Optimal Transport Theory and Its Applications

Toshiaki Yachimura (谷地村 敏明)

2024年12月17日
MathCCS交流会

My research topics

shape optimization problem, inverse problem, **Optimal Transport theory**, (single-cell) data analysis

My research on optimal transport theory

Entropic Gaussian mixture optimal transport and single-cell trajectory inference

- scEGOT: single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport, BMC Bioinformatics, accepted.

Optimal transport theory in persistent homology and its application to machine learning

- Topological Node2vec: Enhanced Graph Embedding via Persistent Homology, JMLR, 2024.

Dimensionality reduction in Wasserstein space via Sinkhorn MDS

- Visualizing Shape Functionals via Sinkhorn Multidimensional Scaling, arXiv:2409.14687

Theoretical study on Tsallis entropic regularized optimal transport

- Convergence rate of Tsallis entropic regularized optimal transport, arXiv:2304.06616
- Information-geometric perspectives and convergence analysis of Tsallis entropic regularized optimal transport, in preparation.

My research topics

shape optimization problem, inverse problem, **Optimal Transport theory**, (single-cell) data analysis

My research on optimal transport theory

Entropic Gaussian mixture optimal transport and single-cell trajectory inference

- scEGOT: single-cell trajectory inference framework based on entropic Gaussian mixture optimal transport, BMC Bioinformatics, accepted.

Optimal transport theory in persistent homology and its application to machine learning

- Topological Node2vec: Enhanced Graph Embedding via Persistent Homology, JMLR, 2024.

Dimensionality reduction in Wasserstein space via Sinkhorn MDS

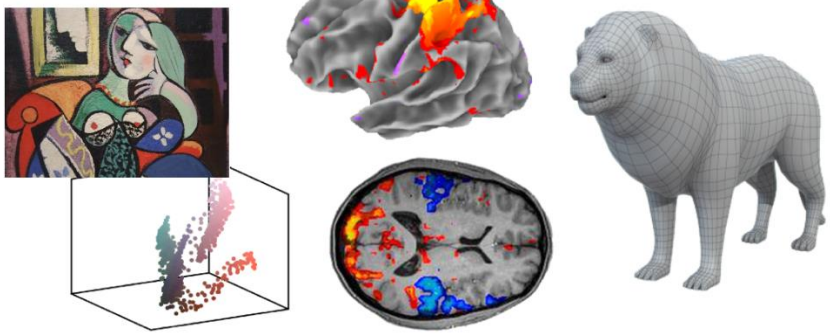
- Visualizing Shape Functionals via Sinkhorn Multidimensional Scaling, arXiv:2409.14687

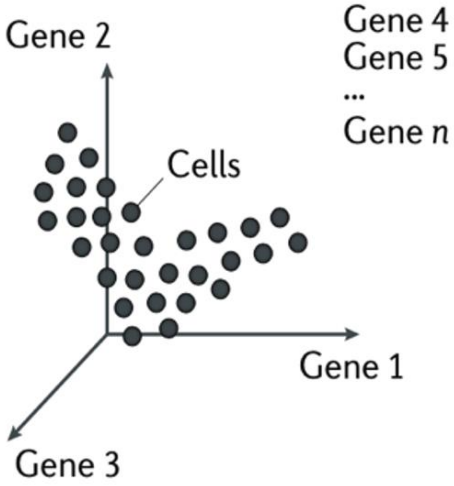
Theoretical study on Tsallis entropic regularized optimal transport

- Convergence rate of Tsallis entropic regularized optimal transport, arXiv:2304.06616
- Information-geometric perspectives and convergence analysis of Tsallis entropic regularized optimal transport, in preparation.

Short introduction to optimal transport

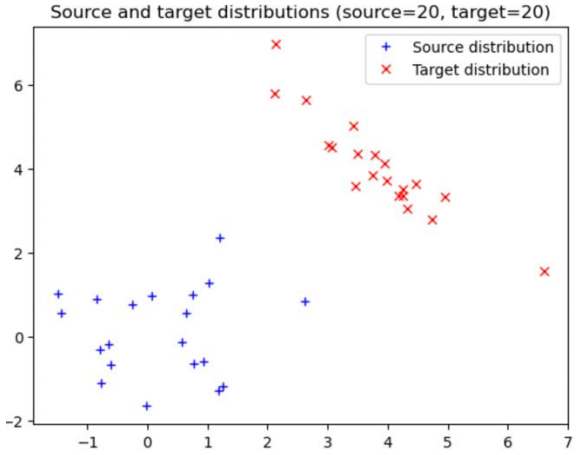
Images, vision, graphics, etc --> Most data can be viewed as probability distribution

