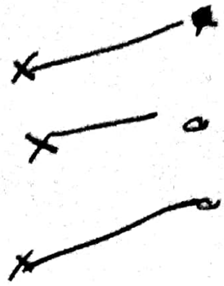
13/11

# Transport Optimal

①

S.QUERK

<u>Input</u> : $(x_i)_{1\le i\le n}$ , $(y_j)_{1\le j\le n}$



$$\min_{\sigma} \sum_{i=1}^{n} C(x_i, y_{\sigma(i)})$$

$$\sum_{i=1}^{n} C_{i\,\sigma(i)} \qquad C_{ij} = \text{cost}\,(i \to j)$$

appareillement optimal – optimal matching.

how to adress mass spliting ?

<u>1D</u> : Si $C_{ij} = |x_i - y_i|^p$ $p \geq 1$

<u>Prop</u> : $\sigma^*$ is monotonic, not unique.

without loss of generality

$x_1 \leq x_2 \leq \dots \leq x_n$

$y_1 \leq y_2 \leq \dots \leq y_n$

We just take $\sigma = id_{[\![1,n]\!]}$.

Le resultat reste vrai dans le cas $C_{ij} = h(x_i - y_j)$ avec $h$ convexe.

<u>Complexité</u> : $O(n \log n)$.

<u>papier of Monge : cas de la dimension 2</u>

<u>resultat</u> : " optimal assignment do not cross ".


propriété géometrique. Non suffisant.

## General case / General Monge problem.

Soit $\mathcal{X}$, $\mathcal{Y}$ deux espaces métriques complets

$\alpha$, $\beta$ deux mesures de Radon (= Borel).

duales de fonctions continues.

Pour $A \subset \mathcal{X}$, $\alpha(A) \in \mathbb{R}_+$.

$A, B \subseteq \mathcal{X}$, $A \cap B = \emptyset$, $\alpha(A \cup B) = \alpha(A) + \alpha(B)$.

On travail avec des mesures de proba.

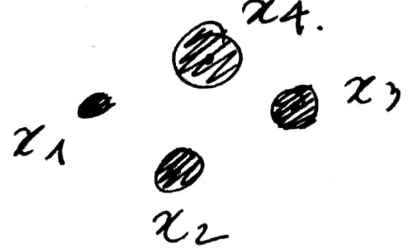
$\alpha(\mathcal{X}) = 1$. On peut calculer des intégrales

$\underbrace{\int f(x)\alpha(dx)}_{\text{" produit scalaire "}}$ pour toute fonction mesurable $f$.

" produit scalaire ", noté $\langle f, \alpha \rangle := \int f d\alpha$.

<u>Mesures discrètes</u> : somme de Dirac.

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \qquad \delta_x(A) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon.} \end{cases}$$




$$\int f d\delta_x = f(x)$$

On a $\displaystyle\int_{\mathcal{X}} f(x)\alpha(dx) = \sum_{i=1}^{n} f(x_i) a_i$

$\alpha$ est une mesure de probabilité si

$\forall i \; a_i \geq 0$ et $\displaystyle\sum_{i=1}^{n} a_i = 1$. histogramme.

$\underline{\{a \in \mathbb{R}_+^n, \sum_{i=1}^{n} a_i = 1\}}$

simplexe.

# Transport Optimal



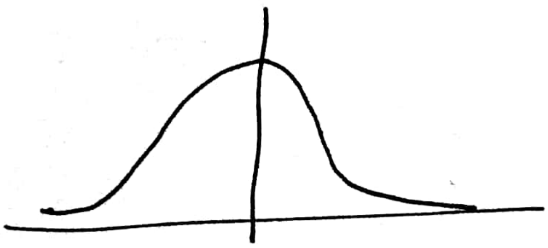
Une densité $\rho$ est une fonction positive $\rho(x) \geq 0$ pour une mesure de référence $\gamma \in \mathcal{M}^+(\mathcal{X})$.

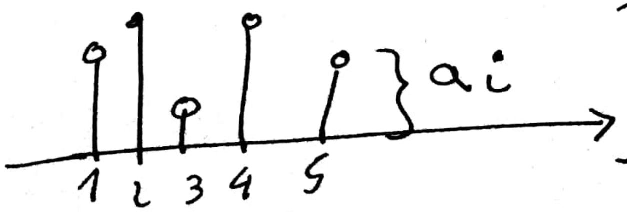
**Def** : La mesure $\alpha$ a une densité $\rho = \frac{d\alpha}{d\gamma}$ par rapport à $\gamma$

ssi $$\int f(x)\, d\alpha(x) = \int f(x)\,\rho(x)\, d\gamma(x)$$

**Exemple classique** : $$\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} = \frac{d\alpha}{dx}$$

élément d'intégration de la mesure de Lebesgue.






densité de $\alpha = \sum_{i=1}^{n} a_i \delta_{x_i}$

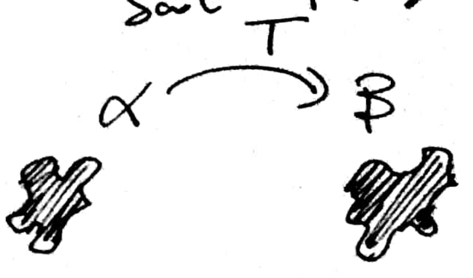
## Mesure image (Push Forward)

$T : \mathcal{X} \to \mathcal{Y}$

$T_\# : \alpha \to \beta$

$\mathcal{M}(\mathcal{X}) \to \mathcal{M}(\mathcal{Y})$

Soit $T : \mathcal{X} \to \mathcal{Y}$ mesurable.




**Def** : $T_\# \delta_x = \delta_{T(x)}$.

$T_\#$ est linéaire.

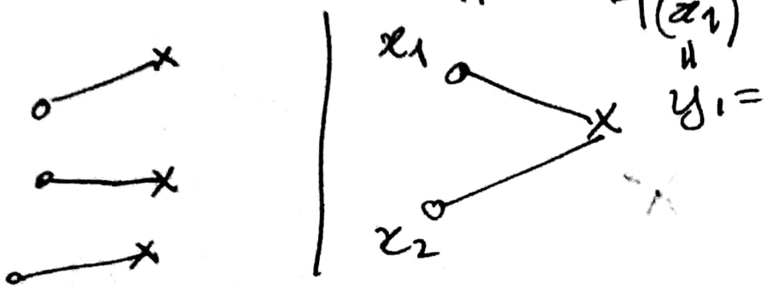
$\beta := T_\# \alpha$.

Par linéarité, $T_\# \underbrace{\left(\sum_{i=1}^{n} a_i \delta_{x_i}\right)}_{\alpha} = \sum_{i=1}^{n} a_i \delta_{T(x_i)} = \beta$

On ne peut pas augmenter le nombre de points.

En effet $\mathrm{supp}(\alpha) = \{x_i\}$ $\quad \mathrm{supp}(\beta) = \{T(x_i)\}$

Alors $\# \mathrm{supp}(\beta) \leq \# \mathrm{supp}(\alpha)$.

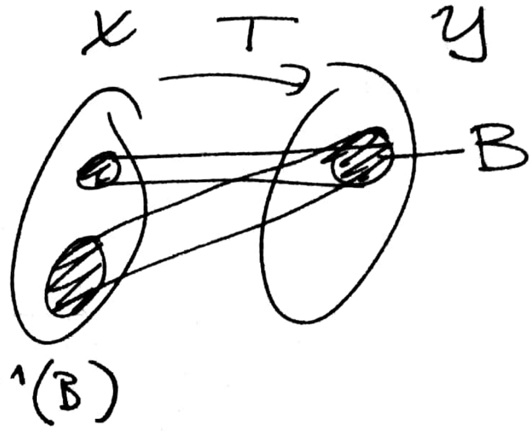



$\delta_{x_1} \neq \delta_{x_2}$, $T(x_1) = T(x_2) = y_1$ $\qquad \delta_{y_1} + \delta_{y_2} = 2\delta_{y_1}$.

Def : $\beta = T_\# \alpha$.

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

$T^{-1}(B) = \{x,\ T(x) \in B\}$.




Définition alternative :

$\forall g \in \mathcal{C}(\mathcal{Y})$. $\quad \int g(y)\, d\beta(y) := \int g(T(x))\, d\alpha(x)$

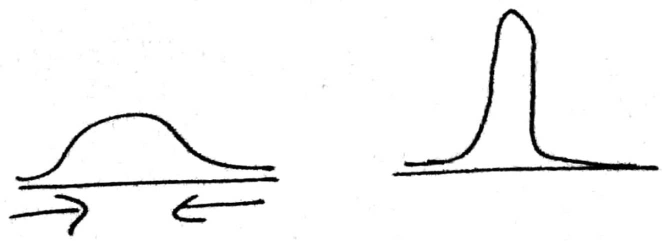
On retrouve la définition précédente en choisissant $g = \mathbb{1}_B$.

Scanné avec CamScanner

13/11. # Transport Optimal



$$\frac{d\alpha}{dx}(x) = \rho_\alpha(x)$$



$$\frac{d\beta}{dy}(y) = \rho_\beta(y)$$

$$y = T(x) \qquad \text{"} dy = |\det(T'(x))|\, dx \text{"}$$

$$T'(x) = \partial T(x) = J_T(x)$$

$T : \mathbb{R}^d \to \mathbb{R}^{q}$ ! $\quad \partial T(x) : \mathbb{R}^d \to \mathbb{R}^q$ linear.

$$T(x) = \begin{pmatrix} T_1(x) \\ | \\ T_q(x) \end{pmatrix} \qquad \partial T(x) = \begin{bmatrix} \frac{\partial T_1}{\partial x_1}(x) & \cdots & \frac{\partial T_1}{\partial x_d}(x) \\ \vdots & \ddots & \vdots \\ \frac{\partial T_q}{\partial x_1}(x) & \cdots & \frac{\partial T_q}{\partial x_d}(x) \end{bmatrix}$$

<u>Prop</u> : Si $T$ est une bijection régulière.

$$\boxed{\rho_\alpha(x) = |\det \partial T(x)|\, \rho_\beta(T(x))}$$

connection avec les equations differentielles.

## connection avec les probas

$\mathcal{X}$ : espace métrique, $X$ variable vecteur aléatoire.

$\alpha \in \mathcal{M}_1^+(\mathcal{X}) \equiv \mathcal{P}(\mathcal{X})$.

$\alpha(A) = \mathbb{P}(X \in A)$.

$\mathbb{E}(f(X)) = \int f(x)\, \alpha(dx)$

$\mathbb{E}(X) = \text{mean}(X)$

$= \int x\, \alpha(dx)$

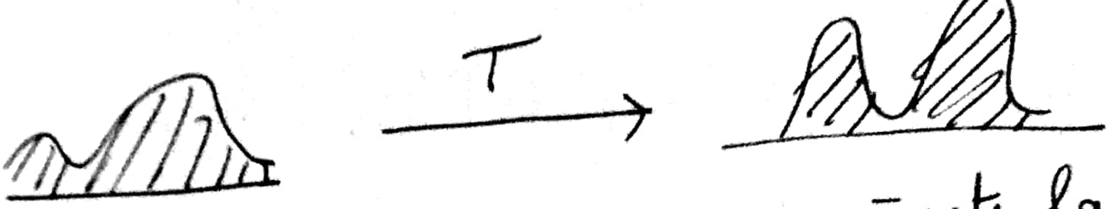
$= \sum_{i=1}^{n} a_i x_i$.

# General Monge Problem

Soit $\alpha \in \mathcal{P}_1^+(\mathcal{X})$, $\beta \in \mathcal{P}_1^+(\mathcal{Y})$, $c(x,y)$.

**Problème de Monge :**

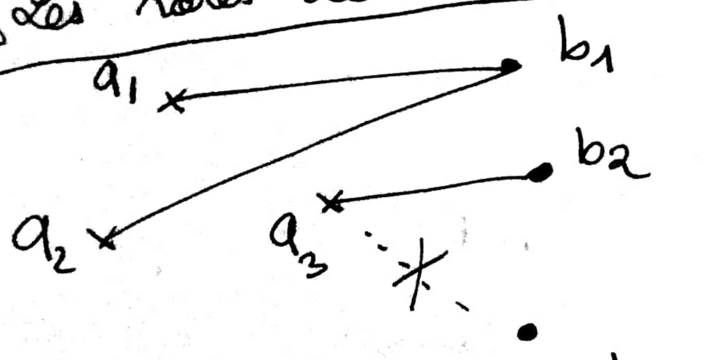
$$\inf \left\{ \int_{\mathcal{X}} c(x, T(x))\, \alpha(dx) \;\middle|\; \begin{array}{l} T : \mathcal{X} \to \mathcal{Y} \\ T\# \alpha = \beta \end{array} \right\}$$




L'équation $T\#\alpha = \beta$ représente la conservation de la masse.

