Introduction

Le but de ce cours est d'étudier des problèmes d'optimisation du type

$$\inf_{x \in K} f(x),$$

où f est une fonction (suffisamment régulière) de $\mathbb{R}^n$ à valeur dans $\mathbb{R}$ et $K \subset \mathbb{R}^n$. Le choix de K traduit l'existence de contrainte sur le problème d'optimisation (argument bornés par exemple). On appelle K le domaine admissible du problème d'optimisation.

Ce type de problème correspond à de nombreuses question rencontrée en mathématiques mais aussi dans la vie "quotidienne" :

- Régression (linéaire ou polynomiale)
- Estimation paramétrique et maximum de vraisemblance
- Inégalités fonctionnelles (théorie spectrale)
- Maximisation d'un volume à surface fixée
- Problème isopérimétrie

Les questions que nous allons étudier dans ce cours sont les suivantes :

1. Y a-t-il existence d'une solution au problème d'optimisation ? Unicité ?
2. S'il y a existence d'une solution, comment caractériser celle-ci, voire un calcul explicite ?
3. Comment approcher les solutions de manière numérique ? Avec quelle efficacité (terminaison, vitesse) ?

Chapitre 1

Rappel et complément de calcul différentiel

1.1 Calcul différentiel

1.1.1 Fonction d'une variable

On considère une fonction $f : \Omega \to \mathbb{R}^p$ où Ω est un ouvert de $\mathbb{R}$.

Définition 1.1.1. *On dit que f est dérivable au point $x \in \Omega$ si le taux d'accroissement*

$$\frac{f(x+h) - f(x)}{h}$$

admet une limite finie quand $h \to 0$. Cette limite est alors notée $f'(x) \in \mathbb{R}^p$. Si f est dérivable sur tout Ω, on définit alors une fonction $f' : \Omega \to \mathbb{R}^p$, la dérivée de f. Si enfin f' est continue sur Ω, on dit que f est de classe $\mathcal{C}^1$.

La formule de Taylor-Young apporte une autre formulation de la définition de dérivée, plus adaptée au cadre général en dimension supérieure. La fonction f est dérivable au point x si il existe $f'(x)\mathbb{R}^p$ tel que

$$f(x+h) = f(x) + f'(x)h + o(h).$$

Cette formulation traduit la dérivée comme pente de la meilleure approximation affine de f au point x : $T_x : h \mapsto f'(x)h + f(x)$.

Remarque. *Si on note $f(x) = \begin{pmatrix} f_1(x) \\ \vdots \\ f_p(x) \end{pmatrix} \in \mathbb{R}^p$, on remarque que f est dérivable en x si est seulement si tout les f_i le sont et dans ce cas $f'(x) = \begin{pmatrix} f_1'(x) \\ \vdots \\ f_p'(x) \end{pmatrix} \in \mathbb{R}^p$.*

Exemple 1.1.2. *La fonction $f : x \mapsto \begin{pmatrix} x^2 \\ e^x - x \end{pmatrix}$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et pour tout $x \in \mathbb{R}$, on a : $f'(x) = \begin{pmatrix} 1 \\ e^x - 1 \end{pmatrix}$*

Remarque.
- *Dérivable en x implique continue en x.*
- *Dérivable sur Ω N'IMPLIQUE PAS de classe $\mathcal{C}^1$ sur Ω. Le contre exemple classique est la fonction*

$$f : x \in \mathbb{R} \mapsto \begin{cases} x^2 \sin(x) & si x \neq 0 \\ 0 & sinon \end{cases}$$

qui est dérivable sur $\mathbb{R}$ mais dont la dérivée n'admet pas de limite en 0.

1.1.2 Fonctions à plusieurs variables

On considère une fonction $f : \Omega \to \mathbb{R}^p$ où Ω est un ouvert de $\mathbb{R}^n$. Pour $x = (x_1, \ldots, x_n) \in \Omega$, on note

$$f(x_1, \ldots, x_n) = \begin{pmatrix} f_1(x1, \ldots, x_n) \\ \vdots \\ f_p(x_1, \ldots, x_n) \end{pmatrix} \in \mathbb{R}^p.$$

On cherche toujours à décrire le comportement locale de la fonction f autour d'un point x, c'est à dire ses variations lorsque l'on s'éloigne du point x. Contrairement au cas unidimensionnel, le déplacement est ici un vecteur. On ne peut donc pas généraliser la définition 1.1.1 sans tenir compte du choix de direction.

Définition 1.1.3. *Soit $v \in \mathbb{R}^n$. On dit que f est dérivable en un point $x \in \Omega$ selon la direction v si le taux d'accroissement*

$$\frac{f(x+tv) - f(x)}{t}$$

admet une limite finie quand $t \to 0$. Dans ce cas, cette limite est appelée dérivée directionnelle de f en x selon le vecteur v et est notée $D_v f(x)$.

Remarque. *La fonction f est dérivable en x selon v si et seulement si la fonction $\varphi : t \mapsto f(x+tv)$ est dérivable en 0 et on a $D_v f(x) = \varphi'(0)$.*

Dans $\mathbb{R}^n$, on dispose d'une base canonique, notée $(e_i)_{1 \leq i \leq n}$. La dérivée directionnelle $D_{e_i} f$ est appelée la dérivée partielle selon la i-ème coordonnée et est notée $\partial_i f$ ou $\frac{\partial f}{\partial x_i}$. Par définition, on a

$$\partial_i f(x) = \lim_{t \to 0} \frac{f(x_1, \ldots, x_i + t, \ldots, x_n) - f(x_1, \ldots, x_n)}{t}$$

ce qui revient à dire que $\partial_i f$ est calculée en gelant les variables $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$ et en dérivant par rapport à la variable x_i.

Exemple 1.1.4. *Soit $f : (x_1, x_2) \in \mathbb{R}^2 \mapsto \begin{pmatrix} x_1^2 + 3x_2 \\ x_1 x_2 \end{pmatrix}$. Les dérivées partielles existent en tout point et on a : $\partial_1 f(x) = \begin{pmatrix} 2x_1 \\ x_2 \end{pmatrix}$ et $\partial_2 f(x) = \begin{pmatrix} 3 \\ x_1 \end{pmatrix}$.*

La dérivée directionnelle ne décrit le comportement de f que le long d'un vecteur. Si l'on cherche une approximation affine de f, un nombre ne va plus suffire, il faut une application affine de $\mathbb{R}^n$ à valeur dans $\mathbb{R}^p$. L'application linéaire sous-jacente est la différentielle de f.

Définition 1.1.5. *On dit que f est différentiable en x s'il existe une application linéaire, notée $df_x : \mathbb{R}^n \to \mathbb{R}^p$ telle que*

$$f(x+h) = f(x) + df_x(h) + o(\|h\|).$$

Dans ce cas cette application linéaire est unique et on l'appelle différentielle de f au point x.

Remarque. *Une fonction différentiable en un point est une fonction qui vérifie une formule de Taylor-Young, à l'ordre 1, en ce point.*

On appelle jacobienne de f en x, notée $J_f(x) \in \mathcal{M}_p, n(\mathbb{R})$ la matrice associée :

$$df_x(h) = J_f(x).h, \, \forall h \in \mathbb{R}^n.$$

Proposition 1.1.6. *Si f est diférentiable en $x \in \Omega$ alors f est dérivable selon toutes les directions en ce point et on a $D_v f(x) = df_x(v)$. En particlulier, la matrice a pour coefficients*

$$J_f(x) = \left(\partial_1 f(x) \middle| \cdots \middle| \partial_p f(x)\right) = \begin{pmatrix} \partial_1 f_1(x) & \dots & \partial_n f_1(x) \\ \vdots & & \vdots \\ \partial_1 f_p(x) & \dots & \partial_n f_p(x) \end{pmatrix}$$

Remarque. *La réciproque est fausse. Soit $f : x \in \mathbb{R}^2 \mapsto \begin{cases} 1 & si x_2 = x_1^2 \neq 0 \\ 0 & sinon \end{cases}$. Cette fonction n'est pas continue et a fortiori pas différentiable en $(0,0)$ mais elle dérivable dans toutes les directions en ce point.*

Il existe cependant une réciproque sous des hypothèses supplémentaire.

Proposition 1.1.7. *Si toutes les dérivées partielles de f en x existent et si de plus les applications $\partial_i f : \Omega \to \mathbb{R}^p$ sont continues alors f est différentiable en x.*

Définition 1.1.8. *Si les dérivées partielles existent en tout point de Ω et sont continues, on dit alors que f est de classe $\mathcal{C}^1$ sur Ω.*

Exemple 1.1.9. *Toutes les fonctions polynomiales sont de classe $\mathcal{C}^1$. La fonction de l'exemple 1.1.4 est ainsi de classe $\mathcal{C}^1$ et on a : $\begin{pmatrix} 2x_1 & 3 \\ x_2 & x_1 \end{pmatrix}$.*

Proposition 1.1.10 (Différentiation de la composée)**.** *Si $f : \mathbb{R}^n \to \mathbb{R}^p$ est différentiable en x et $g : \mathbb{R}^p \to \mathbb{R}^q$ est différentiable en $f(x)$ alors $g \circ f : \mathbb{R}^n \to \mathbb{R}^q$ est différentiable en x et on a : $d(g \circ f)_x = dg_{f(x)} \circ df_x$.*

Remarque. *En terme de jacobienne, on a*

$$J_{g\circ f}(x) = \underbrace{J_g(f(x))}_{\in \mathcal{M}_{q,p}(\mathbb{R})} \underbrace{J_f(x)}_{\in \mathcal{M}_{p,n}(\mathbb{R})} \in \mathcal{M}_{q,n}(\mathbb{R}).$$

Exemple 1.1.11. *Soit $g : \mathbb{R}^n \to \mathbb{R}^p$ de classe $\mathcal{C}^1$. Soient $x, v \in \mathbb{R}^n$ et $ft \in \mathbb{R} \mapsto x + tv$, on pose $\varphi : g \circ f$. Alors φ est de classe $\mathcal{C}^1$ et on a*

$$\varphi'(t) = J_{g\circ f}(t) = J_g(x + tv) J_f(t) = J_g(x + tv)v.$$

1.1.3 Le cas de fonction numérique

On s'intéresse au cas particulier $p = 1$, c'est à dire $f : \mathbb{R}^n \to \mathbb{R}$.

