

Report for CSE 5339 2018 — (OTMLSA)  
Optimal Transport in Machine Learning and Shape Analysis

Algorithms to automatically quantify the geometric similarity of  
anatomical surfaces

Ying Yin

## 1 Motivation

To understand physical and biological phenomena (e.g. speciation, evolutionary adaption, etc.), one can quantify the similarity or dissimilarity of objects affected by the phenomena. One approach is to match landmarks between objects. In standard morphologists' practice, 10 to 100 points will be identified as landmarks. By comparing these landmarks, similarity and dissimilarity between patterns of shapes can be determined. However, the difficulty in acquiring personal knowledge of morphological evidence limits our understanding of the evolutionary significance of morphological diversity. The purpose of this paper is to develop an automatic tool to decide similarity or dissimilarity between objects, and hence, provides more insights on the phenomenon.

By tools in geometry, the authors map 3-D scans of body parts into a 2-D space, therefore, reduce the complexity of processing the objects.

## 2 Definitions

**Definition 1** Given two 2-dimensional smooth surfaces  $\mathcal{S}, \mathcal{S}'$ , a map  $\varphi : \mathcal{S} \rightarrow \mathcal{S}'$  is conformal if for any  $s \in \mathcal{S}$  and for any curves  $\Gamma_1, \Gamma_2$  on  $\mathcal{S}$  that intersects at  $s$ , the angle given by the tangent lines  $\ell_1, \ell_2$  is the same as the angle given by tangent lines  $\ell'_1, \ell'_2$  of curves  $\varphi(\Gamma_1), \varphi(\Gamma_2)$  on  $\mathcal{S}'$ .

**Definition 2** A disk-preserving Möbius transformation is a conformal bijective automorphism on  $D^2$ .

**Remark 1** If  $\varphi$  is conformal, then  $\varphi^{-1}$  is also conformal.

**Remark 2** Suppose the surfaces  $\mathcal{S}, \mathcal{S}'$  are smooth and has boundaries. By Riemann's uniformization theorem,  $\mathcal{S}, \mathcal{S}'$  can be conformally mapped to 2-disk  $D^2$ . Since both conformal map and Möbius transformation preserve angles, for any conformal maps  $\varphi : \mathcal{S} \rightarrow D^2, \varphi' : \mathcal{S}' \rightarrow D^2$  and a Möbius transformation  $m$ , the composition  $\varphi'^{-1} \circ m \circ \varphi$  is a conformal map between  $\mathcal{S}, \mathcal{S}'$ .

**Remark 3** A disk-preserving Möbius transformation has a closed-form formula  $m(z) = e^{i\theta} \frac{z-\alpha}{1-\bar{\alpha}z}$ , where  $\theta \in [0, 2\pi), |\alpha| < 1$

**Definition 3** A Riemannian metric on  $M$  is a family of positive definite inner products  $g_p : T_p M \times T_p M \rightarrow \mathbb{R}$  for  $p \in M$  such that for all differentiable vector fields  $X, Y$  on  $M$ ,  $p \mapsto g_p(X(p), Y(p))$ .

**Definition 4** Two Riemannian metrics  $g$  and  $h$  on a smooth manifold  $M$  are called conformally equivalent if  $g = fh$  for some positive function  $f$  on  $M$ . The function  $f$  is called the conformal factor.

**Remark 4** Conformal factor measures area distortion of a conformal map. Moreover, let  $g$  be an Riemannian metric. A diffeomorphism  $f : M \rightarrow N$  between two Riemannian manifolds is an isometry if for all  $p \in M$  and for all  $u, v \in T_p M$ ,  $g_p(u, v) = g_{f(p)}(df_p(u), df_p(v))$ .

**Definition 5**

$$d\eta(x, y) = [1 - (x^2 + y^2)]^{-2} dx dy$$

is the hyperbolic measure.

**Remark 5** Then given a conformal factor  $f(x, y)$ , we let  $\mathbf{f}(x, y) = [1 - (x^2 + y^2)] f(x, y)$ .

**Definition 6** Let  $\mu$  be a probability measure, and  $\tau$  be a differentiable bijection from  $D^2$  to itself, the mass distribution  $\mu' = \tau_*\mu$  defined by  $\mu(u) = \mu'(\tau(u))J_\tau(u)$  where  $J_\tau$  is the Jacobian of  $\tau$  is the transportation (or push-forward) of  $\mu$  by  $\tau$ .

