

Calcul différentiel : le cas à une variable

Florent Capelli

12/10/2026

1 Fonctions de $\mathbb{R}$ dans $\mathbb{R}$

1.1 Notion de limites et notations $o(\cdot)$, $O(\cdot)$

Soit $f: I \rightarrow \mathbb{R}$ où I est un intervalle, c'est-à-dire une ensemble de la forme $[a, b]$, $[a, b[$, $]a, b]$ ou $]a, b[$ et $a < b$ (et $a, b \in \mathbb{R} \cup \{-\infty, \infty\}$).

Soit $a \in I$. On dit que f admet une limite ℓ en a si et seulement si pour tout $\epsilon > 0$, il existe $\eta > 0$ tel que pour tout b , si $|b - a| < \eta$ alors $|f(b) - \ell| < \epsilon$. En d'autres termes, plus b se rapproche de a , plus $f(b)$ s'approche de ℓ . Si cette valeur existe, on note alors $\ell = \lim_{b \rightarrow a} f(b)$.

Si $a = +\infty$, on dit que f admet une limite ℓ à l'infini si pour tout $\epsilon > 0$, il existe $N > 0$ tel que pour tout b , si $b > N$ alors $|f(b) - \ell| < \epsilon$. C'est-à-dire que plus b grandit, plus $f(b)$ s'approche de ℓ . On note alors $\ell = \lim_{b \rightarrow +\infty} f(b)$. On définit de même $\lim_{b \rightarrow -\infty} f(b)$ en remplaçant $b > N$ par $b < -N$ dans la définition.

Exemple 1.1. On montre que $\lim_{x \rightarrow \infty} \frac{1}{x} = 0$. On peut en effet choisir $N = 1/\epsilon$ dans la définition. Donc si $b > N$, alors $f(b) = \frac{1}{b} < \epsilon$ et donc $|f(b) - 0| < \epsilon$.

Soit f et g deux fonctions définies sur un intervalle I et $a \in I$ (a est possiblement $-\infty$ ou $+\infty$). Alors on écrit :

- $f = o_a(g)$ si $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 0$. On dit alors que f est négligeable devant g au voisinage de a .
- $f = O_a(g)$ s'il existe $A < B$ finis et $\eta > 0$ tels que pour tout $h < \eta$, $A < f(a+h)/g(a+h) < B$. C'est-à-dire que f/g est borné au voisinage de a .
- $f \simeq_a g$ si $\lim_{x \rightarrow a} \frac{f(x)}{g(x)} = 1$. On dit alors que f est équivalente à g au voisinage de a .

On note simplement $f = o(g)$ quand $a = 0$. On note cependant plutôt $f \simeq g$ et $f = O(g)$ quand $a = \infty$.

Exemple 1.2. • $x = o_\infty(x^2)$ mais $x^2 = o(x)$.

- $\ln(1+x) \simeq_0 x$.
- $3n^2 + 2n + 4 = O(n^2)$.

1.2 Continuité

On dit qu'une fonction est continue en $a \in \mathbb{R}$ si $f(a) = \lim_{b \rightarrow a} f(b)$.

La fonction $x \mapsto \text{signe}(x)$ (en prenant la convention $\text{signe}(0) = 1$) n'est par exemple pas continue en 0. En effet, $|f(b) - f(0)| = 2$ pour tout $b < 0$ donc on ne peut pas s'approcher arbitrairement de $f(0)$ ici. Graphiquement, cela se traduit par un saut dans la représentation.

Exercice 1.1. Montrez que $x \mapsto x$ est continue en tout point $a \in \mathbb{R}$. Même question pour $x \mapsto \alpha x + \beta$.

Correction 1. Pour $x \mapsto x$, on choisit $\eta = \epsilon$ dans la définition. Pour $x \mapsto \alpha x + \beta$, on a choisit $\eta = \frac{\epsilon}{|\alpha|}$. On a alors $|f(b) - f(a)| = |\alpha(b - a)| \leq |\alpha| \cdot |b - a| \leq |\alpha| \eta \leq \epsilon$.

