Notes du cours 2

17 septembre 2024

Table des matières

1 Introduction

Ce document se compose de deux parties. La première partie est une introduction à l'optimisation, une explication des conditions de Karush-Kuhn-Tucker (KKT). La seconde partie est une introduction à la convexité.

2 Optimisation avec contraintes

Définition 1. ***Problème d'optimisation sous contraintes***
On considère le problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ respectant } \begin{cases} c_i(x) = 0, i \in \mathcal{E} \\ c_j(x) \leq 0, j \in \mathcal{I} \end{cases}$$

avec :
- $f : \mathbb{R}^n \to \mathbb{R}$ *est la fonction objectif.* $f \in \mathcal{C}^1$
- $\mathcal{E}, \mathcal{I}$ *sont des ensembles finis d'indexes de contraintes.*
- $c : \mathbb{R}^n \to \mathbb{R}$ *sont des contraintes.* $c \in \mathcal{C}^1$

On définit le domaine des contraintes :

$$\Omega = \{x \in \mathbb{R}^n | \forall i \in \mathcal{E}, j \in \mathcal{I}, c_i(x) = 0, c_j(x) \leq 0\}$$

Dans le contexte des contraintes d'inégalité $c_j(x) \leq 0$, une contrainte est dite **active** (ou saturée) en un point $x \in \Omega$ si elle est exactement satisfaite à la frontière, c'est-à-dire si $c_j(x) = 0$. Cela signifie que le point x se situe sur la frontière définie par la contrainte.

Une contrainte est dite **inactive** en x si $c_j(x) < 0$, ce qui implique que le point x se trouve strictement à l'intérieur du domaine admissible défini par cette contrainte, la frontière étant exclue.

Exemple 1. *Nous nous intéressons à la fonction suivante :*

$$f(x, y) = x^3 y + 3y + 2$$

- ***Contrainte d'égalité*** *: Prenons la contrainte $x^2 + y^2 = 1$, qui définit le cercle unité. La contrainte s'écrit $c(x, y) = x^2 + y^2 - 1 = 0$.*
- ***Contrainte d'inégalité*** *: Prenons maintenant la contrainte $x^2 + y^2 \leq 1$, qui définit le disque unité (incluant sa frontière). La contrainte s'écrit $c(x, y) = x^2 + y^2 - 1 \leq 0$.*

  Dans ce cas, la frontière du disque (où $x^2 + y^2 = 1$) correspond aux points où la contrainte est saturée ($c(x, y) = 0$) et donc active. Les points à l'intérieur du disque (où $x^2 + y^2 < 1$) sont ceux pour lesquels la contrainte est inactive ($c(x, y) < 0$).

Théorème 1. *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction à minimiser et Ω le domaine des contraintes. Un point $x^* \in \mathbb{R}^n$ est une solution locale du problème si et seulement si $x^* \in \Omega$ et il existe un ouvert $U \in \mathbb{R}^n$ contenant x^* tel que pour tout $x \in U \cap \Omega$, $f(x^*) \leq f(x)$.*

Démonstration. Par définition d'un minimum local sur l'ensemble Ω, il suffit de considérer un voisinage ouvert U de x^* tel que, pour tout $x \in U \cap \Omega$, on ait $f(x^*) \leq f(x)$. Cela signifie qu'il n'existe aucun point $x \in U \cap \Omega$ pour lequel $f(x)$ soit inférieur ou égal à $f(x^*)$, ce qui prouve que x^* est bien un minimum local sur l'ensemble Ω. □

L'utilisation, dans ce théorème, d'un ouvert non forcément contenu dans Ω est nécessaire pour garantir l'existence de voisinage de x^* dans lequel la fonction objectif est bien définie. Il se peut que Ω soit un ensemble fermé et que x^* soit sur la frontière, dans ce cas, il n'y a pas d'ensemble ouvert Ω.

Remarque 1. *Gradient de fonction et équation de surface.*
Par la suite, nous prenons $f : \mathbb{R}^n \to \mathbb{R}$ une fonction quelconque de classe $\mathcal{C}^1$.

