and concise examples of `moscot`.

Tutorials

Check out how to use `moscot` for data analysis.

Manuscript

Please have a look at our manuscript [Klein *et al.*, 2023] to learn more.

with backend

Perspective

These solvers are used as subroutines, tricks to speed/stabilize them matter

- (Nonlinear) acceleration of fixed point iterations, and their differentiation
- Faster kernel multiplications (with caveat that they must remain >0)
- Amortized and/or clever initializations, notably for LR approaches.

Accel

Grid

MetaOT

Theory and statistics help setting right levels of regularization

- Stats: Sinkhorn vs. LP plugin [Altschuler+17, Genevay+18, Mena+19, Stromme+22]
- Debiasing approaches $\hat{W}_{\text{reg}}(\mu,\nu) = W_{\text{reg}}(\mu,\nu) - \frac{1}{2}(W_{\text{reg}}(\mu,\mu) + W_{\text{reg}}(\nu,\nu))$

  [Genevay+17, Feydy+18, Chizat+22] SinkDiv

More structured losses, beyond linear/quadratic [Alvarez-Melis+19, Redko+20]

Choosing cost-functions adaptively [Cuturi+11, Paty+19, Lin+20, Chuang+23]

Outline of the Tutorial

Prelude Warm-Up: Starting with Optimal Matchings

Part 1 Kantorovich Formulation of OT and Computations

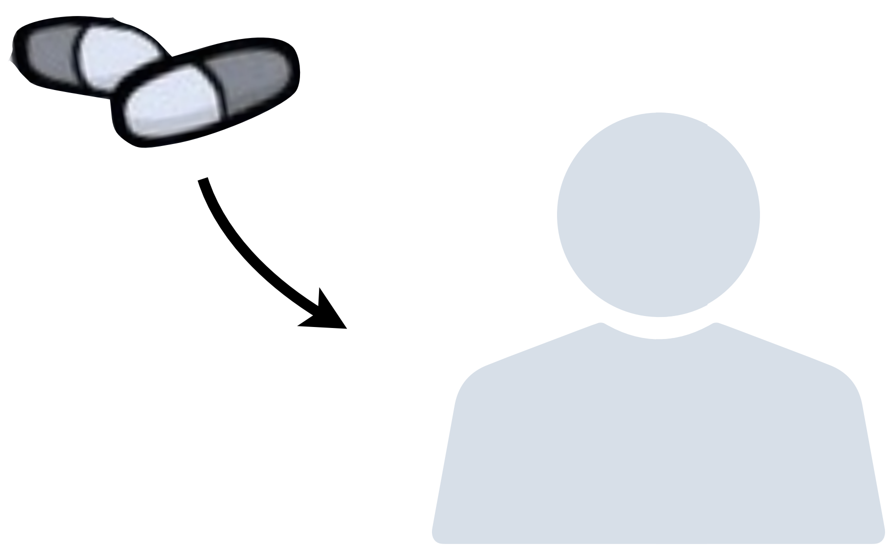
Part 2 **Duality, Monge Formulations and Brenier Theorems**

- Kantorovich duality
- *c*-concavity, Gangbo-McCann and Brenier theorems
- Entropic map estimators, learning neural Monge maps

Part 3 Modeling Measure Dynamics with Optimal Transport

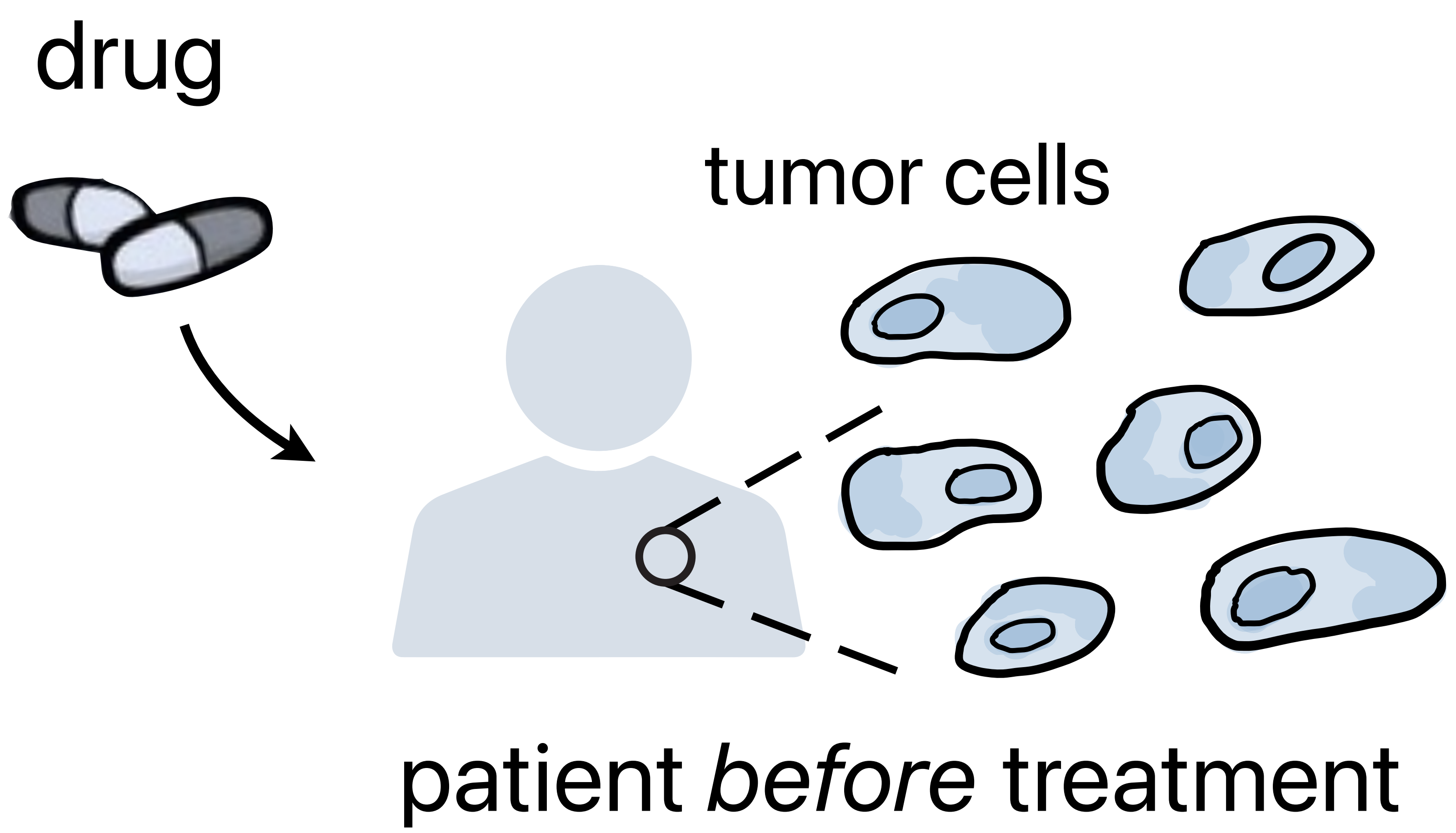
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile



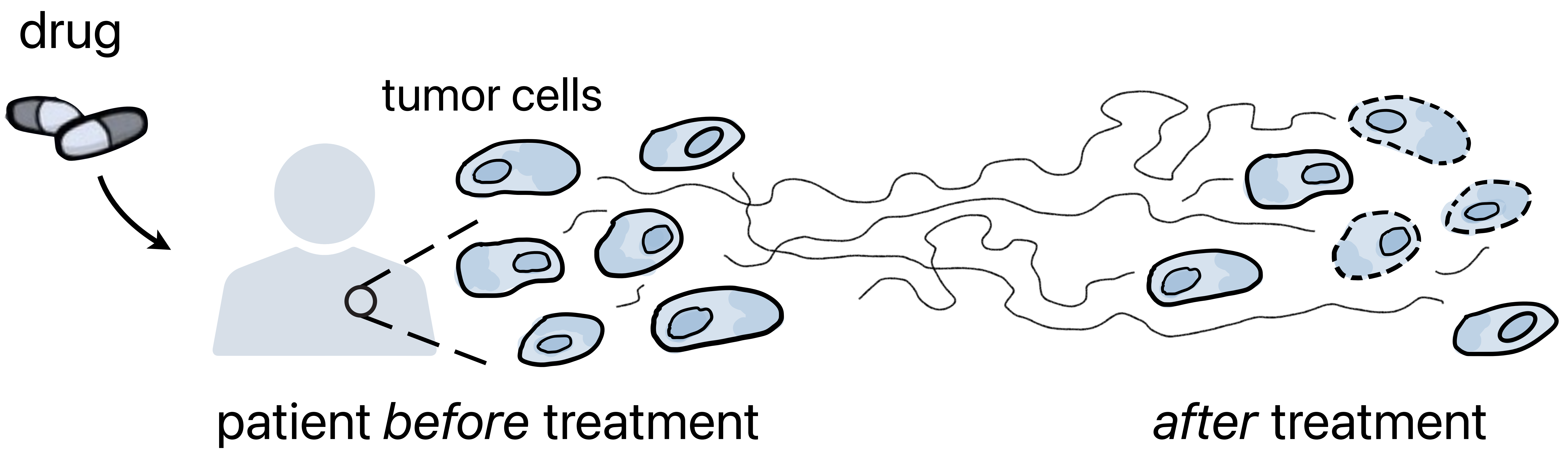
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile



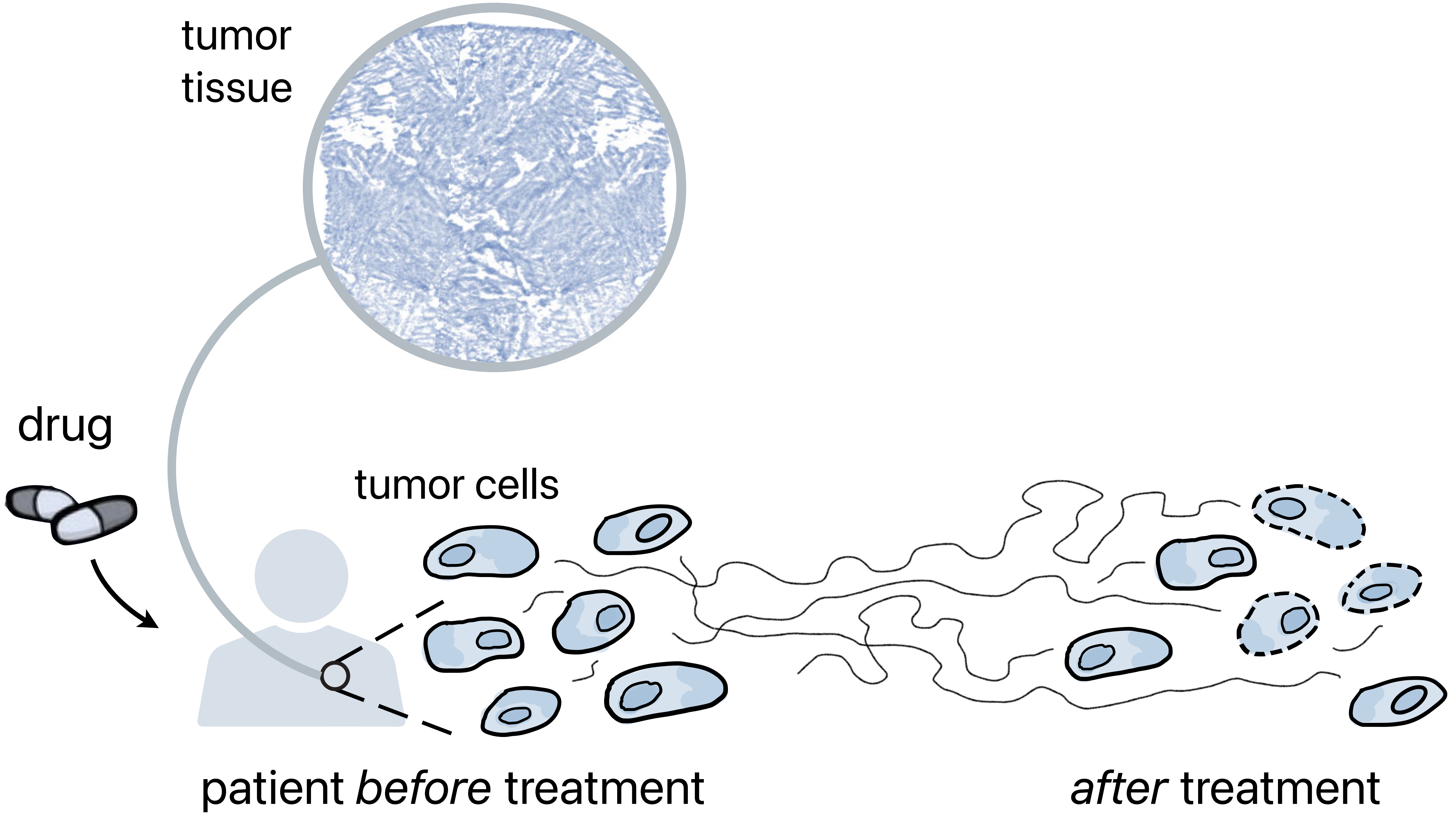
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile



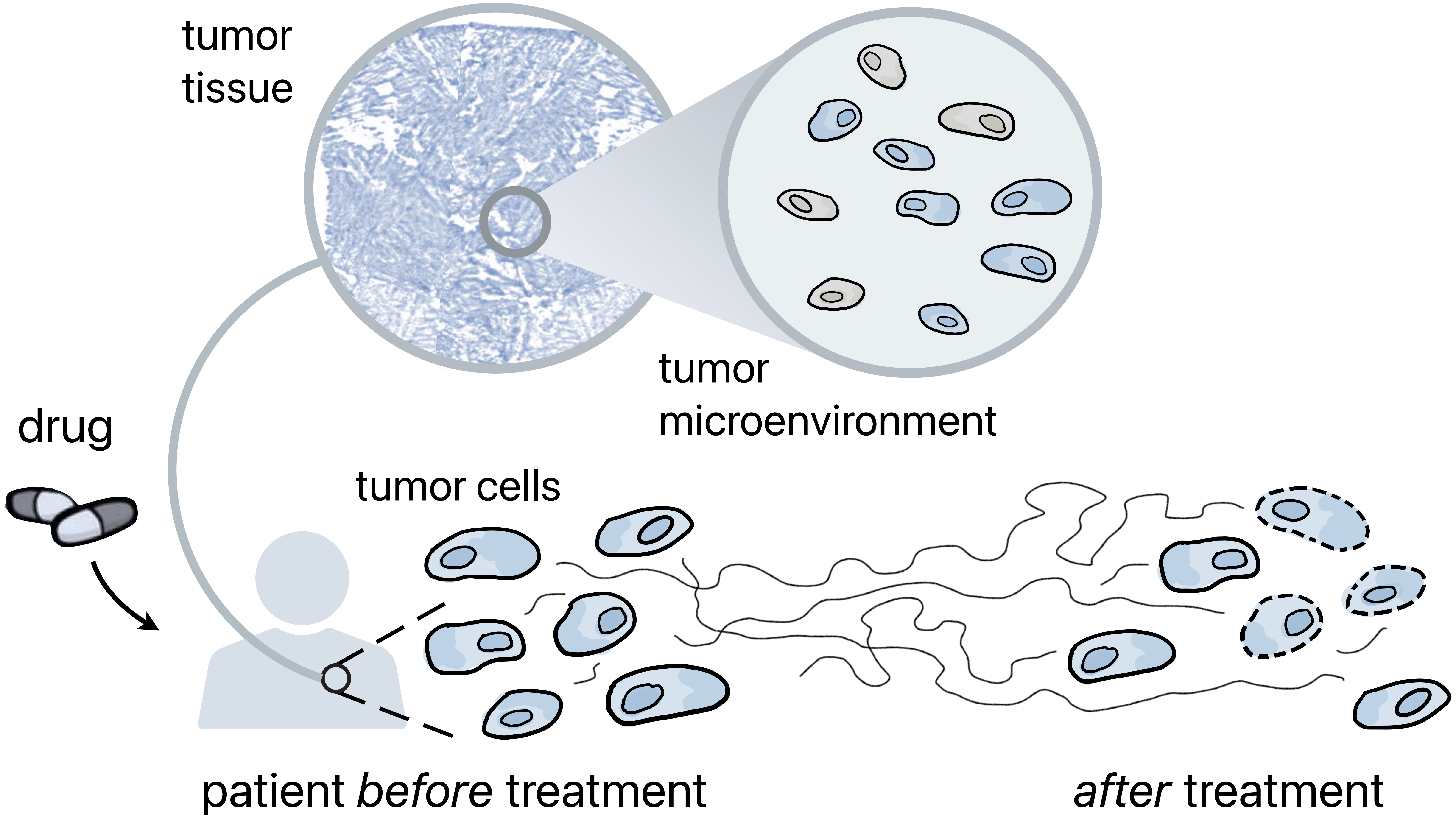
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile



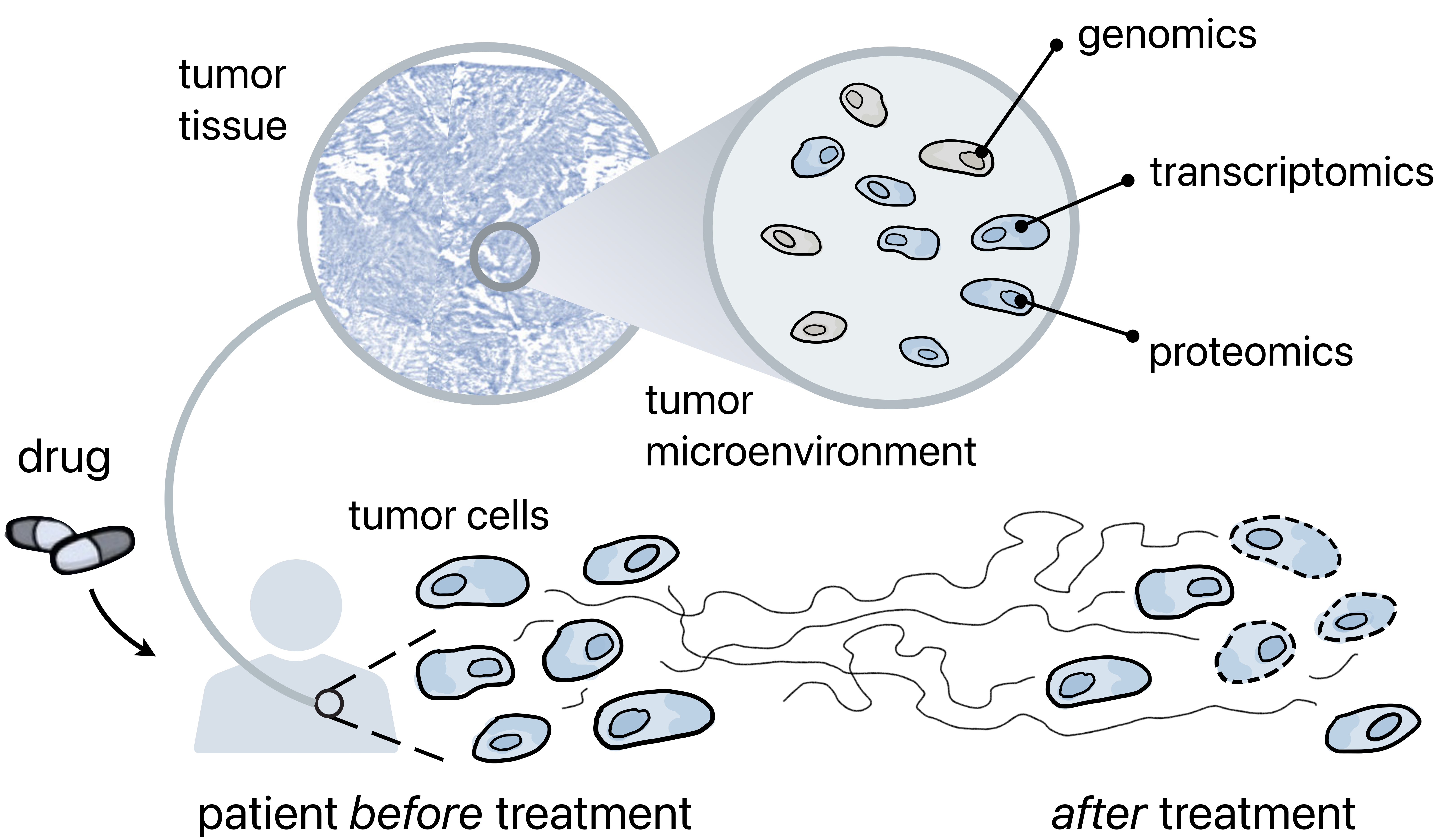
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile



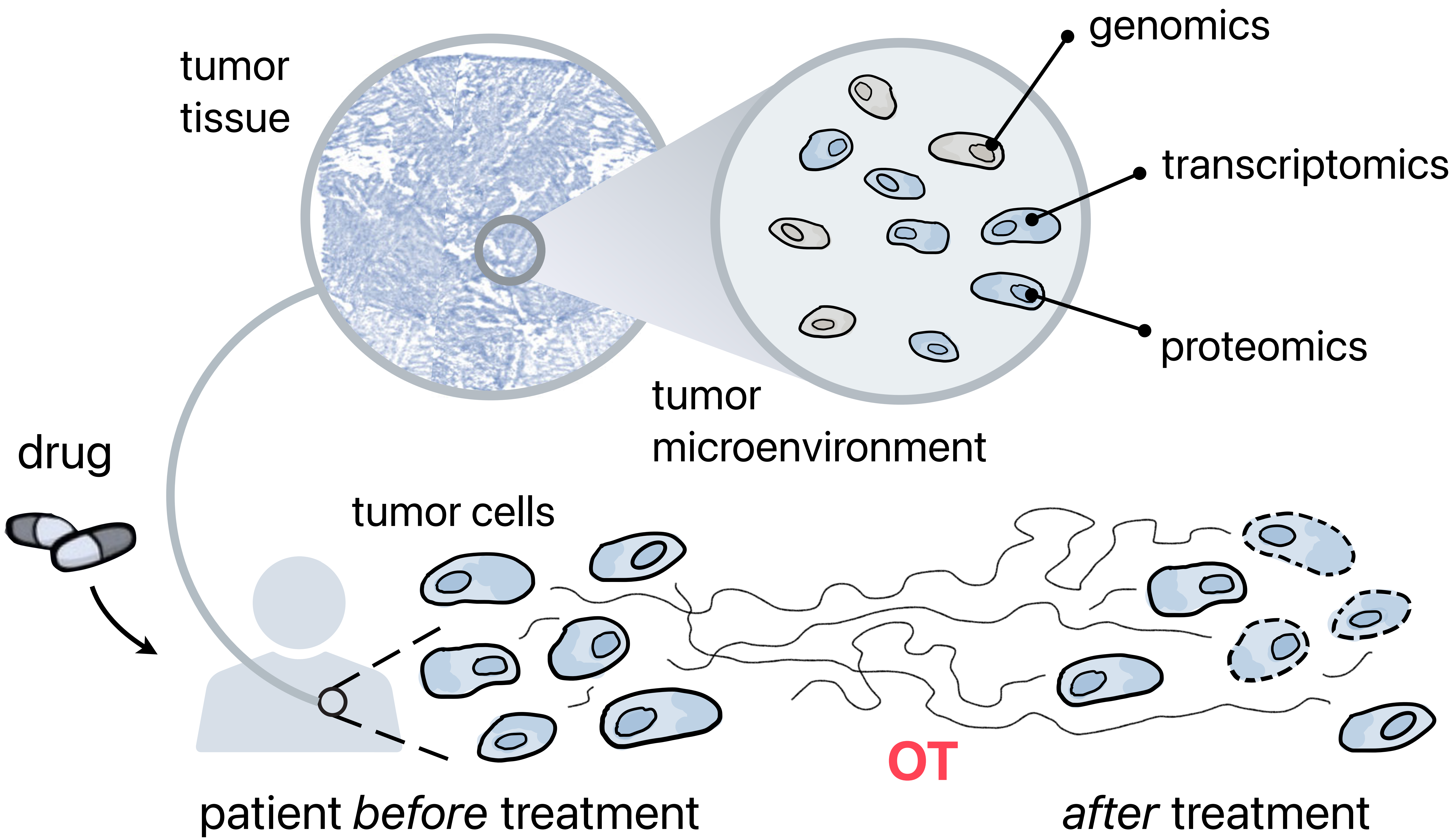
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile



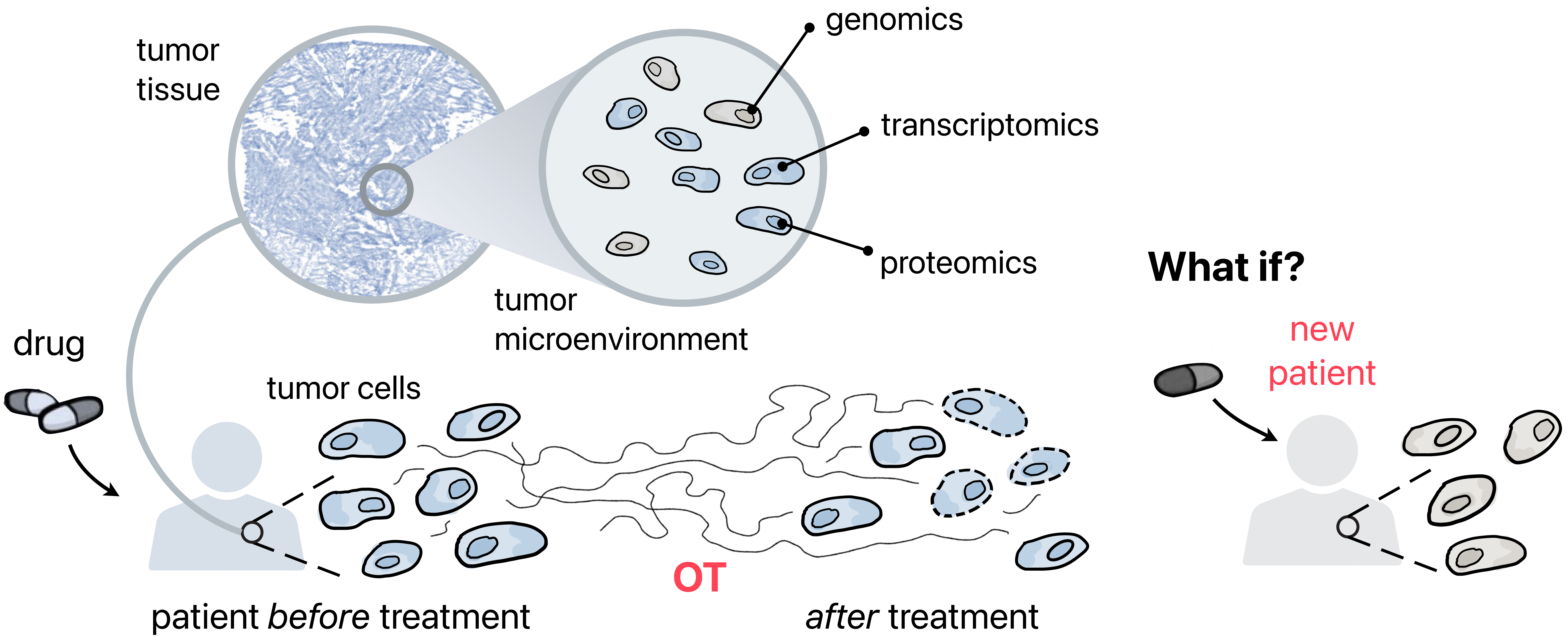
Motivating Example: Personalized Medecine

Goal: give patients **targeted therapies** tailored to their molecular profile

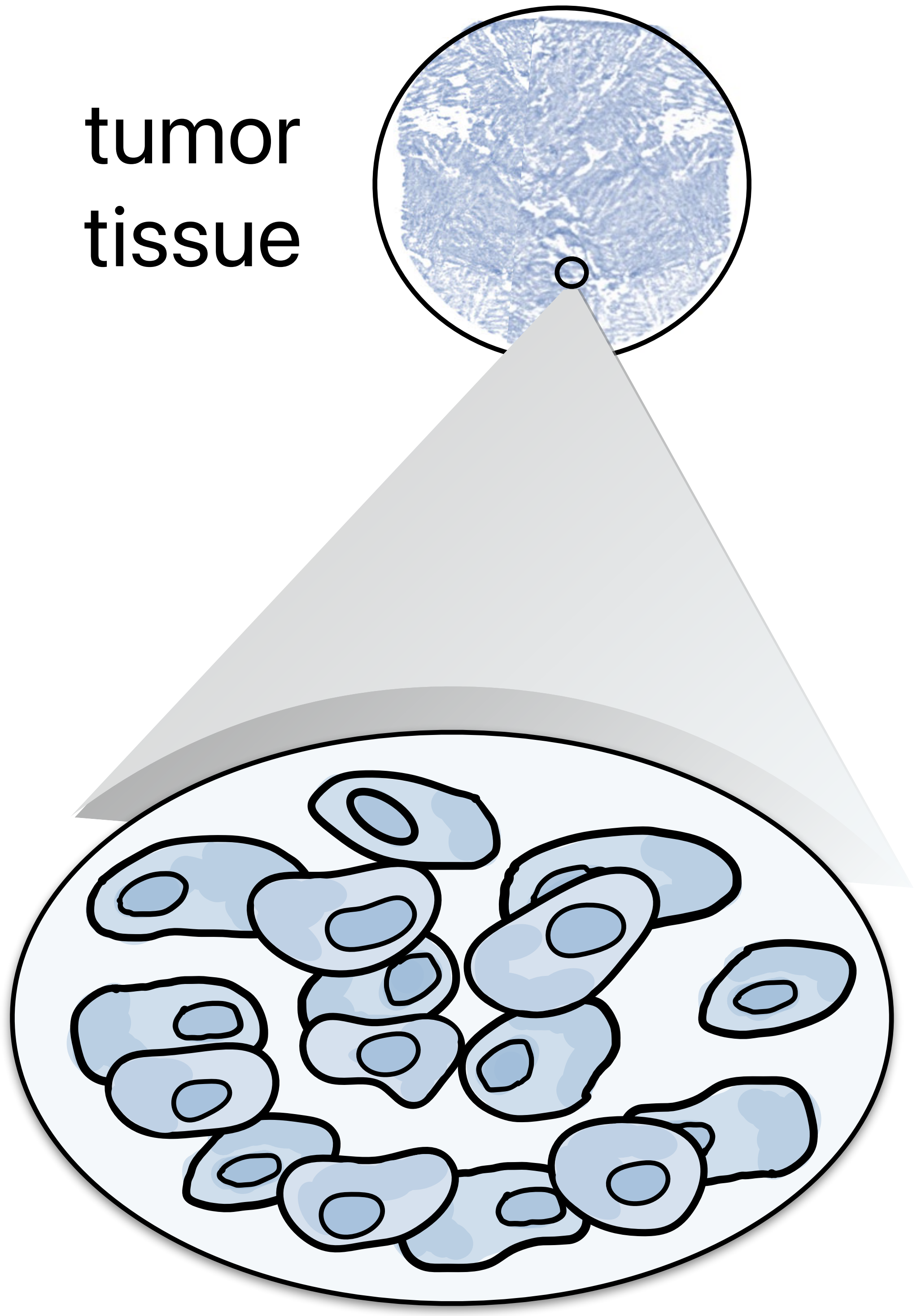


Motivating Example: Personalized Medecine

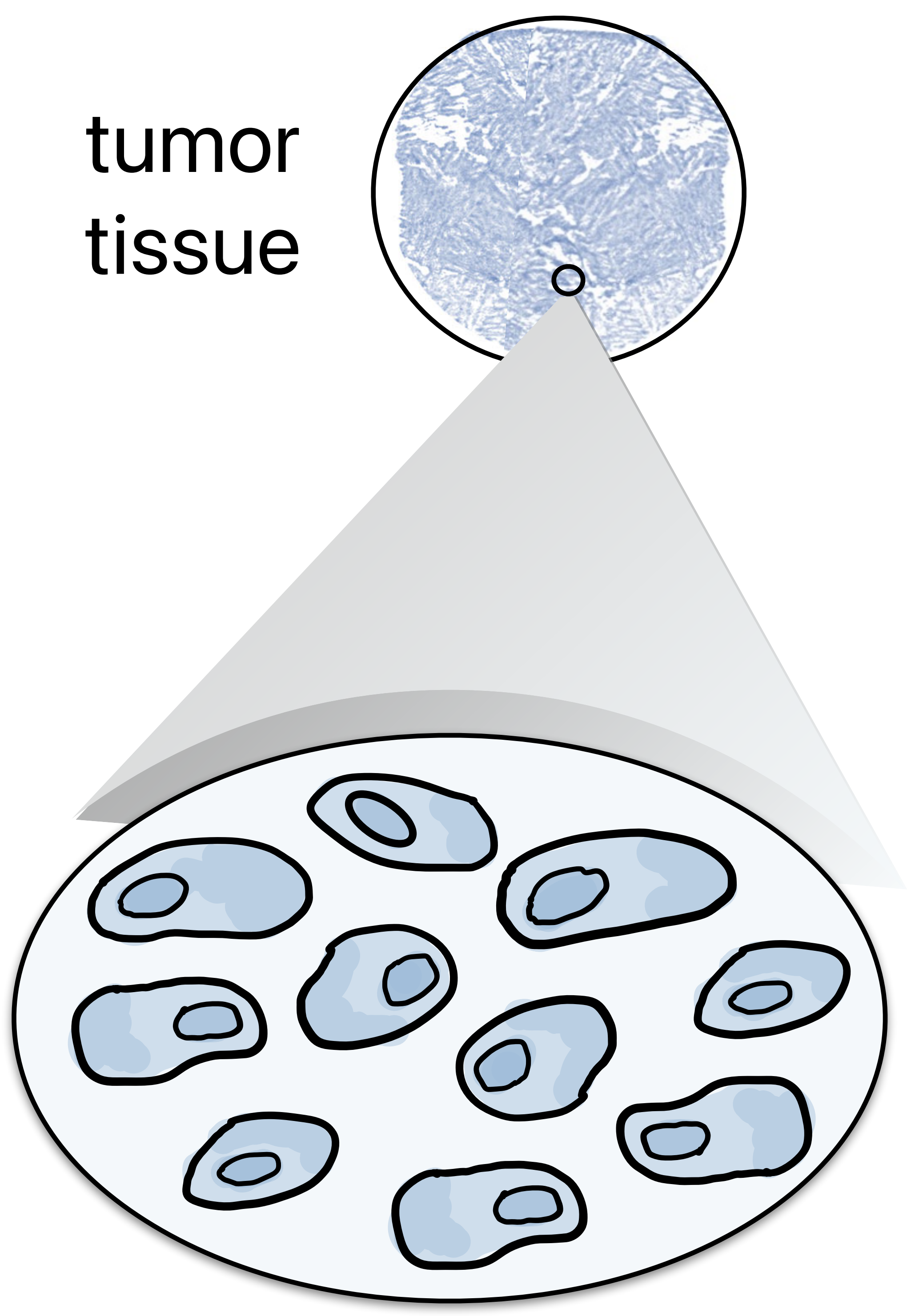
Goal: give patients **targeted therapies** tailored to their molecular profile

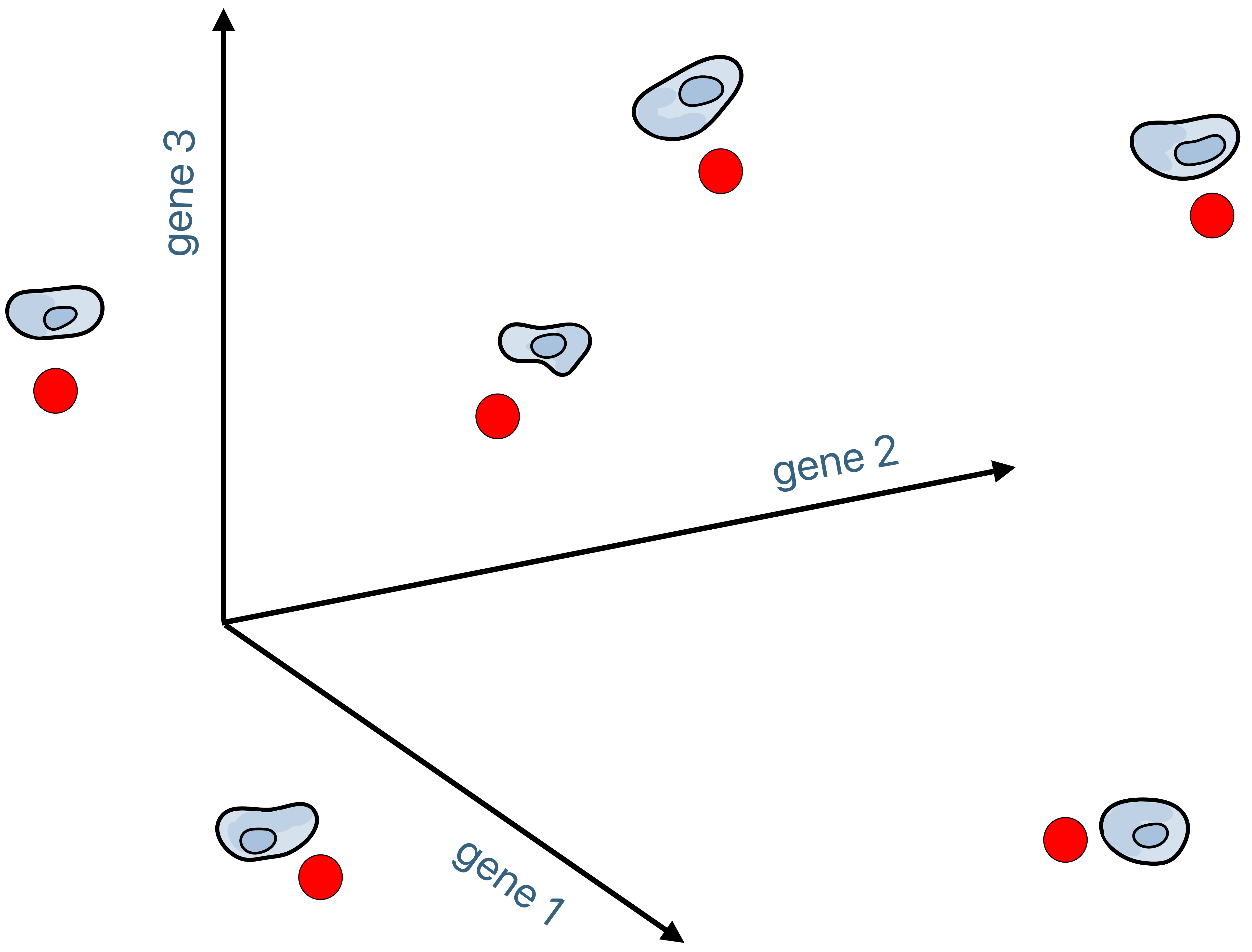


Motivating Example: Personalized Medicine



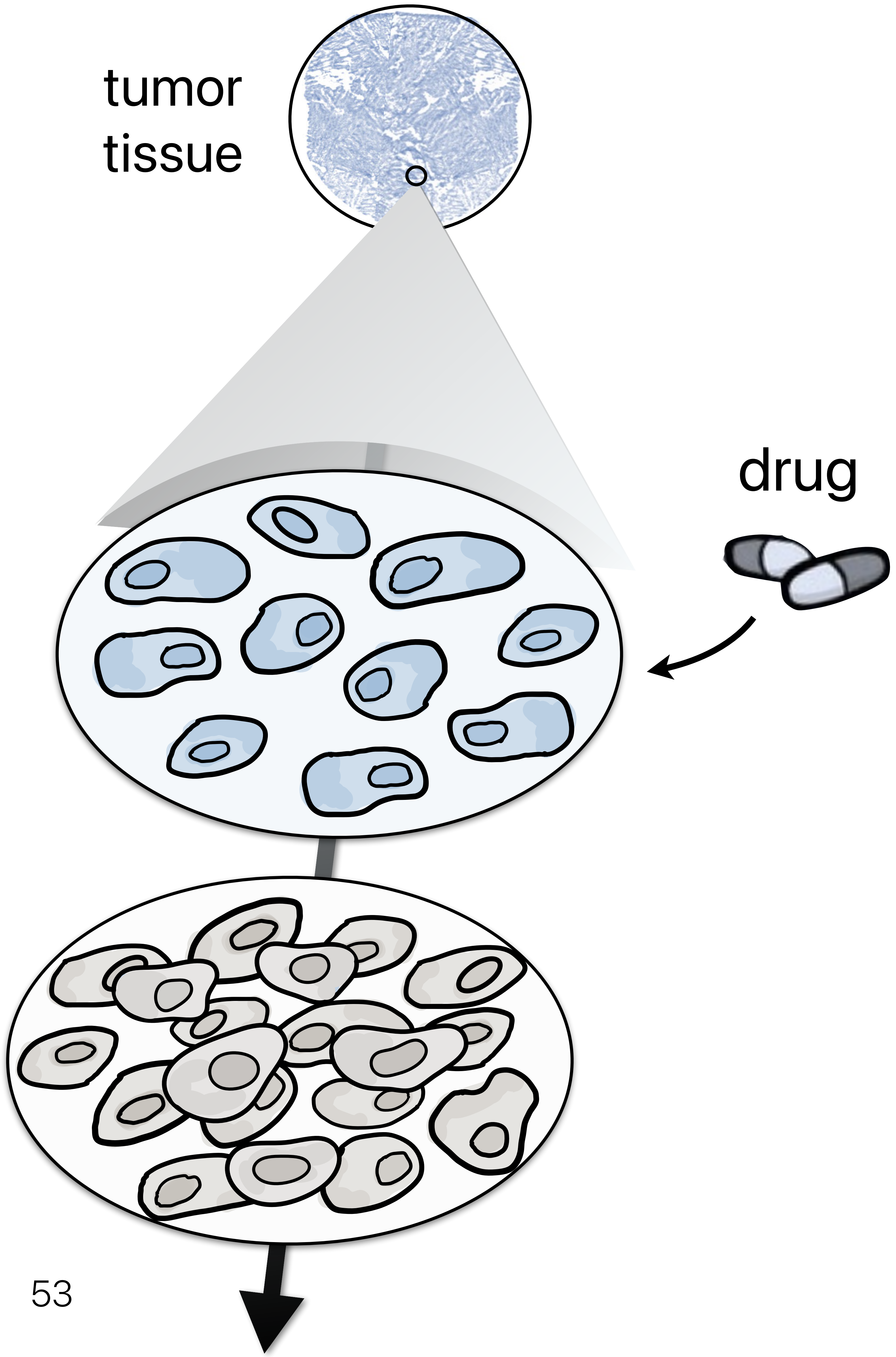
Motivating Example: Personalized Medicine

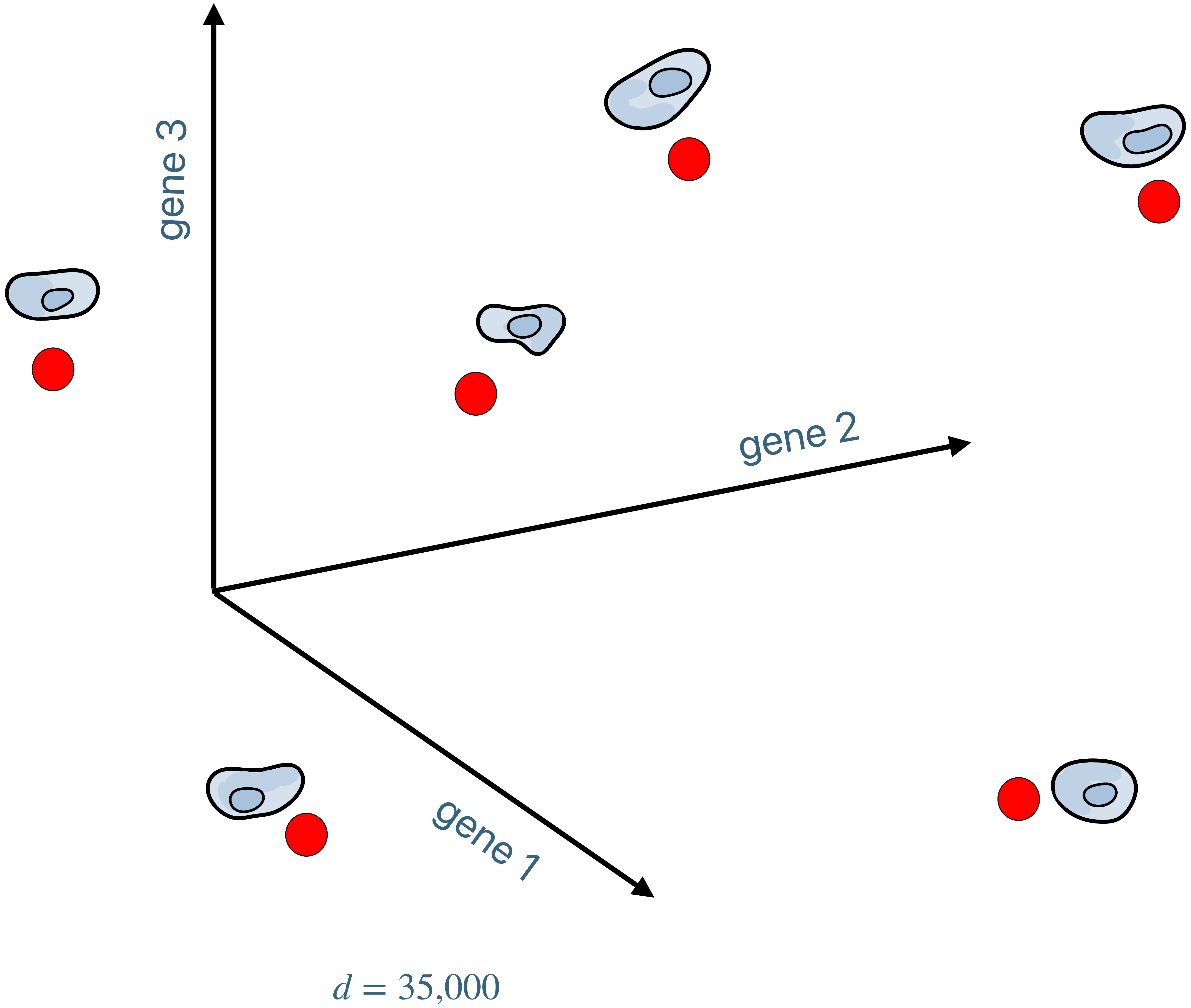




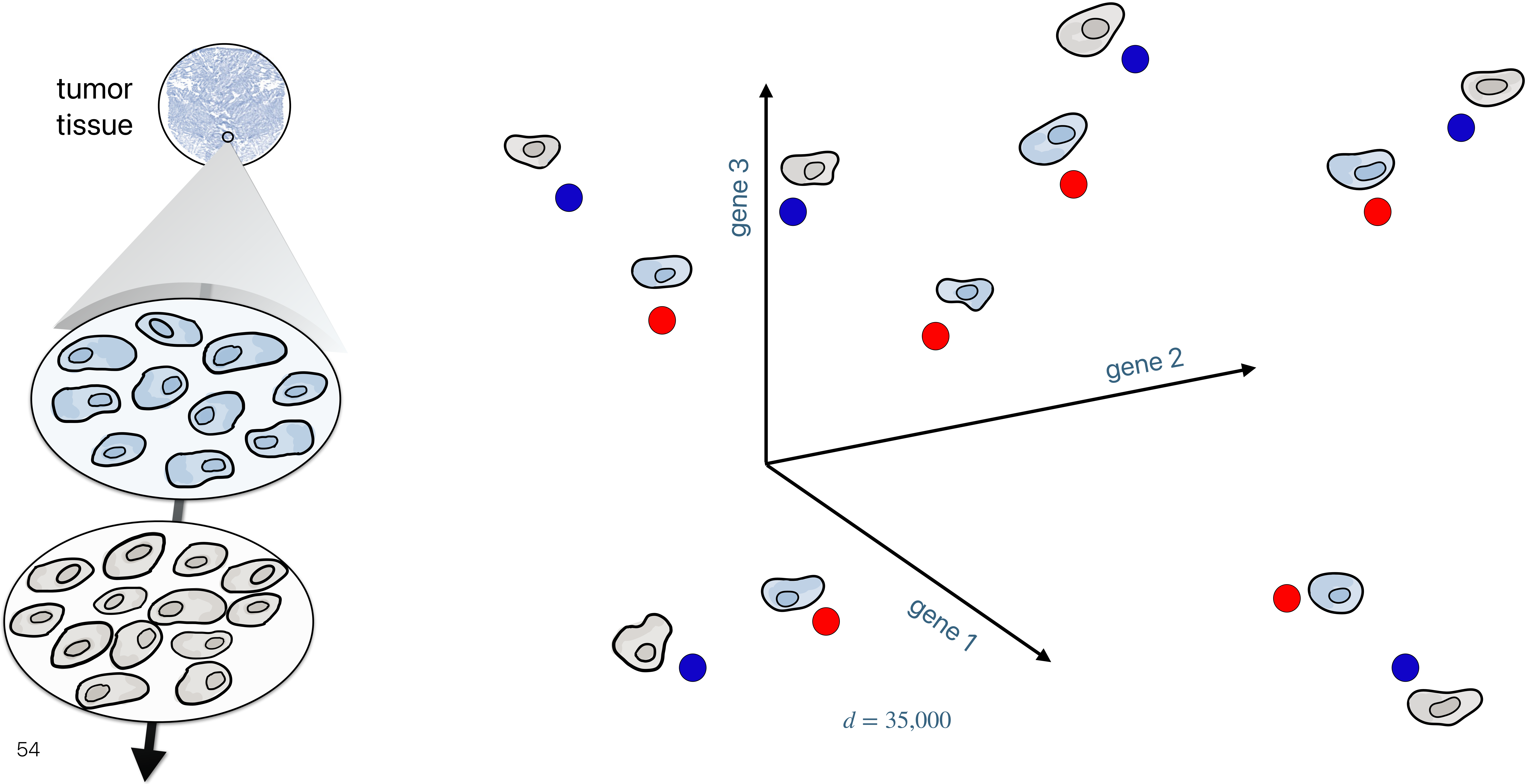
$d = 35{,}000$

Motivating Example: Personalized Medicine

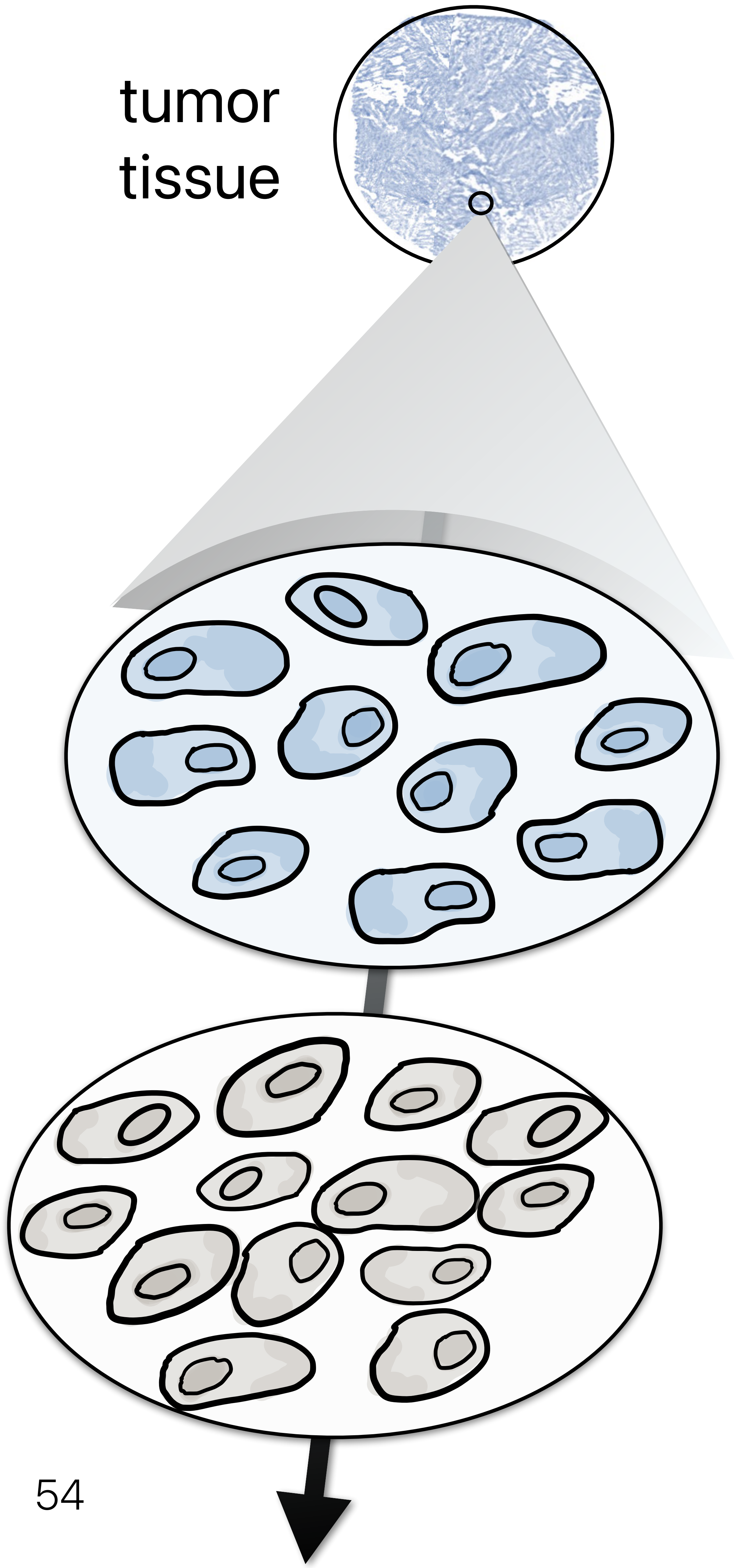


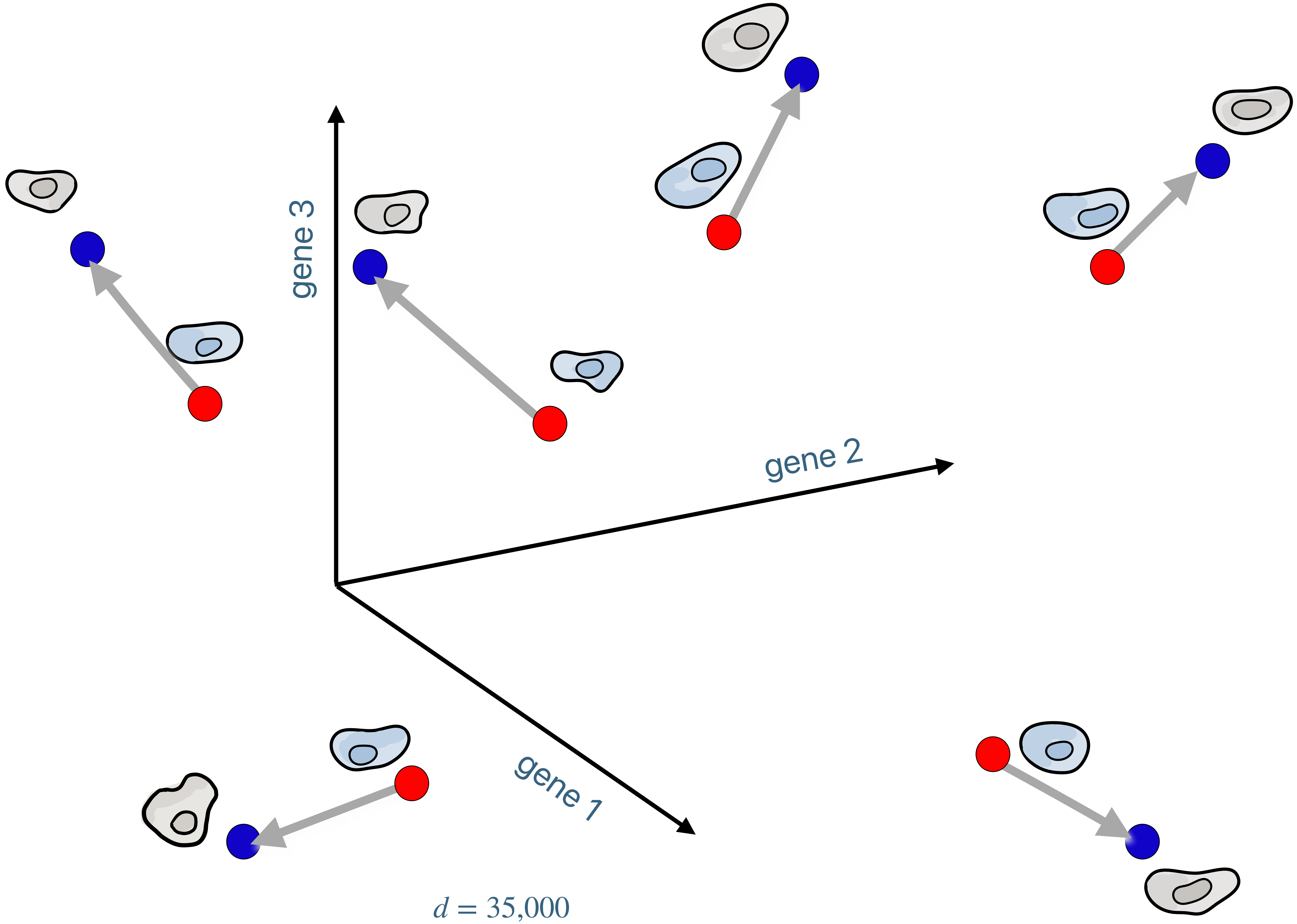


Motivating Example: Personalized Medicine

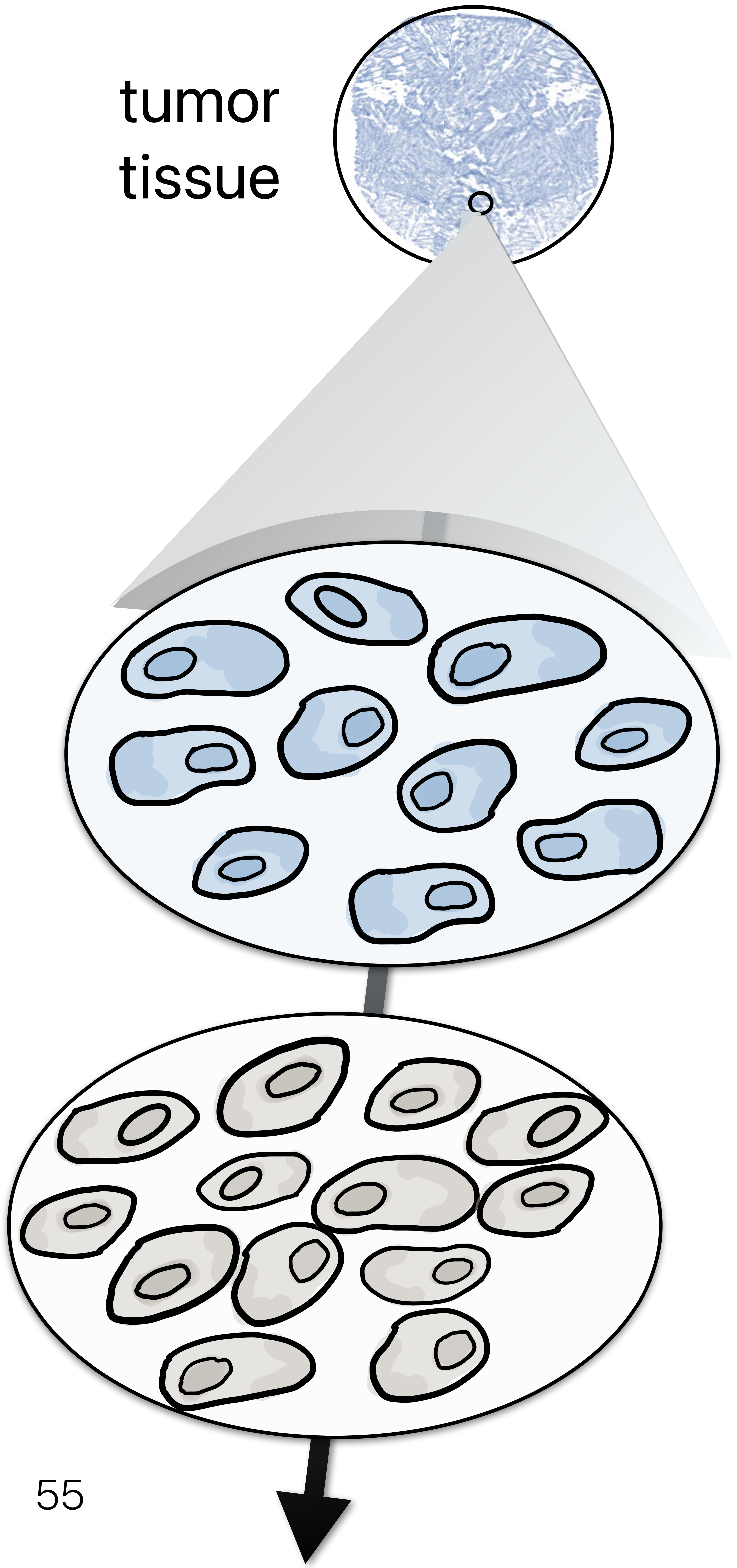


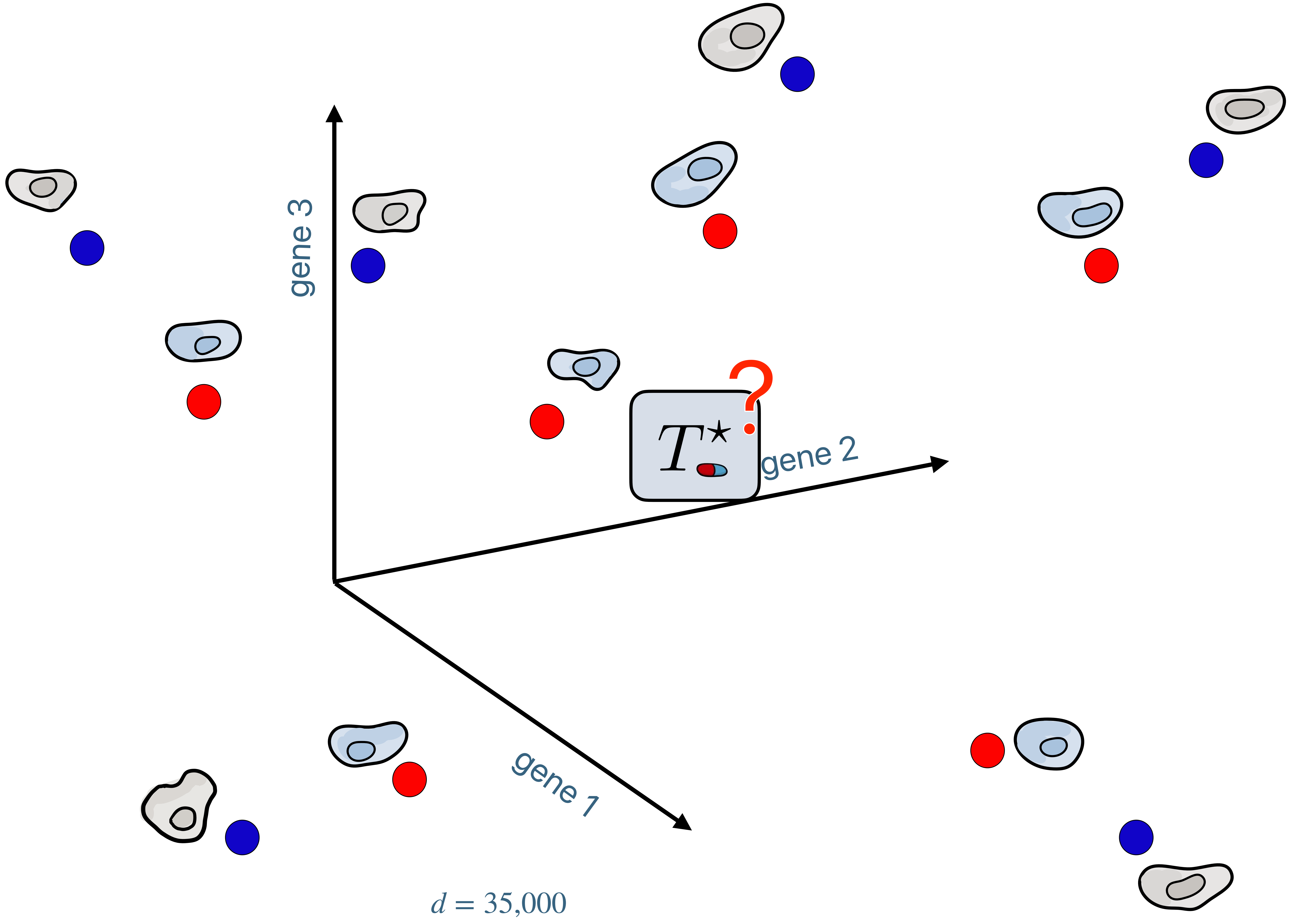
Motivating Example: Personalized Medicine



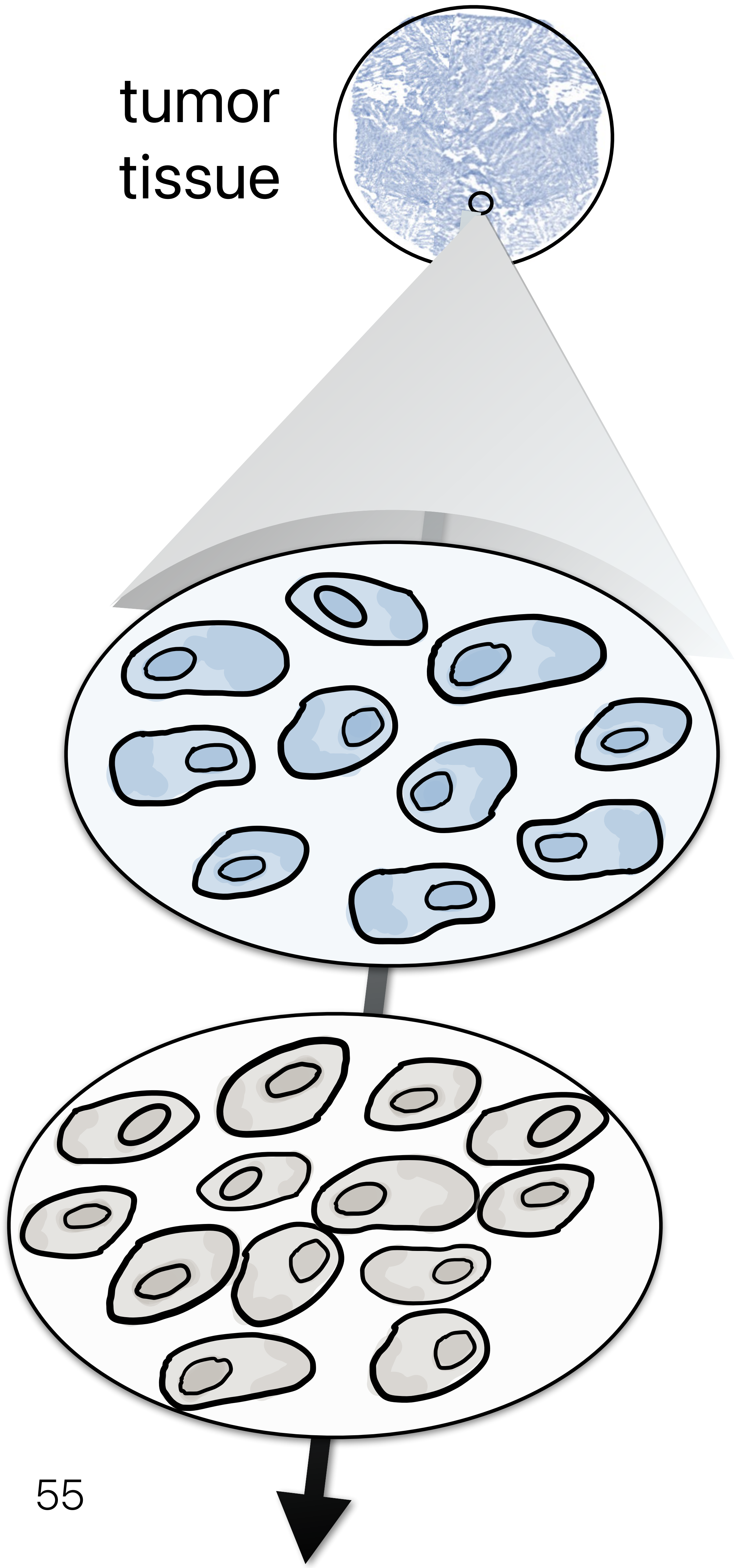


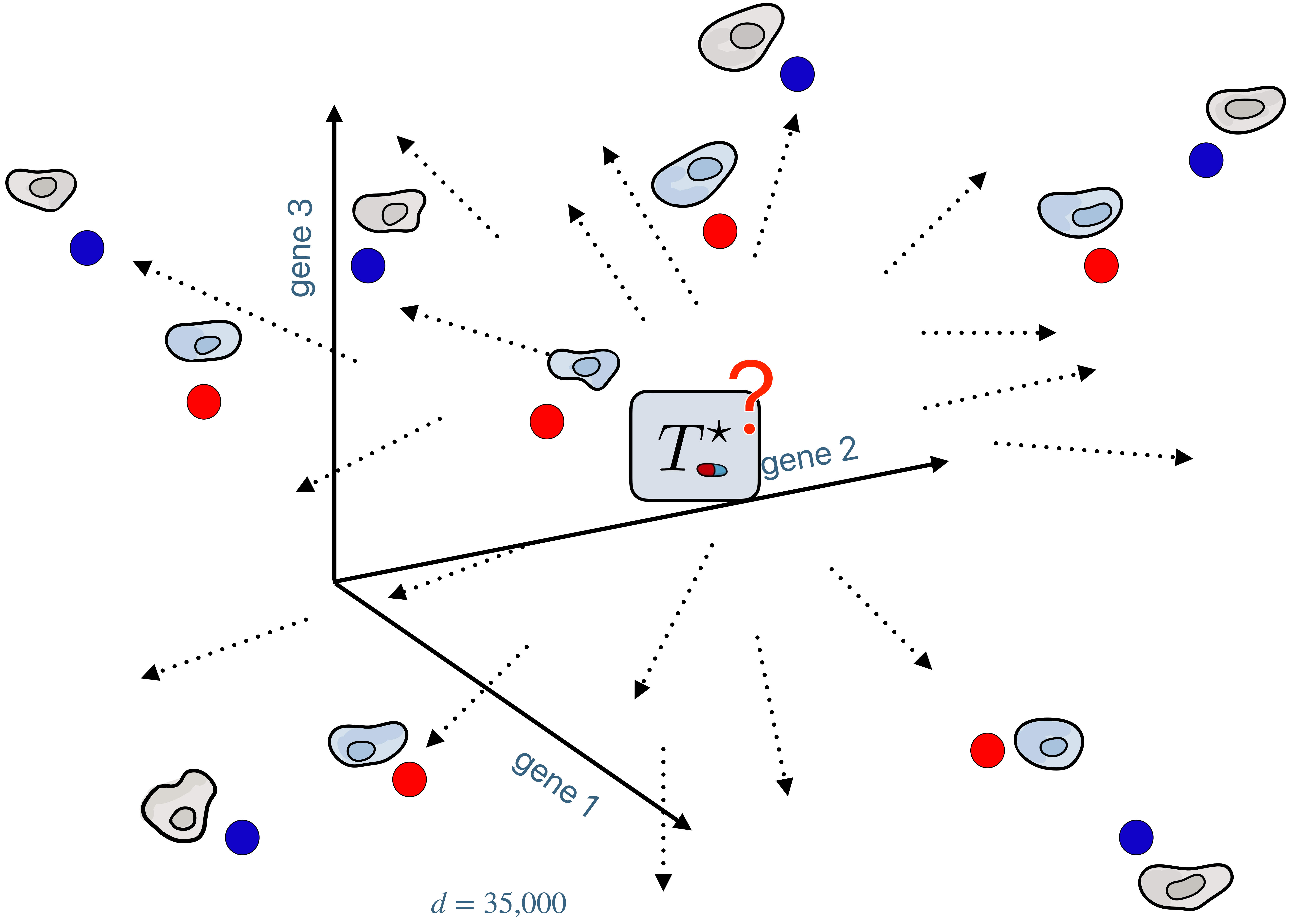
Motivating Example: Personalized Medicine



