

TD IV – Convexité

1 Généralités

Exercice 1.1 (TOPOLOGIE ET CONVEXITÉ). Soit C une partie convexe de \mathbb{R}^n .

- Soient $x, y \in \overline{C}$ et $t \in [0; 1]$. On considère une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de C qui converge vers x et une suite $(y_n)_{n \in \mathbb{N}}$ d'éléments de C qui converge vers y . Alors, pour tout $n \in \mathbb{N}$,

$$z_n = x_n + t(y_n - x_n) \in C$$

et $z_n \rightarrow x + t(y - x)$ quand $n \rightarrow +\infty$, donc $x + t(y - x) \in \overline{C}$.

- Soient $x, y \in \overset{\circ}{C}$ et $t \in [0; 1]$. Posons $z = x + t(y - x)$ et considérons $\delta > 0$ tel que $B(y, \delta) \subset C$. Soit $z' \in B(z, t\delta)$. Posons $y' = y + (z' - z)/t \in B(y, \delta) \subset C$. Alors

$$\begin{aligned} x + t(y' - x) &= x + ty + (z' - z) - tx \\ &= x + t(y - x) + z' - z \\ &= z'. \end{aligned}$$

Ainsi, z' est combinaison convexe de deux points de C donc est lui-même dans C . Autrement dit, $B(z, t\delta) \subset C$ et z est donc bien dans l'intérieur de C . En conclusion, l'intérieur d'un convexe est convexe.

- Soient C et D deux parties convexes. Soient $x, y \in C \cap D$ et $t \in [0; 1]$. Alors, $x + t(y - x) \in C$ par convexité, et de même $x + t(y - x) \in D$. Donc $x + t(y - x) \in C \cap D$, ce qu'il fallait démontrer.

Exercice 1.2 (INÉGALITÉ DES PENTES). 1. On a

$$g_a(x) = \frac{f(x) - f(a)}{x - a} = \frac{f(a) - f(x)}{a - x} = g_x(a).$$

- Posons $\lambda = \frac{y-x}{y-a}$. C'est un nombre compris entre 0 et 1 par hypothèse, et

$$\lambda a + (1 - \lambda)y = \frac{y-x}{y-a}a + \frac{x-a}{y-a}y = \frac{ya - xa + xy - ya}{y-a} = x.$$

On a alors, par convexité de f ,

$$\begin{aligned} f(x) &= f(\lambda a + (1 - \lambda)y) \\ &\leq \lambda f(a) + (1 - \lambda)f(y) \\ &= \frac{y-x}{y-a}f(a) + \frac{x-a}{y-a}f(y). \end{aligned}$$

- Le raisonnement est le même qu'à la question précédente, excepté qu'il faut cette fois poser $\lambda = \frac{y-a}{y-x}$.

4. Soient $x < y$ et supposons que $a < x$. Alors, d'après une question précédente,

$$\begin{aligned}
g_a(x) &= \frac{f(x)}{x-a} - \frac{f(a)}{x-a} \\
&\leq \frac{f(y)}{y-a} + \frac{y-x}{(y-a)(x-a)} f(a) - \frac{f(a)}{x-a} \\
&= \frac{f(y)}{y-a} + \frac{y-x-(y-a)}{(y-a)(x-a)} f(a) \\
&= \frac{f(y)}{y-a} - \frac{f(a)}{y-a} \\
&= g_a(y).
\end{aligned}$$

On procède de même dans les cas $x < a < y$ et $x < y < a$.

Exercice 1.3 (EXEMPLES DE FONCTIONS CONVEXES). 1. Les dérivées partielles de f sont

$$\frac{\partial f}{\partial x} = \alpha x^{\alpha-1} y^{1-\alpha} \quad \& \quad \frac{\partial f}{\partial y} = (1-\alpha) x^\alpha y^{-\alpha},$$

d'où

$$H_f(x, y) = \begin{pmatrix} \alpha(\alpha-1)x^{\alpha-2}y^{1-\alpha} & \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} \\ \alpha(1-\alpha)x^{\alpha-1}y^{-\alpha} & -\alpha(1-\alpha)x^\alpha y^{-\alpha-1} \end{pmatrix} = \alpha(1-\alpha) \begin{pmatrix} -x^{\alpha-2}y^{1-\alpha} & x^{\alpha-1}y^{-\alpha} \\ x^{\alpha-1}y^{-\alpha} & -x^\alpha y^{-\alpha-1} \end{pmatrix}.$$

Le déterminant de cette dernière matrice est

$$\det(H_f(x, y)) = x^{2\alpha-2}y^{-2\alpha} - x^{2\alpha-2}y^{-2\alpha} = 0.$$

Par conséquent, cette matrice n'a qu'une seule valeur propre non nulle si $(x, y) \neq (0, 0)$. Comme sa trace est égale à $-x^{\alpha-2}y^{1-\alpha} - x^\alpha y^{-\alpha-1} < 0$, on en déduit que cette seule valeur propre est négative, donc que la fonction f est concave.

