Séance 7 : Bases du calcul différentiel

Dans le programme d'agrégation, le calcul différentiel s'effectue dans des evn de dimension finie.
Néanmoins, la notion de différentielle sera donnée dans un evn quelconque.

1) Différentiabilité d'une fonction

Def : soit $f : (E, \| \ \|_E) \to (F, \| \ \|_F)$ et $a \in E$. On dit que f est différentiable en a s'il existe $df_a \in \mathcal{L}_c(E,F)$ (l'ensemble des applications linéaires et continues de E dans F) telle que

$$\boxed{\forall h \in E,\ f(a+h) = f(a) + df_a(h) + o(\|h\|)}$$

où $o(\|h\|) = \varepsilon(h)$ est telle que $\lim\limits_{h \to 0} \frac{\varepsilon(h)}{\|h\|} = 0$

(dans ce cas df_a est unique et appelée différentielle de f en a)

Remarques :

1.) $E = F = \mathbb{R}$, on retrouve la dérivée classique $df_a(h) = f'(a).h$ (DL ordre 1).

2) si E et F sont de dimension finie, $\mathcal{L}_c(E,F) = \mathcal{L}(E,F)$

3) si f différentiable en a, f continue en a.

On dispose de théorèmes généraux pour calculer des différentielles :

Proposition : (i) si f, g différentiables en a, alors $d(f+g)_a = df_a + dg_a$

(ii) si $f : E \to F$ et $g : E \to \mathbb{R}$ différentiables en a, alors $\underbrace{d(fg)_a}_{\in \mathcal{L}_c(E,F)} = \underbrace{g(a)}_{\mathbb{R}} \underbrace{df_a}_{\mathcal{L}_c(E,F)} + \underbrace{dg_a}_{\mathcal{L}_c(E,\mathbb{R})} \cdot \underbrace{f(a)}_{F}$

(iii) si $f : E \to F$ et $g : F \to G$, alors $\underbrace{d(g \circ f)_a}_{\in \mathcal{L}_c(E,G)} = \underbrace{dg_{f(a)}}_{\in \mathcal{L}_c(F,G)} \circ \underbrace{df_a}_{\mathcal{L}_c(E,F)}$ (chaîne)

preuve : (exercice, à partir de l'identité de la différentielle)

Exemple : $f : \mathbb{R}^n \to \mathbb{R}$ tq $f(x) = \frac{1}{2} \langle x, Ax \rangle - \langle b, x \rangle$ où $A \in \mathcal{S}_n(\mathbb{R})$, $b \in \mathbb{R}^n$

On montre que f est différentiable et $df_a(h) = \langle Aa - b, h \rangle$ (exercice)

2) Cas de la dimension finie

* Si $f : \mathbb{R}^n \to \mathbb{R}$, alors $df_a \in \mathcal{L}(\mathbb{R}^n, \mathbb{R})$ (forme linéaire) et s'identifie à un vecteur de $\mathbb{R}^n$ noté $\nabla f(a)$ (gradient de f en a) : $\forall h \in \mathbb{R}^n$, $df_a(h) = \langle \nabla f(a), h \rangle$

* Si $f : \mathbb{R}^n \to \mathbb{R}^p$ $(p > 1)$, df_a

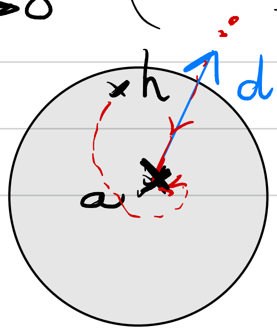
s'identifie à une matrice $Jf_a \in \mathcal{M}_{p,n}(\mathbb{R})$ (Jacobienne de f en a) :

$$df_a(h) = Jf_a \cdot h$$

On introduit alors la notion de dérivée directionnelle :

def : $f : E \to F$, $a \in E$, $d \in E \setminus \{0\}$. On appelle dérivée de f en a suivant la direction d, la quantité (si elle existe) :

$$\lim_{t \to 0} \left(\frac{f(a+td) - f(a)}{t} \right)$$



Il s'agit d'une notion plus faible que la différentiabilité. Par exemple)

$$f \begin{pmatrix} \mathbb{R}^2 \to \mathbb{R} \\ (x,y) \longmapsto \begin{cases} \dfrac{x|y|}{\sqrt{x^2+y^2}} \\ 0 \text{ si } (x,y) = (0,0) \end{cases} \end{pmatrix}$$