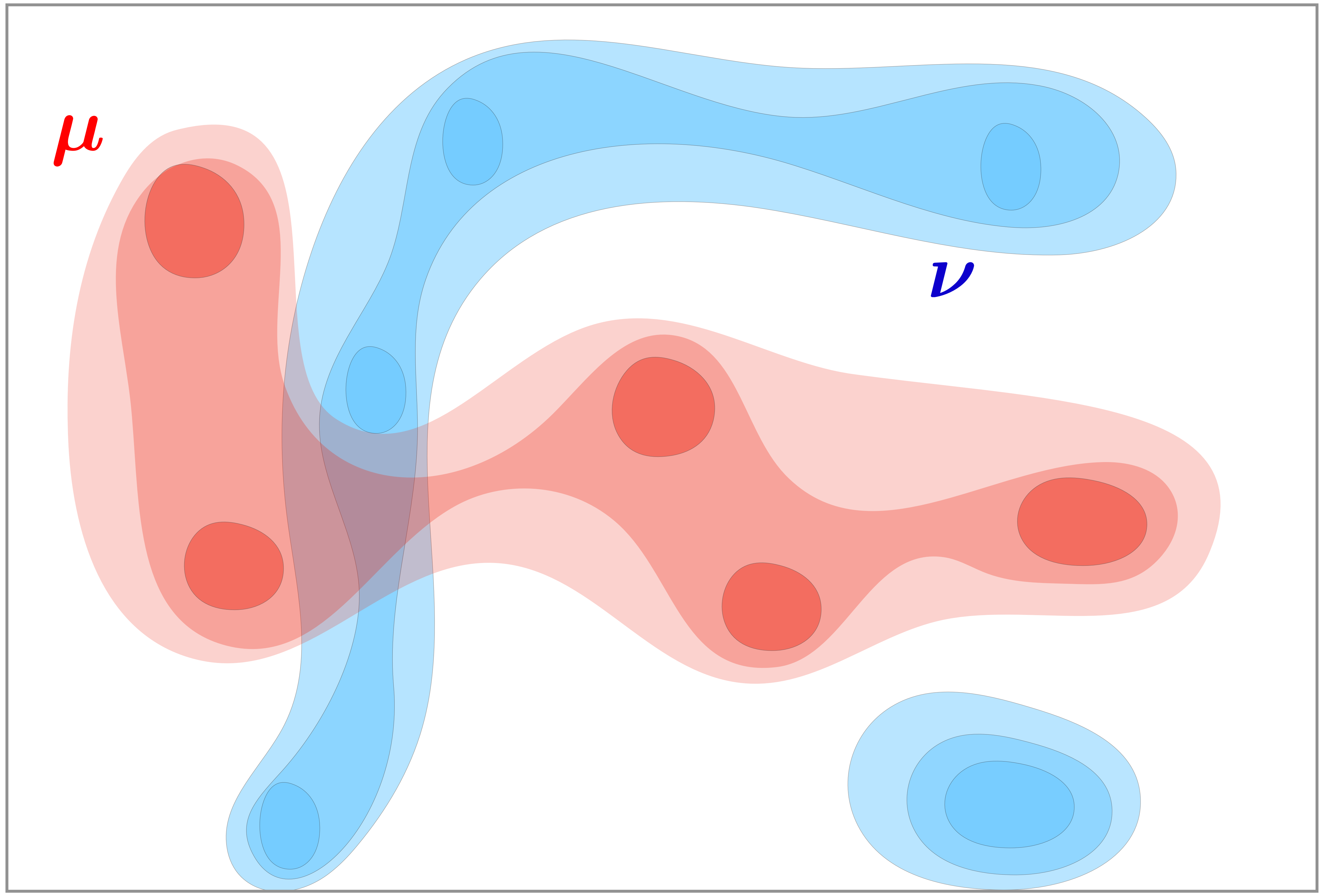


Motivating Example: Personalized Medicine



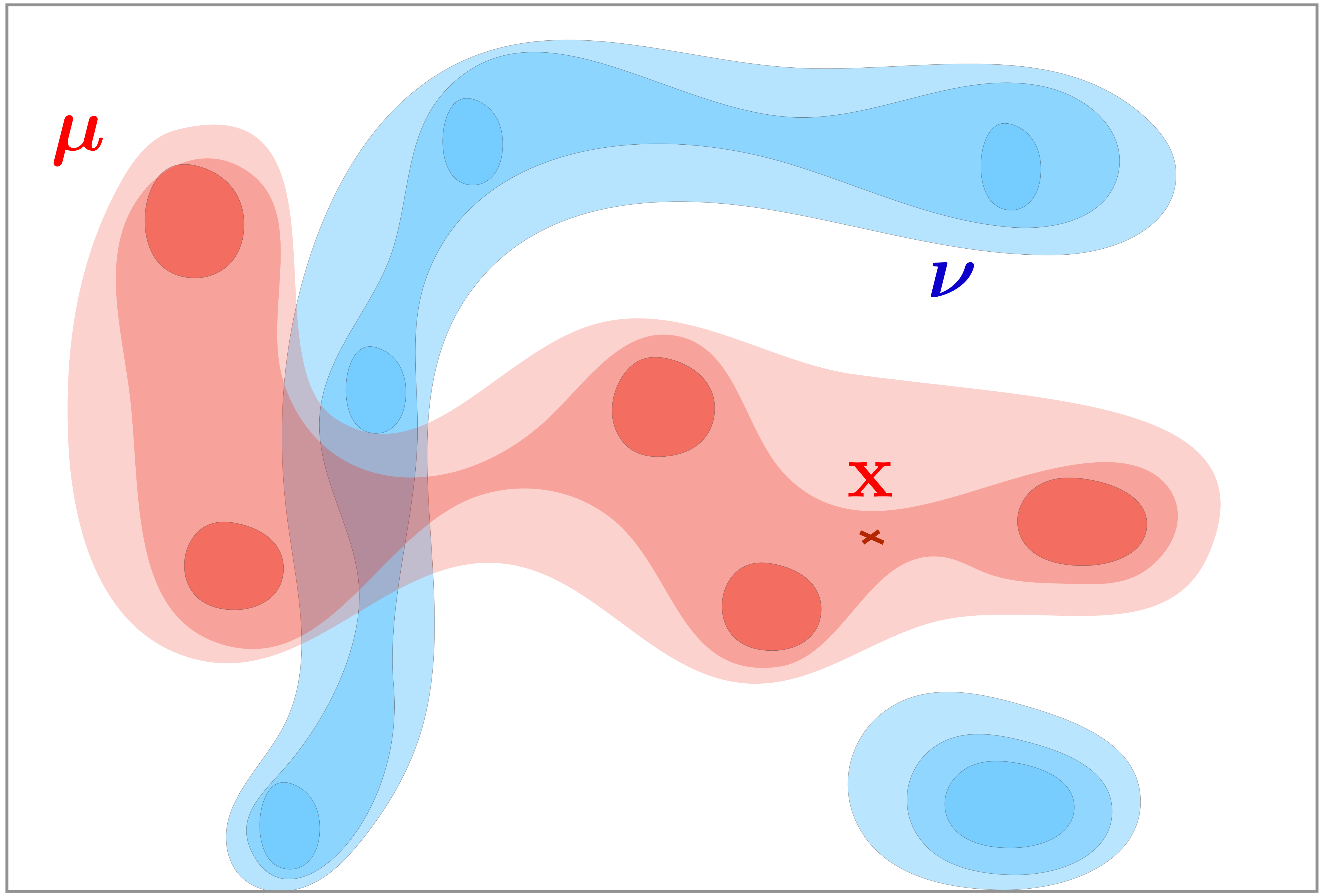


Monge Maps



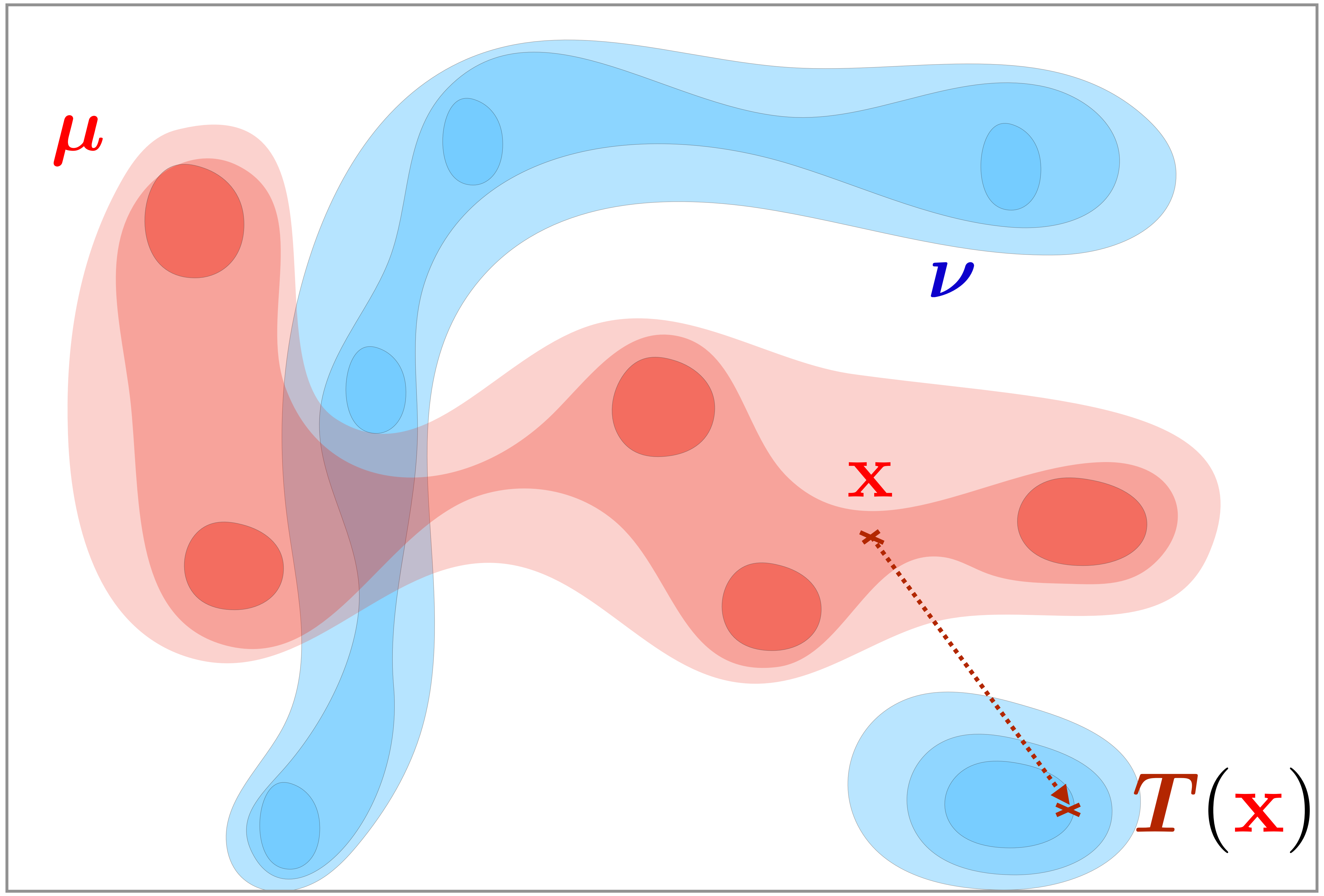
Monge Brenier Gangbo McCann Caffarelli

Monge Maps



Monge Brenier Gangbo McCann Caffarelli

Monge Maps



Monge

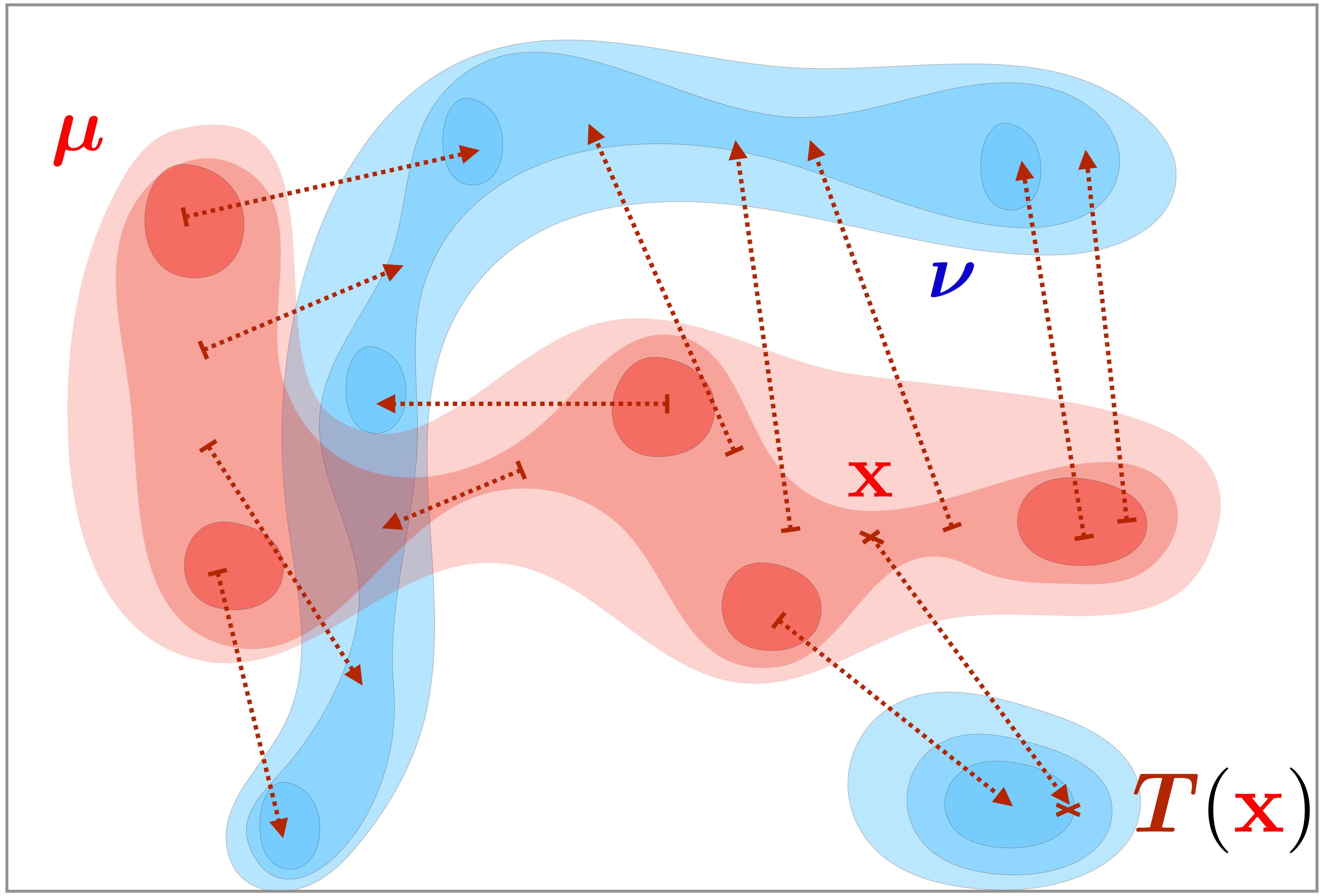
Brenier

Gangbo

McCann

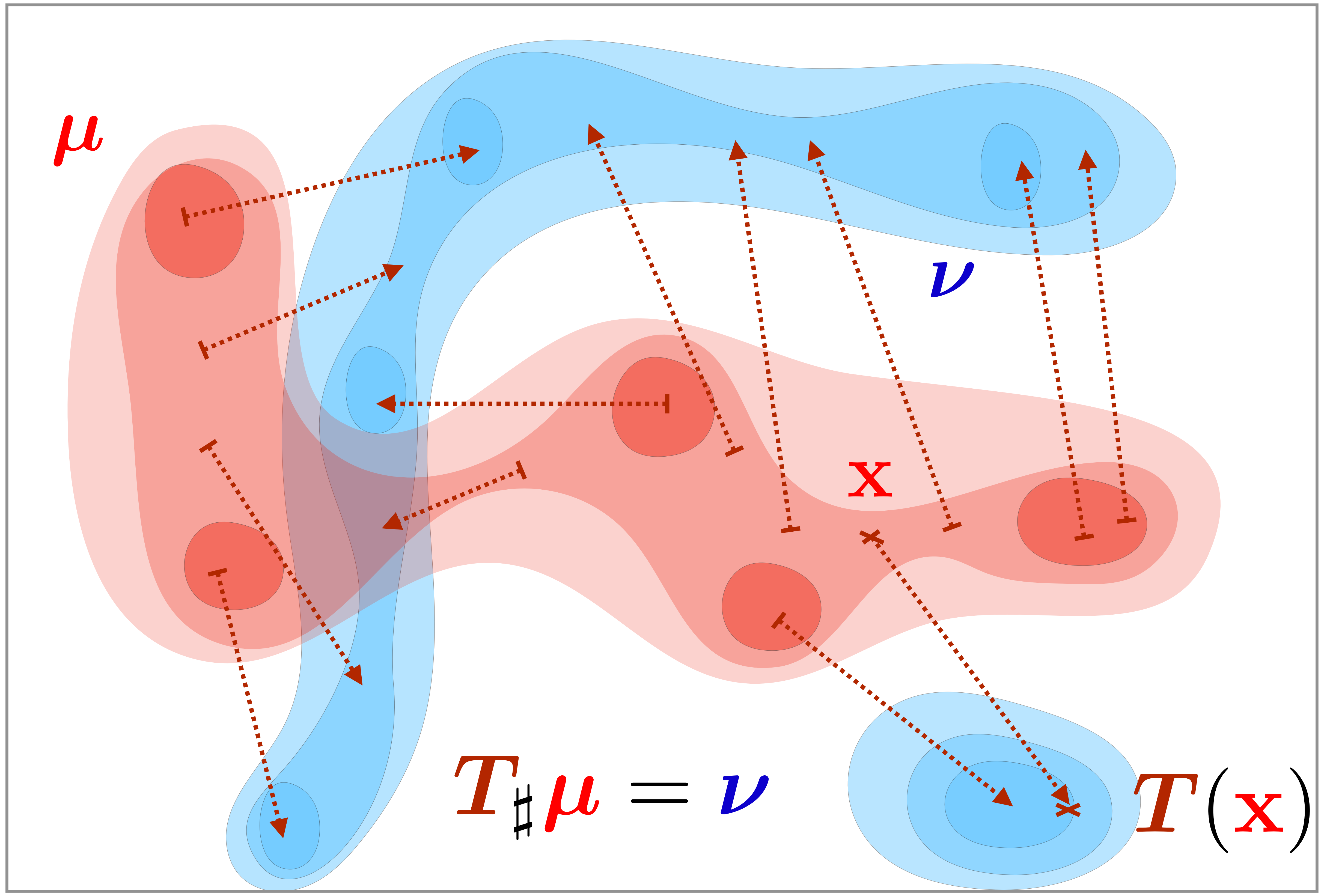
Caffarelli

Monge Maps


μ
ν
$\mathbf{x}$
$T(\mathbf{x})$


Monge
Brenier
Gangbo
McCann
Caffarelli

Monge Maps



μ
ν
$\mathbf{x}$
$T_\sharp \mu = \nu$
$T(\mathbf{x})$

Monge

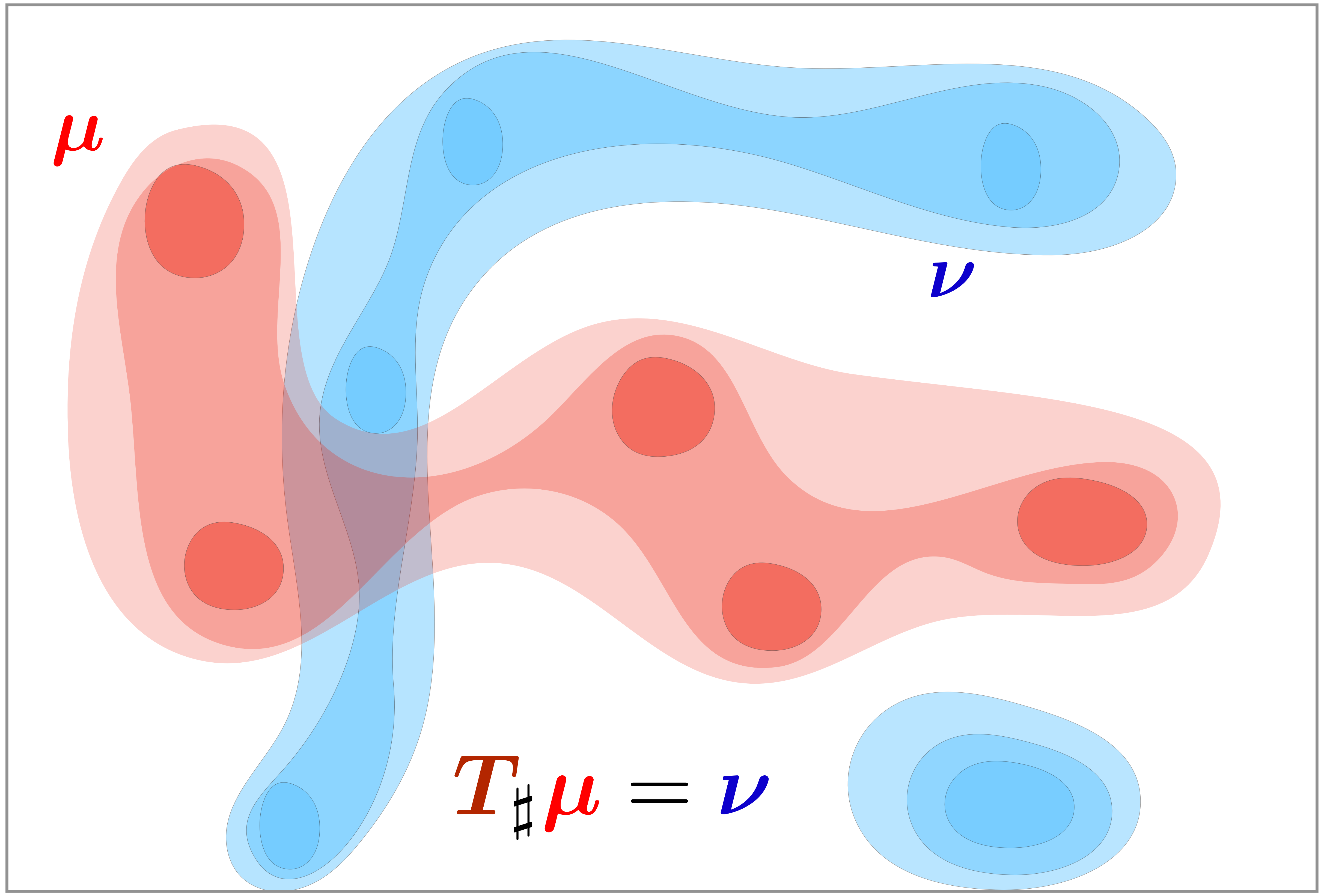
Brenier

Gangbo

McCann

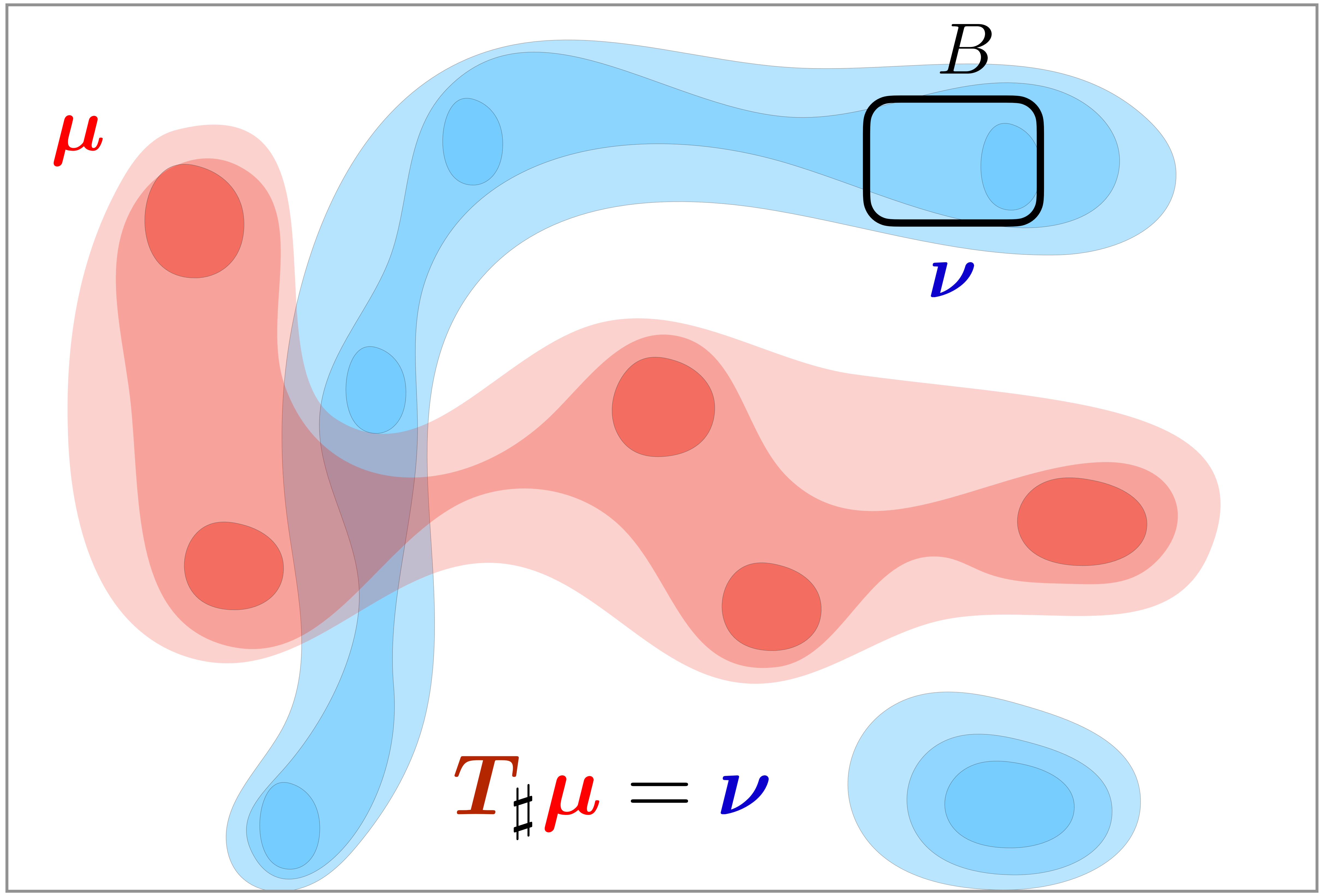
Caffarelli

Monge Maps



$$T_\sharp \mu = \nu$$
$$\Updownarrow$$
$$\forall B \subset \mathbb{R}^d, \nu(B) = \mu(T^{-1}(B))$$

Monge Maps

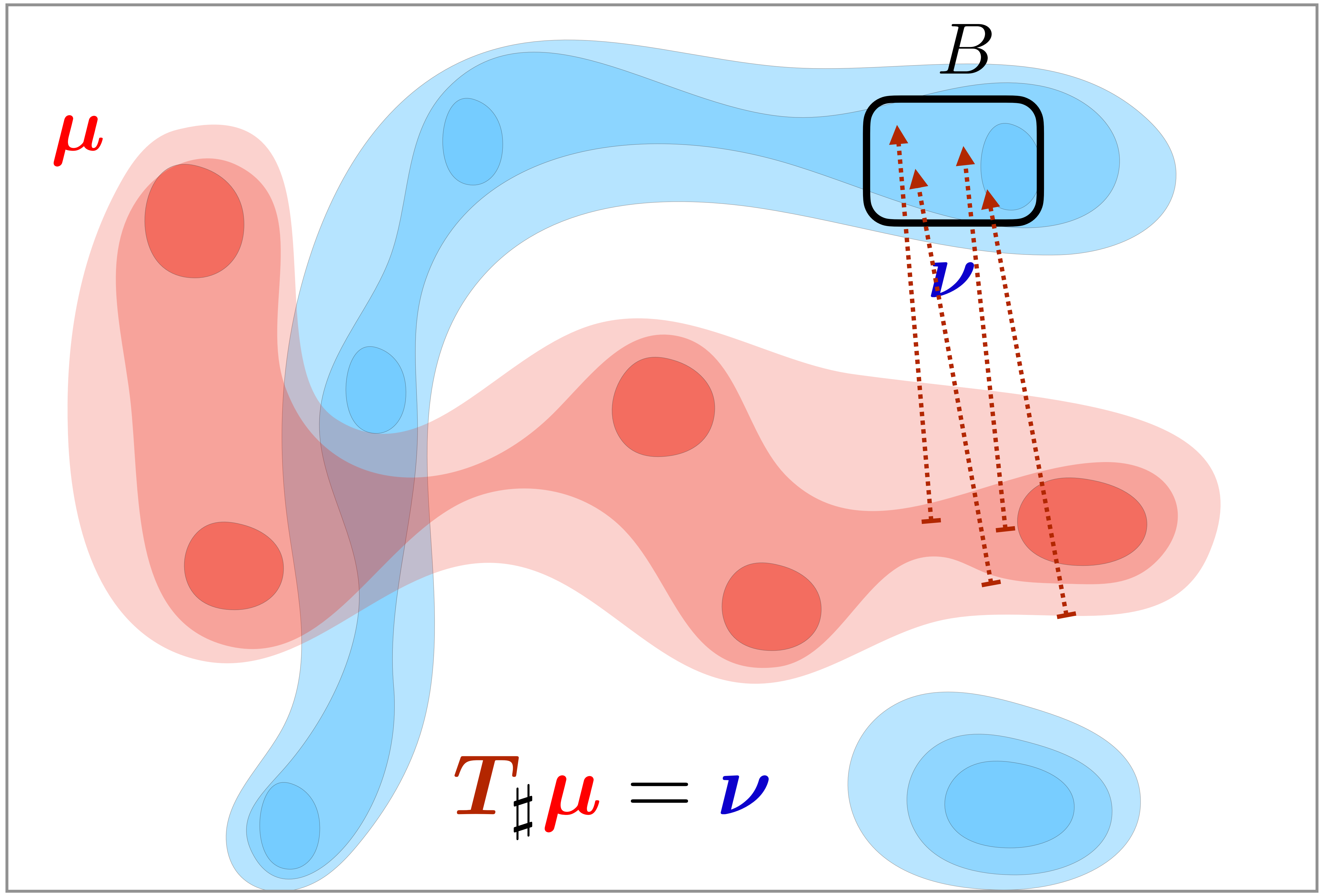


$$T_\sharp \mu = \nu$$

$$\Updownarrow$$

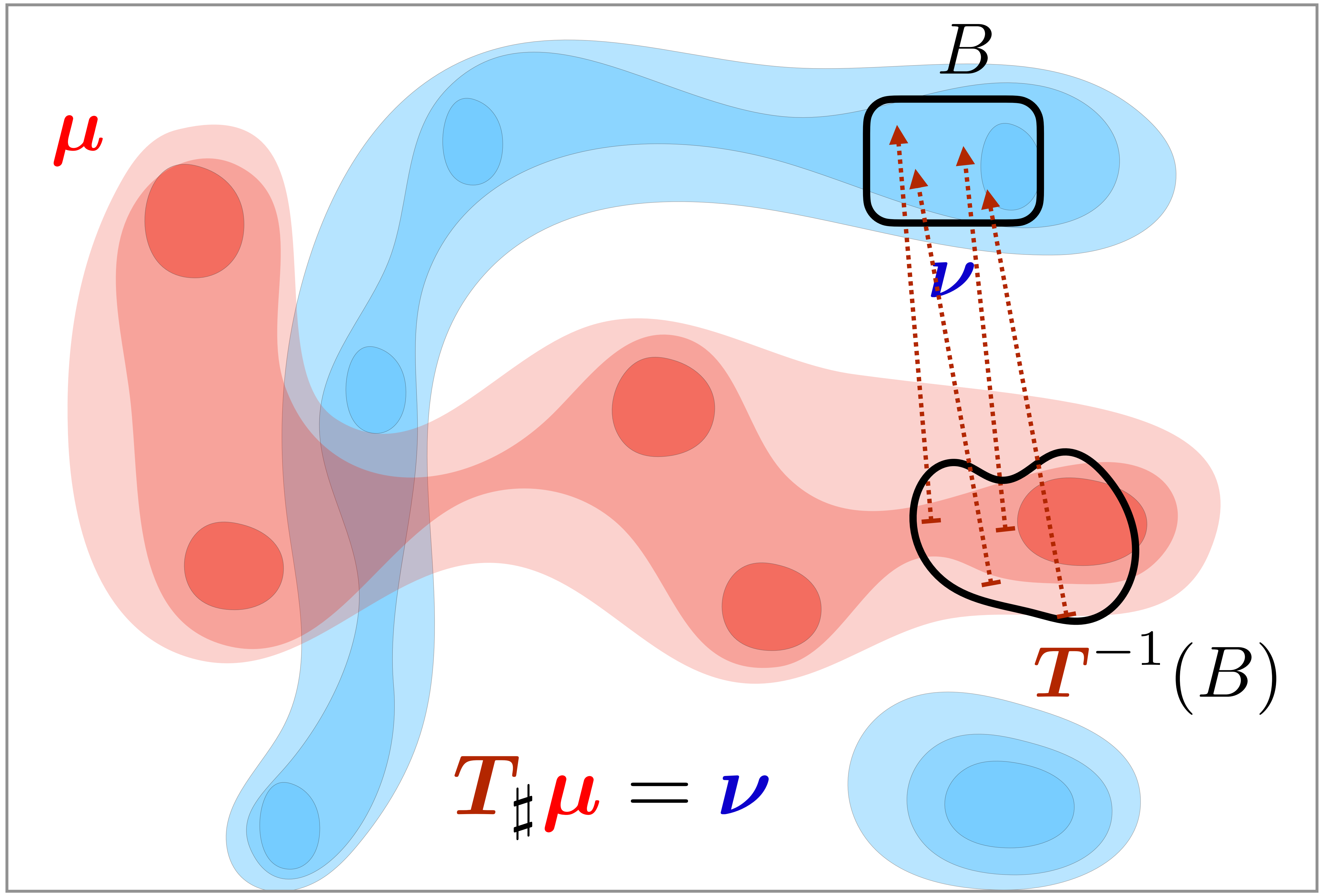
$$\forall B \subset \mathbb{R}^d, \nu(B) = \mu(T^{-1}(B))$$

Monge Maps



$$T_\sharp \mu = \nu$$
$$\Updownarrow$$
$$\forall B \subset \mathbb{R}^d, \nu(B) = \mu(T^{-1}(B))$$

Monge Maps

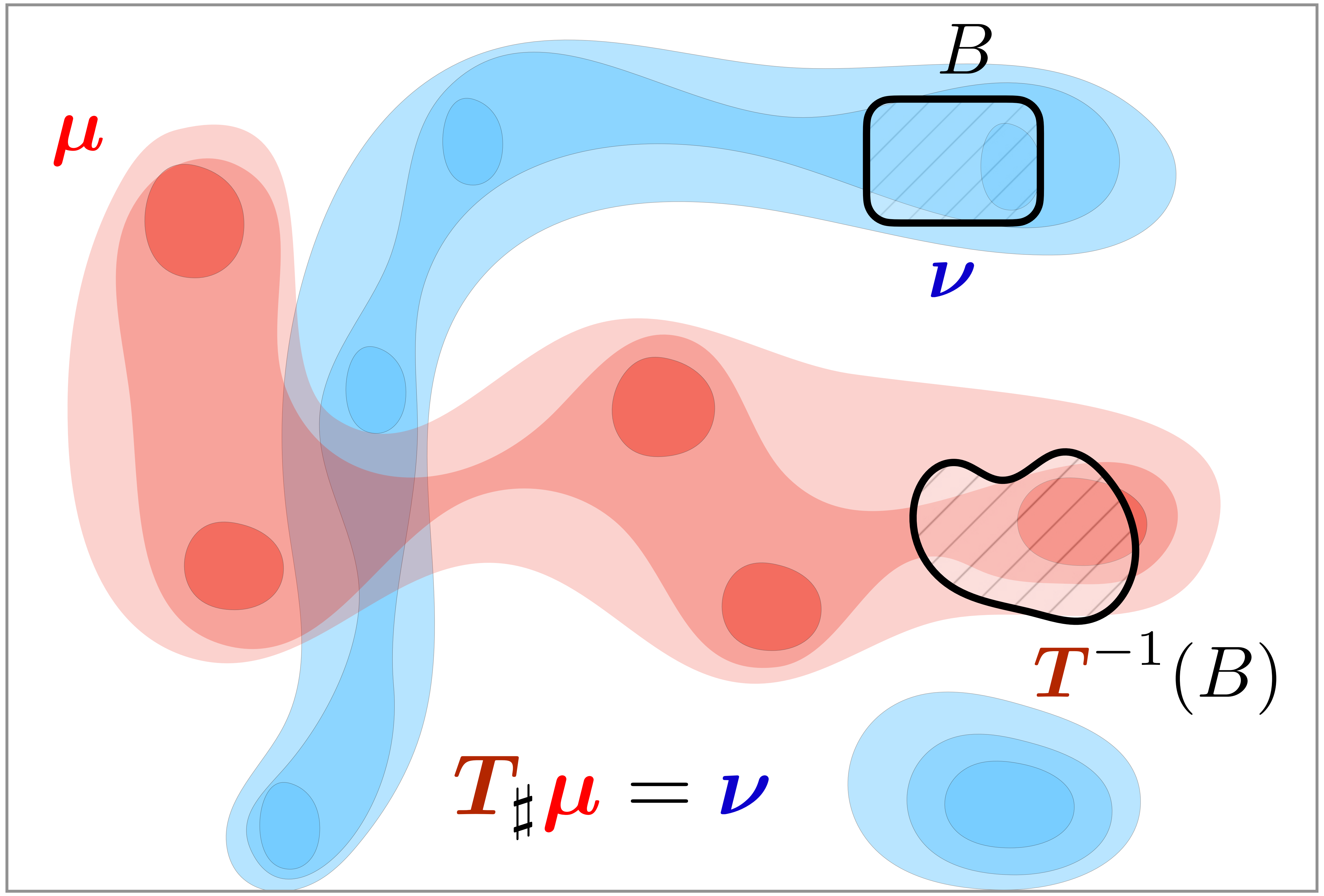


$$T_\sharp \mu = \nu$$

$$\Updownarrow$$

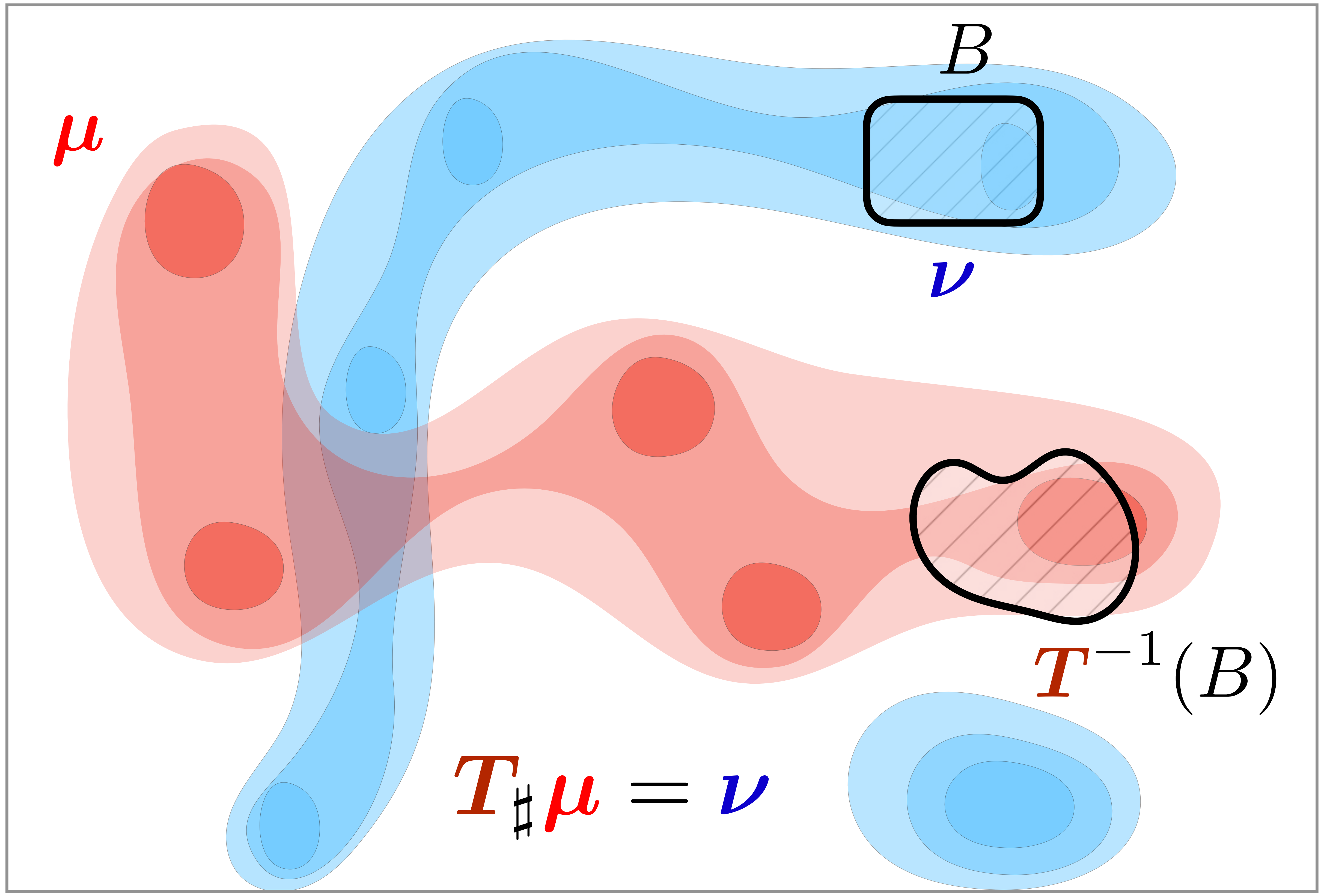
$$\forall B \subset \mathbb{R}^d, \nu(B) = \mu(T^{-1}(B))$$

Monge Maps



$$T_\sharp \mu = \nu$$
$$\Updownarrow$$
$$\forall B \subset \mathbb{R}^d, \nu(B) = \mu(T^{-1}(B))$$

Monge Maps



$$T_\sharp\mu = \nu$$

$$\Updownarrow$$

$$\forall B \subset \mathbb{R}^d, \nu(B) = \mu(T^{-1}(B))$$

$$T^\star = \underset{\substack{T:\mathbb{R}^p\to\mathbb{R}^p \\ T_\sharp\mu=\nu}}{\operatorname{argmin}} \int c(\mathbf{x}, T(\mathbf{x}))\mathrm{d}\mu(x)$$

From Couplings to Monge through Duality

$$\min_{\substack{\mathbf{P} \in \mathbb{R}_+^{n \times m} \\ \mathbf{P}\mathbf{1}_m = \mathbf{a}, \mathbf{P}^T\mathbf{1}_n = \mathbf{b}}} \langle \mathbf{P}, M_{\mathbf{X}\mathbf{Y}} \rangle$$

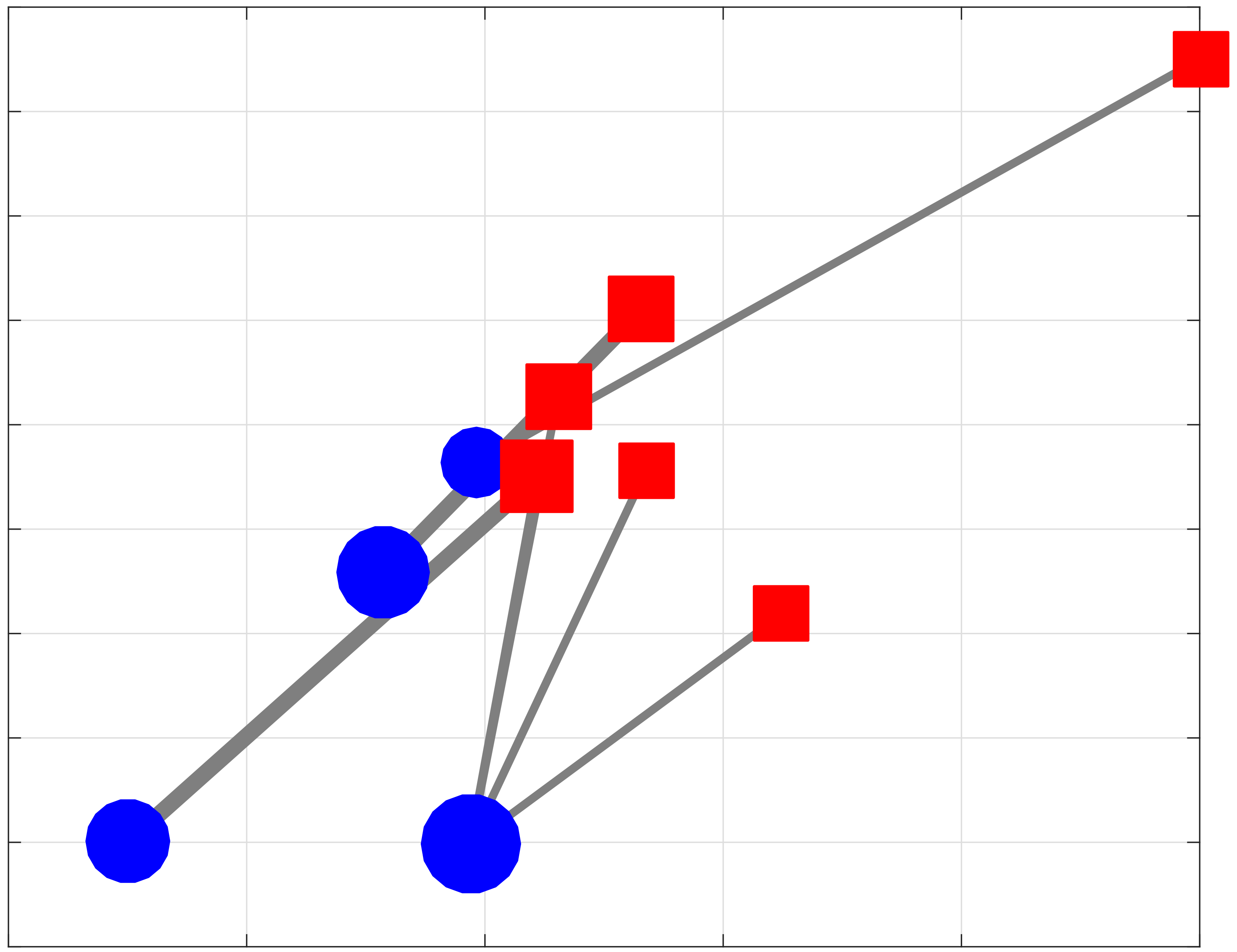
From Couplings to Monge through Duality

$$\min_{\substack{\boldsymbol{P}\in\mathbb{R}_+^{n\times m}\\ \boldsymbol{P}\mathbf{1}_m=\boldsymbol{a},\boldsymbol{P}^T\mathbf{1}_n=\boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle$$

$$\max_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{\boldsymbol{n}},\boldsymbol{\beta}\in\mathbb{R}^{\boldsymbol{m}}\\ \boldsymbol{\alpha}_{\boldsymbol{i}}+\boldsymbol{\beta}_{\boldsymbol{j}}\leq c(\boldsymbol{x}_{\boldsymbol{i}},\boldsymbol{y}_{\boldsymbol{j}})}} \boldsymbol{\alpha}^T\boldsymbol{a} + \boldsymbol{\beta}^T\boldsymbol{b}$$

From Couplings to Monge through Duality

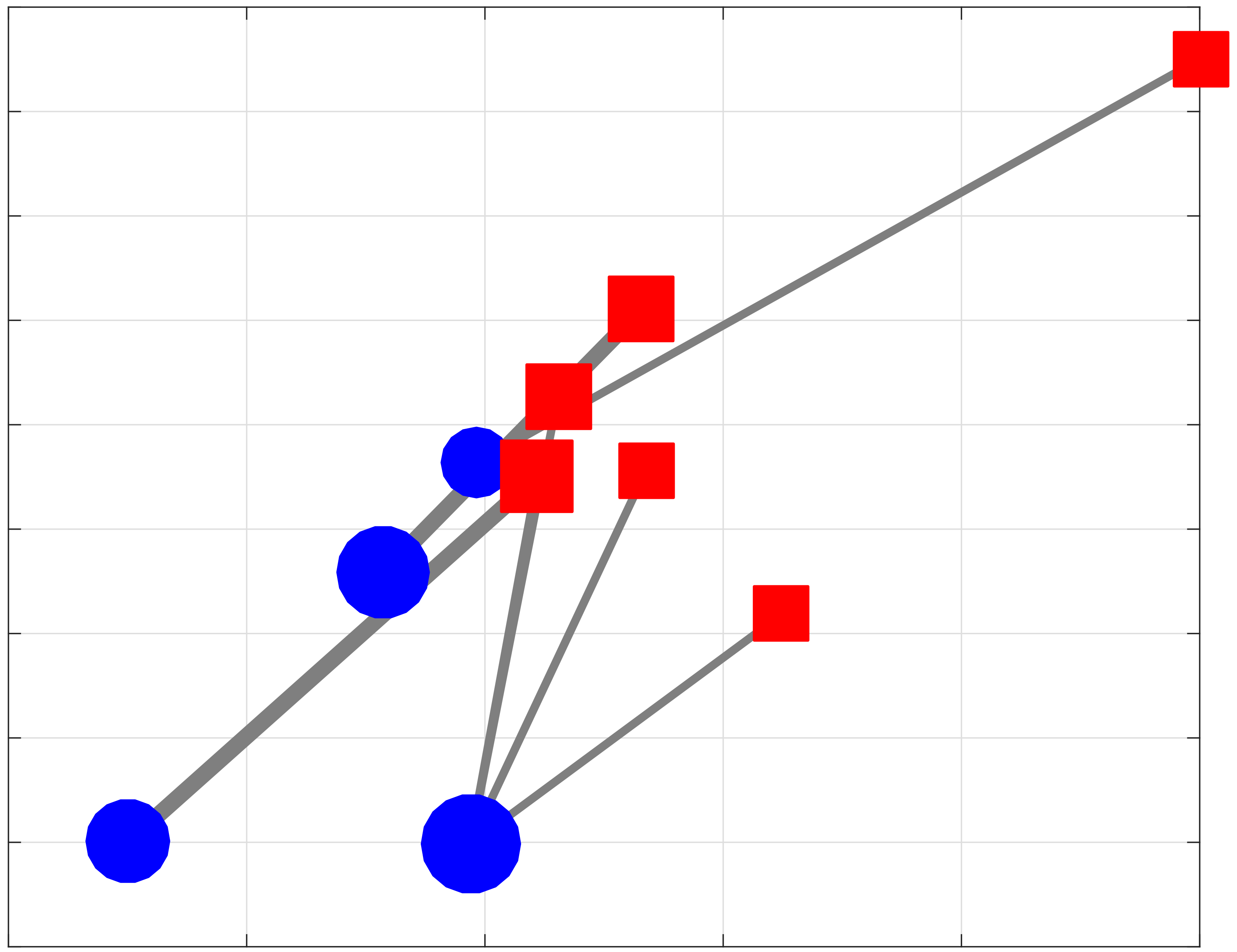
$$\min_{\substack{\boldsymbol{P}\in\mathbb{R}_+^{n\times m}\\ \boldsymbol{P}\mathbf{1}_m=\boldsymbol{a},\boldsymbol{P}^T\mathbf{1}_n=\boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{X}\boldsymbol{Y}} \rangle$$



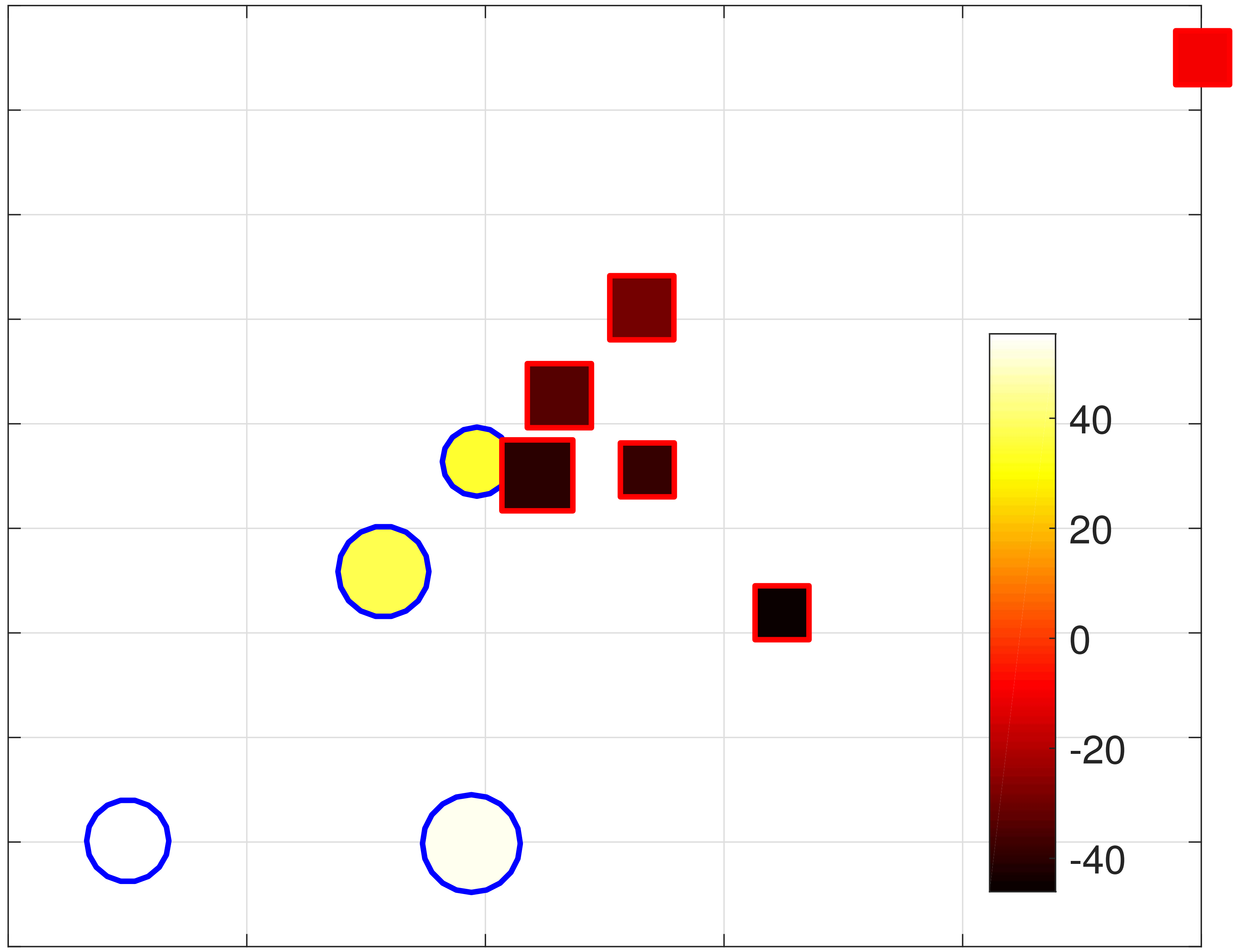
$$\max_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{\boldsymbol{n}},\boldsymbol{\beta}\in\mathbb{R}^{\boldsymbol{m}}\\ \boldsymbol{\alpha_i}+\boldsymbol{\beta_j}\leq c(\boldsymbol{x_i},\boldsymbol{y_j})}} \boldsymbol{\alpha}^T\boldsymbol{a}+\boldsymbol{\beta}^T\boldsymbol{b}$$

From Couplings to Monge through Duality

$$\min_{\substack{\mathbf{P}\in\mathbb{R}_+^{n\times m}\\ \mathbf{P}\mathbf{1}_m=\mathbf{a},\mathbf{P}^T\mathbf{1}_n=\mathbf{b}}} \langle \mathbf{P}, M_{\mathbf{X}\mathbf{Y}} \rangle$$



$$\max_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{\boldsymbol{n}},\boldsymbol{\beta}\in\mathbb{R}^{\boldsymbol{m}}\\ \boldsymbol{\alpha}_{\boldsymbol{i}}+\boldsymbol{\beta}_{\boldsymbol{j}}\leq c(\boldsymbol{x}_{\boldsymbol{i}},\boldsymbol{y}_{\boldsymbol{j}})}} \boldsymbol{\alpha}^T\mathbf{a}+\boldsymbol{\beta}^T\mathbf{b}$$



Roadmap to Solve Monge Problem

Monge

$$\boldsymbol{T}^\star = \underset{\substack{\boldsymbol{T}:\mathbb{R}^p \to \mathbb{R}^p \\ \boldsymbol{T}_\sharp \boldsymbol{\mu} = \boldsymbol{\nu}}}{\operatorname{argmin}} \int \boldsymbol{c}(\mathbf{x}, \boldsymbol{T}(\mathbf{x}))\mathrm{d}\boldsymbol{\mu}(\boldsymbol{x})$$

Kantorovich

$$\min_{\substack{\boldsymbol{P} \in \mathbb{R}_+^{n \times m} \\ \boldsymbol{P}\mathbf{1}_m = \boldsymbol{a}, \boldsymbol{P}^T \mathbf{1}_n = \boldsymbol{b}}} \langle \boldsymbol{P}, M_{\boldsymbol{XY}} \rangle$$

Kantorovich-Dual

$$\max_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^{\boldsymbol{n}}, \boldsymbol{\beta} \in \mathbb{R}^{\boldsymbol{m}} \\ \boldsymbol{\alpha}_{\boldsymbol{i}} + \boldsymbol{\beta}_{\boldsymbol{j}} \leq \boldsymbol{c}(\boldsymbol{x}_{\boldsymbol{i}}, \boldsymbol{y}_{\boldsymbol{j}})}} \boldsymbol{\alpha}^T \boldsymbol{a} + \boldsymbol{\beta}^T \boldsymbol{b}$$

Roadmap to Solve Monge Problem

Monge

$$T^\star = \underset{\substack{T:\mathbb{R}^p \to \mathbb{R}^p \\ T_\sharp \mu = \nu}}{\operatorname{argmin}} \int c(\mathbf{x}, T(\mathbf{x}))\mathrm{d}\mu(x)$$

Kantorovich

$$\min_{\substack{P \in \mathbb{R}_+^{n\times m} \\ P\mathbf{1}_m = a, P^T\mathbf{1}_n = b}} \langle P, M_{\mathbf{XY}} \rangle \longrightarrow \inf_{P \in \Pi(\mu,\nu)} \iint c(x,y)P(\mathrm{d}x, \mathrm{d}y).$$

Kantorovich-Dual

$$\max_{\substack{\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^m \\ \alpha_i + \beta_j \leq c(x_i, y_j)}} \alpha^T a + \beta^T b \longrightarrow \sup_{\substack{f,g:\mathbb{R}^d \to \mathbb{R} \\ f(x)+g(y) \leq c(x,y)}} \int_{\mathbb{R}^d} f\,\mathrm{d}\mu + \int_{\mathbb{R}^d} g\,\mathrm{d}\nu$$

Roadmap to Solve Monge Problem

Monge

$$T^\star = \underset{\substack{T:\mathbb{R}^p \to \mathbb{R}^p \\ T_\sharp \mu = \nu}}{\operatorname{argmin}} \int c(\mathbf{x}, T(\mathbf{x}))\mathrm{d}\mu(x)$$

Kantorovich

$$P^\star = \arg \inf_{P \in \Pi(\mu,\nu)} \iint c(x, y) P(\mathrm{d}x, \mathrm{d}y).$$

Kantorovich-Dual

$$f^\star, g^\star = \arg \sup_{\substack{f, g : \mathbb{R}^d \to \mathbb{R} \\ f(x)+g(y) \leq c(x,y)}} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu + \int_{\mathbb{R}^d} g \,\mathrm{d}\nu\,.$$

Duality: c-transforms

$$\sup_{\substack{f, g : \mathbb{R}^d \to \mathbb{R} \\ f(x) + g(y) \leq c(x, y)}} \int_{\mathbb{R}^d} f \, \mathrm{d}\mu + \int_{\mathbb{R}^d} g \, \mathrm{d}\nu \, .$$

Suppose we fix function f, can we propose a "good" function for g?

Duality: c-transforms

$$\sup_{\substack{f,g:\mathbb{R}^d\to\mathbb{R}\\ f(x)+g(y)\leq c(x,y)}} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu + \int_{\mathbb{R}^d} g \,\mathrm{d}\nu \,.$$

Suppose we fix function f, can we propose a "good" function for g?

$$f(x) + g(y) \leq c(x, y)$$

Duality: c-transforms

$$\sup_{\substack{f, g : \mathbb{R}^d \to \mathbb{R} \\ f(x) + g(y) \leq c(x, y)}} \int_{\mathbb{R}^d} f \, d\mu + \int_{\mathbb{R}^d} g \, d\nu \, .$$

Suppose we fix function f, can we propose a “good” function for g?

$$f(x) + g(y) \leq c(x, y)$$

$$g(y) \leq c(x, y) - f(x)$$

Duality: c-transforms

$$\sup_{\substack{f, g : \mathbb{R}^d \to \mathbb{R} \\ f(x) + g(y) \le c(x,y)}} \int_{\mathbb{R}^d} f \, \mathrm{d}\mu + \int_{\mathbb{R}^d} g \, \mathrm{d}\nu \,.$$

Suppose we fix function f, can we propose a “good” function for g?

$$f(x) + g(y) \le c(x, y)$$

$$g(y) \le c(x, y) - f(x)$$

$$g(y) \le \inf_{x} c(x, y) - f(x)$$

Duality: c-transforms

$$\sup_{\substack{f,g:\mathbb{R}^d\to\mathbb{R}\\ f(x)+g(y)\le c(x,y)}} \int_{\mathbb{R}^d} f \,\mathrm{d}\mu + \int_{\mathbb{R}^d} g \,\mathrm{d}\nu \,.$$

Suppose we fix function f, can we propose a "good" function for g?

$$f(x) + g(y) \le c(x,y)$$

$$g(y) \le c(x,y) - f(x)$$

$$g(y) \le \inf_x c(x,y) - f(x)$$

c-transform: $\overline{f}(y) \stackrel{\text{def}}{=} \inf_x c(x,y) - f(x).$

Duality: c-transforms

$$\sup_{\substack{f,g:\mathbb{R}^d\to\mathbb{R}\\ f(x)+g(y)\le c(x,y)}} \int_{\mathbb{R}^d} f\,\mathrm{d}\mu + \int_{\mathbb{R}^d} g\,\mathrm{d}\nu\,.$$

Suppose we fix function f, can we propose a "good" function for g?

$$f(x) + g(y) \le c(x,y)$$

$$g(y) \le c(x,y) - f(x)$$

$$g(y) \le \inf_x c(x,y) - f(x)$$

c-transform: $$\overline{f}(y) \stackrel{\text{def}}{=} \inf_x c(x,y) - f(x).$$

Given a potential, one can get automatically the "best" other one using c-transforms.

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

definition

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily** **c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

definition

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \overset{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \overset{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

definition

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$
$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$
$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

definition

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \overset{\text{def}}{=} \inf_x c(x, y) - f(x).$$

$$\overline{g}(x) \overset{\text{def}}{=} \inf_y c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily** **c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow \overline{f^\star}(y_0) = c(x_0, y_0) - f^\star(x_0).$$

definition

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow \overline{f^\star}(y_0) = c(x_0, y_0) - f^\star(x_0).$$

$$\text{yet, } \overline{f^\star}(y_0) = \inf_z c(z, y_0) - f^\star(z).$$ definition

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a c-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily** c-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow \overline{f^\star}(y_0) = c(x_0, y_0) - f^\star(x_0).$$

$$\text{yet, } \overline{f^\star}(y_0) = \inf_z c(z, y_0) - f^\star(z).$$ definition

$$\Rightarrow x_0 \text{ is minimizer, } \nabla_1 c(x_0, y_0) = \nabla f^\star(x_0)$$

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily** **c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow \overline{f^\star}(y_0) = c(x_0, y_0) - f^\star(x_0).$$

$$\text{yet, } \overline{f^\star}(y_0) = \inf_z c(z, y_0) - f^\star(z).$$ definition

$$\Rightarrow x_0 \text{ is minimizer, } \nabla_1 c(x_0, y_0) = \nabla f^\star(x_0)$$

$$\text{therefore } y_0 = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \overset{\text{def}}{=} \inf_{x} c(x,y) - f(x).$$

$$\overline{g}(x) \overset{\text{def}}{=} \inf_{y} c(x,y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$T^\star(x) = \nabla_1 c(x,\cdot)^{-1} \circ \nabla f^\star(x)$$

Primal-dual relationship (KKT)

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

$$\Leftrightarrow \overline{f^\star}(y_0) = c(x_0, y_0) - f^\star(x_0).$$

$$\text{yet, } \overline{f^\star}(y_0) = \inf_z c(z, y_0) - f^\star(z).$$ definition

$$\Rightarrow x_0 \text{ is minimizer, } \nabla_1 c(x_0, y_0) = \nabla f^\star(x_0)$$

$$\text{therefore } y_0 = \nabla_1 c(x,\cdot)^{-1} \circ \nabla f^\star(x)$$

Gangbo-McCann Theorem

c-transform:

$$\overline{f}(y) \overset{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \overset{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily** **c**-concave

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

$$P^\star(x_0, y_0) > 0 \Leftrightarrow f^\star(x_0) + g^\star(y_0) = c(x_0, y_0).$$

Primal-dual relationship (KKT)

$$\Leftrightarrow f^\star(x_0) + \overline{f^\star}(y_0) = c(x_0, y_0).$$

c-transforms

$$\Leftrightarrow \overline{f^\star}(y_0) = c(x_0, y_0) - f^\star(x_0).$$

$$\text{yet, } \overline{f^\star}(y_0) = \inf_z c(z, y_0) - f^\star(z).$$

definition

$$\Rightarrow x_0 \text{ is minimizer, } \nabla_1 c(x_0, y_0) = \nabla f^\star(x_0)$$

$$\text{therefore } y_0 = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

Translation invariant costs generated by a convex functional

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

Translation invariant costs generated by a convex functional

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x,y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x,y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$c(x,y) = h(x-y), h : \mathbb{R}^d \to \mathbb{R}, \text{convex}$$

Translation invariant costs generated by a convex functional

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$c(x, y) = h(x - y), h : \mathbb{R}^d \to \mathbb{R}, \text{convex}$$

$$\nabla_1 c(x, y) = \nabla h(x - y)$$

Translation invariant costs generated by a convex functional

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$c(x, y) = h(x - y), h : \mathbb{R}^d \to \mathbb{R}, \text{convex}$$

$$\nabla_1 c(x, y) = \nabla h(x - y)$$

$$\nabla_1 c(x, \cdot) = \nabla h(x - \cdot)$$

Translation invariant costs generated by a convex functional

c-transform:

$$\overline{f}(y) \overset{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \overset{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^{\star}$ is **necessarily c**-concave

$$c(x, y) = h(x - y), h : \mathbb{R}^d \to \mathbb{R}, \text{convex}$$

$$\nabla_1 c(x, y) = \nabla h(x - y)$$

$$\nabla_1 c(x, \cdot) = \nabla h(x - \cdot)$$

$$\nabla_1 c(x, \cdot)^{-1} = x - (\nabla h)^{-1}(\cdot)$$

Translation invariant costs generated by a convex functional

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

$$c(x, y) = h(x - y), h : \mathbb{R}^d \to \mathbb{R}, \text{convex}$$

$$\nabla_1 c(x, y) = \nabla h(x - y)$$

$$\nabla_1 c(x, \cdot) = \nabla h(x - \cdot)$$

$$\nabla_1 c(x, \cdot)^{-1} = x - (\nabla h)^{-1}(\cdot)$$

$$\nabla_1 c(x, \cdot)^{-1} = x - \nabla h^*(\cdot)$$

$$h^*(z) = \sup_x \langle x, z \rangle - h(x)$$

Gangbo-McCann and Brenier Theorems

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^{\star}$ is **necessarily c**-concave

General cost function c $\quad T^{\star}(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^{\star}(x)$

Translation invariant cost h $\quad T^{\star}(x) = x - \nabla h^{*} \circ \nabla f^{\star}(x)$

Squared-Euclidean Cost $\quad T^{\star}(x) = x - \nabla f^{\star}(x) = \nabla(\frac{1}{2}\|\cdot\|_2^2 - f^{\star})(x)$

$h = \frac{1}{2}\|\cdot\|_2^2 \rightarrow \nabla h = \nabla h^{*} = \mathrm{Id}$

Gangbo-McCann and Brenier Theorems

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily** **c**-concave

General cost function c

$$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

Translation invariant cost h

$$T^\star(x) = x - \nabla h^* \circ \nabla f^\star(x)$$

Squared-Euclidean Cost

$$h = \tfrac{1}{2}\|\cdot\|_2^2 \rightarrow \nabla h = \nabla h^* = \mathrm{Id}$$

$$T^\star(x) = x - \nabla f^\star(x) = \nabla \underbrace{\left(\tfrac{1}{2}\|\cdot\|_2^2 - f^\star\right)}_{\text{can prove this is convex}}(x)$$

Gangbo-McCann and Brenier Theorems

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$
$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

General cost function c $\quad T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$

Translation invariant cost h $\quad T^\star(x) = x - \nabla h^* \circ \nabla f^\star(x)$

Squared-Euclidean Cost $\quad T^\star(x) = \nabla \varphi^\star(x)$

$h = \frac{1}{2}\|\cdot\|_2^2 \to \nabla h = \nabla h^* = \mathrm{Id}$

Gangbo-McCann and Brenier Theorems

c-transform:

$$\overline{f}(y) \stackrel{\text{def}}{=} \inf_{x} c(x, y) - f(x).$$

$$\overline{g}(x) \stackrel{\text{def}}{=} \inf_{y} c(x, y) - g(y).$$

f is a **c**-concave function
if exists g such that $f = \overline{g}$.

$f^\star$ is **necessarily c**-concave

General cost function c $$T^\star(x) = \nabla_1 c(x, \cdot)^{-1} \circ \nabla f^\star(x)$$

Translation invariant cost h $$T^\star(x) = x - \nabla h^* \circ \nabla f^\star(x)$$

Squared-Euclidean Cost $$T^\star(x) = \nabla \varphi^\star(x)$$

$h = \frac{1}{2}\|\cdot\|_2^2 \to \nabla h = \nabla h^* = \mathrm{Id}$

$$\varphi^\star = \underset{\varphi \text{ cvx}}{\operatorname{argmin}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

First Attempt at Computing OT Maps, using Sinkhorn

$$T^\star(x) = x - \nabla h^* \circ \nabla f^\star(x)$$

- Use two samples of points from each measure + regularization parameter,

$$\mathbf{x}_1, \dots, \mathbf{x}_n, \mathbf{y}_1, \dots, \mathbf{y}_m, \gamma$$

- Run Sinkhorn algorithm to get $f_i^\star, g_j^\star$ point-wise dual values (these correspond to the $\alpha_i^\star, \beta_j^\star$ values in log-space returned by Sinkhorn).

