Lecture 3

Recall:

$t\#\mu = \nu \iff \nu(B) = \mu(t^{-1}(B)) \quad \forall \; B \in \mathcal{B}(Y)$

$\iff \int_X \varphi(t(x))\, d\mu(x) = \int_X \varphi(y)\, d\nu(y) \quad \forall\; \varphi \in L^1(\nu)$

Borel meas.

Monge's Optimal Transport Problem:

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

constraint set

$$\min_{\{t : t\#\mu = \nu\}} \underbrace{\int d(x, t(x))\, d\mu(x)}_{\text{objective function}}$$

Reasons the Monge problem is difficult:

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

<u>Def</u>: $\mu \in \mathcal{P}(X)$ is an <u>empirical measure</u> if $\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}$ for $\{x_i\}_{i=1}^{N} \subseteq \mathbb{R}^d$.

Exercise: If μ is an empirical measure, then for any transp. map t, $t\#\mu$ is an emp. meas.

Difficulty #2: solutions may not be unique

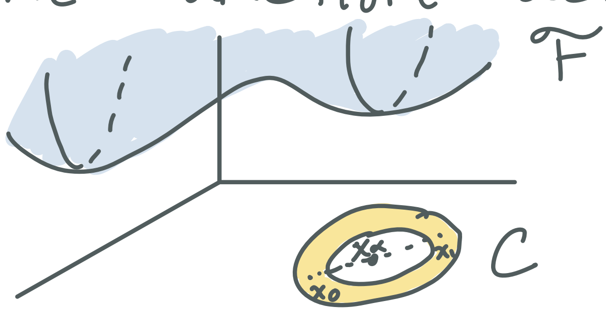
Difficulty #3: the constraint set is nonconvex (along linear interpolations)

Recall:

Def: A subset C of a vector space X is convex (along lin. interp.) if, $\forall\, x_0, x_1 \in C$,

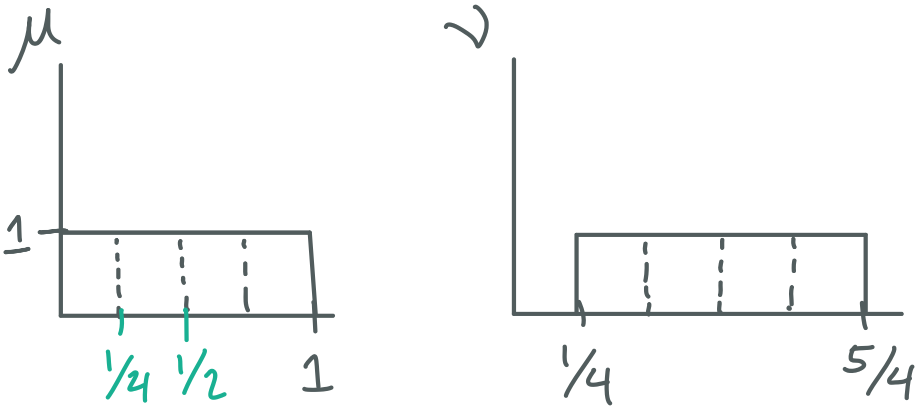
$$x_\alpha := (1-\alpha)x_0 + \alpha x_1 \in C, \quad \forall\ \alpha \in [0,1].$$

Generally, in optimization, we want our constraint set C to be convex, since our normal strategy is to take an initial guess, perturb it, and see if the objective function decreases.



(linear) perturbations can kick us out of the constraint set

Ex: Consider "books on a shelf" example from last time.



$t_0(x) = x + \frac{1}{4}$

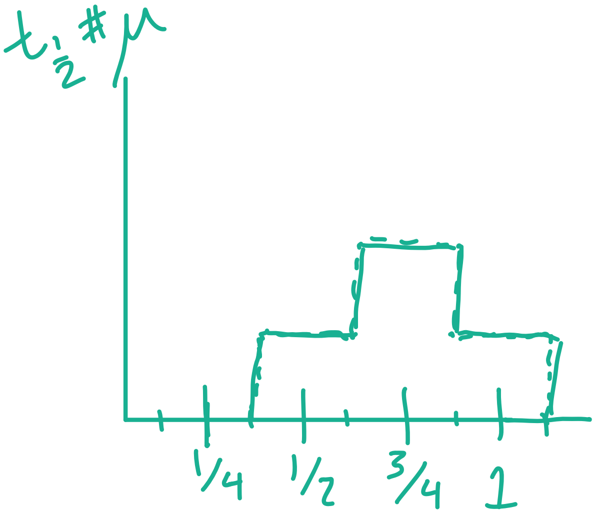
$$t_1(x) = \begin{cases} x + 1 & \text{if } x \in [0, 1/4) \\ x & \text{otherwise} \end{cases}$$

Consider a convex combination:

$t_\alpha(x) = (1-\alpha) t_0(x) + \alpha t_1(x)$

For example, $t_{\frac{1}{2}}(x) = \begin{cases} x + \frac{5}{8} & \text{if } x \in [0, \frac{1}{4}] \\ x + \frac{1}{8} & \text{otherwise} \end{cases}$

Then,



Moral: Even though $t_0\#\mu=\nu$ and $t_1\#\mu=\nu$, we do not have $t_\alpha\#\mu=\nu$ for all $\alpha\in[0,1]$.

