Lecture 4

Recall:

Monge's Optimal Transport Problem

Given $\mu, \nu \in \mathcal{P}(X)$, solve

$$\min_{\substack{t : X \to X \text{ measurable} \\ t\#\mu = \nu}} \overbrace{\int d(x, t(x))\, d\mu(x)}^{\text{effort}}$$

<u>Difficulty #1</u>: the constraint set can be empty.

<u>Difficulty #2</u>: Solutions may not be unique.

<u>Difficulty #3</u>: The constraint set is nonconvex

Recall basic convexity facts:

vector space $\mathcal{X}$

$C \subseteq \mathcal{X}$ is convex if $\forall\, x_0, x_1 \in C$,

$$x_\alpha := (1-\alpha)x_0 + \alpha x_1 \in C, \quad \forall \alpha \in [0,1]$$

$f : C \to \mathbb{R} \cup \{+\infty\}$ is...

- convex if $f(x_\alpha) \leq (1-\alpha) f(x_0) + \alpha f(x_1)$
- concave $\geq$
- strictly convex $<$

...for all $x_0, x_1 \in C$, $\alpha \in (0,1)$.

If f is convex and concave, it is affine linear.

Relax the problem.

Leonid Kantorovich, 1942

"On the translocation of masses"

Notation:

Projection maps:

$\pi_X : X \times Y \to X, \quad \pi_X(x,y) = x$

$\pi_Y : X \times Y \to Y, \quad \pi_Y(x,y) = y$

$A \in \mathcal{B}(X)$

Marginals: For $\gamma \in \mathcal{P}(X \times Y)$, define

first marginal $\pi_X \# \gamma(A) = \gamma(\pi_X^{-1}(A)) = \gamma(A \times Y)$

second marginal $\pi_Y \# \gamma$

Def (transport plan): Given $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$, the set of transport plans from μ to ν is

$$\Gamma(\mu,\nu) = \{\gamma \in \mathcal{P}(X \times Y) : \pi_X \# \gamma = \mu, \pi_Y \# \gamma = \nu\}$$

We will use transport plans as a new way to model rearranging mass in μ to look like ν. For $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(Y)$,

$\gamma(A \times B) =$ amt of mass from $\mu(A)$ that is sent to $\nu(B)$.

How do transport plans relate to transport maps?

Notation: $id: X \to X$, $id(x) = x$

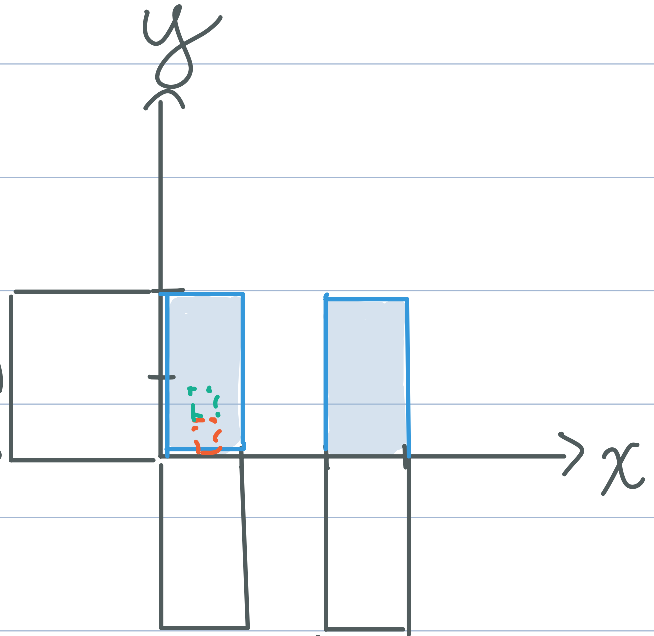
Lemma: Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, if $t_{\#}\mu = \nu$, then

$$\gamma := (id \times t)_{\#}\mu \in \Gamma(\mu, \nu).$$

$id \times t : X \to X \times Y$. $(id \times t)(x) = (x, t(x))$

$$d\gamma(x,y) = \tfrac{1}{2}\left(1_{[0,1]\times[0,1]} + 1_{[2,3]\times[0,1]}\right)(x,y)\, d\lambda^2(x,y)$$

