CORRECTION DU TD 3 : ANALYSE CONVEXE

COURS D'APPRENTISSAGE, ECOLE NORMALE SUPÉRIEURE, AUTOMNE 2014

Rémi Lajugie
remi.lajugie@ens.fr

1. Exercice 1 : Séparation des convexes compacts

Par compacité du produit $C \times D$, il existe deux points $x \in C$ et $y \in D$ tels que $\|x - y\|_2$ soit minimale. Comme $C \cap D = \emptyset$, on a $A = \frac{y+x}{2}$ qui est un point n'appartenant ni à C ni à D. Un dessin convainc assez rapidement que l'hyperplan orthogonal à la direction vectorielle $(y - x)$ et passant par A est un hyperplan séparateur. Montrons le.
On considère l'hyperplan affine $H_A = A + H$. Cet hyperplan vérifie l'équation $u \in H_A \Leftrightarrow (y - x)^T(u + \frac{y+x}{2}) = 0$.

On doit ensuite vérifier que, $\forall u \in C, (y-x)^T(u-\frac{y+x}{2}) < 0$ et $\forall u \in D, (y-x)^T(u-\frac{y+x}{2}) > 0$. Faisons le pour D, sur C le raisonnement sera analogue.

Il suffit juste de vérifier une propriété du minimiseur y de la distance avec C qui est que tout autre point de l'ensemble D fera un angle obtu avec $x - y$.

Considérons l'ensemble D et montrons tout d'abord que pour tout point $u \in D$, on a $(u-y)^T(x-y) \leq 0$. Supposons, par l'absurde qu'il y ait un point u^* tel que $(u-y)^T(x-y) > 0$. Considérons la fonction univariée $f(t) = \|x - y - t(u - y)\|_2^2$. On a $f'(0) = -2 < 0$ par hypothèse donc la fonction est localement décroissante donc il existe un $1 \geq t > 0$ tel que $\|x - y - t(u - y)\|$ soit plus petit que $f(0) = \|x - y\|_2^2$. Mais $y + t(u - y) \in C$ par convexité ce qui est contradictoire avec la définition de x et y comme minimisant la distance de C à D. Donc $\forall u \in D, (u - y)^T(x - y) \leq 0$. Cela nous permet donc de constater que $\forall u \in D, (y - x)^T(u - \frac{y+x}{2}) = (y - x)^T(u - x - \frac{y-x}{2}) \geq 0$.

Ainsi l'hyperplan sépare bien strictement les deux ensembles convexes.
Il existe une vaste littérature concernant la séparation des ensembles convexes, cf chapitre 11 de [1]. *Au delà des espaces de dimensions finis, la notion de séparation des convexes est une notion très importante qui, permet de fonder correctement l'analyse fonctionnelle (théorèmes de Hahn Banach).*

2. Exercice 2 : Convexité des fonctions usuelles

1) On considère cette fonction sur deux ensembles convexes distincts, les ensembles $A^+ = \{(x, y) \in \mathbb{R}^2, y > 0\}$ et $A^- = \{(x, y) \in \mathbb{R}^2, y < 0\}$.
La convexité se vérifie en calculant la matrice Hessienne de cette fonction.

$$H = \frac{2}{y^3}\begin{pmatrix} y^2 & -xy \\ -xy & x^2 \end{pmatrix}.$$

On remarque ainsi que cette Hessienne est définie positive pour $y > 0$ et négative sinon.

2) Il suffit de remarquer que l'indicatrice vérifie la définition d'une fonction convexe, à savoir que $\forall x, y \in \mathbb{R}^n, \lambda \in [0,1], I_C(\lambda x + (1-\lambda)y)$.

3) Soit $Q \in \mathbb{R}^n$. La forme quadratique f peut s'écrire $\frac{1}{2}x^T(Q+Q^T)x$. La fonction f est donc convexe si et seulement si sa partie symétrique $\frac{Q+Q^T}{2}$ est semi définie positive.

4) L'infimum de deux fonctions convexes n'est clairement pas convexe. Prenons par exemple le sup des fonctions définies pour tout x dans $\mathbb{R}$ par $f_1(x) = x$ et $f_2(x) = -x$. L'inf de ces deux fonctions linéaires est $f(x) = -|x|$ qui n'est pas convexe.

5) Soit I une famille de fonctions convexes. Elle vérifie :

$$\begin{aligned}
\forall i \in I, \forall \lambda \in [0,1], \forall x, y \in \mathbb{R}^n,\ & f_i(\lambda x + (1-\lambda)y) \leq \lambda f_i(x) + (1-\lambda) f_i(y) \\
\text{donc}\ & \sup_{i \in I} f_i(\lambda x + (1-\lambda)y) \leq \sup_{i \in I} \lambda f_i(x) + (1-\lambda) f_i(y) \\
\text{d'où},\ & \sup_{i \in I} \left[f_i(\lambda x + (1-\lambda)y) \right] \leq \lambda \sup_{i \in I} f_i(x) + (1-\lambda) \sup_{j \in I} f_i(y)
\end{aligned}$$

6) On remarque que, $\forall S \in S_+^n,\ \lambda_{\max}(S) = \max_{u \in \mathbb{R}^n} \frac{u^T A u}{u^T u}$. Par la question précédente cette fonction est donc bien convexe.

3. Exercice 3 : Dualité Lagrangienne

7) On commence par introduire les variables de Lagrange $\lambda \in {\mathbb{R}_+}^n$ et $\nu \in \mathbb{R}^m$. On écrit ensuite la fonction associée :

$$L(x, \lambda, \nu) = c^T x - \lambda^T x - \nu^T (Ax - b). \tag{1}$$

La minimisation en la variable primale x conduit à la fonction duale suivante :

$$g(\lambda, \nu) = \begin{cases} b^T \nu \text{ si} A^T \nu - \lambda + c = 0 \\ -\infty \text{ sinon} \end{cases}.$$

8)Comme suggéré dans l'énoncé introduisons une variables auxiliaire (souvent appelées dans la littérature anglophone des "slack variables") $t = \max_i a_i^T x + b_i$. On obtient le problème primal suivant :

$$\begin{aligned}
\min_{x \in \mathbb{R}^n} \quad & t \\
& t = \max_i a_i^T x + b_i.
\end{aligned}$$

Ce problème est également équivalent au problème suivant (on peut s'en convaincre par un dessin) :

$$\min_{x\in\mathbb{R}^n} \quad t$$
$$\forall i, t \geq a_i^T x + b_i.$$

C'est un programme linéaire sous forme canonique.

9) En passant à la variable y, le problème devient équivalent à $\min_{y\in\{-1,1\}^n} y^T W y$. Ou encore de manière équivalente :

$$\min_{y\in\mathbb{R}^n} y^T W y \tag{2}$$

$$t.q,\ \forall i, y_i^2 = 1. \tag{3}$$

Si diag() représente l'operateur créant une matrice diagonale à partir d'un vecteur, on écrit même :

$$\min_{y\in\mathbb{R}^n} y^T W y$$
$$t.q,\ \forall i, \mathrm{diag(y)} y = \mathbf{1}.$$

Le Lagrangien est ainsi : $L(y,\lambda) = y^T W y - \lambda^T(\mathrm{diag}(y)y - \mathbf{1})$. Le dual devient donc :

$$g(\lambda) = \begin{cases} \mathbf{1}^T\lambda, & \text{si } W - \mathrm{diag}(\lambda) \text{ est positive semi définie} \\ -\infty, & \text{sinon} \end{cases}.$$

Références

[1] R Tyrrell Rockafellar. *Convex analysis*. Number 28. Princeton university press, 1997.