CHAPITRE III. — DIFFERENTIELLES D'ORDRE SUPERIEUR

Comme d'habitude, tous nos espaces vectoriels sont de dimension finie. On note $(\mathrm{E}, ||.||_{\mathrm{E}})$ et $(\mathrm{F}, ||.||_{\mathrm{F}})$ des espaces vectoriels normés sur $\mathbb{R}$, U un ouvert non vide de E et $f\ : \mathrm{U} \to \mathrm{F}$ une application. Par un choix de bases pour E et F, on se ramène à $f : \mathbb{R}^p \to \mathbb{R}^q$.

5 Différentielles d'ordre 2

Définition 5.1. — *On dit que $f : \mathrm{U} \to \mathrm{F}$ est deux fois différentiable au point $a \in \mathrm{U}$ si f est différentiable sur un voisinage V de a et si l'application $\mathrm{D}f : \mathrm{V} \to \mathcal{L}(\mathrm{E}, \mathrm{F})$ définie par $x \mapsto \mathrm{D}_x f$ est différentiable au point a. On la note $\mathrm{D}^2_a f \in \mathcal{L}(\mathrm{E}, \mathcal{L}(\mathrm{E}, \mathrm{F}))$. L'application est deux fois différentiable sur U si f est deux fois différentiable en tout point de U. Elle est de classe $\mathcal{C}^2$ si f est deux fois différentiable et si la différentielle seconde $a \mapsto \mathrm{D}^2_a f$ est continue.*

Si f est deux fois différentiable, alors $\mathrm{D}^2_a f : \mathrm{E} \to \mathcal{L}(\mathrm{E}, \mathrm{F})$ associe linéairement à un vecteur $k \in \mathrm{E}$ une application linéaire $\mathrm{L}_k : \mathrm{E} \to \mathrm{F}$. On a ainsi

$$\mathrm{D}_{a+k} f = \mathrm{D}_a f + \mathrm{L}_k + \|k\|\varepsilon(k)$$

pour $a + k \in \mathrm{U}$ et où $\lim_{k \to 0} \varepsilon(k) = 0$. Autrement dit, on écrit $\mathrm{D}^2_a f(k)(h) = \mathrm{L}_k(h)$. Le lemme 5.2 ci-dessous nous permet d'identifier $\mathcal{L}(\mathrm{E}, \mathcal{L}(\mathrm{E}, \mathrm{F}))$ avec l'espace vectoriel $\mathcal{L}^2(\mathrm{E}, \mathrm{F})$ des applications bilinéaires $b : \mathrm{E} \times \mathrm{E} \to \mathrm{F}$, de sorte que l'on écrira $\mathrm{D}^2_a f(k, h) = \mathrm{D}^2_a f(k)(h)$.

Lemme 5.2. — *On a un isomorphisme canonique $\Phi : \mathcal{L}(\mathrm{E}, \mathcal{L}(\mathrm{E}, \mathrm{F})) \to \mathcal{L}^2(\mathrm{E}, \mathrm{F})$ préservant les normes défini par $\Phi(f)(x, y) = f_x(y)$ où $(f : x \mapsto f_x) \in \mathcal{L}(\mathrm{E}, \mathcal{L}(\mathrm{E}, \mathrm{F}))$.*

DÉMONSTRATION. — Tout d'abord on vérifie que si $f \in \mathcal{L}(\mathrm{E}, \mathcal{L}(\mathrm{E}, \mathrm{F}))$, alors $\Phi(f)$ est bien bilinéaire: prenons $x, x', y, y' \in \mathrm{E}$ et $\lambda, \mu \in \mathbb{R}$. On écrit

$$\begin{aligned}
\Phi(f)(x + \lambda x', y + \mu y') &= f_{x+\lambda x'}(y + \mu y') \\
&= f_x(y + \mu y') + \lambda f_{x'}(y + \mu y') \quad \text{par linéarité de } x \mapsto f_x \\
&= f_x(y) + \mu f_x(y') + \lambda f_{x'}(y) + \lambda\mu f_{x'}(y') \quad \text{par linéarité de } f_x \text{ et } f_{x'} \\
&= \Phi(f)(x, y) + \lambda\Phi(f)(x', y) + \mu\Phi(x, y') + \lambda\mu\Phi(f)(x', y')\,.
\end{aligned}$$

En prenant $y' = 0$, on montre la linéarité par rapport à la première variable et en prenant $x' = 0$, la linéarité par rapport à la seconde.

Enfin, on montre facilement que $\Phi(f + \lambda g) = \Phi(f) + \lambda\Phi(g)$. Si $\Phi(f)$ est l'application bilinéaire nulle, alors, pour chaque $x \in \mathrm{E}$, l'application $x \mapsto f_x$ est nulle, donc f est nulle. Ceci montre l'injectivité de Φ. Pour montrer la surjectivité, on considère une application bilinéaire $b : \mathrm{E} \times \mathrm{E} \to \mathrm{F}$ et on pose $f : x \mapsto (f_x : y \mapsto f(x, y))$ et on vérifie que $\Phi(f) = b$.

Enfin, si $f \in \mathcal{L}(\mathrm{E}, \mathcal{L}(\mathrm{E}, \mathrm{F}))$, alors

$$\|f\| = \sup_{\|x\|=1} \|f_x\| = \sup_{\|x\|=1} \left(\sup_{\|y\|=1} \|f_x(y)\| \right) = \sup_{\|x\|=\|y\|=1} \|\Phi(f)(x, y)\|\,.$$

■

Lemme 5.3. — *Supposons f deux fois différentiable au point $a \in \mathrm{U}$ et fixons-nous $h \in \mathrm{E}$. Alors l'application $g : x \mapsto \mathrm{D}_x f(h)$ est différentiable au point a et $\mathrm{D}_a g(k) = \mathrm{D}^2_a f(k, h)$. Autrement dit, $\mathrm{D}_a(\mathrm{D}f(h))(k) = \mathrm{D}^2_a f(k, h)$.*

DÉMONSTRATION. — En effet, l'application d'évaluation $\mathrm{ev} : \mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathrm{F}) \mapsto \mathrm{L}(h) \in \mathrm{F}$ est linéaire (donc différentiable) et $g = \mathrm{ev} \circ \mathrm{D}f$ puisque $g(x) = \mathrm{ev} \circ \mathrm{D}_x f = \mathrm{D}_x f(h)$. Donc

$$\mathrm{D}_a g(k) = (\mathrm{ev} \circ \mathrm{D}_a(\mathrm{D}f))(k) = \mathrm{D}^2_a f(k)(h) = \mathrm{D}^2_a f(k, h)\,.$$

■

Remarque 5.4. — *En prenant* $\mathrm{E} = \mathbb{R}^p$, *et en écrivant*

$$\partial^2_{i,j} f(a) = \frac{\partial^2 f}{\partial x_i \partial x_j}(a) \overset{def.}{=} \mathrm{D}^2_a f(\mathrm{e}_i, \mathrm{e}_j) = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(a)$$

on obtient, à partir de $\mathrm{D}_a f(h) = \sum \partial_j f(a) h_j$,

$$\mathrm{D}^2_a f(k,h) = \sum_{i,j=1}^{p} \partial^2_{i,j} f(a) k_i h_j \,.$$

5.1 Théorème de Schwarz

Théorème 5.5 (Schwarz). — *Si* f *est deux fois différentiable au point* a*, alors* $\mathrm{D}^2_a f$ *est une application bilinéaire symétrique: pour tous* $h, k \in \mathrm{E}$*, on a* $\mathrm{D}^2_a f(h,k) = \mathrm{D}^2_a f(k,h)$.

