

# Table des matières

I	Optimisation unidimensionnelle . . . . .	3
1	Méthodes itératives utilisant la dérivée . . . . .	6
1.1	Avec la dérivée seconde : La méthode de Newton . . . . .	7
1.2	Sans la dérivée seconde : La méthode de la sécante . . . . .	10
1.3	Dichotomie avec dérivée . . . . .	12
1.4	Méthode de la fausse position . . . . .	13
2	Dichotomies sans dérivée . . . . .	14
2.1	Principe général : dichotomie ou plutôt "trichotomie" . . . . .	14
2.2	Stratégie du petit epsilon : la plus efficace sur une étape donnée . . . . .	15
2.3	Stratégie du recyclage . . . . .	15
2.4	Recyclage basique : la méthode du nombre d'or . . . . .	16
2.5	Recyclage sophistiqué : la méthode de Fibonacci . . . . .	16
II	Optimisation multidimensionnelle sans contraintes . . . . .	17
1	Méthode de Newton . . . . .	17
2	Méthodes de descente directionnelle . . . . .	18
2.1	Idée générale . . . . .	18
2.2	Méthode du gradient . . . . .	18
2.3	Méthode des coordonnées cycliques . . . . .	19
2.4	Méthode de Hooke and Jeeves . . . . .	19
3	Méthode de Hooke and Jeeves à pas discret . . . . .	19

## Préambule : Atteindre l'inatteignable

Les nombres réels  $x \in \mathbb{R}$  sont par nature des *limites*. Cela signifie qu'on peut très bien s'en approcher, mais jamais vraiment les attraper... Même si, parfois, des formules apprises en cours de mathématiques donnent l'illusion de le faire. Prenons une équation familière, telle que  $x^2 = 2$ . Sa solution positive est  $x = \sqrt{2}$ . Mais qu'est-ce que  $\sqrt{2}$ ? C'est simplement une notation pour désigner "la solution positive de l'équation  $x^2 = 2$ ". Pour avoir une idée de ce que vaut  $\sqrt{2}$  concrètement, on n'a pas d'autre choix que de trouver un processus pour s'en approcher, par exemple une approximation en nombres décimaux

$$x \approx 1.41421356\dots$$

Cela n'est pas si grave : dans la vie pratique, on se satisfera toujours d'un résultat à *un certain degré de précision près*, préalablement fixé. Plutôt qu'un résultat algébrique exact, une réponse tout à fait satisfaisante à un problème peut donc prendre la forme d'une suite  $x_k$  qui se rapproche dudit résultat. Une telle "réponse" n'est pas unique, il y en a une infinité... On peut discerner deux critères qui permettent d'évaluer le degré d'efficacité de la suite choisie :

1. il faut que  $(x_k)$  converge effectivement vers le bon résultat.
2. il faut aussi que  $(x_k)$  converge suffisamment *vite* : on souhaite se rapprocher au mieux du résultat sans avoir à calculer des millions de termes...

Cette notion de "vitesse de convergence" peut être définie mathématiquement.



### Définition

On dit que la convergence de la suite  $(x_n)_{n \in \mathbb{N}}$  vers sa limite  $x$  est au moins d'*ordre*  $p$  si :

$$\exists M \in \mathbb{R}_+^*, \exists n_0 \in \mathbb{N}^*, \forall n \in \mathbb{N}, n \geq n_0 \Rightarrow |x_{n+1} - x| \leq M|x_n - x|^p$$

avec  $M < 1$  si  $p = 1$ .

Plus  $p$  est grand, plus la suite converge vite. Pour  $p = 1$ , la convergence est dite *linéaire* ou *géométrique*. Pour  $p = 2$ , la convergence est *quadratique*.



### Exemple

Dans l'exemple suivant, la suite de la colonne de gauche converge définie sur  $\mathbb{N}^*$  par  $x_n = 10^{-n}$  converge linéairement vers 0, alors que la suite de colonne de droite définie sur  $\mathbb{N}^*$  par  $y_n = 10^{-2^n}$  a un ordre de convergence quadratique.

convergence :	linéaire (M=0.1)	quadratique (M=1)
	$x_1 = 1$	$y_0 = 1$
	$x_1 = 0,1$	$y_0 = 0,01$
	$x_1 = 0,01$	$y_1 = 0,0001$
	$x_2 = 0,001$	$y_2 = 0,00000001$
	$x_3 = 0,0001$	$y_3 = 0,000000000000000001$
	...	...

## I Optimisation unidimensionnelle

Dans cette première partie, on cherche à minimiser la valeur d'une fonction  $f$  à une variable réelle, définie sur un intervalle  $[a; b]$  dans l'ensemble des réels :

$$f : [a, b] \rightarrow \mathbb{R}.$$

Nous allons décrire différentes méthodes itératives pour se rapprocher du point

$$x^* \in [a, b]$$

où  $f$  atteint son minimum. Notons que toutes ces méthodes seront évidemment valables aussi pour maximiser une fonction (puisqu'on maximiser  $f$  est la même chose que minimiser  $-f$ ).



### Définition

La fonction  $f$  admet un minimum en  $x^*$  si et seulement si pour tout  $x \in [a, b]$ , on a  $f(x^*) \leq f(x)$ .  
La fonction  $f$  admet un minimum local en  $x^*$  si et seulement s'il existe  $\epsilon > 0$  tel que pour tout  $x \in [x^* - \epsilon, x^* + \epsilon] \cap [a, b]$ , on ait  $f(x^*) \leq f(x)$ .



### Définition

Une fonction  $f$  est dite *unimodale* sur  $[a, b]$  si et seulement s'il existe  $x^* \in [a, b]$  tel que  $f$  est strictement décroissante sur  $[a, x^*]$  et  $f$  est strictement croissante sur  $[x^*, b]$ .



### Remarques

- Dans ce cas,  $x^*$  est le minimum de  $f$ , est c'est aussi son unique minimum local...
- Si  $f$  est dérivable sur  $]a; b[$  et unimodale, alors

$$\forall x \in ]a; x^*], f'(x) \leq 0 \quad \text{et} \quad \forall x \in [x^*; b[, f'(x) \geq 0$$

Dans toute cette partie, nous avons appliqué les méthodes décrites dans le cours sur tableur Excel. Pour cela, nous avons choisi une fonction  $f$  relativement simple (polynomiale), mais dont le minimum ne peut pas être calculé algébriquement<sup>1</sup> :

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

Cette fonction atteint son minimum en  $x^* \approx 0,75\dots$ . De plus, elle est unimodale sur tout intervalle  $[a, b]$  contenant  $x^*$ .

Les différentes méthodes utilisées pour déterminer une valeur approchée de  $x^*$  sont basées sur le théorème du point fixe suivant :

### Théorème du point fixe - (admis)

Soit  $[a; b]$  un intervalle de  $\mathbb{R}$  et  $g$  une application de  $[a; b]$  dans  $\mathbb{R}$  telle que :

- $g([a; b]) \subset [a; b]$
- $g$  est contractante, c'est à dire :

$$\exists K \in [0; 1[, \forall (x; y) \in [a; b]^2, |g(x) - g(y)| \leq K|x - y|$$

Dans ce cas,  $g$  admet un unique point fixe  $x^* \in [a; b]$  et la suite  $(x_n)$  définie sur  $\mathbb{N}$  par :

$$\begin{cases} x_0 \in [a; b] \\ \forall n \in \mathbb{N}, x_{n+1} = g(x_n) \end{cases}$$

converge vers  $x^*$

De plus :

$$\forall p \in \mathbb{N}, |x_p - x^*| \leq \frac{K^p}{1 - K} |x_1 - x_0|$$

### démonstration

- **L'unicité** du point fixe est donné par le caractère contractant de  $g$ .