- Extend these values *out-of-sample*, with a "soft" h-transform, to create the estimator:

$$\hat{f}_n(\mathbf{x}) = -\gamma \log \sum_j e^{\frac{-h(\mathbf{x}-\mathbf{y}_j)+g_j^\star}{\gamma}}$$

[Pooladian+21]

Monge / Bregman Entropic Maps

$$T^\star(\boldsymbol{x}) = \boldsymbol{x} - \nabla \boldsymbol{h}^* \circ \nabla \boldsymbol{f}^\star(\boldsymbol{x})$$

$$\nabla \hat{\boldsymbol{f}}_{\boldsymbol{n}}(\mathbf{x}) = \sum_j p_j(\mathbf{x}) \nabla \boldsymbol{h}(\mathbf{x} - \mathbf{y}_{\boldsymbol{j}})$$

$$p_j(\mathbf{x}) := \frac{e^{(-\boldsymbol{h}(\mathbf{x}-\mathbf{y}_{\boldsymbol{j}})+\boldsymbol{g}_{\boldsymbol{j}}^\star)/\gamma}}{\sum_k e^{(-\boldsymbol{h}(\mathbf{x}-\mathbf{y}_{\boldsymbol{k}})+\boldsymbol{g}_{\boldsymbol{k}}^\star)/\gamma}}$$

Monge / Bregman Entropic Maps

$$T^\star(\boldsymbol{x}) = \boldsymbol{x} - \nabla \boldsymbol{h}^* \circ \nabla \boldsymbol{f}^\star(\boldsymbol{x})$$

$$\nabla \hat{\boldsymbol{f}}_{\boldsymbol{n}}(\mathbf{x}) = \textstyle\sum_j p_j(\mathbf{x}) \nabla \boldsymbol{h}(\mathbf{x} - \mathbf{y}_{\boldsymbol{j}})$$

$$p_j(\mathbf{x}) := \frac{e^{(-\boldsymbol{h}(\mathbf{x}-\mathbf{y}_{\boldsymbol{j}})+\boldsymbol{g}_{\boldsymbol{j}}^\star)/\gamma}}{\sum_k e^{(-\boldsymbol{h}(\mathbf{x}-\mathbf{y}_{\boldsymbol{k}})+\boldsymbol{g}_{\boldsymbol{k}}^\star)/\gamma}}$$

$$D_{\boldsymbol{h}}(\mathbf{z}|\mathbf{w}) = \boldsymbol{h}(\mathbf{z}) - \boldsymbol{h}(\mathbf{w}) - \langle \nabla \boldsymbol{h}(\mathbf{w}), \mathbf{z} - \mathbf{w} \rangle$$

Bregman divergence

$$\nabla \boldsymbol{h}^* \left(\textstyle\sum_i \lambda_i \nabla \boldsymbol{h}(\mathbf{z}_i)\right) = \operatorname*{argmin}_{\mathbf{z}} \sum_i \lambda_i D_{\boldsymbol{h}}(\mathbf{z}|\mathbf{z_i})$$

Bregman barycenter

Monge / Bregman Entropic Maps

$$T^\star(\boldsymbol{x}) = \boldsymbol{x} - \nabla \boldsymbol{h}^* \circ \nabla \boldsymbol{f}^\star(\boldsymbol{x})$$

$$\nabla \hat{\boldsymbol{f}}_{\boldsymbol{n}}(\mathbf{x}) = \sum_j p_j(\mathbf{x}) \nabla \boldsymbol{h}(\mathbf{x} - \mathbf{y}_{\boldsymbol{j}})$$

$$p_j(\mathbf{x}) := \frac{e^{(-\boldsymbol{h}(\mathbf{x}-\mathbf{y}_{\boldsymbol{j}})+\boldsymbol{g}_{\boldsymbol{j}}^\star)/\gamma}}{\sum_k e^{(-\boldsymbol{h}(\mathbf{x}-\mathbf{y}_{\boldsymbol{k}})+\boldsymbol{g}_{\boldsymbol{k}}^\star)/\gamma}}$$

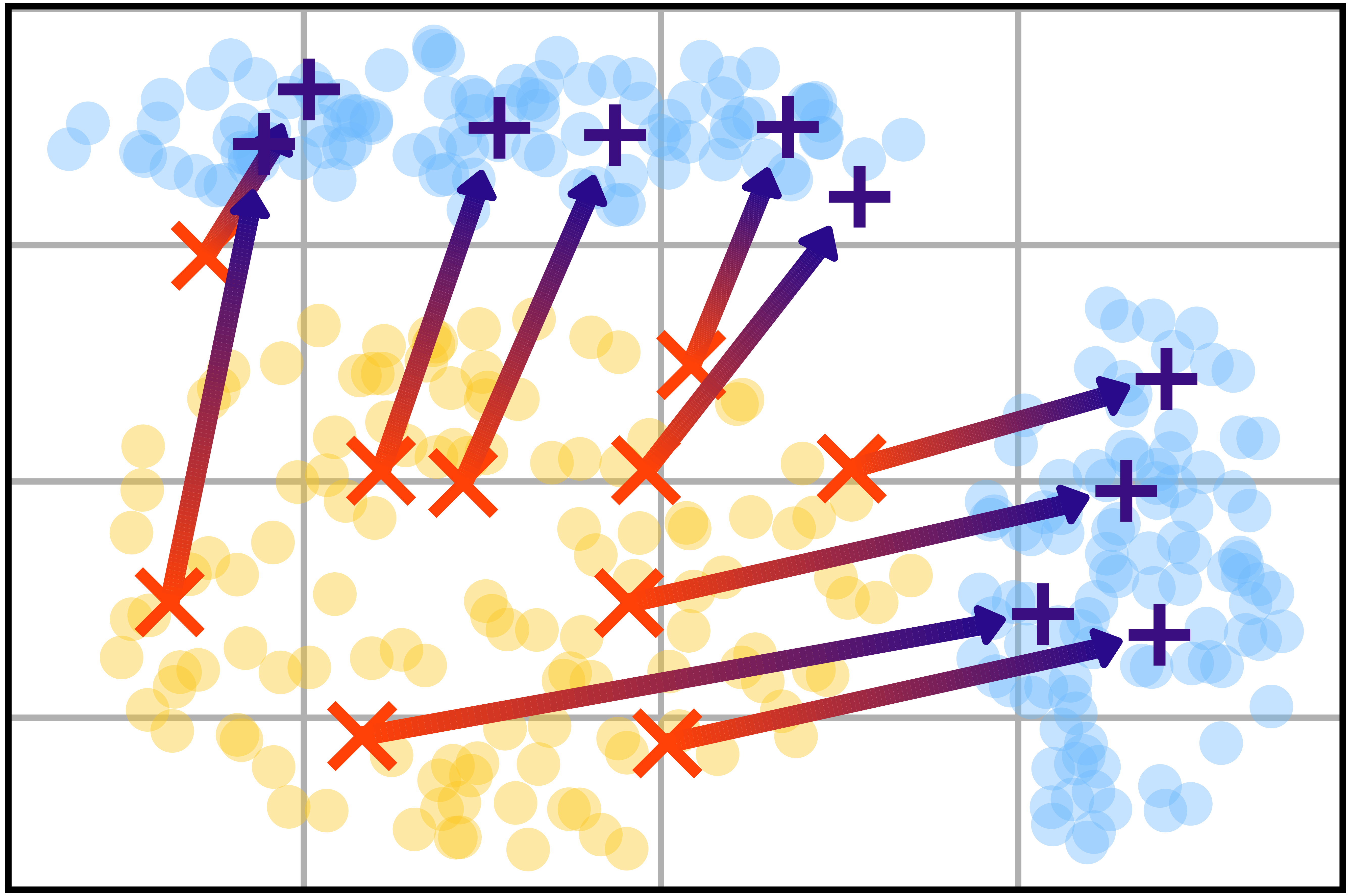
$$D_{\boldsymbol{h}}(\mathbf{z}|\mathbf{w}) = \boldsymbol{h}(\mathbf{z}) - \boldsymbol{h}(\mathbf{w}) - \langle \nabla \boldsymbol{h}(\mathbf{w}), \mathbf{z} - \mathbf{w} \rangle$$ Bregman divergence

$$\nabla \boldsymbol{h}^* \left(\sum_i \lambda_i \nabla \boldsymbol{h}(\mathbf{z}_i)\right) = \operatorname*{argmin}_{\mathbf{z}} \sum_i \lambda_i D_{\boldsymbol{h}}(\mathbf{z}|\mathbf{z_i})$$ Bregman barycenter

Displacement $\nabla \boldsymbol{h}^* \circ \nabla \hat{\boldsymbol{f}}_n(\mathbf{x})$ is a Bregman barycenter

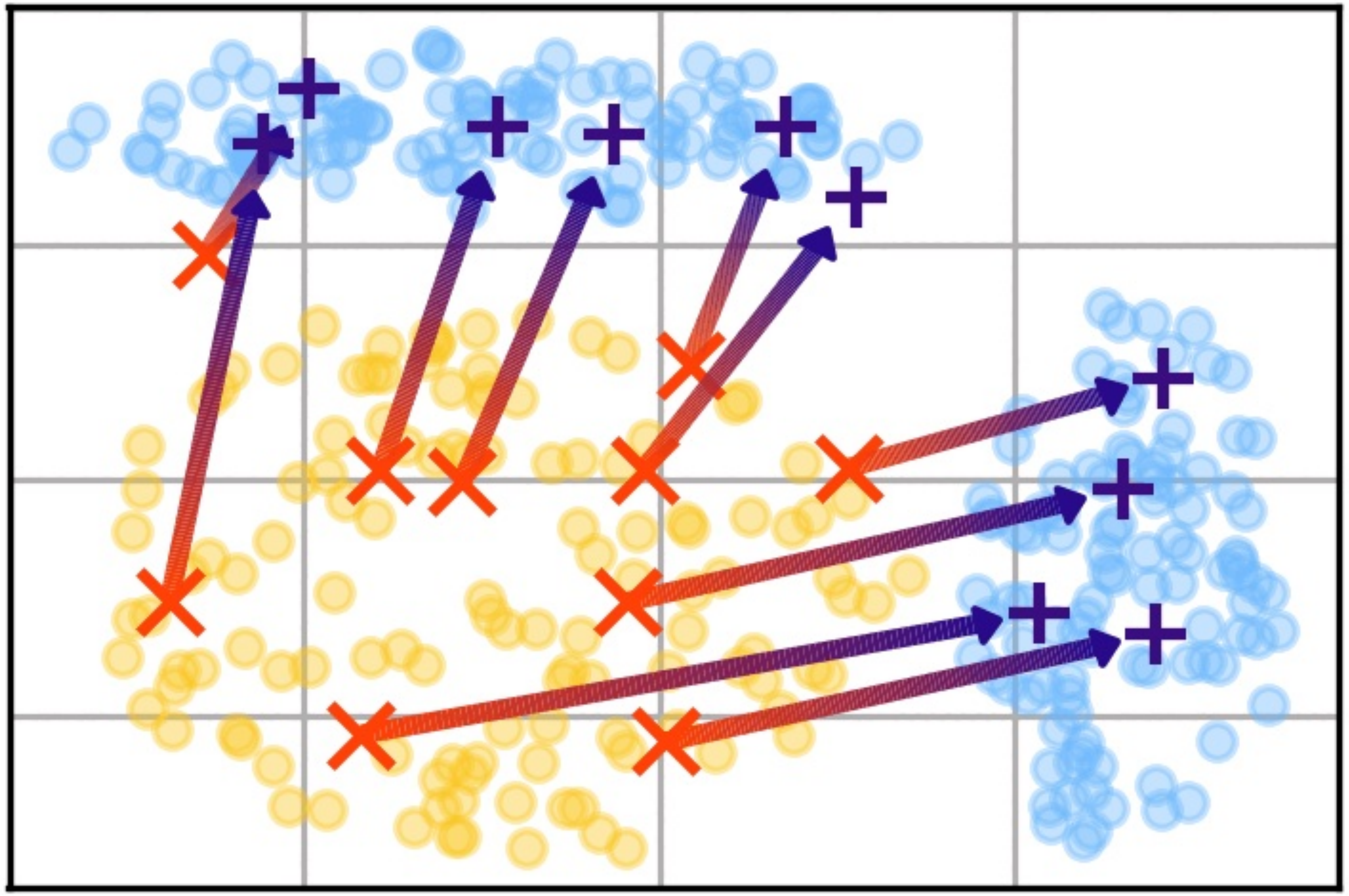
Monge / Bregman Entropic Maps

$$h(\mathbf{z}) = \frac{1}{2}\|\mathbf{z}\|_2^2$$

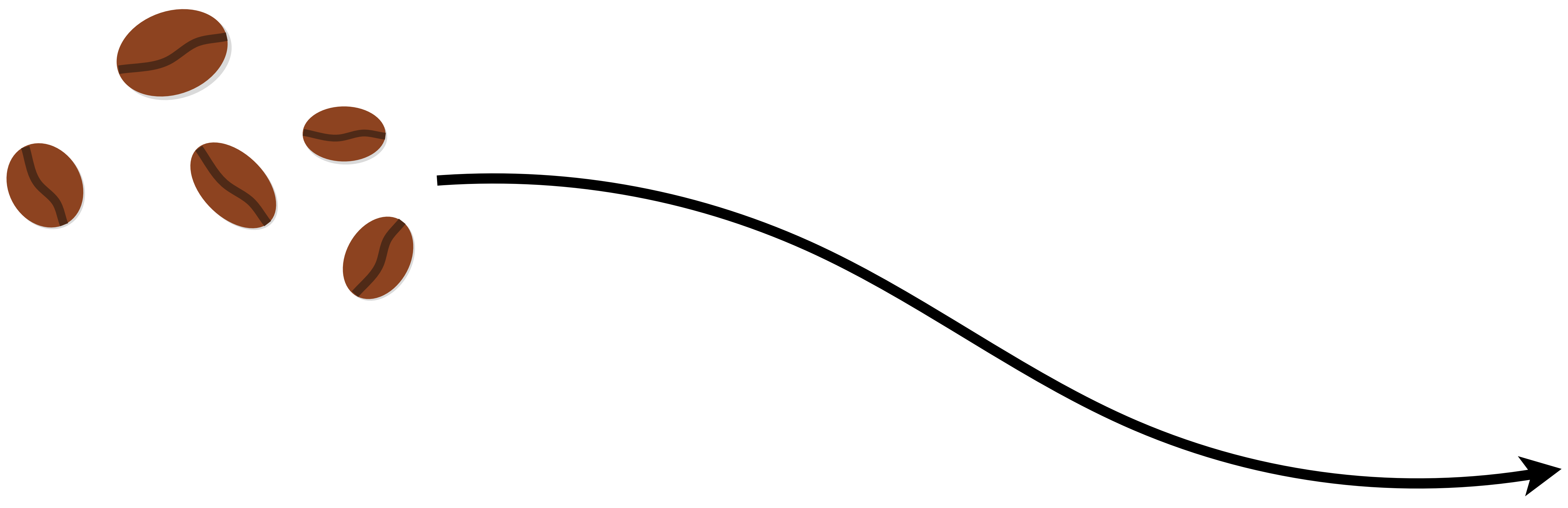


Monge / Bregman Entropic Maps

$$h(\mathbf{z}) = \frac{1}{2}\|\mathbf{z}\|_2^2 + 0\|\mathbf{z}\|_{\text{ov}k}^2$$



Coffee Break



… we will continue with **neural optimal transport**, and **dynamic** and **stochastic control formulations**!

Outline of the Tutorial

Prelude Warm-Up: Starting with Optimal Matchings

Part 1 Kantorovich Formulation of OT and Computations

Part 2 **Duality, Monge Formulations and Brenier Theorems**

- Kantorovich duality
- c-concavity, Gangbo-McCann and Brenier theorems
- Entropic map estimators, learning neural Monge maps

Part 3 Modeling Measure Dynamics with Optimal Transport

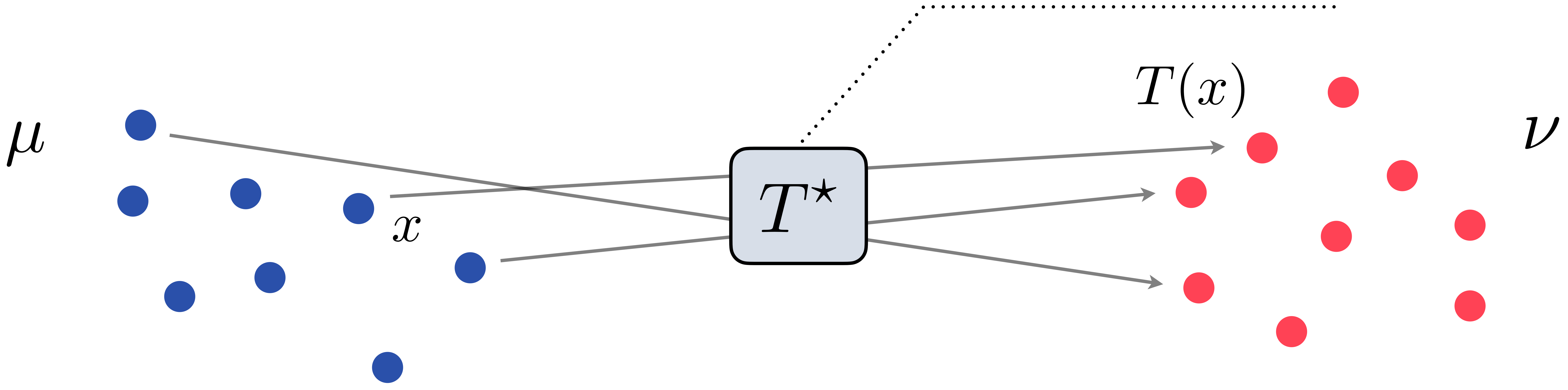
Predicting Out-of-Sample

So far: optimal transport between given point clouds

Predicting Out-of-Sample

So far: optimal transport between given point clouds

$$T^\star = \underset{T : T_\sharp \mu \approx \nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x)$$

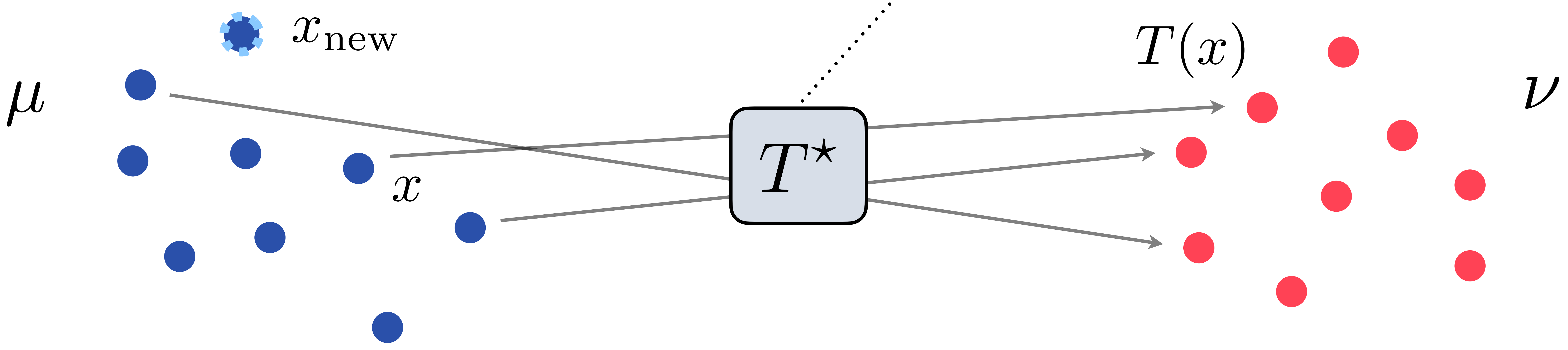


Predicting Out-of-Sample

So far: optimal transport between given point clouds

What if we want to make **out-of-sample predictions**?

$$T^\star = \operatorname*{argmin}_{T:T_\sharp\mu\approx\nu} \int_{\mathbb{R}^d} c(x, T(x))\,d\mu(x)$$

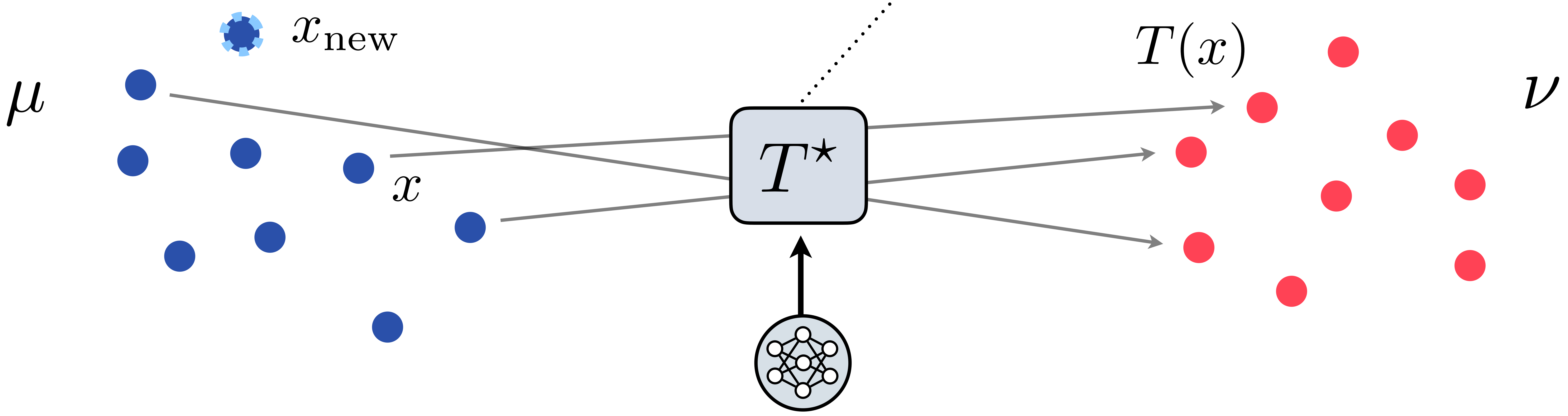


Predicting Out-of-Sample

So far: optimal transport between given point clouds

What if we want to make **out-of-sample predictions**?

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} c(x, T(x))d\mu(x)$$



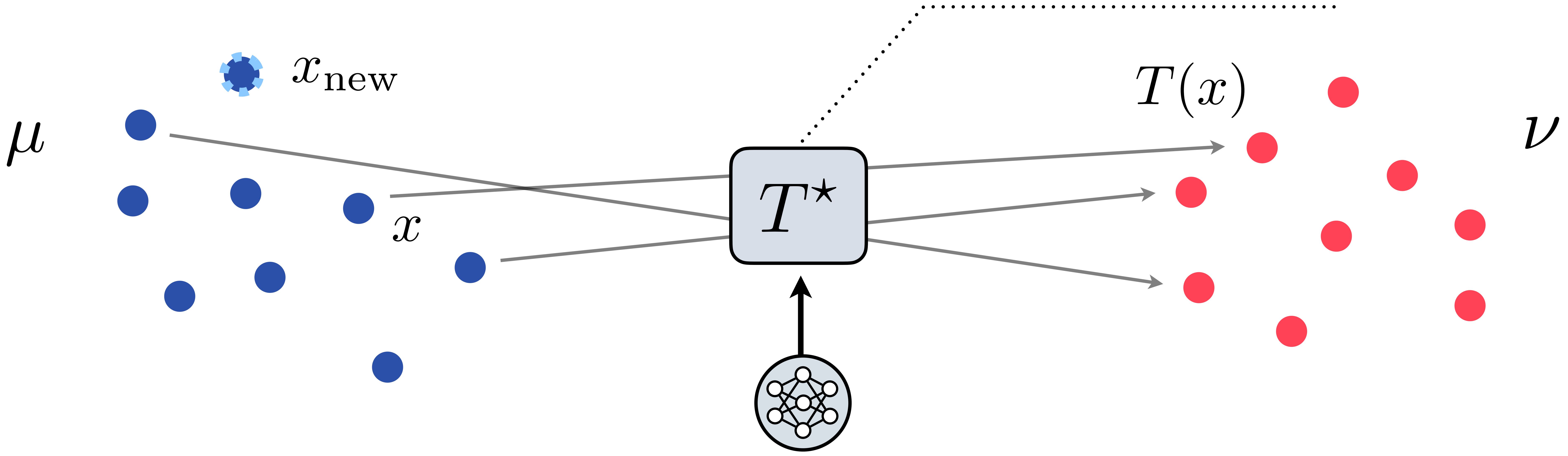
parameterize with neural network

Predicting Out-of-Sample

So far: optimal transport between given point clouds

What if we want to make **out-of-sample predictions**?

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} c(x, T(x))d\mu(x)$$

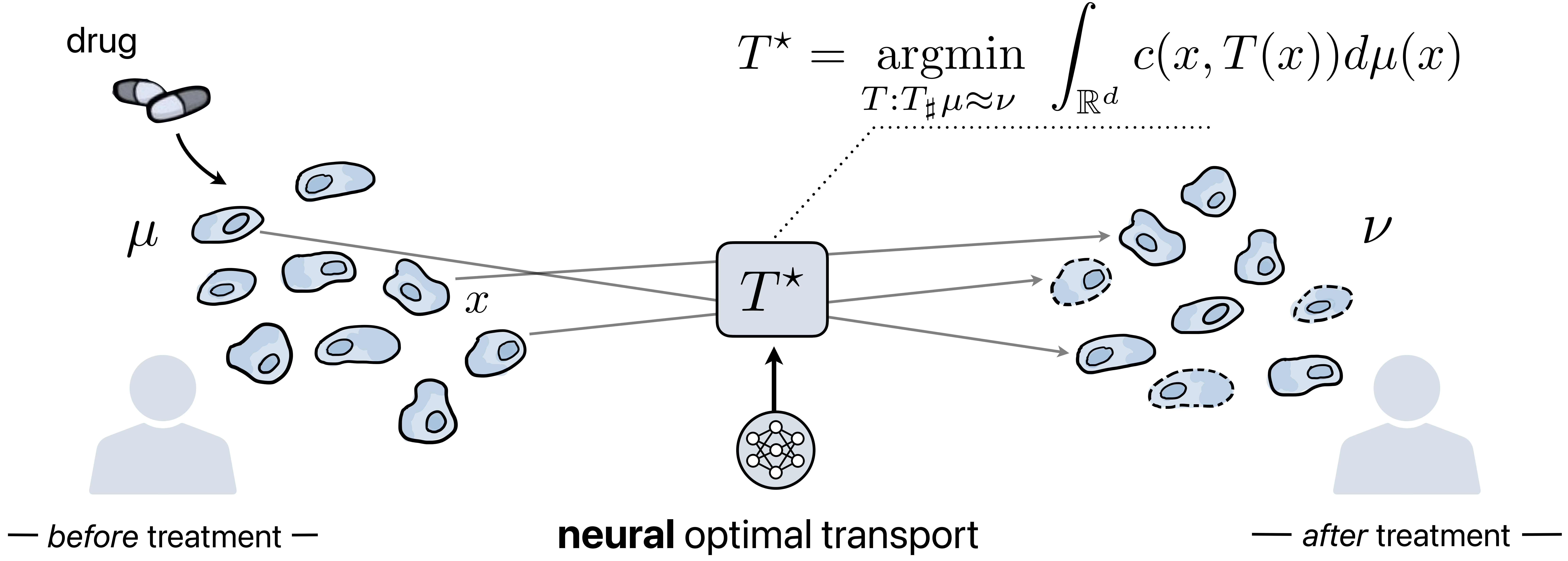


neural optimal transport

Motivation: **Predicting Patient Treatment Responses**

So far: optimal transport between given point clouds

What if we want to make **out-of-sample predictions**?



Motivation: **Predicting Patient Treatment Responses**

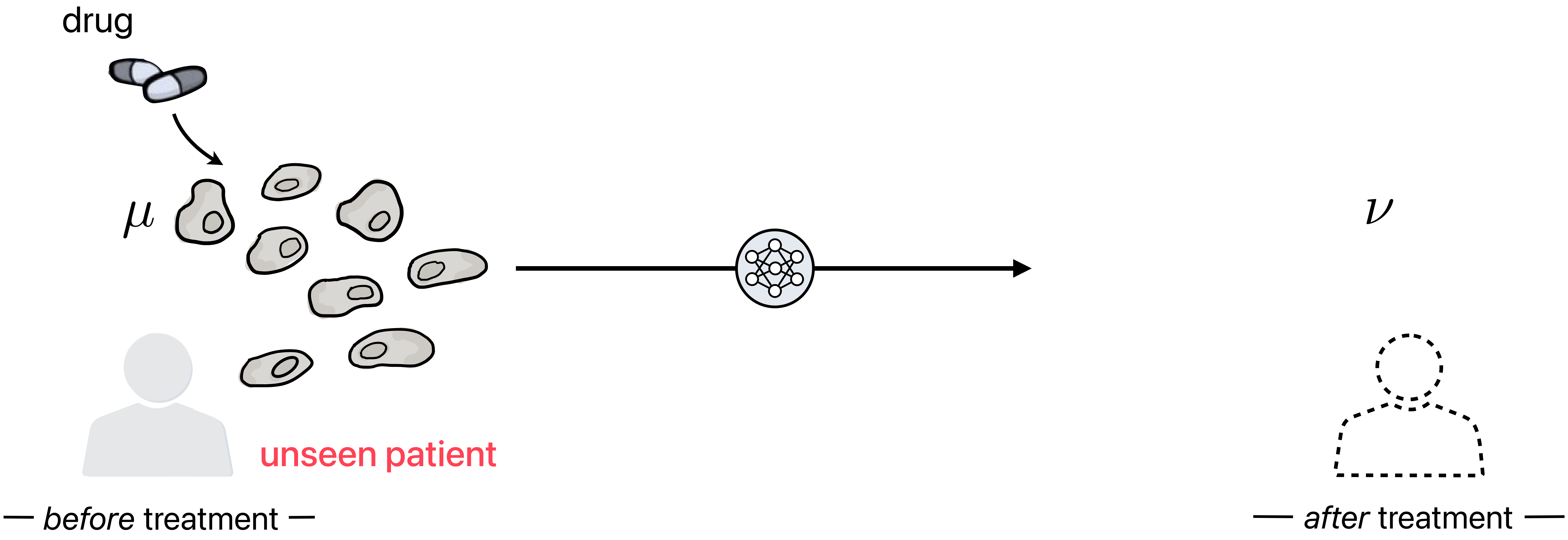
So far: optimal transport between given point clouds

What if we want to make **out-of-sample predictions**?

Motivation: **Predicting Patient Treatment Responses**

So far: optimal transport between given point clouds

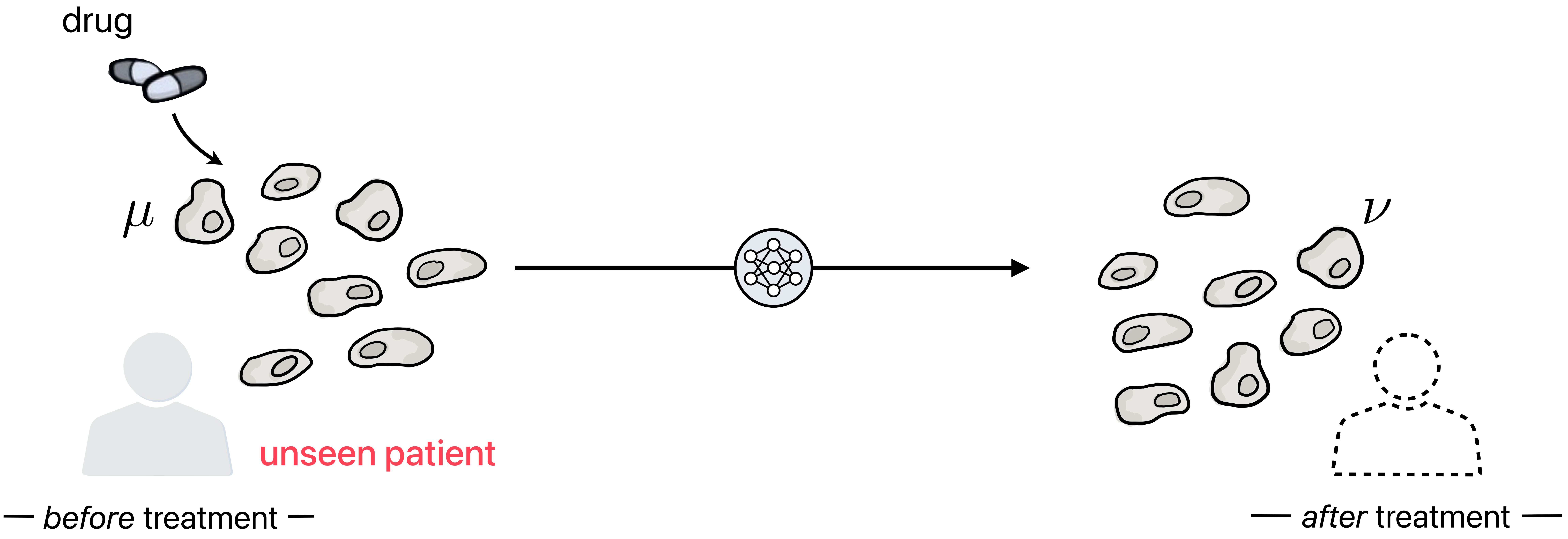
What if we want to make **out-of-sample predictions**?



Motivation: **Predicting Patient Treatment Responses**

So far: optimal transport between given point clouds

What if we want to make **out-of-sample predictions**?



Other Prominent Use Case: Generative Modeling

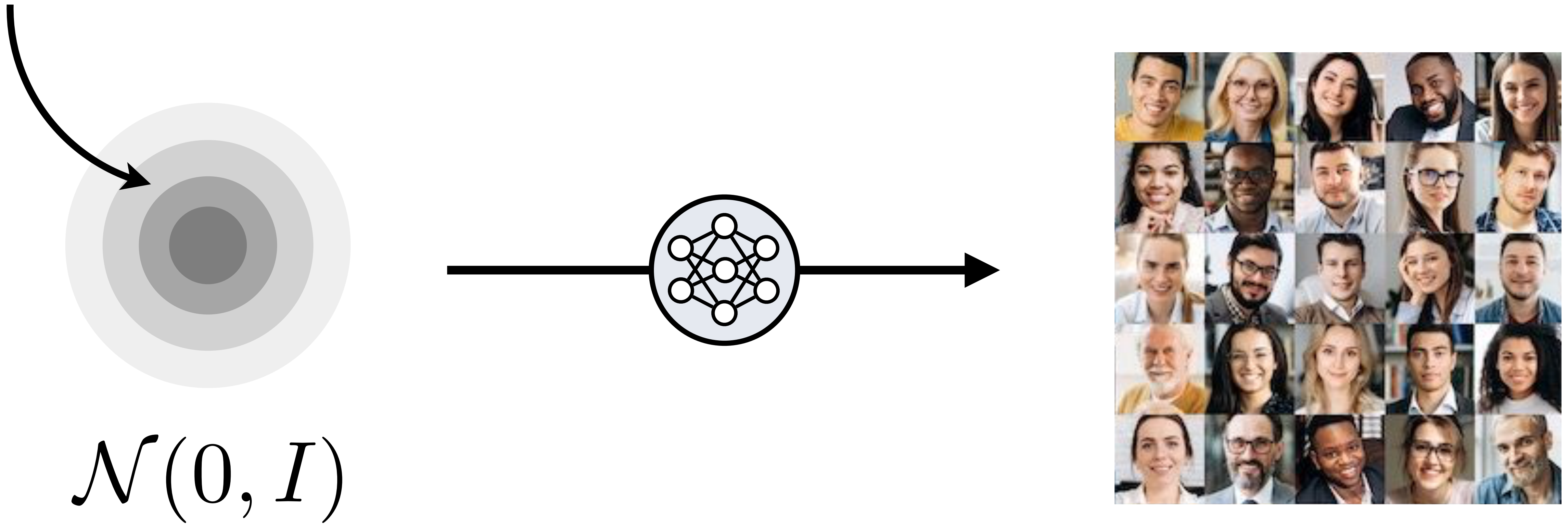
In generative modeling, OT takes different roles:

[Ajovsky+17, Genevay+18, Salimans+18, Korotin+20, Rout+22]

Other Prominent Use Case: Generative Modeling

In generative modeling, OT takes different roles:

1. OT distance serves as **loss function** between data and generated samples
2. OT maps as **generative models**

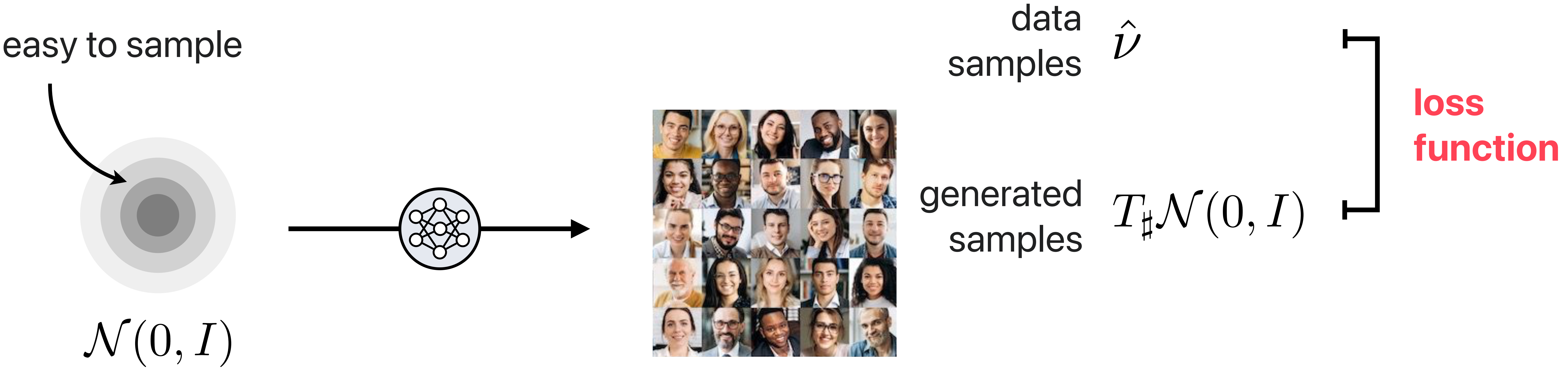


[Ajovsky+17, Genevay+18, Salimans+18, Korotin+20, Rout+22]

Other Prominent Use Case: Generative Modeling

In generative modeling, OT takes different roles:

1. OT distance serves as **loss function** between data and generated samples
2. OT maps as **generative models**

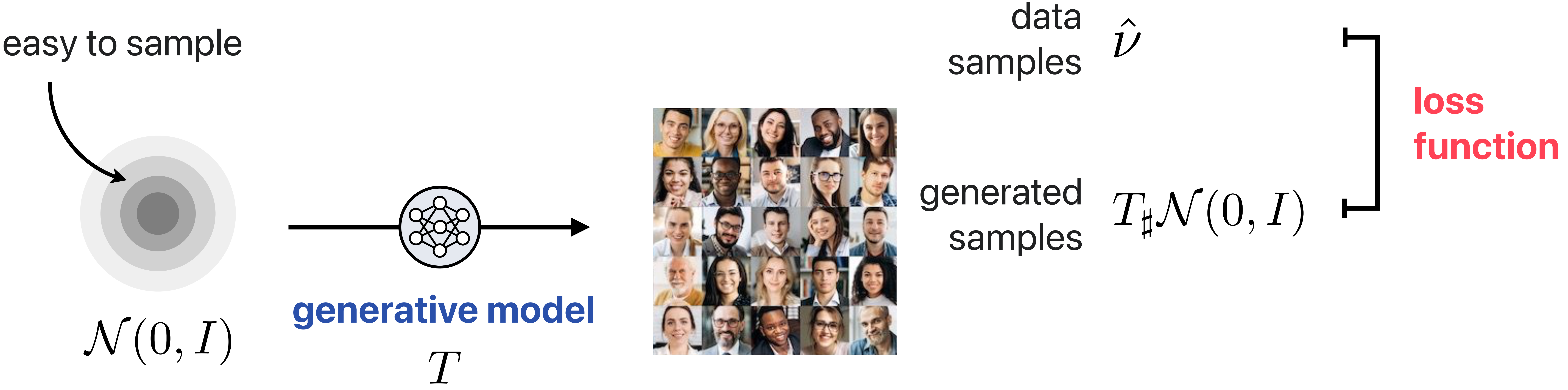


[Ajovsky+17, Genevay+18, Salimans+18, Korotin+20, Rout+22]

Other Prominent Use Case: Generative Modeling

In generative modeling, OT takes different roles:

1. OT distance serves as **loss function** between data and generated samples
2. OT maps as **generative models**



[Ajovsky+17, Genevay+18, Salimans+18, Korotin+20, Rout+22]

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

MONGE

$$T^\star = \operatorname*{argmin}_{T:T_\sharp\mu\approx\nu} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^{\star} = \underset{T:T_{\sharp}\mu \approx \nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

SEMI-DUAL

$$\varphi^\star = \underset{\varphi \text{convex}}{\operatorname{arginf}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

where $T(x) = \nabla\varphi(x)$

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

SEMI-DUAL

$$\varphi^\star = \underset{\varphi\text{convex}}{\operatorname{arginf}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

$$\sup_{x\in\mathbb{R}^d} \langle x, y\rangle - \varphi(x)$$

convex conjugate

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \underset{T : T_\sharp \mu \approx \nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Approach: learn T

SEMI-DUAL

$$\varphi^\star = \underset{\varphi \text{convex}}{\operatorname{arginf}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

$$\sup_{x \in \mathbb{R}^d} \langle x, y \rangle - \varphi(x)$$

convex conjugate

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \operatorname*{argmin}_{T:T_\sharp\mu\approx\nu} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Approach: learn T

SEMI-DUAL

$$\varphi^\star = \operatorname*{arginf}_{\varphi \text{convex}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

Approach: learn φ

$$\sup_{x\in\mathbb{R}^d} \langle x, y\rangle - \varphi(x)$$

convex conjugate

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \operatorname*{argmin}_{T : T_\sharp \mu \approx \nu} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Approach: learn T

- with [Uscidda+23] or without regularization [Yang+19]

SEMI-DUAL

$$\varphi^\star = \operatorname*{arginf}_{\varphi \text{convex}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

Approach: learn φ

$$\sup_{x \in \mathbb{R}^d} \langle x, y \rangle - \varphi(x)$$

convex conjugate

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Approach: learn T

- with [Uscidda+23] or without regularization [Yang+19]

SEMI-DUAL

$$\varphi^\star = \underset{\varphi \text{ convex}}{\operatorname{arginf}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

Approach: learn φ

- convex φ and *(near-)exact* conjugate [Taghvaei+19, Amos23]
- convex φ and *approximate* conjugate [Makkuva+21, Korotin+20]

$$\sup_{x\in\mathbb{R}^d} \langle x, y\rangle - \varphi(x)$$

convex conjugate

Various Neural Optimal Transport Solvers

How to parameterize optimal transport?

Euclidean distance as transport cost

MONGE

$$T^\star = \operatorname*{argmin}_{T:T_\sharp\mu\approx\nu} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x)$$

Approach: learn T

- with [Uscidda+23] or without regularization [Yang+19]

SEMI-DUAL

$$\varphi^\star = \operatorname*{arginf}_{\varphi \text{ convex}} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

Approach: learn φ

- convex φ and *(near-)exact* conjugate [Taghvaei+19, Amos23]
- convex φ and *approximate* conjugate [Makkuva+21, Korotin+20]
- arbitrary φ [Nhan Dam+19, Korotin+21, Rout+21]

$$\sup_{x\in\mathbb{R}^d} \langle x, y\rangle - \varphi(x)$$

convex conjugate

Neural Optimal Transport Solvers

... via the Brenier Theorem

$$\varphi^\star = \underset{\varphi \text{convex}}{\arg\inf} \int \varphi \mathrm{d}\mu + \int \varphi^* \mathrm{d}\nu$$

$$\sup_{x \in \mathbb{R}^d} \langle x, y \rangle - \varphi(x)$$

convex conjugate

Approach: learn φ

Approximate $\varphi^\star$ via convex function g

Neural Solver

[Makkuva+21, Korotin+20]

Neural Optimal Transport Solvers

Part 2

... via the Brenier Theorem

Approach: learn φ

$$\varphi^\star = \arg\inf_{\varphi \text{ convex}} \int \varphi \, \mathrm{d}\mu + \int \varphi^* \, \mathrm{d}\nu$$

$$\sup_{x \in \mathbb{R}^d} \langle x, y \rangle - \varphi(x)$$

convex conjugate

Approximate $\varphi^\star$ via convex function g

$$\arg \sup_{\varphi_\theta \text{ convex}} \inf_{g_\phi \text{ convex}} - \int \varphi_\theta(x) \mathrm{d}\mu(x) - \int \langle y, \nabla g_\phi(y) \rangle - \varphi_\theta(\nabla g_\phi(y)) \, \mathrm{d}\nu(y)$$

Neural Solver

[Makkuva+21, Korotin+20]

Neural Optimal Transport Solvers

... via the Brenier Theorem

$$\varphi^\star = \arg\inf_{\varphi \text{ convex}} \int \varphi \, \mathrm{d}\mu + \int \varphi^* \, \mathrm{d}\nu$$

Approach: learn φ

$$\sup_{x \in \mathbb{R}^d} \langle x, y \rangle - \varphi(x)$$

convex conjugate

Approximate $\varphi^\star$ via convex function g

$$\arg \sup_{\varphi_\theta \text{ convex}} \inf_{g_\phi \text{ convex}} - \int \varphi_\theta(x) \mathrm{d}\mu(x) - \int \langle y, \nabla g_\phi(y) \rangle - \varphi_\theta(\nabla g_\phi(y)) \, \mathrm{d}\nu(y)$$

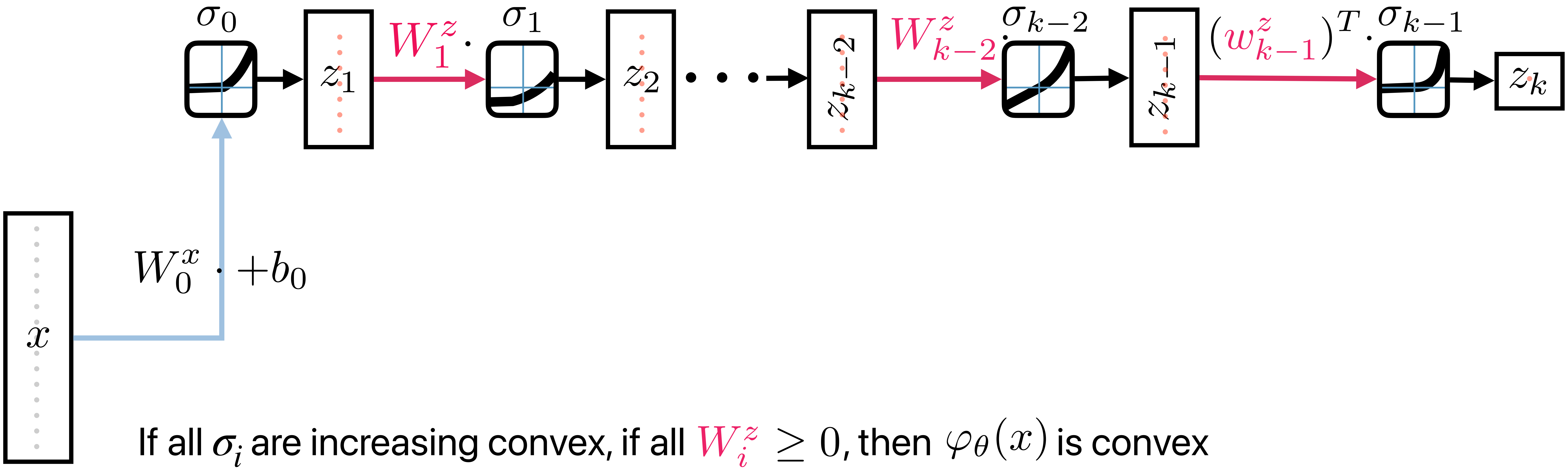
↳ parameterize via **input convex neural networks** [Amos+17]

Neural Solver [Makkuva+21, Korotin+20]

Input Convex Neural Networks (ICNN)

[Amos+17]

$$z_{i+1} = \sigma_i \left(W_i^x x + W_i^z z_i + b_i \right) \qquad \varphi_\theta(x) = z_k$$

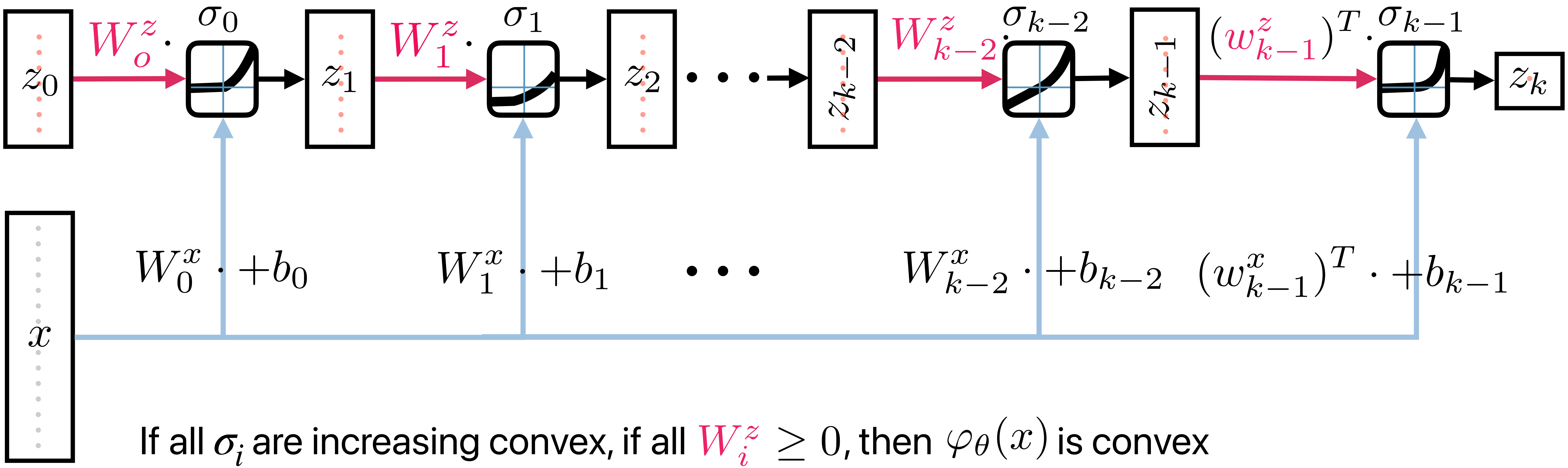


If all σ_i are increasing convex, if all $W_i^z \geq 0$, then $\varphi_\theta(x)$ is convex

Input Convex Neural Networks (ICNN)

[Amos+17]

$$z_0 = \mathbf{0} \qquad z_{i+1} = \sigma_i \left(W_i^x x + W_i^z z_i + b_i \right) \qquad \varphi_\theta(x) = z_k$$



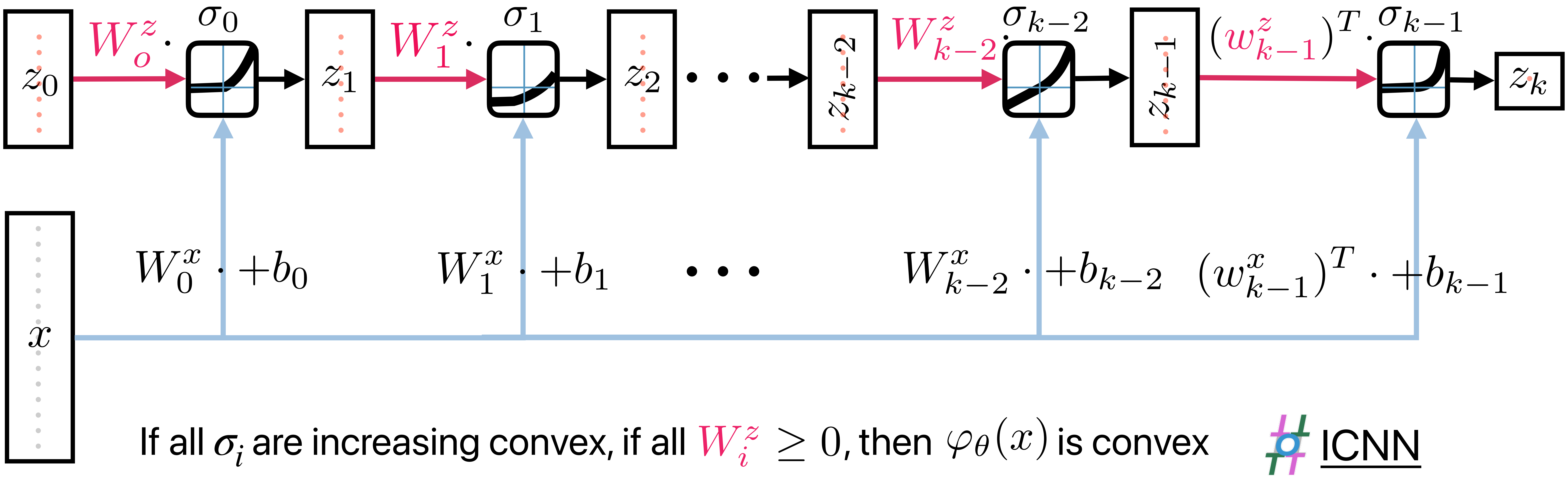
If all σ_i are increasing convex, if all $W_i^z \geq 0$, then $\varphi_\theta(x)$ is convex

Input Convex Neural Networks (ICNN)

Part 2

[Amos+17]

$$z_0 = \mathbf{0} \qquad z_{i+1} = \sigma_i \left(W_i^x x + W_i^z z_i + b_i \right) \qquad \varphi_\theta(x) = z_k$$



If all σ_i are increasing convex, if all $W_i^z \geq 0$, then $\varphi_\theta(x)$ is convex

ICNN

Neural Optimal Transport Solvers

Part 2

… via the Brenier Theorem

Approximate φ^* via convex function g

$$\arg \sup_{\varphi_\theta \text{ convex}} \inf_{g_\phi \text{ convex}} - \int \varphi_\theta(x) \mathrm{d}\mu(x) - \int \langle y, \nabla g_\phi(y) \rangle - \varphi_\theta(\nabla g_\phi(y)) \, \mathrm{d}\nu(y)$$

→ parameterize via **input convex neural networks** [Amos+17] [Makkuva+21, Korotin+20]

However, can we find approaches without

- **constraints** in the weight matrices, [Korotin+21]
- the challenge to **handle conjugate**, [Amos23]
- the problem of **initialization**? [Bunne+22]

Neural Optimal Transport Solvers

… via a Monge Gap for OT-like Maps

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} c(x, T(x))d\mu(x)$$

Approach: learn T

- For **any cost** function c

[Uscidda+23]

Neural Optimal Transport Solvers

… via a Monge Gap for OT-like Maps

$$T^\star = \operatorname*{argmin}_{T:T_\sharp\mu\approx\nu} \int_{\mathbb{R}^d} c(x, T(x))d\mu(x)$$

Approach: learn T

- For **any cost** function c
- Parameterize T via neural network and a **regularizer**

[Uscidda+23]

Neural Optimal Transport Solvers

... via a Monge Gap for OT-like Maps

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\mathrm{argmin}} \int_{\mathbb{R}^d} c(x, T(x))d\mu(x)$$

Approach: learn T

- For **any cost** function c
- Parameterize T via neural network and a **regularizer**

Monge gap regularizer $= \int c(x, T_\theta(x))\mathrm{d}\mu(x) - W_c(\mu, T_\theta\sharp\mu)\,.$

gap quantifies how far a map T_θ deviates
from ideal properties we expect from a c-OT map

[Uscidda+23]

Neural Optimal Transport Solvers

... via a Monge Gap for OT-like Maps

$$T^\star = \underset{T:T_\sharp\mu\approx\nu}{\operatorname{argmin}} \int_{\mathbb{R}^d} c(x, T(x))d\mu(x)$$

Approach: learn T

- For **any cost** function c
- Parameterize T via neural network and a **regularizer**

Monge gap regularizer $= \int c(x, T_\theta(x))\mathrm{d}\mu(x) - W_c(\mu, T_\theta\sharp\mu)\,.$

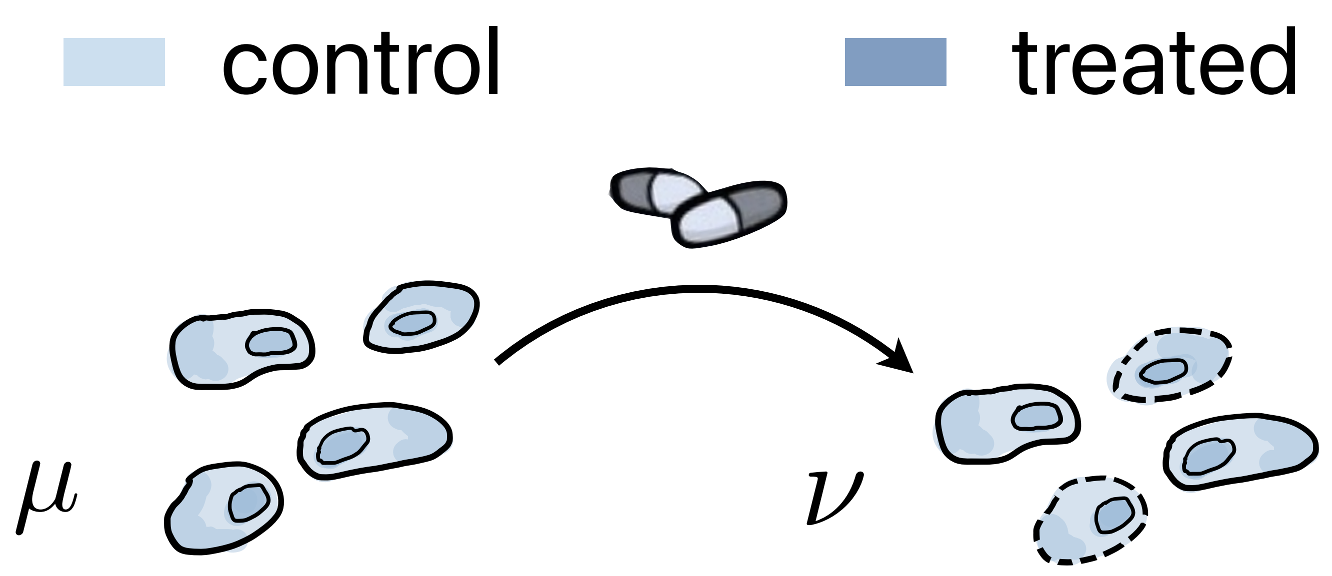
gap quantifies how far a map T_θ deviates from ideal properties we expect from a c-OT map

[Uscidda+23]

Monge Gap

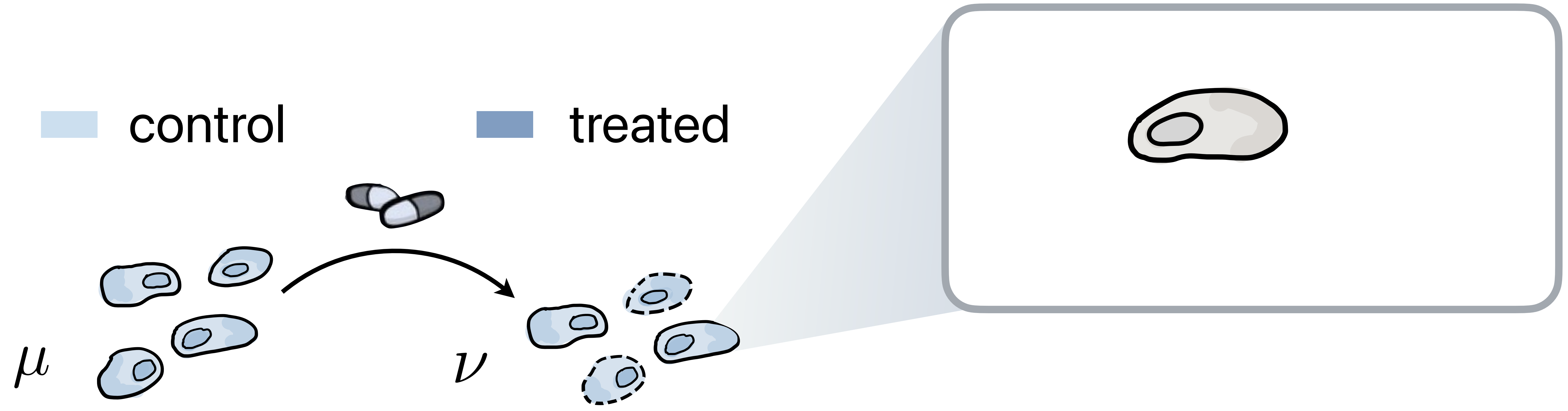
Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**



Application: **Predicting Treatment Responses to Drugs**

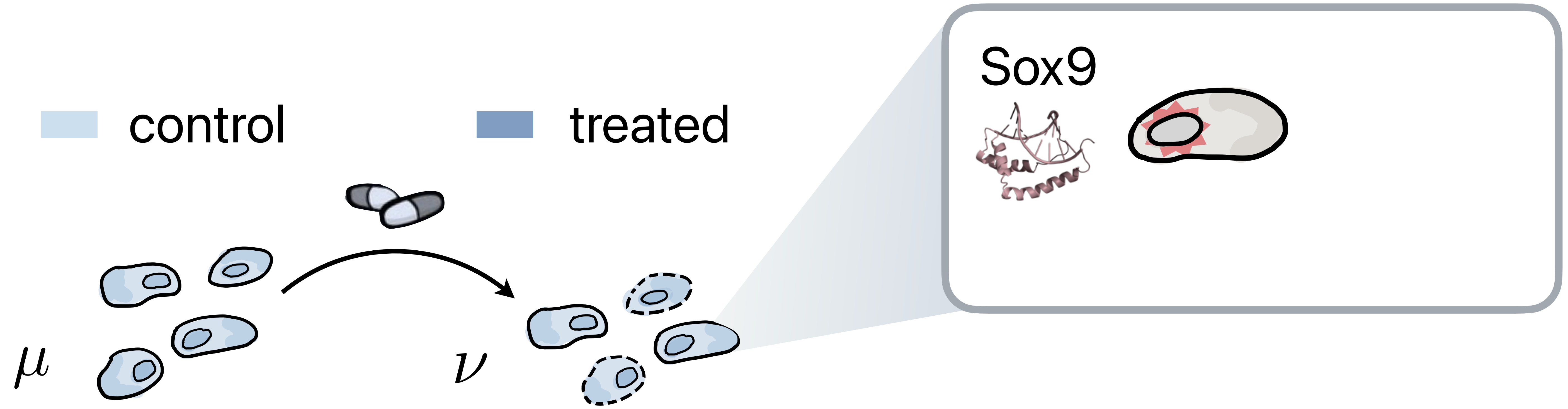
Example of a **drug** on a **mixture of two cell lines**



4i microscopy
[Gut+18]

Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**

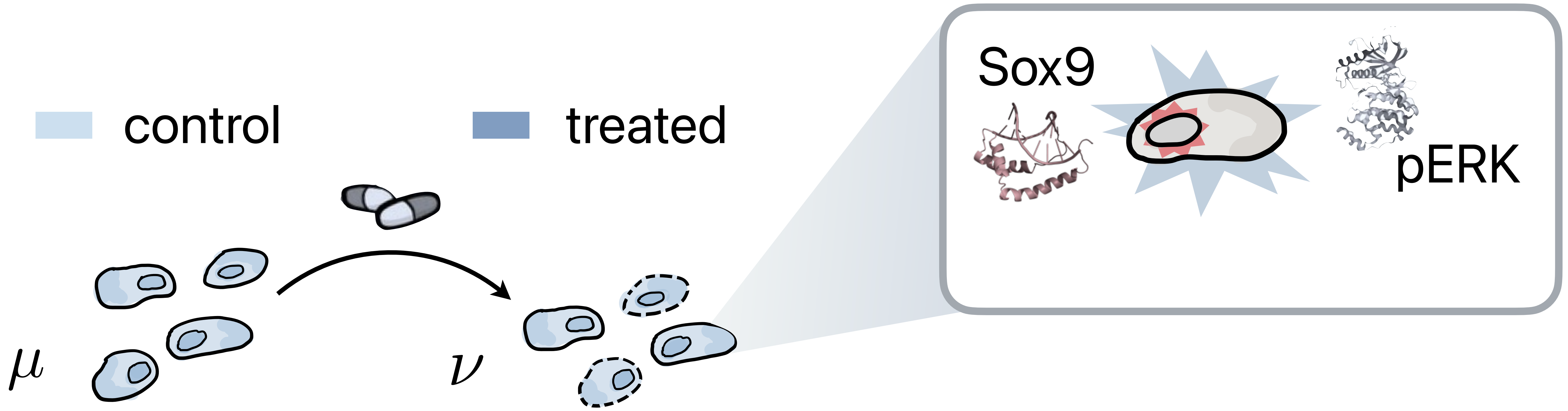


4i microscopy

[Gut+18]

Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**

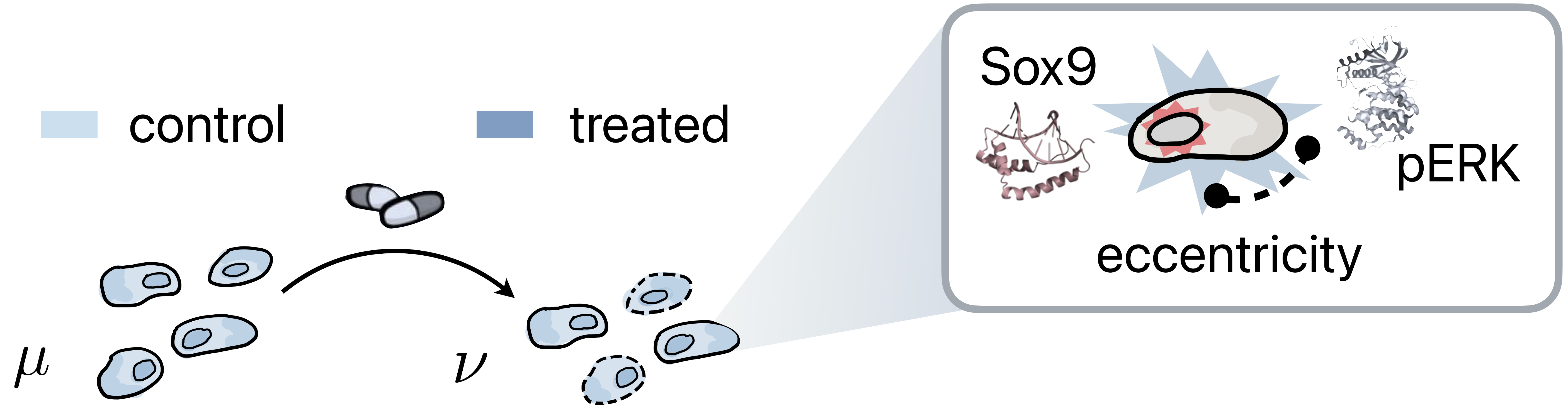


4i microscopy

[Gut+18]

Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**

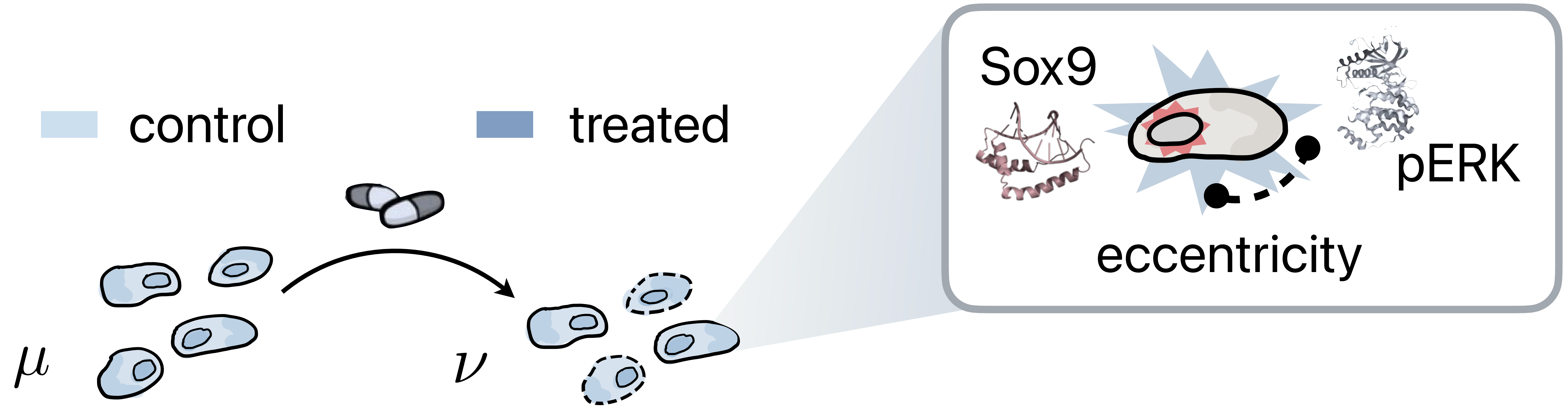


4i microscopy

[Gut+18]

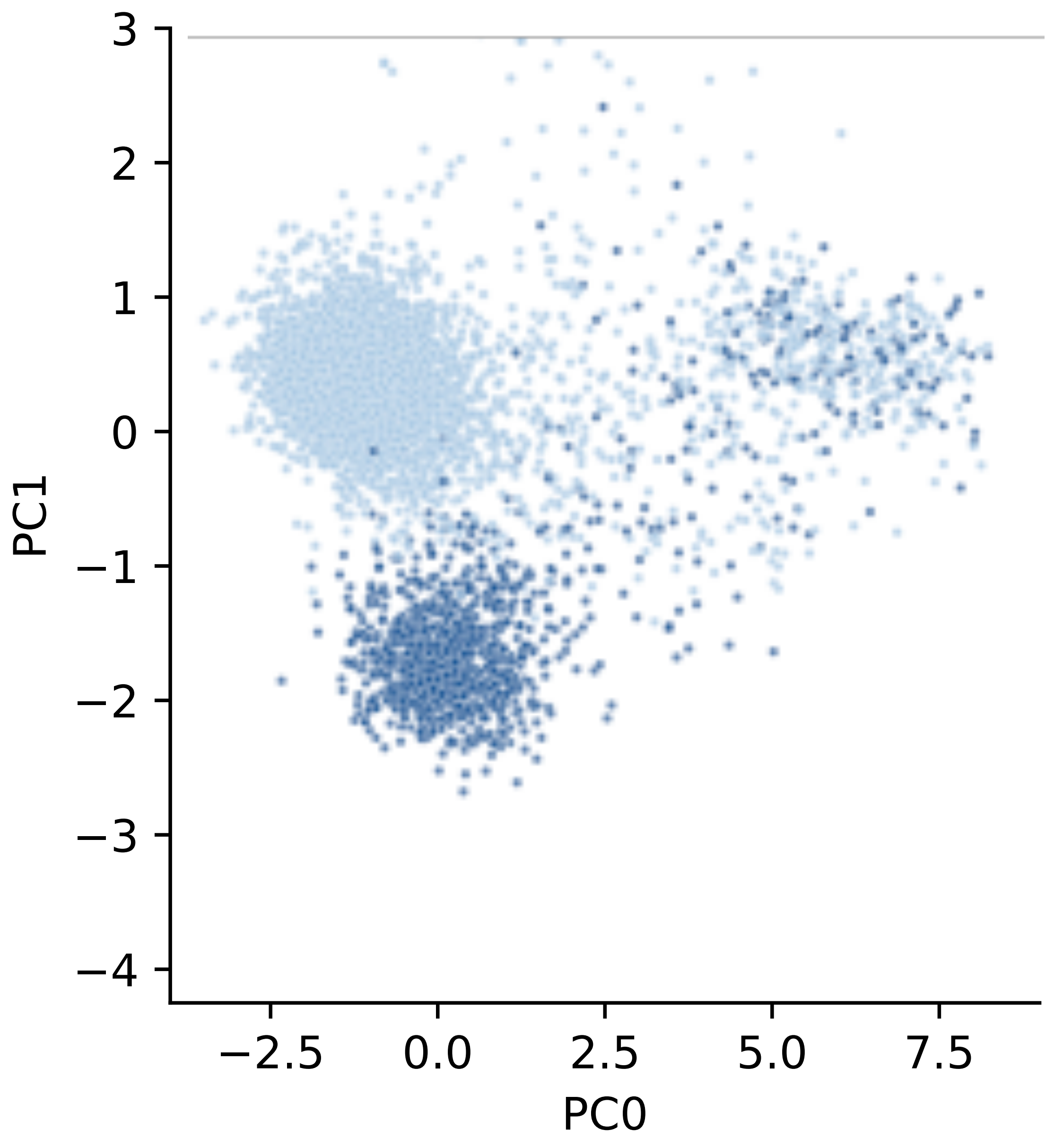
Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**



4i microscopy

[Gut+18]

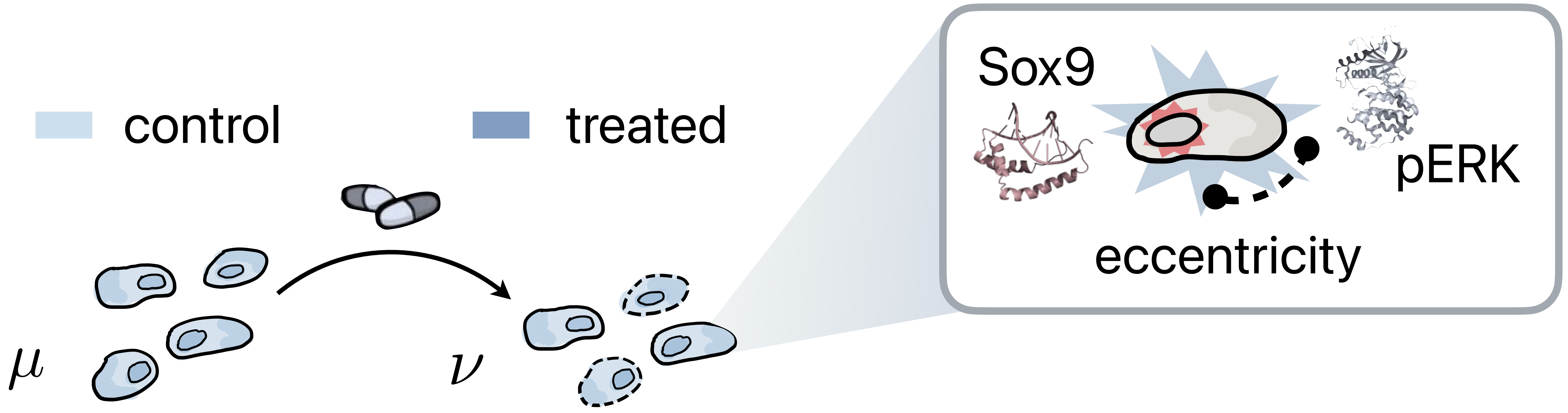


control

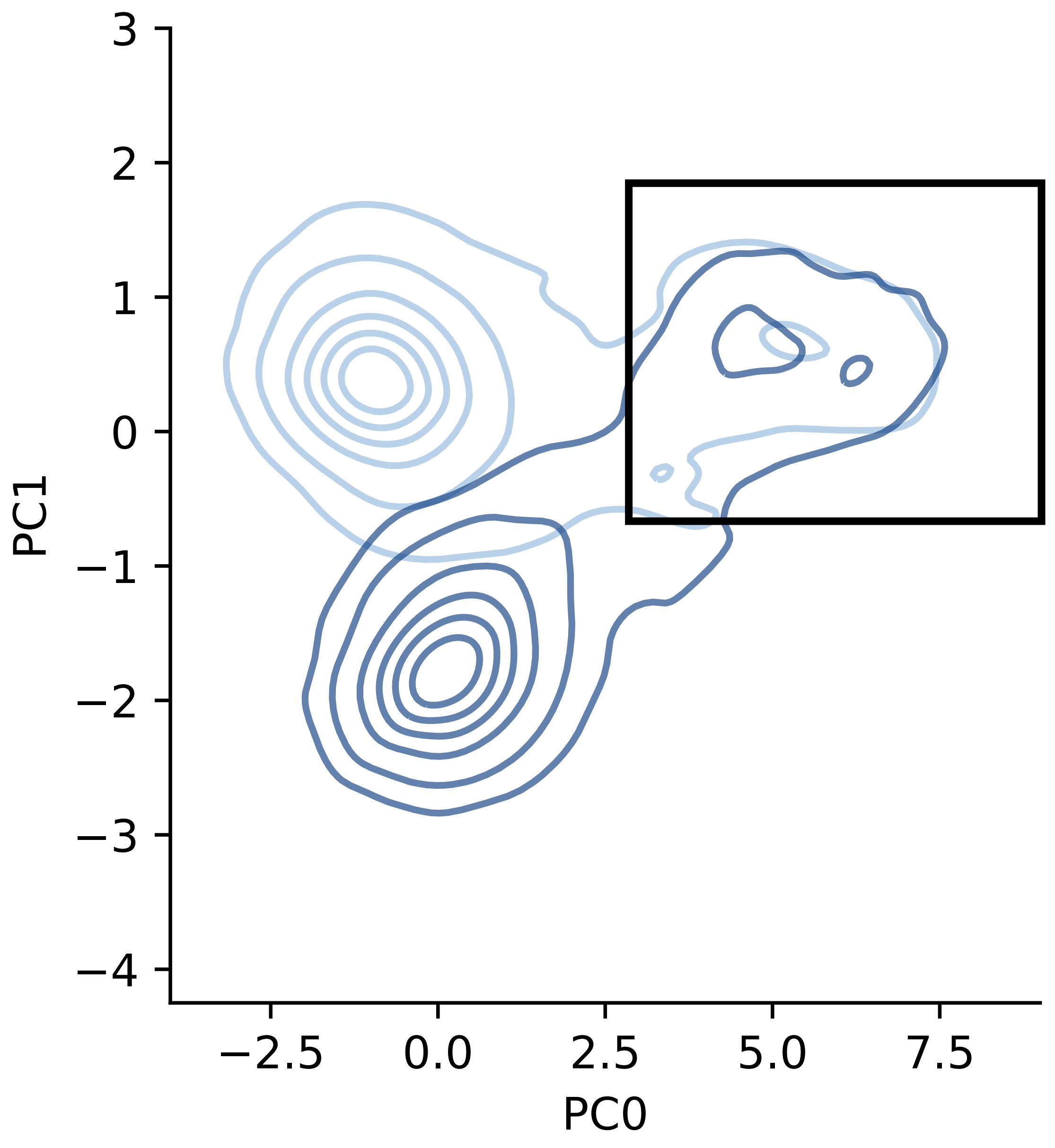
treated

Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**



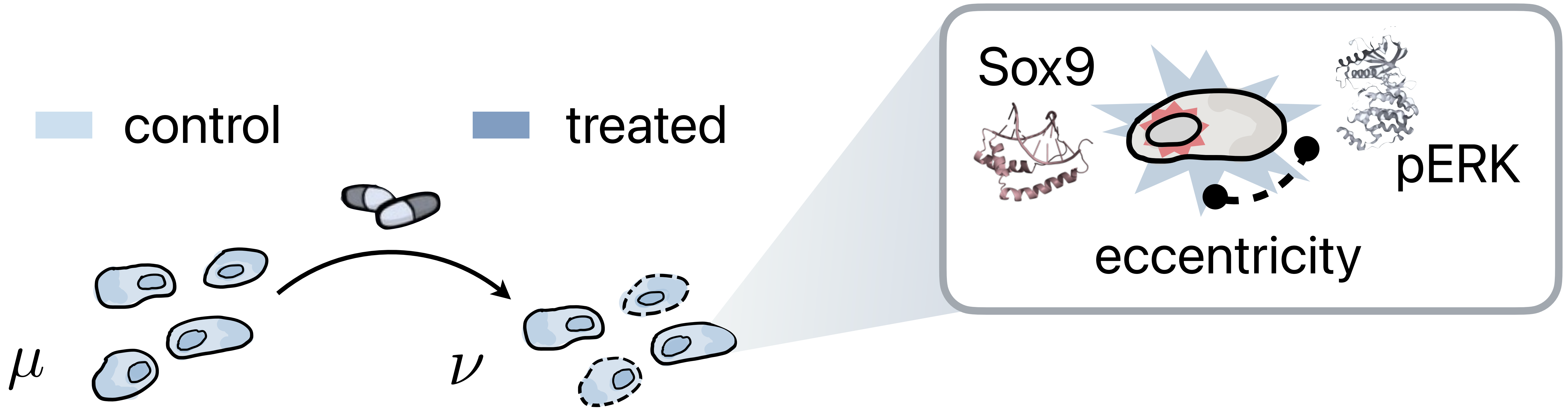
4i microscopy
[Gut+18]



control
treated

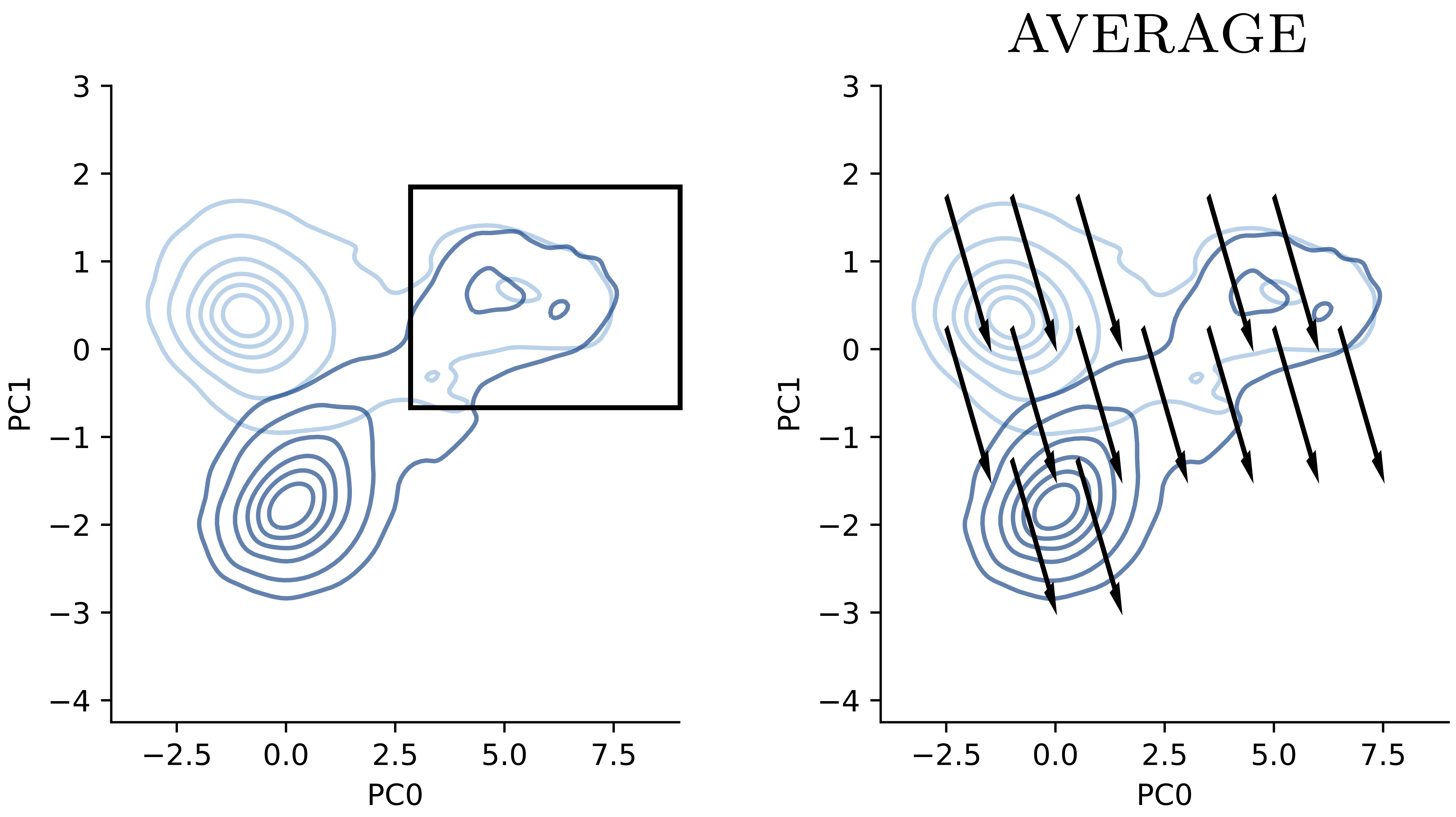
Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**



