Rényi regularized OT—Interpolating between KL and unregularized OT

DWCAA 24 - Canazei, Italy

Jonas Bresch, Co-work with Viktor Stein
TU Berlin, Institut für Mathematik

September 18, 2024

Table of Contents

– **Contribution.** Regularize OT between probability distributions $\mu, \nu \in \mathcal{P}(\mathbb{X})$, $\mathbb{X} \subseteq \mathbb{R}^d$

$$\operatorname*{arg\,min}_{\pi \in \Pi(\mu,\nu)} \langle c, \pi \rangle := \int_{\mathbb{X} \times \mathbb{X}} c(x,y) \, \mathrm{d}\pi(x,y), \qquad \text{with cost } c \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$$

using the α-Rényi-divergences R_α for $\alpha \in (0,1)$.

- **Contribution.** Regularize OT between probability distributions $\mu, \nu \in \mathcal{P}(\mathbb{X})$, $\mathbb{X} \subseteq \mathbb{R}^d$

$$\operatorname*{arg\,min}_{\pi \in \Pi(\mu,\nu)} \langle c, \pi \rangle := \int_{\mathbb{X} \times \mathbb{X}} c(x,y) \, \mathrm{d}\pi(x,y), \qquad \text{with cost } c \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}_+$$

  using the α-Rényi-divergences R_α for $\alpha \in (0,1)$.
- **Prior work.** Regularization with $\mathrm{KL} = \lim_{\alpha \nearrow 1} R_\alpha$ [3] and with q-Tsallis divergence $T_q = (q-1)^{-1} [\exp((q-1)R_q) - 1]$ for $q > 0$ [4].

- **Method.** Solve primal problem with mirror descent [1] and dual problem with subgradient descent.

- **Method.** Solve primal problem with mirror descent [1] and dual problem with subgradient descent.
- **Result.** Rényi-regularized OT plans outperform KL / Tsallis regularized OT plans on real and synthetic data.

Introduction II

- **Method.** Solve primal problem with mirror descent [1] and dual problem with subgradient descent.
- **Result.** Rényi-regularized OT plans outperform KL / Tsallis regularized OT plans on real and synthetic data.
- **Novelty.** $R_\alpha \notin \{f\text{-divergence}, \text{Bregman divergence}\}$ and R_α not "separable" due to the logarithm.

Definition I

- The transport polytope is

$$\Pi(\mu,\nu) := \{\pi \in \mathcal{P}(\mathbb{X}\times\mathbb{X}) : \pi(A\times\mathbb{X}) = \mu(A),\ \pi(\mathbb{X}\times B) = \nu(B) \quad \forall A, B \in \mathcal{B}(\mathbb{X})\}$$

Definition I

- The transport polytope is

$$\Pi(\mu,\nu) := \{\pi \in \mathcal{P}(\mathbb{X}\times\mathbb{X}) : \pi(A\times\mathbb{X}) = \mu(A),\ \pi(\mathbb{X}\times B) = \nu(B) \quad \forall A, B \in \mathcal{B}(\mathbb{X})\}$$

- $\mathcal{D}_p := \{c = d^p \mid d\colon\ \mathbb{X}\times\mathbb{X} \mapsto \mathbb{R}_+$ is lower semicontinuous metric$\}$ the set of cost functions, where $p \in [1,\infty)$

Definition I

- The transport polytope is

$$\Pi(\mu,\nu) := \{\pi \in \mathcal{P}(\mathbb{X}\times\mathbb{X}) : \pi(A\times\mathbb{X}) = \mu(A),\ \pi(\mathbb{X}\times B) = \nu(B) \quad \forall A, B \in \mathcal{B}(\mathbb{X})\}$$

- $\mathcal{D}_p := \{c = d^p \mid d\colon\ \mathbb{X}\times\mathbb{X} \mapsto \mathbb{R}_+$ is lower semicontinuous metric$\}$ the set of cost functions, where $p \in [1,\infty)$
- $\mathcal{P}_p(\mathbb{X})$ is the set of probability measures with finite p-th moment.

