Fonctions convexes

Propriété 0.1

Pour deux réels $x \leqslant y$ et pour $\lambda \in [0, 1]$,

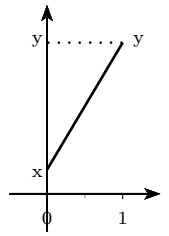
$$(1-\lambda)x + \lambda y \in [x, y].$$

De plus, si on pose $f : \lambda \mapsto (1-\lambda)x + \lambda y$, alors f réalise une bijection continue et strictement croissante de $[0, 1]$ sur $[x, y]$ et de $]0, 1[$ sur $]x, y[$.

Explications

On peut donner beaucoup d'interprétations différentes du barycentre.

1. **Difféomorphisme** (ou paramétrage normal) : on parcourt $[x, y]$ à vitesse constante pendant un temps unité.
   On considère un segment $[x, y]$, avec $x \leqslant y$ et $\lambda \in [0, 1]$. On veut aller de x à y à vitesse constante en un temps $T = 1$..
   À l'instant 0 on est en x et à l'instant 1 on est en y.
   On construit donc une fonction affine :

   L'équation de la droite s'obtient simplement par

   $$u(\lambda) = \frac{y-x}{1-0} \times \lambda + x = (y-x)\lambda + x = (1-\lambda)x + \lambda y.$$

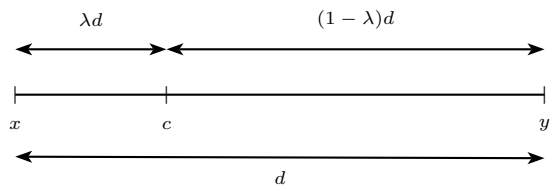
   Ainsi, $u(\lambda)$ décrit le segment $[x, y]$ lorsque λ varie de 0 à 1.

2. **Mécanique :** point d'équilibre avec des poids positifs aux extrémités.
   D'un point de vue physique, On considère deux masses M_1 et M_2 placées respectivement en x et en y sur une barre (de poids nul).
   L'unité de mesure pour les masses n'importe pas pour déterminer le point d'équilibre (que l'on compte en kg, en gramme ou autre, ne change pas la position d'équilibre). On peut donc choisir comme unité de masse[1] $M_1 + M_2$.
   Cela revient à affecter y de la masse $\lambda = \frac{M_2}{M_1+M_2} \in [0, 1]$ et x de la masse $\frac{M_1}{M_1+M_2} = 1 - \lambda \in [0, 1]$.
   Le point d'équilibre correspond à l'annulation des moments (chaque poids tend à faire tourner la barre vers lui[2]). :

   $$(1-\lambda)(c-x) = \lambda(y-c).$$

1. On suppose que les masse ne sont pas toutes les deux nulles.
2. On a simplifié par la constante de gravité de part et d'autre.

On trouve alors $c = (1-\lambda)x + \lambda y$.
Le point c correspond donc au barycentre (centre de gravité) des points x et y affectés respectivement des masses $(1-\lambda)$ et λ.



⋆ Ainsi, le point c sera d'autant plus proche de x que la masse affectée à x est importante par rapport à celle affectée à y, c'est-à-dire pour λ proche de 0.
⋆ Quand λ s'approche de 1, la masse sur y devient largement prépondérante et c est proche de y.
⋆ Pour $\lambda = \frac{1}{2}$, on observe que x et y sont affectés des mêmes masses et que le barycentre est exactement au milieu : $x = \frac{x+y}{2}$.
Dans le cas général, si on note $d = y - x$ la distance entre x et y, alors $c = (1-\lambda)x + \lambda y = x + \lambda(y-x) = x + \lambda d$.
On voit bien alors que λ désigne la proportion du segment entre x et c.

