CORRECTION DU TD 3 : ANALYSE CONVEXE

COURS D'APPRENTISSAGE, ECOLE NORMALE SUPÉRIEURE, OCTOBRE 2017

Aude Genevay
aude.genevay@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 que pour tout point $u \in D$, on a $(u-\frac{x+y}{2})^T(y-x) > 0$. Supposons, par l'absurde qu'il y ait un point u^* tel que $(u^* - \frac{x+y}{2})^T(y-x) \leq 0$. Considérons la fonction univariée $f(t) = \|x - y - t(u-y)\|_2^2$. On a $f'(0) = -2(x-y)^T(u^*-y)$. Or $(x-y)^T(u^*-y) = (x-y)^T(u^* - \frac{x+y}{2}) + (x-y)^T(\frac{x+y}{2} - y)$. On a donc que $(x-y)^T(u^*-y) > 0$, et par suite $f'(0) < -2$. 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 $\forall u \in D, (u - \frac{x+y}{2})^T(y-x) > 0$.

Ainsi l'hyperplan sépare bien les deux ensembles convexes. La séparation est de plus stricte car, par un raisonnement analogue au précédent, on trouve que y et x minimisent les distances respectives à D et C respectivement.
Il existe une vaste littérature concernant la séparation des ensembles convexes, cf chapitre 11 de [**?**]. *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) Soit I une famille de fonctions convexes. Elle vérifie :

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

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) On remarque que, $\forall S \in S^n,\ \lambda_{\max}(S) = \max_{u \in \mathbb{R}^n} \frac{u^T S u}{u^T u}$. Par la question précédente cette fonction est donc bien convexe car la fonction $g(x, S) = \frac{u^T S u}{u^T u}$ est linéaire en S. On peut aussi déduire de cela que la valeur propre minimale est une fonction concave.

3. Exercice 3 : Dualité Lagrangienne

6) 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 :

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

Le problème dual avec contraintes explicites devient donc :

$$\begin{gathered} \max_{\nu \in \mathbb{R}^n} \quad -b^T \nu \\ \text{tel que } A^T \nu + c \geq 0 \end{gathered}$$

7)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 alors le problème primal suivant :

$$\min_{x\in\mathbb{R}^n, t\in\mathbb{R}} \quad t$$
$$\text{tel que } t = \max_i a_i^T x + b_i.$$

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

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

C'est un programme linéaire sous forme canonique.

8) 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 :

$$\min_{y\in\mathbb{R}^n} y^T W y$$
$$t.q, \ \forall i, y_i^2 = 1.$$

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

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

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

$$g(\lambda) = \begin{cases} -\mathbf{1}^T\lambda, & \text{si } W + \text{diag}(\lambda) \succeq 0 \\ -\infty, & \text{sinon.} \end{cases} \quad .$$

On alors le problème dual suivant (avec contraintes explicites) :

$$\max_{\lambda\in\mathbb{R}^n} \quad -\mathbf{1}^T\lambda$$
$$\text{tel que } W + \text{diag}(\lambda) \succeq 0.$$

Remarque : on peut poser $\lambda = -\lambda_{\min}(W)\mathbf{1}$. Ce point est faisable (car $W - \text{diag}(\lambda_{\min}(W)\mathbf{1}) \succeq 0$) et il nous permet d'obtenir une borne inférieure à la valeur optimale du problème primal :

$$p^\star \geq n\lambda_{\min}(W).$$

9) Notons tout d'abord que nous avons : $x^T A x = \frac{1}{2}(A + A^T)$. De ce fait nous supposons dans la suite que A est symétrique. Notons aussi que ce problème n'est pas convexe (contrainte d'égalités non affines). Cependant la solution de ce problème est connue, la valeur optimale est la valeur propre minimale de A, atteinte pour x étant le vecteur propre de norme 1 correspondant. Regardons maintenant si nous avons dualité forte au sens de Lagrange. Ecrivons maintenant le lagrangien du problème :

$$\mathcal{L}(x,\lambda) = x^T Ax + \lambda(x^T x_1) = x^T(A + \lambda I_n)x - \lambda,$$

où I_n est la matrice identité de taille n. En minimisant par rapport à la variable x on obtient :

$$g(\lambda) = \begin{cases} \lambda_{\min} & \text{si } A + \lambda I_n \succeq 0 \\ -\infty & \text{sinon} \end{cases},$$

où $\lambda_{\min}$ est la valeur propre minimale de A. On en déduit que le problème dual a pour valeur optimale $\lambda_{\min}$. Nous avons donc bien dualité forte.