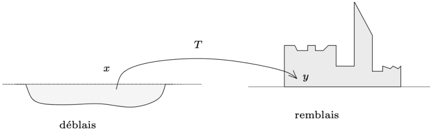
Relaxed intro to Optimal Transport

BAMS

December 9, 2025

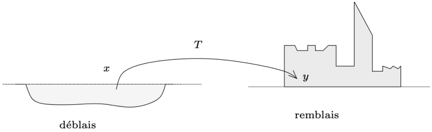
Introduction

- Problem of transporting mass in an efficient way
- Monge 1781: most efficient way of transporting soil from the ground to a given place
- Kantorovich 1975: Nobel prize for Economics "for their contributions to the theory of optimum allocation of resources"
- Many formulations: we will focus on two, discrete and continuous versions
- Many applications!!!



Introduction

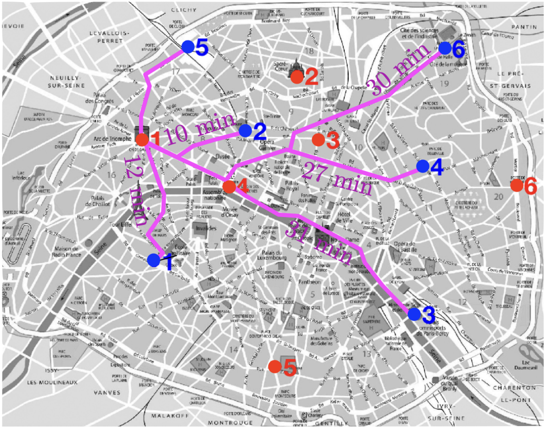
- Problem of transporting mass in an efficient way
- Monge 1781: most efficient way of transporting soil from the ground to a given place
- Kantorovich 1975: Nobel prize for Economics "for their contributions to the theory of optimum allocation of resources"
- Many formulations: we will focus on two, discrete and continuous versions
- Many applications!!!



Disclaimer: attempted to be a soft and not 100% formal introduction to OT! :)

Why are we doing this I

Fom bakeries to cafés

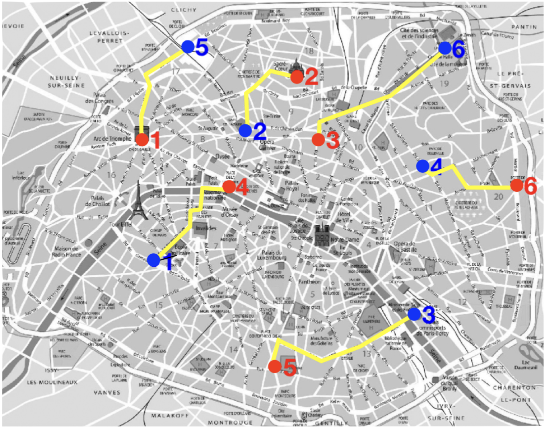


c_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

Figure: Supplying all cafes from bakery 1

Why are we doing this II

Fom bakeries to cafés



c_{ij}	y_1	y_2	y_3	y_4	y_5	y_6
x_1	12	10	31	27	10	30
x_2	22	7	25	15	11	14
x_3	19	7	19	10	15	15
x_4	10	6	21	19	14	24
x_5	15	23	14	24	31	34
x_6	35	26	16	9	34	15

Figure: Organizing supply using optimal transport!

