

Chap III - EXTREMA LIÉS - OPTIMISATION SOUS CONTRAINTE

Où les contraintes ne sont pas si contraignantes,
du moment qu'elles sont C^1

1 Introduction

Un grand nombre de problèmes peuvent se mathématiser par:

Soit $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction, $C \subset U$ le sous-ensemble des points "admissibles", on veut

$$(\mathcal{P}) \begin{cases} \text{Minimiser } f(x) \\ x \in C \end{cases} \quad \text{ou} \quad (\mathcal{Q}) \begin{cases} \text{Maximiser } f(x) \\ x \in C \end{cases}$$

Un point $a \in C$ est solution de (\mathcal{P}) (resp (\mathcal{Q})) si, pour tout $x \in C$, $f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$).

- *En microéconomie*, les consommateurs veulent maximiser l'utilité u que leur procurent leurs achats, sous contrainte de ne pas dépasser leur budget w . Ils résolvent:

$$(\mathcal{Q}) \begin{cases} \text{Maximiser } u(x_1, \dots, x_n) \\ x \in C = \{x \in \mathbb{R}^n, x_i \geq 0, p_1x_1 + \dots + p_nx_n \leq w\} \end{cases}$$

- *En théorie des jeux*, mettons dans un jeu à deux joueurs, chaque joueur cherche à maximiser ses gains, sous contrainte de la stratégie choisie par l'autre.

On représente les gains de chacun des 2 joueurs par deux matrices de gains $P, Q \in \mathcal{M}_{m,n}(\mathbb{R})$, et les stratégies disponibles aux joueurs sont respectivement des éléments de

$$X = \{x \in \mathbb{R}^m, x_i \geq 0, \sum_{i=1}^m x_i = 1\}$$
$$Y = \{y \in \mathbb{R}^n, y_j \geq 0, \sum_{j=1}^n y_j = 1\}$$

Les gains du premier joueur sont donnés par $p(x, y) = {}^t x P y$. Il cherche donc à résoudre

$$(\mathcal{Q}) \begin{cases} \text{Maximiser } \min_{y \in Y} p(x, y) \\ x \in X \end{cases}$$

- *En statistiques*, l'estimateur du maximum de vraisemblance est solution d'un problème d'optimisation. Si X suit une loi dépendant d'un paramètre $\theta \in \mathcal{D}_\theta$, et f_θ est la densité de probabilité de cette loi, alors la vraisemblance au vu d'un échantillon (x_1, \dots, x_n) est la fonction

$$\mathcal{L}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_\theta(x_i)$$

et on cherche

$$(\mathcal{Q}) \begin{cases} \text{Maximiser } \mathcal{L}(x_1, \dots, x_n; \theta) \\ \theta \in \mathcal{D}_\theta \end{cases}$$

Rappelons ce qu'on sait déjà en terme d'extrema:

Optimisation sans contraintes.

\rightsquigarrow Le calcul différentiel est un outil puissant pour étudier les extremas d'une fonction $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ sur un ouvert U .

CS0 Si $K \subset \mathbb{R}^n$ est compact, et $f : K \rightarrow \mathbb{R}$ est continue, alors f est bornée et atteint ses bornes sur K . Autrement dit, il existe $a_{max}, a_{min} \in K$ tels que, pour tout $x \in K$,

$$f(x_{min}) \leq f(x) \leq f(x_{max}).$$

CN1 Si f a un extremum local en $a \in U$ et si $Df(a)$ existe, alors $Df(a) = 0_{(\mathbb{R}^n)^*}$ (c'est-à-dire $\nabla f(a) = 0_{\mathbb{R}^n}$). On dit que a est un **point critique** de f .

CN2 Si f a un minimum local en $a \in U$ et si f est deux fois différentiable en a , alors $\nabla f(a) = 0_{\mathbb{R}^n}$ et la Hessienne $Hf(a)$ est une matrice symétrique positive.

CS2 Si f est deux fois différentiable en $a \in U$ avec $\nabla f(a) = 0$ et $Hf(a)$ définie positive, alors f a un minimum local en a .

Question: Comment résoudre le problème d'optimisation

$$\begin{cases} \text{Minimiser (resp. maximiser) } f(x) \\ x \in U \cap \Gamma \end{cases}$$

lorsque Γ n'est pas ouvert de \mathbb{R}^n ?

2 TEL - une seule contrainte

Cas "linéaire": Γ est un hyperplan affine

Supposons que Γ soit un ensemble affine, décrit par une équation linéaire:

$$\Gamma = \{x \in \mathbb{R}^n, \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n = c\}$$

où les α_i sont des coefficients réels *non tous nuls*. On va supposer, pour faire simple, que $\alpha_n \neq 0$. On a donc:

$$x \in \Gamma \iff x_n = \frac{1}{\alpha_n}(c - \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_{n-1} x_{n-1})$$

\rightsquigarrow Posons

$$\varphi(x_1, \dots, x_{n-1}) = \frac{1}{\alpha_n}(c - \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_{n-1} x_{n-1})$$

et

$$\tilde{U} = \{x' = (x_1, \dots, x_{n-1}) \in \mathbb{R}^{n-1}, (x', \varphi(x')) \in U\}$$

Exercice: Montrer que c'est un ouvert de \mathbb{R}^{n-1} .

On peut alors définir la fonction

$$\tilde{f} : x' \in \tilde{U} \mapsto f\left(x', \frac{1}{\alpha_n}(c - \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_{n-1} x_{n-1})\right) \in \mathbb{R}$$

Alors

$$(\mathcal{P}_\Gamma) \begin{cases} \text{Minimiser } f(x) \\ x \in \Gamma \end{cases} \iff (\mathcal{P}_\Gamma) \begin{cases} \text{Minimiser } \tilde{f}(x') \\ x' \in \tilde{U} \end{cases}$$

C'est devenu un problème d'optimisation sans contrainte !

\rightsquigarrow Pour tout $i = 1 \dots n-1$, notons $x = \left(x', \frac{1}{\alpha_n}(c - \alpha_1 x_1 + \dots + \alpha_{n-1} x_{n-1})\right)$.

