Regularized Optimal Transport is Ground Cost Adversarial

MokaMeeting

May 11, 2022

François-Pierre Paty
francoispierrepaty.github.io

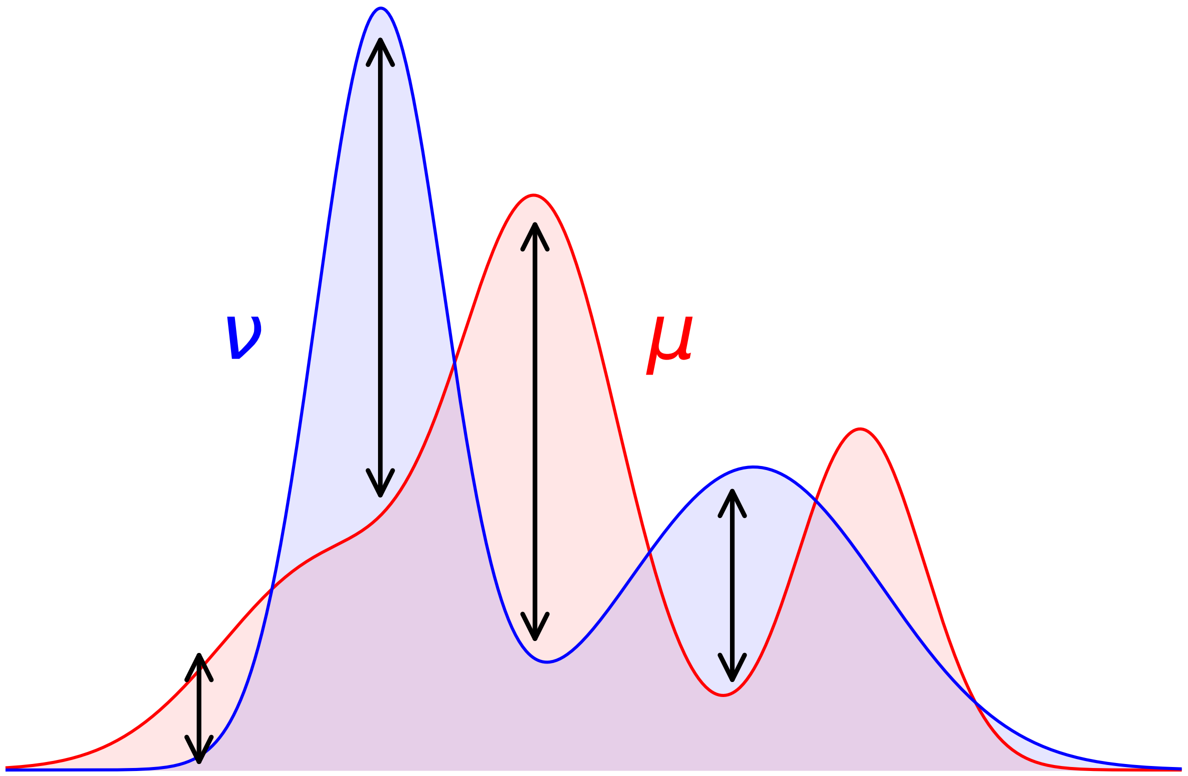
Based on a joint work with ***Marco Cuturi***

COMPARING DISTRIBUTIONS

1. Vertical comparison

Look at the difference, or the ratio of the densities

e.g. Total Variation distance, Kullback Leibler divergence, etc.