Démonstration. — il suffit de montrer que $\|\mathrm{D}^2_a f(h,k) - \mathrm{D}^2_a f(k,h)\| = o((\|h\| + \|k\|)^2)$. En effet, si $u, v \in \mathrm{E}$ sont quelconques (mais non nuls) alors

$$\|\mathrm{D}^2_a f(u,v) - \mathrm{D}^2_a f(v,u)\| = \frac{1}{t^2}\|\mathrm{D}^2_a f(tu,tv) - \mathrm{D}^2_a f(tv,tu)\| = \frac{1}{t^2} o(t^2) = o(1)$$

donc $\mathrm{D}^2_a f(u,v) = \mathrm{D}^2_a f(v,u)$.

On se fixe $k, h \in \mathrm{E}$ assez petits de sorte que la fonction auxiliaire $g : [0,1] \to \mathrm{F}$ définie par

$$g(t) = f(a + th + k) - f(a + th) - t\mathrm{D}^2_a f(k,h)$$

soit effectivement bien définie. Nous souhaitons appliquer le théorème de la moyenne à g. Cette fonction est différentiable comme fonction composée d'applications différentiables et

$$g'(t) = \mathrm{D}_{a+th+k} f(h) - \mathrm{D}_{a+th} f(h) - \mathrm{D}^2_a f(k,h) = (\mathrm{D}_{a+th+k} f - \mathrm{D}_a f)(h) - (\mathrm{D}_{a+th} f - \mathrm{D}_a f)(h) - \mathrm{D}^2_a f(k,h)\,.$$

En remarquant que $\mathrm{D}^2_a f(k,h) = \mathrm{D}^2_a f(k + th, h) - \mathrm{D}^2_a f(th, h)$, on obtient

$$g'(t) = (\mathrm{D}_{a+th+k} f - \mathrm{D}_a f - \mathrm{D}^2_a f(k + th))(h) - (\mathrm{D}_{a+th} f - \mathrm{D}_a f - \mathrm{D}^2_a f(th))(h)\,.$$

Soit $\varepsilon > 0$. La différentiabilité de $\mathrm{D}f$ au point a implique l'existence de $r > 0$ tel que si $u \in \mathrm{B}_\mathrm{E}(0,r)$, alors

$$\|\mathrm{D}_{a+u} f - \mathrm{D}_a f - \mathrm{D}^2_a f(u)\| \leq \varepsilon\|u\|\,.$$

On se fixe $h, k \in \mathrm{B}_\mathrm{E}(0, r/2)$ de sorte que, pour tout $t \in [0,1]$, on ait $\|k + th\| \leq \|k\| + |t|\|h\| < r$. De même, on a $\|th\| = |t|\|h\| < r$. Donc, pour tout $t \in [0,1]$, on obtient

$$\begin{aligned}
\|g'(t)\| &\leq \|(\mathrm{D}_{a+th+k} f - \mathrm{D}_a f - \mathrm{D}^2_a f(k + th))(h)\| + \|(\mathrm{D}_{a+th} f - \mathrm{D}_a f - \mathrm{D}^2_a f(th))(h)\| \\
&\leq \|\mathrm{D}_{a+th+k} f - \mathrm{D}_a f - \mathrm{D}^2_a f(k + th)\|\|h\| + \|\mathrm{D}_{a+th} f - \mathrm{D}_a f - \mathrm{D}^2_a f(th)\|\|h\| \\
&\leq \varepsilon(\|th + k\|\|h\| + \|th\|\|h\|) \leq \varepsilon(\|h\| + \|k\| + \|h\|)\|h\| \\
&\leq 2\varepsilon(\|h\| + \|k\|)^2\,.
\end{aligned}$$

On applique maintenant le théorème de la moyenne à g afin d'obtenir

$$\|g(1) - g(0)\| \leq \sup_{0 \leq t \leq 1} \|g'(t))\| \leq 2\varepsilon(\|h\| + \|k\|)^2\,.$$

On obtient donc

$$\|f(a + h + k) - f(a + h) - f(a + k) + f(a) - \mathrm{D}^2_a f(k,h)\| \leq 2\varepsilon(\|h\| + \|k\|)^2\,.$$

Par symétrie du rôle de h et k, on en déduit

$$\|\mathrm{D}^2_a f(h,k) - \mathrm{D}^2_a f(k,h)\| \leq 4\varepsilon(\|h\| + \|k\|)^2\,.$$

■

Corollaire 5.6. — *Soit* $\mathrm{U} \subset \mathbb{R}^p$ *un ouvert. Si* $f : \mathrm{U} \to \mathrm{F}$ *est deux fois différentiable au point* $a \in \mathrm{U}$*, alors, pour tous* $i, j \in \{1, \dots, p\}$*, on a*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)\,.$$

Démonstration. — D'après le théorème de Schwarz, on a

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \mathrm{D}^2_a f(\mathrm{e}_i, \mathrm{e}_j) = \mathrm{D}^2_a f(\mathrm{e}_j, \mathrm{e}_i) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)\,.$$

■

Corollaire 5.7. — *Sous les mêmes conditions, si toutes les dérivées secondes* $(\partial^2_{i,j} f)$*,* $i, j \in \{1, \dots, p\}$*, existent et sont continues au voisinage au point* a *alors*

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)\,.$$

Démonstration. — Il découle des hypothèses que les dérivées partielles existent et sont continues, donc le théorème 4.6 implique que $f$est de classe $\mathcal{C}^1$. De plus, les dérivées partielles de $\mathrm{D}f$ correspondent aux dérivées partielles secondes de f, donc le même théorème implique que f est de classe $\mathcal{C}^2$ au voisinage de a. Du coup, le corollaire 5.6 permet de conclure. ■

Exemple 5.8. — *Si* f *n'est pas deux fois différentiable, alors l'ordre des dérivées partielles est important. Prenons par exemple* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par* $f(0, 0) = 0$ *et si* $(x, y) \neq (0, 0)$*,*

$$f(x, y) = \frac{xy(x^2 - y^2)}{x^2 + y^2}\,.$$

On peut montrer que les dérivées partielles $\partial_1 f$ *et* $\partial_2 f$ *existent en tout point et* $\partial_1 \partial_2 f(0) = 1$ *alors que* $\partial_2 \partial_1 f(0) = -1$*.*

Plus généralement, si $f(x, y) = xyg(x, y)$ *si* $(x, y) \neq (0, 0)$ *et* $f(0) = 0$*, où* g *admet des dérivées partielles secondes en* 0*, alors les dérivées partielles secondes de* f *en zéro ne commutent pas si*

$$\lim_{x \to 0} \left(\lim_{y \to 0} g(x, y) \right) \neq \lim_{y \to 0} \left(\lim_{x \to 0} g(x, y) \right)\,.$$

5.2 Rappels sur les formes bilinéaires

En préparation pour la suite, on fait quelques rappels sur les formes bilinéaires. Une *forme bilinéaire* sur un espace vectoriel E est une application $b : \mathrm{E} \times \mathrm{E} \to \mathbb{R}$, linéaire par rapport à chaque variable. Si E est de dimension finie et $\mathcal{B} = (\mathrm{e}_1, \dots, \mathrm{e}_n)$ est une base de E, on lui associe la matrice

$$\mathrm{Mat}(b, \mathcal{B}) = (b(\mathrm{e}_i, \mathrm{e}_j))_{1 \le i,j \le n} \in \mathcal{M}_n(\mathbb{R})\,.$$

On peut vérifier que b est entièrement déterminée par $\mathcal{B}$: si $x = \sum_i x_i \mathrm{e}_i$ et $y = \sum_j y_j \mathrm{e}_j$ alors

$$b(x, y) = \sum_{1 \le i,j \le n} x_i y_j b(\mathrm{e}_i, \mathrm{e}_j)\,.$$

Matriciellement, cela s'écrit

$$b(x, y) = \begin{pmatrix} x_1 & \dots & x_n \end{pmatrix} \mathrm{Mat}(b, \mathcal{B}) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = {}^t\!x\, \mathrm{Mat}(b, \mathcal{B}) y\,.$$

Si $\mathcal{B}'$ est une autre base de E et P est la matrice de passage de $\mathcal{B}$ vers $\mathcal{B}'$, alors on a

$$\mathrm{Mat}(b, \mathcal{B}') = {}^t\mathrm{P}\, \mathrm{Mat}(b, \mathcal{B}) \mathrm{P}\,.$$

Enfin, on dit que b est *symétrique* si $b(x, y) = b(y, x)$, ce qui est équivalent à la matrice de b à être symétrique. Si b est symétrique, alors il existe une base dans laquelle b est diagonale.

