Université Claude Bernard - Lyon 1 **Semestre automne 2021-2022**
Préparation à l'agrégation interne J. Melleray

Petit papa Noël, calculons des différentielles....

Ces notes très denses reprennent/résument une grande partie du programme de calcul différentiel de l'agrégation interne. Pour les accompagner, on peut éventuellement regarder les vidéos sur le sujet que j'avais faites lors de la saison 1 du confinement (pour occuper des étudiants préparant l'externe), accessibles en cliquant sur le code ci-contre dans le pdf, ou en le scannant si vous utilisez une version imprimée.

1.1 Définitions et notations

Commençons par revoir la notion de fonction différentiable ; l'idée est d'approcher localement (quand c'est possible) une fonction compliquée par une fonction plus simple. Ici, près de x_0, on espère pouvoir approcher $f - f(x_0)$ par une fonction linéaire. Les fonctions linéaires de $\mathbb{R}$ dans $\mathbb{R}$ sont de la forme $h \mapsto \alpha h$, où $\alpha \in \mathbb{R}$; sur $\mathbb{R}$, essayer d'approcher $f(x_0+h)$ par $f(x_0)+\alpha h$, quand h tend vers 0, nous amènerait à considérer un taux d'accroissement, et à poser $\alpha = f'(0)$ (quand la dérivée existe). En dimension supérieure, la situation est plus compliquée, parce que l'espace des applications linéaires est plus riche.
Pour alléger un peu les notations et comme il ne devrait pas y avoir de risque de confusion, on notera toutes les normes par $\|\cdot\|$ (en particulier, il arrivera qu'on utilise la notation $\|\cdot\|$ à la fois pour désigner une norme sur $\mathbb{R}^n$ et une norme sur $\mathbb{R}^m$).

Définition 1.1

Soit $n, m \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, $x \in U$, $f \colon U \to \mathbb{R}^m$ et $g \colon U \to \mathbb{R}$ deux fonctions.
On dit que f *est négligeable devant* g *au voisinage de* x, et on note

$$f(y) = \underset{y \to x}{o}(g(y))$$

s'il existe un ouvert V tel que $x \in V \subseteq U$, et une fonction $\varepsilon \colon V \to \mathbb{R}^+$ telle que

- $\|f(y)\| = \varepsilon(y)|g(y)|$ pour tout $y \in V$.
- $\lim\limits_{y \to x} \varepsilon(y) = 0$.

Quand il n'y a pas de risque de confusion on notera simplement $f = o(g)$.

La notation $f = \underset{y \to x}{o}(1)$ signifie simplement que $\lim\limits_{y \to x} f(x) = 0$; si g ne s'annule pas sur un ouvert contenant x, $f = \underset{y \to x}{o}(g(y))$ est équivalent à $\lim\limits_{y \to x} \dfrac{f(y)}{g(y)} = 0$, et c'est comme cela qu'il faut y penser.

Définition 1.2

Soit $n, m \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, et $f\colon U \to \mathbb{R}^m$ une fonction.
On dit que f est *différentiable* en $x \in U$ s'il existe une application linéaire $L\colon \mathbb{R}^n \to \mathbb{R}^m$ telle que

$$f(y) = f(x) + L(y-x) + \underset{y\to x}{o}(\|y-x\|)$$

On écrira aussi

$$f(x+h) = f(x) + L(h) + \underset{h\to 0}{o}(\|h\|) \text{ ou simplement } f(x+h) = f(x) + L(h) + o(\|h\|)$$

Quand elle existe, l'application linéaire L ci-dessus est unique. En effet, si L_1, L_2 satisfont toutes les deux la condition ci-dessus, on a $(L_1 - L_2)(h) = o(h)$; s'il existe u tel que $x = L_1(u) - L_2(u) \neq 0$, alors on a $(L_1 - L_2)(tu) = tx$, qui n'est pas négligeable devant tu quand t tend vers 0.

Définition 1.3

Soit $n, m \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}^m$ une fonction, et $x \in U$.
Si f est différentiable en x, on appelle *différentielle de f en x* l'unique application linéaire L telle que

$$f(x+h) = f(x) + L(h) + o(\|h\|)$$

On note $L = Df(x)$.

En dimension 1, on retrouve la notion de dérivée : $f\colon \mathbb{R} \to \mathbb{R}$ est dérivable en x si, et seulement si, elle est différentiable en x ; et la différentielle de f en x est l'application $h \mapsto f'(x)h$.
Plus généralement, quand U est un ouvert de $\mathbb{R}$, $d \in \mathbb{N}^*$ et $f\colon U \to \mathbb{R}^d$ est une fonction différentiable, la différentielle de f est de la forme $h \mapsto hy$, où y est un vecteur de $\mathbb{R}^d$ (toutes les applications linéaires de $\mathbb{R}$ vers $\mathbb{R}^d$ sont de cette forme) ; et on note $y = f'(x)$.
Autrement dit, la différentielle de f en x est l'application $h \mapsto hf'(x)$, comme dans le cas des fonctions de $\mathbb{R}$ vers $\mathbb{R}$ (mais cette fois $f'(x)$ est un vecteur de $\mathbb{R}^d$).
Ici, il est important de remarquer que, quand on considère $f(x+h) - f(x)$, on peut travailler « coordonnée par coordonnée » : comprendre la différentiabilité des fonctions à valeurs réelles permet donc de comprendre la définition en général.

Exercice 1.4

Soit $n, m \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}^m$ une fonction, et $x \in U$.
On note $f = (f_1, \ldots, f_m)$, avec $f_i\colon U \to \mathbb{R}$ la i-ième coordonnée de f.
Montrer que f est différentiable en x si, et seulement si, chaque f_i est différentiable en x.
De plus, pour tout $h \in \mathbb{R}^n$, on a $Df(x)(h) = (Df_1(x)(h), \ldots, Df_m(x)(h))$.

Comme pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$, on peut calculer la différentielle d'une composée ; la formule donnant la différentielle de la composée est sans doute la chose la plus importante à comprendre, dans la partie sur le calcul différentiel, pour aborder les écrits (où le calcul différentiel joue rarement une place prépondérante, mais sait-on jamais...).

Théorème 1.5 (Règle de la chaîne)
Soit $n, m, p \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, V un ouvert de $\mathbb{R}^m$, ainsi que $g\colon U \to V$, $f\colon V \to \mathbb{R}^p$ deux fonctions.
Soit $x \in U$ tel que g soit différentiable en x et f soit différentiable en $g(x)$. Alors $f \circ g$ est différentiable en x, et on a

$$D(f \circ g)(x) = Df(g(x)) \circ Dg(x)$$

Démonstration. Pour $h \in \mathbb{R}^n$, de norme suffisamment petite pour que $x + h$ appartienne à U, on a

$$g(x+h) = g(x) + Dg(x)(h) + \varepsilon(h)\|h\|$$

où ε est une fonction qui tend vers 0 quand h tend vers 0.
Pour h assez petit, $g(x+h)$ est arbitrairement proche de $g(x)$, donc appartient à V ; et puisque f est différentiable en $g(x)$ il existe une fonction δ telle que

$$f(g(x)+v) = f(g(x)) + Df(g(x))(v) + \delta(v)\|v\|$$

où δ tend vers 0 quand v tend vers 0.
Notons $h' = Dg(x)(h) + \varepsilon(h)\|h\|$.
Comme Dg est linéaire continue, il existe une constante M telle que $\|Dg(x)(h)\| \leq M\|h\|$, en particulier h' tend vers 0 quand h tend vers 0 ; et $\|h'\| \leq (M+1)\|h\|$ pour h suffisamment petit.
Pour h suffisamment petit, on a donc

$$\begin{aligned} f(g(x+h)) &= f\left(g(x) + h'\right) \\ &= f(g(x)) + Df(g(x))(h') + \delta(h')\|h'\| \end{aligned}$$