Définition 1.1.12 (Gradient)**.** *Si f est différentiable en $x \in \Omega$, on appelle gradient de f en x l'unique vecteur $\nabla f(x) \in \mathbb{R}^n$ tel que pour tout $h \in \mathbb{R}^n$, $df_x(h) = \langle \nabla f(x), h \rangle$.*

Le vecteur $\nabla f(x)$ est le dual de la forme linéaire df_x. On a ainsi : $\nabla f(x) = J_f(x)^t$, soit en coordonnées

$$\nabla f(x) = \begin{pmatrix} \partial_1 f(x) \\ \vdots \\ \partial_n f(x) \end{pmatrix}.$$

La formule de Taylor-Young au point x s'écrit alors

$$f(x + h) = f(x) = \langle \nabla f(x), h \rangle + o(\|h\|).$$

Rappel. Le produit scalaire sur $\mathbb{R}^n$ est définit comme suit : $\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^t y$ pour tout $x, y \in \mathbb{R}^n$.

Exemple 1.1.13. *1. La fonction $f : x \in \mathbb{R}^2 \mapsto x_1^3 + x_1x_2^2 + 3x_2$ est de classe $\mathcal{C}^1$ et pour tout $x \in \mathbb{R}^n$, $\nabla f(x) = \begin{pmatrix} 3x_1^2 + x_2^2 \\ 2x_1x_2 + 3 \end{pmatrix}$.*

2. Si $b \in \mathbb{R}^n$, la fonction $f : x \in \mathbb{R}^n \mapsto \langle b, x\rangle$ est de classe $\mathcal{C}^1$ et pour tout $x \in \mathbb{R}^n$, $\nabla f(x) = b$.

3. La fonction $f : x \in \mathbb{R}^n \mapsto \|x\|^2$ est $\mathcal{C}^1$ et $\nabla f(x) = 2x$.

4. Plus généralement, si $A \in \mathcal{S}_n(\mathbb{R})$ la fonction $f : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle Ax, x\rangle$ est $\mathcal{C}^1$ et $\nabla f(x) = Ax$.

Définition 1.1.14 (Hessienne)**.** *Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$. On dit que f est deux fois différentiable en $x \in \Omega$ si ∇f est différentiable en x. Dans ce cas, ∇f admet des dérivées partielles, que l'on note $\partial_i(\partial_j f)(x) = \partial^2_{ij} f(x)$ et appelées dérivées partielles secondes en x. On appelle hessienne de f en x la jacobienne de ∇f en x que l'on note $H_f(x)$ ou $\nabla^2 f(x)$.*

La hessienne de f est donc une matrice de $\mathcal{M}_n(\mathbb{R})$ dont les coordonnées sont

$$\nabla^2 f(x) = \begin{pmatrix} \partial^2_{11} f(x) & \dots & \partial^2_{n1} f(x) \\ \vdots & & \vdots \\ \partial^2_{1n} f(x) & \dots & \partial^2_{nn} f(x) \end{pmatrix}.$$

On dit que f est de classe $\mathcal{C}^2$ si f deux fois différentiable et que ∇f est de classe $\mathcal{C}^1$. En d'autre terme, f est de classe $\mathcal{C}^2$ si les dérivées partielle $\partial_1 f$ admettent des dérivées partielles continues ou encore si les dérivées secondes sont des fonctions continues.

Exemple 1.1.15. *1. Les fonction polynomiales sont de classe $\mathcal{C}^2$.*

2. La fonction $f : x \in \mathbb{R}^2 \mapsto x_1^3 + x_1x_2^2 + 3x_2$ est de classe $\mathcal{C}^2$ et on a $\nabla^2 f(x) = \begin{pmatrix} 6x_1 & 2x_2 \\ 2x_2 & 2x_1 \end{pmatrix}$.

3. Soient $A \in \mathcal{S}_n(\mathbb{R})$ et $b \in \mathbb{R}^n$. La fonction $f : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle$ est de classe $\mathcal{C}^2$ et on a $\nabla^2 f(x) = Ax + b$.

Théorem 1.1.16 (Schwartz)**.** *Si f est deux fois différentiable en $x \in \Omega$ alors pour tout ij on a $\partial^2_{ij} f(x) = \partial^2_{ji} f(x)$. En d'autres termes, la matrice $\nabla^2 f(x)$ est symétrique.*

Remarque. *Attention, si f n'est pas deux fois différentiables tout en ayant des dérivées partielles secondes alors le résultats ne tient plus.*

Proposition 1.1.17 (Formule de Taylor-Young)**.** *Si $f : \Omega \to \mathbb{R}$ est deux fois différentiable en $x \in \Omega$, alors*

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

1.2 Fonctions convexes

1.2.1 Définitions

Définition 1.2.1. *On dit que $C \subset \mathbb{R}^n$ est un ensemble convexe si pour tout $x, y \in C$ le segment $[x, y]$ est inclus dans C i.e.*

$$\forall x, y \in C, \forall \lambda \in [0, 1], \lambda x + (1 - \lambda)y \in C.$$

Définition 1.2.2. *Soit $C \subset \mathbb{R}^n$ un ensemble convexe. On dit que $f : C \to \mathbb{R}$ est*

— *convexe si pour tout $x, y \in C$ et $\lambda \in [0, 1]$,*

$$f(\lambda x + (1 - \lambda)y \leq \lambda f(x) + (1 - \lambda)f(y).$$

— *strictement convexe si pour tout $x \neq y \in C$ et $\lambda \in]0,1[$,*

$$f(\lambda x + (1-\lambda)y < \lambda f(x) + (1-\lambda)f(y).$$

— *fortement convexe s'il existe $\alpha > 0$ tel que pour tout $x, y \in C$ et $\lambda \in [0,1]$,*

$$f(\lambda x + (1-\lambda)y \leq \lambda f(x) + (1-\lambda)f(y) - \frac{\alpha}{2}\lambda(1-\lambda)\|y-x\|^2.$$

Si f est fortement convexe pour une constante α on dira que f est α-convexe.

Remarque. *fortement convexe $\Rightarrow$ strictement convexe $\Rightarrow$ convexe.*

Exemple 1.2.3. — *Les fonctions affines sont convexes mais pas strictement convexes.*
— *$x \in \mathbb{R} \mapsto x^2$ et $x \in \mathbb{R} \mapsto e^x$ sont strictement convexes.*

1.2.2 Le cas différentiable

Soient $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ une fonction différentiable sur l'ouvert Ω, et $\mathcal{C}$ un convexe inclus dans Ω. Dans le cas de la dimension 1, on sait que si f est dérivable alors f est convexe si et seulement si sa dérivée f' est croissante ce qui se reformule ainsi : pour tout $x, y\mathcal{C}$

$$f'(x) - f'(y))(x-y) \geq 0.$$

Cette caractérisation se généralise en dimension supérieure.

Proposition 1.2.4 ("dérivée croissante")**.**

$$\begin{aligned} f \text{ convexe} &\Leftrightarrow \langle \nabla f(x) - \nabla f(y), x-y\rangle \geq 0,\ \forall x, y \in \mathcal{C} \\ f \text{ strictement convexe} &\Leftrightarrow \langle \nabla f(x) - \nabla f(y), x-y\rangle > 0,\ \forall x \neq y \in \mathcal{C} \\ f\ \alpha\text{-convexe} &\Leftrightarrow \langle \nabla f(x) - \nabla f(y), x-y\rangle \geq \alpha\|x-y\|^2,\ \forall x, y \in \mathcal{C} \end{aligned}$$

Remarque. *En dimension 1, f est α-convexe si et seulement si pour tout $x \neq y \in \mathcal{C}$ on a*

$$\frac{f'(x) - f'(y)}{x-y} \geq \alpha,$$

c'est à dire si et seulement si ls taux d'accroissement sont uniformément minorés par $\alpha > 0$.

Exemple 1.2.5. — *$x \in \mathbb{R} \mapsto x^2$ est 2-convexe.* $\frac{f'(x)-f'(y)}{x-y} = 2$.
— *$x \in \mathbb{R} \mapsto e^x$ n'est pas fortement convexe.* $\frac{f(x)-f(0)}{x} \underset{-\infty}{\rightarrow} 0$.