Théorème 1.1. Si f et g sont continues en a alors $f + g$ et $f \times g$ sont continues en a . Si f est continue en $g(a)$, alors $(f \circ g)$ est continue en a .

Exemple 1.3. • $x \mapsto P(x)$ où P est un polynôme est continue partout. En effet $x \mapsto c_i x^i$ est continue pour tout $c_i \in \mathbb{R}$ partout car c'est un produit de i fonctions continues et $P = \sum_{i=1}^n c_i x^i$ est une somme de fonctions continues.

- $x \mapsto \cos(x^{24} + x^3 - 3x)$ est continue en tout point $a \in \mathbb{R}$ parce que $\cos$ est une fonction continue partout.

Même si cela reste un cas pathologique Il existe des fonctions qui ne sont continues nulle part. Par exemple, la fonction de Dirichlet qui associe x à 1 si $x \in \mathbb{Q}$ et x à 0 sinon n'est continue nulle part dans $\mathbb{R}$.

On note $\mathcal{C}_0(I, \mathbb{R})$ l'ensemble des fonctions continues de I dans $\mathbb{R}$ où $I \subseteq \mathbb{R}$ est un intervalle.

1.3 Dérivabilité

La notion de dérivabilité se fonde sur la remarque suivante : pour beaucoup de fonctions f , si on les observe dans un intervalle très proche de $a \in \mathbb{R}$, elles “ressemblent” à une fonction linéaire, c'est-à-dire à une droite. La dérivée $f'(a)$ de f en a est alors la “pente” de cette droite, c'est-à-dire la valeur telle que $f(a + h) \simeq f(a) + f'(a)h$ pour h proche de a . Alternativement, on peut voir cette droite comme la tangente à la courbe. On verra que la première intuition se généralise mieux à des fonctions de plusieurs variables.

En réécrivant l'équation de la droite ci-dessus (attention, cela n'est pas très formel, nous développons pour l'intuition), on se rend compte que $f'(a) \simeq \frac{f(a+h) - f(a)}{h}$ quand h est petit (voir dessins au tableau).

On définit donc la valeur suivante : soit $f: I \rightarrow \mathbb{R}$ une fonction et $a \in I$. On dit

que f est *dérivable en a* si la valeur suivante existe :

$$\ell := \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

On dit alors que ℓ est la *dérivée de f en a* . On la note $f'(a)$, ou bien $\frac{\partial f}{\partial x}(a)$.

Notons ici que cette limite n'est pas toujours bien définie. En effet, si on regarde la fonction $f(x) = |x|$ en 0, on voit que si $h < 0$, alors $\frac{f(0+h)-f(0)}{h} = \frac{|h|}{h} = -1$ alors que si $h > 0$, $\frac{f(0+h)-f(0)}{h} = 1$. Donc la limite en 0 n'existe pas. Cela veut simplement dire que f n'est pas dérivable en 0 (elle l'est partout ailleurs).

On note $\mathcal{C}_1(I, \mathbb{R})$ l'ensemble des fonctions de I dans $\mathbb{R}$ qui sont dérivables en tout point de I .

Exercice 1.2. Montrez que la fonction $x \mapsto ax + b$ est dérivable partout et que $f'(x) = a$ pour tout $x \in \mathbb{R}$.

Correction 2. Pour tout h , on a $f(x+h) - f(x) = f(x) + f(h) - f(x) = ah$ donc $\lim_{h \rightarrow 0} \frac{f(a+h)-f(h)}{h}$ est bien définie et vaut a .

Exercice 1.3. Montrez que la fonction $x \mapsto ax^2$ est dérivable partout et que $f'(x) = 2ax$ pour tout $x \in \mathbb{R}$.

Correction 3. Pour tout h , on a $f(x+h) - f(x) = a(x^2 + h^2 + 2ah - x^2)$ donc $\frac{f(a+h)-f(h)}{h} = ah + 2a \xrightarrow{h \rightarrow 0} 2a$.

