

# A Wasserstein-type distance for Gaussian Mixtures on Vector Bundles

Michael Wilson

Florida State University

*mwilson5@fsu.edu*

May 3, 2024

# Overview

## 1 Introduction

- Wasserstein-type Distance

## 2 Applications

- Punctured Sphere
- Kendall's Shape Space of Triangles
- Nanoparticle data

# Introduction

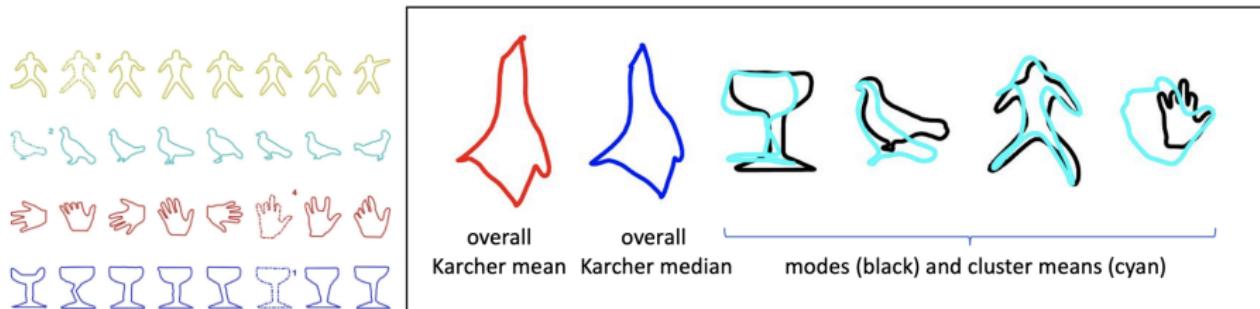


Figure: Karcher means of mixtures of shape aren't representative of data

**Goal:** Compare mixture distributions of shapes

**Approach:** Model shapes with Wrapped Gaussian (WG) mixtures of Preshape

- ① **Estimate Components:** K-modes Kernel Mixtures
- ② **Estimate Parameters:** Karcher Mean & Tangent Space Covariance
- ③ **Compare Distributions:** Wasserstein-type distance for Gaussian Mixtures

# Wasserstein Distance

- $(\mathcal{X}, d)$  a metric space.
- $\mathcal{P}_p(\mathcal{X}) = \{\mu : \int_{\mathcal{X}} d(x_0, x)^p d\mu(x) < \infty \ \forall x_0\}$

The Wasserstein Distance between  $\mu_0, \mu_1 \in \mathcal{P}_p$  is given by;

$$W_p^{\mathcal{X}}(\mu_0, \mu_1) := \left( \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathcal{X} \times \mathcal{Y}} d(x, y)^p d\gamma(x, y) \right)^{\frac{1}{p}}$$

- $d$ : distance/ transport cost
- $\gamma$ : coupling/ transport plan

# Distance Computation

- $|supp(\mu_k)| = n_k, k = 0, 1$
- $x_i \in supp(\mu_0), y_j \in supp(\mu_1)$

$$W_2^2(\mu_0, \mu_1) = \min_{T \in \Pi(\mu_0, \mu_1)} \sum_{i,j} d(x_i, y_j)^2 T_{ij}$$

which can be solved with linear programming

**We can directly calculate Wasserstein distances between finitely supported measures on metric spaces**

# Wasserstein Distance between Gaussians

- $\mathcal{G}(\mathbb{R}^d) := \{\mu = N(m, \Sigma) : m \in \mathbb{R}^d, \Sigma \in \text{Sym}_d^+\}$
- $\mu_0, \mu_1 \in \mathcal{G}(\mathbb{R}^d)$

$$W_2^2(\mu_0, \mu_1) = \|m_0 - m_1\|^2 + \text{tr}(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{\frac{1}{2}} \Sigma_1 \Sigma_0^{\frac{1}{2}})^{\frac{1}{2}})$$

and  $(\mathcal{G}(\mathbb{R}^d), W_2^2)$  is a metric space.