That is, $\{t : t\#\mu=\nu\}$ is not convex.

Solution: consider transport plans

In fact, finding t s.t. $t\#\mu=\nu$ relates to well-known problems in geometric PDE.

<u>Prop</u> (change of variables formula): Suppose $\mu\in L^1(\mathbb{R}^d)\cap P(\mathbb{R}^d)$ and $t:\mathbb{R}^d\to\mathbb{R}^d$ is C^1, <u>one to one</u> and $\det Dt\neq 0$. Then,

$t^{-1}\circ t = id$
$t^{-1}(t(x))=x$

$$d(t\#\mu)(y) = \frac{\mu}{|\det Dt|} \circ t^{-1}(y) \underbrace{\mathbb{1}_{t(\mathbb{R}^d)}(y)\, dy}_{d\lambda(y)|_{t(\mathbb{R}^d)}}$$

Pf: By definition of the pushforward, for all bounded, meas φ,

$$\int_{\mathbb{R}^d} \varphi(y)\, d(t\#\mu)(y)$$

$$= \int_{\mathbb{R}^d} \varphi \circ t(x)\, d\mu(x)$$

$$= \int_{\mathbb{R}^d} \varphi \circ t(x)\, \mu(x)\, dx$$

$$= \int_{\mathbb{R}^d} \varphi \circ t(x)\, \mu \circ t^{-1} \circ t(x) \frac{|\det Dt(x)|}{|\det Dt| \circ t^{-1} \circ t(x)}\, dx$$

change of variables thm

$y = t(x)$

$dy = |\det Dt|(x) dx$

$$= \int_{t(\mathbb{R}^d)} \varphi(y)\, \mu \circ t^{-1}(y) \frac{1}{|\det Dt| \circ t^{-1}(y)}\, dy$$

$$= \int_{\mathbb{R}^d} \varphi(y) \left(\frac{\mu}{|\det Dt|}\right) \circ t^{-1}(y)\, \mathbb{1}_{t(\mathbb{R}^d)}(y)\, dy$$

In particular, for $\varphi = \mathbb{1}_B$, $B \in \mathcal{B}(x)$.

Note that if $B \subseteq t(\mathbb{R}^d)^c$, then $(t\#\mu)(B) = 0$.

Exercise: Suppose $\mu \in L^1(\mathbb{R}^d)$ and $t(x) = ax+b$ for $a>0$, $b \in \mathbb{R}^d$. Prove that $d(t\#\mu)(y) = \frac{1}{a^d}\mu\left(\frac{y-b}{a}\right)dy$.

Thus, if $t = \nabla\varphi$, the above proposition ensures that $t\#\mu = \nu$ iff

$$\frac{\mu(x)}{|\det D^2\varphi(x)|} = \nu(\nabla\varphi(x)) \underset{\text{if } \nu>0 \text{ on } t(\mathbb{R}^d)}{\Leftrightarrow} |\det D^2\varphi|(x) = \frac{\mu(x)}{\nu(\nabla\varphi(x))}$$

Given μ, ν, we would seek to solve for φ such that the above equation holds. We will often restrict to φ s.t. $\det D^2\varphi(x) > 0$.

This is a type of Monge Ampére equation:

$$\begin{cases} \text{Find } \varphi \text{ s.t. } \det D^2\varphi(x) = F(x, \varphi(x), \nabla\varphi(x)). \end{cases}$$

How can we get around the difficulties of Monge's problem?

Relax the problem.

Red Plenty

Leonid Kantorovich, 1942
"On the translocation of masses"

Notation:

Projection maps: for $i=1,2$, define $\pi^i : \mathcal{X}\times\mathcal{X}\to\mathcal{X}$ by $\pi^1(x^1,x^2)=x^1$, $\pi^2(x^1,x^2)=x^2$.

Marginals: for $\gamma\in\mathcal{P}(\mathcal{X}\times\mathcal{X})$, its first and second marginals are $\pi^1\#\gamma(A)=\gamma((\pi^1)^{-1}(A))$ $=\gamma(A\times\mathcal{X})$ and $\pi^2\#\gamma(B)=\gamma(\mathcal{X}\times B)$.