**Remark 6**  $\tau_*\mu = \mu \circ \tau^{-1}$ .

**Remark 7** Note that for any (well-behaved) function  $F$  on  $D^2$ ,  $\int_{D^2} F(u)\mu'(u)du = \int_{D^2} F(\tau(u))\mu(u)du$ .

**Definition 7** The total transport effort  $\varepsilon_\tau = \int_{D^2} d(u, \tau(u))\mu(u)du$  where  $d(u, v)$  is the distance between  $u, v$  in  $D^2$ .

By infimizing  $\varepsilon_\tau$  over all measurable bijections  $\tau$  from  $D^2$  to itself, we solve the Monge problem. Alternatively, since the bijections are hard to search, consider the Kantorovitch problem, i.e.

**Definition 8** for all continuous functions  $F, G$  on  $D^2$ , let  $\pi$  be a coupling with marginals  $\mu, \nu$  satisfying that  $\int_{D^2 \times D^2} F(u)d\pi(u, v) = \int_{D^2} F(u)\mu(u)du$  and  $\int_{D^2 \times D^2} G(v)d\pi(u, v) = \int_{D^2} G(v)\nu(v)dv$ , we find the Wasserstein distance as

$$d_W = \inf_{\pi} \int_{D^2 \times D^2} d(u, v)d\pi(u, v)$$

## 3 Distances

### 3.1 Conformal Wasserstein distances (cW)

Instead of comparing two surfaces  $\mathcal{S}, \mathcal{S}'$ , one can compare two conformal factors  $\mathbf{f}, \mathbf{f}'$  obtained by conformally flattening  $\mathcal{S}, \mathcal{S}'$ . Let  $m$  be a disk-preserving Möbius transformation, then  $\mathbf{f}$  and  $m_*\mathbf{f} = \mathbf{f} \circ m^{-1}$  are both conformal factors for  $\mathcal{S}$ .

Then we define the conformal Wasserstein distance to be

$$\mathcal{D}_{cW}(\mathcal{S}, \mathcal{S}') = \inf_{m \in \mathcal{M}} \left[ \inf_{\pi \in \Pi(m_*\mathbf{f}, \mathbf{f}')} \int_{D^2 \times D^2} \tilde{d}(z, z')d\pi(z, z') \right]$$

, where  $\tilde{d}(\cdot, \cdot)$  is the hyperbolic distance on  $D^2$ .

**Remark 8**  $\mathcal{D}_{cW}$  is a metric.

**Remark 9** This is similar to the Wasserstein distance computed by infimizing over rigid motions between two shapes. Instead of rigid motions, we consider Möbius transformations since they conserve conformality.

**Remark 10** *However, computing  $\mathcal{D}_{cW}$  involves solving a Kantorovitch problem for every  $m$ .*

To reduce the complexity of computing cW, we consider the following distance.

### 3.2 Conformal Wasserstein neighborhood dissimilarity distance (cWn)

We quantify how *dissimilar* the "landscapes" are with a *measure of neighborhood dissimilarity*.

Let  $N(0, R)$  be a neighborhood at 0, i.e.,  $N(0, R) = \{z; |z| < R\}$ .

For any  $m \in \mathcal{M}$  s.t.  $z = m(0)$ ,  $N(z, R)$  is the image of  $N(0, R)$  under  $m$ .

Then we define the dissimilarity between  $\mathbf{f}$  at  $z$  and  $\mathbf{f}'$  at  $z'$  by

$$\mathbf{d}_{\mathbf{f}, \mathbf{f}'}^R(z, z') = \inf_{m \in \mathcal{M}, m(z)=z'} \left[ \int_{N(z, R)} |\mathbf{f}(w) - \mathbf{f}'(m(w))| d\eta(w) \right]$$

We defined the dissimilarity between  $\mathbf{f}$  at  $z$  and  $\mathbf{f}'$  at  $z'$  by

$$\mathbf{d}_{\mathbf{f}, \mathbf{f}'}^R(z, z') = \inf_{m \in \mathcal{M}, m(z)=z'} \left[ \int_{N(z, R)} |\mathbf{f}(w) - \mathbf{f}(m(w))| d\eta(w) \right]$$

The conformal Wasserstein neighborhood dissimilarity distance between  $\mathbf{f}$  and  $\mathbf{f}'$  is

$$\mathcal{D}_{cWn}^R(\mathcal{S}, \mathcal{S}') = \inf_{\pi \in \Pi(\mathbf{f}, \mathbf{f}')} \int_{D^2 \times D^2} \mathbf{d}_{\mathbf{f}, \mathbf{f}'}^R(z, z') d\pi(z, z')$$

Note that the cost of transport is given by  $|\mathbf{f}(w) - \mathbf{f}'(m(w))|$  which measures how two conformal maps distort area.