Alors

$$\frac{\partial \tilde{f}}{\partial x_i}(x') = \frac{\partial f}{\partial x_i}(x) - \frac{\alpha_i}{\alpha_n} \frac{\partial f}{\partial x_n}(x)$$

donc, d'après [CN1], si $a = (a', \varphi(a'))$ est un extremum de f sur $U \cap \Gamma$, alors pour tout $i = 1, \dots, n-1$,

$$(\star) \quad \frac{\partial \tilde{f}}{\partial x_i}(a') = 0 \iff \frac{\partial f}{\partial x_i}(a) = \frac{\alpha_i}{\alpha_n} \frac{\partial f}{\partial x_n}(a)$$

Remarquons que, si on note g la fonction décrivant Γ :

$$g : x \in \mathbb{R}^n \mapsto \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n - c \in \mathbb{R} \rightsquigarrow \Gamma = \{x \in \mathbb{R}^n, g(x) = 0\}$$

Alors g est différentiable et on a, pour tout $x \in \mathbb{R}^n$

$$\nabla g(x) = (\alpha_1, \dots, \alpha_n)$$

et donc (\star) donne

$$\frac{\partial \tilde{f}}{\partial x_i}(a') = 0 \iff \frac{\partial f}{\partial x_i}(a) = \underbrace{\frac{1}{\alpha_n} \frac{\partial f}{\partial x_n}(a)}_{:=\lambda} \frac{\partial g}{\partial x_i}(a)$$

On a donc trouvé que, si a est un extremum de f sur Γ , alors il existe $\lambda \in \mathbb{R}$ tel que

$$\forall i = 1, \dots, n \quad \frac{\partial f}{\partial x_i}(a) = \lambda \frac{\partial g}{\partial x_i}(a) \text{ i.e. } \nabla f(a) = \lambda \nabla g(a)$$

et pour ça, il nous a fallu reformuler le problème d'optimisation contraint par l'équation de Γ en un problème d'optimisation sur un ouvert, en exprimant une des variables de cette équation en fonction des autres.

\rightsquigarrow C'est ce résultat que le théorème des extrema liés va généraliser.

Cas général à une contrainte : Théorème des extrema liés 1

Soit U un ouvert de \mathbb{R}^n et $f : U \rightarrow \mathbb{R}$. On cherche les extremas de f sur un ensemble du type $\Gamma = \{x \in U, g(x) = 0\}$, où $g : U \rightarrow \mathbb{R}$ est une autre application.

On s'intéresse donc aux problèmes d'optimisation du type

$$(\mathcal{P}_\Gamma) \begin{cases} \text{Minimiser } f(x) \\ x \in \Gamma \end{cases} \quad \text{ou} \quad (\mathcal{Q}_\Gamma) \begin{cases} \text{Maximiser } f(x) \\ x \in \Gamma \end{cases}$$

Théorème 1 (TEL1)

Soit $U \subset \mathbb{R}^n$ ouvert, $f, g : U \rightarrow \mathbb{R}$ deux fonctions \mathcal{C}^1 . On note

$$\Gamma = g^{-1}(\{0\}) = \{x \in U, g(x) = 0\}$$

Supposons que

- $f|_{\Gamma}$ admet un extremum local en $a \in \Gamma$
- Le vecteur $\nabla g(a)$ est non nul.

Alors il existe un unique $\lambda \in \mathbb{R}$ tel que $\nabla f(a) = \lambda \nabla g(a)$.

On appelle λ le *multiplicateur de Lagrange*.

Démonstrons-le sans plus attendre. A quoi vous fait penser cette histoire de reformuler $g(x) = 0$ en $x_n = \varphi(x_1, \dots, x_{n-1})$?

Preuve du TEL1

Puisque $\nabla g(a) \neq 0_{(\mathbb{R}^n)^*}$, au moins l'une des dérivées partielles de g en a est non nulle. Supposons que $\frac{\partial g}{\partial x_n}(a) \neq 0$.

Pour $x \in \mathbb{R}^n$, on notera $x = (x', x_n) \in \mathbb{R}^{n-1} \times \mathbb{R}$, avec $x' = (x_1, \dots, x_{n-1})$. On note donc $a = (a', a_n)$.

Puisque $\frac{\partial g}{\partial x_n}(a) \neq 0$, par le théorème des fonctions implicites, il existe un voisinage V de a' dans \mathbb{R}^{n-1} , un voisinage W de a_n dans \mathbb{R} et une application \mathcal{C}^1 $\varphi : V \rightarrow W$ telle que

$$\underbrace{x = (x', x_n) \in (V \times W) \cap \Gamma}_{\text{i.e. } g(x', x_n) = 0} \iff x_n = \varphi(x')$$

Pour $x' \in V$, on pose $F(x') = f(x', \varphi(x'))$. Alors $f|_{\Gamma}$ a un extremum local en a ssi $F : V \subset \mathbb{R}^{n-1} \rightarrow \mathbb{R}$ a un extremum local en a' .

↪ On est donc ramené à un problème d'extremum sur un ouvert.

Or, $F = f \circ \Psi$, où $\Psi : x' \in V \mapsto (x', \varphi(x')) \in U$, avec:

$$\text{Jac } f(x) = \left(\frac{\partial f}{\partial x_1}(x), \dots, \frac{\partial f}{\partial x_n}(x) \right),$$

$$\text{Jac } \Psi(x') = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \\ \hline \frac{\partial \varphi}{\partial x_1}(x') & \dots & \frac{\partial \varphi}{\partial x_{n-1}}(x') \end{pmatrix}$$

Donc

$$\begin{aligned} DF(x')(h') &= Df(x', \varphi(x'))(h', D\varphi(x')(h')) \\ &= \frac{\partial f}{\partial x_1}(x)h'_1 + \dots + \frac{\partial f}{\partial x_{n-1}}(x)h'_{n-1} \\ &\quad + \frac{\partial f}{\partial x_n}(x) \left(\frac{\partial \varphi}{\partial x_1}(x')h'_1 + \dots + \frac{\partial \varphi}{\partial x_{n-1}}(x')h'_{n-1} \right) \\ &= \sum_{i=1}^{n-1} \left(\frac{\partial f}{\partial x_i}(x) + \frac{\partial f}{\partial x_n}(x) \frac{\partial \varphi}{\partial x_i}(x') \right) h'_i \end{aligned}$$

Comme on a supposé que a est un extremum de $f|_{\Gamma}$, a' est un extremum de F , donc on a $DF(a') = 0_{(\mathbb{R}^{n-1})^*}$, autrement dit, pour tout $i \in \llbracket 1, n-1 \rrbracket$,

$$\frac{\partial f}{\partial x_i}(a) + \frac{\partial f}{\partial x_n}(a) \frac{\partial \varphi}{\partial x_i}(a') = 0$$

Par ailleurs, pour tout $x' \in V$, on a $g(x', \varphi(x')) = 0$ donc par un calcul similaire, pour tout $i \in \llbracket 1, n-1 \rrbracket$,

$$\frac{\partial g}{\partial x_i}(a) + \frac{\partial g}{\partial x_n}(a) \frac{\partial \varphi}{\partial x_i}(a') = 0$$

Donc, pour tout $i \in \llbracket 1, n-1 \rrbracket$,

$$\frac{\partial f}{\partial x_i}(a) = -\frac{\partial f}{\partial x_n}(a) \frac{\partial \varphi}{\partial x_i}(a') = -\underbrace{\frac{\frac{\partial f}{\partial x_n}(a)}{\frac{\partial g}{\partial x_n}(a)}}_{:=\lambda} \frac{\partial g}{\partial x_n}(a) \frac{\partial \varphi}{\partial x_i}(a') = \lambda \frac{\partial g}{\partial x_i}(a)$$

