Éléments de correction pour le TD d'optimisation sous contrainte(s) d'égalité

Exercice 1
Dans tous les cas étudiés, les fonctions et contraintes sont clairement $\mathcal{C}^\infty$, et leurs domaines de définition sont des ouverts.

1. Traité en cours.

2. Le seul point critique de la fonction $(x, y, z) \mapsto x^2+y^2$ est l'origine, qui ne vérifie pas la contrainte, et on peut donc appliquer le théorème des multiplicateurs de Lagrange. Si (x, y) réalise un extremum de $f(x, y, z) = 3x - y$ sous la contrainte $x^2 + y^2 = 5$, il existe un multiplicateur de Lagrange λ tel que le système suivant est vérifié :
$$\begin{cases} 3 = 2\lambda x, \\ -1 = 2\lambda y, \\ x^2 + y^2 = 5. \end{cases}$$
On déduit des deux premières lignes que λ est non nul et que $x = -3y$, et par conséquent $10y^2 = 5$. La fonction f ne peut donc atteindre un extremum local sous contrainte qu'en les deux points suivants : $(-3/\sqrt{2}, 1/\sqrt{2})$ ou $(3/\sqrt{2}, -1/\sqrt{2})$. Comme l'espace de contrainte est compacte et que f est continue, comme précédemment, f atteint nécessairement un minimum et un maximum global à trouver parmi ces deux points. Le calcul des valeurs de f en ces deux points suffit alors à conclure.

3. Le gradient de la fonction $(x, y) \mapsto x + 2y$ ne s'annule jamais, donc on peut donc appliquer le théorème des multiplicateurs de Lagrange. Si (x, y) réalise un extremum de $f(x, y, z) = x^2 + y^2$ sous la contrainte $x + 2y = 6$, il existe un multiplicateur de Lagrange λ tel que le système suivant est vérifié :
$$\begin{cases} 2x = \lambda, \\ 2y = 2\lambda, \\ x + 2y = 5. \end{cases}$$
Les deux premières lignes impliques que $2x = y$, et par suite $x = 1$. Le seul point éventuel d'extremum local sous contrainte de f est donc $(1, 2)$, et le multiplicateur de Lagrange est alors égal à 2. Le Lagrangien $L(x, y, \lambda) = x^2 + y^2 - \lambda(x + 2y - 5)$ est une fonction convexe de (x, y) (quel que soit le signe de λ), donc on peut en conclure que le point obtenu est un point de minimum local strict de f sous la contrainte envisagée.

4. Ici f est définie sur $(\mathbb{R}^+)^2$, strictement positive sur $(\mathbb{R}^+_\star)^2$ et nulle en dehors de cet ensemble (c'est-à-dire sur les deux demis axes de coordonnées). Le gradient de la fonction $(x, y) \mapsto 2x + 3y$ ne s'annule jamais, donc on peut donc appliquer le théorème des multiplicateurs de Lagrange pour des points de l'ouvert $(\mathring{\mathbb{R}^+})^2 = (\mathbb{R}^+_\star)^2$. Si $(x, y) \in (\mathbb{R}^+_\star)^2$ réalise un extremum de $f(x, y, z) = (xy)^a$ sous la contrainte $2x + 3y = 12$, il existe un multiplicateur de Lagrange λ tel que le système suivant est vérifié :
$$\begin{cases} ax^{a-1}y^a = 2\lambda, \\ ax^a y^{a-1} = 3\lambda, \\ 2x + 3y = 12. \end{cases}$$
Comme a, x et y sont non nuls, on peut réécrire le système précédent sous la forme suivante :
$$\begin{cases} \frac{y}{2} = \frac{\lambda}{a(xy)^{a-1}}, \\ \frac{x}{3} = \frac{\lambda}{a(xy)^{a-1}}, \\ 2x + 3y = 12. \end{cases}$$

Sous cette forme, on lit aisément qu'alors $3y = 2x$, et la contrainte impose alors que $x = 3$ et $y = 2$. Pour conclure, on constate que l'espace de contrainte est un segment du quart de plan fermé $(\mathbb{R}^+)^2$, et que f continue atteint donc un minimum et un maximum sur ce segment. Les extrémités de ce segment sont sur les demi-axes, où f vaut zéro, et f est strictement positive sur $(\mathbb{R}^+_\star)^2$, donc son minimum est atteint en ces points. Son maximum est donc nécessairement atteint en un point (x, y) appartenant au fragment de segment contenu dans $(\mathbb{R}^+_\star)^2$, et on peut donc lui appliquer ce qui précède : la seule possibilité est que f atteigne son maximum global (et local) sous contrainte au point $(3, 2)$.

