Chapitre XV Calcul différentiel

On étudie dans ce chapitre des fonctions d'un *evn* E vers un *evn* F de dimensions finies $\dim E = p$ et $\dim F = n$. Quitte à choisir une base de E et une base de F :

— cela reviendra à étudier une fonction de $\mathbb{R}^p$ vers $\mathbb{R}^n$ qui associera à chaque point $(x_1, \cdots, x_p) \in \mathbb{R}^p$ un vecteur $f(x_1, \cdots, x_p) \in \mathbb{R}^n$;

— on sait que toutes les normes sur E, *resp.* sur F, seront équivalentes et que, d'après le corollaire 19 du chapitre XI, une fonction sera continue *ssi* chacune de ses coordonnées dans la base de F est continue.

On appelle ces fonctions des champs de vecteurs : le champ électrique, le champ magnétique, un champ de gradients (figure XV.1), etc. En particulier, si $\dim F = 1$, alors on parle de champs scalaires : la pression, la température, l'altitude, etc.

XV.1 Dérivées partielles

Définition 1

Soient un ouvert $D \subset \mathbb{R}^2$ et une fonction $f : D \to \mathbb{R}, (x,y) \mapsto f(x,y)$. Soit un point $(a,b) \in D$.

1. Si
$$\lim_{h\to 0} \frac{f(a+h,b)-f(a,b)}{h}$$
existe et est finie, alors ce nombre réel est noté $\partial_1 f(a,b)$ ou $\frac{\partial f}{\partial x}(a,b)$ et est appelé la première **dérivée partielle** de f en (a,b).
2. Si
$$\lim_{k\to 0} \frac{f(a,b+k)-f(a,b)}{k}$$
existe et est finie, alors ce nombre réel est noté $\partial_2 f(a,b)$ ou $\frac{\partial f}{\partial y}(a,b)$ et est appelé la deuxième **dérivée partielle** de f en (a,b).
3. Soit $i \in [\![1,2]\!]$. Si $\partial_i f(a,b)$ existe pour tout $(a,b) \in D$, alors la fonction $\partial_i f : D \to \mathbb{R}$ est appelée la $i-$ème **dérivée partielle** de f.
4. On dit que la fonction f est de classe $\mathcal{C}^1$ sur D si les 2 dérivées partielles $\partial_1 f$ et $\partial_2 f$ existent et sont continues sur D.

Exercice 2 (Une fonction qui possède des dérivées partielles mais n'est pas continue) — *Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par*
$$f(x,y) = \frac{xy}{x^2+y^2} \quad \textit{si} \quad (x,y) \neq (0,0) \quad \textit{et} \quad f(0,0) = 0.$$
Montrer que la fonction f possède des dérivées partielles $\partial_1 f(0,0)$ et $\partial_2 f(0,0)$ en $(0,0)$ mais que f n'est pas continue en $(0,0)$.

Exercice 3 (Une fonction qui possède des dérivées partielles non continues) — *Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par*
$$f(x,y) = x^2 \sin\left(\frac{y}{x}\right) \quad \textit{si} \quad x \neq 0 \quad \textit{et} \quad f(0,y) = 0.$$

1. *Montrer que f est continue sur $\mathbb{R}^2$.*
2. *Montrer que les dérivées partielles $\partial_1 f(x,y)$ et $\partial_2 f(x,y)$ existent si $x \neq 0$ et les calculer.*
3. *Montrer que les dérivées partielles $\partial_1 f(0,y)$ et $\partial_2 f(0,y)$ existent et les calculer.*
4. *Montrer que la fonction f n'est pas de classe $\mathcal{C}^1$.*

THÉORÈME 4 (**Formule de Taylor & Young à l'ordre 1**)
Soient un ouvert $D \subset \mathbb{R}^2$ et une fonction $f : D \to \mathbb{R}$. Si f est de classe $\mathcal{C}^1$, alors, pour tout $(a, b) \in D$,

$$f(a+h, b+k) = f(a, b) + h\partial_1 f(a, b) + k\partial_2 f(a, b) + \|(h, k)\|\varepsilon(h, k)$$

où $\varepsilon(h, k) \underset{(h,k)\to 0}{\longrightarrow} (0, 0)$.

Preuve —

$$\begin{array}{rcl} f(a+h,b+k)-f(a,b) & = & f(a+h,b+k)-f(a,b+k) \\ & + & f(a,b+k)-f(a,b). \end{array}$$

On applique le théorème des accroissements finis à chaque différence :

$$\begin{array}{rcl} f(a+h,b+k)-f(a,b) & = & h\partial_1 f(a+\theta_1 h,b+k) \\ & + & k\partial_2 f(a,b+\theta_2 k), \end{array}$$

où θ_1 et θ_2 appartiennent à $]0,1[$. Or chaque dérivée partielle est continue (car f est de classe $\mathcal{C}^1$), d'où :

$$\begin{array}{rcl} f(a+h,b+k)-f(a,b) & = & h\left[\partial_1 f(a,b)+\varepsilon_1(h,k)\right] \\ & + & k\left[\partial_2 f(a,b)+\varepsilon_2(h,k)\right], \end{array}$$

où $\varepsilon_1(h,k)$ et $\varepsilon_2(h,k)$ tendent vers zéro quand $(h,k)\to(0,0)$.

D'où $f(a+h,b+k)-f(a,b) = h\partial_1 f(a,b)+k\partial_2 f(a,b)+h\varepsilon_1(h,k)+k\varepsilon_2(h,k)$.

Or $|h\varepsilon_1(h,k)+k\varepsilon_2(h,k)| \le |h\varepsilon_1(h,k)|+|k\varepsilon_2(h,k)| \le \sqrt{h^2+k^2}\cdot\underbrace{\left(|\varepsilon_1(h,k)|+|\varepsilon_2(h,k)|\right)}_{\longrightarrow 0\ ((h,k)\to(0,0))}$. □

REMARQUE 5 — *1. (fonctions de $\mathbb{R}^p$ vers $\mathbb{R}$) Soit $p \in \mathbb{N}^*$. De même que les fonctions de $\mathbb{R}^2$ vers $\mathbb{R}$, on peut étudier une fonction de $\mathbb{R}^p$ vers $\mathbb{R}$, définir ses p dérivées partielles $\partial_1 f, \cdots, \partial_p f$ si elles existent et dire que la fonction f est de classe $\mathcal{C}^1$ sur un ouvert $U \subset \mathbb{R}^p$ si ses p dérivées partielles sont continues.*
D'après la formule de Taylor & Young, si la fonction f est de classe $\mathcal{C}^1$ sur U, alors : en tout point $a = (a_1, \cdots, a_p) \in U$,

$$f(a+h) = f(a) + h_1\partial_1 f(a) + \cdots + h_p\partial_p f(a) + \|h\|\varepsilon(h),$$

où $h = (h_1, \cdots, h_p)$ et $\varepsilon(h) \underset{h\to 0}{\longrightarrow} 0$;

2. (fonctions de $\mathbb{R}^p$ vers $\mathbb{R}^n$) Soient $p \in \mathbb{N}^*$ et $n \in \mathbb{N}^*$, un ouvert $U \subset \mathbb{R}^p$ et une fonction

$$f : U \to \mathbb{R}^n,\ x = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \mapsto \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}.$$

Soit $a \in U$. Si chacune des n fonctions $f_i : U \to \mathbb{R},\ x \mapsto f_i(x)$ est de classe $\mathcal{C}^1$ sur U, alors

$$\forall i \in [\![1, n]\!],\quad f_i(a+h) = f_i(a) + \partial_1 f_i(a)h_1 + \cdots + \partial_p f_i(a)h_p + \|h\|\varepsilon(h).$$

Matriciellement :

$$\underbrace{\begin{pmatrix} f_1(a+h) \\ \vdots \\ f_n(a+h) \end{pmatrix}}_{f(a+h)} = \underbrace{\begin{pmatrix} f_1(a) \\ \vdots \\ f_n(a) \end{pmatrix}}_{f(a)} + \underbrace{\begin{array}{c} \begin{array}{ccc} \partial_1 f(a) & \cdots & \partial_p f(a) \\ \downarrow & & \downarrow \end{array} \\ \begin{pmatrix} \partial_1 f_1(a) & \cdots & \partial_p f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_n(a) & \cdots & \partial_p f_n(a) \end{pmatrix} \end{array}}_{J_f(a)} \underbrace{\begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}}_{h} + \|h\| \underbrace{\begin{pmatrix} \varepsilon_1(h) \\ \vdots \\ \varepsilon_n(h) \end{pmatrix}}_{\varepsilon(h)}.$$

DÉFINITION 6
Soit $f : U \to \mathbb{R}^n,\ x = (x_1, \cdots, x_p) \mapsto \big(f_1(x), \cdots, f_n(x)\big)$ définie sur un ouvert $U \subset \mathbb{R}^p$. Si chaque fonction f_i admet p dérivées partielles en a, alors la matrice $J_f(a) \in \mathcal{M}_{np}(\mathbb{R})$ définie par $(J_f(a))_{ij} = \dfrac{\partial f_i}{\partial x_j}(a)$ pour chaque $(i,j) \in [\![1,n]\!] \times [\![1,p]\!]$ est appelée **la jacobienne** de f en a.

