Lecture 2

Recall:

Given $\mu \in \mathcal{P}(X)$, $B \in \mathcal{B}(X)$,
$\mu(B)$ = amt. of dirt in the pile μ that lies in B.

If $(X,d) = (\mathbb{R}^d, |\cdot|)$ and $\mu \ll \lambda$, $d\mu(x) = \mu(x)dx$
and $\mu(B) = \int_B \mu(x)dx$ = amt. of dirt in B.

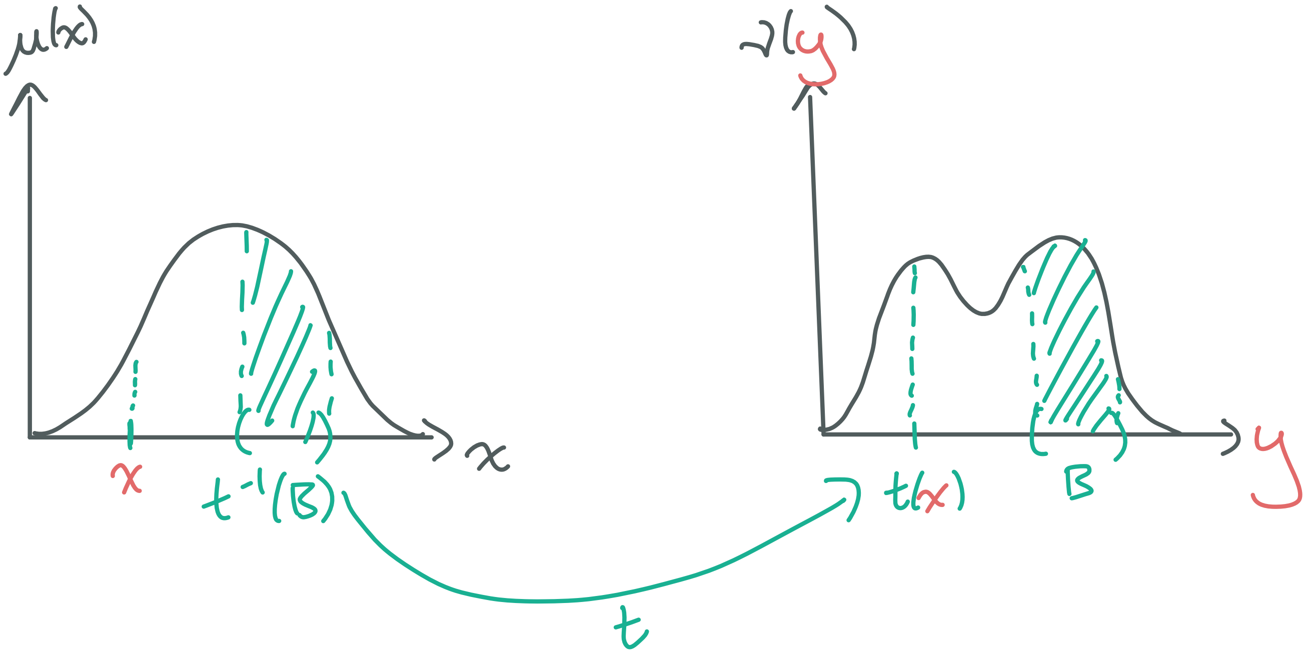
What does it mean to "rearrange one probability measure to look like another"?

$\forall B \in \mathcal{B}(X)$, $t^{-1}(B) \in \mathcal{B}(X)$

Def (transport map): Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$, a measurable function $t: X \to Y$ transports μ to ν if

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

We call ν the pushforward of μ under t, writing $\nu = t\#\mu$, and we call t the transport map from μ to ν



"the amount of mass that ν assigns to B equals the amount of mass sent there from μ"

Informally, "mass starting at location x in μ is sent to location $t(x)$ in ν"

<u>Sanity check</u>: if $\mu \in \mathcal{P}(X)$ and $t: X \to Y$ is measurable, is $t\#\mu$ always a prob measure?