Puisque $\|h'\| \leq (M+1)\|h\|$ pour h suffisamment petit, et δ tend vers 0 en 0, $\delta(h')\|h'\| = o(\|h\|)$.
Par ailleurs,

$$\begin{aligned} Df(g(x))(h') &= Df(g(x))(Dg(x)(h) + \varepsilon(h)\|h\|) \\ &= Df(g(x))(Dg(x)(h)) + \|h\|Df(g(x))(\varepsilon(h)) \\ &= Df(g(x)) \circ Dg(x)(h) + o(\|h\|). \end{aligned}$$

En regroupant tout cela, on obtient finalement

$$f(g(x+h)) = f(g(x)) + Df(g(x)) \circ Dg(x)(h) + o(\|h\|)$$

ce qui montre que $f \circ g$ est différentiable en x, de différentielle égale à $Df(g(x)) \circ Dg(x)$. □

1.2 Gradient, jacobienne, dérivées partielles

Il est particulièrement important de bien comprendre le cas des fonctions à valeurs réelles ; comme on l'a vu plus haut, si U est un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}^m$ est une fonction et qu'on note ses coordonnées $(f_1, \ldots, f_m)$, alors f est différentiable en $x \in U$ si, et seulement si, chaque f_i est différentiable en x ; et qu'on a alors la formule

$$Df(x)(h) = (Df_1(x)(h), \ldots, Df_m(x)(h))$$

Si U est un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}$ est différentiable en un certain $x \in U$, alors $Df(x)$ est une application linéaire de $\mathbb{R}^n$ vers $\mathbb{R}$; si on note $h = (h_1, \dots, h_n)$, il existe donc des réels $a_1, \dots, a_n$ tels que

$$f(x+h) = f(x) + \sum_{i=1}^{n} a_i h_i + o(\|h\|)$$

En notant $a = (a_1, \dots, a_n)$, on a donc la formule

$$Df(x)(h) = \langle a, h \rangle \qquad \text{(le produit scalaire de } a \text{ et de } h \text{)}$$

Définition 1.6

Soit U un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}$ une fonction et $x \in U$.
Si f est différentiable en x, on appelle *gradient de f en x*, et on note $\nabla f(x)$, l'unique élément de $\mathbb{R}^n$ tel que pour tout $h \in \mathbb{R}^n$ on ait

$$Df(x)(h) = \langle \nabla f(x), h \rangle$$

Le gradient de f pointe dans « la direction où f augmente le plus vite » ; et sa norme mesure à quelle vitesse f peut varier au voisinage de x_0. Ci-dessous, un cas particulier important (en géométrie par exemple, pour étudier des courbes ou des surfaces, mais aussi quand on manipule des équations différentielles et qu'on souhaite en dresser un portrait de phase) de la règle de la chaîne

Lemme 1.7

Soit I un ouvert de $\mathbb{R}$, U un ouvert de $\mathbb{R}^n$, $\gamma\colon I \to U$ et $f\colon U \to \mathbb{R}$.
On suppose que γ est dérivable en $t_0 \in I$ et f est différentiable en $\gamma(t_0)$.
Alors $t \mapsto f(\gamma(t))$ est dérivable en t_0, et on a la formule

$$(f \circ \gamma)'(t_0) = \langle \nabla f(\gamma(t_0)), \gamma'(t_0) \rangle$$

Démonstration. C'est une simple application de la règle de la chaîne.
En effet, γ est différentiable en t_0, de différentielle $h \mapsto h\gamma'(t_0)$; et f est différentiable en $\gamma(t_0)$, de différentielle $v \mapsto \langle \nabla f(\gamma(t_0)), v \rangle$.
La règle de la chaîne nous permet alors d'affirmer que $f \circ \gamma$ est différentiable en t_0, de différentielle égale à

$$h \mapsto \langle \nabla f(\gamma(t_0)), h\gamma'(t_0) \rangle = h\langle \nabla f(\gamma(t_0)), \gamma'(t_0) \rangle$$

Autrement dit, $f \circ \gamma$ est dérivable en t_0 et sa dérivée en t_0 est égale à $\langle \nabla f(\gamma(t_0)), \gamma'(t_0) \rangle$. □

Définition 1.8

Soit $n, m \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, et $f\colon U \to \mathbb{R}^m$ une fonction.
Si f est différentiable en $x \in U$, on appelle *matrice jacobienne* de f la matrice de $Df(x)$ relativement aux bases canoniques de $\mathbb{R}^n$ et $\mathbb{R}^m$

Les coefficients de la matrice jacobienne peuvent s'écrire à l'aide des dérivées partielles des coordonnées de f, sur lesquelles on va revenir.

Définition 1.9

Soit U un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}$ et $x \in U$.
Pour $h \in U$, on dit que f admet une *dérivée directionnelle* dans la direction h si l'application $t \mapsto f(x+th)$ est dérivable en 0.
Dans le cas particulier où $h = e_i$ (le i-ième vecteur de la base canonique), on utilise la notation

$$\frac{\partial f}{\partial x_i}(x)$$

pour désigner la dérivée directionnelle de f dans la direction e_i (quand elle existe) et on l'appelle *dérivée partielle de f par rapport à x_i*.

Si on écrit $x = (x_1, \ldots, x_n)$, la définition qu'on vient de donner revient à dire que $\dfrac{\partial f}{\partial x_i}(x)$ est la dérivée en 0 de $t \mapsto f(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_n)$.
Si jamais U est un ouvert de $\mathbb{R}^n$, et f est différentiable en $x \in U$, alors f admet des dérivées partielles dans toutes les directions ; en effet, $t \mapsto (x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_n)$ est dérivable, de dérivée $(0, \ldots, 0, 1, 0, \ldots, 0)$ (où le seul coefficient non nul est en i-ième position), c'est-à-dire e_i, le i-ième vecteur de base de la base canonique de $\mathbb{R}^n$.
La règle de la chaîne nous permet alors de voir que $t \mapsto f(x_1, \ldots, x_{i-1}, x_i + t, x_{i+1}, \ldots, x_n)$ est dérivable en 0, de dérivée égale à $Df(x)(e_i)$.
On vient d'établir la relation suivante, valide dès que f est différentiable en x :

$$\frac{\partial f}{\partial x_i}(x) = Df(x)(e_i)$$

En termes de matrice jacobienne, on voit que si U est un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}$ est différentiable en x, sa matrice jacobienne est la matrice ligne (matrice d'une application linéaire de $\mathbb{R}^n$ vers $\mathbb{R}$) donnée par

$$Jf(x) = \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(x) & \ldots & \dfrac{\partial f}{\partial x_n}(x) \end{pmatrix}$$

Par définition de la matrice d'une application linéaire, on a

$$\begin{aligned} Df(x)(h_1, \ldots, h_n) &= Jf(x) \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} \\ &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x) h_i \end{aligned}$$

Puisque $Df(x)(h) = \langle \nabla f(x), h \rangle$, on voit que $\nabla f(x)$ est donné par la formule

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right)$$

Vu ce qu'on a dit sur la différentielle d'une application à valeurs dans $\mathbb{R}^m$, il suit également que si U est un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}^m$ est différentiable en x, de coordonnées $(f_1, \ldots, f_m)$, la matrice jacobienne de f en x est donnée par

$$Jf(x) = \left(\frac{\partial f_i}{\partial x_j}(x) \right)_{1 \le i \le m, 1 \le j \le n}$$

Remarque 1.10

Il est très tentant de supposer que, si toutes les dérivées partielles de f existent en un point x, alors f est différentiable en ce point. Ce n'est pas le cas en général ! Pour le voir, considérez par exemple la fonction $f\colon \mathbb{R}^2 \to \mathbb{R}$ définie par

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

Alors f admet des dérivées partielles en $(0,0)$, qui sont nulles. Donc si f était différentiable en 0 sa différentielle serait la fonction nulle, et on aurait pour tout $u \in \mathbb{R}^2$ $f(u) = \|u\|_\infty \varepsilon(u)$, où ε tend vers 0 quand u tend vers 0. On devrait donc avoir

$$\lim_{u\to 0} \frac{f(u)}{\|u\|_\infty} = 0 .$$

Puisque $f(\frac{1}{n}, \frac{1}{n}) = \frac{1}{n}$, on voit qu'on n'a pas $\lim\limits_{u\to 0} \dfrac{f(u)}{\|u\|_\infty} = 0$, donc f n'est pas différentiable en 0.

