

OT/WGAN

The fathers (and grandfathers of OT):



Monge



Kantorovich



Koopmans



Dantzig



Brenier



Otto



McCann



Villani



Figalli

Nobel '75

Fields '10

Fields '18

Monge 1781

666. MÉMOIRES DE L'ACADEMIE ROYALE

MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.



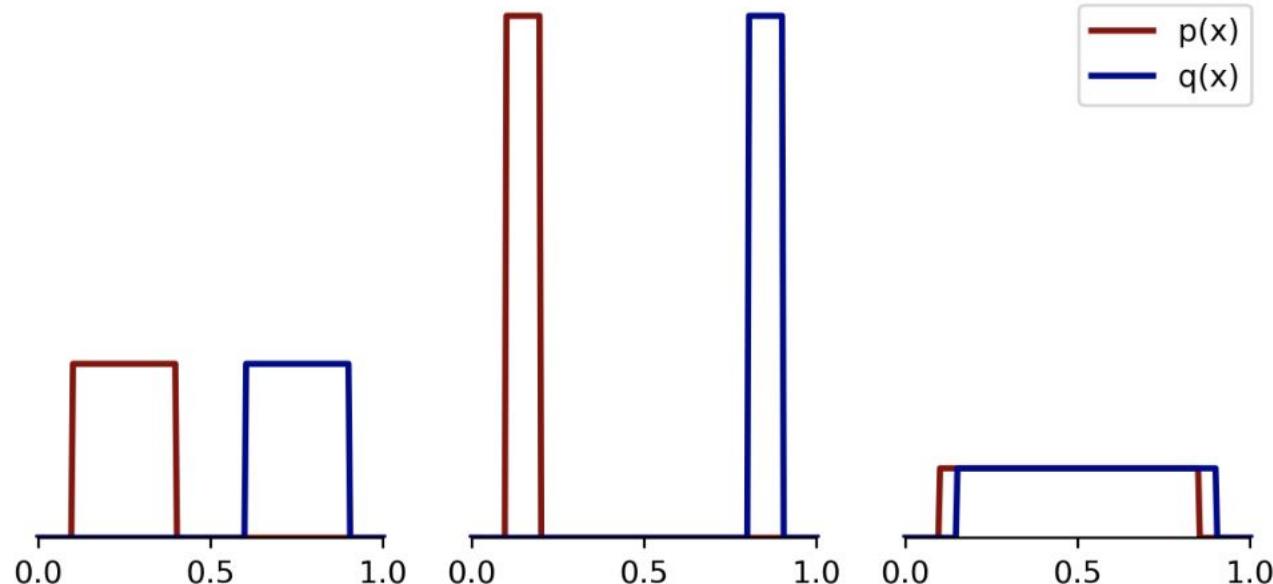
Problem [Monge, 1781]

- How to move dirt from one place (déblais) to another (remblais) while minimizing the effort ?
- Find a mapping T between the two distributions of mass (transport).
- Optimize with respect to a displacement cost $c(x, y)$ (optimal).

OT

- Why OT?
- A fundamental problem in statistics and machine learning is to come up with useful measures of “distance” between pairs of probability distributions. Two desirable properties of a distance function are symmetry and the triangle inequality.
- KL-divergence is not symmetric.
- JS
https://en.wikipedia.org/wiki/Jensen%E2%80%93Shannon_divergence

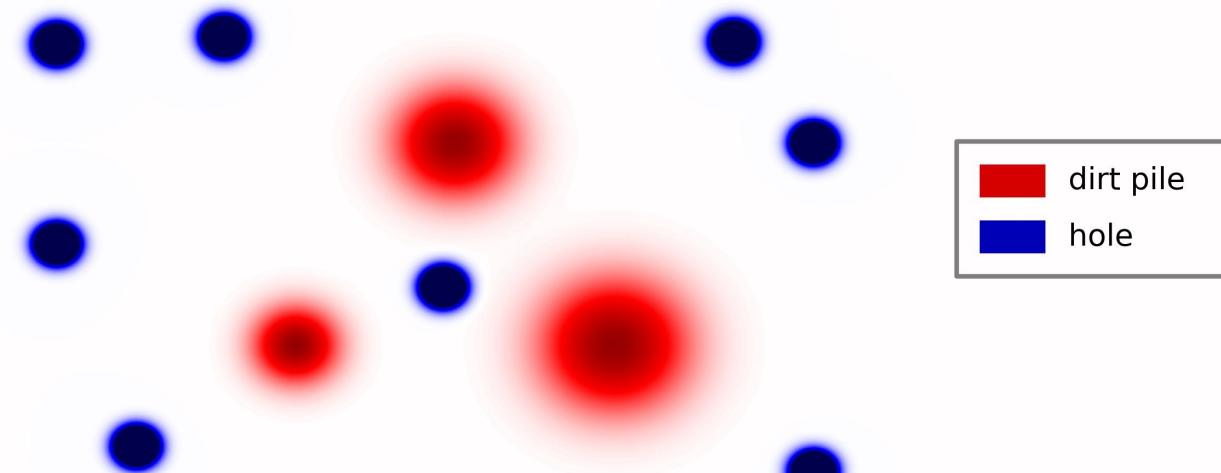
KL-div between these?