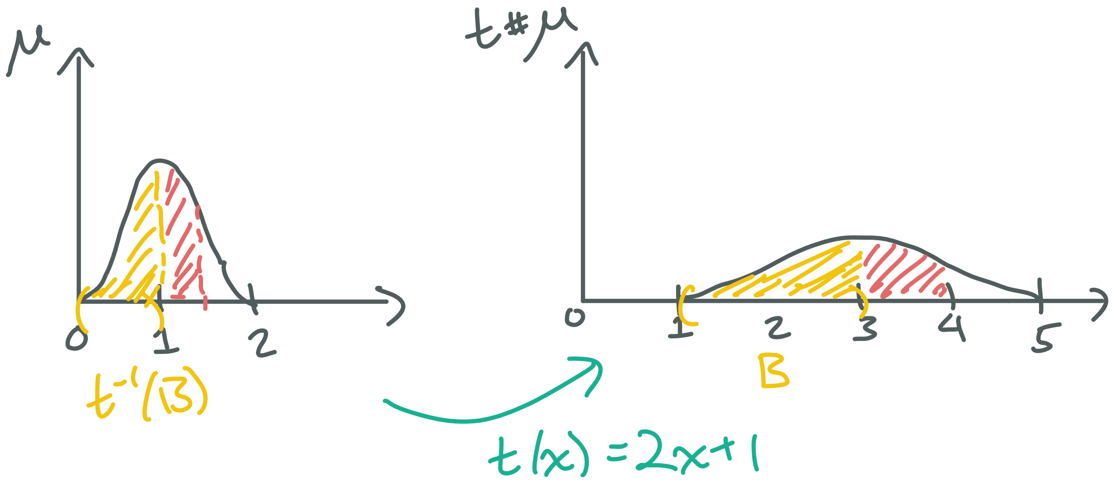
$$(t\#\mu)(Y) = \mu(t^{-1}(Y)) = \mu(X) = 1.$$

Ex (translation/dilation): Suppose $(X,d) = (\mathbb{R}^d, |\cdot|)$. Fix $a > 0$, $b \in \mathbb{R}^d$ and $t(x) = ax + b$.

Then for any $\mu \in \mathcal{P}(X)$, $t\#\mu$ satisfies

dilation translation

$$(t\#\mu)(B) = \mu(t^{-1}(B)) = \mu\left(\frac{B-b}{a}\right) = \mu\left(\left\{\frac{y-b}{a} : y \in B\right\}\right), \ \forall B \in \mathcal{B}(X).$$



<u>Lemma</u> (equiv characterization of transp. map)

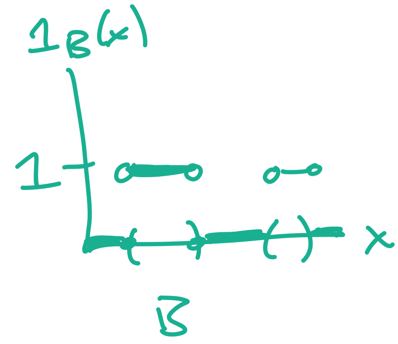
Given $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$ and $t: X \to Y$ measurable, then $t\#\mu = \nu$ if and only if

$$\int_X \varphi(t(x))\, d\mu(x) = \int_Y \varphi(y)\, d\nu(y) \quad \text{for all } \varphi: Y \to \mathbb{R} \text{ measurable, } \varphi \in L^1(\nu)$$

(*)

Before we prove the lemma, recall:

- For any $B \in \mathcal{B}(\mathcal{X})$, define the indicator fn



$$1_B(x) = \begin{cases} 0 & \text{if } x \notin B \\ 1 & \text{if } x \in B \end{cases}$$

- For any ~~bdd, meas~~ nonneg, integrable fn φ, there exists a sequence of simple functions $f_n(x) = \sum_{i=1}^{n} c_{i,n} 1_{B_{i,n}}(x)$ s.t. $f_n \nearrow \varphi$ pointwise.

Pf: First, note that if φ is an indicator fn, then, using the fact

$$\varphi(t(x)) = 1_B(t(x)) = \begin{cases} 0 & \text{if } t(x) \notin B \\ 1 & \text{if } t(x) \in B \end{cases} = 1_{t^{-1}(B)}(x),$$

equation (*) becomes

$$\int 1_B(t(x))\, d\mu(x) = \int 1_{t^{-1}(B)}(x)\, d\mu(x) = \mu(t^{-1}(B))$$

$$\parallel$$

$$\int 1_B(y)\, d\nu(y) = \nu(B).$$

Thus:

(i) If eqn (*) holds for all φ measurable with $\varphi \in L^1(\nu)$, then it must hold for

$\varphi = 1_B$, so the above remark gives
$\nu(B) = \mu(t^{-1}(B))$.

(ii) If $t\#\mu = \nu$, the above remark shows that (*) holds for all indicator functions.

Thus, (*) for all φ meas w/ $\varphi \in L^1(\nu)$ implies $t\#\mu = \nu$.

