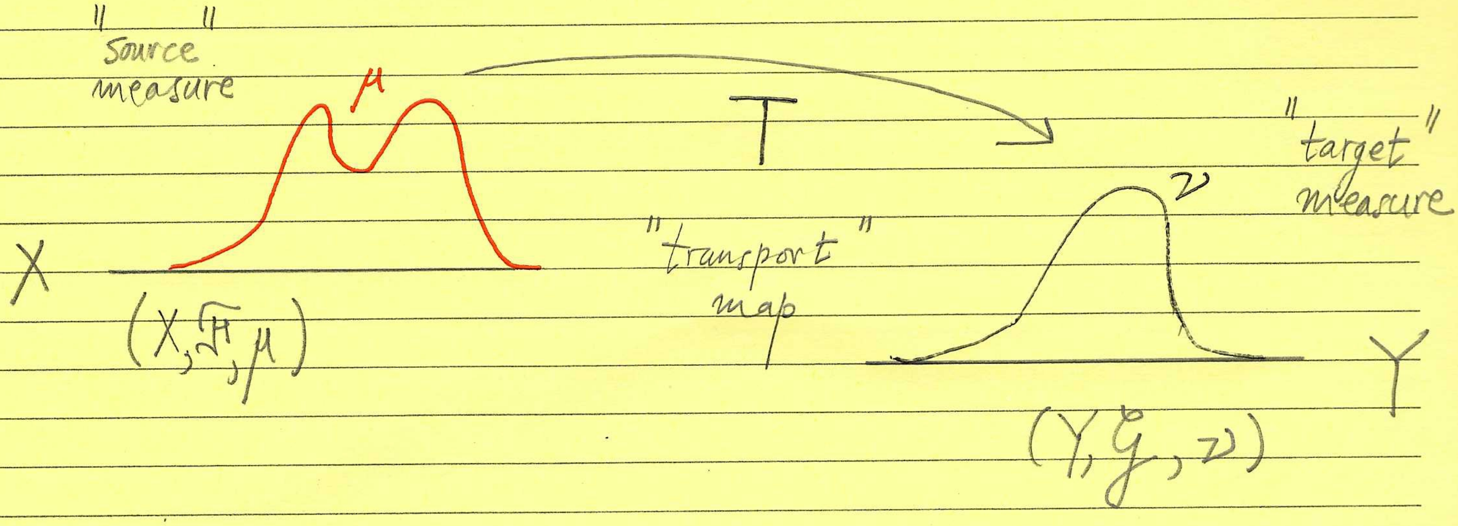
1

NOTIONS OF OPTIMAL TRANSPORT



Suppose there is a pile of soil in some location X and that, for one reason or another, it needs to be transported to a different location Y. (Construction managers and civil engineers face such questions a lot.) You know how the pile is "distributed" over X, and how you want it to be distributed over Y, once you have managed to transport it.

But transportation is costly: you need hands to load the soil on trucks in X, drivers and gas to drive to Y, then hands to upload in Y the way you want. So suppose it costs $c(x,y) \geq 0$ to transport a unit of soil from $x \in X$ to $y \in Y$. Denote the "source" distribution of soil by a probability measure μ on the source space $(X, \mathcal{F})$ (the total amount of soil is finite, so you assume $\mu(\Omega) = 1$), and the "target" distribution by a probability measure ν on $(Y, \mathcal{G})$ ($\nu(Y)=1$, because you assume all soil is transported, without loss or gain of mass). And you devise a "transport map" $T: X \to Y$ which is measurable, i.e., $T^{-1}(B) \in \mathcal{F}$, $\forall\ B \in \mathcal{G}$, and has the property

(1) $$\nu(B) = \mu(T^{-1}(B)), \quad \forall\ B \in \mathcal{G},$$

meaning that

(2) $$\nu = \mu \circ T^{-1} = T \# \mu$$ is the "push-forward" (or "induced") measure on $\mathcal{G}$, through the action of T on μ.