Three examples with infinite KL divergence. These density functions are infinitely far apart according to KL divergence, since in each case there exist intervals of x where $q(x)=0$ but $p(x)>0$, leading to division by zero.

An example transport problem in 2D

- Suppose we are given the task of filling several holes in the ground. The image below shows an overhead 2D view of this scenario — the three red regions correspond to dirt piles, and the eight blue regions correspond to holes.



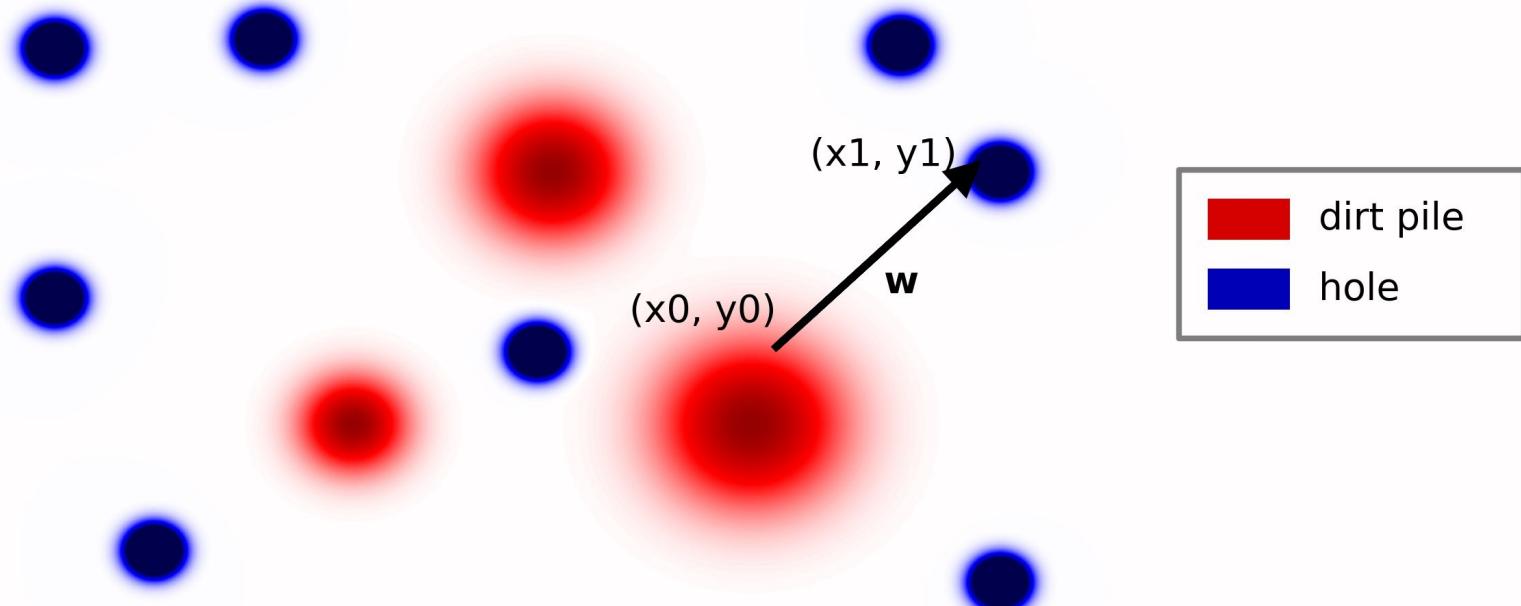
- The “most efficient” plan is the one that minimizes the total transportation cost. To quantify this, let’s say the **transportation cost** of moving 1 unit of dirt from $(x_0, y_0) \rightarrow (x_1, y_1)$ is given by the squared Euclidean distance:

$$C(x_0, y_0, x_1, y_1) = (x_0 - x_1)^2 + (y_0 - y_1)^2$$

- Now we’ll define the **transportation plan** T , which tells us how many units of dirt to move from $(x_0, y_0) \rightarrow (x_1, y_1)$

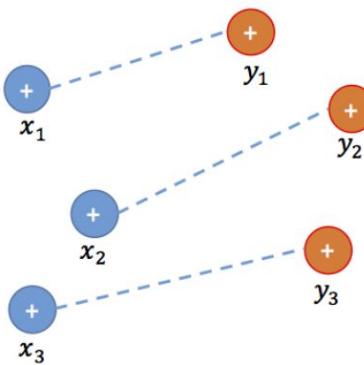
$$T(x_0, y_0, x_1, y_1) = w$$

- Then we intend to move w units of dirt from position $(x_0, y_0) \rightarrow (x_1, y_1)$



Example transport path. The arrow schematizes w units of dirt being transported from location (x_0, y_0) to (x_1, y_1) . A complete transport plan specifies transport paths like this over all pairs of locations.

