Optimal Transport

Arnav Hiray
Mentored by Sebastian Gutierrez Hernandez

Fall 2023 Georgia Tech Directed Reading Program (Mathematics)

December 5^{th}, 2023

1 Defining Optimal Transport

2 The Monge Map

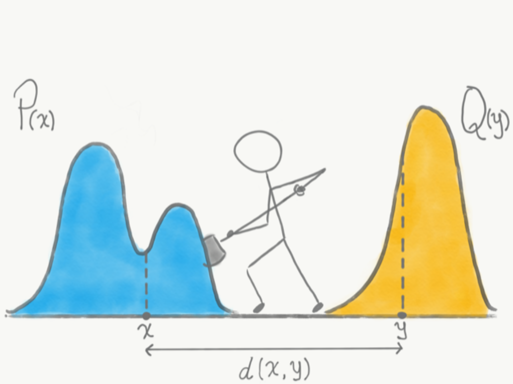
3 Kantorovich Relaxation

Gaspard Monge's Interpretation

- A worker with a shovel, moving a pile of sand into a certain shape.
- Goal: *Minimize* the amount of work required.
- *Transport* the sand in an *optimal* manner.

Optimal Transport

- Optimal Transport can be used to compare two probability distributions.
- This can define a *distance* between distributions.



Where is Optimal Transport Used?

- *Machine Learning and Data Science:*
  - Domain Adaptation and Transfer Learning
- *Economics/Logistics*
- Other applications (Meteorology, Quantum Mechanics etc.)

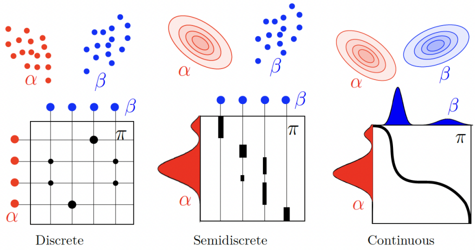


Figure: Image from (Peyré and Cuturi 2020)

Optimal Transport with Graphs

GOT: An Optimal Transport framework for Graph comparison (Maretic et al. 2019)

TLDR: A framework for comparing graphs using Optimal Transport to explicitly express the Wasserstein distance in terms of the graph Laplacian matrices.

Dynamic Optimal Transport on Networks (Berger, Humpert and Pietschmann 2022)

TLDR: In this context, characterized by a transportation cost function that incorporates both the edge traversal cost and a scaling factor based on the transported mass, an optimal transport scenario is defined, minimizing the cumulative transport cost under these constraints.

Optimal Transport Graph Neural Networks (Maretic, Hermina Petric, et al. 2019)

TLDR: Utilizing optimal transport to compute graph embeddings improves performance on various tasks.

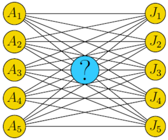
1 Defining Optimal Transport

2 The Monge Map

3 Kantorovich Relaxation

The Assignment Problem

- A company has n workers and n jobs.
- Cost matrix $\mathbf{C}_{n\times n}$ denotes the cost matrix where $\mathbf{C}_{i,j}$ is the cost it takes to pay worker i to do job j.



- Given the set of n permutations $\mathrm{Perm}(n)$

$$\min_{\sigma\in\mathrm{Perm}(n)} \frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_{i,\sigma(i)}$$

Monge Problem

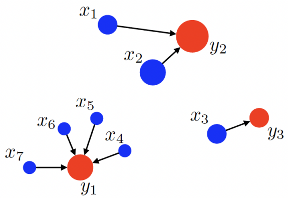
- Find the **Monge Map** T that minimizes the cost $c(x, y)$ (e.g., $||x - y||^2$) in transporting measure α to β.

