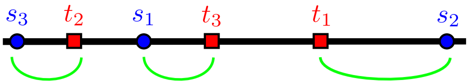
Properties of Discrete Sliced Wasserstein Losses

Eloi Tanguy, Rémi Flamary, Julie Delon

MAP5, Université Paris-Cité

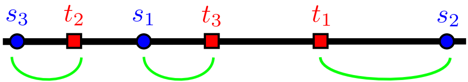
June 2023

1D Wasserstein and Sliced Wasserstein

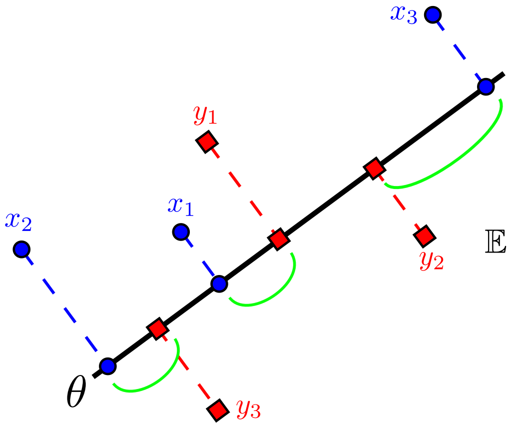


$$\mathrm{W}_2^2(\gamma_S, \gamma_T) = \frac{1}{n}\sum_{i=1}^{n} \left|s_{\sigma(i)} - t_{\tau(i)}\right|^2$$

1D Wasserstein and Sliced Wasserstein



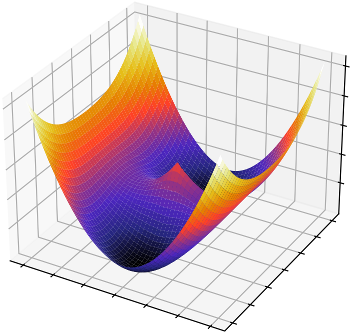
$$\mathrm{W}_2^2(\gamma_S, \gamma_T) = \frac{1}{n}\sum_{i=1}^{n} \left|s_{\sigma(i)} - t_{\tau(i)}\right|^2$$



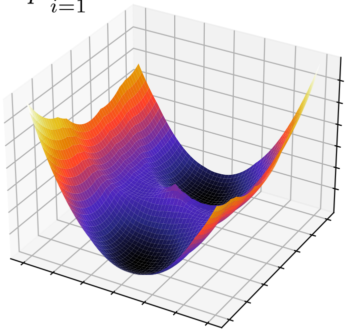
$$\mathrm{SW}_2^2(\gamma_X, \gamma_Y) = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)}\left[\mathrm{W}_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)\right]$$

Monte-Carlo Approximation

$$\mathcal{E}(X) = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} \left[\mathrm{W}_2^2(\theta \# \gamma_X, \theta \# \gamma_Y) \right]$$

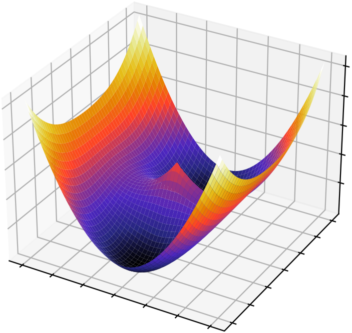


$$\mathcal{E}_p(X) := \frac{1}{p} \sum_{i=1}^{p} \mathrm{W}_2^2(\theta_i \# \gamma_X, \theta_i \# \gamma_Y)$$

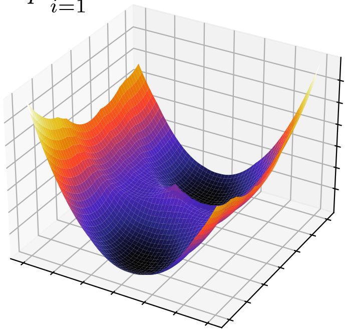


Monte-Carlo Approximation

$$\mathcal{E}(X) = \mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^d)}\left[\mathrm{W}_2^2(\theta\#\gamma_X, \theta\#\gamma_Y)\right]$$



$$\mathcal{E}_p(X) := \frac{1}{p}\sum_{i=1}^{p}\mathrm{W}_2^2(\theta_i\#\gamma_X, \theta_i\#\gamma_Y)$$

