CALCUL DIFFÉRENTIEL

I. Fonctions de classe $\mathcal{C}^1$.

Dans tout ce paragraphe, U est un ouvert de $\mathbb{R}^p$, et f est une application de U vers $\mathbb{R}$. L'espace vectoriel $\mathbb{R}^p$ sera toujours muni de sa norme euclidienne canonique notée $\|\cdot\|$.

1. Dérivée selon un vecteur.

Soit a un point de U, soit v un vecteur de $\mathbb{R}^p$. Comme U est ouvert, il existe un réel $\delta > 0$ tel que $\forall t \in [-\delta, \delta] \quad a + tv \in U$.

En effet, le point a est intérieur à U donc il existe $r > 0$ tel que la boule fermée $\overline{B}(a,r) = \big\{x \in \mathbb{R}^p \;\big|\; \|x-a\| \leq r\big\}$ soit incluse dans U. On a donc $a + tv \in U$ dès que $|t| \leq \dfrac{r}{\|v\|}$ (si v est non nul, et s'il est nul c'est évident!).

L'application $\varphi_v : t \mapsto f(a+tv)$ est donc définie au moins sur l'intervalle $[-\delta, \delta]$.

Définition. Si l'application $\varphi_v : t \mapsto f(a+tv)$ est dérivable en 0, on dit que f admet une **dérivée au point a selon le vecteur v**, et on pose

$$D_v f(a) = \varphi_v'(0)\ .$$

Ainsi, sous réserve d'existence,

$$D_v f(a) = \varphi_v'(0) = \left[\frac{\mathrm{d}}{\mathrm{d}t}\big(f(a+tv)\big)\right]_{t=0} = \lim_{t\to 0} \frac{f(a+tv) - f(a)}{t}\ .$$

Interprétation. Lorsque t décrit $\mathbb{R}$, le point $m = a + tv$ décrit, dans $\mathbb{R}^p$, la droite affine passant par a et de vecteur directeur v (si ce vecteur est non nul). La dérivée de f en a selon v donne une idée de comment varie $f(m)$ lorsque le point m s'éloigne du point a dans la direction du vecteur v. Si cette dérivée directionnelle existe, on a en effet le développement limité à l'ordre un:

$$\varphi_v(t) = f(a+tv) = f(a) + D_v f(a)\cdot t + o(t)\ .$$

2. Dérivées partielles.

Définition 1. Soit $a = (a_1, \cdots, a_p)$ un point de U, soit $i \in [\![1, p]\!]$. L'application

$$\varphi_i : x \mapsto \ \varphi_i(x) = f(a_1, \cdots, a_{i-1}, x, a_{i+1}, \cdots, a_p)$$

est la i-ième **application partielle** de f au point a.

Commentaire. Cette application est définie sur une partie ouverte (*on l'admettra*) de $\mathbb{R}$, à savoir l'ensemble des réels x tels que le p-uplet $(a_1, \cdots, a_{i-1}, x, a_{i+1}, \cdots, a_p)$ appartienne à U. On a, en particulier, $\varphi_i(a_i) = f(a)$.

Définition 2. Avec les notations introduites ci-dessus, si cette application partielle φ_i est dérivable au point a_i, on dit que f admet au point a une **dérivée partielle d'ordre 1** par rapport à la i-ième variable. On note alors

$$\partial_i f(a) = \varphi_i'(a_i)\ , \quad \text{ou encore} \quad \frac{\partial f}{\partial x_i}(a) = \varphi_i'(a_i)\ .$$

Commentaire. Les dérivées partielles sont donc les dérivées des applications partielles (qui sont, elles, des fonctions "ordinaires" d'une variable réelle à valeurs réelles).

Commentaire. En revenant à la définition de la dérivée d'une fonction d'une variable, et en faisant éventuellement une translation de la variable, on a ainsi, sous réserve d'existence,

$$\begin{aligned}
\partial_i f(a) &= \left[\frac{\mathrm{d}}{\mathrm{d}x}\, f(a_1, \cdots, a_{i-1}, x, a_{i+1}, \cdots, a_p)\right]_{x=a_i} \\
&= \lim_{t\to 0} \frac{f(a_1, \cdots, a_{i-1}, a_i + t, a_{i+1}, \cdots, a_p) - f(a_1, \cdots, a_{i-1}, a_i, a_{i+1}, \cdots, a_p)}{t}
\end{aligned}$$

$$= \lim_{t\to 0} \frac{f(a+t\,e_i)-f(a)}{t} = D_{e_i}f(a)\ ,$$

en notant $(e_1,\cdots,e_p)$ la base canonique de $\mathbb{R}^p$. Autrement dit, cette dérivée partielle de f en a par rapport à la i-ème variable est aussi la dérivée de f en a selon le i-ème vecteur e_i de la base canonique de $\mathbb{R}^p$.

Remarque. L'existence de dérivées partielles n'entraîne pas la continuité. Soit en effet la fonction $f:\mathbb{R}^2\to\mathbb{R}$ définie par $f(x,y)=\dfrac{xy}{x^2+y^2}$ si $(x,y)\neq(0,0)$ et $f(0,0)=0$. Cette fonction est discontinue à l'origine puisque, pour x non nul, $f(x,x)=\dfrac{1}{2}\neq f(0,0)=0$. Elle admet pourtant des dérivées partielles en tout point du plan: en effet, par opérations algébriques (produit, quotient avec dénominateur non nul), on obtient, pour $(x,y)\neq(0,0)$, $\dfrac{\partial f}{\partial x}(x,y)=\dfrac{y(y^2-x^2)}{(x^2+y^2)^2}$ et $\dfrac{\partial f}{\partial y}(x,y)=\dfrac{x(x^2-y^2)}{(x^2+y^2)^2}$. Par ailleurs, les applications partielles $x\mapsto f(x,0)$ et $y\mapsto f(0,y)$ sont nulles sur $\mathbb{R}$ tout entier, d'où l'existence de dérivées partielles à l'origine avec $\dfrac{\partial f}{\partial x}(0,0)=\dfrac{\partial f}{\partial y}(0,0)=0$.

3. Fonctions de classe $\mathcal{C}^1$.

Définition. Soit U un ouvert de $\mathbb{R}^p$. Une fonction $f:U\to\mathbb{R}$ est dite **de classe $\mathcal{C}^1$** sur U si elle admet, en tout point de U, des dérivées partielles d'ordre 1, et si les p applications dérivées partielles $\partial_1 f,\ \cdots,\ \partial_p f$ sont continues sur U.

Opérations. On montre sans difficulté que, si $f:U\to\mathbb{R}$ et $g:U\to\mathbb{R}$ sont de classe $\mathcal{C}^1$ sur U, alors il en est de même de $\alpha f+g$, où α est un réel. L'ensemble $\mathcal{C}^1(U,\mathbb{R})$ des fonctions de classe $\mathcal{C}^1$ sur U est alors muni d'une structure de $\mathbb{R}$-espace vectoriel. On a, de plus, pour tout $i\in[\![1,p]\!]$, la relation

$$\partial_i(\alpha f+g)=\alpha\,\partial_i f+\partial_i g\ .$$

Avec les mêmes hypothèses, fg est de classe $\mathcal{C}^1$ sur U avec $\partial_i(fg)=f\cdot\partial_i g+g\cdot\partial_i f$, ou encore avec d'autres notations, $\dfrac{\partial(fg)}{\partial x_i}(a)=\dfrac{\partial f}{\partial x_i}(a)\cdot g(a)+f(a)\cdot\dfrac{\partial g}{\partial x_i}(a)$.

4. Développements limités d'ordre 1.

Définition. Soit $f:U\to\mathbb{R}$ avec U un ouvert de $\mathbb{R}^p$, soit $a\in U$. On dit que f admet un **développement limité d'ordre 1** au point a (en abrégé un $\mathrm{DL}_1(a)$) s'il existe une forme linéaire λ sur $\mathbb{R}^p$ telle que

$$f(a+h)=f(a)+\lambda(h)+o\big(\|h\|\big)$$

lorsque $h\in\mathbb{R}^p$ tend vers 0.

