Nearly Tight Convergence Bounds for Semi-discrete Entropic Optimal Transport

Alex Delalande

Université Paris-Saclay & Inria Saclay

November 2021

Introduction

Quadratic Optimal Transport with Entropic Regularization

- Let $\mathcal{X}, \mathcal{Y}$ be compact subsets of $\mathbb{R}^d$ and $\rho \in \mathcal{P}(\mathcal{X}), \mu \in \mathcal{P}(\mathcal{Y})$.
- For $\varepsilon \geq 0$, Quadratic OT problem between ρ and μ with entropic regularization:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^2 \, \mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\rho \otimes \mu),} \qquad (\mathrm{P}_\varepsilon)$$

where

$$\mathrm{KL}(\pi|\rho\otimes\mu) = \begin{cases} \int_{\mathcal{X}\times\mathcal{Y}} \left(\log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\rho\otimes\mu}(x,y)\right) - 1\right) \mathrm{d}\pi(x,y) + 1 & \text{if } \pi \ll \rho\otimes\mu, \\ +\infty & \text{else.} \end{cases}$$

Introduction

Quadratic Optimal Transport with Entropic Regularization

- Let $\mathcal{X}, \mathcal{Y}$ be compact subsets of $\mathbb{R}^d$ and $\rho \in \mathcal{P}(\mathcal{X}), \mu \in \mathcal{P}(\mathcal{Y})$.
- For $\varepsilon \geq 0$, Quadratic OT problem between ρ and μ with entropic regularization:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x - y\|^2 \, \mathrm{d}\pi(x, y) + \varepsilon \mathrm{KL}(\pi|\rho \otimes \mu).} \qquad (\mathrm{P}_\varepsilon)$$

- **If $\varepsilon = 0$:** value of (P_ε) defines $\mathrm{W}_2^2(\rho, \mu)$.
  Possibly hard to solve numerically, bad sample complexity.

Introduction

Quadratic Optimal Transport with Entropic Regularization

- Let $\mathcal{X}, \mathcal{Y}$ be compact subsets of $\mathbb{R}^d$ and $\rho \in \mathcal{P}(\mathcal{X}), \mu \in \mathcal{P}(\mathcal{Y})$.
- For $\varepsilon \geq 0$, Quadratic OT problem between ρ and μ with entropic regularization:

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^2 \,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\rho\otimes\mu).} \qquad (\mathrm{P}_\varepsilon)$$

- **If $\varepsilon = 0$:** value of (P_ε) defines $\mathrm{W}_2^2(\rho,\mu)$.
  Possibly hard to compute, bad sample complexity.
- **If $\varepsilon > 0$:**
  - (P_ε) is a ε-strongly-convex min. problem $\implies$ fast optimization algorithms (see e.g. Cuturi (2013); Altschuler et al. (2017); Dvurechensky et al. (2018); Peyré and Cuturi (2019); Schmitzer (2019); Genevay et al. (2016); Bercu and Bigot (2020)).
  - improved sample complexity for the value of (P_ε) (Genevay et al. (2019); Mena and Niles-Weed (2019)).

Introduction

Quadratic Optimal Transport with Entropic Regularization

$$\boxed{\min_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^2 \,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\rho\otimes\mu).} \qquad (\mathrm{P}_\varepsilon)$$

- The case $\varepsilon > 0$ is generally easier to solve than the case $\varepsilon = 0$.
- For $\varepsilon > 0$ and a solution $\pi^{(\mathrm{P}_\varepsilon)}$ to (P_ε), introduce

$$\boxed{\mathrm{W}_{2,\varepsilon}(\rho,\mu) := \left(\int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^2 \,\mathrm{d}\pi^{(\mathrm{P}_\varepsilon)}(x,y)\right)^{1/2}.}$$

Introduction

Quadratic Optimal Transport with Entropic Regularization

$$\boxed{\min_{\pi\in\Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^2 \,\mathrm{d}\pi(x,y) + \varepsilon \mathrm{KL}(\pi|\rho\otimes\mu).} \qquad (\mathrm{P}_\varepsilon)$$

- For $\varepsilon > 0$ and a solution $\pi^{(\mathrm{P}_\varepsilon)}$ to (P_ε), introduce

$$\boxed{\mathrm{W}_{2,\varepsilon}(\rho,\mu) := \left(\int_{\mathcal{X}\times\mathcal{Y}} \|x-y\|^2 \,\mathrm{d}\pi^{(\mathrm{P}_\varepsilon)}(x,y)\right)^{1/2}.}$$

- Hope that:

$$\boxed{\mathrm{W}_{2,\varepsilon}(\rho,\mu) \approx \mathrm{W}_{2,0}(\rho,\mu) = \mathrm{W}_2(\rho,\mu).}$$

How good is this approximation?

Introduction

Quadratic Optimal Transport with Entropic Regularization

- $\boxed{\mathrm{W}_{2,\varepsilon} \xrightarrow[\varepsilon \to 0]{} \mathrm{W}_2}$ is established in general settings (see e.g. Mikami (2004); Léonard (2012); Bernton et al. (2021); Nutz and Wiesel (2021)).

Introduction

Quadratic Optimal Transport with Entropic Regularization

- $\boxed{\mathrm{W}_{2,\varepsilon} \xrightarrow[\varepsilon\to 0]{} \mathrm{W}_2}$ is established in general settings (see e.g. Mikami (2004); Léonard (2012); Bernton et al. (2021); Nutz and Wiesel (2021)).
- **In the continuous setting** (where ρ, μ are a.c.): Adams et al. (2011); Duong et al. (2013); Erbar et al. (2015); Pal (2019); Conforti and Tamanini (2021) gave $1^{st}/2^{nd}$ **order asymptotics**:

  If the densities of ρ, μ are bounded, then

$$\boxed{\begin{aligned}\mathrm{W}_{2,\varepsilon}^2(\rho,\mu) + \varepsilon \mathrm{KL}(\pi^{(\mathrm{P}_\varepsilon)}|\rho\otimes\mu) = \mathrm{W}_2^2(\rho,\mu) - \frac{\varepsilon}{2}\left(\mathrm{KL}(\rho|\lambda)+\mathrm{KL}(\mu|\lambda)\right) \\ - \frac{\varepsilon}{2} d\log(\pi\varepsilon) + \frac{\varepsilon^2}{16} I(\rho,\mu) + o(\varepsilon^2),\end{aligned}}$$

  where
  - *λ is the Lebesgue measure on $\mathbb{R}^d$.*
  - *$I(\rho,\mu)$ is the integrated Fisher information along the 2-Wasserstein geodesic connecting ρ and μ.*

Introduction

Quadratic Optimal Transport with Entropic Regularization

- $\boxed{\mathrm{W}_{2,\varepsilon} \xrightarrow[\varepsilon\to 0]{} \mathrm{W}_2}$ is established in general settings (see e.g. Mikami (2004); Léonard (2012); Bernton et al. (2021); Nutz and Wiesel (2021)).
- **In the continuous setting** (where ρ, μ are a.c.): Adams et al. (2011); Duong et al. (2013); Erbar et al. (2015); Pal (2019); Conforti and Tamanini (2021) gave $1^{st}/2^{nd}$ **order asymptotics**.
- **In the discrete setting** (where ρ, μ are discrete): Cominetti and Martín (1994); Weed (2018) showed an **exponential convergence rate**:

$$\boxed{0 \leq \mathrm{W}_{2,\varepsilon}^2(\rho,\mu) - \mathrm{W}_2^2(\rho,\mu) \leq C_1(\rho,\mu)\exp\left(-\frac{C_2(\rho,\mu)}{\varepsilon}\right),}$$

where $C_1(\rho,\mu), C_2(\rho,\mu)$ are explicit positive constants that depend on ρ and μ.

