Optimal Transport for Signal Processing

A tutorial at MLSP 2024

Felipe Tobar[1] Laetitia Chapel[2]

[1]Initiative for Data & Artificial Intelligence, Univeridad de Chile

[2]IRISA, Obelix team, Institut Agro Rennes-Angers

22 September, 2024

Overview

Speaker's presentation

Photos and mini-bio

Why Optimal Transport

State main advantages and show applications in different fields. Emphasis on why OT and not other techniques

Origins of OT

First uses: moving mass/dirt

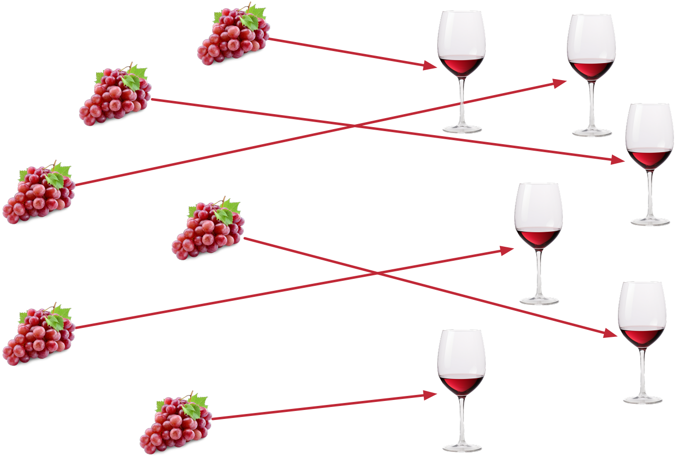
Brief history of OT

From Monge to today

Intuition

Motivation: vineyards transporting grapes from harvest site to processing plants

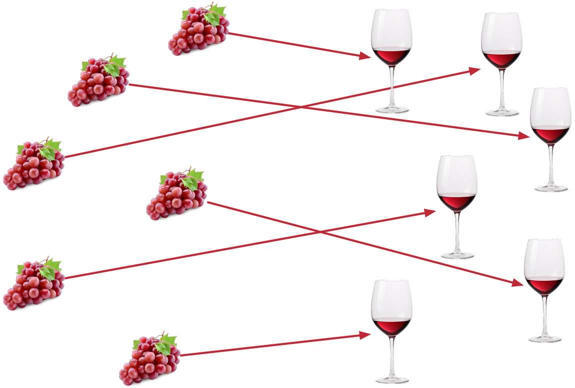
The assigment problem



Build the assignment problem from intuition. Use the above figure to explain all possible ways to assign: straight lines, what's the cost, *cheapest transport*.

The assigment problem: encoding real-world

- Weighted masses
- Different number of sources/targets
- Straight path is not possible
- New sample becomes available



Monge formulation[1]

Objective: Move a pile of mass from one location to another at a minimum effort

Let us first set up our notation

- **Piles of mass** are probability distributions, μ and ν, corresponding to random variables $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$.
- **Moving procedure** is a function $T : x \in \mathcal{X} \mapsto Y \in \mathcal{Y}$.
- **Moving cost** encoded as $c : (x, y) \in \mathcal{X} \times \mathcal{Y} \mapsto c(x, y) \in \mathbb{R}$.
- Optimise the total transport cost

$$\sum_{x \in \mathcal{X}} c(x_i, T(x_i)) \tag{1}$$

over $M_{X,Y} = \{T : \mathcal{X} \to \mathcal{Y},\ s.t.,\ T_{\#\mu} = \nu\}$.

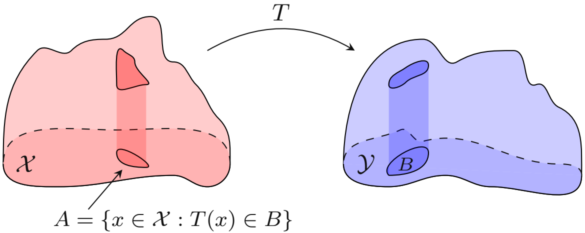
[1] Monge, G. (1781). Mémoire sur la théorie des déblais et des remblais. De l'Imprimerie Royale.

