# 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 $x^\star$ :

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

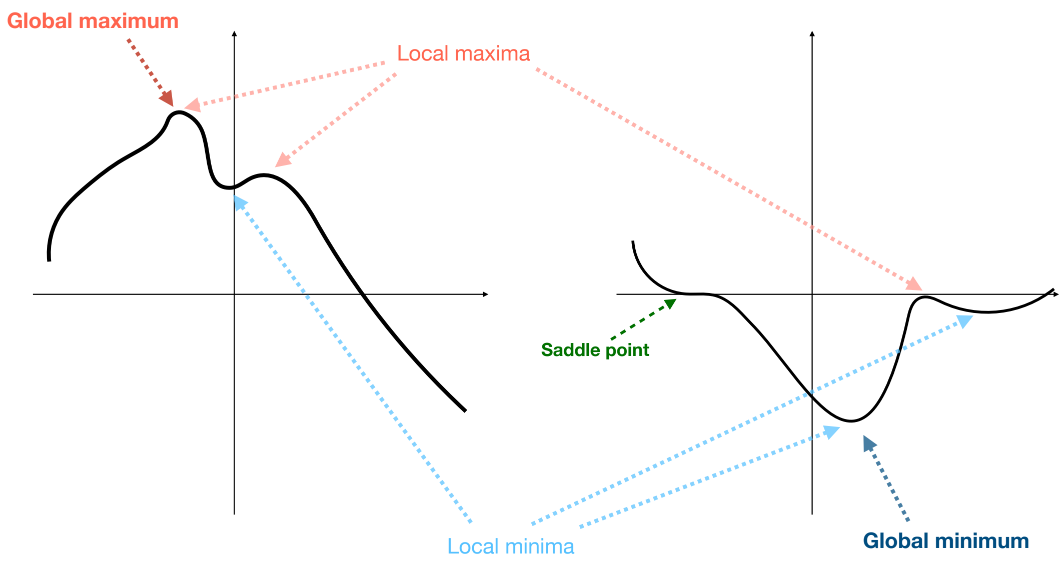



FIGURE 1 – Exemples de points critiques

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.

