FACULTE DES SCIENCES BEN M'SICK
UNIVERSITE HASSAN II DE CASABLANCA

Algorithmes d’optimisation

Pr. Faouzia Benabbou (faouzia.benabbou@univh2c.ma)
Département de mathématiques et Informatique
Master Data Science & Big Data
2024-2025

Plan du Module: Algorithmes d'optimisation

01 Introduction

02 Rappels mathématiques

03 Algorithmes d'Optimisation Sans Contraintes

04 Algorithmes d'Optimisation avec Contraintes

05 Programmation Linéaire et non Linéaire

06 Optimisation combinatoire et globale

07 Méthodes heuristiques et métaheuristiques

Rappels mathématiques

Matrices

- **Matrice Définie +/-**

  Soit A une matrice carrée réelle de taille n×n.

  On dit que A est:

  - **Définie positive** si pour tout vecteur $x \in \mathbb{R}^{*n}$, $x^T A x > 0$.
  - **Semi-définie positive** si pour tout vecteur $x \in \mathbb{R}^{n}$, $x^T A x \geq 0$.
  - **Définie négative** si pour tout vecteur $x \in \mathbb{R}^{*n}$, $x^T A x < 0$.
  - **Semi-définie négative** si pour tout vecteur $x \in \mathbb{R}^{n}$, $x^T A x \leq 0$.

- $x^T A x$ est appelée la forme **quadratique** ou **l'énergie** d'une matrice.

Matrices

- **Matrice Définie +/-**
  - **Exemple.**
    - ✓ X=(x,y) et A=$\begin{bmatrix} a & b \\ b & c \end{bmatrix}$.

$$x^{\top}Ax = (x, y)\begin{bmatrix} a & b \\ b & c \end{bmatrix}(x, y) = ax^2 + 2bxy + cy^2$$

    - ✓ X=(x,y) et A=$\begin{bmatrix} 1 & 6 \\ 6 & -1 \end{bmatrix}$.

$$x^{\top}Ax = (x, y)\begin{bmatrix} 1 & 6 \\ 6 & -1 \end{bmatrix}(x, y) = x^2 + 12xy - y^2$$

Matrices

- **Matrice Définie +/-**
  - **Théorème.** Une matrice symétrique réelle A est définie positive si et seulement si toutes ses **valeurs propres** sont strictement positives.

  - **Décomposition de Cholesky.** Si A est symétrique et définie positive, alors elle peut se décomposer sous forme suivante:

$$A = LL^T$$

L une matrice triangulaire inférieure.

Matrices

- ## Matrice Définie +/-

  - **Critères de détermination.** Il existe plusieurs critères pour déterminer si une matrice symétrique est définie positive.
  - Les plus courants sont :

    1. Calculer les **valeurs propres** de la matrice A. Si toutes les valeurs propres sont strictement positives, alors la matrice est définie positive.
    2. Critère de Sylvester. Calculer les **déterminants** de toutes les sous-matrices principales de A. Si tous ces déterminants sont strictement positifs, alors la matrice est définie positive.
    3. Décomposition de **Cholesky**. Si la matrice A admet une décomposition de Cholesky, alors A est définie positive.

Matrices

- **Matrice Définie +/-**
  - **Critères de détermination.** Ces propriétés sont importantes car elles permettent de déterminer certaines **caractéristiques** de la matrice et de la forme quadratique associée.
  - Par exemple :
    - ✓ Les matrices définies positives sont **inversibles**.
    - ✓ Les **valeurs propres** d'une matrice définie positive sont toutes strictement positives.
    - ✓ La matrice **identité** est définie positive.
    - ✓ Une matrice **diagonale** dont tous les éléments diagonaux sont positifs est définie positive.
    - ✓ Une matrice de **covariance** est semi-définie positive.

