Optimal Transport in a Nutshell (3h!)

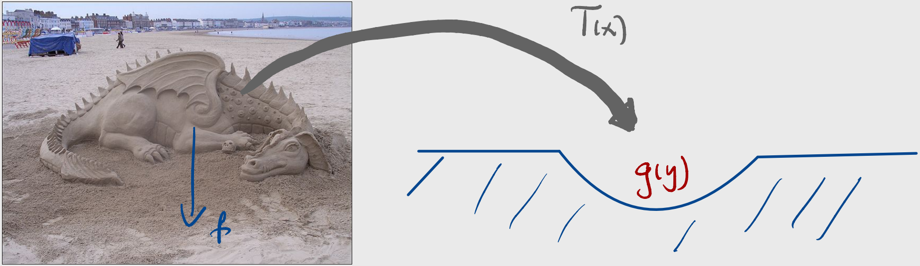
Marie-Therese Wolfram

Doomed to fail
Part 1

ICMS workshop on
'Connections between interacting particle dynamics and data science'

Monge's problem (1781):

How to move a pile of sand to a hole (both having the same volume) at minimial cost?



More mathematical: given two positive densities f and g, with $\int_{\mathbb{R}^d} f(x)dx = \int_{\mathbb{R}^d} g(y)dy = 1$, find a map $T : \mathbb{R}^d \to \mathbb{R}^d$ minimising the cost

displacement of particle located at x

$$M(T) := \int_{\mathbb{R}^d} |T(x) - x| f(x)dx,$$

subject to the constraint

$$\underbrace{\int_A g(y)dy}_{\text{mass dumped in } A} = \overbrace{\int_{T^{-1}(A)} f(x)dx}^{\text{mass taken from pre-image of } A} \text{ for any Borel subset } A \subset \mathbb{R}^d.$$

Timeline

Monge — 1781

Kantorovich — 1945

Villani — 2010

Figalli — 2018

Show properties of T

Considered a relaxation

PDE
applied prob.
economics

Computational OT & machine learning

TODAY:

- Monge
- Kantorovich
- Dual problem

TOMORROW:

- Wasserstein metric
- Geodesics
- Gradient flows
- Comput. OT

Notation

- More general, let's replace f and g by measures

$$\mu \in X \text{ and } \nu \in Y,$$

where X and Y are Polish spaces.

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

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

$c(x,y) = h(x-y)$ h convex

- Consider a general cost function $c : X \times Y \to \mathbb{R}$
- Spaces: $\mathcal{M}(X)$ space of finite measures on X

$$\mathscr{M}^+(X) := \{\mu \in \mathscr{M}(X) \colon \mu \geq 0\},$$
$$\mathscr{P}(X) := \{\mu \in \mathscr{M}^+(X) \colon \mu(X) = 1\}. \Leftarrow \text{probability measures}$$

- Push forward operator: Let $\mu, \nu \in \mathscr{P}(X)$ and $T : X \to Y$ be a measureable map. The push-forward $T_\# \mu$ is defined as

$$\nu(A) = \mu(T^{-1}(A))$$

for all measureable sets $A \subseteq Y$.

Equivalent: $\quad d\mu = f\,dx \qquad d\nu = g\,dv$

$$\int_Y \varphi(y)\,d\nu(y) = \int_X \varphi(T(x))\,d\mu(x) \qquad \varphi : Y \to \mathbb{R} \text{ measurable bdd.}$$

$$\varphi = \mathbb{1}_A \qquad \int_A g(y)\,dy = \int_{T^{-1}(A)} f(x)\,dx$$

The Monge Problem

Given two probability measures $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ and a cost function $c : X \times Y \to [0, \infty]$ find

$$\inf \left(M(T) := \left(\int c(x, T(x)) d\mu(x) \right) \right) \tag{MP}$$

over all measureable maps $T : X \to Y$ such that $\underbrace{T_{\#}\mu = \nu}_{(*)}$.