- *Le gradient $\nabla f(x)$ indique la direction de la plus grande pente de f au point x, en considérant f comme une fonction réelle de plusieurs variables.*
- *Lorsque l'on considère la surface définie par l'équation $f(x) = 0$, le gradient $\nabla f(x)$ est orthogonal (normal) à la surface de f. Il est donc important de faire cette distinction entre la fonction f ayant un graphe dans $\mathbb{R}^{n+1}$ et la surface de $\{x \in \mathbb{R}^n, f(x) = 0\}$ ayant un graphe dans $\mathbb{R}^n$. De plus, la signification de $\nabla f \in \mathbb{R}^n$ est différente entre les deux cas*

Théorème 2. ***Caractérisation de la normale à une surface***
Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction quelconque de classe $\mathcal{C}^1$.
Soit $\Omega = \{x \in \mathbb{R}^n | f(x) = 0\}$.
Pour tout $x \in \Omega$, $\nabla f(x)$ est normal à la surface Ω en x.

Démonstration. Soit $\mu : \mathbb{R} \to \mathbb{R}^n$ tel que $\forall t \in \mathbb{R}, \mu(t) \in \Omega$, une courbe de surface de Ω. On a donc $x \in \Omega$ tel que $\mu(0) = x$. On veut ici montrer que le vecteur de la droite tangente à μ en x est orthogonal à $\nabla f(x)$.

Ainsi, on calcule :

$$\nabla f(x).\nabla\mu(0) = \nabla f(\mu(0)).\nabla\mu(0) = \frac{d}{dt} f(\mu(t))|_{t=0} = 0$$

On a que $\frac{d}{dt} f(\mu(t))|_{t=0} = 0$ car $\mu(t) \in \Omega$ et $\forall x \in \Omega, f(x) = 0$

On vient de montrer que $\nabla f(x).\nabla\mu(0)$ sont orthogonaux. Or cette démonstration est valable pour toute courbe de surface de Ω. Ainsi, $\nabla f(x)$ est la normale à la surface $f(x) = 0$.

□

Exemple 2. *Soit $f(x_1, x_2) = x_1 + x_2$ la fonction à minimiser. Son gradient est $\nabla f(x) = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.*

Nous nous intéressons à la contrainte $c(x) = x_1^2 + x_2^2 - 2$ qui définit le cercle ou le disque de rayon $\sqrt{2}$.

- ***Contrainte d'égalité** : $c(x) = 0$. On a $\nabla c(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = 2x$ La condition d'optimalité devient :*

$$\exists \lambda, \nabla f(x) = \lambda \nabla c(x)$$

  Cette formule signifie que le gradient de f est colinéaire au gradient de c, ce dernier étant perpendiculaire à la frontière du cercle. Cette colinéarité est nécessaire afin de trouver un optimal, tout en respectant la contrainte.
- ***Contrainte d'inégalité** : $c(x) \leq 0$. Cette fois-ci, la condition nécessaire d'optimalité pour un point x^* est :*

$$\exists \lambda_1^*, \begin{cases} \nabla f(x^*) = -\lambda_1^* \nabla c_1(x^*) \\ \lambda_1^* \geq 0 \\ \lambda_1^* c_1(x^*) = 0 \end{cases}$$

Remarque 2. *Reprenons les conditions d'optimalités ci-dessus :*

$$\exists \lambda_1^*, \begin{cases} \nabla f(x^*) = -\lambda_1^* \nabla c_1(x^*) \\ \lambda_1^* \geq 0 \\ \lambda_1^* c_1(x^*) = 0 \end{cases}$$

Essayons de donner du sens à ces dernières.

Voici une première façon de voir cela :

On peut distinguer deux cas que classe $\lambda c_1(x^*) = 0$, impliquant soit $\lambda = 0$ soit $c_1(x^*) = 0$. Entrons dans les détails :

1. ***La contrainte est active*** *(*$c_1(x^*) = 0, \lambda \neq 0$*) :*

   Dans ce cas, le point x^* se situe sur la frontière de c_1.

   Nous souhaitons ici que x^* se situe sur le bout de la surface de c par rapport à l'augmentation de f. Cela se traduit par le fait que $\nabla f(x^*)$ et $\nabla c_1(x^*)$ soient colinéaires. Mathématiquement, cela se traduit par le fait :

   $$\exists \lambda \text{ tel que } \nabla f(x^*) = \lambda \nabla c_1(x^*)$$

   En gardant un λ qui puisse prendre des valeurs positives, on inclut les bouts de la surface de c_1 qui maximisent localement f. Or, dans notre cas, on désire uniquement minimiser la fonction f. La contrainte devient alors :

   $$\exists \lambda \geq 0 \text{ tel que } \nabla f(x^*) = -\lambda \nabla c_1(x^*)$$

   Cela indique que le gradient de f au point x^* pointe dans la direction opposée au gradient de c_1.