# Gaussian Mixtures

A measure  $\mu$  on  $\mathbb{R}^d$  is a Gaussian mixture measure if it can be written as

$$\mu = \sum_{k=1}^K w_k N(m_k, \Sigma_k) \text{ and } \sum_{k=1}^K w_k = 1, \quad w_k \geq 0, \quad \forall k, K < \infty$$

# Identifiability of Gaussian mixtures

**Identifiability:** There exists a bijection between space of measures and parameter space (For  $G(\mathbb{R}^d)$ , see Yakowitz & Spragins, 1968 [3])

- $\mu \in G(\mathcal{X})$
- $(m, \Sigma) \in \mathcal{X} \times Sym_d^+$

$$\sum_k w_k N(m_k, \Sigma_k) \iff \sum_k w_k \delta_{\mu_i^k}(\mu)$$

On spaces  $\mathcal{X}$  where finite Gaussian mixtures are identifiable, they can be uniquely represented as discrete measures on  $G(\mathcal{X})$

# A Distance for Gaussian mixtures

- $(\mathcal{G}(\mathbb{R}^d), W_2^2)$  is a metric space
- $\mu_i^k = N(m_i^k, \Sigma_i^k)$
- $\mu_i = \sum_k^K w_k \mu_i^k$

$$D(\mu_0, \mu_1) := \min_{T \in \Pi(\mu_0, \mu_1)} \sum_{k, \ell} W_2^2(\mu_0^k, \mu_1^\ell) T_{ij}$$

is a distance on Gaussian mixtures. (Chen et. al, 2018 [1])

# Wasserstein-type Distance between Gaussian Mixtures

It turns out that this distance can be seen as an optimal transport *on the domain of the distributions* (not just on the parameter space), where the set of admissible couplings are restricted to be Gaussian mixtures;

$$MW_2^2(\mu_0, \mu_1) = \inf_{\pi \in \Pi(\mu_0, \mu_1) \cap GMM} \int_{\mathcal{X} \times \mathcal{Y}} ||x - y||^2 d\pi(x, y)$$

(Delon & Desolneux, 2019 [2])

$$MW_2^2(\mu_0, \mu_1) = \min_{T \in \Pi(\mu_0, \mu_1)} \sum_{k, \ell} W_2^2(\mu_0^k, \mu_1^\ell) T_{ij}$$

# Trivial Vector Bundles

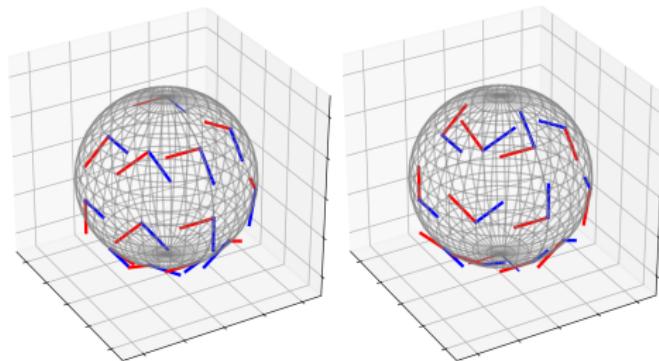


Figure: Illustrations of two trivializations of the punctured sphere.

Vector Bundle:  $\mathcal{E} \rightarrow \mathcal{M}$

Trivialization:  $\varphi : \mathcal{E} \rightarrow \mathcal{M} \times \mathbb{R}^d$

# Gaussians on Trivial Vector Bundles

A Borel measure on the vector bundle  $\mathcal{E}$  is called a Gaussian measure if it is a mean-zero Gaussian measure on the inner product space  $(\mathcal{E}_m, \langle \cdot, \cdot \rangle_m)$ , for some  $m \in \mathcal{M}$ .

The collection of Gaussian measures on  $\mathcal{M} \times \mathbb{R}^d$  is identifiable, and is denoted  $G(\mathcal{M} \times \mathbb{R}^d)$ .

# Wasserstein Distance between Gaussians on Trivial Vector Bundles

$$d_{M \times \mathbb{R}^d}((m_0, v_0), (m_1, v_1))^2 = d_M(m_0, m_1)^2 + \|v_0 - v_1\|^2$$

$$W_2^{M \times \mathbb{R}^d}(\eta_0, \eta_1)^2$$

$$= d_M(m_0, m_1)^2 + \text{tr}\left(\Sigma_0 + \Sigma_1 - 2\left(\Sigma_0^{\frac{1}{2}}\Phi_{m_0, m_1}^{-1}\Sigma_1\Phi_{m_0, m_1}\Sigma_0^{\frac{1}{2}}\right)^{\frac{1}{2}}\right)$$

where

$$\Phi_{m_0, m_1} = \varphi_{m_1}^{-1} \circ \varphi_{m_0} : \mathcal{E}_{m_0} \rightarrow \mathcal{E}_{m_1}$$

