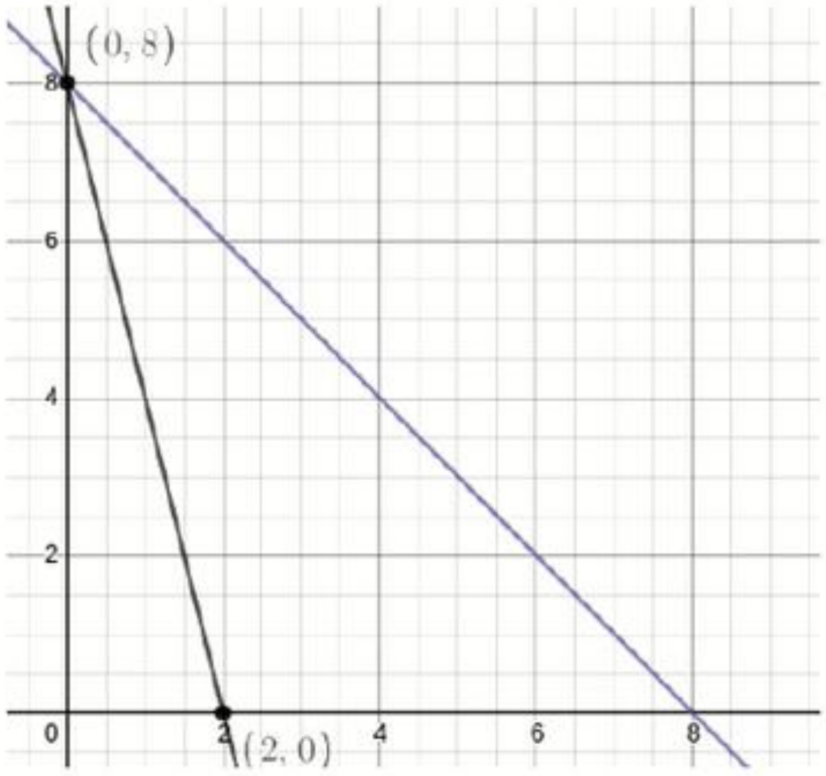
Optimisation sans contraintes

Partie 1

Programmation linéaire vs Programmation non linéaire

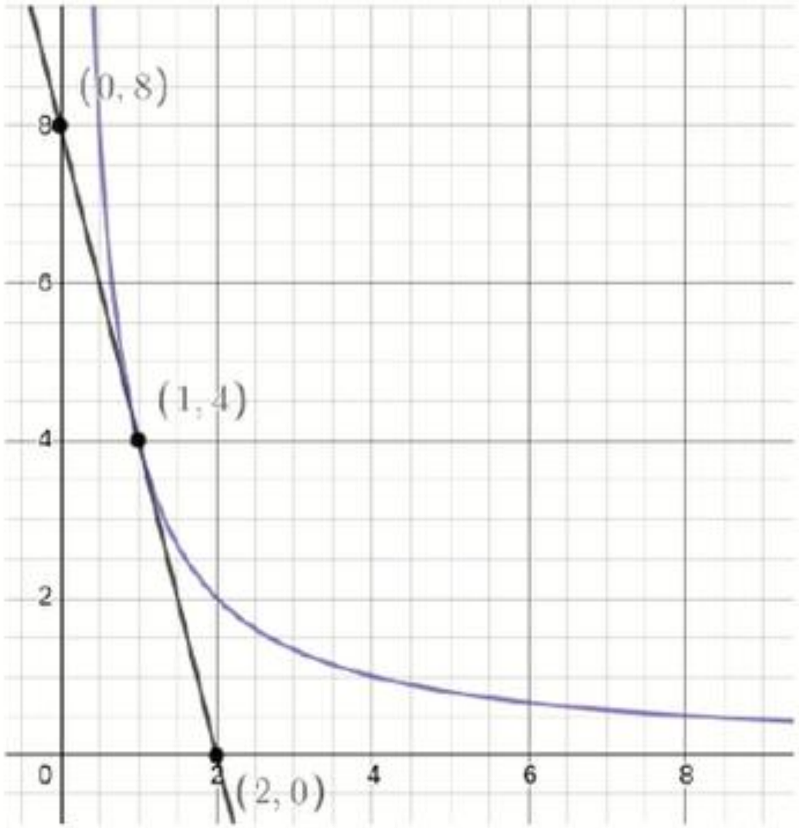
$$\max z = x + y$$

$$4x \quad + y \quad \leq \quad 8$$

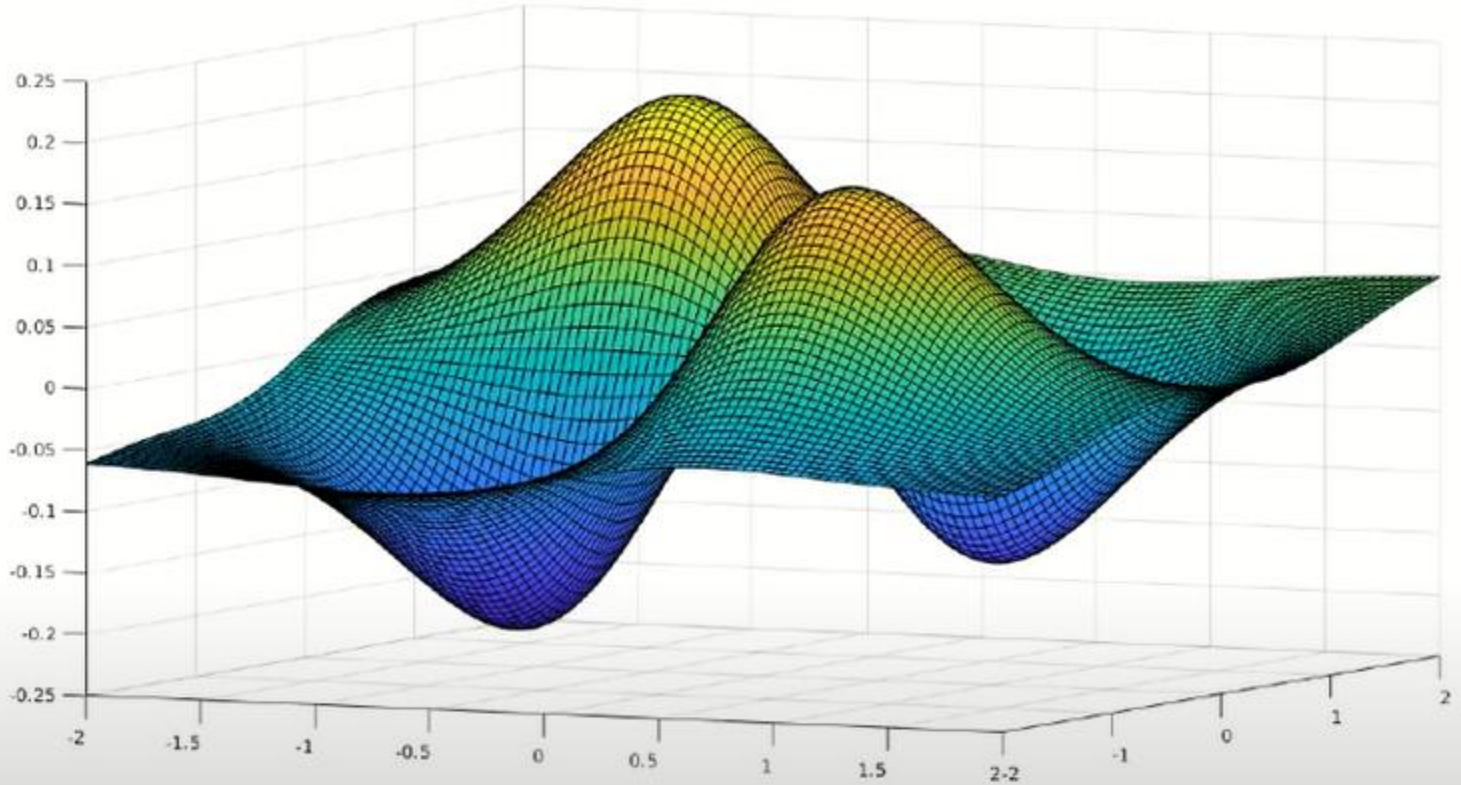


$$\max z = x.y$$

$$4x \quad + y \quad \leq \quad 8$$



Programmation non linéaire : plusieurs optimaux locaux



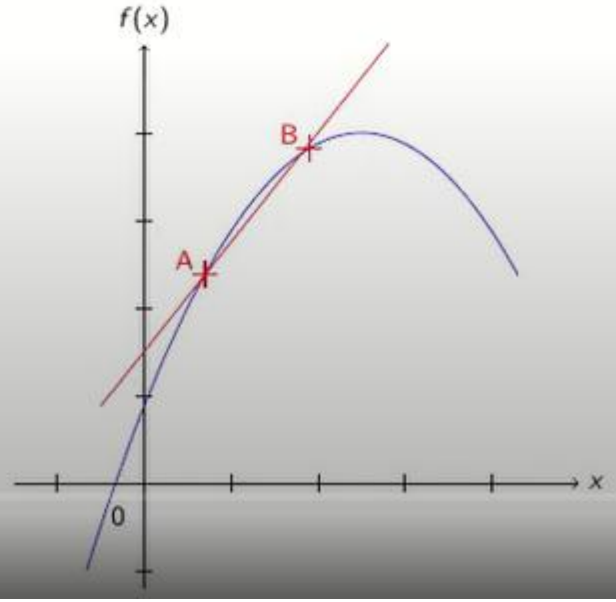
$$z = f(x, y) = 0.03 * y - xy * e^{(-x^2-y^2)}$$

Nombre dérivé

- Soit une fonction $f(x) : I \to \mathbb{R}$ définie sur un intervalle (ouvert ou fermé) I.
- Soit un réel a appartenant à I.
- Soit les points A de coordonnée $(a, f(a))$ et B de coordonnée $(a+h, f(a+h))$ qui sont sur la courbe représentant f.

Le coefficient directeur de la droite (AB) est

$$\frac{y_B - y_A}{x_B - x_A} = \frac{f(a+h) - f(a)}{a+h-a} = \frac{f(a+h) - f(a)}{h}$$



Dérivée

- Une fonction $f(x) : I \to \mathbb{R}$ est dérivable en a s'il existe un nombre réel L tel que $\lim\limits_{h \to 0} \frac{f(a+h)-f(a)}{h} = L$.
- L est alors appelé le nombre dérivé de $f(x)$ en a.
- Une fonction est dérivable sur I si elle est dérivable en tout $a \in I$.
- Pour f dérivable, on appelle dérivée de f la fonction qui associe à tout réel $x \in A$ le nombre dérivée de f en a.
- On note la fonction dérivée $f'(x)$ ou encore $\dfrac{df}{dx}(x)$.

Propriété: Une fonction dérivable sur I est continue sur I.

On note $\mathscr{C}^1$ l'ensemble des fonctions dérivables dont la fonction dérivée est elle-même continue.

