# Robust Analysis via Optimal Transport Methods

Yifan Jiang

University of Oxford

April 6, 2022

# Contents

# What is optimal transport?

## Motivation

Given two spaces $X$ and $Y$, we want to find the 'optimal' way to transport some mass under distribution $\mu$ to $\nu$ with respect to some cost function $c$.

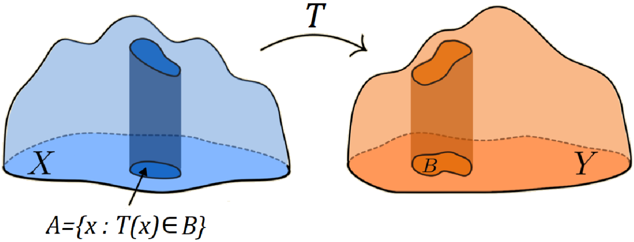


Figure: Optimal transport [Tho18].

# What is optimal transport?

## Motivation

Given two spaces $X$ and $Y$, we want to find the 'optimal' way to transport some mass under distribution $\mu$ to $\nu$ with respect to some cost function $c$.

## Assumption

- $X$ and $Y$ are Polish.
- $\mu$ and $\nu$ are probability measures on $X$ and $Y$ respectively.
- Cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is lower-semi continuous and bounded by below.

# Monge's formulation

## Definition

We say $T : X \to Y$ is a transport map from $\mu$ to $\nu$ if

$$\nu = T_{\#}\mu := \mu \circ T^{-1}.$$

## Monge's optimal transport

$$\text{Minimize} \quad \mathbb{M}(T) = \int_X c(x, T(x))\mu(\mathrm{d}x),$$

over all the transport maps $T$ such that $T_{\#}\mu = \nu$.

# Monge's formulation

## Definition

We say $T : X \to Y$ is a transport map from $\mu$ to $\nu$ if

$$\nu = T_{\#}\mu := \mu \circ T^{-1}.$$

## Monge's optimal transport

$$\text{Minimize} \quad \mathbb{M}(T) = \int_X c(x, T(x))\mu(\mathrm{d}x),$$

over all the transport maps $T$ such that $T_{\#}\mu = \nu$.

Bad news: such a $T$ may not exist! (e.g. $\mu$ is an atom.)

# Kantorovich's formulation

## Definition

Given $\mu$ and $\nu$, the set of transport plans $\Pi(\mu,\nu)$ is given by

$$\Pi(\mu,\nu) = \{\pi \in \mathcal{P}(X \times Y) : \pi_1 = \mu, \pi_2 = \nu\},$$

where $\pi_i$ are the marginals of $\pi$.

## Kantorovich's optimal transport

$$\text{Minimize} \quad \mathbb{K}(\pi) = \int_{X \times Y} c(x,y)\pi(\mathrm{d}x, \mathrm{d}y),$$

over all the transport plans $\pi \in \Pi(\mu,\nu)$.

# Kantorovich's formulation

## Definition

Given $\mu$ and $\nu$, the set of transport plans $\Pi(\mu, \nu)$ is given by

$$\Pi(\mu, \nu) = \{\pi \in \mathcal{P}(X \times Y) : \pi_1 = \mu, \pi_2 = \nu\},$$

where $\pi_i$ are the marginals of $\pi$.

## Kantorovich's optimal transport

$$\text{Minimize} \quad \mathbb{K}(\pi) = \int_{X \times Y} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y),$$

over all the transport plans $\pi \in \Pi(\mu, \nu)$.

Good news: not only is $\Pi(\mu, \nu)$ nonempty, but also the minimizer $\pi^*$ is attainable!