Dans le cadre de la dimension 1 on dispose également d'une caractérisation des fonctions dérivables convexes comme au-dessus de leur tangente :

$$f(y) \geq f(x) + f'(x)(y-x).$$

Proposition 1.2.6 ("tangente-corde")**.**

$$\begin{aligned} f \text{ convexe} &\Leftrightarrow f(y) - f(x) \geq \langle \nabla f(x), y-x\rangle,\ \forall x, y \in \mathcal{C} \\ f \text{ strictement convexe} &\Leftrightarrow f(y) - f(x) > \langle \nabla f(x), y-x\rangle,\ \forall x \neq y \in \mathcal{C} \\ f\ \alpha\text{-convexe} &\Leftrightarrow f(y) - f(x) \geq \langle \nabla f(x), y-x\rangle + \frac{\alpha}{2}\|x-y\|^2,\ \forall x, y \in \mathcal{C} \end{aligned}$$

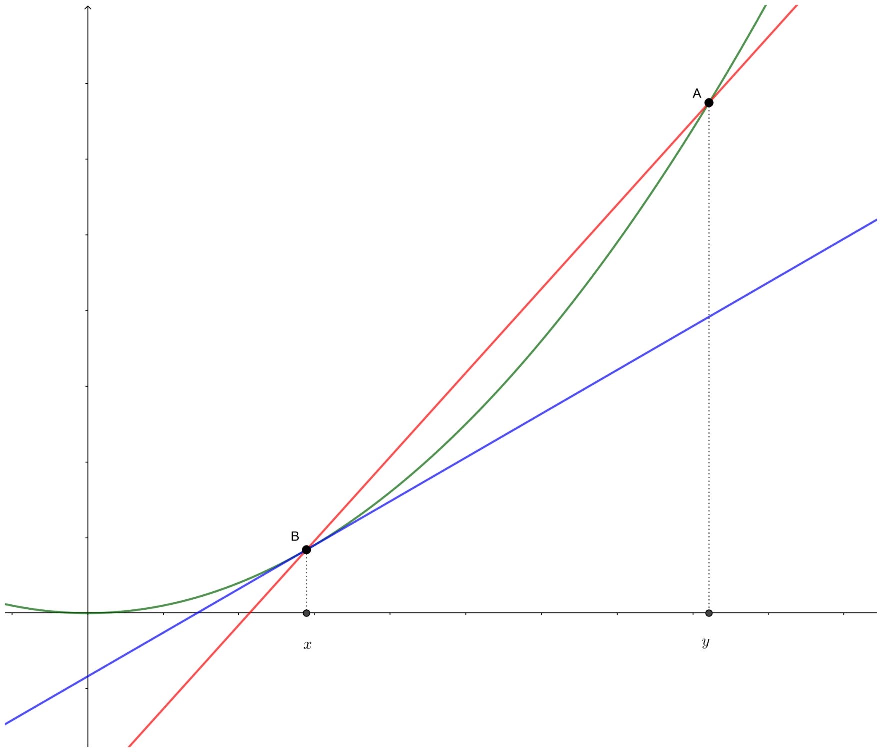

FIGURE 1.1 – La pente de la corde est plus importante que la pente de la tangente.

1.2.3 Le cas deux fois dérivable

On suppose maintenant que $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ une fonction deux fois différentiable sur l'ouvert Ω, et $\mathcal{C}$ un convexe inclus dans Ω.

Dans le cas de la dimension 1, on sait que f est convexe si et seulement si $f'' \geq 0$ sur $\mathcal{C}$. De plus, si $f'' > 0$ sur $\mathcal{C}$ alors f est strictement convexe mais la réciproque est fausse : $x \in \mathbb{R} \mapsto x^4$ est strictement convexe car sa dérivée est strictement croissante mais sa dérivée seconde est nulle en 0.

Proposition 1.2.7. *Si $\mathcal{C}$ est un convexe ouvert, on a*

1. *f est convexe si et seulement si $\nabla^2 f(x)$ est positive pour tout $x \in \mathcal{C}$ (i.e si le spectre de $\nabla^2 f(x)$ est positif pour tout $x \in \mathcal{C}$).*
2. *Si pour tout $x \in \mathcal{C}$, $\nabla^2 f(x)$ est définie positive (i.e si son spectre est strictement positif) alors f est strictement convexe.*
3. *f est α-convexe si et seulement si $\langle \nabla^2 f(x)h, h\rangle \geq \alpha \|h\|^2$ pour tout $x \in \mathcal{C}$ et tout $h \in \mathbb{R}^n$ (i.e si le spectre de $\nabla^2 f(x)$ est inclus dans $[\alpha, +\infty[$).*

Dans le cas où $\mathcal{C}$ n'est pas ouvert, on n'a pas d'information sur toutes les directions. La proposition se reformule donc ainsi.

Proposition 1.2.8. *Si $\mathcal{C}$ est un convexe, on a*

1. *f est convexe si et seulement si pour tout $x, y \in \mathcal{C}$, $\langle \nabla^2 f(x)(y-x), y-x\rangle \geq 0$.*
2. *Si pour tout $x \neq y \in \mathcal{C}$, $\langle \nabla^2 f(x)(y-x), y-x\rangle > 0$ alors f est strictement convexe.*
3. *f est α-convexe si et seulement si pour tout $x, y \in \mathcal{C}$, $\langle \nabla^2 f(x)(y-x), y-x\rangle \geq \alpha\|y-x\|^2$.*

Exemple 1.2.9. *Soient $A \in \mathcal{S}_n(R)$ et $b \in \mathbb{R}^n$. On considère la fonction*

$$f : x \in \mathbb{R}^n \mapsto \frac{1}{2}\langle Ax, x\rangle + \langle b, x\rangle.$$

Comme $\nabla^2 f(x) = A$ *pour tout* $x \in \mathbb{R}^n$, f *est*

- *convexe si et seulement si* $A \in \mathcal{S}_n^+(\mathbb{R})$
- *strictement convexe si et seulement si fortement convexe si et seulement si* $A \in \mathcal{S}_n^{++}(\mathbb{R})$.

Chapitre 2

Généralité sur les problèmes d'optimisation

2.1 Définitions

On considère une fonction $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ et $K \subset \Omega$.

2.1.1 Extrema globaux

Définition 2.1.1. *On dit que x^* est un minimum (resp. maximum) global de f sur K si pour tout $x \in K$ on a*

$$f(x^*) \leq f(x) \quad (resp. f(x^*) \geq f(x)). \tag{2.1}$$

On note alors $f(x^*) = \min_{x\in K} f(x)$ ou encore $x^* \in \operatorname{argmin}_{x\in K} f(x)$. Un extremum global de f sur K est un minimum ou un maximum global de f sur K. Un extremum est dit strict si l'inégalité dan l'équation 2.1 est stricte.

Remarque.
- *Le minimum de f sur K peut ne pas exister ($f = \ln$ sur $K =]0, \infty[$).*
- *Le minimum peut être atteint en plusieurs point ($\mapsto (x^2-1)^2$ sur $\mathbb{R}$).* argmin *et* argmax *désignent des ensembles.*

Exemple 2.1.2. *Si $f : x \in \mathbb{R}^n \mapsto 2 - \|x\|^2$ et $B(0,1)$ désigne la boule unité alors*

$$\min_{x\in B(0,1)} f(x) = 1 \quad et \quad \operatorname{argmin}_{x\in B(0,1)} f(x) = \mathbb{S}(0,1).$$

2.1.2 Extrema locaux

Définition 2.1.3. *On dit que x^* est un minimum local de f sur K s'il existe un voisinage V de x^* dans K tel que x^* qoit un minimum global de f sur V. En d'autre termes, x^* est un minimum local de f sur K s'il esxiste $r > 0$ tel que*

$$\forall x \in B(x^*, r) \cap K,\ f(x^*) \leq f(x).$$

Exemple 2.1.4. *Si $f : x \in \mathbb{R} \mapsto 2x^3 - 3x^2$, 1 est un minimum local sur $\mathbb{R}$ mais f n'admet pas de minimum global.*

2.2 Optimisation et contraintes

On appelle problème d'optimisation le problème consistant à chercher les minima (globaux ou locaux) de $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ sur un sous-ensemble $K \subset \Omega$:

$$f(x^*) = \inf_{x\in K} f(x). \tag{P}$$

On appelle :

— f la *fonction coût* ou *fonction objectif*, ou *critère* du problème,
— K le domaine admissible associé au problème d'optimisation P,
— si $K = \Omega$ le problème sera dit *sans contrainte*, sinon on dira qu'il est contraint,
— toute solution x^* est un *minimiseur* de P ;

L'étude d'un problème d'optimisation et la caractérisation de ses minimiseurs dépend fortement du type de contrainte. Dans ce cours, nous étudierons trois contraintes particulières en plus du cas sans contrainte :

— Contrainte de type égalité

$$K = \{x \in \Omega, g_1(x) = 0, \dots, g_p(x) = 0\},$$

avec $g_i : \mathbb{R}^n \to \mathbb{R}$, pour $1 \leq i \leq p$, de classe $\mathcal{C}^1$.

— Contrainte de type inégalité

$$K = \{x \in \Omega, g_1(x) \leq 0, \dots, g_m(x) \leq 0\},$$

avec $g_i : \mathbb{R}^n \to \mathbb{R}$, pour $1 \leq i \leq m$, de classe $\mathcal{C}^1$.

— Contrainte mixte

$$K = \{x \in \Omega, g_1(x) = 0, \dots, g_p(x) = 0\}, g_{p+1}(x) \leq 0, \dots, g_{p+m}(x) \leq 0\},$$

avec $g_i : \mathbb{R}^n \to \mathbb{R}$, pour $1 \leq i \leq p + m$, de classe $\mathcal{C}^1$.

2.3 Résultats d'existence et d'unicité

2.3.1 Résultats d'existence

Nous allons examiner des situations où l'on peut établir que la fonction $f : \Omega \to \mathbb{R}$ admet un minimum global sur un sous-ensemble $K \subset \Omega$ non-vide.