MONGE (1781) TRANSPORTATION PROBLEM:

Given a cost function $c: X \times Y \to [0, \infty)$, and two probability measures μ (on $(X, \mathcal{F})$, "source") and ν (on $(Y, \mathcal{G})$, "target"), find a transport map $\underbrace{T: X \to Y}_{\text{(as in (2))}}$ that minimizes the total cost

(3) $$G(T) \triangleq \int_X c(x, T(x))\, \mu(dx)$$ over all transport maps.

For reasons we shall try to explain, this is a difficult problem; two centuries had to pass before the first major, significant result about it was proved by Y. BRENIER (1987).

Let us present, then, a different, "relaxed" formulation of this problem which DOES admit general solution.

Linear Programming: A motivational example

Suppose X consists of N bakeries at locations $x_1, \dots, x_N$, and Y of M coffee shops at locations $y_1, \dots, y_M$. The i^{th} bakery produces an amount $\alpha_i \geq 0$ of "bread", and the j^{th} coffee shop needs an amount $\beta_j \geq 0$; assume that "supply equals demand", and normalize: $\sum_{i=1}^{N} \alpha_i = 1 = \sum_{j=1}^{M} \beta_j$.

Unlike in the MONGE formulation (where all the supply of a given bakery needs to be sent to only one coffee shop), we allow now each bakery to supply several coffee shops; and each coffee shop to buy from multiple bakeries.

Assuming a cost $c(x_i, y_j)$ for moving one unit from x_i to y_j, let us look for a matrix $\Gamma = (\gamma_{ij})_{\substack{1 \leq i \leq N \\ 1 \leq j \leq M}}$ such that

- the amount of bread from x_i to y_j is $\gamma_{ij} \geq 0$,
- $\alpha_i = \sum_{j=1}^{M} \gamma_{ij}$, $\forall\, i = 1, \dots, N$ (all supply is consumed)
- $\beta_j = \sum_{i=1}^{N} \gamma_{ij}$, $\forall\, j = 1, \dots, M$ (all demand is met)
- Γ minimizes the total transportation cost

$$\sum_{i=1}^{N} \sum_{j=1}^{M} \gamma_{ij}\, c(x_i, y_j).$$

This leads to the KANTOROVICH (1942, 1948) relaxation of the MONGE problem.

<u>COUPLINGS</u>: With the probability spaces $(X, \mathcal{F}, \mu)$ and $(Y, \mathcal{G}, \nu)$ as before, we introduce now the product measurable space $(X \times Y, \mathcal{F} \otimes \mathcal{G})$ and on it the collection of all probability measures π with the properties

(4) $$\pi(A \times Y) = \mu(A), \ \forall A \in \mathcal{F}; \quad \pi(X \times B) = \nu(B), \ \forall B \in \mathcal{G}.$$

We call such π a <u>coupling</u> of μ, ν.

An equivalent, and often more convenient, formulation of property (4) is

(5) $$\iint_{X \times Y} \left[\varphi(x) + \psi(y)\right] \pi(dx, dy) = \int_X \varphi(x)\,\mu(dx) + \int_Y \psi(y)\,\nu(dy)$$

for every $\varphi \in \mathbb{L}^1(\mu)$, $\psi \in \mathbb{L}^1(\nu)$ (or equivalently $\varphi \in \mathbb{L}^\infty(\mu)$, $\psi \in \mathbb{L}^\infty(\nu)$ and very often $\varphi \in C_b(X)$, $\psi \in C_b(Y)$, depending on context).

The collection of coupling probability measures

(6) $$\Pi(\mu, \nu) \triangleq \{\pi \in \mathcal{P}(X \times Y) : (4) \text{ holds}\}$$

is non-empty: it contains (at least) the product measure $\mu \otimes \nu$. Every supplier should ship uniformly over all consumers; every consumer should select uniformly over all suppliers.

KANTOROVICH (1942, 1948) TRANSPORT PROBLEM

Minimize over $\pi \in \Pi(\mu, \nu)$ the linear transportation cost

(7) $$I(\pi) \triangleq \iint_{X \times Y} c(x,y)\, \pi(dx,dy).$$

<u>Ambition</u>: Use the optimal transport cost

(8) $$\inf_{\pi \in \Pi(\mu,\nu)} I(\pi) =: \mathcal{T}_c(\mu,\nu) \leq \underbrace{\underset{\substack{\uparrow \\ \text{MONGE}}}{\mathcal{M}_c(\mu,\nu)} := \inf_{\substack{T: X \to Y \\ T\# \mu = \nu}} C(T)}_{\text{Can easily be } \equiv \infty}$$

as a kind of "distance" between the probability measures μ and ν.

