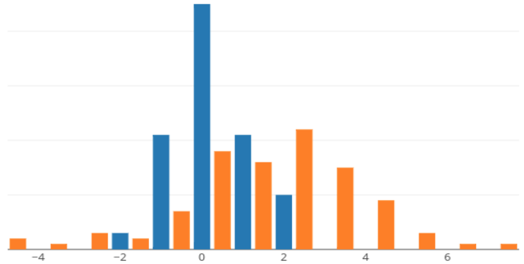
Linear Partial Gromov Wasserstien Embedding

Yikun Bai, Abihith Kothapalli, Hengrong Du, Rocio Diaz Martin, Soheil Kolouri

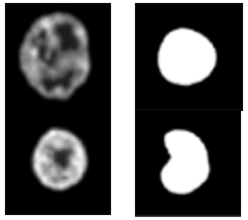
Postdoctoral Scholar, MINT Lab

Vanderbilt University, Nashville, TN

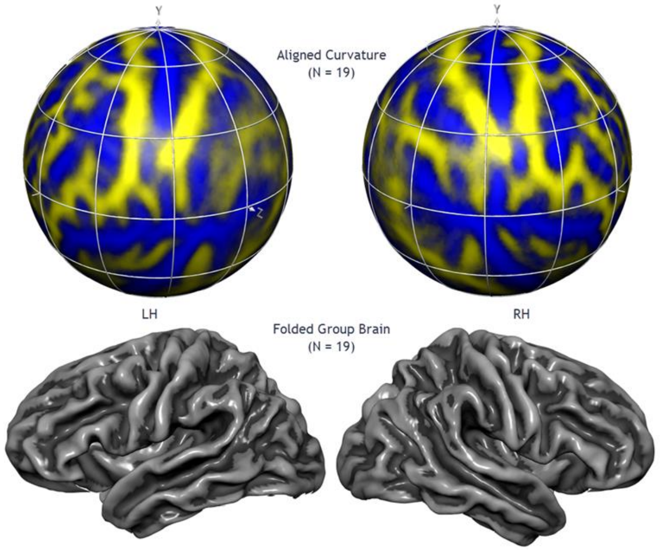
Motivation: Data representation in Machine learning



