

A Survey on the Acceleration of GW

HCP Based And Ball Tree Based

Junyi Lin, Dunyao Xue

Institute of Statistics and Big Data

2024.12.24

Wasserstein Distance

For $(\mu_1, \mu_2) \in \mathcal{M}_1^+(\mathbb{R}^d)^2$, we define the L^2 -Wasserstein distance $\text{W}_{\mathbb{R}^d}(\mu_1, \mu_2)^2$ to be equal to

$$\inf \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x_1 - x_2\|^2 \, d\gamma(x_1, x_2); \gamma \in C(\mu_1, \mu_2) \right\}$$

where $C(\mu_1, \mu_2) = \{\gamma \in \mathcal{M}_1^+(\mathbb{R}^d \times \mathbb{R}^d); \Pi_i \sharp \gamma = \mu_i, i = 1, 2\}$

Sliced Wasserstein Distance[Bonneel et al.(2015)]

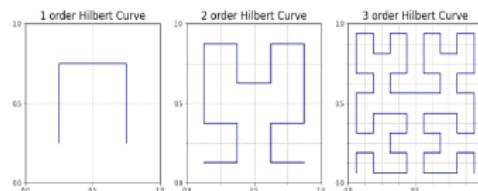
Sliced Wasserstein distance is defined as

$$\text{SW}_{\mathbb{R}^d}(\mu_1, \mu_2)^2 = \text{W}_{\Omega^d}(R\mu_1, R\mu_2)^2 = \int_{\mathbb{S}^{d-1}} \text{W}_{\mathbb{R}}(P_\theta \sharp \mu_1, P_\theta \sharp \mu_2)^2 \, d\theta.$$

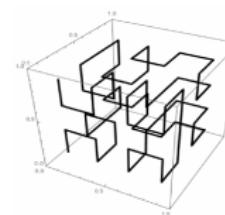
where $d\theta$ is the uniform measure on \mathbb{S}^{d-1} , normalized so that

$$\int_{\mathbb{S}^{d-1}} d\theta = 1.$$

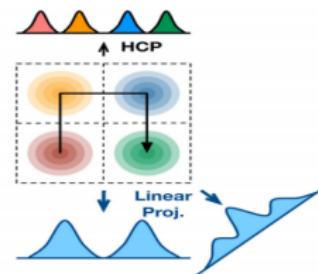
Hilbert Curve



(a) 2Dhcurve



(b) 3Dhcurve



(c) HCP VS SWD

HCP Distance [Li et al.(2024)]

For $p \in \mathbb{Z}_+$, the p -order Hilbert curve projection distance is defined as:

$$\text{HCP}_p(\mu, \nu) = \left(\int_0^1 \|H_\mu(g_\mu^{-1}(t)) - H_\nu(g_\nu^{-1}(t))\|_p^p dt \right)^{\frac{1}{p}}.$$

Gromov Wasserstein Distance

GW[Mémoli(2011)]

(2, 2)-GW distance between Euclidean mm spaces $(\mathbb{R}^{d_x}, \|\cdot\|, \mu)$ and $(\mathbb{R}^{d_y}, \|\cdot\|, \nu)$ of different dimensions:

$$D(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \left(\int_{\mathbb{R}^{d_x} \times \mathbb{R}^{d_y}} \int_{\mathbb{R}^{d_x} \times \mathbb{R}^{d_y}} |\Delta|^2 d\pi \otimes \pi(x, y, x', y') \right)^{\frac{1}{2}}$$

where $\Delta = \|x - x'\|^2 - \|y - y'\|^2$.

