Lecture 2

Cyclical Monotonicity & the Kantorovich Problem

Previously:

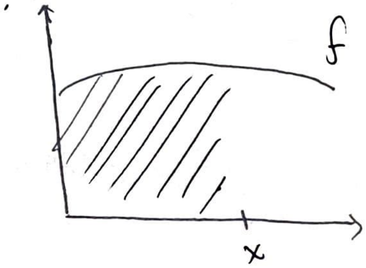
- we saw that ~~##~~ in 1-D the optimal map must be monotone

TODAY: Looking 1D continuous situation

In 1-Dimension

- sol'n is monotone
- can we construct it? (a map)

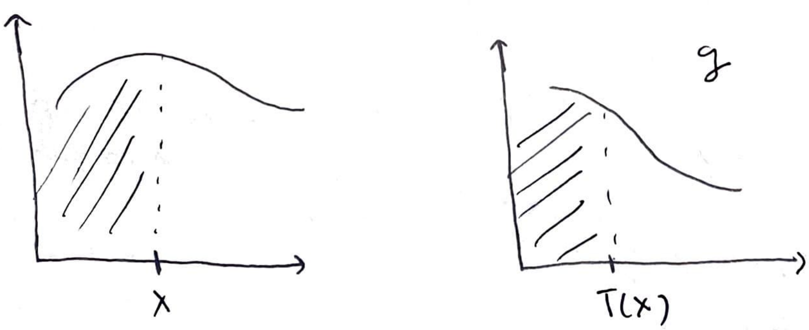
→ using CDFs, given any point x, how much mass came before



$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

* CDFs are obviously monotone

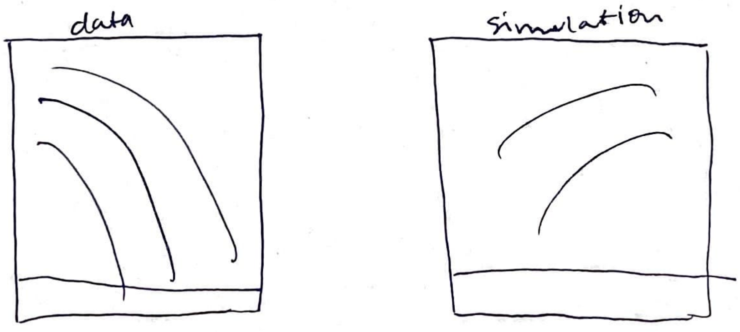
$$G(y) = \int_{-\infty}^{y} g(t)\,dt$$



Expect: ~~that~~ $F(x) = G(T(x))$ ⟹ The mass before x should equals mass before T(x)

→ This gives exact sol'n:

$$T(x) = G^{-1}(F(x))$$



instead of comparing whole things, cut into 1 dimensional little slices & compare them consider continuous problem in higher dimension $\mathbb{R}^n$ (meaning we have density func)

GOAL: Find a map $T(x)$ to minimize

$$\frac{1}{2}\int_X |x-T(x)|^2 f(x)\,dx$$

s.th

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

mass conservation constraint

REVIEW CHANGE OF VARIABLES

want to characterize constraint & learn properties ~~of map~~

So, we do change of variables $y = T(x)$

determinant of Jacobian

$$\int_{T^{-1}(A)} f(x)\,dx = \int_{T^{-1}(A)} g(T(x))\det(\nabla T(x))\,dx$$

* this C.o.V should work for any set

$\therefore$ Expect: $g(T(x))\det(\nabla T(x)) = f(x)$

NOTE we haven't used OT yet. still just conservation of mass

choose a collection of points $x_1, x_2, \ldots, x_N \in X$

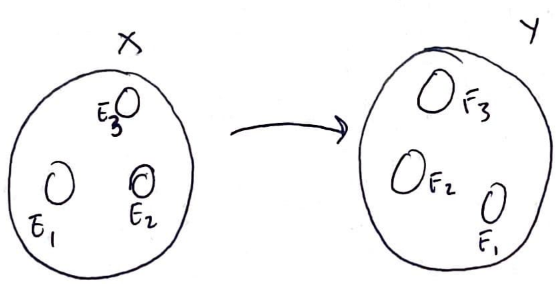
suppose that T is optimal

Let $y_i = T(x_i)$

Let E_i be a ball centered @ x_i

s.th $\int_{E_i} f(x)\,dx = \varepsilon$

and let $F_i = T(E_i)$



Now want to create a new mapping that's measure preserving everywhere except inside little balls, $E_1, E_2, \ldots$
Inside little balls we switch things around

Now we permute mapping cyclically $E_2 \to F_3$, $E_1 \to F_2$, $E_3 \to F_1$

- Let's create a new map $\tilde{T}$ that is still measure preserving and we want:

$$\tilde{T}(x_i) = y_{i+1}$$

$$\tilde{T}(E_i) = F_{i+1}$$

$$\tilde{T}(x) = T(x) \text{ if } x \notin \bigcup_{i=1}^{N} E_i$$ when $N = 3$ in this case

T is optimal, so we know

$$\frac{1}{2}\int_X |x - T(x)|^2 \cancel{f(x)dx} \le \frac{1}{2}\int_X |x - \tilde{T}(x)|^2 f(x)\, dx$$

write down cross terms and divide through by epsilon, so

- As in 1-D:

$$\frac{1}{\varepsilon}\sum_{i=1}^{N}\int_{E_i} x\cdot(\tilde{T}(x) - T(x))\, f(x)\, dx \le 0$$

How is monotonicity implied here???