Les dérivées analytiques usuelles

Fonctions	Dérivée	définie sur	dérivable sur
$x^n, n \in IN^*$	nx^{n-1}	IR	IR
$\frac{1}{x^n}, n \in IN^*$	$\frac{-n}{x^{n+1}}$	IR^*	IR^*
$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$	$[0, +\infty[$	$]0, +\infty[$
e^x	e^x	IR	IR
$a^x (a > 0)$	$a^x ln(a)$	IR	IR
$ln(x)$	$\frac{1}{x}$	$]0, +\infty[$	$]0, +\infty[$
$sin(x)$	$cos(x)$	IR	IR
$cos(x)$	$-sin(x)$	IR	IR
$tan(x)$	$1 + tan^2(x) = \frac{1}{cos^2(x)}$	$x \neq \frac{\pi}{2} + k\pi, k \in \mathbb{Z}$	
$arcos(x)$	$\frac{1}{\sqrt{1-x^2}}$	$-1 \leq x \leq 1$	$-1 < x < 1$
$arsin(x)$	$-\frac{1}{\sqrt{1-x^2}}$	$-1 \leq x \leq 1$	$-1 < x < 1$
$arctan(x)$	$\frac{1}{1+x^2}$	IR	IR

Opérations sur les dérivées

A partir des dérivées analytiques usuelles, on peut déterminer les dérivées de nombreuses fonctions par les opérations suivantes.

Pour deux fonctions f et g dérivables sur I,

- $f + g$ est dérivable sur I et $(f+g)' = f' + g'$
- fg est dérivable sur I et $(fg)' = f'g + fg'$
- si g ne s'annule pas sur I
  $\frac{f}{g}$ est dérivable sur I et $\left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$

Pour une fonction f dérivable sur I, g dérivable sur J telles que $f(x) \in J$ pour tout $x \in I$

- $g \circ f$ est dérivable sur I et $(g \circ f)' = f'g' \circ f$
  i.e. Dérivée de $g(f(x))$ est $f'(x)g'(f(x))$

Variations d'une fonction

- Soit f une fonction $\mathscr{C}^1$ définie sur un intervalle I. On a les équivalences suivantes :
  - f est croissante sur I (i.e., $f(x) \leq f(y)\ \forall x \leq y$ sur I)
    si et seulement si $f'(x) \geq 0$ pour tout $x \in I$
  - f est décroissante sur I (i.e., $f(x) \geq f(y)\ \forall x \leq y$ sur I)
    si et seulement si $f'(x) \leq 0$ pour tout $x \in I$
- Lorsque la dérivée s'annule en un point a de I, c'est que la pente de la tangente en ce point est horizontale.
- Un point $a \in I$ où $f'(a) = 0$ est appelé point stationnaire (ou point critique) .
- En un point stationnaire, la courbe représentant la fonction f présente une courbure horizontale.
- Si la fonction est croissante à gauche du point, puis croissante à sa droite (ou décroissante à gauche et décroissante à droite) on parle alors de **palier**.

Extremum local

Le réel a est appelé un maximum local (resp. minimum local) s'il existe un "petit" intervalle ouvert $I =]a-\epsilon, a+\epsilon[$ centré en a, tel que $\forall x \in I$, $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Propriété: Si a est un extremum local de f, alors $f'(a) = 0$

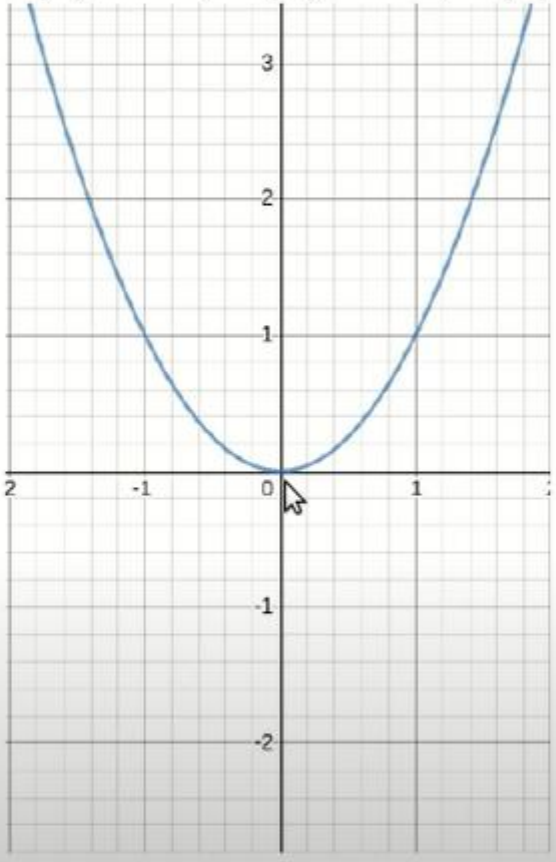
Attention: Inversement, une annulation de la dérivée en a n'indique pas nécessairement un extremum (cela peut-être un palier).

Propriété

Si $f'(a) = 0$ et si $f'(x)$ n'a pas le même signe
sur les intervalles $I =]a-\epsilon, a]$ et $I = [a, a+\epsilon]$
Alors a est un extremum local sur I.

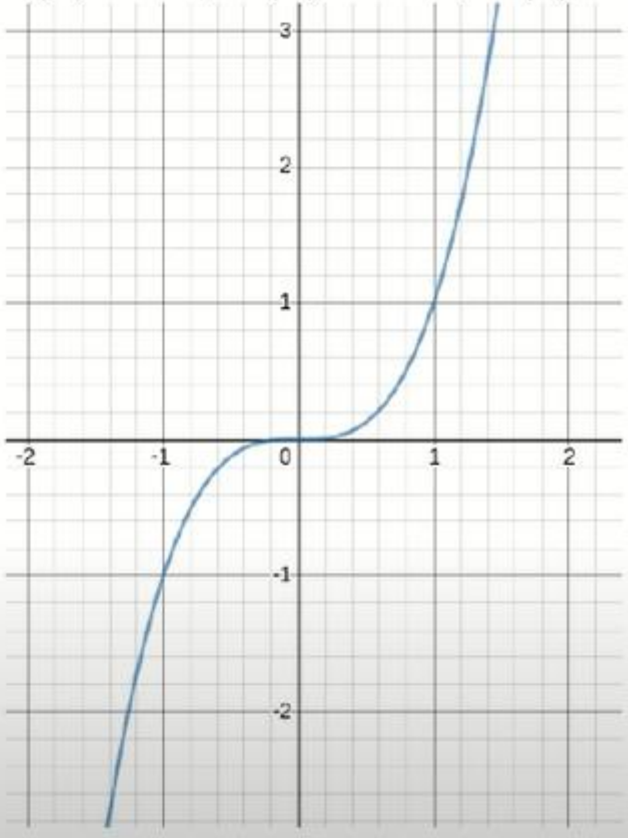
Exemple

$f(x) = x^2, f'(x) = 2x, f'(0) = 0$



Optimum local en 0.

$f(x) = x^3, f'(x) = 3x^2, f'(0) = 0$



Pas d'optimum local en 0.

La dérivée

- La dérivée mesure le taux de variation instantané d'une fonction par rapport à sa variable indépendante.

- Elle est notée généralement $f'(x)$ et se définit comme la limite du taux de variation moyen lorsque l'intervalle de variation tend vers zéro.

La dérivée a plusieurs interprétations :

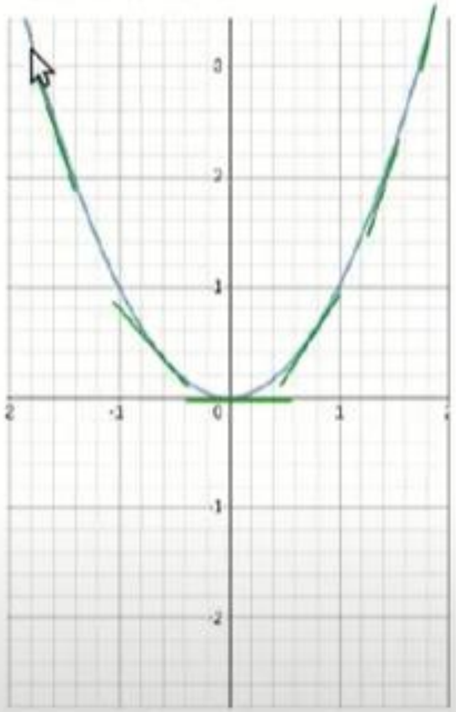
- **Taux de variation instantané :** Elle mesure à quel rythme la fonction change en un point particulier.
- **Pente de la tangente :** En géométrie, la dérivée représente la pente de la tangente à la courbe de la fonction en un point donné.
- **Vitesse instantanée :** Si la fonction représente la position d'un objet par rapport au temps, la dérivée de la fonction par rapport au temps donne la vitesse instantanée de l'objet.

Dérivée seconde

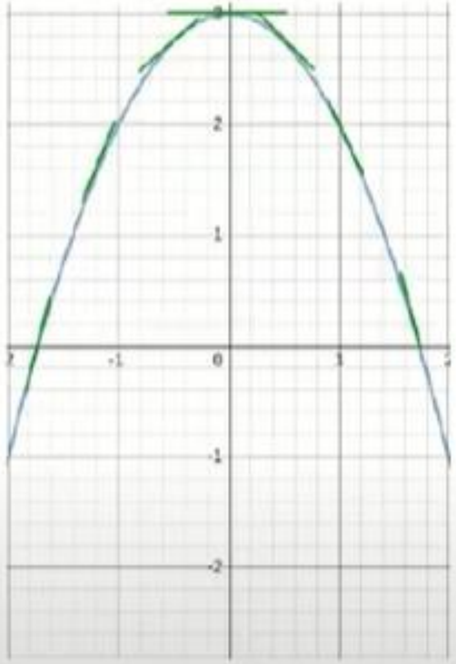
- Si une fonction f est de classe $\mathscr{C}^1$, et si sa dérivée f' est à son tour de classe $\mathscr{C}^1$, on peut considérer la dérivée seconde de f.
- On la note f'' ou encore $\dfrac{d^2 f}{dx^2}$.
- Si f'' est elle-même continue, on dit alors que f est de classe $\mathscr{C}^2$.
- La dérivée seconde indique la variation de la pente de la courbe représentative de la fonction et permet de mesurer la concavité locale de la fonction :
  - si elle est positive sur un intervalle, la pente augmente, la courbure est vers le haut, la fonction est dite « convexe » sur cet intervalle ;
  - si elle est négative sur un intervalle, la pente diminue, la courbure est vers le bas, la fonction est dite « concave » sur cet intervalle ;
  - si elle est nulle, la courbe est localement rectiligne ;
  - si la dérivée seconde s'annule et change de signe, on a un point d'inflexion, la courbure de la courbe s'inverse.

