Méthodes d'optimisation

David Hébert
hebert.iut@gmail.com

2023

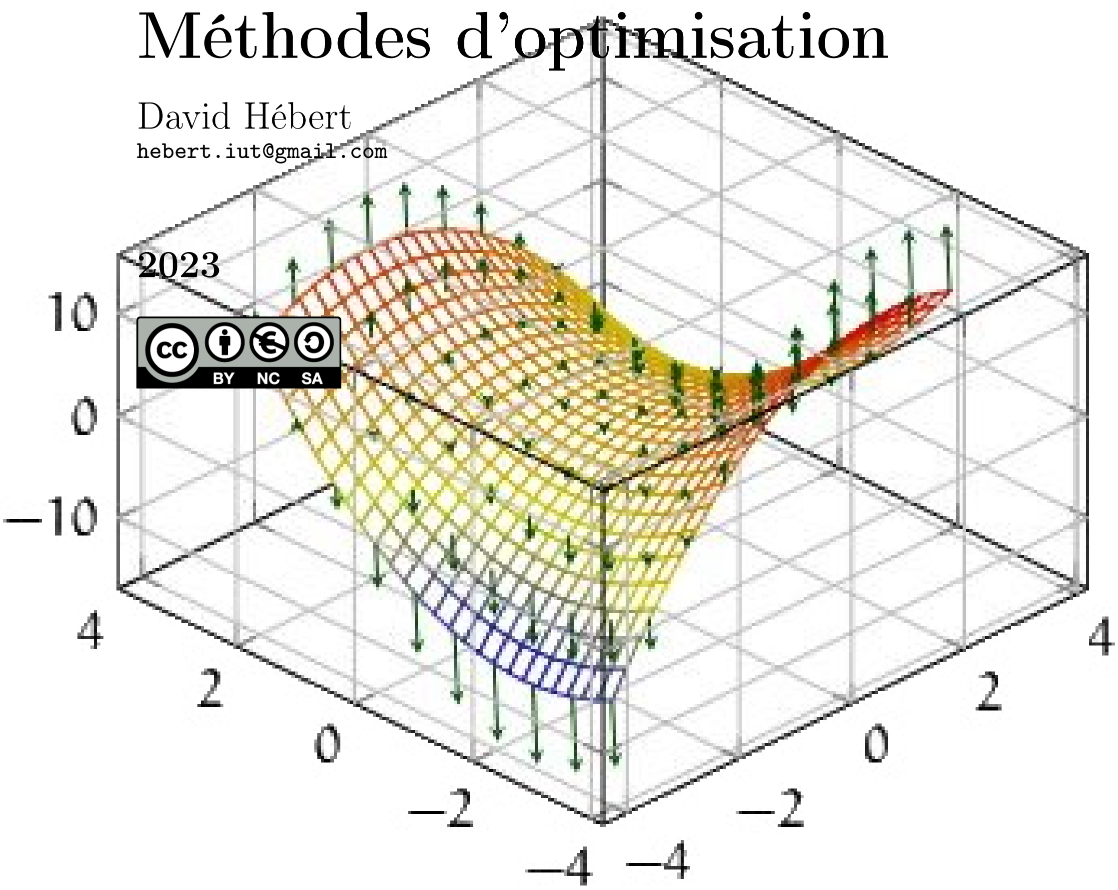


Table des matières

1. Introduction

L'optimisation est le processus consistant à trouver les meilleurs paramètres d'une fonction en fonction d'un certain critère, ou à atteindre un objectif donné, en utilisant des méthodes mathématiques et algorithmiques.

Exemple 1 : Quelle sont les dimensions d'un rectangle de périmètre 10 ayant la plus grand aire possible ?

Il se trouve qu'un carré sera la solution maximisera l'aire !

Exemple 2 : Je dispose de 100€ et de $2\mathrm{h}$ soit 120 minutes de temps à perdre et je souhaite acheté :

- Soit des cartes Pokémon valant 3€ l'unité mais me satisfaisant qu'à 1 (pour une certaine unité de *satisfaction*). L'achat d'une carte me prenant 1 minutes.
- Soit des bonbons valant 1€ l'unité mais me satisfaisant à 2. L'achat d'un bonbon me prenant 7 minutes.

Comment dois-je dépensez mon argent pour avoir la plus grande satisfaction ?

Problème un peu plus difficile mais nous pourront démontrer que 29 cartes pokémon et 13 bonbons sera le choix le plus satisfaisant ?

Mais comment avons nous résolu ces problèmes ?

Pour le premier, cela se fait à l'aide d'un polynôme de degrés 2. Il fait partie de la catégorie des problèmes dis *non linéaire* (car il y a un x^2).

Le second fait partie des problèmes dis *linéaire.*

Dans le cadre de ce cours, nous allons nous intéresser à la résolution de ces problèmes. Cela nous amènera à introduire les fonctions à plusieurs variables qui sera le cœur mathématiques de ce cours.

2. Problème linéaire à deux variables

2.1 Exemples de problème

1. Une usine produit deux modèles de machines, l'une que l'on appellera modèle A exige 2 kg de matière première et de 30 heures de fabrication et donne un bénéfice de 7€. L'autre que l'on appellera B exige 4 kg de matière première et de 15 heures de fabrication et donne un bénéfice de 6€. On dispose de 200 kg de matière première et de 1200 h de travail.

   Quelle production doit on avoir pour obtenir un bénéfice maximal ?
2. Renault fabrique des voitures et des camions. Chaque véhicule doit être traité dans l'atelier de peinture et dans l'atelier de carrosserie. La capacité de l'atelier de peinture permet de traiter 40 camions par jour (si l'on ne peint que des camions), ou 60 voitures par jour (si l'on ne peint que des voitures). De la même façon la capacité de l'atelier de carrosserie est limitée à 50 camions par jour et à 50 voitures par jour. Chaque camion produit rapporte 3000€, et chaque voiture 1000€.

   Quel plan de production quotidien doit adopter Renault pour maximiser le profit de l'entreprise ?

   Et si en plus, le fabriquant impose à ses employés de faire sortir de son usine au moins 30 camions et 20 voitures par jour ?

2.2 Mise en inéquation

Reprenons les exemples précédents et traduisons les en langage mathématiques.

1. La question est de savoir combien de produit A et B peut-on produire au maximum. Notons simplement A le nombre de produit A et B le nombre de produit B. Les contraintes de l'énoncé se traduisent ainsi :
   - **Domaine :** A et B sont des entiers positifs
   - **La matière première :** 2 kg pour le produit A, 4 kg pour le produit B et le tout ne doit pas dépasser la quantité maximale de matière première soit 200 kg :

     $$2A + 4B \leqslant 200$$

   - **Temps de fabrication :** 30 h pour le produit A, 15 h pour le produit B ; le tout ne dépassant pas 1200 h :

     $$30A + 15B \leqslant 1200$$

   - **Le bénéfice :** de 7€, pour le A et de 6€, pour le B. Bien sur on souhaite que ce bénéfice soit maximal ; on cherche donc A et B tel que

     $$7A + 6B$$

     soit maximal.
2. On note c le nombre de camion produit quotidiennement et v le nombre de voiture.
   - **Domaine :** c et v sont des entiers positifs.
   - **Capacité des ateliers :** en un jour on ne peut prendre qu'un maximum de 40 camions dans l'atelier de peinture. Le temps de traitement d'un camion est de $\frac{1}{40}$ de jour et par la même méthode le temps de traitement d'une voiture est $\frac{1}{60}$ de jour. La contrainte sur la capacité de l'atelier de peinture s'exprime donc de la manière suivante

     $$\frac{1}{40}c + \frac{1}{60}v \leqslant 1$$

     En raisonnant de même, la contrainte sur l'atelier de carrosserie s'exprime par

     $$\frac{1}{50}c + \frac{1}{50}v \leqslant 1$$

• **Le bénéfice :** on doit le maximiser c'est à dire trouver la valeur maximale de

$$3000c + 1000v$$

•**... et si en plus :** la condition supplémentaire se traduit simplement par

$$c \geqslant 30, \qquad v \geqslant 20$$

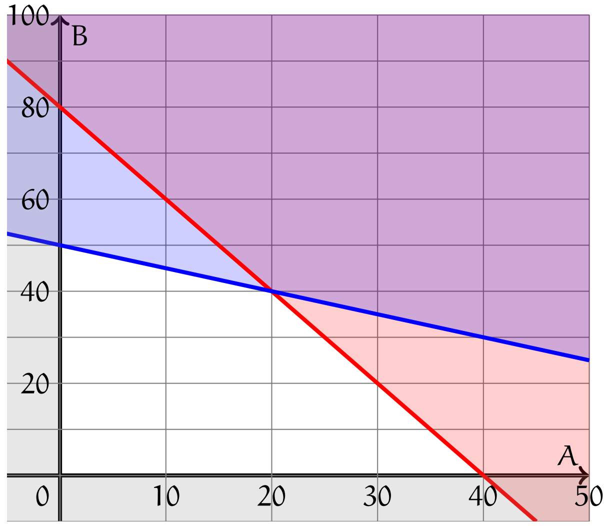
Avant de résoudre ces problèmes en nombres entiers, c'est à dire en satisfaisant toutes les contraintes de domaine, nous allons détailler les méthodes qui permettent de les résoudre en continue, c'est à dire dans $\mathbb{R}$.

2.3 Résolution graphique

Lorsque le problème posé fait apparaitre deux variables on peut développer une méthode de résolution (en continue) par analyse graphique. Prenons l'exemple 1. La formulation mathématique du problème a fait apparaitre 4 inéquations du plan :

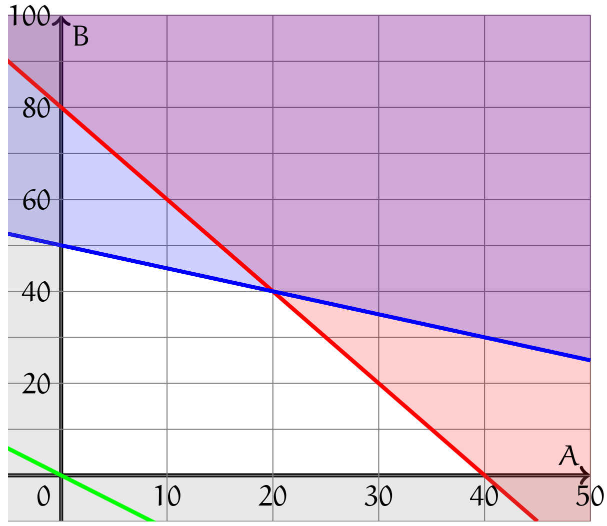
$$\begin{aligned} A &\geqslant 0 \\ B &\geqslant 0 \\ 2A + 4B &\leqslant 200 \\ 30A + 15B &\leqslant 1200 \end{aligned}$$

Représentons ce domaine (A en abscisse et B en ordonnée) ; les droites dessinées correspondent aux droites d'équations $A \longmapsto B = \dfrac{200 - 2A}{4}$ et $A \longmapsto B = \dfrac{1200 - 30A}{15}$



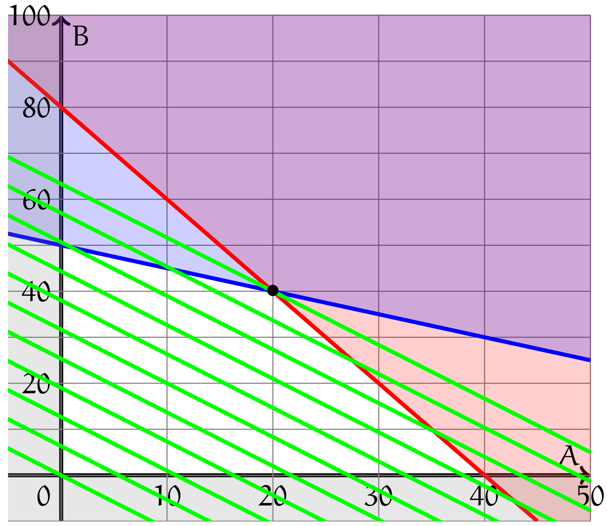
On observe que ce domaine défini un polyèdre convexe (la partie la plus claire). En fait ça sera toujours le cas.