Proposition 2.3.1. *Si f est continue et K compact, alors f admet un minimum global sur K.*

Démonstration. Par définition de l'infimum il existe une suite minimisante $(x_n)_{n\in\mathbb{N}}$ dans K i.e une suite telle que $f(x_n) \to \inf_{x\in K} f(x)$. Comme K est compact, on peut extraire une sous-suite $(x_{\varphi(n)})_n$ convergente vers $x^* \in K$. Par continuité, $f(x_{\varphi(n)}) \to f(x^*)$ et donc

$$f(x^*) = \inf_{x\in K} f(x).$$

□

Corollaire 2.3.2. *On suppose que f est continue et que K est fermé. S'il existe $\bar{x} \in K$ tel que le sous-niveau*

$$\{x \in K, f(x) \leq f(\bar{x})\}$$

soit borné, alors f admet un minimum global sur K.

Démonstration. Comme f est continu $f^{-1}\left([-\infty, f(\bar{x})]\right)$ est fermé. De plus, puisque K est fermé, le sous-niveau, K', est fermé, et borné par hypothèse, donc compact. La fonction f admet donc un minimiseur global x^* sur K'. Par ailleurs, on a $f(x^*) \leq f(\bar{x})$ par définition de sous-niveau et pour tout $x \in K \setminus K'$, on a $f(x^*) \leq f(\bar{x}) \leq f(x)$. Donc x^* est un minimiseur global de f sur K. □

Remarque. *En remplaçant f par $-f$, on montre qu'il existe des maxima globaux s'il existe un sur-niveau*

$$\{x \in K, f(x) \geq f(\bar{x})\}.$$

Exemple 2.3.3. *Il existe un pavé de $\mathbb{R}^3$ de surface $S > 0$ et de volume maximal.*

Démonstration. On note $(x_1, x_2, x_3) \in \mathbb{R}^3$ les dimensions du pavé. On cherche à maximiser la fonction coût $f : (x_1, x_2, x_3) \in \mathbb{R}^3 \mapsto x_1 x_2 x_3$ sur

$$K = \{(x_1, x_2, x_3) \in (\mathbb{R}_+)^3, 2(x_1x_2 + x_1x_3 + x_2x_3) = S\}.$$

On remarque que pour $i \neq j$, $x_i x_j \leq S$, on en déduit que pour $x_i \neq 0$, alors

$$\begin{aligned} f(x) &= \frac{x_i^2 x_j x_k}{x_i} \\ &= \frac{1}{x_i}(x_i x_j)(x_i x_k) \\ &\leq \frac{S^2}{x_i} \end{aligned}$$

Soit $\bar{x} \in K$ tel que $f(\bar{x}) > 0$ (il suffit que toutes les coordonnées soient non nulles et il en existe). On cherche $R > 0$ tel que le sur-niveau

$$K' = \{x \in K, f(x) \geq f(\bar{x})\} \subset [0, R]^3.$$

Heuristiquement, pour $x \in K'$, on a

$$f(\bar{x}) < f(x) \leq \frac{S^2}{x_i}$$

et donc $x_i \leq \frac{S^2}{f(\bar{x})}$. On pose alors $R = \frac{S^2}{f(\bar{x})}$ et pour tout $x \in K \setminus [0, R]^3$, il existe i tel que $x_i > R > 0$. Dans ce cas, on a

$$f(x) \leq \frac{S^2}{x_i} < \frac{S^2}{R} = f(\bar{x}),$$

et $x \notin K'$. On a donc montré que $K' \subset [0, R]^3$ est borné. Comme f est continu et K fermé, f admet un maximum global sur K. □

Définition 2.3.4. *On dit que f est coercive sur K si pour tout $M \in \mathbb{R}$, il existe $R > 0$ tel que*

$$\|x\| > R \Rightarrow f(x) > M.$$

En d'autres termes, $\lim_{\|x\| \to \infty, x \in K} f(x) = +\infty$.

Proposition 2.3.5. *On suppose que K est fermé et f est continue et coercive sur K. Alors f admet un minimum global sur K.*

Démonstration. Soit $\bar{x} \in K$, on pose $M = f(\bar{x})$. Comme f est coercive, il existe $R > 0$ tel que le sous-niveau $\{x \in K, f(x) \leq M\}$ soit inclus dans $B(0, R)$ qui est borné. f étant continue sur le fermé K, on conclu grâce au corollaire précédent. □

La notion de convexité est aussi un argument commode en optimisation comme l'illustre les deux critères suivants.

Proposition 2.3.6. *On suppose que f est différentiable sur Ω. Si K est un convexe fermé et f est fortement convexe sur K alors f admet un minimum global sur K.*

Démonstration. On fixe $\bar{x} \in K$. D'après la Proposition 1.2.6 , on a

$$\begin{aligned} f(x) &\geq f(\bar{x}) + \langle \nabla f(\bar{x}), x - \bar{x} \rangle + \frac{\alpha}{2}\|x - \bar{x}\|^2 \\ &\geq f(\bar{x}) - \|\nabla f(\bar{x})\| \|x - \bar{x}\| + \frac{\alpha}{2}\|x - \bar{x}\|^2 \end{aligned}$$

Ceci implique la coercivité de f sur K. □

La notion de convexité permet le passage du local au global.

Proposition 2.3.7. *On suppose que K est convexe et que f est convexe sur K. Si f admet un minimum local sur K, alors il est global.*

Démonstration. Soit x^* un minimum local de f sur K. on dispose de $r > 0$ tel que le minimum soit global sur $K \cap B(x^*, r)$. Soit $x \in K$. Pour $\lambda \in]0, 1[$ suffisamment petit, $(1 - \lambda)x^* + \lambda x \in K \cap B(x^*, r)$. Par conséquent, on a

$$f(x^*) \leq f((1 - \lambda)x^* + \lambda x) \leq (1 - \lambda)f(x^*) + \lambda f(x)$$

autrement dit $f(x^*) \leq f(x)$ et x^* est un minimum global sur K. □

2.3.2 Résultats d'unicité

Pour les questions d'unicité, la convexité est l'argument clef.

Proposition 2.3.8. *Si K est convexe et f est strictement convexe sur K alors f admet au plus un minimum global sur K.*

Démonstration. Supposons qu'il existe $x^* \neq y^* \in K$ tels que $f(x^*) = f(y^*) = \min_{x \in K} f(x)$. Pour $\lambda \in]0, 1[$, on a alors par stricte convexité

$$f(\lambda x^* + (-1\lambda)y^*) < \lambda f(x^*) + (-1\lambda)f(y^*) = f(x^*).$$

Ceci est absurde. □

Corollaire 2.3.9. *Si f est différentiable sur Ω est fortement convexe sur K convexe fermé, alors f admet un unique minimum global*

Le tableau suivant récapitule les résultats d'existence et unicité en fonction de la convexité.

	Éxistence	Unicité	local $\to$ global
cvx	Non ($x \mapsto x$)	Non ($x \mapsto 0$)	Oui
str cvx	Non ($x \mapsto e^x$)	Oui	Oui
α-cvx	Oui	Oui	Oui

Chapitre 3

Optimisation sans contrainte

On considère une fonction f d'un ouvert $\Omega \subset \mathbb{R}^n$ dans $\mathbb{R}$. Nous allons étudier des conditions nécessaires et des conditions suffisantes, selon la régularité de f, pour qu'elle admette un extremum local en un point $x^* \in \Omega$.

3.1 Conditions nécessaires d'optimalité

3.1.1 Condition d'ordre 1

Proposition 3.1.1. *Si f admet un extremum local en $x^* \in \Omega$ et si f est différentiable en x^* alors*

$$\nabla f(x^*) = 0 \tag{3.1}$$

Remarque. *L'équation (3.1) est appelée équation d'Euler et ses solutions sont les points critiques de f.*

Démonstration. Soit $v \in \mathbb{R}^n$. La formule de Taylor-Young en $x*$ permet d'écrire

$$f(x^* + tv) - f(x^*) = t\langle \nabla f(x^*), v\rangle + o(t).$$

Sans perdre de généralité, on peut supposer que f admet un minimum en x^* de sorte que $f(x^* + tv) - f(x^*) \geq 0$ pour t assez petit. Par conséquent, on a

$$\langle \nabla f(x^*), v\rangle = \lim_{t\to 0} \underbrace{\frac{f(x^* + tv) - f(x^*)}{t}}_{\substack{\geq 0 \,\mathrm{si}\, t > 0 \\ \leq 0 \,\mathrm{si}\, t < 0}} = 0.$$

Comme $\langle \nabla f(x^*), v\rangle = 0$ pour tout $v \in \mathbb{R}^n$ alors $\nabla f(x^*) = 0$. □

Exemple 3.1.2. *Pour $f : (x, y) \in \mathbb{R}^2 \mapsto x^2 + y^2 - xy$, le seul point critique est $(0, 0)$. Par ailleurs, f est coercive car on a $2xy \leq x^2 + y^2$ et ainsi $f(x, y) \geq \|(x, y)\|^2/2$. La fonction f admet donc un minimum global, qui est nécessairement sont unique point critique $(0, 0)$.*

Remarque. *Cette condition n'est pas suffisante. $x \mapsto x^3$ admet un point critique en 0 sans y admettre un extremum.*

Dans le cas des fonction convexe cependant, la condition est nécessaire et suffisante.

Proposition 3.1.3. *On suppose Ω convexe et f convexe et différentiable sur Ω. f admet un minimum global en $x^* \in \Omega$ si et seulement si $\nabla f(x^*) = 0$.*

Démonstration. On suppose que x^* est un point critique de f. D'après la proposition 1.2.6, on a

$$f(x) \geq f(x^*) + \langle \nabla f(x^*), x - x^* \rangle = f(x^*).$$

Donc f admet un minimum global en x^*. □

3.1.2 Conditions d'ordre 2

Nous allons maintenant étudier un critère plus précis faisant intervenir la hessienne de f. Ce critère fait la différence entre maximum local en minimum local.