En effet, soient  $x^*$  et  $\bar{x}$  deux points fixes distincts. Dans ce cas :

$$\begin{aligned} |x^* - \bar{x}| &\leq K|g(x^*) - g(\bar{x})| \\ &\leq K|x^* - \bar{x}| \end{aligned}$$

$$\text{d'où } (1 - K)|x^* - \bar{x}| \leq 0$$

or  $1 - K < 0$  donc  $|x^* - \bar{x}| = 0$  et par conséquent  $x^* = \bar{x}$

- **Pour l'existence**, en considérant la suite  $(x_n)$  définie par  $x_{n+1} = g(x_n)$  :

Soit  $n \in \mathbb{N}$

$$|x_{n+1} - x_n| = |g(x_n) - g(x_{n-1})|, \text{ donc } |x_{n+1} - x_n| \leq K|x_n - x_{n-1}|$$

1. Plus précisément, il n'existe pas de formule algébrique pour déterminer les zéros de la fonction dérivée, qui est un polynôme de degré cinq. Voir "Théorème d'Abel (algèbre)" sur Wikipédia.

D'où,  $|x_{n+1} - x_n| \leq K^n |x_1 - x_0|$

Donc :

$$\begin{aligned} \sum_{p=0}^{n-1} |x_{p+1} - x_p| &\leq \sum_{p=0}^{n-1} K^p |x_1 - x_0| \\ &\leq |x_1 - x_0| \frac{1 - K^n}{1 - K} \\ &\leq \frac{|x_1 - x_0|}{1 - K} \end{aligned}$$

La série de terme général  $(x_{n+1} - x_n)_{n \in \mathbb{N}}$  est donc absolument convergente et par conséquent la série  $\sum (x_p - x_{p+1})$  est convergente.

Or par télescopage,  $x_n - x_0 = \sum_{k=1}^n (x_k - x_{k-1})$ , donc la suite  $(x_n)$  converge dans  $[a; b]$ .

Soit  $l$  sa limite.

$g$  étant continue car  $K$ -lipschitzienne, on a  $g(l) = l$ , donc  $l$  est un point fixe de  $g$  et par conséquent  $l = x^*$

D'autre part, en utilisant le télescopage et l'inégalité triangulaire, on obtient pour tout  $(n, p) \in \mathbb{N}^2$  tel que  $p < n$  :

$$\begin{aligned} |x_n - x_p| &\leq \sum_{i=p}^{n-1} |x_{i+1} - x_i| \\ &\leq \sum_{i=p}^{n-1} K^i |x_1 - x_0| \\ &\leq K^p \sum_{i=0}^{n-p-1} K^i |x_1 - x_0| \\ &\leq \frac{K^p}{1 - K} |x_1 - x_0| \end{aligned}$$

En passant à la limite quant  $n$  tend vers  $+\infty$ , on obtient la dernière inégalité du théorème qui permet d'avoir une majoration de l'erreur.

### Théorème

Soit  $g$  une fonction de  $C^1([a; b[, \mathbb{R})$ .

Si il existe un point fixe  $x^* \in ]a; b[$  tel que  $g'(x^*) < 1$ , alors il existe  $\alpha \in \mathbb{R}$  tel que  $g$  est stable contractante sur  $[x^* - \alpha, x^* + \alpha]$ .

### démonstration

Soient  $g$  une fonction  $C^1$  ( $]a; b[, \mathbb{R}$ ) et  $x^* \in ]a; b[$  tel que  $g'(x^*) < 1$ .

$g'$  étant continue sur  $x^* \in ]a; b[$ , il existe  $K \in [0; 1[$  et  $\alpha \in ]0; \beta[$  tel que :

pour tout  $x \in V = ]x^* - \alpha; x^* + \alpha[$ ,  $g'(x) \leq K$ .

Donc, d'après l'inégalité des accroissements finis, pour tout  $(x, y) \in V^2$ , on a :

$$|g(x) - g(y)| \leq K|x - y|$$

$g$  est donc contractante sur  $V$ .

De plus  $g(x^*) = x^*$ , donc pour tout  $x \in V$ ,  $|g(x) - x^*| \leq K|x - x^*|$ , c'est à dire que pour tout  $x \in V$ ,  $g(x) \in V$ .  $V$  est par conséquent stable par  $g$ ,

### Théorème (Corollaire)

En conservant les notations du théorème précédent, et en supposant  $g$  vérifie les proposition suivantes :

- $g([a; b]) \subset [a; b]$
- $\exists n_0 \in \mathbb{N}^*$  tel que  $g^{n_0}$  soit contractante où  $g^{n_0} = g \circ g \circ \dots \circ g$ , itérée  $n_0$  fois.

Alors  $g$  admet un unique point fixe.

### démonstration

- **Unicité** : Soit  $x^*$  et  $\bar{x}$  deux points fixes par  $g$ , alors  $g(x^*) = g(\bar{x})$ , et par conséquent pour tout  $n \in \mathbb{N}^*$ ,  $g^n(x^*) = g^n(\bar{x})$ , d'où  $g^{n_0}(x^*) = g^{n_0}(\bar{x})$ .

Or on a  $g^{n_0}$  est qui est contractante, donc  $x^* = \bar{x}$ .

- **Existence** : En appliquant le théorème du point fixe à  $g^{n_0}$ , il existe un unique point fixe  $x^*$  tel que  $g^{n_0}(x^*) = x^*$ .

Donc  $(g \circ g^{n_0})(x^*) = g(x^*)$ , d'où  $g^{n_0}(g(x^*)) = g(x^*)$  et par conséquent  $g(x^*)$  est aussi un point fixe de  $g^{n_0}$ , or ce point fixe est unique, donc  $g(x^*) = x^*$ .

## 1 Méthodes itératives utilisant la dérivée

Dans cette section, on suppose que la fonction  $f$  est dérivable et que l'on connaît l'expression de sa dérivée  $f'$ . On sait que si  $f$  atteint son minimum en  $x^*$ , alors sa dérivée s'annule en ce point :  $f'(x^*) = 0$ .

Attention, rappelons que la réciproque n'est pas vraie : en général, la dérivée peut s'annuler ailleurs qu'en un minimum ou maximum local... Voir par exemple  $f(x) = x^3$ .

