An in depth look at the Procrustes-Wasserstein distance: properties and barycenters

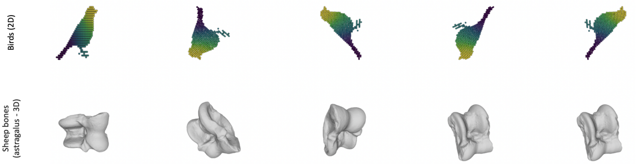
Davide Adamo[1,2]
joint work with M. Corneli[1,2], M. Vuillien[1] and E. Vila[3]

[1] Université Côte d'Azur, UMR 7264 CEPAM, CNRS, Nice, France
[2] Université Côte d'Azur, Inria, CNRS, Laboratoire J.A.Dieudonné, Maasai team, Nice, France
[3] Université Lumière Lyon II, UMR 5133 Archéorient CNRS, Lyon, France

ORCID: 0009-0004-4994-6427

Motivation

- Problem: distributions in different poses (2D/3D)



- We consider **rotations** & **reflections**
- " Can we define an isometry invariant distance ? "
- " Can we compute a mean shape ? "

Idea: yes using OT → **Procrustes-Wasserstein***

*Grave et al. *Unsupervised alignment of embeddings with wasserstein procrustes*, AISTATS 2019

Procrustes-Wasserstein distance

$$PW_2^2(\mu, \overline{\mu}) := \min_{P \in \mathcal{O}(d)} \min_{T \in \mathcal{U}(\mu, \overline{\mu})} \mathbb{E}_{X,Y \sim T} \left[||X - PY||_2^2 \right]$$

where P is an **orthogonal** matrix in $\mathbb{R}^{d \times d}$ and T is any **joint probability mass function** with support on $\mathcal{X} \times \mathcal{Y}$ and marginals h and $\overline{h}$, respectively.

Main question: is PW_2 an OT distance?

Lemma

- $PW_2(\mu, \overline{\mu}) \leq W_2(\mu, \overline{\mu})$, since $I_d \in \mathcal{O}(d)$
- it holds that

$$PW_2(\mu_X, \mu_Y) = W_2(\mu_X, \mu_{P^* Y})$$

Theorem

PW_2 is a distance on $\mathcal{M}_d / \sim$ (e.g. the quotient space of discrete measures under rigid transformations)

Procrustes-Wasserstein barycenters

Given M point clouds/random variables $Y_1, \ldots, Y_M$ we look for their PW barycenter

$$\min_{\mu_X} \left(\frac{1}{M} \sum_{j=1}^{M} PW_2^2(\mu_X, \mu_{Y_j}) \right)$$

where the minimum means that we look for both an *histogram* h_X and *locations* D_X (number of points fixed by the user).

Three alternating steps

1. solve M *independent* problems $PW_2^2(\mu_X, \mu_{Y_j})$
2. update h_X identical to [Cuturi and Doucet 2014].
3. update D_X

$$D_X = D_X + \frac{1}{M} \left(\sum_{i=1}^{M} \Gamma_i D_{Y_i} P_i \right) \mathrm{diag}(h_X) \qquad \text{(Newton-Raphson)}$$

Barycenters comparison

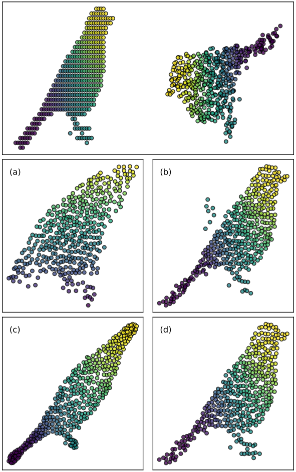
OT barycenters using

(a) Exact Free Wasserstein [Cuturi and Doucet 2014]

(b) Gromov-Wasserstein [Peyré et al. 2016] with MDS

(c) Gromov-Wasserstein with t-SNE

(d) PW (our)

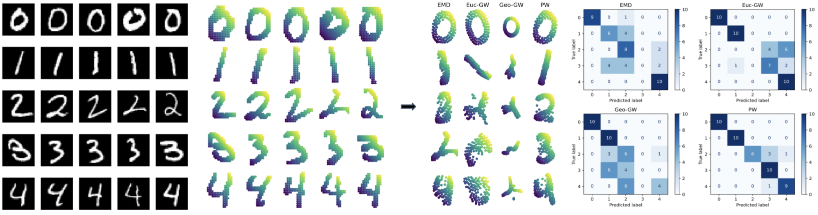


Hidden benefits: k-means with OT metrics

Tested methods

- Earth mover distance
- Euclidean GW
- Geodesic GW
- PW

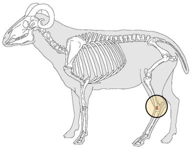
ALGOROTHM	TIME (S)	ARI	NMI
EMD	**9.18**	0.4069	0.5652
EUC-GW	675.19	0.5500	0.6815
GEO-GW	378.82	0.3797	0.5724
PW	130.11	**0.7669**	**0.8361**



MNIST images · MNIST point-clouds · centroids · confusion matrices

Zoo-archaeology

Interpolation between two bones of sheep (*Ovis aries*):
an **archaeological** and a **modern** one



$$\min_X \left(\eta PW_2^2(\mu_X, \mu_{Y_1}) + (1-\eta) PW_2^2(\mu_X, \mu_{Y_2}) \right)$$

with $\eta \in (0, 1)$