On a aussi $\frac{\partial f}{\partial x_n}(a) = \lambda \frac{\partial g}{\partial x_n}(a)$, donc $\nabla f(a) = \lambda \nabla g(a)$. □

Exemple: Cherchons les extremas de $f : (x, y) \in \mathbb{R}^2 \mapsto x + y \in \mathbb{R}$ sur l'ensemble

$$\Gamma = \{(x, y) \in \mathbb{R}^2, x^4 + y^4 = 1\} = g^{-1}(\{0\}), \quad \text{avec } g : (x, y) \in \mathbb{R}^2 \mapsto x^4 + y^4 - 1 \in \mathbb{R}$$

On a, pour tout $(x, y) \in \Gamma$,

$$\nabla g(x, y) = {}^t(4x^3, 4y^3) \neq 0_{\mathbb{R}^2}$$

Par le théorème des extremas liés, si $f|_{\Gamma}$ admet un extremum en $(a, b) \in \Gamma$, alors il existe $\lambda \in \mathbb{R}$ tel que

$$\nabla f(a, b) = \lambda \nabla g(a, b) \iff \begin{cases} 1 = 4\lambda a^3 \\ 1 = 4\lambda b^3 \end{cases}$$

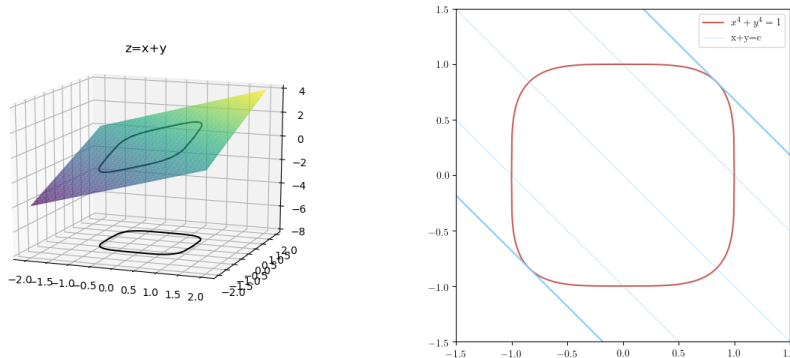
d'où $\lambda \neq 0$ et $a = (4\lambda)^{-\frac{1}{3}} = b$. Or $(a, b) \in \Gamma$ donc $2a^4 = 1$, ce qui donne deux candidats à l'extrémitude:

$$(a_1, b_1) = (-2^{-\frac{1}{4}}, -2^{-\frac{1}{4}}), \quad (a_2, b_2) = (2^{-\frac{1}{4}}, 2^{-\frac{1}{4}})$$

Or, puisque f est continue sur le compact Γ , on sait que f admet sur Γ un maximum global et un minimum global, qui sont forcément parmi nos candidats. Or:

$$f(-2^{-\frac{1}{4}}, -2^{-\frac{1}{4}}) = -2^{\frac{3}{4}} < f(2^{-\frac{1}{4}}, 2^{-\frac{1}{4}}) = 2^{\frac{3}{4}}$$

Donc f admet un minimum global en $(-2^{-\frac{1}{4}}, -2^{-\frac{1}{4}})$ et un maximum global en $(2^{-\frac{1}{4}}, 2^{-\frac{1}{4}})$.



3 TEL - Plusieurs contraintes

On s'intéresse maintenant aux ensembles décrits par, non pas une, mais p équations:

$$\Gamma = \{x \in U, g_1(x) = 0, g_2(x) = 0, \dots, g_p(x) = 0\}$$

On note $g : x \in U \subset \mathbb{R}^n \mapsto (g_1(x), \dots, g_p(x)) \in \mathbb{R}^p$.

Cas "linéaire"

Supposons que pour tout $j = 1, \dots, p$, g_j est une fonction affine:

$$\begin{aligned} g_1(x) &= m_{1,1}x_1 + \dots + m_{n,1}x_n - b_1 \\ g_2(x) &= m_{1,2}x_1 + \dots + m_{n,2}x_n - b_2 \\ &\vdots \\ g_p(x) &= m_{1,p}x_1 + \dots + m_{n,p}x_n - b_p \end{aligned}$$

Autrement dit $g(x) = Mx - b$, où $M = (m_{i,j}) \in \mathcal{M}_{p,n}(\mathbb{R})$ et $b \in \mathbb{R}^p$.

De plus, g est différentiable sur \mathbb{R}^n , et sa jacobienne est M : en particulier, pour $j = 1 \dots p$, $\nabla g_j(a)$ est donné par la j -ième ligne de M .

Alors le problème qui nous intéresse est

$$(\mathcal{P}_\Gamma) \begin{cases} \text{Minimiser } f(x) \\ Mx = b \\ x \in U \end{cases}$$

Si $\text{rg}(M) = p$, alors les n colonnes de M forment une famille génératrice de \mathbb{R}^p . On peut donc en extraire une base: supposons que les p dernières colonnes forment une base de \mathbb{R}^p

$$C_{n-p} = \begin{pmatrix} m_{1,n-p+1} \\ \vdots \\ m_{p,n-p+1} \end{pmatrix}, \dots, C_n = \begin{pmatrix} m_{1,n} \\ \vdots \\ m_{p,n} \end{pmatrix}$$

Alors la matrice carrée $B = (C_{n-p} \mid \dots \mid C_n)$ est inversible. Notons A l'autre sous-matrice de M , celle qui contient les $n - p$ premières colonnes: $A = (C_1 \mid \dots \mid C_{n-p})$. On a alors

$$x \in \Gamma \iff Mx = b \iff Ax_A + Bx_B = b \iff x_B = B^{-1}(b - Ax_A)$$

en notant $x_A = (x_1, \dots, x_{n-p}) \in \mathbb{R}^{n-p}$, $x_B = (x_{n-p+1}, \dots, x_n)$. Et du coup notre problème devient équivalent à

$$(\mathcal{P}'_{\Gamma}) \begin{cases} \text{Minimiser } f(x_A, B^{-1}(b - Ax_A)) \\ x_A \in \mathbb{R}^{n-p} \end{cases} \iff \begin{cases} \text{Minimiser } \tilde{f}(u) \\ u \in \mathbb{R}^{n-p} \end{cases}$$

si on pose $\tilde{f} : u \in \mathbb{R}^{n-p} \mapsto f(u, B^{-1}(b - Au)) \in \mathbb{R}$.

\rightsquigarrow C'est un problème sans contrainte ! Plus qu'à voir quand $\nabla \tilde{f}(u) = 0_{\mathbb{R}^{n-p}}$.

Notons $\nabla f(x) = \begin{pmatrix} \nabla f_A(x) \\ \nabla f_B(x) \end{pmatrix}$ avec $\nabla f_A(x) \in \mathbb{R}^{n-p}$, $\nabla f_B(x) \in \mathbb{R}^p$.

