programmation non linéaire

Pr A. Melit
Université de Jijel

Introduction

- **Formulation générale des problèmes d'optimisation non linéaire:**

La forme générale d'un problème d'optimisation est la suivante :

$$(PC)\quad \begin{cases} \underset{x\in R^n}{Min} \quad \mathrm{f}(x) \\ \quad sujet\,\text{à} : \\ \quad \mathrm{g(x)} \leq 0 \\ \quad h(x) = 0 \end{cases}$$

où les fonctions f, g et h sont typiquement non-linéaires.

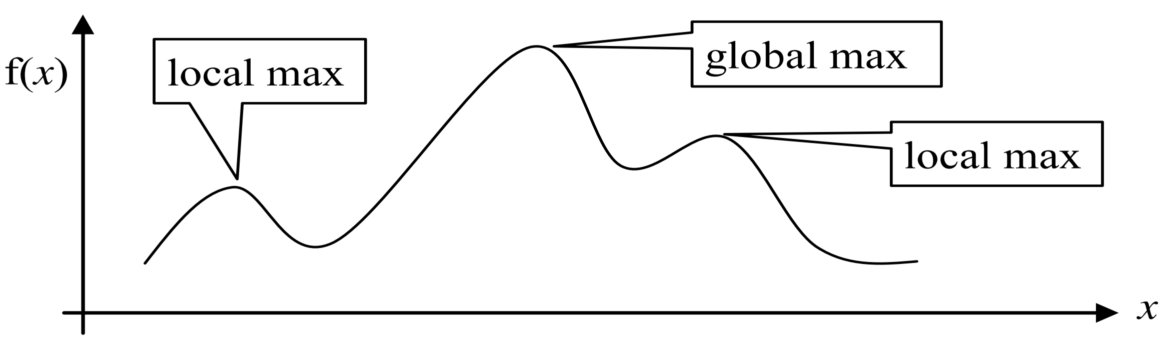
- L'objet de ce cours est la présentation de techniques permettant de résoudre le problème *(PC),* ainsi que des problèmes où soit un seul des deux types de contraintes est présent, soit des problèmes n'y a pas de contraintes du tout. Nous noterons ces types de problèmes ainsi :
  - *(PC) problème général, avec contraintes d'inégalité et d'égalité,*
  - *(PCE) problème avec contraintes d'égalité,*
  - *(PCI) problème avec contraintes d'inégalité,*
  - *(P) problème sans contraintes.*

Introduction

- Un programme non linéaire est similaire à un programme linéaire, car il est constitué:
  - D’une fonction objectif
  - Des contraintes et des bornes sur les variables
- Toutefois ces équations n’ont pas à être linéaires
- Elles sont généralement données comme des fonctions (analytiques)
- Il y a de nombreuses applications pratiques de la PNL:
  - Problème d’optimisation en finance
  - Problème d’optimisation en énergie
  - Problème d’optimisation en (pétro) chimie
  - …
- Un PNL est toutefois beaucoup plus difficile à résoudre qu’un PL.

