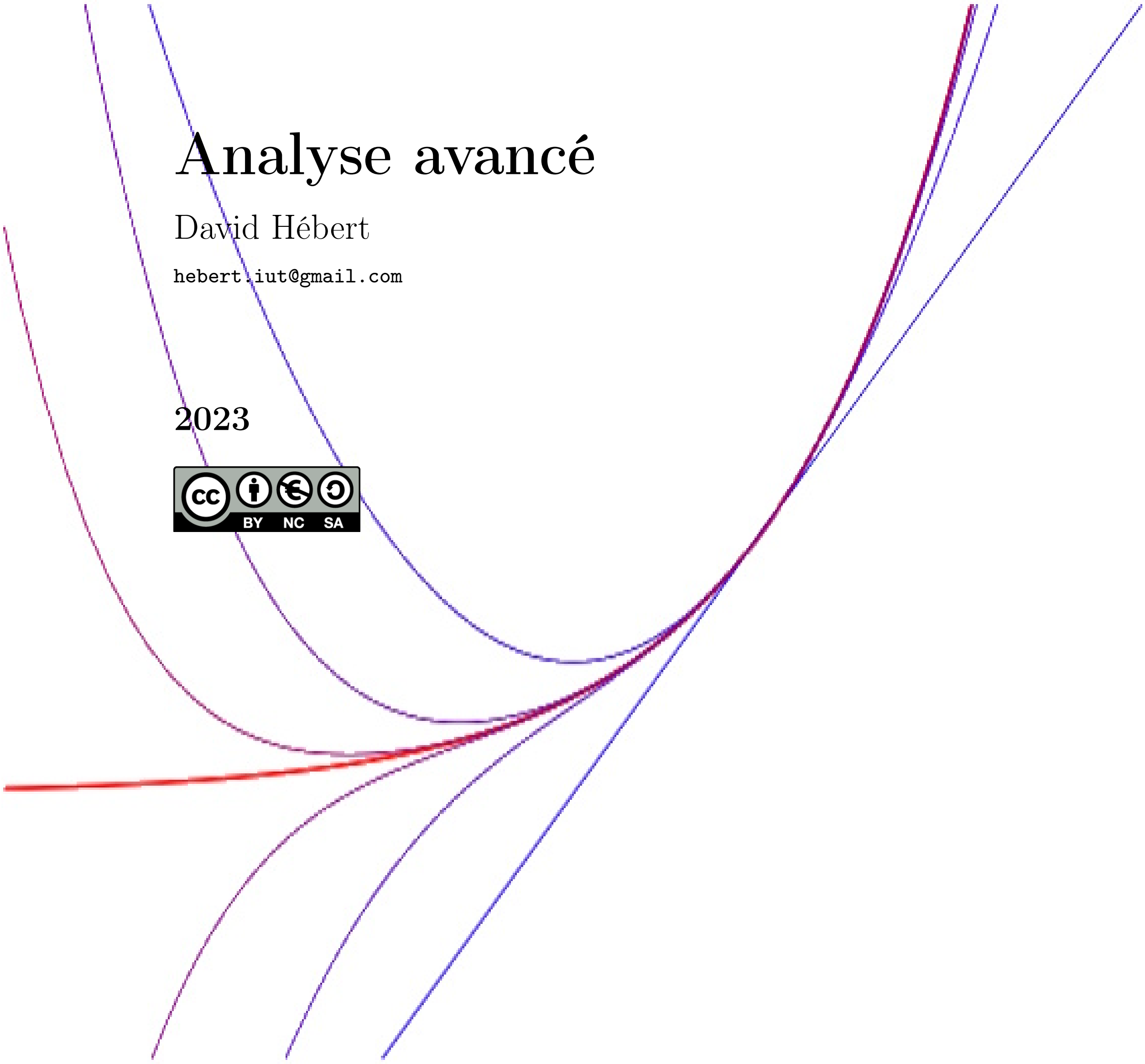


Analyse avancé

David Hébert

hebert.iut@gmail.com

2023

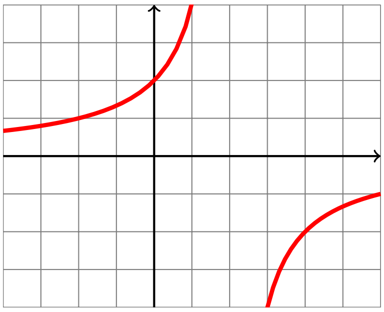
Table des matières

1. Fonctions à une variable

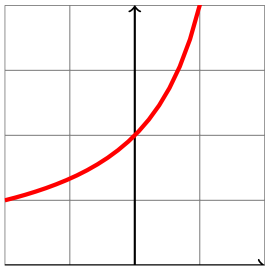
1.1 Introduction

L'objectif de ce cours est d'étudier les fonctions. Mais contrairement à faire une étude globale comme cela est exploré en terminale (dérivé, tableau de variation etc), on s'intéresse ici à faire une étude locale. C'est à dire à **comprendre simplement** ce qu'est une fonction.

Prenons par exemple la fonction $f(x) = \dfrac{1}{1-x}$. Cette élégante fonction est définie sur $\mathbb{R} - \{1\}$.



Comme nous l'avons signalé, nous cherchons à développer des outils pour comprendre cette fonction localement. Prenons par exemple en $a = 0$.



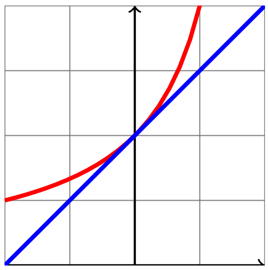
La problématique est d'essayer de trouver, localement, une forme simple à cette fonction. Une *forme simple*, pour les penseurs de l'époque, étaient les polynômes. En effet, l'univers des polynômes est bien connue et sans (presque) aucun mystère : on sait les dériver, les intégrer, les calculer, les transformer... bref! Lorsqu'un mathématicien travaille avec un polynôme, il est en terrain conquit.

Les étudiants se rappelant de leur cours de terminale devrait déjà un voir une prémisse aux solutions : la tangente!

En effet la tangente en un point de la courbe représentative d'une fonction est une droite. Plus savamment dis : c'est un polynôme de degrés 1 et la formule permettant de la calculer est bien connue **lorsque la fonction est dérivable**. L'équation de la tangente à f en un point a est

$$T_a : \qquad y = f(a) + f'(a)(x-a)$$

Dans notre exemple, la tangente en 0 donne $y = 1 + x$

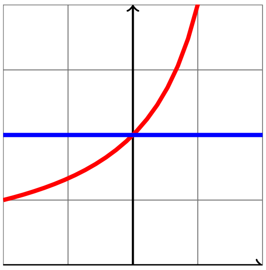


Mais la fonction $x \mapsto \dfrac{1}{1-x}$ a une autre vertu. On se rappel que la somme des termes d'une suite géométrique de raison x est

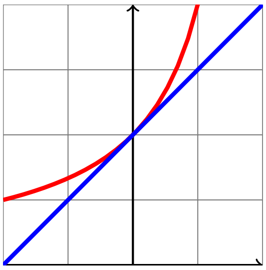
$$\sum_{k=0}^{n} x^k = 1 + x + x^2 + x^3 + \cdots + x^n = \frac{1-x^{n+1}}{1-x} = \frac{1}{1-x} - \frac{x^{n+1}}{1-x} = f(x) - \frac{x^{n+1}}{1-x}$$

Puisque nous somme *local* autour de 0, on peut dire que $\dfrac{x^{n+1}}{1-x}$ est négligeable devant $\dfrac{1}{1-x} = f(x)$. Plus précisément $\lim\limits_{x\to 0} \dfrac{x^{n+1}}{1-x} = 0$.

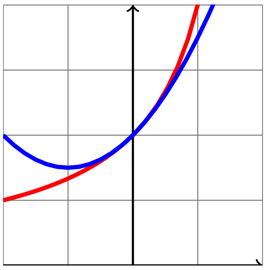
Autrement dis le polynôme $P_n(x) = \sum\limits_{k=0}^{n} x^k$ approche bien la fonction f.