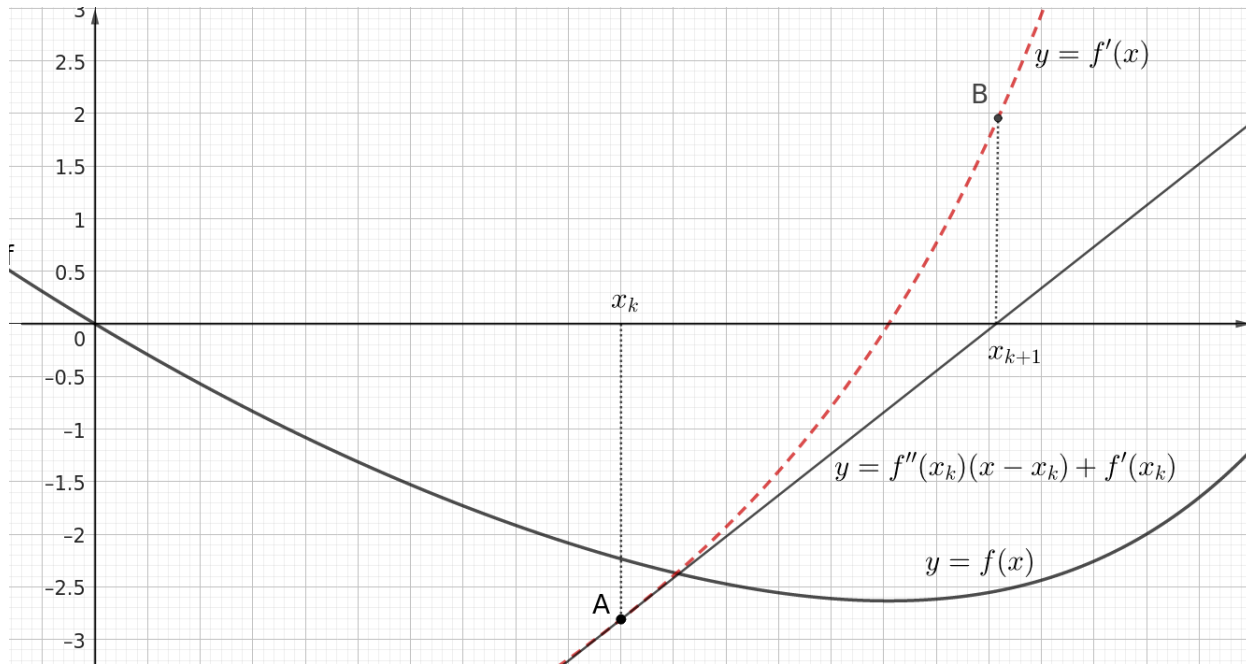
Mais si de plus  $f$  est unimodale, alors le minimum  $x^*$  cherché sera bien l'*unique* solution de l'équation

$$f'(x^*) = 0.$$

On va donc supposer que  $f$  est unimodale sur son intervalle de définition  $[a; b]$ .

### 1.1 Avec la dérivée seconde : La méthode de Newton

On cherche à construire une suite  $(x_k)$  qui converge vers  $x^*$ , à partir d'un réel  $x_0$  choisi arbitrairement dans l'intervalle de définition de la fonction (si possible assez proche tout de même de  $x^*$ ). La méthode de Newton consiste à obtenir  $x_{k+1}$  à partir de  $x_k$  en approximant  $f'$  par sa tangente en  $x_k$ ... Puisqu'on cherche la solution de  $f'(x) = 0$ , on définit alors  $x_{k+1}$  comme étant le point où cette tangente coupe l'axe des abscisses.



Passons au calculs : la tangente à la courbe de  $f'$  en  $x = x_k$  a pour équation

$$y = f''(x_k)(x - x_k) + f'(x_k).$$

Donc  $x_{k+1}$  est la solution de l'équation

$$0 = f''(x_k)(x_{k+1} - x_k) + f'(x_k),$$

soit

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

#### ♥ Théorème

Soient  $f \in C^3 ]a; b[; \mathbb{R}$  et  $x^* \in ]a; b[$  tel que  $f'(x^*) = 0$ .

1. Si  $x^*$  est de multiplicité 1 pour  $f'$ , (c'est à dire que  $f''(x^*) \neq 0$ ), alors il existe  $\alpha \in \mathbb{R}_+^*$  tel que la suite  $(x_n)_{n \in \mathbb{N}}$  définie par :

$$\begin{cases} x_0 \in ]x^* - \alpha; x^* + \alpha[ \\ \forall k \in \mathbb{N}, x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} \end{cases}$$

est bien définie et converge vers  $x^*$  de manière quadratique.

2. Si  $x^*$  est de multiplicité 2 pour  $f'$ , (c'est à dire que  $f''(x^*) = 0$  et  $f'''(x^*) \neq 0$ ), alors il existe  $\alpha \in \mathbb{R}_+^*$  tel que la suite  $(x_n)_{n \in \mathbb{N}}$  définie par :

$$\begin{cases} x_0 \in ]x^* - \alpha; x^* + \alpha[ \\ \forall k \in \mathbb{N}^*, \begin{cases} x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)} & \text{si } x_k \neq x^* \\ x_{k+1} = x^* & \text{si } x_k = x^* \end{cases} \end{cases}$$

est bien définie et converge vers linéairement vers  $x^*$ .

### démonstration

Soient  $f \in C^3 ]a; b[; \mathbb{R}$  et  $x^* \in ]a; b[$  tel que  $f'(x^*) = 0$  et  $g$  la fonction définie par

$$g(x) = x - \frac{f'(x)}{f''(x)}$$

1. Tout d'abord,  $g(x^*) = x^*$  donc  $x^*$  est un point fixe de  $g$  appartenant à  $]a; b[$ .  
D'autre part,  $f$  est de classe  $C^3$  donc  $f''$  est continue, or  $f''(x^*) \neq 0$ , donc il existe un  $\beta \in \mathbb{R}_+^*$  tel que pour tout  $x \in U = ]x^* - \beta; x^* + \beta[$ ,  $f''(x) \neq 0$ .  
Par conséquent,  $g$  est définie et dérivable sur  $U$  et on a :

$$\begin{aligned} g'(x) &= 1 - \frac{f''(x) \times f''(x) - f'(x) \times f'''(x)}{(f''(x))^2} \\ &= 1 - 1 + \frac{f'(x) \times f'''(x)}{(f''(x))^2} \\ &= \frac{f'(x) \times f'''(x)}{(f''(x))^2} \end{aligned}$$

or  $f'(x^*) = 0$ , donc  $g'(x^*) = 0$  et par conséquent il existe  $\alpha \in ]0; \beta[$  et  $V = ]x^* - \alpha; x^* + \alpha[ \subset U$ , tel que  $g$  est contractante et stable sur  $V$  et par conséquent, en utilisant le théorème du point fixe, la suite  $(x_n)$  converge vers l'unique point fixe de  $g$  qui est  $x^*$ .

