Hyperbolic Sliced-Wasserstein via Geodesic and Horospherical Projections

Clément Bonet[1], Laetitia Chapel[2], Lucas Drumetz[3], Nicolas Courty[2]

[1]Université Bretagne Sud, LMBA; [2]Université Bretagne Sud, IRISA; [3]IMT Atlantique, Lab-STICC

Optimal Transport

Wasserstein distance. Let $(\mathcal{M}, d)$ be a Riemannian manifold, $p \geq 1$, $\mu, \nu \in \mathcal{P}_p(\mathcal{M})$, then

$$W_p^p(\mu,\nu) = \inf_{\gamma\in\Pi(\mu,\nu)} \int_{\mathcal{M}\times\mathcal{M}} d(x,y)^p \,\mathrm{d}\gamma(x,y).$$

In practice: $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\hat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$ and we compute $W_p^p(\hat{\mu}_n, \hat{\nu}_n)$.
Complexity *w.r.t* number of samples n: $O(n^3 \log n)$

Wasserstein distance on $\mathbb{R}$. Let $\mu,\nu \in \mathcal{P}(\mathbb{R})$, $p \geq 1$,

$$W_p^p(\mu,\nu) = \int_0^1 |F_\mu^{-1}(u) - F_\nu^{-1}(u)|^p \,\mathrm{d}u.$$

Complexity *w.r.t* number of samples n: $O(n\log n)$

Sliced-Wasserstein distance. Let $p \geq 1$, $\mu,\nu \in \mathcal{P}_p(\mathbb{R}^d)$,

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

where $P^\theta: x \mapsto \langle x, \theta\rangle$ and $\#$ is the push-forward operator.
In practice: Monte-Carlo approximation $\widehat{SW}_p^p(\mu,\nu) = \frac{1}{L}\sum_{\ell=1}^L W_p^p(P^{\theta_\ell}_\#\mu, P^{\theta_\ell}_\#\nu)$.
Complexity *w.r.t* number of samples n and projections L: $O\big(Ln(\log n + d)\big)$.

Hyperbolic Spaces

Hyperbolic spaces are Riemannian manifolds of constant negative curvatures.
Different possible parametrizations (up to isometry):

- **Lorentz model:** $\mathbb{L}^d = \{(x_0,\dots x_d)\in\mathbb{R}^{d+1},\ \langle x,x\rangle_{\mathbb{L}} = -1,\ x_0 > 0\}$ where for all $x,y\in\mathbb{R}^{d+1}$, $\langle x,y\rangle_{\mathbb{L}} = -x_0y_0 + \sum_{i=1}^d x_iy_i$ is the Minkowski inner-product.
  Geodesic distance on $\mathbb{L}^d$: $\forall x,y\in\mathbb{L}^d$, $d_{\mathbb{L}}(x,y) = \mathrm{arccosh}(-\langle x,y\rangle_{\mathbb{L}})$
  Tangent space at $x\in\mathbb{L}^d$: $T_x\mathbb{L}^d = \{v\in\mathbb{R}^{d+1},\ \langle v,x\rangle_{\mathbb{L}} = 0\}$
  Geodesic line $\gamma_{\mathbb{L}}$ passing through $x^0 = (1,0,\dots,0)$ in direction $v\in T_{x^0}\mathbb{L}^d\cap S^d$:
  $$\forall t\in\mathbb{R},\ \gamma_{\mathbb{L}}(t) = \cosh(t)x^0 + \sinh(t)v$$
- **Poincaré ball:** $\mathbb{B}^d = \{x\in\mathbb{R}^d,\ \|x\|_2 < 1\}$
  Geodesic distance on $\mathbb{B}^d$: $\forall x,y\in\mathbb{B}^d$, $d_{\mathbb{B}}(x,y) = \mathrm{arccosh}\left(1 + 2\frac{\|x-y\|_2^2}{(1-\|x\|_2^2)(1-\|y\|_2^2)}\right)$
  Tangent space at $x\in\mathbb{B}^d$: $T_x\mathbb{B}^d = \mathbb{R}^d$
  Geodesic line $\gamma_{\mathbb{B}}$ passing through 0 in direction $\tilde{v}\in S^{d-1}$:
  $$\forall t\in\mathbb{R},\ \gamma_{\mathbb{B}}(t) = \tanh(t/2)\tilde{v}$$