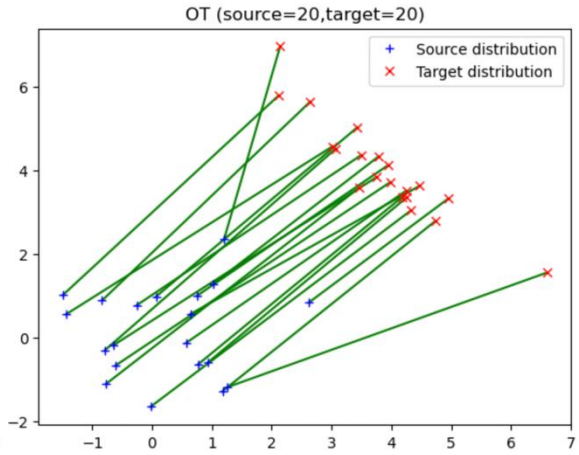


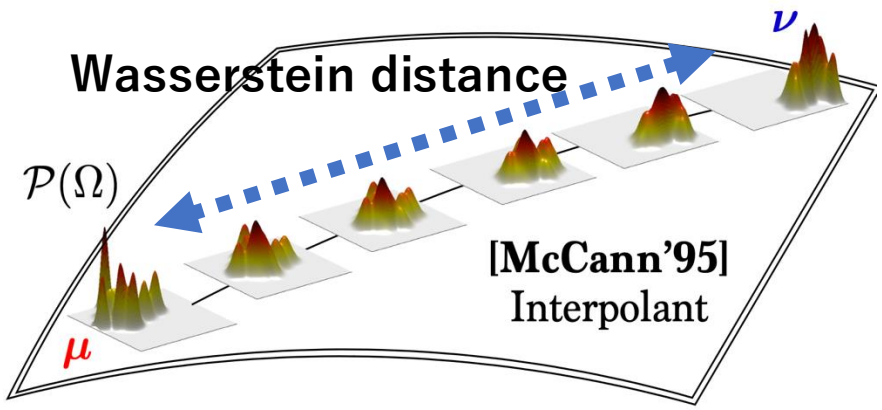


https://optimaltransport.github.io/

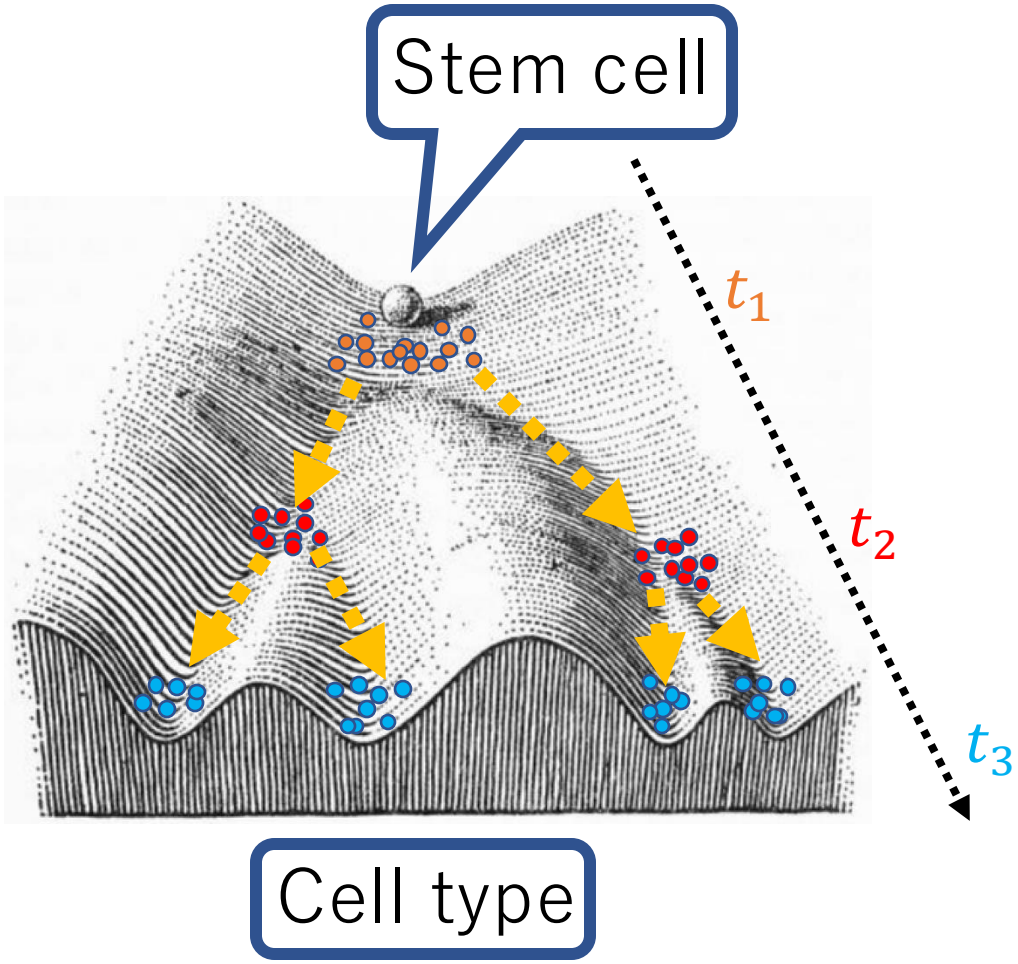
Optimal Transport theory: Provide the (**Wasserstein**) **distance** and **matching** (optimal transport plan) between source probability distribution and target probability distribution