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \qquad \beta = \sum_{j=1}^{m} b_j \delta_{y_j}$$

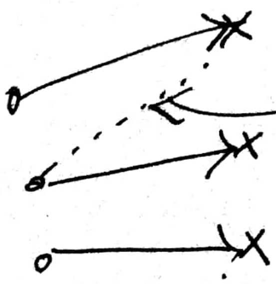
⚠ Les rôles de $\alpha$ et $\beta$ ne sont pas symétriques.




$a_1 + a_2 + a_3 = b_1$

$a_3 = b_2$.

cas où : $n = m$ et $a_i = b_i = \frac{1}{n}$. (pb de Monge discret).



impossible.

mesure uniforme avec un nombre fini de points

$\alpha = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$, $\beta = \frac{1}{n} \sum_{i=1}^{n} \delta_{y_i}$ $\Rightarrow$ on minimise sur les permutations.

Dans ce cas spécial, $T_\# \alpha = \beta \iff \exists \sigma,\ T(x_i) = y_{\sigma(i)}$.

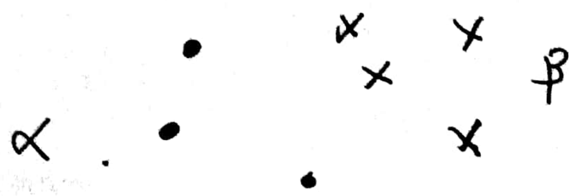
# Transport Optimal



$$(M) \Leftrightarrow \min_{\sigma \in \text{Perm}_n} \frac{1}{n} \sum_{i=1}^{n} c(x_i, \underbrace{T(x_i)}_{= y_{\sigma(i)}}) = \frac{1}{n} \sum_{i=1}^{n} C_{i,\sigma(i)}$$

où $C_{ij} = c(x_i, y_j)$.

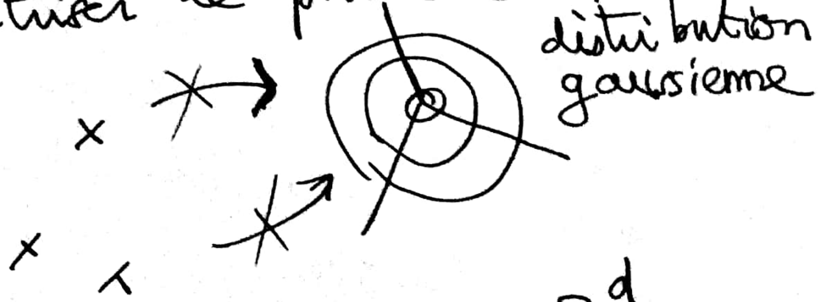
Il se peut que $T_\# \alpha = \beta$ n'ait pas de solution.




Le pb de Monge de $\beta$ à $\alpha$ a un sens, mais celui de $\alpha$ à $\beta$ n'en a pas.

Comment symétriser le problème ?

<u>autre pb</u> :




<u>Brenier's theorem.</u> $\mathcal{X} = \mathcal{Y} = \mathbb{R}^d$

$c = \|\cdot\|_2^2$.

<u>hypothèse</u> : $\alpha$ a une densité par rapport à la mesure de Lebesgue.

Il existe une unique solution $T$ pour (M).

De plus, $T$ est caractérisé par

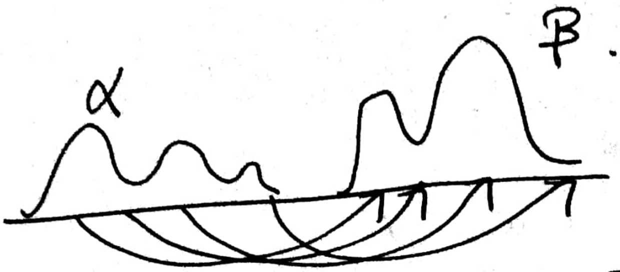
$T_\# \alpha = \beta$ et il existe $\varphi : \mathbb{R}^d \to \mathbb{R}$

convexe telle que $T(x) = \nabla\varphi(x)$.

Rmq : Une fonction convexe $\varphi$ est différentiable presque partout.

donc $T = \nabla\varphi$ est bien définie.

En dim 1 : $\nabla\varphi = \varphi'$ est monotone.




Proposition. Si $T(x) = \nabla\varphi(x)$.

alors $T$ est un opérateur monotone.

Pour tout $x, y$ $\langle T(x) - T(y), x - y \rangle \geq 0$

La réciproque n'est pas vrai.

Il y a des champs de vecteurs monotones qui ne sont pas les gradients de fonctions convexes.

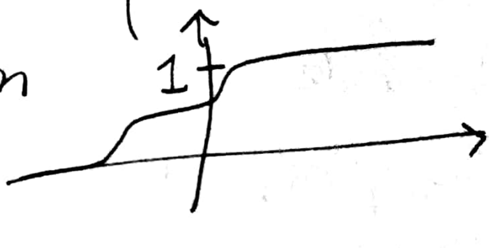
ex : Si $T$ est une petite rotation.

Cas général 1D. hyp : $\alpha$ admet une densité de support $I = [a, b)$

$\rho_\alpha(x) > 0$ sur $I$

définition : fonction de répartition

$$C_\alpha(x) = \int_{-\infty}^{x} d\alpha(x)$$

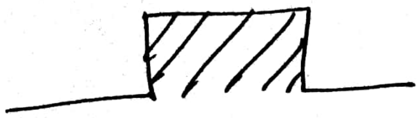


13/11 **Transport Optimal.** 

On définit $Q_\alpha = C_\alpha^{-1}$ dans le cas où $C_\alpha$ est une bijection entre $[a, b]$ et $[0, 1]$.

**Proposition (exercice).** $(C_\alpha)_\# \alpha = \mathcal{U}([0,1]) = \mathbb{1}_{[0,1]} \mathcal{L}$

$$\boxed{(Q_\alpha)_\# \mathbb{1}_{[0,1]} = \alpha}.$$



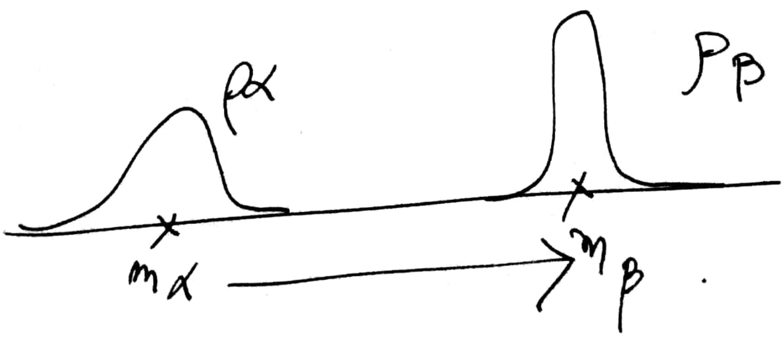
$$\alpha \xrightarrow[\#]{C_\alpha} \mathbb{1}_{[0,1]} \xrightarrow[\#]{Q_\beta} \beta$$

**Théorème** : $T = Q_\beta \circ C_\alpha$ est l'unique transport optimal pour $\alpha$ & $\beta$

**Cas gaussien de dimension $d$.**

$$\rho_\alpha(x) = \frac{1}{\sqrt{2\pi}\,\sigma_\alpha} \exp\left(-\frac{(x - m_\alpha)^2}{2\sigma_\alpha^2}\right).$$

$\sigma_\alpha = 0 \iff \alpha = \delta_{m_\alpha}$ loi normale dégénérée.




$T(x) = ax + b$ ?

**Prop** : $T(x) = \frac{\sigma_\beta}{\sigma_\alpha}(x - m_\alpha) + m_\beta$.

Alors $T_\# \alpha = \beta$ (preuve par changement de variable).

Scanné avec CamScanner

<u>Théorème</u> : Le transport optimal $T^*$ entre deux gaussiennes 1 d.

$$\int_{x \in \mathbb{R}} \|x - T^*(x)\|^2 \alpha(dx) = (m_\alpha - m_\beta)^2 + (\sigma_\alpha - \sigma_\beta)^2$$

$=$ distance euclidienne.

20/11/23 O.T 

$$\tilde{W}(\alpha,\beta) = \left[\inf_{T\#\alpha=\beta} \int \|x - T(x)\|^p d\alpha(x)\right]^{\frac{1}{p}}$$

$$\tilde{W}_p(\alpha,\beta) \geqslant 0$$

Prop : $\tilde{W}_p$ satisfait l'inégalité triangulaire et $\tilde{W}_p(\alpha,\beta) = 0 \Leftrightarrow \alpha=\beta$.

$$\tilde{W}_p(\alpha,\beta) = +\infty \text{ si } \{T : T_\# \alpha = \beta\} = \emptyset.$$

1.D case : $C_\alpha(x) = \int_{-\infty}^{x} d\alpha(x)$

$$\alpha \xrightarrow{C_\alpha} \mathbb{1}[0,1] \xrightarrow{C_\beta^{-1}} \beta \qquad C_\beta^{-1} = Q_\beta.$$

Prop : $\tilde{W}_p(\alpha,\beta) = \left[\int_0^1 |C_\alpha^{-1}(t) - C_\beta^{-1}(t)|^p dt\right]^{\frac{1}{p}}$

closed-form formula.

$p=2$. $\tilde{W}_2(\alpha,\beta) = \|C_\alpha^{-1} - C_\beta^{-1}\|_{L^2([0,1])}$

Prop : $\tilde{W}_1(\alpha,\beta) = \|C_\alpha - C_\beta\|_{L^1}$

$$= \int_{\mathbb{R}} \left|\int_{-\infty}^{x} d(\alpha-\beta)\right| dx.$$

$p = 1$. $\tilde{W}_1(\alpha, \beta) = \| \alpha - \beta \|_{W_1}$

works for $d \geq 1$.

$\boxed{\text{Gaussian measures}}$ Rappel : $d = 1$ : $T(x) = \frac{\sigma_\beta}{\sigma_\alpha}(x - m_\alpha) + m_\beta$.

$$\tilde{W}_2(\alpha, \beta)^2 = (\sigma_\alpha - \sigma_\beta)^2 + (\mu_\alpha - \mu_\beta)^2$$

exercice à faire.

Prop : $\tilde{W}_2(\alpha, \beta)^2 = \tilde{W}_2^2(\alpha_0, \beta_0) + \| m_\alpha - m_\beta \|^2$

$\zeta : x \mapsto x - m_\alpha$ $\quad \alpha_0 \triangleq \zeta_\# \alpha$

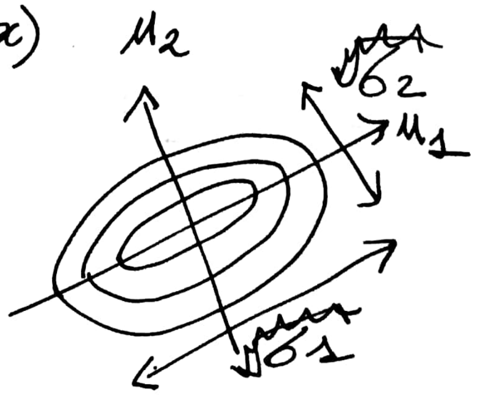
$\quad$ " $= \alpha(\cdot - m_\alpha)$ "

$\Sigma_\alpha$ covariance of $\alpha$.

$m_\alpha = \alpha$ $\quad \Sigma_\alpha = \int x x^T d\alpha(x)$

$\in \mathbb{R}^{d \times d}$.

$$\Sigma_\alpha = U_\alpha \, \mathrm{diag}(\underbrace{\sigma_\alpha^2}_{\in \mathbb{R}_+^d}) \, U_\alpha^T$$




Try : $A(x - m_\alpha) + m_\beta =: T(x)$.

$$A = \Sigma_\beta \Sigma_\alpha^{-1} ?$$

Rappel : $\frac{d\alpha}{d\mathcal{L}} = \frac{1}{\sqrt{2\pi |\Sigma_\alpha|}} e^{-\underbrace{\frac{(x - m_\alpha)^T \Sigma_\alpha^{-1}(x - m_\alpha)}{2}}_{\frac{1}{2}\langle \Sigma_\alpha^{-1} x, x \rangle}}$

**Prop** : $T_\# \alpha$ is gaussian $\mathcal{N}(m_\beta, A^T \Sigma_\alpha A)$

Pour que $T$ soit un transport optimal, il suffit que $T$ soit le gradient d'une fonction convexe. Pour cela il faut que $T$ soit le gradient d'une quadratique

**Prop** : $A$ gradient of convex function
$$\Leftrightarrow A^T = A \succcurlyeq 0.$$

$$f(x) = \frac{1}{2} \langle Bx, x \rangle + \langle a, x \rangle$$
$$\partial f(x) = (B + B^T) x + \cancel{\langle a_k}$$
$$\partial^2 f(x) = B + B^T \succcurlyeq 0.$$

---

$T(x) = A(x - m_\alpha) + m_\alpha$ so that :

**Prop** : Ricatti equation, Kalman filtering.

unique OT $\Leftrightarrow \begin{cases} A \Sigma_\alpha A = \Sigma_\beta \\ A = A^T \succcurlyeq 0 \end{cases}$ At most one matrix $A$.