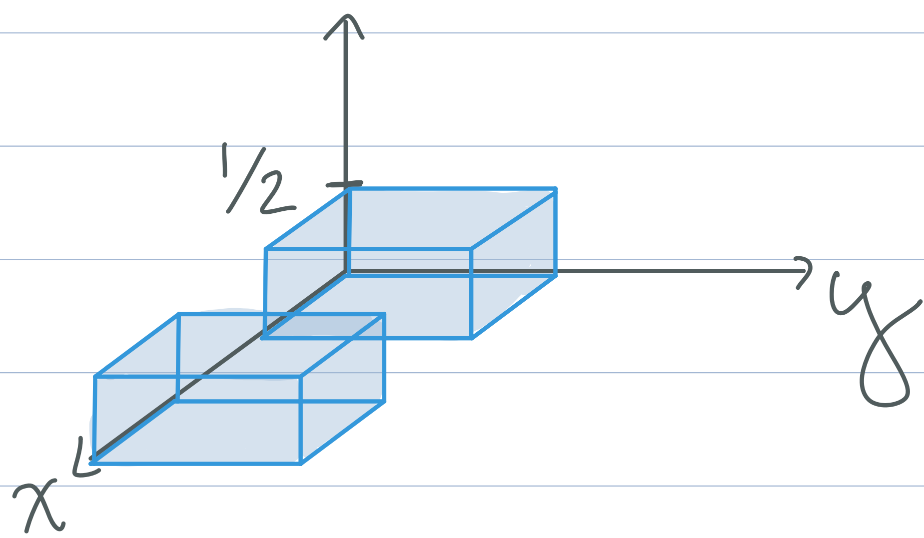
Bird's eye view:



$d\nu(y) = 1_{[0,1]}(y)\, d\lambda(y)$

$d\mu(x) = \tfrac{1}{2}\left(1_{[0,1]} + 1_{[2,3]}\right)(x)\, d\lambda(x)$

Side view:



This is a special case of the fact that..

For any $\mu, \nu \in \mathcal{P}(X)$, the transport plan
$$\gamma = \mu \otimes \nu \in \Gamma(\mu, \nu)$$
"takes mass from any location x_0 in μ and distributes it across ν, in proportion to the amount of mass ν assigns to each location."

Moral: (i) For any μ, ν, $\Gamma(\mu, \nu) \neq \emptyset$.
(ii) transport plans can "split mass"

Ex: For $\mu = \frac{1}{2}(\mathbb{1}_{[0,1]} + \mathbb{1}_{[2,3]})$, $\nu = \frac{1}{2}(\mathbb{1}_{[0,2]})$,

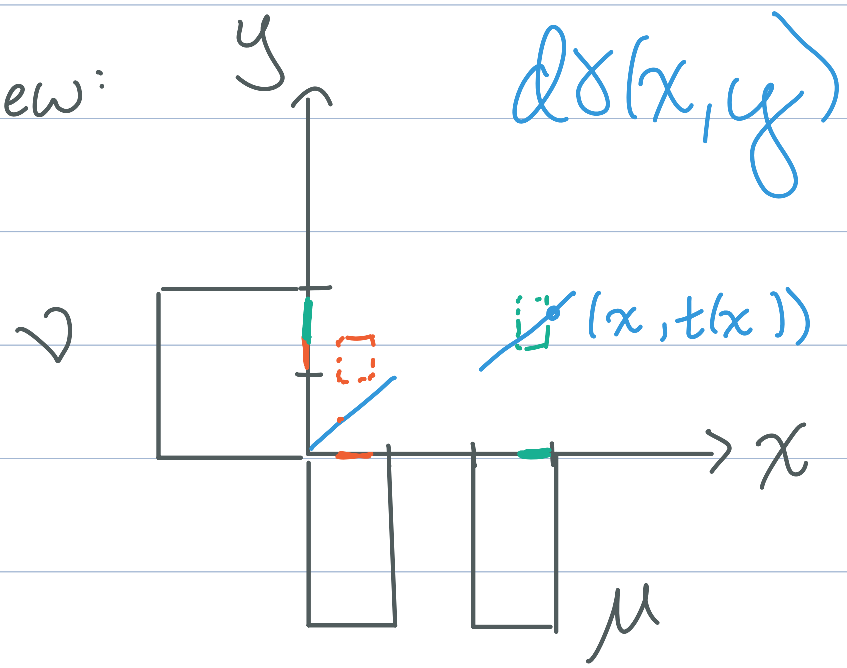
consider the transport map

$$t(x) = \begin{cases} x & \text{if } x \in [0,1] \\ x-1 & \text{otherwise} \end{cases}$$

