Classes préparatoires aux grandes écoles

Filière scientifique • Mathématiques

Calcul différentiel et optimisation

Mouad El Assadi

Professeur agrégé de mathématiques

Ingénieur d'État en génie civil — École Hassania des Travaux Publics (EHTP)

elassadimouad1@gmail.com

Table des matières

I Rappels sur les fonctions vectorielles **2**
I.1 Extension des notations o et O . . . 2
I.2 Dérivation de fonctions vectorielles d'une variable réelle . . . 5
I.3 Fonctions vectorielles de classe C^k . . . 7

II Dérivées partielles et différentielles **7**
II.1 Dérivées selon un vecteur . . . 8
II.2 Différentielles . . . 15
II.2.1 Matrice de la différentielle . . . 19
II.3 Gradient . . . 22
II.4 Opérations sur les applications différentiables . . . 26

III Applications de classe C^n **30**
III.1 Applications de classe C^1 . . . 30
III.2 Plan tangent et noyau de la différentielle . . . 33
III.3 Applications de classe C^k . . . 36
III.4 Hessienne et formule de Taylor-Young à l'ordre 2 . . . 38
III.5 Équations aux dérivées partielles (EDP) . . . 39

IV Optimisation **41**
IV.1 Principes généraux . . . 41
IV.2 Condition nécessaire du premier ordre . . . 42
IV.3 Optimisation sous contrainte . . . 43
IV.4 Étude à l'ordre 2 . . . 47

I. Rappels sur les fonctions vectorielles

I.1 Extension des notations o et O

Dans cette sous-section, E est un $\mathbb{R}$-e.v.n. de dimension finie, F est un K-e.v.n. de dimension finie, et $D \subset E$ est une partie quelconque de E.

Définition I.1 (Extension des notations o et O)**.** *Soit f est une application de D dans F et $g : D \to \mathbb{R}$. Soit $x_0 \in D$. On dira que*

- $f(x) = \underset{x_0}{} o(g(x))$ *s'il existe $\alpha : D \to F$ tel que*

$$\exists V \in \mathcal{V}(x_0), \quad \forall x \in V \cap D, \quad f(x) = g(x)\alpha(x),$$

et

$$\lim_{x \to x_0} \alpha(x) = 0;$$

- $f(x) = \underset{x_0}{} O(g(x))$ *s'il existe $\alpha : D \to F$ tel que*

$$\exists V \in \mathcal{V}(x_0), \quad \forall x \in V \cap D, \quad f(x) = g(x)\alpha(x),$$

et α bornée sur $V \cap D$.

Remarque I.2. *Le terme d'erreur sera toujours contrôlé par une fonction à valeur dans $\mathbb{R}$, et non une fonction vectorielle.*

Proposition I.3 (Caractérisation de o et O par les coordonnées)**.** *Soit $f : I \to F$ et $g : I \to \mathbb{R}$, et $B = (b_1, \dots, b_n)$ une base de F, et pour tout $x \in I$*

$$f(x) = \sum_{i=1}^{n} f_i(x) b_i.$$

Alors

- $f(x) = \underset{x_0}{} o(g(x))$ *si et seulement si pour tout $i \in [\![1, n]\!]$, $f_i(x) = \underset{x_0}{} o(g_i(x))$;*
- $f(x) = \underset{x_0}{} O(g(x))$ *si et seulement si pour tout $i \in [\![1, n]\!]$, $f_i(x) = \underset{x_0}{} O(g_i(x))$.*

Éléments de preuve. Écrire α avec ses coordonnées, et utiliser la norme $\| \cdot \|_{B,\infty}$. □

Exercice I.4. *1.* $\begin{pmatrix} t \\ t^2 \\ \ln(t) \end{pmatrix} =_{+\infty} o(t^3)$*;*

2. $\begin{pmatrix} e^t \\ \sin t \\ \frac{1}{1+t} \\ \arctan t \end{pmatrix} =_0 \begin{pmatrix} 1+t \\ t \\ 1-t \\ t \end{pmatrix} + O(t^2)$*;*

3. Pour $A \in M_n(\mathbb{C})$*,* $e^A = \sum_{k=0}^{n} \frac{A^k}{k!} + o(\|A\|^n)$ *quand* $A \to 0$*.*

Solution:

1. *Pour montrer que c'est un* $o(t^3)$ *quand* $t \to +\infty$*, vérifions chaque composante :*
   - $\frac{t}{t^3} = \frac{1}{t^2} \to 0$
   - $\frac{t^2}{t^3} = \frac{1}{t} \to 0$
   - $\frac{\ln(t)}{t^3} \to 0$ *(par croissance comparée)*

   Donc $\begin{pmatrix} t \\ t^2 \\ \ln(t) \end{pmatrix} =_{+\infty} o(t^3)$*.*

2. *Développons chaque composante au voisinage de 0 :*

$$e^t = 1 + t + \frac{t^2}{2} + O(t^3)$$
$$\sin t = t - \frac{t^3}{6} + O(t^5) = t + O(t^3)$$
$$\frac{1}{1+t} = 1 - t + t^2 + O(t^3)$$
$$\arctan t = t - \frac{t^3}{3} + O(t^5) = t + O(t^3)$$

   Ainsi :

$$\begin{pmatrix} e^t \\ \sin t \\ \frac{1}{1+t} \\ \arctan t \end{pmatrix} = \begin{pmatrix} 1+t \\ t \\ 1-t \\ t \end{pmatrix} + \begin{pmatrix} \frac{t^2}{2} + O(t^3) \\ O(t^3) \\ t^2 + O(t^3) \\ O(t^3) \end{pmatrix}$$

   Le vecteur reste est un $O(t^2)$*, donc :*

$$\begin{pmatrix} e^t \\ \sin t \\ \frac{1}{1+t} \\ \arctan t \end{pmatrix} =_0 \begin{pmatrix} 1+t \\ t \\ 1-t \\ t \end{pmatrix} + O(t^2)$$

3. *Pour* $A \in M_n(\mathbb{C})$*, considérons la série de l'exponentielle :*

$$e^A = \sum_{k=0}^{+\infty} \frac{A^k}{k!}$$

   On peut écrire :

$$e^A = \sum_{k=0}^{n} \frac{A^k}{k!} + \sum_{k=n+1}^{+\infty} \frac{A^k}{k!}$$

Soit $\|\cdot\|$ une norme d'algèbre sur $M_n(\mathbb{C})$ (c'est-à-dire une norme sous-multiplicative : $\|AB\| \leq \|A\|\|B\|$). Majorons le reste :

$$\left\| \sum_{k=n+1}^{+\infty} \frac{A^k}{k!} \right\| \leq \sum_{k=n+1}^{+\infty} \frac{\|A\|^k}{k!}$$

Utilisons maintenant l'exponentielle réelle :

$$\sum_{k=n+1}^{+\infty} \frac{\|A\|^k}{k!} = e^{\|A\|} - \sum_{k=0}^{n} \frac{\|A\|^k}{k!}$$

Or, on sait que pour $x \to 0$,

$$e^x - \sum_{k=0}^{n} \frac{x^k}{k!} = o(x^n)$$

Donc quand $\|A\| \to 0$, on a :

$$\sum_{k=n+1}^{+\infty} \frac{\|A\|^k}{k!} = o(\|A\|^n)$$

Ce qui implique que :

$$\left\| \sum_{k=n+1}^{+\infty} \frac{A^k}{k!} \right\| = o(\|A\|^n)$$

Ainsi, quand $A \to 0$:

$$e^A = \sum_{k=0}^{n} \frac{A^k}{k!} + o(\|A\|^n)$$

On en déduit, pour $t \in \mathbb{R}$: $e^{At} = I_n + tA + o(t)$.

La situation la plus fréquente sera le cas d'un terme d'erreur en $\|x\|^\alpha$, x étant une variable de l'e.v.n. E de dimension finie.

Lemme I.5 (Invariance vis-à-vis de la norme en dimension finie)**.** *Soit E de dimension finie et $\|\cdot\|_1$, $\|\cdot\|_2$ deux normes quelconques sur E. Soit $D \subset E$ et $f : E \to F$, et $x_0 \in F$. Soit $\alpha \in \mathbb{R}$. Alors*

$$f(x) = \underset{x \to x_0}{o}(\|x\|_1^\alpha) \iff f(x) = \underset{x \to x_0}{o}(\|x\|_2^\alpha),$$

et

$$f(x) = \underset{x \to x_0}{O}(\|x\|_1^\alpha) \iff f(x) = \underset{x \to x_0}{O}(\|x\|_2^\alpha).$$

Pour $h \in E$, on trouve parfois par abus la notation $o(h)$ à la place de $o(\|h\|)$, et de même pour $O(h)$, dans le but d'alléger un peu les notations.
Lorsque $F = \mathbb{R}$ ou $\mathbb{C}$, on peut former des quotients, et on retrouve la caractérisation usuelle, généralisée au cas d'une variable vectorielle.

Proposition I.6 (Caractérisation de o par le quotient)**.** *Si $g : D \to \mathbb{R}$ ne s'annule pas au voisinage de x_0, les propositions suivantes sont équivalentes:*

1. $f(x) = o(g(x))$;
2. $\frac{f(x)}{g(x)} \to 0$.

Proposition I.7 (Caractérisation de O par le quotient)**.** *Si $g : D \to \mathbb{R}$ ne s'annule pas au voisinage épointé de x_0, les propositions suivantes sont équivalentes:*

1. $f(x) = O(g(x))$;
2. $\frac{f(x)}{g(x)}$ *est bornée au voisinage de* x_0.

Les propriétés opératoires sur les o et O sont les mêmes que dans le cas réel (sommation, transitivité etc).

I.2 Dérivation de fonctions vectorielles d'une variable réelle

Dans cette sous-section, on suppose que $E = \mathbb{R}$ et $I \subset E$ est un intervalle ouvert. On considère donc $f : I \to F$ une fonction d'une variable réelle à valeurs dans E.

Définition I.8 (Fonction dérivable en a)**.** *Soit $f : I \to F$ et $a \in I$. On dit que f est dérivable en a si le taux d'accroissement*

$$\tau_{x_0} f(x) = \frac{f(x) - f(a)}{x - a}$$

admet une limite finie $\ell \in F$ lorsque x tend vers a. Le vecteur dérivé en a est alors défini par :

$$f'(a) = \lim_{x \to x_0} \frac{f(x) - f(a)}{x - a} = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

La deuxième limite est prise en considérant $h \in I - x_0 = \{x - x_0, x \in I\}$. Ainsi, si x_0 est la borne inférieure de I, il s'agit d'une limite lorsque h tend vers 0^+.

Définition I.9 (Fonction dérivable sur I, fonction dérivée)**.** *Soit $f : I \to F$. On dit que f est dérivable sur I si f est dérivable en tout $x_0 \in I$. On définit alors la fonction dérivée $f' : I \to F$ par :*

$$\forall x \in I, \quad f'(x) = \lim_{y \to x} \frac{f(y) - f(x)}{y - x}.$$

Exemple I.10.

1. *Dérivée de $f : t \mapsto tX_0$, $X_0 \in K$.*
2. *Plus généralement, dérivée de $f : t \mapsto g(t)X_0$, $g : I \to \mathbb{R}$ étant une fonction dérivable en a, et X_0 un vecteur de F.*
3. *Dérivée de $t \mapsto e^{1/t}$.*

Proposition I.11 (Caractérisation de la dérivabilité par les coordonnées)**.** *Soit $f : I \to F$, $x_0 \in I$, et B une base de F. On note, pour tout $x \in I$,*

$$f(x) = \sum_{i=1}^{n} f_i(x) b_i,$$

la décomposition de f suivant ses composantes sur la base B. Les propositions suivantes sont équivalentes:

1. *f est dérivable en x_0;*
2. *pour tout $i \in [\![1, n]\!]$, f_i est dérivable en x_0.*

Dans ce cas, on a

$$f'(x_0) = \sum_{i=1}^{n} f_i'(x_0) b_i.$$

Éléments de preuve. Former le taux d'accroissement coordonnée par coordonnée, et utiliser la caractéristique similaire pour les limites. □

La caractérisation par les coordonnées permet de montrer qu'en définissant la dérivabilité à gauche et à droite comme dans le cas d'une fonction réelle, et en utilisant la caractérisation de la dérivée par les dérivées à gauche et à droite sur chacune de coordonnée, cela permet d'obtenir une caractérisation similaire dans le cas vectoriel.

Corollaire I.12 (Opérations sur les dérivées)**.** *Soit I et J des intervalles de $\mathbb{R}$, et F un e.v.n. de dimension finie.*

1. *La dérivation est linéaire : si $f, g : I \to F$ et $\lambda \in K$, alors si f et g sont dérivables en $a \in I$, $\lambda f + g$ aussi, de dérivée*
$$(\lambda f + g)'(a) = \lambda f'(a) + g'(a).$$
2. *Si $\phi : J \to I$ est une fonction réelle dérivable, et $f : I \to F$ une fonction vectorielle dérivable, alors*
$$\forall t \in J, \quad (f \circ \phi)'(t) = f' \circ \phi(t) \cdot \phi'(t).$$

Éléments de preuve. Voir coordonnées par coordonnées. □

Corollaire I.13 (Dérivabilité à gauche, à droite). *Soit $x_0 \in I$ et $f : I \to F$. La fonction f est dérivable en x_0 si et seulement si elle est dérivable à gauche et à droite en x_0 et $f'_g(x_0) = f'_d(x_0)$.*

Corollaire I.14 (Dérivable implique continu). *Soit $f : I \to \mathbb{R}$ et $x_0 \in I$. Si f est dérivable en x_0, alors f est continue en x_0.*

Éléments de preuve. La caractérisation précédente nous ramène au cas des fonctions d'une variable. On peut aussi utiliser directement la propriété suivante (DL à l'ordre 1, qu'on peut ensuite restreindre à l'ordre 0). □

Proposition I.15 (Caractérisation de la dérivabilité par DL à l'ordre 1). *Soit $f : I \to F$ et $x_0 \in I$. Les propositions suivantes sont équivalentes:*

1. *f est dérivable en x_0;*
2. *il existe A et B dans F tels que $f(x_0 + h) =_{h\to 0} A + hB + o(h)$.*

Éléments de preuve. Caractérisation de o par quotient, puisque t ne s'annule pas au voisinage épointé de 0. □

On donne une première conséquence, qui est plutôt un exemple, suffisamment important pour qu'il soit un résultat du cours (il servira pour l'étude des équations différentielles).

Corollaire I.16 (Dérivée de $t \mapsto \exp(At)$). *Soit $A \in M_n(\mathbb{C})$. La fonction $f : t \mapsto \exp(At)$ est dérivable sur $\mathbb{R}$, et*
$$\forall t \in \mathbb{R}, \quad f'(t) = A \exp(At).$$

Une autre conséquence de la caractérisation par les DL est l'interprétation du vecteur dérivé en terme d'approximation de la trajectoire.

Corollaire I.17 (Tangente à la trajectoire). *Soit $f : I \to F$ et $x_0 \in I$. On suppose que f est dérivable en x_0 et que $f'(x_0) \neq 0$. Alors la droite affine $f(x_0) + \mathbb{R}f'(x_0)$ est tangente à la trajectoire (dans le sens où c'est la droite qui approche au mieux la trajectoire au voisinage de $f(x_0)$).*

Remarque I.18 (Interprétation cinétique de la dérivée). *Si f représente la trajectoire d'un point dans F, et si f est dérivable en x_0 :*

- *$\|f'(x_0)\|_F$ est la vitesse instantanée du point (dépendant de la norme choisie sur F)*
- *le vecteur $f'(x_0)$ est tangent à la trajectoire, et indique le sens de déplacement.*

Exemple I.19. *Décrire la trajectoire, et la vitesse instantanée d'un point, dont la position est définie pour $t \in \mathbb{R}_+$ par*
$$f(t) = (\cos t, \sin t, t).$$

Théorème I.20 (Caractérisation des fonctions constantes). *Soit $f : I \to F$, dérivable sur l'intervalle I, et à valeurs dans l'e.v.n. de dimension finie F. Les propositions suivantes sont équivalentes:*

1. *f est constante sur I*
2. *$f' = 0$ sur I.*

Éléments de preuve. Voir dans une base. □

I.3 Fonctions vectorielles de classe C^k

On considère toujours dans cette sous-section le cas d'une fonction $f : I \to F$ définie sur un intervalle I ouvert, et à valeurs dans un e.v.n. F de dimension finie. Soit $B = (b_1, \dots, b_n)$ une base de F, et, pour $x \in I$,

$$[f(x)]_B = \begin{pmatrix} f_1(x) \\ \vdots \\ f_n(x) \end{pmatrix}.$$

Si f est dérivable sur I, alors elle définit une fonction dérivée

$$\forall x \in I, \quad f'(x) = \sum_{k=1}^{n} f_k'(x) b_k, \quad \text{i.e.} \quad [f'(x)]_B = \begin{pmatrix} f_1'(x) \\ \vdots \\ f_n'(x) \end{pmatrix}.$$

On a encore $f' : I \to F$, et on peut donc étudier la dérivabilité de f', et ainsi de suite, comme dans le cas réel.
Si f est dérivable p fois, on note $f^{(k)}$ sa dérivée d'ordre k.

Définition I.21 (Fonction de classe C^k)**.** *La fonction $f : I \to \mathbb{R}$ est de classe C^k si elle est k fois dérivable et si $f^{(k)} = I \to F$ est continue.*

Définition I.22 (Caractérisation de la classe C^k et expression de la dérivée)**.** *Avec les notations introduites ci-dessus.*

1. *f est k fois dérivable si et seulement si chaque f_i est k fois dérivable, et dans ce cas,*

$$\forall x \in I, \quad f^{(k)}(x) = \sum_{i=1}^{n} f_i^{(k)}(x) \quad \textit{i.e.} \quad [f^{(k)}(x)]_B = \begin{pmatrix} f_1^{(k)}(x) \\ \vdots \\ f_n^{(k)}(x) \end{pmatrix};$$

2. *f est de classe C^k si et seulement si chaque f_i l'est.*

Proposition I.23 (Règles opératoires sur les fonctions de classe C^k)**.** *Soit I et J deux intervalles de $\mathbb{R}$ et F un K-e.v.n. de dimension finie. Soit $f, g : I \to \mathbb{R}$, $\phi : J \to I$ et $\lambda \in K$.*

1. *Si f et g sont de classe C^k, alors $f = \lambda g$ est de classe C^k, et*

$$(f + \lambda g)^{(k)} = f^{(k)} + \lambda g^{(k)}.$$

2. *Si f et ϕ sont de classe C^k, $f \circ \phi$ est de classe C^k. Il n'y a pas de formule simple en général.*
3. *Si f est de classe C^k et $\varphi : t \mapsto at + b$, alors $(f \circ \varphi)^{(k)} = a^k (f^{(k)} \circ \varphi)$.*

Éléments de preuve. Ces propriétés sont bien connues pour les fonctions à valeurs réelles. S'y ramener en travaillant sur les coordonnées. □

II. Dérivées partielles et différentielles

On s'intéresse maintenant à la situation plus générale d'une fonction $f : \Omega \to F$, où Ω est un ouvert d'un $\mathbb{R}$-e.v.n. E de dimension finie, et F est un K-e.v.n. de dimension finie.

