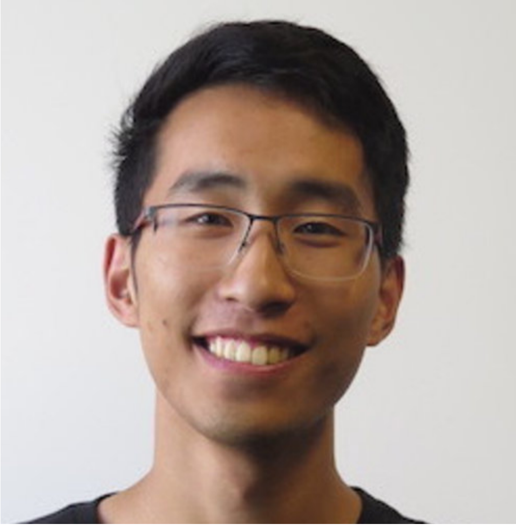
Solving the trajectory inference problem with the help of optimal transport

Hugo Lavenant

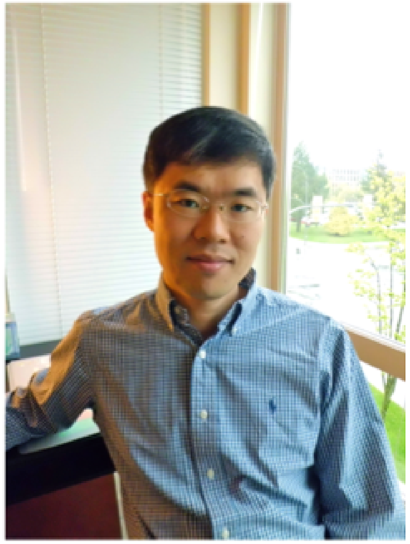
Bocconi University

Workshop ”Optimal Transport, Mean-Field Models and Machine Learning”
Munich, April 26, 2023

Joint work with:



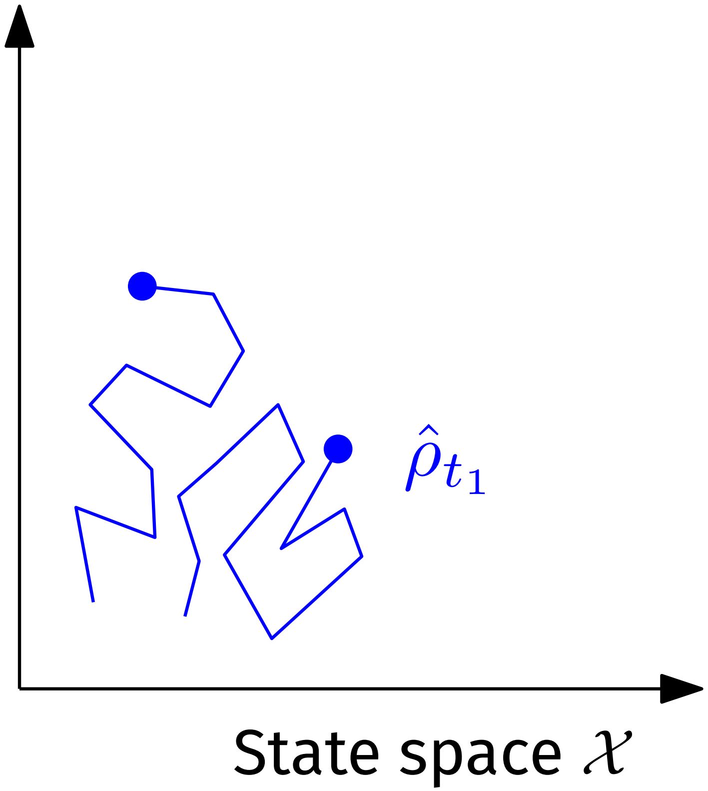
Stephen Zhang



Young-Heon Kim

Geoff Schiebinger

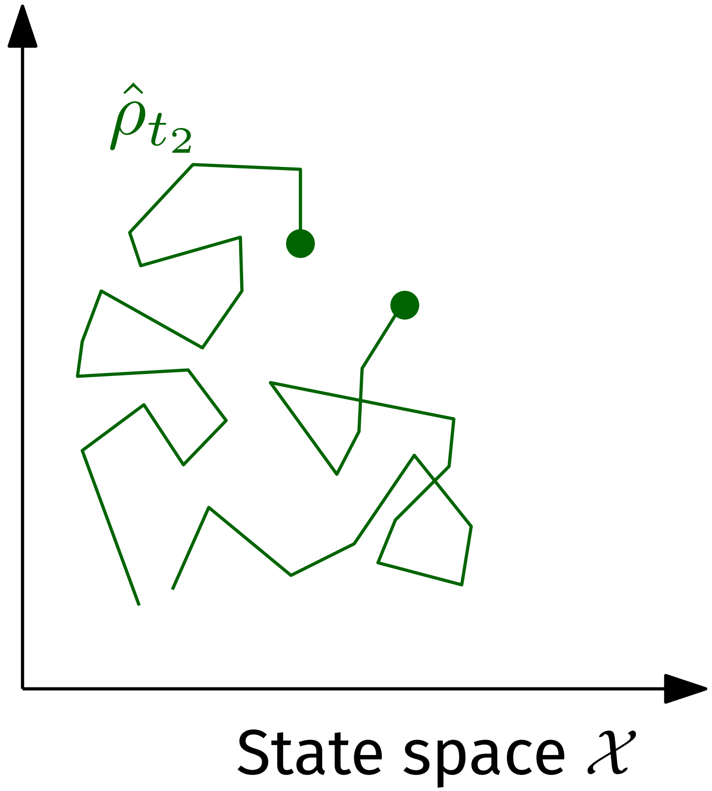
A preview of the mathematical problem



Stochastic process X_t

Samples from law of X_{t_1}

A preview of the mathematical problem

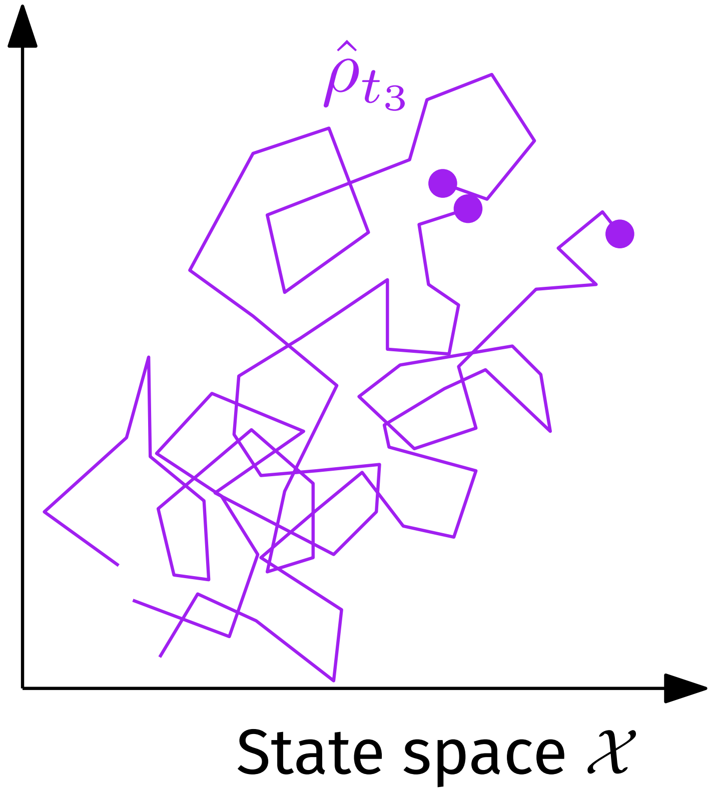


Stochastic process X_t

Samples from law of X_{t_2}

(independent from the previous samples)

A preview of the mathematical problem

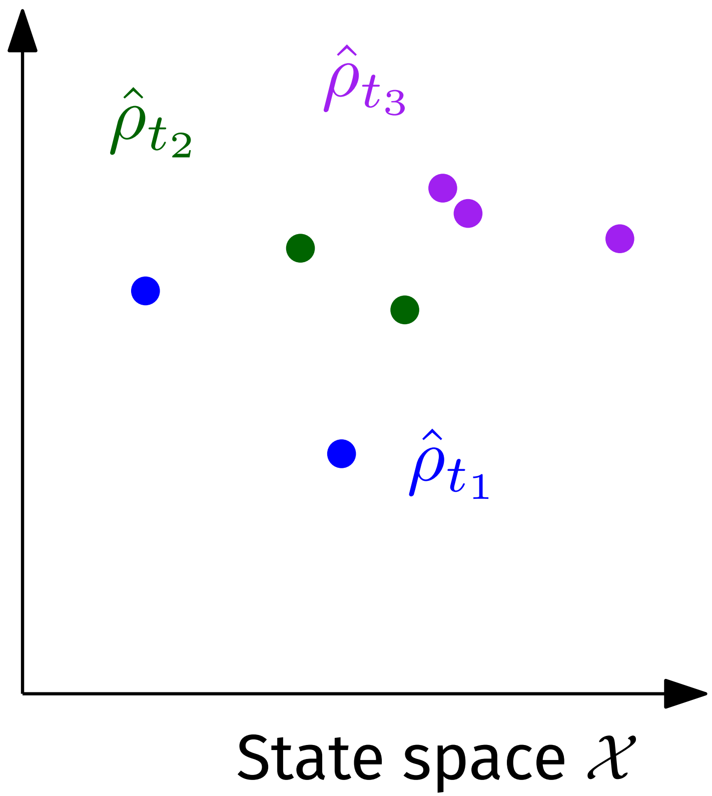


Stochastic process X_t

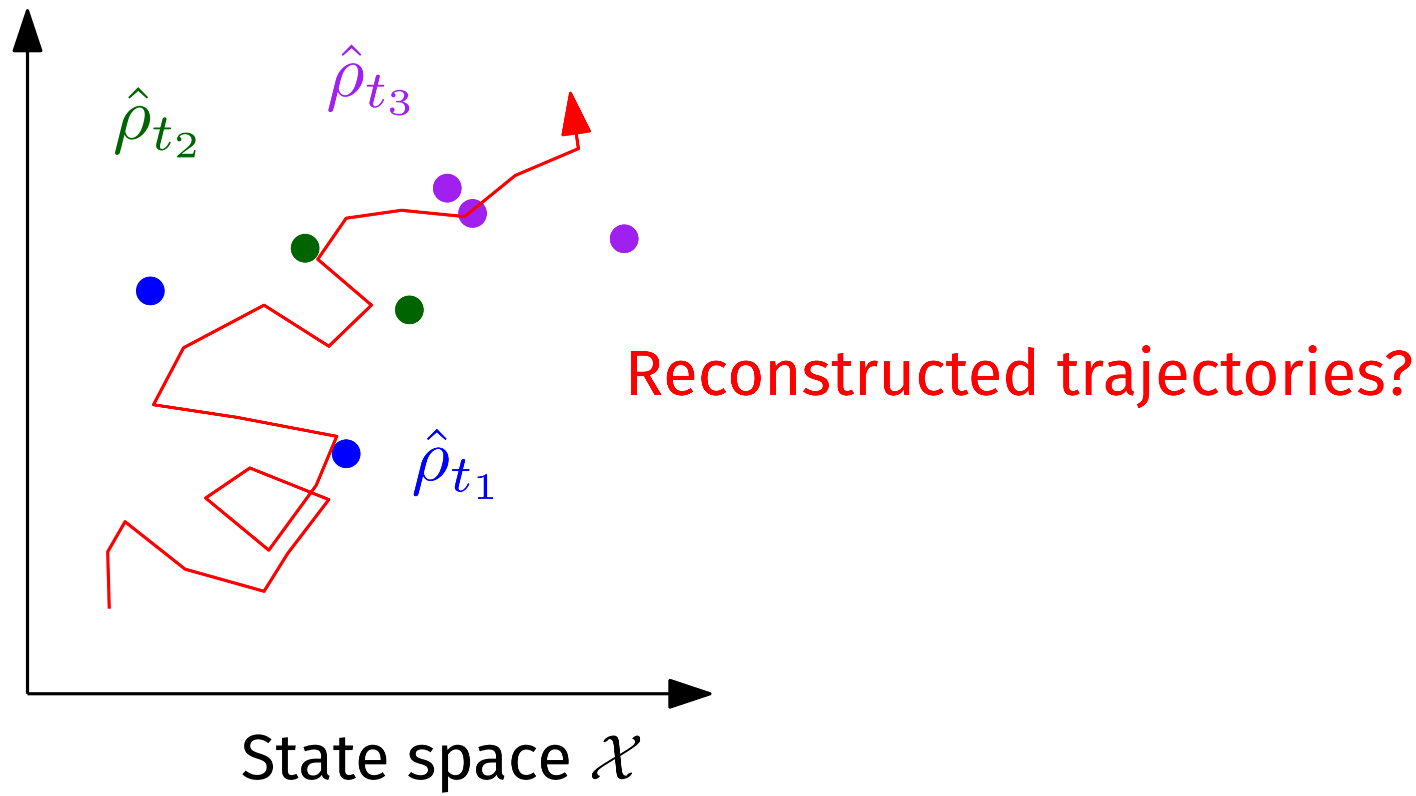
Samples from law of X_{t_3}

(independent from the previous samples)

A preview of the mathematical problem

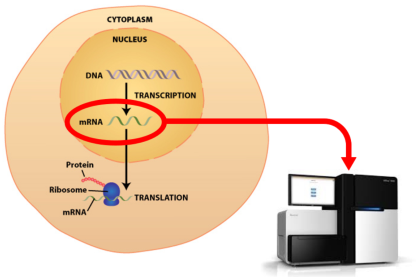


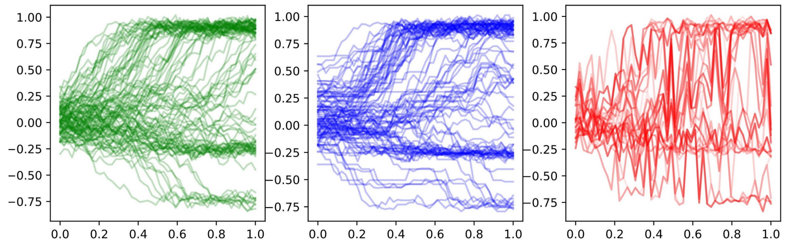
A preview of the mathematical problem



Goal: reconstruct the law of the trajectories X_t from samples of the temporal marginals.