# Monge (1746-1818) meets Kantorovich (1912-1986)

One can easily notice that

$$\inf \mathbb{K}(\pi) \leq \inf \mathbb{M}(T).$$

# Monge (1746-1818) meets Kantorovich (1912-1986)

One can easily notice that

$$\inf \mathbb{K}(\pi) \leq \inf \mathbb{M}(T).$$

**Theorem (Brenier's Theorem [Vil09])**

*Let $X$ and $Y$ be $\mathbb{R}^d$ and $c(x,y) = |x-y|^2$. Assume $\mu$ is absolutely continuous to the Lebesgue measure, then there exists a minimizer $T^*$ for the Monge problem. Moreover,*

$$\pi^* := (\mathrm{Id}, T^*)_{\#}\mu \in \Pi(\mu, \nu)$$

*is a minimizer of the Kantorovich problem and*

$$\mathbb{K}(\pi^*) = \mathbb{M}(T^*).$$

# Casual optimal transport

## Motivation

If we have extra knowledge on the information flow, how can we find the 'optimal' transport without using the knowledge from the future?

## Assumption

- $X$ and $Y$ are Polish spaces with right continuous filtration $(\mathcal{F}_t^X)_{t\in I}$ and $(\mathcal{F}_t^Y)_{t\in I}$. Here, $I$ can be $[N]$, $\mathbb{N}$, $[0,1]$ or $\mathbb{R}_{\geq 0}$.
- $\mu$ and $\nu$ are probability measures on $X$ and $Y$.
- Cost function $c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ is lower-semi continuous and bounded by below.

# Monge's formulation

## Definition

We say $T : (X, (\mathcal{F}_t^X)) \to (Y, (\mathcal{F}_t^Y))$ is a causal transport map from $\mu$ to $\nu$ if

- $T_\# \mu = \nu$.
- $T$ is non-anticipative, i.e., for any $A \in \mathcal{F}_t^Y$ we have $T^{-1}(A) \in {}^\mu\mathcal{F}_t^X$.

## Monge's casual optimal transport

$$\text{Minimize} \quad \mathbb{M}(T) = \int_X c(x, T(x))\mu(\mathrm{d}x),$$

over all the causal transport maps $T$.

# Kantorovich's formulation

## Definition

Given $\mu$ and $\nu$, the set of causal transport plans $\Pi_c(\mu, \nu)$ is given by

$$\Pi_c(\mu, \nu) = \{\pi \in \Pi(\mu, \nu) : \theta(\cdot, A) \in {}^{\mu}\mathcal{F}_t^X \quad \text{for any } A \in \mathcal{F}_t^Y\},$$

where $\theta$ is the disintegration of $\pi$, i.e., $\pi(\mathrm{d}x, \mathrm{d}y) = \theta(x, \mathrm{d}y)\mu(\mathrm{d}x)$.

## Kantorovich's causal optimal transport

$$\text{Minimize} \quad \mathbb{K}(\pi) = \int_{x \times y} c(x, y)\pi(\mathrm{d}x, \mathrm{d}y),$$

over all the causal transport plans $\pi \in \Pi_c(\mu, \nu)$.

# Good and bad news

Is the minimizer for Kantorovich's causal optimal transport still attainable?

# Good and bad news

Is the minimizer for Kantorovich's causal optimal transport still attainable? Yes and no. We need to pay some price to ask the filtration to be compatible to the topology.

## Compactness of $\Pi_c(\mu, \nu)$

[Las18] $(\mathcal{F}_t^Y)_{t \in I}$ satisfies some extra regularity.

[ABVZ20] $\mu$ has some weak continuity with respect to $(\mathcal{F}_t^X)_{t \in I}$.

Under either of the above cases, $\Pi_c(\mu, \nu)$ is compact.

## Remark

Let $X = Y = C([0,1], \mathbb{R})$ the continuous path space equipped with natural filtration and uniform norm. Then both of the above assumptions are satisfied.

# Wasserstein distance

## Definition

Given two probability measures $\mu$, $\nu$ on (filtered) Polish space $X$, their Wasserstein distance is given by

$$\mathrm{d}(\mu,\nu) = \left\{ \inf_{\Pi(\mu,\nu)} \mathbb{K}(\pi) \right\}^{1/p},$$

and their adapted Wasserstein 'distance' is given by

$$\mathrm{d}_c(\mu,\nu) = \left\{ \inf_{\Pi_c(\mu,\nu)} \mathbb{K}(\pi) \right\}^{1/p}.$$

Here, $1 \leq p < \infty$ and the cost function has the form $c = \rho^p$ where $\rho$ is a lower-semi continuous metric on $X$.