D'autre part, l'égalité de Taylor-Lagrange sur  $f'$  à l'ordre 2 entre  $x^*$  et  $x_n$ , donne l'exis-



tence de  $y_n$  tel que  $|y_n - x^*| < |x_n - x^*|$  et :

$$f'(x^*) = f'(x_n) + (x^* - x_n)f''(x_n) + \frac{1}{2}(x^* - x_n)^2 f'''(y_n)$$

or  $f'(x^*) = 0$ , d'où :

$$f'(x_n) = (x_n - x^*)f''(x_n) - \frac{1}{2}(x_n - x^*)^2 f'''(y_n)$$

D'où, en utilisant la relation de récurrence du théorème, on obtient :

$$\begin{aligned} x_{n+1} - x^* &= x_n - \frac{f'(x_n)}{f''(x_n)} - x^* \\ &= x_n - x^* - \frac{(x_n - x^*)f''(x_n) - \frac{1}{2}(x_n - x^*)^2 f'''(y_n)}{f''(x_n)} \\ &= (x_n - x^*) \times \left( 1 - \frac{f''(x_n) - \frac{1}{2}(x_n - x^*)f'''(y_n)}{f''(x_n)} \right) \\ &= (x_n - x^*) \times \left( \frac{\frac{1}{2}(x_n - x^*)f'''(y_n)}{f''(x_n)} \right) \\ &= \frac{1}{2}(x_n - x^*)^2 \frac{f'''(y_n)}{f''(x_n)} \end{aligned}$$

la convergence est donc quadratique.

2. Supposons maintenant que  $f''(x^*) = 0$  et que  $f'''(x^*) \neq 0$ .  
Comme dans la démonstration précédente, il existe  $\beta \in \mathbb{R}_+$  tel que :  
pour tout  $x \in U = ]x^* - \beta; x^* + \beta[$ , tel que  $x \neq x^*$ , on a  $f''(x) \neq 0$ .  
Soit  $g$  la fonction définie sur  $U$  par :

$$g(x) = \begin{cases} x^* & \text{si } x = x^* \\ x - \frac{f'(x)}{f''(x)} & \text{si } x \neq x^* \end{cases}$$

$g$  ainsi définie est  $C^1$ ,  $x^*$  est un point fixe de  $g$  sur  $]a; b[$  et  $g'(x^*) = \frac{1}{2}$

En effet,  $g(x^*) = x^*$  et en écrivant les formules de Taylor-Young autour de  $x^*$  pour  $f'$ ,  $f''$

et  $f'''$ , on obtient :

$$\begin{cases} f'(x) = \frac{(x - x^*)^2}{2} f'''(x^*) + o((x - x^*)^2) \\ f''(x) = (x - x^*) f'''(x^*) + o((x - x^*)) \\ f'''(x) = f'''(x^*) + o((x - x^*)) \end{cases}$$

d'où  $\lim_{x \rightarrow x^*} g(x) = x^*$  donc  $g$  est continue sur  $]a; b[$  et dérivable sur  $]a; b[ \setminus \{x^*\}$ .

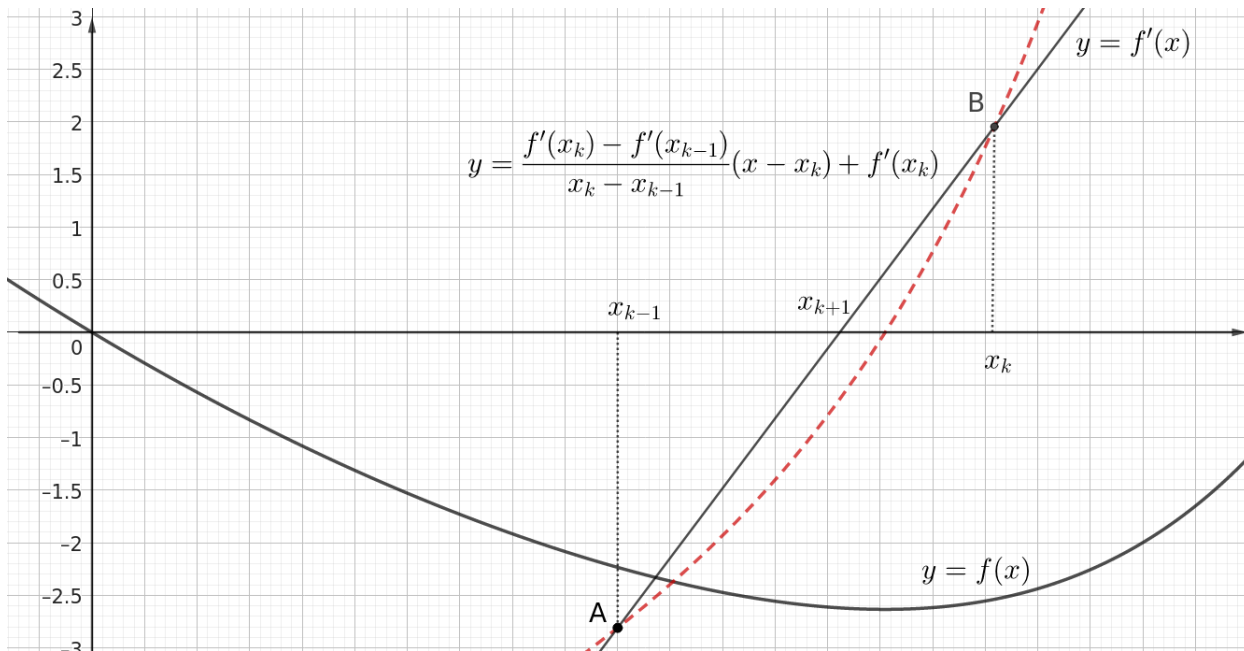
De plus pour tout  $x \in U \setminus \{x^*\}$ , on a :

$$g'(x) = 1 - \frac{f''^2(x) - f'(x)f'''(x)}{f''^2(x)} = \frac{f'(x)f'''(x)}{f''^2(x)}$$

D'où  $\lim_{x \rightarrow x^*} g'(x) = \frac{1}{2}$  et par conséquent  $g$  est dérivable en  $x^*$  et  $g'(x^*) = \frac{1}{2}$ .  
 $g$  est donc contractante et stable sur  $V$  et par conséquent, en utilisant le théorème du point fixe, la suite  $(x_n)$  converge vers l'unique point fixe  $x^*$  de  $g$  sur  $]a; b[$  et sa convergence est au moins linéaire.

## 1.2 Sans la dérivée seconde : La méthode de la sécante

Si l'on ne dispose pas de la dérivée seconde (ou si l'on est trop paresseux pour la calculer), on peut utiliser la méthode de la sécante.



Passons au calculs : la droite passant par les points  $A(x_k; f(x_k))$  et  $B(x_{k-1}; f(x_{k-1}))$  a pour équation

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

Donc  $x_{k+1}$  est la solution de l'équation

$$0 = \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}(x_{k+1} - x_k) + f'(x_k),$$

soit

$$\begin{aligned} x_{k+1} &= x_k - f'(x_k) \times \frac{x_k - x_{k-1}}{f'(x_k) - f'(x_{k-1})} \\ &= \frac{f'(x_k)x_{k+1} - f'(x_{k+1})x_k}{f'(x_k) - f'(x_{k-1})} \end{aligned}$$

### ♥ Théorème

Soient  $f \in C^3 ]a; b[; \mathbb{R}$  et  $x^* \in ]a; b[$  tel que  $f'(x^*) = 0$  et  $f''(x^*) \neq 0$ .  
il existe un voisinage  $V$  de  $x^*$  tel que la suite  $(x_n)_{n \in \mathbb{N}}$  définie par :

$$\left\{ \begin{array}{l} (x_0, x_1) \in V^2 \text{ tel que } x_0 \neq x_1 \\ \forall k \in \mathbb{N}, \left\{ \begin{array}{l} x_{k+1} = \frac{f'(x_k)x_{k+1} - f'(x_{k+1})x_k}{f'(x_k) - f'(x_{k-1})} \quad \text{si } f'(x_k) - f'(x_{k-1}) \neq 0 \\ x_{k+1} = x^* \quad \text{si } f'(x_k) - f'(x_{k-1}) = 0 \end{array} \right. \end{array} \right.$$

est correctement définie et converge vers  $x^*$ .

