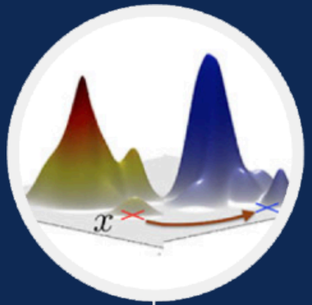
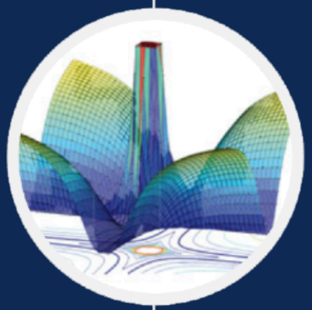
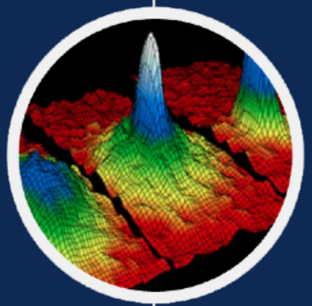
A crash-course on

Entropic Regularized Optimal Transport







Simone Di Marino

Department of Mathematics
Machine Learning Genoa Center
University of Genoa

Augusto Gerolin

Department of Mathematical and Statistics
Dep. of Chemistry and Biomolecular Sciences
University of Ottawa

uOttawa

Chaires de recherche du Canada — Canada Research Chairs

Canada

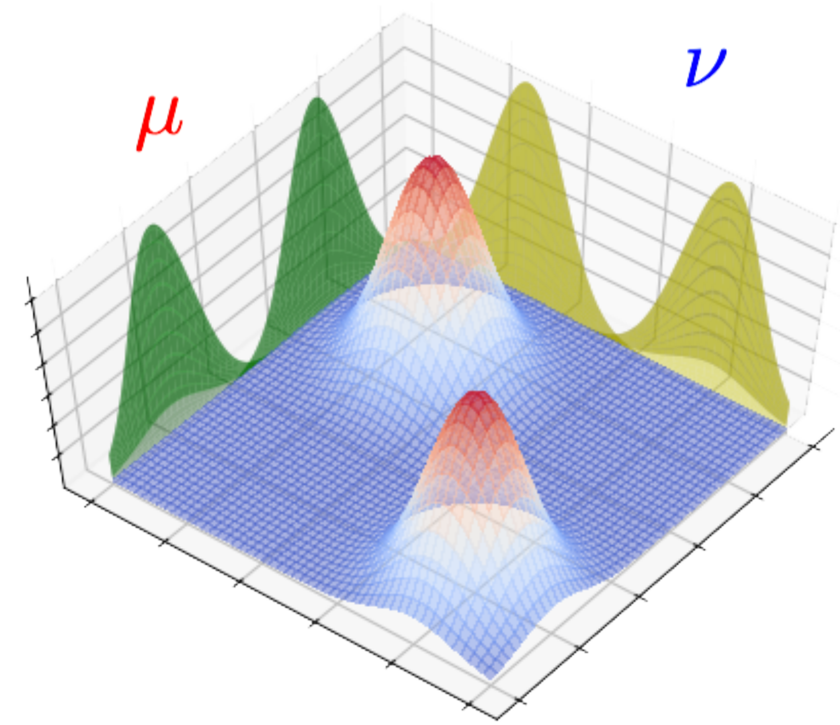
School *Mathematics of Machine Learning*
Centro De Giorgi, Pisa - 9-16th January 2023

PACE

Monge-Kantorovich formulation
of Optimal Transport

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

$c(x,y) = |x-y|^2$



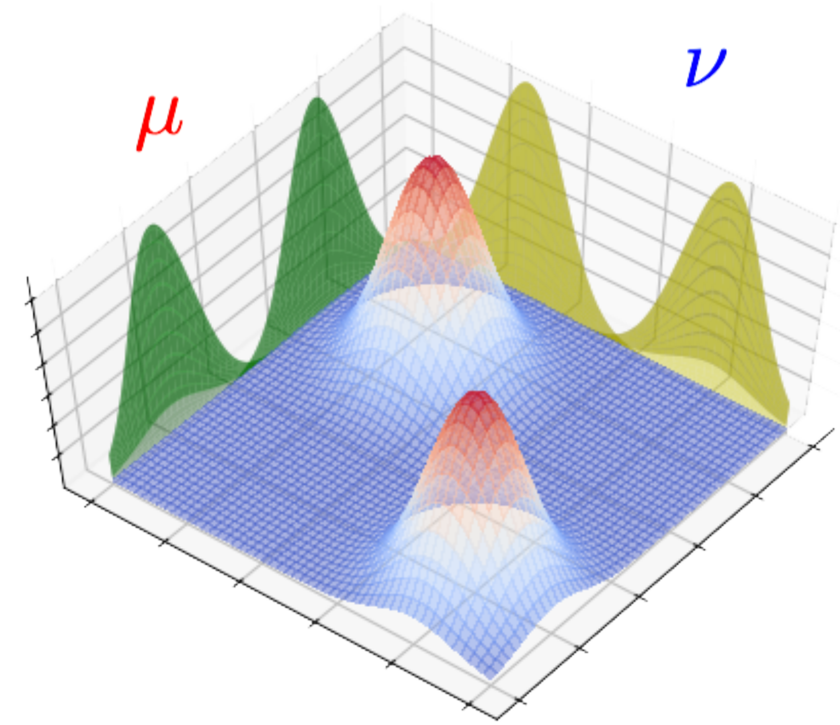
<u>Problem</u>: *Find the best coupling* γ between μ and ν.

$$\min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \gamma \in \Pi(\mu,\nu) \right\}$$

Monge-Kantorovich formulation
of Optimal Transport

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

$c(x,y) = |x-y|^2$



Problem: *Find the best coupling* γ between μ and ν.

$$\min\left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \gamma \in \Pi(\mu,\nu) \right\}$$

γ is a probability in $\mathbb{R}^d \times \mathbb{R}^d$

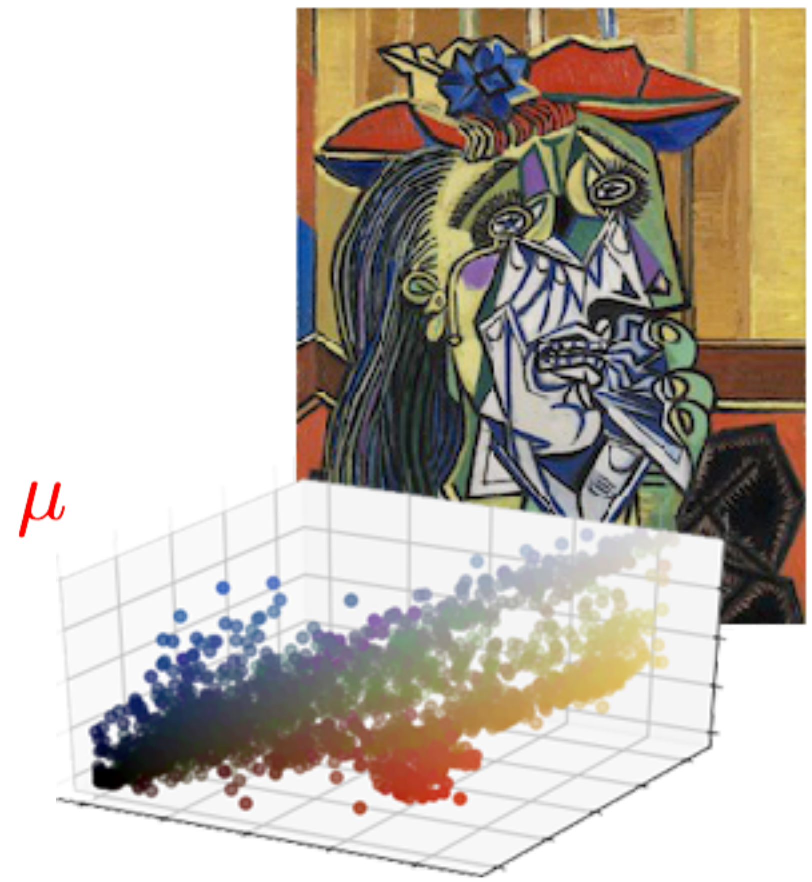
such that $(e_1)_\sharp\gamma = \mu$ and $(e_2)_\sharp\gamma = \nu$

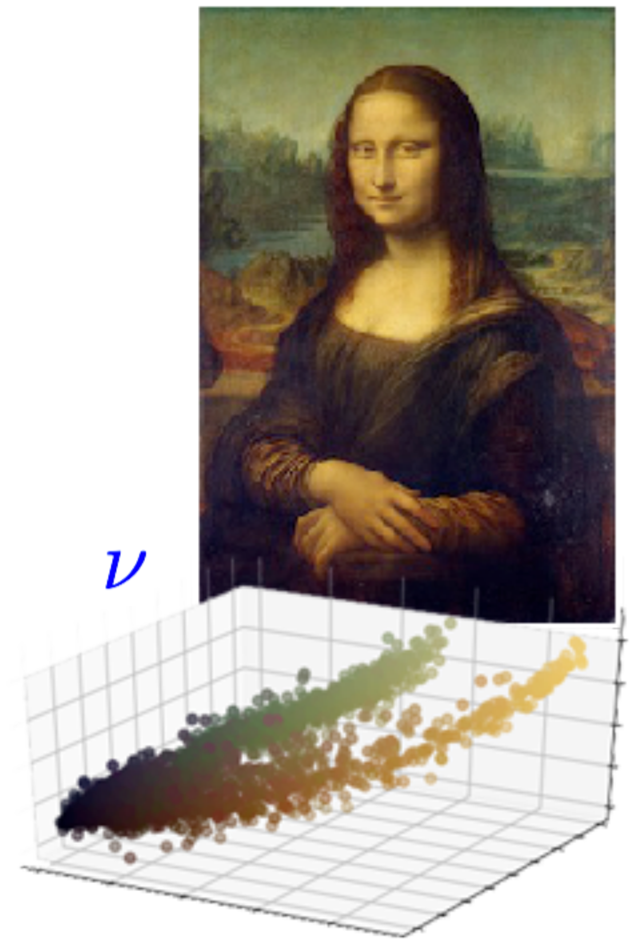
uOttawa

Monge-Kantorovich formulation
of Optimal Transport

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

$c(x, y) = |x - y|^2$





Problem: *Find the best coupling* γ between μ and ν.

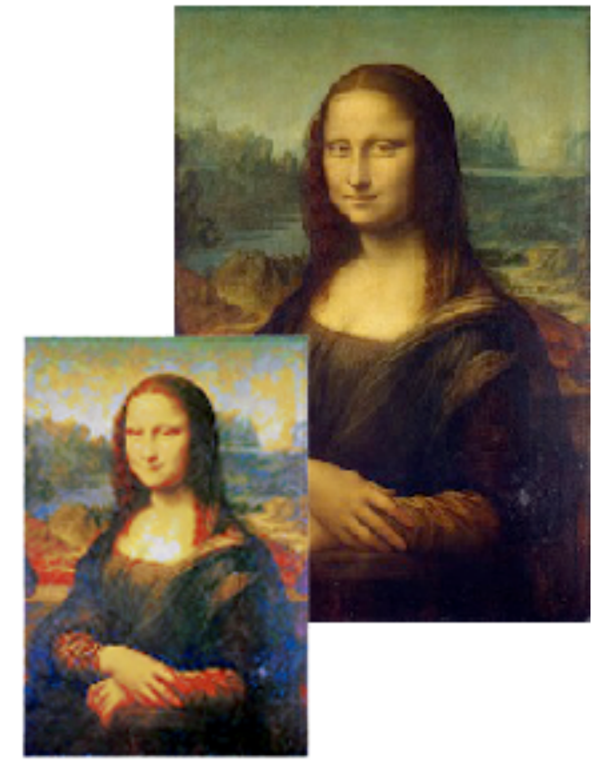
$$\min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 d\gamma(x, y) : \gamma \in \Pi(\mu, \nu) \right\}$$

uOttawa

Monge-Kantorovich formulation
of Optimal Transport

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

$c(x,y) = |x-y|^2$



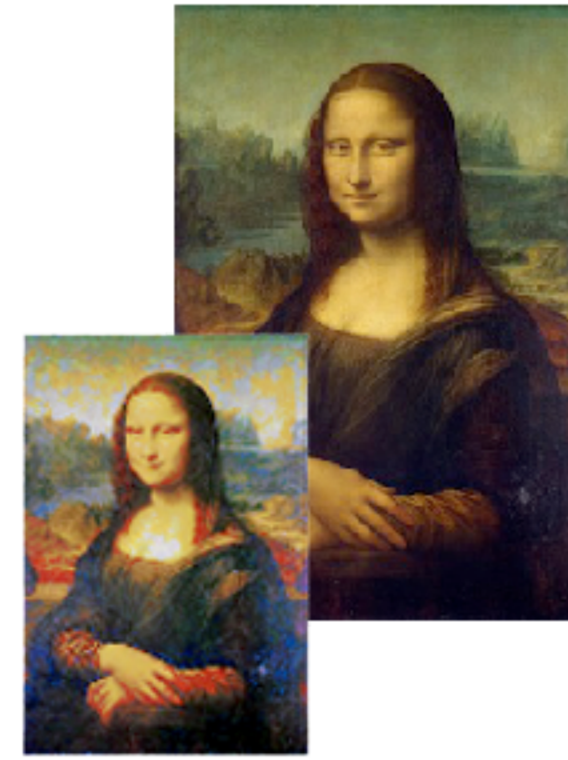
Problem: *Find the best coupling* γ between μ and ν.

$$\min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \gamma \in \Pi(\mu, \nu) \right\}$$

Monge-Kantorovich formulation
of Optimal Transport

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

$c(x,y) = |x-y|^2$



Problem: *Find the best coupling* γ between μ and ν.