# Wasserstein distance

## Example

Let $X = \mathbb{R} \times \mathbb{R}$ equipped with $\mathcal{F}_1 = \mathcal{B}(\mathbb{R}) \times \{\mathbb{R}, \emptyset\}$ and $\mathcal{F}_2 = \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$. Cost function $c : X \times X \to \mathbb{R}$ is given by

$$(x_1, x_2) \mapsto |x_1 - x_2|^2_{\mathbb{R}^2}.$$

Let $\mu = \frac{1}{2}\delta_{(0,1)} + \frac{1}{2}\delta_{(0,-1)}$ and $\mu_\varepsilon = \frac{1}{2}\delta_{(\varepsilon,1)} + \frac{1}{2}\delta_{(-\varepsilon,-1)}$.
Then, as $\varepsilon \to 0$

$$\mathrm{d}(\mu, \mu_\varepsilon) \to 0,$$

but

$$\mathrm{d}_c(\mu, \mu_\varepsilon) \to 1.$$

# Robust optimization

## Framework

$$V(\delta) = \inf_{a \in A} \sup_{\nu \in B_\delta(\mu)} \int_X f(x,a)\nu(\mathrm{d}x).$$

- $X$ is a (filtered) Polish state space.
- $A$ is an admissible control set.
- $B_\delta(\mu)$ is some perturbation of $\mu$ with strength $\delta$.

## What do we want?

$$V'(0) = \lim_{\delta \to 0_+} \frac{V(\delta) - V(0)}{\delta}.$$

# Some results

- KL divergence [Lam16]
  - $X = \mathbb{R}^d$, $A = \{a\}$, $B_\delta^{KL}(\mu)$ KL ball.
  - $V'(0) = \sqrt{\mathrm{Var}_\mu(f(X,a))}$.

# Some results

- KL divergence [Lam16]
  - $X = \mathbb{R}^d$, $A = \{a\}$, $B_\delta^{KL}(\mu)$ KL ball.
  - $V'(0) = \sqrt{\mathrm{Var}_\mu(f(X,a))}$.
- Wasserstein distance [BDOW21]
  - $X = \mathbb{R}^d$, $A$ convex subset of $\mathbb{R}^n$, $B_\delta^W(\mu)$ Wasserstein-2 ball.
  - $V'(0) = \inf_{a^* \in A_0^*} \left( \int_X \|\nabla_x f(x,a^*)\|^2 \mu(\mathrm{d}x) \right)^{1/2}$, where $A_0^*$ is the set of minimizer of
$$\inf_{a\in A} \int_X f(x,a)\mu(\mathrm{d}x).$$

# Some results

- KL divergence [Lam16]
  - $X = \mathbb{R}^d$, $A = \{a\}$, $B_\delta^{KL}(\mu)$ KL ball.
  - $V'(0) = \sqrt{\mathrm{Var}_\mu(f(X,a))}$.
- Wasserstein distance [BDOW21]
  - $X = \mathbb{R}^d$, $A$ convex subset of $\mathbb{R}^n$, $B_\delta^W(\mu)$ Wasserstein-2 ball.
  - $V'(0) = \inf_{a^* \in A_0^*} \left( \int_X \|\nabla_x f(x,a^*)\|^2 \mu(\mathrm{d}x) \right)^{1/2}$, where $A_0^*$ is the set of minimizer of
  $$\inf_{a \in A} \int_X f(x,a)\mu(\mathrm{d}x).$$
- Adapted Wasserstein distance
  - $X = \mathbb{R}^d \times \mathbb{R}^d$ equipped with the natural filtration, $A$ convex subset of $\mathbb{R}^n$, $B_\delta^{AW}(\mu)$ adapted Wasserstein-2 ball.
  - $V'(0) = \inf_{a^*} \left( \int_X \|f_1(x_1,a^*)\|^2 + \|\nabla_{x_2} f(x_1,x_2,a^*)\|^2 \mu(\mathrm{d}x_1, \mathrm{d}x_2) \right)^{1/2}$, where $f_1(x_1,a^*) = \int_{\mathbb{R}^d} \nabla_{x_1} f(x_1,x_2,a^*)\mu_2(x_1,\mathrm{d}x_2)$ and $\mu(\mathrm{d}x_1,\mathrm{d}x_2) = \mu_2(x_1,\mathrm{d}x_2)\mu_1(\mathrm{d}x_1)$ is the disintegration.