1 - Biological Context



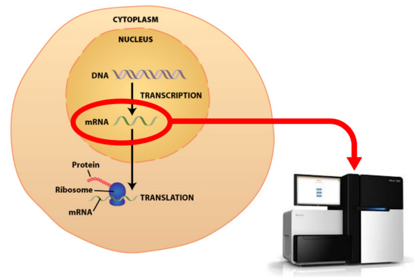


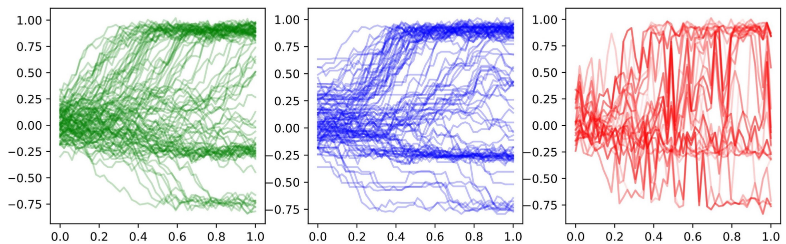
2 - Algorithms and results

3 - Theoretical analysis

$$\mathrm{d}X_t = \mathbf{v}(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t$$

1 - Biological Context





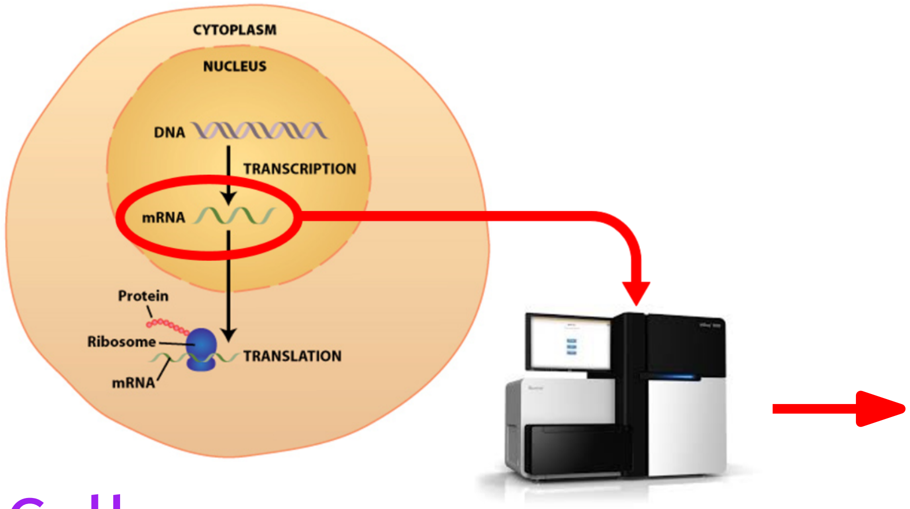
2 - Algorithms and results

3 - Theoretical analysis

$$\mathrm{d}X_t = \mathbf{v}(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t$$

Single-cell RNA sequencing

Gene expression profile

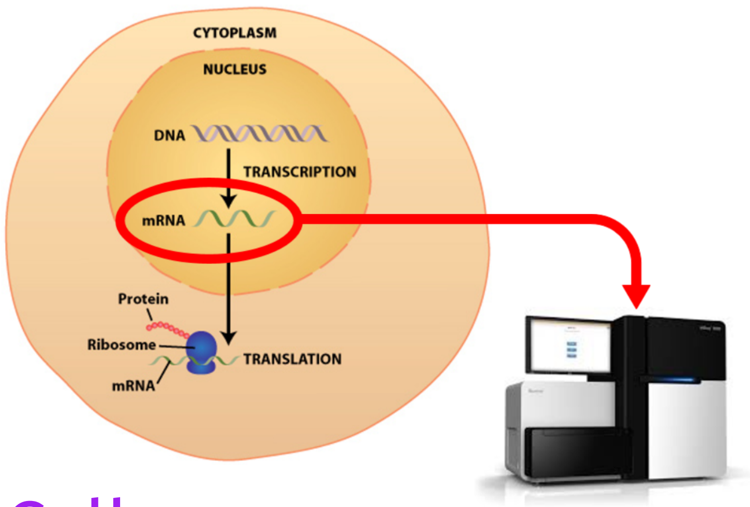


Cell

$$\begin{pmatrix} \text{Number RNA gene 1} \\ \text{Number RNA gene 2} \\ \vdots \\ \text{Number RNA gene N} \end{pmatrix}$$

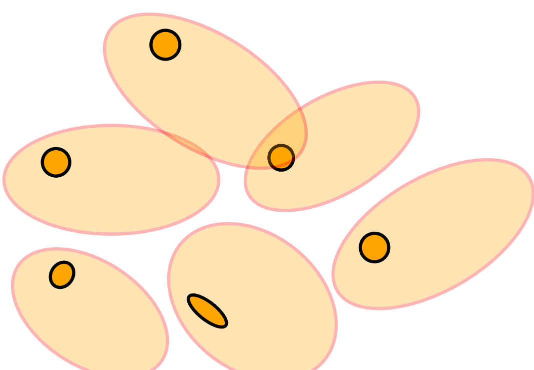
Single-cell RNA sequencing

Gene expression profile

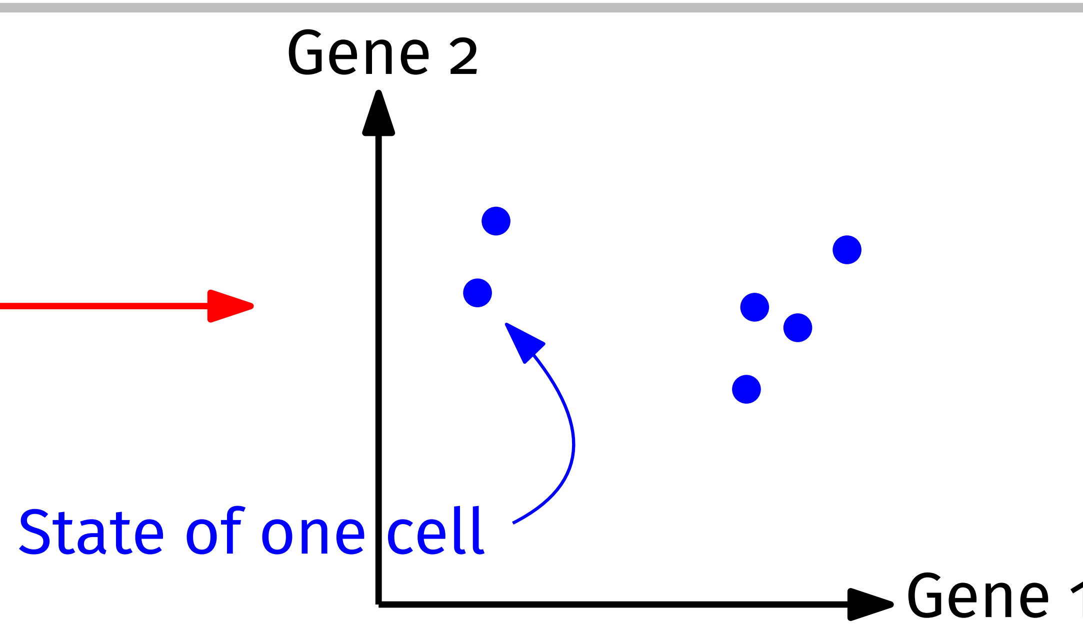


Cell

$$\begin{pmatrix} \text{Number RNA gene 1} \\ \text{Number RNA gene 2} \\ \vdots \\ \text{Number RNA gene N} \end{pmatrix}$$

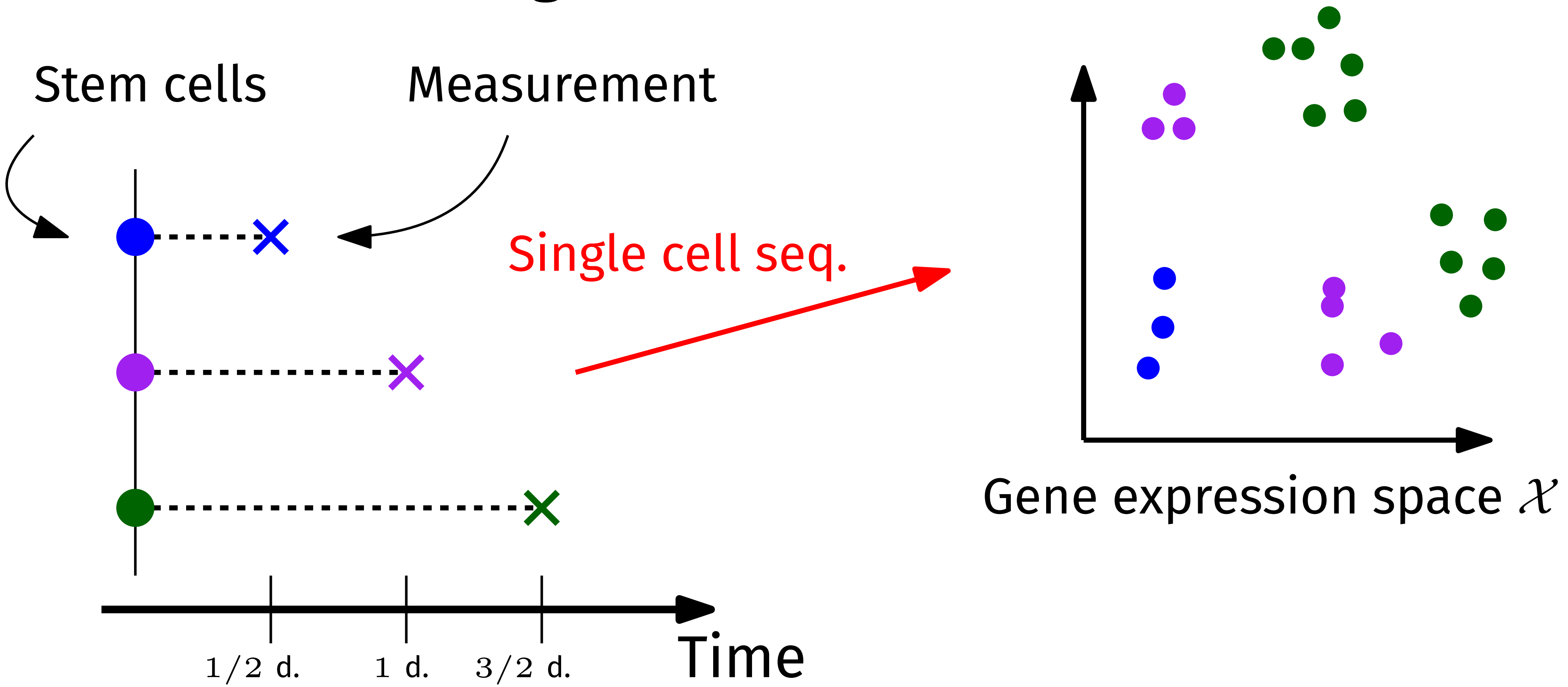


Population of cells



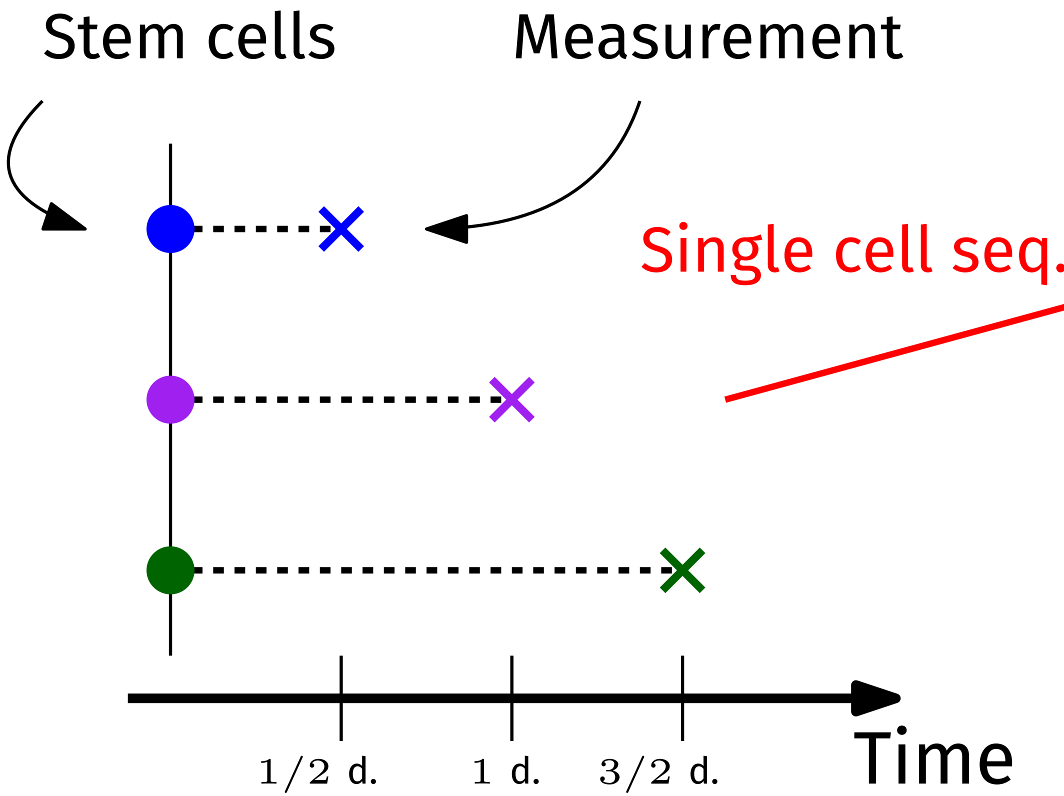
Gene expression space $\mathcal{X}$

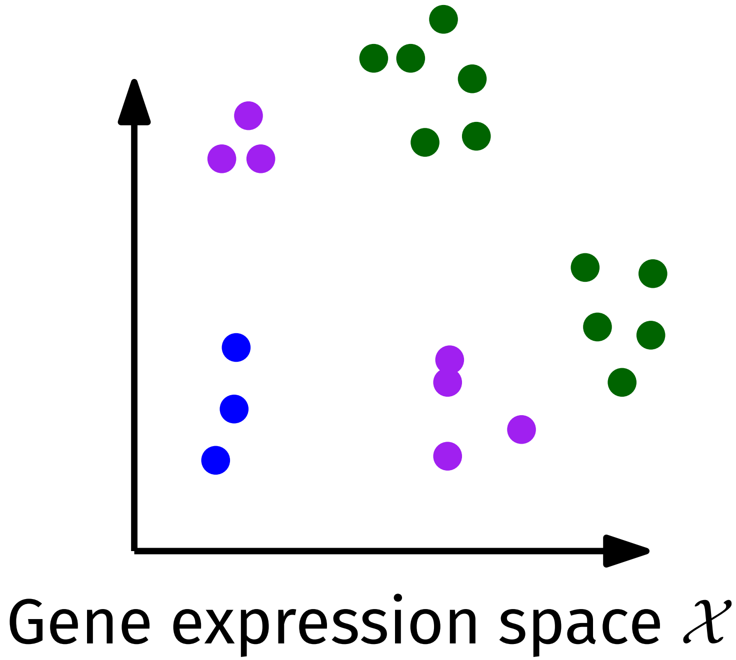
Investing cell differentiation



Schiebinger et al. (2019). Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming.

