CHAPITRE 2

Fonctions de plusieurs variables : Calcul différentiel

La condition de continuité pour une fonction de plusieurs variables est une notion de régularité pratique et naturelle mais elle est trop générale et recouvre un ensemble important de phénomèmes très différents. Dasn ce chapitre on va discuter une condition de régularité plus restreinte : la différentiabilité.

1. Différentiabilité

La différentiabilité généralise aux fonctions de plusieurs variables la notion dérivabilité pour les fonctions d'une variable. Cependant pour énoncer cette généralisation il est utile de reformuler la notion de dérivabilité de manière adéquate.

1.1. Rappels sur la dérivabilité. Traditionellement, on dit qu'une fonction $f(x)$ d'une variable est dérivale en $x_0 \in \mathbf{R}$ si la limite du taux d'accroissement en x_0 existe : f est dérivable en x_0 si et seulement si

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{f(x) - f(x_0)}{x - x_0}$$

existe; on appelle alors cette limite *la dérivée* de f en x_0 et on la note $f'(x_0)$ On montre alors que si f est dérivable en x_0 alors elee est continue, ainsi la notion de dérivabilité est un notion de régularité plus forte que la continuité. Pour passer aux fonctions de plusieurs variables, il sera commode de formuler la dérivabiiltée sous une forme équivalente : l'expression

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0) \tag{1.1}$$

est équivalente [1] à l'identité

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + (x - x_0)\varepsilon(x - x_0),$$

avec $\varepsilon(h) \to 0$ quand $h \to 0$. En particulier, cette formulation montre immédiatement que f est continue en x_0 ; elle s'intèrprète en disant que *quand x est proche de x_0 la fonction $f(x)$ est bien approximée par la fonction affine*

$$x \mapsto f(x_0) + f'(x_0)(x - x_0).$$

1. le vérifier !

REMARQUE 1.1. On a vu que la continuité (en x_0)

$$f(x) = f(x_0) + \varepsilon_1(x - x_0)\text{avec } \varepsilon_1(h) \to 0 \text{ quand } h \to 0.$$

s'interprète en disant que *quand x est proche de x_0 $f(x)$ est approximée par la fonction constante $x \mapsto f(x_0)$*. La dérivabilité implique la continuité car la fonction

$$h \mapsto f'(x_0)h + h\varepsilon(h)$$

tend bien vers 0 quand h tend vers 0 cependant la dérivabilité dit plus car pour la continuité, l'erreur d'approximation par une fonction constante est simplement une fonction $\varepsilon_1(h)$ qui tend vers 0 en $h = 0$, alrs que pour la dérivabilité l'erreur d'approximation par une fonction affine est une fonction de la forme $h\varepsilon(h)$ qui tend vers 0 en $h = 0$ mais *a priori plus vite* puisque produit de deux fonction qui tendent vers 0 : $\varepsilon(h)$ et la fonction h. On parle pour la continuité d'approximation à l'ordre 0 et pour la dérivabilité d'approximation à l'ordre 1.

1.2. Définition de la différentiabilité. Soit $n \geqslant 1$ et une fonction de n variables

$$f : \vec{x} = (x_1, \dots, x_n) \in D(f) \subset \mathbf{R}^n \mapsto f(\vec{x}) \in \mathbf{R}$$

DÉFINITION 2.1. *On dit que f est différentiable en $\vec{x}_0 = (x_{0,1}, \dots, x_{0,n}) \in D(f)$, si il existe $\vec{a} = (a_1, \dots, a_n) \in \mathbf{R}^n$ et une fonction $\varepsilon(\vec{h})$ qui tend vers 0 quand $\vec{h} \to \vec{0}$ tel que on ait*

$$f(\vec{x}) = f(\vec{x}_0) + a_1(x_1 - x_{0,1}) + \cdots + a_n(x_n - x_{0,n}) + \|\vec{x} - \vec{x}_0\|\varepsilon(\vec{x} - \vec{x}_0). \tag{1.2}$$

De manière équivalente en écrivant $\vec{x} = \vec{x}_0 + \vec{h}$

$$f(\vec{x}_0 + \vec{h}) = f(\vec{x}_0) + a_1h_1 + \cdots + a_nh_n + \|\vec{h}\|\varepsilon(\vec{h}).$$

On voit que pour $n = 1$ cette définition n'est autre que celle de la dérivabilité avec $a_1 = f'(x_0)$.

De plus si f est différentiable en $\vec{x}_0$ alors elle est continue au même point : est effet, considéront la fonction

$$\varepsilon_1 : \vec{h} = (h_1, \dots, h_n) \mapsto a_1h_1 + \cdots + a_nh_n + \|\vec{h}\|\varepsilon(\vec{h});$$

alors $\varepsilon_1(\vec{h})$ tend vers 0 quand $\vec{h} \to 0$ et on a $f(\vec{x}) = f(\vec{x}_0) + \varepsilon_1(\vec{x} - \vec{x}_0)$.

