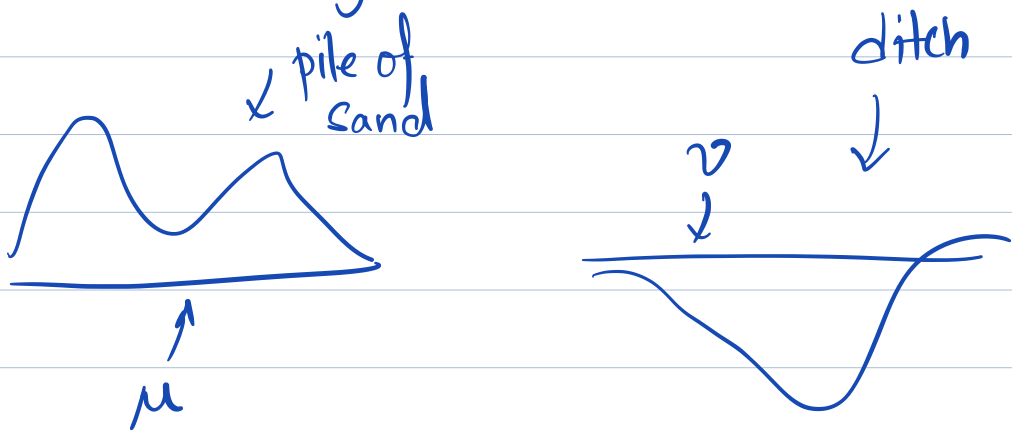
→ Introduction

→ Organization of lectures

Lecture 1:

- → Formulations of OT problem
- → Wasserstein distances
- → Connecting formulations (Brenier's theorem)
- → Duality

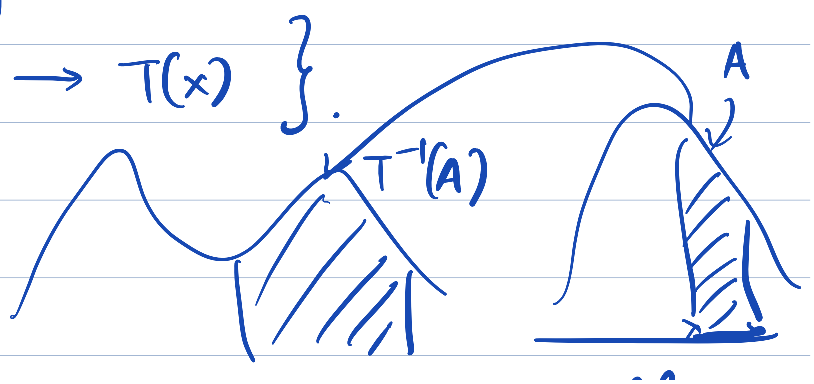
1781 - Monge's formulation



Transporting: $T: \mathbb{R}^d \to \mathbb{R}^d$

$x \to T(x)$ }.

<u>Valid map:</u>



μ ν

$\rightarrow \underline{\mu(T^{-1}(A)) =} \nu(A)$ for every set A

$X \sim \mu$

$T(X) \sim \nu.$

$\rightarrow \quad c(x, T(x)) = \|x - T(x)\|$

Monge's formulation:

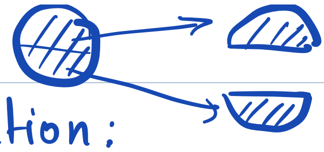
$$\boxed{\text{Find } T: \quad \inf_{\substack{T:\ \text{valid} \\ \underbrace{\boxed{T_\# \mu = \nu}}}} \int \underbrace{\|x - T(x)\|}\, d\mu} \quad \star$$

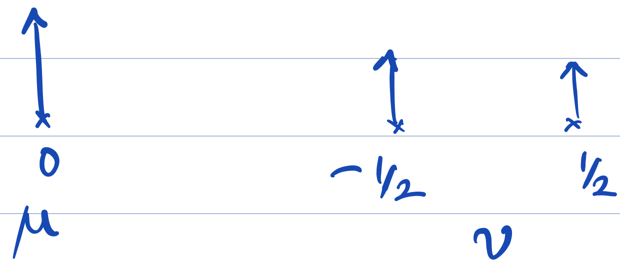
1. Are there always maps?
2. What can I say about the OT map?
   (stable, regular, unique?)

constraint: $\{T_\# \mu = \nu\}$

non-linear, non-convex constraint

for all x : $\{\mu(x) = \nu(T(x)) |\det(\nabla T(x))|\}$

not always a solution: 



No valid map. We need to split atoms.