Although Brenier theorem has nothing to do with linear algebra, in the G. case it's an Algebra theorem.

Scanné avec CamScanner

**Rmq** : Si $A \geqslant 0$, $\exists B = \sqrt{A} \geqslant 0$ $B^2 = A$.

**Prop** : $A = U \operatorname{diag}(\lambda) U^T$.

$$\sqrt{A} = U \operatorname{diag}(\sqrt{\lambda}) U^T$$

$\sqrt{A}\, A = (\sqrt{A})^3 = A\sqrt{A}$ donc $A$ et $\sqrt{A}$ commutent.

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

$$\sqrt{\Sigma_\alpha} A \sqrt{\Sigma_\alpha} \sqrt{\Sigma_\alpha} A \sqrt{\Sigma_\alpha} = \sqrt{\Sigma_\alpha}\, \Sigma_\beta \sqrt{\Sigma_\alpha}$$

$$\left(\sqrt{\Sigma_\alpha} A \sqrt{\Sigma_\alpha}\right)^2 = \sqrt{\Sigma_\alpha}\, \Sigma_\beta \sqrt{\Sigma_\alpha}$$

$$\Leftrightarrow \quad A = \sqrt{\Sigma_\alpha}^{-1} \left(\sqrt{\Sigma_\alpha}\, \Sigma_\beta \sqrt{\Sigma_\alpha}\right)^{\frac{1}{2}} \sqrt{\Sigma_\alpha}^{-1}$$

**d = 1** :
$$\frac{1}{\sqrt{\sigma_\alpha^2}} \left(\sqrt{\sigma_\alpha^2}\, \sigma_\beta^2 \sqrt{\sigma_\alpha^2}\right)^{\frac{1}{2}} \frac{1}{\sqrt{\sigma_\alpha^2}}$$

$$= \frac{\sigma_\beta}{\sigma_\alpha}$$

---

**Thm** : $\tilde{W}_2^2(\alpha, \beta) = \| m_\alpha - m_\beta \|$

$+ \; \mathcal{B}^2(\Sigma_\alpha, \Sigma_\beta)$ $\quad$ $\mathcal{B}$ Bures' metric.

arises in Quantum mechanics

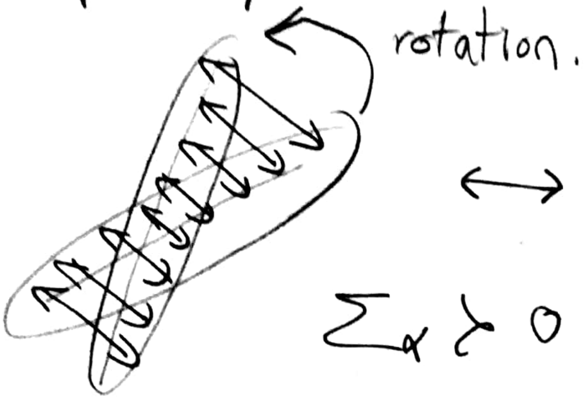
$$\mathcal{B}(\Sigma_\alpha, \Sigma_\beta) = \operatorname{Tr}\left[\Sigma_\alpha + \Sigma_\beta - 2\left(\Sigma_\alpha^{\frac{1}{2}} \Sigma_\beta \Sigma_\alpha^{\frac{1}{2}}\right)^{\frac{1}{2}}\right]$$

20/11/23

(3)

# O.T

Un transport optimal n'est jamais une rotation




$\longleftrightarrow$ transport optimal.

$\Sigma_\alpha \succ 0$ eigh$(\Sigma_\alpha) \neq 0$.

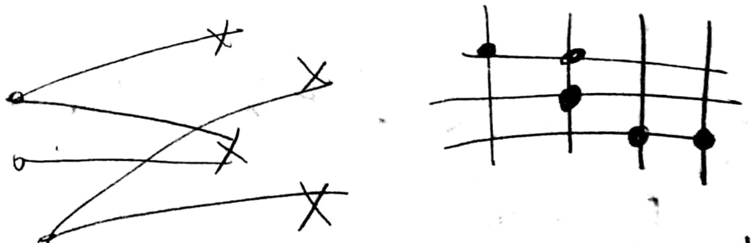
$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ $\longleftrightarrow$ ellipse infiniment fine.

## Kantorovitch

## plan de transport / Coupling

$$\begin{cases} \alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \\ \beta = \sum_{j=1}^{m} b_j \delta_{y_j} \end{cases}$$

$P \in \mathbb{R}_+^{n \times m}$ $\quad P_{ij} > 0$ means we transfert $P_{ij}$ mass from $x_i$ to $y_j$.



structure de graphe bipartie.

Révolution par rapport à Monge.

## Conservation de la masse

$$\forall i \quad \sum_j P_{ij} = a_i, \quad \forall j \quad \sum_i P_{ij} = b_j$$

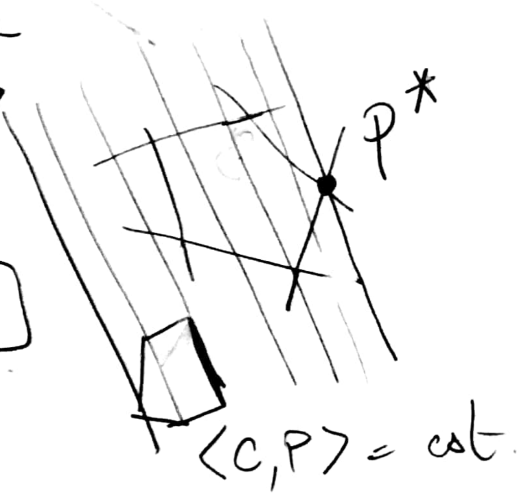
La formulation de Kantorovitch produit ensemble convexe de contraintes.

$$P \geq 0. \quad P \mathbb{1}_m = a \quad P^T \mathbb{1}_n = b \Big) \begin{array}{l} m+n \\ \text{equations.} \end{array}$$

$$\boxed{K} \quad \min_{P \geq 0} \left\{ \underbrace{\sum_{i,j} P_{ij} C_{ij}}_{} \;\middle|\; \begin{array}{l} P \mathbb{1}_m = a \\ P^T \mathbb{1}_n = b \end{array} \right\}$$

Cas $n = m$.
$a = b = \mathbb{1}$

fonction linéaire $= \langle P, C \rangle$




dantzig

Time Complexity

$O(n^3 \log(n)^2)$ $\boxed{n = m}$

$$\boxed{M} \quad \min_{\sigma \in Perm_n} \sum_i C_{i\sigma(i)}$$

$$\Leftrightarrow \min_{P \in \mathcal{P}_n} \langle C, P \rangle \quad \mathcal{P}_n = \text{"permutation matrices".}$$

$$P \in \mathcal{P}_n \Leftrightarrow \exists \sigma \in Perm \text{ tq } P_{ij} = \begin{cases} 1 & \text{si } j = \sigma(i) \\ 0 & \text{sinon.} \end{cases}$$

= a single 1 per row and column.

20/11/23

# O.T

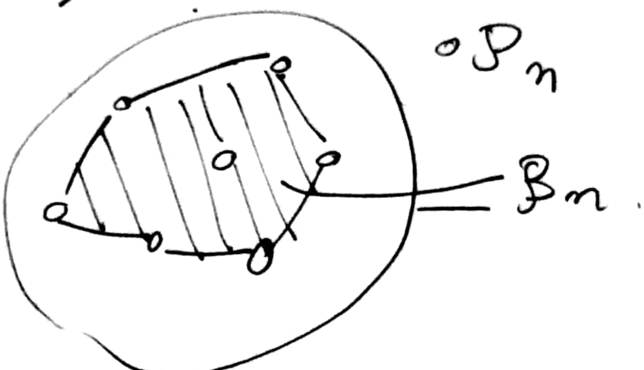
**Kanto** : $P\mathbb{1} = \mathbb{1}$ , $P^T\mathbb{1} = \mathbb{1}$ P matrice bistochastique

$$\mathcal{B}_n \triangleq \left\{ P \in \mathbb{R}_+^{n \times n} \text{ tq } \begin{array}{l} P\mathbb{1} = \mathbb{1} \\ P^T\mathbb{1} = \mathbb{1} \end{array} \right\}$$

On peut reformuler le problème de Kantorovitch

$$\boxed{K} : \min_{P \in \mathcal{B}_n} \langle P, C \rangle$$

$\# \mathcal{P}_n = n!$




On a l'inclusion $\mathcal{P}_n \subseteq \mathcal{B}_n$

**Proposition** : $K \leq M$.

**Théorème** : $K = M$. dans le cas $\begin{array}{l} n = m \\ a = b = \mathbb{1} \end{array}$.

$$C_{ij} = c(x_i, x_j).$$

$$\left\{ \begin{array}{l} P\mathbb{1} = a \\ P^T\mathbb{1} = b \end{array} \right.$$

**Prop** : $\mathcal{U} \triangleq \left\{ P \geq 0 \begin{array}{l} P\mathbb{1} = a \\ P^T\mathbb{1} = b \end{array} \right\}$

$$a \otimes b = a b^T = (a_i b_j)_{ij}.$$

$$\cancel{a \otimes b = \mathcal{U}}.$$

**Lemma** : $\mathcal{U} \neq \emptyset$ is compact.

**Preuve** : $\sum P_{ij} = a_i \quad P_{ij} \geq 0$

$$\Rightarrow 0 \leq P_{ij} \leq a_i$$

**Cor** : Kanto always has at least 1 solution.

**Proof** : Continuity + compacity.

**Def** : $P \in Ext(\mathcal{U}) \Leftrightarrow \left\{ \text{if } \frac{Q+R}{2} = P \Rightarrow P = Q = R, \quad P, Q, R \in \mathcal{U} \right.$.

$$P \notin Ext(\mathcal{U}) \Leftrightarrow \exists\, Q \neq R \quad \left\{ P = \frac{Q+R}{2} \right.$$


**Prop** : If $\mathcal{E}$ compact $Ext(\mathcal{E}) \neq \emptyset$.

**B.V theorem** :

$$Ext(B_n) = P_n.$$

**Thm** : $\underset{P \in \mathcal{U}}{\mathrm{Argmin}} \langle P, C \rangle \cap Ext(\mathcal{U}) \neq \emptyset$. if $\mathcal{U}$ is compact.

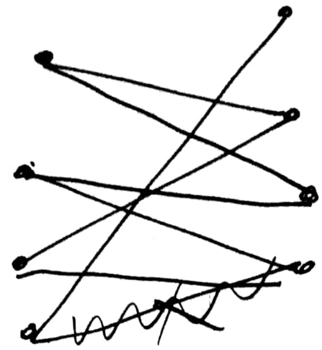
**Corollaires** : Une solution du pb de Kantorovitch appartient à $P_n$.

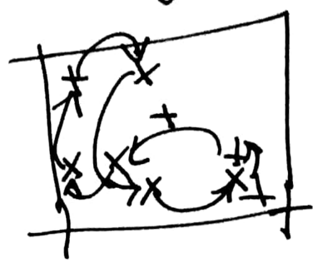
20/11/23

O.T

Adjacency matrix $P_{ij}$

Démonstration de B.V :



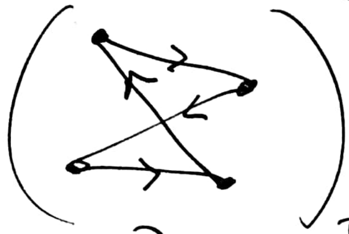


edges st $P_{ij} = 1$ $\quad 0 < P_{ij} < 1$.

are isolated.

Si 2 arêtes sont connectées, alors elles ont un poids $\in (0,1)$.

Soit on a une arête ( •——• ), soit on a un cycle



$\mathcal{P}_n = B_n \cap \{0,1\}^{n \times n}$.

1) $\mathcal{P}_n \subseteq \mathcal{E}xtr(B_n)$

$\mathcal{E}xtr([0,1]) = \{0,1\}$.

$P \in \mathcal{P}_n$.

$$\underset{\{0,1\}}{\overset{\cap}{P_{ij}}} = \frac{\overset{\in\{0,1\}}{Q_{ij}} + \overset{\in\{0,1\}}{R_{ij}}}{2} \Rightarrow P = Q = R.$$

2) $\mathcal{E}xtr(B_n) \subseteq \mathcal{P}_n$. Mq $\mathcal{P}_n^c \subseteq \mathcal{E}xtr(B_n)^c$.

$\Leftrightarrow B_n \setminus \mathcal{P}_n \subseteq B_n \setminus \mathcal{E}xtr(B_n)$

Sq $P \in \mathcal{B}_n$, $P \notin \mathcal{P}_n$.

Mq $\exists Q, R$ $\quad P = \frac{Q+R}{2}$ $\quad Q, R \neq P$ $\in \mathcal{B}_n$.

On se donne un plus petit cycle.

$[i_1 j_1, i_2 j_2 \dots i_K j_K]$ $2K$ edges in the cycle.