Exemple

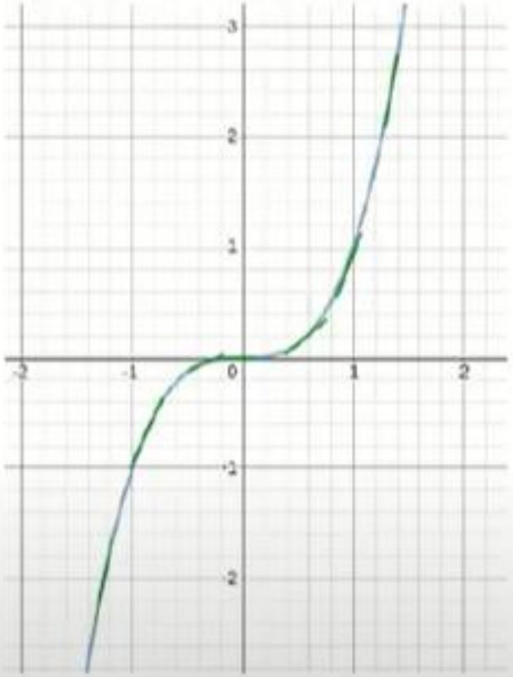
Convexe :



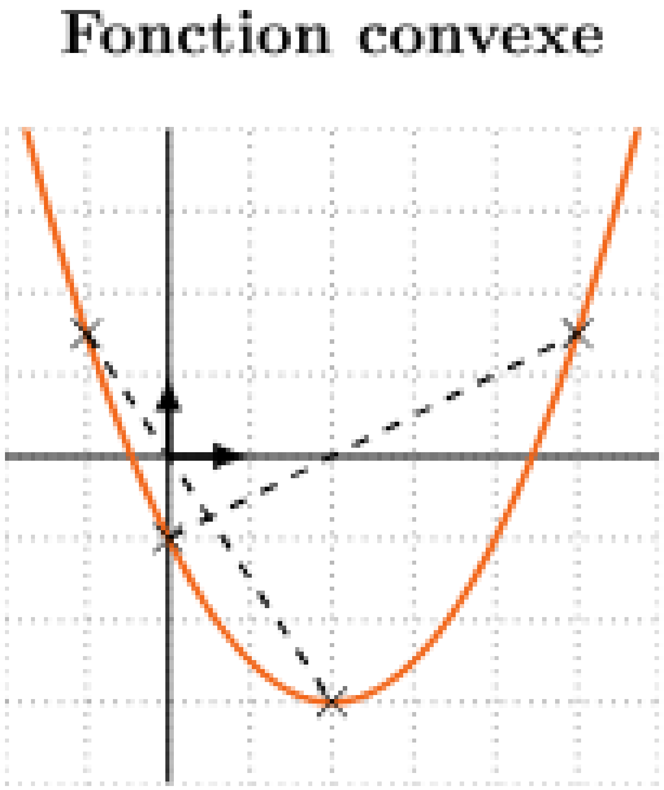
Concave :

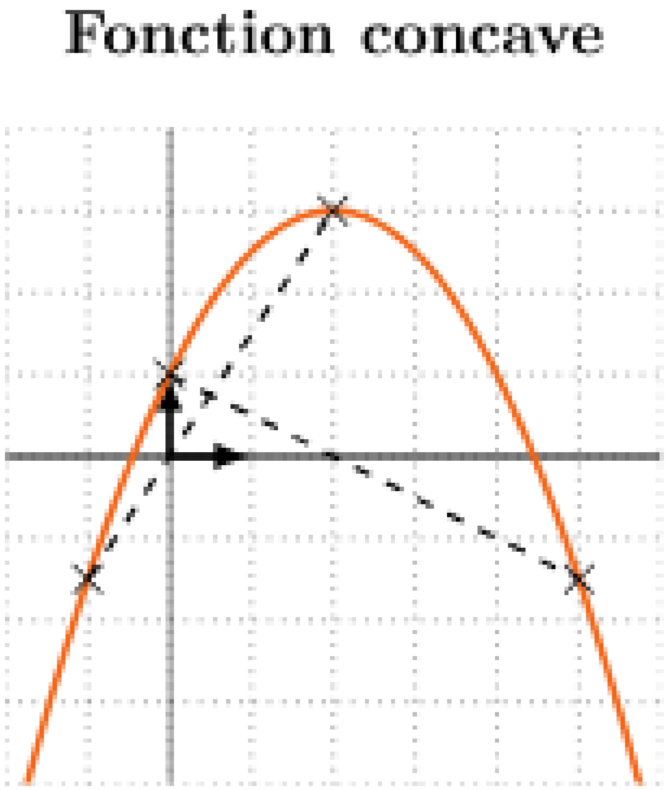


Point d'inflexion :



Convexité, concavité





On dit que f est convexe sur un intervalle I si tout segment reliant deux points de la courbe se trouve au-dessus de la courbe

On dit que f est concave sur un intervalle I si tout segment reliant deux points de la courbe se trouve en-dessous de la courbe

Condition suffisante d'optimalité locale

Theorem

Soit f une fonction de classe $\mathscr{C}^2$ sur un intervalle I.
Soit a un point intérieur de I.

- *Si $f'(a) = 0$ et $f''(a) < 0$, alors a est un maximum local.*
- *Si $f'(a) = 0$ et $f''(a) > 0$, alors a est un minimum local.*
- *Si $f'(a) = 0 = f''(a)$, on ne peut pas conclure.*

Preuve:
En effet, pour a intérieur à I, la dérivée d'une fonction en a indique la pente de la tangente à cette fonction en a.
Donc $f''(a)$ indique la pente de la tangente à f' en a.
Ainsi pour une fonction dérivée qui est décroissante (resp. croissante) au voisinage d'un point a, $f''(a) < 0$ (resp. $f''(a) > 0$).

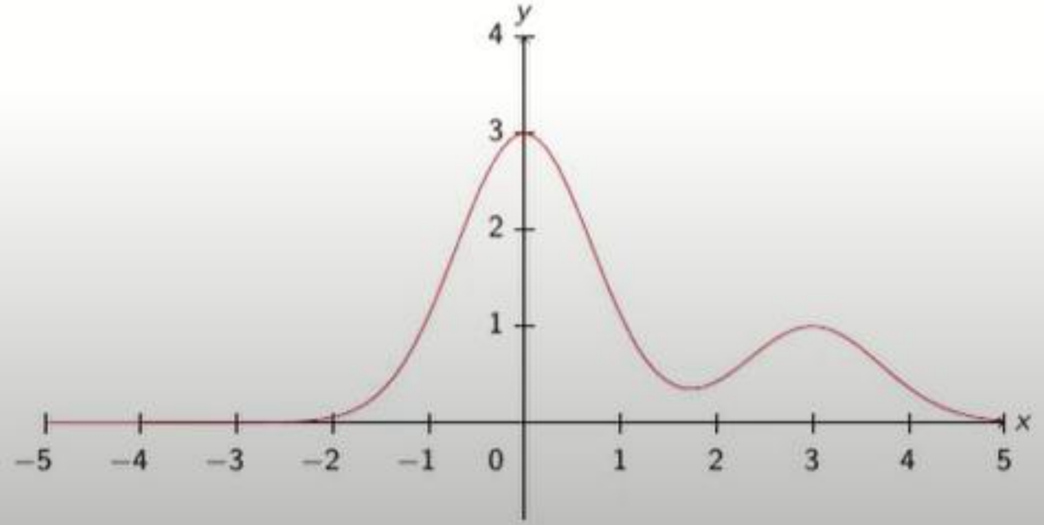
Donc pour un point a où $f'(a) = 0$,
quand $f''(a) < 0$, $f'(x)$ est positive avant a et négative après a.
quand $f''(a) > 0$, $f'(x)$ est négative avant a et positive après a. $\square$

Extremum global

- Soit une fonction $f(x) : I \to \mathbb{R}$ définie sur un intervalle I.
- Un réel $b \in X$ est un un maximum (resp. minimum) global sur I si $\forall x \in I,\ f(x) \leq f(b)$ (resp. $f(x) \geq f(b)$).

Remarque: Un extremum global est un extremum local. Mais un extremum local peut ne pas être un extremum global.

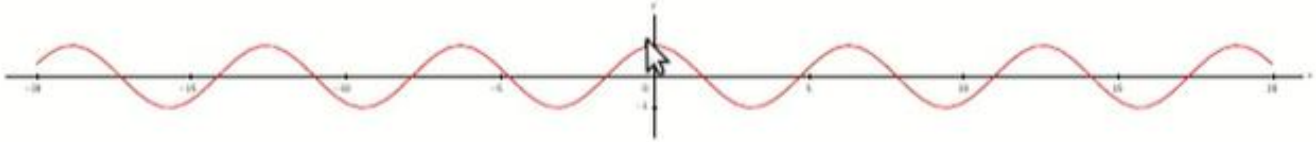
Cette courbe de la fonction définie sur IR par $f(x) = 3e^{-x^2} + e^{-(x-3)^2}$ comporte un maximum local en $x = 3$ qui n'est n'est pas un maximum global.
Le maximum global de cette fonction est en $x = 0$.



Extremum global

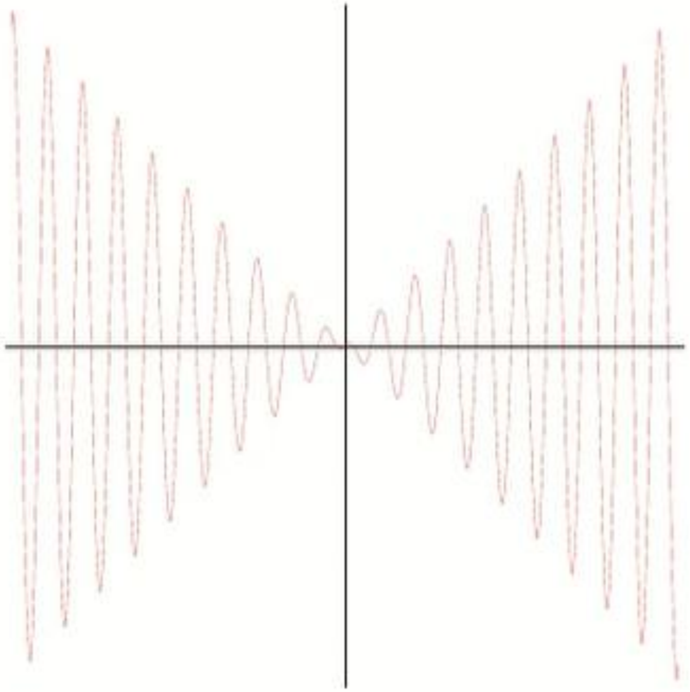
- Une fonction peut avoir une infinité d'extrema globaux.

Par exemple la fonction $f(x) = cos(x)$:



- Une fonction peut avoir une infinité d'extrema locaux sur des intervalles de taille non négligeable; et ne pas posséder pourtant d'extremum global.

Par exemple la fonction $f(x) = xcos(x)$.



Extremum global

Dans le cas simple des intervalles fermés, on a le résultat naturel suivant :

Théorème de Weierstrass dans $\mathbb{R}$:
Une fonction f continue sur un intervalle fermé $[a, b]$ possède un maximum global (resp. minimum global).

Dans le cas d'un intervalle ouvert $]a, b]$, $[a, b[$ ou $]a, b[$, comme dans le cas d'une union d'intervalles ouverts, une fonction f continue sur ce domaine de définition peut ne pas avoir d'optimum global.

C'est le cas de la fonction définie par $f(x) = \dfrac{x^2 - 1}{x + 2}$ qui est continue sur l'intervalle $]-2; +\infty[$ mais qui tend vers $+\infty$ sur le deux extrémités de l'intervalle.

Exemple 1 : Fonction à une variable

Considérons la fonction :

$$f(x) = x^3 - 3x + 1$$

- Trouver l'optimum de f .

Fonctions à plusieurs variables

Définitions et notations

Soit Ω un sous-ensemble de IR^n.
On considère une fonction de Ω dans IR.

