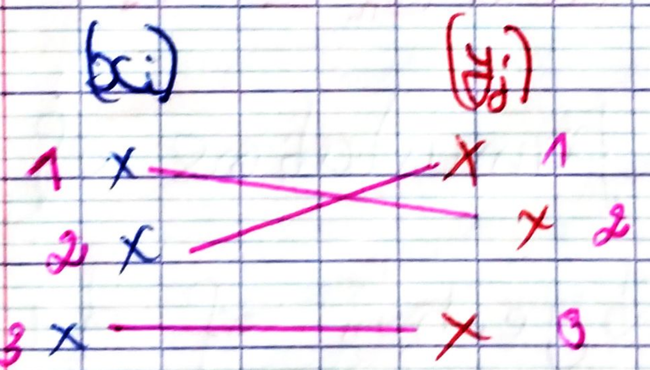
17.09.24

Computational optimal transport

But : Comparer 2 distributions de proba

I : Optimal Transport between points clouds

1) Matching problem



$\sigma \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$

Rq : $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix} = Id$

Monge's problème : Minimize the cost ~~of~~ of matching 2 clouds of points

Let $(\underset{\overset{\parallel}{\mathbb{R}^d}}{X}, d)$ be a metric space

$\forall \ (x_i)_i, (y_i)_i \in \mathcal{X}^n,$

$$\boxed{\bar{\sigma} \in \underset{\sigma \in \mathcal{J}_n}{\arg\min} \ \frac{1}{n} \sum C_{i,\sigma(i)}} \quad (M)$$

~~C_{ij}~~ $C_{i,j} \in \mathbb{R}^{n \times n}$

example: $C_{i,j} = d(x_i, y_j)$

Rq: $\operatorname{card} \mathcal{J}_n = n!$

σ_n permutations of $\{1, .., n\}$

$=$ bijection of $\{1, .., n\}$

2: 1D case $(d = 1)$

On suppose $C_{ij} = h(x_i - y_j)$

avec $h : \mathbb{R} \longrightarrow \mathbb{R}_+$ convexe

Ex: $C_{i,j} = |x_i - y_j|^p, \ p \geq 1$

$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$ $\mapsto$ indice trié, $\mapsto$ indice original

$\mathcal{J}_n$ est le groupe de permutation

WLOG = without loss of generality

map = application

NB: Les normes p sont convexes $(p \geq 1)$ donc en 1D $|\cdot|^p$ est convexe

Lemma: Any optimal map $\bar{\sigma}$ of (M) defines an increasing ~~map~~

map: $x_i \mapsto y_{\sigma(i)}$

$$\forall (i,j), \ (x_i - x_j)(y_{\bar{\sigma}(i)} - y_{\bar{\sigma}(j)}) \geq 0.$$

Proof: Let $\bar{\sigma}$ be an optimal map such that

$$\exists i,j: \ (x_i - x_j)(y_{\bar{\sigma}(i)} - y_{\bar{\sigma}(j)}) < 0.$$

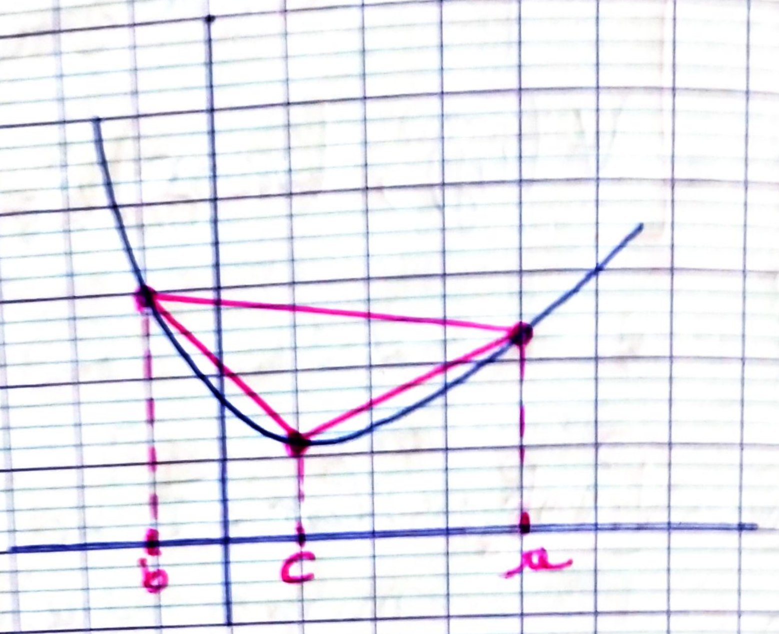
WLOG: $\begin{cases} x_i \geq x_j \\ \underbrace{y_{\bar{\sigma}(i)}}_{c} < y_{\bar{\sigma}(j)} \end{cases}$

TPg: $\overbrace{h(x_i - y_{\bar{\sigma}(j)})}^{c} + \overbrace{h(x_j - y_{\bar{\sigma}(i)})}^{d}$

$< \underbrace{h(x_i - y_{\bar{\sigma}(i)})}_{a} + \underbrace{h(x_j - y_{\bar{\sigma}(j)})}_{b}$

Reminder : Pour toute fonction convexe,

On a : $f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y)$



$b < c < a$,

$$\frac{f(c)-f(b)}{c-b} \leq \frac{f(a)-f(b)}{a-b} \leq \frac{f(a)-f(c)}{a-c}$$

- $\begin{cases} -y_{\sigma(i)} > -y_{\sigma(j)} \\ x_i = x_i \end{cases}$

$x_i - y_{\sigma(i)} > x_i - y_{\sigma(j)}$

$d > c$

- $\begin{cases} x_j \leq x_i \\ -y_{\sigma(j)} = -y_{\sigma(j)} \end{cases}$

$$x_j - y_{\sigma(j)} \leq x_i - y_{\sigma(j)}$$
$$b \leq c$$

Donc : $\underline{b \leq c \leq a}$.

- $\begin{cases} x_j \leq x_i \\ -y_{\sigma(i)} = -y_{\sigma(i)} \end{cases}$

$$x_j - y_{\sigma(i)} \leq x_i - y_{\sigma(i)}$$
$$d \leq a$$

- $\begin{cases} x_j = x_j \\ -y_{\sigma(i)} > -y_{\sigma(j)} \end{cases}$

$$x_j - y_{\sigma(i)} > x_j - y_{\sigma(j)}$$
$$d > b$$

Donc $\underline{b < d \leq a}$

Par convexité de f, on a :

$$\begin{cases} \dfrac{f(c)-f(b)}{c-b} \le \dfrac{f(a)-f(b)}{a-b} \le \dfrac{f(a)-f(c)}{a-c} \\[2ex] \dfrac{f(d)-f(b)}{d-b} \le \dfrac{f(a)-f(b)}{a-b} \le \dfrac{f(a)-f(d)}{a-d} \end{cases}$$

On a donc $\dfrac{f(c)-f(b)}{c-b} \le \dfrac{f(a)-f(d)}{a-d}$

or $c-b = a-d = x_i - x_j$

Car $\begin{cases} c-b = x_i - y_{\sigma(i)} - x_j + y_{\sigma(i)} = x_i - x_j \\ a-d = x_i - y_{\sigma(i)} - x_j + y_{\sigma(i)} = x_i - x_j \end{cases}$

$\Rightarrow f(c) - f(b) \le f(a) - f(d)$

⚠ On vient d'infirmer l'optimalité

d'où la preuve.

NB. Le transport optimal a 1D est la même chose que le tri.

Algo. 1)

$$\begin{cases} \sigma_X : x_{\sigma_X(1)} \le \dots \le x_{\sigma_X(n)} \to \Theta(n\log n) \\ \sigma_Y : x_{\sigma_Y(1)} \le \dots \le x_{\sigma_Y(n)} \to \Theta(n\log n) \end{cases}$$

Optimal match :

$$\bar{\sigma} = \sigma_Y \circ \sigma_X^{-1} \to \frac{-1 \; \Theta(n)}{\Theta(n\log n)}$$

Il est vrai aussi que
$c(x,y) = h(|\varphi(x) - \varphi(y)|)$ avec
$\varphi \nearrow +$

- Ça ne marche pas en 2D
- si h est concave OK $\to \Theta(n^2)$
- $d \geq 2$ algo Hungarian $\to \Theta(n^3)$

NB : L'ordre de l'algo $= n\log n$

Optimal map :

$\sigma : \{1, .., n\} \longrightarrow \{1, .., n\}$

$x_i = x_{\sigma_X^{-1}(i)}$

$y_i = y_{\sigma_Y^{-1}(i)}$

$$\begin{cases} i = \sigma_X(k) \Rightarrow k = \sigma_X^{-1}(i) \\ j = \sigma_Y(k) \Rightarrow = \sigma_Y(\sigma_X^{-1}(i)) \end{cases}$$

Continuous Optimal transport.

II : Monge problem between measure : Beyond point clouds?

1) Reminders on measures

$\boxed{\text{histogram}}$ probability vector:

$$a \in \Delta^n = \{a \in \mathbb{R}^n \mid a_i \geq 0, \sum_{i=1}^{n} a_i = 1\}$$

$\boxed{\text{Discrete | empirical measure}}$

- A discrete measure α with weights $a \in \mathbb{R}^n$, positions $(x_i)_{i=1}^n$, $x_i \in X$.

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \rightarrow \text{Dirac measure.}$$

Eulerian discretisation

- A discrete probability measure if $a \in \Delta^n$.

- An empirical measure

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i} \rightarrow \text{Lagrangian discretisation.}$$

General Measure (on Borelian)

$(X, d) \rightsquigarrow \mathcal{B}$ Borelian

α : Borele measure :

$\alpha(A)$ is finite and $A \in \mathcal{B}$.

- $\mathcal{M}_+(X)$ positive measure
- $\mathcal{M}_+^1(X)$ probability measure

$\boxed{\text{Random Measure}}$

Lebesgue integration, $\alpha \in \mathcal{M}(X)$

$f \in \mathcal{C}(X)$: $\langle f, \alpha \rangle = \int_X f(x)\, d\alpha(x)$.

In particular, if $\alpha = \sum_{i=1}^{n} a_i \delta_{x_i}$

then, $\langle f, \alpha \rangle = \sum_{i=1}^{n} a_i f(x_i)$

$$\begin{cases} f \longmapsto \int f\, d\alpha \\ \varphi(X), \|\ \|_\infty \longrightarrow \mathbb{R} \end{cases}$$ Linear continuous

$\left|\int f\, d\alpha\right| \leq \int |f|\, d\alpha$

Théo : Riez

If X is compact then any

$\ell : f \longmapsto \ell(f)$ linear continuous can be written as

$\ell(f) = \int f\, d\alpha, \quad \alpha \in \mathcal{M}(X)$

Théo : $(\mathcal{M}(X), \|\cdot\|_{TV})$ is a Banach space. $\alpha \in \mathcal{M}(X)$,

$\|\alpha\|_{TV} \triangleq \|\ell\|_{\varphi(X)}$

$= \sup_{f \in \varphi(X)} \{\langle f, \alpha\rangle : \|f\|_\infty \leq 1\}$

$= |\alpha|$

And $|\alpha|(A) = \sup_{A = \cup B_i} \sum_i |\alpha(B_i)|$

- Si $\alpha = \sum a_i \delta_{x_i}$, $|\alpha| = \sum |a_i| \delta_{x_i}$
- Si $d\alpha(x) = \rho_\alpha(x)\, dx$,

$d|\alpha|(x) = |\rho(x)|\, dx$

- Densité : si $\alpha \ll dx$,

$d\alpha(x) = \rho_\alpha(x)\, dx$ où

$\rho_\alpha = \frac{d\alpha}{dx}$ (Random-Nikodym)

- $\forall h \in \mathcal{C}(\mathbb{R}^d)$, $\int_{\mathbb{R}^d} h(x)\, d\alpha(x) = \int_{\mathbb{R}^d} h(x) \rho_\alpha(x)\, dx$.