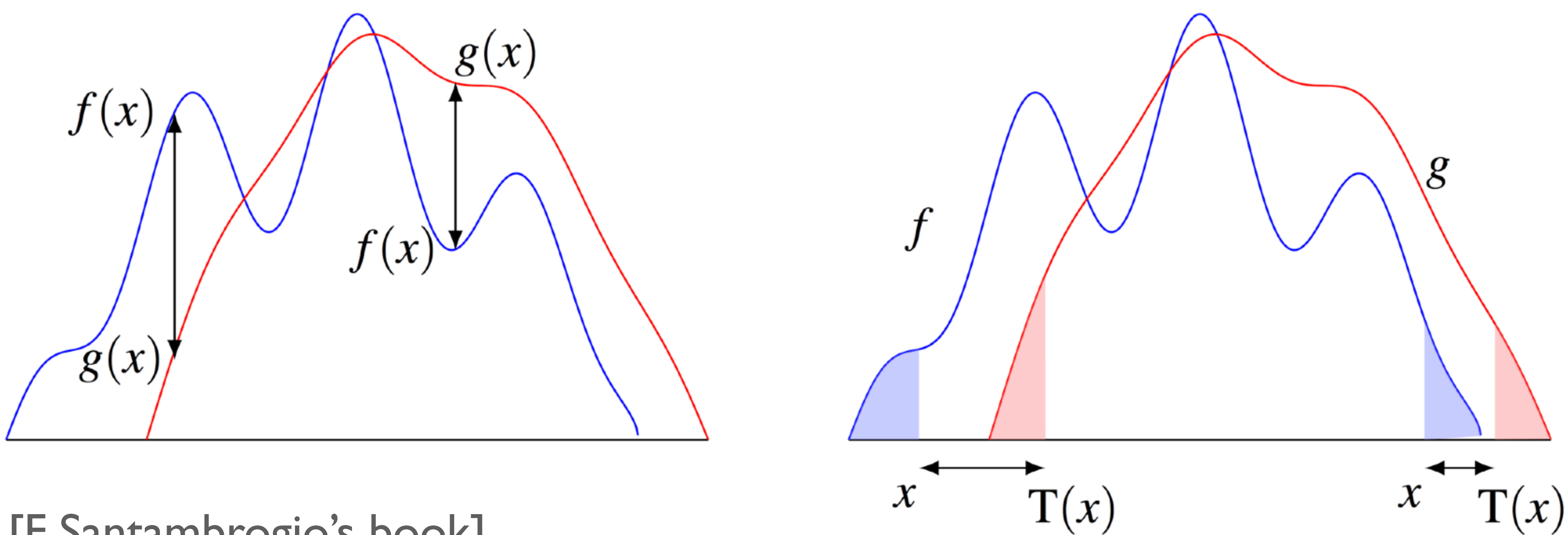
$$W_2^2(\mu,\nu) = \min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \gamma \in \Pi(\mu,\nu) \right\}$$

uOttawa

Why Optimal Transport?

It is define the Wasserstein distance on the space of probability measures.

$$W_2^2(\mu,\nu) = \min\left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \gamma\in\Pi(\mu,\nu)\right\}$$

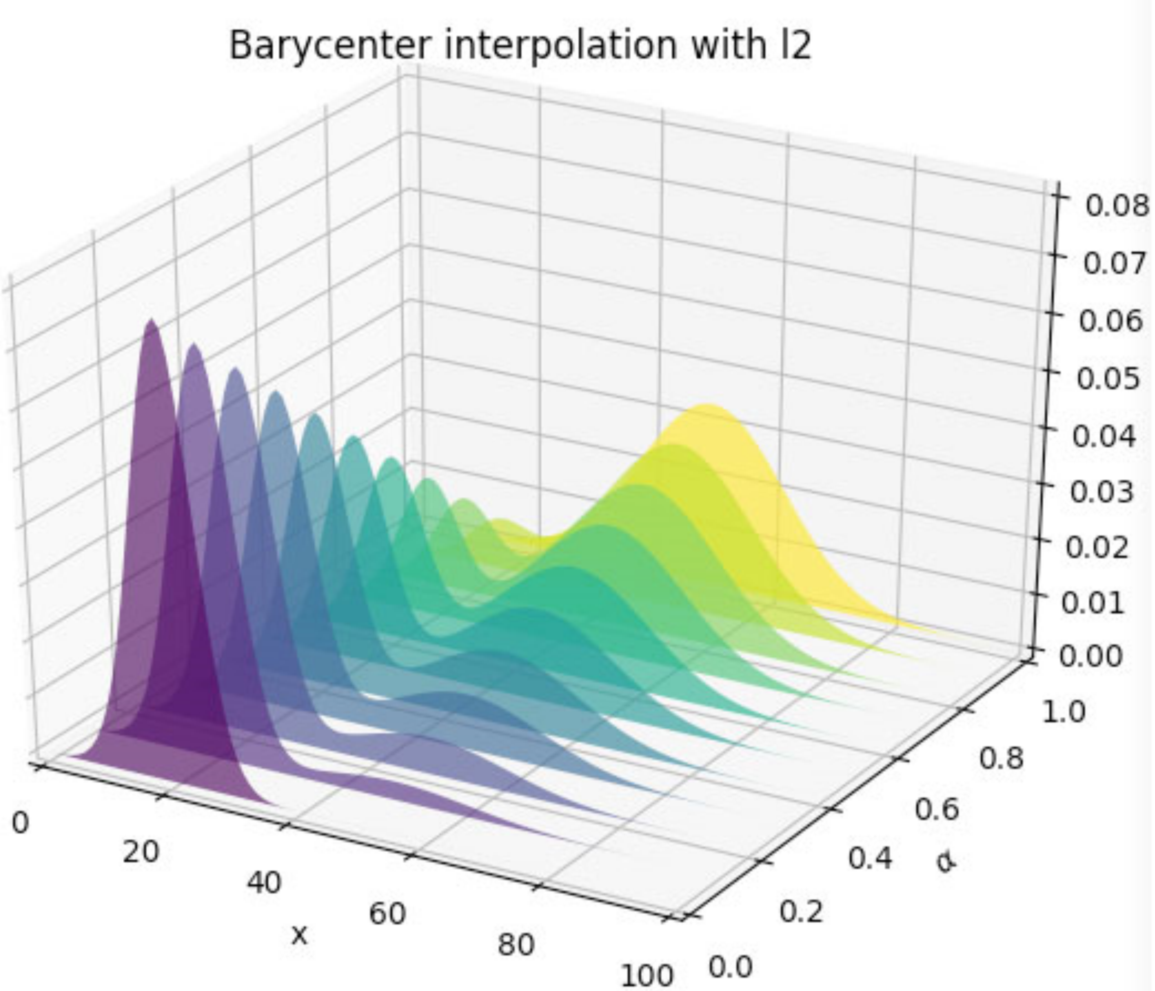


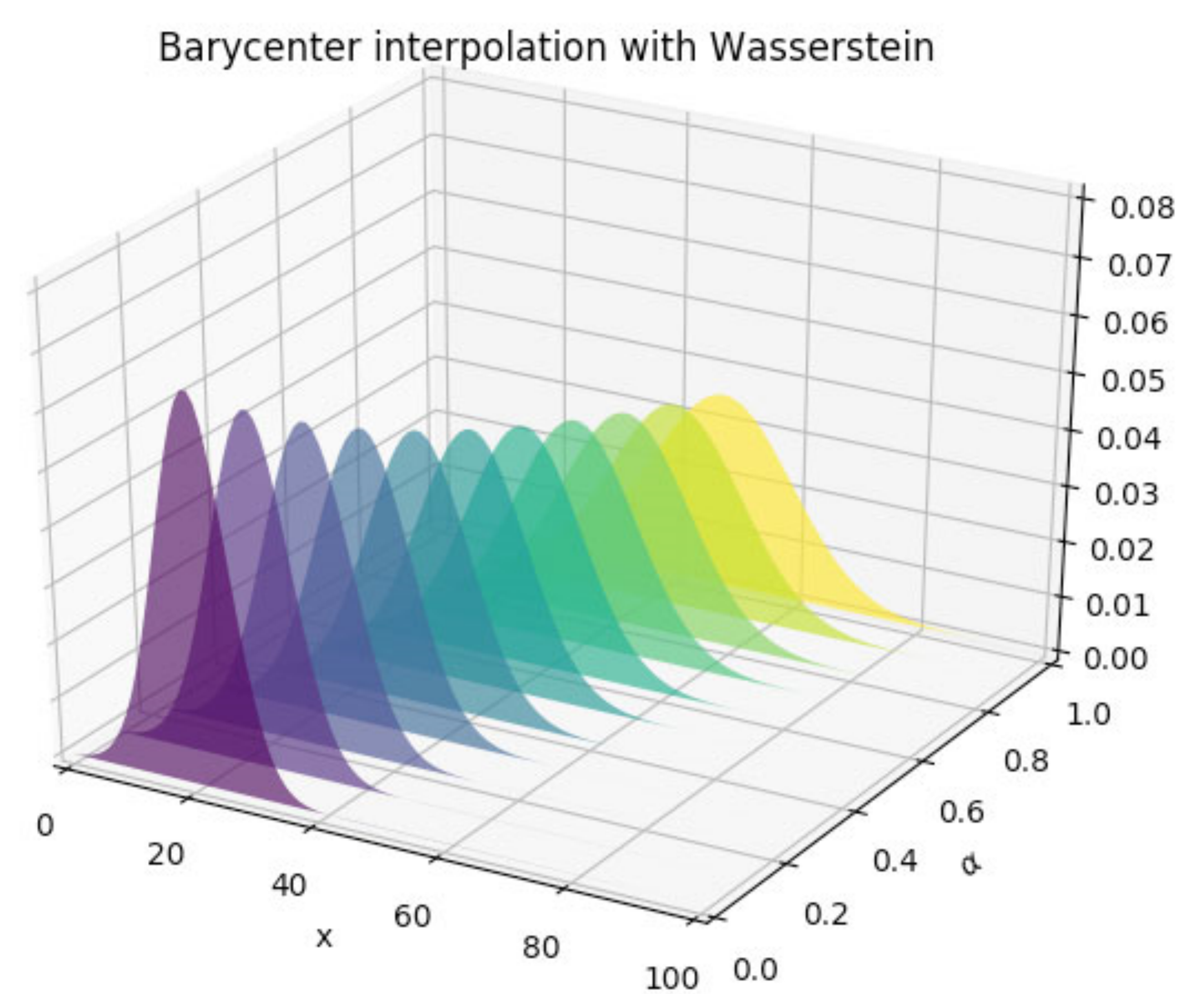
[F. Santambrogio's book]

Why Optimal Transport?

It is define the Wasserstein distance on the space of probability measures.

$$W_2^2(\mu,\nu) = \min\left\{ \int_{\mathbb{R}^d\times\mathbb{R}^d} |x-y|^2 d\gamma(x,y) : \gamma \in \Pi(\mu,\nu) \right\}$$

