Leveraging Optimal Transport via Projections on Subspaces for Machine Learning Applications

Clément Bonet

Under the supervision of François Septier, Nicolas Courty and Lucas Drumetz

PhD Defense

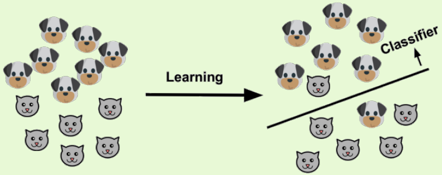
16/11/2023

Machine Learning

Goal: learn a model from data

Example

- Classification



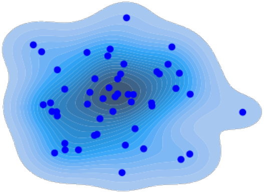
From (Goyal, 2018)

- Generative model

Samples from Stable Diffusion (Rombach et al., 2022)

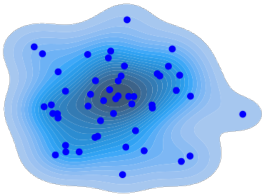
Probability Distributions

- Data: $x_1, \ldots, x_n \in \mathbb{R}^d \longleftrightarrow$ probability distribution $\mu_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$

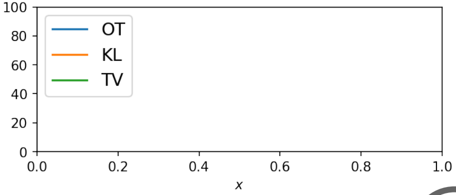


Probability Distributions

- Data: $x_1, \ldots, x_n \in \mathbb{R}^d \longleftrightarrow$ probability distribution $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$



- **Goals**:
  - Compare distributions using some discrepancy D
  - Learn distributions by minimizing D (*e.g.* for generative models)



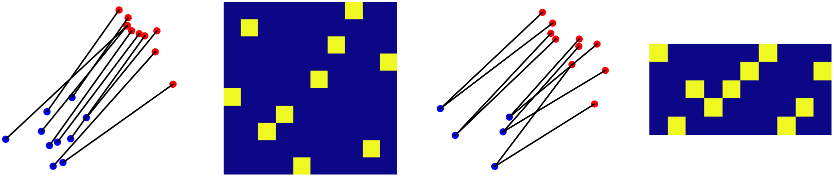
Optimal Transport

Kantorovich Problem

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$,

$$\mathrm{OT}_c(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int c(x, y) \, \mathrm{d}\gamma(x, y),$$

$$\Pi(\mu, \nu) = \left\{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d), \ \forall A \in \mathcal{B}(\mathbb{R}^d), \ \gamma(A \times \mathbb{R}^d) = \mu(A), \ \gamma(\mathbb{R}^d \times A) = \nu(A)\right\}$$



Optimal Transport

Wasserstein Distance

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$W_2^2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int \|x - y\|_2^2 \, \mathrm{d}\gamma(x, y)$$

Properties:

- W_2 distance
- Metrizes the weak convergence

Condition to have a deterministic coupling, *i.e.* $\gamma = (\mathrm{Id}, T)_\# \mu$ with $T_\# \mu = \nu$ where $\forall A \in \mathcal{B}(\mathbb{R}^d),\ T_\# \mu(A) = \mu\big(T^{-1}(A)\big)$: **Brenier's theorem** (Brenier, 1991)

$\mu \ll \mathrm{Leb} \implies$ Optimal coupling γ^* unique and $\gamma^* = (\mathrm{Id}, \nabla\varphi)_\# \mu$ with φ convex

Solving the OT Problem

Let $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbb{R}^d$, $\alpha, \beta \in \Sigma_n$, $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$, $\nu = \sum_{i=1}^n \beta_i \delta_{y_i}$,

$$W_2^2(\mu, \nu) = \min_{P \in \mathbb{R}_+^{n \times n},\ P\mathbb{1}_n = \alpha,\ P^T \mathbb{1}_n = \beta} \langle C, P \rangle_F \quad \text{with} \quad C = \left(\|x_i - y_j\|_2^2\right)_{i,j}$$

Computational Complexity (Pele and Werman, 2009)

Numerical computation: **Linear program** in $O(n^3 \log n)$

Sample Complexity (Boissard and Le Gouic, 2014)

For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $x_1, \ldots, x_n \sim \mu$, $y_1, \ldots, y_n \sim \nu$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$,

$$\mathbb{E}\left[\left|W_2(\hat{\mu}_n, \hat{\nu}_n) - W_2(\mu, \nu)\right|\right] = O(n^{-1/d})$$

Solving the OT Problem

Let $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}^d$, $\alpha, \beta \in \Sigma_n$, $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$, $\nu = \sum_{i=1}^n \beta_i \delta_{y_i}$,

$$W_2^2(\mu, \nu) = \min_{P \in \mathbb{R}_+^{n \times n},\ P\mathbb{1}_n = \alpha,\ P^T \mathbb{1}_n = \beta} \langle C, P \rangle_F \quad \text{with} \quad C = \left(\|x_i - y_j\|_2^2\right)_{i,j}$$

Computational Complexity (Pele and Werman, 2009)

Numerical computation: **Linear program** in $O(n^3 \log n)$

Sample Complexity (Boissard and Le Gouic, 2014)

For $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, $x_1, \dots, x_n \sim \mu$, $y_1, \dots, y_n \sim \nu$, $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ and $\hat{\nu}_n = \frac{1}{n} \sum_{i=1}^n \delta_{y_i}$,

$$\mathbb{E}\left[|W_2(\hat{\mu}_n, \hat{\nu}_n) - W_2(\mu, \nu)|\right] = O(n^{-1/d})$$

Proposed solutions:

- Entropic regularization + Sinkhorn (Cuturi, 2013)
- Minibatch estimator (Fatras et al., 2020)
- Sliced-Wasserstein (Rabin et al., 2011b; Bonnotte, 2013)

1D OT Problem

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R})$,

- Cumulative distribution function:

$$\forall t \in \mathbb{R},\ F_\mu(t) = \mu\big(]-\infty, t]\big) = \int \mathbb{1}_{]-\infty,t]}(x)\ \mathrm{d}\mu(x)$$

- Quantile function:

$$\forall u \in [0,1],\ F_\mu^{-1}(u) = \inf\{x \in \mathbb{R},\ F_\mu(x) \geq u\}$$

1D Wasserstein Distance

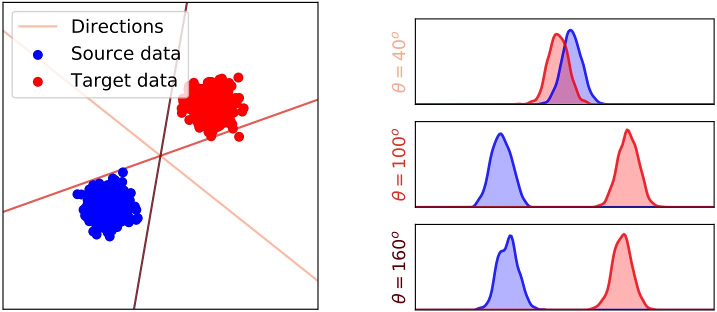
$$W_2^2(\mu,\nu) = \int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^2\ \mathrm{d}u = \|F_\mu^{-1} - F_\nu^{-1}\|_{L^2([0,1])}^2$$

Let $x_1 < \dots < x_n,\ y_1 < \dots < y_n,\ \mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i},\ \nu = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$,

$$W_2^2(\mu,\nu) = \frac{1}{n}\sum_{i=1}^n (x_i - y_i)^2$$

$\rightarrow O(n \log n)$

Sliced-Wasserstein Distance



Definition (Sliced-Wasserstein (Rabin et al., 2011b))

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\mathrm{SW}_2^2(\mu, \nu) = \int_{S^{d-1}} W_2^2(P^\theta_\# \mu, P^\theta_\# \nu)\ \mathrm{d}\lambda(\theta),$$

where $P^\theta(x) = \langle x, \theta \rangle$, λ uniform measure on S^{d-1}.

Properties of the Sliced-Wasserstein Distance

Let $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}^d$, $\alpha, \beta \in \Sigma_n$, $\mu = \sum_{i=1}^n \alpha_i \delta_{x_i}$, $\nu = \sum_{i=1}^n \beta_i \delta_{y_i}$.

Approximation via Monte-Carlo:

$$\widehat{\mathrm{SW}}_{2,L}^2(\mu, \nu) = \frac{1}{L} \sum_{\ell=1}^L W_2^2(P_\#^{\theta_\ell}\mu, P_\#^{\theta_\ell}\nu),$$

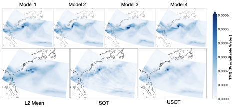
$\theta_1, \dots, \theta_L \sim \lambda$.

Properties:

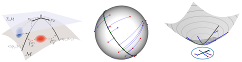
- Computational complexity: $L \cdot O\big(\mathrm{sort}(n)\big) + Ln \cdot O\big(\mathrm{projection}(d)\big)$
- Sample complexity: independent of the dimension (Nadjahi et al., 2020)
- SW_2 distance (Bonnotte, 2013)
- Topologically equivalent to the Wasserstein distance (Nadjahi et al., 2019), *i.e.* $\lim_{n\to\infty} \mathrm{SW}_2^2(\mu_n, \mu) = 0 \iff \lim_{n\to\infty} W_2^2(\mu_n, \mu) = 0$.
- Differentiable, Hilbertian

Summary of the Contributions

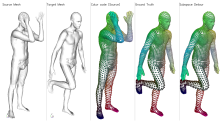
Sliced Unbalanced OT



SW on Manifolds

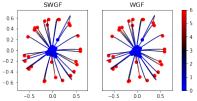


Subspace Detours for GW

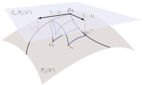


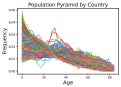
OT via Projections

SW Gradient Flows



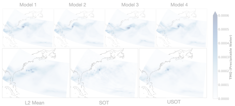
PCA on Wasserstein Space



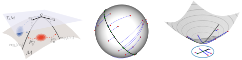


Summary of the Contributions

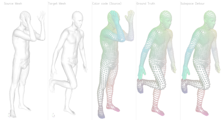
Sliced Unbalanced OT



SW on Manifolds

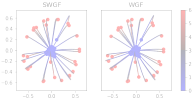


Subspace Detours for GW



OT via Projections

SW Gradient Flows



PCA on Wasserstein Space

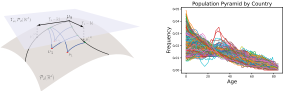


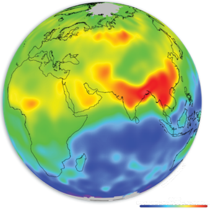
Table of Contents

Riemannian Manifolds in Machine Learning

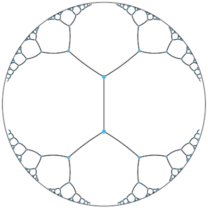
Data often lie on manifolds or have an underlying structure which can be captured on manifolds.

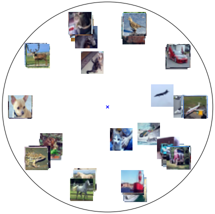
Example

- Directional data, Earth data, cyclic data on the sphere S^{d-1}
- Hierarchical data (trees, graphs, words, images) on Hyperbolic spaces
- M/EEG data on the space of Symmetric Positive Definite Matrices (SPDs)



Source: ESA





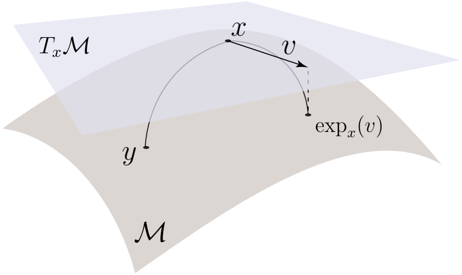
Riemannian Manifolds

Definition

A Riemannian manifold $(\mathcal{M}, g)$ of dimension d is a space that behaves locally as a linear space diffeomorphic to $\mathbb{R}^d$.

