# Sliced Gromov-Wasserstein

Titouan Vayer[1], Rémi Flamary[2], Romain Tavenard[3], Laetitia Chapel[1], Nicolas Courty[1]

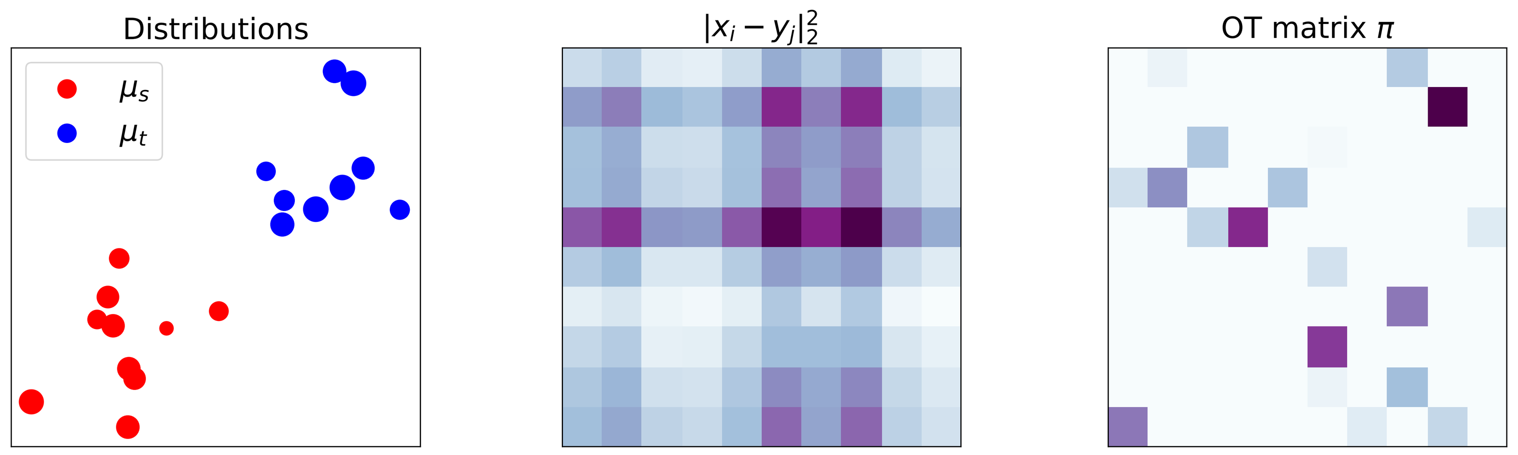
[1]Université de Bretagne Sud, IRISA, UMR 6074, CNRS, [2]Université Côte d'Azur Lagrange, UMR 7293 , CNRS, OCA [3]Université de Rennes LETG, CNRS

## Objective: Propose a scalable sliced version of the Gromov-Wasserstein distance

## Wasserstein distance (W)

Let $\mu_s = \sum_{i=1}^{n_s} a_i \delta_{x_i}$ and $\mu_t = \sum_{j=1}^{n_t} b_j \delta_{y_j}$ two discrete probability measures with $x_i, y_j \in \mathbb{R}^p$. Let $\Pi(a,b) = \{\boldsymbol{\pi} \in (\mathbb{R}^+)^{n_s \times n_t} | \boldsymbol{\pi} \mathbf{1}_{n_t} = a, \boldsymbol{\pi}^T \mathbf{1}_{n_s} = b\}$ all couplings of $a = (a_i), b = (b_j)$

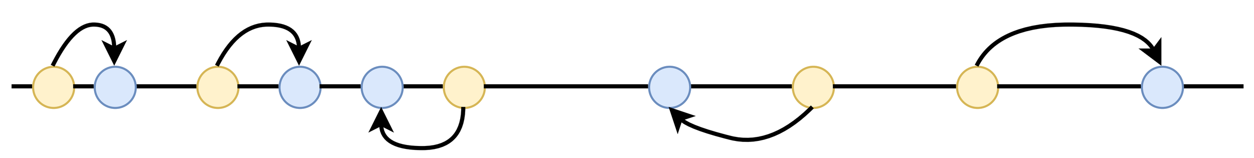
$$W_2^2(\mu_s, \mu_t) = \min_{\boldsymbol{\pi} \in \Pi(a,b)} \sum_{i,j} \pi_{i,j} \|x_i - y_j\|_2^2$$




Linear program: $O(n^3 \log(n))$. With entropic regularization near-linear time approximation [1].