4i microscopy

[Gut+18]

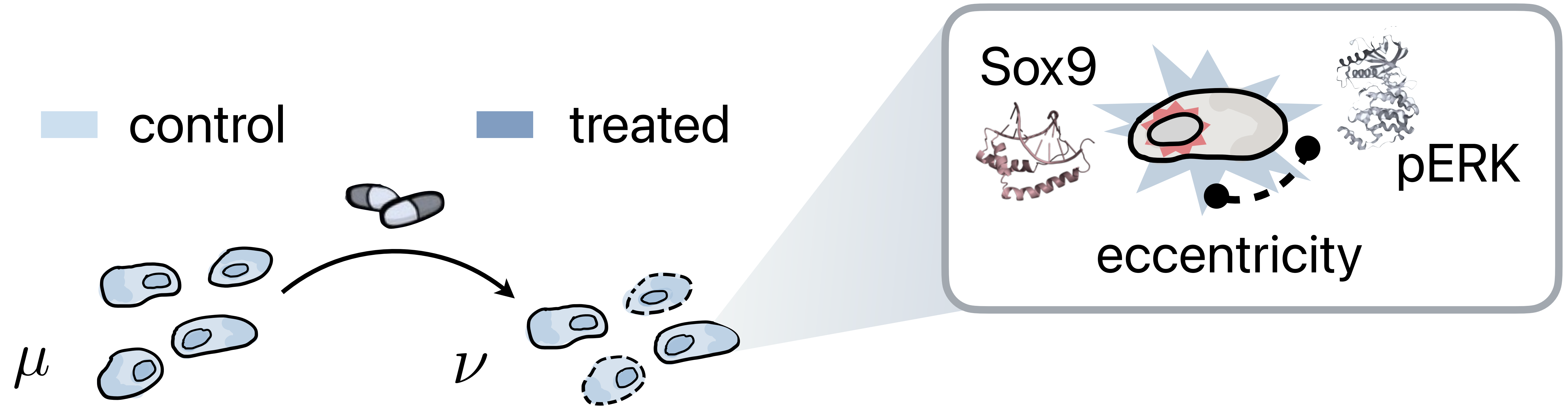


control

treated

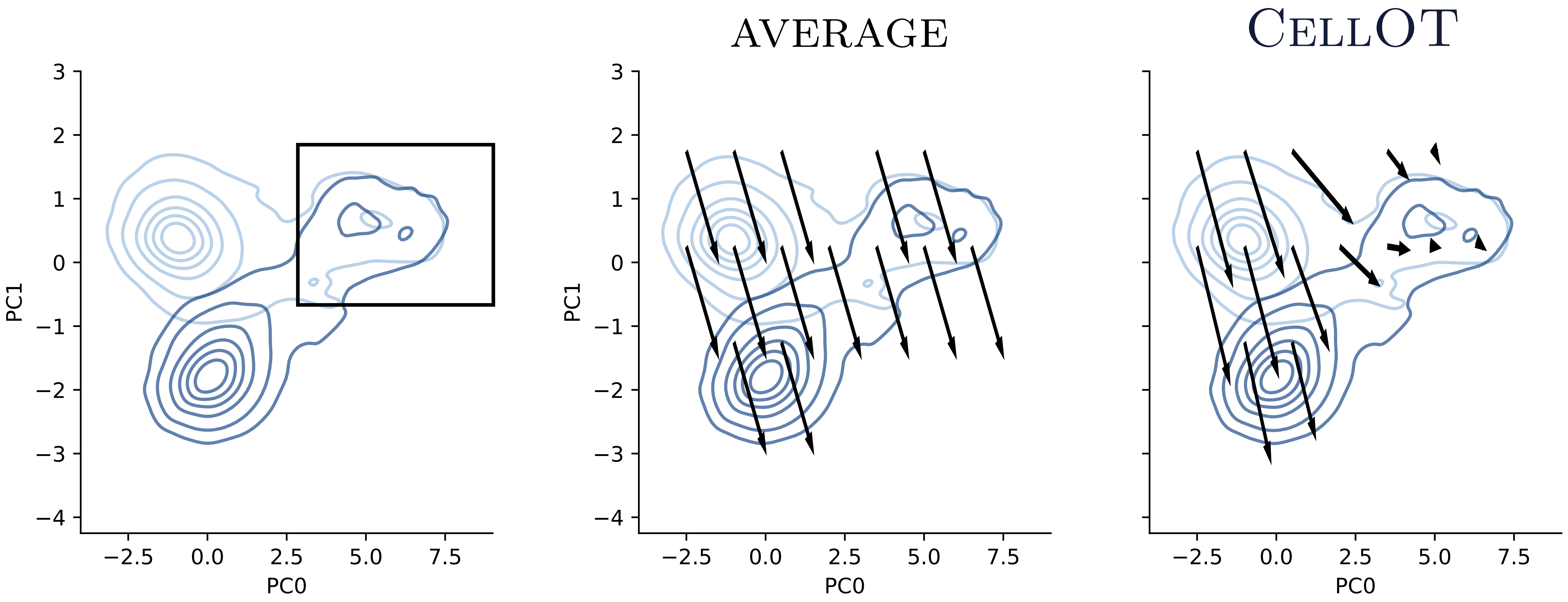
Application: **Predicting Treatment Responses to Drugs**

Example of a **drug** on a **mixture of two cell lines**



4i microscopy

[Gut+18]



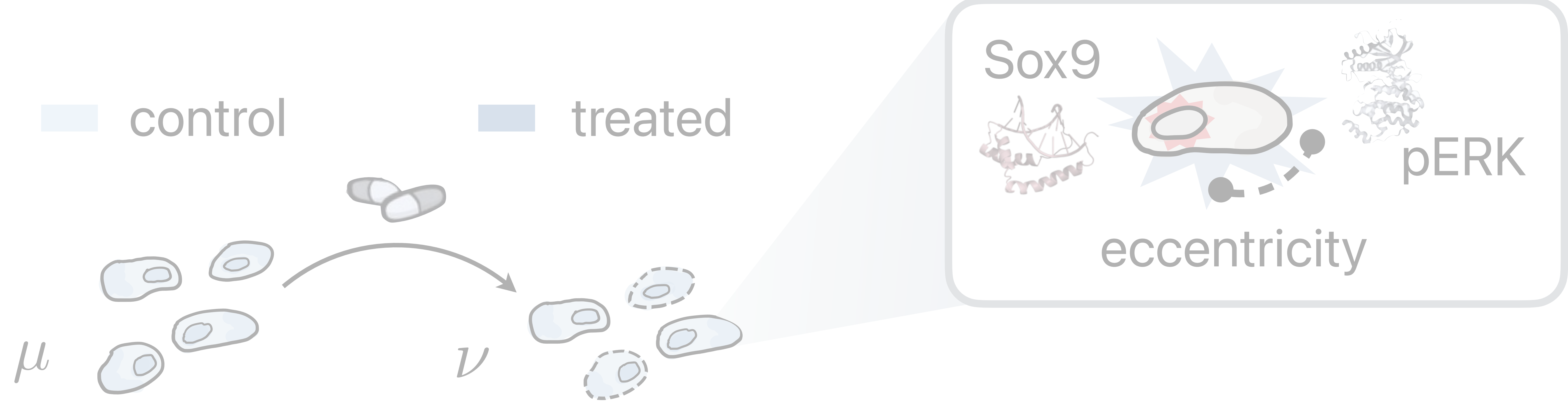
[Bunne+23]

control

treated

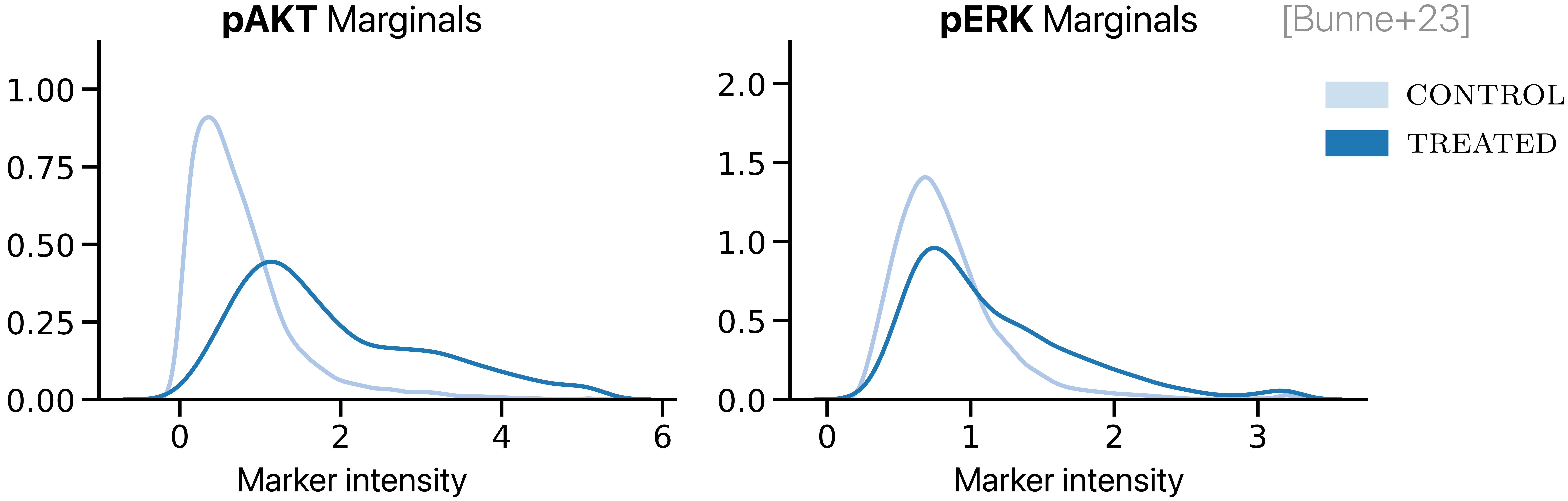
Application: **Predicting Treatment Responses to Drugs**

… do we capture **changes in molecular markers**?



4i microscopy
[Gut+18]

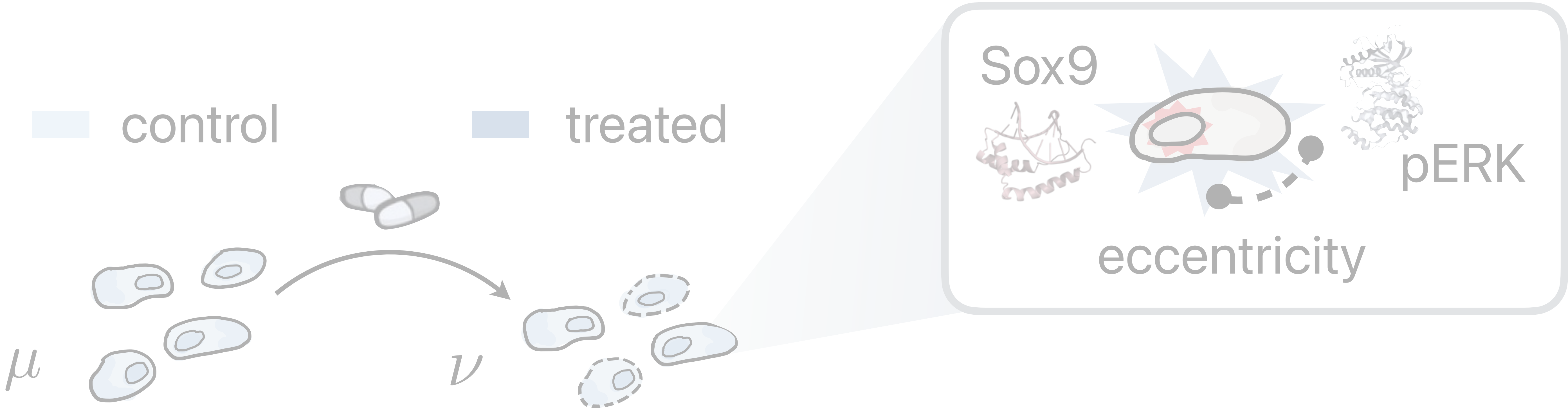
[Bunne+23]



Application: **Predicting Treatment Responses to Drugs**

Part 2

... do we capture **changes in molecular markers**?

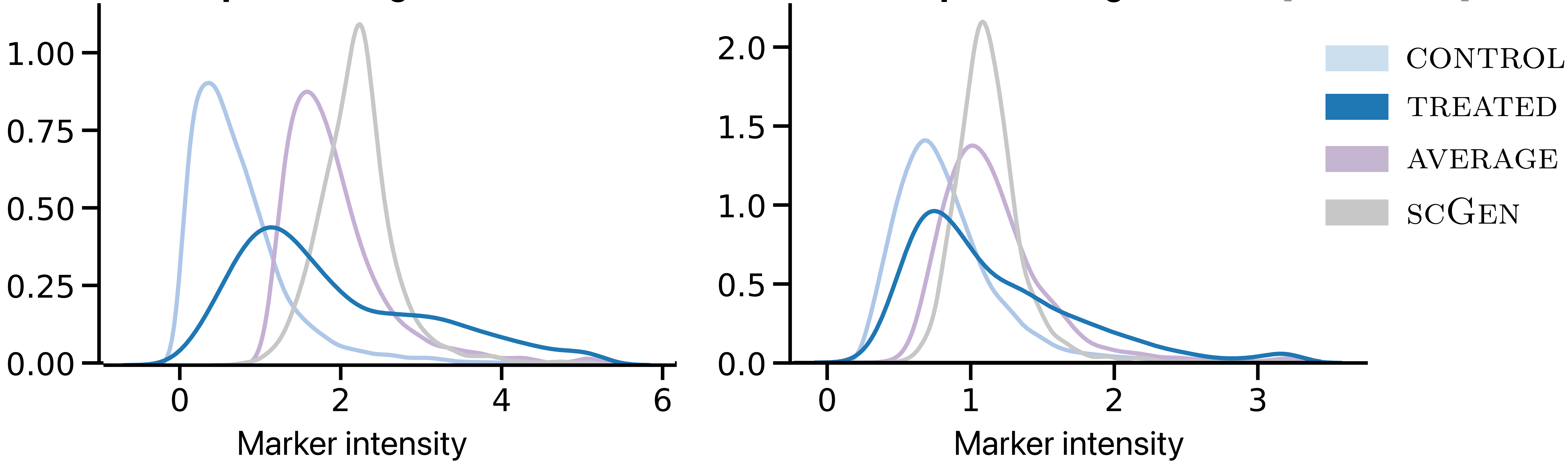


4i microscopy
[Gut+18]

pAKT Marginals

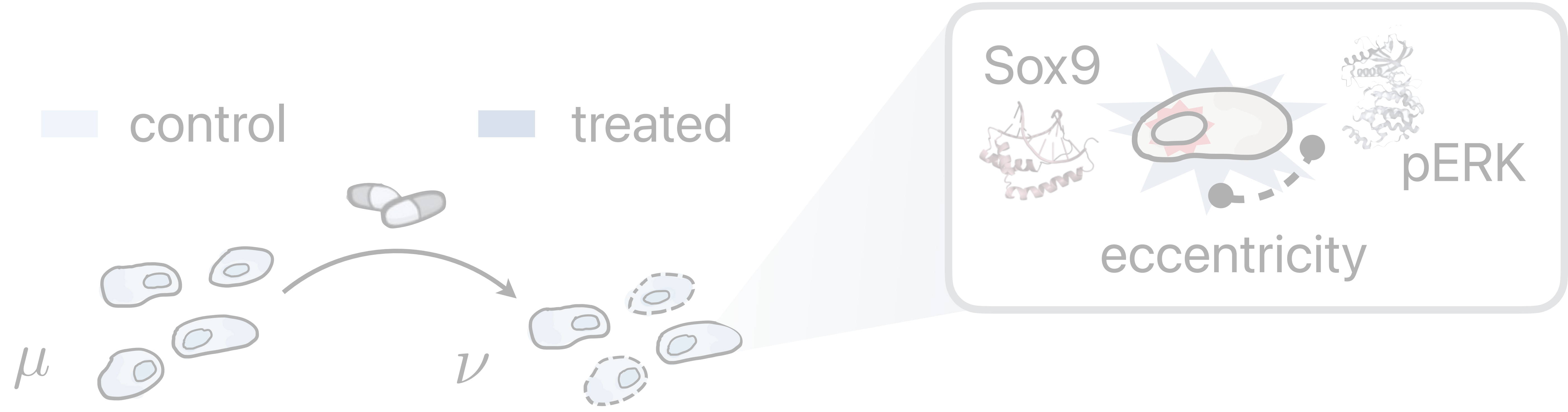
pERK Marginals

[Bunne+23]



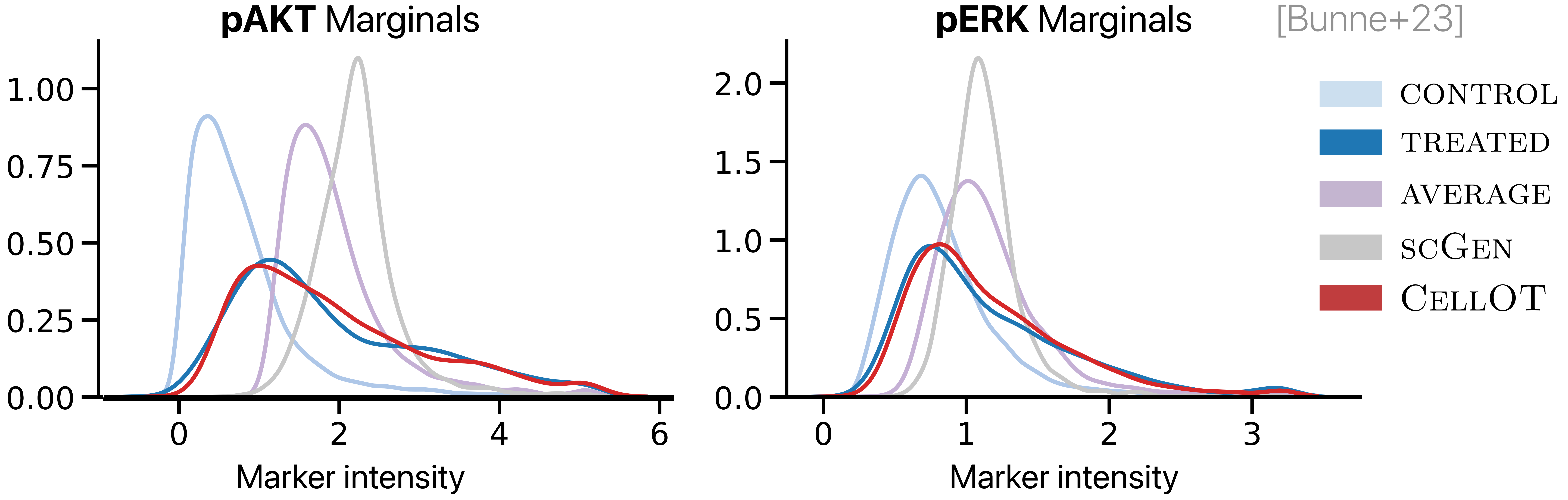
Application: **Predicting Treatment Responses to Drugs**

... do we capture **changes in molecular markers**?



4i microscopy
[Gut+18]

[Bunne+23]



Perspective

Benchmark [Korotin+21]

Extensions
- to structured and general cost functionals [Klein+23, Korotin+22, Fan+21]
- to unbalanced problems [Lübeck+22, Eyring+22]
- to incomparable spaces [Nekrashevich+23, Bunne+19, Dumont+22]
- to conditional settings [Bunne+22]
- to barycenter estimation [Korotin+22, Fan+21]
- to Riemannian manifolds [Rezende+21, Scarvelis+22]

Statistics of entropic maps [Rigollet+22, Pooladian+21]

Applications
- e.g., image super-resolution [Gazdieva+22]

Outline of the Tutorial

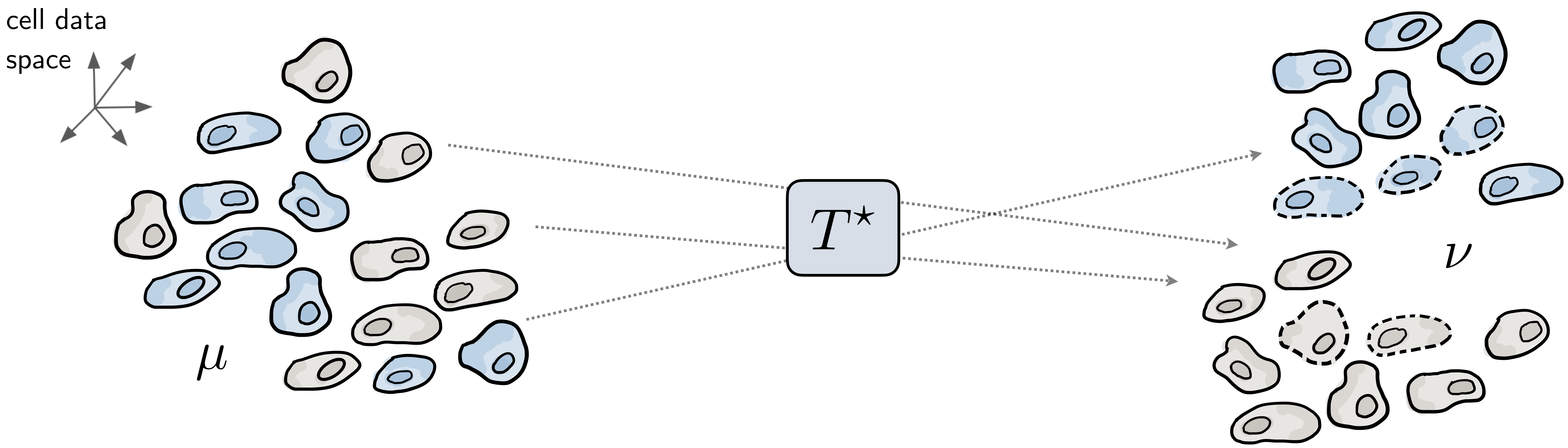
A Dynamic Perspective on Optimal Transport

Part 3

So far: **Static Optimal Transport**

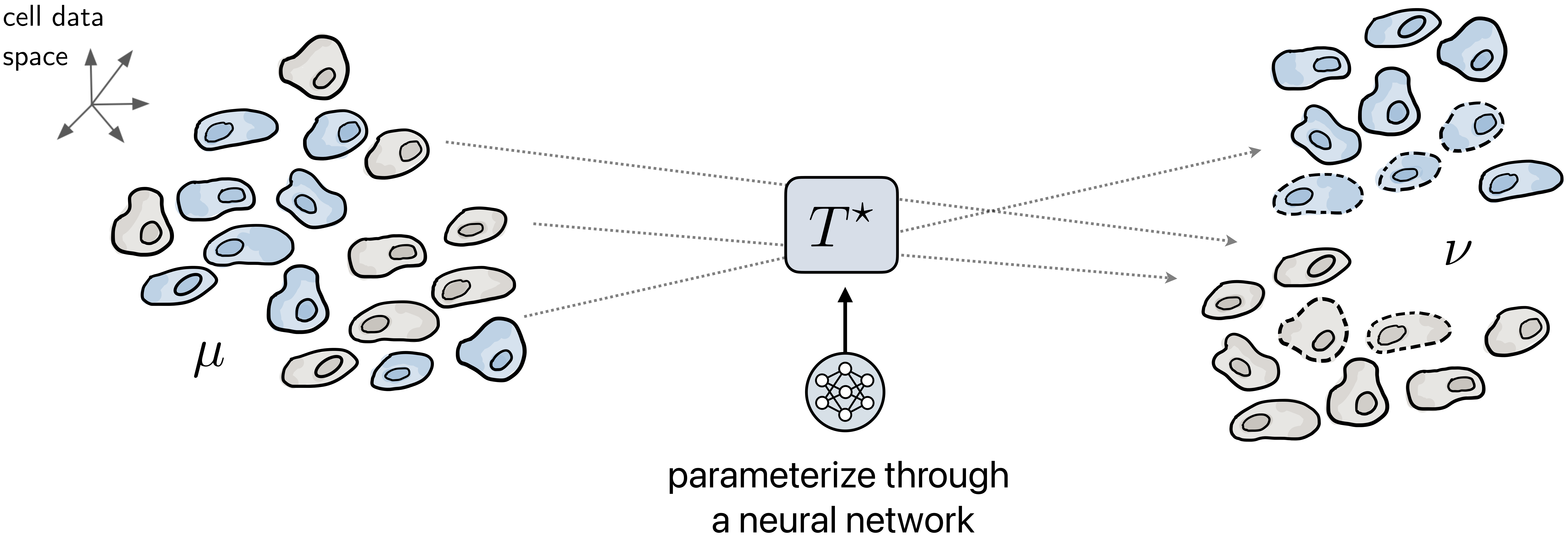
A Dynamic Perspective on Optimal Transport

So far: **Static Optimal Transport**



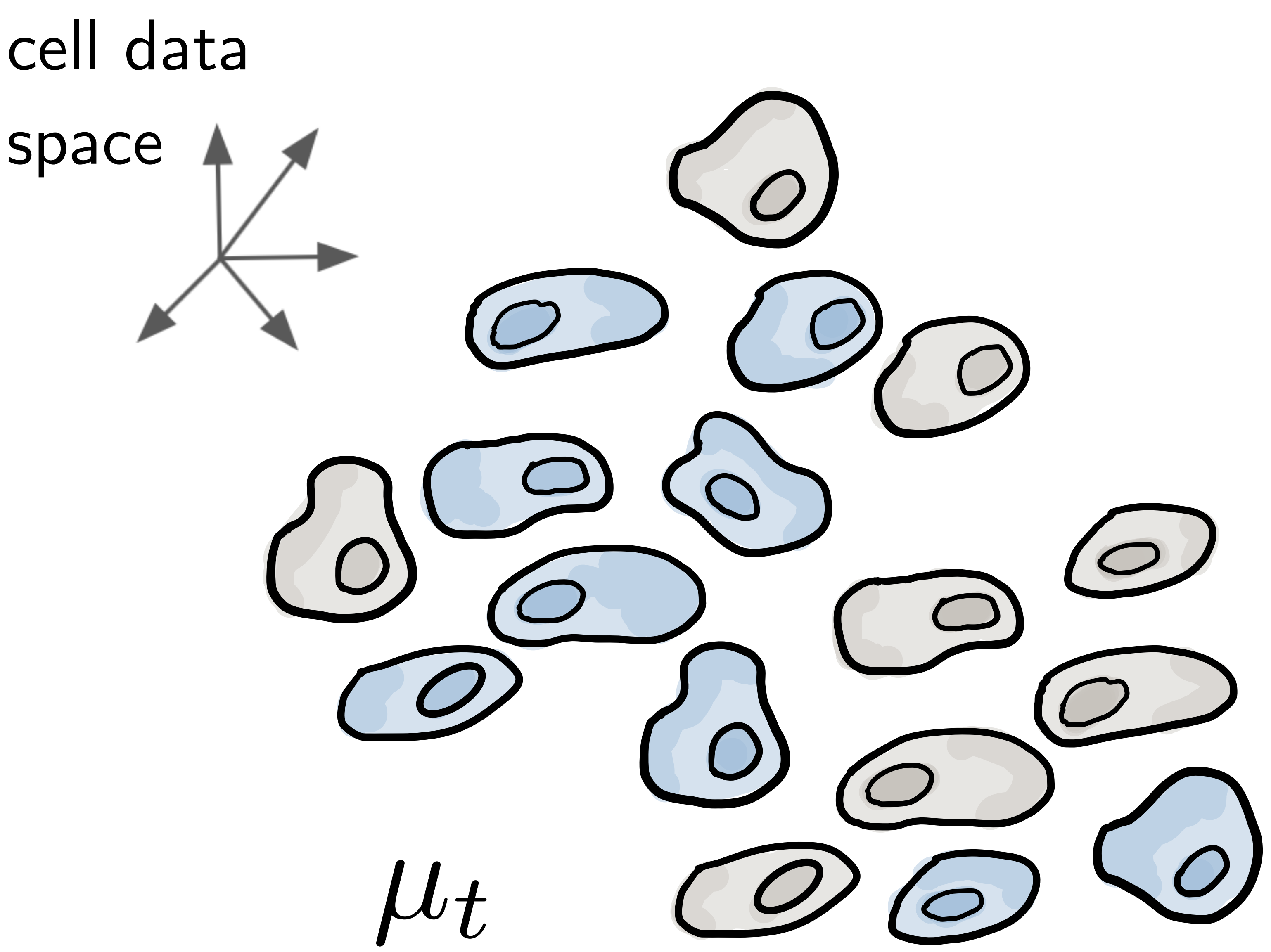
A Dynamic Perspective on Optimal Transport

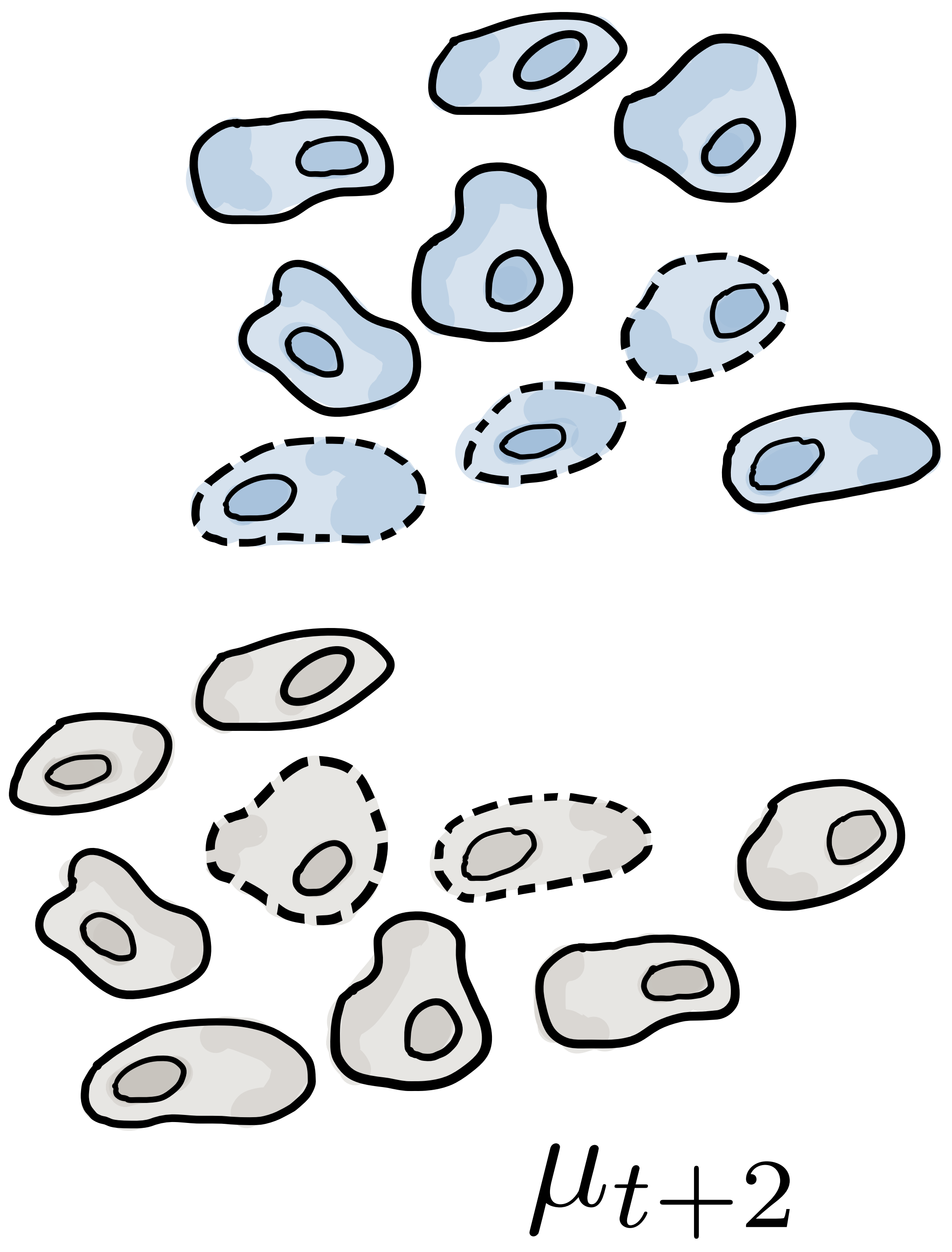
So far: **Static Optimal Transport**



A Dynamic Perspective on Optimal Transport

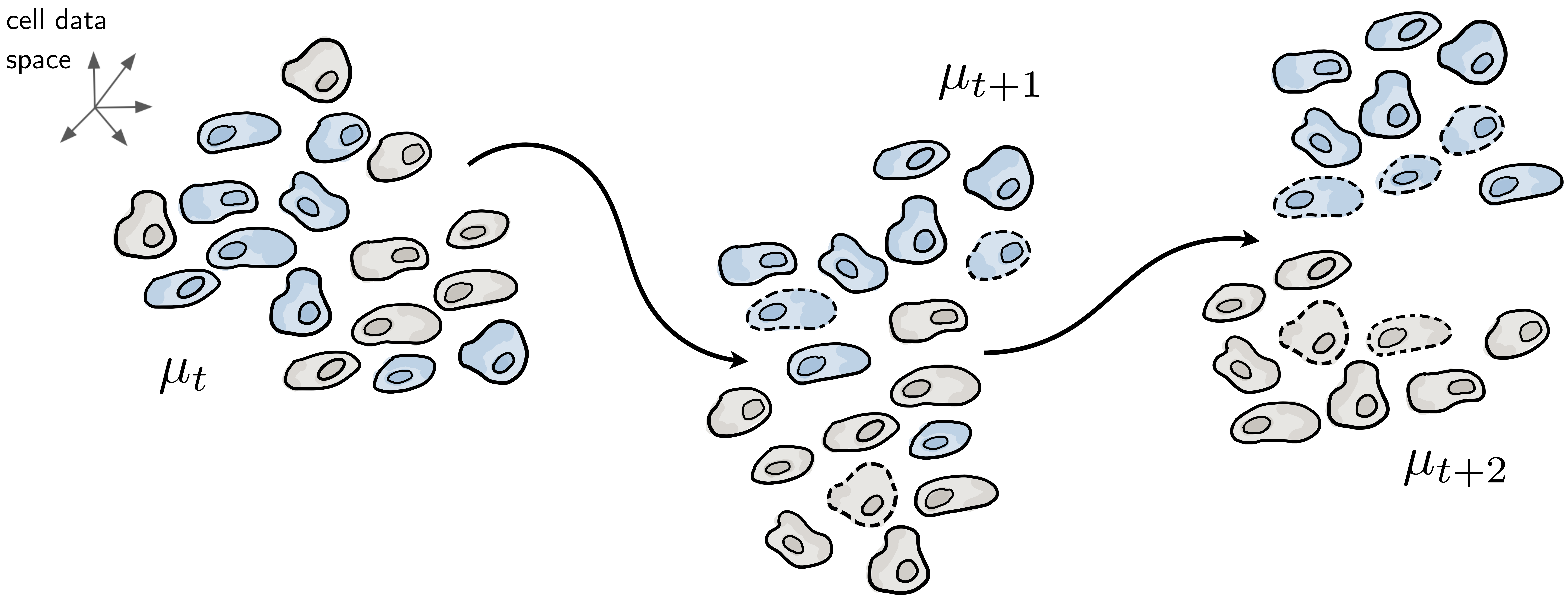
Next: **Dynamic Optimal Transport**





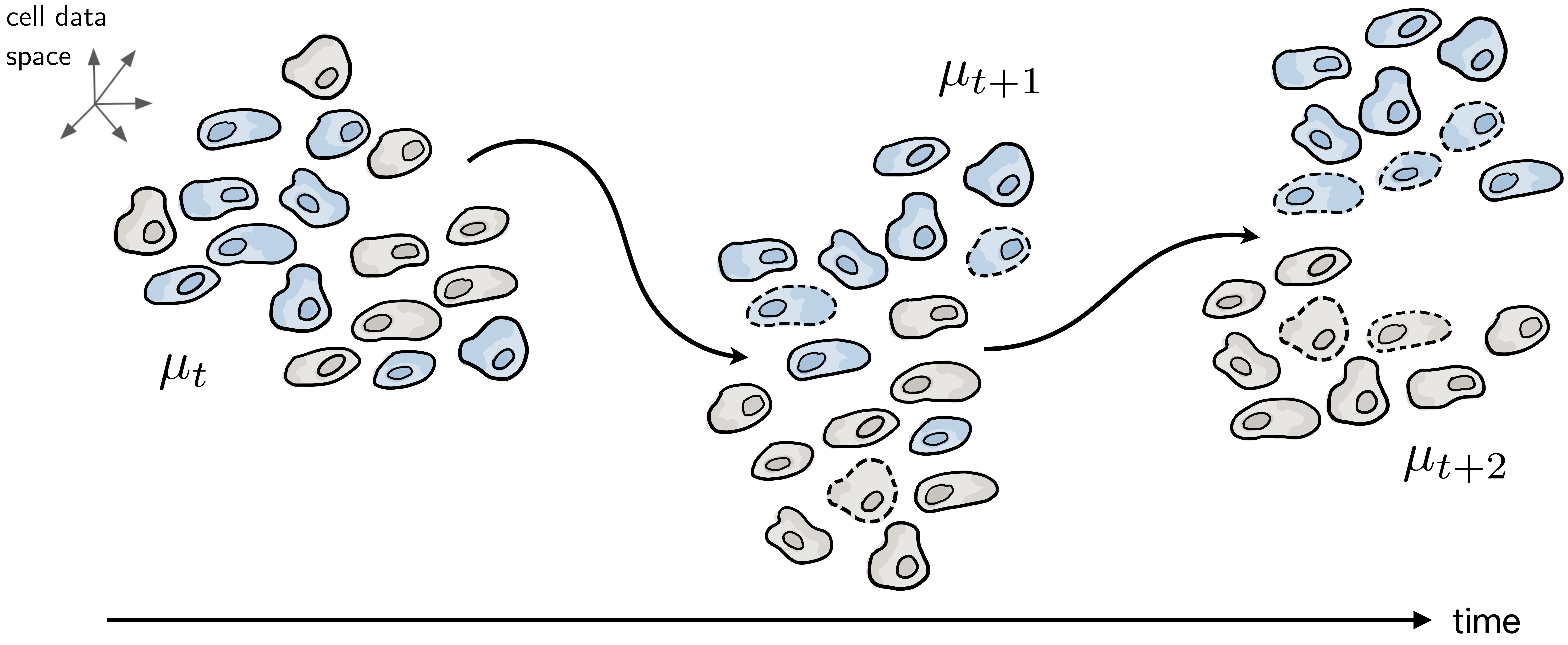
A Dynamic Perspective on Optimal Transport

Next: **Dynamic Optimal Transport**



A Dynamic Perspective on Optimal Transport

Next: **Dynamic Optimal Transport**



Outline of the Tutorial

Prelude Warm-Up: Starting with Optimal Matchings

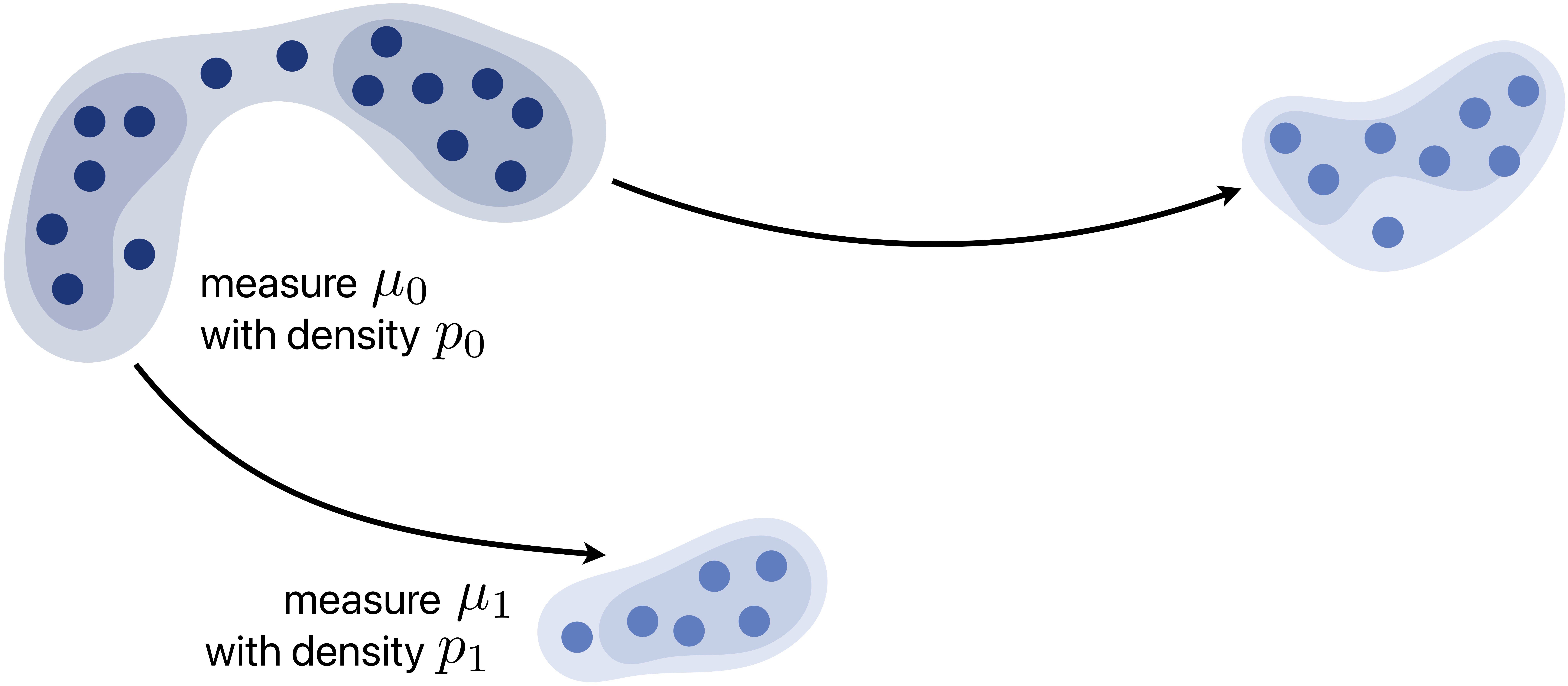
Part 1 Kantorovich Formulation of OT and Computations

Part 2 Duality, Monge Formulations and Brenier Theorems

Part 3 **Modeling Measure Dynamics with Optimal Transport**

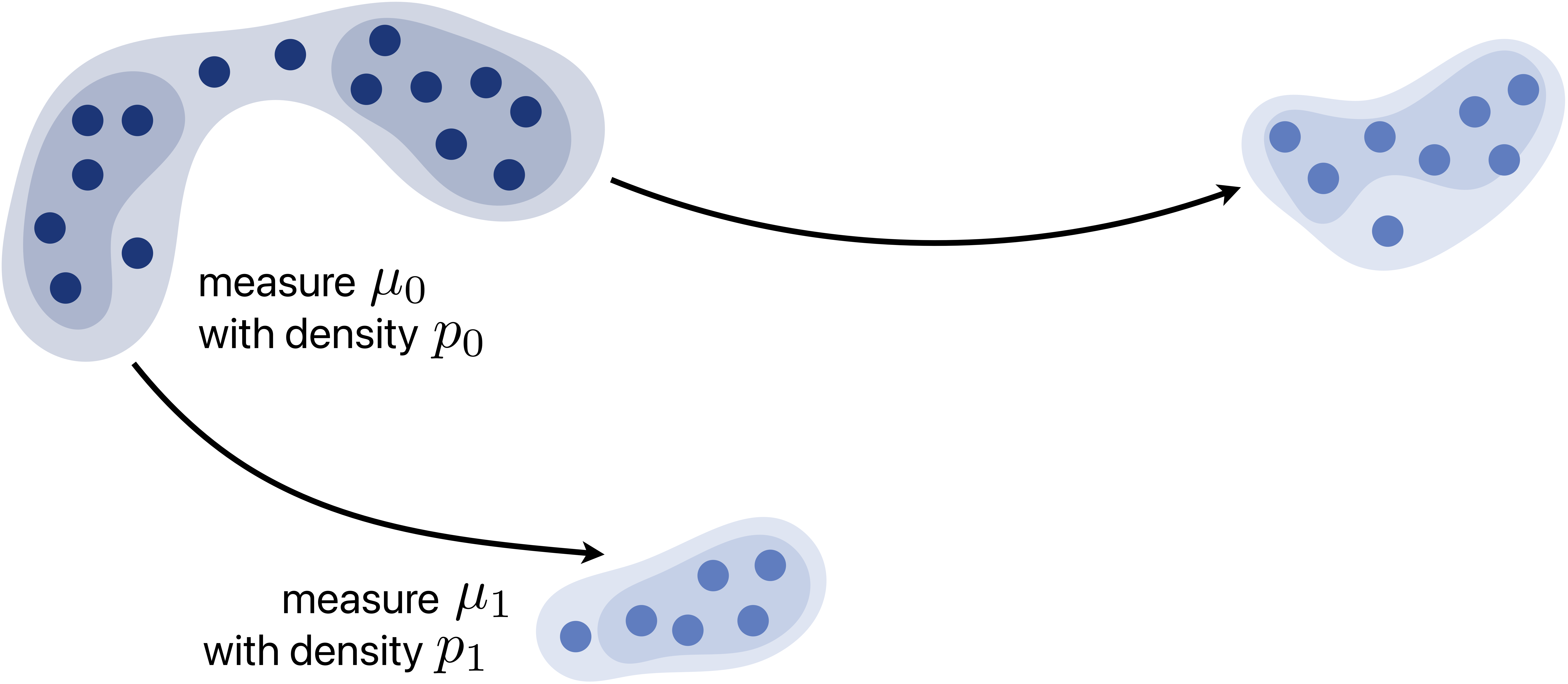
- Benamou-Brenier Formulation
- Jordan-Kinderlehrer-Otto Flows
- Stochastic Control Perspective and Schrödinger Bridges

Monge-Ampère Equation



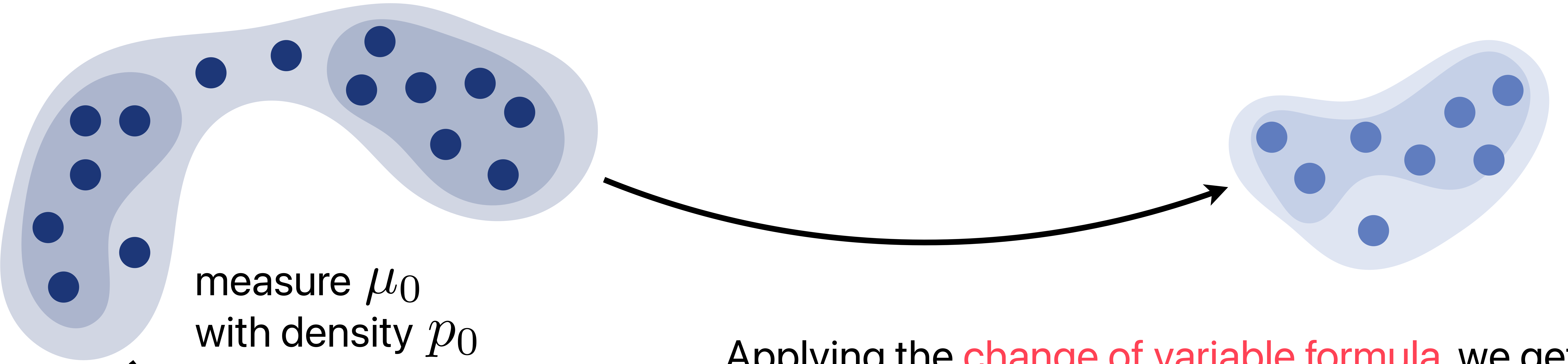
Monge-Ampère Equation

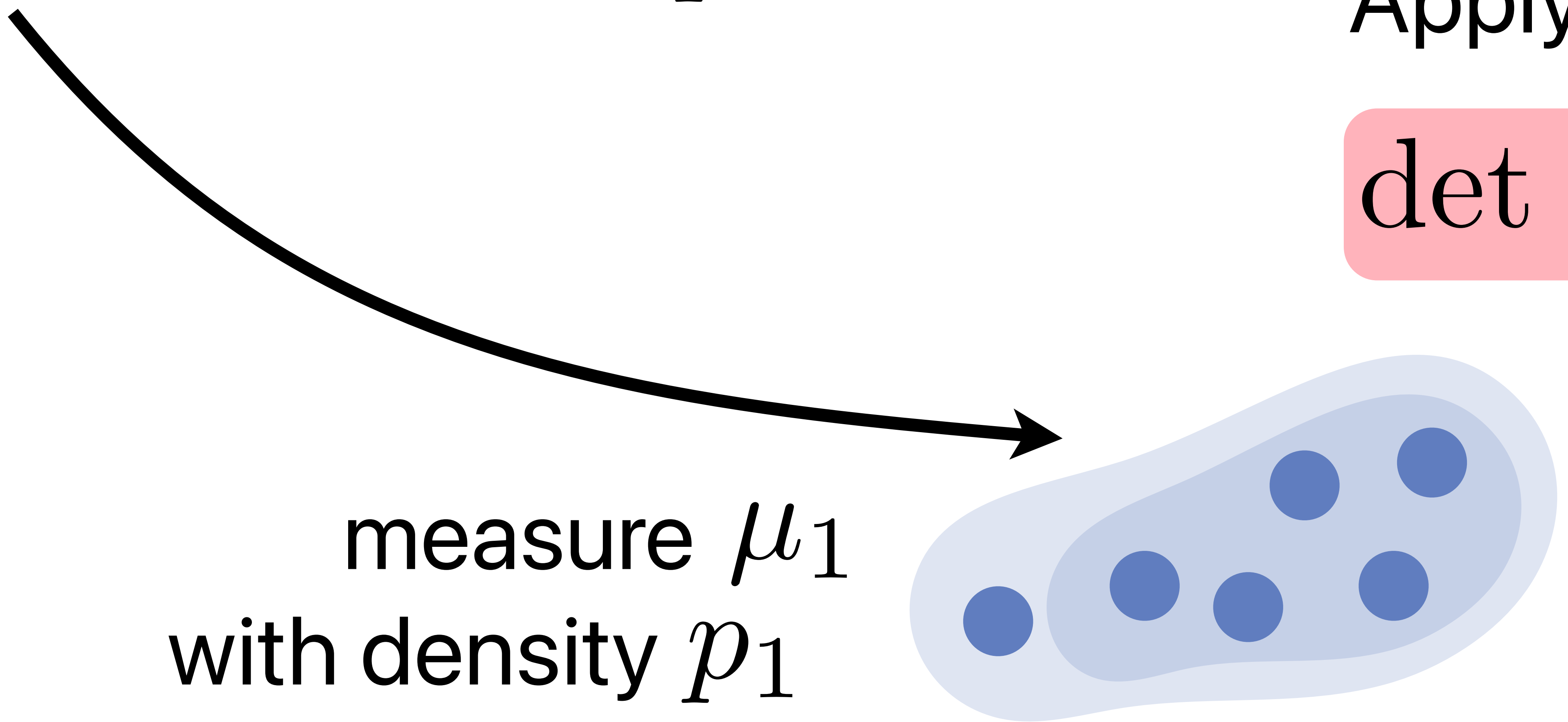
Brenier's Theorem $T(x) = \nabla\varphi(x) \longrightarrow \nabla\varphi_{\#}\mu_0 = \mu_1$



Monge-Ampère Equation

Brenier's Theorem $T(x) = \nabla\varphi(x) \longrightarrow \nabla\varphi_{\#}\mu_0 = \mu_1$





Applying the change of variable formula, we get

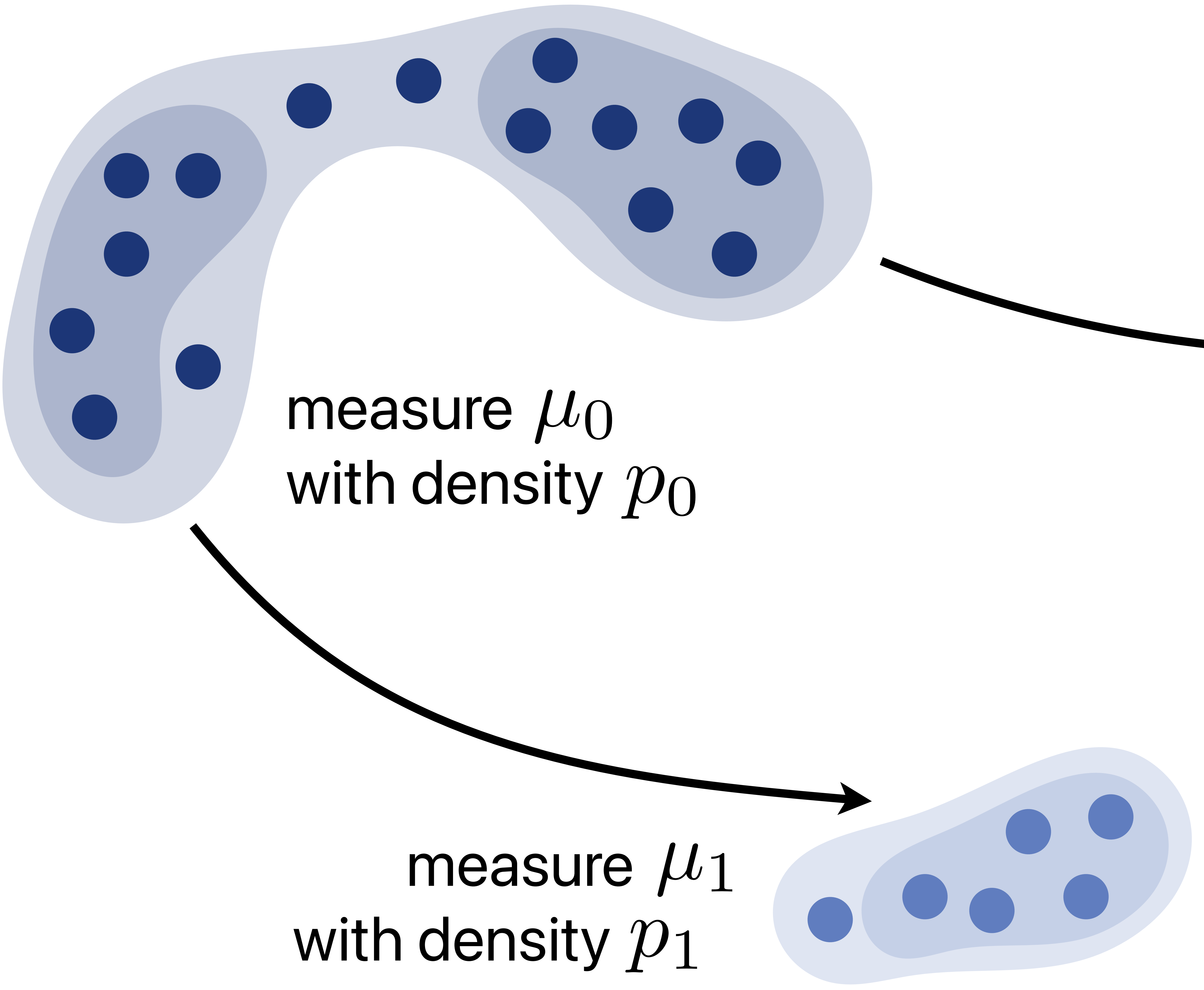
$$\det\left(\partial^2\varphi(x)\right) p_1(\nabla\varphi(x)) = p_0(x)$$

the nonlinear **Monge-Ampère PDE**

describes evolution from μ_0 to μ_1.

Monge-Ampère Equation

$\det\left(\partial^2\varphi(x)\right) p_1(\nabla\varphi(x)) = p_0(x)$ describes evolution from μ_0 to μ_1.




Abel Prize'23

Luis Caffarelli

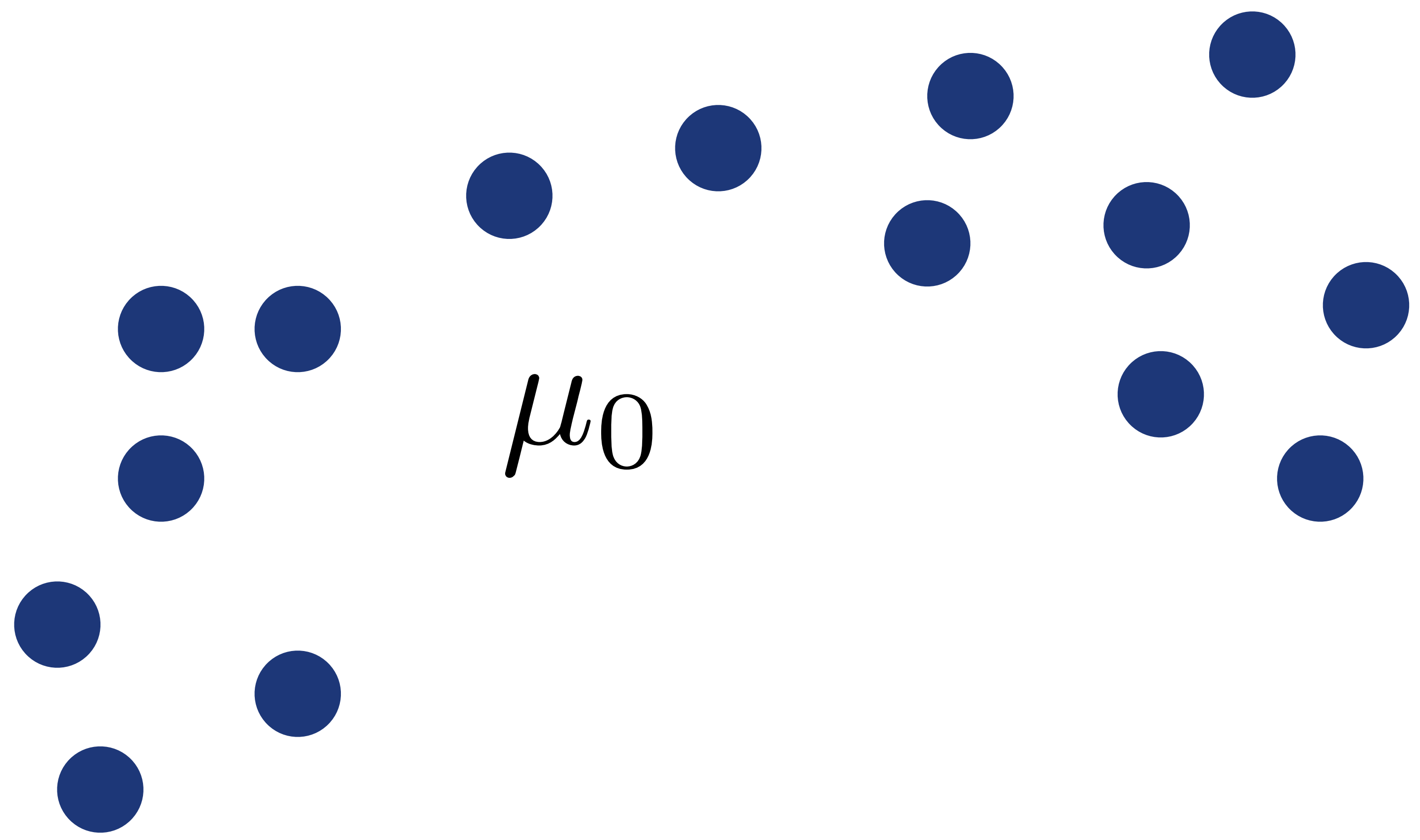
[Caffarelli90]

Fields Medal'18

Alessio Figalli

[Figalli17]

The Benamou-Brenier Formulation

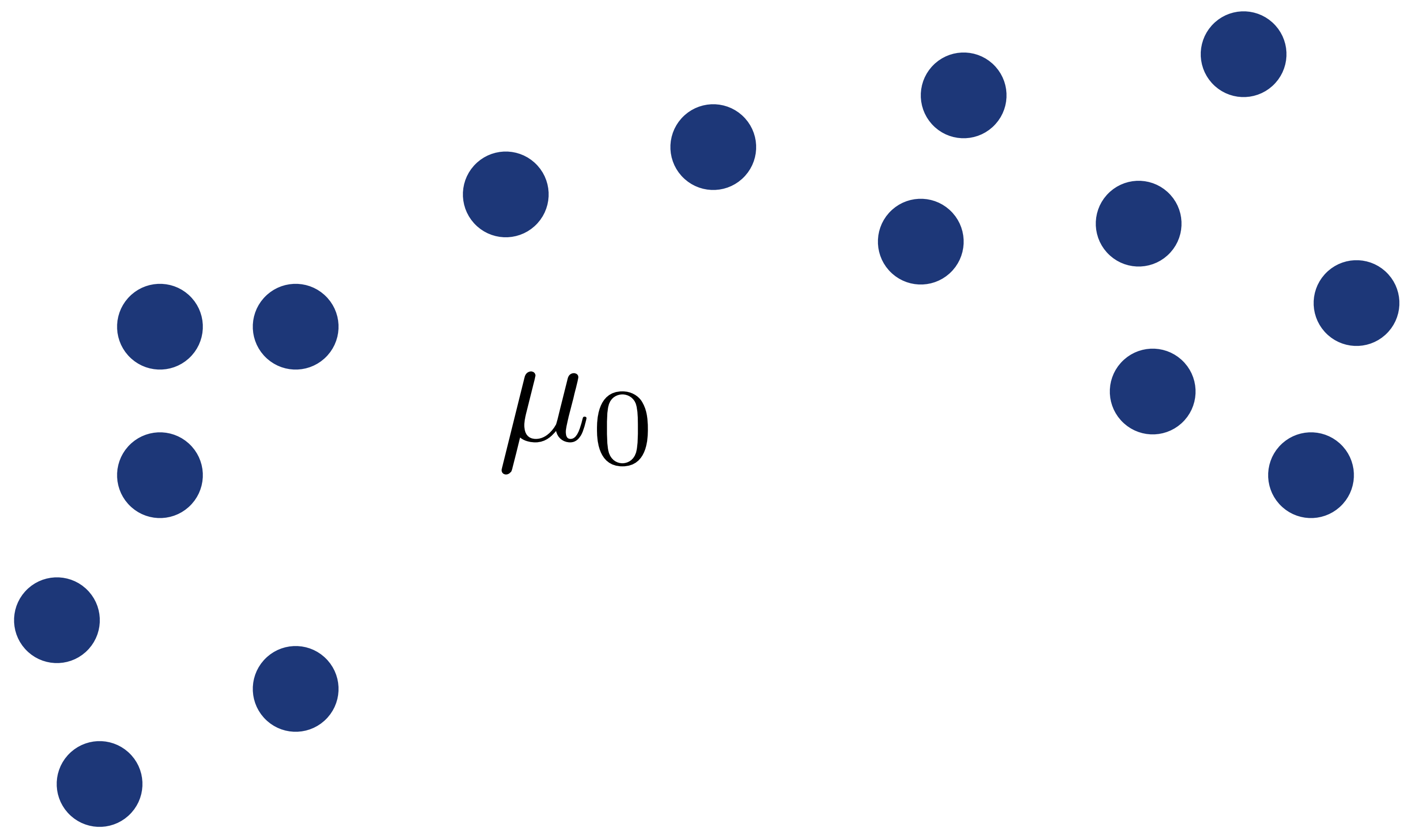


[Benamou+00]

The Benamou-Brenier Formulation

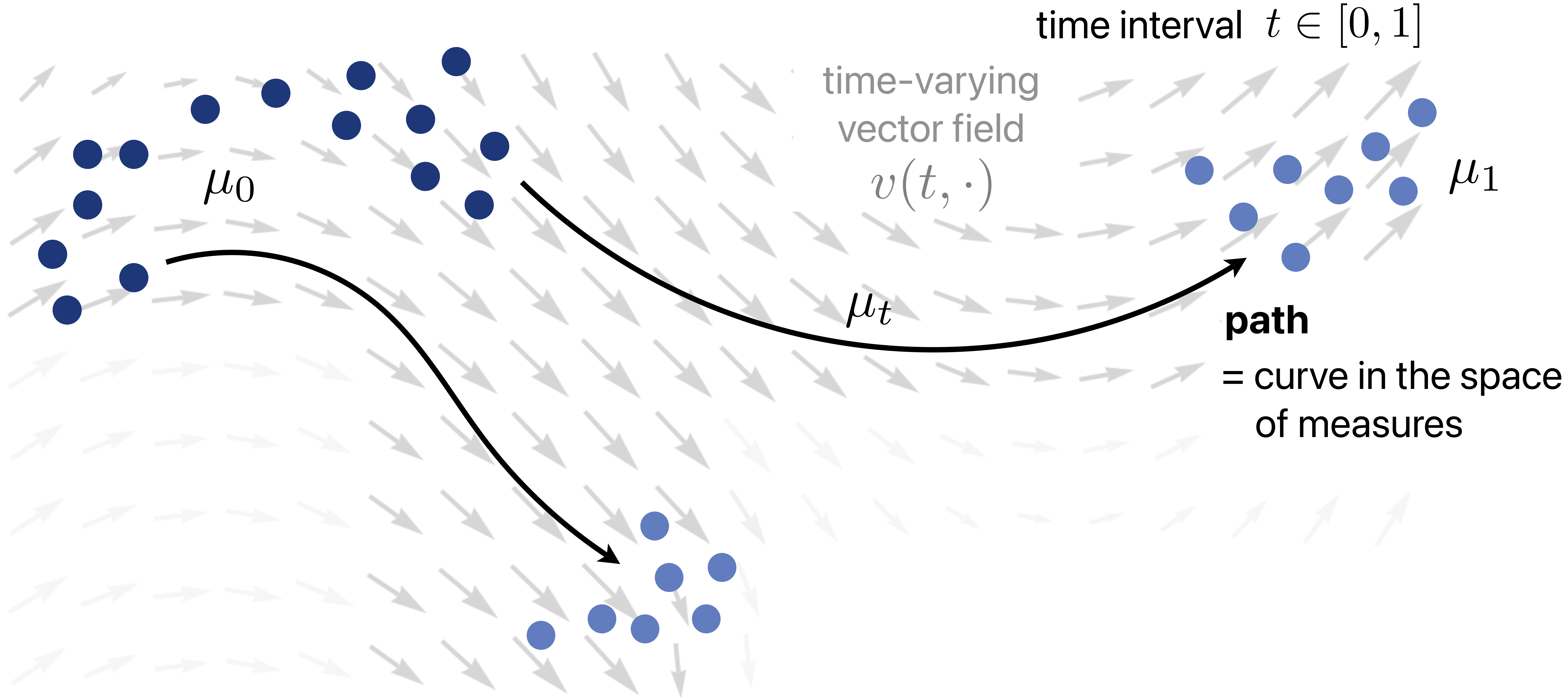
Part 3

time interval $t \in [0, 1]$



[Benamou+00]

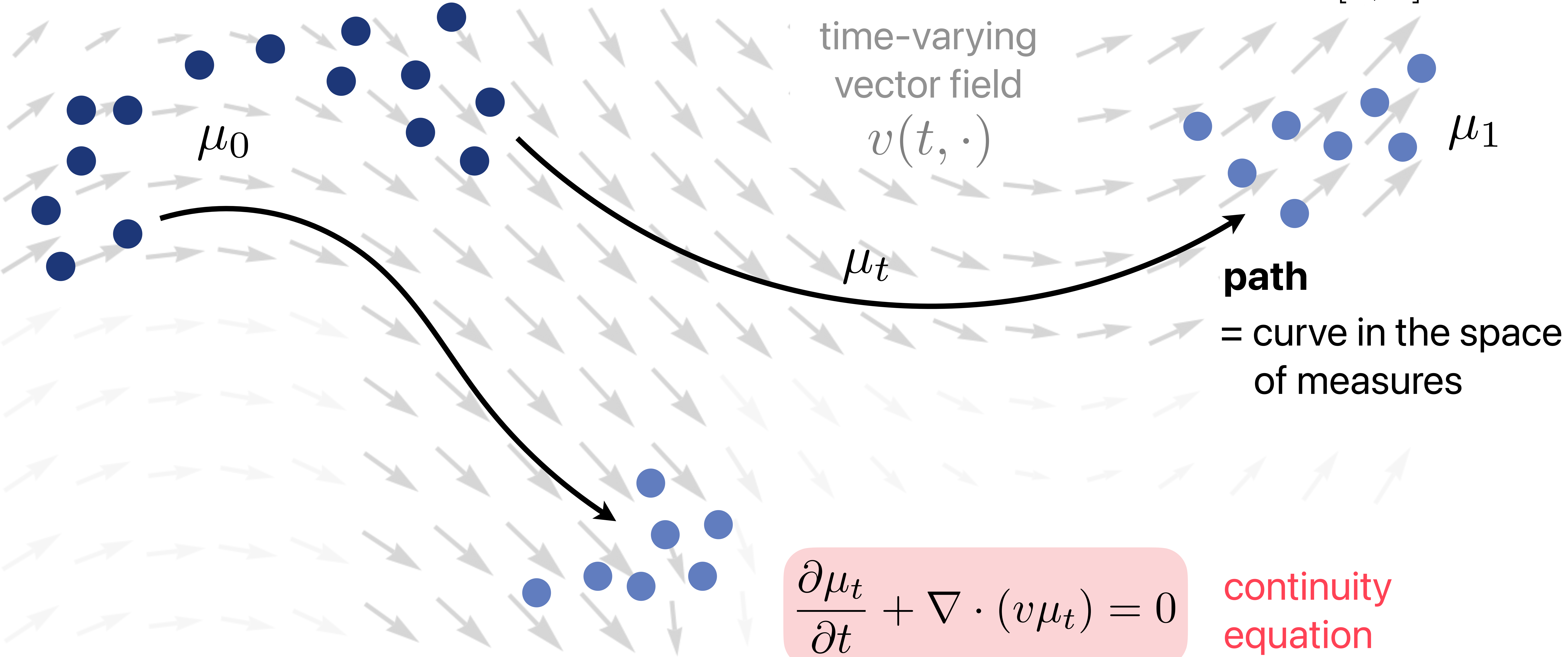
The Benamou-Brenier Formulation



[Benamou+00]

