Fonctions convexes

Prendera due piccioni con una fava :
L'objectif cet épisode est double [1]. D'une part, on cherchera à se familiariser avec la notion de fonction (réelle) convexe, et à en découvrir tout un tas de propriétés hautement sympathiques. C'est pourquoi cet épisode pourra sembler un peu plus décousu que les précédents, et se présente plus sous la forme d'un listing de propriétés que d'une randonnée logiquement articulée autour d'un beau fil rouge. D'autre part, on constatera tout au long de ce qui suit que parfois (mais en fait, c'est très souvent vrai (pour ne pas dire tout le temps)), tout est dans le dessin : intuition, raisonnement, preuve. Les meilleures armes sont souvent des crayons ou des craies de couleurs...

1 Définition

Définition 1. Une fonction $f : \mathbb{R} \to \mathbb{R}$ est dite *convexe* sur $[a, b]$, si la corde prise entre a et b est au-dessus du graphe de f sur tout l'intervalle $[a, b]$

Puisque l'on va s'appuyer essentiellement sur le dessin, autant commencer tout de suite, une fonction convexe, c'est donc un truc de cette tête là :

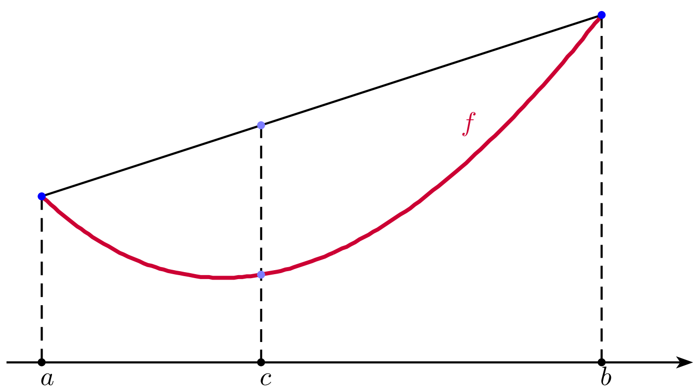

FIGURE 1 – Définition visuelle

On peut retrouver une définition formelle et calculatoire à partir de cela. Quelque soit $c \in [a, b]$, il existe $t \in [0, 1]$ tel que $c = ta + (1 - t)b$ (c est vu comme un barycentre de a et b). L'équation de la corde prise entre a et b est :

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

Soit, en c :

$$y = \frac{f(b) - f(a)}{b - a}(ta + (1 - t)b - a) + f(a) = \frac{f(b) - f(a)}{b - a}(1 - t)(b - a) + f(a) = tf(a) + (1 - t)f(b)$$

On en déduit la propriété suivante [2] :

Propriété 2. *f est convexe si et seulement si $\forall t \in [0, 1], f(ta + (1 - t)b) \leq tf(a) + (1 - t)f(b)$, autrement dit, ssi l'image du barycentre est plus petite que le barycentre des images.*

1. *Littéralement, la superbe phrase en italien juste au-dessus signifie "Prendre deux oiseaux avec une graine". C'était juste histoire de la caser.*
2. *propriété qui, à vrai dire, est la définition usuelle de la convexité.*

Exemple La fonction $x \mapsto x^2$ est convexe sur $\mathbb{R}$. On peut le prouver par le calcul (vous en profiterez au passage pour constater comme la chose est pénible et peu trépidante). Soit $a, b \in \mathbb{R}$ et $t \in [0,1]$, on a :

$$\begin{gathered}(ta+(1-t)b)^2 \leq ta^2+(1-t)b^2 \\ \iff t^2a^2+2t(1-t)ab+(1-t)^2b^2 \leq ta^2+(1-t)b^2 \\ \iff (t^2-t)a^2+2t(1-t)ab+((1-t)^2-(1-t))b^2 \leq 0 \\ \iff t(t-1)a^2-2t(t-1)ab+t(t-1)b^2 \leq 0 \\ \iff t(t-1)(a-b)^2 \leq 0\end{gathered}$$

Si $t=0$ ou $t=1$, c'est clair, et si $t \in]0,1[$, on a $t(t-1)<0$, donc c'est encore vrai.