EGW[Rioux et al.(2023)]

GW distance with entropic regularization, which, for $\varepsilon > 0$:

$$S_\varepsilon(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \|\Delta\|_{L^2(\pi \otimes \pi)}^2 + \varepsilon D_{KL}(\pi \|\mu \otimes \nu).$$

Existence of a Gromov-Monge map[Dumont et al.(2024)]

Let $n \geq d$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^d)$ two measures with compact supports. Let π^* be a solution of $(2, 2)$ -GW and $M^* := \int y' x'^\top d\pi^*(x', y')$. Then:

- (i) if $\text{rk } M^* = d$, there exists an optimal correspondence 2-map, which is also a map/anti-map;
- (ii) if $\text{rk } M^* = d - 1$, there exists an optimal correspondence 2-map;
- (iii) if $\text{rk } M^* \leq d - 2$, there exists an optimal correspondence map.

Therefore, regardless of the rank of M^* , there always exists an optimal correspondence 2-map, i.e. a plan that is supported on the graph of two maps.

SGW[Vayer et al.(2019)]

Let $\mathbf{S}^{q-1} = \{\theta \in \mathbb{R}^q : \|\theta\|_{2,q} = 1\}$ be the q -dimensional hypersphere, $P_\theta(x) = \langle x, \theta \rangle$. For a linear map $\Delta \in \mathbb{R}^{q \times p}$, define the Sliced Gromov Wasserstein as follows:

$$\begin{aligned} SGW_\Delta(\mu, \nu) &= \mathbb{E}_{\theta \sim \lambda_{q-1}} [GW_2^2(d^2, P_\theta \# \mu_\Delta, P_\theta \# \nu)] \\ &= \frac{1}{\text{vol}(\mathbf{S}^{q-1})} \int_{\mathbf{S}^{q-1}} GW_2^2(d^2, P_\theta \# \mu_\Delta, P_\theta \# \nu) d\theta \end{aligned}$$

where $\mu_\Delta = \Delta \# \mu \in \mathcal{P}(\mathbb{R}^q)$.

RISGW

A variant of SGW that does not depends on the choice of Δ called Rotation Invariant SGW (RISGW):[Vayer et al.(2019)]

$$RISGW(\mu, \nu) = \min_{\Delta \in \mathbb{V}_p(\mathbb{R}^q)} SGW_\Delta(\mu, \nu)$$



Decomposition of EGW[Zhang et al.(2024)]

EGW functional can be split into two terms as

$$S_\varepsilon(\mu, \nu) = S^1(\mu, \nu) + S_\varepsilon^2(\mu, \nu),$$

where

$$\begin{aligned} S^1(\mu, \nu) := & \int \|x - x'\|^4 d\mu \otimes \mu(x, x') + \int \|y - y'\|^4 d\nu \otimes \nu(y, y') \\ & - 4 \int \|x\|^2 \|y\|^2 d\mu \otimes \nu(x, y) \end{aligned}$$

$$\begin{aligned} S_\varepsilon^2(\mu, \nu) := & \inf_{\pi \in \Pi(\mu, \nu)} \int -4\|x\|^2 \|y\|^2 d\pi(x, y) \\ & - 8 \sum_{\substack{1 \leq i \leq d_1 \\ 1 \leq j \leq d_2}} \left(\int x_i y_j d\pi(x, y) \right)^2 + \varepsilon D_{KL}(\pi \| \mu \otimes \nu). \end{aligned}$$

EGW duality

Fix $\varepsilon > 0$, let $(\mu, \nu) \in \mathcal{P}_4(\mathbb{R}^{d_x}) \times \mathcal{P}_4(\mathbb{R}^{d_y})$, and define $M_{\mu, \nu} := \sqrt{M_2(\mu)M_2(\nu)}$. We have