However, both cW and cWn are blind to isometric embedding of a surface in 3D. Therefore, we will introduce two extrinsic metric.

### 3.3 Standard Procruste distance

The standard Procrustes distance is between discrete sets of points  $\mathbf{X} = (X_n)_{n=1, \dots, N} \subset \mathcal{S}$  and  $\mathbf{Y} = (Y_n)_{n=1, \dots, N} \subset \mathcal{S}'$  by

$$d_p(\mathbf{X}, \mathbf{Y}) = \min_{R \text{ rigid motions}} \left[ \left( \sum_{n=1}^N |R(X_n) - Y_n|^2 \right)^{1/2} \right]$$

where  $|\cdot|$  is the standard Euclidean norm.

Often  $\mathbf{X}$  and  $\mathbf{Y}$  are sets of landmarks on two surfaces.

Again, one may find that the standard Procruste distance  $d_p$  is constructed in a similar fashion to the Wasserstein distance which was infimized over rigid motions. The cost is determined by how different the sets of landmarks are under isometry mapping.

**Remark 11**  $d_p(\mathbf{X}, \mathbf{Y})$  depends on choices of the sets of landmarks.

**Remark 12** Small number of  $N$  landmarks disregards a wealth of geometric data.

**Remark 13** Identifying and recording  $X_n, Y_n$  requires time and expertise.

Thus, we want to introduce a similar distance that does not require a pre-labelled set of landmarks.

### 3.4 Continuous Procruste distance (cP)

Instead, we consider a family of continuous maps  $a : \mathcal{S} \rightarrow \mathcal{S}'$  and use optimization to find the "best"  $a$ . We require  $a$  to be an *area-preserving* diffeomorphism as a relaxation of being isometry as in the Procrustes distance.

We denote the set of all area-preserving diffeomorphisms by  $\mathcal{A}(\mathcal{S}, \mathcal{S}')$ . And let

$$\mathbf{d}(\mathcal{S}, \mathcal{S}', a)^2 = \min_{R \text{ rigid motions}} \int_{\mathcal{S}} |R(x) - a(x)|^2 dA_{\mathcal{S}}$$

Then we define the *continuous Procrustes distance between  $\mathcal{S}$  and  $\mathcal{S}'$*  by

$$\mathbf{D}_p(\mathcal{S}, \mathcal{S}') = \inf_{a \in \mathcal{A}(\mathcal{S}, \mathcal{S}')} \mathbf{d}(\mathcal{S}, \mathcal{S}', a).$$

Since  $\mathbf{D}_p$  is infimizing over all area-preserving map, it is finding the an area-preserving diffeomorphism that is the closest to being an isometry. And since isometry means conformal and area-preserving, this "best" map  $a$  then is also close to being conformal.

**Remark 14** *There exists closed form formulas for minimizing over rigid motions. But it is hard to infimize over  $\mathcal{A}(\mathcal{S}, \mathcal{S}')$ .*

**Remark 15** *Thus it suffices to only explore a smaller space of maps obtained by small deformations of conformal maps.*

### 3.5 Computation of cP

We modify the search as follows:

Let  $m \in \mathcal{M}$ , then  $m$  is a conformal map. Let  $\varrho$  be a smooth map that roughly aligns high density peaks and  $\chi$  be a special deformation s.t.  $\chi \circ \varrho \circ m$  is area-preserving (up to approximation error).

For each choice of peaks  $p, p'$  in the conformal factors of  $\mathcal{S}, \mathcal{S}'$

1. runs through the 1-parameter family of  $m$  that maps  $p$  to  $p'$
2. constructs a map  $\varrho$  that aligns the other peaks, as best possible
3. compute  $\mathbf{d}(\mathcal{S}, \mathcal{S}', \varrho \circ m)$ .

Repeat for all choices of  $p, p'$ . Choose  $\varrho \circ m$  s.t. it minimizes  $\mathbf{d}$  and deform it to be area-preserving. Then the map  $a = \chi \circ \varrho \circ m$  is the approximate to correspondance map and  $\mathbf{d}(\mathcal{S}, \mathcal{S}', a)$  is the approximate to  $\mathbf{D}_p(\mathcal{S}, \mathcal{S}')$ .

## 4 Experiments

### 4.1 Setup

There are three independent data sets:

1. 116 second mandibular molars (teeth) of prosimian primates and non-primate close relatives
2. 57 proximal first metatarsals (bones behind big toe) of prosimian primates, New and Old World monkeys
3. 45 distal radii (bone in forearm) of apes and humans

For each shape, geometric morphometricians collected landmarks s.t. the points are biologically and evolutionarily meaningful. Then one can compute the Procrustes distances with the landmarks, producing Observer-Determined Landmarks Procrustes (ODLP) distances.

