Lecture 4

Recall:

Monge's Optimal Transport Problem:
Given $\mu, \nu \in \mathcal{P}(X)$, solve

$$\min_{t : t\#\mu = \nu} \underbrace{\int d(t(x), x)^p \, d\mu(x)}_{\mathbb{M}_p(t)} , \quad p \geq 1$$

If a transport map t attains the minimum, we call it an <u>optimal transport map</u> from μ to ν.

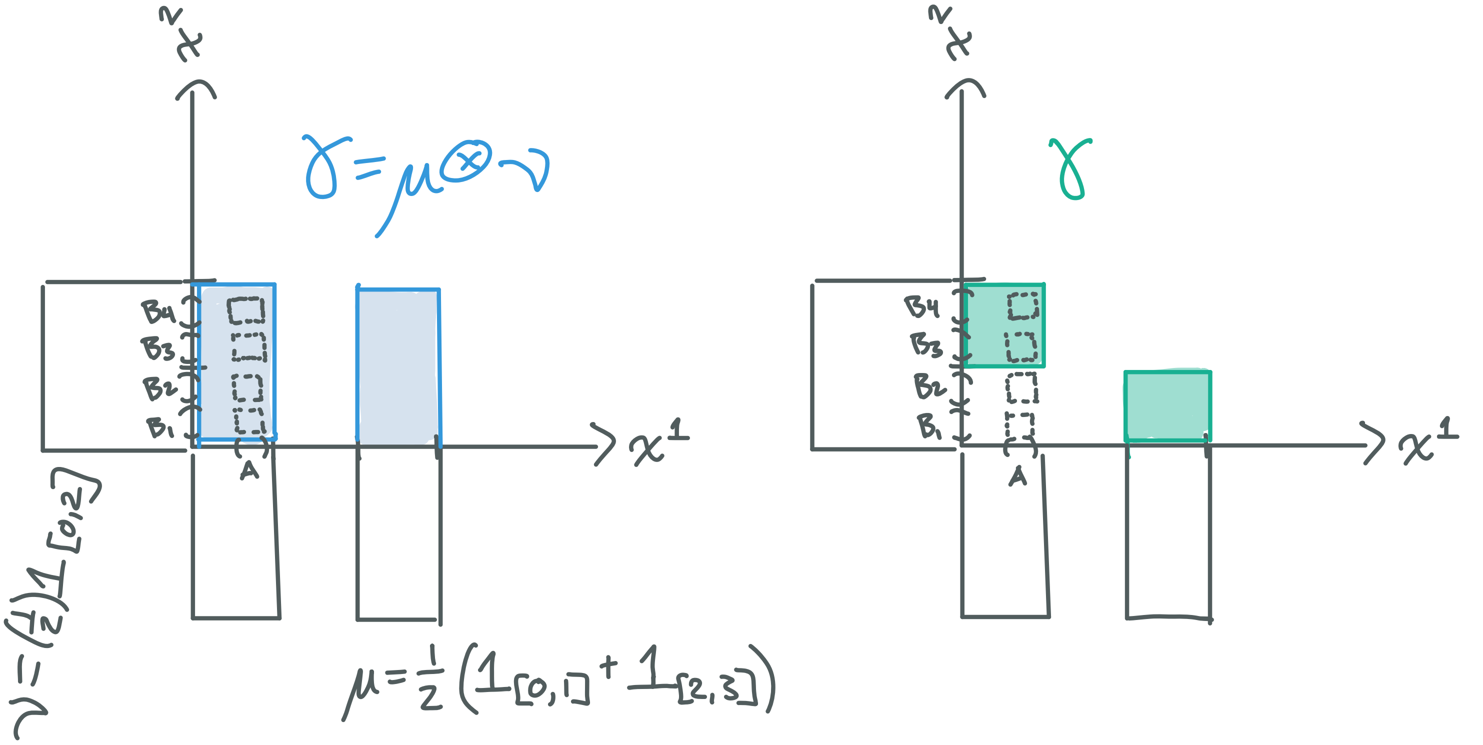
<u>Def</u>: Given $\mu, \nu \in \mathcal{P}(X)$, the set of <u>transport plans</u> from μ to ν is

$$\Gamma(\mu, \nu) = \{ \gamma \in \mathcal{P}(X \times X) : \pi^1 \# \gamma = \mu, \ \pi^2 \# \gamma = \nu \}$$

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

<u>Lemma</u>: If $t \# \mu = \nu$, then $\gamma := (id \times t) \# \mu \in \Gamma(\mu, \nu)$.

Ex: Consider two transport plans.



Exercise: What is a formula for γ? Verify directly from the definition that $\gamma \in \Gamma(\mu, \nu)$.