- **Definition.** The α-Rényi-divergence of *order* $\alpha \in (0, 1)$ is

$$R_\alpha \colon \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \to \mathbb{R}_+, \quad (\mu \mid \nu) \mapsto \frac{1}{\alpha - 1} \ln\left(\int_{\mathbb{X}} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\tau}(x) \right)^{\alpha} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\tau}(x) \right)^{1-\alpha} \mathrm{d}\tau(x) \right),$$

- **Definition.** The α-Rényi-divergence of *order* $\alpha \in (0,1)$ is

$$R_\alpha\colon \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \to \mathbb{R}_+, \quad (\mu \mid \nu) \mapsto \frac{1}{\alpha - 1} \ln\left(\int_{\mathbb{X}} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\tau}(x)\right)^{\alpha} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\tau}(x)\right)^{1-\alpha} \mathrm{d}\tau(x)\right),$$

- where τ is a finite *reference* measure, e.g. $\tau = \mu + \nu \in \mathcal{M}(\mathbb{X})_+$, is chosen such that $\mu, \nu \ll \tau$,

Definition II

- **Definition.** The α-Rényi-divergence of *order* $\alpha \in (0,1)$ is

$$R_\alpha \colon \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \to \mathbb{R}_+, \quad (\mu \mid \nu) \mapsto \frac{1}{\alpha - 1} \ln\left(\int_{\mathbb{X}} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\tau}(x)\right)^{\alpha} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\tau}(x)\right)^{1-\alpha} \mathrm{d}\tau(x) \right),$$

- where τ is a finite *reference* measure, e.g. $\tau = \mu + \nu \in \mathcal{M}(\mathbb{X})_+$, is chosen such that $\mu, \nu \ll \tau$,
- and $\ln(0) := -\infty$.

Definition II

- **Definition.** The α-Rényi-divergence of *order* $\alpha \in (0,1)$ is

$$R_\alpha \colon \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \to \mathbb{R}_+, \quad (\mu \mid \nu) \mapsto \frac{1}{\alpha - 1} \ln\left(\int_{\mathbb{X}} \left(\frac{\mathrm{d}\mu}{\mathrm{d}\tau}(x)\right)^{\alpha} \left(\frac{\mathrm{d}\nu}{\mathrm{d}\tau}(x)\right)^{1-\alpha} \mathrm{d}\tau(x) \right),$$

- where τ is a finite *reference* measure, e.g. $\tau = \mu + \nu \in \mathcal{M}(\mathbb{X})_+$, is chosen such that $\mu, \nu \ll \tau$,
- and $\ln(0) := -\infty$.
- **Remark.** $R_\alpha(\cdot \mid \mu \otimes \nu)$ is well defined on $\Pi(\mu, \nu)$, i.e. $\pi \in \Pi(\mu,\nu)$ implies that $\pi \ll \mu \otimes \nu$.

Table of Contents

- For regularization parameter $\gamma \in [0, \infty]$ and $\alpha \in (0,1)$, the *restricted transport polytope*,

$$\Pi_\gamma^\alpha(\mu, \nu) := \{\pi \in \Pi(\mu, \nu) : R_\alpha(\pi \mid \mu \otimes \nu) \leq \gamma\},$$

is weakly compact, since $R_\alpha(\cdot \mid \mu \otimes \nu)$ is weakly lsc. and $\Pi(\mu, \nu)$ is weakly compact.

- For regularization parameter $\gamma \in [0, \infty]$ and $\alpha \in (0,1)$, the *restricted transport polytope*,

$$\Pi_\gamma^\alpha(\mu,\nu) := \{\pi \in \Pi(\mu,\nu) : R_\alpha(\pi \mid \mu \otimes \nu) \leq \gamma\},$$

  is weakly compact, since $R_\alpha(\,\cdot \mid \mu \otimes \nu)$ is weakly lsc. and $\Pi(\mu,\nu)$ is weakly compact.

- **Definition.** The **Rényi-Sinkhorn distance** between $\mu, \nu \in \mathcal{P}_p(\mathbb{X})$ with cost $c \in \mathcal{D}_p$ is

$$d_{c,\gamma,\alpha} \colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to \mathbb{R}, \qquad (\mu,\nu) \mapsto \min\left\{\langle c, \pi\rangle^{\frac{1}{p}} : \pi \in \Pi_\gamma^\alpha(\mu,\nu)\right\}. \tag{1}$$