On peut écrire cela sous la forme

$$f(a+h)=f(a)+\lambda(h)+h\,\varepsilon(h)\ ,\quad \text{avec}\quad \lim_{h\to 0}\varepsilon(h)=0\ .$$

L'existence d'un $\mathrm{DL}_1(a)$ correspond à la notion de **différentiabilité** (*terme hors programme*) de la fonction f au point a.

Un exemple élémentaire. Soit $f:\mathbb{R}^2\to\mathbb{R}$ définie par $f(x,y)=xy^2-y$ (c'est une fonction "polynomiale" à deux variables), soit $a=(1,3)\in\mathbb{R}^2$. Par un simple calcul algébrique, on peut obtenir un $\mathrm{DL}_1(a)$ de la fonction f. Soit un vecteur "petit déplacement" $h=(u,v)\in\mathbb{R}^2$, alors

$$\begin{aligned} f(a+h) &= f(1+u,3+v) \\ &= (1+u)(3+v)^2-(3+v) \\ &= 6+9u+5v+6uv+v^2+uv^2 \\ &= 6+(9u+5v)+o\big(\|h\|\big)\ . \end{aligned}$$

En effet, on vérifie facilement que, lorsque $h=(u,v)\to(0,0)$, on a $uv=o\big(\|h\|\big)$ et $uv^2=o\big(\|h\|\big)$, par exemple en choisissant $\|h\|=\|h\|_\infty=\max\big\{|u|,|v|\big\}$ (mais peu importe puisqu'en dimension finie, les normes sont toutes équivalentes). L'application $\lambda: h=(u,v)\mapsto 9u+5v$ est une forme linéaire sur $\mathbb{R}^2$, canoniquement représentée par la matrice-ligne $L=\begin{pmatrix}9 & 5\end{pmatrix}$. Enfin, $f(a)=6$. On a donc obtenu un développement limité à l'ordre un de f au voisinage du point $a=(1,3)$. On peut remarquer que les coefficients de ce développement limité sont les dérivées partielles premières de f au point a, en effet $\dfrac{\partial f}{\partial x}(a)=9$ et $\dfrac{\partial f}{\partial y}(a)=5$.

<u>Propriété d'unicité.</u> Si f admet un développement limité d'ordre 1 en un point a de U, la forme linéaire λ apparaissant dans l'écriture de ce développement est unique.

Preuve. Supposons que $f(a+h)=f(a)+\lambda(h)+o\big(\|h\|\big)$ et $f(a+h)=f(a)+\mu(h)+o\big(\|h\|\big)$ lorsque h tend vers 0 dans $\mathbb{R}^p$, où λ et μ sont deux formes linéaires sur $\mathbb{R}^p$. Par différence, en posant $\alpha=\lambda-\mu$ (c'est une nouvelle forme linéaire sur $\mathbb{R}^p$), on a $\alpha(h)=o\big(\|h\|\big)$ lorsque $h\to 0$, ce qui signifie que $\displaystyle\lim_{h\to 0}\frac{\big|\alpha(h)\big|}{\|h\|}=0$. Si x est un vecteur non nul de $\mathbb{R}^p$ et t un réel, comme $\displaystyle\lim_{t\to 0}tx=0$, par composition de limites, on aura $\displaystyle\lim_{t\to 0}\frac{\big|\alpha(tx)\big|}{\|tx\|}=0$, autrement dit $\dfrac{|t|\,\big|\alpha(x)\big|}{|t|\,\|x\|}=\dfrac{\big|\alpha(x)\big|}{\|x\|}\underset{t\to 0}{\longrightarrow}0$, soit $\alpha(x)=0$. Donc α est la forme linéaire nulle, donc $\lambda=\mu$.

Une forme linéaire λ sur $\mathbb{R}^p$ est canoniquement représentée par une matrice-ligne $L=\begin{pmatrix}\alpha_1 & \cdots & \alpha_p\end{pmatrix}\in\mathcal{M}_{1,p}(\mathbb{R})$, c'est alors l'application de $\mathbb{R}^p$ vers $\mathbb{R}$ telle que

$$\forall h=(h_1,\cdots,h_p)\in\mathbb{R}^p\qquad \lambda(h)=\alpha_1h_1+\cdots+\alpha_ph_p=\sum_{i=1}^p\alpha_ih_i\ .$$

L'existence d'un développement limité d'odre 1 au point $a=(a_1,\cdots,a_p)\in U$ pour une fonction $f:U\to\mathbb{R}$ équivaut donc à l'existence de p réels α_1, $\cdots$, α_p tels que

$$f(a+h)=f(a_1+h_1,\cdots,a_p+h_p)=f(a)+\sum_{i=1}^p\alpha_ih_i+o\big(\|h\|\big)$$

lorsque $h=(h_1,\cdots,h_p)\to(0,\cdots,0)$ dans $\mathbb{R}^p$. Les réels α_i (alors uniques) sont les **coefficients** de ce développement limité.

5. <u>Différentielle en un point d'une fonction de classe $\mathcal{C}^1$</u>.

On admettra le théorème suivant:

<u>Théorème.</u> Toute fonction de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^p$ admet, en tout point a de U, un développement limité d'ordre 1, dont les coefficients sont les dérivées partielles d'ordre 1 de f au point a, i.e.

$$f(a+h) = f(a) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(a)\, h_i + o\big(\|h\|\big)$$

lorsque $h = (h_1, \cdots, h_p) \to (0, \cdots, 0)$ dans $\mathbb{R}^p$.

<u>Définition.</u> La "partie linéaire" de ce développement limité, i.e. la forme linéaire sur $\mathbb{R}^p$ définie par

$$h = (h_1, \cdots, h_p) \mapsto \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(a)\, h_i = \sum_{i=1}^{p} h_i\, \partial_i f(a)$$

est appelée **différentielle** de f en a, et notée $\mathrm{d}f(a)$.

<u>Notation.</u> Si f est de classe $\mathcal{C}^1$, si $a \in U$, si $h = (h_1, \cdots, h_p) \in \mathbb{R}^p$, le réel $\displaystyle\sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(a)\, h_i$, qu'il serait logique de noter $\mathrm{d}f(a)(h)$, est couramment noté $\mathrm{d}f(a) \cdot h$. Ainsi, le $\mathrm{DL}_1(a)$ d'une fonction f de classe $\mathcal{C}^1$ se mettra sous la forme

$$\textbf{(*)} \qquad f(a+h) = f(a) + \mathrm{d}f(a) \cdot h + o\big(\|h\|\big)\ .$$

<u>Remarque.</u> Le nombre $\mathrm{d}f(a) \cdot h = \displaystyle\sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(a)\, h_i$ est aussi la dérivée de f au point a selon le vecteur h, que nous avions notée précédemment $D_h f(a)$. En effet, il résulte de **(*)** que, si le vecteur h est fixé, l'application de variable réelle $\varphi_h : t \mapsto f(a+th)$ admet en 0 un développement limité d'ordre un, qui est $\varphi_h(t) = f(a) + t\, \mathrm{d}f(a) \cdot h + o(t)$, ce qui montre que φ_h est dérivable en 0, avec par définition $D_h f(a) = \varphi_h'(0) = \mathrm{d}f(a) \cdot h$.

<u>Conséquence.</u> Toute fonction de classe $\mathcal{C}^1$ sur U est continue sur U.

Preuve. En effet, du $\mathrm{DL}_1(a)$ ci-dessus, il résulte immédiatement (notamment en utilisant la continuité des applications linéaires en dimension finie) que $\lim_{h \to 0} f(a+h) = f(a)$.