On peut écrire la différentiabilité de manière plus compacte en terme de produit scalaire : en effet

$$a_1(x_1 - x_{0,1}) + \cdots + a_n(x_n - x_{0,n}) = \vec{a}.(\vec{x} - \vec{x}_0)$$

et donc

$$f(\vec{x}) = f(\vec{x}_0) + \vec{a}.(\vec{x} - \vec{x}_0) + \|\vec{x} - \vec{x}_0\|\varepsilon(\vec{x} - \vec{x}_0)$$

ou encore

$$f(\vec{x}_0 + \vec{h}) = f(\vec{x}_0) + \vec{a}.\vec{h} + \|\vec{h}\|\varepsilon(\vec{h})$$

1.3. Différentielle ; dérivées partielles. Soit f une fonction différentiable en $\vec{x}_0$, alors les coefficients $a_1, \ldots, a_n$ (ou ce qui revient au même le vecteur $\vec{a}$) est définit de manière unique : soit $i = 1 \ldots_n$, définissons la fonction d'une variable

$$f_i : x \mapsto f(x_{0,1}, \ldots, x, \ldots, x_{0,n})$$

ou la variable x est placée à la i-ième coordonnée : on a en particulier

$$f_i(x_{0,i}) = f(x_{0,1}, \ldots, x_{0,i}, \ldots, x_{0,n}) = f(\vec{x}_0)$$

et on voit par (1.2)

$$\begin{aligned} f_i(x) = f_i(x_{0,i}) &+ a_1.0 + \cdots + a_i.(x - x_{0,i}) + \cdots + a_n.0 \\ &+ \|(0, \ldots, x - x_{0,i}, \ldots, 0)\| \varepsilon((0, \ldots, x - x_{0,i}, \ldots, 0)) \\ = f_i&(x_{0,i}) + a_i.(x - x_{0,i}) + |x - x_{0,i}| \varepsilon_i(x - x_{0,i}) \end{aligned}$$

en posant

$$\varepsilon_i(h) = \varepsilon((0, \ldots, h, \ldots, 0)).$$

Cette fonction tend vers 0 quand h tend vers 0 et donc la fonction $x \mapsto f_i(x)$ est une fonction dérivable en $x_{0,i}$ de dérivée

$$f_i'(x_{0,i}) = a_i.$$

En d'autres termes le coefficient a_i est la dérivée au point $x_{0,i}$ de la fonction d'une variable obtenue à partir de $f(x_1, \ldots, x_i, \ldots, x_n)$ en fixant toutes les coordonnées x_j pour $j \neq i$ à la valeur $x_{0,j}$ et en replaçant la i-ième coordonnée par x. On appelle le coefficient a_i (ie la dérivéee de $f_i(x)$ en $x_{0,i}$) la *dérivée partielle de f dans la direction x_i au point $\vec{x}_0$* et on la note

$$a_i = f_i'(x_{0,i}) = \frac{\partial f}{\partial x_i}(\vec{x}_0).$$

Exemple 1.3.1. Soit $f(x, y) = x^2y + xy^3 + 2xy$; on va voir que cette fonction est différentiable en tout point (x_0, y_0) de $\mathbf{R}^2$. calculons ses dérivées partielles

$$\frac{\partial f}{\partial x}(x_0, y_0), \ \frac{\partial f}{\partial y}(x_0, y_0).$$

Pour calculer $\frac{\partial f}{\partial x}(x_0, y_0)$, on dérive suivant x en considérant y_0 comme une constante et on évalue le résultat en x_0 : on regarde la dérivée en x_0 de la fonction

$$x \mapsto f(x, y_0) = x^2y_0 + xy_0^3 + 2xy_0.$$

Cette dérivée vaut

$$(x \mapsto f(x, y_0) = x^2y_0 + xy_0^3 + 2xy_0)' = 2xy_0 + y_0^3 + 2y_0$$

et donc

$$\frac{\partial f}{\partial x}(x_0, y_0) = 2x_0y_0 + y_0^3 + 2y_0.$$

De même pour la dérivée en y on considère la dérivée en y

$$(x \mapsto f(x_0, y) = x_0^2y + x_0y^3 + 2x_0y)' = x_0^2 + 3x_0y^2 + 2x_0$$

et donc

$$\frac{\partial f}{\partial y}(x_0, y_0) = x_0^2 + 3x_0y_0^2 + 2x_0.$$

1.3.2. *Règles de différentiation.* On dispose comme pour les fonctions d'une variable des critères suivants pour déterminer si une fonction est différentiable :

(1) Les fonctions constantes et les fonctions coordonnées

$$x_i : \vec{x} = (x_1, \ldots, x_n) \mapsto x_i$$

sont différentiables sur $\mathbf{R}^n$; les dérivées partielles des fonctions constantes sont nulles et

$$\frac{\partial x_i}{\partial x_j}(\vec{x}) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}.$$

(2) Sommes, produits : Si $f, g : \mathbf{R}^n \mapsto \mathbf{R}$ sont différentiables en $\vec{x}_0$ alors $f + g$ et $f \times g$ sont différentiables en $\vec{x}_0$.

(3) Quotients : Si de plus $g(\vec{x}_0) \neq 0$ alors $\dfrac{f}{g}$ est différentiable en $\vec{x}_0$.

(4) Composition : Si $f : \mathbf{R}^n \mapsto \mathbf{R}$ différentiable en $\vec{x}_0$ et $\varphi : \mathbf{R} \mapsto \mathbf{R}$ est dérivable (il suffit en fait qu'elle soit dérivable au point $f(\vec{x}_0)$) alors

$$\varphi \circ f : \vec{x} \mapsto \varphi(f(\vec{x}))$$

est différentiable en $\vec{x}_0$.

