Cours M2Ri :

CV II, thm lim,

& transport optimal

Séance 14

15 Novembre 2021

III - Remarques autour de la dualité de Kantorovich & conséquences.

① Coût vs surplus

Remarquons que si $c(x,y) \leq a(x) + b(y)$, $(a,b) \in L^1(\mu) \times L^1(\nu)$

Alors $s(x,y) := a(x) + b(y) - c(x,y)$ "Surplus"

satisfait

$$\inf_{\pi \in \Pi(\mu,\nu)} \int c(x,y)\, \pi(dx,dy) = \mu(a) + \nu(b) - \sup_{\pi \in \Pi(\mu,\nu)} \int s(x,y)\, \pi(dx\,dy)$$

⇝ Minimiser un coût est équivalent à maximiser un surplus. Et la dualité de Kantorovich passe trivialement à cette formulat°.

<u>Lemme</u> :

$$\sup_{\pi \in \Pi(\mu,\nu)} \int s(x,y)\, \pi(dx,dy) = \inf_{f(x) + g(y) \geq s(x,y)} \mu(f) + \nu(g)$$

<u>Preuve</u>:

$$\sup_{\pi \in \Pi(\mu,\nu)} \int s(x,y)\, \pi(dx\,dy)$$

$$= \mu(a) + \nu(b) - \inf I(\pi)$$

Dualité de K.

$$\mu(a) + \nu(b) - \sup_{(\varphi,\psi)\in\Phi_c} \mu(\varphi) + \nu(\psi)$$

$$= \inf \mu(\underbrace{a-\varphi}_{f}) + \nu(\underbrace{b-\psi}_{g})$$

avec $f(x) + g(y) = a(x) - \varphi(x) + b(y) - \psi(y)$

$$\geq a(x) + b(y) - c(x,y) = s(x,y)$$

$\boxtimes$

② c-concavité et l'astuce de double convexificat°

Def [c-concavité]

$c : \mathcal{X}\times\mathcal{Y} \longrightarrow \mathbb{R}\cup\{+\infty\}$ coût arbitraire

Une fonct° $\varphi : \mathcal{X} \longrightarrow \mathbb{R}\cup\{-\infty\}$ est dite c-concave s'il existe $\psi : \mathcal{Y} \longrightarrow \mathbb{R}\cup\{-\infty\}$, $\psi \not\equiv -\infty$ (non identiquement $-\infty$) tq

$$\varphi(x) = \inf_{y\in\mathcal{Y}} c(x,y) - \psi(y)$$

On note $\varphi = \psi^c$

Exemple important : $\mathcal{X} = \mathcal{Y} = \mathbb{R}^n$

$$c(x,y) = \frac{1}{2}\|x-y\|^2 = \frac{\|x\|^2}{2} + \frac{\|y\|^2}{2} - \underbrace{\langle x,y\rangle}_{s(x,y)}$$

* Si $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}\cup\{-\infty\}$,

Alors $$\varphi^c(y) = \inf_{x\in\mathbb{R}^n} \frac{1}{2}\|x-y\|^2 - \varphi(x)$$

$$= \inf_{x \in \mathbb{R}^n} \frac{1}{2}\|x\|^2 - \varphi(x) + \frac{\|y\|^2}{2} - \langle x, y \rangle$$

$$\Longleftrightarrow \quad \underbrace{\left(\varphi^c - \frac{\|\cdot\|^2}{2}\right)}(y) = \inf_{x \in \mathbb{R}^n} - \langle x, y \rangle - \underbrace{\left(\varphi - \frac{\|\cdot\|^2}{2}\right)}(x)$$

$$\Longleftrightarrow \quad f^c(y) = \inf_{x \in \mathbb{R}^n} - \langle x, y \rangle - f(x)$$

C'est la concavité habituelle.

$$-f^c(y) = \sup_{x \in \mathbb{R}^n} \langle x, y \rangle - (-f)(x)$$

est la transformée convexe de Legendre.

* Transformee coût en surplus renverse concavité et convexité

$$\sup_{\pi \in \Pi(\mu,\nu)} \int \langle x, y \rangle \, \pi(dx, dy) = \inf_{f(x) + g(y) \geqslant \langle x, y \rangle} \mu(f) + \nu(g)$$

$$\overset{\text{astuce de convexificat°}}{=} \inf_{f} \mu(f) + \nu(f^*)$$

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

transformée de Legendre.

Passons maintenant à l'astuce de double convexification (en fait 2x concavificat° mais ce mot est laid...) vue dans

La preuve de la dualité de Kantorovich.
(Etape 2, c borné) :

Dans l'égalité $\inf I(\pi) = \sup_{(\varphi,\psi)\in\Phi_c} \mu(\varphi)+\nu(\psi)$

on peut restreindre l'optimisat° à

$(\varphi^{cc}, \varphi^c)$.