## Sliced Wasserstein (SW)

When $\mu_s, \mu_t \in \mathcal{P}(\mathbb{R})$ are 1D probability measures with uniform weights: very simple algorithm to compute the W transport in $O(n \log n)$ by sorting both $x_i$ and $y_j$



$\mathbf{S}^{p-1}$ the $p$-dimensional hypersphere. $P_\theta(x) = \langle x, \theta \rangle$ the projection on $\theta$. SW is defined as [2]:

$$SW_2^2(\mu_s, \mu_t) = \mathop{\mathbb{E}}_{\theta \sim \text{unif}(\mathbf{S}^{p-1})} [W_2^2(P_\theta \# \mu_s, P_\theta \# \mu_t)]$$

where $P_\theta \# \mu_s = \frac{1}{n_s} \sum_{i=1}^{n_s} \delta_{\langle x_i, \theta \rangle} \in \mathcal{P}(\mathbb{R})$ is the mesure "projected" on $\theta$. Monte-Carlo estimation: $O(Ln \log(n))$ with $L$ the number of projections.

## References

[1] J. Altschuler, J. Weed, and P. Rigollet. Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In NeurIPS, 2017.

[2] N. Bonneel and D. Coeurjolly. Spot: Sliced partial optimal transport. ACM Trans. Graph., 2019.

[3] C. Bunne, D. Alvarez-Melis, A. Krause, and S. Jegelka. Learning generative models across incomparable spaces. In ICML, 2019.

[4] F. Memoli. Gromov wasserstein distances and the metric approach to object matching. Foundations of Computational Mathematics, 2011.

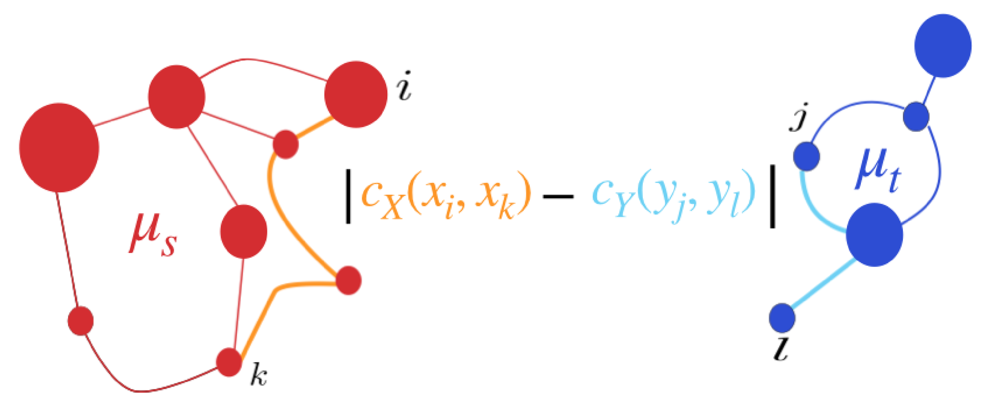
[5] G. Peyré, M. Cuturi, and J. Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In ICML, 2016.

[6] T. Vayer, L. Chapel, R. Flamary, R. Tavenard, and N. Courty. Optimal Transport for structured data with application on graphs. In ICML, 2019.

## Gromov-Wasserstein distance (GW)

What if supports of $\mu_s$, $\mu_t$ are not in the same metric space ? *e.g* $x_i \in \mathbb{R}^p, y_j \in \mathbb{R}^q$. $(GW)$ distance [4] is defined as:

$$GW_2^2(c_X, c_Y, \mu_s, \mu_t) = \min_{\pi \in \Pi(a,b)} \sum_{i,j,k,l} \left| c_X(x_i, x_k) - c_Y(y_j, y_l) \right|^2 \pi_{i,j} \pi_{k,l}.$$




Non-convex Quadratic Program (QP) = notoriously hard. Conditional Gradient [6]: $O(kn^3)$, with entropic regularization [5] nearly $O(n^2)$ but final cost $O(n^3)$.