<u>Remarque. Approximation du premier ordre.</u> Si f est de classe $\mathcal{C}^1$ sur U, le vecteur $h = (h_1, \cdots, h_p)$ intervenant dans l'écriture du $\mathrm{DL}_1(a)$ de f est souvent interprété comme un vecteur "petit déplacement", on le note parfois $\mathrm{d}a = (\mathrm{d}x_1, \cdots, \mathrm{d}x_p)$. Le $\mathrm{DL}_1(a)$ s'écrit alors

$$f(a + \mathrm{d}a) = f(a) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(a)\, \mathrm{d}x_i + o\big(\|\mathrm{d}a\|\big)\ .$$

On dit parfois que, "dans une approximation du premier ordre", on a

$$f(a + \mathrm{d}a) \simeq f(a) + \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}(a)\, \mathrm{d}x_i\ .$$

II. <u>Règle de la chaîne.</u>

1. <u>Dérivée le long d'un arc</u>.

Soit U un ouvert de $\mathbb{R}^p$, soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$. On peut interpréter f comme un **champ de scalaires** sur U.

Soit d'autre part I un intervalle de $\mathbb{R}$ et $x_1, \cdots, x_p$ des fonctions de classe $\mathcal{C}^1$ de I vers $\mathbb{R}$ telles que
$$\forall t \in I \qquad \big(x_1(t), \cdots, x_p(t)\big) \in U \ .$$

Si, pour $t \in I$, on pose $\varphi(t) = \big(x_1(t), \cdots, x_p(t)\big)$, alors la fonction vectorielle $\varphi : I \to U \subset \mathbb{R}^p$ peut être interprétée comme un **arc paramétré** de classe $\mathcal{C}^1$, à support dans U.

Pour tout $t \in I$, posons enfin $g(t) = f\big(\varphi(t)\big) = f\big(x_1(t), \cdots, x_p(t)\big)$. On a ainsi $g = f \circ \varphi$, qui est une application de I vers $\mathbb{R}$.

<u>Théorème.</u> Alors g est de classe $\mathcal{C}^1$ sur I et on a
$$\forall t \in I \qquad g'(t) = \sum_{i=1}^{p} x_i'(t)\ \partial_i f\big(x_1(t), \cdots, x_p(t)\big) = \sum_{i=1}^{p} \frac{\partial f}{\partial x_i}\big(\varphi(t)\big)\ x_i'(t)\ .$$

On peut réécrire cela sous la forme
$$\frac{\mathrm{d}}{\mathrm{d}t}\Big(f\big(x_1(t), \cdots, x_p(t)\big)\Big) = \sum_{i=1}^{p} x_i'(t)\ \frac{\partial f}{\partial x_i}\big(x_1(t), \cdots, x_p(t)\big)\ ,$$

ou encore
$$\boxed{\frac{\mathrm{d}}{\mathrm{d}t}\Big(f\big(\varphi(t)\big)\Big) = \mathrm{d}f\big(\varphi(t)\big) \cdot \varphi'(t)\ .}$$

Preuve. Pour $t \in I$ et h réel "petit" (en tous cas tel que $t + h \in I$), on a
$$g(t+h) = f\big(\varphi(t+h)\big) = f\big(\varphi(t) + h\,\varphi'(t) + h\,\varepsilon(h)\big) \quad \text{avec} \quad \lim_{h\to 0} \varepsilon(h) = 0\ ,$$
en utilisant le développement limité d'ordre 1 de la fonction vectorielle φ au point t. Mais f admet aussi un DL_1 au point $\varphi(t)$, qui s'écrit $f\big(\varphi(t)+k\big) = f\big(\varphi(t)\big) + \mathrm{d}f\big(\varphi(t)\big)\cdot k + o\big(\|k\|\big)$, où la lettre k représente ici un vecteur "petit déplacement" dans $\mathbb{R}^p$. Comme $\lim\limits_{h\to 0} \big(h\ \varphi'(t) + h\ \varepsilon(h)\big) = 0$, il est légitime de remplacer k par cette expression dans ce dernier DL, on a donc
$$\begin{aligned} g(t+h) &= f\big(\varphi(t) + h\,\varphi'(t) + h\,\varepsilon(h)\big) \\ &= f\big(\varphi(t)\big) + \mathrm{d}f\big(\varphi(t)\big) \cdot \big(h\,\varphi'(t) + h\,\varepsilon(h)\big) + o\big(\|k\|\big) \\ &= g(t) + h\ \mathrm{d}f\big(\varphi(t)\big) \cdot \varphi'(t) + \textit{des cacahuètes}\ . \end{aligned}$$

On a utilisé la linéarité de la différentielle $\mathrm{d}f\big(\varphi(t)\big)$ pour scinder le terme du milieu en deux termes. Le lecteur méticuleux s'assurera que les termes restants non explicités et relégués au rang de "cacahuètes" sont bien "négligeables dans une approximation du premier ordre", autrement dit qu'ils peuvent s'écrire sous la forme $o(h)$.

On vient de prouver que la fonction $g : I \to \mathbb{R}$ admet en tout point de l'intervalle I un développement limité à l'ordre 1, elle est donc dérivable sur I. Sa dérivée au point

$t \in I$ est le coefficient du terme du premier ordre, ce qui donne bien la formule attendue:

$$\forall t \in I \quad g'(t) = \mathrm{d}f\big(\varphi(t)\big) \cdot \varphi'(t) = \sum_{i=1}^{p} x_i'(t)\ \frac{\partial f}{\partial x_i}\big(\varphi(t)\big).$$

Enfin, comme f est de classe $\mathcal{C}^1$, ses dérivées partielles $\dfrac{\partial f}{\partial x_i}$ sont continues, on voit alors que g' est continue comme composée, produit et somme de fonctions continues. Donc g est de classe $\mathcal{C}^1$ sur I.

2. <u>Application aux dérivées partielles d'une fonction composée</u>.

Soient U un ouvert de $\mathbb{R}^n$, Ω un ouvert de $\mathbb{R}^p$, soit $X : U \to \Omega$ une application (un "changement de variables") de classe $\mathcal{C}^1$: pour tout $u = (u_1, \cdots, u_n) \in U$, posons

$$X(u) = \big(x_1(u_1, \cdots, u_n), \cdots, x_p(u_1, \cdots, u_n)\big)\ ,$$

dire que l'application X est de classe $\mathcal{C}^1$ signifie que ses p fonctions coordonnées $x_1, \cdots, x_p$ sont de classe $\mathcal{C}^1$ sur U. Soit d'autre part $f : \Omega \to \mathbb{R}$ une application de classe $\mathcal{C}^1$. Soit enfin l'application

$$g = f \circ X : \begin{cases} \quad U & \to\ \mathbb{R} \\ u = (u_1, \cdots, u_n) & \mapsto\ g(u) = f\big(x_1(u_1, \cdots, u_n), \cdots, x_p(u_1, \cdots, u_n)\big) \end{cases}.$$

Alors g est de classe $\mathcal{C}^1$ sur U, et ses dérivées partielles se calculent par les formules

$$\boxed{\forall i \in [\![1, n]\!] \qquad \frac{\partial g}{\partial u_i} = \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}\ \frac{\partial x_j}{\partial u_i}\ .}$$

La preuve est identique à celle du paragraphe précédent.