Then $t\#\mu = \nu$, so by lemma,

$$\gamma := (id \times t)\#\mu \in \Gamma(\mu,\nu).$$

Bird's eye view:

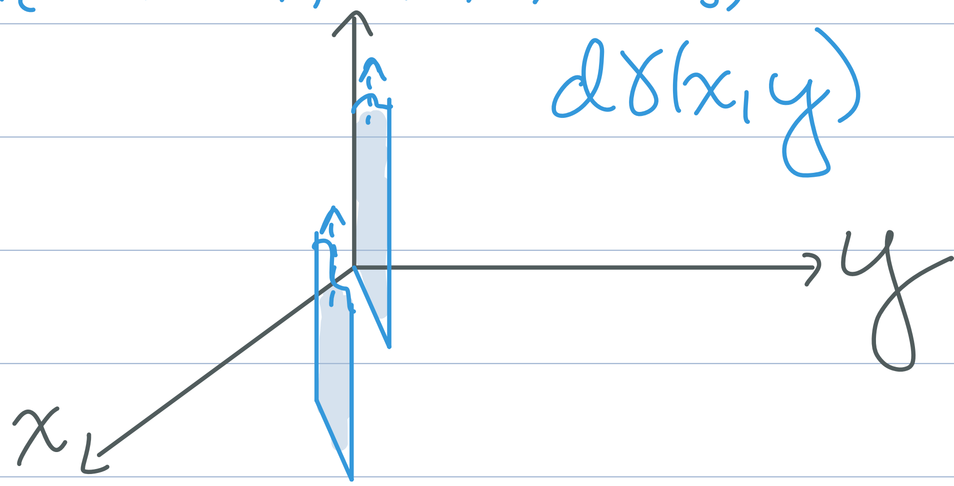


γ is the uniform probability measure supported on

$$\{(x, t(x)) : x \in [0,1] \cup [2,3]\}$$

$$\gamma(A\times B) = \mu\left((\mathrm{id}\times t)^{-1}(A\times B)\right)$$
$$= \mu\left(\{x\in X : (x,t(x))\in A\times B\}\right)$$

Side view:



Foreshadowing: When $\mu \ll \lambda^1$, we will see that γ is an optimal transport plan from μ to ν iff it is supported on $\{(x,t(x)) : x\in\mathbb{R}\}$ for an increasing function $t(x)$. $c(x,y) = d(x,y)^p,\ p\geq 1$

Using transport plans, we can now state...

Kantorovich's Optimal Transport Problem

Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$

$c : X \times Y \to \mathbb{R} \cup \{+\infty\}$ lower semicts and bdd below

$$\min_{\gamma : \gamma \in \Gamma(\mu,\nu)} \underbrace{\int_{X \times Y} c(x,y)\, d\gamma(x,y)}_{\mathbb{K}_c(\gamma)}$$

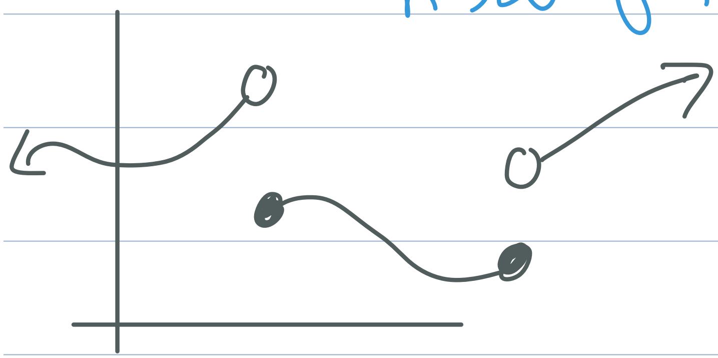
cost of moving mass from location x to y

If γ_* attains the minimum, we will call it an <u>optimal transport plan</u>.