- Proba : Random variable :

$X(\Omega, \mathcal{F}, \mathbb{P}) \longrightarrow (\mathcal{X}, d)$

$$\mathbb{P}(X \in A) = \alpha(A) = \int_A d\alpha(x)$$

$\boxed{\text{PUSH - FORWARD}}$

For any continuous $T : \mathcal{X} \longrightarrow \mathcal{Y}$, the push-forward $T_\# : \mathcal{M}(\mathcal{X}) \longrightarrow \mathcal{M}(\mathcal{Y})$

Ex. $x \in \mathcal{X}$: $T_\# \delta_x = \underbrace{\delta_{\overset{\in \mathcal{Y}}{T(x)}}}_{\in \mathcal{M}(\mathcal{Y})}$

- $T_\# \sum a_i \delta_{x_i} = \sum a_i \delta_{T(x_i)}$: linearization of T

General case :

for $T : \mathcal{X} \to \mathcal{Y}$ continuous, the push-forward of $\alpha \in \mathcal{M}(\mathcal{X})$

~~$\forall h \in \mathcal{P}(\mathcal{X})$~~

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

$\forall B \subset \mathcal{Y}$ measurable.

$\beta(B) = \alpha(\{x \in \mathcal{X} ; T(x) \in B\})$.

01.10.2024 « Suite du cours »

Def: for $T : X \to Y$, the push-forward of $\alpha \in \mathcal{M}_+(X)$ is defined as

$$\beta = T_\# \alpha \in \mathcal{M}_+(Y).$$

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

$\Leftrightarrow$ $\forall B \subset Y$ measurable,

$$\beta(B) = \alpha(\{x \in X : T(x) \in B\}) = \alpha(T^{-1}(B))$$

$\mathcal{C}(Y)$: ensemble de fonction continue sur Y.

Rq: $T_\# \delta_x = \delta_{T(x)}$

Cas si $\alpha = \sum a_i \delta_{x_i}$

$$T_\# \alpha = \sum a_i \delta_{T(x_i)}$$

!!! In particular $\alpha \in \mathcal{M}_+^1(X) \Rightarrow T_\# \alpha \in \mathcal{M}_+^1(Y)$

Densities :

$T : X \to Y$ diffeomorphism (= bijecti° C^1 et inverse C^1)

(A) $\begin{cases} d\alpha = \rho_\alpha \, dx \\ d\beta = \rho_\beta \, dx \end{cases}$ where ρ_α and ρ_β are densities.

Let $y = T(x) \sim dy = |\det T'(x)| dx$
where T' is the Jacobian matrix.

Let $h \in \mathcal{C}(Y)$.

$$\int_Y h(y) \, \rho_\beta(y) \, dy = \int_Y h(y) \, d\beta(y)$$

$$= \int_X h(T(x)) \, d\alpha(x) \quad \text{(By def)}$$

$$= \int_X h(T(x)) \, \rho_\alpha(x) \, dx \quad \text{(By (A))}$$

But $y = T(x)$. So

$$= \int h(y) \frac{dy}{|\det T'(T^{-1}(y))|} \rho_\alpha(T^{-1}(y))$$

$\Rightarrow \rho_\alpha(x) = |\det T'(x)| \, \rho_\beta(T(x))$

$\begin{cases} \text{Rq: if } f, g \in \mathcal{C}(X) \text{ and} \\ \forall h \in \mathcal{C}(X), \int h f = \int h g \\ \Rightarrow f = g \end{cases}$

3) Continuous Monge's Problem

Discret problem:

$$\min_{\sigma \in \mathcal{S}_n} \sum_{i=1}^{n} C_{i,\sigma(i)} \quad ; \quad C_{ij} \in \mathbb{R}^{n\times m}$$

Continuous problem

Cost function

$c : X \times X \longrightarrow \mathbb{R}$

$\alpha \in \mathcal{M}_+^1(X)$ source measure

$\beta \in \mathcal{M}_+^1(X)$ target measure

$$\inf_{T_\# \alpha = \beta} \int c(x, T(x))\, d\alpha(x)$$

Rq: Optimal matching

$$C_{ij} = c(x_i, y_j)$$

- if $\alpha = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$, $\beta = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$

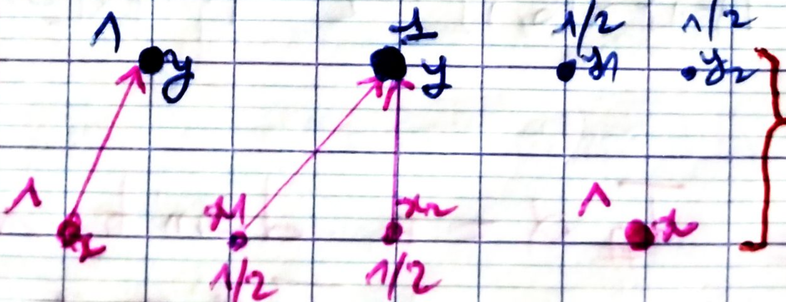
$$T_\# \alpha = \beta \Rightarrow ?$$

We have

$$T_\# \alpha = \beta \Rightarrow \frac{1}{n} \sum_{i=1}^{n} \delta_{T(x_i)} = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$$

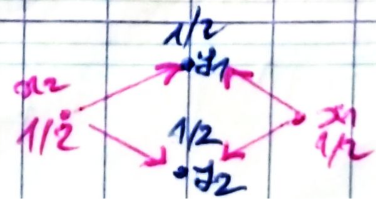
$$\Leftrightarrow \{T(x_i)\}_{i=1}^{n} = \{y_i\}_{i=1}^{n}$$

Ex: Dirac case.

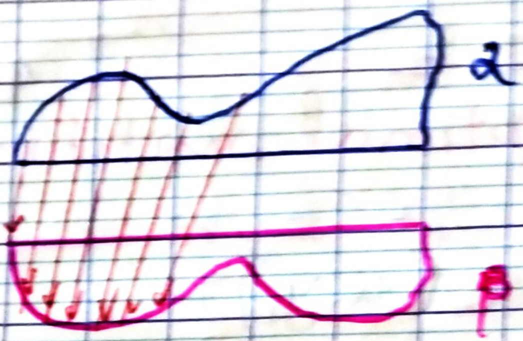


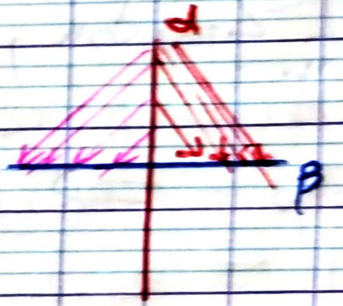
Not Monge's problem case

Many possibilities

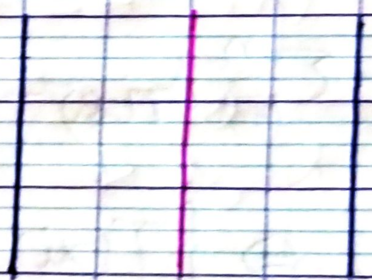


Continuous case





there are many possibilities.



such $T_\#$ doesn't exist!

$\{0\}\times[0,1]$

$$\frac{1}{2} dx_{|\{-1\}\times[0,1]} + \frac{1}{2} dx_{|\{1\}\times[0,1]}$$

Rq $T_\# \alpha = \beta$ doesn't means $T_\# \beta = \alpha$!

$\triangleq$ (by definition)

NB: $\inf \emptyset = +\infty$

$\|f\|_{L^p} = \left(\int |f(x)|^p \, d\alpha\right)^{1/p}$

• **Monge "Distance"**

Let $d : X \times X \to \mathbb{R}_+$ be a distance on X.

$c(x,y) = d(x,y)^p$, $p \geq 1$.

$$\tilde{W}_p(\alpha,\beta)^p = \inf_{T_\# \alpha = \beta} \mathcal{E}_\alpha(T)$$

$$\triangleq \begin{cases} \int_X d(x,T(x))^p \, d\alpha(x) & \text{if } \{T_\#\alpha = \beta\} \neq \emptyset \\ +\infty & \text{if } \{T_\#\alpha = \beta\} = \emptyset \end{cases}$$

Properties.

$\tilde{W}_p$ is a pseudo-distance on $\mathcal{M}(X)$.

Proof:

• $\tilde{W}_p(\alpha,\beta) = 0 \implies \int_X d(x,T(x))^p \, d\alpha(x) =$

$\Rightarrow \forall x \in Supp(\alpha), d(x, T(x)) = 0$

$\Rightarrow T_{|supp \alpha} = Id$

$\Rightarrow \alpha = \beta$. ~~because $\beta = T_\# \alpha$~~

~~$S_\# T_\# \alpha = \beta$~~

Rq: $supp\, \alpha$ = support of α.

• Triangular inequality.

Let $\alpha, \beta \in M(X)$.

* If $\tilde{W}_p(\alpha, \beta) = +\infty$

$\Rightarrow \tilde{W}_p(\alpha, \gamma) = +\infty$ or $\tilde{W}_p(\gamma, \beta) = +\infty$

Otherwise, $\exists$ S and T such that

$$\begin{cases} S_\# \alpha = \gamma \\ T_\# \gamma = \beta \end{cases} \Rightarrow (T \circ S)_\# \alpha = \beta$$

And then $\tilde{W}_p(\alpha, \beta) \leq E_\alpha(T \circ S) < +\infty$

* if $\tilde{w}_p(\alpha, \beta) < +\infty$ and $\begin{cases} \tilde{w}_p(\alpha,\gamma) < +\infty \\ \tilde{w}_p(\gamma,\beta) < +\infty \end{cases}$

otherwise trivial

Let $\varepsilon > 0$. Consider ε-minimizers.

$S_\# \alpha = \gamma$, $T_\# \gamma = \beta \Rightarrow (T \circ S)_\# \alpha = \beta$.

The we know that

$$\begin{cases} \mathcal{E}_\alpha(S)^{1/p} \le \tilde{w}_p(\alpha,\gamma) + \varepsilon \\ \mathcal{E}_\gamma(T)^{1/p} \le \tilde{w}_p(\gamma,\beta) + \varepsilon \end{cases},$$

On the other hand, we have

$$\tilde{w}_p(\alpha,\beta) \le \mathcal{E}_\alpha(T \circ S)^{1/p} = \left(\int d(x, T(S(x)))^p d\alpha(x) \right)^{1/p}$$

$$\le \left(\int (d(x, S(x)) + d(S(x), T(S(x))))^p d\alpha(x) \right)^{1/p}$$

Where we applied triangular inequality of d.

And by Minkowski inequality we obtained

$$\tilde{W}_p(\alpha,\beta) \leq \overbrace{\left[\int d(x,S(x))^p \, d\alpha(x)\right]^{1/p}}^{\mathcal{E}_\alpha(S)^{1/p}} + \underbrace{\left[\int d(S(x),T[S(x)])^p \, d\alpha(x)\right]^{1/p}}_{\mathcal{E}_\gamma(T)^{1/p}}$$

But, since

$$\begin{cases} \mathcal{E}_\alpha(S)^{1/p} = \left(\int d(x,S(x))^p \, d\alpha(x)\right)^{1/p} \\ \mathcal{E}_\gamma(T)^{1/p} = \left(\int d(y,T(y))^p \, d\beta(y)\right)^{1/p} \end{cases}$$

we obtained

$$\tilde{w}_p(\alpha,\beta) \leq \tilde{w}_p(\alpha,\gamma) + \tilde{W}_p(\gamma,\beta) + 2\varepsilon$$

So we conclude that $\tilde{w}_p$ is a pseudo-distance.