Investing cell differentiation



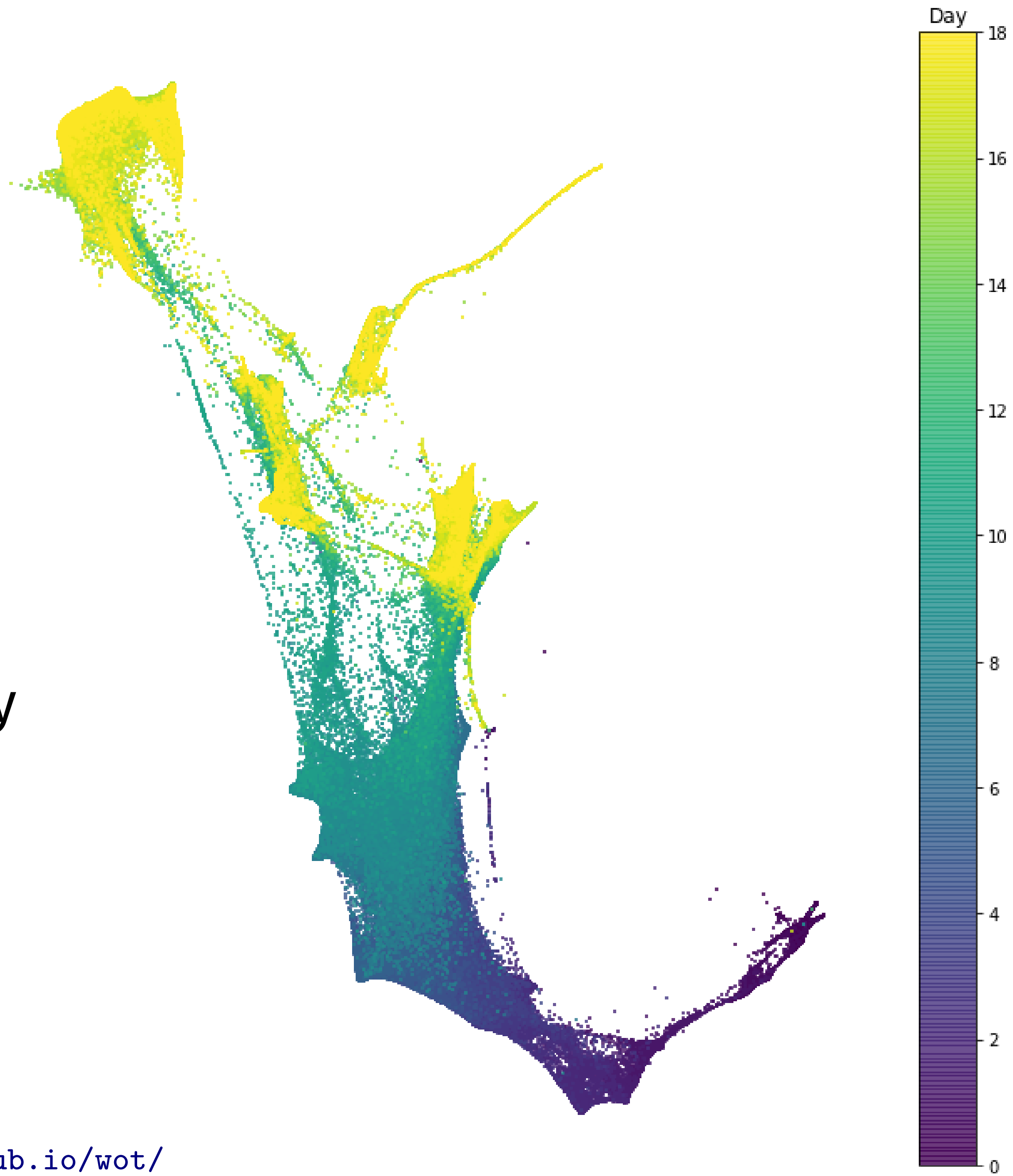


Schiebinger et al.
- 39 time points.
- Total $250,000$ cells measured.

(Biological) goal: reconstruct fate of cells, unravel the regulatory network.

Schiebinger et al. (2019). Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression sheds light on cellular reprogramming.

Dataset

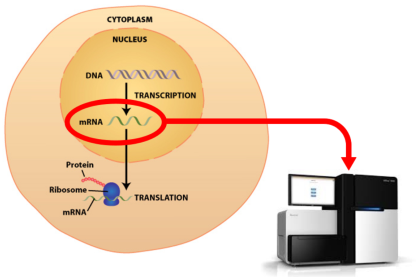


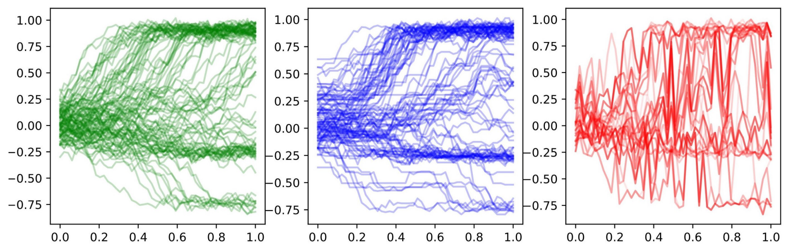
Displayed with **Force Layout Embedding** (FLE), a dimensionality reduction technique

https://broadinstitute.github.io/wot/

Disclaimer: in this presentation, we ignore cell division.

1 - Biological Context



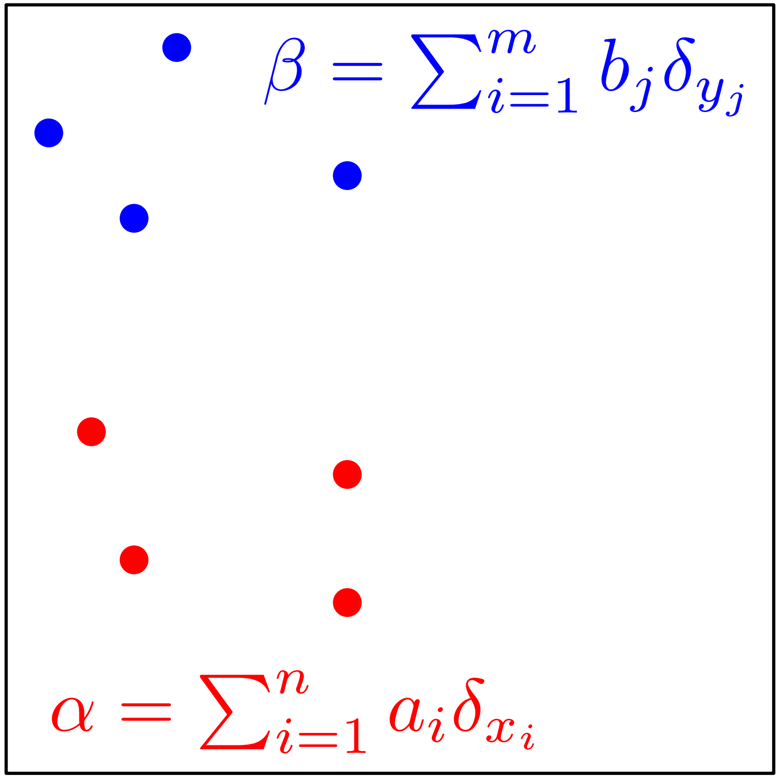


2 - Algorithms and results

3 - Theoretical analysis

$$\mathrm{d}X_t = \mathbf{v}(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t$$

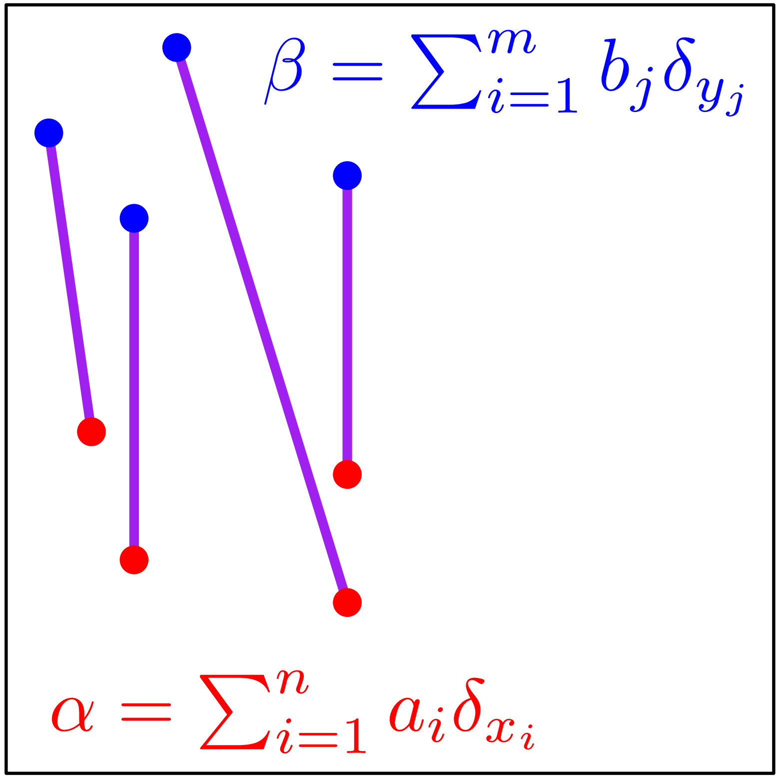
(Entropic) optimal transport



Probability distributions:

$\sum_i a_i = \sum_j b_j = 1$

(Entropic) optimal transport



Probability distributions:

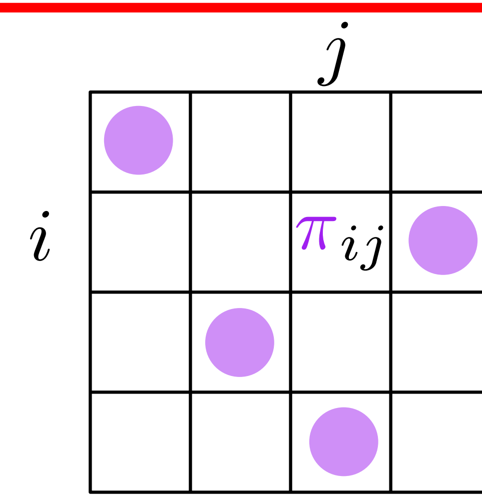
$\sum_i a_i = \sum_j b_j = 1$

Find $\pi \geq 0$ such that

$$\begin{cases} \sum_j \pi_{ij} = a_i \\ \sum_i \pi_{ij} = b_j \end{cases}$$

and which minimizes

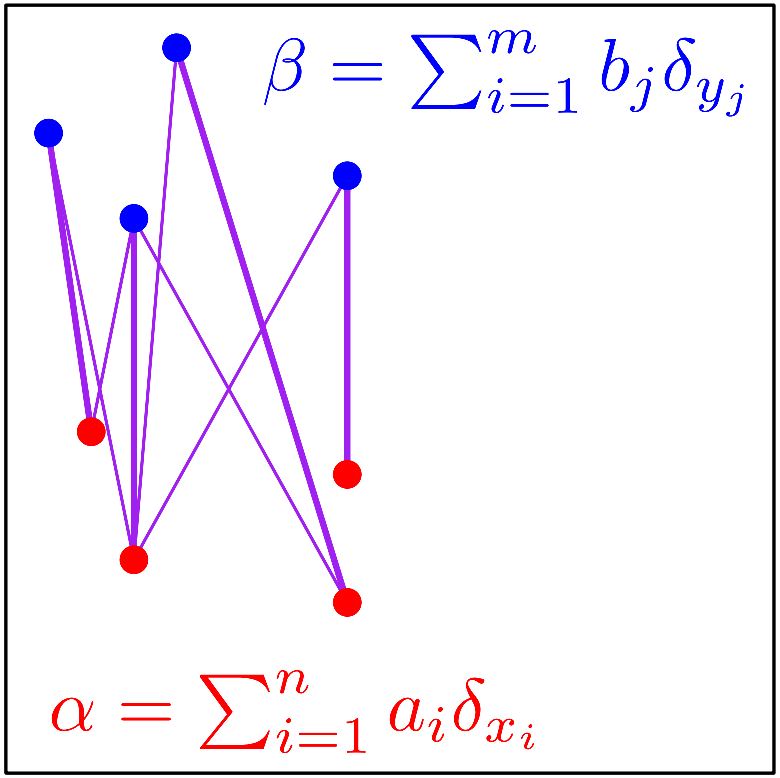
$$\sum_{ij} \pi_{ij} |x_i - y_j|^2$$



Matrix π

Peyré and Cuturi. (2019). Computational Optimal Transport.

(Entropic) optimal transport



Probability distributions:

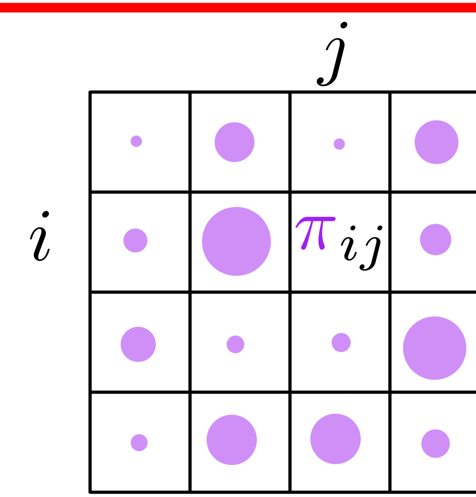
$\sum_i a_i = \sum_j b_j = 1$

Find $\pi \geq 0$ such that

$$\begin{cases} \sum_j \pi_{ij} = a_i \\ \sum_i \pi_{ij} = b_j \end{cases}$$

and which minimizes

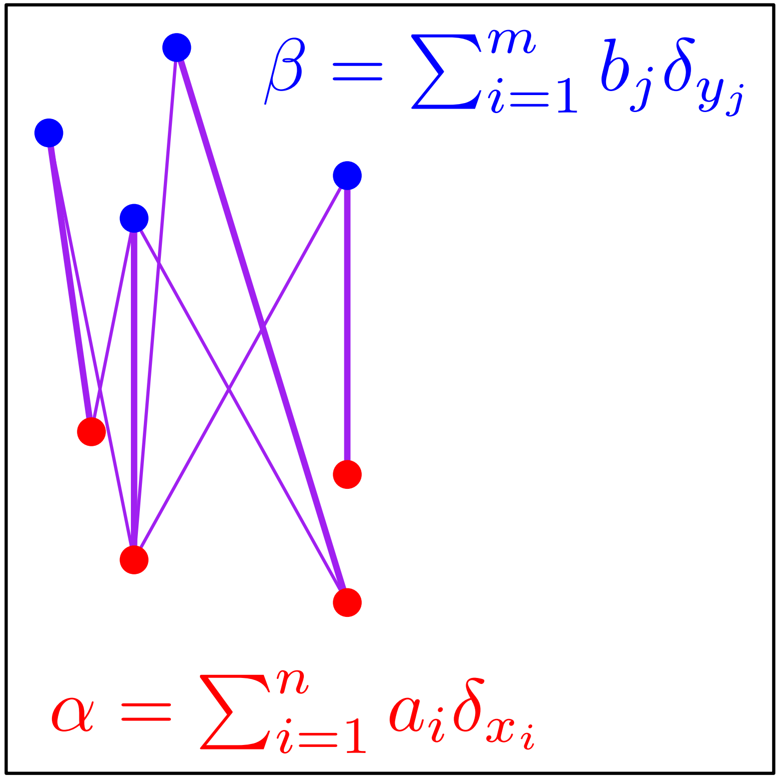
$$\sum_{ij} \pi_{ij} |x_i - y_j|^2 + \varepsilon \sum_{ij} \pi_{ij} \log \pi_{ij}$$



Matrix π

Peyré and Cuturi. (2019). Computational Optimal Transport.