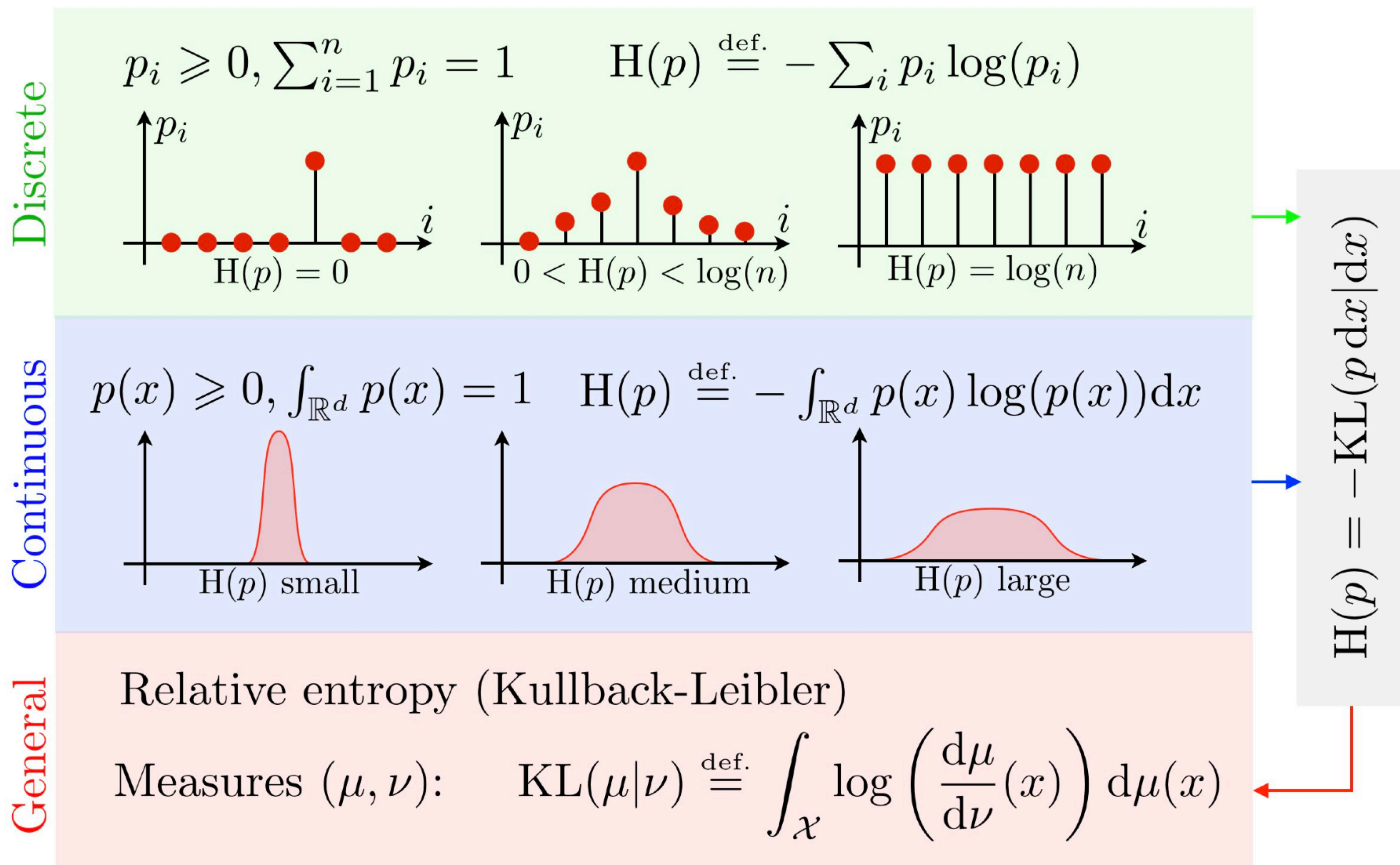




uOttawa

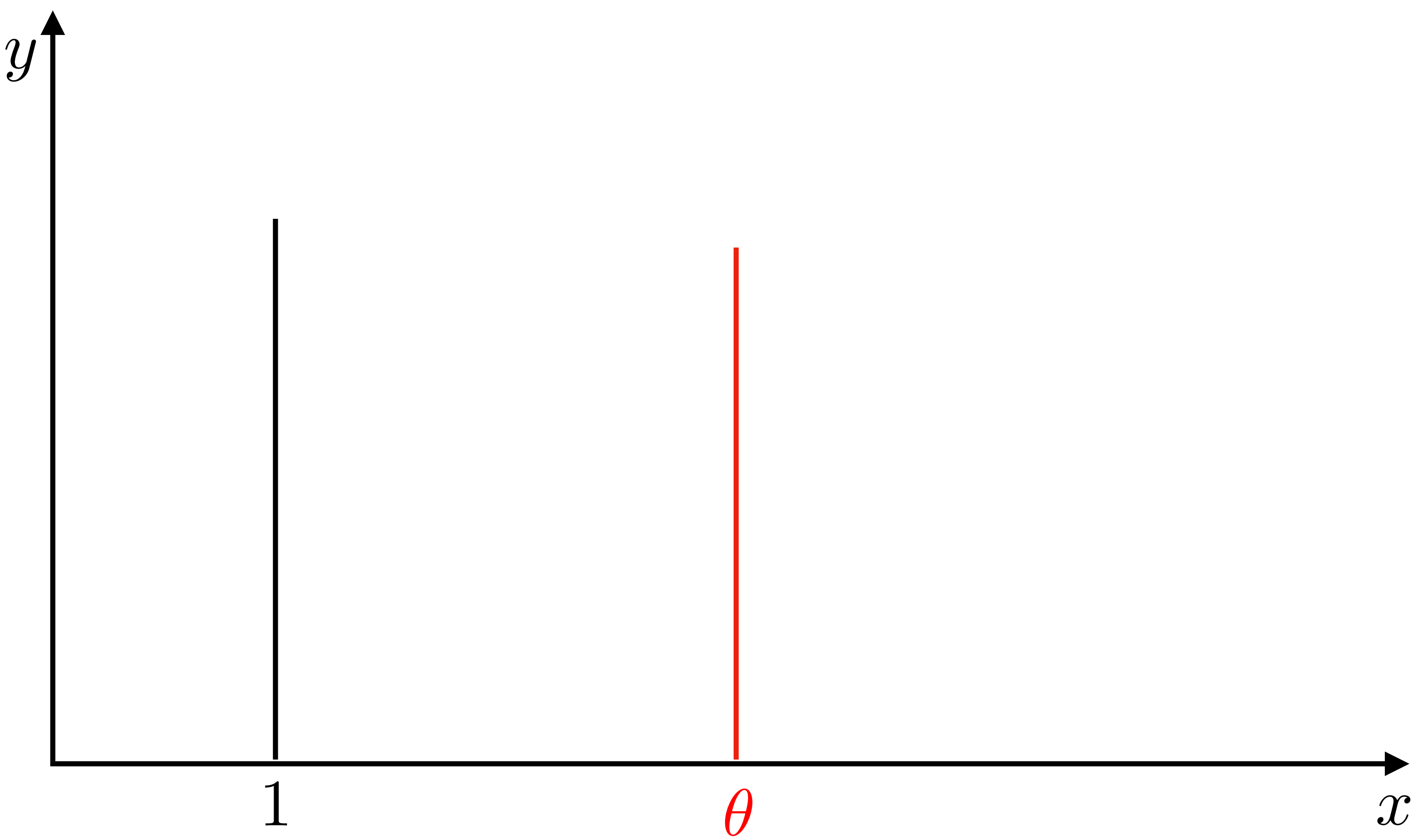
Relative Entropy between probability measures

[*Gabriel Péyre, Twitter*]



uOttawa

An enlightening example



$$P_0(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad P_\theta(x) = \begin{cases} 1 & \text{if } x = \theta \\ 0 & \text{otherwise} \end{cases}$$

Kullback-Leibler divergence

$$\mathrm{KL}(\mathrm{P}_0|\mathrm{P}_\theta) = \begin{cases} 0 & \text{if } \theta = 0 \\ \infty & \text{otherwise} \end{cases}$$

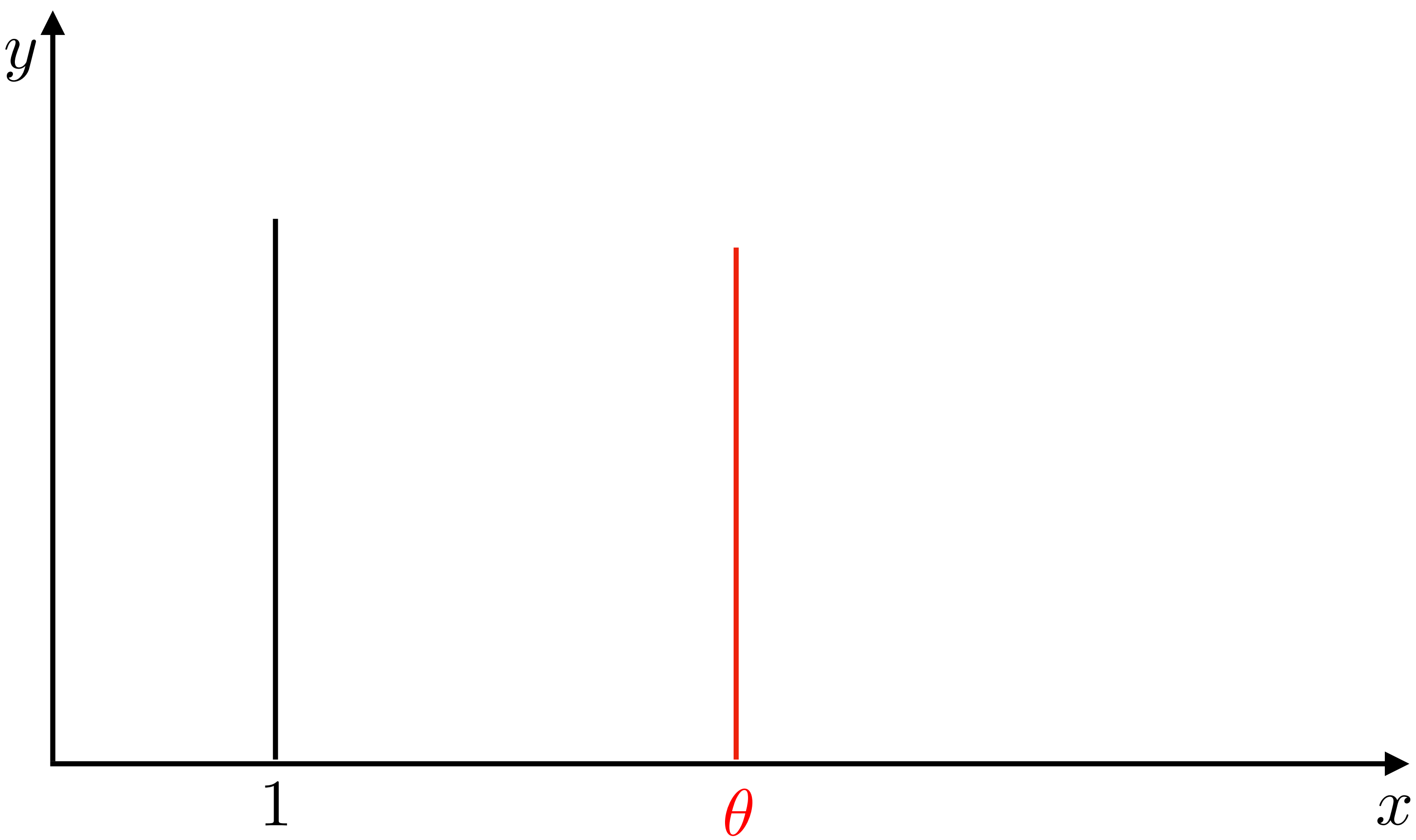
Jensen–Shannon divergence

$$\mathrm{JSD}(\mathrm{P}_0|\mathrm{P}_\theta) = \begin{cases} 0 & \text{if } \theta = 0 \\ \log 2 & \text{otherwise.} \end{cases}$$

[Ian Goodfellow et al, Generative Adversarial Nets, 2014]
[Martin Arjovsky et al, Wasserstein GAN, 2017]
[Ishann Gulrajani et al, WGAN-GP, 2017]
[Genevay et al, Sinkhorn GAN, 2017]

uOttawa

An enlightening example



$$P_0(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases} \qquad P_\theta(x) = \begin{cases} 1 & \text{if } x = \theta \\ 0 & \text{otherwise} \end{cases}$$

Kullback-Leibler divergence

$$\mathrm{KL}(\mathrm{P}_0|\mathrm{P}_\theta) = \begin{cases} 0 & \text{if } \theta = 0 \\ \infty & \text{otherwise} \end{cases}$$

Jensen–Shannon divergence

$$\mathrm{JSD}(\mathrm{P}_0|\mathrm{P}_\theta) = \begin{cases} 0 & \text{if } \theta = 0 \\ \log 2 & \text{otherwise.} \end{cases}$$

2-Wasserstein distance

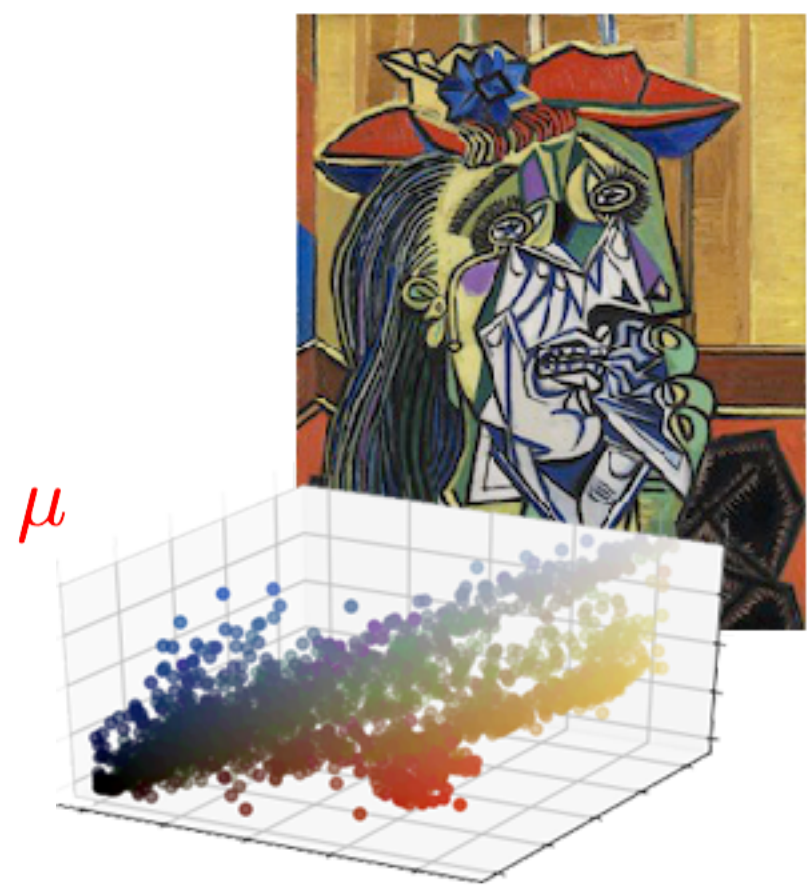
$$W_2^2(P_0, P_\theta) = |\theta|^2$$

[Ian Goodfellow et al, Generative Adversarial Nets, 2014]
[Martin Arjovsky et al, Wasserstein GAN, 2017]
[Ishann Gulrajani et al, WGAN-GP, 2017]
[Genevay et al, Sinkhorn GAN, 2017]

Monge-Kantorovich formulation
of Optimal Transport

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

$c(x,y) = |x-y|^2$



Problem: *Find the best coupling* γ between μ and ν.

$$\min \left\{ \int_{\mathbb{R}^d \times \mathbb{R}^d} |x-y|^2 d\gamma(x,y) \,:\, \gamma \in \Pi(\mu,\nu) \right\}$$

