Scalable Unbalanced Optimal Transport by Slicing

Kimia Nadjahi

Joint work with Thibault Séjourné, Clément Bonet, Kilian Fatras, Nicolas Courty

Séminaire Monge – Université Gustave Eiffel
April 18, 2024

$\mathcal{M}_+^1(\mathbb{R}^d)$: set of probability measures on $\mathbb{R}^d$

Kantorovich OT problem

For $(\alpha, \beta) \in \mathcal{M}_+^1(\mathbb{R}^d) \times \mathcal{M}_+^1(\mathbb{R}^d)$,

$$\mathrm{OT}(\alpha, \beta) \triangleq \inf_{\pi \in \Gamma(\alpha, \beta)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{C}(x, y) \, \mathrm{d}\pi(x, y)$$

$\Gamma(\alpha, \beta) \triangleq \left\{\pi \in \mathcal{M}_+^1(\mathbb{R}^d \times \mathbb{R}^d) \text{ with marginals } (\pi_1, \pi_2) = (\alpha, \beta)\right\}$

Limitations:

- Restricted to *probability* measures
- Lack of robustness to *outliers*

Unbalanced Optimal Transport

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

Unbalanced OT (static formulation) [Liero et al., 2018]

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

$$\mathrm{UOT}(\alpha, \beta) \triangleq \inf_{\pi \in \mathcal{M}_+(\mathbb{R}^d \times \mathbb{R}^d)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \mathrm{C}(x, y)\, \mathrm{d}\pi(x, y) + \mathrm{D}_{\boldsymbol{\rho}\varphi_1}(\pi_1 | \alpha) + \mathrm{D}_{\boldsymbol{\rho}\varphi_2}(\pi_2 | \beta)$$

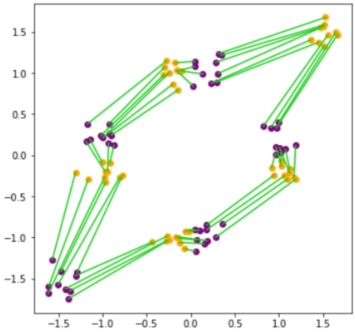
φ-divergences

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

φ "*entropy* function", $\varphi'_\infty \triangleq \lim_{x \to +\infty} \varphi(x)/x$. $\alpha = (\mathrm{d}\alpha/\mathrm{d}\beta)\beta + \alpha^\perp$

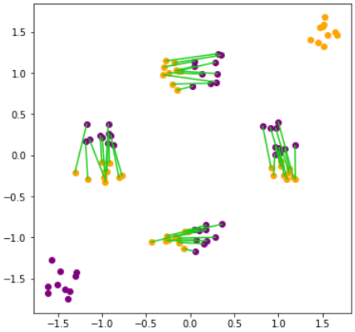
Examples: Kullback-Leibler divergence (KL, $\varphi(t) = t\log(t) - t + 1$),
Total Variation (TV, $\varphi(t) = |t - 1|$)

Balanced vs. Unbalanced OT



(a) OT matching

$(\pi_1, \pi_2) = (\alpha, \beta)$



(b) Unbalanced OT matching

$\rho \mathrm{TV}(\pi_1|\alpha) + \rho \mathrm{TV}(\pi_2|\beta)$

Figure from: T. Séjourné, G. Peyré, and F.-X. Vialard. *Unbalanced optimal transport, from theory to numerics*. 2022

Solving (unbalanced) OT in practice

Solving OT

- Linear programming
- Entropic regularization [Cuturi, 2013]
- Slicing [Rabin et al., 2011]

Solving unbalanced OT

- Various strategies, reviewed in [Séjourné et al., 2022]
- Entropic regularization [Chizat et al., 2018]
- Translation-invariant dual formulation + Frank-Wolfe strategy [Séjourné et al., 2022]
- Slicing → for one specific setting and lack of theoretical insights [Bai et al., 2022]

Background on Sliced Optimal Transport

Optimal Transport in 1D

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

$$\mathrm{OT}(\alpha, \beta) = \int_0^1 \left| F_\alpha^{-1}(t) - F_\beta^{-1}(t) \right| \mathrm{d}t$$

$F_\alpha^{-1}, F_\beta^{-1}$: quantile functions of α, β [Rachev and Rüschendorf, 1998]

$$\alpha = \tfrac{1}{n} \textstyle\sum_{i=1}^n \delta_{x_i} \,, \quad \beta = \tfrac{1}{n} \textstyle\sum_{i=1}^n \delta_{y_i}$$