On a alors les valeurs suivants pour les dérivées partielles : pour tout $j = 1 \ldots n$

(1)

$$\frac{\partial f + g}{\partial x_j}(\vec{x}_0) = \frac{\partial f}{\partial x_j}(\vec{x}_0) + \frac{\partial g}{\partial x_j}(\vec{x}_0).$$

(2)

$$\frac{\partial f \times g}{\partial x_j}(\vec{x}_0) = \frac{\partial f}{\partial x_j}(\vec{x}_0)g(\vec{x}_0) + f(\vec{x}_0)\frac{\partial g}{\partial x_j}(\vec{x}_0).$$

(3)

$$\frac{\partial f/g}{\partial x_j}(\vec{x}_0) = \frac{1}{g(\vec{x}_0)^2}(\frac{\partial f}{\partial x_j}(\vec{x}_0)g(\vec{x}_0) - f(\vec{x}_0)\frac{\partial g}{\partial x_j}(\vec{x}_0)).$$

(4)

$$\frac{\partial \varphi \circ f}{\partial x_j}(\vec{x}_0) = \varphi'(f(\vec{x}_0))\frac{\partial f}{\partial x_j}(\vec{x}_0).$$

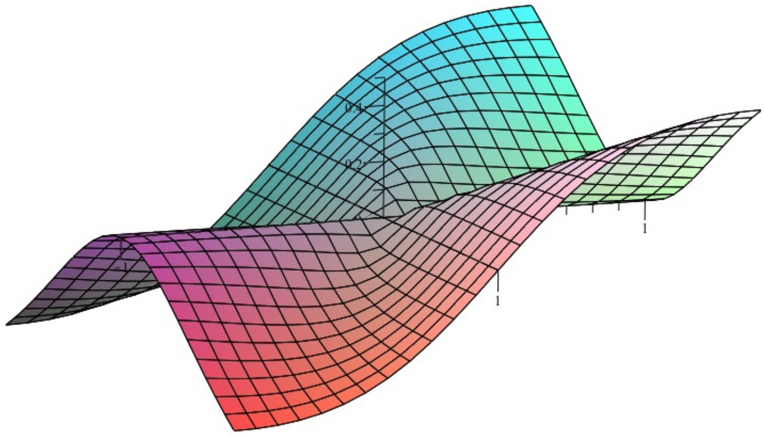

FIGURE 1. $f(x,y) = x^2y/(x^2+y^2)$

1.4. Différentiabilité et dérivées partielles : ATTENTION !. Comme on vient de la voir la différentiabilité d'une fonction $f : D(f) \subset \mathbf{R}^n \to \mathbf{R}$ en $\vec{x}_0$ implique l'existence des dérivées partielles en ce point

$$\frac{\partial f}{\partial x_j}(\vec{x}_0) = \lim_{x_j \to x_{0,j}} \frac{f_j(x_j) - f(\vec{x}_0)}{x_j - x_{0,j}}, \ j = 1 \dots n.$$

ATTENTION la réciproque n'est PAS vraie : par exemple la fonction

$$f(x,y) = \frac{x^2y}{x^2+y^2}, \ (x,y) \neq (0,0), \ f(0,0) = 0$$

est différentiable sur $\mathbf{R}^2 - \{(0,0)\}$ mais en $(0,0)$ elle n'est pas différentiable bien que ses deux derivées partielles existent en $(0,0)$

$$\frac{\partial f}{\partial x}(0,0) = \lim_{x\to 0} \frac{f(x,0)-f(0,0)}{x} = 0, \ \frac{\partial f}{\partial y}(0,0) = \lim_{y\to 0} \frac{f(0,y)-f(0,0)}{y} = 0.$$

Supposons cette fonction différentiable en $(0,0)$, on aura alors pour (x,y) proche de $(0,0)$

$$f(x,y) = f(0,0) + \frac{\partial f}{\partial x}(0,0)x + \frac{\partial f}{\partial y}(0,0)y + \|(x,y)\|\varepsilon(x,y) = \|(x,y)\|\varepsilon(x,y).$$

Prenons $(x,y) = (x,x)$ et calculons

$$\lim_{x\to 0} \frac{f(x,x)-f(0,0)}{x}.$$

D'après la dernière égalité cela vaut ($\|(x,y)\| = |x|\|(1,1)\|$)

$$\lim_{x\to 0} \frac{|x|\varepsilon(x,x)}{x} = 0.$$

D'autre part
$$\frac{f(x,x)-f(0,0)}{x}=\frac{1}{x}\times\frac{x^2\times x}{x^2+x^2}=\frac{1}{2},$$
contradiction. □

On a cependant le résultat plus positif suivant en faisant un hypothèse de régularité des dérivées partielles :

THÉORÈME 2.1. *Soit $f : D(f)\subset \mathbf{R}^n \to \mathbf{R}$ une fonction de n variables. On suppose que en tout point $\vec{x}_0$ de $D(f)$ les dérivées partielles*
$$\frac{\partial f}{\partial x_j}(\vec{x}_0)=\lim_{x_j\to x_{0,j}}\frac{f_j(x_j)-f(\vec{x}_0)}{x_j-x_{0,j}},\ j=1\dots n$$
existent et que les fonctions
$$\frac{\partial f}{\partial x_j}:\vec{x}_0\to\frac{\partial f}{\partial x_j}(\vec{x}_0)$$
sont continues en $\vec{x}_0$, alors f est différentiable sur $D(f)$.