2. ***La contrainte est inactive*** *(*$\lambda = 0, c_1(x^*) < 0$*) :*

   Ici, le point x^* est alors en dehors de la frontière de c_1.

   Dans ce cas-ci, la seule chose qui importe, c'est que le minimum ait un gradient nul. Pour ce faire, on reprend la formule précédente, mais on pose $\lambda = 0$, forcé par la troisième condition, ce qui donne :

   $$\nabla f(x^*) = -0 \times \nabla c_1(x^*) = 0$$

Ainsi, la première et troisième condition encode la disjonction de cas d'une façon élégante. La deuxième impose le fait qu'il faut prendre les extremums qui minimisent sans ceux qui la maximisent.

Nous pouvons généraliser ce résultat pour un ensemble de contraintes.

Théorème 3. ***Conditions de Karush-Kuhn-Tucker (KKT)***
Soit $f : \mathbb{R}^n \to \mathbb{R}$ la fonction à minimiser et c_i des contraintes.
On pose la fonction lagrangienne :

$$\mathcal{L}(x, \lambda) = f(x) + \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

Si x^* est un minimum global, alors, il existe les multiplicateurs de Lagrange λ_i^* tel que :

$$\begin{cases} \nabla_x \mathcal{L}(x^*, \lambda^*) = 0 \\ c_i(x^*) = 0, \forall i \in \mathcal{E} \\ c_j(x^*) \leq 0, \forall j \in \mathcal{I} \\ \lambda_i^* \geq 0, \forall i \in \mathcal{I} \\ \lambda_i^* c_i(x^*) = 0, \forall i \in \mathcal{E} \cup \mathcal{I} \end{cases}$$

Démonstration. Trivial. Le trouver est plus dur... □

Définition 2. ***Qualification des contraintes de Mangasarian-Fromovitz (MFCQ)***

Les contraintes satisfont la condition MFCQ si les gradients des contraintes d'égalité sont linéairement indépendants en x^*, et s'il existe un vecteur $d \in \mathbb{R}^n$ tel que $\nabla c_i(x^*)^T d < 0$ pour toutes les contraintes inégalitaires actives et $\nabla c_j(x^*)^T d = 0$ pour toutes les contraintes d'égalité.

Théorème 4. *Si x^* est une solution locale et que la qualification des contraintes de Mangasarian-Fromovitz est vérifiée en x^*, alors les conditions de KKT sont vérifiées en x^*.*

Démonstration. On considère le problème d'optimisation suivant :

$$\min_{x \in \mathbb{R}^n} f(x) \text{ sous les contraintes } \begin{cases} h_i(x) = 0, i \in \mathcal{E} \\ g_j(x) \leq 0, j \in \mathcal{I} \end{cases}$$

avec :
- $f : \mathbb{R}^n \to \mathbb{R}$ est la fonction objectif, $f \in \mathcal{C}^1$.
- $\mathcal{E}, \mathcal{I}$ sont des ensembles finis d'index de contraintes.
- h et $g : \mathbb{R}^n \to \mathbb{R}$ sont respectivement des contraintes d'égalités et d'inégalités, avec $h, g \in \mathcal{C}^1$.

Montrons que si x^* est une solution locale et que la condition MFCQ est vérifiée en x^*, alors les conditions de KKT sont également vérifiées en x^*.

Puisque la condition MFCQ est satisfaite en x^* :

1. Les gradients des contraintes d'égalité h_j sont linéairement indépendants, ce qui implique l'existence des multiplicateurs de Lagrange λ_j associés à ces contraintes.
2. Il existe un vecteur $d \in \mathbb{R}^n$ tel que :
   - $\nabla_x h_j(x^*) d = 0$ pour toutes les contraintes d'égalité j.
   - $\nabla_x g_i(x^*) d < 0$ pour toutes les contraintes d'inégalité actives, c'est-à-dire les contraintes pour lesquelles $g_i(x^*) = 0$.