Exercice 1.4. Montrez que la fonction $x \mapsto \sqrt{x}$ est dérivable en tout $a > 0$ et que $f'(a) = \frac{1}{2\sqrt{a}}$.

Correction 4. Pour tout h , on a

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

Donc

$$\begin{aligned} \frac{f(a+h) - f(a)}{h} &= \frac{1}{\sqrt{a+h} + \sqrt{a}} \\ &\xrightarrow{h \rightarrow 0} \frac{1}{2\sqrt{a}} \end{aligned}$$

L'intérêt de la dérivée est qu'on peut souvent la calculer de manière formelle.

Théorème 1.2. *On a les résultats suivants :*

- $(af)' = af'$ pour une constante a .
- $(f + g)' = f' + g'$
- $(f \times g)' = f'g + fg'$
- $(f \circ g)' = (f' \circ g) \cdot g'$

Ainsi que les résultats présentés dans le tableau 1.

On remarque que si f est dérivable en a alors f est nécessairement continue. Autrement dit, $\mathcal{C}_1(I, \mathbb{R}) \subseteq \mathcal{C}_0(I, \mathbb{R})$, mais cette inclusion est stricte (voir par exemple, la fonction $x \mapsto |x|$).

Il est bien connu qu'on peut comprendre la variation d'une fonction en fonction du signe de sa dérivée : si $f'(x_0) > 0$ alors la fonction est localement croissante. Si $f'(x_0) < 0$, alors elle est localement décroissante. En effet, si $f'(x_0) > 0$, alors $\frac{f(x_0+h)-f(x_0)}{h} > 0$ pour h suffisamment petit. Donc pour $h > 0$ suffisamment petit, on a $f(x_0 - h) < f(x_0) < f(x_0 + h)$ ce qui montre bien la croissance de f . Si $f'(x_0) < 0$, l'inégalité est inversée et alors f est localement décroissante.

Lorsque $f'(x_0)$ est nulle, on ne peut rien conclure. Cependant, on sait la chose suivante : si x_0 est un minimum local, alors $f'(x_0) = 0$ (le contraire n'est pas forcément vrai, considérez par exemple $x \mapsto x^3$ en 0). Pour une fonction $f: I \rightarrow \mathbb{R}$, un *minimum local* est une valeur $x_0 \in I$ telle qu'il existe $\eta > 0$ pour lequel si $|h| < \eta$, alors $f(x_0) \leq f(x_0 + h)$. Si $f \in \mathcal{C}_1(I, \mathbb{R})$, alors on peut voir que $f'(x_0) = 0$. En effet, on a

$$\frac{f(x_0 + h) - f(x_0)}{h} \begin{cases} \geq 0 & \text{si } h \leq 0 \\ \leq 0 & \text{sinon} \end{cases}$$

donc la limite $f'(x_0)$ quand $h \rightarrow 0$ (qui existe) doit être nul.

Notation. Un point de notation qui pourra être utile pour plus tard. La dérivation d'une fonction composée est centrale en apprentissage machine. On a écrit $(f \circ g)'(x) = f'(g(x))g'(x)$. Dans le cas à une variable, cela est plutôt lisible. Lorsque notre fonction dépend de plusieurs variables (on verra cela la semaine prochaine), on a besoin parfois d'être plus explicite : on veut dériver une fonction selon une de ses variables, en considérant les autres comme des constantes. Dans ce cas, la notation de Leibniz est intéressante. On note $f' = \frac{\partial f}{\partial x}$. Intuitivement, on regarde la *variation de f* (df) lorsqu'on fait varier x (dx). Si on écrit : $y = g(x)$, la dérivée de $f \circ g$ revient à regarder df en fonction de dy et la variation dy en fonction de x . En notation :

$$\frac{\partial f}{\partial x} = \frac{\partial f}{\partial y} \times \frac{\partial y}{\partial x}.$$