3. **Moyenne arithmétique**
La vision mécanique peut être interprétée en terme de moyenne arithmétique. Les masses correspondent alors aux coefficients affectés aux notes (par exemple) x et y.
La moyenne est alors $c = (1-\lambda)x + \lambda y$ où $\lambda = \frac{\text{coeff } y}{\text{coeff } x + \text{coeff } y}$.

On remarque également que l'on peut obtenir l'intervalle ouvert en prenant la restriction à $]-1, 1[$.
Ce *paramétrage* est une bijection de classe $\mathscr{C}^\infty$ de $[0, 1]$ sur $[x, y]$.
Il permet de décrire facilement n'importe quel segment et il est immédiatement généralisable en dimensions supérieures avec

$$\overrightarrow{OC} = (1-\lambda)\overrightarrow{OA} + \lambda\overrightarrow{OB}.$$

Définition 0.2 *(Fonction convexe)*

Soit I un intervalle de $\mathbf{R}$ et $f : I \to \mathbf{R}$.
f est **convexe** sur I si, et seulement si

$$\forall (x, y) \in I^2, \forall \lambda \in [0, 1],\ f(\lambda x + (1-\lambda)y) \leqslant \lambda f(x) + (1-\lambda) f(y).$$

⚠ La notion de convexité n'a de sens que sur un **intervalle** (sinon, il n'y a aucune raison que le point dont on recherche l'image soit dans I).
On rappelle que $+(1-\lambda)y \in [x, y]$.

Explications

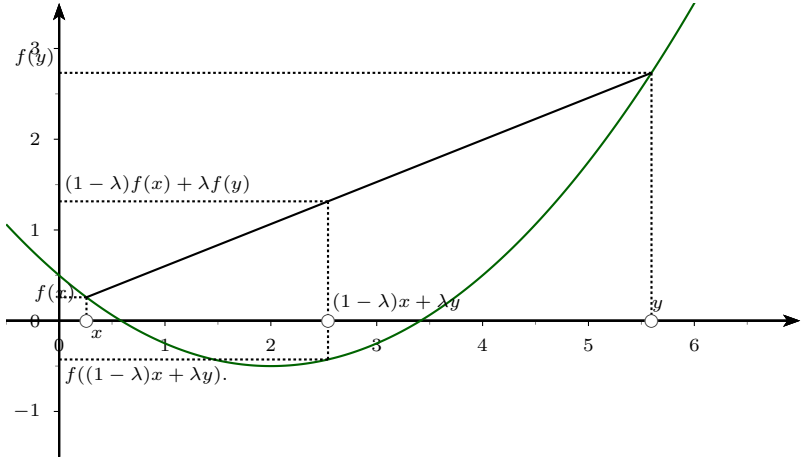
Géométriquement, une fonction est convexe, lorsqu'elle est sous ses cordes.
La corde à la courbe entre x et y est définie par l'ensembles des barycentres des points du plan $(x, f(x))$ et $(y, f(y))$ qui s'écrivent en coordonnées barycentriques sous la forme

$$(\,(1-\lambda)x + \lambda y\ ,\ (1-\lambda)f(x) + \lambda f(y)\,).$$

Le fait que la courbe de la fonctions soit sous cette corde se traduit alors simplement

$$f\,(\,(1-\lambda)x + \lambda y\,) \leqslant (1-\lambda)f(x) + \lambda f(y).$$

C'est bien la définition d'une fonction convexe.



On peut interpréter f en terme de moyennes pondérées : Si f est convexe, alors l'image de la moyenne est inférieure à la moyenne des images.

Définition 0.3

Une fonction f définie sur un intervalle I est **concave** sur I, si $-f$ est convexe sur I.

Exemple

$x \mapsto x^2$ est convexe sur $\mathbf{R}$.
En effet, pour $(x, y) \in \mathbf{R}^2$ et $\lambda \in [0, 1]$,

$$((1-\lambda)x + \lambda y)^2 = (1-\lambda)^2 x^2 + \lambda^2 y^2 + 2(1-\lambda)\lambda xy.$$

Or $2|xy| \leqslant x^2 + y^2$, donc

$$((1-\lambda)x + \lambda y)^2 \leqslant \left((1-\lambda)^2 + (1-\lambda)\lambda\right) x^2 + \left(\lambda^2 + (1-\lambda)\lambda\right) y^2 = (1-\lambda)\, x^2 + \lambda y^2.$$

Donc la fonction est bien convexe.

