

Analyse III

Fonctions de plusieurs variables

MAT2019L séquence 4

CM7

automne 2017

cours de
Francis Clarke

1

utilisation de la matrice hessienne pour écrire le théorème de Lagrange

3

**rappel: Partiel le 15 novembre
(en amphi)**

venez nombreux !

2

$$f(a+h, b+k) = f(a, b) + f_x(a, b)h + f_y(a, b)k + \frac{1}{2} \{ f_{xx}h^2 + 2f_{xy}hk + f_{yy}k^2 \},$$

où les dérivées d'ordre deux sont évaluées en $(a+th, b+tk)$ pour un certain $t \in]0, 1[$.

Il est utile d'adopter une notation vectorielle/matricielle pour écrire la conclusion:

$$f(a+h, b+k) = f(a, b) + \nabla f(a, b) \cdot \begin{bmatrix} h \\ k \end{bmatrix} + \frac{1}{2} \nabla^2 f(a+th, b+tk) \begin{bmatrix} h \\ k \end{bmatrix} \cdot \begin{bmatrix} h \\ k \end{bmatrix}.$$

notation matricielle

Ici, $\nabla^2 f$ est la **matrice hessienne** de la fonction f :

$$\nabla^2 f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$$

Il s'agit d'une matrice 2×2 , symétrique lorsque f est C^2 (par Clairaut).

Rq: un point dans $\mathbb{R}^2$ est une colonne

4

Soit

Exemple

$$f(x, y) = 1 + x - 2y - x^2 + 8xy + 3y^2 + 5x^2y.$$

On calcule

$$\nabla f = \begin{bmatrix} 1 - 2x + 8y + 10xy \\ -2 + 8x + 6y + 5x^2 \end{bmatrix}$$

$$\nabla^2 f = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix} = \begin{bmatrix} -2 + 10y & 8 + 10x \\ 8 + 10x & 6 \end{bmatrix}$$

5

Soit

Exemple

$$f(x, y) = 1 + x - 2y - x^2 + 8xy + 3y^2 + o(x^2 + y^2)$$

le DL en $(0, 0)$ d'ordre deux d'une fonction f .

On écrit ce DL avec le gradient et la hessienne :

$$\begin{aligned} f(x, y) &= 1 + x - 2y - x^2 + 8xy + 3y^2 + o(x^2 + y^2) \\ &= 1 + \begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} -1 & 4 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + o(x^2 + y^2) \\ &= 1 + \begin{bmatrix} 1 \\ -2 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -2 & 8 \\ 8 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + o(x^2 + y^2) \\ &= f(0, 0) + \nabla f(0, 0) \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \frac{1}{2} \nabla^2 f(0, 0) \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + o(x^2 + y^2) \\ &= f(0, 0) + \left\langle \nabla f(0, 0), \begin{bmatrix} x \\ y \end{bmatrix} \right\rangle + \frac{1}{2} \left\langle \nabla^2 f(0, 0) \begin{bmatrix} x \\ y \end{bmatrix}, \begin{bmatrix} x \\ y \end{bmatrix} \right\rangle + o(x^2 + y^2) \end{aligned}$$

7

calcul de la hessienne : un exemple avec $n = 3$

Soit

$$f(x, y, z) = 3x + z \sin y + xz - xyz.$$

On calcule

$$\nabla f = \begin{bmatrix} 3 + z - yz \\ z \cos y - xz \\ \sin y + x - xy \end{bmatrix}$$

$$\nabla^2 f = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix} = \begin{bmatrix} 0 & -z & 1 - y \\ -z & -z \sin y & \cos y - x \\ 1 - y & \cos y - x & 0 \end{bmatrix}$$

6

Le cas général $n > 1$:

Théorème Soit $f : U \rightarrow \mathbb{R}$ une fonction deux fois continûment différentiable sur un ouvert convexe $U \subset \mathbb{R}^n$, et soient x et y deux points distincts dans U . Alors il existe z dans le *segment* ouvert entre x et y (donc z , qui est dans U , est de la forme $x + t(y - x)$ pour $t \in]0, 1[$) tel que

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2} \nabla^2 f(z) (y - x) \cdot (y - x).$$

Ici, $\nabla^2 f(x)$ est la matrice hessienne de la fonction f , la matrice $n \times n$ symétrique définie par

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] \quad (i, j = 1, 2, \dots, n).$$

8

Application à l'optimisation :
étude locale des points critiques

9

Conditions nécessaires ou suffisantes pour un minimum local

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement C^2 autour du point x_* , celui-ci étant un point critique.

Théorème

Conditions nécessaires: Si f atteint un minimum local en x_* , alors on a

$$\nabla^2 f(x_*) \geq 0 \text{ (sdp)}.$$

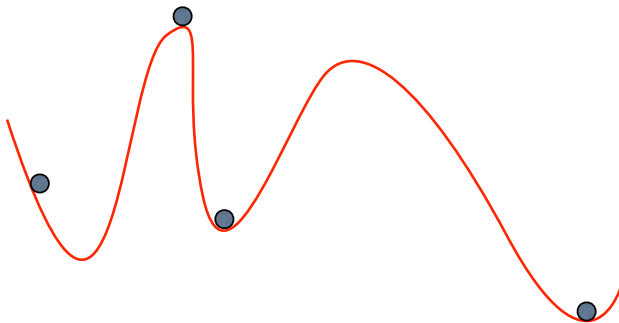
Conditions suffisantes: Réciproquement, si

$$\nabla^2 f(x_*) > 0 \text{ (dp)},$$

alors f atteint un minimum local en x_* .

