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 Implementation Details

Proofs are left to Section 2.

1.1 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

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

1.2 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,

$$WMW_2^2(\mu_0, \mu_1) = \min_{W \in \Pi(w_0, w_1)} \sum_{\ell, k} W_{\ell k} W_2^2(\mu_0^\ell, \mu_1^k) \quad (1)$$

2 Theoretical Details

2.1 Wrapped Gaussian Wasserstein Distance

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,

Proof:

Let $\tilde{X} = \tilde{X}_0$ and $\tilde{Y} = Exp_{m_0}(R_\alpha(Exp^{-1}_{m_1}(\tilde{X}_1)))$. Then,

$$W_2^2(\tilde{\mu}_0, \tilde{\mu}_1) = \inf_{\gamma \in \Pi(\tilde{\mu}_0, \tilde{\mu}_1)} \int_{TS^{d-1}} cos^{-1}(\langle \tilde{x}, \tilde{y} \rangle) \; d\tilde{\gamma}(\tilde{x}, \tilde{y})$$

$$= d^2_{S^{d-1}}(E\tilde{X}_1, E\tilde{X}_2) + \inf_{\gamma \in \Pi(\tilde{\mu}_0, \tilde{\mu}_1)} \int_{T_{m_0}(S^{d-1})} \sqrt{\langle Exp^{-1}_{m_0}(\tilde{x}) - R_\alpha(Exp^{-1}_{m_1}(\tilde{y}), Exp^{-1}_{m_0}(\tilde{x}) - R_\alpha(Exp^{-1}_{m_1}(\tilde{y})) \rangle} \; d\tilde{\gamma}(\tilde{x}, \tilde{y}) \tag{2}$$

$$= cos^{-1}(\langle m_0, m_1 \rangle) + \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^{d-1} \times \mathbb{R}^{d-1}} \sqrt{\langle x - y, x - y \rangle} \; d\gamma(x, y) \tag{3}$$

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

Where the equality in equation 1 follows from Proposition 2 in [5], using m_0 as p_{ref}, and the equality in equation 2 holds since $\forall \gamma \in \Pi(\mu_0, \mu_1)$, if $\tilde{\gamma} = Exp(m_0, m_1)_{\#}\gamma$ we have that

$$= \int_{TS^{d-1}} cos^{-1}(\langle \tilde{x}, \tilde{y} \rangle) \; d\tilde{\gamma}(\tilde{x}, \tilde{y})$$

$$= \int_{S^{d-1} \times S^{d-1}} \sqrt{\langle Exp^{-1}_{m_0}(\tilde{x}) - R_\alpha(Exp^{-1}_{m_1}(\tilde{y}), Exp^{-1}_{m_0}(\tilde{x}) - R_\alpha(Exp^{-1}_{m_1}(\tilde{y})) \rangle} \; d\tilde{\gamma}(\tilde{x}, \tilde{y})$$

$$= \int_{\mathbb{R}^{d-1} \times \mathbb{R}^{d-1}} \sqrt{\langle x - y, x - y \rangle} \; d\gamma(x, y)$$

$$\implies \inf_{\gamma \in \Pi(\tilde{\mu}_0, \tilde{\mu}_1)} \int_{TS^{d-1}} cos^{-1}(\langle \tilde{x}, \tilde{y} \rangle) \; d\tilde{\gamma}(\tilde{x}, \tilde{y}) = \inf_{\gamma \in \Pi(\mu_0, \mu_1)} \int_{\mathbb{R}^{d-1} \times \mathbb{R}^{d-1}} \sqrt{\langle x - y, x - y \rangle} \; d\gamma(x, y) \tag{4}$$

Furthermore, because we know that the optimal coupling in the Euclidian case is a Normal distribution, we see that the optimal coupling in the spherical case is therefore a wrapped Gaussian distribution.

2.2 Wrapped Gaussian Mixture Wasserstein-like 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 $WGMM(*)$.

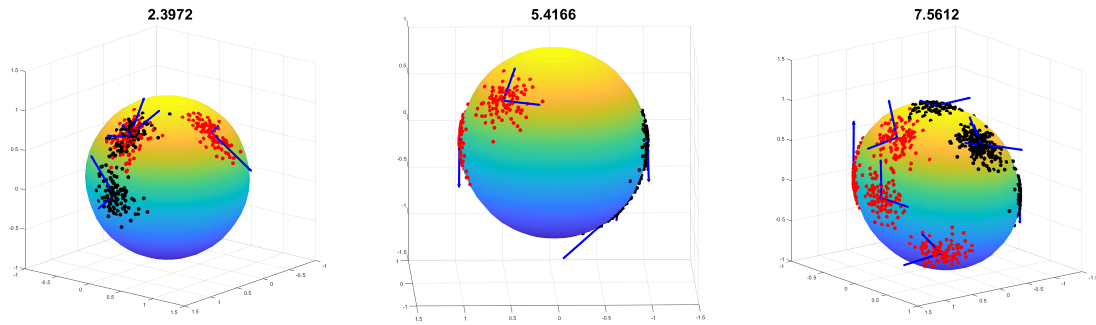

Figure 1: Examples of Wrapped Gaussian Mixture Wasserstein-type distances on the sphere for simulated data

3 Applications

3.1 Simulated data

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

3.2 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 (1).

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.

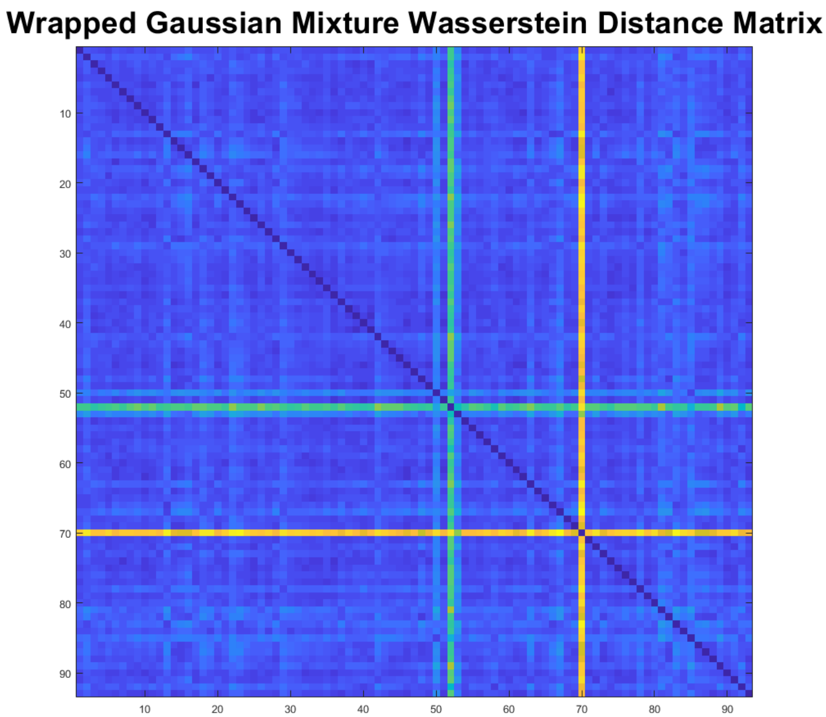

Figure 2: Wrapped Gaussian Mixture Wasserstein-type Distance matrix. The first 42 columns/row correspond to non-PTSD subjects, columns/rows 43:93 correspond to PTSD subjects.

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

[3] [6] [4] [1] [2] [5]