Notons que la relation encadrée est un peu abrégée: en effet, elle n'indique pas en quel(s) point(s) les différentes dérivées partielles sont évaluées. Cette formule, complètement explicitée, s'écrirait

$$\forall i \in [\![1, n]\!] \quad \forall u \in U \qquad \frac{\partial g}{\partial u_i}(u) = \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}\big(X(u)\big)\ \frac{\partial x_j}{\partial u_i}(u)\ .$$

Un cas particulier est à connaître, c'est celui du "**passage en coordonnées polaires**": soit $f : \mathbb{R}^2 \to \mathbb{R}$ de classe $\mathcal{C}^1$. Posons d'autre part $x(r,\theta) = r\ \cos\theta$ et $y(r,\theta) = r\ \sin\theta$ pour tout couple $(r,\theta) \in \mathbb{R}^2$. Soit enfin $g : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$\forall (r,\theta) \in \mathbb{R}^2 \qquad g(r,\theta) = f(r\ \cos\theta, r\ \sin\theta) = f\big(x(r,\theta), y(r,\theta)\big)\ .$$

Alors g est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$ et on a les relations

$$\frac{\partial g}{\partial r} = \frac{\partial f}{\partial x}\ \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\ \frac{\partial y}{\partial r} = (\cos\theta)\ \frac{\partial f}{\partial x} + (\sin\theta)\ \frac{\partial f}{\partial y}$$

$$\frac{\partial g}{\partial \theta} = \frac{\partial f}{\partial x}\ \frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\ \frac{\partial y}{\partial \theta} = (-r\ \sin\theta)\ \frac{\partial f}{\partial x} + (r\ \cos\theta)\ \frac{\partial f}{\partial y}\ .$$

3. Caractérisation des fonctions constantes.

Proposition. Soit U un ouvert convexe de $\mathbb{R}^p$, soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$. Alors f est constante sur U si et seulement si $\forall a \in U \quad \mathrm{d}f(a) = 0$.

La condition énoncée est que, en tout point a de U, la différentielle de f en a est la forme linéaire nulle sur $\mathbb{R}^p$. Comme cette différentielle s'exprime à l'aide des dérivées partielles d'ordre 1 de f, un énoncé équivalent est de dire (toujours si f est de classe $\mathcal{C}^1$ sur un ouvert convexe U) que f est constante sur U si et seulement si

$$\forall a \in U \quad \forall i \in [\![1, p]\!] \qquad \frac{\partial f}{\partial x_i}(a) = 0 \ .$$

Preuve. Le sens direct est évident: si f est constante sur U, toutes ses dérivées partielles sont nulles en tout point de U.

Réciproquement, si $\mathrm{d}f = 0$ sur U, soient $a \in U$ et $b \in U$, montrons que $f(a) = f(b)$. Comme U est convexe, le segment $[a, b]$ est inclus dans U, i.e. $\forall t \in [0, 1] \ \ (1 - t)a + tb \in U$. Posons $x(t) = (1 - t)a + tb$ pour tout $t \in [0, 1]$, alors la fonction vectorielle x est de classe $\mathcal{C}^1$ sur $[0, 1]$, à valeurs dans U, avec $x'(t) = b - a$ pour tout $t \in [0, 1]$. Soit maintenant la fonction composée $g = f \circ x$, à savoir $g : [0, 1] \to \mathbb{R}$ telle que $g(t) = f\big((1-t)a+tb\big)$ pour tout $t \in [0, 1]$. D'après la règle de la chaîne (dérivée le long d'un "arc" qui est ici un segment de droite), cette fonction g est de classe $\mathcal{C}^1$ sur $[0, 1]$ avec $\forall t \in [0, 1] \quad g'(t) = \mathrm{d}f\big(x(t)\big) \cdot x'(t) = 0$. Cette fonction g d'une variable, dérivable et de dérivée nulle, est donc constante sur $[0, 1]$, d'où $g(0) = g(1)$, i.e. $f(a) = f(b)$.

III. Gradient.

Définition. Soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$, où U est un ouvert de $\mathbb{R}^p$. On munit $\mathbb{R}^p$ de sa structure euclidienne canonique. Pour tout $a \in U$, la différentielle de f en a, notée $\mathrm{d}f(a)$, est une forme linéaire sur $\mathbb{R}^p$. D'après le théorème de représentation de Riesz, il existe alors un unique vecteur v de $\mathbb{R}^p$ tel que

$$\forall h \in \mathbb{R}^p \qquad \mathrm{d}f(a) \cdot h = (v|h) \ .$$

Ce vecteur v est appelé **gradient** de f en a, et noté $\nabla f(a)$.

On a ainsi $\boxed{\forall h \in \mathbb{R}^p \qquad \mathrm{d}f(a) \cdot h = \big(\nabla f(a) \bigm| h\big)\ .}$

Comme d'autre part, si $h = (h_1, \cdots, h_p) = \sum_{i=1}^{p} h_i e_i$, on a $\mathrm{d}f(a) \cdot h = \sum_{i=1}^{p} \partial_i f(a) h_i$, on déduit que les coordonnées du vecteur gradient $\nabla f(a)$, dans la base canonique $(e_1, \cdots, e_p)$ de $\mathbb{R}^p$, sont les dérivées partielles premières de f au point a, i.e.

$$\nabla f(a) = \sum_{i=1}^{p} \partial_i f(a)\, e_i = \big(\partial_1 f(a), \cdots, \partial_p f(a)\big) \ .$$

On peut alors réécrire différents résultats des paragraphes précédents:

- si $f : U \to \mathbb{R}$ est de classe $\mathcal{C}^1$, alors son développement limité d'ordre 1 en $a \in U$ s'écrit

$$f(a + h) = f(a) + \big(\nabla f(a) \bigm| h\big) + o\big(\|h\|\big) \ .$$

La dérivée de f au point a selon le vecteur h est donc aussi $D_h f(a) = \big(\nabla f(a) \bigm| h\big)$.

- "dérivée le long d'un arc": si, de plus, on se donne un arc paramétré de classe $\mathcal{C}^1$ à support dans U, i.e. une application $\varphi : I \to U$, $t \mapsto m = \varphi(t)$, de classe $\mathcal{C}^1$, et si on pose $g(t) = f\big(\varphi(t)\big)$ pour tout $t \in I$, alors g est de classe $\mathcal{C}^1$ sur l'intervalle I avec

$$\forall t \in I \qquad g'(t) = \Big(\nabla f\big(\varphi(t)\big) \mid \varphi'(t)\Big) = \Big(\nabla f(m) \mid \frac{\mathrm{d}m}{\mathrm{d}t}\Big) ,$$

i.e. $g'(t)$ est le produit scalaire du vecteur gradient de f au point $m = \varphi(t)$ avec le vecteur vitesse instantanée du point mobile m à l'instant t, dans une interprétation cinématique.

- si U est un ouvert convexe de $\mathbb{R}^p$ et si $f : U \to \mathbb{R}$ est de classe $\mathcal{C}^1$, alors f est constante sur U si et seulement si $\forall a \in U \quad \nabla f(a) = \overrightarrow{0}$.

Expression du gradient en coordonnées polaires.

Soit $f : \mathbb{R}^2 \to \mathbb{R}$ de classe $\mathcal{C}^1$. Posons $g(r,\theta) = f(r\,\cos\theta, r\,\sin\theta)$ pour $(r,\theta) \in \mathbb{R}^2$. Alors g est de classe $\mathcal{C}^1$ sur $\mathbb{R}^2$, et on a obtenu dans le paragraphe **II.2.** les relations

$$\begin{cases} \dfrac{\partial g}{\partial r} = (\cos\theta)\,\dfrac{\partial f}{\partial x} + (\sin\theta)\,\dfrac{\partial f}{\partial y} \\ \dfrac{\partial g}{\partial \theta} = (-r\,\sin\theta)\,\dfrac{\partial f}{\partial x} + (r\,\cos\theta)\,\dfrac{\partial f}{\partial y} \end{cases} .$$

En inversant ce système linéaire, dont le déterminant est $\begin{vmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{vmatrix} = r$, non nul si on se place en un point autre que l'origine, on obtient

$$\begin{cases} \dfrac{\partial f}{\partial x} = \cos\theta\,\dfrac{\partial g}{\partial r} - \dfrac{\sin\theta}{r}\,\dfrac{\partial g}{\partial \theta} \\ \dfrac{\partial f}{\partial y} = \sin\theta\,\dfrac{\partial g}{\partial r} + \dfrac{\cos\theta}{r}\,\dfrac{\partial g}{\partial \theta} \end{cases} .$$

On a alors, en notant (e_1, e_2) la base canonique de $\mathbb{R}^2$,

$$\begin{aligned} \nabla f &= \frac{\partial f}{\partial x}\,e_1 + \frac{\partial f}{\partial y}\,e_2 \\ &= \frac{\partial g}{\partial r}\,(\cos\theta\,e_1 + \sin\theta\,e_2) + \frac{1}{r}\,\frac{\partial g}{\partial \theta}\,(-\sin\theta\,e_1 + \cos\theta\,e_2) \\ &= \frac{\partial g}{\partial r}\,u_r + \frac{1}{r}\,\frac{\partial g}{\partial \theta}\,u_\theta , \end{aligned}$$

en notant (u_r, u_θ) la "base mobile" des coordonnées polaires, i.e. la base orthonormale directe obtenue par rotation d'angle θ à partir de la base canonique.