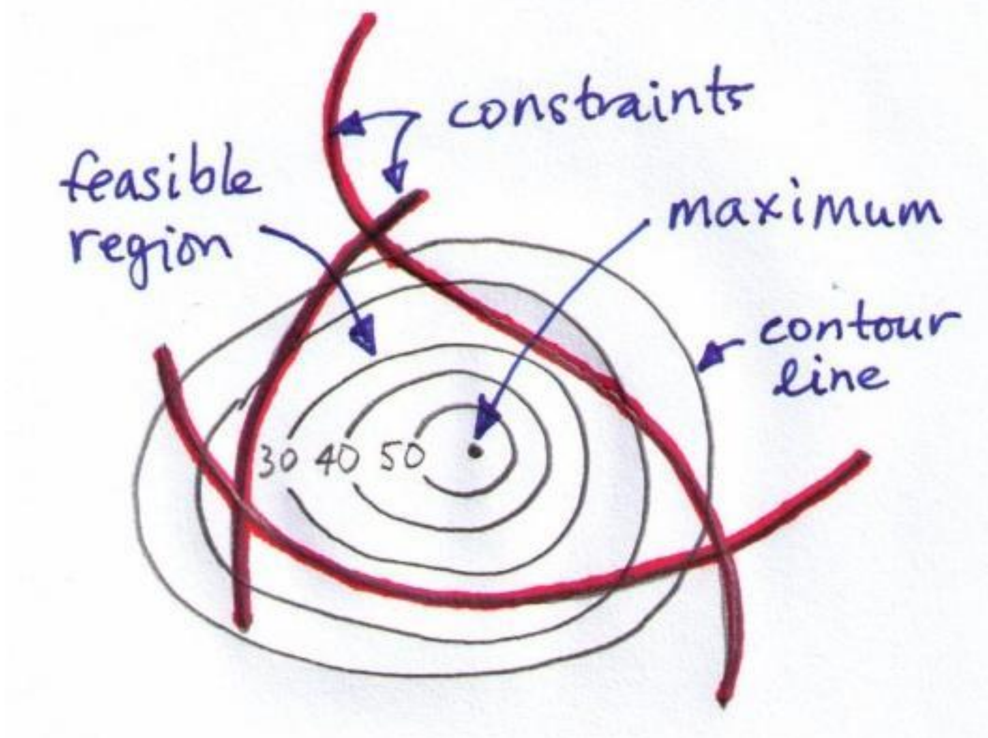
Introduction

- En l’absence de convexité (comme pour les PL) les méthodes de résolution de PNL ne peuvent garantir d’identifier la solution optimale
- Elles font généralement une recherche dans l’espace, basée sur:
  - Un point courant
  - La valeur de l’objectif à ce point
  - La valeur du gradient de la fonction objectif
  - La matrice Hessien (dérivée seconde) de la fonction objectif
- Ces informations permettent d’identifier un maximum (ou minimum) local, mais pas de dire s’il est global.



Introduction

- En PL, s'il existe une solution optimale, alors il existe un point extrême du polyèdre qui est optimal.
- En PNL, l'optimum peut se trouver n'importe où
  - En rouge les lignes de la région réalisables
  - En noir des courbes de niveau de la fonction objectif.



Introduction

- Comme on ne peut pas garantir l'obtention d'un optimum global, le point d'où on débute la recherche joue un rôle important sur la solution que nous pouvons trouver.

- En effet la plupart des solveurs non linéaires font une recherche locale
  - À partir d'une solution de départ ou solution courante
  - On détermine la meilleure direction (avec le gradient)
  - On fait un « pas » dans cette direction

- Une des solutions est de redémarrer le solveur régulièrement à partir de points différents.

Définitions

- On cherche à résoudre

$$\min_{x \in \mathbb{R}^n} \{f(x) : x \in \Omega\}$$

  avec $f : \mathbb{R}^n \to \mathbb{R}$ différentiable et $\Omega \subseteq \mathbb{R}^n$.

- Un point réalisable $x^* \in \Omega$ est un minimum global de la fonction f sur le domaine Ω si

$$f(x^*) \leq f(x) \text{ pout tout } x \in \Omega \text{ .}$$

- Un point réalisable $x^* \in \Omega$ est un minimum local de f sur Ω s'il existe $\varepsilon > 0$ tel que

$$f(x^*) \leq f(x) \text{ pout tout } x \in \Omega \cap \mathcal{B}_\varepsilon(x^*)$$

  avec $\mathcal{B}_\varepsilon(x^*) = \{x \in \mathbb{R}^n : \|x - x^*\| < \varepsilon\}$.

Définitions

Définition I:

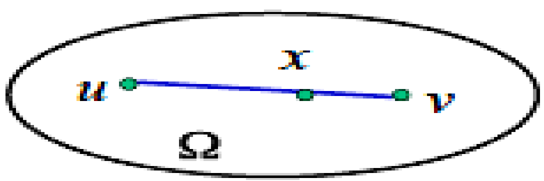
I.2.1 Ensemble convexe

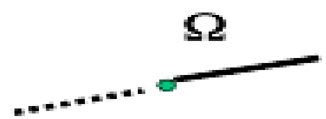
$\Omega \subset \mathbf{R}^n$ est convexe

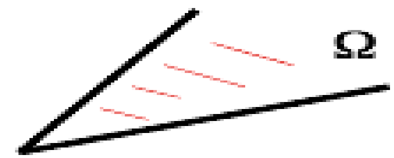
ssi $x = \alpha\, u + (1-\alpha) v \in \Omega \quad \forall u, v \in \Omega$ et $\alpha \in [0,1]$

Exemples:

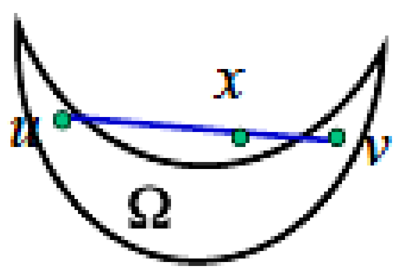
Ensembles convexes







Ensembles non convexes



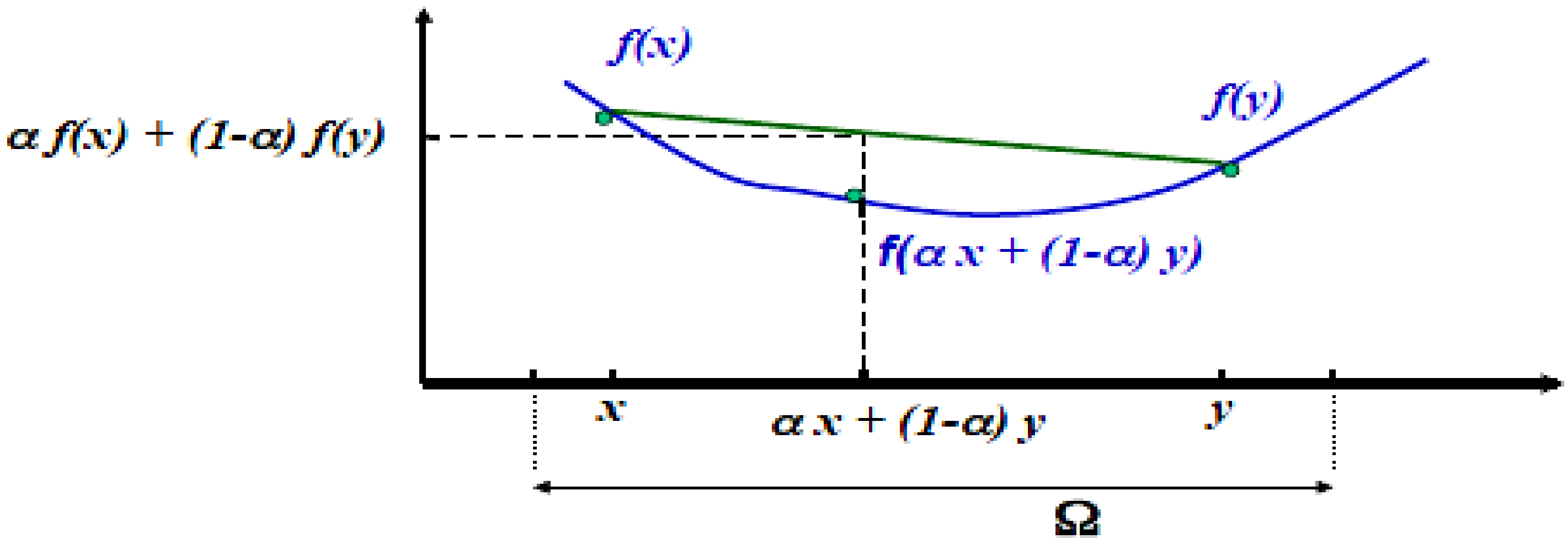
Définitions

Définition II

$$f : \Omega \rightarrow \mathbf{R} \qquad \Omega \subset \mathbf{R}^n \quad \text{convexe ssi}$$

$$\forall\, x, y \in \Omega \qquad \forall \alpha \in [0,1]$$

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



Définitions

Dérivées

- Gradient de f en $x \in \mathbb{R}^n$:

$$\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right) \in \mathbb{R}^n .$$

- Dérivée directionnelle de f en $x \in \mathbb{R}^n$ dans la direction unitaire $d \in \mathbb{R}^n$:

$$f'_d(x) = \lim_{t \downarrow 0} \frac{f(x+td) - f(x)}{t} = d^\top \nabla f(x) .$$

- Si les dérivées secondes de f existent et sont continues, alors la matrice Hessienne en x s'écrit

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

Définitions

- **Exemple:**

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

$$\nabla f = \left[15 - 3(x_3)^2 \quad 6(x_2)^2 \quad -6x_1x_3\right]$$

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

Définitions

Direction de descente

- $d \in \mathbb{R}^n$ est une direction de descente de f en $x \in \mathbb{R}^n$ si

$$f'_d(x) = d^\top \nabla f(x) < 0$$

- Pour tout $\alpha \in \mathbb{R}$ petit, on aura $h(\alpha) = f(x + \alpha d) < f(x)$ et $h'(0) < 0$.
- Principe de la *line search* (recherche linéaire) : Trouver α tel que $h'(\alpha) = 0$.

Définitions

Signe d'une matrice

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

- semi-définie positive si $y^\top A y \geq 0$ pour tout $y \in \mathbb{R}^n$;
- définie positive si $y^\top A y > 0$ pour tout $y \in \mathbb{R}^n \neq 0$;
- semi-définie négative si $y^\top A y \leq 0$ pour tout $y \in \mathbb{R}^n$;
- définie négative si $y^\top A y < 0$ pour tout $y \in \mathbb{R}^n \neq 0$;
- indéfinie sinon.

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

Définitions

- 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)

Définitions

- **Matrices définie positive**

$$\det\{\mathbf{A}_k\} > 0 \qquad k = 1, \cdots, n$$

$$\mathbf{A}_1 = a_{11} \quad \mathbf{A}_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \quad \mathbf{A}_3 = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} \quad \cdots \quad \mathbf{A}_n = \mathbf{A}$$

Définitions

Exemple: Soit la matrice A

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

Les valeurs propres de A sont:

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

Cette matrice est indéfinie.