The running times for a pair of surfaces are:

1. cP:  $\sim 20$  sec.
2. cWn:  $\sim 5$  min.

## 4.2 Mantel correlation analysis

To quantitatively measure the differences between distance matrices, they first correlate the entries in the two square matrices, and then compute the fraction, among all possible relabelings of the rows/columns for one of them. The table is shown in Figure 1. From the table, one can see that for all data sets, cP shows a stronger correlation with ODLP. That is, cP is more similar to ODLP. Then we conclude that cP matches ODLP better than cWn.

**Table 1. Results of Mantel correlation analysis for cP and cWn versus ODLP distances**

Dataset	Obs. 1/cP		Obs. 2/cP		Obs. 1/cWn		Obs. 2/cWn	
	$r$	$P$	$r$	$P$	$r$	$P$	$r$	$P$
Teeth	0.690	0.0001	not applicable		0.373	0.0001	not applicable	
First metatarsal	0.640	0.0001	0.620	0.0001	0.365	0.0001	0.392	0.0001
Radius	0.240	0.0001	not applicable		0.075	0.166	not applicable	

Figure 1: For all data sets and all correlation coefficients available, the numbers associated with cP is much larger than the ones associated with cWn.

## 4.3 Distance matrices

Another way to compare the distance matrices is by observing the distance matrices. In Figure 2, the lower triangular region represents ODLP whereas the upper triangular is either cWn or cP. Looking at the zoomed-in squares at the bottom of Figure 2, one can see that the squares taken from cP looks more symmetric than the squares from cWn. Then again, we can conclude that cP matches with ODLP better.

## 4.4 Leave one out

Yet another way of comparing the performances of different distances is by comparing scores in taxonomic classification. The procedure is done in the following fashion: each specimen (treated as unknown) is assigned to the taxonomic group of its nearest neighbor in the remaining specimens in the data set (treated as training set). Figure 3 shows the results. Again, one may observe that in all cases, cP performs better than cWn and produces results that are very close to the ones by ODLP.

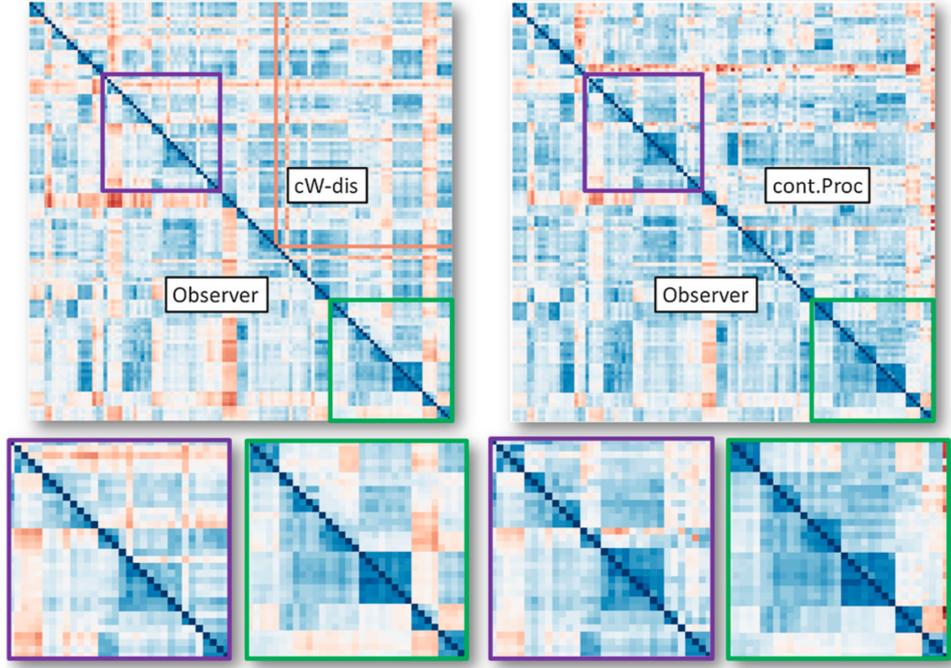


Figure 2: Note that in the zoomed-in parts, we see that the two squares on the right are more symmetric than the two squares on the left.

