Introduction à l'optimisation

MU4MEN01 : CALCUL SCIENTIFIQUE ET TNS – 3 ECTS

RESPONSABLE D'UE ET COURS : FLORENCE OSSART

Cours du 29/09/2021

→ Ch. 2 : Minimisation sans contrainte dans $\mathbb{R}^n$

- 2.1 Introduction
- 2.2 Approche analytique : les points critiques

Cours du 15/09

- 2.3 Approche numérique

Aujourd'hui

II. Minimisation sans contrainte dans $\mathbb{R}^n$

RAPPELS SUR LES FONCTIONS DE PLUSIEURS VARIABLES

II.1 Formulation du problème

Trouver x^* tel que :

$$J(x^*) = \inf \{ J(x), x \in A \}$$
$$A \subset \mathbb{R}^n$$

→ J : critère d'évaluation :

- Performance d'un jeu de variables de décision
- $$J : \begin{array}{l} A \longrightarrow \mathbb{R} \\ x \longmapsto J(x) \end{array}$$

→ x : variables de décision :

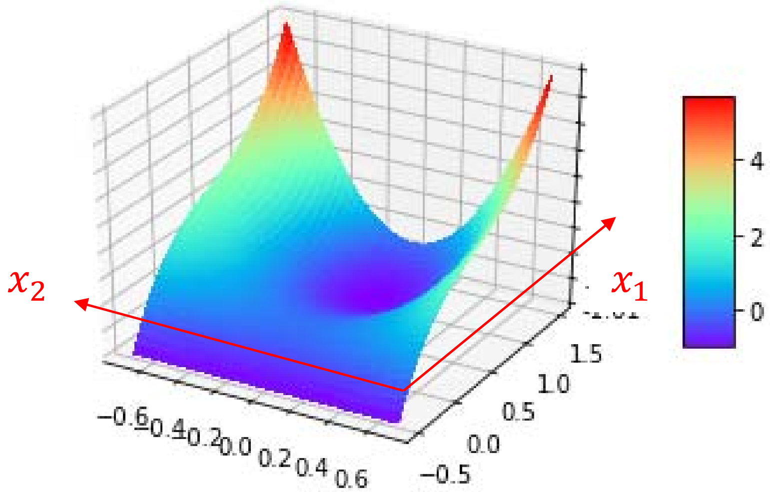
- Variables par rapport auxquelles se fait l'optimisation
- Notation : $x = (x_1, x_2, \dots, , x_n)$ $x \in A$
- A est le domaine de définition de J

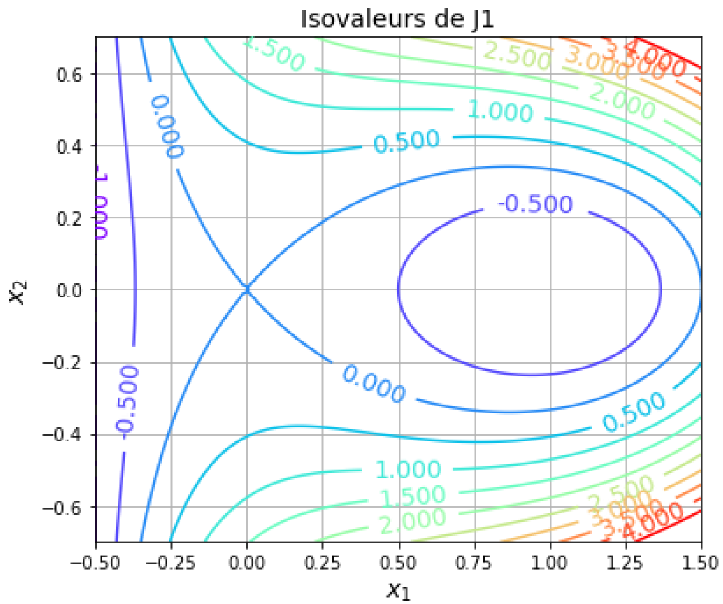
→ Il s'agit donc d'une « bête » recherche de minimum…

II.2.c Approche analytique : exercice

→ Soit la fonction $J(x_1, x_2) = 2x_1^3 - 3x_1^2 + 6\, x_1 x_2^2 + 3x_2^2$

→ D'après les représentations ci-dessous, combien y a-t-il de points critiques ? Faire le calcul.