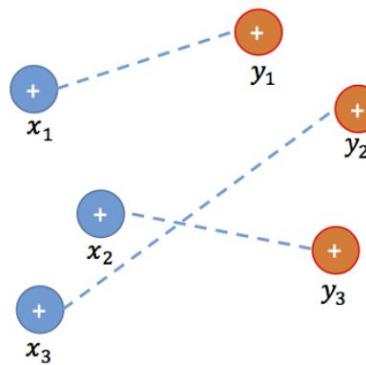
- The optimal transport problem seeks the most efficient way of transporting one distribution of mass into another.



$$x_1 \begin{bmatrix} y_1 & y_2 & y_3 \\ 1 & 0 & 0 \end{bmatrix}$$

$$x_2 \begin{bmatrix} y_1 & y_2 & y_3 \\ 0 & 1 & 0 \end{bmatrix}$$

$$x_3 \begin{bmatrix} y_1 & y_2 & y_3 \\ 0 & 0 & 1 \end{bmatrix}$$

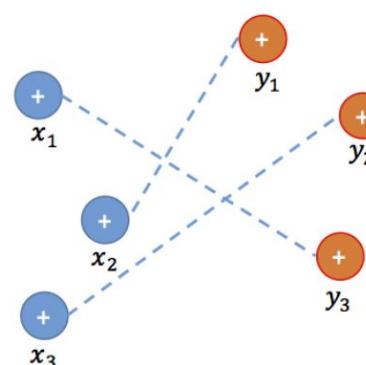


$$x_1 \begin{bmatrix} y_1 & y_2 & y_3 \\ 1 & 0 & 0 \end{bmatrix}$$

$$x_2 \begin{bmatrix} y_1 & y_2 & y_3 \\ 0 & 0 & 1 \end{bmatrix}$$

$$x_3 \begin{bmatrix} y_1 & y_2 & y_3 \\ 0 & 1 & 0 \end{bmatrix}$$

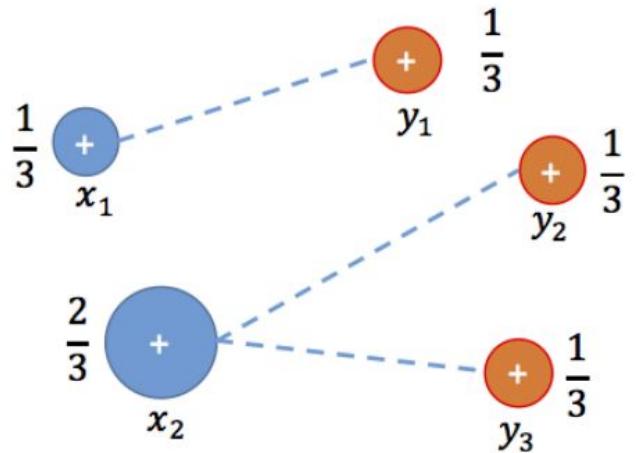
...



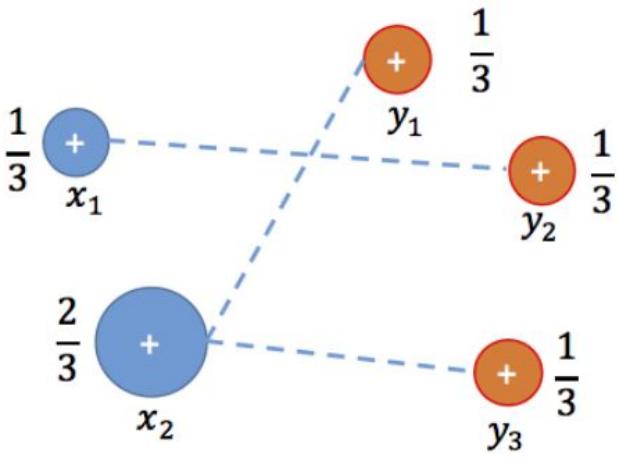
$$x_1 \begin{bmatrix} y_1 & y_2 & y_3 \\ 0 & 1 & 0 \end{bmatrix}$$