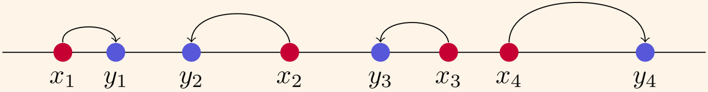


Figure inspired by [Peyré and Cuturi, 2019]

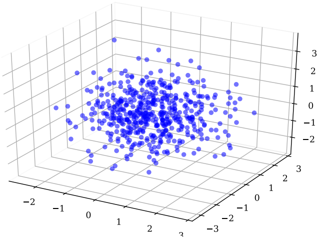
$$\mathrm{OT}(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^n |x_{(i)} - y_{(i)}|$$

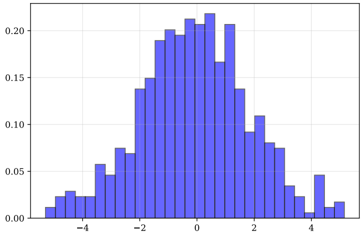
with $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)} \,, \quad y_{(1)} \leq y_{(2)} \leq \cdots \leq y_{(n)}$

$\Rightarrow$ $\mathcal{O}(n \log(n))$ operations

Slicing Distributions

compute $\langle \theta, y_i \rangle$ (for some $\theta \in \mathbb{R}^3$)





$$\beta = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$$

$$\theta^\star_\sharp \beta = \frac{1}{n} \sum_{i=1}^{n} \delta_{\langle \theta, y_i \rangle}$$

$\theta^\star_\sharp \beta$ is the *pushforward measure* of β by $\theta^\star$, with

$$\theta^\star : \mathbb{R}^d \to \mathbb{R}, \quad \theta^\star(y) = \langle \theta, y \rangle$$

Sliced OT [Rabin et al., 2011]

$\mathbb{S}^{d-1} \triangleq \{\theta \in \mathbb{R}^d \ : \ \|\theta\| = 1\}$ with uniform measure $\boldsymbol{\sigma}$

For $\theta \in \mathbb{S}^{d-1}$, $\theta^\star : \mathbb{R}^d \to \mathbb{R}$ s.t. $\forall x \in \mathbb{R}^d$, $\theta^\star(x) \triangleq \langle \theta, x \rangle$

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

$$\mathrm{SOT}(\alpha, \beta) \triangleq \int_{\mathbb{S}^{d-1}} \mathrm{OT}(\theta^\star_\sharp \alpha, \theta^\star_\sharp \beta) \mathrm{d}\boldsymbol{\sigma}(\theta)$$

with $\theta^\star_\sharp \alpha$, $\theta^\star_\sharp \beta$: *pushforward measures* of α, β by $\theta^\star$

Monte Carlo approximation based on $\theta_1, \ldots, \theta_K \overset{i.i.d}{\sim} \boldsymbol{\sigma}$

$\Rightarrow \mathcal{O}\Big(\underbrace{Kdn}_{\text{projecting}} + \underbrace{Kn\log(n)}_{\text{sorting}} \Big)$ operations

Sliced OT

Various applications

- Barycenters of measures [Rabin et al., 2012; Bonneel et al., 2015]
- Kernel functions for topological data analysis, pattern recognition [Kolouri et al., 2016; Carrière et al., 2017]
- Generative modeling [Kolouri et al., 2018; Deshpande et al., 2018; Liutkus et al., 2019; Wu et al., 2019; Kolouri et al., 2019b; Dai and Seljak, 2021]

Theoretical analysis

- Metric axioms and equivalence with OT [Bonnotte, 2013]
- Convergence guarantees [Nadjahi et al., 2019; Xi and Niles-Weed, 2022; Tanguy et al., 2024]
- Sample complexity [Manole et al., 2019; Nadjahi et al., 2020]
- Sliced-OT flows [Park and Slepčev, 2023]

Extensions [Kolouri et al., 2019a; Deshpande et al., 2019; Ohana et al., 2023; Nguyen et al., 2021]

Scalable Unbalanced Optimal Transport by Slicing

Towards Slicing *Unbalanced* OT

For $\varphi : \mathbb{R} \to \mathbb{R}$ convex, denote by φ^* its *Legendre transform*, *i.e.*,

$$\forall x \in \mathbb{R}, \quad \varphi^*(x) \triangleq \sup_{y \in \mathbb{R}} xy - \varphi(y)$$

Strong duality [Liero et al., 2018]

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

$$\mathrm{UOT}(\alpha, \beta) = \sup_{f \oplus g \leq \mathrm{C}} \mathcal{D}(f, g)$$

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

UOT is not invariant to translation [Séjourné et al., 2022]