Introduction

Quadratic Optimal Transport with Entropic Regularization

- $\boxed{W_{2,\varepsilon} \xrightarrow[\varepsilon \to 0]{} W_2}$ is established in general settings (see e.g. Mikami (2004); Léonard (2012); Bernton et al. (2021); Nutz and Wiesel (2021)).
- **In the continuous setting** (where ρ, μ are a.c.): Adams et al. (2011); Duong et al. (2013); Erbar et al. (2015); Pal (2019); Conforti and Tamanini (2021) gave $1^{st}/2^{nd}$ **order asymptotics**.
- **In the discrete setting** (where ρ, μ are discrete): Cominetti and Martín (1994); Weed (2018) showed an **exponential convergence rate**.

What about the intermediate semi-discrete setting (where ρ is a.c. and μ is discrete)?

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

- **Semi-discrete setting:**
  - Let $\mathcal{X} \subset \mathbb{R}^d$ compact and $\rho \in \mathcal{P}_{a.c.}(\mathcal{X})$ absolutely continuous.
  - Let $\mathcal{Y} = \{y_1, \ldots, y_N\} \subset \mathbb{R}^d$ and $\mu = \sum_{i=1}^N \mu_i \delta_{y_i} \in \mathcal{P}(\mathcal{Y})$.
- Note that (P_ε) is equivalent to a regularized "maximum covariance" problem:

$$\boxed{\max_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \langle x|y\rangle \mathrm{d}\pi(x,y) - \varepsilon \mathrm{KL}(\pi|\rho \otimes \sigma),} \qquad (\mathrm{P}'_\varepsilon)$$

where $\sigma = \sum_{i=1}^N \delta_{y_i}$, with the relation:

$$(\mathrm{P}_{2\varepsilon}) = M_2(\rho) + M_2(\mu) - 2\varepsilon\mathcal{H}(\mu) - 2 \times (\mathrm{P}'_\varepsilon).$$

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

- **Semi-discrete setting:**
  - Let $\mathcal{X} \subset \mathbb{R}^d$ compact and $\rho \in \mathcal{P}_{a.c.}(\mathcal{X})$ absolutely continuous.
  - Let $\mathcal{Y} = \{y_1, \ldots, y_N\} \subset \mathbb{R}^d$ and $\mu = \sum_{i=1}^N \mu_i \delta_{y_i} \in \mathcal{P}(\mathcal{Y})$.
- Note that (P_ε) is equivalent to a regularized "maximum covariance" problem:

$$\boxed{\max_{\pi \in \Pi(\rho,\mu)} \int_{\mathcal{X}\times\mathcal{Y}} \langle x|y\rangle \mathrm{d}\pi(x,y) - \varepsilon \mathrm{KL}(\pi|\rho\otimes\sigma).} \qquad (\mathrm{P}'_\varepsilon)$$

- $(\mathrm{P}'_\varepsilon)$ admits a **unique solution** π^ε
  - by ε-strong concavity of $(\mathrm{P}'_\varepsilon)$ if $\varepsilon > 0$.
  - by Brenier (1991) theorem (ρ is a.c.) if $\varepsilon = 0$.

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

- (Semi-)**Dual formulation** with strong duality (Genevay et al. (2016); Bercu and Bigot (2020)):

$$\boxed{\min_{\psi \in \mathbb{R}^N} \int_{\mathcal{X}} \psi^{c,\varepsilon} \mathrm{d}\rho + \langle \psi | \mu \rangle + \varepsilon,} \tag{D_ε}$$

  where $\psi^{c,\varepsilon}$ is the (c,ε)/Legendre transform of ψ:

$$\psi^{c,\varepsilon}(x) = \begin{cases} \varepsilon \log \left(\sum_{i=1}^N e^{\frac{\langle x | y_i \rangle - \psi_i}{\varepsilon}} \right) & \text{if } \varepsilon > 0, \\ \max_{i=1,\dots,N} \langle x | y_i \rangle - \psi_i = \psi^*(x) & \text{if } \varepsilon = 0. \end{cases}$$

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

- (Semi-)**Dual formulation**:

$$\boxed{\min_{\psi \in \mathbb{R}^N} \int_{\mathcal{X}} \psi^{c,\varepsilon} \mathrm{d}\rho + \langle \psi | \mu \rangle + \varepsilon.} \qquad (\mathrm{D}_\varepsilon)$$

- Notice that ψ and $\psi + c\mathbb{1}_N$ yield the same value:

$$\boxed{\min_{\psi \in \mathbb{R}^N, \langle \psi | \mathbb{1}_N \rangle = 0} \int_{\mathcal{X}} \psi^{c,\varepsilon} \mathrm{d}\rho + \langle \psi | \mu \rangle + \varepsilon.} \qquad (\mathrm{D}_\varepsilon)$$

- By strict convexity, (D_ε) admits a **unique solution** $\psi^\varepsilon \in (\mathbb{1}_N)^\perp$.

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

$$\min_{\psi \in \mathbb{R}^N, \langle \psi | \mathbb{1}_N \rangle = 0} \int_{\mathcal{X}} \psi^{c,\varepsilon} \mathrm{d}\rho + \langle \psi | \mu \rangle + \varepsilon. \qquad (\mathrm{D}_\varepsilon)$$

- **First order condition:**

$$\mu(\{y_i\}) = \begin{cases} \int_{x \in \mathcal{X}} e^{\frac{\langle x | y_i \rangle - \psi_i^\varepsilon - (\psi^\varepsilon)^{c,\varepsilon}(x)}{\varepsilon}} \mathrm{d}\rho(x) & \text{if } \varepsilon > 0, \\ \int_{x \in \mathcal{X}} \mathbb{1}_{\mathrm{Lag}_i(\psi^0)}(x) \mathrm{d}\rho(x) = \rho(\mathrm{Lag}_i(\psi^0)) & \text{if } \varepsilon = 0, \end{cases}$$

where $\mathrm{Lag}_i(\psi) = \{x \in \mathcal{X} | \forall j, \langle x | y_i \rangle - \psi_i \geq \langle x | y_j \rangle - \psi_j\}$.

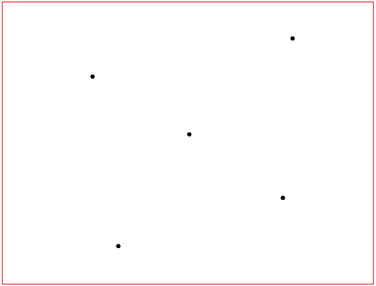
- **Primal-dual relationship:** $\forall A \subset \mathcal{X}, i \in \{1, \ldots, N\}$,

$$\pi^\varepsilon(A, \{y_i\}) = \begin{cases} \int_{x \in A} e^{\frac{\langle x | y_i \rangle - \psi_i^\varepsilon - (\psi^\varepsilon)^{c,\varepsilon}(x)}{\varepsilon}} \mathrm{d}\rho(x) & \text{if } \varepsilon > 0, \\ \int_{x \in A} \mathbb{1}_{\mathrm{Lag}_i(\psi^0)}(x) \mathrm{d}\rho(x) = \rho(\mathrm{Lag}_i(\psi^0) \cap A) & \text{if } \varepsilon = 0. \end{cases}$$

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

- Laguerre cells: a 2-dimensional example



$\rho = \mathbb{1}_{[a,b]\times[c,d]}$ and $\mu = \sum_{i=1}^{N} \mu_i \delta_{y_i}$.