$+ \; i_{K+1} j_{K+1}$ $\quad \left\{ \begin{array}{l} i_1 = i_{K+1} \\ j_{K+1} = j_1 \end{array} \right\}$

$$Q_{ij} = \begin{cases} P_{ij} & \text{if } (i j) \notin \text{cycle} \\ P_{ij} - \varepsilon & \text{if } (ij) = (i_{s+1}, j_s). \\ P_{ij} + \varepsilon & \text{if } (i,j) = (i_s j_s) \end{cases}$$

$$R_{ij} = \begin{cases} & \text{———} \\ P_{ij} + \varepsilon & \text{———} \\ P_{ij} - \varepsilon & \text{———} \end{cases}$$

Then $P = \frac{Q+R}{2}$ we recover $P$.

$\varepsilon < \cancel{\varepsilon \leq 1 A} \inf \left\{ \begin{array}{l} P_{ij} \\ 1 - P_{ij} \end{array} \; (i,j) \in \text{cycle} \right\}$

$Q, R \geq 0 \quad Q1 = 1 \quad R1 = 1.$

Scanné avec CamScanner

**Kanto** : $\inf_{\pi} \left\{ \iint c(x,y)\,\pi(dx,dy) \quad \begin{matrix} \pi_1 = \alpha \\ \pi_2 = \beta \end{matrix} \right\}$

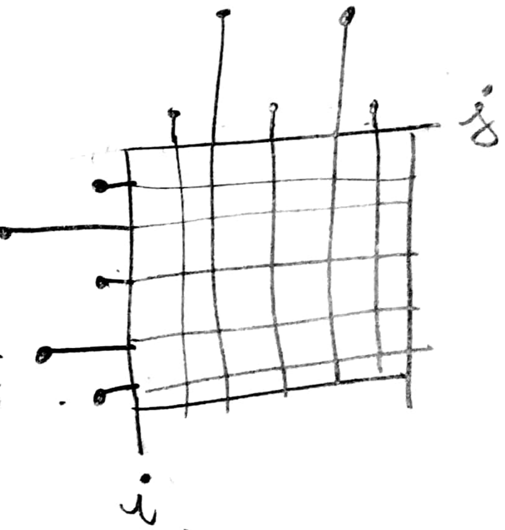
**Cas discret** : $\alpha = \sum a_i \delta_{x_i} \quad \beta = \sum \beta_j \delta_{y_j}$

**Prop** : Si $\pi_1 = \alpha$, $\pi_2 = \beta$

$$\operatorname{supp} \pi \subseteq \left\{ (x_i, y_j)_{\substack{1 \le i \le n \\ 1 \le j \le m}} \right\}$$

**Prop** : Si $\pi_1 = \alpha$, $\pi_2 = \beta$ Alors $\exists P \in \mathbb{R}_+^{n \times m}$

$$\pi = \sum_{i,j} P_{ij}\, \delta_{(x_i, y_j)}$$




$$\begin{cases} \pi_1 = \alpha \iff \sum_j P_{ij} = a_i \\ \pi_2 = \beta \iff \sum_i P_{ij} = b_j \end{cases}$$

$$\iint c\, d\pi = \sum C_{ij} P_{ij} = \langle C, P \rangle$$

$$\int f\, d\mu = \langle f, \mu \rangle.$$

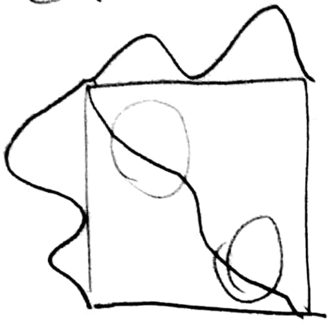
# Interpretation probabiliste

$X \sim \alpha$, $Y \sim \beta$. $\qquad \mathbb{P}(X \in A) = \alpha(A)$.

$$\int f(x)\,\alpha(dx) = \mathbb{E}(f(X))$$

Un couplage correspond à la loi du couple $(X,Y)$

On note $\pi$ la loi du couple.



$$\boxed{K} \equiv \inf_{(X,Y) \text{ v.a}} \left\{ \mathbb{E}(d(X,Y)) \;\; \begin{matrix} X \sim \alpha \\ Y \sim \beta \end{matrix} \right\}$$

$(X,Y)$ a la loi de $\pi$

**Remarque** : On définit $\alpha \otimes \beta \in \mathcal{M}_+(\mathcal{X} \times \mathcal{Y})$

$$\iint f(x,y)\, d(\alpha \otimes \beta(x,y)) = \int \left( \int f(x,y)\, \alpha(dx) \right) \beta(dy)$$

$$(\alpha \otimes \beta)(A \times B) = \alpha(A)\, \beta(B)$$

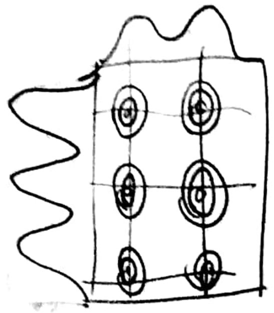
**Cas discret** $\Leftrightarrow$ $P_{ij} = a_i b_j \qquad P = ab^T$

**Proposition** : $\alpha \otimes \beta \in \mathcal{C}(\alpha, \beta) \stackrel{\Delta}{=} \left\{ \pi \;\middle|\; \begin{matrix} \pi_1 = \alpha \\ \pi_2 = \beta \end{matrix} \right\}$

Rmq : $(X,Y) \sim \pi = \alpha \otimes \beta$

$$\mathbb{P}((X,Y) \in A \times B) = \mathbb{P}(X \in A)\, \mathbb{P}(Y \in B)$$

Le transport optimal repose sur une dépendance forte entre les marginales. Le couplage optimal est loin de la mesure produit.



Topologie faible weak$^*$ $\Leftrightarrow$ convergence en loi

1) $\pi \longmapsto \iint c \, d\pi$ continue si $c$ est continue

2) Si $\mathcal{X}, \mathcal{Y}$ sont compacts, $\mathcal{C}(\alpha, \beta)$ est compact.

$\Rightarrow$ existence de solution au problème de Kantorovitch.

## Espace de Wasserstein

$\mathcal{X} = \mathcal{Y}$, $c(x,y) = d(x,y)^p$ / $p \geq 1$, $d(x,y)$ distance

$d(x,y) = d(y,x)$ , $d(x,y) = 0 \Leftrightarrow x = y$

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

**Ex:** $\mathcal{X} = \mathbb{R}^d$ $d(x,y) = \|x - y\|$

**Déf:** $\mathcal{W}_p(\alpha, \beta) = \left[ \inf_{\pi \geq 0, \pi_1 = \alpha, \pi_2 = \beta} \iint d(x,y)^p \, \pi(dx, dy) \right]^{\frac{1}{p}}$

distance de Wasserstein.

Théorème : $W_p$ est une distance

$\Leftrightarrow \quad W_p(\alpha, \beta) = W_p(\beta, \alpha) \geq 0$

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

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

Discrete case

Let's step back a little and understand what $W_p$ corresponds to

Ex : $\mathcal{X} = \mathbb{R}^d$, $d(x, y) = \|x - y\|$.

$\pi(x, y)$ optimal $(\alpha, \beta) \Leftrightarrow \pi(y, x)$ optimal $(\beta, \alpha)$

$$\left[\alpha = \sum a_i \delta_{x_i}, \quad \beta = \sum b_k \delta_{x_k}\right]$$

Si on a $x_i, y_i$ on pose

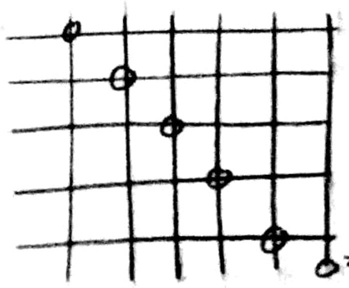
$x \leftarrow (x_i)_i \cup (y_j)_j$ et on met des zéros.


$$\gamma = \sum_j c_j \delta_{x_j}$$

$D_{ij} := d(x_i, x_j)$

$C_{ij} = D_{ij}$,

On est dans un espace métrique discret (un graphe) $D$



0 sur la diagonale

# Transport optimal



$$\sum P^*_{ik} D^p_{ij} = 0 \qquad W_p(\alpha,\beta)$$

On choisit $P^* = \underset{P}{\operatorname{argmin}} \langle C, P \rangle$

$$\iff P^*_{ik} D^p_{ij} = 0 \qquad P^* \mathbb{1} = a,\ P^{*T} \mathbb{1} = b.$$

$$\iff P^k_{ij} = 0 \quad i \neq j \qquad P^* = \mathrm{diag}(h)$$

$$\iff h = a = b.$$

$$\begin{tikzcd} \cdot \arrow[r, bend left, "T"] \arrow[rr, bend right, "T' \circ T"'] & \cdot \arrow[r, bend left, "T'"] & \cdot \end{tikzcd}$$

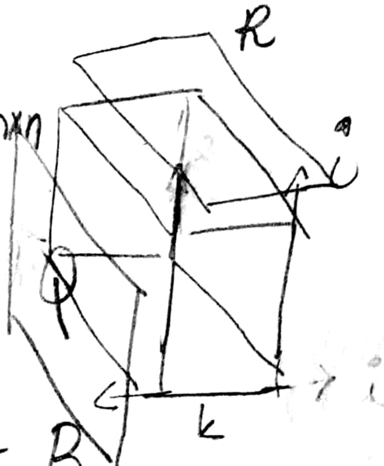
$Q \mathbb{1} = a,\ Q^T \mathbb{1} = b$

$$\begin{tikzcd} a \arrow[r, bend left] & b \arrow[r, bend left] & c \end{tikzcd}$$

$Q \in \mathcal{C}(\alpha,\beta)$ $R \in \mathcal{C}(\alpha,\beta)$

---

**Gluing lemma** : Étant donné $Q, R \in \mathbb{R}^{n \times n}$




$\exists \Sigma \in \mathbb{R}^{n \times n \times n}$

tq $\displaystyle\sum_k \Sigma_{\cdot,\cdot,k} = Q$, $\displaystyle\sum_i \Sigma_{i,\cdot,\cdot} = R$.

$$P = \sum_j \Sigma_{\cdot,j,\cdot} \qquad \begin{cases} \Sigma_{ijk} := \dfrac{Q_{ij} R_{jk}}{b_j} & \text{si } b_j \neq 0 \\ \Sigma_{ijk} = 0 & \text{sinon} \end{cases}$$

Rmq : $P = R @ \mathrm{diag}(1/b) @ Q$

$$W_p(\alpha,\gamma) = \left[\min_{\tilde P} \langle \tilde P, D^p \rangle\right]^{\frac{1}{p}} \leq \langle P, D^p \rangle^{\frac{1}{p}}$$

Soit $Q, R$ optimal entre $\alpha \leftrightarrow \beta$, $\beta \leftrightarrow \gamma$ respectivement

$$W_p(\alpha,\gamma) \leq \left[\sum_{ik} D_{ik}^p \sum_j \underbrace{\frac{Q_{ij} R_{jk}}{b_j}}_{\Sigma_{ijk}}\right]^{\frac{1}{p}} \quad D_{ik} \leq D_{ij} + D_{jk}$$

$$\leq \left[\sum_{ijk} (D_{ij} + D_{jk})^p \Sigma_{ijk}\right]^{\frac{1}{p}}$$

Inégalité de Hölder : $\left(\sum_\ell (a_\ell + b_\ell)^p \lambda_\ell\right)^{\frac{1}{\ell}}$

$\boxed{p \geq 1}$

$$\leq \left(\sum_\ell a_\ell^p \lambda_\ell\right)^{\frac{1}{p}} + \left(\sum_\ell b_\ell^p \lambda_\ell\right)^{\frac{1}{p}}$$

$\|\cdot\|_p$ est une norme.

$$W_p(\alpha,\beta) \leq \left[\sum_{i,j,k} D_{ij}^p \Sigma_{ijk}\right]^{\frac{1}{p}} + \left[\sum_{ijk} D_{jk}^p \Sigma_{ijk}\right]^{\frac{1}{p}}$$

$$\left(\sum_{ij} D_{ij}^p \overbrace{\sum_k \Sigma_{ijk}}^{Q_{ij}}\right)^{\frac{1}{p}} \qquad\qquad \updownarrow$$

$$\| \qquad\qquad\qquad\qquad W_p(\gamma,\beta).$$

$W_p(\alpha,\gamma)$

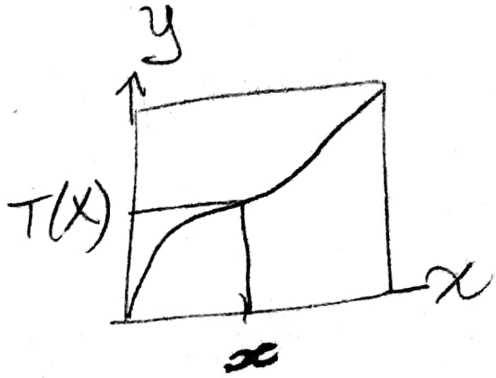
27/11/23 | Transport Optimal | 4

→ Faire les TPs, commencer à écrire le rapport

Théorème de Brenier : $c(x,y) = \|x-y\|^2$

Si $\alpha$ a une densité par rapport à la mesure de Lebesgue alors $T = \nabla\varphi$ $\varphi$ convexe

et $\boxed{M} = \boxed{K}$ . sparsity




$T$, optimal pour Monge
$\Leftrightarrow \pi = (Id, T)\#\alpha$

$(Id, T) : \mathcal{X} \ni x \mapsto (x, T(x)) \in \mathcal{X}\times\mathcal{Y} = \mathbb{R}^d \times \mathbb{R}^d$.