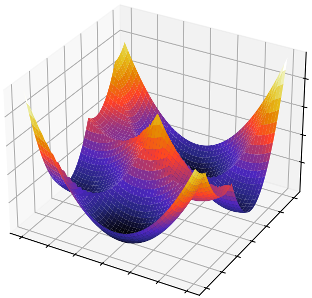


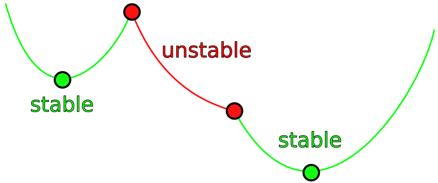
Uniform Convergence

For $\mathcal{K} \subset \mathbb{R}^{n\times d}$ compact, $\mathbb{P}\left(\|\mathcal{E}_p - \mathcal{E}\|_{\infty,\mathcal{K}} \xrightarrow[p\to+\infty]{} 0\right) = 1.$

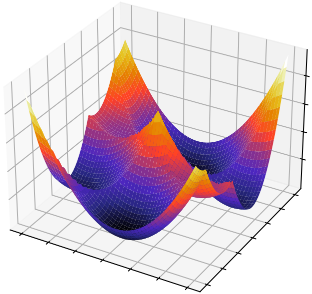
1. The Discrete Sliced Wasserstein Distance
2. Optimisation Properties
3. SGD Convergence

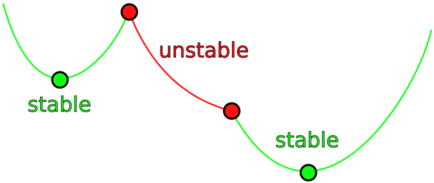
$\mathcal{E}_p$ Cell Decomposition





$\mathcal{E}_p$ Cell Decomposition

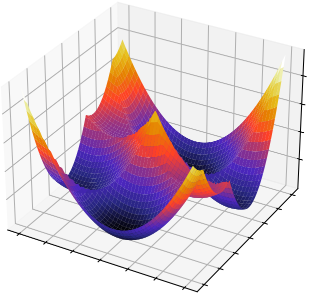


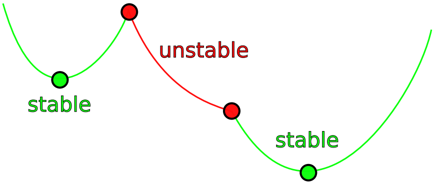


Cell Optima

$\nabla \mathcal{E}_p(X) = 0 \iff X$ is min of a stable cell $\iff X$ is a local min.

$\mathcal{E}_p$ Cell Decomposition



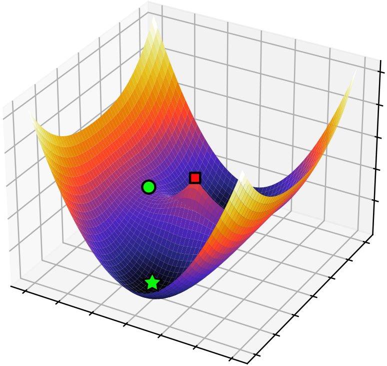


Cell Optima

$\nabla \mathcal{E}_p(X) = 0 \Longleftrightarrow X$ is min of a stable cell $\Longleftrightarrow X$ is a local min.

As $p \longrightarrow +\infty$, $\mathcal{E}_p \approx \mathcal{E}$, more local optima but better optimisation.

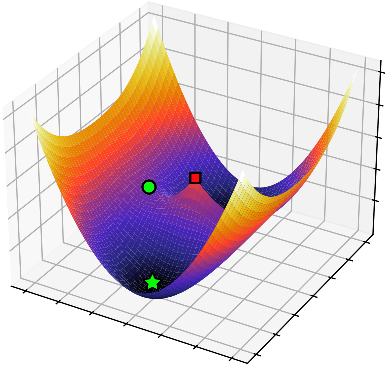
$\mathcal{E}$ Differentiable Critical Points



Critical Points of $\mathcal{E}$

$$\forall X \in \mathcal{D}_{\mathcal{E}},$$
$$\nabla \mathcal{E}(X) = 0 \iff F(X) = X$$

$\mathcal{E}$ Differentiable Critical Points



Critical Points of $\mathcal{E}$

$$\forall X \in \mathcal{D}_{\mathcal{E}},$$
$$\nabla \mathcal{E}(X) = 0 \iff F(X) = X$$

