# Notes d'optimisation différentiable

Hicham Janati
hicham.janati@inria.fr

Janvier 2020

*Disclaimer : ces notes constituent un résumé et non un substitut du cours d'Optimisation différentiable.*

## 1 Calcul différentiel

**Objectifs :**

1. Différentielle et gradient
2. Dérivées partielles et leur continuité
3. Chain rule
4. Hessienne et approximation de second ordre

**notations**

- On note $\langle x, y\rangle = \sum_{i=1}^n x_i y_i = x^\top y$ le produit scalaire Euclidien de $x, y \in \mathbb{R}^n$
- La transposée d'une matrice $A$ est notée $A^\top$.
- La notation $x = o(h^p)$ est équivalente à $x = \|h\|^p \varepsilon(h)$ où $\varepsilon$ est une fonction $\mathbb{R}^n \to \mathbb{R}$ continue en 0 et $\varepsilon(\mathbf{0}) = 0$.
- $\|.\|$ denote la norme Euclidienne : $\|x\| = \sqrt{\sum_i x_i^2}$.
- Pour une matrice symmétrique, $H \succ 0$ signifie que $H$ est définie-positive, càd $x^\top H x > 0$ pour tout $x$ non nul (ou encore toutes ses valeurs propres sont strictement positives).

**Définition 1** (Différentielle et Jacobienne)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ et $x \in \mathbb{R}$. On dit que $f$ est différentiable en $x$ si et seulement s'il existe une application linéaire $J_x : \mathbb{R}^n \to \mathbb{R}^m$ tel que pour tout $h \in \mathbb{R}^n$ :*

$$f(x+h) = f(x) + J_x(h) + o(h) \tag{1}$$

*L'application $J_x$ est dite différentielle de $f$ en $x$.*

*Comme $J_x$ est linéaire, elle peut être représentée par une matrice $J_f(x) \in \mathbb{R}^{m\times n}$ appelée Jacobienne de $f$ et on a $J_f(x)_{ij} = \frac{\partial f_i}{\partial x_j}(x)$.*

**Exemple 1.** *Soit $A \in \mathbb{R}^{m,n}$. La fonction linéaire $f : x \mapsto Ax$ est différentiable et sa hessienne est donnée par $J_f(x) = A$ pour tout $x \in \mathbb{R}^n$. En effet, $f(x+h) = A(x+h) = Ax + Ah = f(x) + Ah$.*

**Définition 2** (Gradient d'une fonction scalaire)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}$ et $x \in \mathbb{R}$ une fonction différentiable. La différentielle de $f$ en $x$ est une application linéaire donc il existe $a_x \in \mathbb{R}^n$ tel que pour tout $h \in \mathbb{R}^n$ :*

$$f(x+h) = f(x) + \langle a_x, h\rangle + o(h) \tag{2}$$

*Le vecteur $a_x$ est dit gradient de $f$ en $x$ et on note, pour tout $x$ où $f$ est différentiable : $\nabla f(x) = a_x$. En plus, les coordonnées de $\nabla f(x)$ sont données par les dérivées partielles de $f$ : $\nabla f(x)_i = \frac{\partial f}{\partial x_i}(x)$.*

*Réciproque : si les dérivées partielles $x \to \frac{\partial f}{\partial x_i}(x)$ existent et sont **continues**, alors $f$ est différentiable et son gradient est donné par $\nabla f(x) = \left(\frac{\partial f}{\partial x_i}(x)\right)$.*

**Exemple 2.** *La fonction $f : x \mapsto \frac{1}{2}\|x\|^2$ est différentiable et $\nabla f(x) = x$. En effet, $\frac{1}{2}\|x+h\|^2 = \frac{1}{2}\|x\|^2 + \langle x, h\rangle + \frac{1}{2}\|h\|^2 = f(x) + \langle x, h\rangle + o(h)$.*

Pour chercher la dérivée partielle d'une fonction $f$ en $x_0$ suivant une direction $u \neq 0$, il suffit de calculer la limite :

$$\lim_{h\to 0} \frac{f(x_0 + hu) - f(x_0)}{h} \tag{3}$$

L'exercice suivant montre que l'existence des dérivées partielles (sans leur continuité) ne garantit pas la continuité de $f$ – et donc sa différentiabilité.

**Exercice 1** (Mi-parcours 2018)**.** *Montrer que les deux fonctions suivantes admettent des dérivées partielles en (0, 0) dans toutes les directions de $\mathbb{R}^2$ sans pour autant être continues en (0, 0).*

1. $f(x, y) = \begin{cases} y^2 \log(|x|) \text{ if } x \neq 0 \\ 0 \text{ if } x = 0 \end{cases}$
2. $f(x, y) = \begin{cases} \frac{x^2 y}{x^4+y^2} \text{ if } (x, y) \neq (0, 0) \\ 0 \text{ if } (x, y) = (0, 0) \end{cases}$

L'existence des dérivées partielles sans leur continuité est donc en général insuffisante pour avoir la différentiabilité de $f$. De même, si une fonction est différentiable, ses dérivées partielles ne sont pas forcément continues. Lorsque c'est le cas, il s'agit du cas (plus fort) d'une fonction de classe $\mathcal{C}^1$ :

**Définition 3** (Fonction de classe $\mathcal{C}^1$)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}$. On dit que $f$ est de Classe $\mathcal{C}^1$ si $f$ est différentiable et ses dérivées partielles sont continues.*

**Proposition 1** (Chain rule)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}^m$ et $g : \mathbb{R}^m \to \mathbb{R}^p$ deux applications différentiables. Leur composée $h = gof$ est différentiable et sa Jacobienne est donnée par le produit matriciel des Jacobiennes :*

$$J_{gof}(x) = J_g(f(x))J_f(x) \tag{4}$$

*Remark* 1. Pour une fonction scalaire différentiable (càd à valeurs dans $\mathbb{R}$), la Jacobienne est la transposée du gradient. Dans la proposition ci-dessus, si $g$ est scalaire ($p = 1$) alors $h$ l'est aussi est on a :

$$\nabla h(x) = J_f(x)^\top \nabla g(f(x)) \tag{5}$$

**Exemple 3** (Changement de variable linéaire)**.** *Dans la remarque ci-dessus, si $f : x \mapsto Ax$ avec $A \in \mathbb{R}^{m,n}$ alors : $\nabla h(x) = A^\top \nabla g(Ax)$.*

**Exercice 2** (Les classiques)**.** *On dit que $x$ est un point stationnaire (ou point critique) d'une fonction $f$ différentiable en $x$ si et seulement si $\nabla f(x) = 0$. Soit $a \in \mathbb{R}^n$, $b \in \mathbb{R}^m$ et $A \in \mathbb{R}^{m\times n}$. Soit une fonction différentable $g : \mathbb{R}^m \to \mathbb{R}$. Les fonctions suivantes sont-elles différentiables ? Donnez le gradient (Là où il existe) et les points critiques éventuels des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ suivantes en fonction de $a, b, A$ et $\nabla g$.*

1. $x \mapsto \langle a, x\rangle = a^\top x$
2. $x \mapsto \|x\|^2$
3. $x \mapsto \|Ax - b\|^2$
4. $x \mapsto g(Ax)$
5. $x \mapsto \|x\|$

**Définition 4** (Hessienne d'une fonction scalaire)**.** *Soit $f : \mathbb{R}^n \to \mathbb{R}$ et $x \in \mathbb{R}$ une fonction différentiable. On dit que $f$ est deux fois différentiable en $x$ si et seulement s'il existe une forme bilinéaire symétrique $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ tel que pour tout $h \in \mathbb{R}^n$ :*

$$f(x + h) = f(x) + \langle \nabla f(x), h\rangle + \frac{1}{2} S_x(h, h) + o(h^2) \tag{6}$$

*Comme $S_x$ est bilinéaire symétrique, elle admet une représentation matricielle donnée par : $H_f(x)_{ij} = \frac{\partial^2 f}{\partial x_i \partial x_j}(x)$. $H_f(x)$ est la Hessienne de $f$ en $x$ et (6) devient :*

$$f(x + h) = f(x) + \nabla f(x)^\top h + \frac{1}{2} h^\top H_f(x)h + o(h^2) \tag{7}$$

*Remark* 2 (Hessienne comme Jacobienne). L'écriture matricielle permet de voir la Hessienne comme la Jacobienne de $\nabla f$. Souvent, il est plus facile de retrouver la Hessienne à partir du gradient. S'il existe une application linéaire $J_x$ telle que :

$$\nabla f(x+h) = \nabla f(x) + J_x(h) + o(h) \tag{8}$$

Alors $H_f(x) = J_x$.

**Exercice 3.** *Calculez la Hessienne des fonctions 1 - 2 - 3 de l'exercice 2.*

# 2 Optimisation sans contraintes

**Objectifs :**

1. Utiliser la coercivité pour montre l'existence de solution
2. Étude de points critiques avec les critères de premier et second ordre
3. Convexité et courbure d'une fonction

## 2.1 Existence

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable sur $\mathbb{R}^n$. On s'intéresse aux problème :

$$\min_{x \in \mathbb{R}^n} f(x) \tag{9}$$

Le problème (9) peut éventuellement ne pas admettre de solution, si par exemple $f$ n'est pas minorée. Une condition suffisante pour qu'une solution existe est la coercivité

**Définition 5** (Coercivité)**.** *On dit que $f$ est coercive si et seulement si :* $\lim_{\|x\| \to +\infty} f(x) = +\infty$

**Proposition 2.** *Si une fonction continue $f$ est coercive, alors le problème (9) admet une solution.*

**Exemple 4.** *Soit $a \in \mathbb{R}^n$ La fonction $f : x \mapsto \|x\|^2 - \langle x, a\rangle$ est coercive. Par l'inégalité de Cauchy-Schwartz : $f(x) \geq \|x\|^2 - \|x\|\|a\| \to +\infty$ when $\|x\| \to +\infty$.*

## 2.2 Étude de points critiques

Souvent, on se contente de trouver des solutions locales au problème. S'il existe $x^\star$ tel que $f(x^\star) \leq f(x) \quad \forall x$ alors $x^\star$ est un minimiseur global de $f$, solution de (9). S'il existe $x^\star$ et $r > 0$ tel que $f(x^\star) \leq f(x) \quad \forall x \; \|x - x^\star\| \leq r$ alors $x^\star$ est un minimiseur local de $f$.

**Proposition 3.** *Soit $x^\star$ un minimiseur local de $f$ alors $\nabla f(x^\star) = 0$.*

PROOF. Il existe un voisinage $\mathcal{N}$ de $x^\star$ tel que $\forall x \in \mathcal{N} \quad f(x^\star) \leq f(x)$. Soit $a \in \mathbb{R}^n$ et $t > 0$, on définit l'interpolation : $x_t = ta + x^\star$. On voit que si $t$ est assez petit, $x_t \in \mathcal{N}$ et donc, pour $t$ assez petit on peut écrire l'équation de premier ordre de $f$ en $x_t$ autour de $x^\star$ :

$$\begin{aligned}
&f(x^\star) \leq f(x_t) \\
&\Rightarrow f(x^\star) \leq f(x^\star + ta) \\
&\Rightarrow f(x^\star) \leq f(x^\star) + \langle \nabla f(x^\star), ta\rangle + o(ta) \\
&\Rightarrow 0 \leq \langle \nabla f(x^\star), ta\rangle + o(t)\|a\| \\
&\Rightarrow 0 \leq \langle \nabla f(x^\star), a\rangle + \frac{o(t)}{t}\|a\| \\
&\Rightarrow 0 \leq \langle \nabla f(x^\star), a\rangle
\end{aligned}$$

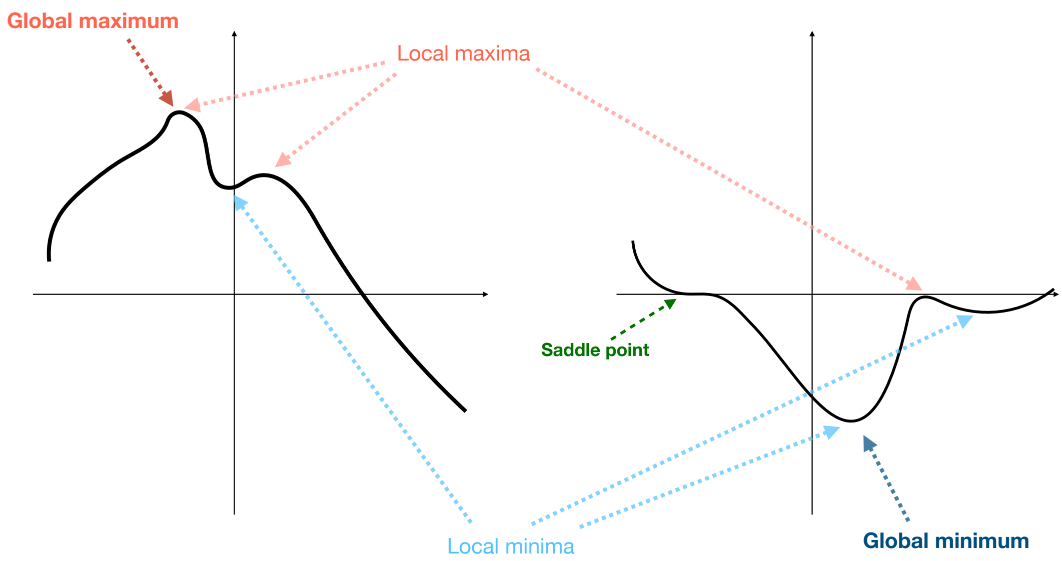



Figure 1 – Exemples de points critiques

où on a divisé par $t$ avant de passer à la limite $t \to 0$.

