Slicing Unbalanced Optimal Transport

Clément Bonet[1]

Joint work with Kimia Nadjahi[2], Thibault Séjourné[3], Kilian Fatras[4], Nicolas Courty[5]

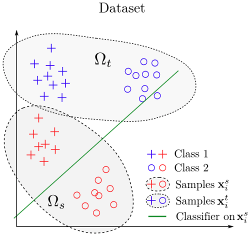
[1]ENSAE, CREST
[2]CNRS, ENS
[3]EPFL, LTS4
[4]McGill University, Mila
[5]Université Bretagne Sud, IRISA

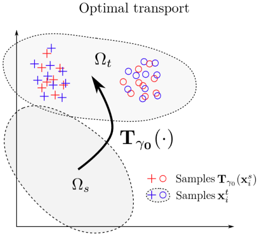
SIAM AN
30/07/2025

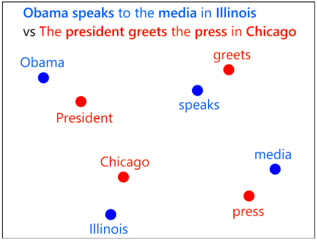
Motivations

Optimal Transport: Meaningful way to compare distributions

- Domain Adaptation (Courty et al., 2016)
- Generative Models (*e.g.* WGAN (Arjovsky et al., 2017))
- Document Classification (Kusner et al., 2015)







Optimal Transport

Definition (Wasserstein distance)

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

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

$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\ \pi^1_\# \gamma = \mu,\ \pi^2_\# \gamma = \nu\}$ and $\pi^1(x, y) = x$, $\pi^2(x, y) = y$, $\pi^1_\# \gamma = \gamma \circ (\pi^1)^{-1}$.

Properties:

- W_2 distance
- Metrizes the weak convergence
- Riemannian structure

Optimal Transport

Definition (Wasserstein distance)

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

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

$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\ \pi^1_\# \gamma = \mu,\ \pi^2_\# \gamma = \nu\}$ and $\pi^1(x, y) = x$, $\pi^2(x, y) = y$, $\pi^1_\# \gamma = \gamma \circ (\pi^1)^{-1}$.

Limitations:

- Computational cost
- Curse of dimensionality

Optimal Transport

Definition (Wasserstein distance)

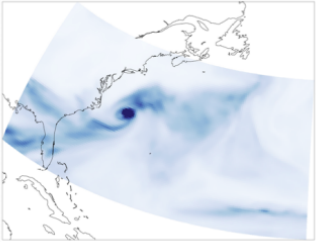
Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

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

$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\ \pi^1_\# \gamma = \mu,\ \pi^2_\# \gamma = \nu\}$ and $\pi^1(x, y) = x$, $\pi^2(x, y) = y$, $\pi^1_\# \gamma = \gamma \circ (\pi^1)^{-1}$.

Limitations:

- Computational cost
- Curse of dimensionality
- Restricted to **probability measures**



Optimal Transport

Definition (Wasserstein distance)

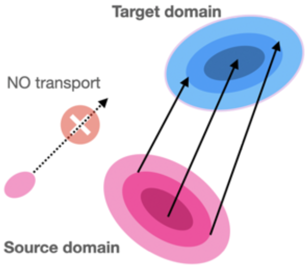
Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$,

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

$\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d),\ \pi^1_\# \gamma = \mu,\ \pi^2_\# \gamma = \nu\}$ and $\pi^1(x, y) = x$, $\pi^2(x, y) = y$, $\pi^1_\# \gamma = \gamma \circ (\pi^1)^{-1}$.

Limitations:

- Computational cost
- Curse of dimensionality
- Restricted to **probability measures**
- Lack of robustness to **outliers**



Solving the OT Problem

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

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

Computational Complexity (Pele and Werman, 2009)

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

Solving the OT Problem

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

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

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

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

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 $\hat{\nu}_n = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$,

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

Proposed solutions:

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

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

1D Wasserstein Distance

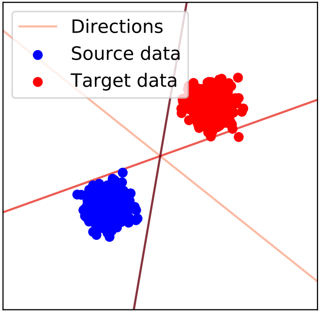
$$W_2^2(\mu,\nu) = \int_0^1 \left|F_\mu^{-1}(u) - F_\nu^{-1}(u)\right|^2\ \mathrm{d}u = \left\|F_\mu^{-1} - F_\nu^{-1}\right\|^2_{L^2([0,1])}$$

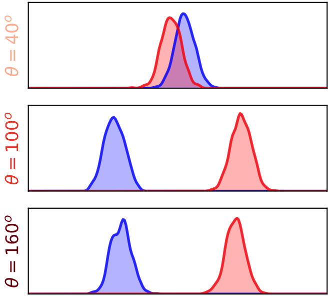
Let $x_1 < \cdots < x_n,\ y_1 < \cdots < y_n,\ \mu = \frac{1}{n}\sum_{i=1}^n \delta_{x_i},\ \nu = \frac{1}{n}\sum_{i=1}^n \delta_{y_i}$,

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

$\to O(n\log n)$

Sliced-Wasserstein Distance





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

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

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

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: $O(Ln \log n + Lnd)$
- 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.* $\lim_{n\to\infty} \mathrm{SW}_2^2(\mu_n, \mu) = 0 \iff \lim_{n\to\infty} W_2^2(\mu_n, \mu) = 0$.

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: $O(Ln\log n + Lnd)$
- 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.* $\lim_{n\to\infty} \mathrm{SW}_2^2(\mu_n, \mu) = 0 \iff \lim_{n\to\infty} W_2^2(\mu_n,\mu) = 0$.

Limitations

- Restricted to probability measures
- Not robust to outliers

Table of Contents

Unbalanced Optimal Transport

$\mathcal{M}_+(\mathbb{R}^d)$: set of positive Radon measures on $\mathbb{R}^d$

UOT Problem (static formulation) (Liero et al., 2018)

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

$$\mathrm{UOT}(\alpha, \beta) \triangleq \inf_{\gamma \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \int \|x-y\|_2^2 \mathrm{d}\gamma(x,y) + \rho_1 \mathrm{D}_{\varphi_1}(\pi^1_{\#}\gamma|\alpha) + \rho_2 \mathrm{D}_{\varphi_2}(\pi^2_{\#}\gamma|\beta)$$

φ-**divergence**:

$$\mathrm{D}_\varphi(\alpha|\beta) \triangleq \int_{\mathbb{R}^d} \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right) \mathrm{d}\beta(x) + \varphi'_\infty \int_{\mathbb{R}^d} \mathrm{d}\alpha^\perp(x)\,,$$

where $\alpha^\perp$ is defined as $\alpha = \frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\beta + \alpha^\perp$, $\varphi'_\infty = \lim_{t\to+\infty} \frac{\varphi(x)}{x}$, φ: "entropy function".

Example

- Kullback-Leibler divergence: $\varphi(t) = t\log t - t + 1$
- Total Variation distance: $\varphi(t) = |t-1|$

Unbalanced Optimal Transport

$\mathcal{M}_+(\mathbb{R}^d)$: set of positive Radon measures on $\mathbb{R}^d$

UOT Problem (static formulation) (Liero et al., 2018)

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

$$\mathrm{UOT}(\alpha, \beta) = \inf_{(\pi_1, \pi_2) \in \mathcal{M}_+(\mathbb{R}^d) \times \mathcal{M}_+(\mathbb{R}^d)} \mathrm{OT}(\pi_1, \pi_2) + \rho_1 \mathrm{D}_{\varphi_1}(\pi_1 | \alpha) + \rho_2 \mathrm{D}_{\varphi_2}(\pi_2 | \beta)$$

φ-**divergence**:

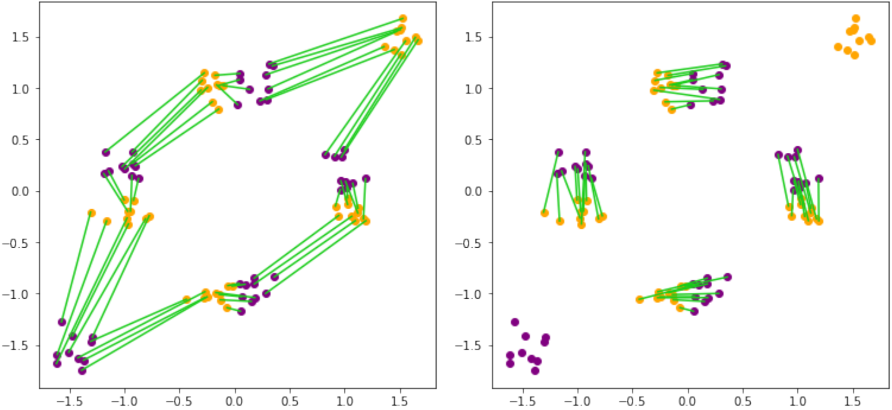
$$\mathrm{D}_\varphi(\alpha|\beta) \triangleq \int_{\mathbb{R}^d} \varphi\left(\frac{\mathrm{d}\alpha}{\mathrm{d}\beta}(x)\right) \mathrm{d}\beta(x) + \varphi'_\infty \int_{\mathbb{R}^d} \mathrm{d}\alpha^\perp(x),$$

where $\alpha^\perp$ is defined as $\alpha = \frac{\mathrm{d}\alpha}{\mathrm{d}\beta}\beta + \alpha^\perp$, $\varphi'_\infty = \lim_{t \to +\infty} \frac{\varphi(x)}{x}$, φ: "entropy function".

Example

- Kullback-Leibler divergence: $\varphi(t) = t \log t - t + 1$
- Total Variation distance: $\varphi(t) = |t - 1|$

Balanced vs Unbalanced OT



OT matching
$(\pi^1_\# \gamma, \pi^2_\# \gamma) = (\alpha, \beta)$

UOT matching
$\mathrm{TV}(\pi^1_\# \gamma | \alpha) + \mathrm{TV}(\pi^2_\# \gamma | \beta)$

From (Séjourné et al., 2023)

Solving the UOT Problem

Various solutions in particular cases:

- Entropic regularization (Chizat et al., 2018) $\rightarrow O(n^2/\varepsilon)$ (Pham et al., 2020)
- Translation-invariant Sinkhorn (Séjourné et al., 2022)
- Translation-invariant + Frank-Wolfe in 1D (Séjourné et al., 2022) $\rightarrow O(n \log n)$
- Maximization-Minimization algorithm for Bregman divergences (Chapel et al., 2021) $\rightarrow O(n^3)$

Slicing approaches: only in the case TV

- No mass destruction in the source measure (Bonneel and Coeurjolly, 2019)
- Restricted to equal weights (Bai et al., 2023)

Solving the UOT Problem

Various solutions in particular cases:

- Entropic regularization (Chizat et al., 2018) $\rightarrow O(n^2/\varepsilon)$ (Pham et al., 2020)
- Translation-invariant Sinkhorn (Séjourné et al., 2022)
- Translation-invariant + Frank-Wolfe in 1D (Séjourné et al., 2022) $\rightarrow O(n \log n)$
- Maximization-Minimization algorithm for Bregman divergences (Chapel et al., 2021) $\rightarrow O(n^3)$

Slicing approaches: only in the case TV

- No mass destruction in the source measure (Bonneel and Coeurjolly, 2019)
- Restricted to equal weights (Bai et al., 2023)

Contributions

- Slicing UOT
- Unbalancing SOT

Table of Contents

Slicing and Unbalancing OT

2 strategies:

Sliced Unbalanced OT

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

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

$\rightarrow$ generalization of sliced partial OT (Bai et al., 2023) (in the case $\mathcal{D}_\varphi = \mathrm{TV}$)
$\rightarrow$ particular case of sliced divergence (Nadjahi et al., 2020)

Slicing and Unbalancing OT

2 strategies:

Sliced Unbalanced OT

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

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

$\rightarrow$ generalization of sliced partial OT (Bai et al., 2023) (in the case $\mathcal{D}_\varphi = \mathrm{TV}$)
$\rightarrow$ particular case of sliced divergence (Nadjahi et al., 2020)