$$x_2 \begin{bmatrix} y_1 & y_2 & y_3 \\ 0 & 0 & 1 \end{bmatrix}$$

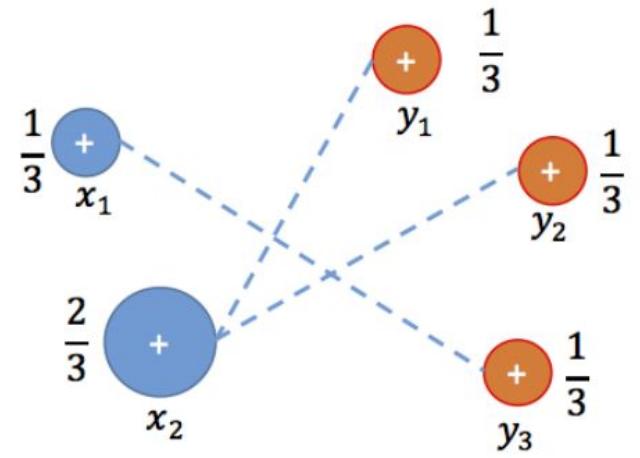
$$x_3 \begin{bmatrix} y_1 & y_2 & y_3 \\ 1 & 0 & 0 \end{bmatrix}$$



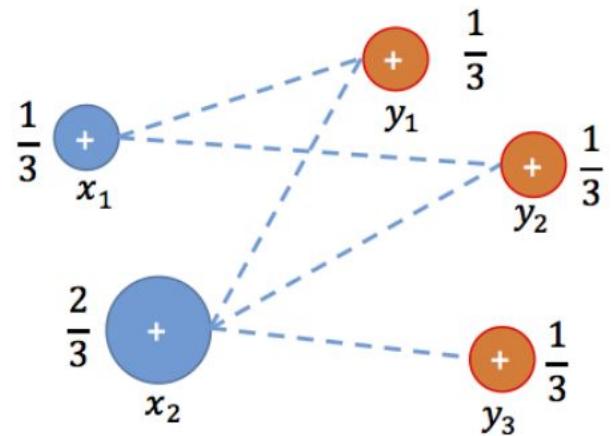
$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & \left[\begin{array}{ccc} \frac{1}{3} & 0 & 0 \end{array} \right] \\ x_2 & \left[\begin{array}{ccc} 0 & \frac{1}{3} & \frac{1}{3} \end{array} \right] \end{matrix}$$



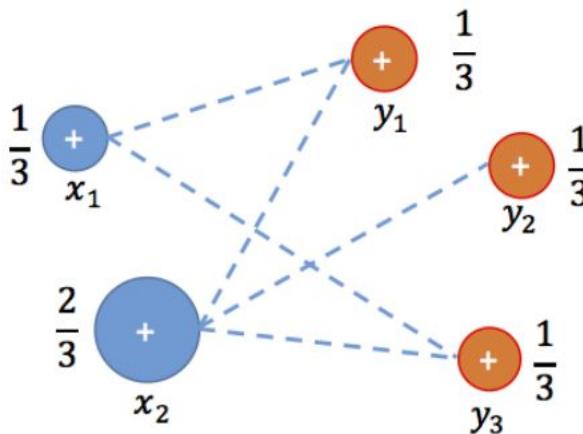
$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & \left[\begin{array}{ccc} 0 & \frac{1}{3} & 0 \end{array} \right] \\ x_2 & \left[\begin{array}{ccc} \frac{1}{3} & 0 & \frac{1}{3} \end{array} \right] \end{matrix}$$



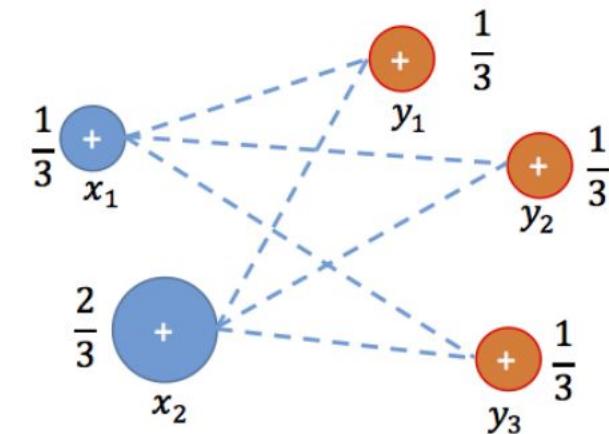
$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & \left[\begin{array}{ccc} 0 & 0 & \frac{1}{3} \end{array} \right] \\ x_2 & \left[\begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & 0 \end{array} \right] \end{matrix}$$