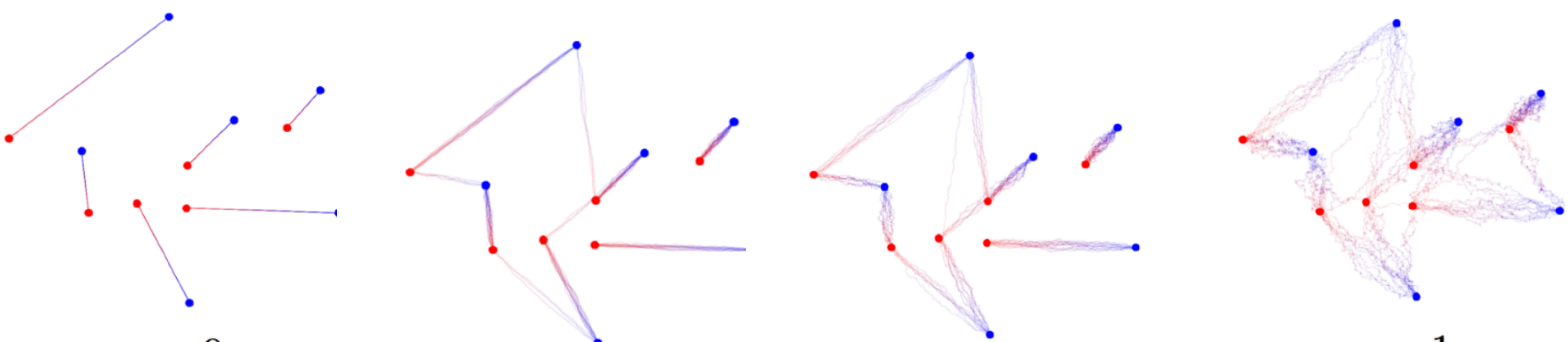
Shannon entropy regularized Optimal Transport

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y)d\left(\rho_1\otimes\rho_2\right).$$

After the works of I. Czisar, L. Ruschendorf, J. M. Borwein, A. S. Lewis and R. D. Nussbaum, C. Léonard, T. Georgiou, Y. Chen, M. Pavon, N. Gigli and L. Tamanini …

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$



$\varepsilon = 0$ $\quad \varepsilon = .05$ $\quad \varepsilon = 0.2$ $\quad \varepsilon = 1$

[Gabriel Péyre, Twitter] $c(x,y) = |x-y|^2$

uOttawa

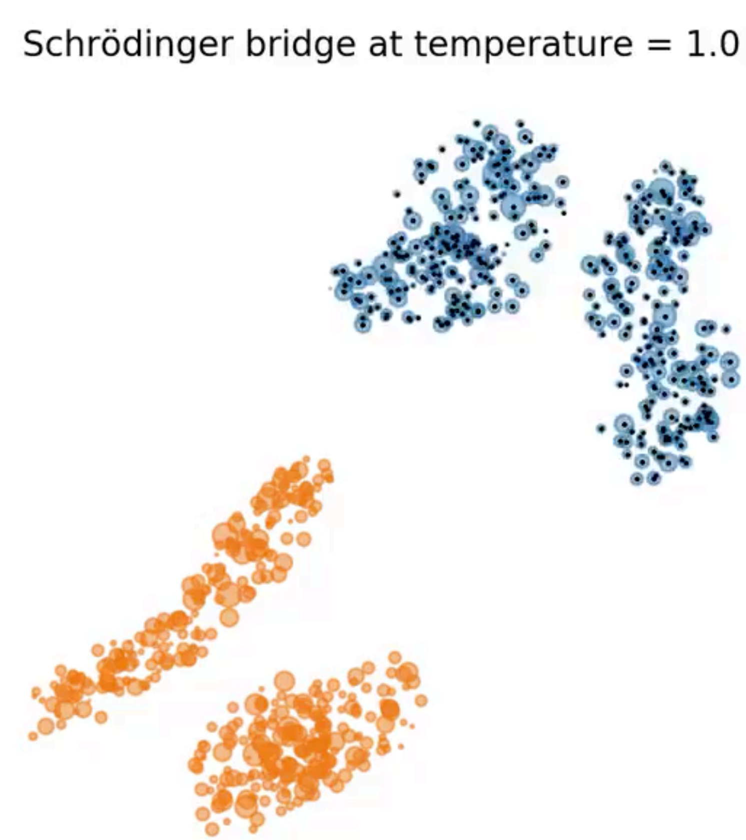
Shannon entropy regularized Optimal Transport

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y)d\left(\rho_1\otimes\rho_2\right).$$

After the works of I. Czisar, L. Ruschendorf, J. M. Borwein, A. S. Lewis and R. D. Nussbaum, C. Léonard, T. Georgiou, Y. Chen, M. Pavon, N. Gigli and L. Tamanini …

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$



[Lénaïc Chizat, Twitter] $c(x,y) = |x-y|^2$

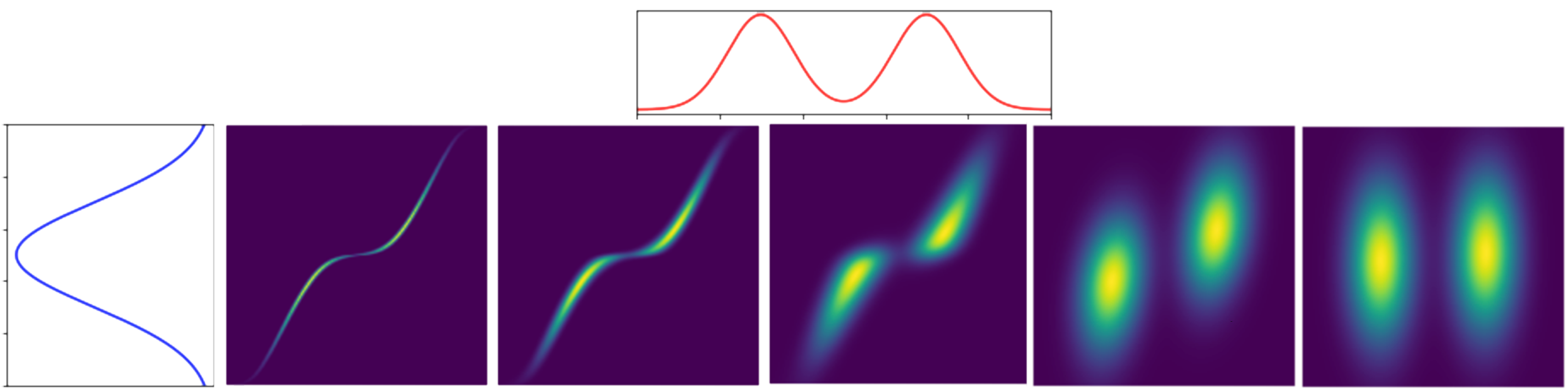
Shannon entropy regularized Optimal Transport

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y) d\left(\rho_1\otimes\rho_2\right).$$

After the works of I. Czisar, L. Ruschendorf, J. M. Borwein, A. S. Lewis and R. D. Nussbaum, C. Léonard, T. Georgiou, Y. Chen, M. Pavon, N. Gigli and L. Tamanini ...

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$



$c(x,y) = |x-y|^2$ $\quad\varepsilon$ *increases*

Shannon entropy regularized Optimal Transport

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad\qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y)d\left(\rho_1\otimes\rho_2\right).$$

After the works of I. Czisar, L. Ruschendorf, J. M. Borwein, A. S. Lewis and R. D. Nussbaum, C. Léonard, T. Georgiou, Y. Chen, M. Pavon, N. Gigli and L. Tamanini …

There exists a unique minimizer γ^ε *(strictly convex)* $\quad \gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

[E. Schrödinger, 1932]

$$\begin{cases} a^\varepsilon(x)\displaystyle\int_Y b^\varepsilon(y)e^{-c(x,y)/\varepsilon}d\rho_2(y) = 1 \\ b^\varepsilon(y)\displaystyle\int_X a^\varepsilon(y)e^{-c(x,y)/\varepsilon}d\rho_1(x) = 1 \end{cases}$$

Shannon entropy regularized Optimal Transport

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y) d\left(\rho_1\otimes\rho_2\right).$$

After the works of I. Czisar, L. Ruschendorf, J. M. Borwein, A. S. Lewis and R. D. Nussbaum, C. Léonard, T. Georgiou, Y. Chen, M. Pavon, N. Gigli and L. Tamanini …

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

[E. Schrödinger, 1932]

$$\begin{cases} a^\varepsilon(x) \displaystyle\int_Y b^\varepsilon(y)e^{-c(x,y)/\varepsilon} d\rho_2(y) = 1 \\ b^\varepsilon(y) \displaystyle\int_X a^\varepsilon(y)e^{-c(x,y)/\varepsilon} d\rho_1(x) = 1 \end{cases}$$

Explicit solutions when $c(x,y) = |x-y|^2$ *and* ρ_1, ρ_2 *are (multi-dimensional) Gaussians.*

[AG, Grossi, Gori-Giorgi, 2019]
[Mallasto, AG, Mihn, 2020]
[del Barrio, Loubes, 2020]
[Janati, Muzellec, Peyré, Cuturi, 2020]

Shannon entropy regularized Optimal Transport

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y)d\left(\rho_1\otimes\rho_2\right).$$

After the works of I. Czisar, L. Ruschendorf, J. M. Borwein, A. S. Lewis and R. D. Nussbaum, C. Léonard, T. Georgiou, Y. Chen, M. Pavon, N. Gigli and L. Tamanini …

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

[E. Schrödinger, 1932]

$$\begin{cases} a^\varepsilon(x)\displaystyle\int_Y b^\varepsilon(y)e^{-c(x,y)/\varepsilon}d\rho_2(y) = 1 \\ b^\varepsilon(y)\displaystyle\int_X a^\varepsilon(y)e^{-c(x,y)/\varepsilon}d\rho_1(x) = 1 \end{cases}$$

Computed numerically via Sinkhorn algorithm…

Sinkhorn algorithm

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x) b^\varepsilon(y) e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

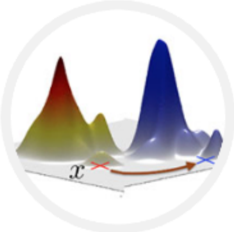
$$\begin{cases} a^\varepsilon(x) \displaystyle\int_Y b^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_2(y) = 1 \\ b^\varepsilon(y) \displaystyle\int_X a^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_1(x) = 1 \end{cases}$$

The algorithm: $a^0(x) = b^0(y) = 1$

Define the sequences (iteratively) $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ by

$$b^n(y) = \frac{1}{\int e^{-c(x,y)/\varepsilon} a^{n-1}(x) d\rho_1(x)},$$

$$a^n(x) = \frac{1}{\int e^{-c(x,y)/\varepsilon} b^n(y) d\rho_2(y)}.$$



Sinkhorn algorithm

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x) b^\varepsilon(y) e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

$$\begin{cases} a^\varepsilon(x) \displaystyle\int_Y b^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_2(y) = 1 \\ b^\varepsilon(y) \displaystyle\int_X a^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_1(x) = 1 \end{cases}$$

The algorithm: $a^0(x) = b^0(y) = 1$

Define the sequences (iteratively) $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ by

$$b^n(y) = \frac{1}{\int e^{-c(x,y)/\varepsilon} a^{n-1}(x) d\rho_1(x)},$$

$$a^n(x) = \frac{1}{\int e^{-c(x,y)/\varepsilon} b^n(y) d\rho_2(y)}.$$

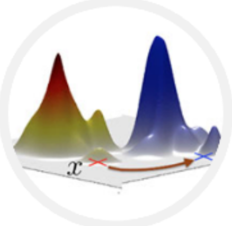
Theoretical guarantees of convergence

[R. Sinkhorn, 1964]
[J. Franklin and J. Lorenz, 1989]
[L. Ruschendorf, 1995]
[Y. Chen, T. Georgiou and M. Pavon, 2016]

Approximates OT distances in near-linear time

[J. Altschuler, J. Weed, P. Rigollet, 2017]

uOttawa



Sinkhorn algorithm

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x) b^\varepsilon(y) e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system
$$\begin{cases} a^\varepsilon(x) \displaystyle\int_Y b^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_2(y) = 1 \\ b^\varepsilon(y) \displaystyle\int_X a^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_1(x) = 1 \end{cases}$$

The algorithm: $a^0(x) = b^0(y) = 1$

Define the sequences (iteratively) $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ by

$$b^n(y) = \frac{1}{\int e^{-c(x,y)/\varepsilon} a^{n-1}(x) d\rho_1(x)},$$

$$a^n(x) = \frac{1}{\int e^{-c(x,y)/\varepsilon} b^{n}(y) d\rho_2(y)}.$$

Theoretical guarantees of convergence

[R. Sinkhorn, 1964]
[J. Franklin and J. Lorenz, 1989]
[L. Ruschendorf, 1995]
[Y. Chen, T. Georgiou and M. Pavon, 2016]

Approximates OT distances in near-linear time

[J. Altschuler, J. Weed, P. Rigollet, 2017]

Martin Idel (2016)

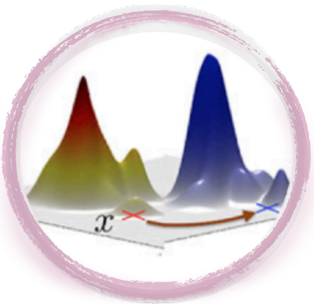
A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps

Probabilistic / Large Deviations / Stochastic control

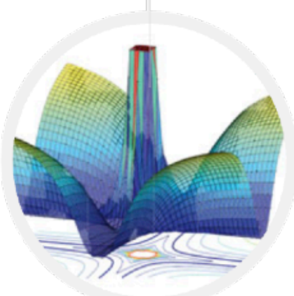
Cristian Léonard (2013)
Y. Chen, T. Georgiou and M. Pavon (2020)

uOttawa

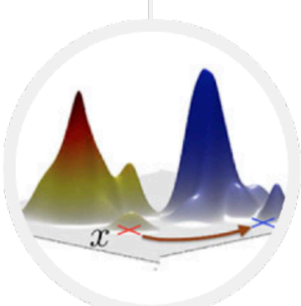
Outline



Duality approach for the Shannon Entropy-Regularized Optimal Transport



Convergence proof of the Sinkhorn algorithm (variational)



Applications

Shannon Entropy-Regularized Optimal Transport

A duality approach

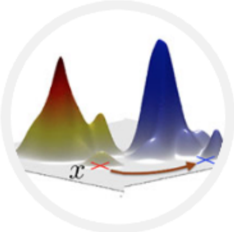
Setting: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures and $\varepsilon > 0$ be a positive number.