11

Les min/max locaux sont d'intérêt dans certains contextes, par exemple en mécanique, la théorie des jeux, les équilibres économiques



cinq points critiques

10

On dit que M est *semi-définie positive* (**sdp**) lorsque $\langle Mx, x \rangle \geq 0 \quad \forall x \in \mathbb{R}^n$.

On écrit $M \geq 0$ dans ce cas.

Une matrice M symétrique $n \times n$ est dite *définie positive* (**dp**) si

$$\langle Mx, x \rangle > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

(La forme quadratique induite par M est strictement positive.) On écrit $M > 0$.

(pour $n=1$, c'est des inégalités large ou stricte comme d'habitude)

12

Conditions nécessaires ou suffisantes pour un minimum local

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement C^2 autour du point x_* , celui-ci étant un point critique.

Théorème

Conditions nécessaires: Si f atteint un minimum local en x_* , alors on a

$$\nabla^2 f(x_*) \geq 0 \text{ (sdp)}.$$

Conditions suffisantes: Réciproquement, si

$$\nabla^2 f(x_*) > 0 \text{ (dp)},$$

alors f atteint un minimum local en x_* .

Démonstration :

Rappel sur les matrices symétriques

Une matrice M symétrique $n \times n$ est dite *définie positive* (**dp**) si

$$\langle Mx, x \rangle > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

(La forme quadratique induite par M est strictement positive.) On écrit $M > 0$.

Il découle de la diagonalisation que $M > 0$ ssi les valeurs propres de M sont toutes strictement positives.

L'idée: un changement de variables $x \rightarrow y$ fait que la forme quadratique ci-dessus devient

$$\lambda_1 y_1^2 + \lambda_2 y_2^2 \cdots + \lambda_n y_n^2,$$

où les λ_i sont les valeurs propres.

13

15

$$f(y) - f(x_*) = \cancel{\nabla f(x_*) \cdot (y - x_*)} + \frac{1}{2} \nabla^2 f(z)(y - x_*) \cdot (y - x_*)$$

1. On suppose que f est minimisée en x_* par rapport à la boule $B(x_*, r)$. Posons $y := x_* + tv$, où v est un vecteur unitaire quelconque et $t \in]0, r[$. Alors

$$\begin{aligned} f(y) - f(x_*) &= \frac{1}{2} t^2 \nabla^2 f(z) v \cdot v \geq 0 \\ \implies \nabla^2 f(z) v \cdot v &\geq 0. \end{aligned}$$

On laisse tendre t vers 0, obtenant $\nabla^2 f(x_*) v \cdot v \geq 0$.

CN Puisque v est un vecteur unitaire quelconque, on en déduit $\nabla^2 f(x_*) \geq 0$.

2. On suppose que $\nabla^2 f(x_*) > 0$. Alors il existe une boule $B(x_*, r)$ telle que

$$\nabla^2 f(z) > 0 \quad \forall z \in B(x_*, r)$$

(exercice). Ainsi, pour $y \in B(x_*, r)$, on a

$$\begin{aligned} f(y) - f(x_*) &= \frac{1}{2} \nabla^2 f(z)(y - x_*) \cdot (y - x_*) \\ &\geq 0. \end{aligned}$$

Il suit que x_* minimise f par rapport à la boule $B(x_*, r)$.

CS 

14

On dit que M est *semi-définie positive* (**sdp**) lorsque $\langle Mx, x \rangle \geq 0 \quad \forall x \in \mathbb{R}^n$, et on écrit $M \geq 0$ dans ce cas. Il suit que $M \geq 0$ ssi les valeurs propres de M sont toutes positives (au sens large).

De même $M < 0$ et $M \leq 0$.

Mais **attention**: il est faux que M doit satisfaire soit $M > 0$, soit $M \leq 0$.

16

La matrice **Exemple**

$$\begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix}$$

est définie positive (même si les entrées ne sont pas toutes positives!)

Preuve:

$$\begin{aligned} \begin{bmatrix} 1 & -2 \\ -2 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} x - 2y \\ -2x + 5y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \\ &= x^2 - 2xy - 2xy + 5y^2 \\ &= x^2 - 4xy + 5y^2 \\ &= x^2 - 4xy + 4y^2 + y^2 \\ &= (x - 2y)^2 + y^2 \end{aligned}$$

ce qui est > 0 si $(x, y) \neq (0, 0)$.

17

La matrice **Exemple**

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

n'est pas définie positive (même si les entrées sont toutes positives!)

Preuve:

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} x + 2y \\ 2x + y \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \\ &= x^2 + 2xy + 2xy + y^2 \\ &= x^2 + 4xy + y^2 \end{aligned}$$

ce qui est < 0 si $(x, y) = (1, -1)$.

