Calcul différentiel : notes de cours et compléments

2. Théorèmes généraux

2.1. Opérations algébriques.

On désigne par E, F, F_1, F_2, G des $\mathbf{R}-$ev de dimensions finies, U ouvert de E, x_0 un point de U.

Proposition 6 (Opérations algébriques).

1. *Soient $f, g : U \to F$, $\lambda, \mu \in \mathbf{R}$ des scalaires.*
   Supposons
   hyp *f et g différentiables en x_0 (resp. diff. sur U, C^1 sur U)*
   Alors
   $C\ell$ *$\lambda f + \mu g$ est différentiable en x_0 (resp. diff. sur U, C^1 sur U), avec*
   $\boxed{\mathrm{d}(\lambda f + \mu g)(x_0) = \lambda \mathrm{d}f(x_0) + \mu \mathrm{d}g(x_0)}$.
2. *Plus généralement, soient $f, g : U \to F$, $\lambda, \mu\, U \to \mathbf{R}$ des fonctions scalaires.*
   Supposons
   hyp *f, g, λ et μ différentiables en x_0 (resp. diff. sur U, C^1 sur U)*
   Alors
   $C\ell$ *$x \mapsto \lambda(x)f(x) + \mu(x)g(x)$ est différentiable en x_0 (resp. diff. sur U, C^1 sur U), avec pour tout $h \in E$*
   $\boxed{\mathrm{d}(\lambda f + \mu g)(x_0)(h) = \lambda(x_0).\mathrm{d}f(x_0)(h) + \mathrm{d}\lambda(x_0)(h).f(x_0) + \mu(x_0).\mathrm{d}g(x_0)(h) + \mathrm{d}\mu(x_0)(h).g(x_0)}$.
3. *Soient $f_i : U \to F_i$ $(i = 1, 2)$ deux applications et $B : F_1 \times F_2 \to G$ une application bilinéaire.*
   Supposons
   hyp *f_1 et f_2 différentiables en x_0 (resp. diff. sur U, C^1 sur U)*
   Alors
   $C\ell$ *$f : x \mapsto B(f_1(x), f_2(x))$ est différentiable en x_0 (resp. diff. sur U, C^1 sur U), avec pour tout $h \in E$*
   $\boxed{\mathrm{d}f(x_0)(h) = B(\mathrm{d}f_1(x_0)(h), f_2(x_0)) + B(f_1(x_0), \mathrm{d}f_2(x_0)(h))}$.
4. *Soient $f, g : U \to \mathbf{K}$.*
   Supposons
   hyp1 *f et g différentiables en x_0 (resp. diff. sur U, C^1 sur U)*
   hyp2 $g(x_0) \neq 0$
   Alors
   $C\ell$ *f/g est différentiable en x_0 (resp. diff. sur U, C^1 sur U), avec pour tout $h \in E$*
   $$\boxed{\mathrm{d}\left(\frac{f}{g}\right)(x_0)(h) = \frac{g(x).\mathrm{d}f(x_0)(h) - f(x_0).\mathrm{d}g(x_0)(h)}{g(x_0)^2}}.$$
5. *Soit $f : U \to F$ une application et $V \subset U$ un ouvert de U. Supposons*
   hyp *f différentiable en x_0 (resp. diff. sur U, C^1 sur U)*
   Alors
   $C\ell$ *La restriction de f à V, $f_{|V}$, est différentiable en x_0 (resp. diff. sur V, C^1 sur V), avec*
   $\boxed{\mathrm{d}f_{|V}(x_0) = \mathrm{d}f(x_0)}$.

Démonstration.