II.1 Dérivées selon un vecteur

Définition II.1 (Fonctions partielles). *Soit $f : \Omega \to F$, $a \in \Omega$, et $\vec{u} \in E \setminus \{0\}$. La fonction partielle $f_{a,\vec{u}}$ au point a et de direction $\vec{u}$, est définie, en tout $t \in \mathbb{R}$ tel que $a + t\vec{u} \in \Omega$, par*

$$f_{a,\vec{u}}(t) = f(a + t\vec{u}).$$

Remarque II.2.
1. *La fonction partielle $f_{a,\vec{u}}$ est bien définie au voisinage de 0, car Ω est ouvert.*
2. *La flèche sur $\vec{u}$ est présente ici uniquement pour faire remarquer que c'est la structure affine naturelle de E qui est en jeu ici. Les éléments de E jouent donc un rôle double de points et de vecteurs. Les éléments en lesquels f est évalué correspondent plutôt à des points, alors que $\vec{u}$ donne une direction de translation, et doit plutôt être considéré comme un vecteur. La fonction $f_{a,\vec{u}}$ correspond plus ou moins, après reparamétrage, à la restriction de f à la droite affine passant par a et dirigée par le vecteur $\vec{u}$.*

Définition II.3 (Dérivée selon un vecteur). *Soit $f : \Omega \to F$, $a \in \Omega$, et $\vec{u} \in E \setminus \{0\}$.*

1. *La fonction f est dérivable selon le vecteur $\vec{u}$ si la fonction partielle $f_{a,\vec{u}}$ est dérivable en 0.*
2. *La dérivée de f en a selon $\vec{u}$ est alors défini par*

$$D_{\vec{u}}f(a) = f'_{a,\vec{u}}(0) = \lim_{t\to 0} \frac{f(a + t\vec{u}) - f(a)}{t} \in F.$$

3. *Si f est dérivable selon $\vec{u}$ en tout point a de Ω, on peut définir la fonction $D_{\vec{u}}f : \Omega \to F$.*

Remarque II.4.
1. *Dans la suite, afin d'alléger les notations, on omettra la flèche sur u, en notant simplement $D_u f(a)$.*
2. *L'existence d'une dérivée selon u en a n'implique pas l'existence de dérivées selon d'autres vecteurs non colinéaires; en revanche, si $v \neq 0$ est colinéaire à u, l'existence de $D_u f(a)$ équivaut à celle de $D_v f(a)$, et on peut relier les deux quantités.*

Exemple II.5. *Étudier l'existence des dérivées en 0 selon $u \in \mathbb{R}^2$ de $f : \mathbb{R}^2 \to \mathbb{R}$ définie par*

$$f(x, y) = \mathbb{I}_{\mathbb{R}^2\setminus\{0\}}(x, y).$$

Solution:
La fonction $f : \mathbb{R}^2 \to \mathbb{R}$ est définie par :

$$f(x, y) = \mathbb{I}_{\mathbb{R}^2\setminus\{0\}}(x, y) = \begin{cases} 1 & \text{si } (x, y) \neq (0, 0) \\ 0 & \text{si } (x, y) = (0, 0) \end{cases}$$

On étudie l'existence de la dérivée directionnelle de f en $(0, 0)$ selon un vecteur $u = (h, k) \in \mathbb{R}^2$, avec $u \neq (0, 0)$.
Par définition, la dérivée directionnelle en $(0, 0)$ selon u est :

$$D_u f(0, 0) = \lim_{t\to 0} \frac{f(tu) - f(0, 0)}{t} = \lim_{t\to 0} \frac{f(th, tk)}{t}$$

Par définition de la dérivée directionnelle $u \neq (0, 0)$ (c'est-à-dire $h \neq 0$ ou $k \neq 0$).
Pour $t \neq 0$, on a $f(th, tk) = 1$ puisque $(th, tk) \neq (0, 0)$ lorsque $u \neq (0, 0)$. Ainsi :

$$D_u f(0, 0) = \lim_{t\to 0} \frac{1 - 0}{t} = \lim_{t\to 0} \frac{1}{t}$$

Cette limite n'existe pas car :

$$\lim_{t\to 0^+} \frac{1}{t} = +\infty \quad \textit{et} \quad \lim_{t\to 0^-} \frac{1}{t} = -\infty$$

Donc la dérivée directionnelle selon tout vecteur non nul n'existe pas (même pas à l'infini, car les limites à droite et à gauche sont de signes opposés).
Conclusion : *La fonction f n'admet de dérivée directionnelle en $(0,0)$ dans aucune direction non nulle. En particulier, elle n'est pas différentiable en $(0,0)$.*

Proposition II.6 (Linéarité de D par rapport à u sur chaque droite)**.** *Soit $u \in E\setminus\{0\}$ et $f : \Omega \to F$ dérivable en $a \in \Omega$ selon u. Soit $\lambda \in \mathbb{R}^*$. Alors f est dérivable en a selon λu et*

$$D_{\lambda u} f(a) = \lambda D_u f(a).$$

Éléments de preuve. Revenir à la définition par taux d'accroissement. □

Remarque II.7 (Avertissement)**.** *L'existence de toutes les dérivées directionnelles en a n'est pas suffisante pour assurer la continuité en a.*

Exemple II.8. *Étudier les dérivées directionnelles en $0_{\mathbb{R}^2}$ et la continuité de $f : \mathbb{R}^2 \to \mathbb{R}$ définie par*

$$f(x,y) = \begin{cases} \frac{x^2}{y} & \textit{si } y \neq 0, \\ 0 & \textit{si } y = 0. \end{cases}$$

Solution
Étudions les dérivées directionnelles en $(0,0)$ et la continuité de la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par :

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

Dérivées directionnelles en $(0,0)$
Soit $u = (h,k) \in \mathbb{R}^2$ avec $u \neq (0,0)$. La dérivée directionnelle de f en $(0,0)$ selon u est :

$$D_u f(0,0) = \lim_{t\to 0} \frac{f(th,tk) - f(0,0)}{t} = \lim_{t\to 0} \frac{f(th,tk)}{t}$$

Distinguons plusieurs cas :

1. ***Cas 1 :*** *$k = 0$ (direction horizontale).*
   Pour $k = 0$, on a $f(th,0) = 0$ par définition (car $y = 0$). Ainsi :

   $$D_{(h,0)} f(0,0) = \lim_{t\to 0} \frac{0}{t} = 0$$

   Donc f admet une dérivée directionnelle nulle selon toute direction horizontale.

2. ***Cas 2 :*** *$k \neq 0$ (direction non horizontale).*
   Pour $t \neq 0$, on a $tk \neq 0$ donc :

   $$f(th,tk) = \frac{(th)^2}{tk} = \frac{t^2h^2}{tk} = t \cdot \frac{h^2}{k}$$

   Ainsi :

   $$D_{(h,k)} f(0,0) = \lim_{t\to 0} \frac{t \cdot \frac{h^2}{k}}{t} = \lim_{t\to 0} \frac{h^2}{k} = \frac{h^2}{k}$$

   Donc f admet une dérivée directionnelle selon toute direction non horizontale, et elle vaut $\frac{h^2}{k}$.

Conclusion : *f admet des dérivées directionnelles en $(0,0)$ dans toutes les directions. Cependant, ces dérivées ne dépendent pas linéairement de (h,k), car pour une direction fixée, $D_{(h,k)}f(0,0) = \frac{h^2}{k}$ est quadratique en h et non linéaire.*
Continuité en $(0,0)$ *Pour étudier la continuité de f en $(0,0)$, utilisons la caractérisation séquentielle. Considérons deux suites particulières qui convergent vers $(0,0)$:*

1. ***Première suite :*** $\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right)$

$$f\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right) = \frac{\left(\frac{1}{\sqrt{n}}\right)^2}{\frac{1}{n}} = \frac{\frac{1}{n}}{\frac{1}{n}} = 1$$

   Ainsi, $f\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right) \to 1$ quand $n \to +\infty$.

2. ***Deuxième suite :*** $\left(\frac{1}{\sqrt{n}}, \frac{1}{2n}\right)$

$$f\left(\frac{1}{\sqrt{n}}, \frac{1}{2n}\right) = \frac{\left(\frac{1}{\sqrt{n}}\right)^2}{\frac{1}{2n}} = \frac{\frac{1}{n}}{\frac{1}{2n}} = 2$$

   Ainsi, $f\left(\frac{1}{\sqrt{n}}, \frac{1}{2n}\right) \to 2$ quand $n \to +\infty$.

On a donc deux suites $\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right)$ et $\left(\frac{1}{\sqrt{n}}, \frac{1}{2n}\right)$ qui convergent toutes deux vers $(0,0)$, mais telles que :

$$\lim_{n\to+\infty} f\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right) = 1 \quad \textit{et} \quad \lim_{n\to+\infty} f\left(\frac{1}{\sqrt{n}}, \frac{1}{2n}\right) = 2$$

Ces deux limites étant différentes, la fonction f ne peut pas être continue en $(0,0)$.

- *La fonction f admet des dérivées directionnelles en $(0,0)$ dans toutes les directions.*
- *La fonction f n'est pas continue en $(0,0)$, comme le montre la caractérisation séquentielle avec les suites $\left(\frac{1}{\sqrt{n}}, \frac{1}{n}\right)$ et $\left(\frac{1}{\sqrt{n}}, \frac{1}{2n}\right)$.*

Remarque II.9. *Cet exemple illustre qu'une fonction peut admettre des dérivées directionnelles dans toutes les directions sans être continue, et a fortiori sans être différentiable. La différentiabilité est une propriété plus forte qui requiert non seulement l'existence de dérivées directionnelles, mais aussi leur dépendance linéaire par rapport à la direction.*

Définition II.10 (Dérivées partielles dans une base)**.** *Soit $f : \Omega \to F$, $a \in \Omega$ et $B = (b_1, \dots, b_n)$ une base de E. Pour $x \in E$, on note $X = (x_1, \dots, x_n)$ son vecteur coordonnées.*

1. *Soit $k \in [1, n]$. La k-ième dérivée partielle de f dans la base B est la dérivée de f selon la direction b_k en a, et est notée $\partial_{B,k} f(a)$, ou, s'il n'y a pas d'ambiguïté sur la base utilisée, simplement $\partial_k f(a)$. Ainsi*

$$\partial_k f(a) = \partial_{B,k} f(a) = D_{b_k} f(a) = \lim_{t \to 0} \frac{f(a + tb_k) - f(a)}{t}.$$

2. *Si f admet une k-ième dérivée partielle dans la base B en tout point de Ω, cela définit une application $\partial_k f : \Omega \to F$.*

Remarque II.11 (Identification à une fonction de plusieurs variables)**.** *Soit $f : \Omega \to F$ et $B = (b_1, \dots, b_n)$ une base de E. Pour $x \in \Omega$, on note $[x]_B = (x_1, \dots, x_n)$ ses coordonnées relativement à la base B. On note*

$$\tilde{\Omega} = \left\{ (x_1, \dots, x_n) \mid \sum_{k=1}^{n} x_k b_k \in \Omega \right\}.$$

Alors $\tilde{\Omega}$ est un ouvert de $\mathbb{R}^n$, et on peut identifier la fonction f à la fonction $\tilde{f} : \tilde{\Omega} \to F$ définie par

$$\tilde{f}(x_1, \dots, x_n) = f\left(\sum_{k=1}^{n} x_k b_k \right).$$

On obtient facilement, du fait de la possibilité de choisir la norme qu'on veut (en dimension finie) et donc notamment par exemple la norme $\| \cdot \|_{B,\infty}$ que :

- *f est continue en x ssi $\tilde{f}$ est continue en $[x]_B$;*
- *f admet en x une dérivée partielle selon b_k si $\tilde{f}$ admet en $[x]_B$ une dérivée partielle selon e_k (vecteur de la base canonique), et alors*

$$D_{b_k}(f)(a) = D_{e_k}(\tilde{f})(a).$$

Lorsque Ω est un ouvert de $\mathbb{R}^n$, on note aussi $\frac{\partial f}{\partial x_k}(a)$ la k-ième dérivée partielle par rapport à la base canonique. Ainsi, pour $a = (a_1, \dots, a_n)$, sous réserve d'existence,

$$\frac{\partial f}{\partial x_k}(a) = \lim_{t \to 0} \frac{f(a_1, \dots, a_k + t, \dots, a_n) - f(a_1, \dots, a_n)}{t}.$$

On l'appelle aussi dérivée partielle par rapport à la k-ième variable.

Proposition II.12 (Identification des dérivées partielles)**.** *Avec les notations de la remarque 16.2.10,*

$$\frac{\partial \tilde{f}}{\partial x_k}(a) = D_{e_k}(f)(a) = \partial_{B,k} f(a).$$

Après avoir fixé la base B et identifié f à $\tilde{f}$, on pourra donc aussi identifier la k-ième dérivée partielles par rapport à la base B à la dérivée partielle par rapport à la k-ième variable de $\tilde{f}$, et on s'autorisera à écrire, s'il n'y a pas d'ambiguïté sur la base :

$$\partial_k f(a) = \frac{\partial f}{\partial x_k}(a).$$

Par ailleurs, il arrive souvent, dans le cadre d'une application d'un ouvert de $\mathbb{R}^n$ dans $\mathbb{R}^p$ (situation à laquelle on se ramène par choix de bases), de travailler coordonnée par coordonnée. Dans ce cadre, on est ramené au calcul de dérivées partielles d'une fonction à valeurs dans $\mathbb{R}$:

Proposition II.13 (Expression de la dérivée partielle d'une fonction à valeurs dans $\mathbb{R}^p$)**.** *Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}^p$, qu'on écrit $f = (f_1, \ldots, f_p)$. Ainsi, pour tout $x \in \Omega$, $f(x) = (f_1(x), \ldots, f_p(x)) \in \mathbb{R}^p$, et pour tout $k \in [1, p]$, $f_k : \Omega \to \mathbb{R}$. Alors f admet une dérivée partielle par rapport à la k-ième variable en a ssi c'est le cas de chacun des f_i, et dans ce cas,*

$$\frac{\partial f}{\partial x_k}(a) = \left(\frac{\partial f_1}{\partial x_k}(a), \ldots, \frac{\partial f_p}{\partial x_k}(a)\right).$$

Par ailleurs, on rappelle comment, concrètement, calculer les dérivée partielles d'une fonction dont on connaît l'expression :

Méthode II.14 (Calcul d'une dérivée partielle)**.** *Soit f une application de $\Omega \subset \mathbb{R}^n$ dans $\mathbb{R}$ (ce à quoi on peut se ramener d'après toutes les propriétés précédentes d'identification ou de travail par coordonnées). Pour calculer la dérivée partielle par rapport à x_k, on utilise les règles usuelles de calcul de la dérivée en utilisant l'expression de f, dans laquelle on dérive par rapport à la variable x_k, en considérant que toutes les autres variables sont des constantes.*

Exemple II.15. *Calculer la dérivée partielle par rapport à x de $f : (x, y, z) \mapsto (xy\ln(x+e^z), \sin(xe^z)\cos(y))$.*
***Solution**:*
Soit $f : \mathbb{R}^3 \to \mathbb{R}^2$ définie par :

$$f(x, y, z) = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \end{pmatrix} = \begin{pmatrix} xy\ln(x + e^z) \\ \sin(xe^z)\cos(y) \end{pmatrix}$$

Nous voulons calculer la dérivée partielle $\dfrac{\partial f}{\partial x}(x, y, z)$.
1. Approche par dérivée directionnelle et taux d'accroissement
Par définition, la dérivée partielle par rapport à x est la dérivée directionnelle selon le vecteur $e_1 = (1, 0, 0)$:

$$\frac{\partial f}{\partial x}(x, y, z) = D_{e_1} f(x, y, z) = \lim_{t \to 0} \frac{f(x + t, y, z) - f(x, y, z)}{t}$$

Pour calculer cette limite, faisons un développement limité en $t = 0$ des deux composantes.
***Développement limité de la première composante** Pour $f_1(x+t, y, z) = (x+t)y\ln((x+t)+e^z)$:*
Développons $\ln((x + t) + e^z)$ au voisinage de $t = 0$:

$$\ln((x + t) + e^z) = \ln(x + e^z + t) = \ln(x + e^z) + \ln\left(1 + \frac{t}{x + e^z}\right)$$

En utilisant le développement limité $\ln(1 + u) = u - \frac{u^2}{2} + o(u^2)$ avec $u = \frac{t}{x+e^z}$:

$$\ln\left(1 + \frac{t}{x + e^z}\right) = \frac{t}{x + e^z} - \frac{t^2}{2(x + e^z)^2} + o(t^2)$$

Donc :

$$\ln((x + t) + e^z) = \ln(x + e^z) + \frac{t}{x + e^z} - \frac{t^2}{2(x + e^z)^2} + o(t^2)$$

En multipliant par $(x + t)y = xy + ty$:

$$\begin{aligned} f_1(x + t, y, z) &= (xy + ty)\left[\ln(x + e^z) + \frac{t}{x + e^z} - \frac{t^2}{2(x + e^z)^2} + o(t^2)\right] \\ &= xy\ln(x + e^z) + \frac{xyt}{x + e^z} + ty\ln(x + e^z) + o(t) \end{aligned}$$

Ainsi :

$$f_1(x + t, y, z) - f_1(x, y, z) = t\left[\frac{xy}{x + e^z} + y\ln(x + e^z)\right] + o(t)$$

Développement limité de la seconde composante *Pour* $f_2(x+t,y,z) = \sin((x+t)e^z)\cos(y)$ *:*
Développons $\sin((x+t)e^z)$ *au voisinage de* $t=0$ *:*

$$\sin((x+t)e^z) = \sin(xe^z + te^z)$$

En utilisant le développement limité $\sin(a+u) = \sin a + u\cos a - \frac{u^2}{2}\sin a + o(u^2)$ *avec* $a = xe^z$ *et* $u = te^z$ *:*

$$\sin(xe^z + te^z) = \sin(xe^z) + te^z\cos(xe^z) - \frac{t^2e^{2z}}{2}\sin(xe^z) + o(t^2)$$

En multipliant par $\cos(y)$ *:*

$$f_2(x+t,y,z) = \cos(y)\sin(xe^z) + t\left[e^z\cos(y)\cos(xe^z)\right] + o(t)$$

Ainsi :

$$f_2(x+t,y,z) - f_2(x,y,z) = t\left[e^z\cos(y)\cos(xe^z)\right] + o(t)$$

2. Calcul de la limite du taux d'accroissement
En combinant les deux développements :

$$\frac{f(x+t,y,z)-f(x,y,z)}{t} = \begin{pmatrix} \dfrac{xy}{x+e^z} + y\ln(x+e^z) \\ e^z\cos(y)\cos(xe^z) \end{pmatrix} + o(1)$$

Quand $t \to 0$*, on obtient donc :*

$$\frac{\partial f}{\partial x}(x,y,z) = \lim_{t\to 0}\frac{f(x+t,y,z)-f(x,y,z)}{t} = \begin{pmatrix} \dfrac{xy}{x+e^z} + y\ln(x+e^z) \\ e^z\cos(y)\cos(xe^z) \end{pmatrix}$$

3. Vérification par dérivation usuelle
On peut vérifier ce résultat en dérivant directement chaque composante :
Pour $f_1(x,y,z) = xy\ln(x+e^z)$ *:*

$$\frac{\partial f_1}{\partial x}(x,y,z) = y\ln(x+e^z) + xy\cdot\frac{1}{x+e^z} = y\ln(x+e^z) + \frac{xy}{x+e^z}$$

Pour $f_2(x,y,z) = \sin(xe^z)\cos(y)$ *:*

$$\frac{\partial f_2}{\partial x}(x,y,z) = e^z\cos(xe^z)\cos(y)$$

Ce qui confirme notre résultat obtenu par développement limité du taux d'accroissement.
4. Conclusion *La dérivée partielle de* f *par rapport à* x *est :*

$$\boxed{\frac{\partial f}{\partial x}(x,y,z) = \begin{pmatrix} y\ln(x+e^z) + \dfrac{xy}{x+e^z} \\ e^z\cos(y)\cos(xe^z) \end{pmatrix}}$$

Remarque II.16 (Avertissement)**.** *L'existence des dérivées partielles par rapport à une base n'implique pas l'existence de dérivées selon toutes les directions en un point* a*.*

Exemple II.17. *Étudier les dérivées directionnelles en* $(0,0)$ *de* $f : (x,y) \mapsto \sqrt{|xy|}$*.*
Solution
Étudions les dérivées directionnelles en $(0,0)$ *de la fonction* $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par :*

$$f(x,y) = \sqrt{|xy|}$$

1. Rappel de la définition

Soit $u = (h, k) \in \mathbb{R}^2$ avec $u \neq (0, 0)$. La dérivée directionnelle de f en $(0, 0)$ selon u est :

$$D_u f(0,0) = \lim_{t\to 0} \frac{f(tu) - f(0,0)}{t} = \lim_{t\to 0} \frac{f(th, tk)}{t}$$

avec $f(0, 0) = \sqrt{|0 \cdot 0|} = 0$.