On rappel que le problème est de trouver A et B de sorte que le bénéfice $7A + 6B$ soit maximal. Imaginons que nous souhaitons un bénéfice nul (pourquoi pas ?). Cela se traduit par le fait de trouver A et B dans le domaine convexe tel que $7A + 6B = 0$. On trace donc la droite $A \longmapsto B = \dfrac{0 - 7A}{6}$.



On observe que cette droite touche le domaine en l'origine, c'est à dire lorsque $A = B = 0$; ce qui est normal : on ne produit rien, on a donc aucun bénéfice.

Cela nous amène alors à considérer la familles de droites $\left\{A \longmapsto B = \dfrac{M - 7A}{6}\right\}_{M \in \mathbb{R}_+}$ qui sont toutes parallèles (puisqu'elles possèdent le même coefficient directeur). Lorsque le M grandit, c'est à dire lorsque le bénéfice augmente, la droite se translate vers le haut puisque le M représente l'ordonnée à l'origine.



Par lecture graphique on observe donc que la valeur maximale de M est atteint au point d'intersection des droites $A \longmapsto B = \dfrac{200-2A}{4}$ et $A \longmapsto B = \dfrac{1200-30A}{15}$; ceci se produit lorsque $A = 20$ donc $B = 40$ et $M = 7A + 6B = 380$.

C'est la solution optimale, car si le bénéfice augmente on sort du domaine imposé par les contraintes.

En procédant de la sorte pour n'importe qu'elle type de tel problème on en déduit nécessairement que la solution est sur "coin" du polyèdre convexe.

Remarque 2.3.1 : Lorsque l'on est face à un tel problème (linéaire et ne faisant apparaitre que deux variables) on représente le domaine et on détermine les "coins" du polyèdre.

Reprenons notre exemple :

$$\begin{array}{rcl} A & \geqslant & 0 \\ B & \geqslant & 0 \\ 2A+4B & \leqslant & 200 \\ 30A+15B & \leqslant & 1200 \\ \mathrm{Max}(7A+6B) & & \end{array}$$

Pour chaque "coin" on calcul la valeur de la fonction à optimiser, dans notre exemple $7A + 6B$ et on cherche la plus grande valeur (car il s'agit d'un maximum, si c'est un minimum on va bien sur prendre le minimum).

(x, y)	$7A + 6B$
$(0, 0)$	0
$(40, 0)$	280
$(0, 50)$	300
$(20, 40)$	380

On arrive donc à la même conclusion que précédemment : il faut produire 20 produit A et 40 produit B pour obtenir un bénéfice maximum de 380€.

3. Vers les fonctions a plusieurs variables

Bien connue des élèves de terminales revenons de manière sibylline sur les fonctions à une variables Voici la définition donnée par L. Euler.

> Si certaines quantités dépendent d'autres quantités de telle manière que si les autres changent, ces quantités changent aussi, alors on a l'habitude de nommer ces quantités fonctions de ces dernières ; cette dénomination a la plus grande étendue et contient en elle-même toutes les manières par lesquelles une quantité peut être déterminée par d'autre. Si, par conséquent, x désigne une quantité variable, alors toutes les autres quantités qui dépendent de x de n'importe quelle manière, ou qui sont déterminées par x, sont appelées fonctions de x.

3.1 Vocabulaire

L'image d'un réel x par une fonction f est le réel $f(x)$. Avec la fonction $f(x) = 3x + 4$, au lieu de dire "la fonction f transforme le nombre réel -1 en 1" on dira "l'image de -1 par f est 1" ce qui est donc la même chose que d'écrire $f(-1) = 1$.

Un antécédent d'un réel y par une fonction f est un réel x tel que $f(x) = y$. Avec la fonction $f(x) = 3x+4$, au lieu de dire "la fonction f transforme le nombre réel -1 en 1" on dira "un antécédent 1 par f est -1" ce qui est donc la même chose que d'écrire $f(-1) = 1$.

le domaine de définition est l'ensemble des valeurs possible pour la variable dans la définition de la fonction.

Si par exemple $f(x) = \dfrac{\sqrt{x-10}}{x-20}$ alors il faut d'une part ne pas annuler le dénominateur, ce qui n'arrive que lorsque $x = 20$ et ne jamais avoir $x - 10<0$ ce qui arrive pour les réels de l'intervalle $]-\infty; 10[$. En définitive il faut enlever à $\mathbb{R}$ ces impossibilités. On en conclut alors que le domaine de définition de cette fonction est $[10; 20[\cup]20; +\infty[$.

Attention ! On parle bien de LA image mais de UN antécédent. En effet, si on prend un nombre réel x dans le domaine de définition alors il a une et une seule image. Tandis que la recherche d'antécédent donne naissance à une équation qui peut avoir aucune, une, voir plusieurs solutions ! Par exemple, si g est une fonction définie par $g(x) = x^2 - 3x$ et que nous cherchons les antécédents (éventuels) de -2 alors nous sommes amené, par définition d'antécédent, à résoudre l'équation $g(x) = -2$ soit encore $x^2 - 3x = -2$ soit encore $x^2-3x+2 = 0$ ce qui se résout à l'aide d'un discriminant et aboutis à 1 et 2 comme solution. D'ailleurs on vérifie sans peine que $g(1) = 1^2 - 3(1) = -2$ et que $g(2) = 2^2 - 3(2) = -2$.

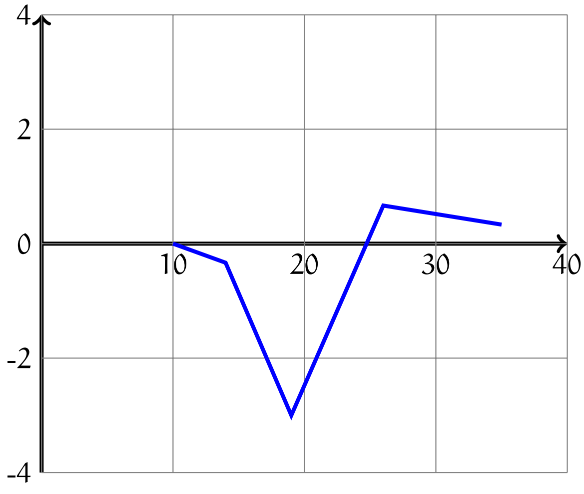
3.2 Représentation

Pour travailler avec les fonctions, il est d'accoutumer de les représenter dans un repère cartésien. Pour chaque valeur de x, dans le domaine de définition de la fonction, on marque, d'un point (ou d'une croix ou peu importe) le point du repère de coordonnée $(x, f(x))$ et on relie les points consécutifs.

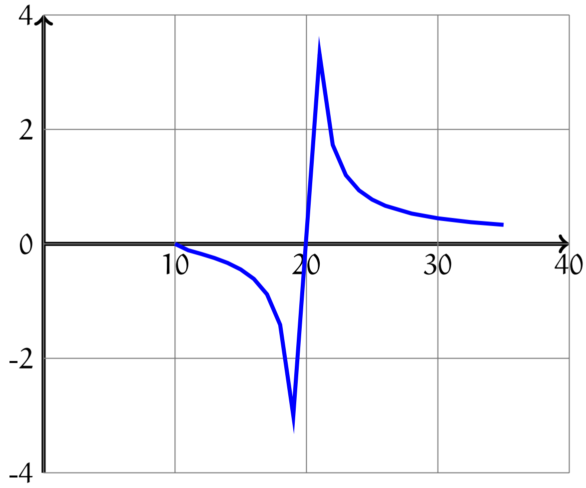
Reprenons l'exemple de la fonction $f(x) = \dfrac{\sqrt{x-10}}{x-20}$. On prend différente valeur et on place ces valeurs dans le repère :

x	10	14	19	26	35
$f(x)$	0	$-\dfrac{1}{3}$	-3	$\dfrac{2}{3}$	$\dfrac{1}{3}$

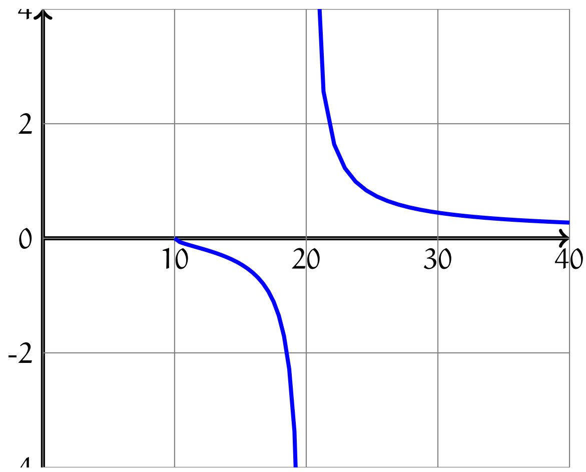
Ce qui correspond aux points suivants :



Plus on va avoir de point, plus la fonction sera dessiner de manière précise.



Avec un peu plus de point on arrive à ce qui est appelé le *graphe* de la fonction f ou *courbe représentative de la fonction* f :



3.3 Limites

Si une fonction n'a pas de valeurs interdites alors elle n'a pas de problème pour le calcul de ses limites. La notion qui est caché derrière ce *bon comportement* des fonctions est la notion de continuité.

Théorème

Soit f une fonction définie (et continue) en $a \in \mathbb{R}$ dans le domaine de définition de la fonction, alors

$$\lim_{x \to a} f(x) = f(a)$$

Les cas problématiques viennent en générale des infinis.

Théorème

$$\lim_{x \to 0^+} \frac{1}{x} = +\infty \qquad \lim_{x \to 0^-} \frac{1}{x} = -\infty$$

$$\lim_{x \to +\infty} \frac{1}{x} = 0 \qquad \lim_{x \to -\infty} \frac{1}{x} = 0$$

Inversion de limites. Inverser une limite est relativement facile si on garde en tête que l'inverse de 0 c'est de l'infini et que l'inverse de l'infini c'est 0.

$\lim_{x\to a} f(x)$	$\lim_{x\to a} \frac{1}{f(x)}$
$l \neq 0$	$\frac{1}{l}$
0^+	$+\infty$
0^-	$-\infty$
$\pm\infty$	0

Somme de limite.

$\lim_{x\to a} f(x)$	$\lim_{x\to a} g(x)$	$\lim_{x\to a} f(x) + g(x)$
l	l'	$l + l'$
l	$\pm\infty$	$\pm\infty$
$+\infty$	$+\infty$	$+\infty$
$-\infty$	$-\infty$	$-\infty$
$+\infty$	$-\infty$	F.I.

Produit d'une limite par un nombre.

$\lim_{x\to a} f(x)$	$\lim_{x\to a} \lambda f(x)$
l	λl
$+\infty$	$+\infty$ si $\lambda > 0$
$+\infty$	$-\infty$ si $\lambda < 0$
$-\infty$	$-\infty$ si $\lambda > 0$
$-\infty$	$+\infty$ si $\lambda < 0$

Produit de deux limites.

$\lim_{x\to a} f(x)$	$\lim_{x\to a} g(x)$	$\lim_{x\to a} f(x) \times g(x)$
l	l'	$l \times l'$
$+\infty$	$+\infty$	$+\infty$
$-\infty$	$-\infty$	$+\infty$
$+\infty$	$-\infty$	$-\infty$
$l > 0$	$+\infty$	$+\infty$
$l < 0$	$+\infty$	$-\infty$
$l > 0$	$-\infty$	$-\infty$
$l < 0$	$-\infty$	$+\infty$
0	$\pm\infty$	F.I.

