

# Properties of Discrete Sliced Wasserstein Losses

**Eloi Tanguy**, Rémi Flamary, Julie Delon

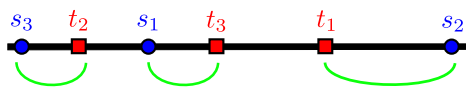
MAP5, Université Paris-Cité

June 2023



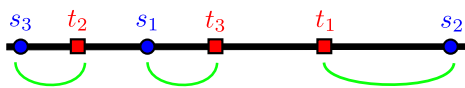
- 1 The Discrete Sliced Wasserstein Distance
- 2 Optimisation Properties
- 3 SGD Convergence

## 1D Wasserstein and Sliced Wasserstein



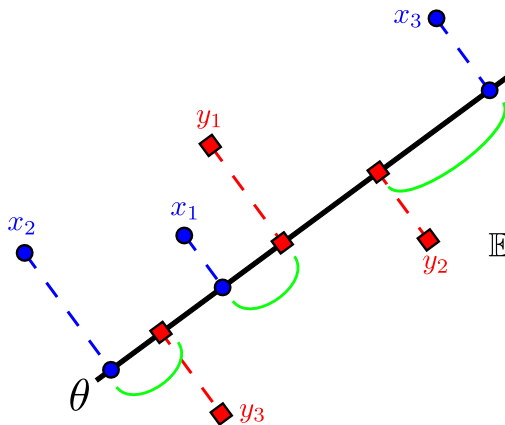
$$W_2^2(\gamma_S, \gamma_T) = \frac{1}{n} \sum_{i=1}^n |s_{\sigma(i)} - t_{\tau(i)}|^2$$

## 1D Wasserstein and Sliced Wasserstein



A horizontal line representing a 1D space. On the left, there are three blue dots labeled  $s_3, s_1, s_2$  from left to right. On the right, there are three red squares labeled  $t_2, t_3, t_1$  from left to right. Green curved lines connect  $s_3$  to  $t_2$ ,  $s_1$  to  $t_3$ , and  $s_2$  to  $t_1$ .

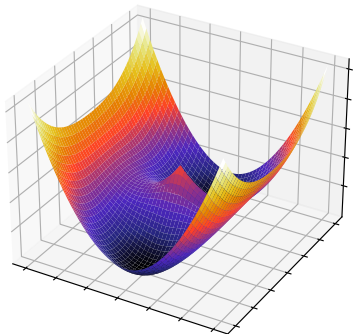
$$W_2^2(\gamma_S, \gamma_T) = \frac{1}{n} \sum_{i=1}^n |s_{\sigma(i)} - t_{\tau(i)}|^2$$



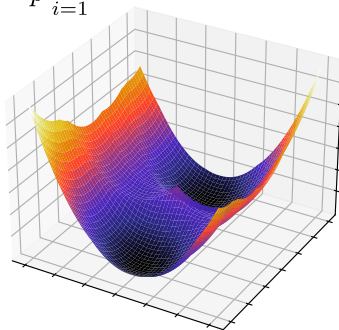
$$SW_2^2(\gamma_X, \gamma_Y) = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} [W_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)]$$

# Monte-Carlo Approximation

$$\mathcal{E}(X) = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} [W_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)]$$

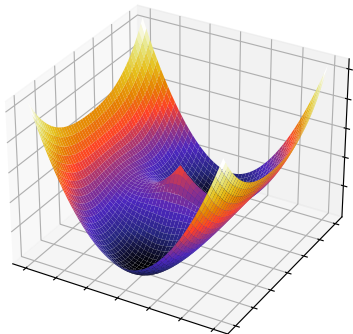


$$\mathcal{E}_p(X) := \frac{1}{p} \sum_{i=1}^p W_2^2(\theta_i \# \gamma_X, \theta_i \# \gamma_Y)$$

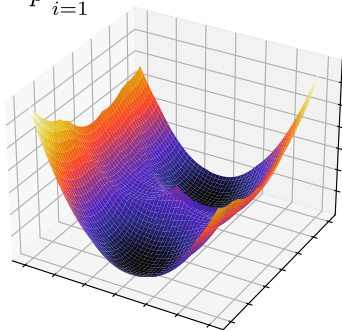


# Monte-Carlo Approximation

$$\mathcal{E}(X) = \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} [W_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)]$$



$$\mathcal{E}_p(X) := \frac{1}{p} \sum_{i=1}^p W_2^2(\theta_i \# \gamma_X, \theta_i \# \gamma_Y)$$

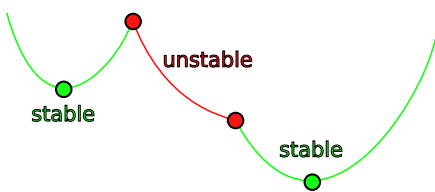
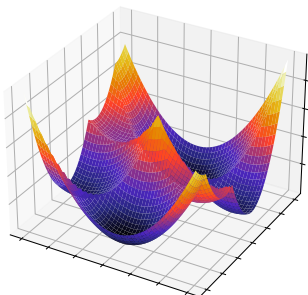


## Uniform Convergence

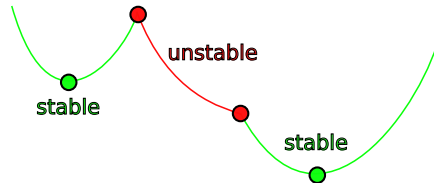
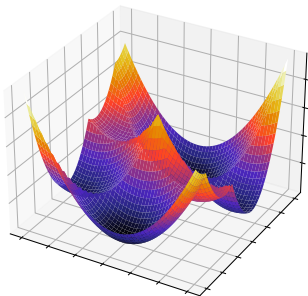
For  $\mathcal{K} \subset \mathbb{R}^{n \times d}$  compact,  $\mathbb{P} \left( \|\mathcal{E}_p - \mathcal{E}\|_{\infty, \mathcal{K}} \xrightarrow{p \rightarrow +\infty} 0 \right) = 1$ .

