Linearization of the 2-Wasserstein space and stability of Optimal Transport maps

Alex Delalande
with Quentin Mérigot and Frédéric Chazal

Poster #39
OTML Workshop - NeurIPS 2019

Decemeber 13, 2019

Hilbert space embedding of $(\mathcal{P}(\mathcal{Y}), \mathrm{W}_2)$

Let $\mathcal{X}, \mathcal{Y}$ compact & convex subsets of $\mathbb{R}^d$.

Wasserstein distances:

- Give a geometry to $\mathcal{P}(\mathcal{Y})$.
- Are costly to compute.
- Are not Hilbertian.

Embedding of $(\mathcal{P}(\mathcal{Y}), \mathrm{W}_2)$:
Fix ρ **a.c.** in $\mathcal{P}(\mathcal{X})$ and define for $\mu \in \mathcal{P}(\mathcal{Y})$,

$$\mu \mapsto \left(T_\mu := \arg \min_{\substack{T \in \mathcal{X} \to \mathcal{Y}; \\ T_\# \rho = \mu}} \int_{\mathcal{X}} \|x - T(x)\|^2 \rho(x) \mathrm{d}(x) \right).$$

- $T_\mu \in \mathrm{L}^2(\rho)$ and $\mathrm{W}_{2,\rho}(\mu, \nu) := \| T_\mu - T_\nu \|_{\mathrm{L}^2(\rho)}$ is Hilbertian.
- For a dataset of k measures, need to solve only k OT problems.
- $\mathrm{W}_{2,\rho}$ does not distort too much W_2.

Metric distortion of $\mathrm{W}_{2,\rho}$ w.r.t W_2 (1)

Discriminative power:

- $\mu \mapsto T_\mu$ is injective ($\mu = T_{\mu\#}\rho$)
- $\mu \mapsto T_\mu$ is reverse-Lipschitz: $\|T_\mu - T_\nu\|_{\mathrm{L}^2(\rho)} = \mathrm{W}_{2,\rho}(\mu,\nu) \geq \mathrm{W}_2(\mu,\nu)$

Continuity:

- [Villani, 2003] $\mu \mapsto T_\mu$ is continuous (not quantitative)
- $\mu \mapsto T_\mu$ is **at best** $\frac{1}{2}$-Hölder continuous
- $\mu \mapsto T_\mu$ is indeed $\frac{1}{2}$-Hölder continuous **near a regular measure**
- [Berman, 2018] In general,

$$\mathrm{W}_{2,\rho}(\mu,\nu) \lesssim \mathrm{W}_2(\mu,\nu)^{\frac{1}{(d+2)2^{(d-1)}}}$$

$\implies$ **dimension-dependent Hölder behavior**

Metric distortion of $\mathrm{W}_{2,\rho}$ w.r.t W_2 (2)

Theorem. For $\rho \equiv 1$ on $\mathcal{X}$ with unit volume, for any $\mu, \nu \in \mathcal{P}(\mathcal{Y})$,

$$\boxed{\mathrm{W}_{2,\rho}(\mu, \nu) \lesssim \mathrm{W}_2(\mu, \nu)^{2/15}.}$$

- Dimension independent.
- No assumptions on μ, ν. Main assumptions: $\mathcal{X}$ and $\mathcal{Y}$ have finite diameters.
- No regularization.

Illustrations

Distance approximation:

Use $\mathrm{W}_{2,\rho}(\mu,\nu)$ instead of $\mathrm{W}_2(\mu,\nu)$:

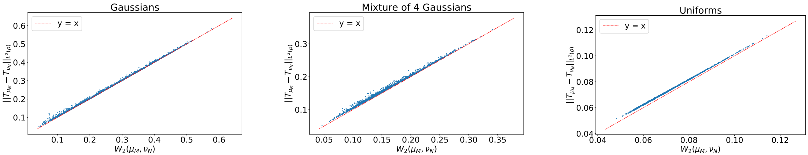

Figure 1: $\mathrm{W}_{2,\rho}$ vs. W_2 on different datasets of measures.

Barycenter approximation:

Use $\left(\sum_{s=1}^{S} \lambda_s T_{\mu_s}\right)_{\#} \rho$ instead of $\arg\min_{\mu} \sum_{s=1}^{S} \lambda_s \mathrm{W}_2^2(\mu, \mu_s)$:

Figure 2: Pushforwards of some Monge embedding barycenters.

References I

Berman, R. J. (2018).
Convergence rates for discretized monge-ampère equations and quantitative stability of optimal transport.
arXiv preprint 1803.00785.

Villani, C. (2003).
Topics in optimal transportation.
American Mathematical Soc.