

Computational Optimal Transport

Universidad Nacional de Colombia

Daniel Melo

G. Peyré y M. Cuturi, *Computational Optimal Transport*, Foundations and Trends in Machine Learning, 2019.

Index

1. Histograms and Measures	§2.1
2. Assignment and Monge Problem	§2.2
3. Kantorovich Relaxation	§2.3
4. Metric Properties of Optimal Transport	§2.4
5. Dual Problem	§2.5
6. The Kantorovich Linear Programs	§3.1

Two ways to describe mass

Histograms. Mass as a vector of weights:

$$\Sigma_n := \left\{ \mathbf{a} \in \mathbb{R}_+^n : \sum_{i=1}^n a_i = 1 \right\}.$$

Discrete measures. Weights placed at positions:

$$\alpha = \sum_{i=1}^n a_i \delta_{x_i}, \quad x_i \in \mathcal{X}.$$

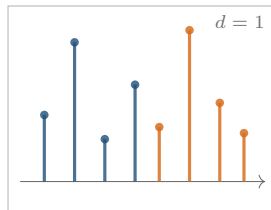
General measures. $\alpha \in \mathcal{M}_+^1(\mathcal{X})$, a Radon measure

$$\int_{\mathcal{X}} f \, d\alpha \left(= \sum_i a_i f(x_i) \text{ if } \alpha \text{ is discrete} \right).$$



α

β



α

β

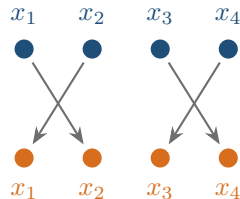
Assignment

Approach. Given a cost matrix $(C_{i,j})_{i \in \llbracket n \rrbracket, j \in \llbracket m \rrbracket}$ and assuming $n = m$, solve

$$\min_{\sigma \in \text{Perm}(n)} \frac{1}{n} \sum_{i=1}^n C_{i, \sigma(i)}.$$

Combinatorial problem: $n!$ candidates.

Is the solution unique?



Monge Problem

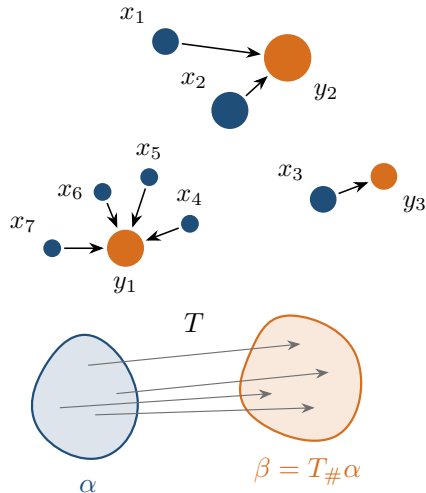
Discrete Monge. Find $T : \{x_i\} \rightarrow \{y_j\}$ that *pushes* α onto β :

$$\forall j, b_j = \sum_{i: T(x_i)=y_j} a_i \iff T_{\#}\alpha = \beta.$$

$$\min_T \left\{ \sum_i c(x_i, T(x_i)) : T_{\#}\alpha = \beta \right\}$$

General Monge (1781).

$$\min_T \left\{ \int_{\mathcal{X}} c(x, T(x)) d\alpha(x) : T_{\#}\alpha = \beta \right\}.$$

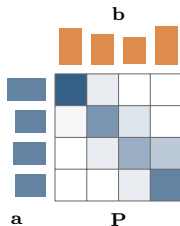
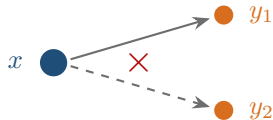


Coupling Approach

We change the Transport Map T for a transport matrix $\mathbf{P} \in \mathbb{R}_+^{n \times m}$

$$\mathbf{U}(\mathbf{a}, \mathbf{b}) := \{\mathbf{P} \in \mathbb{R}_+^{n \times m} : \mathbf{P}\mathbf{1}_m = \mathbf{a}, \mathbf{P}^\top \mathbf{1}_n = \mathbf{b}\}.$$

- $\mathbf{U}(\mathbf{a}, \mathbf{b})$ is a convex polytope.
- $\mathbf{a}\mathbf{b}^\top \in \mathbf{U}(\mathbf{a}, \mathbf{b})$.
- $\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b}) \iff \mathbf{P}^\top \in \mathbf{U}(\mathbf{b}, \mathbf{a})$.