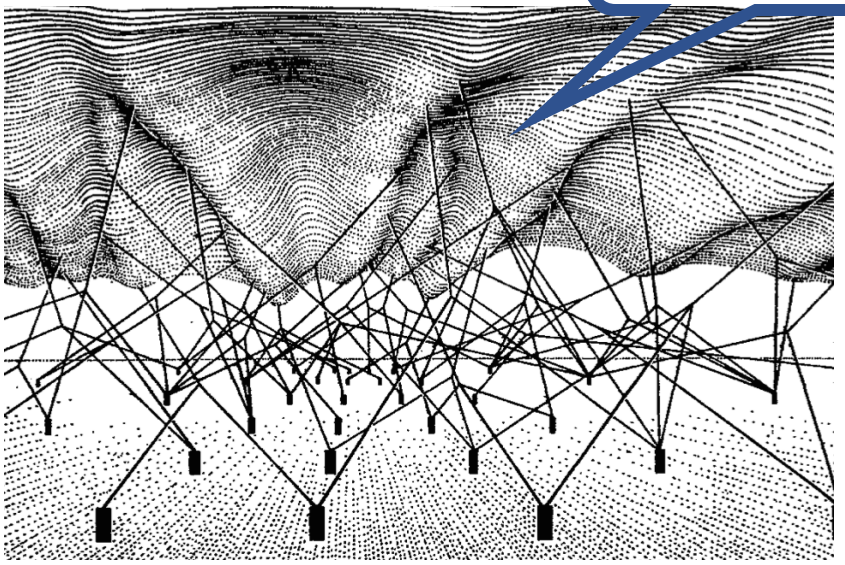




Waddington's landscape (1957)
Conceptual figure of the dynamics for cell differentiation