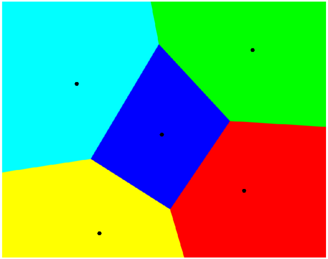
Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

$$\mathrm{Lag}_i(\psi^0) = \{x \in \mathcal{X} | \forall j, \langle x|y_i\rangle - \psi_i^0 \geq \langle x|y_j\rangle - \psi_j^0\}.$$

(Figures inspired from Peyré and Cuturi (2019))

$\varepsilon = 0$:



Each color corresponds to one of $\left(\mathrm{Lag}_i(\psi^0)\right)_{i=1,\ldots,N}$.

- First-order condition: $\mu(\{y_i\}) = \rho(\mathrm{Lag}_i(\psi^0))$.
- Primal-dual relation: $\forall A \subset \mathcal{X}, \pi^0(A, \{y_i\}) = \rho(\mathrm{Lag}_i(\psi^0) \cap A)$

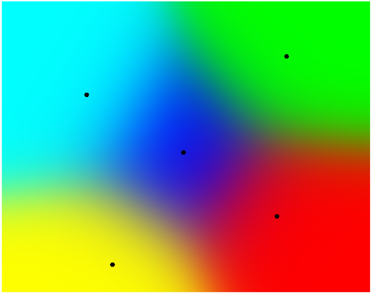
$$\Longrightarrow \boxed{\mathrm{Lag}_i(\psi^0) = \left(T_{\rho\to\mu}\right)^{-1}(\{y_i\}).}$$

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

$$\pi^{\varepsilon}_{x,i} := \frac{e^{(\langle x|y_i\rangle - \psi^{\varepsilon}_i)/\varepsilon}}{\sum_j e^{(\langle x|y_j\rangle - \psi^{\varepsilon}_j)/\varepsilon}}.$$

$\varepsilon = 7.5 \times 10^{-2}$:



Each color represents one of $\left(x \mapsto \pi^{\varepsilon}_{x,i}\right)_{i=1,\ldots,N}$, level of transparency = value.

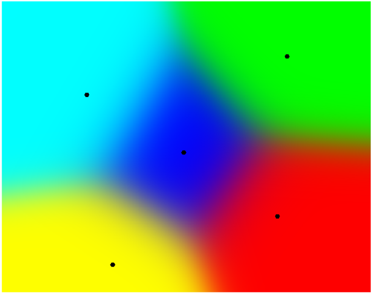
- First-order condition: $\mu(\{y_i\}) = \int_{x\in\mathcal{X}} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.
- Primal-dual relation: $\forall A \subset \mathcal{X}, \pi^{\varepsilon}(A, \{y_i\}) = \int_{x\in A} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

$$\pi^{\varepsilon}_{x,i} := \frac{e^{(\langle x|y_i\rangle - \psi^{\varepsilon}_i)/\varepsilon}}{\sum_j e^{(\langle x|y_j\rangle - \psi^{\varepsilon}_j)/\varepsilon}}.$$

$\varepsilon = 5 \times 10^{-2}$:



Each color represents one of $\left(x \mapsto \pi^{\varepsilon}_{x,i}\right)_{i=1,\dots,N}$, level of transparency = value.

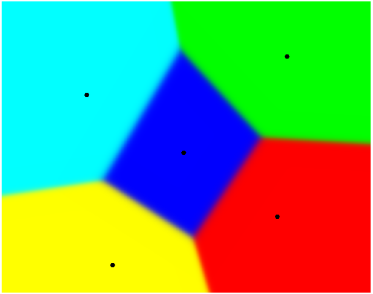
- First-order condition: $\mu(\{y_i\}) = \int_{x\in\mathcal{X}} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.
- Primal-dual relation: $\forall A \subset \mathcal{X}, \pi^{\varepsilon}(A, \{y_i\}) = \int_{x\in A} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

$$\pi^{\varepsilon}_{x,i} := \frac{e^{(\langle x|y_i\rangle - \psi^{\varepsilon}_i)/\varepsilon}}{\sum_j e^{(\langle x|y_j\rangle - \psi^{\varepsilon}_j)/\varepsilon}}.$$

$\varepsilon = 7.5 \times 10^{-3}$:



Each color represents one of $\left(x \mapsto \pi^{\varepsilon}_{x,i}\right)_{i=1,\ldots,N}$, level of transparency = value.

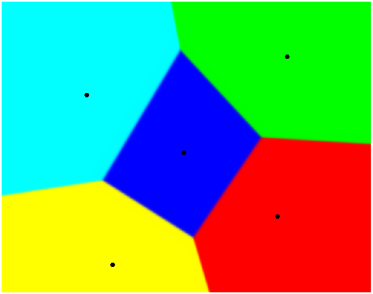
- First-order condition: $\mu(\{y_i\}) = \int_{x\in\mathcal{X}} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.
- Primal-dual relation: $\forall A \subset \mathcal{X}, \pi^{\varepsilon}(A, \{y_i\}) = \int_{x\in A} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.

Convergence Bounds for Entropic SDOT

Entropic Semi-discrete Optimal Transport

$$\pi^{\varepsilon}_{x,i} := \frac{e^{(\langle x|y_i\rangle - \psi^{\varepsilon}_i)/\varepsilon}}{\sum_j e^{(\langle x|y_j\rangle - \psi^{\varepsilon}_j)/\varepsilon}}.$$

$\varepsilon = 2.5 \times 10^{-3}$:



Each color represents one of $\left(x \mapsto \pi^{\varepsilon}_{x,i}\right)_{i=1,\dots,N}$, level of transparency = value.

- First-order condition: $\mu(\{y_i\}) = \int_{x\in\mathcal{X}} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.
- Primal-dual relation: $\forall A \subset \mathcal{X}, \pi^{\varepsilon}(A, \{y_i\}) = \int_{x\in A} \pi^{\varepsilon}_{x,i} \mathrm{d}\rho(x)$.

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

- Recent result from Altschuler, Niles-Weed and Stromme (2021) (Theorem 1.1): *under regularity assumptions on ρ,*

$$\mathrm{W}^2_{2,\varepsilon}(\rho,\mu) = \mathrm{W}^2_2(\rho,\mu) + \varepsilon^2 \frac{\pi^2}{12} \sum_{i<j} \frac{w_{ij}}{\|y_i - y_j\|} + o(\varepsilon^2),$$

where $w_{ij} = \int_{\mathrm{Lag}_i(\psi^0)\cap \mathrm{Lag}_j(\psi^0)} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x)$.

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

- Recent result from Altschuler, Niles-Weed and Stromme (2021) (Theorem 1.1): *under regularity assumptions on* ρ,

$$\mathrm{W}^2_{2,\varepsilon}(\rho,\mu) = \mathrm{W}^2_2(\rho,\mu) + \varepsilon^2 \frac{\pi^2}{12} \sum_{i<j} \frac{w_{ij}}{\|y_i - y_j\|} + o(\varepsilon^2),$$

where $w_{ij} = \int_{\mathrm{Lag}_i(\psi^0)\cap \mathrm{Lag}_j(\psi^0)} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x)$.

- Showed using (Theorem 1.3 in Altschuler et al. (2021)):

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(\psi^\varepsilon - \psi^0) = \dot{\psi^\varepsilon}\Big|_{\varepsilon=0} = 0,$$

where $\dot{\psi^\varepsilon} = \frac{\partial}{\partial\varepsilon}\psi^\varepsilon$.