   Cela garantit qu'il existe une direction de descente faisable.

Pour démontrer la stationnarité, on utilise le lemme de Farkas.

Lemme 1. *Soit $A \in \mathbb{R}^{m \times n}$, l'un des deux systèmes suivants a une solution, mais pas les deux :*

1. *$\exists x \in \mathbb{R}^n$, tel que $Ax \leq b$*
2. *$\exists y \in \mathbb{R}^m$, tel que $y^T A \geq 0$ et $y^T b < 0$.*

Par le lemme de Farkas et la condition MFCQ, il existe des multiplicateurs $\mu_i \geq 0$ et λ_j tels que la stationnarité est vérifiée :

$$\nabla_x \mathcal{L}(x^*, \lambda, \mu) = \nabla_x f(x^*) + \sum_i \mu_i \nabla g_i(x^*) + \sum_j \lambda_j \nabla h_j(x^*) = 0.$$

Enfin, la condition de complémentarité $\mu_i g_i(x^*) = 0$ est satisfaite en raison de la construction des multiplicateurs μ_i associés aux contraintes actives. De plus, x^* respecte les contraintes, c'est-à-dire $g_i(x^*) \leq 0$ et $h_j(x^*) = 0$.

Ainsi, sous l'hypothèse que MFCQ est vérifiée en x^*, les conditions de KKT sont également satisfaites en x^*.

$\square$

3 Convexité

Définition 3. ***Ensemble convexe***
*X est un ensemble dit **convexe** si et seulement si :*

$$\forall x, y \in X, [x, y] = \{x + \lambda(y - x) | \lambda \in [0, 1]\} \in X$$

Définition 4. ***Fonction convexe***
*Soit $f : X \subset \mathbb{R}^n \to \mathbb{R}$ une fonction f est dite **convexe** si et seulement si :*

$$\forall x, y \in X, \forall t \in [0, 1], f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)$$

*Si l'inégalité est stricte dans $t \in]0, 1[$, on parle de fonction **strictement convexe**.*

Définition 5. ***Fonction μ-fortement convexe****.*
*Une fonction $f : \mathbb{R}^n \to \mathbb{R}$ est dite **μ-fortement convexe** si, pour $\forall \mu > 0, \forall x, y \in \mathbb{R}^n$, on a :*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2} \|y - x\|^2.$$

Cela signifie que la fonction f est convexe avec une courbure minimale μ.

Théorème 5. *Soit $f : X \subset \mathbb{R}^n \to \mathbb{R}$ une fonction de classe $\mathcal{C}^2$. Alors :*
- *f est convexe si et seulement si sa Hessienne est semi-définie positive sur l'intérieur de X.*
- *f est strictement convexe si et seulement si sa Hessienne est définie positive sur l'intérieur de X.*

Démonstration. La preuve se fait en montrant que la dérivée seconde de f sur une corde donne la définition d'une matrice définie semi-positive.
Posons :

$$\phi(t) = f(x + t(y - x))$$

$$\frac{d}{dt}\phi(t) = \nabla f(x + t(y - x))^t (y - x)$$

$$\frac{d^2}{dt^2}\phi(t) = (\nabla^2 f(x+t(y-x))(y-x))^t)^t(y-x) = (y-x)^t \nabla^2 f(x+t(y-x))(y-x)$$

Ainsi, si la Hessienne est semi-définie positive, alors la dérivée seconde est positive pour tout t. De ce fait, la fonction est convexe. A l'inverse, si la fonction est convexe, alors la dérivée seconde d'une corde est positive pour tout t. Ainsi, la Hessienne est semi-définie positive.

Pour la stricte convexité, il suffit de remplacer les inégalités par des inégalités strictes. □

Propriétés 1. *Les affirmations suivantes sont vraies :*
1. *Une fonction linéaire est convexe.*

2. *Une combinaison linéaire positive de fonction convexe est convexe.*
3. *Le maximum de fonctions convexes est convexe.*

Démonstration. 1. Soit $f(x) = a^T x$ une fonction linéaire. On a :

$$f(tx+(1-t)y) = a^T(tx+(1-t)y) = ta^T x+(1-t)a^T y = tf(x)+(1-t)f(y)$$

La fonction est donc convexe.