Exercice: Montrer, à grand coup de composition des différentielles, que

$$\nabla \tilde{f}(u) = \nabla f_A(x) - \nabla f_B(x) \cdot B^{-1} \cdot A$$

où $x = (u, B^{-1}(b - Au)) \in \mathbb{R}^n$.

Du coup, on obtient que x est un extremum local de $f|_{\Gamma}$ ssi

$$\text{Jac } f_A(x) = \text{Jac } f_B(x) \cdot B^{-1} \cdot A$$

et on a bien sûr aussi

$$\text{Jac } f_B(x) = \text{Jac } f_B(x) \cdot B^{-1} \cdot B$$

Remarquons que $\text{Jac } f_B(x) \cdot B^{-1}$ est une matrice ligne à p colonnes (vérifiez-le !) qu'on va noter

$$\text{Jac } f_B(x) \cdot B^{-1} = \Lambda = (\lambda_1 \quad \dots \quad \lambda_p)$$

donc la matrice-ligne $\text{Jac } f(x) = (\text{Jac } f_A(x) \quad \text{Jac } f_B(x))$ est de la forme

$$\text{Jac } f(x) = (\Lambda A \quad \Lambda B) = \Lambda M = \Lambda \text{Jac } g(x)$$

ou encore

$$\nabla f(x) = {}^t M^t \Lambda = \lambda_1 \nabla g_1(x) + \dots + \nabla g_p(x)$$

Ca ressemble beaucoup à ce qu'on avait obtenu dans le cas à une contrainte: sauf que maintenant, il ne suffit plus que la matrice de la contrainte g soit non nulle: il faut qu'elle soit de rang p .

Et on se retrouve avec p multiplicateurs de Lagrange.

TEL2 - Cas général

Théorème 2 (TEL2)

Soient $f : U \rightarrow \mathbb{R}$ et $g : U \rightarrow \mathbb{R}^p$ deux applications \mathcal{C}^1 . On note $\Gamma = g^{-1}(\{0\}) = \{x \in U, g_1(x) = \dots = g_p(x) = 0\}$. Supposons que

- $f|_{\Gamma}$ admet un extremum local en $a \in \Gamma$
- L'application linéaire $Dg(a) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ est surjective.

Alors il existe un unique $(\lambda_1, \dots, \lambda_p) \in \mathbb{R}^p$ tel que

$$\nabla f(a) = \lambda_1 \nabla g_1(a) + \dots + \lambda_p \nabla g_p(a)$$

Les λ_i sont appelés *multiplicateurs de Lagrange*.

Preuve du TEL2

L'idée est virtuellement la même que pour le cas à une contrainte, avec une bonne dose d'algèbre linéaire pour inverser des matrices pour faire comme dans le cas linéaire.

Puisque $Dg(a)$ est surjective, on a $\text{rg}(Dg(a)) = \dim \Im(Dg(a)) = p$. Donc, si on considère la matrice jacobienne de g , c'est-à-dire la matrice de $Dg(a)$ dans la base canonique,

$$\begin{pmatrix} \frac{\partial g_1}{\partial x_1}(a) & \dots & \frac{\partial g_1}{\partial x_{n-p}}(a) & \frac{\partial g_1}{\partial x_{n-p+1}}(a) & \dots & \frac{\partial g_1}{\partial x_n}(a) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial x_1}(a) & \dots & \frac{\partial g_p}{\partial x_{n-p}}(a) & \frac{\partial g_p}{\partial x_{n-p+1}}(a) & \dots & \frac{\partial g_p}{\partial x_n}(a) \end{pmatrix} = \left(\begin{array}{c|c|c|c} \frac{\partial g}{\partial x_1}(a) & \frac{\partial g}{\partial x_2}(a) & \dots & \frac{\partial g}{\partial x_n}(a) \end{array} \right)$$

on en déduit que

$$\dim \text{Vect} \left(\frac{\partial g}{\partial x_i}(a) \right)_{1 \leq i \leq n} = p$$

On extrait une base de cette famille génératrice: quitte à renuméroter les variables, on suppose que la famille $\left(\frac{\partial g}{\partial x_i}(a) \right)_{n-p+1 \leq i \leq n}$ est libre.

Pour $x \in \mathbb{R}^n$, on pose $x = (y, z) \in \mathbb{R}^{n-p} \times \mathbb{R}^p$, et on note $a = (b, c)$. Avec ces choix de notation, on a donc

$$\left(\begin{array}{ccc|ccc} \frac{\partial g_1}{\partial y_1}(a) & \cdots & \frac{\partial g_1}{\partial y_{n-p}}(a) & \frac{\partial g_1}{\partial z_1}(a) & \cdots & \frac{\partial g_1}{\partial z_p}(a) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_p}{\partial y_1}(a) & \cdots & \frac{\partial g_p}{\partial y_{n-p}}(a) & \frac{\partial g_p}{\partial z_1}(a) & \cdots & \frac{\partial g_p}{\partial z_p}(a) \end{array} \right)$$

$\underbrace{\hspace{15em}}_{D_y g(a)}$
 $\underbrace{\hspace{15em}}_{D_z g(a)}$

où $D_z g(a) : F \rightarrow \mathbb{R}^p$ inversible, puisque ses colonnes forment une famille libre.

On peut donc appliquer le Théorème des Fonctions Implicites à l'équation $g(y, z) = 0$: il existe

- V voisinage de b dans \mathbb{R}^{n-p}
- W voisinage de c dans \mathbb{R}^p
- $\varphi : V \rightarrow W$ une application \mathcal{C}^1

tels que

$$(y, z) \in \underbrace{\Gamma \cap (V \times W)}_{\text{i.e. } y \in V, z \in W, g(y, z) = 0} \iff z = \varphi(y)$$

Alors a est un extremum local de $f|_{\Gamma}$ si, et seulement si b est un extremum local de

$$F : V \rightarrow \mathbb{R}, y \mapsto f(y, \varphi(y))$$

Dans ce cas, c'est un point critique de F :

$$DF(b) = D_y f(a) + D_z f(a) \circ D\varphi(b) = 0.$$

De plus, pour tout $y \in V$, $g(y, \varphi(y)) = 0$ donc, en $a = (b, c)$,

$$D_y g(a) + D_z g(a) \circ D\varphi(b) = 0.$$

On a donc

$$\begin{aligned} D_y f(a) &= -D_z f(a) \circ D\varphi(b) \\ &= -D_z f(a) \circ (D_z g(a)^{-1} \circ D_z g(a)) \circ D\varphi(b) \\ &= (D_z f(a) \circ D_z g(a)^{-1}) \circ (-D_z g(a) \circ D\varphi(b)) \\ &= (D_z f(a) \circ D_z g(a)^{-1}) \circ D_y g(a). \end{aligned}$$

