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

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

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

## Contributions

**Goal:** Compare labeled datasets represented as $\mathbb{P} = \frac{1}{n}\sum_{i=1}^n \delta_{(x_i,\varphi(y_i))} \in \mathcal{P}\big(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)\big)$, $x_i \in \mathbb{R}^d$, $y_i \in \{1,\dots,C\}$, $\varphi(y) = \frac{1}{n_y}\sum_{i=1}^n \delta_{x_i}1_{\{y_i=y\}}$.

- Study how to extend geodesics on $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$
- Busemann function on $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$
- Define Sliced-Wasserstein distances on $\mathcal{P}\big(\mathbb{R}^d \times \mathcal{P}(\mathbb{R}^d)\big)$

## Optimal Transport

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

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

→ For $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\hat{\nu}_n = \frac{1}{n}\sum_{j=1}^n \delta_{y_j} \in \mathcal{P}_2(\mathbb{R}^d)$, complexity $\mathcal{O}(n^3\log n)$

For $\mu,\nu \in \mathcal{P}_2(\mathbb{R})$, $\mathrm{W}_2^2(\mu,\nu) = \|F_\mu^{-1} - F_\nu^{-1}\|^2_{L^2([0,1])}$
→ For $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, $\hat{\nu}_n = \frac{1}{n}\sum_{j=1}^n \delta_{y_j} \in \mathcal{P}_2(\mathbb{R})$, complexity $\mathcal{O}(n\log n)$

If $\mu \ll \mathrm{Leb}$, $\mathrm{W}_2^2(\mu,\nu) = \inf_{\mathrm{T}_\#\mu=\nu} \int \|x-\mathrm{T}(x)\|_2^2\ \mathrm{d}\mu(x) = \|\mathrm{T}_\mu^\nu - \mathrm{Id}\|^2_{L^2(\mu)}$

**Riemannian structure.**
Geodesic between $\mu_0,\mu_1$: $\forall t\in[0,1]$, $\mu_t = \big((1-t)\pi^1 + t\pi^2\big)_\#\gamma$, $\gamma \in \Pi_o(\mu_0,\mu_1)$
Satisfies for all $s,t\in[0,1]$, $\mathrm{W}_2(\mu_s,\mu_t) = |t-s|\mathrm{W}_2(\mu_0,\mu_1)$
If $\mu_0 \ll \mathrm{Leb}$, $\forall t\in[0,1]$, $\mu_t = \big((1-t)\mathrm{Id} + t\mathrm{T}_{\mu_0}^{\mu_1}\big)_\#\mu_0$
If $\mu_0,\mu_1 \in \mathcal{P}_2(\mathbb{R})$, $\forall t\in[0,1]$, $\mu_t = \big((1-t)F_{\mu_0}^{-1} + tF_{\mu_1}^{-1}\big)_\#\mathrm{Unif}([0,1])$

Conditions to be a **geodesic ray**, *i.e.* to extend $t\mapsto\mu_t$ on $\mathbb{R}_+$:

- If $\mu_0 = \delta_{x_0}$, $x_0\in\mathbb{R}^d$, true for all $\nu\in\mathcal{P}_2(\mathbb{R}^d)$
- If $\mu_0 \ll \mathrm{Leb}$, true if and only if $\mathrm{T}_{\mu_0}^{\mu_1} = \nabla u$, $u:\mathbb{R}^d\to\mathbb{R}^d$ 1-strongly convex
- For $\mu_0 = \mathcal{N}(m_0,\Sigma_0)$, $\mu_1 = \mathcal{N}(m_1,\Sigma_1)$: equivalent to $(\Sigma_0^{\frac12}\Sigma_1\Sigma_0^{\frac12})^{\frac12} \succeq \Sigma_0$
- For $\mu_0,\mu_1\in\mathcal{P}_2(\mathbb{R})$, true if and only if $F_{\mu_1}^{-1} - F_{\mu_0}^{-1}$ non-decreasing
- For $\mu_0 = \mathcal{N}(m_0,\sigma_0^2)$, $\mu_1 = \mathcal{N}(m_1,\sigma_1^2)$: equivalent to $\sigma_1 \geq \sigma_0$

## Busemann Function

**Busemann function** associated to a geodesic ray $t\in\mathbb{R}_+ \mapsto \mu_t$:

$$\forall\nu\in\mathcal{P}_2(\mathbb{R}^d),\ B^\mu(\nu) = \lim_{t\to+\infty} \mathrm{W}_2(\mu_t,\nu) - t\mathrm{W}_2(\mu_0,\mu_1)$$