Une question usuelle en maths, lorsque l'on s'intéresse à des objets qui correspondent à un sous-ensemble particulier d'un plus gros ensemble (ici l'ensemble des fonctions réelles), est de savoir par quel(s) opérateur(s) ce sous-ensemble est stable. Dans notre cas précis, on peut regarder pour l'addition et la composition, qui sont les opérations naturelles sur les fonctions.

Propriété 3. *Si f et g sont deux fonctions convexes, alors $f+g$ est une fonction convexe*

Démonstration. Il suffit de s'appuyer sur la définition calculatoire, et de sommer les deux inégalités... □

Propriété 4. *Si f et g sont deux fonctions convexes sur $[a,b]$, alors $f \circ g : x \mapsto f(g(x))$ n'est pas nécessairement convexe. Une condition nécessaire est que f soit croissante.*

Démonstration. Un bon contre-exemple est $f : x \mapsto -x$ et $g : x \mapsto x^2$ qui sont facilement convexes, alors que $f \circ g : x \mapsto -x^2$ ne l'est clairement pas.
Supposons f croissante, soit $t \in [0,1]$. On a $g(ta+(1-t)b) \leq tg(a)+(1-t)g(b)$ par convexité de g, et $f(tg(a)+(1-t)g(b)) \leq tf(g(a))+(1-t)f(g(b))$ par convexité de f. De plus, comme f est croissante, on a $f(g(ta+(1-t)b)) \leq f(tg(a)+(1-t)g(b))$. D'où, au final, $f(g(ta+(1-t)b)) \leq tf(g(a))+(1-t)f(g(b))$, soit $f \circ g$ convexe. □

C'était ici la dernière fois que l'on montrait quelque chose par le calcul. Pour être rigoureux, il faudrait le faire à chaque fois, mais on va considérer désormais que si l'on voit bien l'idée sur le dessin, la seule difficulté restante est de ne pas s'embrouiller en les différents points, les différents t, mais que la surmonter est plus une question de technicité que d'intelligence à proprement parler.

2 Premières propriétés

Propriété 5. *Soit $f : I \to \mathbb{R}$ convexe et $\alpha \in I$. Alors* $g_\alpha : \left\{\begin{array}{l} I \backslash \{\alpha\} \to \mathbb{R} \\ x \mapsto \frac{f(x)-f(\alpha)}{x-\alpha}\end{array}\right.$ *est croissante.*

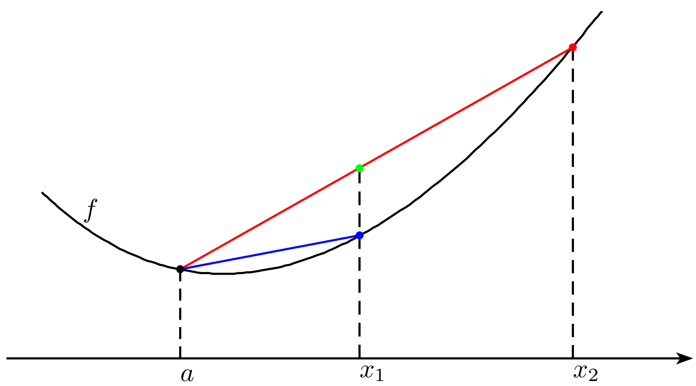

FIGURE 2 – Croissance des pentes

Démonstration. La preuve est à lire sur le dessin. On prend $x_1 < x_2$, on trace les cordes correspondantes. Montrer que $g_\alpha(x_1) < g_\alpha(x_2)$ revient à montrer que la pente rouge est plus forte que la bleue. Or on peut exprimer x_1 comme un barycentre de a et x_2. On déduit de la définition de la convexité que le point • (en tant que barycentre des images de a et x_2) est au-dessus du point • (l'image du barycentre). La pente bleue est donc nécessairement plus faible que la rouge. □

Si vous n'êtes pas encore convaincu par l'intérêt et la rigueur du dessin, essayez de faire la preuve en calculant le bon t, puis les images des différents points, trompez-vous, pleurez un bon coup, et vous verrez la suprématie du crayon de couleur vous apparaître de façon lumineuse...

