Caractérisation de fonctions convexes différentiables.

Le but ici est de généraliser les caractérisations bien connues pour une fonction $f \in \mathcal{F}(\mathbb{R},\mathbb{R})$:

1) Si $f \in \mathcal{C}^1$, alors f est convexe ssi f' est croissante.
2) Si $f \in \mathcal{C}^2$, alors f est convexe ssi f'' est positive.

lemmes-rappels

115.1 Si $f \in \mathcal{F}(\mathbb{R}^n, \mathbb{R}^m)$ est différentiable en $x \in \mathbb{R}^n$ alors $\forall h \in \mathbb{R}^n$, $Df(x)(h) = \lim\limits_{\lambda \to 0} \dfrac{f(x+\lambda h) - f(x)}{\lambda}$

115.2 En particulier si $f \in \mathcal{F}(\mathbb{R}^n, \mathbb{R})$ est différentiable en $x \in \mathbb{R}^n$ alors $\forall h \in \mathbb{R}^n$ $\langle \nabla f(x) \mid h \rangle = \lim\limits_{\lambda \to 0} \dfrac{f(x+\lambda h) - f(x)}{\lambda}$

115.3 Si $g \in \mathcal{F}(\mathbb{R}^n, \mathbb{R}^n)$ est différentiable en $x \in \mathbb{R}^n$ et vérifie $\forall h \in \mathbb{R}^n$ $\langle g(x+h) - g(x) \mid h \rangle \geq 0$ alors on a $\forall h \in \mathbb{R}^n$ $\langle Dg(x)(h) \mid h \rangle \geq 0$. De plus si $g \in \mathcal{C}^1$ on a aussi la réciproque.

Preuve • Par définition de la différentiabilité en x on a

$$\forall h \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}, \quad f(x+\lambda h) = f(x) + Df(x)(\lambda h) + \underset{\lambda\|h\| \to 0}{o}(\lambda \|h\|)$$
$$= f(x) + \lambda\, Df(x)(h) + \underset{\lambda \to 0}{o}(\lambda)$$

Donc $\forall h \in \mathbb{R}^n, \forall \lambda \in \mathbb{R}$ $\quad \dfrac{f(x+\lambda h) - f(x)}{\lambda} = Df(x)(h) + \underset{\lambda \to 0}{o}(1)$

Donc $\forall h \in \mathbb{R}^n$ $\quad \lim\limits_{\lambda \to 0} \dfrac{f(x+\lambda h) - f(x)}{\lambda} = Df(x)(h)$ $\square$

• La définition de $\nabla f(x)$ est justement que $\forall h \in \mathbb{R}^n, \langle \nabla f(x) \mid h \rangle = Df(x)(h)$ donc on réutilise 115.1 $\square$

• Commençons par la réciproque dans le cas $g \in \mathcal{C}^1$. Pour $x \in \mathbb{R}^n$ et $h \in \mathbb{R}^n$
$\langle g(x+h) - g(x) \mid h \rangle = \langle \int_0^1 Dg(x+sh)(h)\, ds \mid h \rangle = \int_0^1 \underbrace{\langle Dg(x+sh)(h) \mid h \rangle}_{\geq 0 \text{ par hypothèse}} ds \geq 0$
D'où 115.3 $\Leftarrow$.

Montrons maintenant le sens direct sous l'hypothèse g diff. en x.

- Par hypothèse $\forall \lambda\in\mathbb{R}\ \forall h\in\mathbb{R}^n\ \langle g(x+\lambda h)-g(x)\ |\ \lambda h\rangle \geq 0$

donc on a $\forall \lambda\in\mathbb{R}^*, \forall h\in\mathbb{R}^n\ \langle \frac{g(x+\lambda h)-g(x)}{\lambda}\ |\ h\rangle \geq 0$

car on a divisé par $\lambda^2\in\mathbb{R}^{+*}$.

En passant à la limite quand $\lambda\to 0$, on a d'après 115.1

$\forall h\in\mathbb{R}^n \langle Dg(x)(h)\ |\ h\rangle \geq 0$. $\square$.