The transport map (aka the *pushforward* operator $T_{\#}$)

T *transports* mass from $\mathcal{X}$ to $\mathcal{Y}$, meaning that for any subset $A \in \mathcal{Y}$, one has

$$\nu(A) = \mu(T^{-1}(A)), \tag{2}$$

where $T^{-1}(A) = \{x \in \mathcal{X}, s.t.\ T(x) \in A\}$ is the preimage of A under T.



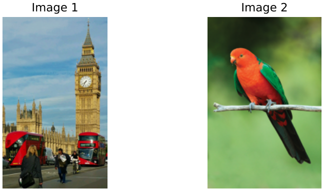
$T\#\mu = \nu$
i.e., $\mu(A) = \nu(B)$

Figure adapted from Thorpe's book.[2]

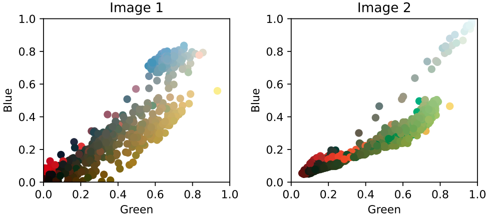
[2] Infinite thanks to Elsa Cazelles (IRIT, CNRS) for kindly sharing these beautiful `tikz` figures for this part.

Example: Colour transfer

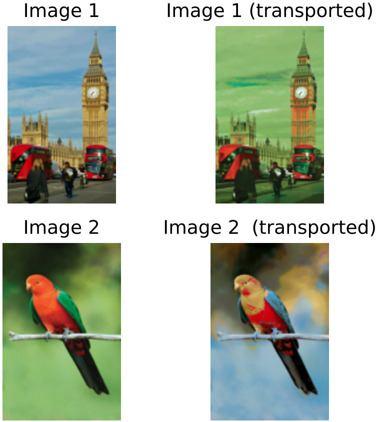
Original images



Histograms



Histograms

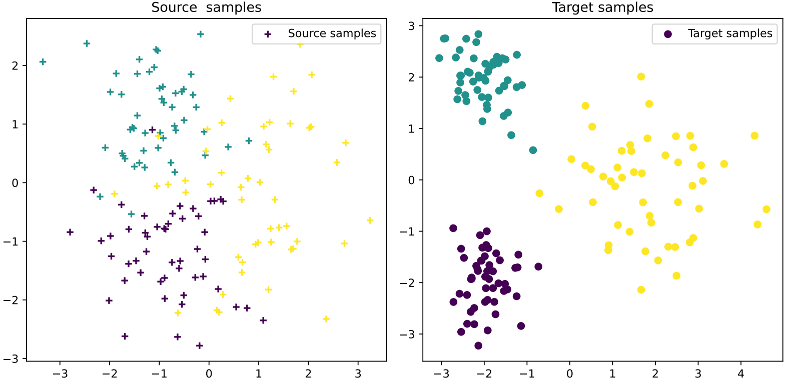


• Ferradans, S., Papadakis, N., Peyre, G., & Aujol, J. F. (2014). Regularized discrete optimal transport. SIAM Journal on Imaging Sciences, 7(3), 1853-1882.

•Code adapted from POT documentation

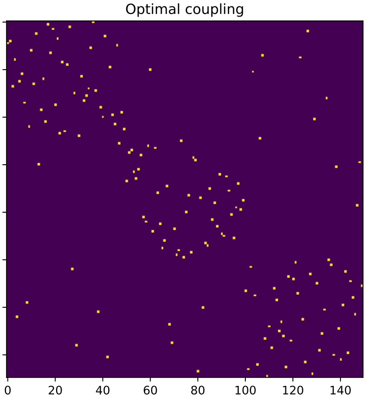
Example of domain adaptation

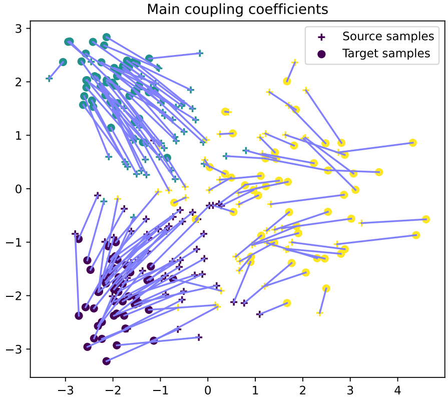
Let us consider the following source and target classification problems