Why are we doing this III

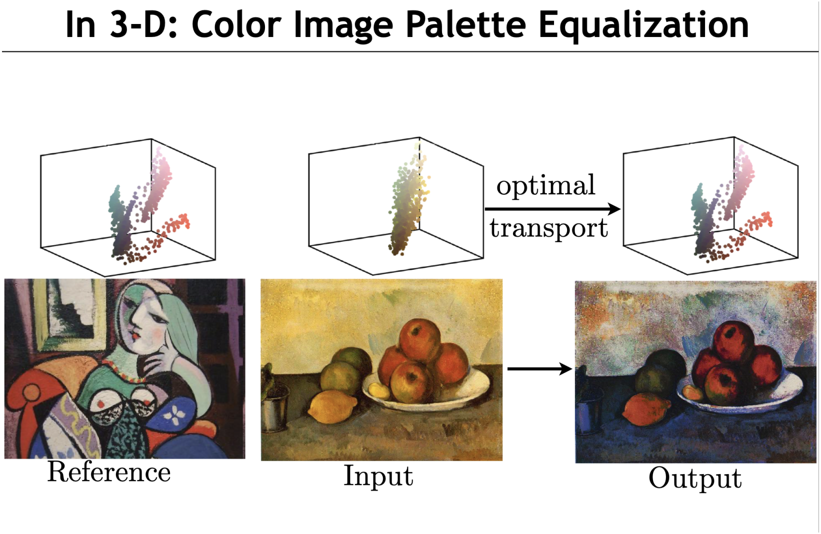


Figure: Optimally allocating color

Convexity I

Convex Set

A set $C \subseteq \mathbb{R}^n$ is convex if for any $x, y \in C$ and any $\lambda \in [0, 1]$, we have

$$\lambda x + (1 - \lambda)y \in C.$$

Intuitively, any line segment between two points in C stays inside C.

Extremal Points

A point $x \in C$ is an *extremal point* of a convex set C if

$$x = \lambda y + (1 - \lambda)z \;\Rightarrow\; y = z$$

for $y, z \in C$ and $\lambda \in (0, 1)$.
E.g. the extremal points of a polygon are its vertices.

Convexity II

Convex Function

A function $f : C \to \mathbb{R}$ defined on a convex set C is convex if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$$

for all $x, y \in C$, $\lambda \in [0,1]$. Geometrically, its graph lies below the chord joining any two points.

Polyhedra

A polyhedron is the intersection of finitely many hyperplanes, i.e. for $b \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times n}$

$$P = \{x \in \mathbb{R}^n : Ax \leq b\}.$$

If P is bounded, it is a polytope.

Convexity III

Polytope

A polytope is a bounded convex set obtained as the convex hull of finitely many points. Example: a triangle, cube, or any convex polygon/polyhedron.

Convex Optimization

A convex optimization problem is one of the form

$$\min_{x \in C} f(x)$$

where C is convex and f is convex. Nice property: any local minimum is also a global minimum.

Discrete Monge formulation

- Two clouds of points $\{x_i\}_{i=1,\cdots,n}$, $\{y_i\}_{i=1,\cdots,n}$
- Two probability measures (distribution of masses) on them
- The cost of transporting mass x_i to y_j is C_{ij}

Permutation

Given an interval $I \subset \mathbb{N}, I = \{1, \cdots, n\}$, a permutation is a bijective function $\sigma : I \to I$.
We indicate Sym(I) or just Sym(n) the set of all permutations on $I = \{1, \cdots, n\}$.

Discrete Monge formulation

- Two clouds of points $\{x_i\}_{i=1,\cdots,n}$, $\{y_i\}_{i=1,\cdots,n}$
- Two probability measures (distribution of masses) on them
- The cost of transporting mass x_i to y_j is C_{ij}

Permutation

Given an interval $I \subset \mathbb{N}, I = \{1, \cdots, n\}$, a permutation is a bijective function $\sigma : I \to I$.
We indicate $\mathsf{Sym}(I)$ or just $\mathsf{Sym}(n)$ the set of all permutations on $I = \{1, \cdots, n\}$.

Monge's formulation

$$\min_{\sigma \in \mathsf{Sym}(n)} \sum_{i,j} C_{i,\sigma(j)} \tag{D1}$$

Discrete Kantorovich formulation

Kantorovich's extension: we are allowed to "divide" masses! yay!

Bistochastic matrices

A matrix $P \subset \mathbb{R}_+^{n \times n}$ is a bistochastic matrix if $\sum_j P_{ij} = 1$, $\sum_i P_{ij} = 1$, i.e. its rows and its columns sum to 1.
We indicate B_n the set of bistochastic matrices in $\mathbb{R}_+^{n \times n}$.

Discrete Kantorovich formulation

Kantorovich's extension: we are allowed to "divide" masses! yay!