Regularization models

- For regularization parameter $\gamma \in [0, \infty]$ and $\alpha \in (0,1)$, the *restricted transport polytope*,
$$\Pi_\gamma^\alpha(\mu,\nu) := \{\pi \in \Pi(\mu,\nu) : R_\alpha(\pi \mid \mu \otimes \nu) \leq \gamma\},$$
is weakly compact, since $R_\alpha(\,\cdot \mid \mu \otimes \nu)$ is weakly lsc. and $\Pi(\mu,\nu)$ is weakly compact.
- **Definition.** The **Rényi-Sinkhorn distance** between $\mu, \nu \in \mathcal{P}_p(\mathbb{X})$ with cost $c \in \mathcal{D}_p$ is
$$d_{c,\gamma,\alpha} \colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to \mathbb{R}, \qquad (\mu,\nu) \mapsto \min\left\{ \langle c, \pi \rangle^{\frac{1}{p}} : \pi \in \Pi_\gamma^\alpha(\mu,\nu) \right\}. \tag{1}$$
- **Remark.** (1) is a convex problem, with unique minimizer.

- For regularization parameter $\gamma \in [0, \infty]$ and $\alpha \in (0, 1)$, the *restricted transport polytope*,

$$\Pi_\gamma^\alpha(\mu, \nu) := \{\pi \in \Pi(\mu, \nu) : R_\alpha(\pi \mid \mu \otimes \nu) \le \gamma\},$$

  is weakly compact, since $R_\alpha(\,\cdot \mid \mu \otimes \nu)$ is weakly lsc. and $\Pi(\mu, \nu)$ is weakly compact.
- **Definition.** The **Rényi-Sinkhorn distance** between $\mu, \nu \in \mathcal{P}_p(\mathbb{X})$ with cost $c \in \mathcal{D}_p$ is

$$d_{c,\gamma,\alpha} \colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to \mathbb{R}, \qquad (\mu, \nu) \mapsto \min\left\{\langle c, \pi\rangle^{\frac{1}{p}} : \pi \in \Pi_\gamma^\alpha(\mu, \nu)\right\}. \qquad (1)$$

- **Remark.** (1) is a convex problem, with unique minimizer.

Theorem 1 (B., Stein [2])

$\mathcal{P}_p(\mathbb{X})^2 \ni (\mu, \nu) \mapsto \mathbf{1}_{[\mu \neq \nu]} d_{c,\gamma,\alpha}(\mu, \nu)$ is a metric for $\alpha \in (0, 1)$, $\gamma \in [0, \infty]$, $c \in \mathcal{D}_p$.

The Dual Point of View - Penalizing the Constraint

- Instead of restricting the domain of the optimization problem, one can penalize the Rényi divergence constraint in (1).

- Instead of restricting the domain of the optimization problem, one can penalize the Rényi divergence constraint in (1).
- **Definition.** The **dual Rényi-Divergence-Sinkhorn distance** for $\alpha \in (0,1)$, $\lambda \in (0,\infty]$ is

$$d_c^{\alpha,\lambda} \colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to \mathbb{R}, \qquad (\mu,\nu) \mapsto \langle c, \pi_c^{\alpha,\lambda}(\mu,\nu)\rangle^{\frac{1}{p}},$$

$$\text{where } \pi_c^{\alpha,\lambda}(\mu,\nu) \in \arg\min \left\{ \langle c, \pi \rangle + \frac{1}{\lambda} R_\alpha(\pi \mid \mu \otimes \nu) : \pi \in \Pi(\mu,\nu) \right\}. \tag{2}$$

- Instead of restricting the domain of the optimization problem, one can penalize the Rényi divergence constraint in (1).
- **Definition.** The **dual Rényi-Divergence-Sinkhorn distance** for $\alpha \in (0,1)$, $\lambda \in (0,\infty]$ is

$$d_c^{\alpha,\lambda}\colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to \mathbb{R}, \qquad (\mu,\nu) \mapsto \langle c, \pi_c^{\alpha,\lambda}(\mu,\nu)\rangle^{\frac{1}{p}},$$

$$\text{where } \pi_c^{\alpha,\lambda}(\mu,\nu) \in \arg\min\left\{\langle c,\pi\rangle + \frac{1}{\lambda} R_\alpha(\pi \mid \mu \otimes \nu) : \pi \in \Pi(\mu,\nu)\right\}. \tag{2}$$

THEOREM 2 (B., STEIN [2])

Problem (2) has a unique solution for every $\alpha \in (0,1), \lambda > 0$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{X})$.