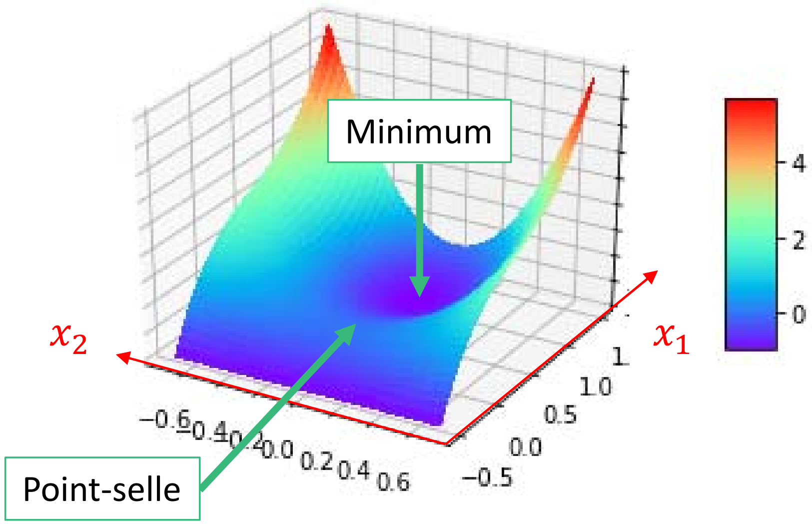


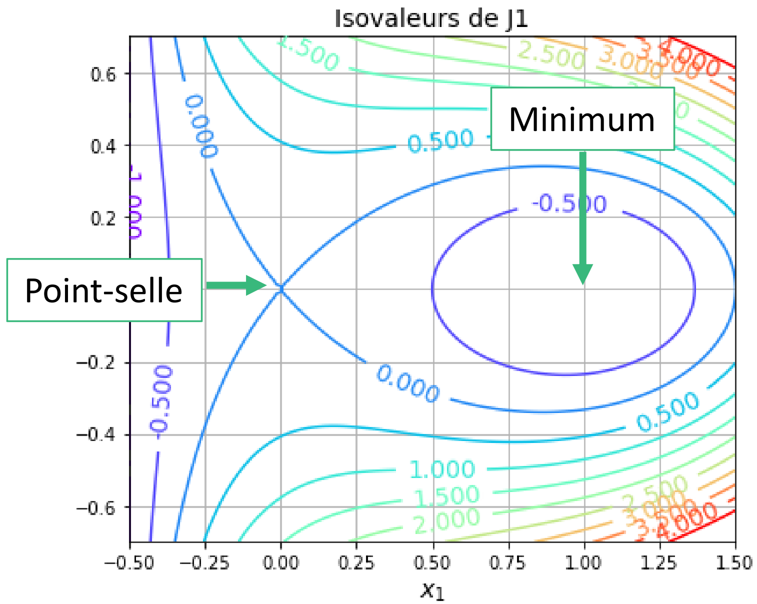


II.2.c Approche analytique : exercice

→ Soit la fonction $J(x_1, x_2) = 2x_1^3 - 3x_1^2 + 6\, x_1 x_2^2 + 3x_2^2$

→ D'après les représentations ci-dessous, combien y a-t-il de points critiques ? Faire le calcul.

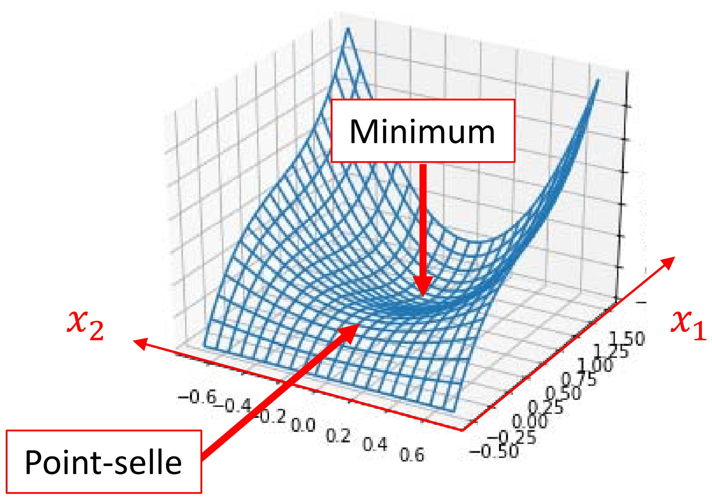




II.2 Approche analytique : exercice

→ Soit la fonction $J(x_1, x_2) = 2x_1^3 - 3x_1^2 + 6\,x_1 x_2^2 + 3x_2^2$

→ Calculs au tableau



II.2.d Rappel : à quoi ressemble un point-selle ?

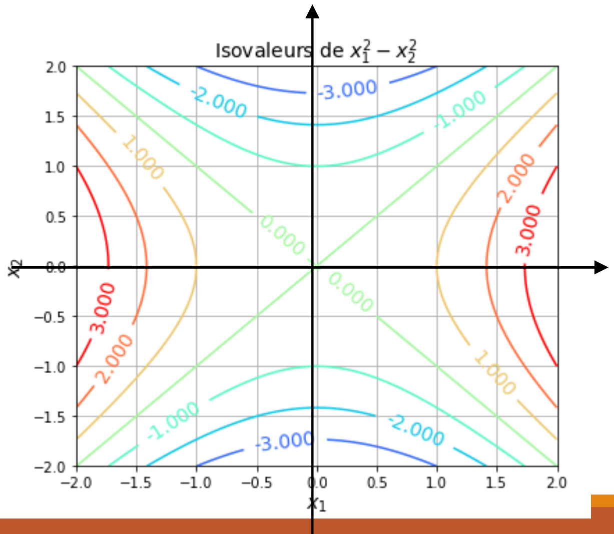
→ Exemple ultra basique : $J(x_1, x_2) = x_1^2 - x_2^2$

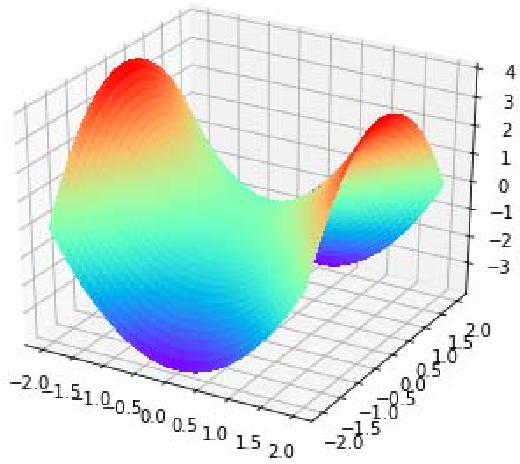
→ $\nabla J(x_1, x_2) = 2\begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$ => unique point critique $(0{,}0)$