Unbalanced Sliced OT

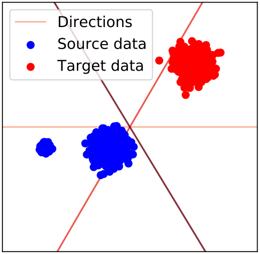
Let $\alpha, \beta \in \mathcal{M}_+(\mathbb{R}^d)$,

$$\mathrm{USW}_2^2(\alpha, \beta) = \inf_{\pi_1, \pi_2 \in \mathcal{M}_+(\mathbb{R}^d)} \mathrm{SW}_2^2(\pi_1, \pi_2) + \mathrm{D}_{\varphi_1}(\pi_1 | \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 | \beta)$$

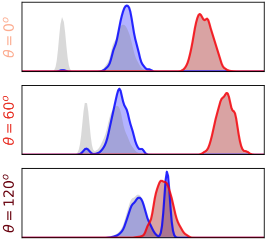
$\rightarrow$ new!

SUOT vs USOT

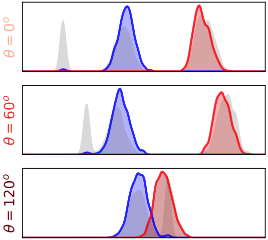
In the rest of the talk: $\mathcal{D}_\varphi = \mathrm{KL}$



Samples and directions



$\mathrm{SUOT}\ (\rho_1 = 0.01,\ \rho_2 = 1)$



$\mathrm{USOT}\ (\rho_1 = 1,\ \rho_2 = 1)$

- SUOT: penalize the marginals of $\pi_\theta \in \Pi(P^\theta_\# \alpha, P^\theta_\# \beta)$
- USOT: penalize the marginals of $\gamma \in \Pi(\alpha, \beta)$

Theoretical Properties

- Under mild assumptions, solutions of USOT, SUOT exist
- SUOT and USOT metrize the weak convergence

Metric properties

- UOT (pseudo) metric on $\mathbb{R} \implies$ SUOT (pseudo) metric on $\mathbb{R}^d$
- USOT pseudo-metric
- Sample complexity independent of the dimension (shown theoretically for SUOT and empirically for USOT)

$$\mathrm{SUOT}(\alpha, \beta) \leq \mathrm{USOT}(\alpha, \beta) \leq \mathrm{UOT}(\alpha, \beta)$$

Table of Contents

Frank-Wolfe Algorithm

Frank-Wolfe algorithm:

$$\min_{x \in \mathrm{X}} \; h(x),$$

$\mathrm{X} \subset \mathbb{R}^d$ **compact** and convex, $h : \mathrm{X} \to \mathbb{R}$ convex and differentiable

Algorithm – `Frank-Wolfe`

Input: $x^{(0)}, (\gamma_t)_t$
Output: $h(x^{(T)})$

for $t = 1, \ldots, T$ **do**
$y^{(t)} \leftarrow \mathrm{argmin}_{y \in \mathrm{X}} \, \langle y, \nabla h(x^{(t)}) \rangle$
$x^{(t+1)} \leftarrow (1 - \gamma_t) x^{(t)} + \gamma_t y^{(t)}$
end for

UOT Problem in 1D

For $\varphi : \mathbb{R} \to \mathbb{R}$ convex, note $\varphi^*(t) = \sup_{s\in\mathbb{R}} \; st - \varphi(s)$

Dual of UOT problem (Liero et al., 2018)

Let $\alpha, \beta \in \mathcal{M}_+(\mathbb{R})$,

$$\mathsf{UOT}(\alpha, \beta) = \sup_{f \oplus g \leq c} \mathcal{D}(f, g; \alpha, \beta)$$

with $\mathcal{D}(f, g; \alpha, \beta) \triangleq - \int \varphi_1^*\big(-f(x)\big)\mathrm{d}\alpha(x) - \int \varphi_2^*\big(-g(y)\big)\mathrm{d}\beta(y)$.