Proposition 3.1.4. *Si f admet un minimum local en x^* et si f est différentiable sur un voisinage de x^* et deux fois différentiable en x^*, alors $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*)$ est positive.*

Démonstration. La condition sur le gradient est gratuite. Soit $v \in \mathbb{R}^n$, d'après la formule de Taylor-Young à l'ordre 2, on a

$$f(x^* + tv) - f(x^*) = \frac{t^2}{2} \langle \nabla^2 f(x^*) v, v \rangle + o(t^2).$$

On a donc pour tout $v \in \mathbb{R}^n$,

$$\langle \nabla^2 f(x^*) v, v \rangle = \lim_{t \to 0} \frac{2(f(x^* + tv) - f(x^*))}{t^2} \geq 0,$$

soit $\nabla^2 f(x^2) \in \mathcal{S}_n^+(\mathbb{R})$. □

Remarque. — *Pour un maximum local, il faut que la hessienne soit négative.*
— *Cette condition n'est pas suffisante comme le montre $x \mapsto x^3$.*

3.2 Conditions suffisantes d'optimalités

Un hypothèse plus forte sur la hessienne permet d'obtenir une condition suffisante d'optimalité.

Proposition 3.2.1. *Soit $x^* \in \Omega$ tel que f soit différentiable sur un voisinage de x et deux fois différentiable en x^*. Si $\nabla f(x^*) = 0$ et $\nabla^2 f(x^*) \in \mathcal{S}_n^{++}(\mathbb{R})$ alors f admet un minimum local en x^*.*

Démonstration. On note ρ la plus petite valeur propre de $\nabla^2 f(x^*)$. On a donc

$$\forall v \in \mathbb{R}^n,\ \langle \nabla^2 f(x^*) v, v \rangle \geq \rho \|v\|^2.$$

D'après la formule de Taylor-Young à l'ordre 2 en x^*, on a

$$\begin{aligned} f(x + v) &= f(x^*) + \frac{1}{2} \langle \nabla^2 f(x^*) v, v \rangle + o(\|v\|^2) \\ &\geq \frac{\rho}{2} \|v\|^2 + o(\|v\|^2) \end{aligned}$$

Ainsi, il existe $r > 0$ tel que pour $\|v\| < r$, $f(x^* + v) - f(x^*) \geq 0$ i.e, x^* est un minimum local. □

Remarque. — *Cette condition n'est pas nécessaire comme le montre $x \mapsto x^4$.*

— *Pour un maximum local, il faut* $\nabla^2 f(x^*)$ *définie négative.*

Pour résoudre un problème d'optimisation sans contrainte, "la" marche à suivre est la suivante :

— Chercher les point critiques.
— En déduire les extrema locaux.
— Remonter du local au global.

Exemple 3.2.2. *On cherche les extrema locaux et globaux de la fonction*

$$f : (x, y) \in \mathbb{R}^2 \mapsto x^2 + y^3 + y^2.$$

Pour $(x, y) \in \mathbb{R}^2$ *on a*

$$\nabla f(x, y) = \begin{pmatrix} 2x \\ 3y^2 + 2y \end{pmatrix}$$

La fonction f *admet donc deux points critiques :* $(0, 0)$ *et* $(0, -2/3)$. *On calcula alors la hessienne de* f *en ces points et on a*

$$\nabla^2 f(0, 0) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \quad \nabla^2 f(0, -2/3) = \begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}.$$

On en déduit donc que

— $(0, 0)$ *est un minimum local car* $\nabla^2 f(0, 0) \in \mathcal{S}_n^{++}(\mathbb{R})$. *En revanche ce n'est pas un minimum global car* f *n'en admet pas (en effet* $f(0, t) \to -\infty$ *quand* $t \to -\infty$*).*
— $(0, -2/3)$ *n'est pas un extremum local car* $\nabla^2 f(0, -2/3)$ *n'est ni dans* $\mathcal{S}_n^{+}(\mathbb{R})$ *ni dans* $\mathcal{S}_n^{-}(\mathbb{R})$. *C'est un point-selle. En effet,* $t \mapsto f(t, -2/3)$ *est minimal en* 0 *et* $t \mapsto f(0, t-2/3)$ *est maximal en* 0.

Chapitre 4

Optimisation sous contrainte

Soient $f : \Omega \to \mathbb{R}$, où Ω est un ouvert de ^n_R, et $K \subset \Omega$. On s'intéresse à la minimisation de la fonction f sur le domaine K, que l'on appellera *domaine admissible* associé au problème d'optimisation.

On remarque que si f admet un minimum local en un point x^* dans l'intérieur du domaine admissible et si f est suffisamment régulière alors les conditions d'optimalité du chapitre précédent sont encore valable. C'est aussi le cas si l'on ajoute la contrainte $x \in \mathring{K}$.

La difficulté de ce type de problème survient lorsque le minimum local est situé sur le bord du domaine admissible, ∂K, ce qui arrive très fréquemment. On dit alors que l'optimum *sature* la contrainte.

4.1 Inéquation d'Euler

Nous commençons par introduire la notion de direction admissible en un point $x \in K$. Il s'agit des directions "asymptotiquement rentrantes" au point x.

Définition 4.1.1. *On dit que $j \in \mathbb{R}^n$ est une direction admissible en $x \in K$ s'il existe une suite $(\varepsilon_k)_{k\in\mathbb{N}}$ dans $\mathbb{R}_+^*$ convergeant vers 0 et une suite $(h_k)_{k\in\mathbb{N}}$ convergeant vers h telle que pour tout $k \in \mathbb{N}$*

$$x + \varepsilon_k h_k \in K.$$

On note $K(x)$ l'ensemble des directions admissibles en $x \in K$.

De manière plus géométrique, on peut le reformuler comme l'existence d'une suite $(x_k)_{k\in\mathbb{N}}$ dans Kconvergeant vers x et dont la direction converge vers celle de h

$$\frac{x_k - x}{\|x_k - x\|} \to \frac{h}{\|h\|}.$$

On retrouve la définition précédente en posant $h_k = \frac{\|h\|}{\|x_k - x\|}(x_k - x)$ et $\varepsilon_k = \frac{\|x_k - x\|}{\|h\|}$.

Remarque. *Si $x \in \mathring{K}$ alors $K(x) = \mathbb{R}^n$. Il suffit de poser $H_k = h$ et $\varepsilon_k = 1/k$. En effet, $x + h/k$ appartient à K à partir d'un certain rang.*

Proposition 4.1.2. *Si K est convexe alors l'ensemble des directions admissibles en $x \in K$ se réécrit*

$$K(x) = \overline{\{\lambda(x-y), \lambda > 0, y \in K\}}.$$

Démonstration. Cf TD □

Pour les problèmes d'optimisation avec contraintes, les points critiques ne sont plus liés à des minimums de la fonction. Ceci est lié au fait que le minimum est recherché selon seulement une partie des directions. On obtient donc une version plus faible de l'équation d'Euler.

Théorem 4.1.3 (Inéquation d'Euler)**.** *Si f est différentiable en $x^* \in K$ et admet un minimum local sur K en x^*, alors pour tout $h \in K(x^*)$*

$$\langle \nabla f(x^*), h \rangle \geq 0.$$

Démonstration. Soit $h \in K(x^*)$ et soient $(\varepsilon_k)_k$ et $(h_k)_k$ associées. Comme x^* est un minimum local de f alors à partir d'un certain rang, on a $f(x^*) \leq f(x^* + \varepsilon_k h_k)$. D'autre part, la différentiabilité de f en x^* donne

$$\frac{f(x^* + \varepsilon_k h_k) - f(x^*)}{\varepsilon_k} \to \langle \nabla f(x^*), h \rangle.$$

Ce qui conclut la démonstration. □

L'inéquation d'Euler n'est pas assez forte pour donner une caractérisation des extrema. Pour obtenir une condition nécessaire, on va avoir besoin d'une hypothèse supplémentaire sur les contraintes : la qualification. Cette condition prendra différente forme selon le type de contrainte. Nonobstant le shéma d'approche des problèmes contraints sera semblable :

— Identifier les contraintes
— Montrer que les contraintes sont qualifiées (souvent un argument de liberté est suffisant)
— Appliquer la condition nécessaire adaptée (extrema lié, KKT...)

Nous allons maintenant définir ce qu'est une contrainte qualifiée dans le cas des contraintes de types égalité, de type inégalité et mixtes.

4.2 Contrainte de type égalité

On suppose que f est de classe $\mathcal{C}^1$ sur Ω et on s'intéresse au cas où le domaine admissible est décrit par des contrainte de type égalité, i.e

$$K = \{x \in \Omega, g_1(x) = 0, \ldots, g_p(x) = 0\},$$

où les fonctions g_i sont de classe $\mathcal{C}^1$ sur Ω à valeur dans $\mathbb{R}$. On écrira de manière plus compacte $g = {}^t(g_1, \ldots, g_p) : \mathbb{R}^n \to \mathbb{R}^p$.

Exemple 4.2.1. — $K = \{x \in \mathbb{R}^2, x_1^2 + x_2^2 - 1 = 0\}$ *(optimisation sur le cercle unité)*
— $K = \{x \in \mathbb{R}^3, x_1^2 + x_2^2 + x_3^2 - 1 = 0, x_3 - 1/2 = 0\}$

Remarque. *Les fonctions g_i étant continues, le domaine admissible est alors toujours fermé. De plus, il sera presque toujours d'intérieur vide (sauf cas dégénéré). Cependant le caractère purement local des arguments en jeux permet de traiter de la même manière les domaines admissibles, plus nécessairement fermés, de la forme*

$$K = \{x \in \Omega, g_i(x) = 0\},$$

où Ω est ouvert. Pour un même problème, on aura parfois besoin de jouer entre différente reformulation.