$\Longrightarrow d = (d_1, d_2)$

$$\frac{f(td_1, td_2) - 0}{t} = \frac{t\cancel{|t|}\,|d_1|\,|d_2|}{t\cancel{|t|}\sqrt{d_1^2+d_2^2}}$$

$\longrightarrow$ Cste quand $t \to 0$

f possède une dérivée directionnelle nulle en 0 dans toutes les directions

$\Longrightarrow$ f n'est pas différentiable : si

$$f(x,y) = 0 + \langle \nabla f(0,0), (x,y) \rangle + o(\|(x,y)\|)$$

Avec $(x,y)=(x,0)$ puis $(0,y)$, on trouve $\nabla f(0,0)=(0,0)$. Alors :

$$f\left(\frac{1}{n},\frac{1}{n}\right)=\frac{1}{\sqrt{2}n}\neq o\left(\frac{1}{n}\right)$$

Ceci est absurde

En dimension finie, on définit la notion de dérivée partielle : si $f:\mathbb{R}^n\to\mathbb{R}$

$$\frac{\partial f}{\partial x_i}(a)=\lim_{t\to 0}\frac{f(a_1,\dots,a_i+t,\dots,a_n)-f(a)}{t}$$

On peut alors faire le lien suivant si f différentiable :

* si $f:\mathbb{R}^n\to\mathbb{R}$, $\nabla f(a)=\begin{pmatrix}\frac{\partial f}{\partial x_1}(a)\\ \vdots\\ \frac{\partial f}{\partial x_n}(a)\end{pmatrix}$
* si $f:\mathbb{R}^n\to\mathbb{R}^p$, $Jf_a \equiv \left[\frac{\partial f_i}{\partial x_j}(a)\right]$

Attention, pas de réciproque :

$$f(x,y)=\begin{cases}\dfrac{xy}{x^2+y^2} & \text{si } (x,y)\neq(0,0)\\ 0 \text{ si } (x,y)=(0,0)\end{cases}$$

f admet des dérivées partielles en $(0,0)$ (nulles) mais f n'est pas continue (donc non différentiable) en 0.

* Les théorèmes généraux peuvent se réécrire en dimension finie avec les dérivées partielles : en particulier, si $f:\mathbb{R}^n\to\mathbb{R}^p$ et $g:\mathbb{R}^p\to\mathbb{R}$,

$$\frac{\partial g\circ f}{\partial x_i}(a)=\sum_{k=1}^{p}\frac{\partial g}{\partial x_k}(f(a))\frac{\partial f_k}{\partial x_i}(a)$$

Par exemple si $h(t,x) = u(t, x+ct)$

où $u : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ et $c \in \mathbb{R}$

$h : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$

On veut calculer $\frac{\partial h}{\partial t}(t,x)$:

$$(t,x) \xmapsto{f} (t, x+ct) \xrightarrow{g} u(t,x+ct)$$

$$\frac{\partial h}{\partial t}(t,x) = \frac{\partial u}{\partial t}(t,x+ct)\times 1 + \dots \frac{\partial u}{\partial x}(t,x+ct)\times c$$

3) Théorème des accroissements finis

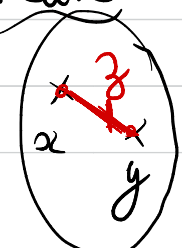
Le théorème sous sa forme réelle ne subsiste que sous forme inégalité

De plus si $f : U \subset E \to F$ où U ouvert, alors U doit également être convexe.