For $\mu_0 = \delta_{x_0}, \mu_1 = \delta_{x_1}$, $B^\mu(\nu) = -\int \langle y - x_0, \frac{x_1-x_0}{\|x_1-x_0\|_2}\rangle\ \mathrm{d}\nu(y)$
→ extends the notion of inner product to $(\mathcal{P}_2(\mathbb{R}^d), \mathrm{W}_2)$
→ extends the notion of projection on geodesic

**Computation as a Linear Program.** For any $\nu\in\mathcal{P}_2(\mathbb{R}^d)$,

$$B^\mu(\nu) = \inf_{\tilde{\gamma}\in\Gamma(\mu_0,\mu_1,\nu)} -\frac{1}{\mathrm{W}_2(\mu_0,\mu_1)}\int \langle x_1 - x_0, y - x_0\rangle\ \mathrm{d}\tilde{\gamma}(x_0,x_1,y),$$

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

If $\mu_0 \ll \mathrm{Leb}$, $\mathrm{T}_0^1$ OT map between $\mu_0$ and $\mu_1$, for any $\nu\in\mathcal{P}_2(\mathbb{R}^d)$,

$$B^\mu(\nu) = \inf_{\gamma\in\Pi(\mu_0,\nu)} -\frac{1}{\|\mathrm{T}_0^1 - \mathrm{Id}\|_{L^2(\mu_0)}}\int \langle \mathrm{T}(x_0) - x_0, y - x_0\rangle\ \mathrm{d}\gamma(x_0,y).$$

**Closed-Forms** for the Busemann function.

- For $t\mapsto\mu_t$ geodesic ray on $\mathcal{P}_2(\mathbb{R})$ such that $\|F_{\mu_0} - F_{\mu_1}\|_{L^2([0,1])} = 1$,
$$\forall\nu\in\mathcal{P}_2(\mathbb{R}),\ B^\mu(\nu) = -\langle F_{\mu_1}^{-1} - F_{\mu_0}^{-1}, F_\nu^{-1} - F_{\mu_0}^{-1}\rangle_{L^2([0,1])}$$
→ $\mathcal{O}(n\log n)$
- For $t\mapsto\mu_t$ geodesic ray such that $\mu_0 = \mathcal{N}(m_0,\Sigma_0)$, $\mu_1 = \mathcal{N}(m_1,\Sigma_1)$ and $\mathrm{W}_2(\mu_0,\mu_1) = 1$, $\nu = \mathcal{N}(m,\Sigma)$,
$$B^\mu(\nu) = -\langle m_1 - m_0, m - m_0\rangle + \mathrm{Tr}\big(\Sigma_0(A_{\mu_0}^{\mu_1} - I_d)\big) - \mathrm{Tr}\Big(\big(\Sigma^{\frac12}(\Sigma_0 - \Sigma_0 A_{\mu_0}^{\mu_1} - A_{\mu_0}^{\mu_1}\Sigma_0 + \Sigma_1)\Sigma^{\frac12}\big)^{\frac12}\Big),$$
where $A_\mu^\nu = \Sigma_\mu^{-\frac12}(\Sigma_\mu^{\frac12}\Sigma_\nu\Sigma_\mu^{\frac12})^{\frac12}\Sigma_\mu^{-\frac12}$.

## Sliced-Wasserstein between Datasets

**Sliced-Wasserstein distance.** Let $\mu,\nu\in\mathcal{P}_2(\mathbb{R}^d)$, $P^\theta:\mathbb{R}^d\to\mathbb{R}$ a projection (typically, $P^\theta(x) = \langle x,\theta\rangle$),

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

→ Monte-Carlo approximation between $\hat{\mu}_n$, $\hat{\nu}_n$: $\mathcal{O}\big(Ln(\log n + d)\big)$