Par ailleurs, on a bien sûr $D_z f(a) = (D_z f(a) \circ D_z g(a)^{-1}) \circ D_z g(a)$, donc

$$Df(a) = (D_z f(a) \circ D_z g(a)^{-1}) \circ Dg(a)$$

Or, $D_z f(a) \circ D_z g(a)^{-1} : F \rightarrow \mathbb{R}$ est une forme linéaire sur F , qui est de dimension p . Notons $(\lambda_1, \dots, \lambda_p)$ sa matrice dans la base (f_1, \dots, f_p) de F , alors on a obtenu, pour tout $k \in F$

$$\begin{aligned} Df(a)(k) &= (\lambda_1, \dots, \lambda_p) \cdot \begin{pmatrix} Dg_1(a)(k) \\ \vdots \\ Dg_p(a)(k) \end{pmatrix} \\ &= \lambda_1 Dg_1(a)(k) + \dots + \lambda_p Dg_p(a)(k) = \left(\sum_i \lambda_i Dg_i(a) \right) (k), \end{aligned}$$

comme requis. □

Exemple: Cherchons les extrema de $f : (x, y, z) \in \mathbb{R}^3 \rightarrow x + z$ sur $\Gamma = g^{-1}(\{0\})$, où

$$g : (x, y, z) \in \mathbb{R}^3 \mapsto (x^2 + y^2 + z^2 - 1, z - y) \in \mathbb{R}^2$$

Alors, pour tout $u = (x, y, z) \in \Gamma$,

$$Dg(u) = \begin{pmatrix} 2x & 2y & 2z \\ 0 & -1 & 1 \end{pmatrix}$$

est de rang 2. Donc, si u est un extremum de $f|_{\Gamma}$, alors il existe λ_1 et λ_2 tels que

$$\nabla f(u) = \lambda_1 \nabla g_1(u) + \lambda_2 \nabla g_2(u) \iff \begin{cases} 1 &= 2\lambda_1 x \\ 0 &= -\lambda_2 + 2\lambda_1 y \\ 1 &= \lambda_2 + 2\lambda_1 z \end{cases}$$

On en déduit que $\lambda_1 \neq 0$ et que $2\lambda_1 x = 1 = 2\lambda_1(y + z)$, d'où $x = y + z$.

De plus,

$$u \in \Gamma \iff \begin{cases} y = z \\ x^2 + y^2 + z^2 = 1 \end{cases} \quad \text{donc} \quad \begin{cases} y = z, & x = 2z \\ 6z^2 = 1 \end{cases}$$

Il y a donc deux candidats à l'extrémitude:

$$u_1 = \left(-\frac{2\sqrt{6}}{6}, -\frac{\sqrt{6}}{6}, -\frac{\sqrt{6}}{6} \right), u_2 = \left(\frac{2\sqrt{6}}{6}, \frac{\sqrt{6}}{6}, \frac{\sqrt{6}}{6} \right)$$

Réciproquement, puisque f est continue et Γ compact, f est bornée et atteint ses bornes: les deux candidats sont donc un min global et un max global.

Plus précisément, $f(u_1) = -\frac{\sqrt{6}}{2} < \frac{\sqrt{6}}{2} = f(u_2)$, donc u_1 est un min et u_2 un max.

Annexe: Interprétation géométrique

Question: Comment décrire une courbe dans \mathbb{R}^2 ? Une surface dans \mathbb{R}^3 ?

Courbes dans le plan

Il y a deux façons de décrire une courbe dans \mathbb{R}^2 :

- *Avec une paramétrisation:* Une courbe Γ dans \mathbb{R}^2 peut être vue comme une trajectoire: à chaque "instant" t on se situe en un point $(x(t), y(t))$ du plan, et le déplacement au cours du temps trace une courbe, décrite par une paramétrisation

$$\gamma : t \in I \mapsto \gamma(t) = (x(t), y(t)) \in \mathbb{R}^2$$

\rightsquigarrow Si γ est C^1 , pour tout $t_0 \in I$, on dit que le vecteur $\gamma'(t_0) = (x'(t_0), y'(t_0))$ est le vecteur tangent à la courbe Γ en t_0 .

Illustration: <https://www.geogebra.org/m/vq24zdsb>.

\rightsquigarrow La droite tangente à la courbe en t_0 est la droite affine de direction $\gamma'(t_0)$ qui passe par $\gamma(t_0)$.

- *Avec une équation:* Une courbe Γ de \mathbb{R}^2 peut aussi être décrite par une équation:

$$\Gamma = \{(x, y) \in \mathbb{R}^2, g(x, y) = 0\}$$

\rightsquigarrow Si on a les deux, on a donc, pour tout $t \in I$, $g(\gamma(t)) = 0$.

Si g et γ sont C^1 , cela donne

$$Dg(\gamma(t))(\gamma'(t)) = 0 \text{ i.e } \langle \nabla g(\gamma(t)), \gamma'(t) \rangle = 0$$

La droite tangente à g est la droite affine orthogonale à $\text{Vect}(\nabla g(t_0))$ qui passe par $\gamma(t_0)$. Autrement dit, pour tout $(x, y) \in \Gamma$, la droite tangente est l'ensemble

$$(x, y) + \text{Vect}(\nabla g(x, y))^\perp$$

Surfaces dans \mathbb{R}^3

On peut décrire une surface dans \mathbb{R}^3 par une équation :

$$\mathcal{S} = \{u = (x, y, z) \in \mathbb{R}^3, g(u) = 0\}.$$

Définition 3

On dit qu'un vecteur $v \in \mathbb{R}^3$ est **tangent** à \mathcal{S} en $u_0 \in \mathcal{S}$ s'il existe une courbe $\mathcal{C}^1 \gamma :]-1, 1[\rightarrow \mathcal{S}$ telle que $\gamma(0) = u_0, \gamma'(0) = v$.

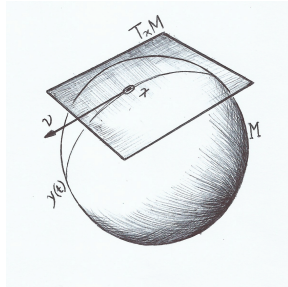
Dans ce cas, pour tout $t \in]-1, 1[$, $g(\gamma(t)) = 0$ donc, par la formule de composition,

$$\langle \nabla g(u_0), \gamma'(0) \rangle = 0$$

\rightsquigarrow L'ensemble de tous les vecteurs tangents forme un plan dans \mathbb{R}^3 . Et c'est l'espace affine $\{u_0\} + \nabla g(u_0)^\perp$.

Exemples

1. $S^2 = \{(x, y, z) \in \mathbb{R}^3, x^2 + y^2 + z^2 = 1\}$ la sphère unité, alors le plan tangent en (x, y, z) passe par (x, y, z) et est orthogonal à $(2x, 2y, 2z)$: il est perpendiculaire au vecteur (x, y, z) .



2. Si $\mathcal{P} = \{(x, y, z) \in \mathbb{R}^3, ax + by + cz = 0\}$ est un plan vectoriel, son plan tangent en tout point est l'orthogonal de (a, b, c) : c'est \mathcal{P} lui-même.