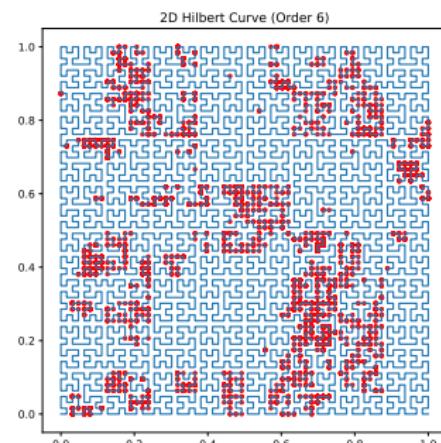
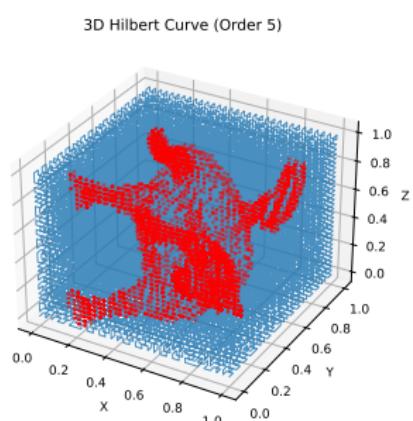
$$\text{S}_{\varepsilon}^2(\mu, \nu) = \inf_{\mathbf{A} \in \mathbb{R}^{d_x \times d_y}} 32\|\mathbf{A}\|_{\text{F}}^2 + \text{OT}_{\mathbf{A}, \varepsilon}(\mu, \nu) \quad (1)$$

- where $\text{OT}_{\mathbf{A}, \varepsilon}$ is the EOT problem with cost function $c_{\mathbf{A}} : (x, y) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \mapsto -4\|x\|^2\|y\|^2 - 32x^{\top}\mathbf{A}y$.
- Infimum is achieved at $\mathbf{A}^* \in \mathcal{D}_{M_{\mu, \nu}} := [-M_{\mu, \nu}/2, M_{\mu, \nu}/2]^{d_x \times d_y}$.

Connection Between GWD And WD

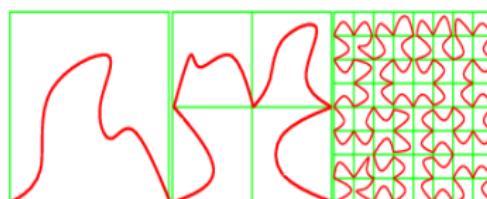
If μ_0 and μ_1 are centered and π_* is optimal for the original EGW , then $\mathbf{A}^* = \frac{1}{2} \int xy^{\top} d\pi_*(x, y)$ is optimal for (1) and $\pi_* = \pi_{\mathbf{A}^*}$, where $\pi_{\mathbf{A}^*}$ is the unique EOT coupling for $\text{OT}_{\mathbf{A}^*, \varepsilon}(\mu_0, \mu_1)$.

Limitations of HCP

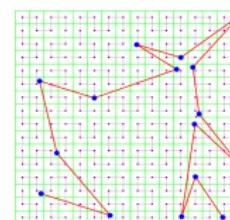


We projected a 3D point cloud of a horse onto a Hilbert curve. When the Hilbert curve was flattened into two dimensions, the original horse structure was disrupted. This suggests that the Hilbert curve cannot consistently capture information from different spaces.