Now, assume $t\#\mu = \nu$.
We have (*) for all indicator fns, $\varphi = 1_B$.
Furthermore, by linearity of the integral, (*) holds for all simple functions φ.

Next, suppose φ is a ~~bdd, meas~~ nonneg, integrable fn.
Choose a sequence φ_n of simple fns so that $\varphi_n \nearrow \varphi$ pointwise. Thus by the ~~dominated~~ monotone convergence theorem,

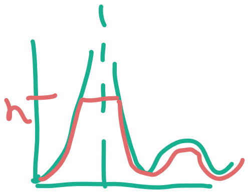
$$\int \varphi(t(x))\,d\mu(x) = \lim_{n\to\infty} \int \varphi_n(t(x))\,d\mu(x)$$

$$\overset{(*)}{=} \lim_{n\to\infty} \int \varphi_n(y)\,d\nu(y)$$

$$= \int \varphi(y)\,d\nu(y)$$

Thus, (*) holds for all φ ~~bdd, meas~~ nonneg, integrable.

Next, suppose φ is a nonnegative, meas fn in $L^1(\nu)$, and define



$$\varphi_n(x) = \varphi(x) \wedge n = \min(\varphi(x), n)$$

Then, by the monotone convergence theorem,

$$\int \varphi(t(x))\, d\mu(x) = \lim_{n\to\infty} \int \varphi_n(t(x))\, d\mu(x)$$

$$\overset{(*)}{=} \lim_{n\to\infty} \int \varphi_n(y)\, d\nu(y)$$

$$= \int \varphi(y)\, d\nu(y)$$

Finally, for φ an arbitrary meas fn in $L^1(\nu)$, the result holds by writing,

$$\varphi(x) = \varphi_+(x) - \varphi_-(x) = \underbrace{\varphi(x) \vee 0}_{\max(\varphi(x),0)} - (-\varphi) \vee 0. \square$$

Now, we know what it means to "rearrange one measure to look like another", or, more precisely, to transport one measure to another.

Back to original question: how can this be done in the most efficient way?

Monge's Optimal Transport Problem:

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

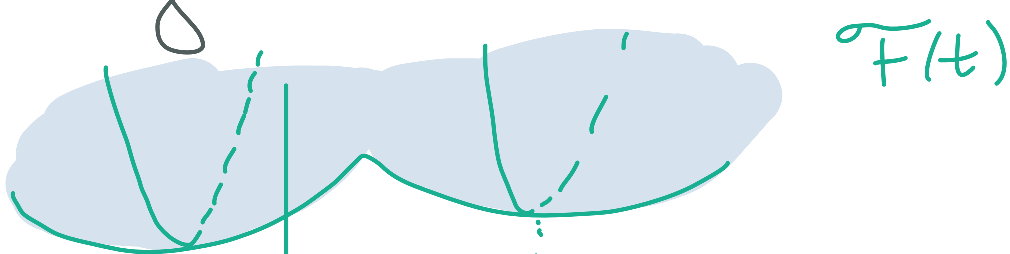
$$\min_{\substack{t: X \to X \text{ measurable} \\ t\#\mu = \nu}} \underbrace{\left\{ \int \overbrace{d(x, t(x))}^{\text{how far dirt is moved}} \underbrace{d\mu(x)}_{\text{how much dirt}} \right\}}_{\text{effort to rearrange } \mu \text{ to look like } \nu \text{ via the transport map } t}$$

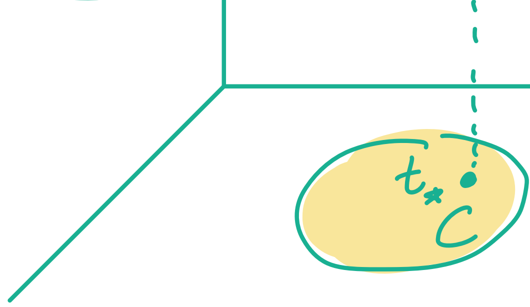
Throughout the course, we'll see many optimization problems of this form:

$$\min_{t \in C} \overbrace{\tilde{F}(t)}^{\text{objective function}}$$

constraint set

Mental image:





Unfortunately, Monge's problem is a horrible optimization problem!

Sudakov 1979, Ambrosio and Pratelli 2001
Evans and Gangbo 1999

Reasons the Monge Problem is difficult:

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

That is, given $\mu, \nu \in P(X)$, there doesn't necessarily exist any t meas s.t. $t\#\mu = \nu$.