Matrices

- **Matrice Définie +/-**
  - **Exemple de matrice DP.** La matrice de **covariance** est un outil important qui permet de :
    - ✓ **Visualiser les relations:** Elle permet de voir rapidement quelles variables sont **corrélées** entre elles et dans quelle direction.
    - ✓ **Réduire la dimensionnalité:** Dans les problèmes avec de nombreuses variables, elle aide à identifier les groupes de variables qui **varient ensemble**, ce qui permet de simplifier l'analyse.
    - ✓ **Analyse en composantes principales (ACP):** Elle est utilisée dans l'ACP pour transformer des variables corrélées en un plus petit nombre de variables non corrélées (réduction de la dimension).
    - ✓ **Modélisation statistique:** Elle est essentielle dans de nombreux modèles statistiques, notamment les modèles gaussiens et les modèles de régression multiple.

Matrices

▪ Matrice Définie +/-

• Exemple de covariance.

- ✓ La matrice de covariance se calcule à l'aide de la formule suivante :
  $$Cov(x, y) = \Sigma[(xi - \bar{x})(yi - \bar{y})] / (n - 1)$$
- ✓ xi et yi sont les valeurs individuelles des variables X et Y
- ✓ $\bar{x}$ et $\bar{y}$ sont les moyennes des variables x et y, n est le nombre d'observations.
- ✓ Supposons que nous ayons les données suivantes pour trois variables : la température (T), l'humidité (H) et la consommation d'énergie (E) :

T (°C)	H (%)	E (kWh)	
20	60	150	
25	50	180	
30	70	200	
25	60	176,67	Moyenne

Matrices

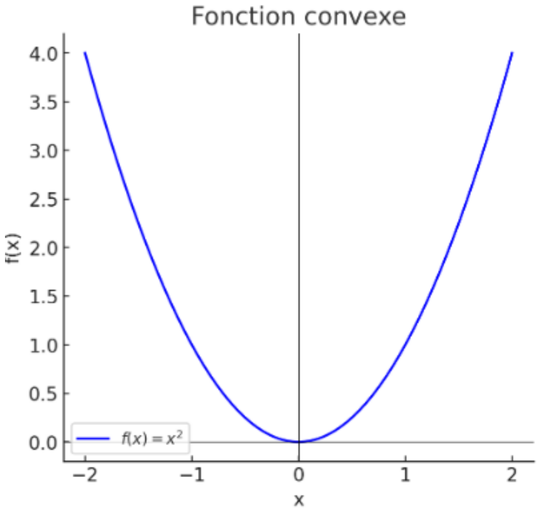
- **Matrice Définie +/-**
  - **Exemple de covariance.**
  - La matrice de covariance est donnée par :

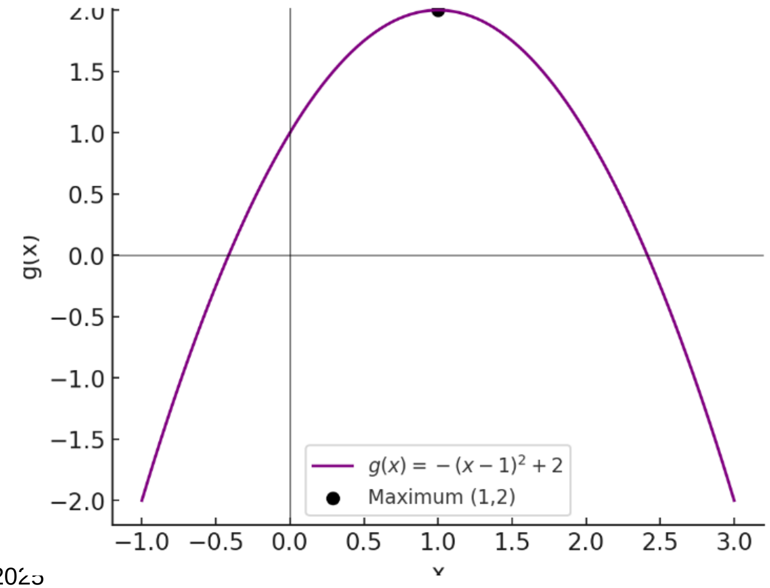
$$\begin{array}{cccc} & T & H & E \\ T & 25 & 25 & 125 \\ H & 25 & 100 & 100 \\ E & 125 & 100 & 633{,}333 \end{array}$$

  - On peut voir par exemple que Cov(T,H)=25 → Légère corrélation positive entre la température et l'humidité alors que Cov(T,E)=125 → forte corrélation positive entre la température et la consommation d'énergie.
  - Voir aussi la corrélation de Pearson, qui est normalisée pour être comprise entre -1 et 1.