Critical Point Approximation

For X_p critical points of $\mathcal{E}_p$, $\quad X_p - F(X_p) \xrightarrow[p \longrightarrow +\infty]{\mathbb{P}} 0.$

Convergence of Interpolated Trajectories

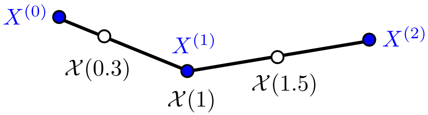
SGD on $\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^d)}\left[\underbrace{\mathrm{W}_2^2(\theta\#\gamma_X, \theta\#\gamma_Y)}_{w_\theta(X)}\right]$:

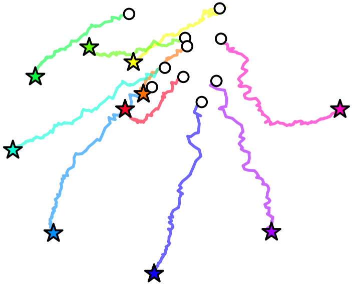
$$X^{(k+1)} = X^{(k)} - \alpha\nabla w_{\theta^{(k+1)}}(X^{(k)})$$

Convergence of Interpolated Trajectories

SGD on $\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^d)}\left[\underbrace{\mathrm{W}_2^2(\theta\#\gamma_X, \theta\#\gamma_Y)}_{w_\theta(X)}\right]$:

$$X^{(k+1)} = X^{(k)} - \alpha\nabla w_{\theta^{(k+1)}}(X^{(k)})$$

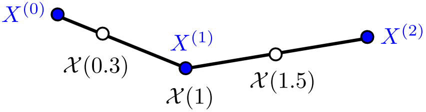


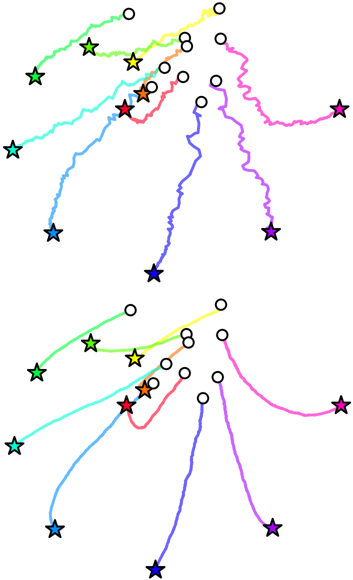


Convergence of Interpolated Trajectories

SGD on $\mathbb{E}_{\theta\sim\mathcal{U}(\mathbb{S}^d)}\left[\underbrace{\mathrm{W}_2^2(\theta\#\gamma_X, \theta\#\gamma_Y)}_{w_\theta(X)}\right]$:

$$X^{(k+1)} = X^{(k)} - \alpha\nabla w_{\theta^{(k+1)}}(X^{(k)})$$





Convergence of Interpolations

$$d(\mathcal{X}_\alpha, \mathcal{S}) \xrightarrow[\alpha\to 0]{\mathbb{P}} 0.$$

With $\mathcal{S} = \left\{\mathcal{X} \;\middle|\; \frac{\mathrm{d}\mathcal{X}}{\mathrm{d}t}(t) \in -\partial_C\mathcal{E}(\mathcal{X}(t))\right\}$.

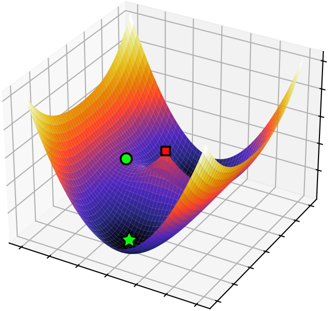
Convergence of Noised Trajectories

Noised SGD: $X^{(k+1)} = X^{(k)} - \alpha \nabla w_{\theta^{(k+1)}}(X^{(k)}) + \alpha \varepsilon^{(k+1)}$.

Convergence of Noised SGD

$$\overline{\lim_{k \longrightarrow +\infty}} d(X_\alpha^{(k)}, \mathcal{Z}) \xrightarrow[\alpha \longrightarrow 0]{\mathbb{P}} 0.$$

With $\mathcal{Z} = \left\{ X \in \mathbb{R}^{n \times d} \mid 0 \in -\partial_C \mathcal{E}(X) \right\}$.



Thank You