Définitions

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:

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

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

Définitions

I.4.4 Contours d' une fonction

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

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

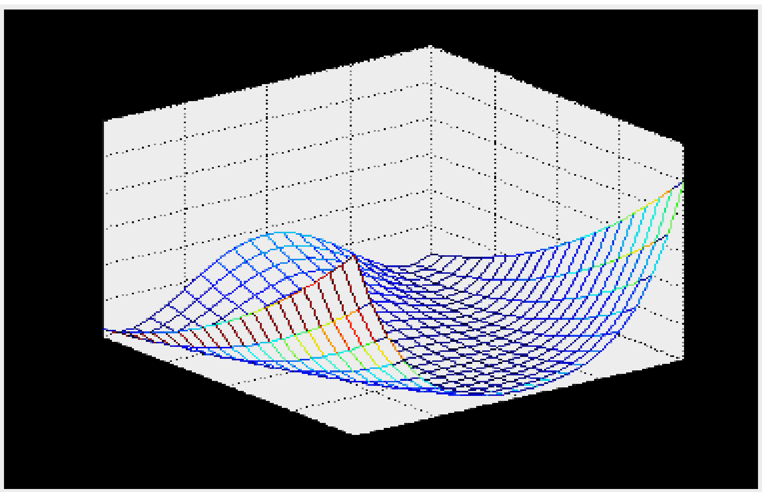
- $f(x) = c$, avec c constant

définissent des courbes de niveau ou des contours de cette surface

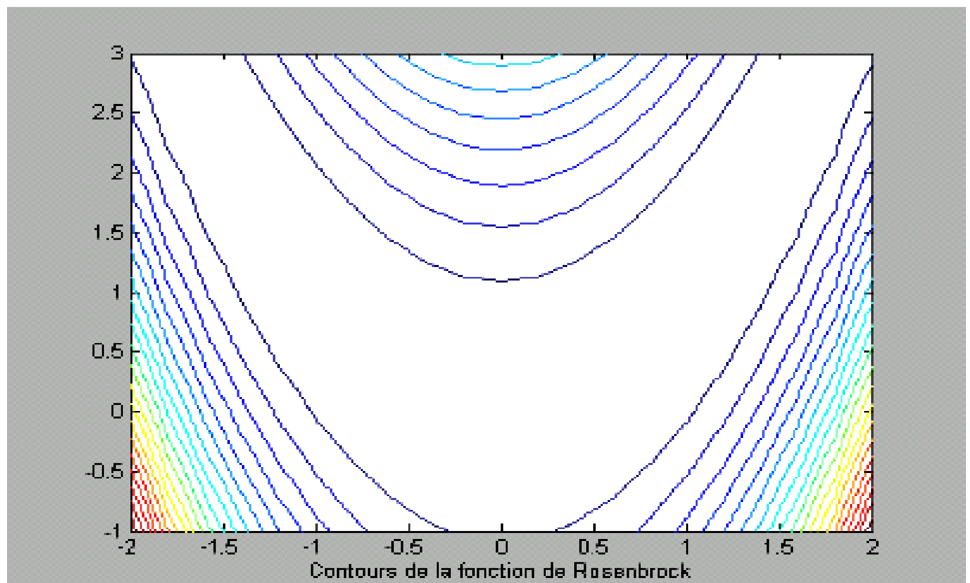
Un contour de niveau c est l' ensemble des points $S(c) = \{x \mid f(x) = c\}$

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

- 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 est "définie positive" alors le point est un minimum local.
   - Si la matrice est "définie négative" alors le point est un maximum local.

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.

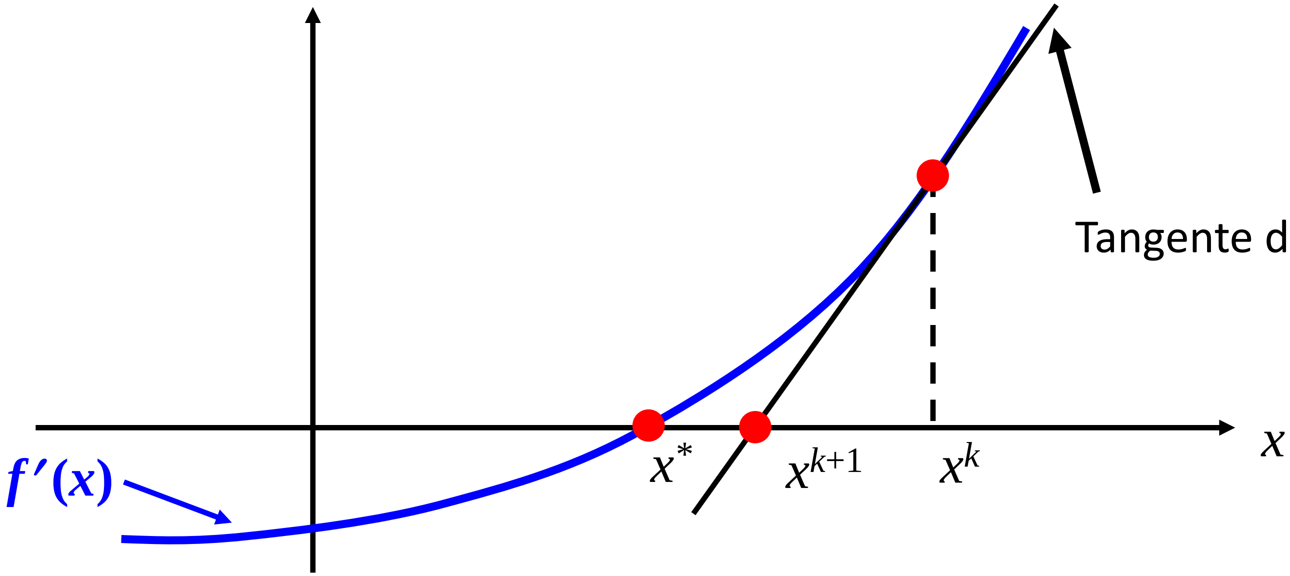
Optimisation non linéaire sans contraintes