Comme $a$ est arbitraire dans $\mathbb{R}^n$, on a $\nabla f(x^\star) = 0$. □

On remplaçant $f$ par $-f$, on voit que tout maximiseur local de $f$ annule également le gradient de $f$. Les points annulant le gradient de $f$ sont appelés *points critiques* ou *points stationnaires*. Pour connaitre leur nature, il faut aller au second ordre et évaluer la Hessienne de $f$ :

**Proposition 4.** *Soit $x^\star$ un point critique de $f$. Alors :*

1. $H_f(x^\star) \succ 0 \Rightarrow x^\star$ *est un minimiseur local.*
2. $H_f(x^\star) \prec 0 \Rightarrow x^\star$ *est un maximiseur local.*

PROOF. Soit $a \in \mathbb{R}^n$ et $t > 0$. On a :

$$\begin{aligned} f(x^\star + ta) - f(x^\star) &= \overbrace{\langle \nabla f(x^\star), ta\rangle}^{=0} + \frac{1}{2}t^2 a^\top H_f(x^\star)a + \|a\|^2 o(t^2) \\ \Rightarrow \frac{f(x^\star + ta) - f(x^\star)}{t^2} &= \frac{1}{2}a^\top H_f(x^\star)a + \|a\|^2 o(1) \end{aligned}$$

Donc pour $t$ assez petit, le signe du membre de gauche est le signe de $a^\top H_f(x^\star)a$, et comme $a$ est arbitraire on obtient 1 et 2. □

**Exercice 4** (Examen 2018)**.** *Calculer le gradient et la Hessienne des fonctions suivantes. En déduire les points critiques des fonctions suivantes et déterminer leur nature*

1. $f : (x, y) \in \mathbb{R} \times \mathbb{R}_+^* \mapsto x^2 - \sqrt{y}$
2. $f : (x, y) \in \mathbb{R}_+^* \times \mathbb{R}_+^* \mapsto \sqrt{xy}$
3. $f : (x, y) \in \mathbb{R} \times \mathbb{R} \mapsto x^2 + y^2$

La proposition 3 donne une condition suffisante pour déterminer si un point critique est un minimiseur ou maximiseur local. Si en revanche la Hessienne a des valeurs propres de signe opposé ou une valeur propre nulle, ce critère de deuxième ordre ne permet pas de déterminer la nature du point critique. En effet, si par exemple la Hessienne a une valeur propre nulle, il faudra aller à un ordre d'approximation supérieur pour évaluer le signe de la courbure de la fonction. En pratique, pour un problème de minimisation sans contraintes, après avoir énuméré tous les points critiques, il suffit d'évaluer $f$ en ces points et comparer les valeurs obtenues : car si un minimiseur global existe, il doit être parmi ces points critiques. L'exercice suivant illustre cette situation.

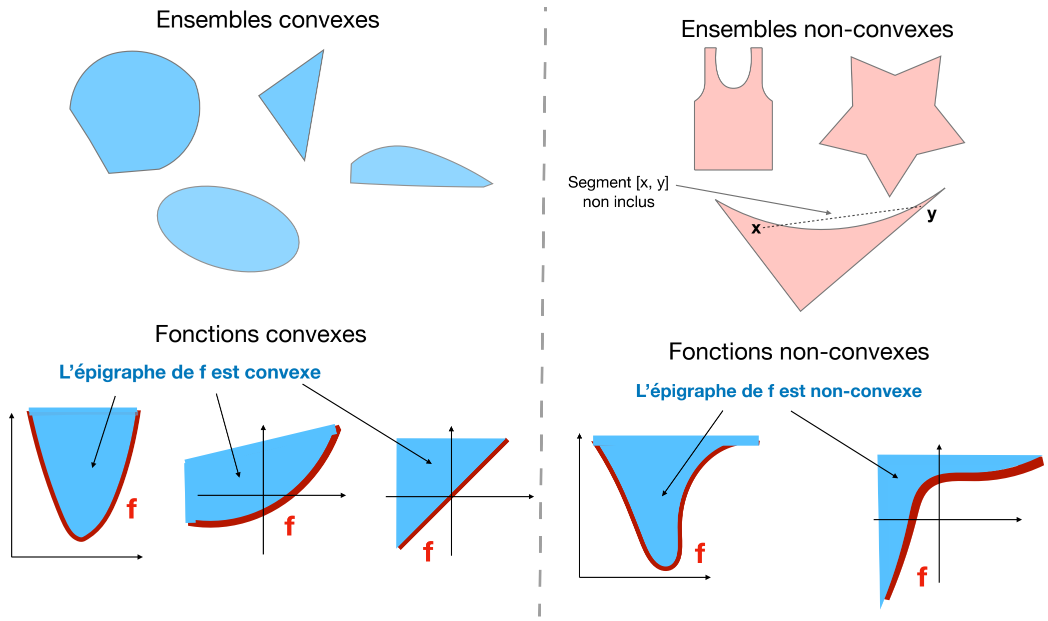



FIGURE 2 – Exemples d'ensembles et fonctions convexes

**Exercice 5.** *Déterminez les points critiques (et leur nature) des fonctions suivantes.*

1. $f(x, y) = x^3 + 3xy^2 - 15x - 12y$.
2. $g(x, y) = 3x^3 + xy^2 - xy$.
3. $h(x, y) = x^4 + \frac{1}{3}y^3 - 4y - 2$.
4. $k(x, y) = x^3 + xy^2 - x^2y - y^3$.

**Exercice 6.** *Soit f :* $\mathbb{R}^2 \to \mathbb{R}$ *telle que :*

$$f(x, y) = x^4 + y^4 - 2x^2 - 2y^2 + 4xy$$

*On s'intéresse au problème*

$$\min_{\mathbb{R}^2} f(x, y) \qquad (P)$$

1. *Montrez que (P) admet une solution*
2. *Résoudre (P)*

**Exercice 7.** *Soit f :* $\mathbb{R}^2 \to \mathbb{R}$ *telle que :*

$$f(x, y) = \frac{1}{4}x^4 - xy + y^2$$

1. *Montrez que f est coercive.*
2. *Calculez les points critiques de f.*
3. *Résoudre* $\min f$.

## 2.3 Cas d'une fonction convexe

Un ensemble $C$ est convexe si et seulement si pour tout $x, y \in C$, le segment liant $x$ à $y$ (formellement tout point $tx + (1-t)y$ pour tout $t \in [0, 1]$) est inclus dans $C$. On dit qu'une fonction $f$ est convexe si son épigraphe est convexe. L'épigraphe d'une fonction est tout simplement l'ensemble des points au-dessus de son graphe : $\text{epi}_f = \{(x_1, \ldots, x_n, y) \in \mathbb{R}^{n+1} | f(x) \leq y\}$. Cette définition admet d'autres formulations équivalentes :

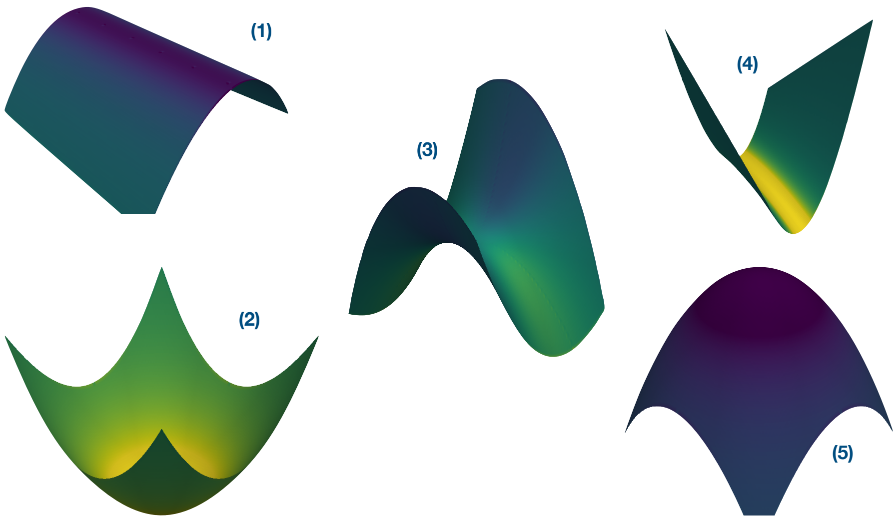



FIGURE 3 – Surfaces de fonctions quadratiques - Exercice 8

**Proposition 5.** *Soit $f$ une fonction deux fois différentiable de $\mathbb{R}^n$ dans $\mathbb{R}$. Les assertions suivantes sont équivalentes :*

1. *$f$ est convexe*
2. *l'ensemble $epi_f = \{(x_1, \dots, x_n, y) \in \mathbb{R}^{n+1} | f(x) \leq y\}$ est convexe.*
3. *$\forall (x, y) \quad \forall t \in [0, 1] f(tx + (1-t)y) \leq tf(x) + (1-t)f(y)$*
4. *$\forall (x, x_0) \quad f(x) \geq f(x_0) + \langle x - x_0, \nabla f(x_0) \rangle$ (f est supérieure à toutes ses tangeantes)*
5. *La hessienne de $f$ est semi-définie positive pour tout $x$ : $H_f(x) \succcurlyeq 0$.*

Lors que $f$ est convexe, elle admet **au plus** un minimum global, atteint en potentiellement **plusieurs minimiseurs**. Si elle est strictement convexe, alors **s'il existe, ce minimiseur est unique**. En effet, l'approximation au premier ordre d'une fonction convexe permet de montrer que tout point critique est un minimiseur global. C'est donc une caractérisation des solutions du problème $\min f$.

**Proposition 6.** *Soit $f$ une fonction convexe et différentiable de $\mathbb{R}^n$ dans $\mathbb{R}$. $x^\star$ est un minimiseur global de $f$ si et seulement si $\nabla F(x^\star) = 0$.*

## 2.4 Fonction quadratique

La proposition 3 montre que l'étude des points critiques passe par l'étude de la fonction quadratique $h \rightarrow h^\top H_f h$. Les fonctions quadratiques donnent l'approximation de second ordre de toute fonction deux fois différentiable. Comprendre le lien entre la courbure d'une fonction quadratique, sa convexité et sa Hessienne est crucial en optimisation. Ceci est l'object de l'exercice suivant.

**Exercice 8.** *Soit $S$ une matrice symétrique dans $\mathbb{R}^{n,n}$ et $b \in \mathbb{R}^n$. On s'intéresse à la fonction quadratique : $f(x) = \frac{1}{2} x^\top S x + b^\top x$.*

1. *Calculez le gradient et la Hessienne de $f$.*
2. *Quels sont les points critiques de $f$ ? Discutez leur nature.*
3. *Montrez que si $S$ a une valeur propre strictement négative, $f$ ne peut pas être coercive.*
4. *On suppose que $S \succcurlyeq 0$. Trouvez une condition nécéssaire et suffisante sur $b$ et $S$ pour qu'un minimum de $f$ existe.*
5. *Prenons $n = 2$. La figure 3 visualise la surface de $f$ pour différentes matrices $A$. Déterminez le signe des valeurs propres de $A$ dans les cas suivants :*

**Exercice 9.** *Soit $A$ une matrice dans $\mathbb{R}^{m,n}$ et $b \in \mathbb{R}^m$. On s'intéresse à la fonction quadratique : $f(x) = \frac{1}{2}\|Ax - b\|^2$.*

1. *Calculez le gradient et la Hessienne de $f$.*
2. *$f$ est-elle convexe ?*
3. *$f$ est-elle coercive ?*
4. *Résoudre* $\min f$

# 3 Optimisation sous contraintes

**Objectifs :**

1. Étudier la qualification des contraintes
2. Appliquer KKT et visualiser les problèmes dans $\mathbb{R}^2$
3. Conclure par convexité ou existence de solution
4. Trouver un problème dual équivalent

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable sur $\mathbb{R}^n$. Et $K$ un fermé de $\mathbb{R}^n$. On s'intéresse au problème d'optimisation suivant :

$$\min_{x \in K} f(x) \tag{10}$$

Nous allons nous restreindre aux cas où $K$ peut s'exprimer sous forme de $E$ contraintes d'égalité et $I$ contraintes d'inégalités avec des fonctions différentiables $g : \mathbb{R}^n \to \mathbb{R}^E$ et $h : \mathbb{R}^n \to \mathbb{R}^I$ : $K = \{x \in \mathbb{R}^n | g_1(x) = \dots g_E(x) = 0, h_1(x) \leq 0 \dots, h_I(x) \leq 0\}$. Pour simplifier les notations, on dira qu'un vecteur $y \in \mathbb{R}^p$ est négatif si toutes ses coordonnées le sons : $y \leq 0 \Leftrightarrow y_1 \leq 0, \dots, y_p \leq 0$. Ainsi, (10) devient :

$$\min_{\substack{g(x)=0 \\ h(x) \leq 0}} f(x) \tag{11}$$

## 3.1 Qualification des contraintes

Lorsque le problème d'optimisation est sous contraintes, on verra que les conditions nécéssaires d'optimalité ne s'appliquent qu'aux points $x \in K$ où les contraintes $g, h$ vérifient certaines propriétés de régularité. Ces propriétés dites *de qualification*, traduisent le fait que l'on peut entièrement décrire la géométrie locale de $K$ en $x$ à l'aide des jacobiennes $J_g(x)$ et $J_h(x)$. En pratique, on ne cherchera pas à exprimer ces propriétés explicitement mais on se contentera de vérifier des conditions suffisantes qui les garantissent. On énumère dans cette section ces différentes conditions suffisantes en allant du plus particulier au plus général.