Le problème dans l'exemple ci-dessus est que les dérivées partielles de f ne sont pas continues en 0.

Théorème 1.11

Soient $n, m \geq 1$ deux entiers, U un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}^m$ une fonction. On note de nouveau $f = (f_1, \ldots, f_m)$.
Si les dérivées partielles $\dfrac{\partial f_i}{\partial x_j}$ existent sur un voisinage de $x \in U$ **et sont continues** en x alors f est différentiable en $x \in U$.

Démonstration. Comme précédemment, il suffit de traiter le cas des fonctions à valeurs dans $\mathbb{R}$. Dans l'espoir que l'idée de la preuve soit claire, on va se contenter de la donner pour des fonctions de deux variables. Soit donc $f\colon U \to \mathbb{R}$ une fonction de deux variables dont les dérivées partielles existent et sont continues au voisinage d'un point $(x_0, y_0) \in U$. Pour tout (s,t) suffisamment petit, $(x_0 + s, y_0 + t)$ appartient à ce voisinage et on a

$$\begin{aligned} f(x_0+s, y_0+t) &= f(x_0+s, y_0) + (f(x_0+s, y_0+t) - f(x_0+s, y_0)) \\ &= f(x_0, y_0) + (f(x_0+s, y_0) - f(x_0, y_0)) + (f(x_0+s, y_0+t) - f(x_0+s, y_0)) \\ &= f(x_0, y_0) + s\frac{\partial f}{\partial x}(x_0+cs, y_0) + t\frac{\partial f}{\partial y}(x_0+s, y_0+dt) \end{aligned}$$

pour une certaine paire (c,d) d'éléments de $]0,1[$ (la dernière égalité résulte du théorème des accroissements finis, appliqué à des fonctions d'une variable réelle à valeurs réelles). Par continuité de $\dfrac{\partial f}{\partial x}$ et $\dfrac{\partial f}{\partial y}$, on a

$$\frac{\partial f}{\partial x}(x_0+cs, y_0) = \frac{\partial f}{\partial x}(x_0, y_0) + \varepsilon_1(s), \text{ où } \lim_{s\to 0} \varepsilon_1(s) = 0 .$$

De même, on a

$$\frac{\partial f}{\partial y}(x_0+cs, y_0+dt) = \frac{\partial f}{\partial y}(x_0,y_0) + \varepsilon_2(s,t), \text{ où } \lim_{(s,t)\to 0} \varepsilon_2(s,t) = 0 \ .$$

Tout ceci nous donne finalement

$$f(x_0+s, y_0+t) = f(x_0,y_0) + s\frac{\partial f}{\partial x}(x_0,y_0) + t\frac{\partial f}{\partial y}(x_0,y_0) + s\varepsilon_1(s) + t\varepsilon_2(s,t) \ .$$

L'application $(s,t) \mapsto s\dfrac{\partial f}{\partial x}(x_0,y_0) + t\dfrac{\partial f}{\partial y}(x_0,y_0)$ est linéaire, et on a

$$\lim_{(s,t)\to(0,0)} \frac{s\varepsilon_1(s) + t\varepsilon_2(s,t)}{\|(s,t)\|_\infty} = 0 \ .$$

On vient de démontrer que f est différentiable en (x_0,y_0), et que sa différentielle est l'application $(s,t) \mapsto s\dfrac{\partial f}{\partial x}(x_0,y_0) + t\dfrac{\partial f}{\partial y}(x_0,y_0)$. □

Exercice 1.12

Prouver le résultat précédent pour des fonctions de $\mathbb{R}^n$ dans $\mathbb{R}$ avec $n \geq 2$ un entier quelconque.

Exemple 1.13

L'espace $M_n(\mathbb{R})$ est un espace de dimension finie ; on va s'intéresser à la différentielle du déterminant, dans le cas $n \geq 2$ (qu'on va se contenter de calculer en une matrice inversible). La méthode de calcul qu'on présente ici n'est pas la plus élégante, mais elle a l'intérêt de passer par la notion de dérivée partielle pour calculer la différentielle.

Le déterminant de A s'écrit sous la forme d'un polynôme en les coefficients de A, ce qui nous permet de voir que $\det$ est de classe $\mathcal{C}^\infty$ (ici on pense aux coefficients de A comme étant les coordonnées de A dans une base de $M_n(\mathbb{R})$; explicitement la base en question est formée par les matrices $(E_{i,j})_{1\leq i,j\leq n}$, où $E_{i,j}$ a tous ses coefficients nuls, sauf celui sur la ligne i et la colonne j, qui vaut 1).

Ceci ne nous donne pas une formule explicite pour calculer la différentielle du déterminant en une matrice A donnée ; commençons par le cas où $A = I_n$ (la matrice identité de $M_n(\mathbb{R})$).

En reprenant les matrices $E_{i,j}$ introduites ci-dessus, on a pour tout $t \in \mathbb{R}$ la formule

$$\det(I + tE_{i,j}) = \begin{cases} 1 & \text{si } i \neq j \\ 1+t & \text{si } i = j \end{cases} = det(I) + t\,\mathrm{tr}(E_{i,j})$$

Dont la dérivée directionnelle de $\det$ dans la direction $E_{i,j}$ existe (on le savait déjà) et vaut $\mathrm{tr}(E_{i,j})$; autrement dit, $D\det(I)(E_{i,j}) = \mathrm{tr}(E_{i,j})$. Comme les $(E_{i,j})_{1\leq i,j\leq n}$ forment une base de $M_n(\mathbb{R})$, on en déduit par linéarité de $D\det$ que $D\det(I)(M) = \mathrm{tr}(M)$ pour toute $M \in M_n(\mathbb{R})$.

Soit maintenant A une matrice inversible, et $M \in M_n(\mathbb{R})$. On a

$$\begin{aligned} \det(A+M) &= \det(A(I_n + A^{-1}M)) \\ &= \det(A)\det(I_n + A^{-1}M) \\ &= \det(A)(1 + \mathrm{tr}(A^{-1}M) + o(\|A^{-1}M\|)) \end{aligned}$$

Ci-dessus, on a muni $M_n(\mathbb{R})$ d'une norme subordonnée $\|\cdot\|$; alors $\|A^{-1}M\| \leq \|A^{-1}\| \cdot \|M\|$ donc une fonction négligeable devant $\|A^{-1}M\|$ est aussi négligeable devant $\|M\|$.
On peut donc écrire

$$\det(A+M) = \det(A) + \det(A)\operatorname{tr}(A^{-1}M) + o(\|M\|)$$

On en conclut que $\det$ est différentiable en A, et que sa différentielle y est donnée par la formule

$$D\det(A)(M) = \det(A)\operatorname{tr}(A^{-1}M)$$

1.3 Différentielles itérées et fonctions de classe $\mathcal{C}^k$

Définition 1.14

Soit $n, m \in \mathbb{N}^*$, U un ouvert de $\mathbb{R}^n$, et $f\colon U \to \mathbb{R}^m$.
On dit que f est *de classe* $\mathcal{C}^1$ si chaque $\dfrac{\partial f_i}{\partial x_j}$ existe et est continue sur U.

De manière plus abstraite, cela revient à dire que l'application Df, définie sur U et à valeurs dans l'espace $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ des applications linéaires de $\mathbb{R}^n$ vers $\mathbb{R}^m$, est continue.

Définition 1.15

Soient $n, m \geq 1$ deux entiers, U un ouvert de $\mathbb{R}^n$ et $f = (f_1, \ldots, f_m)\colon U \to \mathbb{R}^m$ une fonction. Si toutes les dérivées partielles de chaque f_i existent et sont continues sur U alors on dit que f est *de classe* C^1 sur U.
Plus généralement, on définit par récurrence les applications de classe C^k : f est de classe C^{k+1} sur U si toutes les dérivées partielles des f_i existent sur U et sont de classe C^k.

