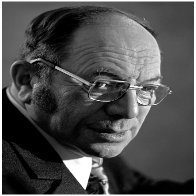
Lecture 32: Optimal Transport

Lec 32 Optimal transport problems

- Transport plans
- Optimal transport problems

Optimal transport

Monge

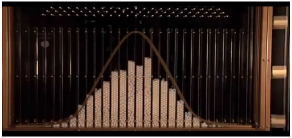


Kantorovich
(1975 Nobel prize in Economics)

Brenier, Caffarelli, McCann, Otto, **Villlani (2008 Fields Medal)**, Ambrosio, Trudinger, Wang, ... , **Figalli (2018 Fields Medal)**, and many many others are working in this area.

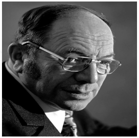
Lesson: It is important to work on a **good problem**.

(Probability) distributions



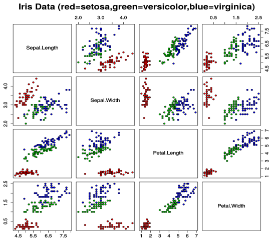
- Randomness
- Distributions





  - distribution of mass, temperature, etc.
  - Images

- Data sets



Applications of Optimal Transport

- Image processing

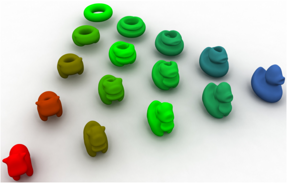
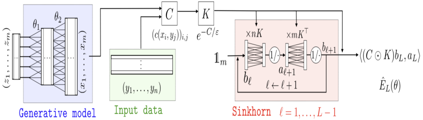


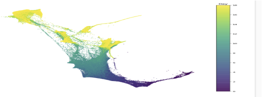
Image provided by Marco Cuturi

- Machine learning



from [Genevay-Peyr'e-Cuturi 2017]

- Stem cell research



(Image from G. Schiebinger)

- Many others....

Probability distributions in discrete setting:

- $\mu : \{1, 2, ..., m\} \to \mathbb{R}$ gives a **probability distribution** if
  - $\sum_{i=1}^{m} \mu_i = 1$
  - $\mu_i \geq 0$ for all i.

  The same as saying $\vec{\mu} = (\mu_1, \cdots, \mu_m)$ is a stochastic vector.
- $\nu : \{1, 2, ..., n\} \to \mathbb{R}$ gives a **probability distribution** if
  - $\sum_{i=j}^{n} \nu_j = 1$
  - $\nu_j \geq 0$ for all j.

  The same as saying $\vec{\nu} = (\nu_1, \cdots, \nu_m)$ is a stochastic vector.

Example

- μ=distribution of buyers over preferences $i = 1, ..., m$.
- ν= distribution of sellers over locations $j = 1, ..., n$.

Example

- μ=distributions of students over locations $i = 1, ..., m$.
- ν= distribution of capacities for schools $j = 1, ..., n$.

Example

- μ= pixels in one photo
- ν= pixels in another photo

Example

- μ= a bacteria population today
- ν= the bacteria population tomorrow.

Transport plan π between probability distributions μ, ν as a probability distribution on the product space $X \times X$ with marginals μ and ν.

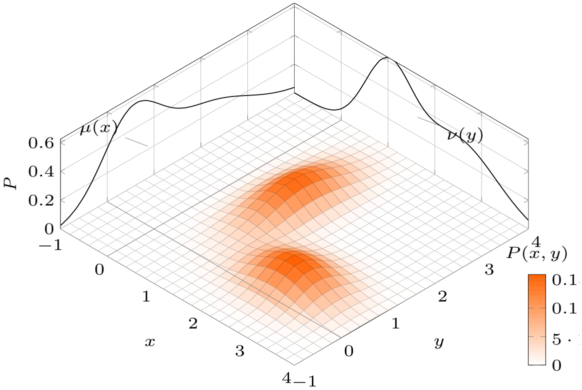


Image provided by Marco Cuturi

$$\pi \in \Pi(\mu, \nu) \iff \text{the marginals } \pi^1 = \mu, \quad \pi^2 = \nu$$

Example

- μ= distribution of buyers,
- ν= distribution of sellers,
- π = how to match the buyers and sellers

Example

- μ= pixels in one photo
- ν= pixels in another photo
- π=how to match these pixels in different photos. (questions in image processing)

Example

- μ= a bacteria population today
- ν= a bacteria population tomorrow
- π = how the bacterias change from today to tomorrow.

Transport plans in discrete setting

(See the board.)