De plus, si  $f'''(x^*) \neq 0$ , alors il existe  $M \in ]0; 1[$  tel que pour tout  $k \in \mathbb{N}$ , on a  $\|x_{k+1} - x^*\| \leq M\|x_k - x^*\|^\alpha$ . avec  $\alpha = \frac{1 + \sqrt{5}}{2}$

La suite  $(x_k)$  ainsi définie converge un peu moins rapidement qu'avec la méthode de Newton. Plus précisément, si  $f'''(x^*) \neq 0$ , l'ordre de convergence est égal au nombre d'or, 1.618.... Cela reste tout de même très rapide : c'est mieux qu'une convergence linéaire.

### 🔪 démonstration

On a  $f''(x^*) \neq 0$ .

Soit  $I$  un voisinage de  $x^*$  tel qu'il existe  $m \in \mathbb{R}_+^*$  tel que :

$$\forall x \in I, |f''(x^*)| \geq m$$

Cela signifie que  $f'$  est strictement monotone sur  $I$ , ce qui est bien sur le cas d'une fonction unimodale.

On note  $M = \max_{x \in I} |f'''(x)|$

Soit  $\beta \in \mathbb{R}_+^*$  tel que  $0 < \frac{2M\beta}{m} < 1$  et que l'intervalle  $J = ]x^* - \beta; x^* + \beta[$  est inclus dans  $I$ . Montrons alors par récurrence que la suite définie dans le théorème de la sécante est bien définie, c'est à dire que :

$$\forall n \in \mathbb{N}, x_n \in J \quad \text{et} \quad (x_n \neq x_{n-1} \text{ ou } x_n = x_{n-1} = x^*)$$

○ **Initialisation**

On a bien  $(x_0, x_1) \in V^2$ , d'où l'initialisation.

○ **Hérédité**

soit  $n \in \mathbb{N}$  et supposons que la proposition est vraie aux rang  $n$  et au rang  $n - 1$ .

La stricte monotonie de  $f'$  sur  $J$  permet de définir  $x_{n+1}$ .

En utilisant deux fois le théorème des accroissements finis dans le cas où  $f'(x_n) \neq f'(x_{n-1})$ , on obtient :

$$\begin{cases} \exists y_{1,n} \in ]x_n, x^*[ \quad \text{tel que} \quad f'(x_n) = (x_n - x^*)f''(y_{1,n}) \\ \exists y_{2,n} \in ]x_n, x^*[ \quad \text{tel que} \quad f'(x_n) - f'(x_{n-1}) = (x_n - x_{n-1})f''(y_{2,n}) \end{cases}$$

$$\text{D'où, } x_{n+1} = x_n - \frac{(x_n - x_{n-1})(x_n - x^*)f''(y_{1,n})}{(x_n - x_{n-1})f''(y_{2,n})} = x_n - (x_n - x^*)\frac{f''(y_{1,n})}{f''(y_{2,n})}$$

et par conséquent :

$$x_{n+1} - x^* = (x_n - x^*) \left( \frac{f''(y_{2,n}) - f''(y_{1,n})}{f''(y_{2,n})} \right)$$

Donc il existe  $y_{3,n} \in ]y_{1,n}, y_{2,n}[$  tel que :

$$x_{n+1} - x^* = (x_n - x^*) \left( \frac{f'''(y_{3,n})}{f''(y_{2,n})} \right)$$

donc  $|x_{n+1} - x^*| \leq |x_n - x^*| \frac{M}{m} 2\beta$ , d'où l'hérédité.

○ **Conclusion**  $\forall n \in \mathbb{N}, x_n \in J$  et  $|x_{n+1} - x^*| \leq |x_n - x^*| \frac{M}{m} 2\beta$

On déduit de ce qui précède que la suite  $(x_n)$  est convergente et que sa convergence est au moins linéaire.

### 1.3 Dichotomie avec dérivée

Le principe est simple. Puisque  $f$  est dérivable et unimodale, on sait que  $f'(x)$  est négative si  $x < x^*$  et positive si  $x > x^*$ .

On part d'un intervalle  $[a_0, b_0]$  tel que  $f'(a_0) < 0$  et  $f'(b_0) > 0$ . On sait alors que  $x^* \in [a_0, b_0]$ . On construit à partir de là une suite d'intervalles  $[a_k, b_k]$  de plus en plus petits. À chaque étape de cette méthode itérative, on découpe l'intervalle  $[a_k, b_k]$  en deux en posant  $c_k = (a_k + b_k)/2$ . Si  $f(c_k) \geq 0$ , alors on pose  $[a_{k+1}, b_{k+1}] := [a_k, c_k]$ . Sinon, on pose  $[a_{k+1}, b_{k+1}] := [c_k, b_k]$ . On s'arrête lorsque la longueur de l'intervalle obtenu  $(b_k - a_k)$  est inférieure au degré de précision  $\varepsilon$  souhaité.

### 1.4 Méthode de la fausse position

On peut adapter la méthode de la sécante pour avoir une méthode de la sécante permettant de déterminer  $x^*$  si  $f$  est juste  $C^1$  de  $]a; b[$  dans  $\mathbb{R}$  avec un intervalle permettant de maximiser l'erreur commise comme dans la Dichotomie.

#### ♥ Théorème

Soient  $f \in C^1(]a; b[; \mathbb{R})$  unimodale et  $x^* \in ]a; b[$  l'unique réel tel que  $f'(x^*) = 0$ .

Les suites  $(a_n)$ ,  $(b_n)$  et  $(x_n)$  définie par :

$$\left\{ \begin{array}{l} a_0 = a \text{ et } b_0 = b \\ \forall k \in \mathbb{N}, \left\{ \begin{array}{l} x_k = a_k - f'(a_k) \times \frac{a_k - b_k}{f'(a_k) - f'(b_k)} \text{ si } a_k \neq b_k \text{ et } x_k = a_k \text{ si } a_k = b_k \\ a_{k+1} = a_k \text{ et } b_{k+1} = x_k \text{ si } f'(a_k)f'(x_k) < 0 \\ a_{k+1} = x_k \text{ et } b_{k+1} = b_k \text{ si } f'(a_k)f'(x_k) > 0 \\ a_{k+1} = x_k \text{ et } b_{k+1} = x_k \text{ si } f'(a_k)f'(x_k) = 0 \end{array} \right. \end{array} \right.$$

sont correctement définies et convergent vers  $x^*$  de manière linéaire.

De plus, pour tout  $k \in \mathbb{N}$ , on a  $x^* \in [a_k; b_k]$

#### 🔪 démonstration

$f$  étant unimodale et  $C^1([a; b], \mathbb{R})$ , il existe  $x^* \in [a; b]$  tel que :

$$\left\{ \begin{array}{l} f'(x^*) = 0 \\ \forall x \in [a; x^*[ , f'(x) < 0 \\ \forall x \in ]x^*, b] , f'(x) < 0 \end{array} \right.$$

On montre que  $(a_n)$  et  $(b_n)$  sont des suites adjacentes convergent vers  $x^*$ .