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

- Recent result from Altschuler, Niles-Weed and Stromme (2021) (Theorem 1.1): *under regularity assumptions on ρ,*

$$\boxed{\mathrm{W}_{2,\varepsilon}^2(\rho,\mu) = \mathrm{W}_2^2(\rho,\mu) + \varepsilon^2 \frac{\pi^2}{12} \sum_{i<j} \frac{w_{ij}}{\|y_i - y_j\|} + o(\varepsilon^2),}$$

  where $w_{ij} = \int_{\mathrm{Lag}_i(\psi^0)\cap \mathrm{Lag}_j(\psi^0)} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x)$.

- Showed using (Theorem 1.3 in Altschuler et al. (2021)):

$$\boxed{\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(\psi^\varepsilon - \psi^0) = \dot{\psi}^\varepsilon\Big|_{\varepsilon=0} = 0.}$$

This result can be extended and quantified with a non-asymptotic analysis.

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

- **Assumption:** *The compact set* $\mathcal{X}$ *is convex. The source density* ρ *is* α*-Hölder continuous for some* $\alpha \in (0, 1]$ *and verifies on* $\mathcal{X}$*:*

$$0 < m_\rho \leq \rho \leq M_\rho < +\infty.$$

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

- **Assumption:** *The compact set $\mathcal{X}$ is convex. The source density ρ is α-Hölder continuous for some $\alpha \in (0,1]$ and verifies on $\mathcal{X}$:*

$$0 < m_\rho \leq \rho \leq M_\rho < +\infty.$$

Theorem (D., 2021): *The mapping $\varepsilon \mapsto \psi^\varepsilon$ from $\mathbb{R}_+^*$ to $(\mathbb{1}_N)^\perp$ is $\mathcal{C}^1$. For any $\varepsilon > 0, \alpha' \in (0, \alpha)$,*

$$\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1),$$

where $\dot{\psi}^\varepsilon = \frac{\partial}{\partial \varepsilon}\psi^\varepsilon$ and $\lesssim$ hides multiplicative constants that depend on $\mathcal{X}, \rho, \mathcal{Y}, \mu$.

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

Theorem (D., 2021): *The mapping $\varepsilon \mapsto \psi^\varepsilon$ from $\mathbb{R}_+^*$ to $(\mathbb{1}_N)^\perp$ is $\mathcal{C}^1$. For any $\varepsilon > 0, \alpha' \in (0, \alpha)$,*

$$\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1),$$

where $\dot{\psi}^\varepsilon = \frac{\partial}{\partial \varepsilon}\psi^\varepsilon$ and $\lesssim$ hides multiplicative constants that depend on $\mathcal{X}, \rho, \mathcal{Y}, \mu$.

- **Remark:** A rough upperbound on the constant is

$$C(d) \times \frac{N}{\underline{\mu}} \frac{M_\rho}{m_\rho} e^{R_\mathcal{Y} \operatorname{diam}(\mathcal{X})} \Bigg(N R_\mathcal{X} \operatorname{diam}(\mathcal{Y}) + \log \frac{1}{\underline{\mu}} \\ + N^2 M_\rho \operatorname{diam}(\mathcal{X})^{d-1} (1 + \frac{C_\rho}{\delta^\alpha} + R_\mathcal{X} \operatorname{diam}(\mathcal{Y}) + \log \frac{1}{\underline{\mu}}) \\ + N^3 M_\rho \frac{\operatorname{diam}(\mathcal{X})^{d-2} \operatorname{diam}(\mathcal{Y})^4}{\cos(\theta/2)\delta^4} (1 + R_\mathcal{X} \operatorname{diam}(\mathcal{Y}) + \log \frac{1}{\underline{\mu}}) \Bigg),$$

where $\mathcal{X} \subset B(0, R_\mathcal{X}), \quad Y \subset B(0, R_\mathcal{Y}), \quad \underline{\mu} = \min_i \mu_i > 0, \quad \delta = \min_{i \neq j} \left\| y_i - y_j \right\| > 0,$

$\theta = \max_{i,j,k} \{\angle y_i y_j y_k \,|\, \angle y_i y_j y_k < \pi\}, \quad \forall x, x' \in \mathcal{X}, \left|\rho(x) - \rho(x')\right| \leq C_\rho \left\|x - x'\right\|^\alpha.$

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

Corollary (D., 2021): *Let* $0 < \varepsilon' \leq \varepsilon$. *For any* $\alpha' \in (0, \alpha)$,

$$\left\| \psi^{\varepsilon} - \psi^{\varepsilon'} \right\|_{\infty} \lesssim \varepsilon^{\alpha'} (\varepsilon - \varepsilon').$$

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

Corollary (D., 2021): *Let* $0 < \varepsilon' \leq \varepsilon$. *For any* $\alpha' \in (0, \alpha)$,

$$\left\| \psi^{\varepsilon} - \psi^{\varepsilon'} \right\|_{\infty} \lesssim \varepsilon^{\alpha'} (\varepsilon - \varepsilon').$$

- **Remark:** May "justify" ε-scaling heuristic, where ε is decreased over the iterations of an algorithm that estimates ψ^0.

Convergence Bounds for Entropic SDOT

Non-asymptotic Behavior of Potentials

Corollary (D., 2021): *Let $0 < \varepsilon' \leq \varepsilon$. For any $\alpha' \in (0, \alpha)$,*

$$\left\|\psi^\varepsilon - \psi^{\varepsilon'}\right\|_\infty \lesssim \varepsilon^{\alpha'}(\varepsilon - \varepsilon').$$

In particular, letting ε' go to 0 yields

$$\left\|\psi^\varepsilon - \psi^0\right\|_\infty \lesssim \varepsilon^{1+\alpha'}.$$

Additionally, for ρ-a.e. $x \in \mathcal{X}$,

$$\left|\pi^\varepsilon(x, \cdot) - \pi^0(x, \cdot)\right| \lesssim e^{-c_x/\varepsilon},$$

where $c_x = \min_{i \in \{1,\ldots,N\}} \{(\psi^0)^*(x) - \langle x|y_i\rangle + \psi_i^0 \quad | \quad \langle x|y_i\rangle - \psi_i^0 \neq (\psi^0)^*(x)\} > 0$.

Convergence Bounds for Entropic SDOT

Non-asymptotic Expansion of the Difference of Costs

Theorem (D., 2021): *For any $\alpha' \in (0, \alpha)$ and $\varepsilon > 0$,*

$$\left| \mathrm{W}_{2,\varepsilon}^2(\rho, \mu) - \mathrm{W}_2^2(\rho, \mu) - \varepsilon^2 \frac{\pi^2}{12} \sum_{i<j} \frac{w_{ij}}{\|y_i - y_j\|} \right| \lesssim \varepsilon^{2+\alpha'}.$$

This inequality is tight.

- **Remark:** No third-order expansion.

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

A governing O.D.E.

- Dual formulation:

$$\min_{\psi\in\mathbb{R}^N, \langle\psi|\mathbb{1}_N\rangle=0} \int_{\mathcal{X}} \psi^{c,\varepsilon}\mathrm{d}\rho + \langle\psi|\mu\rangle + \varepsilon. \tag{D_ε}$$

- Regularized Kantorovich's functional: $\forall\psi\in\mathbb{R}^N$,

$$\begin{aligned}\mathcal{K}^\varepsilon(\psi) &= \int_{\mathcal{X}} \psi^{c,\varepsilon}\mathrm{d}\rho + \varepsilon \\ &= \int_{\mathcal{X}} \varepsilon \log\left(\sum_{i=1}^N \exp\left(\frac{\langle x|y_i\rangle - \psi_i}{\varepsilon}\right)\right)\mathrm{d}\rho(x) + \varepsilon.\end{aligned}$$

- $\mathcal{K}^\varepsilon$ strictly convex on $(\mathbb{1}_N)^\perp$. First-order condition for (D_ε):

$$\nabla\mathcal{K}^\varepsilon(\psi^\varepsilon) = -\mu.$$

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

A governing O.D.E.