**Is there a way to define a sliced version of $GW$ ?**

## GW in 1D

**Theorem 1** (Solving a QAP in 1D). *For real numbers $x_1 \leq ... \leq x_n$ and $y_1 \leq ... \leq y_n$,*

$$\min_{\sigma \in S_n} \sum_{i,j} -(x_i - x_j)^2 (y_{\sigma(i)} - y_{\sigma(j)})^2 \quad (1)$$

*is achieved either by the identity permutation $\sigma(i) = i$ or the anti-identity permutation $\sigma(i) = n + 1 - i$.*

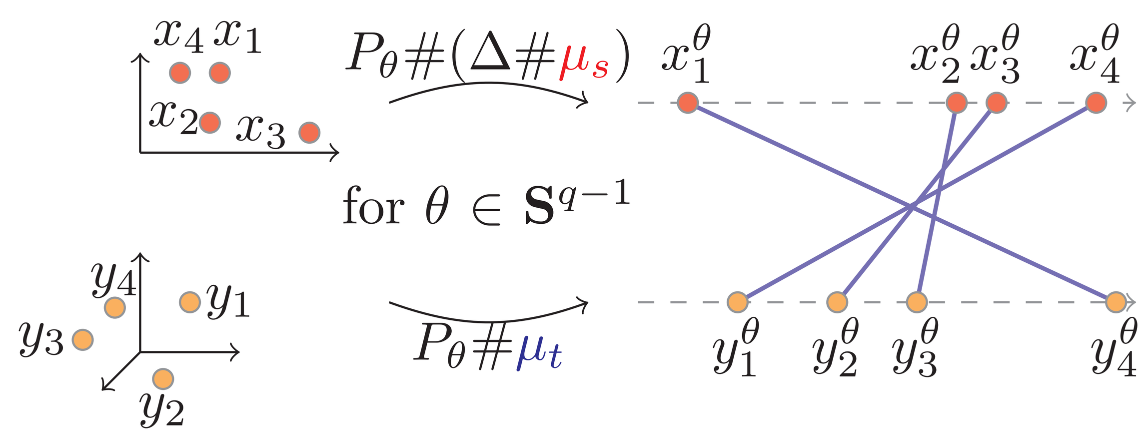
**Theorem 2** (Closed form for GW between 1D discrete measures). *For $\mu = \frac{1}{n} \sum_{i=1}^n \delta_{x_i} \in \mathcal{P}(\mathbb{R})$ and $\nu = \frac{1}{n} \sum_{i=1}^n \delta_{y_j} \in \mathcal{P}(\mathbb{R})$ the GW distance can be computed in $O(n \log(n))$ using simple sorts with $c_X = c_Y = |.|^2$.*

## Sliced Gromov-Wasserstein (SGW)

If $x_i \in \mathbb{R}^p, y_j \in \mathbb{R}^q, p \leq q, a_i = b_j = \frac{1}{n}$. For a linear map $\Delta \in \mathbb{R}^{q \times p}$ we define (SGW) as follows:

$$SGW_\Delta(\mu_s, \mu_t) = \mathop{\mathbb{E}}_{\theta \sim \text{unif}(\mathbf{S}^{p-1})} [GW_2^2(\,|.|^2, P_\theta \# {\mu_s}^\Delta, P_\theta \# \mu_t)] \quad (2)$$

where ${\mu_s}^\Delta = \Delta \# \mu_s \in \mathcal{P}(\mathbb{R}^q)$. Can be computed in $O(Ln \log(n))$ as SW.