4) Existence & uniqueness of Monge maps

* Theorem (Brenier, 1991)

Let $X = \mathbb{R}^d$ endowed with Euclidean structure, $p = 2$, $c(x,y) = \|x-y\|^2$.

$$\alpha, \beta \in \mathcal{M}(\mathbb{R}^d), \begin{cases} \int x^2 d\alpha(x) < +\infty \\ \int x^2 d\beta(x) < +\infty \end{cases}$$

And $\alpha \ll dx \Leftrightarrow d\alpha = \rho_\alpha(x)\, dx$.

Then, there exist T^* such that

$$T^* \in \underset{T_\# \alpha = \beta}{\operatorname{argmin}} \int c(x, T(x))\, d\alpha(x)$$

Moreover

- ① there existe a convex function $\varphi : \mathbb{R}^d \to \mathbb{R}$: $\nabla \varphi = T^*$

• ② if $T \in \underset{T_\# \alpha = \beta}{\operatorname{argmin}} \int c(x, T(x))\, d\alpha(x)$

(uniqueness)

Then $T = T^*$ α-almost everywhere

Rq

① φ Convex $\Rightarrow \nabla\varphi$ monotone

i.e $\langle \nabla\varphi(x) - \nabla\varphi(y), x-y \rangle \geq 0$

" generalizati° of increasing map for optimal matching."

⚠ T monotone $\not\Rightarrow T = \nabla\varphi$

Ex: $T_\theta : \mathbb{R}^2 \to \mathbb{R}$, $|\theta| \leq \frac{\pi}{2}$.

– Monotone:

$\langle T_\theta x - T_\theta y, x-y \rangle$

$= \langle T_\theta(x-y), x-y \rangle$

$= \langle T_\theta v, v \rangle$

$T : x \longmapsto Ax$ be the gradient of A symmetric.

$$= \langle \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} v, v \rangle$$

$$= \langle \begin{pmatrix} v_1\cos\theta - v_2\sin\theta \\ v_1\sin\theta + v_2\cos\theta \end{pmatrix} ; \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \rangle$$

$$= v_1^2\cos\theta - v_1 v_2 \sin\theta + v_1 v_2 \sin\theta + v_2^2 \cos\theta$$
$$= (v_1^2 + v_2^2)\cos\theta \geq 0 .$$

But not a transport map.

Because $R_\theta^T \neq R_\theta$.

02.10.24

5/ OT in 1D : $\alpha, \beta \in \mathcal{M}_+^1(\mathbb{R})$

* Cumulative distribution function (C

$C_\alpha : \mathbb{R} \to [0,1]$; $C_\alpha(x) = \int_{-\infty}^{x} d\alpha(x)$

* Quantile function

$C_\alpha^{-1} : [0,1] \to \mathbb{R}$; $C_\alpha^{-1}(r) = \inf_x \{ x : C_\alpha(x) \geq r \}$

Proposition.

Quantile = push-forward of an uniforme.

For $\alpha \in \mathcal{M}_+^1(\mathbb{R})$, $(C_\alpha^{-1})_\# \mathcal{U} = \alpha$.

If $\alpha \ll dx$, $(C_\alpha)_\# \alpha = \mathcal{U}$
where $\mathcal{U}_{[0,1]}$ is an uniforme on $[0,1]$.

Proof: case $\begin{cases} \alpha \ll dx \Rightarrow d\alpha(x) = \rho_\alpha(x)\, dx \\ \rho_\alpha(x) > 0, \ \forall x. \end{cases}$

- Define $\gamma = (C_\alpha^{-1})_\# \mathcal{U}$

$$C_\gamma(x) = \int_{-\infty}^{x} d\gamma(z)$$
$$= \int_{\mathbb{R}} 1_{]-\infty;x]}(z)\, d\left((C_\alpha^{-1})_\# \mathcal{U}\right)(z)$$
$$= \int_{\mathbb{R}} 1_{]-\infty;x]}(C_\alpha^{-1}(z))\, d\mathcal{U}(z)$$
$$= \int_0^1 1_{]-\infty;x]}(C_\alpha^{-1}(z))\, dz$$

Rq: $F_\alpha^{-1}(z) \in]-\infty; x]$

$\Leftrightarrow 0 < F_\alpha^{-1}(z) \leq x$

When we apply F_α at each element and we obtain

$$0 < z \leq F_\alpha(x).$$

So, $F_\gamma(x) = \int_0^1 \mathbb{1}_{[0, F_\alpha(x)]}(u)\, du$

$$F_\gamma(x) = F_\alpha(x).$$

If $\alpha \ll dx$, $T = F_\beta^{-1} \circ F_\alpha$ & $\beta = T_\# \alpha$

Rq: T is the optimal transport.
Because $T = F_\beta^{-1} \circ F_\alpha$ is increasing $\Rightarrow \nabla \varphi = T$
Where φ is convex function.
And we apply Brenier theorem
$\Rightarrow T$ is optimal and uniq.

First approach to Wasserstein distance

$$W_p(\alpha, \beta)^p = \int_{\mathbb{R}} |T(x) - x|^p \, d\alpha(x)$$

where T is solution of transport problem.

$$W_p(\alpha,\beta)^p = \int_{\mathbb{R}} \left| (\mathcal{C}_\beta^{-1} \circ \mathcal{C}_\alpha)(x) - x \right|^p d\alpha(x)$$

Let $x = \mathcal{C}_\alpha^{-1}(r)$. so,

$$W_p(\alpha,\beta)^p = \int_0^1 \left| \mathcal{C}_\beta^{-1}(r) - \mathcal{C}_\alpha^{-1}(r) \right|^p dr$$

$$= \left\| \mathcal{C}_\beta^{-1} - \mathcal{C}_\alpha^{-1} \right\|^p_{L^p[0,1]}$$

The optimal transport interpolation $\alpha_0 \leftrightarrow \alpha_1$

$$\forall t \in [0,1],\ \mathcal{C}_{\alpha_t}^{-1} = (1-t)\, \mathcal{C}_{\alpha_0}^{-1} + t\, \mathcal{C}_{\alpha_1}^{-1}$$

6| OT for Gaussians

① 1D case : $\alpha = \mathcal{N}(m_\alpha, \sigma_\alpha^2)$
$\beta = \mathcal{N}(m_\beta, \sigma_\beta^2)$

Density: $p_\alpha(x) = \dfrac{1}{\sigma_\alpha \sqrt{2\pi}} e^{-\frac{(x-m_\alpha)^2}{2\sigma_\alpha^2}}$

let $T : x \longmapsto \dfrac{\sigma_\beta}{\sigma_\alpha}(x - m_\alpha) + m_\beta$

Satisfies $T_\# \alpha = \beta$?

Let $h \in \mathcal{C}$

1) $X \sim \alpha \Rightarrow T(X) \sim T_\# \alpha$

$\Rightarrow \int h(T(x))\, d\alpha(x) = \int h(y)\, d\beta(y).$

let $X \sim \mathcal{N}(m_\alpha, \sigma_\alpha^2)$.

Then, $\dfrac{\sigma_\beta}{\sigma_\alpha}(X - m_\alpha) + m_\beta$

$= \sigma_\beta \left(\dfrac{X - m_\alpha}{\sigma_\alpha} \right) + m_\beta$.

$Z = \dfrac{X - m_\alpha}{\sigma_\alpha} \sim \mathcal{N}(0, 1)$

So $\sigma_\beta Z + m_\beta \sim \mathcal{N}(m_\beta, \sigma_\beta^2)$

2)

$$\int_{\mathbb{R}} h(T(x))\, d\alpha(x)$$

$$= \int_{\mathbb{R}} h\left(\frac{\sigma_\beta}{\sigma_\alpha}(x - m_\alpha) + m_\beta\right) \frac{1}{\sigma_\alpha\sqrt{2\pi}}\, e^{-\frac{(x-m_\alpha)^2}{2\sigma_\alpha^2}}\, dx$$

Let $u = T(x) \Rightarrow du = \frac{\sigma_\beta}{\sigma_\alpha}\, dx$

So $\int_{\mathbb{R}} h(T(x))\, d\alpha(x)$

$$= \int h(u) \frac{1}{\sigma_\alpha\sqrt{2\pi}}\, \frac{\sigma_\alpha}{\sigma_\beta}\, e^{-\frac{(T^{-1}(u) - m_\alpha)^2}{2\sigma_\alpha^2}}\, du$$

~~But~~ $\cancel{T^{-1}(u) = x \Leftrightarrow x = \frac{\sigma_\beta}{\sigma_\alpha}(u - m_\alpha) + m_\beta}$

But $T^{-1}(u) = \frac{\sigma_\alpha}{\sigma_\beta}(u - m_\beta) + m_\alpha$

So $\int_{\mathbb{R}} h(T(x))\, d\alpha(x)$

$$= \int_{\mathbb{R}} h(u) \frac{1}{\sigma_\beta\sqrt{2\pi}}\, e^{-\frac{(u-m_\beta)^2}{2\sigma_\beta^2}}\, du$$

$$= \int_{\mathbb{R}} h(u)\, \rho_\beta(u)\, dx.$$

Since $\frac{\sigma_\beta}{\sigma_\alpha} > 0$, T is increasing

$\varepsilon \, \Delta \varphi(x)$
laplacian de $\varphi(x)$

And then $T = \nabla \varphi$ with φ a convex funct°.
By Brenier theo T is the uniq OT.

$\star$ $$W_2^2(\alpha,\beta) = \int_{\mathbb{R}} |x - T(x)|^2 \, d\alpha(x)$$

$$= \int_{\mathbb{R}} |T(x) - x|^2 \, d\alpha(x)$$

$$= \int_{\mathbb{R}} \left| \frac{\sigma_\beta}{\sigma_\alpha}(x - m_\alpha) + m_\beta - x \right|^2 d\alpha(x)$$

$$= \int_{\mathbb{R}} \left(\left(\frac{\sigma_\beta}{\sigma_\alpha} - 1\right) x - \left(\frac{\sigma_\beta}{\sigma_\alpha} m_\alpha - m_\beta\right) \right)^2 d\alpha(x)$$

$$= \left(\frac{\sigma_\beta}{\sigma_\alpha} - 1\right)^2 \int_{\mathbb{R}} x^2 \, d\alpha(x) - 2\left(\frac{\sigma_\beta}{\sigma_\alpha} - 1\right)\left(\frac{\sigma_\beta}{\sigma_\alpha} m_\alpha - m_\beta\right) \int_{\mathbb{R}} x \, d\alpha(x)$$
$$+ \left(\frac{\sigma_\beta}{\sigma_\alpha} m_\alpha - m_\beta\right)^2 \int_{\mathbb{R}} d\alpha(x)$$

Since $$\begin{cases} \int_{\mathbb{R}} x^2 \, d\alpha(x) = V(X) + E(X)^2 = \sigma_\alpha^2 + m_\alpha^2 \\ \int_{\mathbb{R}} x \, d\alpha(x) = E(X) = m_\alpha \\ \int_{\mathbb{R}} d\alpha(x) = 1 \end{cases}$$

$W_2^2(\alpha,\beta) = (\sigma_\alpha - \sigma_\beta)^2 + (m_\alpha - m_\beta)^2$, after simplification.

Rq: In $\mathbb{R}^d$: $\left\| \begin{pmatrix} m_\alpha \\ \sigma_\alpha \end{pmatrix} - \begin{pmatrix} m_\beta \\ \sigma_\beta \end{pmatrix} \right\|_2^2 = W_2^2(\alpha,\beta)$.

09.10.2024 Cas $\mathbb{R}^d$ $\begin{cases} \alpha \sim \mathcal{N}(m_\alpha, \Sigma_\alpha) \\ \beta \sim \mathcal{N}(m_\beta, \Sigma_\beta) \end{cases}$

$T : x \longmapsto m_\beta + A(x - m_\alpha)$

Condition sur A tq $T_\# \alpha = \beta$?