Monge is hard (*) highly nonlinear

$d\mu = f\,dx \quad d\nu = g\,dx$, $T \in C^1$ diffeomorphisms

$T_{\#}\mu = \nu \quad \underset{\text{change of variable}}{\Longrightarrow} \quad f(x) = g(T(x))\,|\det T(x)|$

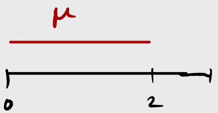
If $T = \nabla\varphi$

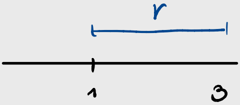
$\det \nabla^2 \varphi = \dfrac{f(x)}{g(\nabla\varphi)} \Leftarrow$ Monge-Ampere equation

Figalli, Caffarelli, ...

Example [Non-uniqueness] Let $X = [0,2]$, $Y = [1,3]$, $\mu = \frac{1}{2}\mathbf{1}_{[0,2]}$, $\nu = \frac{1}{2}\mathbf{1}_{[1,3]}$ and $c(x,y) = |x-y|$. Let T_1 and T_2 denote two transportation maps, given by $T_1(x) = x+1$ and

$$T_2(x) = \begin{cases} x+2 & \text{if } x \in [0,1] \\ x & \text{if } x \in (1,2]. \end{cases}$$





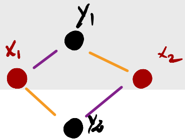
T_1 ... Linear translation

T_2 ... only move mass from $[0,1]$ to $[2,3]$

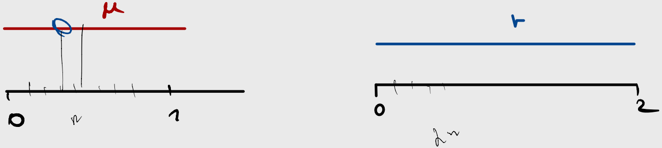
$$\left.\begin{aligned} \pi(T_1) &= \frac{1}{2}\int_0^2 |x+1-x|\,dx = 1 \\ \pi(T_2) &= 1 \end{aligned}\right\} \pi(T) = 1 \text{ the best possible}$$

$\Rightarrow$ T might not be UNIQUE

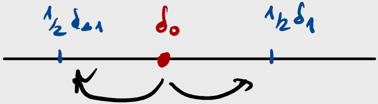
Ex2.



Example [Non-existence] Let $X = [0,1]$ and $Y = [0,2]$ and $\mu = \mathbf{1}_{[0,1]}$ and $\nu = \frac{1}{2}\mathbf{1}_{[0,2]}$ and cost $c(x,y) = |x-y|^{\frac{1}{2}}$.



$\Rightarrow$ mass can't be split



Why is (MP) difficult?

- Class of admissible transportation maps might be empty.
- Solution may not be unique.
- There's no notion of convergence, which makes the class of admissible transportation maps sequentially closed and compact.

The Kantorovich problem:
Given two probability measures $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ and a cost $c : X \times Y \to [0, \infty]$ find

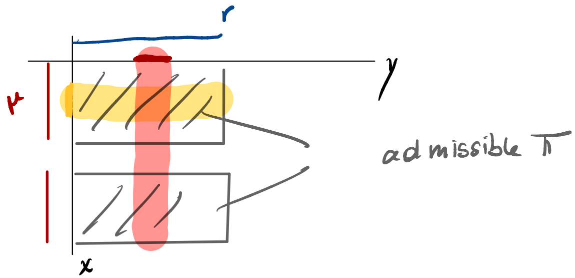
$$\inf \left(K(\pi) := \int_{X \times Y} c(x, y) d\pi(x, y) \right) \tag{KP}$$

among all admissible transportation plans $\pi \in \Pi(\mu, \nu)$, where

$$\Pi(\mu, \nu) := \{\pi \in \mathscr{P}(X \times Y) : (P_X)_{\#}\pi = \mu, (P_Y)_{\#}\pi = \nu\}$$

and P_X and P_Y are the projections of $X \times Y$ onto X and Y, respectively.

Transportation plan $\pi(x,y)$ how much mass was moved from x to y,



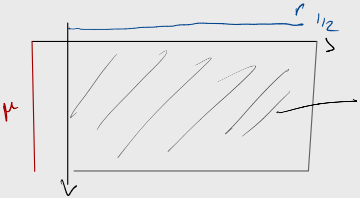
Why is (KP) easier?