As $\varepsilon \to 0$

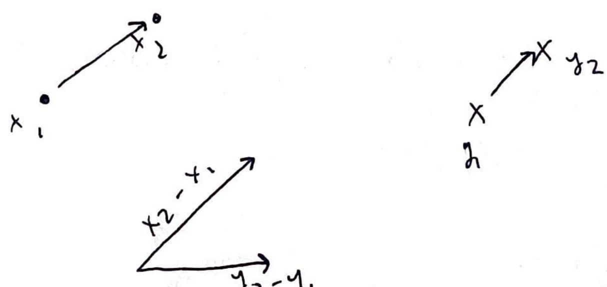
$$\Rightarrow \sum_{i=1}^{N} x_i(y_{i+1} - y_i) \le 0$$

This ~~property~~ is called cyclical monotonicity

If $N = 2$:

$$x_1(y_2 - y_1) + x_2(y_1 - y_2) \le 0$$

$$\Rightarrow (x_2 - x_1)\cdot(y_2 - y_1) \ge 0$$



- They make an acute angle

→ This must hold for all $N \Rightarrow$ placing restriction on how much we can "twist" mass around when we move it.

This is a stronger condition than being irrotational.

NOTE: An irrotational map can be written as the gradient of a scalar

Thm (Rockefelar): A ~~cyclact~~ cyclically monotone map can be expressed as the gradient of a ~~conv~~ convex function. (???)

this is important in leading us to the Monge-Ampere eqn

Def (convex function)

A function f is convex if $\forall$ $x, y \in$ domain (should be convex sets) and $\lambda \in [0, 1]$

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

→* in other words, a convex function has secant lines that are above the function

from Rockerfeller thm

We can write $T(x) = \nabla U(x)$. where U is convex. ← *This is the part that comes from our optimality condition

From mass-preservation:

$$\det(\nabla T(x)) = f(x)/g(T(x))$$

hessian (2nd order gradient)

$$\Rightarrow \quad \boxed{\det(D^2 U(x)) = f(x)/g(\nabla U(x))}$$

2nd order elliptic eqn

Monge - Ampere eqn

← takes optimization problem and pushes into the realm of partial differential eqn's (PDE)

BIG IDEAS

- optimal mappings have some sort of monotonicity (cyclical, in this case)
- optimal mappings are equal to the gradient of a convex func.

Now we will consider a more rigorous explanation.

Monge vs. Kantorovich

- Monge
  - → more intuitive
  - → feasibility issues
  - → hard to work w/
- Kantorovich
  - → more complex
  - → nicer to work with b/c problems are feasible & problem is sort of linear

Lec3

3

Kantorovich Problem

This measure tells us how much stuff moves from x to y

$$\inf \left\{ \int_{X \times Y} c(x,y)\, d\pi(x,y) \,\middle|\, \pi \in \Pi(\mu, \nu) \right\}$$

measure-preserving constraint where $\Pi(\mu,\nu)$ has the set of measures which contain proper marginals

→ if we take the measure and integrate away all the x dependence it should look like the same measure we started w/ that only depended on the y variable

$\Pi(\mu,\nu)$ is the set of measure whose marginals on X & Y are μ, and ν

- Is this problem feasible? what must be true about data to make it so?

→ This is feasible IFF we have mass balance (i.e $\mu = \nu$)

→ The reason the Kantorovich formulation is more generally feasible is b/c I can rearrange & split mass in any way desired

What makes this case ???

- Is there an infimum?

→ As long as $c(x,y)$ is bounded below an X, Y are bounded, then we have an infimum that is finite

(???)

- Do we actually have a minimum? ← we'll need to use some compactness argument to argue that if we have a sequence that converges to the infimum that the sequence of measures π converges to an actual minimum.

Thm (Weierstrass)

If $f: U \to \mathbb{R}$ is cont. and if U is compact then f attains a minimum on U.

NOTE A compact set is one that's closed & bounded

→ every bounded sequence has a convergent subsequence ???

*Thm *

Suppose $X, Y \subseteq \mathbb{R}^n$ are compact, and that $c(x,y)$ is continuous. Then the Kantorovich Problem attains a min.