$\implies (G(M \times \mathbb{R}^d), W_2^{M \times \mathbb{R}^d})$  is a metric space

# Wasserstein-type Distance between Gaussian Mixtures on Trivial Vector Bundles

$$\mu_i = \sum_{k=1}^{K_i} w_i^k \eta_i^k, \quad \eta_i^k = N_{\mathcal{E}}(m_i^k, \Sigma_i^k), \quad i \in \{0, 1\}$$

can be expressed as

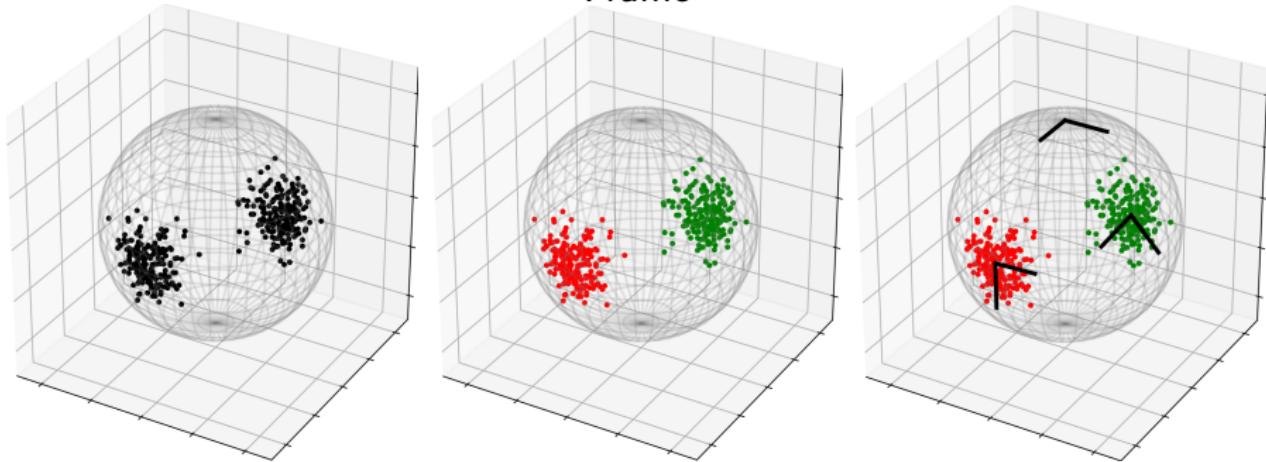
$$MW_2^\varphi(\mu_0, \mu_1)^2 = \min_{\omega \in \Pi(\mu_0, \mu_1)} \sum_{k, \ell} \omega_{k\ell} W_2^{\mathcal{M} \times \mathbb{R}^d}(\varphi \# \eta_0^k, \varphi \# \eta_1^\ell)^2 \quad (1)$$

