Optimal transport, economic exchanges and healthcare facilities

Xavier Bacon

Exposé au GTMod, Centre Borelli, ENS Paris-Saclay

4 nov. 2024

Introduction

Post-doc SafePaw project (**S**ocietal **A**ssets **F**or **E**-healthcare **P**atient p**A**th**W**ays), under the supervision of Julien Randon-Furling.

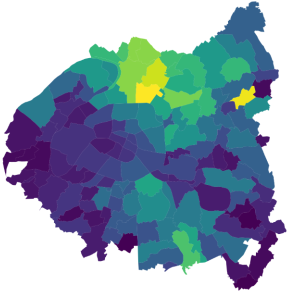
PhD A few models in vector-valued optimal transport, under the supervision of Bruno Nazaret.

Table of Contents

Data available

Patients' diagnosis : information collected when a patient enters a health care facility, recorded as **CMD** (**C**atégorie **M**ajeure de **D**iagnostic) :

01 = Affections du système nerveux,
02 = Affections de l'œil,
03 = Affections des oreilles, du nez, de la gorge, de la bouche et des dents
...



The Kullback-Leibler divergence is used to measure the statistical difference in the CMD's frequency distribution, that is

$$\mathrm{div}_{KL}(\mathrm{CMD}[\text{area A}] \| \mathrm{CMD}[\text{area B}]),$$

In the (classical) pointwise model, area A = municipality and area B = France.

Multiscalar Lens Model (Olteanu & co, 2019)

- **Capturing both the national and the local scale.**

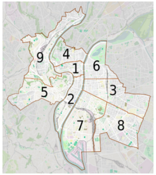
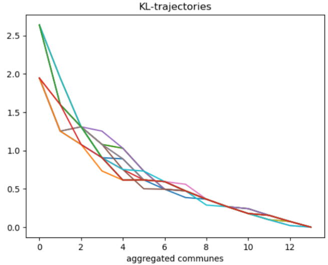


Figure: Districts in Lyon (source : Wikipedia)

Aggregation process : successively computing the divergence of the commune, then the divergence of commune gathered with its nearest neighbor etc. :

- $\mathrm{div}_{KL}(\mathrm{Lyon}_1 \cup \mathrm{Lyon}_4 | \mathrm{France})$
- $\mathrm{div}_{KL}(\mathrm{Lyon}_1 \cup \mathrm{Lyon}_4 | \mathrm{France})$
- $\mathrm{div}_{KL}(\mathrm{Lyon}_1 \cup \mathrm{Lyon}_4 \cup \mathrm{Lyon}_6 | \mathrm{France})$

The **speed of convergence** of the curve tells us how far we have to move away from our starting municipality to look like the national mean.



- (in progress) Take the socio-demographic information into account : Poisson regression, causal inference with observational data.

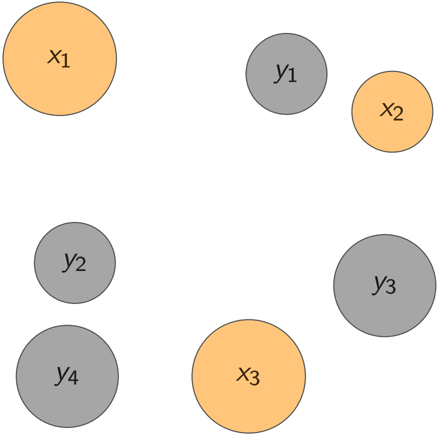
Table of Contents

What is Optimal Transport? 1/2

How can we describe the transportation of mass ?

What is Optimal Transport? 1/2

How can we describe the transportation of mass ?

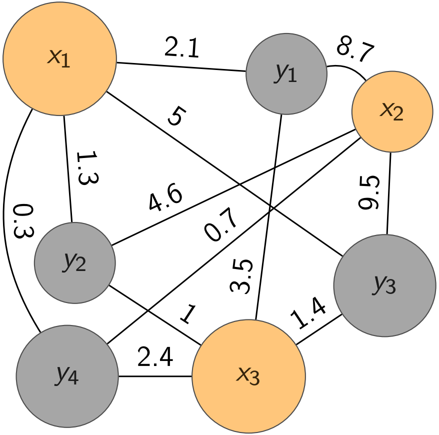


$X = \{x_1, x_2, x_3\}, \; Y = \{y_1, y_2, y_3, y_4\}$

μ and ν **two weighted point clouds**

What is Optimal Transport? 1/2

How can we describe the transportation of mass ?



