Busemann Functions in the Wasserstein Space: Existence, Closed-Forms and Applications

Clément Bonet[1], Elsa Cazelles[2], Lucas Drumetz[3], Nicolas Courty[4]

[1]Ecole Polytechnique, CMAP, Institut Polytechnique de Paris
[2]CNRS, Université de Toulouse, IRIT
[3]IMT Atlantique, Lab-STICC
[4]Université Bretagne Sud, IRISA

Imaging in Paris
07/04/2026

Motivation

Busemann function:

- Major tool in geometry (Busemann, 1955; Bridson and Haefliger, 2013)
- Many applications in Machine Learning (Chami et al., 2021; Ghadimi Atigh et al., 2021; Bonet et al., 2023; Berg et al., 2024)
  → PCA, Classification, Projections...
- Applications restricted to finite dimensional spaces

Space of probability distributions:

- Datasets of Distributions
  → Documents (Kusner et al., 2015), point clouds (Geuter et al., 2025), images (Rubner et al., 2000), single-cells (Bellazzi et al., 2021)
  → Labeled distributions (Alvarez-Melis and Fusi, 2020), Gaussian mixtures (Delon and Desolneux, 2020)...
- Rich geometry with Optimal Transport (Ambrosio et al., 2008)

Table of Contents

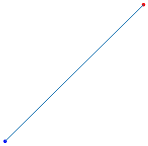
Geodesic Metric space (Bridson and Haefliger, 2013)

Let (X, d) be a metric space.

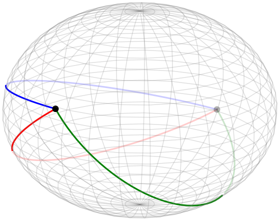
- Let $x, y \in X$. A continuous map $\gamma : [0,1] \to X$ is a (constant-speed) geodesic between x and y if

$$\begin{cases} \gamma(0) = x \\ \gamma(1) = y \\ \forall t, s \in [0,1],\ d\big(\gamma(t), \gamma(s)\big) = |t-s| d(x,y) \end{cases}$$

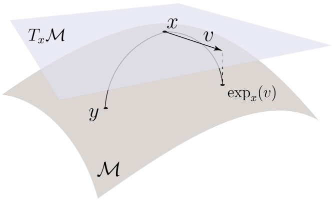
  $\to \gamma$ minimizes the length between x and y
- (X, d) is a geodesic metric space if any two points are joined by a geodesic



$X = \mathbb{R}^d$, $d(x,y) = \|x-y\|_2$, $\gamma(t) = (1-t)x + ty$



$X = S^{d-1}$



$X = \mathcal{M}$, $\gamma(t) = \exp_x(tv)$

Geodesic Lines and Rays

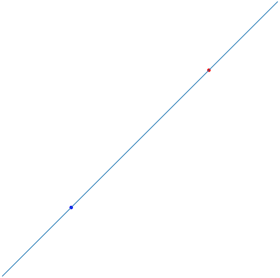
Let $\gamma : [0,1] \to X$ be a geodesic, $\kappa = d\big(\gamma(0), \gamma(1)\big)$ its speed.

- **Geodesic line**: extension of γ to $\mathbb{R}$ such that

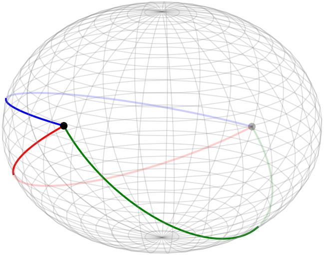
$$\forall t, s \in \mathbb{R},\ d\big(\gamma(t), \gamma(s)\big) = \kappa|t - s|$$

- **Geodesic ray**: extension of γ to $[0, +\infty[$ such that

$$\forall t, s \in [0, +\infty[,\ d\big(\gamma(t), \gamma(s)\big) = \kappa|t - s|$$



$\forall s, t \in \mathbb{R},\ \gamma(t) = (1-t)x + ty$



No geodesic ray or line

Sufficient Conditions for Geodesic Rays?

- Curvature $\leq 0 \iff$ for all $x \in X$, $t \in [0,1]$,

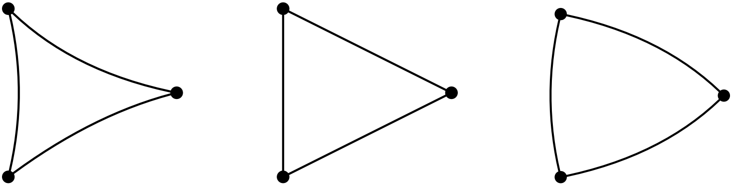
$$d^2\big(x,\gamma(t)\big) \leq (1-t)d^2\big(x,\gamma(0)\big) + td^2\big(x,\gamma(1)\big) - t(1-t)d^2\big(\gamma(0),\gamma(1)\big)$$

  $\rightarrow$ geodesics are **geodesics lines**

- Curvature $> 0 \iff$ for all $x \in X$, $t \in [0,1]$,

$$d^2\big(x,\gamma(t)\big) > (1-t)d^2\big(x,\gamma(0)\big) + td^2\big(x,\gamma(1)\big) - t(1-t)d^2\big(\gamma(0),\gamma(1)\big)$$

  $\rightarrow$ **no guarantees**



Negatively curved — No curvature — Positively curved

Examples of Non-Positively Curves Spaces

- **Euclidean spaces** $(\mathbb{R}^d, \|\cdot\|_2)$: $\forall x \in \mathbb{R}^d$, (parallelogram rule)

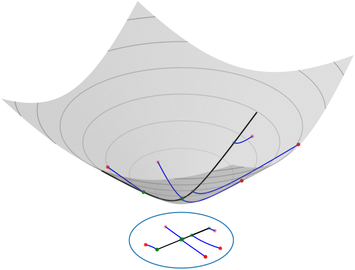
$$\|x-\gamma(t)\|_2^2 = (1-t)\|x-\gamma(0)\|_2^2 + t\|x-\gamma(1)\|_2^2 - t(1-t)\|\gamma(0)-\gamma(1)\|_2^2$$

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

$$d_{\mathbb{L}}(x,y) = \mathrm{arccosh}\big(-\langle x,y\rangle_{\mathbb{L}}\big), \quad \langle x,y\rangle_{\mathbb{L}} = -x_0y_0 + \sum_{i=1}^d x_iy_i$$

Geodesic: $\forall t\in\mathbb{R},\ \gamma(t) = \cosh(\|v\|_{\mathbb{L}})x + \sinh(t\|v\|_{\mathbb{L}})\frac{v}{\|v\|_{\mathbb{L}}}$

$\rightarrow$ (constant) negative curvature



Busemann Function

Let (X, d) be a geodesic metric space and $\gamma : [0, +\infty[\to X$ be a geodesic ray.

Busemann function associated to γ:

$$\forall x \in X,\ B^\gamma(x) = \lim_{t \to +\infty} d\big(x, \gamma(t)\big) - d\big(\gamma(0), \gamma(t)\big)$$
$$= \lim_{t \to +\infty} d\big(x, \gamma(t)\big) - t \cdot d\big(\gamma(0), \gamma(1)\big)$$

Busemann Function

Let (X, d) be a geodesic metric space and $\gamma : [0, +\infty[\to X$ be a geodesic ray.

Busemann function associated to γ:

$$\begin{aligned} \forall x \in X,\ B^\gamma(x) &= \lim_{t \to +\infty} d\big(x, \gamma(t)\big) - d\big(\gamma(0), \gamma(t)\big) \\ &= \lim_{t \to +\infty} d\big(x, \gamma(t)\big) - t \cdot d\big(\gamma(0), \gamma(1)\big) \end{aligned}$$

For $X = \mathbb{R}^d$, $\gamma(t) = x_0 + tv$, $x_0, v = x_1 - x_0 \in \mathbb{R}^d$,

$$\begin{aligned} d\big(x, \gamma(t)\big) - d\big(\gamma(0), \gamma(t)\big) &= \|x - x_0 - tv\|_2 - \|x_0 - x_0 - tv\|_2 \\ &= t\|v\|_2 \sqrt{1 - \frac{2}{t\|v\|_2^2}\langle v, x - x_0\rangle + o(t^{-1})} - t\|v\|_2 \\ &= t\|v\|_2 \left(1 - \frac{1}{t\|v\|_2^2}\langle v, x - x_0\rangle + o(t^{-1})\right) - t\|v\|_2 \\ &= -\left\langle x - x_0, \frac{v}{\|v\|_2} \right\rangle + o(1) \end{aligned}$$

$\to B^\gamma(x) = -\left\langle x - x_0, \frac{v}{\|v\|_2} \right\rangle$

Busemann Function on $\mathbb{R}^d$

On $\mathbb{R}^d$: for $\gamma(t) = x_0 + tv$ with $x_0, v \in \mathbb{R}^d$,

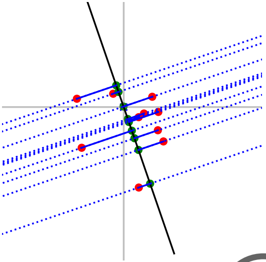
$$\forall x \in \mathbb{R}^d,\ B^\gamma(x) = -\left\langle x - x_0, \frac{v}{\|v\|_2} \right\rangle$$