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

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

We will use transport plans as a new way to model rearranging mass in μ to look like ν.

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

How do transport plans relate to transport maps?

Lemma: Given $\mu, \nu \in \mathcal{P}(\mathcal{X})$, if $t \# \mu = \nu$, then $\gamma := (id \times t) \# \mu \in \Gamma(\mu, \nu)$.

Pf: By definition,

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

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

$$= \mu(\{x \in A : t(x) \in B\})$$

Then, for all $A \in \mathcal{B}(\mathcal{X})$

$(\pi^1 \# \gamma)(A) = \gamma(A \times \mathcal{X}) = \mu(A)$, so $\pi^1 \# \gamma = \mu$.

Similarly, $\pi^2 \# \gamma(B) = \gamma(\mathcal{X} \times B) = \mu(t^{-1}(B)) = \nu(B)$ for all $B \in \mathcal{B}(\mathcal{X})$, so $\pi^2 \# \gamma = \nu$.

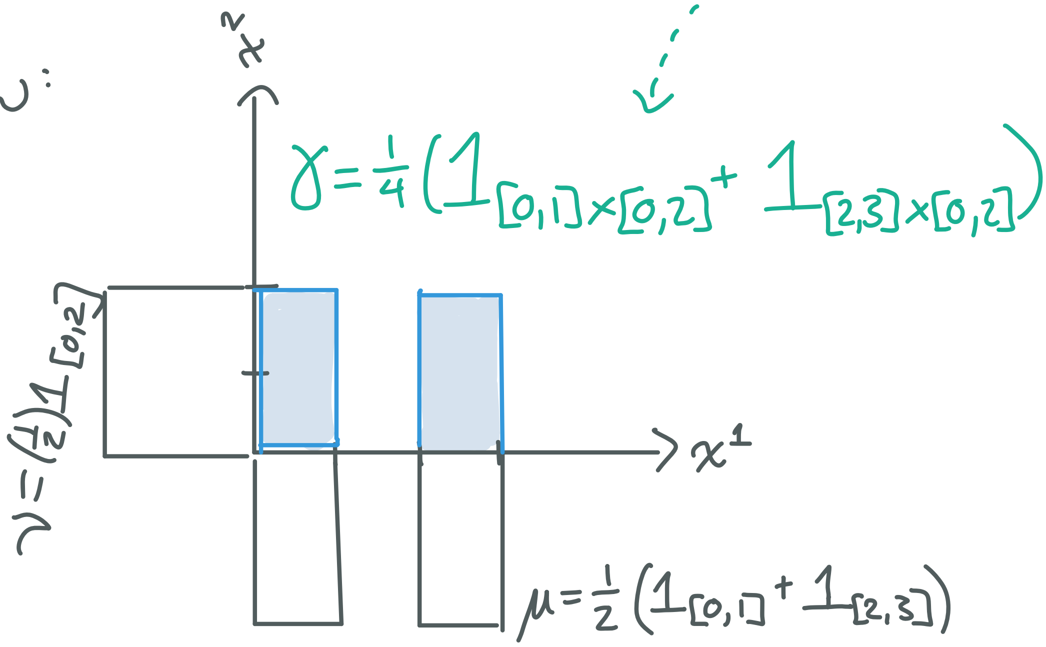
$\square$

Visualizing transport plans

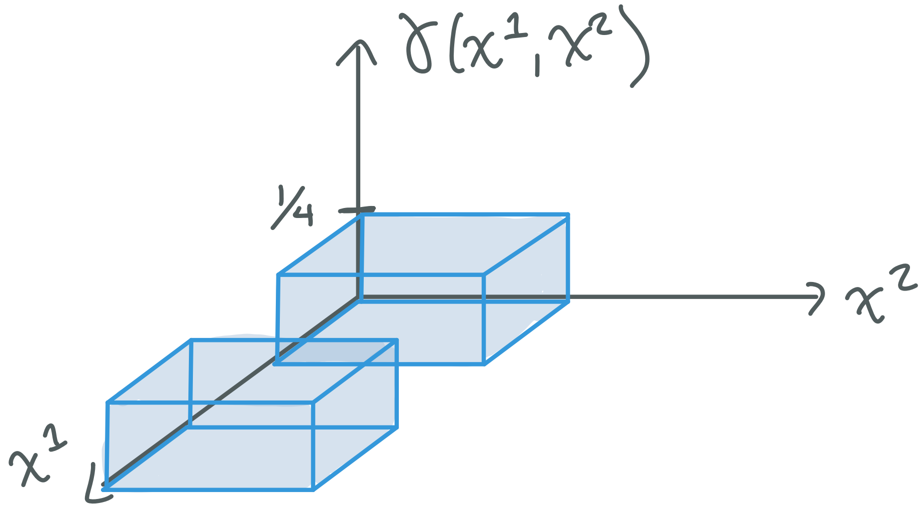
Ex: For μ, ν as below, consider the transport plan "where all mass starting at location x_0 in μ is distributed evenly in ν."