Convexité

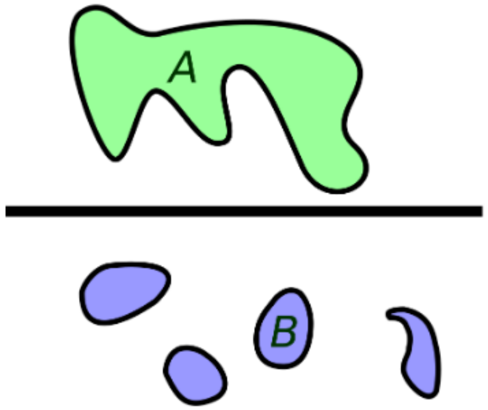
- Dans les problèmes d’optimisation, une notion joue un rôle très important : celle de convexité.
- En effet, pour la plupart des algorithmes, la convergence vers un optimum global ne pourra être démontrée qu’avec des hypothèses de convexité.





Convexité

- **Ensemble convexe**
  - **Définition.** Soit l' ensemble C$\subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, C est **convexe** si $\forall$ x, y $\in$ C, $\forall \lambda \in [0,1]$: $\lambda x+(1-\lambda)y \in C$.
  - D'un point de vue géométrique, un convexe est donc un ensemble qui, lorsqu'il contient deux points, contient nécessairement le segment les reliant.



A et B non connexe

Convexité

- **Exemple d’ensemble convexes.**
  - Un **disque** est convexe, car tout segment reliant deux points du disque reste à l’intérieur.
  - Un **triangle** est convexe, car tout segment joignant deux de ses points reste dans le triangle.
  - Une forme en **fer** à cheval n’est pas convexe, car il existe des segments reliant deux points de la forme qui sortent de l’ensemble.

Convexité

- **Ensemble convexe**

Théorème. L'image d'un connexe par une fonction continue est un connexe.

Proposition. Soient C, C_1, C_2 des convexes de $\mathbb{R}^n$, I, J $\subset \mathbb{R}$; on a les propriétés suivantes:

- Si et $\lambda_1, \lambda_2 \in \mathbb{R}$, alors $\lambda_1 C_1 + \lambda_2 C_2$ est un convexe de $\mathbb{R}^n$
- Si $(C_j)_{j\in J}$ est une famille quelconque de convexes de $\mathbb{R}^n$, alors $\cap_{j\in J} C_j$ est un convexe de $\mathbb{R}^n$.
- Si C est un convexe de $\mathbb{R}^n$ et f : $\mathbb{R}^n \rightarrow \mathbb{R}^m$ est une application affine de type f(x) = Ax + b pour $A \in \mathbb{R}^{mxn}$ et $b \in \mathbb{R}^m$, alors f(C) est un convexe de $\mathbb{R}^m$.

Convexité

- **Fonction convexe.**

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

  - $\forall\, x, y \in C, \forall\, \lambda \in [0,1]\ f(\lambda x+(1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$.
  - f est strictement convexe si on peut mettre l'inégalité stricte pour $\lambda \in]0, 1[$ et $x \neq y$.
  - Une fonction f est dite (strictement) concave si $-f$ est (strictement) convexe.

- L'interprétation géométrique de cette définition est que le graphe d'une fonction convexe est toujours en dessous du segment reliant les points $(x, f(x))$ et $(y, f(y))$.

Convexité

- **Fonction convexe.**

  Propriété. Si f : C$\subset \mathbb{R}^n$ $\rightarrow \mathbb{R}$ est deux fois continument dérivable sur C convexe, alors:

  - f est convexe si et seulement si $f''(X) \geq 0, \forall X \in C$
  - f est concave si $f''(X) \leq 0, \forall X \in C$.
  - f est strictement convexe si $f''(X) > 0, \forall X \in C$.