Quotient de deux limites. Un quotient n'est rien d'autre qu'un produit avec un inverse : $\frac{f(x)}{g(x)} = f(x)\frac{1}{g(x)}$.

Retenez les 4 formes indéterminées :

$$\infty - \infty \qquad 0 \times \infty \qquad \frac{0}{0} \qquad \frac{\infty}{\infty}$$

3.4 Dérivation

Définition

Soit f une fonction définie en $a \in \mathbb{R}$ de son domaine de définition. Si

$$\lim_{x \to a^+} \frac{f(x) - f(a)}{x - a} = \lim_{x \to a^-} \frac{f(x) - f(a)}{x - a}$$

et que ces deux limites sont finies, alors on appellera la valeur de ces limites la *dérivé de la fonction* f *en* a que l'on notera $f'(a)$. Précisément :

$$f'(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}$$

Quelque soit la fonction que nous traitons, cette limite est toujours une forme indéterminée (de la forme $0/0$). Ca serait beaucoup moins intéressant sinon ^_^.

Par exemple dérivons en $a = 3$ la fonction $f(x) = x^2$. Revenons à la définition et essayons de simplifier la limite, qui est malheureusement par construction, toujours une forme indéterminée. Ici l'idée est d'utiliser une identité remarquable.

$$\begin{aligned} f'(3) &= \lim_{x \to 3} \frac{x^2 - 3^2}{x - 3} \\ &= \lim_{x \to 3} \frac{(x-3)(x+3)}{x-3} \\ &= \lim_{x \to 3} (x+3) \\ &= 6 \end{aligned}$$

Conclusion $f'(3) = 6$.

Nous venons de voir la définition qui en tant que tel n'est pas évidente de prime abord. Et nous espérons que vous l'aurez compris, ça sert à calculer la vitesse instantanée. Mais tout ce travail pour calculer une vitesse serait un peu trop de gâchis de concept mathématiques. Puis surtout, en pratique comme procède t-on au calcul ? Les réponses arrivent !

3.5 Tangente

Définition

L'équation de la tangente à la courbe représentative d'une fonction f définie en un réel a est donnée par

$$T_a \quad : \quad Y = f'(a)(X - a) + f(a)$$

Par exemple la tangente de la fonction carré en $a = 3$ est $T_3 : \; y = 6(x-3) + 3^2 = 6x - 9$

3.6 Dérivées usuelles et fonctions dérivées

Dans le calcul que nous avions fait dans les paragraphes précédents pour calculer la dérivée de la fonction carré en 3, que l'on ai mis un 3 ou un 42 ça ne change pas le travail :

$$\begin{aligned} f'(a) &= \lim_{x \to a} \frac{x^2 - a^2}{x - a} \\ &= \lim_{x \to a} \frac{(x-a)(x+a)}{x-a} \\ &= \lim_{x \to a} (x+a) \\ &= 2a \end{aligned}$$

Le cas de la fonction carré est réglé ! C'est l'occasion de parler de **fonction dérivée**.

Définition

La fonction dérivée d'une fonction f est la fonction notée f' tel que la valeur de f' en a est le nombre dérivé de la fonction f en a.

Nous avons par exemple montré que la fonction dérivée de la fonction carré est $f'(x) = 2x$.

Qu'en est-il de la fonction cube $g(x) = x^3$. Abracadabra ! Voici une formule sorti de nul part : $x^3 - y^3 = (x - y)(x^2 + xy + y^2)$. Attaquons à présent le calcul de la limite :

$$\begin{aligned} g'(a) &= \lim_{x \to a} \frac{x^3 - a^3}{x - a} \\ &= \lim_{x \to a} \frac{(x - a)(x^2 + ax + a^2)}{x - a} \\ &= \lim_{x \to a} (x^2 + ax + a^2) \\ &= 3a^2 \end{aligned}$$

En conclusion $g'(x) = 3x^2$.

Commençons par faire peur : OUI ! Pour calculer des dérivées (en tant que fonction ou valeur numérique) il n'y a pas d'autre solution que de passer par la définition de limite ! Mais pas d'inquiétude. Beaucoup de mathématiciens ont établis un formulaire recensant les dérivées des fonctions usuelles (et oui, il faut les apprendre et les connaitre par cœur) :

Proposition

Soit n un entier strictement positif et k un nombre réel alors

$$\begin{aligned} f(x) = k &\quad \Rightarrow \quad f'(x) = 0 \\ f(x) = x^n &\quad \Rightarrow \quad f'(x) = nx^{n-1} \\ f(x) = \frac{1}{x^n} &\quad \Rightarrow \quad f'(x) = -\frac{n}{x^{n+1}} \\ f(x) = \sqrt{x} &\quad \Rightarrow \quad f'(x) = \frac{1}{2\sqrt{x}} \end{aligned}$$

On retrouve par exemple les formules que nous avons trouvé à l'aide du calcul de limite sur les fonctions carré et cube.

3.7 Opérations sur les dérivées

Et si nous souhaitions dériver la fonction $x + 1$? Nous savons dérivé $x = x^1$ qui admet $1x^0 = 1$ comme dérivé et la dérivé du nombre réel 1 est 0. Qu'en est-il de la somme ?

Proposition

$$(f + g)' = f' + g'$$

Démonstration. En passant à la limite dans cette formule on prouve le résultat.

$$\frac{[f(x) + g(x)] - [f(a) + g(a)]}{x - a} = \frac{f(x) - f(a)}{x - a} + \frac{f(x) - f(a)}{x - a}$$

□

Ainsi la fonction dérivée de $x^2 + x + 1$ est $2x + 1$.

Qu'en est-il de la fonction dérivée de la $x^2 - 3x + \frac{1}{x}$. Les précédents résultats nous donnent

$$\left(x^2 - 3x + \frac{1}{x}\right)' = \left(x^2\right)' - (3x)' + \left(\frac{1}{x}\right)' = 2x - (3x)' - \frac{1}{x^2}$$

Réglons le problème de $(3x)'$ à coup de théorème !

Proposition

$$(\lambda f)' = \lambda f'$$

Démonstration. En passant à la limite dans cette formule on prouve le résultat.

$$\frac{\lambda f(x) - \lambda f(a)}{x - a} = \lambda \frac{f(x) - f(a)}{x - a}$$

□

En d'autre terme la dérivé de $3x$ c'est 3 fois la dérivée de x (qui est 1). Au final

$$\left(x^2 - 3x + \frac{1}{x}\right)' = 2x - 3 - \frac{1}{x^2}$$

Et qu'en est-il de la fonction dérivée de $(\sqrt{x} + 1) \times \frac{1}{x}$. Bon soyons claire : ca va commencer a devenir n'importe quoi, mais ça marche !

Proposition

$$(fg)' = f'g + g'f$$

Démonstration. En passant à la limite dans cette égalité on prouve le résultat :

$$\begin{aligned}
\frac{f(x)g(x) - f(a)g(a)}{x - a} &= \frac{f(x)g(x) - f(x)g(a) + f(x)g(a) - f(a)g(a)}{x - a} \\
&= \frac{f(x)g(x) - f(x)g(a)}{x - a} + \frac{f(x)g(a) - f(a)g(a)}{x - a} \\
&= f(x)\frac{g(x) - g(a)}{x - a} + g(a)\frac{f(x) - f(a)}{x - a}
\end{aligned}$$

□

Ainsi

$$\begin{aligned}
\left((\sqrt{x} + 1)\frac{1}{x}\right)' &= (\sqrt{x} + 1)'\frac{1}{x} + (\sqrt{x} + 1)\left(\frac{1}{x}\right)' \\
&= \frac{1}{2\sqrt{x}}\frac{1}{x} + (\sqrt{x} + 1)\left(-\frac{1}{x^2}\right)
\end{aligned}$$

En conclusion $\left((\sqrt{x} + 1)\frac{1}{x}\right)' = \frac{1}{2x\sqrt{x}} - \frac{\sqrt{x} + 1}{x^2}$ (les plus savant pour chercher à simplifier davantage cette dérivée et montrer que $\left((\sqrt{x} + 1)\frac{1}{x}\right)' = -\frac{\sqrt{x} + 2}{2x^2}$ en utilisant le fait que $\sqrt{x}^2 = x$).

Et qu'en est-il de la fonction dérivée de la fonction $\frac{x}{x + 1}$?

Proposition

$$\left(\frac{f}{g}\right)' = \frac{f'g - g'f}{g^2}$$

Démonstration. En passant à la limite dans cette égalité on prouve le résultat :

$$\begin{aligned}
\frac{\frac{f(x)}{g(x)} - \frac{f(a)}{g(a)}}{x-a} &= \frac{\frac{f(x)g(a)}{g(x)g(a)} - \frac{g(x)f(a)}{g(x)g(a)}}{x-a} \\
&= \frac{\frac{f(x)g(a) - g(x)f(a)}{g(x)g(a)}}{x-a} \\
&= \frac{f(x)g(a) - g(x)f(a)}{g(x)g(a)(x-a)} \\
&= \frac{f(x)g(a) - f(a)g(a) + f(a)g(a) - g(x)f(a)}{g(x)g(a)(x-a)} \\
&= \frac{f(x)g(a) - f(a)g(a)}{g(x)g(a)(x-a)} + \frac{f(a)g(a) - g(x)f(a)}{g(x)g(a)(x-a)} \\
&= \frac{f(x)g(a) - f(a)g(a)}{g(x)g(a)(x-a)} - \frac{f(a)g(x) - g(a)f(a)}{g(x)g(a)(x-a)} \\
&= g(a)\frac{f(x) - f(a)}{g(x)g(a)(x-a)} - f(a)\frac{g(x) - g(a)}{g(x)g(a)(x-a)} \\
&= \frac{g(a)\frac{f(x)-f(a)}{x-a} - f(a)\frac{g(x)-g(a)}{x-a}}{g(x)g(a)}
\end{aligned}$$

□

Sur notre exemple

$$\begin{aligned}
\left(\frac{x}{x+1}\right)' &= \frac{(x)'(x+1) - (x+1)'(x)}{(x+1)^2} \\
&= \frac{1(x+1) - 1(x)}{(x+1)^2} \\
&= \frac{1}{(x+1)^2}
\end{aligned}$$

Et qu'en est-il de la fonction dérivé de $\sqrt{x^2+x}$?

Proposition

$$(f(g))' = f'(g) \times g'$$

Démonstration. On a

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

D'un coté $\lim\limits_{x\to a} \dfrac{g(x) - g(a)}{x-a} = g'(a)$ de l'autre coté en effectuant le changement de variable $y = g(x)$ on a, par définition du nombre dérivée $\lim\limits_{x\to a} \dfrac{f(g(x)) - f(g(a))}{g(x) - g(a)} = \lim\limits_{y\to g(a)} \dfrac{f(y) - f(g(a))}{y - g(a)} = f'(g(a))$. □

Ainsi $(\sqrt{x^2+x})' = \dfrac{1}{2\sqrt{x^2+x}} \times (x^2+x)' = \dfrac{1}{2\sqrt{x^2+x}} \times (2x+1) = \dfrac{2x+1}{2\sqrt{x^2+x}}$

En particulier ce dernier résultat permet d'obtenir un petit peu plus de formule de dérivation (il n'y en avait pas assez comme ça !)

