

Computational Optimal Transport: A (biased) Overview

Stefano Gualandi

Università di Pavia, Dipartimento di Matematica

email: stefano.gualandi@unipv.it
twitter: [@famo2spaghi](https://twitter.com/famo2spaghi)
blog: <http://stegua.github.com>
web: <http://matematica.unipv.it/gualandi/>

Comparing images

Imagine that we have to answer to the following question:



(A)

(B)

(C)

e.g.: *A is **closer** to B or C?*

Comparing images

Imagine that we have to answer to the following question:



(A)



(B)

(C)

e.g.: *A is **closer** to B or C?*

Comparing images

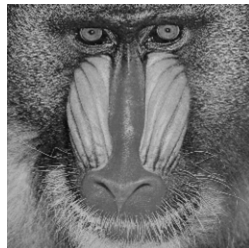
Imagine that we have to answer to the following question:



(A)



(B)



(C)

e.g.: *A is **closer** to B or C?*

Comparing images

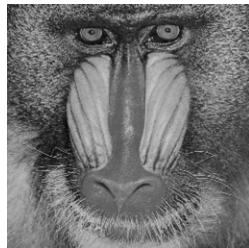
Imagine that we have to answer to the following question:



(A)



(B)

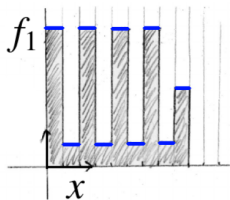


(C)

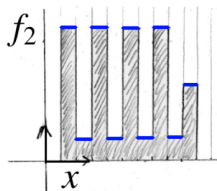
e.g.: *A is **closer** to B or C?*

We need a notion of **DISTANCE** between images.
For instance, we could use the classical L_2 norm.

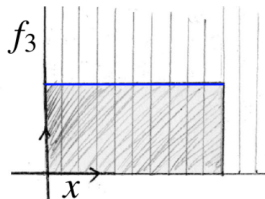
Comparing functions



(A)



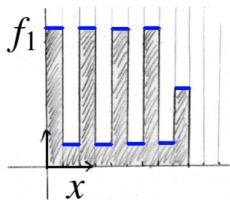
(B)



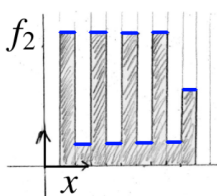
(C)

Is f_1 closer to f_2 or to f_3 ?

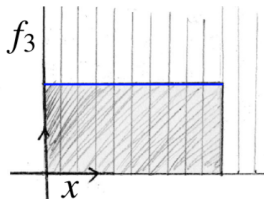
Comparing functions



(A)



(B)

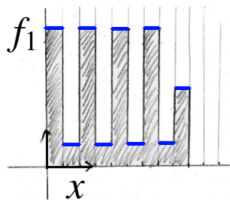


(C)

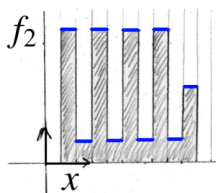
Is f_1 **closer** to f_2 or to f_3 ?

Using the L_2 norm, f_1 is nearer to f_3 than to f_2 .

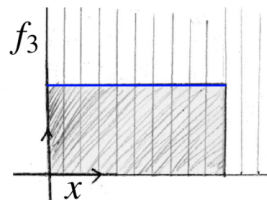
Comparing functions



(A)



(B)



(C)

Is f_1 **closer** to f_2 or to f_3 ?

Using the L_2 norm, f_1 is nearer to f_3 than to f_2 .

We need a new notion of **DISTANCE**, that is able to account for shifts or deformations of f_1 .

Images from: B. Levy, E.L. Schwindt. Notion of optimal transport theory and how to implement them on a computer. Computer and Graphics, 72, pp. 135–148, 2018.

Discrete Probability measures

We define the Dirac δ -measure centered at $x \in X \subseteq \mathbb{R}^q$ via

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \subseteq X \\ 0 & \text{if } x \notin A \subseteq X. \end{cases}$$

Discrete Probability measures

We define the Dirac δ -measure centered at $x \in X \subseteq \mathbb{R}^q$ via

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \subseteq X \\ 0 & \text{if } x \notin A \subseteq X. \end{cases}$$

Let μ and ν be two discrete probability measures defined on a spaces $X, Y \subseteq \mathbb{R}^q$

$$\mu(A) = \sum_{i=1}^m \mu_i \delta_{x_i}(A) \quad \text{and} \quad \nu(A) = \sum_{j=1}^n \nu_j \delta_{y_j}(A)$$

where $\sum_{i=1}^m \mu_i = 1$ and $\sum_{j=1}^n \nu_j = 1$.