The Benamou-Brenier Formulation

time interval $t \in [0, 1]$



$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$

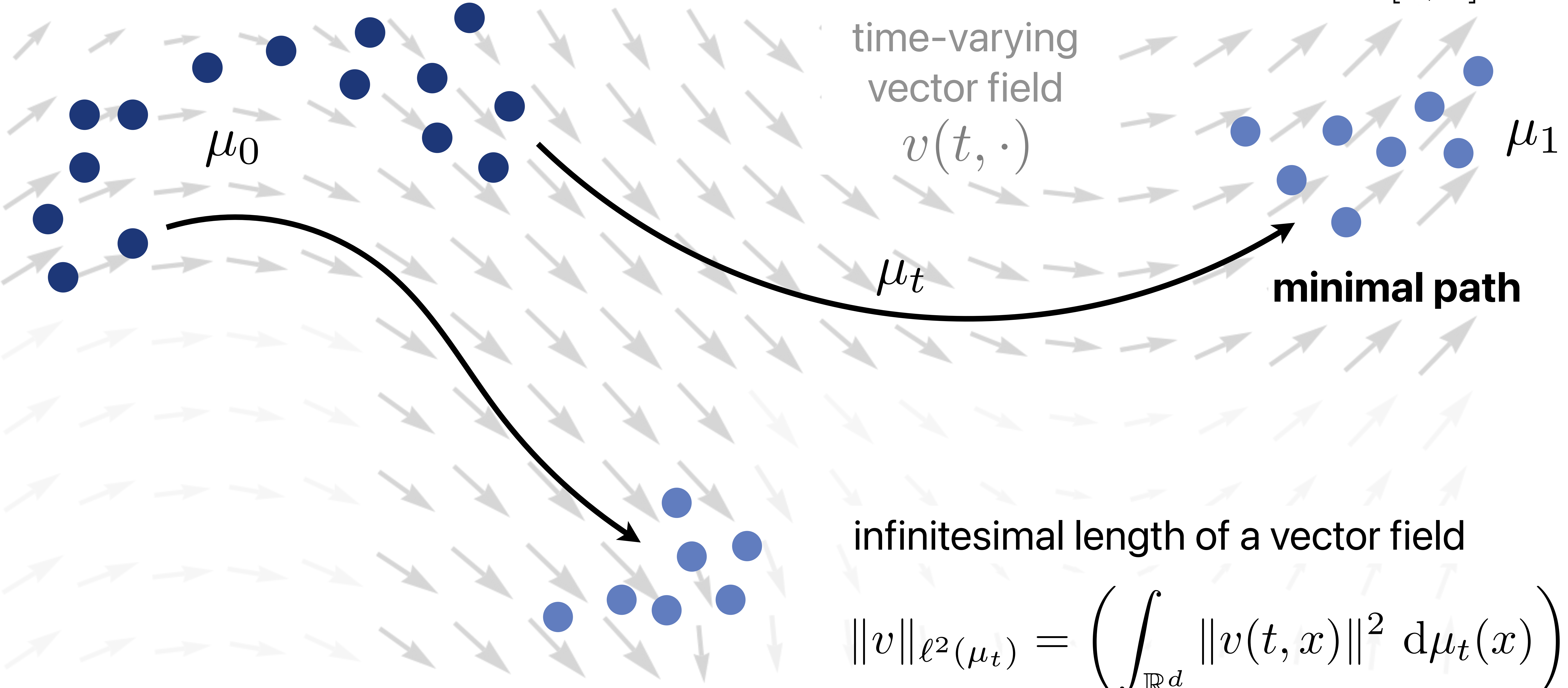
continuity equation

[Benamou+00]

conservation of mass $\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$

The Benamou-Brenier Formulation

time interval $t \in [0, 1]$



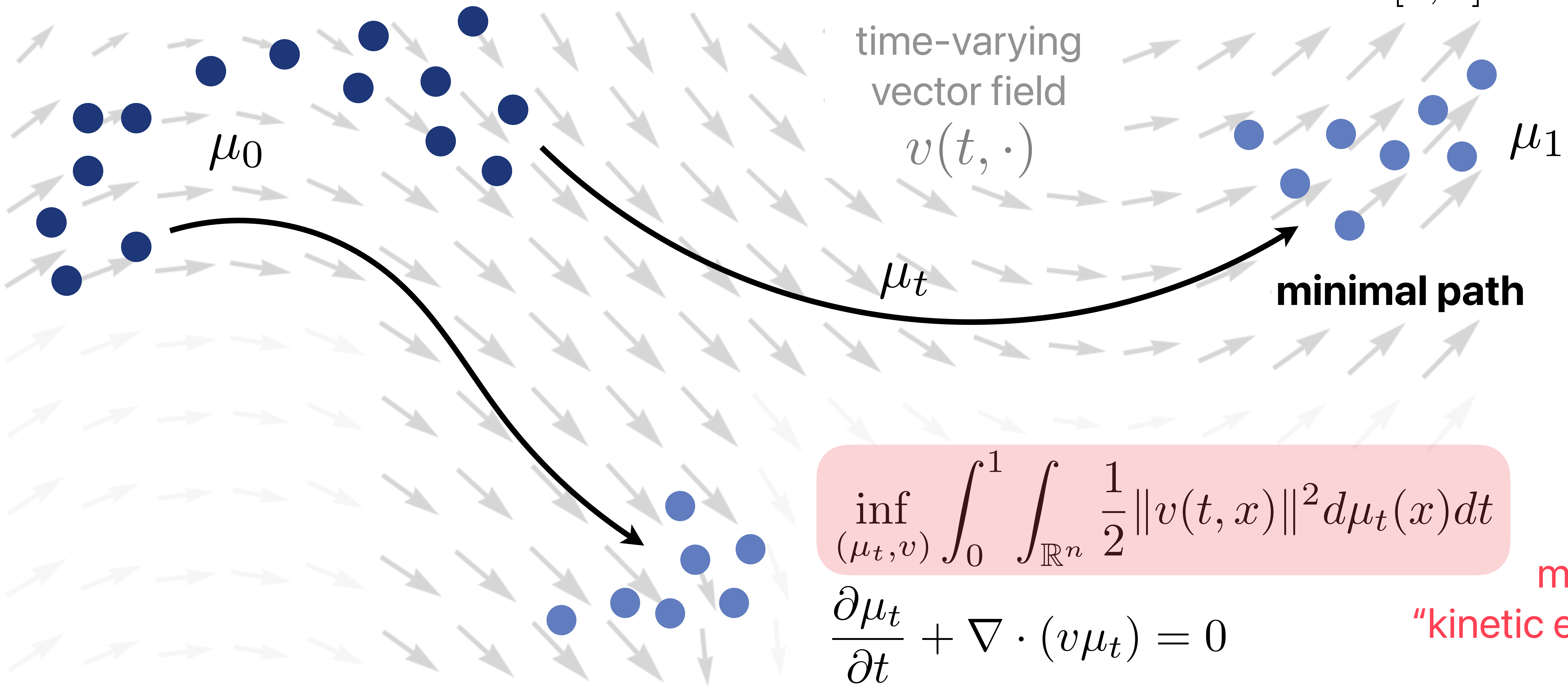
infinitesimal length of a vector field

$$\|v\|_{\ell^2(\mu_t)} = \left(\int_{\mathbb{R}^d} \|v(t, x)\|^2 \, \mathrm{d}\mu_t(x) \right)^{1/2}$$

[Benamou+00]

The Benamou-Brenier Formulation

time interval $t \in [0, 1]$



$$\inf_{(\mu_t, v)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|v(t, x)\|^2 d\mu_t(x) dt$$

minimal "kinetic energy"

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$

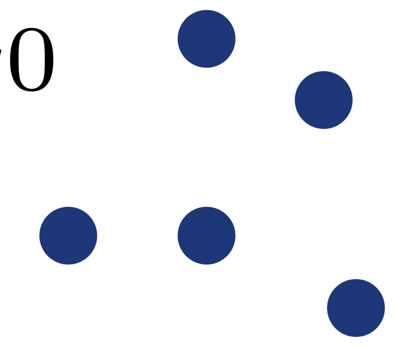
$$\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$$

[Benamou+00]

Connections to Flow Matching

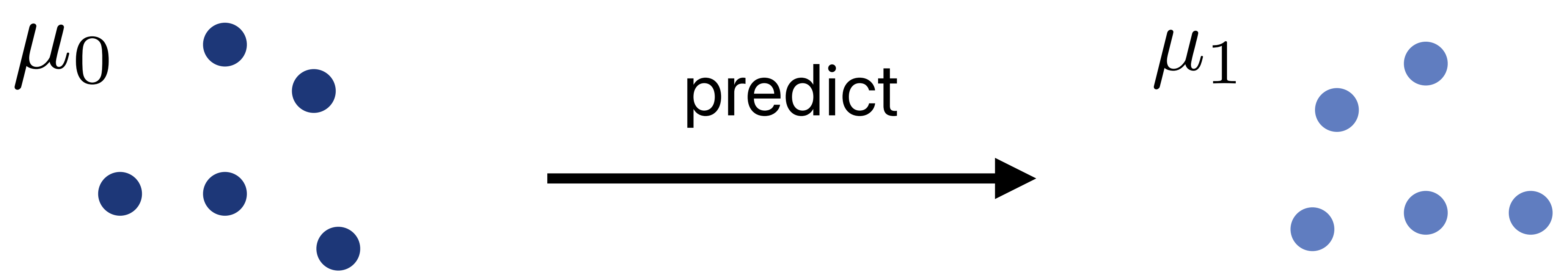
Learn vector field $v_\theta(t, x)$

μ_0



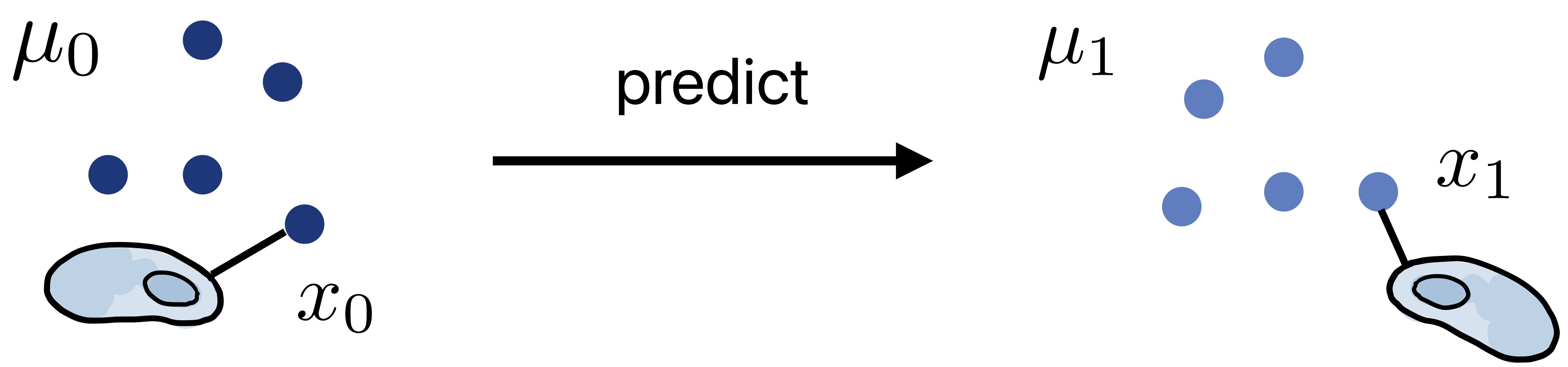
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



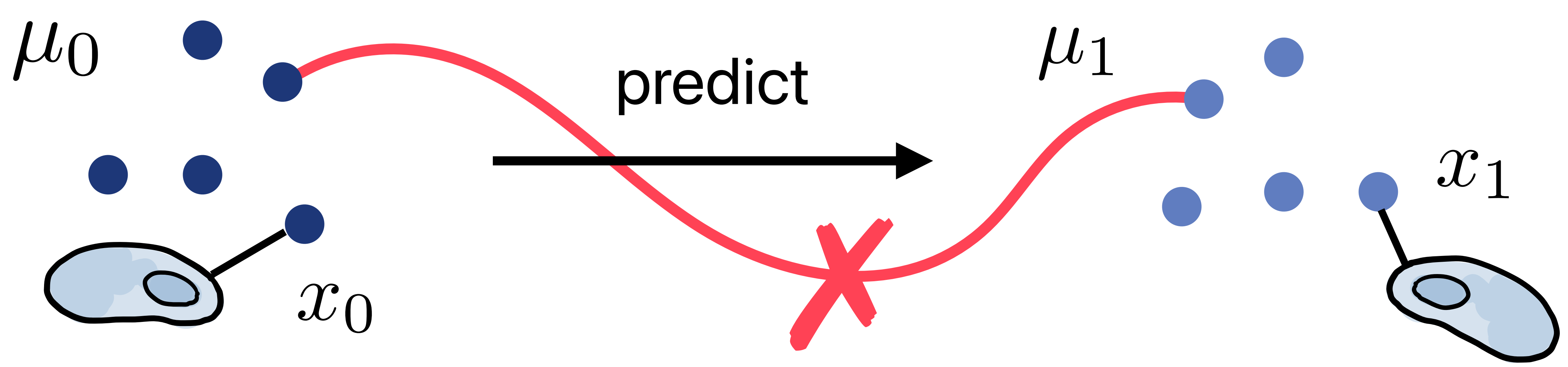
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



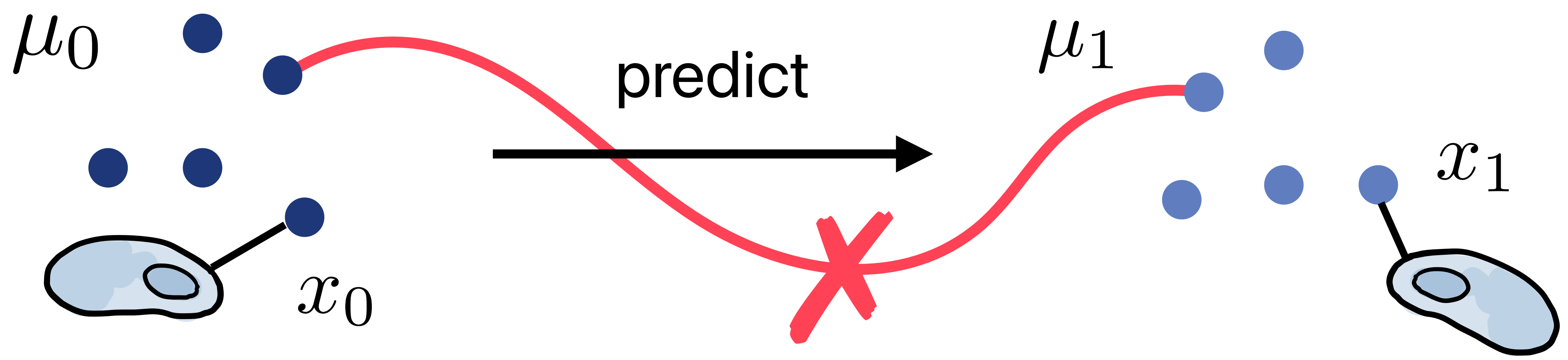
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Connections to Flow Matching

Learn vector field $v_\theta(t, x)$

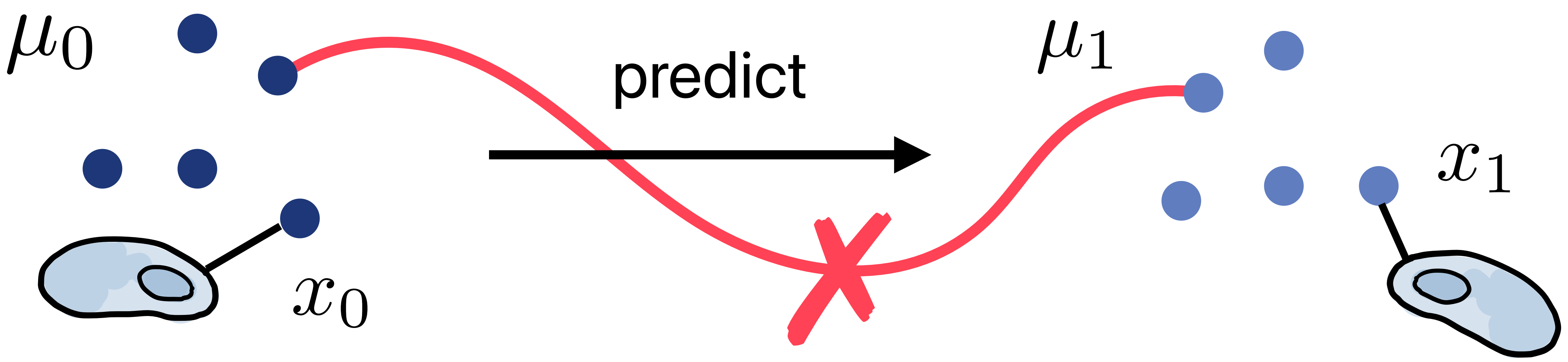


Continuous Normalizing Flows (CNF)

[Chen+23]

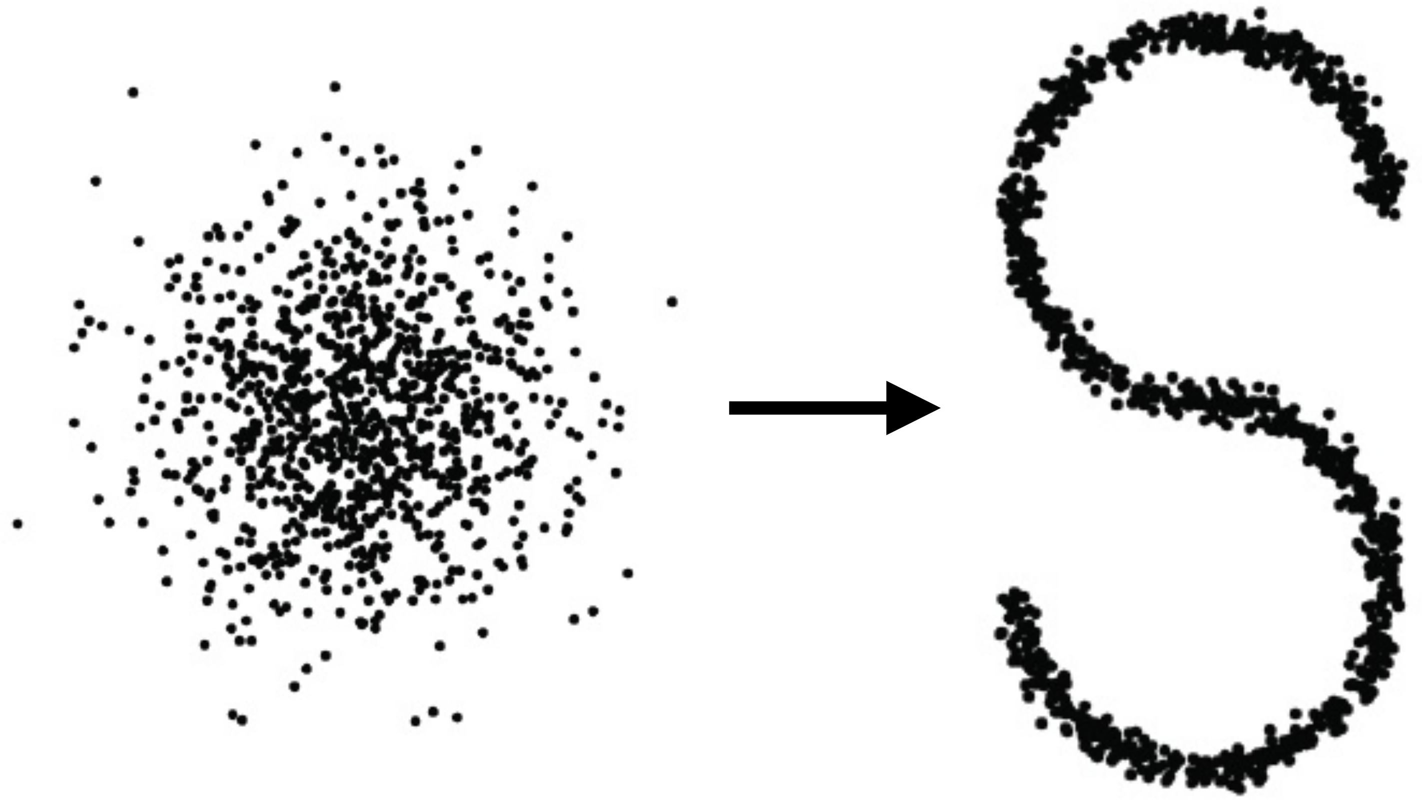
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Continuous Normalizing Flows (CNF)

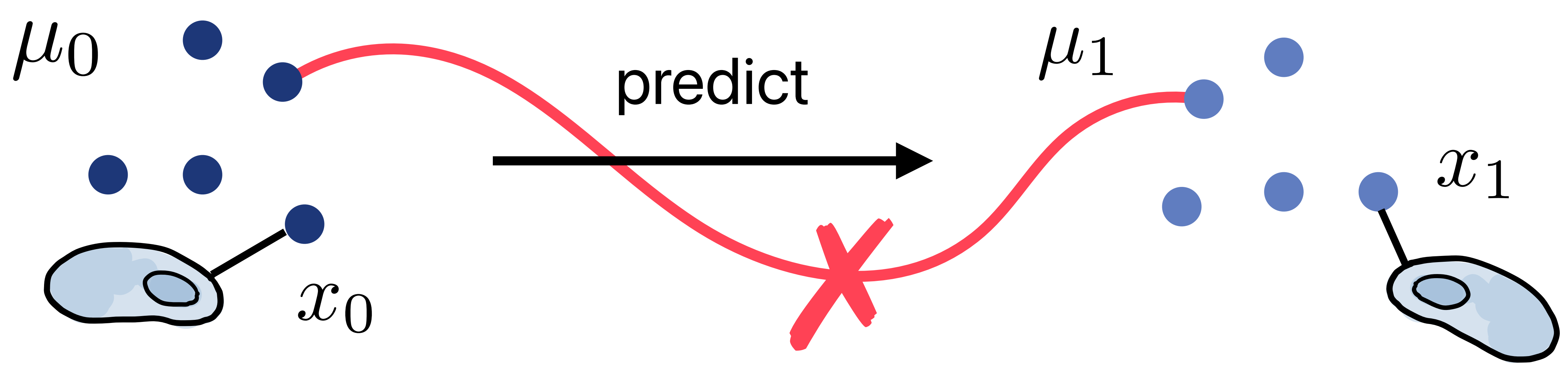
[Chen+23]



[Tong+23]

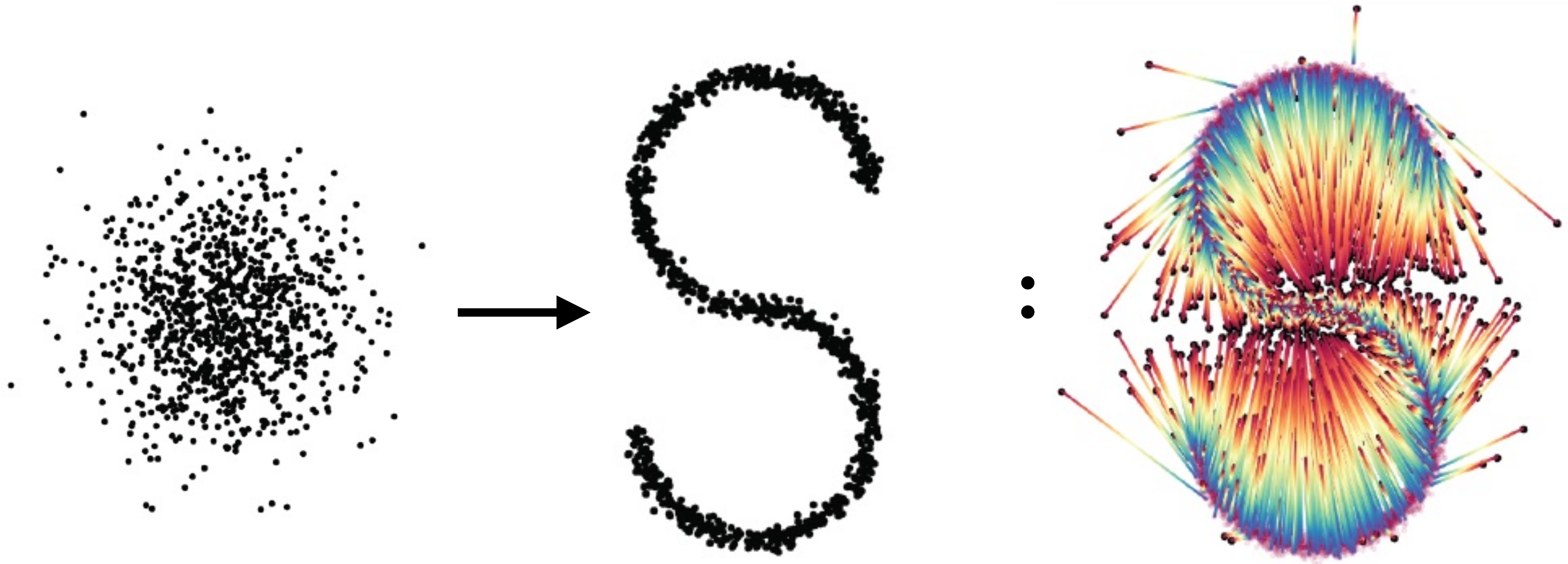
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Continuous Normalizing Flows (CNF)

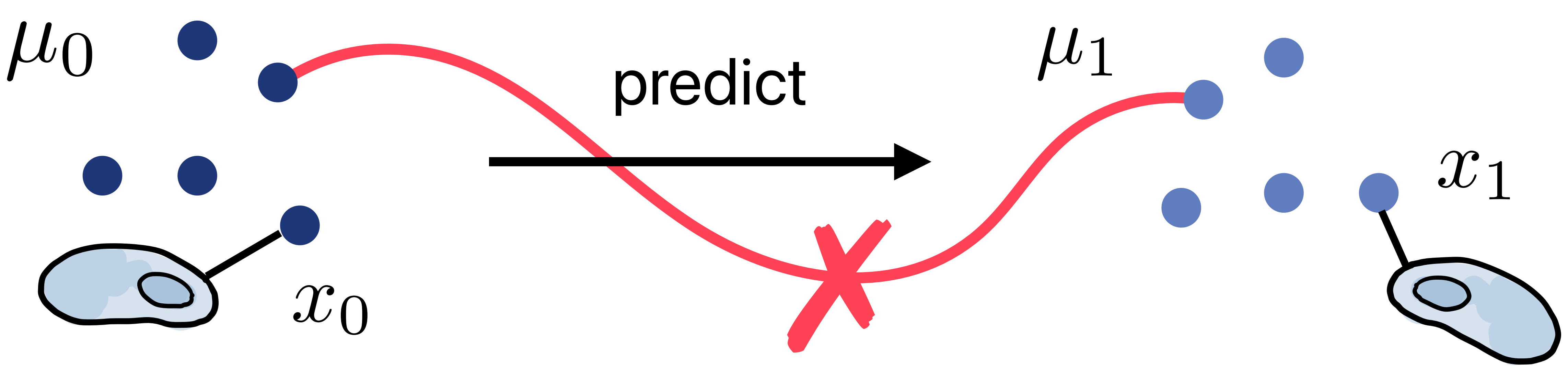
[Chen+23]



[Tong+23]

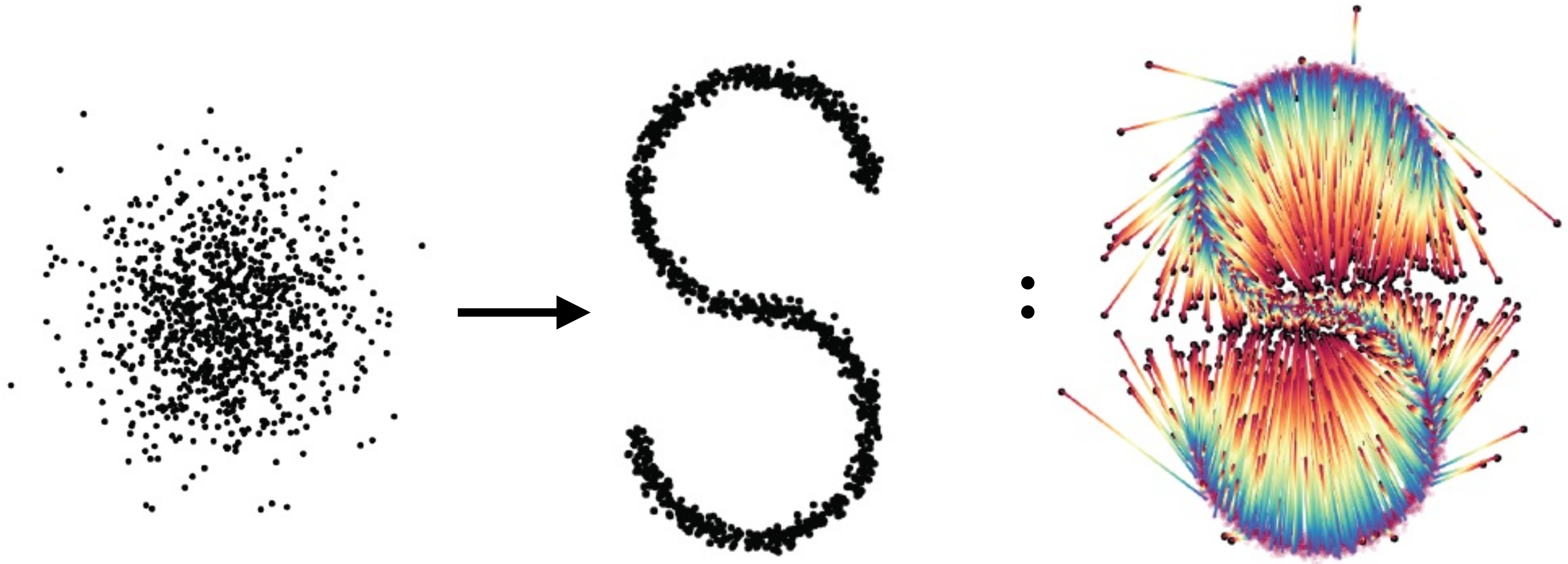
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Continuous Normalizing Flows (CNF)

[Chen+23]



[Tong+23]

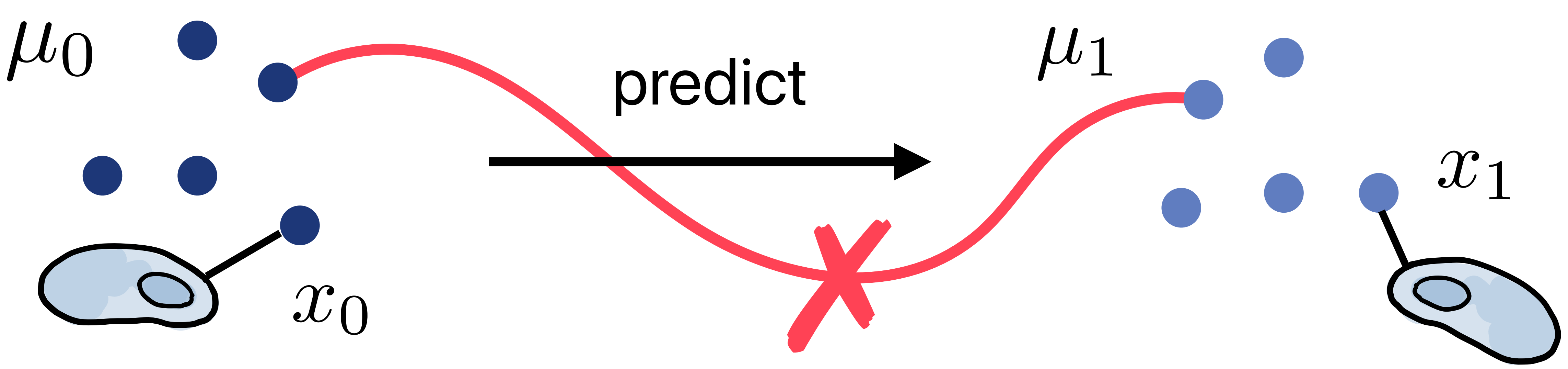
McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

[McCann97]

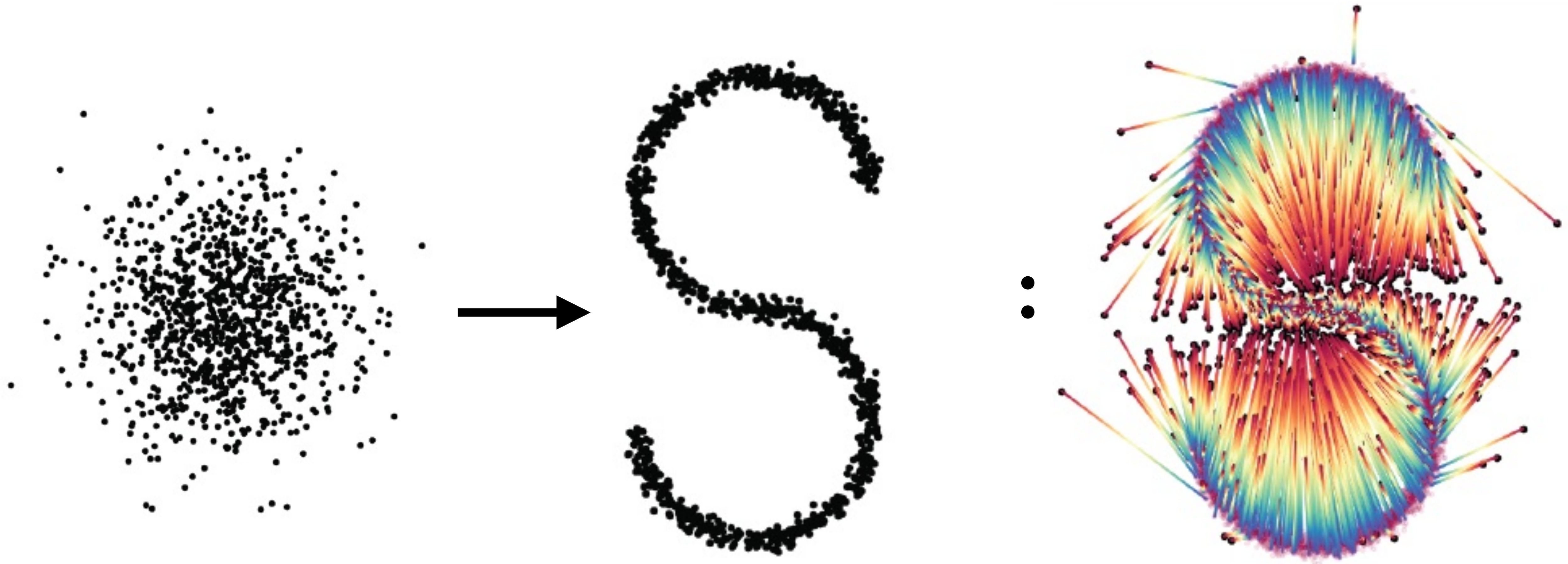
Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Continuous Normalizing Flows (CNF)

[Chen+23]



[Tong+23]

McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

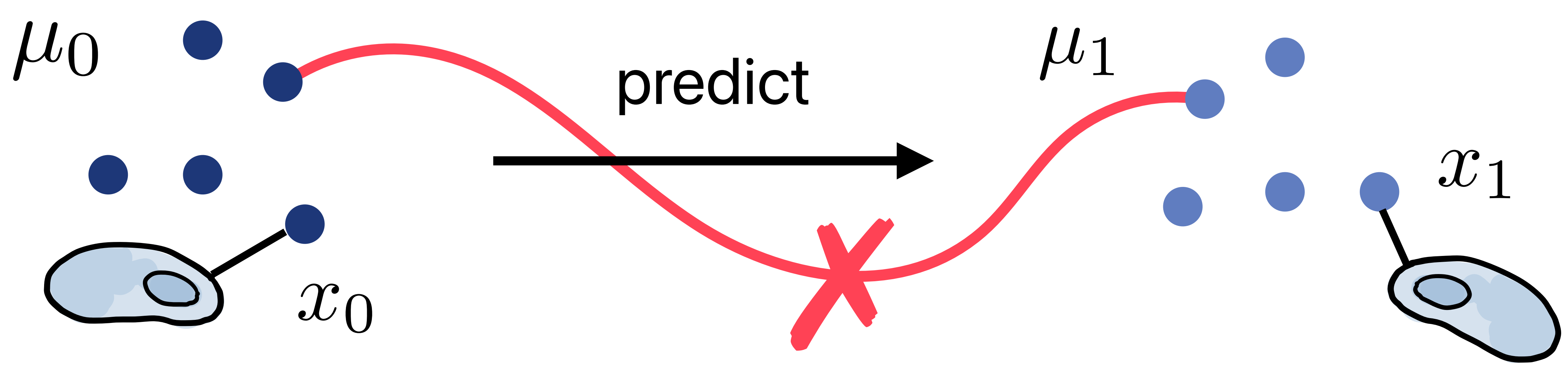
$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

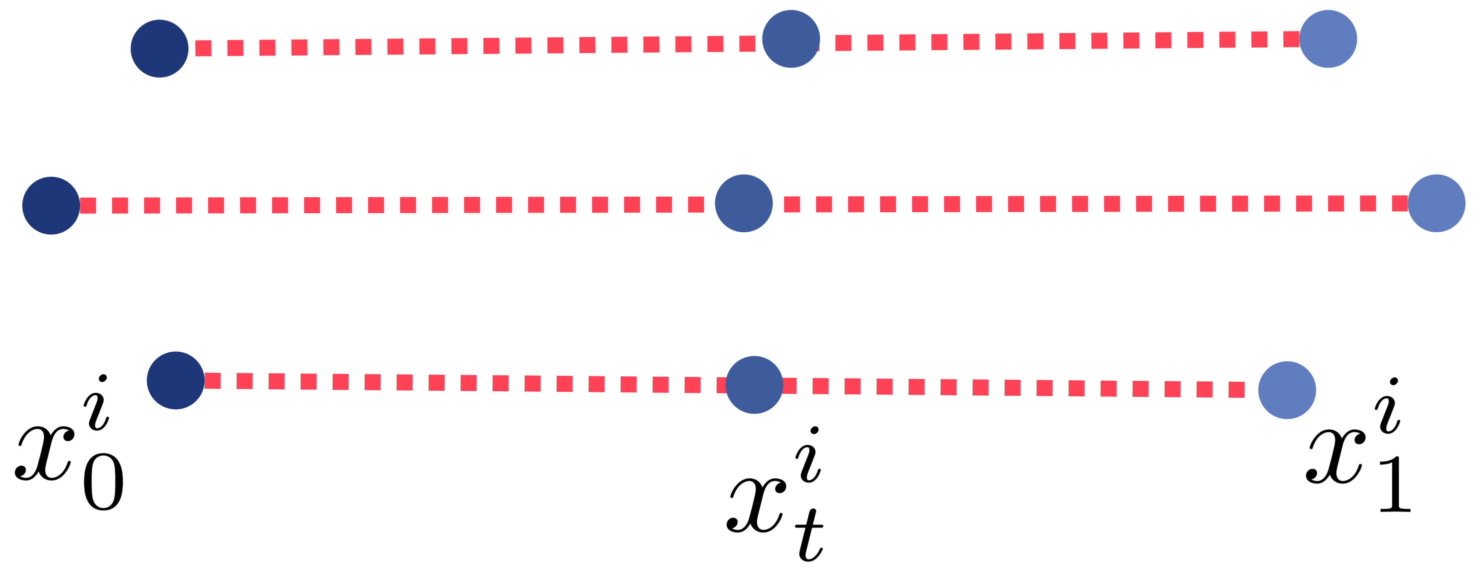
→ particles move in straight line trajectories and with constant speed

[McCann97]

Connections to Flow Matching

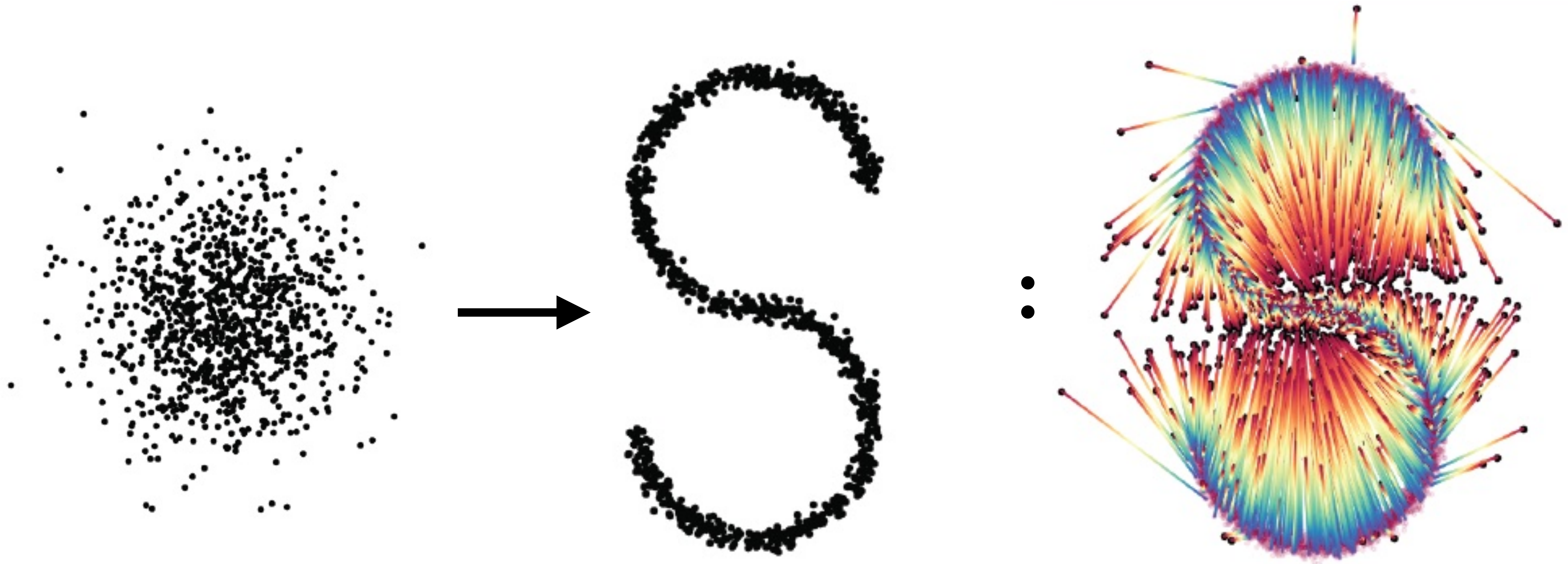
Learn vector field $v_\theta(t, x)$





Continuous Normalizing Flows (CNF)

[Chen+23]



[Tong+23]

McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

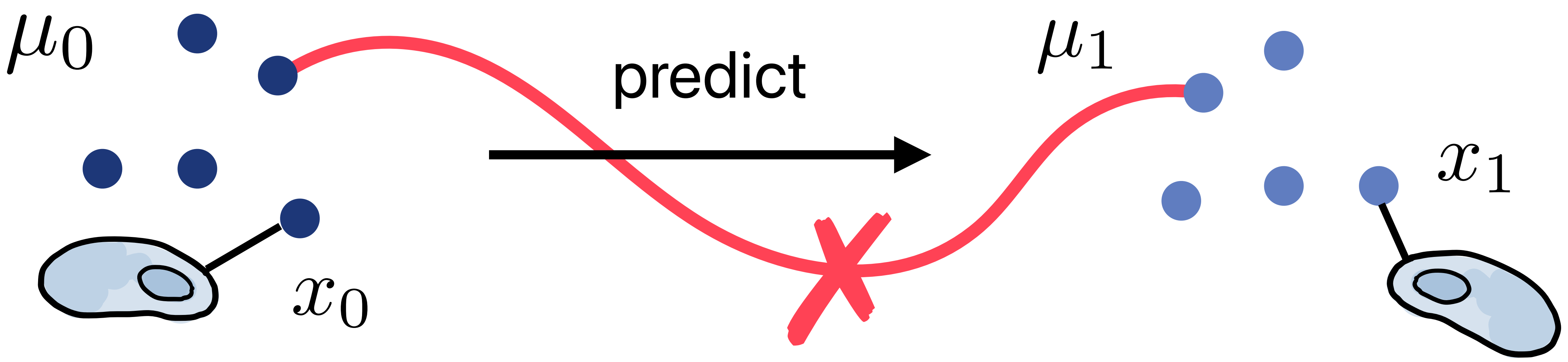
→ particles move in straight line trajectories and with constant speed

[McCann97]

Connections to Flow Matching

Part 3

Learn vector field $v_\theta(t, x)$



McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

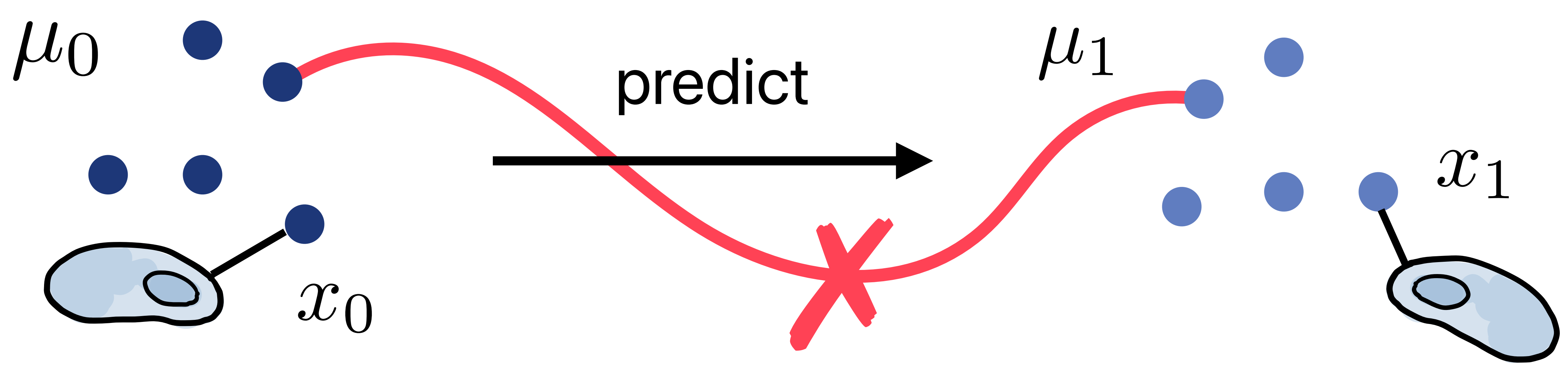
$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

→ particles move in straight line trajectories and with constant speed

[McCann97]

Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Conditional Flow Matching (CFM) = *regression*

$$\mathbb{E}_{t,\mu_1(x_1),\mu_t(x|x_1)} \left\| v_\theta(t, x) - \overset{?}{u_t}(x \mid x_1) \right\|^2$$

McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

→ particles move in straight line trajectories and with constant speed

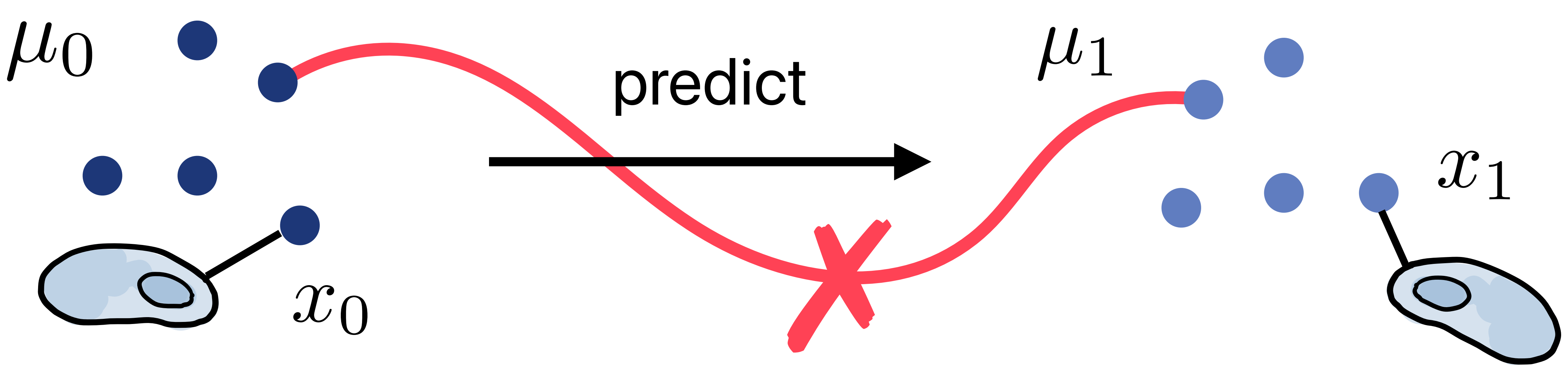
[McCann97]

[Albergo+23, Lipman+23, Pooladian+23, Liu+22]

Connections to Flow Matching

Part 3

Learn vector field $v_\theta(t, x)$



Conditional Flow Matching (CFM) = *regression*

$$\mathbb{E}_{t,\mu_1(x_1),\mu_t(x|x_1)} \|v_\theta(t,x) - u_t(x \mid x_1)\|^2$$

?

$$= \frac{x_1 - x}{1-t}$$

McCann's displacement interpolation

in Euclidean space and $c(x,y) = \|x-y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

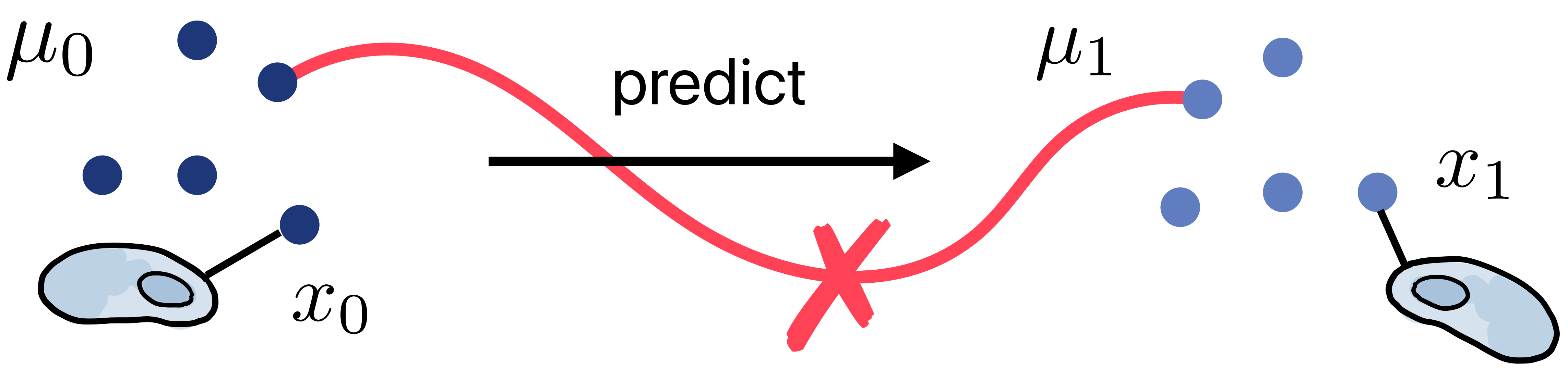
→ particles move in straight line trajectories and with constant speed

[McCann97]

[Albergo+23, Lipman+23, Pooladian+23, Liu+22]

Connections to Flow Matching

Learn vector field $v_\theta(t, x)$

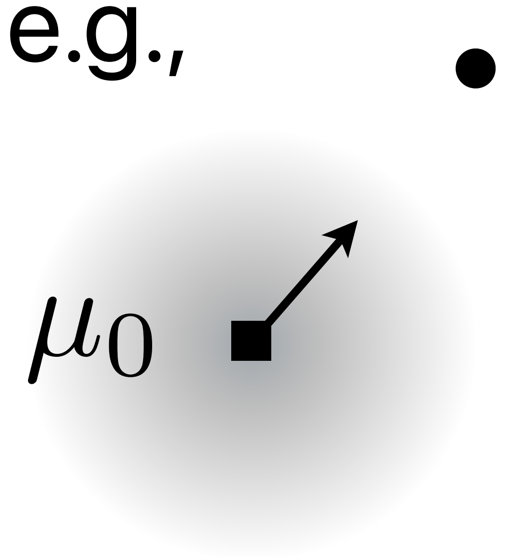


Conditional Flow Matching (CFM) *= regression*

$$\mathbb{E}_{t,\mu_1(x_1),\mu_t(x|x_1)} \|v_\theta(t, x) - \overset{?}{u_t}(x \mid x_1)\|^2$$

$$= \frac{x_1 - x}{1 - t}$$

e.g.,



[Albergo+23, Lipman+23, Pooladian+23, Liu+22]

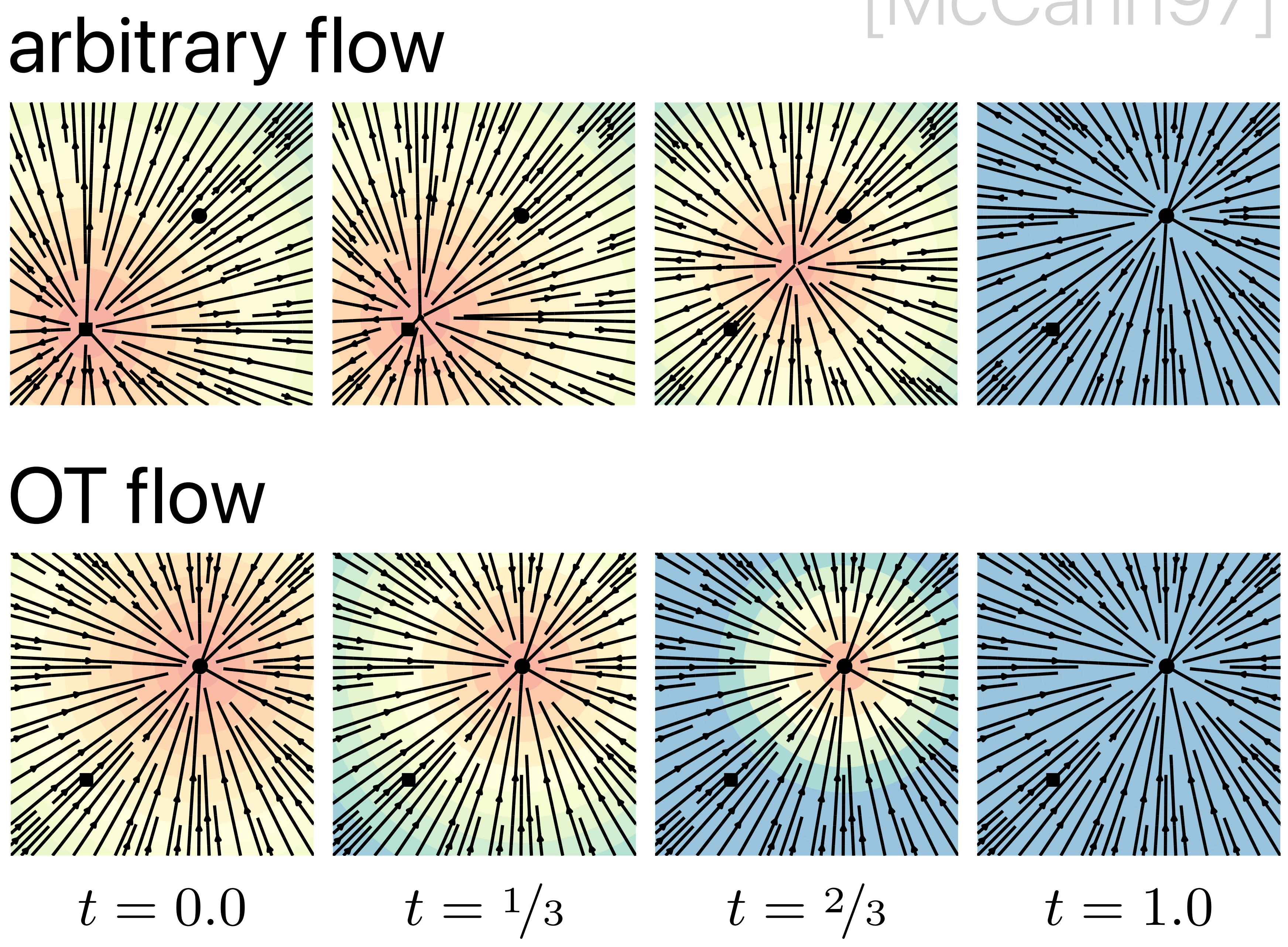
McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

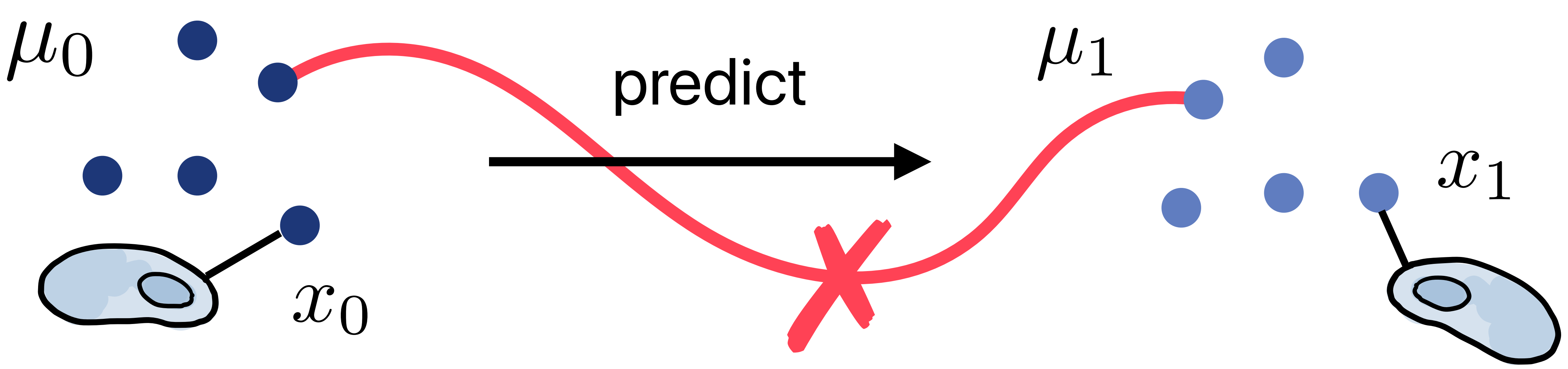
⟶ particles move in straight line trajectories and with constant speed

[McCann97]



Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Conditional Flow Matching (CFM) = *regression*

$$\mathbb{E}_{t,\mu_1(x_1),\mu_t(x|x_1)} \left\| v_\theta(t,x) - \overset{?}{u_t}(x \mid x_1) \right\|^2$$

$$= \frac{x_1 - x}{1-t}$$

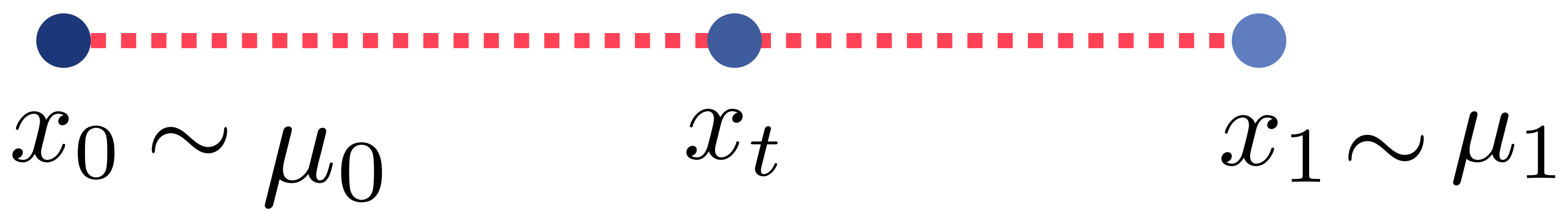
McCann's displacement interpolation

in Euclidean space
and $c(x,y) = \|x-y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

→ particles move in straight line trajectories and with constant speed

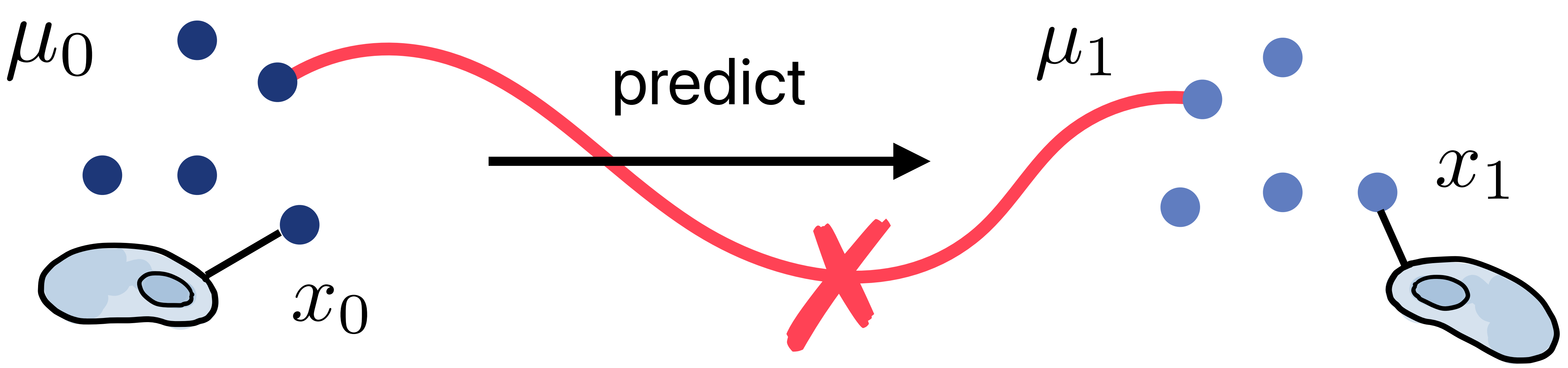
[McCann97]



[Albergo+23, Lipman+23, Pooladian+23, Liu+22]

Connections to Flow Matching

Learn vector field $v_\theta(t, x)$



Conditional Flow Matching (CFM) = *regression*

$$\mathbb{E}_{t,\mu_1(x_1),\mu_t(x|x_1)} \left\| v_\theta(t,x) - \overset{?}{u_t}(x \mid x_1) \right\|^2$$

$$= \frac{x_1 - x}{1 - t}$$

[Albergo+23, Lipman+23, Pooladian+23, Liu+22]

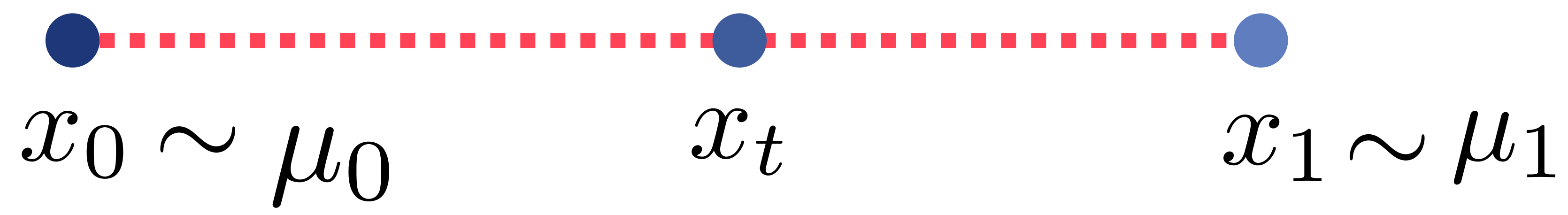
McCann's displacement interpolation

in Euclidean space
and $c(x,y) = \|x-y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

→ particles move in straight line trajectories and with constant speed

[McCann97]

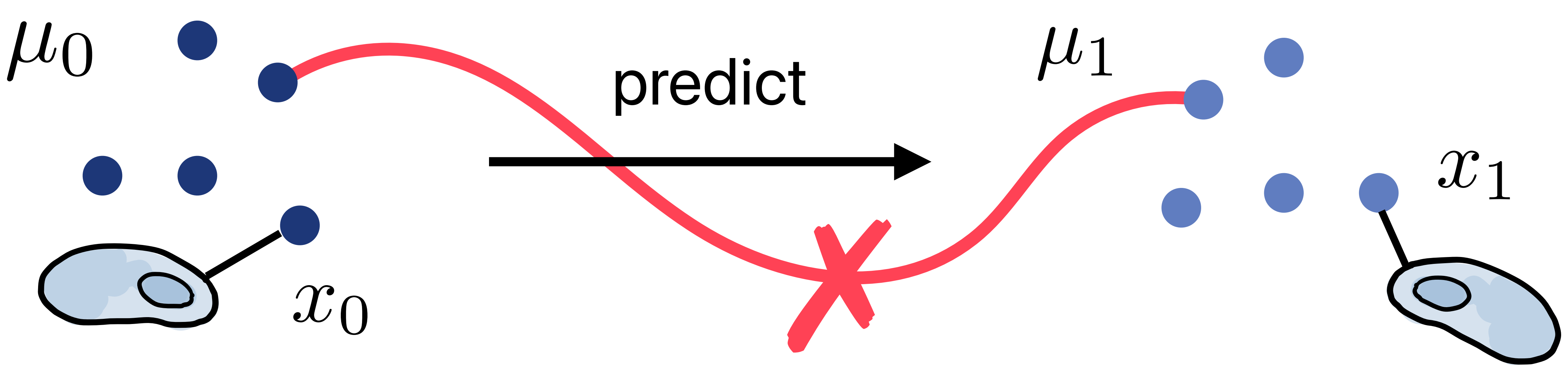


Instead, find optimal coupling

Connections to Flow Matching

Part 3

Learn vector field $v_\theta(t, x)$



Conditional Flow Matching (CFM) = *regression*

$$\mathbb{E}_{t,\mu_1(x_1),\mu_t(x|x_1)} \, \|v_\theta(t,x) - \overset{?}{u_t}(x \mid x_1)\|^2$$

$$= \frac{x_1 - x}{1 - t}$$

$(x_0, x_1) \sim \boldsymbol{P}^\star$ ←

[Albergo+23, Lipman+23, Pooladian+23, Liu+22]

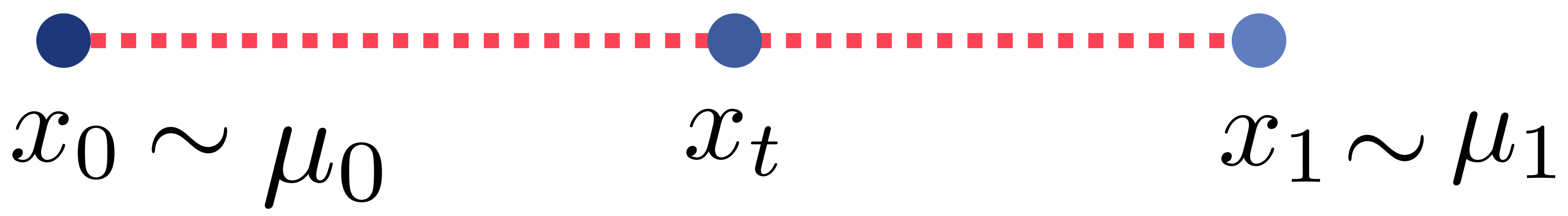
McCann's displacement interpolation

in Euclidean space
and $c(x, y) = \|x - y\|^2$:

$$\mu_t = [(1-t)I + tT]_\sharp \mu_0$$

→ particles move in straight line trajectories and with constant speed

[McCann97]