$Range(H) = \{(x, T(x)), x \in \mathbb{R}^d\}$

regarder la preuve du théorème de Brenier.

Topologie du transport optimal

L'espace $\mathcal{X}$ est supposé compact. Faire attention au cas gaussien qui marche malgré tout car la densité est à décroissance rapide.

weak* topology = convergence in law

<u>définition</u> : $\alpha_n \xrightarrow[n\to+\infty]{*} \alpha$

ssi $\left[\forall f \in \mathcal{C}(\mathcal{X}) \quad \int f d\alpha_n \underset{n\to+\infty}{\longrightarrow} \int f d\alpha \right]$

$\langle f, \alpha_n \rangle \underset{n\to+\infty}{\longrightarrow} \langle f, \alpha \rangle$

inner product convergence

<u>Point de vue probabiliste</u> :

$(X_n)_n \xrightarrow{\mathcal{L}} X \Leftrightarrow \mathbb{E}[f(X_n)] \underset{n\to+\infty}{\longrightarrow} \mathbb{E}[f(X)]$

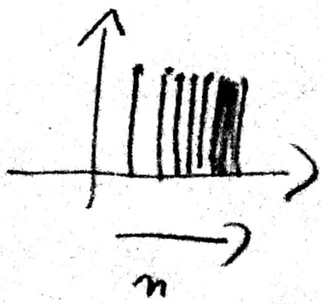
$\forall f \in \mathcal{C}(\mathcal{X})$.

Les $X_n$ peuvent être définis sur des espaces de proba différents $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)_n$.

<u>Thm</u> : $\alpha_n \xrightarrow{*} \alpha \Leftrightarrow W_p(\alpha_n, \alpha) \underset{n\to+\infty}{\longrightarrow} 0$

$W_p$ maîtrise la convergence en loi.

<u>Exemple</u> : $\alpha_n = \delta_{x_n} \longrightarrow \alpha$



Mq $\alpha = \delta_x$, $x \in \mathcal{X}$

$\forall f \in \mathcal{C}(\mathcal{X}) \quad \int f d\alpha_n \longrightarrow \int f d\alpha \qquad \alpha_n \to \delta_x$

$\qquad\qquad\qquad\qquad \overset{\shortparallel}{f(x_n)} \longrightarrow \int f d\alpha. \qquad \Leftrightarrow$

$x_n \to x$.

**Ex :**

$$\alpha = \delta_x, \quad \beta = \delta_y \qquad \mathcal{U}(\alpha,\beta) = \{\delta_{(x,y)}\}$$

$$W_p(\alpha,\beta) = \left[d(x,y)^p\right]^{\frac{1}{p}} = \delta(x,y)$$

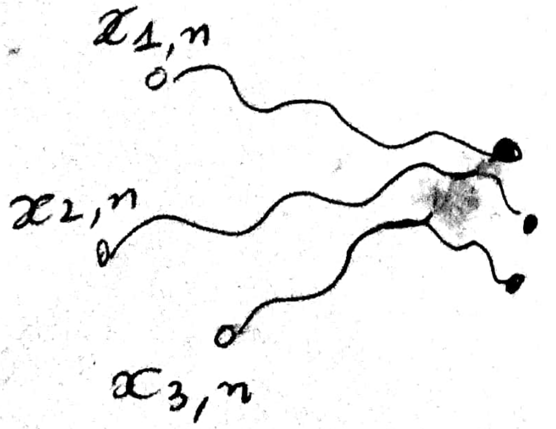
Exemple très simple mais important.

$W_p$ est une distance entre des nuages de points pondérés.

**Théorème :** Si $D(\alpha,\beta)$ est $\begin{cases} \text{Convexe en } (\alpha,\beta) \\ \text{Continue} \\ D(\delta_x,\delta_y) = d(x,y) \end{cases}$

Alors $\exists p \geq 1$ tq $D = W_p$.

Kantorovitch duality



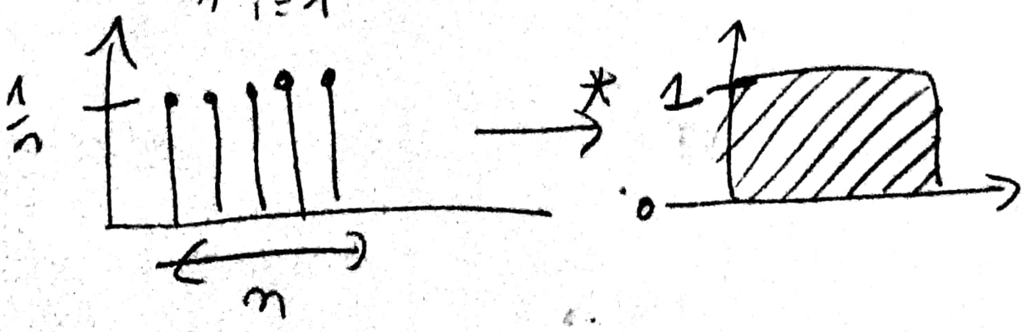

$x_{1,n} \to x_1$

$x_{2,n} \to x_2$

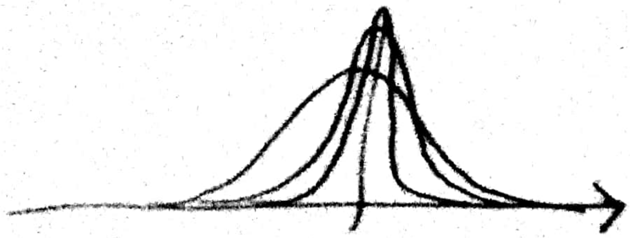
$x_{3,n} \to x_3$

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\frac{i}{n}} \xrightarrow{*} \mathcal{U}([0,1]) \quad \text{car} \quad \frac{1}{n}\sum_{k} f\left(\frac{k}{n}\right) \to \int_0^1 f(x)\,dx$$

somme de Riemann



$$\mathcal{N}(0, \tfrac{1}{n}) \xrightarrow{*} \delta_0$$



$\boxed{p = 2}$ $W_2(\mathcal{N}(m_1, \sigma_1), \mathcal{N}(m_2, \sigma_2))$

$= \| m_1 - m_2 \|^2 + \| \sigma_1 - \sigma_2 \|^2$

$$\sqrt{n}(\overline{X_n} - m) \xrightarrow{*} \mathcal{N}(0, 1)$$

# Optimal Transport



On a introduit la distance de Wasserstein :

$$W_p(\alpha,\beta)^p = \inf_{\pi \geqslant 0} \left\{ \iint d^p(x,y)\, d\pi(x,y) \;\middle|\; \begin{array}{l}\pi_1=\alpha\\ \pi_2=\beta\end{array} \right\}$$

**Théorème** : c'est une distance

et $\alpha_n \xrightarrow[n\to+\infty]{*} \alpha \iff W_p(\alpha_n,\alpha) \xrightarrow[n\to+\infty]{} 0$

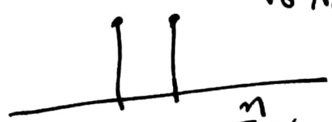
**Théorème Central limite** : $X_1,\dots,X_n$ iid.

$$\Psi_n = \frac{X_1+\dots+X_n}{\sqrt{n}} \qquad \underbrace{\mathbb{E}|X_1|^2}_{=1} < +\infty \qquad \mathbb{E}(X_i)=0$$

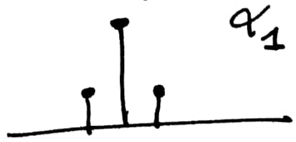
$\Psi_n \xrightarrow{*} N(0,1)$ La loi de $\Psi_n$, $\alpha_n$, converge vers $N(0,1)$.

Chaque $X_i$ peut être définit sur un espace de proba distinct ?

$\alpha_0$ $X_0$



$X_0+X_1$ $\alpha_1$



$X_0+X_1+X_2$ $\alpha_2$

$\sum_{i=1}^n X_i$

$\alpha_n$ :

$n$

Si $X \sim \mathcal{B}(\frac{1}{2})$ Bernoulli.

**Théorème** (Berry-Esseen) si $\mathbb{E}(|X_i|^3) < +\infty$.

$$W_p(\alpha_n, N(0,1)) \leqslant C_p \frac{\mathbb{E}[|X_i|^3]}{\sqrt{n}}$$

Echantillonnage / Méthodes génératives.

en pratique Wasserstein / GAN.

Algorithme d'échantillonnage

densité cible $\alpha(x) \propto e^{-f(x)}$

<u>Langevin</u> : $X_{n+1} = X_n - \tau \nabla f(X_n) + 2\sqrt{\tau}\, \varepsilon_n$

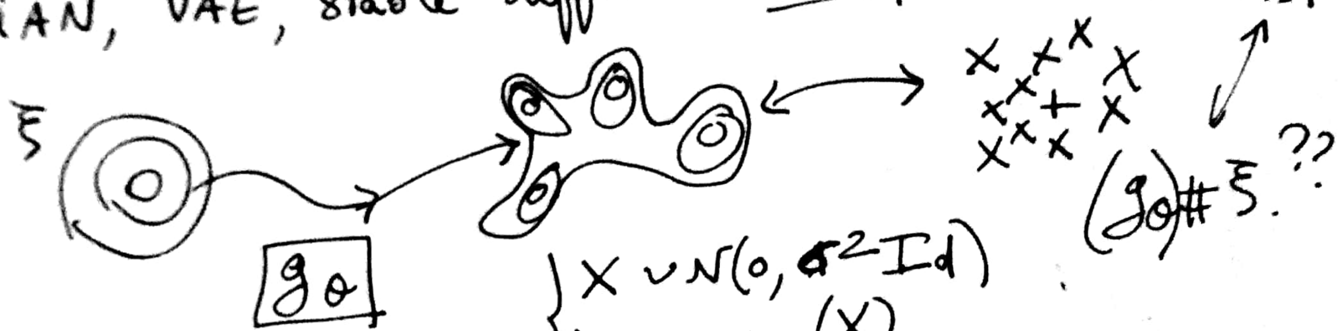
$\varepsilon_n \sim \mathcal{N}(0, Id)$.

$$\begin{matrix}\tau \to 0 \\ n \to +\infty\end{matrix} \qquad X_n \xrightarrow{\mathcal{L}} e^{-f}$$

Dans les modèles génératifs classiques, on dispose d'échantillons $(X_n)_n \sim e^{-f}$ mais $e^{-f}$ n'est pas connu. Comment simuler selon $e^{-f}$.

GAN, VAE, Stable diffusion. <u>Principe</u> :

$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}$




$$\begin{cases} X \sim \mathcal{N}(0, \sigma^2 Id) \\ Y = g_\theta(X). \end{cases}$$

$$\min_{\theta} W_p\big((g_\theta)\#\xi, \hat{\beta}\big)$$

<u>Forte & faible topologies.</u>

<u>Théorème</u> : norme en Variation totale.

$$|\alpha|_{TV} = |\alpha|(\mathcal{X}) = \int d|\alpha|.$$

Si $d\alpha = f(x)\,dx$

alors $|\alpha|_{TV} = \int |f(x)|\,dx.$

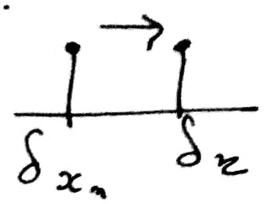
# Transport Optimal

$$\text{Si } \alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \qquad |\alpha|_{TV} = \sum_{i=1}^{n} |a_i| = \|a\|_{\ell^1}.$$

~~Forte~~ **topologie forte** : $\alpha_n \xrightarrow[n\to+\infty]{} \alpha \Leftrightarrow \|\alpha_n - \alpha\|_{TV} \xrightarrow[n\to+\infty]{} 0$

Elle ne permet pas de comparer deux densités discrète et continue.

**Exemple** :



$x \neq x_n$

$\|\alpha - \alpha_n\|_{TV} = 2.$

Pour la topologie forte $\delta_{x_n} \not\to \delta_x$

$\mathrm{Supp}(\alpha_n) \neq \mathrm{Supp}(\alpha) \Rightarrow \|\cdot\|_{TV} = 2$

Si 2 mesures admettent des densités par rapport à des mesures différentes, on dit qu'elles sont mutuellement singulières et dans ce cas $\|\cdot\|_{TV} = 2$.