**Proposition 7** (Contraintes affines)**.** *Si $g$ et $h$ sont affines, alors elles sont qualifiées en tout point de $K$.*

**Proposition 8** (Slater)**.** *Si $g$ est affine, $h_1, \dots, h_r$ sont affines et $h_r, \dots, h_I$ convexes, alors s'il existe $x_0 \in K$ tel que $h_j(x_0) < 0$ pour tout $j \in [\![r, I]\!]$ alors $h, g$ sont qualifiées en tout point de $K$.*

**Proposition 9** (Indépendence linéaire)**.** *Soit $x \in K$. S'il existe $k$ tel que $h_k(x) = 0$ , alors quitte à réindexer les $(h_j)_j$, supposons que $h_1(x) = \dots h_k(x) = 0$ et $h_{k+1}(x) \dots h_I(x) < 0$. Si la famille des gradients $\{\nabla g_1(x), \dots, \nabla g_E(x), \nabla h_1(x), \dots \nabla h_r(x)\}$ est libre, alors $h, g$ sont qualifiées en $x$.*

En particulier, on remarque que si l'on n'a que des contraintes d'égalité, la condition d'indépendence linéaire se résume en l'indépendance linéaire des gradients de $g$ en $x$, et donc la surjectivité de la Jacobienne de $g$ en $x$ :

**Proposition 10** (Indépendence linéaire - égalité)**.** *Soit $x \in K = \{x, g(x) = 0\}$. Si la famille des gradients $\{\nabla g_1(x), \dots, \nabla g_E(x)\}$ est libre, alors $g$ est qualifiée en $x$.*

**Proposition 11** (Condition de Mangasarian-Fromovitz (MF))**.** *Soit $x \in K$. S'il existe $k$ tel que $h_k(x) = 0$, alors quitte à réindexer les $(h_j)_j$, supposons que $h_1(x) = \dots h_k(x) = 0$ et $h_{k+1}(x) \dots h_I(x) < 0$. Si $\{\nabla g_1(x), \dots, \nabla g_E(x)\}$ est libre, et s'il existe $v \in K$ tel que $(\forall i) \quad \nabla g_i(x)^\top v = 0$ et $(\forall i > r) \quad \nabla h_i(x)^\top v < 0$, alors $h, g$ sont qualifiées en $x$.*

Notons $I_0(x) \subset [\![1, I]\!]$ l'ensemble des indices des contraintes $h_i$ tel que $h_i(x) = 0$, dites contraintes actives en $x$. Comme on cherche à décrire l'ensemble $K$ localement en $x$, par continuité de $h$, les contraites inactives ($h_j(x) < 0$) ne participent pas à la description locale de $K$ en x. On peut donc donner une formulation générale des propriétés 7 et 8 en restreignant les conditions aux contraintes d'inégalité actives. Ceci a par contre l'inconvénient de devoir traiter la qualification en un point $x \in K$. Pour la première condition de qualification, cela donne :

**Contraintes affines - (actives)** Soit $x \in K$. Si $g$ et $h_i \forall i \in I_0(x)$ sont affines, alors les contraintes sont qualifiées en $x$.

## 3.2 Contraintes d'égalité

Considérons tout d'abord le cas de contraintes d'égalité uniquement.

$$\min_{g(x)=0} f(x) \tag{12}$$

On rappelle que $g$ est une fonction différentiable $\mathbb{R}^n \to \mathbb{R}^E$, exprimant $E$ contraintes d'égalité $g_1(x) = \cdots = g_E(x) = 0$. On rappelle également que la Jacobienne de $g$ en $x \in \mathbb{R}^n$ est la matrice donnée par :

$$J_g(x) = \left(\frac{\partial g_i}{\partial x_j}(x)\right)_{i,j} = \begin{pmatrix} \nabla g_1(x) \\ \dots \\ \nabla g_i(x) \\ \dots \\ \nabla g_E(x) \end{pmatrix} \tag{13}$$

On commence par donner une condition nécessaire d'optimalité, vérifiée par tout minimum local du problème (12) qui vérifie une condition de qualification.

### 3.2.1 KKT - condition nécessaire

**Proposition 12** (KKT - condition nécessaire)**.** *Soit $x \in K$. Si $x$ est un minimum local de $f$ tel que $g$ est qualifiée en $x$, alors il existe $\lambda \in \mathbb{R}^E$ tel que :*

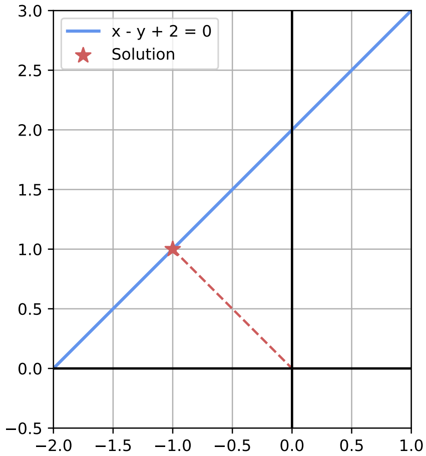
$$\begin{aligned} &\begin{cases} \nabla f(x) + J_g(x)^\top \lambda = 0 \\ g(x) = 0 \end{cases} \\ \Leftrightarrow &\begin{cases} \nabla f(x) + \sum_{i=1}^E \lambda_i \nabla g_i(x) = 0 \\ g(x) = 0 \end{cases} \end{aligned} \tag{14}$$

Par la proposition 10, la surjectivité de $J_g(x)$ est une condition suffisante de qualification de $g$ en $x$. Avec (13), on voit que pour chercher les points où cette condition de qualification est vérifiée, il suffit de chercher les $x$ tels que la famille des gradients $(\nabla g_i(x))_i$ est libre.

**Exemple 5.** *On considère le problème :*

$$\min_{x-y+2=0} x^2 + y^2 \tag{15}$$

*Graphiquement, on cherche un $(x, y)$ appartenant à la droite d'équation $y = x + 2$ qui minimise la norme $\|(x, y)\|^2$, autrement dit le point le plus proche de $(0, 0)$ vérifiant la contrainte.*




*On a une seule contrainte $g(x) = x - y + 2$. $\nabla g(x) = (1, -1) \neq 0$. Une famille constituée d'un seul élément est libre si et seulement si ce dernier est non nul, donc $J_g(x, y)$ est inversible quelque soit $x, y$ ; $g$ est donc qualifiée pour tout $x, y$. Cherchons $x, y \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}$ tels que :*

$$\nabla f(x, y) + \lambda \nabla g(x, y) = 0 \tag{16}$$

$$\Rightarrow \begin{cases} 2x + \lambda = 0 \\ 2y - \lambda = 0 \end{cases} \tag{17}$$

$$\Rightarrow x + y = 0 \tag{18}$$

*Or la contrainte d'égalité s'écrit $x - y + 2 = 0$, avec (18) on obtient $x = -1$ et $y = 1$, et enfin $\lambda = 2$.*

*Conclusion 1 : si un minimum existe alors, c'est forcément $(x, y) = (-1, 1)$.*

*En revanche, la restriction de $f$ sur $K$ est clairement coercive, donc un minimum global existe.*

*Conclusion, $(x, y) = (-1, 1)$ est un minimiseur global.* □

**Exercice 10.** *Résoudre les problèmes d'optimisation suivants :*

1. $\min\limits_{x^2+y^2=1} x + y$
2. $\min\limits_{x+y=1} x^4 + y^4$
3. $\min\limits_{x+2y=1} x^2 + y^2 + xy$

**Exercice 11.** *Soit $p \geq 1$. Étudier les problème d'optimisation selon $p$ :*

$$\min_{x^{2p}+y^{2p}=1} x^2 + y^2$$

*Et*

$$\max_{x^{2p}+y^{2p}=1} x^2 + y^2$$

### 3.2.2 Problème convexe - condition suffisante

**Définition 6** (Problème convexe)**.** *on dit que le problème $\min\limits_{g(x)=0} f(x)$ est convexe si $f$ est convexe et $g$ est affine.*

**Proposition 13** (KKT - condition suffisante)**.** *Si le problème $\min\limits_{g(x)=0} f(x)$ est convexe, alors :*

*S'il existe $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}^E$ solutions du système KKT (14) alors $x$ est un minimum global de $\min\limits_{g(x)=0} f(x)$.*

**Exercice 12.** *Soit $A \in \mathbb{R}^{m,n}$ et $b \in \mathbb{R}^m$. On s'intéresse au problème $\min\limits_{Ax=b} \|x\|^2$.*

1. *S'agit-il d'un problème convexe ?*
2. *Écrire les équations KKT du problème*

## 3.3 Contraintes d'égalité et d'inégalité

Revenons au cas général :

$$\min_{\substack{g(x)=0 \\ h(x)\leq 0}} f(x) \tag{19}$$

### 3.3.1 KKT - condition nécessaire

**Proposition 14** (KKT - condition nécessaire)**.** *Soit $x \in K$ où $g$ et $h$ sont qualifiées. Si $x$ est un minimum local de $f$, alors il existe $\lambda \in \mathbb{R}^E$ et $\mu \in \mathbb{R}^I$ tel que :*

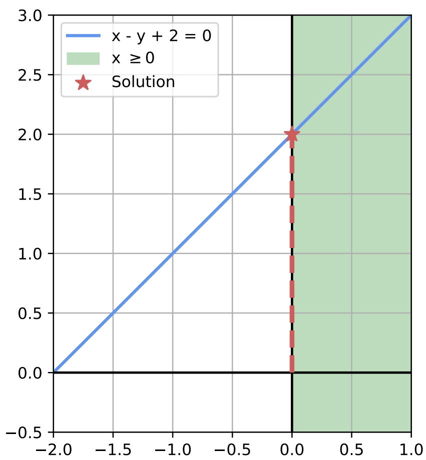
$$\begin{cases} \nabla f(x) + \sum_{i=1}^{E} \lambda_i \nabla g_i(x) + \sum_{i=1}^{I} \mu_i \nabla h_i(x) = 0 \\ g(x) = 0 \\ h(x) \leq 0 \\ \mu \geq 0 \\ h_i(x)\mu_i = 0 \quad 1 \leq i \leq I \end{cases} \tag{20}$$

La condition $h_i(x)\mu_i = 0 \quad \forall 1 \leq i \leq I$ est appelée condition de *complémentarité*. Elle implique que toute contrainte $h_i$ est soit active en $x$ i.e $h_i(x) = 0$ soit son multiplicateur associé $\mu_i$ est nul, et dans ce cas le gradient $\nabla h_i(x)$ ne participe pas dans l'équation d'optimalité et donc la contrainte $h_i \leq 0$ est inutile vis-à-vis de l'optimalité de $x$. Intuitivement, cela traduit le fait que si une contrainte $h_i$ est inactive $(h_i(x) < 0)$ alors comme $h_i$ est continue, la description locale du cône tangeant de $K$ en $x$ ne dépend pas de $h_i$.

**Exemple 6.** *On considère le problème :*

$$\min_{\substack{x-y+2=0 \\ x\geq 0}} x^2 + y^2 \tag{21}$$

*Graphiquement, on cherche un $(x, y)$ appartenant à la droite d'équation $y = x + 2$ avec une abscisse $x \geq 0$ qui minimise la norme $\|(x, y)\|^2$, autrement dit le point le plus proche de $(0, 0)$ vérifiant la contrainte.*




*Le problème s'écrit en forme standard avec les contraintes $g(x, y) = x - y + 2$ et $h(x, y) = -x$. Les contraintes $g$ et $h$ sont affines, par la proposition 7, elles sont qualifiées pour tout $x, y$. Cherchons $x, y \in \mathbb{R}^n$,*

*$\lambda \in \mathbb{R}$ et $\mu \geq 0$ tels que :*

$$\left\{\begin{array}{l} \nabla f(x,y) + \lambda \nabla g(x,y) + \mu \nabla h(x,y) = 0 \\ \mu \geq 0 \\ h(x,y)\mu = 0 \\ g(x,y) = 0 \\ h(x,y) \leq 0 \end{array}\right. \tag{22}$$

$$\Rightarrow \left\{\begin{array}{l} \left\{\begin{array}{l} 2x + \lambda - \mu = 0 \\ 2y - \lambda = 0 \end{array}\right. \\ \mu \geq 0 \\ x\mu = 0 \\ x - y + 2 = 0 \\ x \geq 0 \end{array}\right. \tag{23}$$

$$\tag{24}$$

*La contrainte de complémentarité $x\mu = 0$ implique $\mu = 0$ ou $x = 0$. Si $\mu = 0$, on obtient (comme à l'exemple 5) $(x,y) = (-1,1)$ qui ne vérifie pas la contrainte $x \geq 0$. Si $x = 0$, on obtient avec la contrainte d'égalité $y = 2$ puis $\lambda = \mu = 4 \geq 0$.*

*Conclusion 1 : si un minimum existe alors, c'est forcément $(x,y) = (0,2)$.*

*En revanche, comme $f$ est clairement coercive, un minimum global existe.*

*Conclusion, $(x,y) = (0,2)$ est un minimiseur global.* □

**Exercice 13.** *Résoudre les problèmes d'optimisation suivants :*

1. $\min\limits_{\substack{x \geq 0 \\ y \geq 0 \\ x+2y=4}} x^2 + y^2$
2. $\min\limits_{\substack{x \geq 0 \\ y \geq 0 \\ x+2y=4 \\ x^2+y^2 \leq 16}} x^2 + y^2$
3. $\min\limits_{\substack{x^2+y^2 \leq 1 \\ x \geq 0}} x - y$
4. $\min\limits_{\substack{e^x+e^y \leq 20 \\ x \geq 0}} e^{x-y}$

### 3.3.2 Problème convexe - condition suffisante

**Définition 7** (Problème convexe)**.** *on dit que le problème $\min\limits_{\substack{g(x)=0 \\ h(x) \leq 0}} f(x)$ est convexe si $f$ est convexe, les $h_i$ sont convexes et $g$ est affine.*

**Proposition 15** (KKT - condition suffisante)**.** *Si le problème $\min\limits_{\substack{g(x)=0 \\ h(x) \leq 0}} f(x)$ est convexe, alors :*

*S'il existe $x \in \mathbb{R}^n$ et $\lambda \in \mathbb{R}^E, \mu \in \mathbb{R}^I$ solutions du système KKT (20) alors $x \in K$ est un minimum global de $f$.*

On remarque pour un problème convexe, par la condition de qualification de Slater (Proposition 8), il suffit de montrer qu'il existe $x_0 \in K$ tel que $h(x_0) < 0$ pour avoir la qualification en tout point de $K$.