En effet, démontrons par récurrence que pour tout  $n \in \mathbb{N}$ ,  $a_n \leq a_{n+1} \leq x^* \leq b_{n+1} \leq b_n$

- **Initialisation**

- **Hérédité**

Soit  $n \in \mathbb{N}$  tel que  $a_n \leq a_{n+1} \leq x^* \leq b_{n+1} \leq b_n$ .

Supposons que  $a_{n+1} \neq b_{n+1}$ , dans ce cas :

$$x_{n+1} = a_{n+1} - f'(a_{n+1}) \times \frac{a_{n+1} - b_{n+1}}{f'(a_{n+1}) - f'(b_{n+1})}$$

\* si  $f'(a_{n+1})f'(x_{n+1}) < 0$ , alors on a :

$$\begin{aligned} b_{n+2} - b_{n+1} &= a_{n+1} - f'(a_{n+1}) \times \frac{a_{n+1} - b_{n+1}}{f'(a_{n+1}) - f'(b_{n+1})} - b_{n+1} \\ &= a_{n+1} + \frac{b_{n+1}f'(b_{n+1}) - a_{n+1}f'(a_{n+1})}{f'(a_{n+1}) - f'(b_{n+1})} \end{aligned}$$

- Conclusion

## 2 Dichotomies sans dérivée

Dans cette section, on suppose que  $f$  n'est pas dérivable ou qu'on ne connaît pas l'expression de sa dérivée.

### 2.1 Principe général : dichotomie ou plutôt "trichotomie"

Tout est basé sur la propriété suivante qui est une conséquence directe de la définition des fonctions unimodales.

#### ♥ Théorème

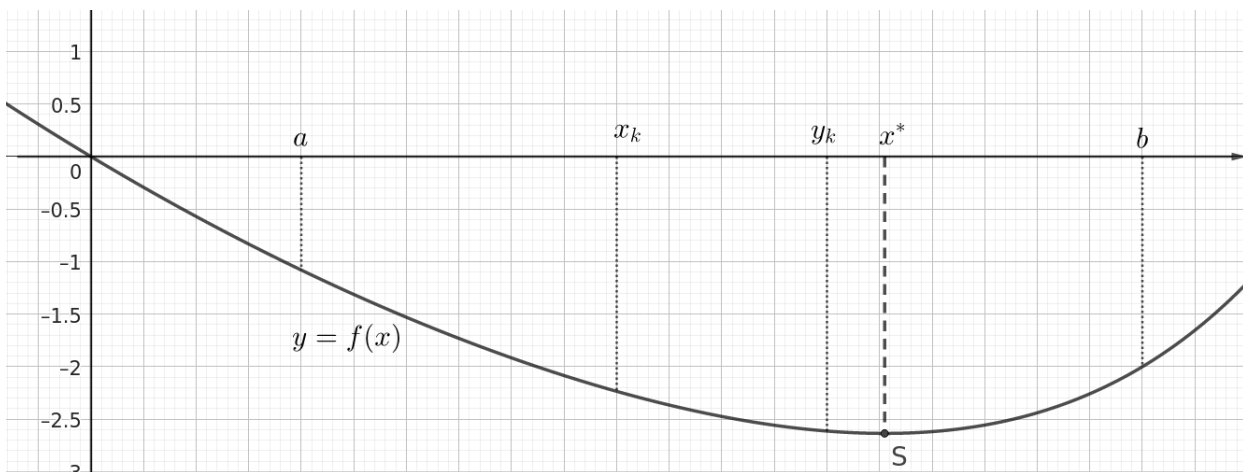
Soit  $f$  une fonction unimodale sur  $[a, b]$ , et deux nombres  $x$  et  $y$  tels que  $a < x < y < b$ .

- Si  $f(x) \leq f(y)$ , alors  $x^* \in [a, y]$ ,
- Si  $f(x) \geq f(y)$ , alors  $x^* \in [x, b]$ .

#### 🔪 démonstration

En effet, si  $f(x) \leq f(y)$ , alors  $f$  n'est pas décroissante sur  $[a, y]$ , or elle est unimodale donc décroissante sur  $[a, x^*]$  et par conséquent  $x^* < y$ , d'où  $x^* \in [a, y]$

D'autre part, si  $f(x) \geq f(y)$ , alors  $f$  n'est pas croissante sur  $[x, b]$ , or elle est unimodale donc croissante sur  $[x^*, b]$  et par conséquent  $x < x^*$ , d'où  $x^* \in [x, b]$



L'idée est donc de découper nos intervalles successifs  $[a_k, b_k]$  en trois parties, en choisissant  $x_k$  et  $y_k$  tels que  $a_k < x_k < y_k < b_k$ . Le calcul de  $f(x_k)$  et  $f(y_k)$  permettra donc, par application du Théorème 2.1, de réduire l'intervalle contenant  $x^*$  :

- Si  $f(x_k) \leq f(y_k)$ , alors  $[a_{k+1}, b_{k+1}] = [a_k, y_k]$ ,
- Sinon,  $[a_{k+1}, b_{k+1}] = [x_k, b_k]$ .

Toute la question est de choisir  $x_k$  et  $y_k$  à l'intérieur de nos intervalles  $[a_k, b_k]$ ... Comment les choisir efficacement ?

Une première exigence, qui semble naturelle, est de choisir  $x_k$  et  $y_k$  symétriques par rapport au centre de l'intervalle  $[a_k, b_k]$ . Par conséquent, le choix de  $x_k$  détermine  $y_k$  et réciproquement. De plus, la symétrie implique que la taille de l'intervalle  $[a_{k+1}, b_{k+1}]$  ne dépend pas du résultat du test.

## 2.2 Stratégie du petit epsilon : la plus efficace sur une étape donnée

Si l'on raisonne sur une seule étape, on minimisera la taille de l'intervalle suivant en prenant  $x_k$  et  $y_k$  au plus proche du centre de l'intervalle  $[a_k, b_k]$  :

$$x_k := \frac{a_k + b_k}{2} - \varepsilon \quad \text{et} \quad y_k := \frac{a_k + b_k}{2} + \varepsilon,$$

avec  $\varepsilon > 0$  aussi petit que possible. Avec cette méthode, la taille des intervalles est "presque" divisée par deux à chaque étape. En fait, on se rapproche de la méthode de dichotomie avec dérivée : ici, le signe de la dérivée au centre de l'intervalle est remplacé par le signe de la variation de  $f$  sur un très petit intervalle autour du centre...

## 2.3 Stratégie du recyclage

Si la stratégie du petit epsilon est effectivement la plus efficace sur une itération donnée, il se trouve qu'il existe de meilleurs choix globaux. Voyons ce qu'il se passe si les itérations de l'algorithme sont un peu moins individualistes et décident de coopérer entre elles<sup>2</sup>.