La version abstraite de cette définition, également par récurrence, est la suivante : f est de classe $\mathcal{C}^{k+1}$ si et seulement si elle est différentiable et sa différentielle $Df\colon U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ est de classe $\mathcal{C}^k$.
Grâce à la règle de la chaîne, on vérifie par récurrence qu'une composée d'applications de classe C^k est encore de classe C^k.

Théorème 1.16 (Théorème de Schwarz)

Soient n un entier, U un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}$ une fonction de classe C^2 sur U. Alors on a, pour tout $i, j \in \{1, \ldots, n\}$ et tout $x \in U$:

$$\frac{\partial}{\partial x_i}\frac{\partial f}{\partial x_j}(x) = \frac{\partial}{\partial x_j}\frac{\partial f}{\partial x_i}(x)\ .$$

En d'autres termes, pour des fonctions de classe C^2, l'ordre dans lequel on effectue les dérivations n'a pas d'influence sur le résultat. On note alors $\dfrac{\partial^2 f}{\partial x_i \partial x_j}$ la fonction obtenue en dérivant une fois par rapport à x_i et une fois par rapport à x_j ; dans le cas où $i = j$, on note $\dfrac{\partial^2 f}{\partial x_i^2}$ la fonction obtenue en dérivant deux fois par rapport à x_i.

Démonstration. On va donner la preuve pour une fonction définie sur un ouvert U de $\mathbb{R}^2$ (le cas général s'en déduit facilement : si on considère deux dérivées partielles en un point, il n'y a que deux variables qui ne sont pas fixées !). Fixons $(x_0, y_0) \in U$. Pour (s,t) proche de 0, $(x_0+s, y_0+t) \in U$, et on pose

$$F(s,t) = f(x_0+s, y_0+t) - f(x_0+s, y_0) + f(x_0, y_0) - f(x_0, y_0+t) \ .$$

A s, t fixés, on pose $\varphi(x) = f(x, y_0+t) - f(x, y_0)$. On a a lors $F(s,t) = \varphi(x_0+s) - \varphi(x_0)$. Alors, φ est dérivable, de dérivée

$$\varphi'(x) = \frac{\partial f}{\partial x}(x, y_0+t) - \frac{\partial f}{\partial x}(x, y_0) \ .$$

On peut appliquer le théorème des accroissements finis (pour les fonctions de $\mathbb{R}$ dans $\mathbb{R}$) à φ sur le segment d'extrémités x_0, x_0+s, et obtenir $c_1 \in]0,1[$ tel que

$$\varphi(x_0+s) - \varphi(x_0) = s\varphi'(x_0+c_1 s) \ ,$$

c'est-à-dire

$$F(s,t) = s\Big(\frac{\partial f}{\partial x}(x_0+c_1 s, y_0+t) - \frac{\partial f}{\partial x}(x_0+c_1 s, y_0)\Big) \ .$$

Le théorème des accroissements finis, appliqué cette fois à $y \mapsto \dfrac{\partial f}{\partial x}(x_0+c_1 s, y)$ sur le segment d'extrémités y_0, y_0+t, nous donne un $d_1 \in]0,1[$ tel que

$$F(s,t) = st\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(x_0+c_1 s, y_0+d_1 t) \ .$$

En appliquant exactement le même raisonnement avec la fonction $\psi\colon y \mapsto f(x_0+s, y) - f(x_0, y)$, on obtient l'existence de $c_2, d_2 \in]0,1[$ tels que

$$F(s,t) = st\frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x_0+c_2 s, y_0+d_2 t) \ .$$

On a donc (dès que s et t sont tous les deux non nuls et suffisamment proches de 0) :

$$\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(x_0+c_1 s, y_0+d_1 t) = \frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x_0+c_2 s, y_0+d_2 t) \ .$$

En faisant tendre (s,t) vers 0 et en utilisant la continuité de $\dfrac{\partial}{\partial y}\dfrac{\partial f}{\partial x}$ et de $\dfrac{\partial}{\partial x}\dfrac{\partial f}{\partial y}$ (qui fait partie des hypothèses du théorème !), on obtient finalement

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

$\square$

Définition 1.17
Soit $n \geq 1$ un entier, A une partie de $\mathbb{R}^n$ et $f\colon A \to \mathbb{R}$ une fonction. On dit que f a un *extremum* en $x \in A$ si $f(x)$ est le maximum, ou le minimum, de f sur A. Si pour tout $y \in A \setminus \{x\}$ on a $f(y) < f(x)$ alors on dit que x est un maximum *strict* de f sur A. On définit de même la notion de minimum strict.
S'il existe un ouvert U contenant x et tel que $f_{|U \cap A}$ admette un extremum en x, on dit que x est un *extremum local* de f. On définit de même la notion d'*extremum local strict*

1.4 Points critiques, extrema et matrice hessienne

Théorème 1.18
Soit $n \geq 1$ un entier, U un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}$ une fonction différentiable en $x \in U$. Si f admet un extremum en $x \in U$, alors $\nabla f(x) = 0$. On dit alors que x est un *point critique* de f sur U.

Démonstration. Pour tout vecteur $u \in \mathbb{R}^n$, considérons la fonction $f_u\colon t \mapsto f(x+tu)$. Cette fonction est définie sur un intervalle ouvert contenant 0, et admet un extremum local en 0. Par conséquent, on doit avoir $f_u'(0) = 0$. Par la règle de la chaîne, on a

$$f_u'(0) = Df_x(u) = \langle \nabla f(x), u \rangle \ .$$

Par conséquent, on a $\langle \nabla f(x), u \rangle \ = 0$ pour tout $u \in \mathbb{R}^n$, et ceci n'est possible que si $\nabla f(x) = 0$. $\square$

La réciproque n'est pas vraie : on peut avoir $\nabla f(x) = 0$ sans que x soit un extremum local pour f ; c'est déjà le cas pour des fonctions de $\mathbb{R}$ dans $\mathbb{R}$, considérez par exemple $f\colon x \mapsto x^3$. Alors $f'(0) = 0$ mais 0 n'est pas un extremum local pour f.
Si on cherche les extrema d'une fonction f sur un ouvert U, on peut donc commencer par chercher les éléments x tels que $\nabla f(x) = 0$. Puisque les différentielles ne donnent qu'une information locale (elles ne disent rien sur ce que fait f « loin » de x !), on ne peut de toute façon pas espérer qu'elles nous suffisent à décider si un point est un extremum sur U. Par contre, on peut essayer d'utiliser des dérivées pour savoir si x est un extremum local pour f ; pour des fonctions de $\mathbb{R}$ dans $\mathbb{R}$, on utiliserait un développement limité à l'ordre 2 : si x est tel que $f'(x) = 0$ et f est deux fois dérivable en x, alors on a, par la formule de Taylor-Young,

$$f(x+h) = f(x) + \frac{f''(x)}{2}(y-x)^2 + o((y-x)^2) \ .$$

Ainsi, pour des fonctions de $\mathbb{R}$ dans $\mathbb{R}$, si $f'(x) = 0$ et $f''(x) > 0$, alors f admet un minimum local strict en x ; si $f'(x) = 0$ et $f''(x) < 0$, alors f admet un maximum local strict en x. Si $f'(x) = 0$ et $f''(x) = 0$, alors le développement limité à l'ordre 2 ne nous permet pas de conclure.
La situation est similaire pour les fonctions de plusieurs variables, mais la notion de dérivée seconde est plus compliquée : il y a beaucoup de dérivées secondes possibles, puisqu'on peut d'abord dériver par rapport à la variable x_i, puis dériver une nouvelle fois par rapport à la variable x_j... La bonne approche consiste à regrouper toutes ces dérivées dans une matrice, et à étudier les propriétés de cette matrice.

