1.5 Théorème de Fenchel-Rockafellar (208, 219, 229, 253) [4]

Voici un petit (gros !) théorème d'optimisation duale dans les espaces vectoriels normés. Je dis optimisation duale car d'un problème de minimisation dans l'espace vectoriel normé E, on passe à un problème de maximisation dans l'espace dual E', qui a **toujours** une solution ! Ce théorème se base sur **le** gros théorème de convexité en dimension infinie : Hahn-Banach géométrique. Le mieux pour comprendre ce théorème est de se replonger un peu dans l'analyse convexe en dimension infinie et ça peut être un investissement ultra rentable pour les leçons 219, 229 et 253 ! Bref. Avant de parler du théorème il faut introduire un objet : la fonction conjuguée d'une fonction.

Définition 1.11 (Conjuguée d'une fonction)**.** Soient E un espace vectoriel normé et $\varphi : E \longrightarrow (-\infty, +\infty]$ une fonction telle que $D(\varphi) := \{x \in E \mid \varphi(x) < +\infty\}$ ne soit pas vide. On définit alors la fonction conjuguée de φ, notée φ^* ainsi :

$$\begin{array}{rccl} \varphi^* & : E' & \longrightarrow & (-\infty, +\infty] \\ & f & \longmapsto & \sup\limits_{x \in E} \left(f(x) - \varphi(x) \right). \end{array}$$

Théorème 1.12 (Fenchel-Rockafellar)**.** Soient E un espace vectoriel normé, φ et ψ deux fonctions convexes de E dans $(-\infty, +\infty]$ telles qu'il existe $x_0 \in E$ tel que :

$$\varphi(x_0) < +\infty, \quad \psi(x_0) < +\infty, \quad \text{et} \quad \varphi \text{ est continue en } x_0.$$

Alors on a :

$$\inf_{x \in E} \left(\varphi(x) + \psi(x) \right) = \sup_{f \in E'} \left(-\varphi^*(-f) - \psi^*(f) \right) = \max_{f \in E'} \left(-\varphi^*(-f) - \psi^*(f) \right).$$

Démonstration. Dans la preuve de ce théorème, nous aurons besoin du lemme suivant :

Lemme 1.13. Soit $C \subset E$ un ensemble convexe. Alors :

1. $\mathring{C}$ est convexe.
2. Si de plus $\mathring{C} \neq \emptyset$, alors $\overline{C} = \overline{\mathring{C}}$.

Démonstration du lemme.

1. Soient $x, y \in \mathring{C}$ et soit $t \in [0,1]$. Montrons que $z := tx + (1-t)y \in \mathring{C}$. Puisque x et y sont dans $\mathring{C}$, il existe $\varepsilon_x > 0$ et $\varepsilon_y > 0$ tels que :

$$B(x, \varepsilon_x) \subset C \quad \text{et} \quad B(y, \varepsilon_y) \subset C.$$

Posons $\varepsilon := \min(\varepsilon_x, \varepsilon_y)$ et montrons que $B(z, \varepsilon) \subset C$. Soit $a \in B(0, \varepsilon)$. On a alors :

$$z + a = t \underbrace{(x+a)}_{\in C} + (1-t) \underbrace{(y+a)}_{\in C} \in C \quad \text{par convexité de } C.$$

Ainsi, $B(z, \varepsilon) \subset C$ et donc $z \in \mathring{C}$.

2. On a déjà que $\overline{\mathring{C}} \subset \overline{C}$. Montrons le sens réciproque. Soit $x \in \overline{C}$ et soit $y \in \mathring{C}$ (supposé non-vide). On va montrer un résultat plus fort qui est le suivant :

$$[y, x) := \{ty + (1-t)x \mid t \in (0,1]\} \subset \mathring{C}.$$

Soit alors $t \in (0,1]$ et soit $z_t := ty + (1-t)x$. Montrons qu'il existe $\varepsilon > 0$ assez petit tel que :

$$z_t + \varepsilon B(0,1) \subset C.$$

On a :