1. Immédiat en partant de la définition, par somme de DL.
2. Se démontre à partir du point 3 : l'application $B : \mathbf{R} \times F \to F$, $(\lambda, y) \mapsto \lambda y$ est en effet bilinéaire, donc par composition $x \mapsto B(\lambda(x), f(x))$ est différentiable en x_0 de différentielle $h \mapsto B(\mathrm{d}\lambda(x_0)(h), f(x_0)) + B(\lambda(x_0), \mathrm{d}f(x_0)(h)) = \lambda(x_0).\mathrm{d}f(x_0)(h) + \mathrm{d}\lambda(x_0)(h).f(x_0)$. On conclut par somme de différentielles.
3. On écrit par bilinéarité
$$f(x+h) = B(f_1(x+h), f_2(x+h)) = B\Big(f_1(x_0)+\mathrm{d}f_1(x_0)(h)+o(h), f_2(x_0)+\mathrm{d}f_2(x_0)(h)+o(h)\Big)$$
$$= B\Big(f_1(x_0), f_2(x_0)\Big)+B\Big(\mathrm{d}f_1(x_0)(h), f_2(x_0)\Big)+B\Big(f_1(x_0), \mathrm{d}f_2(x_0)(h)\Big)+$$
$$B\Big(f_1(x_0)+\mathrm{d}f_1(x_0)(h)+o(h), o(h)\Big)+B\Big(o(h), f_2(x_0)+\mathrm{d}f_2(x_0)(h)+o(h)\Big)+B\Big(\mathrm{d}f_1(x_0)(h), \mathrm{d}f_2(x_0)(h)\Big).$$
Les trois derniers termes sont des $o(h)$ en vertu de l'inégalité $\|B(x_1, x_2)\| \leq C\|x_1\|.\|x_2\|$. Par exemple pour g continue en x_0 :
$\|B(g(x), \|h\|\varepsilon(h))\| \leq C\|g(x)\|.\|h\|.\|\varepsilon(h))\| \leq C'\|h\|.\|\varepsilon(h))\|$ par bornitude de g au voisinage de x_0 et
$\|B\Big(\mathrm{d}f_1(x_0)(h), \mathrm{d}f_2(x_0)(h)\Big)\| \leq C\|\mathrm{d}f_1(x_0)(h)\|.\|\mathrm{d}f_2(x_0)(h)\| \leq C'\|h\|^2$ (par continuité des appl. linéaires que sont les différentielles en x_0)
4. On calcule le taux d'accroissement en x_0 comme dans le cas des fonctions d'une variable $\dfrac{f(x_0+h)}{g(x_0+h)} - \dfrac{f(x_0)}{g(x_0)} = \dfrac{f(x_0+h)g(x_0) - g(x_0+h)f(x_0)}{g(x_0+h)g(x_0)}$ et l'on injecte les DL de f et g en x_0. Il reste
$$\frac{g(x).\mathrm{d}f(x_0)(h) - f(x_0).\mathrm{d}g(x_0)(h)}{g(x_0+h)g(x_0)} + o(h)$$

□

Deux cas particuliers intéressants :
• Si $f_1, f_2 : U \to F$ où F est euclidien, on peut considérer $g(x) = \Big(f_1(x) \,|\, f_2(x)\Big)$ – ce qui revient à appliquer le point 3 avec $B(y_1, y_2) = \Big(y_1 \,|\, y_2\Big)$. On obtient alors $\mathrm{d}g(x_0)(h) = \Big(\mathrm{d}f_1(x_0)(h) \,|\, f_2(x_0)\Big) + \Big(f_1(x_0) \,|\, \mathrm{d}f_2(x_0)(h)\Big)$.
• Encore plus particulièrement, si $F = \mathbf{R}$, on peut prendre $g(x) = f_1(x).f_2(x)$, et il vient alors $\mathrm{d}g(x_0)(h) = f_2(x_0).\mathrm{d}f_1(x_0)(h) + f_1(x_0).\mathrm{d}f_2(x_0)(h)$.