Discrete Probability measures

We define the Dirac δ -measure centered at $x \in X \subseteq \mathbb{R}^q$ via

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \subseteq X \\ 0 & \text{if } x \notin A \subseteq X. \end{cases}$$

Let μ and ν be two discrete probability measures defined on a spaces $X, Y \subseteq \mathbb{R}^q$

$$\mu(A) = \sum_{i=1}^m \mu_i \delta_{x_i}(A) \quad \text{and} \quad \nu(A) = \sum_{j=1}^n \nu_j \delta_{y_j}(A)$$

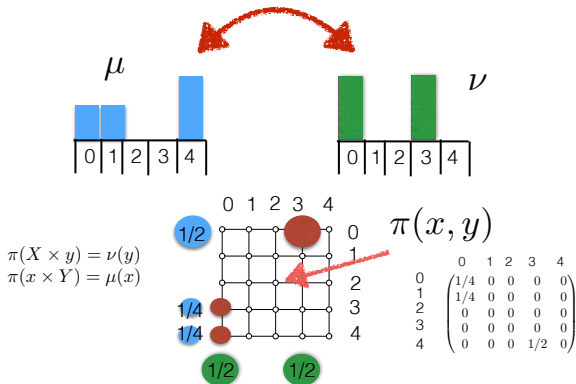
where $\sum_{i=1}^m \mu_i = 1$ and $\sum_{j=1}^n \nu_j = 1$.

Example: Greyscale images are discrete measures: $x_i = (x_{i1}, x_{i2})$ is the pixel position, and μ_i is the pixel intensity at x_i .

(Discrete) Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$ and

- μ and ν are two **(discrete) probability measures** on X and Y
- $c : X \times Y \rightarrow \mathbb{R}_+$ is a cost function.



Discrete Kantorovich-Rubinstein is LP

Definition 1 (Kantorovich-Rubinstein Functional)

The **Kantorovich-Rubinstein functional** in the discrete setting is the following Linear Programming problem (a special case of the Hitchcock Problem)

$$\begin{aligned} \mathcal{W}_c(\mu, \nu) = \min \quad & \sum_{x \in X} \sum_{y \in Y} c(x, y) \pi(x, y) \\ \text{s.t.} \quad & \sum_{y \in Y} \pi(x, y) = \mu(x) & \forall x \in X = \{x_1, \dots, x_m\} \\ & \sum_{x \in X} \pi(x, y) = \nu(y) & \forall y \in Y = \{y_1, \dots, y_n\} \\ & \pi(x, y) \geq 0 \end{aligned}$$

Discrete Kantorovich-Rubinstein is LP

Definition 1 (Kantorovich-Rubinstein Functional)

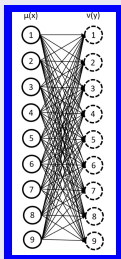
The **Kantorovich-Rubinstein functional** in the discrete setting is the following Linear Programming problem (a special case of the Hitchcock Problem)

$$\mathcal{W}_c(\mu, \nu) = \min \sum_{x \in X} \sum_{y \in Y} c(x, y) \pi(x, y)$$

$$\text{s.t.} \quad \sum_{y \in Y} \pi(x, y) = \mu(x) \quad \forall x \in X = \{x_1, \dots, x_m\}$$

$$\sum_{x \in X} \pi(x, y) = \nu(y) \quad \forall y \in Y = \{y_1, \dots, y_n\}$$

$$\pi(x, y) \geq 0$$

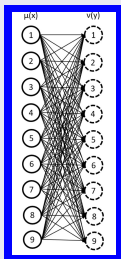


Discrete Kantorovich-Rubinstein is LP

Definition 1 (Kantorovich-Rubinstein Functional)

The **Kantorovich-Rubinstein functional** in the discrete setting is the following Linear Programming problem (a special case of the Hitchcock Problem)

$$\mathcal{W}_c(\mu, \nu) = \min \sum_{x \in X} \sum_{y \in Y} c(x, y) \pi(x, y)$$



$$\text{s.t.} \quad \sum_{y \in Y} \pi(x, y) = \mu(x) \quad \forall x \in X = \{x_1, \dots, x_m\}$$

$$\sum_{x \in X} \pi(x, y) = \nu(y) \quad \forall y \in Y = \{y_1, \dots, y_n\}$$

$$\pi(x, y) \geq 0$$

This problem reduces to an Uncapacitated Min Cost Flow problem on a bipartite graph with $N = n + m$ nodes and (nm) arcs, and can be solved in $O(N^3 \log N)$ time ?? . Indeed, the **Network Simplex** is the fastest exact algorithm.