$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & \frac{2}{9} & \frac{1}{9} & 0 \\ x_2 & \frac{1}{9} & \frac{2}{9} & \frac{1}{3} \end{matrix}$$



$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & \frac{1}{9} & 0 & \frac{2}{9} \\ x_2 & \frac{2}{9} & \frac{1}{3} & \frac{1}{9} \end{matrix}$$



$$\begin{matrix} & y_1 & y_2 & y_3 \\ x_1 & \frac{1}{9} & \frac{1}{9} & \frac{1}{9} \\ x_2 & \frac{2}{9} & \frac{2}{9} & \frac{2}{9} \end{matrix}$$

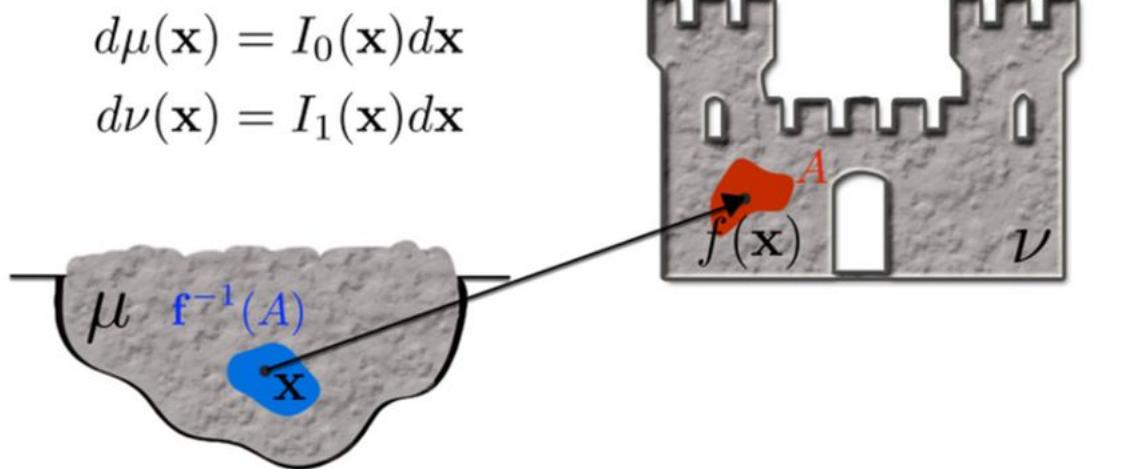
There are infinitely many transportation plans!

A little bit of history!

- The problem was originally studied by Gaspard Monge in the 18'th century.



Gaspard Monge
1746-1818



Le mémoire sur les déblais et les remblais
(The note on land excavation and infill)

Where $p(\cdot, \cdot)$ and $q(\cdot, \cdot)$ are density functions encoding the units of dirt and hole depth at each 2D location. Intuitively, the first constraint says that the amount of piled dirt at (x_0, y_0) is “used up” or transported somewhere. The second constraint says that the hole at (x_1, y_1) is “filled up” with the required amount of dirt (no more and no less).

$$\int \int T(x_0, y_0, x, y) dx dy = p(x_0, y_0) \quad \text{for all starting locations } (x_0, y_0).$$

$$\int \int T(x, y, x_1, y_1) dx dy = q(x_1, y_1) \quad \text{for all destinations } (x_1, y_1).$$

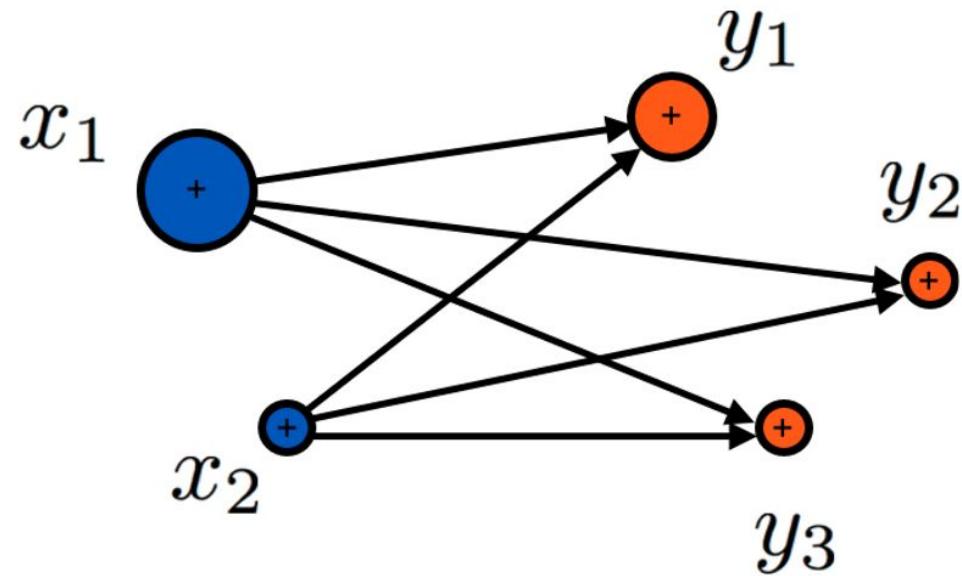
$$\sum_{y \in Y} \gamma(x_i, y) = p_0(x_i),$$

$$\sum_{x \in X} \gamma(x, y_j) = p_1(y_j).$$

- Working on optimal allocation of scarce resources during World War II, Kantorovich revisited the optimal transport problem in 1942.