Example of domain adaptation

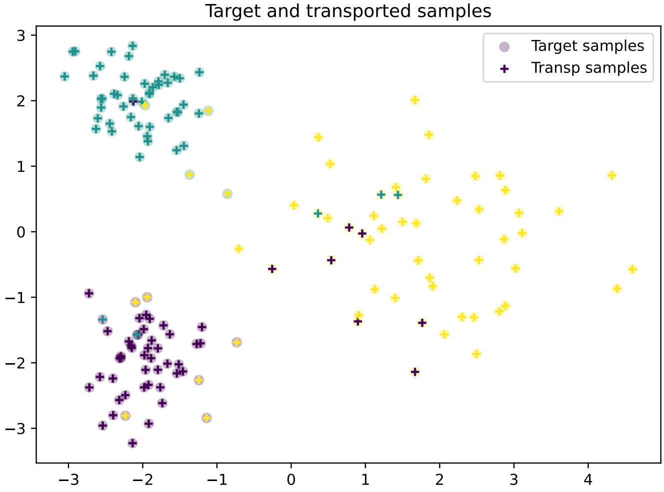
Let us consider the following source and target classification problems





Example of domain adaptation

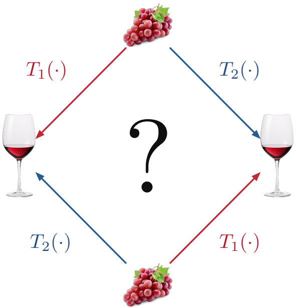
Let us consider the following source and target classification problems

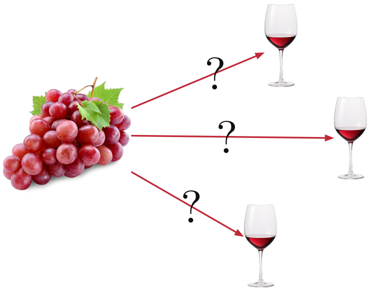


Observation

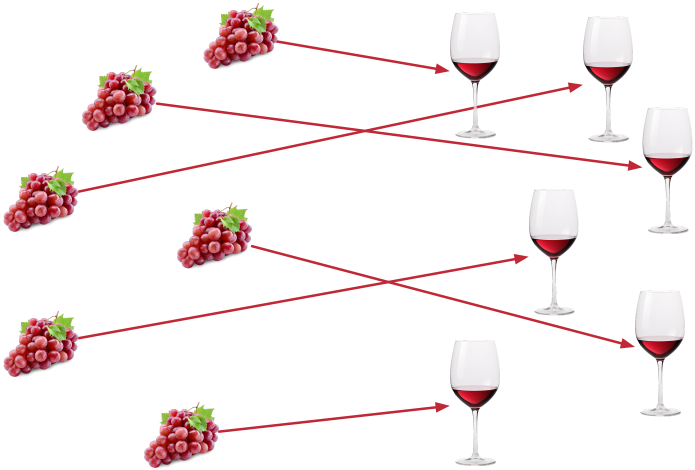
Not all problems feature *atoms*, sometimes we have weighted samples