**Correction de l'exercice 1 :**

$c(x,y) = \mathbb{1}_{x \neq y}. \Rightarrow W_p = \dfrac{\|\cdot\|_{TV}}{2}.$

**ex 2** : $\mathcal{N}(m, \sigma) \xrightarrow{*} \mathcal{N}(m', \sigma') \Leftrightarrow \begin{cases} m \to m' \\ \sigma \to \sigma' \end{cases}$

**ex 3** : $\underbrace{\alpha}_{W} \longmapsto \underbrace{\int f \, d\alpha}_{\mathbb{R}}$ est continue $\Leftrightarrow$ $f$ est continue.

$\pi \longmapsto \int c \, d\pi)$ continue si $c$ est continue.

Rmq : Si $c$ est continue

$\pi \longmapsto \int c \, d\pi$ est continue

$\mathcal{X}$ compacte $\Rightarrow$ $\mathcal{P}(\mathcal{X})$ compacte pour la topologie faible.

continuité + compacité $\rightarrow$ existence de solution de des pb. d'optimisation.

Exo 4.  transport optimal : fonction affine.

Exo pour $W_2$ $\quad \tilde{c}(x,y) = c(x,y) + u(x) + v(y)$

$$\int c \, d\pi = \int \tilde{c} \, d\pi + \text{cste}$$

$$\|x - y\|^2 \longleftrightarrow -\langle x, y \rangle$$

minimiser ~~maximiser~~ la norme $\ell^2$ revient à maximiser la corrélation.

$m_\alpha = 0$

$$W_2^2(\alpha, \beta) = W_2^2(\alpha_0, \beta_0) + \|m_\alpha - m_\beta\|^2$$

$$\alpha \underset{\text{translation}}{\longleftrightarrow} \beta = t\#\alpha \qquad \underset{\text{transport optimal}}{t : x \longmapsto x + \text{cst}}.$$

Sinkhorn.

$$\alpha = \sum_{i=1}^{n} a_i \delta_{x_i} \qquad \beta = \sum_{j=1}^{m} b_j \delta_{y_j} \qquad C_{ij} = c(x_i, y_j)$$

$$P\mathbb{1} = a \qquad P^T\mathbb{1} = b$$

$$\min \langle C, P \rangle \longrightarrow \min_{P \geq 0} \langle C, P \rangle + \varepsilon H(P)$$

température pour Schrödinger.

$P \geq 0$

$$H(P) = \sum_{ij} P_{ij} \log P_{ij}$$

neg-entropie. $0 \log 0 = 0$.

04/12/23 # Transport Optimal



<u>Théorème</u> : $P$ est la solution au problème de Schrödinger ssi

$$\begin{cases} P\mathbb{1} = a, \quad P^T\mathbb{1} = b \\ \exists u, v \in \mathbb{R}_+^n, \mathbb{R}_+^m, \quad P_{ij} = \underbrace{e^{-\frac{C_{ij}}{\varepsilon}}}_{\triangleq K_{i,j} \text{ (Gibbs kernel)}} u_i v_j \end{cases}$$

La démo du théorème repose sur les multiplicateurs de Lagrange.

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

---

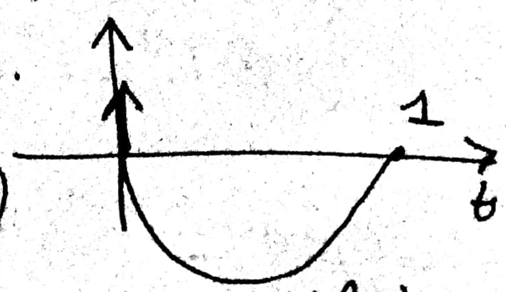
Si $a = b = \mathbb{1}$.

<u>Problème</u> : $K > 0$ matrice carrée tq $K_{i,j} > 0$
Comment normaliser $K$ en une matrice $P$ bistochastique ? C'est toujours possible en divisant les lignes et les colonnes par un paramètre d'échelle.

$$\boxed{\underline{\text{Thm}} \Leftrightarrow \exists (u, v) \mid P = \operatorname{diag}(u) K \operatorname{diag}(v)}$$

<u>Preuve</u> : Montrons d'abord que $P^*_{ij} > 0$, $i,j > 0$
(La solution est $> 0$) Sq $P^*_{ij} = 0$.

$$P_t = (1-t)P^* + t\, ab^T \quad (P^* \text{ solution})$$


$t \mapsto t \log t$.

$$f(t) = \langle P_t, C \rangle + \varepsilon H(P_t)$$

$$f'(0) = -\infty \Rightarrow \exists t,\ f(t) < f(0).$$

$$\inf_{P} \left\{ \sum_{ij} P_{ij} C_{ij} + \varepsilon P_{ij} \log P_{ij} \;/\; \begin{matrix} P\mathbb{1} = a \\ P^T\mathbb{1} = b \end{matrix} \right\}$$

On introduit les variables duales $f$ et $g$ associées aux contraintes d'égalité.

$$\min_{P} \max_{f,g} \underbrace{\varepsilon(P) + \langle P\mathbb{1} - a, f \rangle + \langle P^T\mathbb{1} - b, g \rangle}_{\mathcal{L}(P,(f,g))}$$

$$\nabla_P \mathcal{L} = 0 \Leftrightarrow C + \varepsilon \log P - \varepsilon + f\mathbb{1}^T + \mathbb{1}g^T$$

$$C_{ij} + \varepsilon \log P_{ij} - \varepsilon + f_i + g_j = 0$$

$$P_{ij} = e^{-\frac{C_{ij}}{\varepsilon}} \underbrace{e^{-f_i/\varepsilon}}_{u_i} \underbrace{e^{-g_j/\varepsilon} e^{-\varepsilon}}_{v_j}.$$

$$P = \mathrm{diag}(u) K \mathrm{diag}(v).$$

$$\mathrm{diag}(u) \underbrace{K \underbrace{\mathrm{diag}(v)\mathbb{1}}_{v}}_{Kv} = a$$

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

<u>Algorithme de Sinkhorn :</u>

$v \leftarrow \mathbb{1}.$

$u \leftarrow a/Kv$

$v \leftarrow b/K^T u$

<u>Complexité</u> : $O(n^2 \times T)$

$T$ : nb d'itération

1964 : preuve de convergence.

<u>Thm</u> : $\begin{pmatrix} a \\ b \\ C \end{pmatrix} \mapsto P^*_\varepsilon$

$C^\infty$

# Transport Optimal

## Théorème de Perron-Frobenius

$K > 0 \quad \exists ! \, v^* > 0$ tq $K v^* = v^*$ (point fixe).

$K^n v \to v^*$.

---

$\varphi(u) = Ku. \quad u \to a/Ku \to \dfrac{b}{K^T\left(\frac{a}{Ku}\right)} = \varphi(u).$

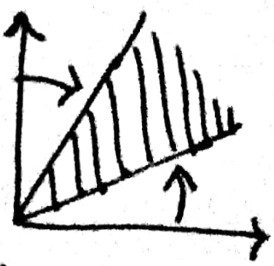
### Hilbert metric

def : $d_H(u, u') := \| \log(u) - \log(u') \|_V$

$\| v \|_V = \max(v) - \min(v)$

Proposition : $(\mathbb{R}^n_+ / \sim, d_H)$ est un espace métrique complet.

$\left( u \sim u' \Leftrightarrow \exists \lambda, \; u = \lambda u' \right.$ ou $\left. u' = \lambda u \right)$



Théorème : (Birkhoff)

$\mathcal{C}_0 = \mathbb{R}^n_+ \qquad K \in \mathbb{R}_{++}^{n \times n}$

$K \mathcal{C}_0$ alors $K$ est une contraction pour $(\mathbb{R}^n_+/\sim, d_H)$ pour la métrique de Hilbert.

$$d_H(Ku, Ku') \leq \eta \, d_H(u, u')$$

$$\eta = \frac{\sqrt{s} - 1}{\sqrt{s} + 1} < 1$$

$$s \hat{=} \max_{ik,l} \frac{K_{ik} K_{jl}}{K_{jk} K_{il}} \geq 1$$

Corollaire :

$d_H(K^t u_0, u^*) \leq \eta^t d_H(u_0, u^*)$

Rmq : $d_H\left(\frac{a}{u}, \frac{a}{u'}\right) = d_H(u, u')$

$$\phi(u) = a/K\left[b/K^T u\right].$$

Cv : $d_H(\varphi^t(u_0), u^*) \leq \eta_\varepsilon^{2t} d_H(u, u^*)$

$$\eta \sim 1 - e^{-\frac{\|C\|_\infty}{\varepsilon}} \underset{\varepsilon \to 0}{\longrightarrow} 1.$$

Régularisation entropique de Kantorovitch:

$$\min_{P\geqslant 0} \left\{ \langle P, C\rangle + \varepsilon \overbrace{\sum_{i,j} P_{ij}\log P_{ij}}^{H(P)},\ \begin{matrix} P\mathbb{1}=a \\ P^T\mathbb{1}=b \end{matrix} \right\}.$$

Def : Kullback-Leibler / Relative entropy.

C'est une "divergence", terme un peu flou.

↳ Bregman divergence "$L^2$"

↳ f-Ciszar divergence "$L^1$"

$$KL(U|V) : \sum_k U_k \log\left(\frac{U_k}{V_k}\right) + \underbrace{U_k - V_k}_{=0 \text{ si } \sum U_k = \sum V_k}.$$

$H(P) = KL(P|\mathbb{1})_{+cst.}$ cas particulier.

$KL(u|v)$ variable de référence (⚠ sauf en maximum de vraissemblance).

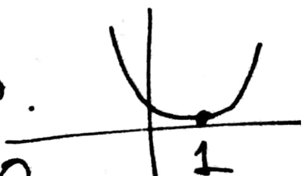
variable d'intérêt, à optimiser.

Convention : ⚠ $0\log 0 = 0$ si $v_k \neq 0$, $u_k = 0$ $\Rightarrow 0 \log 0$.

Si $\exists k \mid v_k = 0,\ u_k \neq 0$

$$KL(u|v) = +\infty.$$

Si $f : \mathbb{R}^+ \to \mathbb{R}^+$ convexe tq $f(1) = 0$. et $f(x) > 0$ pour $x \neq 1$.




La f-divergence de u et v :

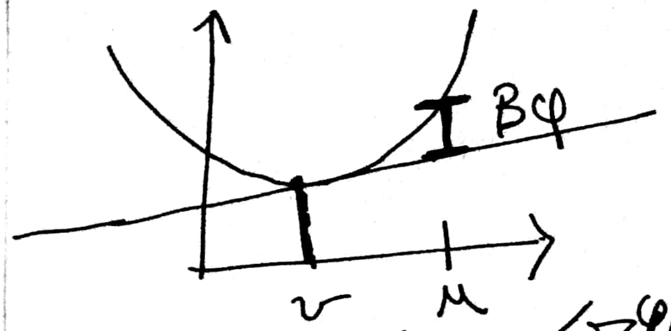
$$D_f(u|v) = \sum_k f\left(\frac{u_k}{v_k}\right) v_k.$$

Proposition : $D_f(u|v) \geqslant 0$, $D_f(u|v) = 0 \Leftrightarrow u = v$.

$D_f(u|v)$ est convexe sur $\mathbb{R}^n_+ \times \mathbb{R}^n_+$.

**Rmq** : Si $f(t) = t\log(t) + t - 1$ on retrouve KL.

**Bregman - Divergence** : $\varphi : \mathbb{R}^d \to \mathbb{R}$ strictement convexe.



$B\varphi$ approximation de Taylor.

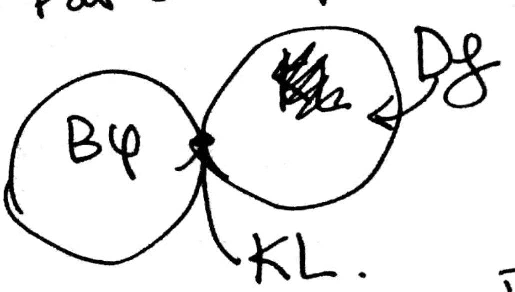
$f(u) = f(v) + \langle \nabla f(v), u - v \rangle$

$B_\varphi(u|v) \triangleq \varphi(u) - \varphi(v) - \langle \nabla\varphi(v), u - v\rangle \geqslant 0$ par construction.

$KL = B\varphi = f$.

$B\varphi(u|v) = 0 \Leftrightarrow u = v$.

⚠ Pas de symetrie.



$\varphi(x) = \frac{\|x\|_2^2}{2}$

$B\varphi(x|y) = \frac{\|x-y\|_2^2}{2}$.

$f(t) = |t-1|$, $D_f(x|y) = \|x - y\|_1$

$|\sqrt{t} - 1|^2$ ? Hellinger distance.

$\varepsilon \sum_{ij} P_{ij} \frac{C_{ij}}{\varepsilon} + \varepsilon P_{ij} \log P_{ij}$

$- \log(e^{-\frac{C_{ij}}{\varepsilon}})$

On veut réecrire cette somme comme une divergence KL.