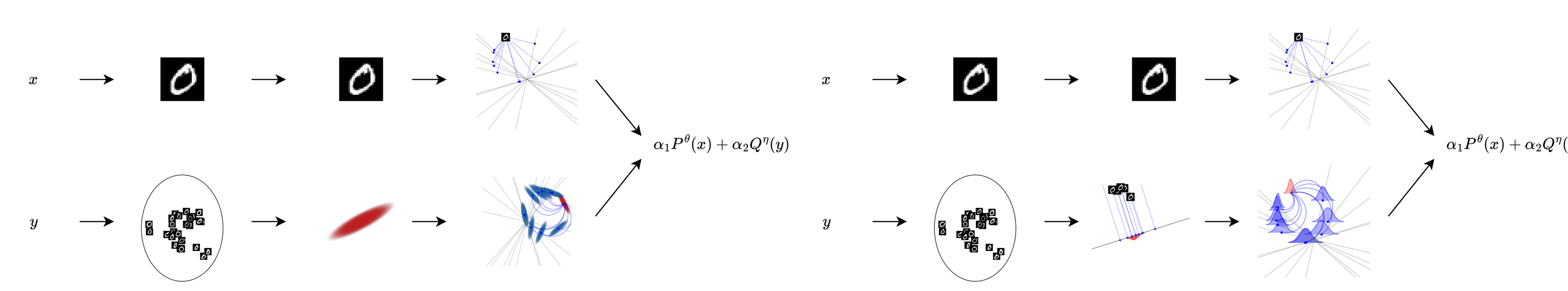
**Optimal Transport between Datasets (OTDD) (Alvarez-Melis and Fusi, 2020).** Optimal Transport problem on $\mathcal{P}_2\big(\mathbb{R}^d\times\mathcal{P}_2(\mathbb{R}^d)\big)$ with ground cost

$$d\big((x,\mu),(y,\nu)\big)^2 = \|x-y\|_2^2 + \mathrm{W}_2^2(\mu,\nu)$$

**Sliced-Wasserstein distance between Datasets.**
Define $P^{\alpha,\theta,\eta}:\mathbb{R}^d\times\mathcal{P}(\mathbb{R})\to\mathbb{R}$ with $\alpha\in S^1$, $\theta\in S^{d-1}$, as

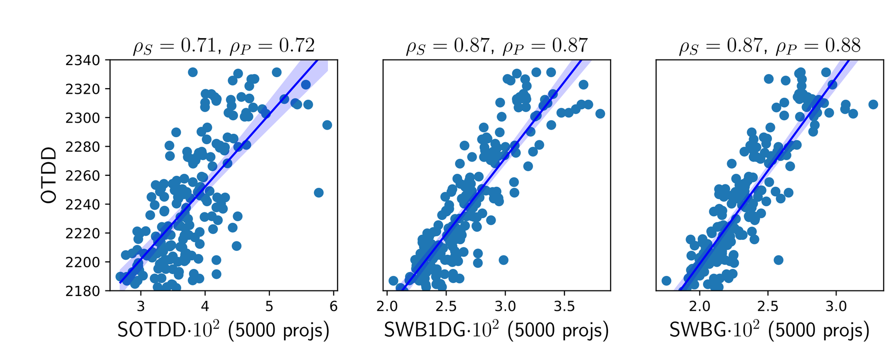
$$\forall x\in\mathbb{R}^d,\ \mu\in\mathcal{P}_2(\mathbb{R}^d),\ P^{\alpha,\theta,\eta}(x,\mu) = \alpha_1 P^\theta(x) + \alpha_2 Q^\eta(\mu)$$

Use $Q^\eta(\mu) = B^\eta(P^\theta_\#\mu)$ or $Q^\eta(\mu) = B^\eta\big(\mathcal{N}(m(\mu),\Sigma(\mu))\big)$ with $\eta$ a geodesic ray




## Experiments

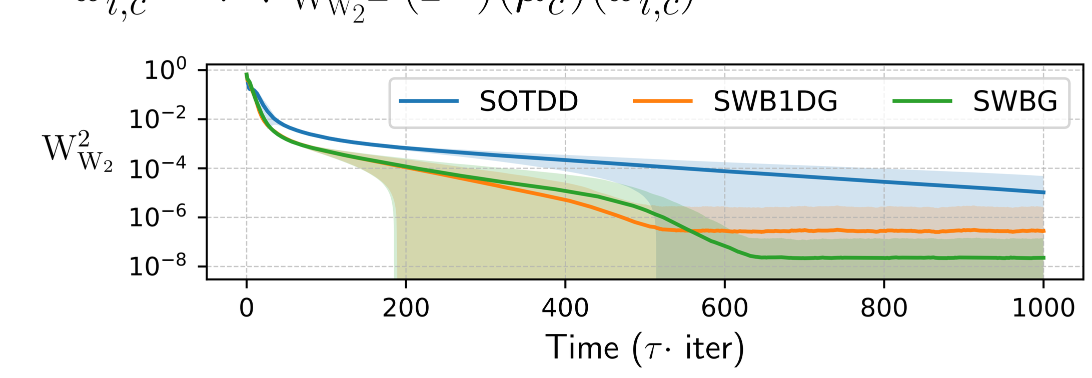
**Correlation with OTDD.** Compute distances between random subpairs of CIFAR10.
→ SWBG/SWB1DG more correlated than SOTDD (Nguyen et al., 2025)