Properties:

- To any $x \in \mathcal{M}$, associate a tangent space $T_x\mathcal{M}$ with a smooth inner product $\langle \cdot, \cdot \rangle_x : T_x\mathcal{M} \times T_x\mathcal{M} \to \mathbb{R}$.
- Geodesic between x and y: shortest path minimizing the length $\mathcal{L}$
- Geodesic distance: $d(x, y) = \inf_{\gamma} \mathcal{L}(\gamma)$
- Exponential map: $\forall x \in \mathcal{M},\ \exp_x : T_x\mathcal{M} \to \mathcal{M}$



Cartan-Hadamard Manifolds

Particular case of Riemannian manifold: **Cartan-Hadamard** manifolds $(\mathcal{M}, g)$

Definition: Non-positive curvature, complete and connected

Properties:

- Geodesically complete: Any geodesic $\gamma : [0,1] \to \mathcal{M}$ between $x \in \mathcal{M}$ and $y \in \mathcal{M}$ can be extended to $\mathbb{R}$
- For any $x \in \mathcal{M}$, $\exp_x : T_x\mathcal{M} \to \mathcal{M}$ diffeomorphism

Example

- Euclidean spaces
- Hyperbolic spaces (Nickel and Kiela, 2017, 2018; Khrulkov et al., 2020)
- SPDs endowed with specific metrics (Sabbagh et al., 2019, 2020; Pennec, 2020)
- Product of Cartan-Hadamard manifolds (Gu et al., 2019; Skopek et al., 2019)

Hyperbolic Space

Hyperbolic space: Riemannian manifold of constant negative curvature

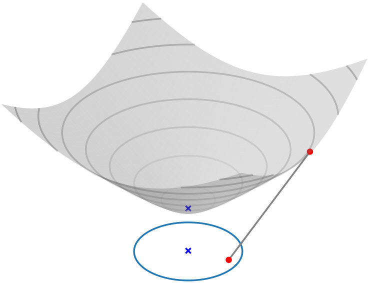
Different isometric models:

- **Lorentz model** $\mathbb{L}^d = \big\{(x_0, \dots, x_d) \in \mathbb{R}^{d+1},\ \langle x, x\rangle_{\mathbb{L}} = -1, x_0 > 0\big\}$,

$$d_{\mathbb{L}}(x, y) = \operatorname{arccosh}(-\langle x, y\rangle_{\mathbb{L}}), \quad \langle x, y\rangle_{\mathbb{L}} = -x_0 y_0 + \sum_{i=1}^{d} x_i y_i$$

- **Poincaré ball** $\mathbb{B}^d = \{x \in \mathbb{R}^d,\ \|x\|_2 < 1\}$,

$$d_{\mathbb{B}}(x, y) = \operatorname{arccosh}\left(1 + 2\frac{\|x - y\|_2^2}{(1 - \|x\|_2^2)(1 - \|y\|_2^2)}\right)$$



Optimal Transport on Riemannian Manifolds

Let $(\mathcal{M}, g)$ be a Riemannian manifold, d its geodesic distance.

Definition (Wasserstein distance)

Let $\mu, \nu \in \mathcal{P}_2(\mathcal{M})$, then

$$W_2^2(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int d(x, y)^2 \, \mathrm{d}\gamma(x, y)$$

In practice: same drawbacks of the Euclidean case.

SW on Cartan-Hadamard Manifolds

Goal: defining SW discrepancy on Cartan-Hadamard manifolds taking care of geometry of the manifold

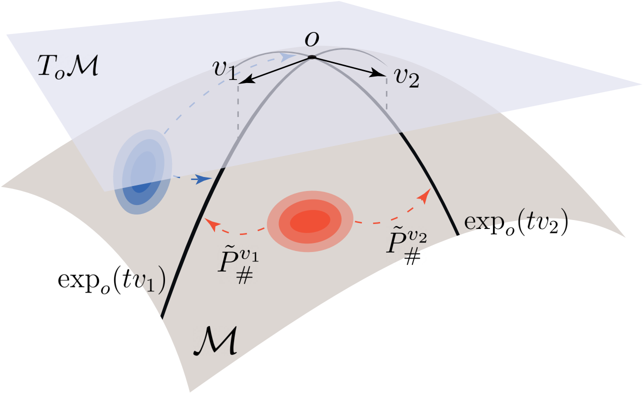
	SW	CHSW
Closed-form of W	Line	?
Projection	$P^\theta(x) = \langle x, \theta \rangle$	?
Integration	S^{d-1}	?

Projecting on Geodesics

- Generalization of straight lines on manifolds: **geodesics**

$$\forall v \in T_o\mathcal{M},\ \mathcal{G}^v = \{\exp_o(tv),\ t \in \mathbb{R}\}$$

- Geodesics isometric to $\mathbb{R}$
- Integrate along all possible directions on $S_o = \{v \in T_o\mathcal{M},\ \|v\|_o = 1\}$



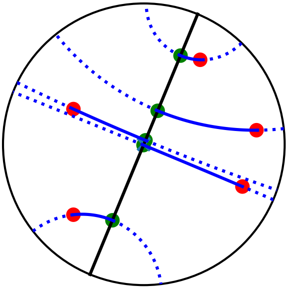
Projections

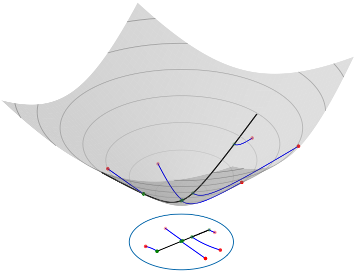
1. **Geodesic projections**:
   - On Euclidean space: For $\theta \in S^{d-1}$, $\mathcal{G}_\theta = \{t\theta,\ t \in \mathbb{R}\}$,

$$\forall x \in \mathbb{R}^d,\ P^\theta(x) = \langle x, \theta \rangle = \operatorname*{argmin}_{t \in \mathbb{R}}\ \|x - t\theta\|_2$$

   - On Cartan-Hadamard manifold: For $v \in T_o\mathcal{M}$, $\mathcal{G}_v = \{\exp_o(tv),\ t \in \mathbb{R}\}$,

$$\forall x \in \mathcal{M},\ P^v(x) = \operatorname*{argmin}_{t \in \mathbb{R}}\ d\big(x, \exp_o(tv)\big)$$





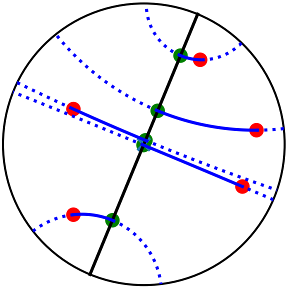
Projections

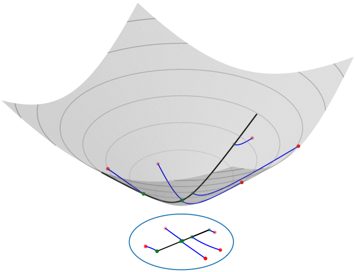
1. **Geodesic projections**:
   - On Euclidean space: For $\theta \in S^{d-1}$, $\mathcal{G}_\theta = \{t\theta,\ t \in \mathbb{R}\}$, $\exp_0(t\theta) = 0 + t\theta = t\theta$,

$$\forall x \in \mathbb{R}^d,\ P^\theta(x) = \langle x, \theta \rangle = \operatorname*{argmin}_{t\in\mathbb{R}}\ \|x - t\theta\|_2 = \operatorname*{argmin}_{t\in\mathbb{R}}\ d\big(x, \exp_0(t\theta)\big)$$

   - On Cartan-Hadamard manifold: For $v \in T_o\mathcal{M}$, $\mathcal{G}_v = \{\exp_o(tv),\ t \in \mathbb{R}\}$,

$$\forall x \in \mathcal{M},\ P^v(x) = \operatorname*{argmin}_{t\in\mathbb{R}}\ d\big(x, \exp_o(tv)\big)$$



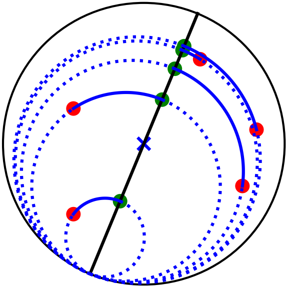


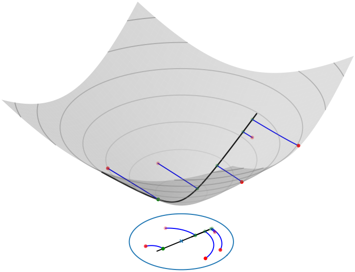
Projections

1. **Geodesic projections**: $\forall x \in \mathcal{M},\ P^v(x) = \operatorname*{argmin}_{t\in\mathbb{R}}\ d\big(x, \exp_o(tv)\big)$
2. **Horospherical projections**: following level sets of the Busemann function

$$B^\gamma(x) = \lim_{t\to\infty}\ d\big(x, \gamma(t)\big) - t$$

- On Euclidean space: $B^\theta(x) = -\langle x, \theta\rangle$
- On Cartan-Hadamard manifold: $B^v(x) = \lim_{t\to\infty}\ d\big(x, \exp_o(tv)\big) - t$





Cartan-Hadamard Sliced-Wassertein

Let $(\mathcal{M}, g)$ a Hadamard manifold with o its origin. Denote λ the uniform distribution on $S_o = \{v \in T_o\mathcal{M},\ \|v\|_o = 1\}$.

Geodesic-Cartan Hadamard Sliced-Wasserstein

$$\forall \mu, \nu \in \mathcal{P}_2(\mathcal{M}),\ \mathrm{GCHSW}_2^2(\mu, \nu) = \int_{S_o} W_2^2(P^v_\# \mu, P^v_\# \nu)\ \mathrm{d}\lambda(v)$$

Horospherical-Cartan Hadamard Sliced-Wasserstein

$$\forall \mu, \nu \in \mathcal{P}_2(\mathcal{M}),\ \mathrm{HCHSW}_2^2(\mu, \nu) = \int_{S_o} W_2^2(B^v_\# \mu, B^v_\# \nu)\ \mathrm{d}\lambda(v)$$

$\mathrm{CHSW} = \mathrm{GCHSW}$ or HCHSW

General Properties

Some properties:

- Pseudo distance on $\mathcal{P}_2(\mathcal{M})$ $\to$ open question: distance?
- $\forall \mu, \nu \in \mathcal{P}_2(\mathcal{M}),\ \mathrm{CHSW}_2^2(\mu,\nu) \le W_2^2(\mu,\nu)$
- Sample complexity independent of the dimension
- Computational complexity: $L \cdot O\big(\mathrm{sort}(n)\big) + Ln \cdot O\big(\mathrm{projection}(d)\big)$
- CHSW_2 is Hilbertian

Proposition

Define $K : \mathcal{P}_2(\mathcal{M}) \times \mathcal{P}_2(\mathcal{M}) \to \mathbb{R}$ *as* $K(\mu,\nu) = \exp\big(-\gamma \mathrm{CHSW}_2^2(\mu,\nu)\big)$ *for* $\gamma > 0$*. Then* K *is a positive definite kernel.*

Proposition

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{B}^d)$ *and denote* $\tilde{\mu} = (P_{\mathbb{B}\to\mathbb{L}})_\#\mu$, $\tilde{\nu} = (P_{\mathbb{B}\to\mathbb{L}})_\#\nu$*. Then,*

$$\mathrm{HHSW}_2^2(\mu,\nu) = \mathrm{HHSW}_2^2(\tilde{\mu},\tilde{\nu}),$$
$$\mathrm{GHSW}_2^2(\mu,\nu) = \mathrm{GHSW}_2^2(\tilde{\mu},\tilde{\nu}).$$