Rq Si $g\in\mathcal{F}(\mathbb{R}^n,\mathbb{R}^n)$ est différentiable et vérifie $\forall x\in\mathbb{R}^n, \forall h\in\mathbb{R}^n \langle g(x+h)-g(x)\ |\ h\rangle\geq 0$ alors en tout point la différentielle de g est une application positive (au sens de la positivité pour les applications linéaires, pour les matrices : "A est positive ssi $\forall X\in\mathbb{R}^n\ {}^tXAX\geq 0$").
On retrouve la généralisation de la propriété bien connue pour les fonctions réelles : "$f\nearrow \Leftrightarrow f'\geq 0$".
Ici l'hypothèse de croissance s'exprime de manière peut-être un peu surprenante mais sur $\mathbb{R}$ la croissance n'est rien d'autre que dire que pour un déplacem^t $h\geq 0$, f varie aussi de manière ≥ 0,
à l'inverse ——— ≤ 0 ——— ≤ 0.
Autrement dit la variation de la fonction est "dans le m̂ sens" que celle de l'argument. Ici on relâche un peu cela : on ne demande pas que h (la variat° de l'argument) et $f(x+h)-f(x)$ (la varia° de la fonction) soient positivement liées mais seulem^t qu'elles aillent "grosso modo" dans la même direction, c-à-d qu'elles soient de produit scalaire positif.

115.4 **Pté** Soit $f\in\mathcal{C}^1(\mathbb{R}^n,\mathbb{R})$.

f est convexe $\overset{(a)}{\Longleftrightarrow} \forall (x,y)\in(\mathbb{R}^n)^2,\ f(y)-f(x)\geq \langle\nabla f(x).\,y-x\rangle$

$\overset{(b)}{\Longleftrightarrow} \forall (x,h)\in(\mathbb{R}^n)^2,\ \langle\nabla f(x+h)-\nabla f(x)\ |\ h\rangle\geq 0$

analogue de ∇f croissant.

Preuve $\boxed{\overset{a}{\Rightarrow}}$ On utilise 115.2, pour $(x,y)\in(\mathbb{R}^n)^2$

$$\nabla f(x)(y-x) = \lim_{\lambda\to 0} \frac{f(x+\lambda(y-x))-f(x)}{\lambda}$$

IIS.1

Or par convexité de f on a, pour $\lambda \in]0,1[$,

$$\frac{f(x+\lambda(y-x))-f(x)}{\lambda} = \frac{f((1-\lambda)x+\lambda y)-f(x)}{\lambda}$$
$$\leq \frac{(\cancel{1}-\lambda)f(x)+\lambda f(y)-\cancel{f(x)}}{\lambda} = f(y)-f(x).$$

Donc nécessairement $\langle \nabla f(x) | (y-x)\rangle \leq f(y)-f(x)$ □.

$\boxed{\overset{a}{\Leftarrow}}$ Soit $(x,y)\in(\mathbb{R}^n)^2$. Soit $\lambda \in]0,1[$. On note $z = \lambda x+(1-\lambda)y$.

$$f(z)\leq \lambda f(x)+(1-\lambda)f(y) \iff \lambda f(z)+(1-\lambda)f(z) \leq \lambda f(x)+(1-\lambda)f(y)$$
$$\iff \lambda[f(z)-f(x)]+(1-\lambda)[f(z)-f(y)]\leq 0$$

Or par hypothèse $f(z)-f(x)\leq \langle \nabla f(z) | \underbrace{z-x}_{=(\lambda-1)x+(1-\lambda)y\,=\,(1-\lambda)(y-x)}\rangle$ (car $\langle\nabla f(z)|x-z\rangle \leq f(x)-f(z)$)

et $f(z)-f(y)\leq \langle \nabla f(z) | \underbrace{z-y}_{=\lambda x-\lambda y\,=\,-\lambda(y-x)}\rangle$ de même

Donc $\lambda[f(z)-f(x)]+(1-\lambda)[f(z)-f(y)] \leq \lambda(1-\lambda)(y-x)+(1-\lambda)(-\lambda)(y-x) = 0$

Ainsi on a bien $f(z)\leq \lambda f(x)+(1-\lambda)f(y)$.

