A Wasserstein-type distance in the space of Wrapped Gaussian Mixtures

Michael Wilson

Abstract

We present a closed form expression for the Wasserstein Distance between two Wrapped Gaussian Distributions on the sphere. We then show how to use this distance to extend some results for mixtures of Gaussians in $\mathbb{R}^d$ to mixtures of Wrapped Gaussians on the sphere.

1 Introduction

Statistical shape analysis is a field that allows the analysis of complex, high-dimensional data sets with a non-Euclidean form of variation known as domain warping, with applications in a, b, and medical imaging. The Elastic functional data analysis framework provides a Riemannian metric in the shape space, or the quotient space of functions modulo the time-warping group. Although the shape space is a metric space, it is not a complete, separable metric space, which makes it difficult to define distributions of shapes. Usually, because distances in the shape space correspond to distances in the pre-shape space after optimal alignment, and the pre-shape space is a complete metric space, distributions of shapes are defined for aligned functions in the pre-shape space. Because the pre-shape space is a unit sphere, one can compare shape distributions in this way by comparing distributions of optimally aligned functions on the unit sphere.

Still, this does not solve the problem of how to compare these distributions on the sphere. The approach of this paper is to compare distributions of shapes by approximating the distributions as (possibly unimodal) Wrapped Gaussian mixtures, and then calculating a Wasserstein-type distance between them. For Gaussian distributions, a closed form solution for the Wasserstein distance is available, with a similar result holding for Wrapped Gaussian distributions on non-linear manifolds, such as the sphere. By combining a Wasserstein-type distance based on this closed form solution with the theory necessary for extending it to the sphere yields a computationally simple method for calculating Wasserstein-type distances between Gaussian mixture distributions of shapes.

The paper proceeds as follows; in section 2, we cover background information on the Elastic functional data analysis framework, Optimal Transport and the Wasserstein distance, and Wasserstein-type distances for Gaussians and Wrapped Gaussians. In section 3, we present a closed form expression for a Wasserstein

type-distance for Wrapped Gaussian Mixtures. In section 4, we present our computational implementation for calculating the Wasserstein-type distance for Wrapped Gaussian Mixtures on the sphere, and in section 5 we show applications for real and simulated data. Section 6 concludes.

2 Background

2.1 Function Registration with the Fisher-Rao Metric

We provide a brief introduction to Function Registration with the Fisher-Rao metric. For the complete details, see [6].

Let $\mathcal{F} = \{f \in \mathbb{L}^2([0,1],\mathbb{R}) : f \text{ abs. cont.}\}$. For each $f \in \mathcal{F}$, define its square-root velocity function (SRVF) representation to be $q(t) = \dot{f}/\sqrt{\dot{f}}$. One nice property of the SRVF transformation is that the (parameterization invariant) Fisher-Rao distance between two absolutely continuous functions is equal to the $\mathbb{L}^2$ distance between their corresponding SRVF's;

$$d_{FR}(f_1, f_2) = \sqrt{\int_0^1 (q_1(t) - q_2(t))^2 dt} = ||q_1 - q_2||$$

Now, consider the time-warping group $\Gamma = \{\gamma : [0,1] \to [0,1], \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ diffeo}\}$. Most approaches to function registration seek to align functions by solving $\inf_{\gamma\in\Gamma} ||f_1 - f_2 \circ \gamma||$. However, this formulation is not symmetric (and thus does not define a proper distance), and leads to issues of 'pinching' and 'tearing.' By aligning functions with respect to the Fisher-Rao metric, that is by instead focusing on the problem $\inf_{\gamma\in\Gamma} d_{FR}(f_1, f_2 \circ \gamma)$, one finds a Riemannian metric that is invariant to re-parameterization.

To see this, we note that the SRVF of $f(\gamma(t))$ is $q(\gamma(t))\sqrt{\dot{\gamma}(t)}$, and thus defining the group action of Γ on $\mathbb{L}^2$ to be $q * \gamma = q(\gamma(t))\sqrt{\dot{\gamma}(t)}$, we have that

$$\inf_{\gamma\in\Gamma} d_{FR}(f_1 \circ \gamma, f_2) = \inf_{\gamma\in\Gamma} ||q_1 * \gamma - q_2||$$

Because Γ is a group, we can define the orbit of an SRVF $q \in \mathbb{L}^2$ to be $[q] = closure\{q * \gamma : \gamma \in \Gamma\}$, referred to as a 'shape orbit', and define the shape space $\mathcal{S}$ to be the set of all such shape orbits. Thus, defining

$$d_{\mathcal{S}}([q_1], [q_2]) = \inf_{\gamma\in\Gamma} ||q_1 - q_2 * \gamma|| \tag{1}$$

it can be shown that $(\mathcal{S}, d_{\mathcal{S}})$ is a metric space.