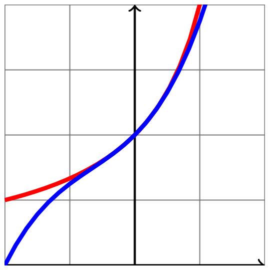
$n = 0$



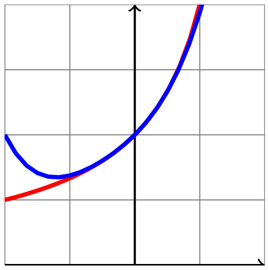
$n = 1$



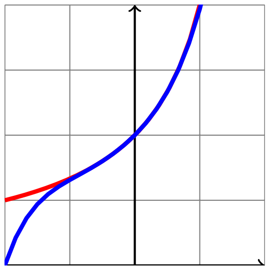
$n = 2$



$n = 3$



$n = 4$



$n = 5$

Cette dernière approximation nous donne que, **localement autour de** 0, la fonction $f(x) = \dfrac{1}{1-x}$ est proche de $P_5(x) = 1 + x + x^2 + x^3 + x^4 + x^5$. On peut d'ailleurs s'en convaincre en prenant une valeur de x, autour de 0, comme par exemple -0.01. D'un coté $f(-0.01) \simeq 0.9900990099009$ d'autre part $P_5(-0.01) = 0.9900990099$. Dix décimales correctes !

Mais y a-t-il mieux ? Comment faire pour trouver la *meilleure approximation polynômiale* ? C'est ce que nous allons explorer.

1.2 Termes d'erreurs *aka* Poubelles

Avant de parler de ce qui nous intéresse, nous allons introduire une notation à faible rigueur apparente mais qui va nous permettre d'ajouter du confort de calcul et de rédaction.

La nature de notre recherche est, rappelons-le, **locale**. On ne s'intéresse qu'a une approximation autour d'un point. Dans notre exemple introductif, nous avons approché la fonction $f(x) = \dfrac{1}{1-x}$ par le polynôme $P_5(x) = 1 + x + x^2 + x^3 + x^4 + x^5$ **autour de** 0 et pour nous en convaincre nous avons comparé ces deux expressions en -0.01. Mais si nous nous éloignons de 0, comme en -1 par exemple, d'un coté nous avons $f(-1) = 0.5$ et d'un autre cote $P_5(-1) = 0$ et l'approximation est très mauvaise.

Pour comprendre ce phénomène locale, regardons de plus près ce qui nous a amené à ce résultat. Nous avons $f(x) = P_5(x) - \dfrac{x^6}{1-x}$ ce qui permet de mesurer l'erreur d'approximation commise ; c'est $-\dfrac{x^6}{1-x}$ que nous allons préférer écrire $x^5\left(\dfrac{-x}{1-x}\right)$.

Le principe de *localité* dont nous avons besoin est de dire que *lorsque l'on se rapproche de* 0*, le terme d'erreur tend vers* 0. La caractérisation de la *vitesse* de cette convergence vers 0 est mesurée par le x^5. Dans notre exemple, on pourrait dire que le terme d'erreur tend vers 0, aussi vite que x^5.

Définition

On dira qu'une fonction f localement définie autour de $a \in \mathbb{R}$ est un terme d'erreur d'ordre $n \in \mathbb{N}$, si

$$f(x) = (x-a)^n \varepsilon(x-a)$$

pour une certaine fonction ε définie autour de 0, tel que $\lim_{x \to 0} \varepsilon(x) = 0$.

Dans la suite de ce cours, pour des raisons que nous détaillerons plus tard, nous appellerons un terme d'erreur une **poubelle**.

Proposition

1. Toutes les poubelles d'ordre n sont des poubelles d'ordre $0 \leqslant p \leqslant n$.
2. Une combinaison linéaire de poubelles d'ordre n est une poubelle d'ordre n.
3. Le produit d'une poubelle d'ordre n par une poubelle d'ordre m est une poubelle d'ordre $n+m$.

Démonstration.

1. Simplement parce que $(x-a)^n\varepsilon(x-a) = (x-a)^p\left((x-a)^{n-p}\varepsilon(x-a)\right) = (x-a)\eta(x-a)$ et on a bien $\lim_{\to 0} \eta(x) = 0$.
2. Simplement parce que $\sum_{i=0}^{N} \alpha_i (x-a)^n \varepsilon_i(x-a) = (x-a)^n \left(\sum_{i=0}^{N} \alpha_i \varepsilon_i(x-a)\right) = (x-a)^n \eta(x-a)$ et $\lim_{x \to 0} \eta(x) = 0$.
3. Car $(x-a)^n\varepsilon_1(x-a)(x-a)^m\varepsilon_2(x-a) = (x-a)^{n+m}\left(\varepsilon_1(x-a)\varepsilon_2(x-a)\right) = (x-a)^{n+m}\eta(x-a)$ où $\lim_{x \to 0} \eta(x) = 0$.

□

Ces règles vont nous permettre de noter **TOUTES** les poubelles par le symbole ε.

1.3 Développement limité

Motivé par notre introduction nous posons la définition suivante.

Définition

Soient $f : I \longrightarrow \mathbb{R}$ une fonction définie sur un intervalle $I \subset \mathbb{R}$, $a \in I$ et $n \in \mathbb{N}$. On dira que f admet un **développement limité** d'ordre n en a, noté $DL_n(a)$, s'il existe un polynôme P de degrés au plus n tel que

$$f(x) = P(x-a) + (x-a)^n \varepsilon(x-a)$$

où ε est une poubelle, c'est à dire $\lim_{x \to 0} \varepsilon(x) = 0$.

Si nous reprenons l'exemple de l'introduction, nous avons déterminer un $DL_5(0)$ à la fonction $f(x) = \dfrac{1}{1-x}$. En effet nous avons trouvé

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + x^5 + x^5\varepsilon(x)$$

où $\varepsilon(x) = \dfrac{-x}{1-x}$ qui tend bien 0 lorsque x tend vers 0.

Quitte à réaliser un changement de variable $t \leftarrow x-a$, on peut se ramener à n'étudier que les $DL_n(0)$ que nous noterons simplement DL_n.

Théorème Unicité

Si une fonction admet un DL_n alors il est unique.

Démonstration. On rappel que l'on note toutes les poubelles par la notation $\varepsilon(x)$.

Considérons deux DL_n d'une fonction f : $f(x) = P(x)+x^n\varepsilon(x) = Q(x)+x^n\varepsilon(x)$. Soit encore $R(x) = x^n\varepsilon(x)$ où $R = P - Q$. Puisque P et Q sont des polynômes de degrés au plus n, il en va de même pour R. En particulier $R(x) = \sum_{k=0}^{n} a_k x^k$ pour certains réels a_k. En calculant la limite en 0, on trouve que $R(0) = 0$ et donc $a_0 = 0$. C'est à dire que $R(x) = \sum_{k=1}^{n} a_k x^k$. En simplifiant par x dans l'expression $R(x) = x^n\varepsilon(x)$ on trouve $\sum_{k=0}^{n-1} a_{k+1}x^k = x^{n-1}\varepsilon(x)$. En prenant la limite lorsque x tend vers 0, on trouve de $a_1 = 0$. En divisant par x encore une fois et en prenant encore la limite on trouve que proche en proche que tous les a_k sont nuls et donc que $R = 0$ soit encore que $P = Q$ et donc que le DL_n est unique. □

Nous pourrons donc parler **du** DL_n au lieu **d'un** DL_n.

Proposition

Soient $n \in \mathbb{N}$ et $f(x) = P(x) + x^n\varepsilon(x)$ et $g(x) = Q(x) + x^n\varepsilon(x)$ des DL_n.

1. Pour tout $0 \leqslant p \leqslant n$, f admet un DL_p.
   Précisément le DL_p de f est le polynôme P tronqué au delà du degrés p.
2. Pour tout réel α et β, la fonction $\alpha f + \beta g$ admet un DL_n.
   Précisément le DL_n de $\alpha f + \beta g$ est $\alpha P + \beta Q$.
3. La fonction $f \times g$ admet un DL_n.
   Précisément le DL_n de $f \times g$ est $P \times Q$.
4. Si $g(0) \neq 0$ alors la fonction $\frac{f}{g}$ admet un DL_n.
   Précisément le DL_n de $\frac{f}{g}$ est le polynôme R quotient de la division euclidienne de P par Q suivant les puissances croissantes.
5. Si $g(0) = 0$ alors la fonction $f \circ g$ admet un DL_n.
   Précisément le DL_n de $f \circ g$ est $P \circ Q$.