La matrice est indéfinie ; ses valeurs propres sont -1 et 3

18

Conditions nécessaires ou suffisantes pour un minimum local

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement C^2 autour du point x_* , celui-ci étant un point critique.

Théorème

Conditions nécessaires: Si f atteint un minimum local en x_* , alors on a

$$\nabla^2 f(x_*) \geq 0 \text{ (sdp)}.$$

Conditions suffisantes: Réciproquement, si

$$\nabla^2 f(x_*) > 0 \text{ (dp)},$$

alors f atteint un minimum local en x_* .

19

De façon plus générale:

Conditions nécessaires ou suffisantes pour un extrémum local

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement C^2 autour du point x_* , celui-ci étant un point critique.

Théorème

Conditions nécessaires: Si f atteint un minimum [maximum] local en x_* , alors on a $\nabla^2 f(x_*) \geq 0$ (sdp) [≤ 0 (sdn)].

Conditions suffisantes: Réciproquement, si $\nabla^2 f(x_*) > 0$ (dp) [< 0 (dn)], alors f atteint un minimum [maximum] local en x_* .

20

Classification des points critiques (suite)

Le critère de la hessienne ne détermine pas toujours à lui seul la nature d'un point critique.

Si $\nabla^2 f(x_*) > 0$ ou $\nabla^2 f(x_*) < 0$, il s'agit d'un extrémum local.

Si le point critique x_* ne satisfait ni $\nabla^2 f(x_*) \geq 0$ ni $\nabla^2 f(x_*) \leq 0$, alors il est ni un minimum ni un maximum local.

Par contre, un point critique x_* qui satisfait $\nabla^2 f(x_*) \geq 0$ (par exemple) peut correspondre à un minimum local, un maximum local, un point selle, ou (lorsqu'il est dégénéré) aucune de ces possibilités.

Une analyse *ad hoc* doit être faite pour le classifier.

21

Exemple Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction

$$f(x, y) = 2x^3 - 6x^2 - 8y^2 + 2.$$

1. f admet-elle un minimum global? un maximum global?
2. Trouver ses points critiques, et calculer la matrice hessienne à chacun d'eux.
3. Classifier chacun des points critiques. ?

22

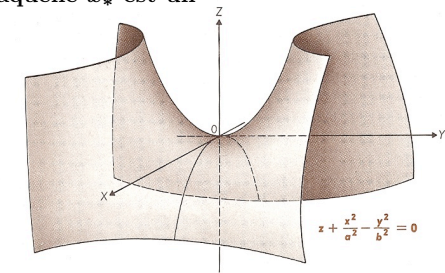
Classification des points critiques

Lorsqu'un point critique x_* satisfait $\det \nabla^2 f(x_*) \neq 0$, on dit qu'il est *non dégénéré*.

(aucune valeur propre = 0)

Dans ce contexte, x_* est soit un minimum local (quand $\nabla^2 f(x_*) > 0$), un maximum local (quand $\nabla^2 f(x_*) < 0$), ou un *point selle*.

Ce dernier type de point critique admet une direction par rapport à laquelle x_* est un minimum local, et une autre direction par rapport à laquelle x_* est un maximum local.



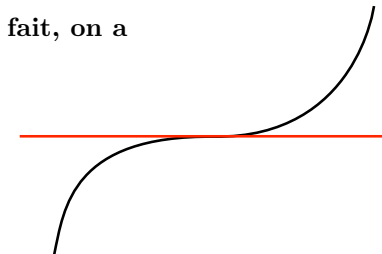
une paraboloïde hyperbolique

23

Exemple. La fonction $f(x, y) := x^3 + y^3$ admet un point critique en $(0, 0)$ qui est ni un min ni un max local, ni un point selle.

Il est forcément dégénéré. En fait, on a

$$\nabla^2 f(0, 0) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$



Question:

Que peut-on déduire lorsque

$$\nabla f(x_*) = 0 \text{ et } \nabla^2 f(x_*) \geq 0 \text{ (sdp)} ?$$

rien, sans une analyse ad hoc...mais nous n'allons pas considérer les points critiques dégénérés

24

Exemple

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ la fonction $f(x, y) = 2x^3 - 6x^2 - 8y^2 + 2$.

1. f admet-elle un minimum global? un maximum global?
2. Trouver ses points critiques, et calculer la matrice hessienne en chacun d'eux.
3. Classifier chacun des points critiques.

1. Non, car...

2. On trouve

$$\nabla f = \begin{bmatrix} 6x^2 - 12x \\ -16y \end{bmatrix} = \begin{bmatrix} 6x(x - 2) \\ -16y \end{bmatrix}$$

$\Rightarrow$ les points critiques sont $(0, 0)$ et $(2, 0)$.