2. *Soit*$\{f_1, f_2, ..., f_n\}$ des fonctions convexes et $\{\lambda_1, \lambda_2, ..., \lambda_n\}$ des réels positifs. On a $\forall f_i, \lambda_i$:

$$\lambda_i f_i(tx + (1-t)y) \leq \lambda_i(tf(x) + (1-t)f(y))$$

Alors :

$$\lambda_i f_i(tx + (1-t)y) \leq t\lambda_i f(x) + (1-t)\lambda_i f(y)$$

De plus, en sommant les inégalités, on obtient :

$$\sum_{i=1}^{n} \lambda_i f_i(tx + (1-t)y) \leq t \sum_{i=1}^{n} \lambda_i f(x) + (1-t) \sum_{i=1}^{n} \lambda_i f(y)$$

La combinaison linéaire positive de fonction convexe est donc convexe.

3. Soit f_1, f_2 deux fonctions convexes. On a :

$$g(x) = \max(f_1(x), f_2(x))$$

Pour tout $x, y \in X$ et $t \in [0, 1]$, on a :

$$g(tx + (1-t)y) = \max(f_1(tx + (1-t)y), f_2(tx + (1-t)y))$$

En utilisant la propriété de convexité de f_1 et f_2, on obtient :

$$g(tx + (1-t)y) \leq \max(tf_1(x) + (1-t)f_1(y), tf_2(x) + (1-t)f_2(y))$$

Puisque le maximum de deux valeurs est toujours inférieur ou égal à la somme des maxima, on a :

$$g(tx + (1-t)y) \leq t \max(f_1(x), f_2(x)) + (1-t) \max(f_1(y), f_2(y))$$

Ainsi, on conclut que g est convexe.

□

Théorème 6. *Soit $f : \mathbb{R}^n \to \mathbb{R}$. f est strictement convexe sur un ensemble convexe $X \subseteq \mathbb{R}^n$ alors f admet au plus un minimum global sur X.*

Démonstration. Supposons par l'absurde qu'il existe deux minimums globaux en x_1 et x_2.
Alors par convexité, pour tout y dans le voisinage ouvert de x_1 ou x_2, on a $f(y) < f(x_1)$ et $f(y) < f(x_2)$.
Observons la fonction de la corde entre x_1 et x_2, qui est convexe.
Cette dernière part de x_1, puis croît en valeur en passant par un y du voisinage de x_1 et finit en x_2, qui est en dessous.
Ainsi, la fonction croît puis décroit. Ce qui est absurde par stricte convexité. □

Théorème 7. *Soit $f : X \subseteq \mathbb{R}^n \to \mathbb{R}$ une fonction convexe sur un ensemble convexe $X \subseteq \mathbb{R}^n$. Alors, tout minimum local est un minimum global. Si X est ouvert, alors $\nabla f(x^*) = 0$ et équivalent à dire que x^* est un minimum global.*

Démonstration. Soient x_1 et x_2 deux minimums locaux.
Par convexité, pour tout y dans le voisinage ouvert de x_1 ou x_2, on a $f(y) \geq f(x_1)$ et $f(y) \geq f(x_2)$.
Observons la fonction de la corde entre x_1 et x_2, qui est convexe.

— Soit $f(x_1) = f(x_2)$, dans ce cas, la fonction est constante et x_1 et x_2 sont des maximums globaux.

— Soit $f(x_1) > f(x_2)$, alors il existe t tel que $y = tx_1 + (1-t)x_2$.
On a alors :

$$f(y) = f(tx_1 + (1-t)x_2) = f(x_1) > tf(x_1) + (1-t)f(x_2)$$

car $f(x_1) > f(x_2)$.
Dans ce cas, la fonction n'est pas convexe sur X.

Ainsi, si f est convexe, alors tout minimum local est un minimum global.

Si X est ouvert, alors $\nabla f(x^*) = 0$ et équivalent à dire que x^* est un minimum global.
X étant ouvert, on peut donc calculer la dérivée de f en tout point de X.
Si pour $x^* \in \mathbb{R}^n$, $\nabla f(x^*) = 0$, alors x^* est un minimum local.
Or par le précédent résultat, tout minimum local est un minimum global.
Ainsi, x^* est un minimum global. □