Dual functional Let $u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2)$

$$D_\varepsilon(u,v) = \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2).$$

Dual (Kantorovich) problem: $\sup\left\{ D_\varepsilon(u,v) \ : \ u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2) \right\}.$

where $L^{\exp}_{\varepsilon}(X,\rho_1) = \left\{ u : X \to [-\infty, \infty[: u \text{ is a measurable function in } (X,\rho_1) \text{ and } 0 < \int_X e^{u/\varepsilon}\, d\rho_1 < \infty \right\}.$



Shannon Entropy-Regularized Optimal Transport

A duality approach

Setting: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures and $\varepsilon > 0$ be a positive number.

Dual functional Let $u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2)$

$$D_\varepsilon(u, v) = \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X \times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2).$$

Dual (Kantorovich) problem: $\sup \left\{ D_\varepsilon(u, v) \ : \ u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2) \right\}.$

where $L^{\exp}_{\varepsilon}(X, \rho_1) = \left\{ u : X \to [-\infty, \infty[: u \text{ is a measurable function in } (X, \rho_1) \text{ and } 0 < \int_X e^{u/\varepsilon} d\rho_1 < \infty \right\}.$

Entropic Transform: The entropic (c, ε)-transforms $\mathcal{F}^{(c,\varepsilon)}_1 : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}^{(c,\varepsilon)}_2 : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}^{(c,\varepsilon)}_1(u)(y) := -\varepsilon \log \left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1(x) \right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}^{(c,\varepsilon)}_2(v)(x) := -\varepsilon \log \left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}} d\rho_2 \right)$$

Shannon Entropy-Regularized Optimal Transport

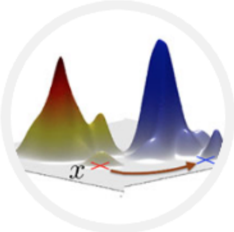
A duality approach

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

In particular we can say that $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_\varepsilon(u, v) \ : \ v \in L^{\exp}_\varepsilon(\rho_2)\}$.

Proof idea:



Shannon Entropy-Regularized Optimal Transport

A duality approach

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

In particular we can say that $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_\varepsilon(u, v) \ : \ v \in L^{\exp}_\varepsilon(\rho_2)\}$.

Proof idea:

$$D_\varepsilon(u, v) = \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2),$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

In particular we can say that $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_\varepsilon(u, v) \ : \ v \in L^{\exp}_\varepsilon(\rho_2)\}$.

Proof idea:

$$D_\varepsilon(u, v) = \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2),$$

$$= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_Y e^{\frac{v(y)}{\varepsilon}} \left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) d\rho_2,$$

uOttawa

Shannon Entropy-Regularized Optimal Transport

A duality approach

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

In particular we can say that $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_\varepsilon(u, v) \ : \ v \in L^{\exp}_\varepsilon(\rho_2)\}$.

Proof idea:

$$\begin{aligned}
D_\varepsilon(u, v) &= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2), \\
&= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_Y e^{\frac{v(y)}{\varepsilon}} \left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) d\rho_2, \\
&= \int_X u(x) d\rho_1(x) + \int_Y v(y) - \varepsilon e^{\frac{v(y)-u^{(c,\varepsilon)}(y)}{\epsilon}} d\rho_2(y).
\end{aligned}$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

In particular we can say that $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_\varepsilon(u, v) \ : \ v \in L^{\exp}_\varepsilon(\rho_2)\}$.

Proof idea:

$$\begin{aligned}
D_\varepsilon(u, v) &= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2), \\
&= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_Y e^{\frac{v(y)}{\varepsilon}} \left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) d\rho_2, \\
&= \int_X u(x) d\rho_1(x) + \int_Y v(y) - \varepsilon e^{\frac{v(y)-u^{(c,\varepsilon)}(y)}{\epsilon}} d\rho_2(y).
\end{aligned}$$

Therefore, for any $v \in L^{\exp}_\varepsilon(\rho_1)$, $D_\varepsilon(u, v) \leq D_\varepsilon(u, u^{(c,\varepsilon)})$, since the function $g(t) = t - \varepsilon e^{(t-a)/\varepsilon}$ is strictly concave and attains its maximum in $t = a$.

Shannon Entropy-Regularized Optimal Transport

A duality approach

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

In particular we can say that $u^{(c,\varepsilon)} \in \operatorname{argmax}\{D_\varepsilon(u, v) \ : \ v \in L^{\exp}_\varepsilon(\rho_2)\}$.

Proof idea:

$$\begin{aligned}
D_\varepsilon(u, v) &= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_{X\times Y} e^{\frac{u(x)+v(y)-c(x,y)}{\varepsilon}} d(\rho_1 \otimes \rho_2), \\
&= \int_X u(x) d\rho_1(x) + \int_Y v(y) d\rho_2(y) - \varepsilon \int_Y e^{\frac{v(y)}{\varepsilon}} \left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1 \right) d\rho_2, \\
&= \int_X u(x) d\rho_1(x) + \int_Y v(y) - \varepsilon e^{\frac{v(y)-u^{(c,\varepsilon)}(y)}{\epsilon}} d\rho_2(y).
\end{aligned}$$

Therefore, for any $v \in L^{\exp}_\varepsilon(\rho_1)$, $D_\varepsilon(u, v) \leq D_\varepsilon(u, u^{(c,\varepsilon)})$, since the function $g(t) = t - \varepsilon e^{(t-a)/\varepsilon}$ is strictly concave and attains its maximum in $t = a$.

In particular, $D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v)$ if and only if $v = u^{(c,\varepsilon)}$.

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}}\, d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}}\, d\rho_2\right)$$

Proposition: Let (X,d_X), (Y,d_Y) be Polish spaces, $u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2)$ and $\varepsilon > 0$. $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost,

$(i) u^{(c,\varepsilon)}(y) \in L^{\infty}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\infty}(\rho_1)$.

(ii) $u^{(c,\varepsilon)}(y) \in L^{\exp}_{\varepsilon}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\exp}_{\varepsilon}(\rho_1)$. Moreover $|\lambda_{u^{(c,\varepsilon)}(y)} + \lambda_u| \leq \|c\|_\infty$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}} d\rho_2\right)$$

Proposition: Let (X, d_X), (Y, d_Y) be Polish spaces, $u \in L^{\exp}_{\varepsilon}(\rho_1)$, $v \in L^{\exp}_{\varepsilon}(\rho_2)$ and $\varepsilon > 0$. $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost,

(i) $u^{(c,\varepsilon)}(y) \in L^{\infty}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\infty}(\rho_1)$.

(ii) $u^{(c,\varepsilon)}(y) \in L^{\exp}_{\varepsilon}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\exp}_{\varepsilon}(\rho_1)$. Moreover $|\lambda_{u^{(c,\varepsilon)}(y)} + \lambda_u| \leq \|c\|_{\infty}$

Proof idea: (i) If $u \in L^{\exp}_{\varepsilon}(\rho_1)$ then

$$u^{(c,\varepsilon)}(y) = -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1\right)$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}} d\rho_2\right)$$

Proposition: Let (X,d_X), (Y,d_Y) be Polish spaces, $u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2)$ and $\varepsilon > 0$. $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost,

(i) $u^{(c,\varepsilon)}(y) \in L^{\infty}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\infty}(\rho_1)$.

(ii) $u^{(c,\varepsilon)}(y) \in L^{\exp}_{\varepsilon}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\exp}_{\varepsilon}(\rho_1)$. Moreover $|\lambda_{u^{(c,\varepsilon)}(y)} + \lambda_u| \leq \|c\|_{\infty}$

Proof idea: (i) If $u \in L^{\exp}_{\varepsilon}(\rho_1)$ then

$$u^{(c,\varepsilon)}(y) = -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1\right) \leq -\varepsilon \log\left(e^{\frac{-\|c\|_{\infty}}{\varepsilon}} \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1\right) = \|c\|_{\infty} - \varepsilon \log\left(\int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1\right).$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}}\, d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}}\, d\rho_2\right)$$

Proposition: Let (X,d_X), (Y,d_Y) be Polish spaces, $u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2)$ and $\varepsilon > 0$. $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost,

(i) $u^{(c,\varepsilon)}(y) \in L^{\infty}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\infty}(\rho_1)$.

(ii) $u^{(c,\varepsilon)}(y) \in L^{\exp}_{\varepsilon}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\exp}_{\varepsilon}(\rho_1)$. Moreover $|\lambda_{u^{(c,\varepsilon)}(y)} + \lambda_u| \leq \|c\|_{\infty}$

Proof idea: (i) If $u \in L^{\exp}_{\varepsilon}(\rho_1)$ then

$$u^{(c,\varepsilon)}(y) = -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1\right) \leq -\varepsilon \log\left(e^{\frac{-\|c\|_\infty}{\varepsilon}} \int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1\right) = \|c\|_\infty - \varepsilon \log\left(\int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1\right).$$

Also, for the lower bound we have

$$u^{(c,\varepsilon)}(y) = -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1\right) \geq -\|c\|_\infty - \varepsilon \log\left(\int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1\right).$$

uOttawa

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}} d\rho_2\right)$$

Proposition: Let (X,d_X), (Y,d_Y) be Polish spaces, $u \in L^{\exp}_{\varepsilon}(\rho_1), v \in L^{\exp}_{\varepsilon}(\rho_2)$ and $\varepsilon > 0$. $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost,

(i) $u^{(c,\varepsilon)}(y) \in L^{\infty}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\infty}(\rho_1)$.

(ii) $u^{(c,\varepsilon)}(y) \in L^{\exp}_{\varepsilon}(\rho_2)$ and $v^{(c,\varepsilon)}(x) \in L^{\exp}_{\varepsilon}(\rho_1)$. Moreover $|\lambda_{u^{(c,\varepsilon)}(y)} + \lambda_u| \leq \|c\|_\infty$

Proof idea: Let us prove (ii), i.e. $u^{(c,\varepsilon)} \in L^{\infty}(\rho_2)$.

$$\int_Y e^{\frac{u^{(c,\varepsilon)}(y)}{\varepsilon}} d\rho_2(y) \leq \int_Y e^{\|c\|_\infty/\varepsilon}\left(\int_X e^{\frac{u(x)}{\varepsilon}} d\rho_1(x)\right)^{-1} d\rho_2(y) < +\infty.$$

uOttawa

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L_\varepsilon^{\exp}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L_\varepsilon^{\exp}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log \left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} \, d\rho_1(x) \right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log \left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}} \, d\rho_2 \right)$$

Proposition:

(i) if c is L-Lipschitz, then $u^{(c,\varepsilon)}$ is L-Lipschitz;

(ii) if $|c| \leq M$ then $\operatorname{osc}(u^{(c,\varepsilon)}) \leq 2M$.

(iii) if $|c| \leq M$, then $u^{(c,\varepsilon)} : L^\infty(\rho_1) \to L^p(\rho_2)$ is a 1-Lip. compact operator.

Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_\varepsilon(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_\varepsilon(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}}\, d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}}\, d\rho_2\right)$$

Proposition: (i) if c is L-Lipschitz, then $u^{(c,\varepsilon)}$ is L-Lipschitz;

Proof idea: Let $u \in L^{\exp}_\varepsilon(\rho_1)$, $y_1, y_2 \in Y$. We can assume WLOG that $u^{(c,\varepsilon)}(y_1) \geq u^{(c,\varepsilon)}(y_2)$;

$$|u^{(c,\varepsilon)}(y_1) - u^{(c,\varepsilon)}(y_2)| = \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_2)}{\varepsilon}}\, d\rho_1\right) - \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}}\, d\rho_1\right)$$

Shannon Entropy-Regularized Optimal Transport
A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L_\varepsilon^{\exp}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L_\varepsilon^{\exp}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}}\, d\rho_1(x)\right)$$

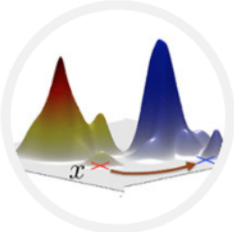
$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}}\, d\rho_2\right)$$

Proposition: (i) if c is L-Lipschitz, then $u^{(c,\varepsilon)}$ is L-Lipschitz;

Proof idea: Let $u \in L_\varepsilon^{\exp}(\rho_1)$, $y_1, y_2 \in Y$. We can assume WLOG that $u^{(c,\varepsilon)}(y_1) \geq u^{(c,\varepsilon)}(y_2)$;

$$\begin{aligned}|u^{(c,\varepsilon)}(y_1) - u^{(c,\varepsilon)}(y_2)| &= \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_2)}{\varepsilon}}\, d\rho_1\right) - \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}}\, d\rho_1\right)\\ &= \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)+c(x,y_1)-c(x,y_2)}{\varepsilon}}\, d\rho_1\right) - \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}}\, d\rho_1\right)\end{aligned}$$

uOttawa



Shannon Entropy-Regularized Optimal Transport

A duality approach

Entropic Transform: The entropic (c,ε)-transforms $\mathcal{F}_1^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_1) \to L^0(\rho_2)$, $\mathcal{F}_2^{(c,\varepsilon)} : L^{\exp}_{\varepsilon}(\rho_2) \to L^0(\rho_1)$ are defined by

$$u^{(c,\varepsilon)}(y) := \mathcal{F}_1^{(c,\varepsilon)}(u)(y) := -\varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y)}{\varepsilon}} d\rho_1(x)\right)$$

$$v^{(c,\varepsilon)}(x) := \mathcal{F}_2^{(c,\varepsilon)}(v)(x) := -\varepsilon \log\left(\int_Y e^{\frac{v(y)-c(x,y)}{\varepsilon}} d\rho_2\right)$$

Proposition: (i) if c is L-Lipschitz, then $u^{(c,\varepsilon)}$ is L-Lipschitz;

Proof idea: Let $u \in L^{\exp}_{\varepsilon}(\rho_1)$, $y_1, y_2 \in Y$. We can assume WLOG that $u^{(c,\varepsilon)}(y_1) \geq u^{(c,\varepsilon)}(y_2)$;

$$\begin{aligned}
|u^{(c,\varepsilon)}(y_1) - u^{(c,\varepsilon)}(y_2)| &= \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_2)}{\varepsilon}} d\rho_1\right) - \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1\right) \\
&= \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)+c(x,y_1)-c(x,y_2)}{\varepsilon}} d\rho_1\right) - \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1\right) \\
&\leq \varepsilon \log\left(e^{\frac{Ld(y_1,y_2)}{\varepsilon}} \int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1\right) - \varepsilon \log\left(\int_X e^{\frac{u(x)-c(x,y_1)}{\varepsilon}} d\rho_1\right) \\
&= Ld(y_1, y_2)
\end{aligned}$$

uOttawa

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The supremum

$$\sup\left\{D_\varepsilon(u,v)\ :\ u\in L^{\exp}_\varepsilon(\rho_1), v\in L^{\exp}_\varepsilon(\rho_2)\right\}$$

is attained for a unique couple (u_0,v_0) (up to the trivial tranformation$(u,v)\mapsto(u+a,v-a)$).
In particular we have $u_0\in L^\infty(\rho_1)$ and $v_0\in L^\infty(\rho_2)$;

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The supremum

$$\sup\{D_\varepsilon(u,v)\ :\ u\in L^{\exp}_\varepsilon(\rho_1), v\in L^{\exp}_\varepsilon(\rho_2)\}$$

is attained for a unique couple (u_0,v_0) (up to the trivial tranformation$(u,v)\mapsto(u+a,v-a)$).
In particular we have $u_0\in L^\infty(\rho_1)$ and $v_0\in L^\infty(\rho_2)$;

Lemma

There exist $u^*\in L^{\exp}_\varepsilon(\rho_1)$ and $v^*\in L^{\exp}_\varepsilon(\rho_2)$ such that $D_\varepsilon(u,v)\le D_\varepsilon(u^*,v^*)$ with $\|v^*\|_\infty,\|u^*\|_\infty\le 3\|c\|_\infty/2$.
Moreover, there exist $a\in\mathbb{R}$ such that $u^*=(v+a)^{(c,\varepsilon)}$ and $v^*=(u^*)^{(c,\varepsilon)}$.

uOttawa

Shannon Entropy-Regularized Optimal Transport
A duality approach

Theorem (S. Di Marino & AG, 2020)

The supremum

$$\sup\left\{D_\varepsilon(u,v)\ :\ u\in L^{\exp}_\varepsilon(\rho_1), v\in L^{\exp}_\varepsilon(\rho_2)\right\}$$

is attained for a unique couple (u_0,v_0) (up to the trivial tranformation $(u,v)\mapsto(u+a,v-a)$).
In particular we have $u_0\in L^\infty(\rho_1)$ and $v_0\in L^\infty(\rho_2)$;

Lemma

There exist $u^*\in L^{\exp}_\varepsilon(\rho_1)$ and $v^*\in L^{\exp}_\varepsilon(\rho_2)$ such that $D_\varepsilon(u,v)\leq D_\varepsilon(u^*,v^*)$ with $\|v^*\|_\infty,\|u^*\|_\infty\leq 3\|c\|_\infty/2$.
Moreover, there exist $a\in\mathbb{R}$ such that $u^*=(v+a)^{(c,\varepsilon)}$ and $v^*=(u^*)^{(c,\varepsilon)}$.

Proof idea: Let $(u_n)_{n\in\mathbb{N}}\subset L^{\exp}_\varepsilon(\rho_1)$ and $(v_n)_{n\in\mathbb{N}}\subset L^{\exp}_\varepsilon(\rho_2)$ be maximizing sequences.
We can suppose that $u_n\in L^\infty(\rho_1)$, $v_n\in L^\infty(\rho_2)$ and $\|u_n\|_\infty,\|v_n\|_\infty\leq\frac{3}{2}\|c\|_\infty$.
By Banach-Alaoglu theorem there exists subsequences $u_{n_k}\rightharpoonup\overline{u}$ and $v_{n_k}\rightharpoonup\overline{v}$.

Then,

$$\sup_{u,v}D_\varepsilon(u,v)=\lim_{n\to\infty}D_\varepsilon(u_n,v_n)\leq\lim_{n\to\infty}\left\{\int_X u_n d\rho_1+\int_Y v_n d\rho_2\right\}-\varepsilon\liminf_{n\to\infty}\left\{\int_{X\times Y}e^{\frac{u_n+v_n-c}{\varepsilon}}d(\rho_1\otimes\rho_2)\right\}$$

$$\leq\int_X\overline{u}d\rho_1+\int_Y\overline{v}d\rho_2-\varepsilon\int_{X\times Y}e^{\frac{\overline{u}+\overline{v}-c}{\varepsilon}}d(\rho_1\otimes\rho_2)=D(\overline{u},\overline{v}).$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

Let $\varepsilon > 0$ be a positive number, (X, d_X) and (Y, d_Y) be Polish metric spaces, $c : X \times Y \to \mathbb{R}$ be a cost function (not necessarily bounded), $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then for every $\gamma \in \Pi(\rho_1, \rho_2)$, $u \in L^{\exp}_{\varepsilon}(\rho_1)$ and $v \in L^{\exp}_{\varepsilon}(\rho_2)$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_{\varepsilon}(u, v) + \varepsilon, \qquad \text{with equality iff } \gamma = e^{\frac{u+v-c}{\varepsilon}}.$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

Let $\varepsilon > 0$ be a positive number, (X, d_X) and (Y, d_Y) be Polish metric spaces, $c : X \times Y \to \mathbb{R}$ be a cost function (not necessarily bounded), $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then for every $\gamma \in \Pi(\rho_1, \rho_2)$, $u \in L^{\exp}_{\varepsilon}(\rho_1)$ and $v \in L^{\exp}_{\varepsilon}(\rho_2)$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_{\varepsilon}(u, v) + \varepsilon, \qquad \text{with equality iff } \gamma = e^{\frac{u+v-c}{\varepsilon}}.$$

Proof idea:

$$\begin{aligned}
\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) &= \int_{X \times Y} c d\gamma + \varepsilon \int_{X \times Y} \gamma \log \gamma d\left(\rho_1 \otimes \rho_2\right) \\
&= \int_{X \times Y} (c + \varepsilon \log \gamma - u - v) \cdot \gamma d\rho_1 \otimes \rho_2 + \int_X u d\rho_1 + \int_Y v d\rho_2
\end{aligned}$$

uOttawa

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

Let $\varepsilon > 0$ be a positive number, (X, d_X) and (Y, d_Y) be Polish metric spaces, $c : X \times Y \to \mathbb{R}$ be a cost function (not necessarily bounded), $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then for every $\gamma \in \Pi(\rho_1, \rho_2)$, $u \in L^{\exp}_{\varepsilon}(\rho_1)$ and $v \in L^{\exp}_{\varepsilon}(\rho_2)$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon, \qquad \text{with equality iff } \gamma = e^{\frac{u+v-c}{\varepsilon}}.$$

Proof idea:

$$\begin{aligned}
\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) &= \int_{X\times Y} c\, d\gamma + \varepsilon \int_{X\times Y} \gamma \log \gamma\, d\left(\rho_1 \otimes \rho_2\right) \\
&= \int_{X\times Y} (c + \varepsilon \log \gamma - u - v) \cdot \gamma\, d\rho_1 \otimes \rho_2 + \int_X u\, d\rho_1 + \int_Y v\, d\rho_2 \\
&= \int_X u\, d\rho_1 + \int_Y v\, d\rho_2 + \int_{X\times Y} (\varepsilon \log \gamma + c - u - v) \cdot \gamma\, d\left(\rho_1 \otimes \rho_2\right) \\
&\geq \int_X u\, d\rho_1 + \int_Y v\, d\rho_2 - \varepsilon \int_{X\times Y} e^{\frac{u+v-c}{\varepsilon}}\, d\left(\rho_1 \otimes \rho_2\right) + \varepsilon \\
&= D_\varepsilon(u, v) + \varepsilon,
\end{aligned}$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

Let $\varepsilon > 0$ be a positive number, (X, d_X) and (Y, d_Y) be Polish metric spaces, $c : X \times Y \to \mathbb{R}$ be a cost function (not necessarily bounded), $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then for every $\gamma \in \Pi(\rho_1, \rho_2)$, $u \in L^{\exp}_{\varepsilon}(\rho_1)$ and $v \in L^{\exp}_{\varepsilon}(\rho_2)$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon, \qquad \text{with equality iff } \gamma = e^{\frac{u+v-c}{\varepsilon}}.$$

Proof idea:

$$\begin{aligned}
\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) &= \int_{X\times Y} c d\gamma + \varepsilon \int_{X\times Y} \gamma \log \gamma d\left(\rho_1 \otimes \rho_2\right) \\
&= \int_{X\times Y} (c + \varepsilon \log \gamma - u - v) \cdot \gamma d\rho_1 \otimes \rho_2 + \int_X u d\rho_1 + \int_Y v d\rho_2 \\
&= \int_X u d\rho_1 + \int_Y v d\rho_2 + \int_{X\times Y} (\varepsilon \log \gamma + c - u - v) \cdot \gamma d\left(\rho_1 \otimes \rho_2\right) \\
&\geq \int_X u d\rho_1 + \int_Y v d\rho_2 - \varepsilon \int_{X\times Y} e^{\frac{u+v-c}{\varepsilon}} d\left(\rho_1 \otimes \rho_2\right) + \varepsilon \\
&= D_\varepsilon(u, v) + \varepsilon,
\end{aligned}$$

$ts + \varepsilon t \ln t - \varepsilon \geq -\varepsilon e^{-s/\varepsilon}$
with equality if $t = e^{-s/\varepsilon}$

uOttawa

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

Shannon Entropy-Regularized Optimal Transport

A duality approach

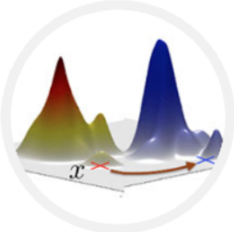
Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$$1 \implies 2$$



Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$1 \implies 2$

We have $D_\varepsilon(u^*, (u^*)^{(c,\varepsilon)}) \geq D_\varepsilon(u^*, v^*)$;

However, by the maximality of u^*, v^* we have also $D_\varepsilon(u^*, (u^*)^{(c,\varepsilon)}) \leq D_\varepsilon(u^*, v^*)$;

So we conclude that $D_\varepsilon(u^*, (u^*)^{(c,\varepsilon)}) = D_\varepsilon(u^*, v^*)$.

Therefore $v^* = (u^*)^{(c,\varepsilon)}$.

uOttawa

Shannon Entropy-Regularized Optimal Transport
A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$$2 \implies 3$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$2 \implies 3$ For every $u \in L_\varepsilon^{\exp}(\rho_1)$ and $v \in L_\varepsilon^{\exp}(\rho_2)$ we have

$$(\pi_1)_\sharp(e^{(u+v-c)/\varepsilon}\rho_1 \otimes \rho_2) = e^{(u-v^{(c,\varepsilon)})/\varepsilon}\rho_1$$

$$(\pi_2)_\sharp(e^{(u+v-c)/\varepsilon}\rho_1 \otimes \rho_2) = e^{(v-u^{(c,\varepsilon)})/\varepsilon}\rho_2$$

So if we assume 2 then we have $\gamma^* = e^{(u^*+v^*-c)/\varepsilon} \in \Pi(\rho_1, \rho_2)$.

Shannon Entropy-Regularized Optimal Transport
A duality approach

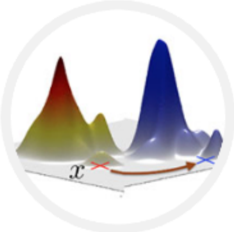
Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$3 \implies 4$



Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$3 \implies 4$ Since $\gamma^* \in \Pi(\rho_1, \rho_2)$, we have

$$\varepsilon \mathrm{KL}(\gamma^* | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon \qquad \forall\, u \in L^{\exp}_\varepsilon(\rho_1), v \in L^{\exp}_\varepsilon(\rho_2)$$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u^*, v^*) + \varepsilon \qquad \forall\, \gamma \in \Pi(\rho_1, \rho_2).$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$3 \implies 4$ Since $\gamma^* \in \Pi(\rho_1, \rho_2)$, we have

$$\varepsilon \mathrm{KL}(\gamma^* | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon \qquad \forall\, u \in L^{\exp}_\varepsilon(\rho_1), v \in L^{\exp}_\varepsilon(\rho_2)$$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u^*, v^*) + \varepsilon \qquad \forall\, \gamma \in \Pi(\rho_1, \rho_2).$$

Since by definition $\gamma^* = e^{u^*/\varepsilon} e^{v^*/\varepsilon} e^{-c/\varepsilon}$, one also has $\varepsilon \mathrm{KL}(\gamma^* | e^{-c/\varepsilon}) \geq D_\varepsilon(u^*, v^*) + \varepsilon$.