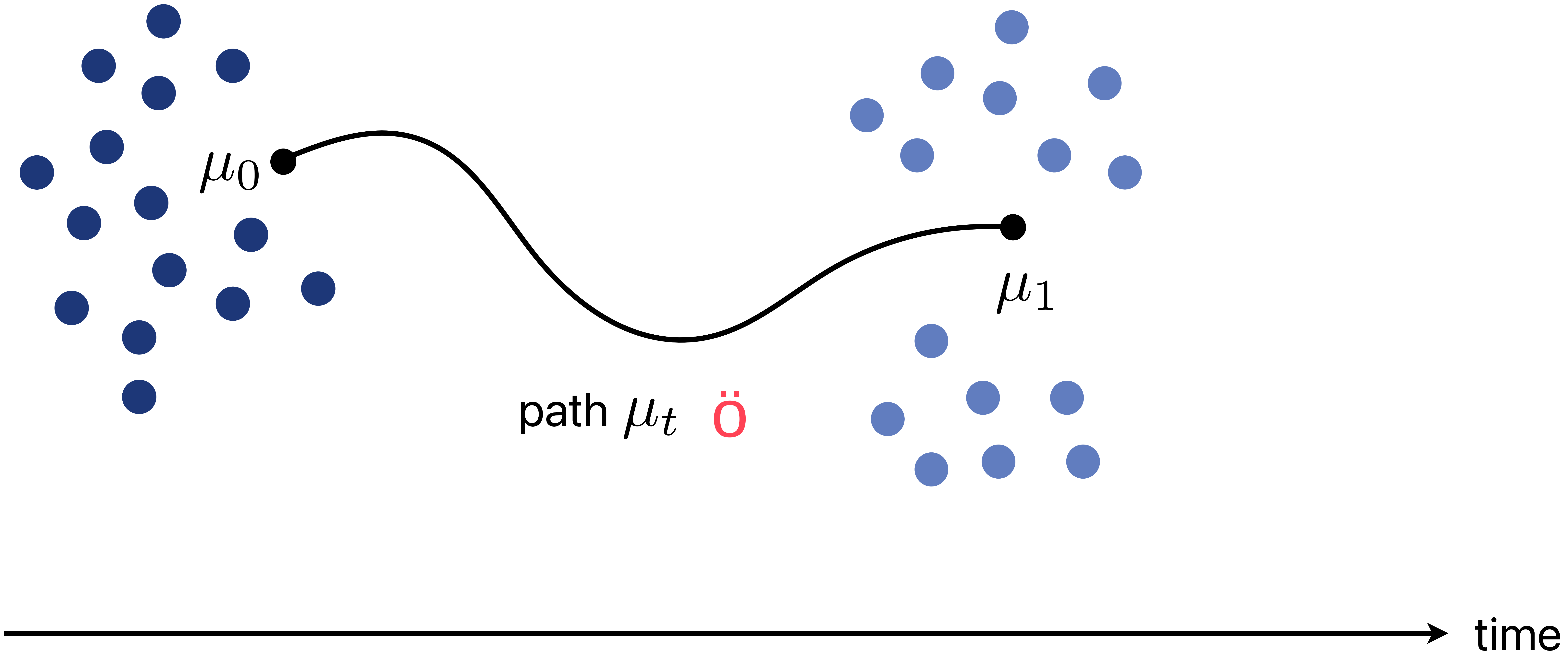


Instead, find optimal coupling

From Interpolation to Extrapolation

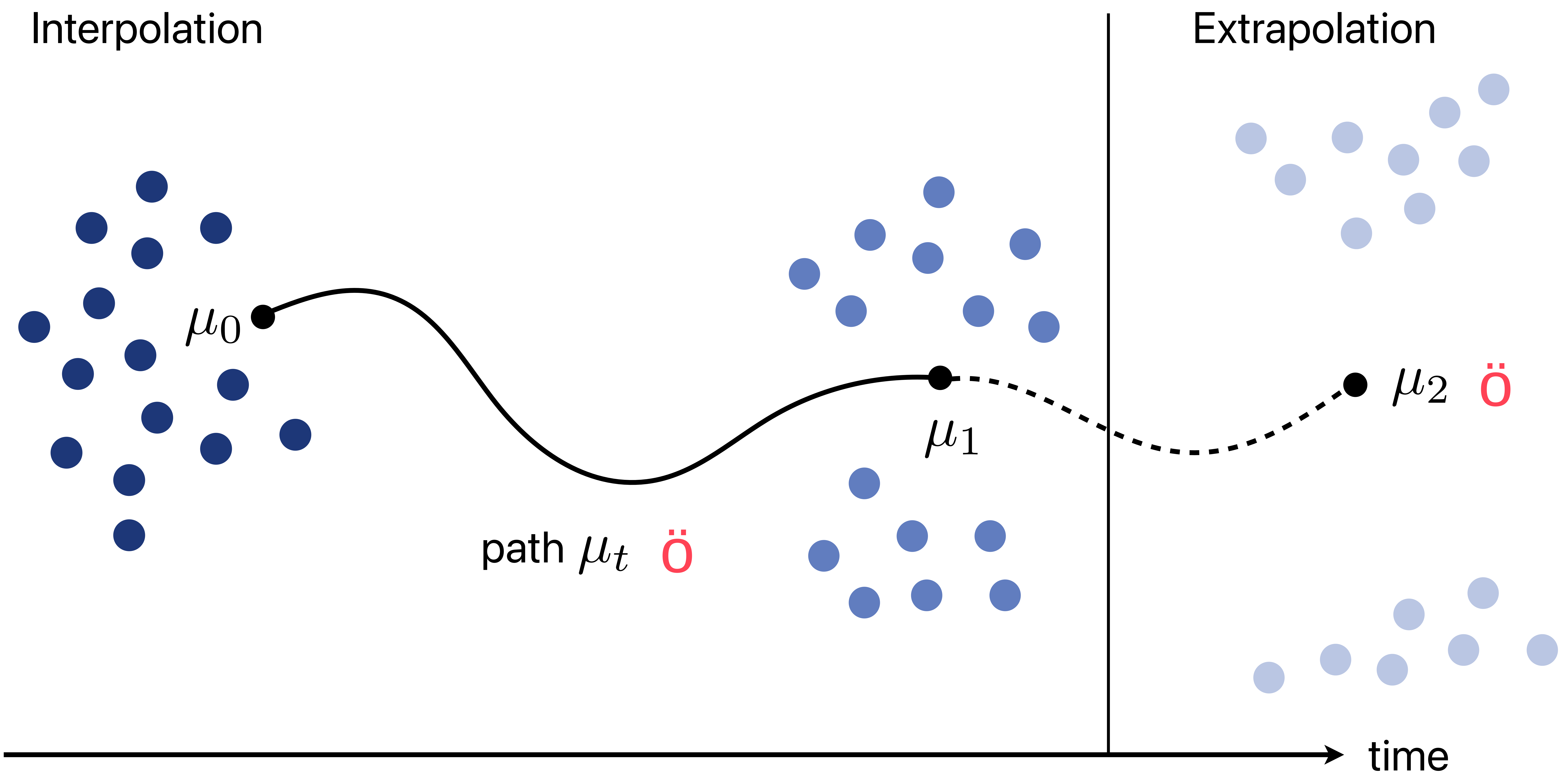
Part 3

Interpolation



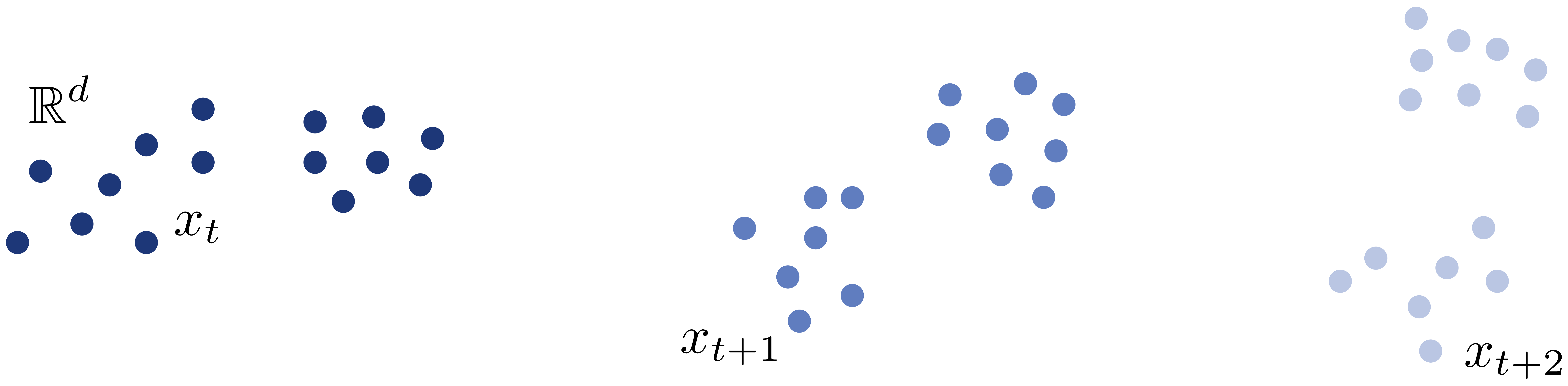
From Interpolation to Extrapolation

Part 3



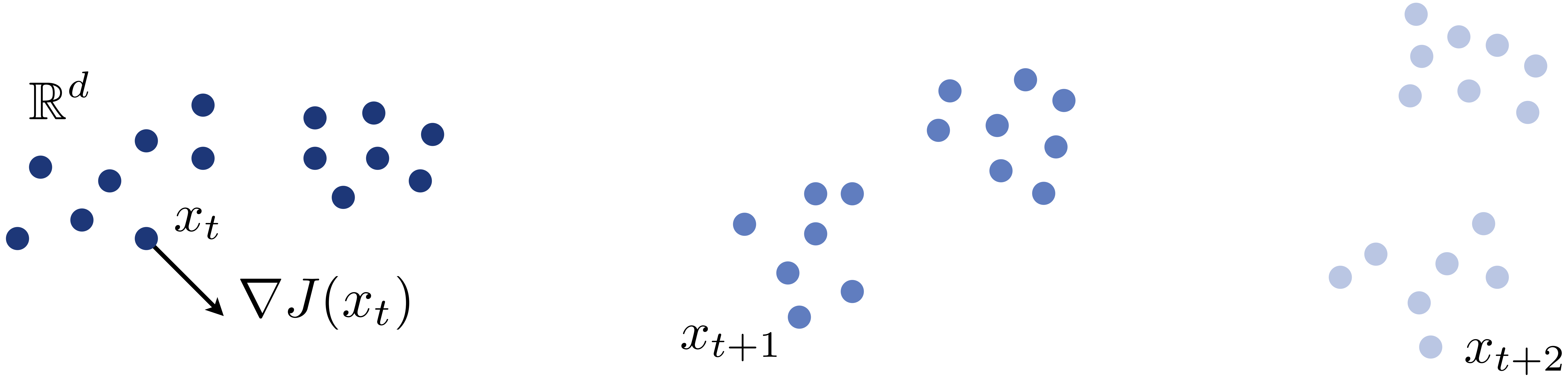
Population Dynamics in Euclidean Spaces

Dynamics of particles are described by minimizing a potential function $J(x)$



Population Dynamics in Euclidean Spaces

Dynamics of particles are described by minimizing a potential function $J(x)$

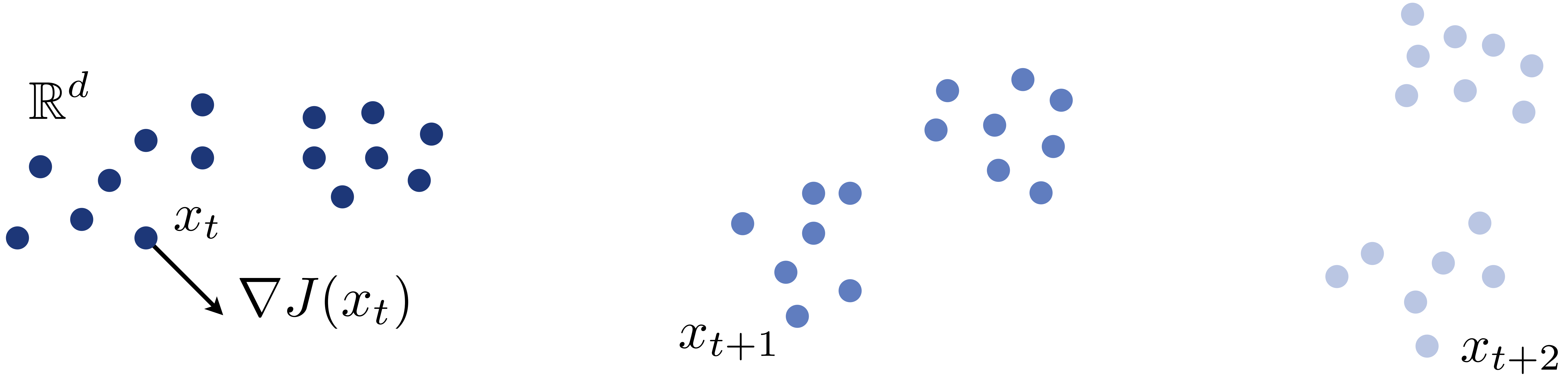


forward scheme $$x_{t+1} := x_t - \tau \nabla J(x_t)$$

can be *unstable* and requires *differentiability* of J

Population Dynamics in Euclidean Spaces

Dynamics of particles are described by minimizing a potential function $J(x)$



forward scheme $x_{t+1} := x_t - \tau \nabla J\left(x_t\right)$

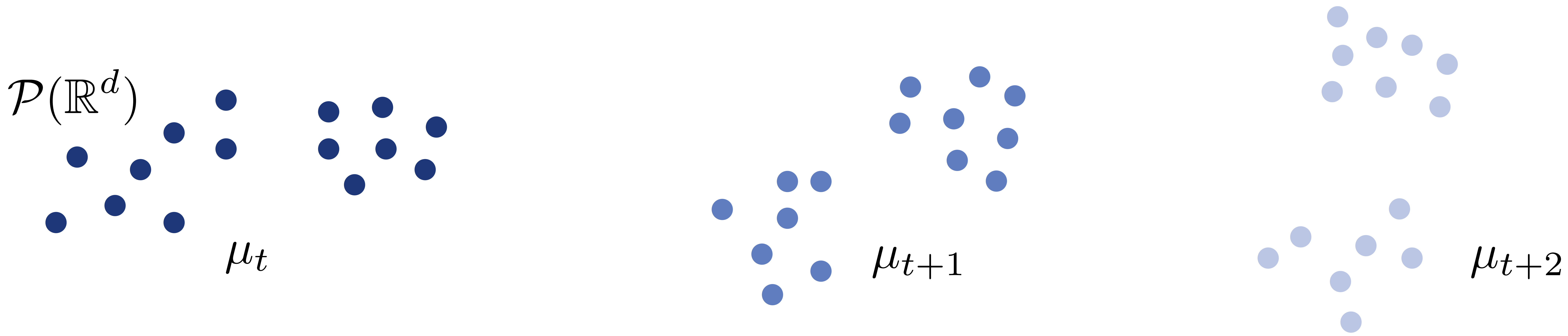
can be *unstable* and
requires *differentiability* of J

backward scheme $x_{t+1} := \underset{x}{\operatorname{argmin}} \frac{1}{2\tau} \left\| x - x_t \right\|^2 + J(x)$

proximity operator

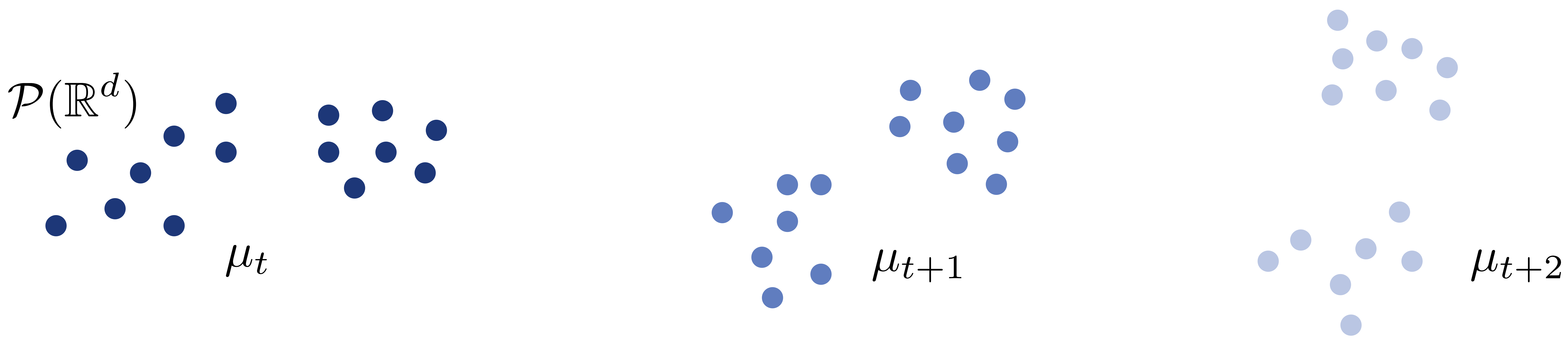
Population Dynamics in Wasserstein Spaces

Dynamics of particles are described by minimizing a potential function $J(\mu)$



Population Dynamics in Wasserstein Spaces

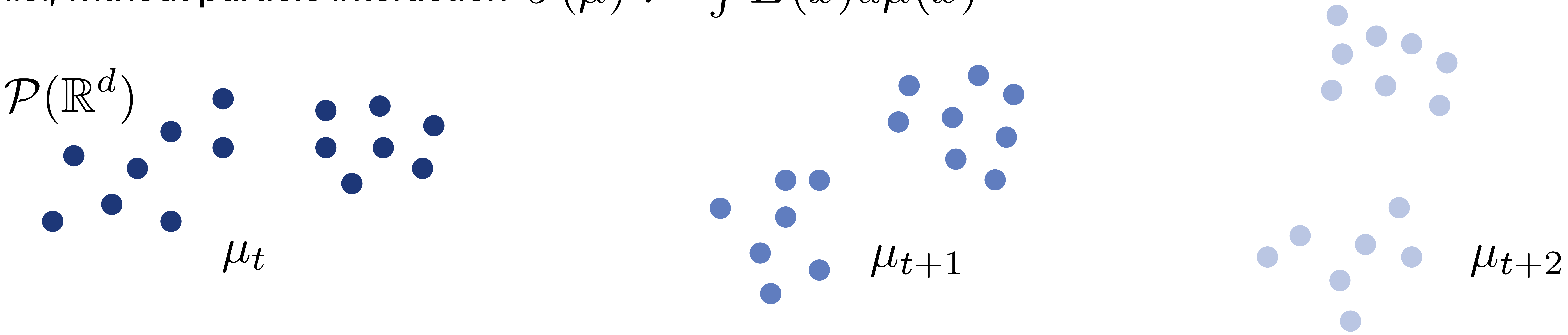
Dynamics of particles are described by minimizing a potential function $J(\mu)$ ← taking measures as input



Population Dynamics in Wasserstein Spaces

Dynamics of particles are described by minimizing a potential function $J(\mu)$ ← taking measures as input

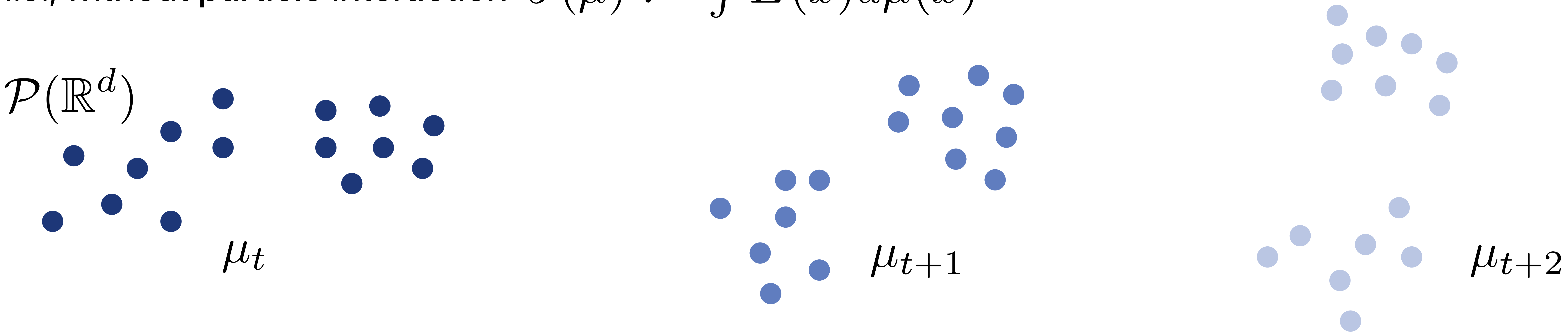
i.e., without particle interaction $J(\mu) := \int E(x) d\mu(x)$



Population Dynamics in Wasserstein Spaces

Dynamics of particles are described by minimizing a potential function $J(\mu)$ ← taking measures as input

i.e., without particle interaction $J(\mu) := \int E(x) d\mu(x)$

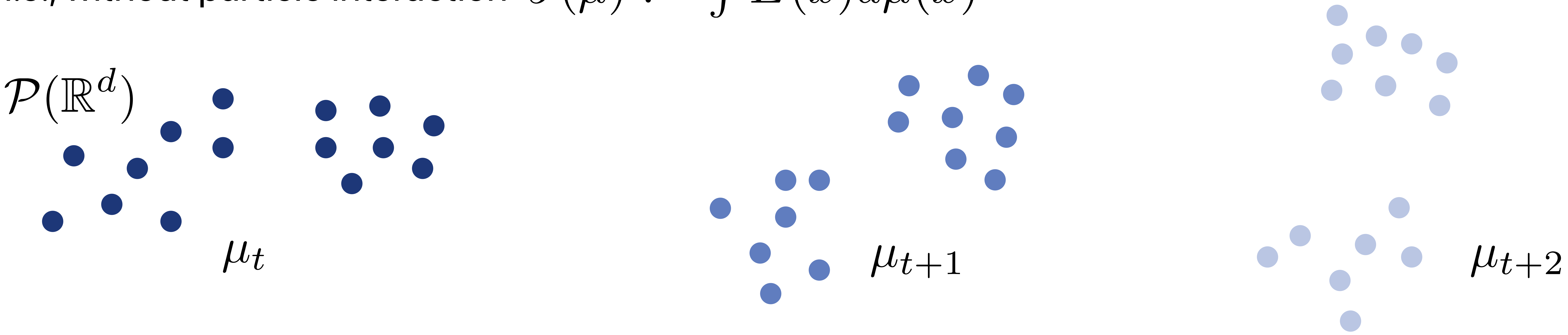


forward scheme $\mu_{t+1} := (I - \tau \nabla E)_{\#} \mu_t$

Population Dynamics in Wasserstein Spaces

Dynamics of particles are described by minimizing a potential function $J(\mu)$ ← taking measures as input

i.e., without particle interaction $J(\mu) := \int E(x) d\mu(x)$



forward scheme $\mu_{t+1} := (I - \tau \nabla E)_{\#} \mu_t$

backward scheme $\mu_{t+1} := \underset{\mu}{\operatorname{argmin}} \; \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$

proximity operator

allows for general and more complex J

[Jordan+98]

Jordan-Kinderlehrer-Otto (JKO) Flows


Richard Jordan

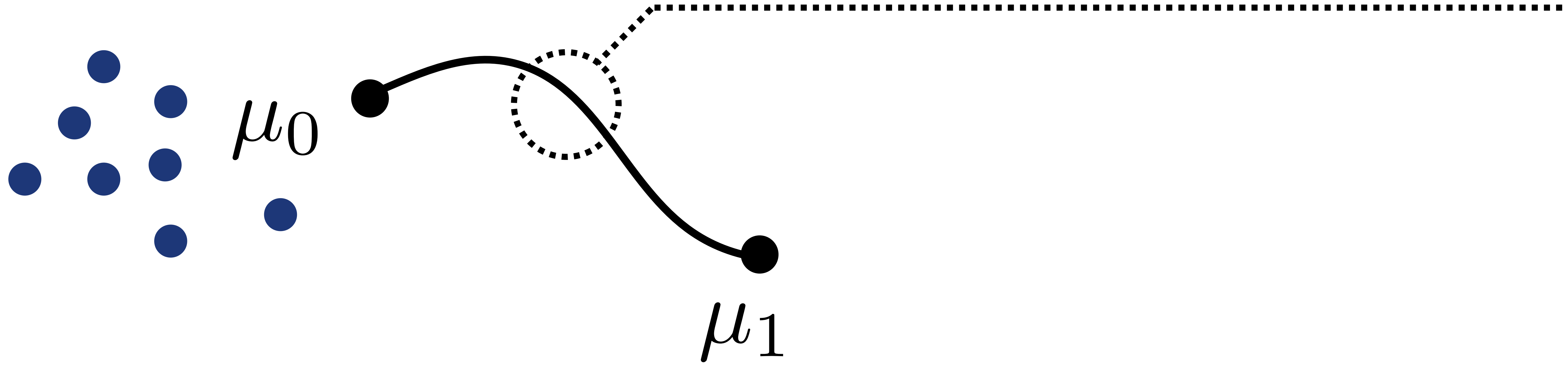

David Kinderlehrer


Felix Otto

Jordan-Kinderlehrer-Otto (JKO) Flows

$$\mu_{t+1} := \underset{\mu}{\text{argmin}} \ \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$$

proximity operator



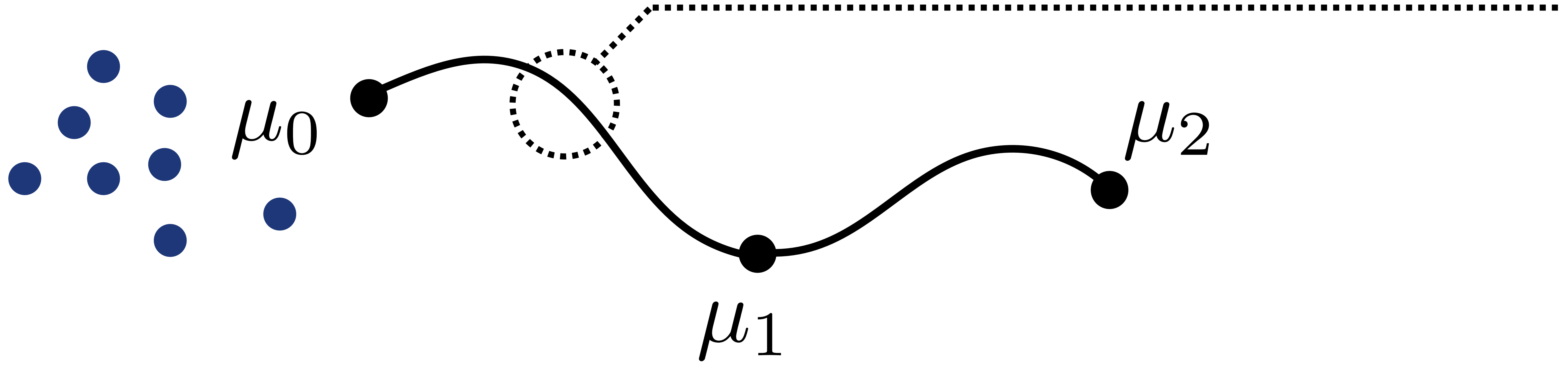
Richard Jordan David Kinderlehrer Felix Otto

Jordan-Kinderlehrer-Otto (JKO) Flows

Part 3

$$\mu_{t+1} := \underset{\mu}{\operatorname{argmin}} \; \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$$

proximity operator



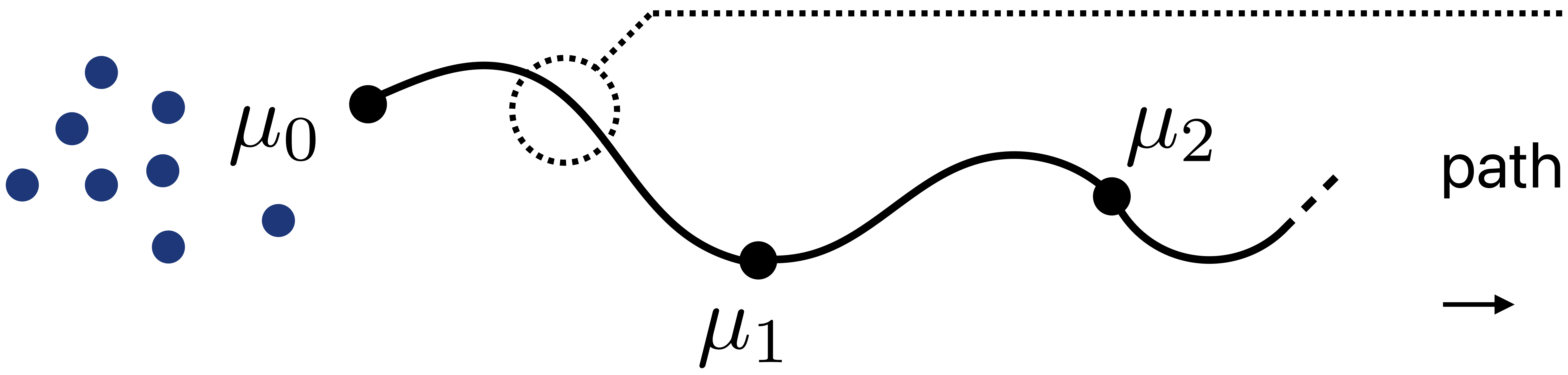
Richard Jordan David Kinderlehrer Felix Otto

Jordan-Kinderlehrer-Otto (JKO) Flows

Part 3

$$\mu_{t+1} := \underset{\mu}{\operatorname{argmin}} \; \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$$

proximity operator



path μ_t

→ Wasserstein gradient flow a.k.a. JKO flow


Richard Jordan


David Kinderlehrer


Felix Otto

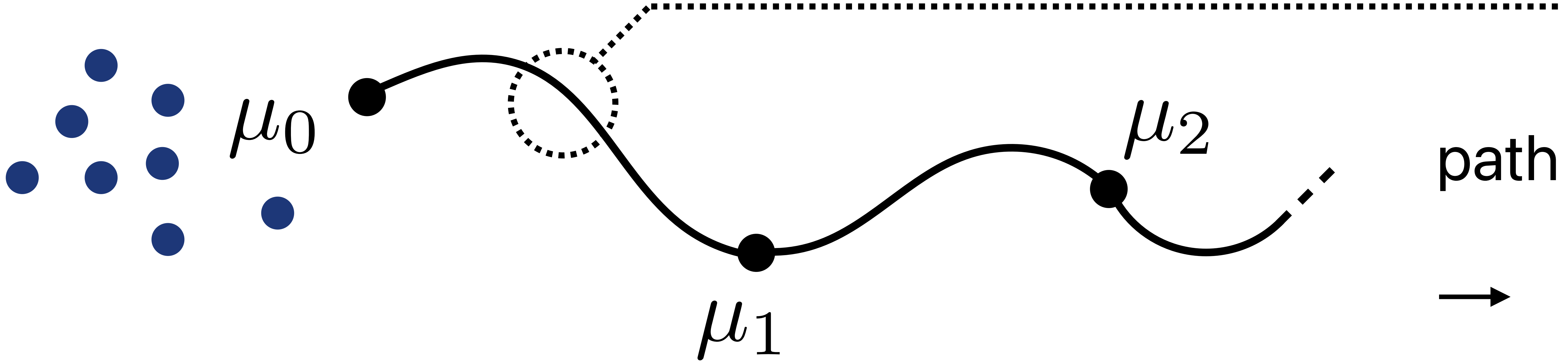
[Jordan+98]

Jordan-Kinderlehrer-Otto (JKO) Flows

Part 3

$$\mu_{t+1} := \underset{\mu}{\operatorname{argmin}} \; \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$$

proximity operator



path μ_t

Richard Jordan David Kinderlehrer Felix Otto

→ Wasserstein gradient flow a.k.a. JKO flow

[Jordan+98]

… **heat equation**

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t$$

energy potential

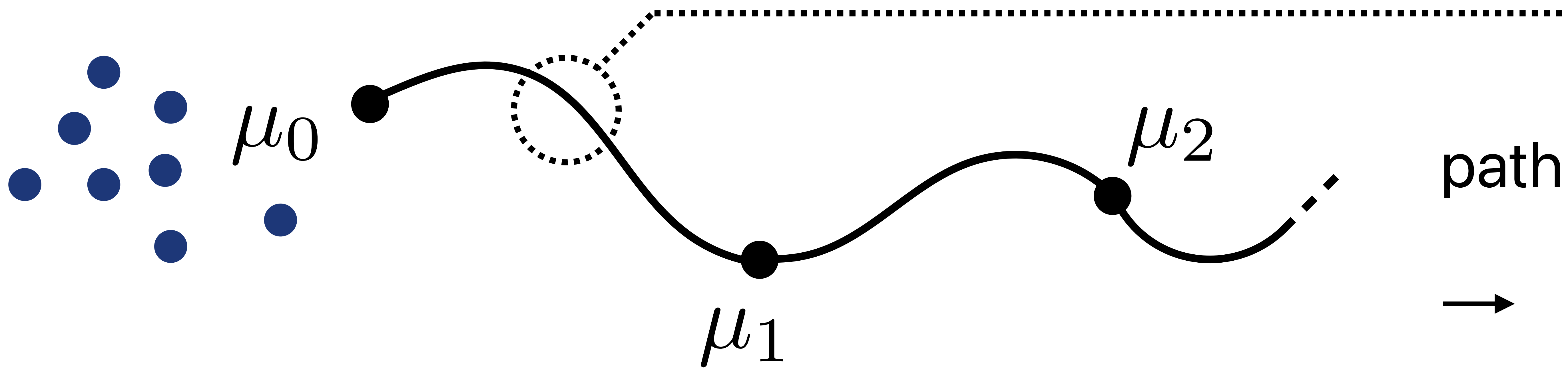
$$J(\mu_t) = \int \mu_t(x) \log \mu_t(x) \mathrm{d}x$$

[Santambrogio15]

Jordan-Kinderlehrer-Otto (JKO) Flows

$$\mu_{t+1} := \underset{\mu}{\operatorname{argmin}} \; \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$$

proximity operator



path μ_t

→ Wasserstein gradient flow a.k.a. JKO flow

Richard Jordan David Kinderlehrer Felix Otto

[Jordan+98]

… **heat equation**

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t$$

energy potential

$$J(\mu_t) = \int \mu_t(x) \log \mu_t(x) \mathrm{d}x$$

… linear **Fokker-Planck equation**

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \nabla \cdot (\mu_t \nabla V)$$

energy potential

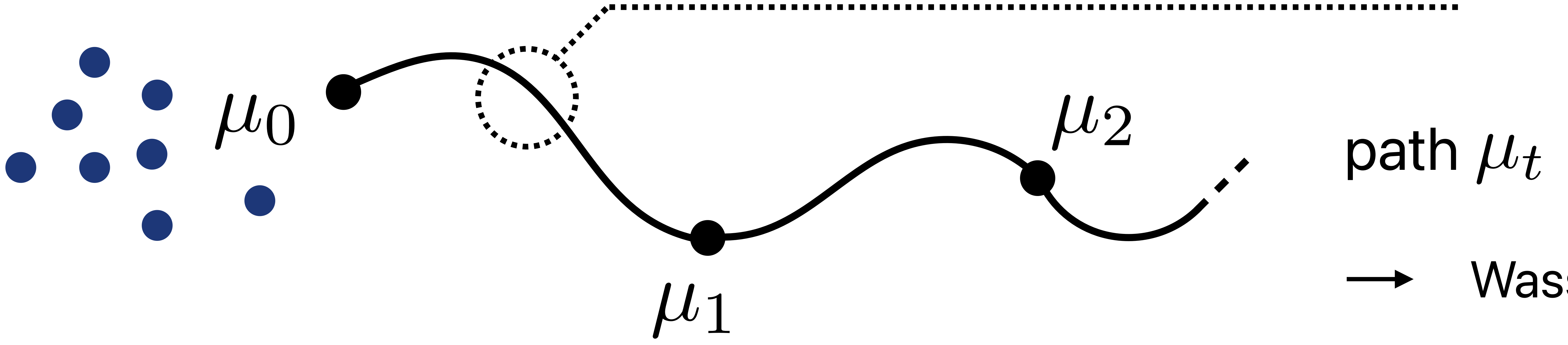
$$J(\mu_t) = \int \mu_t(x) \log \mu_t(x) \mathrm{d}x + \int V(x) \mu_t(x) \mathrm{d}x$$

[Santambrogio15]

Jordan-Kinderlehrer-Otto (JKO) Flows

$$\mu_{t+1} := \underset{\mu}{\operatorname{argmin}} \; \frac{1}{2\tau} W(\mu, \mu_t) + J(\mu)$$

proximity operator



Richard Jordan   David Kinderlehrer   Felix Otto

→ Wasserstein gradient flow a.k.a. JKO flow

[Jordan+98]

... **heat equation**

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t$$

energy potential

$$J(\mu_t) = \int \mu_t(x) \log \mu_t(x) \mathrm{d}x$$

... linear **Fokker-Planck equation**

$$\frac{\partial \mu_t}{\partial t} = \Delta \mu_t + \nabla \cdot (\mu_t \nabla V)$$

energy potential

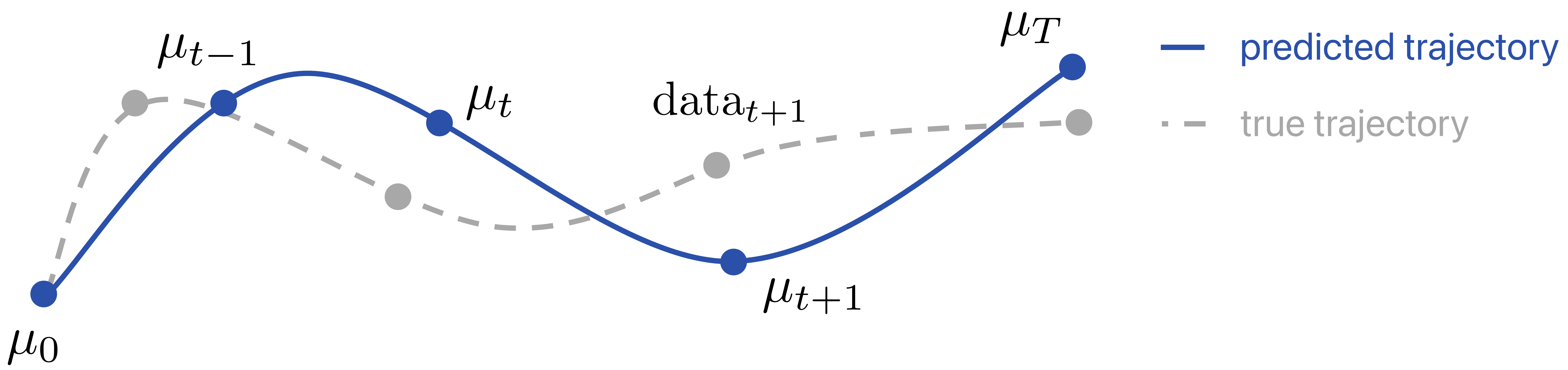
$$J(\mu_t) = \int \mu_t(x) \log \mu_t(x) \mathrm{d}x + \int V(x) \mu_t(x) \mathrm{d}x$$

... **unknown PDE** and **unknown potential function** J ?

[Santambrogio15]

Proximal Optimal Transport Model

Model trajectories as collective realizations of a causal JKO flow of measures



[Benamou+16, Peyré15, Korotin+21, Bunne+22, Alvarez-Melis+22]

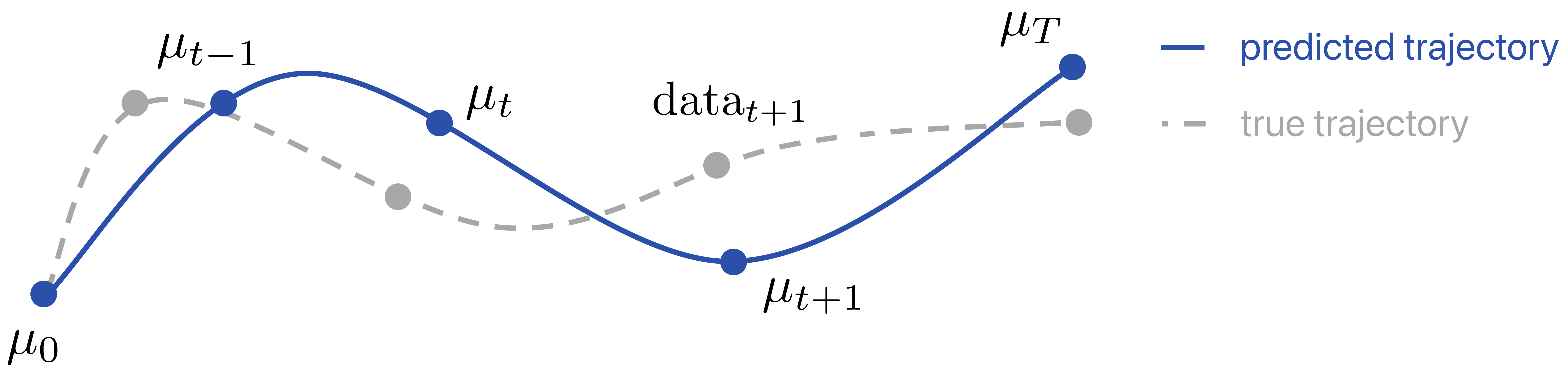
Proximal Optimal Transport Model

Model trajectories as collective realizations of a causal JKO flow of measures

and **learn the energy functional**

$$J_\xi(\mu) := \int E_\xi(x) d\mu(x)$$

where $E_\xi : \mathbb{R}^d \to \mathbb{R}$



[Benamou+16, Peyré15, Korotin+21, Bunne+22, Alvarez-Melis+22]

Proximal Optimal Transport Model

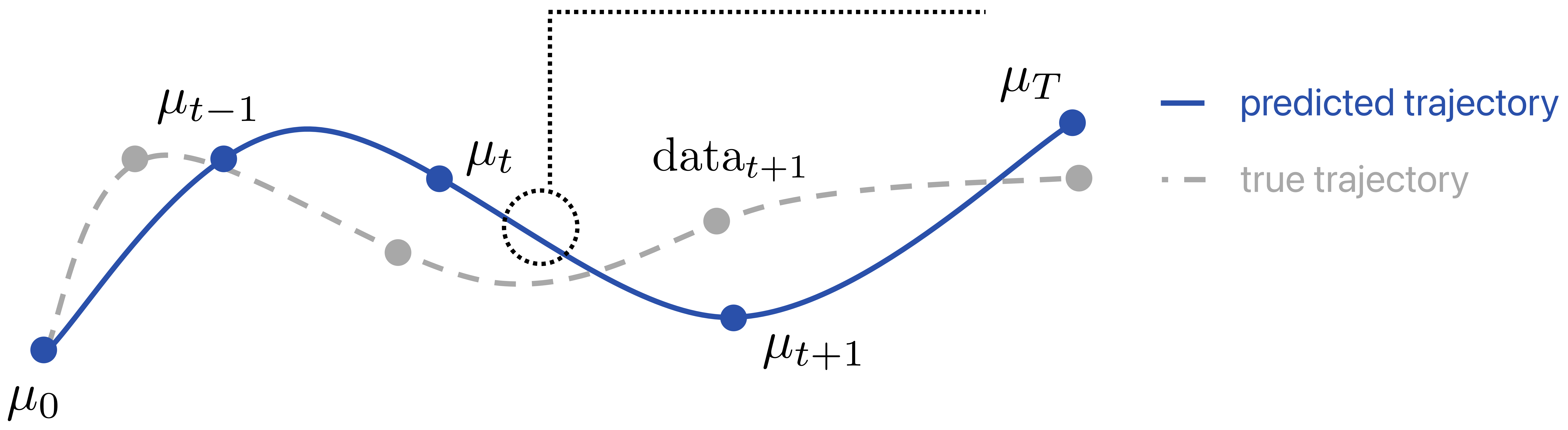
Model trajectories as collective realizations of a causal JKO flow of measures

and **learn the energy functional**

$$J_\xi(\mu) := \int E_\xi(x) d\mu(x)$$

where $E_\xi : \mathbb{R}^d \to \mathbb{R}$

$$\arg\min_\mu \frac{1}{2\tau} W(\mu, \mu_t) + J_\xi(\mu)$$



[Benamou+16, Peyré15, Korotin+21, Bunne+22, Alvarez-Melis+22]

Proximal Optimal Transport Model

Part 3

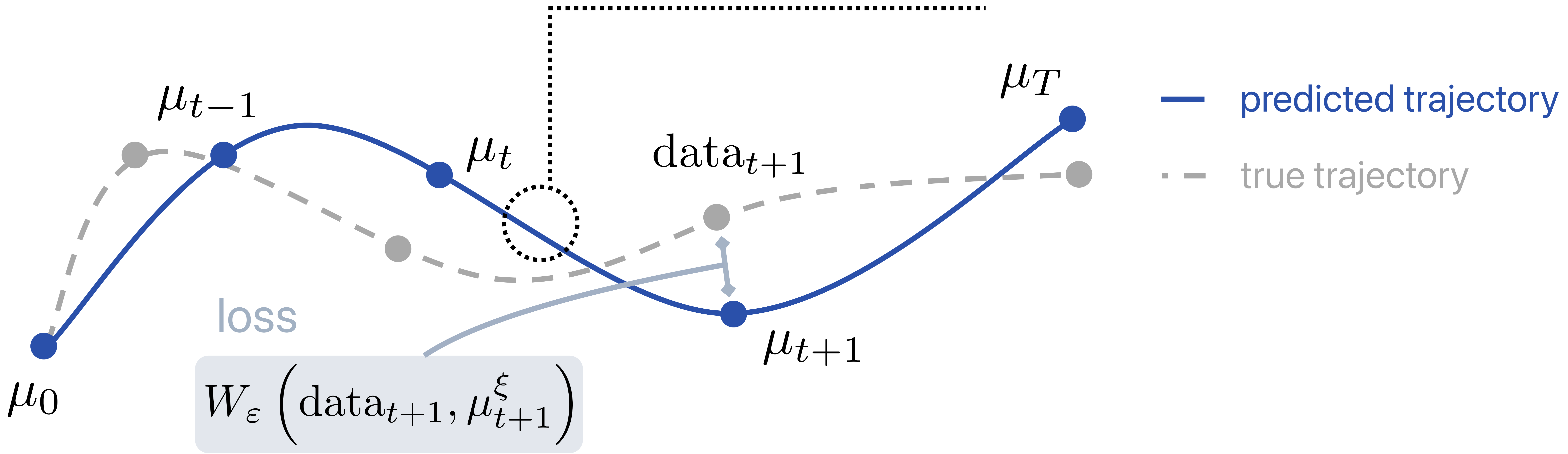
Model trajectories as collective realizations of a causal JKO flow of measures

and **learn the energy functional**

$$J_\xi(\mu) := \int E_\xi(x) d\mu(x)$$

where $E_\xi : \mathbb{R}^d \to \mathbb{R}$

$$\arg\min_\mu \frac{1}{2\tau} W(\mu, \mu_t) + J_\xi(\mu)$$



[Benamou+16, Peyré15, Korotin+21, Bunne+22, Alvarez-Melis+22]

Proximal Optimal Transport Model

Model trajectories as collective realizations of a causal JKO flow of measures

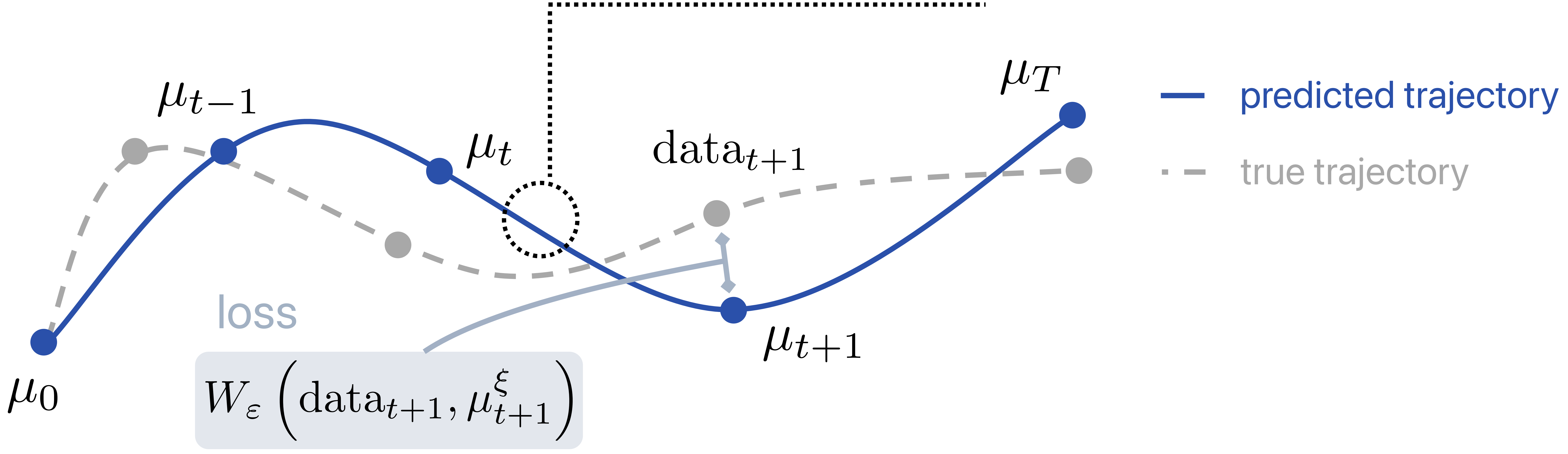
and **learn the energy functional**

$$J_\xi(\mu) := \int E_\xi(x) d\mu(x)$$

where $E_\xi : \mathbb{R}^d \to \mathbb{R}$

$$\arg\min_{\mu} \frac{1}{2\tau} W(\mu, \mu_t) + J_\xi(\mu)$$

How to solve JKO step numerically?



[Benamou+16, Peyré15, Korotin+21, Bunne+22, Alvarez-Melis+22]

Proximal Optimal Transport Model

Model trajectories as collective realizations of a causal JKO flow of measures

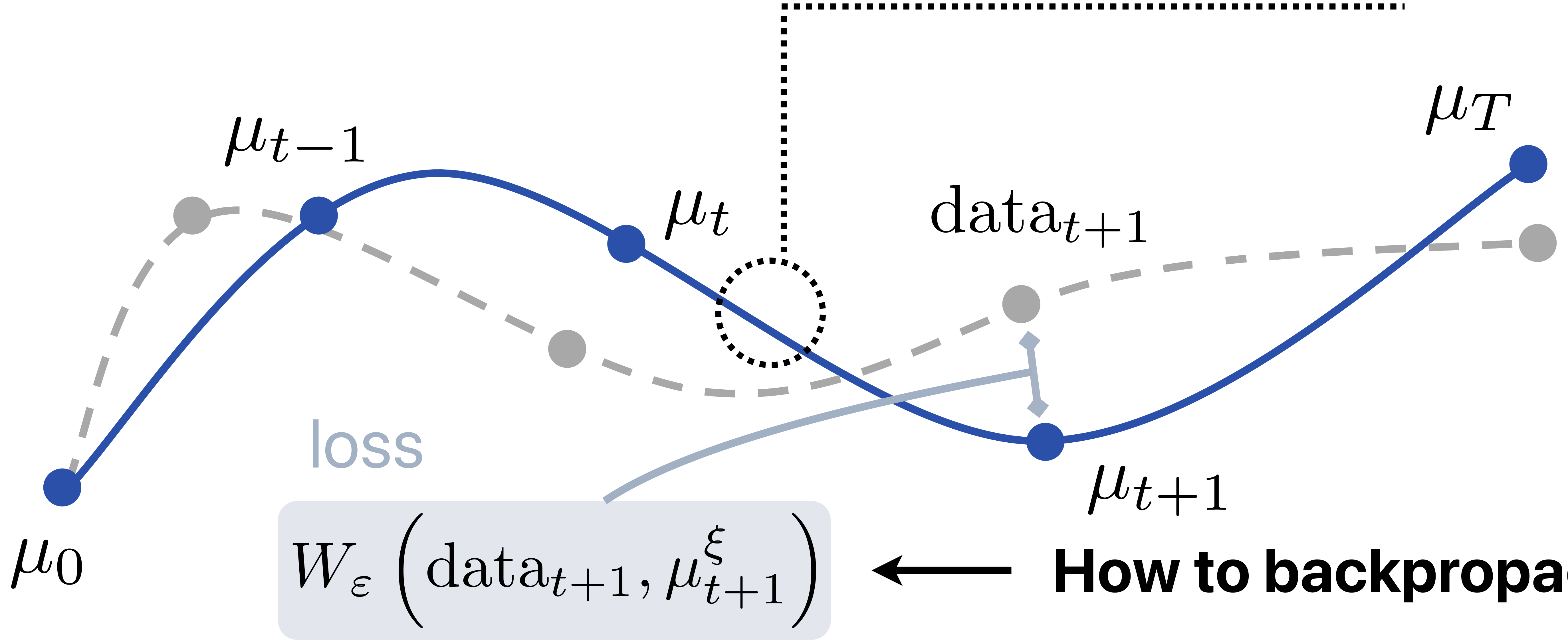
and **learn the energy functional**

$$J_\xi(\mu) := \int E_\xi(x) d\mu(x)$$

where $E_\xi : \mathbb{R}^d \to \mathbb{R}$

$$\arg\min_{\mu} \frac{1}{2\tau} W(\mu, \mu_t) + J_\xi(\mu)$$

How to solve JKO step numerically?



How to backpropagate?

predicted trajectory

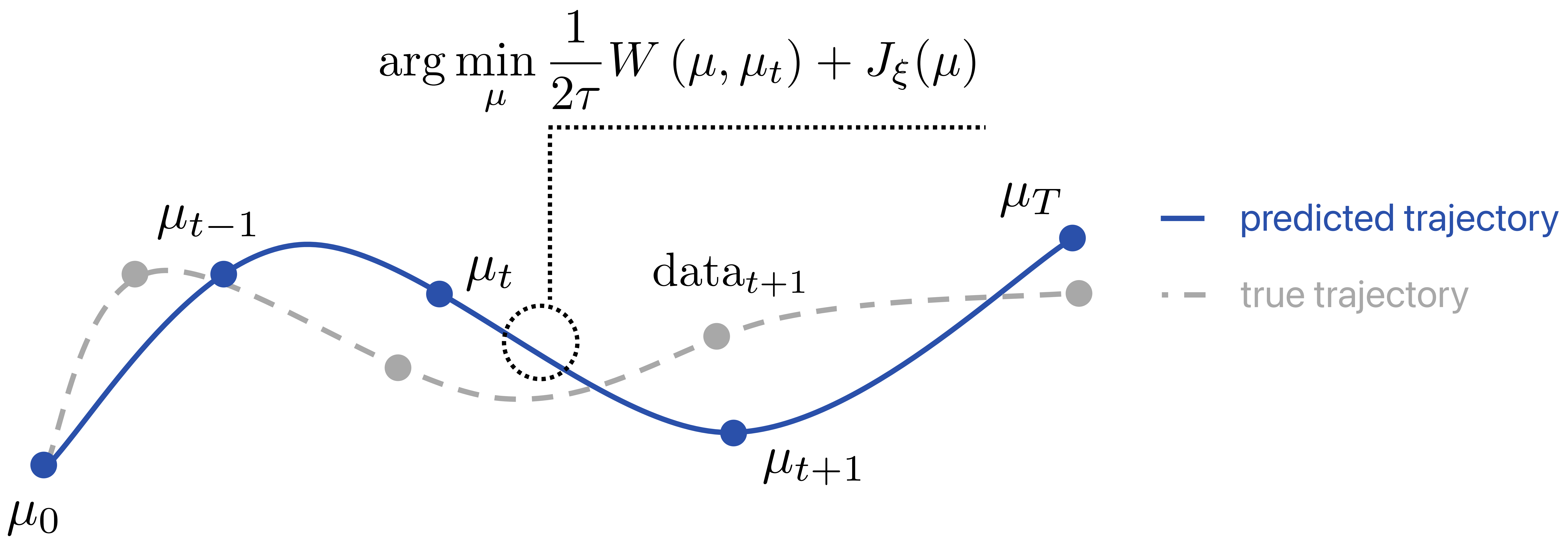
true trajectory

[Benamou+16, Peyré15, Korotin+21, Bunne+22, Alvarez-Melis+22]

Reformulating the JKO Objective

Given Brenier's theorem

$$T(x) = \nabla\psi(x)$$



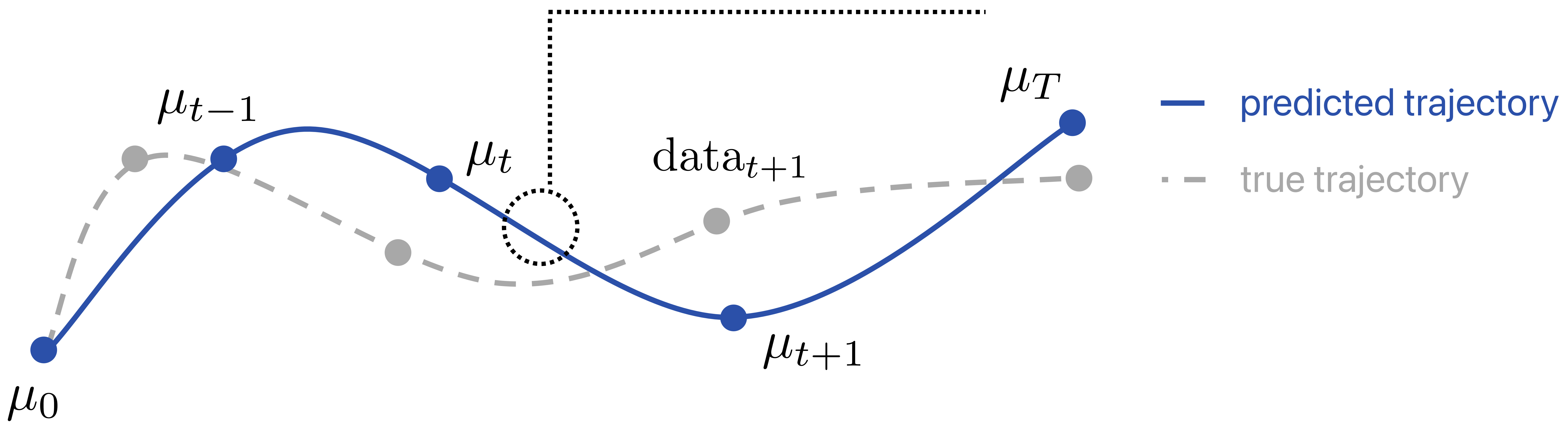
[Benamou+16, Korotin+21, Bunne+22, Alvarez-Melis+22]

Reformulating the JKO Objective

Given Brenier's theorem

$$T(x) = \nabla\psi(x)$$

~~$$\arg\min_{\mu} \frac{1}{2\tau} W(\mu, \mu_t) + J_\xi(\mu)$$~~



[Benamou+16, Korotin+21, Bunne+22, Alvarez-Melis+22]

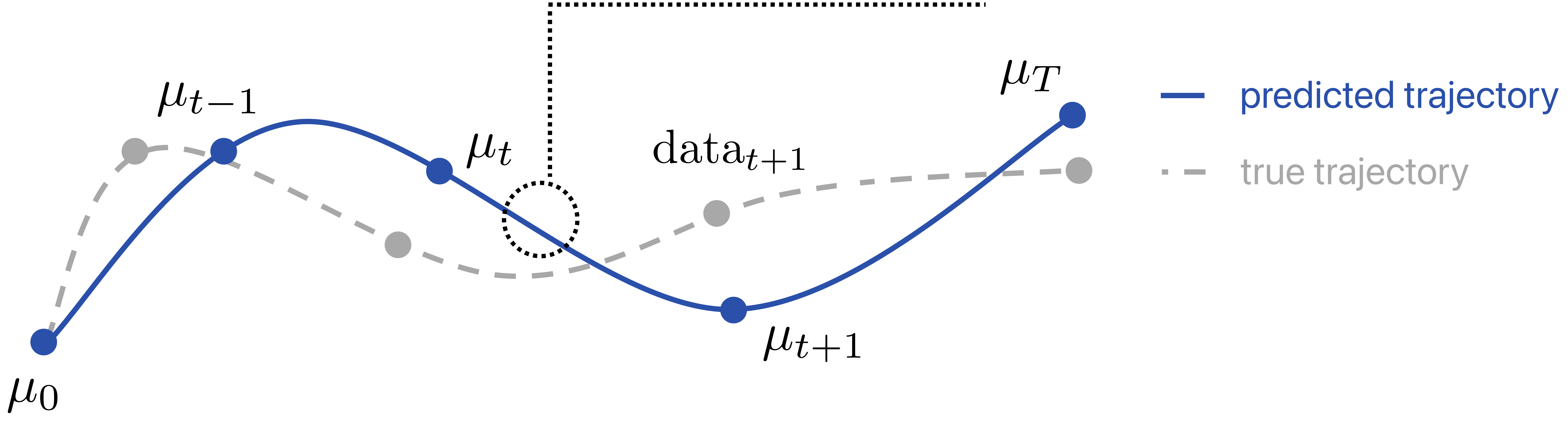
Reformulating the JKO Objective

Given Brenier's theorem

$$T(x) = \nabla\psi(x)$$

$$\arg\min_{\theta} \frac{1}{2\tau} \int \|x - \nabla\psi_{\theta}(x)\|^2 \mathrm{d}\mu_t + J_{\xi}(\nabla\psi_{\theta\#}\mu_t)$$

~~$$\arg\min_{\mu} \frac{1}{2\tau} W(\mu, \mu_t) + J_{\xi}(\mu)$$~~



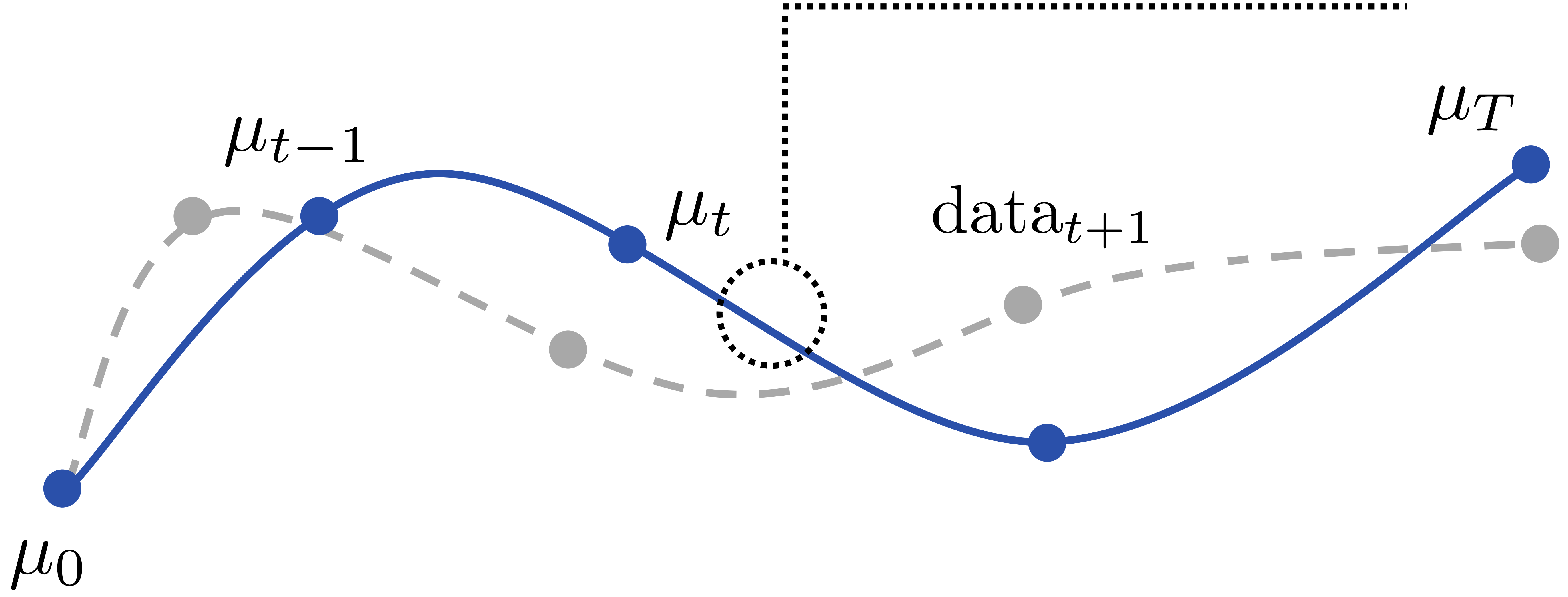
[Benamou+16, Korotin+21, Bunne+22, Alvarez-Melis+22]

Proximal Optimal Transport Model

Part 3

$$\arg\min_{\theta} \frac{1}{2\tau} \int \|x - \nabla\psi_\theta(x)\|^2 \mathrm{d}\mu_t + J_\xi(\nabla\psi_{\theta\#}\mu_t)$$

$$\mu_{t+1}^{\xi} := \nabla\left(\psi_{\theta^\star}\right)_{\#}\mu_t$$

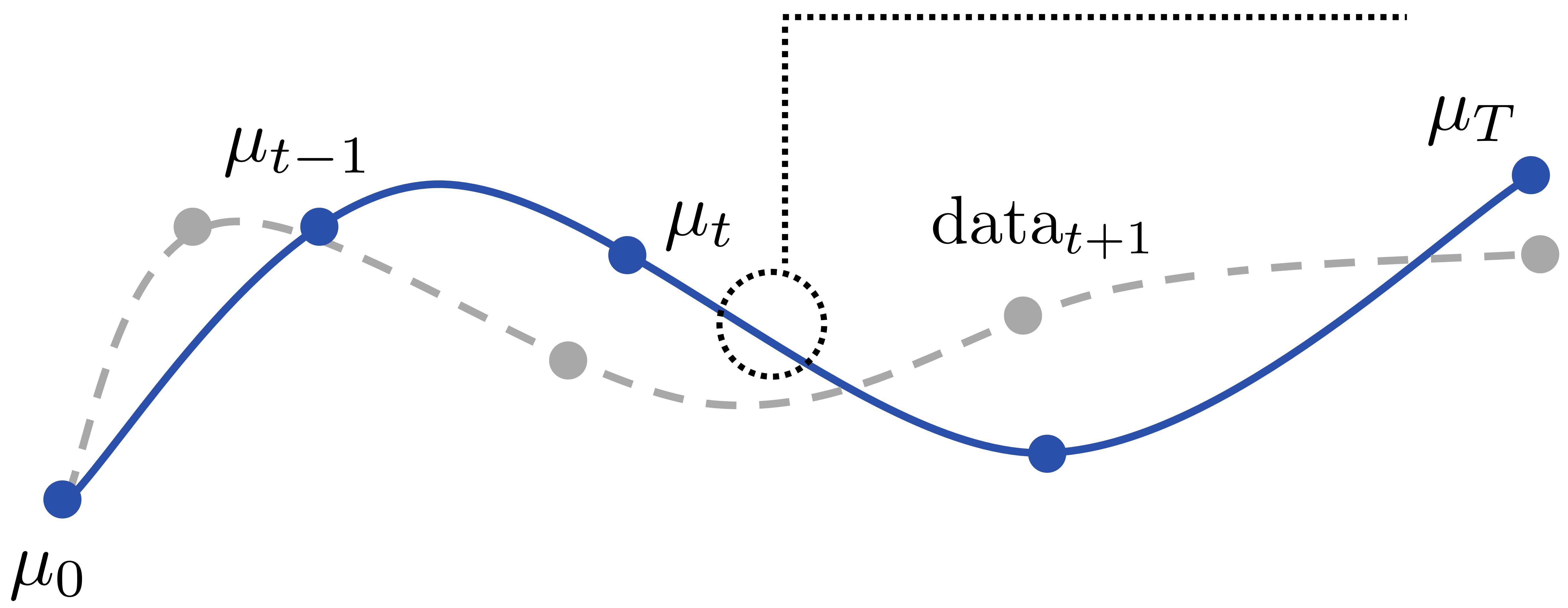


[Bunne+22]

Proximal Optimal Transport Model

$$\arg\min_{\theta} \frac{1}{2\tau} \int \|x - \nabla\psi_\theta(x)\|^2 \mathrm{d}\mu_t + J_\xi(\nabla\psi_{\theta\#}\mu_t)$$

$$\mu_{t+1}^{\xi} := \nabla\left(\psi_{\theta^\star}\right)_{\#} \mu_t$$



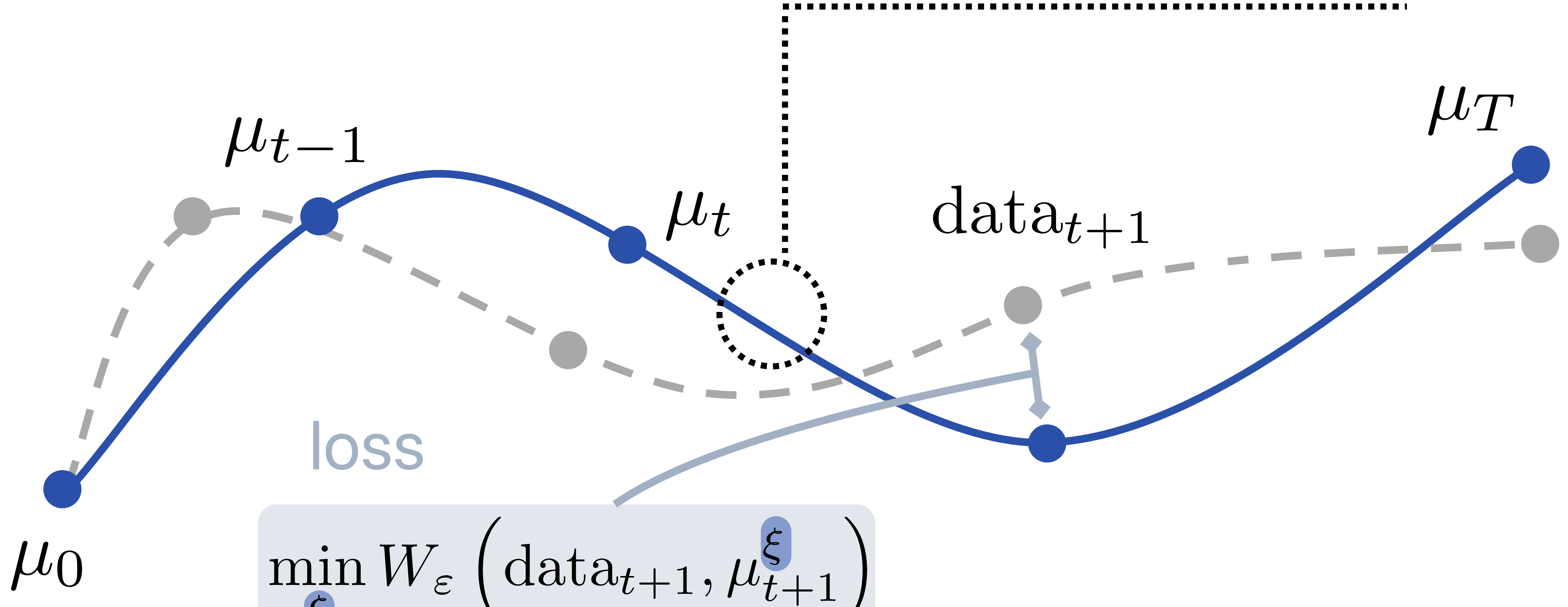
lower level optimization w.r.t. ψ_θ

[Bunne+22]

Proximal Optimal Transport Model

$$\arg\min_{\theta} \frac{1}{2\tau} \int \|x - \nabla \psi_\theta(x)\|^2 d\mu_t + J_\xi(\nabla \psi_{\theta\#}\mu_t)$$

$$\mu^{\xi}_{t+1} := \nabla \left(\psi_{\theta^\star}\right)_{\#} \mu_t$$



$$\min_{\xi} W_\varepsilon \left(\text{data}_{t+1}, \mu^{\xi}_{t+1}\right)$$

lower level optimization w.r.t. ψ_θ

upper level optimization w.r.t. J_ξ

[Bunne+22]

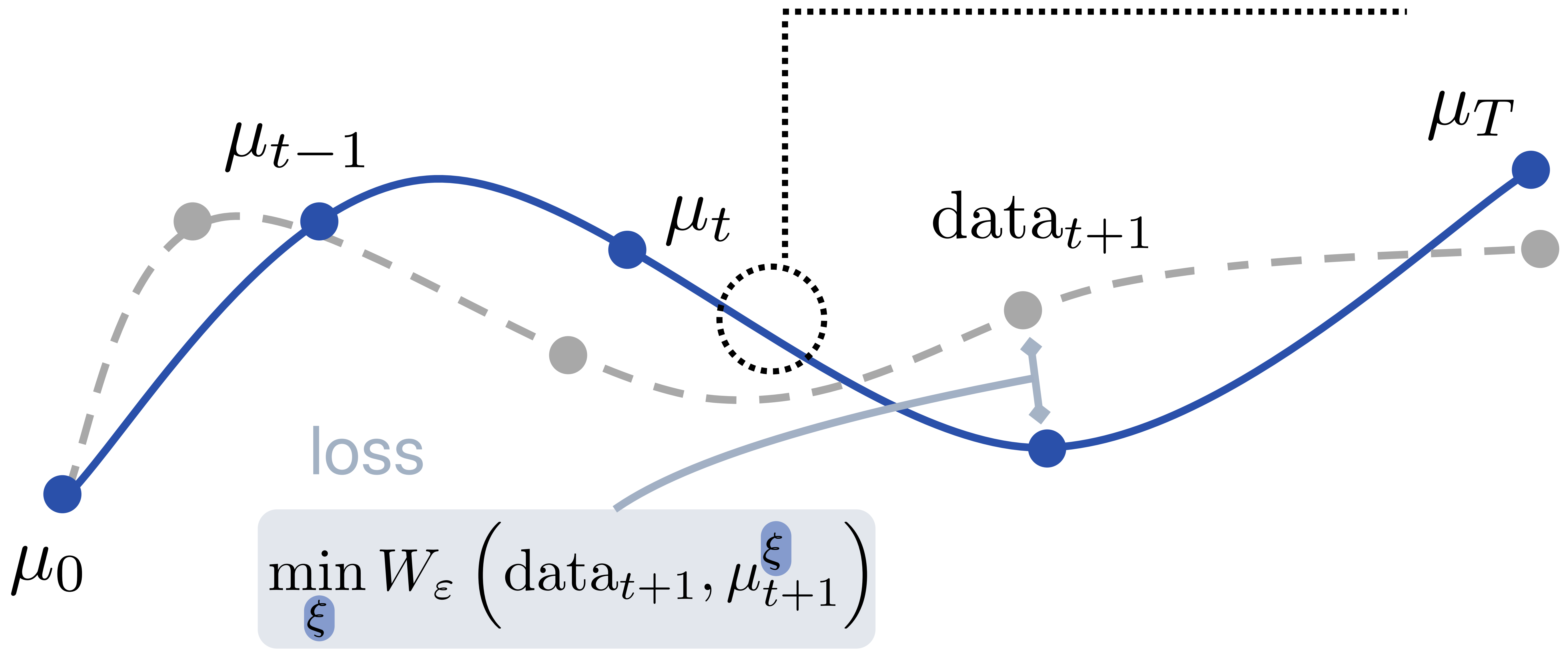
Proximal Optimal Transport Model

Part 3

influence of ξ in gradient updates of θ

$$\arg\min_{\theta} \frac{1}{2\tau} \int \|x - \nabla\psi_\theta(x)\|^2 \mathrm{d}\mu_t + J_\xi(\nabla\psi_{\theta\#}\mu_t)$$

$$\mu^{\xi}_{t+1} := \nabla\left(\psi_{\theta^\star}\right)_{\#}\mu_t$$



lower level optimization w.r.t. ψ_θ

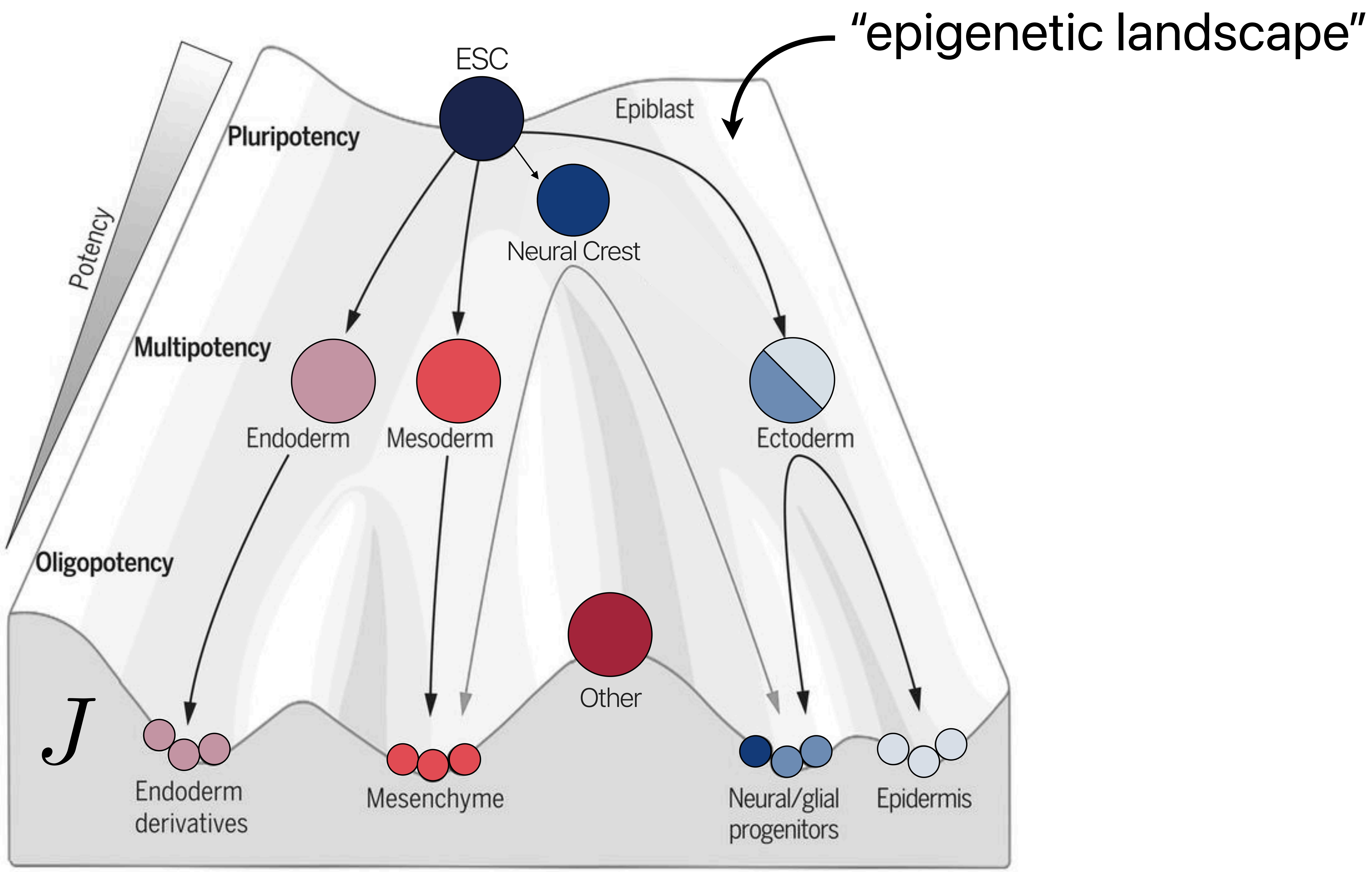
upper level optimization w.r.t. J_ξ

[Bunne+22]