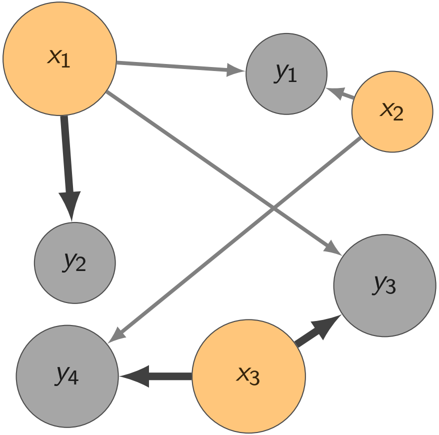
$X = \{x_1, x_2, x_3\}, \ Y = \{y_1, y_2, y_3, y_4\}$

μ and ν **two weighted point clouds**

$c = c(x, y)$ **transport cost**

What is Optimal Transport? 1/2

How can we describe the transportation of mass ?



$X = \{x_1, x_2, x_3\},\ Y = \{y_1, y_2, y_3, y_4\}$

μ and ν **two weighted point clouds**

$c = c(x, y)$ **transport cost**

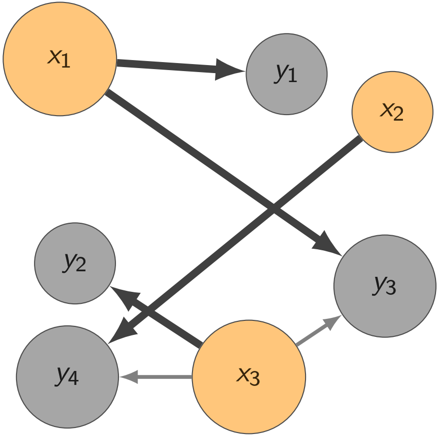
$\gamma =$ **transference plan**

0.1	0.2	0.1	0
0.1	0	0	0.1
0	0	0.2	0.2

$\mathbf{Cost}_c(\gamma) = \langle c | \gamma \rangle$

What is Optimal Transport? 1/2

How can we describe the transportation of mass ?



$X = \{x_1, x_2, x_3\}, \ \ Y = \{y_1, y_2, y_3, y_4\}$

μ and ν **two weighted point clouds**

$c = c(x,y)$ **transport cost**

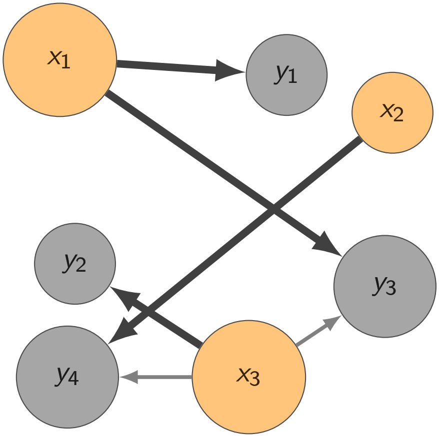
$\gamma =$ **transference plan**

0.2	0	0.2	0
0	0	0	0.2
0	0.2	0.1	0.1

$\text{Cost}_c(\gamma) = \langle c \,|\, \gamma \rangle$

What is Optimal Transport? 1/2

Among all transference, which is the best?



$X = \{x_1, x_2, x_3\}, \; Y = \{y_1, y_2, y_3, y_4\}$

μ and ν **two weighted point clouds**

$c = c(x, y)$ **transport cost**

$\gamma =$

0.2	0	0.2	0
0	0	0	0.2
0	0.2	0.1	0.1

transference plan

$\textbf{Cost}_c(\gamma) = \langle c | \gamma \rangle$

What is Optimal Transport? 2/2

- $\mu \in \mathscr{P}(X)$, $\nu \in \mathscr{P}(Y)$ are respectively the source and the target distributions.
- $c = c(x, y)$ is the **transport cost**.

The (Static) Optimal Transport problem is defined as follows

$$\mathrm{OT}_c(\mu, \nu) = \min \left\{ \int_{X \times Y} c(x, y)\, d\gamma(x, y) \text{ s.t. } \gamma \in \Pi(\mu, \nu) \right\},$$

where $\Pi(\mu, \nu)$ denotes the set of all $\gamma \in \mathscr{P}(X \times Y)$ having μ and ν as marginals, called **transference plans**.