$\pi = \mu \otimes \nu$

- The set $\Pi(\mu, \nu)$ is not empty, and compact and convex wrt to the narrow topology.
- The mapping $\pi \to \int c(x, y) d\pi(x, y)$ is linear.
- Transportation plans π 'include' transportation maps T.
- Problem is symmetric.

$$\begin{tikzcd} \pi \arrow[r, bend left, "\text{under some conditions!}"] & T \arrow[l, bend left, "\text{true}"] \end{tikzcd}$$

Example Let $X = [0, 1]$ and $Y = [0, 2]$ and $\mu = \mathbf{1}_{[0,1]}$ and $\nu = \frac{1}{2}\mathbf{1}_{[0,2]}$ and cost $c(x, y) = |x - y|^{\frac{1}{2}}$.



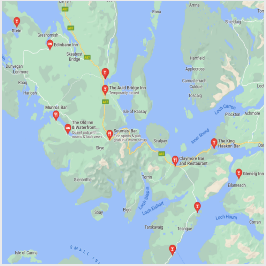
$\pi = \frac{1}{2} \mathbb{1}_{[0,1] \times [0,2]}$ admissible

A Scottish example [Discrete measures] Consider two discrete measures, where all x_i and y_j are different

$$\mu = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \text{ and } \nu = \frac{1}{n}\sum_{j=1}^{n}\delta_{y_j}.$$

For example, x_i may denote the locations of distilleries on Skye while y_j are the location of pubs.

(a) Distilleries



(b) Pubs

⇒ Linear assignment problem

Monge

$$\min \frac{1}{n} \sum c(x_i, T(x_i))$$

$$\Downarrow$$

$$\min \frac{1}{n} \sum c(x_i, y_{\sigma(i)}) \qquad \sigma \ldots \text{ permutation}$$

Kantorovich

$$\min \frac{1}{n} \sum c_{ij} \pi_{ij}$$

$$\Pi(\mu,\nu) = \left\{ \pi \in \mathbb{R}^{n \times n} : \sum_i \pi_{ij} = \sum_j \pi_{ij} = \frac{1}{n} \right\}$$

$\Rightarrow$ Birkhoff's theorem: extremal points of the set of bistochastic matrices are induced by permutation

Monge and Kantorovich agree

Relax, Monge!

Every transportation map T (in the sense of Monge) induces a transportation plan π:

$$\pi_T := (\mathrm{Id}, T)_{\#}\mu. \quad \Leftarrow \quad d\pi_T = \delta_{y=T(x)}\, d\mu$$

If T^+ is optimal $\Rightarrow$ π_{T^+}

$$\pi_{T^+}(A \times Y) = \int_A \delta_{y=T(x)}\, d\mu(x) = \mu(A) \quad ✓$$

$$\pi_{T^+}(X \times B) = \nu(B) \quad ✓$$

marginal constr. satisfied

Since

$$\overbrace{\int_{X\times Y} c(x,y)\, d\pi(x,y)}^{K(\pi)} = \int_X \int_Y c(x,y)\, \delta_{y=T(x)}\, dy\, d\mu$$

$$= \int_X c(x, T(x))\, d\mu(x) = \mathcal{M}(T)$$

Therefore

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

$\Rightarrow$ Transport map induces transport plan π

Converse? If an plan π can be written

$$d\pi(x,y) = \delta_{y=T(x)} \, d\mu(x)$$

$\Rightarrow$ T is an OT map

Theorem (Existence of transportation plans)

Let X and Y be compact metric spaces, $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ and $c : X \times Y \to \mathbb{R}$ be a lower semi-continuous function. Then (KP) admits a unique solution.

