Différentielle et convexité

Leçons : 215, 229.

Énoncés : • prop : $U \subset \mathbb{R}^n$ ouvert, $f : U \to \mathbb{R}$.

→ Si f différentiable en $x \in U$ et convexe, pour tout $y \in U$, $f(y) - f(x) \geq D_x f(y-x)$;

si f différentiable : f convexe ssi $\forall x, y \in U$, $f(y) - f(x) \geq D_x f(y-x)$.

→ Si f deux fois différentiable : f convexe ssi $\forall x \in U$, la fbl $D_x^2 f$ est positive.

→ Si f différentiable en $x \in U$ et convexe : si $D_x f = 0$ alors x est un min global.

• apli : $A \in S_n^+(\mathbb{R})$, $b \in \mathbb{R}^n$, $f : x \mapsto \langle \frac{1}{2} A x - b, x \rangle$. f est convexe et atteint son min exactement en les sols de l'éq $Ax = b$.

⊛ Prop.

• Supposons d'abord f dif en $x \in U$ et convexe. Soit $y \in U$. On déf $g : \begin{cases} [0;1] \to \mathbb{R} \\ t \mapsto f((1-t)x + ty) \end{cases}$:

$g'(0) h = D_x f(h(-x+y))$ pour $h \in \mathbb{R}$ cad $g'(0) = D_x f(y-x)$. Mais par convexité, pour $0 < t \leq 1$, $g(t) \leq (1-t) f(x) + t f(y)$ cad $g(t) - g(0) \leq -t f(x) + t f(y)$ et $\frac{g(t) - g(0)}{t} \leq f(y) - f(x)$.

En passant à la limite on obtient $D_x f(y-x) \leq f(y) - f(x)$.

Pour le 2e point il n'y a alors que la réciproque à faire. Soient $x, y \in U$, $0 \leq t \leq 1$, $z_t = (1-t)x + ty$.

On applique l'hyp à (z_t, x) et (z_t, y) : $D_{z_t} f(x - z_t) \leq f(x) - f(z_t)$, $D_{z_t} f(y - z_t) \leq f(y) - f(z_t)$.

On multiplie la 1e par $1-t$, la 2e par t, et on ajoute : $D_{z_t} f((1-t)x + ty - z_t) \leq (1-t) f(x) + t f(y) - f(z_t)$.

Le membre de gauche est $D_{z_t} f(0) = 0$, d'où la convexité.

• Supposons f convexe et prenons $x \in U$. Soit $h \in \mathbb{R}^n$: pour $t \in \mathbb{R}$ au voisinage de 0, la formule de Taylor-Young donne $f(x + th) = f(x) + t D_x f(h) + \frac{t^2}{2} D_x^2 f(h,h) + t^2 \varepsilon(t)$ avec $\varepsilon(t) \underset{t \to 0}{=} o(1)$.

D'après ce qui précède par convexité, $\frac{t^2}{2} D_x^2 f(h,h) + t^2 \varepsilon(t) = f(x+th) - f(x) - D_x f(th) \geq 0$,

ou encore $D_x^2 f(h,h) + 2 \varepsilon(t) \geq 0$. Faisant $t \to 0$, $D_x^2 f(h,h) \geq 0$.

Réciproquement on fixe $x \in U$ et $h \in U - x$: on va utiliser la caractérisation du premier point. On pose encore $g : \begin{cases} [0;1] \to \mathbb{R} \\ t \mapsto f(x+th) \end{cases}$, qui est deux fois différentiable. On lui applique la formule de Taylor-Lagrange à l'ordre 1 entre 0 et 1 : $g(1) = g(0) + g'(0) + \frac{g''(s)}{2}$ pour un certain $0 < s < 1$. Si $0 \leq t \leq 1$, $g'(t) = D_{x+th} f(h)$ et $g''(t) = D_{x+th}^2 f(h,h)$ (le h à droite vient du fait que $t \mapsto x+th$ a pour dérivée h). On peut alors réécrire $f(x+h) - f(x) - D_x f(h) = \frac{1}{2} D_{x+sh}^2 f(h,h) \geq 0$, d'où la convexité.

• Il suffit d'appliquer le point 1 : pour tout $y \in U$, $f(y) - f(x) \geq D_x f(y-x) = 0$ donc $f(y) \geq f(x)$.

□

⊗ Appli.

Si $x, h \in \mathbb{R}^n$: $f(x+h) = \langle \frac{1}{2}A(x+h) - b, x+h \rangle = f(x) + \langle \frac{1}{2}Ax - b, h \rangle + \langle \frac{1}{2}Ah, x \rangle + \langle \frac{1}{2}Ah, h \rangle$.