Si $n = 1$, alors la jacobienne est une matrice ligne avec p colonnes :

$$f(a+h) = f(a) + \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(a) & \cdots & \dfrac{\partial f}{\partial x_p}(a) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix} + \|h\|\varepsilon(h).$$

Si $p = 1$, alors la jacobienne est une matrice colonne avec n lignes :

$$\begin{pmatrix} f_1(t+h) \\ \vdots \\ f_n(t+h) \end{pmatrix} = \begin{pmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{pmatrix} + h \underbrace{\begin{pmatrix} f_1'(t) \\ \vdots \\ f_n'(t) \end{pmatrix}}_{f'(t)} + |h| \begin{pmatrix} \varepsilon_1(h) \\ \vdots \\ \varepsilon_n(h) \end{pmatrix}$$

et $f'(t)$ est le vecteur-vitesse.

XV.2 La différentielle d'une fonction

Dans toute la suite, E et F sont des *evn* de dimensions finies $p = \dim E$ et $n = \dim F$. Si l'on choisit une base de E et une base de F, alors on pourra confondre $x \in E$ et $(x_1, \cdots, x_p) \in \mathbb{R}^p$ d'une part, $f(x) \in F$ et $(f_1(x), \cdots, f_n(x)) \in \mathbb{R}^n$ d'autre part.

PROPOSITION-DÉFINITION 7
Soient U un ouvert de E, un point $a \in U$ et une fonction $f : U \to F$. On dit que f est **différentiable** en a s'il existe une application linéaire $\ell_a : E \to F$ telle que

$$f(a+h) = f(a) + \ell_a(h) + \underbrace{\|h\|\varepsilon(h)}_{=o(h)}, \qquad \text{où} \quad \varepsilon(h) \underset{h \to 0_E}{\longrightarrow} 0_F.$$

Si f est différentiable en a, alors :

(i) l'application ℓ_a est unique, on l'appelle **la différentielle de** f **en** a et on la note

$$\ell_a = df(a) : E \to F,\ h \mapsto \ell_a(h) = df(a) \cdot h\ ;$$

(ii) les dérivées partielles de f en a existent et

$$\forall h \in E,\ df(a) \cdot h = h_1 \partial_1 f(a) + \cdots h_p \partial_p f(a) = \sum_{j=1}^{p} h_i \partial_i f(a).$$

Preuve — *Une première preuve de l'unicité* : si L_1 et L_2 sont deux applications linéaires telles que $f(a+h) = f(a) + L_1(h) + \|h\|\varepsilon_1(h) = f(a) + L_2(h) + \|h\|\varepsilon_2(h)$, alors $L_2(h) - L_1(h) = \|h\|\varepsilon(h)$. D'où, pour chaque $t \in \mathbb{R}$, $L_2(th) - L_1(th) = \|th\|\varepsilon(th)$. Or $L_2(th) = tL_2(h)$ et $L_1(th) = tL_1(h)$ car L_1 et L_2 sont linéaires. D'où, pour tout $t \in \mathbb{R}^*$, $L_2(h) - L_1(h) = \|h\|\varepsilon(th) \underset{t \to 0}{\longrightarrow} 0_F$. D'où $L_2(h) = L_1(h)$. C'est vrai pour chaque $h \in E$, donc $L_1 = L_2$.

Une seconde preuve de l'unicité : si f est différentiable, alors $f(a+h) - f(a) = df(a) \cdot h + o(h)$. En particulier, si $h = te_i$, alors $df(a) \cdot h = df(a) \cdot te_i = t df(a) \cdot e_i$ car l'application $df(a)$ est linéaire. D'où $f(a_1, \cdots, a_i + t, \cdots, a_p) - f(a_1, \cdots, a_i, \cdots, a_p) = t df(a) \cdot e_i + o(t)$. Donc $\frac{f(a_1, \cdots, a_i+t, \cdots, a_p) - f(a_1, \cdots, a_i, \cdots, a_p)}{t} \underset{t \to 0}{\longrightarrow} df(a) \cdot e_i$, ce qui prouve que $\partial_i f(a)$ existe et est égal à $df(a) \cdot e_i$. Enfin, si $h = h_1 e_1 + \cdots + h_p e_p$, alors, par linéarité, $df(a) \cdot h = h_1 df(a) \cdot e_1 + \cdots + h_p df(a) \cdot e_p$, ce qui prouve à la fois l'unicité et la formule de $df(a) \cdot h$ pour tout vecteur $h \in E$. $\square$

EXEMPLE 8 —

1. *La fonction* $f : \mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R}),\ A \mapsto A^2$ *est différentiable en chaque point* $A \in \mathcal{M}_n(\mathbb{R})$ *et*
$$df(A) : \mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R}),\ H \mapsto AH + HA = df(A) \cdot H.$$
Preuve — $f(A+H) = (A+H)(A+H) = A^2 + AH + HA + H^2 = f(A) + AH + HA + H^2$. *Pour montrer que le reste* $R(H) = H^2$ *est un* $o(H)$*, choisissons une norme sous-multiplicative (équivalente à toute autre norme, car* $\mathcal{M}_n(\mathbb{R})$ *est de dimension finie) :* $\|R(H)\| = \|H^2\| \leq \|H\|^2$*. D'où* $R(H) = o(H)$*. Donc l'application linéaire* $\mathcal{M}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R}),\ H \mapsto AH + HA$ *est la différentielle de* f *en* A. □

2. *Une norme* $\|\cdot\|$ *sur un ev* E *est toujours continue sur* E *mais n'est jamais différentiable en* 0_E.
Preuve — *On sait que toute norme est continue car* 1*-lipschitzienne (exemple 36 du chapitre XI). Mais elle n'est pas différentiable en* 0 *car (par l'absurde) : s'il existe une application linéaire* $L : E \to \mathbb{R}$ *telle que* $\forall h \in E,\ \|0_E + h\| = \|0_E\| + L(h) + o(\|h\|)$*, alors* $L(th) + o(th) = L(-th) + o(-th)$*, d'où* $2tL(h) = \|th\|\varepsilon(th)$*, d'où (en divisant par* $2t$*) :* $L(h) = \|h\|\varepsilon(th) \underset{t\to 0}{\longrightarrow} 0$*. D'où* $L(h) = 0$*. C'est vrai pour chaque* $h \in E$*, donc* $L = 0$*. C'est absurde car, alors,* $\|h\| = o(h)$. □

EXERCICE 9 —

1. *Soient* $H \in \mathcal{M}_n(\mathbb{R})$ *et* I_n *la matrice identité. Montrer que*
$$\det(I_n + H) = 1 + \mathrm{tr}(H) + o(H).$$
2. *En déduire que la fonction* $\det : \mathcal{M}_n(\mathbb{R}) \to \mathbb{R},\ M \mapsto \det(M)$ *est différentiable en* I_n*. Quelle est sa différentielle en* I_n *?*
3. *Soit* $A \in \mathcal{M}_n(\mathbb{R})$ *une matrice inversible. Montrer que* $\det$ *est différentiable en* A*. Quelle est sa différentielle en* A *?*

PROPOSITION 10

$$\begin{array}{ccc} f \text{ est différentiable en } a & \Longrightarrow & f \text{ est continue en } a \\ \Downarrow \not\Uparrow & & \\ f \text{ possède des dérivées partielles en } a & \not\Longrightarrow & f \text{ est continue en } a \end{array}$$

Preuve — D'après la proposition 7, si f est différentiable en a, alors f possède des dérivées partielles en a.