(Entropic) optimal transport



Find $\pi \geq 0$ such that

$$\begin{cases} \sum_j \pi_{ij} = a_i \\ \sum_i \pi_{ij} = b_j \end{cases}$$

and which minimizes

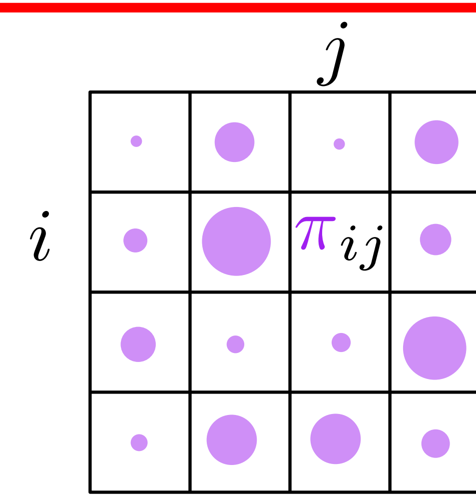
$$\sum_{ij} \pi_{ij} |x_i - y_j|^2 + \varepsilon \sum_{ij} \pi_{ij} \log \pi_{ij}$$

Probability distributions:

$\sum_i a_i = \sum_j b_j = 1$

π law of (X, Y) with $X \sim \alpha$ and $Y \sim \beta$:

$$\mathbb{P}(X = x_i, Y = y_j) = \pi_{ij}$$

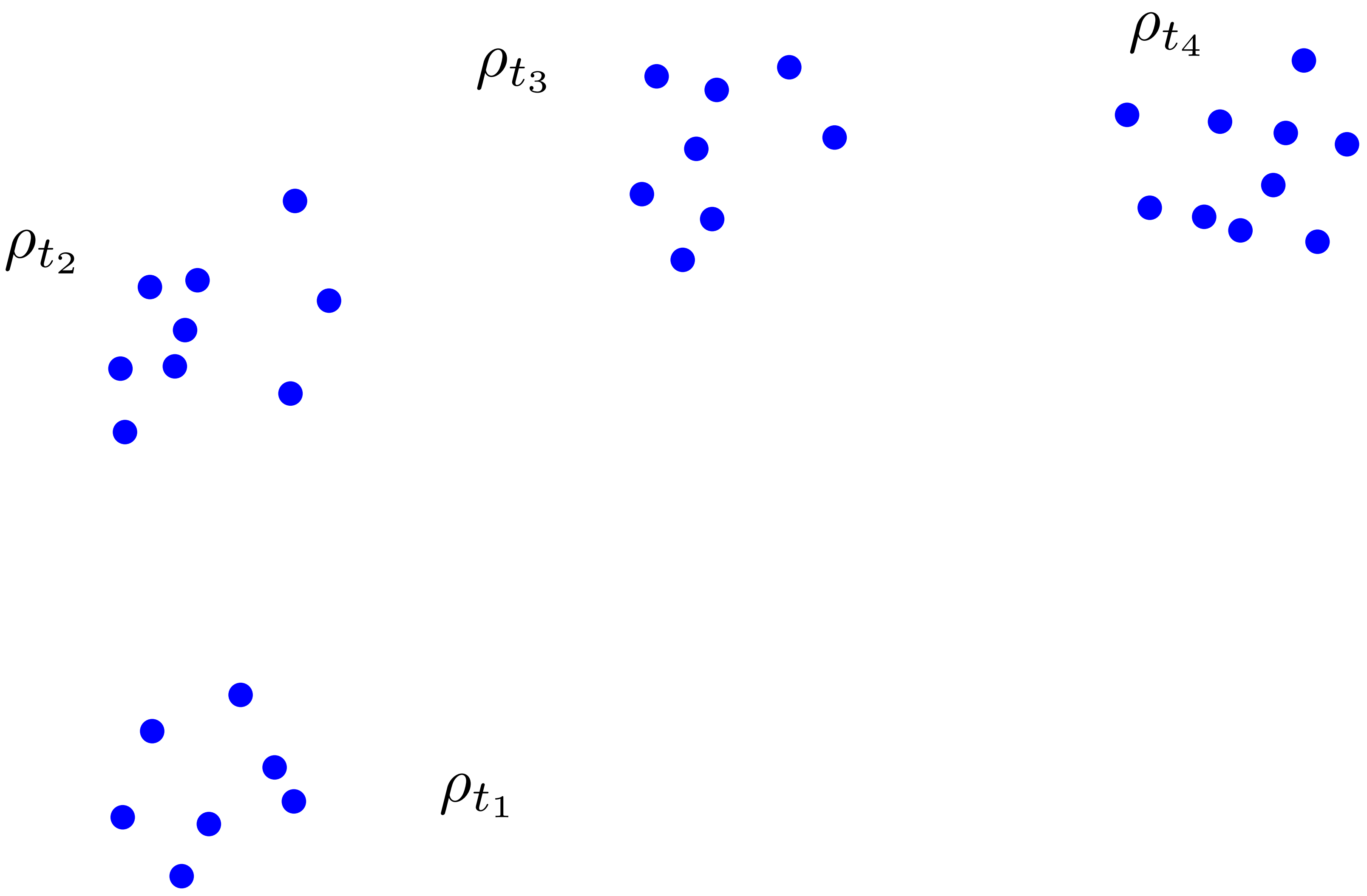


Matrix π

Peyré and Cuturi. (2019). Computational Optimal Transport.

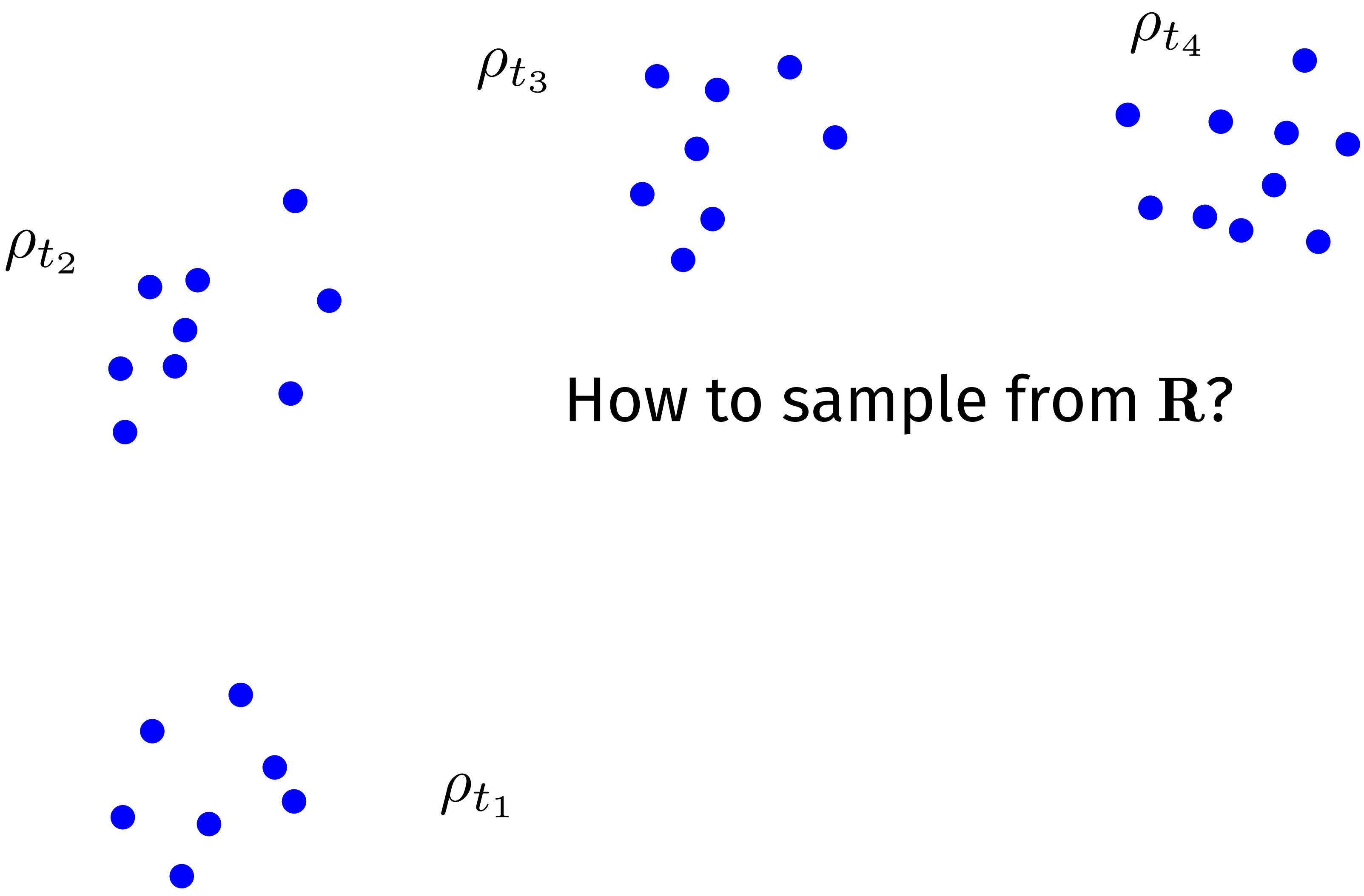
A description of (a simplified) Waddington OT

Input: $\rho_{t_1}, \rho_{t_2}, \ldots \rho_{t_T}$ probability measures
Ouptut: R law of reconstructed trajectories.



A description of (a simplified) Waddington OT

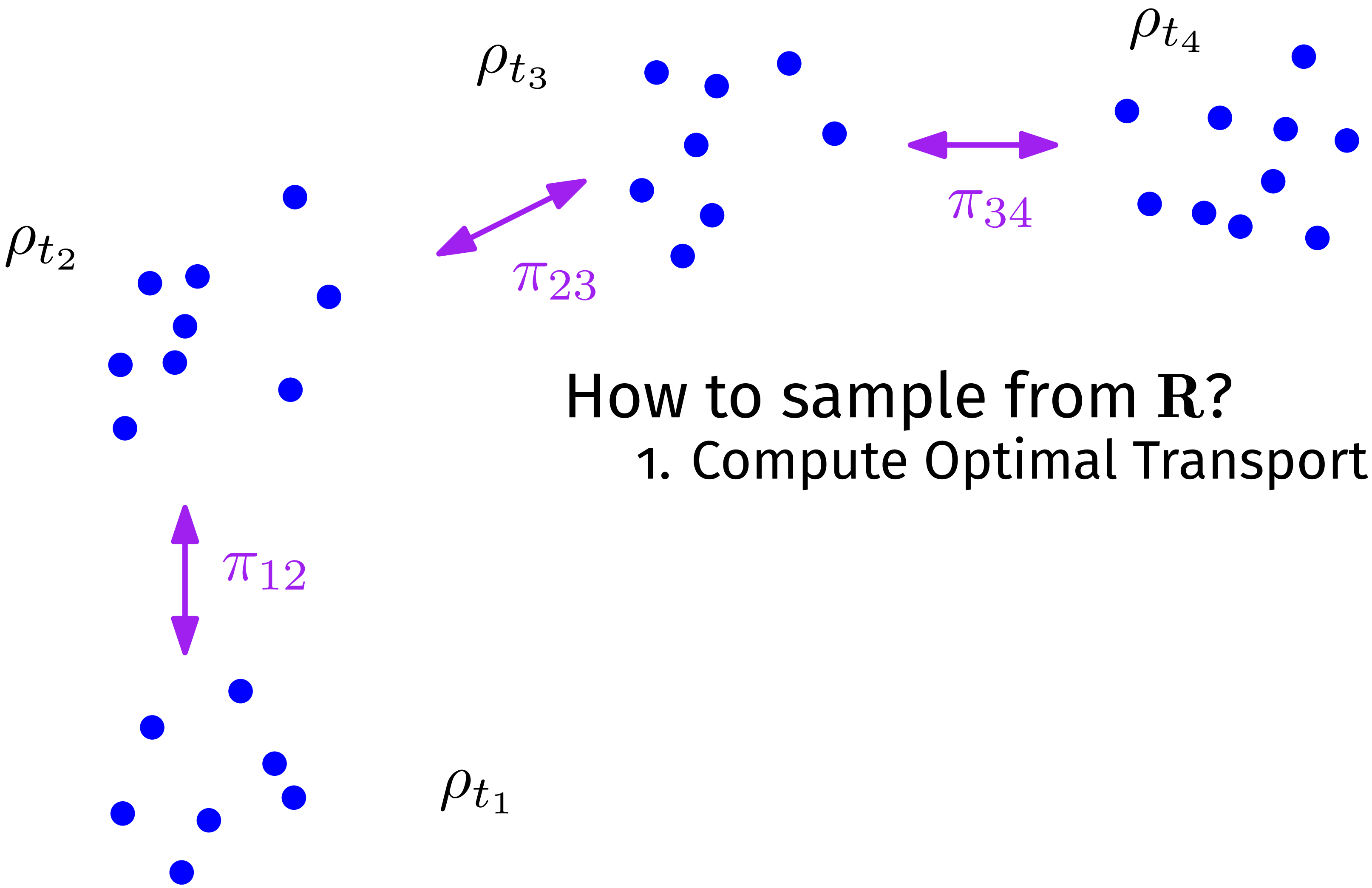
Input: $\rho_{t_1}, \rho_{t_2}, \dots \rho_{t_T}$ probability measures
Ouptut: $\mathbf{R}$ law of reconstructed trajectories.



How to sample from $\mathbf{R}$?

A description of (a simplified) Waddington OT

Input: $\rho_{t_1}, \rho_{t_2}, \ldots \rho_{t_T}$ probability measures
Ouptut: $\mathbf{R}$ law of reconstructed trajectories.

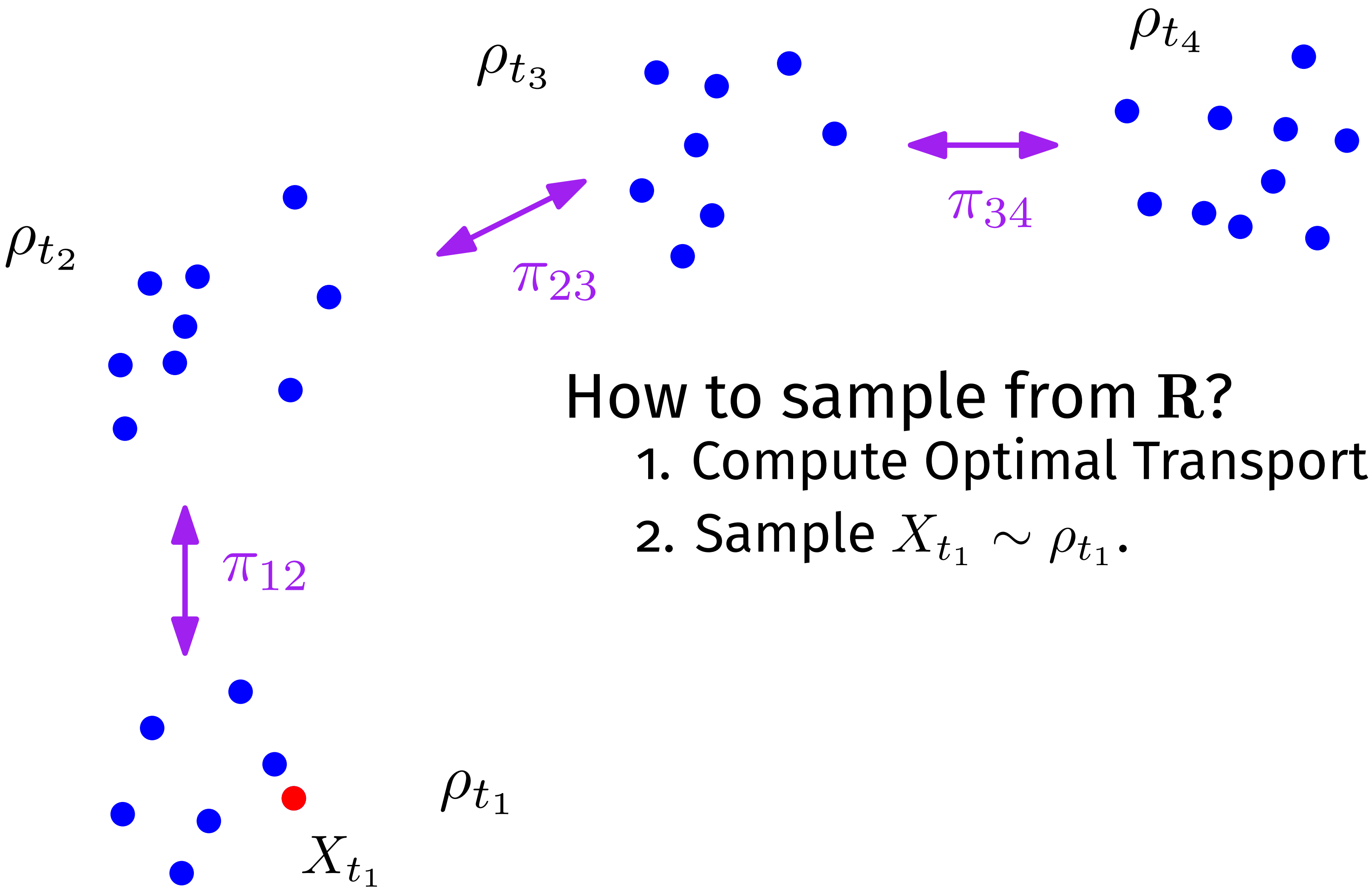


How to sample from $\mathbf{R}$?