# Applications

# Applications

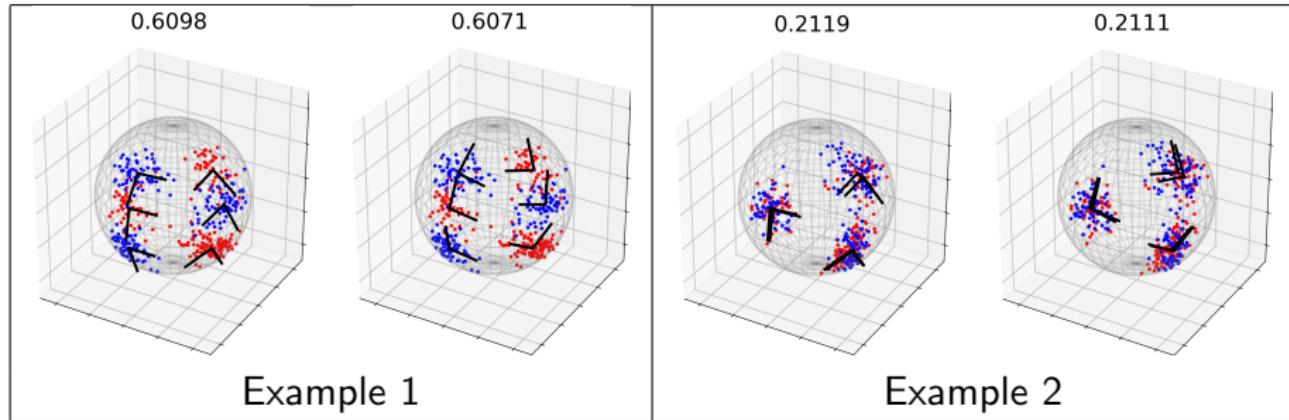
# Density Estimation

Identifying Gaussian Mixture Parameters with Respect to Orthogonal Frame



**Figure:** Steps for estimating Gaussian mixture parameters with respect to a moving frame on a punctured <sup>2</sup>. Left: Sphere with sample data; Middle: Sample data colored by K-means cluster assignment; Right: Orthogonal frame transported to Fréchet means of clusters.

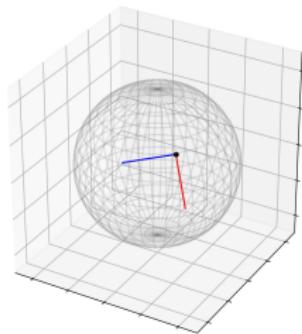
# Distance Calculation



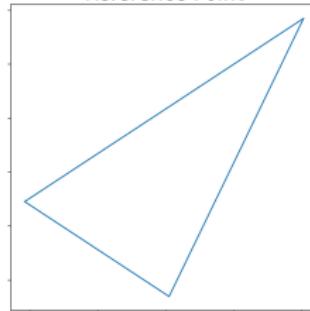
**Figure:** Examples of Wasserstein-type Distances between wrapped Gaussian mixtures on a punctured  $S^2$  computed using two different moving frames. Color denotes a WG mixture; plot titles show calculated distances. Example 1: Samples from two WG mixtures with different means, covariances, and weights. Example 2: Samples from two WG mixtures with same means and covariances, but different weights.