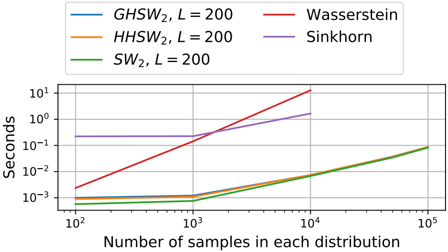
Runtime and Complexity (Bonet et al., 2023c)

Closed-forms for P^v and B^v on $\mathbb{B}^d$ and $\mathbb{L}^d$:

$$\forall v \in T_{x^0}\mathbb{L}^d \cap S^d,\ x \in \mathbb{L}^d,$$
$$P^v(x) = \operatorname{arctanh}\left(-\frac{\langle x, v\rangle_{\mathbb{L}}}{\langle x, x^0\rangle_{\mathbb{L}}}\right)$$
$$B^v(x) = \log\big(-\langle x, x^0 + v\rangle_{\mathbb{L}}\big)$$

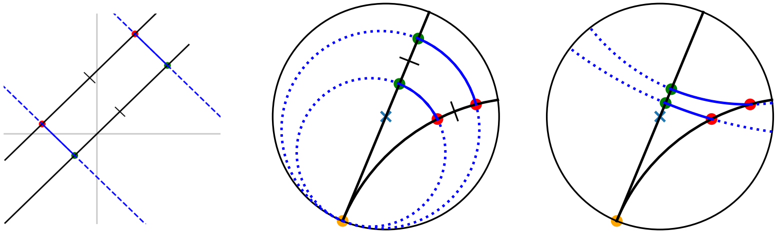
$$\forall \tilde{v} \in S^{d-1},\ y \in \mathbb{B}^d,$$
$$P^{\tilde{v}}(y) = 2\operatorname{arctanh}\big(s(y)\big)$$
$$B^{\tilde{v}}(y) = \log\left(\frac{\|\tilde{v} - y\|_2^2}{1 - \|y\|_2^2}\right)$$

Method	Complexity
Wasserstein + LP	$O(n^3 \log n + n^2 d)$
Sinkhorn	$O(n^2 d)$
SW	$O\big(Ln(d + \log n)\big)$
GHSW	$O\big(Ln(d + \log n)\big)$
HHSW	$O\big(Ln(d + \log n)\big)$



Comparison of the Projections

- Property of the Horospherical projection: conserves the distance between points on a parallel geodesic (Chami et al., 2021)

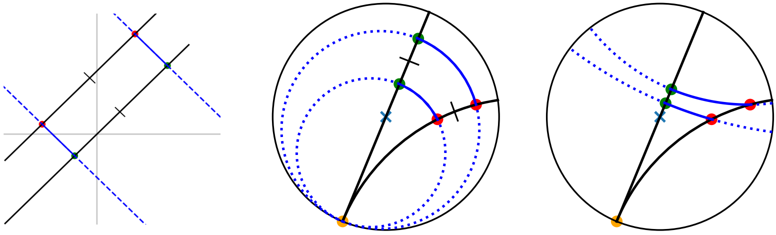


Horospherical projection

Geodesic projection

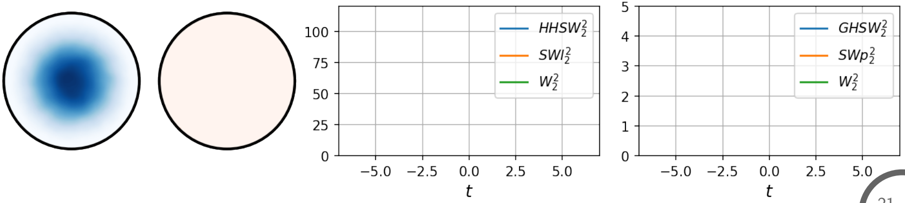
Comparison of the Projections

- Property of the Horospherical projection: conserves the distance between points on a parallel geodesic (Chami et al., 2021)



Horospherical projection   Geodesic projection

- Let $\mu = \mathrm{WND}(0, I_d)$, $\nu_t = \mathrm{WND}(x_t, I_d)$,



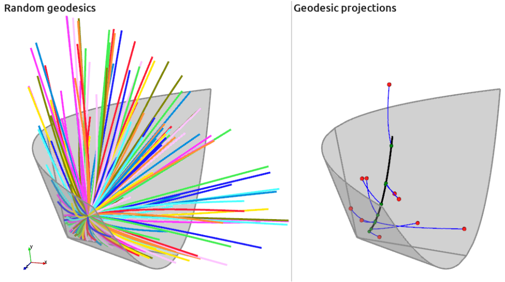
Manifold of SPD Matrices (Bonet et al., 2023a)

- **Symmetric Positive Definite (SPD) Matrices**:

$$S_d^{++}(\mathbb{R}) = \left\{M \in S_d(\mathbb{R}),\ \forall x \in \mathbb{R}^d \setminus \{0\},\ x^T M x > 0\right\}$$

- Log-Euclidean distance: $\forall X, Y \in S_d^{++}(\mathbb{R}),\ d_{LE}(X,Y) = \|\log X - \log Y\|_F$
- Tangent space: $T_{I_d} S_d^{++}(\mathbb{R}) \cong S_d(\mathbb{R})$
- Projection on geodesics $\mathcal{G}_A = \{\exp(tA),\ t \in \mathbb{R}\}$ for $A \in S_{I_d}$:

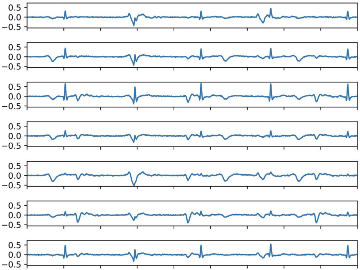
$$\forall M \in S_d^{++}(\mathbb{R}),\ P^A(M) = -B^A(M) = \langle A, \log M\rangle_F$$



M/EEG data (with Benoît Malézieux)

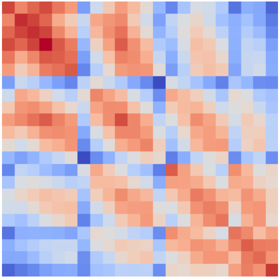
M/EEG data:

- Recorded from the brain
- Multivariate time series $X \in \mathbb{R}^{N \times T}$
- Transform X into SPDs



Data X with T time samples

$\longrightarrow$

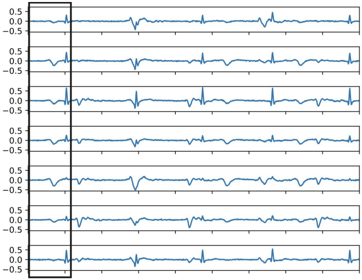


SPD matrix

M/EEG data (with Benoît Malézieux)

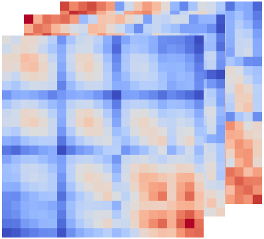
M/EEG data:

- Recorded from the brain
- Multivariate time series $X \in \mathbb{R}^{N \times T}$
- Transform X into distribution of SPDs



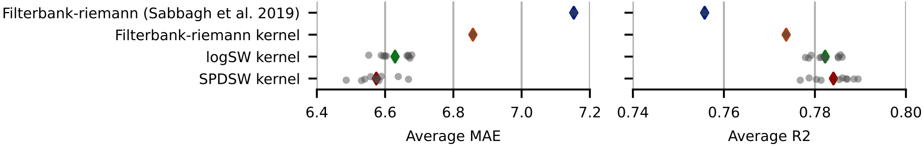
Data X with T time samples

$\longrightarrow$



Distribution of SPD matrices

Brain-Age Prediction



Positive definite Gaussian Kernel with SPDSW

$$K(\mu,\nu) = e^{-\gamma \mathrm{SPDSW}_2^2(\mu,\nu)} = e^{-\gamma \|\Phi(\mu)-\Phi(\nu)\|_{\mathcal{H}}^2}$$

Known feature map Φ, no need for expensive quadratic computations

$\rightarrow$ **Kernel Ridge** regression

Conclusion

Conclusion:

- General formulation of SW on Cartan-Hadamard manifolds
- Specification to Hyperbolic Spaces and SPDs with Log-Euclidean metric
- Applications to ML

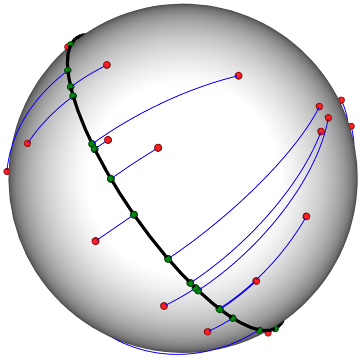
Publications:

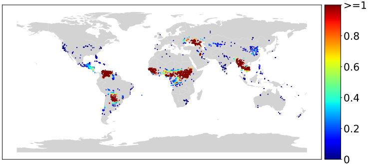
- Clément Bonet, Laetitia Chapel, Lucas Drumetz, and Nicolas Courty. *Hyperbolic Sliced-Wasserstein via Geodesic and Horospherical Projections*. Annual Workshop on Topology, Algebra, and Geometry in Machine Learning, 2023.
- Clément Bonet, Benoît Malézieux, Alain Rakotomamonjy, Lucas Drumetz, Thomas Moreau, Matthieu Kowalski, and Nicolas Courty. *Sliced-Wasserstein on Symmetric Positive Definite Matrices for M/EEG Signals*. International Conference of Machine Learning, 2023.
- Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. *Spherical Sliced-Wasserstein*. International Conference on Learning Representations, 2023.

Perspectives

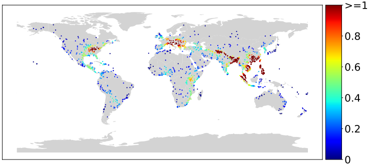
Other type of manifolds?

- Sphere (Bonet et al., 2023b)
  - Not a Cartan-Hadamard manifold (positively curved)
  - Geodesics are great circles
  - Dedicated algorithm to compute Wasserstein on the circle (Delon et al., 2010)
  - Application to Geophysical data





Fire



Flood

Perspectives

Other type of manifolds?

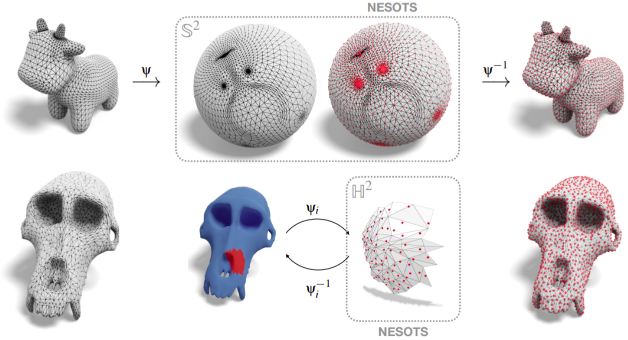
- Sphere (Bonet et al., 2023b)
- Manifolds of non-constant curvature
  - SPD Matrices with Affine-Invariant metric (Pennec, 2020)
  - Siegel space (Cabanes and Nielsen, 2021)
- Manifolds with no closed-forms for the projections
- Finsler manifolds (López et al., 2021a,b; Pouplin et al., 2023)

Perspectives

Other type of manifolds?