Si f est différentiable en a, alors f est continue en a car $f(a+h) - f(a) = df(a)\cdot h + \|h\|\varepsilon(h) \underset{h\to 0_E}{\longrightarrow} 0_F$. En effet, $\|h\|\varepsilon(h) \longrightarrow 0_F$. Et $df(a)\cdot h \longrightarrow df(a)\cdot 0_E = 0_F$. car $df(a)$ est continue comme toute application linéaire sur un *ev* de dimension finie. Mais l'existence de dérivées partielles n'implique pas la continuité (ni *a fortiori* la différentiabilité), comme le montre l'exercice 2. □

REMARQUE 11 — *Soient un intervalle ouvert* $I \subset \mathbb{R}$ *et* $a \in I$*. Une fonction* $f : I \to \mathbb{R}$ *est dérivable a si, et seulement si, elle est différentiable en* a*. Et alors* $f'(a) = df(a) \cdot 1$.

Preuve — *Si* f *est dérivable en* a,
$$\lim_{h\to 0} \frac{f(a+h) - f(a)}{h} = f'(a), \quad \text{d'où} \quad f(a+h) = f(a) + hf'(a) + h\varepsilon(h),$$
donc f *est différentiable en* a *et* $df(a)\cdot h = hf'(a)$*. Réciproquement, si* f *est différentiable en* a*, alors* $f(a+h) = f(a) + df(a)\cdot h + h\varepsilon(h)$ *et* $df(a)\cdot h = hdf(a)\cdot 1$ *car* df_a *est linéaire, d'où* $\frac{f(a+h)-f(a)}{h}$ *tend vers* $df(a)\cdot 1$ *quand* h *tend vers 0, donc* f *est dérivable en* a *et* $f'(a) = df(a)\cdot 1$. □

DÉFINITION 12

Si une fonction $f : E \to F$ est différentiable en chaque point $a \in U$ d'une partie $U \subset E$, alors on dit que f est différentiable sur U et l'application
$$df : U \to \mathcal{L}(E,F),\ a \mapsto df(a)$$
est appelée la différentielle de f sur U.

Si chacune des n fonctions f_i est de classe $\mathcal{C}^1$ sur U, alors on dit que f est de classe $\mathcal{C}^1$ sur U.

Proposition 13 1. f est de classe $\mathcal{C}^1$ sur $U \underset{\not\Leftarrow}{\Longrightarrow} f$ est différentiable sur U.

2. f est $\mathcal{C}^1$ sur U si, et seulement si, f est différentiable sur U et sa différentielle df est continue sur U.

Preuve — D'après la formule de Taylor & Young à l'ordre 1 (théorème 4), si f est de classe $\mathcal{C}^1$ en a, alors f est différentiable en a. La récipoque est fausse, comme le prouve l'exercice suivant. La deuxième partie de la proposition est admise. □

Exercice 14 (Une fonction différentiable mais pas $\mathcal{C}^1$) — *Montrer que la fonction f définie sur $\mathbb{R}^2$ par*

$$f(x,y) = x^2 \sin\frac{y}{x} \quad \text{si} \quad (x,y) \neq (0,0) \qquad \text{et} \qquad f(0,y) = 0$$

est différentiable sur $\mathbb{R}^2$ mais n'est pas de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$.

Exemple 15 — *1. La norme « deux » définie sur $\mathbb{R}^p$ par $N(x) = \sqrt{x_1^2 + \cdots + x_p^2}$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^p \setminus \{0_{\mathbb{R}^p}\}$ et*

$$\forall a \in \mathbb{R}^p \setminus \{0_{\mathbb{R}^p}\},\ \forall h \in \mathbb{R}^p, \qquad dN(a)\cdot h = \frac{\langle a,h\rangle}{N(a)},$$

où $\langle a,h\rangle = \displaystyle\sum_{i=1}^p a_i h_i$ est le produit scalaire des vecteurs a et h.

Preuve — *Soit un vecteur non nul $a \in \mathbb{R}^p$: $\partial_i N(a) = \dfrac{2a_i}{2\sqrt{a_1^2+\cdots+a_p^2}} = \dfrac{a_i}{N(a)}$ d'où la fonction $\partial_i N$ est continue sur $\mathbb{R}^p \setminus \{0_{\mathbb{R}^p}\}$, donc N est de classe $\mathcal{C}^1$ sur $\mathbb{R}^p \setminus \{0_{\mathbb{R}^p}\}$ et $dN(a)\cdot h = \displaystyle\sum_{i=1}^p \frac{a_i}{N(a)} h_i$.* □

2. La fonction déterminant

$$\det\ :\ \mathcal{M}_n(\mathbb{R}) \to \mathbb{R},\ A \mapsto \det(A)$$

est de classe $\mathcal{C}^1$ et

$$\forall A \in \mathcal{M}_n(\mathbb{R}),\ \forall H \in \mathcal{M}_n(\mathbb{R}), \quad \det(A+H) = \det(A) + \mathrm{tr}({}^tBH) + o(H),$$

où B est la comatrice de A.

Preuve — *Le déterminant d'une matrice $A = (a_{ij})$ est une fonction des n^2 variables a_{ij}. On développe $\det(A)$ en suivant la $j-$ème colonne :*

$$\det(A) = (-1)^{1+j} a_{1j}\Delta_{1j} + \cdots + (-1)^{i+j} a_{ij}\Delta_{ij} + \cdots + (-1)^{n+j} a_{nj}\Delta_{nj},$$

où Δ_{ij} est le mineur de a_{ij}. D'où $\dfrac{\partial \det(A)}{\partial a_{ij}} = (-1)^{i+j}\Delta_{ij} = b_{ij}$. La matrice $B = (b_{ij})$ est la comatrice de A. Ses coefficients b_{ij} sont des fonctions continues des variables a_{ij}, d'où $\det$ est de classe $\mathcal{C}^1$, donc $\det$ est différentiable et

$$\forall A \in \mathcal{M}_n(\mathbb{R}), \quad d\det(A)\ :\ H \mapsto \sum_{i=1}^p \sum_{j=1}^p b_{ij} h_{ij} = \mathrm{tr}(\,{}^tBH).$$ □

Remarque 16 — *Du dernier exemple, il résulte que la fonction $\det\ :\ \mathcal{M}_n(\mathbb{R}) \to \mathbb{R}$ est continue (car $\mathcal{C}^1 \implies$ différentiable $\implies \mathcal{C}^0$). D'où $GL_n(\mathbb{R}) = \det^{-1}(\mathbb{R}^*)$ est l'image réciproque, par la fonction continue $\det$, de l'ouvert $\mathbb{R}^*$, donc l'ensemble $GL_n(\mathbb{R})$ des matrices inversibles est un ouvert de $\mathcal{M}_n(\mathbb{R})$. On avait aussi montré (exercice 26 du chapitre XI) que $GL_n(\mathbb{R})$ est dense dans $\mathcal{M}_n(\mathbb{R})$.*

XV.3 Le gradient

L'espace vectoriel E était jusqu'ici normé et de dimension finie. On le munit désormais d'un produit scalaire : E est donc un espace euclidien.

PROPOSITION-DÉFINITION 17
Si $f : E \to \mathbb{R}$ est différentiable en $a \in E$, alors il existe un unique vecteur de E, appelé le **gradient** de f en a et noté $\nabla f(a)$ ou $\operatorname{grad} f(a)$, tel que :

$$\forall h \in E, \quad df(a) \cdot h = \langle h, \nabla f(a) \rangle$$

est le produit scalaire du vecteur déplacement h et du gradient de f en a. Si $(e_1, \cdots, e_p)$ est une base orthonormée de E, alors

$$\nabla f(a) = \partial_1 f(a) e_1 + \cdots + \partial_p f(a) e_p.$$

Preuve — La différentielle de f en a est une application linéaire de E vers $\mathbb{R}$, c'est-à-dire une forme linéaire. Or E est supposé euclidien, d'où l'existence et l'unicité de $\nabla f(a)$, d'après le théorème de représentation de Riesz (théorème 29 du chapitre VIII). De plus, d'après la proposition 7, pour tout $h \in E$, $df(a) \cdot h = h_1 \partial_1 f(a) + \cdots + h_p \partial_p f(a) = \sum_{j=1}^{p} h_i \partial_i f(a)$, qui est bien égal à $\langle h_1 e_1 + \cdots + h_p e_p, \partial_1 f(a) e_1 + \cdots + \partial_p f(a) e_p \rangle$ si la base $(e_1, \cdots, e_p)$ de E est orthonormée. □

EXEMPLE 18 — *La fonction f définie par $f(x,y) = \ln(x^2+y^2)$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2 \setminus \{(0,0)\}$ et*

$$\forall (x,y) \neq (0,0), \quad \nabla f(x,y) = \left(\frac{2x}{x^2+y^2}, \frac{2y}{x^2+y^2} \right)$$

La figure XV.1 représente en chaque point (x,y) différent de l'origine le gradient de f en (x,y). On obtient ainsi un champ de vecteurs.