2. Calcul pour une direction générale

Pour $t \neq 0$, on a :

$$f(th, tk) = \sqrt{|(th)(tk)|} = \sqrt{|t^2 hk|} = |t|\sqrt{|hk|}$$

car $\sqrt{t^2} = |t|$.

Ainsi, le taux d'accroissement s'écrit :

$$\frac{f(th, tk)}{t} = \frac{|t|\sqrt{|hk|}}{t} = \frac{|t|}{t}\sqrt{|hk|}$$

3. Analyse selon la direction

Distinguons deux cas :

1. ***Cas 1 :*** *$hk = 0$, c'est-à-dire $h = 0$ ou $k = 0$.*

   Si $h = 0$ (direction verticale) ou $k = 0$ (direction horizontale), alors $\sqrt{|hk|} = 0$. Ainsi :

   $$D_{(h,k)} f(0,0) = \lim_{t\to 0} 0 = 0$$

2. ***Cas 2 :*** *$hk \neq 0$, c'est-à-dire $h \neq 0$ et $k \neq 0$.*

   Alors $\sqrt{|hk|} > 0$ et :

   $$\frac{f(th, tk)}{t} = \frac{|t|}{t}\sqrt{|hk|} = \operatorname{sgn}(t)\sqrt{|hk|}$$

   où $\operatorname{sgn}(t)$ est le signe de t.

   Cette expression a des limites différentes à droite et à gauche :

   $$\lim_{t\to 0^+} \frac{f(th, tk)}{t} = \sqrt{|hk|}$$

   $$\lim_{t\to 0^-} \frac{f(th, tk)}{t} = -\sqrt{|hk|}$$

   Puisque ces deux limites sont différentes (l'une est positive, l'autre négative), la limite bilatérale n'existe pas. Donc $D_{(h,k)} f(0, 0)$ n'existe pas.

4. Synthèse des résultats

- *Pour les directions horizontales ($k = 0$) et verticales ($h = 0$) :*

  $$D_{(h,0)} f(0,0) = 0 \quad \text{et} \quad D_{(0,k)} f(0,0) = 0$$

- *Pour toutes les autres directions ($h \neq 0$ et $k \neq 0$) :*

  $$D_{(h,k)} f(0,0) \text{ n'existe pas}$$

Remarque II.18. *Cependant, sous des hypothèses supplémentaires de régularité, les dérivées partielles suffisent à retrouver les dérivées directionnelles. Par exemple, dans le cas des fonctions de 2 variables, vous avez montré l'année dernière que si f est de classe C^1 au voisinage de a (dans le sens où ses deux dérivées partielles existent et sont continues au voisinage de a), alors, en notant $\nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x}(a) \\ \frac{\partial f}{\partial y}(a) \end{pmatrix}$:*

- *pour tout $u = (u_x, u_y)$, $D_u f(a) = \langle \nabla f(a), u \rangle = u_x \frac{\partial f}{\partial x}(a) + u_y \frac{\partial f}{\partial y}(a)$,*
- *au voisinage de $a = (a_x, a_y)$, en notant $X = (x, y)$,*

$f(X) = f(a) + (x - a_x)\frac{\partial f}{\partial x}(a) + (y - a_y)\frac{\partial f}{\partial y}(a) + o(\|X - a\|) = f(a) + \langle X - a, \nabla f(a)\rangle + o(\|X - a\|)$ *(formule de Taylor-*

II.2 Différentielles

En particulier, d'après la remarque clôturant le paragraphe précédent, les fonctions de 2 variables suffisamment régulières se comportent à première approximation comme la fonction affine

$$X \mapsto f(a) + \langle X - a, \nabla f(a)\rangle,$$

donc la partie linéaire est l'application linéaire est

$$Y \mapsto \langle Y, \nabla f(a)\rangle = D_Y f(a).$$

Ainsi, sous certaines conditions de régularité, la dérivée directionnelle en a est une application linéaire en la direction, et permet de définir un développement limité à l'ordre 1.
C'est l'existence de cette approximation linéaire qui définit la notion de différentiabilité d'une application (tout comme la dérivabilité d'une fonction d'une variable équivaut à l'existence d'un DL1).

Définition II.19 (Différentiabilité et différentielle de f)**.** *Soit $f : \Omega \to F$, et $a \in \Omega$.*

1. *On dit que f est différentiable en a s'il existe une application linéaire $u \in \mathcal{L}(E, F)$ telle que au voisinage de a,*
$$f(x) = f(a) + u(x - a) + o(\|x - a\|).$$
2. *Dans ce cas, l'application linéaire u est notée $df(a)$, et est appelée différentielle de f au point a, ou application linéaire tangente à f en a.*

Ainsi, la différentielle $df(a)$ est caractérisée par :

1. *sa linéarité ($df(a) \in \mathcal{L}(E, F)$)*
2. *le DL $f(a + h) = f(a) + [df(a)](h) + o(\|h\|)$.*

Pour alléger les notations, on écrira souvent $o(h)$ au lieu de $o(\|h\|)$ (et de même aux ordres supérieurs), et on omettra les crochets et parenthèses en écrivant

$$[df(a)](h) = df(a) \cdot h.$$

Remarque II.20. *Attention à bien comprendre les significations des différentes parenthèses. Si f est différentiable sur tout Ω par exemple,*

$$df : \Omega \to \mathcal{L}(E, F)$$

est une fonction qui à tout point de Ω associe une application linéaire, qui est la partie linéaire de l'approximation affine de f en a. Pour tout a,

$$df(a) : E \to F$$

est donc une application linéaire, qu'on peut ensuite évaluer en $h \in E$. Nous verrons plus loin que, comme dans le cas d'une fonction de deux variables, $df(a) \cdot h$ est la dérivée directionnelle en a selon u.

Cette définition est à rapprocher de la caractérisation de la dérivabilité des fonctions d'une variable réelle par le DL1. D'ailleurs, cette caractérisation permet d'en déduire le rapport entre dérivée et différentielle dans le cas d'une fonction d'un intervalle I de $\mathbb{R}$ dans $\mathbb{R}$.

Proposition II.21 (Différentielle d'une application d'une variable réelle)**.** *Soit Ω un ouvert de $\mathbb{R}$, $f : \Omega \to \mathbb{R}$ et $a \in \mathbb{R}$. Alors f est différentiable en a si et seulement si f est dérivable en a, et dans ce cas :*

$$df(a) : \begin{array}{ccc} \mathbb{R} & \longrightarrow & F \\ h & \longmapsto & f'(a) \cdot h \end{array}$$

Cette proposition permet de voir la plupart des résultats suivants comme généralisation de la situation réelle bien connue.

Méthode II.22 (Montrer que f est différentiable, méthode 1)**.** *Pour montrer que f est différentiable en a et calculer sa différentielle, on peut effectuer un DL1 en a.*

Exemple II.23. *Différentiabilité et différentielle en $(0,2,1)$ de la fonction définie sur $\mathbb{R}^3$ par $f(x,y,z) = \sin(x) - \frac{1}{1+x-y-z}$.*

2. *Différentiabilité en M et différentielle de $M \mapsto M^k$, définie sur $M_n(\mathbb{R})$ à valeurs dans $M_n(\mathbb{R})$.*
3. *Différentiabilité et différentielle en 0 de $\exp : M_n(\mathbb{R}) \to M_n(\mathbb{R})$.*
Solution

1. ***Différentiabilité et différentielle en** $(0,2,1)$ **de** $f(x,y,z) = \sin(x) - \frac{1}{1+x-y-z}$*

(a) ***Différentiabilité :***
La fonction f est définie comme la somme de deux fonctions :

- *$f_1(x,y,z) = \sin(x)$ est de classe C^∞ sur $\mathbb{R}^3$ car c'est la composée de $\sin : \mathbb{R} \to \mathbb{R}$ (de classe C^∞) avec la projection sur la première coordonnée (linéaire donc C^∞).*
- *$f_2(x,y,z) = -\frac{1}{1+x-y-z}$ est une fonction rationnelle. Elle est définie et de classe C^∞ sur l'ouvert $U = \{(x,y,z) \in \mathbb{R}^3 : 1+x-y-z \neq 0\}$.*

En $(0,2,1)$, on a $1+0-2-1 = -2 \neq 0$, donc $(0,2,1) \in U$. Ainsi, f_2 est de classe C^∞ au voisinage de $(0,2,1)$.
La somme de deux fonctions de classe C^1 étant de classe C^1, f est différentiable en $(0,2,1)$ (et même C^∞ au voisinage de ce point).

(b) ***Différentielle :***
Calculons les dérivées partielles en $(0,2,1)$:

$$\begin{aligned}
\frac{\partial f}{\partial x}(x,y,z) &= \cos(x) + \frac{1}{(1+x-y-z)^2} \\
\frac{\partial f}{\partial y}(x,y,z) &= \frac{1}{(1+x-y-z)^2} \\
\frac{\partial f}{\partial z}(x,y,z) &= \frac{1}{(1+x-y-z)^2}
\end{aligned}$$

En $(0,2,1)$: $1+x-y-z = -2$, donc :

$$\begin{aligned}
\frac{\partial f}{\partial x}(0,2,1) &= \cos(0) + \frac{1}{(-2)^2} = 1 + \frac{1}{4} = \frac{5}{4} \\
\frac{\partial f}{\partial y}(0,2,1) &= \frac{1}{4} \\
\frac{\partial f}{\partial z}(0,2,1) &= \frac{1}{4}
\end{aligned}$$

Ainsi, la différentielle de f en $(0,2,1)$ est l'application linéaire :

$$\mathrm{d}f(0,2,1) : \mathbb{R}^3 \to \mathbb{R}, \quad \mathrm{d}f(0,2,1)\cdot(h,k,\ell) = \frac{5}{4}h + \frac{1}{4}k + \frac{1}{4}\ell$$

Sous forme matricielle dans la base canonique :

$$\mathrm{d}f(0,2,1) = \begin{pmatrix} \frac{5}{4} & \frac{1}{4} & \frac{1}{4} \end{pmatrix}$$

2. ***Différentiabilité de*** $M \mapsto M^k$ ***sur*** $M_n(\mathbb{R})$
Considérons $f : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ *définie par* $f(M) = M^k$ *avec* $k \in \mathbb{N}^*$.

(a) ***Différentiabilité en*** M ***:***
Pour $H \in M_n(\mathbb{R})$ *petit, développons* $f(M+H)$ *:*

$$f(M+H) = (M+H)^k = \underbrace{(M+H)(M+H)\cdots(M+H)}_{k \text{ fois}}$$

En développant ce produit, on obtient des termes de différentes puissances en H*. Plus précisément :*

$$(M+H)^k = M^k + \sum_{i=0}^{k-1} M^i H M^{k-1-i} + \text{termes en } H^2, H^3, \ldots, H^k$$

Pour justifier cette formule, considérons l'expression développée. Chaque terme du développement de $(M+H)^k$ *est un produit de* k *facteurs, chacun étant soit* M *soit* H*.*
Les termes linéaires en H *sont ceux où exactement un des* k *facteurs est* H*. Si* H *est en position* $i+1$ *(avec* $0 \leq i \leq k-1$*), alors on a* i *facteurs* M *avant* H *et* $k-1-i$ *facteurs* M *après* H*, ce qui donne* $M^i H M^{k-1-i}$*.*
Il y a k *choix pour la position de* H*, donc la somme de tous ces termes linéaires est :*

$$\sum_{i=0}^{k-1} M^i H M^{k-1-i}$$

(b) ***Expression de la différentielle :***
Posons $L(H) = \sum_{i=0}^{k-1} M^i H M^{k-1-i}$*. Cette application est :*

- ***Linéaire en*** H *: c'est une combinaison linéaire d'applications de la forme* $H \mapsto M^i H M^{k-1-i}$ *qui sont linéaires.*
- ***Continue*** *: car toute application linéaire en dimension finie est continue.*

Montrons que le reste $R(H) = f(M+H) - f(M) - L(H)$ *est un* $o(\|H\|)$*. En effet,* $R(H)$ *est une somme de termes contenant au moins deux facteurs* H*. Chacun de ces termes est de la forme* $AHBHC$ *où* A, B, C *sont des produits de matrices* M *et* H*. Par sous-multiplicativité de la norme,* $\|AHBHC\| \leq \|A\|\|B\|\|C\|\|H\|^2$*.*
Puisque $\|A\|, \|B\|, \|C\|$ *restent bornés quand* $H \to 0$ *(car ils dépendent de* M *qui est fixe), on a* $\|R(H)\| = O(\|H\|^2) = o(\|H\|)$*.*
Ainsi, f *est différentiable en* M *et :*

$$\boxed{\mathrm{d}f(M) \cdot H = \sum_{i=0}^{k-1} M^i H M^{k-1-i}}$$

(c) ***Cas particulier*** $k = 2$ ***:*** $\mathrm{d}f(M) \cdot H = MH + HM$

(d) ***Cas particulier*** $k = 3$ ***:*** $\mathrm{d}f(M) \cdot H = M^2 H + MHM + HM^2$

3. ***Différentiabilité et différentielle en*** 0 ***de*** $\exp : M_n(\mathbb{R}) \to M_n(\mathbb{R})$

(a) ***Différentiabilité :***
Par définition, $\exp(M) = \sum_{p=0}^{+\infty} \frac{M^p}{p!}$*. Pour* H *petit, développons :*

$$\begin{aligned}\exp(H) &= I + H + \frac{H^2}{2!} + \frac{H^3}{3!} + \cdots + \frac{H^p}{p!} + \cdots \\ &= I + H + \sum_{p=2}^{+\infty} \frac{H^p}{p!}\end{aligned}$$

L'application $L(H) = H$ est linéaire. Le reste est :

$$R(H) = \sum_{p=2}^{+\infty} \frac{H^p}{p!}$$

Majorons la norme de ce reste. Soit $\|\cdot\|$ une norme d'algèbre sur $M_n(\mathbb{R})$ (sous-multiplicative) :

$$\|R(H)\| \leq \sum_{p=2}^{+\infty} \frac{\|H\|^p}{p!} = e^{\|H\|} - 1 - \|H\|$$

Or, on sait que $e^x - 1 - x = \frac{x^2}{2} + o(x^2) = o(x)$ quand $x \to 0$. Donc $\|R(H)\| = o(\|H\|)$.
Ainsi, $\exp$ est différentiable en 0 et :

$$\mathrm{d}\exp(0) \cdot H = H$$

Proposition II.24 (Différentiable implique continu)**.** *Soit $f : \Omega \to F$ une application différentiable en a. Alors f est continue en a.*

Éléments de preuve. Par DL! Remarquer que $df(a)$ est continue en 0 puisque linéaire en dimension finie. □

Proposition II.25 (Différentielle et dérivées directionnelles)**.** *Soit $f : \Omega \to F$ une application différentiable en a. Alors f admet en a des dérivées selon toute direction $u \in E$, et*

$$D_u f(a) = df(a) \cdot u.$$

En particulier, si $B = (b_1, \ldots, b_n)$ est une base de E, et si f est identifiée à une fonction des variables $x_1, \ldots, x_n$ coordonnées dans la base B,

$$\forall k \in [1, n], \quad \frac{\partial f}{\partial x_k}(a) = df(a) \cdot b_k.$$

Éléments de preuve. Encore par DL. □

II.2.1 Matrice de la différentielle

Soit $f : \Omega \to F$ une application différentiable en a, et $\mathcal{B}$ et $\mathcal{C}$ deux bases de E et F respectivement. Soit $(f_1, \ldots, f_p)$ les fonctions coordonnées de f, définies sur E à valeurs dans $\mathbb{K}$. Alors

$$\mathrm{Mat}_{\mathcal{B},\mathcal{C}}(\mathrm{d}\, f(a)) = \left(\frac{\partial f_i}{\partial x_j}\right)_{\substack{1\leq i\leq p\\ 1\leq j\leq n}} \in \mathcal{M}_{p,n}(\mathbb{K}).$$

Exemple II.26. *Considérons la fonction $f : \mathbb{R}^2 \to \mathbb{R}^2$ définie par :*

$$f(x, y) = \begin{pmatrix} x^2 + 3y \\ xy + e^x \end{pmatrix}$$

Choix des bases :

- *Pour l'espace de départ $E = \mathbb{R}^2$: prenons la base canonique $\mathcal{B} = (e_1, e_2)$ où $e_1 = (1, 0)$ et $e_2 = (0, 1)$.*
- *Pour l'espace d'arrivée $F = \mathbb{R}^2$: prenons également la base canonique $\mathcal{C} = (e'_1, e'_2)$ où $e'_1 = (1, 0)$ et $e'_2 = (0, 1)$.*

Calcul des dérivées partielles :
Les fonctions coordonnées sont $f_1(x, y) = x^2 + 3y$ et $f_2(x, y) = xy + e^x$.
Calculons les dérivées partielles au point $a = (x_0, y_0)$:

$$\frac{\partial f_1}{\partial x} = 2x, \qquad \frac{\partial f_1}{\partial y} = 3$$
$$\frac{\partial f_2}{\partial x} = y + e^x, \qquad \frac{\partial f_2}{\partial y} = x$$

3. Matrice de la différentielle :
D'après le théorème, la matrice de $\mathrm{d}f(x_0, y_0)$ dans les bases $\mathcal{B}$ et $\mathcal{C}$ est :

$$\mathrm{Mat}_{\mathcal{B},\mathcal{C}}(\mathrm{d}f(x_0, y_0)) = \begin{pmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{pmatrix}_{(x_0,y_0)} = \begin{pmatrix} 2x_0 & 3 \\ y_0 + e^{x_0} & x_0 \end{pmatrix}$$

5. Cas particulier :
Au point $a = (1, 2)$, on a :

$$\mathrm{d}f(1, 2) = \begin{pmatrix} 2 & 3 \\ 2 + e & 1 \end{pmatrix}$$

Pour $h = (3, -1)$, on obtient :

$$\mathrm{d}f(1, 2) \cdot (3, -1) = \begin{pmatrix} 2 & 3 \\ 2 + e & 1 \end{pmatrix} \begin{pmatrix} 3 \\ -1 \end{pmatrix} = \begin{pmatrix} 6 - 3 \\ (2 + e) \cdot 3 - 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 5 + 3e \end{pmatrix}$$

Ce résultat peut être interprété comme l'approximation linéaire au voisinage de $(1, 2)$:

$$f(1 + 3, 2 - 1) \approx f(1, 2) + \mathrm{d}f(1, 2) \cdot (3, -1) = \begin{pmatrix} 7 \\ 2 + e \end{pmatrix} + \begin{pmatrix} 3 \\ 5 + 3e \end{pmatrix} = \begin{pmatrix} 10 \\ 7 + 4e \end{pmatrix}$$

tandis que la valeur exacte est $f(4, 1) = \begin{pmatrix} 16 \\ 4 + e^4 \end{pmatrix}$.

