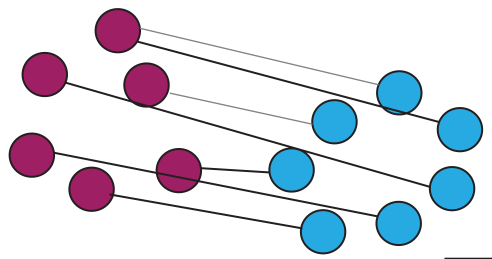
Contact Me Full Paper

Wasserstein Wormhole - Scalable Optimal Transport Distance with Transformers

Doron Haviv[1,2], Russell Z. Kunes[1,3], Thomas Dougherty[1,2], Cassandra Burdziak[1], Tal Nawy[1], Anna Gilbert[4], Dana Pe'er[1,5]

[1]MSKCC [2]Cornell, Weill [3]Columbia [4]Yale [5]HHMI

Optimal Transport seeks the ideal mapping between point-clouds:



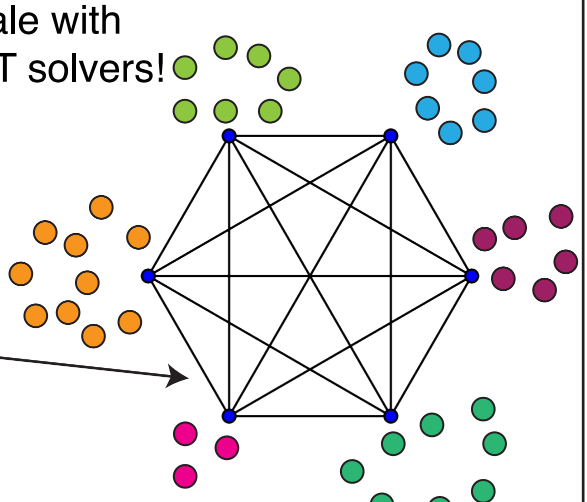
Ground cost distance between points

$$\min_{P \in U(\nu,\mu)} \sum_{i,j} c(x_i, y_j) P_{i,j} - \varepsilon H(P)$$

Mass preserving pairings

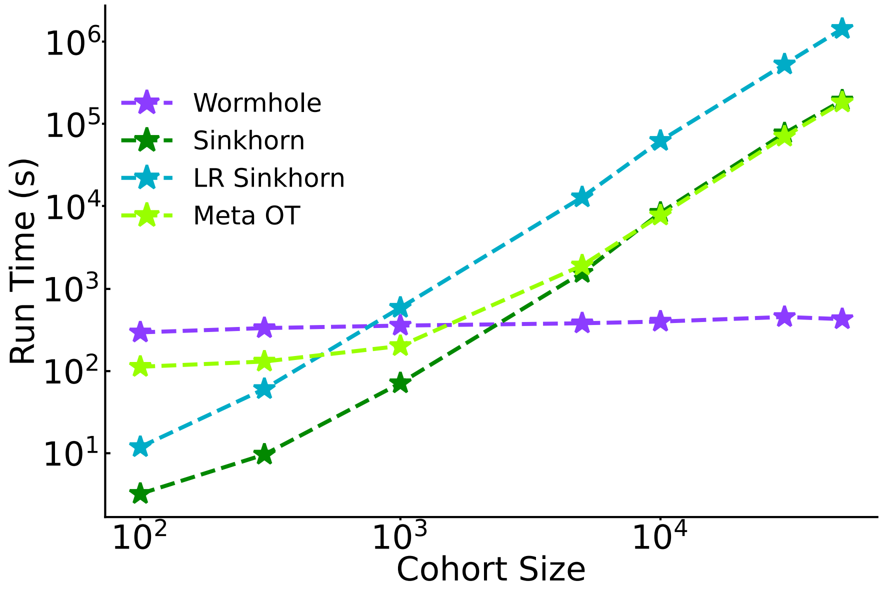
Entropic regularization

But it does not scale with cohort size - N^2 OT solvers!