- Regularized Kantorovich's functional: $\forall \psi \in \mathbb{R}^N$,

$$\mathcal{K}^\varepsilon(\psi) = \int_{\mathcal{X}} \varepsilon \log\left(\sum_{i=1}^{N} \exp\left(\frac{\langle x|y_i\rangle - \psi_i}{\varepsilon}\right)\right) \mathrm{d}\rho(x) + \varepsilon.$$

- $\mathcal{K}^\varepsilon$ strictly convex on $(\mathbb{1}_N)^\perp$. First-order condition for (D_ε):

$$\nabla \mathcal{K}^\varepsilon(\psi^\varepsilon) = -\mu.$$

- **Implicit function theorem:**
  - $\psi \mapsto \mathcal{K}^\varepsilon(\psi)$ is a $\mathcal{C}^2$ strictly convex mapping from $(\mathbb{1}_N)^\perp$ to $\mathbb{R}$.
  - $\varepsilon \mapsto \nabla \mathcal{K}^\varepsilon(\psi)$ is a $\mathcal{C}^1$ mapping from $\mathbb{R}_+^*$ to $\mathbb{R}$.

$\implies$ $\varepsilon \mapsto \psi^\varepsilon$ is a $\mathcal{C}^1$ mapping from $\mathbb{R}_+^*$ to $(\mathbb{1}_N)^\perp$ and

$$\nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon)\dot{\psi}^\varepsilon + \frac{\partial}{\partial \varepsilon}(\nabla \mathcal{K}^\varepsilon)(\psi^\varepsilon) = 0.$$

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

A governing O.D.E.

- The potential ψ^ε satisfies

$$\boxed{\nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon)\dot{\psi}^\varepsilon + \frac{\partial}{\partial \varepsilon}(\nabla \mathcal{K}^\varepsilon)(\psi^\varepsilon) = 0.}$$

- $\implies$ An upper bound on $\left\|\dot{\psi}^\varepsilon\right\|$ may be obtained from:
  - A lower bound on $\nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) \to$ **Prékopa-Leindler inequality**.
  - An upper bound on $\frac{\partial}{\partial \varepsilon}(\nabla \mathcal{K}^\varepsilon)(\psi^\varepsilon) \to$ **"Laplace's method"**.

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any* $\varepsilon > 0$ *and* $v \in \mathbb{R}^N$,

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any $\varepsilon > 0$ and $v \in \mathbb{R}^N$,*

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_\mathcal{Y} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- **Remark:** this result is mainly deduced from the Prékopa-Leindler inequality:
  Let $0 < \lambda < 1$ and $f, g, h : \mathbb{R}^N \to \mathbb{R}_+$. Assume that $\forall x, y \in \mathbb{R}^N$,

  $$h((1-\lambda)x + \lambda y) \geq f(x)^{1-\lambda} g(y)^\lambda.$$

  Then

  $$\|h\|_1 \geq \|f\|_1^{1-\lambda} \|g\|_1^\lambda .$$

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any $\varepsilon > 0$ and $v \in \mathbb{R}^N$,*

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- **Remark:** this result is deduced mainly from the Prékopa-Leindler inequality.
- **Remark:** as $\varepsilon \to 0$, recover a similar estimate in the unregularized case that was proved using the Brascamp-Lieb inequality (D. and Mérigot, 2021):

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} \right) \langle v | \nabla^2 \mathcal{K}^0(\psi^0) v \rangle.$$

Sketch of proof for $\left\|\dot{\psi^\varepsilon}\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

Bound on the second term $\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)$

Theorem (D., 2021): *For any $\varepsilon > 0$,*

$$\left\|\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)\right\|_\infty \lesssim \min\left(\varepsilon^{\alpha'}, \frac{1}{\varepsilon}\right).$$

Sketch of proof for $\left\|\dot{\psi}^\varepsilon\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$

Bound on the second term $\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)$

Theorem (D., 2021): *For any $\varepsilon > 0$,*

$$\left\|\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)\right\|_\infty \lesssim \min\left(\varepsilon^{\alpha'}, \frac{1}{\varepsilon}\right).$$

- **Proof idea:** We have

$$[\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)]_i = \int_{\mathcal{X}} \sum_{j\neq i} \left(\frac{f_i^\varepsilon(x) - f_j^\varepsilon(x)}{\varepsilon^2}\right) \pi^\varepsilon_{x,j}\pi^\varepsilon_{x,i}\mathrm{d}\rho(x),$$

where $\forall j,\ f_j^\varepsilon(x) = \langle x|y_j\rangle - \psi_j^\varepsilon$ and $\pi^\varepsilon_{x,j} = \frac{\exp(\frac{f_j^\varepsilon(x)}{\varepsilon})}{\sum_k \exp(\frac{f_k^\varepsilon(x)}{\varepsilon})}$.

$\implies$ Control $\left|f_i^\varepsilon(x) - f_j^\varepsilon(x)\right|$, $\pi^\varepsilon_{x,i}$, $\pi^\varepsilon_{x,j}$ depending on the position of x in $\mathcal{X} = \bigcup_i \mathrm{Lag}_i(\psi^\varepsilon)$.

Numerical illustrations

Behavior of $\varepsilon \mapsto \psi^\varepsilon$

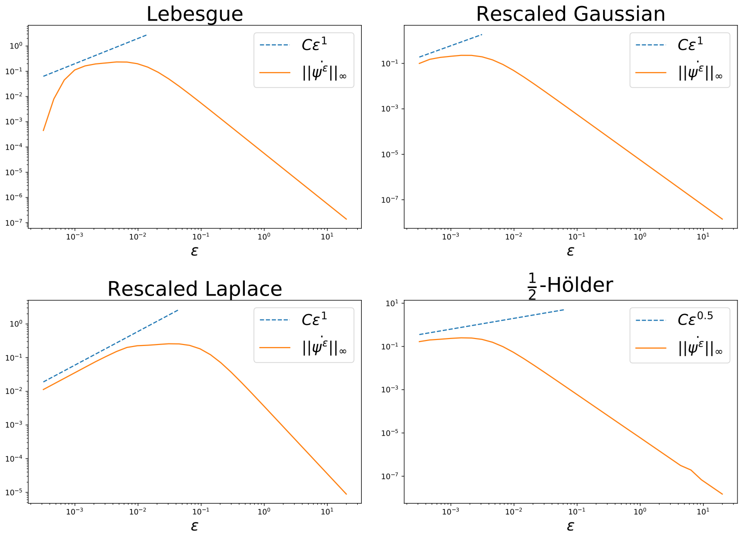
- Let $\mathcal{X} = [-1, 1]$, ρ symmetric on $\mathcal{X}$ and $\mu = \frac{1}{5} \sum_{i=1}^{5} \delta_{y_i}$, where $\{y_1, \dots, y_5\} \subset \mathcal{X}$.

- Consider 4 different sources:
  - Lebesgue: $\rho(x) \propto \mathbb{1}_{[-1,1]}(x)$,
  - Rescaled Gaussian: $\rho(x) \propto e^{-x^2/2\sigma^2} \mathbb{1}_{[-1,1]}(x)$,
  - Rescaled Laplace: $\rho(x) \propto e^{-|x|} \mathbb{1}_{[-1,1]}(x)$,
  - $\frac{1}{2}$-Hölder density: $\rho(x) \propto (1 - |x|^{1/2}) \mathbb{1}_{[-1,1]}(x)$.