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

# $\mathcal{E}_p$ Cell Decomposition

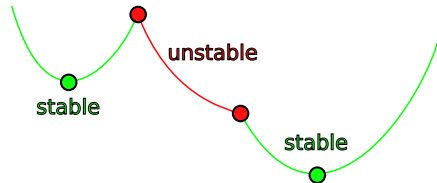
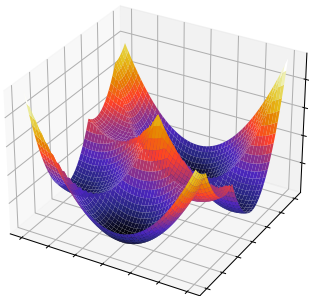




$\mathcal{E}_p$  Cell Decomposition

## Cell Optima

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

$\mathcal{E}_p$  Cell Decomposition

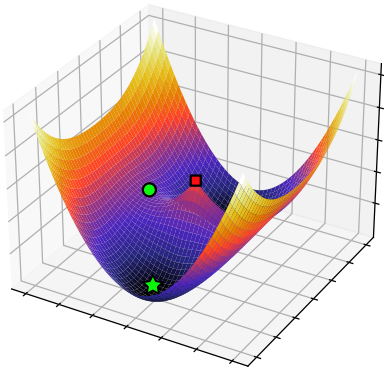
## Cell Optima

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



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

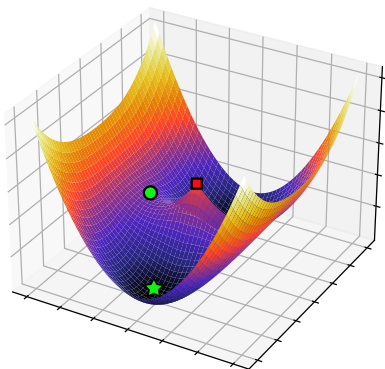
# $\mathcal{E}$ Differentiable Critical Points



## 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

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

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

## Convergence of Interpolated Trajectories

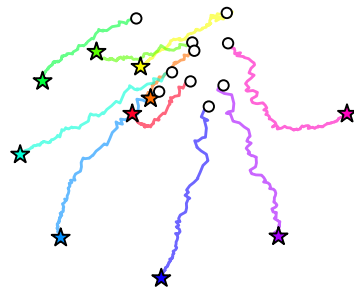
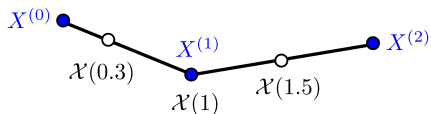
SGD on  $\mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} \left[ \underbrace{W_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)}_{w_\theta(X)} \right] :$

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

# Convergence of Interpolated Trajectories

$$\text{SGD on } \mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} \left[ \underbrace{W_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)}_{w_\theta(X)} \right] :$$

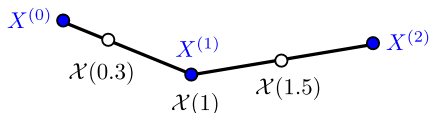
$$X^{(k+1)} = X^{(k)} - \alpha \nabla w_{\theta^{(k+1)}}(X^{(k)})$$



## Convergence of Interpolated Trajectories

SGD on  $\mathbb{E}_{\theta \sim \mathcal{U}(\mathbb{S}^d)} \left[ \underbrace{W_2^2(\theta \# \gamma_X, \theta \# \gamma_Y)}_{w_\theta(X)} \right] :$

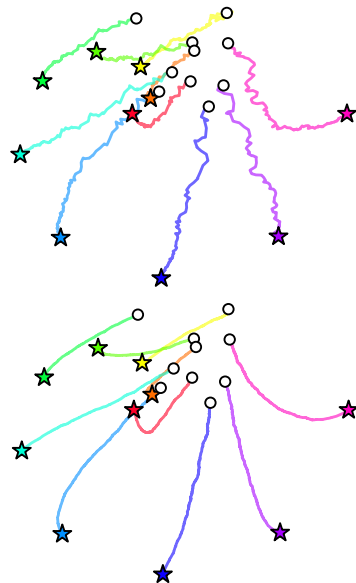
$$X^{(k+1)} = X^{(k)} - \alpha \nabla w_{\theta^{(k+1)}}(X^{(k)})$$



## Convergence of Interpolations

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

$$\text{With } \mathcal{S} = \left\{ \mathcal{X} \mid \frac{d\mathcal{X}}{dt}(t) \in -\partial_C \mathcal{E}(\mathcal{X}(t)) \right\}.$$





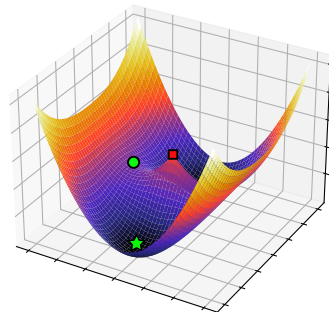
# Convergence of Noised Trajectories

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

## Convergence of Noised SGD

$$\overline{\lim}_{k \rightarrow +\infty} d(X_{\alpha}^{(k)}, \mathcal{Z}) \xrightarrow{\alpha \rightarrow 0} 0.$$

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



*Thank You*