- **Exemple.** $f(x)=x^2$, $\mathbb{R}$ est convexe ,
  - f est 2 fois dérivable
  - $f''(x) = 2 > 0$, f est convexe

Convexité

- **Fonction convexe.**

  Théorème fondamental : Si une fonction f est **convexe** sur un ensemble ouvert convexe C$\subset \mathbb{R}^n$, alors :

  a) Si f admet en $X_0 \in C$ un **minimum local**, alors f admet en X_0 un minimum **global**.

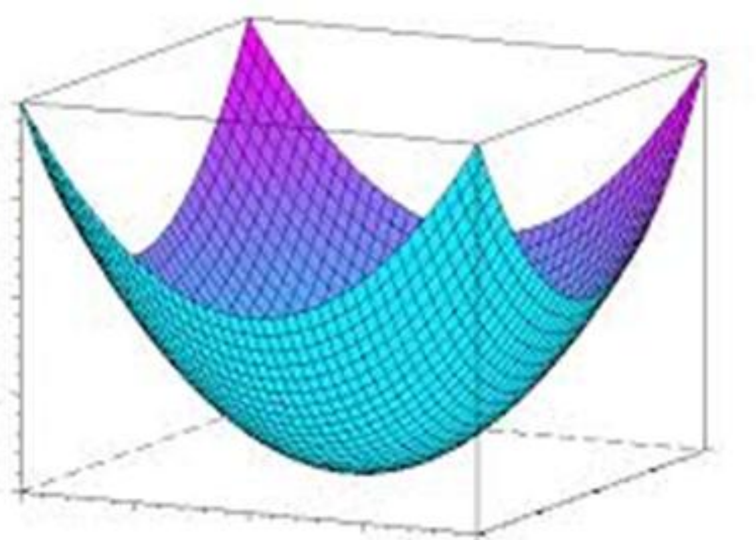
  b) Si f est de classe $\mathbf{C^1}$ sur C ; alors si $\nabla f(X_0) = 0$, alors X_0 est un minimum global sur C.

Extremums

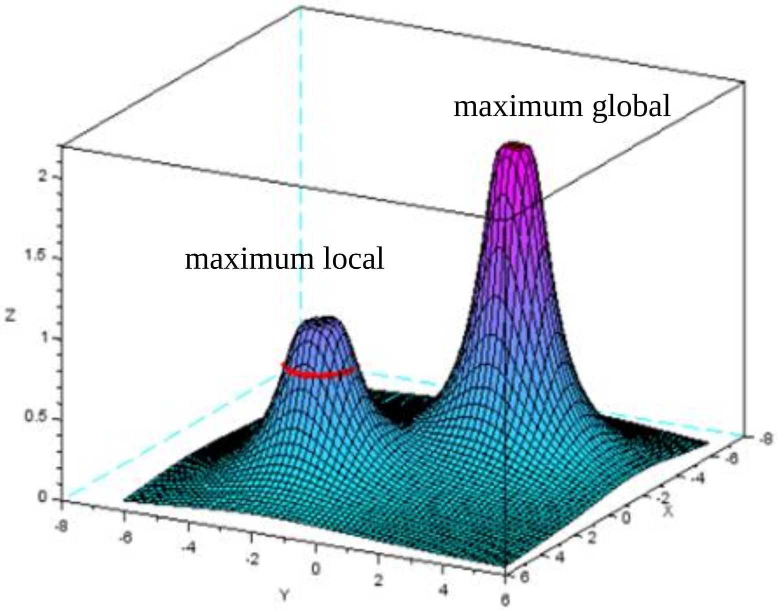
- **Définition des Extremums local/ global.** Soit f une fonction définie sur une partie U de $\mathbb{R}^n$ et à valeurs réelles, un point a $\in$U:
  - a est un minimum local (ou relatif) de f s'il existe un voisinage υ_a de a ouvert dans U tel que: $f(x) \geq f(a)$ pour tout $x \in \upsilon_a$.
  - a est un maximum local (ou relatif) de f s'il existe un voisinage Va de a ouvert dans U tel que: $f(x) \leq f(a)$ pour tout $x \in \upsilon_a$.
  - a est un extremum local si f admet un maximum local ou bien un minimum local en ce point.