1. Compute Optimal Transport couplings.

A description of (a simplified) Waddington OT

Input: $\rho_{t_1}, \rho_{t_2}, \ldots \rho_{t_T}$ probability measures
Ouptut: $\mathbf{R}$ law of reconstructed trajectories.

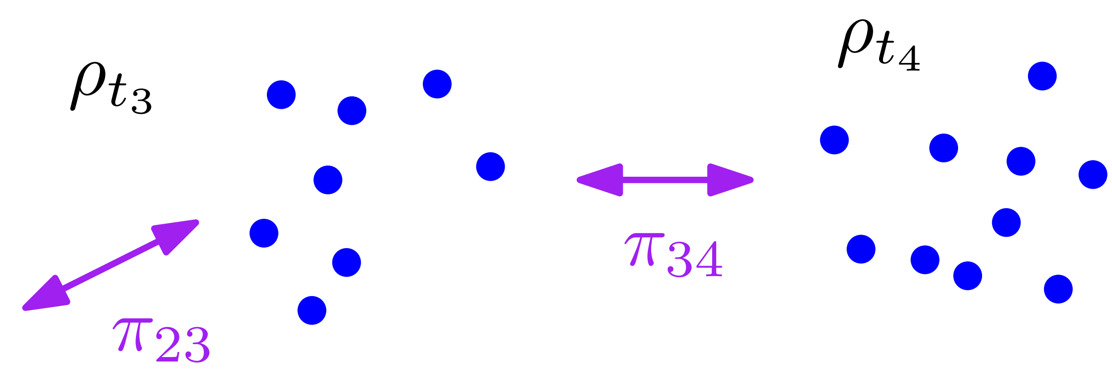


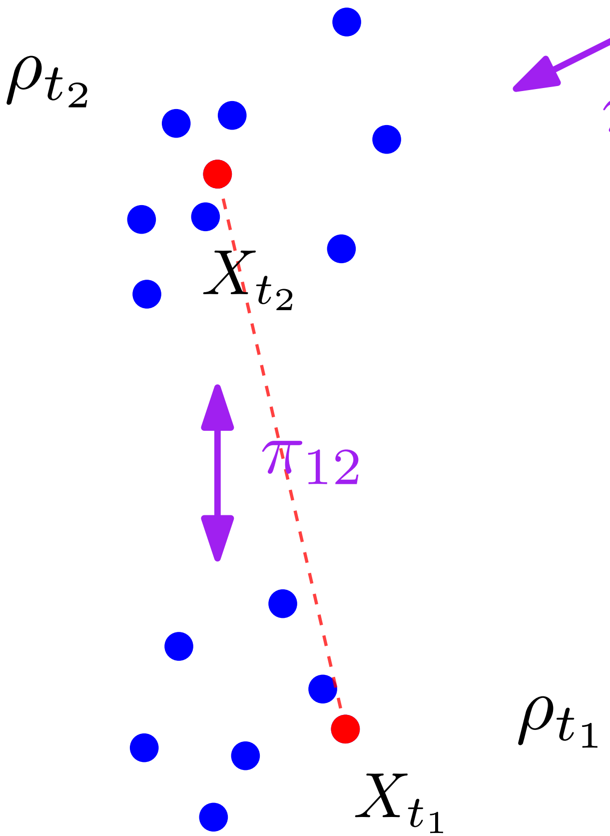
How to sample from $\mathbf{R}$?

1. Compute Optimal Transport couplings.
2. Sample $X_{t_1} \sim \rho_{t_1}$.

A description of (a simplified) Waddington OT

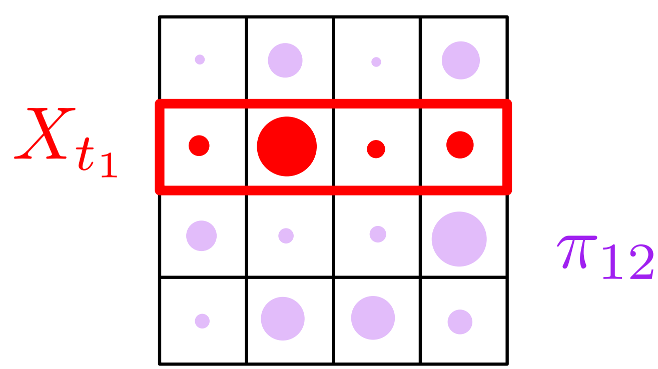
Input: $\rho_{t_1}, \rho_{t_2}, \ldots \rho_{t_T}$ probability measures
Ouptut: $\mathbf{R}$ law of reconstructed trajectories.





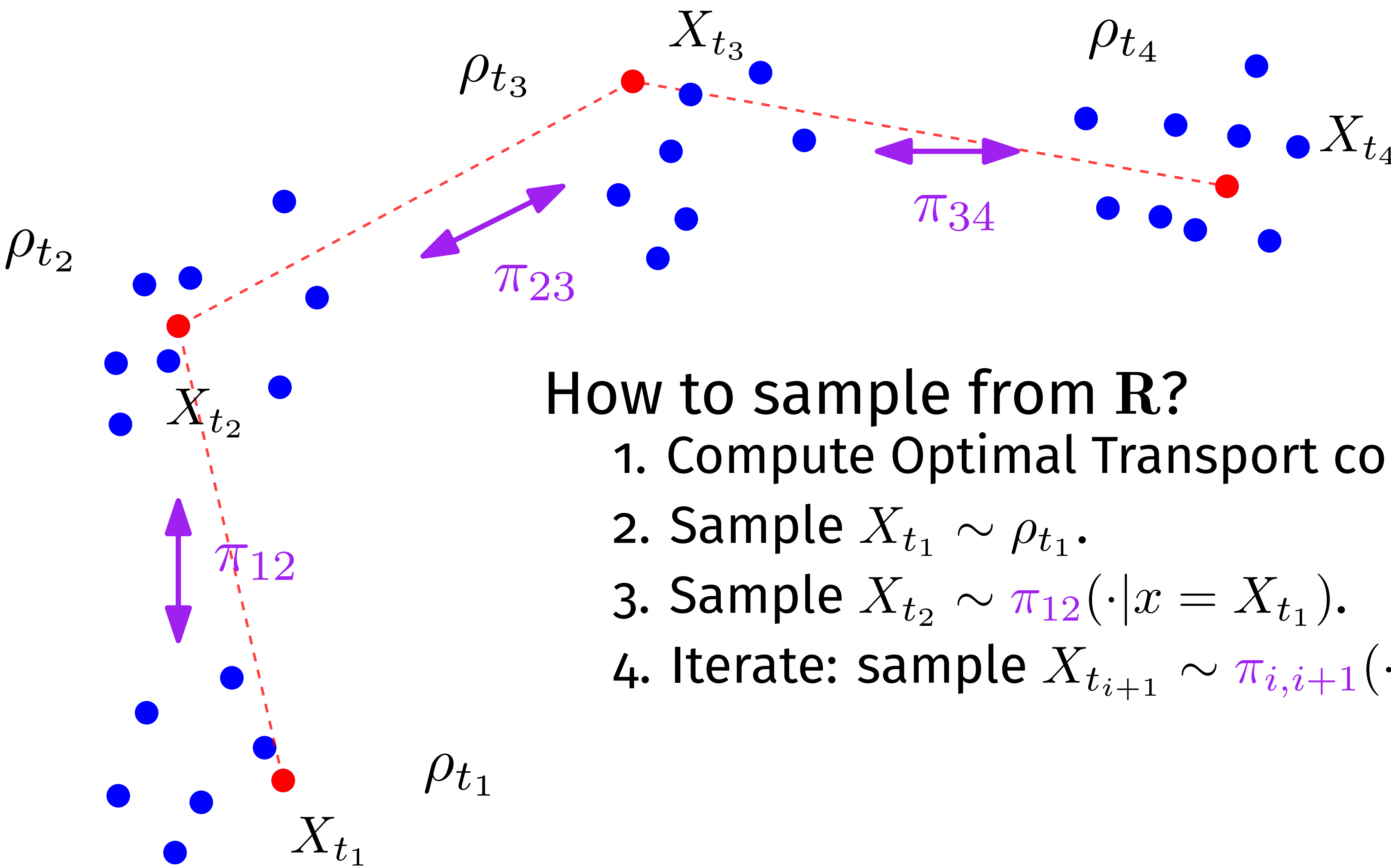
How to sample from $\mathbf{R}$?

1. Compute Optimal Transport couplings.
2. Sample $X_{t_1} \sim \rho_{t_1}$.
3. Sample $X_{t_2} \sim \pi_{12}(\cdot | x = X_{t_1})$.



A description of (a simplified) Waddington OT

Input: $\rho_{t_1}, \rho_{t_2}, \ldots \rho_{t_T}$ probability measures
Ouptut: $\mathbf{R}$ law of reconstructed trajectories.

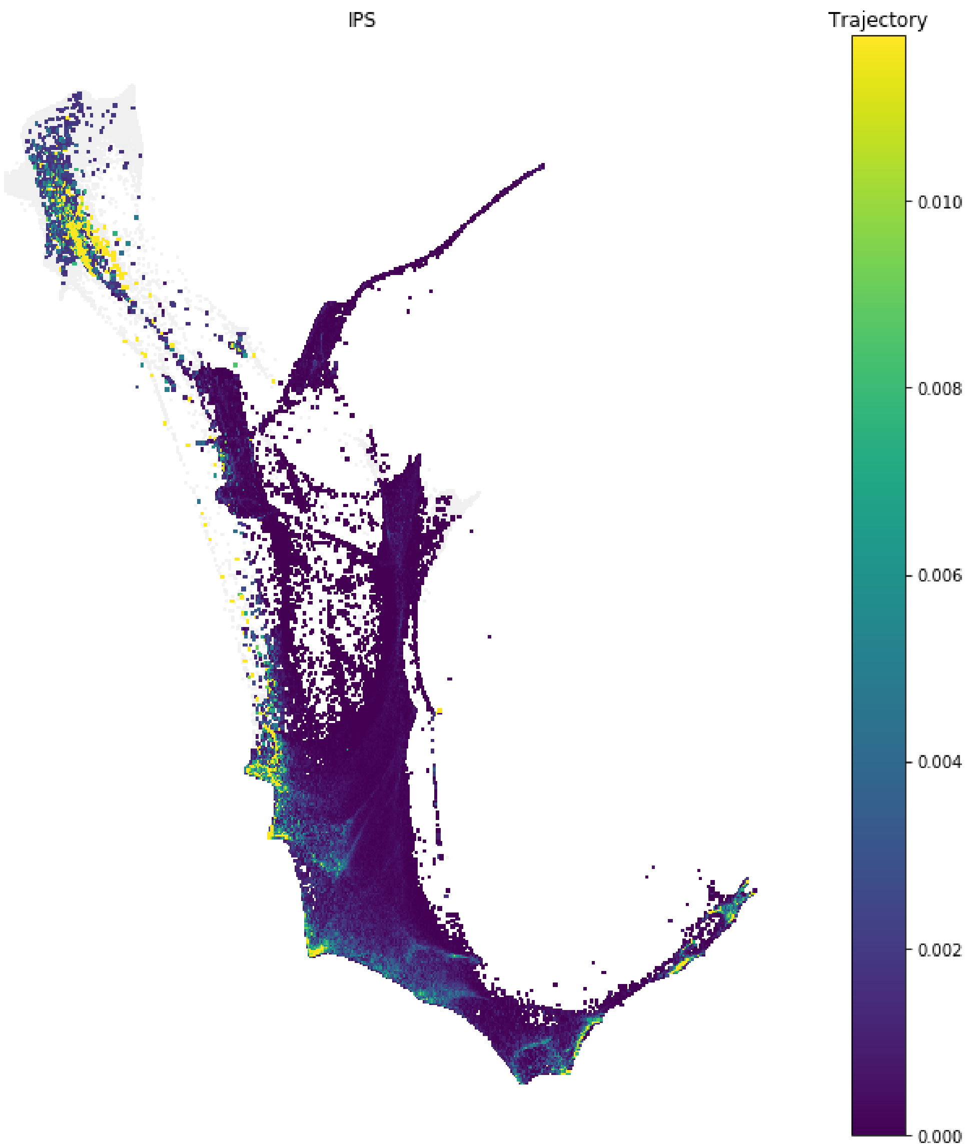


How to sample from $\mathbf{R}$?

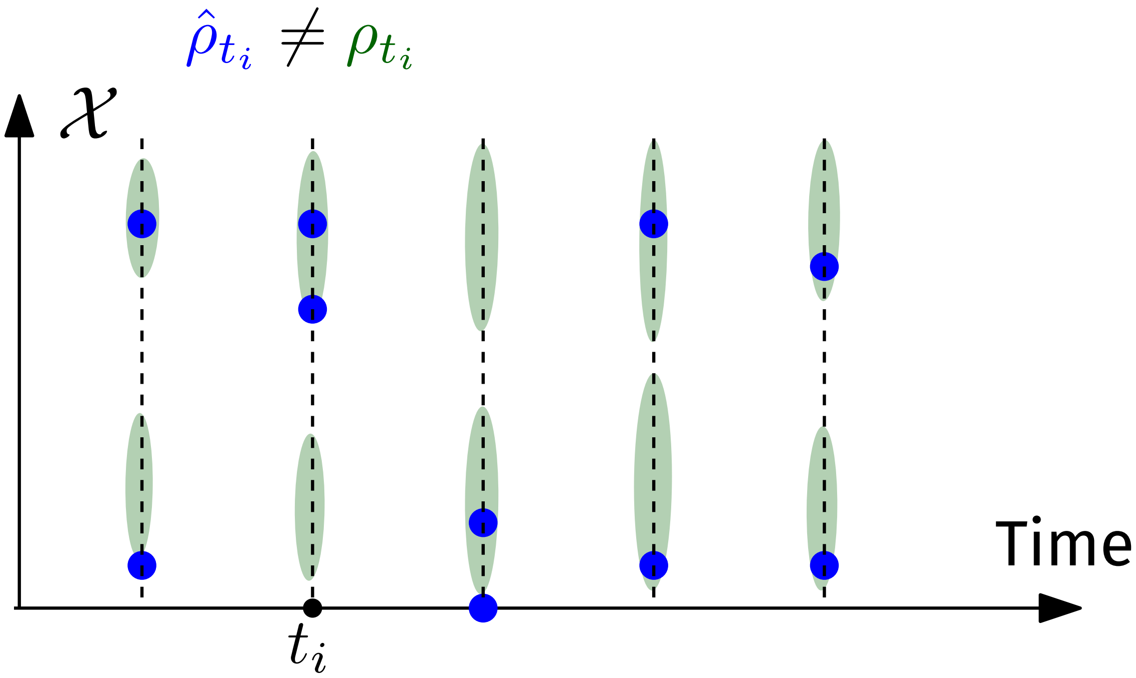
1. Compute Optimal Transport couplings.
2. Sample $X_{t_1} \sim \rho_{t_1}$.
3. Sample $X_{t_2} \sim \pi_{12}(\cdot | x = X_{t_1})$.
4. Iterate: sample $X_{t_{i+1}} \sim \pi_{i,i+1}(\cdot | x = X_{t_i})$.