On calcule

$$\nabla^2 f = \begin{bmatrix} 12x - 12 & 0 \\ 0 & -16 \end{bmatrix}$$

3. Conclusion:

$(0, 0)$ est un max local,
 $(2, 0)$ est un point selle

d'où

$$\nabla^2 f(0, 0) = \begin{bmatrix} -12 & 0 \\ 0 & -16 \end{bmatrix}, \quad \nabla^2 f(2, 0) = \begin{bmatrix} 12 & 0 \\ 0 & -16 \end{bmatrix}$$

25

Comment reconnaître qu'une matrice
symétrique est dp ou sdp ?

On peut utiliser la définition, et étudier
la forme quadratique de M

ou (comme dans l'exemple ci-dessus) :

Théorème (critère des valeurs propres) Soit M une matrice $n \times n$ symétrique. Alors $M > 0$ (définie positive) ssi toutes ses valeurs propres sont strictement positives.

On a $M \geq 0$ (semi-définie positive) ssi toutes ses valeurs propres sont positives au sens large.

26

Exemple

La matrice

$$\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$$

n'est pas définie positive (même si les entrées sont toutes positives!)

Preuve: Le polynôme caractéristique est $\lambda^2 - 2\lambda - 3$, d'où les valeurs propres sont 3 et -1 .

La matrice est indéfinie.

27

Comment reconnaître qu'une matrice
symétrique est dp ou sdp ?

Théorème (critère de Sylvester) Soit M une matrice $n \times n$ symétrique. Alors $M > 0$ (définie positive) ssi ses n sous-matrices principales *initiales* sont de déterminant strictement positif.

Il s'agit de commencer en haut à gauche :

$$[m_{11}], \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix}, \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \dots \text{etc.}$$

28

Comment reconnaître qu'une matrice symétrique est dp ou sd ?

Théorème (critère de Sylvester) Soit M une matrice $n \times n$ symétrique. Alors $M > 0$ (définie positive) ssi ses n sous-matrices principales *initiales* sont de déterminant strictement positif.

Exemple La matrice

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

est définie positive, puisque

$$\det [2] = 2 > 0,$$

$$\det \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 3 > 0,$$

$$\det \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = 4 > 0.$$

On ne sera pas responsable du critère de Sylvester

29

Comment reconnaître qu'une matrice symétrique est dp ou sd ?

Cas spécial : les matrices symétriques 2×2

Une matrice M symétrique 2×2 est définie positive ssi son déterminant $\det M$ et sa trace $\text{tr } M$ satisfont

$$\det M > 0, \quad \text{tr } M > 0.$$

Elle est semi-définie positive ssi

$$\det M \geq 0, \quad \text{tr } M \geq 0.$$

$$\begin{array}{l} \det M > 0 \\ \text{tr } M < 0 \end{array}$$

(preuve: $\det = \lambda_1 \lambda_2$
et $\text{tr} = \lambda_1 + \lambda_2 \dots$)

Quel est le résultat analogue pour définie négative ?

30

résumé :

- cas général du théorème de Lagrange, écriture matricielle
- conditions nécessaires/suffisantes pour un extremum local : matrices dp, sd...
- classification des points critiques

31

Le cas général $n > 1$:

Théorème Soit $f : U \rightarrow \mathbb{R}$ une fonction deux fois continûment différentiable sur un ouvert convexe $U \subset \mathbb{R}^n$, et soient x et y deux points distincts dans U . Alors il existe z dans le *segment* ouvert entre x et y (donc z , qui est dans U , est de la forme $x + t(y - x)$ pour $t \in]0, 1[$) tel que

$$f(y) = f(x) + \nabla f(x) \cdot (y - x) + \frac{1}{2} \nabla^2 f(z) (y - x) \cdot (y - x).$$

Ici, $\nabla^2 f(x)$ est la matrice hessienne de la fonction f , la matrice $n \times n$ symétrique définie par

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x) \right] \quad (i, j = 1, 2, \dots, n).$$

32

Conditions nécessaires ou suffisantes pour un minimum local

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ localement C^2 autour du point x_* , celui-ci étant un point critique.

Théorème

Conditions nécessaires: Si f atteint un minimum local en x_* , alors on a

$$\nabla^2 f(x_*) \geq 0 \text{ (sdp)}.$$

Conditions suffisantes: Réciproquement, si

$$\nabla^2 f(x_*) > 0 \text{ (dp)},$$

alors f atteint un minimum local en x_* .

(et de façon analogue pour un max local)

33

Comment reconnaître qu'une matrice symétrique est dp ou sdp ?

On peut utiliser la définition, et étudier la forme quadratique de M

ou :

Théorème (critère des valeurs propres) Soit M une matrice $n \times n$ symétrique. Alors $M > 0$ (définie positive) ssi toutes ses valeurs propres sont strictement positives.

On a $M \geq 0$ (semi-définie positive) ssi toutes ses valeurs propres sont positives au sens large.

35

On dit que M est *semi-définie positive* (sdp) lorsque $\langle Mx, x \rangle \geq 0 \quad \forall x \in \mathbb{R}^n$.

On écrit $M \geq 0$ dans ce cas.

Une matrice M symétrique $n \times n$ est dite *définie positive* (dp) si

$$\langle Mx, x \rangle > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

(La forme quadratique induite par M est strictement positive.) On écrit $M > 0$.