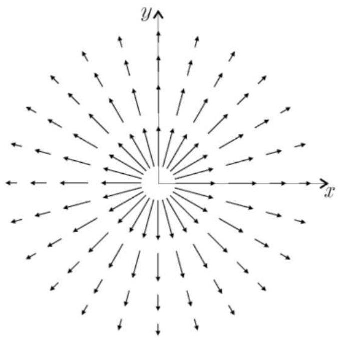

FIGURE XV.1 – LE CHAMP DES GRADIENTS DE $(x,y) \mapsto \ln(x^2+y^2)$

On se déplace dans un espace euclidien E, le long d'une courbe paramétrée par $M : t \mapsto M(t)$ et on évalue, à chaque instant t, la valeur $f(M(t))$ prise par une fonction scalaire f en le point $M(t)$.

LEMME 19 (**Règle de la chaîne**)
Soient deux fonctions $f : U \to \mathbb{R}$, $x \mapsto f(x)$ définie sur un ouvert U de l'espace euclidien E (ici de dimension 2, de même en dimension p) et $M : I \to E$, $t \mapsto M(t)$ définie sur un intervalle I de $\mathbb{R}$. Si $M(I) \subset U$ et les fontions f et M sont différentiables, alors la fonction $f \circ M : I \to \mathbb{R}$, $t \mapsto f(M(t))$ est différentiable et, pour tout $t \in I$,

$$\begin{aligned}(f \circ M)'(t) = \frac{d}{dt} f(M(t)) = \frac{d}{dt} f(x(t), y(t)) &= x'(t) \frac{\partial f}{\partial x}(M(t)) + y'(t) \frac{\partial f}{\partial y}(M(t)) \\ &= \langle M'(t), \nabla f(M(t)) \rangle \\ &= df(M(t)) \cdot M'(t)\end{aligned}$$

est le produit scalaire du vecteur-vitesse $M'(t)$ et du gradient de f en $M(t)$.

De plus, si f et M sont de classe $\mathcal{C}^1$, alors $f \circ M$ l'est aussi.

Preuve — Pour tout $t \in I$,

$$x(t+u) = x(t) + ux'(t) + u\varepsilon_1(u) \qquad \text{et} \qquad y(t+u) = y(t) + uy'(t) + u\varepsilon_2(u)$$

car x et y sont dérivables sur I. Pour tout $(a,b) \in U$,

$$f(a+h, b+k) = f(a,b) + h\partial_1 f(a,b) + k\partial_2 f(a,b) + \|(h,k)\|\varepsilon(h,k)$$

car f est différentiable sur D.

D'où $f \circ M(t+u) = f \circ M(t) \;+\; u\big[x'(t)\partial_1 f(x(t), y(t)) + y'(t)\partial_2 f(x(t), y(t))\big] + u\varepsilon_3(u)$.

Par suite $\dfrac{f \circ M(t+u) - f \circ M(t)}{u} \underset{u \to 0}{\longrightarrow} x'(t)\partial_1 f(x(t), y(t)) + y'(t)\partial_2 f(x(t), y(t))$.

Donc $f \circ M$ est dérivable. Et cette dérivée est continue si les fonctions x', y', $\partial_1 f$ et $\partial_2 f$ sont continues. □

On en déduit que : $(f \circ M)'(t) = 0$ si, et seulement si, le vecteur vitesse $M'(t)$ est orthogonal au gradient de f en $M(t)$ à l'instant $t \in I$.

En particulier, si la fonction f est constante le long de la trajectoire, *i.e.* si la trajectoire $M(I)$ est incluse dans une courbe de niveau de f, alors le gradient de f est orthogonal au vecteur-vitesse en chaque point de la trajectoire.

> Définition 20
> On dit qu'un vecteur $v \in E$ est **tangent à une partie** $C \subset E$ en un point $a \in C$ s'il existe une application dérivable $M : I \to E$ telle que $M(I) \subset C$ et s'il existe $t_0 \in I$ tel que $M(t_0) = a$ et $M'(t_0) = v$. On note $T_a C$ l'ensemble des vecteurs tangents à C en a.

On dit qu'un vecteur est orthogonal à la partie C s'il est orthogonal à $T_a C$.

> Théorème 21 (admis)
> Soient une fonction $f : E \to \mathbb{R}$ et une partie $C \subset E$ telle que la fonction f est constante sur C. Soit $a \in C$: si f est de classe $\mathcal{C}^1$ et si $\nabla f(a) \neq 0_E$, alors un vecteur $v \in E$ est tangent à C si, et seulement si, v est orthogonal à $\nabla f(a)$.
>
> Autrement dit : $T_a C$ est l'hyperplan $[\mathrm{Vect} \nabla f(a)]^{\perp} = \mathrm{Ker}\, df(a)$.

Le gradient de f en a est donc orthogonal à C : $\nabla f(a) \perp T_a C$. Ainsi le champ électrostatique est orthogonal aux équipotentielles, la ligne de plus grande pente est orthogonale aux lignes de niveau, *etc.*

Exercice 22 — *Déterminer une équation de la droite tangente à l'hyperbole d'équation $x^2 - y^2 = 1$ au point $(\sqrt{2}, 1)$.*

Exemple 23 — *La sphère $S \subset \mathbb{R}^3$ de rayon 1 et de centre $(0,0,0)$ est la surface paramétrée par*

$$\forall (\varphi, \theta) \in \mathbb{R}^2, \quad \begin{cases} x(\varphi,\theta) = \cos\theta\, \cos\varphi \\ y(\varphi,\theta) = \cos\theta\, \sin\varphi \\ z(\varphi,\theta) = \sin\theta \end{cases}.$$

Les vecteurs $\frac{\partial M}{\partial \varphi}(\varphi,\theta) = \begin{pmatrix} -\cos\theta \sin\varphi \\ -\cos\theta\cos\varphi \\ 0 \end{pmatrix}$ *et* $\frac{\partial M}{\partial \theta}(\varphi,\theta) = \begin{pmatrix} -\sin\theta\cos\varphi \\ \sin\theta\sin\varphi \\ \cos\theta \end{pmatrix}$ *sont tangents à la sphère en le point $M(\varphi,\theta)$. Le plan tangent à la sphère au point (a,b,c) est orthogonal au vecteur*

$$\frac{\partial M}{\partial \varphi}(\varphi,\theta) \wedge \frac{\partial M}{\partial \theta}(\varphi,\theta) = \begin{pmatrix} \cos^2\theta\cos\varphi \\ \cos^2\theta\sin\varphi \\ \cos\theta\sin\theta \end{pmatrix} = \cos\theta \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

*qui est non nul si $\cos\theta \neq 0$ (*i.e. *si le point $M(\varphi,\theta) = (a,b,c)$ n'est ni au pôle Sud ni au pôle Nord).*

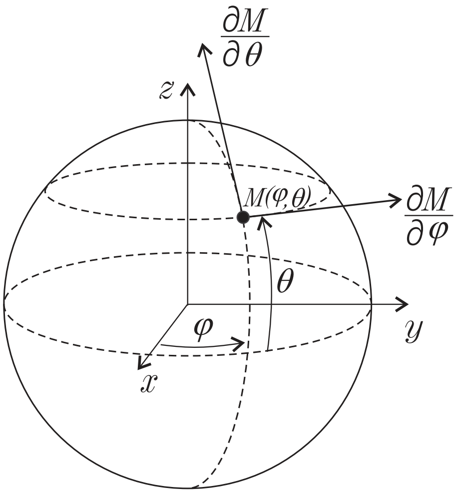

FIGURE XV.2 – Une sphère paramétrée

La même sphère S est la surface de niveau 1 de la fonction $f : \mathbb{R}^3 \to \mathbb{R},\ (x, y, z) \mapsto x^2 + y^2 + z^2$. En chaque point (a, b, c) de la sphère, le plan tangent a pour équation $ax + by + cz = 1$.

