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

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

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,

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

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 the Wasserstein distance between $\tilde{\mu_0}$ and $\tilde{\mu}_1$ can be calculated as;

$$W_2^2(\tilde{\mu}_0, \tilde{\mu}_1) = cos^{-1}(\langle m_0, m_1\rangle) + (\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}})^{\frac{1}{2}})$$

Proof:

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

$$W_2^2(\tilde{\mu}_0, \tilde{\mu}_1) = \inf_{\gamma\in\Pi(\tilde{\mu}_0,\tilde{\mu}_1)} E||\tilde{X}_1 - \tilde{X}_2||$$

$$= ||E\tilde{X}_1 - E\tilde{X}_2|| + \inf_{\gamma\in\Pi(\tilde{\mu}_0,\tilde{\mu}_1)} E||\tilde{X} - \tilde{Y}|| \tag{1}$$

$$= cos^{-1}(\langle m_0, m_1\rangle) + \inf_{\gamma\in\Pi(\mu_0,\mu_1)} E||X - Y|| \tag{2}$$

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

Where the equality in equation 1 holds since;

???

(I think for someone who understands geometry, a proof following that of [5] would be trivial. The issue with my approach is that parallel transport on the sphere does not correspond to a simple translation in the tangent space at a given point: A wrapped normal centered at $m_1 \neq m_0$ does not get pulled back to a normal distribution in $T_{m_0}(S^{d-1})$ by $Exp^{-1}(m_0)^{\#}$.)

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

$$E||\tilde{X} - \tilde{Y}||$$

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

$$= \int_{S^{d-1}\times S^{d-1}} \sqrt{\langle Exp_{m_0}^{-1}(\tilde{x}_0) - R_\alpha(Exp_{m_1}^{-1}(\tilde{x}_1), Exp_{m_0}^{-1}(\tilde{x_0}) - R_\alpha(Exp_{m_1}^{-1}(\tilde{x_1}))\rangle}\ d\tilde{\gamma}(\tilde{x_0}, \tilde{x_1})$$

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

$$= E||X-Y||$$

$$\implies \inf_{\tilde{\gamma}\in\Pi(\tilde{\mu}_0,\tilde{\mu}_1)} E||\tilde{X}-\tilde{Y}|| = \inf_{\gamma\in\Pi(\mu_0,\mu_1)} E||X-Y|| \tag{3}$$

Furthermore, The optimal coupling is also 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 $WGMM(*)$.

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

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