34

Comment reconnaître qu'une matrice symétrique est dp ou sdp ?

Pour les matrices symétriques 2×2

Une matrice M symétrique 2×2 est définie positive ssi son déterminant $\det M$ et sa trace $\text{tr } M$ satisfont

$$\det M > 0, \quad \text{tr } M > 0.$$

Elle est semi-définie positive ssi

$$\det M \geq 0, \quad \text{tr } M \geq 0.$$

36

Classification des points critiques

Lorsqu'un point critique x_* satisfait $\det \nabla^2 f(x_*) \neq 0$, on dit qu'il est *non dégénéré*.

(aucune valeur propre = 0)

Dans ce contexte, x_* est soit :

- un minimum local (quand $\nabla^2 f(x_*) > 0$) (toute valeur propre > 0)
- un maximum local (quand $\nabla^2 f(x_*) < 0$) (toute valeur propre < 0)
- un *point selle* (dans les autres cas).

Un point selle admet une direction par rapport à laquelle x_* est un minimum local, et une autre direction par rapport à laquelle x_* est un maximum local.

(au moins une valeur propre de chaque signe)

37

Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y),$$

$x, y \in \mathbb{R}^n, t \in [0,1].$

combinaison convexe de x et y (où $t \in [0,1]$)

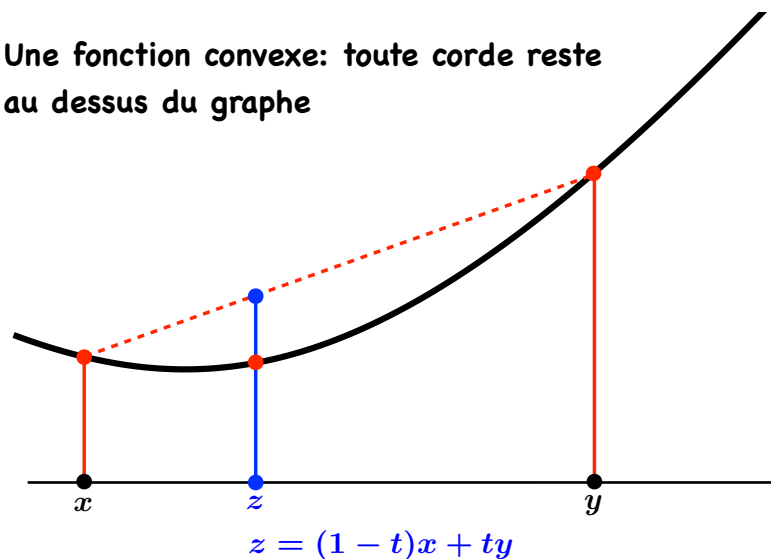
39

Les fonctions convexes

(sujet assez récent)

38

Une fonction convexe: toute corde reste au dessus du graphe



40

Proposition f est convexe ssi

$$f\left(\sum_{i=1}^m t_i x_i\right) \leq \sum_{i=1}^m t_i f(x_i)$$

pour toute combinaison convexe de points

$$x_i \in \mathbb{R}^n, m \geq 2, t_i \geq 0, \sum_{i=1}^m t_i = 1.$$

Une fonction g est **concave** quand la fonction $-g$ est convexe.

Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y), \\ x, y \in \mathbb{R}^n, t \in [0,1].$$

Proposition

1. Une fonction affine $x \mapsto \zeta \cdot x + c$ est convexe.

Preuve :

$$\|(1-t)x + ty\| \leq \|(1-t)x\| + \|ty\| \text{ (inégalité du triangle)} \\ = (1-t)\|x\| + t\|y\| \text{ (car } \|\lambda y\| = |\lambda|\|y\|)$$

41

43

Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y), \\ x, y \in \mathbb{R}^n, t \in [0,1].$$

Preuve :

$$\begin{aligned} f((1-t)x + ty) &= \zeta \cdot ((1-t)x + ty) + c \\ &= \zeta \cdot ((1-t)x) + \zeta \cdot (ty) + (1-t)c + tc \\ &= (1-t)\zeta \cdot x + (1-t)c + t\zeta \cdot y + tc \\ &= (1-t)[\zeta \cdot x + c] + t[\zeta \cdot y + c] \\ &= (1-t)f(x) + tf(y) \end{aligned}$$

Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y), \\ x, y \in \mathbb{R}^n, t \in [0,1].$$

Proposition

1. Une fonction affine $x \mapsto \zeta \cdot x + c$ est convexe.
2. Une norme $x \mapsto \|x\|$ définit une fonction convexe.
3. Si g est convexe sur $\mathbb{R}^p$, alors $f(x) := g(Lx + w)$ est convexe sur $\mathbb{R}^n$ (L une matrice $p \times n$, $w \in \mathbb{R}^p$).
4. Si la fonction $x \mapsto f(x)$ est convexe, alors la fonction $(x, y) \mapsto f(x)$ est convexe.
5. Une combinaison linéaire positive de fonctions convexes est convexe.

Exemple La fonction $(x, y) \mapsto |2x - y| - 3y$ est convexe.