- Sphere (Bonet et al., 2023b)
- Manifolds of non-constant curvature
- Manifolds with no closed-forms for the projections
- Finsler manifolds (López et al., 2021a,b; Pouplin et al., 2023)

Other applications to ML



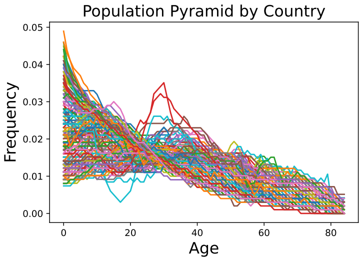
Non-Euclidean Sliced Optimal Transport Sampling (Genest, Courty, and Coeurjolly, 2023)

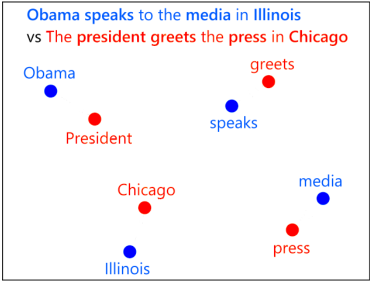
Table of Contents

Datasets of Distributions

Example

- Histograms (*e.g.* age distributions of countries, financial assets...)
- Documents: distributions of words (Kusner et al., 2015)
- Cells: distributions of genes (Bellazzi et al., 2021)
- Embedding of words in Gaussian distributions (Vilnis and McCallum, 2015)





→ Perform dimensionality reduction to identify modes of variations
e.g. PCA on Wasserstein space (Seguy and Cuturi, 2015; Cazelles et al., 2018)

Busemann in Wasserstein Space

Goal: compute and use

$$\forall \nu \in \mathcal{P}_2(\mathbb{R}^d),\ B^\mu(\nu) = \lim_{t\to\infty}\ W_2^2(\mu_t, \nu) - t$$

$\rightarrow$ provides a way to project distributions on lower-dimensional spaces

Contributions:

- Identify for which geodesics B^μ is well defined
- Derive closed-form in 1D and between Gaussians
- Application to PCA

Geodesics in Wasserstein Space

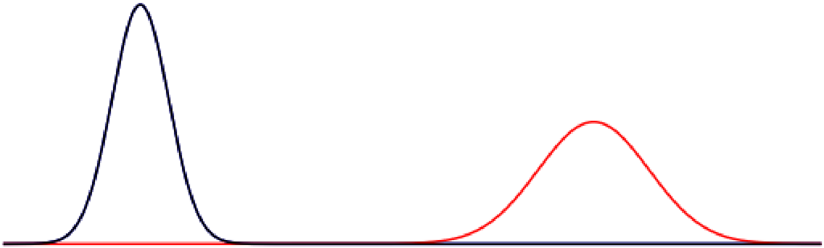
Geodesics

- For $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, geodesic = displacement interpolation:

$$\forall t \in [0,1],\ \mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\# \gamma \quad \text{for} \quad \gamma \in \Pi_o(\mu, \nu)$$

- If $\mu_1 = T_\# \mu_0$ with T the OT map,

$$\forall t \in [0,1],\ \mu_t = \big((1-t)\mathrm{Id} + tT\big)_\# \mu_0$$



Geodesics in Wasserstein Space

Geodesics

- For $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, geodesic = displacement interpolation:

$$\forall t \in [0,1],\ \mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\# \gamma \quad \text{for} \quad \gamma \in \Pi_o(\mu, \nu)$$

- If $\mu_1 = T_\# \mu_0$ with T the OT map,

$$\forall t \in [0,1],\ \mu_t = \big((1-t)\mathrm{Id} + tT\big)_\# \mu_0$$

- Busemann function well defined for geodesic rays, *i.e.* such that $t \mapsto \mu_t$ defined on $[0, +\infty[$
- $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ positively curved

$\rightarrow$ Conditions to extend $t \mapsto \mu_t$ to $\mathbb{R}_+$?

Geodesic Rays in Wasserstein Space

Proposition

Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ with μ_0 absolutely continuous with respect to the Lebesgue measure and consider $c(x,y) = \frac{1}{2}\|x-y\|_2^2$. Then, the optimal transport map T between μ_0 and μ_1 is the gradient of a 1-convex function u if and only if $\mu_t = \big((1-t)\mathrm{Id} + tT\big)_{\#}\mu_0$ is a geodesic ray.

Example

- 1D Gaussian: Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$, $T(x) = \frac{\sigma_1}{\sigma_0}(x - m_0) + m_1$

$$T'(x) - 1 \geq 0 \iff \sigma_1 \geq \sigma_0$$

- General Gaussian: Let $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$, $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$, $T(x) = A(x - m_0) + m_1$ with $A = \Sigma_0^{-\frac{1}{2}}\big(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\big)^{\frac{1}{2}}\Sigma_0^{-\frac{1}{2}}$,

$$\nabla T(x) - I_d \succeq 0 \iff \big(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}}\big)^{\frac{1}{2}} \succeq \Sigma_0$$

Geodesic Rays in 1D Wasserstein Space

For $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$, quantile F_t^{-1} of the geodesic $t \mapsto \mu_t$ characterized as

$$\forall t \in [0,1],\ F_t^{-1} = (1-t)F_0^{-1} + tF_1^{-1}.$$

Proposition (Kloeckner, 2010)

Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$ and denote F_0^{-1}, F_1^{-1} their quantile functions. Denote for any $t \in [0,1]$, $\mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\# \gamma$ with $\gamma = (F_0^{-1}, F_1^{-1})_\# \mathrm{Unif}([0,1])$ the optimal coupling between μ_0 and μ_1. Then, $t \mapsto \mu_t$ is a geodesic ray if and only if $F_1^{-1} - F_0^{-1}$ is non-decreasing.

Example

- 1D Gaussian: $F_1^{-1} - F_0^{-1}$ non-decreasing iff $\sigma_1 \geq \sigma_0$
- For $\mu_0 = \delta_{x_0}$ and $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $t \mapsto \mu_t$ geodesic ray

Closed-Forms of the Busemann Function

Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R})$,

$$W_2^2(\mu_0, \mu_1) = \int_0^1 |F_{\mu_0}^{-1}(u) - F_{\mu_1}^{-1}(u)|^2 \, \mathrm{d}u = \|F_{\mu_0}^{-1} - F_{\mu_1}^{-1}\|_{L^2([0,1])}^2$$

Proposition (Closed-from for the Busemann function on $\mathcal{P}_2(\mathbb{R})$)

Let $(\mu_t)_{t\geq 0}$ be a unit-speed geodesic ray in $\mathcal{P}_2(\mathbb{R})$, then

$$\begin{aligned}\forall \nu \in \mathcal{P}_2(\mathbb{R}),\ B^\mu(\nu) &= -\int_0^1 \big(F_{\mu_1}^{-1}(u) - F_{\mu_0}^{-1}(u)\big)\big(F_{\nu}^{-1}(u) - F_{\mu_0}^{-1}(u)\big) \, \mathrm{d}u \\ &= -\langle F_{\mu_1}^{-1} - F_{\mu_0}^{-1}, F_{\nu}^{-1} - F_{\mu_0}^{-1}\rangle_{L^2([0,1])}.\end{aligned}$$

Closed-Forms of the Busemann Function

Let $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$, $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$,

$$W_2^2(\mu_0, \mu_1) = \|m_0 - m_1\|_2^2 + \mathrm{Tr}\left(\Sigma_0 + \Sigma_1 - 2(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}})^{\frac{1}{2}}\right)$$

Proposition (Closed-form for the Busemann function on $BW(\mathbb{R}^d)$)

Let $(\mu_t)_{t\geq 0}$ be a unit-speed geodesic ray characterized by $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$ and $\mu_1 = \mathcal{N}(m_1, \Sigma_1)$. Then, for any $\nu = \mathcal{N}(m, \Sigma)$,

$$\begin{aligned} B^\mu(\nu) = -\langle m_1 - m_0, m - m_0\rangle + \mathrm{Tr}\big(\Sigma_0(A - I_d)\big) \\ - \mathrm{Tr}\big((\Sigma^{\frac{1}{2}}(\Sigma_0 - \Sigma_0 A - A\Sigma_0 + \Sigma_1)\Sigma^{\frac{1}{2}})^{\frac{1}{2}}\big), \end{aligned}$$

where $A = \Sigma_0^{-\frac{1}{2}}(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}})^{\frac{1}{2}}\Sigma_0^{-\frac{1}{2}}$.

Application to PCA

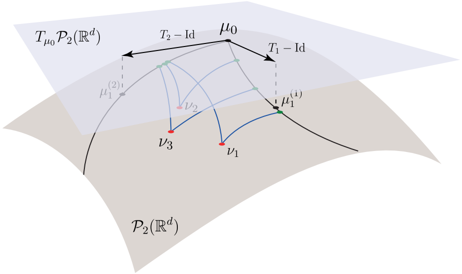
Euclidean PCA:

$$\forall i \geq 1,\ \theta_i \in \underset{\theta \in S^{d-1} \cap \mathrm{span}(\theta_1, \dots, \theta_{i-1})^\perp}{\mathrm{argmax}} \ \mathrm{Var}\big((\langle \theta, x_k \rangle)_k\big) = \mathrm{Var}\left(\big(B^\theta(x_k)\big)_k\right)$$

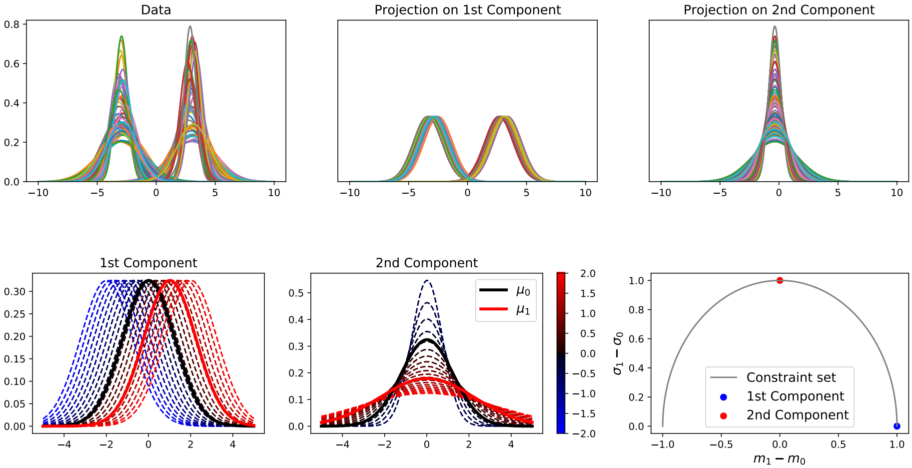
Busemann Wasserstein PCA: Let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $\nu_1, \dots, \nu_n \in \mathcal{P}_2(\mathbb{R}^d)$,

$$\forall i \geq 1,\ \mu_1^{(i)} \in \underset{\mu_1}{\mathrm{argmax}}\ \mathrm{Var}\left(\big(B^\mu(\nu_k)\big)_k\right) \quad \text{such that} \quad \begin{cases} W_2^2(\mu_0, \mu_1) = 1 \\ t \mapsto \mu_t \text{ is a geodesic ray} \\ v \in \mathrm{span}\big((v_j)_{1 \leq j \leq i-1}\big)^\perp, \end{cases}$$

with $\forall j$, $v_j = T_j - \mathrm{Id}$, T_j OT map from μ_0 to $\mu_1^{(j)}$.



Gaussian Dataset



Conclusion and Discussion

Conclusion:

- Closed-form for the Busemann function
- Application to PCA

Challenges and Perspectives:

- Projections on geodesics not always well defined
- Application for Gaussian distributions
- Closed-form for other classes of distributions?
- Extension to more general distributions, *e.g.* with Sliced-Busemann function: for $t \mapsto \mu_t$ geodesic ray in $\mathcal{P}_2(\mathbb{R})$ and $\nu \in \mathcal{P}_2(\mathbb{R}^d)$,

$$SB^\mu(\nu) = \int_{S^{d-1}} B^\mu(P^\theta_\# \nu) \, \mathrm{d}\lambda(\theta)$$

Thank you for your attention!