1.5. Différentielle ; Gradient. Le vecteur des dérivées partielles
$$\vec{a}=(a_1,\dots,a_n)=(\frac{\partial f}{\partial x_1}(\vec{x}_0),\dots,\frac{\partial f}{\partial x_n}(\vec{x}_0))$$
est appelé *le gradient* de f au point $\vec{x}_0$ et est noté
$$\nabla(f)(\vec{x}_0)=(\frac{\partial f}{\partial x_1}(\vec{x}_0),\dots,\frac{\partial f}{\partial x_n}(\vec{x}_0)).$$
On trouve aussi comme notation du gradient
$$\overrightarrow{\mathrm{Grad}}(f)(\vec{x}_0).$$

La *différentielle* de f en $\vec{x}_0$ est la fonction sur $\mathbf{R}^n$
$$df(\vec{x}_0):\vec{h}=(h_1,\dots,h_n)\mapsto\vec{\nabla}(f)(\vec{x}_0).\vec{h}=\frac{\partial f}{\partial x_1}(\vec{x}_0)h_1+\cdots+\frac{\partial f}{\partial x_n}(\vec{x}_0)h_n.$$
C'est une fonction linéaire (également appelée *forme linéaire*) :
$$df(\vec{x}_0)(\vec{h}+\vec{h}')=df(\vec{x}_0)(\vec{h})+df(\vec{x}_0)(\vec{h}'),\ df(\vec{x}_0)(\lambda\vec{h})=\lambda df(\vec{x}_0)(\vec{h}).$$
On écrira de manières équivalentes
$$f(\vec{x})=f(\vec{x}_0)+df(\vec{x}_0)(\vec{x}-\vec{x}_0)+\|\vec{x}-\vec{x}_0\|\varepsilon(\vec{x}-\vec{x}_0),$$
$$f(\vec{x})=f(\vec{x}_0)+\nabla(f)(\vec{x}_0).(\vec{x}-\vec{x}_0)+\|\vec{x}-\vec{x}_0\|\varepsilon(\vec{x}-\vec{x}_0),$$
$$f(\vec{x}_0+\vec{h})=f(\vec{x}_0)+df(\vec{x}_0)(\vec{h})+\|\vec{h}\|\varepsilon(\vec{h}).$$
$$f(\vec{x}_0+\vec{h})=f(\vec{x}_0)+\nabla(\vec{x}_0).\vec{h}+\|\vec{h}\|\varepsilon(\vec{h}).$$

1.6. Approximation linéaire. La fonction

$$\vec{x} \mapsto f(\vec{x}_0) + \frac{\partial f}{\partial x_1}(\vec{x}_0).(x_1 - x_{0,1}) + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0).(x_n - x_{0,n})$$

est une fonction affine (somme d'une constante et d'une fonction linéaire). La définition de la différentiabilité

$$f(\vec{x}) = f(\vec{x}_0) + \frac{\partial f}{\partial x_1}(\vec{x}_0).(x_1 - x_{0,1}) + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0).(x_n - x_{0,n}) + \|\vec{x} - \vec{x}_0\|\varepsilon(\vec{x} - \vec{x}_0)$$

s'interprète en disant que la fonction $\vec{x} \mapsto f(\vec{x})$ est bien approximée par cette fonction affine quand $\vec{x}$ est proche de $\vec{x}_0$. Notons que (si le vecteur $df(\vec{x}_0) = \nabla f(\vec{x}_0)$ est non-nul, le terme d'erreur fait dans cette approximation $\|\vec{x} - \vec{x}_0\|\varepsilon(\vec{x} - \vec{x}_0)$ est un produit de deux termes tendant vers 0 et en tous cas, tend vers zéro "plus vite" que le terme $\frac{\partial f}{\partial x_1}(\vec{x}_0)(x_1 - x_{0,1}) + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0)(x_n - x_{0,n})$.

2. Plan tangent

2.1. Hyperplans de $\mathbf{R}^n$.

Définition 2.2. *Soit $\vec{a} = (a_1, \cdots, a_n)$, $\vec{x}_0 = (x_{0,1}, \cdots, x_{0,n}) \in \mathbf{R}^n$ deux vecteurs ; on suppose que $\vec{a} \neq 0$, l'hyperplan $HP(\vec{a}, \vec{x}_0)$de $\mathbf{R}^n$ perpendiculaire à $\vec{a}$ et passant par $\vec{x}_0$ est l'ensemble des vecteurs $\vec{x}$ tels que $\vec{x} - \vec{x}_0$ est perpendiculaire à $\vec{a}$, c'est à dire qui vérifient*

$$0 = \vec{a}.(\vec{x} - \vec{x}_0) = a_1(x_1 - x_{0,1}) + \cdots + a_n(x_n - x_{0,n}).$$

Remarque 2.1. Si $\vec{a} = \vec{0}$, l'équation devient

$$0 = 0$$

et on obtient $\mathbf{R}^n$ en entier, c'est porquoi on exclut ce cas dans la suite.

Alternativement, on voit en posant $\vec{x} = \vec{x}_0 + \vec{h}$ ($x_i = x_{0,i} + h_i$) que $HP(\vec{a}, \vec{x}_0)$ est obtenu à partir de l'hyperplan $HP(\vec{a}, \vec{0})$ formé des vecteurs perpendiculaires à $\vec{a}$ (les vecteurs $\vec{h} = (h_1, \cdots, h_n)$ tels que $\vec{a}.\vec{h} = a_1h_1 + \cdots + a_n.h_n = 0$),

$$HP(\vec{a}, \vec{x}_0) = \vec{x}_0 + HP(\vec{a}, \vec{0}).$$

Le vecteur $\vec{a}$ est appelé un *vecteur normal* à l'hyperplan, et $\vec{x}_0$ est un *vecteur de translation* associé à cet hyperplan. On dit encore que l'hyperplan "passe" par $\vec{x}_0$ puisque $\vec{x}_0$ est contenu dedans.