Application: **Cell Differentiation on the Epigenetic Landscape**

Part 3

Differentiation of embryonic stem cells into diverse cell lineages

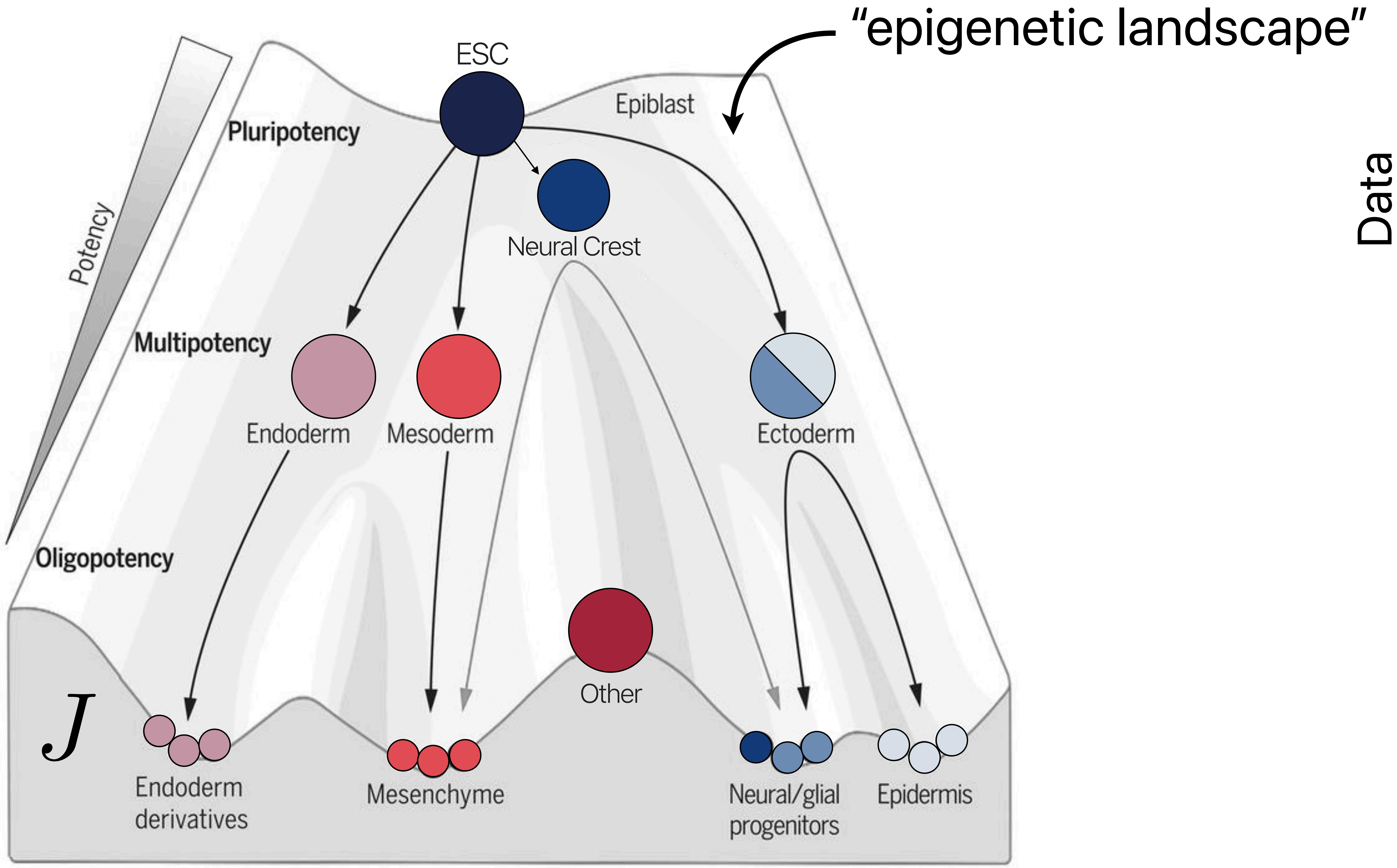


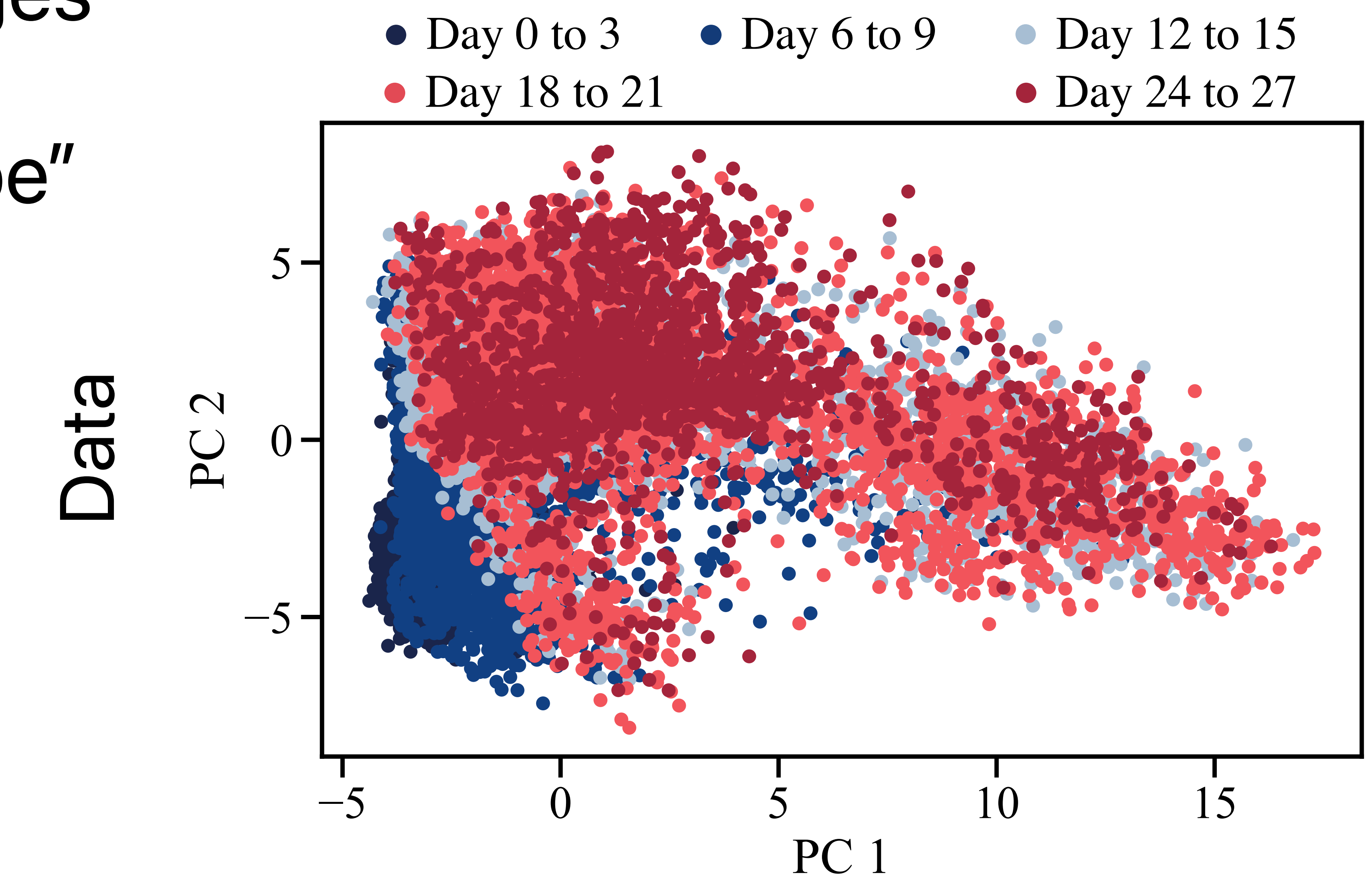
[Bunne+22, Moon+19, Zalc+21, Tong+21, Hashimoto+16]

Application: **Cell Differentiation on the Epigenetic Landscape**

Part 3

Differentiation of embryonic stem cells into diverse cell lineages



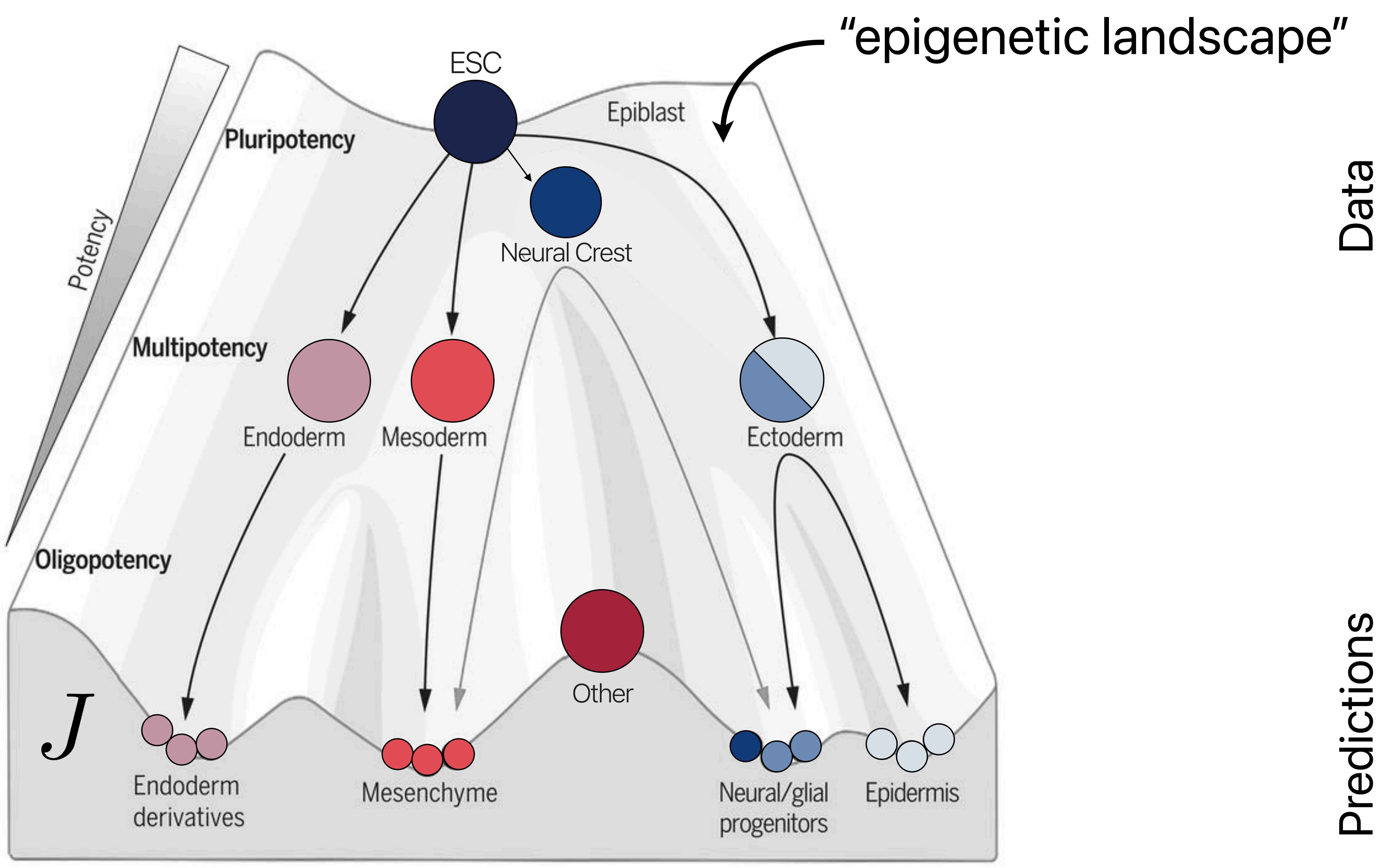


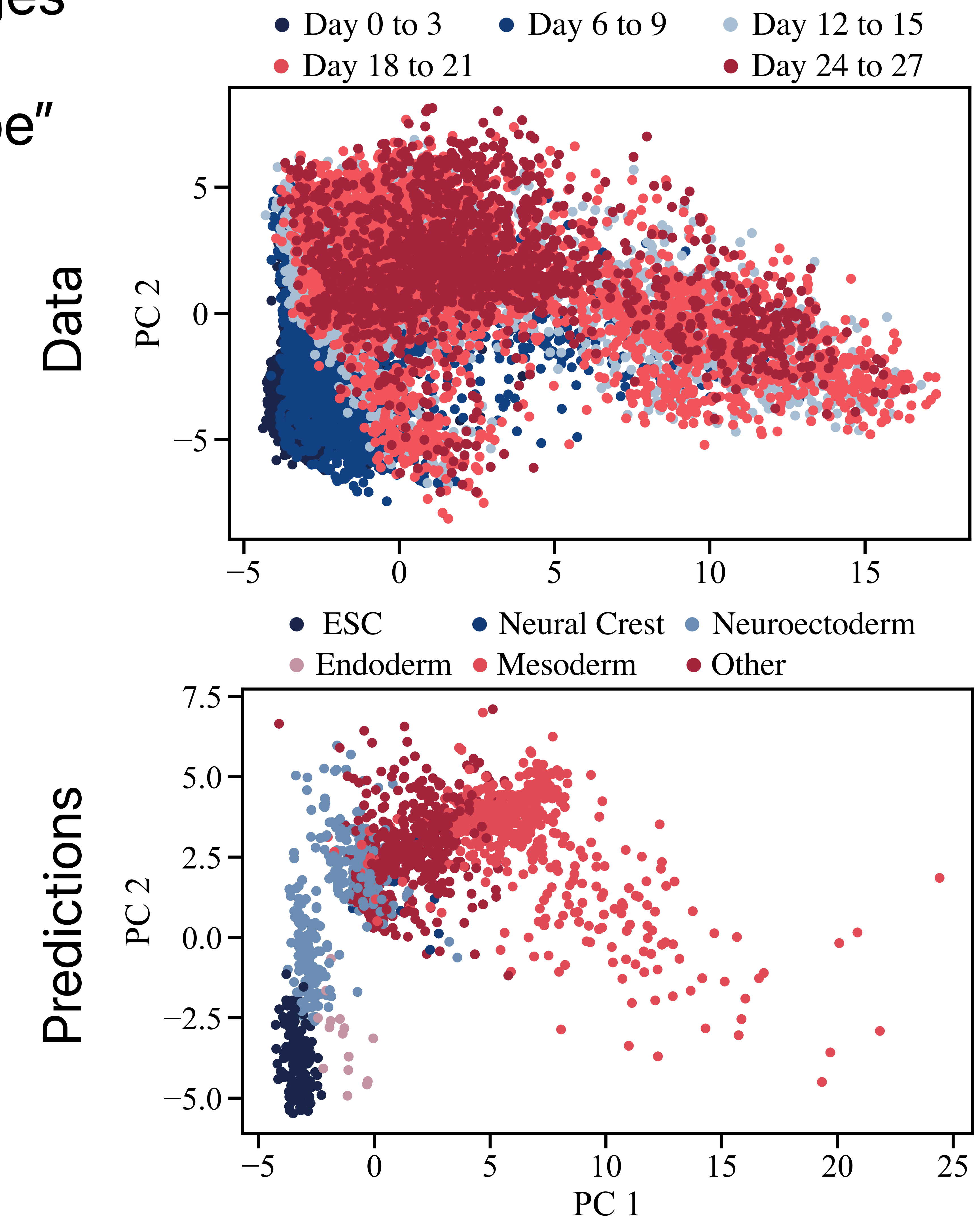
[Bunne+22, Moon+19, Zalc+21, Tong+21, Hashimoto+16]

Application: **Cell Differentiation on the Epigenetic Landscape**

Part 3

Differentiation of embryonic stem cells into diverse cell lineages





[Bunne+22, Moon+19, Zalc+21, Tong+21, Hashimoto+16]

A Stochastic Control Perspective

Part 3

Benamou-Brenier

$$\inf_{(\mu_t, v)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|v(t,x)\|^2 d\mu_t(x) dt$$

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$ continuity equation

$$\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$$

[Mikami+08, Chen+16, Chen+21]

A Stochastic Control Perspective

Benamou-Brenier

$$\inf_{(\mu_t, v)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|v(t,x)\|^2 d\mu_t(x) dt$$

→ minimal kinetic energy in *fluid dynamics*

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$ continuity equation

$$\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$$

[Mikami+08, Chen+16, Chen+21]

A Stochastic Control Perspective

Part 3

Benamou-Brenier

$$\inf_{(\mu_t, v)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|v(t,x)\|^2 d\mu_t(x) dt$$

→ minimal kinetic energy in *fluid dynamics*

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$ continuity equation

$$\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$$

Stochastic Control

feedback control

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \left\| v(t, X_t) \right\|^2 \mathrm{d}t \right\}$$

→ steer system with minimal cost

$$\mathrm{d}X_t = v(t, X_t) \, \mathrm{d}t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

[Mikami+08, Chen+16, Chen+21]

A Stochastic Control Perspective

Part 3

Benamou-Brenier

$$\inf_{(\mu_t, v)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|v(t,x)\|^2 d\mu_t(x) dt$$

→ minimal kinetic energy in *fluid dynamics*

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$ continuity equation

$$\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$$

Stochastic Control

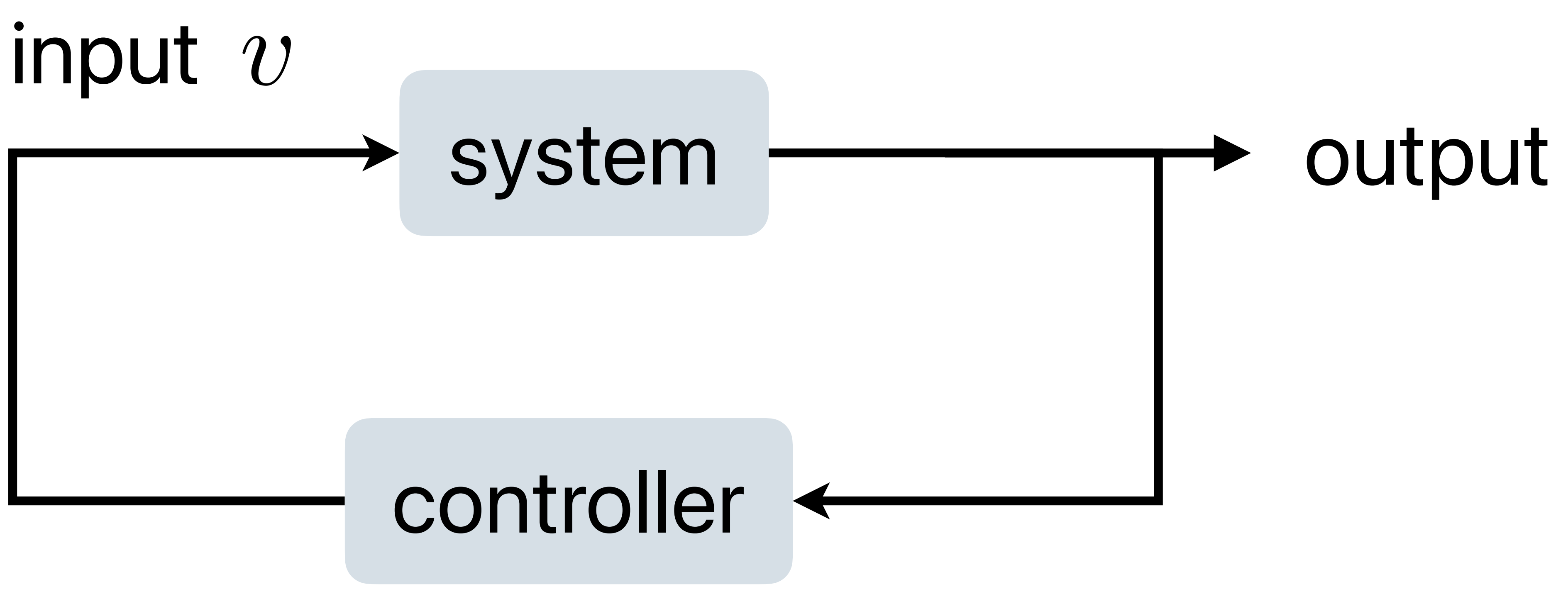
feedback control

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \|v(t, X_t)\|^2 \, \mathrm{d}t \right\}$$

→ steer system with minimal cost

$$\mathrm{d}X_t = v(t, X_t)\,\mathrm{d}t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$



[Mikami+08, Chen+16, Chen+21]

A Stochastic Control Perspective

Benamou-Brenier

$$\inf_{(\mu_t, v)} \int_0^1 \int_{\mathbb{R}^n} \frac{1}{2} \|v(t,x)\|^2 d\mu_t(x) dt$$

→ minimal kinetic energy in *fluid dynamics*

$$\frac{\partial \mu_t}{\partial t} + \nabla \cdot (v \mu_t) = 0$$ continuity equation

$$\mu_{t=0} = \mu_0, \mu_{t=1} = \mu_1$$

Stochastic Control

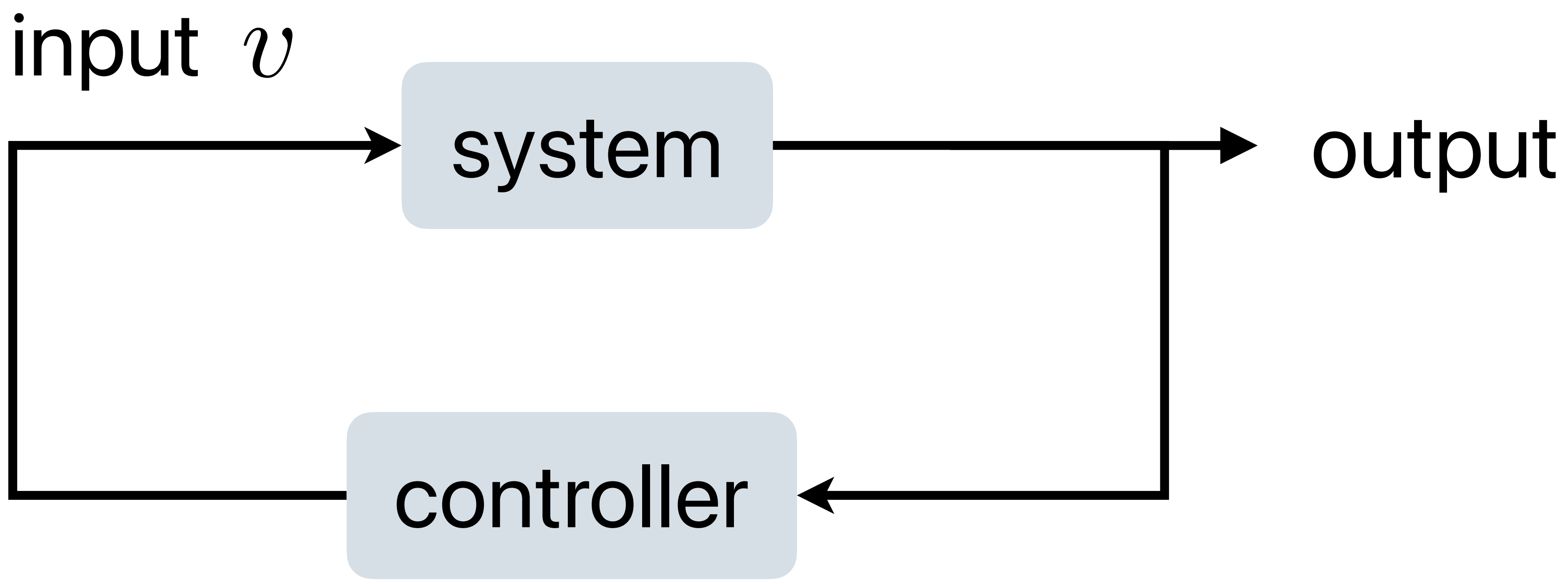
feedback control

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \|v(t, X_t)\|^2 \, \mathrm{d}t \right\}$$

→ steer system with minimal cost

$$\mathrm{d}X_t = v(t, X_t)\, \mathrm{d}t$$ state distribution

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$



[Mikami+08, Chen+16, Chen+21]

A Stochastic Control Perspective

Stochastic Control of OT

$$\inf_{v\in\mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\left\|v\left(t, X_t\right)\right\|^2 \, \mathrm{d}t\right\}$$

$$\mathrm{d}X_t = v\left(t, X_t\right) \mathrm{d}t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

Stochastic Control of Regularized OT

$$\inf_{v\in\mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\left\|v\left(t, X_t\right)\right\|^2 \, \mathrm{d}t\right\}$$

stochastic diffusion process

$$\mathrm{d}X_t = v\left(t, X_t\right) \mathrm{d}t + \sigma \mathrm{d}\mathbb{W}_t$$

standard Wiener process

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

[Chen+21]

A Stochastic Control Perspective

Stochastic Control of OT

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \left\| v\left(t, X_t\right) \right\|^2 \, \mathrm{d}t \right\}$$

$$\mathrm{d}X_t = v\left(t, X_t\right) \mathrm{d}t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

Stochastic Control of Regularized OT

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \left\| v\left(t, X_t\right) \right\|^2 \, \mathrm{d}t \right\}$$

stochastic diffusion process

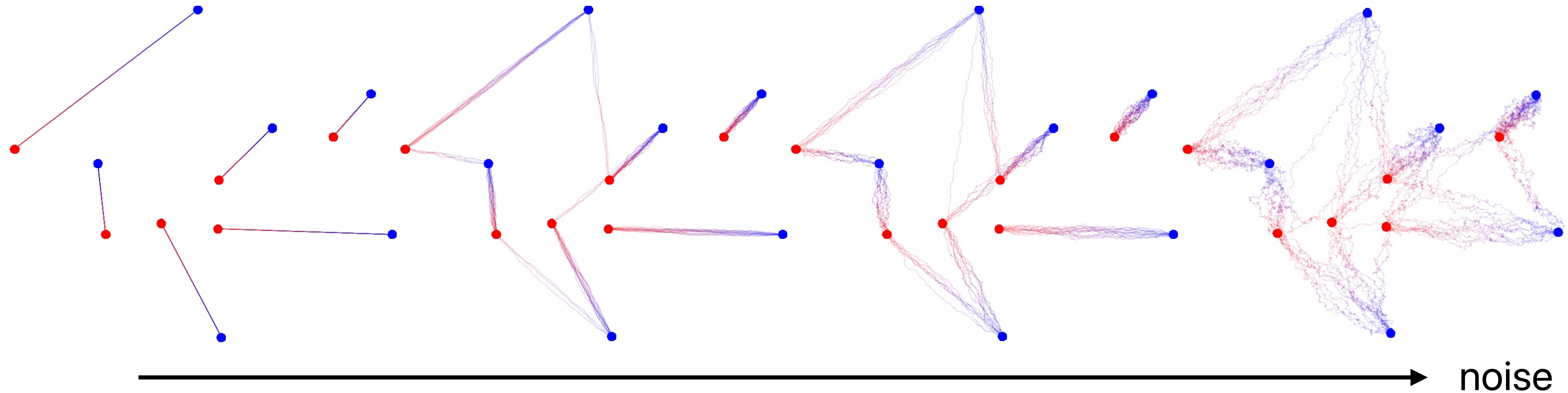
$$\mathrm{d}X_t = v\left(t, X_t\right) \mathrm{d}t + \sigma \mathrm{d}\mathbb{W}_t$$

standard Wiener process

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

● $t = 0$

● $t = 1$



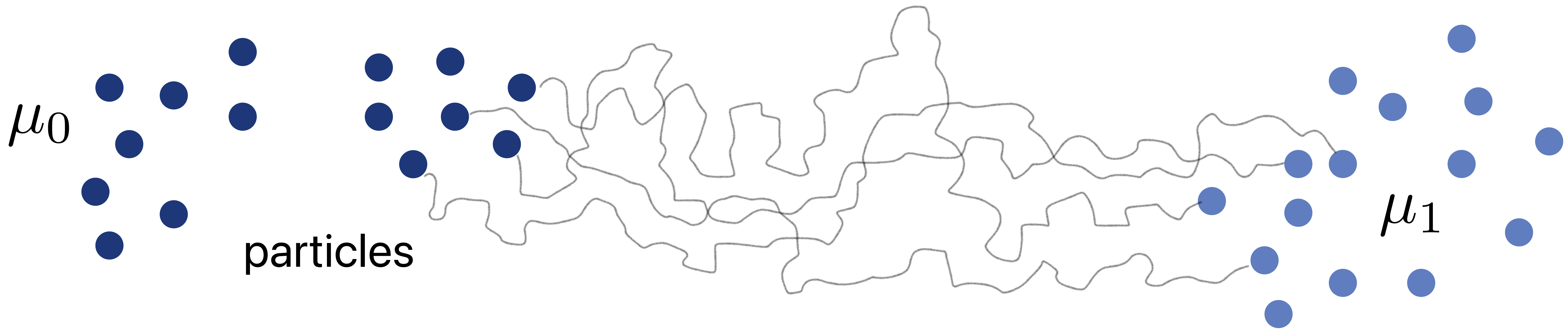
[Peyré+19]

[Chen+21]

A Connection to Particle Physics

Gedankenexperiment

Most likely random evolution between two point clouds of diffusive particles?



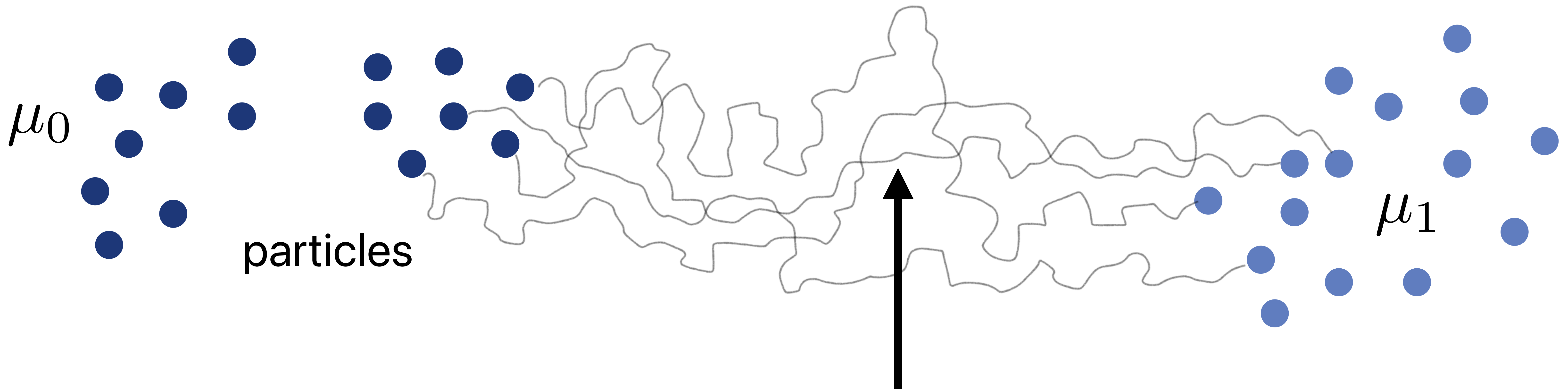
Erwin Schrödinger

[Schrödinger31,32]

A Connection to Particle Physics

Gedankenexperiment

Most likely random evolution between two point clouds of diffusive particles?




Erwin Schrödinger

... find stochastic process $\mathbb{P}_t$ on $[0,1]$ s.t. $\mathbb{P}_0 = \mu_0$ and $\mathbb{P}_1 = \mu_1$

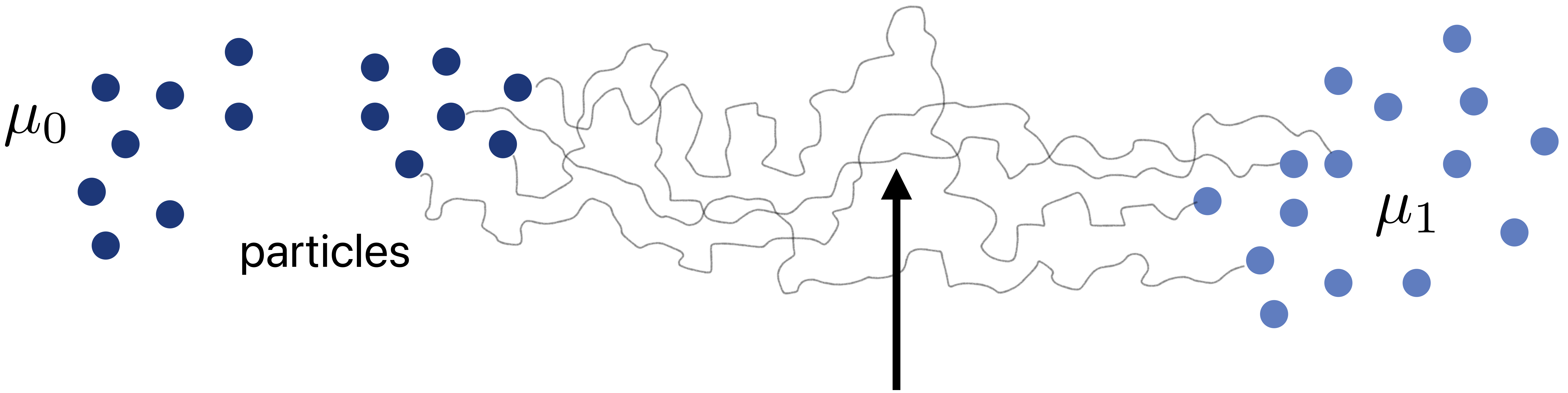
given some *prior* knowledge on a **reference process** $\mathbb{Q}_t$

[Schrödinger31,32]

A Connection to Particle Physics

Gedankenexperiment

Most likely random evolution between two point clouds of diffusive particles?



Erwin Schrödinger

... find stochastic process $\mathbb{P}_t$ on $[0,1]$ s.t. $\mathbb{P}_0 = \mu_0$ and $\mathbb{P}_1 = \mu_1$

given some *prior* knowledge on a **reference process** $\mathbb{Q}_t$

$$\min_{\mathbb{P}_t} D_{\mathrm{KL}} \left(\mathbb{P}_t \| \mathbb{Q}_t \right)$$

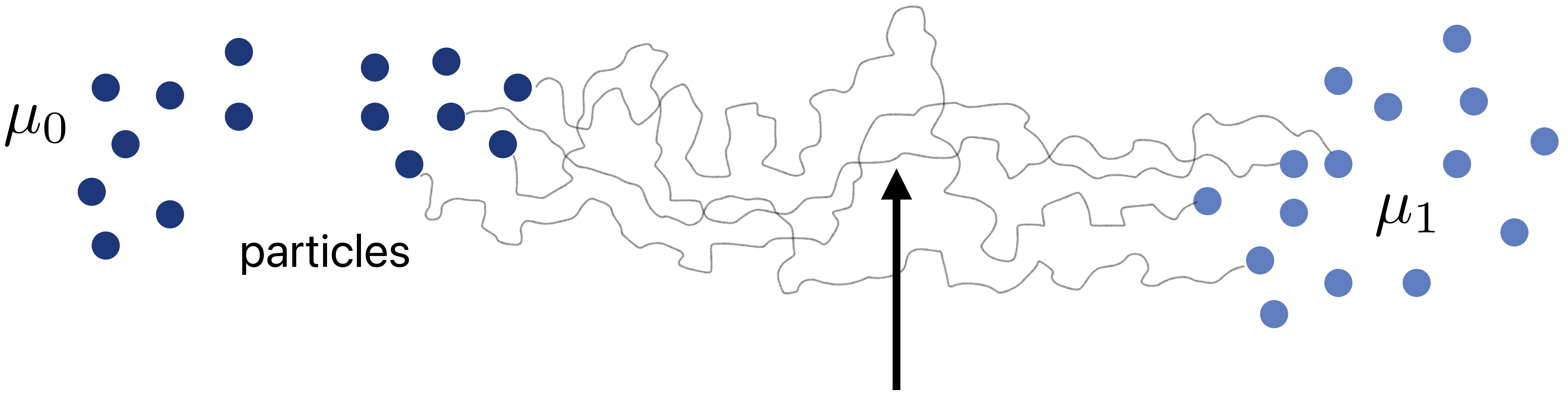
i.e., minimize overall *relative entropy*

[Schrödinger31,32]

A Connection to Particle Physics

Gedankenexperiment

Most likely random evolution between two point clouds of diffusive particles?



Erwin Schrödinger

... find stochastic process $\mathbb{P}_t$ on $[0,1]$ s.t. $\mathbb{P}_0 = \mu_0$ and $\mathbb{P}_1 = \mu_1$

given some *prior* knowledge on a **reference process** $\mathbb{Q}_t$

$$\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right)$$

i.e., minimize overall *relative entropy*

e.g., **Brownian motion**

[Schrödinger31,32]

A General SDE Class

Stochastic Control
of **Regularized** OT

$$\inf_{v \in \mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\left\|v\left(t, X_t\right)\right\|^2 \mathrm{~d}t\right\}$$

$$\mathrm{d}X_t = v\left(t, X_t\right) \mathrm{d}t + \sigma \mathrm{d}\mathbb{W}_t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

[Caluya+21, Chen+22]

A General SDE Class

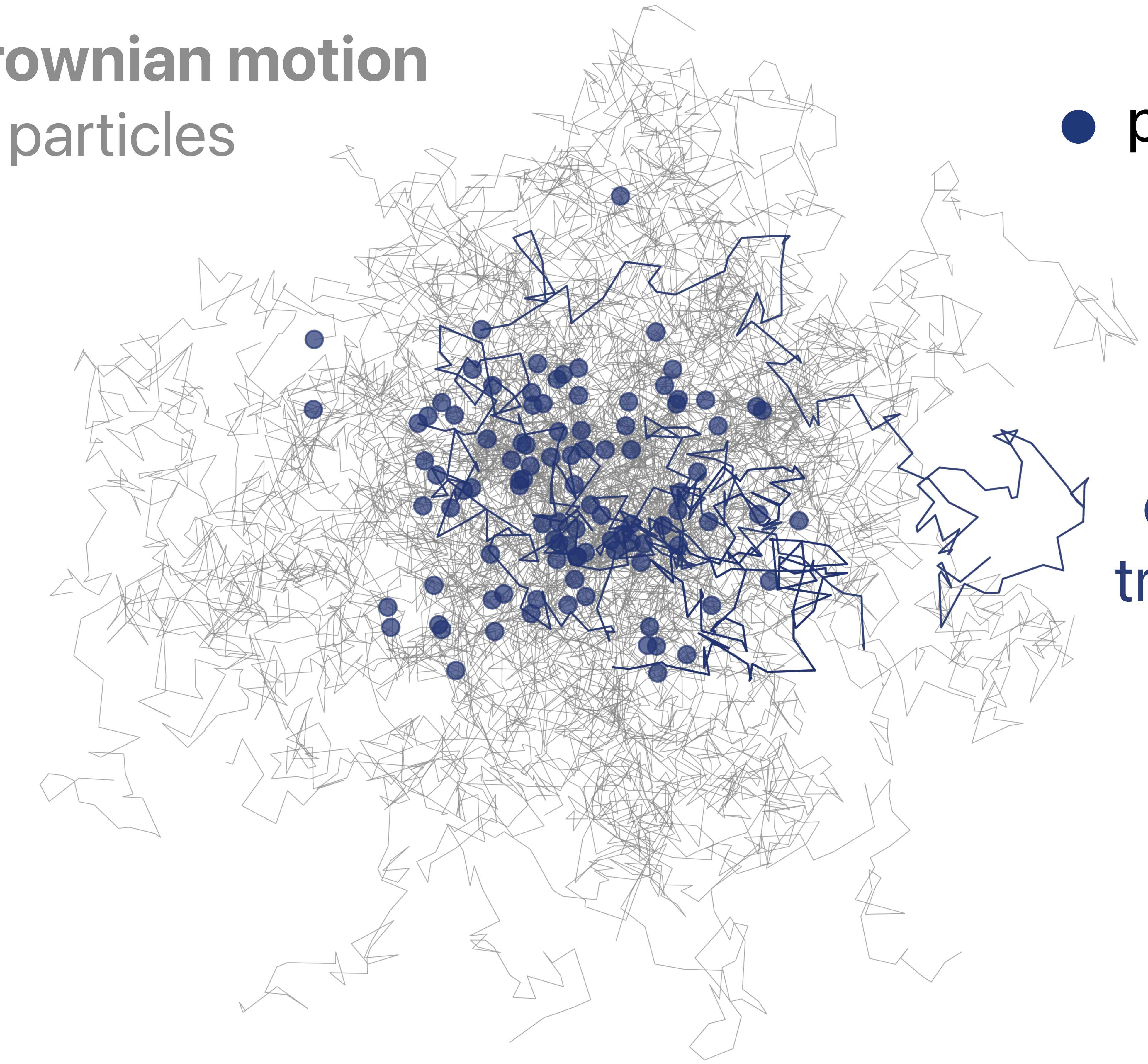
Stochastic Control
of **Regularized** OT

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \| v(t, X_t) \|^2 \, \mathrm{d}t \right\}$$

$$\mathrm{d}X_t = v(t, X_t) \, \mathrm{d}t + \sigma \mathrm{d}\mathbb{W}_t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

Brownian motion
of particles



[Caluya+21, Chen+22]

A General SDE Class

Part 3

Stochastic Control
of **Regularized** OT

$$\inf_{v \in \mathcal{V}} \mathbb{E} \left\{ \int_0^1 \frac{1}{2} \left\| v\left(t, X_t\right) \right\|^2 \, \mathrm{d}t \right\}$$

$$\mathrm{d}X_t = v\left(t, X_t\right) \mathrm{d}t + \sigma \mathrm{d}\mathbb{W}_t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$

$$\mathrm{d}X_t = \left[f\left(t, X_t\right) + g(t) v\left(t, X_t\right)\right] \mathrm{d}t + g(t) \mathrm{d}\mathbb{W}_t$$

with $\sigma = 1$

drift: $\mathbb{R}^d \to \mathbb{R}^d$

diffusion: $\mathbb{R}$

[Caluya+21, Chen+22]

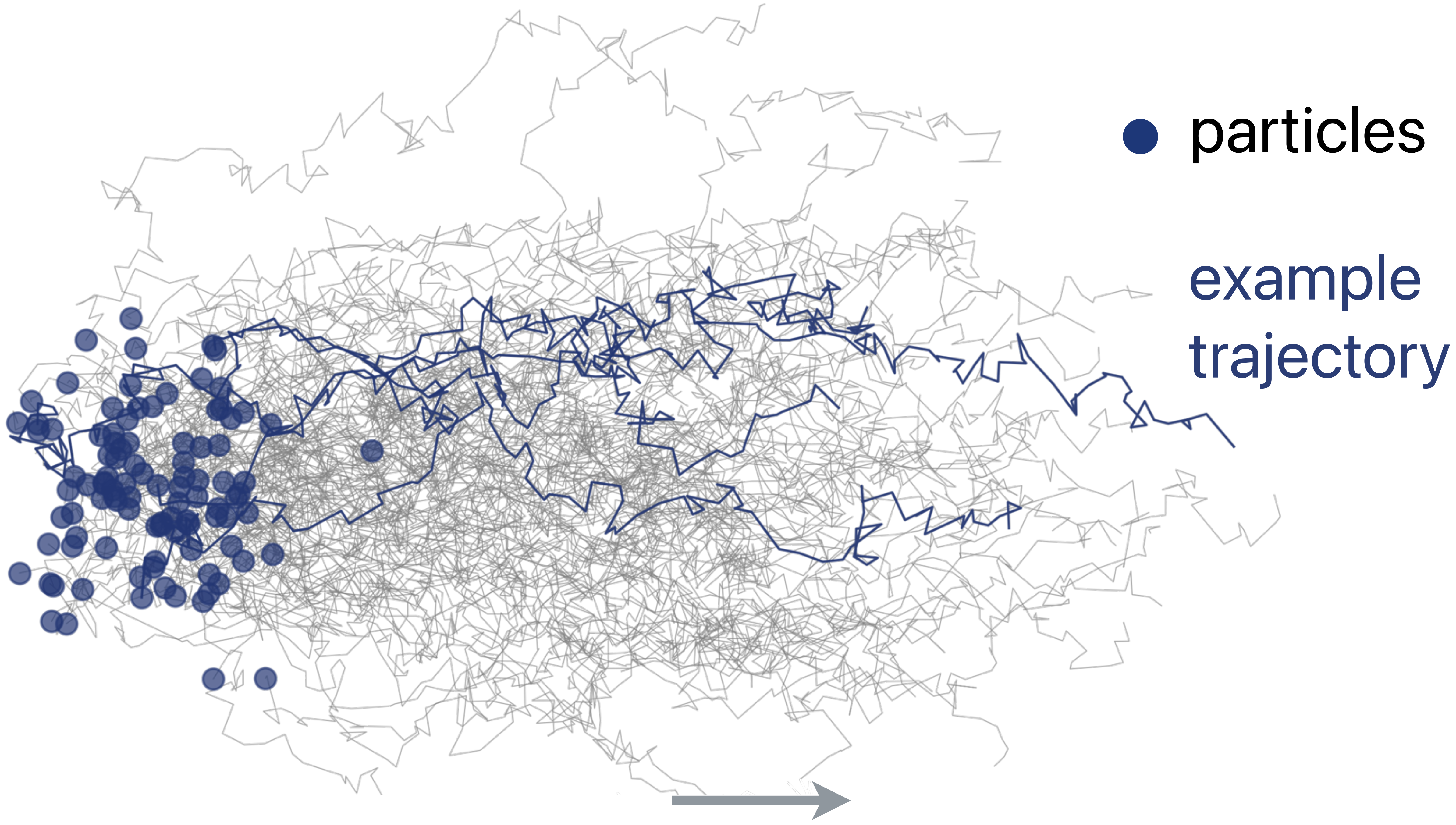
A General SDE Class

Stochastic Control of **Regularized** OT

$$\inf_{v \in \mathcal{V}} \mathbb{E}\left\{\int_0^1 \frac{1}{2}\|v(t, X_t)\|^2 \, \mathrm{d}t\right\}$$

$$\mathrm{d}X_t = v(t, X_t)\,\mathrm{d}t + \sigma \mathrm{d}\mathbb{W}_t$$

$$X_0 \sim \mu_0, \quad X_1 \sim \mu_1$$



particle evolution with **drift** f and diffusivity g

$$\mathrm{d}X_t = [f(t, X_t) + g(t) v(t, X_t)]\,\mathrm{d}t + g(t)\mathrm{d}\mathbb{W}_t$$

with $\sigma = 1$

drift $\mathbb{R}^d \to \mathbb{R}^d$

diffusion $\mathbb{R}$

[Caluya+21, Chen+22]

Solving Schrödinger Bridges

Schrödinger bridge optimality given by [Léonard+13]

$$* \begin{cases} \dfrac{\partial \Phi}{\partial t} = -\nabla \Phi^\top f - \dfrac{1}{2}\sigma^2 g^2 \Delta \Phi \\ \dfrac{\partial \widehat{\Phi}}{\partial t} = -\nabla \cdot (\widehat{\Phi} f) + \dfrac{1}{2}\sigma^2 g^2 \Delta \widehat{\Phi} \end{cases} \quad \text{s.t.} \quad \begin{aligned} \Phi(0,\cdot)\widehat{\Phi}(0,\cdot) &= \mu_0 \\ \Phi(1,\cdot)\widehat{\Phi}(1,\cdot) &= \mu_1 \end{aligned}$$

Solution of $*$ can be expressed via two SDEs of the form

$$\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\left(t, X_t\right)\right] \mathrm{d}t + g \; \mathrm{d}\mathbb{W}_t, \quad X_0 \sim \mu_0$$

$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\left(t, X_t\right)\right] \mathrm{d}t + g \; \mathrm{d}\mathbb{W}_t, \quad X_1 \sim \mu_1$$

Solving Schrödinger Bridges

Schrödinger bridge optimality given by [Léonard+13]

$$
* \begin{cases} \dfrac{\partial \Phi}{\partial t} = -\nabla \Phi^\top f - \dfrac{1}{2}\sigma^2 g^2 \Delta \Phi \\ \dfrac{\partial \widehat{\Phi}}{\partial t} = -\nabla \cdot (\widehat{\Phi} f) + \dfrac{1}{2}\sigma^2 g^2 \Delta \widehat{\Phi} \end{cases} \quad \text{s.t.} \quad \begin{aligned} \Phi(0,\cdot)\widehat{\Phi}(0,\cdot) &= \mu_0 \\ \Phi(1,\cdot)\widehat{\Phi}(1,\cdot) &= \mu_1 \end{aligned}
$$

Solution of $*$ can be expressed via two SDEs of the form

forward policy

$$
\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t, \quad X_0 \sim \mu_0
$$

$$
\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t, \quad X_1 \sim \mu_1
$$

Solving Schrödinger Bridges

Schrödinger bridge optimality given by [Léonard+13]

$$
* \begin{cases} \dfrac{\partial \Phi}{\partial t} = -\nabla \Phi^\top f - \dfrac{1}{2}\sigma^2 g^2 \Delta \Phi \\ \dfrac{\partial \widehat{\Phi}}{\partial t} = -\nabla \cdot (\widehat{\Phi} f) + \dfrac{1}{2}\sigma^2 g^2 \Delta \widehat{\Phi} \end{cases} \quad \text{s.t.} \quad \begin{aligned} \Phi(0,\cdot)\widehat{\Phi}(0,\cdot) &= \mu_0 \\ \Phi(1,\cdot)\widehat{\Phi}(1,\cdot) &= \mu_1 \end{aligned}
$$

Solution of $*$ can be expressed via two SDEs of the form

forward policy

$$
\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t, \quad X_0 \sim \mu_0
$$

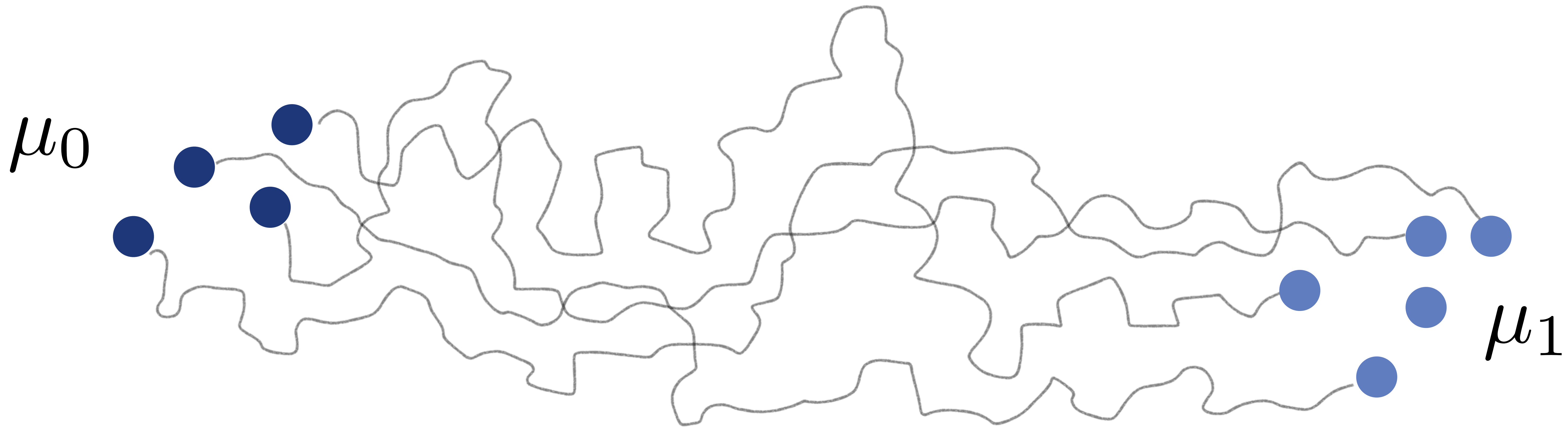
$$
\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t, \quad X_1 \sim \mu_1
$$

backward policy

Solving Schrödinger Bridges

The solution of the SB is a system of a forward and backward SDE

$$\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\left(t, X_t\right)\right] \mathrm{d}t + g \ \mathrm{d}\mathbb{W}_t$$



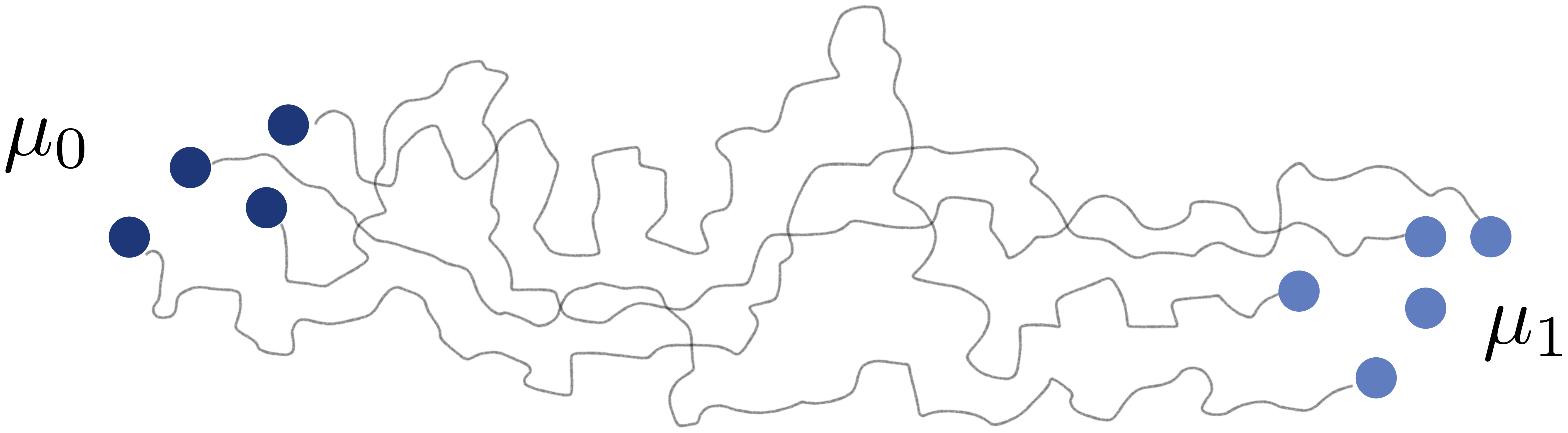
reverse *SDE* $$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\left(t, X_t\right)\right] \mathrm{d}t + g \ \mathrm{d}\mathbb{W}_t$$

[Anderson82]

Solving Schrödinger Bridges

The solution of the SB is a system of a forward and backward SDE

$$X_0 \sim \mu_0 \qquad \mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$



reverse *SDE* $$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$

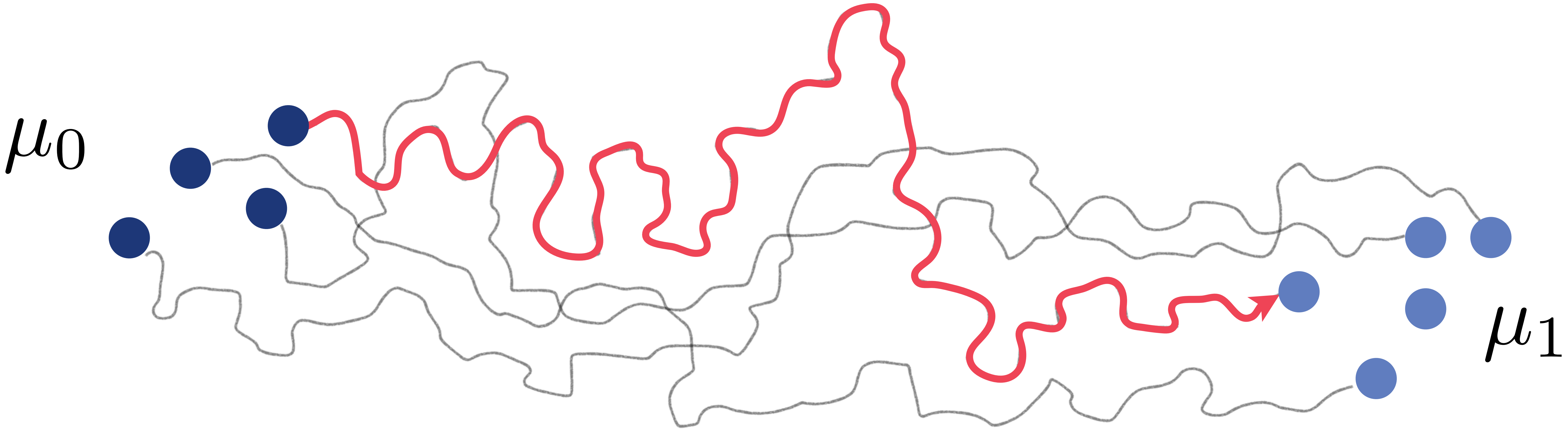
[Anderson82]

Solving Schrödinger Bridges

The solution of the SB is a system of a forward and backward SDE

forward policy

$$X_0 \sim \mu_0 \qquad \mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\left(t, X_t\right)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$



reverse *SDE* $$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\left(t, X_t\right)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$

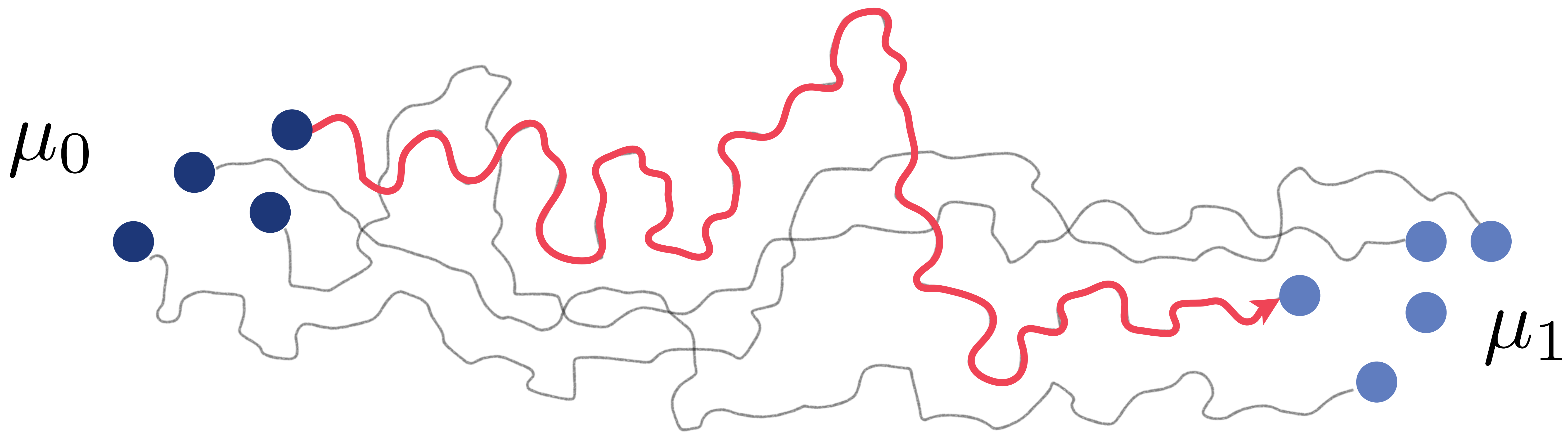
[Anderson82]

Solving Schrödinger Bridges

The solution of the SB is a system of a forward and backward SDE

forward policy

$$X_0 \sim \mu_0 \qquad \mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\left(t, X_t\right)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$



reverse *SDE* $$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\left(t, X_t\right)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t \qquad X_1 \sim \mu_1$$

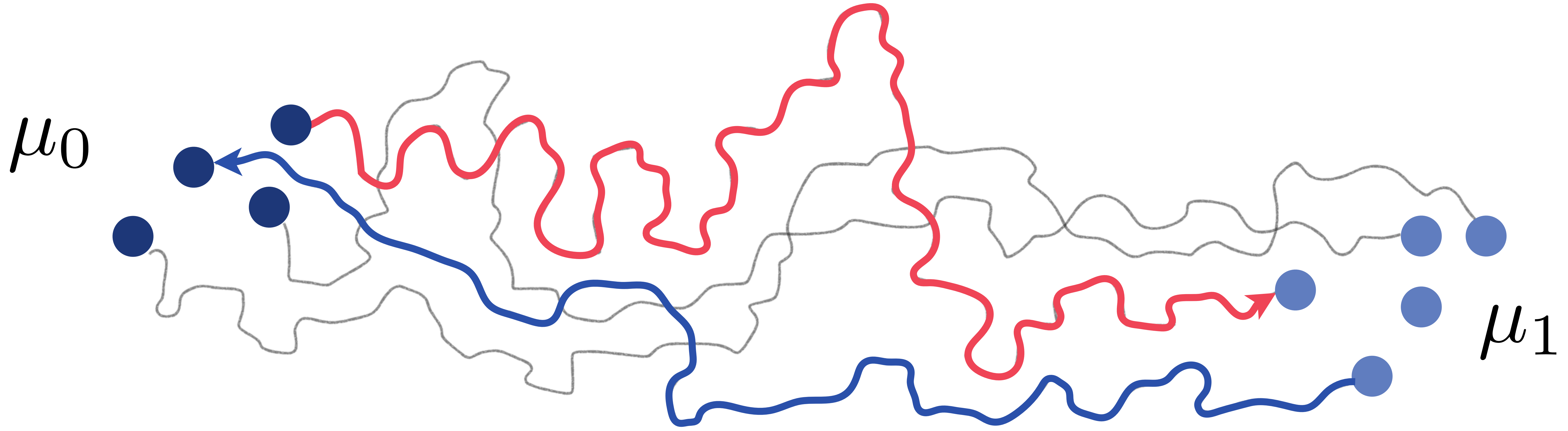
[Anderson82]

Solving Schrödinger Bridges

The solution of the SB is a system of a forward and backward SDE

forward policy

$$X_0 \sim \mu_0 \qquad \mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\left(t, X_t\right)\right] \mathrm{d}t + g \ \mathrm{d}\mathbb{W}_t$$



reverse *SDE* $\qquad \mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\left(t, X_t\right)\right] \mathrm{d}t + g \ \mathrm{d}\mathbb{W}_t \qquad X_1 \sim \mu_1$

backward policy

[Anderson82]

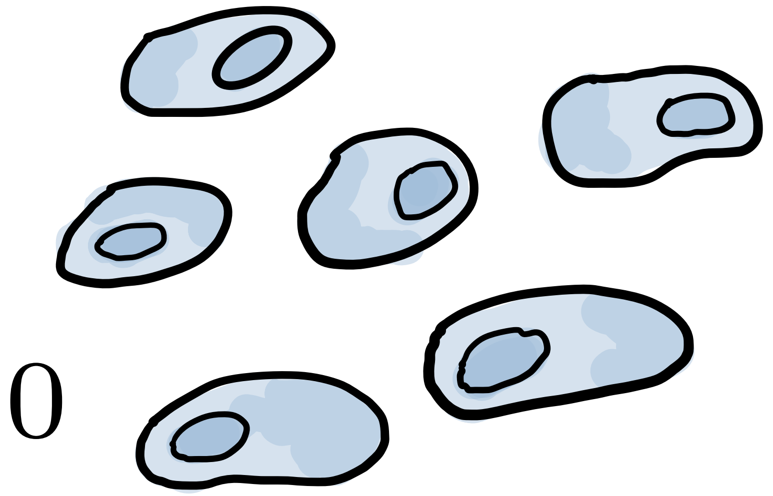
A Link to Score-Based Generative Models

Class of powerful generative models: **Score-Based Generative Models**

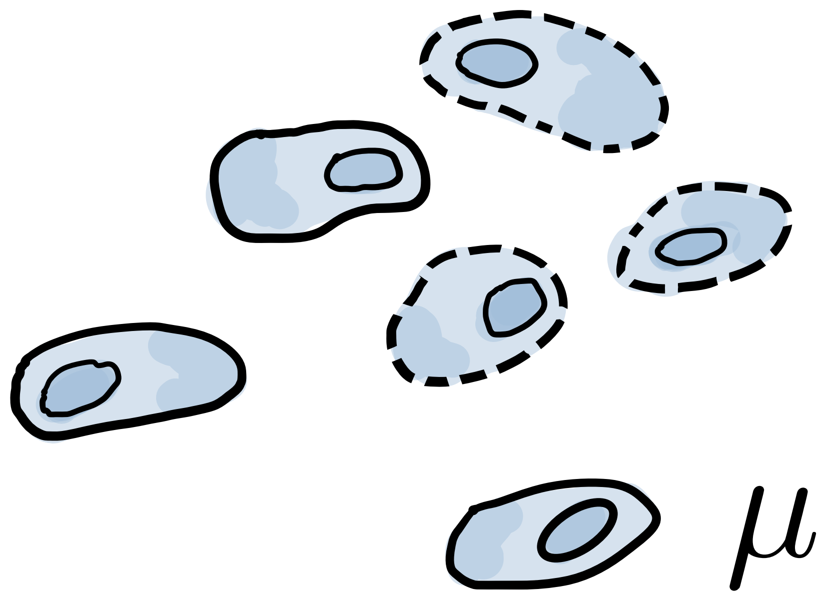
[Song+21, Ho+20, Hyvärinen05]

forward drift

$$\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\left(t, X_t\right)\right] \mathrm{d}t + g \; \mathrm{d}\mathbb{W}_t$$



μ_0



μ_1

$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}\left(t, X_t\right)\right] \mathrm{d}t + g \; \mathrm{d}\mathbb{W}_t$$

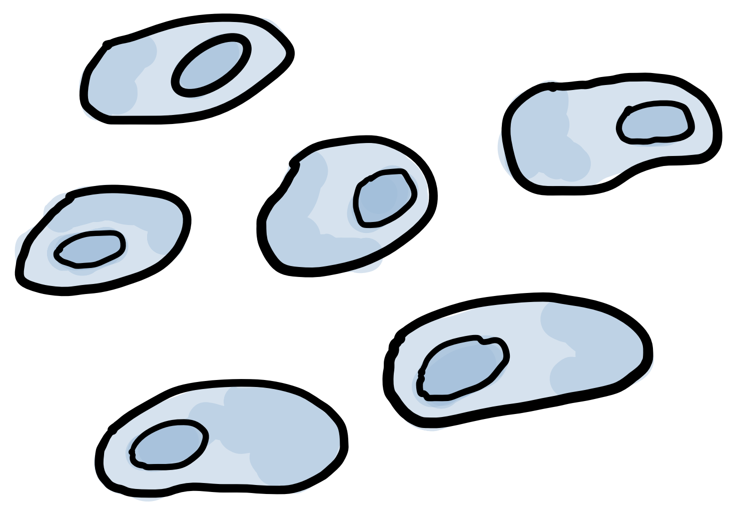
backward drift

A Link to Score-Based Generative Models

Class of powerful generative models: **Score-Based Generative Models**

[Song+21, Ho+20, Hyvärinen05]

data distribution



μ

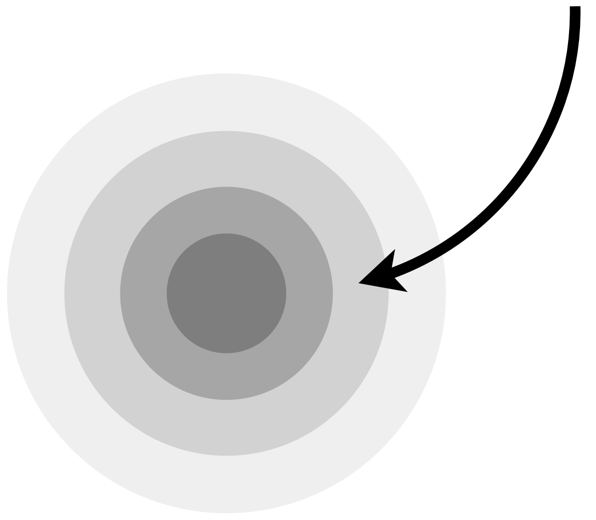
forward drift

$$dX_t = \left[f + g^2 \nabla \log \Phi\,(t, X_t)\right] dt + g\, d\mathbb{W}_t$$

$$dX_t = \left[f - g^2 \nabla \log \widehat{\Phi}\,(t, X_t)\right] dt + g\, d\mathbb{W}_t$$

backward drift

easy to sample



$\mathcal{N}(0, I)$

A Link to Score-Based Generative Models

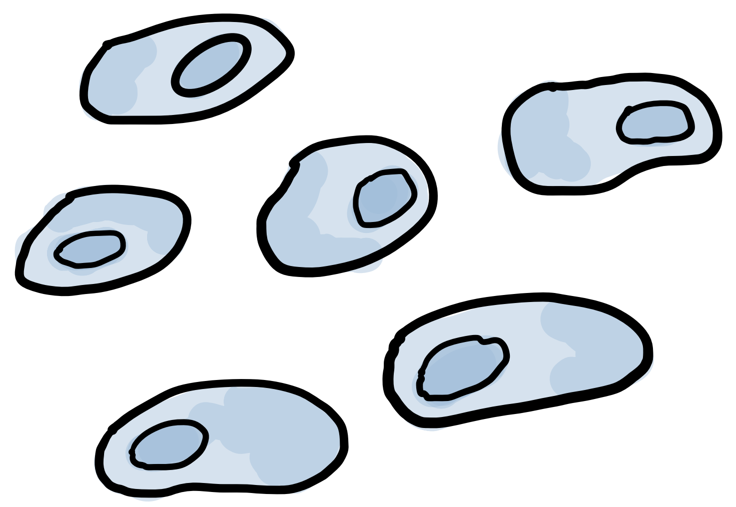
Class of powerful generative models: **Score-Based Generative Models**

[Song+21, Ho+20, Hyvärinen05]

forward drift

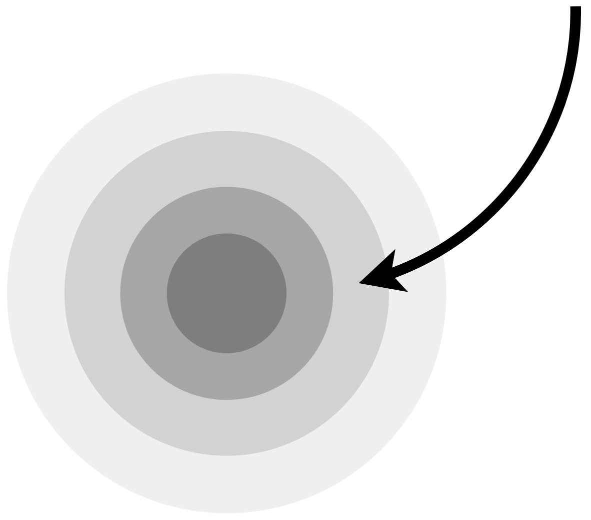
$$\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi(t, X_t)\right] \mathrm{d}t + g \, \mathrm{d}\mathbb{W}_t$$

data distribution



μ

easy to sample



$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \widehat{\Phi}(t, X_t)\right] \mathrm{d}t + g \, \mathrm{d}\mathbb{W}_t$$

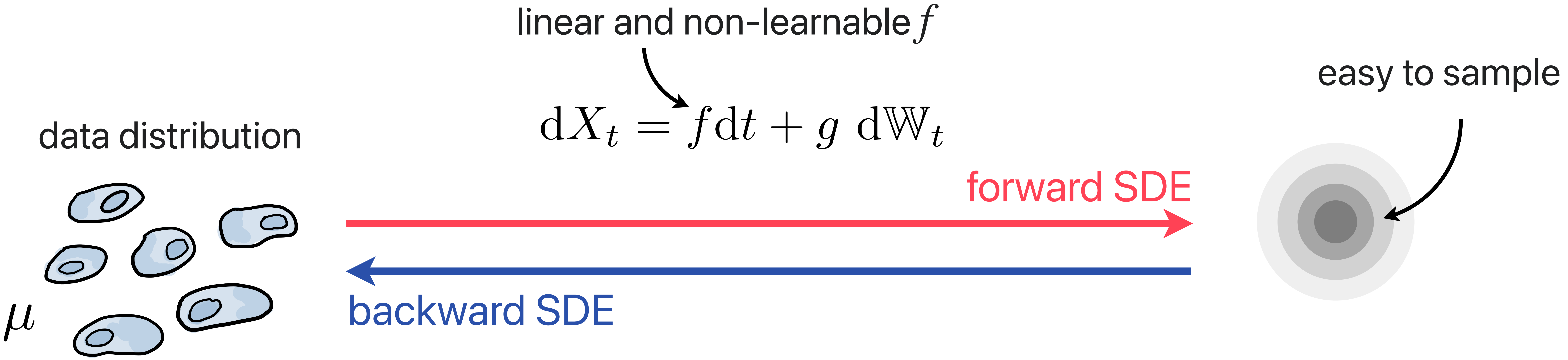
$\mathcal{N}(0, I)$

backward drift

A Link to Score-Based Generative Models

Class of powerful generative models: **Score-Based Generative Models**

[Song+21, Ho+20, Hyvärinen05]



A Link to Score-Based Generative Models

Class of powerful generative models: **Score-Based Generative Models**

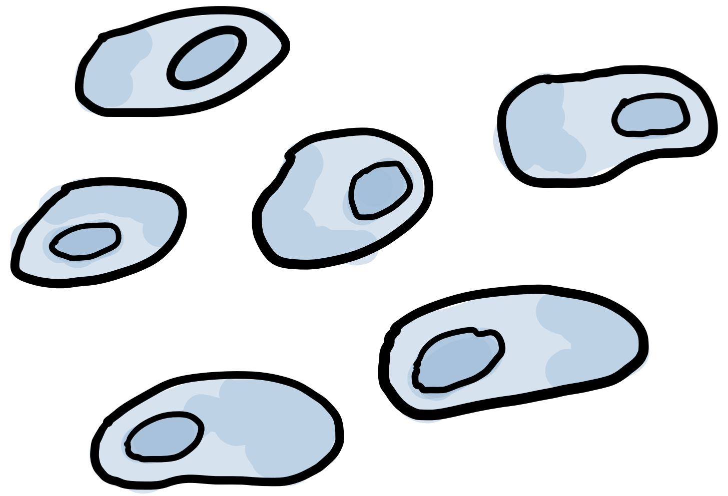
[Song+21, Ho+20, Hyvärinen05]

linear and non-learnable f

$$\mathrm{d}X_t = f\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t$$

easy to sample

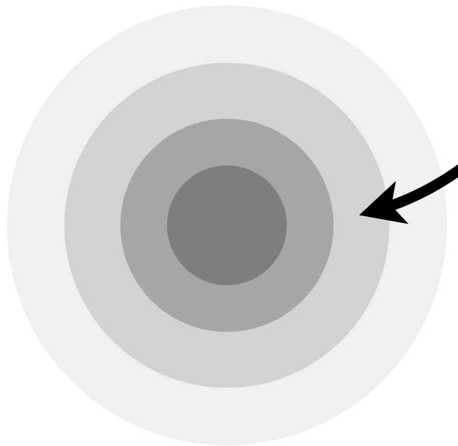
data distribution



μ

forward SDE

backward SDE



$\mathcal{N}(0, I)$

$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t$$

score function

= marginal density of forward diffusion at time t

tractable objective

[Anderson82]

A Link to Score-Based Generative Models

Class of powerful generative models: **Score-Based Generative Models**

[Song+21, Ho+20, Hyvärinen05]

linear and non-learnable f

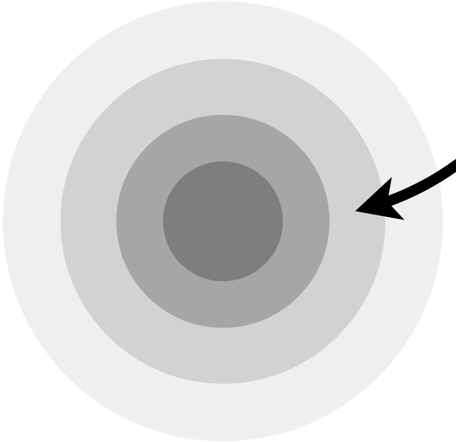
$$\mathrm{d}X_t = f\,\mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$

forward SDE

backward SDE

$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$

easy to sample



$\mathcal{N}(0, I)$

tractable objective

score function

= marginal density of forward diffusion at time *t*

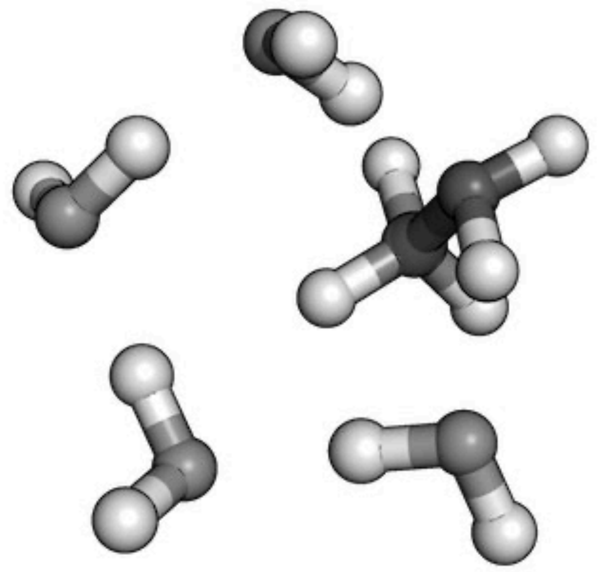
[Anderson82]

A Link to Score-Based Generative Models

Class of powerful generative models: **Score-Based Generative Models**

[Song+21, Ho+20, Hyvärinen05]