**Exercice 14.** *Soit $A \in \mathbb{R}^{m,n}$ et $b \in \mathbb{R}^m$. On s'intéresse au problème des moindres carrés positif : $\min\limits_{\substack{Ax=b \\ x \geq 0}} \|x\|^2$.*

1. *S'agit-il d'un problème convexe ?*
2. *Écrire les équations KKT du problème*

### 3.4 Dualité Lagrangienne

Le Lagrangien du problème d'optimisation $(P) \min_{\substack{g(x)=0 \\ h(x)\leq 0}} f(x)$ est défini par la fonction $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^E \times \mathbb{R}^I_+ \to \mathbb{R}$ telle que :

$$\mathcal{L}(x, \lambda, \mu) = f(x) + \sum_{i=1}^{E} \lambda_i g_i(x) + \sum_{i=1}^{I} \mu_i h_i(x)$$

L'idée derrière cette définition est de trouver un problème équivalent à (P) en cherchant des bornes inférieures à la solution du problème (P). En effet, si $x \in K$, on a $g_i(x) = 0$ et $h_i(x) \leq 0$. Comme $\mu \geq 0$, on obtient :

$$(\forall \lambda, \mu \geq 0, x \in K) \quad \mathcal{L}(x, \lambda, \mu) \leq f(x).$$

Ainsi, en passant à l'inf sur $x \in K$ :

$$(\forall \lambda, \mu \geq 0) \inf_{x \in K} \mathcal{L}(x, \lambda, \mu) \leq \inf_{x \in K} f(x)$$

Or par définition de l'inf on a $\inf_{x \in \mathbb{R}^n} \mathcal{L}(x, \lambda, \mu) \leq \inf_{x \in K} \mathcal{L}(x, \lambda, \mu)$, donc :

$$(\forall \lambda, \mu \geq 0) \inf_{x} \mathcal{L}(x, \lambda, \mu) \leq \inf_{x \in K} f(x)$$

Et on peut donc "réduire" l'écart de l'inégalité ci-dessus en maximisant sur $\lambda, \mu$ pour mieux "approcher" notre problème original. On obtient alors l'inégalité dite de dualité faible :

$$\max_{\lambda, \mu \geq 0} \inf_{x} \mathcal{L}(x, \lambda, \mu) \leq \inf_{x \in K} f(x)$$

L'objectif est de trouver des conditions suffisantes pour que l'inégalité ci-dessus soit une égalité. On obtiendra donc un problème de maximisation équivalent au problème original.

**Définition 8.** *Fonction et problème dual En gardant les mêmes notations ci-dessus, supposons que* $\inf_x \mathcal{L}(x, \lambda, \mu)$ *soit bien défini et notons un de ses minimiseurs* $x^\star(\lambda, \mu)$*. La fonction* $g : (\lambda, \mu) \mapsto \inf_x \mathcal{L}(x, \lambda, \mu) = \mathcal{L}(x^\star(\lambda, \mu), \lambda, \mu)$ *est appelée fonction duale et le problème de maximisation :*

$$\max_{\substack{\lambda \in \mathbb{R}^E \\ \mu \in \mathbb{R}^I_+}} g(\lambda, \mu) \tag{25}$$

*est dit problème dual.*

L'intérêt de la dualité Lagrangienne est de résoudre un problème avec contraintes potentiellement compliquées en passant à un problème équivalent avec une seule contrainte de signe ($\mu \geq 0$). Par contre, cette équivalence (dualité forte) n'a pas toujours lieu. Une condition suffisante est le cas d'un problème convexe différentiable :

**Proposition 16** (Dualité Forte)**.** *Si le problème (P) est convexe i.e* $f, h_1, \ldots, h_I$ *sont convexes et* $g_1, \ldots, g_E$ *sont affines, alors on a dualité forte :*

$$\max_{\substack{\lambda \in \mathbb{R}^E \\ \mu \in \mathbb{R}^I_+}} g(\lambda, \mu) = \min_{x \in K} f(x)$$

**En pratique (valeur optimale)**

En pratique, si (P) est convexe, la proposition 16 nous permet de résoudre le problème en suivant les étapes suivantes :

1. Calculer la fonction duale $g(\lambda, \mu) = \inf_x \mathcal{L}(x, \lambda, \mu)$ (problème sans contraintes ; donc résoudre $\nabla_x \mathcal{L}(x, \lambda, \mu) = 0$. En effet, pour un problème convexe $\mathcal{L}$ reste convexe en $x$ (avec $\mu \geq 0$) donc tout point critique est un minimum global.
2. Résoudre le problème dual en appliquant KKT (ce qui donne des solutions duales $(\lambda^\star, \mu^\star)$) et la valeur optimale du problème $f(x^\star) = g(\lambda^\star, \mu^\star)$.

*Remark* 3. La fonction duale $g$ est toujours concave (voir exercice **??**). Et donc $-g$ est convexe. Par conséquent, KKT est suffisant pour avoir l'optimalité à l'étape 2.

On peut donc retrouver la valeur optimale $f(x^\star)$ en passant par la dualité Lagrangienne. En revanche, la proposition 16 ne nous donne pas un moyen de retrouver le / les minimiseur $x^\star$ du problème initial. Pour cela, il faut d'abord vérifier la qualification des contraintes et utiliser la condition de complémentarité :

**En pratique (minimiseurs optimaux)**

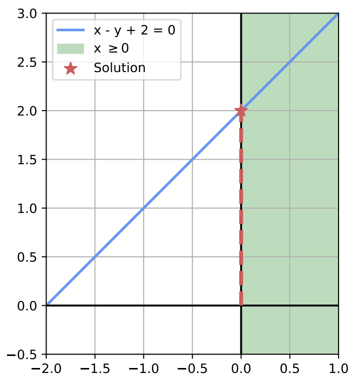
En pratique, si (P) est convexe, pour lequel les contraintes sont qualifiées (Souvent avec la condition de Slater, étant la plus facile à établir en optimisation convexe), pour retrouver les minimisurs $x^\star$ :

3. Suivre les étapes 1. et 2. dans l'encadré ci-dessus pour trouver des solutions duales $(\lambda^\star, \mu^\star)$.
4. Pour chaque pair de solutions duales $(\lambda^\star, \mu^\star)$, trouver les $x^\star$ tels que $\nabla_x \mathcal{L}(x^\star, \lambda^\star, \mu^\star) = 0$ et $\mu_i^\star h_i(x^\star) = 0 \, \forall 1 \leq i \leq I$. Cette liste englobe toutes les solutions primales $x^\star$.

**Exemple 7.** *Reprenons l'exemple 6 :*

$$\min_{\substack{x-y+2=0 \\ x \geq 0}} x^2 + y^2 \tag{26}$$

*En utilisant KKT, on avait montré que la solution du problème était donnée par (0, 2) et donc la valeur optimale est 4. Essayons de retrouver ce résultat par dualité Lagrangienne en appliquant les étapes 1 à 4 énumérees ci-dessus. Tout d'abord, il s'agit bien d'un problème convexe (f et h convexe et g affine). En plus, les contraintes sont linéaires donc qualifiées partout.*




*Le problème s'écrit en forme standard avec les contraintes $g(x, y) = x - y + 2$ et $h(x, y) = -x$. Notons $z = [x, y]^\top$. Le Lagrangien du problème s'écrit pour tout $\lambda \in \mathbb{R}$ et $\mu \in \mathbb{R}_+$ : $\mathcal{L}(z, \lambda, \mu) = x^2 + y^2 + \lambda(x - y + 2) - \mu x$. Pour trouver la fonction duale, on résout :*

$$\begin{aligned} & \nabla_z \mathcal{L}(z, \lambda, \mu) = 0 \\ & \Rightarrow \begin{cases} 2x + \lambda - \mu = 0 \\ 2y - \lambda = 0 \end{cases} \Rightarrow \begin{cases} x = \frac{\mu - \lambda}{2} \\ y = \frac{\lambda}{2} \end{cases} \Rightarrow z^\star = \frac{1}{2}[\mu - \lambda, \lambda]^\top \end{aligned}$$

*Par conséquent, la fonction duale est donnée par :*

$$\begin{aligned} g(\lambda, \mu) &= \mathcal{L}(z^\star, \lambda, \mu) \\ &= \frac{1}{4}(\lambda^2 + (\mu - \lambda)^2) + \lambda(\frac{1}{2}\mu - \lambda + 2) - \frac{1}{2}\mu(\mu - \lambda) \\ &= -\frac{\lambda^2}{2} - \frac{\mu^2}{4} + \frac{\mu\lambda}{2} + 2\lambda \end{aligned}$$

*Maintenant, il faut résoudre le problème dual :*

$$\max_{\lambda, \mu \geq 0} g(\lambda, \mu) = - \min_{\lambda, \mu \geq 0} \frac{\lambda^2}{2} + \frac{\mu^2}{4} - \frac{\mu\lambda}{2} - 2\lambda$$

.

*Comme ici l'objectif dual est une fonction quadratique, on peut facilement vérifier sa convexité. En calculant les dérivées partielles secondes, la Hessienne de -g est donnée par :* $\nabla^2(-g)(\lambda, \mu) = \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}$ *Sa trace est* $\frac{3}{2}$ *et son déterminant est* $\frac{1}{4}$*. Elle admet donc deux valeurs propres strictement positives :* $-g$ *est stricement convexe. KKT est donc suffisant pour avoir l'optimalité.* $\mu, \lambda$ *est une solution duale si et seulement il existe* $a \in \mathbb{R}$ *tel que :*

$$\begin{cases} (i) \lambda - \frac{\mu}{2} - 2 = 0 \\ (ii) \frac{\mu}{2} - \frac{\lambda}{2} - a = 0 \\ a\mu = 0 \\ \mu \geq 0; a \geq 0 \end{cases}$$

*Si* $\mu = 0$ *alors (i) donne* $\lambda = 2$ *et (ii) donne* $\mu = -\frac{\lambda}{2} < 0$ *impossible. Donc forcément* $a = 0$ *et on a* $\lambda = \mu = 4$*. L'unique solution duale est* $(\lambda^\star, \mu^\star) = (4, 4)$*. On obtient comme valeur optimale* $g(4, 4) = 4$*.*

***Ètape 4, retour aux solutions primales*** *:*

*Avec la solution duale* $(4, 4)$ *le seul* $(x, y)$ *annulant le gradient du Lagragien est* $(x^\star, y^\star) = (0, 2)$*. Comme il vérifie également la contrainte de complémentarité* $\mu^\star h(x^\star, y^\star) = \mu^\star \times -x^\star = 4 \times 0 = 0$*, c est bien l'unique solution primale.*

*Conclusion : la solution est (0, 2).*

**Exercice 15.** *Résoudre les problèmes d'optimisation suivants en passant par la dualité Lagrangienne :*

1. $\min\limits_{\substack{x \geq 0 \\ y \geq 0 \\ x + 2y = 4}} x^2 + y^2$
2. $\min\limits_{\substack{x^2 + y^2 \leq 1 \\ x \geq 0}} x - y$
3. $\min\limits_{\substack{e^x + e^y \leq 20 \\ x \geq 0}} e^{x-y}$

**Exercice 16.** *Soit* $A \in \mathbb{R}^{n \times n}$ *symétrique définie positive et* $b \in \mathbb{R}^n$*.*

$$\begin{aligned} \min \quad & x^\top Ax + b^\top x \\ & \|x\|_\infty \leq 1 \end{aligned} \tag{(P)}$$

*1. Ecrire ce problème sous forme standard (problème d'optimisation différentiable sous contraintes d'inégalités / d'égalités).*

*2. Donner le lagrangien de (P).*

*3. Donner le problème dual de (P).*

# 4 Algorithmes d'optimisation numérique

# 5 Banque d'exercices et problèmes

## 5.1 Exercices

**Exercice 17.** *(Fonction quadratique - Optim sans contraintes) Soit $a \neq 0$ et $c \in \mathbb{R}$ On définit la fonction $f$ de $\mathbb{R}^2$ dans $\mathbb{R}$ par :*

$$f(x, y) = ax^2 + 2xy + y^2 + c$$

1. *Étudier les extrema de $f$ selon $a$ et $c$*
2. *si $a > 1$ montrez que $f$ est coercive.*
3. *Conclure quant à la nature des extrema si $a = 2$.*

**Exercice 18** (Élimination de contrainte)**.** *Le but de cet exercice est de montrer qu'il faut être très prudent lorsqu'on élimine des contraintes. Soit $a > 0$. On considère le problème*

$$\min_{ax^2=y} x^2 + (y-1)^2 \quad (P)$$

1. *Montrez que (P) admet une solution et résoudre (P).*
2. *Si dans la fonction à minimiser on remplace $x^2$ en utilisant la contrainte, on obtient :*

$$\min_y \frac{1}{a} y + (y-1)^2 \quad (Q)$$

*Résoudre (Q).*

3. *Que constatez-vous ?*

**Exercice 19** (Fonctions quadratiques - Opt sans contraintes - Régression linéaire)**.** *Soit $A \in \mathbb{R}^{n,p}$ et $b \in \mathbb{R}^n$ Et $f : \mathbb{R}^p \mapsto \mathbb{R} : x \to \frac{1}{2}\|Ax - b\|^2$.*