La fonction f associe à tout point

$$x = (x_1, \ldots, x_n) \in \Omega$$

un réel

$$f(x) \in IR$$

Exemples:

- La fonction définie sur IR^2
  par $f(x_1, x_2) = \frac{x_1 - x_2}{1 + x_1^2 + x_2^2}$
- La fonction définie sur $\{(x_1, x_2) \in IR^2 \mid (x_1 - x_2) > 0\}$
  par $f(x_1, x_2) = \frac{x_1}{\sqrt{x_1 - x_2}}$

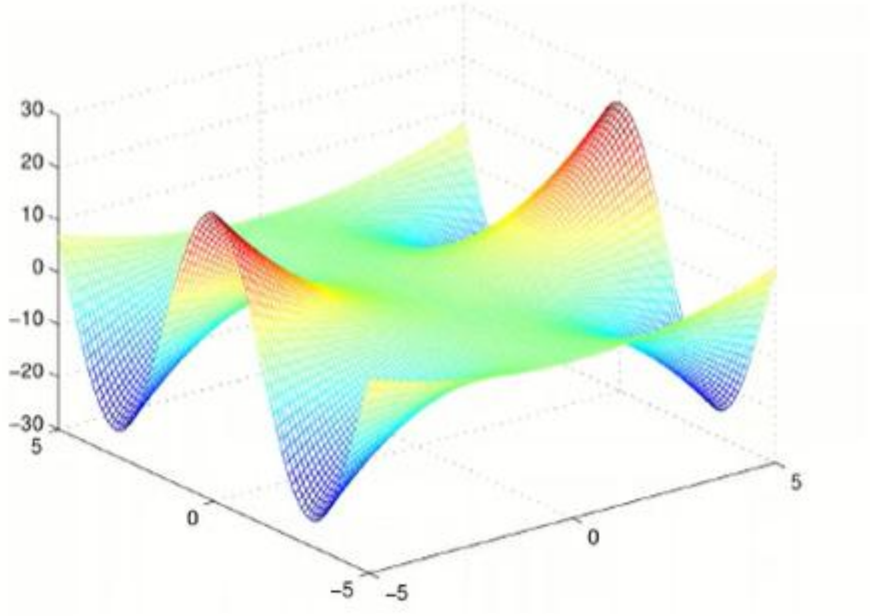
Dans IR^2, surface représentative

- Dans le cas particulier des fonctions définies sur $\Omega \in IR^2$, on peut définir la surface représentative de f comme l'ensemble des points de IR^3

$$\{(x_1, x_2, f(x_1, x_2)),\ \text{tq}\ (x_1, x_2) \in \Omega\}$$

- Cela revient à dire qu'on utilise la troisième dimension pour représenter la valeur de la fonction, on peut ajouter des couleurs du plus froid (bleu) au plus chaud (rouge) pour indiquer visuellement les petites et grandes valeurs de f.

Surface représentative de $f(x_1, x_2) = x_1^2 cos(x_2)$



Dans IR^2, courbes de niveau

- Dans le cas des fonctions définies sur $\Omega \in IR^2$, on peut définir la courbe de niveau de f en $\lambda \in IR$ comme l'ensemble des points de IR^2

$$\{(x_1, x_2) \in \Omega \text{ tq } f(x_1, x_2) = \lambda\}$$

- Si f n'atteint jamais la valeur λ, la courbe est vide.
- Dans le cas d'une valeur λ possible pour f, la courbe représente tous les couples (x_1, x_2) qui réalisent cette valeur.

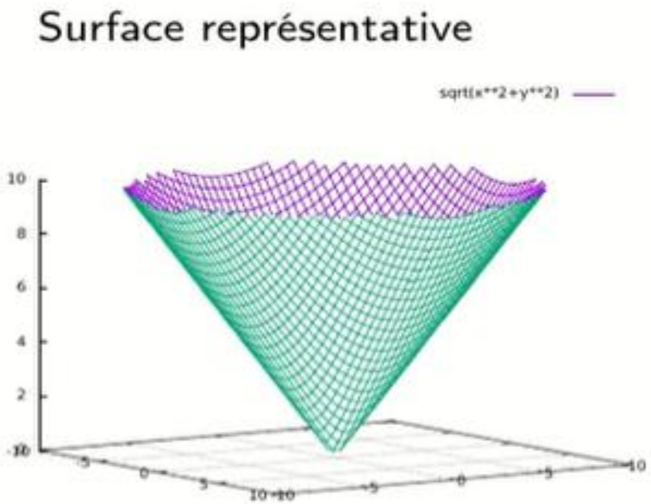
Dans IR^2, courbes de niveau

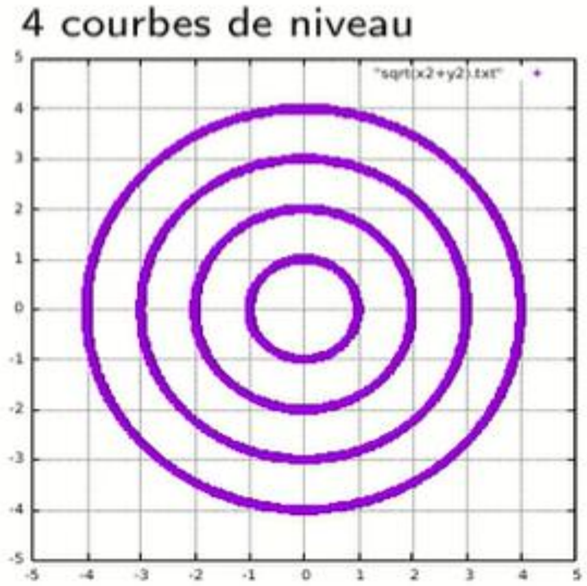
Voici la surface représentative de $f(x) = \sqrt{x^2 + y^2}$ c'est un cône centré en 0 dont la base est un cercle.

Dessous, voici la représentation de ses courbes de niveau pour $\lambda \in \{0, 1, 2, 3, 4\}$.

On peut remarquer qu'une courbe de niveau en λ est la "coupe" à $z = \lambda$ de la surface représentative.
Pour un $\lambda > 0$ donné, la courbe est un cercle de rayon λ.

Ici, on a représenté les 4 courbes sur un même plan : elles apparaissent concentriques.

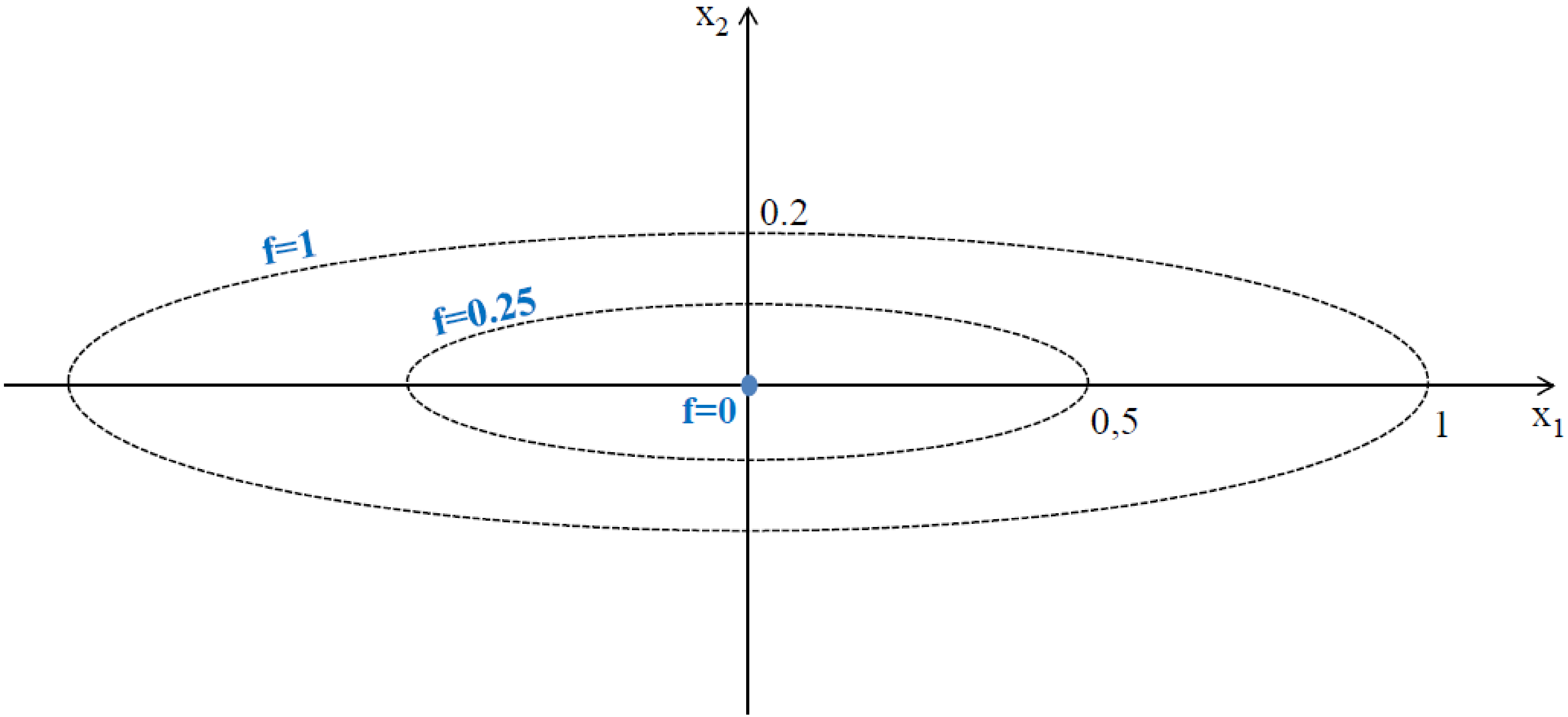




Exemple :

Ligne de niveau : fonction quadratique

$$f(x_1, x_2) = x_1^2 + 25x_2^2$$



Dérivées partielles

- Soit une fonction f définie sur un ouvert $\Omega \in IR^n$.
- On note e_i le point de IR^n tel que $e_j = 1$ si $j = i$ et 0 sinon.
- On appelle $(e_1, e_2, \ldots, e_n)$ la base canonique de IR^n et e_i est le $i$ème vecteur canonique.
- Soit un point $a = (a_1, \ldots, a_n) \in \Omega$.
  On considère alors la fonction de IR dans IR suivante
  $$t \mapsto f(a + te_i)$$
  Notez que $f(a + te_i) = f(a_1, \ldots, a_{i-1}, a_i + te_i, a_{i+1}, \ldots, a_n)$.
- On dit que la $i$ème dérivée partielle de f existe en a si la fonction $t \mapsto f(a + te_i)$ est dérivable en a.
- La $i$ème dérivée partielle existe sur Ω si elle existe en tout point de Ω.
- On note alors $\dfrac{\partial f}{\partial x_i}$ la $i$ème dérivée partielle.
- On note $\mathscr{C}^1$ l'ensemble des fonctions dont les n dérivées partielles existent et dont les n fonctions dérivées partielles sont elles-mêmes continues.

Remarque :
Pour dériver par rapport à une variable x_i, on considère les autres variables comme des constantes et on dérive comme on en a l'habitude par rapport à x_i.