Définition 1.19

Soit $n \geq 1$ un entier, U un ouvert de $\mathbb{R}^n$ et $f\colon U \to \mathbb{R}$ une fonction de classe C^2 sur U. On définit la *matrice hessienne* $H(f)(x)$ de f en $x \in U$ comme étant la matrice dont le coefficient sur la i-ième ligne et la j-ième colonne est égal à $\frac{\partial^2 f}{\partial x_i \partial x_j}(x)$:

$$H(f)(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_i \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_i \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{pmatrix}$$

C'est une matrice carrée $n \times n$. Evidemment, quand $n = 1$, la matrice hessienne de f en x est une matrice 1×1, dont le coefficient vaut $f''(x)$, donc on ne fait que se compliquer la vie en y pensant comme étant une matrice - mais en dimension supérieure, il faut bien prendre en compte toutes les dérivées secondes possibles. Grâce au théorème de Schwarz, on sait que la matrice hessienne est symétrique ; par conséquent, elle est diagonalisable sur $\mathbb{R}$ et est la matrice d'une forme bilinéaire symétrique. Ce sont les propriétés de la forme quadratique associée, que je vais aussi noter $H(f)$, qui jouent un rôle dans l'étude des extrema de f.

On peut, dans ce contexte aussi, écrire une formule de Taylor-Young (le programme, et ces notes, s'arrêtent à l'ordre 2, mais bien sûr on pourrait aller à un ordre supérieur pour peu que la fonction étudiée soit différentiable suffisamment de fois) ; pour la comprendre, gardons les notations de la définition précédente, puis fixons $x \in U$ et r tel que $B(x,r) \subset U$. Fixons aussi un vecteur u de norme plus petite que r.

La fonction $g\colon t \mapsto f(x+tu)$ est bien définie sur $[-1,1]$, de classe $\mathcal{C}^2$, et on obtient tout d'abord

$$\forall t \quad g'(t) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x+tu)u_i$$

Puis, en dérivant une seconde fois :

$$g''(t) = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x+tu)u_i u_j$$

On obtient en particulier les formules

$$\begin{aligned} g'(0) &= \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x)u_i = Df(x)(u) \quad \text{(on s'y attendait !)} \\ g''(0) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}(x)u_i u_j = H(f)(u) \end{aligned}$$

En appliquant la formule de Taylor-Young à g en 0, on obtient ainsi

$$\begin{aligned} f(x+tu) &= g(t) \\ &= g(0) + tg'(0) + \frac{t^2}{2}g''(0) + o(t^2) \\ &= f(x) + tDf(x)(u) + \frac{t^2}{2}H(f)(u) + o(t^2) \\ &= f(x) + Df(x)(tu) + \frac{1}{2}H(f)(tu) + o(t^2) \end{aligned}$$

La formule de Taylor-Young à l'ordre 2 est une petite généralisation de ce qu'on vient de prouver (on ne s'est déplacé que le long d'un segment, il faudrait estimer le reste plus précisément) : si f est de classe $\mathcal{C}^2$ sur U, alors pour tout $x \in U$ on a

$$f(x+u) = f(x) + Df(x)(u) + \frac{1}{2}H(f)(u) + o(\|u\|^2)$$

On en tire en particulier le résultat suivant, utile pour la recherche d'extrema.

Théorème 1.20

Soit $n \geq 1$ un entier, U un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}$ une fonction de classe C^2 sur U, et $x \in U$ tel que $\mathrm{grad}(f)(x) = 0$. Alors :

1. Si la matrice hessienne de f en x est définie positive (i.e. si toutes ses valeurs propres sont strictement positives) alors x est un minimum local de f.
2. Si la matrice hessienne de f en x est définie négative (i.e. si toutes ses valeurs propres sont strictement négatives) alors x est un maximum local de f.
3. Si la matrice hessienne de f en x a une valeur propre strictement positive et une valeur propre strictement négative, alors x n'est pas un extremum local de f ; on dit alors que x est un *point selle* de f.
4. Si l'on n'est pas dans un des cas précédents, alors on ne peut pas savoir si x est, ou non, un extremum local de f.

Exercice 1.21

En appliquant la formule de Taylor-Young à l'ordre 2, démontrer ce résultat ; dans le cas particulier où $n = 2$, reformuler les conditions ci-dessus en utilisant la trace et le déterminant de la matrice hessienne de f en x.

1.5 Accroissements finis

On fixe deux entiers $n, m \in \mathbb{N}^*$, des normes $\|\cdot\|$ sur $\mathbb{R}^n, \mathbb{R}^m$ respectivement, et $|||\cdot|||$ la norme subordonnée associée.

Théorème 1.22 (Inégalité des accroissements finis)

Soit U un ouvert de $\mathbb{R}^n$, $f\colon U \to \mathbb{R}^m$ une application différentiable, et $x, y \in U$. On suppose que le segment

$$[x, y] = \{x + t(y - x)\colon t \in [0, 1]\}$$

est contenu dans U (cette hypothèse est en particulier vérifiée si U est convexe).
On suppose aussi que

$$M = \sup_{z \in [x,y]} |||Df(z)|||$$

est fini. Alors on a l'inégalité

$$\|f(y) - f(x)\| \leq M \cdot \|y - x\|$$

La démonstration est hors-programme ; elle n'est pas particulièrement difficile, et passe par l'étude d'une fonction auxiliaire d'un intervalle de $\mathbb{R}$ dans $\mathbb{R}$ (quand on se déplace sur $[x, y]$, on ne voit qu'un segment ...). La version au programme est le cas où f est $\mathcal{C}^1$, qui admet une autre peuve, plus courte, via le théorème fondamental de l'analyse.

Attention!

Dans le cas des fonctions à valeurs dans $\mathbb{R}$, vous connaissez déjà une *égalité* des accroissements finis ; si $f\colon [a, b] \to \mathbb{R}$ est dérivable sur $]a, b[$, continue sur $[a, b]$ alors il existe $c \in]a, b[$ tel que $f(b) - f(a) = f'(x)(b - a)$.
Pourquoi ne pas avoir établi une telle égalité dans le cadre des fonctions à valeurs dans $\mathbb{R}^n$? Tout simplement parce qu'elle est fausse en général.
La preuve de l'égalité des accroissements finis en dimension 1 est basée sur le lemme de Rolle, qui repose lui même sur l'existence d'un maximum pour toute fonction continue définie sur un segment, à valeurs réelles ; la notion de maximum n'a de sens que parce que $\mathbb{R}$ est muni d'un ordre. Et on ne peut pas munir $\mathbb{R}^n$ d'une structure d'ordre avec d'aussi bonnes propriétés que celles de l'ordre sur $\mathbb{R}$.
Voyons un contre-exemple explicite : la fonction $f\colon t \mapsto (\cos(t), \sin(t))$ est bien définie sur $\mathbb{R}$ à valeurs dans $\mathbb{R}^2$.
De plus, f est dérivable en tout t, et $f'(t) = (-\sin(t), \cos(t))$. Donc $f'(t)$ ne s'annule jamais.
Pourtant, on a $f(2\pi) - f(0) = 0$: il ne peut pas exister de c tel que $f(2\pi) - f(0) = 2\pi f'(c)$.

Corollaire 1.23

Soit U un ouvert connexe de $\mathbb{R}^n$, et $f\colon U \to \mathbb{R}^n$ une fonction différentiable de différentielle nulle. Alors f est constante sur U.

Démonstration. Soit $x_0 \in U$. Montrons que $A = \{x \in U\colon f(x) = f(x_0)\}$ est ouvert et fermé (dans U) et non vide ; on en conclura que c'est U tout entier, puisque U est supposé connexe. Le fait que A soit fermé est une conséquence de la continuité de f ; et $A \neq \emptyset$ puisque $x_0 \in A$.
Montrons que A est ouvert : soit $y \in A$. Comme U est ouvert, il existe $r > 0$ tel que la boule $B(y, r)$ soit contenue dans U. Et pour tout $z \in B(y, r)$, l'inégalité des accroissements finis couplée à l'hypothèse selon laquelle Df est nulle sur U donne $\|f(y) - f(z)\| \leq 0$, autrement dit $f(z) = f(y) = f(x_0)$. Par conséquent $B(y, r) \subset A$, et on conclut comme espéré que A est ouvert. $\square$

1.6 Difféomorphismes, inversion locale, fonctions implicites

Définition 1.24

Soient U, V deux ouverts de $\mathbb{R}^n$, $k \geq 1$ et $f: U \to \mathbb{R}^n$ une fonction. On dit que f est un *difféomorphisme de classe C^k* de U sur V si :

1. f est une bijection de U sur V (i.e. f est injective sur U, et $f(U) = V$)
2. f est de classe C^k sur U.
3. f^{-1} est différentiable sur U.