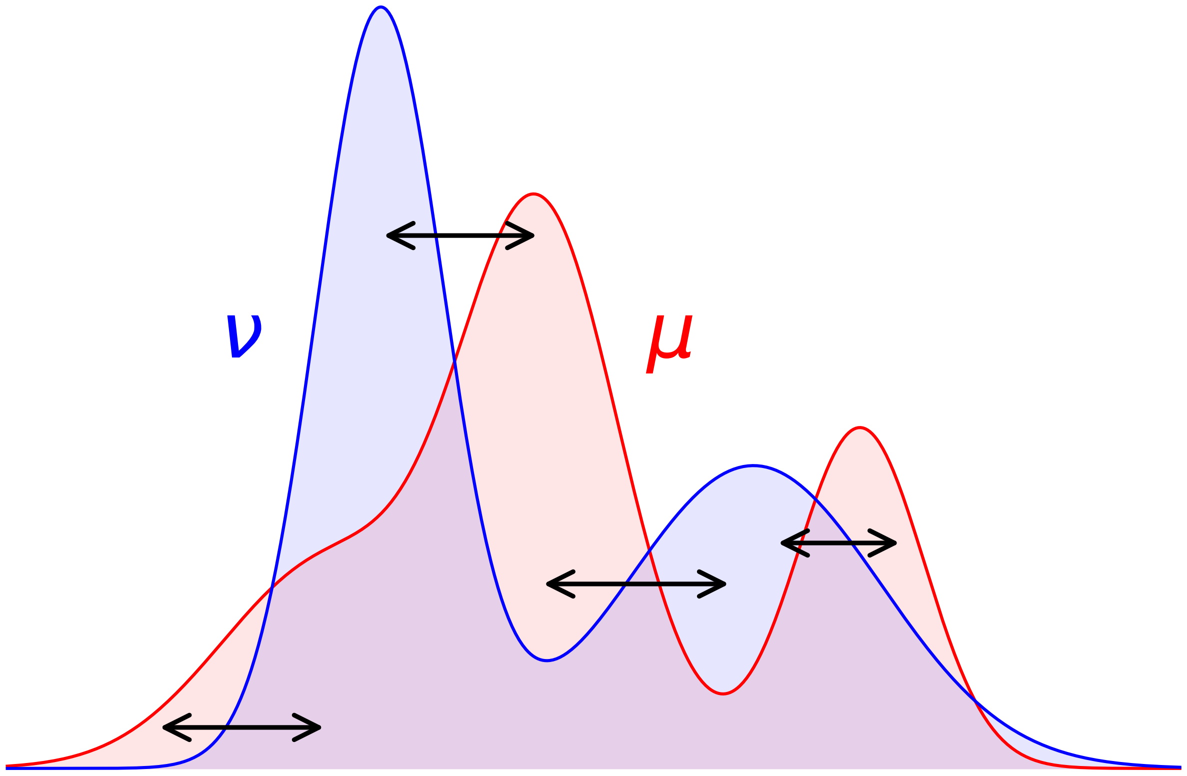


Comparing Distributions

2. Horizontal comparison aka Optimal Transport

Move the mass across the ground space

⚠️ *Need for a notion of **displacement cost** on the ground space*



OPTIMAL TRANSPORT

Leonid Kantorovich

Optimal Transport

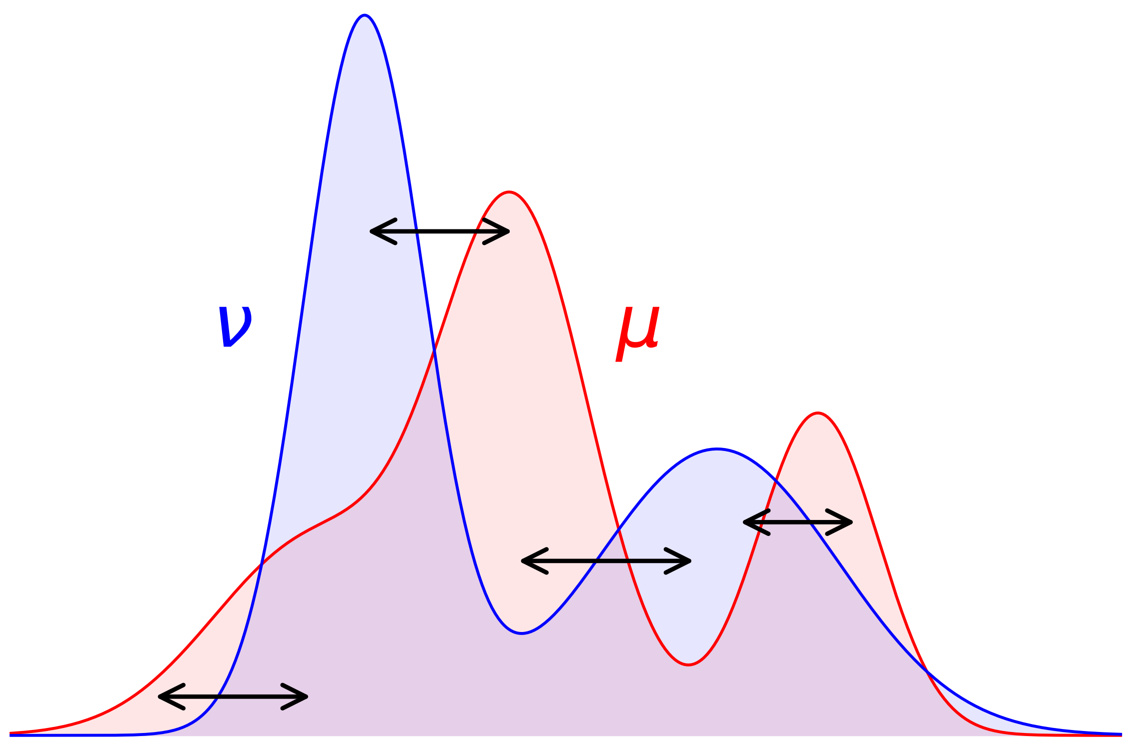
Data:

Two distributions μ and ν over $\mathbb{R}^d$

Parameter:

A (countinuous) cost function

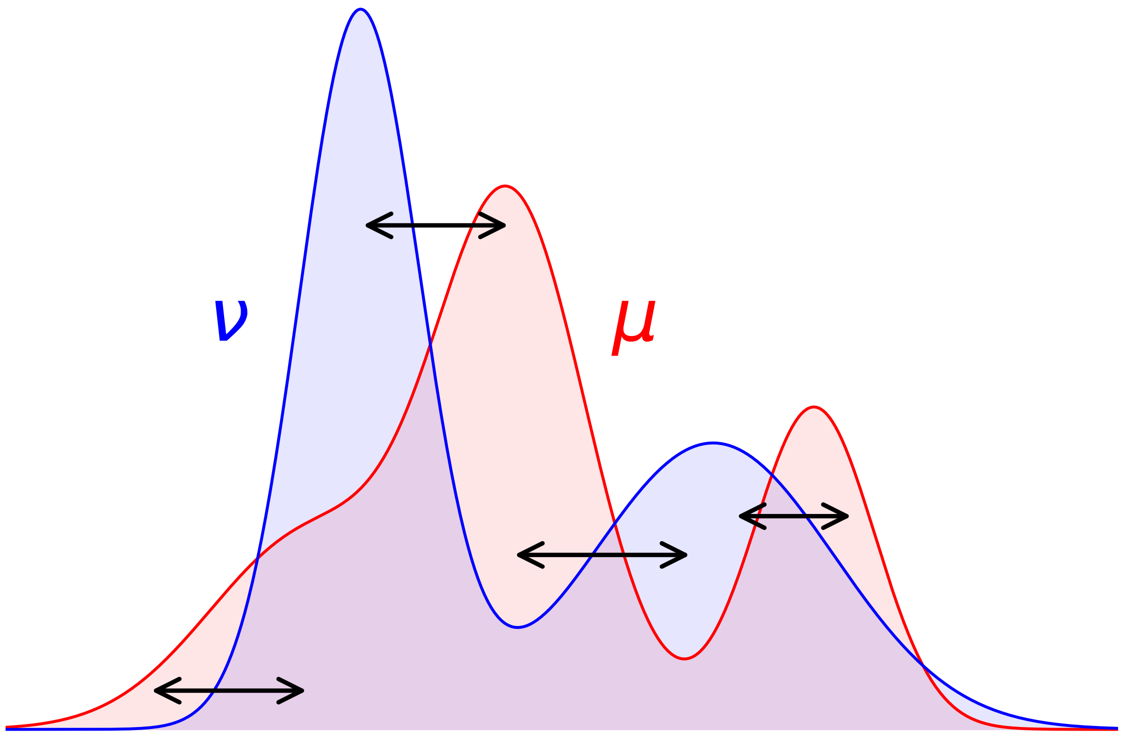
$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$



Optimal Transport

Data:

Two distributions μ and ν over $\mathbb{R}^d$



Parameter:

A (countinuous) cost function

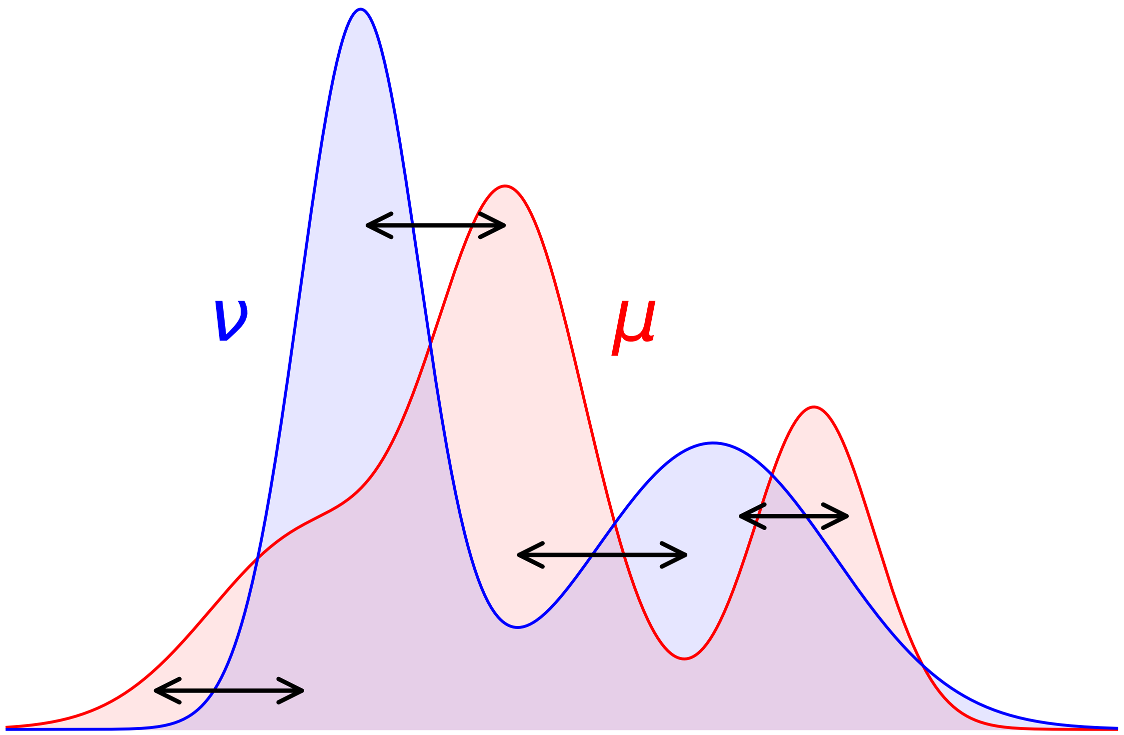
$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

Definition of Optimal Transport (OT):

Optimal Transport

Data:

Two distributions μ and ν over $\mathbb{R}^d$



Parameter:

A (countinuous) cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

Definition of Optimal Transport (OT):

$$c(x, y)\, d\pi(x, y)$$

Optimal Transport

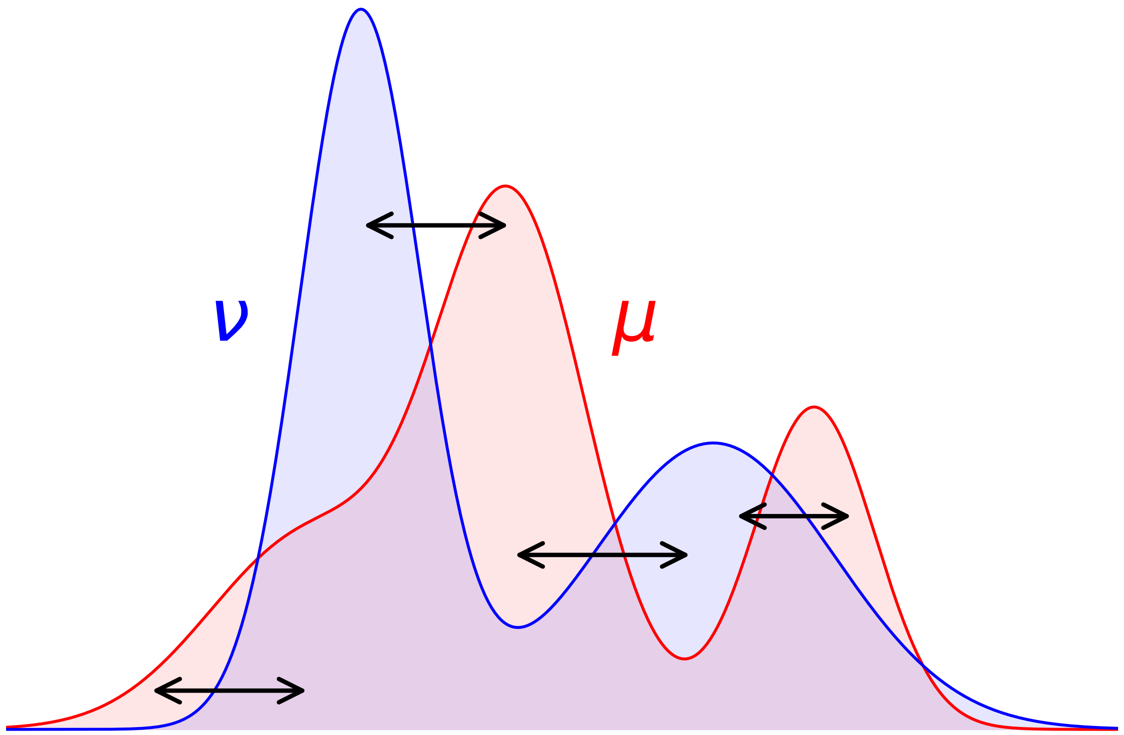
Data:

Two distributions μ and ν over $\mathbb{R}^d$

Parameter:

A (countinuous) cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$



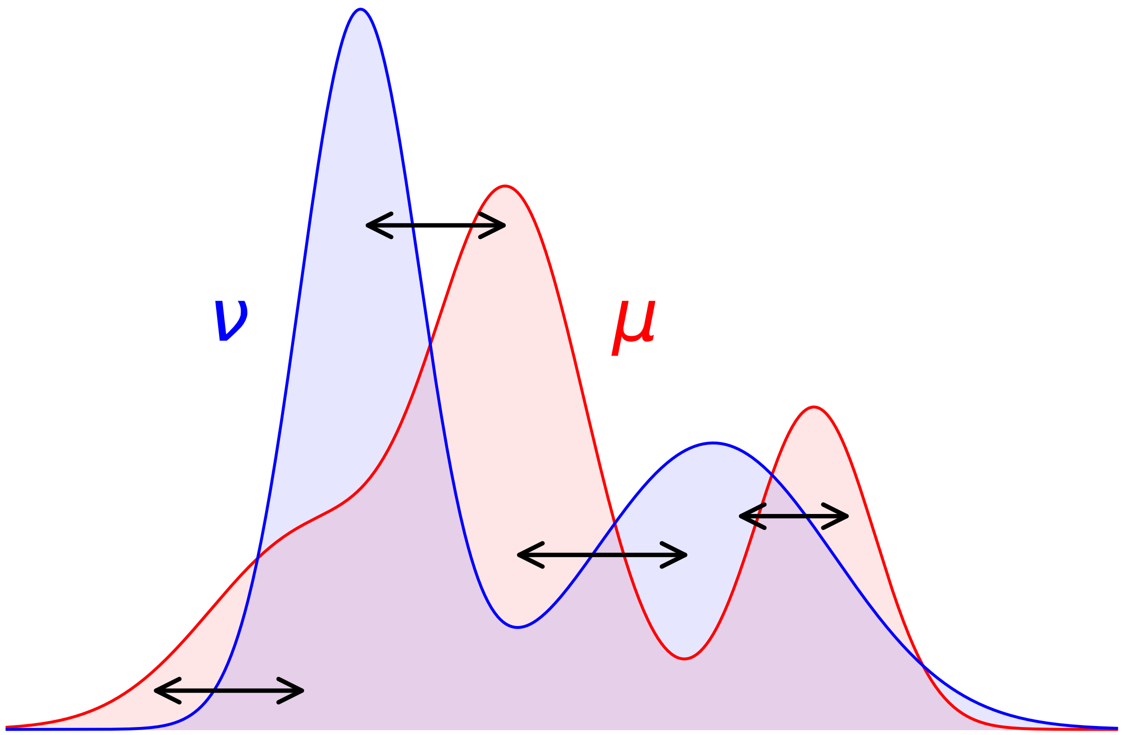
Definition of Optimal Transport (OT):

$$\iint c(x, y)\, d\pi(x, y)$$

Optimal Transport

Data:

Two distributions μ and ν over $\mathbb{R}^d$



Parameter:

A (countinuous) cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

Definition of Optimal Transport (OT):

$$\inf_{\pi} \iint c(x, y) \, d\pi(x, y)$$

Optimal Transport

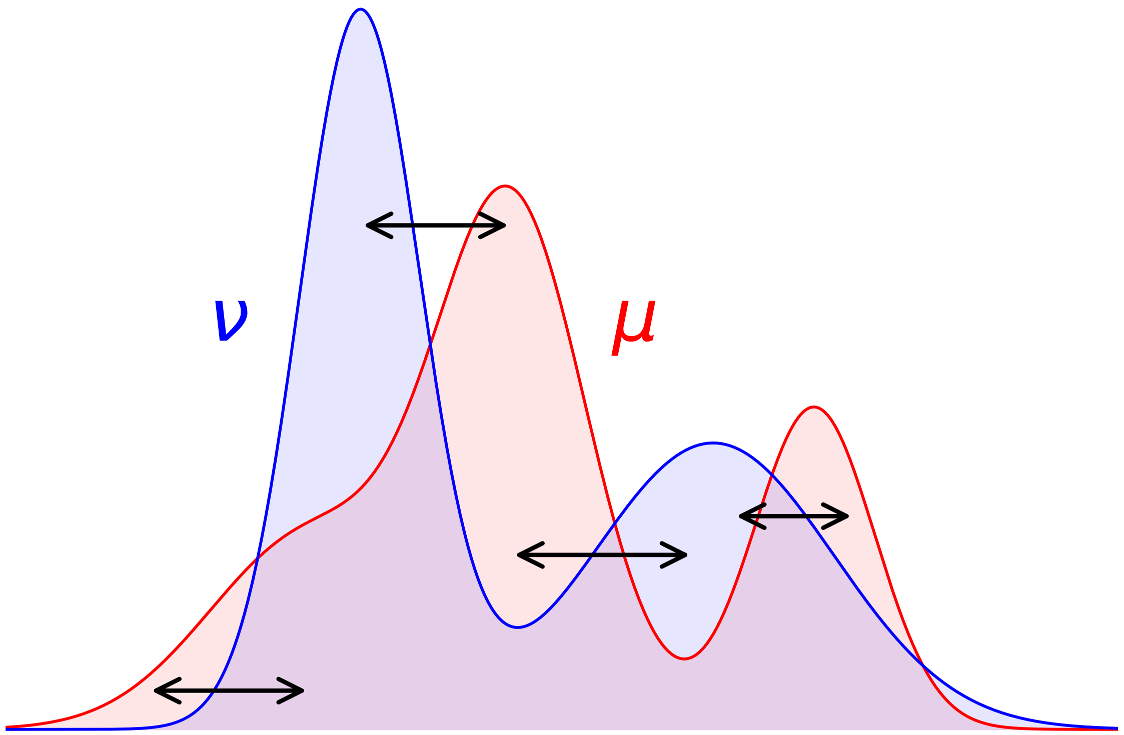
Data:

Two distributions μ and ν over $\mathbb{R}^d$

Parameter:

A (countinuous) cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$



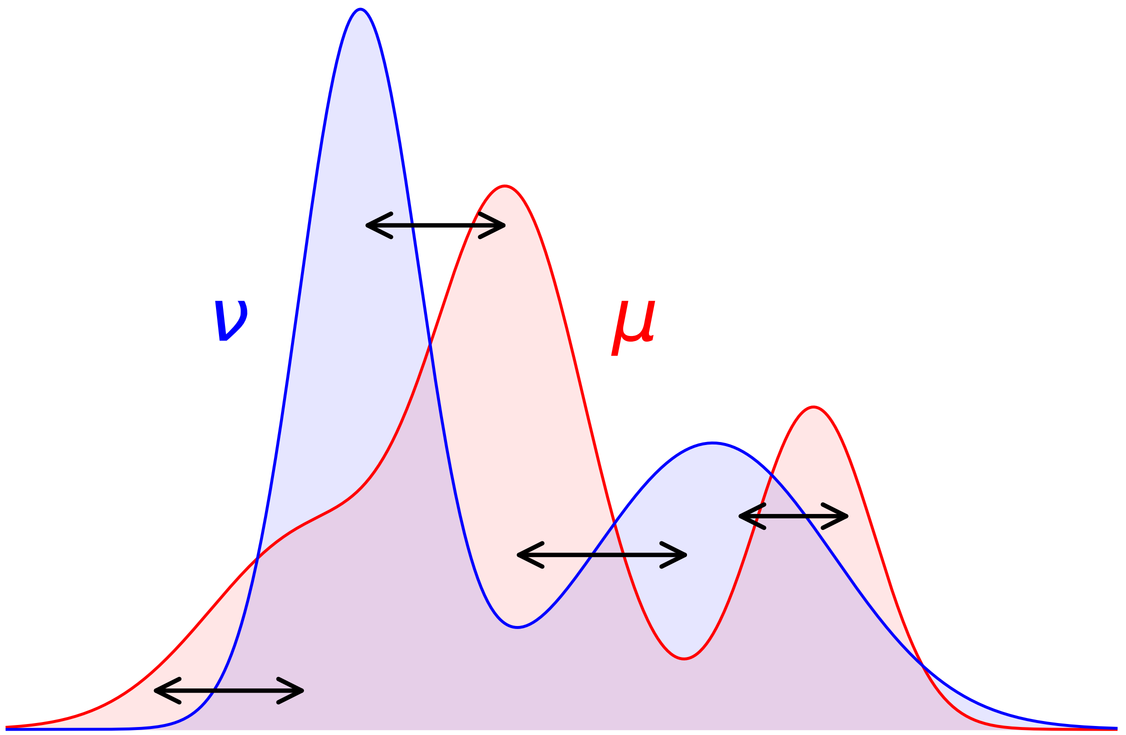
Definition of Optimal Transport (OT):