Neither existence nor uniqueness is guaranteed

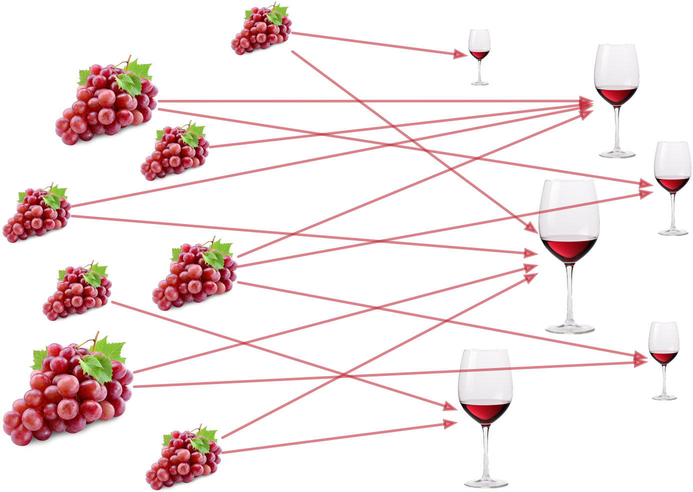




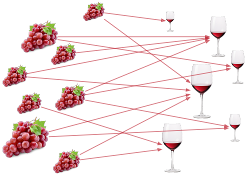
Kantorovich formulation: mass splitting



Kantorovich formulation: mass splitting



Transport plan

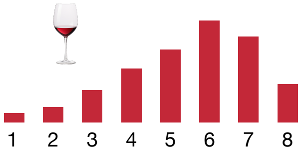


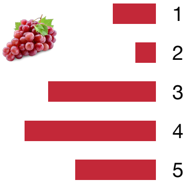
$$\inf_{P \in \Pi_{\mu,\nu}} \langle P, C \rangle = \sum_{i,j}^{n,m} C_{ij} P_{ij}$$

where $\Pi_{\mu,\nu}\langle P, C\rangle = \{P \in [0,1]^{m \times n} : \sum_{i=1}^{m} P_{ij} = \nu_j, \sum_{j=1}^{n} P_{ij} = \mu_i\}$

Transport plan

P





	1	2	3	4	5	6	7	8
1	?	?	?	?	?	?	?	?
2	?	?	?	?	?	?	?	?
3	?	?	?	?	?	?	?	?
4	?	?	?	?	?	?	?	?
5	?	?	?	?	?	?	?	?

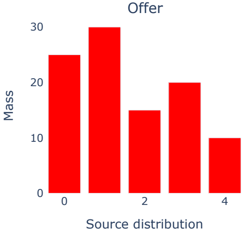
Cost Matrix

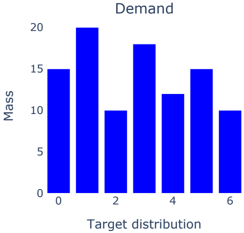
C

	1	2	3	4	5	6	7	8
1	$	$	$	$	$	$	$	$
2	$	$	$	$	$	$	$	$
3	$	$	$	$	$	$	$	$
4	$	$	$	$	$	$	$	$
5	$	$	$	$	$	$	$	$

Example:

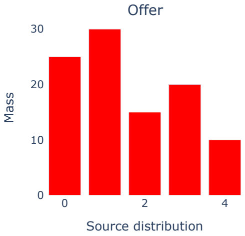
Consider the following source and target distributions

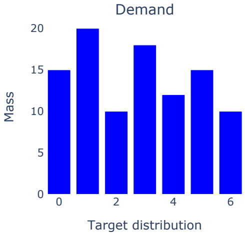


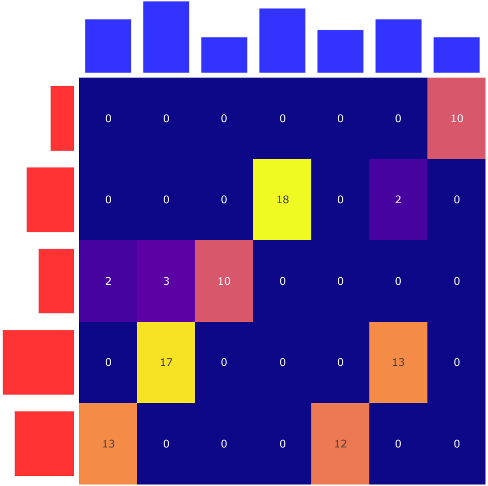


Example:

Consider the following source and target distributions

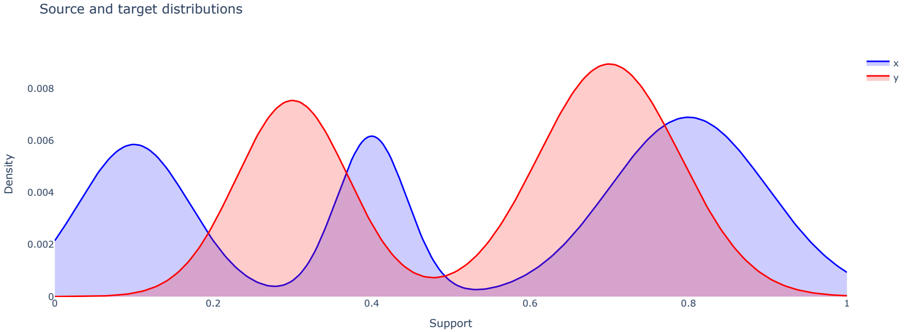






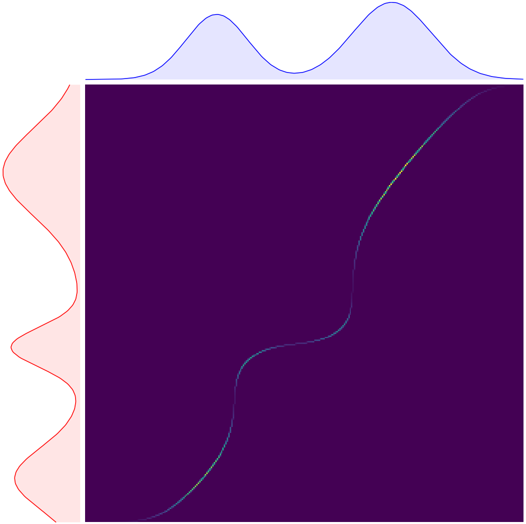
Example (Continuous)

Let us now consider two distributions over a continuous support



Example (Continuous)

Let us now consider two distributions over a continuous support



Observe that the plan remained *sparse*, i.e., the mass did not spread much

This motivates the following results

Observations

- Let us consider a cost $c(x, y) = |x - y|^p, p \geq 1$. Then, if μ and ν are absolutely continuous wrt the Lebesgue measure, the Kantorovich problem as a unique solution. Furthermore, this solution is the same solution of the Monge problem.
- If $p = 2$, the optimal map is the gradient of a convex function
- In some cases the optimal plan will require to split mass (e.g., in the case of atomic measures) and thus Monge's solution may fail to exist.
- Luckily, from a (Kantorovich) transport plan we can always extract a tranport map, e.g., via the barycentric projection

Example: Domain adaptation

Cite some papers and hopefully show some simulations

Dual formulation

Formualtion and illustration with the factories

Examples

See which ones make sense here: Domain Adaptation, OT for fairness, Colour-transfer, WGANs.

Motivation

We need a distance, OT provides one. Show how some *strong* topologies cannot be used for learning systems

The Wasserstein distance

Definition, properties

The Wasserstein space

Definition, properties

Convergence properties, weak topology

Revise the above example

Geodesic properties

Interpolation

The Wasserstein barycenter

Definition, properties, example

Examples

See which ones make sense here: OT spectral transport, Wasserstein Bays, VAEs.

What we did not see

multimarginal, unbalanced OT, partial OT, Gromov-Wasserstein,

Conclusions & the future

Optimal Transport for Signal Processing

A tutorial at MLSP 2024

Felipe Tobar[1] Laetitia Chapel[2]

[1]Initiative for Data & Artificial Intelligence, Univeridad de Chile

[2]IRISA, Obelix team, Institut Agro Rennes-Angers

22 September, 2024