Cours M2Ri :

CV II, thm lim,

& transport optimal

Séance 15

22 Novembre
2021

La minimisat° du problème régularisé a de meilleures propriétés. En effet, la convexité donne de meilleures garanties pour les approches différentielles.

De plus, les minimiseurs pour $\varepsilon \to 0$ (et $\varepsilon \to \infty$) se comportent très bien.

Propositi° : L'unique solut° P_ε de ⊛ converge

* $\lim_{\varepsilon \to 0} P_\varepsilon = \underset{P}{\text{Argmin}} \left\{ -H(P) \mid P \in \Pi(\mu,\nu),\ \langle P, C\rangle = W_C(\mu,\nu) \right\}$

solut° optimale pour le problème de Kantorovich $\underline{\underline{\text{et}}}$ d'entropie maximale.

* $\lim_{\varepsilon \to +\infty} P_\varepsilon = \mu \otimes \nu$ couplage indépendant

Par conséquent P_ε interpole entre deux régimes naturels.

Preuve : Comme $\Pi(\mu,\nu)$ compact (au sens strict! Pas de Prokhorov ici !), on peut supposer $P_\varepsilon \to P^*$ convergent.

Mq la limite est unique et comme annoncée

+ $\varepsilon \to +\infty$: Par optimalité de P_ε, on a

$$\frac{1}{\varepsilon}\langle C, P_\varepsilon\rangle - H(P_\varepsilon) \leq \frac{1}{\varepsilon}\langle C, P\rangle - H(P)$$

$\forall P \in \Pi(\mu, \nu)$

En passant à la limite, $-H(P^*) \leq -H(P)$

$\forall P$.

Donc $P^* = \underset{P \in \Pi(\mu,\nu)}{\operatorname{Argmin}} -H(P) \overset{\text{exo ?}}{\underset{\text{connu}}{=}} \mu \otimes \nu$

unique limite donc $\lim P_\varepsilon = \mu \otimes \nu$.

* $\varepsilon \to 0$: Comme P_ε minimiseur de $W_C^\varepsilon(\mu,\nu)$

$\forall P \quad \langle C, P_\varepsilon\rangle - \varepsilon H(P_\varepsilon) \leq \langle C, P\rangle - \varepsilon H(P)$

donc par limite :

$\forall P, \quad \langle C, P^*\rangle \leq \langle C, P\rangle$

donc $P^* = \underset{P \in \Pi(\mu,\nu)}{\operatorname{Argmin}} \langle C, P\rangle$ Minimiseur pour $W_C(\mu,\nu)$

Maintenant on se restreint aux minimiseurs $P \in \Pi(\mu,\nu)$

$$0 \leq \langle C, P_\varepsilon\rangle - \langle C, P\rangle \leq \cancel{\varepsilon}\left(H(P_\varepsilon) - H(P)\right)$$

P minimiseur

Donc $H(P^*) \geq H(P) \rightsquigarrow P^*$ minimiseur d'entropie maximale.

(unique).

$\blacksquare$

③ Algorithme de Sinkhorn.

Il est aisé d'analyser la structure du minimiseur P_ε par analyse différentielle :

Propositᵒ : Posons $K = \left(K_{ij} = e^{-\frac{C_{ij}}{\varepsilon}} \right)_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$

La solut° P_ε de $(\star)$ est unique et de la forme

$$[P_\varepsilon]_{ij} = u_i \, K_{ij} \, v_j$$

où $(u, v) \in \mathbb{R}_+^m \times \mathbb{R}_+^n$ sont deux vecteurs uniques.

On écrira aussi $P_\varepsilon = \operatorname{diag}(u) \, K . \operatorname{diag}(v)$.

Preuve : Ecrivons l'optimisat° sous contraintes avec un Lagrangien

$$\mathcal{L}(P, f, g) = \underbrace{\langle C, P \rangle - \varepsilon H(P)}_{\text{Fonct° objectif}} \underbrace{- \langle f, P \mathbb{1}_m - \mu \rangle - \langle g, P^T \mathbb{1}_n - \nu \rangle}_{\text{Contraintes}}$$

Multiplicateurs de Lagrange

Point critique du Lagrangien :

$$\frac{\partial \mathcal{L}}{\partial P_{ij}}(P, f, g) = C_{ij} + \varepsilon \log P_{ij} - f_i - g_j = 0$$

$$\iff P_{ij} = \exp\left(\frac{f_i + g_j - C_{ij}}{\varepsilon}\right)$$

$$\iff P = \Big[\underbrace{e^{\frac{f_i}{\varepsilon}}}_{u_i} K_{ij} \underbrace{e^{\frac{g_j}{\varepsilon}}}_{v_j}\Big]_{i,j}$$

$\blacksquare$

Rmk: Les u_i, v_j sont des $\exp\left(\frac{\cdot}{\varepsilon}\right)$ de multiplicateurs de Lagrange.