Corollaire II.27 (Détermination de la différentielle par les dérivées partielles)**.** *Si f est différentiable en a, ses dérivées partielles selon une base B déterminent la différentielle. Plus précisément, pour tout $h \in E$ tel que $[h]_B = (h_1, \dots, h_n)$,*

$$df(a) \cdot h = \sum_{k=1}^{n} h_k \partial_k f = \sum_{k=1}^{n} h_k \frac{\partial f}{\partial x_k},$$

les dérivées partielles et les coordonnées étant prises relativement à la base B.

Démonstration. Soit $f : \Omega \subset E \to F$ différentiable en $a \in \Omega$. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E et $\mathcal{C}$ une base de F.
Pour tout $h \in E$, on peut écrire $h = \sum_{k=1}^{n} h_k e_k$ où $(h_1, \dots, h_n) = [h]_{\mathcal{B}}$ sont les coordonnées de h dans la base $\mathcal{B}$.
Par linéarité de la différentielle $\mathrm{d}f(a) : E \to F$ (qui est une application linéaire) :

$$\mathrm{d}f(a) \cdot h = \mathrm{d}f(a) \cdot \left(\sum_{k=1}^{n} h_k e_k \right) = \sum_{k=1}^{n} h_k \, \mathrm{d}f(a) \cdot e_k$$

Or, par définition de la dérivée partielle selon la direction e_k :

$$\mathrm{d}f(a) \cdot e_k = \partial_k f(a) = \frac{\partial f}{\partial x_k}(a)$$

où $\partial_k f(a)$ est la dérivée partielle de f en a selon le k-ième vecteur de la base $\mathcal{B}$.
Ainsi, on obtient bien :

$$\mathrm{d}f(a) \cdot h = \sum_{k=1}^{n} h_k \, \partial_k f(a) = \sum_{k=1}^{n} h_k \frac{\partial f}{\partial x_k}(a)$$

Ceci montre que la différentielle $\mathrm{d}f(a)$ est entièrement déterminée par les dérivées partielles $\partial_k f(a)$ pour $k = 1, \dots, n$. □

Interprétation géométrique
La matrice $\mathrm{Mat}_{\mathcal{B},\mathcal{C}}(\mathrm{d}f(a))$ représente comment l'application linéaire $\mathrm{d}f(a)$ transforme les vecteurs de la base $\mathcal{B}$:

- La première colonne est l'image de e_1 : $\mathrm{d}f(a) \cdot e_1 = \frac{\partial f}{\partial x_1}(a)$
- La deuxième colonne est l'image de e_2 : $\mathrm{d}f(a) \cdot e_2 = \frac{\partial f}{\partial x_2}(a)$
- De manière générale, la k-ième colonne est $\mathrm{d}f(a) \cdot e_k = \frac{\partial f}{\partial x_k}(a)$

Définition II.28 (Jacobienne d'une application de $\Omega \subset \mathbb{R}^n$ dans $\mathbb{R}^p$)**.** *Si $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$, la matrice de $df(a)$ relativement aux bases canoniques est appelée matrice jacobienne de f en a, et notée $Jf(a)$. Ainsi :*

$$Jf(a) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_n} \end{pmatrix} \in \mathcal{M}_{p,n}(\mathbb{K}),$$

où $f = (f_1, \dots, f_p)$.

Méthode II.29 (Montrer que f est différentiable, méthode 2)**.** *Dans le cas où $E = \mathbb{R}^n$, $F = \mathbb{R}^p$, étudier l'existence des dérivées partielles en a, et considérer l'application linéaire v canoniquement associée à la matrice jacobienne. Si*

$$f(a + h) - f(a) - v(h) = o(h),$$

f est différentiable en a de différentielle $df(a) = v$.

- *Adaptation immédiate dans les autres situations après avoir fixé des bases.*

Exemple II.30. *Étudier la différentiabilité en $(0,0)$ de $f:(x,y)\mapsto\sqrt{x^2+y^2}$, prolongée par 0 en 0.*
NB : en exercice, assurez-vous que trouver un DL1 de façon directe n'est pas aisé du tout !

Proposition II.31 (Différentielle d'une application constante)**.** *Soit $f:\Omega\to\mathbb{R}$ une application constante. Alors f est différentiable sur Ω, de différentielle nulle.*
Autrement dit, pour tout $a\in\Omega$, $df(a)$ est l'application linéaire nulle de E dans F.

Proposition II.32 (Différentielle d'une AL)**.** *Soit $f\in\mathcal{L}(E,F)$. Alors f est différentiable en tout point $a\in E$, et pour tout $a\in E$,*

$$df(a)=f.$$

Ainsi, df est constante de valeur f.

Les notations dx, dy, etc, très utilisée par les physiciens notamment, peuvent être définies rigoureusement comme des cas particuliers de la proposition précédente.

Définition II.33 (Notation dx_i)**.** *Étant donné un espace E de dimension n, muni d'une base $B=(b_1,\dots,b_n)$, on note dx_i la différentielle de la projection*

$$\pi_i:\begin{array}{ccc}E&\longrightarrow&\mathbb{R}\\x&\longmapsto&x_i\end{array}$$

où x_i est la coordonnée de x selon b_i.
Ainsi, en tout $a\in E$,

$$dx_i(a)\cdot h=h_i,$$

où h_i est la coordonnée selon b_i de h sur la base B.

Remarque II.34. *L'expression de $df(a)$ en fonction des dérivées partielles se réécrit donc :*

$$df(a)=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}(a)\,dx_i(a),$$

ou encore, sur un domaine de différentiabilité de f :

$$df=\sum_{i=1}^{n}\frac{\partial f}{\partial x_i}\,dx_i,$$

relation prisée des physiciens.
La notation dx_i, bien qu'adhérente au programme et utile pour justifier les notations des physiciens, n'est pas explicitement au programme.

Démonstration de la remarque
Soit E un $\mathbb{R}$-espace vectoriel normé de dimension finie n, soit U un ouvert de E et soit $f:U\to\mathbb{R}$ une application différentiable en un point $a\in U$.
Soit $(e_1,\dots,e_n)$ une base de E. Pour tout $h\in E$, on écrit

$$h=\sum_{i=1}^{n}h_ie_i,$$

où les h_i sont les composantes de h dans cette base.

1. Différentielle comme forme linéaire. La différentielle $df(a)$ est une forme linéaire sur E, donc elle est entièrement déterminée par ses valeurs sur les vecteurs de base. Par définition, la dérivée partielle selon le i-ème vecteur de base est

$$\frac{\partial f}{\partial e_i}(a)=df(a)(e_i).$$

2. Linéarité de la différentielle. Par linéarité de $df(a)$, pour tout $h = \sum_{i=1}^{n} h_i e_i$,

$$df(a)(h) = \sum_{i=1}^{n} h_i \, df(a)(e_i) = \sum_{i=1}^{n} \frac{\partial f}{\partial e_i}(a) \, h_i.$$

3. Définition des formes coordonnées dx_i. On définit les formes linéaires coordonnées dx_i : $E \to \mathbb{R}$ associées à la base $(e_1, \ldots, e_n)$ par

$$dx_i(h) = h_i,$$

où h_i est la i-ème composante de h. Ces formes sont les éléments de la base duale $(dx_1, \ldots, dx_n)$ de E^* associée à la base $(e_1, \ldots, e_n)$.

4. Écriture de $df(a)$ dans la base duale. L'égalité précédente devient, pour tout $h \in E$,

$$df(a)(h) = \sum_{i=1}^{n} \frac{\partial f}{\partial e_i}(a) \, dx_i(h).$$

Ainsi, en tant qu'éléments de E^*, on a l'égalité

$$df(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial e_i}(a) \, dx_i.$$

5. Cas où $E = \mathbb{R}^n$ avec la base canonique. Si $E = \mathbb{R}^n$ et si $(e_1, \ldots, e_n)$ est la base canonique, alors

$$\frac{\partial f}{\partial e_i}(a) = \frac{\partial f}{\partial x_i}(a),$$

et dx_i est la forme linéaire qui à $h = (h_1, \ldots, h_n)$ associe h_i. On retrouve ainsi l'écriture

$$df(a) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) \, dx_i.$$

6. Écriture globale. Si f est différentiable sur un ouvert $U \subset E$, on peut écrire cette relation en tout point de U sous la forme

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial e_i} \, dx_i,$$

où maintenant df et les dx_i sont vues comme des applications de U dans E^* (plus précisément comme des 1-formes différentielles sur U).

II.3 Gradient

On généralise dans cette section la notion de gradient introduite l'année dernière dans le cadre de fonctions de fonctions de deux variables réelles à valeurs dans $\mathbb{R}$. Dans ce contexte, le gradient a été défini en un point a comme le vecteur de $\mathbb{R}^2$ formé des deux dérivées partielles en a, autrement dit des deux dérivées selon les vecteurs d'une base.
On pourrait définir le gradient de la même manière, par rapport à une base, mais c'est plus commode de le définir sans référence à une base particulière. Pour cela, on se place dans le contexte où E est un espace euclidien et $F = \mathbb{R}$, et on utilise le lemme suivant, qu'on a déjà utilisé pour justifier l'existence de l'adjoint d'un endomorphisme.

Lemme II.35 (Théorème de représentation de Riesz)**.** *Soit $(E, \langle\cdot,\cdot\rangle)$ un espace euclidien, et $\phi \in E^*$ une forme linéaire. Alors il existe un unique vecteur $a \in E$ tel que*

$$\forall x \in E, \quad \phi(x) = \langle a, x\rangle.$$

Proposition II.36 (Définition du gradient d'une application numérique)**.** *Soit $(E, \langle\cdot,\cdot\rangle)$ un espace euclidien, et Ω un ouvert de E. Soit $f : \Omega \to \mathbb{R}$, et $a \in \Omega$ tel que f est différentiable en a. Alors $df(a) \in E^*$. On définit alors le gradient de f en a comme étant l'unique vecteur $\nabla f(a) \in E$ tel que*

$$\forall h \in E, \quad df(a) \cdot h = \langle \nabla f(a), h\rangle.$$

Théorème II.37 (Expression du gradient en b.o.n.)**.** *Soit $(E, \langle\cdot,\cdot\rangle)$ un espace euclidien, et Ω un ouvert de E. Soit $f : \Omega \to \mathbb{R}$, et $a \in \Omega$ tel que f est différentiable en a. Soit $B = (b_1, \dots, b_n)$ une base orthonormale de E. Alors*

$$\nabla f(a) = \sum_{k=1}^{n} \partial_k f(a) b_k,$$

les dérivées partielles étant prises dans la base B. Ainsi :

$$[\nabla f(a)]_B = \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_n f(a) \end{pmatrix}.$$

Corollaire II.38 (Gradient d'une fonction de n variables réelles)**.** *On munit $\mathbb{R}^n$ (identifié à $M_{n,1}(\mathbb{R})$) de sa structure euclidienne canonique. Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ une application différentiable en $a \in \Omega$. Alors*

$$\nabla f(a) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(a) \\ \vdots \\ \frac{\partial f}{\partial x_n}(a) \end{pmatrix}.$$

Remarque II.39. *Si $\Omega \subset \mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ est différentiable en a, alors la jacobienne en a est*

$$J_f(a) = \mathrm{Mat}_{bc}(\mathrm{d}f(a)) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)\right) \in \mathcal{M}_{1,n}(\mathbb{R}).$$

Ainsi, $J_f(a) = \nabla f(a)^T$.

Proposition II.40 (Première interprétation géométrique du gradient)**.** *Soit $\Omega \subset \mathbb{R}^n$, et $f : \Omega \to \mathbb{R}$ différentiable en $a \in \Omega$. Le gradient $\nabla f(a)$ est positivement colinéaire au vecteur unitaire u tel que $D_u f(a)$ soit maximal :*

$$\max_{\|u\|=1} D_u f(a) \iff u = \frac{\nabla f(a)}{\|\nabla f(a)\|}.$$

Éléments de preuve. C'est l'inégalité de Cauchy-Schwarz et son cas d'égalité, le choix du signe étant dicté par le fait que $D_{-u} f(a) = -D_u f(a)$, par linéarité de $df(a)$. Ainsi, en tout point de différentiabilité, le gradient indique la direction de plus forte pente. □

Démonstration

Soit $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ différentiable en $a \in \Omega$. On note $\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)\right)^\top$ son gradient.

1. Expression de la dérivée directionnelle.

Pour tout vecteur unitaire $u \in \mathbb{R}^n$ (c'est-à-dire $\|u\| = 1$), on a

$$D_u f(a) = df(a)(u) = \nabla f(a) \cdot u,$$

où $\cdot$ désigne le produit scalaire euclidien.

2. Maximisation sous contrainte.

On cherche à maximiser $D_u f(a) = \nabla f(a) \cdot u$ parmi les vecteurs u de norme 1.

3. Application de l'inégalité de Cauchy–Schwarz.

Par l'inégalité de Cauchy–Schwarz,

$$|\nabla f(a) \cdot u| \leq \|\nabla f(a)\| \, \|u\| = \|\nabla f(a)\|.$$

Donc

$$D_u f(a) \leq \|\nabla f(a)\|.$$

L'égalité dans Cauchy–Schwarz se produit exactement lorsque u est colinéaire à $\nabla f(a)$ (et de même sens, car on veut maximiser et non la valeur absolue).

4. Condition d'égalité.

L'égalité $\nabla f(a) \cdot u = \|\nabla f(a)\|$ a lieu si et seulement si :

- u est colinéaire à $\nabla f(a)$: $u = \lambda \nabla f(a)$ pour un certain $\lambda \in \mathbb{R}$;
- $\|u\| = 1$.

De $u = \lambda \nabla f(a)$ et $\|u\| = 1$, on tire

$$|\lambda| \|\nabla f(a)\| = 1 \quad \text{donc} \quad \lambda = \frac{1}{\|\nabla f(a)\|} \quad \text{(en prenant } \lambda > 0 \text{ pour maximiser)}.$$

Ainsi

$$u = \frac{\nabla f(a)}{\|\nabla f(a)\|}.$$

5. Conclusion.

Le maximum de $D_u f(a)$ pour $\|u\| = 1$ est atteint lorsque u est le vecteur unitaire dans la direction de $\nabla f(a)$. De plus, ce maximum vaut précisément $\|\nabla f(a)\|$, et on a

$$\max_{\|u\|=1} D_u f(a) = \|\nabla f(a)\|,$$

atteint uniquement pour

$$u = \frac{\nabla f(a)}{\|\nabla f(a)\|} \quad \text{(lorsque } \nabla f(a) \neq 0).$$

Calcul du gradient du déterminant en l'identité
On considère l'application

$$f : \mathcal{M}_n(\mathbb{R}) \to \mathbb{R}, \quad f(A) = \det A.$$

On cherche le gradient de f en $A = I_n$ pour le produit scalaire usuel

$$\langle X, Y \rangle = \mathrm{tr}(X^\top Y).$$

1. Différentielle en I_n
Soit $H \in \mathcal{M}_n(\mathbb{R})$ de petite norme. On trigonalise H (dans $\mathbb{R}$ ou $\mathbb{C}$ selon les valeurs propres). Il existe P inversible telle que $P^{-1}HP = T$ avec T triangulaire supérieure de diagonale $(\lambda_1, \ldots, \lambda_n)$, où λ_i sont les valeurs propres de H.
Alors :

$$\det(I_n + H) = \det(P^{-1}(I_n + H)P) = \det(I_n + T) = \prod_{i=1}^{n}(1 + \lambda_i).$$

En développant à l'ordre 1 :

$$\prod_{i=1}^{n}(1 + \lambda_i) = 1 + \sum_{i=1}^{n} \lambda_i + o(\|H\|) = 1 + \mathrm{tr}(H) + o(\|H\|),$$

car $\sum \lambda_i = \mathrm{tr}(H)$ et les termes d'ordre supérieur (comme $\lambda_i\lambda_j$) sont en $O(\|H\|^2)$.
On choisit une norme sous-multiplicative et subordonnée à la norme 1 pour contrôler $|\lambda_i| \leq \|H\|$.
Alors $|\lambda_i\lambda_j| \leq \|H\|^2$, donc le reste est bien en $o(\|H\|)$.
On obtient :

$$df(I_n)(H) = \mathrm{tr}(H).$$

2. Gradient dans la base canonique orthonormée
Le produit scalaire usuel sur $\mathcal{M}_n(\mathbb{R})$ s'écrit, dans la base canonique $(E_{ij})_{1\leq i,j\leq n}$ (où E_{ij} est la matrice avec un 1 en position (i, j) et 0 ailleurs) :

$$\langle E_{ij}, E_{kl} \rangle = \delta_{ik}\delta_{jl}.$$

Cette base est donc orthonormée.
La différentielle $df(I_n) : H \mapsto \mathrm{tr}(H)$ est une forme linéaire. Son gradient $\nabla f(I_n)$ est l'unique matrice G telle que

$$\langle G, H \rangle = df(I_n)(H) \quad \forall H \in \mathcal{M}_n(\mathbb{R}).$$

Or $\mathrm{tr}(H) = \sum_{i=1}^{n} H_{ii} = \langle I_n, H \rangle$, car

$$\langle I_n, H \rangle = \mathrm{tr}(I_n^\top H) = \mathrm{tr}(H).$$

Ainsi :

$$\boxed{\nabla f(I_n) = I_n}.$$

Problème de l'ordre de la base

Si on avait choisi la base canonique dans un ordre différent (par exemple ordonnée par colonnes, ou par lignes), la matrice représentant la forme linéaire $df(I_n)$ changerait, mais le gradient $\nabla f(I_n)$ en tant qu'élément de $\mathcal{M}_n(\mathbb{R})$ reste le même : I_n. Cependant ses coordonnées vont évidemment changer avec le changement de la base (Ordre des éléments de la base).
En effet, le gradient est défini indépendamment de la base : c'est le vecteur (ici matrice) qui représente la forme linéaire $df(I_n)$ via le produit scalaire. L'expression des coordonnées de $\nabla f(I_n)$ dans une base orthonormée (e_k) est

$$(\nabla f(I_n))_k = df(I_n)(e_k).$$

Dans la base canonique (E_{ij}), on trouve :

$$df(I_n)(E_{ij}) = \mathrm{tr}(E_{ij}) = \begin{cases} 1 & \text{si } i = j, \\ 0 & \text{sinon,} \end{cases}$$

ce qui donne bien les coordonnées de I_n.
Ainsi, le résultat $\nabla f(I_n) = I_n$ est intrinsèque et ne dépend pas de l'ordre choisi pour énumérer la base, pourvu que celle-ci soit orthonormée.

II.4 Opérations sur les applications différentiables

On se place à nouveau dans la situation générale : E est un e.v.n. de dimension finie sur $\mathbb{R}$, F est un e.v.n. de dimension finie de K et Ω est un ouvert de E.