Leonid Kantorovich
1912-1986



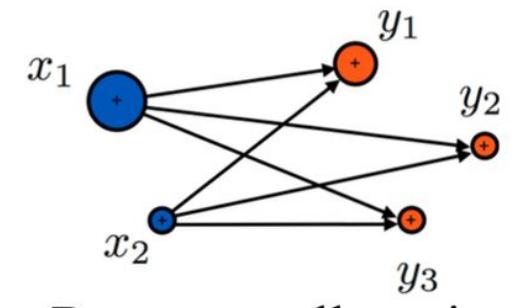
Resource allocation

- In 1975, he shared the Nobel Memorial Prize in Economic Sciences with Tjalling Koopmans "for their contributions to the theory of optimum allocation of resources."



Leonid Kantorovich
1912-1986

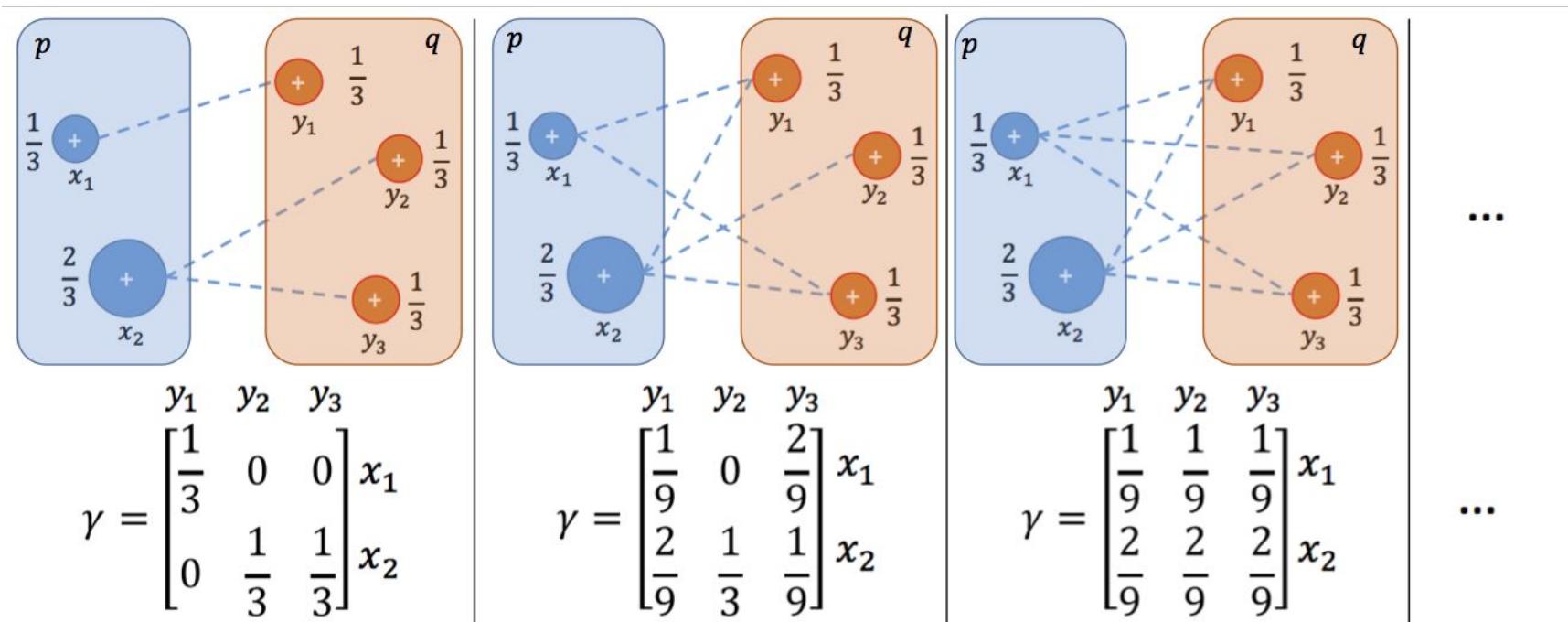
Tjalling Koopmans
1910-1985



Resource allocation

Linear programming
is born!