Corollaire

Soient n un entier strictement positif u une fonction

$$f(x) = u(x)^n \qquad \Rightarrow \qquad f'(x) = nu(x)^{n-1} \times u'(x)$$

$$f(x) = \frac{1}{u(x)^n} \qquad \Rightarrow \qquad f'(x) = -\frac{n}{u(x)^{n+1}} \times u'(x)$$

$$f(x) = \sqrt{u(x)} \qquad \Rightarrow \qquad f'(x) = \frac{1}{2\sqrt{u(x)}} \times u'(x)$$

Par exemple déterminons l'équation de la tangente à la courbe représentative de la fonction f définie sur $\mathbb{R}$ par $f(x) = \frac{1}{2}x^2 - x + 1$ en $a = 0$. D'après les paragraphes précédents $T_0 :\ y = f'(0)(x-0) + f(0)$. D'une part $f'(x) = x - 1$ d'où $f'(0) = -1$ et d'autre part $f(0) = 1$. En conclusion l'équation de la tangente est $y = -x + 1$.

3.8 Variations et dérivations

Lorsque l'on se donne une fonction, on souhaite l'étudier. Ce que nous avons déja explorer avec l'étude du domaine de définition ou des translations de fonction de références. D'autre informations son a extraire de la fonction. L'une d'elle est ce que l'on appelle les *variations* de la fonction ce qui se traduit vulgairement par : quand est-ce que la fonction monte et quand-est-ce qu'elle descend.

Définition

- On dira qu'une fonction f est **strictement croissante** sur un intervalle I si pour tout réels α et β de I tel que $\alpha < \beta$ alors $f(\alpha) < f(\beta)$.
- On dira qu'une fonction f est **croissante** sur un intervalle I si pour tout réels α et β de I tel que $\alpha < \beta$ alors $f(\alpha) \leqslant f(\beta)$.
- On dira qu'une fonction f est **strictement décroissante** sur un intervalle I si pour tout réels α et β de I tel que $\alpha < \beta$ alors $f(\alpha) > f(\beta)$.
- On dira qu'une fonction f est **décroissante** sur un intervalle I si pour tout réels α et β de I tel que $\alpha < \beta$ alors $f(\alpha) \geqslant f(\beta)$.

Plaçons nous par exemple sur l'intervalle $]0;+\infty[$ et étudions les variations de la fonction $f(x) = x^2$. Soit $0 < \alpha < \beta$ alors

$$f(\alpha) - f(\beta) = \alpha^2 - \beta^2 = (\alpha - \beta)(\alpha + \beta)$$

Puisque α et β sont tout deux positifs alors $\alpha + \beta > 0$ et puisque $\alpha < \beta$ alors $\alpha - \beta < 0$. La règle des signes nous donne donc que $\alpha^2 - \beta^2$ est négatif soit encore $f(\alpha) < f(\beta)$ et donc que la fonction carré est strictement croissante sur $]0;+\infty[$

On montrerai de même que la fonction carré est strictement décroissante sur $]-\infty;0[$.

Pour compiler ces informations de variation on réalise un *tableau de variation* qui comporte deux lignes : une ligne pour spécifier les valeurs de x classer dans l'ordre croissant, comme pour les tableaux de signe et une seconde ligne représentant les variations par des flèches qui montent pour signaler la croissance ou qui descende pour signaler la décroissance.

Le tableau de variation de la fonction carré est donc le suivant :

x	$-\infty$		0		$+\infty$
x^2		$\searrow$		$\nearrow$	

On complète souvent le tableau de variation en mettant les valeurs des limites au bout des flèches :

x	$-\infty$		0		$+\infty$
x^2	$+\infty$	$\searrow$	0	$\nearrow$	$+\infty$

Cet exemple était assez facile. Les droites, c'est à dire les fonctions affines ont aussi des variations faciles à étudier.

Proposition

Soit f une fonction affine de coefficient directeur a.

- Si $a > 0$ la droite est strictement croissante.

x	$-\infty$		$+\infty$
f	$-\infty$	$\nearrow$	$+\infty$

- Si $a < 0$ la droite est strictement décroissante.

x	$-\infty$		$+\infty$
f	$+\infty$	$\searrow$	$-\infty$

- Si $a = 0$ la droite est constante.

x	$-\infty$		$+\infty$
f	k	$\longrightarrow$	k

où $f(x) = k$

Théorème

- Si f' existe et est positive sur un intervalle I alors la fonction f est croissante sur I.
- Si f' existe et est négative sur un intervalle I alors la fonction f est décroissante sur I.

Autrement dis : étudier les variations d'une fonction reviens à étudier le signe de sa dérivé.

La dérivé de la fonction $f(x) = x^2$ est $f'(x) = 2x$. Naturellement $2x$ est positif lorsque x est positif donc la fonction f sera croissante sur cet intervalle ce que nous avions déjà déterminer.

Exemple

Étudions la fonction $f(x) = x + \dfrac{x}{x^2 - 1}$.

Nous voyons une fraction, il faut donc que le dénominateur, $x^2 - 1$, ne s'annule pas. En utilisant une identité remarquable ou une les formules de calculs des racines pour les polynômes de degrés 2, nous pouvons rapidement conclure que le domaine de définition de cette fonction est $\mathbb{R} - \{-1; 1\}$. Le but étant de déterminer le signe de la dérivé, c'est à dire de résoudre une inéquation de la forme $f'(x) > 0$, nous allons calculer la dérivée et la factoriser au maximum.

$$
\begin{aligned}
f'(x) &= 1+\frac{(x)'(x^2-1)-(x^2-1)x}{(x^2-1)^2}\\
&= \frac{(x^2-1)^2+1(x^2-1)-2xx}{(x^2-1)^2}\\
&= \frac{(x^2-1)^2+x^2-1-2x^2}{(x^2-1)^2}\\
&= \frac{x^4-2x^2+1+x^2-1-2x^2}{(x^2-1)^2}\\
&= \frac{x^4-3x^2}{(x^2-1)^2}\\
&= \frac{x^2(x^2-3)}{(x^2-1)^2}
\end{aligned}
$$

Nous allons a présent dresser le tableau de signe et par la même le tableau de variation. La forme est donc la suivante.

x	$-\infty$	$-\sqrt{3}$	-1	0	1	$\sqrt{3}$	$+\infty$
x^2							
x^2-3							
$(x^2-1)^2$							
f'							
f							

Dans le tableau de variation comme dans les tableaux de signe on signale les valeurs interdites par une double barre.

x	$-\infty$	$-\sqrt{3}$	-1	0	1	$\sqrt{3}$	$+\infty$
x^2				0			
x^2-3		0				0	
$(x^2-1)^2$			0		0		
f'		0	$\|\|$	0	$\|\|$	0	
f			$\|\|$		$\|\|$		

En complétant les signes on arrive à

x	$-\infty$		$-\sqrt{3}$		-1		0		1		$\sqrt{3}$		$+\infty$
x^2		$+$		$+$		$+$	0	$+$		$+$		$+$	
x^2-3		$+$	0	$-$		$-$		$-$		$-$	0	$+$	
$(x^2-1)^2$		$+$		$+$	0	$+$		$+$	0	$+$		$+$	
f'		$+$	0	$-$	$\|$	$-$	0	$-$	$\|$	$-$	0	$+$	
f					$\|$				$\|$				

D'après le cours, une dérivée positive donner une fonction croissante et inversement, ce qui permet d'arriver aux variations suivantes :

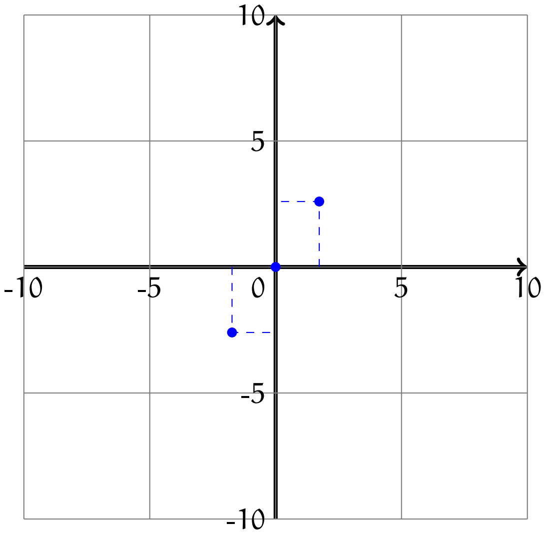
x	$-\infty$		$-\sqrt{3}$		-1		0		1		$\sqrt{3}$		$+\infty$
x^2		$+$		$+$		$+$	0	$+$		$+$		$+$	
x^2-3		$+$	0	$-$		$-$		$-$		$-$	0	$+$	
$(x^2-1)^2$		$+$		$+$	0	$+$		$+$	0	$+$		$+$	
f'		$+$	0	$-$	$\|$	$-$	0	$-$	$\|$	$-$	0	$+$	
f		$\nearrow$		$\searrow$	$\|$		$\searrow$		$\|$	$\searrow$		$\nearrow$	

Une étude de limite permet alors de finaliser le tableau :

x	$-\infty$		$-\sqrt{3}$		-1		0		1		$\sqrt{3}$		$+\infty$
x^2		$+$		$+$		$+$	0	$+$		$+$		$+$	
x^2-3		$+$	0	$-$		$-$		$-$		$-$	0	$+$	
$(x^2-1)^2$		$+$		$+$	0	$+$		$+$	0	$+$		$+$	
f'		$+$	0	$-$	$\|$	$-$	0	$-$	$\|$	$-$	0	$+$	
f	$-\infty$	$\nearrow$	$\frac{-3\sqrt{3}}{2}$	$\searrow$	$-\infty$ $\|$ $+\infty$	$\searrow$	0	$\searrow$	$-\infty$ $\|$ $+\infty$	$\searrow$	$\frac{3\sqrt{3}}{2}$	$\nearrow$	$+\infty$

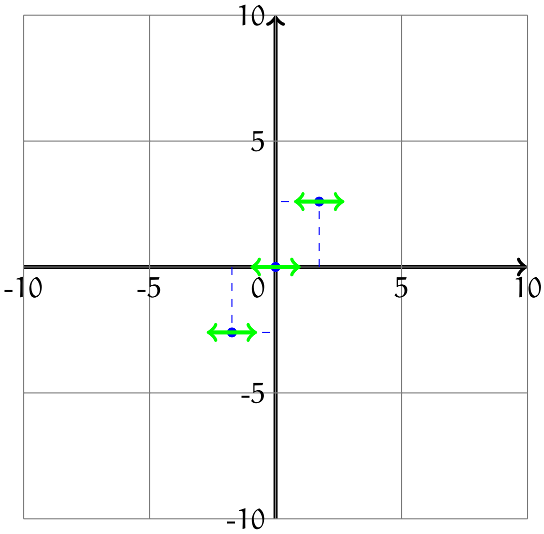
Le tableau de variation suffit en générale pour donner l'allure de la courbe représentative de la fonction f.

D'une part les valeurs finie de la fonction donne des point de passage. Dans notre exemple nous savons que la courbe passera par les points $\left(-\sqrt{3}, -\frac{3\sqrt{3}}{2}\right)$, $(0,0)$ et $\left(\sqrt{3}, \frac{3\sqrt{3}}{2}\right)$.

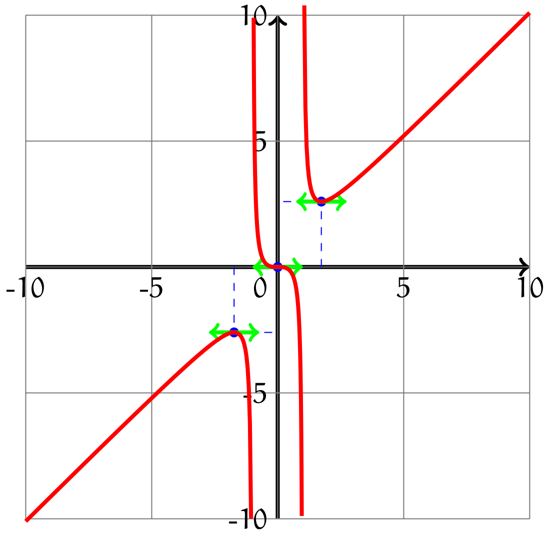


De plus, par construction, le coefficient directeur d'une tangente est le nombre dérivé. Lorsque la dérivée s'annule, on en déduit que la tangente est horizontale. Cela donnera une information sur le comportement autour de la tangente puisque par définition de *tangente*, cette droite "frôle" la courbe. Ainsi la courbe sera horizontale autour des tangente horizontale !