Recall... given a metric space (Z, d_Z),

Def: $f: Z \to \mathbb{R} \cup \{+\infty\}$ is <u>lower semicontinuous</u> if $z_n \xrightarrow{d} z$ implies

$$\liminf_{n\to\infty} f(z_n) \geq f(z)$$



Reasons Kantorovich Problem Better:

① the constraint set is always nonempty
② the constraint set is convex (Ex 8)
③ the objective fn is linear, hence convex (Ex 9)

Remark: Since Kantorovich's problem is the minimization of a linear objective function, subject to affine linear equality and inequality constraints, it is a linear program... hence is a convex optimization problems.

Ex (finite dim'l linear program)

$$\min_{x} \; A_2 x$$

$$A_0 x = b_0$$

$$A_1 x \geq b_1$$

(4) Kantorovich's problem has a dual

(5) We can easily prove existence of solns to Kantorovich prob via...

The Direct Method of the Calculus of Variations

Setup: $C \subseteq \mathcal{X}$, $(\mathcal{X}, d)$ metric space

Goal: prove that $\min_{x \in C} F(x)$ exists

Step 0: Show $\inf_{x \in C} F(x) < +\infty$.

Step 1: Take a minimizing sequence.

That is, choose $\{x_n\}_{n=1}^{\infty} \subseteq C$ s.t.

$$\lim_{n \to \infty} F(x_n) = \inf_{x \in C} F(x).$$

Step 2: Prove that C is compact, so $\exists$ subsequence $\{x_{n_k}\}_{k=1}^{\infty} \subseteq C$ so that $x_{n_k} \xrightarrow{k \to \infty} x_* \in C$

Step 3: Prove that F is lsc, so

$$\inf_{x\in C} F(x) = \lim_{n\to\infty} F(x_n) = \lim_{k\to\infty} F(x_{n_k})$$

$$= \ldots = \liminf_{k\to\infty} F(x_{n_k}) \geq F(x_*)$$

Thus x_* is a minimizer of F on C.

Key challenge: choosing the right metric d on C.

- weak enough to get compactness
- strong enough to get lsc

What is the right topology for Kantorovich's problem?

First: consider compactness of constraint set.

Thm (Prokhorov)

Given a Polish space (Z, d_Z) and $\mathcal{K} \subseteq \mathcal{P}(Z)$,

- $\mathcal{K}$ is relatively compact in narrow topology

$\Updownarrow$

- $\mathcal{K}$ is tight, $\forall \varepsilon > 0$, $\exists K_\varepsilon \subset\subset Z$ s.t. $\mu(Z \setminus K_\varepsilon) \leq \varepsilon \ \forall \mu \in \mathcal{K}$

An immediate corollary is...

Cor: If (Z, d_Z) is a Polish space, then for any $\sigma \in \mathcal{P}(Z)$, $\{\mu\}$ is tight.

Other key step...

<u>Lemma</u>: Given Polish spaces (X, d_X) and (Y, d_Y) and $\{\mu_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(X)$ narrowly converging to $\mu \in \mathcal{P}(X)$, then for any continuous function $t: X \to Y$, $t\#\mu_n$ narrowly converges to $t\#\mu$.

Pf: See exercise 5.

<u>Prop</u>: Given Polish spaces (X, d_X) and (Y, d_Y), $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, then $\Gamma(\mu, \nu)$ is compact in the narrow topology.

Pf: By corollary, $\forall \varepsilon>0$, $\exists\, K_\varepsilon^\mu \subset\subset X$, $K_\varepsilon^\nu \subset\subset Y$ s.t.