La relation

$$0 = a_1(x_1 - x_{0,1}) + \cdots + a_n(x_n - x_{0,n})$$

définit l'équation de cet hyperplan et posant

$$b = -a_1x_{0,1} + \cdots - a_nx_{0,n}$$

cette equation s'écrit

$$0 = a_1x_1 + \cdots + a_nx_n + b.$$

Réciproquement tout relation linéaire de cette forme (avec $(a_1, \cdots, a_n) = \vec{a} \neq \vec{0}$) définit un hyperplan de vecteur normal $\vec{a}$ et dont un vecteur de translation est

$$\vec{x}_0 = (0, \cdots, -b/a_i, \cdots, 0)$$

pour i tel que $a_i \neq 0$.

2.2. Plan tangent au graphe d'une fonction. Comme nous l'avons expliqué, si une fonction est différentiable en un point $\vec{x}_0$ alors "au voisinage" de $\vec{x}_0$, la fonction $f(\vec{x})$ est bien approximée par la fonction affine

$$\vec{x} \mapsto f(\vec{x}_0) + \vec{a}.(\vec{x} - \vec{x}_0) \tag{2.1}$$

avec

$$\vec{a} = (\frac{\partial f}{\partial x_1}(\vec{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}_0)) = \nabla f(\vec{x}_0).$$

Géométriquement cela signifie encore que dans $\mathbf{R}^{n+1}$, "au voisinage" du point $(\vec{x}_0, f(\vec{x}_0))$, le graphe de la fonction (2.1) est "proche" du graphe de f. Ce dernier graphe est un hyperplan dont l'équation est (les coordonnées de $\mathbf{R}^{n+1}$ étant notées $(x_1, \ldots, x_n, x_{n+1})$

$$x_{n+1} = z_0 + \vec{a}.(\vec{x} - \vec{x}_0), \tag{2.2}$$

où on a posé $z_0 = f(\vec{x}_0)$. C'est hyperplan est appelé *hyperplan tangent* au graphe de f au point

$$(\vec{x}_0, f(\vec{x}_0)) = (x_{0,1}, \ldots, x_{0,n}, z_0).$$

2.3. Vecteur normal au graphe d'une fonction. L'q́aution du plan tangent se réecrit

$$0 = \vec{a}.(\vec{x} - \vec{x}_0) - (z - z_0) = (a_1, \cdots, a_n, -1).(x_1 - x_{0,1}, \cdots, x_n - x_{0,n}, x_{n+1} - z_0).$$

C'est donc l'hyperplan associé au vecteur normal

$$\vec{n}_f(\vec{x}_0) = (\nabla(f)(\vec{x}_0), -1) = (\frac{\partial f}{\partial x_1}(\vec{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}_0), -1)$$

et passant par le point $(\vec{x}_0, f(\vec{x}_0))$.

DÉFINITION 2.3. *Le vecteur $\vec{n}_f(\vec{x}_0)$ est le vecteur normal au graphe de f au(-dessus du) point $\vec{x}_0$.*

2.4. Plan tangent en une variété de niveau. On a déjà vu que le gradient permet de former le *vecteur normal* au graphe de f en $(\vec{x}_0, f(\vec{x}_0))$:

$$\vec{n}_f(\vec{x}_0) = (\nabla(f)(\vec{x}_0), -1).$$

Soit $C = f(\vec{x}_0)$ et considérons maintenant la variété de niveau

$$V_f(C) = \{\vec{x} \in \mathbf{R}^n,\ f(\vec{x}) = C\}$$

Cette variété est contenue dans $\mathbf{R}^n$ et contient évidemment le point $\vec{x}_0$. On voudrait décrire la structure de la variété de niveau $V_f(C)$ au voisinage de $\vec{x}_0$: soit $\vec{x}$ sur $V_f(C)$ et proche de $\vec{x}_0$; $\vec{x}$ vérifie

$$f(\vec{x}) = f(\vec{x}_0)$$

(ie. est dans la variété de niveau) d'autre part, part la formule d'approximation linéaire, on a

$$f(\vec{x}) = f(\vec{x}_0) = f(\vec{x}_0) + \nabla(f)(\vec{x}_0).(\vec{x} - \vec{x}_0) + \|\vec{x} - \vec{x}_0\|\varepsilon(\vec{x} - \vec{x}_0)$$

soit en simplifiant

$$\nabla(f)(\vec{x}_0).(\vec{x} - \vec{x}_0) = -\|\vec{x} - \vec{x}_0\|\varepsilon(\vec{x} - \vec{x}_0)$$

ou encore si $\vec{x} \neq \vec{x}_0$

$$\nabla(f)(\vec{x}_0).\frac{\vec{x} - \vec{x}_0}{\|\vec{x} - \vec{x}_0\|} = \varepsilon(\vec{x} - \vec{x}_0) \to 0$$

quand $\vec{x} \to \vec{x}_0$: en d'autres termes le vecteur de longueur 1 $\frac{\vec{x}-\vec{x}_0}{\|\vec{x}-\vec{x}_0\|}$ est "presque" orthogonal au vecteur $\nabla(f)(\vec{x}_0)$; si $\nabla(f)(\vec{x}_0) \neq 0$ cela signifie aussi que $\vec{x}$ est preseque contenu dans l'hyperplan $HP(\nabla f(\vec{x}_0), \vec{x}_0)$ passant par $\vec{x}_0$ et perpendiculaire à $\nabla f(\vec{x}_0)$.