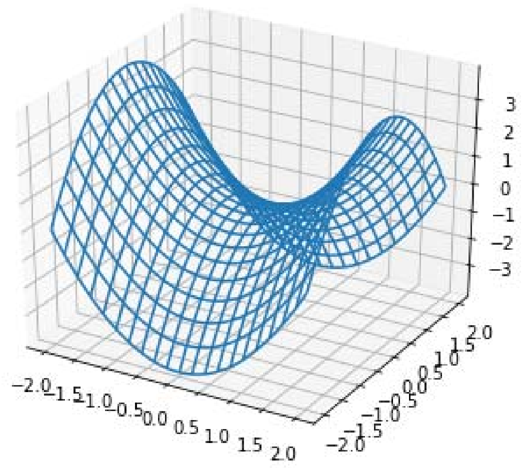
→ $H_J(x) = 2\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ => $(0{,}0)$ est un point-selle

Commentaire sur les représentations ci-dessous :

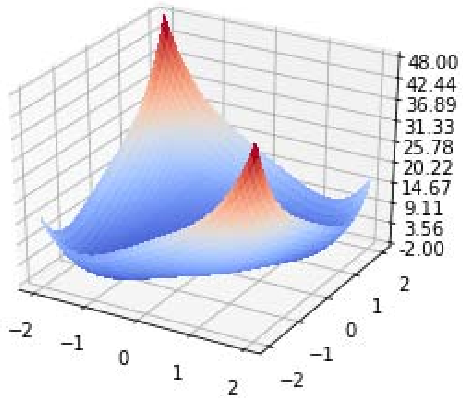
- *La représentation par surface montre la forme du point-selle*
- *La représentation par isovaleurs permet de positionner plus précisément le point-selle*







II.2.e Quelques résultats d'existence

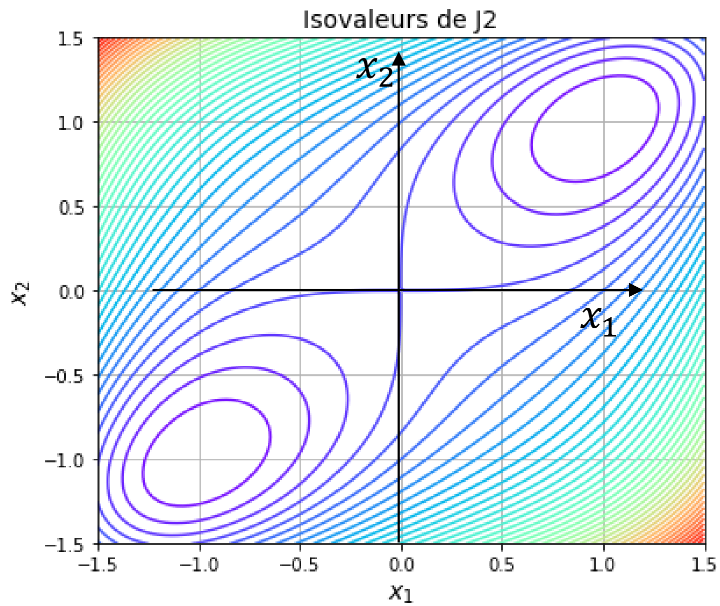


→ Si J est une fonction continue « coercive » :

- Définition : J est coercive ssi : $\lim_{\|x\| \to \infty} J(x) = \infty$
- Alors il existe au moins 1 minimum global (et éventuellement des minima locaux)

→ Exemple :

- $J(x_1, x_2) = {x_1}^4 - 4\, x_1 x_2 + {x_2}^4$
- J continue et coercive => au moins un minimum global
- D'après les isovaleurs ci-contre, combien de points critiques ?
- Faire le calcul.



II. Minimisation sans contrainte dans $\mathbb{R}^n$

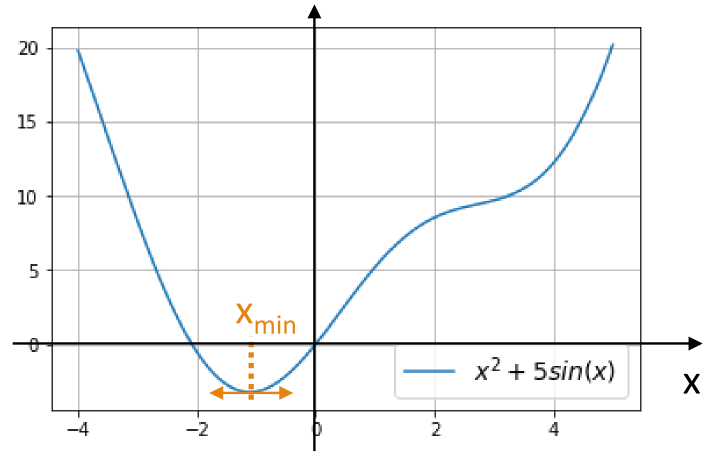
II.3 QUELQUES MÉTHODES NUMÉRIQUES

POUR COMMENCER : FONCTION D'UNE VARIABLE

II.3 Approche numérique

→ Exemple tout simple d'une fonction d'une variable :

$$f(x) = x^2 + 5\sin(x)$$



→ D'après le graphe de la fonction, il existe un minimum

→ Dérivée : $f'(x) = 2x + 5\cos(x)$

$$\boxed{f'(x_{min}) = 0 \Rightarrow 2x_{min} = -5\cos(x_{min})}$$

→ Pour déterminer x_{min}, il faut résoudre une équation dite « transcendentale » dont on ne sait pas exprimer analytiquement la solution.