Numerical illustrations

Behavior of $\varepsilon \mapsto \psi^\varepsilon$

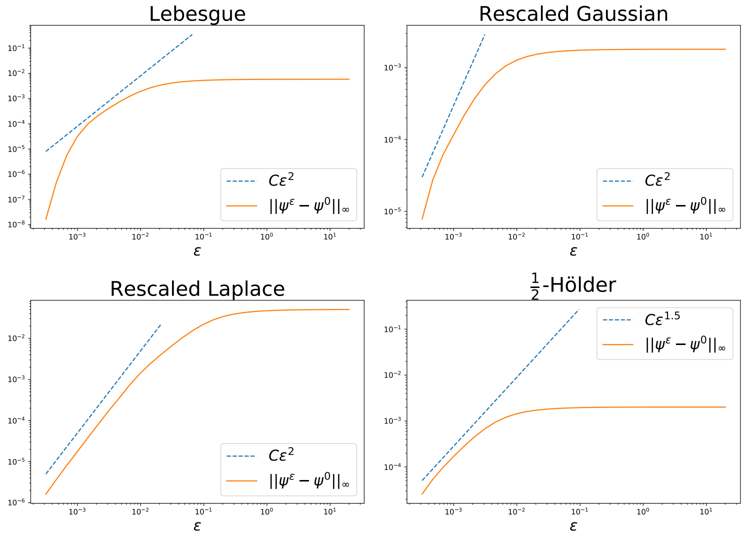
- Observe $\left\|\dot{\psi^\varepsilon}\right\|_2 \lesssim \min(\varepsilon^{\alpha'}, 1)$:



Numerical illustrations

Behavior of $\varepsilon \mapsto \psi^\varepsilon$

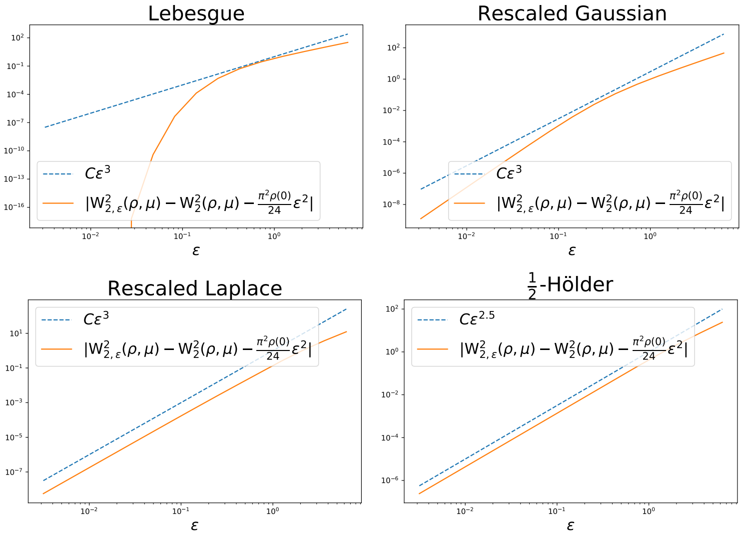
- Observe $\left\|\psi^\varepsilon - \psi^0\right\|_2 \lesssim \varepsilon^{1+\alpha'}$:



Numerical illustrations

Difference of Costs

- Observe $\left|W^2_{2,\varepsilon}(\rho,\mu) - W^2_2(\rho,\mu) - \varepsilon^2 \frac{\pi^2}{12} \sum_{i<j} \frac{w_{ij}}{\|y_i - y_j\|}\right| \lesssim \varepsilon^{2+\alpha'}$:



Thank you for your attention!

References I

Adams, S., Dirr, N., Peletier, M. A., and Zimmer, J. (2011). From a Large-Deviations Principle to the Wasserstein Gradient Flow: A New Micro-Macro Passage. *Communications in Mathematical Physics*, 307(3):791.

Altschuler, J., Niles-Weed, J., and Stromme, A. (2021). Asymptotics for semi-discrete entropic optimal transport.

Altschuler, J., Weed, J., and Rigollet, P. (2017). Near-linear time approximation algorithms for optimal transport via sinkhorn iteration. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., editors, *Advances in Neural Information Processing Systems 30*, pages 1964–1974. Curran Associates, Inc.

Bercu, B. and Bigot, J. (2020). Asymptotic distribution and convergence rates of stochastic algorithms for entropic optimal transportation between probability measures. *Annals of Statistics*.

Bernton, E., Ghosal, P., and Nutz, M. (2021). Entropic optimal transport: Geometry and large deviations.

References II

Bertsekas, D. P. and Eckstein, J. (1988). Dual coordinate step methods for linear network flow problems. *Mathematical Programming*, 42(1):203–243.

Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Communications on Pure and Applied Mathematics*, 44(4):375–417.

Cominetti, R. and Martín, J. S. (1994). Asymptotic analysis of the exponential penalty trajectory in linear programming. *Mathematical Programming*, 67(1):169–187.

Conforti, G. and Tamanini, L. (2021). A formula for the time derivative of the entropic cost and applications. *Journal of Functional Analysis*, 280(11):108964.

Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2*, NIPS'13, pages 2292–2300.

References III

Duong, M. H., Laschos, V., and Renger, M. (2013). Wasserstein gradient flows from large deviations of many-particle limits. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(4):1166–1188.

Dvurechensky, P., Gasnikov, A., and Kroshnin, A. (2018). Computational optimal transport: Complexity by accelerated gradient descent is better than by sinkhorn's algorithm. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 1367–1376, Stockholmsmässan, Stockholm Sweden. PMLR.

Erbar, M., Maas, J., and Renger, M. (2015). From large deviations to Wasserstein gradient flows in multiple dimensions. *Electronic Communications in Probability*, 20(none):1 – 12.

Feydy, J. (2020). *Analyse de données géométriques, au delà des convolutions*. Theses, Université Paris-Saclay.

References IV

Genevay, A., Chizat, L., Bach, F., Cuturi, M., and Peyré, G. (2019). Sample complexity of sinkhorn divergences. In Chaudhuri, K. and Sugiyama, M., editors, *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 1574–1583. PMLR.

Genevay, A., Cuturi, M., Peyré, G., and Bach, F. (2016). Stochastic optimization for large-scale optimal transport. In *Advances in neural information processing systems*, pages 3440–3448.

Kosowsky, J. and Yuille, A. (1994). The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks*, 7(3):477–490.

Léonard, C. (2012). From the schrödinger problem to the monge-kantorovich problem. *Journal of Functional Analysis*, 262.

Mena, G. and Niles-Weed, J. (2019). Statistical bounds for entropic optimal transport: sample complexity and the central limit theorem. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

References V

Mikami, T. (2004). Monge's problem with a quadratic cost by the zero-noise limit of h-path processes. *Probability Theory and Related Fields*, 129(2):245–260.

Nutz, M. and Wiesel, J. (2021). Entropic optimal transport: Convergence of potentials.

Pal, S. (2019). On the difference between entropic cost and the optimal transport cost. *arXiv: Probability*.

Peyré, G. and Cuturi, M. (2019). Computational optimal transport. *Foundations and Trends in Machine Learning*, 11(5-6):355–607.

Schmitzer, B. (2019). Stabilized sparse scaling algorithms for entropy regularized transport problems. *SIAM J. Sci. Comput.*, 41(3):A1443–A1481.

Weed, J. (2018). An explicit analysis of the entropic penalty in linear programming. In Bubeck, S., Perchet, V., and Rigollet, P., editors, *Conference On Learning Theory, COLT 2018, Stockholm, Sweden, 6-9 July 2018*, volume 75 of *Proceedings of Machine Learning Research*, pages 1841–1855. PMLR.