- Dual of UOT not translation invariant (Séjourné et al., 2022), *i.e.* there exists $\lambda \in \mathbb{R}$ s.t.

$$\mathcal{D}(f + \lambda, g - \lambda) \neq \mathcal{D}(f, g)$$

- Can't apply Frank-Wolfe as constraint set is unbounded (Séjourné et al., 2022), *i.e.* $f \oplus g \leq c \implies f + \lambda \oplus g - \lambda \leq c$

UOT Problem in 1D

For $\varphi : \mathbb{R} \to \mathbb{R}$ convex, note $\varphi^*(t) = \sup_{s\in\mathbb{R}}\ st - \varphi(s)$

Translation-Invariant Dual of UOT problem (Séjourné et al., 2022)

Let $\alpha, \beta \in \mathcal{M}_+(\mathbb{R})$,

$$\mathsf{UOT}(\alpha,\beta) = \sup_{f\oplus g \leq c} \mathcal{H}(f,g;\alpha,\beta)$$

with $\mathcal{H}(f,g;\alpha,\beta) \triangleq \sup_{\lambda\in\mathbb{R}}\ \mathcal{D}(f+\lambda, g-\lambda;\alpha,\beta)$.

- Dual of UOT not translation invariant (Séjourné et al., 2022), *i.e.* there exists $\lambda \in \mathbb{R}$ s.t.

$$\mathcal{D}(f+\lambda, g-\lambda) \neq \mathcal{D}(f,g)$$

- Can't apply Frank-Wolfe as constraint set is unbounded (Séjourné et al., 2022), *i.e.* $f \oplus g \leq c \implies f+\lambda \oplus g-\lambda \leq c$
- 1D UOT by Frank-Wolfe algorithm (for $D_\varphi = \mathrm{KL}$)
  $\rightarrow$ complexity $O(Tn\log n)$ with T number of Frank-Wolfe iterations

UOT Problem in 1D

For $\varphi : \mathbb{R} \to \mathbb{R}$ convex, note $\varphi^*(t) = \sup_{s\in\mathbb{R}} \; st - \varphi(s)$

Translation-Invariant Dual of UOT problem (Séjourné et al., 2022)

Let $\alpha, \beta \in \mathcal{M}_+(\mathbb{R})$,

$$\mathrm{UOT}(\alpha, \beta) = \sup_{f \oplus g \leq c} \mathcal{H}(f, g; \alpha, \beta)$$

with $\mathcal{H}(f, g; \alpha, \beta) \triangleq \sup_{\lambda\in\mathbb{R}} \; \mathcal{D}(f + \lambda, g - \lambda; \alpha, \beta)$.

- Dual of UOT not translation invariant (Séjourné et al., 2022), *i.e.* there exists $\lambda \in \mathbb{R}$ s.t.

$$\mathcal{D}(f + \lambda, g - \lambda) \neq \mathcal{D}(f, g)$$

- Can't apply Frank-Wolfe as constraint set is unbounded (Séjourné et al., 2022), *i.e.* $f \oplus g \leq c \implies f + \lambda \oplus g - \lambda \leq c$
- 1D UOT by Frank-Wolfe algorithm (for $D_\varphi = \mathrm{KL}$)
  $\to$ complexity $O(Tn \log n)$ with T number of Frank-Wolfe iterations
- SUOT solved in $O(LTn \log n)$ by doing L Frank-Wolfe in parallel

Solving USOT

Dual of USOT

$$\mathrm{USOT}(\alpha,\beta) = \sup_{\forall \ell, f_{\theta_\ell} \oplus g_{\theta_\ell} \leq c} \mathcal{D}\left(\frac{1}{L}\sum_{\ell=1}^{L} f_{\theta_\ell} \circ P^{\theta_\ell}, \frac{1}{L}\sum_{\ell=1}^{L} g_{\theta_\ell} \circ P^{\theta_\ell}; \alpha, \beta\right)$$

$\rightarrow$ solve $\max_{\bar{f}=\frac{1}{L}\sum_{\ell=1}^{L} f_\ell, \bar{g}=\frac{1}{L}\sum_{\ell=1}^{L} g_\ell, f_\ell \oplus g_\ell \leq c} \mathcal{H}(\bar{f}, \bar{g}; \alpha, \beta)$