$$\frac{d\mathbf{x}_{\text{cell}}}{dt} = f(\mathbf{x}_{\text{cell}})$$

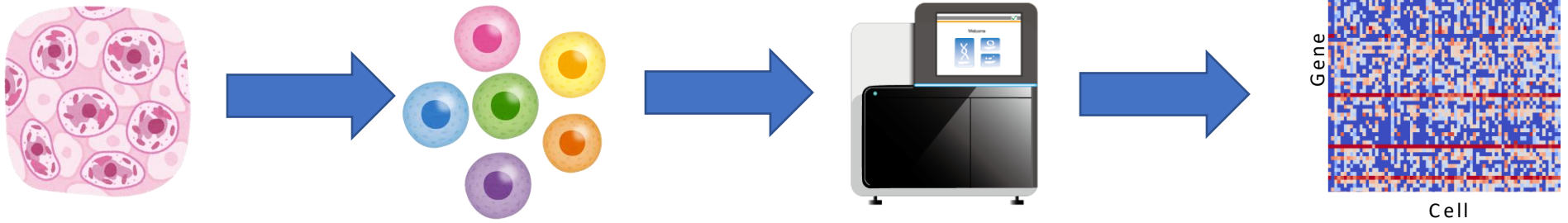


Gene Regulatory Networks (**GRN**s)
behind Waddington's landscape

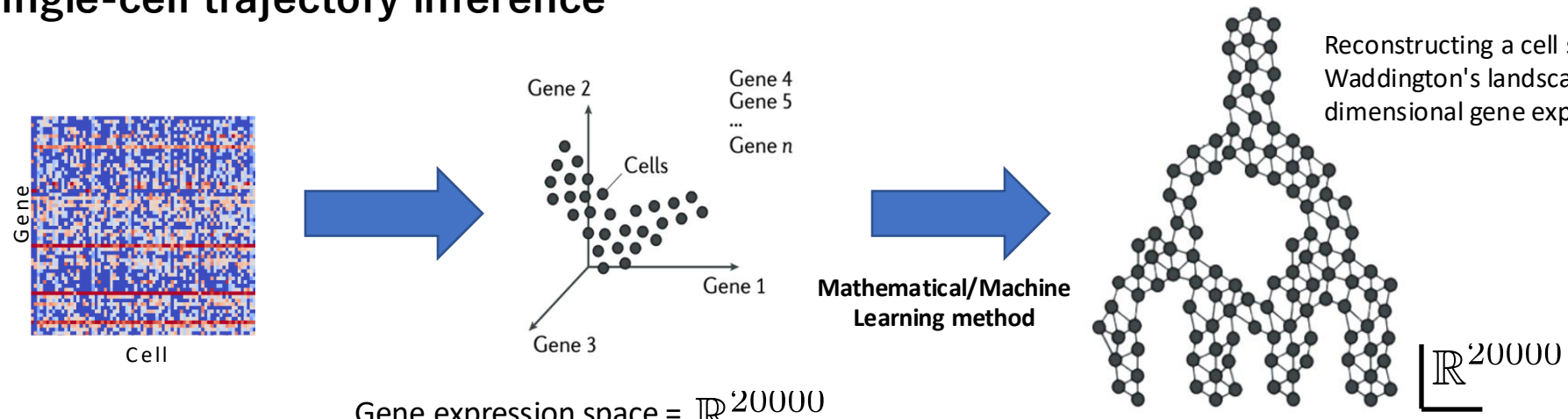
- What is the **"shape"** of Waddington's landscape in real cell differentiation?
- What are **marker genes that shape it**?

Infer trajectory and dynamics of cell differentiation by scRNA-seq data

Single-cell RNA sequencing (scRNA-seq)



Single-cell trajectory inference

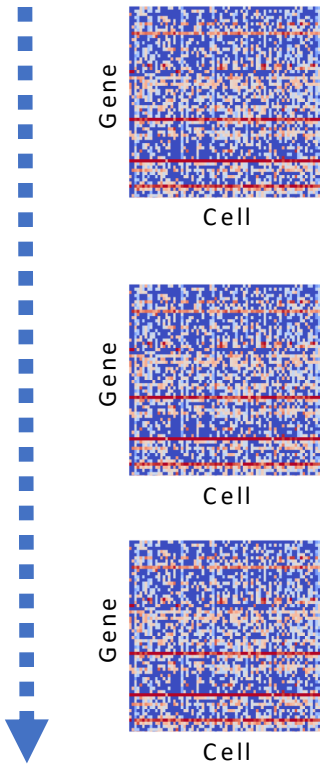


Gene expression space = $\mathbb{R}^{20000}$

Problems

It is difficult to infer the dynamics (such as the velocity and intermediate states of cell differentiation) of the cell differentiation process.