Remarque : On peut aussi calculer la matrice des dérivées secondes par rapport à x et y du Lagrangien $L(x, y, \lambda) = (xy)^a - \lambda(2x + 3y - 12)$ au point $(3, 2)$ (comme la contrainte est linéaire, la matrice est indépendante de λ) :

$$H = \underbrace{a(xy)^{a-2}}_{>0} \begin{pmatrix} (a-1)y^2 & axy \\ axy & (a-1)x^2 \end{pmatrix}$$

puis déterminer le signe du produit scalaire $Hv \cdot v$ où v est un vecteur orthogonal au gradient de la contrainte $(2, 3)$, par exemple $v = (-3, 2)$. Le calcul se déroule ainsi :

$$\begin{pmatrix} -3 & 2 \end{pmatrix} \begin{pmatrix} (a-1)y^2 & axy \\ axy & (a-1)x^2 \end{pmatrix} \begin{pmatrix} -3 \\ 2 \end{pmatrix} = 9(a-1)y^2 + 4(a-1)x^2 - 12axy = -72 \text{ si } (x, y) = (3, 2).$$

Ceci confirme que le point obtenu un maximum local sous contrainte de la fonction f ; mais attention, un tel raisonnement ne permet pas de conclure au niveau global. La méthode de la hessienne est donc ici comme en général plus systématique mais locale, et non globale.

5. Le Lagrangien s'écrit $L(x, y, \lambda) = xy^2 - \lambda(x^2 + 4y^2 - 6)$, et a six points critiques : $(\pm\sqrt{6}, 0, 0)$, $(x = \pm\sqrt{\frac{6}{5}}, x, x/4)$ et $(x = \pm\sqrt{\frac{6}{5}}, x, x/4)$. La contrainte est compacte, donc la fonction f continue est bornée et atteint ses bornes sous cette contrainte. Le gradient de la contrainte s'annule uniquement en l'origine, qui ne vérifie pas la contrainte. Les points où f atteint son extremum sous contraintes se trouvent donc nécessairement parmi les points critiques du Lagrangien. On compare les valeurs obtenues en chaque point pour conclure quant aux extrema globaux : le maximum global est atteint en $(\sqrt{\frac{6}{5}}, \pm\sqrt{\frac{6}{5}})$ et vaut $\sqrt{\frac{6}{5}}^3$; le minimum global est atteint en $(-\sqrt{\frac{6}{5}}, \pm\sqrt{\frac{6}{5}})$ et vaut $-\sqrt{\frac{6}{5}}^3$. En $(\sqrt{6}, 0)$, la fonction atteint un minimum local (nul) car f est positive au voisinage de ce point. En $(-\sqrt{6}, 0)$, la fonction atteint un maximum local (nul) puisque f est négative au voisinage de ce point.

6. Le Lagrangien s'écrit $L(x, y, z, \lambda_1, \lambda_2) = \frac{1}{3}x^3 + y + z^2 - \lambda_1(x + y + z) - \lambda_2(x + y - z)$. Les gradients des deux fonctions de contrainte forment une famille libre.

   Les deux points critiques sont $(1, -1, 0, \frac{1}{2}, \frac{1}{2})$ et $(-1, 1, 0, \frac{1}{2}, \frac{1}{2})$. Ils ne peuvent pas être des extrema globaux puisque lorsque $x \to \pm\infty$, $f(x, -x, 0) \to \pm\infty$; f est non bornée sous la contrainte. Leur étude locale peut être complétée en considérant les dérivées secondes du Lagrangien dans la direction de la contrainte, voir prochain cours.

7. Les gradients des deux fonctions de contrainte forment une famille libre. L'unique point critique du Lagrangien $L(x, y, z, t, \lambda_1, \lambda_2) = x^2 + y^2 + z^2 + t^2 - \lambda_1(x + y - 1) - \lambda_2(z + t)$ est $(1, 1, 0, 0, 2, 0)$. Comme $(x, y, z, t) \mapsto L(x, y, z, t, 2, 0)$ est convexe, ce point est un minimum global sous la contrainte de f.

8. Le Lagrangien s'écrit $L(x, y, z, \lambda_1, \lambda_2) = x^2 + (y-1)^2 + z^2 - \lambda_1(x + y - \sqrt{2}) - \lambda_2(x^2 + y^2 - 1)$. Il n'a pas de point critique. Les gradients des contraintes sont liés aux points de la forme $(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, z)$! En fait la contrainte peut se réécrire $(x = y = \frac{\sqrt{2}}{2})$, et le problème revient donc à optimiser la fonction d'une variable $g(z) = \frac{1}{2} + (\frac{\sqrt{2}}{2} - 1)^2 + z^2$, qui atteint évidemment son minimum global en $z = 0$ est n'est pas majorée.