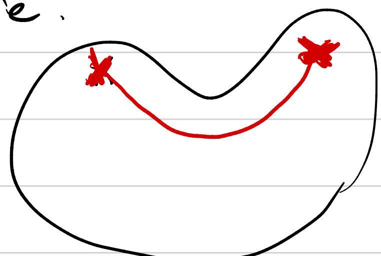
Théorème : si $f : U \to F$, U ouvert convexe. Soit $x, y \in U$, on suppose que f est différentiable pour tout $z \in [x,y]$. Alors

$$\boxed{\|f(y) - f(x)\|_F \leq \sup_{z\in[x,y]} |||df_z||| \,.\, \|y-x\|_E}$$

preuve : formule de la chaîne : $t \mapsto f(tx+(1-t)y)$ et inégalité des accroissements finis sur $[0,1]$.



Corollaire : si $f : U \to F$, U convexe (et même connexe) est différentiable sur U et $df_a = 0$ $\forall a \in U$, alors f est constante.



4) Régularité, fonction $C^1, C^2, \ldots C^\infty$

Def : $f : U$ ouvert $\subset E \to F$ est de classe C^1 sur U si df_a existe en tout point et est une fonction continue en a : $df \in \mathcal{C}(U, \mathcal{L}_c(E,F))$

En dimension finie, on a l'équivalence :

Théorème : soit $f : U \subset \mathbb{R}^n \to \mathbb{R}^p$ on a

f C^1 sur $U \iff f$ possède des dérivées partielles continues en tout point de U.

On itère ensuite en définissant des différentielles d'ordre supérieur : f C^2 si df est C^1. On note alors $d^2 f_a \in \mathcal{L}_c(E, \mathcal{L}_c(E,F))$ la différentielle seconde de f en a

En dimension finie, on a l'équivalence :

f C^2 sur $U \iff f$ possède des dérivées partielles secondes continues telles que :

$$\frac{\partial^2 f(a)}{\partial a_i \partial a_j} = \frac{\partial^2 f(a)}{\partial a_j \partial a_i}$$

Si $f : \mathbb{R}^n \to \mathbb{R}$, on identifie :

$$\begin{cases} df_a \text{ à } \nabla f(a) \in \mathbb{R}^n & : \text{gradient} \\ d^2 f_a \text{ à } \operatorname{Hess} f(a) \equiv \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right] & \end{cases}$$

: matrice Hessienne

De même, pour des ordres supérieurs.

<u>Exemple</u> : ~~si~~ $f : \mathbb{R}^n \to \mathbb{R}$ tq

$f(x) = \frac{1}{2}\langle x, Ax\rangle - \langle b, x\rangle$, on a

vu que $\nabla f(x) = Ax - b$

ce qui donne facilement

$d^2 f_x(h) = Ah$, soit

$\operatorname{Hess} f(x) = A$

On le retrouve avec les dérivées partielles :

$$f(x_1, \dots x_n) = \frac{1}{2} \sum_{i,j} x_i A_{i,j} x_j - \sum_i b_i x_i$$

$$\Rightarrow \nabla f(x_1, \dots x_n) = \begin{pmatrix} \sum_j A_{1,j} x_j - b_1 \\ \vdots \\ \sum_j A_{n,j} x_j - b_n \end{pmatrix}$$

puis

$$\operatorname{Hess}(f)(a) = \begin{pmatrix} A_{11} & \dots & A_{1,n} \\ \vdots & & \\ A_{n,1} & \dots & A_{n,n} \end{pmatrix}$$

5) Formules de Taylor :

On s'intéresse au cas d'une fonction $f : U$ ouvert $\subset \mathbb{R}^n \longrightarrow \mathbb{R}$, qu'on supposera C^1, C^2, ... voire C^∞.

* La définition de la différentielle correspond à la formule de Taylor Young à l'ordre 1 : si f C^1

$$f(a+h) = \underbrace{f(a) + \langle \nabla f(a), h \rangle}_{\text{approximation d'ordre 1 près de } a} + o(\|h\|)$$

* Si f est C^2 on peut écrire la formule de Taylor Young à l'ordre 2 :

$$f(a+h) = \underbrace{f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} \langle h, \mathrm{Hess} f(a) . h \rangle}_{\text{approximation à l'ordre 2}} + o(\|h\|^2)$$

De manière générale, on peut écrire la formule de Taylor Young à l'ordre n : si f C^n,

$$f(a+h) = f(a) + \sum_{k=1}^{n} \frac{1}{k!} d^k f_a(h, h, \dots h) + o(\|h\|^n)$$

La formule de Taylor-Lagrange n'est plus valable. On peut néanmoins obtenir une formule globale avec la formule de Taylor intégrale en se ramenant au cas réel.