→ Il faut donc mettre en œuvre une méthode numérique, par approximations successives.

II.3.a Méthode de Newton

→ Résolution numérique de $f'(x) = 0$:

→ Construction d'une suite d'approximations successives de la solution, notées x_n

→ A l'itération n, on dispose d'une solution approchée x_n

→ On approxime f' par $\widetilde{f}'$, sa tangente au point x_n : $\widetilde{f}'(x) = f'(x_n) + f''(x_n)(x - x_n)$

→ La solution approchée à l'itération $n + 1$ est obtenue en résolvant $\widetilde{f}'(x_{n+1}) = 0$

→ On obtient ainsi $x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$ (à condition que$f''(x_n) \neq 0$, bien sûr)

→ Algorithme simple, qui converge bien pour des fonctions convexes, et en particulier au voisinage de la solution. Mais attention si la fonction n'est pas convexe !

II.3.a Méthode de Newton, bis

→ Autre manière de voir les choses : approximer f' par sa tangente au point x_n est équivalent à approximer f par son développement de Taylor d'ordre 2

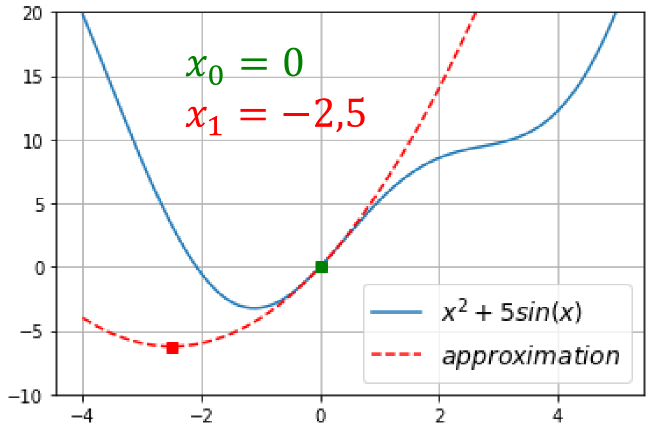
→ Au point x_n : $\tilde{f}(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2}f''(x_n)(x - x_n)^2$

→ **Exemple** : supposons qu'à l'itération 0, on part de $x_0 = 0$

→ ----------- Approximation parabolique

→ Minimum de la parabole : situé en $x_1 = -2.5$

→ On réapplique l'algorithme à partir de $x_1 = -2.5$



II.3.a Méthode de Newton, bis

→ Autre manière de voir les choses : approximer f' par sa tangente au point x_n est équivalent à approximer f par son développement de Taylor d'ordre 2

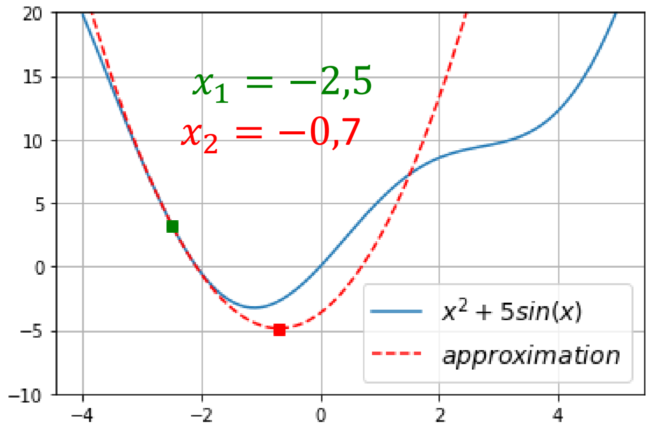
→ Au point x_n : $\tilde{f}(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{1}{2}f''(x_n)(x - x_n)^2$

→ A l'itération 1, x_1 =-2,5 (point vert)

→ ----------- Approximation parabolique

→ Minimum de la parabole situé en $x_2 = -0.7$

→ On réapplique l'algorithme à partir de $x_2 = -0.7$

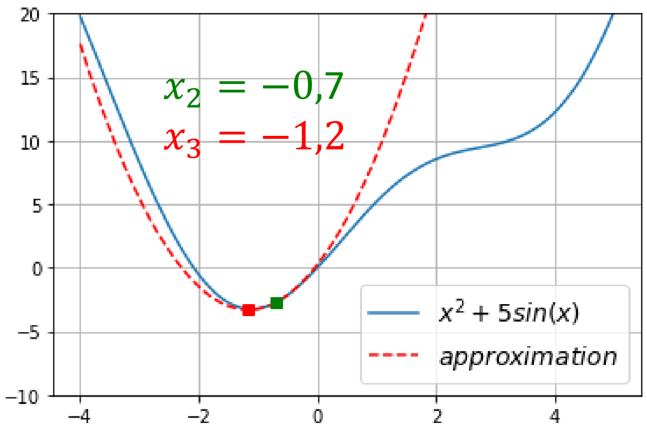


II.3.a Méthode de Newton, bis

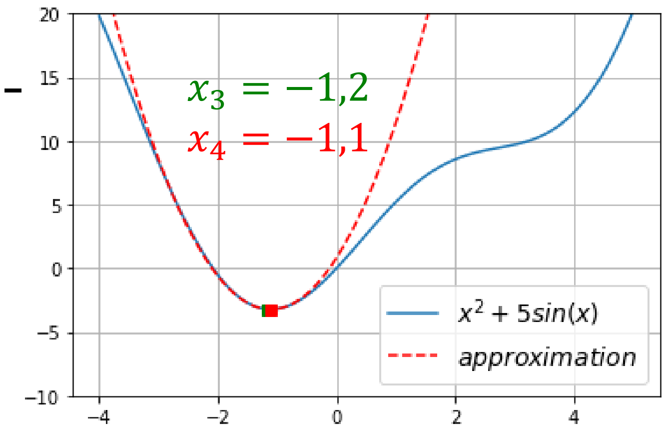
→ Les figures ci-dessous montrent les itérations suivantes