Interprétation géométrique du TEL

Dans le cas à une contrainte, le théorème des extrema liés dit que, si $f|_\Gamma$ a un extremum local en $a \in \Gamma$, alors il existe $\lambda \in \mathbb{R}$ tel que

$$\nabla f(a) = \lambda \nabla g(a)$$

Géométriquement : Dans ce cas $\nabla f(a)$ et $\nabla g(a)$ sont colinéaires, donc

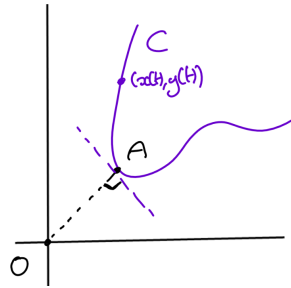
$$\nabla f(a)^\perp = \nabla g(a)^\perp \text{ donc } \{a\} + \nabla f(a)^\perp = \{a\} + \nabla g(a)^\perp$$

Or, $a + \nabla g(a)^\perp$ est l'espace tangent de Γ en a et $a + \nabla f(a)^\perp$ est l'espace tangent de l'ensemble $\{x \in U, f(x) = f(a)\}$ en a .

\rightsquigarrow A un extremum local de f , la surface Γ est tangente à une ligne de niveau de la fonction f .

Exemple : Soit C une courbe dans \mathbb{R}^2 , quels sont les points de C les plus proches de l'origine ?

\rightsquigarrow Ceci revient à chercher $\min_C(f(x, y))$ avec $f(x, y) = x^2 + y^2$.



▷ Si C est décrite par une paramétrisation $\gamma(t) = (x(t), y(t)) : I \rightarrow \mathbb{R}^2$, alors il s'agit de minimiser $f \circ \gamma : t \in I \mapsto x(t)^2 + y(t)^2 \in \mathbb{R}$. Or

$$(f \circ \gamma)'(t) = 2x(t)x'(t) + 2y(t)y'(t)$$

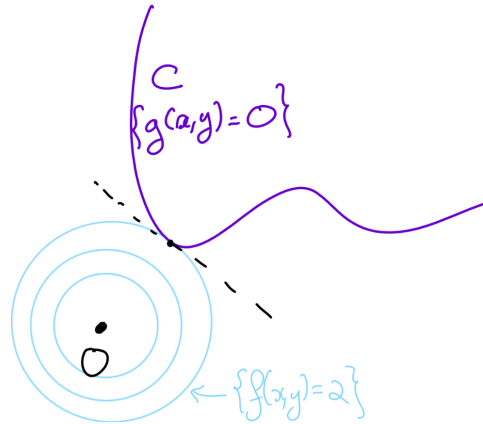
ce qui donne 0 ssi $\nabla f(\gamma(t)) \perp \gamma'(t)$.

\rightsquigarrow Autrement dit, le minimum, s'il existe, est un point $A = (x_0, y_0)$ de C tel que $\nabla f(A)$ est perpendiculaire au vecteur tangent à C en A .

▷ Si C est définie implicitement par une équation $g(x, y) = 0$, on cherche le minimum de $x^2 + y^2$ lorsque x et y sont liés par cette relation.

Graphiquement, on pourrait envisager de tracer des cercles centrés en 0, d'équation $x^2 + y^2 = \alpha$, de plus en plus grands (on augmente α), jusqu'à toucher C .

\rightsquigarrow Cela se produit à un point où $\{f(x, y) = \alpha\}$ est tangent à $C = \{g(x, y) = 0\}$.



Applications à l'économie

Typiquement, les problèmes d'optimisation qui se présentent en économie sont des problèmes de *maximisation de l'utilité* et de *minimisation des coûts*.

....Et notre cadre pourrait sembler inadapté pour les traiter, car les contraintes sur ces problèmes sont typiquement des inégalités $g(x) \leq 0$, et non des égalités $g(x) = 0$.

On va voir comment, dans certains cas, on peut se ramener à des contraintes d'égalité pour pouvoir utiliser quand même notre beau théorème des extrema liés.

Et, par souci d'honnêteté, on va aussi voir des cas où ça ne fonctionne pas.

Une application à la théorie du consommateur

On suppose qu'une personne consomme deux biens en quantités x_1 et x_2 , qui lui procurent une utilité décrite par la fonction $u(x_1, x_2) = x_1 x_2$.

Son revenu est noté $I > 0$ (on lui souhaite) et les prix des biens sont $p_1, p_2 > 0$ (hélas).

Il s'agit donc de résoudre le problème d'optimisation

$$(\mathcal{U}) \begin{cases} \text{Maximiser } u(x) \\ x \in \mathcal{B} = \{x = (x_1, x_2) \in \mathbb{R}^2, p_1 x_1 + p_2 x_2 \leq I, x_1 \geq 0, x_2 \geq 0\} \end{cases}$$

\rightsquigarrow l'ensemble \mathcal{B} décrit par les contraintes budgétaires est fermé et borné dans \mathbb{R}^2 qui est de dimension finie, donc c'est un compact: la fonction continue u admet donc un maximum global $x^* = (x_1^*, x_2^*)$ sur \mathcal{B} .

...L'ennui, c'est que ces contraintes budgétaires sont des inégalités, et nous nous trouvons donc fort démunis pour déterminer x^* .

Remarquons, cependant, que

- $x_{\text{moit}'\text{-moit}'} = (\frac{I}{2p_1}, \frac{I}{2p_2}) \in \mathcal{B}$ et vérifie $u(x) > 0$. Donc le maximum x^* vérifie $u(x^*) \geq u(x_{\text{moit}'\text{-moit}'}) > 0$.
- or, si $x_1 = 0$ ou $x_2 = 0$, on a $u(x) = 0$, donc le maximum qu'on cherche n'est pas atteint sur les points de \mathcal{B} tels que $x_1 = 0$ ou $x_2 = 0$.
- et d'un autre côté, si $x_1, x_2 > 0$ et $p_1x_1 - p_2x_2 < I$, alors $x \in \mathring{\mathcal{B}}$, donc il existe $r > 0$ tel que $(x_1 + r, x_2) \in \mathcal{B}$. Mais dans ce cas, $u(x_1 + r, x_2) > u(x_1, x_2)$ et x n'est pas un maximum.

\rightsquigarrow Le maximum est forcément atteint en un point de

$$\begin{aligned} \tilde{\mathcal{B}} &= \{x = (x_1, x_2) \in \mathbb{R}^2, p_1x_1 + p_2x_2 = I, x_1 > 0, x_2 > 0\} \\ &= \underbrace{(\mathbb{R}_+^*)^2}_{\text{ouvert}} \cap \{x = (x_1, x_2) \in \mathbb{R}^2, \underbrace{p_1x_1 + p_2x_2}_{=g(x)} = I\} \end{aligned}$$

ce qui nous met confortablement dans le cadre du TEL.