Example on the dataset of Schiebinger et al.

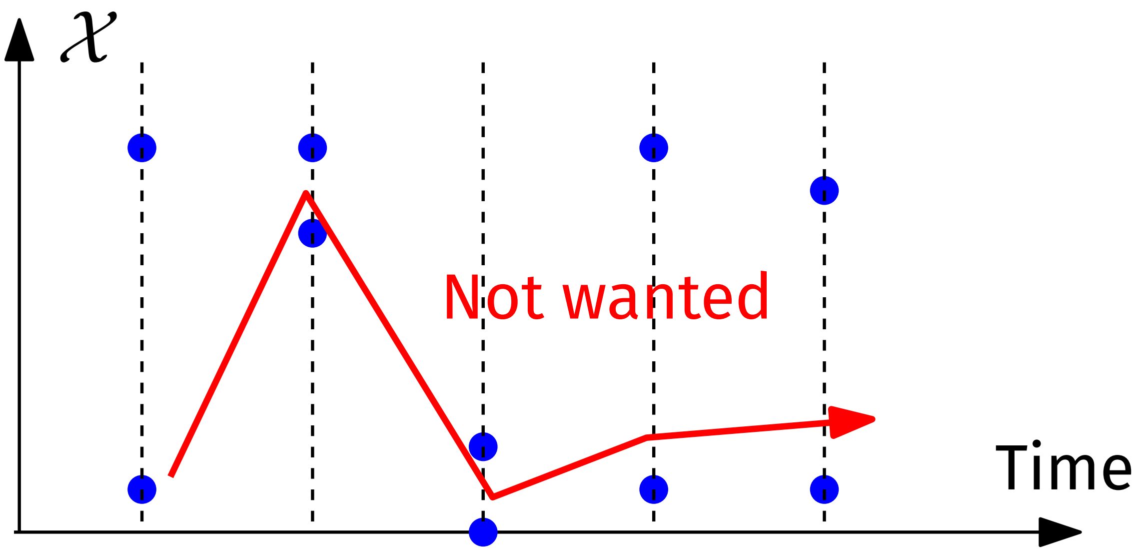
Looking at the ancestors
of a particular set of cells



"Sparse data" framework

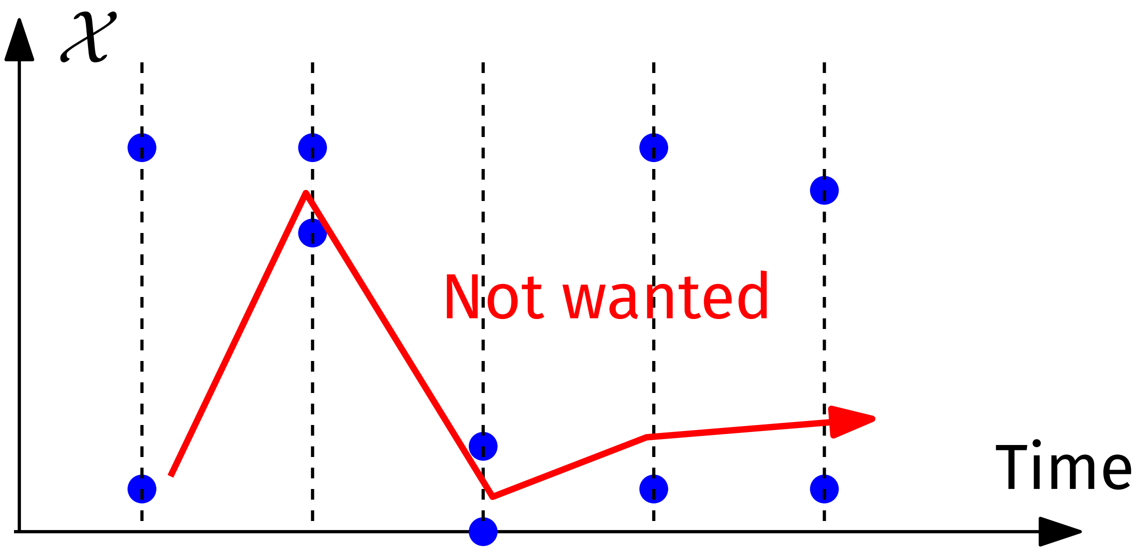


“Sparse data” framework



Few samples per time point, need to share information across time points.

“Sparse data” framework



Few samples per time point, need to share information across time points.

Idea: **data fitting + regularization**

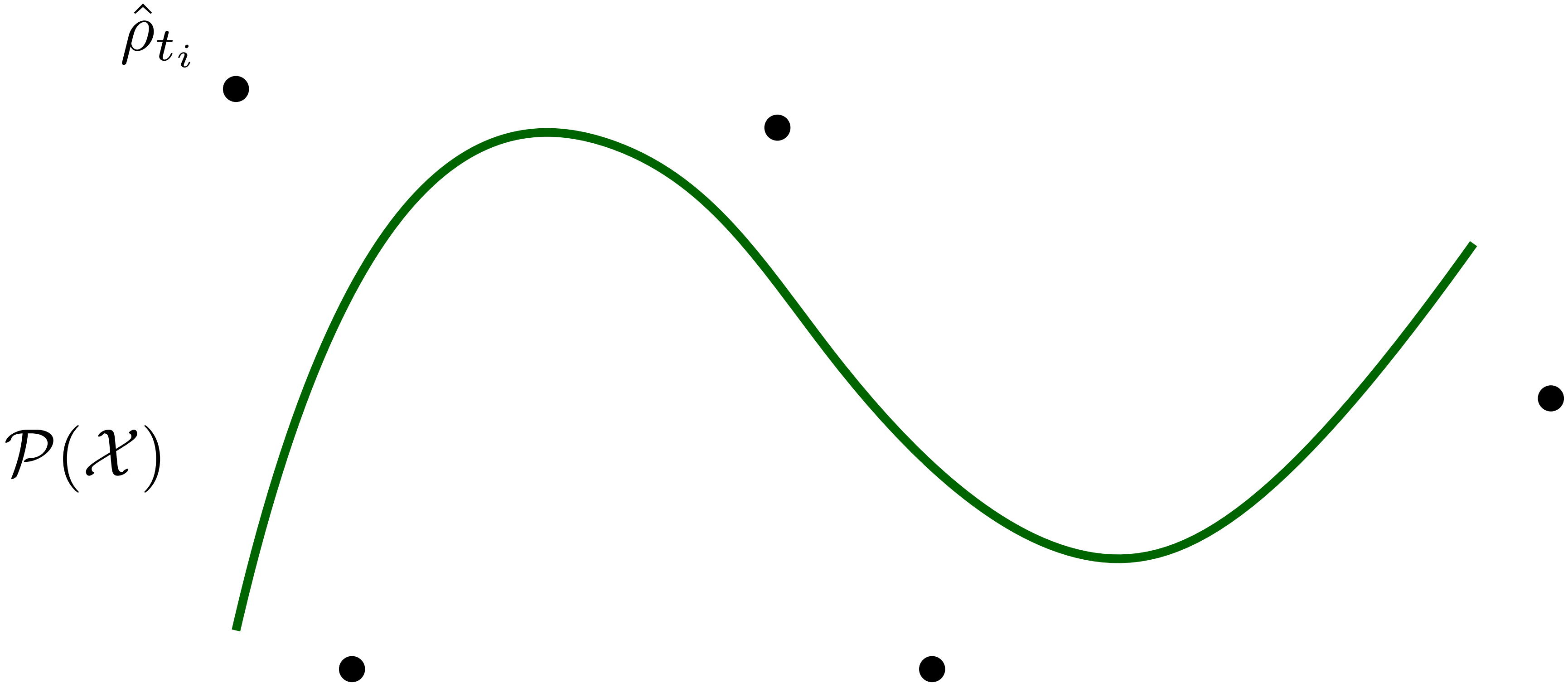
Cross entropy $H(\hat{\rho}_{t_i}|\mathbf{R}_{t_i})$ between data $\hat{\rho}_{t_i}$ and reconstructed marginal $\mathbf{R}_{t_i}$

Sum of optimal transport distances

Schmitzer, Schäfers, Wirth. (2019). Dynamic Cell Imaging in PET With Optimal Transport Regularization.

Global Waddington OT

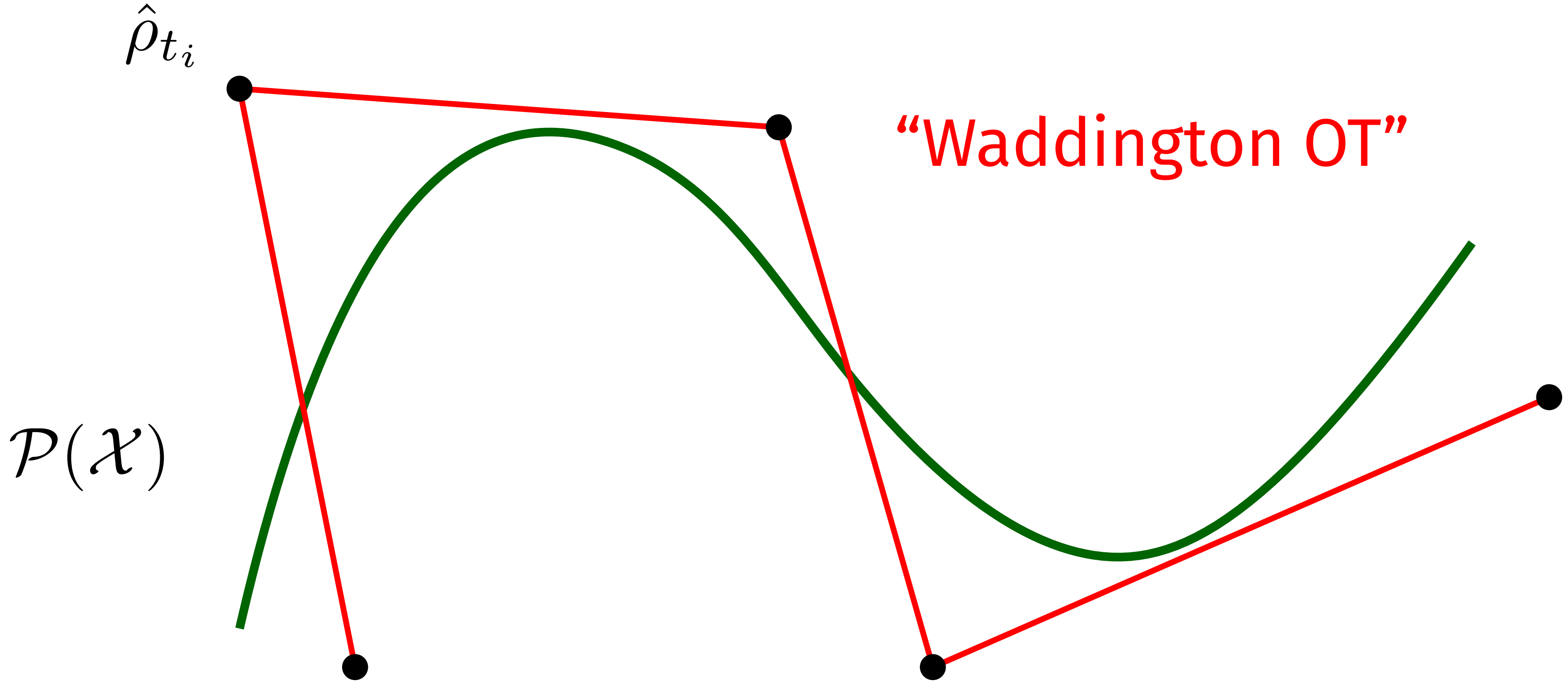
Unknowns: marginals $\mathbf{R}_{t_i}$,



Lavenant, Zhang, Kim, Schiebinger. (2021). Towards a mathematical theory of trajectory inference.

Global Waddington OT

Unknowns: marginals $\mathbf{R}_{t_i}$,



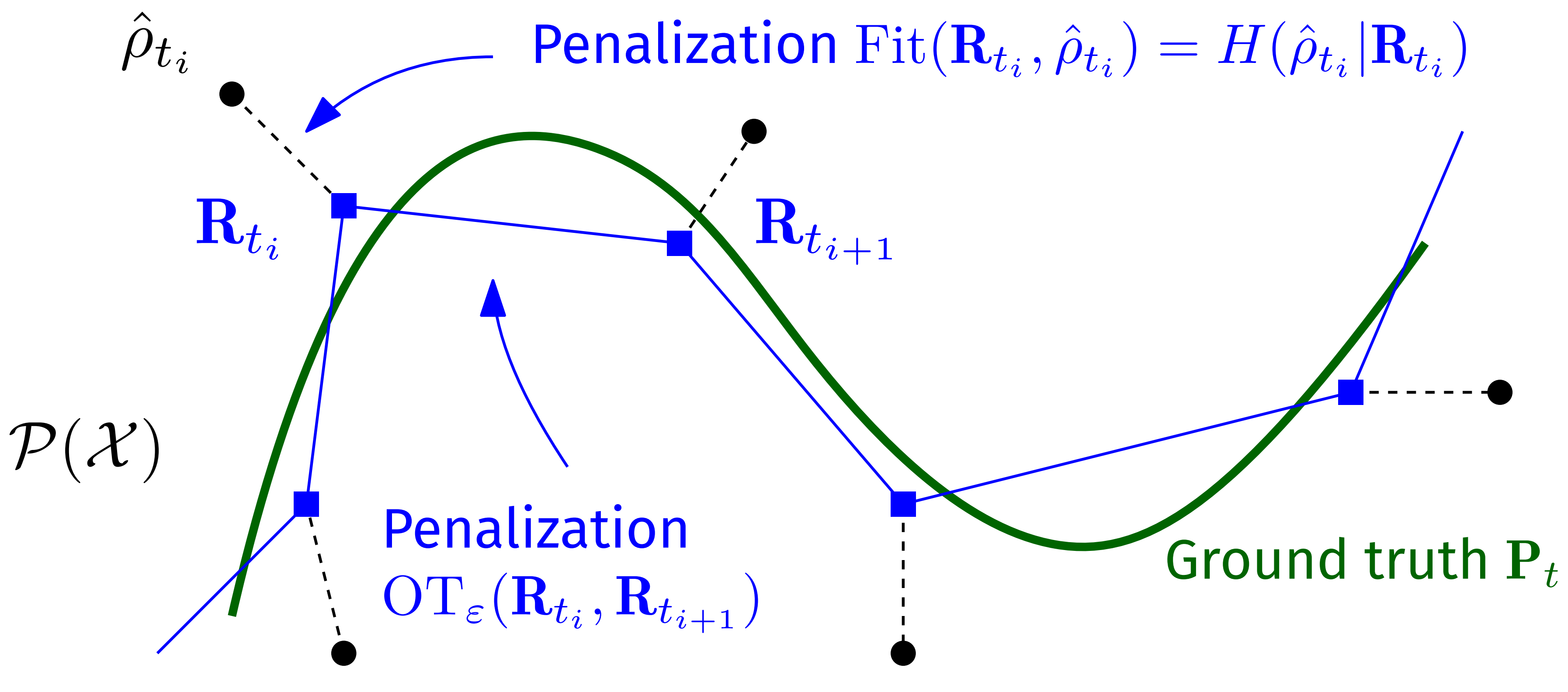
Lavenant, Zhang, Kim, Schiebinger. (2021). Towards a mathematical theory of trajectory inference.

