Optimal Transport Introductio (Simon's Institute)

Topics

→ Introduction

→ Kantorovich's Duality

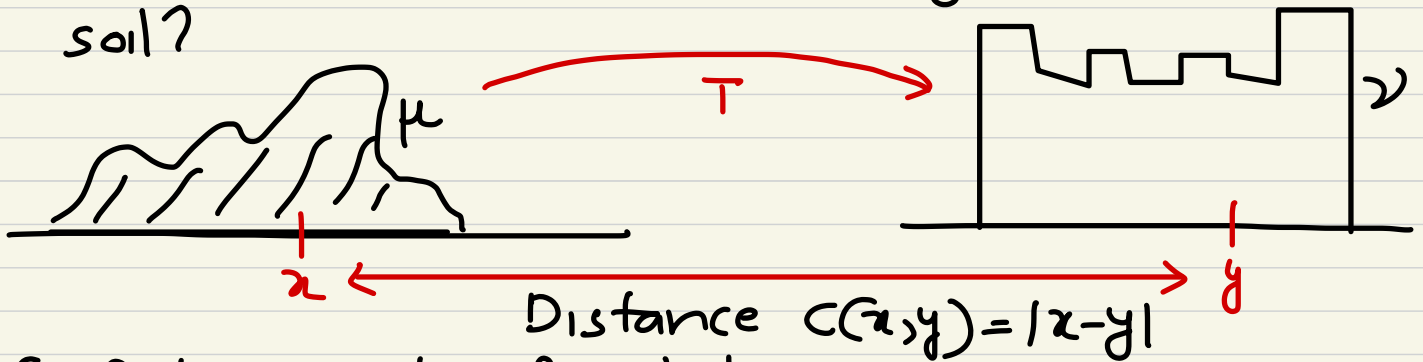
→ Brenier's Theorem

→ Wasserstien Distances

⊛ Monge's Problem

Question Assume one extracts soil for fortification

What is the cheapest possible way to transport the soil?