- $T = \nabla \varphi$, φ Convex

$\Leftrightarrow$

$\varphi(x) = \langle m_\beta, x \rangle + \frac{1}{2} \langle A(x - m_\alpha), x - m_\alpha$

$A \in S_n^+$ (= symetric semi-definit-posit

Proposition

$$T_\# \alpha = \beta \Leftrightarrow A \Sigma_\alpha A = \Sigma_\beta$$

(1) $X \sim \mathcal{N}(m_\alpha, \Sigma_\alpha) \Rightarrow T(X) \sim \mathcal{N}(\overline{m}, \overline{\Sigma}$

NB: Transformation affine d'un gaussien est un gaussien.

Soit $Y \sim \mathcal{N}(m_\beta, \Sigma_\beta)$.

Cq $\mathbb{E}(Y) = \mathbb{E}(T(X))$ et $\mathrm{cov}(Y) = \mathrm{cov}(T(X))$

- $\mathbb{E}[T(X)] = \mathbb{E}(m_\beta + A(X - m_\alpha))$

$$= m_\beta + A(\underbrace{\mathbb{E}(X) - m_\alpha}_{=0}) = m_\beta$$

(linéarité de $\mathbb{E}$).

Par déf de cov,

$$\mathrm{Cov}(T(X)) = \mathbb{E}((T(X) - m_\beta)(T(X) - m_\beta)^T)$$

$$= \mathbb{E}[(A(X - m_\alpha))(A(X - m_\alpha))^T]$$

$$= \mathbb{E}[A(X - m_\alpha)(X - m_\alpha)^T A^T]$$

$$= A\,\mathbb{E}[(X - m_\alpha)(X - m_\alpha)^T]A^T$$

$$\mathrm{Cov}(T(X)) = A\,\Sigma_\alpha A$$

- let $\alpha \sim \mathcal{N}(m_\alpha, \Sigma_\alpha)$

$\beta \sim \mathcal{N}(m_\beta, \Sigma_\beta)$ such that

$\Sigma_\alpha, \Sigma_\beta \in S_n^{++}$.

Rappel : $\forall A \in S_n^{++} \Leftrightarrow \exists !\, M \in S_n^{++}$
sq : $A = M^2$
We note $M = A^{1/2}$.

$$\Sigma_\beta = A \Sigma_\alpha A$$

$$\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2} = \Sigma_\alpha^{1/2} A \Sigma_\alpha A \Sigma_\alpha^{1/2}$$

$$= \Sigma_\alpha^{1/2} A \left(\Sigma_\alpha^{1/2}\right)^2 A \Sigma_\alpha^{1/2}$$

$$\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2} = \left(\Sigma_\alpha^{1/2} A \Sigma_\alpha^{1/2}\right)^2$$

$$\Rightarrow A = \Sigma_\alpha^{-1/2} \left(\Sigma_\alpha^{1/2} \Sigma_\beta \Sigma_\alpha^{1/2}\right)^{1/2} \Sigma_\alpha^{-1/2}$$

D'après Brenier T est l'uniq OT

One can show that

$$\mathcal{W}(\alpha, \beta) = \| m_\alpha - m_\beta \|^2 + \mathcal{B}^2(\Sigma_\alpha, \Sigma_\beta)$$

where $\mathcal{B}^2(\Sigma_\alpha, \Sigma_\beta)$

$$= \text{trace}\left(\Sigma_\alpha + \Sigma_\beta - 2\left(\Sigma_\alpha^{1/2}\Sigma_\beta\Sigma_\alpha^{1/2}\right)\right)$$

III | Kantorovich relaxation

1) Discrete relaxation.

NB : For Monge Matching, there is not uniqueness and not ~~exist~~ necessary existence with $n \neq m$

- In the continuous case, the same problem appear if we are in a non-convex case.

Deterministic T Monge.

Replace $\begin{cases} \sigma \in \mathfrak{S}_n \\ T : X \to Y \end{cases}$ with coupling P

$P \in \mathbb{R}^{n \times m}$ départ, arrivée

P_{ij} = masse transportée de $i \to j$

Constraint

$\alpha = \sum a_i \delta_{x_i}$

$\beta = \sum b_j \delta_{y_j}$

Admissible coupling

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

$P^T \mathbb{1}_n = b$

where $\mathbb{1}_n = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n . \}$

Rq :

(1) $P\mathbb{1}_n = a \Leftrightarrow \sum_j P_{ij} = a_i \quad \forall i$

$P^T\mathbb{1}_m = b \Leftrightarrow \sum_i P_{ij} = b_j \quad \forall j$

(2) $P \in U(a,b) \Leftrightarrow P^T \in U(b,a)$

(3) $U(a,b) \subsetneq \mathbb{R}_+^{n\times m}$ <u>borné</u>

$n+m$ linéar inequality $\Rightarrow$ Convex polyhedre $\Rightarrow \exists$

$P_1, \dots P_p \in \mathbb{R}_+^{n\times m}$ tq $U(a,b) = \mathrm{conv}(P)$

• Si S est un ensemble convexe, $\mathrm{conv}(S)$ est le plus petit ensemble convexe contenant S

$$\mathrm{conv}(S) = \left\{ x = \sum_{i=1}^{n} \lambda_i x_i,\ n \in \mathbb{N},\ \lambda_i \geq 0,\ \sum \lambda_i = 1,\ x_i \in S \right\}$$

NB: La formulation de KP est symétrique et est un Problème Convexe.

Kantorovich formulation of OT
(K.P)

$$\min_{P \in U(a,b)} \langle C, P \rangle \quad : C \in \mathbb{R}^{n \times m}$$

$$\langle C, P \rangle = \sum_{i,j} C_{ij} P_{ij}$$

linear programming

⚠ Pas forcement unicité, mais

$\exists p^*$ solution de (K.P)

avec $\#\{(i,j) : P_{ij} \neq 0\} \leq n+m-1$.

1D : case

$C_{ij} = c(x_i, y_j) = |x_i - y_j|^p$,

$n = m$

$a_i = b_i = \frac{1}{n}$. $\rightsquigarrow$ Optimal Matching

$n = m$

$a = b = \mathbb{1}_n$,

$\min_{\sigma \in \Sigma} \sum_{i=1}^{n} a_{i,\sigma(i)}$

$\min_{P \in U(a,b)} \langle C, P \rangle$

$U(1,1) = \{ P \in \mathbb{R}_+^{n \times m} : P\mathbb{1}_m = \mathbb{1}_n,\ P^T \mathbb{1}_n = \mathbb{1}_m \}$

Matrice de permutati:

$$P_n = \{ P_\sigma \in \{0,1\}^{n \times m} : \sigma \in S_n \}$$

$$\# P_n = \# S_n = n!$$

Rq: $\langle C, P_\sigma \rangle = \sum_{ij} C_{ij} (P_\sigma)_{ij}$

$$= \sum_{\substack{ij \\ j=\sigma(i)}} C_{ij} (P_\sigma)_{ij} + \underbrace{\sum_{\substack{ij \\ j \neq \sigma(i)}} C_{ij} (P_\sigma)_{ij}}_{=0}$$

$$\langle C, P_\sigma \rangle = \sum_{i=1}^{n} C_{i,\sigma(i)}$$

~~[illegible]~~ $\cancel{\min_{\sigma \in S_n} \sum C_{i,\sigma(i)} = \min_{P \in P_n(a,b)}}$

$$\min_{\sigma \in S_n} \sum C_{i,\sigma(i)} = \min_{P \in P_n} \langle C, P \rangle$$

Rq : Bistochastic matrix.

Ensemble de Matrices bistochastiques.

$B_n = U(1,1) = \left\{ P \in \mathbb{R}_+^{n \times m} : \begin{array}{l} P 1\!\!1_n = 1\!\!1 \\ P 1\!\!1_m = 1\!\!1 \end{array} \right\}$

$\Rightarrow P_n = B_n \cap \{0,1\}^{n \times m}$

Rappel :

$\min_{x \in C} f(x) \geq \min_{x \in D} f(x)$

$C \subseteq D$

Donc :

$\underbrace{\min_{P \in P_n} \langle C, P \rangle}_{\text{Monge}} \geq \underbrace{\min_{P \in B_n} \langle C, P \rangle}_{\text{Kantorovich}}$

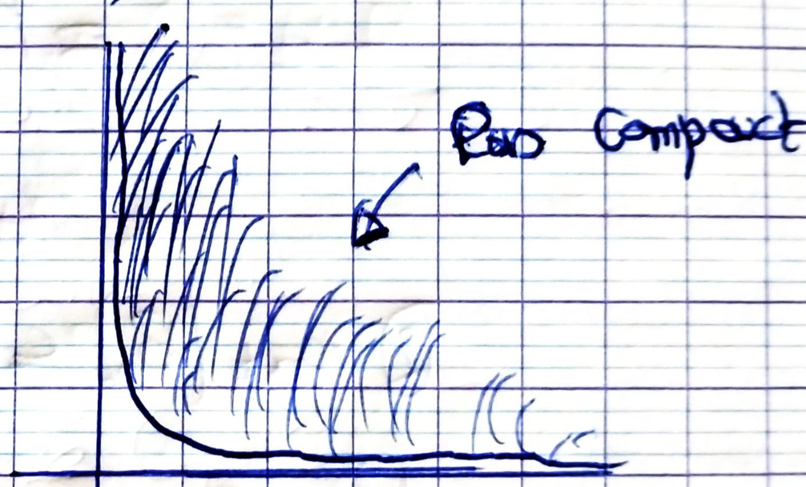
Detour pour les geo convex

Def: External points

$$Ext(\mathcal{C}) = \{ P \mid \forall Q, R \in \mathcal{C}^2, P = \frac{Q+R}{2} \Rightarrow Q = R \}$$

Proposition:

If $\mathcal{C}$ compact, $Ext(\mathcal{C}) \neq \emptyset$



$\{(x,y) \in \mathbb{R}_+^2 : xy \geq 1\}$

Prop) If $\mathcal{C}$ compact,

$$Ext(\mathcal{C}) \cap \underset{p \in \mathcal{C}}{\operatorname{argmin}} \langle C, p \rangle \neq \emptyset$$

Theo: Birkhoff - von Neuman

$$Ext(B_n) = P_n$$

Proof :

(1) $P_n \subseteq Ext(B_n)$?

$Ext([0,1]) = \{0,1\}$

let $P \in P_n$ if $P = \frac{Q+R}{2}$, $Q_{ij} \in [0,1]$, $R_{ij} \in [0,1]$

$\Rightarrow P_{ij} \in [0,1] \Rightarrow \frac{Q_{ij}+R_{ij}}{2} \in \{0,1\}$

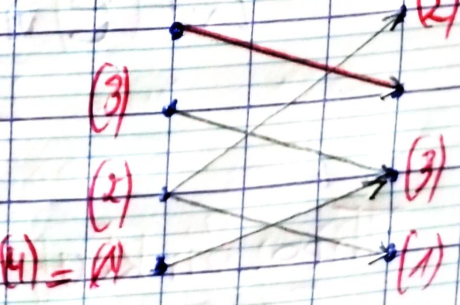
$\Leftrightarrow P = Q = R$

(2) $Ext(B_n) \subseteq P_n \Leftrightarrow P_n^{c(B_n)} \subseteq Ext(B_n)^c$

let $P_n \in B_n \setminus P_n$ (Goal, find $Q, R \in B_n$: $P = \frac{Q+R}{2}$)

Constructi proof:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 2/3 & 0 & 0 & 1/3 \\ 0 & 0 & 1/3 & 2/3 \end{pmatrix}$$



$P_{ij} = 1 \Leftrightarrow$ Only one edge from left to right.

$P_{ij} \in [0,1] \Leftrightarrow \geq 2$ edge from left to right.

Consider the set of "non solated edges.