Ex: For μ and ν as above, consider the transport map

$$t(x) = \begin{cases} x & \text{if } x \in [0,1] \\ x-1 & \text{if } x \in [2,3] \end{cases}$$

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

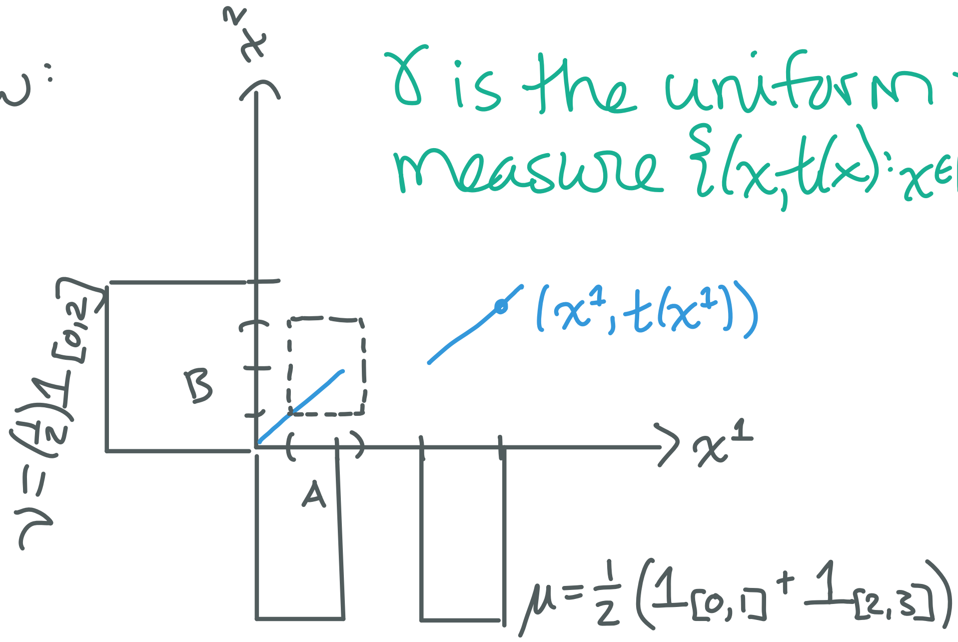
$$\gamma(A \times B) = \mu\left((id \times t)^{-1}(A \times B)\right)$$

$$= \mu\left(\{x : (x, t(x)) \in A \times B\}\right)$$

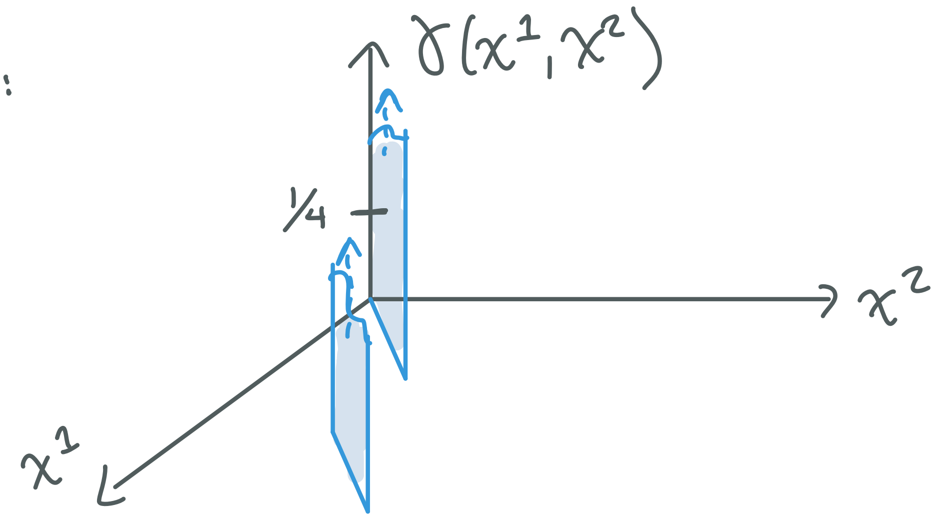
$$= \frac{1}{2}\left|\{x \in [0,1] : (x, x) \in A \times B\}\right| + \frac{1}{2}\left|\{x \in [2,3] : (x, x-1) \in A \times B\}\right|$$

Bird's eye view:

γ is the uniform prob. measure $\{(x, t(x)) : x \in [0,1] \cup [2,3]\}$.



Side view:



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

Using transport plans, we can now state...

Kantorovich's Optimal Transport Problem

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

$$\min_{\gamma : \gamma \in \Gamma(\mu,\nu)} \underbrace{\int_{X \times X} d(x^1,x^2)^p \underbrace{d\gamma(x^1,x^2)}_{\text{"how much mass that } \gamma \text{ sends from } x^1 \text{ to } x^2\text{"}}}_{\mathbb{K}_p(\gamma) :=}, \quad p \geq 1$$

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

Reasons this is a better behaved problem:

① $\forall \mu, \nu \in \mathcal{P}(X)$, the constraint set is nonempty (since $\mu \otimes \nu \in \Gamma(\mu,\nu)$)

② The constraint set is convex.