(Discrete setting) $\mu = \sum_{i=1}^{n} \alpha_i \delta_{x_i}$, $\nu = \sum_{j=1}^{m} \beta_j \delta_{y_j}$, with $\sum_{i=1}^{n} \alpha_i = \sum_{j=1}^{m} \beta_j = 1$, $\alpha_i, \beta_j \geq 0$,

$$\mathrm{OT}_c(\mu, \nu) = \mathrm{OT}_c(\alpha, \beta) = \min \left\{ \langle c \,|\, \gamma \rangle \text{ s.t. } \gamma \cdot \mathbb{1} = \alpha, \gamma^T \cdot \mathbb{1} = \beta \right\}.$$

Wasserstein distance

p-Wasserstein distance

If (X,d) is a metric space and $c = d^p$, then

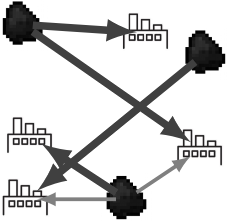
$$\mathrm{OT}_c \text{ is a distance on } \mathscr{P}(X) \text{ if } 0 < p \leqslant 1,$$

$$\mathrm{OT}_c^{1/p} \text{ is a distance on } \mathscr{P}(X) \text{ if } p > 1.$$

- OT_c metrizes the weak-* convergence of probabilities (convergence in Law).
- If $p = 2$, OT_2 is a geodesical distance: any pair of probabilities μ and ν can be connected by a continuous path of length $\mathrm{OT}_2(\mu,\nu)$.

A short digression about duality 1/2

An external provider (say Alice) offers to transport the coal with her own trucks according to the following contract.



Alice chooses the price $\varphi(x)$ for loading in x.
Alice chooses the price $\psi(y)$ for unloading in y.
Alice assures me that it will cost less than doing it myself

$$\varphi(x) + \psi(y) \leqslant c(x, y).$$

Alice charges me $\langle \varphi \,|\, \mu \rangle + \langle \psi \,|\, \nu \rangle$.

$$\mathrm{Dual}(\mu, \nu) = \max \left\{ \langle \varphi \,|\, \mu \rangle + \langle \psi \,|\, \nu \rangle : \varphi(x) + \psi(y) \leqslant c(x, y) \right\}. \qquad \text{(Alice problem)}$$

A short digression about duality 2/2

$$\mathrm{Dual}(\mu,\nu) = \max\{\langle\varphi|\mu\rangle + \langle\psi|\nu\rangle : \varphi(x)+\psi(y) \leqslant c(x,y)\}$$

What is the link between $\mathrm{Dual}(\mu,\nu)$ and $\mathrm{OT}(\mu,\nu)$?

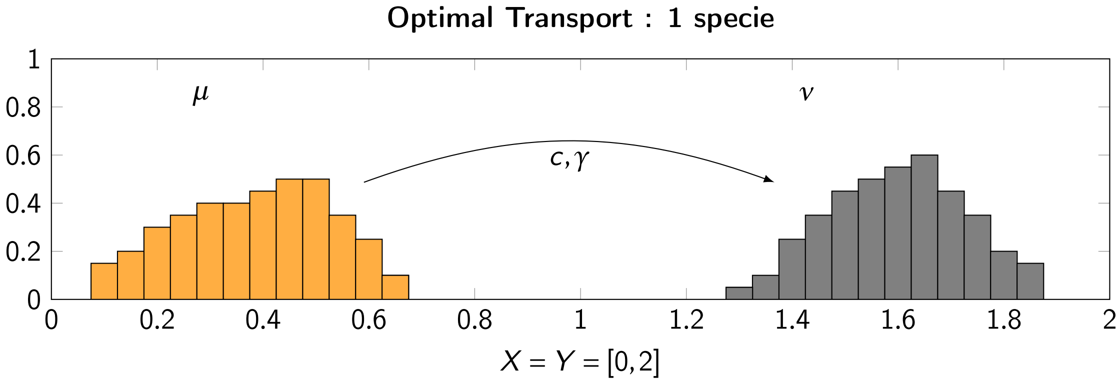
Theorem (strong duality)

$\mathrm{Dual}(\mu,\nu) = \mathrm{OT}(\mu,\nu)$ and any solution of $\mathrm{Dual}(\mu,\nu)$ leads to a solution of $\mathrm{OT}(\mu,\nu)$.