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

- β is the *push-forward* of α under T if

$$\forall j \in \{1, ..., m\}, \mathbf{b}_j = \sum_{i : T(x_i) = y_j} \mathbf{a}_i$$

Image from (Peyré and Cuturi 2020)



1 Defining Optimal Transport

2 The Monge Map

3 Kantorovich Relaxation

Mines and Factories

- A company has n warehouses and m factories.
- Each warehouse i stores $\mathbf{a}_i$ units of a raw material needed for factories to run.
- Each factory requires $\mathbf{b}_j$ units of raw material to run.
- Transporting a unit of raw material from warehouse i to factory j costs $\mathbf{C}_{i,j}$.

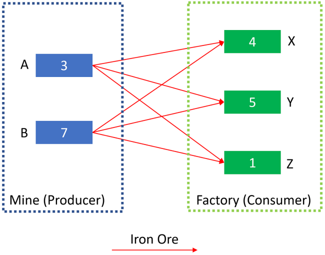


Figure: Image From (Wang 2020)

Mines and Factories

- A company has n warehouses and m factories.
- Each warehouse i stores $\mathbf{a}_i$ units of a raw material needed for factories to run.
- Each factory requires $\mathbf{b}_j$ units of raw material to run.
- Transporting a unit of raw material from warehouse i to factory j costs $\mathbf{C}_{i,j}$.

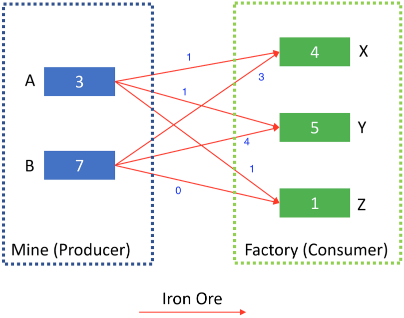


Figure: Image From (Wang 2020)

Kantorovich's Idea

- "Relax" the deterministic nature of transportation.
- What if x_i can be *split* across several locations?
- Now referred to as *mass splitting*.

Generalized Kantorovich Problem

Denote $\mathcal{L}_c(\alpha, \beta)$ as the value of the optimization problem associated to the optimal transport with cost c. Then,

$$\mathcal{L}_c(\alpha, \beta) \overset{\text{def.}}{=} \min_{\pi \in \mathcal{U}(\alpha, \beta)} \int_{\mathcal{X} \times \mathcal{Y}} c(x, y) d\pi(x, y)$$

Georgia Institute of Technology

References

- Peyré, Gabriel, and Marco Cuturi. “Computational Optimal Transport.” 2020. arXiv, arXiv:1803.00567.
- Chen, Benson, et al. “Optimal Transport Graph Neural Networks.” 2021. arXiv, arXiv:2006.04804.
- Maretic, Hermina Petric, et al. "GOT: An Optimal Transport Framework for Graph Comparison." *NeurIPS*, 2019. arXiv, arXiv:1906.02085. *NeurIPS*, 2019
- Burger, Martin, Ina Humpert, and Jan-Frederik Pietschmann. "Dynamic Optimal Transport on Networks." 2022. arXiv, arXiv:2101.03415.

References (Cont.)

- “Measuring Dataset Similarity Using Optimal Transport”, David Alvarez-Melis and Nicolo Fusi, Microsoft Research Blogs
- Jianfeng Wang, ”A Simple Introduction on Sinkhorn Distances.” *Medium*, `https://amsword.medium.com/a-simple-introduction-on-sinkhorn-distances-d01a4ef4f085`.

Kantorovich's Probabilistic Approach

Couplings

We define a coupling measure π between α and β such that

- for any $A \subset \mathcal{X}$, $\pi(A \times \mathcal{Y}) = \alpha(A)$
- for any $B \subset \mathcal{Y}$, $\pi(\mathcal{X} \times B) = \beta(B)$

Kantorovich Problem Between Arbitrary Measures

Consider couplings $\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y})$. Let $\mathcal{U}(\alpha, \beta)$ be a set of coupling measures such that

$$\mathcal{U}(\alpha, \beta) \stackrel{\text{def.}}{=} \left\{\pi \in \mathcal{M}_+^1(\mathcal{X} \times \mathcal{Y}) : P_{\mathcal{X}_\#}\pi = \alpha \text{ and } P_{\mathcal{Y}_\#}\pi = \beta\right\}$$

where $P_{\mathcal{X}_\#}$ and $P_{\mathcal{Y}_\#}$ are push forwards such that $P_{\mathcal{X}}(x, y) = x$ and $P_{\mathcal{Y}}(x, y) = y$

Kantorovich's Probabilistic Approach (Cont.)