1- Méthode de Newton: Lors de la résolution de $f'(x) = 0$ pour trouver un maximum ou un minimum, on peut faire utiliser les itérations suivantes:

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)}$$

- Avec x_0 donné.
- k est l'itération courante
- $|x^{k+1} - x^k| < \varepsilon$ où ε est une précision donnée.

Optimisation non linéaire sans contraintes



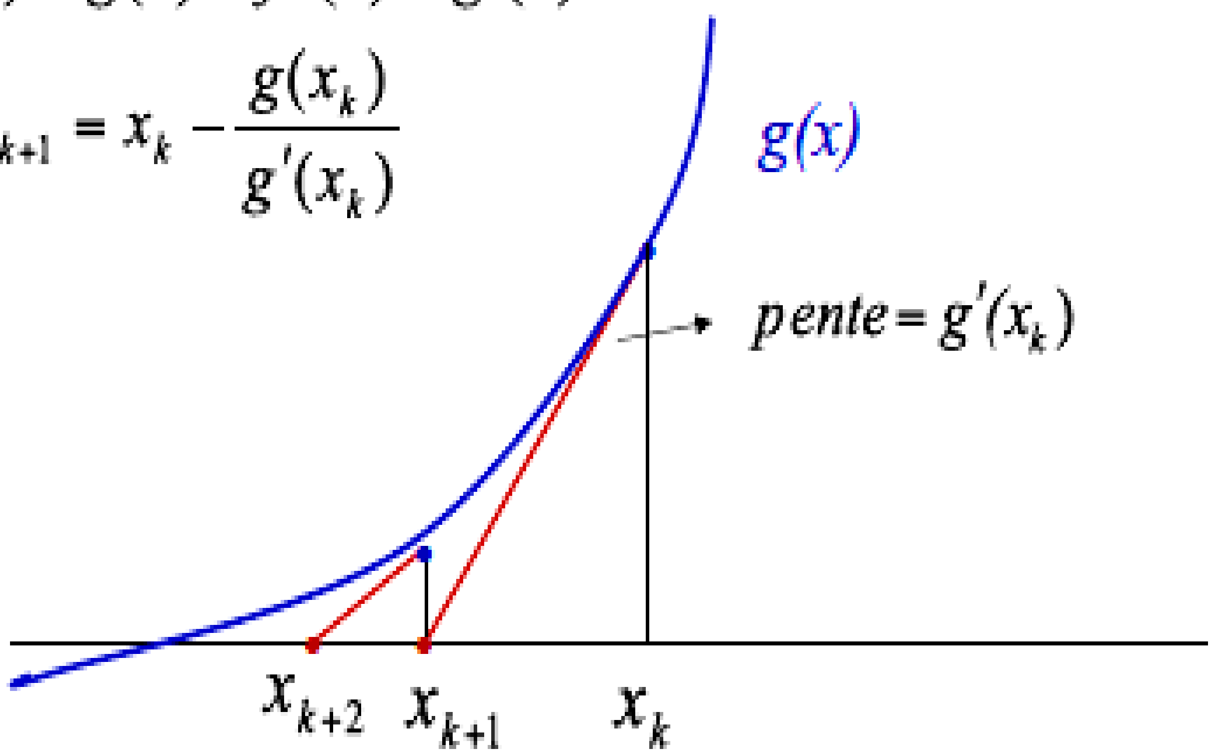
La méthode de Newton approxime $f'(x)$ avec une droite au point x^k et on obtient le point (x^{k+1}), qui est utilisé pour approximer la fonction au prochain point. Jusqu'à ce qu'on rapproche suffisamment de x^*.