- Given $\lambda \in \mathbb{R}$, one may have $\mathcal{D}(f + \lambda, g - \lambda) \neq \mathcal{D}(f, g)$
- This can slow down the computation of $\mathrm{UOT}(\alpha, \beta)$

Strong duality [Liero et al., 2018]

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

$$\mathrm{UOT}(\alpha, \beta) = \sup_{f \oplus g \leq \mathrm{C}} \mathcal{D}(f, g)$$

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

Translation-invariant (TI) dual [Séjourné et al., 2022]

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

$$\mathrm{UOT}(\alpha, \beta) = \sup_{f \oplus g \leq \mathrm{C}} \mathcal{H}(f, g) \tag{1}$$

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

1-D UOT: Frank-Wolfe algorithm to solve (1) [Séjourné et al., 2022]

$\Rightarrow \mathcal{O}(Tn\log(n))$ operations, with T : number of F-W iterations

Slicing and Unbalancing Optimal Transport

We propose **two strategies** [Séjourné et al., 2023]

Sliced Unbalanced OT

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

$$\mathrm{SUOT}(\alpha, \beta) \triangleq \int_{\mathbb{S}^{d-1}} \mathrm{UOT}(\theta^\star_\sharp \alpha, \theta^\star_\sharp \beta) \mathrm{d}\boldsymbol{\sigma}(\theta)$$

$$\mathrm{UOT}(\theta^\star_\sharp \alpha, \theta^\star_\sharp \beta) = \inf_{\pi_\theta \in \mathcal{M}_+(\mathbb{R}\times\mathbb{R})} \int \mathrm{C}(x,y)\, \mathrm{d}\pi_\theta(x,y) + \mathrm{D}_{\varphi_1}((\pi_\theta)_1 | \theta^\star_\sharp \alpha) + \mathrm{D}_{\varphi_2}((\pi_\theta)_2 | \theta^\star_\sharp \beta)$$

Unbalanced Sliced OT

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

$$\mathrm{USOT}(\alpha, \beta) \triangleq \inf_{(\pi_1, \pi_2) \in \mathcal{M}_+(\mathbb{R}^d)^2} \mathrm{SOT}(\pi_1, \pi_2) + \mathrm{D}_{\varphi_1}(\pi_1 | \alpha) + \mathrm{D}_{\varphi_2}(\pi_2 | \beta)$$

SUOT: generalization of sliced partial OT [Bai et al., 2022], particular case of sliced divergences [Nadjahi et al., 2020]

USOT: new!

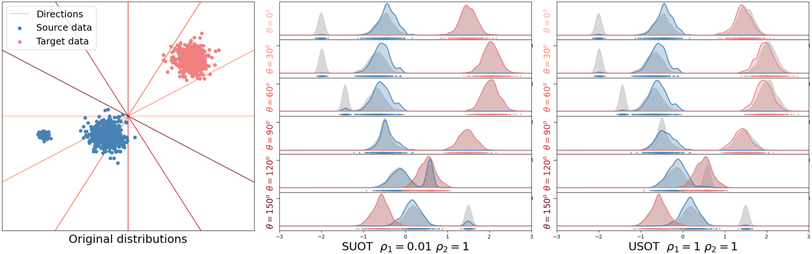


Figure 1: *(left)* input distributions; *(middle)* $(\pi_\theta)_1, (\pi_\theta)_2$ by **SUOT**; *(right)* $\theta^\star_\sharp \pi_1, \theta^\star_\sharp \pi_2$ by **USOT**; projected distributions in grey

Theoretical Properties

Existence of solutions

Under mild assumptions on C and (φ_1, φ_2), the solution of $\mathrm{USOT}(\alpha, \beta)$ and $\mathrm{SUOT}(\alpha, \beta)$ exist, *i.e.*,

- There exists π attaining the infimum in $\mathrm{USOT}(\alpha, \beta)$
- For $\theta \in \mathbb{S}^{d-1}$, there exists π_θ attaining the inf. in $\mathrm{UOT}(\theta^\star_\sharp \alpha, \theta^\star_\sharp \beta)$

Metric axioms

- UOT is a (pseudo-)metric on $\mathbb{R}$ $\Rightarrow$ SUOT is a (pseudo-)metric on $\mathbb{R}^d$ (consistent with [Nadjahi et al., 2020])
- USOT is a pseudo-metric on $\mathbb{R}^d$