On peut, grâce à la propriété précédente, en démontrer une un petit peu plus forte :

Propriété 6. *Soient $a < b, x < y$, alors $\frac{f(x)-f(a)}{x-a} \leq \frac{f(y)-f(b)}{y-b}$.*

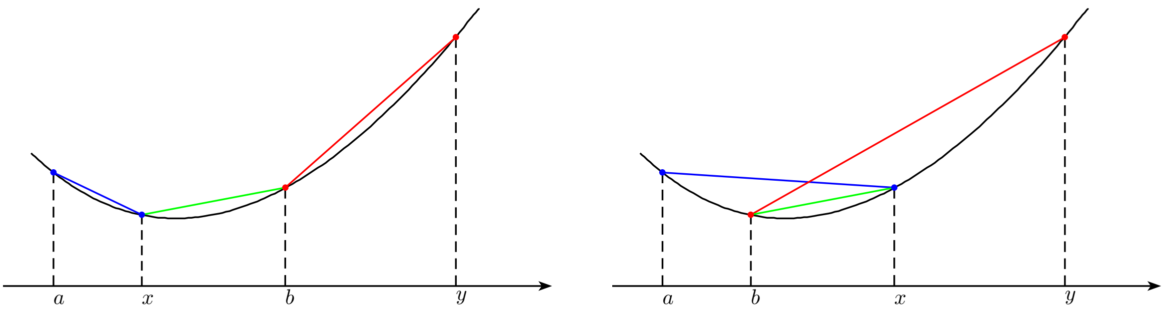

FIGURE 3 – Croissance des pentes, par paires

Démonstration. L'idée est très simple : ici, notre propriété porte sur des paires de points. Or la propriété 5 ne peut nous apporter des informations que pour deux pentes ayant un point commun. On va donc prendre un point commun à nos deux pentes et appliquer deux fois la propriété 5. On rajoute dans les deux cas la corde de b à x. D'après la propriété 5, comme $a < b$, on a $g_x(a) \leq g_x(b)$, et $x < y$ nous donne $g_b(x) \leq g_b(y)$. Or $g_x(b) = \frac{f(b)-f(x)}{b-x} = \frac{f(x)-f(b)}{x-b} = g_b(x)$. D'où au final $g_x(a) \leq g_x(b) \leq g_b(y)$, et en particulier $\frac{f(x)-f(a)}{x-a} \leq \frac{f(y)-f(b)}{y-b}$ □

Toujours à l'aide de la propriété 5 (ce qui montre bien son importance), on va en montrer une assez complétement évidente, mais pénible à prouver en partant juste de la définition de base.

Propriété 7. *Si f est convexe, le graphe de f est au-dessus de toute droite sécante à l'extérieure des intersections. Formellement, si D est une droite (notons h la fonction affine correspondante) qui coupe C_f en a et b, quelque soit $x \notin]a,b[$, $h(x) \leq f(x)$.*

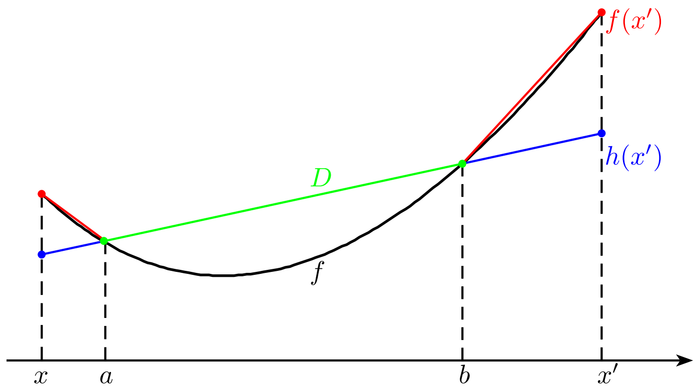

FIGURE 4 – Sécante au graphe de f

Démonstration. On commence par faire un beau dessin plein de couleur. Jusqu'à présent, nous ne disposons que de propriété portant sur l'intérieur des intersection avec des sécantes/cordes. L'idée est donc de s'appuyer sur la valeur des pentes pour en déduire celle des image des points. En effet, il est clair que si la pente $g_b(x')$ (resp. $g_a(x)$) est supérieure (resp. inférieure) à celle de la droite D, alors on aura $h(x') \leq f(x')$ (resp. $h(x) \leq f(x)$), puisque les deux courbes se coupent au point d'abscisse b (resp. a). Or la pente de la droite D est la même tout au au long de la droite, et vaut a fortiori $g_a(b)$. Comme $a < x'$, on a d'après la propriété 5 $\frac{h(x')-h(b)}{x'-b} = g_b(a) \leq g_b(x')$, donc $h(x') \leq f(x')$, c.q.f.d. □