Contributions

- Sliced-Wasserstein distance intrinsically defined on Hyperbolic spaces
- Comparisons on gradient flows and classification

Geodesic Hyperbolic Sliced-Wasserstein

- Draw a direction of geodesic ($v\in T_{x^0}\mathbb{L}^d\cap S^d \cong S^{d-1}$ or $\tilde{v}\in S^{d-1}$)
- Geodesic projection:

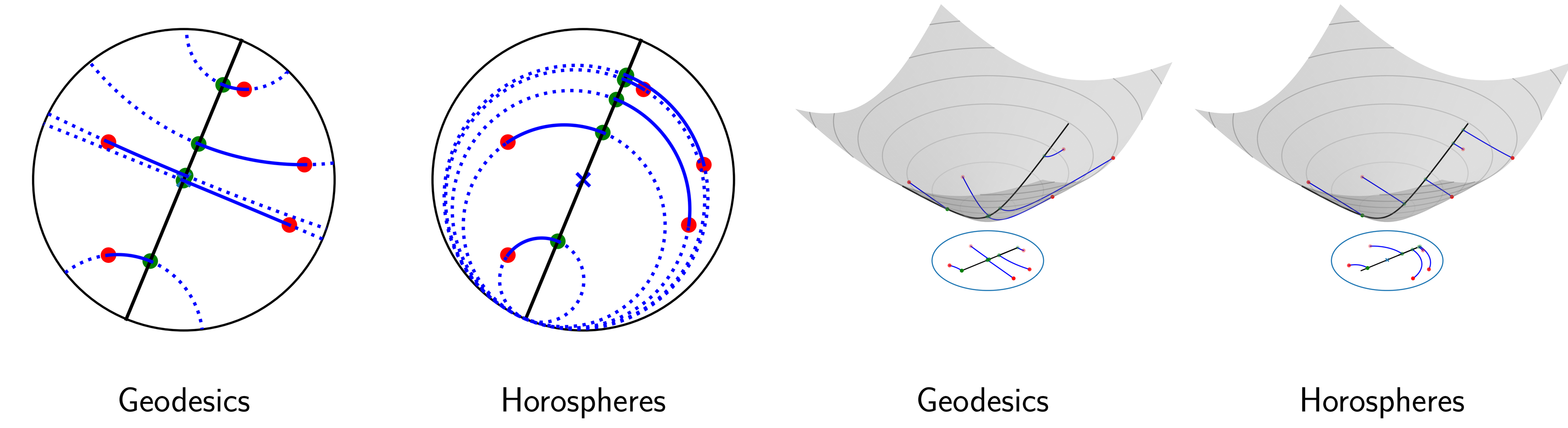
$$\forall x\in\mathbb{L}^d,\ P^v(x) = \operatorname*{argmin}_{t\in\mathbb{R}}\ d_{\mathbb{L}}\big(\gamma_{\mathbb{L}}(t),x\big) = \mathrm{arctanh}\left(-\frac{\langle x,v\rangle_{\mathbb{L}}}{\langle x,x^0\rangle_{\mathbb{L}}}\right),$$
$$\forall x\in\mathbb{B}^d,\ P^{\tilde{v}}(x) = \operatorname*{argmin}_{t\in\mathbb{R}}\ d_{\mathbb{B}}\big(\gamma_{\mathbb{B}}(t),x\big) = 2\,\mathrm{arctanh}\big(s(x)\big),$$

where $s(x) = \frac{1+\|x\|_2^2 - \sqrt{(1+\|x\|_2^2)^2 - 4\langle x,\tilde{v}\rangle^2}}{2\langle x,\tilde{v}\rangle}\mathbb{1}_{\{\langle x,\tilde{v}\rangle\neq 0\}}$.