# Kendall's Shape Space of Triangles

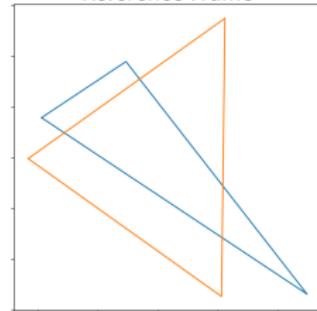
Reference Frame on 2-sphere



Reference Point

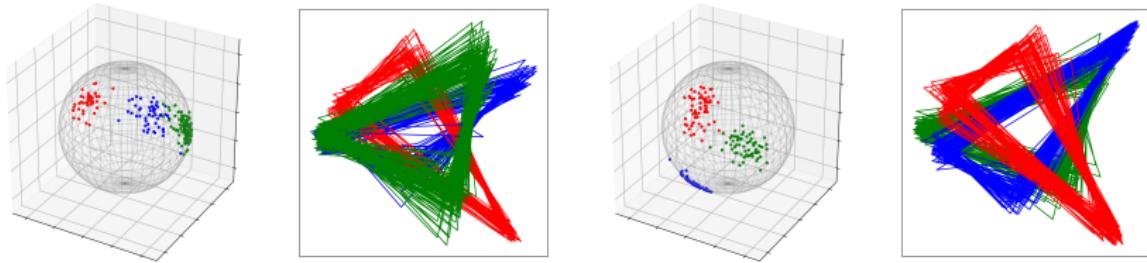


Reference Frame



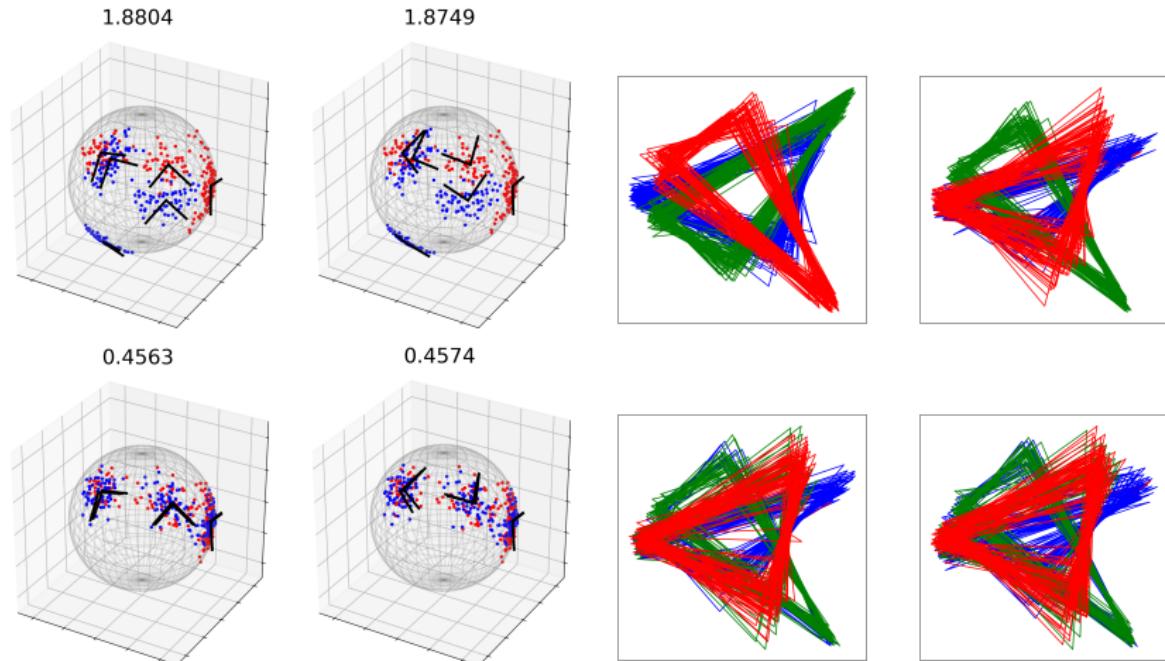
**Figure:** A reference point and a reference frame for Kendall shape space of 2D triangles. Left: Representation of moving frame under Hopf map on  $\mathbb{S}^2$ ; Middle: Triangular representation of reference point; Right: Triangular representations of tangent vectors in reference frame.

# Kendall's Shape Space of Traingles



**Figure:** Samples from two wrapped Gaussian mixtures defined on  $\mathcal{S}_T$ , colored by cluster assignment. The triangles in panel 2 (resp. 4) correspond to the points on the sphere in panel 1 (resp. 3).

# Kendall's Shape Space of Triangles



**Figure:** Comparison of Wasserstein-type Distances between Gaussian mixtures on Kendall's shape space of planar triangles

# Data

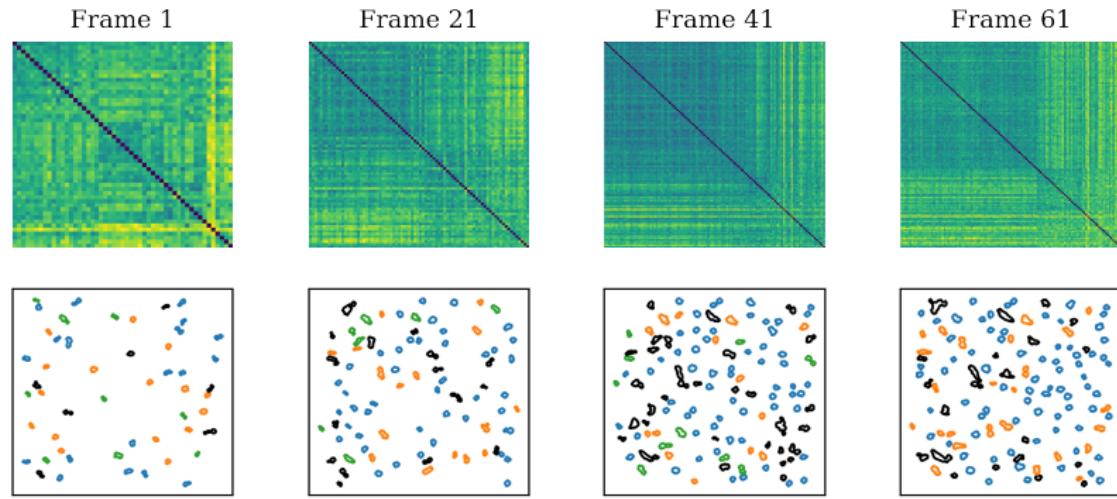
Figure: Recording of nanoparticle manufacturing process

Goal: Test for change in distribution of shapes over time

Approach:

- Represent each image as a Gaussian Mixture
- Calculate between frame Wasserstein-type distances
- E-divisive Change point detection

