Fast Optimal Transport through Sliced Generalized Wasserstein Geodesics

Guillaume Mahey, Laetitia Chapel, Gilles Gasso, Clément Bonet, Nicolas Courty

NEURAL INFORMATION PROCESSING SYSTEMS · Université Bretagne Sud ubs · IRISA · INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES ROUEN NORMANDIE · litis · L'INSTITUT agro Rennes Angers

Background on Optimal Transport

- **The square Wasserstein distance** (WD) between μ_1 and $\mu_2 \in \mathcal{P}_2(\mathbb{R}^d)$ is defined as

$$W_2^2(\mu_1, \mu_2) \stackrel{\text{def}}{=} \inf_{\pi \in \Pi(\mu_1,\mu_2)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\boldsymbol{x} - \boldsymbol{y}\|_2^2 \, d\pi(\boldsymbol{x}, \boldsymbol{y})$$

with $\Pi(\mu_1, \mu_2) = \{\pi \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d)$ such that $\pi(\mathbb{R}^d \times A) = \mu_2(A)$ and $\pi(A \times \mathbb{R}^d) = \mu_1(A), \forall A \subset \mathbb{R}^d \}$.

- **The Space** $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is a geodesic metric space of positive curvature, respecting the following inequality:

$$W_2^2(\mu_1, \mu_2) \geq 2W_2^2(\mu_1, \nu) + 2W_2^2(\nu, \mu_2) - 4W_2^2(\mu^{1\to 2}, \nu)$$

for all measures $\nu \in \mathcal{P}_2(\mathbb{R}^d)$ where $\mu^{1\to 2}$ is the Wasserstein mean between μ_1 and μ_2.

- **Solving OT**

WD between empirical measures $\mu_1 = \frac{1}{n}\Sigma_{i=1}^n \delta_{x_i}$ and $\mu_2 = \frac{1}{n}\Sigma_{i=1}^n \delta_{y_i}$ can be computed in $\mathcal{O}(n^3 \log n)$.

When μ_1 and μ_2 are 1D distributions with uniform mass, computing WD can be done by matching the sorted samples, with a complexity of $\mathcal{O}(n + n \log n)$.

$$W_2^2(\mu_1, \mu_2) = \frac{1}{n}\sum_{i=1}^{n}(x_{\sigma(i)} - y_{\tau(i)})^2$$

with σ and τ two permutation operators such that $x_{\sigma(1)} \leq x_{\sigma(2)} \leq ... \leq x_{\sigma(n)}$ and $y_{\tau(1)} \leq y_{\tau(2)} \leq ... \leq y_{\tau(n)}$.

SWGG with permutation

Let μ_1, μ_2 be n-empirical distributions and $\theta \in \mathbb{S}^{d-1}$. Denote by σ_θ and τ_θ the permutations obtained by sorting the 1D projections $P^\theta_\# \mu_1$ and $P^\theta_\# \mu_2$. SWGG is defined as:

$$\text{SWGG}_2^2(\mu_1, \mu_2, \theta) \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \|\boldsymbol{x}_{\sigma_\theta(i)} - \boldsymbol{y}_{\tau_\theta(i)}\|_2^2.$$

SWGG only involves projection and sorting and comes with a transport map:

$$T(\boldsymbol{x}_i) = \boldsymbol{y}_{\tau_\theta^{-1}(\sigma_\theta(i))}, \quad \forall 1 \leq i \leq n.$$

SWGG with generalized geodesics

Let $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, a generalized geodesic draws a correspondence between μ_1 and μ_2, through the correspondences between μ_1 and ν, and μ_2 and ν:

$$T_\nu^{1\to 2} \stackrel{\text{def}}{=} T^{\nu \to \mu_2} \circ T^{\mu_1 \to \nu} \quad \text{with} \quad (T_\nu^{1\to 2})_\# \mu_1 = \mu_2.$$