Fonction $f(x)$	Dérivée $f'(x)$
k	0
kx	k
x^n	nx^{n-1}
$\frac{1}{x^n} = x^{-n}$	$-nx^{-n-1} = -\frac{n}{x^{n+1}}$
$\sqrt[n]{x} = x^{1/n}$	$\frac{1}{n}x^{1/n-1} = \frac{1}{n\sqrt[n]{x^{n-1}}}$
x^α	$\alpha x^{\alpha-1}$
$\ln x $	$\frac{1}{x}$
$\log_a x $	$\frac{1}{x \ln a}$
e^x	e^x
a^x	$a^x \ln a$
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\frac{1}{\cos^2 x} = 1 + \tan^2 x$
$\cot x$	$-\frac{1}{\sin^2 x} = -1 - \cot^2 x$
$\arcsin x$	$\frac{1}{\sqrt{1-x^2}}$
$\arccos x$	$-\frac{1}{\sqrt{1-x^2}}$
$\arctan x$	$\frac{1}{1+x^2}$
$\sinh x$	$\cosh x$
$\cosh x$	$\sinh x$
$\tanh x$	$\frac{1}{\cosh^2 x} = 1 - \tanh^2 x$
$\coth x$	$-\frac{1}{\sinh^2 x} = 1 - \coth^2 x$
$\operatorname{arsinh} x$	$\frac{1}{\sqrt{1+x^2}}$
$\operatorname{arcosh} x$	$\frac{1}{\sqrt{x^2-1}}$
$\operatorname{artanh} x$	$\frac{1}{1-x^2}$
$\frac{1}{1+e^{-x}}$	$f(x)(1-f(x))$

Table 1: Fonction dérivée principales [source](#)

1.4 Développement en série de Taylor

Soit f une fonction dérivable en a . On définit $g_a(h) = f(a+h) - f(a) - f'(a)h$. On remarque que $\frac{g_a(h)}{h} = \frac{f(a+h)-f(a)}{h} - f'(a) \xrightarrow{h \rightarrow 0} 0$. En d'autres termes, $g_a(h) \in o_0(h)$.

On peut donc écrire $f(a+h) = f(a) + f'(a)h + g_a(h) = f(a) + hf'(a) + o_0(h)$. On doit donc qu'au voisinage de a , la fonction f se comporte comme une fonction linéaire avec un reste négligeable devant h . Ou encore $f(a+h) \simeq_0 f(a) + f'(a)h$. On peut donc, dans beaucoup d'applications, remplacer f par une fonction linéaire lorsqu'on sait que les valeurs qu'elle prendra seront dans un intervalle très petit.

La fonction dérivée peut être elle-même dérivable. On note alors la dérivée de la dérivée f'' et $\mathcal{C}_2(I, \mathbb{R})$ l'ensemble des fonctions dérivables deux fois sur I . Plus généralement, une fonction peut être dérivable n fois, on note alors $f^{(n)}$ la n -ième dérivée. On note $\mathcal{C}_n(I, \mathbb{R})$ les fonctions dérivables n fois sur I et $\mathcal{C}_\infty(I, \mathbb{R})$ les fonctions dérivables n fois sur I pour tout $n \in \mathbb{N}$.

Lorsque f est une fonction de $\mathcal{C}_n(I, \mathbb{R})$ et $a \in I$, on peut "approximer" localement f comme un polynôme défini par :

$$f(a+h) = f(a) + f'(a)h + \frac{f''(a)}{2!}h^2 + \dots + \frac{f^{(n)}(a)}{n!}h^n + o(h^n).$$

On appelle cette forme le développement de Taylor ou le développement limité de f . Cela est très pratique pour comprendre facilement le comportement de certaines fonctions dans des cas limites.