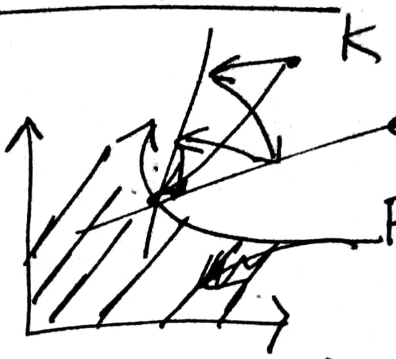
$\min \varepsilon \sum_{ij} P_{ij} \log\left(\frac{P_{ij}}{K_{ij}}\right) + \frac{P_{ij} - K_{ij}}{cst.}$

$\varepsilon \min_{P \geqslant 0} \{ KL(P|K) : P\mathbb{1} = a, P^T\mathbb{1} = b \}$.

11/12/23 . **Transport Optimal** (9)

KL - projection.

**Interpretation :**




$$\mathrm{Proj}_C^{KL/B\varphi}(\varphi) = \underset{P\in C}{\mathrm{argmin}}\, KL(P|\varphi) . \qquad \varphi(t) = \|t\|^2$$

**Projection on convex set (POCS) Bregman**

$\mathcal{C}_1, \mathcal{C}_2$ ensembles convexes.

$$P_0 \leftarrow \text{init}$$

$$\begin{cases} P_{2k+1} = \mathrm{Proj}_{\mathcal{C}_1}^{B\varphi}(P_{2k}) \\ P_{2k+2} = \mathrm{Proj}_{\mathcal{C}_2}^{B\varphi}(P_{2k+1}) . \end{cases}$$

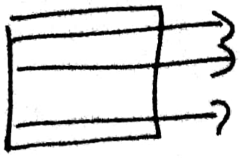
**Théorème** : $P_k \xrightarrow[k\to+\infty]{} P^* \in \mathcal{C}_1 \cap \mathcal{C}_2$.

$$\min_{P\in C_1\cap C_2} KL(P|K) \qquad C_1 = \{P \geq 0, \; P\mathbb{1} = a\}, \quad C_2 = \{P \geq 0, \; P^T\mathbb{1} = b\}$$

Si $C_1$ et $C_2$ sont affine : $P_k \to P^* = \mathrm{proj}_{C_1\cap C_2}^{KL}(P_0)$.

KL : ignore la contrainte $\geq 0$.

Il s'agit de l'algorithme de Sinkhorn.



$\sum_j P_{ij} = a_i$

Prop : $Q = \text{Proj}^{KL}_{C^1}(P)$

$$Q_{ij} = \left(P_{ij} / \sum_l P_{il}\right) \times a_i \Leftrightarrow Q = \text{diag}\left(\frac{a}{P\mathbb{1}}\right) P.$$

preuve : multiplicateurs de Lagrange.

Prop : $Q = \text{Proj}^{KL}_{C^2}(P)$

$$Q_{ij} = P \,\text{diag}\left(b / P^T \mathbb{1}\right).$$

POCS : l'itération peut se réécrire

$$P_k = \text{diag}(u_k)\, K \,\text{diag}(v_k)$$

$(u_k, v_k)$ sont les mêmes que pour l'algo de Sinkhorn.

La preuve de convergence de Bregman est très différente de celle utilisant la distance de Hilbert par Birkhoff. $\rightarrow \rho^k \quad \rho - \rho_\varepsilon$ linear rate

Bregman $\rightarrow \frac{C_\varepsilon}{k}$ sublinear rate.

$$\rho_\varepsilon = 1 - e^{-\frac{\|C\|_\infty}{\varepsilon}} \qquad C_\varepsilon = \frac{\|C\|_\infty}{\varepsilon}.$$

# Transport Optimal.

Généralisation à $\alpha \in \mathcal{M}_+(\mathcal{X})$ $c(x,y)$ continue.
$\beta \in \mathcal{M}_+(\mathcal{Y})$

$H(P) \not\to H(\pi)$. $KL(\pi \mid \pi_0)$.

**Definition** : $\pi_0$ fixé.

$$(K_\varepsilon) \quad \inf \left\{ \iint c(x,y)\,\pi(dx,dy) + \varepsilon\, KL(\pi \mid \pi_0) \;\middle|\; \begin{matrix} \pi_1 = \alpha \\ \pi_2 = \beta \end{matrix} \right\}$$

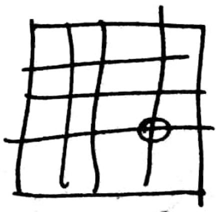
On peut choisir $\pi_0 = \alpha \otimes \beta$ pour se simplifier la vie.

**Proposition** : Si $\begin{cases} \pi_1 = \alpha \\ \pi_2 = \beta \end{cases}$

$$KL(\pi \mid \alpha' \otimes \beta') = KL(\pi \mid \alpha \otimes \beta) - KL(\alpha \mid \alpha') - KL(\beta \mid \beta')$$

**Cas discret** :

$\alpha = \sum a_i \delta_{x_i}$ $\beta = \sum b_j \delta_{y_j}$



$C_{ij}$.

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

$$\min_{P \geq 0} \sum_{ij} P_{ij} C_{ij} + \varepsilon \log\left(\frac{P_{ij}}{a_i b_j}\right) P_{ij}$$

$$\boxed{\begin{aligned} &\underline{\text{def}}: \\ &KL(\pi \mid \pi_0) := \iint \log\left(\frac{d\pi}{d\pi_0}(x,y)\right) d\pi(x,y) + \int d\pi - \int d\pi_0. \end{aligned}}$$

**Déf** : Si $\frac{d\pi}{d\pi_0}$ existe on prend la définition précédente,

sinon $KL(\pi|\pi_0) = +\infty$

Si $\pi = \sum_{ij} P_{ij}\, \delta_{(x_i, y_j)}$ $\quad \pi_0 = \sum_{ij} (P_0)_{ij}\, \delta_{(x_i, y_j)}$

$$KL(\pi|\pi_0) = KL(P|P_0).$$

**Remarque** : $\underbrace{KL(P\,|\,a\otimes b)}_{\text{cas discret.}} = \overbrace{KL(P\,|\,\mathbb{1}\otimes\mathbb{1})}^{H(P).} + \text{cst}.$

$$(K_\varepsilon) : \inf_{\pi} \left\{ \int c\, d\pi + \varepsilon KL(\pi\,|\,\alpha\otimes\beta) \;\middle|\; \begin{matrix}\pi_1 = \alpha \\ \pi_2 = \beta\end{matrix} \right\}$$

Si $\alpha$ et $\beta$ sont des mesures discrètes, le problème est équivalent à : $\min_{P \geq 0} \langle P, C\rangle + \varepsilon H(P)$ $\quad P\mathbb{1} = a,\ P^T\mathbb{1} = b.$

La nouvelle définition du problème est bien cohérente avec celle donnée dans le cas particulier discret.

**Remarque** : Quand $\varepsilon \to 0$, $\pi_\varepsilon \to$ à solut° of Kantorovitch. / $\pi_\varepsilon \to \pi_0 = \operatorname{argmin} \{ KL(\pi|\alpha\otimes\beta),\ \pi$ solution of Kantorovitch $\}$.

**Prop** : $(\varepsilon \to +\infty)$

$$\pi_\varepsilon \xrightarrow[\varepsilon\to+\infty]{} \alpha\otimes\beta$$

**interprétation probabiliste** : $X \sim \alpha$, $Y \sim \beta$

$(X, Y)$ suit $\pi \triangleq \mathcal{L}(X, Y)$

$$I(X,Y) = KL(\pi\,|\,\alpha\otimes\beta)$$

$I(X,Y) = 0 \Leftrightarrow X \perp Y.$

# Transport Optimal.

$\boxed{K_\varepsilon} \Leftrightarrow \mathbb{E}\left(c(X,Y)\right) + \varepsilon I(X,Y)\}$.

$< 10^4$ points : LP possible

$> 10^7$ points : Sinkhorn.

## Problème dual.

$$\boxed{P} \quad \min_{P \geq 0} \{ \langle P, C \rangle : \underbrace{P\mathbb{1} = a}_{f \in \mathbb{R}^n}, \underbrace{P^T\mathbb{1} = b}_{g \in \mathbb{R}^m} \}$$

$$\min_{\substack{P \geq 0 \\ \cap \\ \mathbb{R}^{n \times m}}} \ \max_{f,g \in \mathbb{R}^n \times \mathbb{R}^m} \langle P, C \rangle_{n \times m} + \langle P\mathbb{1} - a, f \rangle_{\mathbb{R}^n} + \langle P^T\mathbb{1} - b, g \rangle$$

On échange min et max ce qui est possible car il s'agit d'un problème linéaire, ce qui implique la dualité forte.

**Théorème.** On peut échanger min et max

$$\sup_{f,g} \langle f, a \rangle + \langle g, b \rangle + \min_{P \geq 0} \langle P, C \rangle - \langle P\mathbb{1}, f \rangle - \langle P^T\mathbb{1}, g \rangle$$

$$\sup_{f,g} \langle f, a \rangle + \langle g, b \rangle + \inf_{P \geq 0} \langle P, \underbrace{C - f\mathbb{1}^T - \mathbb{1}g^T}_{\underbrace{C - f \oplus g}_{= R}} \rangle$$

$$\inf_{P \geq 0} \langle P, R \rangle = \begin{cases} 0 \text{ si } R \geq 0 \\ -\infty \text{ otherwise (if } \exists i \; R_{ij} < 0) \end{cases}$$

$$\sum_{ij} P_{ij} R_{ij}$$

cela nous donne la contrainte $C - f \oplus g \geq 0$.

$\boxed{D}$ $\sup_{f,g} \{\langle f, a \rangle + \langle g, b \rangle, \; f \oplus g \leq C\}$

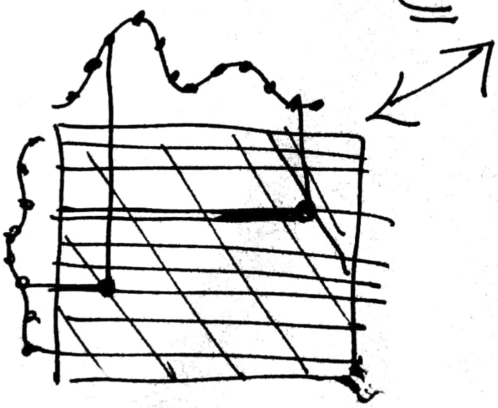
$f$ = prix de départ
$g$ = prix d'arrivée.

<u>Théorème</u> : sup = max

Si $f$ et $g$ sont solutions
alors $(f + c, g - c)$ est solution.

$(P^*, f^*, g^*)$ solution non unique

$$\mathrm{supp}(P^*) = \{(i,j) \mid P^*_{ij} \neq 0\}$$
$$\subseteq \{(i,j) \mid f^*_i + g^*_j = C_{ij}\}$$



$C - f \oplus g \geq 0$.

Scanné avec CamScanner

① **Transport optimal** 18/11/23.

$$\min \langle \tilde{f}, a\rangle + \langle \tilde{g}, b\rangle : \tilde{f}_i + \tilde{g}_j \leq C_{ij} \}$$

$$\sup_{\substack{f\in \mathcal{C}(\mathcal{X})\\ g\in \mathcal{C}(\mathcal{Y})}} \int f(x)\,\alpha(dx) + \int g(y)\,\beta(dy) : f(x)+g(y) \leq c(x,y).$$

$$\alpha = \sum_i a_i \delta_{x_i} \qquad \beta = \sum_j b_j \delta_{y_j} \qquad \tilde{f}_i = f(x_i),\ \tilde{g}_j = g(y_j).$$

$$\mathrm{Supp}(P^*) \subseteq \{(i,j) \mid f_i^* + g_j^* = C_{ij}\}$$

$$\mathrm{Supp}(\pi^*) \subseteq \{(x,y) \mid f^*(x) + g^*(y) = c(x,y)\}.$$

<u>Minimisation alternée</u> On fixe $f$

$$\sup_{f\in \mathcal{C}(\mathcal{Y})} \left\{ \int g(y)\,\beta(dy) : g(y) \leq c(x,y) - f(x) \ \forall (x,y). \right\}$$

$$\forall y,\ g(y) \leq \underbrace{\min_x c(x,y) - f(x)}_{\text{c-transform of } f.}$$

$$\triangleq f^c(y).$$

<u>Def</u>: c-transform of $f(x)$:

$$f^c(y) = \inf_x c(x,y) - f(x).$$

$$f \in \mathcal{C}(\mathcal{X}) \xmapsto[c(x,y)]{} f^c \in \mathcal{C}(\mathcal{Y}).$$

3/02/2023

Scanné avec CamScanner

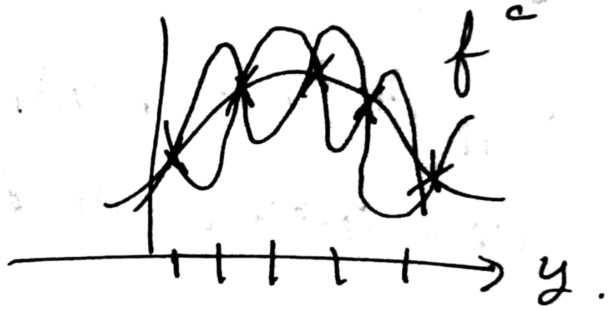
**Déf** : $\bar{c}$-transform. $g(y)$

$$g^{\bar{c}}(x) = \inf_y c(x,y) - g(y).$$

minimisation // $y$ :

$$\sup_{g \in \mathcal{C}(y)} \left\{ \int g(y) \beta(dy) \,,\; g(y) \leq f^c(y) \right\}$$

**Prop** : Si $f$ est fixée.
Tout $g$ optimum vérifie $g(y) = f^c(y)$ $\beta$.a.e.




$c(x,y) = \|x-y\|^2$

$\hookrightarrow c(x,y) = -\langle x,y \rangle$.

$$f^c(y) = \inf_x -\langle x,y\rangle - f(x).$$

$$= \cancel{\inf_x \langle x,y\rangle + f}$$

$$= - \sup_x \langle x,y\rangle - (-f(x))$$

$$= -(-f)^*(x) = f_*(x)$$

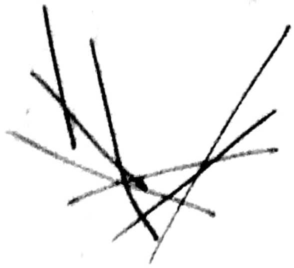
**Déf** : Transformée de Fenchel-Legendre.

$$f^*(y) = \sup_x \langle x,y\rangle - f(x)$$

$f^*$ est toujours convexe. Toute fonction convexe est une transformée de Fenchel Legendre.