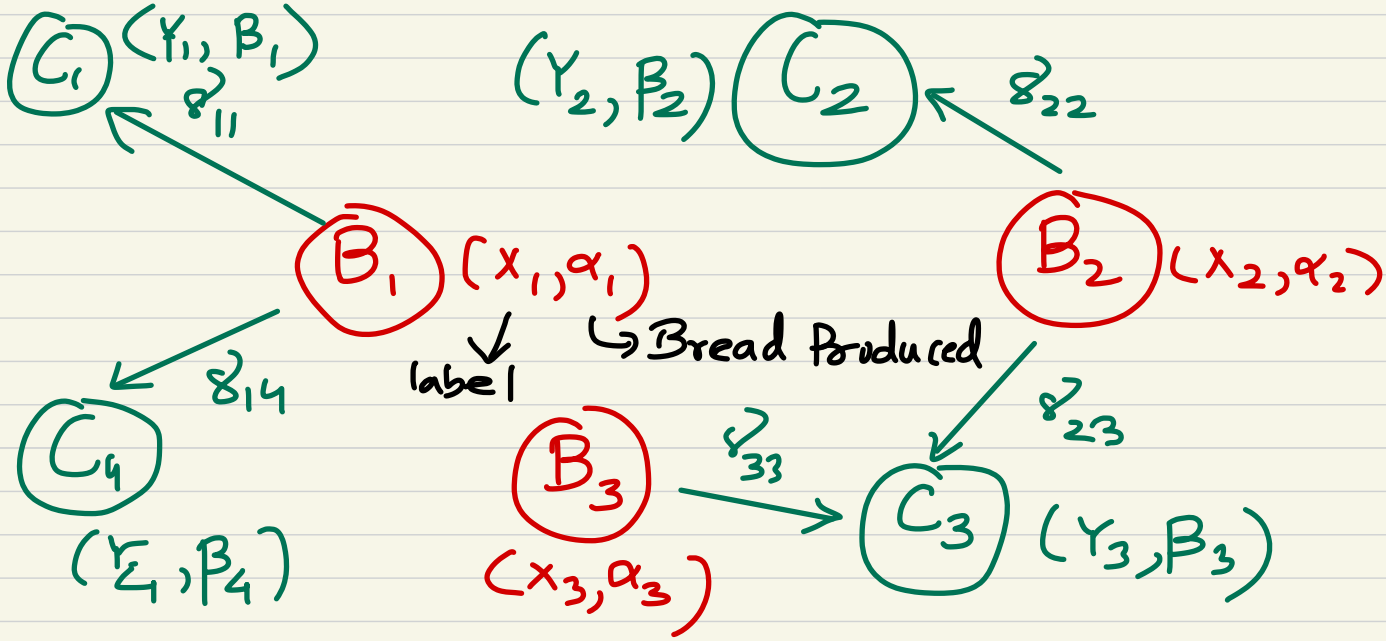


⊛ Deterministic formulation

⊛ Kantorovich's Formulation

⊛ Non-deterministic or

⊛ Bakeries (B_i) $i = 1, 2, \ldots, N$ & Coffe shops (C_j), $j = 1, 2, \ldots, M$



For example $\delta_{33} = \alpha_3$, $\delta_{23} = \beta_3 - \alpha_3$

$\alpha_i \geq 0$ amount of bread produced by B_i

$\beta_j \geq 0$ amount of bread needed by C_j

(*) We assume, $\sum_{j=1}^{M} \beta_j = \sum_{i=1}^{N} \alpha_i = 1$

(*) Given $c(x_i, y_j)$ = cost to move one unit of mass from x_i to y_j

(a) $\gamma_{ij} \geq 0$ (convex constraints)

(b) $\forall i$, $\sum_{j=1}^{M} \gamma_{ij} = \alpha_i$

(c) $\forall j$, $\sum_{i=1}^{N} \gamma_{ij} = \beta_j$

(d) γ_{ij} minimizes, $\sum_{i,j=1}^{N,M} \gamma_{ij}\, c(x_i, y_j)$ ← total cost

Applications

(a) Euler expression, isoperimetric ineq, Sobolev inequality

(b) Evolution of PDEs: $\partial_t u = \lambda u^m \ldots$

(c) Probability & Kinetic Theory

(d) Study of Ricci curvature on Riemanian manifolds

⊛ Commonly, $c(x,y) = \|x-y\|_2^2$ in $\mathbb{R}^d$

⊛ X = space where source measure lives "μ"

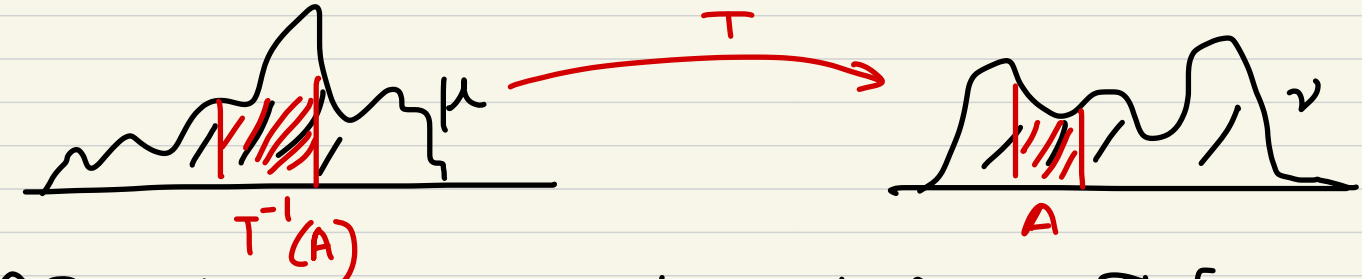
Y = space " target " " "ν"

$P(X)$ = Set of all probability measure over X

Defn: (Push Forward measure) Let $T: X \to Y$, $\mu \in P(X)$

$T_{\#}\mu \in P(Y)$ s.t $(T_{\#}\mu)(A) = \mu(T^{-1}(A))$ $\forall$ A borel set in Y

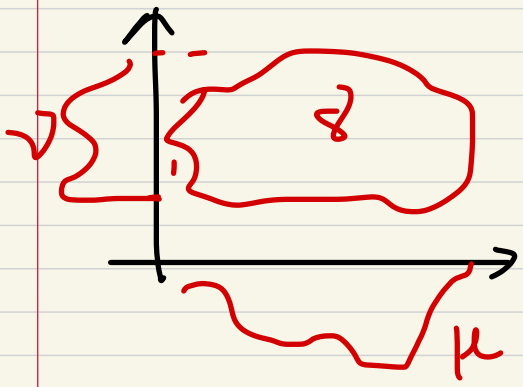
⊛ A map $T: X \to Y$ is called a transport map if $T_{\#}\mu = \nu$



