Optimal Transport: Theory, Computation and Applications

Wenchong Huang

School of Mathematical Sciences,
Zhejiang University.

Dec. 30th, 2024

Optimal Transport

Principal concern: the distance between two probability measures.
First introduced in 1781 by Monge.
Relative subjects: probability theory, geometry, graph theory, machine learning...
Applications:

- Image registration and warping;
- Reflector design;
- Retrieving information from shadowgraphy and proton radiography;
- Seismic tomography and reflection seismology.

Some well-known researchers:

- Gasoard Monge (France);
- Leonid Kantorovich (Russia);
- Yann Brenier (France);
- Xianfeng Gu (顾险峰, China);

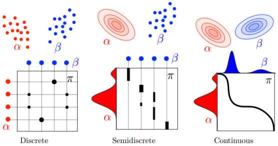


Fig. 1. Three main scenarios for Kantorovich OT

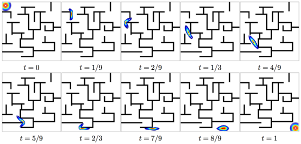


Fig. 2. Solving maze with OT

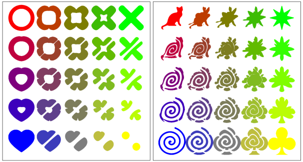


Fig. 3. 2D shape interpolation with OT

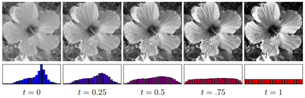


Fig. 4. Histogram equalization with OT

The sand-moving problem

A child wants to make a pile of sand in the shape of a castle.
Cost: 1 kcal per shovel and per meter horizontally.
Target: Minimize the total cost.

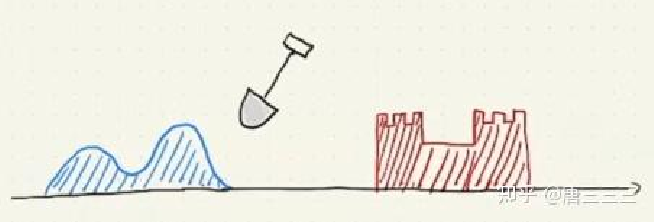


Fig. 5. The sand-moving problem.

The sand-moving problem

A child wants to make a pile of sand in the shape of a castle.
Cost: 1 kcal per shovel and per meter horizontally.
Target: Minimize the total cost.

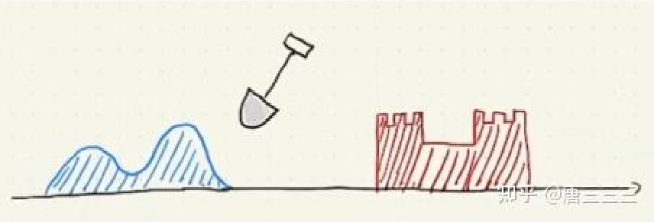


Fig. 5. The sand-moving problem.

Let's denote the source shape by $f(x)$ and the target by $g(x)$. The sand-moving problem cound be formulated as: find a **transport mapping** $T : \mathbb{R} \to \mathbb{R}$ to minimize

$$\int_{\mathbb{R}} |T(x) - x| f(x)\ dx, \tag{1}$$

which satisfies

$$\int_{T(U)} g(x)\ dx = \int_U f(x)\ dx \text{ for all open interval } U \subset \mathbb{R}. \tag{2}$$

The allocation problem

There are some steel coils to be transported from warehouses to factories. The transport cost is $1 per coil and per kilometer. How to minimize the total cost?

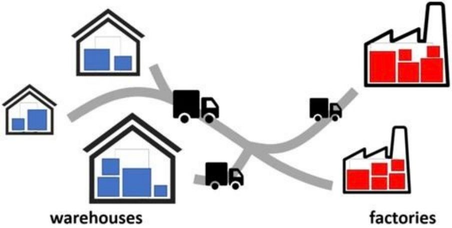


Fig. 6. The allocation problem.

The allocation problem

There are some steel coils to be transported from warehouses to factories. The transport cost is $1 per coil and per kilometer. How to minimize the total cost?

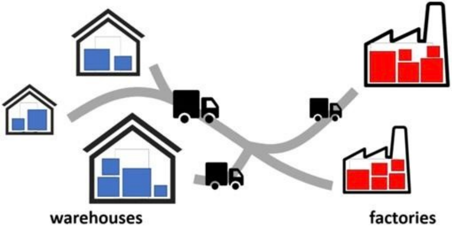


Fig. 6. The allocation problem.