Kantorovich (1941)

Define a coupling of μ & ν

γ → joint distribution
→ first marginal is μ
→ second marginal is ν.

$$\gamma(A \times \mathbb{R}^d) = \mu(A) \quad \forall A$$
$$\gamma(\mathbb{R}^d \times A) = \nu(A) \quad \forall A$$

Denote set of all couplings as Γ

Kantorovich's program:

Find Γ which ⋆

$$\inf_{\gamma \in \Gamma_{\mu,\nu}} \int \underbrace{\|x-y\|}\ \underbrace{d\gamma(x,y)}$$

1. Constraint set is nice

infinite-dimensional linear program

$$\int_{\mathbb{R}^d} \gamma(x,y)\, dy = \mu(x) \quad \checkmark$$

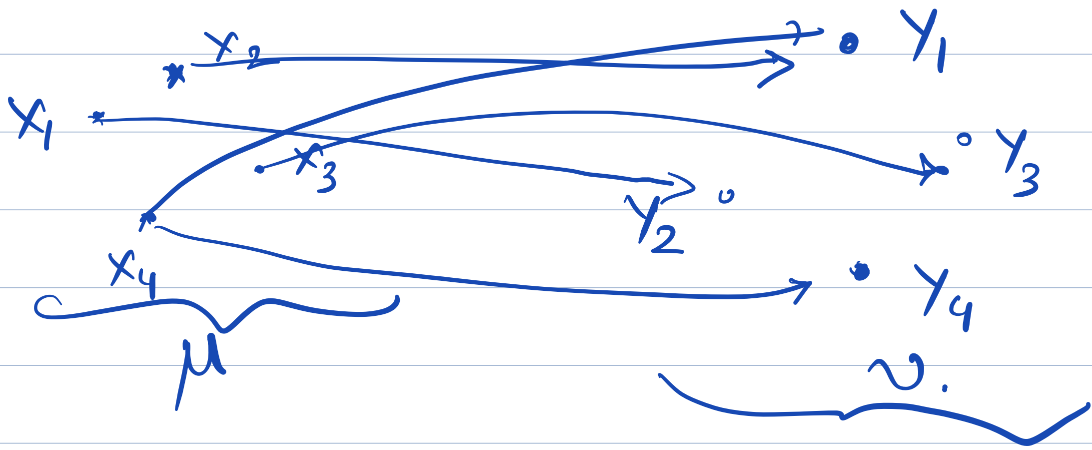
$$\int_{\mathbb{R}^d} \gamma(x,y)\, dx = \nu(y).$$

2. There are always valid couplings

$$\left.\begin{array}{l} X \sim \mu \\ Y \sim \nu \end{array}\right]$$ "independence coupling"

Discrete case:

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \qquad \nu = \frac{1}{n}\sum_{j=1}^{n} \delta_{y_j}$$



$\gamma \to$ matrix $(n \times n)$

constraints: $\gamma_{ij} \geq 0$

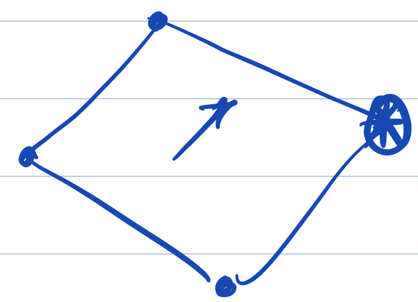
marginal constraints $\to$ $\sum_j \gamma_{ij} = 1/n \quad \forall i$ $\qquad \sum_i \gamma_{ij} = 1/n \quad \forall j$

objective: $c(x,y) = \|x - y\|$

$$\min \sum_{ij} \gamma_{ij}\, c(x_i, y_j)$$

s.t $\gamma \in \Gamma$.

Birkhoff polytope: convex hull of set of permutation matrices.



Monge's problem also has a solution in discrete case.

Wasserstein distances:

Kantorovich's program with ℓ_p cost

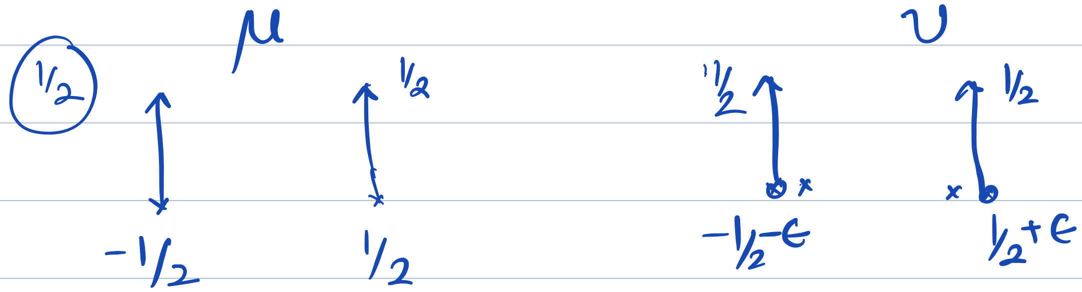
$$\underbrace{c(x,y) = \|x-y\|^p}$$

μ, ν , $\int \|x\|^p d\mu < \infty$, $\int \|y\|^p d\nu < \infty$

$$\boxed{W_p^{\cancel{p}}(\mu,\nu) = \left[\inf_{\gamma \in \Gamma_{\mu,\nu}} \int \|x-y\|^p d\gamma\right]^{1/p}}$$

Wasserstein – cost of the best coupling.

1. Metric on distributions,
2. Respects underlying metric structure



Total variation / KL / χ^2 / Hellinger.

↓ 1. ↓ ∞ — — —

$W_2^2(\mu, \nu) \leq \epsilon^2$.

Monge & Kantorovich

$p = 2$, cost is squared Euclidean

Brenier's theorem: → μ, ν → finite 2^{nd} moments

& μ has a density wrt Lebesgue measure

1. $\exists$ map T_0, $\boxed{T_{0\#}\mu = \nu}$

T_0 is optimal for Monge's program.

→ $T_0 = \nabla \varphi_0$ $\quad \varphi_0$ is a convex function.

$\nabla \varphi_0$ is unique, μ almost everywhere.

2. γ_0 which solves Kantorovich's problem.

$$\gamma_0 = \underbrace{(x, \nabla\varphi_0(x))}.$$

3. If ν also has a density

$$S_{0\,\#}\,\nu = \mu$$

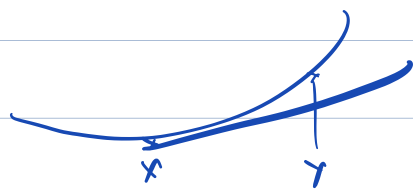
$$S_0 = \nabla\varphi_0^*$$

$$\varphi_0^*(y) = \sup_x \left[\langle x, y\rangle - \varphi_0(x) \right].$$

Why is $T_0 = \nabla\varphi_0$?

Gradient of a convex function $\simeq$ "multivariate monotone".

1D:



$$\rightarrow \quad \left.\begin{aligned} \varphi(y) &\geq \varphi(x) + \nabla\varphi(x)^T (y-x) \\ \varphi(x) &\geq \varphi(y) + \nabla\varphi(y)^T (x-y) \end{aligned}\right\}$$

$$\rightarrow \quad (\nabla\varphi(x) - \nabla\varphi(y))(x-y) \geq 0.$$

$\rightarrow$ if $x \geq y$ then $\underbrace{\nabla\varphi(x) \geq \nabla\varphi(y)}$

(ie) $\nabla\varphi$ is monotonic.

Cyclical monotonicity:

$$x_1 \to x_2 \to x_3 \to \cdots \to x_k$$

$$\boxed{\sum_{i=1}^{k} (x_i - x_{i+1})^T \nabla\varphi(x_i) \geq 0}$$

Duality: (Kantorovich)

Discrete: $\quad \min \sum_{ij} \gamma_{ij} c_{ij}$

$$\alpha_j \to \sum_i \gamma_{ij} = 1/n \quad \forall j$$

$$\beta_i \to \sum_j \gamma_{ij} = 1/n \quad \forall i$$

$$\delta_{ij} \to \gamma_{ij} \geq 0. \quad \forall (i,j)$$

$$\min \max_{\substack{\alpha, \beta, \\ \delta \geq 0}} \sum_{ij} \gamma_{ij} c_{ij} - \sum_j \alpha_j \left(\sum_i \gamma_{ij} - \frac{1}{n} \right) - \sum_i \beta_i \left(\sum_j \gamma_{ij} - 1/n \right) - \sum_{ij} \delta_{ij} \gamma_{ij}$$

constraints:

$$C_{ij} - \alpha_j - \beta_i - \delta_{ij} = 0$$

objective: $\max\limits_{\alpha, \beta} \dfrac{\sum_j \alpha_j}{n} + \dfrac{\sum_i \beta_i}{n}$

$$[\ \underbrace{\alpha_j + \beta_i}\ \leq\ \underbrace{C_{ij}}.\ ]$$

"Shipper's problem".

α_j to unload y_j

β_i to load X_i

→ shipper will solve dual.

strong duality ⇝ clever shipper can set α, β so that I pay full cost.