Wasserstein Distance of order p

Definition 2 (Wasserstein distance)

When $X = Y$ and $c(x, y) = d^p(x, y)$, where d is a (ground) distance on X , the Wasserstein distance of order $p > 0$ is

$$W_p(\mu, \nu) := \mathcal{W}_{d^p}(\mu, \nu)^{\min(\frac{1}{p}, 1)}.$$

Wasserstein Distance of order p

Definition 2 (Wasserstein distance)

When $X = Y$ and $c(x, y) = d^p(x, y)$, where d is a (ground) distance on X , the Wasserstein distance of order $p > 0$ is

$$W_p(\mu, \nu) := \mathcal{W}_{d^p}(\mu, \nu)^{\min(\frac{1}{p}, 1)}.$$

REMARK: To compute the Wasserstein distance between a pair of discrete distributions, we have to solve a Constrained Optimization Problem.

Wasserstein Distance of order p

Definition 2 (Wasserstein distance)

When $X = Y$ and $c(x, y) = d^p(x, y)$, where d is a (ground) distance on X , the Wasserstein distance of order $p > 0$ is

$$W_p(\mu, \nu) := \mathcal{W}_{d^p}(\mu, \nu)^{\min(\frac{1}{p}, 1)}.$$

REMARK: To compute the Wasserstein distance between a pair of discrete distributions, we have to solve a Constrained Optimization Problem.

IN MY RESEARCH PAPERS: We considered Wasserstein distances of order $p = 1$ with ground distances:

p	ground distance	
1	$d(x, y) = \ x - y\ _1$	→ Exact method
1	$d(x, y) = \ x - y\ _\infty$	→ Exact method
1	$d(x, y) = \ x - y\ _2$	→ Exact and approximation
2	$d(x, y) = \ x - y\ _2$	

Wasserstein Distance of order p

Definition 2 (Wasserstein distance)

When $X = Y$ and $c(x, y) = d^p(x, y)$, where d is a (ground) distance on X , the Wasserstein distance of order $p > 0$ is

$$W_p(\mu, \nu) := \mathcal{W}_{d^p}(\mu, \nu)^{\min(\frac{1}{p}, 1)}.$$

REMARK: To compute the Wasserstein distance between a pair of discrete distributions, we have to solve a Constrained Optimization Problem.

IN MY RESEARCH PAPERS: We considered Wasserstein distances of order $p = 1$ with ground distances:

p	ground distance	
1	$d(x, y) = \ x - y\ _1$	→ Exact method
1	$d(x, y) = \ x - y\ _\infty$	→ Exact method
1	$d(x, y) = \ x - y\ _2$	→ Exact and approximation
2	$d(x, y) = \ x - y\ _2$	→ (NeurIPS, 2018)

Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$ and

- μ and ν are two probability measures on X and Y
- $c : X \times Y \rightarrow \mathbb{R}_+$ is a cost function

Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$ and

- μ and ν are two probability measures on X and Y
- $c : X \times Y \rightarrow \mathbb{R}_+$ is a cost function

Definition 3 (Kantorovich-Rubinstein Functional)

The *Kantorovich-Rubinstein functional (I)* between μ and ν is defined as

$$\mathcal{W}_c(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{X \times Y} c(x, y) \pi(dx dy)$$

where $\Pi(\mu, \nu)$ is the set of all probability measures on $X \times Y$ that have marginals μ and ν .

Computational Optimal Transport

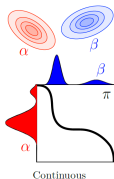
Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$

Computationally, we distinguish three cases:

Computational Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$

Computationally, we distinguish three cases:

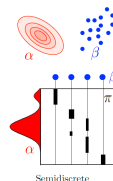
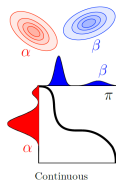


Continuous $\rightarrow$ **Continuous:**
both μ and ν are continuous
functions

Computational Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$

Computationally, we distinguish three cases:



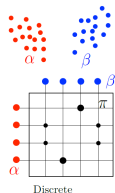
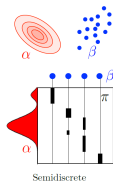
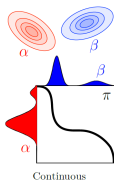
Continuous $\rightarrow$ Continuous:
both μ and ν are continuous
functions