We can extract a cycle of this form.

Along this cycle, $\forall s \in [\![1, p]\!]$,

$$\begin{cases} 0 < P_{i_s, j_s} < 1 & \text{left} \to \text{right} \\ 0 < P_{i_{s+1}, j_s} < 1 & \text{left} \to \text{right.} \end{cases}$$

Shortest $\Rightarrow \forall s, s' \in [\![1, p]\!]$, $i_s \neq i_{s'}$

$j_s \neq j_{s'}$

let $\varepsilon \overset{\text{def}}{=} \min\limits_{0 \leq s \leq p} \{ P_{i_s, j_s} ; P_{i_{s+1}, j_s} ;$

$1 - P_{i_s, j_s} , 1 - P_{i_{s+1}, j_s} \}$

$\in]0, 1/2]$

split

$$\mathcal{L} = \{(i_s, j_s)\}_{s=1}^{P}$$

$$R = \{j_s, i_{s+1}\}_{i=1}^{P}$$

$$Q_{ij} = \begin{cases} P_{ij} & \text{si } (i,j) \notin \mathcal{L} \cup R \\ P_{ij} + \varepsilon/2, & (i,j) \in \mathcal{L} \\ P_{ij} - \varepsilon/2, & (i,j) \in R \end{cases}$$

$$R_{ij} = \begin{cases} P_{ij} & \text{si } (i,j) \notin \mathcal{L} \cup R \\ P_{ij} - \varepsilon/2 & (i,j) \in \mathcal{L} \\ P_{i,j} + \varepsilon/2 & (i,j) \in R \end{cases}$$

Rq : (1) $\forall ij$,

$0 \leq Q_{ij}, R_{ij} \leq 1$, $Q, R \in \mathcal{B}_n$

$$P = \frac{Q+R}{2}$$

a/ Relaxation for arbitrary measures

Trouver un analogue continu à

$$\min_{P \in \mathcal{B}_n} \langle C, P \rangle .$$

<u>Rk</u>: $\mathcal{B}_n = \{ P \in \mathbb{R}_+^{n \times n} \mid P \mathbb{1}_n = P^T \mathbb{1}_n = \mathbb{1} \}$

Soit $\alpha \in \mathcal{M}_1^+(X)$, $\beta \in \mathcal{M}_1^+(Y)$

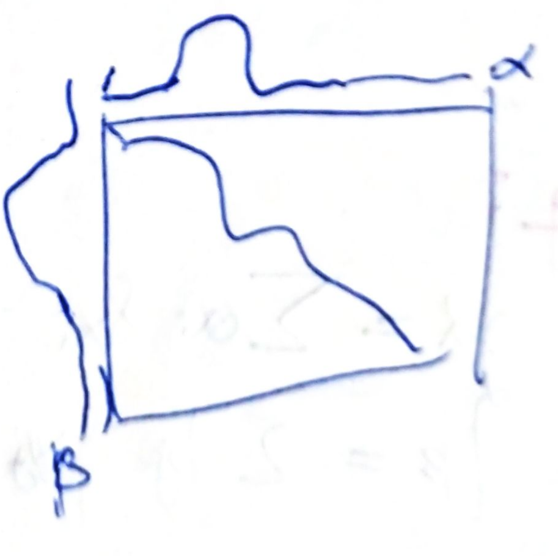
Coupling $\pi \in \mathcal{M}_1^+(X \times Y)$

$$\pi_1 = (P_1)_\# \pi = \alpha$$

$$\pi_2 = (P_2)_\# \pi = \beta$$

$P_1(x, y) = x$

$P_2(x, y) = y$



" Intuitive interpretation

$\sum_i P_{ij} = 1 \rightsquigarrow$ " $\int_Y d\pi(x,y) = d\alpha(x)$ "

$\sum_j P_{ij} = 1 \rightsquigarrow$ " $\int_X d\pi(x,y) = d\beta(y)$ "

Formally

$\forall f \in \mathcal{C}(X), \int_{X \times Y} f(x)\, d\pi(x,y) = \int_X f\, d\alpha$

$\forall g \in \mathcal{C}(Y), \int_{X \times Y} g(y)\, d\pi(x,y) = \int_Y g\, d\beta.$

$P \in \mathcal{B}_n \rightsquigarrow \cancel{\pi \in \mathcal{U}(\alpha,\beta)}$

$\mathcal{U}(\alpha,\beta) \overset{\text{def}}{=} \left\{ \pi \in \mathcal{M}_1^+(X \times Y), \begin{array}{l} \pi_1 = \alpha \\ \pi_2 = \beta \end{array} \right\}$

Rq :

1) $\begin{cases} \alpha = \sum a_i \delta_{x_i} \\ \beta = \sum b_j \delta_{y_j} \end{cases}$

$\pi \mid \begin{array}{l} \pi_1 = \alpha \\ \pi_2 = \beta \end{array} \Rightarrow \pi$ support discret $\{(x_i, y_j)\}_{\substack{i=1 \dots n \\ j=1 \dots m}}$

$\Rightarrow \pi = \sum_{ij} P_{ij}\, \delta_{x_i, y_j}$

$\boxed{\mathcal{U}(\alpha, \beta) \neq \phi}$

car $\alpha \otimes \beta \in \mathcal{M}_1^+(X, Y)$

« $d(\alpha\otimes\beta)(x, y) = d\alpha(x)\, d\beta(y)$ »

formally,

$\forall h \in \varphi(X \times Y)$,

$$\int_{X\times Y} h(x,y)\, d(\alpha\otimes\beta)(x,y)$$

$$= \int_X \left(\int_Y h(x,y)\, d\beta(y) \right) d\alpha(x)$$

$$= \int_Y \left(\int_X h(x,y)\, d\alpha(x) \right) d\beta(y)$$

$\begin{cases} (\alpha\otimes\beta)_1 = \alpha \\ (\alpha\otimes\beta)_2 = \beta \end{cases}$, $\forall f \in \varphi(X)$,

$\int_X f(x)\, d(\alpha\otimes\beta)_1(x)$

$= \int_{X\times Y} f(x)\, d\alpha(x)\, d\beta(y) = \int_X f(x)\, d\alpha(x) \underbrace{\int_Y f(y)\, d}_{=1}$

(3) supposons qu'il existe

$T: X \longrightarrow Y$, $T_{\#}\alpha = \beta$.

$\Rightarrow \overline{\pi} = (Id, T)_{\#}\alpha \in \Pi(\alpha, \beta)$.

$\forall h \in \varphi(X \times Y)$,

$$\int_{X \times Y} h(x, y) \, d\pi(x, y)$$

$$= \int_X h(F(x)) \, d\alpha(x)$$

$$= \int_X h(x, T(x)) \, d\alpha(x)$$

with $X \longrightarrow X \times Y$

$x \mapsto (x, T(x))$

$$F(x) = (x, T(x))$$

27.10.2024 OT : course ②

Continuouse Kantorovich problem

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

"Infinite dimensional linear program on the space of measure"

If X, Y compact, inf = min

then everything work well.

$$\boxed{\text{Rappel : } X \sim \alpha \Rightarrow T(X) \sim T_\# \alpha}$$

Probabilistic interpretation

$X \sim \alpha$, $Y \sim \beta$ $\Leftarrow$ $(X,Y) \sim \pi$ statistical dependency

if $\pi = \alpha \otimes \beta \Rightarrow X \perp\!\!\!\perp Y$.

$(\mathbb{R}^d, \|\cdot\|_2)$, Brenier théo $\sim T$.

$\pi = (Id, T)_{\#}\alpha \longrightarrow Y = T(X)$

Corollary

Si $\alpha \ll dx$, $c(x,y) = \|x-y\|^2$,

$\int x^2 d\alpha, \int y^2 d\beta < +\infty$.

T unique application de transport,

Alors, $\pi = (Id, T)_{\#}\alpha$ unique

couplage optimal

$$\inf_{T_{\#}\alpha = \beta} \mathcal{E}_\alpha(T) = \mathcal{L}_c(\alpha, \beta)$$

③ Metric properties of OT

Discrete case

Lemma (Gluing lemma)

Given $(a,b,c) \in \Sigma_n \times \Sigma_p \times \Sigma_m$,
$P \in U(a,b)$, $Q \in U(a,b)$.
Then it exists $S \in \mathbb{R}_+^{n\times p\times m}$ such that 2D marginals

$$\sum_k S_{i,j,k} = P_{i,j} \quad , \quad \sum_i S_{i,j,k} = Q_{j,k}$$

[In particular 1D marginals are (a,b,c) $\sum_{j,k} S_{i,j,k} = a_i$)

Proof

$$S_{i,j,k} = \begin{cases} \dfrac{P_{ij}\, Q_{jk}}{b_j} & \text{si } b_j \neq 0 \\ 0 & \text{sinon} \end{cases}$$

Rappel :

$$\begin{cases} L_C(a,b) = \min_{P \in U(a,b)} \langle P, C \rangle, \quad C \in \mathbb{R}^{n \times m} \\ U(a,b) = \left\{ P \in \mathbb{R}_+^{n \times m} \;\middle|\; \begin{array}{l} P\mathbb{1} = a \\ P^T\mathbb{1} = b \end{array} \right\} \end{cases}$$

Proposition

Supposons $p \geq 1$, $C = D^p = (D_{ij}^p)_{\substack{1 \leq i \leq n \\ 1 \leq j \leq n}} \in \mathbb{R}_+^{n \times n}$

$D \in \mathbb{R}_+^{n \times m}$ et D est une distance sur $\{1, \dots, n\}$

- (1) D est symétrique
- (2) $D_{i,j} = 0 \iff i = j$
- (3) $\forall (i,j,k) \in \{1, \dots, n\}^3$, $D_{i,k} \leq D_{i,j} + D_{j,k}$.

Alors, $W_p(a,b) = L_{D^p}(a,b)^{1/p}$ (distance de Wasserstein)

29.10.24 OT: cours ③

W_p est une distance sur Σ_r :

① $W_p(a,b) = W_p(b,a) \quad \forall a,b \in \Sigma_r$

② $W_p(a,b) = 0 \iff a = b$

③ $W_p(a,c) \leq W_p(a,b) + W_p(b,c),$

$\forall a,b,c \in \Sigma_r.$

Preuve

① symetric

$\left.\begin{array}{l} * \ C \text{ symetriq} \\ * \ P \in U(a,b) \iff P^T \in U(b,a) \end{array}\right\} \Rightarrow L_C(a,b) = L_C(b,a)$

② Defini

$* \ W_p(a,b) = 0 \iff L_C(a,b) = 0 \iff \left\{\begin{array}{l} \langle C,P\rangle = 0 \\ P \in U(a,b) \end{array}\right.$

$P^* = \mathrm{diag}(a) = \mathrm{diag}(b)$

$\Updownarrow$

$a = b$

$\sum_{ij} C_{ij} P_{ij} = 0$; $C_{ij} \geq 0$, $P_{ij} \geq 0$

$\Downarrow$

$\forall i,j$, $C_{ij} = 0$ ou $P_{ij} = 0$

$\Downarrow$

$i = j$

$\Downarrow$

$\forall i \neq j$, $P_{ij} = 0 \Rightarrow P = \mathrm{diag}(a) = \mathrm{diag}(b)$

③ Triangular inequality

Soit $a, b, c \in \Sigma_n$

$P \in \underset{M \in U(a,b)}{\operatorname{argmin}} \langle C, M \rangle$,

$Q \in \underset{M \in U(b,c)}{\operatorname{argmin}} \langle C, M \rangle$

Gluing Lemma

$\exists S \in \mathbb{R}^{n \times n \times n}$ tq $\begin{cases} \sum_k S_{\cdot,\cdot,k} = P \\ \sum_i S_{i,\cdot,\cdot} = Q \end{cases}$

$$\begin{tikzcd} a \arrow[r, "P"] \arrow[rr, bend right, "R"'] & b \arrow[r, "Q"] & c \end{tikzcd}$$