Théorème 1.25

Soient U, V deux ouverts de $\mathbb{R}^n$, $k \geq 1$ et $f: U \to \mathbb{R}^n$ un difféomorphisme de classe C^k de U sur V. Alors l'application inverse f^{-1} de f est un difféomorphisme de classe C^k de V sur U, la différentielle de f est inversible en tout $x \in U$, et on a, pour tout $x \in U$:

$$Df^{-1}(f(x) = (Df(x))^{-1} \ .$$

Démonstration. La formule permettant de calculer la différentielle de f^{-1} est une conséquence immédiate de la règle de la chaîne : pour tout $x \in U$ on a, par définition de f^{-1}, que $f^{-1} \circ f(x) = x$. En différentiant cette égalité, et en utilisant le fait que la différentielle de l'application $x \mapsto x$ est l'application identité, on obtient, pour tout $x \in U$:

$$Df^{-1}(f(x)) \circ Df(x) = I \ .$$

On en déduit= que $Df(x)$ est inversible, d'inverse $Df^{-1}(f(x))$; donc $Df^{-1}(f(x)) = (Df(x))^{-1}$. Reste à vérifier que f^{-1} est de classe C^k. Pour cela, on rappelle que la *comatrice* $c(A)$ d'une matrice carrée A est la matrice dont le coefficient sur la i-ième ligne et la j-ième colonne est le déterminant de la matrice obtenue en enlevant de A sa i-ième ligne et sa j-ième colonne ; si A est inversible, alors l'inverse de A est égale à $\frac{1}{\det(A)} c(A)^T$. Les coefficients apparaissant dans la matrice jacobienne de $Df^{-1}(y)$ sont donc des quotients de de classe $\mathcal{C}^{k-1}$ dont le dénominateur ne s'annule pas ; ce sont donc des fonctions de classe $\mathcal{C}^{k-1}$, ce qui montre que f^{-1} est de classe C^k.

□

Remarque 1.26

Si U est un ouvert non vide de $\mathbb{R}^n$, et V est un ouvert de $\mathbb{R}^m$, alors il ne peut exister de bijection différentiable et d'inverse différentiable de U sur V que si $n = m$: en effet, la différentielle de f en un point $x \in U$ quelconque devrait être une bijection linéaire de $\mathbb{R}^n$ sur $\mathbb{R}^m$, et une telle application ne peut exister que si $n = m$.
Plus généralement, il est impossible qu'un ouvert U de $\mathbb{R}^n$ non vide soit *homéomorphe* à un ouvert V de $\mathbb{R}^m$ si $n \neq m$. C'est le *théorème d'invariance du domaine*, beaucoup plus général que celui qu'on vient d'énoncer, et hors de portée pour nous (pouvoir démontrer cela est une bonne raison pour développer des notions de topologie algébrique).

Théorème 1.27 (Théorème d'inversion locale)

Soit U un ouvert de $\mathbb{R}^n$, et $f\colon U \to \mathbb{R}^n$ une fonction de classe C^1. Supposons que $x \in U$ soit tel que $Df(x)$ soit inversible. Alors il existe deux ouverts U_1, V_1 tels que $x \in U_1$, $f(x) \in V_1$, et $f_{|U_1}$ soit un difféomorphisme de classe C^1 de U_1 sur V_1.

La preuve de ce théorème est hors-programme.

Théorème 1.28 (Théorème d'inversion globale)

Soit U un ouvert de $\mathbb{R}^n$, et $f\colon U \to \mathbb{R}^n$ une fonction de classe C^1 et injective. Si la différentielle de f est inversible en tout point de U, alors $f(U)$ est ouvert et f est un difféomorphisme de classe C^1 de U sur $f(U)$.

Remarque 1.29

L'hypothèse selon laquelle la différentielle de f est inversible en tout point de U est essentielle : sans cela le théorème est faux. Par exemple, l'application $x \mapsto x^3$ est une bijection de classe C^1 de $\mathbb{R}$ sur $\mathbb{R}$, mais sa fonction réciproque $x \mapsto x^{1/3}$ n'est pas dérivable en 0.

Démonstration. Montrons d'abord que $f(U)$ est ouvert : si $y \in f(U)$, alors il existe $x \in U$ tel que $f(x) = y$. Par le théorème d'inversion locale, il existe un ouvert $U_1 \ni x$ et un ouvert $V_1 \ni y$ tels que $f_{|U_1}$ soit un difféomorphisme de U_1 sur V_1. Alors V_1 est ouvert, contient y, et est contenu dans $f(U)$. Ceci prouve que $f(U)$ est ouvert.
Ensuite, le théorème d'inversion locale nous assure que f^{-1} est différentiable en $f(x)$ pour tout $x \in U$; autrement dit f^{-1} est différentiable en y pour tout $y \in f(U)$, et toutes les hypothèses définissant un difféomorphisme sont vérifiées. $\square$

Une raison pour laquelle les difféomorphismes de classe $\mathcal{C}^1$ sont importants pour nous est que ce sont les fonctions qu'on utilise pour faire des changements de variables. Par exemple, l'application $\varphi\colon (r,\theta) \mapsto (r\cos(\theta), r\sin(\theta))$ (le passage en *coordonnées polaires*) réalise un difféomorphisme de classe $\mathcal{C}^1$ de $]0,+\infty[\times]0,2\pi[$ sur $\mathbb{R}^2 \setminus \{(x,0)\colon x \geq 0\}$ (on doit enlever une demi-droite pour obtenir une bijection, mais ce n'est pas grave du point de vue du calcul d'intégrale puisque l'aire de cette demi-droite est nulle...).
Il nous reste à voir un dernier énoncé. Si on étudie l'ensemble des solutions de $f(x_1,\ldots,x_n) = a$, avec $f\colon \mathbb{R}^n \to \mathbb{R}^m$ et $a \in \mathbb{R}^m$, on s'attend à voir un objet « de dimension $n-m$ » : l'équation considérée impose m contraintes à un vecteur qui a n degrés de libertés.
Une formalisation de cette idée passerait par la notion de sous-variété (hors programme). Le théorème des fonctions implicites nous dit que, sous les bonnes hypothèses sur f, si l'on regarde les solutions de notre équation près d'une solution x, on peut (sous les bonnes hypothèses sur f) isoler $n-m$-coordonnées de x et écrire les autres coordonnées comme une fonction de ces $n-m$ coordonnées privilégiées. Par exemple, quand on regarde un cercle, on a affaire à l'équation $x^2 + y^2 = 1$. Près d'une solution avec $y \neq 0$, on peut écrire $y = \sqrt{1-x^2}$ ou $y = -\sqrt{1-x^2}$; et près d'une solution avec $y = 0$, c'est-à-dire près de $(1,0)$ ou $(-1,0)$, on peut écrire $x = \sqrt{1-y^2}$ ou $x = -\sqrt{1-y^2}$.

Théorème 1.30 (Théorème des fonctions implicites)

Soient $n, m \geq 1$ deux entiers, U un ouvert de $\mathbb{R}^n \times \mathbb{R}^m$ et $f\colon U \to \mathbb{R}^m$ une fonction de classe C^k. Soit $(x_0, y_0) \in U$ tel que $f(x_0, y_0) = 0$, et la différentielle de l'application $y \mapsto f(x_0, y)$ soit inversible en y_0. Alors il existe un ouvert O contenant x_0, un ouvert W contenant (x_0, y_0), et une application $\varphi\colon U_0 \to \mathbb{R}^m$ de classe C^k tels que :

$$\forall (x,y) \in W \; f(x,y) = 0 \Leftrightarrow x \in O \text{ et } \varphi(x) = y \; .$$

(en particulier $\varphi(x_0) = y_0$)
On dit alors que l'équation $f(x,y) = 0$ *définit implicitement* y *en fonction de* x au voisinage de (x_0, y_0).