Theoretical Properties

Equivalence

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

If $(\alpha, \beta) \in \mathcal{M}_+(\mathsf{X}) \times \mathcal{M}_+(\mathsf{X})$ with $\mathsf{X} \subseteq \mathbb{R}^d$ compact, and $\mathrm{D}_{\varphi_1} = \mathrm{D}_{\varphi_2} = \rho\mathrm{KL}$ or $\rho\mathrm{TV}$,

$$\mathrm{UOT}(\alpha, \beta) \leq c\, \mathrm{SUOT}(\alpha, \beta)^{\frac{1}{d+1}}$$

with c : constant depending on $(\rho, \mathrm{radius}(\mathsf{X}), m(\alpha), m(\beta))$

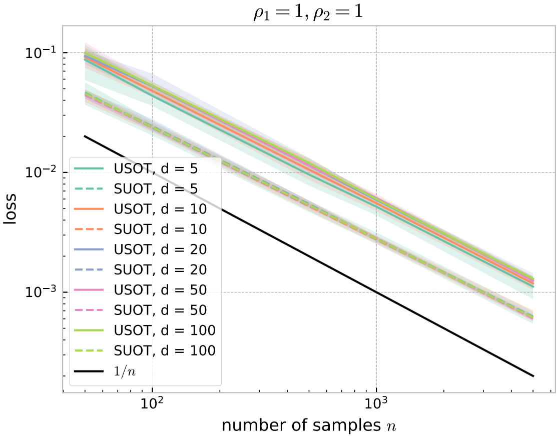
$\Rightarrow$ Equivalence of SUOT, USOT and UOT if $m(\alpha), m(\beta) \leq M$

Weak* metrization

Under the same assumptions,

$$\alpha_n \rightharpoonup \alpha \;\Leftrightarrow\; \lim_{n\to+\infty} \mathrm{SUOT}(\alpha_n, \alpha) = 0 \;\Leftrightarrow\; \lim_{n\to+\infty} \mathrm{USOT}(\alpha_n, \alpha) = 0$$

Statistical Efficiency



Sample complexity: $\mathbb{E}[|\mathcal{L}(\hat{\alpha}_n, \hat{\beta}_n) - \mathcal{L}(\alpha, \beta)|]$ vs. n, with $\mathcal{L} = \text{SUOT}$ or USOT and $\alpha = \beta = \mathcal{N}(\mathbf{0}_d, \mathbf{I}_d)$

$\Rightarrow$ Convergence rate in $\mathcal{O}(1/n)$ for *any* dimension d

Strong Duality

- $(\alpha, \beta) \in \mathcal{M}_+(\mathbb{R}^d) \times \mathcal{M}_+(\mathbb{R}^d)$
- $\varphi_i^\circ(x) \triangleq -\varphi_i^*(-x)$, with φ_i^* : Legendre transform of φ_i

Primal problems

$$\begin{aligned}
\mathrm{UOT}(\alpha, \beta) &= \inf_{(\pi_1,\pi_2)\in\mathcal{M}_+(\mathbb{R}^d)^2} \mathrm{OT}(\pi_1, \pi_2) + \mathrm{D}_{\varphi_1}(\pi_1|\alpha) + \mathrm{D}_{\varphi_2}(\pi_2|\beta) \\
\mathrm{SUOT}(\alpha, \beta) &= \int_{\mathbb{S}^{d-1}} \mathrm{UOT}\big(\theta^\star_\sharp \alpha, \theta^\star_\sharp \beta\big) \mathrm{d}\boldsymbol{\sigma}(\theta) \\
\mathrm{USOT}(\alpha, \beta) &= \inf_{(\pi_1,\pi_2)\in\mathcal{M}_+(\mathbb{R}^d)^2} \mathrm{SOT}(\pi_1, \pi_2) + \mathrm{D}_{\varphi_1}(\pi_1|\alpha) + \mathrm{D}_{\varphi_2}(\pi_2|\beta)
\end{aligned}$$

Strong duality

$$\mathrm{UOT}(\alpha, \beta) = \sup_{f\oplus g\leq \mathrm{C}} \langle \varphi_1^\circ \circ f, \alpha\rangle + \langle \varphi_2^\circ \circ g, \beta\rangle$$

$$\mathrm{SUOT}(\alpha, \beta) = \sup_{f_{\theta_k}\oplus g_{\theta_k}\leq \mathrm{C}} \frac{1}{K}\sum_{k=1}^{K} \Big[\big\langle \varphi_1^\circ \circ f_{\theta_k}, (\theta_k^\star)_\sharp \alpha\big\rangle + \big\langle \varphi_2^\circ \circ g_{\theta_k}, (\theta_k^\star)_\sharp \beta\big\rangle \Big]$$

$$\mathrm{USOT}(\alpha, \beta) = \sup_{f_{\theta_k}\oplus g_{\theta_k}\leq \mathrm{C}} \langle \varphi_1^\circ \circ (\frac{1}{K}\sum_{k=1}^{K} f_{\theta_k}), \alpha\rangle + \langle \varphi_2^\circ \circ (\frac{1}{K}\sum_{k=1}^{K} g_{\theta_k}), \beta\rangle$$