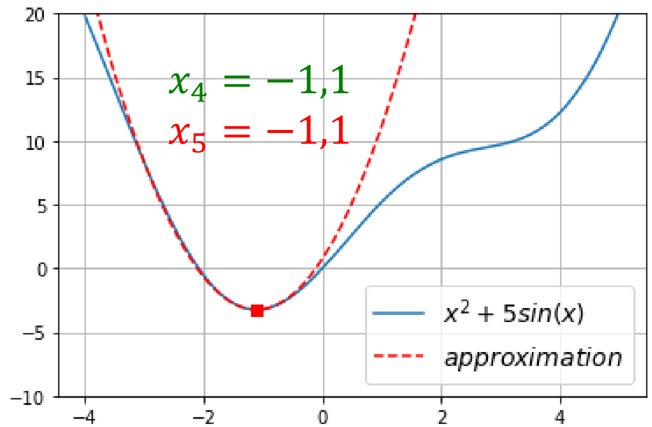
→ Le minimum de la parabole se rapproche progressivement du minimum de la fonction

→ Au voisinage du minimum, la parabole est quasiment confondue avec la courbe.



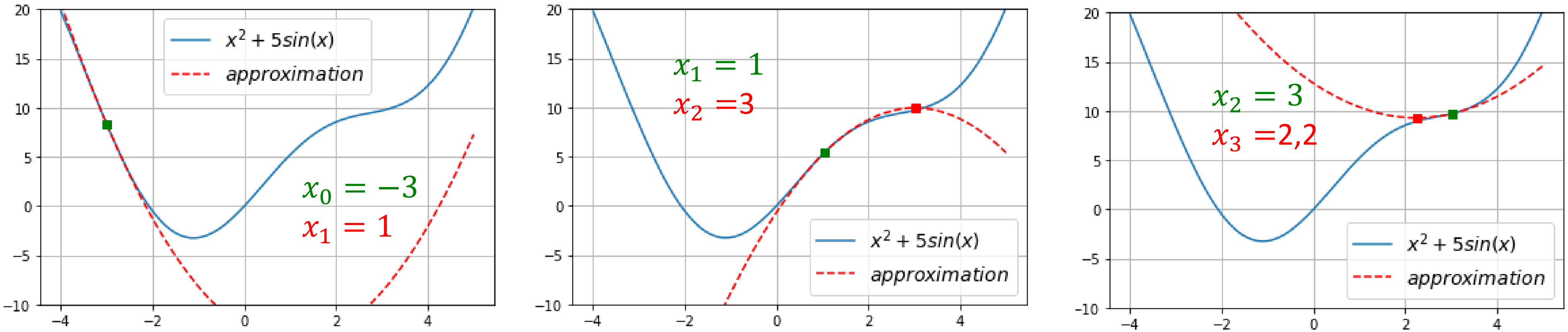
1 -





II.3.a Méthode de Newton, bis

→ Que se passe-t-il si on part d'un autre point ? $x_0 = -3$, par exemple



Nb : en x_1, la parabole est très éloignée de f

Nb : la parabole n'est pas dans le bon sens et x_2 correspond à un maximum !

Nb : la convergence parait mal partie… mais ça finira par converger

→ Attention : si la fonction n'est pas convexe, l'approximation parabolique peut être très mauvaise

II.3.a Méthode de Newton, ter

Bilan de la méthode de Newton :

→ Très efficace au voisinage de la solution

→ Très efficace pour des fonctions convexes, convergence garantie

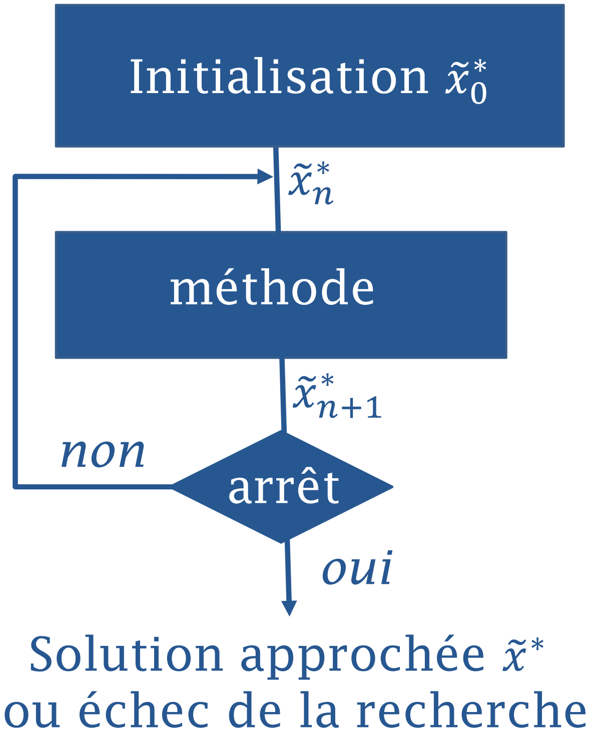
→ Danger s'il y a des points d'inflexion

→ A appliquer à partir d'un point initial bien choisi

II.3.b Structure générale d'une recherche numérique

→ Recherche itérative, par *approximations successives*

→Schéma général :



Le **critère d'arrêt** regroupe deux conditions :

1. $|x_{n+1} - x_n| \leq \varepsilon$, avec $\varepsilon = 10^{-6}$, par exemple

Précision attendue

2. Nombre maximal d'itérations atteint.

Arrêt de la recherche si ça ne converge pas

II.3.c Recherche par dichotomie, pourquoi ?