- $B^\gamma(x)$: coincides with the coordinate of the geodesic projection (up to a sign)

$$t = \operatorname*{argmin}_{s\in\mathbb{R}} \big\|x - (x_0 + sv)\big\|_2^2 = \left\langle x - x_0, \frac{v}{\|v\|_2} \right\rangle$$

  $\rightarrow$ Projection: $\gamma(-B^\gamma(x))$

- Level sets of B^γ are hyperplanes orthogonal to v



Busemann Function on $\mathbb{R}^d$

On $\mathbb{R}^d$: for $\gamma(t) = x_0 + tv$ with $x_0, v \in \mathbb{R}^d$,

$$\forall x \in \mathbb{R}^d,\ B^\gamma(x) = -\left\langle x - x_0, \frac{v}{\|v\|_2} \right\rangle$$

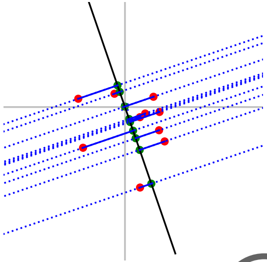
- $B^\gamma(x)$: coincides with the coordinate of the geodesic projection (up to a sign)

$$t = \underset{s\in\mathbb{R}}{\mathrm{argmin}} \left\| x - (x_0 + sv) \right\|_2^2 = \left\langle x - x_0, \frac{v}{\|v\|_2} \right\rangle$$

  $\rightarrow$ Projection: $\gamma(-B^\gamma(x))$

- Level sets of B^γ are hyperplanes orthogonal to v

$\rightarrow$ **What about on other spaces?**



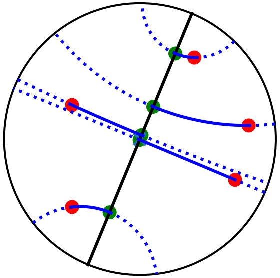
Busemann Function on Hyperbolic Space

Let $x^0 = (1, 0, \dots, 0) \in \mathbb{R}^{d+1}$, $v \in T_{x^0}\mathbb{L}^d \cap S^d$,

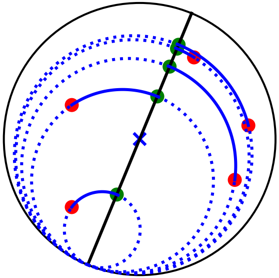
$$\forall x \in \mathbb{L}^d,\ B^\gamma(x) = \log\big(-\langle x, x^0 + v\rangle_{\mathbb{L}}\big)$$

$\rightarrow$ different from geodesic projection (Bonet et al., 2023)

Horospheres: Levels sets of B^γ: $(B^\gamma)^{-1}(\{t\})$ for all $t \in \mathbb{R}$
$\rightarrow$ Second generalization of hyperplanes



Projections along geodesics submanifolds

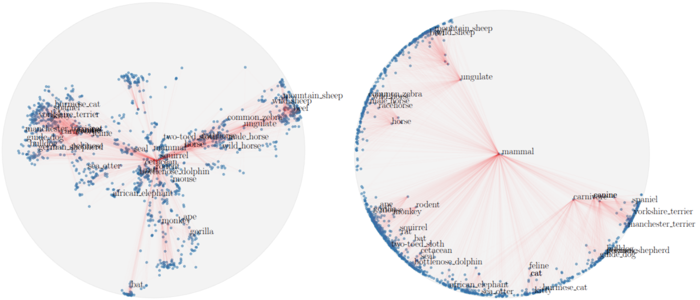


Projections along horospheres

Applications in Machine Learning

- Busemann used as a projection
  - HoroPCA (Chami et al., 2021)

    → Project on a geodesic subspace along horospheres

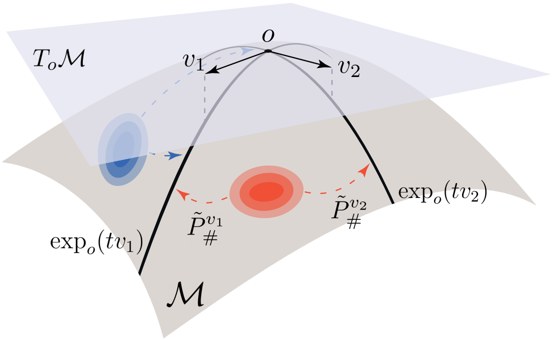


(a) PGA (average distortion: 0.534) (b) HOROPCA (average distortion: 0.078)

From (Chami et al., 2021)

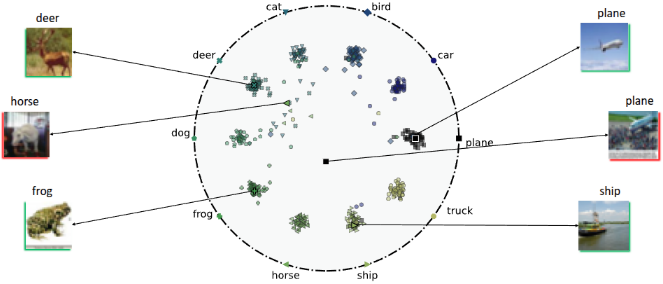
Applications in Machine Learning

- Busemann used as a projection
  - HoroPCA (Chami et al., 2021)
    → Project on a geodesic subspace along horospheres
  - Sliced-Wasserstein on Cartan-Hadamard manifolds (Bonet et al., 2023, 2025a)
    → Project on geodesics with the Busemann function



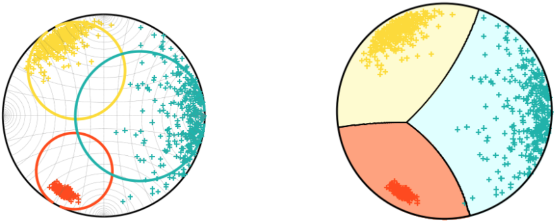
Applications in Machine Learning

- Busemann used as a projection
  - HoroPCA (Chami et al., 2021)
    → Project on a geodesic subspace along horospheres
  - Sliced-Wasserstein on Cartan-Hadamard manifolds (Bonet et al., 2023, 2025a)
    → Project on geodesics with the Busemann function
- Busemann used for alignments
  → Embedding with prototypes (Ghadimi Atigh et al., 2021)



Applications in Machine Learning

- Busemann used as a projection
  - HoroPCA (Chami et al., 2021)
    → Project on a geodesic subspace along horospheres
  - Sliced-Wasserstein on Cartan-Hadamard manifolds (Bonet et al., 2023, 2025a)
    → Project on geodesics with the Busemann function
- Busemann used for alignments
  → Embedding with prototypes (Ghadimi Atigh et al., 2021)
- Busemann used for classification (Fan et al., 2023; Doorenbos et al., 2024; Berg et al., 2024, 2025)
  → SVM, Random Forests, Logistic Regression...



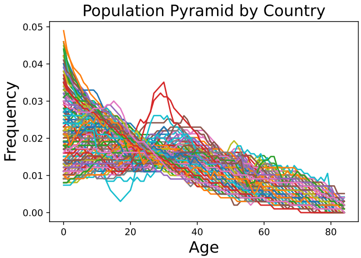
From (Berg et al., 2024)

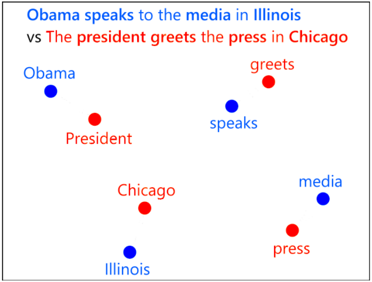
Table of Contents

Datasets of Distributions

Examples

- 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)





Wasserstein Geometry (Ambrosio et al., 2008)

Definition (Wasserstein distance)

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and denote by $\Pi(\mu,\nu)$ the set of coupling between μ, ν. Then, the Wasserstein distance is

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

with $\Pi(\mu,\nu) = \{\gamma \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d),\ \pi^1_\#\gamma = \mu,\ \pi^2_\#\gamma = \nu\}$, $\pi^1 : (x,y) \mapsto x$, $\pi^2 : (x,y) \mapsto y$

Remainder: For $T : \mathbb{R}^d \to \mathbb{R}^p$ measurable, $X \sim \mu \implies T(X) \sim T_\#\mu$

Wasserstein Geometry (Ambrosio et al., 2008)

Definition (Wasserstein distance)

Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ and denote by $\Pi(\mu, \nu)$ the set of coupling between μ, ν. Then, the Wasserstein distance is

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

with $\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}_2(\mathbb{R}^d \times \mathbb{R}^d),\ \pi^1_\#\gamma = \mu,\ \pi^2_\#\gamma = \nu\}$, $\pi^1 : (x, y) \mapsto x$, $\pi^2 : (x, y) \mapsto y$

Remainder: For $T : \mathbb{R}^d \to \mathbb{R}^p$ measurable, $X \sim \mu \implies T(X) \sim T_\#\mu$

Properties:

- W_2 distance, $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$: Wasserstein space
- $\mathrm{W}_2(\delta_x, \delta_y) = \|x - y\|_2$
- Riemannian structure
- Geodesic metric space

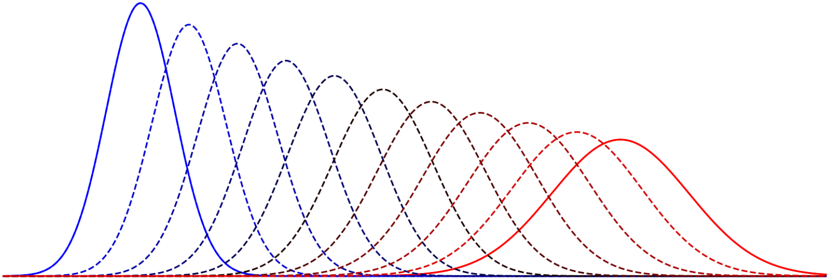
Geodesics in Wasserstein Space

Let $\Pi_o(\mu,\nu) = \{\gamma,\ \gamma \in \mathrm{argmin}_{\gamma\in\Pi(\mu,\nu)} \int \|x-y\|_2^2 \,\mathrm{d}\gamma(x,y)\}$

Geodesics

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

$$\forall t \in [0,1],\ \mu_t = \left((1-t)\pi^1 + t\pi^2\right)_{\#}\gamma \quad \text{for} \quad \gamma \in \Pi_o(\mu_0,\mu_1)$$



Geodesics in Wasserstein Space

Let $\Pi_o(\mu,\nu) = \{\gamma,\ \gamma \in \mathrm{argmin}_{\gamma\in\Pi(\mu,\nu)} \int \|x-y\|_2^2\ \mathrm{d}\gamma(x,y)\}$

Geodesics

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

$$\forall t \in [0,1],\ \mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\#\gamma \quad \text{for} \quad \gamma \in \Pi_o(\mu_0,\mu_1)$$

For all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, $t \in [0,1]$, (Ambrosio et al., 2008)

$$\mathrm{W}_2^2(\mu_t,\nu) \geq (1-t)\mathrm{W}_2^2(\mu_0,\nu) + t\mathrm{W}_2^2(\mu_1,\nu) - t(1-t)\mathrm{W}_2^2(\mu_0,\mu_1)$$

$\rightarrow$ Positively curved space (PC Space)

Geodesics in Wasserstein Space

Let $\Pi_o(\mu,\nu) = \{\gamma,\ \gamma \in \mathrm{argmin}_{\gamma\in\Pi(\mu,\nu)} \int \|x-y\|_2^2 \,\mathrm{d}\gamma(x,y)\}$

Geodesics

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

$$\forall t \in [0,1],\ \mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\# \gamma \quad \text{for} \quad \gamma \in \Pi_o(\mu_0,\mu_1)$$

For all $\nu \in \mathcal{P}_2(\mathbb{R}^d)$, $t\in[0,1]$, (Ambrosio et al., 2008)

$$\mathrm{W}_2^2(\mu_t,\nu) \geq (1-t)\mathrm{W}_2^2(\mu_0,\nu) + t\mathrm{W}_2^2(\mu_1,\nu) - t(1-t)\mathrm{W}_2^2(\mu_0,\mu_1)$$

→ Positively curved space (PC Space)

By (Zhu et al., 2021): always at least one geodesic ray starting from μ_0
→ Conditions to extend $t \mapsto \mu_t$ to $\mathbb{R}_+$?

1D Wasserstein Space

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:

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

1D Wasserstein Distance

The optimal coupling is $\gamma^* = (F_\mu^{-1}, F_\nu^{-1})_\# \mathrm{Unif}([0,1])$, and

$$\mathrm{W}_2^2(\mu,\nu) = \int_0^1 \big|F_\mu^{-1}(u) - F_\nu^{-1}(u)\big|^2\ \mathrm{d}u = \big\|F_\mu^{-1} - F_\nu^{-1}\big\|^2_{L^2([0,1])}$$

1D Wasserstein Space

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:

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

1D Wasserstein Distance

The optimal coupling is $\gamma^* = (F_\mu^{-1}, F_\nu^{-1})_\# \mathrm{Unif}([0,1])$, and

$$\mathrm{W}_2^2(\mu,\nu) = \int_0^1 \left|F_\mu^{-1}(u) - F_\nu^{-1}(u)\right|^2\ \mathrm{d}u = \left\|F_\mu^{-1} - F_\nu^{-1}\right\|_{L^2([0,1])}^2$$

Geodesic between μ and ν: $\forall t \in [0,1]$,

$$\mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\# \gamma^* = \big((1-t)F_\mu^{-1} + tF_\nu^{-1}\big)_\# \mathrm{Unif}([0,1])$$

$\to F_t^{-1} = (1-t)F_\mu^{-1} + tF_\nu^{-1}$

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

$$\forall t \in [0,1],\ F_t^{-1} = (1-t)F_0^{-1} + tF_1^{-1} = F_0^{-1} + t(F_1^{-1} - F_0^{-1}).$$

- For all $t, s \in \mathbb{R}$,

$$\|F_t^{-1} - F_s^{-1}\|_{L^2([0,1])}^2 = (t-s)^2 \|F_0^{-1} - F_1^{-1}\|_{L^2([0,1])}^2$$

$\rightarrow$ ok if F_t^{-1}, F_s^{-1} quantile functions, *i.e.* non-decreasing and left-continuous

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

$$\forall t \in [0,1],\ F_t^{-1} = (1-t)F_0^{-1} + tF_1^{-1} = F_0^{-1} + t(F_1^{-1} - F_0^{-1}).$$

- For all $t, s \in \mathbb{R}$,

$$\|F_t^{-1} - F_s^{-1}\|^2_{L^2([0,1])} = (t-s)^2 \|F_0^{-1} - F_1^{-1}\|^2_{L^2([0,1])}$$

$\rightarrow$ ok if F_t^{-1}, F_s^{-1} quantile functions, *i.e.* non-decreasing and left-continuous

- For all $t \geq 0$, $0 < m < m' < 1$,

$$\begin{aligned} F_t^{-1}(m) - F_t^{-1}(m') &= F_0^{-1}(m) - F_0^{-1}(m') \\ &\quad + t\big(F_1^{-1}(m) - F_0^{-1}(m) - (F_1^{-1}(m') - F_0^{-1}(m'))\big) \leq 0 \end{aligned}$$

Proposition (Kloeckner, 2010)

$t \mapsto \mu_t$ is a geodesic ray if and only if $F_1^{-1} - F_0^{-1}$ is non-decreasing.

Illustrations - 1D Gaussian Case

Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$, ϕ^{-1} quantile function of $\mathcal{N}(0,1)$.

$\rightarrow F_0^{-1} = m_0 + \sigma_0\phi^{-1}$, $F_1^{-1} = m_1 + \sigma_1\phi^{-1}$
$\rightarrow \forall t \in [0,1]$, $\mu_t = \mathcal{N}\big((1-t)m_0 + tm_1, ((1-t)\sigma_0 + t\sigma_1)^2\big)$

Illustrations - 1D Gaussian Case

Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$, ϕ^{-1} quantile function of $\mathcal{N}(0,1)$.

$\rightarrow F_0^{-1} = m_0 + \sigma_0 \phi^{-1}$, $F_1^{-1} = m_1 + \sigma_1 \phi^{-1}$
$\rightarrow \forall t \in [0,1]$, $\mu_t = \mathcal{N}\big((1-t)m_0 + tm_1, ((1-t)\sigma_0 + t\sigma_1)^2\big)$

For all $0 < m < m' < 1$,

$$\big((F_1^{-1} - F_0^{-1})(m') - (F_1^{-1} - F_0^{-1})(m)\big) = (\sigma_1 - \sigma_0)\big(\phi^{-1}(m') - \phi^{-1}(m)\big) \geq 0$$
$$\iff \sigma_1 \geq \sigma_0$$

$t \mapsto \mu_t$ is a geodesic ray if and only if $\sigma_1 \geq \sigma_0$.

Illustrations - 1D Gaussian Case

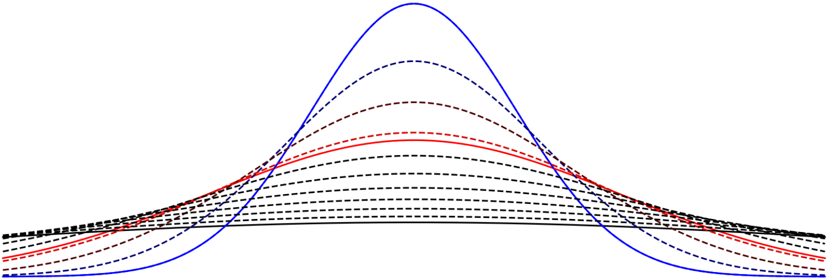
Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$, ϕ^{-1} quantile function of $\mathcal{N}(0,1)$.