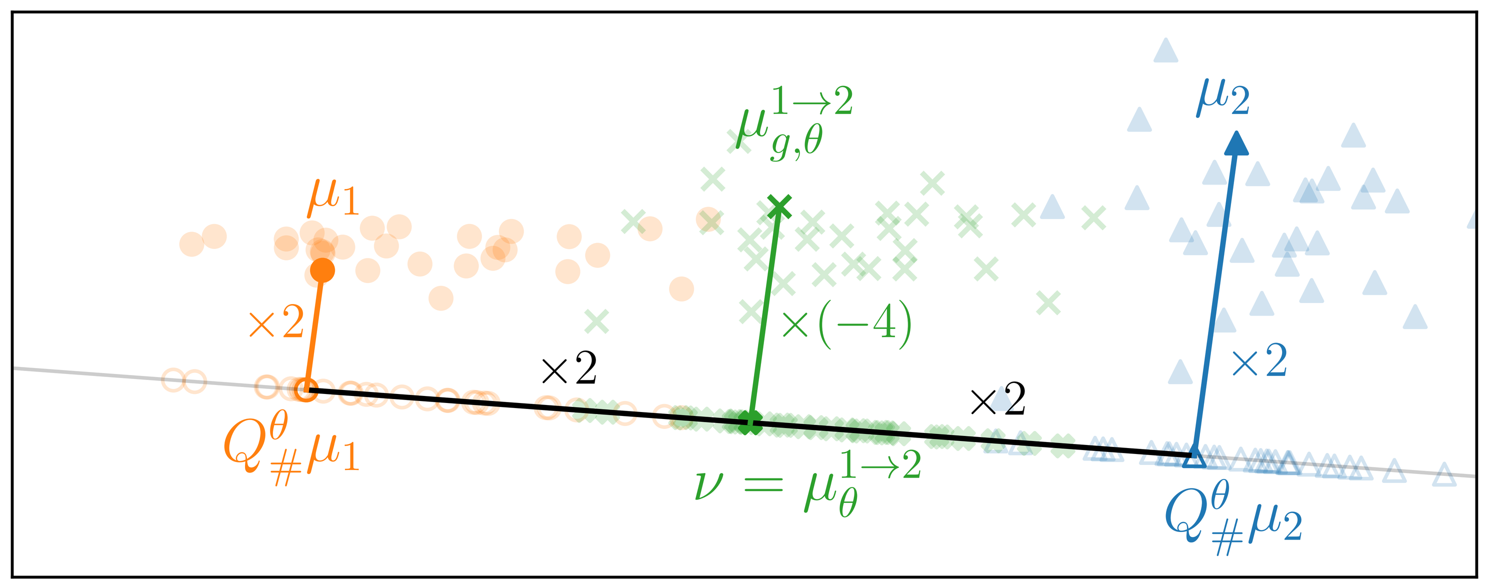
The square ν-Wasserstein distance is then given by:

$$\begin{aligned} W_\nu^2(\mu_1, \mu_2) &\stackrel{\text{def}}{=} \int_{\mathbb{R}^d} \|\boldsymbol{x} - T_\nu^{1\to 2}(\boldsymbol{x})\|_2^2 \, d\mu_1(\boldsymbol{x}) \\ &= 2W_2^2(\mu_1, \nu) + 2W_2^2(\nu, \mu_2) - 4W_2^2(\mu_g^{1\to 2}, \nu). \end{aligned}$$

where $\mu_g^{1\to 2}$ is the middle of the geodesic given by $T_\nu^{1\to 2}$.

When ν is taken to be the middle of the geodesic of $Q^\theta_\# \mu_1$ and $Q^\theta_\# \mu_2$, with $Q^\theta : x \mapsto \theta\langle x, \theta\rangle$, we have:

$$\text{SWGG}_2^2(\mu_1, \mu_2, \theta) = W_\nu^2(\mu_1, \mu_2).$$



Properties

SWGG is **an upper bound** of WD.

SWGG is a **distance** which **metricizes the weak convergence** of measure. Moreover, it has the same behavior with **translation of measure** than WD.

SWGG has a complexity of $\mathcal{O}(dn + n \log n)$ (akin to sliced Wasserstein).

SWGG delivers a **sparse transport plan**.

SWGG definition allows us to show a **closed form for WD** whenever μ_2 is supported on a line.

Optimization

Since it serves as an upper limit for WD, our objective is to minimize SWGG with respect to θ in order to closely approximate WD:

$$\text{min-SWGG}_2^2(\mu_1, \mu_2) \stackrel{\text{def}}{=} \min_{\theta \in \mathbb{S}^{d-1}} \text{SWGG}_2^2(\mu_1, \mu_2, \theta).$$