En utilisant la règle de la chaîne, on peut calculer les différentielles successives de φ en x_0, voir les exemples ci-dessous.
Si l'on considère l'exemple de l'application $f\colon (x,y) \mapsto x^2 + y^2 - 1$, on voit que, au voisinage de $(1,0)$, l'équation $f(x,y) = 0$ ne peut pas définir y comme une fonction de x : pour chaque x proche de 1 il y a deux solutions pour y. La raison pour laquelle le théorème ne s'applique pas est que la dérivée partielle de f par rapport à y en ce point vaut 0...
La preuve, qui utilise le théorème d'inversion locale, est hors programme.
Discutons maintenant quelques exemples.

Exemple 1.31

Considérons la fonction $f\colon (x,y) \mapsto xy + \ln(xy) - 1$, définie sur $U = \{(x,y)\colon x > 0, y > 0\}$. Alors f est de classe C^∞ sur U, on a $f(1,1) = 0$, et $\dfrac{\partial f}{\partial y}(x,y) = x + \dfrac{1}{y}$ donc $\dfrac{\partial f}{\partial y}(1,1) = 2 \neq 0$. Par conséquent, le théorème des fonctions implicites nous permet d'affirmer que l'équation $xy + \ln(xy) = 1$ définit implicitement y comme une fonction φ de x au voisinage de $(1,1)$. Si on doit calculer la dérivée de φ en 1, on écrit : $x\varphi(x) + \ln(x\varphi(x)) = 1$, ce qui se dérive en

$$\varphi(x) + x\varphi'(x) + \frac{1}{x} + \frac{\varphi'(x)}{\varphi(x)} = 0 \; .$$

Comme $\varphi(1) = 1$, l'équation ci-dessus donne $2 + 2\varphi'(1) = 0$, donc $\varphi'(1) = -1$. Ceci nous permettrait par exemple de trouver l'équation de la tangente à la courbe définie par l'équation $xy + \ln(xy) = 1$ au voisinage de $(1,1)$.

Exemple 1.32

Considérons maintenant l'équation $2xy - z + 2xz^3 = 5$. Cette équation définit-elle implicitement z comme une fonction de (x,y) au voisinage de $(1,2,1)$?
Pour le savoir, on pose $f(x,y,z) = 2xy - z + 2xz^3 - 5$, et on calcule la matrice jacobienne de f en (x,y,z), qui vaut $\begin{pmatrix} 2y + 2z^3 & 2x & -1 + 6xz^2 \end{pmatrix}$. En $(1,2,1)$, cela donne $\begin{pmatrix} 4 & 2 & 5 \end{pmatrix}$. Puisque $5 \neq 0$, on voit que l'équation définit bien implicitement z comme fonction de (x,y) au voisinage de $(1,2,1)$. Notons $z = \varphi(x,y)$ et essayons de calculer $\dfrac{\partial^2 \varphi}{\partial y^2}(1,2)$. Pour cela on doit d'abord calculer

les dérivées partielles de φ en $(1,2)$, ce qu'on fait en dérivant l'équation $f(x,y,\varphi(x,y)) = 0$, qui donne par la règle de la chaîne :

$$\begin{pmatrix} 2y + 2\varphi^3(x,y) & 2x & -1 + 6x\varphi^2(x,y) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \dfrac{\partial\varphi}{\partial x}(x,y) & \dfrac{\partial\varphi}{\partial y}(x,y) \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix} .$$

On a donc le système suivant :

$$\begin{cases} 2y + 2\varphi^3(x) + (6x\varphi^2(x) - 1)\dfrac{\partial\varphi}{\partial x}(x,y) & = 0 \\ 2x + (6x\varphi^2(x) - 1)\dfrac{\partial\varphi}{\partial y}(x,y) & = 0 \end{cases}$$

On en déduit les formules suivantes :

$$\frac{\partial\varphi}{\partial x}(x,y) = \frac{2y + 2\varphi^3(x)}{1 - 6x\varphi^2(x)} \quad \text{et} \quad \frac{\partial\varphi}{\partial y}(x,y) = \frac{2x}{1 - 6x\varphi^2(x)} .$$

Ces deux équations nous donnent $\dfrac{\partial\varphi}{\partial x}(1,2) = -\dfrac{6}{5}$ et $\dfrac{\partial\varphi}{\partial y}(1,2) = -\dfrac{2}{5}$.
En redérivant par rapport à y l'équation donnant $\dfrac{\partial\varphi}{\partial y}(x,y)$, on obtient :

$$\frac{\partial^2\varphi}{\partial y^2}(x,y) = 2x\frac{12x\frac{\partial\varphi}{\partial y}(x,y)}{(1 - 6x\varphi^2(x))^2} .$$

En $(1,2)$, on obtient $\dfrac{\partial^2\varphi}{\partial y^2}(1,2) = -\dfrac{48}{125}$.

Exemple 1.33

Un dernier exemple, pour une fonction f de trois variables, à valeurs dans $\mathbb{R}^2$. On considère le système d'équations suivant :

$$\begin{cases} 4xy + 2xz + y + 4y^2 & = 0 \\ x^3y + xz + yz - z & = 0 \end{cases}$$

Essaons de voir si ce système définit (y,z) comme une fonction de x au voisinage de $(0,0,0)$. Pour cela, on considère l'application $f\colon \mathbb{R}^3 \to \mathbb{R}^2$ définie par

$$f(x,y,z) = (4xy + 2xz + y + 4y^2, x^3y + xz + yz - z)$$

Cette fonction est de classe C^∞, et sa matrice jacobienne en (x,y,z) vaut

$$\begin{pmatrix} 4y + 2z & 4x + 8y + 1 & 2x \\ 3x^2y + z & x^3 + z & x + y - 1 \end{pmatrix}$$

En $(0,0,0)$, cette matrice vaut $\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.
Pour décider si cette équation définit implicitement (y,z) comme fonction de x au voisinage de $(0,0,0)$, il nous faut donc décider si la matrice $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ est inversible ; son déterminant vaut -1, donc c'est bien le cas. On peut donc écrire $(y,z)=\varphi(x)$ au voisinage de $(0,0,0)$.
Essayons maintenant de calculer la dérivée de φ en 0 : si on note $\varphi=(\varphi_1,\varphi_2)$; le fait que $f(x,\varphi_1(x),\varphi_2(x))=0$ donne, par la règle de la chaîne, $Df_{(x,\varphi_1(x),\varphi_2(x))}(1,\varphi_1'(x),\varphi_2'(x))=0$, d'où le système suivant :

$$\begin{cases} \varphi_1'(0) & = 0 \\ \varphi_2'(0) & = 0 \end{cases}$$

Autrement dit, $\varphi_1'(0)=\varphi_2'(0)=0$. S'il avait fallu calculer les dérivées de φ en 0 à un ordre supérieur, alors on aurait dû écrire le système suivant (toujours donné par la règle de la chaîne) :

$$\begin{cases} 4\varphi_1(x)+2\varphi_2(x)+(4x+8\varphi_1(x)+1)\varphi_1'(x)+2x\varphi_2'(x) & = 0 \\ 3x^2\varphi_1(x)+\varphi_2(x)+(x^3+\varphi_2(x))\varphi_1'(x)+(x+\varphi_1(x)-1)\varphi_2'(x) & = 0 \end{cases}$$

Pour calculer $\varphi''(0)$, par exemple, il aurait fallu dériver ce système, puis l'écrire en 0 en y substituant le fait que $\varphi_1'(0)=\varphi_2'(0)$, pour obtenir un nouveau système de deux équations à deux inconnues $\varphi_1''(0)$, $\varphi_2''(0)$.
Plutôt que de continuer ces calculs, une dernière question : est-ce que l'équation $f(x,y,z)=(0,0)$ définit implicitement (x,y) en fonction de z au voisinage de $(0,0,0)$? Pour déterminer cela, il nous faut décider si la matrice $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ est inversible. Ce n'est clairement pas le cas (une colonne ne contient que des 0), donc l'équation ne définit pas implicitement (x,y) comme fonction de z au voisinage de $(0,0,0)$.