Computing Sliced Unbalanced (and *vice versa*) OT with the Frank-Wolfe Algorithm

Solving Sliced Unbalanced OT

Frank-Wolfe algorithm

$$\min_{x \in \mathsf{X}} h(x)$$

$\mathsf{X} \subseteq \mathbb{R}^d$ compact and convex, $h : \mathsf{X} \to \mathbb{R}$ convex and differentiable

Initialize $x^{(0)} \in \mathsf{X}$

Repeat for $t = 1, \dots, T$

$$y^{(t)} \leftarrow \operatorname*{argmin}_{y \in \mathsf{X}} \langle y, \nabla h(x^{(t)}) \rangle$$

$$x^{(t+1)} \leftarrow (1 - \gamma_t) x^{(t)} + \gamma_t y^{(t)}$$

Return $h(x^{(T)})$

Application: Solving SUOT(α, β)

$$\max_{f_k \oplus g_k \leq \mathrm{C}} \mathcal{H}(f_k, g_k)$$

with $\mathcal{H}$ the translation-invariant dual of $\mathrm{UOT}\big((\theta_k^\star)_\sharp \alpha, (\theta_k^\star)_\sharp \beta\big)$

Initialize $(f_k^{(0)}, g_k^{(0)})$ s.t. $f_k^{(0)} \oplus g_k^{(0)} \leq \mathrm{C}$

Repeat for $t = 1, \dots, T$

$$(r_k^{(t)}, s_k^{(t)}) \leftarrow \operatorname*{argmax}_{r \oplus s \leq \mathrm{C}} \langle (r, s), \nabla \mathcal{H}(f_k^{(t)}, g_k^{(t)}) \rangle$$

$$(f_k^{(t+1)}, g_k^{(t+1)}) \leftarrow (1 - \gamma_t)(f_k^{(t)}, g_k^{(t)}) + \gamma_t (r_k^{(t)}, s_k^{(t)})$$

Return $\frac{1}{K} \sum_{k=1}^K \mathcal{H}(f_k^{(T)}, g_k^{(T)})$

Solving Sliced Unbalanced OT

Translation-invariant dual of UOT$(\theta_\sharp^\star\alpha, \theta_\sharp^\star\beta)$

$$\mathcal{H}(f_\theta, g_\theta) = \sup_{\lambda\in\mathbb{R}} \mathcal{D}(f_\theta + \lambda, g_\theta - \lambda)$$

with $\mathcal{D}(f_\theta, g_\theta) = \langle \varphi_1^\circ \circ f_\theta, \theta_\sharp^\star\alpha\rangle + \langle \varphi_2^\circ \circ g_\theta, \theta_\sharp^\star\beta\rangle$

If $\mathrm{D}_{\varphi_1} = \mathrm{D}_{\varphi_2} = \rho\mathrm{KL}$,

$$\lambda_\theta^\star \triangleq \operatorname*{argmax}_{\lambda\in\mathbb{R}} \mathcal{D}(f_\theta + \lambda, g_\theta - \lambda) = \frac{\rho^2}{2\rho}\log\left(\frac{\langle e^{-f_\theta/\rho}, \theta_\sharp^\star\alpha\rangle}{\langle e^{-g_\theta/\rho}, \theta_\sharp^\star\beta\rangle}\right)$$

$$\nabla\mathcal{H}(f_\theta, g_\theta) = (\underbrace{\nabla\varphi_1^\circ(f_\theta + \lambda_\theta^\star)\theta_\sharp^\star\alpha}_{\tilde{\alpha}_\theta}, \nabla\,\underbrace{\varphi_2^\circ(g_\theta - \lambda_\theta^\star)\theta_\sharp^\star\beta}_{\tilde{\beta}_\theta})$$

where $\tilde{\alpha}_\theta$, $\tilde{\beta}_\theta$ have the same mass

$$(r_k^{(t)}, s_k^{(t)}) = \operatorname*{argmax}_{r\oplus s\leq \mathrm{C}} \langle (r,s), \nabla\mathcal{H}(f_k^{(t)}, g_k^{(t)})\rangle = \operatorname*{argmax}_{r\oplus s\leq \mathrm{C}} \langle r, \tilde{\alpha}_{\theta_k}\rangle + \langle s, \tilde{\beta}_{\theta_k}\rangle$$