2. On calcule les dérivées partielles

$$\frac{\partial g}{\partial x} = a(a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-\gamma}{\gamma}} \quad \& \quad \frac{\partial g}{\partial y} = (1-a)(ax^\gamma y^{-\gamma} + (1-a))^{\frac{1-\gamma}{\gamma}},$$

d'où

$$\begin{aligned}
\frac{\partial^2 g}{\partial x^2}(x, y) &= a \frac{1-\gamma}{\gamma} (-\gamma(1-a)y^\gamma x^{-\gamma-1}) (a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-2\gamma}{\gamma}} \\
&= -a(1-a)(1-\gamma)y^\gamma x^{-\gamma-1} (a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-2\gamma}{\gamma}} \\
\frac{\partial^2 g}{\partial y^2}(x, y) &= (1-a) \frac{1-\gamma}{\gamma} (-\gamma ax^\gamma y^{-\gamma-1}) (ax^\gamma y^{-\gamma} + (1-a))^{\frac{1-2\gamma}{\gamma}} \\
&= -a(1-a)(1-\gamma)x^\gamma y^{-\gamma-1} x^{1-2\gamma} y^{2\gamma-1} (a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-2\gamma}{\gamma}} \\
&= -a(1-a)(1-\gamma)x^{1-\gamma} y^{\gamma-2} (a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-2\gamma}{\gamma}} \\
\frac{\partial^2 g}{\partial y \partial x}(x, y) &= a \frac{1-\gamma}{\gamma} \gamma (1-a)x^{-\gamma} y^{\gamma-1} (a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-2\gamma}{\gamma}} \\
&= a(1-a)(1-\gamma)x^{-\gamma} y^{\gamma-1} (a + (1-a)y^\gamma x^{-\gamma})^{\frac{1-2\gamma}{\gamma}}
\end{aligned}$$

On en déduit que

$$H_g(x, y) = a(1-a)(1-\gamma) \left(a + (1-a)y^\gamma x^{-\gamma} \right)^{\frac{1-2\gamma}{\gamma}} \begin{pmatrix} -x^\gamma y^{-\gamma-1} & x^{-\gamma} y^{\gamma-1} \\ x^{-\gamma} y^{\gamma-1} & -x^{1-\gamma} y^{\gamma-2} \end{pmatrix}.$$

Le déterminant de la matrice ci-dessus est égal à

$$\det(H_g(x, y)) = xy^{-1} - x^{-2\gamma} y^{2\gamma-2}.$$

Cette quantité est positive si et seulement si $x^{2\gamma+1} \geq y^{2\gamma-1}$ et négative si et seulement si $x^{2\gamma+1} \leq y^{2\gamma-1}$. Comme aucune de ces deux inégalités n'est vérifiée sur tout \mathbb{R}_+^2 , la fonction g n'est ni convexe, ni concave.

2 Optimisation

Exercice 2.1 (ÉCHAUFFEMENT). On considère, pour $a \in \mathbb{R}$, la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f_a(x, y) = x^2 + y^2 - axy - 2x - 2y.$$

1. On calcule d'abord le gradient de f_a :

$$\nabla f_a(x, y) = \begin{pmatrix} 2x - ay - 2 \\ 2y - ax - 2 \end{pmatrix},$$

dont on déduit la Hessienne de f_a :

$$H_{f_a}(x, y) = \begin{pmatrix} 2 & -a \\ -a & 2 \end{pmatrix}.$$

On notera que cette matrice ne dépend pas de x ni de y , ce qui est normal puisque la fonction est quadratique.

2. Le déterminant de la Hessienne est $4 - a^2$, donc pour que cette matrice soit positive il faut que $a \in [-2; 2]$. De plus, la trace de la matrice est $4 > 0$, donc si les valeurs propres sont de même signe elle sont positives. Ainsi, f_a est convexe si et seulement si $a \in [-2; 2]$.
3. Comme f_a est convexe, pour trouver le minimum, il suffit d'annuler le gradient. Rappelons que

$$\nabla f_a(x, y) = \begin{pmatrix} 2x - ay - 2 \\ 2y - ax - 2 \end{pmatrix}.$$

Si ce vecteur est nul, alors en soustrayant les deux lignes on trouve $(2+a)(x-y) = 0$ donc $x = y$, ce qui donne finalement

$$x = \frac{2}{2-a} = y.$$

Le minimum correspondant est alors

$$\begin{aligned} f_a(x, y) &= \frac{8-4a}{(2-a)^2} - \frac{8}{(2-a)} \\ &= \frac{-8+4a}{(2-a)^2}. \end{aligned}$$

4. (a) On a

$$f_a(x, y) - \frac{1}{2} \left\langle H_{f_a} \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} \right\rangle = -2x - 2y.$$

- (b) Si $|a| > 2$, alors la matrice H_{f_a} a une valeur propre strictement négative λ puisque son déterminant est strictement négatif. Si $v = (x_0, y_0)$ est un vecteur propre associé, alors

$$\begin{aligned} f_a(tx_0, ty_0) &= \langle H_{f_a}tv, tv \rangle - 2t(x_0 + y_0) \\ &= \lambda t^2 \|v\|^2 - 2t(x_0 + y_0) \\ &= t(t\lambda(x_0^2 + y_0^2) - 2(x_0 + y_0)). \end{aligned}$$

Cette quantité tend vers $-\infty$ quand t tend vers $+\infty$, donc f_a n'a pas de minimum global.