Proposition II.41 (Linéarité). *Soit f et $g : \Omega \to F$ différentiables en a, et $\lambda, \mu \in K$. Alors $f + \lambda g$ est différentiable en a et*

$$d(f + \lambda g)(a) = d(f)(a) + \lambda dg(a).$$

Éléments de preuve. Revenir à la définition par DL. □

Proposition II.42 (Différentielle d'une forme multilinéaire). *Soit Ω un ouvert de E, $F_1, \dots, F_n$ des K-e.v.n. de dimension finie, et $f_i : \Omega \to F_i$ différentiables en $a \in \Omega$. Soit $M : F_1 \times \cdots \times F_n \to F$ une application multilinéaire. Alors*

$$M(f_1, \dots, f_n) : x \mapsto M(f_1(x), \dots, f_n(x))$$

est différentiable en a, et

$$\forall h \in E, \quad dM(f_1, \dots, f_n)(a) \cdot h = \sum_{k=1}^{n} M(f_1(a), \dots, df_k(a) \cdot h, \dots, f_n(a)).$$

Éléments de preuve. Par DL et majoration de la forme multilinéaire (par continuité). □

Exemple II.43. *Soit $f, g : \Omega \to K$, $\Omega \subset E$. Exprimer la différentielle de fg.*

2. *Si $f : \Omega \to F$ avec F euclidien, et si f est différentiable en a, montrer que $\|f\|^2$ est différentiable en tout point et exprimer sa différentielle.*
3. *En particulier, si E est un espace euclidien, exprimer la différentielle en tout point de $X \mapsto X^2$.*

Solution
Dérivée d'un produit de matrices. Le produit matriciel $M(A, B) = AB$ est une application bilinéaire de $\mathcal{M}_{n,p}(\mathbb{C}) \times \mathcal{M}_{p,q}(\mathbb{C})$ dans $\mathcal{M}_{n,q}(\mathbb{C})$. Par le corollaire sur la dérivée d'une forme multilinéaire, si A et B sont dérivables,

$$\frac{d}{dt}\big(A(t)B(t)\big) = A'(t)B(t) + A(t)B'(t).$$

Exemple :

$$A(t) = \begin{pmatrix} 2 & 4 & 1+t \\ 3 & 2+2t & 2 \\ 1 & t & t \end{pmatrix}, \quad A'(t) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 1 & 1 \end{pmatrix},$$

$$B(t) = \begin{pmatrix} 2t & 4 & -t \\ 0 & 1 & 1-t \\ 2 & 1 & t \end{pmatrix}, \quad B'(t) = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 1 \end{pmatrix},$$

$$\frac{d}{dt}(A(t)B(t)) = \begin{pmatrix}0&0&1\\0&2&0\\0&1&1\end{pmatrix}\begin{pmatrix}2t&4&-t\\0&1&1-t\\2&1&t\end{pmatrix} + \begin{pmatrix}2&4&1+t\\3&2+2t&2\\1&t&t\end{pmatrix}\begin{pmatrix}2&0&-1\\0&0&-1\\0&0&1\end{pmatrix}.$$

Dérivée de la norme. Soit $N(t) = \|X(t)\| = \sqrt{\langle X(t), X(t)\rangle}$. La fonction $t \mapsto \langle X(t), X(t)\rangle$ est dérivable comme composée de X et de la forme bilinéaire symétrique $\langle\cdot,\cdot\rangle$ (cas particulier du corollaire avec $n = 2$ et $f_1 = f_2 = X$) :

$$\frac{d}{dt}\langle X(t), X(t)\rangle = \langle X'(t), X(t)\rangle + \langle X(t), X'(t)\rangle = 2\langle X'(t), X(t)\rangle.$$

Comme $X(t) \neq 0$, $t \mapsto \|X(t)\|$ est dérivable sur I (car $x \mapsto \sqrt{x}$ est dérivable sur $\mathbb{R}_+^*$) et

$$\frac{d}{dt}\|X(t)\| = \frac{1}{2\|X(t)\|}\cdot 2\langle X'(t), X(t)\rangle = \frac{\langle X(t), X'(t)\rangle}{\|X(t)\|}.$$

Dérivée du vecteur unitaire. Écrivons $U(t) = \frac{X(t)}{\|X(t)\|} = \varphi(X(t))$ où $\varphi(y) = \frac{y}{\|y\|}$. C'est aussi $U(t) = X(t)\cdot(\|X(t)\|)^{-1}$. En utilisant la dérivée d'un produit (cas bilinéaire, avec la multiplication scalaire-vecteur) et le résultat précédent :

$$U'(t) = X'(t)\cdot\frac{1}{\|X(t)\|} + X(t)\cdot\frac{d}{dt}\left(\frac{1}{\|X(t)\|}\right).$$

Or $\frac{d}{dt}\left(\frac{1}{\|X(t)\|}\right) = -\frac{\|X(t)\|'}{\|X(t)\|^2} = -\frac{\langle X(t),X'(t)\rangle}{\|X(t)\|^3}$. Donc

$$U'(t) = \frac{X'(t)}{\|X(t)\|} - X(t)\frac{\langle X(t), X'(t)\rangle}{\|X(t)\|^3} = \frac{X'(t)\|X(t)\|^2 - X(t)\langle X(t), X'(t)\rangle}{\|X(t)\|^3}.$$

Dérivée de $X(t)^T AY(t)$. L'application $(X, Y) \mapsto X^T AY$ est bilinéaire (à valeurs scalaires). Par le corollaire,

$$\frac{d}{dt}\big(X(t)^T AY(t)\big) = X'(t)^T AY(t) + X(t)^T AY'(t).$$

Dérivée d'une puissance de matrice. Montrons par récurrence que $\frac{d}{dt}X(t)^n = \sum_{k=0}^{n-1} X(t)^k X'(t) X(t)^{n-1-k}$. Pour $n = 1$: $\frac{d}{dt}X(t) = X'(t)$, formule vérifiée. Supposons la propriété vraie au rang n. Alors

$$\frac{d}{dt}X(t)^{n+1} = \frac{d}{dt}\big(X(t)^n X(t)\big).$$

Le produit matriciel est bilinéaire, donc

$$\frac{d}{dt}\big(X(t)^n X(t)\big) = \left(\frac{d}{dt}X(t)^n\right)X(t) + X(t)^n X'(t).$$

Par hypothèse de récurrence,

$$\frac{d}{dt}X(t)^{n+1} = \left(\sum_{k=0}^{n-1} X^k X' X^{n-1-k}\right)X + X^n X' = \sum_{k=0}^{n-1} X^k X' X^{n-k} + X^n X' = \sum_{k=0}^{n} X^k X' X^{n-k},$$

ce qui est la formule au rang $n+1$.

Dérivée du déterminant. Le déterminant $\det_B$ est une forme n-linéaire alternée. Par le corollaire (cas particulier 3), en dérivant colonne par colonne :

$$\frac{d}{dt}\det(C_1(t), C_2(t), C_3(t)) = \det(C_1'(t), C_2(t), C_3(t)) + \det(C_1(t), C_2'(t), C_3(t)) + \det(C_1(t), C_2(t), C_3'(t)).$$

Ici $C_1(t) = \begin{pmatrix} t-2 \\ 0 \\ 2 \end{pmatrix}$, $C_2(t) = \begin{pmatrix} 3 \\ t+1 \\ 5 \end{pmatrix}$, $C_3(t) = \begin{pmatrix} 6 \\ 2 \\ t-3 \end{pmatrix}$. Leurs dérivées sont $C_1'(t) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $C_2'(t) = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $C_3'(t) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Donc

$$\Delta'(t) = \begin{vmatrix} 1 & 3 & 6 \\ 0 & t+1 & 2 \\ 0 & 5 & t-3 \end{vmatrix} + \begin{vmatrix} t-2 & 0 & 6 \\ 0 & 1 & 2 \\ 2 & 0 & t-3 \end{vmatrix} + \begin{vmatrix} t-2 & 3 & 0 \\ 0 & t+1 & 0 \\ 2 & 5 & 1 \end{vmatrix}.$$

En développant :

$$\begin{aligned} \Delta'(t) &= 1 \cdot \big((t+1)(t-3) - 10\big) + 1 \cdot \big((t-2)(t-3) - 12\big) + 1 \cdot \big((t-2)(t+1) - 0\big) \\ &= (t^2 - 2t - 13) + (t^2 - 5t - 6) + (t^2 - t - 2) \\ &= 3t^2 - 8t - 21. \end{aligned}$$

Proposition II.44 (Différentielle d'une composée)**.** *Soit E et F deux e.v.n. de dimension finie sur $\mathbb{R}$, et G un e.v.n. de dimension finie sur K. Soit Ω un ouvert de E et Ω' un ouvert de f, et $f : \Omega \to F$ et $g : \Omega' \to G$ deux applications. Soit $a \in \Omega$. On suppose que*

1. *f est différentiable en a*
2. *$f(a) \in \Omega'$ et g est différentiable en $f(a)$.*

Alors $g \circ f$ est différentiable en a, et

$$d(g \circ f)(a) = dg(f(a)) \circ df(a).$$

Éléments de preuve. Composer les DL. □

Corollaire II.45 (Composition par une fonction réelle)**.** *Soit Ω un ouvert d'un e.v.n. E, et I un intervalle ouvert de $\mathbb{R}$. Soit $f : \Omega \to \mathbb{R}$ et $g : I \to \mathbb{R}$. Si f est différentiable en $a \in \Omega$, si $f(a) \in I$ et si g dérivable en $f(a)$, alors $g \circ f$ est différentiable en a et*

$$d(g \circ f)(a) = g'(f(a))df(a).$$

Exemple II.46. *Différentiabilité de $x \mapsto \|x\|$ en tout $x \neq 0$ dans un espace euclidien.*

2. *Différentiabilité de $\|f\|$ lorsque f l'est.*

Corollaire II.47 (Dérivée le long d'un arc)**.** *Soit I un intervalle ouvert de $\mathbb{R}$, Ω un ouvert de E. Soit $\gamma : I \to \Omega$ et $f : \Omega \to F$. Si γ est dérivable en $a \in I$, et f différentiable en $\gamma(a)$, alors $f \circ \gamma$ est dérivable en a et*

$$(f \circ \gamma)'(a) = df(\gamma(a)) \cdot \gamma'(a) = \langle \nabla f(\gamma(a)), \gamma'(a) \rangle.$$

Exemple II.48. *Dérivée de $t \mapsto f(a + th)$. Cohérence en 0.*

Corollaire II.49 (Règle de la chaîne)**.** *Avec les notations et hypothèses de la proposition 16.2.51, si $B = (b_1, \ldots, b_n)$ est une base de E, et $C = (c_1, \ldots, c_p)$ une base de F,*

$$\forall k \in [1, n], \quad \partial_{B,k}(g \circ f)(a) = \sum_{i=1}^{p} \partial_{c,i} g(f(a)) \partial_{B,k} f(a).$$

Remarque II.50 (Avertissement)**.** *Attention à ne pas confondre $\partial_i g(f(a))$ et $\partial_i (g \circ f)(a)$. Quelle est la différence entre les deux ?*

Corollaire II.51 (Réexpression de la règle de la chaîne dans $\mathbb{R}^n$)**.** *Si $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$, et si les bases considérées sont les bases canoniques, on obtient, avec les mêmes hypothèses :*

$$\forall k \in [1, n], \quad \frac{\partial(g \circ f)}{\partial x_k}(a) = \sum_{i=1}^{p} \frac{\partial g}{\partial y_i}(f(a)) \frac{\partial f}{\partial x_k}(a),$$

où $(x_1, \ldots, x_n)$ représentent les coordonnées génériques dans $E = \mathbb{R}^n$ et $(y_1, \ldots, y_p)$ les coordonnées génériques dans $F = \mathbb{R}^p$.
De façon équivalente, en exprimant coordonnée par coordonnée, $J_{g\circ f}(a) = J_g(f(a)) \times J_f(a)$.

Exemple II.52. *Soit f une application de $\mathbb{R}^3$ dans $\mathbb{R}$ admettant des dérivées partielles en tout (x, y, z). Exprimer les dérivées partielles de*

$$g(u, v, w) = f(u^2 + v, 2v, 3w - e^v).$$

Solution On applique le corollaire sur la règle de la chaîne dans $\mathbb{R}^n$.
Soit $\Phi : \mathbb{R}^3 \to \mathbb{R}^3$ définie par

$$\Phi(u, v, w) = (f_1(u, v, w), f_2(u, v, w), f_3(u, v, w))$$

avec

$$f_1(u, v, w) = u^2 + v, \qquad f_2(u, v, w) = 2v, \qquad f_3(u, v, w) = 3w - e^v.$$

Alors $g = f \circ \Phi$.
Puisque f admet des dérivées partielles en tout point et que Φ est différentiable (ses composantes sont polynomiales ou exponentielle), le corollaire donne pour tout $k \in \{1, 2, 3\}$:

$$\frac{\partial g}{\partial x_k}(a) = \sum_{i=1}^{3} \frac{\partial f}{\partial y_i}(\Phi(a)) \frac{\partial \Phi_i}{\partial x_k}(a),$$

où $(x_1, x_2, x_3) = (u, v, w)$ sont les coordonnées dans $\mathbb{R}^3$ (variable de dérivation) et $(y_1, y_2, y_3) = (x, y, z)$ sont les coordonnées dans l'espace de départ f.

Calcul des dérivées partielles de Φ :

$$\frac{\partial \Phi_1}{\partial u} = 2u, \quad \frac{\partial \Phi_1}{\partial v} = 1, \quad \frac{\partial \Phi_1}{\partial w} = 0,$$
$$\frac{\partial \Phi_2}{\partial u} = 0, \quad \frac{\partial \Phi_2}{\partial v} = 2, \quad \frac{\partial \Phi_2}{\partial w} = 0,$$
$$\frac{\partial \Phi_3}{\partial u} = 0, \quad \frac{\partial \Phi_3}{\partial v} = -e^v, \quad \frac{\partial \Phi_3}{\partial w} = 3.$$

Pour $k = 1$ (dérivée partielle par rapport à u) :

$$\frac{\partial g}{\partial u} = \frac{\partial f}{\partial x}(\Phi) \cdot \frac{\partial \Phi_1}{\partial u} + \frac{\partial f}{\partial y}(\Phi) \cdot \frac{\partial \Phi_2}{\partial u} + \frac{\partial f}{\partial z}(\Phi) \cdot \frac{\partial \Phi_3}{\partial u}$$

$$= \frac{\partial f}{\partial x}(u^2 + v, 2v, 3w - e^v) \cdot (2u) + \frac{\partial f}{\partial y}(\cdots) \cdot 0 + \frac{\partial f}{\partial z}(\cdots) \cdot 0,$$

d'où

$$\boxed{\frac{\partial g}{\partial u}(u, v, w) = 2u \, \frac{\partial f}{\partial x}(u^2 + v, \ 2v, \ 3w - e^v)}.$$

Pour $k = 2$ (dérivée partielle par rapport à v) :

$$\frac{\partial g}{\partial v} = \frac{\partial f}{\partial x}(\Phi) \cdot \frac{\partial \Phi_1}{\partial v} + \frac{\partial f}{\partial y}(\Phi) \cdot \frac{\partial \Phi_2}{\partial v} + \frac{\partial f}{\partial z}(\Phi) \cdot \frac{\partial \Phi_3}{\partial v}$$

$$= \frac{\partial f}{\partial x}(\cdots)\cdot 1 + \frac{\partial f}{\partial y}(\cdots)\cdot 2 + \frac{\partial f}{\partial z}(\cdots)\cdot(-e^v),$$

d'où

$$\boxed{\frac{\partial g}{\partial v}(u,v,w) = \frac{\partial f}{\partial x}(\cdots) + 2\,\frac{\partial f}{\partial y}(\cdots) - e^v\,\frac{\partial f}{\partial z}(\cdots)}.$$

Pour $k=3$ (dérivée partielle par rapport à w) :

$$\frac{\partial g}{\partial w} = \frac{\partial f}{\partial x}(\Phi)\cdot\frac{\partial \Phi_1}{\partial w} + \frac{\partial f}{\partial y}(\Phi)\cdot\frac{\partial \Phi_2}{\partial w} + \frac{\partial f}{\partial z}(\Phi)\cdot\frac{\partial \Phi_3}{\partial w}$$

$$= \frac{\partial f}{\partial x}(\cdots)\cdot 0 + \frac{\partial f}{\partial y}(\cdots)\cdot 0 + \frac{\partial f}{\partial z}(\cdots)\cdot 3,$$

d'où

$$\boxed{\frac{\partial g}{\partial w}(u,v,w) = 3\,\frac{\partial f}{\partial z}(u^2+v,\ 2v,\ 3w-e^v)}.$$

Proposition II.53 (Différentielle de $(f_1,\dots,f_p)$)**.** *Soit Ω un ouvert de E, et $F_1,\dots,F_q$ des K-e.v.n. de dimension finie, et soit $f_i:\Omega\to F_i$, $i\in[1,q]$. Les propositions suivantes sont équivalentes:*

1. *Pour tout $i\in[\![1,q]\!]$, f_i est différentiable en $a\in\Omega$;*
2. *$f=(f_1,\dots,f_q):E\to F_1\times\cdots\times F_q$ est différentiable en a.*

Si c'est le cas,

$$df(a) = (df_1(a),\dots,df_q(a)) \in \mathcal{L}(E, F_1\times\cdots\times F_q),$$

c'est-à-dire :

$$\forall h\in E,\quad df(a) = (df_1(a)\cdot h,\dots,df_q(a)\cdot h).$$

Remarque II.54. *Retrouver de la sorte l'expression de la jacobienne.*

III. Applications de classe C^n

III.1 Applications de classe C^1

Définition III.1 (Application de classe C^1)**.** *Soit $f:\Omega\to F$. On dit que f est de classe C^1 sur Ω si :*

1. *f est différentiable en tout point de Ω*
2. *$df:\Omega\to\mathcal{L}(E,F)$ est continue.*

Exemple III.2. 1. *Soit $f:E\to F$ une application linéaire, alors f est de classe C^1.*
2. *Soit f une application constante, alors f est de classe C^1.*

Proposition III.3 (Opérations sur les applications de classe C^1)**.** *Soit E,E' des $\mathbb{R}$-e.v.n. de dimension finie, et $F,F_1,\dots,F_q$ des K-e.v.n. de dimension finie. Soit Ω,Ω' des ouverts de E et E' respectivement.*

1. *Soit $f,g:\Omega\to F$ de classe C^1 et $\lambda\in K$. Alors $f+\lambda g$ est de classe C^1.*
2. *Soit pour $i\in[\![1,q]\!]$, $f_i:\Omega\to F_i$, et $M:F_1\times\cdots\times F_q\to F$ une forme multilinéaire. Si les f_i sont de classe C^1, alors $M(f_1,\dots,f_q)$ est de classe C^1.*
3. *Soit $f:\Omega'\to\Omega$ et $g:\Omega\to F$. Si f et g sont de classe C^1 sur Ω' et Ω respectivement, alors $g\circ f$ est de classe C^1 sur Ω.*
4. *Si $f:\Omega\to K$ ne s'annule pas et est de classe C^1, alors $\frac{1}{f}$ aussi.*

5. *Soit pour $i \in [1,q]$, $f_i : \Omega \to F_i$, et $f : \Omega \to F_1 \times \cdots \times F_q$ définie par*

$$\forall x \in \Omega, \quad f(x) = (f_1(x), \ldots, f_q(x)).$$

Alors f est de classe C^1 si et seulement si tout f_i est de classe C^1.