Optimisation non linéaire sans contraintes

La méthode de Newton peut être vue comme une méthode de recherche de zéro d'une fonction $g(x)$ si on pose

$$f'(x) = g(x) \quad f''(x) = g'(x)$$

$$\Rightarrow x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$



Optimisation non linéaire sans contraintes

- Il faut s’assurer que
  - $f(x^{k+1}) < f(x^k)$ pour trouver un minimum.
  - $f(x^{k+1}) > f(x^k)$ pour trouver un maximum.
- Désavantage:
  - Il faut calculer les dérivées premières et secondes
  - La solution initiale est importante, si elle est trop loin de l’optimum la méthode peut ne pas converger.

Optimisation non linéaire sans contraintes

- **Exemple:** calculer la racine qui se trouve entre 0 et 5 de la fonction $f(x)=x^2-1$, en prenant x0 égal 4.

Iteration	x(k)	g(x)=x**2-1	g'(x)=2x	Erreur
0	4,00000000	1,5E+01	8,0000	3,0E+00
1	2,12500000	3,5E+00	4,2500	1,1E+00
2	1,29779412	6,8E-01	2,5956	3,0E-01
3	1,03416618	6,9E-02	2,0683	3,4E-02
4	1,00056438	1,1E-03	2,0011	5,6E-04
5	1,00000016	3,2E-07	2,0000	1,6E-07
6	1,00000000	2,5E-14	2,0000	1,3E-14

Optimisation non linéaire sans contraintes

- **La méthode de la sécante**

La dérivée n'est pas toujours disponible

=> approximation

$$f''(x_k) \cong \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}$$

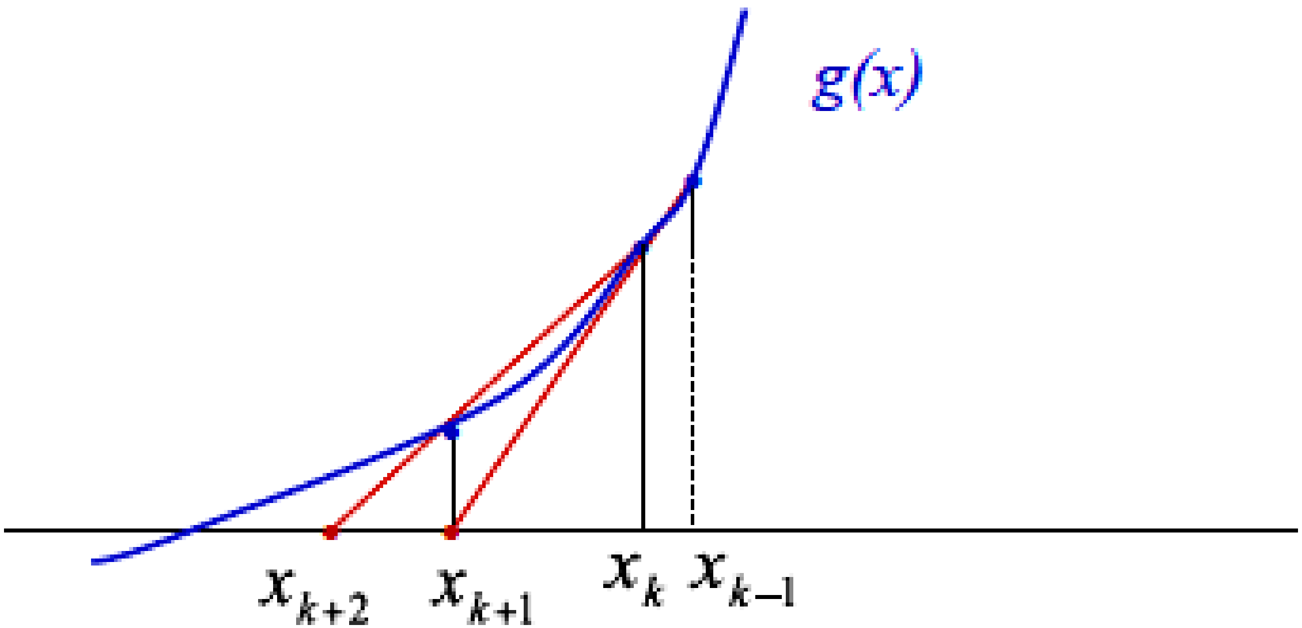
$$\Rightarrow x_{k+1} = x_k - \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} f'(x_k)$$

Cette méthode amène la dérivée de $f(x)$ à zéro

Optimisation non linéaire sans contraintes

Si $f'(x) = g(x)$ cette méthode trouve la racine de $g(x)$

$$x_{k+1} = x_k - \frac{x_k - x_{k-1}}{g(x_k) - g(x_{k-1})} g(x_k)$$



Optimisation non linéaire sans contraintes

- **Exemple:** calculer la racine qui se trouve entre 0 et 5 de la fonction $f(x)=x^2-1$, en prenant x_0 égal 4 et x_1= 5.