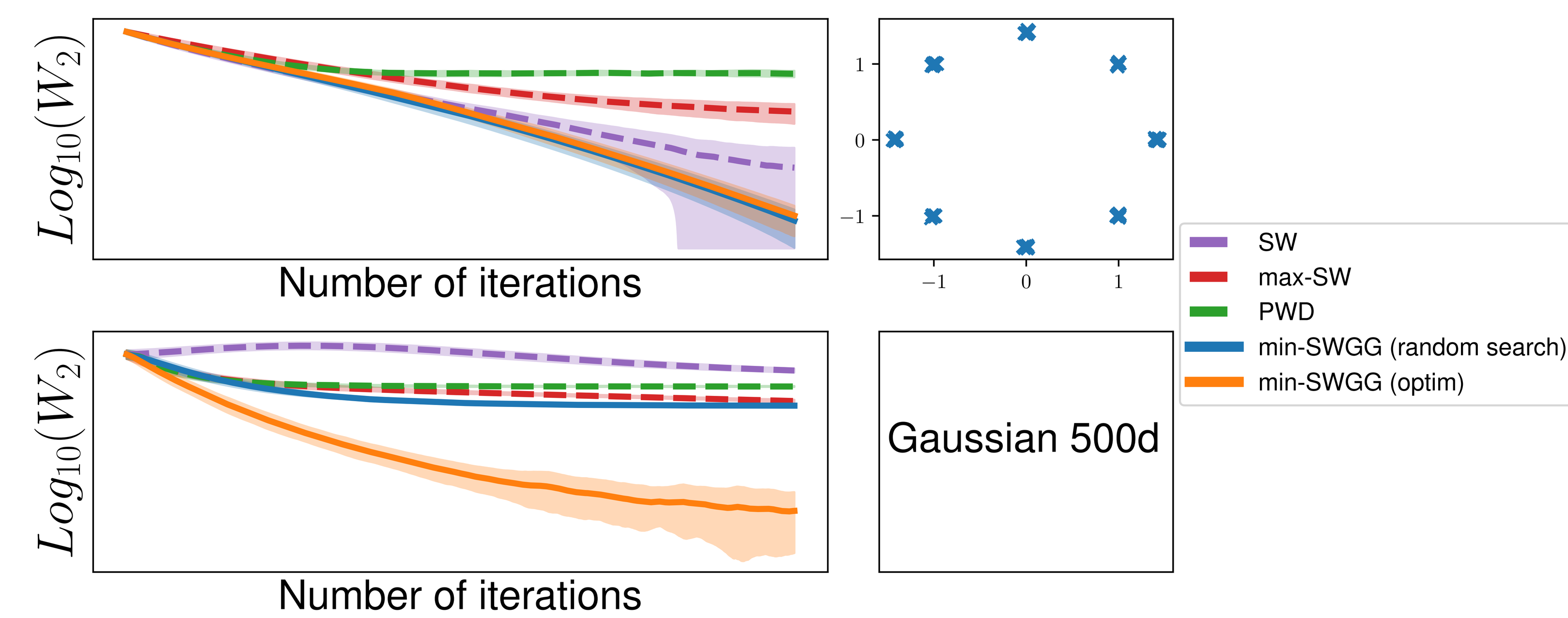
We propose two schemes: i) random search, appropriate in low dimension d ii) gradient descent on $\mathbb{S}^{d-1}$, thanks to the generalized geodesic definition of SWGG, optimization after a smoothing of $\mu_g^{1\to 2}$.

Experiments

Code available at `https://github.com/MaheyG/SWGG`

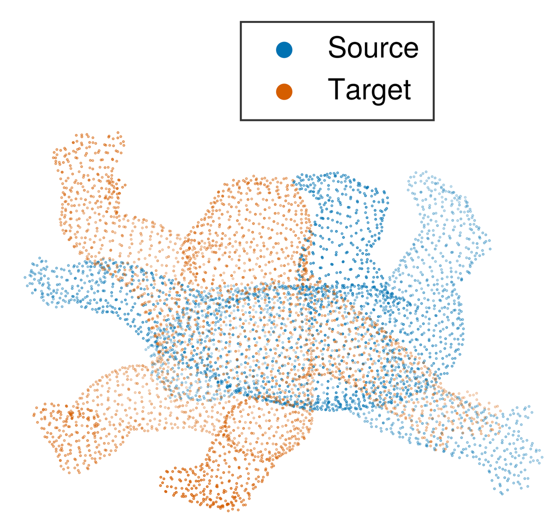
- **Gradients Flows**

Starting from a random initial distribution, we move the particles of a source distribution μ_1 towards a target one μ_2 by reducing min-SWGG(μ_1, μ_2) at each step. We compare both variants of min-SWGG against SW, max-SW and PWD.



- **Point Cloud Registration**

Iterative Closest Point defines a one-to-one correspondence, computes a rigid transformation, moves the source point clouds using the transformation, and iterates the process until convergence. We perform ICP with different matching: NN, OT and min-SWGG transport map.



$n =$	500	3000	150 000
NN	3.54 (**0.02**)	96.9 (**0.30**)	23.3 (**59.37**)
OT	0.32 (0.18)	48.4 (58.46)	·
min-SWGG	**0.05** (0.04)	**37.6** (0.90)	**6.7** (105.75)

Sinkhorn Divergence between final transformation. Timings in seconds are into parenthesis.