**Table 2. Success rates (percentage) of leave-one-out classification, based on the cP, cWn, and ODLP distances**

Dataset	Teeth					First metatarsal					Radius					
	No.	N	cP	Obs. 1	cWn	No.	N	Obs. 1	cP	Obs. 2	cWn	No.	N	cP	Obs. 1	cWn
Genera	24	99	90.9	91.9	68	13	59	76.3	79.9	88.1	50.8	4	45	84.4	77.8	68.9
Family	17	106	92.5	94.3	75.1	9	61	83.6	91.8	93.4	68.9			not applicable		
Above family	5	116	94.8	95.7	83.3	2	61	100	100	100	98.4			not applicable		

Figure 3: For each data set, No. denotes the number of categories in the specified class; N denotes number of specimen tested; column cP, Obs.1/2 and cWn shows the success rate of the leave on out experiment.

## 5 Conclusion

In terms of runtime and the three experiments shown above, we can conclude that the results by cP recovers the results produced by using Procrustes distance on morphologists-determined landmarks better than cWn. And in the leave one out experiment, both ODLP and cP produce a good rate in recovering taxonomic classification.

## References

- [1] C. Villani. Topics in Optimal Transportation. <http://bookstore.ams.org/gsm-58>.
- [2] <https://github.com/rflamary/POT>.
- [3] Remy Flamary's webpage. <http://remi.flamary.com>.
- [4] R. Flamary et al. "Optimal Transport for Domain Adaptation." <http://remi.flamary.com/biblio/courty2016optimal.pdf>. IEEE-TPAMI 2016.
- [5] Philip A. Knight. "The Sinkhorn-Knopp algorithm: convergence." <http://epubs.siam.org/doi/abs/10.1137/060659624>.
- [6] O. Bousquet et al. "From optimal transport to generative modeling: the VEGAN cookbook" <https://arxiv.org/abs/1705.07642>.
- [7] Marco Cuturi's webpage. <http://marcocuturi.net>.
- [8] M. Carriere et al. "Sliced Wasserstein Kernel for Persistence Diagrams". <http://proceedings.mlr.press/v70/carriere17a.html>.
- [9] M. Cuturi et al. "Stochastic Optimization for Large-scale Optimal Transport." <https://papers.nips.cc/paper/6566-stochastic-optimization-for-large-scale-optimal-transport>.
- [10] E. Bernton et al. "Inference in generative models using the Wasserstein distance." <https://arxiv.org/abs/1701.05146>.
- [11] N. Courty et al. "Learning Wasserstein Embeddings." <https://arxiv.org/pdf/1710.07457.pdf>. NIPS 2017.
- [12] Y. Chen et al. "Optimal Transport for Gaussian Mixture Models." [LearningWassersteinEmbeddings](https://arxiv.org/pdf/1705.08044v1.pdf)
- [13] Justin Solomon's webpage. <http://people.csail.mit.edu/jsolomon/>.
- [14] G. Peyre et al. "Gromov-Wasserstein Averaging of Kernel and Distance Matrices." <https://hal.archives-ouvertes.fr/hal-01322992/document>.
- [15] C. Brecheteau. "The DTM-signature for a geometric comparison of metric-measure spaces from samples". <https://arxiv.org/abs/1702.02838>
- [16] F. Memoli. "On the use of the Gromov-Hausdorff distance for Shape Comparison". <http://sites.fas.harvard.edu/~cs277/papers/gromov.pdf>.
- [17] Chazal et al. "Geometric Inference for Measures based on Distance Functions". <https://hal.inria.fr/inria-00383685/file/RR-6930v2.pdf>.
- [18] De Goes et al. "An Optimal Transport Approach to Robust Reconstruction and Simplification of 2D Shapes". <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8659.2011.02033.x/full>.
- [19] Solomon et al. "Entropic Metric Alignment for Correspondence Problems". <https://people.csail.mit.edu/jsolomon/assets/gw.pdf>.
- [20] Cuturi. "SINKHORN DISTANCES: LIGHTSPEED COMPUTATION OF OPTIMAL TRANSPORTATION DISTANCES". <https://arxiv.org/pdf/1306.0895.pdf>.

- [21] Garcia-Trillos et al. "ON THE RATE OF CONVERGENCE OF EMPIRICAL MEASURES IN 1-TRANSPORTATION DISTANCE". <https://arxiv.org/pdf/1407.1157.pdf>.
- [22] Di Marino et al. "THE ENTROPIC REGULARIZATION OF THE MONGE PROBLEM ON THE REAL LINE". <https://arxiv.org/pdf/1703.10457.pdf>.
- [23] <http://marcocuturi.net/dagstuhl.pdf>.