Éléments de preuve. Attention à bien comprendre sur quelle variable porte la continuité (le point $x \in \Omega$ en lequel on considère la différentielle). Cela conditionne la validité de l'argument donné. Par exemple, pour la composée, la différentielle $d(g \circ f)$ s'exprime comme composée de deux différentielles, mais la composée porte sur la variable h, et non sur a qui est à voir plutôt comme un paramètre. Donc on ne peut pas obtenir la continuité de cette différentielle par propriété de continuité d'une composée.
En revanche, puisque la composition se fait entre applications linéaires d'espaces de dimension finie, l'application $(f,g) \mapsto g \circ f$ est une application bilinéaire entre espaces de dimension finie, donc elle-même continue. C'est ce point qui est en jeu ici.
Pour 2, on peut remarquer que si $M : E^n \to F$ est une application multilinéaire, alors $\tilde{M} : E^{k-1} \times \mathcal{L}(E,F) \times E^{n-k} \to \mathcal{L}(E,F)$ définie par

$$\tilde{M}(x_1, \ldots, x_{k-1}, \alpha, x_{k+1}, \ldots, x_n)(x) = M(x_1, \ldots, x_{k-1}, \alpha(x), x_{k+1}, \ldots, x_n)$$

est également multilinéaire, donc continue, tous les espaces en jeu étant de dimension finie. □

Vous avez défini l'année dernière le caractère C^1 des fonctions de 2 variables à l'aide des dérivées partielles. On retrouve ce point de vue via la caractérisation 16.3.5. On démontre d'abord le résultat suivant, qui n'apparaît qu'implicitement au programme (cela peut être vu comme un sous-produit de la caractérisation 16.3.5 et de la définition de la différentielle).

Théorème III.4 (Formule de Taylor-Young à l'ordre 1 pour les dérivées partielles). *Soit $f : \Omega \to K$ une application à valeurs dans $\mathbb{R}$ ou $\mathbb{C}$. Soit B une base de E, et $a \in \Omega$. On suppose que pour tout f, $\partial_k f$ est définie et continue au voisinage de a. Alors*

$$f(a+h) \underset{h \to 0}{=} f(a) + \sum_{k=0}^{n} h_k \partial_k f(a) + o(h).$$

où $(h_1, \ldots, h_n) = [h]_B$.

Éléments de preuve. Utiliser la formule de Taylor pour les fonctions d'une variable, successivement pour chacune des n variables. □

Théorème III.5 (Caractérisation de la classe C^1 par les dérivées partielles). *Soit $f : \Omega \to F$ et $B = (b_1, \ldots, b_n)$ une base de E. Les propositions suivantes sont équivalentes:*

1. *f est de classe C^1 sur Ω;*
2. *f admet des dérivées partielles en tout point selon la base B, et pour tout $k \in [1,n]$, $\partial_k f : \Omega \to F$ est continue.*

Éléments de preuve. Démonstration non exigible.

- Se ramener à $f : \Omega \to K$, en considérant les coordonnées dans une base C de F.
- Utiliser la décomposition $df_i(a) = \sum_{k=1}^{n} \partial_k f_i$.

Cela revient en gros à utiliser le fait que df est continue si et seulement si $\mathrm{Mat}_{B,C}(df)$ est continue, ce qui est la matrice des dérivées partielles des coordonnées dans C (l'analogue de la jacobienne). Dans le sens (ii) $\Rightarrow$ (i), ne pas oublier de montrer d'abord la différentiabilité de f, en utilisant la formule de Taylor-Young. □

Remarque III.6. *La formule de Taylor 16.3.4 se réécrit alors simplement*

$$f(a+h) = f(a) + \langle \nabla f(a), h\rangle + o(h),$$

qui n'est rien de plus que la combinaison de la définition de la différentielle et du gradient ∇f.

Corollaire III.7 (Continuité via les dérivées partielles)**.** *Si $E = \mathbb{R}^n$ et si f admet des dérivées partielles par rapport à tout x_i en a, et que ces dérivées partielles sont continues au voisinage de a, alors f est continue au voisinage de a.*

Remarque III.8 (Avertissement)**.** *Comme on l'a déjà vu, l'existence seule des dérivées partielles (et même plus généralement des dérivées directionnelles) ne suffit pas à établir la continuité.*

Théorème III.9 (Intégrale de la différentielle le long d'un arc)**.** *Soit $f : \Omega \to F$ une application de classe C^1, et $\gamma : [0,1] \to \Omega$ un chemin de classe C^1 tel que $\gamma(0) = a$ et $\gamma(1) = b$. Alors*

$$f(b) - f(a) = \int_0^1 df(\gamma(t)) \cdot \gamma'(t)\, dt.$$

Éléments de preuve. L'intégrande est la dérivée de $f \circ \gamma : [0,1] \to F$. □

Remarque III.10. *Ce théorème peut être vu comme une généralisation du théorème fondamentale de l'analyse, puisqu'en prenant $E = \mathbb{R}$ et $\gamma : t \mapsto a(1-t) + bt$, puis en faisant un changement de variable affine, on retrouve*

$$\int_a^b f(t)\, dt = f(b) - f(a)$$

lorsque $f : [a,b] \to F$ est de classe C^1.

Exemple III.11. *Si f est de classe C^1 sur l'ouvert Ω et si $[a,b] \subset \Omega$, alors*

$$f(b) - f(a) = \int_0^1 df(a(1-t) + bt) \cdot (b-a)\, dt.$$

2. *De façon équivalente, si f est de classe C^1 sur l'ouvert Ω et si $[a, a+u] \subset \Omega$, alors*

$$f(a+u) - f(a) = \int_0^1 df(a+tu) \cdot u\, dt.$$

3. *Si f est de classe C^1 sur l'ouvert Ω et si $\gamma : [0,1] \to \Omega$ est un lacet de classe C^1 (i.e. $\gamma(a) = \gamma(b)$), alors*

$$\int_0^1 df(\gamma(t)) \cdot \gamma'(t)\, dt = 0.$$

Théorème III.12 (Caractérisation des fonctions constantes (1))**.** *Soit Ω un ouvert convexe de E, et $f : \Omega \to F$. Les propositions suivantes sont équivalentes:*

1. *f est constante sur Ω*
2. *f est différentiable et df est nulle sur Ω.*

Éléments de preuve.

- Le sens direct est déjà vu.
- Pour le sens réciproque, utiliser le théorème 16.3.9 sur une paramétrisation du segment $[a,b]$ pour montrer que pour tout $a, b \in \Omega$, $f(a) = f(b)$.

□

Ce théorème admet une généralisation qui est aussi au programme, mais sans sa démonstration (que nous esquissons quand même).

Théorème III.13 (Caractérisation des fonctions constantes (2))**.** *Soit Ω un ouvert connexe par arcs de E, et $f : \Omega \to F$. Les propositions suivantes sont équivalentes:*

1. *f est constante sur Ω*
2. *f est différentiable et df est nulle sur Ω.*

Éléments de preuve. Démonstration hors-programme.
C'est toujours la réciproque qui reste à voir.
On pourrait essayer d'adapter la preuve précédente, en montrant que si Ω est ouvert et connexe par arcs, on peut toujours joindre deux points par un arc de classe C^1. Mais cette construction n'est pas si évidente que cela à faire.
On procède donc autrement, en « grignotant » petit à petit.
Soit $a \in \Omega$, et $\Omega' \subset \Omega$ l'ensemble des points $x \in \Omega$ tels que $f(a) = f(x)$. Si $\Omega' \neq \Omega$, considérer $b \in \Omega \setminus \Omega'$ et joindre a à b. Montrer l'existence d'un point c sur cet arc, étant sur la frontière de Ω'. Trouver une contradiction en utilisant la caractérisation (1) et la convexité d'une petite boule centrée en c. □

III.2 Plan tangent et noyau de la différentielle

Définition III.14 (Vecteur tangent à une partie, figure 16.3)**.** *Soit $X \subset E$ une partie de E, et $x \in X$.*

1. *Un vecteur $v \in E$ est tangent à X en x s'il existe $\varepsilon > 0$ et un arc $\gamma :]-\varepsilon, \varepsilon[\to X$ dérivable en 0 tel que $\gamma'(0) = v$.*
2. *On note T_xX l'ensemble des vecteurs tangents à X en x.*

Exemple III.15. *Quels sont les vecteurs tangents à $\varnothing$?*
Pour qu'un vecteur soit tangent, il faut pouvoir construire un arc γ tracé dans X. Or, $\varnothing$ ne contient aucun point. Il est donc impossible de définir un tel arc. L'ensemble des vecteurs tangents est donc l'ensemble vide : $T_x\varnothing = \varnothing$.

2. *0 est tangent à tout X non vide.*
Soit $x \in X$. On considère l'arc constant $\gamma(t) = x$ pour tout t. Cet arc est bien dérivable, $\gamma(0) = x$, et $\gamma'(0) = 0$. Le vecteur nul est donc toujours tangent.
3. *Décrire T_xX lorsque $x \in X$.*
Si x est un point isolé de X, il existe un voisinage de x ne contenant aucun autre point de X. Le seul arc possible est l'arc constant. Donc $T_xX = \{0\}$. Si x n'est pas isolé, l'ensemble peut être plus grand, comme une droite ou un plan.
4. *Décrire les vecteurs tangents à $\mathbb{R} \times \mathbb{R}_+$ en $(0, 0)$.*
Soit $v = (h, k)$ un vecteur tangent. Il existe un arc $\gamma(t) = (x(t), y(t))$ tel que $\gamma(0) = (0, 0)$ et $\gamma'(0) = (h, k)$. Comme $\gamma(t) \in \mathbb{R} \times \mathbb{R}_+$, on a $y(t) \geq 0$ pour tout t. La dérivée en 0 est donc $y'(0) = \lim_{t\to 0} \frac{y(t)-y(0)}{t} = \lim_{t\to 0} \frac{y(t)}{t}$. Puisque $y(t) \geq 0$, cette limite ne peut être négative. On a donc $k \geq 0$. Réciproquement, pour tout $k \geq 0$, l'arc $\gamma(t) = (ht, kt)$ convient car $kt \geq 0$ pour $t > 0$. Donc $T_{(0,0)}(\mathbb{R} \times \mathbb{R}_+) = \{(h, k) \in \mathbb{R}^2 \mid k \geq 0\}$.

Proposition III.16 (Vecteurs tangents à un sous-espace affine)**.** *Soit X un sous-espace affine de E de direction V. Alors, pour tout $B \in X$, $T_B(X) = V$.*

Éléments de preuve. L'inclusion directe est évidente en paramétrant des droites incluses dans X passant par B.
Réciproquement la dérivée de γ est limite d'un taux d'accroissement qui est dans V. Comme on est en dimension finie, V est fermé. □

Proposition III.17 (Vecteurs tangents à une sphère). *Soit E un espace euclidien, $c \in E$ et $r > 0$. Soit $X = S(c,r)$ et $x \in S(c,r)$. Alors $T_x X = x^\perp$. En d'autres termes, l'ensemble des vecteurs tangents à une sphère en un point x est l'orthogonal du rayon défini par x.*

Éléments de preuve. Considérer les grands cercles passant par x.
Pour l'autre inclusion, vérifier que si γ est à valeurs dans S, sa dérivée est orthogonale au rayon (dériver un produit scalaire). □

Exemple III.18. *Décrire sous forme d'une équation le plan vectoriel $T_A(S)$ où $S = S(0,1) \subset \mathbb{R}^3$ et $A = (a,b,c) \in S$.*
D'après la proposition, $T_A S = A^\perp$. L'équation de ce plan vectoriel est donc $ax+by+cz = 0$. En effet, tout vecteur $v = (x,y,z)$ orthogonal au rayon $OA = (a,b,c)$ vérifie le produit scalaire nul.

Théorème III.19 (Description d'un espace tangent à un graphe). *Soit Ω un ouvert de $\mathbb{R}^2$, $f : \Omega \to \mathbb{R}$, et $X_0 \in \Omega$. On suppose f différentiable en X_0. L'ensemble $T_{X_0} f$ des vecteurs tangents à la courbe de f en $(X_0, f(X_0))$ est le plan d'équation*

$$h\frac{\partial f}{\partial x}(X_0) + k\frac{\partial f}{\partial y}(X_0) = \ell,$$

c'est-à-dire l'ensemble des vecteurs $H = (h,k,\ell) \in \mathbb{R}^3$ vérifiant l'équation ci-dessus.

Éléments de preuve. Comme dans les autres exemples, étudier les propriétés des arcs du graphe, et réciproquement, trouver des arcs adéquats. □

Dans le cas général ($\Omega \subset E$, E de dimension finie), on obtient la description suivante :

Théorème III.20 (Espace tangent à un ensemble défini par une équation). *Soit $f : \Omega \to \mathbb{R}$ une fonction numérique de classe C^1, $X = f^{-1}(\{0\})$. Ainsi X est l'ensemble défini par l'équation $f(x) = 0$.*
Soit $a \in X$. Si $df(a) \neq 0$, l'ensemble des vecteurs tangents à X en a est égal à

$$T_a X = \mathrm{Ker}(df(a)).$$

Éléments de preuve. Hors-programme. □

Exemple III.21. *Retrouver à l'aide de ce théorème la description des vecteurs tangents à un espace affine de $\mathbb{R}^n$, décrit par une équation*

$$a_1 x_1 + \cdots + a_n x_n = b.$$

On pose $f(x_1,\dots,x_n) = a_1x_1 + \cdots + a_nx_n - b$. L'ensemble est $X = f^{-1}(\{0\})$. La différentielle est $df(x)(h) = a_1h_1 + \cdots + a_nh_n$, qui est non nulle si les a_i ne sont pas tous nuls. Le noyau de $df(a)$ est donc l'ensemble des h vérifiant $\sum a_i h_i = 0$. D'après le théorème, $T_aX = \mathrm{Ker}(df(a))$. C'est bien la direction de l'hyperplan affine X, qui est un espace vectoriel car X est un sous-espace affine parallèle à ce noyau.

2. *De même, retrouver la description des vecteurs tangents de la sphère unité de l'espace euclidien $\mathbb{R}^n$.*
On pose $f(x) = \|x\|^2 - 1$. La sphère unité est $X = f^{-1}(\{0\})$. La différentielle est $df(a)(h) = 2\langle a,h\rangle$. Pour a sur la sphère, $df(a) \neq 0$ car $a \neq 0$. Donc $T_aX = \mathrm{Ker}(df(a)) = \{h \in \mathbb{R}^n \mid \langle a,h\rangle = 0\} = a^\perp$.
3. *Retrouver la description de l'espace des vecteurs tangents au graphe d'une fonction $f : \Omega \to \mathbb{R}$ où $\Omega \subset \mathbb{R}^2$, lorsque f est de classe C^1 au voisinage de X_0. Généraliser pour $\Omega \subset \mathbb{R}^n$.*
Soit $g : \Omega \times \mathbb{R} \to \mathbb{R}$ définie par $g(x,z) = f(x) - z$. Le graphe de f est $G = g^{-1}(\{0\})$. La différentielle est $dg(x_0,f(x_0))(h,\ell) = df(x_0)(h) - \ell$. En coordonnées dans $\mathbb{R}^2 \times \mathbb{R}$, si $df(x_0)(h) = \frac{\partial f}{\partial x}(x_0)h_1 + \frac{\partial f}{\partial y}(x_0)h_2$, alors $\mathrm{Ker}(dg(x_0,f(x_0)))$ est l'ensemble des (h,ℓ) vérifiant $\ell = df(x_0)(h)$. Donc $T_{(x_0,f(x_0))}G = \{(h,\ell) \in \mathbb{R}^n \times \mathbb{R} \mid \ell = df(x_0)(h)\}$. Pour $\Omega \subset \mathbb{R}^n$, on a de même $T_{(x_0,f(x_0))}G = \{(h,\ell) \in \mathbb{R}^n \times \mathbb{R} \mid \ell = \nabla f(x_0) \cdot h\}$.

Proposition III.22 (Deuxième interprétation géométrique du gradient)**.** *Avec les hypothèses du théorème 16.3.21, si on suppose de plus que E est un espace euclidien, $T_aX = \nabla f(a)^\perp$.*

Éléments de preuve. Dans le contexte euclidien, cette formulation est équivalente à celle du théorème, par définition du gradient.
On peut le démontrer indépendamment du théorème, en considérant un arc à valeurs dans X, et en dérivant $f \circ \gamma$. □

Exemple III.23. *Retrouver l'équation des hyperplans des vecteurs tangents à la sphère unité de $\mathbb{R}^n$. Pour la sphère unité $S(0,1)$, on a $f(x) = \|x\|^2 - 1$ et $\nabla f(a) = 2a$. La proposition donne $T_aS = \nabla f(a)^\perp = \{h \mid \langle 2a, h\rangle = 0\} = \{h \mid \langle a, h\rangle = 0\} = a^\perp$. L'équation est donc $\langle a, h\rangle = 0$.*

Remarque III.24.
1. *Si $E = \mathbb{R}^2$, X est la ligne de niveau de valeur 0. Plus généralement, en considérant $f - b$ au lieu de f, on peut décrire de la sorte toutes les lignes de niveau. La proposition affirme que le gradient en a est orthogonal à la droite tangente à la ligne de niveau en a, ce qu'on peut réexprimer de façon plus abrégée en disant que le gradient est orthogonal aux lignes de niveau.*
2. *Ainsi, si f est de classe C^1 sur un ouvert de $\mathbb{R}^2$, en combinant les deux interprétations, la direction de plus forte pente au point a est toujours orthogonale à la ligne de niveau en a.*

Exemple III.25 (Figure 16.5)**.** *Soit $f : \mathbb{R}^2 \setminus \{(0,0)\} \to \mathbb{R}$ définie par*

$$f(x,y) = \frac{x^2 + y^3}{x^4 + y^2}.$$

En exprimant le gradient, déterminer la tangente à la courbe de niveau en $(1,1)$. La figure 16.5 illustre cet exemple. Le gradient en $(1,1)$ est le vecteur AB, reporté sur le plan $z = 1$, dans lequel est tracée en rouge la courbe de niveau au point $(1,1)$.
Calculons les dérivées partielles. Posons $u(x,y) = x^2 + y^3$ et $v(x,y) = x^4 + y^2$. Alors $f = u/v$. En $(1,1)$, $u = 2$, $v = 2$.

$$\frac{\partial f}{\partial x} = \frac{u_x v - u v_x}{v^2}, \quad u_x = 2x, \quad v_x = 4x^3.$$

Donc $u_x(1,1) = 2$, $v_x(1,1) = 4$. Ainsi $\frac{\partial f}{\partial x}(1,1) = \frac{2\times 2 - 2\times 4}{4} = \frac{4-8}{4} = -1$.

$$\frac{\partial f}{\partial y} = \frac{u_y v - u v_y}{v^2}, \quad u_y = 3y^2, \quad v_y = 2y.$$

Donc $u_y(1,1) = 3$, $v_y(1,1) = 2$. Ainsi $\frac{\partial f}{\partial y}(1,1) = \frac{3\times 2 - 2\times 2}{4} = \frac{6-4}{4} = \frac{1}{2}$.
Le gradient est $\nabla f(1,1) = \left(-1, \frac{1}{2}\right)$. La tangente à la ligne de niveau $f(x,y) = f(1,1) = 1$ est la droite passant par $(1,1)$ orthogonale au gradient. Soit $M = (x,y)$ un point de cette tangente. Le vecteur $\overrightarrow{AM} = (x-1, y-1)$ doit vérifier $\nabla f(1,1) \cdot \overrightarrow{AM} = 0$. Donc :

$$-1 \cdot (x-1) + \frac{1}{2} \cdot (y-1) = 0.$$

En multipliant par 2 pour simplifier :

$$-2(x-1) + (y-1) = 0.$$

Donc $-2x + 2 + y - 1 = 0$, soit $y - 2x + 1 = 0$, ou encore $y = 2x - 1$. C'est l'équation de la tangente à la ligne de niveau en $(1,1)$.