3 Régularité des fonctions convexes

Nous avons désormais vu suffisament de propriété pour se débrouiller dans presque toute situation mélant une fonction convexe et des cordes. Nous allons maintenant nous intéresser au lien entre convexité, continuité et dérivation. On dispose d'un premier théorème assez fort [3], qui nous dit que la convexité implique la continuité. En revanche, une fonction convexe n'est pas nécessairement dérivable, mais si elle l'est, on peut en déduire certaines propriétés.

Théorème 8. *Soit f convexe sur $[a,b]$. Alors f est continue sur $]a,b[$.*

Il est à noter que la continuité est bien sur l'intervalle ouvert, il peut se passer des choses bizarres au bord sinon :

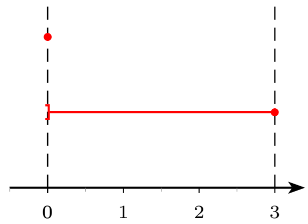

Figure 5 – Fonction convexe non continue au bord

Démonstration. On considère la définition suivante [4] de la continuité : f est continue en a ssi $f(x) \underset{x\to a}{\longrightarrow} f(a)$. L'idée va être de se servir de la convexité pour coincer la fonction au voisinage de a. Partant de là, vous devriez assez naturellement penser à votre théorème des gendarmes préféré, qui sert exactement à ça :

Théorème des gendarmes. Soient $f,g,h \in \mathcal{F}(\mathbb{R},\mathbb{R})$ [5] et $l \in \mathbb{R}$, telles que $\forall x, g(x) \leq f(x) \leq h(x)$, $g(x) \underset{x\to a}{\longrightarrow} l$ et $h(x) \underset{x\to a}{\longrightarrow} l$. Alors $f(x) \underset{x\to a}{\longrightarrow} l$

Il ne reste plus qu'à construire les fonctions g et h qui vont bien. On va se servir pour ça des seuls objets en relation avec f sur lesquels on sache potentiellement des choses, à savoir des sécantes. Assez naturellement on va essayer de les faire passer par $f(a)$, il ne reste plus qu'à construire une fonction en dessous et une au-dessus.

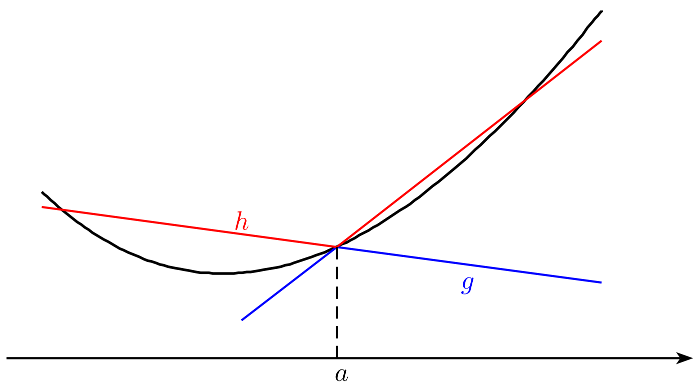

Figure 6 – Théorème des gendarmes appliqué à f

On construit donc g et h comme sur la figure ci-dessus [6] qui vérifient bien toutes les conditions du théorème des gendarmes. On en déduit donc que $f(x) \underset{x\to a}{\longrightarrow} g(a) = h(a) = f(a)$, et donc que f est continue en a. □

Propriété 9. *Supposons f dérivable sur $[a,b]$. Alors f convexe $\Leftrightarrow$ f' croissante.*

Démonstration. On va commencer par le sens calculatoire [7] de la preuve. Supposons donc f' croissante, et montrons que f est convexe. Soient donc x,y. Montrons que $\Phi : t \mapsto tf(x) + (1-t)f(y) - f(tx + (1-t)y)$ est