Exercice 3 : Théorème spectral

1. Ici, c'est la différentielle de F qui est la plus agréable à déterminer :
$$F(x+h) = {}^t xAx + {}^t xAh + {}^t hAx + \underbrace{{}^t hAh}_{\text{terme d'ordre 2}},$$
or ${}^t xAh + {}^t hAx = 2^t xAh = (2Ax \cdot h)$ car A est symétrique et que la transposée d'un scalaire est un scalaire. Le gradient de F est donc égal à $\nabla F(x) = 2Ax$. On peut par exemple remarquer que G est la forme quadratique associée à la matrice identité pour prouver que $\nabla G(x) = 2x$ (c'est un résultat classique).

2. La sphère $\mathbb{S}$ est compacte, et F est une application continue. La fonction F atteint donc ses bornes, minimale et maximale, sur $\mathbb{S}$.

3. Soit x est le point où F admet son minimum sur $\mathbb{S}$. On peut appliquer la méthode des multiplicateurs de Lagrange pour la fonction F sous la contrainte $G = 1$. Il existe un scalaire $\lambda \in \mathbb{R}$ tel que $\nabla F(x) = \lambda \nabla G(x)$, ou encore, en divisant par 2 :
$$Ax = \lambda x,$$
ce qui répond à la question.

Exercice 4 : Convexité et fonction entropie

1. Soient $p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix}$, $q = \begin{pmatrix} q_1 \\ \vdots \\ q_n \end{pmatrix}$ dans $(\mathbb{R}_+^\star)^n$ et t dans $[0,1]$.
$$-H(tp + (1-t)q) = \sum_{k=1}^{n} (tp_k + (1-t)q_k)\ln(tp_k + (1-t)q_k)$$
Or $t \mapsto t\ln t$ est une application convexe de $\mathbb{R}_+^\star$, puisqu'elle est de classe $\mathcal{C}^1$ et que sa dérivée $t \mapsto \ln t + 1$ est strictement croissante. Ceci implique :
$$\begin{aligned} -H(tp + (1-t)q) &\leq \sum_{k=1}^{n} tp_k \ln(p_k) + (1-t)q_k \ln q_k \\ &= t\sum_{k=1}^{n} p_k \ln p_k + (1-t)\sum_{k=1}^{n} q_k \ln q_k \\ &= -tH(p) - (1-t)H(q), \end{aligned}$$
bref, $-H$ est une application convexe sur $(\mathbb{R}_+^\star)^n$. Il reste à observer que E est convexe pour en conclure directement que $-H$ est convexe sur E (en utilisant l'inégalité démontrée précédemment pour des p, q dans E). Pour cela, on constate que si p et q vérifient les conditions d'appartenance à E, tout barycentre les vérifie également.

2. On montre que f est strictement croissante en déterminant le signe de sa dérivée :
$$f'(x) = \sum_{k=1}^{n} \underbrace{(a_k - a)^2}_{\geq 0} \underbrace{e^{(a_k - a)x}}_{>0}.$$
Comme il existe au moins un a_k différent de a (sinon tous les a_k seraient égaux), f' est strictement positive et f est une bijection strictement croissante de $\mathbb{R}$ sur son image. Il reste à déterminer l'image de f.

   Si l'espace E n'est pas vide, comme les a_k sont tous distincts et que a s'écrit comme barycentre des a_k, il existe nécessairement au moins un a_k strictement plus grand que a et un autre strictement inférieur à a. Ceci implique que f tend vers $+\infty$ en $+\infty$ et vers $-\infty$ en $-\infty$.

3. On note $g_1(p) = \sum_{k=1}^n p_k$ et $g_2(p) = \sum_{k=1}^n a_k p_k$. On cherche à maximiser H sous les contraintes $g_1 = 1$ et $g_2 = a$. Les trois fonctions sont de classe $\mathcal{C}^1$ sur l'ouvert $(\mathbb{R}_+^\star)^n$. Comme $\nabla g_1(p) = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ et $\nabla g_2(p) = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ forment toujours une famille libre (sinon les a_k sont égaux), on peut appliquer le théorème des extrema liés : si H atteint son maximum sous contraintes en un point p de l'ouvert $(\mathbb{R}_+^\star)^n$, alors il existe des multiplicateurs de Lagrange λ_1 et λ_2 tels que :

$$\nabla H(p) = \lambda_1 \nabla g_1(p) + \lambda_2 \nabla g_2(p)$$

ou encore :