Pour tout $x = (x_1, x_2) \in \tilde{\mathcal{B}}$, u et g sont \mathcal{C}^1 en x et on a

$$\nabla u(x) = \begin{pmatrix} x_2 \\ x_1 \end{pmatrix}, \quad \nabla g(x) = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$$

\rightsquigarrow On s'aperçoit que la condition $\nabla g(x) \neq 0_{\mathbb{R}^2}$ est vérifiée pour tout $x \in \tilde{\mathcal{B}}$. On peut utiliser la méthode de Lagrange: donc on cherche les solutions (x_1, x_2, λ) de

$$\begin{cases} \nabla u(x) = \lambda \nabla g(x) \\ g(x) = 0 \end{cases} \iff \begin{cases} x_2 & = \lambda p_1 \\ x_1 & = \lambda p_2 \\ p_1x_1 + p_2x_2 & = I \end{cases}$$

\rightsquigarrow si $\lambda = 0$, on trouve $x_1 = x_2 = 0$, ce qui contredit la troisième équation. On peut donc supposer que $\lambda \neq 0$, ce qui nous donne

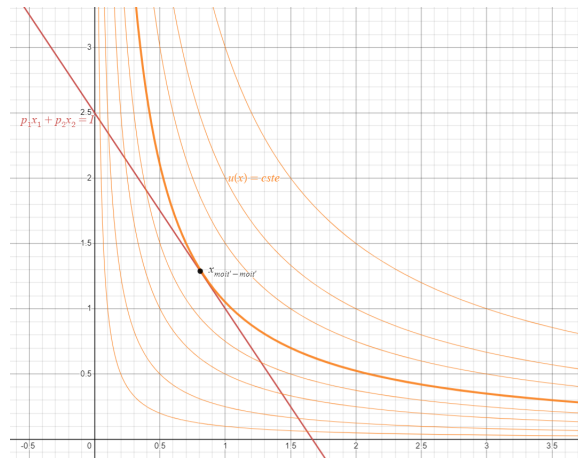
$$\lambda = \frac{x_1}{p_2} = \frac{x_2}{p_1} \Rightarrow x_1 = \frac{p_2x_2}{p_1}$$

et donc avec la troisième équation,

$$p_1 \frac{p_2x_2}{p_1} + p_2x_2 = 2p_2x_2 = I \Rightarrow x_2 = \frac{I}{2p_2}, x_1 = \frac{I}{2p_1}, \lambda = \frac{I}{2p_1p_2}$$

↪ S'il y a un extremum local de u sur $\tilde{\mathcal{B}}$, c'est forcément $x^* = (\frac{I}{2p_1}, \frac{I}{2p_2})$. Or, on sait que u admet un maximum global sur $\tilde{\mathcal{B}}$ d'après les arguments de compacité avancés précédemment: et il n'y a qu'un seul suspect, donc ce maximum est forcément x^* .

Illustration géométrique



↪ On retrouve notre interprétation géométrique: la solution est donnée par le point où une des courbes d'équation $u(x) = cste$ est tangente à la droite de contrainte budgétaire.

Un cas où on ne peut pas faire ça

Considérons un autre consommateur de ces deux mêmes biens, pourvu d'un revenu $I' > 0$, et donc la fonction d'utilité est cette fois $v(x_1, x_2) = x_1 + x_2$. On cherche donc à résoudre

$$(\mathcal{U}) \begin{cases} \text{Maximiser } v(x) \\ x \in \mathcal{B}' = \{x = (x_1, x_2) \in \mathbb{R}^2, p_1x_1 + p_2x_2 \leq I', x_1 \geq 0, x_2 \geq 0\} \end{cases}$$

Mais cette fois, on ne peut pas allègrement remplacer les conditions $x_1 \geq 0, x_2 \geq 0$ par $x_1 > 0, x_2 > 0$. En fait, cette fois, on s'aperçoit que la solution est

$$\begin{cases} \left(\frac{I'}{p_1}, 0\right) & \text{si } p_1 < p_2 \\ \left(0, \frac{I'}{p_2}\right) & \text{si } p_2 < p_1 \end{cases}$$

et si $p_1 = p_2$, n'importe quel point vérifiant $p_1x_1 + p_2x_2 = I'$ est optimal. Autrement dit, dès que les prix sont différents, ce deuxième consommateur

achète le plus possible du bien le moins cher, et pas du tout de l'autre, donc cette fois les cas d'égalités dans $x_1 \geq 0, x_2 \geq 0$ ne peuvent pas être ignorés impunément.

Si on n'avait pas remarqué ça, et si on utilise la méthode de Lagrange avec un enthousiasme excessif, on calcule

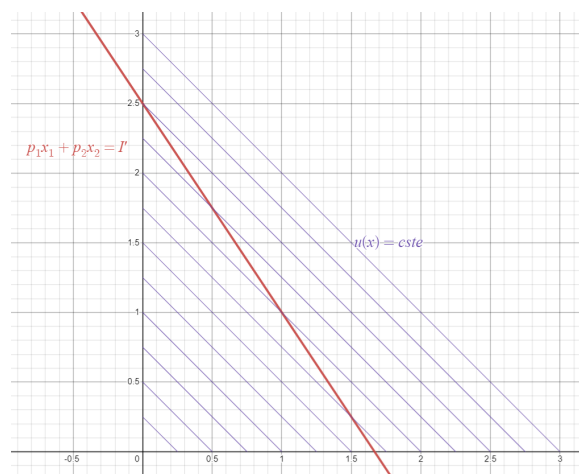
$$\nabla v(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \nabla g(x) = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$$

et on doit résoudre

$$\begin{cases} \nabla v(x) = \lambda \nabla g(x) \\ g(x) = 0 \end{cases} \iff \begin{cases} 1 & = \lambda p_1 \\ 1 & = \lambda p_2 \\ p_1 x_1 + p_2 x_2 & = I \end{cases}$$

qui n'a pas de solution lorsque $p_1 \neq p_2$: la méthode de Lagrange, dans ce cas, ne nous trouve pas le maximum.

Illustration géométrique



\rightsquigarrow Et ça se voit géométriquement: si $p_1 \neq p_2$, aucune des droites d'équation $v(x) = cste$ n'est tangente à la droite de contrainte budgétaire.

Exercice - Théorie du producteur

On s'intéresse maintenant à un problème de minimisation des coûts. Une entreprise utilise deux facteurs x_1, x_2 , de coûts unitaires respectifs $w_1, w_2 >$

0, pour produire un bien y via la fonction de production $y = g(x_1, x_2) = x_1x_2$.

La firme cherche à produire une quantité $\tilde{y} > 0$ au moindre coût possible. On veut donc résoudre

$$(\mathcal{P}) \begin{cases} \text{Minimiser } w_1x_1 + w_2x_2 \\ x \in \mathcal{X} = \{(x_1, x_2) \in \mathbb{R}^2, x_1x_2 = \tilde{y}, x_1 \geq 0, x_2 \geq 0\} \end{cases}$$

1. Justifier que la fonction de coût $c(x) = w_1x_1 + w_2x_2$ admet un minimum global sur

$$\{(x_1, x_2) \in \mathbb{R}^2, x_1x_2 = \tilde{y}, (x_1, x_2) \in [0, \sqrt{\tilde{y}}]^2\}$$

et en déduire que c admet un minimum global sur \mathcal{X} .