Publications:

- Clément Bonet, Titouan Vayer, Nicolas Courty, François Septier, and Lucas Drumetz. *Subspace Detours meet Gromov-Wasserstein*. Algorithms, 2021.
- Clément Bonet, Nicolas Courty, François Septier, and Lucas Drumetz. *Efficient Gradient Flows in Sliced-Wasserstein Space*. Transactions on Machine Learning Research, 2022.
- Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. *Spherical Sliced-Wasserstein*. International Conference on Learning Representations, 2023.
- Clément Bonet, Laetitia Chapel, Lucas Drumetz, and Nicolas Courty. *Hyperbolic Sliced-Wasserstein via Geodesic and Horospherical Projections*. Annual Workshop on Topology, Algebra, and Geometry in Machine Learning, 2023.
- Clément Bonet, Benoît Malézieux, Alain Rakotomamonjy, Lucas Drumetz, Thomas Moreau, Matthieu Kowalski, and Nicolas Courty. *Sliced-Wasserstein on Symmetric Positive Definite Matrices for M/EEG Signals*. International Conference of Machine Learning, 2023.
- Guillaume Mahey, Laetitia Chapel, Gilles Gasso, Clément Bonet, and Nicolas Courty. *Fast Optimal Transport through Sliced Wasserstein Generalized Geodesics*. Advances in Neural Information Processing Systems, 2023.
- Thibault Séjourné, Clément Bonet, Kilian Fatras, Kimia Nadjahi, and Nicolas Courty. *Unbalanced Optimal Transport meets Sliced-Wasserstein*. Submitted, 2023.

References I

Francis Bach. Effortless optimization through gradient flows, 2020. URL https://francisbach.com/gradient-flows/.

Riccardo Bellazzi, Andrea Codegoni, Stefano Gualandi, Giovanna Nicora, and Eleonora Vercesi. The gene mover's distance: Single-cell similarity via optimal transport. *arXiv preprint arXiv:2102.01218*, 2021.

Emmanuel Boissard and Thibaut Le Gouic. On the mean speed of convergence of empirical and occupation measures in wasserstein distance. In *Annales de l'IHP Probabilités et statistiques*, volume 50, pages 539–563, 2014.

Clément Bonet, Benoît Malézieux, Alain Rakotomamonjy, Lucas Drumetz, Thomas Moreau, Matthieu Kowalski, and Nicolas Courty. Sliced-Wasserstein on Symmetric Positive Definite Matrices for M/EEG Signals. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 2777–2805. PMLR, 23–29 Jul 2023a.

Clément Bonet, Titouan Vayer, Nicolas Courty, François Septier, and Lucas Drumetz. Subspace detours meet gromov–wasserstein. *Algorithms*, 14(12):366, 2021.

References II

Clément Bonet, Nicolas Courty, François Septier, and Lucas Drumetz. Efficient gradient flows in sliced-wasserstein space. *Transactions on Machine Learning Research*, 2022.

Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. Spherical sliced-wasserstein. In *The Eleventh International Conference on Learning Representations*, 2023b.

Clément Bonet, Laetitia Chapel, Lucas Drumetz, and Nicolas Courty. Hyperbolic sliced-wasserstein via geodesic and horospherical projections. In *Proceedings of 2nd Annual Workshop on Topology, Algebra, and Geometry in Machine Learning (TAG-ML)*, pages 334–370. PMLR, 2023c.

Nicolas Bonnotte. *Unidimensional and evolution methods for optimal transportation*. PhD thesis, Paris 11, 2013.

Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

References III

Yann Cabanes and Frank Nielsen. Classification in the siegel space for vectorial autoregressive data. In *Geometric Science of Information: 5th International Conference, GSI 2021, Paris, France, July 21–23, 2021, Proceedings 5*, pages 693–700. Springer, 2021.

Elsa Cazelles, Vivien Seguy, Jérémie Bigot, Marco Cuturi, and Nicolas Papadakis. Geodesic pca versus log-pca of histograms in the wasserstein space. *SIAM Journal on Scientific Computing*, 40(2):B429–B456, 2018.

Ines Chami, Albert Gu, Dat P Nguyen, and Christopher Ré. Horopca: Hyperbolic dimensionality reduction via horospherical projections. In *International Conference on Machine Learning*, pages 1419–1429. PMLR, 2021.

Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

Tim R. Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, and Jakub M. Tomczak. Hyperspherical variational auto-encoders. In Amir Globerson and Ricardo Silva, editors, *Proceedings of the Thirty-Fourth Conference on Uncertainty in Artificial Intelligence, UAI 2018, Monterey, California, USA, August 6-10, 2018*, pages 856–865. AUAI Press, 2018.

References IV

Julie Delon, Julien Salomon, and Andrei Sobolevski. Fast transport optimization for monge costs on the circle. *SIAM Journal on Applied Mathematics*, 70(7): 2239–2258, 2010.

Kilian Fatras, Younes Zine, Rémi Flamary, Remi Gribonval, and Nicolas Courty. Learning with minibatch wasserstein : asymptotic and gradient properties. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 2131–2141. PMLR, 26–28 Aug 2020.

Kilian Fatras, Younes Zine, Szymon Majewski, Rémi Flamary, Rémi Gribonval, and Nicolas Courty. Minibatch optimal transport distances; analysis and applications. *arXiv preprint arXiv:2101.01792*, 2021.

Baptiste Genest, Nicolas Courty, and David Coeurjolly. Non-euclidean sliced optimal transport sampling. *Submitted*, 2023.

Anil Goyal. *Learning a Multiview Weighted Majority Vote Classifier: Using PAC-Bayesian Theory and Boosting*. PhD thesis, Lyon, 2018.

References V

Albert Gu, Frederic Sala, Beliz Gunel, and Christopher Ré. Learning mixed-curvature representations in product spaces. In *International Conference on Learning Representations*, 2019.

Shayan Hundrieser, Marcel Klatt, and Axel Munk. The statistics of circular optimal transport. *arXiv preprint arXiv:2103.15426*, 2021.

Richard Jordan, David Kinderlehrer, and Felix Otto. The variational formulation of the fokker–planck equation. *SIAM journal on mathematical analysis*, 29(1): 1–17, 1998.

Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6418–6428, 2020.

Benoit Kloeckner. A geometric study of wasserstein spaces: Euclidean spaces. *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze*, 9(2):297–323, 2010.

Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. From word embeddings to document distances. In *International conference on machine learning*, pages 957–966. PMLR, 2015.

References VI

Federico López, Beatrice Pozzetti, Steve Trettel, Michael Strube, and Anna Wienhard. Symmetric spaces for graph embeddings: A finsler-riemannian approach. In *International Conference on Machine Learning*, pages 7090–7101. PMLR, 2021a.

Federico López, Beatrice Pozzetti, Steve Trettel, Michael Strube, and Anna Wienhard. Vector-valued distance and gyrocalculus on the space of symmetric positive definite matrices. *Advances in Neural Information Processing Systems*, 34:18350–18366, 2021b.

Boris Muzellec and Marco Cuturi. Subspace detours: Building transport plans that are optimal on subspace projections. *Advances in Neural Information Processing Systems*, 32, 2019.

Kimia Nadjahi, Alain Durmus, Umut Simsekli, and Roland Badeau. Asymptotic guarantees for learning generative models with the sliced-wasserstein distance. *Advances in Neural Information Processing Systems*, 32, 2019.

Kimia Nadjahi, Alain Durmus, Lénaïc Chizat, Soheil Kolouri, Shahin Shahrampour, and Umut Simsekli. Statistical and topological properties of sliced probability divergences. *Advances in Neural Information Processing Systems*, 33: 20802–20812, 2020.

References VII

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, pages 3779–3788. PMLR, 2018.

Ofir Pele and Michael Werman. Fast and robust earth mover's distances. In *2009 IEEE 12th international conference on computer vision*, pages 460–467. IEEE, 2009.

Xavier Pennec. Manifold-valued image processing with spd matrices. In *Riemannian geometric statistics in medical image analysis*, pages 75–134. Elsevier, 2020.

Gabriel Peyré, Marco Cuturi, and Justin Solomon. Gromov-wasserstein averaging of kernel and distance matrices. In *International Conference on Machine Learning*, pages 2664–2672. PMLR, 2016.

Alison Pouplin, David Eklund, Carl Henrik Ek, and Søren Hauberg. Identifying latent distances with finslerian geometry. *Transactions on Machine Learning Research*, 2023. ISSN 2835-8856.

References VIII

Michael Quellmalz, Robert Beinert, and Gabriele Steidl. Sliced optimal transport on the sphere. *arXiv preprint arXiv:2304.09092*, 2023.

Julien Rabin, Julie Delon, and Yann Gousseau. Transportation distances on the circle. *Journal of Mathematical Imaging and Vision*, 41(1):147–167, 2011a.

Julien Rabin, Gabriel Peyré, Julie Delon, and Marc Bernot. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 435–446. Springer, 2011b.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10684–10695, 2022.

Boris Rubin. Inversion and characterization of the hemispherical transform. *Journal d'Analyse Mathématique*, 77(1):105–128, 1999.

David Sabbagh, Pierre Ablin, Gaël Varoquaux, Alexandre Gramfort, and Denis A Engemann. Manifold-regression to predict from meg/eeg brain signals without source modeling. *Advances in Neural Information Processing Systems*, 32, 2019.

David Sabbagh, Pierre Ablin, Gaël Varoquaux, Alexandre Gramfort, and Denis A Engemann. Predictive regression modeling with meg/eeg: from source power to signals and cognitive states. *NeuroImage*, 222:116893, 2020.

Meyer Scetbon, Marco Cuturi, and Gabriel Peyré. Low-rank sinkhorn factorization. In *International Conference on Machine Learning*, pages 9344–9354. PMLR, 2021.

Vivien Seguy and Marco Cuturi. Principal geodesic analysis for probability measures under the optimal transport metric. *Advances in Neural Information Processing Systems*, 28, 2015.

Ondrej Skopek, Octavian-Eugen Ganea, and Gary Bécigneul. Mixed-curvature variational autoencoders. *arXiv preprint arXiv:1911.08411*, 2019.

Titouan Vayer, Rémi Flamary, Nicolas Courty, Romain Tavenard, and Laetitia Chapel. Sliced gromov-wasserstein. *Advances in Neural Information Processing Systems*, 32:14753–14763, 2019.

Luke Vilnis and Andrew McCallum. Word representations via gaussian embedding. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

References X

Jiacheng Xu and Greg Durrett. Spherical latent spaces for stable variational autoencoders. In Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii, editors, *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Brussels, Belgium, October 31 - November 4, 2018*, pages 4503–4513. Association for Computational Linguistics, 2018.

Or Yair, Felix Dietrich, Ronen Talmon, and Ioannis G Kevrekidis. Domain adaptation with optimal transport on the manifold of spd matrices. *arXiv preprint arXiv:1906.00616*, 2019.

Table of Contents

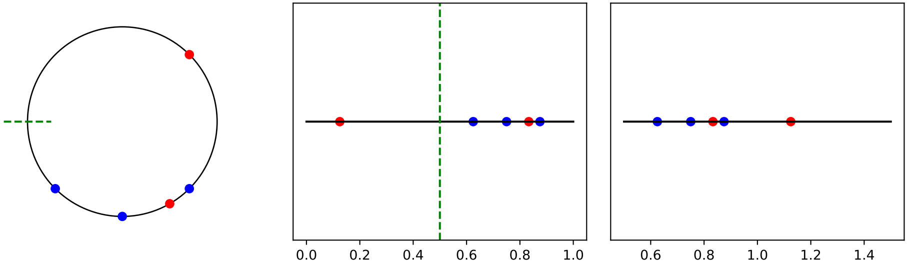
Wasserstein on the Circle

Let $\mu, \nu \in \mathcal{P}(S^1)$ where $S^1 = \mathbb{R}/\mathbb{Z}$.