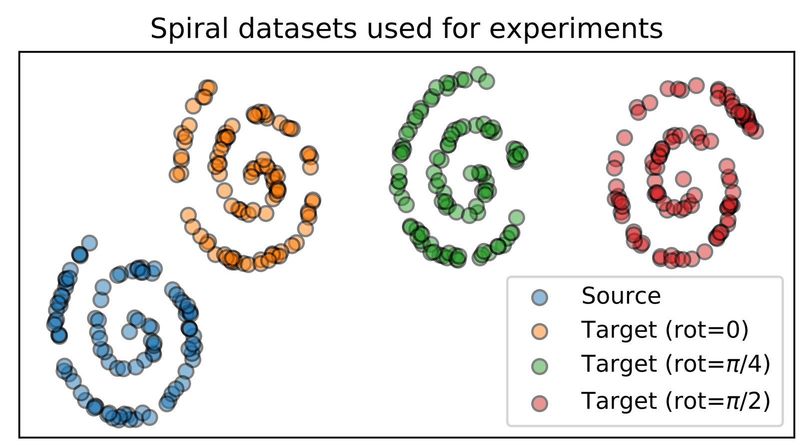

## Properties

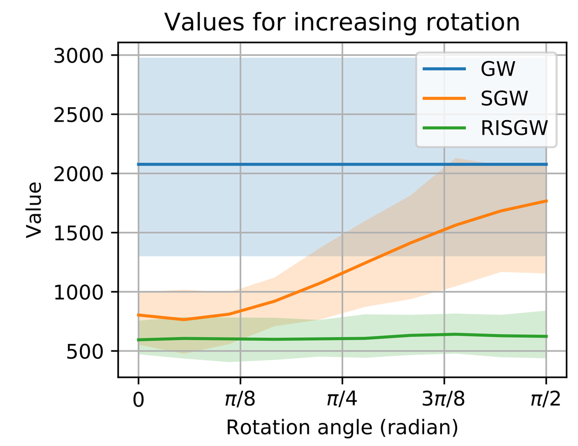
We also define Rotation Invariant SGW ($RISGW$):

$$RISGW(\mu_s, \mu_t) = \min_{\Delta \in \mathbb{V}_p(\mathbb{R}^q)} SGW_\Delta(\mu_s, \mu_t) \quad (3)$$

We propose to minimize $\Delta$ in the Stiefel Manifold.

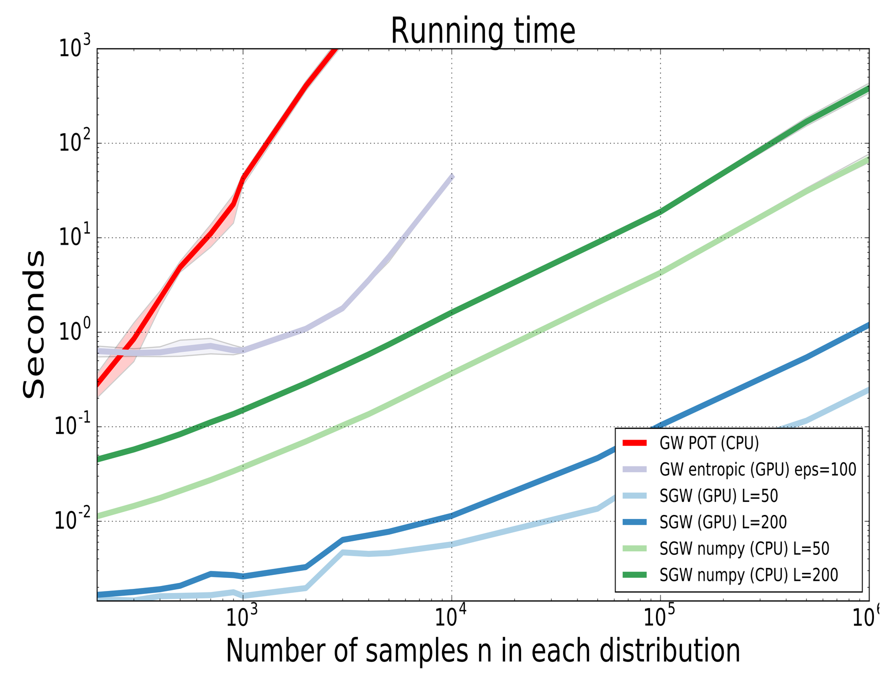
$SGW$ and $RISGW$ holds various properties of $GW$: they are pseudo distances, translation invariant, $RISGW$ is rotational invariant and if $RISGW$ and $SGW$ vanish then $\mu_s$ and $\mu_t$ are isomorphic.