Proposition 7 (Différentiabilité et applications coordonnées)**.**

1. *Soient $f_i : U \to F_i$ des application $(i = 1, \ldots, p)$.*
Alors en notant $f : U \to F = F_1 \times \cdots \times F_p$ l'application donnée par $f(x) = \Big(f_1(x), \ldots, f_p(x)\Big)$ on a l'équivalence
   - *(i) chaque f_i $(i = 1, \ldots, p)$ différentiable en x_0 (resp. diff. sur U, C^1 sur U)*
   - *(ii) f est différentiable en x_0 (resp. diff. sur U, C^1 sur U)*

   Si tel est le cas on la relation $\boxed{\mathrm{d}f(x_0)(h) = \Big(\mathrm{d}f_1(x_0)(h), \ldots, \mathrm{d}f_p(x_0)(h)\Big)}$
2. *Soient $f : U \to F$ une application et $\mathcal{B}_F = (\varepsilon_1, \ldots, \varepsilon_p)$ une base de F. Alors en notant f_i les applications coordonnées de f, i.e. de sorte que $f(x) = f_1(x)\varepsilon_1 + \cdots + f_p(x)\varepsilon_p$ pour tout x, on a l'équivalence*
   - *(i) chaque f_i $(i = 1, \ldots, p)$ différentiable en x_0 (resp. diff. sur U, C^1 sur U)*
   - *(ii) f est différentiable en x_0 (resp. diff. sur U, C^1 sur U)*

Si tel est le cas on la relation $\boxed{\mathrm{d}f(x_0)(h) = \sum_{i=1}^{p} \mathrm{d}f_1(x_0)(h)\varepsilon_1 + \cdots + \mathrm{d}f_p(x_0)(h)\varepsilon_p}$

Démonstration. On traite seulement le premier point, le 2nd étant en tout point similaire.
Dans le sens direct, on écrit $f(x_0+h) = \Big(f_1(x_0+h), \dots, f_p(x_0+h)\Big) = \Big(f_1(x_0), \dots, f_p(x_0)\Big) + \Big(\mathrm{d}f_1(x_0)(h), \dots, \mathrm{d}f_p(x_0)(h)\Big) + o(h)$ et le tour est joué.
Pour le sens réciproque, on écrit $f_i = \pi_i \circ f$ où π_i est la projection associant à x sa $i$ème coordonnée ; on s'en sort alors par composition, ou directement en notant que si L est linéaire et f différentiable en x_0 alors $L \circ f$ est différentiable en x_0 avec $\mathrm{d}(L \circ f)(x_0) = L \circ \mathrm{d}f(x_0)$. En effet, on a
$L(f(x_0+h)) = L(f(x_0) + \mathrm{d}f(x_0)(h) + \|h\|\varepsilon(h)) = L(f(x_0)) + L(\mathrm{d}f(x_0)(h)) + \underbrace{\|h\|L(\varepsilon(h))}_{=o(h)}$. □

2.2. Composition

Proposition 8. *Soient*

- *E, F, G des $\mathbf{R}$-ev de dimension finie,*
- *$U \subset E$, $V \subset G$ des ouverts,*
- *$f : U \to F$ et $g : V \to G$ des applications telles que $f(U) \subset V$.*
- *$x_0 \in U$ et $y_0 = f(x_0)$*

Supposons

hyp1 *f différentiable en x_0 (resp. diff. sur U, C^1 sur U).*

hyp2 *g différentiable en y_0(resp. diff. sur V, C^1 sur V).*

Alors *$C\ell$* *$g \circ f$ est différentiable en x_0 (resp. diff. sur U, C^1 sur U) et l'on a*

$$\boxed{\mathrm{d}(g \circ f)(x_0) = \mathrm{d}g(y_0) \circ \mathrm{d}f(x_0) = \mathrm{d}g\Big(f(x_0)\Big) \circ \mathrm{d}f(x_0)}$$