- Parametrize S^1 by $[0, 1[$
- $\forall x, y \in [0, 1[,\ d_{S^1}(x, y) = \min(|x - y|, 1 - |x - y|)$
- $\forall \mu, \nu \in \mathcal{P}(S^1)$, (Rabin et al., 2011a)

$$W_p^p(\mu, \nu) = \inf_{\alpha \in \mathbb{R}} \int_0^1 |F_\mu^{-1}(t) - (F_\nu - \alpha)^{-1}(t)|^p \, \mathrm{d}t.$$

- To find α: binary search (Delon et al., 2010)



Particular Cases

- For $p = 1$, (Hundrieser et al., 2021)

$$W_1(\mu,\nu) = \int_0^1 |F_\mu(t) - F_\nu(t) - \mathrm{LevMed}(F_\mu - F_\nu)| \, \mathrm{d}t,$$

where

$$\mathrm{LevMed}(f) = \inf \left\{ t \in \mathbb{R},\ \mathrm{Leb}(\{x \in [0,1[,\ f(x) \leq t\}) \geq \frac{1}{2} \right\}.$$

- For $p = 2$ and $\nu = \mathrm{Unif}(S^1)$,

$$W_2^2(\mu,\nu) = \int_0^1 |F_\mu^{-1}(t) - t - \hat{\alpha}|^2 \, \mathrm{d}t \quad \text{with} \quad \hat{\alpha} = \int x \, \mathrm{d}\mu(x) - \frac{1}{2}.$$

In particular, if $x_1 < \cdots < x_n$ and $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, then

$$W_2^2(\mu_n,\nu) = \frac{1}{n}\sum_{i=1}^n x_i^2 - \left(\frac{1}{n}\sum_{i=1}^n x_i\right)^2 + \frac{1}{n^2}\sum_{i=1}^n (n+1-2i)x_i + \frac{1}{12}.$$

Sliced-Wasserstein on the Sphere

- Great circle: Intersection between 2-plane and S^{d-1}
- Parametrize 2-plane by the Stiefel manifold

$$\mathbb{V}_{d,2} = \{U \in \mathbb{R}^{d\times 2},\ U^T U = I_2\}$$

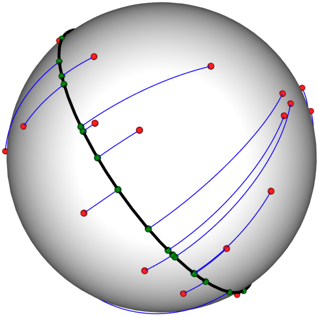
- Projection on great circle C: For a.e. $x \in S^{d-1}$,

$$P^C(x) = \underset{y\in C}{\operatorname{argmin}}\ d_{S^{d-1}}(x,y),$$

where $d_{S^{d-1}}(x,y) = \arccos(\langle x, y\rangle)$.

- For $U \in \mathbb{V}_{d,2}$, $C = \operatorname{span}(UU^T) \cap S^{d-1}$,

$$P^U(x) = U^T \underset{y\in C}{\operatorname{argmin}}\, d_{S^{d-1}}(x,y) = \frac{U^T x}{\|U^T x\|_2}.$$



Spherical Sliced-Wasserstein (Bonet et al., 2023b)

Definition (Spherical Sliced-Wasserstein)

Let $p \geq 1$, $\mu, \nu \in \mathcal{P}_p(S^{d-1})$ absolutely continuous *w.r.t.* Lebesgue measure,

$$SSW_p^p(\mu, \nu) = \int_{\mathbb{V}_{d,2}} W_p^p(P^U_\# \mu, P^U_\# \nu) \, \mathrm{d}\sigma(U),$$

with σ the uniform distribution over $\mathbb{V}_{d,2}$.

	SW	SSW
Closed-form of W	Line	(Great)-Circle
Projection	$P^\theta(x) = \langle x, \theta \rangle$	$P^U(x) = \frac{U^T x}{\|U^T x\|_2}$
Integration	S^{d-1}	$\mathbb{V}_{d,2}$

Table: Comparison SW-SSW

Is SSW a Distance?

Question: Is SSW a distance?

Proposition

Let $p \geq 1$, then SSW_p is a pseudo-distance on $\mathcal{P}_{p,ac}(S^{d-1})$.

- Lacking property (for now): indiscernibility property, *i.e.* $SSW_p(\mu,\nu) = 0 \implies \mu = \nu$.
- Need to show that $P^U_\# \mu = P^U_\# \nu$ for σ-ae $U \in \mathbb{V}_{d,2}$ implies $\mu = \nu$.
- Idea: relate P^U to a well chosen (injective) Radon transform which integrates along $\{x \in S^{d-1},\ P^U(x) = z\}$ for $U \in \mathbb{V}_{d,2}$ and $z \in S^1$.

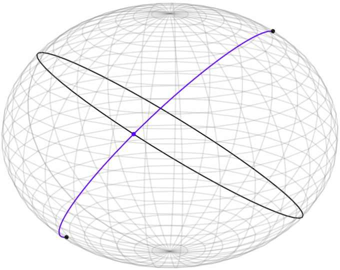
Projections Sets

Proposition

Let $U \in \mathbb{V}_{d,2}$, $z \in S^1$. *The projection set on* $z \in S^1$ *is*

$$\{x \in S^{d-1},\ P^U(x) = z\} = \{x \in F \cap S^{d-1},\ \langle x, Uz \rangle > 0\},$$

where $F = \mathrm{span}(UU^T)^\perp \oplus \mathrm{span}(Uz)$.



A Spherical Radon Transform

Definition (Spherical Radon Transform)

Let $f \in L^1(S^{d-1})$, then we define a Spherical Radon transform $\tilde{R} : L^1(S^{d-1}) \to L^1(S^1 \times \mathbb{V}_{d,2})$ as

$$\forall z \in S^1,\ \forall U \in \mathbb{V}_{d,2},\ \tilde{R}f(z,U) = \int_{S^{d-1}} f(x)\ \mathrm{d}\sigma_d^z(x),$$

with σ_d^z a suitable measure on $\{x \in S^{d-1},\ P^U(x) = z\}$.

Results on the injectivity of $\tilde{R}$ so far:

- In our work: linked it with the Hemispherical Radon transform studied in (Rubin, 1999)
- In (Quellmalz et al., 2023): showed that it a distance on S^2

Density Estimation

Goal: learn a normalizing flow T such that $T_{\#}\mu = p_Z$ with $p_Z = \mathrm{Unif}(S^{d-1})$:

$$\operatorname*{argmin}_T \ \mathrm{SSW}_2^2(T_{\#}\mu, p_Z),$$

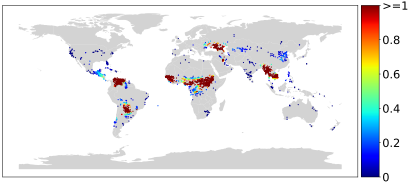
where we have access to μ through samples.

Density:

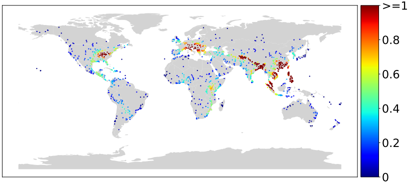
$$\forall x \in S^{d-1},\ f_\mu(x) = p_Z\big(T(x)\big)\big|\det J_T(x)\big|.$$

Table: Negative test log likelihood.

	Earthquake	Flood	Fire
SSW	$\mathbf{0.84_{\pm 0.07}}$	$\mathbf{1.26_{\pm 0.05}}$	$\mathbf{0.23_{\pm 0.18}}$
SW	$0.94_{\pm 0.02}$	$1.36_{\pm 0.04}$	$0.54_{\pm 0.37}$
Stereo	$1.91_{\pm 0.1}$	$2.00_{\pm 0.07}$	$1.27_{\pm 0.09}$



Fire



Flood

Wasserstein Autoencoders

SSWAE:

$$\mathcal{L}(f,g) = \int c\big(x, g(f(x))\big)\mathrm{d}\mu(x) + \lambda SSW_2^2(f_{\#}\mu, p_Z),$$

Much interest in using a spherical latent space (Davidson et al., 2018; Xu and Durrett, 2018), *e.g.* uniform.

Wasserstein Autoencoders

Autoencoder with spherical latent space (Davidson et al., 2018; Xu and Durrett, 2018)

SSWAE:

$$\mathcal{L}(f,g) = \int c\big(x, g(f(x))\big)\mathrm{d}\mu(x) + \lambda SSW_2^2(f_\# \mu, p_Z),$$

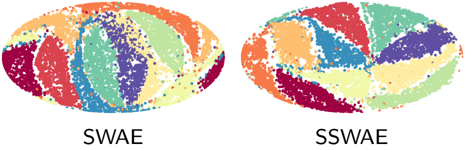


Table: FID on MNIST (Lower is better).

Method / Prior	Unif(S^{10})
SSWAE	**14.91 ± 0.32**
SWAE	15.18 ± 0.32
WAE-MMD IMQ	18.12 ± 0.62
WAE-MMD RBF	20.09 ± 1.42
SAE	19.39 ± 0.56
Circular GSWAE	15.01 ± 0.26

Table of Contents

Classification with Prototypes

- $(x_i, y_i)_{i=1}^n$ training set, $y_i \in \{1, \dots, C\}$, $\forall c \in \{1, \dots, C\},\ p_c$ prototype.
- $\forall i,\ z_i = \exp_0\big(f_\theta(x_i)\big)$
- Loss:

$$\ell(\theta) = \frac{1}{n}\sum_{i=1}^n B^{p_{y_i}}(z_i) + \lambda D\left(\frac{1}{n}\sum_{i=1}^n \delta_{z_i}, \frac{1}{C}\sum_{c=1}^C \mathcal{G}(\alpha_c p_c, \beta I_d)\right)$$

Table: Test Accuracy on deep classification with prototypes (best performance in bold)

	CIFAR10		CIFAR100		
Dimensions	2	4	3	5	10
PeBuse	$90.64_{\pm 0.06}$	$90.59_{\pm 0.11}$	$49.28_{\pm 1.95}$	$53.44_{\pm 0.76}$	$59.19_{\pm 0.39}$
GHSW	$91.39_{\pm 0.23}$	$91.66_{\pm 0.27}$	$\mathbf{53.97}_{\pm 1.35}$	$60.64_{\pm 0.87}$	$61.45_{\pm 0.41}$
HHSW	$91.28_{\pm 0.26}$	$\mathbf{91.98}_{\pm 0.05}$	$53.88_{\pm 0.06}$	$\mathbf{60.69}_{\pm 0.25}$	$\mathbf{62.80}_{\pm 0.09}$
SWp	$\mathbf{91.84}_{\pm 0.31}$	$91.68_{\pm 0.10}$	$53.25_{\pm 3.27}$	$59.77_{\pm 0.81}$	$60.36_{\pm 1.26}$
SWl	$91.13_{\pm 0.14}$	$91.74_{\pm 0.12}$	$53.88_{\pm 0.02}$	$60.62_{\pm 0.39}$	$62.30_{\pm 0.23}$
W	$91.67_{\pm 0.18}$	$91.83_{\pm 0.21}$	$50.07_{\pm 4.58}$	$57.49_{\pm 0.94}$	$58.82_{\pm 1.66}$
MMD	$91.47_{\pm 0.10}$	$91.68_{\pm 0.09}$	$50.59_{\pm 4.44}$	$58.10_{\pm 0.73}$	$58.91_{\pm 0.91}$

Gradient Flows

$$\underset{(x_i)_i}{\mathrm{argmin}}\ \mathrm{HSW}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \nu\right),$$

where we have access to ν through samples, *i.e.* $\hat{\nu}_m = \frac{1}{m}\sum_{j=1}^m \delta_{y_j}$ with $y_1, \dots, y_m \sim \nu$.