3. *En effet, la continuité est une propriété très recherchée en générale, qui permet de s'assurer un cadre de travail agréable.*
4. *il en existe d'autres caractérisation avec des suites ou des ϵ, que vous connaissez peut-être*
5. *Notation désignant l'ensemble des fonctions de $\mathbb{R}$ dans $\mathbb{R}$*
6. *Et c'est là qu'est caché le fait que l'on est sur l'intérieur de l'intervalle. En effet, sinon, on ne peut pas forcément construire une corde de chaque côté...*
7. *Non, je n'ai pas dit pénible !*

positive sur $[0,1]$. On constate que $\Phi(0)=\Phi(1)=0$. De plus, $\Phi'(t)=f(x)-f(y)+(y-x)f'((x-y)t+y)$. Or $y-x>0$, donc on a $\Phi'(t)=k+af'(y-at)$, avec $a>0$. Cette fonction est décroissante, puisque f' est croissante. Si $\Phi'(1)>0$, alors quelque soit $t\in[0,1]$, $\Phi'(t)>0$ et $\Phi(1)=\Phi(0)+\int_0^1\Phi'(t)\,\mathrm{d}t>\Phi(0)=\Phi(1)$, ce qui est absurde. D'où $\Phi'(1)\leq 0$. Par un raisonnement analogue, on montre que $\Phi'(0)\geq 0$. De là, on déduit facilement [8] que $\forall t\in[0,1], \Phi(t)\geq 0$ (sinon, en raisonnant sur les valeurs de la dérivée notamment, on trouve rapidement une contradiction). Donc f est convexe.

Réciproquement, supposons f croissant et montrons que f' est croissante. Il suffit pour cela de revenir à la définition du nombre dérivé : $f'(a)=\lim\limits_{x\to a}\frac{f(x)-f(a)}{x-a}=\lim\limits_{x\to a}g_a(x)$. f étant dérivable, cette limite est parfaitement définie et est la même à droite et à gauche, i.e. $f'(a)=\lim\limits_{x\to a,x>a}g_a(x)=\lim\limits_{x\to a,x<a}g_a(x)$ Il ne reste plus qu'à faire un dessin, pour changer.

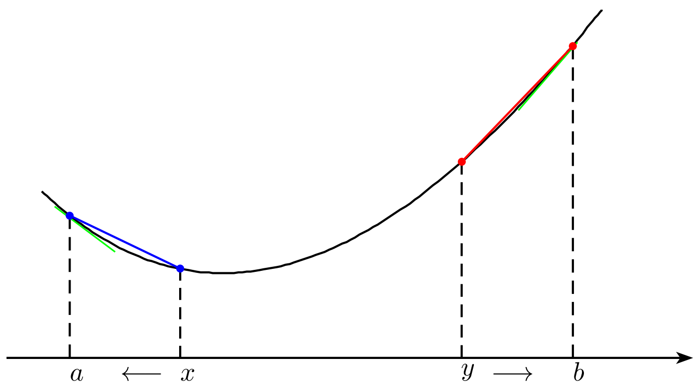

FIGURE 7 – Dérivées d'une fonction convexe

On a donc, par croissance de g_a et g_b que $\forall x>a$, $f'(a)=\lim\limits_{x\to a,x>a}g_a(x)\leq g_a(x)$ et $\forall y<b$, $f'(b)\geq g_b(y)$. Or, comme $a<x<y<b$, on a d'après la propriété 6 que $g_a(x)\leq g_b(y)$, d'où, au final, $f'(a)\leq f'(b)$. $\square$

Corollaire 10. *Si f est deux fois dérivable ($f\in\mathcal{C}^2$), alors (f convexe $\Leftrightarrow f''\geq 0$).*

En s'appuyant sur les idées de la preuve ci-dessus, on obtient facilement la propriété suivante des fonctions convexes :

Propriété 11. *Si f est convexe et dérivable, alors f est au-dessus de ses tangentes*

Démonstration. Il suffit d'exprimer la tangente en a comme étant la limite des cordes de a à x ($x>a$) lorsque $x\to a$. D'après la propriété 5, les cordes sont au-dessus de la tangente. Et d'après la propriété 7, le graphe de f est au-dessus de la sécante après x. Donc, en faisant tendre $x\to a$, on obtient le résultat voulu. Si vous n'êtes pas convaincu, faites le dessin ! $\square$

On termine avec une dernière propriété calculatoire, qui généralise la définition barycentrique à un barycentre de plusieurs points, et qui permet, comme vous allez le voir juste après sur l'exemple, de prouver de jolies inégalités assez inattendues (souvent issues de problèmes géométriques).