Iteration	x(k)	g(x)=x**2-1	dg/dx	Erreur
	5,00000000	2,4E+01		4,0E+00
0	4,00000000	1,5E+01	9,0000	3,0E+00
1	2,33333333	4,4E+00	6,3333	1,3E+00
2	1,63157895	1,7E+00	3,9649	6,3E-01
3	1,21238938	4,7E-01	2,8440	2,1E-01
4	1,04716672	9,7E-02	2,2596	4,7E-02
5	1,00443349	8,9E-03	2,0516	4,4E-03
6	1,00010193	2,0E-04	2,0045	1,0E-04
7	1,00000023	4,5E-07	2,0001	2,3E-07
8	1,00000000	2,3E-11	2,0000	1,1E-11

Optimisation non linéaire sans contraintes

- **La Méthode de regula-falsi:** Cette méthode nécessite deux points x^a & x^b qui encadre la solution de l'équation $f'(x) = 0$.

$$x^c = x^b - \frac{f'(x^b)\cdot(x^b - x^a)}{f'(x^b) - f'(x^a)}$$

où x^c sera entre x^a & x^b.

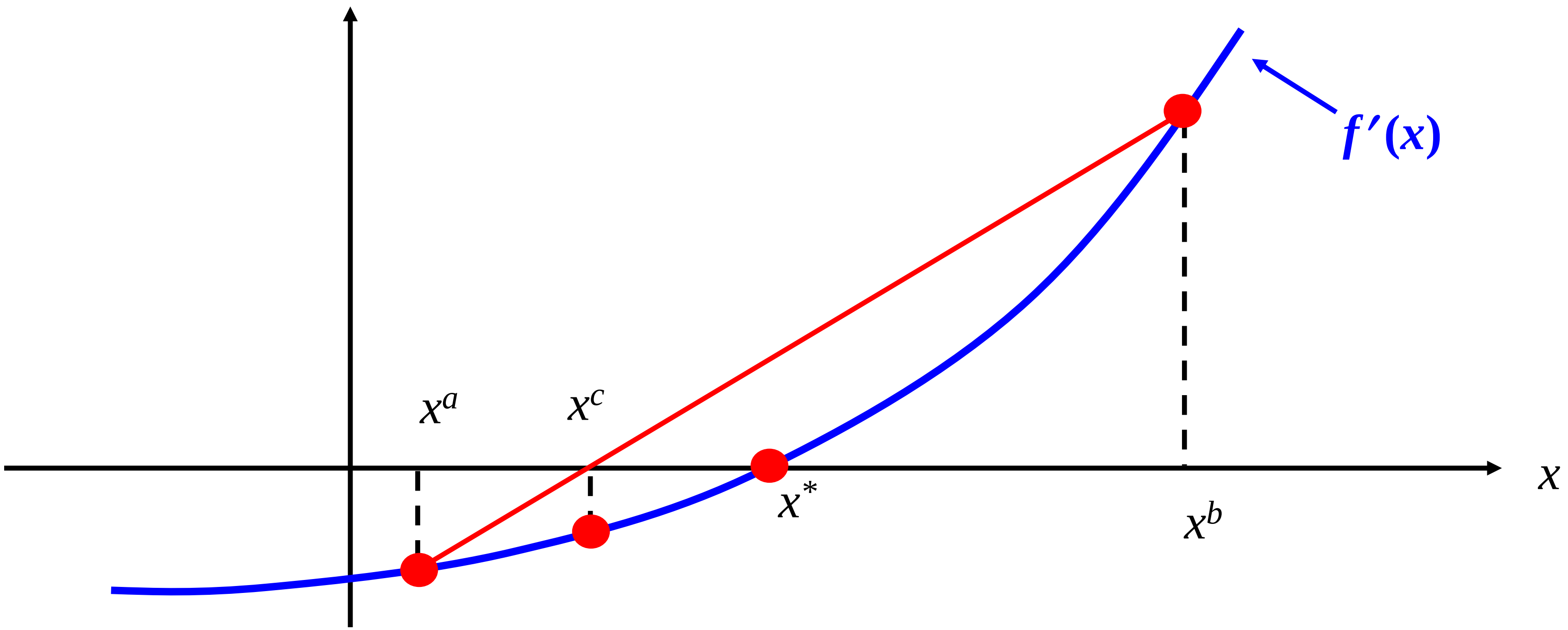
Le prochain intervalle sera défini par x^c et soit x^a ou x^b, celui dont le signe sera différent de x^c.

Optimisation non linéaire sans contraintes

- Si $f'(x)=g(x)$, cette méthode calcule la racine de $g(x)$.
- **Exemple:** calculer la racine qui se trouve entre 0 et 2 de la fonction $f(x)=x^2-1$, en prenant $x_a = 0$ et $x_b = 2$.

iteration	xc	xa	xb
0		0	2
1	0.5	0.5	2
2	0.8	0.8	2
3	0.9285714	0.9285714	2
4	0.9756098	0.9756098	2
5	0.9918033	0.9918033	2
6	0.990859	0.990859	2
7	0.9996952	0.9996952	2
8	0.9999984	0.9999984	2

Optimisation non linéaire sans contraintes



La méthode de Regula-Falsi approxime fonction $f'(x)$ avec une droite et utilise l'interpolation pour trouver la racine.