- **Rényi-Sinkhorn distance** $d_{c,\gamma,\alpha}(\mu,\nu)$ and **dual Rényi-Sinkhorn distance** $d_c^{\alpha,\lambda}(\mu,\nu)$ are equivalent:

$$\text{for } \mu,\nu \in \mathcal{P}_p(\mathbb{X}) \text{ and } \gamma > 0, \text{ there exists } \lambda \in [0,\infty], \text{ such that}$$
$$\langle c, \pi_c^{\alpha,\lambda}(\mu,\nu)\rangle = d_{c,\gamma,\alpha}(\mu,\nu)^p.$$

- **Rényi-Sinkhorn distance** $d_{c,\gamma,\alpha}(\mu,\nu)$ and **dual Rényi-Sinkhorn distance** $d_c^{\alpha,\lambda}(\mu,\nu)$ are equivalent:

$$\text{for } \mu,\nu \in \mathcal{P}_p(\mathbb{X}) \text{ and } \gamma > 0, \text{ there exists } \lambda \in [0,\infty], \text{ such that}$$
$$\langle c, \pi_c^{\alpha,\lambda}(\mu,\nu)\rangle = d_{c,\gamma,\alpha}(\mu,\nu)^p.$$

- The problem (2) yields the regularized OT problem

$$\mathrm{OT}_{\frac{1}{\lambda},\alpha}\colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to [0,\infty),\ (\mu,\nu) \mapsto \min_{\pi\in\Pi(\mu,\nu)} \langle c,\pi\rangle + \frac{1}{\lambda} R_\alpha(\pi \mid \mu\otimes\nu). \tag{3}$$

Lagrangian point of view and pre-metric

- **Rényi-Sinkhorn distance** $d_{c,\gamma,\alpha}(\mu,\nu)$ and **dual Rényi-Sinkhorn distance** $d_c^{\alpha,\lambda}(\mu,\nu)$ are equivalent:

$$\text{for } \mu,\nu \in \mathcal{P}_p(\mathbb{X}) \text{ and } \gamma > 0, \text{ there exists } \lambda \in [0,\infty], \text{ such that}$$
$$\langle c, \pi_c^{\alpha,\lambda}(\mu,\nu)\rangle = d_{c,\gamma,\alpha}(\mu,\nu)^p.$$

- The problem (2) yields the regularized OT problem

$$\mathrm{OT}_{\frac{1}{\lambda},\alpha}\colon \mathcal{P}_p(\mathbb{X}) \times \mathcal{P}_p(\mathbb{X}) \to [0,\infty),\ (\mu,\nu) \mapsto \min_{\pi \in \Pi(\mu,\nu)} \langle c, \pi\rangle + \frac{1}{\lambda} R_\alpha(\pi \mid \mu \otimes \nu). \tag{3}$$

Theorem 3 (B., Stein [2])

$\mathcal{P}(\mathbb{X})^2 \ni (\mu,\nu) \mapsto \mathbf{1}_{[\mu\neq\nu]}\,\mathrm{OT}_{\frac{1}{\lambda},\alpha}(\mu,\nu)$ is a metric for $\alpha \in (0,1), \lambda \in (0,\infty]$, $c \in \mathcal{D}_p$.

Theory

– The aim is to employ the duality theory by Fenchel-Rockefellar.

- The aim is to employ the duality theory by Fenchel-Rockefellar.
- Therefore, we can prove a Pre-conjugate formulation of the Rényi divergence w.r.t. the first component.

- The aim is to employ the duality theory by Fenchel-Rockefellar.
- Therefore, we can prove a Pre-conjugate formulation of the Rényi divergence w.r.t. the first component.

THEOREM 4 (B., STEIN [2] — PRE-CONVEX CONJUGATE OF $R_\alpha(\cdot \mid \theta)$)

For $\theta \in \mathcal{M}(\mathbb{X})$ we have

$$\big[R_\alpha(\cdot \mid \theta)\big]_*(g) = \begin{cases} \ln\left(\langle |g|^{\frac{\alpha}{\alpha-1}}, \theta\rangle\right) + const_\alpha, & \text{if } g \in \mathcal{C}_0(\mathbb{X}),\ g < 0, \\ +\infty & \text{else.} \end{cases}$$

Theory

- The aim is to employ the duality theory by Fenchel-Rockefellar.
- Therefore, we can prove a Pre-conjugate formulation of the Rényi divergence w.r.t. the first component.