* Si U convexe ouvert $\subset \mathbb{R}^n$ et $f : U \longrightarrow \mathbb{R}$, C^{n+1}, si $x, y \in U$, on note

$$h : \left(\begin{array}{ccc} [0,1] & \longrightarrow & \mathbb{R} \\ t & \longmapsto & f(\underbrace{ty + (1-t)x}_{z_t}) \end{array} \right)$$

On a alors :

$$\underbrace{h(1)}_{f(y)} = \underbrace{h(0)}_{f(x)} + \sum_{k=1}^{n} \frac{1}{k!} h^{(k)}(0) + \int_0^1 \frac{(1-t)^n}{n!} h^{(n+1)}(t)\,dt$$

En particulier :

$$\begin{cases} h'(0) = \langle \nabla f(x), y-x \rangle \\ h''(0) = \langle y-x, \operatorname{Hess} f(x)(y-x) \rangle \end{cases}$$

et

$$\begin{cases} h'(t) = \langle y-x, \nabla f(z_t) \rangle \\ h''(t) = \langle y-x, \operatorname{Hess} f(z_t)(y-x) \rangle \end{cases}$$

On retrouve l'inégalité des accroissements finis si f est C^1. Si f est C^2, on a :

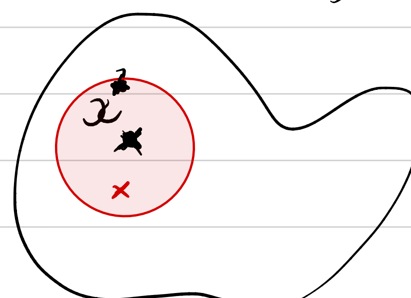
$$\| f(y) - f(x) - \langle \nabla f(x), y-x \rangle \| \leq C \|y-x\|^2$$

6) Conditions d'optimalité

Soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$, C^1 voire C^2

On cherche à calculer, ou à caractériser d'éventuels minima ou maxima locaux de f :

Déf : $x^* \in U$ ouvert de $\mathbb{R}^n$ est un minimum local de f, s'il existe $r > 0$ tq $\forall x \in B(x^*, r)$, $f(x) \geq f(x^*)$



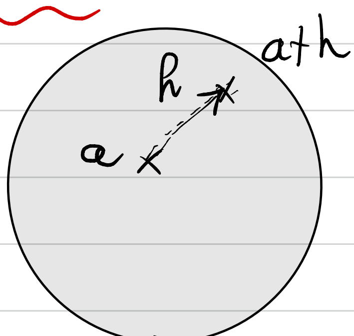
(pour le max, on considère le min de $-f$)

Il existe 3 résultats donnant des conditions (nécessaires ou suffisantes) pour que x^* soit un minimum local de f :

(CN1)

Théorème : soit $f : U \text{ ouvert} \subset \mathbb{R}^n \to \mathbb{R}$, $f\ \mathcal{C}^1$ et $x^* \in U$ minimum local de f sur U.
Alors $\nabla f(x^*) = 0$.

preuve : $h \in \mathbb{R}^n$ fixé. Par Taylor Young à l'ordre 1 :



$$f(a+th) = f(a) + \langle th, \nabla f(a) \rangle + o(t) \quad (t \to 0)$$

En divisant par $t > 0$, on a :

$$\frac{f(a+th) - f(a)}{t} = \langle h, \nabla f(a) \rangle + o(1)$$

Or si t petit, $\dfrac{f(a+th) - f(a)}{t} \geq 0$

Par passage à la limite quand $t \to 0^+$, on a

$$0 \leq \langle h, \nabla f(a) \rangle$$

De même $h \to -h$, on a

$$0 \leq \langle -h, \nabla f(a) \rangle$$

$$\Rightarrow \langle \nabla f(a), h \rangle = 0 \quad \forall h \in \mathbb{R}^n$$

D'où $\nabla f(a) = 0$ //

Remarques :

1) Cette condition vaut aussi si x^* maximum local de f.