III.3 Applications de classe C^k

Soit $B = (b_1, \dots, b_n)$ une base de E, $(i,j) \in [1,n]$, et $f : \Omega \to F$ admettant une dérivée partielle $\partial_i f$ sur tout Ω. Si la fonction $\partial_i f : \Omega \to F$ admet elle-même une j-ième dérivée partielle en $a \in \Omega$, on peut donc définir $\partial_j \partial_i f(a)$, la j-ième dérivée partielle de $\partial_i f$. De façon plus générale, on définit par récurrence sur k les dérivées d'ordre k :

Définition III.26 (Dérivées partielles d'ordre k)**.** *On définit, si c'est possible, les dérivées d'ordre k de f par récurrence sur k :*

- *Si $k = 1$, il s'agit des dérivées partielles $\partial_i f$.*
- *Soit $k > 1$ et $(i_1, \dots, i_k) \in [\![1,n]\!]^k$. La fonction f admet en $a \in \Omega$ une dérivée par rapport successivement aux $i_1, \dots, i_{k-1}$ et i_k-ème vecteurs de la base B, notée $\partial_{i_k} \partial_{i_{k-1}} \dots \partial_{i_1} f(a)$, si et seulement si $\partial_{i_{k-1}} \dots \partial_{i_1} f(a)$ existe au voisinage (relatif dans Ω) de a et est dérivable selon b_{i_k} en a. Ainsi,*
$$\partial_{i_k} \partial_{i_{k-1}} \dots \partial_{i_1} f(a) = \partial_{i_k} \left(\partial_{i_{k-1}} \dots \partial_{i_1} f \right)(a).$$

On trouve aussi les notations suivantes :

- *$\partial_{i_1,\dots,i_k} f(a)$ (attention à l'ordre des indices !),*
- *$\frac{\partial^k f}{\partial x_{i_k} \cdots \partial x_{i_1}}(a)$, si $(x_1, \dots, x_n)$ représentent les coordonnées génériques dans la base B, en particulier si $E = \mathbb{R}^n$ muni de la base canonique.*

On peut définir des différentielles d'ordre supérieur, mais cette notion est hors programme, et les objets manipulés deviennent de plus en plus complexes. Pour cette raison, nous définissons la classe C^k à l'aide des dérivées partielles d'ordre k. Cette définition est une généralisation de la caractérisation de la classe C^1 par les dérivées partielles premières.

Définition III.27 (Applications de classe C^k)**.** *Soit $f : \Omega \to F$ une application définie sur l'ouvert Ω de E.*

1. *On dit que f est de classe C^k sur Ω si toutes les dérivées d'ordre k de f existent (il y en a n^k), et sont continues sur Ω.*
2. *On dit que f est de classe C^k au voisinage de $a \in \Omega$, s'il existe un ouvert $V \subset \Omega$ tel que $f|_V$ soit de classe C^k sur V.*
3. *On note $C^k(\Omega, F)$ l'ensemble des fonctions de classe C^k sur Ω, à valeurs dans F. Par convention, $C^0(\Omega, F)$ est l'ensemble des fonctions continues.*
4. *On dit que f est de classe C^∞ sur Ω si elle est de classe C^k pour tout $k \in \mathbb{N}$.*
5. *On note $C^\infty(\Omega, F)$ l'ensemble des fonctions de classe C^∞ sur Ω, à valeurs dans F.*

Remarque III.28. *Si f est de classe C^k, alors :*

- *f est de classe C^j pour tout $j \leq k$;*
- *plus généralement, si $\ell \leq k$ et $(i_1, \dots, i_\ell) \in [\![0,n]\!]^\ell$, $\partial_{i_1,\dots,i_\ell} f$ est de classe $C^{k-\ell}$.*

Pour étudier la classe C^k, il y a donc a priori n^k dérivées à étudier. Mais en pratique, cela en fait moins, car sous des conditions assez peu restrictives, on va retrouver plusieurs fois la même, dans le sens où l'ordre de dérivation (i.e. le choix de l'ordre des variables par rapport auxquelles on dérive) n'importe pas. C'est ce que dit le théorème suivant.

Théorème III.29 (Théorème de Schwarz)**.** *Soit $f : \Omega \to F$, $B = (b_1, \dots, b_n)$ une base de E, et $a \in \Omega$. Si f est de classe C^2 au voisinage de a, alors*
$$\forall (i,j) \in [1,n], \quad \partial_i \partial_j f(a) = \partial_j \partial_i f(a).$$

Éléments de preuve. Démonstration non exigible.
En fixant toutes les autres variables, on est ramené au cas de fonctions de 2 variables. Exprimer un DL2 de $f(x+h,y+k)-f(x,y)$ en $H=(h,k)$, de deux manières différentes, en intercalant $f(x+h,y)$, ou $f(x,y+k)$, et en se ramenant à la formule de Taylor-Young à l'ordre 2 pour les fonctions d'une variable (en fixant l'autre). Identifier les deux DL obtenus (justifier proprement cette identification). □

Corollaire III.30 (Théorème de Schwarz pour la classe C^k). *Soit $f:\Omega\to F$, $B=(b_1,\dots,b_n)$ une base de E, et $a\in\Omega$. Si f est de classe C^2 au voisinage de a, alors pour tout $(i_1,\dots,i_k)\in[1,n]^k$, et toute permutation $\sigma\in S_k$,*

$$\partial_{i_1,\dots,i_k}f(a)=\partial_{i_{\sigma(1)},\dots,i_{\sigma(k)}}f(a).$$

Éléments de preuve. Les dérivées partielles successives commutent 2 à 2. On peut donc ramener d'abord i_1 en position initiale, par échange successif, pour faire ensuite une récurrence. Ou directement utiliser le fait que S_n est engendré par les transpositions $(i\ i+1)$. □

Ainsi, l'ordre des variables de dérivation importe peu, dès lors que la fonction est de classe C^k, ce qui permet de regrouper éventuellement les dérivées selon les variables, afin d'écrire ces dérivées sous la forme :

$$\frac{\partial^k f}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}},\quad \alpha_1+\cdots+\alpha_n=k.$$

Cette notation signifie qu'on a dérivé α_1 fois par rapport à x_1, α_2 fois par rapport à x_2 etc. Certains α_i peuvent être nuls et peuvent être omis dans la notation. On notera par exemple $\frac{\partial^3 f}{\partial x\partial z^2}$ la dérivée d'ordre 3 d'une fonction de 3 variables x,y,z, obtenue en dérivant une fois par rapport à x et 2 fois par rapport à z, mais aucune fois par rapport à y.

Proposition III.31 (Opérations sur les applications de classe C^k). *Soit E,E' des $\mathbb{R}$-e.v.n. de dimension finie, et $F,F_1,\dots,F_q$ des K-e.v.n. de dimension finie. Soit Ω,Ω' des ouverts de E et E' respectivement.*

1. *Soit $f,g:\Omega\to F$ de classe C^k et $\lambda\in K$. Alors $f+\lambda g$ est de classe C^k.*
2. *Soit pour $i\in[1,q]$, $f_i:\Omega\to F_i$, et $M:F_1\times\cdots\times F_q\to F$ une forme multilineaire. Si les f_i sont de classe C^k, alors $M(f_1,\dots,f_q)$ est de classe C^k.*
3. *Soit $f:\Omega'\to\Omega$ et $g:\Omega\to F$. Si f et g sont de classe C^k sur Ω' et Ω respectivement, alors $g\circ f$ est de classe C^k sur Ω.*
4. *Si $f:\Omega\to K$ ne s'annule pas et est de classe C^k, alors $\frac{1}{f}$ aussi.*
5. *Soit pour $i\in[1,q]$, $f_i:\Omega\to F_i$, et $f:\Omega\to F_1\times\cdots\times F_q$ définie par*

$$\forall x\in\Omega,\quad f(x)=(f_1(x),\dots,f_q(x)).$$

f est de classe C^k si et seulement si tout f_i est de classe C^k.

Éléments de preuve. Démonstrations non exigibles, un peu fastidieuse mais sans difficulté. Faire une récurrence, en dérivant une première fois. □

Exemple III.32. 1. *Les fonctions polynomiales $f:\mathbb{R}^n\to K$, de n variables réelles et à coefficients dans K, sont de classe C^∞.*
2. *Les fractions rationnelles aussi, sur leur domaine de définition.*
3. $\det:M_n(\mathbb{R})\to\mathbb{R}$ *est de classe C^∞.*
4. $M\mapsto M^{-1}$ *définie sur* $\mathrm{GL}_n(\mathbb{R})$ *est de classe C^∞.*

Dans le cas de fonctions d'une seule variable réelle, on rajoute une propriété, concernant la dérivée itérée d'une forme bilinéaire. Lorsqu'on applique cette formule au cas du produit de deux réels, on retrouve la formule de Leibniz classique.

Théorème III.33 (Formule de Leibniz vectorielle)**.** *Soit E, F, G des K-evn de dimension finie, et $B : E \times F \to G$ des K-evn. Soit I un intervalle, et $f : I \to E$ et $g : I \to F$ deux applications. Si f et g sont de classe C^k, alors $B(f,g) : t \mapsto B(f(t), g(t))$ est aussi de classe C^k, et*

$$\forall t \in I, \quad B(f,g)^{(k)}(t) = \sum_{i=1}^{k} \binom{k}{i} B(f^{(i)}(t), g^{(k-i)}(t))\, dt.$$

III.4 Hessienne et formule de Taylor-Young à l'ordre 2

Définition III.34 (Matrice hessienne de f)**.** *Soit Ω un ouvert de $\mathbb{R}^n$, et $f : \Omega \to \mathbb{R}$ de classe C^2. La matrice hessienne de f en $a \in \Omega$ est*

$$H_f(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{1 \leq i,j \leq n} = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(a) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(a) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(a) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(a) \end{pmatrix}.$$

Proposition III.35 (Symétrie de $H_f(a)$)**.** *Sous les mêmes hypothèses, $H_f(a) \in S_n(\mathbb{R})$.*

Éléments de preuve. C'est une réexpression du théorème de Schwarz. □

Lemme III.36 (Expression de la dérivée seconde par rapport à un vecteur)**.** *Soit Ω un ouvert de $\mathbb{R}^n$, et $f : \Omega \to \mathbb{R}$ de classe C^2. Soit $a \in \Omega$ et $U \in \mathbb{R}^n$. Alors, la dérivée seconde selon le vecteur U au point a s'exprime à l'aide de la hessienne :*

$$f''_{a,U}(0) = U^T H_f(a) U = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i u_j \frac{\partial^2 f}{\partial x_i \partial x_j}(a), \quad \text{où } U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}.$$

La matrice hessienne va nous permettre d'exprimer de façon élégante la formule de Taylor-Young, c'est-à-dire l'expression d'un DL à l'ordre 2 à l'aide des dérivées d'ordre 1 et 2, sous les hypothèses idoines de régularité.
Nous avons déjà obtenu une première version de cette formule, lorsque f est une fonction de 2 variables de classe C^2 au voisinage d'un point a, lors de la démonstration du théorème de Schwarz :

$$f(x+h, y+k) = f(x,y) + h\frac{\partial f}{\partial x}(x,y) + k\frac{\partial f}{\partial y}(x,y) + \frac{1}{2}h^2\frac{\partial^2 f}{\partial x^2}(x,y) + hk\frac{\partial^2 f}{\partial x \partial y}(x,y) + \frac{1}{2}k^2\frac{\partial^2 f}{\partial y^2}(x,y) + o(\|H\|^2),$$

où $H = \begin{pmatrix} h \\ k \end{pmatrix}$. À l'aide du lemme précédent, on obtient une réexpression synthétique de cette formule, en posant $X = \begin{pmatrix} x \\ y \end{pmatrix}$:

$$f(X+H) = f(X) + \langle \nabla f(X), H \rangle + \frac{1}{2} H^\top H_f(x) H + o(\|H\|^2).$$

Plus généralement :

Théorème III.37 (Formule de Taylor-Young à l'ordre 2)**.** *Soit Ω un ouvert de $\mathbb{R}^n$ et $f : \Omega \to \mathbb{R}$ de classe C^2 au voisinage de $a \in \Omega$. Alors f admet un DL à l'ordre 2 en a, s'exprimant :*

$$f(a+h) = f(a) + \langle \nabla f(a), h\rangle + \frac{1}{2} h^\top H_f(a) h + o(\|h\|^2).$$

Éléments de preuve. Démonstration non exigible. On peut procéder comme dans le cas de 2 variables, en utilisant la formule de Taylor-Young pour les fonctions d'une variable, successivement sur chaque variable. On peut aussi exploiter la symétrie de $H_f(a)$ en considérant d'abord une b.o.n. de diagonalisation $(b_1, \dots, b_n)$, et en exploitant la formule de Taylor à l'ordre 2 pour les fonctions partielles associées. Les exprimer par ε, et exprimer le cas général par linéarité et bilinéarité. □

III.5 Équations aux dérivées partielles (EDP)

Une équation aux dérivées partielles (EDP) est une équation d'inconnue une fonction f de plusieurs variables, reliant ses différentes dérivées partielles. L'ordre de l'EDP est l'ordre de dérivation le plus important intervenant dans cette équation. Par exemple, l'équation suivante, pour une fonction inconnue f de 2 variables, l'équation

$$\frac{\partial^3 f}{\partial x^3} = \frac{\partial^2 f}{\partial y^2} + xy \frac{\partial^2 f}{\partial x \partial y}$$

est une EDP d'ordre 3. La résolution d'équations aux dérivées partielles est un problème en général dur. Nous nous contentons ici de donner quelques pistes, sur des exemples simples.
Le principe va souvent être de se ramener à des équations simples. Pour cela, nous commençons par une forme simple d'équation de degré 1.

Proposition III.38 (Résolution de $\frac{\partial f}{\partial x_i} = 0$)**.** *Soit $i \in [1, n]$. Les propositions suivantes sont équivalentes:*

1. *$f : \mathbb{R}^n \to \mathbb{R}$ est de classe C^1 et vérifie $\frac{\partial f}{\partial x_i} = 0$ sur Ω;*
2. *il existe $g \in C^1(\mathbb{R}^{n-1}, \mathbb{R})$ tel que pour tout $(x_1, \dots, x_n) \in \mathbb{R}^n$,*

$$f(x_1, \dots, x_n) = g(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n).$$

En d'autres termes, les solutions de l'EDP $\frac{\partial f}{\partial x_i} = 0$ sont les fonctions ne dépendant pas de la variable x_i et de classe C^1 par rapport aux autres variables.

Exemple III.39. *Résoudre $a\frac{\partial f}{\partial x} + b\frac{\partial f}{\partial y} = 0$, $(a, b) \neq (0, 0)$, $f \in C^1(\mathbb{R}^2, \mathbb{R})$.*
Solution:
Analyse: *Supposons qu'une solution f existe. On cherche un changement de variables linéaire qui transforme l'opérateur différentiel en une dérivée partielle simple. Posons :*

$$u = bx - ay, \quad v = ax + by, \quad F(u, v) = f(x, y)$$

Ce choix est motivé par le fait que $a\frac{\partial u}{\partial x} + b\frac{\partial u}{\partial y} = a \cdot b + b \cdot (-a) = 0$ (u est constant le long des courbes caractéristiques) et que (u, v) forme une base de $\mathbb{R}^2$ car le déterminant est $b^2 + a^2 \neq 0$.
Calculons les dérivées de f en fonction de F :

$$\frac{\partial f}{\partial x} = b\frac{\partial F}{\partial u} + a\frac{\partial F}{\partial v}, \quad \frac{\partial f}{\partial y} = -a\frac{\partial F}{\partial u} + b\frac{\partial F}{\partial v}$$

Substituons dans l'équation :

$$a\left(b\frac{\partial F}{\partial u} + a\frac{\partial F}{\partial v}\right) + b\left(-a\frac{\partial F}{\partial u} + b\frac{\partial F}{\partial v}\right) = (a^2 + b^2)\frac{\partial F}{\partial v} = 0$$

Comme $a^2+b^2\neq 0$, on obtient $\frac{\partial F}{\partial v}=0$.
Synthèse: *De $\frac{\partial F}{\partial v}=0$ on déduit que F ne dépend pas de v :*

$$\frac{\partial F}{\partial v}=0 \quad \textit{donc} \quad F(u,v)=H(u)$$

où H est une fonction arbitraire C^1.
En revenant aux variables initiales, on obtient la solution générale :

$$\boxed{f(x,y)=H(bx-ay)} \quad \textit{avec } H\in C^1(\mathbb{R},\mathbb{R}) \textit{ arbitraire.}$$

Exemple III.40. *Résoudre $\frac{\partial^2 f}{\partial x\partial y}=0$, $f\in C^2(\mathbb{R}^2,\mathbb{R})$.*
Solution:
Posons $g=\frac{\partial f}{\partial y}$. Alors $\frac{\partial g}{\partial x}=0$ donc $g(x,y)=\phi(y)$. Donc $\frac{\partial f}{\partial y}=\phi(y)$ donc $f(x,y)=\int\phi(y)\,dy+\psi(x)=\Phi(y)+\psi(x)$.
Ainsi $\boxed{f(x,y)=\Phi(y)+\psi(x)}$ avec $\Phi,\psi\in C^2(\mathbb{R},\mathbb{R})$ arbitraires.

2. *Résoudre $\frac{\partial^2 f}{\partial x^2}-a^2\frac{\partial^2 f}{\partial y^2}=0$, $a>0$, $f\in C^2(\mathbb{R}^2,\mathbb{R})$.*
Solution:
Analyse:
Cherchons un changement de variables linéaire qui simplifie l'équation. L'opérateur $\frac{\partial^2}{\partial x^2}-a^2\frac{\partial^2}{\partial y^2}$ est un opérateur d'onde. On cherche des variables p,q telles que l'équation devienne $\frac{\partial^2 F}{\partial p\partial q}=0$.
Considérons le changement :

$$p=x+ay,\quad q=x-ay,\quad F(p,q)=f(x,y)$$

Ce choix correspond aux courbes caractéristiques de l'équation des ondes.
Calculons les dérivées premières :

$$\frac{\partial f}{\partial x}=\frac{\partial F}{\partial p}+\frac{\partial F}{\partial q},\quad \frac{\partial f}{\partial y}=a\frac{\partial F}{\partial p}-a\frac{\partial F}{\partial q}$$

Calculons les dérivées secondes :

$$\frac{\partial^2 f}{\partial x^2}=\frac{\partial}{\partial x}\left(\frac{\partial F}{\partial p}+\frac{\partial F}{\partial q}\right)=\frac{\partial^2 F}{\partial p^2}+2\frac{\partial^2 F}{\partial p\partial q}+\frac{\partial^2 F}{\partial q^2}$$

$$\frac{\partial^2 f}{\partial y^2}=\frac{\partial}{\partial y}\left(a\frac{\partial F}{\partial p}-a\frac{\partial F}{\partial q}\right)=a^2\frac{\partial^2 F}{\partial p^2}-2a^2\frac{\partial^2 F}{\partial p\partial q}+a^2\frac{\partial^2 F}{\partial q^2}$$

Substituons dans l'équation $\frac{\partial^2 f}{\partial x^2}-a^2\frac{\partial^2 f}{\partial y^2}=0$:

$$\left(\frac{\partial^2 F}{\partial p^2}+2\frac{\partial^2 F}{\partial p\partial q}+\frac{\partial^2 F}{\partial q^2}\right)-a^2\left(a^2\frac{\partial^2 F}{\partial p^2}-2a^2\frac{\partial^2 F}{\partial p\partial q}+a^2\frac{\partial^2 F}{\partial q^2}\right)=0$$

En simplifiant (ou en vérifiant directement pour $a=1$), on constate que cette expression se réduit à :

$$4a^2\frac{\partial^2 F}{\partial p\partial q}=0 \quad \textit{donc} \quad \frac{\partial^2 F}{\partial p\partial q}=0$$

Synthèse:
De $\frac{\partial^2 F}{\partial p\partial q}=0$ on déduit, comme à la question 1, que F est somme d'une fonction de p et d'une fonction de q :

$$\frac{\partial^2 F}{\partial p\partial q}=0 \quad \textit{donc} \quad F(p,q)=\Phi(p)+\Psi(q)$$

où Φ et Ψ sont des fonctions C^2 arbitraires.
En revenant aux variables initiales, on obtient la solution générale de l'équation des ondes :

$$\boxed{f(x,y)=\varphi(x+ay)+\psi(x-ay)} \quad \textit{avec } \varphi,\psi\in C^2(\mathbb{R},\mathbb{R}) \textit{ arbitraires.}$$

IV. Optimisation

IV.1 Principes généraux

L'optimisation consiste en l'étude des extrema d'une fonction f (globaux ou locaux).