Direct method of calculus of variations

Kontorovich

$$\min_{v \in V} F(v) \qquad \min_{\pi \in \Pi} K(\pi)$$

- Check that $\{v \in V : F(v) < \infty\} \neq 0$.
- Consider a minimising sequence $\{v_n\}$ in V with $F(v_n) \to \inf F(v)$.
- Show that V is compact in a suitable topology. ← the weaker the topology the easier conv. of subsequ.
- F is lower semicontinuous wrt this topology. ← the stronger the easier the set

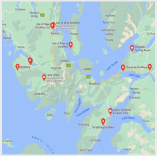
Put together

$$\inf\{F(v) \mid v \in V\} \overset{\text{min. sequ.}}{=} \lim_{n\to\infty} F(v_n) \overset{\text{conv. subseq.}}{=} \lim_{k\to\infty} F(v_{n_k}) \overset{\text{low. semi cont}}{\geq} F(v^+) \geq \inf\{F(v) \mid v \in V\}$$

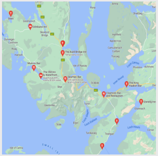
$$F(v^+) = \inf\{F(v) \mid v \in V\}$$

Example

- *We wish to transport bottles of whiskey from a given number of distilleries and a given number of pubs.*
- *Let $c(x, y)$ denote the cost of transporting one unit of whiskey from distillery x to pub y.*



(c) Distilleries



(d) Pubs

Outsourcing to contractor

o) $\psi(x)$ price to pick up whiskey at x

o) $\varphi(y)$ — " — drop off — " — y

$$\psi(x) + \varphi(y) \le c(x,y) \quad \forall\, x,y$$

The Dual Problem

Given two probability densities $\mu \in \mathscr{P}(X)$ and $\nu \in \mathscr{P}(Y)$ and a cost function $c : X \times Y \to [0, \infty)$ consider

$$\max_{\varphi,\psi\in\Phi_c} \left(\int_X \varphi(x) d\mu(x) + \int_Y \psi(y) d\nu(y) d : \varphi \in C_b((X), \psi \in C_b(Y) \right), \tag{DP}$$

where $\Phi_c = \{(\varphi, \psi) \in C_b(X) \times C_b(Y) : \varphi(x) + \psi(y) \leq c(x, y)\}$.

Since $\varphi(x) + \psi(y) \leq c(x,y)$ $\qquad \varphi \oplus \psi = \varphi(x) + \psi(y)$

$$\int_X \varphi(x) d\mu + \int_Y \psi(y) d\mu = \int_{X\times Y} (\varphi \oplus \psi) d\pi$$

$$\leq \int_{X\times Y} c(x,y) d\pi$$

$\sup DP \leq \min KP$

If $\sup DP = \min KP$ $\Leftarrow$ strong duality

Optimality & Monotonicity

$\exists \pi$: π_{ij} amount of whiskey going to x_i to y_j

Assume my cost is too high

BENEFIT

o) $x_1 \to y_1$ decide $x_1 \to y_2$

$c(x_1, y_1) - c(x_1, y_2)$

o) $x_2 \to y_2$ decide $x_2 \to y_3$

$c(x_2, y_2) - c(x_2, y_3)$

$\vdots$

$+ \, c(x_N, y_N)$

$$c(x_1, y_2) + \dots + c(x_N, y_1) \leq c(x_1, y_1) + \dots$$

$\Rightarrow \pi$ CAN'T BE OPTIMAL

Definition (C-Cyclical monotonicity)

Let X, Y be a arbitrary sets and $c\colon X \times Y \to (-\infty, \infty]$ be a function. A subset $\Gamma \subset X \times Y$ is called c-cyclically monotone if for any $m \in \mathbb{N}$ and any family $(x_1, y_1), (x_2, y_2) \ldots, (x_N, y_N)$ of points in Γ

$$\sum_{i=1}^{N} c(x_i, y_i) \leq \sum_{i=1}^{N} c(x_i, y_{i+1})$$

with $y_{N+1} = y_1$. A transportation plan is c-cyclically monotone if its concentrated on a c-cyclically monotone set.

$\Rightarrow$ π is optimal $\rightarrow$ spt(π) is a c-CM set

Definition (C-Transforms)

Let $\varphi : X \to \mathbb{R} \cup \{\infty\}$, then its c-transform $\varphi^c : Y \to \mathbb{R} \cup \{\infty\}$ is defined as

$$\varphi^c(y) = \inf_{x \in X} c(x, y) - \varphi(x)$$

$\varphi(x) + \psi(y) \leq c(x,y)$

The $\bar{c}$-transform of $\psi : Y \to \mathbb{R} \cup \{\infty\}$ is given by $\psi^{\bar{c}}(x) = \inf_{y \in Y} c(x, y) - \psi(y)$.