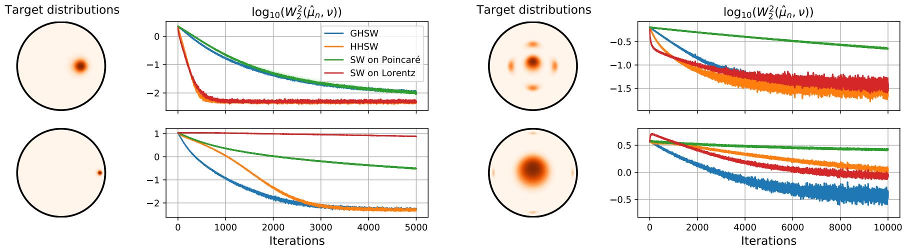


Table of Contents

Domain Adaptation in BCI

Learning a map f_θ between a source μ and a target ν

$$\min_\theta \ \mathrm{SPDSW}_2^2\big((f_\theta)_\# \mu, \nu\big)$$

Minimizing over the particles

$$\min_{(x_i)_{i=1}^n} \ \mathrm{SPDSW}_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}, \nu\right)$$

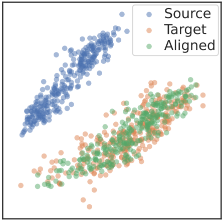


Table: Accuracy and Runtime for Cross Session.

Subjects	Source	AISOTDA (Yair et al., 2019)	SPDSW	LogSW	LEW	LES	SPDSW	LogSW	LEW	LES
			Transformations in $S_d^{++}(\mathbb{R})$				Descent over particles			
1	82.21	80.90	84.70	84.48	84.34	84.70	85.20	85.20	77.94	82.92
3	79.85	87.86	85.57	84.10	85.71	86.08	87.11	86.37	82.42	81.47
7	72.20	82.29	81.01	76.32	81.23	81.23	81.81	81.73	79.06	73.29
8	79.34	83.25	83.54	81.03	82.29	83.03	84.13	83.32	80.07	85.02
9	75.76	80.25	77.35	77.88	77.65	77.65	80.30	79.02	76.14	70.45
Avg. acc.	77.87	82.93	82.43	80.76	82.24	82.54	83.71	83.12	79.13	78.63
Avg. time (s)	-	-	**4.34**	**4.32**	11.41	12.04	**3.68**	**3.67**	8.50	11.43

Manifold of SPD Matrices with Affine-Invariant metric

Symmetric Positive Definite (SPD) Matrices:

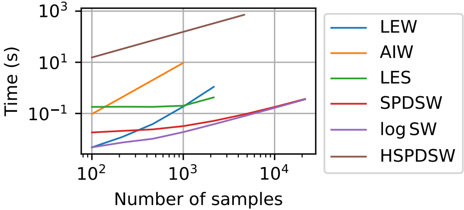
$$S_d^{++}(\mathbb{R}) = \{M \in S_d(\mathbb{R}),\ \forall x \in \mathbb{R}^d \setminus \{0\},\ x^T M x > 0\}$$

- Affine-Invariant distance:
  $\forall X, Y \in S_d^{++}(\mathbb{R})$,

$$d_{AI}(X,Y) = \sqrt{\mathrm{Tr}\big(\log(X^{-1}Y)^2\big)}$$

- Tangent space: $T_{I_d} S_d^{++}(\mathbb{R}) \cong S_d(\mathbb{R})$
- Geodesics through I_d:

$$\mathcal{G}^A = \{\exp(tA),\ t \in \mathbb{R}\}$$



- Closed-form for the geodesic projection?
- Busemann function:

$$\forall M \in S_d^{++}(\mathbb{R}),\ B^A(M) = -\langle A, \log\big(\pi_A(M)\big)\rangle_F,$$

with π_A projection on the space of matrices commuting with A.

Table of Contents

Pullback Euclidean Metric

Let $(\mathcal{N}, \langle \cdot, \cdot \rangle)$ an Euclidean space, $\phi : \mathcal{M} \to \mathcal{N}$ a diffeomorphism.

- $(\mathcal{M}, g^\phi)$ Riemannian manifold with $g_x^\phi(u, v) = \langle \phi_{*,x}(u), \phi_{*,x}(v) \rangle$ for $x \in \mathcal{M}$, $u, v \in T_x\mathcal{M}$
- Geodesic distance: $d_\mathcal{M}(x, y) = \|\phi(x) - \phi(y)\|$
- Geodesic through $o \in \mathcal{M}$ with direction $v \in T_o\mathcal{M}$: $\forall t \in \mathbb{R},\ \gamma_v(t) = \phi^{-1}\big(\phi(o) + t\phi_{*,o}(v)\big)$

Proposition

Let $v \in S_o = \{v \in T_o\mathcal{M},\ \|v\|_o = \|\phi_{*,o}(v)\| = 1\}$, then the projection coordinate on $\mathcal{G}^v = \{\gamma_v(t),\ t \in \mathbb{R}\}$ is

$$\forall x \in \mathcal{M},\ P^v(x) = -B^v(x) = \langle \phi(x) - \phi(o), \phi_{*,o}(v) \rangle.$$

Example

- Mahalanobis distance: $P^v(x) = \langle A^{\frac{1}{2}}x, A^{\frac{1}{2}}v \rangle$
- Squared geodesic distance where $\langle u, v \rangle_x = u^T A(x) v$
- SPD with ($O(n)$-Invariant) Log-Euclidean metric, Log-Cholesky metric

Table of Contents

Gradient Flows

Goal: $\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{F}(\mu)$ for $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$.

Example

- $\mathcal{F}(\mu) = \mathrm{KL}(\mu || \nu)$ for sampling from $\nu \propto e^{-V(x)}$
- $\mathcal{F}(\mu) = D(\mu, \nu)$ for sampling from ν

Definition (Gradient Flow)

A gradient flow is a curve $x : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$ which decreases as much as possible along the functional $\mathcal{F}$.

Gradient Flows

Goal: $\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{F}(\mu)$ for $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$.

Example

- $\mathcal{F}(\mu) = \mathrm{KL}(\mu||\nu)$ for sampling from $\nu \propto e^{-V(x)}$
- $\mathcal{F}(\mu) = D(\mu, \nu)$ for sampling from ν

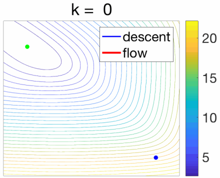
Definition (Gradient Flow)

A gradient flow is a curve $x : [0, T] \to \mathcal{P}_2(\mathbb{R}^d)$ which decreases as much as possible along the functional $\mathcal{F}$.

- Need to solve a PDE of the form

$$\partial_t \mu_t - \mathrm{div}\big(\mu_t \nabla_{W_2} \mathcal{F}(\mu_t)\big) = 0.$$

- Or approximate it by a time discretization
- Analogous to Gradient Descent/Proximal Gradient Descent



From (Bach, 2020)

Sliced-Wasserstein Gradient Flows (Bonet et al., 2022)

- Endowing $\mathcal{P}_2(\mathbb{R}^d)$ with W_2, JKO scheme (Jordan et al., 1998):

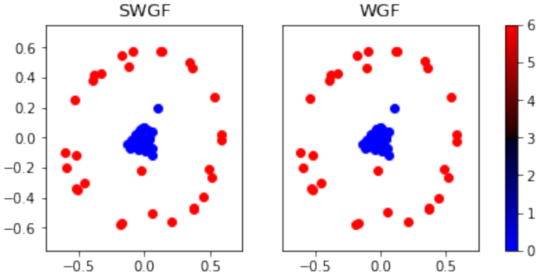
$$\mu_{k+1}^\tau \in \underset{\mu\in\mathcal{P}_2(\mathbb{R}^d)}{\mathrm{argmin}} \ \frac{1}{2\tau} W_2^2(\mu,\mu_k^\tau) + \mathcal{F}(\mu)$$

  $\rightarrow$ costly to compute in practice

- Endowing $\mathcal{P}_2(\mathbb{R}^d)$ with SW_2:

$$\mu_{k+1}^\tau \in \underset{\mu}{\mathrm{argmin}} \ \frac{d}{2\tau} \mathrm{SW}_2^2(\mu,\mu_k^\tau) + \mathcal{F}(\mu)$$

Similar trajectories:



Ornstein Uhlenbeck Process

Problem:

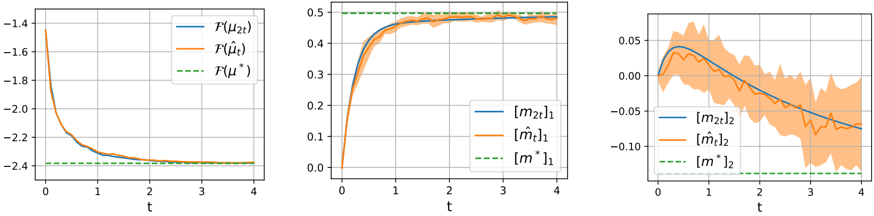
$$\min_{\mu \in \mathcal{P}_2(\mathbb{R}^d)} \mathcal{F}(\mu) = \int V \mathrm{d}\mu + \int \log(\rho(x))\rho(x)\mathrm{d}x$$

with $V(x) = \frac{1}{2}(x-m)^T A(x-m)$, $\mu^* \propto e^{-V}$, *i.e.* $\mu^* = \mathcal{N}(m, A^{-1})$.

If $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$, closed-form for the WGF of the form $\mu_t = \mathcal{N}(m_t, \Sigma_t)$

$$\begin{cases} m_t = m + e^{-tA}(m_0 - m) \\ \Sigma_t = e^{-tA}\Sigma_0(e^{-tA})^T + A^{-\frac{1}{2}}(I_d - e^{-2tA})(A^{-\frac{1}{2}})^T. \end{cases}$$

At each step k, $\mu_k^\tau = (g_\theta^k)_\# p_Z$ with g_θ^k normalizing flow.



Fokker-Planck

Problem:

$$\min_{\mu\in\mathcal{P}_2(\mathbb{R}^d)} \mathcal{F}(\mu) = \int V\mathrm{d}\mu + \int \log(\rho(x))\rho(x)\mathrm{d}x$$

with $V(x) = \frac{1}{2}(x-m)^T A(x-m)$, $\mu^* \propto e^{-V}$, *i.e.* $\mu^* = \mathcal{N}(m, A^{-1})$.

At each step k, $\mu_k^\tau = (g_\theta^k)_\# p_Z$ with g_θ^k normalizing flow.