Preuve — *Soient $A = (a, b, c)$ un point de la sphère :*

$$\begin{aligned} \forall M = (x,y,z) \in \mathbb{R}^3, \quad \overrightarrow{AM} \perp \nabla f(A) \quad &\iff \quad \langle \overrightarrow{AM}, \nabla f(A)\rangle = 0 \\ &\iff \quad \begin{pmatrix} x-a \\ y-b \\ z-c \end{pmatrix} \cdot \begin{pmatrix} 2a \\ 2b \\ 2c \end{pmatrix} = 0 \\ &\iff \quad ax + by + cz = a^2 + b^2 + c^2. \end{aligned}$$

Or $a^2 + b^2 + c^2 = 1$ car le point A appartient à la sphère. □

DÉFINITION 24
On dit qu'une partie D :

- d'un espace vectoriel est **convexe** si, pour tout $(a, b) \in D^2$, $\forall t \in [0, 1],\ a+t(b-a) = (1-t)a+tb \in D$;
- d'un *evn* E est **connexe par arcs** si, pour tout $(a, b) \in D^2$, il existe une application continue $M : [0, 1] \to E,\ t \mapsto M(t)$ telle que $M(0) = a,\ M(1) = b$ et $\forall t \in [0, 1],\ M(t) \in D$.

REMARQUE 25 — *1. Le **segment** $[a, b]$ est l'ensemble des vecteurs $M(t) = a + t(b - a) = (1 - t)a + tb$, où $t \in [0, 1]$. Par suite : une partie D est convexe si, et seulement si, pour tout $(a, b) \in D^2$, $[a, b] \subset D$.*

2. *Toute partie convexe est connexe par arcs. La réciproque est fausse ; mais vraie dans $\mathbb{R}$: une partie de $\mathbb{R}$ est un intervalle si, et seulement si, elle est convexe si, et seulement si, elle est connexe par arcs.*
3. *L'image directe $f(D)$ d'une partie D connexe par arcs de E par une application continue $f : E \to F$ est une partie connexe par arcs de F.*

   Preuve — *Soit x' et y' dans $f(D)$. Il existe alors $(x, y) \in D^2$ tel que $x' = f(x)$ et $y' = f(y)$. Or D est connexe, il existe donc une fonction continue $M : [0, 1] \to E$ telle que $M(0) = x$ et $M(1) = y$ et $M([0, 1]) \subset D$. Par suite $f \circ M : [0, 1] \to F$ est une application continue telle que $f \circ M(0) = x'$, $f \circ M(1) = y'$ et $f \circ M([0, 1]) \subset f(D)$.* □
4. *Des deux remarques précédentes, il résulte que toute fonction continue sur une partie connexe par arcs vérifie le théorème des valeurs intermédiaires.*

EXERCICE 26 — *1. Montrer que toute boule de tout evn est convexe.*

2. *Montrer que l'hyperbole d'équation $y^2 - x^2 = 1$ n'est pas connexe par arcs.*
3. *Montrer que le produit cartésien $D_1 \times D_2$ de deux parties connexes par arcs $D_1 \subset E_1$ et $D_2 \subset E_2$ est une partie connexe par arcs de $E_1 \times E_2$. En déduire que la sphère $\{x \in \mathbb{R}^3 \mid \|x\|_2 = 1\}$ est une partie connexe par arcs de $\mathbb{R}^3$ et que, pour aller d'un pôle à l'autre, il faut traverser l'équateur.*

Proposition 27
Soit f une fonction de classe $\mathcal{C}^1$ sur un convexe D. La fonction f est constante sur D si, et seulement si, son gradient est nul sur D. De même si D est connexe par arcs.

Preuve — Si la fonction f est constante, alors toutes ses dérivées partielles (qui existent car f est $\mathcal{C}^1$) sont nulles, donc le gradient est nul en tout point de D. Réciproquement : on le prouve si D est convexe et on l'admet si D est connexe par arcs. Soient deux points a et b de D et, pour chaque $t \in [0,1]$ le point $M(t) = a + t(b-a)$. La partie D est convexe, d'où le segment $[a,b] = M([0,1])$ est inclus dans D. D'après la règle de la chaîne (lemme 19), la fonction $f \circ M$ est de classe $\mathcal{C}^1$ sur $[0,1]$, d'où :

$$f(b) - f(a) = f \circ M(1) - f \circ M(0) = \int_0^1 (f \circ M)'(t)\, dt = \int_0^1 \langle M'(t), \nabla f(M(t)) \rangle\, dt.$$

Si le gradient est nul en tout point de D, alors $f(a) = f(b)$ car $f(b) - f(a)$ est l'intégrale d'une fonction nulle. C'est vrai pour tout $(a,b) \in D^2$, donc la fonction f est constante sur D. □

XV.4 La différentielle d'une composée

Proposition 28
Soient trois espaces vectoriels normés E, F et G. Soient deux ouverts $U \subset E$ et $V \subset F$. Soient une fonction $f : U \to F$ et une fonction $g : V \to G$ telles que $f(U) \subset V$:

$$\begin{array}{ccccc} E \supset U & \longrightarrow & F \supset V & \longrightarrow & G \\ x & \longmapsto & y = f(x) & \longmapsto & g(y) = (g \circ f)(x) \end{array}$$

Si f est différentiable en $a \in U$ et g est différentiable en $b = f(a)$, alors $g \circ f$ est différentiable en a et

$$d(g \circ f)(a) = dg(b) \circ df(a).$$

Preuve — Soit $b = f(a)$: $g \circ f(a+h) = g(b+k)$, où $k = f(a+h) - f(a)$.

Or f est différentiable en a : $f(a+h) = f(a) + df(a) \cdot h + \|h\| \varepsilon_1(h)$. D'où $k = df(a) \cdot h + \|h\| \varepsilon_1(h)$.

Et g est différentiable en b : $g(b+k) = g(b) + dg(b) \cdot k + \|k\| \varepsilon_2(k)$. D'où $g \circ f(a+h) = g \circ f(a) + dg(b) \cdot k + \|k\| \varepsilon_2(k)$.

D'une part, $dg(b) \cdot k = dg(b) \cdot [df(a) \cdot h + \|h\| \varepsilon_1(h)] = dg(b) \circ df(a) \cdot h + \|h\| dg(b) \cdot \varepsilon_1(h)$ car $dg(b)$ est linéaire.

D'autre part, $\|k\| \varepsilon_2(k) = \|df(a) \cdot h + \|h\| \varepsilon_1(h)\|\ \varepsilon_2(k) = \|h\| \cdot \left\| df(a) \cdot \frac{h}{\|h\|} + \varepsilon_1(h) \right\| \varepsilon_2(k)$.

Donc $g \circ f(a+h) = g \circ f(a) + dg(b) \circ df(a) \cdot h + \|h\| \varepsilon(h)$ et $\varepsilon(h) = dg(b) \cdot \varepsilon_1(h) + \left\| df(a) \cdot \frac{h}{\|h\|} + \varepsilon_1(h) \right\| \varepsilon_2(k)$ tend vers 0_G quand h tend vers 0_E. En effet, $df(a)$ et $dg(b)$ sont des applications linéaires sur des *ev* de dimensions finies et sont donc continues. Par suite, $df(a)$ est bornée sur la sphère unité (théorème 37 du chapitre XI) et $dg(b) \cdot \varepsilon_1(h) \longrightarrow 0_G$. □

D'après la définition 6, la jacobienne $J_f(a)$ d'une fonction $f : E \to F$ différentiable en $a \in E$ est la matrice (dans des bases de E et de F) de la différentielle $df(a)$ de f en a.