$\rightarrow F_0^{-1} = m_0 + \sigma_0 \phi^{-1}$, $F_1^{-1} = m_1 + \sigma_1 \phi^{-1}$
$\rightarrow \forall t \in [0,1]$, $\mu_t = \mathcal{N}\big((1-t)m_0 + tm_1, ((1-t)\sigma_0 + t\sigma_1)^2\big)$

For all $0 < m < m' < 1$,

$$\big((F_1^{-1} - F_0^{-1})(m') - (F_1^{-1} - F_0^{-1})(m)\big) = (\sigma_1 - \sigma_0)\big(\phi^{-1}(m') - \phi^{-1}(m)\big) \geq 0$$
$$\iff \sigma_1 \geq \sigma_0$$

$t \mapsto \mu_t$ is a geodesic ray if and only if $\sigma_1 \geq \sigma_0$.



Illustrations - 1D Gaussian Case

Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$, ϕ^{-1} quantile function of $\mathcal{N}(0,1)$.

$\rightarrow F_0^{-1} = m_0 + \sigma_0 \phi^{-1}$, $F_1^{-1} = m_1 + \sigma_1 \phi^{-1}$

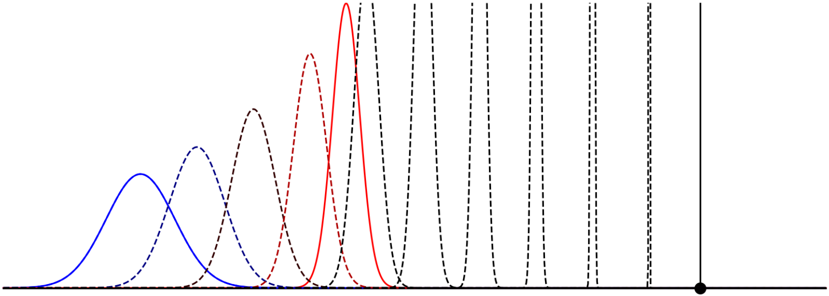
$\rightarrow \forall t \in [0,1]$, $\mu_t = \mathcal{N}\big((1-t)m_0 + tm_1, ((1-t)\sigma_0 + t\sigma_1)^2\big)$

For all $0 < m < m' < 1$,

$$\big((F_1^{-1} - F_0^{-1})(m') - (F_1^{-1} - F_0^{-1})(m)\big) = (\sigma_1 - \sigma_0)\big(\phi^{-1}(m') - \phi^{-1}(m)\big) \geq 0$$

$$\iff \sigma_1 \geq \sigma_0$$

$t \mapsto \mu_t$ is a geodesic ray if and only if $\sigma_1 \geq \sigma_0$.



Illustrations - Starting from a Dirac

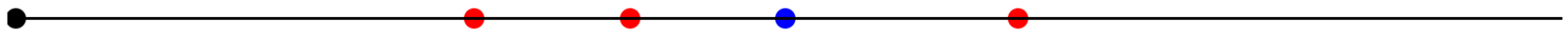
Let $x_0 \in \mathbb{R}$, $\mu_0 = \delta_{x_0}$, $F_0^{-1}(p) = x_0$ for all $1 \geq p > 0$.
$\rightarrow$ For all $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $F_1^{-1} - F_0^{-1}$ non-decreasing
$\rightarrow \gamma^* = \mu_0 \otimes \mu_1$, $\mu_t = \big((1-t)x_0 + t\mathrm{Id}\big)_{\#}\mu_1$

For all $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $t \mapsto \mu_t$ is a geodesic ray.

Illustrations - Starting from a Dirac

Let $x_0 \in \mathbb{R}$, $\mu_0 = \delta_{x_0}$, $F_0^{-1}(p) = x_0$ for all $1 \geq p > 0$.
$\rightarrow$ For all $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $F_1^{-1} - F_0^{-1}$ non-decreasing
$\rightarrow \gamma^* = \mu_0 \otimes \mu_1$, $\mu_t = \big((1-t)x_0 + t\mathrm{Id}\big)_{\#}\mu_1$

For all $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $t \mapsto \mu_t$ is a geodesic ray.



Illustrations - Starting from a Dirac

Let $x_0 \in \mathbb{R}$, $\mu_0 = \delta_{x_0}$, $F_0^{-1}(p) = x_0$ for all $1 \geq p > 0$.
$\rightarrow$ For all $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $F_1^{-1} - F_0^{-1}$ non-decreasing
$\rightarrow \gamma^* = \mu_0 \otimes \mu_1$, $\mu_t = \big((1-t)x_0 + t\mathrm{Id}\big)_{\#}\mu_1$

For all $\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $t \mapsto \mu_t$ is a geodesic ray.
$\rightarrow$ extends to $\mathbb{R}^d$ (Bertrand and Kloeckner, 2016, Lemma 2.1)

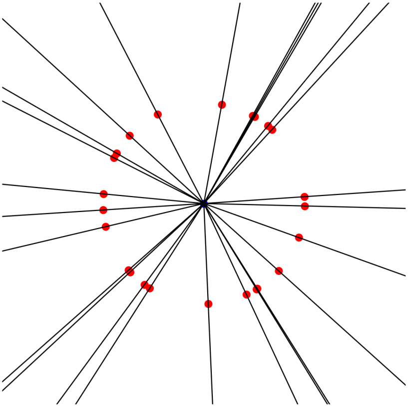


Illustration - Discrete Distributions

Let $x_1 < \cdots < x_n$, $y_1 < \cdots < y_n$, $\mu_0 = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\mu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$.

$t \mapsto \mu_t$ is a geodesic ray if and only if for all $j > i$, $y_i - x_i \leq y_j - x_j$.

Illustration - Discrete Distributions

Let $x_1 < \cdots < x_n$, $y_1 < \cdots < y_n$, $\mu_0 = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\mu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$.

$t \mapsto \mu_t$ is a geodesic ray if and only if for all $j > i$, $y_i - x_i \leq y_j - x_j$.

$$\mu_0 = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_0, \quad \mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{1.5}$$

$\to y_1 - x_1 = 1 < y_2 - x_2 = 1.5$

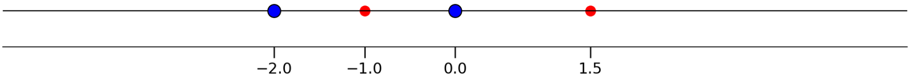


Illustration - Discrete Distributions

Let $x_1 < \cdots < x_n$, $y_1 < \cdots < y_n$, $\mu_0 = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\mu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$.

$t \mapsto \mu_t$ is a geodesic ray if and only if for all $j > i$, $y_i - x_i \leq y_j - x_j$.

$$\mu_0 = \frac{1}{2}\delta_{-2} + \frac{1}{2}\delta_0, \quad \mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{1.5}$$

$\rightarrow y_1 - x_1 = 1 < y_2 - x_2 = 1.5$

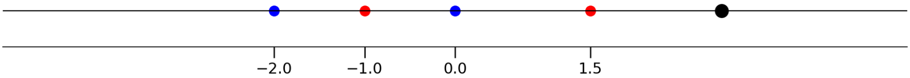


Illustration - Discrete Distributions

Let $x_1 < \cdots < x_n$, $y_1 < \cdots < y_n$, $\mu_0 = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$, $\mu_1 = \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$.

$t \mapsto \mu_t$ is a geodesic ray if and only if for all $j > i$, $y_i - x_i \leq y_j - x_j$.

$$\mu_0 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_4, \quad \mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{1.5}$$

$\rightarrow y_1 - x_1 = -1 > y_2 - x_2 = -2.5$

Not a ray (particles cross at $t > 0$)

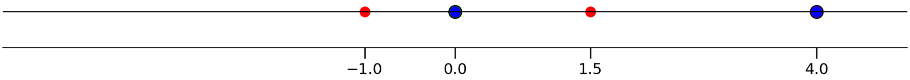


Illustration - Discrete Distributions

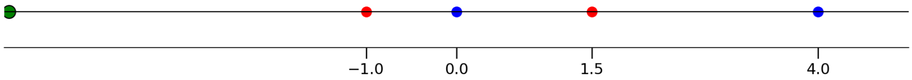
Let $x_1 < \cdots < x_n$, $y_1 < \cdots < y_n$, $\mu_0 = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\mu_1 = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$.