IV. Applications à la géométrie différentielle.

1. Courbes du plan.

Soit U un ouvert de $\mathbb{R}^2$, soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$. On peut alors définir la **courbe d'équation** $f(x,y) = 0$, soit $\Gamma = \big\{(x,y) \in U \mid f(x,y) = 0\big\} = f^{-1}\big(\{0\}\big)$. *On notera que, sans hypothèses supplémentaires, cette "courbe" n'aura pas forcément l'aspect d'une courbe dans le sens courant du terme. Par exemple, si f est une fonction constante, cette "courbe" est vide, ou bien est l'ouvert U tout entier.*

Un point $m_0 = (x_0, y_0)$ de Γ est dit **point régulier** si $\nabla f(m_0) \neq 0$, autrement dit s'il n'est pas un **point critique** de f.

Au voisinage d'un point régulier m_0, nous admettrons l'existence d'un paramétrage local de classe $\mathcal{C}^1$ de Γ, c'est-à-dire l'énoncé suivant: il existe $r > 0$ tel que $\Gamma \cap B(m_0, r)$ soit le support d'un arc paramétré de classe $\mathcal{C}^1$ injectif $\varphi : \begin{cases} I \to \mathbb{R}^2 \\ t \mapsto \varphi(t) = \big(x(t), y(t)\big) \end{cases}$ tel que $m_0 = \varphi(t_0)$ avec $t_0 \in I$ et $\varphi'(t_0) \neq 0$. On a alors $\forall t \in I \ \ f \circ \varphi(t) = f\big(x(t), y(t)\big) = 0$. En dérivant cette relation, on obtient

$$\forall t \in I \qquad \frac{\partial f}{\partial x}\big(x(t), y(t)\big) \cdot x'(t) + \frac{\partial f}{\partial y}\big(x(t), y(t)\big) \cdot y'(t) = 0 \ ,$$

soit, pour $t = t_0$, $\big(\nabla f(m_0) \mid \varphi'(t_0)\big) = 0$. Comme le vecteur $\varphi'(t_0)$ est non nul, il dirige la tangente à Γ en m_0, on voit que ce vecteur $\varphi'(t_0)$ est orthogonal au vecteur gradient $\nabla f(m_0)$, de coordonnées $\dfrac{\partial f}{\partial x}(m_0)$ et $\dfrac{\partial f}{\partial y}(m_0)$.

On en déduit l'**équation de la tangente $\mathcal{T}_{m_0}$ à la courbe Γ : $f(x,y) = 0$ en un point régulier** m_0: si T est un point du plan de coordonnées (x, y), alors

$$\begin{aligned} T \in \mathcal{T}_{m_0} \quad &\Longleftrightarrow \quad \big(\nabla f(m_0) \mid \overrightarrow{m_0 T}\big) = 0 \\ &\Longleftrightarrow \quad \frac{\partial f}{\partial x}(m_0) \cdot (x - x_0) + \frac{\partial f}{\partial y}(m_0) \cdot (y - y_0) = 0 \ . \end{aligned}$$

Plus généralement, soit λ un réel, soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$. L'ensemble

$$\Gamma_\lambda = f^{-1}\big(\{\lambda\}\big) = \big\{(x, y) \in U \mid f(x, y) = \lambda\big\}$$

est appelé **ligne de niveau** λ de l'application f. Ce qui précède s'applique toujours (en remplaçant la fonction f par $f - \lambda$, ce qui ne modifie pas ses dérivées partielles). On en déduit le résultat suivant:

Proposition. **En un point où il est non nul, le gradient de f est orthogonal aux lignes de niveau $f(x, y) = \lambda$.**

On notera par ailleurs que, **lorsqu'il est non nul, le gradient de f est orienté dans le sens des valeurs croissantes de f**. En effet, soit $m_0 \in U$ un point régulier, i.e. $\nabla f(m_0) \neq 0$, soit $h \in \mathbb{R}^2$ un vecteur unitaire, observons les variations de f lorsque l'on part du point m_0 dans la direction du vecteur h, i.e. plus précisément lorsqu'on se trouve en un point m de la forme $m = m_0 + th$, avec t réel positif petit. Posons donc $g(t) = f(m_0 + th)$, cela a un sens pour t réel assez proche de 0 (puisque U est ouvert, donc le point m_0 est intérieur à U, il existe donc $\delta > 0$ tel que $m_0 + th \in U$ pour t vérifiant $|t| < \delta$). Alors g est de classe $\mathcal{C}^1$ sur $]-\delta, \delta[$, avec $g'(t) = \big(\nabla f(m_0 + th) \mid h\big)$, en particulier $g'(0) = \big(\nabla f(m_0) \mid h\big)$. L'inégalité de Cauchy-Schwarz donne $\big|g'(0)\big| \leq \big\|\nabla f(m_0)\big\|$ (puisque h est unitaire), et on a $g'(0) = \big\|\nabla f(m_0)\big\|$ (valeur maximale) lorsque h est de même direction et de même sens que $\nabla f(m_0)$. C'est donc en partant du point m_0 dans la direction du vecteur-gradient $\nabla f(m_0)$ que l'on obtient la plus forte "croissance infinitésimale" de f.

2. Surfaces dans l'espace.

Soit U un ouvert de $\mathbb{R}^3$, soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$. On peut alors définir la **surface d'équation** $f(x,y,z)=0$, soit $\Sigma = \big\{(x,y,z)\in U \mid f(x,y,z)=0\big\} = f^{-1}\big(\{0\}\big)$.

Un point $m_0=(x_0,y_0,z_0)$ de Σ est dit **point régulier** si $\nabla f(m_0)\neq 0$, autrement dit s'il n'est pas un **point critique** de f.

Soit $\varphi : \begin{cases} I \to U \\ t \mapsto \varphi(t) = \big(x(t),y(t),z(t)\big) \end{cases}$ un arc paramétré de classe $\mathcal{C}^1$ à support dans U.

Si on a $\varphi(t)\in\Sigma$ pour tout $t\in I$, i.e. si

$$\forall t\in I \qquad f\big(x(t),y(t),z(t)\big)=0\ ,$$

on dit que cet arc est une **courbe tracée sur la surface** Σ.

Exemple. La sphère Σ de centre O et de rayon 1 est la surface d'équation $x^2+y^2+z^2=1$. Soit $\alpha\in]-\pi,\pi]$, alors l'arc $\varphi:[0,2\pi]\to\mathbb{R}^3$ défini par

$$\forall t\in[0,2\pi] \qquad \varphi(t)=(\cos\alpha\ \sin t\ ,\ \sin\alpha\ \sin t\ ,\ \cos t)$$

est une courbe tracée sur Σ, c'est ce qu'un géographe appellerait un "cercle méridien".