Propriété 12 (*Inégalité de Jensen*)**.** *Soient $\lambda_1,\ldots,\lambda_n$ des coefficients tels que $\sum\limits_{i=1}^n\lambda_i=1$. Alors*

$$f(\sum_{i=1}^n\lambda_i x_i)\leq\sum_{i=1}^n\lambda_i f(x_i)$$

Démonstration. La preuve se fait sans surprise par récurrence, mais n'est pas spécialement marrante. C'est de la technique uniquement, ou presque. On s'en passera ici. $\square$

Exemple Soient $a,b,c>0$. Alors

$$\frac{3}{2}\leq\frac{a}{b+c}+\frac{b}{a+c}+\frac{c}{a+b}$$

8. *Ce n'est pas vraiment dur, il suffit de faire de beaux dessins avec les dérivées, mais ce n'est pas trop l'objet du jour.*

En premier lieu, on constate que l'expression de droite est homogène, c'est-à-dire que si l'on multiplie a, b et c par $\alpha \neq 0$, la valeur ne change pas. Quitte à diviser toutes les valeurs par $a+b+c$, on peut donc supposer que $a+b+c=1$, et donc réécrire ainsi le problème :

$$\frac{3}{2} \leq \frac{a}{1-a} + \frac{b}{1-b} + \frac{c}{1-c}$$

Soit, en posant $f : x \neq 1 \mapsto \frac{x}{1-x}$ et en passant le trois de l'autre côté,

$$\frac{1}{2} \leq \frac{f(a)+f(b)+f(c)}{3}$$

Là, c'est le moment de penser à la convexité et à l'inégalité de Jensen. En effet, supposons que f soit convexe, en prenant tous les coefficients à $\frac{1}{3}$, on obtient le résultat espéré :

$$\frac{f(a)+f(b)+f(c)}{3} \geq f(\frac{a+b+c}{3}) = f(\frac{1}{3}) = \frac{1}{2}$$

Ne reste plus qu'à montrer la convexité de f. Ce qui est facile grâce au corollaire sur la dérivée seconde, puisque $f'(x) = \frac{(1-x)-(-x))}{(1-x)^2} = \frac{1}{(1-x)^2}$ et $f''(x) = \frac{2}{(1-x)^3} \geq 0$ sur $]0,1[$.

4 Bonus track

Un premier petit complément nécessaire est de dire que la notion de convexité se généralise fort bien au fonction à plusieurs variables, il suffit de tout faire avec des gradients à la place de la dérivée, et on s'en sort. Et qu'accessoirement, on dit qu'une forme est convexe si dès lors celle-ci contient deux points, elle contient le segment les reliants. Encore une histoire de barycentre.

Ma foi, c'est bien joli tout ça, me direz-vous, mais cela sert-il vraiment à quelque chose ? En fait, oui. D'une part, en analyse, la propriété de convexité est relativement recherchée, puisque comme vous venez de le voir, elle offre tout un bon nombre de propriétés agréables sur la fonction. Et d'autre part, dans un grand nombre de problèmes d'optimisation (ce qui est assez courant dans la vraie vie), on cherche systématiquement à minimiser des quantités correspondant à des fonctions convexes. En effet, dès lors, des méthodes assez similaires à la méthode de Newton pour la résolution des équations $f(x) = 0$ fonctionnent bien et permettent des résolutions simples. Si ça vous amuse, prenez une fonction convexe, essayer de voir le lien entre minimum local et minimum global, et essayer de trouver une méthode systématique de descente vers le minimum. Au besoin, je peux aider !

Un petit dessin vaut mieux qu'un long discours
[Napoléon Bonaparte]