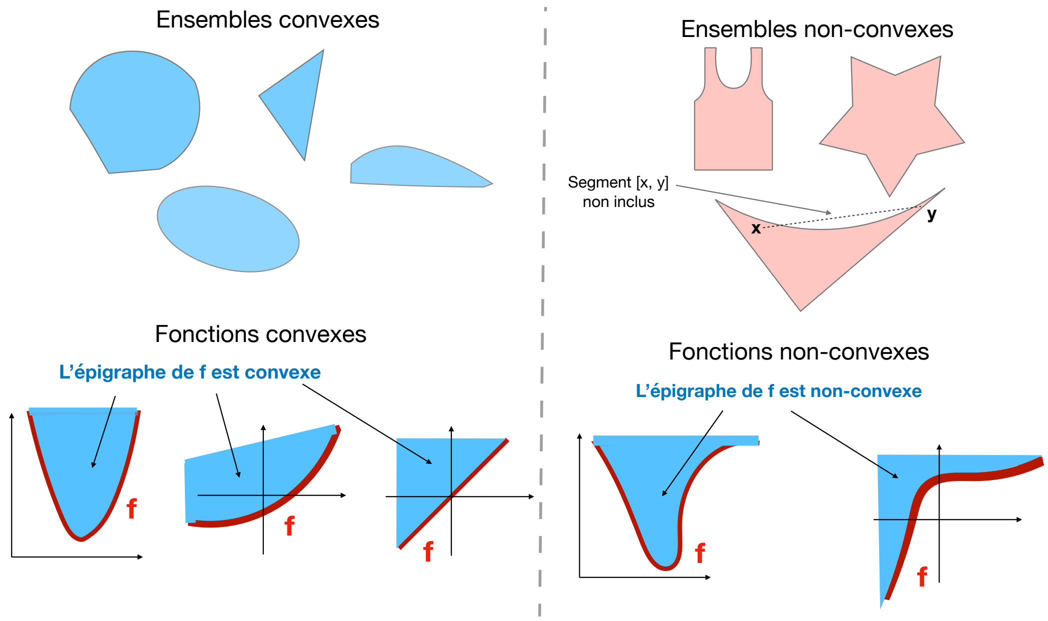



FIGURE 2 – Exemples d'ensembles et fonctions convexes

**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 :

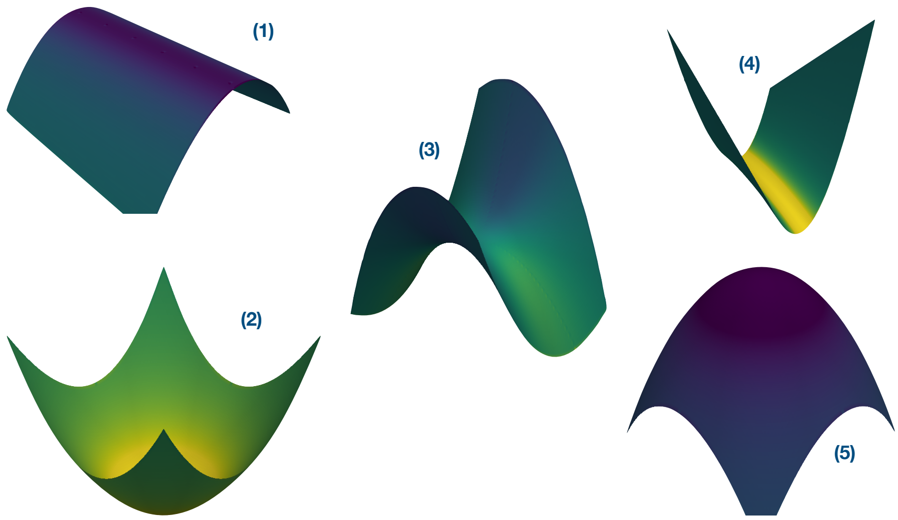



Figure 3 – Surfaces de fonctions quadratiques - Exercice 8

**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 $h \to h^\top H_f h$. 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 : $f(x) = \frac{1}{2}x^\top Sx + b^\top x$.*

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

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'indépendence 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 :*

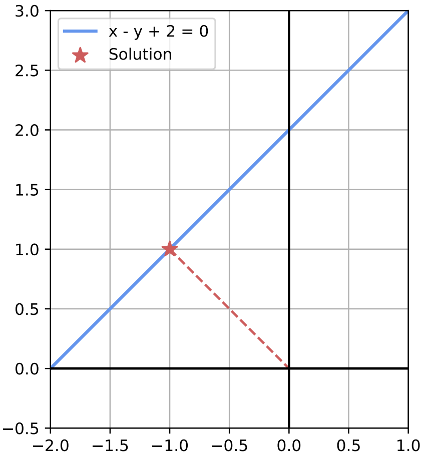
$$\begin{aligned} &\begin{cases} \nabla f(x) + J_g(x)^\top \lambda = 0 \\ g(x) = 0 \end{cases} \\ \Leftrightarrow &\begin{cases} \nabla f(x) + \sum_{i=1}^E \lambda_i \nabla g_i(x) = 0 \\ g(x) = 0 \end{cases} \end{aligned} \tag{14}$$

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 :*

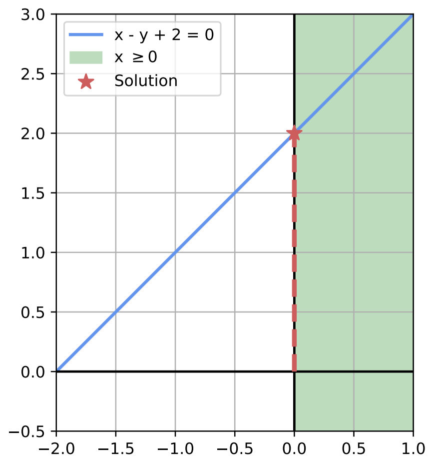
$$\begin{cases} \nabla f(x) + \sum_{i=1}^{E} \lambda_i \nabla g_i(x) + \sum_{i=1}^{I} \mu_i \nabla h_i(x) = 0 \\ g(x) = 0 \\ h(x) \leq 0 \\ \mu \geq 0 \\ h_i(x)\mu_i = 0 \quad 1 \leq i \leq I \end{cases} \tag{20}$$