On dit qu'une forme bilinéaire symétrique b est *non dégénérée* si $\mathrm{Mat}(b, \mathcal{B})$ est inversible. La forme b, symétrique, est *positive* si, pour tout $x \in \mathrm{E}$, on a $b(x, x) \geq 0$; elle est *définie positive* si, pour tout $x \neq 0$, on a $b(x, x) > 0$. On a des définitions analogues pour b négative et définie négative.

On remarque que b est positive si les valeurs propres de $\mathrm{Mat}(b, \mathcal{B})$ sont positives, et b est définie positive si les valeurs propres sont strictement positives.

5.3 Hessienne

On suppose dans ce paragraphe que U est un ouvert de $\mathbb{R}^n$ et $f : \mathrm{U} \to \mathbb{R}$ est une application à valeurs réelles.

Définition 5.9. — *Si $f : \mathrm{U} \to \mathbb{R}$ est deux fois différentiable au point $a \in \mathrm{U}$, alors on définit la hessienne $\mathrm{H}_f(a)$ au point a par la matrice de terme général $(\partial^2_{i,j} f(a))_{i,j}$.*

La hessienne est donc la matrice de $\mathrm{D}^2_a f$ dans la base canonique. Il découle du théorème de Schwarz (et de ses corollaires) que la hessienne est une forme bilinéaire symétrique et on a

$$\mathrm{D}^2_a f(h, k) = {}^t\!h \mathrm{H}_f(a) k$$

pour $h, k \in \mathbb{R}^n$.

Proposition 5.10. — *Si $f : \mathrm{U} \to \mathbb{R}$ est deux fois différentiable au point a, alors il existe $\varepsilon : \mathrm{U} \to \mathbb{R}$ telle que*

$$f(x) = f(a) + \mathrm{D}_a f(x - a) + \frac{1}{2}\, {}^t(x - a) \mathrm{H}_f(a)(x - a) + \|x - a\|^2 \varepsilon(x)$$

où $\lim_{x \to a} \varepsilon(x) = 0$.

Démonstration. — On considère la fonction auxiliaire $g : \mathrm{U} \to \mathbb{R}$ définie par

$$g(x) = f(x) - f(a) - \mathrm{D}_a f(x - a) - \frac{1}{2}\, {}^t(x - a) \mathrm{H}_f(a)(x - a)$$

que l'on différentie pour x assez proche de a en se souvenant que le dernier terme est bilinéaire en x, cf. § 3.1:

$$\mathrm{D}_x g(h) = \mathrm{D}_x f(h) - \mathrm{D}_a f(h) - {}^t(x - a) \mathrm{H}_f(a) h = \mathrm{D}_x f(h) - \mathrm{D}_a f(h) - \mathrm{D}^2_a f(x - a, h)\,.$$

Soit $\varepsilon > 0$; comme $\mathrm{D}f$ est différentiable au point a, il existe $\delta > 0$ telle que, si $\|x - a\| < \delta$, alors $\|\mathrm{D}_x g\| \leq \varepsilon \|x - a\|$. Du coup, le théorème de la moyenne implique, pour $x \in \mathrm{B}(a, \delta)$,

$$\begin{aligned} \|f(x) - f(a) - \mathrm{D}_a f(x - a) - \frac{1}{2}\, {}^t(x - a) \mathrm{H}_f(a)(x - a)\| &= \|g(x) - g(a)\| \\ &\leq \sup_{y \in [x,a]} \|\mathrm{D}_y g\| \cdot \|x - a\| \\ &\leq \varepsilon \|x - a\|^2\,. \end{aligned}$$

Ceci conclut la démonstration. ∎

Corollaire 5.11 (extrema locaux). — *Soit $f : \mathrm{U} \to \mathbb{R}$ une application deux fois différentiables. Si a est un maximum local (resp. un minimum local) alors $\mathrm{D}_a f = 0$ et $\mathrm{H}_f(a)$ est négative (resp. positive). Réciproquement, si $\mathrm{D}_a f = 0$ et $\mathrm{H}_f(a)$ est définie positive (resp. définie négative) alors a est un minimum local strict (resp. maximum local strict).*

Remarque 5.12. — $\mathrm{H}_f(a)$ *est (définie) positive si ses valeurs propres sont (strictement) positives, et $\mathrm{H}_f(a)$ est (définie) négative si ses valeurs propres sont (strictement) négatives. Si les valeurs propres sont de signe différent, alors a n'est pas un extremum.*

Démonstration. — Supposons que a est un maximum local. La proposition 3.12 nous affirme que $\mathrm{D}_a f = 0$. Du coup, sur un voisinage $\mathrm{B}(a,r)$ suffisamment petit de a, la proposition précédente nous affirme, que pour $h \in \mathrm{E}$ et $t \in [0, r/\|h\|]$,

$$f(a+th) = f(a) + \frac{t^2}{2}\,{}^t h \mathrm{H}_f(a) h + t^2\|h\|^2\varepsilon(th) \leq f(a)$$

d'où

$$\frac{1}{2}\,{}^t h \mathrm{H}_f(a) h \leq -\|h\|^2\varepsilon(th)\,.$$

En prenant la limite quand t tend vers 0, on obtient ${}^t h\mathrm{H}_f(a)h \leq 0$. Comme h est arbitraire, on en déduit que $\mathrm{H}_f(a)$ est négative.

Si a est un minimum local, le même argument montre ${}^t h\mathrm{H}_f(a)h \geq 0$ pour tout $h \in \mathrm{E}$, impliquant que $\mathrm{H}_f(a)$ est positive.

Supposons maintenant $\mathrm{D}_a f = 0$ et $\mathrm{H}_f(a)$ définie positive. Alors, pour h assez petit, on a

$$f(a+h) = f(a) + \frac{1}{2}\,{}^t h\mathrm{H}_f(a)h + \|h\|^2\varepsilon(h)\,.$$

D'après le lemme 5.13 ci-dessous, il existe $c > 0$ tel que ${}^t h\mathrm{H}_f(a)h \geq c\|h\|^2$. Prenons maintenant h assez proche de 0 pour que $\|\varepsilon(h)\| \leq (c/4)$. Il vient

$$f(a+h) \geq f(a) + \frac{c}{4}\|h\|^2 > f(a)\,.$$

Donc a est un minimum local strict.