Indication: Chercher l'inspiration dans la feuille d'exercices: par exemple, le 4.

2. Montrer qu'il existe un ouvert $\mathcal{U} \subset \mathbb{R}^2$ tel que le problème d'optimisation (\mathcal{P}) soit équivalent à

$$(\mathcal{P}') \begin{cases} \text{Minimiser } w_1x_1 + w_2x_2 \\ x \in \tilde{\mathcal{X}} = \mathcal{U} \cap \{(x_1, x_2) \in \mathbb{R}^2, x_1x_2 = \tilde{y}\} \end{cases}$$

3. Vérifier que, pour tout $x \in \tilde{\mathcal{X}}$, $\nabla g(x) \neq 0_{\mathbb{R}^2}$, ce qui nous met dans le cadre du TEL.
4. A grands coups de méthode de Lagrange, montrer qu'il y a un seul suspect pour être un extremum local de c sur $\tilde{\mathcal{X}}$,

$$x^* = \left(\sqrt{\frac{w_2\tilde{y}}{w_1}}, \sqrt{\frac{w_1\tilde{y}}{w_2}} \right)$$

5. Justifier que x^* est le minimum qu'on cherche.
6. Montrer que si la fonction de production est donnée par $\tilde{g}(x) = x_1 + x_2$, alors c admet (au moins) un minimum global sur \mathcal{X} , et pourtant la méthode de Lagrange ne nous permet pas de le trouver.

Justifier abondamment.

Complément - Théorème des extrema liés convexe

Comme dans le cas de l'optimisation sans contrainte, pour les fonctions convexes, les conditions nécessaires sont aussi suffisantes. Ainsi, pour rappel:

CNS convexe: Soit $U \subset \mathbb{R}^n$ un ouvert convexe. Soit $f : U \rightarrow \mathbb{R}$ une fonction convexe (resp. concave). Soit $a \in U$ tel que f est différentiable en a .

Alors a est un minimum (resp. maximum) global de f ssi $Df(a) = 0$.

Et quand il y a des contraintes, on a:

Théorème 4

Soit $U \subset \mathbb{R}^n$ un ouvert convexe, $f : U \rightarrow \mathbb{R}$ et $g : U \rightarrow \mathbb{R}^p$. On note $\Gamma = g^{-1}(\{0\})$. Soient de plus $a \in \Gamma$ et $\lambda_1, \dots, \lambda_p$ des réels tels que

1. f, g sont différentiables en a ;
2. $f + \sum_i \lambda_i g_i$ est convexe;
3. $Df(a) + \sum_i \lambda_i Dg_i(a) = 0$

Alors $f|_{\Gamma}$ a un minimum global en a .

Remarques:

- On n'a pas besoin de supposer f ou $g \in \mathcal{C}^1$, car la preuve ne repose pas sur le TFI, mais sur l'application de la CNS à $h = f + \sum_i \lambda_i g_i$.
- La condition 2. est vérifiée par exemple si f et les g_i sont convexes et $\lambda_i \geq 0$ pour tout i .

Exemple:

Soit $(a, b) \neq (0, 0)$. Cherchons les extrema de $f(x, y) = ax + by$ sur $\Gamma = \{(x, y), x^2 + y^2 = 1\} = g^{-1}(\{0\})$, où $g(x, y) = x^2 + y^2 - 1$. Pour $\lambda \in \mathbb{R}$, on a

$$Df(x, y) + \lambda Dg(x, y) = 0 \iff \begin{cases} a + 2\lambda x = 0 \\ b + 2\lambda y = 0 \end{cases}$$

ce qui donne, pour $(x, y) \in \Gamma$, $4\lambda^2(x^2 + y^2) = 4\lambda^2 = a^2 + b^2 \neq 0$. Donc $\lambda = \pm \frac{\sqrt{a^2 + b^2}}{2}$.

- Pour $\lambda = \frac{\sqrt{a^2 + b^2}}{2}$, $f + \lambda g$ est convexe et $f|_{\Gamma}$ a un minimum global en

$$u_1 = \left(\frac{a}{\sqrt{a^2 + b^2}}, \frac{b}{\sqrt{a^2 + b^2}} \right)$$

- Pour $\lambda = -\frac{\sqrt{a^2+b^2}}{2}$ $f + \lambda g$ est concave, donc $-(f + \lambda g)$ est convexe, et admet un minimum global en

$$u_2 = \left(-\frac{a}{\sqrt{a^2+b^2}}, -\frac{b}{\sqrt{a^2+b^2}} \right).$$

On en déduit que $f|_{\Gamma}$ admet un maximum global en u_2 .

Exercice - Inégalité de Hölder

Soit $p > 1$ et $(b_1, \dots, b_n) \in (\mathbb{R}_+^*)^n$. On note $q = \frac{p}{p-1}$ le réel tel que $\frac{1}{p} + \frac{1}{q} = 1$.
On cherche le maximum de la fonction

$$f : (x_i)_{1 \leq i \leq n} \mapsto \sum_{i=1}^n b_i x_i \text{ sur } \mathcal{K} = \left\{ (x_i)_{1 \leq i \leq n} \in (\mathbb{R}_+^*)^n; \underbrace{\sum_{i=1}^n x_i^p - 1}_{:=g(x)} = 0 \right\}.$$

Les fonctions f et g sont \mathcal{C}^1 sur l'ouvert $\mathcal{U} = (\mathbb{R}_+^*)^n$.

⚠ p n'est pas forcément un entier, et donc g n'est pas, a priori, un polynôme.

1. Calculer, pour tout $x \in \mathcal{U}$, $\nabla f(x)$ et $\nabla g(x)$, et montrer que $\nabla g(x) \neq 0_{\mathbb{R}^n}$.
2. Montrer que si $f|_{\mathcal{K}}$ a un extremum en $a = (a_1, \dots, a_n) \in \mathcal{K}$, alors

$$\forall j = 1, \dots, n, a_j = \left(\frac{b_j^q}{\sum_{i=1}^n b_i^q} \right)^{\frac{1}{p}} \text{ et } f(a) = \left(\sum_{i=1}^n b_i^q \right)^{\frac{1}{q}}$$

3. Montrer que \mathcal{K} est convexe et que, pour tout $\lambda > 0$, la fonction $f - \lambda g$ est concave.
4. En déduire que le point qu'on a trouvé est bien un maximum de $f|_{\mathcal{K}}$
5. En déduire que, pour tout $(a_i)_{1 \leq i \leq n} \in (\mathbb{R}_+^*)^n$ et $(b_i)_{1 \leq i \leq n} \in (\mathbb{R}_+^*)^n$,

$$\sum_{i=1}^n x_i b_i \leq \left(\sum_{i=1}^n x_i^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^n b_i^q \right)^{\frac{1}{q}}.$$