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

Optimal Transport

Data:

Two distributions μ and ν over $\mathbb{R}^d$



Parameter:

A (countinuous) cost function

$$c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

Definition of Optimal Transport (OT):

$$\mathscr{T}_c(\mu, \nu) = \inf_{\pi} \iint c(x, y)\, d\pi(x, y)$$

over all π such that

$$\begin{cases} \int d\pi(x,y) = d\mu(x) \ \ \forall x \\ \int d\pi(x,y) = d\nu(y) \ \ \forall y \end{cases}$$

Optimal Transport

Two main questions in practice

Optimal Transport

Two main questions in practice

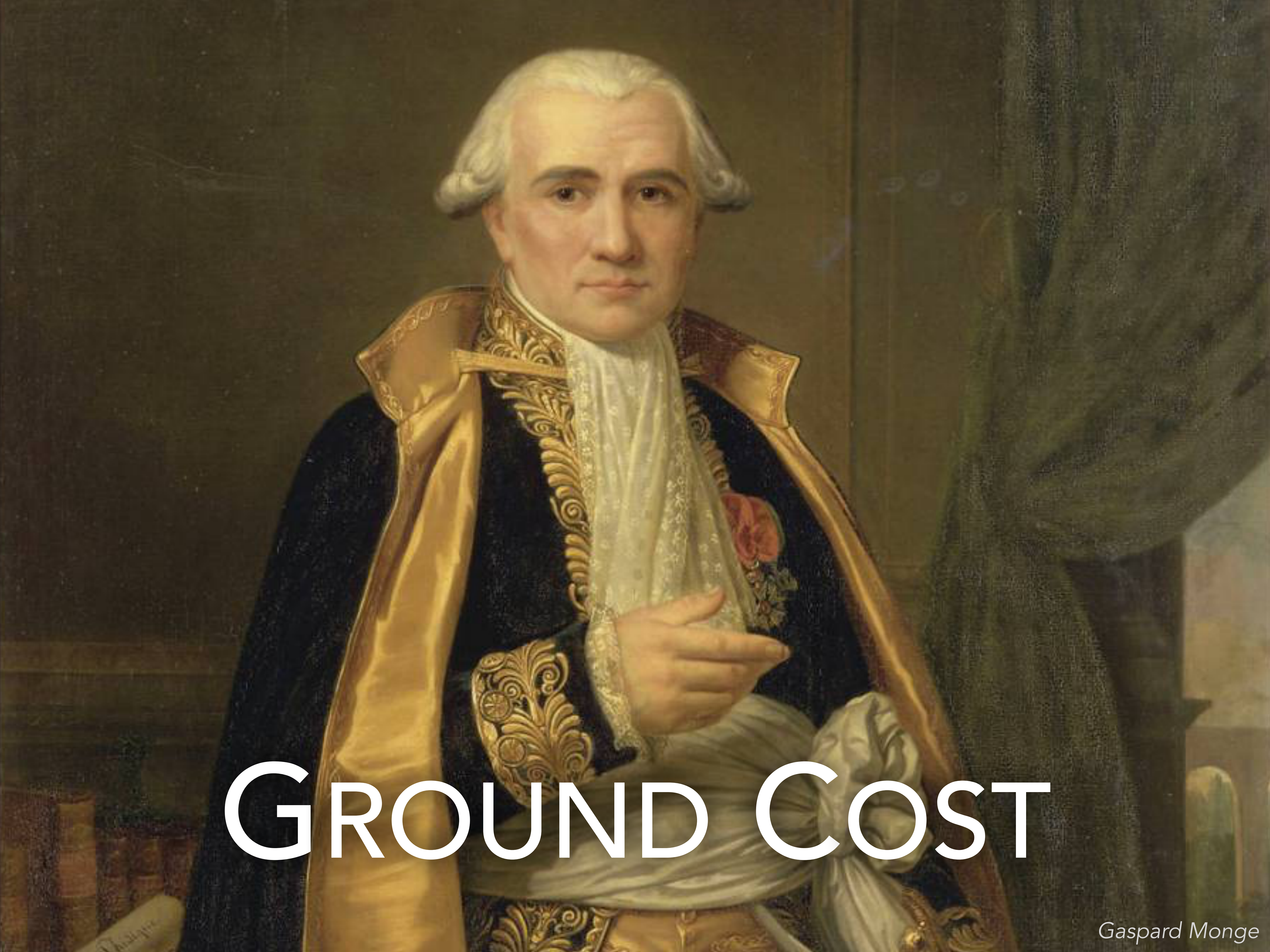
1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

OPTIMAL TRANSPORT

Two main questions in practice

1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

2. How to compute/approximate the OT cost $\mathscr{T}_c(\mu, \nu)$, at least when the measures are discrete *(i.e. are finite sums of Dirac masses) in a scalable way*?


GROUND COST
Gaspard Monge

Ground Cost

1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

Ground Cost

1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

1. Monge initially proposed $c(x, y) = \|x - y\|$

Ground Cost

1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

1. Monge initially proposed $c(x, y) = \|x - y\|$
2. This was generalized to cost functions of the form

$$c(x, y) = \|x - y\|^p \quad \text{where} \quad p \geq 1$$

(in this case, we say that $\mathscr{T}_c^{1/p}$ *is the p-Wasserstein distance)*

GROUND COST

1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

1. Monge initially proposed $c(x, y) = \|x - y\|$
2. This was generalized to cost functions of the form

$$c(x, y) = \|x - y\|^p \quad \text{where} \quad p \geq 1$$

(in this case, we say that $\mathscr{T}_c{}^{1/p}$ *is the p-Wasserstein distance)*

But does it make sense when the ground space is high-dimensional ?