5. Pour $a = 2$, on observe que ∇f_2 ne s'annule jamais. En effet, on aurait sinon $2(x - y) = 2 = 2(y - x)$. Pour $a = -2$, on peut remarquer que f_{-2} se factorise de la façon suivante :

$$\begin{aligned} f_{-2}(x, y) &= x^2 + y^2 + 2xy - 2(x + y) \\ &= (x + y)^2 - 1 - 2(x + y) \\ &= (x + y - 1)^2 - 2. \end{aligned}$$

On voit alors que f_{-2} a pour minimum -2 , qui est atteint en tout point tel que $x + y = 1$.

Exercice 2.2 (MINIMISATION SUR UNE BOULE). Pour $a \in \mathbb{R}^n$ et $p > 0$, on considère la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$f(x) = e^{\|x\|^2} + \langle a, x \rangle.$$

On cherche à minimiser f sur la boule fermée $B_f(0, p)$.

1. La boule fermée étant compacte et f étant continue, cette dernière y admet en particulier un minimum global.
2. Le plus simple est de poser $h(x) = \|x\|^2 - p^2$.
3. Il faut d'abord calculer les gradients, en notant que le gradient de la fonction $x \mapsto \|x\|^2 = \langle x, x \rangle$ est la fonction $x \mapsto 2x$. Ainsi, les conditions KKT sont

$$\begin{cases} 2e^{\|x\|^2}x + a &= 2\mu x \\ \|x\|^2 - p^2 &\leq 0 \\ \mu(\|x\|^2 - p^2) &= 0 \\ \mu &\leq 0 \end{cases}$$

4. On suppose $\mu \neq 0$.

- (a) Comme $\mu \neq 0$, on a d'après la troisième condition $\|x\|^2 = p^2$ et donc par positivité $\|x\| = p$. La première condition devient alors

$$(2e^{p^2} - 2\mu)x = -a,$$

qui en prenant la norme (et en utilisant le fait que $\mu \leq 0$ donc $2e^{p^2} - 2\mu > 0$) donne la relation attendue.

- (b) La relation précédente donne

$$\mu = -\frac{\|a\|}{2p} + e^{p^2},$$

qui réinjectée dans la première condition donne finalement

$$x = -\frac{p}{\|a\|}a.$$

On remarquera que ce point est bien dans $B_f(0, p)$.

(c) Il faut et il suffit que $\|a\| \geq 2pe^{p^2}$.

5. On suppose maintenant $\mu = 0$.

(a) Il suffit de prendre la norme de la première condition puisque le terme avec μ s'annule.

(b) La fonction φ est continue et même dérivable. De plus, on a $\varphi'(t) = 2e^{t^2} + 4t^2e^{t^2} > 0$, donc φ est strictement croissante. Comme en outre $\varphi(0) = 0$ et $\varphi(t) \rightarrow +\infty$ quand $t \rightarrow +\infty$, on conclut que φ est bijective par le THÉORÈME DES VALEURS INTERMÉDIAIRES.

(c) La première condition nous dit que x est négativement colinéaire à a , donc

$$x = -\frac{\|x\|}{\|a\|}a.$$

Or, d'après ce qui précède, $\|x\| = \varphi^{-1}(\|a\|)$, donc

$$x = -\frac{\varphi^{-1}(\|a\|)}{\|a\|}a.$$

(d) Pour vérifier la contrainte, il faut que $\|x\| \leq p$, c'est-à-dire que $\varphi^{-1}(\|a\|) \leq p$. Comme φ est croissante, ceci est équivalent à $\|a\| \leq \varphi(p) = 2pe^{p^2}$.

6. Remarquons que la fonction $x \mapsto \langle a, x \rangle$ est convexe car elle est linéaire, et qu'il suffit donc de montrer que $x \mapsto e^{\|x\|^2}$ est convexe puisqu'une somme de fonctions convexes est convexe. Or, la fonction $x \mapsto \|x\|^2$ est convexe et la fonction exponentielle est convexe et croissante, donc leur composée est convexe.

7. Par convexité, dès qu'un point vérifie les conditions KKT, on a un extremum global. Or, si $\|a\| \geq 2pe^{p^2}$ alors le point $-pa/\|a\|$ vérifie les conditions, et si $\|a\| \leq 2pe^{p^2}$, le point $-\varphi^{-1}(\|a\|)a/\|a\|$ vérifie les conditions. Nous avons donc dans tous les cas trouvé un point auquel le minimum global est atteint. Dans le premier cas, ce minimum vaut $e^{-p^2} - p\|a\|$ et dans le second cas, il vaut

$$e^{\varphi^{-1}(\|a\|)^2} - \|a\|\varphi^{-1}(\|a\|) = \frac{\|a\|}{2\varphi^{-1}(\|a\|)} - \|a\|\varphi^{-1}(\|a\|).$$

Exercice 2.3 (RADIO LONDRES). 1. Le Lagrangien est

$$\mathcal{L}(x_1, \dots, x_n, \mu, \mu_1, \dots, \mu_n) = \sum_{i=1}^n \ln(\alpha_i + x_i) - \mu \left(\sum_{i=1}^n x_i - 1 \right) + \sum_{i=1}^n \mu_i x_i$$

et les conditions KKT correspondantes s'écrivent

$$\begin{cases} \frac{1}{\alpha_i + x_i} = \mu - \mu_i \\ \mu_i x_i = 0 \end{cases}$$

La fonction f est concave et les contraintes sont affines, donc en particulier convexes. D'après le cours, les conditions KKT sont alors suffisantes pour avoir un maximum global.

