Problem Statement

Single-cell technologies are destructive making it impossible to measure a cell lineage.

The goal is to model individual cells as dynamic and continuously evolving entities over gene space from population level data.

This gives a continuous picture of lineage from coarse-grained time measurements.

Main Idea

Constrain continuous normalizing flows to match assumptions on cellular dynamics.

Background

Dynamic Optimal Transport [1] Defined as the flow with the minimum energy from a source distribution to a target distribution.

This relates the kinetic energy of a fluid flow to the Wasserstein distance.


Method

We approximate dynamic optimal transport over gene space with continuous normalizing flows. We add the following regularizations to build in priors on cellular dynamics.

\( R_{\text{energy}} \) – Minimize kinetic energy and particle time derivative terms encouraging more Euclidean optimal and energy efficient paths.

\( R_{\text{energy}} \) – At random time interpolations minimize distance to K nearest neighbors for each point preferring trajectories close to the manifold of observed cells.

\( R_{\text{velocity}} \) – Minimize angle between RNA velocity data and flow at every measured cell building in knowledge of RNA splicing data.

Finally, we learn a growth function \( g \) for modeling the cell growth rate using unbalanced optimal transport (see Fig. 3).

Conclusions

Our results relate continuous normalizing flows to dynamic optimal transport and create a flexible model for single cell population modeling.

Our model allows for interpolation of trajectories to unobserved cell states inferring future and past states of individuals from population data.

Further exploration is needed in efficient learning of stochastic and unbalanced models of cell populations.

References


Further information

Email: alexander.tong@yale.edu
Code: github.com/KrishnaswamyLab
Supported by IVADO (l'institut de valorisation des données) and the Chan-Zuckerberg Initiative.