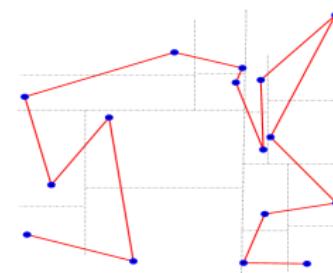
Connection Between HCP And KD-Tree



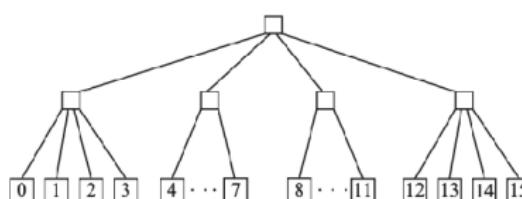
(d) 2Dhcurve



(e) 3Dhcurve



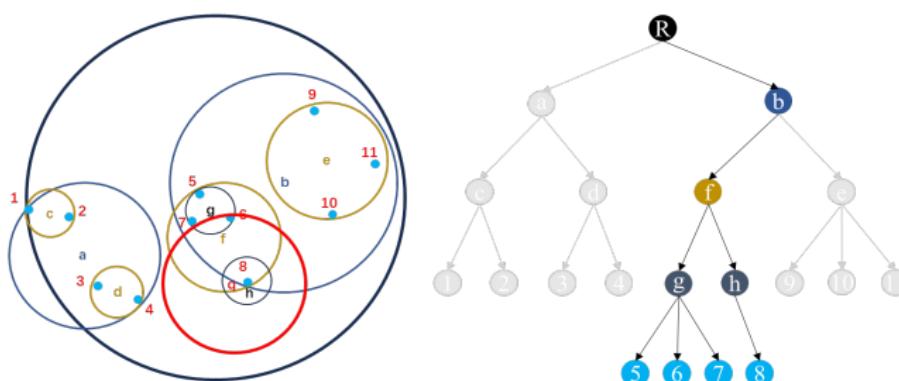
(f) HCP VS SWD



5	6	9	10
4	7	8	11
3	2	13	12
0	1	14	15

Fig. 1.5 Quadtree representation of a regular 4×4 grid, and a sequential order that avoids jumps

Ball Tree



Difference Between KD Tree And Ball Tree

KD tree depends on the coordinate system of Euclidean space, Ball Tree only depends on the internal relationship between points and the metric of the space.

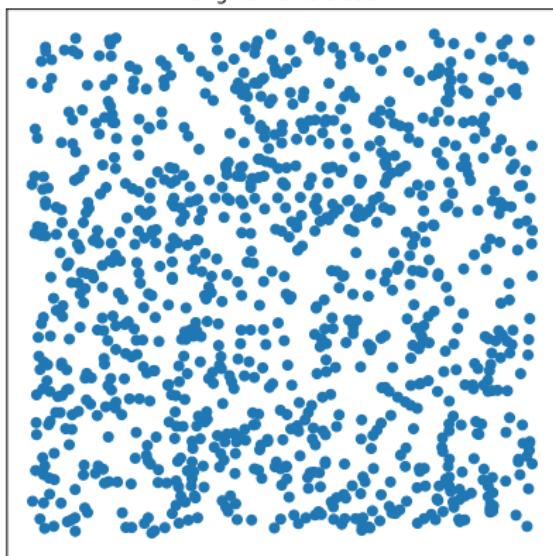
Ball Tree Sorting

Algorithm. 1 Ball Tree Construction Algorithm

- 1: Compute the mean of all data points and set it as the root node \mathcal{R} of the ball tree;
- 2: Find the point p_1 farthest from the root \mathcal{R} , and then find the point p_2 farthest from p_1 . Set the distance from \mathcal{R} to p_1 as the radius r of the current hypersphere;
- 3: Assign each data point to the closer of p_1 or p_2 based on their distances, resulting in two subsets $subset_1$ and $subset_2$;
- 4: Recursively repeat steps 1 to 3 on $subset_1$ and $subset_2$ until the tree is fully constructed.

Time complexity: $O(N \log(N)) < O(N^2)!$

Original Point Cloud



Sorted Point Cloud

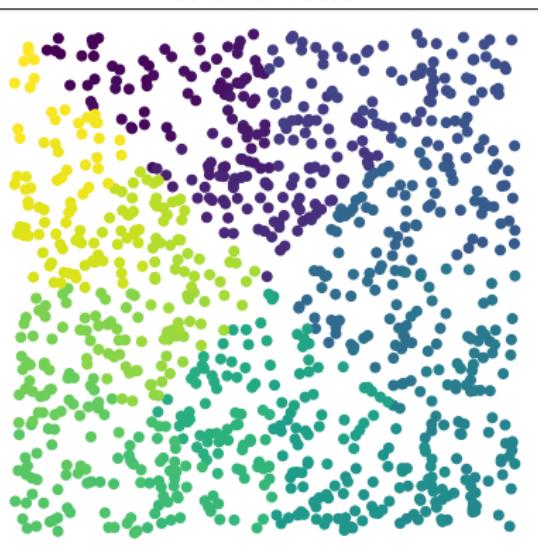


Fig. 1. A simulation in the unit square $[0, 1] \times [0, 1]$. Uniformly distributed points are sorted using the Ball Tree algorithm, with lighter colors indicating earlier positions and darker colors later positions in the ordering.