Geodesic Hyperbolic Sliced-Wasserstein

Let $\mu,\nu\in\mathcal{P}_p(\mathbb{L}^d)$, $\tilde{\mu},\tilde{\nu}\in\mathcal{P}_p(\mathbb{B}^d)$, $p\geq 1$,

$$GHSW_p^p(\mu,\nu) = \int_{T_{x^0}\mathbb{L}^d\cap S^d} W_p^p(P^v_\#\mu, P^v_\#\nu)\,\mathrm{d}\lambda(v)$$
$$GHSW_p^p(\tilde{\mu},\tilde{\nu}) = \int_{S^{d-1}} W_p^p(P^{\tilde{v}}_\#\tilde{\mu}, P^{\tilde{v}}_\#\tilde{\nu})\,\mathrm{d}\lambda(\tilde{v}).$$



Horospherical Hyperbolic Sliced-Wasserstein

Busemann function associated to a geodesic line γ:

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

On $\mathbb{R}^d$, $\gamma(t) = t\theta$ for $\theta\in S^{d-1}$ and $B^\gamma(x) = -\langle x,\theta\rangle$
Level sets of B^γ: horospheres which can be seen as generalizations of hyperplanes

- Horospherical projection:

$$\forall x\in\mathbb{L}^d,\ B^v(x) = \log(-\langle x, x^0+v\rangle_{\mathbb{L}}),\quad \forall x\in\mathbb{B}^d,\ B^{\tilde{v}}(x) = \log\left(\frac{\|\tilde{v}-x\|_2^2}{1-\|x\|_2^2}\right).$$

Horospherical Hyperbolic Sliced-Wasserstein

Let $\mu,\nu\in\mathcal{P}_p(\mathbb{L}^d)$, $\tilde{\mu},\tilde{\nu}\in\mathcal{P}_p(\mathbb{B}^d)$, $p\geq 1$,

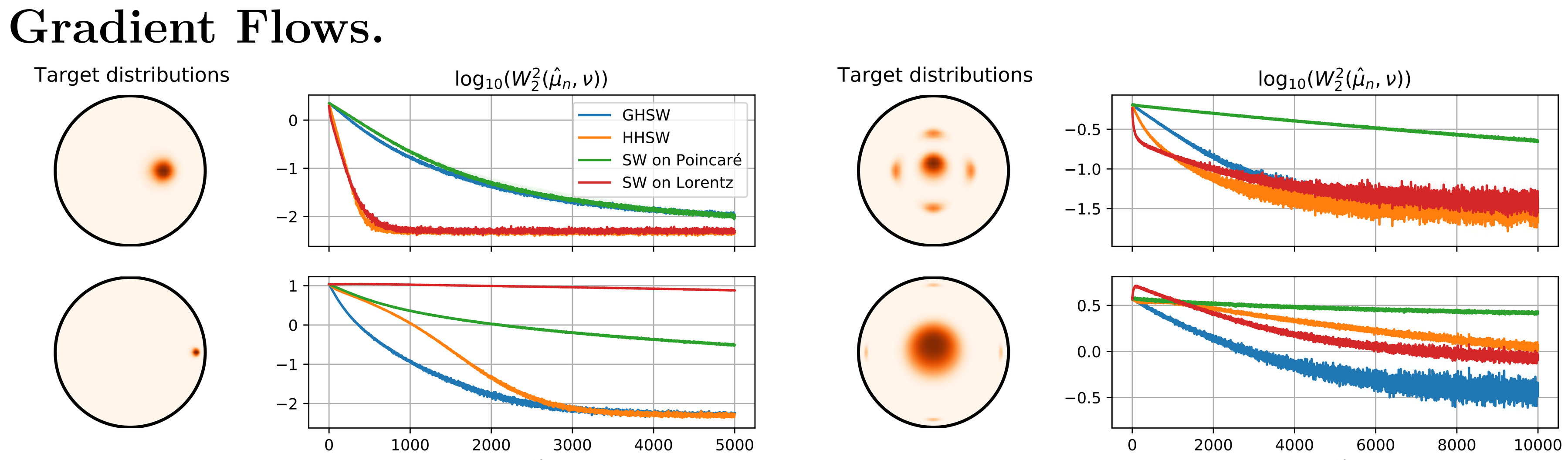
$$HHSW_p^p(\mu,\nu) = \int_{T_{x^0}\mathbb{L}^d\cap S^d} W_p^p(B^v_\#\mu, B^v_\#\nu)\,\mathrm{d}\lambda(v)$$
$$HHSW_p^p(\tilde{\mu},\tilde{\nu}) = \int_{S^{d-1}} W_p^p(B^{\tilde{v}}_\#\tilde{\mu}, B^{\tilde{v}}_\#\tilde{\nu})\,\mathrm{d}\lambda(\tilde{v}).$$