*On s'intéresse au problème :*

$$(P) \qquad \min_{x \in \mathbb{R}^p} f(x)$$

1. *Montrez que (P) admet en moins une solution.*
2. *Justifiez la différentiabilité de f et donner son gradient.*
3. *Quelle est la relation entre le gradient et les dérivées partielles de f ?*
4. *f est-elle de classe $\mathcal{C}^2$ ? Si oui, déterminez da matrice Hessienne.*
5. *Étudier les points critiques de f.*
6. *Sous quelle condition (P) admet une solution unique ?*
7. *En statistiques, (P) sert à estimer quel type de modèle ? Donnez un exemple en précisant à quoi correspondent p, n, A et b.*
8. *Supposons que la condition (6) soit vérifiée. Proposez un algorithme pour résoudre (P).*

**Exercice 20** (Problèmes Inverses - Optim sous contraintes - dualité)**.** *On souhaite inférer une certaine grandeur physique $x \in \mathbb{R}^p$ étant donnés $n$ observations ou mesures $y \in \mathbb{R}^n$ supposés linéairement liés à $x$. Formellement : $y = Ax$ où $A \in \mathbb{R}^{n \times p}$ est supposée connue. Ce type de problème – dit problème inverse – est souvent mal conditionné, c'est à dire que $n < p$ et $A$ n'est donc pas de rang plein. On suppose dorénavant que $n \ll p$.*

1. *Quelle est la condition nécessaire sur $A$ et $y$ pour que l'ensemble des solutions soit non vide ?*

2. *Si cette condition est vérifiée, on a l'existence d'une solution $x_0$. Quel est l'ensemble des solutions en fonction de $x_0$ ?*
3. *On rencontre ce genre de problème en physique (où, par exemple on essaie de reconstruire des propriétés sur un emetteur radio avec des mesures à une certaine distance) – en imagerie médicale (voir Compressed sensing) – en statistiques en haute dimension (génomique : on a très peu d'observations sur les cellules, et on veut reconstruire l'expression génétique avec des dizaines de milliers de gènes ...).*

   *On suppose dorénavant qu'une solution existe. Quel problème se pose en pratique si on se contente de la solution donnée par un algorithme de minimisation de la fonction $\|Ax - y\|^2$ ?*
4. *Pour restreindre l'ensemble des solutions possibles. On procède souvent par régularisation : on privilégie les solutions les plus naturelles selon le problème donné. Une régularisation classique consiste à chercher la solution du problème avec la plus faible norme 2 :*

$$\begin{aligned} \min \quad & \|x\|_2^2 \\ & Ax = y \end{aligned}$$

   *Résoudre ce problème d'optimisation en résolvant les équations KKT.*
5. *Écrire le lagrangien du problème et le problème dual pour n'importe quelle norme $\|x\|_{2p}^{2p}$. Résoudre le cas général.*

**Exercice 21** (Optim sans contraintes - Matrix factorization)**.** *En Data science, on cherche souvent à compresser les données de manière à réduire la taille occupée en mémoire tout en conservant leur "essence". Ceci est particulièrement intéressant si les données contiennent une forme de redondance. Supposons qu'on souhaite compresser une matrice symétrique très large $A \in S_n$. On va chercher à trouver la meilleure approximation de $A$ ayant un rang égal à 1. Ceci est intéressant car toute matrice $B$ de rang 1 s'écrit comme un produit colonne $\times$ ligne : $B = xy^\top$ où $x, y \in \mathbb{R}^n$. Une matrice de rang 1 est donc caractérisée par $2n$ éléments au lieu de $n^2$. Si en plus $B$ est symétrique, alors il existe $x \in \mathbb{R}^n, \alpha \in \mathbb{R}$ tel que $|\alpha| = 1$ et $B = \alpha xx^\top$, et on passe donc à une caractérisation à $n$ + 1 éléments.*

*On munit $\mathbb{R}^{n,n}$ du produit scalaire $\langle A, B\rangle = \sum_{i,j}^n A_{ij}B_{ij} = \mathrm{trace}(A^\top B)$ et sa norme associée (norme de Frobenius) $\|A\|_F^2 = \langle A, A\rangle = \sum_{i,j}^n A_{ij}^2$.*

*On va chercher la meilleure compression symétrique de $A$ de rang 1 en minimisant :*

$$\min_{x\in\mathbb{R}^n} f(x) = \min_{x\in\mathbb{R}^n} \|A - xx^\top\|_F^2 \quad (P) \tag{27}$$

*On suppose dans le reste de l'énoncé que <u>$A$ est symétrique.</u>*

1. *Soit $B \in \mathbb{R}^{n,n}$. Montrez que rang$(B) = 1 \Leftrightarrow \exists x, y \in \mathbb{R}^n$ tel que $B = xy^\top$. Si en plus $B$ est symétrique, montrez qu'il existe $\alpha = \pm 1$ et $x \in \mathbb{R}^n$ tels que $B = \alpha xx^\top$.*
2. *Montrez que (P) admet une solution.*
3. *Simplifiez l'écriture de $f(x)$ puis déterminez le gradient de $f$ et sa Hessienne.*
4. *Montrez que toute solution du problème est soit le vecteur nul, soit le vecteur propre associée à la valeur propre maximale $\lambda_\infty$ et de norme $\sqrt{\lambda_\infty}$.*
5. *Montrez la réciproque de 4.*

**Exercice 22** (Quotient de Rayleigh.)**.** *On note $\|.\|$ la norme Euclidienne de $\mathbb{R}^n$. Soit $A \in S_n$ une matrice symétrique dans $\mathbb{R}^{n\times n}$. Le quotient de Rayleigh de $A$ en $x\mathbb{R}_*^n$ est défini par :*

$$q(x) = \frac{\langle Ax, x\rangle}{\|x\|^2}$$

*On s'intéresse aux points stationnaires de $q$ (et plus particulièrement à ses éventuels minimiseurs et maximiseurs globaux). En effet, comme le montre cet exercice, il existe un lien entre ses points stationneurs et les vecteurs / valeurs propres de $A$. Avec un algorithme d'optimisation numérique, on peut calculer (de manière approchée) les valeurs propres / vecteurs propres d'une matrice symétrique $A$.*

*Considérons les problèmes :*

$\min_x \frac{\langle Ax,x\rangle}{\|x\|^2}$ $(P_1)$ *et* $\min_{\|x\|^2=1}\langle Ax,x\rangle$ $(P_2)$

*Comme $A$ est symétrique, on peut noter ses valeurs propres dans un ordre croissant $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.*

*1. Dans quel sens les problèmes $P_1$ et $P_2$ sont-ils équivalents ?*

*2. Montrez que $P_1$ et $P_2$ admettent une solution.*

*3. Soit $(x, \lambda) \in \mathbb{R}^n \times \mathbb{R}$. Montrez que $x$ est un vecteur propre de norme 1 de $A$ associé à la valeur propre $\lambda$ si, et seulement si, $(x, \lambda)$ est un point stationnaire (solution KKT) de $(P_2)$. De plus, $\lambda$ est la valeur (de $f$) critique associée.*

*4. Montrez que les solutions de $P_2$ sont les vecteurs propres unitaires de $A$ de valeur propre minimale $\lambda_0$. En particulier,*

$$\langle Ax, x\rangle \geq \lambda_1 \|x\|^2$$

*5. On rappelle qu'on définit la norme d'opérateur de $A$ : par : $\|A\| = \sup_{\|x\|=1} \|Ax\|$.*

*Montrez que $\|A\| = \max_i |\lambda_i|$.*

*Et qu'en particulier si $A$ est semi-définie positive : $\|A\| = \lambda_n$.*

**Exercice 23** (Entrainement - Qualification des contraintes)**.** *Pour chaque ensemble $\mathcal{K}$ suivant, déterminez le sous-ensemble $\mathcal{K}' \subset \mathcal{K}$ des points où les contraintes sont qualifiées.*

*1.* $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x + y = 1, y + z \geq 4\}$

*2.* $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 \leq 1, y + z = 4\}$

*3.* $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 = 1, y^2 + z^2 = 4\}$

*4.* $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 = 1, y^2 + z^2 \leq 4\}$

# 6 Corrigés

*Merci à François-Pierre Paty pour les corrigés 10, 11 et 12.*

**Corrigé Exercice 10**

1. Résoudre

$$\min_{x^2+y^2=1} x + y.$$

(a) La fonction $f : (x, y) \mapsto x + y$ est continue, et l'ensemble $K = \{(x, y)\,,\, x^2 + y^2 = 1\}$ est compact, donc $f$ atteint ses bornes sur $K$, en particulier le minimum existe.

(b) La fonction $g : (x, y) \mapsto x^2 + y^2 - 1$ n'est pas affine, donc on ne peut pas appliquer directement les théorèmes de qualification. On utilise alors la condition de "liberté des gradients" : le point $(x, y)$ est qualifié dès lors que la famille $\{\nabla g(x, y)\}$ est libre, *i.e.* dès que $\nabla g(x, y) \neq 0$, soit $(2x, 2y) \neq (0, 0)$. En particulier, comme $(0, 0) \notin K$, tous les points satisfaisant la contrainte sont qualifiés pour KKT.

(c) On écrit le Lagrangien du problème : $\mathcal{L}(x, y, \lambda) = x + y + \lambda(x^2 + y^2 - 1)$. Le système KKT s'écrit alors :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x}(x, y, \lambda) = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(x, y, \lambda) = 0 \\ x^2 + y^2 = 1 \end{cases} \Leftrightarrow \begin{cases} 1 + 2\lambda x = 0 \\ 1 + 2\lambda y = 0 \\ x^2 + y^2 = 1 \end{cases} \Leftrightarrow \begin{cases} \lambda = \frac{-1}{2x} \\ y = x \\ x^2 + y^2 = 1 \end{cases} \Leftrightarrow \begin{cases} x = y = \frac{\sqrt{2}}{2} \\ \lambda = -\frac{\sqrt{2}}{2} \end{cases} \textbf{ ou } \begin{cases} x = y = -\frac{\sqrt{2}}{2} \\ \lambda = \frac{\sqrt{2}}{2} \end{cases}$$

Le système KKT possède donc deux solutions $(x, y, \lambda)$ : $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$ et $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$.

(d) En comparant les valeurs aux solutions du système KKT, on déduit que le minimum est atteint en un unique point $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$ et vaut $-\sqrt{2}$.

2. Résoudre

$$\min_{x+y=1} x^4 + y^4.$$

(a) La fonction $f : (x, y) \mapsto x + y$ est continue, l'ensemble $K = \{(x, y)\,,\, x + y = 1\}$ est fermé (mais pas compact). De plus, $f$ est coercive. En effet, quand $||(x, y)|| \to \infty$, $x^4 \to +\infty$ ou $y^4 \to +\infty$ si bien que $x^4 + y^4 \to +\infty$. Donc $f$ atteint son minimum.

(b) La fonction $g : (x, y) \mapsto x + y - 1$ est affine, et la contrainte est non vide. Donc tous les points sont qualifiés pour KKT.

(c) On écrit le Lagrangien du problème : $\mathcal{L}(x, y, \lambda) = x^4 + y^4 + \lambda(x + y - 1)$. Le système KKT s'écrit alors :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x}(x, y, \lambda) = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(x, y, \lambda) = 0 \\ x + y = 1 \end{cases} \Leftrightarrow \begin{cases} 4x^3 + \lambda = 0 \\ 4y^3 + \lambda = 0 \\ x + y = 1 \end{cases} \Leftrightarrow \begin{cases} 4x^3 + \lambda = 0 \\ x = y \\ x = 1 - y \end{cases} \Leftrightarrow \begin{cases} x = y = \frac{1}{2} \\ \lambda = -4 \times \frac{1}{8} = -\frac{1}{2} \end{cases}$$

Le système KKT possède donc une unique solution $(x, y, \lambda)$ : $\left(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)$.

(d) Le minimum est atteint en un unique point $\left(\frac{1}{2}, \frac{1}{2}\right)$ et vaut $\frac{1}{2^4} + \frac{1}{2^4} = \frac{1}{8}$.

3. Résoudre

$$\min_{x+2y=1} x^2 + y^2 + xy.$$

(a) La fonction $f : (x, y) \mapsto x^2 + y^2 + xy$ est continue, l'ensemble $K = \{(x, y)\,,\, x + 2y = 1\}$ est fermé (mais pas compact). De plus, $f$ est coercive. En effet, $f(x, y) = \frac{1}{2}(x + y)^2 + \frac{x^2+y^2}{2} \geq \frac{1}{2}\|(x, y)\|^2 \to +\infty$ quand $||(x, y)|| \to \infty$. Donc $f$ atteint son minimum.

(b) La fonction $g : (x, y) \mapsto x + 2y - 1$ est affine, et la contrainte est non vide. Donc tous les points sont qualifiés pour KKT.

(c) On écrit le Lagrangien du problème : $\mathcal{L}(x, y, \lambda) = x^2 + y^2 + xy + \lambda(x + 2y - 1)$. Le système KKT s'écrit alors :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x}(x, y, \lambda) = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(x, y, \lambda) = 0 \\ x + 2y = 1 \end{cases} \Leftrightarrow \begin{cases} 2x + y + \lambda = 0 \\ 2y + x + 2\lambda = 0 \\ x + 2y = 1 \end{cases} \Leftrightarrow \begin{cases} 2x + y + \lambda = 0 \\ 1 + 2\lambda = 0 \\ 2y = 1 - x \end{cases} \Leftrightarrow \begin{cases} 2x + y = \frac{1}{2} \\ \lambda = -\frac{1}{2} \\ 2y = 1 - x \end{cases} \Leftrightarrow \begin{cases} 2 - 3y = \frac{1}{2} \\ \lambda = -\frac{1}{2} \\ x = 1 - 2y \end{cases} \Leftrightarrow \begin{cases} y = \frac{1}{2} \\ \lambda = -\frac{1}{2} \\ x = 0 \end{cases}$$

Le système KKT possède donc une unique solution $(x, y, \lambda)$ : $\left(0, \frac{1}{2}, -\frac{1}{2}\right)$.