# K-Modes Kernel Mixture



Estimate Mixture Components: K-mode Kernel mixtures clustering returns index set  $\mathcal{I}$ ;

$$\{q_i, i = 1, \dots, N\} \rightarrow \mathcal{I}$$

# Get Pre-shape Representation

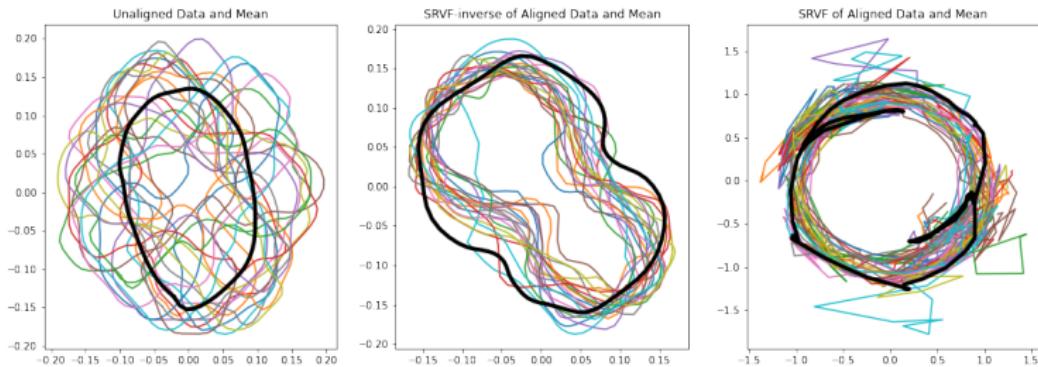


Figure:  $\beta$ 's, before and after alignment, and aligned SRVF's

Estimate Component Means: For each cluster, calculate shape mean and align curves

$$C_{k,t} = \{q_i : \mathcal{I}^t(q_i) = k\} \rightarrow (\mu_k, \{\tilde{q}_i\}) \quad (2)$$

# Get Coordinate Representation

Define an orthonormal functional basis  $B$ , project data into basis to get coordinates;

$$(\mu_k^*)_j = \int B_j(s) \mu_k(s) ds$$

$$(\tilde{q}_i^*)_j = \int B_j(s) \tilde{q}_i(s) ds$$

We also need coordinates for the basis (trivial, but a choice)

$$(p_{ref}^*)_j = \int B_j(s) B_1(s) ds$$

$$(F_{ref}^*)_{j,\ell} = \int B_j(s) B_\ell(s) ds$$

## Estimate Parameters for Gaussian

We estimate the Coordinate Representation for the kth Gaussian component for frame  $t$  as;

$$\nu_k = N(\mu_k^*, \Sigma_k^*)$$

with

$$\Sigma_k^* = \frac{1}{N - 1} X^t X$$

where

$$X_{:,i} = R_{(p_{ref} \rightarrow \mu_k^*)}(F_{ref}^*)^T \text{Exp}_{\mu_k^*}^{-1}(\tilde{q}_i^*) \quad (3)$$

# Wrapped Gaussian Mixture

We estimate the wrapped Gaussian mixture for image  $t$  as (note, we've added a subscript for frame number);

$$\nu_t = \sum_{k=1}^{K_t} w_k \nu_{t,k}$$

$$w_{t,k} = |\{q_{t,i} : \mathcal{I}(q_{t,i}) = k\}| / N_t$$

# Wasserstein-type Distance between wrapped Gaussian mixtures

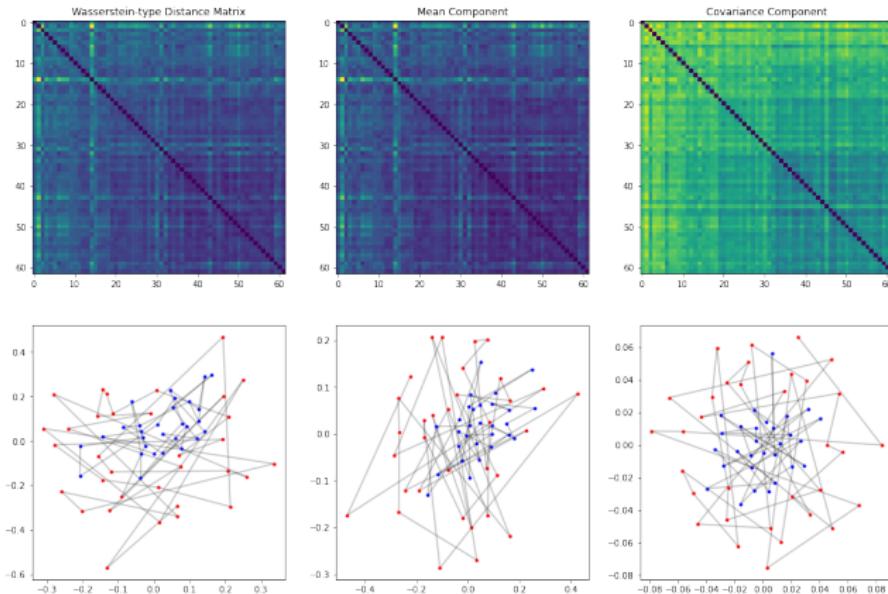
Given wrapped Gaussian mixture distributions for each image  $t = 1, \dots, T$ ,