Ground Cost

1. How to choose the ground cost c in a way that makes sense for the data distributions μ and ν ?

1. Monge initially proposed $c(x, y) = \|x - y\|$
2. This was generalized to cost functions of the form

$$c(x, y) = \|x - y\|^p \quad \text{where} \quad p \geq 1$$

(in this case, we say that $\mathscr{T}_c{}^{1/p}$ is the p-Wasserstein distance)

But does it make sense when the ground space is high-dimensional ?

But does it make sense when the data lives on a low-dimensional manifold ?

Ground Cost

Idea: Find a ground cost c that is **adversarial**, *i.e.* that best separates the two distributions by maximizing the OT cost

$$\max_{c \in \mathscr{C}} \mathscr{T}_c(\mu, \nu)$$ where $\mathscr{C}$ is a convex class of functions

Ground Cost

Idea: Find a ground cost c that is **adversarial**, *i.e.* that best separates the two distributions by maximizing the OT cost

$$\max_{c \in \mathscr{C}} \mathscr{T}_c(\mu, \nu)$$ where $\mathscr{C}$ is a convex class of functions

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c)$$ for some convex f

$$f(c) = \begin{cases} 0 & \text{if } c \in \mathscr{C} \\ +\infty & \text{if } c \notin \mathscr{C} \end{cases}$$

Ground Cost

Idea: Find a ground cost c that is **adversarial**, *i.e.* that best separates the two distributions by maximizing the OT cost

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c) \quad \text{for some convex } f$$

- Links with the Robust Optimization literature
- Links with the matchings literature in Economics
- Initially proposed by Genevay *et al.* in 2017 to learn generative models
- When $\mathscr{C}$ is the set of Mahalanobis distances, it defines the Subspace Robust Wasserstein distances (ICML 2019, cf. in a few slides)

In Economics

Data: Two probability distributions μ and ν representing two groups of people (*e.g.* men and women), and a matching between them π_0 (*e.g.* marriage/dating data)

Problem: Explain/understand the observed matching π_0

Method: Assume π_0 is optimal for a certain ground-cost $c_\star$, which we can then interpret. We just have to solve:

$$\sup_c \mathscr{T}_c(\mu, \nu) - \int c \, d\pi_0$$

In practice, economists assume that

$$c_\star \in \left\{ d_\Omega^2 : (x, y) \mapsto (x - y)^\top \Omega (x - y) \,|\, \Omega \succeq 0, \|\Omega\| \leq 1 \right\}$$

and seek Ω, *i.e.* rather solve

$$\sup_c \mathscr{T}_c(\mu, \nu) - \int c \, d\pi_0 - R^*(c)$$


REGULARIZATION
Alessio Figalli

Regularization

2. How to compute/approximate the OT cost $\mathcal{T}_c(\mu, \nu)$?

1. This is a Linear Program $\longrightarrow \mathcal{O}(n^3)$ complexity

2. Entropic regularization $\longrightarrow \mathcal{O}(n^2)$ Sinkhorn algorithm, GPU-friendly, differentiable…

$$\inf_\pi \iint c(x, y)\, d\pi(x, y) + \varepsilon R(\pi)$$

where $R(\pi) = \mathrm{KL}(\pi || \mu \otimes \nu)$

Other regularizations have been proposed: *e.g.* quadratic, group-lasso, capacity constraints, with different algorithms and effects on the OT plan / value

REGULARIZATION

2. How to compute/approximate the OT cost $\mathscr{T}_c(\mu, \nu)$?

1. This is a Linear Program $\longrightarrow \mathcal{O}(n^3)$ complexity

2. Entropic regularization $\longrightarrow \mathcal{O}(n^2)$ Sinkhorn algorithm, GPU-friendly, differentiable…

$$\inf_{\pi} \iint c(x, y)\, d\pi(x, y) + \varepsilon R(\pi)$$

How can we interpret the effect of the regularization ?

Two views of the same phenomenon

Cédric Villani

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c) \quad = \max_{c} \min_{\pi \in \Pi(\mu, \nu)} \int c \, d\pi - f(c)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_{c}(\mu, \nu) - f(c) \quad = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathcal{T}_{c}(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^{*}(\pi)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_c(\mu,\nu) - f(c) = \max_{c} \min_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi\in\Pi(\mu,\nu)} \max_{c} \int c\, d\pi - f(c)$$

$$= \min_{\pi\in\Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c - c_0}{\varepsilon}\right)$ where R is convex:

Ground-cost Adversarial Transport

$$\max_{c} \mathcal{T}_c(\mu,\nu) - f(c) = \max_{c} \min_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi\in\Pi(\mu,\nu)} \max_{c} \int c\, d\pi - f(c)$$

$$= \min_{\pi\in\Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^*\left(\frac{c - c_0}{\varepsilon}\right)$ where R is convex:

$$f^*(\pi) = \sup_{c} \int c\, d\pi - \varepsilon R^*\left(\frac{c-c_0}{\varepsilon}\right)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$ where R is convex:

$$f^*(\pi) = \sup_{c} \int c \, d\pi - \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right) = \sup_{d} \int (c_0 + \varepsilon d) \, d\pi - \varepsilon R^*(d)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathcal{T}_c(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$ where R is convex:

$$f^*(\pi) = \sup_{c} \int c \, d\pi - \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right) = \sup_{d} \int (c_0 + \varepsilon d) \, d\pi - \varepsilon R^*(d)$$

$$= \int c_0 \, d\pi + \varepsilon \sup_{d} \int d \, d\pi - R^*(d)$$

Ground-cost Adversarial Transport

$$\max_c \mathscr{T}_c(\mu, \nu) - f(c) = \max_c \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_c \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$ where R is convex:

$$f^*(\pi) = \sup_c \int c \, d\pi - \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right) = \sup_d \int (c_0 + \varepsilon d) \, d\pi - \varepsilon R^*(d)$$

$$= \int c_0 \, d\pi + \varepsilon \sup_d \int d \, d\pi - R^*(d)$$

$$= \int c_0 \, d\pi + \varepsilon R^{**}(\pi)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathcal{T}_c(\mu,\nu) - f(c) = \max_{c} \min_{\pi\in\Pi(\mu,\nu)} \int c\,d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi\in\Pi(\mu,\nu)} \max_{c} \int c\,d\pi - f(c)$$

$$= \min_{\pi\in\Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^*\left(\frac{c-c_0}{\varepsilon}\right)$ where R is convex:

$$f^*(\pi) = \sup_c \int c\,d\pi - \varepsilon R^*\left(\frac{c-c_0}{\varepsilon}\right) = \sup_d \int (c_0+\varepsilon d)\,d\pi - \varepsilon R^*(d)$$

$$= \int c_0\,d\pi + \varepsilon \sup_d \int d\,d\pi - R^*(d)$$

$$= \int c_0\,d\pi + \varepsilon R^{**}(\pi) \quad = \int c_0\,d\pi + \varepsilon R(\pi)$$