Ceci étant pour tout x,y,λ, on a bien f convexe □.

$\boxed{\overset{b}{\Rightarrow}}$ On suppose ici que $\forall (x,y)\in(\mathbb{R}^n)^2$, $f(y)-f(x) \geq \langle\nabla f(x) | y-x\rangle$

Donc pour $x\in\mathbb{R}^n$ et $h\in\mathbb{R}^n$ $\quad f(x+h)-f(x) \geq \langle \nabla f(x) | h\rangle$

et $\quad f(x)-f(x+h) \geq \langle \nabla f(x+h) | -h\rangle = \langle -\nabla f(x+h) | h\rangle$

Donc en sommant on a $0 \geq \langle \nabla f(x) - \nabla f(x+h) | h\rangle$

soit $\langle \nabla f(x+h) - \nabla f(x) | h \rangle \geq 0$ □

$\boxed{\overset{b}{\Leftarrow}}$ Ici on utilise le caractère $\mathcal{C}^1$ de f ; pour $(x,y)\in(\mathbb{R}^n)^2$

$$f(y)-f(x) = \int_0^1 \langle \nabla f(x+s(y-x)) | y-x\rangle \, ds$$
$$= \int_0^1 \frac{1}{s} \underbrace{\langle \nabla f(x+s(y-x)) | s(y-x)\rangle}_{\geq \langle \nabla f(x) | s(y-x)\rangle} ds$$
$$\geq \int_0^1 \frac{1}{\cancel{s}} \langle \nabla f(x) | \cancel{s}(y-x)\rangle \, ds$$
$$= \langle \nabla f(x) | y-x\rangle$$

□

115.5 **Rappel** Si $f \in \mathcal{F}(\mathbb{R}^n, \mathbb{R})$, alors $\forall x \in \mathbb{R}^n$ $\mathcal{D}f(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}) = \mathbb{R}^{n\prime}$
Or d'après le th. de représentation de Riesz $\mathbb{R}^{n\prime}$ s'identifie à $\mathbb{R}^n$, ce qui revient à identifier $\mathcal{D}f(x)$ et $\nabla f(x)$.
Poursuivant cette identification on entendra par $D^2 f(x)$ plutôt $\mathcal{D}[\nabla f](x)$, et comme $\nabla f \in \mathcal{F}(\mathbb{R}^n, \mathbb{R}^n)$, $D^2 f(x) \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ et est donc identifiable à une matrice carrée de taille n (dans la même base que celle utilisée pour identifier $\nabla f(x)$; c'est à dire dans la base qui définit notre produit scalaire) qu'on appelle matrice hessienne.
Dans ce cas la condition $\forall h \in \mathbb{R}^n, \langle \mathcal{D}^2 f(x)(h) \mid h \rangle \geq 0$ se réécrit $\forall h, {}^t h A h \geq 0$ où A représente $\mathcal{D}^2 f(x)$, ce qui revient à dire que A est une matrice positive.
ce qu'on notera $\underline{A \geq 0}$

115.6 **Th** Si $f \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R})$
f est convexe $\Leftrightarrow \forall x, \ D^2 f(x) \geq 0$.

Preuve.
$$f \text{ cvx} \overset{115.4}{\Longleftrightarrow} \forall (x,h) \in (\mathbb{R}^n)^2 \ \langle \nabla f(x+h) - \nabla f(x) \mid h \rangle \geq 0$$
$$\overset{115.3}{\Longleftrightarrow} \forall (x,h) \in (\mathbb{R}^n)^2, \ \underbrace{\langle \mathcal{D}(\nabla f)(x)(h) \mid h \rangle}_{\langle \mathcal{D}^2 f(x)(h) \mid h \rangle} \geq 0$$
$$\overset{\text{Rappel}}{\Longleftrightarrow} \forall x \ \ \mathcal{D}^2 f(x) \text{ est une matrice } \geq 0.$$
(est une forme quadratique positive).