$$-\begin{pmatrix} 1 + \ln p_1 \\ \vdots \\ 1 + \ln p_n \end{pmatrix} = \lambda_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Pour chaque $1 \leq k \leq n$, on a donc $\ln p_k = -(1 + \lambda_1 + \lambda_2 a_k)$. Comme p vérifie la contrainte $g_1 = 1$, on obtient :

$$1 = \sum_{k=1}^n p_k = e^{-(1+\lambda_1)} \sum_{k=1} e^{-\lambda_2 a_k}.$$

Par ailleurs, comme p vérifie la contrainte $g_2 = a$, on obtient :

$$a = \sum_{k=1}^n a_k p_k = e^{-(1+\lambda_1)} \sum_{k=1} a_k e^{-\lambda_2 a_k}$$

En multipliant les deux extrémités de cette équation par $e^{(1+\lambda_1)}$, on obtient :

$$a e^{(1+\lambda_1)} = \sum_{k=1} a_k e^{-\lambda_2 a_k},$$

mais puisque $e^{(1+\lambda_1)} = \sum_{k=1} e^{-\lambda_2 a_k}$, ceci s'écrit, en multipliant les deux membres par $e^{a\lambda_2}$:

$$\sum_{k=1} a e^{-\lambda_2 (a_k - a)} = \sum_{k=1} a_k e^{-\lambda_2 (a_k - a)},$$

autrement dit $f(-\lambda_2) = 0$. Puisque f est une bijection, cette équation admet une unique solution $-\lambda_2 = f^{-1}(0)$. On détermine ensuite λ_1 en utilisant l'équation $e^{(1+\lambda_1)} = \sum_{k=1} e^{f^{-1}(0) a_k}$, puis les p_k comme $e^{-(1+\lambda_1+\lambda_2 a_k)}$.

4. De deux choses l'une : soit on applique le cours, en déterminant la matrice des dérivées secondes du lagrangien $L(p, \lambda_1, \lambda_2) = H(p) - \lambda_1(g_1(p) - 1) - \lambda_2(g_2(p) - a)$. Comme les dérivées secondes de g_1 et g_2 sont nulles, cette matrice est exactement la matrice hessienne de H. En l'occurence, c'est une matrice diagonale (pas de terme croisé dans l'écriture de H) avec toutes ses valeurs strictement négatives (puisque $-H$ est convexe). On a donc un maximum local strict de H sous les contraintes $g_1 = 1$, $g_2 = a$. Comme $-H$ est convexe sur E, on en conclut que c'est un maximum global de H par le même argument qu'en dimension 1.

Exercice 5 : Inégalité arithmético-géométrique

1. On cherche à optimiser la fonction $f : \mathbb{R}_+^n \to \mathbb{R}$ définie par $f(x_1, \cdots, x_n) = x_1 \cdots x_n$ sous la contrainte $x_1 + \cdots x_n = 1$.

   La fonction f est continue, et même de classe $\mathcal{C}^1$ sur $\mathbb{R}_+^n$. L'espace définissant la contrainte est $\{(x_1, \cdots x_n) | x_1 + \cdots x_n = 1 \text{ et } x_i \geq 0 \ \forall i\}$. Cet espace est fermé puisque les conditions le définissant sont fermées (égalité et inégalités larges). Comme $x_1 + \cdots x_n = 1$ et les x_i sont positifs,on a nécessairement $0 \leq x_i \leq 1$ pour tout i, ce qui implique que l'espace de contrainte est borné.Comme la fonction f est continue et comme l'espace de contraintes est compact, f est bornée et atteint ses bornes sous la contrainte.

La fonction f est positive sur $\mathbb{R}_+^n$ mais n'est pas constamment nulle sur l'espace de contraintes (elle est strictement positive par exemple en $x_1 = \cdots x_n = 1/n$). Le maximum de f sous la contrainte est donc strictement positif, ce qui implique qu'aucune des coordonnées d'un point l'atteignant ne peut être nulle (sinon f s'annule). Si f admet un maximum sous contrainte en un point x, on a donc que x appartient à l'ouvert $(\mathbb{R}_+^\star)^n$.

La fonction f s'annule si et seulement si au moins une coordonnée x_i est nulle. Il existe des points d'annulation de f dans l'espace de contrainte (par exemple $x_1 = 1$, $x_2 = \cdots = x_n = 0$). Comme f est positive, ceci implique que le minimum de f sous contrainte est zéro. Ce minimum est atteint en tout point de la contrainte ayant au moins une coordonnée nulle.