Extremums

- **Définition des Extremums local/ global.** Un extremum est dit strict si l'inégalité est stricte, c'est à dire f(x) > f(a), pour tout $\boldsymbol{x} \neq \boldsymbol{a}$.



Minimum global



Extremums

- **Condition d’existence d’un extremum.** L’existence d’un extremum dépend énormément de la fonction en question et du domaine sur lequel on cherche cet optimum.
- Plusieurs théorèmes ont permettent de montrer leu existence sous certaines conditions.
- Ces théorèmes sont cruciaux en optimisation, car ils fournissent une condition **suffisante ou suffisante** pour l'existence d'une solution optimale pour de nombreux problèmes.

Extremums

Cas des fonctions d'une variable réelle.

- **Théorème de Fermat (condition nécessaire du premier ordre)**. Si une fonction *f* est dérivable en un point x_0 et que x_0 est un extremum local de *f*, alors le gradient de *f* en x_0 est nul : $\boldsymbol{\nabla f(x_0)=0}$.
- Autrement dit, si a est un extremum local alors c'est un point critique.
- La réciproque n'est pas toujours vraie. Par exemple, pour $f:x \rightarrow x^3$, le point a = 0 est un point critique, mais ce n'est ni un maximum local ni un minimum local (c'est un point d'inflexion).

Extremums

Cas des fonctions d'une variable réelle.

- **Théorème (Conditions suffisantes du second ordre)**. Soit f une fonction deux fois dérivable sur un intervalle ouvert I. Soit a un point de I tel que $f'(a) = 0$ (c'est-à-dire, a est un point critique), alors :
  - Si $f''(a) > 0$, alors f admet un minimum local strict en a.
  - Si $f''(a) < 0$, alors f admet un maximum local strict en a.
  - Si $f''(a) = 0$, alors on ne peut pas conclure à l'aide de ce théorème. Il faut approfondir l'étude (en utilisant des dérivées d'ordre supérieur ou d'autres méthodes).

Extremums

Cas des fonctions d'une variable réelle.

- **Théorème de Weierstrass**. Soient K un ensemble **compact** (fermé et borné) non vide de $\mathbb{R}^n$ et f : K → $\mathbb{R}$ une fonction **continue** sur K, alors f admet est bornée et atteint ses bornes sur K.
- Autrement il existe toujours un point dans cet intervalle où la fonction atteint sa valeur maximale, et un autre point où elle atteint sa valeur minimale.
- C'est à dire, il existe a* tel que : $f(a^*)=\min_{x \in K} f(x)$,

  ou $f(a*) = \max_{x \in K} f(x)$

Extremums

Cas des fonctions d'une variable réelle.

- **Théorème de Weierstrass**. Soient K un ensemble **compact** (fermé et borné) non vide de $\mathbb{R}^n$ et f : K → ℝ une fonction **continue** sur K, alors f admet est bornée et atteint ses bornes sur K.
- Autrement il existe toujours un point dans cet intervalle où la fonction atteint sa valeur maximale, et un autre point où elle atteint sa valeur minimale.
- C'est à dire, il existe a* tel que : f(a*)=$\min\limits_{x \in K} f(x)$, ou $f(a*) = \max\limits_{x \in K} f(x)$

Extremums

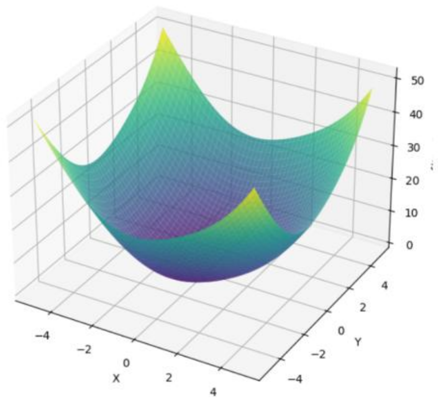
Cas des fonctions d'une variable réelle.