(d) Le minimum est atteint en un unique point $\left(0, \frac{1}{2}\right)$ et vaut $0^2 + \frac{1}{2^2} + 0 \times \frac{1}{2} = \frac{1}{4}$.

**Corrigé Exercice 11** Résoudre, pour $p \geq 1$ entier,

$$\min_{x^{2p}+y^{2p}=1} x^2 + y^2$$

et

$$\max_{x^{2p}+y^{2p}=1} x^2 + y^2.$$

1. La fonction $f : (x, y) \mapsto x^2 + y^2$ est continue, l'ensemble $K = \{(x, y)\,,\, x^{2p} + y^{2p} = 1\}$ est compact. Donc $f$ atteint ses bornes.
2. Soit $g : (x, y) \mapsto x^{2p} + y^{2p} - 1$ la fonction de contrainte. Comme il y a une seule contrainte d'égalité, un point $(x, y)$ sera qualifié pour KKT dès que $\nabla g(x, y) \neq 0$, *i.e.* dès que $(x, y) \neq (0, 0)$, ce qui est toujours le cas. Donc tous les points de la contrainte sont qualifiés.
3. On considère le cas $p > 1$. On écrit le Lagrangien du problème : $\mathcal{L}(x, y, \lambda) = x^2 + y^2 + \lambda(x^{2p} + y^{2p} - 1)$. Le système KKT s'écrit alors :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x}(x, y, \lambda) = 0 \\ \frac{\partial \mathcal{L}}{\partial y}(x, y, \lambda) = 0 \\ x^{2p} + y^{2p} = 1 \end{cases} \Leftrightarrow \begin{cases} 2x + 2p\lambda x^{2p-1} = 0 \\ 2y + 2p\lambda y^{2p-1} = 0 \\ x^{2p} + y^{2p} = 1 \end{cases} \Leftrightarrow \begin{cases} x\left(1 + p\lambda x^{2p-2}\right) = 0 \\ y\left(1 + p\lambda y^{2p-2}\right) = 0 \\ x^{2p} + y^{2p} = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} x = 0 \\ y = \pm 1 \\ \lambda = \dots \end{cases} \textbf{ ou } \begin{cases} x = \pm 1 \\ y = 0 \\ \lambda = \dots \end{cases} \textbf{ ou } \begin{cases} x = \pm y \\ 1 + p\lambda x^{2p-2} = 0 \\ x^{2p} = \frac{1}{2} \Leftrightarrow x = \pm\left(\frac{1}{2}\right)^{1/2p} \end{cases}$$

Le système KKT possède donc huit solutions $(x, y)$ : $(0, \pm 1), (\pm 1, 0), \left(\pm\left(\frac{1}{2}\right)^{1/2p}, \pm\left(\frac{1}{2}\right)^{1/2p}\right)$.

4. On compare les valeurs de l'objectif $f(x,y) = x^2 + y^2$ pour les couples $(x,y)$ solutions de KKT. On voit que $f(0,1) = f(0,-1) = f(1,0) = f(-1,0) = 1$ et que $f\left(\pm\left(\frac{1}{2}\right)^{1/2p}, \pm\left(\frac{1}{2}\right)^{1/2p}\right) = 2^{\frac{p-1}{p}} \geq 1$. Donc le minimum est atteint aux points $(0,\pm 1), (\pm 1, 0)$ et vaut 1, et le maximum est atteint aux points $\left(\pm\left(\frac{1}{2}\right)^{1/2p}, \pm\left(\frac{1}{2}\right)^{1/2p}\right)$ et vaut $2^{\frac{p-1}{p}}$. Si $p = 1$, $f$ est constante sur $K$ et tous les points sont des minimiseurs et des maximiseurs.

**Corrigé Exercice 12** Le problème suivant est-il convexe ? Écrire les conditions KKT.

$$\min_{Ax=b} \|x\|^2.$$

Le problème est convexe car les contraintes d'égalité sont affines et l'objectif $f(x) = \|x\|^2$ est une fonction convexe. **Attention, pour que le problème soit convexe, il faut que les contraintes soient AFFINES ! Avoir des contraintes d'égalité convexe ne suffit pas !**

On écrit le Lagrangien du problème : $\mathcal{L}(x,\lambda) = \|x\|^2 + \langle \lambda, Ax - b\rangle$. Les conditions KKT s'écrivent alors :

$$\begin{cases} \nabla_x \mathcal{L}(x,\lambda) = 0 \\ Ax = b \end{cases} \Leftrightarrow \begin{cases} 2x + A^\top \lambda = 0 \\ Ax = b \end{cases}$$

Dans le cas où $AA^\top$ est inversible, on peut alors résoudre le système KKT :

$$\begin{cases} 2x + A^\top \lambda = 0 \\ Ax = b \end{cases} \Leftrightarrow \begin{cases} 2x + A^\top \lambda = 0 \\ 2Ax + AA^\top \lambda = 0 \\ Ax = b \end{cases} \Leftrightarrow \begin{cases} 2x + A^\top \lambda = 0 \\ AA^\top \lambda = -2b \end{cases} \Leftrightarrow \begin{cases} x = -\frac{1}{2} A^\top \lambda \\ \lambda = -2\left(AA^\top\right)^{-1} b \end{cases} \Leftrightarrow \begin{cases} x = A^\top \left(AA^\top\right)^{-1} b \\ \lambda = -2\left(AA^\top\right)^{-1} b \end{cases}$$

**Exercice 13** Résoudre les problèmes d'optimisation suivants :

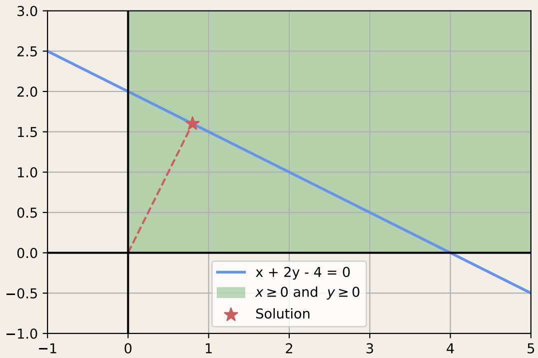
1. $\min\limits_{\substack{x\geq 0 \\ y \geq 0 \\ x+2y=4}} x^2 + y^2$
2. $\min\limits_{\substack{x\geq 0 \\ y \geq 0 \\ x+2y=4 \\ x^2+y^2 \leq 16}} x^2 + y^2$
3. $\min\limits_{\substack{x^2+y^2\leq 1 \\ x \geq 0}} x - y$
4. $\min\limits_{\substack{e^x+e^y\leq 20 \\ x \geq 0}} e^{x-y}$

1. **(problem 1)** Le problème s'écrit en forme standard : $\min\limits_{\substack{h_1(x,y)\leq 0 \\ h_2(x,y) \leq 0 \\ g(x,y)=0}} x^2 + y^2$

   avec $h_1(x,y) = -x$, $h_2(x,y) = -y$ et $g(x,y) = x + 2y - 4$.

**Intuition / Interprétation :**

Graphiquement, on cherche un $(x, y)$ appartenant à la droite d'équation $x+2y-4 = 0$ avec $x \geq 0$ et $y \geq 0$ qui minimise la norme $\|(x, y)\|^2$, autrement dit le point le plus proche de $(0, 0)$ vérifiant la contrainte. Il s'agit donc de la projection orthogonale de (0, 0) sur la droite bleue.




Les fonctions $h_1, h_2$ et $g$ sont affines, elles sont donc qualifiées pour tout $x, y \in \mathbb{R}$. On peut donc appliquer le théorème KKT :
Si $(x, y)$ est solution, alors il existe $\lambda, \mu_1, \mu_2 \in \mathbb{R}$ tels que :

$$\begin{cases} \nabla f(x,y) + \lambda \nabla g(x,y) + \mu_1 \nabla h_1(x,y) + \mu_2 \nabla h_2(x,y) = 0 \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ h_1(x,y)\mu_1 = 0 \\ h_2(x,y)\mu_2 = 0 \\ g(x,y) = 0 \\ h_1(x,y) \leq 0 \\ h_2(x,y) \leq 0 \end{cases} \tag{28}$$

$$\Rightarrow \begin{cases} \begin{cases} 2x + \lambda - \mu_1 = 0 \\ 2y + 2\lambda - \mu_2 = 0 \end{cases} \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ x\mu_1 = 0 \\ y\mu_2 = 0 \\ x + 2y - 4 = 0 \\ x \geq 0 \\ y \geq 0 \end{cases} \tag{29}$$

- si $\mu_1 > 0$ alors $x = 0$ et la contrainte d'égalité donne $y = 2$. Ainsi, on a forcément $\mu_2 = 0$ donc $\lambda = -2 < 0$ et $\mu_1 = \lambda = -2 < 0$; contradiction. A fortiori. on a $\mu_1 = 0$.
- si $\mu_2 > 0$ alors $y = 0$ et la contrainte d'égalité donne $x = 4$. Puis comme $\mu_1 = 0$, on obtient $\lambda = -8$ et $\mu_2 = 2\lambda < 0$; contradiction. A fortiori on a $\mu_2 = 0$.
- Comme $\mu_1 = \mu_2 = 0$, on peut à présent éliminer $\lambda$ en combinant les deux premières équations :

$$\begin{cases} 2x + \lambda = 0 \\ 2y + 2\lambda = 0 \\ x + 2y - 4 = 0 \end{cases} \Rightarrow \begin{cases} 4x + 2\lambda = 0 \\ 2y + 2\lambda = 0 \\ x + 2y - 4 = 0 \end{cases} \Rightarrow \begin{cases} 2x - y = 0 \\ x + 2y - 4 = 0 \end{cases} \tag{30}$$

$$\Rightarrow (x, y) = (\frac{4}{5}, \frac{8}{5}) \tag{31}$$

L'unique solution du sytème KKT est donc $(\frac{4}{5}, \frac{8}{5})$.
En revanche, comme $f : (x, y) \mapsto x^2 + y^2$ est convexe et les contraintes sont affines, le problème est convexe et KKT suffit pour garantir l'optimalitié. $(\frac{4}{5}, \frac{8}{5})$ est donc un minimiseur global.

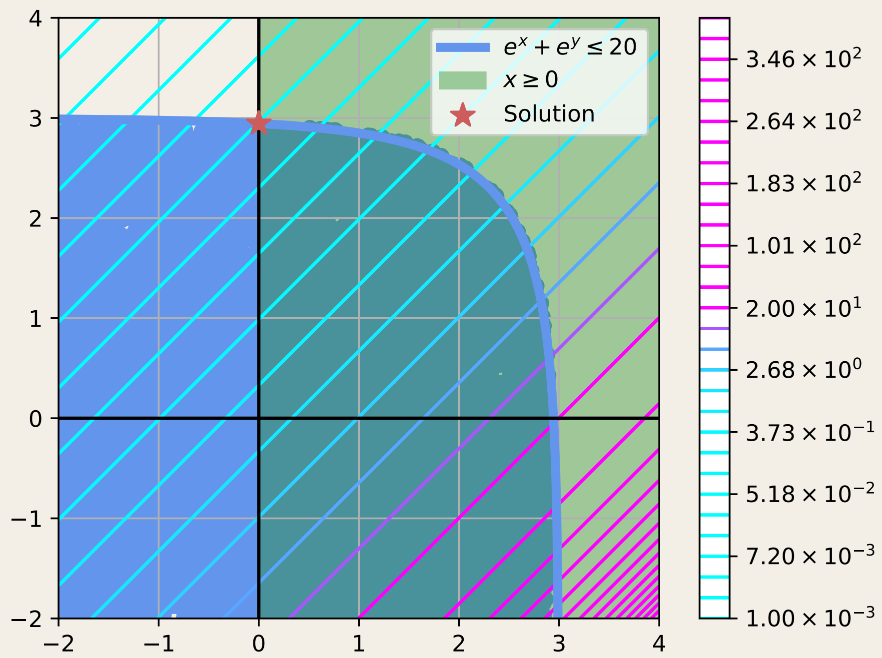
2. **(problème 2)**
3. **(problème 3)**

4. **(problem 4)** Le problème s'écrit en forme standard : $\min\limits_{\substack{h_1(x,y)\leq 0\\ h_2(x,y)\leq 0}} e^{x-y}$

avec $h_1(x,y) = e^x + e^y - 20$, $h_2(x,y) = -x$.

**Intuition / Interprétation :**

Graphiquement, on cherche à minimiser la fonction $(x,y) \mapsto e^{x-y}$ dont on représente les courbes de niveau ci-dessous. Comme il s'agit de l'exponentielle de la fonction linéaire $(x,y) \mapsto x-y$, l'ensemble des points où la fonction est constante est toujours une droite d'équation $y = x + cte$. Par ailleurs, pour minimiser $f$, il suffit de prendre $x \to -\infty$ et $y \to +\infty$ pour que $f$ tende vers 0 ($f$ n'est donc pas coercive). Sans contraintes, on voudrait donc aller le plus loin possible dans la direction $(-1,1)$ indéfiniment pour approcher la valeur minimale de 0. En revanche, l'ensemble des contraintes (intersection du vert et bleu) "coupe" la partie où f tend vers 0 indéfiniment. On s'attend donc à ce que le coin de l'ensemble des contraintes soit la solution. Les contraintes jouent alors un rôle dans le système KKT (on aura par conséquent des $\mu$ optimaux strictement positifs).