L'idée du recyclage est de faire en sorte que l'un des  $x_k$  ou  $y_k$  non utilisé pour être l'une des extrémités de l'intervalle  $[a_{k+1}, b_{k+1}]$  soit réutilisé à l'étape  $k + 1$ ...

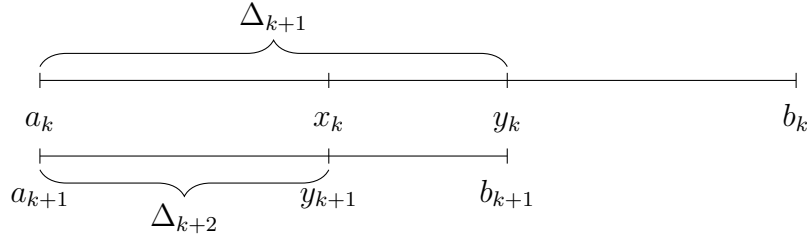
On évalue donc  $f$  deux fois (en  $x_0$  et  $y_0$ ) lors de la première étape  $k = 0$ , puis grâce au recyclage, on évalue  $f$  une seule fois à chaque étape suivante. Étant donné que c'est le calcul de  $f$  qui demande le gros du travail dans l'algorithme de dichotomie, on comprend pourquoi cette stratégie est intelligente.

Posons  $\Delta_k := b_k - a_k$  la longueur de l'intervalle à l'étape  $k$ . On peut montrer (voir figure ci-dessous) que le principe de symétrie ( $b_k - x_k = y_k - a_k$  ou encore  $x_k - a_k = b_k - y_k$ ) couplé avec le principe de recyclage implique la relation :

$$\Delta_k = \Delta_{k+1} + \Delta_{k+2}. \tag{1}$$

---

2. C'est un problème social bien connu : les meilleures stratégies d'un point de vue individuel sont parfois loin d'être optimales si elles sont appliquées à l'échelle collective...



### démonstration

En effet, En supposant que  $f(x_k) \leq f(y_k)$  et  $f(x_{k+1}) \leq f(y_{k+1})$  avec  $f(y_{k+1}) = f(x_k)$ , on obtient :

$$\begin{aligned}
 \Delta_{k+1} + \Delta_{k+2} &= b_{k+1} - a_{k+1} + b_{k+2} - a_{k+2} \\
 &= b_{k+1} - a_k + y_{k+1} - a_k \\
 &= b_{k+1} - a_k + x_k - a_k \\
 &= b_{k+1} - a_k + b_k - y_k \quad \text{d'après le principe de symétrie} \\
 &= b_{k+1} - a_k + b_k - b_{k+1} \\
 &= b_k - a_k \\
 &= \Delta_k.
 \end{aligned}$$

## 2.4 Recyclage basique : la méthode du nombre d'or

Supposons que le rapport entre les tailles de deux intervalles consécutifs soit une constante  $\alpha > 0$  :

$$\Delta_{k+1} = \alpha \Delta_k.$$

Alors  $\Delta_{k+2} = \alpha \Delta_{k+1} = \alpha^2 \Delta_k$ . Donc l'équation (1) implique  $1 = \alpha + \alpha^2$ . La solution positive de cette équation est

$$\alpha = \frac{-1 + \sqrt{5}}{2} \approx 0.618\dots$$

En pratique, on commence donc par construire la suite  $\Delta_k = \alpha^k \Delta_0$  avec  $\Delta_0 = b_0 - a_0$ , et à chaque étape  $k$  on pose

- Si  $f(x_k) \leq f(y_k)$ , alors  $x_{k+1} = b_{k+1} - \Delta_{k+2}$  et  $y_{k+1} := x_k$ ,
- Sinon,  $x_{k+1} := y_k$  et  $y_{k+1} = a_{k+1} + \Delta_{k+2}$ .

## 2.5 Recyclage sophistiqué : la méthode de Fibonacci

C'est la méthode la plus compliquée (une sorte de mixture entre la stratégie du petit epsilon et le recyclage), mais c'est celle qui permet un nombre minimal d'évaluations de la fonction  $f$  pour une *précision finale donnée*.



On part donc de la précision finale (appelons  $k = N$  l'étape finale), qui correspond à la longueur  $\Delta_N$ . En posant  $F_k := \Delta_{N-k}$ , on a  $F_0 = \Delta_N$  et  $F_{k+2} = F_{k+1} + F_k$ . La suite  $F_k$  est donc de type "Fibonacci".

L'idée de la méthode de Fibonacci est qu'on n'a pas besoin de considérer le principe du "recyclage" pour la dernière étape  $k = N$  (puisque'on arrive à la fin du processus). On va donc optimiser la découpe de l'avant-dernier intervalle  $\Delta_{N-1}$  en suivant la méthode du petit epsilon :

$$\Delta_N = \frac{1}{2}\Delta_{N-1} + \varepsilon.$$

Cette dernière relation implique que

$$F_1 = \Delta_{N-1} = 2(\Delta_N - \varepsilon).$$

En pratique, on commence donc par construire la suite  $F_k$  (elle est uniquement déterminée par les deux premiers termes  $F_0, F_1$ , et la relation de récurrence), puis on prend cette suite "à l'envers" pour obtenir la suite  $\Delta_k$ . Enfin, on construit les  $x_k, y_k$  comme dans la méthode du nombre d'or.

## II Optimisation multidimensionnelle sans contraintes

Dans cette seconde partie, on cherche à minimiser la valeur d'une fonction  $f$  à plusieurs variables réelles, sans contraintes :

$$f : \mathbb{R}^n \rightarrow \mathbb{R}.$$

On note  $\mathbf{x} = (x_1, \dots, x_n)$  les vecteurs de  $\mathbb{R}^n$ . Dans les exercices en pratique, on prendra souvent  $n = 2$ , et dans ce cas les coordonnées pourront aussi être notées  $\mathbf{x} = (x, y)$ .

### 1 Méthode de Newton

La méthode de Newton vue dans la partie 1 peut être généralisée aux fonctions à plusieurs variables. Dans ce cas :

- La dérivée  $f'(x)$  devient le *vecteur gradient* :

$$\nabla_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\mathbf{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\mathbf{x}) \end{bmatrix}$$

- La dérivée seconde  $f''(x)$  devient la *matrice hessienne*

$$H_f(\mathbf{x}) = \left[ \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right]$$

(la notation ci-dessus signifie que l'entrée  $(i, j)$  de la matrice est donnée par le coefficient  $\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x})$ ).

On rappelle que la matrice hessienne est toujours symétrique :

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}).$$

Finalement, la formule de récurrence établie dans la section 1.1 devient :

$$\mathbf{x}_{k+1} := \mathbf{x}_k - H_f^{-1}(\mathbf{x}) \cdot \nabla f(\mathbf{x}).$$

La suite  $\mathbf{x}_k$  ainsi construite itérativement à partir d'un point initial  $\mathbf{x}_0$  converge quadratiquement vers un point sur lequel le vecteur gradient s'annule. En supposant que la fonction  $f$  possède un unique minimum local, c'est bien en ce point critique que  $f$  atteint son minimum.

### Remarque

appelons une formule utile pour calculer l'inverse d'une matrice de taille  $2 \times 2$  :

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

## 2 Méthodes de descente directionnelle

### 2.1 Idée générale

L'idée est de se ramener à des problèmes d'optimisation unidimensionnelle, en recherchant le minimum de la fonction  $f$  selon différentes directions...