Si $\mathrm{H}_f(a)$ est définie négative, alors $-\mathrm{H}_f(a)$ est définie positive, donc il existe $c > 0$ telle que ${}^t h\mathrm{H}_f(a)h \leq -c\|h\|^2$. Du coup, l'argument précédent montre

$$f(a+h) \leq f(a) - \frac{c}{4}\|h\|^2 < f(a)\,.$$

Donc a est un maximum local strict. ∎

Lemme 5.13. — *Soit $b : \mathrm{E}\times\mathrm{E} \to \mathbb{R}$ une forme bilinéaire définie positive définie sur un evn de dimension finie. Il existe $c > 0$ telle que, pour tout $x \in \mathrm{E}$, on ait $b(x,x) \geq c\|x\|^2$.*

Démonstration. — Soit $\mathrm{S} = \{x \in \mathrm{E}, \|x\| = 1\}$. L'application $x \mapsto b(x,x)$ est continue d'après la proposition 2.14, donc atteint son minimum sur S qui est compact en un vecteur $x_0 \neq 0$. Du coup, on a, pour tout $x \in \mathrm{E}$, $b(x,x) \geq b(x_0,x_0)\|x\|^2$, avec $b(x_0,x_0) > 0$. ∎

Corollaire 5.14 (Lagrange). — *Soit U un ouvert de $\mathbb{R}^2$ et soit $f : \mathrm{U} \to \mathbb{R}$ une application deux fois différentiable en $a \in \mathrm{U}$ tel que $\mathrm{D}_a f = 0$. On pose*

$$r = \frac{\partial^2 f}{\partial x^2}(a), \quad s = \frac{\partial^2 f}{\partial x \partial y}(a), \quad t = \frac{\partial^2 f}{\partial y^2}(a)\,.$$

- *Si $rt - s^2 > 0$ alors f admet au point a un extremum local strict. Si $r > 0$, alors a est un minimum et si $r < 0$, alors a est un maximum.*
- *Si $rt - s^2 < 0$, alors f n'admet pas d'extremum local au point a.*

Démonstration. — On vérifie que $\mathrm{H}_f(a)$ est définie positive ou négative si $rt - s^2 > 0$, avec $\mathrm{H}_f(a)$ positive si $r > 0$ et négative si $r < 0$. En effet, le polynôme caractéristique de $\mathrm{H}_f(a)$ est

$$\chi(\lambda) = \lambda^2 - (r+t)\lambda + (rt - s^2)\,.$$

Du coup, le discriminant vaut $(r+t)^2 - 4(rt-s^2) = (r-t)^2 + 4s^2 \geq 0$ et on a bien deux valeurs propres réelles, dont le produit est $rt - s^2$. Si le produit est strictement positif, alors les deux valeurs propres sont du même signe (et non nul), donc $\mathrm{H}_f(a)$ est définie positive ou négative. Si le produit est strictement négatif, alors les valeurs propres sont de signe opposé, donc $\mathrm{H}_f(a)$ n'est ni positive ni négative. Du coup a ne peut être un extremum local d'après la condition nécesaire du corollaire 5.11.

Si $rt - s^2 > 0$ alors r et t sont de même signe, donc le signe de la somme des valeurs propres est celui de r. ∎

6 Différentielles d'ordre quelconque

On adapte le paragraphe précédent aux différentielles d'ordre supérieur. On commence par identifier leur nature.

Lemme 6.1. — *Soient* E, F *deux espaces vectoriels normés. On note* $\mathcal{L}_1(\mathrm{E}, \mathrm{F}) = \mathcal{L}(\mathrm{E}, \mathrm{F})$*, puis par récurrence* $\mathcal{L}_{n+1}(\mathrm{E}, \mathrm{F}) = \mathcal{L}(\mathrm{E}, \mathcal{L}_n(\mathrm{E}, \mathrm{F}))$*. Pour tout* $n \geq 1$*, il existe un isomorphisme canonique entre* $\mathcal{L}_n(\mathrm{E}, \mathrm{F})$ *et l'espace* $\mathcal{L}^n(\mathrm{E}, \mathrm{F})$ *des applications* n*-linéaires* $f : \mathrm{E}^n \to \mathrm{F}$ *préservant les normes.*

On rappelle que $\mathcal{L}^n(\mathrm{E}, \mathrm{F})$ est muni de la norme

$$\|\mathrm{L}\| = \sup_{\|x_1\|=\|x_2\|=\ldots=\|x_n\|=1} \|\mathrm{L}(x_1, \ldots, x_n)\| .$$

Démonstration. — On procède par récurrence. Les cas $n = 1, 2$ ont déjà été traités. Supposons que la conclusion du lemme soit vraie en rang n, et montrons-la au rang $n + 1$. On définit $\Phi : \mathcal{L}_{n+1}(\mathrm{E}, \mathrm{F}) \to \mathcal{L}^{n+1}(\mathrm{E}, \mathrm{F})$ en posant

$$\Phi(\mathrm{L})(x_1, x_2, \ldots, x_{n+1}) = \mathrm{L}(x_1)(x_2, \ldots, x_{n+1}) .$$

La linéarité par rapport à la première variable découle du fait que $\mathrm{L} \in \mathcal{L}(\mathrm{E}, \mathcal{L}_n(\mathrm{E}, \mathrm{F}))$, et la linéarité par rapport aux autres variables vient de l'hypothèse de récurrence.

Le reste de la démonstration suit celle du lemme 5.2. ∎

Définition 6.2. — *Soit* $p \geq 2$*. On dit que* $f : \mathrm{U} \to \mathrm{F}$ *est* p *fois différentiable au point* $a \in \mathrm{U}$ *si* f *est* $(p-1)$ *différentiable sur un voisinage* V *de* a *et si la différentielle de rang* $(p-1)$ $\mathrm{D}^{p-1} f : \mathrm{V} \to \mathcal{L}_{p-1}(\mathrm{E}, \mathrm{F})$ *est différentiable au point* a*. L'application est* p *fois différentiable sur* U *si* f *est* p *fois différentiable en tout point de* U*. Elle est de classe* $\mathcal{C}^p$ *si* f *est* p *fois différentiable et si la différentielle* p*-ième* $a \mapsto \mathrm{D}^p_a f$ *est continue. On dit que* f *est de classe* $\mathcal{C}^\infty$ *sur* U *si* f *est différentiable* p *fois pour tout* $p \geq 1$*.*

Si f est p fois différentiable, on obtient par récurrence

$$\mathrm{D}_a(\mathrm{D}(\ldots(\mathrm{D}(\mathrm{D}f(h_1))(h_2))\ldots)(h_{p-1}))(h_p) = \mathrm{D}^p_a f(h_p, \ldots, h_1) .$$

Remarque 6.3. — *Si* f *est* p*-différentiable au voisinage de* a *et si* $\mathrm{D}^p f$ *est* q *fois différentiable, alors* f *est* $(p + q)$ *fois différentiable et* $\mathrm{D}^{p+q}_a f = \mathrm{D}^q_a(\mathrm{D}^p f)$*.*

6.1 Symétrie des différentielles et dérivées partielles

Une application k-linéaire $f : \mathrm{E}^k \to \mathrm{F}$ est *symétrique* si, pour toute permutation $\sigma \in \mathfrak{S}_k$ et tout $(x_1, \ldots, x_k) \in \mathrm{E}^k$, on a

$$f(x_1, \ldots, x_k) = f(x_{\sigma(1)}, \ldots, x_{\sigma(k)}) .$$

Théorème 6.4 (de Schwarz généralisé). — *Si* f *est* p *fois différentiable au point* a*, alors* $\mathrm{D}^p_a f$ *est une application* p*-linéaire symétrique.*

Démonstration. — On procède par récurrence sur p. Le cas $p = 2$ est connu, donc supposons $p \geq 3$ et les cas $j < p$ connus. Soit $(h_1, \ldots, h_p) \in \mathrm{E}^p$. Si f est p fois différentiable, alors l'application $x \mapsto \mathrm{D}^{p-2}_x f(h_3, \ldots, h_p)$ est deux fois différentiable au point a, donc le théorème de Schwarz implique

$$\mathrm{D}^p_a f(h_1, h_2, \ldots, h_p) = \mathrm{D}^p_a f(h_2, h_1, \ldots, h_p) .$$