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 :*

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

$$\Rightarrow \begin{cases} \begin{cases} 2x + \lambda - \mu = 0 \\ 2y - \lambda = 0 \end{cases} \\ \mu \geq 0 \\ x\mu = 0 \\ x - y + 2 = 0 \\ x \geq 0 \end{cases} \tag{23}$$

$$\tag{24}$$

*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 $(P)$ $\min\limits_{\substack{g(x)=0 \\ h(x) \leq 0}} f(x)$ 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 au min sur $x \in K$ puis au max sur $(\lambda, \mu)$ :

**Exercice 15.** *Soit $A \in \mathbb{R}^{n\times n}$ symétrique définie positive et $b \in \mathbb{R}^n$.*

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

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

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

*4. Donner le problème dual de (P).*

# 4 Corrigés

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

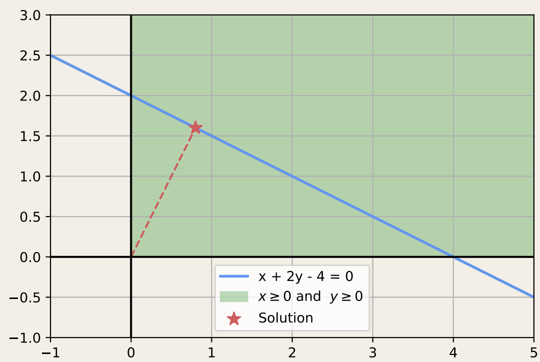
1. $\min\limits_{\substack{x\geq 0\\ y\geq 0\\ x+2y=4}} x^2+y^2$
2. $\min\limits_{\substack{x\geq 0\\ y\geq 0\\ x+2y=4\\ x^2+y^2\leq 16}} x^2+y^2$
3. $\min\limits_{\substack{x^2+y^2\leq 1\\ x\geq 0}} x-y$
4. $\min\limits_{\substack{e^x+e^y\leq 20\\ x\geq 0}} e^{x-y}$

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 :

$$\begin{cases} \nabla f(x,y) + \lambda \nabla g(x,y) + \mu_1 \nabla h_1(x,y) + \mu_2 \nabla h_2(x,y) = 0 \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ h_1(x,y)\mu_1 = 0 \\ h_2(x,y)\mu_2 = 0 \\ g(x,y) = 0 \\ h_1(x,y) \leq 0 \\ h_2(x,y) \leq 0 \end{cases} \tag{25}$$

$$\Rightarrow \begin{cases} \begin{cases} 2x + \lambda - \mu_1 = 0 \\ 2y + 2\lambda - \mu_2 = 0 \end{cases} \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ x\mu_1 = 0 \\ y\mu_2 = 0 \\ x + 2y - 4 = 0 \\ x \geq 0 \\ y \geq 0 \end{cases} \tag{26}$$

- 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 :

$$\begin{cases} 2x + \lambda = 0 \\ 2y + 2\lambda = 0 \\ x + 2y - 4 = 0 \end{cases} \Rightarrow \begin{cases} 4x + 2\lambda = 0 \\ 2y + 2\lambda = 0 \\ x + 2y - 4 = 0 \end{cases} \Rightarrow \begin{cases} 2x - y = 0 \\ x + 2y - 4 = 0 \end{cases} \tag{27}$$

$$\Rightarrow (x,y) = (\frac{4}{5}, \frac{8}{5}) \tag{28}$$

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.