$t \mapsto \mu_t$ is a geodesic ray if and only if for all $j > i$, $y_i - x_i \leq y_j - x_j$.

$$\mu_0 = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_4, \quad \mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{1.5}$$

$\to y_1 - x_1 = -1 > y_2 - x_2 = -2.5$

Not a ray (particles cross at $t > 0$)



Geodesic Rays in Brenier's Setting

Proposition (Brenier's Theorem (Brenier, 1991))

$\mu_0 \ll \mathrm{Leb} \implies$ *Optimal coupling* γ^* *unique and* $\gamma^* = (\mathrm{Id}, \nabla\varphi)_\#\mu_0$ *with* φ *convex*

In this setting:

- Geodesic between $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ and $\mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ unique
- For all $t \in [0,1]$, $\mu_t = \big((1-t)\mathrm{Id} + t\mathrm{T}\big)_\#\mu_0$ with $\mathrm{T}_\#\mu_0 = \mu_1$, $\mathrm{T} = \nabla\varphi$, φ convex

Geodesic Rays in Brenier's Setting

Proposition (Brenier's Theorem (Brenier, 1991))

$\mu_0 \ll \mathrm{Leb} \implies$ *Optimal coupling* γ^* *unique and* $\gamma^* = (\mathrm{Id}, \nabla\varphi)_\# \mu_0$ *with* φ *convex*

In this setting:

- Geodesic between $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$ and $\mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$ unique
- For all $t \in [0,1]$, $\mu_t = \big((1-t)\mathrm{Id} + t\mathrm{T}\big)_\# \mu_0$ with $\mathrm{T}_\#\mu_0 = \mu_1$, $\mathrm{T} = \nabla\varphi$, φ convex
- If $t \mapsto \mu_t$ is a geodesic ray:

$$\begin{aligned} \forall s \geq 0,\ \mathrm{W}_2^2(\mu_s, \mu_0) &= s^2 \mathrm{W}_2^2(\mu_1, \mu_0) \\ &= \int \|s(x - \nabla\varphi(x))\|_2^2 \, \mathrm{d}\mu_0(x) \\ &= \int \|x - (1-s)x - s\nabla\varphi(x)\|_2^2 \, \mathrm{d}\mu_0(x) \end{aligned}$$

$\rightarrow (1-s)x + s\nabla\varphi(x) = \nabla\big(x \mapsto (1-s)\frac{\|x\|_2^2}{2} + s\varphi(x)\big)(x)$ must be the gradient of a convex function for all $s \geq 0$

$\rightarrow$ true if and only if $\varphi - \frac{\|x\|_2^2}{2}$ convex

Geodesic Rays in Brenier's Setting

Let $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, $\mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, $\mu_1 = (\nabla\varphi)_\#\mu_0$, φ convex.
$\to t \mapsto \mu_t$ geodesic ray if and only if φ 1-strongly convex

Examples

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

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

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

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

Busemann Function in the Wasserstein Space

Let $\mu_0 \in \mathcal{P}_2(\mathbb{R}^d)$, $t \mapsto \mu_t$ a geodesic ray starting from μ_0

Busemann function in $\big(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2\big)$:

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

with $\kappa_\mu = \mathrm{W}_2(\mu_0, \mu_1)$.

$\rightarrow$ Computation?

Computing the Busemann Function

Define $\Gamma(\mu_0, \mu_1, \nu) = \{\tilde{\gamma} \in \Pi(\mu_0, \mu_1, \nu),\ \pi^{1,2}_{\#}\tilde{\gamma} \in \Pi_o(\mu_0, \mu_1)\}$

- General case:

$$B^\mu(\nu) = \inf_{\tilde{\gamma}\in\Gamma(\mu_0,\mu_1,\nu)} -\kappa_\mu^{-1} \int \langle x_1 - x_0, y - x_0 \rangle \, \mathrm{d}\tilde{\gamma}(x_0, x_1, y)$$

Computing the Busemann Function

Define $\Gamma(\mu_0,\mu_1,\nu) = \big\{\tilde{\gamma} \in \Pi(\mu_0,\mu_1,\nu),\ \pi^{1,2}_{\#}\tilde{\gamma} \in \Pi_o(\mu_0,\mu_1)\big\}$

- General case:

$$B^\mu(\nu) = \inf_{\tilde{\gamma}\in\Gamma(\mu_0,\mu_1,\nu)} \ -\kappa_\mu^{-1} \int \langle x_1 - x_0, y - x_0\rangle \ \mathrm{d}\tilde{\gamma}(x_0,x_1,y)$$

- If $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, $\mu_1 = (\nabla\varphi)_\#\mu_0$, φ 1-strongly convex:

$$B^\mu(\nu) = \inf_{\gamma\in\Pi(\mu_0,\nu)} \ -\kappa_\mu^{-1} \int \langle \nabla\varphi(x_0) - x_0, y - x_0\rangle \ \mathrm{d}\gamma(x_0,y)$$

$\rightarrow$ equivalent to OT problem with cost $c(x_0,y) = \|\nabla\varphi(x_0) - x_0 - y\|_2^2$

Computing the Busemann Function

Define $\Gamma(\mu_0, \mu_1, \nu) = \{\tilde{\gamma} \in \Pi(\mu_0, \mu_1, \nu),\ \pi^{1,2}_{\#}\tilde{\gamma} \in \Pi_o(\mu_0, \mu_1)\}$

- General case:

$$B^\mu(\nu) = \inf_{\tilde{\gamma} \in \Gamma(\mu_0, \mu_1, \nu)} -\kappa_\mu^{-1} \int \langle x_1 - x_0, y - x_0 \rangle \, \mathrm{d}\tilde{\gamma}(x_0, x_1, y)$$

- If $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, $\mu_1 = (\nabla\varphi)_{\#}\mu_0$, φ 1-strongly convex:

$$B^\mu(\nu) = \inf_{\gamma \in \Pi(\mu_0, \nu)} -\kappa_\mu^{-1} \int \langle \nabla\varphi(x_0) - x_0, y - x_0 \rangle \, \mathrm{d}\gamma(x_0, y)$$

- If $\mu_0 = \delta_{x_0}$, $x_0 \in \mathbb{R}^d$, $\pi^{1,2}_{\#}\tilde{\gamma} = \mu_0 \otimes \mu_1$,

$$B^\mu(\nu) = \inf_{\gamma \in \Pi(\mu_1, \nu)} -\kappa_\mu^{-1} \int \langle x_1 - x_0, y - x_0 \rangle \, \mathrm{d}\gamma(x_1, y)$$

→ Equivalent to $\mathrm{W}_2^2(\mu_1, \nu)$
→ For $\mu_1 = \delta_{x_1}$, $\theta = x_1 - x_0 \in S^{d-1}$, $\gamma(t) = x_0 + t\theta$, $B^\mu(\nu) = \int B^\gamma(y)\mathrm{d}\nu(y)$

Computing the Busemann Function

Define $\Gamma(\mu_0,\mu_1,\nu) = \left\{\tilde{\gamma} \in \Pi(\mu_0,\mu_1,\nu),\ \pi^{1,2}_{\#}\tilde{\gamma} \in \Pi_o(\mu_0,\mu_1)\right\}$

- General case:

$$B^\mu(\nu) = \inf_{\tilde{\gamma}\in\Gamma(\mu_0,\mu_1,\nu)} -\kappa_\mu^{-1} \int \langle x_1 - x_0, y - x_0\rangle \ \mathrm{d}\tilde{\gamma}(x_0,x_1,y)$$

- If $\mu_0 \in \mathcal{P}_{2,\mathrm{ac}}(\mathbb{R}^d)$, $\mu_1 = (\nabla\varphi)_\#\mu_0$, φ 1-strongly convex:

$$B^\mu(\nu) = \inf_{\gamma\in\Pi(\mu_0,\nu)} -\kappa_\mu^{-1} \int \langle \nabla\varphi(x_0) - x_0, y - x_0\rangle \ \mathrm{d}\gamma(x_0,y)$$

- If $\mu_0 = \delta_{x_0}$, $x_0 \in \mathbb{R}^d$, $\pi^{1,2}_{\#}\tilde{\gamma} = \mu_0 \otimes \mu_1$,

$$B^\mu(\nu) = \inf_{\gamma\in\Pi(\mu_1,\nu)} -\kappa_\mu^{-1} \int \langle x_1 - x_0, y - x_0\rangle \ \mathrm{d}\gamma(x_1,y)$$

$\rightarrow$ **Linear programs**

Busemann Function in the 1D Wasserstein Space

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

$$\mathrm{W}_2^2(\mu_0, \mu_1) = \int_0^1 |F_0^{-1}(u) - F_1^{-1}(u)|^2 \, \mathrm{d}u = \|F_0^{-1} - F_1^{-1}\|^2_{L^2([0,1])}$$

$\rightarrow$ Hilbert structure

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

$$\begin{aligned} \forall \nu \in \mathcal{P}_2(\mathbb{R}),\ B^\mu(\nu) &= -\int_0^1 \big(F_1^{-1}(u) - F_0^{-1}(u)\big)\big(F_\nu^{-1}(u) - F_0^{-1}(u)\big) \, \mathrm{d}u \\ &= -\langle F_1^{-1} - F_0^{-1}, F_\nu^{-1} - F_0^{-1} \rangle_{L^2([0,1])}. \end{aligned}$$

Example

Let $\mu_0 = \mathcal{N}(m_0, \sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1, \sigma_1^2)$, $\nu = \mathcal{N}(m, \sigma^2)$,

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

Busemann Function on the Bures-Wasserstein Space

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

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

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 $\nu = \mathcal{N}(m, \Sigma)$,

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

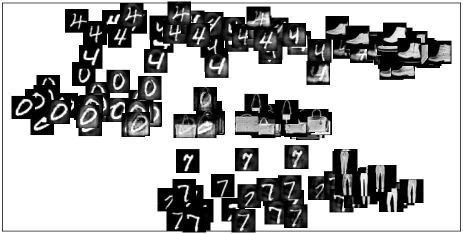
where $A = \Sigma_0^{-\frac{1}{2}}(\Sigma_0^{\frac{1}{2}}\Sigma_1\Sigma_0^{\frac{1}{2}})^{\frac{1}{2}}\Sigma_0^{-\frac{1}{2}}$.