Ground-cost Adversarial Transport

$$\max_{c} \mathscr{T}_c(\mu, \nu) - f(c) = \max_{c} \min_{\pi \in \Pi(\mu,\nu)} \int c \, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi \in \Pi(\mu,\nu)} \max_{c} \int c \, d\pi - f(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$ where R is convex:

Ground-cost Adversarial Transport

$$\max_{c} \mathcal{T}_c(\mu,\nu) - f(c) = \max_{c} \min_{\pi\in\Pi(\mu,\nu)} \int c\, d\pi - f(c)$$

Sion's minimax theorem

$$= \min_{\pi\in\Pi(\mu,\nu)} \max_{c} \int c\, d\pi - f(c)$$

$$= \min_{\pi\in\Pi(\mu,\nu)} f^*(\pi)$$

Take $f(c) = \varepsilon R^*\left(\frac{c-c_0}{\varepsilon}\right)$ where R is convex:

$$\inf_{\pi\in\Pi(\mu,\nu)} \iint c_0(x,y)\, d\pi(x,y) + \varepsilon R(\pi)$$

$$= \sup_{c} \mathcal{T}_c(\mu,\nu) - \varepsilon R^*\left(\frac{c-c_0}{\varepsilon}\right)$$

Ground Cost Robustness ⇔ Regularization

Theorem: Regularized OT is ground cost adversarial in the following sense

$$\inf_{\pi \in \Pi(\mu,\nu)} \iint c_0(x,y)\, d\pi(x,y) + \varepsilon R(\pi)$$
$$= \sup_{c} \mathscr{T}_c(\mu,\nu) - \varepsilon R^* \left(\frac{c - c_0}{\varepsilon} \right)$$

where R is a convex regularizer

and R^* is the convex conjugate of R:

$$R^*(c) = \sup_{\pi} \int c\, d\pi - R(\pi)$$

Examples: Entropic OT

$$\min_{\pi \in \Pi(\mu,\nu)} \int c_0 \, d\pi + \varepsilon \mathrm{KL}(\pi \| \mu \otimes \nu)$$

Examples: Entropic OT

$$\min_{\pi \in \Pi(\mu,\nu)} \int c_0 \, d\pi + \varepsilon \mathrm{KL}(\pi \| \mu \otimes \nu)$$

$$= \sup_{c} \mathscr{T}_c(\mu, \nu) - \varepsilon \int \exp\left(\frac{c - c_0}{\varepsilon} \right) \, d\mu \otimes \nu + \varepsilon$$

Examples: Subspace Robust Wasserstein

$$\mathcal{S}_k^2(\mu, \nu) = \min_{\pi \in \Pi(\mu, \nu)} \sum_{l=1}^{k} \lambda_l \left(\iint (x - y)(x - y)^\top \, d\pi(x, y) \right)$$

Examples: Subspace Robust Wasserstein

l^{th} largest eigenvalue

$$\mathcal{S}_k^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \sum_{l=1}^{k} \lambda_l \left(\iint (x-y)(x-y)^\top \, d\pi(x,y) \right)$$

Examples: Subspace Robust Wasserstein

l^{th} largest eigenvalue

$$\mathcal{S}_k^2(\mu,\nu) = \min_{\pi\in\Pi(\mu,\nu)} \underbrace{\sum_{l=1}^{k} \lambda_l \left(\iint (x-y)(x-y)^\top \, d\pi(x,y) \right)}_{\textit{convex function of } \pi}$$

Examples: Subspace Robust Wasserstein

l^{th} largest eigenvalue

$$\mathcal{S}_k^2(\mu, \nu) = \min_{\pi \in \Pi(\mu,\nu)} \underbrace{\sum_{l=1}^{k} \lambda_l \left(\iint (x-y)(x-y)^\top d\pi(x,y) \right)}_{\textit{convex function of } \pi}$$

$$= \max_{\substack{0 \preceq \Omega \preceq I \\ \mathrm{trace}(\Omega)=k}} \mathscr{T}_{d_\Omega^2}(\mu, \nu)$$

Examples: Subspace Robust Wasserstein

l^{th} largest eigenvalue

$$\mathcal{S}_k^2(\mu,\nu) = \min_{\pi \in \Pi(\mu,\nu)} \underbrace{\sum_{l=1}^{k} \lambda_l \left(\iint (x-y)(x-y)^\top d\pi(x,y) \right)}_{\text{convex function of } \pi}$$

$$= \max_{\substack{0 \preceq \Omega \preceq I \\ \text{trace}(\Omega)=k}} \mathcal{T}_{d_\Omega^2}(\mu,\nu)$$

Where $d_\Omega^2(x,y) = (x-y)^\top \Omega (x-y)$ *is the squared Mahalanobis distance*

Examples: in Economics

$$\sup_{c} \mathscr{T}_c(\mu, \nu) - \int c \, d\pi_0 - R^*(c)$$

Examples: in Economics

$$R^*(c) = \iota\left(\exists \Omega \succeq 0, \|\Omega\| \leq 1, c = d_\Omega^2\right)$$

$$\sup_{c} \mathscr{T}_c(\mu, \nu) - \int c \, d\pi_0 - R^*(c)$$

Examples: in Economics

$$R^*(c) = \iota\left(\exists \Omega \succeq 0, \|\Omega\| \leq 1, c = d_\Omega^2\right)$$

$$\sup_{c} \mathscr{T}_c(\mu, \nu) - \int c \, d\pi_0 - R^*(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} R(\pi - \pi_0)$$

Examples: in Economics

$$R^*(c) = \iota\left(\exists \Omega \succeq 0, \|\Omega\| \leq 1, c = d_\Omega^2\right)$$

$$\sup_{c} \mathscr{T}_c(\mu, \nu) - \int c \, d\pi_0 - R^*(c)$$

$$= \min_{\pi \in \Pi(\mu,\nu)} R(\pi - \pi_0)$$

Is the adversarial cost $c_\star$ an interesting dissimilarity measure on the ground space ?