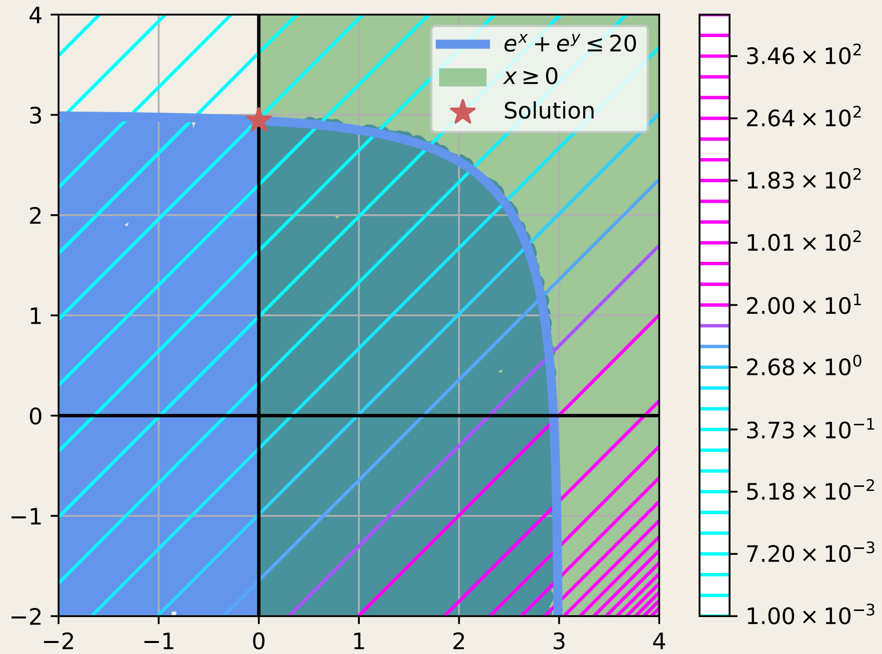
2. **(problème 2)**
3. **(problème 3)**
4. **(problem 4)** Le problème s'écrit en forme standard : $\min\limits_{\substack{h_1(x,y) \leq 0 \\ h_2(x,y) \leq 0}} e^{x-y}$

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 $y = x + cte$. On voit qu'en l'occurrence lorsque x - y est très petit, $f$ tend vers 0 (x, y tous les deux très petits par ex). $f$ n'est donc pas coercive. Comme on minimise $f$, on veut aller le plus loin possible dans la direction $(-1, 1)$ indéfiniment. En revanche, l'ensemble des contraintes (représenté en vert et bleu) "coupe" la partie où f tend vers 0 indéfiniment. On s'attend donc à ce que les contraintes jouent un rôle dans le système KKT (on aura une solution à la frontière, représentée en rouge).

**Intuition / Interprétation :**




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 : $\nabla_2(x,y) = \begin{pmatrix} e^x & 0 \\ 0 & e^y \end{pmatrix}$ 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 :

$$\begin{cases} \nabla f(x,y) + \mu_1 \nabla h_1(x,y) + \mu_2 \nabla h_2(x,y) = 0 \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ h_1(x,y)\mu_1 = 0 \\ h_2(x,y)\mu_2 = 0 \\ h_1(x,y) \leq 0 \\ h_2(x,y) \leq 0 \end{cases} \tag{29}$$

$$\Rightarrow \begin{cases} \begin{cases} e^{x-y} + \mu_1 e^x - \mu_2 = 0 \\ -e^{x-y} + \mu_1 e^y = 0 \end{cases} \\ \mu_1 \geq 0 \\ \mu_2 \geq 0 \\ x\mu_1 = 0 \\ y\mu_2 = 0 \\ e^x + e^y - 20 \leq 0 \\ x \geq 0 \end{cases} \tag{30}$$

- 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 $y = \log(19)$.

L'unique solution du sytème KKT est donc $(\frac{4}{5}, \frac{8}{5})$.
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.