Preuve

On note $m = \min\limits_{1\leqslant k\leqslant n} x_k$ et $M = \max\limits_{1\leqslant k\leqslant n} x_k$ de sorte que $\forall k \in [\![1, n]\!]$, $m \leqslant x_k \leqslant M$.
Or $\lambda_k \geqslant 0$, donc $\lambda_k m \leqslant \lambda_k x_k \leqslant \lambda_k M$.
Donc par somme d'inégalités

$$m = \sum_{k=1}^{n} \lambda_k m \leqslant \sum_{k=1}^{n} \lambda_k x_k \leqslant \sum_{k=1}^{n} \lambda_k M = M.$$

Or $m \in I$ et $M \in I$ qui est un intervalle, donc $\sum\limits_{k=1}^{n} \lambda_k x_k \in I$. ■

Théorème 0.4 *(Inégalité de Jensen)*

Soit f une fonction convexe sur un intervalle I et $n \geqslant 1$ un entier.
Pour tous points $(x_1, x_2, \cdots, x_n)$ de I
et pour tous réels positifs $(\lambda_1, \lambda_2, \cdots, \lambda_n)$ tels que $\sum\limits_{k=1}^{n} \lambda_k = 1$, on a

$$\sum_{k=1}^{n} \lambda_k x_k \in I \qquad \text{et} \qquad f\left(\sum_{k=1}^{n} \lambda_k x_k\right) \leqslant \sum_{k=1}^{n} \lambda_k f(x_k).$$

Pour $n = 2$, on retrouve la définition de la convexité en prenant $\lambda = \lambda_2$ et donc $\lambda_1 = 1 - \lambda$.

Preuve

On prouve l'inégalité par récurrence sur $n \geqslant 1$.
Initialisation : pour $n = 1$, le résultat est trivial.
Hérédité : On suppose le résultat vrai à un rang $n \geqslant 1$ et on veut le montrer au rang $n + 1$.
On considère donc $(x_1, \cdots, x_{n+1})$ des points de I et $(\lambda_1, \cdots, \lambda_{n+1})$ des réels positifs de somme 1.
Si pour tout $i \in [\![1, n]\!]$, $\lambda_i = 0$, alors $\lambda_{n+1} = 1$ et on se ramène au cas de l'initialisation.
On suppose donc à présent qu'il existe $i \in [\![1, n]\!]$ tel que $\lambda_i \neq 0$.
Ainsi, par positivité des λ_i, on sait que $\sum\limits_{k=1}^{n} \lambda_k \neq 0$ ce qui permet de poser :

$$x = \frac{1}{\sum\limits_{k=1}^{n} \lambda_k} \sum_{k=1}^{n} \lambda_k x_k, \quad \text{et} \quad y = x_{n+1}.$$

On note également $\mu = \lambda_{n+1}$ de sorte que $1 - \mu = \sum\limits_{k=1}^{n} \lambda_k$.
Ainsi on peut écrire

$$\sum_{k=1}^{n+1} \lambda_k x_k = (1 - \mu)x + \mu y.$$

On remarque que x est lui-même combinaison convexe de $(x_1, \cdots, x_n)$ affecté des poids $\left(\frac{\lambda_1}{\sum_{i=1}^{n} \lambda_i}, \cdots, \frac{\lambda_n}{\sum_{i=1}^{n} \lambda_i}\right)$ donc $x \in I$ par hypothèse de récurrence.
Or le barycentre $(1 - \mu)x + \mu x_{n+1} \in I$ car $\mu \geqslant 0$.