2. Pour tout i on a

$$\mu = \frac{1}{\alpha_i + x_i} + \mu_i.$$

Comme $\mu_i \geq 0$, on a bien $\mu \geq 0$.

3. Supposons tout d'abord $\mu_i = 0$. Alors,

$$x_i = \frac{1}{\mu} - \alpha_i$$

mais cette quantité n'est positive que si $\mu \leq 1/\alpha_i$. Ainsi, si $\mu > 1/\alpha_i$, on ne peut avoir $\mu_i = 0$ et par conséquent on a $x_i = 0$. Pour résumer

$$x_i = \begin{cases} \frac{1}{\mu} - \alpha_i & \text{si } \mu \leq 1/\alpha_i \\ 0 & \text{si } \mu > 1/\alpha_i \end{cases}$$

4. Il suffit d'observer que le résultat de la question précédente peut s'écrire

$$x_i = \max\left(0, \frac{1}{\mu} - \alpha_i\right).$$

5. La fonction

$$F : t \mapsto \sum_{i=1}^n \max(0, t - \alpha_i)$$

est continue et strictement croissante. De plus, $F(0) = 0$ et $F(t) \rightarrow +\infty$ quand $t \rightarrow +\infty$, donc par le THÉORÈME DES VALEURS INTERMÉDIAIRES, il existe un unique $t_0 > 0$ tel que $F(t_0) = 1$. Alors, $\mu = 1/t_0$ donne l'unique maximum de f .

Exercice 2.4 (MÉTHODE DE PÉNALISATION). Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction strictement convexe telle que

$$\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty.$$

Pour une matrice $H \in M_{np}(\mathbb{R})$, on s'intéresse à l'optimisation de f sous les contraintes $Hx \leq d$.

1. Par linéarité, si $x, y \in \mathcal{D}$ et $t \in [0; 1]$, on a

$$\begin{aligned} H(tx + (1-t)y) &= tHx + (1-t)Hy \\ &\leq td + (1-t)d \\ &= d. \end{aligned}$$

. Ainsi, \mathcal{D} est bien convexe.

2. L'unicité du minimum découle de la convexité stricte de f .

3. Pour $\epsilon > 0$, on définit une fonction $f_\epsilon : U \rightarrow \mathbb{R}$ par

$$f_\epsilon(x) = f(x) + \frac{1}{2\epsilon} \|\max(Hx - d, 0)\|^2.$$

(a) L'application $x \mapsto Hx - d$ est convexe, de même que la norme au carré et que l'application $t \in \mathbb{R} \mapsto t^2$. Il suffit donc de montrer que l'application $s \in \mathbb{R} \mapsto \max(s, 0)$ est convexe pour conclure par composition. Nous allons pour ce faire distinguer plusieurs cas :

- Si $s_1, s_2 < 0$, alors $ts_1 + (1-t)s_2 < 0$ donc l'inégalité de convexité se réduit à $0 = 0$.
- De même, si $s_1, s_2 \geq 0$, alors l'inégalité de convexité est trivialement vérifiée.

- Si maintenant $s_1 \leq 0$ et $s_2 \geq 0$, on a pour $t \in [0; 1]$

$$\begin{aligned} \max(ts_1 + (1-t)s_2, 0) &\leq ts_1 + (1-t)s_2 \\ &\leq (1-t)s_2 \\ &= t \max(s_1, 0) + (1-t) \max(s_2, 0) \end{aligned}$$

et l'inégalité de convexité est encore satisfaite.

- (b) Considérons la fonction $F : s \in \mathbb{R} \mapsto \max(s, 0)^2$. Elle est dérivable en tout point de \mathbb{R}^* , de dérivée $s \mapsto 2 \max(s, 0)$ puisqu'elle coïncide alors soit avec la fonction nulle soit avec la fonction carré. En 0 on a

$$\frac{F(h) - F(0)}{h} = \begin{cases} h & \text{si } h \geq 0 \\ 0 & \text{si } h \leq 0 \end{cases}$$

qui tend vers 0. Ainsi, F est dérivable sur \mathbb{R} et $F'(s) = 2 \max(s, 0)$, ce qui signifie qu'elle est même de classe \mathcal{C}^1 . Considérons maintenant, pour $1 \leq k \leq n$, la fonction $G_k : x \mapsto F((Hx - d)_k)$. Elle est de classe \mathcal{C}^1 par composition, et

$$\begin{aligned} \frac{\partial G_k}{\partial x_i}(x) &= \frac{\partial}{\partial x_i} F \left(\sum_{j=1}^n H_{kj} x_j - d_k \right) \\ &= H_{ki} 2 \max((Hx - d)_k, 0). \end{aligned}$$