Les fonctions $h_1, h_2$ sont convexes. En effet, $h_2$ est linéaire donc convexe. En calculant les dérivées partielles de $h_1$, il s'avère que $h_1$ est de classe $\mathcal{C}^\infty$ et sa hessienne est donnée par : $\nabla^2 h_1(x,y) = \begin{pmatrix} e^x & 0 \\ 0 & e^y \end{pmatrix}$ qui est définie-positive pour tout $x, y$. D'après le critère de Slater, il suffit de trouver un $x, y$ tel que $x > 0$ et $e^x + e^y < 20$ ; en l'occurrence $(x,y) = (1,1)$ vérifie le critère de Slater. Les contraintes sont donc qualifiées en tout point. Si $(x,y)$ est solution, alors il existe $\mu_1, \mu_2 \in \mathbb{R}$ tels que :

$$\begin{cases} \nabla f(x,y) + \mu_1 \nabla h_1(x,y) + \mu_2 \nabla h_2(x,y) = 0 \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ h_1(x,y)\mu_1 = 0 \\ h_2(x,y)\mu_2 = 0 \\ h_1(x,y) \leq 0 \\ h_2(x,y) \leq 0 \end{cases} \tag{32}$$

$$\Rightarrow \begin{cases} \begin{cases} e^{x-y} + \mu_1 e^x - \mu_2 = 0 \\ -e^{x-y} + \mu_1 e^y = 0 \end{cases} \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ x\mu_1 = 0 \\ y\mu_2 = 0 \\ e^x + e^y - 20 \leq 0 \\ x \geq 0 \end{cases} \tag{33}$$

- Si $\mu_1 = 0$ alors $-e^{x-y} = 0$ ce qui est absurde. Donc $\mu_1 > 0$ et a fortiori la première contrainte est saturée : $e^x + e^y - 20 = 0$.
- De même, si $\mu_2 = 0$ alors $e^{-y} + \mu_1 = 0$ ce qui est impossible avec $\mu_1 \geq 0$ donc a fortiori on a $\mu_2 > 0$ et donc $x = 0$.

Les deux contraintes d'inégalités sont saturées, on s'atteint donc à une solution en coin (à la frontière de $K$). Il suffit de remplacer avec $x = 0$ dans $h_1(x,y) = 0$ pour obtenir $y = \log(19)$.

Finalement, on n'oublie pas de vérifier que $\mu_1, \mu_2$ sont positifs : on obtient avec $x = 0, y = \log(19)$ : $\mu_1 = e^{-2y} > 0$ et $\mu_2 = \mu_1 + e^{-y} > 0$.

L'unique solution du sytème KKT est donc $(0, \log(19))$.

En revanche, comme $f$ et les contraintes d'inégalité sont convexes, le problème est convexe et KKT suffit pour garantir l'optimalitié. $(0, \log(19))$ est donc un minimiseur global.

**Corrigé exercice 14** Soit $A \in \mathbb{R}^{m,n}$ et $b \in \mathbb{R}^m$. On s'intéresse au problème des moindres carrés positif : $\min\limits_{\substack{Ax=b \\ x\geq 0}} \|x\|^2$.

1. La fonction objective est clairement convexe (strictement convexe), en effet sa Hessienne est donnée par $2I_n$ qui est définie positive. Les contraintes sont linéaires donc il s'agit bien d'un problème convexe.
2. Tout d'abord il faut garder en tête que la contrainte d ;égalité $Ax = b$ exprime en fait $m$ contraintes d'égalité données par les lignes de $A$, on s'attend donc à avoir une variable duale $\lambda \in \mathbb{R}^m$. De mme, il y a $n$ contraintes d'inégalité données par $x \geq 0$, et donc une variable duale associée $\mu \in \mathbb{R}^n$. Formellement, on a $g(x) = Ax - b$ et $h(x) = -x$. En toute généralité, la première équation KKT s'écrit en fonction des Jacobiennes des contraintes :

$$\nabla f(x) + J_g(x)^\top \lambda + J_h(x)^\top \mu = 0$$

Pour s'en convaincre, il suffit de voir que le terme $i$ du vecteur $J_g(x)^\top \lambda$ est donné par $\sum_{j=1}^m \frac{\partial g_j}{\partial x_i}(x)\lambda_j = \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i}(x) = \sum_{j=1}^m \lambda_j [\nabla g_j(x)]_i = [\sum_{j=1}^m \nabla g_j(x)\lambda_j]_i$

En l'occurrence, on $\nabla f(x) = 2x$ et la Jacobienne d'une fonction linéaire est donnée par sa matrice (voir exemple 1), donc $J_g(x) = A$ et $J_h(x) = -I_n$. Ainsi on obtient le système KKT :

$$\begin{cases} 2x + A^\top \lambda - \mu = 0 \\ Ax - b = 0 \\ x \geq 0 \\ \mu \geq 0 \\ \mu^\top x = 0 \end{cases}$$

**Corrigé exercice 19**

**1** . (P) peut être réecrit :

$$(P) \qquad \min_{v \in \Im(A)} \|v - b\|^2$$

qui correspond à la projection orthogonale de $b$ sur l'image $\Im(A)$ – convexe et fermé (car sous-espace vectoriel de dimension finie) – donc (P) admet une solution.

**2** . Réponses possibles :

a) $f$ est une fonction quadratique / polynomiale, elle est donc différentiable (de Classe $\mathcal{C}^\infty$ même).

b) $f$ est différentiable comme composée de deux fonctions différentiables : une fonction linéaire et la norme Euclidienne au carré

et on a pour tout $h$ :

$$f(x + h) = f(x) + \langle A^\top (Ax - b), h\rangle + h^\top A^\top Ah$$

Comme $h^\top A^\top Ah = \|Ah\|^2 \leq \||A|\|^2 \|h\|^2 = o(h)$, le gradient de f est donné par :

$$\nabla f(x) = A^\top (Ax - b)$$

**3** . Le gradient de f en $x$ est le vecteur des dérivées partielles de f en $x$.

**4** . $f$ est polynomiale en $x$, elle est donc $\mathcal{C}^2$ OU BIEN Les dérivées partielles de f sont linéaires, elles sont donc de Classe $\mathcal{C}^1$ donc f est de Classe $\mathcal{C}^2$.

On a $\nabla f(x + h) = \nabla f(x) + A^\top Ah$ :

Comme $h \to A^\top Ah$ est linéaire continue (continue car en dimension finie toute application linéaire est continue), f est deux fois différentiable et sa hessienne est donnée par

$$\nabla^2 f(x) = A^\top A.$$

**5** . Les points critiques annulent le gradient de f. Ils sont donnés par $x$ tq :

$$A^\top (Ax - b) = 0 \tag{34}$$

En tout point critique $x$ et tout $h$ on a :

$$f(x + h) - f(x) = h^\top A^\top Ah = \|Ah\|^2 \geq 0$$

donc tout point critique est un minimiseur global de $f$.

Lorsque f est convexe, f a au plus une valeur minimale et tout point critique est solution, ici en effet, la hessienne de f est semi-définie positive pour tout $x$ (car $\forall x, x^\top A^\top Ax = \|Ax\|^2 \geq 0$), donc f est convexe, la condition d'optimailité $\nabla f(x) = 0$ est donc suffisante : toute solution de (34) est solution de (P).

**6** . Comme (P) admet une solution, elle est unique si (34) admet une solution unique, i.e si $A^\top A$ est inversible.

On peut également aller plus loin, et démontrer que $\mathrm{Ker}(A^\top A) = \mathrm{Ker}(A)$ :

$$x \in \mathrm{Ker}(A^\top A) \Rightarrow A^\top Ax = 0 \Rightarrow x^\top A^\top Ax = 0 \Rightarrow \|Ax\|^2 = 0 \Rightarrow Ax = 0 \Rightarrow x \in \mathrm{Ker}(A)$$

$$x \in \mathrm{Ker}(A) \Rightarrow Ax = 0 \Rightarrow A^\top Ax = 0 \Rightarrow x \in \mathrm{Ker}(A^\top A)$$

Donc il suffit que inversible $\mathrm{Ker}(A) = \{0\}$ (A de rang maximal).

De même, si f est strictement convexe, elle a au plus un **minimiseur** et si un point critique existe, alors il est l'unique solution. Comme $A^\top A$ est semi-définie positive, la condition $A^\top A$ inversible est équivalente à $A^\top A$ définie positive qui implique f strictement convexe.

**7** . (P) est un problème des moindres carrés. En stats, il peut éventuellement servir à estimer un modèle de régression linéaire de la variable observée $b$ sur les régresseurs donnés par les colonnes de $A$ : p correspond au nombre de variables, n au nombre d'observations.

Voici un example détaillé :

Let $b_1, \dots b_n$ correspondent aux nombres de coup-francs marqués par Del Piero [1] en $n = 200$ matchs. On souhaite expliquer sa performance en utilisant comme données : le classement de l'équipe adverse, la distance au gardien, la taille moyenne des joueurs dans le mur, l'heure du match, le nombre d'heures dormis la veille, ce qu'il a mangé avant whatever... Pour chaque variable $1 \leq j \leq p = 6$, on note l'observation du match $i$ par $A_{ij}$. Résoudre (P) correspond à chercher la fonction linéaire en les données $A_{.j}$ la plus "proche" de b, c'est à dire, les $x_j$ tel que $\phi_i = \sum_j A_{ij}x_j - b_i$ soit petit pour tout match $i$, on minimise donc $\sum_i \phi_i^2$ i.e (P).

En toute généralité, on suppose que les données observées suivent un modèle linéaire càd l'existence d'un certain $x$ tel que :

$$Ax = b + \varepsilon$$

où $\varepsilon$ est une variable aléatoire centrée modélisant le bruit des mesures / l'inexactitude du modèle. Minimiser la variance empirique du bruit $\varepsilon$ correspond exactement au problème (P).

**8** . Lorsque f est strictement convexe, résoudre (P) c'est résoudre (34) qui a une solution unique. Il suffit d'utiliser un algorithme pour résoudre un système linéaire comme le Pivot de Gauss.

**Corrigé exercice 21**

1. Soit $B$ une matrice de rang 1. Alors l'image de B est une droite engendrée par un vecteur $x \in \mathbb{R}^n$. Ainsi, pour tout $u \in \mathbb{R}^n$ il existe $y_u \in \mathbb{R}$ tel que $Bu = y_u x$. En particulier pour les vecteurs canonique de la base de $\mathbb{R}^n$ : $Be_i = y_i x$, $y_i \in \mathbb{R}$. Par conséquent les colonnes de $B$ sont donnés par $b_i = y_i x$, le terme général $B_{ij} = x_i y_j$ et donc $B = xy^\top$. Réciproquement, pour $B = xy^\top$ on a $Bu = xy^\top u = \langle u, y\rangle x$ pour tout $u \in \mathbb{R}^n$ donc l'image de $B$ est incluse dans la droite engendrée par $x$. Comme x est non nul, le rang de B est 1.
   Si B est symétrique de rang 1, alors par le théorème spectral il existe $P \in O_n$, base orthonormée de vecteurs propres de B notés $p_1, \dots, p_n$ tel que :
   $$B = \begin{pmatrix} p_1 & \dots & p_n \end{pmatrix} \begin{pmatrix} \lambda_1 & \dots & \dots \\ \dots & 0 & \dots \\ \dots & \dots & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ \dots \\ p_n \end{pmatrix} = \lambda_1 p_1 p_1^\top = \alpha x x^\top$$
   En posant $x = \sqrt{|\lambda_1|} p_1$.
2. On développant l'expression de f, on obtient :
   $$f(x) = \|x\|^4 - 2x^\top Ax + \|A\|_F^2 \geq \|x\|^4 - 2\lambda_n(A)\|x\|^2 + \|A\|_F^2.$$
   Le membre de droite est un polynôme en $\|x\|$, le terme $\|x\|^4$ domine lorsque $\|x\| \to +\infty$ donc $f$ est coercive.
3. On peut obtenir le gradient et la Hessienne de $f$ de deux façons. Soit écrire le développement différentiel de $f$ ou ses dérivées partielles. En passant par les dérivées partielles on a affaire à moins de calculs.
   En effet posons $\phi : x \to \|x\|^4 = \left(\sum_{i=1}^n x_i^2\right)^2$. Ainsi étant polynomiale en chaque $x_i$, elle admet des dérivées partielles données par : $\frac{\partial \phi}{\partial x_j}(x) = 4x_j\|x\|^2$. Ses dérivées partielles sont bien continues (polynomiales) donc $\phi$ est différentiable et son gradient est donné par $\nabla\phi(x) = 4\|x\|^2 x$.
   De même, chaque dérivée partielle est polynomiale et on a : $\frac{\partial^2 \phi}{\partial x_i x_j}(x) = \frac{\partial^2 \nabla\phi}{\partial x_i x_j}(x) = 4\|x\|^2\delta_{ij} + 8x_i x_j$ Où $\delta_{ij} = 1$ si $i = j$ et 0 sinon. Ainsi, comme ces fonctions sont bien continues, la Hessienne de $\phi$ est donnée par la matrice $\nabla^2\phi(x) = 4\|x\|^2 I_n + 8xx^\top$.
   Et enfin, en revenant à l'expression de $f$, on obtient :