$$\begin{array}{rcl} z_t + \varepsilon B(0,1) & = & ty + (1-t)x + \varepsilon B(0,1) \\ & \subset & ty + (1-t)\left(C + \varepsilon B(0,1)\right) + \varepsilon B(0,1), \quad \text{car } x \in \overline{C}. \\ & = & (1-t)C + t\left(y + \dfrac{2-t}{t}\varepsilon B(0,1)\right) \\ & \subset & (1-t)C + tC, \quad \text{pour } \varepsilon \text{ assez petit, car } y \in \overset{\circ}{C} \\ & = & C. \end{array}$$

Détaillons peut-être pourquoi $x \in C + \varepsilon B(0,1)$. Puisque $x \in \overline{C}$, alors il existe $x' \in C$ tel que $\|x - x'\|_E < \varepsilon$. Ainsi :

$$x = x' + \underbrace{x - x'}_{\in \varepsilon B(0,1)} \in C + \varepsilon B(0,1).$$

On conclut donc que $x \in \overline{\overset{\circ}{C}}$ et donc $\overline{C} = \overline{\overset{\circ}{C}}$. $\square$

Pour la preuve du théorème, notons a et b les quantités qui nous intéressent :

$$a := \inf_{x \in E} \left(\varphi(x) + \psi(x)\right), \quad \text{et} \quad b := \sup_{f \in E'} \left(-\varphi^*(-f) - \psi^*(f)\right).$$

Par définition de φ^* et ψ^*, on a :

$$\begin{array}{rcl} \forall (x,f) \in E \times E' \quad -\varphi^*(-f) - \psi^*(f) & \leq & -(-f(x)) + \varphi(x) - f(x) + \psi(x) \\ & = & \varphi(x) + \psi(x). \end{array}$$

Ainsi, $\varphi(x) + \psi(x)$ est minoré uniformément en x par $-\varphi^*(-f) - \psi^*(f)$ pour tout $f \in E'$. Ainsi :

$$\forall f \in E', \quad -\varphi^*(-f) - \psi^*(f) \leq a.$$

En passant alors au sup sur E' on a donc :

$$b \leq a.$$

Si $a = -\infty$, la conclusion est immédiate : on a alors $b = -\infty$ et donc :

$$\forall f \in E', \quad -\varphi^*(-f) - \psi^*(f) = -\infty$$

et donc le sup est atteint en tout $f \in E'$. On se place donc dans le cas où $a \in \mathbb{R}$. Puisque $b \leq a$, on veut montrer que $a \leq b$. Pour cela, étant donné qu'on veut également prouver que le sup b est un max, on va montrer qu'il existe $g \in E'$ tel que :

$$-\varphi^*(-g) - \psi^*(g) \geq a.$$

On va prouver un résultat plus fort : on va montrer qu'il existe une constante $\beta \in \mathbb{R}$ tel que :

$$\varphi^*(-g) \leq -\beta, \quad \text{et} \quad \psi^*(g) \leq \beta - a.$$

Au vu de la définition de φ^* et ψ^* cela revient à prouver qu'il existe $\beta \in \mathbb{R}$ tel que :

$$\forall x \in E, \quad -g(x) - \varphi(x) \leq -\beta \quad \text{et} \quad g(x) - \psi(x) \leq \beta - a.$$

i.e.
$$\forall x \in E, \quad \varphi(x) \geq \beta - g(x) \quad \text{et} \quad a - \psi(x) \leq \beta - g(x).$$

On est donc face à un problème de séparation au sens large ! On va donc chercher à appliquer Hahn-Banach géométrique première forme. On cherche donc deux convexes disjoints à séparer, l'un d'entre eux devant être ouvert. On a à disposition des fonctions convexes, dont l'une est continue en un certain $x_0 \in E$. Ainsi, l'épigraphe de φ, que l'on notera C est convexe et d'intérieur non-vide car $(x_0, \varphi(x_0) + 1) \in \overset{\circ}{C}$ par continuité. Prenons alors :

$$A := \overset{\circ}{C}, \quad \text{et} \quad B := \{(x, \lambda) \in E \times \mathbb{R} \mid \lambda \leq a - \psi(x)\}.$$