- Let $\mu = (\mu_1, ..., \mu_m)$, $\nu = (\nu_1, .., \nu_n)$ be probability distributions.
- A **transportation plan** π between μ and ν satisfies

$$\sum_{j=1}^{n} \pi_{ij} = \mu_i \quad \text{(the first marginal of } \pi \text{ is } \mu\text{)},$$
$$\sum_{i=1}^{m} \pi_{ij} = \nu_j \quad \text{(the second marginal of } \pi \text{ is } \nu\text{)},$$
$$\pi_{ij} \geq 0 \quad \text{(for all } i, j\text{'s.)}.$$

- Note that π itself is a probability distribution:
  - $\sum_{i=1}^{m} \sum_{j=1}^{n} \pi_{ij} = 1.$

Notation $\Pi(\mu, \nu)$

The set of transport plans between μ and ν:

$$\Pi(\mu, \nu) = \left\{ \pi \;\middle|\; \sum_{j=1}^{n} \pi_{ij} = \mu_i, \quad \sum_{i=1}^{m} \pi_{ij} = \nu_j, \quad \pi_{ij} \geq 0 \quad \text{(for all } i, j\text{'s.)} \right\}$$

If we view $\pi = [\pi_{ij}]$ as a matrix, then $\pi \in \Pi(\mu, \nu)$ is the same as

- The sum of entries of the i-th row of π is μ_i
- The sum of entries of the j-th column of π is ν_j
- All the entries of π is ≥ 0.

Example

(Board) For $\mu = (1/2, 1/2)$ and $\nu = (1/3, 1/3, 1/3)$,

$$\pi = \begin{bmatrix} 1/6 & 0 & 1/3 \\ 1/6 & 1/3 & 0 \end{bmatrix} \in \Pi(\mu, \nu)$$

Example

(Board) For $\mu = (1/2, 1/2)$ and $\nu = (1/3, 1/3, 1/3)$,

$$\pi = \begin{bmatrix} 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 \end{bmatrix} \in \Pi(\mu, \nu)$$

Theorem

For any probability distributions $\mu = (\mu_1, \ldots, \mu_m)$, $\nu = (\nu_1, .., \nu_n)$, *there is a transport plan* $\pi \in \Pi(\mu, \nu)$.

Proof.

Let $\pi_{ij} = \mu_i \nu_j$. Then, obviously, $\pi_{ij} \geq 0$. Moreover,

$$\sum_{j=1}^{n} \pi_{ij} = \sum_{j=1}^{n} \mu_i \nu_j = \mu_i \sum_{j=1}^{n} \nu_j = \mu_i$$

$$\sum_{i=1}^{m} \pi_{ij} = \sum_{i=1}^{m} \mu_i \nu_j = \nu_j \sum_{i=1}^{m} \mu_i = \nu_j$$

$$\pi_{ij} \geq 0 \quad \text{(for all } i, j\text{'s.)}.$$

Thus, $\pi \in \Pi(\mu, \nu)$. $\square$

Example of μ, ν with only one transport plan

$$\mu = (1), \qquad \nu = (\nu_1, \nu_2, ..., \nu_n).$$

The unique $\pi \in \Pi(\mu, \nu)$ is

$$\pi = [\nu_1, \nu_2, ..., \nu_n].$$

Cost functions

Cost function: $c : \{1, \cdots, m\} \times \{1, \cdots, n\} \to \mathbb{R}$.

- c_{ij} = the cost of moving a unit mass at location i to location j.

Example

For points $\vec{x}_1, ..., \vec{x}_m \in \mathbb{R}^d$, $\vec{y}_1, ..., \vec{y}_n \in \mathbb{R}^d$

- distance cost $c_{ij} = |\vec{x}_i - \vec{y}_j|$
- distance squared cost $c_{ij} = |\vec{x}_i - \vec{y}_j|^2$

Example

- μ= distribution of buyers,
- ν= distribution of sellers,
- π = how to match the buyers and sellers
- c=the “distance” between buyers and sellers
  - the ”distance” can be a distance in some abstract data space, not necessarily the physical distance.

Example

- μ= a bacteria population today
- ν= a bacteria population tomorrow
- π = how the bacterias change from today to tomorrow.
- c= the distance between cells
  - It can be some quantifier to measure the difference between two cells in some data space.

Transportation cost

Cost function: $c : \{1, \cdots, m\} \times \{1, \cdots, n\} \to \mathbb{R}$.

- c_{ij} = the cost of moving a unit mass at location i to location j.