Démonstration. On écrit pour h assez petit
$(g \circ f)(x_0+h) = g\Big(f(x_0) + \underbrace{\mathrm{d}f(x_0)(h) + \|h\|\varepsilon_1(h)}_{=k}\Big) = g(y_0) + \mathrm{d}g(y_0)(k) + \|k\|\varepsilon_2(k)$
En remplaçant k par sa valeur, on en tire
$\mathrm{d}g(y_0)(k) = \mathrm{d}g(y_0)\Big(\mathrm{d}f(x_0)(h) + \|h\|\varepsilon_1(h)\Big) = \mathrm{d}g(y_0)\Big(\mathrm{d}f(x_0)(h)\Big) + \|h\|\mathrm{d}g(y_0)\Big(\varepsilon_1(h)\Big)$: le 2nd terme est donc un $o(h)$
de la même manière, on écrit $\|k\|\varepsilon_2(k) = \|\mathrm{d}f(x_0)(h) + \|h\|\varepsilon_1(h)\|\varepsilon_3(h)$ où $\varepsilon_3(h) \to 0$ par composition, et par IT on en tire
$\|\mathrm{d}f(x_0)(h) + \|h\|\varepsilon_1(h)\| \le \|\mathrm{d}f(x_0)(h)\| + \|h\|.\|\varepsilon_1(h)\| \le C\|h\| + \|h\|.\|\varepsilon_1(h)\|$ par caract. des appl. lin. continues, ce qui donne bien $\|k\|\varepsilon_2(k) = o(h)$. □

4.2. Notion de tangence

Définition 11 : arc et tangence. Soit E un $\mathbf{R}-$ev de dim finie.

- On appelle *arc paramétré* toute application $\gamma : [0,1] \to E$ de classe C^1.
- L'image de γ, ie $\gamma([0,1])$, s'appelle le *support* de l'arc.
- On appelle *point régulier* de l'arc tout couple $(t_0, \gamma(t_0))$ tel que $\gamma'(t_0) \neq 0$.
- La *tangente* en un point régulier $(t_0, \gamma(t_0))$ est la droite passant par $\gamma(t_0)$ et dirigée par $\gamma'(t_0)$.

On peut décrire paramétriquement la tangente en $\gamma(t_0)$ comme l'ensemble des points $x \in E$ tels que $x - \gamma(t_0) = (t - t_0)\gamma'(t_0)$ pour un certain $t \in \mathbf{R}$.

Définition 12 : Vecteur tangent à une partie. Soit E un $\mathbf{R}-$ev de dim finie, $X \subset E$ une partie non vide et $x_0 \in X$. Un vecteur $u \in E$ est dit tangent à X en x_0 s'il existe un arc paramétré γ dont le support est inclus dans X tel que $\gamma(0) = x_0$ et $\gamma'(0) = u$.

L'espace vectoriel tangent à X en x_0 est l'ensemble des vecteurs u tangents à X en x_0. On le note $\overrightarrow{\mathcal{T}}_{x_0}(X)$.
L'espace affine tangent à X en x_0 est le sous-ensemble $\mathcal{T}_{x_0}(X)$ de E parallèle à $\overrightarrow{\mathcal{T}}_{x_0}(X)$, translaté par x_0 : on a donc $\mathcal{T}_{x_0}(X) = x_0 + \overrightarrow{\mathcal{T}}_{x_0}(X)$.
En pratique, si X est suffisamment régulière, $\overrightarrow{\mathcal{T}}_{x_0}(X)$ est un ss-ev de E et $\mathcal{T}_{x_0}(X)$ est un sous-espace affine dirigé par $\overrightarrow{\mathcal{T}}_{x_0}(X)$. Mais cette régularité est difficile à définir (et hors programme de toute façon) : c'est la notion de sous-variété.