Or, notre hypothèse de récurrence implique aussi que pour toute permutation $\sigma \in \mathfrak{S}_{p-1}$, on a

$$\mathrm{D}^{p-1}_x f(h_2, \ldots, h_p) = \mathrm{D}^{p-1}_x f(h_{\sigma(2)}, \ldots, h_{\sigma(p)})$$

pour x assez proche de a, donc en différentiant au point a, on obtient

$$\mathrm{D}^p_a f(h_1, h_2, \ldots, h_p) = \mathrm{D}^p_a f(h_1, h_{\sigma(2)}, \ldots, h_{\sigma(p)}) .$$

Pour conclure, on remarque que $\mathfrak{S}_p$ est engendré par la transposition $(1, 2)$ et les permutations qui fixent 1. On en déduit donc

$$\mathrm{D}^p_a f(h_1, h_2, \ldots, h_p) = \mathrm{D}^p_a f(h_{\sigma(1)}, \ldots, h_{\sigma(p)})$$

pour toute permutation $\sigma \in \mathfrak{S}_p$. ∎

Soient U un ouvert de $\mathbb{R}^n$ et $f : \mathrm{U} \to \mathrm{F}$ une application. Pour $p \geq 1$ et $j_1, \ldots, j_p \in \{1, \ldots, n\}$, on définit la dérivée partielle par le vecteur

$$\partial^p_{j_1,\ldots,j_p} f = \frac{\partial^p f}{\partial x_{j_1} \ldots \partial x_{j_p}} \stackrel{\text{def.}}{=} \partial_{j_1}(\partial_{j_2}(\ldots(\partial_{j_p} f)\ldots))$$

s'il est bien défini. Si f est p fois différentiable, alors on obtient par récurrence

$$\partial^p_{j_1,\ldots,j_p} f(a) = \mathrm{D}^p_a f(\mathrm{e}_{j_1}, \ldots, \mathrm{e}_{j_p}) \,.$$

De plus, si, pour $j \in \{1, \ldots, p\}$, on note

$$h_j = \begin{pmatrix} h_j^1 \\ \vdots \\ h_j^n \end{pmatrix}$$

alors on obtient, sans tenir compte des symétries,

$$\mathrm{D}^p_a f(h_1, \ldots, h_p) = \sum_{j_1,\ldots,j_p=1}^{n} \partial^p_{j_1,\ldots,j_p} f(a) h_1^{j_1} \ldots h_p^{j_p} \,.$$

Cette formule et le théorème de Schwarz généralisé impliquent

Corollaire 6.5. — *Si f est p fois différentiable au point a, alors, pour toute permutation $\sigma \in \mathfrak{S}_p$, on a*

$$\partial^p_{j_1,\ldots,j_p} f(a) = \partial^p_{\sigma(j_1),\ldots,\sigma(j_p)} f \,.$$

6.2 Propriétés et exemples

On énonce quelques propriétés qui découlent sans malice des propriétés déjà établies des applications différentiables en procédant par récurrence.

Proposition 6.6. — *Soient* U *un ouvert d'un evn* E, $a \in \mathrm{U}$, $p \geq 1$ *et* $f : \mathrm{U} \to \mathrm{F}$ *une application à valeurs dans un evn* F.

1. *Si* E *est un produit fini* $\mathrm{E}_1 \times \ldots \times \mathrm{E}_k$ *d'evn, alors f est de classe $\mathcal{C}^k$ si et seulement si les différentielles partielles d'ordre au plus p existent et sont continues le long de chaque sous-espace* E_j.

2. *Si* F *est un produit fini d'espaces* $\mathrm{F}_1 \times \ldots \times \mathrm{F}_k$*, alors $f = (f_1, \ldots, f_k)$ est de classe $\mathcal{C}^p$ si et seulement si f_j est de classe $\mathcal{C}^p$ pour chaque $j \in \{1, \ldots, k\}$.*

3. *La combinaison linéaire d'applications de classe $\mathcal{C}^p$ dans* U *est de classe $\mathcal{C}^p$ dans* U.

4. *Si* $\mathrm{L} : \mathrm{F} \to \mathrm{G}$ *est une application linéaire et f est différentiable p fois au point a, alors $\mathrm{L} \circ f$ est différentiable p fois et $\mathrm{D}^p_a(\mathrm{L} \circ f) = \mathrm{L} \circ \mathrm{D}^p_a f$. De même, si $\mathrm{L} : \mathrm{G} \to \mathrm{E}$ est linéaire, $\mathrm{L}(b) = a$, alors $(f \circ \mathrm{L})$ est p fois différentiable au point b et $\mathrm{D}^p_b(f \circ \mathrm{L})(h_1, \ldots, h_p) = \mathrm{D}^p_a f(\mathrm{L}(h_1), \ldots, \mathrm{L}(h_p)))$.*

Cela nous permet de montrer

Exemple 6.7. — *Une application k-linéaire $f : \mathrm{E}_1 \times \ldots \times \mathrm{E}_k \to \mathrm{F}$ est de classe $\mathcal{C}^\infty$ et les différentielles d'ordre $p \geq k+1$ sont toutes nulles.*

Démonstration. — On procède par récurrence sur k. Pour le cas $k = 1$, on sait déjà que $\mathrm{D}_a f = f$ pour tout $a \in \mathrm{E}_1$, donc $\mathrm{D}_{a+h} f - \mathrm{D}_a f = f - f = 0$. Supposons maintenant le cas $(k-1)$ connu et considérons une transformation k-linéaire. D'après le théorème 4.7, on obtient

$$\mathrm{D}_a f(h_1, \ldots, h_k) = \sum_{j=1}^{k} f(a_1, \ldots, a_{j-1}, \underbrace{h_j}_{j\text{ème place}}, a_{j+1}, \ldots, a_k)$$

soit

$$\mathrm{D}_a f(h_1, \ldots, h_k) = f(h_1, a_2, \ldots, a_k) + f(a_1, h_2, a_3, \ldots, x_k) + \ldots + f(a_1, \ldots, a_{k-1}, h_k) \,.$$

Donc $\mathrm{D}f$ est la somme finie de la composée de la projection de $\mathrm{E}_1 \times \ldots \times \mathrm{E}_k$ sur $\mathrm{E}_1 \times \ldots \times \mathrm{E}_{j-1} \times \mathrm{E}_{j+1} \times \ldots \times \mathrm{E}_k$ et de l'application $(k-1)$-linéaire de $\mathrm{E}_1 \times \ldots \times \mathrm{E}_{j-1} \times \mathrm{E}_{j+1} \times \ldots \times \mathrm{E}_k$ sur $\mathcal{L}(\mathrm{E}_j, \mathrm{F})$ définie par

$$(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_k) \mapsto (h \in \mathrm{E}_j \mapsto f(x_1, , \ldots, x_{j-1}, h, x_{j+1}, \ldots, x_k)) .$$

Donc l'hypothèse de récurrence s'applique et la proposition précédente permet de conclure. ∎

Proposition 6.8. — *Soient* $p \in \mathbb{N} \cup \{\infty\}$, $\mathrm{E}, \mathrm{F}, \mathrm{G}$ *trois evns,* U *un ouvert de* E, V *un ouvert de* F, a *un point de* U, *et* $f : \mathrm{U} \to \mathrm{F}$ *et* $g : \mathrm{V} \to \mathrm{G}$ *deux applications telles que* $f(\mathrm{U}) \subset \mathrm{V}$.

1. *Si* f *est* p *fois différentiable au point* a *et* g *est* p *fois différentiable au point* $f(a)$ *alors* $(g \circ f)$ *est* p *fois différentiable au point* a.