A est convexe d'après le lemme et B est convexe par convexité de ψ. Vérifions que A et B sont disjoints : si $(x, \lambda) \in A$, alors :
$$\lambda > \varphi(x) \geq a - \psi(x).$$

Ainsi, si $(x, \lambda) \in A$ alors $(x, \lambda) \notin B$. On peut donc appliquer le théorème de Hahn-Banach géométrique : il existe $\Phi \in (E \times \mathbb{R})' \setminus \{0\}$ et $\alpha \in \mathbb{R}$ tel que l'hyperplan $\Phi^{-1}(\{\alpha\})$ sépare A et B au sens large :

$$\forall (x, \lambda) \in A, \quad \Phi(x, \lambda) \geq \alpha \quad \text{et} \quad \forall (x, \lambda) \in B, \quad \Phi(x, \lambda) \leq \alpha.$$

Par continuité de Φ, on a également :
$$\forall (x, \lambda) \in \overline{A}, \quad \Phi(x, \lambda) \geq \alpha.$$

Or, d'après le lemme, $\overline{A} = \overline{C}$. Ainsi, on a en particulier :

$$\forall (x, \lambda) \in C, \quad \Phi(x, \lambda) \geq \alpha \quad \text{et} \quad \forall (x, \lambda) \in B, \quad \Phi(x, \lambda) \leq \alpha.$$

Φ étant une forme linéaire continue, on a :

$$\forall (x, \lambda) \in E \times \mathbb{R}, \quad \Phi(x, \lambda) = \Phi(x, 0) + \lambda \Phi(0, 1).$$

Ainsi, Φ s'écrit sous la forme :
$$\Phi(x, \lambda) = f(x) + c\lambda$$

où $f \in E'$ et $c \in \mathbb{R}$. On peut donc réécrire les conditions de séparation ainsi :

$$\forall (x, \lambda) \in C, \quad f(x) + c\lambda \geq \alpha \quad \text{et} \quad \forall (x, \lambda) \in B, \quad f(x) + c\lambda \leq \alpha.$$

Puisque $\varphi(x_0) < +\infty$, pour tout $\lambda \geq \varphi(x_0)$, $(x_0, \lambda) \in C$. Ainsi, si c était strictement négatif, en faisant tendre λ vers $+\infty$ dans l'inégalité :
$$f(x_0) + c\lambda \geq \alpha,$$

on aboutirait à une contradiction ! Ainsi $c \geq 0$. Montrons que $c \neq 0$. Si c était nul, on aurait, étant donné que pour tout $x \in D(\varphi)$ (resp. $D(\psi)$) $(x, \varphi(x)) \in C$ (resp. $(x, a - \psi(x)) \in B$) :

$$\forall x \in D(\varphi), \quad f(x) \geq \alpha, \quad \text{et} \quad \forall x \in D(\psi), \quad f(x) \leq \alpha.$$

Or, $x_0 \in D(\varphi)$ et, puisque φ est continue en x_0, il exite $\varepsilon_0 > 0$ tel que $B(x_0, \varepsilon_0) \subset D(\varphi)$. Ainsi, pour tout $z \in B(0, 1)$:
$$f(x_0 - \varepsilon_0 z) \geq \alpha \quad \text{et donc} f(x_0) \geq \alpha + \varepsilon_0 f(z).$$

En passant au sup sur les $z \in B(0, 1)$, on obtient :

$$f(x_0) \geq \alpha + \varepsilon_0 \|f\|_{E'}.$$

Et, étant donné que $x_0 \in D(\psi)$ également, on a aussi que $f(x_0) \leq \alpha$ et donc :

$$\varepsilon_0 \|f\|_{E'} \leq 0.$$

Ainsi, $f = 0$ et donc Φ est nulle ! **ABSURDE !** En prenant alors $\beta := \dfrac{\alpha}{c}$ et $g := \dfrac{f}{c}$, on a :

$$\forall x \in D(\varphi), \quad \beta - g(x) = \frac{1}{c}\left(\alpha - f(x)\right) \leq \varphi(x) \quad \text{étant donné que } (x, \varphi(x)) \in C.$$