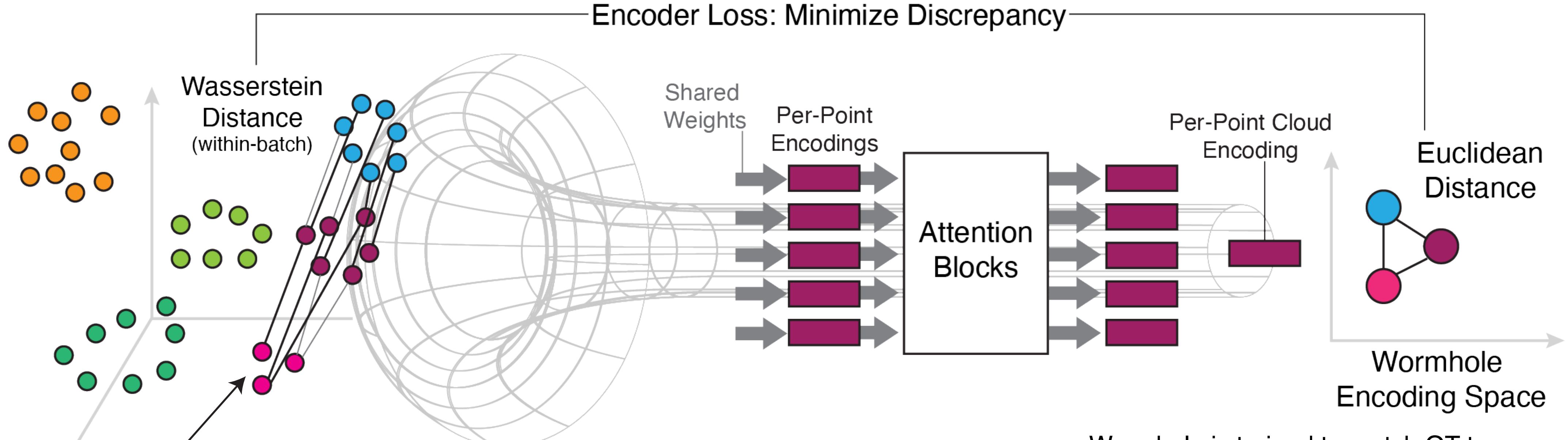
Each OT calculation is **quadratic** in point-cloud sizes

Even on small cohorts, full pairwise OT can take weeks!



Solution - Find OT preserving embedding

Wormhole estimates OT distances in linear time.



OT is permutation equivariant:

$$OT(X,Y) = OT(\Pi_{perm}(X), Y)$$

but so is **Attention**!

Wormhole is trained to match OT to embedding distance:

$$\mathcal{L}_{enc} = \sum_{i,j} \left| \left| \alpha_i - \alpha_j \right|^2 - OT(X_i, X_j) \right|^2$$

But Wasserstein cannot be embedded perfectly...

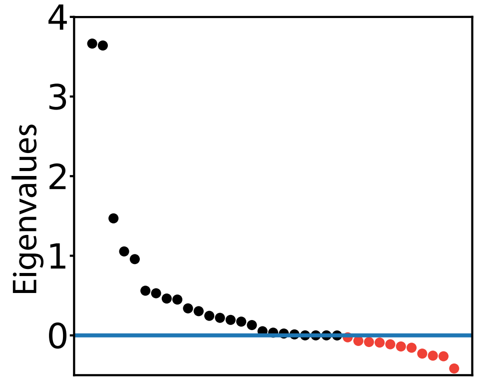
A pairwise distance matrix:

$$D_{i,j} = OT(X_i, X_j)$$

Is Euclidean if:

$$-(I - \frac{\mathbf{1}^T\mathbf{1}}{n}) D (I - \frac{\mathbf{1}^T\mathbf{1}}{n}) \succcurlyeq 0$$

However...



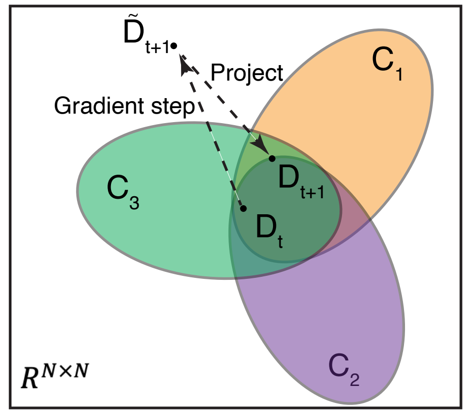
So how close can we get?

$$\min_{D'} |D - D'|^2$$

s.t. $D'_{i,j} \geq 0$ $D'_{i,i} = 0$ $-JD'J \succcurlyeq 0$

C_1: Positive C_2: Hollow C_3: Euclidean

Fully convex! Projected-Gradient is guarenteed optimal!