Similairement au point de vue dual de Kantorovich : $P_{ij} \leq 1 \iff \exp\left(\frac{f_i + g_j - C_{ij}}{\varepsilon}\right) \leq 1$

$$\iff f_i + g_j \leq C_{ij}$$

Lorsque $\varepsilon \to 0$, il est préférable de saturer les inégalités i.e d'être extrémal.

L'idée de l'algorithme de Sinkhorn est de prendre $P_\varepsilon = \text{diag}(u) . K \ \text{diag}(v)$ puis d'ajuster successivement les marginales pour finir dans $\Pi(\mu, \nu)$:

Algorithme de Sinkhorn :

$u^{(0)} \leftarrow \mu$; $v^{(0)} \leftarrow \nu$

Tant que $P = u^{(\ell)} K v^{(\ell)}$ satisfait

$$\| P\mathbb{1} - \mu \| + \| P^T \mathbb{1} - \nu \| \geq \text{erreur}$$

division terme à terme

$$u^{(\ell+1)} \leftarrow \mu / K v^{(\ell)} \qquad (1)$$

$$v^{(\ell+1)} \leftarrow \nu / K^T u^{(\ell+1)} \qquad (2)$$

Fin tant que

En effet, après (1) :

$$\mathrm{diag}\,(u^{(t+1)})\, K\, \mathrm{diag}\,(v^{(t)})\, \mathbb{1}_m$$

$$= \mathrm{diag}\left(\frac{\mu}{K v^{(t)}}\right) . K v^{(t)} = \mu$$ ⤳ bonne 1ère marginale.

Après (2) :

$$\left[\mathrm{diag}\, u^{(t+1)}\, K\, \mathrm{diag}\,(v^{(t+1)})\right]^T \mathbb{1} = \nu .$$

⤳ Site de simulat°

"Numerical Tours" de Gabriel Peyré

$\boxed{V}$ De Kantorovich à Monge : le thm de Brenier

Vu le temps qu'il reste, nous ne pourrons pas démontrer le thm de Brenier.

Tentons d'énoncer le thm.

Ingrédients : * $(\mu, \nu) \in \mathcal{M}_1(\mathbb{R}^n) \times \mathcal{M}_1(\mathbb{R}^n)$

avec second moment $\int |x|^2 \mu(dx) + \int |x|^2 \nu(dx) < +\infty$

* Coût quadratique : $c(x,y) = |x-y|^2$

⤳ Lien fort avec la convexité usuelle

car c-conjugué $\simeq$ Conjuguée de Legendre usuelle.

Thm :

(i) (Critère d'optimalité de Knott - Smith)

$\pi = \pi^* \in \Pi(\mu, \nu)$ optimal ssi l'un des assert° suivantes est vraie (équivalentes) :

- $\exists \varphi$ convexe semi-continue inférieurement telle que

$$\operatorname{supp}(\pi) \subset \operatorname{Graph}(\underline{\underline{\partial \varphi}})$$

$\hookrightarrow$ sous-gradient d'une fonct° convexe.

(Caractérisat° Prop. 2.4

$x \cdot y = \varphi(x) + \varphi^*(y) \iff y \in \partial\varphi(x) \iff x \in \partial \varphi^*(y)$)

- $y \in \partial\varphi(x)$ pour $d\pi$-presque tout (x, y)

De plus (φ, φ^*) minimiseur du problème

$$\inf \{ \mu(\varphi) + \nu(\varphi^*) \mid \forall (x,y),\ \langle x, y\rangle \leq \varphi(x) + \varphi^*(y) \}$$

(ii) (Thm de Brenier)

Supposons que μ ne donne pas de masse aux petits ensemble

$\underset{\text{def}}{\iff}$ Mesure nulle ~~si~~ dim. de Hausdorff au plus $n-1$.

$\Leftarrow$ μ admet une densité.

Dans ce cas $\exists !\ \pi$ optimal avec $\pi(dx, dy) = \mu(dx) \times \delta_{\nabla\varphi(x)}(dy)$

où $T = \nabla\varphi$ est l'unique (μ-pp) gradient de fonct° convexe tq $T_* \mu = \nu$

De plus $\mathrm{supp}(\nu) = \overline{\nabla\varphi(\mathrm{supp}\,\mu)}$