Frank-Wolfe step $\Leftrightarrow$ **Solving** $\mathrm{OT}(\tilde{\alpha}_{\theta_k}, \tilde{\beta}_{\theta_k})$, $k = 1, \dots, K$

$\Rightarrow$ Overall computational complexity in $\mathcal{O}(TKn\log n)$

Solving Unbalanced Sliced OT

Frank-Wolfe algorithm

$$\min_{x \in \mathsf{X}} h(x)$$

$\mathsf{X} \subseteq \mathbb{R}^d$ compact and convex, $h : \mathsf{X} \to \mathbb{R}$ convex and differentiable

Initialize $x^{(0)} \in \mathsf{X}$

Repeat for $t = 1, \dots, T$

$$y^{(t)} \leftarrow \operatorname*{argmin}_{y \in \mathsf{X}} \langle y, \nabla h(x^{(t)}) \rangle$$

$$x^{(t+1)} \leftarrow (1 - \gamma_t) x^{(t)} + \gamma_t y^{(t)}$$

Return $h(x^{(T)})$

Application: Solving USOT(α, β)

$$\max_{\substack{\bar{f} = \frac{1}{K} \sum_{k=1}^K f_k,\ \bar{g} = \frac{1}{K} \sum_{k=1}^K g_k, \\ f_k \oplus g_k \leq \mathrm{C}}} \mathcal{H}(\bar{f}, \bar{g})$$

with $\mathcal{H}$ the translation-invariant dual of USOT(α, β)

Initialize $(f_k^{(0)}, g_k^{(0)})$ s.t. $f_k^{(0)} \oplus g_k^{(0)} \leq \mathrm{C}$

Repeat for $t = 1, \dots, T$

$$(\bar{r}^{(t)}, \bar{s}^{(t)}) \leftarrow \operatorname*{argmax}_{\substack{\bar{r} = \frac{1}{K} \sum_{k=1}^K r_k, \\ \bar{s} = \frac{1}{K} \sum_{k=1}^K s_k, \\ r_k \oplus s_k \leq \mathrm{C}}} \langle (\bar{r}, \bar{s}), \nabla \mathcal{H}(\bar{f}^{(t)}, \bar{g}^{(t)}) \rangle$$

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

Return $\mathcal{H}(\bar{f}^{(T)}, \bar{g}^{(T)})$

Solving Unbalanced Sliced OT

Translation-invariant dual of USOT(α, β)

$$\mathcal{H}(\bar{f}, \bar{g}) = \sup_{\lambda \in \mathbb{R}} \mathcal{D}(\bar{f} + \lambda, \bar{g} - \lambda)$$

with $\mathcal{D}(\bar{f}, \bar{g}) = \langle \varphi_1^\circ \circ \bar{f}, \alpha \rangle + \langle \varphi_2^\circ \circ \bar{g}, \beta \rangle$

If $\mathrm{D}_{\varphi_1} = \mathrm{D}_{\varphi_2} = \rho \mathrm{KL}$,

$$\bar{\lambda}^\star \triangleq \operatorname*{argmax}_{\lambda \in \mathbb{R}} \mathcal{D}(\bar{f} + \lambda, \bar{g} - \lambda) = \frac{\rho^2}{2\rho} \log \left(\frac{\langle e^{-\bar{f}/\rho}, \alpha \rangle}{\langle e^{-\bar{g}/\rho}, \beta \rangle} \right)$$

$$\nabla \mathcal{H}(\bar{f}, \bar{g}) = (\underbrace{\nabla \varphi_1^\circ(\bar{f} + \bar{\lambda}^\star)\alpha}_{\tilde{\alpha}}, \nabla \underbrace{\varphi_2^\circ(\bar{g} - \bar{\lambda}^\star)\beta}_{\tilde{\beta}})$$

where $\tilde{\alpha}$, $\tilde{\beta}$ have the same mass

$$(\bar{r}^{(t)}, \bar{s}^{(t)}) = \operatorname*{argmax}_{r_k \oplus s_k \leq \mathrm{C}} \langle (\bar{r}, \bar{s}), \nabla \mathcal{H}(\bar{f}^{(t)}, \bar{g}^{(t)}) \rangle = \operatorname*{argmax}_{r_k \oplus s_k \leq \mathrm{C}} \langle \bar{r}, \tilde{\alpha} \rangle + \langle \bar{s}, \tilde{\beta} \rangle$$