Let: $R = \sum_j S_{\cdot,j,\cdot}$. Then, $R \in U(a,c)$

(if $b > 0$, $R = \mathrm{diag}(1/b)\, Q$

$\forall b_i > 0$).

$$W_p(a,c) = \left(\min_{\tilde{R} \in U(a,c)} \langle \tilde{R}, C \rangle\right)^{1/p} \text{(def)}$$

$$\leq \underbrace{\langle R, C \rangle}^{1/p} \text{(sous optimal)}$$

$$= \left(\sum_{i,k} C_{ik} R_{ik}\right)^{1/p} \text{(def prod scalaire)}$$

$$= \left(\sum_{i,k} D_{ik}^p \sum_j S_{ij,k}\right)^{1/p} \text{(def } C_{ij} = D_{ij}^p \text{ et } S\text{)}$$

$$\leq \left(\sum_{i,k} (D_{ij} + D_{jk})^p S_{ijk}\right)^{1/p}$$

{car D distance sur $\{1, \ldots, n\}$}

$$\leq \underbrace{\left(\sum_{ijk} D_{ij}^p S_{ijk}\right)^{1/p} + \left(\sum_{ijk} D_{jk}^p S_{ijk}\right)^{1/p}}_{\text{(Minkowski)}}$$

$$= \left(\sum_{ij} D_{ij}^p \sum_k s_{ijk}\right)^{1/p} + \left(\sum_{jk} D_{jk}^p \sum_i S_{ijk}\right)^{1/p}$$

(séparation des ...)

$$= \left(\sum_{ij} D_{ij}^p P_{ij}\right)^{1/p} + \left(\sum_{jk} D_{jk}^p Q_{jk}\right)^{1/p} \quad \text{(def)}$$

$$= W_p(a,b) + W_p(b,c).$$

Continuous Case

Gluing lemma

X, Y, Z → $\mathbb{R}^d$

↓ espace polonais (~~topologique~~)

Soit $(\alpha, \beta, \gamma) \in \mathcal{M}_1^+(X) \times \mathcal{M}_1^+(Y) \times \mathcal{M}_1^+(Z)$

$\pi \in \mathcal{U}(\alpha, \beta)$, $\xi \in \mathcal{U}(\beta, \gamma)$.

Alors il existe un

$\sigma \in \mathcal{M}_1^+(X \times Y \times Z)$ tq

$(P_{X,Y})_{\#} \sigma = \pi$, $(P_{Y,Z})_{\#} \sigma = \xi$

$P_{X,Y}(x,y,z) = (x,y)$, $P_{Y,Z}(x,y,z) = (y,z)$.

Preuve : admis !

Proposition : Let $X = Y = Z$, $p \geq 1$:

$$C(x,y) = d(x,y)^p$$ avec d distance sur X.

(1) $d(x,y) = d(y,x) \geq 0$

(2) $d(x,y) = 0 \iff x = y$

(3) $d(x,z) \leq d(x,y) + d(y,z)$.

Alors $$W_p(\alpha,\beta) = \mathcal{L}_{d^p}(\alpha,\beta)^{1/p}$$

$$= \left(\inf_{\pi \in \mathcal{U}(\alpha,\beta)} \int d^p(x,y) \, d\pi(x,y) \right)^{1/p}$$

est une distance sur $M_1^+(X)$.

(1) $W_p(\alpha,\beta) = W_p(\beta,\alpha) \geq 0, \forall \alpha,\beta \in M_+^1(X)$

(2) $W_p(\alpha,\beta) = 0 \Leftrightarrow \alpha = \beta$

(3) $W_p(\alpha,\gamma) \leq W_p(\alpha,\beta) + W_p(\beta,\gamma)$

Preuve :

(1) Symetric

$(x,y) \longmapsto \pi(x,y)$ est optimal pour $\mathcal{L}_{d^p}(\alpha,\beta)$ alors

$(x,y) \longmapsto \pi(y,x)$ est optimal pour $\mathcal{L}_{d^p}(\beta,\alpha)$.

(2) Defini

Si $\mathcal{L}_{d^p}(\alpha,\beta) = 0 \to$

$$\inf_{\pi \in U(\alpha,\beta)} \int d^p(x,y)\, d\pi(x,y) = 0$$

$\Rightarrow \operatorname{supp}(\pi) \subseteq \{(x,x) : x \in X\} = \Delta$

$\lambda \cancel{= \pi / \Delta}$

let $\lambda = \pi|_{\Delta}$ ← restriction de π sur α.

Δ ← diagonal.

$\forall h \in \varphi(X)$,

$$\int h(x,y)\, d\pi(x,y) = \int h(x,x)\, d\lambda(x).$$

Comme $\pi \in \mathcal{U}(\alpha,\beta)$

$\rightarrow \lambda = \alpha$ et $\lambda = \beta$.

(3) Inegalité triangulaire

optimaux :

$\pi \in \mathcal{U}(\alpha,\beta)$, $\xi \in \mathcal{U}(\beta,\gamma)$

gluing lemma

$\rightarrow \sigma \in \mathcal{M}_1^+(X \times X \times X)$.

On définit $\rho = (P_{X,Z})_{\#}\sigma$

$$\begin{tikzcd} \alpha \arrow[r, "\pi"] \arrow[rr, bend right, "\rho"'] & \beta \arrow[r, "\xi"] & \gamma \end{tikzcd}$$

$$W_p(\alpha,\gamma) = \inf_{\rho\in\Pi(\alpha,\gamma)} \int_{X\times Z} d(x,z)^p \, d\rho(x,z) \Big)^{1/p} \quad (\text{def})$$

$$\le \left(\int_{X\times Z} d(x,z)^p \, d\rho(x,z)\right)^{1/p} \quad (\text{Sous-optimalité})$$

$$= \left(\int_{X\times Y\times Z} d(x,z)^p \, d\sigma(x,y,z)\right)^{1/p} \quad (\text{def de } \rho)$$

$$\le \left(\int_{X\times Y\times Z} \big(d(x,y)+d(y,z)\big)^p \, d\sigma(x,y,z)\right)^{1/p}$$

(d distance pour X)

$$\le \left(\int_{X\times Y\times Z} d(x,y)^p \, d\sigma(x,y,z)\right)^{1/p} + \left(\int_{X\times Y\times Z} d(y,z)^p \, d\sigma(x,y,z)\right)^{1/p}$$

(Minkowski)

$$= \left(\int_{X\times Y} d(x,y)^p \, d\pi(x,y)\right)^{1/p} + \left(\int_{Y\times Z} d(y,z)^p \, d\xi(y,z)\right)^{1/p}$$

(def de σ)

$$= W_p(\alpha,\beta) + W_p(\beta,\gamma)$$

30/10/24 OT : Course ①

convergence in law topology

X espace metriq compact.

Prop:

Pour $p \leq q$

$$W_p(\alpha, \beta) \leq W_q(\alpha, \beta) \leq \operatorname{diam}(X)^{\frac{q-p}{q}} W_p(\alpha, \beta)^{q/p}$$

convergence faible-*

$(\alpha_k)_{k \in \mathbb{N}}$ cv faiblement-* vers $\alpha \in \mathcal{M}_1^+(X)$ (on note $\alpha_k \rightharpoonup \alpha$)

ssi $\forall f \in \mathcal{C}(X), \int_X f \, d\alpha_k \longrightarrow \int_X f \, d\alpha$

Rq: Cas des mesures discrètes

• $\delta_{x_n} \longrightarrow \delta_x \Longleftrightarrow \forall f \in \varphi(X),$

$$\underbrace{\int f \, d\delta_{x_n}}_{= f(x_n)} \longrightarrow \underbrace{\int f \, d\delta_x}_{f(x)}$$

$\Longleftrightarrow \quad x_n \longrightarrow x$

• $\alpha_n = \sum_i^{k} a_i^{(n)} \delta_{x_i^{(n)}}$

$\beta = \sum_j^{m} b_j \delta_{y_j}$

$\alpha_n \longrightarrow \beta \Longleftrightarrow \begin{cases} \exists\, a_i, \ a_i^{(n)} \longrightarrow a_i, \ i = 1 \cdots k \\ \forall a_i > 0, \ x_i^{(n)} \longrightarrow y_{\sigma(j)} \\ \text{Pour une certaine injection } \sigma \\ \sum_{i \mid \sigma(i) = j} a_i = b_j, \ \forall j \end{cases}$

Rk:

$X_n \sim \alpha_n$, $X \sim \alpha$

$\alpha_n \longrightarrow \alpha \Leftrightarrow X_n \xrightarrow{\mathcal{L}} X$

TCL : $(X_1, \dots, X_n)$ iid avec moment d'ordre 2 fini $\begin{cases} \mathbb{E}(X_i) = 0 \\ \mathbb{E}(X_i X_i^T) = I_d \end{cases}$

alors,

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \xrightarrow{\mathcal{L}} \mathcal{N}(0, I_d).$$

$\boxed{\text{Topologie forte}}$ $\alpha, \beta \in \mathcal{M}^+(X)$

variation total

$$\|\alpha - \beta\|_{TV} = \sup_{\substack{\|f\|_\infty \leq 1 \\ f \in \mathcal{C}^\infty(X)}} \int f \, d(\alpha - \beta)$$

$$= |\alpha - \beta| (X)$$

- Si $\alpha - \beta = \rho \, dx$, $\|\alpha - \beta\|_{TV} = \int |\rho(x)| \, dx$

$$= \|\rho\|_{L^1(X)}$$

- Si $\alpha - \beta = \sum a_i \delta_{x_i}$,

$$\|\alpha-\beta\|_{TV} = \sum |u_i| = \|u\|_1$$

- Wasserstein $\longleftrightarrow$ topologie forte?

$$d_{0/1}(x,y) = \begin{cases} 0 & \text{si } x=y \\ 1 & \text{si } x \neq y \end{cases}$$

distance (0/1).

$$W_p(\alpha,\beta) = \frac{1}{2}\|\alpha-\beta\|_{TV}$$

Cette norme n'a pas de bonnes propriétés

(topologie faible)

Rq : • $\delta_{x_n} \rightarrow \delta_x \Leftrightarrow x_n \rightarrow x$

- Si $x_n \neq x$, $\|\delta_{x_n} - \delta_x\|_{TV} = 2$

→ les diracs ne convergent pas (sauf suite constante).

30/10/24

OT : Course

(2)

Prop

Si X est compact $\alpha_n \to \alpha$

$\iff W_p(\alpha_n, \alpha) \longrightarrow 0$.

* $X = [0,1]$,

$$\frac{1}{n}\sum_{k=1}^{n} \delta_{k/n} \longrightarrow \mathcal{U}_{[0,1]}$$

$$\forall f \in \mathcal{C}(X), \quad \int f \, d\frac{1}{n}\sum_{k=1}^{n} \delta_{k/n} \to \int f \, d\mathcal{U}_{[0,1]}$$

$$\frac{1}{n}\sum_{i=1}^{n} f(k/n) \xrightarrow[n \to +\infty]{} \int f(x)\,dx.$$

Berry-Esseen

$(X_i)_{iid}$ $\mathbb{E}(X_i)=0$, $\mathbb{E}[X_iX_i^T]=Id$,

$\mathbb{E}[\|X_i\|^3]<+\infty$

$\Rightarrow W_p(\alpha_n, \alpha) \leq C \dfrac{\mathbb{E}(\|X_i\|^3)}{\sqrt{n}}$.

$\alpha_n = \dfrac{1}{\sqrt{n}} \sum X_i$

α = mesure gaussienne

vitesse de cvgence $\dfrac{1}{\sqrt{n}}$.

IV/ Sinkhorn

1) Entropic regularization

Entropie : $H(P) = -\sum_{i,j} P_{ij} \log(P_{ij})$.