On trace donc des morceaux de droite horizontale lorsque la dérivée s'annule :



Pour obtenir finalement



4. Fonctions à deux variables

Dans le cadre de ce cours nous allons nous intéresser aux fonctions à deux variables à valeur réelle.

Définition

Soit $E \subset \mathbb{R}^2$, une fonction f définie sur E à valeur dans $\mathbb{R}$ est la donnée pour chaque $(x, y) \in E$ d'un nombre réel noté $f(x, y)$

Par exemple :

$$\begin{array}{rcl} f : \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x, y) & \longmapsto & x^2 + y^2 \end{array}$$

Autre exemple :

$$\begin{array}{rcl} f : E & \longrightarrow & \mathbb{R} \\ (x, y) & \longmapsto & \ln\left(\dfrac{x}{y}\right) \end{array}$$

est définie sur $E = \left\{(x, y) \in \mathbb{R}^2 \middle| (x > 0 \wedge y > 0) \vee (x < 0 \wedge y < 0)\right\}$

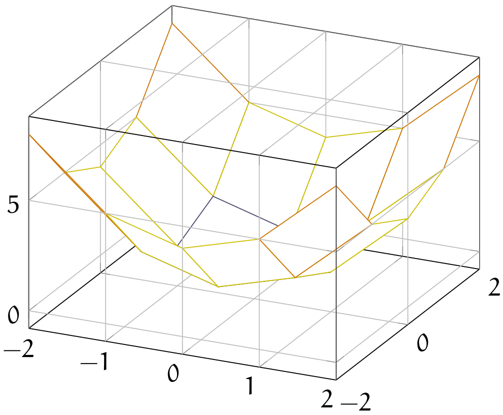
4.1 Représentation

Avec les fonctions à une variable nous pouvions tracer le graphe $(x, f(x))$ pour les valeurs de x dans le domaine de définition de f.

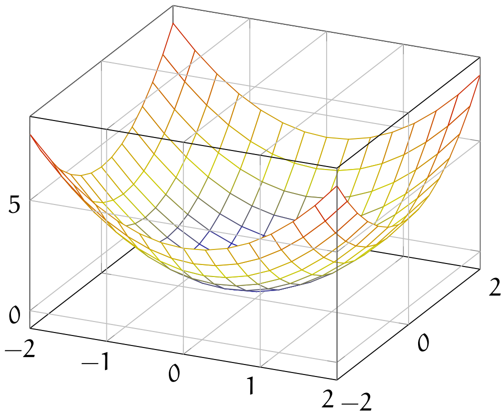
Avec les fonctions à deux variables cela est plus compliqué. Il faut réaliser le graphique en trois dimensions du graphe $(x, y, f(x, y))$. Sans un ordinateur cela semble difficile.

On prend différentes valeurs de x et y et on relie les points. Plus on prend de valeur plus la représentation se raffine. Il est parfois utile d'ajouter un dégradé de couleur.

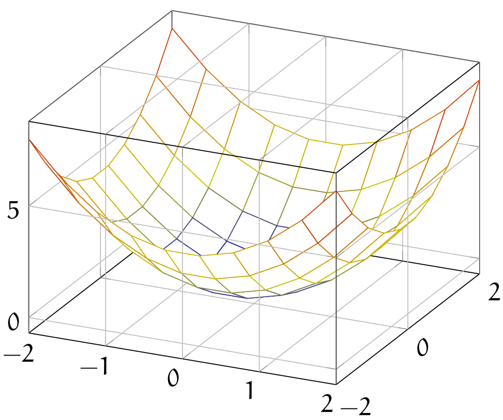
$f(x, y) = x^2 + y^2$ avec 5×5 points



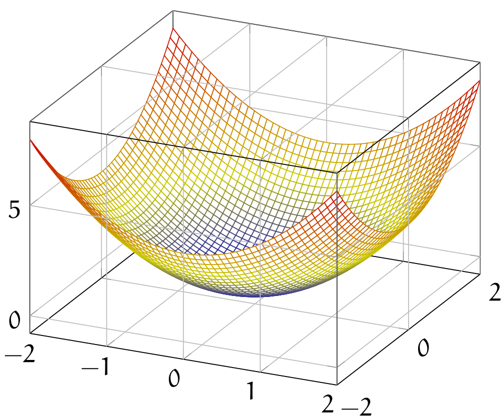
$f(x, y) = x^2 + y^2$ avec 15×15 points



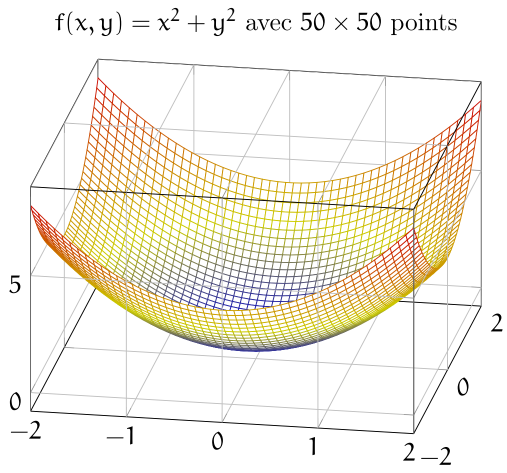
$f(x, y) = x^2 + y^2$ avec 10×10 points

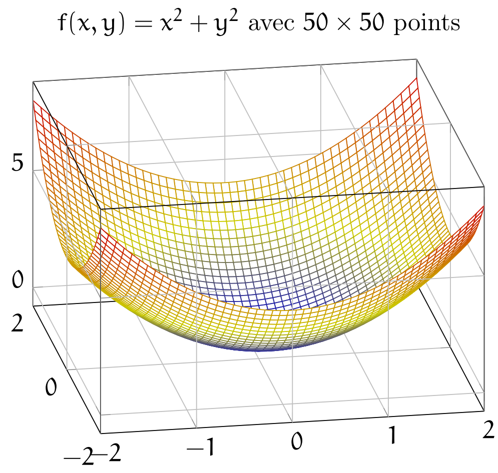


$f(x, y) = x^2 + y^2$ avec 50×50 points



La représentation sous différent angle de vue permet aussi de mieux appréhender la fonction.





4.2 Lignes de niveau et fonctions partielles

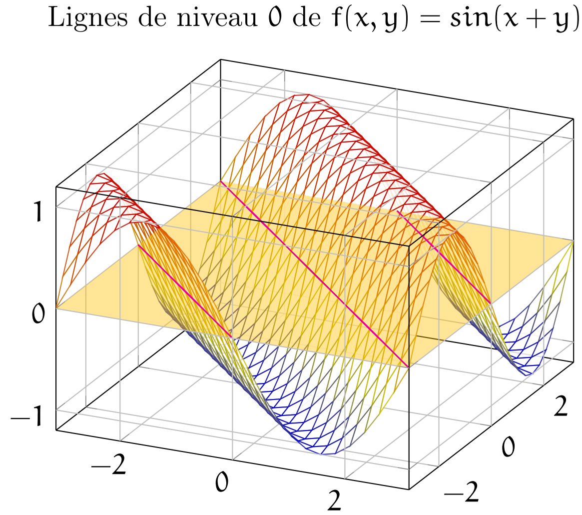
Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $h \in R$. La ligne de niveau h de f est l'ensemble

$$L_h = \left\{ (x,y) \in E \middle| f(x,y) = h \right\}$$

Géométriquement une ligne de niveau correspond à l'intersection de la courbe de f et du plan $z = h$.

Prenons par exemple la fonction définie sur $\mathbb{R}^2$: $f(x,y) = \sin(x+y)$ alors la ligne de niveau 0 est l'ensemble des droite $y = -x + k\pi$ pour $k \in \mathbb{Z}$.



Définition

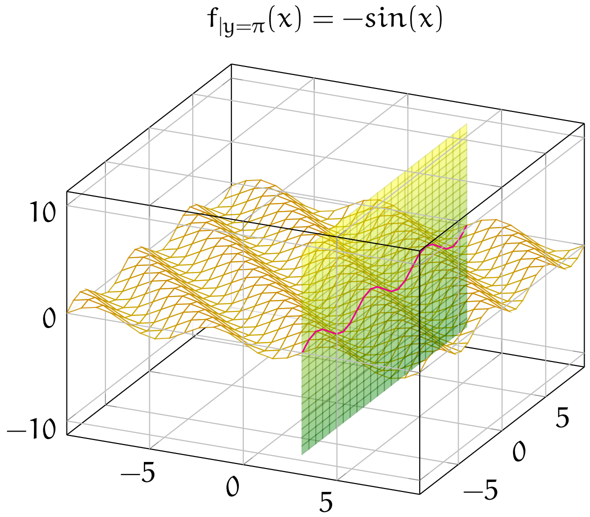
Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a,b) \in E$. Les fonctions partielles de f en (a,b) sont définies par

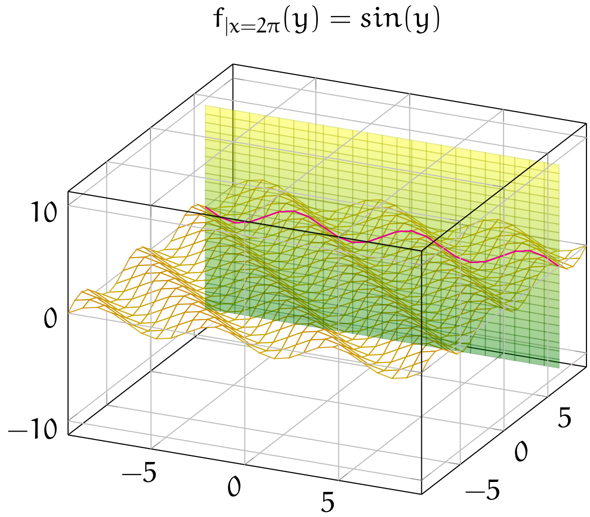
$$f_{|y=b}(x) = f(x,b) \qquad f_{|x=a}(y) = f(a,y)$$

Géométriquement les fonctions partielle correspondent à l'intersection du graphe de f avec les plans $x = a$ ou $y = b$.

Reprenons l'exemple de la fonction $f(x,y) = \sin(x+y)$. Alors

$$f_{|y=\pi}(x) = -\sin(x) \qquad f_{|x=2\pi}(y) = \sin(y)$$





4.3 Dérivées partielles et gradient

L'idée est d'imiter ce que nous savons faire avec les fonctions à une variable et d'obtenir un équivalent de l'outil qu'est la dérivée.

Sauf qu'il y a un problème avec la définition de dérivé. Avec une variable le nombre dérivé d'une fonction f en un point a est défini par la limite $\lim\limits_{x \to a} \dfrac{f(x) - f(a)}{x - a}$. Avec une variable, la notion de limite est assez claire, soit on se rapproche de a par la gauche, ce qui est noté $\lim\limits_{x \to a^-}$, soit par la droite, noté $\lim\limits_{x \to a^+}$. En effet, sur *une ligne*, on peut moralement se rapprocher d'un nombre soit par la gauche soit par la droite.

Mais lorsque nous sommes avec deux variables, quel sens donner à la notion de limite ? Comment se rapprocher d'un point. Il y a plein de manière différente de se rapproche de (a, b) : en spirale, en ligne droite, de manière exponentielle ou logarithmique, comme une parabole etc... Est-ce qu'il existe une *meilleure* manière de se rapprocher d'un point ou est-ce qu'il existe un *moyen* universelle ?