Corollaire 29
$J_{g \circ f}(a) = J_g(b) \cdot J_f(a)$, où $b = f(a)$. Autrement dit, coordonnée par coordonnée :

$$\forall j \in [\![1,p]\!], \quad \frac{\partial (g \circ f)}{\partial x_j}(a) = \sum_{k=1}^{q} \frac{\partial g}{\partial y_k}(b) \cdot \frac{\partial f_k}{\partial x_j}(a).$$

Preuve — L'égalité des applications linéaires $d(g \circ f)(a) = dg(b) \circ df(a)$ déduite du théorème s'écrit matriciellement $J_{g \circ f}(a) = J_g(b) \cdot J_f(a)$:

$$\begin{pmatrix} \partial_1 (g_1 \circ f)(a) & \cdots & \partial_p (g_1 \circ f)(a) \\ \vdots & & \vdots \\ \partial_1 (g_n \circ f)(a) & \cdots & \partial_p (g_n \circ f)(a) \end{pmatrix} = \begin{pmatrix} \partial_1 g_1(b) & \cdots & \partial_q g_1(b) \\ \vdots & & \vdots \\ \partial_1 g_n(b) & \cdots & \partial_q g_n(b) \end{pmatrix} \begin{pmatrix} \partial_1 f_1(a) & \cdots & \partial_p f_1(a) \\ \vdots & & \vdots \\ \partial_1 f_q(a) & \cdots & \partial_p f_q(a) \end{pmatrix}.$$

Autre preuve : pour chaque $j \in [\![1,p]\!]$, on applique la règle de la chaîne (lemme 19) à la fonction $x_j \mapsto g \circ f(x)$, comme dans l'exemple suivant. □

Exemple 30 (**Le gradient en coordonnées polaires**) — *Soit une fonction* $f : \mathbb{R}^2 \to \mathbb{R},\ (x, y) \mapsto f(x, y)$ *différentiable. La fonction* F *définie par*

$$\forall r > 0,\ \forall \varphi \in \mathbb{R}, \quad F(r, \varphi) = f(r \cos \varphi, r \sin \varphi)$$

s'écrit aussi $F = f \circ M$*, où* M *est la fonction*

$$M : \mathbb{R}_+^* \times \mathbb{R} \to \mathbb{R}^2,\ (r, \varphi) \mapsto (x, y) = (r \cos \varphi, r \sin \varphi)$$

qui change les coordonnées polaires (r, φ) *en coordonnées cartésiennes* (x, y)*. Cette fonction* M *possède des dérivées partielles et sa matrice jacobienne est*

$$J_M(r, \varphi) = \begin{pmatrix} \frac{\partial x}{\partial r}(r, \varphi) & \frac{\partial x}{\partial \varphi}(r, \varphi) \\ \frac{\partial y}{\partial r}(r, \varphi) & \frac{\partial y}{\partial \varphi}(r, \varphi) \end{pmatrix} = \begin{array}{c} \begin{array}{cc} \frac{\partial M}{\partial r} & \frac{\partial M}{\partial \varphi} \\ \downarrow & \downarrow \end{array} \\ \begin{pmatrix} \cos \varphi & -r \sin \varphi \\ \sin \varphi & +r \cos \varphi \end{pmatrix} \end{array}.$$

Ses deux vecteurs colonnes sont représentés sur la figure XV.3 :

- *le vecteur* $\frac{\partial M}{\partial r}$ *est tangent à la droite paramétrée par* $r \mapsto M(r, \varphi)$ *;*
- *le vecteur* $\frac{\partial M}{\partial \varphi}$ *est tangent au cercle paramétré par* $\varphi \mapsto M(r, \varphi)$*.*

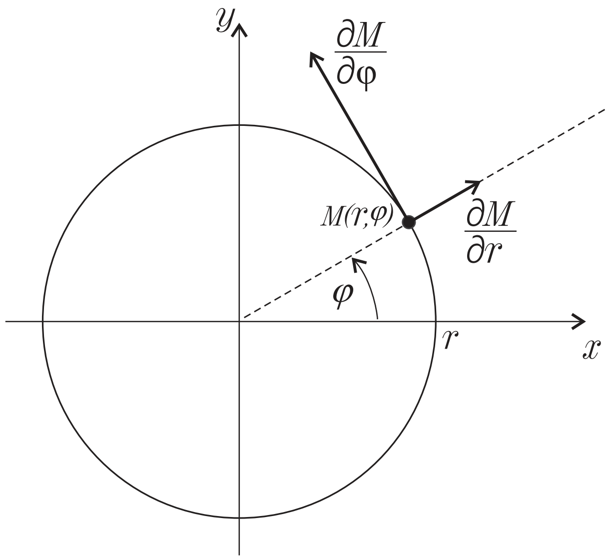

Figure XV.3 – Coordonnées polaires

La fonction F *est différentiable car les fonctions* f *et* M *le sont (*f *par hypothèse et* M *parce qu'elle est* $\mathcal{C}^1$*). D'où*

$$\begin{cases} \frac{\partial F}{\partial r}(r, \varphi) = \cos \varphi \frac{\partial f}{\partial x}(x, y) + \sin \varphi \frac{\partial f}{\partial y}(x, y) \\ \frac{\partial F}{\partial \varphi}(r, \varphi) = -r \sin \varphi \frac{\partial f}{\partial x}(x, y) + r \cos \varphi \frac{\partial f}{\partial y}(x, y) \end{cases} \iff \begin{cases} \frac{\partial f}{\partial x}(x, y) = \cos \varphi \frac{\partial F}{\partial r}(r, \varphi) - \frac{\sin \varphi}{r} \frac{\partial F}{\partial \varphi}(r, \varphi) \\ \frac{\partial f}{\partial y}(x, y) = \sin \varphi \frac{\partial F}{\partial r}(r, \varphi) + \frac{\cos \varphi}{r} \frac{\partial F}{\partial \varphi}(r, \varphi) \end{cases}.$$

On en déduit le gradient en coordonnées polaires :

$$\nabla f(x, y) = \frac{\partial F}{\partial r}(r, \varphi) \begin{pmatrix} \cos \varphi \\ \sin \varphi \end{pmatrix} + \frac{1}{r} \frac{\partial F}{\partial \varphi}(r, \varphi) \begin{pmatrix} -\sin \varphi \\ \cos \varphi \end{pmatrix}.$$

XV.5 Dérivées secondes

Définition 31
Soient $p \in \mathbb{N}^*$, un ouvert $D \subset \mathbb{R}^p$ et une fonction $f : D \to \mathbb{R}$, $(x_1, \cdots, x_p) \mapsto f(x_1, \cdots, x_p)$.

1. Soit un point $a = (a_1, \cdots, a_p) \in D$. Si la $i-$ème dérivée partielle $\partial_i f$ existe sur D et possède une $j-$ème dérivée partielle en a, alors le nombre réel $\partial_j(\partial_i f)(a)$ est une **dérivée partielle seconde** en a et est noté $\partial_j\partial_i f(a)$ ou $\frac{\partial^2 f}{\partial x_j \partial x_i}(a)$.
2. On dit que la fonction f est de classe $\mathcal{C}^2$ sur D si les p^2 fonctions $\partial_i\partial_j f$ sont continues sur D.

Exercice 32 — *Soit la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par*

$$f(x,y) = \frac{xy(x^2-y^2)}{x^2+y^2} \quad \text{si} \quad (x,y) \neq (0,0) \quad \text{et} \quad f(0,0) = 0.$$

Montrer que $\frac{\partial^2 f}{\partial x \partial y}(0,0) = +1$ et $\frac{\partial^2 f}{\partial y \partial x}(0,0) = -1$.

On en déduit que cette fonction f n'est pas de classe $\mathcal{C}^2$ grâce au théorème suivant.

Proposition 33 (**Théorème de Schwarz**)
Soient $p \in \mathbb{N}^*$, un ouvert $D \subset \mathbb{R}^p$ et une fonction $f : D \to \mathbb{R}$. Si f est de classe $\mathcal{C}^2$ sur D, alors

$$\forall (i,j) \in [\![1,p]\!]^2,\ \forall a \in D, \qquad \partial_i\partial_j f(a) = \partial_j\partial_i f(a).$$

Preuve — Dans le cas $p = 2$ d'abord : soient $(a,b) \in D$ et $(h,k) \in \mathbb{R}^2$ tel que $(a+h, b+h) \in D$. On va calculer

$$\Delta(h,k) = f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)$$

de deux manières :

— Soit $u(x) = f(x,b+k) - f(x,b)$.
Alors $\Delta(h,k) = u(a+h) - u(a)$ d'où (théorème des accroissements finis) $\Delta(h,k) = hu'(a+\theta_1 h)$, avec $\theta_1 \in]0,1[$.
Or $u'(x) = \partial_1 f(x,b+k) - \partial_1 f(x,b)$, d'où $\Delta(h,k) = h\left[\partial_1 f(a+\theta_1 h, b+k) - \partial_1 f(a+\theta_1 h, b)\right]$. D'où (encore le théorème des accroissements finis)
$$\Delta(h,k) = hk\partial_2\partial_1 f(a+\theta_1 h, b+\theta_2 k),$$
avec $\theta_2 \in]0,1[$.