- It is easy to construct examples of probability measures μ, ν on $\mathcal{B}(\mathbb{R}^n)$ which charge "small" sets (e.g., line segments, if $n = 2$), such that the optimal transfer plans in the KANTOROVICH problem have to split mass; and thus the inequality in (8) is strict, even with both quantities finite.

<u>Remark</u>: We have the probabilistic interpretation

(9) $$\mathcal{T}_c(\mu,\nu) \triangleq \inf_{\substack{(U,V) \\ U \sim \mu,\, V \sim \nu}} \mathbb{E}^{\mathbb{P}}\big(c(U,V)\big).$$

The special case $X = Y = \mathbb{R}^n$, $c_r(x,y) = |x-y|^r$ for $1 \le r < \infty$ has special significance, in that

(10) $$W_r(\mu,\nu) \triangleq \left(\mathcal{T}_{c_r}(\mu,\nu) \right)^{1/r}, \quad (\mu,\nu) \in \left(\mathcal{P}_r(\mathbb{R}^n) \right)^2$$

defines a <u>distance</u> on $\mathcal{P}_r(\mathbb{R}^n)$, the so-called WASSERSTEIN distance of order r. Very important special case is $r=2$, which corresponds to the quadratic WASSERSTEIN distance.

<u>Remark:</u> We can replace here $\mathbb{R}^n$ by a locally compact, separable metric space

<u>BIG BENEFIT</u> of the KANTOROVICH Formulation:

The "relaxed" problem of (7)-(9) admits a solution under very general conditions.

NOBEL Prize for KANTOROVICH.

Methodologies of <u>convex duality</u>.

THEOREM: KANTOROVICH DUALITY

- $\mathcal{X}, \mathcal{Y}$ Polish spaces (think of $\mathbb{R}^n$);
- $c: \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ a l.s.c. cost function is given;
- $\mu \in \mathcal{P}(\mathcal{X})$ and $\nu \in \mathcal{P}(\mathcal{Y})$ are given. Consider probability measures $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ and functions $\varphi \in \mathbb{L}^1(\mu)$, $\psi \in \mathbb{L}^1(\nu)$ NOT required to be nonnegative;
- $\Pi(\mu, \nu)$ is the collection of probability measures π in $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ with marginals μ, ν as in (6);

$$\Phi_c \triangleq \left\{ (\varphi, \psi) \in \mathbb{L}^1(\mu) \times \mathbb{L}^1(\nu) : \varphi(x) + \psi(y) \leq c(x,y) \text{ for } \mu\text{-a.e. } x \in \mathcal{X},\ \nu\text{-a.e. } y \in \mathcal{Y} \right\}$$

PRIMAL COST: $$I(\pi) \triangleq \iint_{\mathcal{X} \times \mathcal{Y}} c(x,y)\, \pi(dx, dy) \quad \text{as in (7)}$$

DUAL PAYOFF: $$J(\varphi, \psi) \triangleq \int_{\mathcal{X}} \varphi(x)\, \mu(dx) + \int_{\mathcal{Y}} \psi(y)\, \nu(dy)$$

$$= \iint_{\mathcal{X} \times \mathcal{Y}} [\varphi(x) + \psi(y)]\, \pi(dx, dy)$$

from (5).

- PRIMAL PROBLEM: $\underline{V} \triangleq \inf_{\pi \in \Pi(\mu,\nu)} I(\pi)$
- DUAL PROBLEM: $\overline{V} \triangleq \sup_{(\varphi,\psi) \in \Phi_c} J(\varphi, \psi)$

Then the primal and dual value functions are the same, and the primal problem admits a solution: there exists $\pi^* \in \Pi(\mu, \nu)$ with

$$I(\pi^*) = \underline{V} = \overline{V}.$$