D'après la convexité de f sur I appliquée avec μ, on trouve donc

$$f\left(\sum_{k=1}^{n+1} \lambda_k x_k\right) = f\left((1 - \mu)x + \mu y\right) \leqslant (1-\mu)f(x)+\mu f(y) = \left(\sum_{k=1}^{n} \lambda_k\right) f(x)+\lambda_{n+1} f(x_{n+1}).$$

Or, en appliquant l'hypothèse de récurrence aux points $(x_1, \cdots, x_n)$ affectés des coefficients $\left(\frac{1}{\sum\limits_{k=1}^{n} \lambda_k}\lambda_1, \cdots \frac{1}{\sum\limits_{k=1}^{n} \lambda_k}\lambda_n\right)$ on trouve alors

$$f(x) \leqslant \frac{1}{\sum\limits_{k=1}^{n} \lambda_k} \sum_{k=1}^{n} \lambda_k f(x_k).$$

Les deux inégalités mises ensemble donnent donc

$$\begin{aligned} f\left(\sum_{k=1}^{n+1} \lambda_k x_k\right) &\leqslant \left(\sum_{k=1}^{n} \lambda_k\right) f(x) + \lambda_{n+1} f(x_{n+1}) \\ &\leqslant \left(\sum_{k=1}^{n} \lambda_k\right) \left(\frac{1}{\sum\limits_{k=1}^{n} \lambda_k} \sum_{k=1}^{n} \lambda_k f(x_k)\right) + \lambda_{n+1} f(x_{n+1}) \\ &\leqslant \sum_{k=1}^{n} \lambda_k f(x) + \lambda_{n+1} f(x_{n+1}) \leqslant \sum_{k=1}^{n+1} \lambda_k f(x_k). \end{aligned}$$

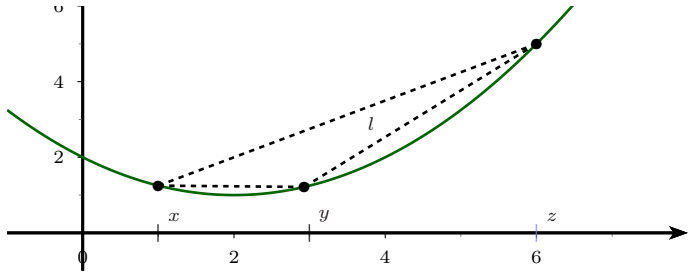
■

Théorème 0.5 *(Croissance des pentes)*

Soit f définie sur un intervalle I.
f est convexe sur I si, et seulement si

$$\forall (x, y, z) \in I^3,\ x < y < z \Rightarrow \frac{f(y) - f(x)}{y - x} \leqslant \frac{f(z) - f(x)}{z - x} \leqslant \frac{f(z) - f(y)}{z - y}.$$

Une seule inégalité des pentes suffit pour avoir la réciproque.



Ce théorème est à apprendre avec son schéma où on observe bien les pentes des différentes cordes à la courbe, que l'on classe aisément par ordre croissant.
Cette propriété est souvent très efficace pour démontrer des résultats sur les fonctions convexes.

Preuve

Sens direct.
Soit f convexe sur I. Comme $y \in]x,\, z[$, il existe $\lambda \in]0,\, 1[$ tel que $y = (1-\lambda)x + \lambda z$.
On remarque alors que $y - x = \lambda(z-x)$. Par convexité de f, on peut écrire

$$f(y) = f\left((1-\lambda)x + \lambda z\right) \leqslant (1-\lambda)f(x) + \lambda f(z).$$