Assume the i-th warehouse has a_i coils and the j-th factory needs b_i coils. And assume the distance between the i-th warehouse and the j-th factory is d_{ij}. The allocation problem could be formulated as: find a **transport matrix** v_{ij} to minimize

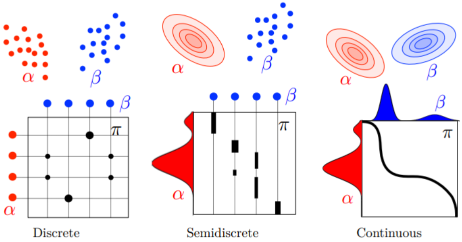
$$\sum_{i,j} d_{ij} v_{ij} \tag{3}$$

which satisfies

$$a_i = \sum_j v_{ij}, \quad \forall i, \qquad \text{and} \qquad b_j = \sum_i v_{ij}, \quad \forall j. \tag{4}$$

The Kantorovich formulation

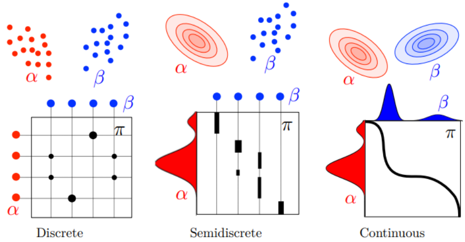
To give an general formulation of OT, we first recall the three main scenarios for OT.



[1] Leonid Kantorovich. "On the transfer of masses". In: *Doklady Akademii Nauk* 37.2 (1942).

The Kantorovich formulation

To give an general formulation of OT, we first recall the three main scenarios for OT.



Given two probability measures μ on $\mathcal{X}$ and ν on $\mathcal{Y}$, and a cost function $c(x, y)$. The Kantorovich formulation[1] of OT is

$$\mathcal{L}_c(\mu, \nu) = \min_{\pi} \int_{\mathcal{X}\times\mathcal{Y}} c(x, y)\ d\pi(x, y), \tag{5}$$

where π is a measure on $\mathcal{X} \times \mathcal{Y}$, whose marginals are μ and ν, that is,

$$\mu = \int_{\mathcal{Y}} \pi(\cdot, y)\ dy, \qquad \nu = \int_{\mathcal{X}} \pi(x, \cdot)\ dx. \tag{6}$$

[1] Leonid Kantorovich. "On the transfer of masses". In: *Doklady Akademii Nauk* 37.2 (1942).

Metric properties of OT

Here we assume $\mathcal{X} = \mathcal{Y}$ and $c(x, y) = d(x, y)^p$ $(p > 1)$, where d is a distance on $\mathcal{X}$. Denote $\mathcal{M}_+^1(\mathcal{X})$ the set of probability measures on $\mathcal{X}$.

[2] Cédric Villani. *Optimal Transport: Old and New*. Vol. 338. Springer Verlag, 2009.

Metric properties of OT

Here we assume $\mathcal{X} = \mathcal{Y}$ and $c(x, y) = d(x, y)^p$ $(p > 1)$, where d is a distance on $\mathcal{X}$. Denote $\mathcal{M}_+^1(\mathcal{X})$ the set of probability measures on $\mathcal{X}$.

Theorem (Wasserstein distance)

Under the above assumptions, $\mathcal{L}_c(\mu, \nu)^{1/p}$ is a distance on $\mathcal{M}_+^1(\mathcal{X})$.

The distance $\mathcal{W}_p(\mu, \nu) := \mathcal{L}_c(\mu, \nu)^{1/p}$ is called p-Wasserstein distance.

[2] Cédric Villani. *Optimal Transport: Old and New*. Vol. 338. Springer Verlag, 2009.

Metric properties of OT

Here we assume $\mathcal{X} = \mathcal{Y}$ and $c(x, y) = d(x, y)^p$ $(p > 1)$, where d is a distance on $\mathcal{X}$. Denote $\mathcal{M}_+^1(\mathcal{X})$ the set of probability measures on $\mathcal{X}$.

Theorem (Wasserstein distance)

Under the above assumptions, $\mathcal{L}_c(\mu, \nu)^{1/p}$ is a distance on $\mathcal{M}_+^1(\mathcal{X})$.

The distance $\mathcal{W}_p(\mu, \nu) := \mathcal{L}_c(\mu, \nu)^{1/p}$ is called p-Wasserstein distance.

Definition (Weak convergence)

Assume $\mathcal{X}$ is compact. Say $(\mu_k)_{k\geq 1} \subset \mathcal{M}_+^1(\mathcal{X})$ converges weakly to $\mu \in \mathcal{M}_+^1(\mathcal{X})$ if

$$\int_{\mathcal{X}} g \; d\mu_k \to \int_{\mathcal{X}} g \; d\mu, \quad \forall g \in \mathcal{C}(\mathcal{X}). \tag{7}$$

Theorem (Wasserstein distance and weak convergence[2])

On a compact domain $\mathcal{X}$, $(\mu_k)_{k\geq 1} \subset \mathcal{M}_+^1(\mathcal{X})$ converges weakly to $\mu \in \mathcal{M}_+^1(\mathcal{X})$ if and only if $\mathcal{W}_p(\mu_k, \nu) \to 0$.

[2] Cédric Villani. *Optimal Transport: Old and New*. Vol. 338. Springer Verlag, 2009.

Thank You