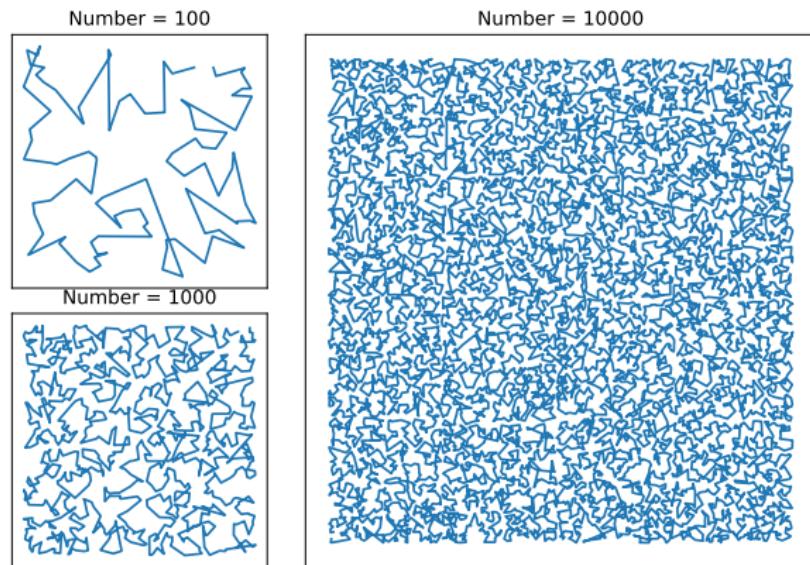


Fig. 2. Lines connecting points ordered by the Ball Tree algorithm, showing results for 100, 1,000, and 10,000 points. The resulting pattern exhibits a space-filling effect similar to a Hilbert curve.

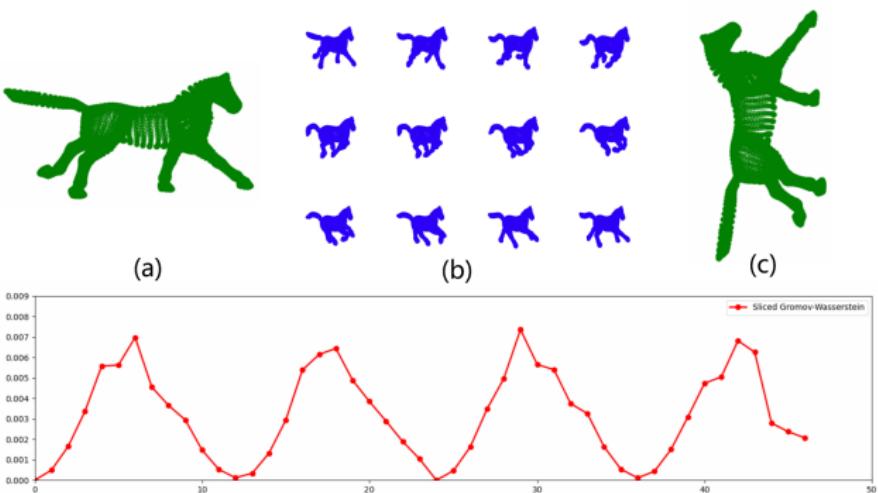


Fig. 3. Horse demo provided by [Vayer et al.(2019)]. Comparison of SGW loss between a static horse(a) and a running horse(b).

Building upon this demo, we further analyze the GW/OT loss between a rotated horse (c) and the running horse (b).