Continuous $\rightarrow$ Discrete:
 μ is continuous, but ν is
a discrete empirical
measure (sum of Diracs)

Computational Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$

Computationally, we distinguish three cases:



Continuous $\rightarrow$ Continuous:
both μ and ν are continuous functions

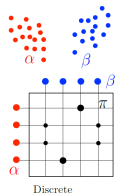
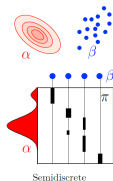
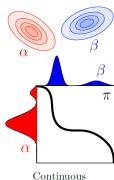
Continuous $\rightarrow$ Discrete:
 μ is continuous, but ν is a discrete empirical measure (sum of Diracs)

Discrete $\rightarrow$ Discrete:
 μ and ν are both discrete empirical measures

Computational Optimal Transport

Problem: find the **Optimal (Mass) Transport** from μ to ν , where moving a unit mass from x to y costs $c(x, y)$

Computationally, we distinguish three cases:



Continuous $\rightarrow$ Continuous:
both μ and ν are continuous
functions

Continuous $\rightarrow$ Discrete:
 μ is continuous, but ν is
a discrete empirical
measure (sum of Diracs)

Discrete $\rightarrow$ Discrete:
 μ and ν are both
discrete empirical
measures

G. Peyr , M. Cuturi. Computational Optimal Transport.
<https://arxiv.org/abs/1803.00567>

Can we reconstruct the initial condition of the Universe?

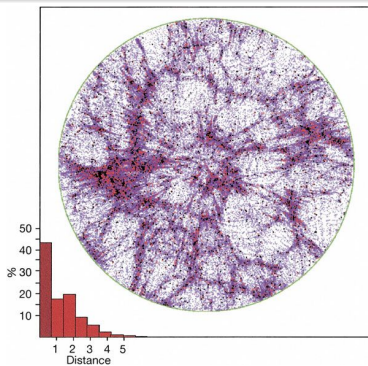


Can we reconstruct the initial condition of the Universe?

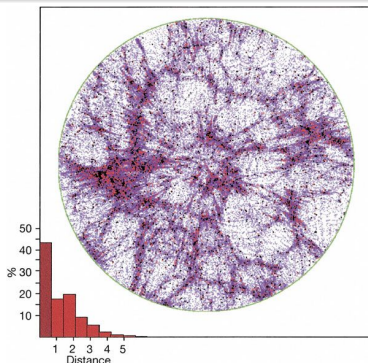


Question: “Can one follow back in time to initial locations the highly structured present distribution of mass in the Universe, as mapped by redshift catalogues of galaxies?”

Can we reconstruct the initial condition of the Universe?



Can we reconstruct the initial condition of the Universe?

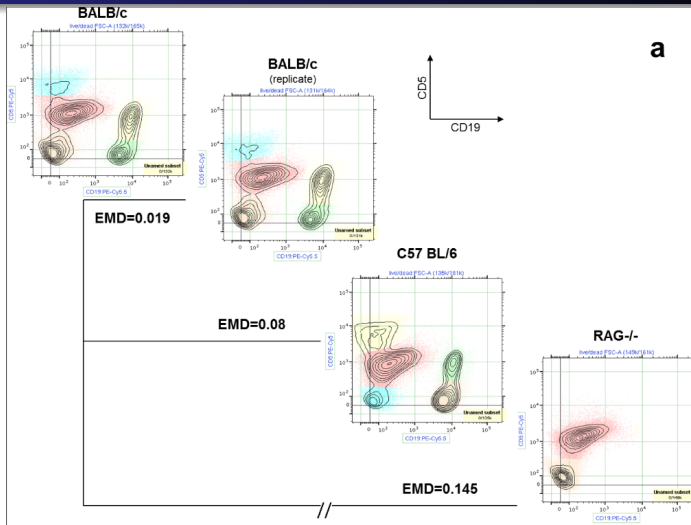


Considering that the initial mass distribution was quasi-uniform, about 14 billion years ago, the initial density field makes reconstruction a well-posed instance of what is called the **Optimal Mass Transportation Problem**.

Frisch, U., Matarrese, S., Mohayaee, R. and Sobolevski, A., 2002. *A reconstruction of the initial conditions of the universe by optimal mass transportation*. Nature, 417(6886), pp. 260–262.

Can we help doctors to diagnose Acute Myeloid Leukemia?

Can we help doctors to diagnose Acute Myeloid Leukemia?