Global Waddington OT

Unknowns: marginals $\mathbf{R}_{t_i}$,

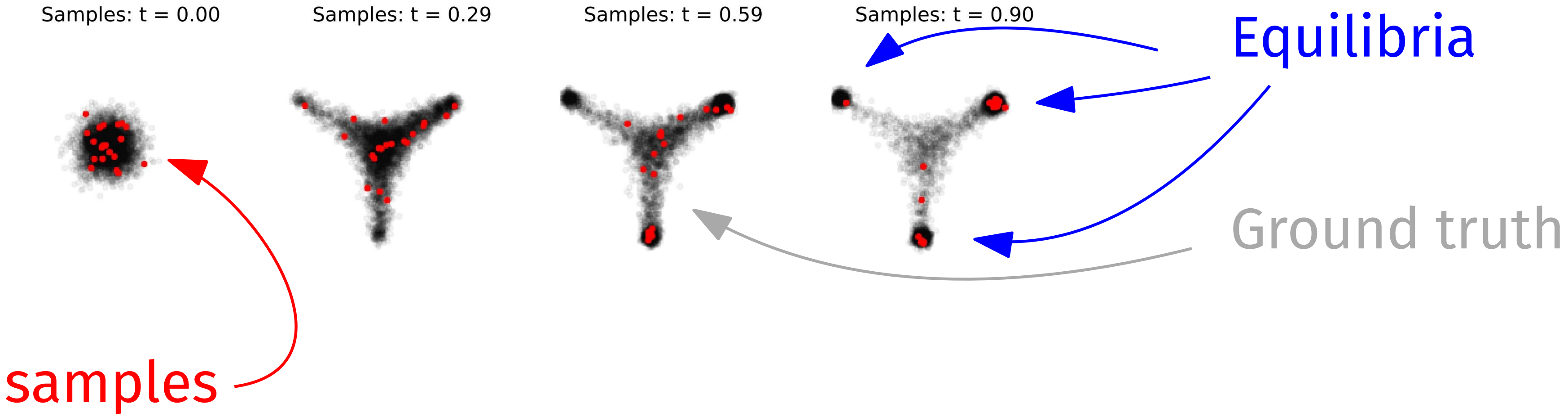
Optimal transport cost

$$\mathrm{Reg}((\mathbf{R}_{t_i})_i) \sim \sum_{i=1}^{T-1} \mathrm{OT}_\varepsilon(\mathbf{R}_{t_i}, \mathbf{R}_{t_{i+1}})$$

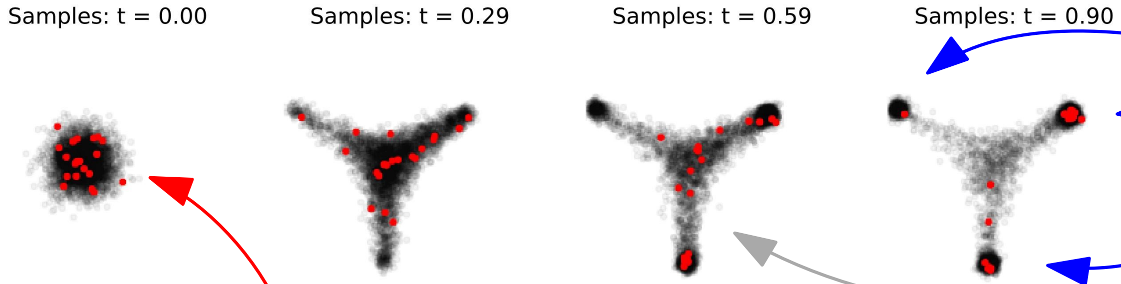


Lavenant, Zhang, Kim, Schiebinger. (2021). Towards a mathematical theory of trajectory inference.

Numerical results (synthetic)



Numerical results (synthetic)

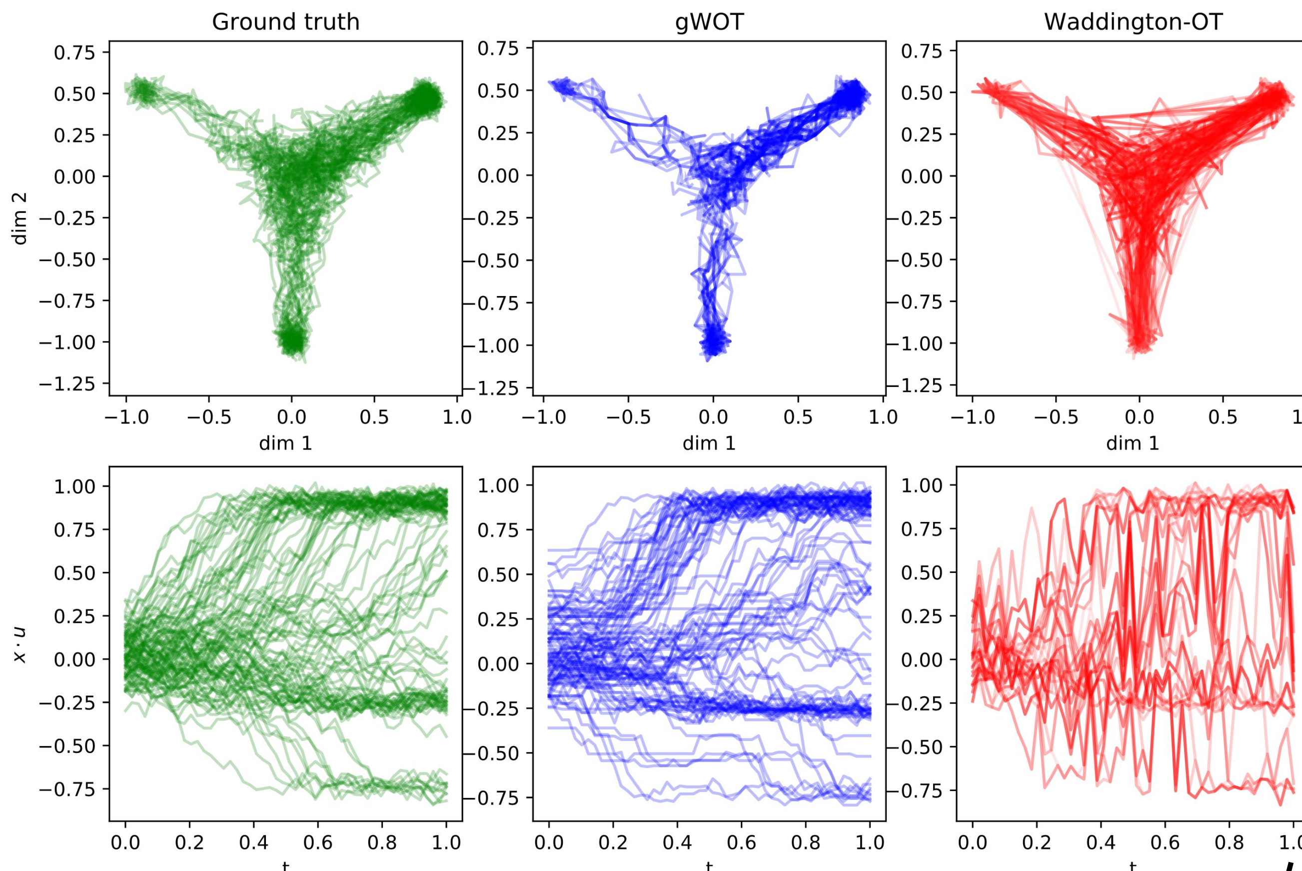


Equilibria

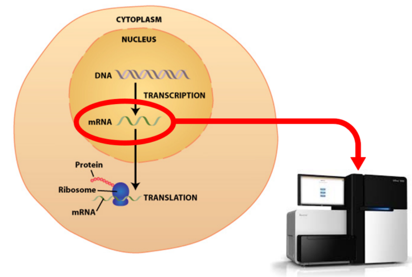
Ground truth

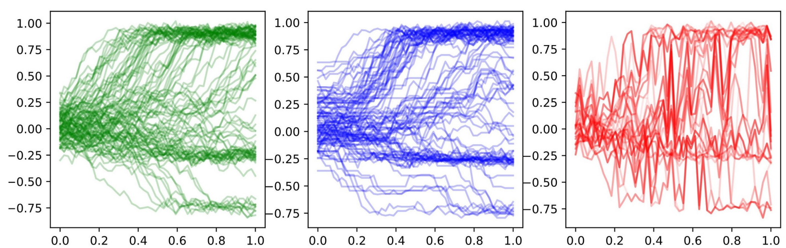
samples

Solving the inference problem



1 - Biological Context





2 - Algorithms and results

3 - Theoretical analysis

$$\mathrm{d}X_t = \mathbf{v}(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t$$

Questions about (global) Waddington OT

In short: temporal couplings are given by optimal transport.

Questions about (global) Waddington OT

In short: temporal couplings are given by optimal transport.

1. How to choose ε entropic parameter?

Questions about (global) Waddington OT

In short: temporal couplings are given by optimal transport.

1. How to choose ε entropic parameter?
2. How can one justify it?

Questions about (global) Waddington OT

In short: temporal couplings are given by optimal transport.

1. How to choose ε entropic parameter?
2. How can one justify it?
3. Does it converge with more and more marginals?

Questions about (global) Waddington OT

In short: temporal couplings are given by optimal transport.

1. How to choose ε entropic parameter?
2. How can one justify it?
3. Does it converge with more and more marginals?

Short answer:

- Works if data is generated by a **potential** Stochastic Differential Equation.
- Choose $\varepsilon = \sigma^2 \Delta t$ with σ noise level in the SDE.

Potential SDEs

The process X_t is a **Stochastic Differential Equation** (SDE):

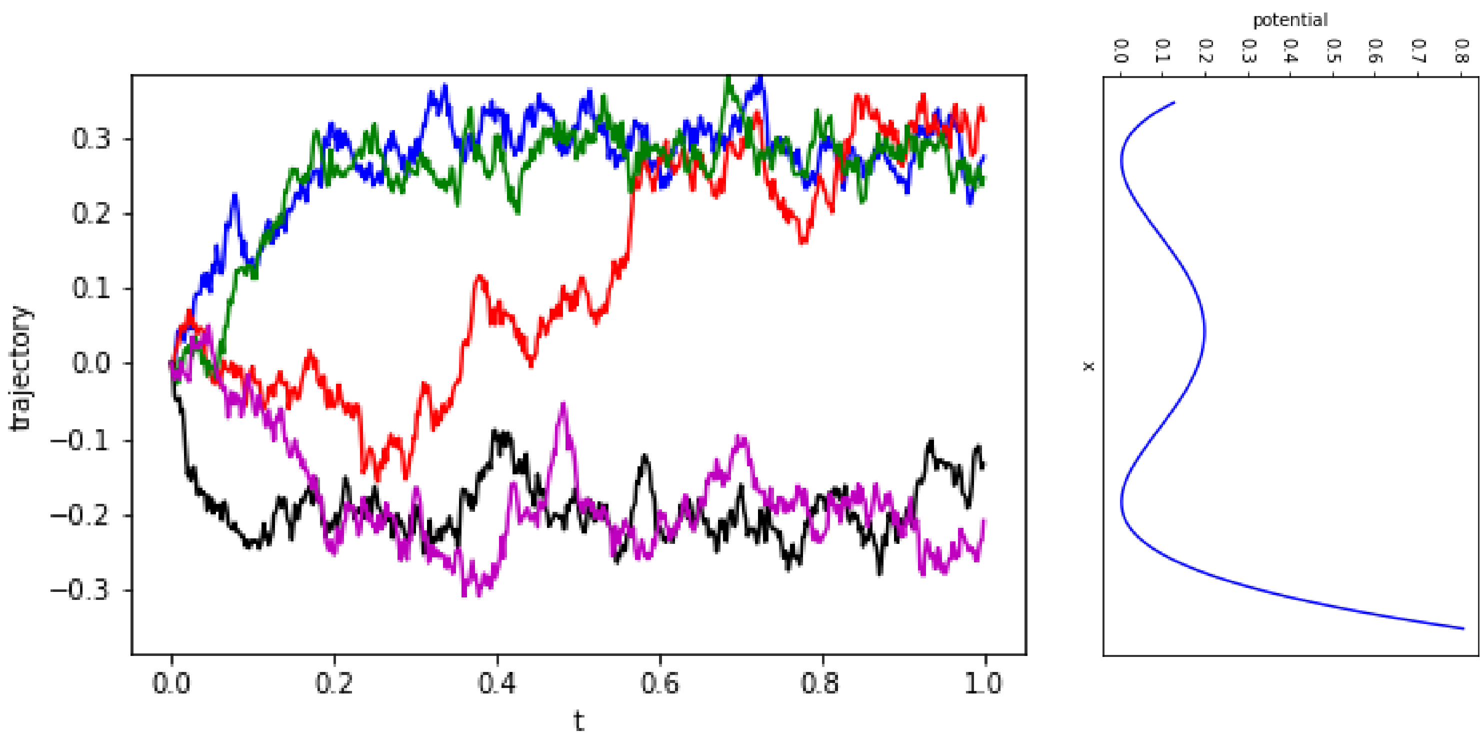
$$\boxed{\mathrm{d}X_t = \mathbf{v}(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t.}$$

Potential SDEs

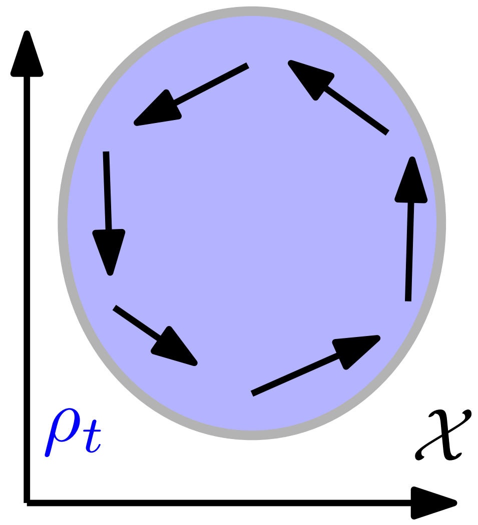
The process X_t is a **Stochastic Differential Equation** (SDE):

$$\mathrm{d}X_t = \mathbf{v}(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t.$$

$\mathbf{v}(t,x) = -\nabla\Psi(t,x)$ for potential $\Psi = \Psi(t,x)$

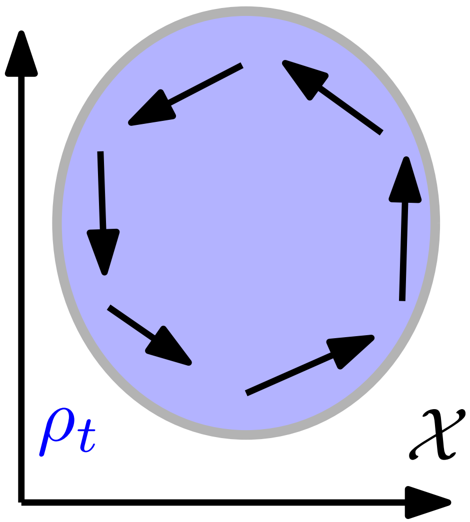


Intuitive explanation: removing identifiability issue



Impossible to distinguish periodic motion from cells at rest.

Intuitive explanation: removing identifiability issue



Impossible to distinguish periodic motion from cells at rest.

Assuming

$$\mathbf{v}(t, x) = -\nabla \Psi(t, x)$$

prevents the velocity field to create periodic motion

Weinreb et al. (2018). Fundamental limits on dynamic inference from single-cell snapshots.

Rigorous result: a variational characterization

$\Omega = C([0,1], \mathcal{X})$, unknown $\mathbf{R} \in \mathcal{P}(\Omega)$.

$$\mathrm{Reg}((\mathbf{R}_{t_i})_i) \sim \sum_{i=1}^{T-1} \mathrm{OT}_{\sigma^2 \Delta t}(\mathbf{R}_{t_i}, \mathbf{R}_{t_{i+1}}) \sim H(\mathbf{R}|\mathbf{W}^\sigma)$$

where $\mathbf{W}^\sigma$ law of Brownian motion with diffusivity σ.