2.1.1 Pre-shape space

While $(\mathcal{S}, d_{\mathcal{S}})$ is a metric space, because it is not a complete, seperable metric space, it is generally not possible to model distributions of shapes in the shape space with the distributions commonly used in statistics.

Typically, to get around this, one simply defines distributions over the 'pre-shape' space between optimally aligned functions.

2.2 Optimal Transport/ Wasserstein distances

2.3 Wasserstein and Wasserstein-type distances for Gaussians and Wrapped Gaussians

3 A Wasserstein-type Distance for Wrapped Gaussian Mixtures

Let $\mu_i = N(0, \Sigma_i)$, $i = \{0, 1\}$ be Gaussian distributions defined on $\mathbb{R}^{d-1}$, with random variables $X_i \sim \mu_i$. For some $m_i \in S^{d-1}$, let $\tilde{\mu}_i = Exp(m_i)_{\#}\mu_i$, (and thus $\tilde{\mu}_i = WN(m_i, \Sigma_i)$) and let $\tilde{X}_i \sim \tilde{\mu}_i$. Then,

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

To establish our next result, it is important to note that the optimal coupling is also a wrapped Gaussian. Both of the above facts are proved in [5].

3.1 Wrapped Gaussian Mixture Wasserstein-type Distance

Our proof is identical to the one presented in section 4.2 of [3], except replacing their W_2^2 with our W_2^2, and their $GMM(*)$ with our $WGMM(*)$.

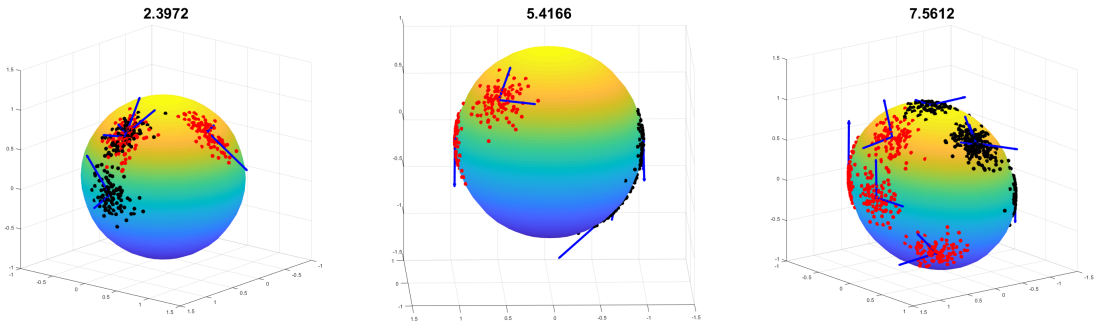

Figure 1: Examples of Wrapped Gaussian Mixture Wasserstein-type distances on the sphere for simulated data

4 Computational Details

4.1 Wrapped Gaussian Wasserstein Distance

Let $\mu_i = WN(m_i, \Sigma_i)$, $i = \{0, 1\}$. Let $U_0 S_0 U_0^t$ denote the SVD of Σ_0, with u_j the jth row of U_0. Define $\Sigma_0^* = U_0^* S_0 U_0^{*t}$, where the jth row of U_0^* is the parallel transport of u_j from $T_{m_0}(S^d)$ to $T_{m_1}(S^d)$;

$$u_j^* = u_j - (2(u_j * m_1^t)(|m_0 + m_1|^2))(m_0 + m_1)$$

Then

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

4.2 Wrapped Gaussian Mixture Wasserstein-like Distance

Let $\mu_i = \Sigma_{k=1}^{K_i} \frac{w_{ik}}{\Sigma_k w_{ik}} WN(m_{ik}, \Sigma_{ik})$, $i = \{0, 1\}$, be two wrapped Gaussian mixtures. Let $\Pi(w_0, w_1) = \{W \in \mathbb{R}^{K_0 \times K_1}, \Sigma_i W_{ij} = w_{1i}, \Sigma_j W_{ij} = w_{0j}\}$. Then,

$$WMW_2^2(\mu_0, \mu_1) = \min_{W \in \Pi(w_0, w_1)} \Sigma_{\ell,k} W_{\ell k} W_2^2(\mu_0^\ell, \mu_1^k) \tag{2}$$

where the minimization is performed with linear programming.

5 Applications

5.1 Simulated data

We present plots of simulated Wrapped Gaussian Mixtures on the sphere, and include the distance calculated.