Orlova, D.Y., et al. Earth Mover's Distance (EMD): a true metric for comparing biomarker expression levels in cell populations. PloS one, 11(3), 2016

Can we improve text queries (aka, text Googling)?

D_1 **Obama** **speaks** to the **media** in **Illinois.**

D_0 The **President** greets the **press** in **Chicago.**

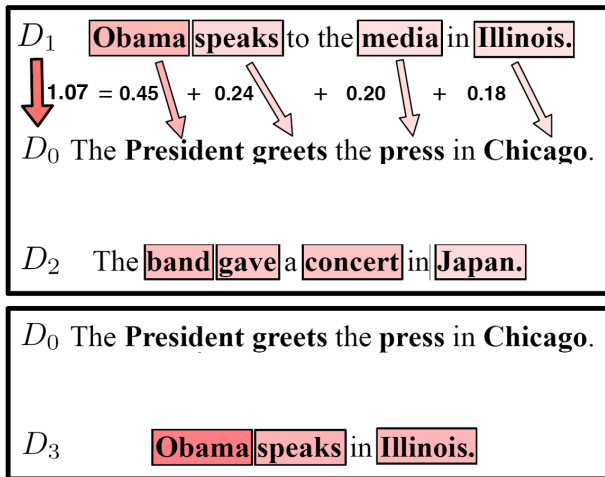
D_2 The **band** **gave** a **concert** in **Japan.**

D_0 The **President** greets the **press** in **Chicago.**

D_3 **Obama** **speaks** in **Illinois.**

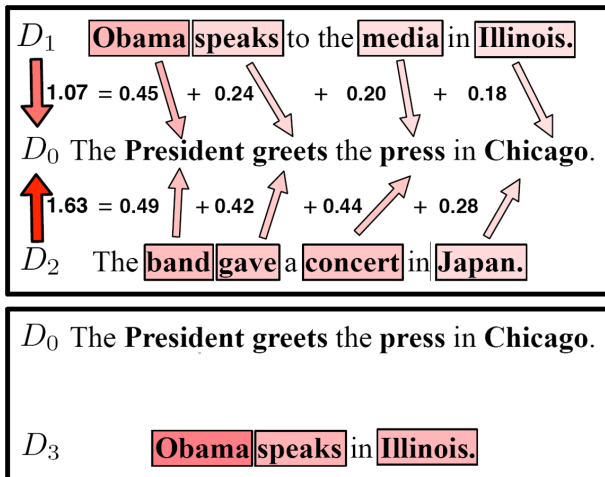
Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In International Conference on Machine Learning (pp. 957-966).

Can we improve text queries (aka, text Googling)?



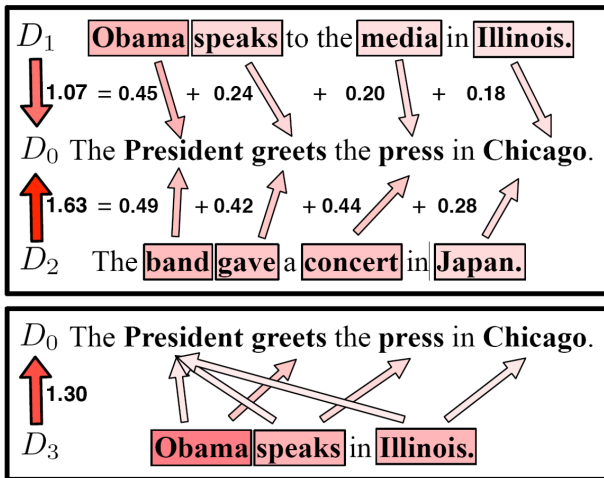
Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In International Conference on Machine Learning (pp. 957-966).

Can we improve text queries (aka, text Googling)?



Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In International Conference on Machine Learning (pp. 957-966).

Can we improve text queries (aka, text Googling)?



Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K., 2015, June. From word embeddings to document distances. In International Conference on Machine Learning (pp. 957-966).

Optimal Transport and Distance between Images

Imagine that we would like to use W_p to answer to the following question:



(A)



(B)

(C)

e.g.: *A is closer to B or C?*

Optimal Transport and Distance between Images

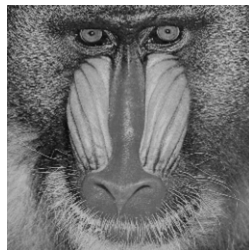
Imagine that we would like to use W_p to answer to the following question:



(A)



(B)



(C)

e.g.: *A is closer to B or C?*

PASSA ALLE ALTRE SLIDES

Thanks for your attention!