Dérivées partielles

Exemple :

- Pour simplifier les notations, on va utiliser x, y, z plutôt que x_1, x_2, x_3.
- Prenons la fonction $f(x, y, z) = 3x^2y + 4\frac{x}{2y+z} + xyz$
- f est $\mathscr{C}^1$ sur $R^3 \setminus \{(x,0,0) \mid x \in IR\}$ avec:

$$\frac{\partial f}{\partial x}(x,y,z) = 6xy + \frac{4}{2y+z} + yz$$

$$\frac{\partial f}{\partial y}(x,y,z) = 3x^2 - 8\frac{x}{(2y+z)^2} + xz$$

$$\frac{\partial f}{\partial z}(x,y,z) = -4\frac{x}{(2y+z)^2} + xy$$

Changement de variable(dérivé par rapport à y):

On change $2y + z = \mu$ alors $\partial y = \frac{1}{2}\partial\mu$

Alors $\frac{df}{d\mu}\frac{4}{\mu} = \frac{df}{d\mu}4 * \mu^{-1} = -1\mu^{-2}$

Alors = $-2 * (2y+z)^{-2} = -\frac{2}{(2y+z)^2}$

Changement de variable(dérivé par rapport à z):

On change $2y + z = \mu$ alors $\partial z = \partial\mu$

Alors $\frac{df}{d\mu}\frac{4}{\mu} = \frac{df}{d\mu}4 * \mu^{-1} = -1\mu^{-2}$

Alors = $-(2y+z)^{-2} = -\frac{1}{(2y+z)^2}$

Notations

On note un point $x \in IR^n$ en vecteur-colonne $x = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_{n-1} \\ x_n \end{pmatrix}$.

Donc la notation transposée ^{T} désignera un vecteur-ligne.

$$x^T = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_{n-1} \\ x_n \end{pmatrix}^T = (x_1, x_2, \dots, x_{n-1}, x_n)$$

Et inversement:

$$x = (x_1, x_2, \dots, x_{n-1}, x_n)^T = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_{n-1} \\ x_n \end{pmatrix}$$

Le gradient

Soit $f(X)$ une fonction scalaire de n variables.

Le gradient de la fonction f qu'on note ∇f est un vecteur ligne dont les éléments sont les dérivées partielles de la fonction f suivant chaque variable.

$$\nabla f = \frac{\partial f}{\partial \mathbf{X}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}$$

Le gradient

Exemple Trouver le gradient de la fonction :

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

Solution

Les dérivées partielles suivant les variables x et y sont :

$$\frac{\partial f}{\partial x} = 2x + y,$$
$$\frac{\partial f}{\partial y} = x + 2y,$$

Le vecteur du gradient est donnée par :

$$\frac{\partial f}{\partial \mathbf{X}} = [2x + y \ , \ x + 2y]$$

Gradient

- Pour une fonction f $\mathscr{C}^1$ sur un ouvert $\Omega \in IR^n$ et $a \in \Omega$, on appelle gradient de f en a le vecteur-colonne

$$\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a) \right)^T$$

- Pour $n = 1$, c'est le nombre dérivée de f en a: donc la pente de la tangente à la courbe représentative à f en a.
- Pour $n = 2$, c'est le vecteur engendrant le plan tangent en a à la surface représentative de f.
- Pour n quelconque, on dira que le vecteur gradient au point a engendre un hyperplan tangent.

Plan tangent

Dans le cas de IR^2, le plan tangent à la surface représentative de la fonction f au point $a = (a_1, a_2)$ est

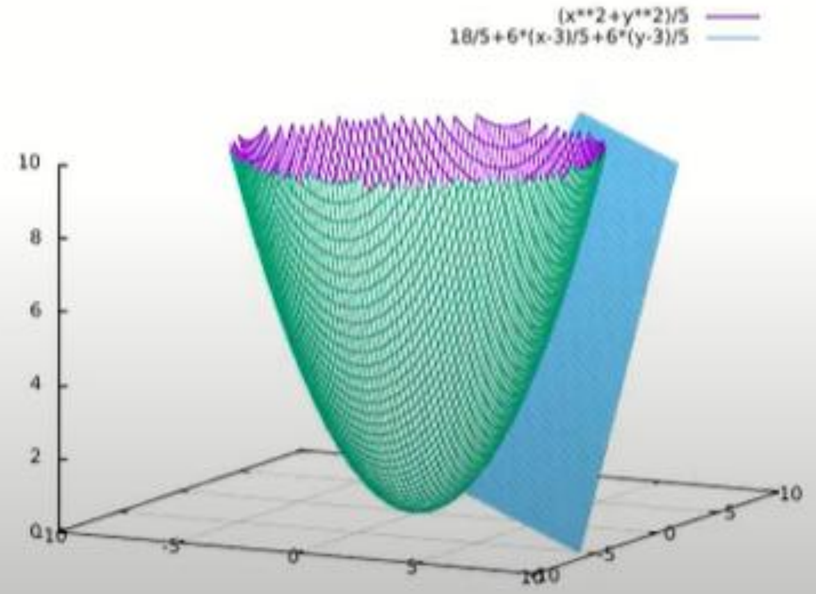
$$z = f(a_1, a_2) + \frac{\partial f}{\partial x_1}(a_1, a_2)(x_1 - a_1) + \frac{\partial f}{\partial x_2}(a_1, a_2)(x_2 - a_2)$$

Exemple: Pour la fonction parabolique $f(x) = \frac{1}{5}(x^2 + y^2)$,
On a $\nabla f(3,3) = (\frac{6}{5}, \frac{6}{5})$ et $f(3,3) = \frac{18}{5}$,

l'équation du plan est

$$z = \frac{18}{5} + \frac{6}{5}(x-3) + \frac{6}{5}(y-3)$$

Ce qui se représente ainsi :



Dérivées partielles d'ordre 2

Si f est $\mathscr{C}^1$ sur un ouvert Ω, sa $i$ème dérivée partielle est une fonction continue sur Ω.
Si elle admet à son tour une $j$ème dérivée partielle sur Ω, on la note:

$$\frac{\partial^2 f}{\partial x_j \partial x_i} = \frac{\partial f}{\partial x_j}\left(\frac{\partial f}{\partial x_i}\right)$$

et lorsque $i = j$

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial f}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right)$$

Si f admet des dérivées partielles d'ordre 2 en tout point de Ω et que ces fonctions sont elles-mêmes continues, on dit que f est $\mathscr{C}^2$ sur Ω.

Théorème de Schwarz :
si f est $\mathscr{C}^2$ sur Ω, alors $\dfrac{\partial^2 f}{\partial x_i \partial x_j} = \dfrac{\partial^2 f}{\partial x_j \partial x_i}$.

Dérivées partielles d'ordre 2

Exemple: Prenons la fonction $f(x, y) = x^3y^2$ qui est $\mathscr{C}^2$ sur IR^2.
On a:

$$\frac{\partial f}{\partial x}(x, y) = 3x^2y^2 \text{ et } \frac{\partial f}{\partial y}(x, y) = 2x^3y$$

Puis

$$\frac{\partial^2 f}{\partial x^2}(x, y) = 6xy^2 \qquad \frac{\partial^2 f}{\partial x \partial y}(x, y) = 6x^2y$$

$$\frac{\partial^2 f}{\partial y \partial x}(x, y) = 6x^2y \qquad \frac{\partial^2 f}{\partial y^2}(x, y) = 2x^3$$

Matrice hessienne

Si f est $\mathscr{C}^2$ sur Ω, et $a \in \Omega$, on définit la matrice hessienne de f en a comme la matrice dont le coefficient (i, j) est $\frac{\partial^2 f}{\partial x_i \partial x_j}(a)$, i.e.

$$\nabla^2 f(a) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(a) & \frac{\partial^2 f}{\partial x_2^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(a) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \frac{\partial^2 f}{\partial x_n \partial x_2}(a) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{pmatrix}$$

Notez que les éléments de cette matrice sont des réels.
Il s'agit d'une matrice carrée $n \times n$ et symétrique par le théorème de Schwarz.

Extremum local et global

Soit f une fonction réelle définie sur un ensemble $\Omega \in IR^n$.

Un réel $b \in X$ est un un maximum (resp. minimum) global sur Ω
si $\forall x \in X$, $f(x) \leq f(b)$ (resp. $f(x) \geq f(b)$).

Un réel a est appelé un maximum local (resp. minimum local)
s'il existe une boule centrée en a de rayon $r > 0$
telle que $\forall x \in B(x, r)$, $f(x) \leq f(a)$ (resp. $f(x) \geq f(a)$).

Cas des fonctions continues sur un compact :
Théorème de Weierstraas :
Si f est continue sur $\Omega \in IR^n$ avec Ω compact, alors f possède un optimum global $x^* \in \Omega$.

Attention: une fonction peut avoir des extrema locaux qui ne sont pas globaux.

La plupart des méthodes détectent uniquement des extrema locaux.

Matrice définie positive

- Une matrice **symétrique** A d'ordre n est définie positive (noté $A \succ 0$) si toutes ses valeurs propres sont strictement positives.
- Autre définition : une matrice **symétrique** A d'ordre n est définie positive si pour tout vecteur colonne non nul x à n éléments réels, on a $x^T Ax > 0$.
- Une matrice symétrique peut être :
  - Définie positive : $A \succ 0$
  - Semi-définie positive : $A \succeq 0$
  - Définie négative : $A \prec 0$ ($\Leftrightarrow -A \succ 0$)
  - Semi-définie négative : $A \preceq 0$ ($\Leftrightarrow -A \succeq 0$)
  - Non-définie.

Cas particulier pour une matrice symétrique $n = 2$:

- Une matrice $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ est définie positive si et seulement si $a > 0$ et $ac - b^2 > 0$.
- Sinon, si $a < 0$ et $ac - b^2 > 0$, A est définie négative.

Autre remarque utile : Le déterminant d'une matrice est égal au produit des valeurs propres.

Condition nécessaire d'optimalité de premier ordre

Dans le cas $\Omega = IR^n$

Theorem

Si x^* est un optimum local de f sur $\mathbb{R}^n$, alors $\nabla f(x^*) = 0$ (x^* est un point critique). (Ou point stationnaire)

- Attention : ce n'est pas une condition suffisante : Un point critique peut être un minimum local, un maximum local, ou bien un point-selle.
- Un point critique x est un point-selle si pour tout $\epsilon > 0$ il existe $a, b \in B(x, \epsilon)$ tels que $f(a) < f(x) < f(b)$.
- Si x n'est pas un point critique, il ne peut pas être un minimum ou un maximum.

Condition nécessaire d'optimalité de second ordre

Theorem

Si x^* est un minimum (resp. maximum) local de f sur $\mathbb{R}^n$, alors $\nabla f(x^*) = 0$ et la matrice hessienne $\nabla^2 f(x^*)$ est semi-définie positive (resp. négative).

Proof.