- **Théorème d'existence.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction **continue**. Si f est **coercive**, c'est-à-dire que $\lim_{||x|| \to \infty} f(x) = +\infty$, Alors f admet au moins un minimum global sur $\mathbb{R}^n$.
- **Théorème d'unicité.** Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction **continue**. Si f est **coercive** et strictement **convexe**, alors il existe un unique minimum $a* \in \mathbb{R}^n$ de f tel que : $\forall x \in \mathbb{R}^n$, $f(x) \geq f(a*)$.
- Pour les fonctions concaves, on recherche des maximas, tandis que pour les fonctions convexes, on recherche des minima, les inégalités sont inversées.

Extremums

Cas des fonctions d'une variable réelle.

- **Exemple.** f: $\mathbb{R}^2 \rightarrow \mathbb{R}$ définie par : $f(x, y) = x^2 + y^2 + \sin(x) + \sin(y)$
  - f est coersive
  - f est continue
  - Cette fonction est coercive et continue, donc d'après le théorème, elle admet un minimum global.
- Exercice : chercher ce minimum globale.



Extremums

Cas des fonctions de plusieurs variables : matrice Hessienne.

- On peut généraliser les résultats précédents aux fonctions de plusieurs variables afin de déterminer la nature locale des points critiques en utilisant la matrice **Hessienne**.

Extremums

Cas des fonctions de plusieurs variables.

- **Définition.** Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ une fonction de n variables. La matrice **Hessienne** de f en $x = (x_1, \ldots, x_n)$ est la matrice $n \times n$ des dérivées partielles d'ordre 2 de f notée aussi $\nabla^2 f(x)$:

$$\nabla^2 f(x) = \boldsymbol{Hf(x)} = \left(\frac{\boldsymbol{\partial^2 f}}{\boldsymbol{\partial x_i \partial x_j}}(x) \right)_{1 \leq i,j \leq n}$$

- La matrice hessienne fournit des informations sur la courbure de la fonction *f*.
- Cela est crucial pour déterminer si un point critique est un minimum, un maximum ou un point de selle.

Extremums

Cas des fonctions de plusieurs variables.

- Dans le cas d'une fonction de deux variables la matrice hessienne est noté $\nabla^2(f)(x)$ est une matrice contenant les dérivées partielles d'ordre 2 en x.
- $\nabla^2 f(x,y) = \mathbf{Hf}\,(x, y) = \begin{pmatrix} \frac{\partial^2 f(x,y)}{\partial x \partial x} & \frac{\partial^2 f(x,y)}{\partial y \partial x} \\ \frac{\partial^2 f(x,y)}{\partial x \partial y} & \frac{\partial^2 f(x,y)}{\partial y \partial y} \end{pmatrix}$
- Exemple: $f(x,y)=x^2+xy$, $\nabla^2 f(x,y) = \mathbf{Hf}\,(x, y) = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}$

Extremums

Cas des fonctions de plusieurs variables. hessienne de

- Soit f : $\mathbb{R}^n \to \mathbb{R}$. Sa hessienne est :

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

Extremums

Cas des fonctions de plusieurs variables :

- **Définition.** Soit $f : U \to \mathbb{R}$ une fonction de deux variables, où U est un ouvert de $\mathbb{R}^n$.
  - On dit que f admet un **maximum local** (resp. minimum local) en $X^* \in U$ s'il existe un voisinage ouvert $D \subset U$, centré en X^*, tel que : $\forall X \in D$ $f(X) \leq f(X^*)$ (resp. $f(X) \geq f(X^*)$).
  - On dit que f admet un extremum local en $f(X^*)$ si elle y admet un maximum local ou un minimum local.

Extremums

Cas des fonctions de plusieurs variables.