Properties

- Independent of the model, *i.e.* for $p\geq 1$, $\tilde{\mu},\tilde{\nu}\in\mathcal{P}_p(\mathbb{B}^d)$, denote $\mu = (P_{\mathbb{B}\to\mathbb{L}})_\#\tilde{\mu}$, $\nu = (P_{\mathbb{B}\to\mathbb{L}})_\#\tilde{\nu}$. Then,
  $$HHSW_p^p(\mu,\nu) = HHSW_p^p(\tilde{\mu},\tilde{\nu}),\quad GHSW_p^p(\mu,\nu) = GHSW_p^p(\tilde{\mu},\tilde{\nu}).$$
- Pseudo-distance, sample complexity independent of the dimension

Applications

Gradient Flows.



Classification with prototypes. Denote $(x_i,y_i)_i$ the training set, $p_{y_i}\in S^{d-1}$ a prototype associated to the label y_i, $z_i = f_\theta(x_i)\in\mathbb{B}^d$.
Idea: regularize with a mixture of Wrapped Normal distribution $MWND$.

Dimensions	3	5	10
PeBuse	$49.28_{\pm 1.95}$	$53.44_{\pm 0.76}$	$59.19_{\pm 0.39}$
GHSW	$\mathbf{53.97}_{\pm 1.35}$	$60.64_{\pm 0.87}$	$61.45_{\pm 0.41}$
HHSW	$53.88_{\pm 0.06}$	$\mathbf{60.69}_{\pm 0.25}$	$\mathbf{62.80}_{\pm 0.09}$
SWp	$53.25_{\pm 3.27}$	$59.77_{\pm 0.81}$	$60.36_{\pm 1.26}$
SWl	$53.88_{\pm 0.02}$	$60.62_{\pm 0.39}$	$62.30_{\pm 0.23}$

Test Accuracy on CIFAR100

$$\text{For } (w_i)_i \sim MWND,\ \ell(\theta) = \frac{1}{n}\sum_{i=1}^n B^{p_i}(z_i) + \lambda HSW_2^2\left(\frac{1}{n}\sum_{i=1}^n \delta_{z_i}, \frac{1}{n}\sum_{i=1}^n \delta_{w_i}\right)$$

References

Ghadimi Atigh, M., Keller-Ressel, M., Mettes, P. Hyperbolic Busemann Learning with Ideal Prototypes. *Neurips*, 2021.

Chami, I., Gu, A., Nguyen, D.P., Ré, C. Horopca: Hyperbolic Dimensionality Reduction via Horospherical Projections. *ICML*, PMLR, 2021.