Kantorovich Problem:

$$L_C(\mathbf{a}, \mathbf{b}) := \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle = \sum_{i,j} \mathbf{C}_{ij} \mathbf{P}_{ij}.$$

The reformulation - Birkhoff

Suppose $n = m$ and $\mathbf{a} = \mathbf{b} = \mathbf{1}_n/n$, the set of permutation matrices are included in $\mathbf{U}(\mathbf{a}, \mathbf{b})$, so

$$L_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) \leq \min_{\sigma} \frac{1}{n} \sum_i \mathbf{C}_{i, \sigma(i)}.$$

Proposition (Kantorovich for matching)

If $n = m$ and $\mathbf{a} = \mathbf{b} = \mathbf{1}_n/n$, then there exists an optimal solution for Kantorovich, which is a permutation matrix associated to an optimal permutation for the assignment problem.

$$L_{\mathbf{C}}(\mathbf{1}_n/n, \mathbf{1}_n/n) = \min_{\sigma \in \text{Perm}(n)} \langle \mathbf{C}, \mathbf{P}_{\sigma} \rangle$$

Marginals and projections

$$P_{\mathcal{X}\#}\pi(A) = \pi(P_{\mathcal{X}}^{-1}(A)) = \pi(A \times \mathcal{Y}).$$

Marginal constraints. $\pi \in \mathbf{U}(\alpha, \beta)$ iff for all measurable $A \subset \mathcal{X}$, $B \subset \mathcal{Y}$:

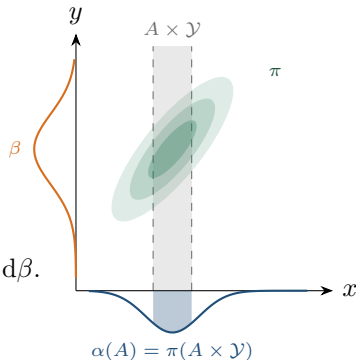
$$\pi(A \times \mathcal{Y}) = \alpha(A), \quad \pi(\mathcal{X} \times B) = \beta(B).$$

With test functions. Equivalently, for all continuous f , g :

$$\int_{\mathcal{X} \times \mathcal{Y}} f(x) \, d\pi(x, y) = \int_{\mathcal{X}} f \, d\alpha, \quad \int_{\mathcal{X} \times \mathcal{Y}} g(y) \, d\pi(x, y) = \int_{\mathcal{Y}} g \, d\beta.$$

Discrete case. $\pi = \sum_{i,j} \mathbf{P}_{i,j} \delta_{(x_i, y_j)}$:

$$\mathbf{P} \mathbf{1}_m = \mathbf{a}, \quad \mathbf{P}^\top \mathbf{1}_n = \mathbf{b}$$

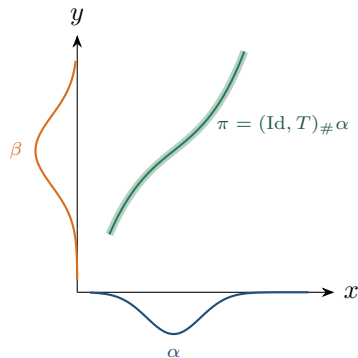


Kantorovich problem between arbitrary measures

With $P_{\mathcal{X}}(x, y) = x$, $P_{\mathcal{Y}}(x, y) = y$ as projections:

$$\mathbf{U}(\alpha, \beta) := \{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : P_{\mathcal{X}\#}\pi = \alpha, P_{\mathcal{Y}\#}\pi = \beta\},$$

$$L_c(\alpha, \beta) := \min_{\pi \in \mathbf{U}(\alpha, \beta)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y).$$



Monge–Kantorovich equivalence

Theorem (Brenier)

Let $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$, $c(x, y) = \|x - y\|^2$, and let α have a density. Then there exists a unique optimal coupling, supported on the graph of a map:

$$\pi^\star = (\text{Id}, T)_\# \alpha, \quad T = \nabla \varphi, \quad \varphi : \mathbb{R}^d \rightarrow \mathbb{R} \text{ convex},$$

and T is the unique gradient of a convex function with $(\nabla \varphi)_\# \alpha = \beta$.