Shannon Entropy-Regularized Optimal Transport
A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$3 \implies 4$ Since $\gamma^* \in \Pi(\rho_1, \rho_2)$, we have

$$\varepsilon \mathrm{KL}(\gamma^* | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon \qquad \forall\, u \in L^{\exp}_\varepsilon(\rho_1), v \in L^{\exp}_\varepsilon(\rho_2)$$

$$\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u^*, v^*) + \varepsilon \qquad \forall\, \gamma \in \Pi(\rho_1, \rho_2).$$

Since by definition $\gamma^* = e^{u^*/\varepsilon} e^{v^*/\varepsilon} e^{-c/\varepsilon}$, one also has $\varepsilon \mathrm{KL}(\gamma^* | e^{-c/\varepsilon}) \geq D_\varepsilon(u^*, v^*) + \varepsilon$.

Putting everything together we get

$$\varepsilon \mathrm{KL}(\gamma | \mathcal{K}) \geq D_\varepsilon(u^*, v^*) + \varepsilon = \varepsilon \mathrm{KL}(\gamma^* | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon;$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$4 \implies 1$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$4 \implies 1$ Since $\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon$, with equality iff $\gamma = e^{\frac{u+v-c}{\varepsilon}}$.

minimizing in γ we find that $\mathrm{OT}_\epsilon(\rho_1, \rho_2) \geq D_\varepsilon(u, v) + \varepsilon \qquad \forall\, u \in L^{\exp}_\varepsilon(\rho_1),\ v \in L^{\exp}_\varepsilon(\rho_2)$;

uOttawa

Shannon Entropy-Regularized Optimal Transport
A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

Proof idea: $1 \implies 2 \implies 3 \implies 4 \implies 1$

$4 \implies 1$ Since $\varepsilon \mathrm{KL}(\gamma | e^{-c/\varepsilon}) \geq D_\varepsilon(u, v) + \varepsilon$, with equality iff $\gamma = e^{\frac{u+v-c}{\varepsilon}}$.

minimizing in γ we find that $\mathrm{OT}_\epsilon(\rho_1, \rho_2) \geq D_\varepsilon(u, v) + \varepsilon \qquad \forall\, u \in L^{\exp}_\varepsilon(\rho_1),\ v \in L^{\exp}_\varepsilon(\rho_2)$;

using that by hypotesis $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$, we get that

$$D_\varepsilon(u^*, v^*) \geq D_\varepsilon(u, v) \qquad \forall\, u \in L^{\exp}_\varepsilon(\rho_1),\ v \in L^{\exp}_\varepsilon(\rho_2).$$

Shannon Entropy-Regularized Optimal Transport

A duality approach

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

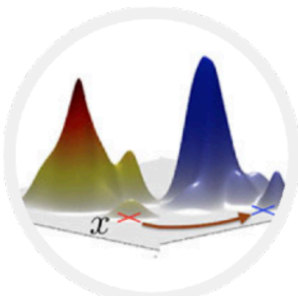
Remark (the role of the reference measure)

Let $(X, d_X, \mathfrak{m}_1)$, $(Y, d_Y, \mathfrak{m}_2)$ be Polish metric measure spaces, $c : X \times Y \to \mathbb{R}$
$\rho_1 \in \mathcal{P}(X)$ and $\rho_2 \in \mathcal{P}(Y)$ be probability measures such that $\mathrm{KL}(\rho_1|\mathfrak{m}_1) + \mathrm{KL}(\rho_2|\mathfrak{m}_2) < \infty$.

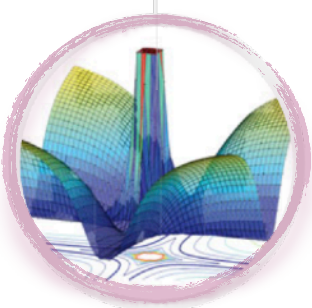
$$
\begin{aligned}
S_\varepsilon(\rho_1, \rho_2; \mathfrak{m}_1, \mathfrak{m}_2) :&= \min_{\gamma \in \Pi(\rho_1, \rho_2)} \left\{ \int_{X \times Y} c \, d\gamma + \varepsilon \, \mathrm{KL}(\gamma | \mathfrak{m}_1 \otimes \mathfrak{m}_2) \right\} \\
&= \mathrm{OT}_\varepsilon(\rho_1, \rho_2) + \varepsilon \, \mathrm{KL}(\rho_1|\mathfrak{m}_1) + \varepsilon \, \mathrm{KL}(\rho_2|\mathfrak{m}_2)
\end{aligned}
$$

Dual potentials: $u_0 - \varepsilon \log \rho_1 \in L^\infty(\mathfrak{m}_1)$ and $v_0 - \varepsilon \log \rho_2 \in L^\infty(\mathfrak{m}_2)$.

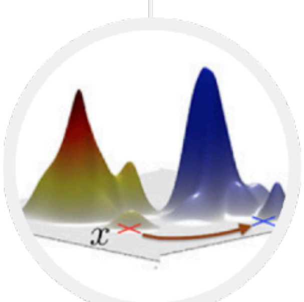
Outline



Duality approach for the Shannon Entropy-Regularized Optimal Transport



Convergence proof of the Sinkhorn algorithm (variational)



Applications

Shannon entropy regularized OT and Sinkhorn algorithm

$$\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = \min_{\gamma \in \Pi(\rho_1,\rho_2)} \varepsilon \int_{X \times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1 \otimes \rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma \in \Pi(\rho_1,\rho_2)} \int_{X \times Y} c(x,y)\, d\gamma(x,y) + \varepsilon \int_{X \times Y} \gamma(x,y) \log \gamma(x,y)\, d(\rho_1 \otimes \rho_2)\,.$$

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x) b^\varepsilon(y) e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

[E. Schrödinger, 1932]

$$\begin{cases} a^\varepsilon(x) \displaystyle\int_Y b^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_2(y) = 1 \\ b^\varepsilon(y) \displaystyle\int_X a^\varepsilon(y) e^{-c(x,y)/\varepsilon} d\rho_1(x) = 1 \end{cases}$$

Shannon entropy regularized OT and Sinkhorn algorithm

$$\mathrm{OT}_\varepsilon(\rho_1,\rho_2) = \min_{\gamma\in\Pi(\rho_1,\rho_2)} \varepsilon \int_{X\times Y} \gamma \log\left(\frac{\gamma}{e^{-c/\varepsilon}}\right) d(\rho_1\otimes\rho_2) \qquad \varepsilon > 0$$

$$= \min_{\gamma\in\Pi(\rho_1,\rho_2)} \int_{X\times Y} c(x,y)d\gamma(x,y) + \varepsilon \int_{X\times Y} \gamma(x,y)\log\gamma(x,y)d\left(\rho_1\otimes\rho_2\right).$$

There exists a unique minimizer γ^ε *(strictly convex)* $\gamma^\varepsilon(x,y) = a^\varepsilon(x)b^\varepsilon(y)e^{-c(x,y)/\varepsilon}$

If and only if solves the Schrödinger system

[E. Schrödinger, 1932]

$$\begin{cases} a^\varepsilon(x)\displaystyle\int_Y b^\varepsilon(y)e^{-c(x,y)/\varepsilon}d\rho_2(y) = 1 \\ b^\varepsilon(y)\displaystyle\int_X a^\varepsilon(y)e^{-c(x,y)/\varepsilon}d\rho_1(x) = 1 \end{cases}$$

The algorithm: $a^0(x) = b^0(y) = 1$

Define the sequences (iteratively) $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ by

$$b^n(y) = \frac{1}{\int e^{-c(x,y)/\varepsilon}a^{n-1}(x)d\rho_1(x)},\ a^n(x) = \frac{1}{\int e^{-c(x,y)/\varepsilon}b^n(y)d\rho_2(y)}.$$

Shannon entropy regularized OT and Sinkhorn algorithm

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

uOttawa

Shannon entropy regularized OT and Sinkhorn algorithm

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

Lemma

There exist $u^* \in L^{\exp}_\varepsilon(\rho_1)$ and $v^* \in L^{\exp}_\varepsilon(\rho_2)$ such that $D_\varepsilon(u, v) \leq D_\varepsilon(u^*, v^*)$ with $\|v^*\|_\infty, \|u^*\|_\infty \leq 3\|c\|_\infty/2$.
Moreover, there exist $a \in \mathbb{R}$ such that $u^* = (v + a)^{(c,\varepsilon)}$ and $v^* = (u^*)^{(c,\varepsilon)}$.

Shannon entropy regularized OT and Sinkhorn algorithm

Lemma: Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\rho_1 \in \mathcal{P}(X)$, $\rho_2 \in \mathcal{P}(Y)$ be probability measures. Then

$$D_\varepsilon(u, u^{(c,\varepsilon)}) \geq D_\varepsilon(u, v) \qquad \forall v \in L^{\exp}_\varepsilon(\rho_2),$$
$$D_\varepsilon(u, u^{(c,\varepsilon)}) = D_\varepsilon(u, v) \text{ if and only if } v = u^{(c,\varepsilon)}.$$

Lemma

There exist $u^* \in L^{\exp}_\varepsilon(\rho_1)$ and $v^* \in L^{\exp}_\varepsilon(\rho_2)$ such that $D_\varepsilon(u, v) \leq D_\varepsilon(u^*, v^*)$ with $\|v^*\|_\infty, \|u^*\|_\infty \leq 3\|c\|_\infty/2$.
Moreover, there exist $a \in \mathbb{R}$ such that $u^* = (v + a)^{(c,\varepsilon)}$ and $v^* = (u^*)^{(c,\varepsilon)}$.

Theorem (S. Di Marino & AG, 2020)

The following are equivalent:

1. *(Maximizers)* u^* and v^* are maximizing potentials for dual problem;
2. *(Maximality condition)* $\mathcal{F}_1^{(c,\varepsilon)}(u^*) = v^*$ and $\mathcal{F}_2^{(c,\varepsilon)}(v^*) = u^*$;
3. *(Schrödinger system)* let $\gamma^* = e^{(u^*(x)+v^*(y)-c(x,y))/\varepsilon} \cdot \rho_1 \otimes \rho_2$, then $\gamma^* \in \Pi(\rho_1, \rho_2)$;
4. *(Duality attainement)* $\mathrm{OT}_\varepsilon(\rho_1, \rho_2) = D_\varepsilon(u^*, v^*) + \varepsilon$.

uOttawa

Convergence proof

Sinkhorn algorithm aka Iterative Proportional fitting procedure (IPFP)

Define the sequences (iteratively) $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ by

$$b^n(y) = \frac{1}{\int e^{-c(x,y)/\varepsilon} a^{n-1}(x) d\rho_1(x)},$$

$$a^n(x) = \frac{1}{\int e^{-c(x,y)/\varepsilon} b^n(y) d\rho_2(y)}.$$

Theorem

If $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ are the IPFP sequences defined above, then there exists $(\lambda^n)_{n\in\mathbb{N}}, \lambda^n \in \mathbb{R}$ such that

$a^n/\lambda^n \to a$ in $L^p(\rho_1)$ and $\lambda^n b^n \to b$ in $L^p(\rho_2)$, $1 \le p < +\infty$, where (a, b) solve the Schrödinger problem.

In particular, the sequence $\gamma^n = a^n b^n e^{-c/\varepsilon}$, converges in $L^p(\rho_1 \otimes \rho_2)$ to γ_ε for any $1 \le p < +\infty$.

Convergence proof

Sinkhorn algorithm aka Iterative Proportional fitting procedure (IPFP)

Define the sequences (iteratively) $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ by

$$b^n(y) = \frac{1}{\int e^{-c(x,y)/\varepsilon} a^{n-1}(x) d\rho_1(x)},$$

$$a^n(x) = \frac{1}{\int e^{-c(x,y)/\varepsilon} b^n(y) d\rho_2(y)}.$$

Theorem

If $(a^n)_{n\in\mathbb{N}}$ and $(b^n)_{n\in\mathbb{N}}$ are the IPFP sequences defined above, then there exists $(\lambda^n)_{n\in\mathbb{N}}, \lambda^n \in \mathbb{R}$ such that

$a^n/\lambda^n \to a$ in $L^p(\rho_1)$ and $\lambda^n b^n \to b$ in $L^p(\rho_2)$, $1 \le p < +\infty$, where (a,b) solve the Schrödinger problem.

In particular, the sequence $\gamma^n = a^n b^n e^{-c/\varepsilon}$, converges in $L^p(\rho_1 \otimes \rho_2)$ to γ_ε for any $1 \le p < +\infty$.