2. *Si* f *et* g *sont de classe* $\mathcal{C}^p$ *alors* $(g \circ f)$ *est de classe* $\mathcal{C}^p$.

Démonstration. — On procède par récurrence: le cas $p = 1$ est connu, et de plus $\mathrm{D}_a(g \circ f) = \mathrm{D}_{f(a)}g \circ \mathrm{D}_a f$. Supposons maintenant la proposition connue au rang $(p-1) \geq 1$ et montrons-la au rang p. Sans perte de généralité, on peut supposer que f est p fois différentiable sur tout U et g sur tout V. On remarque que $x \mapsto \mathrm{D}_x(g \circ f)$ est la composée de $\varphi : \mathrm{U} \to \mathcal{L}(\mathrm{E}, \mathrm{F}) \times \mathcal{L}(\mathrm{F}, \mathrm{G})$ définie par $\varphi(x) = (\mathrm{D}_x f, \mathrm{D}_{f(x)}g)$ et de l'application bilinéaire $\psi : \mathcal{L}(\mathrm{E}, \mathrm{F}) \times \mathcal{L}(\mathrm{F}, \mathrm{G}) \to \mathcal{L}(\mathrm{E}, \mathrm{G})$ définie par $\psi(\mathrm{A}, \mathrm{B}) = \mathrm{B} \circ \mathrm{A}$. D'après l'exemple précédent, ψ est de classe $\mathcal{C}^\infty$, et les applications composantes de φ sont $(p-1)$ fois différentiables, donc φ aussi. On en déduit que $(g \circ f)$ est p fois différentiable. Si les différentielles d'ordre p sont continues, alors on en déduit que $(g \circ f)$ est aussi de classe $\mathcal{C}^p$. ∎

On obtient par récurrence

Proposition 6.9. — *Soient* U *un ouvert d'un evn et* $f, g : \mathrm{U} \to \mathbb{R}$ *des applications* p *fois différentiables (resp. de classe* $\mathcal{C}^p$*), alors le produit* (fg) *est aussi* p *fois différentiable (resp. de classe* $\mathcal{C}^p$*).*

Exemple 6.10. — *Toute application polynomiale ou rationnelle* $f : \mathbb{R}^n \to \mathbb{R}$ *est de classe* $\mathcal{C}^\infty$ *là où elle est définie.*

Exemple 6.11. — *L'application* $f : \mathrm{GL}(\mathrm{E}, \mathrm{F}) \to \mathrm{GL}(\mathrm{F}, \mathrm{E})$ *définie par* $f(u) = u^{-1}$ *est de classe* $\mathcal{C}^\infty$.

Démonstration. — Soit $\Phi : \mathrm{E} \to \mathrm{F}$ un isomorphisme. On considère $g : \mathrm{GL}(\mathrm{E}) \to \mathrm{GL}(\mathrm{E})$ définie par $g(u) = u^{-1}$ et les applications linéaires $\mathrm{A}_\Phi : \mathcal{L}(\mathrm{E}, \mathrm{F}) \to \mathcal{L}(\mathrm{E})$ et $\mathrm{B}_\Phi : \mathcal{L}(\mathrm{E}) \to \mathcal{L}(\mathrm{F}, \mathrm{E})$ définies par $\mathrm{A}_\Phi(u) = \Phi^{-1} \circ u$ et $\mathrm{B}_\Phi(u) = u \circ \Phi^{-1}$. On a vu $\mathrm{D}_u g(h) = -u^{-1} \circ h \circ u^{-1}$. Or $f = \mathrm{B}_\Phi \circ g \circ \mathrm{A}_\Phi$ de sorte que, par linéarité et composition, on obtient aussi $\mathrm{D}_u f(h) = -u^{-1} \circ h \circ u^{-1}$. On l'écrit sous la forme $\mathrm{D}_u f = \psi \circ \varphi$ où $\varphi : \mathrm{GL}(\mathrm{E}, \mathrm{F}) \to \mathcal{L}(\mathrm{F}, \mathrm{E}) \times \mathcal{L}(\mathrm{F}, \mathrm{E})$ est définie par $\varphi(u) = (u^{-1}, u^{-1})$ et $\psi : \mathcal{L}(\mathrm{F}, \mathrm{E}) \times \mathcal{L}(\mathrm{F}, \mathrm{E}) \to \mathcal{L}(\mathcal{L}(\mathrm{E}, \mathrm{F}), \mathcal{L}(\mathrm{F}, \mathrm{E}))$ est définie par $\psi(v_1, v_2)(h) = -v_1 \circ h \circ v_2$. Par récurrence, l'application φ est de classe $\mathcal{C}^p$ si f est de classe $\mathcal{C}^p$ et ψ est de classe $\mathcal{C}^\infty$ car bilinéaire. Donc $\mathrm{D}f$ est de classe $\mathcal{C}^p$ et f de classe $\mathcal{C}^{p+1}$. ∎

7 Développements limités

On adapte les formules de Taylor aux applications de plusieurs variables. Les hypothèses chaque fois sont différentes, et il convient de bien les retenir. On commence par le cas de polynômes où on obtient une formule exacte.

7.1 Polynômes généralisés

Définition 7.1. — *Soient* E, F *deux espaces vectoriels de dimension finie.*

1. *Soit* $k \geq 0$*; une application* $f : \mathrm{E} \to \mathrm{F}$ *est un polynôme homogène de degré* k *s'il existe une application* k*-linéaire symétrique* $g : \mathrm{E}^k \to \mathrm{F}$ *telle que* $f(x) = g(x, \ldots, x)$. *Si* $k = 0$, f *est constante.*

2. *Soit* $d \geq 0$*; un polynôme de degré* d $\mathrm{P} : \mathrm{E} \to \mathrm{F}$ *est une combinaison linéaire de polynômes homogènes de degré* $k \leq d$ *de sorte que le polynôme homogène de degré* d *est non identiquement nul.*

Théorème 7.2. — *Soit* $\mathrm{P} : \mathrm{E} \to \mathrm{F}$ *un polynôme de degré* $d \geq 0$. *Alors*

$$\mathrm{P}(x) = \sum_{k=0}^{d} \frac{1}{k!} \mathrm{D}_0^k \mathrm{P}(\underbrace{x, \ldots, x}_{k \text{ fois}}).$$

On traite d'abord le cas des polynômes homogènes.

Proposition 7.3. — *Soit* $f : \mathrm{E} \to \mathrm{F}$ *un polynôme homogène de degré* $k \geq 0$, *alors* f *est de classe* $\mathcal{C}^\infty$, $\mathrm{D}^p f = 0$ *si* $p > k$ *et, si* $p \leq k$,

$$\mathrm{D}_x^p f(h, \ldots, h) = \frac{k!}{(k-p)!} g(\underbrace{x, \ldots, x}_{k-p \text{ fois}}, \underbrace{h, \ldots, h}_{p \text{ fois}})$$

où $f(x) = g(x, \ldots, x)$.

Démonstration. — On considère pour chaque $j \geq 1$, l'application linéaire $\Delta_j : \mathrm{E} \to \mathrm{E}^j$ définie par $\Delta_j(x) = (\underbrace{x, \ldots, x}_{j \text{ fois}})$ de sorte que $f = g \circ \Delta_k$ pour un polynôme homogène de degré k.

On procède par récurrence sur k. Prenons $k = 1$, de sorte que $f = g \in \mathcal{L}(\mathrm{E}, \mathrm{F})$. Si $p = 0$ alors $\mathrm{D}_x^0 f = f(x)$; si $p = 1$ alors $\mathrm{D}_x f = f = g$ pour tout $x \in \mathrm{E}$. Du coup, $\mathrm{D}_{x+h} f - \mathrm{D}_x f = 0$ impliquant ainsi $\mathrm{D}^2 f = 0$ et plus généralement $\mathrm{D}^p f = 0$ pour tout $p \geq 2$.