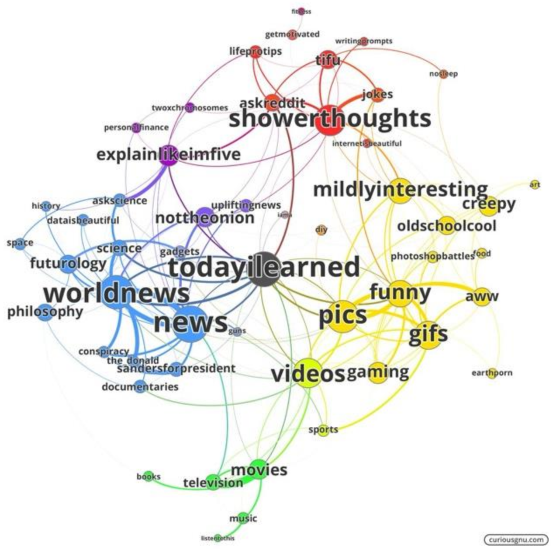
1D –Histograms



2D – Distributions



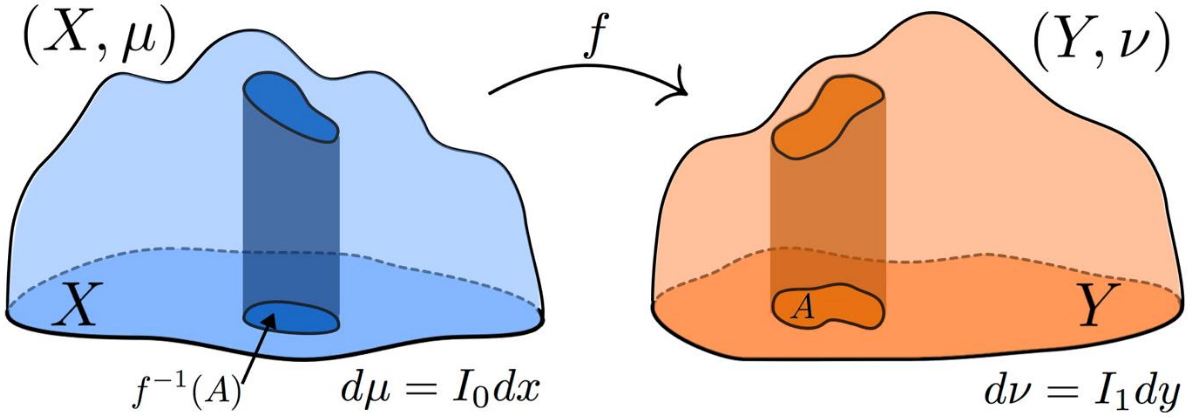
Data in Manifold



Graphic data

nD – Distributions (n>>3)

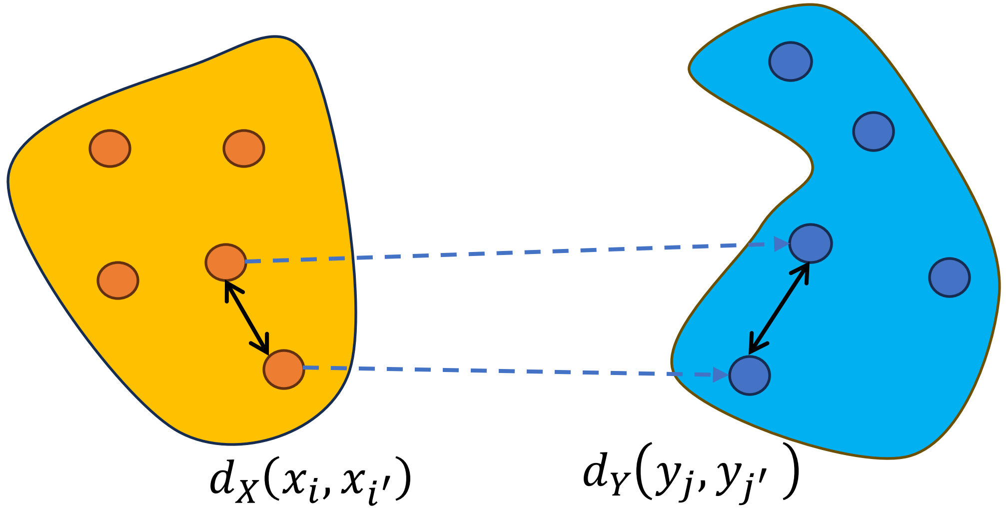
Background: Optimal transport and Gromov Wassersein



Optimal transport problem: Given probability measures $\mu = \sum_{i=1}^{n} p_i \delta_{x_i}, \nu = \sum_{j=1}^{m} q_j \delta_{y_j}$

$$OT(\mu,\nu) \coloneqq \min_{\gamma \in \mathbb{R}^{n\times m}} \sum_{i,j} c(x_i, y_j)\gamma_{ij},$$
$$\text{s.t.}, \gamma 1_m = p, \gamma^\top 1_n = q$$

- OT defines a metric between μ and ν
- Complexity of OT: $\mathcal{O}(n^3)$



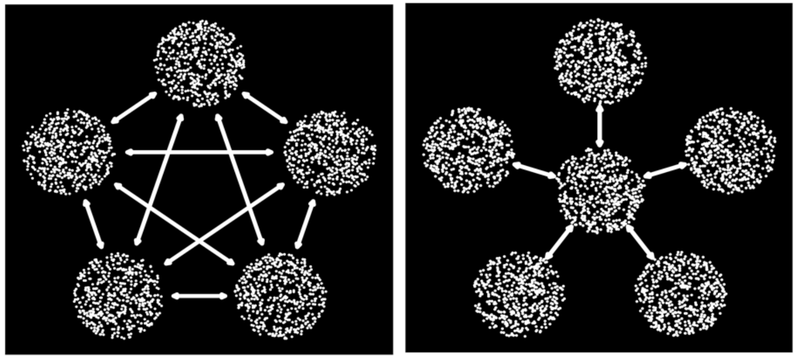
Gromov Wasserstein problem: Given mm-spaces $\mathbb{X} = (X, d_X, \mu), \mathbb{Y} = (Y, d_Y, \nu)$

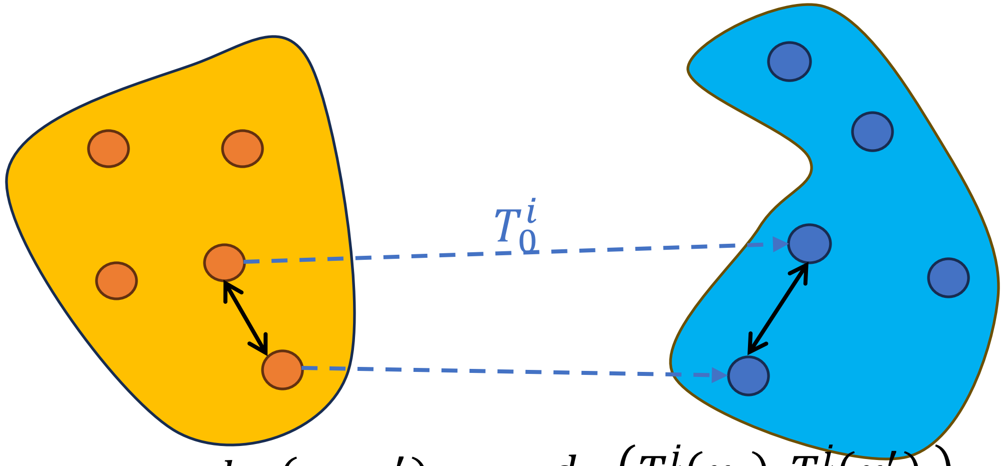
$$GW(\mathbb{X},\mathbb{Y}) \coloneqq \min_{\gamma \in \mathbb{R}_+^{m\times n}} \sum_{i,i',j,j'} |d_X(x_i, x_{i'}) - d_Y(y, y')| \gamma_{ij}\gamma_{i'j'},$$

subject to: $\gamma 1_m = p_\mu, \gamma^\top 1_n = p_\nu$

- GW defines a metric between $\mathbb{X}$ and $\mathbb{Y}$.
- Computational complexity (Frank-Wolfe): $\mathcal{O}\left(\frac{n^4}{\epsilon^2} \cdot n^3 \ln(n)\right)$

Background: Linear Gromov Wasserstein embedding

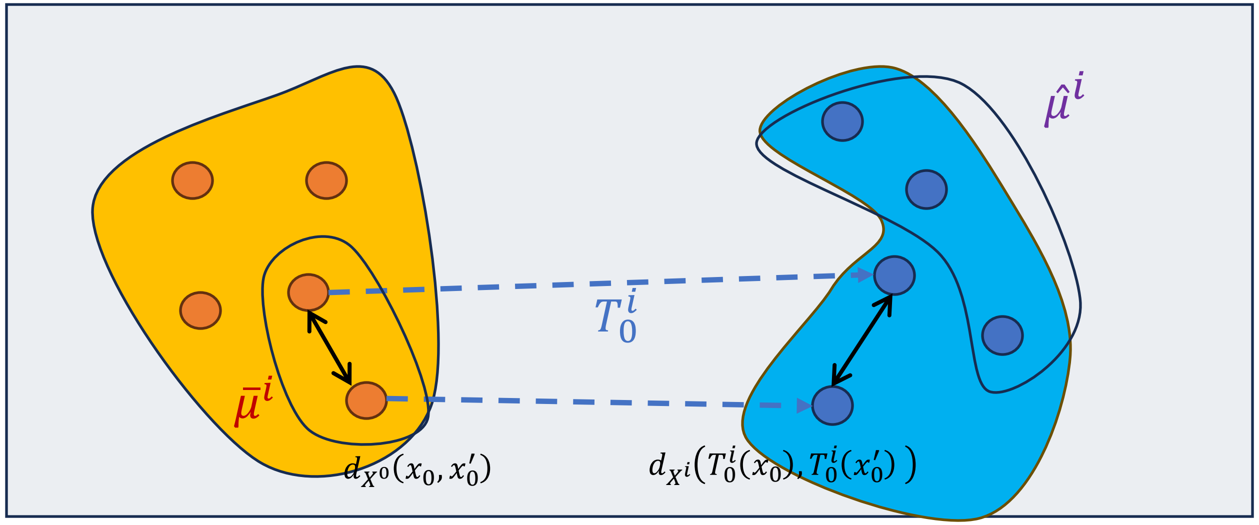




Given K measures, how to efficiently compute their GW distances?

- Linear GW embedding: $\mathbb{X}^i = \left(X^i, d_{X^i}, \mu^i\right) \mapsto v^i \coloneqq d_{X^i}\left(T_0^i(x_0), T_0^i(x_0')\right) - d_{X^0}(x_0, x_0') \in L^2_{sys}\left((X^0)^2\right)$
- Linear GW distance: $\mathrm{GW}\left(\mathbb{X}^i, \mathbb{X}^j\right) \approx LGW\left(\mathbb{X}^i, \mathbb{X}^j\right) \coloneqq \left\|v^i - v^j\right\|^2_{L^2\left((\mu^0)^{\otimes 2}\right)}$.
- Linear GW defines a metric.
- Computational complexity:
  - Embedding construction $\mathrm{K} \cdot \mathcal{O}\left(n^3 \ln(n) \cdot \frac{n^4}{\epsilon^2}\right)$ for K measures:
  - Distance computation: $\binom{K}{2} \cdot \mathcal{O}(n^2)$

Our contribution: Linear Partial Gromov Wasserstein embedding



- Linear Partial GW embedding: Suppose $\gamma = \left(id \times T_0^i\right)_{\#}\mu^0$ is an optimal solution for $PGW\left(\mathbb{X}^0, \mathbb{X}^i\right)$, then the LPGW embedding is defined by:

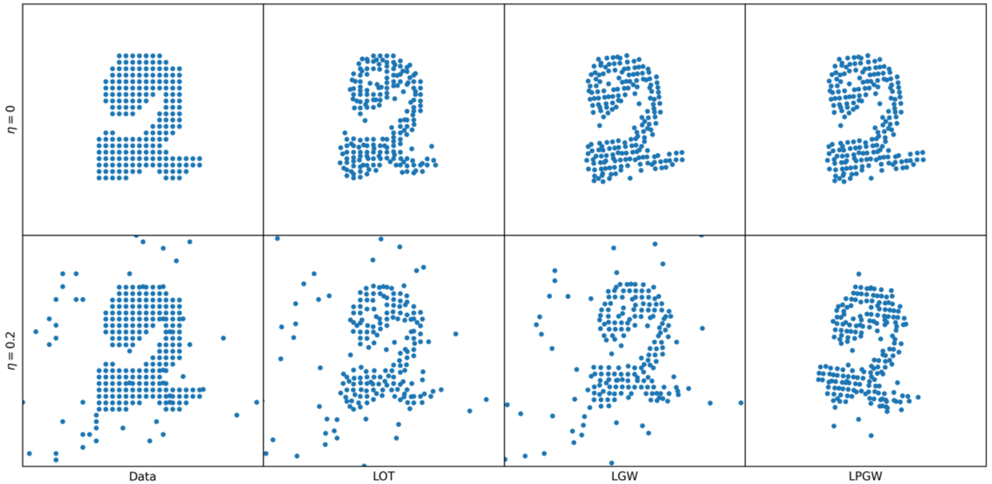
  $\mathbb{X}^i = \left(X^i, d_{X^i}, \mu^i\right) \mapsto \left(v^i, \bar{\mu}^i, \hat{\mu}^i\right)$ *where*

  - $v^i := d_{X^i}\left(T_0^i(x_0), T_0^i(x_0')\right) - d_{X^0}(x_0, x_0')$ (displacement)
  - $\bar{\mu}^i = \gamma_0$ (information for mass destruction/transportation in source domain)
  - $\hat{\mu}^i = \left(\mu^i\right)^{\otimes 2} - (\gamma_1)^{\otimes 2}$ (mass creation in the target domain

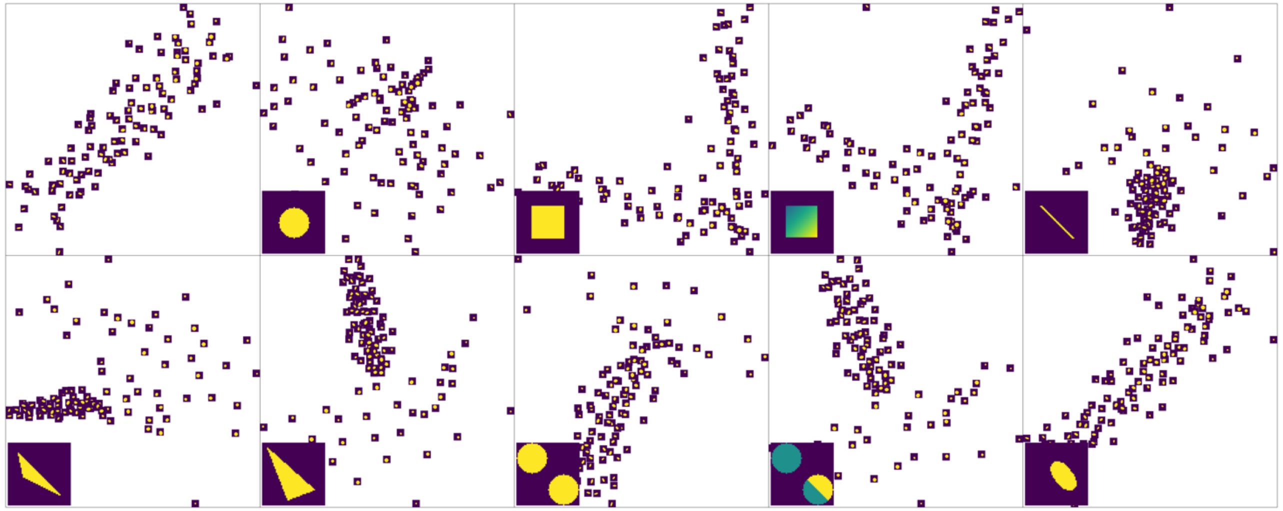
- Linear PGW distance: $\mathrm{PGW}\left(\mathbb{X}^i, \mathbb{X}^j\right) \approx LPGW\left(\mathbb{X}^i, \mathbb{X}^j\right) := \left\|v^i - v^j\right\|^2_{L\left(\left(\bar{\mu}^i \wedge \bar{\mu}^j\right)^{\otimes 2}\right)} + \lambda\left(\left|\left(\bar{\mu}^i\right)^{\otimes 2} - \left(\bar{\mu}^j\right)^{\otimes 2}\right| + \left|\hat{\mu}^i + \hat{\mu}^j\right|\right).$

Properties of Linear Partial Gromov Wasserstein distance

- Linear PGW defines a metric.
- Computational complexity:
  - Embedding construction $K \cdot \mathcal{O}\left(n^3 \ln(n) \cdot \frac{n^4}{\epsilon^2}\right)$ for K measures:
  - Distance computation: $\binom{K}{2} \cdot \mathcal{O}(n^2)$
- LPGW embedding provide a noise-rubost representation for the data

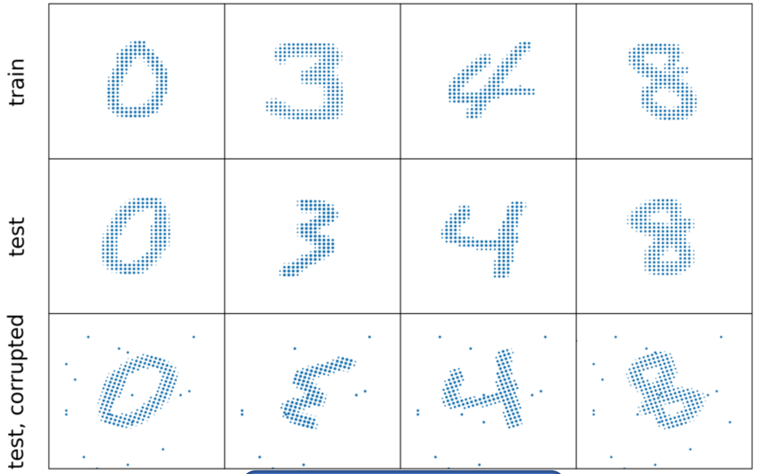


Data reconstruction by embeddings

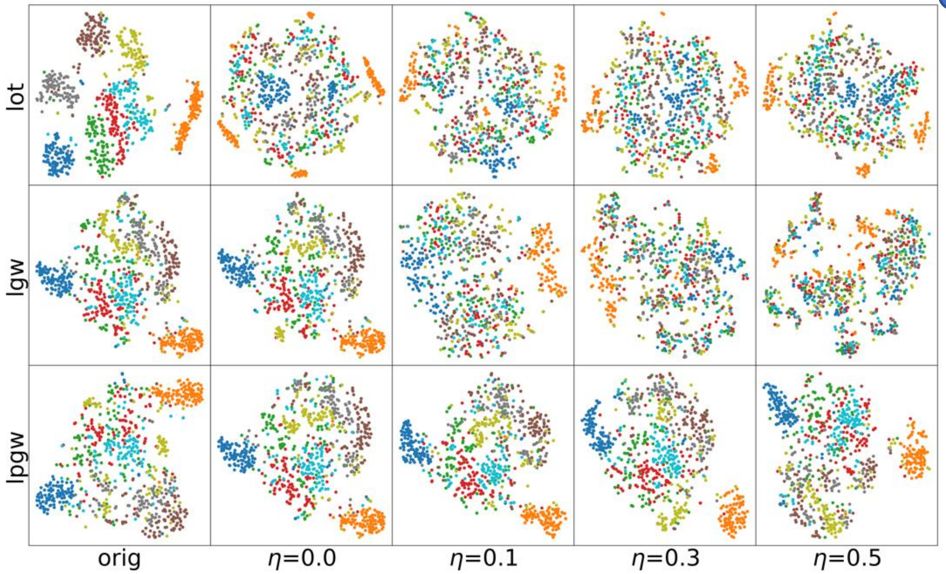


Accuracy analysis

Linear Partial Gromov Wasserstein embedding (Bai et al, ICLR 2024)



Data



Classification

Interpolation

Thank you for your attention

Related References:

- Mémoli, F. (2011). Gromov–Wasserstein Distances and the Metric Approach to Object Matching. *Foundations of Computational Mathematics*, 11(4), 417–487. cs2022.sciencesconf.org+2bibsonomy.org+2media.adelaide.edu.au+2
- Beier, F., Beinert, R., & Steidl, G. (2022). On a Linear Gromov–Wasserstein Distance. *IEEE Transactions on Image Processing*, 31, 7292–7305. static.tu.berlin
- Bai, Y., Kothapalli, A., Du, H., Martin, R. D., & Kolouri, S. (2024). Linear Partial Gromov-Wasserstein Embedding. *International Conference on Learning Representations* (ICLR).
- Bai, Y., Martin, R. D., Du, H., Shahbazi, A., & Kolouri, S. (2024). Partial Gromov-Wasserstein Metric. *International Conference on Learning Representations* (ICLR).