La réponse à cette question est OUI. Mais dans le cadre de ce cours nous n'allons pas introduire cette notion qui nécessite de définir une distance et de parler de développement limité. Une autre manière d'introduire la dérivée avec plusieurs variables est de se servir des fonctions partielles et de les dériver.

Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$.

- La **dérivé partielle** de f par rapport à x en (a, b) est la dérivé de la fonction partielle $f_{|y=b}(x)$ en $x = a$. On la note $\dfrac{\partial f}{\partial x}(a, b)$. Par définition c'est

$$\frac{\partial f}{\partial x}(a, b) = \lim_{x \to a} \frac{f(x, b) - f(a, b)}{x - a}$$

- La **dérivé partielle** de f par rapport à y en (a, b) est la dérivé de la fonction partielle $f_{|x=a}(y)$ en $y = b$. On la note $\dfrac{\partial f}{\partial y}(a, b)$. Par définition c'est

$$\frac{\partial f}{\partial y}(a, b) = \lim_{y \to b} \frac{f(a, y) - f(a, b)}{y - b}$$

On définit, comme pour les fonction à une variables, les fonctions dérivées partielles en (a, b).

En d'autres termes, la dérivée partielle $\dfrac{\partial f}{\partial x}$ revient à "*oublier*" que y est une variable dans l'expression de $f(x, y)$ et de ne dériver qu'en considérant x comme variable et y comme constante. En particulier, toutes les opérations classiques sur les dérivées s'appliquent.

Prenons par exemple la fonction $f(x, y) = \dfrac{x}{x^2 + y^2}$ définie sur $\mathbb{R}^2 - \{(0, 0)\}$. Alors

En utilisant la dérivé de $\frac{u}{v}$ qui est $\frac{u'v - v'u}{v^2}$:

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

En utilisant la dérivé de $\frac{1}{u}$ qui est $-\frac{u'}{u^2}$, le x au numérateur étant considéré comme une constante :

$$\begin{aligned}\frac{\partial f}{\partial y} &= -\frac{x\times 2y}{(x^2+y^2)}\\ &= -\frac{2xy}{(x^2+y^2)}\end{aligned}$$

Avec une seule variable le nombre dérivée représentait la vitesse, précisément le vecteur vitesse (c'est la tangente).

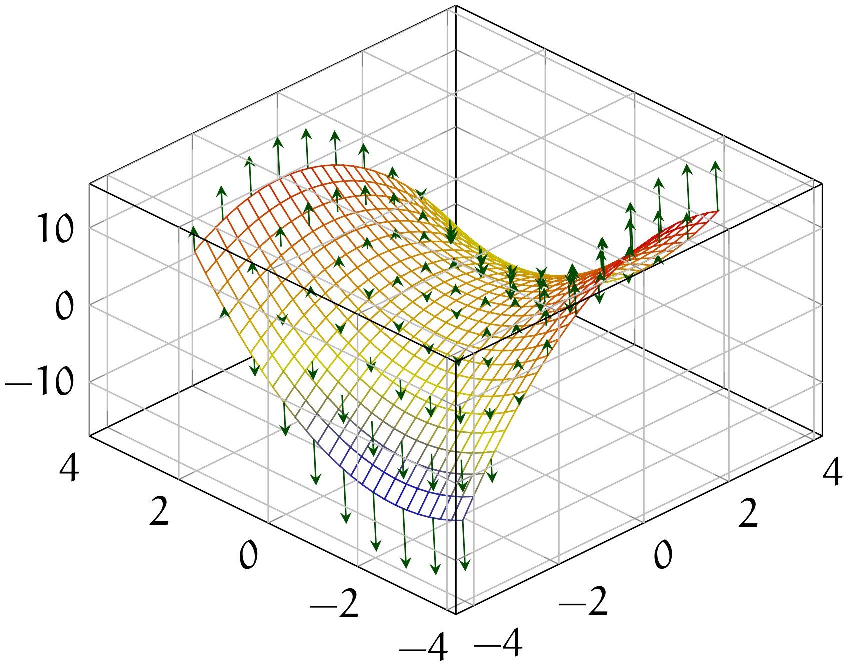
Avec deux variables, c'est la même chose. On parle du gradient.

Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$. Le **gradient** de f en (a, b) est le vecteur

$$\overrightarrow{\mathrm{Grad}}_{(a,b)}(f) = \begin{pmatrix}\frac{\partial f}{\partial x}(a,b)\\ \frac{\partial f}{\partial y}(a,b)\end{pmatrix}$$

Considérons la fonction $f(x, y) = -x^2 - xy + x + y^2$. Le gradient est $\overrightarrow{\mathrm{Grad}}_{(a,b)}(f) = \begin{pmatrix}-2a-b+1\\ -a+2b\end{pmatrix}$ représenté en vert ci dessous.



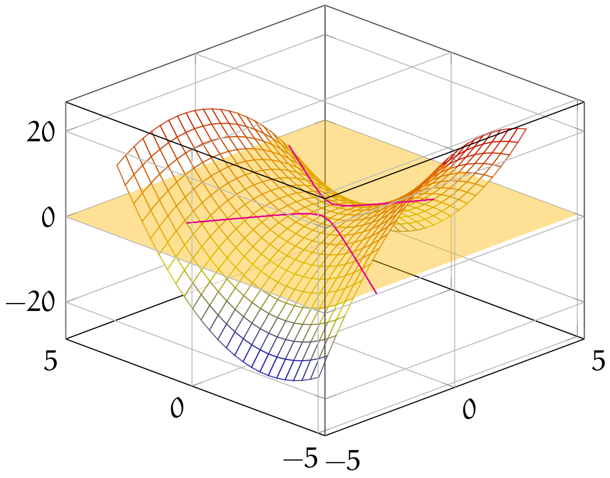
Proposition

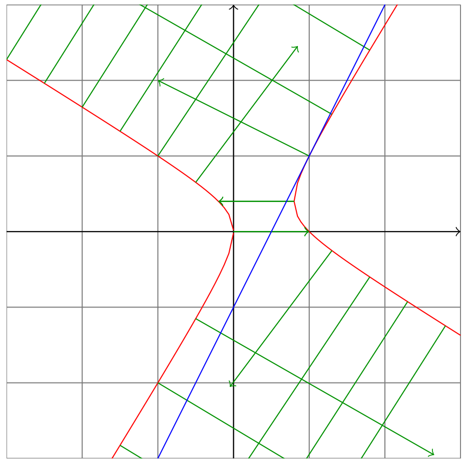
Le gradient d'une fonction en (a, b) est orthogonale à la ligne de niveau $f(a, b)$.

Démonstration. Admise □

Reprenons la fonction précédente. Considérons le point $(a, b) = (1, 1)$. Le gradient est $\overrightarrow{\mathrm{Grad}}_{(1,1)}(f) = \begin{pmatrix}-2\\ 1\end{pmatrix}$

En dessin cela donne (en rouge la ligne de niveau $f(1, 1) = 0$) en verts les gradients.





Corollaire

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$ tel que $\overrightarrow{\mathrm{Grad}}_{(a,b)}(f)$ existe et n'est pas le vecteur nul. Alors la droite tangente à la ligne de niveau $f(a, b)$ a pour équation :

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) = 0$$

Démonstration. Soit $A = (a, b)$. Notons $M = (x, y)$ un point de cette droite. Comme $\overrightarrow{AM}$ est orthogonal à $\overrightarrow{\mathrm{Grad}}_{(a,b)}(f)$ (d'après la proposition précédente) on a $\overrightarrow{\mathrm{Grad}}_{(a,b)}(f) \cdot \overrightarrow{AM} = 0$ ce qui traduit l'équation du corollaire. □

Par exemple l'équation de la tangente en $(1, 1)$ de la ligne de niveau 0 de la fonction $f(x, y) = -x^2 - xy + x + y^2$ est $-2(x - 1) + 1(y - 1) = 0$ soit la droite $y = 2x - 1$ (en bleue sur le graphique précédent).

Que se passe-t-il si le gradient est nul ?

4.4 Extrema globaux et locaux

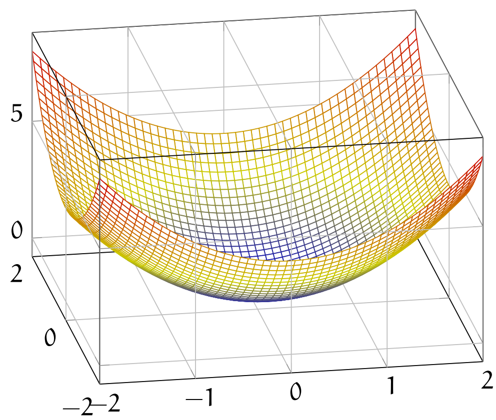
Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$.

- On dira que (a, b) est un maximum (resp. minimum) global si pour tout $(x, y) \in E$, $f(x, y) \leqslant f(a, b)$ (resp. $f(x, y) \geqslant f(a, b)$).
- On dira que (a, b) est un maximum (resp. minimum) global si pour tout $(x, y) \in D$, $f(x, y) \leqslant f(a, b)$ (resp. $f(x, y) \geqslant f(a, b)$).

  Où $D \subset E$ est un (petit) disque autour de (a, b).

Par exemple, il est facile d'observer que $(0, 0)$ est un minimum global de $f(x, y) = x^2 + y^2$.



Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$.
On dira que (a, b) est un point critique si $\overrightarrow{\mathrm{Grad}}_{(a,b)}(f) = \overrightarrow{0}$.

Reprenons l'exemple de la fonction $f(x, y) = -x^2 - xy + x + y^2$ et dont le gradient est $\overrightarrow{\mathrm{Grad}}_{(a,b)}(f) = \begin{pmatrix} -2a - b + 1 \\ -a + 2b \end{pmatrix}$. Trouver les points critiques revient donc à résoudre le système

$$\begin{cases} -2a - b + 1 = 0 \\ -a + 2b = 0 \end{cases}$$

Il est facile de voir que ce système admet une unique solution $(a, b) = \left(\frac{1}{5}, \frac{2}{5}\right)$. En conclusion, l'unique point critique de cette fonction est $\left(\frac{1}{5}, \frac{2}{5}\right)$.

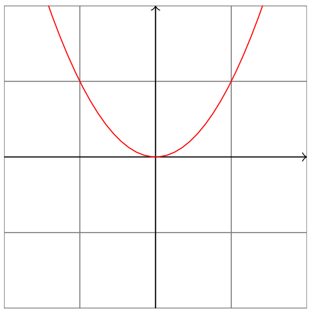
Théorème Fermat

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$.
Si (a, b) est un extrema local alors c'est un point critique.

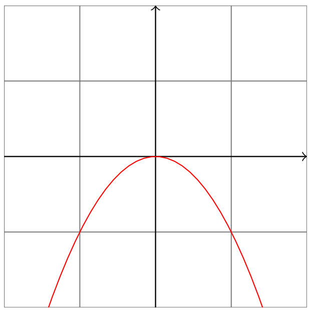
Démonstration. Admise □

En conclusion, pour déterminer un extremum, il suffit de déterminer les points critiques d'une fonction. Mais comment savoir si c'est un maximum ou un minimum. Réalisons quelques exemple à une variables.

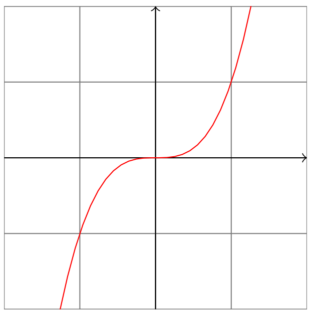
Considérons la fonction $f(x) = x^2$. Sa dérivé est $f'(x) = 2x$ qui s'annule trivialement en $x = 0$. Ainsi 0 est un point critique. Nous savons, en regardant le graphique que c'est un minimum. Regardons la dérivé seconde : $f''(x) = 2$... c'est positif en $x = 0$.



