Lecture #17: Optimal Transport

- Recall that for two measures μ_0, μ_1, we have the geodesic in W_2 given by McCann interpolation. Namely, if T^* is the Monge map $\mathbb{R}^d \to \mathbb{R}^d$ that realizes

$$\min_{\substack{T:\mathbb{R}^d \to \mathbb{R}^d \\ T_\# \mu_0 = \mu_1}} \int_{\mathbb{R}^d} \|T(x) - x\|^2 \, d\mu_0(x),$$

Then the curve of measures $\{\mu_t\}_{t \in [0,1]}$ with $\mu_t := [t \cdot Id + (1-t) \cdot T]_\# \mu_0$ is the constant speed geodesic connecting μ_0 to μ_1.

- So, in some sense μ_t is the "weighted average" between μ_0 and μ_1, with relative weights $(t, 1-t)$.
- We can formalize this and consider notions of "averaging" in W_2 as follows:
- Recall that for reals $\{x_i\}_{i=1}^n$, we define the weighted average with respect to weights $\lambda \in \Delta^n = \{(\lambda_1, \dots, \lambda_n) \mid \lambda_i \geq 0, \sum_{i=1}^n \lambda_i = 1\}$, as

$$\bar{x} = \sum_{i=1}^n \lambda_i x_i .$$

- Note that we could have defined $\bar{x}$ variationally as follows:

$$\bar{x} = \arg\min_{z \in \mathbb{R}} \sum_{i=1}^n \lambda_i |x_i - z|^2 \quad \text{(exercise)}$$

- Given any metric space, we could try to mimic this. Doing so in $\mathcal{W}_2 = (\mathcal{P}_2(\mathbb{R}^d), W_2)$ yields the so-called Wasserstein barycenter:

Defn: Let $\{\mu_i\}_{i=1}^p \subset \mathcal{P}_2(\mathbb{R}^d)$ and let $\lambda \in \Delta^p$. The Wasserstein barycenter associated to $\{\mu_i\}_{i=1}^p$ and $\lambda \in \Delta^p$ is defined as

$$\mu^* = \underset{\mu \in \mathcal{P}(\mathbb{R}^d)}{\arg\min} \sum_{i=1}^p \lambda_i W_2^2(\mu_i, \mu).$$

- Questions of existence, uniqueness, and so on are non-trivial but addressed in [Agueh & Carlier, SIAM Analysis, 2011]. Basically, one just needs at least one of the μ_i to be sufficiently regular (i.e., vanishes on small sets).
- It can be shown that when $p=2$, the barycenter formulation

$$\mu^* = \underset{\mu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min} \sum_{i=1}^p \lambda_i W_2^2(\mu, \mu_i)$$

$$= \underset{\mu \in \mathcal{P}_2(\mathbb{R}^d)}{\arg\min} [\lambda_1] W_2^2(\mu_1, \mu) + [1-\lambda_1] W_2^2(\mu, \mu_2)$$

coincides with the λ_1^{th} time point of the constant speed geodesic connecting μ_1 and μ_2.

ex: Gaussian barycenters: Let $\mu_i \sim \mathcal{N}(0, \Sigma_i)$ ${}^{i=1,\dots,p}$ be Gaussian r.v. with mean 0 and covariance $\Sigma_i \in \mathbb{R}^{m \times m}$. Then: (i) $\mu^* = \underset{\mu \in \mathcal{P}_2(\mathbb{R}^m)}{\arg\min} \sum_{i=1}^p \lambda_i W_2^2(\mu_i, \mu)$ is Gaussian

(ii) The covariance Σ^* of μ^* solves $\sum_{i=1}^p \lambda_i [\Sigma^{*\frac{1}{2}} \Sigma_i \Sigma^{*\frac{1}{2}}]^{\frac{1}{2}} = \Sigma^*$.