Démonstration. Exercice □

1. Nous avons par exemple déterminer le DL_5 de $f(x) = \frac{1}{1-x} = 1 + x + x^2 + x^3 + x^4 + x^5 + x^5\varepsilon(x)$. Alors de DL_3 de f est $f(x) = 1 + x + x^2 + x^3 + x^3\varepsilon(x)$.
2. Nous pourrons, à l'aide d'une formule de Taylor, démontrer que $\exp(x) = 1+x+\frac{x^2}{2}+\frac{x^3}{6}+x^3\varepsilon(x)$. Ainsi le DL_3 de $\frac{1}{1-x} - \exp(x) = \frac{x^2}{2} + \frac{5x^3}{6} + x^3\varepsilon(x)$.
3. De même le DL_3 de $\frac{\exp(x)}{1-x}$ est $\left(1 + x + \frac{x^2}{2} + \frac{x^3}{6}\right)\left(1 + x + x^2 + x^3\right) + x^3\varepsilon(x)$. Dans cette dernière expression, en développant, tous les monômes de degrés strictement supérieur à 3 vont à la poubelle. Ainsi :

$$
\begin{aligned}
\frac{\exp(x)}{1-x} &= \left(1+x+\frac{x^2}{2}+\frac{x^3}{6}\right)\left(1+x+x^2+x^3\right)+x^3\varepsilon(x)\\
&= 1\left(1+x+x^2+x^3\right)\\
&+ x\left(1+x+x^2+x^3\right)\\
&+ \frac{x^2}{2}\left(1+x+x^2+x^3\right)\\
&+ \frac{x^3}{6}\left(1+x+x^2+x^3\right)\\
&+ x^3\varepsilon(x)\\
&= 1+x+x^2+x^3\\
&+ x+x^2+x^3\\
&+ \frac{x^2}{2}+\frac{x^3}{2}\\
&+ \frac{x^3}{6}\\
&+ x^3\varepsilon(x)\\
&= 1+2x+\frac{5}{2}x^2+\frac{8}{3}x^3+x^3\varepsilon(x)
\end{aligned}
$$

4. Pour réaliser le DL_3 de $\dfrac{1}{\exp(x)(1-x)} = \dfrac{f(x)}{\exp(x)}$ nous allons utiliser la deux méthodes.

   Première méthode : la composition des DL. On a

$$
\begin{aligned}
\exp(x)(1-x) &= \left(1+x+\frac{x^2}{2}+\frac{x^3}{6}+x^3\varepsilon(x)\right)(1-x)\\
&= 1\left(1+x+\frac{x^2}{2}+\frac{x^3}{6}\right)-x\left(1+x+\frac{x^2}{2}+\frac{x^3}{6}\right)+x^3\varepsilon(x)\\
&= 1+x+\frac{x^2}{2}+\frac{x^3}{6}-x-x^2-\frac{x^3}{2}+x^3\varepsilon(x)\\
&= 1-\frac{x^2}{2}-\frac{x^3}{3}+x^3\varepsilon(x)
\end{aligned}
$$

En utilisant la composition de ce DL_3 et du DL_3 de $\dfrac{1}{1-X}$ on a

$$
\begin{aligned}
\frac{1}{\exp(x)(1-x)} &= \frac{1}{1-\frac{x^2}{2}-\frac{x^3}{3}+x^3\varepsilon(x)} \\
&= \frac{1}{1-X} \qquad \text{où } X = \frac{x^2}{2}+\frac{x^3}{3}+x^3\varepsilon(x) \\
&= 1+X+X^2+X^3+X^3\varepsilon(X) \\
&= 1 \\
&+ \left(\frac{x^2}{2}+\frac{x^3}{3}\right) \\
&+ \left(\frac{x^2}{2}+\frac{x^3}{3}\right)^2 \\
&+ \left(\frac{x^2}{2}+\frac{x^3}{3}\right)^3 \\
&+ x^3\varepsilon(x) \\
&= 1+\frac{x^2}{2}+\frac{x^3}{3}+x^3\varepsilon(x)
\end{aligned}
$$

Seconde méthode : la division des polynômes. On réalise, comme cela est indiquer dans la proposition, la division euclidienne du polynôme $1+x+x^2+x^3$ par le polynôme $1+x+\frac{x^2}{2}+\frac{x^3}{6}$ selon les puissance croissante. On trouve

$$(1+x+x^2+x^3) = \left(1+x+\frac{x^2}{2}+\frac{x^3}{6}\right)\left(1+\frac{x^2}{2}+\frac{x^3}{3}\right)+\left(-\frac{7x^4}{12}-\frac{x^5}{4}-\frac{x^6}{18}\right)$$

En particulier tous les monômes du reste vont à la poubelle et le quotient donne $1+\frac{x^2}{2}+\frac{x^3}{3}$ (incroyable ! c'est le même résultat qu'avec la précédente méthode).

Proposition

Soit f une fonction dérivable autour de 0 tel que f' admette un DL_n : $f'(x) = \sum_{k=0}^{n} a_k x^k + x^n\varepsilon(x)$

Alors f admet un DL_{n+1}, précisément $f(x) = f(0) + \sum_{k=0}^{n} \frac{a_k}{k+1} x^{k+1} + x^{n+1}\varepsilon(x)$

Démonstration. Admise □

A présent que nous avons un théorème d'unicité, ainsi que des opérations, nous allons chercher un théorème d'existence.

Commençons par un cas simple : les polynômes.

Proposition

Considérons pour tout $x \in \mathbb{R}$, un polynôme $P(x) = \sum a_k x^k$. Alors pour tout $n \in \mathbb{N}$, P admet un DL_n. Précisément

$$P(x) = \sum_{k=0}^{n} a_k x^k + x^n\varepsilon(x)$$

où

Si $n < \deg(P)$ alors $\varepsilon(x) = \displaystyle\sum_{k=0}^{n-\deg(P)} a_{k+n}x^k$.

Si $n \geqslant \deg(P)$ alors $\varepsilon(x) = 0$.

Démonstration. Triviale □

Par exemple le polynôme $P(x) = 2 - x + \frac{x^2}{2} - x^3 + \frac{x^4}{6}$ admet :

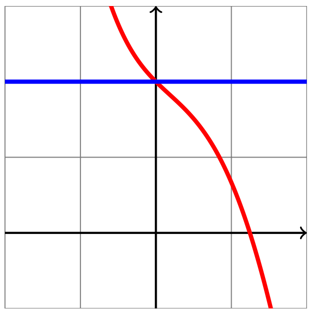
pour DL_0 **:** $P(x) = 2 + \varepsilon(x)$ où $\varepsilon(x) = -x + \frac{x^2}{2} - x^3 + \frac{x^4}{6}$.

pour DL_1 **:** $P(x) = 2 - x + x\varepsilon(x)$ où $\varepsilon(x) = \frac{x}{2} - x^2 + \frac{x^3}{6}$.

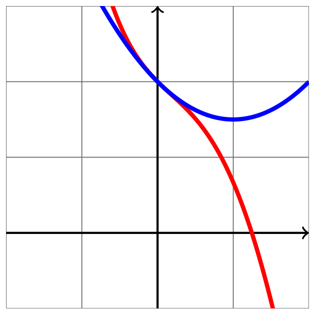
pour DL_2 **:** $P(x) = 2 - x + \frac{x^2}{2} + x^2\varepsilon(x)$ où $\varepsilon(x) = -x + \frac{x^2}{6}$.

pour DL_3 **:** $P(x) = 2 - x + \frac{x^2}{2} - x^3 + x^3\varepsilon(x)$ où $\varepsilon(x) = \frac{x}{6}$.

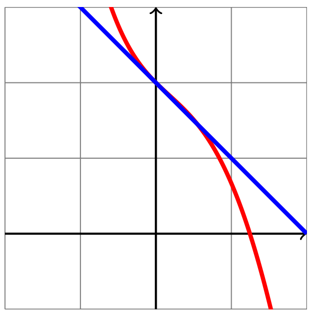
pour DL_n **pour tout** $n \geqslant 4$ **:** $P(x) = 2 - x + \frac{x^2}{2} - x^3 + \frac{x^4}{6} + x^n\varepsilon(x)$ où $\varepsilon(x) = 0$.