Metric Properties of Optimal Transport

Proposition

Let $n = m$, $p \geq 1$, and $\mathbf{C} = \mathbf{D}^p = (\mathbf{D}_{i,j}^p)_{i,j} \in \mathbb{R}^{n \times n}$ (entrywise power), where $\mathbf{D} \in \mathbb{R}_+^{n \times n}$ is a distance on $\llbracket n \rrbracket$, i.e.

- (i) $\mathbf{D}$ is symmetric;
- (ii) $\mathbf{D}_{i,j} = 0$ if and only if $i = j$;
- (iii) $\forall (i, j, k) \in \llbracket n \rrbracket^3$, $\mathbf{D}_{i,k} \leq \mathbf{D}_{i,j} + \mathbf{D}_{j,k}$.

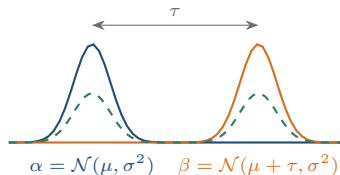
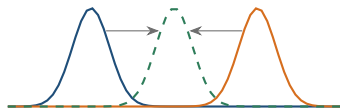
Then

$$\mathcal{W}_p(\mathbf{a}, \mathbf{b}) := L_{\mathbf{D}^p}(\mathbf{a}, \mathbf{b})^{1/p}$$

defines the p -Wasserstein distance on Σ_n : $\mathcal{W}_p$ is symmetric, positive, $\mathcal{W}_p(\mathbf{a}, \mathbf{b}) = 0 \iff \mathbf{a} = \mathbf{b}$, and satisfies the triangle inequality

$$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \Sigma_n, \quad \mathcal{W}_p(\mathbf{a}, \mathbf{c}) \leq \mathcal{W}_p(\mathbf{a}, \mathbf{b}) + \mathcal{W}_p(\mathbf{b}, \mathbf{c}).$$

Measure Metric



Divergences

$$\text{KL}(\alpha \parallel \beta) = \int_{\mathcal{X}} \log \left(\frac{d\alpha}{d\beta} \right) d\alpha$$

$$\|\alpha - \beta\|_{TV} = \int_{\mathcal{X}} |d\alpha - d\beta|$$

$$\mathcal{W}_2(\alpha, \beta) = \left(\min_{\pi \in \mathcal{U}(\alpha, \beta)} \int_{\mathcal{X} \times \mathcal{X}} \|x - y\|^2 d\pi(x, y) \right)^{1/2}$$

Kantorovich Duality (discrete case)

Proposition

$$L_C(\mathbf{a}, \mathbf{b}) = \max_{(\mathbf{f}, \mathbf{g}) \in \mathbf{R}(C)} \langle \mathbf{f}, \mathbf{a} \rangle + \langle \mathbf{g}, \mathbf{b} \rangle, \quad \mathbf{R}(C) := \{(\mathbf{f}, \mathbf{g}) \in \mathbb{R}^n \times \mathbb{R}^m : \mathbf{f}_i + \mathbf{g}_j \leq C_{ij}\}.$$

Example

A shipper charges $\mathbf{f}_i$ to pick up at x_i and $\mathbf{g}_j$ to deliver at y_j . To be competitive, the prices must satisfy $\mathbf{f}_i + \mathbf{g}_j \leq C_{ij}$; maximizing their profit, the shipper exactly matches the optimal cost.

Primal in standard form

We vectorize column-wise: $\mathbf{p} = \text{vec}(\mathbf{P}) \in \mathbb{R}^{nm}$, $\mathbf{c} = \text{vec}(\mathbf{C})$. Both marginal constraints are linear in $\mathbf{p}$:

$$\mathbf{P}\mathbf{1}_m = (\mathbf{1}_m^\top \otimes \mathbf{I}_n) \mathbf{p}, \quad \mathbf{P}^\top \mathbf{1}_n = (\mathbf{I}_m \otimes \mathbf{1}_n^\top) \mathbf{p}.$$

Primal in standard form

$$L_C(\mathbf{a}, \mathbf{b}) = \min_{\mathbf{p} \in \mathbb{R}_+^{nm}} \mathbf{c}^\top \mathbf{p} \quad \text{s.t.} \quad \mathbf{A}\mathbf{p} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \mathbf{1}_m^\top \otimes \mathbf{I}_n \\ \mathbf{I}_m \otimes \mathbf{1}_n^\top \end{bmatrix} \in \mathbb{R}^{(n+m) \times nm}.$$