Remarque Une fonction continue f est convexe ssi

$$f\left(\frac{1}{2}x + \frac{1}{2}y\right) \leq \frac{1}{2}f(x) + \frac{1}{2}f(y) \quad \forall x, y$$

42

44

Attention! La convexité (comme la continuité) ne se vérifie pas une coordonnée à la fois.

La fonction $(x, y) \mapsto xy$ est convexe par rapport à x , pour chaque y , et vice-versa (chaque fonction partielle est linéaire).

Mais elle n'est pas convexe (conjointement).

Il est utile d'avoir certains critères directs pour vérifier la convexité (ou son absence).

Par exemple, la convexité de la fonction $f(x) = x^4$ sur $\mathbb{R}$ saute aux yeux, mais la vérification par la définition est fastidieuse...

On peut localiser la notion de convexité, et ces critères.

45

Soit U un convexe dans $\mathbb{R}^n$

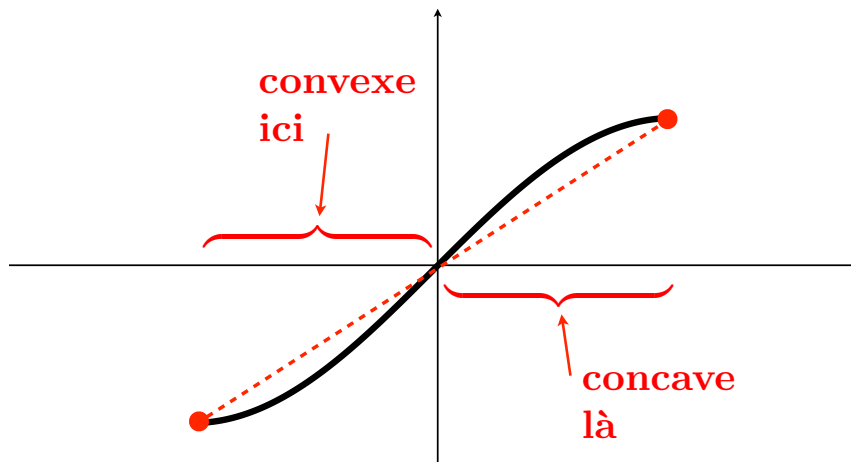
Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe sur U* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

$$\forall x, y \in U, t \in [0,1].$$

47

Une fonction pas très convexe



On peut définir la convexité par rapport à une partie U dans $\mathbb{R}^n$ si celle-ci est convexe...

46

Critères de convexité en termes de dérivées

1. critère de premier ordre

Théorème Soit U une partie convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction continûment dérivable autour de U .

Alors f est convexe sur U ssi

$$f(y) - f(x) \geq \nabla f(x) \cdot (y - x), \quad x, y \in U. \quad (*)$$

L'inégalité sous-différentielle

48

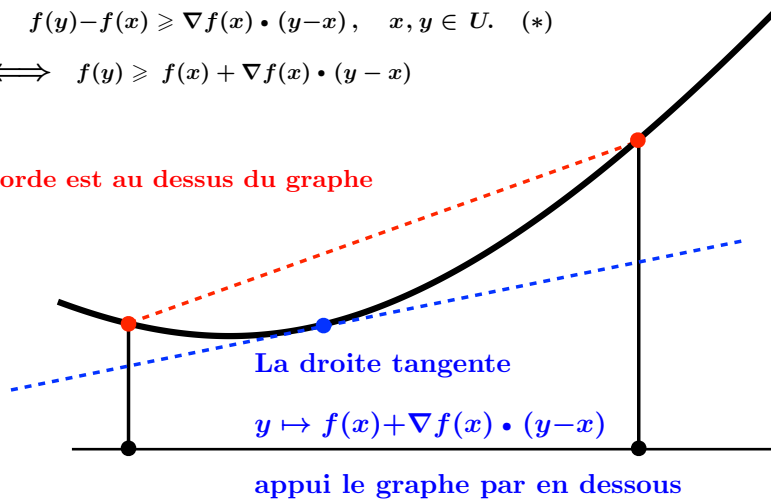
L'inégalité sous-différentielle : interprétation géométrique

f est convexe sur U ssi

$$f(y) - f(x) \geq \nabla f(x) \cdot (y - x), \quad x, y \in U. \quad (*)$$

$$\iff f(y) \geq f(x) + \nabla f(x) \cdot (y - x)$$

La corde est au dessus du graphe



49

Théorème Soit U une partie convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction continûment dérivable autour de U .

Alors f est convexe sur U ssi

$$f(y) - f(x) \geq \nabla f(x) \cdot (y - x), \quad x, y \in U. \quad (*)$$

Démonstration

Lemme Si f est convexe sur U et $x \in U$, alors la fonction

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

est croissante sur le domaine $t > 0$.