Supposons maintenant la proposition établie jusqu'au rang $(k-1) \geq 1$ et prenons un polynôme homogène de degré k de sorte que $f = g \circ \Delta_k$. D'après l'exemple 6.7, on obtient

$$\mathrm{D}_x f(h) = \mathrm{D}_{\Delta_k(x)} g(\Delta_k(h)) = kg(\underbrace{x, \ldots, x}_{k-1}, h) = kg(\Delta_{k-1}(x), \Delta_1(h))$$

par symétrie, donc $\mathrm{D}f$ est un polynôme homogène de degré $(k-1)$ à valeurs dans $\mathcal{L}(\mathrm{E}, \mathrm{F})$. L'hypothèse de récurrence s'applique de sorte que

– Si $p \leq k$, alors
$$\mathrm{D}_x^{p-1}(\mathrm{D}f) \circ \Delta_{p-1} = \frac{k \times (k-1)!}{[(k-1)-(p-1)]!} g(\Delta_{k-p}(x), \Delta_{p-1}, \Delta_1)$$
donc
$$\mathrm{D}_x^p f \circ \Delta_p = \mathrm{D}_x^{p-1}(\mathrm{D}f) \circ \Delta_p = \frac{k!}{(k-p)!} g(\Delta_{k-p}(x), \Delta_p).$$

– Si $p > k$ alors $p - 1 > k - 1$ et $\mathrm{D}_x^p f = \mathrm{D}_x^{p-1}(\mathrm{D}f) = 0$.

■

Démonstration du théorème 7.2. — On considère pour chaque $j \geq 1$, l'application linéaire $\Delta_j : \mathrm{E} \to \mathrm{E}^j$ définie par $\Delta_j(x) = (\underbrace{x, \ldots, x}_{j \text{ fois}})$. Il existe des polynômes homogènes P_k et des applications k-linéaires et symétriques g_k de degré k pour k variant de 0 à d tels que $\mathrm{P}_k = g_k \circ \Delta_k$ et $\mathrm{P} = \sum_{k=0}^d \mathrm{P}_k$.

D'après la proposition précédente, on a

$$\mathrm{D}_x^p \mathrm{P}_k(\Delta_p(h)) = \begin{cases} \frac{k!}{(k-p)!} g_k(\Delta_{k-p}(x), \Delta_p(h)) & \text{si } 0 \leq p \leq k \\ 0 & \text{si } p > k \end{cases}$$

Par conséquent, $\mathrm{D}_0^k \mathrm{P} = k! g_k$ et

$$\sum_{k=0}^{d} \frac{1}{k!} \mathrm{D}_0^k \mathrm{P}(\Delta_k(x)) = \sum_{k=0}^{d} g_k(\Delta_k(x)) = \mathrm{P}(x).$$

■

Si $f : \mathrm{U} \to \mathrm{F}$ est une application p fois différentiable, alors ses différentielles sont des applications multilinéaires symétriques, donc on peut appliquer ce qui précède. On introduit d'abord quelques notations.

Notations.— Afin de simplifier les notations, on introduit le formalisme suivant. Un *multi-indice* est un n-uplet d'entiers naturels $\alpha = (\alpha_1, \dots, \alpha_n)$, $\alpha_j \in \mathbb{N}$. Sa *longueur* vaut $|\alpha| = \sum_{j=1}^n \alpha_j$.

On définit aussi $\alpha! = \alpha_1! \times \ldots \times \alpha_n!$ et $h^\alpha = h_1^{\alpha_1} \ldots h_n^{\alpha_n}$ si $h = (h_1, \dots, h_n)$. Enfin, pour une application p fois différentiable définie sur un ouvert de $\mathbb{R}^n$, on écrit, pour un multi-indice de longueur p

$$\partial_\alpha f = \partial_1^{\alpha_1} \ldots \partial_n^{\alpha_n} f = \frac{\partial^p f}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}$$

où le symbole $\partial_j^{\alpha_j}$ signifie que l'on dérive α_j fois dans la direction e_j.

Lemme 7.4. — *Étant donné un multi-indice $\alpha \in \mathbb{N}^n$ de longueur p, le nombre $m(\alpha)$ de p-uplet $(j_1, \dots, j_p)$ qui comportent exactement α_k termes prenant la valeur k vaut*

$$m(\alpha) = \frac{p!}{\alpha!} = \frac{p!}{\alpha_1! \ldots \alpha_n!}\,.$$

Démonstration. — Le nombre $m(\alpha)$ est déterminé par le choix de α_1 indices parmi p, puis de α_2 indices parmi les $(p - \alpha_1)$ restant, etc. D'où

$$m(\alpha) = \binom{p}{\alpha_1} \times \binom{p-\alpha_1}{\alpha_2} \times \ldots \times \binom{p - \sum_{j=1}^{n-1} \alpha_j}{\alpha_n} = \frac{p!}{\alpha_1! \ldots \alpha_n!}\,.$$

■

Corollaire 7.5 (formule du binôme généralisé). — *On a, pour $x_1, \dots, x_n \in \mathbb{R}$, $k \geq 1$,*

$$\left(\sum_{j=1}^n x_j\right)^k = \sum_{\alpha \in \mathbb{N}^n,\ |\alpha| = k} \frac{k!}{\alpha!} x^\alpha\,.$$

Corollaire 7.6 (expression développée de la différentielle). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application p fois différentiable en un point a d'un ouvert U de $\mathbb{R}^n$. On a, pour tout $h \in \mathbb{R}^n$,*

$$\frac{1}{p!} \mathrm{D}_a^p f(\underbrace{h, \dots, h}_{p}) = \sum_{\alpha \in \mathbb{N}^n, |\alpha| = p} \frac{1}{\alpha!} \partial_\alpha f(a) h^\alpha\,.$$

7.2 Formule de Taylor-Young

Cette formule décrit le comportement local d'une fonction au voisinage d'un point.

Théorème 7.7 (Taylor-Young). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application différentiable p fois au point $a \in \mathrm{U}$. Soit $r > 0$ tel que $\mathrm{B}(a, r) \subset \mathrm{U}$. Il existe $\varepsilon : \mathrm{B}(0, r) \to \mathrm{F}$ telle que $\lim_0 \varepsilon = 0$ et, pour tout $h \in \mathrm{B}(0, r)$,*

$$f(a + h) = \sum_{k=0}^p \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{h, \dots, h}_{k}) + \|h\|^p \varepsilon(h)\,.$$

Démonstration. — On procède par récurrence. Le cas $p = 1$ est connu. On suppose que la formule est établi jusqu'au rang $(p - 1) \geq 1$. On considère la fonction auxiliaire $g : \mathrm{B}(0, r) \to \mathrm{F}$ définie par

$$g(x) = f(a + x) - \sum_{k=0}^p \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{x, \dots, x}_{k})$$

que l'on différentie grâce à la formule établies pour les polynômes homogènes:

$$\mathrm{D}_x g(h) = \mathrm{D}_{a+x} f(h) - \sum_{k=0}^{p} \frac{1}{(k-1)!} \mathrm{D}_a^k f(\underbrace{x, \dots, x}_{k-1}, h)\,.$$

Soit $\varepsilon > 0$; comme $\mathrm{D}f$ est différentiable $(p-1)$ fois au point a, il existe $\delta > 0$ telle que, si $\|x\| < \delta$, alors $\|\mathrm{D}_x g\| \leq \varepsilon \|x\|^{p-1}$. Du coup, le théorème de la moyenne implique, pour $h \in \mathrm{B}(0, \delta)$,

$$\begin{aligned} \left\| f(a+h) - \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{h, \dots, h}_{k}) \right\| &= \|g(h) - g(0)\| \\ &\leq \sup_{x \in \mathrm{B}(0, \|h\|)} \|\mathrm{D}_x g\| \cdot \|h\| \\ &\leq \varepsilon \cdot \|h\|^p\,. \end{aligned}$$