Table of Contents

Labeled Datasets

$\mathcal{D}_1 : \mu_1 = \frac{1}{m}\sum_{i=1}^{m} \delta_{(x_i^1, y_i^1)} \in \mathcal{P}(\mathbb{R}^d \times \{1, \ldots, C\}),$
$\mathcal{D}_2 : \mu_2 = \frac{1}{m}\sum_{j=1}^{m} \delta_{(x_j^2, y_j^2)} \in \mathcal{P}(\mathbb{R}^d \times \{1, \ldots, C\})$
C: number of classes, n: number of sample in each class, $m = nC$

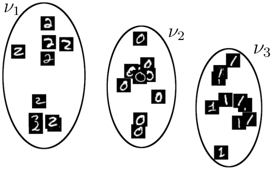
Question: how to compare datasets $\mathcal{D}_1$ and $\mathcal{D}_2$?



OTDD (Alvarez-Melis and Fusi, 2020)

Solution of Alvarez-Melis and Fusi (2020):

- Embed a label (a class) in $\mathcal{P}(\mathbb{R}^d)$ as $c \mapsto \nu_c^k = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^k}\mathbb{1}_{\{y_i^k=c\}}$ for $k=1,2$

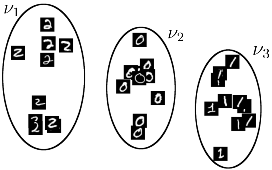


$\rightarrow \mathcal{D}_k : \mu_k = \frac{1}{m}\sum_{i=1}^m \delta_{(x_i^k, \nu^k_{y_i^k})} \in \mathcal{P}\big(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)\big)$

OTDD (Alvarez-Melis and Fusi, 2020)

Solution of Alvarez-Melis and Fusi (2020):

- Embed a label (a class) in $\mathcal{P}(\mathbb{R}^d)$ as $c \mapsto \nu_c^k = \frac{1}{n}\sum_{i=1}^n \delta_{x_i^k} \mathbb{1}_{\{y_i^k = c\}}$ for $k = 1, 2$



$\rightarrow \mathcal{D}_k : \mu_k = \frac{1}{m}\sum_{i=1}^m \delta_{(x_i^k, \nu_{y_i^k}^k)} \in \mathcal{P}\big(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)\big)$

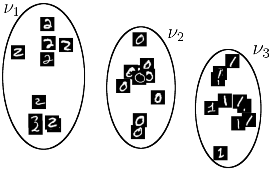
- Cost: $d\big((x,y),(x',y')\big)^2 = \|x - x'\|_2^2 + \mathrm{W}_2^2(\nu_y, \nu_{y'})$
- **Optimal transport distance**: $O\big(C^2 n^3 \log n + n^3 C^3 \log(nC)\big)$

$$\mathrm{OTDD}(\mu_1, \mu_2) = \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int d\big((x,y),(x',y')\big)^2 \, \mathrm{d}\gamma\big((x,y),(x',y')\big).$$

OTDD (Alvarez-Melis and Fusi, 2020)

Solution of Alvarez-Melis and Fusi (2020):

- Embed a label (a class) in $\mathbb{R}^p \times S_p^{++}(\mathbb{R})$ as $c \mapsto \nu_c^k \approx \mathcal{N}(m_c^k, \Sigma_c^k)$ for $k = 1, 2$

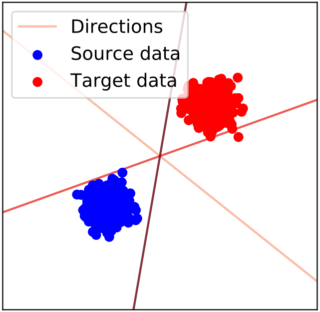


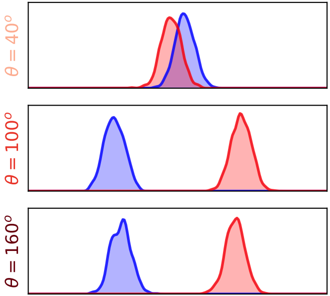
$\rightarrow \mathcal{D}_k : \mu_k = \frac{1}{m} \sum_{i=1}^m \delta_{(x_i^k, m^k_{y_i^k}, \Sigma^k_{y_i^k})} \in \mathcal{P}\big(\mathbb{R}^d \times \mathbb{R}^p \times S_p^{++}(\mathbb{R})\big)$

- Cost: $d\big((x,y),(x',y')\big)^2 = \|x - x'\|_2^2 + \mathrm{BW}_2^2(\nu_y, \nu_{y'})$
- **Optimal transport distance**: approximated in $O\big(C^2 d^3 + n^2 C^2 \log(nC)/\varepsilon^2\big)$

$$\mathrm{OTDD}_\varepsilon(\mu_1, \mu_2) = \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int d\big((x,y),(x',y')\big)^2 \, \mathrm{d}\gamma\big((x,y),(x',y')\big) + \varepsilon \mathcal{H}(\gamma).$$

Sliced-Wasserstein Distance





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

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

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

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

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:

$$\widehat{\mathrm{SW}}_{2,L}^2(\mu, \nu) = \frac{1}{L} \sum_{\ell=1}^{L} \mathrm{W}_2^2(P_\#^{\theta_\ell}\mu, P_\#^{\theta_\ell}\nu),$$

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

$\rightarrow$ Computational complexity: $O\big(Ln(\log n + d)\big)$

Goal: Define a SW distance on $\mathcal{P}_2\big(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)\big)$

Previous Methods

- Sliced-Wasserstein on $\mathcal{P}_2(X \times Y)$ (Nguyen and Ho, 2024):
  - Define 2 projections $P^\theta : X \to \mathbb{R}$, $Q^\phi : Y \to \mathbb{R}$
  - For $\alpha \in S^1$, define

$$\forall (x,y) \in X \times Y,\ P^{\alpha,\theta,\phi}(x,y) = \alpha_1 P^\theta(x) + \alpha_2 Q^\phi(y)$$

  - For $\mu, \nu \in \mathcal{P}_2(X \times Y)$,

$$\mathrm{SW}_2^2(\mu,\nu) = \int \mathrm{W}_2^2(P_\#^{\alpha,\theta,\phi}\mu, P_\#^{\alpha,\theta,\phi}\nu)\ \mathrm{d}\lambda(\alpha,\theta,\phi)$$

Previous Methods

- Sliced-Wasserstein on $\mathcal{P}_2(X \times Y)$ (Nguyen and Ho, 2024):
  - Define 2 projections $P^\theta : X \to \mathbb{R}$, $Q^\phi : Y \to \mathbb{R}$
  - For $\alpha \in S^1$, define

$$\forall (x,y) \in X \times Y,\ P^{\alpha,\theta,\phi}(x,y) = \alpha_1 P^\theta(x) + \alpha_2 Q^\phi(y)$$

  - For $\mu, \nu \in \mathcal{P}_2(X \times Y)$,

$$\mathrm{SW}_2^2(\mu,\nu) = \int \mathrm{W}_2^2(P^{\alpha,\theta,\phi}_\# \mu, P^{\alpha,\theta,\phi}_\# \nu)\ \mathrm{d}\lambda(\alpha,\theta,\phi)$$

- Sliced-Wasserstein on $\mathcal{P}_2\big(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)\big)$ (SOTDD) (Nguyen et al., 2025)
  - For a label $y \in \{1,\dots,C\}$, define $\varphi(y) = \frac{1}{n_y}\sum_{i=1}^n \delta_{x_i}\mathbb{1}_{\{y_i=y\}}$
  - Use for $\alpha \in S^k$,

$$P^{\alpha,\theta,\lambda}(x,y) = \alpha_1 P^\theta(x) + \sum_{i=1}^k \alpha_{i+1}\mathcal{M}^{\lambda_i}\big(P^\theta_\# \varphi(y)\big),$$

  with $P^\theta : \mathbb{R}^d \to \mathbb{R}$, $\mathcal{M}^\lambda : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ the moment transform projection.

Previous Methods

- Sliced-Wasserstein on $\mathcal{P}_2(X \times Y)$ (Nguyen and Ho, 2024):
  - Define 2 projections $P^\theta : X \to \mathbb{R}$, $Q^\phi : Y \to \mathbb{R}$
  - For $\alpha \in S^1$, define

$$\forall (x,y) \in X \times Y,\ P^{\alpha,\theta,\phi}(x,y) = \alpha_1 P^\theta(x) + \alpha_2 Q^\phi(y)$$

  - For $\mu, \nu \in \mathcal{P}_2(X \times Y)$,

$$\mathrm{SW}_2^2(\mu,\nu) = \int \mathrm{W}_2^2(P^{\alpha,\theta,\phi}_\# \mu, P^{\alpha,\theta,\phi}_\# \nu)\ \mathrm{d}\lambda(\alpha,\theta,\phi)$$

- Sliced-Wasserstein on $\mathcal{P}_2\big(\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)\big)$ (SOTDD) (Nguyen et al., 2025)
  - For a label $y \in \{1,\dots,C\}$, define $\varphi(y) = \frac{1}{n_y}\sum_{i=1}^n \delta_{x_i}\mathbb{1}_{\{y_i=y\}}$
  - Use for $\alpha \in S^k$,

$$P^{\alpha,\theta,\lambda}(x,y) = \alpha_1 P^\theta(x) + \sum_{i=1}^k \alpha_{i+1}\mathcal{M}^{\lambda_i}\big(P^\theta_\#\varphi(y)\big),$$

  with $P^\theta : \mathbb{R}^d \to \mathbb{R}$, $\mathcal{M}^\lambda : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ the moment transform projection.