Définition 4.2.2. *On dit que les contraintes $(g_i)_i$ sont qualifiées en un point $x \in K$ si la famille*

$$\nabla g_1, \ldots, \nabla g_p,$$

est libre, autrement dit, si $rg(J_g(x)) = p$.

Remarque. *Si $p = 1$, alors la contrainte est qualifiée en x si et seulement si $\nabla g(x) \neq 0$.*

Proposition 4.2.3. *Si les contraintes sont qualifiées en $x^* \in K$ alors on a*

$$K(x^*) = \{\nabla g_1, \dots, \nabla g_p\}^\perp .$$

Démonstration. On procède par double inclusion.

• Soient $h \in K(x^*)$ et $(\varepsilon_k)_k$, $(h_k)_k$ associées. Pour tout i et tout k, on a $g_i(x^* + \varepsilon_k h_k) = 0$. D'après la formule de Taylor-Young, on a donc

$$0 = \frac{g_i(x^* + \varepsilon_k h_k) - g_i(x^*)}{\varepsilon_k} = \langle \nabla g_i(x^*), h_k \rangle + \|h_k\| \eta(\|\varepsilon_k h_k\|),$$

où $\lim_{t\to 0} \eta(t) = 0$. On en déduit donc que $\langle \nabla g_i(x^*), h \rangle = 0$.

• Soit $h \in \mathbb{R}^n$ tels que pour tout i, $\langle \nabla g_i(x^*), h \rangle = 0$. La contrainte étant qualifiée en x^*, la matrice $J_g(x^*)$ est de rang plein. On pose $J_g(x^*) = (J_1 \| J_2)$ avec $J_1 \in \mathcal{M}_{p,n-p}(\mathbb{R})$ et $J_2 \in \mathcal{M}_p(\mathbb{R})$. Quitte à échanger les coordonnées, on peut supposer que J_2 est inversible et on note alors $x^* = (x_1^*, x_2^*) \in \mathbb{R}^{n-p} \times \mathbb{R}^p$. D'après le théorème des fonctions implicites, il existe un voisinage $U \subset \mathbb{R}^{n-p}$ de x_1^*, un voisinage $V \subset \mathbb{R}^p$ de x_2^* et un $\mathcal{C}^1$-difféomorphisme $\varphi : U \to V$ tel que $\varphi(x_1^*) = x_2^*$ et pour tout $(x_1, x_2) \in U \times V$

$$g(x_1, x_2) = 0) \Leftrightarrow x_2 = \varphi(x_1).$$

On a de plus $J_\varphi(x_1^*) = -J_2^{-1} J_1$.

A partir d'un certain rang, on a $x_1 + h_1/k \in U$ donc $g(x_1 + h_1/k, \varphi(x_1 + h_1/k)) = 0$. Or d'après la formule de Taylor-Young, on a

$$\begin{aligned} \varphi(x_1 + h_1/k) &= \varphi(x_1^*) + \frac{1}{k} J_\varphi(x_1^*) h_1 + \frac{1}{k}\eta(1/k) \\ &= x_2^* - \frac{1}{k} J_2^{-1} J_1 h_1 + \frac{1}{k}\eta(1/k) \end{aligned}$$

où $\eta(t) \to 0$ quand $t \to 0$. Par hypothèse, on a $J_g(x^*)h = 0$, ce qui se traduit par $J_1 h_1 + J_2 h_2 = 0$. On a donc

$$\varphi(x_1 + h_1/k) = x_2^* + \frac{1}{k}\left(h_2 + \eta(1/k)\right).$$

On a donc montré que

$$(x_1^* + \frac{h_1}{k}, \varphi(x_1^* + \frac{h_1}{k})) = (x_1^*, x_2^*) + \frac{1}{k}\left(h_1, h_2 + \eta(1/k)\right) \in K$$

Ainsi, en posant $\varepsilon_k = 1/k > 0$ et $h_k = (h_1, h_2 + \eta(1/k)) \in \mathbb{R}^n$ on a montré que $h \in K(x^*)$. □

Remarque. *On a aussi montré que la première inclusion était vérifiée même quand les contraintes ne sont pas qualifiées.*

On peut maintenant énoncer une condition nécessaire d'optimalité.

Théorem 4.2.4 (Théorème des extrema liés)**.** *Si f admet un extremum local en $x^* \in K$ et si les contraintes sont qualifiées en ce point alors il existe des réels $\lambda_1, \dots, \lambda_p$ tels que*

$$\nabla f(x^*) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x^*) = 0. \tag{4.1}$$

Les λ_i sont appelés multiplicateur de Lagrange associés aux contraintes g_i. L'équation (4.1) est appelée l'équation Euler-Lagrange.

Démonstration. D'après la proposition précédente, $K(x^*)$ est un espace vectoriel. Ainsi on déduit de l'inéquation d'Euler que pour tout $h \in K(x^*)$, $\langle \nabla f(x^*), h\rangle = 0$, ce qui se traduit par pour tout $K(x^*) \subset \nabla f(x^*)^\perp$. On a donc

$$\nabla f(x^*) \in \nabla f(x^*)^{\perp\perp} \subset K(x^*)^{\perp\perp} = K(x^*) = Vect(\nabla g_1, \dots, \nabla g_p).$$

□

Dans la pratique, on cherche donc $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ et $\lambda = (\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ tels que

$$\begin{cases} \nabla f(x) + \sum_{i=1}^{p} \lambda_i \nabla g_i(x) = 0 \\ \forall 1 \leq i \leq p,\, gi(x) = 0 \end{cases}$$

Il s'agit d'un système de $n+p$ équation à $n+p$ inconnues.

Remarque. *L'hypothèse de qualification des contraintes est cruciale. Par exemple, la fonction* $f : (x,y) \in \mathbb{R}^2 \mapsto x + y^2$ *atteint son minimum sur* $\{(x,y) \in \mathbb{R}^2, x^3 - x^2 = 0\}$ *en* $(0,0)$ *(car* f *est positive sur* K*). Cependant, pour tout* $\lambda \in \mathbb{R}$*, on a*

$$\nabla f(0,0) + \lambda \nabla g(0,0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Définition 4.2.5. *On appelle lagrangien du problème de minimisation l'application*

$$\begin{aligned} \mathcal{L} : \mathbb{R}^n \times \mathbb{R}^p &\to \mathbb{R} \\ (x,\lambda) &\mapsto f(x) + \sum_i \lambda_i g_i(x) \end{aligned}$$

Avec ce formalisme, la recherche d'extrema pour le problème d'optimisation avec contrainte de type égalité se ramène à résoudre $\nabla \mathcal{L}(x,\lambda) = 0$.

Exemple 4.2.6. *On revient au problème de pavé de volume maximal à surface fixée* $S > 0$*. On a déjà montré que le problème admet une solution et on cherche à la caractériser. Le domaine admissible en question est*

$$K = \left\{ x \in (\mathbb{R}_+)^3, 2(x_1x_2 + x_1x_3 + x_2x_3) - S = 0 \right\}.$$

Ce domaine n'est pas de type égalité à cause de $\mathbb{R}_+$*. On va commencer par résoudre le problème sur un domaine plus simple et montrer l'équivalence entre les deux problèmes. On pose*

$$\tilde{K} = \left\{ x \in (\mathbb{R}_+^*)^3, 2(x_1x_2 + x_1x_3 + x_2x_3) - S = 0 \right\}.$$

La fonction objectif est $f : x \in \mathbb{R}^3 \mapsto x_1x_2x_3$ *et la contrainte est de type égalité avec* $g : x \in \mathbb{R}^3 \mapsto 2(x_x 1 x_2 + x_1x_3 + x_2x_3) - S$*. On a*

$$\forall x \in \mathbb{R}^3, \nabla g(x) = 2 \begin{pmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix},$$

donc la contrainte est qualifiée sur K *(car* $0 \notin K$*). On cherche donc* $(x,\lambda) \in \mathbb{R}^3 \times \mathbb{R}$ *tel que*

$$\begin{cases} x_2x_3 + \lambda(x_2 + x_3) = 0 \\ x_1x_3 + \lambda(x_1 + x_3) = 0 \\ x_1x_2 + \lambda(x_1 + x_2) = 0 \\ 2(x_1x_2 + x_1x_3 + x_2x_3) = S \end{cases}.$$

On montre alors que nécessairement $x_1 = x_2 = x_3 = \sqrt{\frac{S}{6}}$. Pour finir, si l'une des coordonnées de $x \in K$ est nulle, alors $f(x) = 0$. Ainsi le maximum de f sur K n'est pas atteint en ces points et on a bien

$$\max_{x \in K} f(x) = \max_{x \in \tilde{K}} f(x).$$

Comme f admet un maximum global sur K, il est forcément atteint en $\left(\sqrt{\frac{S}{6}}, \sqrt{\frac{S}{6}}, \sqrt{\frac{S}{6}}\right)$ et est unique.

4.3 Contraintes de type inégalité

Dans cette partie le domaine admissible est de la forme

$$K = \{x \in \mathbb{R}^n, g_1(x) \leq 0, \dots, g_p(x) \leq 0\},$$

où les fonctions $g_i : \mathbb{R}^n \to \mathbb{R}$ sont de classe $\mathcal{C}^1$.