On small examples, Wormhole approaches optimum!

Dataset	Lower	Upper	PGD	Wormhole
Simplex (35)	0.765	6.369	1.117	1.420
Gaussian (128)	0.129	2.552	0.168	0.401
MNIST (256)	0.042	0.616	0.058	0.100

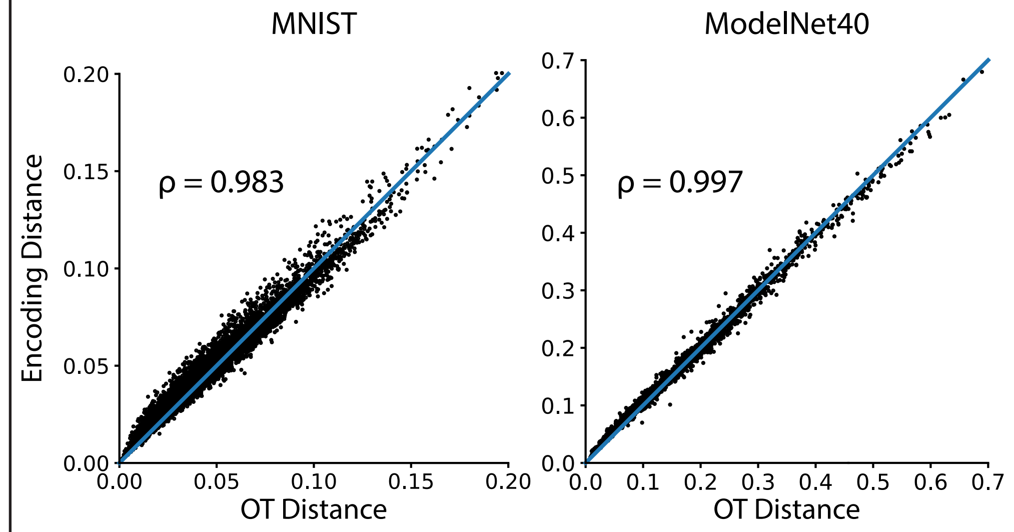
And the bounds are:

$$\sum_{i:\lambda_i<0} \lambda_i^2 \leq \mathcal{L}^* \leq \sum_{i,j} (\Delta g_i + \Delta g_j)^2 + \sum_{i:\lambda_i<0} \lambda_i^2$$

Where:

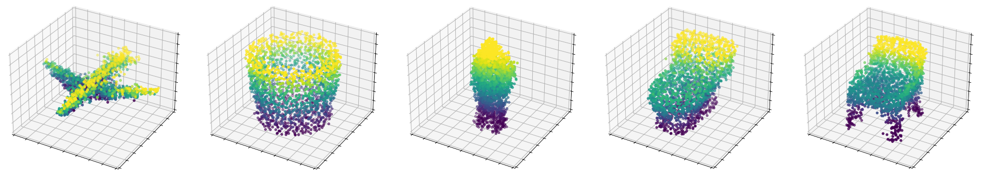
$$\Delta g_i = \frac{1}{2} \sum_{j:\lambda_j<0} \lambda_j v_{i,j}^2$$

Wormhole accurately embeds Wasserstein:

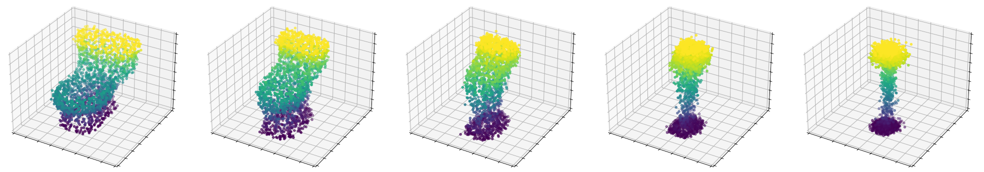


Can quickly solve variational problems:

Wormhole Barycenters



Wormhole Interpolates



And recover niche signal from high-dimensional spatial transcriptomics data: