Summer School Valparaiso

Wasserstein and MMD penalizations for Unbalanced Optimal Transport

Guillaume Mahey

Collaboration with: F. Tobar
Supervision: G. Gasso, L. Chapel

12/01/2024

Objective of the talk

- Recall of Unbalanced Optimal Transport
- Shortcomings of UOT
- Contribution: A new penalization on the marginals

Disclaimer

- Work in progress
- Not technical talk

Unbalanced Optimal Transport

OT: Recall

The Wasserstein distance: optimal displacement of mass

$$\mathrm{OT}(\mu_1, \mu_2) = \inf_{\pi \in \Pi(\mu_1,\mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 d\pi(x,y) \qquad (1)$$

$\Pi(\mu_1, \mu_2) \stackrel{\text{def}}{=} \{\pi \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)$ s.t. $\pi(\mathbb{R}^d \times A) = \mu_2(A)$ and $\pi(A \times \mathbb{R}^d) = \mu_1(A)$, $\forall A \subset \mathbb{R}^d\}$

- Hard (mass-preserving) hard constraint on the marginal $\Rightarrow |\mu_1| = |\mu_2|$
- When $\mu_1 = \sum_{i=1}^n a_i \delta_{x_i}$ and $\mu_2 = \sum_{j=1}^m b_j \delta_{y_j}$:

$$\mathrm{OT}(\mu_1, \mu_2) = \min_{P \in U(a,b)} \langle C, P \rangle \qquad (2)$$

with $C_{ij} = \|x_i - y_j\|_2^2$ and $U(a,b) \stackrel{\text{def}}{=} \{P \in \mathbb{R}_+^{n \times m}$ s.t. $P 1_m = a$ and $P^\top 1_n = b\}$

UOT: Definition

Relax the hard marginal constraint of OT by soft constraint [CPSV15]

$$\mathrm{UOT}(\mu_1, \mu_2) = \inf_{\pi \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 d\pi(x, y) + \lambda_1 \mathrm{KL}(\pi^1, \mu_1) + \lambda_2 \mathrm{KL}(\pi^2, \mu_2) \quad (3)$$

$\pi^1 = \pi(\cdot \times \mathbb{R}^d)$, $\pi^2 = \pi(\mathbb{R}^d \times \cdot)$, $\lambda_1, \lambda_2 > 0$

- Not necessarily $|\mu_1| = |\mu_2|$
- Generally: KL or TV or ℓ_2 penalization
- π^i same support μ_i
- When $|\mu_1| = |\mu_2|$, UOT $\xrightarrow{\lambda \to \infty}$ OT

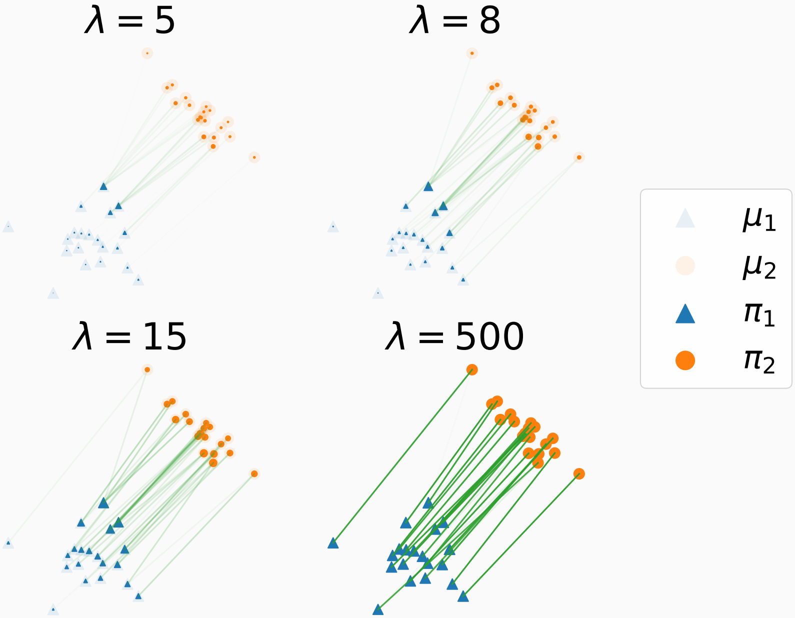
UOT: Definition

Can be rewritten as

$$\mathrm{UOT}(\mu_1, \mu_2) = \inf_{\pi \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \mathrm{OT}(\pi^1, \pi^2) + \lambda \left(\mathrm{KL}(\pi^1, \mu_1) + \mathrm{KL}(\pi^2, \mu_2) \right) \tag{4}$$

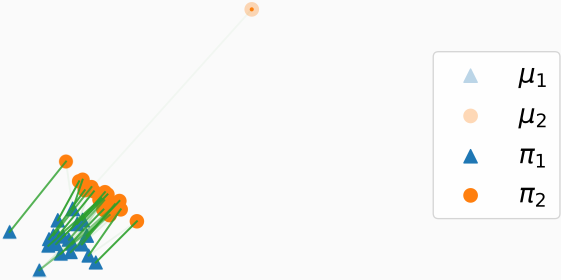
- π^1, π^2 proxies for μ_1, μ_2
- Atoms weights (but not locations) can change from π^i to μ_i
- Masses of $|\mu_i|$ can be informative

UOT: First Example



UOT: Outliers

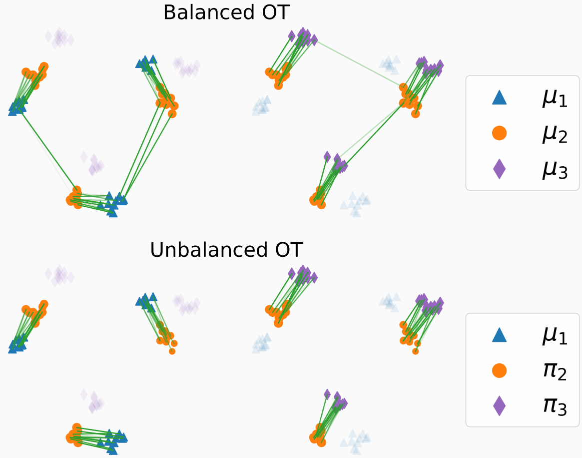
Interesting case: μ_2 is *infected* by outliers



- UOT seems to be robust to outliers

UOT: Birth and Death

Find a trajectory when the size of clusters can vary



- Can deal with birth and death of population [SST+19]

Drawbacks of UOT

UOT

Current penalizations compare π^i with μ_i vertically

$$\mathrm{KL}(\pi^1, \mu_1) = \sum_{i=0}^{n} a'_i \log(\frac{a'_i}{a_i}) - a'_i + a_i, \tag{5}$$

$\mu_1 = \sum a_i \delta_{x_i}$ and $\pi^1 = \sum a'_i \delta_{x_i}$

- No geometric information of support
- Change of weights driven by $\|x_i - y_j\|_2^2$
- π^i can be misleading as proxy of μ_i

Some Toy Example - Outliers

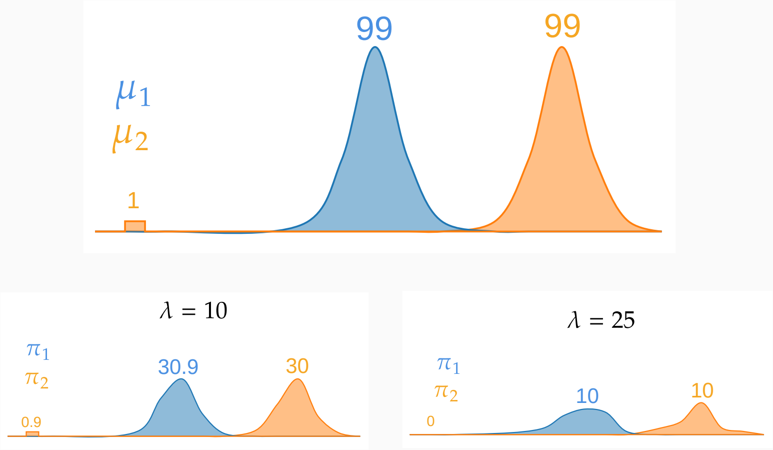


Figure: Source distribution μ_i (up) and proxies π_i for KL penalization (bot)

Some Toy Example - Birth and Death

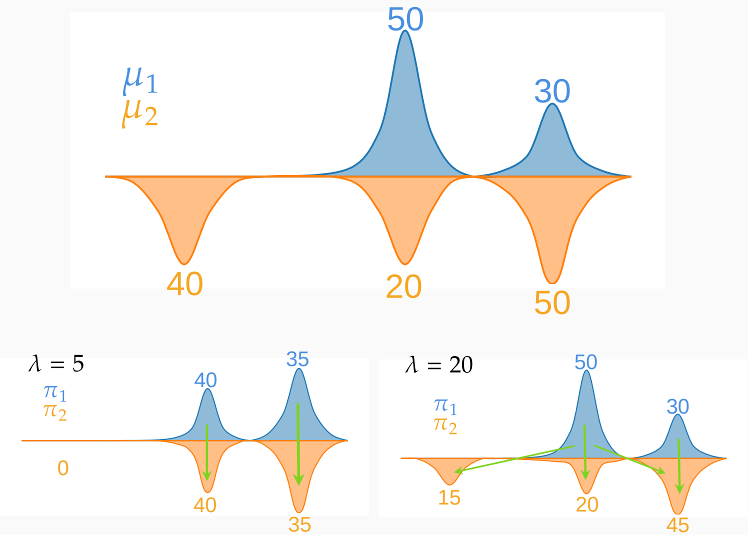


Figure: Source distribution μ_i (up) and proxies π_i for KL penalization (bot)

A New Penalization

Geometric Penalization

Penalization should consider the geometry of the marginals' support

- OT(π^i, μ_i)
- MMD(π^i, μ_i)

where

$$\begin{aligned} \text{MMD}(\mu_1, \pi^1) = & \int_{\mathbb{R}^d \times \mathbb{R}^d} k(x, x') d\mu_1(x) d\mu_1(x') \\ & + \int_{\mathbb{R}^d \times \mathbb{R}^d} k(x, x') d\pi^1(x) d\pi^1(x') \\ & - 2 \int_{\mathbb{R}^d \times \mathbb{R}^d} k(x, x') d\mu_1(x) d\pi^1(x) \end{aligned} \tag{6}$$

- Here Gaussian kernel $k(x, x') = \exp\left(\frac{-\|x-y\|_2^2}{2\sigma^2}\right)$, $\sigma > 0$

Geometric Penalization - OT

$$\mathrm{UOT}(\mu_1, \mu_2) = \inf_{\pi \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 d\pi(x, y) + \lambda_1 \mathrm{OT}(\pi^1, \mu_1) + \lambda_2 \mathrm{OT}(\pi^2, \mu_2) \quad (7)$$

- $|\pi^i| = |\mu_i| \Rightarrow |\mu_1| = |\mu_2|$
- Conserve overall mass: not death/birth but rather *redistribution of mass*
- Avoid cluster collapse

Geometric Penalization - MMD

$$\text{UOT}(\mu_1, \mu_2) = \inf_{\pi \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|_2^2 d\pi(x, y) + \lambda_1 \text{MMD}(\pi^1, \mu_1) + \lambda_2 \text{MMD}(\pi^2, \mu_2) \quad (8)$$

- Not necessarily $|\mu_1| = |\mu_2|$
- Lot of close atoms $\Rightarrow$ expensive to discard
- Few close atom $\Rightarrow$ easy to discard

Toy Examples

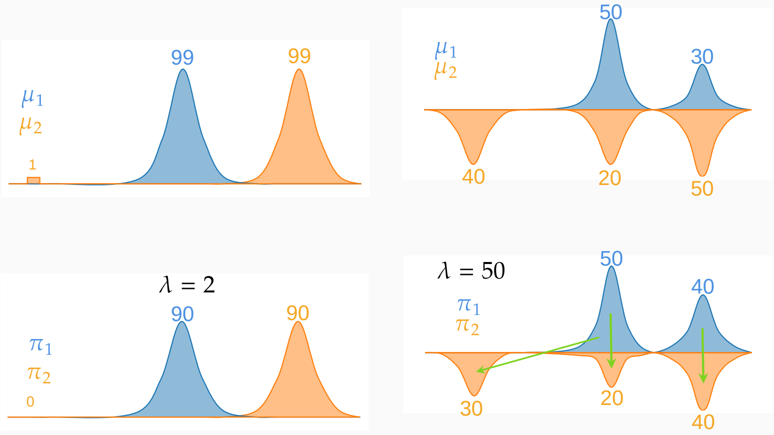


Figure: Source distribution μ_i (up) and proxies π_i for MMD penalization (bot) for Outliers (left) and Birth death (right)

Optimization

- UOT_D is a convex minimization problem
- Sinkhorn iterates [CPSV18]:

$$u^{2l} = \mathrm{prox}^{\epsilon \mathrm{KL}}_{\lambda_1 D}(Gv^{2l-1}) \oslash Gv^{2l-1} \ , \ v^{2l+1} = \mathrm{prox}^{\epsilon \mathrm{KL}}_{\lambda_2 D}(G^t u^{2l}) \oslash G^\top u^{2l}$$

where $G = e^{-\|x-y\|_2^2/\epsilon}$ Gibbs Kernel and:

$$\mathrm{prox}^{\epsilon \mathrm{KL}}_{\lambda D}(z) = \mathrm{argmin}_{u \in \mathbb{R}^n} \epsilon \mathrm{KL}(u|z) + \lambda D(u|a) \quad , \forall z \in \mathbb{R}^n$$

- OT and MMD more expensive than KL

Conclusion

Take home message:

- Proxies π^i must not be to far from μ_i
- Geometries of μ_i can be important
- MMD and OT penalization can be computed via Sinkhorn

Future works:

- Consider other penalizations:

$$\text{NLL}(\pi^i, \mu_i) \stackrel{\text{def}}{=} \int -\log\left(\int e^{\frac{-\|x-y\|^2}{2\sigma^2}} d\pi^i(x)\right) d\mu_i(y) \quad (9)$$

- Relax the fixed support constraint

- Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard.
Unbalanced optimal transport: geometry and kantorovich formulation.
2015.
- Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard.
Scaling algorithms for unbalanced optimal transport problems.
Mathematics of Computation, 87(314):2563–2609, 2018.
- Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al.

Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming.
Cell, 176(4):928–943, 2019.