1.7 Quelques exercices

Exercice 1. Soit $n > 0$ un entier. Montrer qu'une norme N sur $\mathbb{R}^n$ ne peut pas être différentiable en 0.

Exercice 2. Soit $\|\cdot\|$ la norme euclidienne sur $\mathbb{R}^2$ et U un ouvert convexe de $\mathbb{R}^2$. Soit $f\colon U \to \mathbb{R}^2$ une fonction qui vérifie

$$\forall x, y \in U \ \|f(x) - f(y)\| \leq \|x - y\|^2 \ .$$

Montrer que f est constante sur U (on pourra commencer par montrer que f est différentiable et calculer sa différentielle).

Exercice 3. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(x, y) = \begin{cases} x^2y^2 \sin\left(\dfrac{1}{x}\right) & \text{si } x \neq 0 \\ 0 & \text{si } x = 0\ . \end{cases}$$

Montrer que la fonction f est différentiable en tout point de $\mathbb{R}^2$ mais que $\dfrac{\partial f}{\partial x}$ et $\dfrac{\partial f}{\partial y}$ ne sont pas continues en certains points de $\mathbb{R}^2$.

Exercice 4. On considère les fonctions $f\colon \mathbb{R}^2 \to \mathbb{R}^3$ et $g\colon \mathbb{R}^3 \to \mathbb{R}$ définies par

$$f(x, y) = (\sin(xy), \cos x, \exp(y^2)), \quad g(u, v, w) = uvw.$$

1. Calculer explicitement $g \circ f$.
2. En utilisant l'expression trouvée en (1), calculer les dérivées partielles de $g \circ f$.
3. Déterminer les matrices jacobiennes $M_f(x, y)$ et $M_g(u, v, w)$ de f et de g.
4. Retrouver le résultat de (2.) en utilisant un produit approprié de matrices jacobiennes.

Exercice 5. Soit $g\colon \mathbb{R}^3 \to \mathbb{R}^2$ une fonction différentiable, telle que $g(1, -1, 2) = (-1, 5)$ et la matrice jacobienne de g en $(1, -1, 2)$ soit égale à $\begin{pmatrix} 1 & -1 & 0 \\ 4 & 0 & 2 \end{pmatrix}$. Soit f la fonction de $\mathbb{R}^2$ dans $\mathbb{R}^2$ définie par $f(x, y) = (xy, 3x^2 - 2y + 3)$. Calculer les dérivées partielles de chaque coordonnée de $f \circ g$ en $(1, -1, 2)$.

Exercice 6. Soit $f(x, y) = \arcsin\left(\dfrac{1 + xy}{\sqrt{(1 + x^2)(1 + y^2)}}\right)$ et $g(x, y) = \arctan x - \arctan y$.

1. Vérifier que f est définie sur $\mathbb{R}^2$.
2. Calculer les dérivées partielles premières de f et de g.
3. Simplifier f à l'aide de g.

Exercice 7. Soit $f(x, y) = \dfrac{x^3y}{x^2 + y^2}$ si $(x, y) \neq 0$ et $f(0, 0) = 0$.

1. Étudier la continuité de f et de ses dérivées partielles premières sur $\mathbb{R}^2$.
2. Montrer que $\dfrac{\partial^2 f}{\partial x \partial y}(0, 0) \neq \dfrac{\partial^2 f}{\partial y \partial x}(0, 0)$.

Exercice 8. Dans cet exercice, on cherche à déterminer toutes les fonctions continues f telles que

$$\forall x \in \mathbb{R} \quad f(x) - 2\int_0^x f(t)\cos(x-t)\,dt = 1$$

1. On suppose que $g\colon \mathbb{R} \to \mathbb{R}$ est continue, $\theta\colon \mathbb{R}^2 \to \mathbb{R}$ est de classe $\mathcal{C}^1$ et $\alpha, \beta\colon \mathbb{R} \to \mathbb{R}$ sont de classe $\mathcal{C}^1$. Calculer la dérivée de $x \mapsto \displaystyle\int_{\alpha(x)}^{\beta(x)} g(t)\theta(x,t)dt$.
   Indication : quand une variable joue deux rôles différents, il peut être intéressant de considérer une fonction de deux variables $(x,y) \mapsto \Phi(x,y)$ qui soit telle que la fonction qu'on étudie s'écrive sous la forme $x \mapsto \Phi(x,x)$...
2. On considère une fonction f comme ci-dessus. Montrer que $f(0) = 1$, puis que f est dérivable et $f'(0) = 2$, et enfin que f est deux fois dérivable et est solution de l'équation différentielle $y'' = 2y' - y + 1$.
3. En déduire la seule valeur possible pour f, et vérifier que cette fonction est bien solution de l'équation considérée au début de l'exercice.

Exercice 9. Étudier les points critiques de la fonction $f\colon (x,y,z)\colon x\ln(x) + y\ln(y) + z\ln(z)$ sur $(]0,+\infty[)^3$; déterminer les extrema éventuels de cette fonction.

Exercice 10. Montrer que, parmi tous les triangles dans $\mathbb{R}^2$ dont le cercle circonscrit est le cercle unité, les triangles équilatéraux sont ceux dont le périmètre est maximal.
(On pourra se ramener à la recherche du maximum d'une fonction de deux variables bien choisie...)
Généralisation aux polygones convexes dont tous les sommets sont sur le cercle unité ?

Exercice 11. Rechercher les points critiques de $(x,y) \mapsto y(x^2 + (\ln(y))^2)$ sur $\mathbb{R}^2$, et étudier leur nature.
Même question sur $\mathbb{R}^3$ pour $(x,y,z) \mapsto z(e^x - 1) - y^2$.

Exercice 12. Trouver le plus grand ouvert U tel que l'application $(x,y,z) \mapsto (x, xy, xyz)$ définisse un difféomorphisme de classe C^1 de U dans $\mathbb{R}^3$. Déterminer l'image de U par cette application.

Exercice 13. On pose $U = \{(x,y) \in \mathbb{R}^2\colon x \neq 0 \text{ et } y \neq 0\}$.
Montrer que l'application $(x,y) \mapsto (x^2y, xy^2)$ définit un difféomorphisme de classe C^1 de U sur U. Cette application est-elle un difféomorphisme de classe C^1 de $\mathbb{R}^2$ dans $\mathbb{R}^2$?

Exercice 14. Montrer que l'égalité $2e^{x+y} + y - x = 0$ définit $y = \varphi(x)$ au voisinage de $(1,-1)$. Calculer $\varphi'(1)$ et $\varphi''(1)$.

Exercice 15. Considérons la fonction $F\colon \mathbb{R}^4 \to \mathbb{R}^2$ définie par

$$F(x_1,x_2,y_1,y_2) = (x_2y_2 - x_1\cos(y_1), x_2\sin(y_1) + x_1y_2 - 1)\ .$$

Est-ce que l'équation $F(x_1,x_2,y_1,y_2)$ définit implicitement ϕ, ψ telles que $y_1 = \phi(x_1,x_2)$ et $y_2 = \psi(x_1,x_2)$ au voisinage de $(1,1,\frac{\pi}{2},\frac{\pi}{4})$? Si oui, calculer $\dfrac{\partial\phi}{\partial x_1}(1,1)$ et $\dfrac{\partial\phi}{\partial x_2}(1,1)$.
Même question pour l'équation $F(x_1,x_2,y_1,y_2) = (5,1)$ au voisinage de $(0,2,\frac{\pi}{2},\frac{5}{2})$.