Preuve On montre facilement, pour $0 < s < t$, l'équivalence

$$g(s) \leq g(t) \iff f(x + sv) \leq (s/t)f(x + tv) + (1 - (s/t))f(x).$$

Mais cette dernière inégalité découle de la convexité de f , puisque

$$x + sv = \frac{s}{t}(x + tv) + \left(1 - \frac{s}{t}\right)x.$$

Corollaire La dérivée directionnelle unilatérale suivante existe:

$$f'_+(x; v) := \lim_{t \downarrow 0} \frac{f(x + tv) - f(x)}{t}.$$

51

Fixons $x, y \in U$. Si f est convexe alors

$$\nabla f(x) \cdot (y - x) = f'_+(x; y - x) \leq \frac{f(x + (y - x)) - f(x)}{1},$$

d'après le lemme, d'où (*).

Réciproquement: supposons (*). Pour $0 < t < 1$, on pose $z = (1 - t)x + ty$. Alors $z \in U$, car U est convexe. On invoque (*) pour découvrir

$$f(y) - f(z) \geq \nabla f(z) \cdot (y - z) \text{ et } f(x) - f(z) \geq \nabla f(z) \cdot (x - z).$$

On multiplie ces inégalités par t et $1 - t$, et on additionne pour obtenir

$$f((1 - t)x + ty) \leq (1 - t)f(x) + tf(y),$$

ce qui confirme la convexité de f . ■

50

52

Critères de convexité en termes de dérivées

2. critère de second ordre

Théorème Soit U un ouvert convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction deux fois continûment dérivable. Alors f est convexe sur U ssi

$$\nabla^2 f(x) \stackrel{\text{sdp}}{\geq} 0, \quad x \in U. \quad (**)$$

Rq : comme on a vu, on dispose de plusieurs critères pour vérifier qu'une matrice soit définie positive...

53

Soient x, y distincts dans U . On écrit le DL d'ordre deux avec reste de Lagrange :

$$f(y) - f(x) - \nabla f(x) \cdot (y-x) = (1/2) \nabla^2 f(z) (y-x) \cdot (y-x) \\ \text{où } z \in]x, y[.$$

Si $\nabla^2 f$ est partout sdp dans U , le terme à droite est positif; la convexité de f découle de la partie (1), puisque (*) est vérifié.

Réciproque: supposons que f est convexe. Alors le terme à gauche est positif, par (*). Fixons $v \in \mathbb{R}^n$, posons $y = x + tv$, qui appartient à U pour $t > 0$ suffisamment petit. On déduit

$$t^2 \nabla^2 f(z) v \cdot v \geq 0 \quad \text{où } z \in]x, x + tv[.$$

On divise par t^2 et on laisse tendre $t \downarrow 0$. On obtient $\nabla^2 f(x) v \cdot v \geq 0$. Puisque $x \in U$ et $v \in \mathbb{R}^n$ sont arbitraires, il suit que $\nabla^2 f$ est sdp sur U .



55

Critères de convexité en termes de dérivées

2. critère de second ordre

Théorème Soit U un ouvert convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction deux fois continûment dérivable. Alors f est convexe sur U ssi

$$\nabla^2 f(x) \stackrel{\text{sdp}}{\geq} 0, \quad x \in U. \quad (**)$$

Démonstration du théorème :

54

Exemple La fonction $t \mapsto t^4$ est convexe sur $\mathbb{R}$, car $(t^4)'' = 12t^2 \geq 0 \forall t$.

Exemple La fonction $t \mapsto t^3$ est convexe sur $]0, \infty[$, car $(t^3)'' = 6t > 0, t \in]0, \infty[$.

Exemple La fonction $(x, y) \mapsto f(x, y) = xy$ n'est pas convexe sur $\mathbb{R}^2$, car

$$\nabla^2 f(x, y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}. \quad (\text{matrice indéfinie})$$

Exemple

La fonction $(x, y) \mapsto f(x, y) = x^2 + y^2 + xy$ est convexe sur $\mathbb{R}^2$, car

$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}. \quad (\text{matrice définie positive})$$

56

Le rôle de la convexité en optimisation

Proposition Soit U un convexe ouvert, et soit f une fonction convexe et dérivable sur U . On se donne $x_* \in U$. Alors

f atteint un minimum local en x_*

$\iff f$ atteint un minimum global en x_* (par rapport à U)

$\iff \nabla f(x_*) = 0$

Démonstration:

f est convexe sur U ssi

$$f(y) - f(x) \geq \nabla f(x) \cdot (y - x), \quad x, y \in U. \quad (*)$$



Un renforcement de la convexité : la convexité stricte

Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y), \\ x, y \in \mathbb{R}^n, \quad t \in [0,1].$$

Définition La fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *strictement convexe* si

$$x \neq y, \quad t \in]0,1[\implies f((1-t)x + ty) < (1-t)f(x) + tf(y).$$

convexe, mais pas strictement

ici on a $f((1-t)x + ty) = (1-t)f(x) + tf(y)$.

Définition La fonction $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *strictement convexe* si

$$x, y \in U, \quad x \neq y, \quad t \in]0,1[\implies f((1-t)x + ty) < (1-t)f(x) + tf(y).$$

Proposition. Soit U une partie convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction strictement convexe. Alors le minimum de f sur U est atteint en au plus un point.

Remarque. Si f est convexe et g est strictement convexe, alors la somme $f + g$ est strictement convexe.