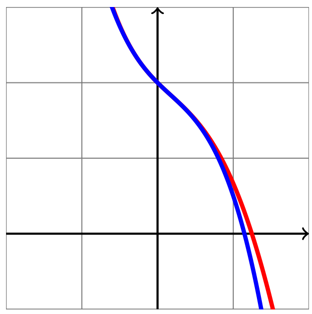
DL_0



DL_2



DL_1



DL_3

Passer par les polynômes peut être une bonne méthode pour trouver les coefficients des polynômes d'un DL_n. Dans l'exemple précédent, nous pouvons observer que 2, premier terme de chacun des DL_n est en fait $P(0)$. Que dire que -1, coefficient en x ? Nous souhaiterions l'exprimer uniquement en fonction de P. Passons par les dérivés.

On rappel que pour une fonction f, on note $f^{[n]}$ sa dérivé n-ième définie de manière récursive par les règles :

$$f^{[0]} = f, \qquad f^{[n+1]} = \left(f^{[n]}\right)'$$

Dans note exemple :

0. $P^{[0]}(x) = 2 - x + \frac{x^2}{2} - x^3 + \frac{x^4}{6}$
1. $P^{[1]}(x) = -1 + 2\frac{x}{2} - 3x^2 + 4\frac{x^3}{6}$
2. $P^{[2]}(x) = 2\frac{1}{2} - 3 \times 2x + 4 \times 3\frac{x^2}{6}$
3. $P^{[3]}(x) = -3 \times 2 + 4 \times 3 \times 2\frac{x}{6}$
4. $P^{[4]}(x) = 4 \times 3 \times 2\frac{1}{6}$

et $P^{[n]} = 0$ pour tout $n > 4$.

En particulier les valeurs en 0 donne :

0. $P^{[0]}(0) = 2 = a_0$
1. $P^{[1]}(0) = -1 = a_1$
2. $P^{[2]}(x) = 2\frac{1}{2} = 2a_2$
3. $P^{[3]}(x) = -3 \times 2 = 3 \times 2a_2$
4. $P^{[4]}(x) = 4 \times 3 \times 2\frac{1}{6} = 4 \times 3 \times 2a_4$

Lemme : Pour tout $x \in \mathbb{R}$, considérons $P(x) = \sum a_k x^k$ un polynôme à coefficient réels.

$$P^{[k]}(0) = a_k k!$$

Démonstration. Exercice □

Ainsi une autre manière de considérer l'écriture d'un polynôme $P(x) = \sum a_k x^k$ est

$$P(x) = \sum \frac{P^{[k]}(0)}{k!} x^k$$

Si cela est vrai pour les polynômes, ça doit être vrai pour n'importe quelle fonction f (dérivable suffisamment de fois).

Définition

Soient $n \in \mathbb{N}$ et $f : I \longrightarrow \mathbb{R}$ une fonction définie sur un intervalle $I \subseteq \mathbb{R}$ et dérivable n fois.

On appel **développement de Maclaurin d'ordre** n, le polynôme

$$P(x) = \sum_{k=0}^{n} \frac{f^{[k]}(0)}{k!} x^k$$

Reprenons notre fonction de l'introduction $f(x) = \dfrac{1}{1-x}$. L'étudiant passionné de récurrence pourra montrer que pour tout k, $f^{[k]}(x) = \dfrac{k!}{(1-x)^{k+1}}$ de sorte que $f^{[k]}(0) = k!$. Ainsi le développement de Maclaurin correspond exactement aux DL_n.

L'objectif est de montrer qu'un développement de Maclaurin d'ordre n est un DL_n. En effet la définition ne permet pas de caractériser le reste d'une égalité $f(x) = \displaystyle\sum_{k=0}^{n} \frac{f^{[k]}(0)}{k!} x^k + x^n \varepsilon(x)$.

1.4 Formules de Taylor

Théorème Formule de Taylor avec reste intégrale

Soient $n \in \mathbb{N}$ et $f : I \longrightarrow \mathbb{R}$ une fonction définie sur un intervalle $I \subseteq \mathbb{R}$ tel que $0 \in I$ et qui est $n+1$ fois dérivables et a dérivés continues sur I.

$$\forall x \in I, \quad f(x) = \sum_{k=0}^{n} \frac{f^{[k]}(0)}{k!} x^k + \int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\ dt$$

Démonstration. Pour tout $k \leqslant n$, posons $I_k(x) = \displaystyle\int_0^x \frac{(x-t)^k}{k!} f^{[k+1]}(t)\ dt$. Réalisons une intégration par partie de cette expression en posant $u'(t) = f^{[k+1]}(t)$ et $v(t) = \dfrac{(x-t)^k}{k!}$. Alors $u(t) = f^{[k]}(t)$ et $v'(t) =$

$-\dfrac{(x-t)^{k-1}}{(k-1)!}$.

$$\begin{aligned}
I_k(x) &= \int_0^x \frac{(x-t)^k}{k!} f^{[k+1]}(t)\ dt \\
&= \left[\frac{(x-t)^k}{k!} f^k(t)\right]_0^x - \int_0^x -\frac{(x-t)^{k-1}}{(k-1)!} f^k(t)\ dt \\
&= \left(\frac{(x-x)^k}{k!} f^k(x) - \frac{(x-0)^k}{k!} f^k(0)\right) + I_{k-1}(x) \\
&= -\frac{f^{[k]}(0)}{k!} x^k + I_{k-1}(x)
\end{aligned}$$

On observe de plus que

$$I_0(x) = \int_0^x \frac{(x-t)^0}{0!} f^{[0+1]}(t)\ dt = \int_0^x f'(t)\ dt = f(x) - f(0)$$

En réécrivant la formule $I_k(x) = -\dfrac{f^{[k]}(0)}{k!} x^k + I_{k-1}(x)$ on a $I_{k-1}(x) - I_k(x) = \dfrac{f^{[k]}(0)}{k!} x^k$. Sommons les termes :

$$\begin{aligned}
\sum_{k=1}^{n} \Big(I_{k-1}(x) - I_k(x)\Big) &= \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k \\
\Big(I_0(x) - I_1(x)\Big) + \Big(I_1(x) - I_2(x)\Big) + \cdots + \Big(I_{n-2}(x) - I_{n-1}(x)\Big) + \Big(I_{n-1}(x) - I_n(x)\Big) &= \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k \\
I_0(x) + \Big(-I_1(x) + I_1(x)\Big) + \Big(-I_2(x) + \cdots + I_{n-2}(x)\Big) + \Big(-I_{n-1}(x) + I_{n-1}(x)\Big) - I_n(x) &= \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k \\
I_0(x) - I_n(x) &= \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k \\
I_0(x) &= \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k + I_n(x) \\
f(x) - f(0) &= \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k + I_n(x) \\
f(x) &= \underbrace{f(0) + \sum_{k=1}^{n} \frac{f^{[k]}(x)}{k!} x^k}\, + I_n(x) \\
&= \sum_{k=0}^{n} \frac{f^{[k]}(x)}{k!} x^k + I_n(x) \\
&= \sum_{k=0}^{n} \frac{f^{[k]}(x)}{k!} x^k + \int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\ dt
\end{aligned}$$

□

Pour pouvoir conclure sur l'existence d'un DL_n, il suffit de montrer que $\displaystyle\int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\ dt = x^n \varepsilon(x)$.

Théorème Formule de Taylor-Lagrange

Soient $n \in \mathbb{N}$ et $f : I \longrightarrow \mathbb{R}$ une fonction définie sur un intervalle $I \subseteq \mathbb{R}$ tel que $0 \in I$ et qui est $n+1$ fois dérivables et a dérivés continues sur I.

Il existe un réel c_x dans le plus petit intervalle contenant 0 et x tel que

$$\forall x \in I, \quad f(x) = \sum_{k=0}^{n} \frac{f^{[k]}(0)}{k!} x^k + \frac{x^{n+1}}{(n+1)!} f^{[n+1]}(c_x)$$

Démonstration. Pour simplifier le raisonnement supposons que $x \geqslant 0$. Ainsi le plus petit intervalle contenant 0 et x est $[0;x]$ (sinon il faudrait travailler sur $[x;0]$ ce qui sera le même raisonnement).

Nous allons partir de la formule de Taylor avec reste intégrale et montrer que ce reste intégrale est de la forme énoncée par ce théorème.