- **Proposition (condition nécessaire du premier ordre).** Si f : U $\subset \mathbb{R}^n \longrightarrow \mathbb{R}$ admet un extremum local en un point ($a_{0,\ldots}$, a_1), alors la condition nécessaire de premier ordre pour que f ait un extremum local à a est que le gradient de f au point a soit nul, c'est à dire :

$$\nabla f(a) = \left(\frac{\partial f}{\partial x1}(a_1, \cdots, a_n), \frac{\partial f}{\partial x2}(a_1, \cdots, a_n), \ldots, \frac{\partial f}{\partial xn}(a_1, \cdots, a_n)\right) = (0, \ldots, 0).$$

- Dans le cas de n=2, cela se traduit par :

$$\frac{\partial f}{\partial x}(a_0, a_1) = 0 \; et \; \frac{\partial f}{\partial y}(a_0, a_1) = 0.$$

Extremums

Cas des fonctions de plusieurs variables: hessienne

- **Théorème (Conditions suffisantes).** Soit f une fonction de classe C^2 sur un ouvert U, et soit $x = (x_1, \ldots, x_n) \in U$ un point critique de f c'est-à-dire que $\nabla f(x) = (0, \ldots, 0)$.
  - Si $\nabla^2 f(x)$ est **définie positive alors** f admet un **minimum strict local** en x.
  - Si $\nabla^2 f(x)$est **définie négative**, alors f admet un **maximum strict local** en x.
  - Si $\nabla^2 f(x)$ est **indéfinie** (c'est-à-dire qu'il existe des vecteurs v et w tels que $v^T Hf(x)\, v > 0$ et $z^T Hf(x) z < 0$), alors x est un **point selle** et il n'existe pas d'extremum local en ce point.

Extremums

Cas des fonctions de plusieurs variables: hessienne

- **Théorème (Conditions nécessaire).** Soit f une fonction de classe C^2 sur un ouvert U, et soit $x = (x_1, \ldots, x_n) \in U$ un point critique de f c'est-à-dire que $\nabla f(x) = (0, \ldots, 0)$.
  - Si x_0 est un minimum local, alors la matrice hessienne de f en x_0 est semi-définie positive : $\nabla^2 f(x_0) \geq 0$.
  - Si x_0 est un maximum local, alors la matrice hessienne de f en x_0 est semi-définie négative: $\nabla^2 f(x_0) \leq 0$.

Extremums

Cas des fonctions de plusieurs variables.

- **Théorème (Critère de Monge).** Soit f : $\mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction de classe C^2 et soit (x_0 , y_0) un point critique de f .
- On pose r = $\frac{\partial^2 f(x,y)}{\partial x \partial x}$s, t=$\frac{\partial^2 f(x,y)}{\partial y \partial y}$, s =$\frac{\partial^2 f(x,y)}{\partial x \partial y}$, ∇^2 f(x,y) **= Hf** (x, y) $=\begin{pmatrix} r & s \\ s & t \end{pmatrix}$
- Alors : Det(∇^2 f(x,y)) = rt − s^2
  - Si Det(∇^2 f(x,y)) > 0 et r > 0, alors (x_0 , y_0) est un **minimum** local de f ,
  - si Det(∇^2 f(x,y)) > 0 et r < 0, alors (x_0 , y_0) est un **maximum** local de f
  - • si Det(∇^2 f(x,y)) = 0, on ne peut pas conclure directement (il faut approfondir l'étude).

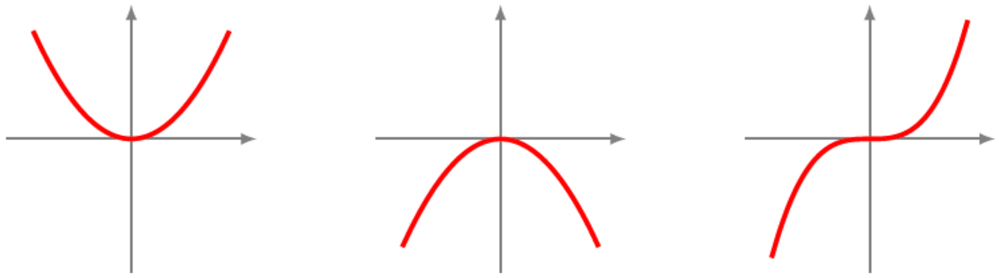
Extremums

- Exemples:

• $f : x \longrightarrow x^2$, a minimum local en 0, on a $f\,'(0) = 0$ et $f\,''(0) > 0$.

• $f : x \longrightarrow -x^2$, a maximum local en 0, on a $f\,'(0) = 0$ et $f\,''(0) < 0$.