A small detour: Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) \quad = \max_{c} \mathscr{T}_c(\mu, \nu) - f^*(c)$$

A small detour: Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{c} \mathscr{T}_c(\mu,\nu) - f^*(c)$$

$$= \sup_{c} \max_{\phi \oplus \psi \leq c} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c)$$

A small detour: Duality

$$\begin{aligned}\min_{\pi \in \Pi(\mu,\nu)} f(\pi) &= \max_{c} \mathscr{T}_c(\mu,\nu) - f^*(c) \\ &= \sup_{c} \max_{\phi \oplus \psi \leq c} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) \\ &= \sup_{c} \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) - \iota(\phi \oplus \psi \leq c)\end{aligned}$$

A small detour: Duality

$$\begin{aligned}
\min_{\pi \in \Pi(\mu,\nu)} f(\pi) &= \max_{c} \mathscr{T}_c(\mu,\nu) - f^*(c) \\
&= \sup_{c} \max_{\phi \oplus \psi \leq c} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) \\
&= \sup_{c} \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) - \iota(\phi \oplus \psi \leq c) \\
&= \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - \inf_{c} f^*(c) + \iota(\phi \oplus \psi \leq c)
\end{aligned}$$

A small detour: Duality

$$
\begin{aligned}
\min_{\pi \in \Pi(\mu,\nu)} f(\pi) &= \max_{c} \mathscr{T}_c(\mu,\nu) - f^*(c) \\
&= \sup_{c} \max_{\phi \oplus \psi \leq c} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) \\
&= \sup_{c} \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) - \iota(\phi \oplus \psi \leq c) \\
&= \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - \inf_{c} f^*(c) + \iota(\phi \oplus \psi \leq c) \\
&= \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - \inf_{c \geq \phi \oplus \psi} f^*(c)
\end{aligned}
$$

A small detour: Duality

$$\begin{aligned}
\min_{\pi\in\Pi(\mu,\nu)} f(\pi) &= \max_{c} \mathscr{T}_c(\mu,\nu) - f^*(c) \\
&= \sup_{c} \max_{\phi\oplus\psi\leq c} \int \phi\, d\mu + \int \psi\, d\nu - f^*(c) \\
&= \sup_{c} \max_{\phi,\psi} \int \phi\, d\mu + \int \psi\, d\nu - f^*(c) - \iota(\phi\oplus\psi\leq c) \\
&= \max_{\phi,\psi} \int \phi\, d\mu + \int \psi\, d\nu - \inf_{c} f^*(c) + \iota(\phi\oplus\psi\leq c) \\
&= \max_{\phi,\psi} \int \phi\, d\mu + \int \psi\, d\nu - \inf_{c\geq\phi\oplus\psi} f^*(c)
\end{aligned}$$

If *e.g.* f^* is increasing, $\inf_{c\geq\phi\oplus\psi} f^*(c) = f^*(\phi\oplus\psi)$ hence:

A small detour: Duality

$$
\begin{aligned}
\min_{\pi \in \Pi(\mu,\nu)} f(\pi) &= \max_{c} \mathscr{T}_c(\mu, \nu) - f^*(c) \\
&= \sup_{c} \max_{\phi \oplus \psi \leq c} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) \\
&= \sup_{c} \max_{\phi, \psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(c) - \iota(\phi \oplus \psi \leq c) \\
&= \max_{\phi, \psi} \int \phi \, d\mu + \int \psi \, d\nu - \inf_{c} f^*(c) + \iota(\phi \oplus \psi \leq c) \\
&= \max_{\phi, \psi} \int \phi \, d\mu + \int \psi \, d\nu - \inf_{c \geq \phi \oplus \psi} f^*(c)
\end{aligned}
$$

If *e.g.* f^* is increasing, $\inf_{c \geq \phi \oplus \psi} f^*(c) = f^*(\phi \oplus \psi)$ hence:

$$
\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi, \psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)
$$

Characterization of the Ground-Cost

Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)$$

Characterization of the Ground-Cost

Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)$$

$$= \max_{\phi,\psi} \mathscr{T}_{\phi \oplus \psi}(\mu, \nu) - f^*(\phi \oplus \psi)$$

Characterization of the Ground-Cost

Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)$$

$$= \max_{\phi,\psi} \mathscr{T}_{\phi\oplus\psi}(\mu,\nu) - f^*(\phi \oplus \psi)$$

$$\leq \max_{c} \mathscr{T}_{c}(\mu,\nu) - f^*(c)$$

Characterization of the Ground-Cost

Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)$$

$$= \max_{\phi,\psi} \mathscr{T}_{\phi \oplus \psi}(\mu, \nu) - f^*(\phi \oplus \psi)$$

$$\leq \max_{c} \mathscr{T}_{c}(\mu, \nu) - f^*(c)$$

Main result

Characterization of the Ground-Cost

Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)$$

$$= \max_{\phi,\psi} \mathscr{T}_{\phi \oplus \psi}(\mu, \nu) - f^*(\phi \oplus \psi)$$

$$\leq \max_{c} \mathscr{T}_{c}(\mu, \nu) - f^*(c)$$

Main result

$$= \min_{\pi \in \Pi(\mu,\nu)} f(\pi)$$

Characterization of the Ground-Cost

Duality

$$\min_{\pi \in \Pi(\mu,\nu)} f(\pi) = \max_{\phi,\psi} \int \phi \, d\mu + \int \psi \, d\nu - f^*(\phi \oplus \psi)$$

$$= \max_{\phi,\psi} \mathscr{T}_{\phi \oplus \psi}(\mu, \nu) - f^*(\phi \oplus \psi)$$

$$\leq \max_{c} \mathscr{T}_{c}(\mu, \nu) - f^*(c)$$

Main result

$$= \min_{\pi \in \Pi(\mu,\nu)} f(\pi)$$

So the inequality is an equality and there exists a separable cost function that is an optimal adversarial ground-cost

CHARACTERIZATION OF THE GROUND-COST

Is the adversarial cost $c_\star$ an interesting dissimilarity measure on the ground space ?

Short answer: In a sense, no.

Characterization of the Ground-Cost

Is the adversarial cost $c_\star$ an interesting dissimilarity measure on the ground space ?

Short answer: In a sense, no.

Theorem: Under some technical assumption on R *(verified for the entropic or quadratic regularizations)*, there exists functions ϕ and ψ such that

$$c : (x, y) \mapsto \phi(x) + \psi(y)$$

is an optimal adversarial cost, *i.e.* is solution to

$$\sup_c \mathscr{T}_c(\mu, \nu) - \varepsilon R^* \left(\frac{c - c_0}{\epsilon} \right)$$

What I could not talk about

- Restriction to nonnegative adversarial costs $\sup_{c \geq 0} \ldots$
- Extension to several measures

Thank you

francoispierrepaty.github.io