1. Messi aurait été un choix trop facile

$$\nabla f(x) = 4\|x\|^2 x - 4Ax$$

et

$$\nabla^2 f(x) = 4\|x\|^2 I_n + 8xx^\top - 4A.$$

4. Les conditions d'optimalité vérifié par un minimiseur $x^\star$ sont le critère du premier ordre $\nabla f(x^\star) = 0 \Leftrightarrow Ax^\star = \|x^\star\|^2 x^\star$ ainsi que le critère du second ordre : pour tout $h$ on a $h^\top \nabla^2 f(x^\star) h \geq 0$. Ainsi, d'une part la solution $x^\star$ du problème est un vecteur propre de $A$ de valeur propre $\lambda = \|x^\star\|^2$. Et d'autre part on a pour tout $h$ : $4\|x^\star\|^2\|h\|^2 + 8(h^\top x^\star)^2 - 4h^\top Ah \geq 0$.
Donc d'abord pour des $h = h_1$ orthogonaux à $x^\star$ on obtient : $\|x^\star\|^2\|h_1\|^2 - h_1^\top Ah_1 \geq 0$. Par conséquent $\lambda\|h_1\|^2 \geq h_1^\top Ah_1$ pour tout $h_1$ tel que $h_1^\top x = 0$.
Et pour $h_0 = \alpha x^\star$, $\alpha \in \mathbb{R}$, la contrainte $\nabla f(x) = 0$ s'écrit également $\lambda\|h_0\|^2 = h_0^\top Ah_0$.
Donc dans les deux cas on a $\lambda\|h\|^2 \geq h^\top Ah$.
Soit $h \in \mathbb{R}^n$ quelconque. En utilisant une projection orthogonale sur la droite dirigée par $x^\star$, il existe $h_0 \in Vect(x^\star)$ et $h_1 \in Vect(x^\star)^\perp$ tel que $h = h_0 + h_1$. On a :
$$h^\top Ah = h_0^\top Ah_0 + 2h_1^\top Ah_0 + h_1^\top Ah_1 \leq \lambda(\|h_0\|^2 + \|h_1\|^2) + 2h_1^\top Ah_0$$
Or comme $h_0$ est colinéaire à $x^\star$, on a par la contrainte d'optimalité $Ax^\star \in Vect(x^\star)$ donc $Ah_0 \in Vect(x^\star)$ donc $h_1 \perp Ah_0$ et donc $h_1^\top Ah_0 = 0$
Ainsi, l'inégalité ci-dessus donne : $h^\top Ah \leq \lambda(\|h_0\|^2 + \|h_1\|^2) = \lambda\|h\|^2$. Où la dernière égalité est obtenue par Pythagore. On a donc :
$$(\forall h \in \mathbb{R}^n) \quad h^\top Ah \leq \lambda\|h\|^2 \quad (O)$$
Ainsi, comme $\lambda = \|x^\star\|^2 \geq 0$, on en déduit que toute solution du problème $x^\star$ vérifie :
Si $A \preceq 0$, alors la plus grande valeur propre de $A$ est négative ou nulle et donc $\lambda = 0$ donc $x^\star = 0$.
Sinon, le seul $\lambda$ vérifiant (O) est $\lambda_n(A)$ donc $\|x^\star\|^2 = \lambda_n(A)$ et $x^\star$ est un vecteur propre associé à $\lambda_n(A)$.
5. Comme la norme de $x^\star$ est fixé, on peut penser que ce dernier est bien défini et que l'on peut conclure par coercivité. Hélas, si $A \succ 0$ et $\lambda_n$ a une multiplicité $m > 1$, l'ensemble des $x^\star$ candidats est infini. Il faut donc une réciproque pour ce cas. Nous allons montrer que $f$ prend la même valeur peu importe le choix de $x^\star$. Comme $A$ est symétrique, il existe $p_1, \dots, p_n$ une base orthonormée de vecteurs propres associés aux valeurs propres $\lambda_1 \leq \dots \leq \lambda_m = \dots = \lambda_n$ de A. Comme toute solution $x^\star$ est un vecteur propre associé à $\lambda_n$ de multiplicité $m$, on peut écrire $x^\star$ dans la base $p_m, \dots p_n$ : $x^\star = \sum_{i=m}^n x_i p_i$ avec $\sum_{i=m}^n x_i^2 = \|x^\star\|^2 = \lambda_n$.
$$x^{\star\top} Ax^\star = x^{\star\top}(\sum_{i=1}^n \lambda_i p_i p_i^\top)x^\star = \sum_{i=m}^n \lambda_i x_i^2 = \lambda_n\|x^\star\|^2 = \lambda_n^2$$
Par ailleurs $\|x^\star\|^4 = \lambda_n^2$, ainsi on obtient :
$$f(x^\star) = \|A\|_F^2 - \lambda_n^2$$
f est constante indépendemment du choix de $x^\star$ vérifiant les conditions nécessaires d'optimalité. Par coercivité, tout $x^\star$ vérifiant ces contraintes est solution.

**Corrigé exercice 22**

1. Soit $\mathbf{x} \in \mathbb{R}^n$. On a $q(x) = (\frac{x}{\|x\|})^\top A(\frac{x}{\|x\|})$. Comme $\frac{x}{\|x\|}$ est de norme 1, les valeurs minimales des deux problèmes $\min_x q(x)$ et $\min_{\|x\|=1} x^\top Ax$ sont égales. Par ailleurs, si $x^\star$ est un minimiseur de$\min_{\|x\|=1} x^\top Ax$ alors, pour $t \in \mathbb{R}$, $tx^\star$ est un minimiseur de $q(x)$.
2. La contrainte $g(x) = \|x\|^2 - 1$ définit la sphère unité qui est un compact dans $\mathbb{R}^n$. f étant continue, le problème admet une solution par le théorème de Weirstrass.
3. La contrainte $g(x) = \|x\|^2 - 1$ est qualifiée sur la sphère unité car $\nabla g(x) = 2x \neq 0$, ainsi $\{\nabla g(x)\}$ est libre pour tout $x$ tq $\|x\| = 1$. Un point stationnaire $(x, \lambda)$ est un point qui vérifie le système KKT : $\nabla f(x) - \lambda\nabla g(x) = 0$ et $\|x\|^2 = 1$, ce qui équivaut à $Ax = \lambda x$ et $\|x\| = 1$ ou encore $x$ est un vecteur propre unitaire associé à la valeur propre $\lambda$. (Remarquez que je me suis permis d'écrire -$\lambda\nabla g(x)$ dans l'équation KKT car $\lambda$ étant dans $\mathbb{R}$, son signe est arbitraire). Dans ce cas, on a bien $x^\top Ax = \lambda\|x\|^2 = \lambda$. $\lambda$ est bien la valeur critique associée.

4. Cette inégalité peut s'obtenir trivialement en utilisation le théorème spectral. Ici en revanche, on souhaite la retrouver par un argument d'optimisation.

   Par la question 2), si $x^\star$ est solution du problème, alors il existe $\lambda^\star$ valeur propre associée au vecteur propre unitaire $x^\star$ tels que $x^{\star\top} A x^\star = \lambda^\star$. Et par optimalité, on a pour tout couple $(x, \lambda)$ stationnaire : $\lambda^\star = x^{\star\top} A x^\star \leq x^\top A x = \lambda$. Par conséquent, comme $\lambda$ est arbitraire, $\lambda^\star = \lambda_1$ (la plus petite valeur propre de $A$) et $x^\star$ est son vecteur propre unitaire associée. Réciproquement si $x^\star$ est le vecteur propre associé à $\lambda_1$, alors on a $x^{\star\top} A x^\star = \lambda_1$ qui est, par 2), la plus petite valeur prise par un point stationnaire. Comme une solution existe, c'est forcément la solution. Ainsi par définition de l'optimalité on a pour tout $x$ : $x^\top A x \geq \lambda_1 \|x\|^2$.
5. On a $\|A\|^2 = \max_{\|x\|=1} \|Ax\|^2 = \max_{\|x\|^2=1} x^\top A^\top A x = -\min_{\|x\|^2=1} x^\top (-A^\top A) x = -\lambda_1(-A^\top A) = -\lambda_1(-A^2) = \lambda_n(A^2)$.

   Or, comme $A$ est diagonalisable, les valeurs propres de $A^2$ sont les carrés des valeurs propres de $A$. Ainsi la plus grande valeur propre de $A^2$ est donnée par : $\lambda_n(A^2) = \max_i |\lambda_i(A)|$.

   Si A est semi-définie positive, alors tous les $\lambda_i(A)$ sont positives et $\max_i |\lambda_i(A)| = \lambda_n$.

**Corrigé exercice 23** L'ensemble des points où les contraintes sont qualifiées est aussi dit ensemble des points réguliers $\mathcal{K}'$.

1. $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x + y = 1, y + z \geq 4\}$

   Les deux contraintes sont affines, l'ensemble des points réguliers est donc l'ensemble tout entier $\mathcal{K}$.
2. $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 \leq 1, y + z = 4\}$

   La contrainte d'égalité est affine et la contrainte d'inégalité est convexe (en effet la Hessienne de $(x, y, z) \mapsto x^2 + y^2 - 1$ est donnée par $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ qui est semi-définie positive). Par la condition de Slater, comme $(x_0, y_0, z_0) = (0, 0.5, 3.5)$ vérifie la contrainte d'égalité et la contrainte d'inégalité strictement, l'ensemble des points réguliers est $\mathcal{K}$.
3. $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 = 1, y^2 + z^2 = 4\}$

   Les deux contraintes d'égalités ne sont pas affines, on ne peut donc pas utiliser la condition de Slater. Prenons $t = (x, y, z) \in \mathcal{K}$. $t$ est régulier si et seulement si la famille des gradients des contraintes en $t$ est libre. On a $\nabla g_1(t) = (2x, 2y, 0)^\top$ et $\nabla g_2(t) = (0, 2y, 2z)^\top$. On rappelle que $\{\nabla g_1(t), \nabla g_2(t)\}$ est libre ssi $\alpha_1 \nabla g_1(t) + \alpha_2 \nabla g_2(t) = 0 \Rightarrow \alpha_1 = \alpha_2 = 0 (*)$. Soient donc $\alpha_1, \alpha_2$ tels que $\alpha_1 \nabla g_1(t) + \alpha_2 \nabla g_2(t) = 0$ :

$$\alpha_1 \nabla g_1(t) + \alpha_2 \nabla g_2(t) = 0 \Rightarrow \begin{cases} \alpha_1 x = 0 (i) \\ (\alpha_1 + \alpha_2) y = 0 (ii) \\ \alpha_2 z = 0 (iii) \end{cases} \tag{35}$$

   - Si $x = 0$ alors par les contrainte d'égalité 1 et 2 $y^2 = 1$ et $z^2 = 3$. Donc par (ii) $\alpha_1 = -\alpha_2$ et par (iii) on a $\alpha_2 = 0$. Donc $\alpha_1 = 0$.
   - Sinon par (i) on a $\alpha_1 = 0$. Si $y = 0$ alors $z^2 = 4$ donc $\alpha_2 = 0$ ; si $y \neq 0$ alors $\alpha_2 = 0$.

   Dans tous les cas on obtient $\alpha_1 = \alpha_2 = 0$ pour tout $t \in \mathcal{K}$, donc $\mathcal{K}' = \mathcal{K}$.
4. $\mathcal{K} = \{(x, y, z) \in \mathbb{R}^3 \,|\, x^2 + y^2 = 1, y^2 + z^2 \leq 4\}$

   Nous avons une contrainte d'inégalité et une contrainte d'égalité non affine, on ne peut qu'utiliser la condition la plus générale du cours : Mangasarian-Fromovotiz. Pour $t \in \mathcal{K}$, MF est vérifiée si les deux conditions suivantes sont vérifiées :
   - la famille des gradients des contraintes d'égalité $\{\nabla g_i(t), i = 1 \dots E\}$ est libre
   - $\exists v$ tel que $\begin{cases} \langle v, \nabla g_i(t) \rangle = 0 \\ h_j(t) = 0 \Rightarrow \langle v, \nabla h_j(t) \rangle < 0 \end{cases} \quad \forall (i, j)$

   La difficulté de MF réside dans le fait qu'il faut discuter si les contraintes d'inégalités sont actives ou pas. Commençons par le premier point :

   **MF 1 : (Liberté des gradients)** le gradient de la contrainte d'égalité est donné $(2x, 2y, 0)^\top$ qui est nul uniquement pour $x = y = 0$, or pour tout $z$, $(0, 0, z)^\top \notin \mathcal{K}$, donc $\{\nabla_g(t)\}$ est libre pour tout $t \in \mathcal{K}$.

   **MF 2 :** Cherchons d'abord les $v = (v_1, v_2, v_3)^\top$ orthogonaux à la $\nabla g(t)$. On a $\langle v, \nabla g(t) \rangle = 0 \Rightarrow v_1 x + v_2 y = 0$ (*).

- Si la contrainte est inactive ($h(t) < 0$), alors il suffit de trouver un $v$ vérifiant (*). En l'occurrence, on peut prendre $(v_1, v_2, v_3) = (y, -x, 0)$.
- Si la contrainte est active ($h(t) = 0$), alors il faut trouver un $v$ vérifiant (*) ainsi que : $\langle v, \nabla h(t) < 0 \Rightarrow v_2 y + v_3 z < 0$ (**). Remarquons par ailleurs que comme $h(t) = 0$, on a forcément $z \neq 0$, en effet, si $z = 0$ alors $y^2 = 4$ qui ne peut pas vérifier la contrainte d'égalité. Dans ce cas, on peut prendre $(v_1, v_2, v_3) = (y, -x, -z + \frac{xy}{z})$ qui vérifie bien (*) et (**). En effet, $v_2 y + v_3 z = -xy - z^2 + xy = -z^2 < 0$.

Ainsi, MF est bien vérifiée pour tout $t \in \mathcal{K}$, $\mathcal{K}$ est donc régulier en entier.