→ La méthode de Newton est puissante, mais *elle nécessite d'avoir accès aux dérivées d'ordre 1 et 2 de la fonction à minimiser*. Ces dérivées peuvent ne pas être disponibles.

→ Pour être efficace, la méthode de Newton doit être appliquée à partir d'un point relativement proche de la solution, ou au moins dans une région où la fonction est localement convexe.

→ Si ces conditions ne sont pas satisfaites, on peut utiliser une méthode de recherche par dichotomie, basique et peu rapide, mais plus sûre.

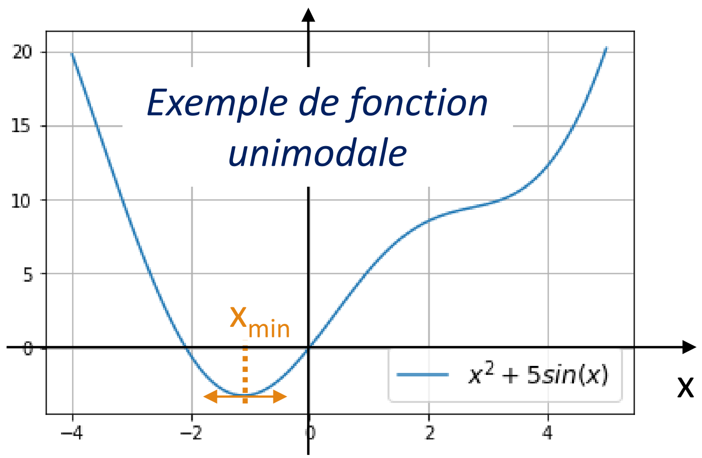
→ On peut combiner recherche par dichotomie pour réduire l'intervalle de recherche de la solution, puis recherche par Newton quand on n'est plus très éloigné de la solution.

II.3.c Recherche par dichotomie, algorithme

→ Cette méthode de recherche part d'un intervalle sur lequel la fonction est « *unimodale* » et présente un minimum unique.

→ La fonction f est *unimodale* sur $[a, b]$ si elle dérivable, et que sa dérivée s'annule et change de signe en un point unique de l'intervalle, noté c.

→ La méthode réduit progressivement la taille de cet intervalle par un processus itératif pour avoir un encadrement de la position du minimum de plus en plus précis.



II.3.c Recherche par dichotomie, algorithme

→ Notations :

- $[x_{1,n-1}, x_{5,n-1}]$: intervalle au *début* de l'itération n, avec $x_{1,n-1} < x_{5,n-1}$
- $[x_{1,n}, x_{5,n}]$: intervalle à la *fin* de l'itération n, avec $x_{1,n} < x_{5,n}$.

→ Procédure $dichotomie$: réduit la taille de l'intervalle

- $[x_{1,n}, x_{5,n}] = dichotomie([x_{1,n-1}, x_{5,n-1}])$,
- avec $(x_{5,n} - x_{1,n}) < (x_{5,n-1} - x_{1,n-1})$

→ Appel de $dichotomie$ tant que $(x_{5,n} - x_{1,n}) > \varepsilon$.

II.3.c Recherche par dichotomie, algorithme

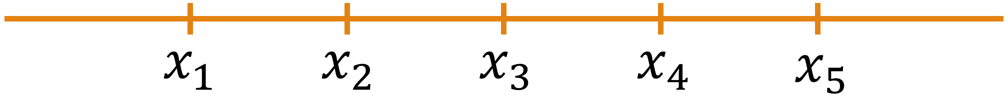
→ Algorithme :

- Choix de l'intervalle initial $[x_{1,0}, x_{5,0}]$
- Initialisation : $[x_{1,n-1}, x_{5,n-1}] \leftarrow [x_{1,0}, x_{5,0}]$
- Tant que $(x_{5,n-1} - x_{1,n-1}) > \varepsilon$:
  - $[x_{1,n}, x_{5,n}] \leftarrow dichotomie([x_{1,n-1}, x_{5,n-1}])$
  - $[x_{1,n-1}, x_{5,n-1}] \leftarrow [x_{1,n}, x_{5,n}]$

Principe de la procédure dichotomie

→ Hypothèse : f *unimodale* dans $[x_1, x_5]$ => *minimum unique* dans $[x_1, x_5]$.

→ Découpage de $[x_1, x_5]$ en 4 sous-intervalles de largeurs identiques.



$x_3 = \frac{1}{2}(x_1 + x_5),\ x_2 = \frac{1}{2}(x_1 + x_3)$ et $x_4 = \frac{1}{2}(x_3 + x_5)$

→ Calcul de $f_i = f(x_i)$ pour $i = 1, \dots, 5$

→ Comparaison de f_1, f_2, f_3, f_4 et f_5 pour éliminer les intervalles qui ne contiennent pas le minimum => nouvel intervalle