(OT_ε) : $\min_{P \in U[a,b]} \langle C, P \rangle - \varepsilon H(P)$.

Prop : P est l'unique solution de (OT_ε) ssi $\exists (u,v) \in \mathbb{R}_+^n \times \mathbb{R}_+^m$ tq

$\forall i,j \in \{1,..,n\} \times \{1,..,m\}$,

$$\boxed{P_{ij} = u_i K_{i,j} v_j \ , \ K_{i,j} = e^{-\frac{C_{ij}}{\varepsilon}}}$$

et $P \in U[a,b]$

Preuve

Sans perte de généralité, $a_i, b_j > 0$.

Il existe un unique P^*

solution de ~~(OT_ε)~~ $(OT)_\varepsilon$

(par strict convexité de l'objectif)

* Supposons qu'il existe i,j tq

$P^*_{i,j} = 0$, on note $t > 0$.

$$P_t = (1-t)P^* + t\, a \otimes b \in \mathcal{U}(a,b)$$

$$\boxed{\underline{\text{Rappel}} : \quad (a \times b) = (a_i b_j)_{ij}}$$

Soit

$$\mathcal{E} : P \mapsto \langle C, P \rangle - \varepsilon H(P)$$

On pose $f(t) = \mathcal{E}(P_t)$

$$f'(t) \overset{t \to 0}{=} o(\log t)$$

$\exists t_0 > 0$, $f(t)$ ↘ sur $]0, t_0[$

$\Longrightarrow$ Contradiction $\longrightarrow P^*_{i,j} \neq 0$.

(P) : $\min_{P \in U(a,b)} \mathcal{E}(P)$

$$\mathcal{L}(P, f, g) = \langle P, C \rangle - \varepsilon H(P) + \langle f, a - P \mathbb{1}_m \rangle + \langle g, b - P^T \mathbb{1}_n \rangle$$

$\hookrightarrow \frac{\partial \mathcal{L}}{\partial P}(P^*) = 0$

$\Leftrightarrow \forall i,j, \quad C_{i,j} + \varepsilon(\log P_{i,j} + 1) - f_i - g_j = 0$

$\Leftrightarrow \frac{f_i + g_j - C_{i,j}}{\varepsilon} - 1 = \log P_{ij}$

$\Leftrightarrow P_{i,j} = e^{\frac{f + g - C_{ij}}{\varepsilon} - 1}$

$\cancel{= \left(e^{\frac{f_i}{\varepsilon/2}}\right)\left(e^{-\frac{C_{ij}}{\varepsilon}}\right)\left(e^{\frac{g_j}{\varepsilon/2}}\right)}$

$$= \underbrace{e^{\left(\frac{f_i}{\varepsilon} - 1\right)}} \times e^{\left(-\frac{C_{ij}}{\varepsilon}\right)} \times e^{\left(\frac{g_j}{\varepsilon}\right)}$$

$= u_i \cdot K_{ij} \cdot v_j$

$P_{ij} = u_i K_{ij} v_j \quad ; \quad K_{ij} = e^{-\frac{c_{ij}}{\varepsilon}}$

$\Leftrightarrow \boxed{P = \operatorname{diag}(u)\, K\, \operatorname{diag}(v)}$

$a = P \mathbb{1}_m = \operatorname{diag}(u)\, K \underbrace{\operatorname{diag}(v) \mathbb{1}_n}_{v}$

$b = P^T \mathbb{1}_m = \operatorname{diag}(v)\, K^T \underbrace{\operatorname{diag}(u) \mathbb{1}_n}_{u}$

D'où : $\begin{cases} \operatorname{diag}(u)\, K v = a \\ \operatorname{diag}(v)\, K^T u = b \end{cases}$

$\Leftrightarrow \begin{cases} u \odot (Kv) = a \\ v \odot (K^T u) = b \end{cases}$

Rq : $x \odot y = \begin{pmatrix} x_1 y_1 \\ \vdots \\ x_n y_n \end{pmatrix}$

$\boxed{\text{Algo}}$ Sinkhorn's algo

$u_0, v_0 \in \mathbb{R}^n_+ \times \mathbb{R}^m_+$

$$l=1 \begin{cases} u^{(\ell+1)} = a \oslash K v^{(\ell)} \\ v^{(\ell+1)} = b \oslash K^T u^{(\ell+1)} \end{cases}$$

$$x \oslash y = \begin{pmatrix} x_1/y_1 \\ \vdots \\ x_n/y_n \end{pmatrix}$$

15/11/24 O.T : cours (1)

3) Reformulation par entropie relative

Divergence de Kullback-Leibler entre $P, Q \in \mathbb{R}_+^{n \times m}$,

$$KL(P|Q) = +\sum_{i,j} P_{ij} \log\left(\frac{P_{ij}}{Q_{ij}}\right) - P_{ij} + Q_{ij}$$

avec $0 \log(0)$.

- Si $\exists i,j,\ Q_{ij} = 0 \Rightarrow KL(P|Q) = +\infty$.
- Si P, Q sont des couplages de proba,

$$KL(P|Q) = \sum_{i,j} P_{ij} \log\left(\frac{P_{ij}}{Q_{ij}}\right)$$

On remarque que

$H(P) = -KL(P|\mathbb{1})$

$$\min_{P \in U(a,b)} \langle C, P \rangle + \varepsilon KL(P|a \otimes b).$$

Prop : Pour $P \in U[a,b]$, $\forall a', b' \in \mathbb{R}_+^n \times \mathbb{R}_+^m$

$\sum a_i \log \frac{a_i}{a'_i} - a_i + a'_i$

$$KL(P|a \otimes b) = KL(P|a' \otimes b') - KL(a|a') - KL(b|b')$$

Preuve

$$KL(P|a \otimes b) = \sum_{i,j} P_{ij} \log\left(\frac{P_{ij}}{a_i b_j}\right)$$

$$= \sum_{ij} P_{ij} \log\left(\frac{P_{ij}\, a'_i b'_j}{a'_i b'_j\, a_i b_j}\right)$$

$$= \underbrace{\sum_{ij} P_{ij} \log\left(\frac{P_{ij}}{a'_i b'_j}\right)}_{KL(P|a' \otimes b')} + \sum_{ij} P_{ij} \log\left(\frac{a'_i b'_j}{a_i b_j}\right)$$

$= KL(P|a' \otimes b') + A$ où

$$A = \sum_{ij} P_{ij} \log\left(\frac{a_i b_j}{a_i' b_j'}\right)$$

$$= \sum_{ij} P_{ij} \log\left(\frac{a_i b_j}{a_i' b_j'}\right)$$

$$= -\sum_{ij} P_{ij} \log\left(\frac{a_i}{a_i'}\right) + P_{ij} \log\left(\frac{b_j}{b_j'}\right)$$

$$= -\sum_{ij} P_{ij} \log\left(\frac{a_i}{a_i'}\right) - \sum_{ij} P_{ij} \log\left(\frac{b_j}{b_j'}\right)$$

$$= -\sum_{i} a_i \log\left(\frac{a_i}{a_i'}\right) - \sum_{j} b_j \log\left(\frac{b_j}{b_j'}\right)$$

$$= -KL(a|a') - KL(b|b').$$

Prop

l'unique $P_\varepsilon = \underset{P \in U(a,b)}{\operatorname{argmin}} \langle C, P \rangle - \varepsilon H(P)$

$= L_C^\varepsilon(a,b)$ cv

quand $\varepsilon \to 0$

~~quand $\varepsilon \to 0$~~

vers la solution du problème de

de Kantorovich d'entropie maximale

$P_\varepsilon \xrightarrow{\varepsilon \to 0} \underset{P}{\operatorname{argmin}} \left\{ -H(P) \,\middle|\, \begin{array}{l} P \in U[a,b] \\ \langle P, C\rangle = L_C(a,b) = \min\limits_{\tilde{P} \in U[a,b]} \langle \tilde{P}, C\rangle \end{array} \right\}$

En particulier,

$L_C^\varepsilon(a,b) \xrightarrow{\varepsilon \to 0} L_C(a,b)$ et

$P_\varepsilon \xrightarrow{\varepsilon \to \infty} a \otimes b$

Rq : $P_\varepsilon = \operatorname{argmin}$ & $L_C^\varepsilon = \min$

Preuve :

—o—

$\boxed{\varepsilon \to 0}$ On prend $(\varepsilon_\ell) \downarrow 0$ avec

$\varepsilon_\ell > 0$,

$P_\ell = P_{\varepsilon_\ell} = \underset{P \in U[a,b]}{\operatorname{argmin}} \langle C, P\rangle - \varepsilon_\ell H(P)$.

$U[a,b]$ compact de $\mathbb{R}_+^{n \times m}$, $\forall \ell$,

$P_\ell \in U[a,b]$

15/11/24 OT : Cours ②

On extrait

$(P_{\varphi(k)}) \longrightarrow \bar{P} \in \mathcal{U}_{[a,b]}$ (B. weierstrass).

On choisit $\tilde{P}$ tq $L_c(a,b) = \langle C, \tilde{P} \rangle$

$\forall \ell, \begin{cases} 0 \leq \langle C, P_\ell \rangle - \varepsilon_\ell H(P_\ell) \leq \langle C, \tilde{P} \rangle - \varepsilon_\ell H(\tilde{P}) \\ 0 \leq \langle C, \tilde{P} \rangle \leq \langle C, P_\ell \rangle \end{cases}$

$0 \leq \langle C, P_\ell \rangle + \varepsilon_\ell KL(P_\ell \mid a \otimes b) \leq \langle C, \tilde{P} \rangle + \varepsilon_\ell KL(\tilde{P} \mid a \otimes b)$

$\langle C, P_\ell \rangle - \langle C, \tilde{P} \rangle \leq \varepsilon_\ell \left(KL(\tilde{P} \mid a \otimes b) - KL(P_\ell \mid a \otimes b) \right)$

Rq $-H(P) = KL(P \mid 1_n \times 1_n)$

$KL(P \mid a \otimes b) - KL(a \mid 1) - KL(b \mid 1)$

(continuité de KL)

$\rightarrow)$ $\varepsilon \to +\infty$, $\langle C, P_\varepsilon \rangle \xrightarrow{\varepsilon \to \infty} \langle C, \tilde{P} \rangle$

$$\Rightarrow \begin{cases} L_C^{\varepsilon}(a,b) \longrightarrow L_C(a,b) \\ \bar{P} \in \underset{P \in U_{[a,b]}}{\operatorname{argmin}} \langle C, P \rangle \end{cases}$$

$$0 \leq \frac{\langle C, P_\varepsilon \rangle - \langle C, \tilde{P} \rangle}{\varepsilon} \leq KL(\tilde{P} | a \otimes b) - KL[P_\varepsilon | a \otimes b)$$

$$\xrightarrow{\varepsilon \to \infty} 0 \leq KL(\tilde{P} | a \otimes b) - KL(\bar{P} | a \otimes b)$$

$$\Leftrightarrow KL(\bar{P} | a \otimes b) \leq KL(\tilde{P} | a \otimes b)$$

$$\Leftrightarrow -H(\bar{P}) \leq -H(\tilde{P})$$

$$\Leftrightarrow \bar{P} \in \operatorname{argmin} \{ -H(P) \mid \langle C, P \rangle = L_C(a,b), \; P \in U_{[a,b]} \}$$

$\boxed{\varepsilon \to \infty}$;

$\overbrace{\langle C, P_\varepsilon \rangle}^{\geq 0} + \varepsilon\, KL(P_\varepsilon | a \otimes b)$

$\leq \langle C, a\otimes b \rangle + \varepsilon \times 0$

$KL(P_\varepsilon | a\otimes b) \leq \frac{1}{\varepsilon} \langle C, a \otimes b \rangle \leq \frac{C}{\varepsilon} \to 0$