Frank-Wolfe step $\Leftrightarrow$ **Solving** $\mathrm{OT}((\theta_k^\star)_\sharp \tilde{\alpha}, (\theta_k^\star)_\sharp \tilde{\beta})$, $k = 1, \ldots, K$

$\Rightarrow$ Overall computational complexity in $\mathcal{O}(TKn \log n)$

Experiments

Color Transfer

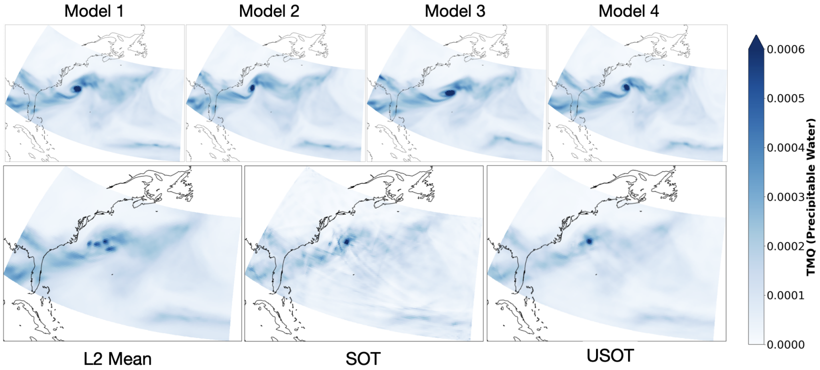
(*1st and 2nd columns*) Source and target

(*3rd column*) SOT gradient flow

(*4th column*) USOT gradient flow

(*last*) % of mass change by USOT (*red*: mass creation, *blue* destruction)

Barycenters for Geophysical Data



(*First row*) 4 climate models (different evolutions of a tropical cyclone)

(*Second row*) Different barycenters

Conclusion

Our contributions

- Definition of two new losses merging unbalanced and sliced OT
- Theoretical analysis
- Efficient and modular Frank-Wolfe algorithm
- Encouraging empirical results

Some perspectives

- Open theoretical questions (sample complexity for USOT? Strong duality for $\boldsymbol{\sigma}$?)
- Address new large-scale applications (challenge: sensitivity to ρ in $\mathrm{D}_\varphi = \rho\mathrm{KL}$)

Link to paper: `https://arxiv.org/abs/2306.07176`

References

Bai, Yikun et al. (2022). “Sliced optimal partial transport”. In: *CVPR*.

Bonneel, Nicolas et al. (2015). “Sliced and Radon Wasserstein Barycenters of Measures”. In: *Journal of Mathematical Imaging and Vision* 1.51, pp. 22–45. DOI: 10.1007/s10851-014-0506-3.

Bonnotte, Nicolas (2013). “Unidimensional and Evolution Methods for Optimal Transportation”. PhD thesis. Paris 11.

Carrière, Mathieu, Marco Cuturi, and Steve Oudot (2017). “Sliced Wasserstein Kernel for Persistence Diagrams”. In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. International Convention Centre, Sydney, Australia: PMLR, pp. 664–673.

Chizat, Lenaic et al. (2018). "Scaling Algorithms for Unbalanced Optimal Transport Problems". In: *Mathematics of Computation* 87.314, pp. 2563–2609. ISSN: 00255718, 10886842. (Visited on 04/08/2024).

Cuturi, Marco (2013). "Sinkhorn distances: Lightspeed computation of optimal transport". In: *Advances in Neural Information Processing Systems*, pp. 2292–2300.

Dai, Biwei and Uros Seljak (2021). "Sliced Iterative Normalizing Flows". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, pp. 2352–2364.

Deshpande, Ishan, Ziyu Zhang, and Alexander G Schwing (2018). "Generative modeling using the sliced Wasserstein distance". In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3483–3491.

Deshpande, Ishan et al. (2019). "Max-Sliced Wasserstein Distance and its use for GANs". In: *IEEE Conference on Computer Vision and Pattern Recognition*.

Kolouri, Soheil, Gustavo K Rohde, and Heiko Hoffmann (2018). "Sliced Wasserstein Distance for Learning Gaussian Mixture Models". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3427–3436.

Kolouri, Soheil, Yang Zou, and Gustavo K Rohde (2016). "Sliced-Wasserstein Kernels for Probability distributions". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4876–4884.

Kolouri, Soheil et al. (2019a). "Generalized Sliced Wasserstein Distances". In: *NeurIPS 2019*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc.

Kolouri, Soheil et al. (2019b). "Sliced Wasserstein Auto-Encoders". In: *International Conference on Learning Representations*.