Theorem 4 (B., Stein [2] — Pre-convex conjugate of $R_\alpha(\cdot \mid \theta)$)

For $\theta \in \mathcal{M}(\mathbb{X})$ we have

$$\left[R_\alpha(\cdot \mid \theta)\right]_*(g) = \begin{cases} \ln\left(\langle |g|^{\frac{\alpha}{\alpha-1}}, \theta\rangle\right) + const_\alpha, & \text{if } g \in \mathcal{C}_0(\mathbb{X}),\ g < 0, \\ +\infty & \text{else.} \end{cases}$$

- **Remark.** The latter theorem holds not only for compact $\mathbb{X}$ but also for $\mathbb{X} = \mathbb{R}^d$.

Theorem 5 (B., Stein [2] — Dual problem)

If $\mathbb{X} \subset \mathbb{R}^d$ ***compact****, then the Fenchel-Rockafellar theorem yields the strong duality*

$$(3) = \max_{\substack{f,g \in \mathcal{C}(\mathbb{X}) \\ f \oplus g < c}} \langle f \oplus g, \mu \otimes \nu \rangle - \frac{1}{\lambda} \ln \left(\left\langle (c - f \oplus g)^{\frac{\alpha}{\alpha - 1}}, \mu \otimes \nu \right\rangle \right) + const_{\alpha,\lambda}. \tag{4}$$

Theorem 5 (B., Stein [2] — Dual problem)

*If $\mathbb{X} \subset \mathbb{R}^d$ **compact**, then the Fenchel-Rockafellar theorem yields the strong duality*

$$(3) = \max_{\substack{f,g \in \mathcal{C}(\mathbb{X}) \\ f \oplus g < c}} \langle f \oplus g, \mu \otimes \nu \rangle - \frac{1}{\lambda} \ln \left(\left\langle (c - f \oplus g)^{\frac{\alpha}{\alpha - 1}}, \mu \otimes \nu \right\rangle \right) + const_{\alpha,\lambda}. \tag{4}$$

Corollary 6 (B., Stein [2] — Representation of $\pi_c^{\alpha,\lambda}$)

For unique solution $\pi_c^{\alpha,\lambda}(\mu, \nu)$ of (2), and optimal dual potentials $\hat{f}, \hat{g} \in \mathcal{C}(\mathbb{X})$ from (4):

$$\pi_c^{\alpha,\lambda} = \frac{(c - \hat{f} \oplus \hat{g})^{\frac{1}{\alpha - 1}}}{\left\langle (c - \hat{f} \oplus \hat{g})^{\frac{1}{\alpha - 1}}, \mu \otimes \nu \right\rangle} \mu \otimes \nu.$$

Theorem 5 (B., Stein [2] — Dual problem)

*If $\mathbb{X} \subset \mathbb{R}^d$ **compact**, then the Fenchel-Rockafellar theorem yields the strong duality*

$$(3) = \max_{\substack{f,g \in \mathcal{C}(\mathbb{X}) \\ f \oplus g < c}} \langle f \oplus g, \mu \otimes \nu \rangle - \frac{1}{\lambda} \ln \left(\langle (c - f \oplus g)^{\frac{\alpha}{\alpha-1}}, \mu \otimes \nu \rangle \right) + const_{\alpha,\lambda}. \tag{4}$$

Corollary 6 (B., Stein [2] — Representation of $\pi_c^{\alpha,\lambda}$)

For unique solution $\pi_c^{\alpha,\lambda}(\mu,\nu)$ of (2), and optimal dual potentials $\hat{f}, \hat{g} \in \mathcal{C}(\mathbb{X})$ from (4):

$$\pi_c^{\alpha,\lambda} = \frac{(c - \hat{f} \oplus \hat{g})^{\frac{1}{\alpha-1}}}{\langle (c - \hat{f} \oplus \hat{g})^{\frac{1}{\alpha-1}}, \mu \otimes \nu \rangle} \mu \otimes \nu.$$

– **Remark.** We have: $\operatorname{supp}(\pi_c^{\alpha,\lambda}(\mu,\nu)) = \operatorname{supp}(\mu \otimes \nu)$ and uniqueness of $(\hat{f}, \hat{g})$ up to additive constants.