DÉFINITION 2.4. *Soit $f : \mathbf{R}^n \to \mathbf{R}$ différentiable en $\vec{x}_0$ et soit $C = f(\vec{x}_0)$. Si $\nabla(f)(\vec{x}_0) \neq \vec{0}$ alors l'hyperlan de $\mathbf{R}^n$ donné par l'équation*

$$0 = \nabla(f)(\vec{x}_0).(\vec{x} - \vec{x}_0) = \frac{\partial f}{\partial x_1}(\vec{x}_0)(x_1 - x_{0,1}) + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0)(x_n - x_{0,n})$$

est appelé plan tangent en $\vec{x}_0$ à la variété de niveau $V_f(C)$. Le gradient $\nabla f(\vec{x}_0)$ est le vecteur normal *en $\vec{x}_0$ à la variété de niveau $V_f(C)$.*

On interprète la discussion précédente en disant que quand on est "près" de $\vec{x}_0$, la variété de niveau $V_f(C)$ "ressemble" à un hyperplan autrement dit :

2.4.1. *La terre est plate.* En effet la surface de la terre est donné par l'équation

$$x^2 + y^2 + z^2 = R^2, \ R = 6378Km$$

et si un humain (0.0018 Km) se tient au point (x_0, y_0, z_0) de la surface terrestre„ il aura l'impression que la surface de la terre ressemble au plan tangent d'équation

$$0 = x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0).$$

2.5. Plan tangent au graphe/plan tangent à une variété de niveau. On notera que cette définition du vecteur normal à une *variété de niveau* en terme du gradient est "compatible" avec la notion de vecteur normal associé au *graphe* d'une fonction. En effet le graphe d'une fonction f de $\mathbf{R}^n$, n'est autre que la variété de niveau 0, $V_F(0)$ associée à la fonction sur $\mathbf{R}^{n+1}$,

$$F(x_1, \ldots, x_n, z) := f(x_1, \ldots, x_n) - z;$$

En effet le graphe de f est définit par l'équation

$$F(x_1, \ldots, x_n, z) = f(x_1, \ldots, x_n) - z = 0.$$

On a, en posant $z_0 = f(\vec{x}_0)$,

$$\nabla F(\vec{x}_0, z_0) = (\frac{\partial f}{\partial x_1}(\vec{x}_0), \ldots, \frac{\partial f}{\partial x_n}(\vec{x}_0), \frac{\partial F}{\partial z}(z_0)) = (\nabla(f)(\vec{x}_0), -1) = n_f(\vec{x}_0).$$

2.6. Point critique. Si $\nabla f(\vec{x}_0) = 0$, le plan tangent à $V_f(f(\vec{x}_0))$ n'est pas définit : on obtient l'équation $0 = 0$ satisfaite par tous les points de $\mathbf{R}^n$: on dit alors que $\vec{x}_0$ est un *point critique*. Voir le chapitre suivant pour une discussion des points critique dans le cas des fonctions de deux variables.

3. Dérivée directionnelle d'une fonction

La i-ème dérivée partielle $\frac{\partial f}{\partial x_i}(\vec{x}_0)$ d'une fonction différentiable décrit la variation de f au voisinage de $\vec{x}_0$ quand on se deplace à partir de $\vec{x}_0$ parallèlelement à l'axe de la coordonnée x_i. La notion de dérivée directionnelle est une extension de cette notion quand on se déplace le long d'un axe général :

Définition 2.5. *Soit f une fonction sur $\mathbf{R}^n$ différentiable en $\vec{x}_0$ et soit $\vec{e} = (e_1, \dots, e_n)$ un vecteur de $\mathbf{R}^n$; la dérivée de f en $\vec{x}_0$ dans la direction $\vec{e}$, $D_{\vec{e}}f(\vec{x}_0)$ est donnée par la limite*

$$D_{\vec{e}}f(\vec{x}_0) = \lim_{\substack{h \to 0 \\ h \neq 0}} \frac{f(\vec{x}_0 + h\vec{e}) - f(\vec{x}_0)}{h}.$$

Notons que cette limite existe : en effet par la formule d'approximation linéaire, on a

$$f(\vec{x}_0 + h\vec{e}) - f(\vec{x}_0) = \nabla f(\vec{x}_0).h\vec{e} + |h|\|\vec{e}\|\varepsilon(h\vec{e})$$

et donc

$$\frac{f(\vec{x}_0 + h\vec{e}) - f(\vec{x}_0)}{h} = \nabla f(\vec{x}_0).\vec{e} + \|\vec{e}\|\varepsilon(h\vec{e}) \to \nabla f(\vec{x}_0).\vec{e} \; h \to 0.$$

Ainsi

$$D_{\vec{e}}f(\vec{x}_0) = \nabla f(\vec{x}_0).\vec{e} = \frac{\partial f}{\partial x_1}(\vec{x}_0)e_1 + \cdots + \frac{\partial f}{\partial x_n}(\vec{x}_0)e_n.$$

Remarque 3.1. Si $\vec{e} = \vec{e}_i = (0, \dots, 1, 0, \dots, 0)$ (le 1 en i-ème position)

$$D_{\vec{e}_i}f = \frac{\partial f}{\partial x_i}.$$

Ainsi la dérivée directionnelle généralise la notion de dérivées partielles.