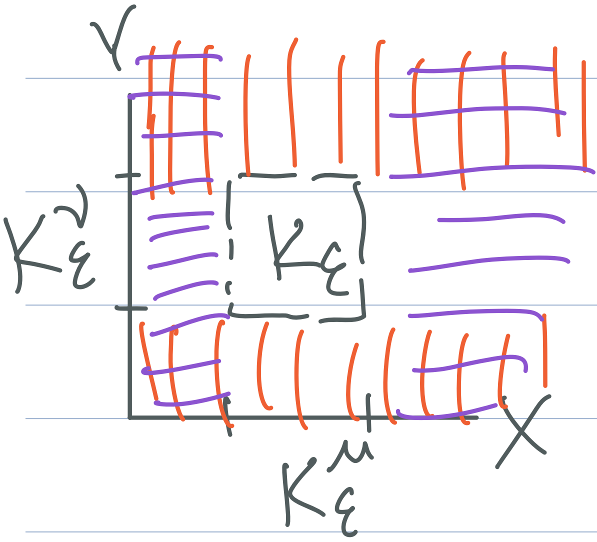
$$\mu(X\setminus K_\varepsilon^\mu)+\nu(Y\setminus K_\varepsilon^\nu)\le \varepsilon.$$

Define $K_\varepsilon := K_\varepsilon^\mu \times K_\varepsilon^\nu \subset\subset X\times Y$.

$\gamma(A\times Y)=\mu(A)$, $\gamma(X\times B)=\nu(B)$

Then, for any $\gamma\in\Gamma(\mu,\nu)$,

$$\gamma((X\times Y)\setminus K_\varepsilon)\le \gamma((X\setminus K_\varepsilon^\mu)\times Y) + \gamma(X\times(K_\varepsilon^\nu\setminus Y))$$



$$= \mu(X\setminus K_\varepsilon^\mu) + \nu(Y\setminus K_\varepsilon^\nu) \le \varepsilon.$$

Thus $\Gamma(\mu,\nu)$ is tight, hence relatively cpt, by Prokhorov.

For any $\{\gamma_n\}_{n=1}^{\infty} \subseteq \Gamma(\mu,\nu)$, $\exists$ $\{\gamma_{n_k}\}_{k=1}^{\infty} \subseteq \Gamma(\mu,\nu)$, $\gamma_* \in \mathcal{P}(X \times Y)$ so that $\gamma_{n_k} \to \gamma_*$ narrowly.

It remains to show $\gamma_* \in \Gamma(\mu,\nu)$.

By continuity of π_X, π_Y and previous lemma,

$$\underbrace{\pi_X \# \gamma_{n_k}}_{\mu} \to \pi_X \# \gamma_* \text{ narrowly}$$

$$\underbrace{\pi_Y \# \gamma_{n_k}}_{\nu} \to \pi_Y \# \gamma_* \text{ narrowly}$$

Hence $\gamma_* \in \Gamma(\mu,\nu)$. $\square$

Next: want to show $\mathbb{K}_p(\gamma)$ is lsc in narrow topology.

<u>Lemma</u>: Given metric space (Z, d_Z), suppose $g: Z \to \mathbb{R} \cup \{+\infty\}$ is lsc and bdd below. Then $\exists$ $\{g_k\}_{k=1}^{\infty} \subseteq C_b(Z)$ s.t.

$$g_k(z) \xrightarrow[k\to\infty]{} \nearrow g(z), \quad \forall z \in Z.$$

<u>Pf</u>: Exercise 9, via Moreau-Yosida regularization.

<u>Thm</u> (Portmanteau): For any $g: Z \to \mathbb{R} \cup \{+\infty\}$ lsc, bdd below, the functional

$$\mu \mapsto \int_Z g\, d\mu$$

is lsc wrt narrow topology.

Pf: By lemma, $\exists \{g_k\}_{k=1}^{\infty} \subseteq C_b(Z)$ s.t. $g_k \nearrow g$ pointwise.

Fix $\{\mu_n\}_{n=1}^{\infty} \subseteq \mathcal{P}(Z)$ converging narrowly to $\mu \in \mathcal{P}(Z)$.

For any $k \in \mathbb{N}$,

$$\liminf_{n\to\infty} \int g\, d\mu_n \geq \liminf_{n\to\infty} \int g_k\, d\mu_n$$

$$= \int g_k\, d\mu.$$

... finish next time ⌣̈.