Problem of snapshots (time-series) of scRNA-seq data

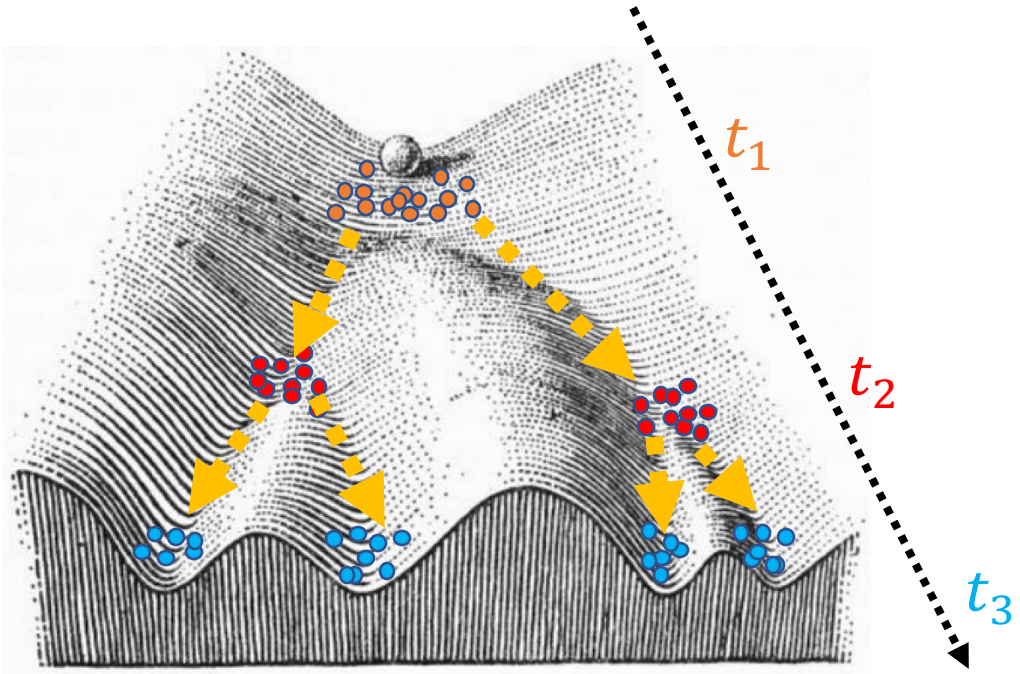


t_1

t_2

t_3

scRNA-seq



Mathematical question:

Infer the dynamics $\frac{d\mathbf{x}_{\text{cell}}}{dt} = f(\mathbf{x}_{\text{cell}})$ by given data

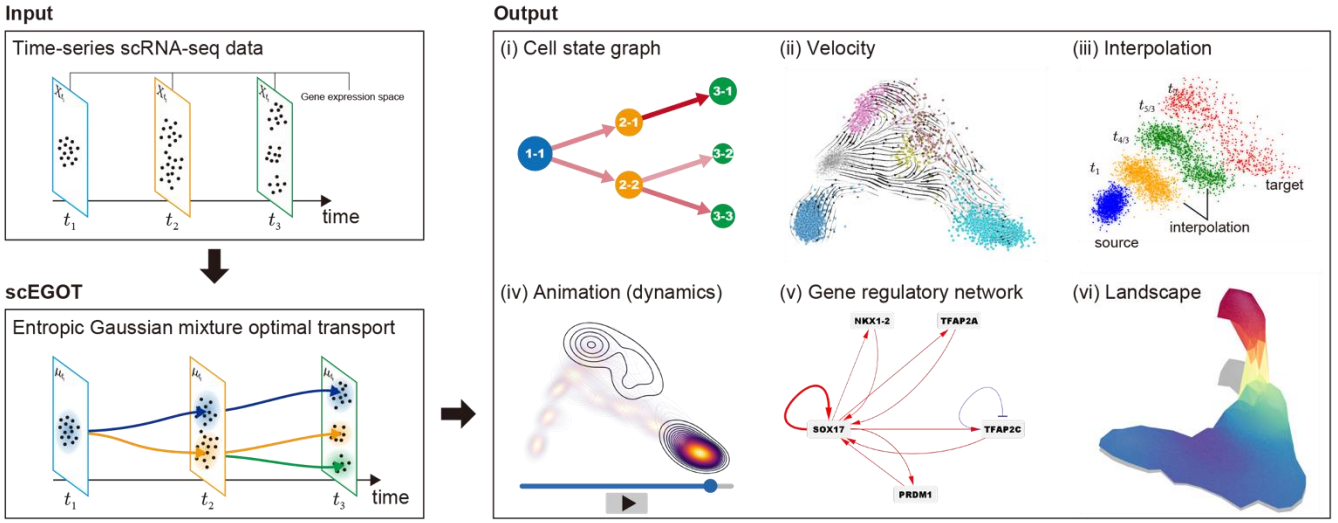
Difficulty: Cells are killed in RNA sequencing (**disappearing labels**)

Connect the most related cells in the sense of minimizing transportation costs (**Optimal Transport, OT**)

Results

Results

1. scEGOT: single cell trajectory inference framework by Entropic Gaussian mixture Optimal Transport(T. Yachimura, et al., BMC Bioinformatics, accepted, github: https://github.com/yachimura-lab/scEGOT)



- Cell populations should differentiate while forming sub-populations (clusters).
- Assuming that the data on cell populations is a Gaussian mixture distributions, we consider OT (**EGOT**) between them. -> OT between clusters

Advantages :

- The computational cost is extremely low because it is the optimal transport between clusters. (Generation of intermediate states in 2D distributions, **EGOT = 41ms**, usual OT = 1h 44min 49s)
- Correspondence with continuous OT (**generative model**) .
- High interpretability in biology (each Gaussian distribution = each cell type)

2. Discovery of marker genes in the precursor cell population of primordial germ cells and genes that inhibits differentiation