Soit x^* un minimum local. Pour toute direction unitaire $\mathbf{d} \in \mathbb{R}^n$ et $t \in \mathbb{R}$ suffisamment petit :

$$\begin{aligned} f(x^*) \leq f(x^* + t\mathbf{d}) &\simeq f(x^*) + t\mathbf{d}^T \nabla f(x^*) + \tfrac{t^2}{2}\mathbf{d}^T \nabla^2 f(x^*)\mathbf{d} \\ &= f(x^*) + \tfrac{t^2}{2}\mathbf{d}^T \nabla^2 f(x^*)\mathbf{d} \end{aligned}$$

$\Rightarrow \mathbf{d}^T \nabla^2 f(x^*)\mathbf{d} \geq 0$ (i.e. $\nabla^2 f(x^*)$ est semi-définie positive). $\square$

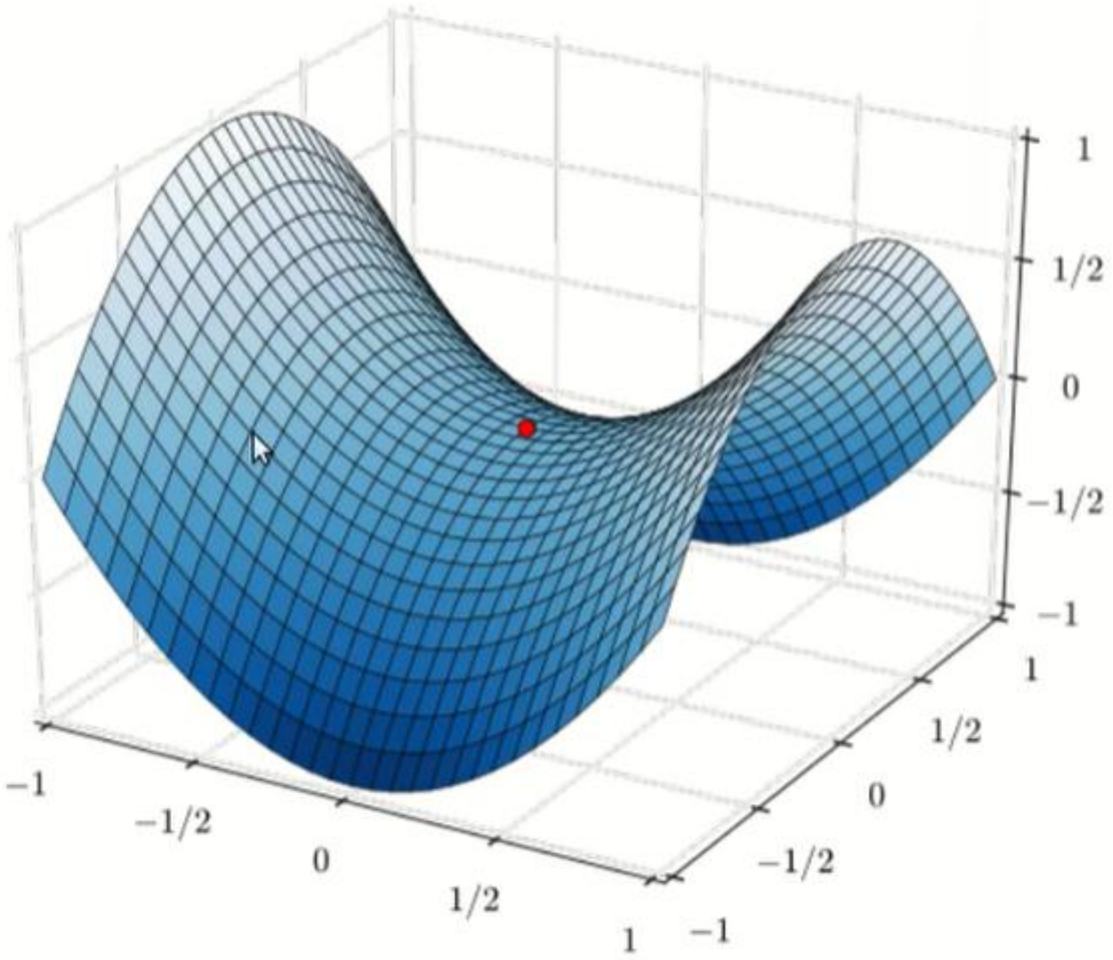
Condition suffisante d'optimalité de second ordre

Si $x^* \in \mathbb{R}^n$ est un point critique et si $\nabla^2 f(x^*)$ est :

- Définie positive : x^* est un minimum local
- Définie négative : x^* est un maximum local
- Indéfinie : x^* est un point-selle
- Semi-définie positive ou semi-définie négative : on ne peut rien dire. Une analyse du voisinage de x^* peut être nécessaire.

Point-selle (aussi appelé point col)

méthodes analytiques



$f(x, y) = x^2 - y^2$, point selle en (0,0).

Ensemble convexe

Un ensemble $S \in IR^2$ est dit convexe si
$\forall x, y \in S$ et $\forall \lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda) y \in S.$$

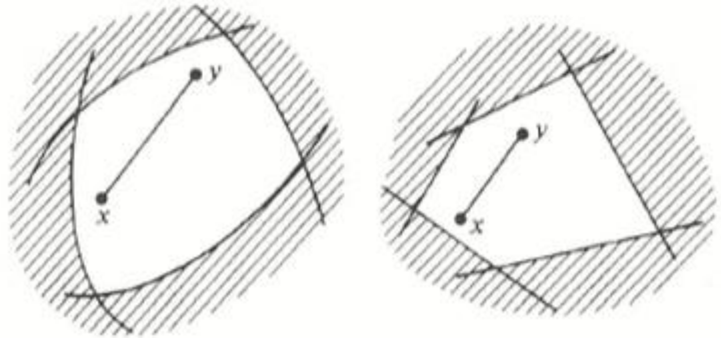
Cette définition revient à dire qu'un ensemble est convexe si, pour toute paire de points, le segment reliant ces deux points est entièrement dans l'ensemble.

Ensemble convexe

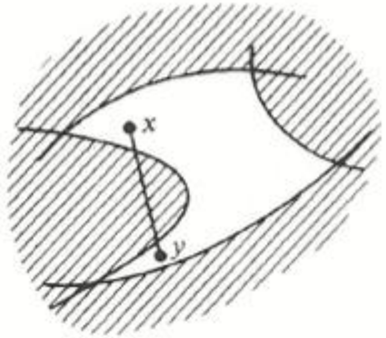
Les dessins ci-dessous montrent (dans IR^2) :

- à gauche un ensemble convexe délimité par des courbes
- à droite, un polyèdre (*i.e.*, intersection de demi-plans) et qui est un ensemble convexe

Dans les deux cas, on peut voir que le segment reliant x et y est entièrement contenu dans l'ensemble.



Ce dessin montre un exemple d'ensemble non-convexe : pour les points x et y proposés, une partie du segment est située à l'extérieur de l'ensemble.



Fonctions convexes

Fonctions à une variable

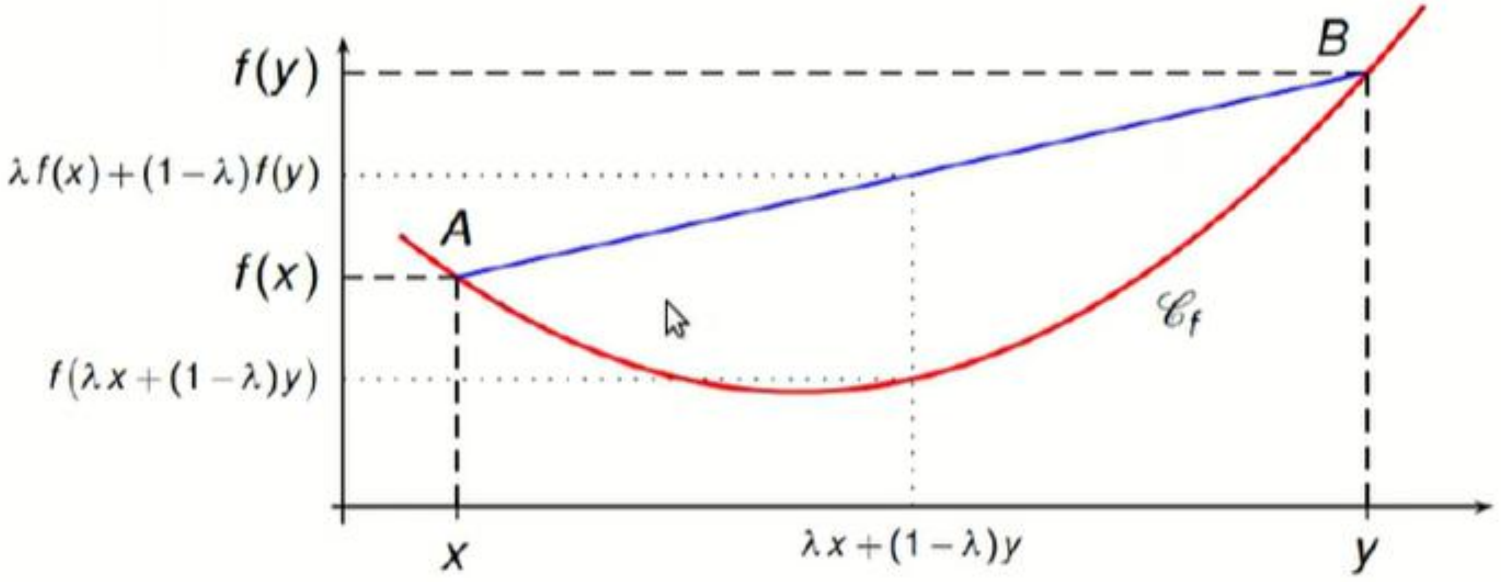
Fonction à une variable : convexité

- Une fonction $f(x) : IR \to IR$ est dite convexe si et seulement si $\forall (x, y) \in IR^2$ et $\forall \lambda \in [0, 1]$,

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

- Une fonction est dite concave si $-f$ est convexe.
- Autre définition :
  On dit que $f(x) : IR \to IR$ est convexe (resp. concave) sur un intervalle I si et seulement si pour tous points A et B de la courbe représentant f, l'arc de courbe est situé au-dessous du segment $[AB]$.

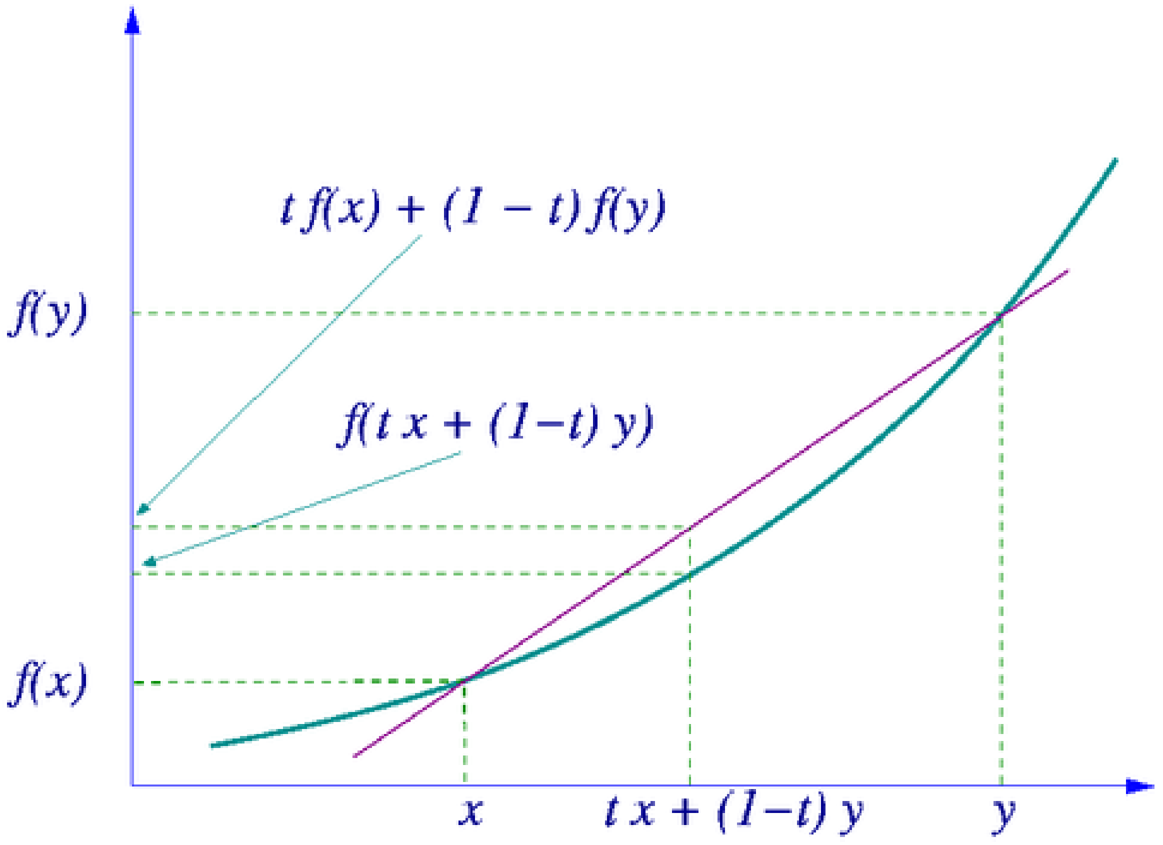
Illustration de la définition



La convexité

En géométrie :

Un ensemble est dit convexe si, pour chaque paire de points à l'intérieur de l'ensemble, la ligne droite reliant ces deux points est également à l'intérieur de l'ensemble.