Considérons la fonction $f(x) = -x^2$. Sa dérivé est $f'(x) = -2x$ qui s'annule trivialement en $x = 0$. Ainsi 0 est un point critique. Nous savons, en regardant le graphique que c'est un maximum. Regardons la dérivé seconde : $f''(x) = -2$... c'est négatif en $x = 0$.



Considérons la fonction $f(x) = x^3$. Sa dérivé est $f'(x) = 3x^2$ qui s'annule trivialement en $x = 0$. Ainsi 0 est un point critique. Regardons la dérivé seconde : $f''(x) = 6x$... nul en $x = 0$.



Vous l'aurez compris, c'est en regardant du coté de la dérivé seconde que nous aurons une réponse.

4.5 La hessienne

Il nous faut définir la *dérivée seconde*. Nous avons les dérivés premiers $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$. Ce sont des fonctions à deux variables que nous pouvons encore dériver par rapport à la première ou la seconde variable.

Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables. Si elles existent, on note

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

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

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

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

Prenons par exemple $f(x, y) = -x^2 - xy + x + y^2$ alors

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x}(-2x - y + 1) = -2$$

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x}(-x + 2y) = -1$$

$$\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y}(-2x - y + 1) = -1$$

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

On observe que $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$. C'est toujours vrai.

Théorème Lemme de Schwartz

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

Démonstration.

$$
\begin{aligned}
\frac{\partial^2 f}{\partial x \partial y}(a,b) &= \lim_{x\to a} \frac{\frac{\partial f}{\partial y}(x,b) - \frac{\partial f}{\partial y}(a,b)}{x-a} \\
&= \lim_{x\to a} \frac{\lim\limits_{y\to b} \frac{f(x,y)-f(x,b)}{y-b} - \lim\limits_{y\to b} \frac{f(a,y)-f(a,b)}{y-b}}{x-a} \\
&= \lim_{x\to a} \frac{\lim\limits_{y\to b} \left[\frac{f(x,y)-f(x,b)}{y-b} - \frac{f(a,y)-f(a,b)}{y-b}\right]}{x-a} \\
&= \lim_{x\to a} \frac{\lim\limits_{y\to b} \left[\frac{(f(x,y)-f(x,b))-(f(a,y)-f(a,b))}{y-b}\right]}{x-a} \\
&= \lim_{y\to b}\lim_{x\to a} \frac{\left[\frac{(f(x,y)-f(x,b))-(f(a,y)-f(a,b))}{y-b}\right]}{x-a} \\
&= \lim_{y\to b}\lim_{x\to a} \frac{\left[\frac{(f(x,y)-f(x,b))-(f(a,y)-f(a,b))}{x-a}\right]}{y-b} \\
&= \lim_{y\to b}\lim_{x\to a} \frac{\left[\frac{(f(x,y)-f(a,y))-(f(x,b)-f(a,b))}{x-a}\right]}{y-b} \\
&= \lim_{y\to b}\lim_{x\to a} \frac{\left[\frac{f(x,y)-f(a,y)}{x-a} - \frac{f(x,b)-f(a,b)}{x-a}\right]}{y-b} \\
&= \lim_{y\to b} \frac{\lim\limits_{x\to a} \left[\frac{f(x,y)-f(a,y)}{x-a} - \frac{f(x,b)-f(a,b)}{x-a}\right]}{y-b} \\
&= \lim_{y\to b} \frac{\lim\limits_{x\to a} \frac{f(x,y)-f(a,y)}{x-a} - \lim\limits_{x\to a} \frac{f(x,b)-f(a,b)}{x-a}}{y-b} \\
&= \lim_{y\to b} \frac{\frac{\partial f}{\partial x}(a,y) - \frac{\partial f}{\partial x}(a,b)}{y-b} \\
&= \frac{\partial f^2}{\partial y \partial x}(a,b)
\end{aligned}
$$

□

Définition

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a,b) \in E$.

La **matrice hessienne** de f est la matrice

$$
H_{(a,b)}(f) = \begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a,b) & \frac{\partial^2 f}{\partial x \partial y}(a,b) \\ \frac{\partial^2 f}{\partial y \partial x}(a,b) & \frac{\partial^2 f}{\partial y^2}(a,b) \end{pmatrix}
$$

Par exemple la hessienne de $f(x,y) = -x^2 - xy + x + y^2$ (en tout point) est $\begin{pmatrix} -2 & -1 \\ -1 & 2 \end{pmatrix}$

Théorème

Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables, (a, b) un point critique et H sa matrice hessienne.

Si $\texttt{det}(H) > 0$ et

si $\texttt{tr}(H) > 0$ alors (a, b) est un minimum local.

si $\texttt{tr}(H) < 0$ alors (a, b) est un maximum local.

Si $\texttt{det}(H) < 0$ alors (a, b) est un *point selle*.

Si $\texttt{det}(H) = 0$ alors on ne peut pas conclure. On dit que (a, b) est un *point critique dégénéré*.

Démonstration. Admise. □

Rappelons que pour une matrice carré en dimension 2, le déterminant $\texttt{det}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$ et la trace $\texttt{tr}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + d$

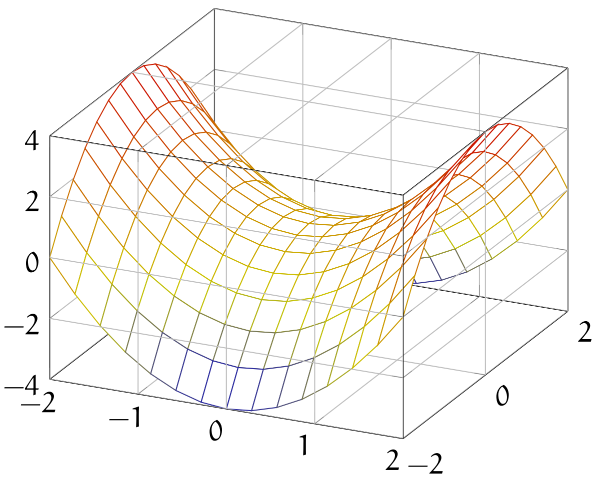
§ On peut démontrer que si $\texttt{det}(H) > 0$ alors la trace est nécessairement non nulle de sorte que nous n'avons pas oublié de traiter un cas dans le théorème précédent.

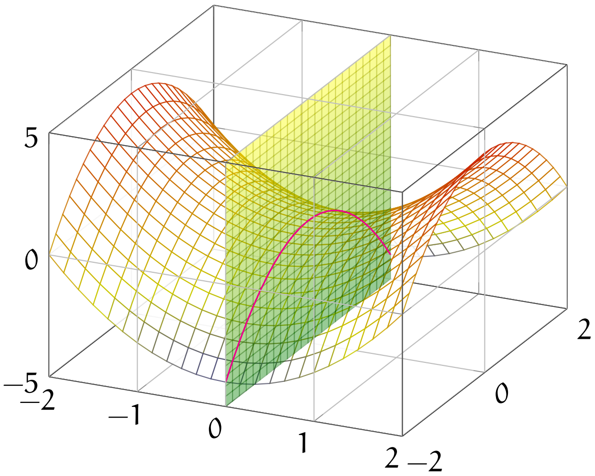
Par exemple nous avions trouver que le point critique de la fonction $f(x, y) = -x^2 - xy + x + y^2$ est $\left(\frac{1}{5}, \frac{2}{5}\right)$.

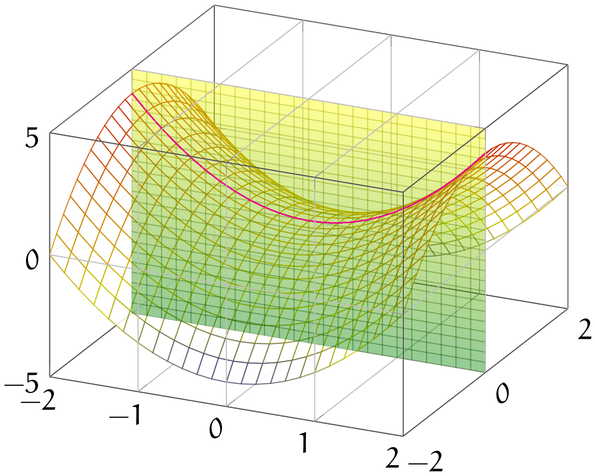
La hessienne est $\begin{pmatrix} -2 & -1 \\ -1 & 2 \end{pmatrix}$. Son déterminant vaut -5. Le point $\left(\frac{1}{5}, \frac{2}{5}\right)$ est un point selle.

§ Si la notion de minimum et de maximum est assez claire, celle de *point selle* est un peu plus exotique. C'est un point qui n'est ni un maximum ni un minimum mais aussi les deux en même temps.

Regardons la fonction définie sur $\mathbb{R}^2$, $f(x, y) = x^2 - y^2$. On vérifie que $\overrightarrow{\texttt{Grad}}f = \begin{pmatrix} 2x \\ -2y \end{pmatrix}$ et on observe trivialement que le seul point critique est $(0, 0)$. On vérifie également que la hessienne $\begin{pmatrix} 2 & 0 \\ 0 & -2 \end{pmatrix}$. D'après le théorème précédent le point critique est un point selle : selon une direction la fonction est croissante et selon une autre elle est décroissante... comme une selle de cheval, d'où le nom.







4.6 Et plus de deux variables ?

Si on dispose d'une fonction de plus de deux variables, les concepts sont les mêmes !

Cependant l'étude du déterminant et de la trace de la hessienne ne suffit pas pour déterminer la nature du point critique.

Le théorème spectrale devrait être un outil puissant, mais hors de propos dans ce cours.

Pour étudier la fonction définie sur $\mathbb{R}^3$ par $f(x, y, z) = x^2 + yz + z$, on commence par chercher les points critique par le calcul du gradient.

On trouve aisément que $\overrightarrow{\mathtt{Grad}}(f) = \begin{pmatrix} 2x \\ z \\ y+1 \end{pmatrix}$ qui s'annule en $(x, y, z) = (0, -1, 0)$. On détermine sa hessienne tout aussi aisément : $H = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$. On a $\mathtt{det}(H) = -2 < 0$ et $\mathtt{tr}(H) = 2 > 0$ et le point $(0, -1, 0)$ est une sorte de point selle : il existe deux directions différentes selon lesquelles la fonction est croissante pour l'une et décroissante pour l'autre.

5. Méthode des moindre carrés

5.1 La statistique bivarié

Dans une promotion de 52 étudiants, on dispose, pour chaque étudiant, de sa moyenne en mathématique et de sa moyenne en informatique.

On représente ces données dans un graphique, où les notes de maths sont placées en abscisse et les notes d'informatique en ordonnée. Chaque point représente donc un étudiant.

Une observation rapide, permet d'observer qu'il y a un alignement.

L'observation de cet alignement nous motive à penser que la note d'informatique d'un étudiant, notée y, est une fonction affine de la note de mathématique, notée x. La formulation mathématique est donc

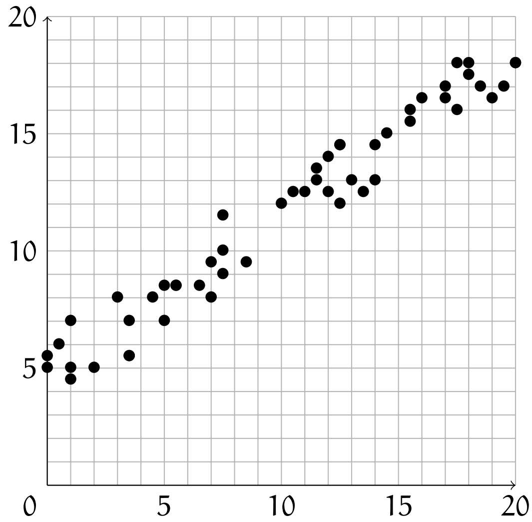
$$y = ax + b$$

On observe aussi que, bien que les données soient alignées elles ne le sont pas parfaitement, en vérité $y = ax + b + \mathrm{erreur}$.