In particular, given $\gamma_0, \gamma_1 \in \Gamma(\mu,\nu)$, define

$$\gamma_\alpha := (1-\alpha)\gamma_0 + \alpha\gamma_1 \quad \text{for } \alpha \in [0,1].$$

Then for any $\alpha \in [0,1]$, $A \in \mathcal{B}(X)$,

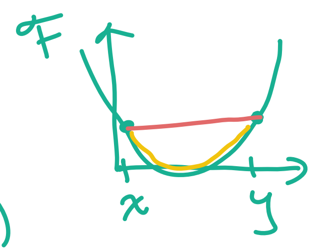
$$\begin{aligned}\pi^1 \# \gamma_\alpha(A) &= \gamma_\alpha(A \times X) \\ &= (1-\alpha)\gamma_0(A \times X) + \alpha \gamma_1(A \times X) \\ &= (1-\alpha)\pi^1 \# \gamma_0(A) + \alpha \pi^1 \# \gamma_1(A) \\ &= (1-\alpha)\mu(A) + \alpha \mu(A) \\ &= \mu(A).\end{aligned}$$

Thus $\pi^1 \# \gamma_\alpha = \mu$. Similarly, $\pi^2 \# \gamma_\alpha = \nu$. We conclude $\gamma_\alpha \in \Gamma(\mu,\nu)$ for all $\alpha \in [0,1]$.

3 The objective function is convex.

Recall:

Def: Given a vector space Y and $C \subseteq Y$ convex, a function $\mathcal{F}: C \to \mathbb{R} \cup \{+\infty\}$ is convex if, for all $x, y \in C$,



$$\mathcal{F}((1-\alpha)x + \alpha y) \leq (1-\alpha)\mathcal{F}(x) + \alpha \mathcal{F}(y)$$

With regard to Kantorovich's problem,

$$\mathbb{K}_p(\gamma_\alpha) = \int d(x^1,x^2)^p \, d\gamma_\alpha(x^1,x^2)$$

$$= \int d(x^1,x^2)^p \, d\big((1-\alpha)\gamma_0 + \alpha\gamma_1\big)(x^1,x^2)$$

$$= (1-\alpha)\int d(x^1,x^2)^p \, d\gamma_0(x^1,x^2) + \alpha \int d(x^1,x^2)^p \, d\gamma_1(x^1,x^2)$$

$$= (1-\alpha)\,\mathbb{K}_p(\gamma_0) + \alpha\,\mathbb{K}_p(\gamma_1)$$

better than convex: linear!

Remark: Since Kantorovich's problem is the minimization of a convex objective function, subject to convex constraints, it is a convex optimization problem.

④ Kantorovich's problem has a dual problem.

will discuss soon

⑤ We can prove there always exists a solution to Kantorovich's problem via...

The Direct Method of the Calculus of Variations

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

Step 1: Take a minimizing sequence, i.e. choose $x_n \in C$ s.t.

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

Step 2: Prove that C is sequentially compact in some topology, so there exists a subsequence s.t. $x_{n_k} \to x_* \in C$.

Step 3: Prove that $F(x)$ is lower semicontinuous in some topology so that

$F: X \to \mathbb{R} \cup \{+\infty\}$ is lsc if $x_n \to x_0$ implies $\liminf_{n\to\infty} F(x_n) \geq F(x_0)$.

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

Thus x_* is a minimizer of F over C.

Key challenge: choosing the right topology
- weak enough to get compactness of C
- strong enough to get lsc of F

What is the right topology for Kantorovich's problem?

<u>Def</u>: A sequence $\mu_n \in \mathcal{M}(X)$ is <u>narrowly convergent</u> to $\mu \in \mathcal{M}(X)$ if

$$\lim_{n\to\infty} \int_X \varphi \, d\mu_n = \int \varphi \, d\mu, \quad \forall \varphi \in C_b(X).$$

<u>Remark</u>: Thinking of $\mathcal{M}(X) \subseteq (C_b(X))^*$, this is convergence in the weak-* topology. Also called "weak convergence" in probability literature.

<u>Remark</u>: The narrow topology is <u>metrizable</u>, so cpt $\Leftrightarrow$ sequentially compact.

From now on, we will suppose that (X,d) is a <u>Polish space</u>.

complete, separable metric space.

Thm (Prokhorov): Given a Polish space (X,d) and $\mathcal{K} \subseteq \mathcal{P}(X)$,

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

$\Updownarrow$

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

(Will give a rough sketch of proof next time.)

(Will explain connection to uniform integrability.)

Cor: Given a Polish space (X,d), for all $\mu \in \mathcal{P}(X)$, $\{\mu\}$ is tight.

Pf: This is an immediate consequence of Prokhorov's theorem, since $\mathcal{K} = \{\mu\}$ is clearly relatively compact.