Exemple 4.3.1.
- $K = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 - 1 \leq 0, y \leq 0\}$ *(demi-disque unité inférieur)*
- $K = \{x \in \mathbb{R}^3, |x| - 1 \leq 0, x_3 - 1/2 \leq 0\}$ *(intersection de la boule unité et d'un demi-espace)*

Pour $x \in K$, on note $\mathbb{I}(x)$ l'ensemble des indices des contraintes saturées en x :

$$\mathcal{I}(x) = \{i \in [\![1, p]\!], g_i(x) = 0\}.$$

Définition 4.3.2. *On dit que les contraintes $(g_i)_i$ sont qualifiées en $x \in K$ s'il existe une direction $v \in \mathbb{R}^n$ telle que*

$$\forall i \in \mathcal{I}(x),\ \langle \nabla g_i(x), v \rangle < 0. \tag{4.2}$$

La direction v est dite rentrante.

Remarque. *Pour $x \in \partial K$ et $i \in \mathcal{I}(x)$, les vecteurs $\nabla g_i(x)$ sont normaux sortants par rapport au $\{g_i = 0\}$. La condition de qualification se traduit donc par l'existence d'une portion de cône de sommet x dans K.*

La condition de qualification (4.2) peut être assez délicate à vérifier. On dispose d'une condition suffisante plus simple à vérifier.

Proposition 4.3.3. *Si la famille $(\nabla g_i(x))_{i \in \mathcal{I}(x)}$ est libre alors les contraintes sont qualifiées en x.*

Démonstration. On va montrer qu'il existe une direction rentrante $v \in Vect\{\nabla g_i(x), i \in \mathcal{I}(x)\}$ telle que

$$\forall i \in \mathcal{I}(x), \langle \nabla g_i(x), v \rangle = -1. \tag{4.3}$$

Soit un tel v. On dispose de $(\alpha_j) \in \mathbb{R}$ tels que

$$v = \sum_{j \in \mathcal{I}(x)} \alpha_j \nabla g_j(x).$$

On note $\mathcal{I}(x) = \{i_1, \dots, i_m\}$. L'équation (4.3) se reformule matriciellement par

$$B \begin{pmatrix} \alpha_{i_1} \\ \vdots \\ \alpha_{i_m} \end{pmatrix} = \begin{pmatrix} -1 \\ \vdots \\ -1 \end{pmatrix},$$

où $B = {}^t M M$ et $M = (\nabla g_{i_1}(x) | \cdots | \nabla g_{i_m}(x))$. Par hypothèse, M est de rang plein, ainsi la matrice B est inversible et il existe une solution à l'équation matrcielle. $\square$

Remarque. *Cette condition n'est pas nécessaire. En fait on peut montrer que la qualification des contraintes est équivalente à ce que la famille* $(\nabla g_i(x))_{i\in\mathcal{I}(x)}$ *soit positivement libre.*

Sous l'hypothèse de qualification, l'espace des direction admissible en x peut être caractérisé.

Lemme 4.3.4. *Sil es contraintes sont qualifiées en $x \in K$ alors on a*

$$K(x) = \{h \in \mathbb{R}^n, \forall i \in \mathcal{I}(x), \langle g_i(x), h\rangle \leq 0\}\,.$$

Démonstration. On procède par double inclusion. • Soit $h \in k(x)$, on dispose des suite (ε_k) et (h_k) comme dans la définition. Comme $x + \varepsilon_k h_k \in K$ alors pour tout i, on a $g_i(x + \varepsilon_k h_k) \leq 0$. De plus, si $i \in \mathcal{I}(x)$, on a $g_i(x) = 0$ et donc d'après la formule de Taylor-Young, on a

$$0 \geq \frac{g_i(x+\varepsilon_k h_k)}{\varepsilon_k} = \langle g_i(x), h_k\rangle + o(1)$$

Ainsi à la limite, on a bien $\langle g_i(x), h\rangle \leq 0$.

• Soit $h \in \mathbb{R}^n$ tel que pour tout $i \in \mathcal{I}(x)$, on a $\langle g_i(x), h\rangle \leq 0$. Les contrainte étant qualifiées en x, on dispose d'une direction rentrante $v \in \mathbb{R}^n$. On pose alors $h_k = h + \frac{v}{k}$.

Si $i \in \mathcal{I}(x)$, on a $g_i(x) = 0$ et $\langle g_i(x), h_k\rangle < 0$. Pour tout $\varepsilon > 0$

$$g_i(x + \varepsilon h_k) = \varepsilon\langle g_i(x), h_k\rangle + o(\varepsilon)$$

On dispose donc de $\varepsilon_k^{(i)} \in]0, 1/k]$ tel que $g_i(x + \varepsilon h_k) \leq 0$ pour tout $\varepsilon < \varepsilon_k^{(i)}$.

Si $i \notin \mathcal{I}(x)$ alors $g_i(x) < 0$ et la continuité de g_i assure qu'il existe $\varepsilon_k^{(i)} \in]0, 1/k]$ tel que $g_i(x + \varepsilon h_k) \leq 0$ pour tout $\varepsilon < \varepsilon_k^{(i)}$.

On pose alors $\varepsilon_k = \min_i\{\varepsilon_k^{(i)}\}$ et on a $\varepsilon_k \to 0$, $h_k \to h$ et $x + \varepsilon_k h_k \in K$ pour tout k. Donc $h \in K(x)$. □

Le dernier argument pour obtenir un équivalent au théorème des extrema liés et le lemme de Farkas.

Lemme 4.3.5 (Farkas)**.** *Soient $u_1, \dots, u_m$ et v des vecteurs de $\mathbb{R}^n$. Les propositions suivant sont équivalentes.*

(i) $\{h \in \mathbb{R}^n, \forall 1 \leq i \leq m\, \langle u_i, h\rangle \geq 0\} \subset \{h \in \mathbb{R}^n, \langle v, h\rangle \geq 0\}$

(ii) le vecteur v est une combinaison linéaire à coefficients positifs des u_i.

Démonstration. On note $\mathcal{C} = \{\sum_{i=1}^m \mu_i u_i, \forall 1 \leq i \leq m\, \mu_i \geq 0\}$. La preuve repose sur le théorème de projection sur un convexe fermé. La convexité de $\mathcal{C}$ est évidente mais son caractère fermé ne l'est pas.

• Montrons que $\mathcal{C}$ est fermé. Supposons pour commencer que la famille (u_i) est libre. Elle forme donc une base de l'espace vectoriel qu'elle engendre. Dans ce cas, si $(h_k)_k$ est une suite dans $\mathcal{C}$ qui converge vers $h \in \mathbb{R}^n$, alors $h \in Vect\{u_i\}$ et les suites de coordonnées $(\mu_i^{(k)}$ converge vers les coordonnées de h dans cette base, $\mu_i \geq 0$. Donc $h \in \mathcal{C}$ et $\mathcal{C}$ est fermé.

Si la famille n'est pas libre, on pose $\mathcal{C}_I = \{\sum_{i\in I} \mu_i u_i, \mu_i \geq 0\}$, pour $I \subset [\![1, m]\!]$. On a montré que si $(u_i)_{i\in I}$ est libre alors $\mathcal{C}_I$ est fermé. On va montrer que $\mathcal{C}$ est l'union des $\mathcal{C}_I$ tel que $(u_i)_{i\in I}$ est libre.

Soit $h = \sum_{i\in J} \mu_i u_i \in \mathcal{C}_J$ avec $\mu_i > 0$ et $(u_i)_{i\in J}$ liée. On dispose donc de $(\alpha_i)_{i\in J}$, non tous nuls, telle que $\sum_{i\in J} \alpha_i u_i = 0$. Pour tout $t \in \mathbb{R}$, on a

$$h = \sum_{i\in J} (\mu_i + t\alpha_i) u_i.$$

On pose alors $\phi : t \in \mathbb{R} \mapsto \min_i\{\mu_i + t\alpha_i\}$. On observe que ϕ est continue, que $\phi(0) > 0$ et que $\lim_{t\to-\infty} \phi(t) = -\infty$. D'après le théorème des valeurs intermédiaire, il existe t_0 tel que

$\phi(t_0) = 0$. Ce qui signifie que $h \in \mathcal{C}_{\tilde{J}}$ avec $\tilde{J} \subsetneq J$. En itérant le procédé, on se ramène à une combinaison de (u_i) indépendants. On a donc montré que $\mathcal{C}$ était une union finie d'ensembles fermés. Donc $\mathcal{C}$ est fermé.

• On remarque que $(ii) \Rightarrow (i)$ est triviale.

• On suppose que $v \notin \mathcal{C}$ et on pose $h = p_{\mathcal{C}}(v) - v$, ou $p_{\mathcal{C}}$ désigne la projection sur le convexe fermé $\mathcal{C}$. Par propriété de la projection[1], $\langle h, v\rangle = 0$. On a donc

$$\langle v, h\rangle = \langle v - p_{\mathcal{C}}(v), h\rangle = -\|h\|^2 < 0,$$

car $v \notin \mathcal{C}$ donc $h \neq 0$. Ainsi, $h \notin \{x \in \mathbb{R}^n, \langle v, x\rangle \geq 0\}$. Par ailleurs, pour tout $1 \leq i \leq m$, par propriété[2] de la projection sur un convexe fermé, on a

$$-\langle h, u_i\rangle = \langle v - p_{\mathcal{C}}(v), u_i - p_{\mathcal{C}}(v)\rangle \leq 0.$$

Ceci implique que $h \in \{x \in \mathbb{R}^n, \forall 1 \leq i \leq m \, \langle x, u_i\rangle \geq 0\}$. Ce qui conclu la preuve. □

Le théorème de Karush-Kuhn-Tucker est le pendant du théorème des extrema liés pour les problèmes de type inégalité et mixte.