— De même, soit $v(y) = f(a+h,y) - f(a,y)$. Alors
$$\Delta(h,k) = v(b+k) - v(b) = kh\partial_1\partial_2 f(a+\theta_4 h, b+\theta_3 k).$$

— En comparant les deux expressions de $\Delta(h,k)$, on obtient
$$\partial_2\partial_1 f(a+\theta_1 h, b+\theta_2 k) = \partial_1\partial_2 f(a+\theta_4 h, b+\theta_3 k)$$
puis, en faisant tendre (h,k) vers $(0,0)$,
$$\partial_2\partial_1 f(a,b) = \partial_1\partial_2 f(a,b)$$
car les dérivées secondes sont continues. Si $p > 2$, alors on raisonne de même avec la fonction
$$(x,y) \mapsto f(a_1, \cdots, a_{i-1}, x, a_{i+1}, \cdots, a_{j-1}, y, a_{j+1}, \cdots, a_p).$$

□

Théorème 34 (**Formule de Taylor & Young à l'ordre 2**, admise)
Soient $p \in \mathbb{N}^*$, un ouvert $D \subset \mathbb{R}^p$ et une fonction scalaire $f : D \to \mathbb{R}$. Si la fonction f est de classe $\mathcal{C}^2$, alors : en tout point $a = (a_1, \cdots, a_p) \in D$,

$$f(a+h) = f(a) + \underbrace{\sum_{i=1}^{p} h_i \partial_i f(a)}_{\ell_a(h)} + \underbrace{\frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} h_i h_j \partial_i\partial_j f(a)}_{q_a(h)} + \|h\|^2 \varepsilon(h),$$

où $h = (h_1, \cdots, h_p)$ et $\varepsilon(h) \underset{h\to 0}{\longrightarrow} 0$.

On sait déjà réécrire la forme linéaire ℓ_a sous les formes : $\ell_a(h) = df(a) \cdot h = \langle \nabla f(a), h \rangle$. On peut aussi réécrire la forme quadratique q_a sous la forme matricielle $q_a(h) = h^T \cdot H_f(a) \cdot h$ en notant h le vecteur-colonne des coordonnées du vecteur h et en utilisant la définition suivante.

DÉFINITION 35
Soit f une fonction admettant des dérivées partielles secondes a. La **hessienne** de f en a est la matrice

$$H_f(a) = (\partial_i \partial_j f(a))_{i,j} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_p}(a) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_p \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_p^2}(a) \end{pmatrix} \in \mathcal{M}_{pp}(\mathbb{R}).$$

D'après le théorème de Schwarz, si f est de classe $\mathcal{C}^2$, alors la hessienne est une matrice symétrique. En particulier, en dimension $p = 2$:

$$f(a+h, b+k) = f(a,b) + h\frac{\partial f}{\partial x}(a,b) + k\frac{\partial f}{\partial y}(a,b) + \frac{1}{2}\underbrace{\left[h^2\frac{\partial^2 f}{\partial x^2}(a,b) + 2hk\frac{\partial^2 f}{\partial x \partial y}(a,b) + k^2\frac{\partial^2 f}{\partial y^2}(a,b)\right]}_{q_a(h,k)} + (h^2+k^2)\varepsilon(h,k)$$

et la forme quadratique q_a peut s'écrire matriciellement :

$$q_a(h,k) = \begin{pmatrix} h & k \end{pmatrix} \underbrace{\begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(a,b) & \frac{\partial^2 f}{\partial x \partial y}(a,b) \\ \frac{\partial^2 f}{\partial x \partial y}(a,b) & \frac{\partial^2 f}{\partial y^2}(a,b) \end{pmatrix}}_{H_f(a)} \begin{pmatrix} h \\ k \end{pmatrix}.$$

XV.6 OPTIMISATION

DÉFINITION 36
Soient une partie $D \subset E$, un point $a \in D$ et une fonction scalaire $f : D \to \mathbb{R}$. On dit que f possède :
1. un minimum global en a si $\quad \forall x \in D, \quad f(x) \geq f(a)$;
2. un minimum local en a si $\quad \exists \varepsilon > 0,\ \forall x \in D, \quad \|x - a\| \leq \varepsilon \implies f(x) \geq f(a)$.

On définit de même un maximum local en a et un maximum global en a. On dit que f possède un extremum (local ou global) en a si f possède un maximum ou un minimum (local ou global) en a. Un extremum global est *a fortiori* local.

THÉORÈME 37
Soit $f : E \to \mathbb{R}$. Si D est une partie fermée et bornée de E et si f est une fonction continue sur D, alors f possède un maximum et un minimum globaux. Autrement dit : toute fonction réelle continue sur un fermé borné est bornée et atteint ses bornes.

Preuve — Voir l'annexe B. □

PROPOSITION-DÉFINITION 38 (Une condition **nécessaire** d'extremum local en un point **intérieur**)
Soient $D \subset \mathbb{R}^p$, un point a intérieur à D et une fonction scalaire $f : D \to \mathbb{R}$ différentiable en a.

(i) On dit que a est un **point critique** de f si $\nabla f(a) = 0_{\mathbb{R}^p}$ (*i.e.* si $df(a)$ est l'application nulle) ;

(ii) f possède un extremum local en le point intérieur $a \underset{\not\Longleftarrow}{\Longrightarrow} a$ est un point critique de f.

Preuve — Soit, pour chaque $i \in [\![1, p]\!]$, la fonction

$$f_i : t \mapsto f(a_1, \cdots, a_{i-1}, t, a_{i+1}, \cdots, a_p).$$

L'ensemble D est un ouvert, d'où il existe $\varepsilon > 0$ tel que la fonction f_i est définie sur $]a_i - \varepsilon, a_i + \varepsilon[\subset \mathbb{R}$. C'est une fonction d'une variable, dérivable en a_i et elle possède un extremum local en a_i, donc sa dérivée est nulle en a_i : $f_i'(a_i) = 0$. Or

$f_i'(a_i) = \partial_i f(a)$. La réciproque est fausse : 0 est un point critique de la fonction $\mathbb{R} \to \mathbb{R}$, $x \mapsto x^3$ qui n'a pourtant pas d'extremum. □

La proposition est fausse si le point n'est pas intérieur, comme le prouve l'exercice 41.

Pour confirmer l'existence d'un extremum local (autrement dit : énoncer une condition suffisante d'extremum) et étudier sa nature (est-ce un minimum ou un maximum ?) on utilise la formule de Taylor & Young à l'ordre 2, sous l'hypothèse que f est de classe $\mathcal{C}^2$. La hessienne $H_f(a)$ est alors une matrice symétrique d'après le théorème de Schwarz, donc diagonalisable en base orthonormée d'après le théorème spectral. Il existe donc des valeurs propres $(\lambda_1, \ldots, \lambda_p) \in \mathbb{R}^p$ (non nécessairement distinctes) telles que $q_a(h) = \sum_{i=1}^{p} \lambda_i h_i'^2$, en notant $(h_1', \cdots, h_p') \in \mathbb{R}^p$ les coordonnées de h dans la nouvelle base orthonormée.

Proposition 39

Soient un ouvert $D \subset \mathbb{R}^p$, un point $a \in D$ et une fonction $f : D \to \mathbb{R}$ de classe $\mathcal{C}^2$. Soient $(\lambda_1, \ldots, \lambda_p)$ les valeurs propres de la hessienne de f en a.

1. (Condition **nécessaire** de minimum local en un point **intérieur**) Si f possède un minimum local en a, alors
$$\nabla f(a) = 0 \quad \text{et} \quad \forall i \in [\![1, p]\!],\ \lambda_i \geq 0 \text{ (autrement dit : } H_f(a) \in \mathcal{S}_n^+(\mathbb{R})\text{)} ;$$
2. (Condition **suffisante** de minimum local en un point **intérieur**) Si
$$\nabla f(a) = 0 \quad \text{et} \quad \forall i \in [\![1, p]\!],\ \lambda_i > 0 \text{ (autrement dit : } H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R})\text{)},$$
alors f possède un minimum local en a.

De même pour un maximum local : il suffit de remplacer f par $-f$ et donc chaque λ_i par $-\lambda_i$ (autrement dit, $H_f(a)$ par $-H_f(a)$).