Generalized Kantorovich Problem

Denote $\mathcal{L}_c(\alpha, \beta)$ as the value of the optimization problem associated to the optimal transport with cost c. Then,

$$\mathcal{L}_c(\alpha, \beta) \stackrel{\text{def.}}{=} \min_{\pi \in \mathcal{U}(\alpha, \beta)} \int_{\mathcal{X}\mathcal{Y}} c(x, y) d\pi(x, y)$$

Generalized Kantorovich Problem (Probabilistic Interpretation)

$$\mathcal{L}_c(\alpha, \beta) = \min_{(X,Y)} \left\{ \mathbb{E}_{(X,Y)}(c(X, Y)) : X \sim \alpha, Y \sim \beta \right\}$$

Georgia Institute of Technology

Monge Problem

- Cost Matrix $(\mathbf{C}_{i,j})_{i\in\{1,\dots,m\},j\in\{1,\dots,n\}}$
- TODO: Assign n entities to n positions/jobs while minimizing the total cost.
- Given the set of n permutations $\mathrm{Perm}(n)$

$$\min_{\sigma\in\mathrm{Perm}(n)} \frac{1}{n}\sum_{i=1}^{n} \mathbf{C}_{i,\sigma(i)}$$

- Naively trying every permutation is impossible!

Preliminaries

Discrete Measures

A **measure** α is **discrete** with weights $\mathbf{a}$ and locations $x_1, ..., x_n \in \mathcal{X}$ if

$$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i}$$

where δ_x is a Dirac at position x.
Such a measure is a **probability measure** if $\sum_{i=1}^{n} \mathbf{a}_i = 1$.

Monge Problem

- Take two discrete measures

$$\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta_{x_i} \quad \beta = \sum_{j=1}^{n} \mathbf{b}_i \delta_{y_j}$$

- The *Monge Problem* seeks a map relating every x_i to a point y_j.
- Such a map must also "push" the mass of α completely towards β.
- Find a map $T : \{x_1, ..., x_n\} \to \{y_1, ..., y_m\}$ such that

$$\forall j \in \{1, ..., m\}, \mathbf{b}_j = \sum_{i : T(x_i) = y_j} \mathbf{a}_i$$

- Can also be written as $T_{\#}\alpha = \beta$.

Push-Forward: General Case

Push-Forward (General Definition)

For a continuous map $T : \mathcal{X} \to \mathcal{Y}$, the push forward measure $\beta = T_{\#}\alpha \in M(\mathcal{Y})$ for some $\alpha \in M(\mathcal{X})$ such that

$$\forall h \in \mathcal{C}(\mathcal{Y}), \quad \int_{\mathcal{Y}} h(y) d\beta(y) = \int_{\mathcal{X}} h(T(x)) d\alpha(x)$$

Monge Problem with Arbitrary Measures

We can now extend the Monge Problem to two arbitrary probability measures (α, β) on two spaces $(\mathcal{X}, \mathcal{Y})$ using a map $T : \mathcal{X} \to \mathcal{Y}$ that minimizes

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

Push-Forward : Discrete Case

- For discrete measures, push-forward is simply moving the positions of all the points in the support of the measure.

Push-Forward (Discrete)

For a continuous map $T : \mathcal{X} \to \mathcal{Y}$, we define a push-forward operator $T_\# : M(\mathcal{X}) \times M(\mathcal{Y})$ such that

$$T_\# \alpha \overset{\text{def.}}{=} \sum_i \mathbf{a}_i \delta_{T(x_i)}$$