Liero, Matthias, Alexander Mielke, and Giuseppe Savaré (2018). "Optimal entropy-transport problems and a new Hellinger–Kantorovich distance between positive measures". In: *Inventiones mathematicae* 211.3, pp. 969–1117.

Liutkus, Antoine et al. (2019). "Sliced-Wasserstein Flows: Nonparametric Generative Modeling via Optimal Transport and Diffusions". In: *Proceedings of the 36th International Conference on Machine Learning*. Ed. by Kamalika Chaudhuri and Ruslan Salakhutdinov. Vol. 97. Proceedings of Machine Learning Research. Long Beach, California, USA: PMLR, pp. 4104–4113.

Manole, Tudor, Sivaraman Balakrishnan, and Larry Wasserman (2019). *Minimax Confidence Intervals for the Sliced Wasserstein Distance*. arXiv: 1909.07862 [math.ST].

Nadjahi, Kimia et al. (2019). "Asymptotic Guarantees for Learning Generative Models with the Sliced-Wasserstein Distance". In: *NeurIPS 2019*. Ed. by H. Wallach et al. Vol. 32. Curran Associates, Inc.

Nadjahi, Kimia et al. (2020). "Statistical and Topological Properties of Sliced Probability Divergences". In: *NeurIPS 2020*. Ed. by H. Larochelle et al. Vol. 33. Curran Associates, Inc., pp. 20802–20812.

Nguyen, Khai et al. (2021). "Distributional Sliced-Wasserstein and Applications to Generative Modeling". In: *International Conference on Learning Representations*.

Ohana, Ruben et al. (2023). “Shedding a PAC-Bayesian Light on Adaptive Sliced-Wasserstein Distances”. In: *Proceedings of the 40th International Conference on Machine Learning*. Ed. by Andreas Krause et al. Vol. 202. Proceedings of Machine Learning Research. PMLR, pp. 26451–26473.

Park, Sangmin and Dejan Slepčev (2023). *Geometry and analytic properties of the sliced Wasserstein space*. arXiv: 2311.05134 [math.AP].

Peyré, Gabriel and Marco Cuturi (2019). “Computational Optimal Transport: With Applications to Data Science”. In: *Foundations and Trends® in Machine Learning* 11.5-6, pp. 355–607. ISSN: 1935-8237. DOI: 10.1561/2200000073.

Rabin, J. et al. (2012). “Wasserstein Barycenter and Its Application to Texture Mixing”. In: *Scale Space and Variational Methods in Computer Vision*. Ed. by Alfred M. Bruckstein et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 435–446. ISBN: 978-3-642-24785-9.

Rabin, Julien, Julie Delon, and Yann Gousseau (2011). "Transportation Distances on the Circle". In: *Journal of Mathematical Imaging and Vision* 41.1–2, 147–167. DOI: 10.1007/s10851-011-0284-0.

Rachev, Svetlozar T and Ludger Rüschendorf (1998). *Mass Transportation Problems: Volume I: Theory*. Vol. 1. Springer Science & Business Media.

Séjourné, Thibault, Gabriel Peyré, and François-Xavier Vialard (2022). "Unbalanced Optimal Transport, from theory to numerics". In: *arXiv preprint arXiv:2211.08775*.

Séjourné, Thibault, François-Xavier Vialard, and Gabriel Peyré (2022). "Faster Unbalanced Optimal Transport: Translation invariant Sinkhorn and 1-D Frank-Wolfe". In: *International Conference on Artificial Intelligence and Statistics, AISTATS 2022, 28-30 March 2022, Virtual Event*. Ed. by Gustau Camps-Valls, Francisco J. R. Ruiz, and Isabel Valera. Vol. 151. Proceedings of Machine Learning Research. PMLR, pp. 4995–5021.

Séjourné, Thibault et al. (2023). *Unbalanced Optimal Transport meets Sliced-Wasserstein*. arXiv: 2306.07176 [cs.LG].

Tanguy, Eloi, Rémi Flamary, and Julie Delon (2024). *Properties of Discrete Sliced Wasserstein Losses*. arXiv: 2307.10352 [stat.ML].

Wu, Jiqing et al. (2019). “Sliced Wasserstein Generative Models”. In: *IEEE Conference on Computer Vision and Pattern Recognition*.

Xi, Jiaqi and Jonathan Niles-Weed (2022). “Distributional Convergence of the Sliced Wasserstein Process”. In: *Advances in Neural Information Processing Systems*. Ed. by S. Koyejo et al. Vol. 35. Curran Associates, Inc., pp. 13961–13973.