Preuve —

1. D'après la proposition 38, le gradient de f est nul en a. Par suite,
$$f(a+h) = f(a) + q(h) + \|h\|_2^2 \varepsilon(h) = f(a) + \sum_{i=1}^{p} \lambda_i h_i'^2 + \|h'\|_2^2 \varepsilon(h').$$
S'il existe $i \in [\![1, p]\!]$ tel que $\lambda_i < 0$, alors : pour tout $h' = \begin{pmatrix} h_1' \\ \vdots \\ h_p' \end{pmatrix}$ tel que $h_j' = 0$ pour tout $j \neq i$,
$$f(a+h) = f(a) + \lambda_i h_i'^2 + h_i'^2 \varepsilon(h_i') = f(a) + h_i'^2 \left(\lambda_i + \varepsilon(h_i')\right).$$
Si h_i' est suffisamment petit et non nul, alors $\lambda_i + \varepsilon(h_i') < 0$, d'où $f(a+h) < f(a)$, donc il n'y a pas de minimum en a.
2. Si le gradient de f en a est nul, alors
$$f(a+h) = f(a) + \sum_{i=1}^{p} \lambda_i h_i'^2 + \|h'\|_2^2 \varepsilon(h').$$
Si tous les λ_i sont strictement positifs, alors, en notant $\lambda = \min\{\lambda_1, \cdots, \lambda_p\}$:
$$f(a+h) - f(a) = \sum_{i=1}^{p} \lambda_i h_i'^2 + \|h'\|_2^2 \varepsilon(h') \geq \|h'\|_2^2 \left[\lambda + \varepsilon(h')\right]$$
est positif si h est suffisamment petit. Donc il y a un minimum local en a.

□

En dimension $p = 2$, pour connaître les signes des valeurs propres λ_1 et λ_2 de la hessienne, il suffit de calculer le déterminant et la trace de la hessienne :
$$\det H_f(a) = \lambda_1 \cdot \lambda_2 \qquad \text{et} \qquad \mathrm{tr} H_f(a) = \lambda_1 + \lambda_2.$$

Méthode 40 (Comment déterminer les extrema locaux en dimension 2) —

(i) On détermine les points critiques de la fonction f.

(ii) Pour chaque point critique a, on calcule le déterminant et la trace de la hessienne $H_f(a)$.

(a) Si le déterminant est strictement positif, alors il y a un extremum local en a :

— *si la trace est strictement positive, alors f admet un minimum local en a ;*

— *si la trace est strictement négative, alors f admet un maximum local en a.*

(b) Si le déterminant est strictement négatif, alors f n'admet pas d'extremum local en a.

(c) Si le déterminant est nul, alors la propositon ne permet pas de conclure.

Exercice 41 — *Etudier les extrema de la fonction*

$$f : \bar{B}(0,1) \to \mathbb{R},\ (x,y) \mapsto x^2 - y^2$$

définie sur la boule $\bar{B}(0,1) = \{(x,y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$ fermée de centre $(0,0)$ et de rayon 1.

On termine avec l'**optimisation sous contrainte**, aussi appelée problème des **extrema liés** : rendre extrémale une valeur $g(x_1, \cdots, x_p)$ sous une contrainte $f(x_1, \cdots, x_p) = \text{cte}$.

Proposition 42 (Théorème des extrema liés : c'est une condition nécessaire)
Soient f et g deux fonctions de classe $\mathcal{C}^1$ de E vers $\mathbb{R}$. Si f est constante sur une partie $C \subset E$, si la restriction de g à C admet un extremum local en $x \in C$ et si x n'est pas un point critique de f, alors

$$\exists \lambda \in \mathbb{R},\ \nabla g(x) = \lambda \nabla f(x).$$

Preuve — D'une part, d'après le théorème 21, le vecteur $\nabla f(x)$ est orthogonal à l'hyperplan T_xC car la fonction f est constante sur C, est de classe $\mathcal{C}^1$ et $\nabla f(x) \neq 0_E$.

D'autre part, le vecteur $\nabla g(x)$ est aussi orthogonal à l'hyperplan T_xC. En effet, si un vecteur v est tangent à C, alors il existe un arc paramétré $M : I \to E,\ t \mapsto M(t)$ tel que $M(I) \subset C$ et un instant $t_0 \in I$ tel que $M(t_0) = x$ et $M'(t_0) = v$ d'après la définition 20. Or $g|_C \circ M = g \circ M$ possède un extremum local en t_0, d'où $0 = (g \circ M)'(t_0) = \langle \nabla g(x), v \rangle$ d'après la régle de la chaîne (lemme 19).

Enfin, $\nabla g(x)$ est peut-être nul mais $\nabla f(x)$ ne l'est pas, par hypothèse. Donc $\exists \lambda \in \mathbb{R},\ \nabla g(x) = \lambda \nabla f(x)$. □

Exemple 43 (figure XV.4) — *Soit C une courbe de niveau d'une fonction $f : \mathbb{R}^2 \to \mathbb{R},\ (x,y) \mapsto f(x,y)$ de classe $\mathcal{C}^1$. Soit A un point du plan. Si la distance AM du point A à un point non critique $M \in C$ posséde un extremum local, alors le vecteur $\overrightarrow{AM}$ est orthogonal à la courbe C.*

Preuve — *Soit K la valeur prise par f sur la courbe C. Soient (a,b) et (x,y) les coordonnées respectives des points A fixé et M parcourant C. Si la valeur $g(x,y) = (x-a)^2 + (y-b)^2$ est extrémale sous la contrainte $f(x,y) = K$, alors (d'après le théorème des extrema liés) : il existe $\lambda \in \mathbb{R}$ tel que $\nabla g(x,y) = 2(x-a, y-b) = 2\overrightarrow{AM} = \lambda \nabla f(x,y)$ si $\nabla f(x,y) \neq (0,0)$. Or $\nabla f(x,y)$ est orthogonal à la courbe C, d'après le théorème 21.* □

Exercice 44 (figure XV.5) — *Déterminer les points du cercle d'équation $x^2+y^2 = 5$ qui rendent extrémale la valeur $2x + y$.*

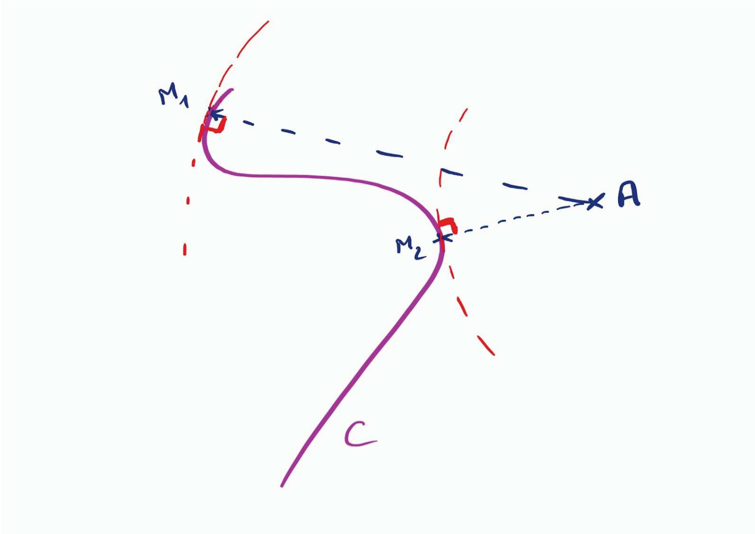

Figure XV.4 – Distances extrémales d'un point à une courbe

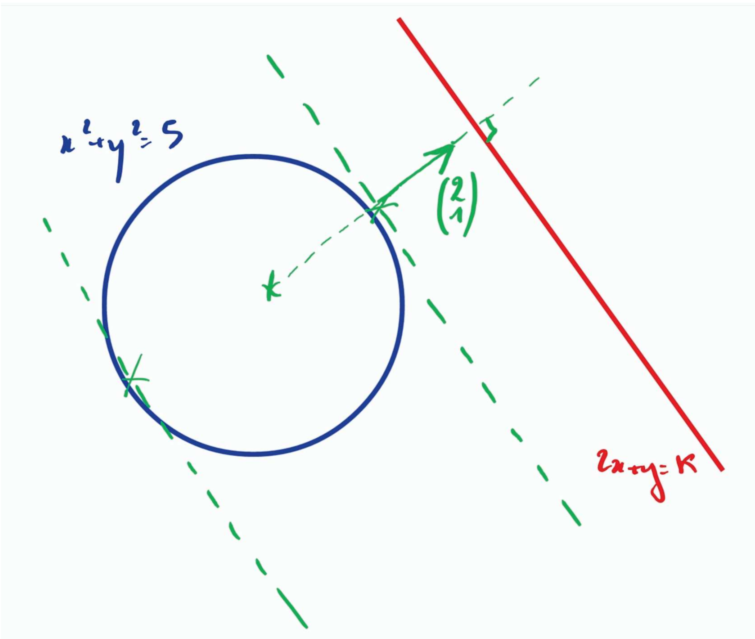

Figure XV.5 – Un cercle et des droites