Bistochastic matrices

A matrix $P \subset \mathbb{R}_+^{n\times n}$ is a bistochastic matrix if $\sum_j P_{ij} = 1$, $\sum_i P_{ij} = 1$, i.e. its rows and its columns sum to 1.
We indicate B_n the set of bistochastic matrices in $\mathbb{R}_+^{n\times n}$.

Kantorovich's formulation

$$\min_{P \in B_n} \sum_{ij} P_{ij} C_{ij} \tag{D2}$$

Connecting Monge and Kantorovich

We can rewrite permutations as in Monge's formulation via permutations matrices

$$P_n := B_n \cap \{0,1\}^{n \times n}$$

Then Monge's problem becomes:

$$\min_{P \in P_n} \sum_{ij} P_{ij} C_{ij} \tag{D1'}$$

Connecting Monge and Kantorovich

We can rewrite permutations as in Monge's formulation via permutations matrices

$$P_n := B_n \cap \{0,1\}^{n \times n}$$

Then Monge's problem becomes:

$$\min_{P \in P_n} \sum_{ij} P_{ij} C_{ij} \tag{D1'}$$

Important remarks

1. $B_n \subset [0,1]^{n \times n}$ is a polyhedra
2. $P_n = \mathrm{Ext(B_n)}$ (Von Neumann)
3. Kantorovich formulation is a **convex relaxation** of Monge's one
4. OT problem as in D2 can be solved efficiently via linear programming

Probability measures

Probability measure

A probability measure on (a σ-algebra $\mathcal{A}$ of) a set X is a measure $\mu : \mathcal{A} \to [0, 1]$ such that $\mu(X) = 1$. We indicate as $\mathcal{P}(X)$ the set of measures on X.

Probability measures

Probability measure

A probability measure on (a σ-algebra $\mathcal{A}$ of) a set X is a measure $\mu : \mathcal{A} \to [0, 1]$ such that $\mu(X) = 1$. We indicate as $\mathcal{P}(X)$ the set of measures on X.

Push-forward measure

Given a function $T : X \to Y$, we define the push forward operator $T_{\#} : \mathcal{P}(X) \to \mathcal{P}(Y)$ by

$$T_{\#}\mu(A) = \mu(T^{-1}(A)) \quad \forall A \in \mathcal{B}(Y)$$

and call $T_{\#}\mu$ push forward measure.

The push forward measure is the measure assigning to a set the measure of its **pre-image** according to a map T.

Transport via functions

Given

- two spaces X and Y
- probability measures $\mu \in \mathcal{P}(X)$, $\nu \in \mathcal{P}(Y)$
- cost function $c : X \to Y$

Monge's formulation

$$\min_{T | T_{\#}\mu = \nu} \int_X c(x, T(x)) \, \mathrm{d}\mu(x) \tag{C1}$$

Transport via transport plans I

Transport plan

A transport plan is a probability measure $\pi \in \mathcal{P}(X \times Y)$ such that $\pi(A \times Y) = \mu(A), \pi(X \times B) = \nu(B)$.
We indicate $\Gamma(\mu, \nu)$ the set of all transport plans for μ and ν.

Transport plans are another way of carrying mass from X to Y, and in particular $\pi(A \times B)$ is the mass that was in $A \subset X$ and has been sent to $B \subset Y$.

For stats buddies: transport plans are like joint probability measures!

Kantorovich's formulation

$$\min_{\pi \in \Gamma(\mu,\nu)} \int_{X \times Y} c(x, y) \ \mathrm{d}\pi(x, y) \tag{C2}$$

Connecting Monge and Kantorovich, again

For any transport *map* T, there exists a transport *plan* π_T:

$$T \longmapsto (\mathrm{id} \times T)_{\#}\mu =: \pi_T$$

Moreover, it can be checked $\mathcal{C}(T) = \mathcal{C}(\pi_T)$ where $\mathcal{C}$ is the total cost attained by a transport map/plan.

Thus,

$$\inf_M \geq \inf_K$$

Useful resources

- Brué Ambrosio Semola Lecture on Optimal Transport
- Villani Optimal Transport Old and New
- Peyré and Cuturi Computational Optimal Transport
- Peyré slides on discrete part
- Mati's notes chapter 3

Beware: there's a sea out there, only check out topics we went over or ask for more!