Nous avons supposer que $f^{[n+1]}$ est continue sur I. En particulier, il existe deux réels m_x et M_x tel que $m_x \leqslant f^{[n+1]}(t) \leqslant M_x$ pour tout $t \in [0;x]$. On a ainsi pour tout $0 \leqslant t \leqslant x$:

$$\begin{aligned}
m_x \leqslant f^{[n+1]}(t) \leqslant M_x \quad & \Longrightarrow \quad m_x \frac{(x-t)^n}{n!} \leqslant \frac{(x-t)^n}{n!} f^{[n+1]}(t) \leqslant M_x \frac{(x-t)^n}{n!} \\
& \Longrightarrow \quad m_x \int_0^x \frac{(x-t)^n}{n!}\, dt \leqslant \int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\, dt \leqslant M_x \int_0^x \frac{(x-t)^n}{n!}\, dt \\
& \Longrightarrow \quad m_x \left[-\frac{(x-t)^{n+1}}{(n+1)!} \right]_0^x \leqslant \int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\, dt \leqslant M_x \left[-\frac{(x-t)^{n+1}}{(n+1)!} \right]_0^x \\
& \Longrightarrow \quad m_x \frac{x^{n+1}}{(n+1)!} \leqslant \int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\, dt \leqslant M_x \frac{x^{n+1}}{(n+1)!} \\
& \Longrightarrow \quad m_x \leqslant \frac{\displaystyle\int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\, dt}{\dfrac{x^{n+1}}{(n+1)!}} \leqslant M_x
\end{aligned}$$

D'après le théorème des valeurs intermédiaire, puisque $f^{[n+1]}$ est continue, il existe un réel $c_x \in [0;x]$ tel que

$$f^{[n+1]}(c_x) = \frac{\displaystyle\int_0^x \frac{(x-t)^n}{n!} f^{[n+1]}(t)\, dt}{\dfrac{x^{n+1}}{(n+1)!}}$$

□

Corollaire Formule de Taylor-Young

Soient $n \in \mathbb{N}$ et $f : I \longrightarrow \mathbb{R}$ une fonction définie sur un intervalle $I \subseteq \mathbb{R}$ tel que $0 \in I$ et qui est $n+1$ fois dérivables et a dérivés continues sur I.

$$\forall x \in I, \quad f(x) = \sum_{k=0}^{n} \frac{f^{[k]}(0)}{k!} x^k + x^n \varepsilon(x)$$

où $\lim\limits_{x \to 0} \varepsilon(x) = 0$

Démonstration. Partant de la formule de Taylor-Lagrange, nous allons montrer que $\dfrac{x^{n+1}}{(n+1)!} f^{[n+1]}(c_x) = x^n \varepsilon(x)$. Le candidat naturel est $\varepsilon(x) = x \dfrac{f^{[n+1]}(c_x)}{(n+1)!}$. Il suffit de montrer que cette quantité tend vers 0. Or $\lim\limits_{x \to 0} f^{[n+1]}(c_x) = f^{[n+1]}(0)$ puisque $f^{[n+1]}$ est continue et que c_x est dans le plus petit intervalle contenant 0 et x donc lorsque x tend vers 0 le nombre c_x tend vers un nombre dans le plus petit intervalle contenant 0 et 0 donc 0. Ainsi $\lim\limits_{x \to 0} x \dfrac{f^{[n+1]}(c_x)}{(n+1)!} = 0$. □

Ca y est ! Nous avons montrer que ce que les polynômes nous avaient laisser entrevoir (Maclaurin) est juste. Attention cependant : pour qu'il existe un DL_n, il est suffisant que la fonction soit $(n+1)$ fois dérivables et a dérivées continues.

Mais ce n'est pas une condition nécessaire : il est possible qu'une fonction admette une DL_n sans qu'elle ne soit $(n+1)$ fois dérivable à dérivé continue.

Par exemple la fonction $f(x) = x^3 \sin\left(\dfrac{1}{x}\right)$ admet un DL_2 mais n'est pas deux fois dérivable en 0.

En effet, on a $f(x) = x^2 \left(x\sin\left(\frac{1}{x}\right) \right)$ et pour tout $x \in \mathbb{R}^*$, $\left|\sin\left(\frac{1}{x}\right)\right| \leqslant 1$ donc $\left|x\sin\left(\frac{1}{x}\right)\right| \leqslant |x|$. En particulier en passant à la limite lorsque x tend vers 0 on arrive à montrer que $x\sin\left(\frac{1}{x}\right)$ tend vers 0 et est donc une poubelle $f(x) = 0 + x^2\varepsilon(x)$. Ceci prouve que la fonction f admet un DL_2.

Par le même procédé (théorème des gendarmes) on peut montrer que f est continue [1] en 0 et que $f(0) = 0$.

En reprenant la définition de dérivé en 0, on a

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to} x^2\sin\left(\frac{1}{x}\right) = 0$$

en raisonnant encore grâce au théorème des gendarmes et donc f est dérivable en 0 et $f'(0) = 0$.

Pour tout $x \neq 0$, en dérivant un produit et une composé, on a $f'(x) = 3x^2\sin\left(\frac{1}{x}\right) + x^3\left(-\frac{1}{x^2}\right)\cos\left(\frac{1}{x}\right) = 3x^2\sin\left(\frac{1}{x}\right) - x\cos\left(\frac{1}{x}\right)$.

Par définition :

$$f''(0) = \lim_{x\to 0} \frac{f'(x) - f'(0)}{x - 0} = \lim_{x\to 0} 3x\sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right)$$

Si, encore par les gendarmes, on peut montrer que $\lim_{x\to 0} 3x\sin\left(\frac{1}{x}\right) = 0$, on peut également montrer que $\lim_{x\to 0} \cos\left(\frac{1}{x}\right) = \lim_{X\to\infty} \cos(X)$ n'existe pas et donc la fonction f n'est pas deux fois dérivable en 0.

Nous avons cependant les résultats réciproques parcellaires suivants.

Proposition

Soit $f : I \longrightarrow \mathbb{R}$ une fonction définie sur un intervalle $I \subseteq \mathbb{R}$ tel que $0 \in I$.

- Si f admet un DL_0 alors f admet une limite finie en 0
- Si f admet un DL_1 alors f est dérivable en 0

Démonstration.

- Soit $f(x) = a + \varepsilon(x)$ un DL_0. En passant à la limite en 0 on prouve le résultat.
- Soit $f(x) = a + bx + x\varepsilon(x)$ un DL_1 alors

$$f'(0) = \lim_{x\to 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x\to 0} \frac{bx + x\varepsilon(x)}{x} = \lim_{x\to 0} b + \varepsilon(x) = b$$

et f est donc dérivable en 0.

□

1.5 Formulaire

En appliquant la formule de Taylor-Young ou en intégrant ou dérivant des DL nous obtenons les formules suivantes.

- $e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + x^n\varepsilon(x)$.
- $\cos(x) = 1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^k\frac{x^{2k}}{(2k)!} + x^{2k}\varepsilon(x)$.
- $\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots + (-1)^k\frac{x^{2k+1}}{(2k+1)!} + x^{2k+1}\varepsilon(x)$.
- $(1+x)^\alpha = 1 + \alpha x + \alpha(\alpha-1)\frac{x^2}{2} + \alpha(\alpha-1)(\alpha-2)\frac{x^3}{3!} + \cdots + \alpha(\alpha-1)\cdots(\alpha-(n-1))\frac{x^n}{n!} + x^n\varepsilon(x)$

1. Plus rigoureusement f admet un prolongement continue

$$\alpha = -1 : \quad \frac{1}{1+x} = 1 - x + x^2 - x^3 + \cdots + (-1)^n x^n + x^n \varepsilon(x)$$

$$\alpha = -1 \ + \ (\textstyle\int) : \quad \ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots + (-1)^{n+1}\frac{x^n}{n} + x^n \varepsilon(x)$$

$$\alpha = -1 \ + \ (X = -x) : \quad \frac{1}{1-X} = 1 + X + X^2 + X^3 + X^4 + \cdots + X^n + X^n \varepsilon(X)$$

$$\alpha = \frac{1}{2} : \quad \sqrt{1+x} = 1 + \frac{1}{2}x + \frac{1}{2}\left(-\frac{1}{2}\right)\frac{x^2}{2!} + \frac{1}{2}\left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\frac{x^3}{3!} + \cdots + \left(\prod_{k=1}^{n} \frac{1-2k}{2}\right)\frac{x^n}{n!} + x^n \varepsilon(x)$$