Reprenons l'étude générale. Si un arc paramétré $\varphi : I\to U$ de classe $\mathcal{C}^1$ est tracé sur la surface $\Sigma : f(x,y,z)=0$, alors en dérivant la relation $f\big(\varphi(t)\big)=0$, on obtient

$$\forall t\in I \qquad \Big(\nabla f\big(\varphi(t)\big)\mid \varphi'(t)\Big)=0\ .$$

On en déduit que, **si l'arc est régulier**, i.e. si $\forall t\in I\quad \varphi'(t)\neq 0$, alors la tangente à cet arc au point $m=\varphi(t)$ est orthogonale au vecteur gradient $\nabla f(m)$. Ceci motive la définition suivante:

Définition. Soit $f : U\to\mathbb{R}$ de classe $\mathcal{C}^1$ avec U ouvert de $\mathbb{R}^3$, soit $m_0=(x_0,y_0,z_0)$ un point régulier sur la surface $\Sigma : f(x,y,z)=0$. On appelle **plan tangent** à Σ au point m_0 le plan affine orthogonal au gradient $\nabla f(m_0)$ et passant par m_0. Ce plan $\mathcal{T}_{m_0}$ a donc pour équation

$$\frac{\partial f}{\partial x}(m_0)\cdot(x-x_0)+\frac{\partial f}{\partial y}(m_0)\cdot(y-y_0)+\frac{\partial f}{\partial z}(m_0)\cdot(z-z_0)=0\ .$$

Le plan tangent à Σ en m_0 contient donc toutes les droites tangentes en m_0 aux courbes régulières tracées sur Σ.

En un point régulier m_0 d'une surface $\Sigma : f(x,y,z)=0$, le vecteur $\nabla f(m_0)$ dirige donc la **normale** à la surface Σ en ce point m_0.

Exemple. Soit Σ la sphère d'équation $f(x,y,z)=0$ avec $f(x,y,z)=x^2+y^2+z^2-1$. En tout point $m_0=(x_0,y_0,z_0)$ de Σ, on a $\nabla f(m_0)=(2x_0,2y_0,2z_0)$, ce vecteur est non nul et il dirige la normale à la sphère en m_0. On constate donc (et cela ne doit pas être une surprise!) que la normale à Σ en un point m_0 est la droite $(O\,m_0)$, où l'origine O est le centre de la sphère.

V. Fonctions de classe $\mathcal{C}^2$.

Soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^1$, où U est un ouvert de $\mathbb{R}^p$. Soit $(i,j) \in [\![1,p]\!]^2$. Si la fonction dérivée partielle première $\partial_j f = \dfrac{\partial f}{\partial x_j}$ admet elle-même en un point a de U une dérivée partielle par rapport à la i-ième variable, on définit alors une **dérivée partielle seconde**

$$\partial^2_{i,j} f(a) = \partial_i(\partial_j f)\,(a) \qquad \text{ou} \qquad \frac{\partial^2 f}{\partial x_i\,\partial x_j}(a) = \frac{\partial}{\partial x_i}\Big(\frac{\partial f}{\partial x_j}\Big)\,(a)\ .$$

Définition. On dit que f est **de classe $\mathcal{C}^2$** sur U si elle admet en tout point de U des dérivées partielles d'ordre deux, et si toutes les fonctions $\partial^2_{i,j} f$, avec $(i,j) \in [\![1,p]\!]^2$, sont continues sur U.

Opérations. On admettra que, si $f : U \to \mathbb{R}$ et $g : U \to \mathbb{R}$ sont de classe $\mathcal{C}^2$ sur U, si α et β sont des réels, alors les fonctions $\alpha f + \beta g$, fg et $\dfrac{f}{g}$ (si g ne s'annule pas) sont de classe $\mathcal{C}^2$ sur U. En particulier, l'ensemble $\mathcal{C}^2(U,\mathbb{R})$ des fonctions de classe $\mathcal{C}^2$ de U vers $\mathbb{R}$ est un sous-espace vectoriel de $\mathcal{F}(U,\mathbb{R})$.

Théorème de Schwarz. Si $f : U \to \mathbb{R}$ est de classe $\mathcal{C}^2$ sur un ouvert U de $\mathbb{R}^p$, alors pour tout couple $(i,j) \in [\![1,p]\!]^2$, on a $\partial^2_{i,j} f = \partial^2_{j,i} f$, ou encore $\dfrac{\partial^2 f}{\partial x_i\,\partial x_j} = \dfrac{\partial^2 f}{\partial x_j\,\partial x_i}$.

On peut donc intervertir l'ordre des dérivations. *Ce théorème est admis.*

On peut chercher à résoudre des **équations aux dérivées partielles** (EDP) d'ordre un ou deux, un changement de variables étant donné. Ces "EDP" sont une généralisation des "équations différentielles ordinaires" (EDO), mais concernant des fonctions de plusieurs variables réelles. Aucun théorème sur ce sujet (notamment concernant l'existence ou l'unicité de solutions avec conditions "initiales" ou autres) n'est à connaître, il s'agit essentiellement de pratiquer, *cf.* donc la feuille d'exercices. Je propose toutefois deux exemples:

Exemple 1. On va résoudre, sur $\mathbb{R}_+^* \times \mathbb{R}$, l'équation aux dérivées partielles du premier ordre

$$x\,\frac{\partial f}{\partial x} + y\,\frac{\partial f}{\partial y} = \sqrt{x^2+y^2}\ .$$

Pour cela, on peut utiliser comme changement de variables le **passage en coordonnées polaires**, i.e. on pose $x = r\,\cos\theta$ et $y = r\,\sin\theta$. Lorsque les coordonnées cartésiennes (x,y) décrivent le demi-plan ouvert $U = \mathbb{R}_+^* \times \mathbb{R}$, on peut considérer que les coordonnées polaires (r,θ) décrivent la demi-bande ouverte $V = \mathbb{R}_+^* \times \Big]-\dfrac{\pi}{2}, \dfrac{\pi}{2}\Big[$. On posera donc $F(r,\theta) = f(r\cos\theta, r\sin\theta)$ pour $(r,\theta) \in V$. Si f est de classe $\mathcal{C}^1$ sur U, alors F est de classe $\mathcal{C}^1$ sur V. Un calcul classique par la règle de la chaîne donne

$$\frac{\partial F}{\partial r} = (\cos\theta)\,\frac{\partial f}{\partial x} + (\sin\theta)\,\frac{\partial f}{\partial y} = \frac{1}{r}\,\Big(x\,\frac{\partial f}{\partial x} + y\,\frac{\partial f}{\partial y}\Big) \qquad \text{et...}$$

... et cela suffit puisque nous retrouvons le premier membre de notre équation aux dérivées partielles qui s'écrit $\dfrac{\partial F}{\partial r} = 1$, et qui se résout en $F(r,\theta) = r + g(\theta)$, où $g : \Big]-\dfrac{\pi}{2}, \dfrac{\pi}{2}\Big[\to \mathbb{R}$ est une fonction arbitraire de classe $\mathcal{C}^1$. *L'équation* $\dfrac{\partial F}{\partial r} = 1$ *s'écrit aussi* $\dfrac{\partial}{\partial r}\big(F(r,\theta) - r\big) = 0$, *elle signifie donc que l'expression $F(r,\theta)-r$ ne dépend que de la variable θ.* En revenant aux coordonnées cartésiennes, on obtient l'expression $f(x,y)=\sqrt{x^2+y^2}+g\Big(\operatorname{Arctan}\dfrac{y}{x}\Big)$, ou $f(x,y)=\sqrt{x^2+y^2}+h\Big(\dfrac{y}{x}\Big)$, où $h:\mathbb{R}\to\mathbb{R}$ est une fonction arbitraire de classe $\mathcal{C}^1$.

__Exemple 2.__ Soit $c>0$. On va rechercher les fonctions $f:\mathbb{R}^2\to\mathbb{R}$, $(x,t)\mapsto f(x,t)$, de classe $\mathcal{C}^2$ vérifiant l'EDP du second ordre

$$c^2\,\frac{\partial^2 f}{\partial x^2}-\frac{\partial^2 f}{\partial t^2}=0$$

(**équation de d'Alembert**, ou **équation des cordes vibrantes**). Pour cela, le changement de variables utilisé sera une transformation affine: on posera $\begin{cases} u\ =\ x+ct\\ v\ =\ x-ct\end{cases}$.