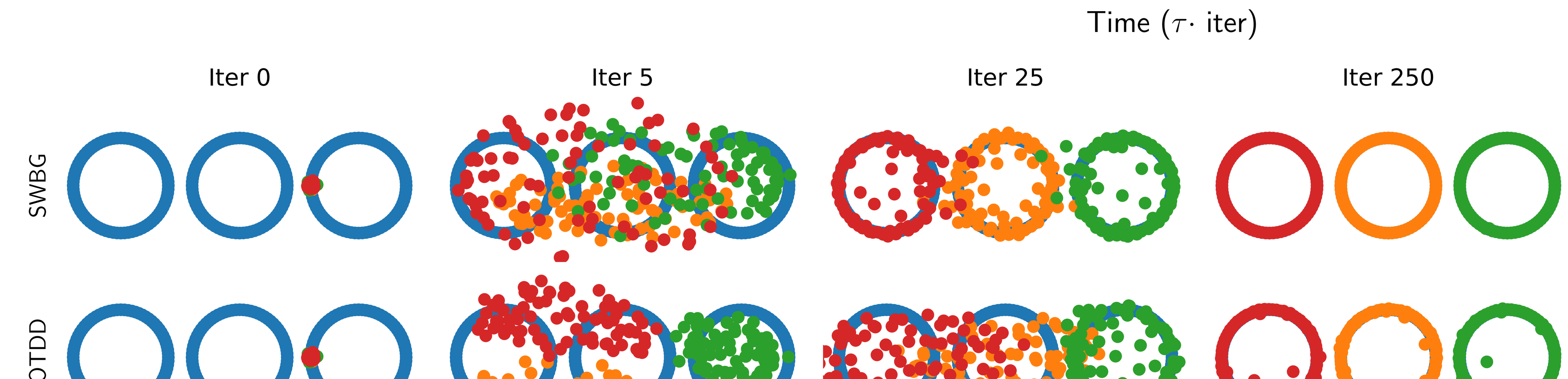



**Flow of Datasets (Bonet at al, 2025).** Minimize $\mathbb{F}(\mathbb{P}) = D(\mathbb{P},\mathbb{Q})$ by Wasserstein over Wasserstein gradient descent.
For $\mathbb{P}^k = \frac{1}{C}\sum_{c=1}^C \delta_{\mu_c^k}$, $\mu_c^k = \frac{1}{n}\sum_{i=1}^n \delta_{x_{i,c}^k}$:

$$\forall k\geq 0,\ x_{i,c}^{k+1} = x_{i,c}^k - \tau\nabla_{\mathrm{W}_{\mathrm{W}_2}}\mathbb{F}(\mathbb{P}^k)(\mu_c^k)(x_{i,c}^k)$$

Ring dataset: $C=3$, $n=80$ samples by class




**Transfer learning ($k$-shot learning).** Augment a dataset $\mathbb{Q} = \frac{1}{C}\sum_{c=1}^C \delta_{\nu_c^k}$ ($k$ small) adding flowed sampled starting from $\mathbb{P}_0 = \frac{1}{C}\sum_{c=1}^C \delta_{\mu_c^n}$ a known dataset ($n$ bigger).

| Dataset | $k$ | Trained on $\mathbb{Q}$ | OTDD | SWB1DG | SOTDD |
|---|---|---|---|---|---|
| M to F | 1 | $26.0_{\pm 5.3}$ | $30.5_{\pm 4.2}$ | $41.3_{\pm 3.4}$ | $\mathbf{43.4}_{\pm 2.6}$ |
| | 5 | $38.5_{\pm 6.7}$ | $59.7_{\pm 1.8}$ | $\mathbf{65.5}_{\pm 1.6}$ | $64.5_{\pm 1.2}$ |
| | 10 | $53.9_{\pm 7.9}$ | $64.0_{\pm 1.4}$ | $66.0_{\pm 0.9}$ | $\mathbf{67.7}_{\pm 0.6}$ |
| | 100 | $71.1_{\pm 1.5}$ | - | $\mathbf{74.1}_{\pm 0.6}$ | $72.0_{\pm 1.9}$ |

## References

David Alvarez-Melis and Nicolo Fusi. Geometric Dataset Distances via Optimal Transport. *NeurIPS*, 2020.

Khai Nguyen, Hai Nguyen, Tuan Pham, and Nhat Ho. Lightspeed Geometric Dataset Distance via Sliced Optimal Transport. *ICML*, 2025.

Clément Bonet, Christophe Vauthier, and Anna Korba. Flowing Datasets with Wasserstein over Wasserstein Gradient Flows. *ICML*, 2025.