D'une part $\langle \frac{1}{2}Ah, h \rangle \underset{h \to 0}{=} o(h)$. D'autre part $\langle \frac{1}{2}Ah, x \rangle = \langle \frac{1}{2}Ax, h \rangle$; on obtient donc $D_x f(h) = \langle Ax - b, h \rangle$. De plus $D_{x+h} f(h) = D_x f(h) + \langle Ah, h \rangle$ donc $D_x^2 f(h,h) = \langle Ah, h \rangle$ (en fait : $x \mapsto D_x f$ affine).

On déduit de ce dernier point que f est convexe (en effet $A \in S_n^+(\mathbb{R})$ donc $\langle Ah, h \rangle \geq 0$ pour tout $h \in \mathbb{R}^n$).

De plus $D_x f = 0$ ssi $Ax = b$. Si c'est le cas, la convexité entraîne que x est un min global. Réciproquement un minimum global est aussi local, et c'est un point critique. f est donc minimale exactement en les sols de $Ax = b$. □

Complément : formule de Taylor-Lagrange. Si $f \in \mathcal{C}^n([a;b], \mathbb{R})$ est $n+1$ fois dérivable sur $]a;b[$ (où $a<b$), il existe $c \in]a;b[$ tq $f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(b-a)^k + \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1}$.

Preuve. L'idée est de se ramener à appliquer le th de Rolle. Soit $g : t \mapsto f(b) - \sum_{k=0}^{n} \frac{f^{(k)}(t)}{k!}(b-t)^k - \alpha(b-t)^{n+1}$ où $\alpha \in \mathbb{R}$ est pris tq $g(a) = 0$. On a $g \in \mathcal{C}^0([a;b], \mathbb{R})$ dérivable sur $]a;b[$, avec de plus $g(b) = 0$: d'après le th de Rolle, il existe $a < c < b$ tq $g'(c) = 0$.

Mais pour $a<t<b$:

$$g'(t) = -\sum_{k=0}^{n} \frac{1}{k!}\left(f^{(k+1)}(t)(b-t)^k - f^{(k)}(t)\,k(b-t)^{k-1}\right) + \alpha(n+1)(b-t)^n$$

$$= -\sum_{k=0}^{n} \frac{f^{(k+1)}(t)}{k!}(b-t)^k + \sum_{l=0}^{n-1} \frac{f^{(l+1)}(t)}{l!}(b-t)^l + \alpha(n+1)(b-t)^n$$

$$= -\frac{f^{(n+1)}(t)}{n!}(b-t)^n + \alpha(n+1)(b-t)^n.$$

Ainsi $g'(c) = 0$ entraîne $\alpha = \frac{f^{(n+1)}(c)}{(n+1)!}$. L'égalité voulue s'obtient par $g(a) = 0$. □

Réf :
- Rouvière : p 119 (exo 42, point 1 de la prop), p 319 (exo 103, point 2 de la prop), p 371 (exo 119, point 3 de la prop).
- Ciarlet : p 188 (supplément sur l'optimisation et la fonction de l'appli).

↳ Attention au sens de $f(h) \underset{h \to 0}{=} o(h^2)$. Cela signifie qu'il existe $\varepsilon(h) \underset{h \to 0}{\longrightarrow} 0$ tq $f(h) = \|h\|^2 \varepsilon(h)$.

↳ Savoir faire les dessins correspondant à la preuve pour le point 1 de la prop. Ils sont faits dans Rouvière.

↳ Analogue strict du point 2 : vrai dans un seul sens : si $\forall x \in U$, $D_x^2 f$ déf positive, alors f strictement convexe (même preuve). En particulier dans l'appli si A déf pos, f est strictement convexe. Contre-ex à la réciproque : $f : \left\{\begin{array}{l}\mathbb{R} \to \mathbb{R} \\ x \mapsto x^4\end{array}\right.$ est strictement convexe mais $f''(0) = 0$.

↳ On utilise dans la réciproque du point 2 la formule de Taylor - Lagrange, qui est prouvée en cpl. À ne pas confondre avec l'inégalité de Taylor - Lagrange, qui en est un corollaire direct.

↳ Si f est supposée $\mathcal{C}^2$, on peut prouver la réciproque du point 2 à l'aide de Taylor reste intégral directement sur f (fait dans Rouvière).

↳ L'appli mq l'on peut approximer les sol de $Ax = b$ par l'optimisation, en cherchant les min de f. En particulier si A est déf pos, on peut approximer $A^{-1}b$ ainsi. La méthode du gradient optimal converge alors pour f (voir Ciarlet).