(*) Remark: Given μ, ν, the set of map $\mathcal{T} = \{T \; T_{\#}\mu = \nu\}$ can be empty i.e. T might not exist always

Ex $\mu = \delta_{x_0}$, $T_{\#}\mu = \delta_{T(x_0)}$ Unless ν is a dirac delta, the set of transport map from μ to ν is empty

Defn (coupling or transport plan) We call $\gamma \in P(X \times Y)$ a coupling of μ and ν if $(\pi_X)_{\#}\gamma = \mu$ & $(\pi_Y)_{\#}\gamma = \nu$ where $\pi_X(x,y) = x$, $\pi_Y(x,y) = y$



Let, $\Gamma(\mu,\nu) = \{$set of all couplings of μ & $\nu\}$

Remark Given μ,ν, the set Γ is always non-empty Eg $\mu\otimes\nu \in \Gamma(\mu,\nu)$

Transport map vs coupling

If $T: X\to Y$ be a transport map, then there is a way to create coupling using T Consider $\mathbb{I}_d \times T : X \to X\times Y$

$\gamma_T = (\mathbb{I}_d \times T)_{\#}\mu \in P(X\times Y)$ & $\gamma_T \in \Gamma(\mu,\nu)$

Monge vs Kantorovich formulation

Let $\mu \in P(X)$ & $\nu \in P(Y)$ $c: X\times Y \to [-c_0, +\infty)$ be

a continuous cost. The Monge & Kantorovich cost are

$$C_M(\mu,\nu) = \inf\left\{ \int_X c(x,T(x))\, d\mu(x) \;\middle|\; T_{\#}\mu = \nu \right\}$$

$$C_K(\mu,\nu) = \inf\left\{ \int_{X\times Y} c(x,y)\, d\gamma(x,y) \;\middle|\; \gamma \in \Gamma(\mu,\nu) \right\}$$

(*) Optimal transport map / coupling is the map/coupling that realizes optimal cost of the above

<u>Remark</u> If $T_{\#}\mu = \nu \Rightarrow \gamma_T \in \Gamma(\mu,\nu)$ in addition T & γ_T have the same cost In general

$$\boxed{C_M(\mu,\nu) \geq C_K(\mu,\nu)}$$

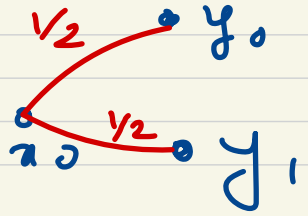
(*) Q Do minimizer's exist? If yes can we characterize them?

Theorem: Let $c : X \times Y \to [-c_0, +\infty)$ be continuous s.t. $\mu \in P(X)$, $\nu \in P(Y)$ $\Rightarrow$ there exists an optimal coupling $\bar{\gamma} \in \Gamma(\mu, \nu)$

Q Is $\bar{\gamma}$ unique?

Q Is $\bar{\gamma}$ induced by a transport map?

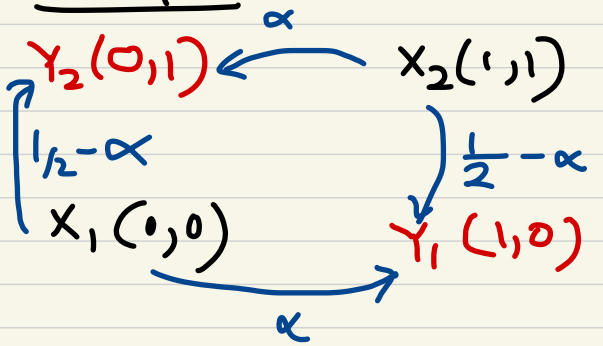
Examples $\mu = \delta_{x_0}$ $\nu = \frac{1}{2}\delta_{y_0} + \frac{1}{2}\delta_{y_1}$



$\exists$ a minimizer/optimal coupling

$\bar{\gamma} = \frac{1}{2}\delta_{(x_0, y_0)} + \frac{1}{2}\delta_{(x_0, y_1)}$ however no transport map

Example: $X = Y = \mathbb{R}^2$, $c(x,y) = |x-y|^2$



Let, $\mu = \frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_2}$

$\nu = \frac{1}{2}\delta_{y_1} + \frac{1}{2}\delta_{y_2}$

Here $\alpha \in [0, \frac{1}{2}]$ Thus,

$$\gamma_\alpha = \alpha\,\delta_{(x_1,y_1)} + (\tfrac{1}{2}-\alpha)\,\delta_{(x_1,y_2)} + (\tfrac{1}{2}-\alpha)\,\delta_{(x_2,y_1)} + \alpha\,\delta_{(x_2,y_2)}$$