En effet $\mu(\varphi) + \nu(\psi)$ avec $\varphi(x)+\psi(y) \leqslant c(x,y)$

$\leqslant \mu(\varphi) + \nu(\varphi^c)$

c'est juste ça ! $\leqslant \mu(\varphi^{cc}) + \nu(\varphi^c)$

Le lemme suivant montre qu'il est inutile de répéter l'opérat°.

Lemme : $\forall \varphi : \mathcal{X} \to \mathbb{R}\cup\{-\infty\}$, $(\varphi^{cc})^c = \varphi^c$

En particulier φ c-concave $\iff \varphi = \varphi^{cc}$

Preuve : * Mq $\varphi^{cc} \geqslant \varphi$

Par def $\varphi^c(y) = \inf_{x\in\mathcal{X}} c(x,y) - \varphi(x)$

donc $\varphi(x) + \varphi^c(y) \leqslant c(x,y) \quad \forall x, y$

De même $\varphi^{cc}(x) = \inf \underbrace{c(x,y) - \varphi^c(y)}_{\geqslant \varphi(x)}$

d'où $\varphi^{cc} \geqslant \varphi$.

* Mq $(\varphi^{cc})^c = \varphi^c$

On a déjà $(\varphi^{cc})^c \geqslant \varphi^c$ (par le point précédent)

Ensuite $\varphi^{ccc}(y) = \inf c(x,y) - \overbrace{\varphi^{cc}(x)}^{\leq -\varphi}$

$\leqslant \inf c(x,y) - \varphi(x) = \varphi^c(y)$

Donc $\varphi^{ccc} \leqslant \varphi^c$.

Ensemble, les deux inégalités donnent $\varphi^{ccc} = \varphi^c$ ☒

③ Cas des distances $c = d$

On retrouve la distance de Kantorovich-Rubinstein (dans une version non bornée).

Thm [Kantorovich-Rubinstein]

Soit $\mathcal{X} = \mathcal{Y}$ espace polonais muni d'une distance d. et

$$W_1(\mu,\nu) := \inf_{\pi \in \Pi(\mu,\nu)} \int d(x,y)\, \pi(dx\, dy)$$

Alors $$W_1(\mu,\nu) = \sup_{\|\varphi\|_{Lip} \leq 1} |\mu(\varphi) - \nu(\varphi)|$$

(et non $\|\varphi\|_{BL} = \|\varphi\|_\infty + \|\varphi\|_{Lip} \leq 1$)

Preuve: Par approximat°, on peut supposer d bornée.

On a :

$$W_1(\mu,\nu) \overset{\substack{\text{dualité de K.}\\ \downarrow}}{=} \sup_{\varphi(x)+\psi(y)\leq d(x,y)} \mu(\varphi) + \nu(\psi)$$

$$\underset{\substack{\text{double convexificat°}\\ \text{(coût borné)}}}{=} \sup_{\varphi \text{ bornée}} \mu(\varphi^d) + \nu(\varphi^{dd})$$

Or $\varphi^d(y_1) = \inf_{x\in X} d(x,y_1) - \varphi(x)$ est 1-Lip

(En effet $\varphi^d(y_1) \overset{IT}{\leq} d(y_1,y_2) + \inf_{x\in X} d(x,y_2) - \varphi(x)$

$= d(y_1,y_2) + \varphi^d(y_2)$)

$\leadsto$ $$-\varphi^d(x) \underset{\varphi^d \text{ 1-Lip}}{\leq} \varphi^{dd}(x) = \inf_y d(x,y) - \varphi^d(y)$$

$$\underset{x=y}{\leq} -\varphi^d(x)$$

$\leadsto$ $$-\varphi^d = \varphi^{dd}$$

Donc $$W_1(\mu,\nu) = \sup_{\varphi \text{ bornée}} \mu(\varphi^d) - \nu(\varphi^d)$$

φ^d 1-Lip $\to$ $$\leq \sup_{\|\varphi\|_{Lip}\leq 1} \mu(\varphi) - \nu(\varphi)$$

φ 1-Lip $\Rightarrow (\varphi,-\varphi)\in\Phi_d$ $\to$ $$\leq \sup_{(\varphi,\psi)\in\Phi_d} \mu(\varphi)+\nu(\psi) = W_1(\mu,\nu)$$

⚠ Nul besoin de vous faire la preuve que W_1 métrise la CV faible !

(Au détail près que $\|.\|_{Lip} \longleftrightarrow \|.\|_{BL}$).