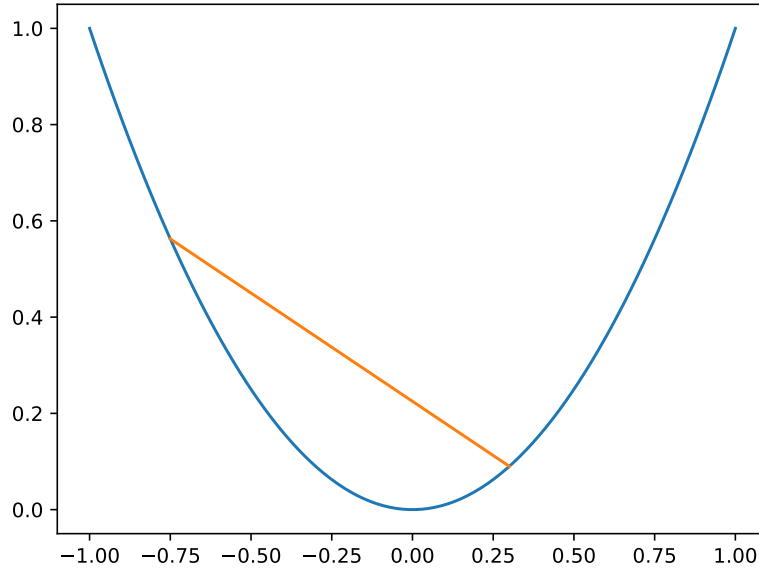
Exemple 1.4. • *Au voisinage de 0, on a $e^x = 1 + x + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + o(x^n)$.*
• *Pour f un polynôme de degré n , le développement limité de f en a est f .*

1.5 Convexité

On dit qu'une fonction de I dans $\mathbb{R}$ est *convexe* si et seulement si on a, pour tout $a < b \in I$ et $t \in [0, 1]$, l'inégalité :

$$f((1-t)a + tb) \leq (1-t)f(a) + tf(b).$$

Pour comprendre cette équation, remarquons la chose suivante : $ta + (1-t)b$ varie de a (quand $t = 0$) à b (quand $t = 1$). Cela couvre donc tous les points entre a et b . L'équation $(1-t)f(a) + tf(b)$ couvre de même tous les points entre $f(a)$ et $f(b)$. Si on regarde le point $t \mapsto (1-t)a + tb$, $(1-t)f(a) + tf(b)$, on est en train de tracer la corde entre $f(a)$ et $f(b)$. La fonction est donc convexe si sa courbe est sous la corde.



Exemple 1.5. • $x \mapsto x^2$ est convexe sur $\mathbb{R}$.
 • $x \mapsto e^x$ est convexe sur $\mathbb{R}$.
 • $x \mapsto -\log x$ est convexe sur $\mathbb{R}_+^*$.

Pour les fonctions $\mathcal{C}_2$, on peut caractériser la convexité par le signe de sa dérivée seconde :

Théorème 1.3. Soit $f \in \mathcal{C}_2(I, \mathbb{R})$. Alors f est convexe sur I si et seulement si $f''(x) \geq 0$ pour tout $x \in I$.

1.6 Optimisation des fonctions convexes

Les fonctions convexes sont intéressantes car elles ont la propriété suivante : tout minimum local de $f \in \mathcal{C}_0(I, \mathbb{R})$ est aussi un minimum global. Un minimum local de f est un point x tel. C'est-à-dire qu'autour de x , $f(x)$ est la plus petite valeur de f . Un minimum est globale si pour tout $y \in I$, $f(x) \leq f(y)$.

On peut donc se poser la question de trouver (ou d'approximer) un minimum globale d'une fonction convexe. Plusieurs algorithmes ont été étudiés pour cela dans l'histoire des mathématiques. Lorsque cela est possible, on peut se contenter de calculer un point où la dérivée f' s'annule. Dans le cas des fonctions convexes, cela suffit à établir que c'est un minimum local (on rappelle que ce n'est pas forcément le cas pour une fonction arbitraire) donc global. Cependant, cela n'est pas toujours possible de calculer cette valeur formellement.

On va donc utiliser un algorithme connu sous le nom de *descente de gradient* et prouver sa correction dans le cas des fonctions convexes. L'algorithme s'écrit ainsi :

- Soit $k = 0$ et $x_0 \in I$ arbitraire.
- Tant que $f'(x_0) > \epsilon$:
 - $\alpha = \text{choix d'un pas de gradient}$
 - $x_0 = x_0 - \alpha f'(x_0)$.

Le seul paramètre à choisir est la façon de calculer α . On peut choisir α comme étant une constante, mais on risque d'osciller autour du minimum.