Recall: δ_{x_0} is the probability measure $\delta_{x_0}(B) = \begin{cases} 0 & \text{if } x_0 \notin B \\ 1 & \text{if } x_0 \in B \end{cases}$

Ex:

$\frac{1}{2}$ $\qquad$ $0 \quad \frac{1}{2} \quad 1$

If $t^{\#}\mu = \nu$, then

$$\lambda(B \cap [0,1]) = \nu(B) = \mu(t^{-1}(B)) = \begin{cases} 0 & \text{if } t(\frac{1}{2}) \notin B \\ 1 & \text{if } t(\frac{1}{2}) \in B \end{cases}$$

There is no such t for which this holds.

Heuristically, the problem is that a transport map t sends all mass starting at a location x_0 to $t(x_0)$. In particular, mass cannot split.

On the other hand, note that $t(x) = \frac{1}{2}$ satisfies

$$(t\#\nu)(B) = \nu(t^{-1}(B)) = \begin{cases} \nu(\emptyset) & \text{if } \frac{1}{2} \notin B \\ \nu(X) & \text{if } \frac{1}{2} \in B \end{cases} = \delta_{\frac{1}{2}}(B) = \mu(B).$$

Two potential solutions to empty constraint set:

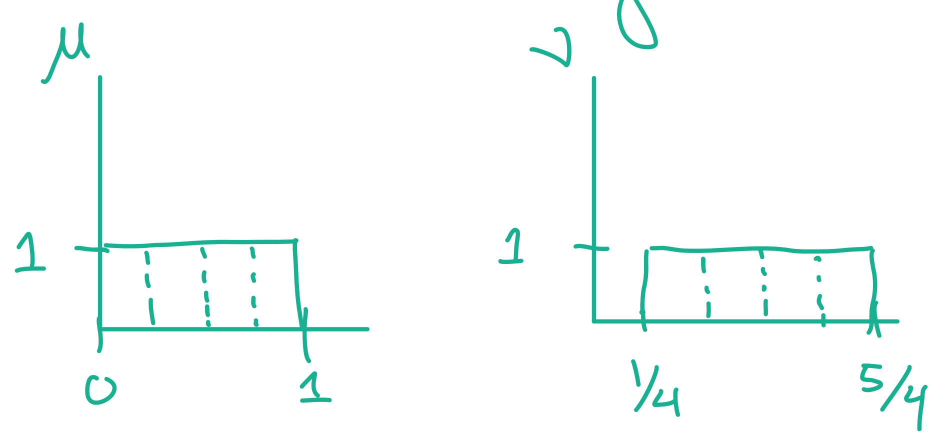
(a) don't allow source measure to concentrate mass on "small sets"

(b) instead of considering transport maps, consider transport plans.

Difficulty #2: Solutions may not be unique.

That is, given $\mu, \nu \in P(X)$, there may exist multiple, distinct optimal transport maps.

Ex: "books on shelf"



Consider $t_0(x) = x + \frac{1}{4}$ "shifting all books to right"

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

Exercise: $t_0 \# \mu = \nu$ and $t_1 \# \mu = \nu$, so both t_0 and t_1 belong to the constraint set.

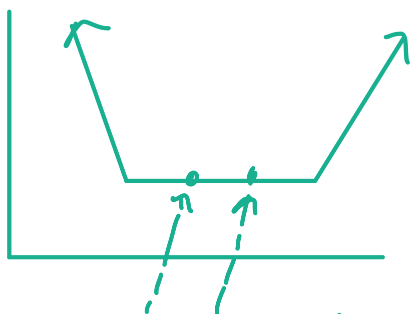
Fact (will show later): t_0 and t_1 are <u>both</u> optimal transport maps.

Potential solution to nonuniqueness of optima:
modify effort to be "strictly convex"

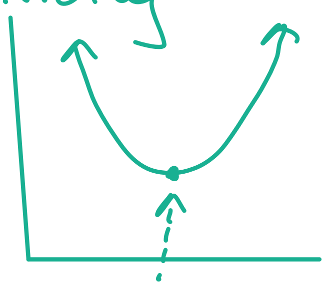
$$\int |t(x)-x| \, d\mu(x) \Rightarrow \left(\int |t(x)-x|^p \, d\mu(x) \right)^{1/p}, \; p>1$$

convex

strictly



many optima



unique optimum

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

Recall:

Def: A subset C of a vector space X is <u>convex</u> 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.