Exemple 1 : graphe d'une fonction. Soit $f : \mathbf{R}^2 \to \mathbf{R}$ supposée C^1, et soit X le graphe de f, ie $X = \{(x, y, f(x, y)) \,:\, (x, y) \in \mathbf{R}^2\} \subset \mathbf{R}^3$. Alors l'espace tangent à X en un point $(x_0, y_0, f(x_0, y_0))$ du graphe $\overrightarrow{\mathcal{T}}_{x_0}(X)$ est le plan d'équation $z = \partial_1 f(x_0, y_0)x + \partial_2 f(x_0, y_0)y$ (donc l'espace affine tangent est le plan affine d'équation $z - z_0 = (x - x_0)\partial_1 f(x_0, y_0) + (y - y_0)\partial_2 f(x_0, y_0)$).

On procède par double inclusion.
$\subset$ Soit γ un arc inclus dans le graphe : on peut donc écrire $\gamma(t) = \Big(x(t), y(t), f\big(x(t), y(t)\big)\Big)$ et donc par composition $\gamma'(t) = \Big(x'(t), y'(t), x'(t)\partial_1 f\big(x(t), y(t)\big) + y'(t)\partial_2 f\big(x(t), y(t)\big)\Big)$. En faisant $t = 0$, il vient qu'en notant $\gamma'(0) = u = (u_x, u_y, u_z)$, on a la relation $\boxed{u_z = u_x\partial_1 f(x_0, y_0) + u_y\partial_2 f(x_0, y_0)}$.
$\supset$ Réciproquement, si $u = (u_x, u_y, u_z)$ vérifie $u_z = u_x\partial_1 f(x_0, y_0) + u_y\partial_2 f(x_0, y_0)$, posons $x(t) = x_0 + tu_x$ et $y(t) = y_0 + tu_y$ et $\gamma(t) = \Big(x(t), y(t), f\big(x(t), y(t)\big)\Big)$. Par dérivation d'un composée, on a $\gamma'(t) = \Big(x'(t), y'(t), x'(t)\partial_1 f\big(x(t), y(t)\big) + y'(t)\partial_2 f\big(x(t), y(t)\big)\Big) = \Big(u_x, u_y, u_x\partial_1 f\big(x(t), y(t)\big) + u_y\partial_2 f\big(x(t), y(t)\big)\Big)$ et donc avec $t = 0$ il vient $\boxed{\gamma'(0) = \Big(u_x, u_y, u_x\partial_1 f\big(x_0, y_0\big) + u_y\partial_2 f\big(x_0, y_0\big)\Big) = (u_x, u_y, u_z)}$ comme voulu.

Exemple 2 : gradient et surface de niveau. Soit $f : \mathbf{R}^3 \to \mathbf{R}$ une fonction C^1, et X_c la surface de niveau $f^{-1}(\{c\})$, ie l'ensemble des $x \in \mathbf{R}^3$ tels que $f(x) = c$. Alors l'espace vectoriel tangent est contenu dans l'orthogonal du gradient de f en x_0.
En effet soit γ un arc de support inclus dans X_c, on a $f(\gamma(t)) = c$ pour tout t, donc en dérivant $\underbrace{\Big(\nabla f(\gamma(t)) \,|\, \gamma'(t)\Big)}_{= \frac{\mathrm{d}}{\mathrm{d}t}\left(f(\gamma(t))\right)} = 0$ et en faisant $t = 0$ il vient $\Big(\nabla f(x_0) \,|\, u\Big) = 0$ et donc u est orthogonal au gradient.
On peut montrer (à l'aide du théorème d'inversion locale, hors programme aujourd'hui) que l'inclusion inverse est vraie si $\nabla f(x_0) \neq 0$, et on retient que le gradient est orthogonal aux (hyper)surfaces de niveau (dirigé vers les grandes valeurs de f).

Ceci motive l'algorithme du gradient (à pas constant, puis optimal), dont le principe est de donner une procédure numérique permettant d'approcher un point en lequel une fonction atteint un minimum ou un maximum, en suivant à partir d'un point initial, la direction opposée à celle donnée par le gradient.