Optimisation Non linéaire sans contraintes

II- Cas d'une fonction à plusieurs variables: soit

$f : \mathbf{R}^n \to \mathbf{R} \quad f \in C^1$

Conditions nécessaires

$$x^* \text{ minimum local} \quad \Rightarrow \begin{cases} \nabla f(x^*) = 0 & \to \text{ordre 1 : point critique ou stationnaire} \\ \nabla^2 f(x^*) \geq 0 & \to \text{ordre 2 : hessien semi-défini positif} \end{cases}$$

- **Un point critique peut être un minimum, un maximum ou un point selle**

Conditions suffisantes

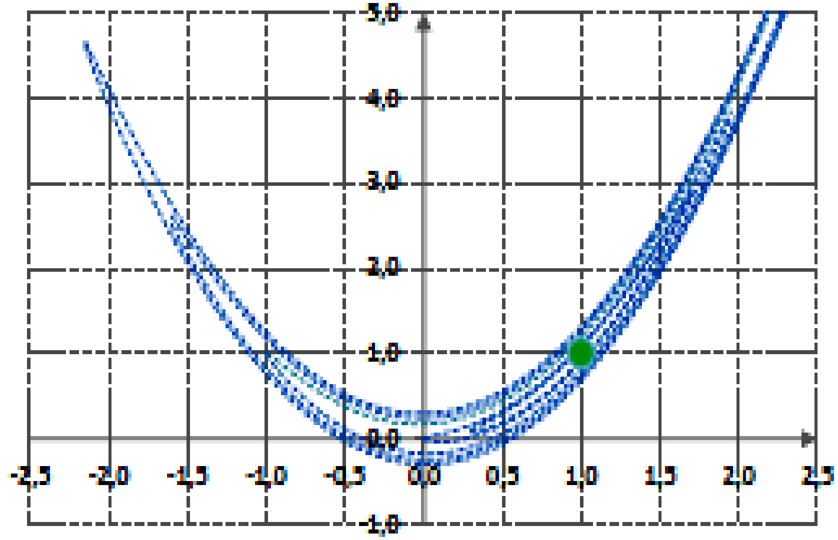
$$\begin{cases} \nabla f(x^*) = 0 & \to \text{ordre 1 : point critique ou stationnaire} \\ \nabla^2 f(x^*) > 0 & \to \text{ordre 2 : hessien défini positif} \end{cases} \quad \Rightarrow x^* \text{ minimum local}$$

Optimisation Non linéaire sans contraintes

Exemple 1

Fonction de Rosenbrock

$$f(x_1, x_2) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$



- Gradient :

$$\nabla f(x_1, x_2) = \begin{pmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ 200x_2 - 200x_1^2 \end{pmatrix}$$

- Hessien :

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

- **Point stationnaire** : $\nabla f(x_1, x_2) = 0 \Rightarrow \begin{cases} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 = 0 \\ 200x_2 - 200x_1^2 = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 1 \\ x_2 = x_1^2 \end{cases} \Rightarrow x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

- **Valeurs propres du hessien** : $\nabla^2 f(x^*) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} \rightarrow \begin{cases} \sigma_1 = 1001.60 \\ \sigma_2 = 0.39936 \end{cases}$

- Condition d'ordre 2 : $\nabla^2 f(x^*)$ est défini positif

- x^* vérifie les conditions suffisantes de minimum local (strict)

  x^* est un **minimum local** de f

Optimisation non linéaire sans contraintes

Exemple 2

Fonction : $f(x_1, x_2) = -x_1^4 - x_2^4$

- Gradient : $\nabla f(x_1, x_2) = \begin{pmatrix} -4x_1^3 \\ -4x_2^3 \end{pmatrix}$

- Hessien : $\nabla^2 f(x_1, x_2) = \begin{pmatrix} -12x_1^2 & 0 \\ 0 & -12x_2^2 \end{pmatrix}$

- Point stationnaire : $\nabla f(x_1, x_2) = 0 \Rightarrow \begin{cases} x_1 = 0 \\ x_2 = 0 \end{cases} \Rightarrow x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$

- Valeurs propres du hessien : $\nabla^2 f(x^*) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \rightarrow \begin{cases} \sigma_1 = 0 \\ \sigma_2 = 0 \end{cases}$
  $\nabla^2 f(x^*)$ est semi-défini positif

- x^* vérifie les conditions nécessaires de minimum local
  x^* ne vérifie pas les conditions suffisantes de minimum local

x^* est en fait un maximum local de f : $f(x_1, x_2) = -x_1^4 - x_2^4 \leq 0$
$\Rightarrow \forall (x_1, x_2),\ f(x_1, x_2) \leq f(0,0)$