Un critère d'ordre deux :

Proposition. Soit U un ouvert convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction dans $C^2(U)$ telle que $\nabla^2 f(x) > 0 \forall x \in U$. Alors f est strictement convexe sur U .

Remarque. Ce critère d'ordre deux est suffisant mais pas nécessaire pour la convexité stricte.

Par exemple, la fonction $t \mapsto f(t) = t^4$ est strictement convexe, mais $f''(0) = 0$.

57

59

Démonstration

Lemme Soit h de classe C^2 sur $\mathbb{R}$ avec $h''(t) > 0 \forall t$. Alors

$$h(t) < (1-t)h(0) + th(1), \quad 0 < t < 1.$$

Preuve. On veut prouver

$$(1-t)[h(t) - h(0)] < t[h(1) - h(t)].$$

Par le théorème des accroissements finis, le terme à gauche vaut $(1-t)th'(a)$ et celui à droite $t(1-t)h'(b)$, où $a < b$. Puisque h' est strictement croissante, on obtient la conclusion.

Preuve de la proposition: Soient x et y deux points distincts dans U . On pose $h(t) := f(x + t(y - x))$, et on trouve

$$h''(t) = \nabla^2 f(x + t(y - x))(y - x) \cdot (y - x) > 0,$$

puisque $\nabla^2 f > 0$. Ceci nous permet d'invoquer le lemme, qui donne directement la conclusion.



58

60

Le rôle de la convexité en optimisation

Exemple Minimiser la fonction

$$f(x, y) = x^2 + 4xy + 9x + 11y + y^3 + 6y^2$$

sur le demi-plan $\{(x, y) : y \geq 0\}$.

Indication: penser à la convexité

61

résumé de la convexité des fonctions

63

Exemple Minimiser la fonction

$$f(x, y) = x^2 + 4xy + 9x + 11y + y^3 + 6y^2$$

sur le demi-plan $\{(x, y) : y \geq 0\}$.

2. On a

$$\nabla^2 f(x, y) = \begin{bmatrix} 2 & 4 \\ 4 & 6y + 12 \end{bmatrix}$$

3. On remarque

$$\det \nabla^2 f = \underbrace{12y + 8}_{> 0 \text{ sur le convexe } U := \{(x, y) : y > -\frac{1}{2}\}} \quad \text{tr } \nabla^2 f = \underbrace{6y + 14}_{> 0 \text{ sur le convexe } U := \{(x, y) : y > -\frac{1}{2}\}}$$

$\Rightarrow f$ est convexe sur U (en fait strictement)

$\Rightarrow f$ atteint un minimum (unique) par rapport à U en $(-13/2, 1)$

Ce point est donc solution (unique) du problème initial.

(C'est une approche **inductive**; on n'a pas à prouver l'existence a priori, étudier la frontière, classifier le point...)

62

Soit U un convexe dans $\mathbb{R}^n$

Définition Une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *convexe sur U* si

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y)$$

$$\forall x, y \in U, t \in [0, 1].$$

Une fonction g est **concave** quand la fonction $-g$ est convexe.

64

critères directs pour vérifier la convexité (ou son absence).

1. critère de premier ordre

Théorème Soit U une partie convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction continûment dérivable autour de U .

Alors f est convexe sur U ssi inégalité sous-différentielle

$$f(y) - f(x) \geq \nabla f(x) \cdot (y - x), \quad x, y \in U. \quad (*)$$

2. critère de second ordre

Théorème Soit U un ouvert convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction deux fois continûment dérivable. Alors f est convexe sur U ssi

$$\nabla^2 f(x) \geq \overset{\text{sdp}}{0}, \quad x \in U. \quad (**)$$

65

Proposition Soit U un convexe ouvert, et soit f une fonction convexe et dérivable sur U .
On se donne $x_* \in U$. Alors

f atteint un minimum local en x_*



f atteint un minimum global en x_* (par rapport à U)



$$\nabla f(x_*) = 0$$

(et si f est strictement convexe, le min est unique)

(P) $\min_A f$

Corollaire Si f est convexe sur un ouvert convexe U qui contient A , et si $\nabla f(x_*) = 0$ pour un point $x_* \in A$, alors x_* est solution du problème $\min_A f$.

(x_* est l'unique solution si f est *strictement* convexe)

67

Définition La fonction $f : U \subset \mathbb{R}^n \rightarrow \mathbb{R}$ est dite *strictement convexe* si

$$x, y \in U, x \neq y, t \in]0,1[\implies f((1-t)x+ty) < (1-t)f(x)+tf(y).$$

Proposition. Soit U une partie convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction strictement convexe. Alors le minimum de f sur U est atteint en au plus un point.

Un critère suffisant d'ordre deux :

Proposition. Soit U un ouvert convexe dans $\mathbb{R}^n$, et soit $f : U \rightarrow \mathbb{R}$ une fonction dans $C^2(U)$ telle que $\nabla^2 f(x) > 0 \forall x \in U$. Alors f est strictement convexe sur U .

66

Ce que couvrira le Partiel :

CM1 au CM6 (inclus)

(pas le CM7, donc pas le calcul différentiel d'ordre deux, ni la convexité)

68

FIN