Rigorous result: a variational characterization

$\Omega = C([0,1],\mathcal{X})$, unknown $\mathbf{R} \in \mathcal{P}(\Omega)$.

$$\mathrm{Reg}((\mathbf{R}_{t_i})_i) \sim \sum_{i=1}^{T-1} \mathrm{OT}_{\sigma^2 \Delta t}(\mathbf{R}_{t_i}, \mathbf{R}_{t_{i+1}}) \sim H(\mathbf{R}|\mathbf{W}^\sigma)$$

where $\mathbf{W}^\sigma$ law of Brownian motion with diffusivity σ.

Take $\mathbf{P} \in \mathcal{P}(\Omega)$ law of the SDE

$$\mathrm{d}X_t = -\nabla\Psi(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t.$$

Then

$$H(\mathbf{P}|\mathbf{W}^\sigma) \leq H(\mathbf{R}|\mathbf{W}^\sigma).$$

for any $\mathbf{R} \in \mathcal{P}(\Omega)$ such that $\mathrm{Law}_{\mathbf{P}}(X_t) = \mathrm{Law}_{\mathbf{R}}(X_t)$ for any $t \in [0,1]$.

Rigorous result: convergence of the reconstructed laws

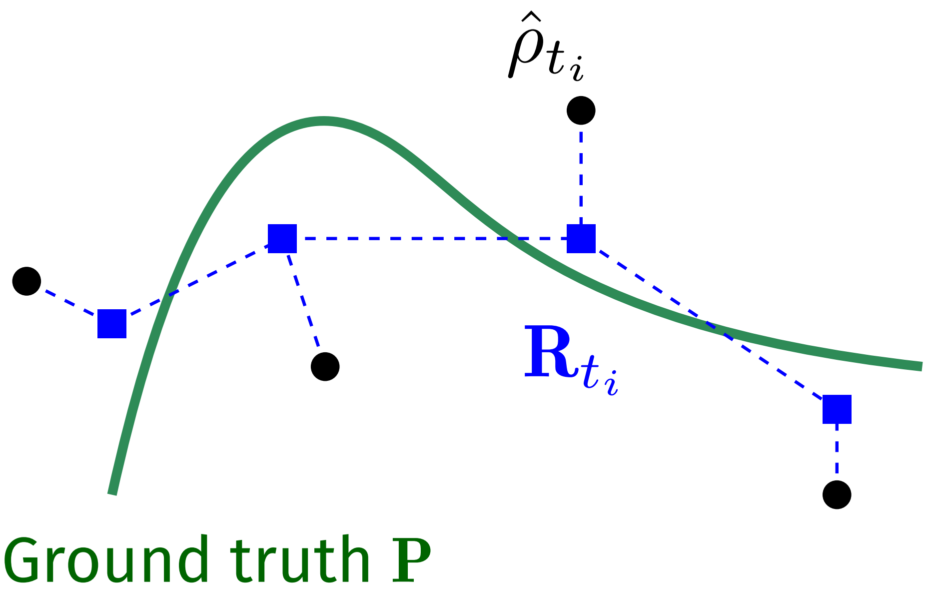
Take $\mathbf{P} \in \mathcal{P}(C([0,1],\mathcal{X}))$ the law of the SDE

$$\mathrm{d}X_t = -\nabla\Psi(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t.$$

Take $(\hat{\rho}_{t_i})_{1 \leq i \leq T}$ samples from X_{t_i}. Call $\mathbf{R}^{T,h,\lambda}$ minimizer of

$$\mathbf{R} \mapsto \text{Data fitting}[(\mathbf{R}_{t_i})_i, (\Phi_h \hat{\rho}_{t_i})_{t_i}] + \lambda \mathrm{Reg}(\mathbf{R})$$

Φ heat flow



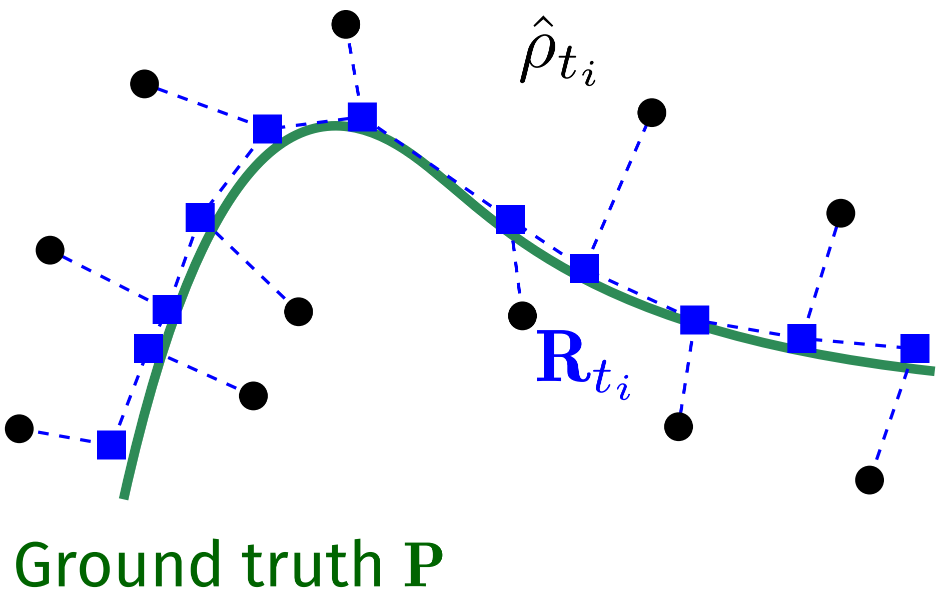
Rigorous result: convergence of the reconstructed laws

Take $\mathbf{P} \in \mathcal{P}(C([0,1],\mathcal{X}))$ the law of the SDE

$$\mathrm{d}X_t = -\nabla\Psi(t, X_t)\mathrm{d}t + \sigma \mathrm{d}B_t.$$

Take $(\hat{\rho}_{t_i})_{1 \leq i \leq T}$ samples from X_{t_i}. Call $\mathbf{R}^{T,h,\lambda}$ minimizer of

$$\mathbf{R} \mapsto \text{Data fitting}[(\mathbf{R}_{t_i})_i, (\Phi_h \hat{\rho}_{t_i})_{t_i}] + \lambda \mathrm{Reg}(\mathbf{R})$$



Then

$$\lim_{h,\lambda \to 0} \left(\lim_{T \to +\infty} \left(\mathbf{R}^{T,h,\lambda}\right) \right) = \mathbf{P}$$

for the topology of narrow convergence.

Conclusion

- Mathematical framework for trajectory inference.
- Guarantees of reconstruction.
- Convex method, but with parameters tuning.

What I have not described

- Extensive numerical experiments.
- How we handle cell division
- Recent work on mesh free numerical optimization

Conclusion

- Mathematical framework for trajectory inference.
- Guarantees of reconstruction.
- Convex method, but with parameters tuning.

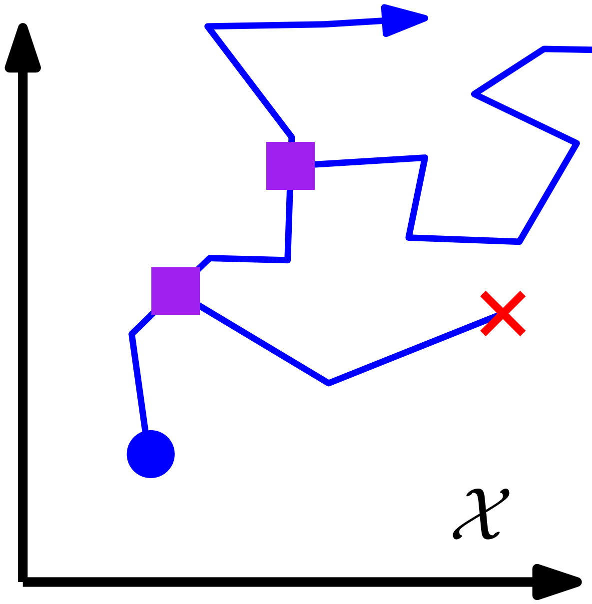
What I have not described

- Extensive numerical experiments.
- How we handle cell division
- Recent work on mesh free numerical optimization

Thank you for your attention

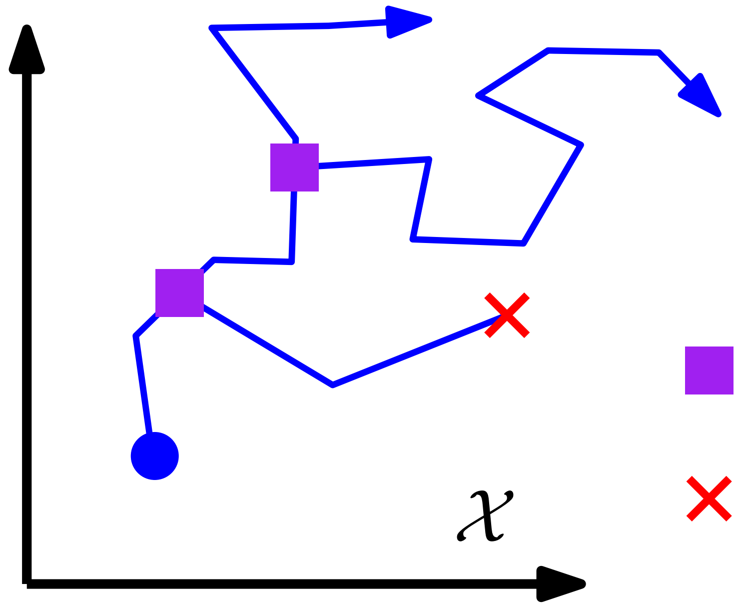
Zhang, Chizat, Heitz and Schiebinger. (2022). Trajectory Inference via Mean-field Langevin in Path Space.

What about branching?



In reality cells divide and die.

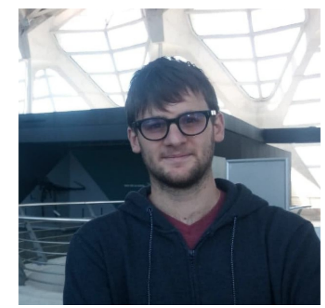
What about branching?



In reality cells divide and die.

■ Branching

× Death



In progress (with Aymeric Baradat): studying entropy minimization with respect to the law of the **Branching Brownian Motion**.

Baradat and L. (2021). Regularized optimal transport as entropy minimization with respect to branching Brownian motion.

Handling growth in our paper: splitting

Before: unknowns marginals $\mathbf{R}_{t_i}$,

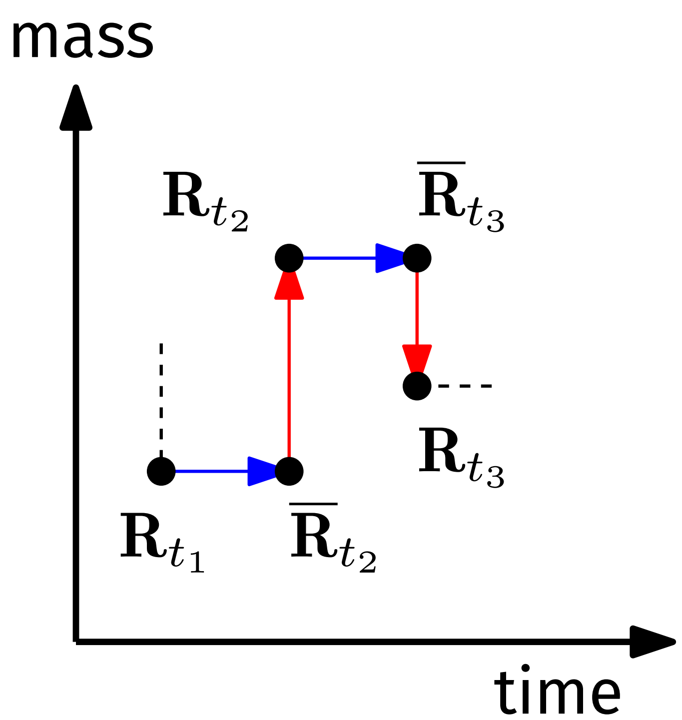
$$\mathrm{Reg}((\mathbf{R}_{t_i})_i) \sim \sum_{i=1}^{T-1} \mathrm{OT}_\varepsilon(\mathbf{R}_{t_i}, \mathbf{R}_{t_{i+1}}).$$

Handling growth in our paper: splitting

Before: unknowns marginals $\mathbf{R}_{t_i}$,

$$\mathrm{Reg}((\mathbf{R}_{t_i})_i) \sim \sum_{i=1}^{T-1} \mathrm{OT}_\varepsilon(\mathbf{R}_{t_i}, \mathbf{R}_{t_{i+1}}).$$

To handle **branching**: alternance of transport and growth phases.

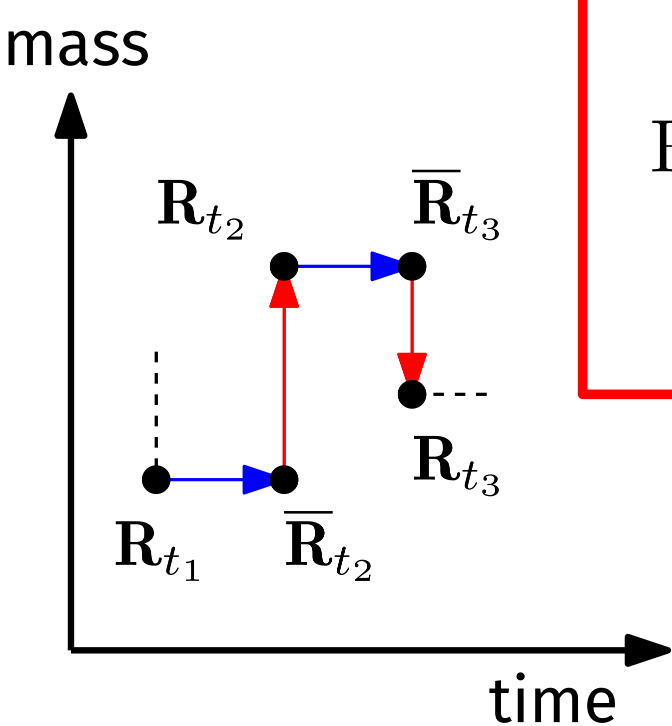


Handling growth in our paper: splitting

Before: unknowns marginals $\mathbf{R}_{t_i}$,

$$\mathrm{Reg}((\mathbf{R}_{t_i})_i) \sim \sum_{i=1}^{T-1} \mathrm{OT}_\varepsilon(\mathbf{R}_{t_i}, \mathbf{R}_{t_{i+1}}).$$

To handle **branching**: alternance of transport and growth phases.



$$\mathrm{Reg}((\mathbf{R}_{t_i})_i, (\overline{\mathbf{R}}_{t_i})_i) = \sum_i \mathrm{OT}_\varepsilon(\mathbf{R}_{t_i}, \overline{\mathbf{R}}_{t_{i+1}}) + G(\overline{\mathbf{R}}_{t_i}, \mathbf{R}_{t_i})$$

$G(\overline{\mathbf{R}}_{t_i}, \mathbf{R}_{t_i})$ measures discrepancy (e.g. KL) between $\overline{\mathbf{R}}_{t_i}(x)\exp(\Delta t\, g(x))$ and $\mathbf{R}_{t_i}(x)$ with $g : \mathcal{X} \to \mathbb{R}$ a priori growth rate.