Transportation cost for π:

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij}$$

Example

For points $\vec{x}_1, ..., \vec{x}_m \in \mathbb{R}^d$, $\vec{y}_1, ..., \vec{y}_n \in \mathbb{R}^d$

- distance cost $c_{ij} = |\vec{x}_i - \vec{y}_j|$
- distance squared cost $c_{ij} = |\vec{x}_i - \vec{y}_j|^2$

Optimal Transport problem

Given μ, ν and c,

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{ij} c_{ij}\pi_{ij}.$$

That is,

$$\begin{array}{cl} \text{Minimize} & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij}\pi_{ij} \\ \text{subject to} & \sum_{j=1}^{n} \pi_{ij} = \mu_i \\ & \sum_{i=1}^{m} \pi_{ij} = \nu_j \\ & \pi_{ij} \geq 0 \end{array}$$

- This is an LP problem!

Example (distance cost example: two points to two points in $\mathbb{R}^1$.)

- $x_1 = 0, x_2 = 1, y_1 = 2, y_2 = 3 \in \mathbb{R}$.
- $\mu = (1/2, 1/2)$,
  mass $1/2$ at x_1 and mass $1/2$ at x_2.
- $\nu = (1/2, 1/2)$,
  mass $1/2$ at y_1 and mass $1/2$ at y_2.
- $c_{ij} = |x_i - y_j|$.

$$\min_{\pi \in \Pi(\mu,\nu)} \sum_{ij} c_{ij} \pi_{ij}.$$

What are the optimal solutions?

Distance cost example: two points to two points in $\mathbb{R}^1$

- $$[c_{ij}] = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}, \quad [\pi_{ij}] = \begin{bmatrix} \pi_{11} & \pi_{12} \\ \pi_{21} & \pi_{22} \end{bmatrix}$$
- $$\sum_{ij} c_{ij}\pi_{ij} = 2\pi_{11} + 3\pi_{12} + \pi_{21} + 2\pi_{22}$$
- Thus the LP problem is

$$\begin{array}{rl} \text{Minimize} & 2\pi_{11} + 3\pi_{12} + \pi_{21} + 2\pi_{22} \\ \text{subject to} & \pi_{11} + \pi_{12} = 1/2 \\ & \pi_{21} + \pi_{22} = 1/2 \\ & \pi_{11} + \pi_{21} = 1/2 \\ & \pi_{12} + \pi_{22} = 1/2 \\ & \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22} \geq 0 \end{array}$$

Distace cost example: two points to two points in $\mathbb{R}^1$

$$\begin{array}{ll}
\text{Minimize} & 2\pi_{11} + 3\pi_{12} + \pi_{21} + 2\pi_{22} \\
\text{subject to} & \pi_{11} + \pi_{12} = 1/2 \\
& \pi_{21} + \pi_{22} = 1/2 \\
& \pi_{11} + \pi_{21} = 1/2 \\
& \pi_{12} + \pi_{22} = 1/2 \\
& \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22} \geq 0
\end{array}$$

is equivalent to

$$\begin{array}{ll}
\text{Minimize} & 2\pi_{11} + 3(1/2 - \pi_{11}) + (1/2 - \pi_{11}) + 2\pi_{22} \\
\text{subject to} & \pi_{11} \leq 1/2 \\
& 1/2 - \pi_{11} + \pi_{22} = 1/2 \\
& \pi_{11} \leq 1/2 \\
& 1/2 - \pi_{11} + \pi_{22} = 1/2 \\
& \pi_{11}, \pi_{22} \geq 0
\end{array}$$

Distance cost example: two points to two points in $\mathbb{R}^1$

$$\begin{array}{ll} \text{Minimize} & 2\pi_{11} + 3(1/2 - \pi_{11}) + (1/2 - \pi_{11}) + 2\pi_{22} \\ \text{subject to} & \pi_{11} \le 1/2 \\ & 1/2 - \pi_{11} + \pi_{22} = 1/2 \\ & \pi_{11} \le 1/2 \\ & 1/2 - \pi_{11} + \pi_{22} = 1/2 \\ & \pi_{11}, \pi_{22} \ge 0 \end{array}$$

is equivalent to

$$\begin{array}{ll} \text{Minimize} & 2\pi_{11} + 3(1/2 - \pi_{11}) + (1/2 - \pi_{11}) + 2\pi_{11} \\ \text{subject to} & \pi_{11} \le 1/2 \\ & \pi_{11} \ge 0 \end{array}$$

which is equivalent to

$$\begin{array}{ll} \text{Minimize} & 2 \\ \text{subject to} & 0 \le \pi_{11} \le 1/2 \end{array}$$

Distance cost example: two points to two points in $\mathbb{R}^1$

$$\begin{array}{ll} \text{Minimize} & 2 \\ \text{subject to} & 0 \leq \pi_{11} \leq 1/2 \end{array}$$

What does this mean?

- Answer:
  Any feasible solution (so, any transport plan between μ and ν) in this case, is optimal.

Warning: Not true in general, even with the case of two points to two points.