Proof idea: Let us write $a^n = e^{u_n/\varepsilon}$, $b^n = e^{v_n/\varepsilon}$

$$\begin{cases} v_{2n+1} = (u_{2n})^{(c,\varepsilon)} \\ u_{2n+1} = u_{2n} \end{cases}, \quad \begin{cases} v_{2n+2} = v_{2n+1} \\ u_{2n+2} = (v_{2n+1})^{(c,\varepsilon)} \end{cases}.$$

$\exists\ \ell_n \in \mathbb{R}$ such that
$\|u_n - \ell_n\|_\infty, \|v_n + \ell_n\| \le \frac{3}{2}\|c\|_\infty$

$$D_\varepsilon(u_n, v_n) \le D_\varepsilon(u_{n+1}, v_{n+1}) \le \cdots \le \mathrm{OT}_\varepsilon(\rho_1, \rho_2) - \varepsilon.$$

$u_n - \ell_n$ and $v_n + \ell_n$ are precompact in every L^p, for $1 \le p < \infty$

One obtain

$$D_\varepsilon(v^{(c,\varepsilon)}, v) - D_\varepsilon(u, v) = \lim_{n_k\to\infty} D_\varepsilon(u_{n_k+1}, v_{n_k+1}) - D_\varepsilon(u_{n_k}, v_{n_k}) = 0.$$

$$D_\varepsilon(u, u^{(c,\varepsilon)}) - D_\varepsilon(u, v) = \lim_{n_k\to\infty} D_\varepsilon(u_{n_k+2}, v_{n_k+2}) - D_\varepsilon(u_{n_k}, v_{n_k}) = 0.$$

uOttawa

Robustness of the Sinkhorn algorithm

Proposition (Continuity with respect to the marginals)

The functional $\mathrm{OT}_\varepsilon(\rho_1, \rho_2)$ is continuous: if $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ are sequences weakly converging respectively to ρ_1 and ρ_2, then the corresponding Kantorovich potentials $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ converges uniformly in L^∞ to (u^*, v^*).

Robustness of the Sinkhorn algorithm

Proposition (Continuity with respect to the marginals)

The functional $\mathrm{OT}_\varepsilon(\rho_1, \rho_2)$ is continuous: if $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ are sequences weakly converging respectively to ρ_1 and ρ_2, then the corresponding Kantorovich potentials $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ converges uniformly in L^∞ to (u^*, v^*).

Proof idea:

Robustness of the Sinkhorn algorithm

Proposition (Continuity with respect to the marginals)

The functional $\mathrm{OT}_\varepsilon(\rho_1, \rho_2)$ is continuous: if $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ are sequences weakly converging respectively to ρ_1 and ρ_2, then the corresponding Kantorovich potentials $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ converges uniformly in L^∞ to (u^*, v^*).

Proof idea:

Consider sequences $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ weakly converging to ρ_1 and ρ_2.

For each $n \in \mathbb{N}$ consider the couple of optimal potentials (u_n, v_n).

WLOG assume that one is the $u_n = (v_n)^{(c,\varepsilon)}$ and $v_n = (u_n)^{(c,\varepsilon)}$.

Robustness of the Sinkhorn algorithm

Proposition (Continuity with respect to the marginals)

The functional $\mathrm{OT}_\varepsilon(\rho_1, \rho_2)$ is continuous: if $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ are sequences weakly converging respectively to ρ_1 and ρ_2, then the corresponding Kantorovich potentials $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ converges uniformly in L^∞ to (u^*, v^*).

Proof idea:

Consider sequences $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ weakly converging to ρ_1 and ρ_2.

For each $n \in \mathbb{N}$ consider the couple of optimal potentials (u_n, v_n).

WLOG assume that one is the $u_n = (v_n)^{(c,\varepsilon)}$ and $v_n = (u_n)^{(c,\varepsilon)}$.

The potentials u_n and v_n are bounded for all $n \in \mathbb{N}$ and then, by Banach-Alaoglu theorem, there exists subsequences $(u_{n_k})_{n_k\in\mathbb{N}}$ and $(v_{n_k})_{n_k\in\mathbb{N}}$ such that $u_{n_k} \to u^*$ and $v_{n_k} \to v^*$ (unif.).

Robustness of the Sinkhorn algorithm

Proposition (Continuity with respect to the marginals)

The functional $\mathrm{OT}_\varepsilon(\rho_1, \rho_2)$ is continuous: if $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ are sequences weakly converging respectively to ρ_1 and ρ_2, then the corresponding Kantorovich potentials $(u_n)_{n\in\mathbb{N}}$ and $(v_n)_{n\in\mathbb{N}}$ converges uniformly in L^∞ to (u^*, v^*).

Proof idea:

Consider sequences $(\rho_1^n)_{n\in\mathbb{N}} \subset \mathcal{P}(X)$ and $(\rho_2^n)_{n\in\mathbb{N}} \subset \mathcal{P}(Y)$ weakly converging to ρ_1 and ρ_2.
For each $n \in \mathbb{N}$ consider the couple of optimal potentials (u_n, v_n).
WLOG assume that one is the $u_n = (v_n)^{(c,\varepsilon)}$ and $v_n = (u_n)^{(c,\varepsilon)}$.

The potentials u_n and v_n are bounded for all $n \in \mathbb{N}$ and then, by Banach-Alaoglu theorem, there exists subsequences $(u_{n_k})_{n_k\in\mathbb{N}}$ and $(v_{n_k})_{n_k\in\mathbb{N}}$ such that $u_{n_k} \to u^*$ and $v_{n_k} \to v^*$ (unif.).

Similarly, one can show that (u^*, v^*) is a maximizer couple for ρ_1 and ρ_2.

uOttawa

Wasserstein distance as a loss function

$$\min_{\nu \in \mathcal{P}(X)} \mathrm{OT}_\varepsilon(\nu, \rho_1)$$

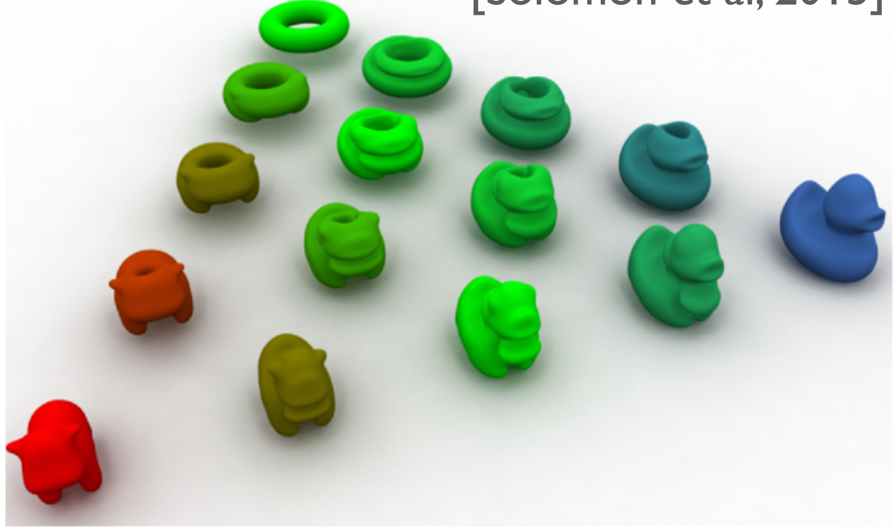
Wasserstein distance as a loss function

$$\min_{\nu \in \mathcal{P}(X)} \mathrm{OT}_\varepsilon(\nu, \rho_1)$$

Wasserstein Barycenters $\rho_1, \cdots, \rho_N$ be probability measures in X and $\lambda_1, \ldots, \lambda_N \geq 0$ such that $\sum_{j=1}^N \lambda_j = 1$.

$$\min_{\nu \in \mathcal{P}(X)} \sum_{i=1}^N \lambda_i \mathrm{OT}_\varepsilon(\nu, \rho_i)$$

[Solomon et al, 2015]



Wasserstein distance as a loss function

$$\min_{\nu \in \mathcal{P}(X)} \mathrm{OT}_\varepsilon(\nu, \rho_1)$$

How to compute "$\nabla \mathrm{OT}_\varepsilon(\nu, \rho_1)$" ?

uOttawa

Wasserstein distance as a loss function

$$\min_{\nu \in \mathcal{P}(X)} \mathrm{OT}_\varepsilon(\nu, \rho)$$

How to compute "$\nabla \mathrm{OT}_\varepsilon(\nu, \rho)$" ?

Theorem

Let (X, d_X), (Y, d_Y) be Polish spaces, $c : X \times Y \to \mathbb{R}$ be a Borel bounded cost, $\nu \in \mathcal{P}(X)$, $\rho \in \mathcal{P}(Y)$ be probability measures.The Fréchet differential of $\mathrm{OT}_\varepsilon(\nu, \rho)$ is given by

$$\nabla \mathrm{OT}_\varepsilon(\nu, \rho) := \left(\frac{\delta \mathrm{OT}_\varepsilon(\nu, \rho)}{\delta \nu}, \frac{\delta \mathrm{OT}_\varepsilon(\nu, \rho)}{\delta \rho} \right) = (u^*, v^*)\,,$$

where (u^*, v^*) are Kantorovich potentials such that $u^* = (v^*)^{(c,\varepsilon)}$

uOttawa

Wasserstein distance as a loss function

Proof idea:

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u_t d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u_t(x) - v_t(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) +$$
$$+ \int_Y v_t d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u_t(x) - v_t(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\limsup_{t\to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u_t d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u_t(x) - v_t(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) +$$
$$+ \int_Y v_t d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u_t(x) - v_t(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

since $u_t \to u$ and $v_t \to v$ uniformly.

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\limsup_{t \to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u d\chi - \varepsilon \int_{X \times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + \\ + \int_Y v d\alpha - \varepsilon \int_{X \times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

since $u_t \to u$ and $v_t \to v$ uniformly.

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\limsup_{t\to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + \\ + \int_Y v d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Since (u, v) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\limsup_{t\to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + \int_Y v d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Since (u, v) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$

$$\frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \geq \int_X u_t d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u_t(x) - v_t(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + \int_Y v_t d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u_t(x) - v_t(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\limsup_{t\to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u\, d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + \int_Y v\, d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Since (u, v) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$

$$\liminf_{t\to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \geq \int_X u\, d\chi - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + \int_Y v\, d\alpha - \varepsilon \int_{X\times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Wasserstein distance as a loss function

Proof idea: Let $\nu^t = \nu + t\chi$, $\rho^t = \rho + t\alpha$ and

(u, v) (resp. (u_t, v_t)) the optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$ (resp. $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$).

Using that (u_t, v_t) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu^t, \rho^t)$.

$$\limsup_{t \to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \leq \int_X u d\chi - \varepsilon \int_{X \times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + $$
$$+ \int_Y v d\alpha - \varepsilon \int_{X \times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Since (u, v) are optimal potentials for $\mathrm{OT}_\varepsilon(\nu, \rho)$

$$\liminf_{t \to 0^+} \frac{\mathrm{OT}_\varepsilon(\nu^t, \rho^t) - \mathrm{OT}_\varepsilon(\nu, \rho)}{t} \geq \int_X u d\chi - \varepsilon \int_{X \times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\chi \otimes \rho) + $$
$$+ \int_Y v d\alpha - \varepsilon \int_{X \times Y} \exp\left(\frac{u(x) - v(y) - c(x,y)}{\varepsilon}\right) d(\nu \otimes \alpha).$$

Combining both inequalities we have $\dfrac{\delta \mathrm{OT}_\varepsilon}{\delta \nu}(\nu, \rho) = u$ and $\dfrac{\delta \mathrm{OT}_\varepsilon}{\delta \rho}(\nu, \rho) = v.$

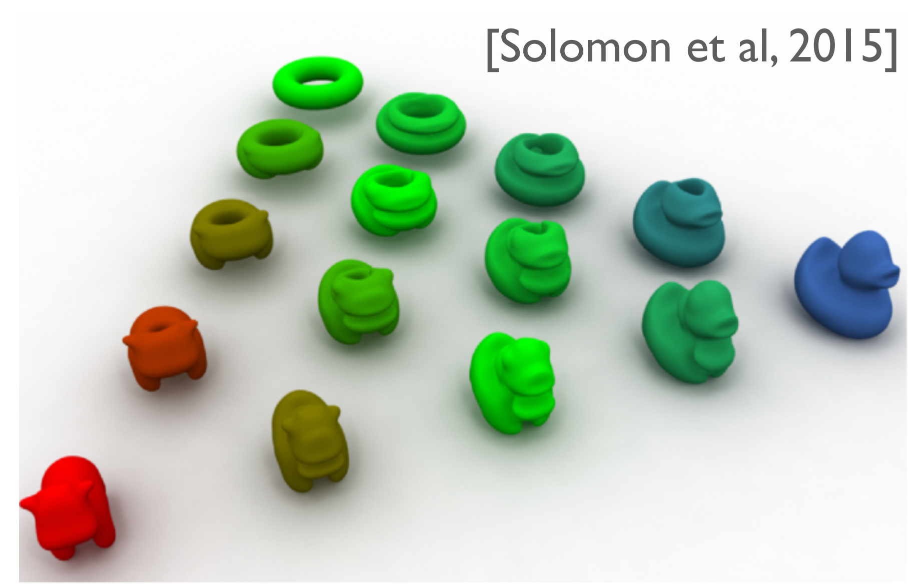
Wasserstein distance as a loss function

$$\min_{\nu \in \mathcal{P}(X)} \mathrm{OT}_{\varepsilon}(\nu, \rho)$$

Wasserstein Barycenters

$\rho_1, \cdots, \rho_N$ be probability measures in X and $\lambda_1, \ldots, \lambda_N \geq 0$ such that $\sum_{j=1}^{N} \lambda_j = 1$.

$$\min_{\nu \in \mathcal{P}(X)} \sum_{i=1}^{N} \lambda_i \mathrm{OT}_{\varepsilon}(\nu, \rho_i)$$



[Solomon et al, 2015]