Table of Contents

Theorem 7 (B., Stein [2] — Interpolation properties I)

For fixed $\alpha \in (0,1)$, *we have*

$$\mathrm{OT}_{\frac{1}{\lambda},\alpha}(\mu,\nu) \to \mathrm{OT}(\mu,\nu) \quad \textit{for} \quad \lambda \to \infty$$

and

$$\mathrm{OT}_{\frac{1}{\lambda},\alpha}(\mu,\nu) \to \langle c, \mu \otimes \nu \rangle \quad \textit{for} \quad \lambda \to 0.$$

with

$$\pi_c^{\alpha,\lambda}(\mu,\nu) \rightharpoonup \arg\min \left\{ R_\alpha(\pi \mid \mu \otimes \nu) : \pi \in \Pi(\mu,\nu),\ \langle c, \pi \rangle = \mathrm{OT}(\mu,\nu) \right\}, \qquad \lambda \to \infty,$$
$$\pi_c^{\alpha,\lambda}(\mu,\nu) \rightharpoonup \mu \otimes \nu, \qquad \lambda \to 0.$$

Theorem 8 (B., Stein [2] — Interpolation properties II)

For fixed $\lambda > 0$, we have

$$\mathrm{OT}_{\frac{1}{\lambda},\alpha}(\mu,\nu) \to \mathrm{OT}(\mu,\nu) \quad \textit{for} \quad \alpha \searrow 0$$

and

$$\mathrm{OT}_{\frac{1}{\lambda},\alpha}(\mu,\nu) \to \mathrm{OT}_{\frac{1}{\lambda},\alpha'}(\mu,\nu) \quad \textit{for } \alpha \nearrow \alpha' \in (0,1].$$

and

$$\pi_c^{\alpha,\lambda}(\mu,\nu) \rightharpoonup \pi \in \Pi(\mu,\nu) \quad \textit{with} \quad \langle c,\pi\rangle = \mathrm{OT}(\mu,\nu) \qquad \alpha \searrow 0, \tag{5}$$

$$\pi_c^{\alpha,\lambda}(\mu,\nu) \rightharpoonup \pi_c^{\alpha',\lambda}(\mu,\nu), \qquad \alpha \nearrow \alpha' \in (0,1]. \tag{6}$$

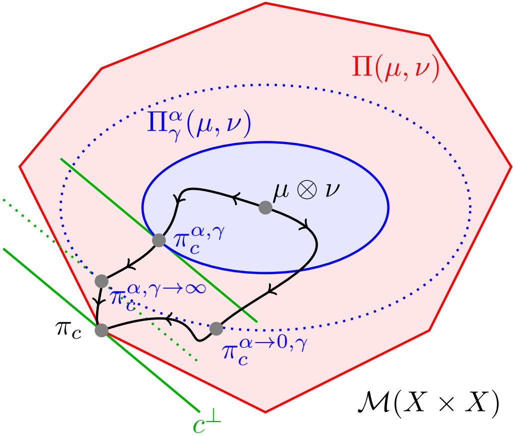


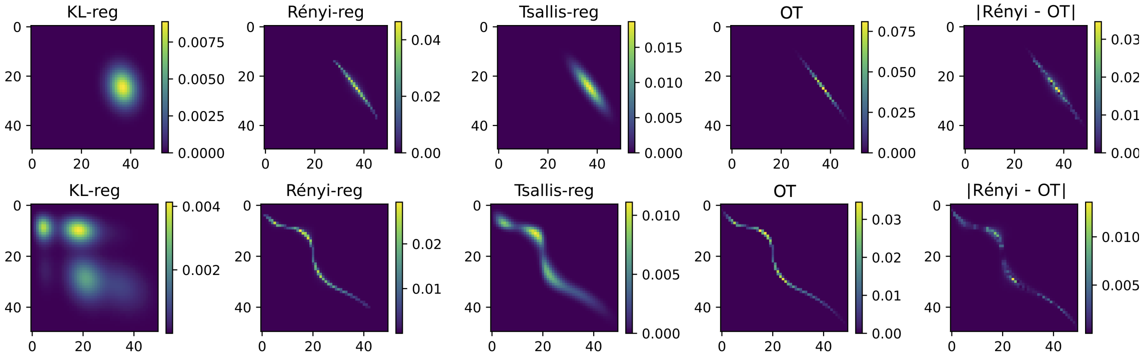
FIGURE 1: Transport polytope $\Pi(\mu, \nu)$ with the restricted transport polytope $\Pi_\gamma^\alpha(\mu, \nu)$. Convergence of $\pi_c^{\alpha,\gamma} \to \pi_c$ to the unregularized OT plan with $\alpha \to 0$ and $\gamma \to \infty$. (Plot inspired by Cuturi [3].)

