Linear Optimal Transport for Gaussian Mixtures

Michael Wilson, Tom Needham, Anuj Srivastava

Florida State University

mwilson5@.fsu.edu

March 24, 2024

DTMRI Data

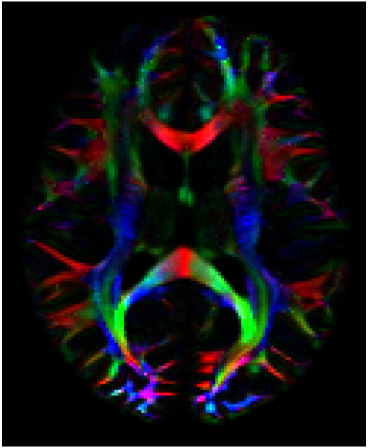


Figure: Slice from DTMRI image

Diffusion Tensor MRI

- DTMRI measures diffusion of water in tissue
- "Diffusion Tensor" is a covariance matrix

Optimal Transport

- Represent images as empirical probability measures on $\mathbb{R}^3 \times Sym_3^+$
- Statistical analysis with LOT

Wasserstein Distance

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

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

$$W_p^p(\mu_0, \mu_1) := \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\Omega^2} d(x, y)^p d\gamma(x, y)$$

- $\Pi(\mu_0, \mu_1)$: set of couplings ($\sim$joint distributions) of μ_0, μ_1
- d: distance/ ground metric
- γ: optimal coupling/ correspondence

Empirical Probability Measures

- $b_\ell \in \Sigma_{m_\ell} = \{u \in R_+^{m_\ell} : ||u||_1 = 1\}$
- $\mathcal{U}(b_0, b_1) = \{P \in \mathbb{R}^{m_0 \times m_1} : \sum_i P_{ij} = b_j, \sum_j P_{ij} = a_i\}$

For empirical probability measures $\mu_\ell = \sum_{j=1}^m b_j \delta_{x_j}, \ell = 1, 2$ supported on (Ω, d),

$$W_2^2(\mu_0, \mu_1) = \min_{\gamma \in \mathcal{U}(b_0, b_1)} \sum_{i,j} \gamma_{ij} d(x_i, y_j)^2$$

- We can directly calculate Wasserstein distances between empirical probability measures on metric spaces

Wasserstein Distance between Gaussians

- $\mathrm{G}(\mathbb{R}^d) := \{\mu = N(m, \Sigma) : m \in \mathbb{R}^d, \Sigma \in \mathrm{Sym}_d^+\}$
- $\mu_i = N(m_i, \Sigma_i)$

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

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

- RHS defines a distance on $R^d \times Sym_d^+$

Gaussian mixtures

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

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

Identifiablity: Gaussian mixtures on $\mathbb{R}^d$ can be uniquely represented as empirical probability measures on $\mathbb{R}^d \times Sym_d^+$.

An Optimal Transport Distance for Gaussian mixtures

- $\mu_i^k = N(m_i^k, \Sigma_i^k)$
- $\mu_i = \sum_k^K w_k \mu_i^k$

$$D(\mu_0, \mu_1) := \min_{\gamma \in \mathcal{U}(b_0, b_1)} \sum_{k,\ell} \gamma_{ij} W_2^2(\mu_0^k, \mu_1^\ell)$$

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

Wasserstein-type Distance between Gaussian Mixtures

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

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

(Delon & Desolneux, 2019 [2])