Theorem

Let X and Y be Polish spaces and assume that $c : X \times Y \to \mathbb{R}$ is uniformly continuous and bounded. The (DP) admits a solution (φ, φ^c) and $\max(DP) = \min(KP)$.

(φ, ψ) $\Leftarrow$ Kantorovich potential

Theorem

Let μ and ν be two probability measures on a compact domain $\Omega \subset \mathbb{R}^d$ and the cost $c(x, y) = h(x - y)$ with h strictly convex.

Then there exists a unique transportation plan π of the form $(id, T)_{\#}\mu$ if μ is absolutely continuous and $\partial\Omega$ negligible. The corresponding Kantorovich potential φ is linked via

$$T(x) = x - (\nabla h)^{-1}(\nabla \varphi(x))).$$

Duality: Assume there exist $\exists$ OT plan π & Kantorovich plan φ

$$\varphi(x) + \varphi^c(y) \leq c(x,y) \quad \text{on } \Omega \times \Omega$$

$$\varphi(x) + \varphi^c(y) = c(x,y) \quad \text{on } \mathrm{spt}(\pi)$$

Consider (x_0, y_0) on $\mathrm{spt}(\pi)$

$$x \mapsto \varphi(x) - c(x, y_0) \text{ is min at } x_0$$

If φ & c diff. $\quad \nabla\varphi(x_0) = \nabla_x c(x_0, y_0) = \nabla_x h(x_0 - y_0)$

If h is strictly convex

$$x_0 - y_0 = (\nabla h)^{-1}(\nabla \varphi(x_0))$$

$$T(x) = x - (\nabla h)^{-1}(\nabla \varphi(x))$$

NOT working $c(x,y) = |x-y|$

Theorem (Brenier's Theorem)

Let $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ with $X, Y \subset \mathbb{R}^d$, assume that both have finite second moments and that μ does not give mass to small sets, and let $c(x, y) = |x - y|^2$.

Then there exists a unique solution π to the Kantorovich problem (KP). This plan is uniquely supported on the graph $(x, T(x))$, that is $\pi = (Id, T)_{\#}\mu$. Furthermore there exists an $L^1(\mu)$, convex, lower-semicontinuous function $\bar{\varphi}$ such that $\pi = (Id \times \nabla\bar{\varphi})_{\#}\mu$.

Quadratic cost.

$$T(x) = x - \nabla\varphi(x) = \nabla \underbrace{\left(\frac{x^2}{2} - \varphi\right)}_{:= u}$$

u convex
Lsc.

$\Rightarrow$ Transportation map is the gradient of a convex & lsc function

$$\begin{tikzcd}
\text{MONGE } T \arrow[rr] & & \boxed{\text{KANTOROVICH}} \\
& \text{DUAL } (\varphi, \varphi^c) \arrow[ul] \arrow[ur] &
\end{tikzcd}$$

References:

- Alfred Galichon, *Optimal transport in economics*, Princeton University Press, 2016
- Gabriel Peyre and Macro Cuturi, *Computational optimal transport*, Foundations and Trends in Machine Learning, 2019
- Fillipo Santambrogio,*Optimal transport for applied mathematicians*, Springer, 2015.
- Cedric Villani, *Topics in optimal transport*, AMS 2003.
- Cedric Villani, *Optimal transport: old and new*, Springer, 2008.

Python libraries

- CVPXY
  Stephan Diamon and Steve Boyd *CVXPY: A Python-embedded modeling language for convex optimization*, Journal of Machine Learning Research, 2016
  URL: `https://www.cvxpy.org/index.html`
- POT Python Optimal Transport
  R. Flamary et al. *POT Python Optimal Transport library,*Journal of Machine Learning Research, 22(78):18, 2021.
  URL: `https://pythonot.github.io/`
- OTT Optimal Transport Toolbox
  M. Cuturi et al. *Optimal Transport Tools (OTT): A JAX Toolbox for all things Wasserstein*, arXiv:2201.12324, 2022
  URL: `https://github.com/ott-jax/ott`