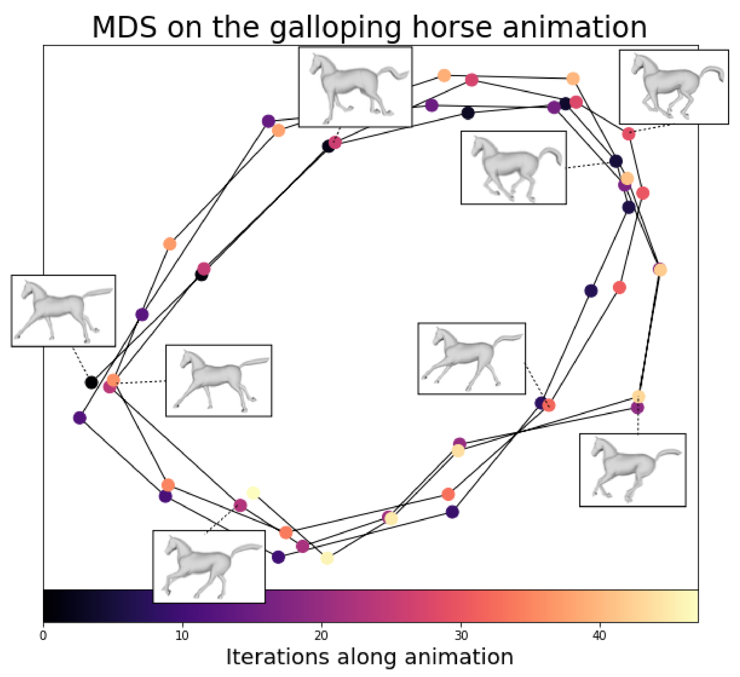



## Experiments

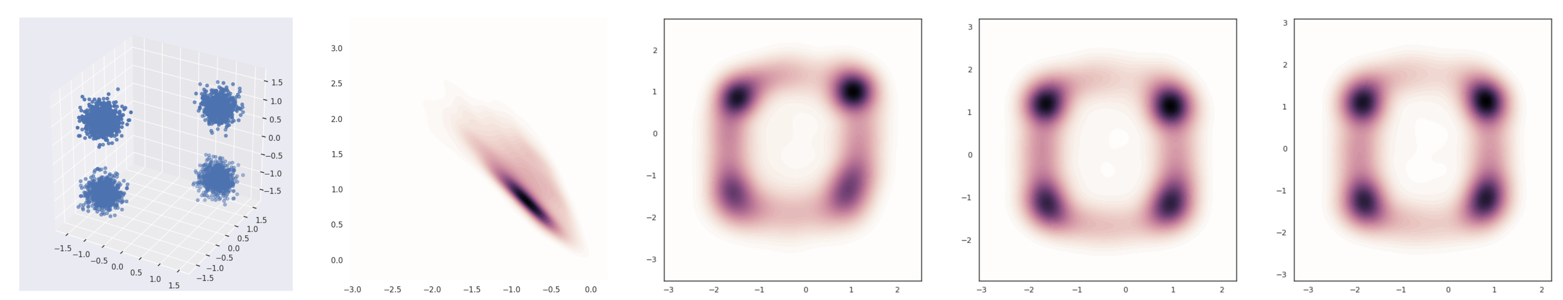
**Runtimes & meshes** **SGW between 1M points in 1 s**




**Generative modeling across incomparable space [3]**

$$G^* = \arg\min SGW_2^2(c_X, c_{G(Z)}, \mu, \nu_G),$$



Code available `https://github.com/tvayer/SGW` and in the Python Optimal Transport toolbox (POT) `https://github.com/rflamary/POT`