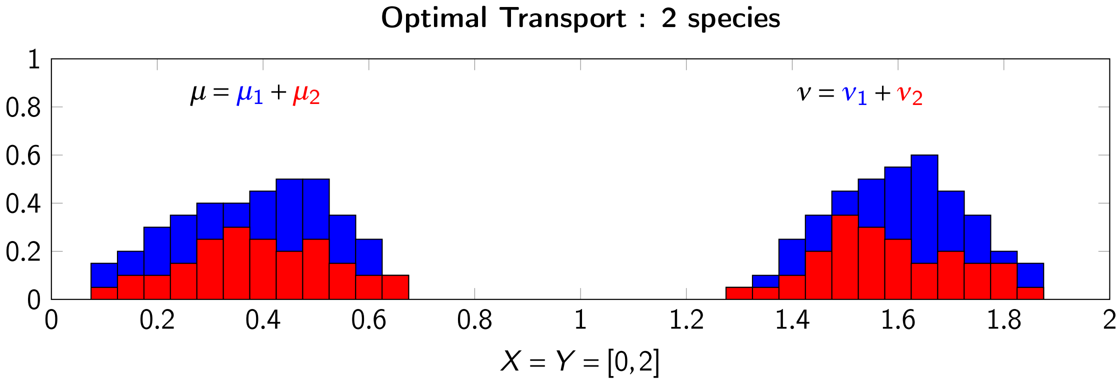
$\mathrm{OT}(\mu,\nu)$ is a optimization problem on $\mathbb{R}^{n\times m}$ while $\mathrm{Dual}(\mu,\nu)$ is on $\mathbb{R}^{n+m}$.

Table of Contents

From scalar OT to multi-species OT

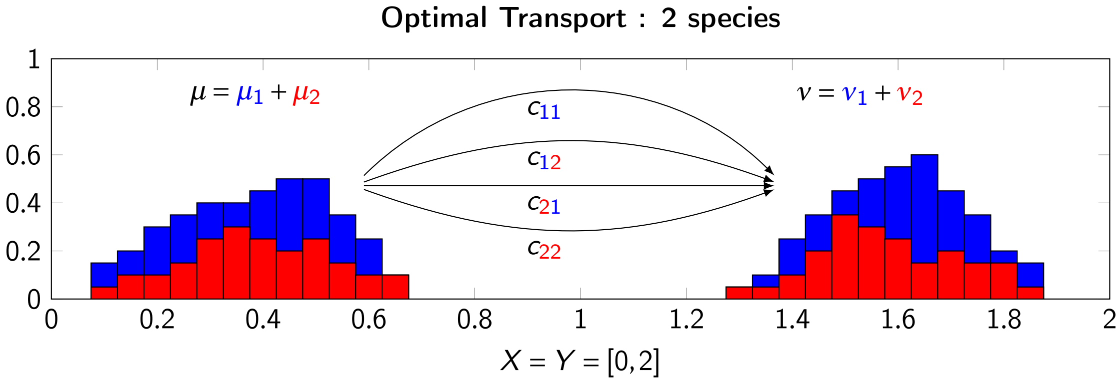


From scalar OT to multi-species OT



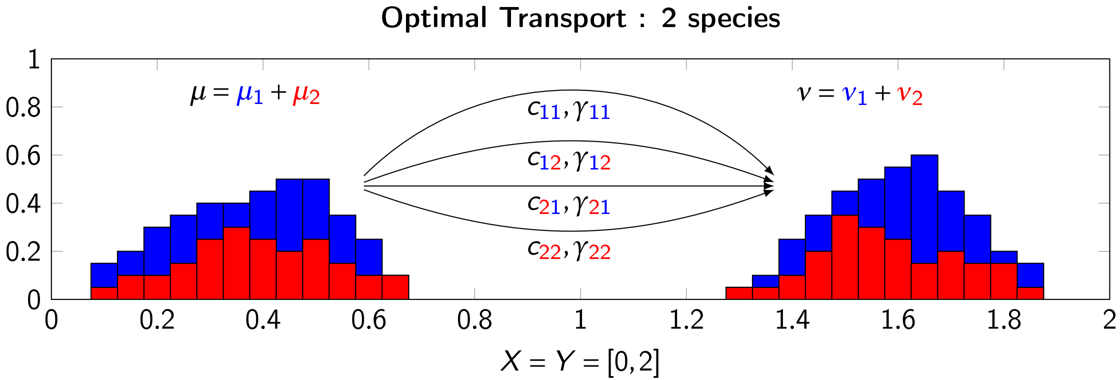
- $\mu = \mu_1 + \mu_2 \in \mathscr{P}(X)$, $\nu = \nu_1 + \nu_2 \in \mathscr{P}(Y)$.