Principe de la procédure dichotomie

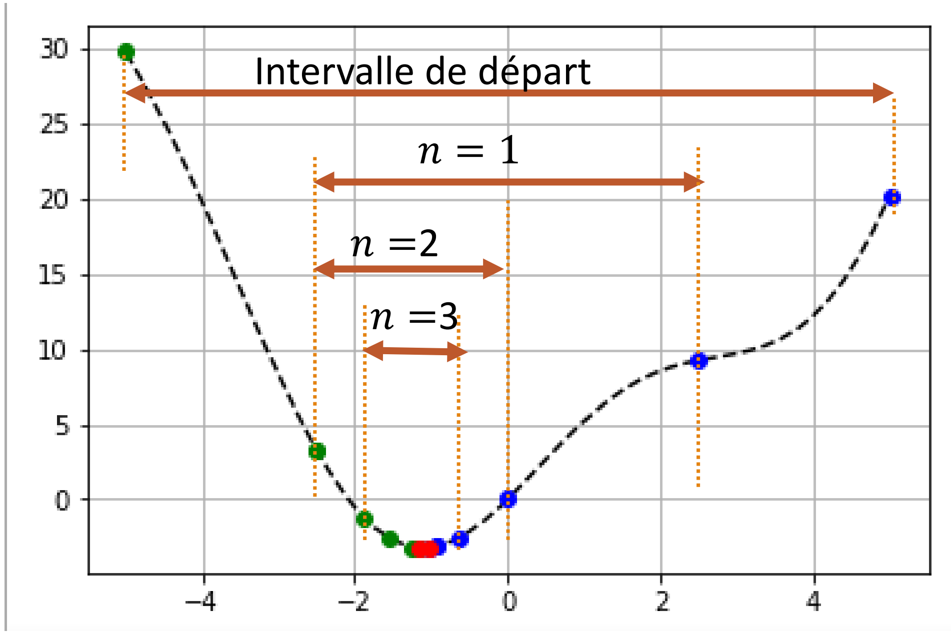
→ Cinq cas de figure peuvent se présenter :

- Si $f_1 < f_2 < f_3 < f_4 < f_5$: le minimum est dans l'intervalle $[x_1, x_2]$
- Si $f_1 > f_2 < f_3 < f_4 < f_5$: le minimum est dans l'intervalle $[x_1, x_3]$
- Si $f_1 > f_2 > f_3 < f_4 < f_5$: le minimum est dans l'intervalle $[x_2, x_4]$
- Si $f_1 > f_2 > f_3 > f_4 < f_5$: le minimum est dans l'intervalle $[x_3, x_5]$
- Si $f_1 > f_2 > f_3 > f_4 > f_5$: le minimum est dans l'intervalle $[x_4, x_5]$

→ On obtient ainsi un nouvel intervalle dont la largeur a été divisée par 4 dans les cas 1 et 5, et par 2 dans les cas 2, 3 et 4.

Exemple d'exécution de la dichotomie

→ Réduction de l'intervalle contenant le minimum :




```
Intervalle de départ :  -5  < X* <  5
 n=1 Intervalle :  -2.5  < X* <  2.5
 n=2 Intervalle :  -2.5  < X* <  0.0
 n=3 Intervalle :  -1.875  < X* <  -0.625
 n=4 Intervalle :  -1.5625  < X* <  -0.9375
 n=5 Intervalle :  -1.25  < X* <  -0.9375
 n=6 Intervalle :  -1.171875  < X* <  -1.015625
```

II. Minimisation sans contrainte dans $\mathbb{R}^n$

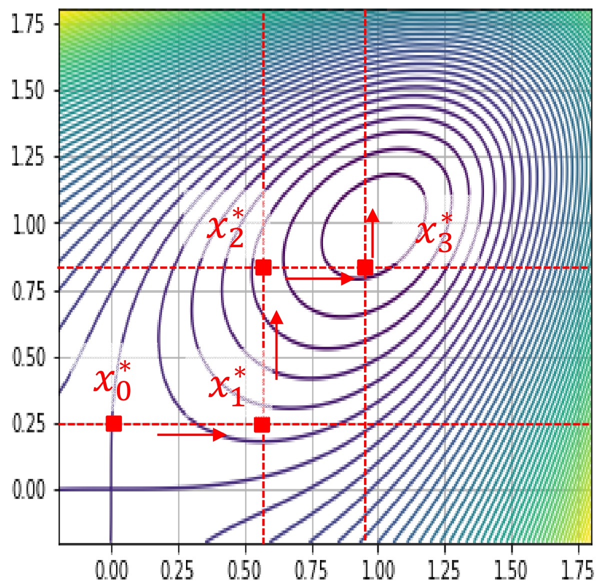
II.3 QUELQUES MÉTHODES NUMÉRIQUES

FONCTION DE PLUSIEURS VARIABLES

II.3.d Méthode de relaxation

→ Revenons à notre cas général : minimiser une fonction de plusieurs variables

> Trouver x^* tel que :
> $$J(x^*) = \inf \{ J(x), x \in A \}$$
> $$A \subset \mathbb{R}^n$$



→ Le principe de la méthode consiste à minimiser J successivement par rapport à chacune des variables

→ Principe simple, mais efficacité limitée, surtout quand le nombre de variables augmente

II.3.e Méthode de gradient (ou de descente)

→ Au voisinage de x : approximation de J par son développement de Taylor d'ordre 1 :

$$J(x + \Delta x) \approx \tilde{J}(x + \Delta x) = J(x) + \nabla J(x)^T . \Delta x$$

→ Le long d'une isovaleur : $\tilde{J}(x + \Delta x) = J(x) \Longrightarrow \nabla J(x)^T . \Delta x = 0$

Conséquence : gradient localement normal aux isovaleurs

→ Dans la direction du gradient : $\nabla J(x)^T . \Delta x = \mp \|\nabla J(x)\| . [\![\Delta x]\!]$