Théorem 4.3.6 (Karush-Kuhn-Tucker)**.** *Si f admet un minimum local en $x^* \in K$ et que les contraintes sont qualifiées en x^* alors il existe $\mu_1, \ldots, \mu_p \in \mathbb{R}_+$ tels que*

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{p} \nabla g_i(x^*) = 0 \\ \mu_i g_i(x^*) = 0 \, \forall 1 \leq i \leq p \end{cases} \tag{4.4}$$

Les μ_i sont appelés multiplicateurs de Karush-Kuhn-Tucker.

Remarque.
— *Les multiplicateurs sont de signe positifs à condition que les contraintes soient formulées sous la forme $g_i(x) \leq 0$.*
— *La seconde équation signifie que les multiplicateurs sont nuls pour les contraintes non saturées. En effet, dans l'intérieur du domaine, l'équation d'Euler est satisfaite.*
— *Pour la recherche d'un maximum, les multiplicateurs sont négatifs.*

Démonstration. Comme x^* est un minimum local de f sur K, l'inéquation d'Euler est satisfaite en x^*. On a donc

$$K(x^*) \subset \{h \in \mathbb{R}^n, \langle \nabla f(x^*), h\rangle \geq 0\}.$$

Par ailleurs, les contraintes étant qualifiées en x^*, on a l'égalité

$$K(x^*) = \{h \in \mathbb{R}^n, \forall i \in \mathcal{I}(x^*) \, \langle \nabla g_i(x^*), h\rangle \leq 0\}.$$

D'après les lemme de Farkas, on a donc montré que $\nabla f(x^*)$ est une combinaison linéaire positive des $-\nabla g_i(x^*)$ pour $i \in \mathcal{I}(x^*)$. On complète alors la famille des μ_i par $\mu_i = 0$ si $i \notin \mathcal{I}(x^*)$. □

4.4 Contraintes mixtes

Le dernier type de contraintes étudié dans ce cours est la contrainte mixte pour laquelle le domine admissible est donné par

$$K = \{x \in \mathbb{R}^n, \, \forall i \in \mathcal{E}, g_i(x) = 0, \quad \forall i \in \mathcal{I}, \, g_i(x) \leq 0\},$$

avec $\mathcal{E}$ et $\mathcal{I}$ formant une partition de $[\![1, p]\!]$ et les fonction g_i de classe $\mathcal{C}^1$.

1. la fonction $\phi : t \mapsto \|v - tp_{\mathcal{C}}(v)\|^2$ est minimale en $t = 1$ donc $\phi'(1) = 0$
2. On montre que pour tout $x \in \mathcal{C}$, $\|x - p_{\mathcal{C}}(v)\|^2 \geq 2\langle v - p_{\mathcal{C}}(v), x - p_{\mathcal{C}}(v)\rangle$ et on l'applique à $x = \lambda x_0 + (1-\lambda)p_{\mathcal{C}}(v)$ avant de faire tendre $\lambda \to 0^+$.

Définition 4.4.1. *On dit que les contraintes sont qualifiées en $x \in K$ si*

— *la famille $(\nabla g_i(x))_{i\in\mathcal{E}}$ est libre*

— *il existe une direction rentrante en x, c'est à dire il existe $v \in \mathbb{R}^n$ tel que*

  — $\forall i \in \mathcal{E}$, $\langle \nabla g_i(x), v\rangle = 0$

  — $\forall i \in \mathcal{I}(x)$, $\langle \nabla g_i(x), v\rangle < 0$

De la même manière que pour les contraintes de type inégalité, on dispose d'une condition nécessaire plus simple.

Proposition 4.4.2. *Si la famille $(\nabla g_i(x))_{i\in\mathcal{E}\cup\mathcal{I}(x)}$ est libre alors les contraintes sont qualifiées en x.*

Lorsque les contraintes sont qualifiées en $x \in K$, l'espace des direction admissible est connu.

Lemme 4.4.3. *Si les contraintes sont quelifiées en $x \in K$, on a*

$$K(x) = \{h \in \mathbb{R}^n, \forall i \in \mathcal{E},\ \langle \nabla g_i(x), h\rangle = 0, \quad \forall i \in \mathcal{I}(x),\ \langle \nabla g_i(x), h\rangle < 0\}\,.$$

Démonstration. L'inclusion de $K(x)$ est toujours vérifiée, même lorsque les contraintes ne sont pas qualifiées. Cela se démontre de la même manière que pour le lemme 4.3.4.

Supposons pour commencer que $h \in \mathbb{R}^n$ vérifiant $\langle \nabla g_i(x), h\rangle = 0$ pour tout $i \in \mathcal{E}$ et $\langle \nabla g_i(x), h\rangle < 0$ pour tout $i \in \mathcal{I}(x)$. En suivant la démonstration de la proposition 4.2.3, on montre qu'il existe des suites $(\varepsilon_k)_k \in \mathbb{R}_+^*$ et $(h_k)_k \in \mathbb{R}^n$ telles que $\varepsilon \to 0$, $h_k \to h$ et $g_i(x + \varepsilon_k h_k) = 0$ pour tout $i \in \mathcal{E}$. De plus, si $i \in \mathcal{I} \setminus \mathcal{I}(x)$, la continuité de g_i assure que pour k assez grand, $g_i(x + \varepsilon_k h_k) \leq 0$. Enfin, soit $i \in \mathcal{I}(x)$, on a $g_i(x) = 0$ et d'après la formule de Taylor-Young

$$\lim_{k\to\infty} \frac{g_i(x + \varepsilon_k h_k)}{\varepsilon_k} = \langle \nabla g_i(x), h\rangle < 0.$$

On en déduit que $g_i(x + \varepsilon_k h_k) \leq 0$ pour k assez grand. Ainsi $k \in K(x)$.

Pour finir, supposons que $h \in \{h \in \mathbb{R}^n, \forall i \in \mathcal{E},\ \langle \nabla g_i(x), h\rangle = 0, \quad \forall i \in \mathcal{I}(x),\ \langle \nabla g_i(x), h\rangle < 0\}$. Les contraintes étant qualifiées, on dispose d'une direction rentrante v. On pose alors $h_\varepsilon = h + \varepsilon v$. Pour tout $\varepsilon > 0$, h_ε vérifie $\langle \nabla g_i(x), h_\varepsilon\rangle = 0$ pour tout $i \in \mathcal{E}$ et $\langle \nabla g_i(x), h_\varepsilon\rangle < 0$ pour tout $i \in \mathcal{I}(x)$. On a donc montré que $h_\varepsilon \in K(x)$. Or $K(x)$ est fermé. on en déduit donc que $h \in K(x)$. □

Dans le cadre des contrainte de type mixte, il faut généraliser le lemme de Farkas.

Lemme 4.4.4 (Farkas généralisé)**.** *Soient $u_1, \ldots, u_p$ et v des vecteurs de $\mathbb{R}^n$ et $\mathcal{E}, \mathcal{I}$ deux sous-ensembles disjoints, de $[\![1, m]\!]$. On a équivalence entre*

(i) $\{h \in \mathbb{R}^n, \forall i \in \mathcal{E}\ \langle u_i, h\rangle = 0 \quad \forall i \in \mathcal{I}\ \langle u_i, h\rangle \geq 0\} \subset \{h \in \mathbb{R}^n, \langle v, h\rangle \geq 0\}$

(ii)

$$v \in C := \left\{ \sum_{i\in\mathcal{E}\cup\mathcal{I}} \mu_i u_i, \forall i \in \mathcal{I}, \mu_i \geq 0 \right\}.$$

Démonstration. On commence par remarquer que C est fermé de la même manière que pour le lemme de Farkas. Pour cela il faut se ramener à des coefficients positifs. On décompose donc $\mu_i = \mu_i^+ - \mu_i^-$ pour $i \in \mathcal{E}$ et on applique la démonstration à la famille (nécessairement non libre) des vecteurs u_i, $i \in \mathcal{I}$, u_i, $i \in \mathcal{E}$ et $-u_i$, $i \in \mathcal{E}$.

On remarque ensuite que $(ii) \Rightarrow (i)$ est triviale. Supposons alors $v \notin C$. On pose alors $h = p_C(v) - v \neq 0$. On montre alors que $\langle v, h\rangle < 0$ et pour tout $u \in C$ on a $\langle u, h\rangle \geq 0$. On en déduit donc que pour tout $i \in \mathcal{E}$, $\langle u_i, h\rangle = 0$ et pour tout $i \in \mathcal{I}$, $\langle u_i, h\rangle \geq 0$. Ce qui achève la preuve. □

On peut alors énoncé le théorème de Karsuch-Kuhn-Tucker généralisé.

Théorem 4.4.5 (Karush-Kuhn-Tucker généralisé)**.** *Si f admet un minimum local sur K en x^* et si les contraintes sont quialifiées en x^* alors il existe des multiplicateurs $\mu_1, \ldots, \mu_p \in \mathbb{R}$ tell que*

$$\begin{cases} \nabla f(x^*) + \sum_{i=1}^{p} \nabla g_i(x^*) = 0 \\ \mu_i g_i(x^*) = 0 \, \forall 1 \in \mathcal{I} \end{cases} \tag{4.5}$$

et pour tout $i \in \mathcal{I}$, $\mu_i \geq 0$.

Démonstration. La démonstration est la même que celle du théorème de Karush-Kuhn-Tucker. □

Chapitre 5

Algorithmes d'optimisation