■

7.3 Formule de Taylor avec reste intégral

Par convention, si $f = (f_1, \dots, f_n) : [a, b] \to \mathbb{R}^n$, est une application continue, on pose

$$\int_a^b f(x)dx = \begin{pmatrix} \int_a^b f_1(x)dx \\ \vdots \\ \int_a^b f_n(x)dx \end{pmatrix}$$

Théorème 7.8 (Taylor avec reste intégral). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application de classe $\mathcal{C}^{p+1}(\mathrm{U})$ définie sur un ouvert convexe. Pour tout $a, b \in \mathrm{U}$, on a*

$$f(b) = \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{b-a, \dots, b-a}_{k}) + \frac{1}{p!} \int_0^1 (1-t)^p \mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \dots, b-a}_{p+1})\, dt\,.$$

Démonstration. — Cela découle de la formule en dimension 1; Soit $g : \mathrm{I} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^{p+1}(\mathrm{I})$ définie sur un intervalle ouvert contenant $[0, 1]$. Alors on a

$$g(1) = \sum_{k=0}^{p} \frac{1}{k!} g^{(k)}(0) + \frac{1}{p!} \int_0^1 (1-t)^p g^{(p+1)}(t)dt \tag{7.1}$$

Montrons (7.1) par récurrence sur p. Si $p = 0$, c'est le fameux théorème fondamental de l'analyse:

$$g(1) - g(0) = \int_0^1 g'(t)dt\,.$$

Supposons la formule (7.1) établie pour les fonctions de classe $\mathcal{C}^p$, et supposons g de classe $\mathcal{C}^{p+1}$. On part donc de

$$g(1) = \sum_{k=0}^{p-1} \frac{1}{k!} g^{(k)}(0) + \frac{1}{(p-1)!} \int_0^1 (1-t)^{p-1} g^{(p)}(t)dt$$

et on intègre par parties le reste intégral en prenant une primitive de $(1-t)^{p-1}$ de sorte que

$$\begin{aligned} \frac{1}{(p-1)!} \int_0^1 (1-t)^{p-1} g^{(p)}(t)dt &= \frac{-1}{p!} (1-t)^p g^{(p)}(t) \Big|_0^1 + \frac{1}{p!} \int_0^1 (1-t)^p g^{(p+1)}(t)dt \\ &= \frac{1}{p!} g^{(p)}(0) + \frac{1}{p!} \int_0^1 (1-t)^p g^{(p+1)}(t)dt\,. \end{aligned}$$

Du coup, on obtient bien (7.1).

La formule de Taylor avec reste intégral s'obtient en appliquant (7.1) à la fonction auxiliaire $g : [0, 1] \to \mathbb{R}^n$ définie par $g(t) = f(a + t(b-a))$. ■

Corollaire 7.9 (Majoration du reste). — *Soit $f : \mathrm{U} \to \mathrm{F}$ une application de classe $\mathcal{C}^{p+1}(\mathrm{U})$ définie sur un ouvert convexe. Si $\|\mathrm{D}^{p+1}f\| \leq \mathrm{M}$ sur U, alors, pour tout $a, b \in \mathrm{U}$, on a*

$$\left\| f(b) - \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) \right\| \leq \frac{\mathrm{M}}{(p+1)!} \|b-a\|^{p+1} .$$

Démonstration. — On estime le reste intégral

$$\begin{aligned}
\left\| \frac{1}{p!} \int_0^1 (1-t)^p \mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1})\, dt \right\| &\leq \frac{1}{p!} \int_0^1 (1-t)^p \|\mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1})\| dt \\
&\leq \frac{1}{p!} \int_0^1 (1-t)^p \mathrm{M}\, \|b-a\|^{p+1}\, dt \\
&\leq \frac{\mathrm{M}}{(p+1)!} \|b-a\|^{p+1}
\end{aligned}$$

■

7.4 Formule de Taylor-Lagrange

Dans ce paragraphe, on s'intéresse à des applications à valeurs réelles $f : \mathrm{U} \to \mathbb{R}$ définies sur un ouvert convexe d'un espace vectoriel normé E.

Théorème 7.10 (Taylor-Lagrange). — *Soit $f : \mathrm{U} \to \mathbb{R}$ une application définie sur un ouvert convexe d'un evn E, différentiable $(p+1)$ fois sur U. Pour tous $a, b \in \mathrm{U}$, il existe $c \in]a, b[$ tel que*

$$f(b) = \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \frac{1}{(p+1)!} \mathrm{D}_c^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}) .$$

Démonstration. — On considère la fonction auxiliaire $g : [0, 1] \to \mathbb{R}$ définie par

$$g(t) = \sum_{k=0}^{p} \frac{(1-t)^k}{k!} \mathrm{D}_{a+t(b-a)}^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \mathrm{K} \frac{(1-t)^{p+1}}{(p+1)!} - f(b)$$

où l'on fixe la valeur de K de sorte que $g(0) = 0$. On différentie g, différentiable par hypothèses:

$$\begin{aligned}
g'(t) &= \sum_{k=0}^{p} \frac{(1-t)^k}{k!} \mathrm{D}_{a+t(b-a)}^{k+1} f(\underbrace{b-a, \ldots, b-a}_{k+1}) - \sum_{k=1}^{p} \frac{(1-t)^{k-1}}{(k-1)!} \mathrm{D}_{a+t(b-a)}^{k} f(\underbrace{b-a, \ldots, b-a}_{k}) - \mathrm{K} \frac{(1-t)^p}{p!} \\
&= \sum_{k=1}^{p+1} \frac{(1-t)^{k-1}}{(k-1)!} \mathrm{D}_{a+t(b-a)}^{k} f(\underbrace{b-a, \ldots, b-a}_{k}) - \sum_{k=1}^{p} \frac{(1-t)^{k-1}}{(k-1)!} \mathrm{D}_{a+t(b-a)}^{k} f(\underbrace{b-a, \ldots, b-a}_{k}) - \mathrm{K} \frac{(1-t)^p}{p!} \\
&= \frac{(1-t)^p}{p!} \mathrm{D}_{a+t(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}) - \mathrm{K} \frac{(1-t)^p}{p!} .
\end{aligned}$$

Or $g(1) = g(0) = 0$, donc le théorème de Rolle implique l'existence de $\theta \in]0, 1[$ tel que $g'(\theta) = 0$, ce qui nous permet d'obtenir

$$\mathrm{K} = \mathrm{D}_{a+\theta(b-a)}^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}) .$$

Le fait que $g(0) = 0$ implique, en notant $c = a + \theta(b-a)$,

$$f(b) = \sum_{k=0}^{p} \frac{1}{k!} \mathrm{D}_a^k f(\underbrace{b-a, \ldots, b-a}_{k}) + \frac{1}{(p+1)!} \mathrm{D}_c^{p+1} f(\underbrace{b-a, \ldots, b-a}_{p+1}) .$$

■

Remarque 7.11. — *Cette formule n'est pas valide pour des fonctions à valeurs vectoriels. Autrement dit, il n'y a pas de raison que l'on puisse choisir la même valeur du segment pour chaque composante. Prenons par exemple* $g : [0,1] \to \mathbb{R}^2$ *définie par* $g(t) = (t^2, t^3)$*. On a donc* $g'(t) = (2t, 3t)$*,* $g(1) - g(0) = (1,1)$ *et il n'existe pas de valeurs de* t *vérifiant* $2t = 3t = 1$*.*