Fast Diffusion Optimal Transport for Manifold-of-Manifold Embeddings

Alexander Tong[1], Manik Kuchroo[1], Guillaume Huguet[2], Ronald Coifman[1], Guy Wolf[2 †] and Smita Krishnaswamy[2 †]

[1] Yale University, [2] University of Montreal; MILA, [†] Equal contribution

Main Idea

Compute the geometric structure of many distributions over the same manifold efficiently by linking the dual Kantorovich distance to graph diffusions.

Problem Statement

There are many datasets where we would like to understand the geometry between distributions.

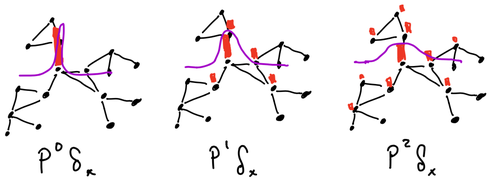
However, useful exact distances between point clouds are prohibitively costly to compute.

Here we approximate the geometry between distributions using the dual formulation of the Wasserstein distribution to first "bin" the space of the graph before computing distribution nearest neighbors.

Background

Graph Diffusion Kernels [1] Centered at each point, we can construct a kernel that is the graph equivalent of a gaussian by powering the natural diffusion operator *P* at each point where

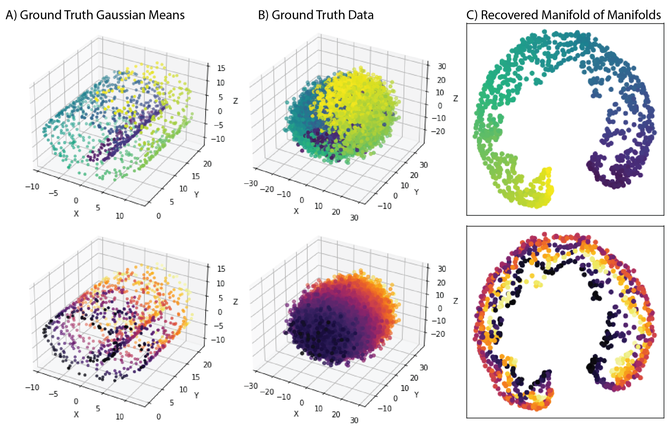
$P = D^{-1}A$



Kantorovich Dual Formulation Transport distances can also be formulated in the dual space as an integral probability measure, namely

$$\mathcal{W}_\alpha(\mu,\nu) := \left(\inf_{\pi\in\Pi(\mu,\nu)}\int_{\mathcal{X}\times\mathcal{X}} d(x,y)^\alpha d\pi(x,y)\right)^{1/\alpha}$$
$$= \left(\sup_{f:|f(x)-f(y)|\le d(x,y)^\alpha}\int_{\mathcal{X}} f(x)d\mu - \int_{\mathcal{X}} f(y)d\nu\right)^{1/\alpha}$$

Results

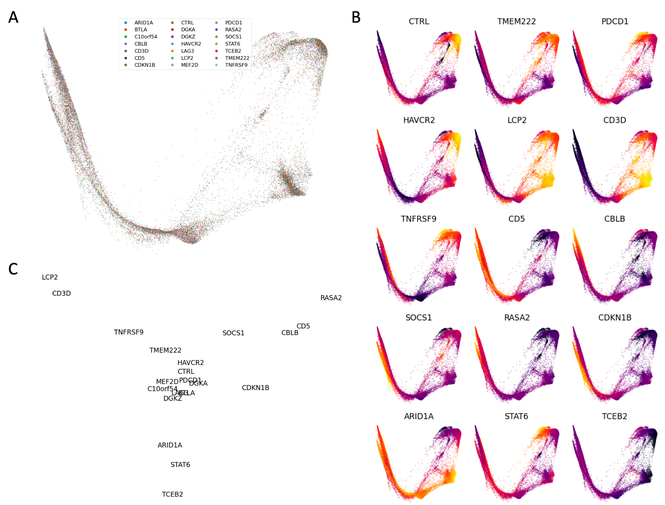


1000 gaussians of 100 points each centered on 2D "Swissroll" manifold in 3D ambient space. We unroll this manifold by constructing distances between distributions using our method which is ~50X faster than computing all exact Wasserstein distances.

Task 1: Embed Gaussians centered on the swiss roll.

Time complexity:
For m distributions each with n points,
Exact: $O(m^2n^3)$ for exact pairwise Wasserstein distances using network flow
Ours: $O(m^2n \log mn)$ calculations.

	Time (seconds)
Diffusion EMD	354.0
Exact	17400.0
Sinkhorn	38200.0



Task 2: Embed CRISPER perturbed distributions of T-Cells to understand perturbation similarity.

Enzymes dataset class exchange preferences vs. experimentally observed. LEGS-FCN learns a function which still preserves exchange preferences between classes. Better than the GCN baseline.

Method

Algorithm:

1) Build a nearest neighbors graph between points
2) Compute and concatenate K diffusion scales using Chebyshev approximation
3) Down sample and reweight scales to remove redundant information in large scales.
4) Compute K-nearest neighbors between distributions in L_1 embedding space induced by scales.

Multiscale Diffusion Distances [2] An alternative characterization of Wasserstein distances on a manifold was explored in [2]. Using the dual formulation, we construct the function f from a set of multiscale diffusion kernels. For a diffusion operator *P*, the Wasserstein distance on the diffusion space can be characterized by the weighted scales of *P* applied to the difference of the distributions.

$$\mathcal{W}_\alpha(\mu,\nu) = \sum_{k\ge 0} 2^{-k\alpha}\|P_{2^{-k}}(\mu-\nu)\|_1$$

This is similar to the work in [3] which explored tree approximations to the Wasserstein distance.

Conclusions

Computing the structure between many distributions using the Wasserstein distance is slow but can be accelerated by embedding distributions into vector spaces where fast nearest neighbor calculation is possible.

In spaces with natural diffusion geometries we can compute these vector spaces using a series of subsampled multiscale diffusion distances extremely quickly.

References

[1] Coifman R. and Lafon S. Diffusion Maps (2006)

[2] Leeb W. Topics in metric approximation (2015)

[3] Le T. Yamada M. Fukumizu K. and Cuturi M. Tree-Sliced Variants of Wasserstein Distances (2019)

Further information

Email: alexander.tong@yale.edu

Work supported by IVADO, the Chan-Zuckerberg Initiative, and the NIH.