Ainsi, on obtient $f(y) - f(x) \leqslant \lambda(f(z) - f(x))$. Or $z - x \neq 0$, donc $\lambda = \frac{y-x}{z-x}$ ce qui donne $f(y) - f(x) \leqslant (y-x)\frac{f(z)-f(x)}{z-x}$.
De plus $y - x > 0$, donc

$$\frac{f(y)-f(x)}{y-x} \leqslant \frac{f(z)-f(x)}{z-x}.$$

On fait de même avec $1-\lambda$, ce qui donne $f(y) - f(z) \leqslant (1-\lambda)(f(x) - f(z)) = \frac{y-z}{x-z}(f(x) - f(z))$.
En divisant alors par $y - z < 0$, on obtient

$$\frac{f(z)-f(y)}{z-y} = \frac{f(y)-f(z)}{y-z} \geqslant \frac{f(x)-f(z)}{x-z} = \frac{f(z)-f(x)}{z-x}.$$

Ce qui donne la deuxième inégalité.
Réciproque : supposons à présent que la fonction f vérifie une des inégalités des pentes sur I.
On pose alors $x < z$ deux points de I et $\lambda \in [0,\, 1]$. On note $y = (1-\lambda)x + \lambda z \in [x,\, z]$.
Si $\lambda = 0$ ou $\lambda = 1$, une simple vérification montre que $f(y) \leqslant (1-\lambda)f(x) + \lambda f(z)$ et qu'on a même l'égalité.
Supposons à présent que $\lambda \in]0,\, 1[$,

- Si f vérifie la première inégalité, alors $y \in]x,\, z[$ et on peut appliquer l'inégalité des pentes.
  Ainsi
  $$\frac{f(y)-f(x)}{y-x} \leqslant \frac{f(z)-f(x)}{z-x}.$$
  Et par produit avec $y - x > 0$, on retrouve $f(y) \leqslant f(x) + \frac{y-x}{z-x}(f(z) - f(x))$.
  Or $\frac{y-x}{z-x} = \lambda$, donc $f(y) \leqslant (1-\lambda)f(x) + \lambda f(z)$ ce qui prouve la convexité de f.
- De même, si f vérifie la deuxième inégalité, alors
  $$\frac{f(z)-f(x)}{z-x} \leqslant \frac{f(z)-f(y)}{z-y}.$$
  Et par produit avec $z - x > 0$, on trouve $-f(y) \geqslant -f(z) + \frac{z-y}{z-x}(f(z) - f(x))$.
  Puis par produit par $-1 < 0$, on obtient $f(y) \leqslant \left(1 - \frac{z-y}{z-x}\right) + \frac{z-y}{z-x}f(x)$.
  Or $\frac{z-y}{z-x} = 1 - \lambda$, donc $f(y) \leqslant (1-\lambda)f(x) + \lambda f(z)$ ce qui prouve la convexité de f.
- Enfin, si f vérifie l'inégalité entre les deux extrêmes, alors
  $$\frac{f(y)-f(x)}{y-x} \leqslant \frac{f(z)-f(y)}{z-y}.$$
  Et par produit avec $y - x > 0$, on trouve $f(y) + \frac{y-x}{z-y}f(y) \geqslant f(x) + \frac{y-x}{z-y}f(z)$.
  C'est-à-dire
  $$\frac{z-x}{z-y}f(y) \leqslant f(x) + \frac{y-x}{z-y}f(z).$$
  Ce qui donne par produit avec $\frac{z-y}{z-x} > 0$,
  $$f(y) \leqslant \frac{z-y}{z-x}f(x) - \frac{y-x}{z-x}f(z).$$
  Or $\frac{z-y}{z-x} = 1 - \lambda$, donc $f(y) \leqslant (1-\lambda)f(x) + \lambda f(z)$ ce qui prouve la convexité de f.

■

Propriété 0.6

Soit f une fonction convexe sur I. Soient x et y deux points de I avec $x < y$.
La courbe de f sur $[x,\, y]$ est sous la corde $[(x,\, f(x)),\, (y,\, f(y))]$ et au dessus en dehors de $[x,\, y]$.
Autrement dit, la courbe de f est en dessous de ses sécantes entre les points de séquence et au dessus en dehors.