2) Il s'agit d'une condition nécessaire mais non suffisante :

($n=1$: $f(x)=x^3$ en $x=0$)

($n=2$: $f(x,y)=xy$ en $(0,0)$)

Si f est C^2, on peut trouver une condition nécessaire à l'ordre 2 (CN2)

Théorème : $f : U \subset \mathbb{R}^n$, ouvert $\to \mathbb{R}$

f C^2, $x^* \in U$ minimum local de f.

Alors : $\nabla f(x^*)=0$ et

$\mathrm{Hess}(f)(x^*)$ est une matrice semi-définie positive : $\forall h \in \mathbb{R}^n$

$\langle h, Hh \rangle \geq 0$

preuve : on s'inspire de la preuve précédente avec la formule de Taylor à l'ordre 2.

Remarque : cette condition est toujours seulement nécessaire mais pas suffisante :

($n=1$, $f(x)=x^3$: $f'(0)=f''(0)=0$)

($n=2$, $f(x,y)=x^3+y^2$

$\nabla f(0,0)=0$

$\mathrm{Hess}\, f(0,0)=\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$ s.d.p.

mais $(0,0)$ n'est pas un minimum local)

Si on renforce légèrement les propriétés sur x^*, on arrive à une condition suffisante :

Théorème (CS2) soit $f : U \subset \mathbb{R}^n \to \mathbb{R}$ f C^2. Si $x^* \in U$ est tel que :

$\nabla f(x^*) = 0$ et $\mathrm{Hess}(f)(x^*)$ définie positive, alors x^* est un minimum local de f.

preuve : on écrit T.Y à l'ordre 2 :

$$f(x^* + th) = f(x^*) + 0 + t^2 \langle h, \mathrm{Hess}(f)(x^*)h \rangle + o(t^2)$$

$$\Rightarrow \frac{f(x^* + th) - f(x^*)}{t^2} = \underbrace{\langle h, H(f)(x^*).h \rangle}_{>0} + o(1)$$

Pour $|t|$ petit, $\langle h, Hf(x^*).h \rangle + o(1) \geq 0$ (uniformément en h dans $B(0,1)$) $\Rightarrow x^*$ minimum local

Remarque : cette condition est suffisante, pas forcément nécessaire :

$n = 1$: $f(x) = x^4$

$n = 2$: $f(x,y) = x^4 + y^2$: $\mathrm{Hess}(f)(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}$

7) Cas des fonctions convexes

$f : U$ convexe $\subset \mathbb{R}^n \to \mathbb{R}$ est **convexe** si $\forall (x,y,t) \in U \times U \times [0,1]$,

$$f(tx + (1-t)y) \leq t f(x) + (1-t) f(y)$$

(+ stricte convexité : inégalité stricte si $t \in]0,1[$ et $x \neq y$)

On a les caractérisations suivantes si f C^1 ou C^2 :

Théorème : * si f est C^1, f est convexe (resp. strictement convexe) ssi $\forall (x,y) \in U^2$,
$f(y) \geqslant f(x) + \langle \nabla f(x), y-x \rangle$
(resp. inégalité stricte et $x \neq y$)

* si f est C^2,
f convexe $\Longleftrightarrow$ $\forall x \in U$, $Hess(f)(x)$ s.d.p
f strictement convexe $\Longleftarrow$ $\forall x \in U$ $Hess(f)(x)$ d.p

Les 3 conditions précédentes d'optimalité se simplifient dans le cas convexe :

Corollaire : si $f : U$ convexe $\subset \mathbb{R}^n \to \mathbb{R}$ et f C^1. Si x^* est tel que $\nabla f(x^*) = 0$, alors x^* est un minimum de f.