• $f : x \longrightarrow x^3$, n'a ni minimum ni maximum local en 0, on a $f\,'(0) = 0$ et $f\,''(0) = 0$.

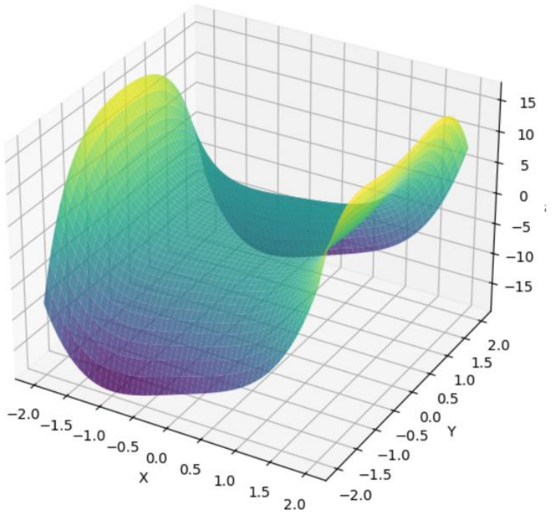


$f(x) = x^2$ $\quad$ $f(x) = -x^2$ $\quad$ $f(x) = x^3$

Extremums

- Exemples:

• $f : (x,y) \rightarrow f(x, y) = xy^2 + x^4 - y^4$, Supposons que f admet un extremum, donc



$$\begin{cases} \frac{\partial f(x,y)}{\partial x} = y^2 + 4x^3 = 0 \\ \frac{\partial f(x,y)}{\partial y} = 2xy - 4y^3 = 0 \end{cases}$$ (0,0) est le seul point critique

$$\nabla^2 f(x,y) = \begin{pmatrix} 12x^2 & 2y \\ 2y & 2x - 12y^2 \end{pmatrix} \quad \nabla^2 f(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

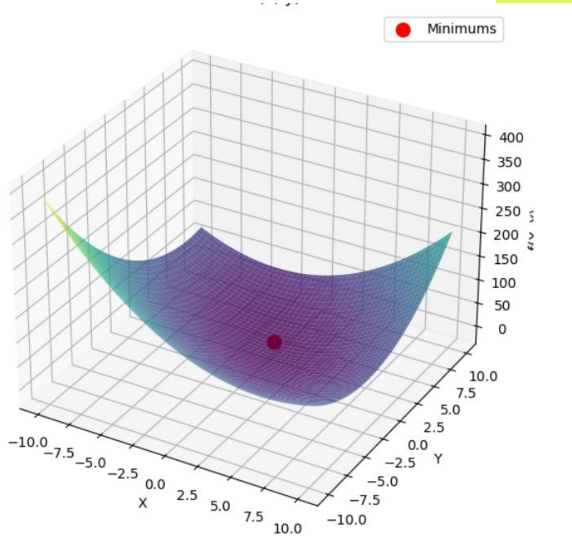
Det($\nabla^2 f(0,0)$)=0 donc (0,0) est un point de selle.

Extremums

- Exemples:

• $f : (x,y) \rightarrow$ f (x, y)= **f(x,y)=x^2+xy+y^2−3x−6y** , Supposons que f admet un extremum, donc

$$\begin{cases} \frac{\partial f(x,y)}{\partial x} = 2x+y-3, = 0 \\ \frac{\partial f(x,y)}{\partial y} = x + 2y - 6 = 0 \end{cases}$$ (0,3) est le seul point critique



$$\nabla^2 f(x,y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \quad \nabla^2 f(0,3) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$$

Det(∇^2f(0, 3))=3, La Hessienne est définie positive (courbure convexe). donc (0,3) est un point minimum local.

Extremums

Important.

- Un extremum est un point critique mais pas la réciproque.
- Un point critique qui n’est ni un maximum local ni un minimum local est nommé point-selle (ou point-col).
- Les points de selle sont importants en optimisation car ils peuvent **piéger** les algorithmes de recherche d'extrema.
- En effet, un algorithme peut converger vers un point de selle au lieu d'un minimum global.