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

Data

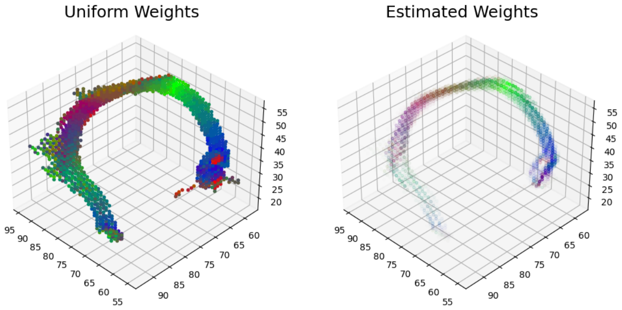
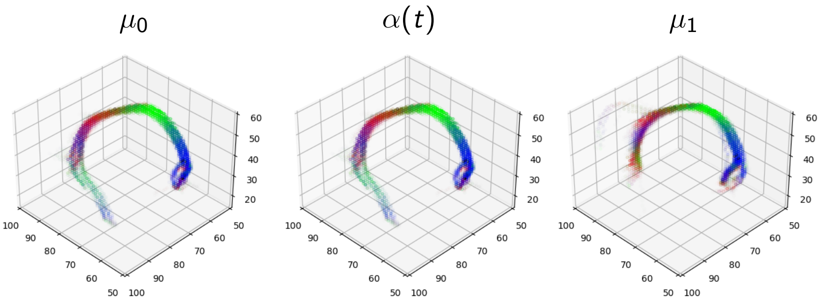


Figure: Left: Segmented Fornix L; Right: Fornix L with estimated weights

$$\mu = \sum_{j=1}^{m} b_j \delta_{y_j(t)}, \;\; b \in \Sigma_m, \;\; y_j \in \mathbb{R}^d \times Sym_d^+$$

Correspondences



$$\alpha(t) = \sum_i \sum_j \gamma_{ij} \delta_{y_{ij}(t)}$$

$$y_{ij}(t) = ((1-t)m_0^i + tm_1^j, Exp_{\Sigma_0^i}(tLog_{\Sigma_0^i}(\Sigma_1^j)))$$

Data Analysis on Riemannian Domains

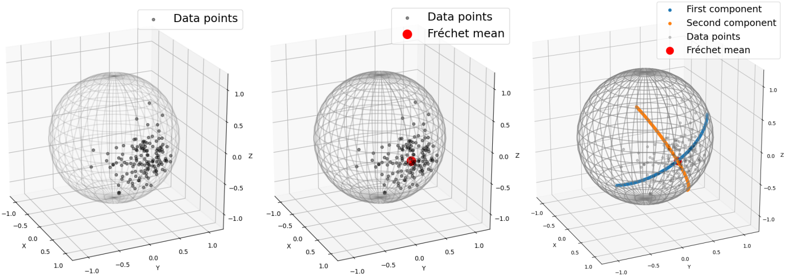


Figure: Left: Data on a Riemannian Manifold (the sphere); Middle; Frechet Mean of data; Right; Principal Geodesics; Images from geomstats website

Basic data analysis pipeline:

1. Calculate a Centroid (Frèchet/ Karcher Mean)
2. Linearize data (Log Map $\Longleftrightarrow$ Riemannian)
3. Apply Euclidean statistical tools to tangent vectors

Barycentric Projection on $\mathbb{R}^d$

- $\nu = \sum_{i=1}^{n} a_i \delta_{x_i}$
- $\mu_\ell = \sum_{j=1}^{m_\ell} b_j \delta_{y_j}$, $\ell = 1, ..., N$
- γ^ℓ the optimal coupling of ν and μ^ℓ