On part d'un point  $\mathbf{x}_0$ , et on détermine à partir ce point une direction  $\mathbf{d}_1$  le long de laquelle la valeur de  $f$  diminue. En utilisant une méthode de recherche unidimensionnelle (comme vu en première partie), on trouve le minimum  $\mathbf{x}_1$  de  $f$  sur la droite passant par  $\mathbf{x}_0$  et dirigée par  $\mathbf{d}_1$  (cette droite est paramétrée par  $\lambda \mapsto \mathbf{x}_0 + \lambda \mathbf{d}_1$ ).

Puis à partir du point  $\mathbf{x}_1$ , on détermine de nouveau une direction  $\mathbf{d}_2$  le long de laquelle  $f$  diminue, *etc...* On continue le processus itérativement et on construit ainsi une suite  $\mathbf{x}_k$  qui se rapproche du minimum. On s'arrête lorsque le critère d'arrêt que l'on s'est préalablement fixé est atteint.

### 2.2 Méthode du gradient

En partant d'un point  $\mathbf{x} \in \mathbb{R}^n$ , la direction le long de laquelle  $f$  diminue le plus vite (également appelée "plus forte pente") est donnée par le vecteur  $-\nabla f(\mathbf{x})$ . (Mathématiquement, cela provient du fait que la dérivée directionnelle de  $f$  en  $\mathbf{x}$  suivant une direction  $\mathbf{d}$  est égale au produit scalaire  $\langle \nabla f(\mathbf{x}), \mathbf{d} \rangle$ .)

La méthode du gradient consiste donc à choisir  $\mathbf{d}_{k+1} = -\nabla f(\mathbf{x}_k)$ .

### 2 .3 Méthode des coordonnées cycliques

On choisit les directions données par les axes de coordonnées, en les faisant varier périodiquement (ou "cycliquement"), d'où le nom "coordonnées cycliques". Par exemple, en dimension 2, on pose

$$\mathbf{d}_1 = (1, 0), \mathbf{d}_2 = (0, 1), \mathbf{d}_3 = (1, 0), \text{ etc...}$$

En dimension 3, on pose

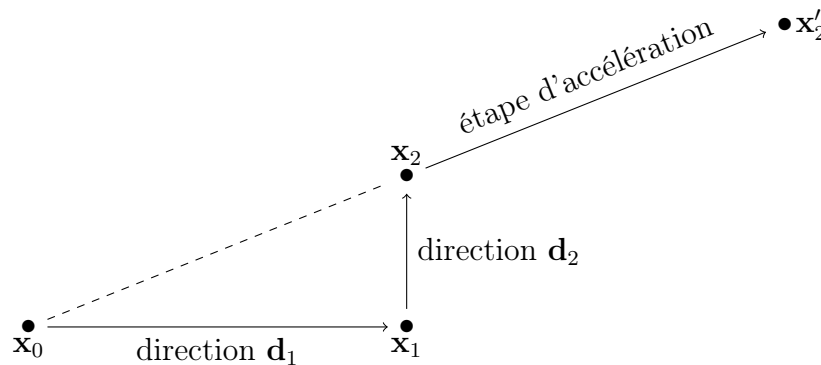
$$\mathbf{d}_1 = (1, 0, 0), \mathbf{d}_2 = (0, 1, 0), \mathbf{d}_3 = (0, 0, 1), \mathbf{d}_4 = (1, 0, 0), \text{ etc...}$$

Un inconvénient de cette méthode est que l'on peut se retrouver "coincé", ou du moins fortement ralenti, au fond d'une vallée étroite... Par exemple en dimension 2, au fond d'une vallée orientée dans la direction  $(1, 1)$ , on ne se déplacera que très lentement dans chaque direction  $(1, 0)$  et  $(0, 1)$ .

### 2 .4 Méthode de Hooke and Jeeves

Pour régler ce problème, l'idée de la méthode de Hooke et Jeeves est d'intercaler une étape dite "d'accélération" après chaque cycle d'exploration (i.e. chaque passage des  $n$  coordonnées).

La direction  $\mathbf{d}$  fixée pour cette étape d'accélération est égale à la direction "globale" du cycle qui vient d'être effectué... Autrement dit,  $\mathbf{d}$  est le vecteur obtenu en faisant la différence "point obtenu à la fin du dernier cycle" - "point obtenu à la fin de l'avant-dernier cycle"... Voir le dessin ci-dessous pour une illustration en dimension 2.



## 3 Méthode de Hooke and Jeeves à pas discret

On reprend cette idée d'alterner des phases d'exploration par coordonnées cycliques avec des étapes d'accélération... mais cette fois, au lieu de faire une recherche unidimensionnelle du minimum à chaque étape (ce qui est très coûteux en énergie), on saute simplement d'un "pas discret" préalablement fixé. Cette méthode nécessite en général plus d'étapes que la méthode précédente, mais chaque étape coûte beaucoup moins cher...

Avant de commencer l'exécution de l'algorithme, on fixe :

- un point de départ  $\mathbf{x}_0 \in \mathbb{R}^n$ ,
- le pas discret  $\Delta > 0$ ,
- un facteur d'accélération  $\alpha \geq 1$ ,
- la précision finale  $\varepsilon > 0$ .

Première phase d'exploration (coordonnées cycliques) :

- Si  $f(\mathbf{x}_0) > f(\mathbf{x}_0 + \Delta \mathbf{d}_1)$ , alors  $\mathbf{x}_1 := \mathbf{x}_0 + \Delta \mathbf{d}_1$
- Sinon, si  $f(\mathbf{x}_0) > f(\mathbf{x}_0 - \Delta \mathbf{d}_1)$ , alors  $\mathbf{x}_1 := \mathbf{x}_0 - \Delta \mathbf{d}_1$
- Sinon,  $\mathbf{x}_1 := \mathbf{x}_0$ .

On continue la phase d'exploration en se dépla

### Exercices de réflexion :

cant suivant les autres directions  $\mathbf{d}_2, \dots, \mathbf{d}_n$ . On complète ainsi la suite  $\mathbf{x}_2, \dots, \mathbf{x}_n$  itérativement.

À la fin de la phase d'exploration :

- Si le point  $\mathbf{x}_n$  obtenu après avoir exploré les  $n$  directions n'a pas bougé ( $\mathbf{x}_n = \mathbf{x}_0$ ), alors on recommence une nouvelle phase d'exploration en divisant la longueur du pas par deux :  $\Delta \leftarrow \Delta/2$ .
- Sinon (si  $\mathbf{x}_n \neq \mathbf{x}_0$ ), on passe à une étape d'accélération dans la direction  $\mathbf{x}_n - \mathbf{x}_0$  : on pose  $\mathbf{x}'_n := \mathbf{x}_n + \alpha(\mathbf{x}_n - \mathbf{x}_0)$ . Et on recommence une nouvelle phase d'exploration à partir du point  $\mathbf{x}'_n$ .

Puisque la longueur du pas  $\Delta$  diminue au cours de l'algorithme et représente le degré de précision de la recherche, on s'arrête lorsque  $\Delta < \varepsilon$ .