2. Fonctions à plusieurs variables

2.1 Introduction

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 :

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

Autre exemple :

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

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

2.2 Représentation

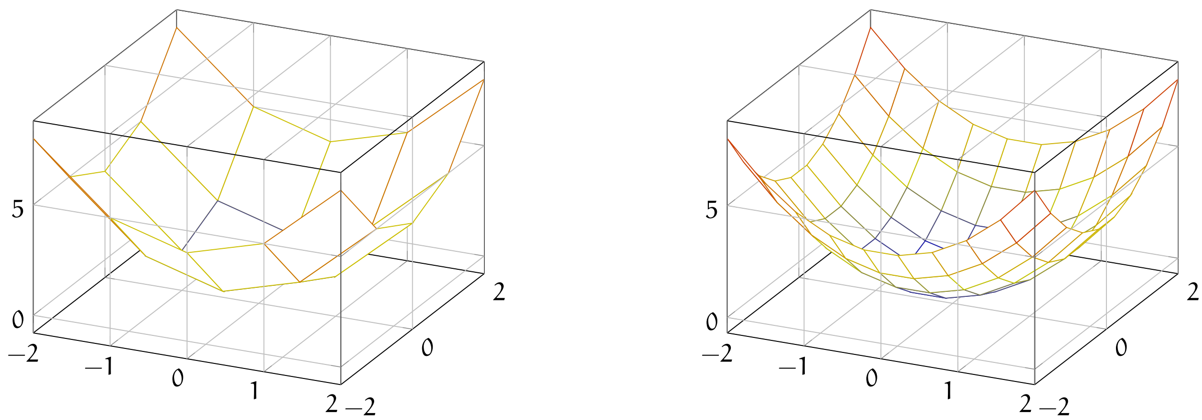
Avec les fonctions à une variable nous pouvions tracer le graphe $(x, f(x))$ pour les valeurs de x dans le domaine de définition de f.

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.

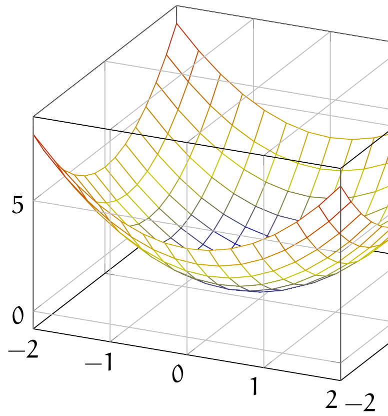
On prend différentes valeurs de x et y et on relie les points. Plus on prend de valeur plus la représentation se raffine. Il est parfois utile d'ajouter un dégradé de couleur.

$f(x, y) = x^2 + y^2$ avec 5×5 points

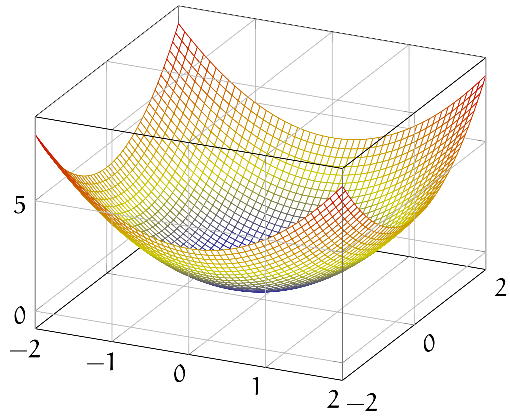
$f(x, y) = x^2 + y^2$ avec 10×10 points



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

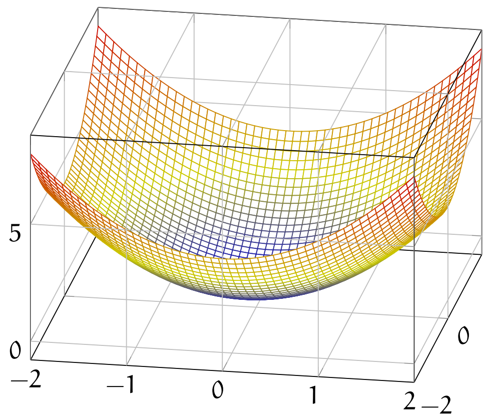


$f(x, y) = x^2 + y^2$ avec 50×50 points

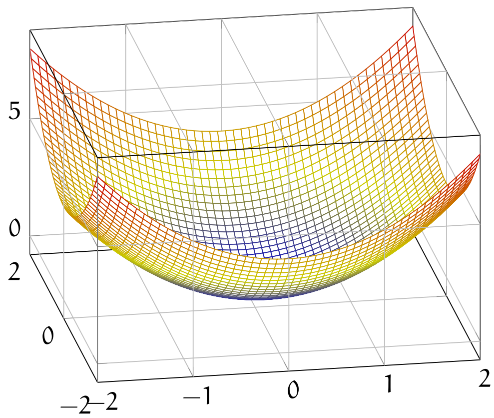


La représentation sous différent angle de vue permet aussi de mieux appréhender la fonction.

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



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



2.3 Lignes de niveau et fonctions partielles

Définition

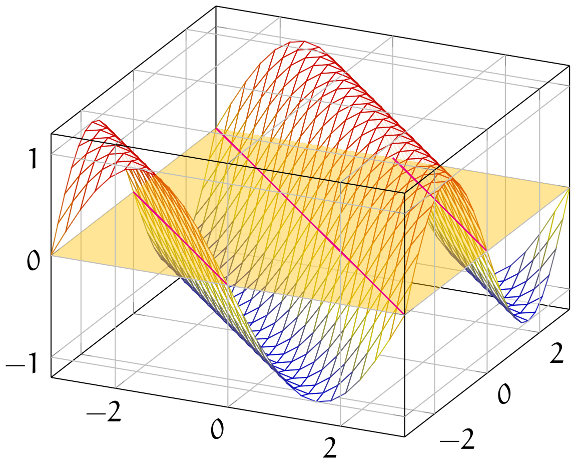
Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $h \in R$. La ligne de niveau h de f est l'ensemble

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

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 $k \in \mathbb{Z}$.

Lignes de niveau 0 de $f(x, y) = \sin(x + y)$



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

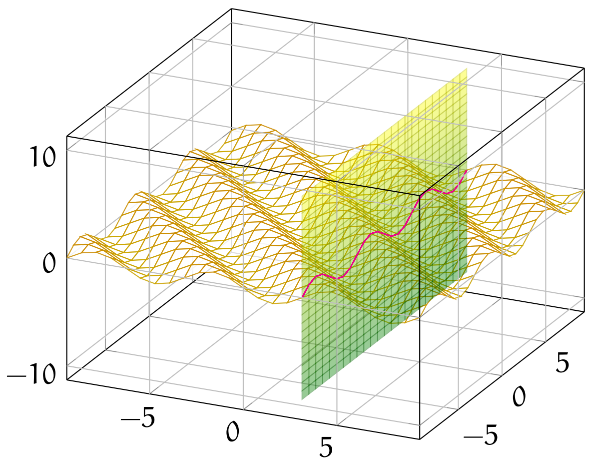
$$f_{|y=b}(x) = f(x, b) \qquad f_{|x=a}(y) = f(a, y)$$

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

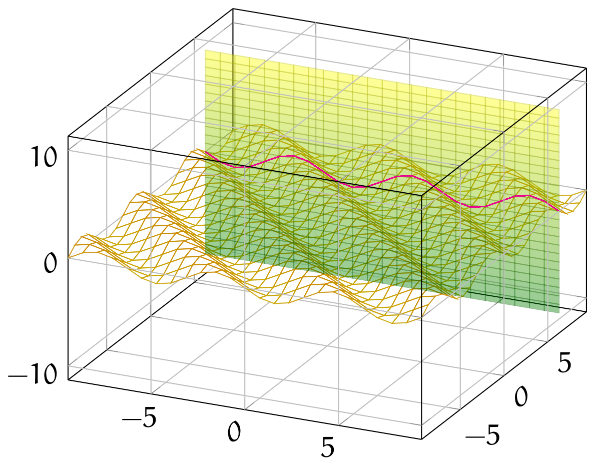
Reprenons l'exemple de la fonction $f(x, y) = \sin(x + y)$. Alors

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

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



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



2.4 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 $\lim_{x \to a} \dfrac{f(x) - f(a)}{x - a}$. Avec une variable, la notion de limite est assez claire, soit on se rapproche de a par la gauche, ce qui est noté $\lim_{x \to a^-}$, soit par la droite, noté $\lim_{x \to a^+}$. 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 $\dfrac{u}{v}$ qui est $\dfrac{u'v - v'u}{v^2}$:

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

En utilisant la dérivé de $\dfrac{1}{u}$ qui est $-\dfrac{u'}{u^2}$, 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).