Résolution: Posons $f(x,y)=g(u,v)$. La règle de la chaîne donne

$$\frac{\partial f}{\partial x}=\frac{\partial g}{\partial u}+\frac{\partial g}{\partial v}\qquad\text{et}\qquad \frac{\partial f}{\partial t}=c\left(\frac{\partial g}{\partial u}-\frac{\partial g}{\partial v}\right),$$

puis

$$\frac{\partial^2 f}{\partial x^2}=\frac{\partial^2 g}{\partial u^2}+2\,\frac{\partial^2 g}{\partial u\partial v}+\frac{\partial^2 g}{\partial v^2}$$

et

$$\frac{\partial^2 f}{\partial t^2}=c^2\left(\frac{\partial^2 g}{\partial u^2}-2\,\frac{\partial^2 g}{\partial u\partial v}+\frac{\partial^2 g}{\partial v^2}\right).$$

L'équation proposée devient $\dfrac{\partial^2 g}{\partial u\partial v}=0$, soit $\dfrac{\partial}{\partial u}\Big(\dfrac{\partial g}{\partial v}\Big)=0$ **(*)**, donc $\dfrac{\partial g}{\partial v}(u,v)=h(v)$ **(**)** où $h:\mathbb{R}\to\mathbb{R}$ est de classe $\mathcal{C}^1$. *En effet, l'équation* **(*)** *signifie que* $\dfrac{\partial g}{\partial v}(u,v)$ *dépend seulement de la variable v.* Puis $g(u,v)=k(u)+H(v)$, où k et H sont deux fonctions arbitraires de $\mathbb{R}$ vers $\mathbb{R}$, de classe $\mathcal{C}^2$. *En effet, en introduisant une primitive H de la fonction h, l'équation* **(**)** *s'écrit* $\dfrac{\partial}{\partial v}\big(g(u,v)-H(v)\big)=0$*, donc signifie que $g(u,v)-H(v)$ dépend seulement de la variable u.* Finalement,

$$f(x,y)=k(x+ct)+H(x-ct)\,.$$

__Définition.__ Soit $f:U\to\mathbb{R}$ de classe $\mathcal{C}^2$, où U est un ouvert de $\mathbb{R}^p$, soit a un point de U. On appelle **matrice hessienne** de f en a, et on note $H_f(a)$, la matrice carrée symétrique d'ordre p dont le coefficient d'indices (i,j) est la dérivée partielle d'ordre deux $\dfrac{\partial^2 f}{\partial x_i\,\partial x_j}(a)$. Ainsi,

$$H_f(a)=\left(\frac{\partial^2 f}{\partial x_i\,\partial x_j}(a)\right)_{1\le i,j\le p}\in\mathcal{S}_p(\mathbb{R})\,.$$

__Formule de Taylor-Young à l'ordre deux.__ **Si $f:U\to\mathbb{R}$ est de classe $\mathcal{C}^2$, alors f admet en tout point a de U un développement limité d'ordre deux, qui s'écrit sous la forme**

$$f(a+h)\underset{h\to 0}{=}f(a)+\nabla f(a)^\top h+\frac{1}{2}\,h^\top H_f(a)\,h+o\big(\|h\|^2\big)\,.$$

Ce résultat sera admis.

Commentaire. Dans ce développement limité, il faut comprendre que l'on identifie le vecteur $h = (h_1, \cdots, h_p)$ de $\mathbb{R}^p$ avec la matrice-colonne $\begin{pmatrix} h_1 \\ \vdots \\ h_p \end{pmatrix}$ de $\mathcal{M}_{p,1}(\mathbb{R})$. De la même façon, le vecteur $\nabla f(a) \in \mathbb{R}^p$ est identifié avec la matrice-colonne $\begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_p f(a) \end{pmatrix} \in \mathcal{M}_{p,1}(\mathbb{R})$.

Une écriture équivalente, utilisant des produits scalaires, est

$$f(a+h) \underset{h \to 0}{=} f(a) + \big(\nabla f(a) \mid h\big) + \frac{1}{2}\big(H_f(a) \cdot h \mid h\big) + o\big(\|h\|^2\big) .$$

VI. Extremums.

1. Définitions générales.

Soit A une partie de $\mathbb{R}^p$, soit $f : A \to \mathbb{R}$, soit $a \in A$.

On dit que f présente un **maximum global** au point a si $\forall x \in A \quad f(x) \leq f(a)$.

On dit que f présente un **maximum local** au point a si

$$\exists r > 0 \quad \forall x \in A \,\cap\, B(a,r) \qquad f(x) \leq f(a) ,$$

où $B(a,r)$ est la boule ouverte de centre a et de rayon r (pour une certaine norme sur $\mathbb{R}^p$).

2. Condition nécessaire de présence d'un extremum dans un ouvert

Définition. Soit $f : U \to \mathbb{R}$ une fonction de classe $\mathcal{C}^1$, où U est un ouvert de $\mathbb{R}^p$, soit $a \in U$. On dit que a est un **point critique** de f si $\nabla f(a) = 0$ (le vecteur gradient de f en a est nul).

On peut aussi dire qu'un point critique de f est un point a tel que $\mathrm{d}f(a) = 0$, i.e. la différentielle de f en a est la forme linéaire nulle.

Ou encore: a est un point critique de f si $\forall i \in [\![1,p]\!] \quad \partial_i f(a) = 0$ (toutes les dérivées partielles d'ordre 1 de f en a sont nulles).

On a alors le résultat suivant:

Théorème. Si une fonction f de classe $\mathcal{C}^1$ sur un ouvert U de $\mathbb{R}^p$ admet un extremum local en un point, alors celui-ci est un point critique de f.

Preuve. Supposons que f admette un maximum local au point $a = (a_1, \cdots, a_p)$ de U. Soit $i \in [\![1,p]\!]$, montrons que $\partial_i f(a) = 0$. Il existe $r > 0$ tel que la boule $B_2(a,r)$ de centre a et de rayon r soit incluse dans U (considérons par exemple la boule pour la norme euclidienne canonique) et tel que $\forall x \in U \,\cap\, B_2(a,r) \quad f(x) \leq f(a)$. En notant $(e_1, \cdots, e_p)$ la base canonique de $\mathbb{R}^p$, on a alors $f(a+te_i) \leq f(a)$ pour $t \in]-r,r[$. La fonction $\varphi_i : t \mapsto f(a+te_i)$ est dérivable sur $]-r,r[$, de dérivée $\varphi_i'(t) = \partial_i f(a+te_i)$, et elle admet en 0 un maximum local, donc $\varphi_i'(0) = 0$, soit $\partial_i f(a) = 0$.

La réciproque de ce résultat est fausse. Le fait que $a \in U$ soit un point critique de f ne suffit pas pour que f présente un extremum local en ce point. Il existe déjà des contreexemples en dimension 1, par exemple la fonction $f : x \mapsto x^3$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ avec $f'(0) = 0$, mais n'admet pas d'extremum local en ce point.

Une situation fréquente avec des fonctions de deux variables réelles est celle du **point-col**: imaginons une fonction $f : \mathbb{R}^2 \to \mathbb{R}$, de classe $\mathcal{C}^1$, telle que l'application partielle $x \mapsto f(x,0)$ admette un minimum pour $x = 0$, alors que l'application partielle $y \mapsto f(0,y)$ admet un maximum pour $y = 0$. Dans ce cas, les deux dérivées partielles $\dfrac{\partial f}{\partial x}$ et $\dfrac{\partial f}{\partial y}$ sont nulles en $(0,0)$ alors que f n'admet pas d'extremum local à l'origine. Un exemple simple de fonction satisfaisant ces conditions est $f : (x,y) \mapsto x^2 - y^2$.