- First lets focus on the common trait of these transportation plans.

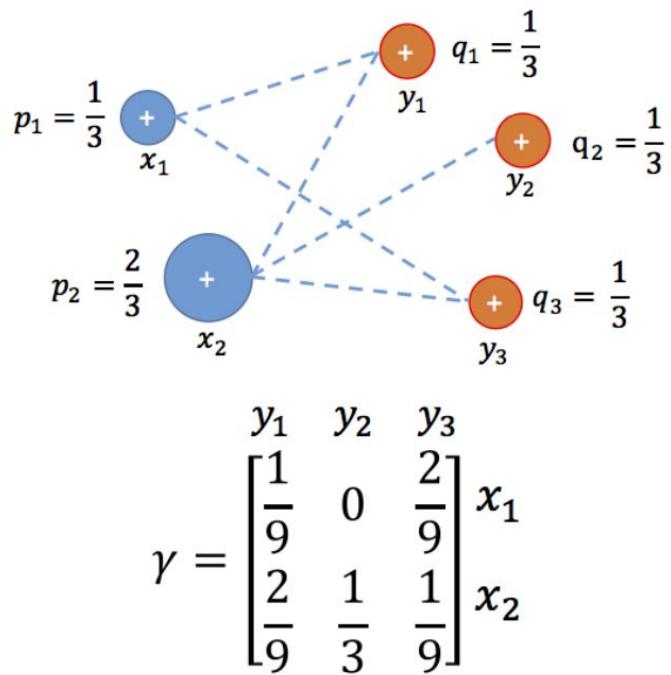


Targ
et

$$\sum_i \gamma_{i\cdot} = \left[\frac{1}{3} \quad \frac{1}{3} \quad \frac{1}{3} \right]$$

$$\sum_j \gamma_{\cdot j} = \begin{bmatrix} \frac{1}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$$

Sourc
e



$$\sum_i \gamma_{ij} = q_j, \sum_j \gamma_{ij} = p_i$$

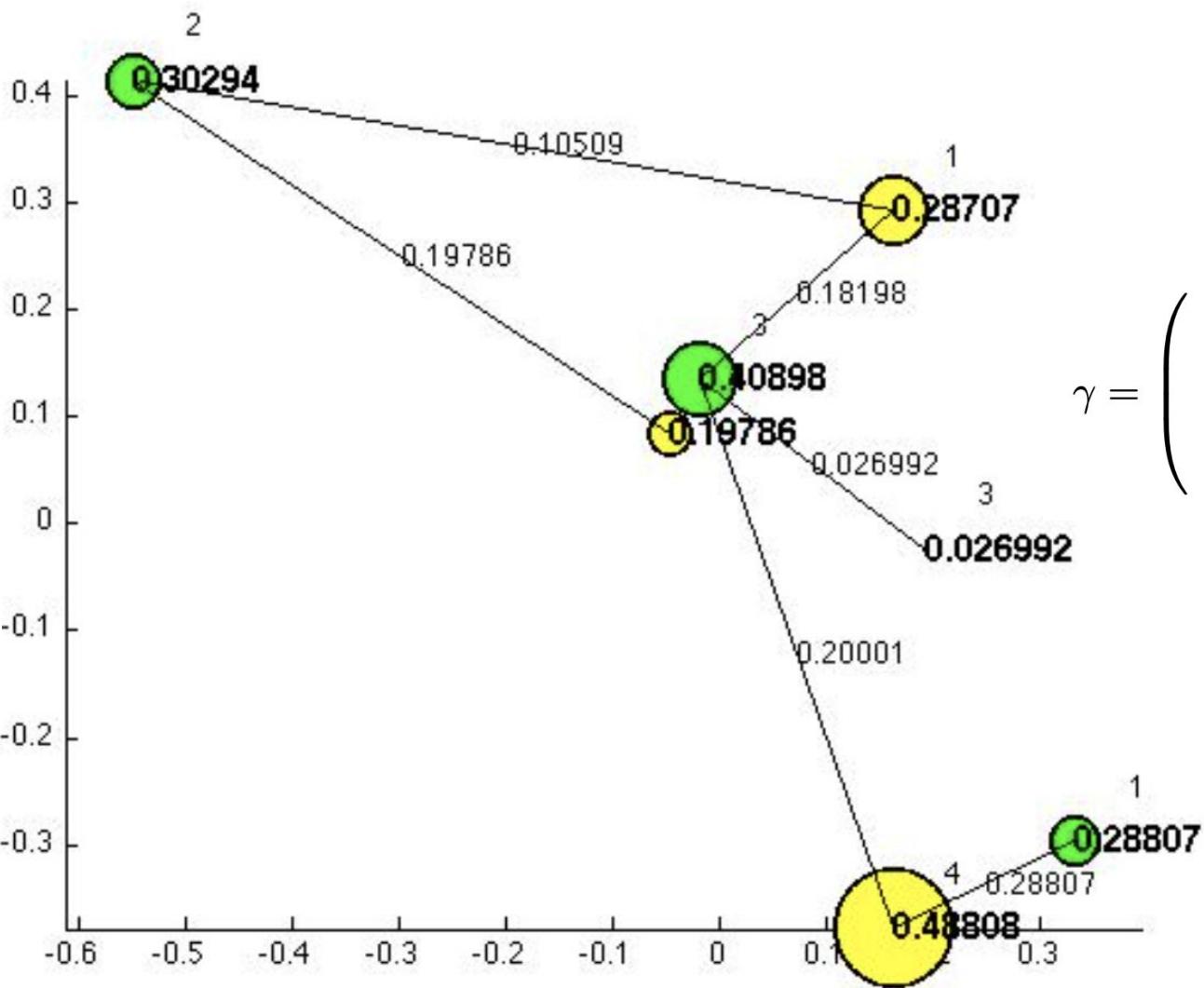
As we mentioned γ_{ij} identifies the amount of mass that is being transported from x_i to y_j