Molecular Conformers



[Corso+22]

linear and non-learnable f

$$\mathrm{d}X_t = f\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t$$

forward SDE

backward SDE

$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t$$

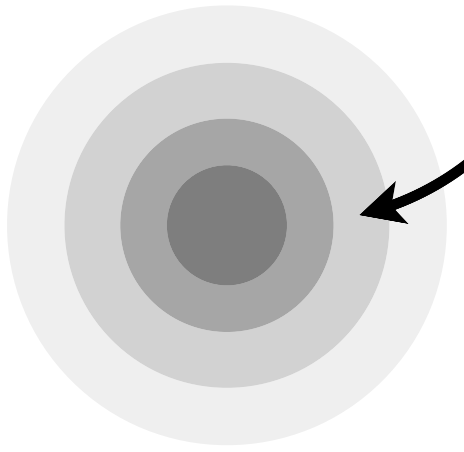
tractable objective

score function

= marginal density of forward diffusion at time t

[Anderson82]

easy to sample



$\mathcal{N}(0, I)$

Solving Schrödinger Bridges

Understand the stochastic process between two data distributions μ_0 and μ_1

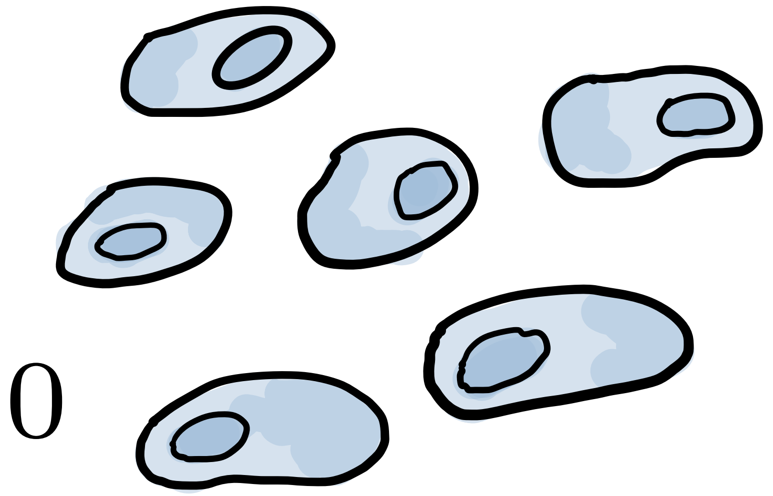
linear and non-learnable f

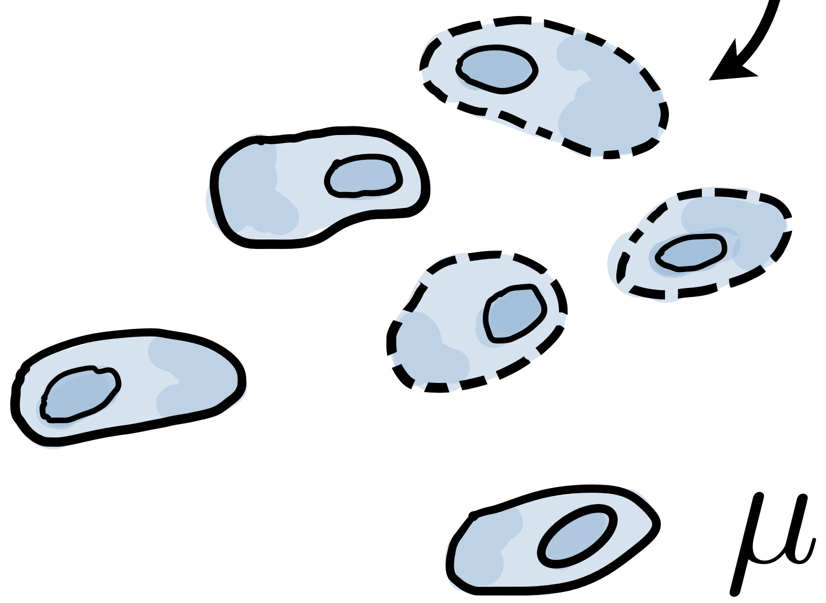
$$\mathrm{d}X_t = f\,\mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t$$

easy to sample

forward SDE

μ_0





μ_1

backward SDE

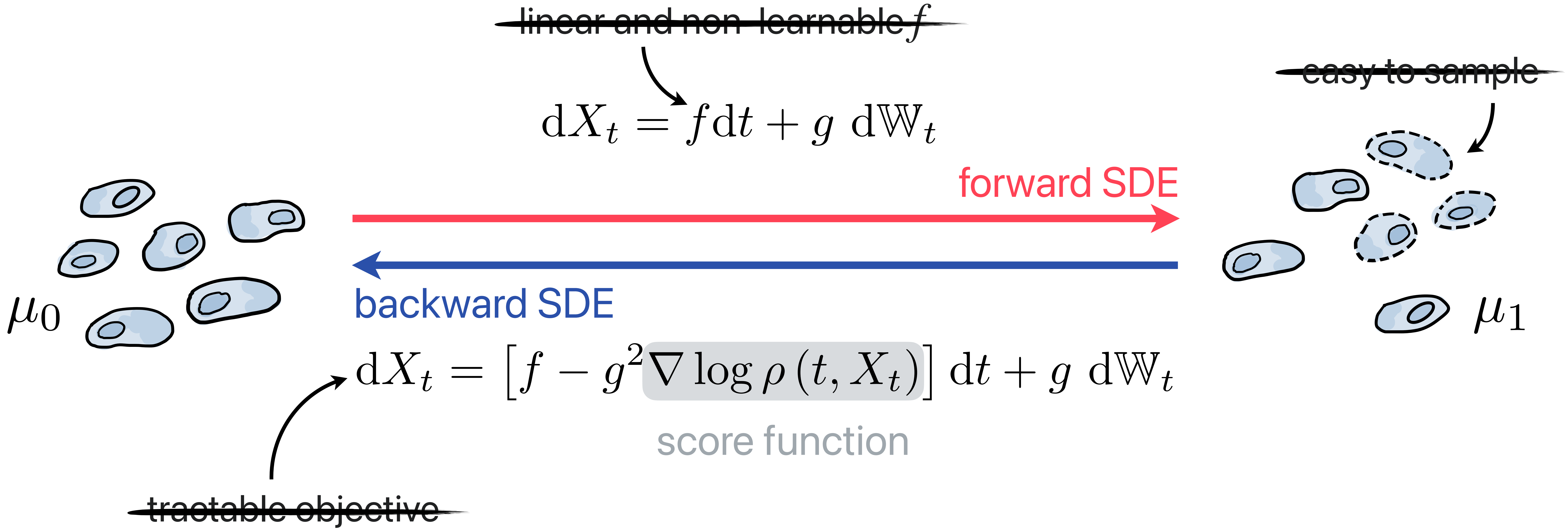
$$\mathrm{d}X_t = \left[f - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\,\mathrm{d}\mathbb{W}_t$$

score function

tractable objective

Solving Schrödinger Bridges

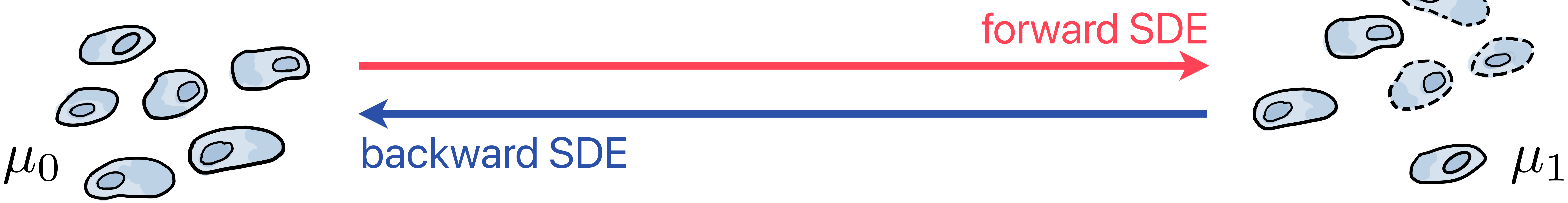
Understand the stochastic process between two data distributions μ_0 and μ_1



Diffusion Schrödinger Bridges

Find a non-linear drift such that $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. [De Bortoli+21, Vargas+21, Chen+22]

$$\mathrm{d}X_t = [f + \square]\mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$



$$\mathrm{d}X_t = \left[f + \square - g^2 \nabla \log \rho\,(t, X_t)\right]\mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$

Diffusion Schrödinger Bridges

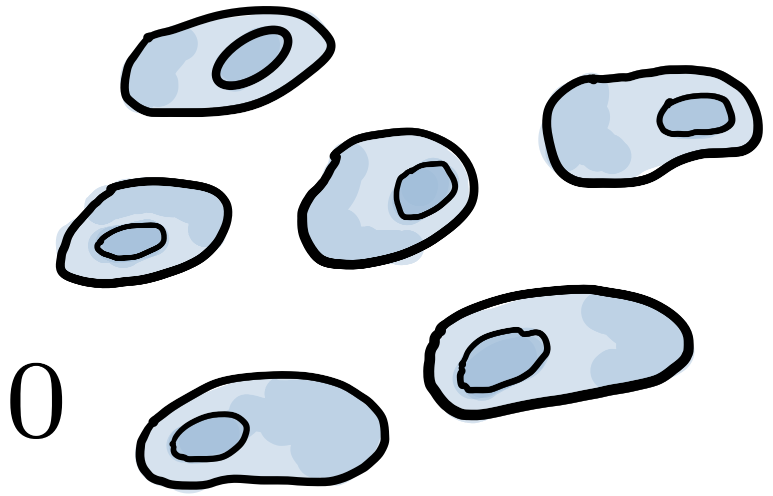
Find a non-linear drift such that $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. [De Bortoli+21, Vargas+21, Chen+22]

$$\square = g^2 \nabla \log \Phi\,(t, X_t)$$

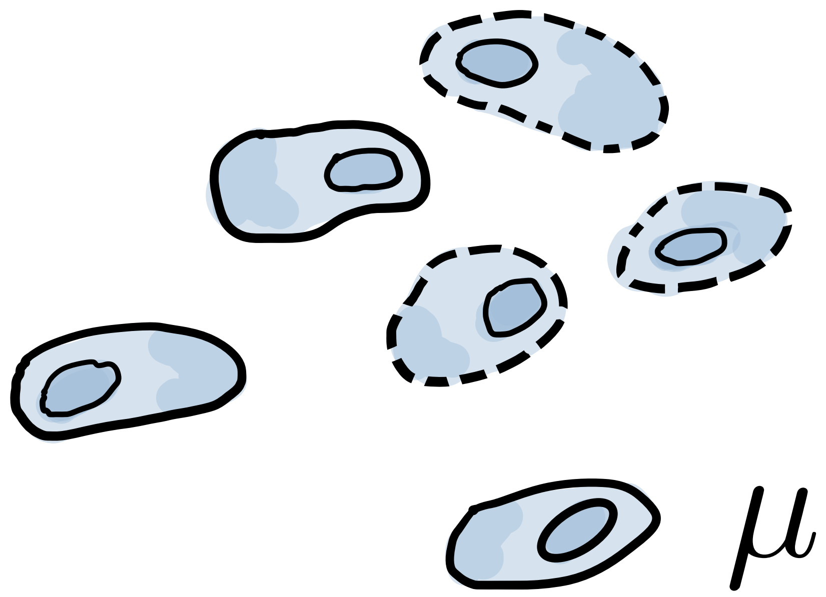
$$\mathrm{d}X_t = [f + \square]\mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$

forward SDE

μ_0



backward SDE



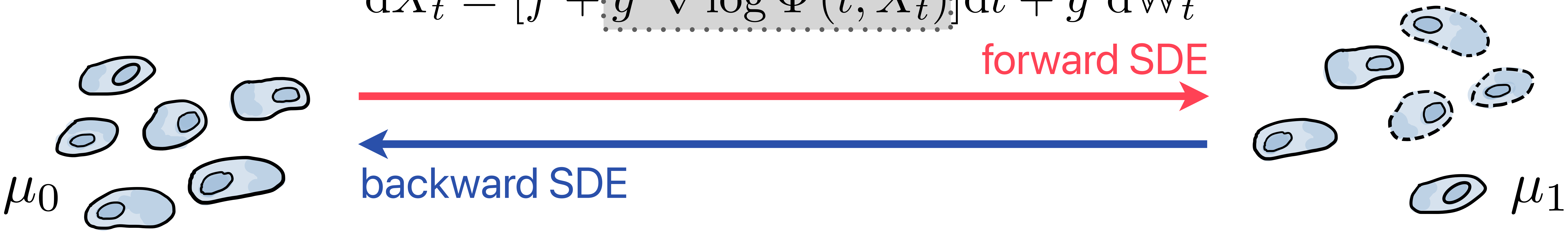
μ_1

$$\mathrm{d}X_t = \left[f + \square - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$

Diffusion Schrödinger Bridges

Find a non-linear drift such that $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. [De Bortoli+21, Vargas+21, Chen+22]

$$\mathrm{d}X_t = \left[f + \boxed{g^2 \nabla \log \Phi\,(t, X_t)}\right]\mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$

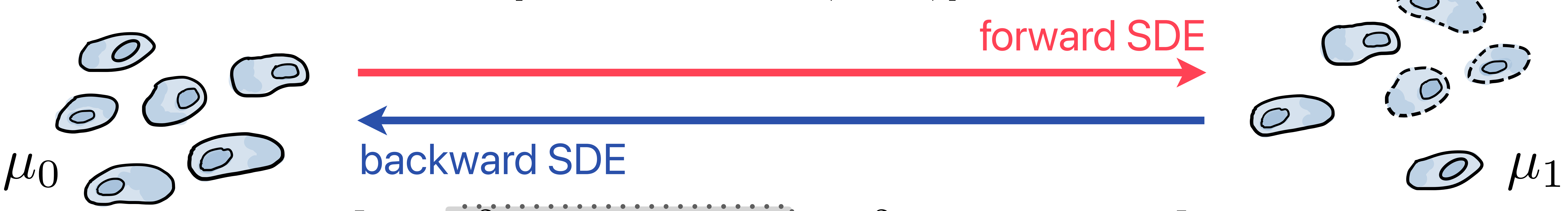


$$\mathrm{d}X_t = \left[f + \boxed{} - g^2 \nabla \log \rho\,(t, X_t)\right]\mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$

Diffusion Schrödinger Bridges

Find a non-linear drift such that $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. [De Bortoli+21, Vargas+21, Chen+22]

$$\mathrm{d}X_t = [f + g^2 \nabla \log \Phi\,(t, X_t)]\mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$



$$\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\,(t, X_t) - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\;\mathrm{d}\mathbb{W}_t$$

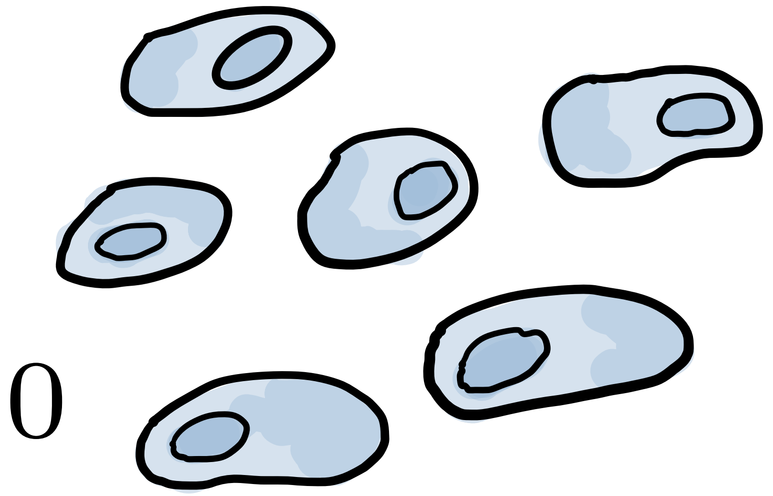
Diffusion Schrödinger Bridges

Find a non-linear drift such that $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. [De Bortoli+21, Vargas+21, Chen+22]

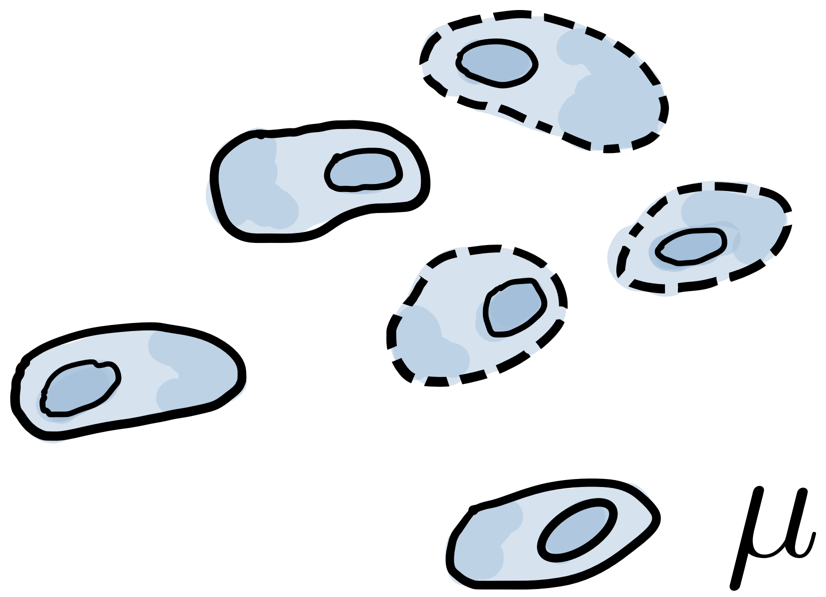
$$dX_t = [f + g^2 \nabla \log \Phi(t, X_t)]dt + g \, d\mathbb{W}_t$$

forward SDE

μ_0



backward SDE



μ_1

$$dX_t = \left[f + g^2 \nabla \log \Phi(t, X_t) - g^2 \nabla \log \rho(t, X_t)\right] dt + g \, d\mathbb{W}_t$$

$$:= -g^2 \nabla \log \hat{\Phi}(t, X_t)$$

Diffusion Schrödinger Bridges

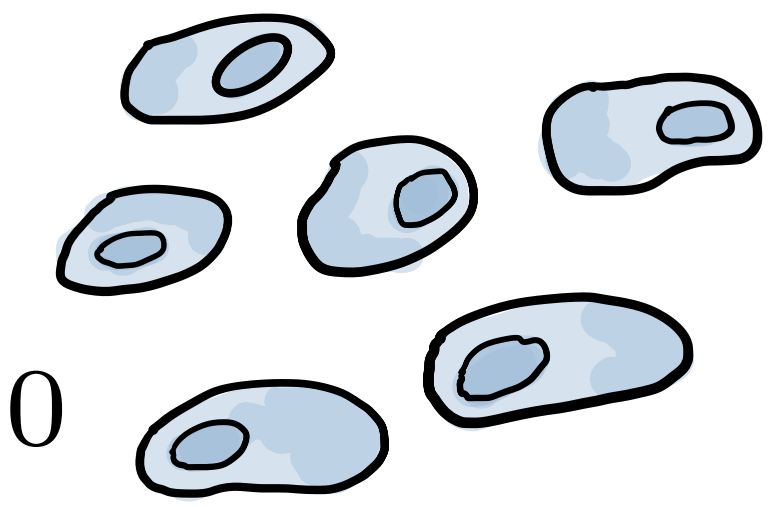
Find a non-linear drift such that $X_0 \sim \mu_0$ and $X_1 \sim \mu_1$. [De Bortoli+21, Vargas+21, Chen+22]

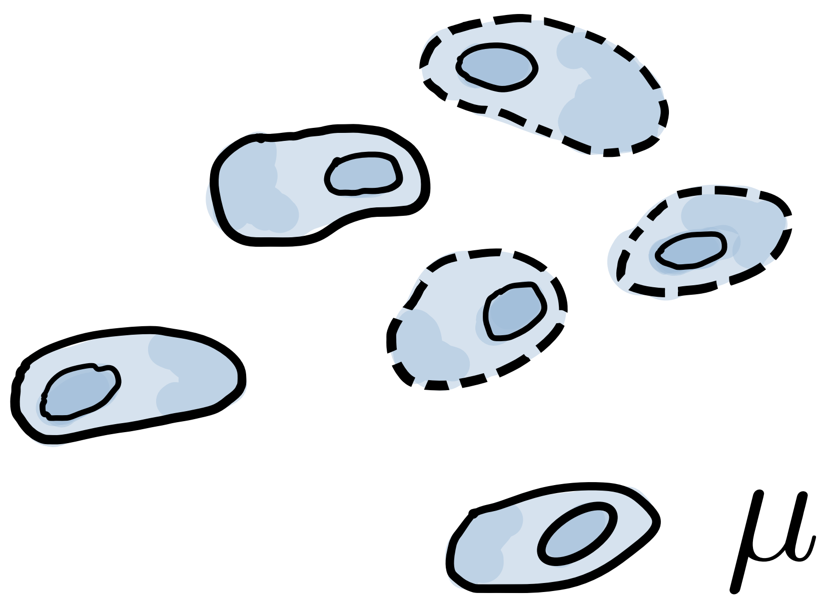
parameterize forward drift Z_θ

$$\mathrm{d}X_t = [f + g^2 \nabla \log \Phi\,(t, X_t)]\mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$

forward SDE

μ_0





μ_1

backward SDE

$$\mathrm{d}X_t = \left[f + g^2 \nabla \log \Phi\,(t, X_t) - g^2 \nabla \log \rho\,(t, X_t)\right] \mathrm{d}t + g\ \mathrm{d}\mathbb{W}_t$$

$$:= -g^2 \nabla \log \hat{\Phi}\,(t, X_t)$$

parameterize backward drift $\hat{Z}_\phi$

Sinkhorn-Style Solvers for Diffusion Models

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \qquad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \quad \text{and} \quad \mathbb{P}_1 = \mu_1$$

Sinkhorn-Style Solvers for Diffusion Models

Part 3

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \ \text{ and } \ \mathbb{P}_1 = \mu_1$$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$

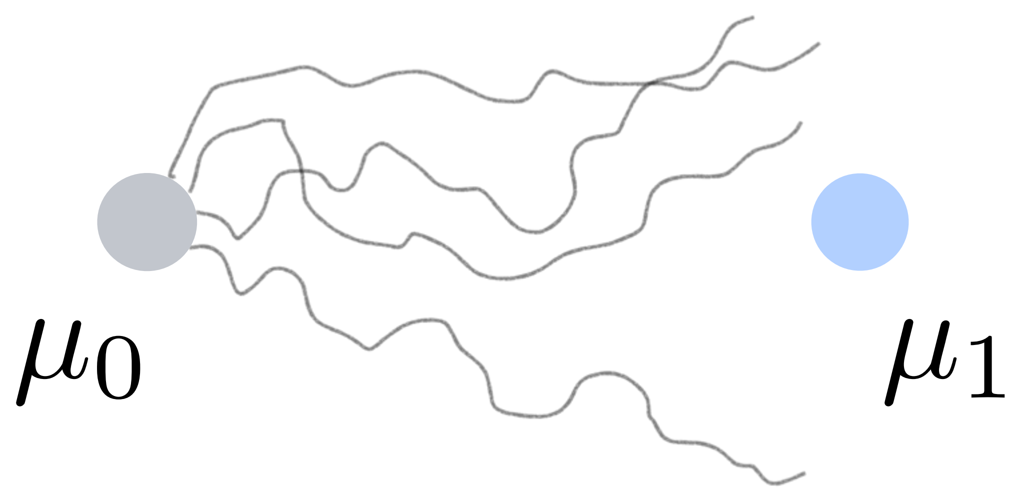
Sinkhorn-Style Solvers for Diffusion Models

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \text{ and } \mathbb{P}_1 = \mu_1$$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$



Sinkhorn-Style Solvers for Diffusion Models

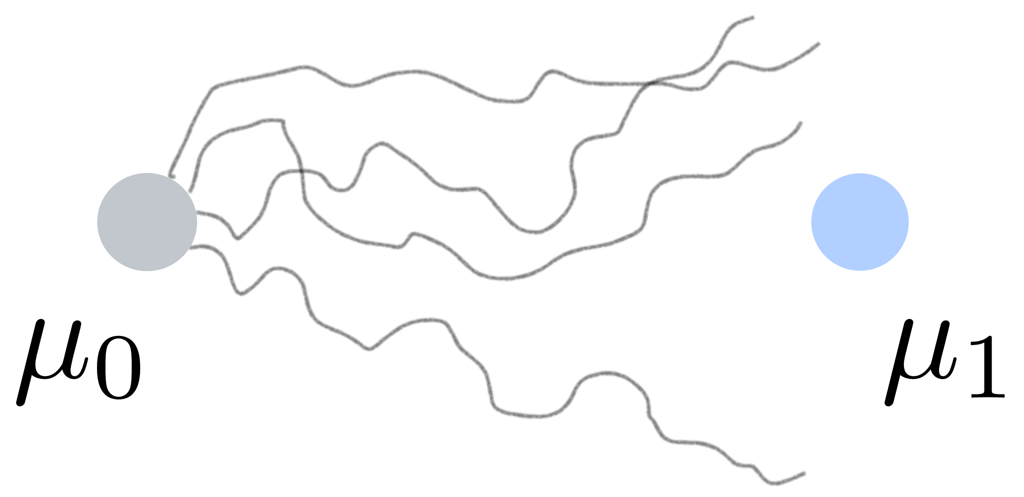
Part 3

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \text{ and } \mathbb{P}_1 = \mu_1$$

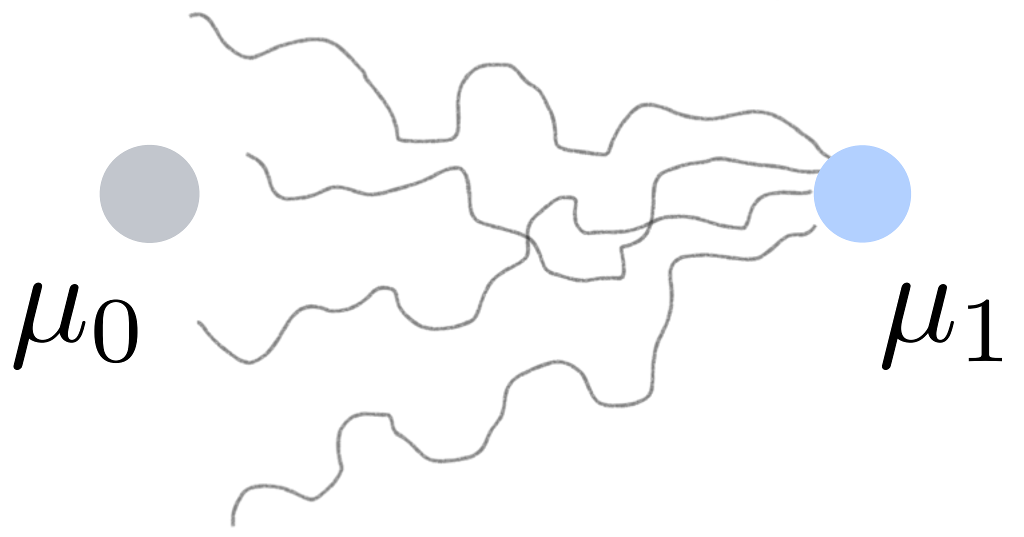
Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$



$$\mathbb{P}_t^{2n+2} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n+1}\right) \quad \text{s.t.} \quad \mathbb{P}_1 = \mu_1$$



Sinkhorn-Style Solvers for Diffusion Models

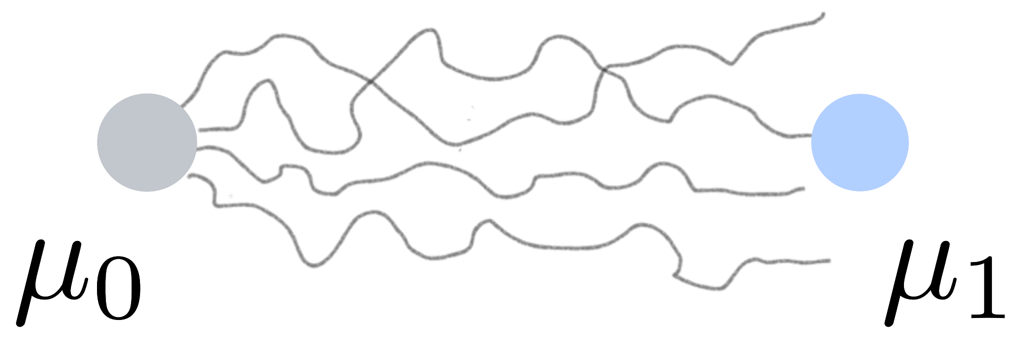
Part 3

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \text{ and } \mathbb{P}_1 = \mu_1$$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$



Sinkhorn-Style Solvers for Diffusion Models

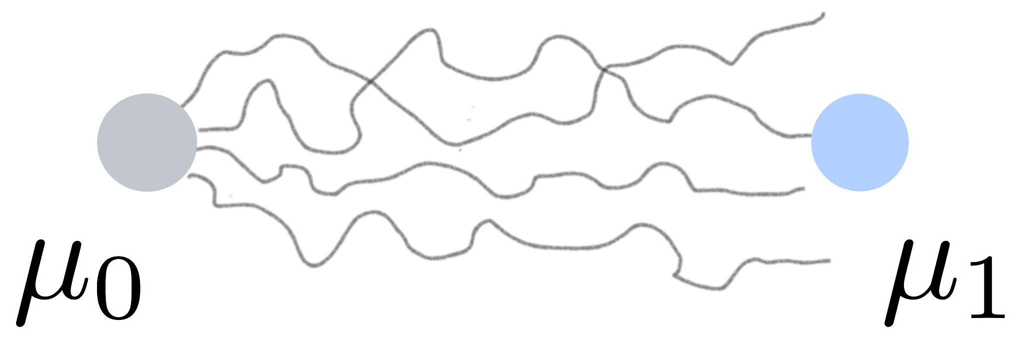
Part 3

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \text{ and } \mathbb{P}_1 = \mu_1$$

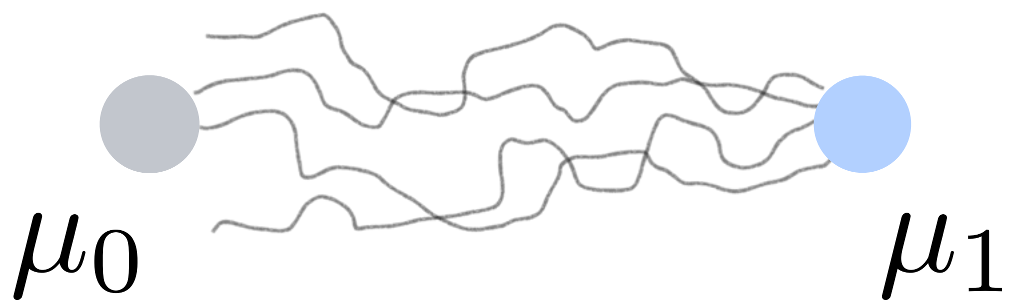
Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$



$$\mathbb{P}_t^{2n+2} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n+1}\right) \quad \text{s.t.} \quad \mathbb{P}_1 = \mu_1$$



Sinkhorn-Style Solvers for Diffusion Models

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right)$$ s.t. $\mathbb{P}_0 = \mu_0$ and $\mathbb{P}_1 = \mu_1$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right)$$ s.t. $\mathbb{P}_0 = \mu_0$

Sinkhorn-Style Solvers for Diffusion Models

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \ \text{ and } \ \mathbb{P}_1 = \mu_1$$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$

$$\mathbb{P}_t^{2n+2} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n+1}\right) \quad \text{s.t.} \quad \mathbb{P}_1 = \mu_1$$

Sinkhorn-Style Solvers for Diffusion Models

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \ \text{ and } \ \mathbb{P}_1 = \mu_1$$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$

$$\mathbb{P}_t^{2n+2} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n+1}\right) \quad \text{s.t.} \quad \mathbb{P}_1 = \mu_1$$

Sinkhorn-Style Solvers for Diffusion Models

Part 3

$$\arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{Q}_t\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0 \text{ and } \mathbb{P}_1 = \mu_1$$

Iterative proportional fitting (IPF) [Fortet49, Kullback68, Sinkhorn64]

set $\mathbb{P}_t^0 = \mathbb{Q}_t$

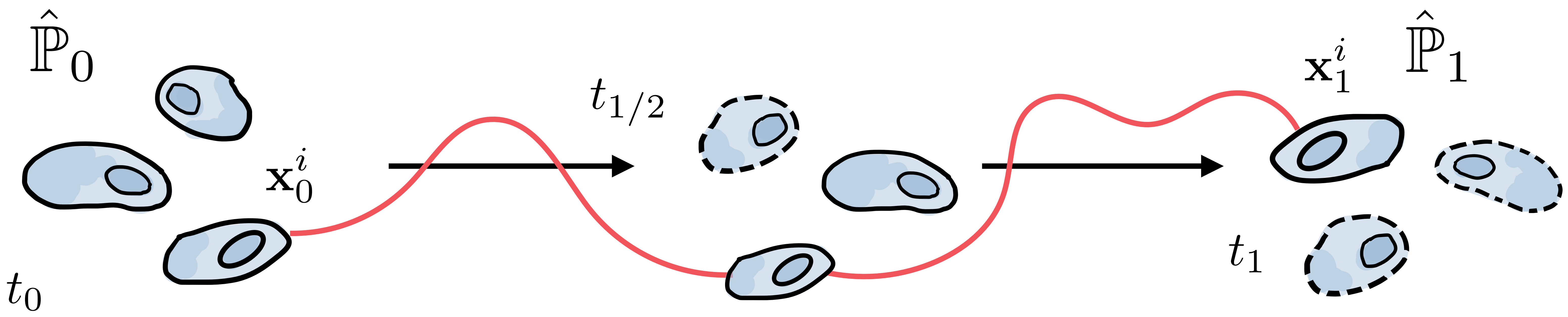
$$\mathbb{P}_t^{2n+1} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n}\right) \quad \text{s.t.} \quad \mathbb{P}_0 = \mu_0$$

$$\mathbb{P}_t^{2n+2} \leftarrow \arg\min_{\mathbb{P}_t} D_{\mathrm{KL}}\left(\mathbb{P}_t \| \mathbb{P}_t^{2n+1}\right) \quad \text{s.t.} \quad \mathbb{P}_1 = \mu_1$$

... estimators are based on Gaussian processes [Vargas+21], dual potentials [Finlay+2020], or neural networks [De Bortoli+21, Chen+22]

Application: **Recovering Continuous Dynamics of Cell Differentiation**

Part 3



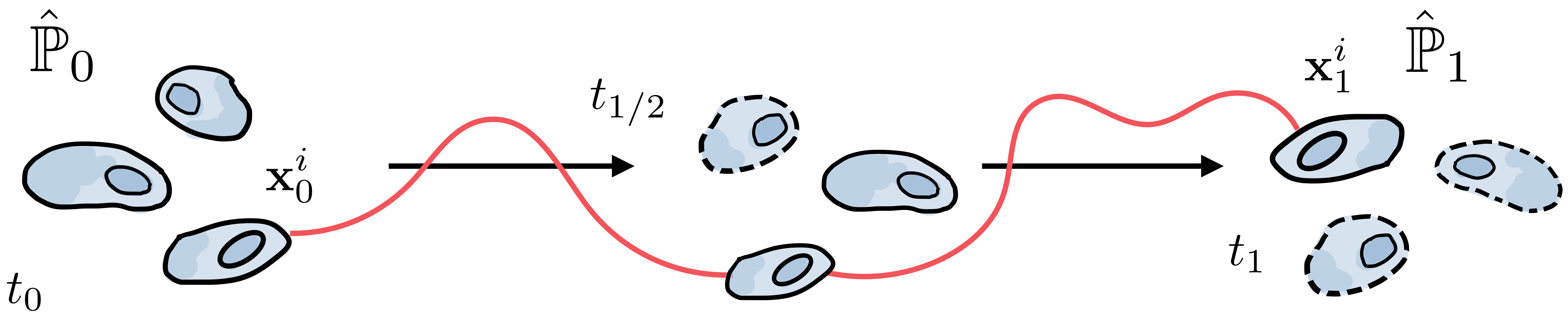
Fate determination in hematopoiesis [Weinreb+20]

- cells on day 2 → $\hat{\mathbb{P}}_0$
- cells on day 4 → $\hat{\mathbb{P}}_1$

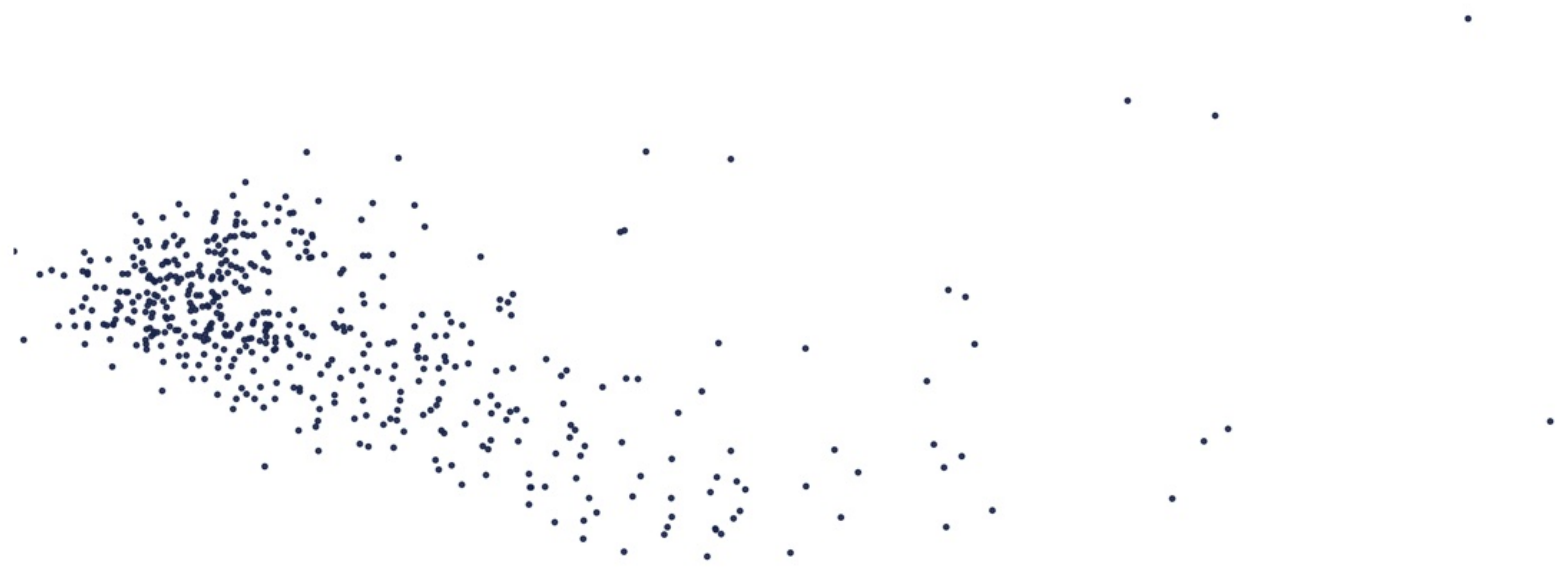
Results in [Somnath+23]

Application: **Recovering Continuous Dynamics of Cell Differentiation**

Part 3



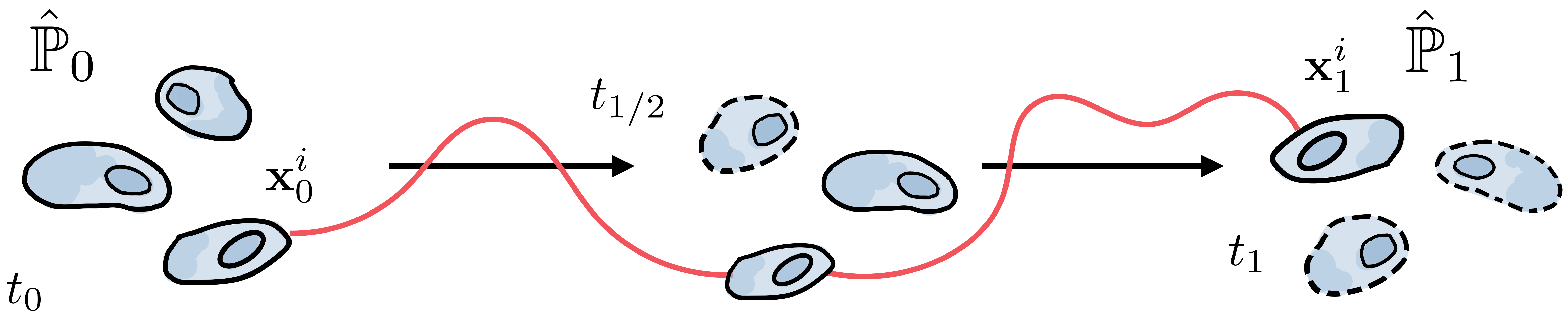
Fate determination in hematopoiesis [Weinreb+20]



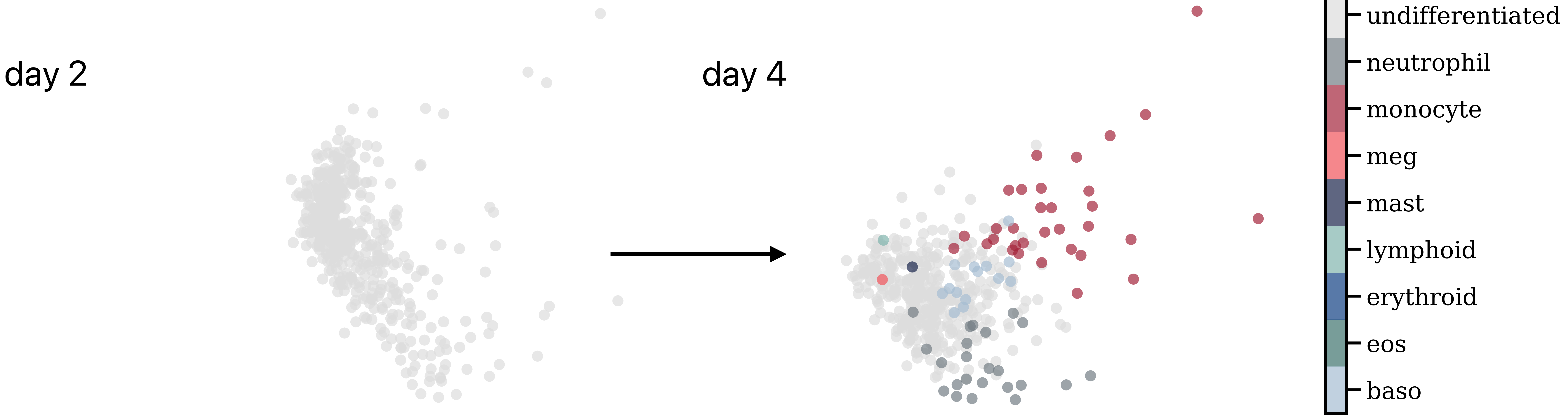
- cells on day 2 → $\hat{\mathbb{P}}_0$
- cells on day 4 → $\hat{\mathbb{P}}_1$

Results in [Somnath+23]

Application: **Recovering Continuous Dynamics of Cell Differentiation**

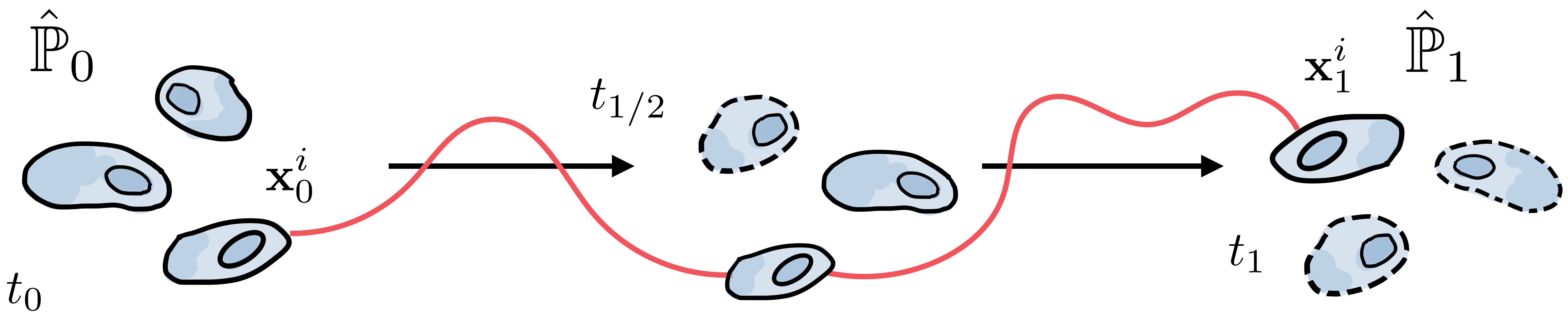


Fate determination in hematopoiesis [Weinreb+20]

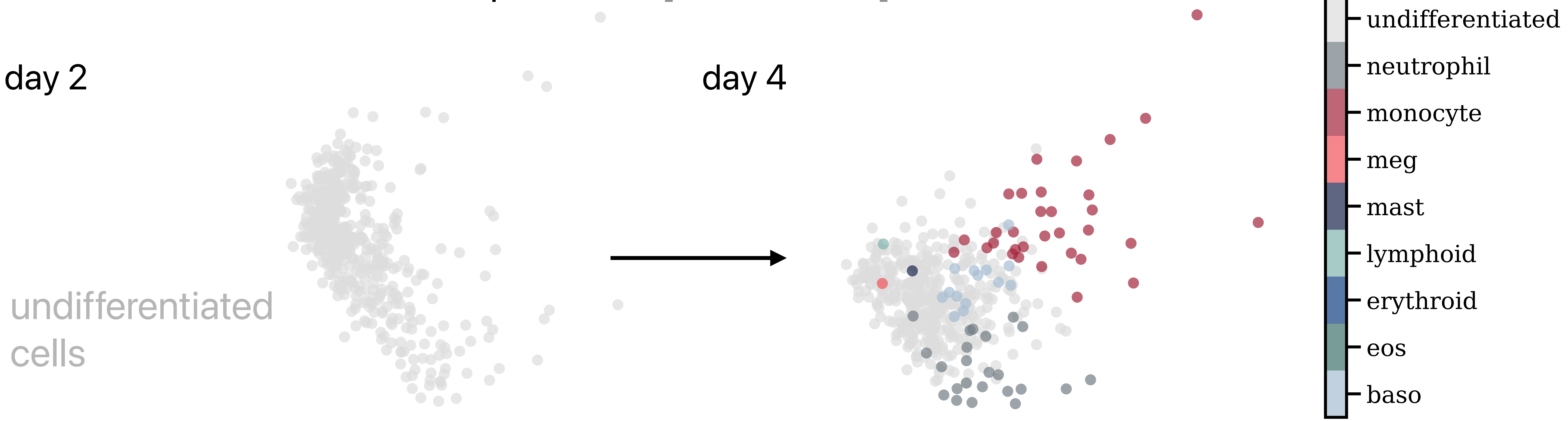


Results in [Somnath+23]

Application: **Recovering Continuous Dynamics of Cell Differentiation**



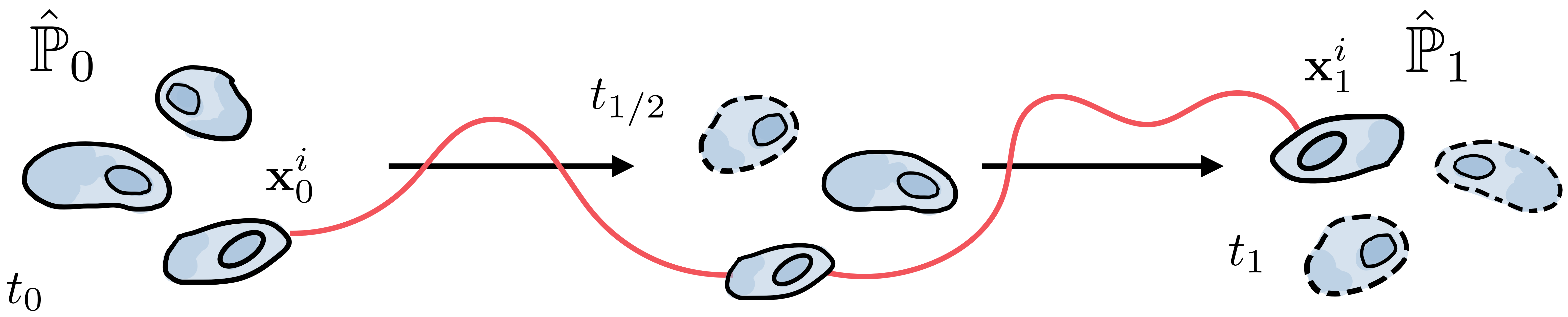
Fate determination in hematopoiesis [Weinreb+20]



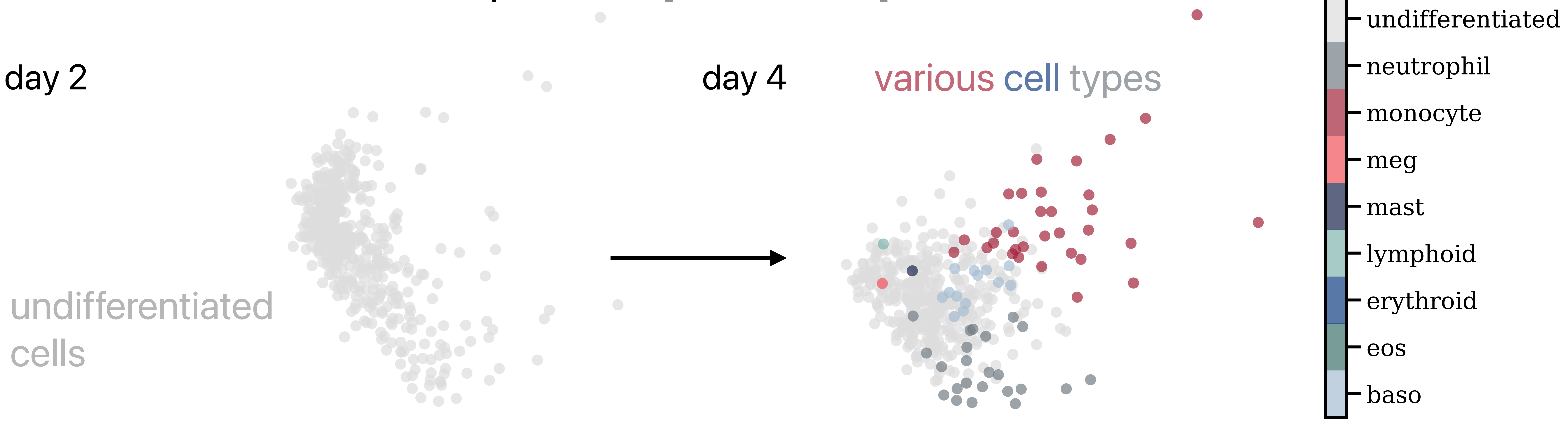
Results in [Somnath+23]

Application: **Recovering Continuous Dynamics of Cell Differentiation**

Part 3



Fate determination in hematopoiesis [Weinreb+20]



Results in [Somnath+23]

Perspective

Alternative Viewpoints

- interpolating spline curves of measures [Chen+18, Benamou+18, Chewi+21]
- links with sampling algorithms [Salim+21, Stromme+23]

Extensions

- to different manifolds [De Bortoli+22, Thornton+22]
- to unbalanced problems [Baradat+21, Chen+22, Pariset+23]
- to multi-marginal problems [Chen+19, Haasler+21, Chen+23, Noble+23]
- to conditional settings [Shi+22]
- to flow matching settings [Shi+23, Liu+23, Somnath+23, Tong+23, Neklyudov+22]

Applications

- to mean-field games [Liu+23]
- in chemistry and physics, e.g., for transition path sampling [Holdijk+22]

Connections of Optimal Transport to Control and Dynamics

Optimal Transport

Maps

Coupling

PDEs **Dynamical Systems**

Gradient Flows

Jordan-Kinderlehrer-Otto Flows

Control Theory

Schrödinger Bridges

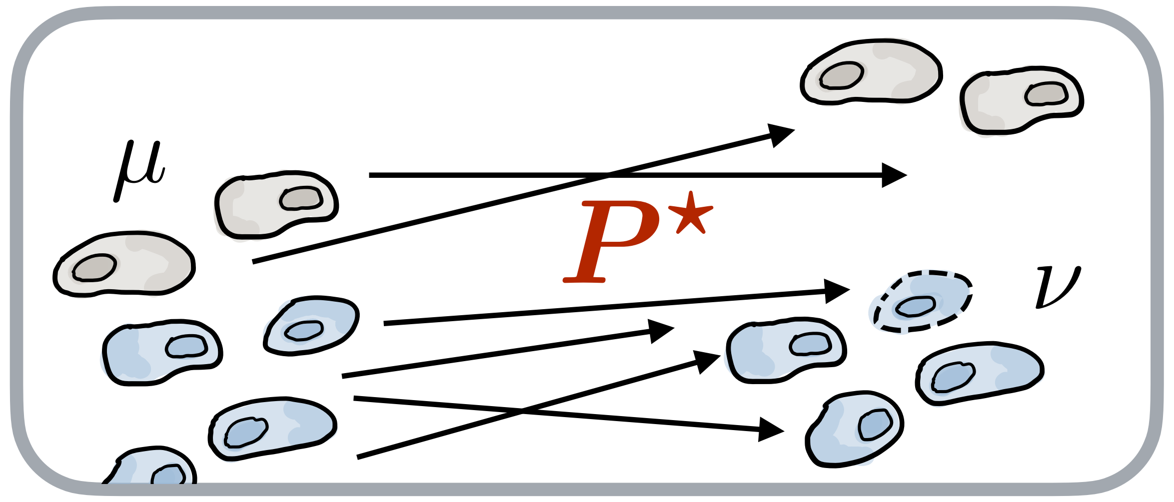
SDEs Diffusion Models

Connections of Optimal Transport to Control and Dynamics

Optimal Transport

Maps

Coupling



PDEs **Dynamical Systems**

Gradient Flows

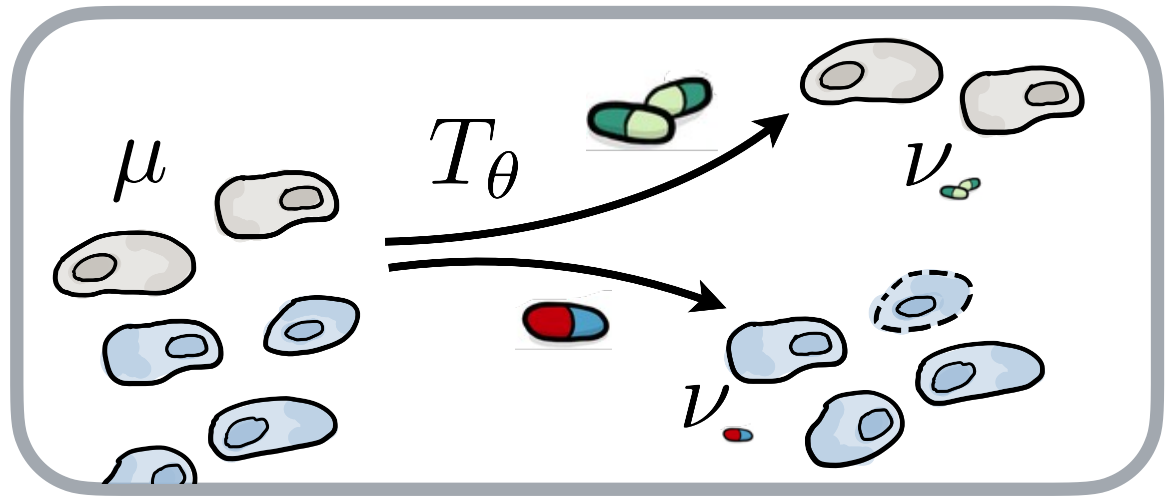
Jordan-Kinderlehrer-Otto Flows

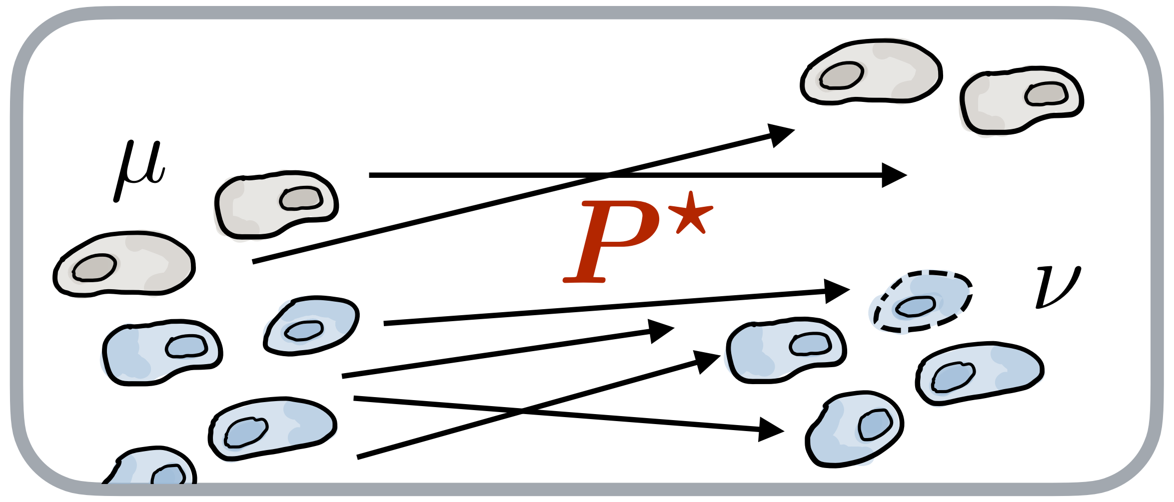
Control Theory

Schrödinger Bridges

SDEs Diffusion Models

Connections of Optimal Transport to Control and Dynamics





Optimal Transport

Maps

Coupling

PDEs **Dynamical Systems**

Gradient Flows

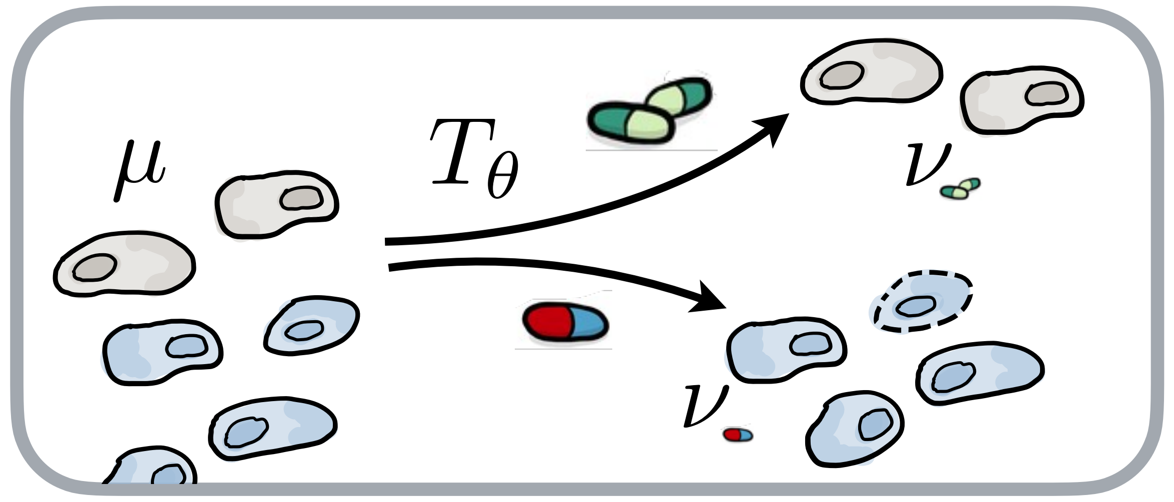
Jordan-Kinderlehrer-Otto Flows

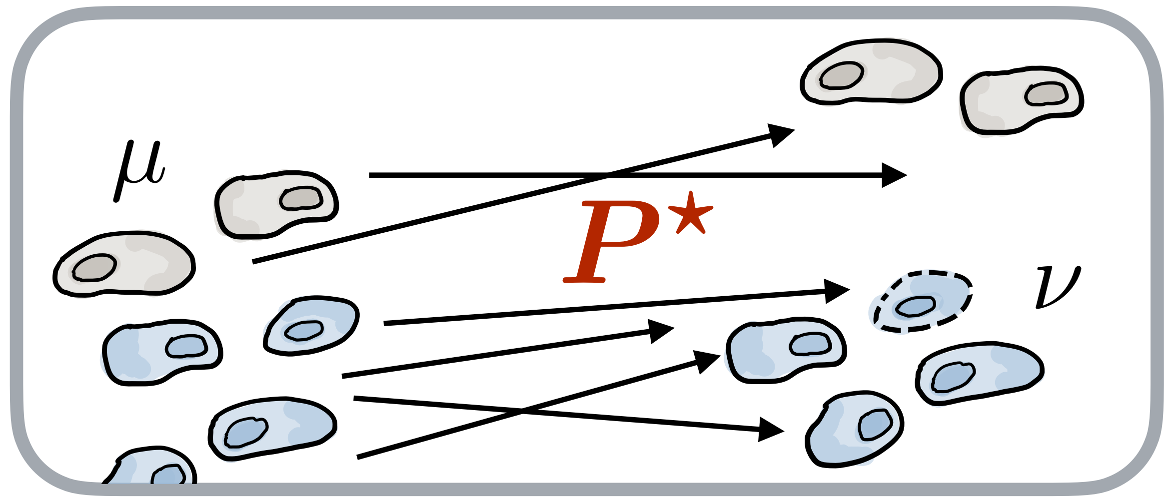
Control Theory

Schrödinger Bridges

SDEs Diffusion Models

Connections of Optimal Transport to Control and Dynamics





Optimal Transport

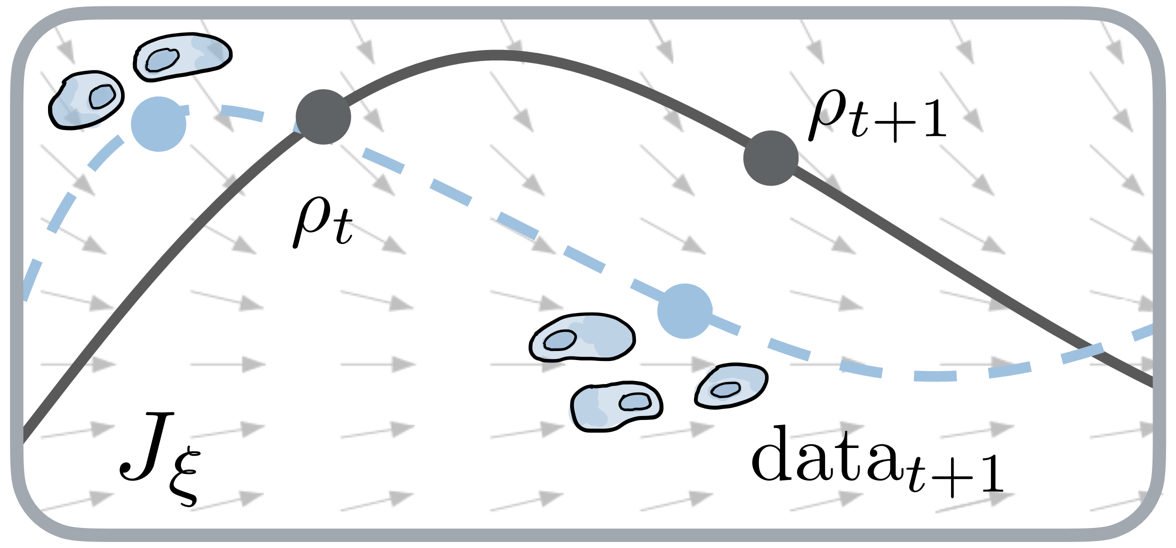
Maps

Coupling

PDEs **Dynamical Systems**

Gradient Flows

Jordan-Kinderlehrer-Otto Flows

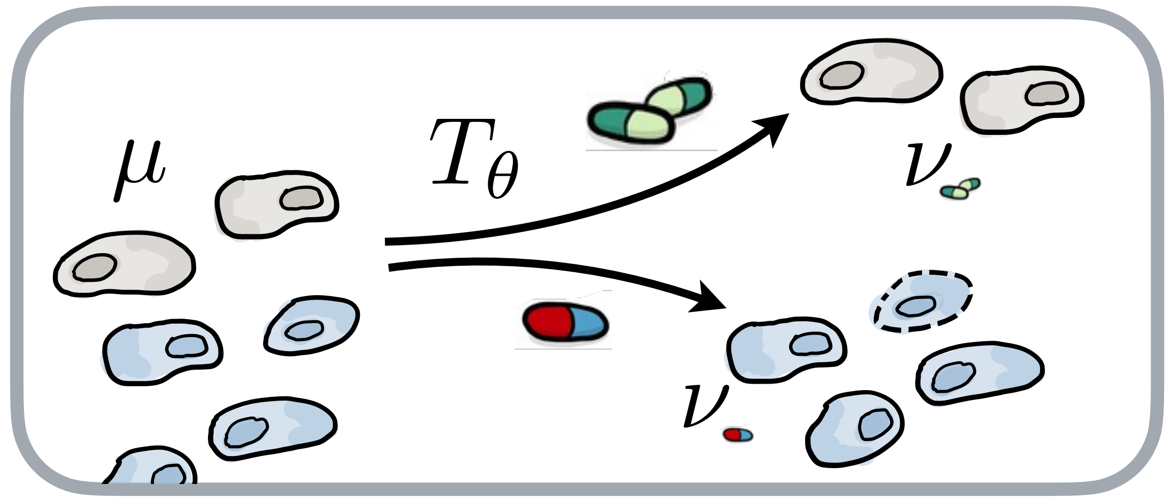


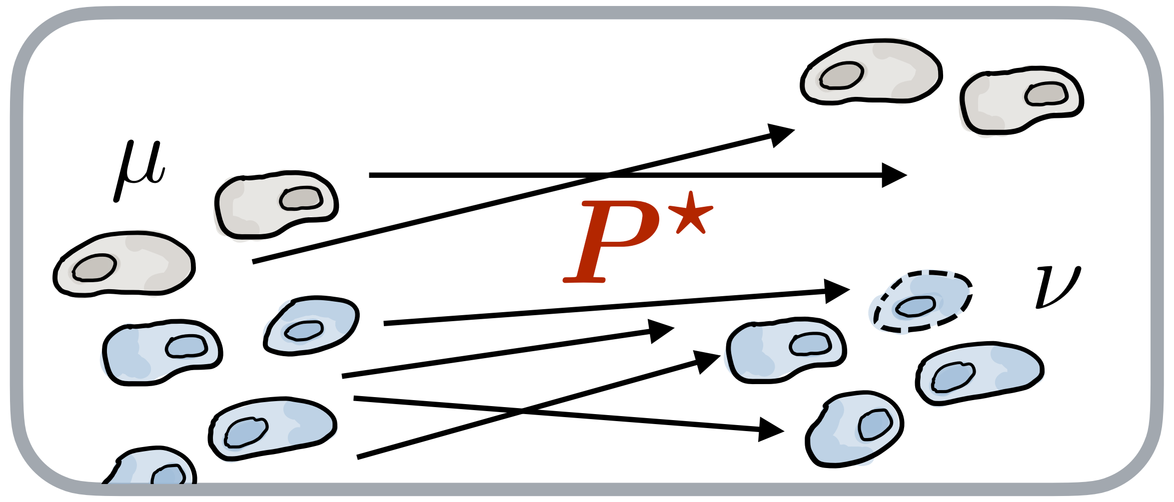
Control Theory

Schrödinger Bridges

SDEs Diffusion Models

Connections of Optimal Transport to Control and Dynamics





Optimal Transport

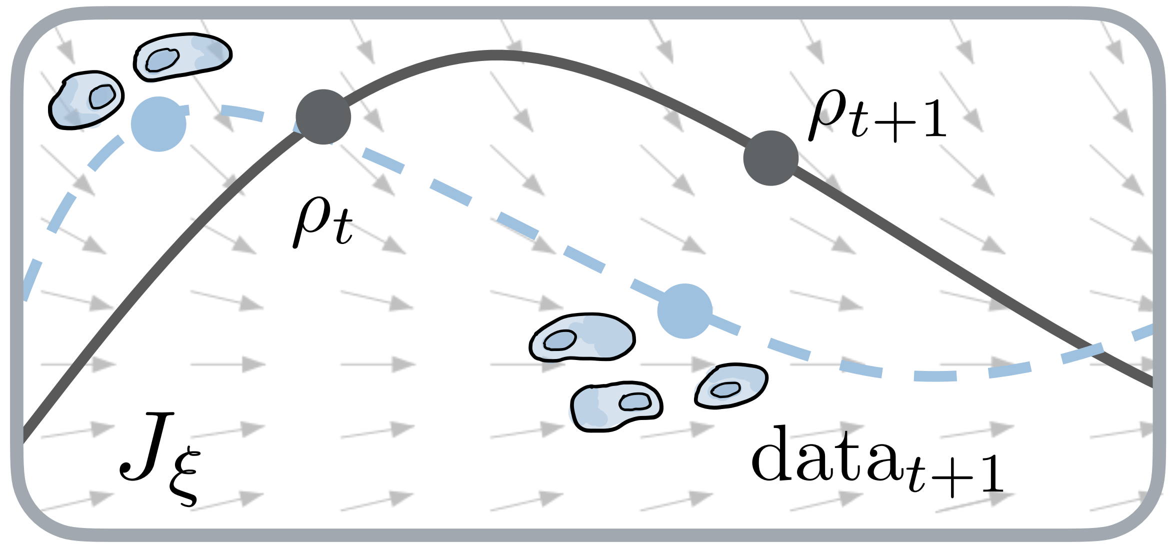
Maps

Coupling

PDEs **Dynamical Systems**

Gradient Flows

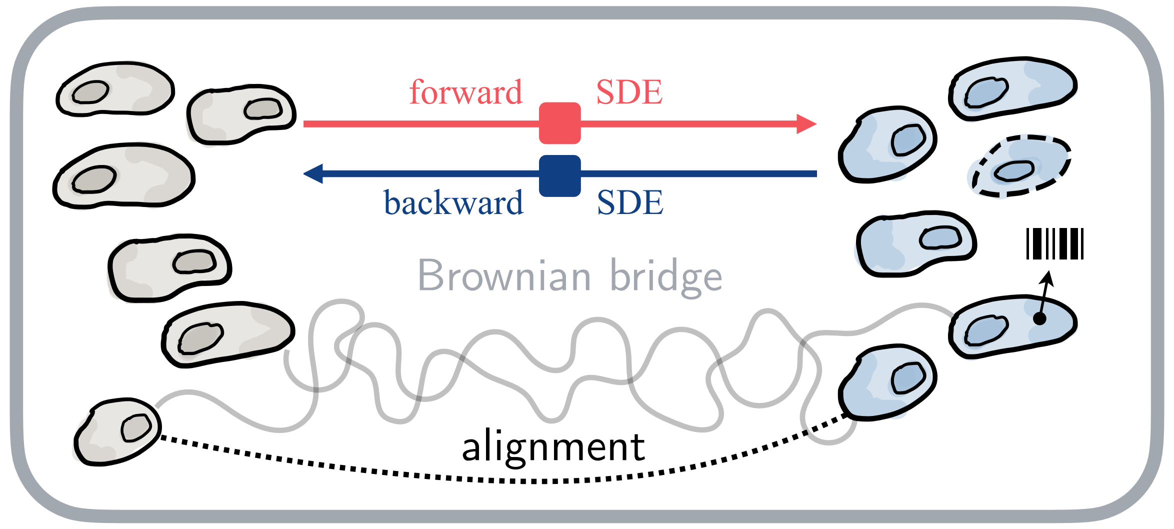
Jordan-Kinderlehrer-Otto Flows



Control Theory

Schrödinger Bridges

SDEs Diffusion Models



Upcoming Initiatives

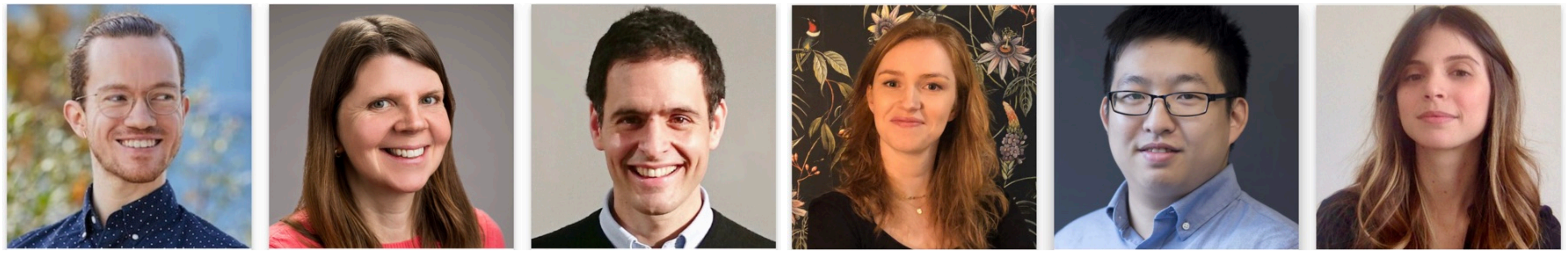
... at ICML 2023

New Frontiers in
Learning, Control, and Dynamical Systems

Workshop at ICML 2023

#Frontiers4LCD

Speakers:



Brandon Amos | Claire Tomlin | Marco Cuturi | Rianne van den Berg | Jiequn Han | Giorgia Ramponi

Upcoming Initiatives

… at ICML 2023

New Frontiers in Learning, Control, and Dynamical Systems

Workshop at ICML 2023

#Frontiers4LCD

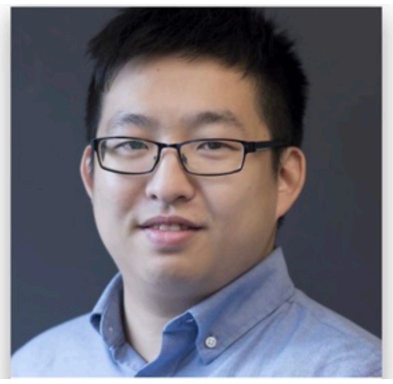
Speakers:

Brandon Amos

Claire Tomlin

Marco Cuturi

Rianne van den Berg



Jiequn Han

Giorgia Ramponi

… at NeurIPS 2023

Optimal Transport and Machine Learning

Workshop at NeurIPS 2023

#OTML

Speakers:

Felix Otto

Laetitia Chapel

Brandon Amos

Arnaud Doucet

Florentina Bunea

Smita Krishnaswamy

Sinho Chewi

Resources

Books

G. Peyré and M. Cuturi. **Computational Optimal Transport: With Applications to Data Science.** Foundations and Trends® in Machine Learning 11.5-6 (2019)

C. Villani. **Topics in Optimal Transportation.** GSM Vol. 58, AMS, 2009.

F. Santambrogio. **Optimal Transport for Applied Mathematicians.** Birkhäuser, 2015.

Review Articles

Chen, Yongxin, Tryphon T. Georgiou, and Michele Pavon. **Optimal Transport in Systems and Control.** *Annual Review of Control, Robotics, and Autonomous Systems* Vol. 4 (2021).

Python Libraries

https://pythonot.github.io/

https://ott-jax.readthedocs.io/