Exemple 1.6. On considère la fonction convexe $f(x) = x^2$ et un pas de gradient constant $\alpha = 1$. Supposons que $x_0 = 0.1$. Alors on a $f'(x_0) = 0.2$. Donc après un pas de gradient, on a $x_0 = 0.1 - 1 \times 0.2 = -0.1$. En faisant à nouveau un pas de gradient, on se retrouve sur $x_0 = 0.1$.

On a dans ce cas deux solutions : choisir un pas α constant mais un ϵ raisonnable pour éviter cette oscillation, soit diminuer α à mesure que l'on progresse. Une façon de faire est chercher localement un α qui garantit que $f'(x_0 - \alpha f'(x_0))$ sera strictement plus petit que $f'(x_0)$. Si on impose cela, on, on progressera à chaque étape mais il se peut que l'on ralentisse tellement que l'on ne converge jamais. Le choix du bon pas de gradient est une question compliquée. En pratique en apprentissage machine, on utilise des méthodes dites adaptatives où le pas est adapté en fonction d'autres propriétés de la fonction, voir par exemple, la méthode ADAM.

Pour ce cours, nous allons nous contenter d'un cas simple où l'on sait que cela marche toujours : le cas des fonctions $\mathcal{C}_2(I, \mathbb{R})$ où on connaît L tel que pour tout $x \in I$, $f''(x) \leq L$.

Dans ce cas, on choisit un pas de gradient constant de $\alpha = \frac{1}{L}$. En effet, si on pose x^* le minimum global de f , on veut montrer qu'à chaque boucle de l'algorithme, $|x - x^*|$ diminue. Autrement dit, si on pose x_k la valeur de x à la k ème itération, $|x_{k+1} - x^*| < |x_k - x^*|$. Comme $x_{k+1} = x_k - \alpha \cdot f'(x_k)$, on veut donc montrer: $|x_k - \alpha \cdot f'(x_k) - x^*| < |x_k - x^*|$. On va montrer plus généralement que pour tout y , on a l'inégalité $|y - \alpha \cdot f'(y) - x^*| < |y - x^*|$.

On admettra ici le théorème suivant :

Théorème 1.4. Si f est dérivable sur $]a, b[$, alors il existe $c \in]a, b[$ tel que $\frac{f(b) - f(a)}{b - a} = f'(c)$.

Appliqué à f' dans notre cas, on sait qu'il existe $c \in [y, x^*]$ si $y < x^*$ ou $c \in [x^*, y]$ tel que $f'(y) - f'(x^*) = f''(c)(y - x^*)$. Comme $f'(x^*) = 0$ (c'est un minimum local), on a $f'(y) = f''(c)(y - x^*)$. Donc

$$\begin{aligned} y - \alpha f'(y) - x^* &= y - x^* - \alpha f''(c)(y - x^*) \\ &= (y - x^*)(1 - \alpha f''(c)) \end{aligned}$$

Comme $\alpha < \frac{1}{L}$, on a $0 \leq 1 - \alpha f''(c) < 1$ et le résultat est ainsi obtenu.

On a même que si x_k est la valeur de x à la k ème itération, alors $|x_k - x^*| \leq c^k |x_0 - x^*|$ pour une constant $c < 1$.

Exemple 1.7. Pour $f(x) = ax^2$, on a $f''(x) = 2a$. On choisit $\alpha < \frac{1}{2a}$. On note $2a\alpha = \eta < 1$. Une itération de l'algorithme s'écrit donc $x := x - 2ax \times \alpha = x(1 - 2a\alpha) = x(1 - \eta)$. Donc si x_k est la valeur à la k ème itération, on a $x_k = (1 - \eta)^k x_0$ qui tend vers 0 quand k grandit. On converge donc bien.

Dans le cas non-convexe, on ne peut a priori rien dire. Cela dit, on observe que l'algorithme fournit de bon résultat en pratique, c'est-à-dire qu'il semble converger vers des minimums locaux, en tout cas dans peut être localement conve