Transportation from x_i to y_j would induce a cost $c_{ij} = c(x_i, y_j)$ (e.g. cost of gas for transportation distance)

- A transportation plan is a joint probability distribution with marginal distributions equal to the original distributions, p and q .

$$p_0 = \begin{pmatrix} 0.28701 \\ 0.19786 \\ 0.026992 \\ 0.48808 \end{pmatrix}$$

$$, p_1 = \begin{pmatrix} 0.28807 \\ 0.30294 \\ 0.40898 \\ 0 \end{pmatrix}$$



$$\gamma = \begin{pmatrix} 0 & 0.10509 & 0.18198 \\ 0 & 0.19786 & 0 \\ 0 & 0 & 0.026992 \\ 0.28807 & 0 & 0.20001 \end{pmatrix}$$

Yellow is
p0
Green is
p1

OT Problem as optimization problem

$$\text{OT}(p, q; C) = \begin{array}{ll} \underset{\mathbf{T}}{\text{minimize}} & \langle \mathbf{T}, \mathbf{C} \rangle \\ \text{subject to} & \mathbf{T}\mathbf{1} = \mathbf{p}, \quad \mathbf{T}^\top \mathbf{1} = \mathbf{q}, \quad \mathbf{T} \geq 0 \end{array}$$

Solved via linear programming with complexity $O(d^3)$
d-variables

$$\begin{aligned} & \min_{\gamma} \sum_i \sum_j c_{ij} \gamma_{ij} \\ s.t. \quad & \sum_i \gamma_{ij} = q_j, \quad \sum_j \gamma_{ij} = p_i, \quad \gamma_{ij} \geq 0 \end{aligned}$$

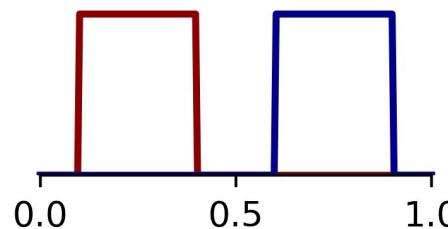
- Letting T^* denote the solution to the above optimization problem, the Wasserstein distance is defined as

$$W_p(p, q) := [\langle T^*, C \rangle]^{1/p} \quad C = \|x - y\|_2^p$$

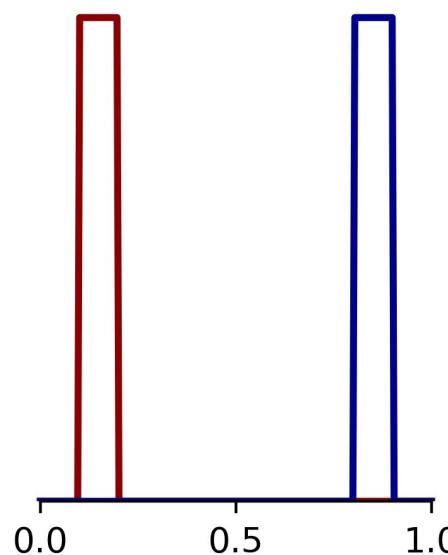
With $p=1$, it is termed as Wasserstein-1 distance and $p=2$, it is W-2 distance

W-1 is also known as EMD

$$\mathcal{W}(P, Q) = 0.503$$



$$\mathcal{W}(P, Q) = 0.704$$



$$\mathcal{W}(P, Q) = 0.05$$