Algorithm – `Frank-Wolfe`

Input: $(f_\ell^{(0)}, g_\ell^{(0)})_\ell$, $(\gamma_t)_t$
Output: $\mathcal{H}(\bar{f}^{(T)}, \bar{g}^{(T)})$

for $t = 1, \dots, T$ **do**
  $(\bar{r}^{(t)}, \bar{s}^{(t)}) \leftarrow \mathrm{argmax}_{\bar{r}, \bar{s}, r_\ell \oplus s_\ell \leq c} \langle (\bar{r}, \bar{s}), \nabla \mathcal{H}(\bar{f}^{(t)}, \bar{g}^{(t)}) \rangle$
  $(\bar{f}^{(t+1)}, \bar{g}^{(t+1)}) \leftarrow (1-\gamma_t)(\bar{f}^{(t)}, \bar{g}^{(t)}) + \gamma_t(\bar{r}^{(t)}, \bar{s}^{(t)})$
end for

Solving USOT

Algorithm – `Frank-Wolfe`

Input: $(f_\ell^{(0)}, g_\ell^{(0)})_\ell,\ (\gamma_t)_t$
Output: $\mathcal{H}(\bar{f}^{(T)}, \bar{g}^{(T)})$

for $t = 1, \dots, T$ **do**

$(\bar{r}^{(t)}, \bar{s}^{(t)}) \leftarrow \operatorname{argmax}_{\bar{r},\bar{s},r_\ell \oplus s_\ell \leq c} \langle (\bar{r}, \bar{s}), \nabla\mathcal{H}(\bar{f}^{(t)}, \bar{g}^{(t)}) \rangle$

$(\bar{f}^{(t+1)}, \bar{g}^{(t+1)}) \leftarrow (1-\gamma_t)(\bar{f}^{(t)}, \bar{g}^{(t)}) + \gamma_t(\bar{r}^{(t)}, \bar{s}^{(t)})$

end for

If $\mathrm{D}_{\varphi_1} = \rho_1 \mathrm{KL}, \mathrm{D}_{\varphi_2} = \rho_2 \mathrm{KL}$:

- $\lambda^* = \operatorname{argmax}_{\lambda \in \mathbb{R}}\ \mathcal{D}(\bar{f} + \lambda, \bar{g} - \lambda) = \frac{\rho_1\rho_2}{\rho_1+\rho_2} \log \left(\frac{\int e^{-\bar{f}(x)/\rho_1} \mathrm{d}\alpha(x)}{\int e^{-\bar{g}(y)/\rho_2} \mathrm{d}\beta(y)} \right)$
- $\nabla\mathcal{H}(\bar{f}, \bar{g}) = (\tilde{\alpha}, \tilde{\beta})$ with $\tilde{\alpha} = \nabla\varphi_1^*(-\bar{f} - \lambda^*)\alpha$ and $\tilde{\beta} = \nabla\varphi_2^*(-\bar{g} + \lambda^*)\beta$

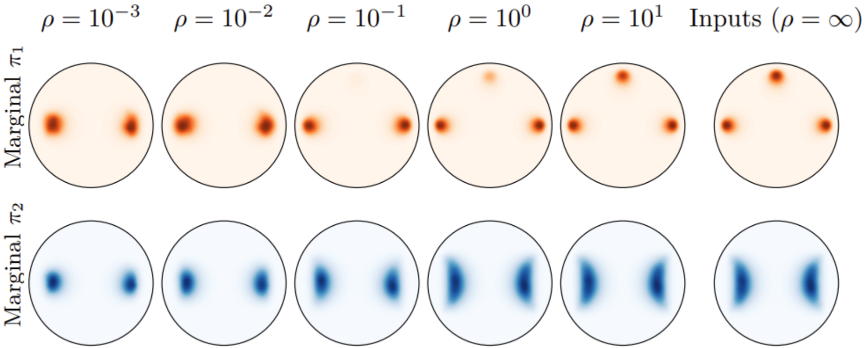
$$(\bar{r}^{(t)}, \bar{s}^{(t)}) = \underset{r_\ell \oplus s_\ell \leq c}{\operatorname{argmax}}\ \langle \bar{r}, \tilde{\alpha} \rangle + \langle \bar{s}, \tilde{\beta} \rangle$$

→ FW step: solve $\mathrm{OT}\big(P_\#^{\theta_\ell}\tilde{\alpha}, P_\#^{\theta_\ell}\tilde{\beta}\big)$
→ Complexity in $O(TLn \log n)$