Fonction Convexe

$\boldsymbol{f}: \boldsymbol{R}^{\boldsymbol{n}} \to \boldsymbol{R}$ **est convexe si :**

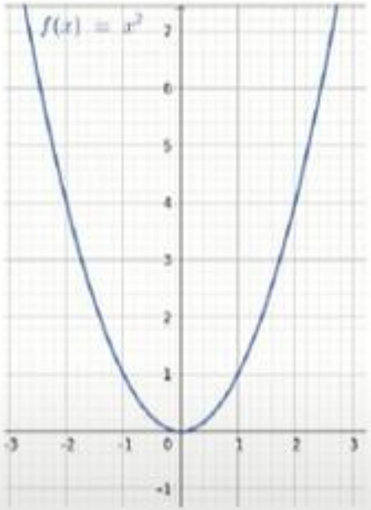
pour tout $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{R}^{\boldsymbol{n}}$ **et tout** $\boldsymbol{t} \in [\mathbf{0};\ \mathbf{1}]$,

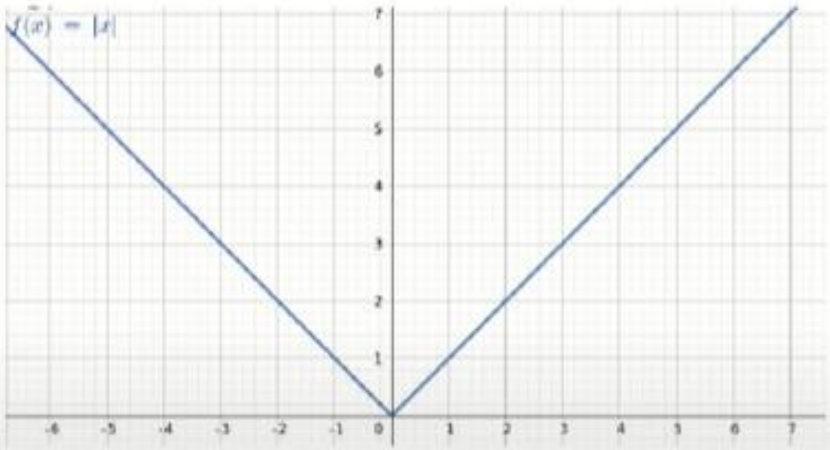
$$\boldsymbol{f}(\boldsymbol{tx} + (\mathbf{1} - \boldsymbol{t})\boldsymbol{y}) \leq \boldsymbol{tf}(\boldsymbol{x}) + (\mathbf{1} - \boldsymbol{t})\boldsymbol{f}(\boldsymbol{y}).$$

- une fonction convexe a une unique valeur minimale locale qui est également la valeur minimale globale.

Exemples

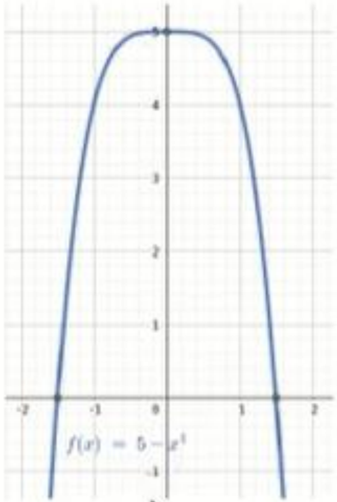
- $f(x) = x^2$ et $f(x) = |x|$ sont convexes :



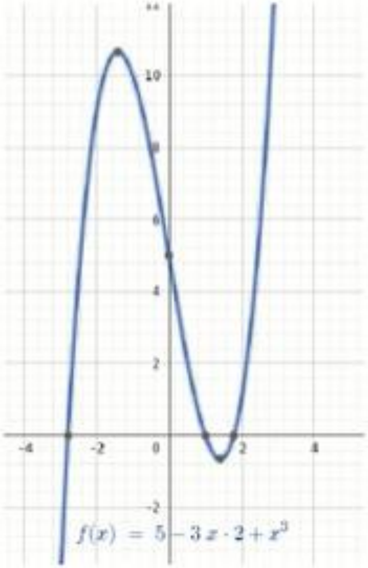


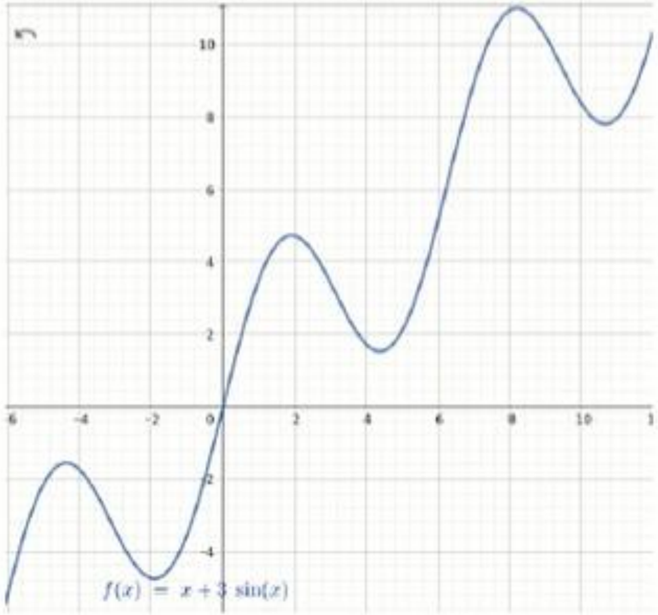
Exemples

- fonction concave :



- fonctions ni convexe ni concave :





Convexité et différentiabilité de seconde ordre

Corollaire 1.4.1

Soit f une fonction deux fois dérivable sur un intervalle I de $\mathbb{R}$. La fonction f est convexe si et seulement si sa dérivée seconde f''est à valeurs positives ou nulles.

$$\text{La fonction } f \text{ est convexe } \Longleftrightarrow f'' \geq 0 \text{ sur } I.$$

$$\text{La fonction } f \text{ est concave} \Longleftrightarrow f'' \leq 0 \text{ sur } I.$$

1. la fonction $f(x) = x^2$ est convexe car $f''(x) = 2 > 0, \forall x \in \mathbb{R}$.
2. la fonction $f(x) = \exp x$ est convexe car $f''(x) = \exp x > 0, \forall x \in \mathbb{R}$.
3. la fonction $f(x) = \ln x$ est concave car $f''(x) = -\frac{1}{x^2} < 0, \forall x \in \left]0, +\infty\right[$.

Fonction convexe et minimum global

Theorem

Tout optimum local d'une fonction convexe est un optimum global.

Preuve: Soit f une fonction convexe $f(x) : IR \to IR$ et un petit intervalle $I \subset IR$ centré autour d'un réel a.
Supposons que a est un minimum local sur I,
et supposons que a n'est pas un minimum global sur IR,
c'est-à-dire qu'il existe $y \in IR$ tel que $f(a) > f(y)$.
Comme f est convexe, alors $\forall \lambda \in [0, 1]$

$$f(\lambda a + (1 - \lambda)y) \leq \lambda f(a) + (1 - \lambda)f(y)$$

Donc

$$f(\lambda a + (1 - \lambda)y) < \lambda f(a) + (1 - \lambda)f(a) = f(a)$$

Or, pour des valeurs λ proche de 1, $\lambda a + (1 - \lambda)y$ représente les voisinages autour de a.
Ainsi on voit que a n'est pas minimum local dans son voisinage, contradiction. $\square$

- **Exemple:**

$$f = 15x_1 + 2(x_2)^3 - 3x_1(x_3)^2$$

$$\nabla f = \begin{bmatrix} 15 - 3(x_3)^2 & 6(x_2)^2 & -6x_1x_3 \end{bmatrix}$$

$$\nabla^2 f = \begin{bmatrix} 0 & 0 & -6x_3 \\ 0 & 12x_2 & 0 \\ -6x_3 & 0 & -6x_1 \end{bmatrix}$$

Méthode 1 :

Signe d'une matrice

$A \in \mathbb{R}^{n \times n}$ symétrique est dite

- Semi-définie positive si $\mathbf{x}^\top A\mathbf{x} \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$
- Définie positive si $\mathbf{x}^\top A\mathbf{x} > 0$ pour tout $\mathbf{x} \in \mathbb{R}^n \neq 0$
- Semi-définie négative si $\mathbf{x}^\top A\mathbf{x} \leq 0$ pour tout $\mathbf{x} \in \mathbb{R}^n$
- Définie négative si $\mathbf{x}^\top A\mathbf{x} < 0$ pour tout $\mathbf{x} \in \mathbb{R}^n \neq 0$
- Indéfinie sinon

En pratique, on peut vérifier le signe d'une matrice en examinant ses valeurs propres ou ses mineurs principaux dominants

Méthode 1 :

- exemple: $A = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$, Donc, nous avons

$$y^T Ay = y_1^2 + 4y_2^2 - 2y_1y_2$$
$$= (y_1 - y_2)^2 + 3y_2^2 > 0 \ \forall (y_1, y_2) \neq (0, 0)$$

donc, la matrice A est définie positive.

Méthode 2 :

- Une matrice est “définie positive” si toutes ces valeurs propres sont positives. (> 0)

- Une matrice est “définie négative” si toutes ces valeurs propres sont négatives. (< 0)

- Une matrice est “semi-définie positive” si toutes ces valeurs propres sont positives ou nulles. (≥ 0)

- Une matrice est “semi-définie négative” si toutes ces valeurs propres sont négatives nulles. (≤ 0)

<u>Méthode 2 :</u>

Exemple: Soit la matrice A

$$A = \begin{bmatrix} 2 & 4 & 5 \\ -5 & -7 & -1 \\ 1 & 1 & 2 \end{bmatrix}$$

- Le polynôme caractéristique de A est le det(A-I.λ)=p(λ)
- $p(\lambda) = -\lambda^3 - 3\lambda^2 + 8\lambda + 20$
- Les valeurs propres de A sont les racines de p(λ) . D'où

$\lambda_1 = -3.702 \qquad \lambda_2 = -2 \qquad \lambda_3 = 2.702$

Donc, Cette matrice est indéfinie.