On a donc :

$$\forall x \in D(\varphi), \quad \varphi(x) \geq \beta - g(x),$$

et l'inégalité est trivialement vérifiée dans le cas où $x \notin D(\varphi)$. Enfin :

$$\forall x \in D(\psi), \quad \beta - g(x) = \frac{1}{c}(\alpha - f(x)) \geq a - \psi(x).$$

Ainsi :

$$\forall x \in D(\psi), \quad a - \psi(x) \leq \beta - g(x),$$

l'inégalité étant trivialement vérifiée dans le cas où $x \notin D(\psi)$. On a donc prouvé ce que l'on voulait ! C'est-à-dire :

$$\varphi^*(-g) \leq -\beta \quad \text{et} \quad \psi^*(g) \leq \beta - a.$$

On a donc :

$$b \geq -\varphi^*(-g) - \psi^*(g) \geq a \geq b.$$

On a donc égalité dans cette chaîne d'inégalités et cela termine la démonstration du théorème ! □

Remarque 1.5.1 (Mais qu'est-ce que c'est que ce théorème ? ?)**.** *Ce théorème est un théorème d'optimisation duale/min-max : si vous voulez connaître l'infimum d'une fonction $\varphi + \psi$, regardez plutôt le max de la fonction $-\varphi^*(-\cdot) - \psi^*$, qui existe toujours. Cela sert en transport optimal pour prouver que le problème de transport de Kantorovitch admet une formulation duale ! Le problème de Kantorovitch est le suivant : pour X et Y des espaces polonais (espaces métriques complets séparables), μ et ν des probabilités sur $(X, \mathscr{B}(X))$ et $(Y, \mathscr{B}(Y))$ on cherche, pour une fonction de coût $c : X \times Y \longrightarrow \mathbb{R}^+ \cup \{+\infty\}$ donnée, à minimiser la quantité :*

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

pour $\pi \in \Pi(\mu, \nu)$ l'ensemble des probabilités sur $(X \times Y, \mathscr{B}(X) \otimes \mathscr{B}(Y))$ de marginale selon X égale à μ et de marginale selon Y égale à ν. Le théorème de dualité de Kantorovitch dis donc que, si c est semi-continue inférieurement sur $X \times Y$ et si Φ_c désigne l'ensemble des couples de fonctions $(\varphi, \psi) \in \mathrm{L}^1(\mu) \times \mathrm{L}^1(\nu)$ telles que $\varphi(x) + \psi(y) \leq c(x, y)$ presque partout, alors :

$$\inf_{\pi \in \Pi(\mu,\nu)} \int_{X \times Y} c(x, y) \mathrm{d}\pi(x, y) = \sup_{(\varphi,\psi) \in \Phi_c} \left(\int_X \varphi \mathrm{d}\mu + \int_Y \psi \mathrm{d}\nu \right).$$

La preuve de ce gros théorème de transport optimal repose sur le fait que le sup est le même que l'on regarde des fonctions φ, ψ dans L^1 ou dans $\mathscr{C}_b^0$ et sur le théorème de représentation de Riesz pour les fonctions continues sur X lorsque X est compact : les formes linéaires continues positives sur $\mathscr{C}^0(X)$ s'identifient aux mesures finies positives sur X lorsque X est compact. On prouve donc le résultat d'abord pour des fonctions c continues et pour X et Y compacts avec le théorème de Fenchel-Rockafellar, puis on se ramène au cas général en considérant pour c le sup de fonctions c_n uniformément continues et bornées. On utilise également le théorème de Prokhorov à un moment, qui est un résultat de compacité sur les mesures de probabilité...