Optimal Transport

Axel Böhm

September 7, 2021

The Monge Problem (1781)

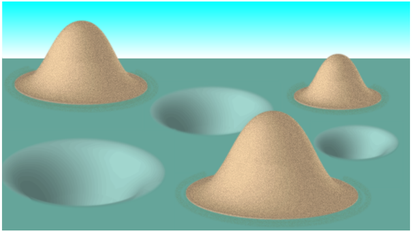


Figure: How to best move piles of sand to fill up holes of the same total volume?

- X, Y are metric spaces
- $\mathcal{B}$ is a Borel σ-Algebra (open sets, countable union, ...)
- μ is a probability measure

Given a cost $c : X \times Y \to \mathbb{R}_+$ find a **transport map** $T : X \to Y$ minimizing

$$M(\mu, \nu) = \inf_T \int_X c(x, Tx)\, \mathrm{d}\mu(x)$$

s.t. the *mass* remains the same:

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

Drawbacks of Monge formulation

Example

$X = (x_1, \ldots, x_m)$ and $Y = (y_1, \ldots, y_n)$ and $\mu = \frac{1}{m}(\delta_{x_1} + \cdots + \delta_{x_m})$ and $\nu = \frac{1}{n}(\delta_{y_1} + \cdots + \delta_{y_m})$

If $m = n$, then $c(\cdot, \cdot) = C \in \mathbb{R}^{m \times n}$ is just a (square) matrix and T is a permutation σ

$$M(\mu, \nu) = \min_T \frac{1}{n} \sum_i c(x_i, T(x_i))$$
$$= \min_\sigma \frac{1}{n} \sum_i c(x_i, y_{\sigma(i)})$$
$$= \min_\sigma \frac{1}{n} \sum_i C_{i,\sigma(i)}$$

If $m \neq n$, for example $\mu = \delta_{x_1}$ and $\nu = \frac{1}{2}(\delta_{y_1} + \delta_{y_2}) \Rightarrow$ No T exists.

Nonuniqueness of solutions

book shifting for $C(x,y) = |x-y|$

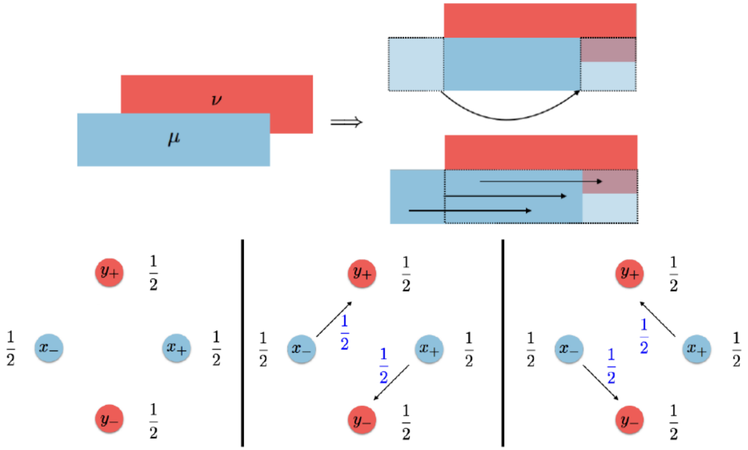


Image courtesy: Mathias Liero.

Figure

Kantorovich's relaxation (1940s)

Before: *Transport map*, $\forall x$ move all amount to some y.
Now: *Transport plan*, $\forall (x, y)$ how much to move from x to y.

Find $\gamma \in \P(X \times Y)$

$$K(\mu, \mu) = \inf_{\gamma} \int_{X \times Y} c(x, y) \, \mathrm{d}\gamma(x, y)$$

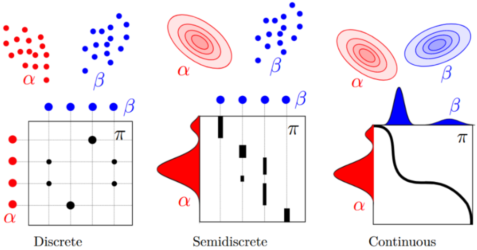
$$\text{s.t.} \gamma(A \times Y) = \mu(A)$$

$$\gamma(X \times B) = \nu(B)$$

for all $A \in \mathcal{B}(X)$ and $B \in \mathcal{B}(Y)$. Note: constraints say that μ and ν are the marginals of γ, i.e. $\gamma \in \Pi(\mu, \nu)$.

Kantorovich's relaxation

Allows for many more settings.



Figure

Example (as before)

$X = (x_1, \ldots, x_m)$ and $Y = (y_1, \ldots, y_n)$ and $\mu = \frac{1}{m}(\delta_{x_1} + \cdots + \delta_{x_m})$ and $\nu = \frac{1}{n}(\delta_{y_1} + \cdots + \delta_{y_m})$

The problem reduces to

$$\min_{\gamma} \sum_{i,j} C_{i,j}\gamma_{i,j},$$

where again $C_{i,j}$ is the cost of moving mass from x_i to y_j Then the space of probability measures on $X \times Y$ is just the set of matrices $[\gamma_{i,j}]_{i=1,\ldots,m,j=1,\ldots,n}$ such that

$$\sum_{i=1}^{m} \gamma_{i,j} = \frac{1}{n} \quad \text{and} \quad \sum_{j=1}^{n} \gamma_{i,j} = \frac{1}{m}$$

called **bistochastic matrices**.

Doubly stochastic matrices and the Birkhoff polytope

If $m = n$, then

$$\sum_{i=1}^{m} \gamma_{i,j} = \frac{1}{n} = \sum_{j=1}^{n} \gamma_{i,j}$$

which called **bistochastic matrices** (scaled by $1\overline{_{n}).}$

The set of all such matrices is called the **Birkhoff polytope**.

Doubly stochastic matrices and the Birkhoff polytope

If $m = n$, then

$$\sum_{i=1}^{m} \gamma_{i,j} = \frac{1}{n} = \sum_{j=1}^{n} \gamma_{i,j}$$

which called **bistochastic matrices** (scaled by $1\overline{n)}$.
The set of all such matrices is called the **Birkhoff polytope**.

Vertices of the Birkhoff polytope

Are given by the permutation matrices.

Doubly stochastic matrices and the Birkhoff polytope

If $m = n$, then

$$\sum_{i=1}^{m} \gamma_{i,j} = \frac{1}{n} = \sum_{j=1}^{n} \gamma_{i,j}$$

which called **bistochastic matrices** (scaled by $1\overline{n)}.$
The set of all such matrices is called the **Birkhoff polytope**.

Vertices of the Birkhoff polytope

Are given by the permutation matrices.

By linearity

$$\min_{\gamma \in \Pi} \sum_{i,j} C_{i,j}\gamma_{i,j}$$

$$= \frac{1}{n} \min_{\gamma \in Perm} \sum_{i,j} C_{i,j}\gamma_{i,j} \quad = \frac{1}{n} \min_{\sigma} \sum_{i,j} C_{i,\sigma(i)}$$

Example (with non-uniform distribution)

Now $\mu = p = (p_1, \dots, p_m)$ and $\nu = q = (q_1, \dots, q_n)$

$$K(\mu, \nu) = \min_{\gamma} \sum C_{i,j} \gamma_{i,j}$$

$$s.t. \begin{cases} \sum_i \gamma_{i,j} = q_j \\ \sum_j \gamma_{i,j} = p_i \\ \gamma_{i,j} \geq 0 \end{cases}$$