Pour conclure, il suffit de remarquer que

$$h(x) = \sum_{k=1}^n G_k(x)$$

et que par conséquent elle est de classe \mathcal{C}^1 et vérifie

$$\begin{aligned} \frac{\partial h}{\partial x_i}(x) &= \sum_{k=1}^n \frac{\partial G_k}{\partial x_i}(x) \\ &= \sum_{k=1}^n H_{ki} 2 \max((Hx - d)_k, 0) \\ &= (H^t 2 \max(Hx - d, 0))_k. \end{aligned}$$

- (c) La fonction f_ϵ est strictement convexe et tend vers l'infini à l'infini, donc elle a un unique minimum. Rappelons pourquoi. Posons $M = f(0)$ et remarquons que comme f tend vers l'infini, il existe $r > 0$ tel que $f(x) \geq f(0)$ dès que $\|f(x)\| > r$. Comme la boule fermée $B_f(0, r)$ est compacte et que f est continue, elle y admet un minimum que nous noterons m . Alors, on a $m \leq M$ et pour tout $x \in U$,

- Si $\|x\| \leq r$, alors $f(x) \geq m$ par définition d'un minimum ;
- Si $\|x\| > r$, alors $f(x) \geq M \geq m$.

Ainsi, m est bien un minimum global de f .

4. (a) On a par définition $f(x_\epsilon) < f_\epsilon(x_\epsilon)$. De plus, par minimalité, $f_\epsilon(x_\epsilon) < f_\epsilon(\tilde{x})$. Mais comme \tilde{x} appartient à \mathcal{D} , on a $\max(H\tilde{x} - d, 0) = 0$, donc $f_\epsilon(\tilde{x}) = f(\tilde{x})$, ce qui conclut.

Supposons maintenant que $(x_\epsilon)_\epsilon$ n'est pas bornée. Alors, il existe une suite $(\epsilon_n)_{n \in \mathbb{N}}$ telle que $\|x_{\epsilon_n}\| \rightarrow +\infty$. Par conséquent, $f(x_{\epsilon_n}) \rightarrow +\infty$, ce qui contredit le fait que cette suite est bornée par $f(\tilde{x})$.

- (b) Considérons la suite $(x_{1/n})_{n \in \mathbb{N}}$. Elle est bornée, donc elle admet une sous-suite convergente.

(c) Comme x_ϵ est un point critique de f_ϵ , on a pour tout $\epsilon > 0$ que

$$2H^t F(Hx_\epsilon - d) = 0$$

et donc en passant à la limite que

$$2H^t F(Hx_0 - d, 0) = 0.$$

Comme H^t est injective, on en déduit que $Hx_0 - d = 0$, c'est-à-dire que $x_0 \in \mathcal{D}$. Mais alors, l'inégalité $f(x_0) \leq f(\tilde{x})$ obtenue en passant à la limite dans la question précédente donne par unicité $x_0 = \tilde{x}$.

Nous avons montré que $(x_\epsilon)_\epsilon$ possède une unique valeur d'adhérence et ceci implique qu'elle converge vers cette valeur.

Exercice 2.5 (THÉORÈME DE SLATER ★). On considère la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = e^{-x}$$

qu'on veut minimiser sous la contrainte $h(x, y) \leq 0$, où $h : \mathbb{R} \times \mathbb{R}_+^* \rightarrow \mathbb{R}$ est définie par

$$h(x, y) = \frac{x^2}{y}.$$

1. La fonction $(x, y) \mapsto -x$ est linéaire donc convexe, et la fonction exponentielle est convexe et croissante, donc f est bien convexe. Quant à h , son gradient est $\nabla h(x, y) = (2x/y, -x^2/y^2)$ et sa Hessienne

$$H_h(x, y) = \begin{pmatrix} 2/y & -2x/y^2 \\ -2x/y^2 & 2x^2/y^3 \end{pmatrix}.$$

Son déterminant vaut 0 et sa trace est positive puisque $y > 0$. Ainsi, $H_h(x, y)$ est une matrice positive, ce qui d'après le cours implique que h est convexe.

Il suffit maintenant de constater que la contrainte ne peut être satisfaite que pour $x = 0$, ce qui donne 1 comme valeur du minimum de f .

2. (a) Le Lagrangien du problème est

$$\mathcal{L}(x, y, \mu) = e^{-x} - \mu \frac{x^2}{y}.$$

Comme $\mu \leq 0$, cette quantité est toujours positive. De plus, elle tend vers e^{-x} quand $y \rightarrow +\infty$, et cette dernière quantité tend vers 0 quand $x \rightarrow -\infty$. Ainsi, on a

$$f^*(\mu) = 0,$$

avec pour seule contrainte $\mu \leq 0$. Le maximum est donc 0, atteint pour n'importe quelle valeur de $\mu \leq 0$.

- (b) Par définition, le saut de dualité est la différence entre les valeurs optimales du problème primal et du problème dual :

$$S = 1 - 0 = 1 > 0.$$

Comme il n'est pas nul, le problème ne vérifie pas la dualité de Lagrange forte, ce qui signifie que les hypothèses du THÉORÈME DE SLATER ne sont pas satisfaites. De fait, il n'existe pas de point tel que $h(x, y) < 0$.