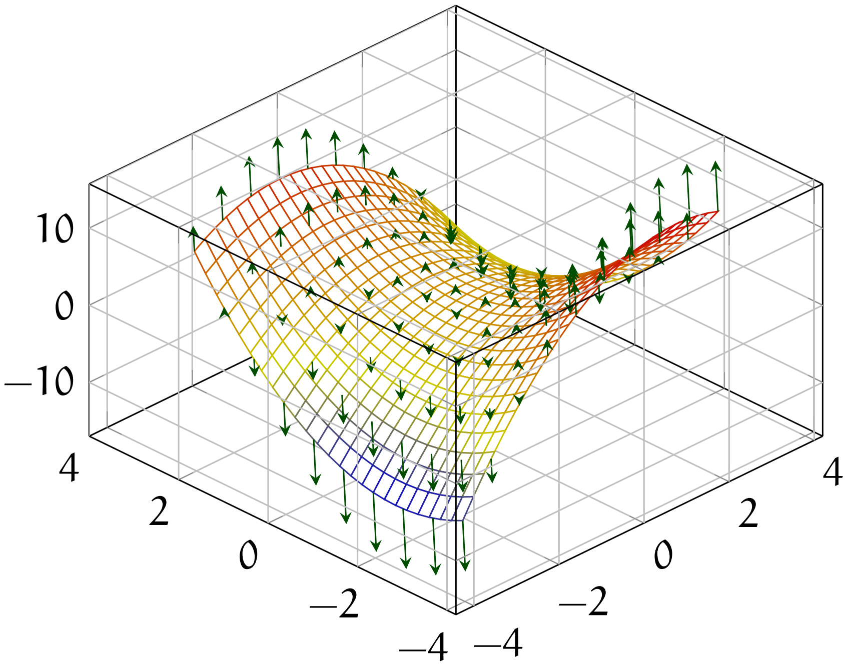
Avec deux variables, c'est la même chose. On parle du gradient.

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

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

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.



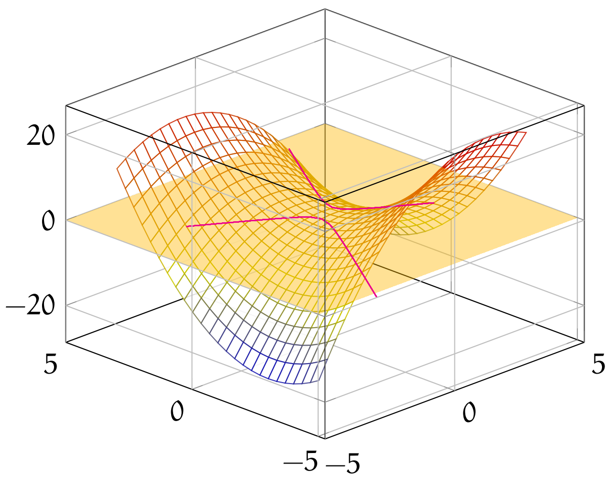
Proposition

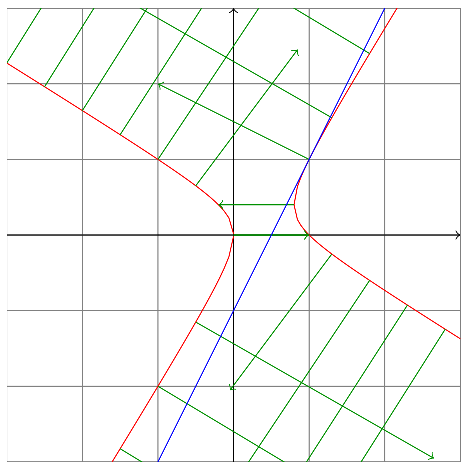
Le gradient d'une fonction en (a,b) est orthogonale à la ligne de niveau $f(a,b)$.

Démonstration. Admise $\square$

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. $\square$

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 ?

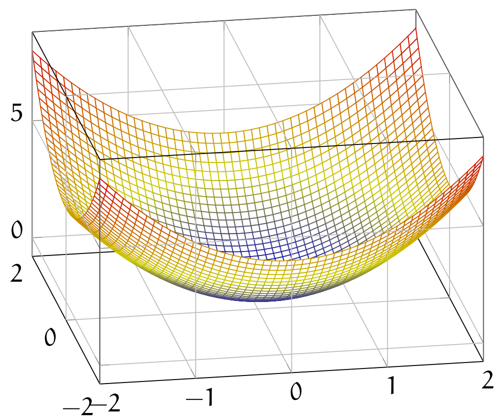
2.5 Extrema globaux et locaux

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

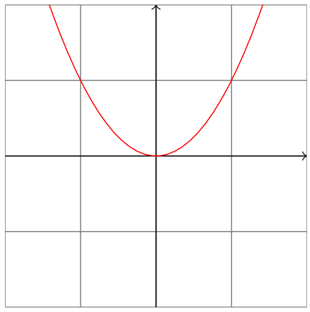
Soient $E \subseteq \mathbb{R}^2$, $f : E \to \mathbb{R}$ une fonction à deux variables et $(a, b) \in E$.

Si (a, b) est un extrema local alors c'est un point critique.

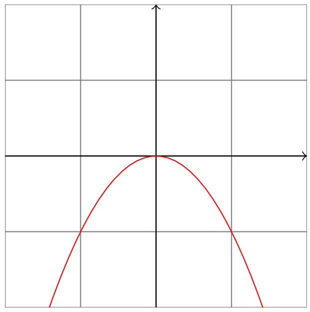
Démonstration. Admise $\square$

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.

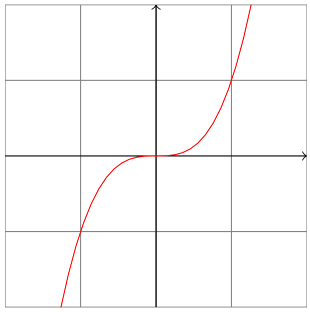
Considérons la fonction $f(x) = x^2$. Sa dérivé est $f'(x) = 2x$ qui s'annule trivialement en $x = 0$. Ainsi 0 est un point critique. Nous savons, en regardant le graphique que c'est un minimum. Regardons la dérivé seconde : $f''(x) = 2$... c'est positif en $x = 0$.



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.

2.6 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.

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

$\square$

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

si $\mathrm{tr}(H) < 0$ 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 $\mathrm{tr}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a + d$

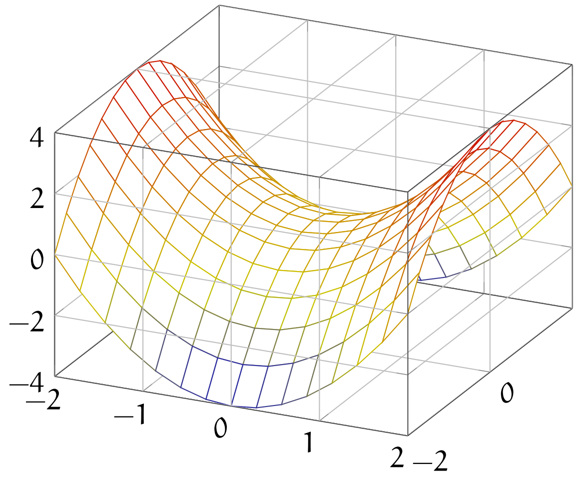
§ On peut démontrer que si $\texttt{det}(H) > 0$ alors la trace est nécessairement non nulle de sorte que nous n'avons pas oublié de traiter un cas dans le théorème précédent.