ε-scaling

Corollary (D., 2021): *Let* $0 < \varepsilon' \leq \varepsilon$. *For any* $\alpha' \in (0, \alpha)$,

$$\left\| \psi^{\varepsilon} - \psi^{\varepsilon'} \right\|_{\infty} \lesssim \varepsilon^{\alpha'} (\varepsilon - \varepsilon').$$

- **Remark:** (ε-scaling) Assume we know an algorithm `approx_psi` such that

$$\widetilde{\psi^{\varepsilon}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon, \psi^{init}) \approx \psi^{\varepsilon}.$$

Fast if ε big or $\left\| \psi^{\varepsilon} - \psi^{init} \right\|$ small.

ε-scaling

Corollary (D., 2021): *Let* $0 < \varepsilon' \leq \varepsilon$*. For any* $\alpha' \in (0, \alpha)$*,*

$$\left\| \psi^{\varepsilon} - \psi^{\varepsilon'} \right\|_{\infty} \lesssim \varepsilon^{\alpha'} (\varepsilon - \varepsilon').$$

- **Remark:** (ε-scaling) Assume we know an algorithm `approx_psi` such that

  $$\widetilde{\psi^{\varepsilon}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon, \psi^{init}) \approx \psi^{\varepsilon}.$$

  Fast if ε big or $\left\| \psi^{\varepsilon} - \psi^{init} \right\|$ small.

  ε**-scaling approximates** ψ^0 **with** $\widetilde{\psi^{\varepsilon_K}}$ **for some** $K \in \mathbb{N}^*$, where:

  - $\varepsilon_0 > 0$ (big), $\widetilde{\psi^{\varepsilon_0}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon_0, \psi^{init})$,
  - $\varepsilon_{k+1} = \varepsilon_k / 2$,
  - $\widetilde{\psi^{\varepsilon_{k+1}}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon_{k+1}, \widetilde{\psi^{\varepsilon_k}})$.

ε-scaling

Corollary (D., 2021): *Let $0 < \varepsilon' \leq \varepsilon$. For any $\alpha' \in (0, \alpha)$,*

$$\left\| \psi^{\varepsilon} - \psi^{\varepsilon'} \right\|_{\infty} \lesssim \varepsilon^{\alpha'}(\varepsilon - \varepsilon').$$

- **Remark:** (ε-scaling) Assume we know an algorithm `approx_psi` such that

$$\widetilde{\psi^{\varepsilon}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon, \psi^{init}) \approx \psi^{\varepsilon}.$$

  Fast if ε big or $\left\| \psi^{\varepsilon} - \psi^{init} \right\|$ small.

  ε-scaling approximates ψ^0 with $\widetilde{\psi^{\varepsilon_K}}$ for some $K \in \mathbb{N}^*$, where:

  - $\varepsilon_0 > 0$ (big), $\widetilde{\psi^{\varepsilon_0}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon_0, \psi^{init})$,
  - $\varepsilon_{k+1} = \varepsilon_k / 2$,
  - $\widetilde{\psi^{\varepsilon_{k+1}}} := \texttt{approx\_psi}(\rho, \mu, \varepsilon_{k+1}, \widetilde{\psi^{\varepsilon_k}})$.

  Hope that $\left\| \widetilde{\psi^{\varepsilon_k}} - \psi^{\varepsilon_{k+1}} \right\|$ gets small as $\varepsilon_k \to 0$. See for instance Kosowsky and Yuille (1994); Schmitzer (2019); Feydy (2020).

ε-scaling

Corollary (D., 2021): *Let* $0 < \varepsilon' \leq \varepsilon$. *For any* $\alpha' \in (0, \alpha)$,

$$\left\| \psi^{\varepsilon} - \psi^{\varepsilon'} \right\|_{\infty} \lesssim \varepsilon^{\alpha'}(\varepsilon - \varepsilon').$$

- **Remark:** (ε-scaling) Originally introduced for Bertsekas' auction algorithm (Bertsekas and Eckstein (1988)) for the N-assignment problem. Reduced worst case complexity:

$$O\left(\frac{N^2}{\varepsilon}\right) \to O\left(N^3 \log\left(\frac{1}{\varepsilon}\right)\right)$$

in order to get an ε-approximate solution.

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any* $\varepsilon > 0$ *and* $v \in \mathbb{R}^N$,

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_\mathcal{Y} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- Notice that

$$\nabla^2 \mathcal{K}^\varepsilon(\psi) = \frac{1}{\varepsilon} \mathbb{E}_{x\sim\rho} \left(\mathrm{diag}(\pi_x^\varepsilon(\psi)) - \pi_x^\varepsilon(\psi) \pi_x^\varepsilon(\psi)^\top \right),$$

where $\pi_x^\varepsilon(\psi) \in \mathbb{R}^N$ and $\forall i \in \{1, \ldots, N\}$,

$$\pi_x^\varepsilon(\psi)_i = \frac{\exp\left(\frac{\langle x | y_i \rangle - \psi_i}{\varepsilon} \right)}{\sum_{j=1}^N \exp\left(\frac{\langle x | y_j \rangle - \psi_j}{\varepsilon} \right)}.$$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any $\varepsilon > 0$ and $v \in \mathbb{R}^N$,*

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- Notice that

$$\nabla^2 \mathcal{K}^\varepsilon(\psi) = \frac{1}{\varepsilon} \mathbb{E}_{x\sim\rho} \left(\mathrm{diag}(\pi_x^\varepsilon(\psi)) - \pi_x^\varepsilon(\psi) \pi_x^\varepsilon(\psi)^\top \right).$$

- Introduce $I : \mathbb{R}^N \to \mathbb{R}, \psi \mapsto \log(\int_{\mathcal{X}} e^{-\psi^{c,\varepsilon}})$. Notice that I is $\mathcal{C}^2$:

$$\begin{aligned}\nabla^2 I(\psi^\varepsilon) = & -\frac{1}{\varepsilon} \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} (\mathrm{diag}(\pi_x^\varepsilon(\psi^\varepsilon)) - \pi_x^\varepsilon(\psi^\varepsilon) \pi_x^\varepsilon(\psi^\varepsilon)^\top) \\ & + \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} \pi_x^\varepsilon(\psi^\varepsilon) \pi_x^\varepsilon(\psi^\varepsilon)^\top - \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} \pi_x^\varepsilon(\psi^\varepsilon) \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} \pi_x^\varepsilon(\psi^\varepsilon)^\top,\end{aligned}$$

where $\tilde{\rho}^\varepsilon := \frac{e^{-(\psi^\varepsilon)^{c,\varepsilon}}}{\int_{\mathcal{X}} e^{-(\psi^\varepsilon)^{c,\varepsilon}}}$.

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any $\varepsilon > 0$ and $v \in \mathbb{R}^N$,*