3 Optimisation quadratique

Exercice 3.1 (MEILLEUR ANTÉCÉDENT PAR UNE MATRICE NON INVERSIBLE). Soit $A \in M_n(\mathbb{R})$ une matrice et $b \in \mathbb{R}^n$. L'équation $Au = b$ n'ayant pas nécessairement de solution, on cherche la "meilleure approximation" possible en minimisant la quantité $\|Au - b\|$ sur u .

1. Il s'agit de minimiser la fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ définie par

$$\begin{aligned} f(x) &= \|Au - b\|^2 \\ &= \langle Au - b, Au - b \rangle \\ &= \langle Au, Au \rangle - 2\langle Au, b \rangle + \langle b, b \rangle \\ &= \langle A^t Au, u \rangle - \langle u, 2A^t b \rangle + \|b\|^2. \end{aligned}$$

On a donc bien un problème d'optimisation quadratique. De plus, on peut oublier le terme $\|b\|^2$, qui n'influe pas sur l'existence d'un minimum ni sur les points auquel il est atteint.

2. La matrice $A^t A$ est symétrique positive, donc le problème est convexe. De plus, u réalise le minimum si et seulement si le gradient s'annule, c'est-à-dire si et seulement si

$$A^t Au = A^t b.$$

3. Soit $x \in \mathbb{R}^n$. On peut le décomposer sur la base de vecteurs propres de $A^t A$:

$$x = \sum_{i=1}^n x_i v_i.$$

Alors d'une part,

$$\begin{aligned} \left(\sum_{i=1}^n \lambda_i v_i v_i^t \right) (x) &= \sum_{i,j=1}^n \lambda_i v_i v_i^t (x_j v_j) \\ &= \sum_{i=1}^n \lambda_i x_i v_i \end{aligned}$$

et d'autre part

$$\begin{aligned} A^t Ax &= \sum_{i=1}^n x_i A^t A v_i \\ &= \sum_{i=1}^n x_i \lambda_i v_i. \end{aligned}$$

d'où l'égalité.

4. On a

$$\begin{aligned} P &= \left(\sum_{i=1}^n \lambda_i v_i v_i^t \right) \left(\sum_{i=i_0+1}^n \frac{1}{\lambda_i} v_i v_i^t \right) \\ &= \left(\sum_{i=i_0+1}^n \lambda_i v_i v_i^t \right) \left(\sum_{i=i_0+1}^n \frac{1}{\lambda_i} v_i v_i^t \right) \\ &= \sum_{i=i_0+1}^n v_i v_i^t. \end{aligned}$$

Comme $\text{Im}(A^t) = \ker(A)^\perp$, $x \in \text{Im}(A^t)$ si et seulement si x est une combinaison linéaire de vecteurs propres associés à des valeurs propres non nulles. Autrement dit, l'image de A^t est engendrée par les vecteurs $(v_i)_{i>i_0}$ qui en forment une base orthonormée. Et P est exactement la projection orthogonale sur ce sous-espace.

5. On sait que le minimum est atteint en un point \tilde{u} tel que

$$A^t A \tilde{u} = A^t b.$$

Or,

$$\begin{aligned} A^t A \left(\sum_{i=i_0+1}^n \frac{1}{\lambda_i} v_i v_i^t \right) A^t b &= P(A^t b) \\ &= A^t b, \end{aligned}$$

d'où le résultat.

Exercice 3.2 (CONTRAINTES QUADRATIQUES). On considère la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = x^2 + y.$$

1. Il suffit de poser

$$A = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix} \quad \& \quad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

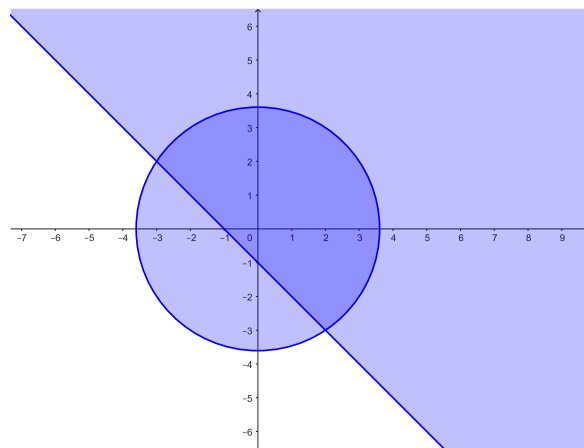
La matrice A est positive, donc f est convexe. Par contre, on a

$$\begin{aligned} f(0, 0) + \frac{1}{2}(f(0, 1) - f(0, 0)) &= \frac{1}{2} \\ &= f\left(0, \frac{1}{2}\right) \\ &= f\left(\frac{1}{2}(0, 0) + \frac{1}{2}[(0, 1) - (0, 0)]\right) \end{aligned}$$

donc f n'est pas strictement convexe.

On considère maintenant l'ensemble $\mathcal{D} = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 13 \ \& \ x + y \geq -1\}$.