④ Cas discret

$\mathcal{X} = \{x_1, \dots, x_m\}$ $\qquad \mu = (\mu_1, \dots, \mu_m) \in \mathbb{R}_+^m$

$\mathcal{Y} = \{y_1, \dots, y_n\}$ $\qquad \nu = (\nu_1, \dots, \nu_n) \in \mathbb{R}_+^n$

$$\Pi(\mu,\nu) = \left\{ P \in M_{m,n}(\mathbb{R}) \;\middle|\; \begin{array}{l} P_{ij} \geq 0 \\ \sum_{j=1}^{n} P_{ij} = \mu_i \\ \sum_{i=1}^{m} P_{ij} = \nu_j \end{array} \right\}$$

$$= \left\{ P \in M_{m,n}(\mathbb{R}_+) \;\middle|\; P\mathbb{1}_n = \mu,\ P^T \mathbb{1}_m = \nu \right\}$$

$$\mathbb{1}_n = \underbrace{(1, \dots, 1)}_{n \text{ fois}}$$

Coût = Matrice de coûts C

$$= \left(C_{ij} \,,\, \begin{array}{l} 1 \leq i \leq m \\ 1 \leq j \leq n \end{array} \right)$$

La dualité suivante de Kantorovich est très intéressante du point de vue computationnel :

$$\begin{array}{ll} \min\limits_{\substack{P \in M_{m,n}(\mathbb{R}_+) \\ P \mathbb{1}_n = \mu \\ P^T \mathbb{1}_m = \nu}} \langle C, P \rangle & = \inf\limits_{\pi \in \Pi(\mu,\nu)} I(\pi) \\ \quad \| & \quad \| \text{ dualité de K.} \\ \max\limits_{\substack{(f,g) \in \mathbb{R}^m \times \mathbb{R}^n \\ f_i + g_j \leq C_{ij}}} \langle \mu, f \rangle + \langle \nu, g \rangle & = \sup\limits_{\varphi(x) + \psi(y) \leq c(x,y)} \mu(\varphi) + \nu(\psi) \end{array}$$

$P \in M_{m,n}(\mathbb{R}_+)$ est bien plus grand à explorer que les $(f,g) \in \mathbb{R}^m \times \mathbb{R}^n$

Clairement une résolut° algorithmique devrait préférer le point de vue dual !

IV. Aspects computationnels.

Comme référence très complète pour les aspects computationnels du transport optimal, nous recommandons

" Computational Optimal Transport ", Peyré, Cuturi (ArXiv 1803.00567)

① Problème historique

D'un point de vue numérique, seul le cas discret est formalisable et programmable sur un ordinateur digital.
Et nous sommes alors confronté à résoudre :

$$\underbrace{\min_{\substack{P\mathbb{1}_n=\mu\\ P^T\mathbb{1}_m=\nu}} \langle C,P\rangle}_{\text{Primal}} = \overbrace{\max_{f_i+g_j\leq C_{ij}} \langle \mu,f\rangle+\langle \nu,g\rangle}^{\text{dual}} =: \boxed{W_C(\mu,\nu)}$$

↳ 1 seul problème d'optimisat°

⇝ Optimisat° linéaire sous contraintes : Explorat° des points extrêmes du polytope des contraintes comme enseigné dans un cours d'analyse numérique (Chap 2 Cuturi Peyré)

② Point de vue moderne : Regularisat° entropique.

L'entropie discrete pour $P \in M_{m,n}(\mathbb{R}_+)$:

$$H(P) = \begin{cases} -\sum_{ij} P_{ij}(\log P_{ij} - 1) & \text{si } P_{ij} > 0 \ \forall ij \\ -\infty & \text{sinon} \end{cases}$$

Il est facile de voir que H est 1-fortement concave car

$$\frac{\partial^2 H}{\partial P} = -\mathrm{diag}(1/P_{ij})$$

$$\xRightarrow{P_{ij} \leq 1} \quad \langle Q, \frac{\partial^2 H}{\partial P^2} Q \rangle \leq -\|Q\|^2$$

1-forte concavité.

Par conséquent une idée est de considérer le problème ε-régularisé :

$$W_C^{\varepsilon}(\mu, \nu) = \inf_{\substack{P \in M_{m,n}(\mathbb{R}_+) \\ P \mathbb{1}_n = \mu \\ P \mathbb{1}_m = \nu}} \underbrace{\underbrace{\langle C, P \rangle}_{\text{linéaire}} - \varepsilon H(P)}$$

$\mathcal{E}$ convexe sur le compact $\Pi(\mu, \nu)$ donc solut° unique P_ε.