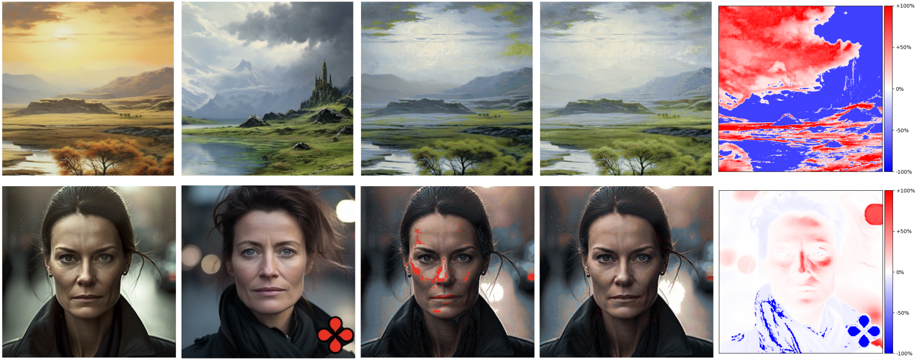
Table of Contents

Hyperbolic Space

Can use any projection on $\mathbb{R}$: *e.g.* projection on geodesics on hyperbolic spaces (Bonet et al., 2023)



Color Transfer



1st column: sources, 2nd column: targets, 3rd column: SOT gradient flow, 4th column: USOT + SOT gradient flow, 5th column: % of mass change by USOT (red: mass creation, blue: mass destruction)

- 300x300 images: $n = 90$K
- Step 1: solve USOT to obtain marginals (π_1, π_2)
- Step 2: gradient flow of SOT

Conclusion

Conclusion

- Introduction 2 new losses merging UOT and SOT
- Theoretical Analysis
- Efficient and modular Frank-Wolfe algorithm
- Encouraging empirical results

Some perspectives:

- Open theoretical questions: sample complexity of USOT, strong duality for $\lambda = \mathrm{Unif}(S^{d-1})$
- New large scale applications
- Challenge: choice of ρ

Conclusion

Conclusion

- Introduction 2 new losses merging UOT and SOT
- Theoretical Analysis
- Efficient and modular Frank-Wolfe algorithm
- Encouraging empirical results

Some perspectives:

- Open theoretical questions: sample complexity of USOT, strong duality for $\lambda = \mathrm{Unif}(S^{d-1})$
- New large scale applications
- Challenge: choice of ρ

Thank you!

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

Yikun Bai, Bernhard Schmitzer, Matthew Thorpe, and Soheil Kolouri. Sliced optimal partial transport. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13681–13690, 2023.

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, 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.

Nicolas Bonneel and David Coeurjolly. Spot: sliced partial optimal transport. *ACM Transactions on Graphics (TOG)*, 38(4):1–13, 2019.

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

References II

Laetitia Chapel, Rémi Flamary, Haoran Wu, Cédric Févotte, and Gilles Gasso. Unbalanced optimal transport through non-negative penalized linear regression. *Advances in Neural Information Processing Systems*, 34:23270–23282, 2021.

Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced optimal transport problems. *Mathematics of computation*, 87(314):2563–2609, 2018.

Nicolas Courty, Rémi Flamary, Devis Tuia, and Alain Rakotomamonjy. Optimal transport for domain adaptation. *IEEE transactions on pattern analysis and machine intelligence*, 39(9):1853–1865, 2016.

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

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.

References III

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.

Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018.

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.

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.

References IV

Khiem Pham, Khang Le, Nhat Ho, Tung Pham, and Hung Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, pages 7673–7682. PMLR, 2020.

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.

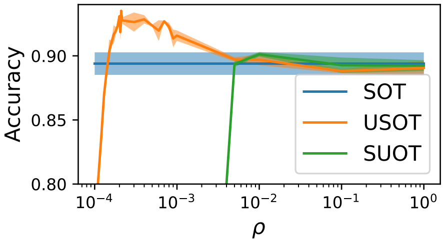
Thibault Séjourné, François-Xavier Vialard, and Gabriel Peyré. Faster unbalanced optimal transport: Translation invariant sinkhorn and 1-d frank-wolfe. In *International Conference on Artificial Intelligence and Statistics*, pages 4995–5021. PMLR, 2022.