2. (a) Voici la figure



- (b) Il s'agit de l'intersection d'un disque et d'un demi-plan, qui sont tous les deux convexes. C'est donc bien une partie convexe.
- (c) On pose, en faisant attention aux signes,

$$h_1(x, y) = x^2 + y^2 - 13 \quad \& \quad h_2(x, y) = -1 - x - y.$$

3. On calcule les gradients des contraintes :

$$\nabla h_1(x, y) = \begin{pmatrix} 2x \\ 2y \end{pmatrix} \quad \& \quad \nabla h_2(x, y) = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

Il suffit de vérifier la qualification quand les contraintes sont actives. Si seule h_1 est active, le gradient est alors non nul, donc c'est bon. De même si seule h_2 est active. Si les deux contraintes sont actives, alors on est sur l'un des deux points d'intersections du disque avec le demi-plan. Ces deux points ont pour coordonnées $(2, -3)$ et $(-3, 2)$ et on voit alors directement que la qualification linéaire des contraintes est vérifiée.

4. Les conditions KKT sont nécessaires puisque les contraintes sont qualifiées et suffisantes parce que la fonction est convexe. Elles s'écrivent

$$\begin{cases} 2x & = & 2x\mu_1 - \mu_2 \\ 1 & = & 2y\mu_1 - \mu_2 \\ \mu_1(x^2 + y^2 - 13) & = & 0 \\ \mu_2(-1 - x - y) & = & 0 \end{cases}$$

5. Il faut distinguer suivant les contraintes qui sont actives.

- Si aucune contrainte n'est active, la seconde équation n'a pas de solution.
- Si les deux contraintes sont actives, alors on est au point $(2, -3)$ ou $(-3, 2)$. Dans le premier cas, on a alors le système

$$\begin{cases} 4 & = & 4\mu_1 - \mu_2 \\ 1 & = & -6\mu_1 - \mu_2 \end{cases}$$

En soustrayant la seconde ligne à la première, on obtient $3 = 10\mu_1$ ce qui contredit $\mu_1 \leq 0$. Dans le second cas, on a le système

$$\begin{cases} -6 & = & -6\mu_1 - \mu_2 \\ 1 & = & 4\mu_1 - \mu_2 \end{cases}$$

En soustrayant cette fois la première ligne à la seconde, on obtient $7 = 10\mu_1$, qui contredit une nouvelle fois $\mu_1 \leq 0$.

- Si seule la première contrainte est active, alors la première équation s'écrit $2(1 - \mu_1)x = 0$, qui n'a pas de solution si $x \neq 0$ puisque $\mu_1 \leq 0$. Par conséquent, $x = 0$. Quant à la seconde équation, comme $\mu_2 = 0$ elle implique alors que $y \leq 0$. On doit donc avoir $y = -\sqrt{13}$, mais le point $(0, -\sqrt{13})$ n'est pas dans \mathcal{D} .
- Si seule la seconde contrainte est active, alors la seconde équation donne $\mu_2 = -1$ et la première par conséquent $x = 1/2$. Comme de plus on doit avoir $x + y = -1$, on a $y = -3/2$, qui est bien dans \mathcal{D} .

En conclusion, il n'y a qu'un point critique sous contraintes, à savoir $(1/2, -3/2)$ et le minimum est par conséquent atteint en ce point et vaut

$$f\left(\frac{1}{2}, -\frac{3}{2}\right) = \frac{1}{4} - \frac{3}{2} = -\frac{5}{4}.$$

- Exercice 3.3** (CONTRAINTES D'ÉGALITÉ LINÉAIRES ★). 1. Les contraintes sont données par les fonctions $g_i : x \mapsto (Gx)_i - c_i$. Ces fonctions sont de classe \mathcal{C}^1 car linéaires, et le gradient de g_i est la fonction constante égale au vecteur L_i^t , où L_i est la i -ème ligne de la matrice G . Par conséquent la famille des gradients des contraintes est libre si et seulement si la famille des lignes de G est libre. Or, celle-ci est libre si et seulement si elle est génératrice, et génératrice si et seulement si G est surjective.
2. Soit λ_1 la plus petite valeur propre de A . Alors, pour tout $x \in \mathbb{R}^n$, $\|f(x)\| \geq \lambda_1 \|x\|^2 - \|b\| \|x\|$. Il suit que la fonction f tend vers $+\infty$ quand $\|x\|$ tend vers $+\infty$, donc elle admet un minimum global. Comme de plus f est strictement convexe puisque A est définie positive, ce minimum est unique.
3. On applique la méthode des multiplicateurs de Lagrange : si le minimum est atteint en $\tilde{x} \in \mathbb{R}^n$, alors il existe $\lambda_1, \dots, \lambda_p \in \mathbb{R}$ tels que

$$\nabla f(\tilde{x}) = \sum_{i=1}^p \lambda_i \nabla g_i(\tilde{x}),$$

c'est-à-dire

$$A\tilde{x} + b = G^t \lambda$$

où $\lambda = (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$. Cette équation donne $\tilde{x} = A^{-1}(G^t \lambda - b)$.