$$\nu_t = \sum_{k=1}^{K_t} w_{t,k} N(\mu_{t,k}^*, \Sigma_{t,k}^*)$$

we calculate the Wasserstein-type distance between them using the formula;

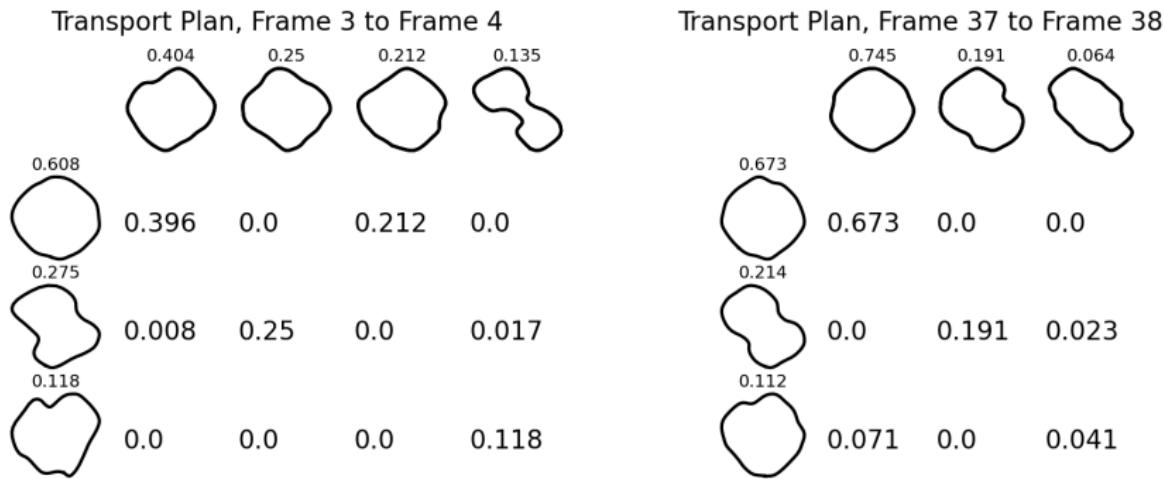
$$MW_2^2(\nu_t, \nu_s) = \inf_{\Pi \in \Gamma(w_t, w_s)} \sum_{k=1}^{K_t} \sum_{\ell=1}^{K_s} \Pi_{k,\ell} W_2^2(\nu_{t,k}, \nu_{s,\ell}) \quad (4)$$

# Results



**Figure:** Top: Wasserstein-type Distance matrix for process, with mean and covariance components; Bottom: MDS embedding of distance matrices into 2 dimensions

# Interpreting Transport Plans



**Figure:** Visual representation of inter-frame transport plans. Gaussian mixture component mean shapes are represented in the margins. Numbers above mean shapes correspond to estimated weight for that component; array of numbers in center of plot are the transport plans, they represent the amount of mass transported between the corresponding marginal components.

## References

-  [Yongxin Chen, Tryphon T Georgiou, and Allen Tannenbaum.  
Optimal transport for gaussian mixture models.  
IEEE Access, 7:6269–6278, 2018.](#)
-  [Julie Delon and Agnes Desolneux.  
A wasserstein-type distance in the space of gaussian mixture models,  
2019.](#)
-  [J Yakowitz Sidney and D Spragins John.  
On the identifiability of finite mixtures.  
The Annals of Mathematical Statistics, 39\(1\):209–214, 1968.](#)