Thibault Séjourné, Clément Bonet, Kilian Fatras, Kimia Nadjahi, and Nicolas Courty. Unbalanced optimal transport meets sliced-wasserstein. *arXiv preprint arXiv:2306.07176*, 2023.

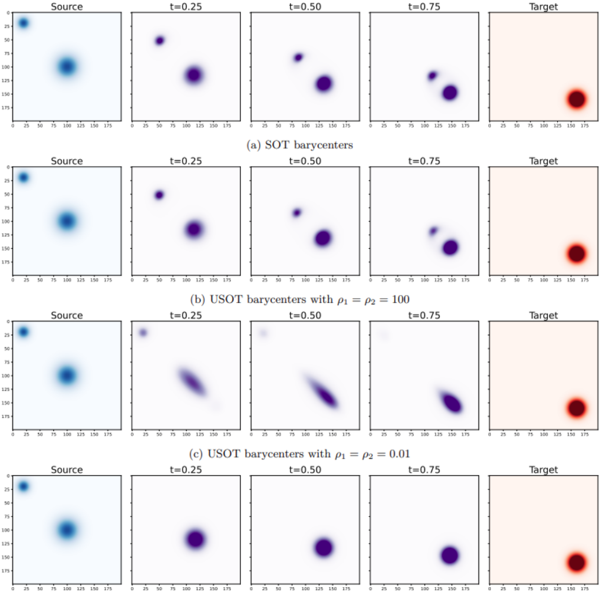
Document Classification

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

	BBCSport		Goodreads		
	Acc	t $(\cdot 10^{-3}$s)	Acc (genre)	Acc (like)	t $(\cdot 10^{-3}$s)
OT	94.55	$3.12_{\pm 1.61}$	55.22	71.00	$440.30_{\pm 250}$
UOT	96.73	$243.39_{\pm 9.24}$	-	-	-
SinkhUOT	95.45	$46.22_{\pm 2.17}$	53.55	67.81	$2021.68_{\pm 356}$
SOT	$89.39_{\pm 0.76}$	$1.80_{\pm 0.22}$	$50.09_{\pm 0.51}$	$65.60_{\pm 0.20}$	$4.49_{\pm 1.44}$
SUOT	$90.12_{\pm 0.15}$	$13.9_{\pm 1.21}$	$50.15_{\pm 0.04}$	$66.72_{\pm 0.38}$	$14.32_{\pm 0.95}$
USOT	$93.52_{\pm 0.04}$	$14.37_{\pm 1.29}$	$52.67_{\pm 0.62}$	$67.78_{\pm 0.39}$	$14.45_{\pm 0.88}$



Barycenters



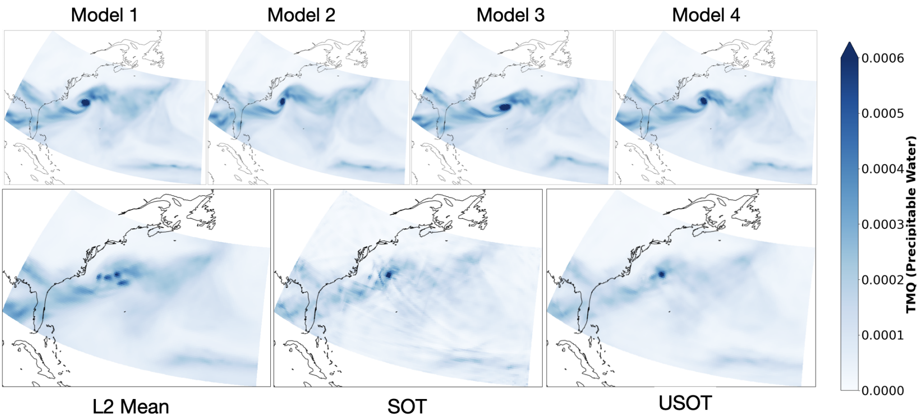
(a) SOT barycenters

(b) USOT barycenters with $\rho_1 = \rho_2 = 100$

(c) USOT barycenters with $\rho_1 = \rho_2 = 0.01$

(d) USOT barycenters with $\rho_1 = 0.01$ and $\rho_2 = 100$

Barycenter on Geophysical Data



- First row: 4 climate models
- Second row: Different barycenters
- Measures of size 100×200