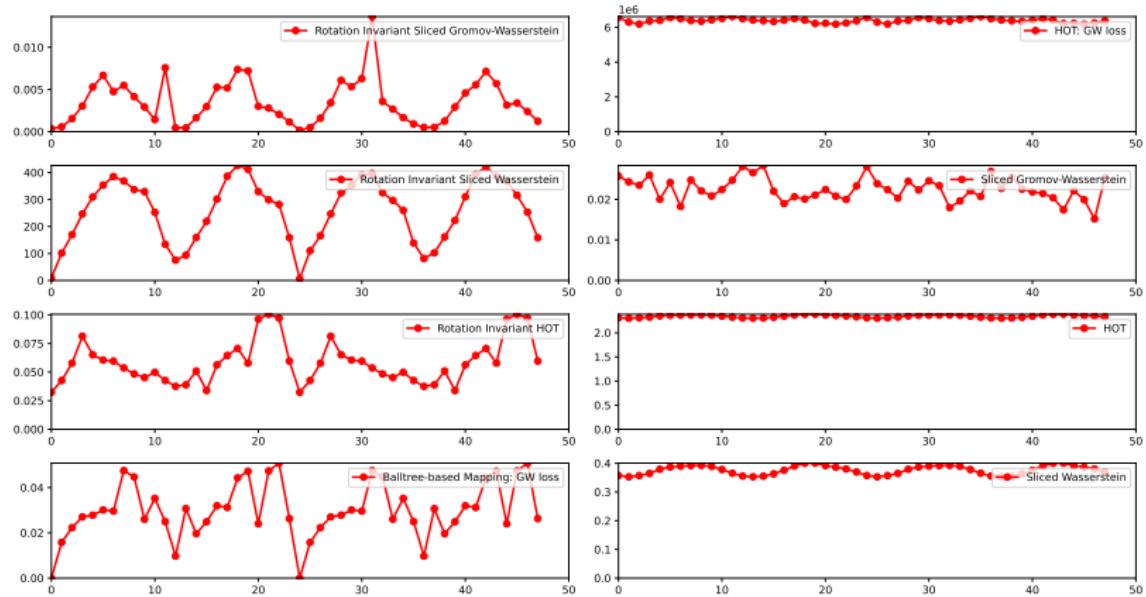


Fig. 4. Results of eight methods applied to the horse demo. Only the methods incorporating the Rotation Invariant module and the Ball Tree-based method exhibit significant fluctuations.

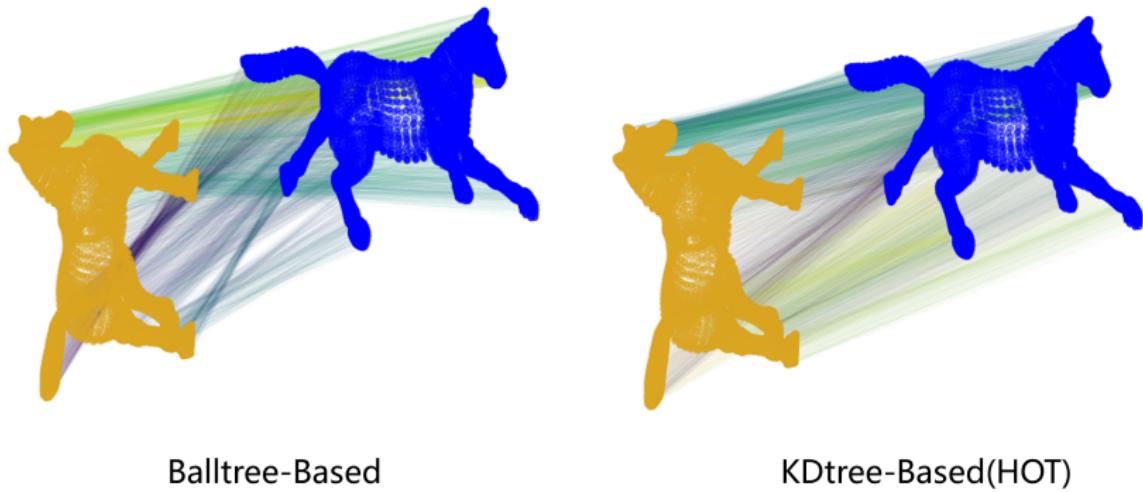


Fig. 5. Comparison of couplings obtained from two tree-based methods.