L'objectif de la *régression linéaire* est d'essayer de trouver a et b tel que l'*erreur* soit la plus petite possible.

Il y a enfin un dernier paramètre à prendre en compte : il n'y a pas qu'une information, mais 52, c'est à dire qu'en fait $y_i = ax_i + b + \mathrm{erreur}_i$ où l'ajout de l'indice i permet d'identifier l'étudiant i.

Finalement, nous cherchons a et b tel que les erreur_i soient les plus petit possible.



L'idée[1] est de faire en sorte que $\sum_i \mathrm{erreur}_i^2$ soient les plus petit possible. De la formule $y_i = ax_i + b + \mathrm{erreur}_i$ on obtient $\mathrm{erreur}_i = y_i - (ax_i + b)$ et on cherche donc le minimum de

$$f(a,b) = \sum_i (y_i - (ax_i + b))^2$$

Déterminons les points critiques éventuels de cette fonction.

$$\frac{\partial f}{\partial a}(a,b) = \sum_i -2x_i\,(y_i - (ax_i + b))$$

$$\frac{\partial f}{\partial b}(a,b) = \sum_i -2\,(y_i - (ax_i + b))$$

Trouver un point critique reviens donc à résoudre le système

$$\left\{\begin{array}{rcl} \frac{\partial f}{\partial a}(a,b) & = & 0 \\ \frac{\partial f}{\partial b}(a,b) & = & 0 \end{array}\right. \Longrightarrow \left\{\begin{array}{rcl} \sum_i - 2x_i\,(y_i - (ax_i + b)) & = & 0 \\ \sum_i - 2\,(y_i - (ax_i + b)) & = & 0 \end{array}\right.$$

Faisons un peut de math, rappelons quelques notation de la statistique : $\overline{x} = \frac{1}{52}\sum_i x_i$ (la moyenne), $\overline{xy} = \frac{1}{52}\sum_i x_i y_i$, $\overline{x^2} = \overline{xx} = \frac{1}{52}\sum_i x_i^2$, $\mathrm{cov}(x,y) = \overline{xy} - \overline{x}\cdot\overline{y} = \frac{1}{52}\sum_i (x_i y_i - \overline{x}\cdot\overline{y})$ (la covariance), $\sigma_x^2 = \mathrm{cov}(x,x)$ (la variance, la racine carré, σ_x est appelé l'*écrat-type*).

1. de Legendre et Gauss

$$\left\{\begin{array}{rcl}\displaystyle\sum_i -2x_i\left(y_i-(ax_i+b)\right) &=& 0\\ \displaystyle\sum_i -2\left(y_i-(ax_i+b)\right) &=& 0\end{array}\right. \Rightarrow \left\{\begin{array}{rcl}\displaystyle\sum_i x_i\left(y_i-(ax_i+b)\right) &=& 0\\ \displaystyle\sum_i \left(y_i-(ax_i+b)\right) &=& 0\end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl}\displaystyle\sum_i x_iy_i-ax_i^2-bx_i &=& 0\\ \displaystyle\sum_i y_i-ax_i-b &=& 0\end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl}\displaystyle\frac{1}{52}\sum_i x_iy_i-ax_i^2-bx_i &=& 0\\ \displaystyle\frac{1}{52}\sum_i y_i-ax_i-b &=& 0\end{array}\right.$$

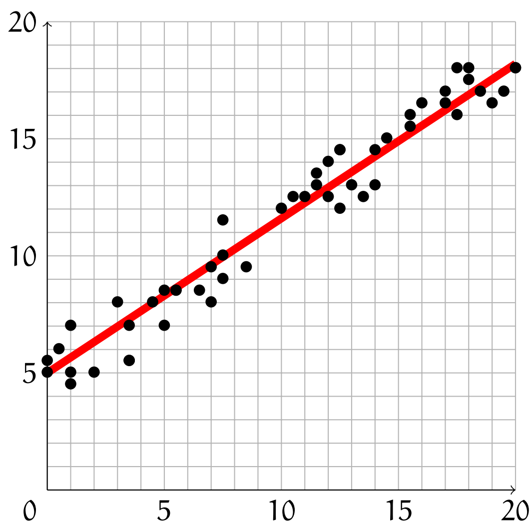
$$\Rightarrow \left\{\begin{array}{rcl}\displaystyle\frac{1}{52}\sum_i x_iy_i-a\frac{1}{52}\sum_i x_i^2-b\frac{1}{52}\sum_i x_i &=& 0\\ \displaystyle\frac{1}{52}\sum_i y_i-a\frac{1}{52}\sum_i x_i-b\frac{1}{52}\sum_i 1 &=& 0\end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl}\overline{xy}-a\overline{x^2}-b\overline{x} &=& 0\\ \overline{y}-a\overline{x}-b &=& 0\end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl}a\overline{x^2}+b\overline{x} &=& \overline{xy}\\ a\overline{x}+b &=& \overline{y}\end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl}a\overline{x^2}+b\overline{x} &=& \overline{xy}\\ a\overline{x}^2+b\overline{x} &=& \overline{y}\cdot\overline{x}\end{array}\right.$$

En faisant la soustraction des deux lignes, on trouve $a(\overline{x^2}-\overline{x}^2)=\overline{xy}-\overline{y}\cdot\overline{x}$, soit avec les outils et notation de la statistique $a\sigma_x^2=\mathrm{cov}(x,y)$ et donc $a=\dfrac{\mathrm{cov}(x,y)}{\sigma_x^2}$, la seconde équation donne $b=\overline{y}-a\overline{x}$. Bref, nous avons trouver un unique point critique. Puisqu'il s'agit d'une somme de carré, c'est nécessairement un minimum. Ce a et ce b donne donc la *meilleure* droite.



On peut se convaincre qu'il s'agit d'un minimum en calculant la hessienne.

$$\begin{aligned}\frac{\partial^2 f}{\partial a^2}(a,b) &= \frac{\partial}{\partial a}\left(\sum_i -2x_i\left(y_i-(ax_i+b)\right)\right)\\ &= \frac{\partial}{\partial a}\left(\sum_i -2x_iy_i+2ax_i^2+2bx_i\right)\\ &= \sum_i 2x_i^2\\ &= 2\times 52\overline{x^2}\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial b^2}(a,b) &= \frac{\partial f}{\partial b}\left(\sum_i -2\left(y_i - (ax_i + b)\right)\right)\\
&= \frac{\partial f}{\partial b}\left(\sum_i -2y_i + 2ax_i + 2b\right)\\
&= \sum_i 2\\
&= 2 \times 52
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial a \partial b}(a,b) &= \frac{\partial f}{\partial a}\left(\sum_i -2\left(y_i - (ax_i + b)\right)\right)\\
&= \frac{\partial f}{\partial a}\left(\sum_i -2y_i + 2ax_i + 2b\right)\\
&= \sum_i 2x_i\\
&= 2 \times 52\overline{x}
\end{aligned}$$

Ainsi la hessienne est $H = \begin{pmatrix} 104\overline{x^2} & 104\overline{x} \\ 104\overline{x} & 104 \end{pmatrix}$. En particulier $\mathtt{det}(H) = (104\overline{x^2})(104) - (104\overline{x})^2 = 104^2\left(\overline{x^2} - \overline{x}\right) = 104^2\sigma_x^2$. Ce déterminant est donc positif.

On a de plus $\mathtt{tr}(H) = 104\overline{x^2} + 104 > 0$ et nous avons bien un minimum.

5.2 Cas général

On dispose de n observations de deux caractères quantitatifs $x = (x_1, \dots, x_n)$ et $y = (y_1, \dots, y_n)$, les observations permettent d'émettre l'hypothèse que pour tout i, les données y_i suivent une certaine fonction $F(x_i, \vartheta)$ où ϑ est un ensemble de paramètre inconnue.

La méthode des moindres carrés, consiste à trouver les paramètres ϑ tel que $(y_i - F(x_i, \vartheta))^2$ soit le plus petit possible.

Dans la pratique, on cherche à minimiser $f(\vartheta) = \sum_i (y_i - F(x_i, \vartheta))^2$.

Dans l'exemple bien connu de statistique bivarié, les paramètres $\vartheta = (a, b)$ et $F(x, \vartheta) = ax_i + b$

5.3 Exemple : un modèle non linéaire

Vous disposez de 100 valeurs x_i et y_i représentées dans le dessin ci dessous.

Un statisticien vous souffle que ce nuage de point semble suivre une courbe $\dfrac{a}{x_i} + \dfrac{b}{x_i^2}$.

Déterminer les *meilleures* estimations de a et b.

Appliquons la méthodes des moindre carrés, et déterminons les minimum de

$$f(a,b) = \sum_i \left(y_i - \frac{a}{x_i} - \frac{b}{x_i^2}\right)^2$$

Pour simplifier les notations, posons $z_i = \dfrac{1}{x_i}$, on a donc, $f(a,b) = \sum_i \left(y_i - az_i - bz_i^2\right)^2$.

Déterminons ses points critiques :

$$\frac{\partial f}{\partial a} = \sum_i -2z_i\left(y_i - az_i - bz_i^2\right)$$

$$\frac{\partial f}{\partial b} = \sum_i -2z_i^2\left(y_i - az_i - bz_i^2\right)$$

$$\left\{\begin{array}{rcl} \dfrac{\partial f}{\partial a} & = & 0 \\ \dfrac{\partial f}{\partial b} & = & 0 \end{array}\right. \Rightarrow \left\{\begin{array}{rcl} \sum_i -2z_i\left(y_i - az_i - bz_i^2\right) & = & 0 \\ \sum_i -2z_i^2\left(y_i - az_i - bz_i^2\right) & = & 0 \end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl} \sum_i z_i\left(y_i - az_i - bz_i^2\right) & = & 0 \\ \sum_i z_i^2\left(y_i - az_i - bz_i^2\right) & = & 0 \end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl} \sum_i z_iy_i - az_i^2 - bz_i^3 & = & 0 \\ \sum_i z_i^2y_i - az_i^3 - bz_i^4 & = & 0 \end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl} \overline{zy} - a\overline{z^2} - b\overline{z^3} & = & 0 \\ \overline{z^2y} - a\overline{z^3} - b\overline{z^4} & = & 0 \end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl} a\overline{z^2} + b\overline{z^3} & = & \overline{zy} \\ a\overline{z^3} + b\overline{z^4} & = & \overline{z^2y} \end{array}\right.$$

$$\Rightarrow \left\{\begin{array}{rcl} a\overline{z^2}\cdot\overline{z^3} + b\overline{z^3}\cdot\overline{z^3} & = & \overline{zy}\cdot\overline{z^3} \\ a\overline{z^3}\cdot\overline{z^2} + b\overline{z^4}\cdot\overline{z^2} & = & \overline{z^2y}\cdot\overline{z^2} \end{array}\right.$$

La différence des deux lignes donne : $b(\overline{z^3}^2 - \overline{z^4}\cdot\overline{z^2}) = \overline{zy}\cdot\overline{z^3} - \overline{z^2y}\cdot\overline{z^2}$ et donc $b = \dfrac{\overline{zy}\cdot\overline{z^3} - \overline{z^2y}\cdot\overline{z^2}}{\overline{z^3}^2 - \overline{z^4}\cdot\overline{z^2}}$ et $a = \dfrac{\overline{zy} - b\overline{z^3}}{\overline{z^2}}$ (il faudrait prendre toutes les précaution mathématiques et s'assurer que les opérations effectuées dans le système ou les dénominateurs des fractions qui apparaissent soient bien définies ; dans la pratique de la statistique les cas d'erreurs sont très rare).