Conséquence : plus grande variation de J, à $\|\Delta x\|$ donné

→ Meilleure direction de recherche *locale* : $-\nabla J(x)$

II.3.e Méthode de gradient à pas optimal

- Soit x_n, la solution approchée en début d'itération
- Direction de recherche : $-\nabla J(x_n)$
- On cherche x_{n+1} dans cette direction : $x_{n+1} = x_n - \alpha.\nabla J(x_n)$
- Pour cela, on minimise $\mathrm{G}(\alpha) = J\big(x_n - \alpha.\nabla J(x_n)\big)$, par rapport à α
- Minimisation de $\mathrm{G}(\alpha)$ = problème à 1 dimension

II.3.e Méthode de gradient à pas optimal

→ Principe simple, mais il faut calculer le gradient

→ Efficacité meilleure que la relaxation, mais limitée car en pratique les directions de recherche successives sont 2 à 2 orthogonales

→ Remarque : on peut se contenter de trouver α tel que :

- $J\big(x_n - \alpha.\nabla J(x_n)\big) < J(x_n)$ on s'assure de toujours « descendre »
- On peut faire du gradient « à pas fixe » : voir TP

II.3.e Méthode du gradient, exemple

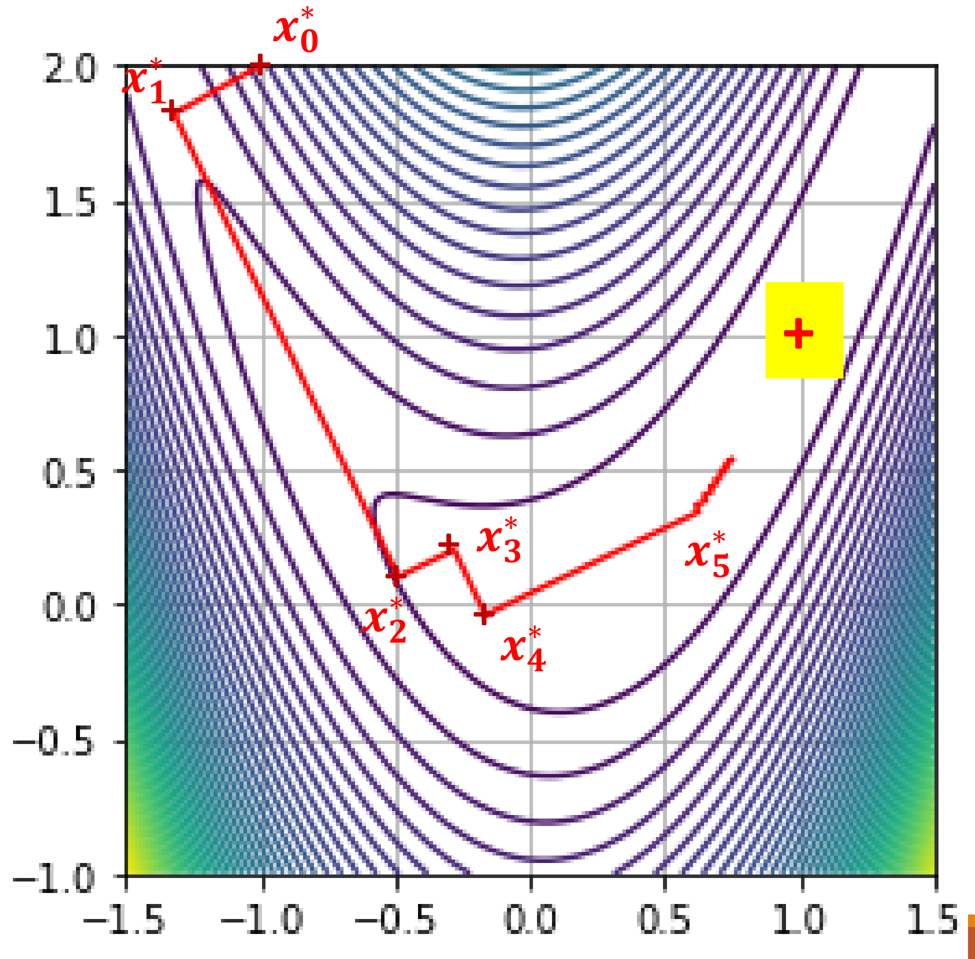
→ Banane de Rosenbrock :

$J(x_1, x_2) = (x_1 - 1)^2 + 10.(x_1{}^2 - x_2)^2$

$J(x_1, x_2) \geq 0$, s'annule si $x_1 - 1 = 0$ et $x_1{}^2 - x_2 = 0$

=> minimum unique en (1,1)

Premières itérations efficaces, puis avancée en zig-zag le long de la vallée (directions de recherche orthogonales)



II.3.f Méthode de Newton, en n-d

- Approximation locale de la fonction par son développement de Taylor d'ordre 2
- $\tilde{J}(x + \Delta x) = J(x) + \nabla J(x)^T . \Delta x + \frac{1}{2} \Delta x^T . H_J(x) . \Delta x$
- Son minimum est alors donné par $\nabla \tilde{J}(x + \Delta x) = 0$
- $\nabla \tilde{J}(x + \Delta x) = \nabla J(x) + H_J(x) . \Delta x$
- $\nabla \tilde{J}(x + \Delta x) = 0 \implies \Delta x$ est solution de : $H_J(x) . \Delta x = -\nabla J(x)$ système linéaire
- Localement très puissant, mais à utiliser avec précaution (peut diverger)
- Nécessite de connaître la dérivée seconde…

Fin de la partie II

→ Les méthodes abordées seront mises en œuvre en TD/TP