# Open problems and partial answers

## General problems

Under what conditions will Monge meet Kantorovich again in causal optimal transport? Moreover, do we have parallel Brenier's Theorem for causal optimal transport?

## Specific problems

Let $X$ be $C([0,1],\mathbb{R})$, $D([0,1],\mathbb{R})$ or $\mathcal{S}^{'}(\mathbb{R})$ equipped with the natural filtration. What is the 'suitable' cost function?
Some 'nice' candidates:

- Uniform norm $c(\omega,\omega') = \|\omega - \omega'\|_\infty$.
- Cameron-Martin norm $c(\omega,\omega') = \|\omega - \omega'\|_H$.
- Variation norm $c(\omega,\omega') = \|\omega - \omega^{'}\|_{BV}$.

# One last thing

What the answer of $V'(0)$ when $X$ is the continuous path space, $c$ is induced by Cameron-Martin norm and $f \in L^2(\mu)$?

# One last thing

What the answer of $V'(0)$ when $X$ is the continuous path space, $c$ is induced by Cameron-Martin norm and $f \in L^2(\mu)$?
Partial answer:

- $\mu$ is Wiener measure $\gamma$, then we have

$$V'(0) = \left( \int_0^1 \mathbb{E}^\gamma \left[ \| \mathbb{E}^\gamma [\mathrm{D} f | \mathcal{F}_s] \|^2 \right] \mathrm{d}s \right)^{1/2},$$

  where $\mathrm{D}$ denotes the Malliavin derivative [NØP08].

- $\mu$ is a continuous martingale measure with the martingale representation property, then we have

$$V'(0) = \left( \int_0^1 \mathbb{E}^\mu \left[ \| \nabla_w \mathbb{E}^\mu [f | \mathcal{F}_s] \|^2 \right] \mathrm{d}s \right)^{1/2},$$

  where $\nabla_w$ denotes the vertical derivative [CF13].

# Main references I


[ABVZ20] Beatrice Acciaio, Julio Backhoff-Veraguas, and Anastasiia Zalashko, *Causal optimal transport and its links to enlargement of filtrations and continuous-time stochastic optimization*, Stochastic Processes and their Applications **130** (2020), no. 5, 2918–2953.

[BDOW21] Daniel Bartl, Samuel Drapeau, Jan Obłój, and Johannes Wiesel, *Sensitivity analysis of Wasserstein distributionally robust optimization problems*, Proceedings of the Royal Society A **477** (2021), no. 2256, 20210176.

[CF13] Rama Cont and David-Antoine Fournié, *Functional Itô calculus and stochastic integral representation of martingales*, The Annals of Probability **41** (2013), no. 1, 109–133.

# Main references II


[Lam16] Henry Lam, *Robust sensitivity analysis for stochastic systems*, Mathematics of Operations Research **41** (2016), no. 4, 1248–1275.

[Las18] Rémi Lassalle, *Causal transport plans and their Monge–Kantorovich problems*, Stochastic Analysis and Applications **36** (2018), no. 3, 452–484.

[NØP08] Giulia Di Nunno, Bernt Øksendal, and Frank Proske, *Malliavin calculus for Lévy processes with applications to finance*, Springer, 2008.

[Tho18] Matthew Thorpe, *Introduction to optimal transport*, March 2018.

[Vil09] Cédric Villani, *Optimal transport: old and new*, vol. 338, Springer, 2009.

# Thank you!