Méthode 2 :

Exemple 2: soit la matrice A:

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

Les valeurs propres de A sont les racines du polynôme caractéristique de A, qui est égal à: $(2-\lambda)(2-4\lambda+\lambda^2)$.

D'où

$$\lambda_1 = 3.414 \qquad \lambda_2 = 0.586 \qquad \lambda_3 = 2$$

Comme toutes les valeurs propres sont > 0, alors la matrice A est définie positive.

Méthode 3 :

- **Les mineurs principaux:**

Si $A=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{12} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{nn} \end{pmatrix}$ alors $\Delta_1 = a_{11}, \Delta_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix}, \ldots, \Delta_n = \det(A)$ sont les mineurs principaux de A.

- A est définie positive ssi $\Delta_k > 0$ $k = 1, \ldots, n$
- A est définie négative ssi $(-1)^k \Delta_k > 0$ $k = 1, \ldots, n$
- si $\Delta_k > 0$ $k = 1, \ldots, n-1$ et $\Delta_n = 0$ alors A est semi-définie positive
- si $(-1)^k \Delta_k > 0$ $k = 1, \ldots, n-1$ et $\Delta_n = 0$ alors A est semi-définie négative
- si $\Delta_n < 0$ et si n est pair alors A est indéfinie

exemple: $A = \begin{pmatrix} 1 & -1 \\ -1 & 4 \end{pmatrix}$, $\Delta_1 = 1, \Delta_2 = 3$. A est donc définie positive.

I.4.4 Contours d'une fonction

$f : \mathbf{R}^n \rightarrow \mathbf{R}$

- $y = f(x)$ définit une surface dans $\mathbf{R}^{n+1}$

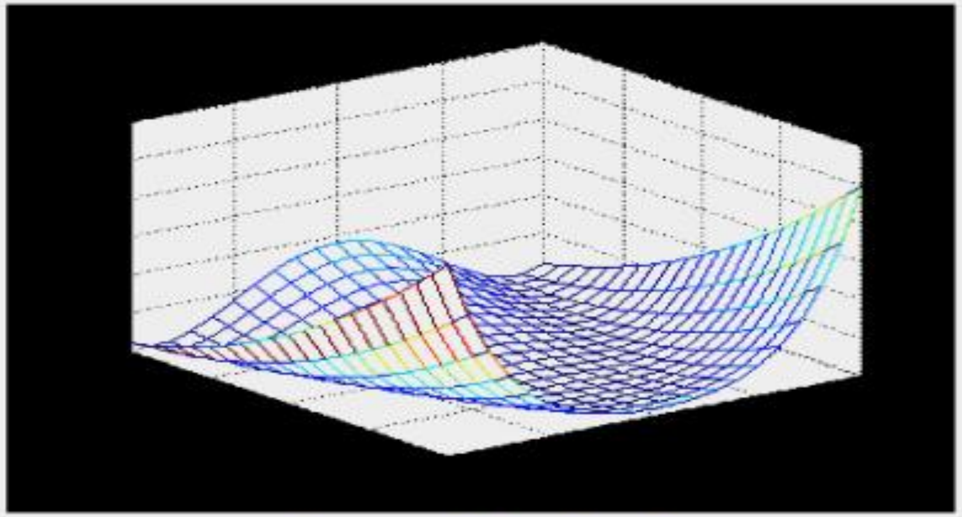
- $f(x) = c$, avec c constant

définissent des courbes de niveau ou des contours de cette surface

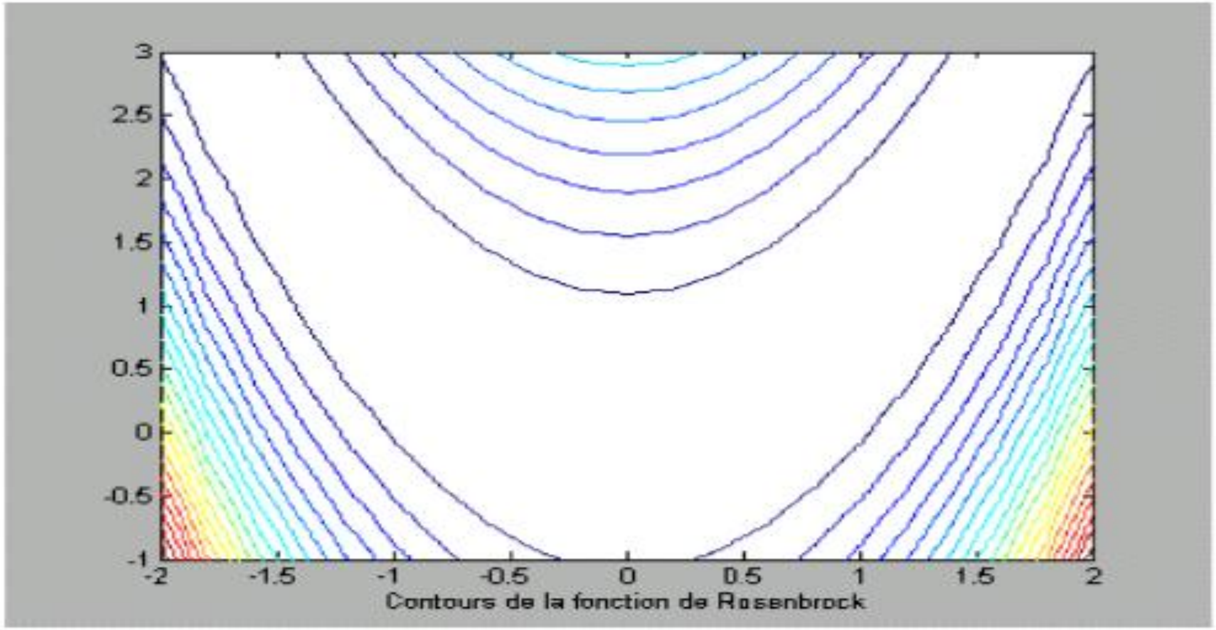
Un contour de niveau c est l'ensemble des points $S(c) = \{x \mid f(x) = c\}$

Exemple: La fonction de Rosenbrock

$$f(x) = 100\left(x_2 - x_1^2\right)^2 + \left(1 - x_1\right)^2$$



Fonction de Rosenbrock



Optimisation non linéaire sans contraintes

I- Cas d’une fonction à une seule variable:

$$\min_{x \in R} f(x)$$

Avec f : R$\rightarrow$R continue et dérivable

- La procédure d’optimisation sur une variables fonctionne comme suit:

1. Résoudre le problème qui consiste à trouver les points pour lesquels la dérivée de f s’annule.

2. Calculer la derivée seconde de la fonction objectif et l’évaluer pour chacun des points obtenus en 1.
   - Si la dérivée seconde est positive alors le point est un minimum local.
   - Si la derivée seconde est négatve alors le point est un maximum local.
   - Si, la dérivée seconde est nulle , alors le point est un point d’inflection.

Optimisation non linéaire sans contraintes

- La procédure d'optimisation sur plusieurs variables fonctionne comme suit:

1. Résoudre le problème qui consiste à trouver les points pour lesquels le gradient est nul.
2. Calculer le Hessien de la fonction objectif et l'évaluer pour chacun des points obtenus en 1.
   - Si la matrice (Hessienne) est "définie positive" alors le point est un minimum local.
   - Si la matrice est (Hessienne) "définie négative" alors le point est un maximum local.
   - Si la matrice (Hessienne) est "semi-définie négative", ou "semi-définie positive" alors, on ne peut rien dire.
   - Si, c'est indéfini, alors le point est un point col ou selle.

Exemple :

Soit $f(x,y) = x^3 - y^3 + 9xy$

- **Quel est le minimum et le maximum de $\boldsymbol{f(x,y)}$?**

$$\frac{\partial f}{\partial x}(x,y) = 3\,x^2 + 9y$$

$$\Leftrightarrow\ \nabla f(x,y) = (3x^2 + 9y, -3y^2 + 9x = 0)$$

$$\frac{\partial f}{\partial y}(x,y) = -3\,y^2 + 9x$$

- $Calculant\ \nabla f(x,y) = 0$

Exemple :

Soit $f(x,y) = x^3 - y^3 + 9xy$

$$\nabla f(x,y) = 0 \Leftrightarrow \begin{cases} 3x^2 + 9y = 0 \\ -3y^2 + 9x = 0 \end{cases} \quad \Leftrightarrow \begin{cases} x^2 + 3y = 0 \\ -y^2 + 3x = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} y = -\dfrac{x^2}{3} \\ -\dfrac{x^4}{9} + 3x = 0 \end{cases} \quad \Leftrightarrow \begin{cases} y = -\dfrac{x^2}{3} \\ x(-\dfrac{x^3}{9} + 3) = 0 \end{cases} \quad \Leftrightarrow \begin{cases} y = -\dfrac{x^2}{3} \\ x = 0 \text{ } ou \text{ } x^3 = 27 \end{cases}$$

$$\Leftrightarrow \begin{cases} x = 0 \\ y = 0 \end{cases} \quad ou \quad \begin{cases} x = 3 \\ y = -3 \end{cases} \quad \Longrightarrow \begin{cases} (3,-3) \text{est un point critique} \\ (0,0) \text{est un point critique} \end{cases}$$

Déterminer la nature des points critiques

$$\nabla f(x,y) = (3x^2 + 9y, -3y^2 + 9x) \quad \Rightarrow \nabla^2 f(x,y) = \begin{pmatrix} 6x & 9 \\ 9 & -6y \end{pmatrix}$$

Pour (*x*, *y*) = (0, 0) on a donc

$$\Rightarrow \nabla^2 f(0,0) = \begin{pmatrix} 0 & 9 \\ 9 & 0 \end{pmatrix}$$ $\Rightarrow$ calculer les valeurs propres de la matrice

$$\begin{pmatrix} -\lambda & 9 \\ 9 & -\lambda \end{pmatrix} \Rightarrow \lambda^2 - 81 = 0 \Rightarrow \lambda^2 = 81 \Rightarrow \lambda_1 = 9 \; ou \; \lambda_2 = -9$$

$$\Rightarrow \begin{cases} \lambda_1 > 0 \\ \lambda_2 < 0 \end{cases}$$ Alors la matrice $\nabla^2 f(0,0)$ est indéfinie donc $(0,0)$ est un point sel ou col

Déterminer la nature des points critiques

$$\nabla f(x,y) = (3x^2 + 9y, -3y^2 + 9x\ ) \quad \Rightarrow \nabla^2\ f(x,y) = \begin{pmatrix} 6x & 9 \\ 9 & -6y \end{pmatrix}$$

Pour (*x*, *y*) = $(\mathbf{3}, -\mathbf{3})$on a donc

$$\Rightarrow \boldsymbol{\nabla^2}\ \boldsymbol{f}(\mathbf{3}, -\mathbf{3}) = \begin{pmatrix} 18 & 9 \\ 9 & 18 \end{pmatrix}$$

$\Delta_1 = 18 > 0\ et\ \Delta_2 = 243 > 0$ alors la matrice $\nabla^2 f(3,-3)$ est définie positive

Alors $(3,-3)$ est un minimum local