3.1. Propriétés de la dérivée directionnelle. Soient f et g des fonction différentiables en $\vec{x}_0$ et λ un réel, on vérifie facilement que la dérivée directionnelle est *linéaire en $\vec{e}$* :

$$D_{\vec{e}+\vec{e}'}(f)(\vec{x}_0) = D_{\vec{e}}(f)(\vec{x}_0) + D_{\vec{e}'}(f)(\vec{x}_0)$$

et

$$D_{\lambda\vec{e}}(f)(\vec{x}_0) = \lambda D_{\vec{e}}(f)(\vec{x}_0).$$

Applquant cela à la décomposition

$$\vec{e} = (e_1, \dots, e_n) = e_1\vec{e}_1 + \cdots + e_n\vec{e}_n$$

on en déduit immédiatement des identités analogues à celles des dérivées partielles $\frac{\partial f}{\partial x_i}$ les identités

(1)
$$D_{\vec{e}}(f+g)(\vec{x}_0) = D_{\vec{e}}(f)(\vec{x}_0) + D_{\vec{e}}(g)(\vec{x}_0).$$
(2)
$$D_{\vec{e}}(f \times g)(\vec{x}_0) = D_{\vec{e}}(f)(\vec{x}_0)g(\vec{x}_0) + f(\vec{x}_0)D_{\vec{e}}(g)(\vec{x}_0).$$
(3)
$$D_{\vec{e}}(f/g)(\vec{x}_0) = \frac{1}{g(\vec{x}_0)^2}(D_{\vec{e}}(f)(\vec{x}_0)g(\vec{x}_0) - f(\vec{x}_0)D_{\vec{e}}(g)(\vec{x}_0).$$
(4)
$$D_{\vec{e}}(\varphi \circ f)(\vec{x}_0) = \varphi'(f(\vec{x}_0))D_{\vec{e}}(f)(\vec{x}_0).$$

3.2. Variation d'une fonction dans une direction donnée. Supposons que $\vec{e}$ est un vecteur *unitaire* : ie un vecteur de longueur euclidienne 1
$$\|\vec{e}\| = \sqrt{e_1^2 + \cdots + e_n^2} = 1.$$
La formule d'approximation linéaire s'écrit
$$f(\vec{x}_0 + h\vec{e}) = f(\vec{x}_0) + hD_{\vec{e}}f(\vec{x}_0) + h\varepsilon(h\vec{e}).$$
Comme $h\varepsilon(h\vec{e})$ est une erreur, la dérivée directionnelle $D_{\vec{e}}f(\vec{x}_0) = \nabla f(\vec{x}_0).\vec{e}$ mesure donc *l'amplitude* de la variation de la fonction f quand la variable $\vec{x}$ se deplace depuis $\vec{x}_0$ dans la direction $\vec{e}$. Par exemple cette amplitude est *nulle* (f varie très peu) si
$$D_{\vec{e}}f(\vec{x}_0) = \nabla f(\vec{x}_0).\vec{e} = 0$$
c'est à dire si la direction $\vec{e}$ du déplacement est *perpendiculaire* au gradient, ou ce qui revient au même, si le deplacement à lieu "le long" du plan tangent à la variété de niveau
$$f(\vec{x}) = f(\vec{x}_0).$$
Inversemment, si $\vec{e}$ est *colinéaire* à $\nabla f(\vec{x}_0)$ alors
$$D_{\vec{e}}f(\vec{x}_0) = \nabla f(\vec{x}_0).\vec{e} = \pm\|\nabla f(\vec{x}_0)\|$$
et le coefficient de variation est le plus grand possible : en effet par l'inégalité de Cauchy-Schwarz on a toujours
$$|\nabla f(\vec{x}_0).\vec{e}| \leqslant \|\nabla f(\vec{x}_0)\|\|\vec{e}\| = \|\nabla f(\vec{x}_0)\|$$
avec égalité si et seulement si $\vec{e}$ et $\nabla f(\vec{x}_0)$ son colinéaires. Ainsi un déplacement de $\vec{x}$ dans la direction du gradient produit une variation *maximale* de la valeur de la fonction f.

Remarque 3.2. Cela est naturellement intuitif : si on regarde une carte topographique, la variation d'altitude est la plus faible si on se déplace parallèlement aux courbes de niveaux et la variation d'altitude est la plus forte si on se déplace perpendiculairement aux courbes de niveaux.

4. Dérivées partielles d'ordre supérieur

La notion de dérivées partielles d'ordre supérieur généralise celle de dérivée d'rodre supérieur (dérivée seconde, troisième,...) pour les fonction d'une variable.

Pour plus de clarté, on restreint la discussion qui vient aux fonctions de deux variables

$$(x, y) \mapsto f(x, y).$$

Etant donné une telle fonction ; supposons qu'elle soit différentiable sur une domaine $D(f)$; on dispose alors en chaque point (x, y) de ce domaine des deux dérivées partielles

$$\frac{\partial f}{\partial x}(x, y), \ \frac{\partial f}{\partial y}(x, y),$$

et on dispose donc de deux fonctions de deux variables, notées $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ sur $D(f)$

$$\begin{aligned} \frac{\partial f}{\partial x} &: (x, y) \mapsto \frac{\partial f}{\partial x}(x, y), \\ \frac{\partial f}{\partial y} &: (x, y) \mapsto \frac{\partial f}{\partial y}(x, y) \end{aligned}$$

on peut alors se démander si ces fonctions sont continues différentiables sur $D(f)$.