18/12/93. **Transport Optimal** ②

Si $f$ est convexe, $f^{**} = f$.



$f^{***} = f^{*}$.

**Théorème** : ① Pour $c$ continue, arbitraire.

$$f^{c\bar{c}c} = f^{c}$$

② Si $c$ est $L$-lipschitz alors $f^{c}$ est $L$-lipschitz.

$f^{c}$ prend la régularité de $c$.

**Def** : Si $f$ tq $\exists g$ $f = g^{\bar{c}}$.

Alors $f$ est $c$-concave.

$$\sup \left\{ \int f \, d\alpha + \int g \, d\beta \quad f \oplus g \leq c \right\}$$
$$f, g \in \mathcal{C}(X) \times \mathcal{C}(Y)$$

$\in$ $c$-concave.

Solution $c$-Lipschitz $\cap$ Bounded fonction

Ascoli $\longrightarrow$ Compact for $L^{\infty}$.

**Corollaire** : Les solutions duales ~~sont~~ existent lipschitzienne. et sont lipschitz.

**Def** : Si $f : (X, d_X) \to \mathbb{R}$ est Lipschitz. $d(f(x), f(y)) \leq L\,d(x,y)$.

$\Leftrightarrow$ $\forall (x,y)$, $d(f(x), f(y))$

$$L^* = \inf_{f(x) \neq f(y)} \frac{d_X(x,y)}{d_Y(f(x), f(y))}$$

Cas où $d_Y = \|\cdot\|$. $\mathrm{Lip}(f) = \sup_x \|\partial f(x)\|$.

Dans le cas où $c(x,y) = -\langle x,y\rangle$.

$(f,g)$. alternate coordinate descent.

$\hookrightarrow (f, f^c) \to (f^{c\bar{c}}, f^c) \to (f^{cc}, f^{ccc})$

$f^{ccc} = f^c$

$$\boxed{f^c = -(-f)^*}$$

Ne pas faire d'optimisation alternée si la fonction objectif des variables à minimiser n'est pas $\mathcal{C}^1$. Ici on est bloqué dans un cycle après 3 itérations.

$\sup \{ \langle f, \alpha\rangle + \langle g, \beta\rangle \quad f \oplus g \leq c \}$

<u>Primal-Dual</u> $(x,y) \in \mathrm{Supp}(\pi) \Rightarrow f(x) + g(y) = c(x,y)$

$$-(-f^c)^*(x) + f^c(y) = -\langle x,y\rangle.$$

$\varphi \triangleq -f^c$ est convexe

$$\varphi^*(x) + \varphi(y) = \langle x,y\rangle \quad (x,y) \in \mathrm{Supp}(\pi).$$

<u>Lemme</u> : (Young) $\varphi^*(x) + \varphi(y) \geq c(x,y)$.

$$\varphi^*(x) + \varphi(y) = \langle x,y\rangle \Leftrightarrow x \in \partial\varphi(y).$$

$x = \nabla\varphi(y)$.

On vient de montrer le théorème de Brenier.

**Thm** : Si $\varphi$ est convexe alors $\varphi$ est différentiable presque partout par rapport à la mesure de Lebesgue.

La transformée de c montre que les solutions duales ne sont pas quelconques mais assez régulières.

$X = Y$.

$c(x,y) = d(x,y) \rightarrow W_1 = \sup\{\langle f, \alpha\rangle + \langle g, \beta\rangle$ tq $f \oplus g \leq d\}$

$d$ distance

**Prop** : $\{f : d\text{-concave}$
$\exists g, f = g^d\}$

$= \{f : Lip(f) = 1\}$.

$= \{f : \|\nabla f\|_\infty \leq 1\}$.

Si $f$ est 1-lipschitz $\Rightarrow f^c = -f$.

$$\sup_g \langle \underset{\overset{\|}{-f^c}}{f^{cc}}, \alpha\rangle + \langle f^c, \beta\rangle$$

$$\sup_{f \in \mathcal{C}} \langle f^c, \beta - \alpha\rangle = \int \underset{\overset{\|}{g}}{f^c(y)}\, d(\beta - \alpha)(y).$$

$$(\alpha,\beta) = \sup_{Lip(g)\le 1} \int g(y)\, d(\beta-\alpha)(y)$$

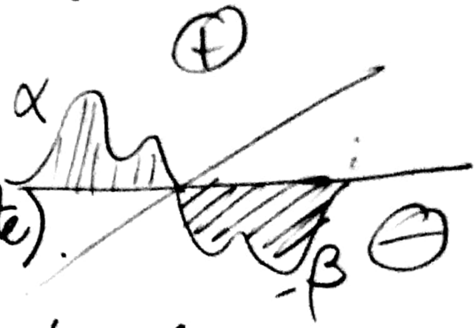
$$= \sup_{Lip(f)\le 1} \int f(x)\, d(\alpha-\beta)(x).$$

$W_1(\alpha,\beta)$ est une norme $= \|\alpha-\beta\|_{W^2}$.

<u>Normes duales</u>. $B$ un ensemble de fonctions.

1) $f \in B \Leftrightarrow -f \in B$.

~~$\|f\|$~~ $\|\alpha-\beta\|_B = \sup_{f\in B} \int f(x)\, d(\alpha-\beta)(x)$.




$B = \{f, \|f\|_\infty \le 1\}$.

$\to \|\alpha-\beta\|_{TV}$. (topologie forte)

$B = \{f, Lip(f)\le 1\} \to W_1$ (topologie faible)

<u>Rmq</u>: for $\|\cdot\|_B$ to be a norm

$B$ boule d'une semi-normee.

$B = \{f, Lip(f) + \|f\|_\infty \le 1\}$.

Unbalanced optimal transport.

$B$ = espace des kernel functions

$= \{ f \mid \|f\|_{RKHS} \le 1\} \longrightarrow \|\alpha-\beta\|_{MMD}$.

18/12/23 # Transport Optimal



$$\min_{\theta} \|\hat{\beta} - \alpha_\theta\|_B .$$

$$\min_{\theta} \max_{f \in B} \int f \, d\alpha_\theta - \int f \, d\hat{\beta}$$

$$\alpha_\theta \triangleq g_{\theta \#} \delta$$

$$\min_{\theta} \max_{\phi} \int f_\phi \, d\alpha_\theta - g_\phi \, d\hat{\beta}$$

$$= \int f_\phi(\underbrace{g_\theta(\varepsilon)}_{x}) \, d\delta(\varepsilon)$$

$$- \frac{1}{N} \sum_{j=1}^{N} f_\phi(y_j)$$

$$\min \max \; xy$$

$\boxed{\text{Régularisation entropique.}}$

$$\inf_{\substack{\pi_1 = \alpha \\ \pi_2 = \beta}} \iint c \, d\pi + \varepsilon \, KL(\pi | \alpha \otimes \beta)$$

$$\inf_{\pi} \sup_{f,g} \iint c \, d\pi + \varepsilon \, KL(\pi | \alpha \otimes \beta)$$
$$+ \langle f, \alpha - \pi_1 \rangle$$
$$+ \langle g, \beta - \pi_2 \rangle .$$

$$\sup_{f,g} \int f \, d\alpha + \int g \, d\beta + \inf_{\pi} \langle \pi, C - f \oplus g \rangle$$
$$+ \varepsilon \, KL(\pi | \alpha \otimes \beta).$$

$$\sup_{f,g} \int f\,d\alpha + \int g\,d\beta - \varepsilon \iint \exp\left(\frac{f(x)+g(y)-c(x,y)}{\varepsilon}\right) \alpha(dx)\beta(dy) + \varepsilon.$$

en minimisant sur $\pi$.

<u>Minimisation alternée</u> : $f$ fixé. $g$ optimal.

$$g(y) = -\varepsilon \log \int e^{\frac{f(x)-c(x,y)}{\varepsilon}} \alpha(dx) \cancel{\beta(dy)}.$$

$$\min{}_\varepsilon(Z) = -\varepsilon \log \int \exp\left(\frac{-Z(y)}{\varepsilon}\right) d\beta(y)$$

$$\min{}_\varepsilon(Z) \to \min_{\mathrm{Supp}(\beta)}(Z) \quad ?$$

$$g(y) = \min_x{}_\varepsilon \left[c(x,y) - f(x)\right]$$

$$= f^{c,\varepsilon}(y).$$

<u>Prop</u> : Algo de minimisation alternée.

$$\boxed{\begin{aligned} f &\leftarrow g^{c,\varepsilon} \\ g &\leftarrow f^{c,\varepsilon} \end{aligned}}$$

"soft c-transform".

# Transport Optimal



<u>Proposition</u> : Take $\begin{cases} u(x) = e^{-\frac{f(x)}{\varepsilon}} \\ v(y) = e^{-\frac{g(y)}{\varepsilon}} \end{cases}$

→ give back Sinkhorn.

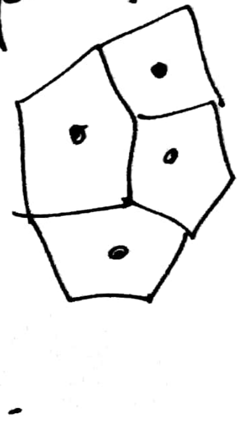
$$-\varepsilon \log \int - \frac{Z(x)}{\varepsilon} \alpha(dx) \Big]$$ log-sum-exp trick.

$$\min_{\varepsilon}(Z) = \min_{\varepsilon}(Z - c) + c$$

avec $c \triangleq \min_0 (Z)$

<u>Numerics</u>: extended Sinkhorn.

## Semi-discrete optimal transport



<u>Thm</u> : $c(x,y) = |x-y|^2$.

$\alpha$ mesure "continue"

↳ $\beta$ mesure discrète.

$$V_i = \{ x \mid f_i + \|x - x_i\|^2 \leq f_j + \|x - x_j\|^2 \}$$

extended Voronoi cell.

$f \neq 0 \longrightarrow$ Laguerre cell / Power diagram.

Thm Dudley 60's $\mathbb{E}[|W_2(\alpha,\beta) - W_2(\hat{\alpha},\hat{\beta})|]$

hyp : $\left| \begin{matrix} x_i \\ y_j \end{matrix} \overset{i.i.d}{\sim} \right| \begin{matrix} \alpha \\ \beta \end{matrix}$ $\leq O\left(\frac{1}{n^{1/d}}\right) C(\alpha,\beta)$

$$\begin{cases} \hat{\alpha}_N = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i} \\ \hat{\beta}_N = \frac{1}{N}\sum_{j=1}^{N} \delta_{y_j}. \end{cases}$$

malédiction de la ~~dimensionalité~~ dépend de la dimension.

$$\mathbb{E}\left[\,\left|\|\alpha-\beta\|^2 - \|\hat{\alpha}-\hat{\beta}\|^2\right|\,\right] \sim \frac{C(\alpha,\beta)}{\sqrt{N}}.$$

régularisation entropique.

$$W_2^{\varepsilon}(\alpha,\beta) \xrightarrow{\varepsilon \to 0} W(\alpha,\beta)$$

$$\Pi^{\varepsilon} \longrightarrow \alpha \otimes \beta$$

$$\iint c(x,y)\, \alpha(dx)\beta(dy)$$

Thm (Genevay)

$$\mathbb{E}[|W^{\varepsilon}(\alpha,\beta) - W^{\varepsilon}(\hat{\alpha},\hat{\beta})|] \leq \frac{C}{\sqrt{N}} \varepsilon^{-\frac{d}{2}}.$$