* For all $\alpha \in [0, \frac{1}{2}]$ the above couplings optimal
  Thus couplings need not be unique
* As Kantorovich's formulation was a linear optimization under convex constraints – Thus not unique

Kantorovich's Duality

Let $c \in C^0(X \times Y)$, bounded from below. Assume,

$$\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{X \times Y} c \, d\gamma < +\infty \text{ then,}$$

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int c \, d\gamma = \max_{\substack{\psi(x)+\psi(y) \\ +c(x,y) \geq 0}} \left\{ \int_X -\psi \, d\mu + \int_Y -\psi \, d\nu \right\}$$

primal problem — dual problem

Special case, $X = \mathbb{R}^d$, $c(x,y) = |x-y|$, then

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int |x-y| \, d\gamma(x,y) = \max_{\psi \text{ 1-lipschitz}} \left[\int_Y \psi \, d\nu - \int_X \psi \, d\mu \right]$$

Brenier Theorem

Let $X = Y = \mathbb{R}^d$ & $c(x,y) = \|x-y\|^2$ Suppose,

$$\int_{\mathbb{R}^d} |x|^2 \, d\mu + \int_{\mathbb{R}^d} |y|^2 \, d\nu < +\infty$$ & assume $\mu \ll$ lebesgue

Then $\exists$ a optimal transport coupling $\gamma \in \Gamma(\mu,\nu)$ & $\gamma = (\mathrm{Id} \times T)_{\#}\mu$, T is an optimal transport map where $T = \nabla\phi$ where $\phi: \mathbb{R}^d \to \mathbb{R}$ is convex

Remark: The above theorem is true in general with $c(x,y) = |x-y|^p$ however T is not gradient of a convex function

Defn: (probability measures with finite p^{th} moment).

$X = \mathbb{R}^d$, $p \in [1, +\infty)$ then

$$P_p(X) = \left\{ \sigma \in P(X) \cdot \int_X |x|^p \, d\sigma(x) < +\infty \right\}$$

Defn (Wasserstein distance) Given two measures $\mu, \nu \in P_p(X)$ then,

$$W_p(\mu,\nu) = \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \underbrace{\int_{X \times X} |x-y|^p \, d\gamma(x,y)}_{C_K} \right)^{1/p}$$

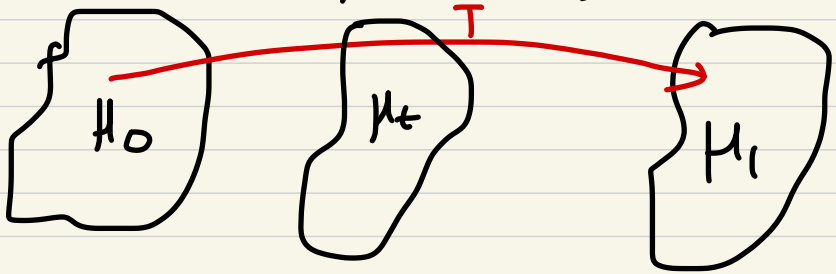
⊛ W_p is a distance on the space $P_p(X)$ Thus, $(P_p(X), W_p)$ are called Wasserstein space

Theorem: $K \subset \mathbb{R}^d$ compact & $p \geq 1$ Let $(\mu_n)_{n \in \mathbb{N}} \in P(K)$ $\mu \in P(K)$, then we call μ_n weakly converges to μ if

$$\mu_n \longrightarrow \mu \iff W_p(\mu_n, \mu) \to 0$$

Construction of Geosedics

Let $X = \mathbb{R}^d$, $\mu_0, \mu_1 \in P(X)$



⊛ T is the OT map between μ_0 & μ_1, then the geodesic μ_t

$$\mu_t = (T_t)_{\#}\mu_0 \quad s.t \quad T_t(x) = (1-t)x - t\,T(x)$$

⊛ Geodesic is the shortest path between two points

⊛ In this case, the geodesic μ_t is constant speed meaning $W_p(\mu_s, \mu_t) = |t-s| W_p(\mu_0, \mu_1) \; \forall s,t \in [0,1]$

⊛ Note that $\mu_t \in P_p(X)$