From scalar OT to multi-species OT



- $\mu = \mu_1 + \mu_2 \in \mathscr{P}(X)$, $\nu = \nu_1 + \nu_2 \in \mathscr{P}(Y)$.
- Four costs c_{11}, c_{12}, c_{21}, c_{22}.

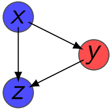
From scalar OT to multi-species OT



- $\mu = \mu_1 + \mu_2 \in \mathscr{P}(X)$, $\nu = \nu_1 + \nu_2 \in \mathscr{P}(Y)$.
- Four costs c_{11}, c_{12}, c_{21}, c_{22}.
- ...And as much transference plans γ_{11}, γ_{12}, γ_{21}, γ_{22}.

Distance on $\mathbb{R}_+^n$-valued measures

- **Classical OT:** if (X,d) is a Polish space, for $p \in [1,\infty)$ and $c = d^p$, $\mathcal{T}^{1/p}$ is a distance on $\mathscr{P}_p(X)$.



- **Multi-species OT:** setting $c_{ij} = d_{ij}^p$, we make the following assumption

$$d_{ik}(x,z) \leqslant d_{ij}(x,y) + d_{jk}(y,z) \quad \textbf{\textit{(Mixed Triangles Inequalities)}}$$

Theorem (B. 2020)

Given n^2 functions (d_{ij}) defined on $X \times X$ and $\mathbb{R}_+$-valued such that
[1] $\forall (i,j) \in [\![1,n]\!]^2, d_{ij}$ is symmetric.
[2] **(MTI)** is satisfied for all $(i,j,k) \in [\![1,n]\!]^3$ et $(x,y,z) \in X^3$.
[3] $\forall i \in [\![1,n]\!], \forall x \in X, d_{ii}(x,x) = 0$.
[4] $\forall (i,j) \in [\![1,n]\!]^2, i \neq j, \forall (x,y) \in X \times X, d_{ij}(x,y) \neq 0$.
Then MultiOT is a distance between multi-valued probabilities.

Table of Contents

Primal problem

- X is a compact metric space.

Primal problem

$$(\mathscr{P}) = \max \Big\{ \qquad\qquad\qquad\qquad \Big\}.$$

Primal problem

- X is a compact metric space.
- $\mu = (\mu_1, \dots, \mu_N) \in \mathscr{M}_+(X)^N$ are the source distributions of goods in region X.

Primal problem

$$(\mathscr{P}) = \max \Big\{ \qquad \Big\}.$$

Primal problem

- X is a compact metric space.
- $\mu = (\mu_1, \ldots, \mu_N) \in \mathscr{M}_+(X)^N$ are the source distributions of goods in region X.
- $\nu = (\nu_1, \ldots, \nu_N) \in \mathscr{M}_+(X)^N$ are the target distributions of goods in region X.

Primal problem

$$(\mathscr{P}) = \max_{\boldsymbol{\nu}} \Big\{ \qquad \text{s.t. } \mu_i(X) = \boldsymbol{\nu_i}(X) \Big\}.$$

Primal problem

- X is a compact metric space.
- $\mu = (\mu_1, \ldots, \mu_N) \in \mathcal{M}_+(X)^N$ are the source distributions of goods in region X.
- $\nu = (\nu_1, \ldots, \nu_N) \in \mathcal{M}_+(X)^N$ are the target distributions of goods in region X.
- $\mathcal{U} = \mathcal{U}(\boldsymbol{\nu})$ is the average utility.

Primal problem

$$(\mathscr{P}) = \max_{\boldsymbol{\nu}} \left\{ \mathcal{U}(\boldsymbol{\nu}) \qquad \text{s.t. } \mu_i(X) = \boldsymbol{\nu_i}(X) \right\}.$$