$\rightarrow$ **Use B^μ for projecting distributions on $\mathbb{R}$**

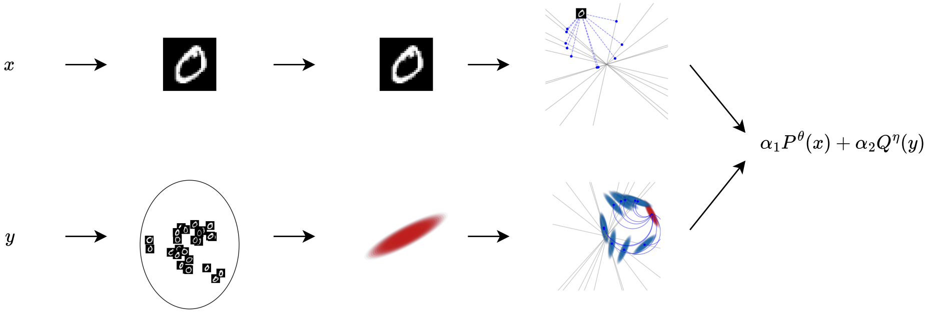
Slicing Datasets with Busemann on Gaussian

With Gaussian approximation:

- Define $\Xi(\mu) = \mathcal{N}\big(m(\mu), \Sigma(\mu)\big)$
- For all $y \in \{1, \dots, C\}$,

$$Q^\eta(y) = B^\eta\big(\Xi(\varphi(y))\big),$$

with η a geodesic ray on $BW(\mathbb{R}^d)$



Computational Complexity: $O\big(LCd^3 + Ln_yC(\log n_yC + d) + d^2Cn_y\big)$

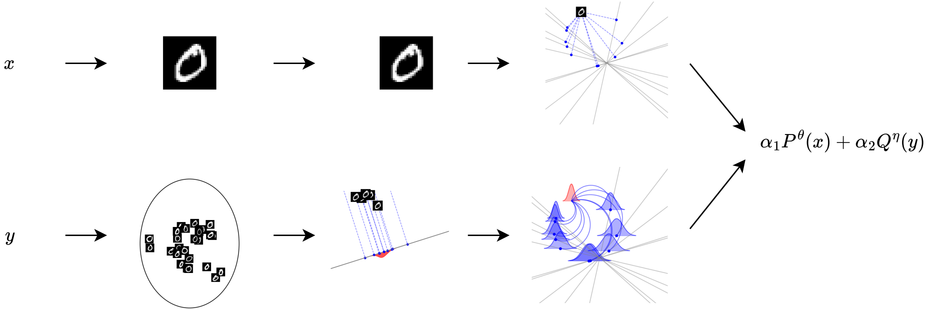
Slicing Datasets with Busemann in 1D

With 1D Projection:

- For all $y \in \{1, \dots, C\}$,

$$Q^{\eta,\theta}(y) = B^\eta\big(P^\theta_\# \varphi(y)\big),$$

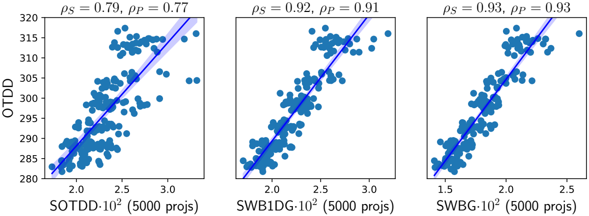
with η a geodesic ray on $\mathcal{P}_2(\mathbb{R})$



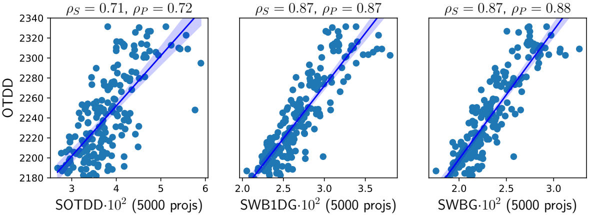
Computation Complexity: $O\big(L n_y C(\log(n_y C + d))\big)$

Correlation vs OTDD

Goal: Measure correlation between sliced distances and OTDD
$\rightarrow$ Compare randomly sampled subdatasets + Spearman and Pearson correlations



MNIST

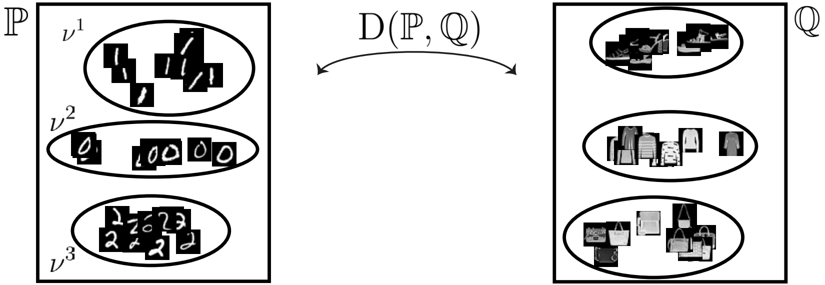


CIFAR10

Flowing Labeled Datasets

- Model datasets as $\mathbb{P} = \frac{1}{C}\sum_{c=1}^{C}\delta_{\nu^c} \in \mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$ where $\nu^c = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i^c}$
- Flow a dataset $\mathbb{P}$ towards $\mathbb{Q}$ by minimizing a discrepancy D on $\mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$

$\rightarrow$ minimization problem on $\mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$



Example

Let $\psi\big(\frac{1}{C}\sum_{c=1}^{C}\delta_{\mu^{n,c}}\big) = \frac{1}{nC}\sum_{c=1}^{C}\sum_{i=1}^{n}\delta_{(x_i^c,\mu^{n,c})}$ where $\mu^{n,c} = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i^c}$.
$\rightarrow \mathrm{D}(\mathbb{P},\mathbb{Q}) = \mathrm{SWB1DG}(\mathbb{P},\mathbb{Q})$

Minimizing on $\mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$ (Bonet et al., 2025b)

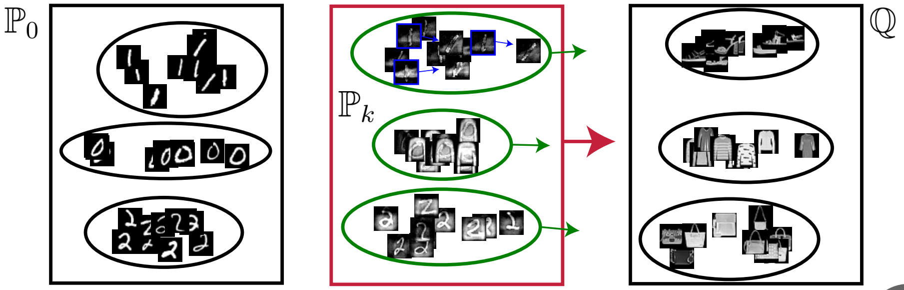
Goal: minimize $\mathbb{F} : \mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big) \to \mathbb{R}$

In practice: For $\mathbb{P}_k = \frac{1}{C}\sum_{c=1}^C \delta_{\mu_k^{c,n}}$ with $\mu_k^{c,n} = \frac{1}{n}\sum_{i=1}^n \delta_{x_{i,k}^c} \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\forall k \geq 0, \text{ particle (image) } i, \text{ class } c, \ x_{i,k+1}^c = x_{i,k}^c - \tau \nabla_{\mathrm{W}_{\mathrm{W}_2}} \mathbb{F}(\mathbb{P}_k)(\mu_k^{c,n})(x_{i,k}^c).$$

$\mathbb{P}_k$: inter-class interaction, $\mu_k^{c,n}$: intra-class interaction, $x_{i,k}^c$ image

$\nabla_{\mathrm{W}_{\mathrm{W}_2}} \mathbb{F}(\mathbb{P}_k)(\mu_k^{c,n})(x_{i,k}^c) = nC[\nabla F(\mathbf{x})]_{i,c}$ with $F(\mathbf{x}) = \mathbb{F}(\mathbb{P}_k)$, $\mathbf{x} = (x_{i,k}^c)_{i,c}$



Synthetic Data

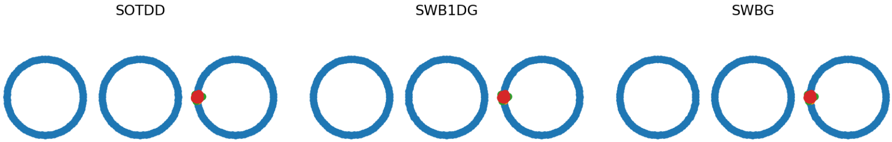
Goal: minimize $\mathbb{F} : \mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big) \to \mathbb{R}$