Preuve

Tout point c du segment $[x, y]$ peut s'écrire sous la forme $c = (1-\lambda)x + \lambda y$ avec $\lambda \in [0,\, 1]$.
Par convexité de f, on en déduit donc que $f(c) \leqslant (1-\lambda)f(x) + \lambda f(y)$.
$f(c)$ désigne l'ordonnée au point c sur la courbe de f et $(1-\lambda)f(x) + \lambda f(y)$ désigne l'ordonnée au point c sur la corde.
Ceci prouve bien que la courbe de f est au dessus de sa corde sur $[x,\, y]$.
Pour $z > y$, l'inégalité des pentes donne $\frac{f(z)-f(y)}{z-y} \geqslant \frac{f(y)-f(x)}{y-x}$.
Et par positivité de $z - y$

$$\underbrace{f(z)}_{\text{sur la courbe}} \geqslant \underbrace{f(y) + (z-y)\frac{f(y)-f(x)}{y-x}}_{\text{sur la corde}}.$$

On obtient la même chose pour un point avant x par un raisonnement similaire (exercice).

■

Théorème 0.7

Si f est dérivable sur un intervalle I et à valeurs réelles.

f est convexe sur I si, et seulement si f' est croissante sur I.

Corollaire 0.8

Si f est deux fois dérivable sur un intervalle I et à valeurs réelles.

f est convexe sur I si, et seulement si f'' est positive sur I.

Preuve

Si f est convexe sur I, on note $x < y$ deux points de I. Pour tout point $x + h \in]x,\, y[$, l'inégalité des pentes donne

$$\frac{f(x+h)-f(x)}{h} \leqslant \frac{f(y)-f(x)}{y-x}.$$

Par passage de l'inégalité à la limite quand $h \to 0$, on trouve donc

$$f'(x) \leqslant \frac{f(y)-f(x)}{y-x}.$$

Par ailleurs, pour $h < 0$ tel que $y + h \in]x, y[$, l'inégalité des pentes donne

$$\frac{f(y)-f(x)}{y-x} \leqslant \frac{f(y)-f(y+h)}{-h} = \frac{f(y+h)-f(y)}{h}.$$

À nouveau, par passage à la limite, on trouve

$$f'(x) \leqslant \frac{f(y)-f(x)}{y-x} \leqslant f'(y).$$

Ce qui prouve bien la croissance de f'.
Réciproquement, si f est croissante sur I, alors pour $x < y < z$, d'après l'égalité des accroissements finis, il existe c_1 et c_2 tels que $x < c_1 < y < c_2 < y$ et

$$\frac{f(y)-f(x)}{y-x} = f'(c_1) \leqslant f'(c_2) = \frac{f(z)-f(y)}{z-y}.$$

Ce qui prouve bien que f est convexe.

■

Théorème 0.9

Si f est convexe sur I, $a \in I$ et f dérivable en a.
Alors la courbe de f est au dessus de sa tangente en a.

Preuve

Cela revient à montrer que pour tout $x \in I$,

$$f(x) \geqslant f(a) + f'(a)(x-a).$$

Pour $x > a$, on pose $a + h \in]a, x[$ et par convexité, on obtient

$$\frac{f(a+h)-f(a)}{h} \leqslant \frac{f(x)-f(a)}{x-a}.$$

En passant l'inégalité à la limite pour $h \to 0$, on obtient $f'(a) \leqslant \frac{f(x)-f(a)}{x-a}$ ce qui se transforme en l'inégalité cherchée.
On fait de même pour $x < a$.
(Pour $x = a$, c'est trivial).

■

Propriété 0.10

Si f est dérivable sur I et si sa courbe est au dessus de toutes ses tangentes, alors f est convexe sur I.

Preuve

En exercice.
La courbe étant au dessus de ses tangentes, on remarque que l'inégalité des pentes est immédiate : pour $a < c < b$,

(la pente entre a et c) $\leqslant$ (pente au point x = pente de la tangente) $\leqslant$ (pente entre c et b).

■