Define

$$T^\ell(x_i) = \sum_j \frac{\gamma_{ij}^\ell}{a_i} y_j$$

Then the barycentric projection of μ w.r.t. ν can be calculated as

$$\tilde{\mu} = T_{\#}^\ell \nu = \sum_{i=1}^{n} a_i \delta_{T^\ell(x_i)}.$$

Linear Optimal Transport

What we really want are vector representations. For this, one can calculate

$$V_i^{\ell} := T^{\ell}(x_i) - x_i$$

$$\implies Vec(V^{\ell}) \in \mathbb{R}^{nd}$$

Barycentric Projection on Riemmanian manifolds

$\mathbb{R}^d$ is a Riemmanian manifold with $Log_x(v) = x - v$. Thus,

$$V_i^\ell = T^\ell(x_i) - x_i$$

$$= \sum_j \frac{\gamma_{ij}}{a_i}(y_j - x_i)$$

$$= \sum_j \frac{\gamma_{ij}}{a_i} Log_{x_i}(y_j)$$

Barycentric Projection - DTMRI

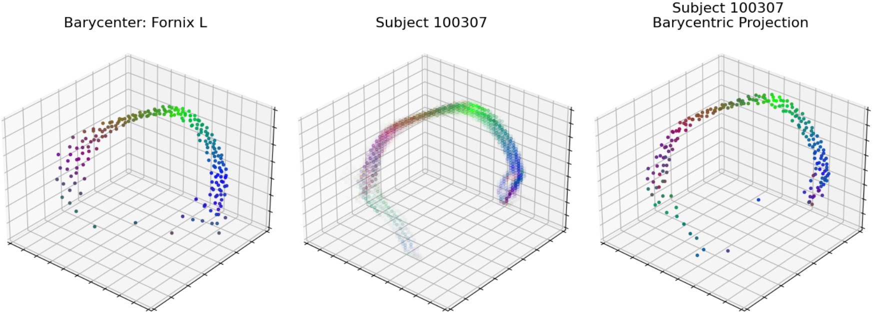


Figure: Points represent locations in $\mathbb{R}^3$, (r, g, b) values correspond to (normalized) diagonal elements of covariance matrices. Left: free support Barycenter of 'Fornix L' in HCP data (200 support points); Middle: Subject 100307's 'Fornix L' in the HCP data (6980 support points); Right: Barycentric Projection of Subject 100307's 'Fornix L' with respect to calculated barycenter (200 support points).

Results

Classification of Gender in HCP-YA/MZ

Region	N	Number of PCs used			Accuracy (%)			
		$\lambda = 0$	$\lambda = \lambda^*$	$\lambda = 1$	$\lambda = 0$	$\lambda = \lambda^*$	$\lambda = 1$	Baseline
Cingulum Frontal Parahippocampal L	513	266	250	46	65.69	65.88	59.84	63.93
Cingulum Frontal Parahippocampal R	476	266	251	50	65.33	62.81	56.51	63.22
Cingulum Frontal Parietal L	786	327	310	78	66.66	65.39	57.88	53.06
Cingulum Frontal Parietal R	785	349	328	84	65.98	66.11	59.61	59.49
Cingulum Parahippocampal L	786	302	311	124	71.37	73.28	61.06	67.17
Cingulum Parahippocampal R	786	306	313	121	70.48	71.75	61.06	63.36
Cingulum Parahippocampal Parietal L	770	329	335	141	66.10	66.88	59.87	63.38
Cingulum Parahippocampal Parietal R	779	329	334	126	66.23	64.31	59.30	57.38
Cingulum Parolfactory L	786	248	229	37	65.52	65.77	54.32	56.85
Cingulum Parolfactory R	786	297	270	38	63.23	65.13	56.23	57.88
Fornix L	771	259	245	93	69.90	70.55	55.25	65.36
Fornix R	768	275	259	95	69.66	68.09	55.46	63.16
Uncinate Fasciculus L	786	290	276	141	65.26	62.84	60.68	55.99
Uncinate Fasciculus R	786	266	256	109	61.57	60.55	61.70	53.80
All* Regions Combined	739	N/A	N/A	N/A	74.96	74.01	69.01	

Table: Leave One Out Cross Validation Accuracy for Support Vector Machine built on PCs capturing 99% of variance for each region, for different values of λ.

Results

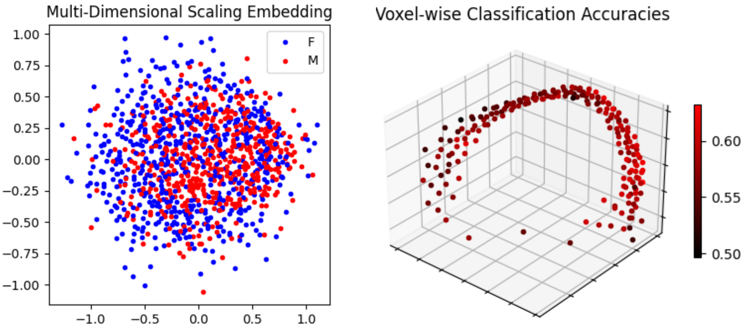


Figure: Left: MDS embedding of LOT embedding; Right: Voxel-wise SVM classification accuracies can be used to localize effects

References

Y. Chen, T. Georgiou, and A. 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, 2020.