PROOF

- Before we can talk about compactness we need a notion of convergence.
  - we identify $U = \Pi(\mu,\nu)$ ~~assume μ & ν~~

    → w/o loss of generality, these are probability measure

* we say that $\gamma_n \to \gamma$ if

$$\int_{X \times Y} g(x,y)\, d\gamma_n(x,y) \xrightarrow{\text{converges}} \int_{X \times Y} g(x,y)\, d\gamma \quad \forall\, g \in C^0(X \times Y)$$

for every g that's continuous

limiting measure (???)

PROOF (cont)

Q1. Is $\Pi(\mu,\nu)$ compact? $\rightarrow$ we have to choose any sequence & show that we can extract a convergent subsequence.

- choose sequence $\pi_n \in \Pi(\mu,\nu)$ $\rightarrow$ * we know π_n is a sequence of prob. measures that are defined on a compact set

  $\rightarrow$ so I know that I can extract something that converges to another probability measure

  $\therefore$ $\pi_{n_k} \rightarrow \pi$, which is also a probability measure

Need to check the marginals & if $\pi \in \Pi(\mu,\nu)$

~~$\rightarrow$ Does $\pi \in \Pi(\mu,$~~ ~~choose any~~

$\rightarrow$ we need to see what these measures do when the ~~y/x~~ y or x dependence is integrated away

choose any $g \in C(X)$ $\leftarrow$ means continuous on X

$\rightarrow$ we hope that if we integrate our limit π against g, we recover the result of integrating μ against g.

$$\boxed{\int_{X\times Y} g(x)\,d\pi(x,y) = \lim_{k\to\infty}\int g(x)\,d\pi_{n_k}(x,y)}$$

we know this is the limit of the ~~#~~ sequence of probability measures, π_{n_k}

this works since $\pi_{n_k} \rightarrow \pi$

so we can show that the marginals are correct

$$= \int_X g(x)\,d\mu(x)$$

since $\pi_{n_k} \in \Pi(\mu,\nu)$, so the marginal over X is μ

same argument $\Rightarrow$ the marginal of π over Y is ν

RECAP

- we started w/ a sequence that lives in the set we care about $\rightarrow \pi_n \in \Pi(\mu,\nu)$
- we found out that there's a subsequence that converges to some other prob. measure
- we showed that the limiting measure π has correct marginals

$\Rightarrow$ all of this shows $\pi \in \Pi(\mu,\nu)$ &

$\therefore \Pi(\mu,\nu)$ is compact

NOTE we are concerned about compactness b/c it is a necessary step in showing that there is a ~~XXXXXX~~ minimum to our cost func. & that the Kantorovich formulation attains that minimum

Q2: Is my value of f continuous?

where

$$f(\pi) = \int_{X\times Y} c(x,y)\, d\pi(x,y) \quad \text{for } \pi \in U$$

def (continuity)

If I have a sequence of π_n that converge to some π, then my $f(\pi_n) \to f(\pi)$

choose any $\pi_n \in \Pi(\mu,\nu)$ s.th $\pi_n \to \pi \in \Pi(\mu,\nu)$.

need to show $f(\pi_n) \to f(\pi)$

$$f(\pi_n) = \int_{X\times Y} c(x,y)\, d\pi_n(x,y)$$

$$\xrightarrow{\text{converges}} \int_{X\times Y} c(x,y)\, d\pi(x,y) \leftarrow \text{since } \pi_n \to \pi \text{ and } c(x,y) \text{ is cont.}$$

$$= f(\pi)$$

$\therefore$ f is continuous.

RECAP

- we are trying to minimize a real-valued function over some set.
  - → the set is compact
  - → the real-valued func is cont.

$\therefore$ we have a minimizer.

- Kantorovich Problem seeks to minimize a cont. real-valued function over a compact set
  - → This is possible by Weierstrass!
  - → so ~~that's~~ we can come up w/ a transport plan that actually minimizes the cost

i.e the Kantorovich problem admits a minimizer.

SUMMARY

Previously, we saw that the OT map in the 1-D discrete case was monotone. We now consider the continuous 1-D case to understand properties of that optimal map. We learn that optimal mappings have some sort of monotonicity and are equal to the gradient of a convex function. Now we take a more detailed look @ the Kantorovich problem for a more rigorous explanation. We learn that the problem is feasible IFF we have mass balance. We also learn that the Kantorovich form. attains a min. if we are working w/ compact set $X, Y \subset \mathbb{R}^n$ & have a continuous cost func. We prove these cond'ns are met & conclude that Kantorovich admits a minimizer.

LEC OUTLINE

- 1-D CONTINUOUS CASE
- SEEK TO ASCERTAIN PROPERTIES OF OPTIMAL MAP
- EXPLORE KANTOROVICH PROBLEM IN DETAIL
  - → FEASIBILITY
  - → INFIMUM
  - → MINIMUM
- SHOW KANTOROVICH FORMULATION ADMITS MINIMIZER