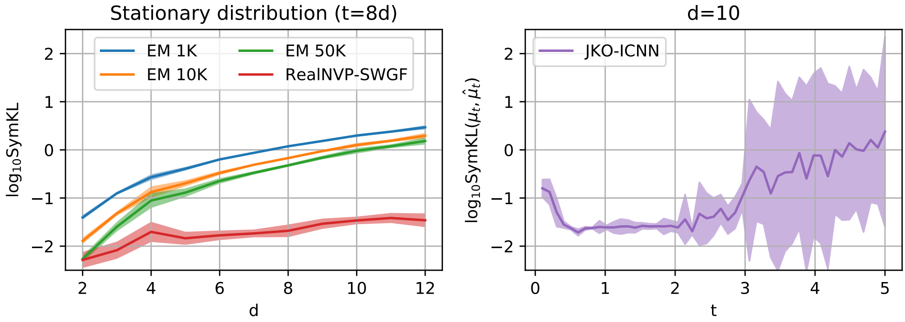


Table of Contents

Sliced-Unbalanced OT

Unbalanced OT problem

Let $\mu, \nu \in \mathcal{M}_+(\mathbb{R}^d)$,

$$\mathrm{UOT}(\mu, \nu) = \inf_{\pi_1, \pi_2 \in \mathcal{M}_+(\mathbb{R}^d)} W_2^2(\pi_1, \pi_2) + \rho_1 \mathrm{KL}(\pi_1 || \mu) + \rho_2 \mathrm{KL}(\pi_2 || \nu)$$

Sliced-UOT problems:

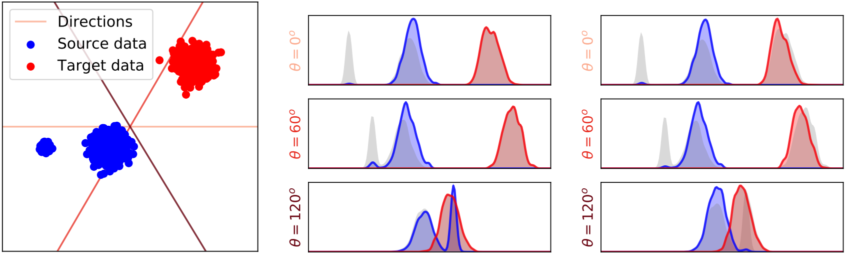
- SUOT:

$$\mathrm{SUOT}(\mu, \nu) = \int_{S^{d-1}} \mathrm{UOT}(P^\theta_\# \mu, P^\theta_\# \nu) \, \mathrm{d}\lambda(\theta)$$

- USOT:

$$\mathrm{USW}_2^2(\mu, \nu) = \inf_{\pi_1, \pi_2 \in \mathcal{M}_+(\mathbb{R}^d)} \mathrm{SW}_2^2(\pi_1, \pi_2) + \rho_1 \mathrm{KL}(\pi_1 || \mu) + \rho_2 \mathrm{KL}(\pi_2 || \nu)$$

Sliced-Unbalanced OT



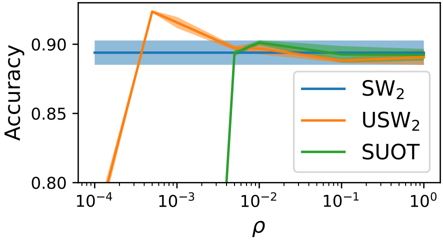
Samples and directions | SUOT $(\rho_1 = 0.01,\ \rho_2 = 1)$ | USW $(\rho_1 = 1,\ \rho_2 = 1)$

- Applicable to any SW distance (*e.g.* HSW)
- Application to document classification and geoclimatic barycenter

Document Classification

- Document $D_k = \sum_{i=1}^{n_k} w_i^k \delta_{x_i^k}$
- Compute $\left(L(D_k, D_\ell)\right)_{k,\ell}$
- k-nearest neighbor classifier

	BBCSport	**Movies**	**Goodreads genre**	**Goodreads like**
W_2	94.55	74.44	55.22	71.00
UOT	96.73	-	-	-
Sinkhorn UOT	93.64	72.48	53.55	67.81
SW_2	$89.39_{\pm 0.76}$	$66.95_{\pm 0.45}$	$50.09_{\pm 0.51}$	$65.60_{\pm 0.20}$
SUOT	$90.12_{\pm 0.15}$	$67.84_{\pm 0.37}$	$50.15_{\pm 0.04}$	$66.72_{\pm 0.38}$
USOT_2	$92.36_{\pm 0.07}$	$69.21_{\pm 0.37}$	$51.87_{\pm 0.56}$	$67.41_{\pm 1.06}$
SUSOT_2	$92.45_{\pm 0.39}$	$69.53_{\pm 0.53}$	$51.93_{\pm 0.53}$	$67.33_{\pm 0.26}$



Barycenter on Geophysical Data

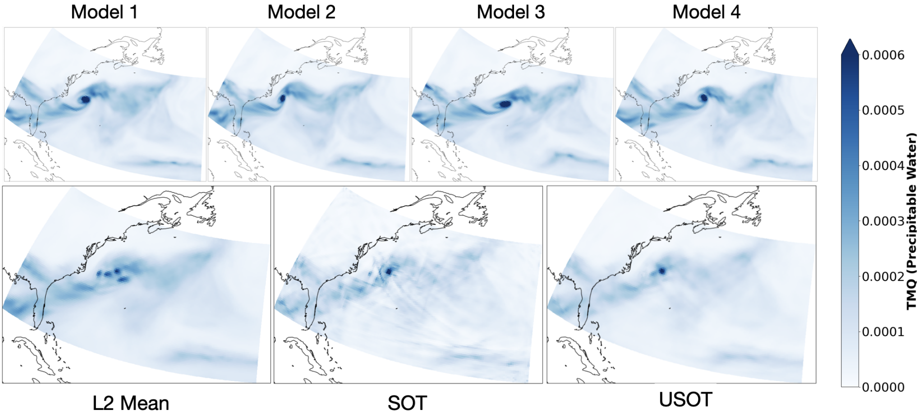


Table of Contents

Well-definedness of the projection

Proposition

Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$ and $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$ two Gaussian defining a unit-speed geodesic ray starting from μ_0 and passing through μ_1 at $t = 1$, i.e. satisfying $\sigma_1 \geq \sigma_0$ and $(m_1 - m_0)^2 + (\sigma_1 - \sigma_0)^2 = 1$. Then, the underlying geodesic ray $t \mapsto \mu_t$ is well defined on $[-\frac{\sigma_0}{\sigma_1 - \sigma_0}, +\infty[$.

- Translation: $\sigma_0 = \sigma_1$ implies that μ_t is defined on $]-\infty, +\infty[$
- Dilation: $\sigma_1 = 1 + \sigma_0$ implies that μ_t is defined on $[-\sigma_0, +\infty[$. But for any $\nu = \mathcal{N}(m, \sigma^2)$,

  $$B^\mu(\nu) = -(m - m_0)(m_1 - m_0) - (\sigma - \sigma_0)(\sigma_1 - \sigma_0) = -(\sigma - \sigma_0),$$

  and thus, $P^\mu(\nu) = -B^\mu(\nu) = \sigma - \sigma_0 < -\sigma_0 \iff \sigma < 0$, which is not possible.
- For other cases: distributions might get out of the geodesic

Table of Contents

Gromov-Wasserstein Distance

Let (X, d_X, μ), (Y, d_Y, ν) be metric-measures spaces (mm-spaces).

Gromov-Wasserstein

$$GW(X,Y) = \inf_{\gamma \in \Pi(\mu,\nu)} \iint \left(c_X(x,x') - c_Y(y,y')\right)^2 \,\mathrm{d}\gamma(x,y)\mathrm{d}\gamma(x',y')$$

with $\Pi(\mu,\nu) = \{\gamma \in \mathcal{P}(X \times Y),\ \pi^1_\# \gamma = \mu,\ \pi^2_\# \gamma = \nu\}$.

Examples

For $X = \mathbb{R}^p$, $Y = \mathbb{R}^q$,

- $c_X(x,x') = d_X(x,x')$, $c_Y(y,y') = d_Y(y,y)$
- $c_X(x,x') = \|x - x'\|_2^2$, $c_Y(y,y') = \|y - y'\|_2^2$
- $c_X(x,x') = \langle x, x' \rangle_p$, $c_Y(y,y') = \langle y, y' \rangle_q$

Gromov-Wasserstein Distance

Let (X, d_X, μ), (Y, d_Y, ν) be metric-measures spaces (mm-spaces).

Gromov-Wasserstein

$$GW(X,Y) = \inf_{\gamma \in \Pi(\mu,\nu)} \iint \left(c_X(x,x') - c_Y(y,y')\right)^2 \mathrm{d}\gamma(x,y)\mathrm{d}\gamma(x',y')$$

with $\Pi(\mu,\nu) = \{\gamma \in \mathcal{P}(X \times Y),\ \pi^1_\# \gamma = \mu,\ \pi^2_\# \gamma = \nu\}$.

Practical Issues:

- Computational Complexity: $O(n^4)$

Proposed Solution:

- Entropic Regularization (Peyré et al., 2016)
- Low rank constraints (Scetbon et al., 2021)
- Sliced Gromov-Wasserstein (Vayer et al., 2019)
- Minibatch estimators (Fatras et al., 2021)

Subspace Detour Approach

Subspace Detour approach (Muzellec and Cuturi, 2019):

- Choose a subspace E
- Project the measures: $\mu_E = \pi^E_\# \mu$ and $\nu_E = \pi^E_\# \nu$
- Take the optimal coupling $\gamma_E^* \in \Pi(\mu_E, \nu_E)$
- Find a coupling $\gamma \in \Pi_E(\mu, \nu) = \{\gamma \in \Pi(\mu, \nu) |\ (\pi^E, \pi^E)_\# \gamma = \gamma_E^*\}$

Coupling on the whole set:

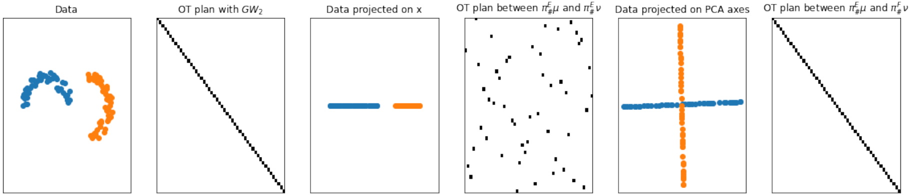
- Monge-Independent plan:

$$\pi_{\mathrm{MI}} = \gamma_E^* \otimes (\mu_{E^\perp|E} \otimes \nu_{E^\perp|E})$$

- Monge-Knothe plan:

$$\pi_{\mathrm{MK}} = \gamma_E^* \otimes \gamma^*_{E^\perp|E}$$

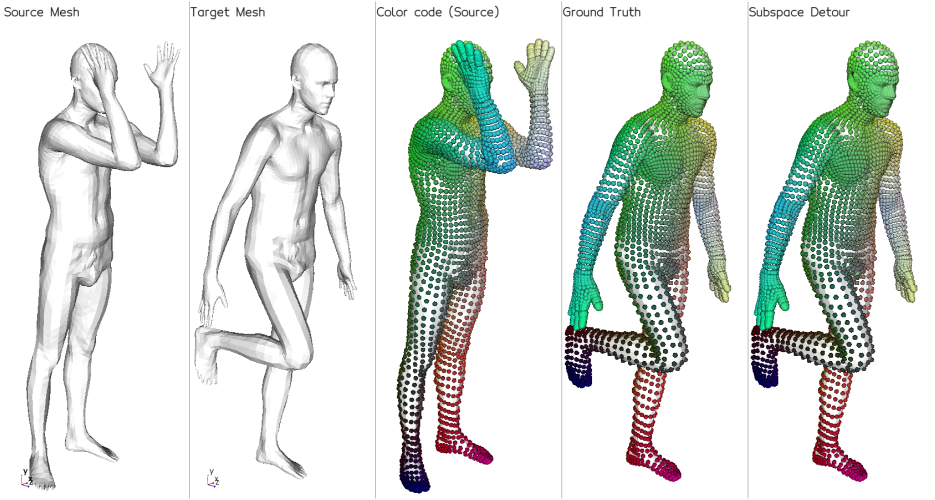
Subspace Detour Approach for GW (Bonet et al., 2021)



Subspace Detour approach for GW: Let $\mu \in \mathcal{P}(\mathbb{R}^p)$, $\nu \in \mathcal{P}(\mathbb{R}^q)$,

- Choose subspace $E \subset \mathbb{R}^p$ and $F \subset \mathbb{R}^q$
- Project the measures: $\mu_E = \pi^E_\# \mu$ and $\nu_F = \pi^F_\# \nu$
- Take the optimal coupling $\gamma^*_{E,F} \in \Pi(\mu_E, \nu_F)$
- Find a coupling $\gamma \in \Pi_{E,F}(\mu, \nu) = \{\gamma \in \Pi(\mu, \nu) |\ (\pi^E, \pi^F)_\# \gamma = \gamma^*_{E \times F}\}$

Application



Application