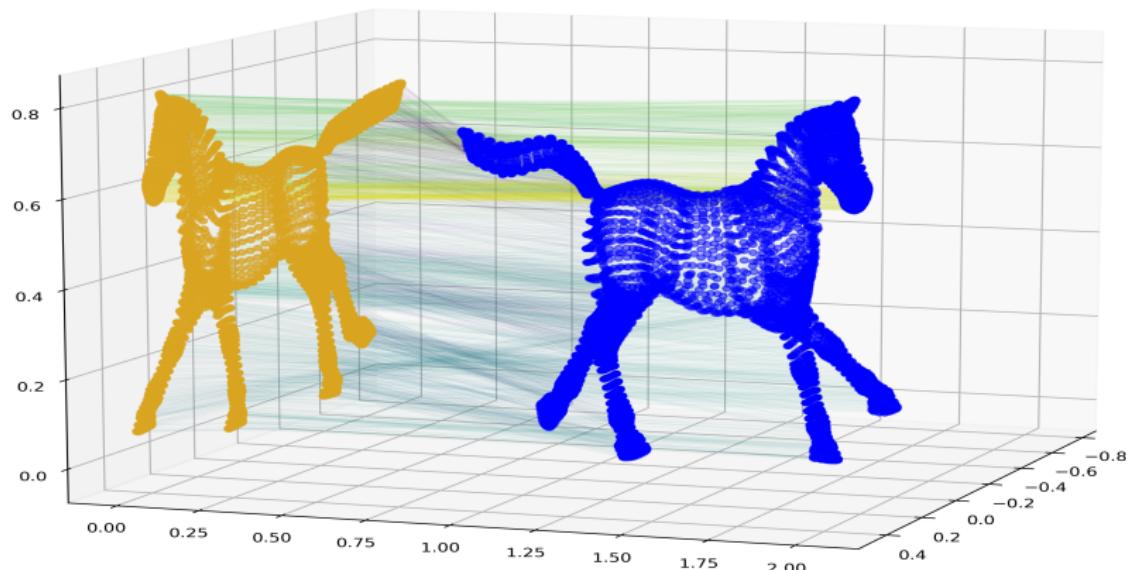


Fig. 6. Coupling between a 2D horse (yellow) and a 3D horse (blue) by the Ball Tree algorithm.

References |

- [Bonneel et al.(2015)] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. 2015.
Sliced and radon wasserstein barycenters of measures.
Journal of Mathematical Imaging and Vision 51 (2015), 22–45.
- [Dumont et al.(2024)] Théo Dumont, Théo Lacombe, and François-Xavier Vialard. 2024.
On the existence of Monge maps for the Gromov–Wasserstein problem.
Foundations of Computational Mathematics (2024), 1–48.

References II

[Li et al.(2024)] Tao Li, Cheng Meng, Hongteng Xu, and Jun Yu. 2024.

Hilbert curve projection distance for distribution comparison.
IEEE Transactions on Pattern Analysis and Machine Intelligence (2024).

[Mémoli(2011)] Facundo Mémoli. 2011.

Gromov–Wasserstein distances and the metric approach to object matching.
Foundations of computational mathematics 11 (2011), 417–487.

References III

[Rioux et al.(2023)] Gabriel Rioux, Ziv Goldfeld, and Kengo Kato. 2023.

Entropic Gromov-Wasserstein Distances: Stability and Algorithms.

arXiv preprint arXiv:2306.00182 (2023).

[Vayer et al.(2019)] Titouan Vayer, Rémi Flamary, Romain Tavenard, Laetitia Chapel, and Nicolas Courty. 2019.

Sliced gromov-wasserstein.

arXiv preprint arXiv:1905.10124 (2019).

References IV

[Zhang et al.(2024)] Zhengxin Zhang, Ziv Goldfeld, Youssef Mroueh, and Bharath K Sriperumbudur. 2024.
Gromov–Wasserstein distances: Entropic regularization, duality and sample complexity.
The Annals of Statistics 52, 4 (2024), 1616–1645.

THANK YOU!