Table of Contents

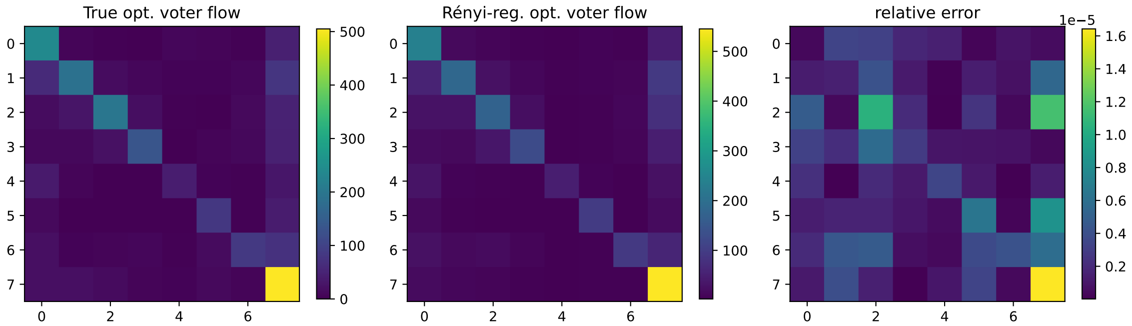
– $\Phi_c^{\alpha,\lambda}$ is not Lipschitz continuous in zero, but locally Lipschitz in

$$\{\pi \in \Pi(\mu,\nu) : \pi|_{\mathrm{supp}(\mu\otimes\nu)} > 0\} = \Pi(\mu,\nu) \cap \mathrm{int}(\mathrm{dom}\, h),$$

where h generates KL divergence $\Rightarrow$ modified mirror descent algorithm with Polyak step size converges [5]. In each iteration one Sinkhorn projection is performed.



Numerical Experiments - Predicting voter migration



distance matrix	regularizer	abs error ± std	KL error	mean squared error	scalar
$\boldsymbol{M}^{\mathrm{Eucl}}$	$\mathrm{KL}_{\lambda=1}$	$2.4221 \times 10^1 \pm 2.848 \times 10^1$	8.422×10^2	9.008×10^4	
	$\mathrm{T}^{\dagger}_{\lambda=1,q=10}$	$9.409 \pm 1.529 \times 10^1$	3.173×10^2	2.063×10^4	10
	OT	$1.845 \times 10^1 \pm 2.358 \times 10^1$	7.655×10^2	5.738×10^4	
	$\mathrm{R}_{\lambda=1,\alpha=0.3}$	$\mathbf{6.611 \pm 7.868}$	2.128×10^2	6.759×10^3	

Thank you for your attention!

References I

[1] Amir Beck and Marc Teboulle. “Mirror descent and nonlinear projected subgradient methods for convex optimization”. In: *Operations Research Letters* 31.3 (2003), pp. 167–175. DOI: 10.1016/S0167-6377(02)00231-6.

[2] Jonas Bresch and Viktor Stein. *Interpolating between Optimal Transport and KL regularized Optimal Transport using Rényi Divergences*. 2024. arXiv: 2404.18834 [math.OC]. URL: https://arxiv.org/abs/2404.18834.

[3] Marco Cuturi. “Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances”. In: *Advances in Neural Information Processing Systems* 26 (2013).

[4] Boris Muzellec et al. “Tsallis regularized optimal transport and ecological inference”. In: *Proceedings of the AAAI conference on artificial intelligence*. Vol. 31. 2017.

References II

[5] Jun-Kai You and Yen-Huan Li. *Two Polyak-Type Step Sizes for Mirror Descent.* 2022. arXiv: `2210.01532 [math.OC]`.

Outlook—Rényi-Sinkhorn Divergence