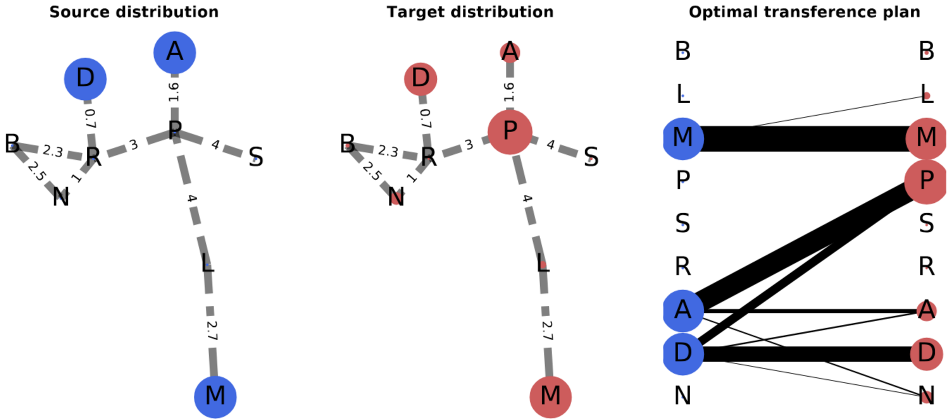
Primal problem

- X is a compact metric space.
- $\mu = (\mu_1, \ldots, \mu_N) \in \mathcal{M}_+(X)^N$ are the source distributions of goods in region X.
- $\nu = (\nu_1, \ldots, \nu_N) \in \mathcal{M}_+(X)^N$ are the target distributions of goods in region X.
- $\mathcal{U} = \mathcal{U}(\boldsymbol{\nu})$ is the average utility.
- $\mathcal{T}(\mu, \boldsymbol{\nu}) = \sum_{i=1}^{N} \mathrm{OT}_{c_i}(\mu_i, \boldsymbol{\nu_i})$ is the transport cost between μ and $\boldsymbol{\nu}$.

Primal problem

$$(\mathcal{P}) = \max_{\boldsymbol{\nu}} \left\{ \mathcal{U}(\boldsymbol{\nu}) - \mathcal{T}(\mu, \boldsymbol{\nu}) \text{ s.t. } \mu_i(X) = \boldsymbol{\nu_i}(X) \right\}.$$

An example


Source distribution
Target distribution
Optimal transference plan
A
D
P
B
R
N
S
L
M
1.6
0.7
2.3
2.5
1
3
4
4
2.7

Economic interpretation

Let (β,φ) be optimal in $(\mathscr{P})$ and $(\mathscr{D})$. Then we have an **equilibrium** for the initial monetary endowment $\boldsymbol{w} = \langle \boldsymbol{\varphi^c} \,|\, \boldsymbol{\alpha} \rangle + \langle \boldsymbol{\varphi} \,|\, \boldsymbol{\beta} \rangle (= \mathscr{T}_c(\alpha,\beta))$ in the sense that:

Sellers at x maximize their profits by exporting their goods $\alpha(x)$:

$$\text{profits}_i(x) = \max_y \varphi_i(y) - c_i(x,y) \quad (= -\varphi_i^{c_i}(x))$$

$$\text{total profits}(x) = \langle \text{profits}(x) \,|\, \alpha(x) \rangle$$

Consumers in y have an initial endowment $w(y)$ and buy $\beta_i(y)$ in such a way:

$$\beta_i(y) = \operatorname*{argmax}_{\beta} \left\{ \boldsymbol{U}(\boldsymbol{y},\beta) \text{ s.t. } \langle \varphi \,|\, \beta \rangle \leqslant \underbrace{\overbrace{\langle -\varphi^c(y) \,|\, \alpha(y) \rangle}^{\text{export profit}} + w(y)}_{\text{total revenue}} \right\}$$

Markets are clear : it exists an optimal transport plan for all target endowments.

Summarizing table

	Unregularized	ε-regularized
Primal	$\overbrace{\sup_{\nu} \mathscr{U}(\nu) - \mathscr{T}(\mu,\nu)}^{(\mathscr{P})}$	$\overbrace{\sup_{\nu} \mathscr{U}(\nu) - \mathscr{T}_\varepsilon(\mu,\nu)}^{(\mathscr{P}_\varepsilon)}$
Dual	$\overbrace{\inf_{\varphi} \mathscr{K}(\varphi) + \mathscr{V}(\varphi)}^{(\mathscr{D})}$	$\overbrace{\inf_{\varphi} \mathscr{V}(\varphi) + \sum_{x\in X} \mu(x)\, softmax_\varepsilon(\varphi - c(x,\cdot))}^{(\mathscr{D}_\varepsilon)}$

In order to solve $(\mathscr{P}) \leftarrow$ introduce $(\mathscr{P}_\varepsilon)$

$\iff$ solving $(\mathscr{D}_\varepsilon)$

$\rightarrow$ we solve it by coordinate ascent (linear convergence)

Conclusion

Merci pour votre attention !