$\gamma(A \times B) = \mu(A)\nu(B)$

Bird's eye view:



Side view:



Remark: This example illustrates the fact that, for any $\mu, \nu \in \mathcal{P}(X)$, there exists $\gamma \in \Gamma(\mu, \nu)$ given by $\gamma := \mu \otimes \nu$.

$$\mu\otimes\nu(A\times B)=\mu(A)\nu(B)$$

For any $\mu,\nu\in\mathcal{P}(X)$, the transport plan $\gamma=\mu\otimes\nu$ "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: ① $\forall\,\mu,\nu\in\mathcal{P}(X)$, $\Gamma(\mu,\nu)\neq\emptyset$

② 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

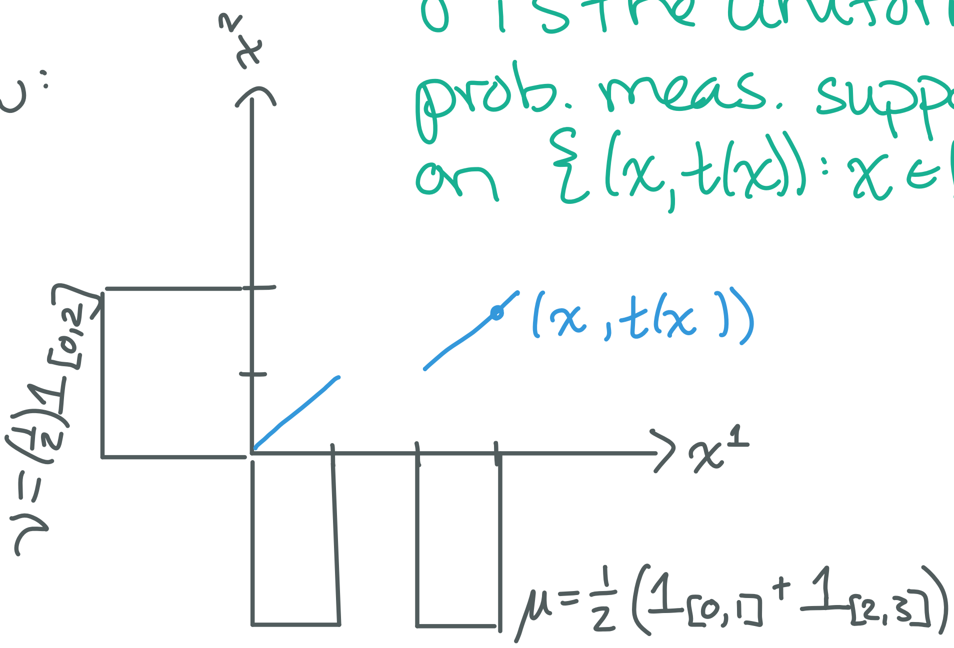
$$t(x)=\begin{cases} x & \text{if } x\in[0,1] \\ x-1 & \text{if otherwise}\end{cases}$$

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

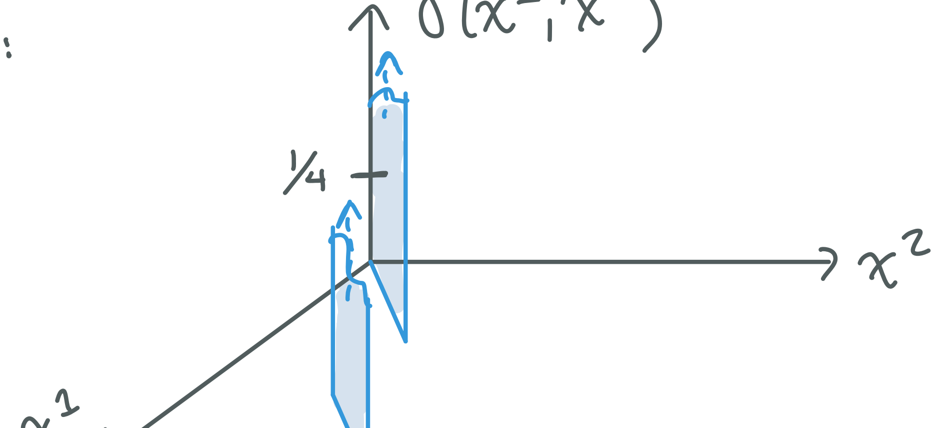
Mass starting at x_0 is only sent to $t(x_0)$.

Bird's eye view:

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



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 $t(x)$.