Supossons qu'elles soient différentiables, on a alors pour chacune d'elle deux dérivées partielles ce qui nous donne quater dérivées partielles *d'ordre* 2

$$\frac{\partial f}{\partial x}(\frac{\partial f}{\partial x}), \ \frac{\partial f}{\partial y}(\frac{\partial f}{\partial x}), \ \frac{\partial f}{\partial x}(\frac{\partial f}{\partial y}), \ \frac{\partial f}{\partial y}(\frac{\partial f}{\partial y}),$$

et à patir de ces quaters fonctions on peut éventuellement former (si chacune d'elle est différentiable) 8 dérivées partielles *d'ordre 3.* et ainsi de suite...

DÉFINITION 2.6. *Une fonction de deux variables $f : D(f) \subset \mathbf{R}^2 \to \mathbf{R}$ est différentiable à l'ordre* 1 *sur un domaine $D(f)$ si elle est différentiable en tout point de ce domaine : elle admet alors deux dérivées partielles $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ qui sont des fonctions sur $D(f)$.*

DÉFINITION 2.7. *Une telle fonction est différentiable à l'ordre k si elle est différentiable et si ses deux dérivées partielles $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ sont différentiables à l'ordre $k-1$. elle admet alors 2^k dérivées partielles d'rodre k qui sont obtenues en appliquant à f k fois f une dérivation dans la direction x ou dans la direction y :*

$$(\frac{\partial}{\partial x}(\frac{\partial}{\partial x}(\frac{\partial}{\partial y}...f)), \ k \ fois.$$

Une function k-fois différentiable pour tout entier $k \geqslant 1$, sera dite infiniment différentiable.

EXEMPLE 4.0.1. Soit $f(x, y) = x\cos(xy)$; cette fonction est infiniment différentiable sur $\mathbf{R}^2$. On a

$$\frac{\partial f}{\partial x}(x, y) = \cos(xy) - xy\sin(xy), \ \frac{\partial f}{\partial y}(x, y) = -x^2\sin(xy)$$

$$\frac{\partial}{\partial x}(\frac{\partial f}{\partial x})(x,y) = -2y\sin(xy) - xy^2\cos(xy),\ \frac{\partial}{\partial y}(\frac{\partial f}{\partial x})(x,y) = -2x\sin(xy) - x^2y\cos(xy)$$

$$\frac{\partial}{\partial x}(\frac{\partial f}{\partial y})(x,y) = -2x\sin(xy) - x^2y\cos(xy),\ \frac{\partial}{\partial y}(\frac{\partial f}{\partial y})(x,y) = -x^3\cos(xy)$$

NOTATION. *Si on effectue plusieurs fois (k fois) une opération de dérivation suivant la même variable (disons x), on notera cette opération*

$$\frac{\partial^k}{\partial x^k}.$$

Ainsi

$$\frac{\partial}{\partial x}(\frac{\partial f}{\partial x}) = \frac{\partial^2 f}{\partial x^2},\ \frac{\partial}{\partial y}(\frac{\partial f}{\partial y}) = \frac{\partial^2 f}{\partial y^2};$$

de même on écrira

$$\frac{\partial}{\partial x}(\frac{\partial f}{\partial y}) = \frac{\partial^2 f}{\partial x \partial y}$$

ou encore

$$\frac{\partial}{\partial x}(\frac{\partial}{\partial x}(\frac{\partial f}{\partial y})) = \frac{\partial^3 f}{\partial^2 x \partial y}$$

4.1. Commutativité de l'ordre des dérivations. Etant donné une fonction f au moins deux fois différentiable : on peut calculer sa dérivée partielle en x puis sa dérivée partielle en y, obtenant la dérivée partielle d'ordre 2

$$\frac{\partial}{\partial y}(\frac{\partial f}{\partial y}) = \frac{\partial^2 f}{\partial y \partial x}$$

alternativement on pourrait commencer par dériver en y puis en x obtenant

$$\frac{\partial}{\partial x}(\frac{\partial f}{\partial y}) = \frac{\partial^2 f}{\partial x \partial y}.$$

On remarque que dans l'exemple précédent ($f(x,y) = x\cos(xy)$) qu'on obtient la même fonction

$$\frac{\partial}{\partial y}(\frac{\partial f}{\partial x}) = \frac{\partial}{\partial x}(\frac{\partial f}{\partial y}) = -2x\sin(xy) - x^2y\cos(xy).$$

C'est en fait un phénomème général :

THÉORÈME 2.2. *Soit* $f : D(f) \subset \mathbf{R}^2 \to \mathbf{R}$ *une fonction différentiable à l'ordre* 2. *Supposons que ses dérivées partielles d'ordre* 2,

$$\frac{\partial^2 f}{\partial x^2},\ \frac{\partial^2 f}{\partial x \partial y},\ \frac{\partial^2 f}{\partial y \partial x},\ \frac{\partial^2 f}{\partial y^2}$$

sont toutes continues (par exemple si f est différentiable à l'ordre 3*) alors on a l'égalité*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

En d'autre termes pour une fonction suffisament régulière, l'ordre dans lequel on effectue les opérations de dérivation partielle ne compte pas.

Généralisant ce théorème on voit que si f est différentiable à l'ordre au moins $k+1$ ses dérivées partielles d'ordre k sont toutes de la forme

$$\frac{\partial^k f}{\partial^{k_1} x \partial^{k_2} y},\ k_1 + k_2 = k.$$