In practice: For $\mathbb{P}_k = \frac{1}{C}\sum_{c=1}^{C} \delta_{\mu_k^{c,n}}$ with $\mu_k^{c,n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_{i,k}^c} \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\forall k \geq 0,\ \text{particle (image) } i,\ \text{class } c,\ x_{i,k+1}^c = x_{i,k}^c - \tau \nabla_{\mathrm{W}_{\mathrm{W}_2}} \mathbb{F}(\mathbb{P}_k)(\mu_k^{c,n})(x_{i,k}^c).$$

Let $\mathbb{Q} = \frac{1}{3}\sum_{c=1}^{3} \delta_{\nu^c}$, ν^c ring

$$\min_{\mathbb{P}}\ \mathbb{F}(\mathbb{P}) = D(\mathbb{P}, \mathbb{Q})$$



Synthetic Data

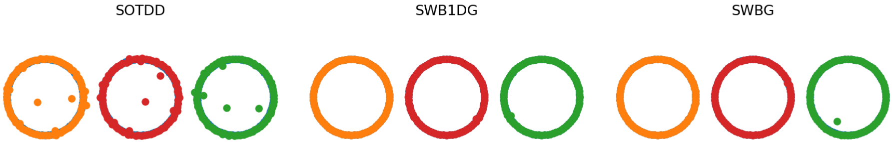
Goal: minimize $\mathbb{F} : \mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big) \to \mathbb{R}$

In practice: For $\mathbb{P}_k = \frac{1}{C}\sum_{c=1}^{C} \delta_{\mu_k^{c,n}}$ with $\mu_k^{c,n} = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_{i,k}^c} \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\forall k \geq 0,\ \text{particle (image) } i,\ \text{class } c,\ x_{i,k+1}^c = x_{i,k}^c - \tau \nabla_{\mathrm{W}_{\mathrm{W}_2}} \mathbb{F}(\mathbb{P}_k)(\mu_k^{c,n})(x_{i,k}^c).$$

Let $\mathbb{Q} = \frac{1}{3}\sum_{c=1}^{3} \delta_{\nu^c}$, ν^c ring

$$\min_{\mathbb{P}}\ \mathbb{F}(\mathbb{P}) = D(\mathbb{P}, \mathbb{Q})$$



Synthetic Data

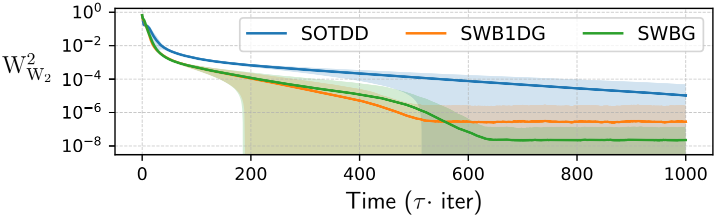
Goal: minimize $\mathbb{F} : \mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big) \to \mathbb{R}$

In practice: For $\mathbb{P}_k = \frac{1}{C}\sum_{c=1}^C \delta_{\mu_k^{c,n}}$ with $\mu_k^{c,n} = \frac{1}{n}\sum_{i=1}^n \delta_{x_{i,k}^c} \in \mathcal{P}_2(\mathbb{R}^d)$:

$$\forall k \geq 0, \text{ particle (image) } i, \text{ class } c, \ x_{i,k+1}^c = x_{i,k}^c - \tau \nabla_{\mathrm{W}_{\mathrm{W}_2}} \mathbb{F}(\mathbb{P}_k)(\mu_k^{c,n})(x_{i,k}^c).$$

Let $\mathbb{Q} = \frac{1}{3}\sum_{c=1}^3 \delta_{\nu^c}$, ν^c ring

$$\min_{\mathbb{P}} \ \mathbb{F}(\mathbb{P}) = D(\mathbb{P}, \mathbb{Q})$$



Conclusion

Conclusion:

- Studied geometry of geodesics on the Wasserstein space
- Studied the existence and computation of the Busemann function on the Wasserstein space
- Defined new SW distances to compare labeled Datasets
- Also in the paper: SW distances to compare Gaussian mixtures
- Related to new sliced distances on Gaussian mixtures (Baouan et al., 2025; Piening and Beinert, 2025a) and on $\mathcal{P}_2\big(\mathcal{P}_2(\mathbb{R}^d)\big)$ (Piening and Beinert, 2025b)

Perspectives:

- Applications to PCA on the Wasserstein space
- Extension to $\mathcal{P}_2(\mathbb{L}^d)$
- Busemann along other curves (Gallouët et al., 2025)

Thank you!

Paper: https://arxiv.org/abs/2510.04579

References I

David Alvarez-Melis and Nicolo Fusi. Geometric Dataset Distances via Optimal Transport. *Advances in Neural Information Processing Systems*, 33: 21428–21439, 2020.

Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. *Gradient flows: in metric spaces and in the space of probability measures*. Springer, 2008.

Ali Baouan, Mathieu Rosenbaum, and Sergio Pulido. An optimal transport based embedding to quantify the distance between playing styles in collective sports. *Journal of Quantitative Analysis in Sports*, 2025.

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.

Paul Berg, Bjoern Michele, Minh-Tan Pham, Laetitia Chapel, and Nicolas Courty. Horospherical Learning with Smart Prototypes. In *British Machine Vision Conference (BMVC)*, 2024.

Paul Berg, Léo Buecher, Björn Michele, Minh-Tan Pham, Laetitia Chapel, and Nicolas Courty. Multi-Prototype Hyperbolic Learning Guided by Class Hierarchy. *International Journal of Computer Vision*, pages 1–16, 2025.

References II

Jérôme Bertrand and Benoît Kloeckner. A geometric study of Wasserstein spaces: isometric rigidity in negative curvature. *International Mathematics Research Notices*, 2016(5):1368–1386, 2016.

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, 2023.

Clément Bonet, Lucas Drumetz, and Nicolas Courty. Sliced-Wasserstein Distances and Flows on Cartan-Hadamard Manifolds. *Journal of Machine Learning Research*, 26(32):1–76, 2025a.

Clément Bonet, Christophe Vauthier, and Anna Korba. Flowing Datasets with Wasserstein over Wasserstein Gradient Flows. In *International Conference on Machine Learning*. PMLR, 2025b.

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

Martin R Bridson and André Haefliger. *Metric Spaces of Non-Positive Curvature*, volume 319. Springer Science & Business Media, 2013.

References III

Herbert Busemann. *The Geometry of Geodesics*. Academic Press, New York, 1955.

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.

Julie Delon and Agnes Desolneux. A Wasserstein-Type Distance in the space of Gaussian Mixture Models. *SIAM Journal on Imaging Sciences*, 13(2):936–970, 2020.

Lars Doorenbos, Pablo Márquez Neila, Raphael Sznitman, and Pascal Mettes. Hyperbolic Random Forests. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856.

Xiran Fan, Chun-Hao Yang, and Baba Vemuri. Horocycle Decision Boundaries for Large Margin Classification in Hyperbolic Space. *Advances in neural information processing systems*, 36:11194–11204, 2023.

Thomas O Gallouët, Andrea Natale, and Gabriele Todeschi. Metric extrapolation in the wasserstein space. *Calculus of Variations and Partial Differential Equations*, 64(5):147, 2025.

References IV

Jonathan Geuter, Clément Bonet, Anna Korba, and David Alvarez-Melis. DDEQs: Distributional Deep Equilibrium Models through Wasserstein Gradient Flows. In *The 28th International Conference on Artificial Intelligence and Statistics*, 2025.

Mina Ghadimi Atigh, Martin Keller-Ressel, and Pascal Mettes. Hyperbolic Busemann Learning with Ideal Prototypes. *Advances in Neural Information Processing Systems*, 34:103–115, 2021.

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.

Khai Nguyen and Nhat Ho. Hierarchical Hybrid Sliced Wasserstein: A Scalable Metric for Heterogeneous Joint Distributions. *Advances in Neural Information Processing Systems*, 37:108140–108166, 2024.

Khai Nguyen, Hai Nguyen, Tuan Pham, and Nhat Ho. Lightspeed Geometric Dataset Distance via Sliced Optimal Transport. In *Forty-second International Conference on Machine Learning*, 2025.

References V

Moritz Piening and Robert Beinert. Slicing the Gaussian Mixture Wasserstein Distance. *Transactions on Machine Learning Research*, 2025a. ISSN 2835-8856.

Moritz Piening and Robert Beinert. Slicing Wasserstein Over Wasserstein Via Functional Optimal Transport. *arXiv preprtint arXiv:2509.22138*, 2025b.

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, 2011.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The Earth Mover's Distance as a Metric for Image Retrieval. *International journal of computer vision*, 40(2): 99–121, 2000.

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.

Guomin Zhu, Wen-Long Li, and Xiaojun Cui. Busemann functions on the wasserstein space. *Calculus of Variations and Partial Differential Equations*, 60 (3):97, 2021.