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- $I : \psi \mapsto \log(\int_{\mathcal{X}} e^{-\psi^{c,\varepsilon}})$ is concave: Let $\psi, \varphi \in \mathbb{R}^N, 0 < \lambda < 1$. For any $u, v \in \mathcal{X}$, we have:

$$\begin{aligned}
&\big(\lambda\psi + (1-\lambda)\varphi\big)^{c,\varepsilon}(\lambda u + (1-\lambda)v) \\
&= \varepsilon \log \left(\sum_{i=1}^N e^{\frac{\langle \lambda u + (1-\lambda)v | y_i \rangle - (\lambda\psi + (1-\lambda)\varphi)(y_i)}{\varepsilon}} \right) \\
&= \varepsilon \log \left(\sum_{i=1}^N \left(e^{\frac{\langle u | y_i \rangle - \psi(y_i)}{\varepsilon}} \right)^\lambda \left(e^{\frac{\langle v | y_i \rangle - \varphi(y_i)}{\varepsilon}} \right)^{1-\lambda} \right) \\
&\leq \varepsilon \log \left[\left(\sum_{i=1}^N e^{\frac{\langle u | y_i \rangle - \psi(y_i)}{\varepsilon}} \right)^\lambda \left(\sum_{i=1}^N e^{\frac{\langle v | y_i \rangle - \varphi(y_i)}{\varepsilon}} \right)^{1-\lambda} \right] \\
&= \lambda\psi^{c,\varepsilon}(u) + (1-\lambda)\varphi^{c,\varepsilon}(v).
\end{aligned}$$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any $\varepsilon > 0$ and $v \in \mathbb{R}^N$,*

$$\mathbb{V}\mathrm{ar}_\mu(v) \le \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- $I : \psi \mapsto \log(\int_{\mathcal{X}} e^{-\psi^{c,\varepsilon}})$ is concave: Let $\psi, \varphi \in \mathbb{R}^N, 0 < \lambda < 1$. For any $u, v \in \mathcal{X}$, we have:

  $$\big(\lambda\psi + (1-\lambda)\varphi\big)^{c,\varepsilon}(\lambda u + (1-\lambda)v) \le \lambda\psi^{c,\varepsilon}(u) + (1-\lambda)\varphi^{c,\varepsilon}(v).$$

  Denoting

  $$h(u) = e^{-(\lambda\psi + (1-\lambda)\varphi)^{c,\varepsilon}(u)},$$
  $$f(u) = e^{-\psi^{c,\varepsilon}(u)}, \quad g(u) = e^{-\varphi^{c,\varepsilon}(u)},$$

  we thus have shown that

  $$h(\lambda u + (1-\lambda)v) \ge f(u)^\lambda g(v)^{1-\lambda}.$$

  $\implies$ Prékopa–Leindler inequality:

  $$\int_{\mathcal{X}} h \ge \left(\int_{\mathcal{X}} f\right)^\lambda \left(\int_{\mathcal{X}} g\right)^{1-\lambda}.$$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any* $\varepsilon > 0$ *and* $v \in \mathbb{R}^N$,

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_{\mathcal{Y}} \mathrm{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- $I : \psi \mapsto \log(\int_{\mathcal{X}} e^{-\psi^{c,\varepsilon}})$ is concave: Prékopa–Leindler inequality:

$$\int_{\mathcal{X}} h \geq \left(\int_{\mathcal{X}} f \right)^\lambda \left(\int_{\mathcal{X}} g \right)^{1-\lambda}.$$

$\implies$ I is concave:

$$\begin{aligned} I(\lambda\psi+(1-\lambda)\varphi) &= \log \left(\int_{\mathcal{X}} h \right) \\ &\geq \lambda \log \left(\int_{\mathcal{X}} f \right) + (1-\lambda) \log \left(\int_{\mathcal{X}} g \right) \\ &= \lambda I(\psi) + (1-\lambda) I(\varphi). \end{aligned}$$

Strong convexity estimate of $\mathcal{K}^\varepsilon$

Theorem (D., 2021): *For any $\varepsilon > 0$ and $v \in \mathbb{R}^N$,*

$$\mathbb{V}\mathrm{ar}_\mu(v) \leq \left(e^{R_\mathcal{Y} \operatorname{diam}(\mathcal{X})} \frac{M_\rho}{m_\rho} + \varepsilon \right) \langle v | \nabla^2 \mathcal{K}^\varepsilon(\psi^\varepsilon) v \rangle.$$

- Notice that

$$\nabla^2 \mathcal{K}^\varepsilon(\psi) = \frac{1}{\varepsilon} \mathbb{E}_{x\sim\rho} \left(\operatorname{diag}(\pi_x^\varepsilon(\psi)) - \pi_x^\varepsilon(\psi) \pi_x^\varepsilon(\psi)^\top \right).$$

- Introduce $I : \mathbb{R}^N \to \mathbb{R}, \psi \mapsto \log(\int_\mathcal{X} e^{-\psi^{c,\varepsilon}})$. Notice that I is $\mathcal{C}^2$ **and concave**:

$$\begin{aligned}\nabla^2 I(\psi^\varepsilon) = &- \frac{1}{\varepsilon} \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} (\operatorname{diag}(\pi_x^\varepsilon(\psi^\varepsilon)) - \pi_x^\varepsilon(\psi^\varepsilon)\pi_x^\varepsilon(\psi^\varepsilon)^\top) \\ &+ \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} \pi_x^\varepsilon(\psi^\varepsilon)\pi_x^\varepsilon(\psi^\varepsilon)^\top - \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} \pi_x^\varepsilon(\psi^\varepsilon) \mathbb{E}_{x\sim\tilde{\rho}^\varepsilon} \pi_x^\varepsilon(\psi^\varepsilon)^\top \leq 0,\end{aligned}$$

where $\tilde{\rho}^\varepsilon := \frac{e^{-(\psi^\varepsilon)^{c,\varepsilon}}}{\int_\mathcal{X} e^{-(\psi^\varepsilon)^{c,\varepsilon}}}$.

Upper bound on $\left\|\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)\right\|_\infty$

Theorem (D., 2021): *For any* $\varepsilon > 0$,

$$\left\|\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)\right\|_\infty \lesssim \varepsilon^{\alpha'}.$$

- We have

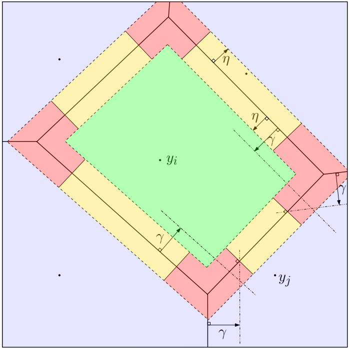
$$[\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)]_i = \int_{\mathcal{X}} \sum_{j\neq i} \left(\frac{f_i^\varepsilon(x) - f_j^\varepsilon(x)}{\varepsilon^2}\right) \pi^\varepsilon_{x,j}\pi^\varepsilon_{x,i}\mathrm{d}\rho(x),$$

where $\forall j$, $f_j^\varepsilon(x) = \langle x|y_j\rangle - \psi_j^\varepsilon$ and $\pi^\varepsilon_{x,j} = \frac{\exp(\frac{f_j^\varepsilon(x)}{\varepsilon})}{\sum_k \exp(\frac{f_k^\varepsilon(x)}{\varepsilon})}$.

Upper bound on $\left\|\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)\right\|_\infty$

We have:

$$[\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi^\varepsilon)]_i = \int_{\mathcal{X}} \sum_{j\neq i} \left(\frac{f_i^\varepsilon(x) - f_j^\varepsilon(x)}{\varepsilon^2}\right) \pi_{x,j}^\varepsilon \pi_{x,i}^\varepsilon \mathrm{d}\rho(x).$$



Key technical result: for any $\eta, \gamma > 0$,

$$\begin{aligned}[\frac{\partial}{\partial\varepsilon}(\nabla\mathcal{K}^\varepsilon)(\psi)]_i \lesssim\ & \frac{1}{\varepsilon^2} e^{-\eta/\varepsilon} + \frac{\eta^{2+\alpha}}{\varepsilon^2} \\ & + \frac{\gamma^2}{\varepsilon^2}\left(\eta + e^{-\eta/\varepsilon}\right) \\ & + \frac{1}{\varepsilon^2} e^{-\tilde{\gamma}/\varepsilon}\Big(\eta + \varepsilon\eta e^{\eta/\varepsilon} \\ & \qquad - \varepsilon^2(e^{\eta/\varepsilon} - 1)\Big),\end{aligned}$$

where $\tilde{\gamma} = \gamma\delta - \frac{\mathrm{diam}(\mathcal{Y})^2}{\delta}\eta$ and $\delta = \min_{i\neq j} \|y_i - y_j\| > 0$.