La fonction $g : (x_1, \cdots, x_n) \mapsto x_1 + \cdots + x_n$ est $\mathcal{C}^1$ sur $\mathbb{R}_+^n$. Le gradient de g est constamment égal au vecteur $(1, \cdots, 1)$ et ne s'annule donc jamais. Le théorème des extrema liés donne que si f admet un extremum sous contraintes en un point $x = (x_1, \cdots, x_n)$ intérieur au domaine $\mathbb{R}_+^n$ (c'est le cas pour son maximum d'après ce qui précède), alors il existe un multiplicateur de Lagrange λ tel que $\nabla f(x) = \lambda \nabla g(x)$, ou de manière équivalente :

$$\left\{ \begin{aligned} x_2 \cdots x_n &= \lambda \\ x_1 x_3 \cdots x_n &= \lambda \\ &\vdots \\ x_1 \cdots x_{n-1} &= \lambda \\ x_1 + \cdots + x_n &= 1 \end{aligned} \right.$$

Si $x = (x_1, \cdots, x_n)$ est un point pour lequel f est maximale sous contrainte, les x_i sont tous non nuls et le système d'équations précédent a comme seule solution $(1/n, \cdots, 1/n)$: les x_i sont égaux deux à deux puisque les produits de $n-1$ coordonnées sont toujours égaux et qu'on peut simplifier par les x_j (tous non nuls) redondants. La valeur de ces x_i est ensuite donnée par la contrainte. Le maximum de f sous contrainte est donc égal à $1/n^n$.

Remarque : Sur la figure suivante, pour $n = 2$, la contrainte est figurée par l'épais segment rouge en biais, les gradients de f sont les flèches vertes pointillées, et les gradients de g sont les flèches rouges et pleines.

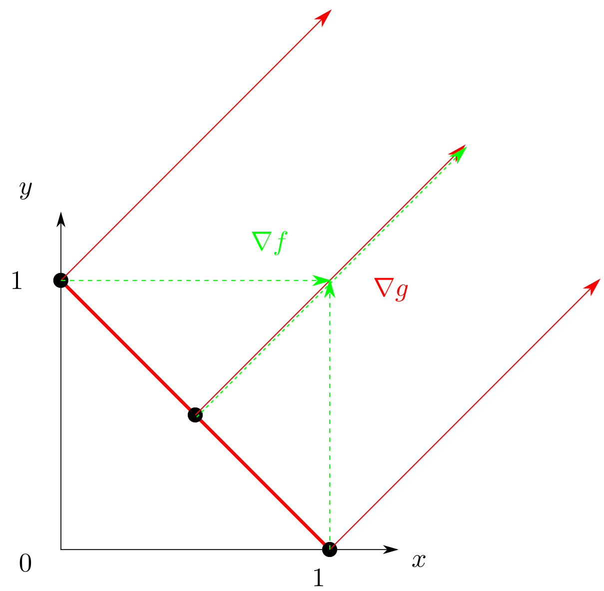

Figure 1: Espace de contrainte et gradient aux points d'extrema

- Au point de maximum $(1/2, 1/2)$, le théorème des extrema liés exprime que le gradient de f est colinéaire à celui de g.
- Le gradient de f n'est pas colinéaire à celui de g aux points de minimum $(1, 0)$ et $(0, 1)$. L'hypothèse manquante à l'application du théorème est celle topologique : les fonctions f et g ne sont pas définies sur un ouvert autour de ces points, et le théorème ne peut donc pas s'appliquer.

2. Soit $(x_1, \cdots, x_n)$ un n-uplet de réels positifs. S'ils sont tous nuls, l'inégalité suivante est triviale :

$$0 = \prod_{i=1}^{n} x_i^{\frac{1}{n}} \leq \frac{\sum_{i=1}^{n} x_i}{n} = 0.$$

S'ils ne sont pas tous nuls, leur somme $S = \sum_{i=1}^{n} x_i$ est strictement positive et l'on peut considérer le n-uplet $(u_1, \cdots, u_n) = (x_1/S, \cdots, x_n/S)$ qui appartient à l'espace de contrainte envisagé dans la question 1. Puisque le maximum de f vaut $1/n^n$, on a l'inégalité suivante :

$$u_1 \cdots u_n \leq 1/n^n$$

qui s'écrit encore :

$$\frac{x_1 \cdots x_n}{S^n} \leq 1/n^n.$$

On obtient l'inégalité arithmético-géométrique en multipliant les deux membres par S^n puis en passant à la racine n-ième :

$$\forall (x_1, \cdots, x_n) \in \mathbb{R}_+^n, \quad \prod_{i=1}^{n} x_i^{\frac{1}{n}} \leq \frac{\sum_{i=1}^{n} x_i}{n}.$$