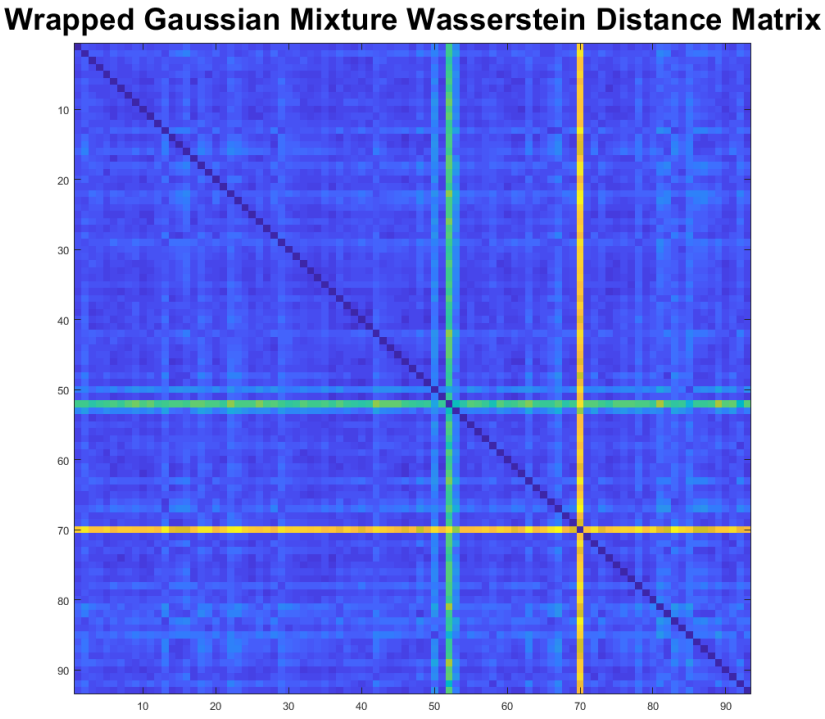

Figure 2: Wrapped Gaussian Mixture Wasserstein-type Distance matrix. The first 42 columns/row correspond to non-PTSD subjects, columns/rows 43:93 correspond to PTSD subjects.

5.2 DTMRI data

One possible application of this approach is the comparison of the distributions of shapes, such as set of DTMRI fiber tracts. The data for this example consists of Fiber tractography data for the right Cingulum Parahippocampal region of 93 subjects in the Grady Trauma Project. 41 are diagnosed with PTSD, 42 are not diagnosed with PTSD. The regions are sets of fiber tracts, represented as curves in $\mathbb{R}^3$. We estimate these distributions by identifying shape modes in the shape space (using k-mode kernel mixture clustering) representing the distributions as Gaussian mixtures in the pre-shape space (which is a sphere) and then calculating the Wrapped Gaussian Mixture Wasserstein-type distance between the distributions of subjects. We then use these distances to classify subjects with respect to their PTSD status.

Specifically, let $X_i = \{f_{ij}(t) \in L^2([0,1] \to \mathbb{R}^3), j = 1, ..., M_j\}$, $i = 1, ..., 93$ correspond to the set of fibers for subject i's right Cingulum Parahippocampal region. For each X_i, we represent it as a wrapped Gaussian mixture $N(m_{ik}, \Sigma_{ik})$, where m_k is the mode kth mode identified using the k-mode kernel mixture algorithm applied to the fibers for subject i, and Σ_{ik} is the covariance of the fibers for subject i assigned to cluster k, calculated in $T_{m_{ik}}(S^{d-1})$, the tangent space of the cluster mode in the pre-shape space. Using these as estimates for wrapped gaussian mixtures, we calculate Wasserstein-type distances using equation (2).

The best classifier we found gets $\sim 60\%$ test set accuracy, which likely isn't significant. This could be because only a subset of mixture components is significant, causing the signal to get drowned out in the calculation of the Wasserstein distance, which takes comparisons of all mixture components down into a single number.

6 Conclusion

References

[1] Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607, 2019.

[2] Luigi Ambrosio and Nicola Gigli. *A User's Guide to Optimal Transport*, pages 1–155. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[3] Julie Delon and Agnes Desolneux. A wasserstein-type distance in the space of gaussian mixture models, 2019.

[4] Clark R. Givens and Rae Michael Shortt. A class of Wasserstein metrics for probability distributions. *Michigan Mathematical Journal*, 31(2):231 – 240, 1984.

[5] Anton Mallasto and Aasa Feragen. Optimal transport distance between wrapped gaussian distributions. *International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering (MaxEnt).*, 38, 2018.

[6] Anuj Srivastava, Wei Wu, Sebastian Kurtek, Eric Klassen, and J. S. Marron. Registration of functional data using fisher-rao metric, 2011.

[7] Asuka Takatsu. On wasserstein geometry of the space of gaussian measures, 2008.

[3] [7] [4] [1] [2] [5]