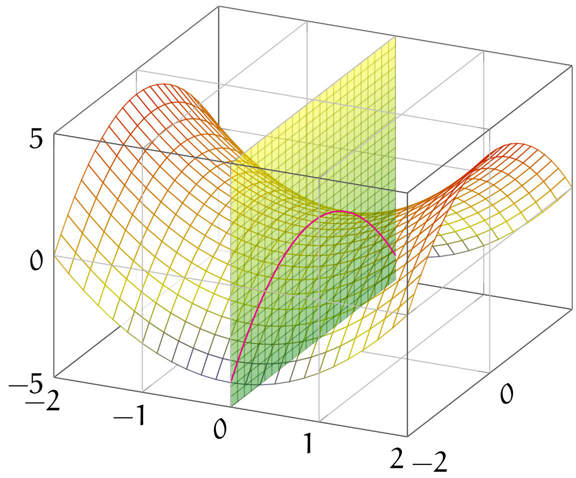
Par exemple nous avions trouver que le point critique de la fonction $f(x, y) = -x^2 - xy + x + y^2$ est $\left(\frac{1}{5}, \frac{2}{5}\right)$.

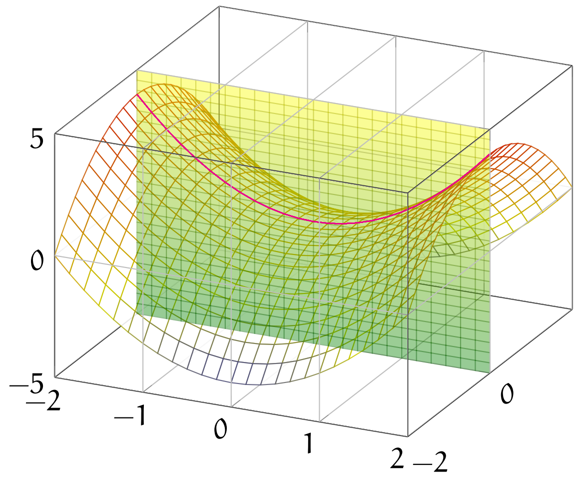
La hessienne est $\begin{pmatrix} -2 & -1 \\ -1 & 2 \end{pmatrix}$. Son déterminant vaut -5. Le point $\left(\frac{1}{5}, \frac{2}{5}\right)$ est un point selle.

§ 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 $\overrightarrow{\mathrm{Grad}}f = \begin{pmatrix}2x\\-2y\end{pmatrix}$ 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.







2.7 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.

En effet la matrice hessienne est une matrice symétrique. Elle est donc diagonalisable. Les vecteurs propre de valeur propre positive donnent les directions suivant lesquelles la fonction admet un minimum et celle qui sont positives donnent des maximums.

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 $\overrightarrow{\mathrm{Grad}}(f) = \begin{pmatrix}2x\\z\\y+1\end{pmatrix}$ qui s'annule en $(x,y,z)=(0,-1,0)$. On détermine sa hessienne tout aussi aisément : $H=\begin{pmatrix}2 & 0 & 0\\0 & 0 & 1\\0 & 1 & 0\end{pmatrix}$.

Diagonalisons cette matrice.

Première étape. Déterminons le polynôme caractéristique.

$$\chi(X) = \mathtt{det}(H - XId) = \begin{vmatrix} 2-X & 0 & 0 \\ 0 & -X & 1 \\ 0 & 1 & -X \end{vmatrix} = -(X-2)(X-1)(X+1)$$

Seconde étape. Déterminons les valeurs propres, solution de $\chi(X) = 0$. Ici on trouve facilement les trois valeurs propres : -1, 1 et 2.

Troisième étape. Déterminons les espaces propres.

Espace propre de la valeur propre -1. Il s'agit de trouver un vecteur $X \in \mathbb{R}^3$ tel que $(H + Id)X = \overrightarrow{0}$.

$$\begin{cases} 3x & & & = & 0 \\ & y & + & z & = 0 \\ & y & + & z & = 0 \end{cases}$$

qui trouve une infinité de solution $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = y \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$.

Espace propre de la valeur propre 1. Il s'agit de trouver un vecteur $X \in \mathbb{R}^3$ tel que $(H - Id)X = \overrightarrow{0}$.

$$\begin{cases} x & & & = & 0 \\ - & y & + & z & = 0 \\ & y & - & z & = 0 \end{cases}$$

qui trouve une infinité de solution $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = y \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.

Espace propre de la valeur propre 2. Il s'agit de trouver un vecteur $X \in \mathbb{R}^3$ tel que $(H - 2Id)X = \overrightarrow{0}$.

$$\begin{cases} & & & 0 & = & 0 \\ - & 2y & + & z & = 0 \\ & y & - & 2z & = 0 \end{cases}$$

qui trouve une infinité de solution $\begin{pmatrix} x \\ y \\ z \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

Conclusion. Dans la base $\left\{ \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$ la hessienne s'écrit $\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$

On en déduit que l'on a une *sorte de point selle* : en coupant la fonction (*ie* ligne de niveau) suivant le plan engendré par les vecteurs $\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$ et $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ la fonction admet un minimum et suivant la droite $\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$ la fonction admet un maximum au point $(0, -1, 0)$.

3. Application : Méthode des moindre carrés

3.1 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

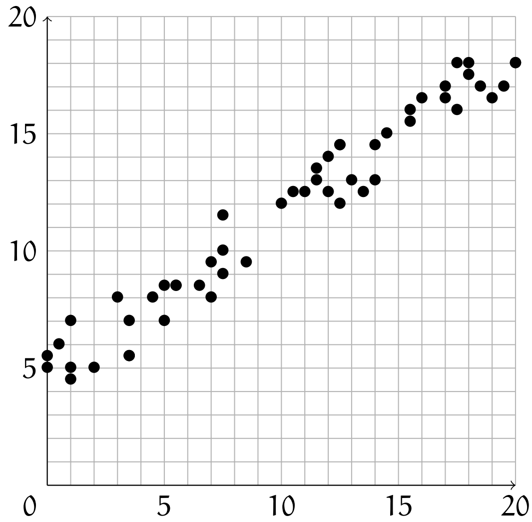
$$y = ax + b$$

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 [2] 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

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

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*).

2. de Legendre et Gauss

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

$$
\Rightarrow \begin{cases} \sum_i x_iy_i - ax_i^2 - bx_i &= 0 \\ \sum_i y_i - ax_i - b &= 0 \end{cases}
$$

$$
\Rightarrow \begin{cases} \frac{1}{52}\sum_i x_iy_i - ax_i^2 - bx_i &= 0 \\ \frac{1}{52}\sum_i y_i - ax_i - b &= 0 \end{cases}
$$

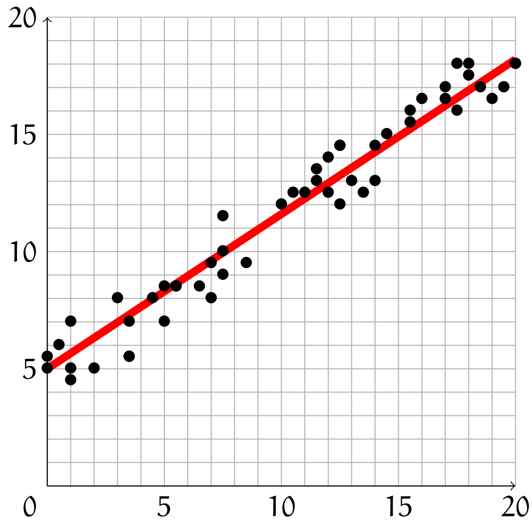
$$
\Rightarrow \begin{cases} \frac{1}{52}\sum_i x_iy_i - a\frac{1}{52}\sum_i x_i^2 - b\frac{1}{52}\sum_i x_i &= 0 \\ \frac{1}{52}\sum_i y_i - a\frac{1}{52}\sum_i x_i - b\frac{1}{52}\sum_i 1 &= 0 \end{cases}
$$

$$
\Rightarrow \begin{cases} \overline{xy} - a\overline{x^2} - b\overline{x} &= 0 \\ \overline{y} - a\overline{x} - b &= 0 \end{cases}
$$

$$
\Rightarrow \begin{cases} a\overline{x^2} + b\overline{x} &= \overline{xy} \\ a\overline{x} + b &= \overline{y} \end{cases}
$$

$$
\Rightarrow \begin{cases} a\overline{x^2} + b\overline{x} &= \overline{xy} \\ a\overline{x}^2 + b\overline{x} &= \overline{y}\cdot\overline{x} \end{cases}
$$

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.

3.2 Cas général

On dispose de n observations de deux caractères quantitatifs $x=(x_1,\ldots,x_n)$ et $y=(y_1,\ldots,y_n)$, 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 $F(x,\vartheta)=ax_i+b$

3.3 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).