$\Rightarrow P_\varepsilon \to a\otimes b$; car

$KL(P | Q) = 0 \Leftrightarrow P = Q$

4) CVgence de Sinkhorn

Propi: On a $\langle P, C \rangle + \varepsilon KL(P | a\otimes b)$
$= \varepsilon KL(P | K) + cst$

Preuve: $\sum_{ij} P_{ij} C_{ij} + \varepsilon \sum_{ij} P_{ij} \log\left(\frac{P_{ij}}{a_i b_j}\right)$

$$= \sum_{ij} P_{ij}\left[\log\left[\exp(c_{ij})\right] + \varepsilon \log\left(\frac{P_{ij}}{a_i b_j}\right)\right]$$

$$\cancel{= \sum_{ij} \varepsilon P_{ij}\left(\log\left(\frac{\exp(c_{ij})}{\exp(\varepsilon)}\right) + \log\left(\frac{P_{ij}}{a_i b_j}\right)\right]}$$

$$\cancel{= \varepsilon \sum_{ij} P_{ij} \log\left[\frac{\exp(c_{ij})}{\exp(\varepsilon)} \cdot \frac{P_{ij}}{a_i b_j}\right)}$$

$$= \sum_{ij} \varepsilon P_{ij}\left[\log\left[\exp\left(\frac{c_{ij}}{\varepsilon}\right)\right] + \log\left(\frac{P_{ij}}{a_i b_j}\right)\right]$$

$$= \varepsilon \sum_{ij} P_{ij} \log\left(\exp\left(\frac{c_{ij}}{\varepsilon}\right) \cdot \frac{P_{ij}}{a_i b_j}\right)$$

$$= \varepsilon \sum_{ij} P_{ij} \log\left(\frac{P_{ij}}{\exp\left(-\frac{c_{ij}}{\varepsilon}\right)} \times \frac{1}{a_i b_j}\right)$$

$$= \varepsilon KL(P|K) + \varepsilon \sum P_{ij} \log\left(\frac{1}{a_i b_j}\right).$$

D'où:

$$P_\varepsilon = \operatorname*{argmin}_{P \in U(a,b)} KL(P|K).$$

15/11/24 OT: Course ③

Posons $A_\varepsilon = \underset{P \in U_{[a,b]}}{\operatorname{argmin}} \|P - K\|_2^2$

d'où :

$$P_\varepsilon = \Pi^{KL}_{U_{[a,b]}}(K)$$

Soit

$$\mathcal{C}_a = \{P \mid P\mathbb{1} = a\}$$

$$\mathcal{C}_b = \{P \mid P^T\mathbb{1} = b\},$$

$$U_{[a,b]} = \mathcal{C}_a \cap \mathcal{C}_b$$

NB : Contraintes affines.

Minimisation alternée

$$\begin{cases} P_{k+\frac{1}{2}} = \Pi^{KL}_{\mathcal{C}_a}(P_k) \\ P_{k+1} = \Pi^{KL}_{\mathcal{C}_b}(P_{k+1/2}) \end{cases}$$

e

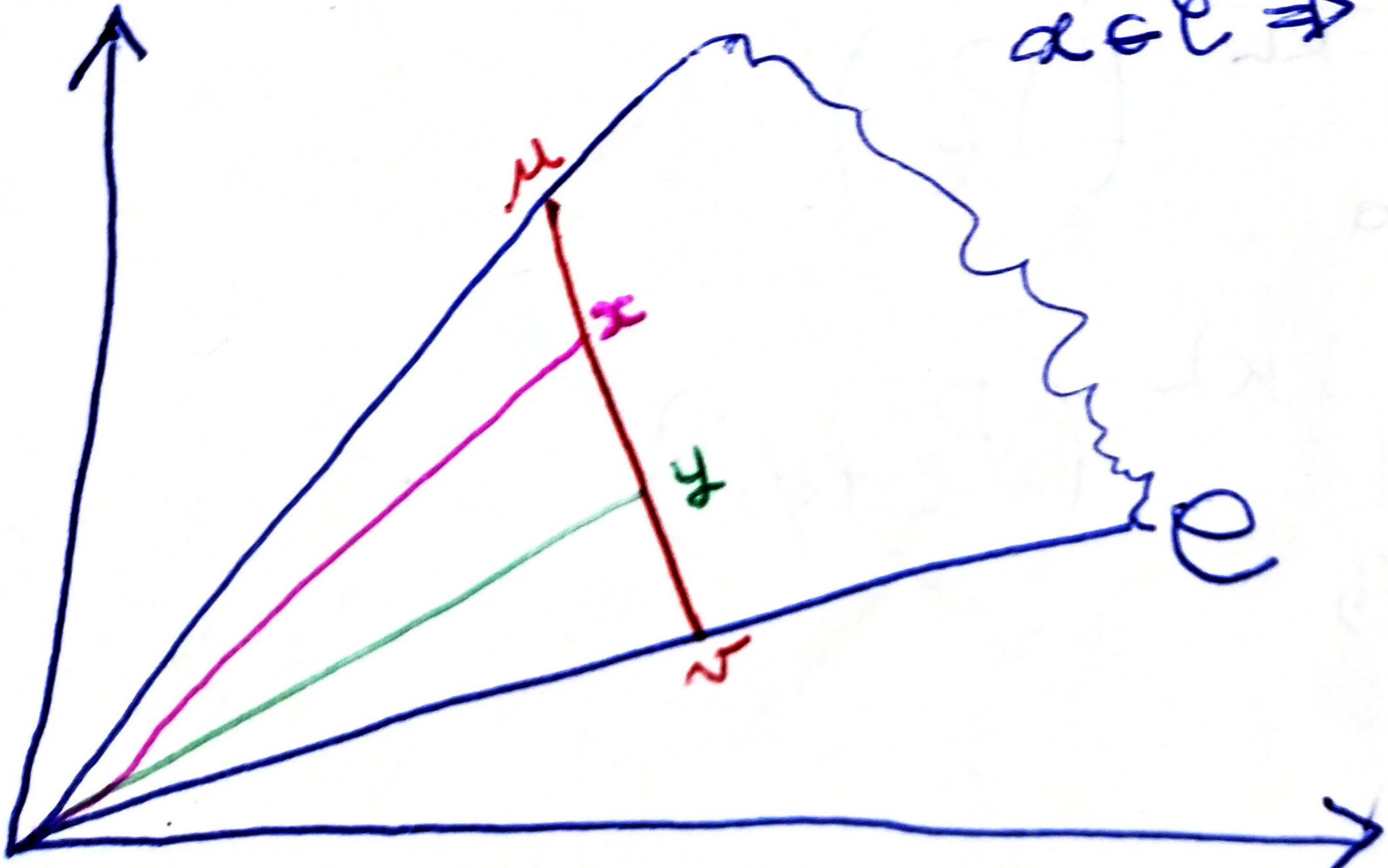
On peut montrer que

$$\begin{cases} \pi^{KL}_{C_a}(P) = \operatorname{diag}\left(\dfrac{a}{P\mathbb{1}_m}\right) P \\ \pi^{KL}_{C_b}(P) = P \operatorname{diag}\left(\dfrac{b}{P^T\mathbb{1}_n}\right). \end{cases}$$

Convergence en métrique d'Hilbert

$\mathcal{C}$: cone convexe

$x \in \mathcal{C} \Rightarrow \lambda x \in \mathcal{C}, \lambda > 0.$



$$d_H(x,y) = \log \frac{M(x,y)}{m(x,y)}$$ Distance d'Hilbert

avec $\begin{cases} M(x,y) = \inf\{\lambda \mid x - \lambda y \in \mathcal{C} \\ m(x,y) = \sup\{\lambda \mid \lambda y - x \in \mathcal{C} \end{cases}$

d_H est une distance projective

1) $d_H(x,y) \geq 0$

2) $d_H(x,y) \not\geq \cancel{d}_H(y,x)$

3) $d_H(x,y) \leq d_H(x,z) + d_H(z,y)$

4) $d_H(x,y) = 0 \Leftrightarrow x = \lambda y$ avec $\lambda > 0$

On peut montrer que

$\begin{cases} M(x,y) = \overline{xv} / \overline{yv} \\ m(x,y) = \overline{ux} / \overline{uy} \end{cases}$

Théo 1) Soit $K \in \mathbb{R}^{n,m}_{+,*}$, alors pour $(v, v') \in (\mathbb{R}^m_{+,*})^2$ avec,

- $d_H(Kv, Kv') \le \lambda(K)\, d_H(v, v')$

$$\begin{cases} \lambda(K) = \dfrac{\sqrt{\eta(K)} - 1}{\sqrt{\eta(K)} + 1} < 1 \\ \eta(K) = \max\limits_{i,j,k,l} \dfrac{K_{ik} K_{jl}}{K_{jk} K_{il}} \end{cases}$$

Théo 2 On a $(u_k, v_k) \to (u^*, v^*)$

et $$\begin{cases} d_H(u_k, u^*) = O(\lambda(K)^{2k}) \\ d_H(v_k, v^*) = O(\lambda(K)^{2k}) \end{cases}$$

et $\|\log P_k - \log P^*\|_\infty \le d_H(u_k, u^*) + d_H(v_k, v^*)$

15/11/2024 ④

Preuve

- On prends $v, v' \in \mathbb{R}^m_{+,*}$,

$$d_H(v, v') = d_H(v/v', \mathbb{1}_n)$$

$$= d_H(\mathbb{1}_n/v, \mathbb{1}_n/v').$$

En effet,

$$d_H\left(\frac{v}{v'}, \mathbb{1}_m\right) = \log \frac{M\left(\frac{v}{v'}, \mathbb{1}_m\right)}{m\left(\frac{v}{v'}, \mathbb{1}_m\right)}$$

et $M\left(\frac{v}{v'}, \mathbb{1}_m\right) = \inf\left\{\lambda \geq 0 \mid \frac{v}{v'} - \lambda \mathbb{1}_m \in \mathbb{R}^m_{+,*}\right\}$

$$\forall j, \frac{v_j}{v_j'} - \lambda > 0$$

$$\forall j, v_j - \lambda v_j' \geq 0$$

$$v - \lambda v' \in \mathbb{R}^m_{+,*}$$

- $d_H(u_{k+1}, v^*) = d_H\left(\frac{a}{K v_k}, \frac{a}{K v^*}\right)$

$$= d_H(K v_k, K v^*)$$

$$\leq \lambda(K)\, d_H(v_k, v^*)$$

$\leq \lambda(k)\,\lambda(k)\; d_H(u_k, u^*)$ en refaisant le même raisonnement

$\leq$

$\Rightarrow\; d_H(u_{k+1}, u^*) \leq \lambda(k)\; d_H(u_0, u^*)$

Monge discret

$$\min_{\sigma \in S_n} \sum C_{i,\sigma(i)}$$

SOLUTIONS

$n = m$

$n = m$ ⇕

Monge continu

$$\min_{T_\# \alpha = \beta} \int C(x, T(x))\, d\alpha(x)$$

Soluti. ? en général non

euclidien, coût quadratique

départ ... mesure de départ

BRENIER

Kantorovich discret

$$\min_{P \in U_{[a,b]}} \langle C, P \rangle$$

$$U_{[a,b]} = \left\{ P \in \mathbb{R}_+^{n \times m} \;\middle|\; \begin{array}{l} P\mathbb{1} = a \\ P^T \mathbb{1} = b \end{array} \right\}$$

Solutions

Kantorovich continu

$$\min_{\pi \in U(a,b)} \int C(x,y)\, d\pi(x,y)$$

$$U(\alpha, \beta) = \left\{ \pi \in \mathcal{M}^+(X, Y) \;\middle|\; \begin{array}{l} \pi_1 = \alpha \\ \pi_2 = \beta \end{array} \right\}$$