4. Il suffit de prendre en compte la contrainte : on doit avoir

$$\begin{aligned} c &= G\tilde{x} \\ &= GA^{-1}G^t \lambda - GA^{-1}b, \end{aligned}$$

ce qui peut également s'écrire $GA^{-1}G^t \lambda = c + GA^{-1}b$. On sait que f admet un minimum, donc un tel λ existe.

5. Pour tout $y \in \mathbb{R}^p$, on a

$$\langle GA^{-1}G^t(y), y \rangle = \langle A^{-1}G^t(y), G^t(y) \rangle > 0$$

puisque A^{-1} est définie positive. Ainsi, $GA^{-1}G^t$ est définie positive et en particulier inversible. On a donc $\lambda = (GA^{-1}G^t)^{-1}(c + GA^{-1}b)$ et en remplaçant dans l'expression précédente, on conclut que

$$\begin{aligned} \tilde{x} &= A^{-1}G^t \lambda - A^{-1}b \\ &= A^{-1}G^t(GA^{-1}G^t)^{-1}c + A^{-1}G^t(GA^{-1}G^t)^{-1}GA^{-1}b - A^{-1}b. \end{aligned}$$

6. Nous avons montré plus haut que f tend vers $+\infty$, donc elle n'a pas de maximum.

Exercice 3.4 (UN PEU DE DUALITÉ ★). On considère le problème de minimisation de la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = x^2 + y^2$$

sous les contraintes $x + y = 1$ et $x \geq 0$.

1. Le Lagrangien du problème est

$$\begin{aligned} \mathcal{L}((x, y), \lambda, \mu) &= x^2 + y^2 - \lambda(x + y - 1) - \mu(-x) \\ &= (x^2 - (\lambda + \mu)x) + (y^2 - \lambda y) + \lambda. \end{aligned}$$

2. Pour déterminer le problème dual, il faut minimiser \mathcal{L} par rapport à la variable (x, y) . Comme on le voit, on peut minimiser indépendamment selon x et y . Dans les deux cas, on doit minimiser une fonction de la forme $t \mapsto t^2 - \alpha t$. La dérivée de cette fonction est $2t - \alpha$, qui s'annule pour $\tilde{t} = \alpha/2$. Ce point critique est un minimum global par convexité de la fonction, et la valeur de ce minimum se calcule directement pour donner $-\alpha^2/4$. Ainsi, comme pour x on a $\alpha = \lambda + \mu$ et pour y on a $\alpha = \lambda$, la fonction duale est

$$f^*(\lambda, \mu) = -\frac{(\lambda + \mu)^2}{4} - \frac{\lambda^2}{4} + \lambda.$$

Aucune contrainte n'est apparue lors de la minimisation, il ne reste donc que la contrainte $\mu \leq 0$.

3. On cherche maintenant à résoudre le problème dual.

- (a) L'ensemble $\{(\lambda, \mu) \mid \mu < 0\}$ est un ouvert, donc si f^* admet un maximum local sur cet ensemble, son gradient s'y annule. En particulier,

$$\begin{aligned} 0 &= \frac{\partial f^*}{\partial \lambda}(\lambda, \mu) \\ &= -\frac{\lambda + \mu}{2} - \frac{\lambda}{2} + 1 \\ &= 1 - \lambda - \frac{\mu}{2}. \end{aligned}$$

De même, on doit avoir

$$\begin{aligned} 0 &= \frac{\partial f^*}{\partial \mu}(\lambda, \mu) \\ &= -\frac{\lambda + \mu}{2}. \end{aligned}$$

Ainsi, $\mu = -\lambda$ et donc $\lambda = 2$ et $\mu = -2$.

L'unique point critique de f^* est $(2, -2)$, et comme f^* est concave elle a un maximum global en ce point et $f^*(2, -2) = 1$.

- (b) Si $\mu = 0$, on cherche le maximum de la fonction $\lambda \mapsto \lambda - \lambda^2/2$. En dérivant on voit que le maximum est atteint en $\lambda = 1$ et vaut $1/2$.
- (c) En comparant les deux valeurs obtenus, on voit que le maximum de f^* sous la contrainte $\mu \leq 0$ est 1.
4. Commençons par vérifier les hypothèses du Théorème de Slater : la fonction f à optimiser est convexe, les contraintes d'égalité sont linéaires et les contraintes d'inégalités sont linéaires donc convexes et de classe \mathcal{C}^1 . De plus,

- (a) Le point $(1, 0)$ satisfait les contraintes et vérifie $h(x) < 0$;
- (b) La fonction g est surjective, donc son image est \mathbb{R} et 0 en est bien un point intérieur.

On peut donc appliquer le THÉORÈME DE SLATER qui affirme que le minimum du problème primal est égal au maximum du problème dual, c'est-à-dire $-1/2$. On sait que ce minimum sera atteint en un point pour lequel les multiplicateurs seront donnés par le maximum du problème dual. Ainsi, en écrivant les équations KKT on doit avoir

$$\begin{cases} 2x &= \lambda - \mu &= 1 \\ 2y &= \lambda &= 1 \end{cases}$$

Autrement dit, $x = y = 1/2$, qui donne bien $1/2$ comme valeur de f .