3. **Recherche d'extremum global sur une partie fermée bornée.**

Je rappelle simplement ici le **théorème des bornes atteintes**: si K est une partie fermée bornée de $\mathbb{R}^p$, si $f : K \to \mathbb{R}$ est une application continue, alors f est bornée sur K et atteint ses bornes, i.e. il existe

$$m = \min_K f = \min_{x \in K} f(x) \qquad \text{et} \qquad M = \max_K f = \max_{x \in K} f(x) \ .$$

Ce théorème peut être combiné au précédent pour rechercher les extremums globaux d'une fonction continue sur une partie fermée bornée de $\mathbb{R}^p$.

Exemple. Soit K la partie fermée bornée du plan:

$$K = \big\{(x,y) \in \mathbb{R}^2 \mid x \geq 0 \ , \ y \geq 0 \ , \ x + y \leq 1\big\} \ .$$

On reconnaît un "triangle plein" (bords compris), de sommets $O(0,0)$, $A(1,0)$ et $B(0,1)$. Soit la fonction $f : (x,y) \mapsto xy(1-x-y)$. Alors f est continue sur K (c'est une fonction polynomiale à deux variables), donc elle atteint un maximum $M = \max\limits_{(x,y)\in K} f(x,y)$ sur cette partie. Comme f est nulle sur le bord de K, i.e. sur les trois côtés du triangle, et qu'elle est strictement positive sur l'intérieur

$$\overset{\circ}{K} = U = \big\{(x,y) \in \mathbb{R}^2 \mid x > 0 \ , \ y > 0 \ , \ x + y < 1\big\} \ ,$$

ce maximum est atteint sur la partie ouverte U. Comme f est de classe $\mathcal{C}^1$ sur U, il est nécessairement atteint en un point critique de f sur U. Recherchons donc les points critiques de f sur U: on calcule $\dfrac{\partial f}{\partial x} = y(y - 2x - 1)$ et $\dfrac{\partial f}{\partial y} = x(x - 2y - 1)$. Pour $(x,y) \in U$, on a

$$\nabla f(x,y) = 0 \iff \begin{cases} 2x + y &= 1 \\ x + 2y &= 1 \end{cases} \iff x = y = \frac{1}{3} \ .$$

La fonction f admet donc un seul point critique dans U, qui est $\left(\dfrac{1}{3}, \dfrac{1}{3}\right)$. C'est donc nécessairement en ce point que le maximum est atteint, donc

$$M = \max_{(x,y)\in K} f(x,y) = f\left(\frac{1}{3}, \frac{1}{3}\right) = \frac{1}{27} \ .$$

4. **Conditions d'ordre deux.**

Le théorème vu dans le paragraphe 2, n'utilisant que les dérivées partielles premières, ne donne qu'une condition nécessaire pour qu'une fonction f admette un extremum local en un point. En considérant les dérivées partielles secondes, on peut énoncer des conditions suffisantes (la condition **(1)** ci-dessous), ce qui permet souvent de savoir, parmi les points critiques de f, lesquels sont effectivement des extremums.

<u>Théorème.</u> Soit $f : U \to \mathbb{R}$ de classe $\mathcal{C}^2$, où U est un ouvert de $\mathbb{R}^p$. Soit $a \in U$ un point critique de f. On a alors les deux résultats suivants:

(1): si $H_f(a) \in \mathcal{S}_p^{++}(\mathbb{R})$, alors f admet un minimum local strict au point a ;

(2): si $H_f(a) \notin \mathcal{S}_p^{+}(\mathbb{R})$, alors f n'admet pas de minimum en a.

<u>Commentaires.</u> Bien sûr, on peut remplacer l'énoncé **(2)** par sa contraposée: si f admet un minimum local en a, alors $H_f(a) \in \mathcal{S}_p^{+}(\mathbb{R})$.

Jouons avec le vocabulaire: Si f est de classe $\mathcal{C}^2$ sur un ouvert U et si $a \in U$ alors, pour que f admette un minimum local en a:

- il est **nécessaire** que la matrice symétrique $H_f(a)$ soit positive ;

- il est **suffisant** que a soit un point critique avec une matrice hessienne définie positive.

Dans les deux énoncés ci-dessus, on peut aussi remplacer “minimum” par “maximum” à condition de remplacer aussi la matrice hessienne $H_f(a)$ par son opposée $-H_f(a)$.

<u>Preuve.</u> Comme a est un point critique de f, le développement limité d'ordre deux de f en a s'écrit

$$f(a+h) - f(a) = \frac{1}{2}\big(H_f(a)\cdot h \mid h\big) + \|h\|^2\,\varepsilon(h)\ ,$$

avec $\lim\limits_{h\to 0} \varepsilon(h) = 0$.

(1): *Si la matrice symétrique $H_f(a)$ est définie positive, notons λ_1 sa plus petite valeur propre, on a alors $\lambda_1 > 0$ puisque $\mathrm{Sp}\big(H_f(a)\big) \subset \mathbb{R}_+^*$, et il résulte du théorème spectral (cf. “un approfondissement utile” à la fin du cours sur les espaces euclidiens) que*

$$\forall h \in \mathbb{R}^p \qquad \big(H_f(a)\cdot h \mid h\big) \geq \lambda_1\,\|h\|^2\ .$$

On a donc, pour $\|h\|$ petit, $f(a+h) - f(a) \geq \left(\dfrac{\lambda_1}{2} + \varepsilon(h)\right)\|h\|^2$. Comme $\lim\limits_{h\to 0}\varepsilon(h) = 0$, il existe $\delta > 0$ tel que, pour $h \in \mathbb{R}^p$ tel que $\|h\| \leq \delta$, on ait $\varepsilon(h) \geq -\dfrac{\lambda_1}{4}$. Pour h non nul et de norme inférieure à δ, on a alors $f(a+h) - f(a) > 0$, ce qui montre la présence d'un minimum local strict de f au point a.

(2): *Si la matrice symétrique $H_f(a)$ n'est pas positive, elle admet alors une valeur propre strictement négative λ_0, soit v_0 un vecteur propre associé, on a $H_f(a)\cdot v_0 = \lambda_0 v_0$ et $v_0 \neq 0$. Puis, pour t réel avec $|t|$ petit, on a*

$$\begin{aligned} f(a+tv_0) - f(a) &= \frac{1}{2}\big(H_f(a)\cdot tv_0 \mid tv_0\big) + \|tv_0\|^2\,\varepsilon(tv_0) \\ &= \frac{t^2}{2}\big(H_f(a)\cdot v_0 \mid v_0\big) + t^2\,\|v_0\|^2\,\varepsilon(tv_0) \\ &= t^2\,\|v_0\|^2\left(\frac{\lambda_0}{2} + \varepsilon(tv_0)\right) \end{aligned}$$

et cette quantité est strictement négative pour t non nul et $|t|$ assez petit car $\lim_{t\to 0} \varepsilon(tv_0) = 0$, *donc il existe* $\alpha > 0$ *tel que, pour* $|t| \leq \alpha$, *on ait* $\varepsilon(tv_0) \leq \dfrac{|\lambda_0|}{4}$, *donc* $\dfrac{\lambda_0}{2} + \varepsilon(tv_0) < 0$. *La fonction f ne présente donc pas de minimum local au point a.*

Cas de la dimension deux. Il résulte de l'étude des matrices symétriques (définies) positives d'ordre deux (tout à la fin du chapitre "espaces euclidiens") que, si $f : U \to \mathbb{R}$ est de classe $\mathcal{C}^2$ avec U ouvert du plan $\mathbb{R}^2$, si $a \in U$ est un point critique de f, alors:

- si $\mathrm{tr}\big(H_f(a)\big) > 0$ et $\det\big(H_f(a)\big) > 0$, alors f admet un minimum local strict en a ;
- si f admet un minimum local en a, alors $\mathrm{tr}\big(H_f(a)\big) \geq 0$ et $\det\big(H_f(a)\big) \geq 0$.