Définition IV.1 (Extrema). *Soit D un sous-ensemble de E, et $f : D \to \mathbb{R}$, et $a \in D$. On dit que :*

- *f admet un minimum global (resp. maximum global) en a si pour tout $x \in D$, $f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$);*
- *f admet un minimum global strict (resp. maximum global strict) en a si pour tout $x \in D \setminus \{a\}$, $f(x) > f(a)$ (resp. $f(x) < f(a)$);*
- *f admet un minimum local (resp. maximum local) en a s'il existe $V \in \mathcal{V}(a)$ tel que pour tout $x \in D \cap V$, $f(x) \geq f(a)$ (resp. $f(x) \leq f(a)$);*
- *f admet un minimum local strict (resp. maximum local strict) en a s'il existe $V \in \mathcal{V}(a)$ tel que pour tout $x \in D \cap V \setminus \{a\}$, $f(x) > f(a)$ (resp. $f(x) < f(a)$).*

Le but de cette section est de donner des principes de recherche des extrema (locaux ou globaux). Les sous-sections suivantes donneront des méthodes pour trouver en quels points de D il existe un extremum, local ou global. Pour commencer, nous nous intéressons à l'existence d'un extremum global (ou local par restriction à un voisinage).
Nous rappelons le théorème de compacité, outil incontournable dans ce contexte.

Théorème IV.2 (Compacté). *Si D est compact et $f : D \to \mathbb{R}$ est continue, alors f admet un minimum et un maximum globaux.*

Il est fréquent d'étudier des fonctions qui ne sont pas définies sur des compacts. Mais il est possible de s'y ramener, soit en restreignant (si D n'est pas borné), soit en prolongeant (si D est borné).

Méthode IV.3 (Prouver l'existence d'un extremum global). *On peut bien entendu utiliser les méthodes de localisation des extrema des paragraphes suivants et comparer les valeurs à toutes les autres, mais il peut être intéressant de savoir justifier a priori l'existence d'un extremum global de $f : D \to \mathbb{R}$.*

1. *Si D est compact et f continue, le théorème de compacité donne l'existence d'un maximum et d'un minimum.*
2. *Si on parvient à trouver un compact $K \subset D$, et $a \in D$ tel que*

$$\forall x \in D \setminus K, \quad f(x) \geq f(a),$$

alors f admet un minimum sur D, atteint dans K.

3. *Si D est borné, $\overline{D}$ est borné. Si f admet une limite en tout point de $\overline{D} \setminus D$, et s'il existe $a \in D$ tel que*

$$\forall b \in \overline{D} \setminus D, \quad \lim_{x \to b} f(x) \geq f(a),$$

alors f admet un minimum sur D.

Exemple IV.4. 1. *Montrer que $f : (x, y) \mapsto \frac{x^2y - 3y + y^2}{(1+x^2+y^2)^2}$ admet un minimum et un maximum sur $\mathbb{R}^2$.*

2. *Montrer que $g : (x, y) \mapsto \frac{e^{-e^{xy}}}{1-(x^2+y^2)}$ admet un minimum sur $B(0, 1)$.*

Solution:
1. Étude de $f(x, y) = (x^2 + y^2 - 1)e^{-x^2-y^2}$
La fonction f est continue sur $\mathbb{R}^2$. Comportement à l'infini :

$$\lim_{\|(x,y)\| \to +\infty} f(x, y) = 0$$

car l'exponentielle négative domine la croissance polynomiale.
Pour $\varepsilon = 1$, il existe $R > 0$ tel que

$$\|(x,y)\| > R \implies |f(x,y)| < 1.$$

Sur la boule fermée $\overline{B}(0,R)$ (compacte), f atteint son minimum m et son maximum M par continuité.
Pour tout $(x,y) \in \mathbb{R}^2$, on a donc :

$$f(x,y) \geq \min(m,-1) \quad \text{et} \quad f(x,y) \leq \max(M,1)$$

ce qui prouve que f est bornée sur $\mathbb{R}^2$.
Puisque $f(0,0) = -1 < 0$ et $f(0,3) = 8e^{-9} > 0$, la fonction prend des valeurs négatives et positives.
Comme elle tend vers 0 à l'infini, ses bornes (inférieure et supérieure) sont nécessairement atteintes sur le compact $\overline{B}(0,R)$ (car en dehors, $|f| < 1$ alors que le minimum est ≤ -1 et le maximum $\geq 8e^{-9} < 1$).
Ainsi, f admet un minimum et un maximum globaux sur $\mathbb{R}^2$.

2. Étude de $g(x,y) = \dfrac{e^{-x^2-y^2}}{1-x^2-y^2}$
g est continue sur $B(0,1)$ et $\lim\limits_{x^2+y^2 \to 1^-} g(x,y) = +\infty$.
Soit $m = \inf_{B(0,1)} g$. Il existe $r \in]0,1[$ tel que :

$$x^2+y^2 > r \implies g(x,y) > m+1.$$

Posons $K = \{(x,y) \mid x^2+y^2 \leq r\}$ (compact). Sur K, g atteint son minimum m_K.
Alors $\inf_{B\setminus K} g \geq m+1 > m \geq m_K$. Donc $m = \min(m_K, \inf_{B\setminus K} g) = m_K$.
Ainsi g admet un minimum global m_K, atteint dans K.

IV.2 Condition nécessaire du premier ordre

Théorème IV.5 (CN du premier ordre pour un extremum local)**.** *Soit $f : D \to \mathbb{R}$. Si f admet un extremum local en $a \in \mathring{D}$ et si f est différentiable en a, alors $df(a) = 0$.*

Remarque IV.6 (Avertissement)**.** *Comme dans le cas d'une fonction d'une variable réelle, $df(a) = 0$ n'est pas une condition suffisante pour avoir un extremum local.*

Exemple IV.7 (Figure 16.8)**.** *Quels sont les points critiques de $f : (x,y) \mapsto xy$? La fonction f admet-elle un extremum local ? Cet exemple montre au passage que le fait que les restrictions de f à toutes les droites passant par a admettent un extremum local en a n'est pas suffisant pour que f admette un extremum local en a.*

Méthode IV.8 (Rechercher les extrema locaux sur un ouvert)**.** *On cherche les points critiques de f, puis on étudie chaque point critique a :*

- *soit par étude directe du signe de $f(x) - f(a)$,*
- *soit en utilisant la CN d'ordre 2 qui fait l'objet d'un paragraphe ultérieur.*

Méthode IV.9 (Rechercher les extrema globaux)**.** *Soit $f : D \to \mathbb{R}$, différentiable sur D.*

- *Rechercher les points critiques sur $\mathring{D}$. Les extrema globaux, s'ils existent, sont un point critique, ou un point de $D \setminus \mathring{D}$.*
- *Si on a pu justifier l'existence d'un extremum global, on peut le trouver en comparant les valeurs prises aux différents points critiques, ainsi que le maximum et/ou le minimum de f restreinte à $D \setminus \mathring{D}$.*

Le paragraphe suivant donne justement des techniques en vue de l'étude des extrema d'une fonction restreinte. Cela permet notamment l'étude aux bords (si D lui-même peut être plongé dans un ouvert Ω de sorte que $D \setminus \mathring{D}$ soit une partie de Ω).

IV.3 Optimisation sous contrainte

Soit $f : \Omega \to \mathbb{R}$, et $X \subset D$. Optimiser f sous la contrainte X consiste à étudier les extrema de f en se restreignant à X (la contrainte). Il s'agit donc d'étudier les extrema d'une restriction. Nous commençons par une CN similaire à celle qu'on a trouvé dans le cas global (qui d'ailleurs peut être vu comme conséquence de celle-ci).

Proposition IV.10 (CN du premier ordre pour une restriction)**.** *Soit Ω un ouvert de E, X une partie de Ω, et $f : \Omega \to \mathbb{R}$. Si $f|_X$ admet un extremum local en $a \in X$, alors*

$$\forall h \in T_a X, \quad df(a) \cdot h = 0.$$

En d'autres termes, $df(a)$ s'annule en tout vecteur tangent à X, c'est-à-dire $T_a X \subset \operatorname{Ker} df(a)$.

Éléments de preuve. Dériver $f(\alpha(t))$, où $\alpha : [-\varepsilon, \varepsilon] \to X$ vérifie $\alpha(0) = a$. □

Remarque IV.11. *En prenant $X = \mathring{D}$, puisque X est un voisinage de a, $T_a X = E$, et on retrouve donc le théorème 16.4.5.*

2. *Si E est euclidien, la proposition 16.4.15 affirme que si $f|_X$ admet un extremum local en a, $\nabla f(a) \perp T_a X$.*

Si l'ensemble X est défini par une équation $g(x) = 0$, on peut combiner le résultat précédent et le théorème 16.3.21.

Théorème IV.12 (Optimisation sous contrainte)**.** *Soit $f, g : \Omega \to \mathbb{R}$ de classe C^1, et $X = \{x \in \Omega, g(x) = 0\}$. Si $f|_X$ admet un extremum local en $a \in X$ tel que $dg(a) \neq 0$, alors $df(a)$ et $dg(a)$ sont colinéaires.*

Exercices d'optimisation sous contrainte

Exemple IV.13. *Déterminer les extrema de $f(x, y) = x^2 + xy - 2y^2$ sur le cercle unité $g(x, y) = x^2 + y^2 - 1 = 0$.*

Solution
Existence. Le cercle unité est compact et f est continue, donc f admet un minimum et un maximum sur ce cercle.
Multiplicateurs de Lagrange. On résout :

$$\begin{cases} \nabla f = \lambda \nabla g \\ g(x, y) = 0 \end{cases}$$

Les gradients :

$$\nabla f = (2x + y,\ x - 4y), \quad \nabla g = (2x,\ 2y)$$

D'où le système linéaire en x, y :

$$\begin{cases} 2x + y = 2\lambda x \\ x - 4y = 2\lambda y \end{cases}$$

Soit :

$$\begin{cases} (2 - 2\lambda)x + y = 0 \\ x + (-4 - 2\lambda)y = 0 \end{cases}$$

Recherche de λ. Pour que ce système homogène admette une solution non nulle $(x, y) \neq (0, 0)$, il faut que son déterminant soit nul :

$$\det \begin{pmatrix} 2 - 2\lambda & 1 \\ 1 & -4 - 2\lambda \end{pmatrix} = 0$$

Calculons :

$$(2-2\lambda)(-4-2\lambda)-1=0$$
$$-8-4\lambda+8\lambda+4\lambda^2-1=0$$
$$4\lambda^2+4\lambda-9=0$$
$$\lambda^2+\lambda-\frac{9}{4}=0$$

Le discriminant est $\Delta=1+9=10$, donc :

$$\lambda=\frac{-1\pm\sqrt{10}}{2}$$

Recherche de (x,y)**.** Pour chaque λ, le système devient une seule équation (les deux sont proportionnelles).
Pour $\lambda=\frac{-1+\sqrt{10}}{2}$: La première équation : $(2-2\lambda)x+y=0$. Calculons $2-2\lambda=2-(-1+\sqrt{10})=3-\sqrt{10}$. Donc $(3-\sqrt{10})x+y=0$, soit $y=(\sqrt{10}-3)x$.
Pour $\lambda=\frac{-1-\sqrt{10}}{2}$: $2-2\lambda=2-(-1-\sqrt{10})=3+\sqrt{10}$. Donc $(3+\sqrt{10})x+y=0$, soit $y=-(3+\sqrt{10})x$.
Contrainte $g(x,y)=0$ **i.e.** $x^2+y^2=1$**.** Pour le premier cas : $y=(\sqrt{10}-3)x$. Alors :

$$x^2+(\sqrt{10}-3)^2x^2=x^2[1+(\sqrt{10}-3)^2]=1$$

Or $(\sqrt{10}-3)^2=10-6\sqrt{10}+9=19-6\sqrt{10}$. Donc :

$$1+19-6\sqrt{10}=20-6\sqrt{10}$$

Ainsi :

$$x^2=\frac{1}{20-6\sqrt{10}},\quad x=\pm\frac{1}{\sqrt{20-6\sqrt{10}}}$$

et $y=(\sqrt{10}-3)x=\pm\frac{\sqrt{10}-3}{\sqrt{20-6\sqrt{10}}}$.
Pour le second cas : $y=-(3+\sqrt{10})x$. Alors :

$$x^2+(3+\sqrt{10})^2x^2=x^2[1+(3+\sqrt{10})^2]=1$$

Or $(3+\sqrt{10})^2=19+6\sqrt{10}$. Donc :

$$1+19+6\sqrt{10}=20+6\sqrt{10}$$

Ainsi :

$$x^2=\frac{1}{20+6\sqrt{10}},\quad x=\pm\frac{1}{\sqrt{20+6\sqrt{10}}}$$

et $y=-(3+\sqrt{10})x=\mp\frac{3+\sqrt{10}}{\sqrt{20+6\sqrt{10}}}$.
Valeurs de f**.** Calculons $f=x^2+xy-2y^2$ pour chaque cas.
Pour le premier cas ($\lambda=\frac{-1+\sqrt{10}}{2}$), on trouve $f=\frac{-1+\sqrt{10}}{2}$.
Pour le second cas ($\lambda=\frac{-1-\sqrt{10}}{2}$), on trouve $f=\frac{-1-\sqrt{10}}{2}$.
Conclusion. Les extremums sont :

$$\boxed{\max f=\frac{-1+\sqrt{10}}{2},\quad \min f=\frac{-1-\sqrt{10}}{2}}$$

Les points où ils sont atteints sont respectivement $(x,y)=\left(\pm\frac{1}{\sqrt{20-6\sqrt{10}}},\ \pm\frac{\sqrt{10}-3}{\sqrt{20-6\sqrt{10}}}\right)$ pour le maximum et $(x,y)=\left(\pm\frac{1}{\sqrt{20+6\sqrt{10}}},\ \mp\frac{3+\sqrt{10}}{\sqrt{20+6\sqrt{10}}}\right)$ pour le minimum.

Exemple IV.14. *Montrer l'inégalité arithmético-géométrique :*

$$\frac{x_1+\cdots+x_n}{n} \geq \sqrt[n]{x_1\cdots x_n} \quad (x_i>0)$$

Solution
On maximise $f(x_1,\ldots,x_n)=\sqrt[n]{x_1\cdots x_n}$ sous la contrainte $g(x_1,\ldots,x_n)=x_1+\cdots+x_n-n=0$, avec $x_i>0$.
Existence. Soit $K=\{x\in\mathbb{R}^n \mid x_i\geq 0 \text{ et } g(x)=0\}$. L'ensemble K est fermé (image réciproque du fermé $\{0\}$ par l'application continue g) et borné (car $x_i\geq 0$ et $\sum x_i=n$ implique $0\leq x_i\leq n$). K est compact car c'est un fermé bonrné en dimension finie. La fonction f est continue sur K (prolongée par continuité en posant $f=0$ dès qu'un $x_i=0$). D'après le théorème des extremums, f atteint son maximum sur K. Puisque $f(x)>0$ lorsque tous les $x_i>0$ et que $f(x)=0$ lorsqu'un $x_i=0$, le maximum est strictement positif et n'est donc pas atteint sur la partie où un x_i s'annule. Par conséquent, le maximum est atteint en un point où tous les $x_i>0$.
Recherche du maximum par la méthode des extrema liés. Au point où le maximum est atteint, le gradient de f est colinéaire au gradient de g. Il existe donc un réel λ tel que :

$$\nabla f=\lambda\nabla g$$

Calculons les dérivées partielles. En écrivant $f=(x_1\cdots x_n)^{1/n}$, on a :

$$\frac{\partial f}{\partial x_i}=\frac{1}{n}(x_1\cdots x_n)^{1/n}\cdot\frac{1}{x_i}=\frac{f}{nx_i}$$

Comme $\frac{\partial g}{\partial x_i}=1$, la condition $\nabla f=\lambda\nabla g$ donne pour tout i :

$$\frac{f}{nx_i}=\lambda\cdot 1 \quad\Rightarrow\quad \frac{f}{nx_i}=\lambda$$

Cette égalité étant vraie pour tout i, on en déduit :

$$\frac{f}{nx_1}=\frac{f}{nx_2}=\cdots=\frac{f}{nx_n}$$

Puisque $f>0$ au point considéré, on simplifie par $\frac{f}{n}$ et on obtient :

$$\frac{1}{x_1}=\frac{1}{x_2}=\cdots=\frac{1}{x_n}$$

Ainsi, tous les x_i sont égaux : $x_1=x_2=\cdots=x_n$. Notons a cette valeur commune. La contrainte $g(x)=0$ s'écrit :

$$a+a+\cdots+a=n \quad\Rightarrow\quad na=n \quad\Rightarrow\quad a=1$$

Le seul point vérifiant la condition est donc $(1,1,\ldots,1)$. En ce point, on a :

$$f(1,\ldots,1)=\sqrt[n]{1\cdots 1}=1$$

C'est donc la valeur maximale de f sous la contrainte.
On a ainsi montré que pour tout $x_1,\ldots,x_n>0$ vérifiant $x_1+\cdots+x_n=n$:

$$\sqrt[n]{x_1\cdots x_n}\leq 1$$

Passage au cas général. Soient maintenant $y_1,\ldots,y_n>0$ quelconques. Posons :

$$x_i=\frac{ny_i}{\sum_{j=1}^n y_j}$$

Alors on vérifie que $x_1 + \cdots + x_n = n$ et $x_i > 0$. En appliquant l'inégalité obtenue à ces x_i :

$$\sqrt[n]{x_1 \cdots x_n} \leq 1$$

En remplaçant x_i par leur expression :

$$\sqrt[n]{\frac{ny_1}{\sum y_j} \cdot \frac{ny_2}{\sum y_j} \cdots \frac{ny_n}{\sum y_j}} \leq 1$$

$$\sqrt[n]{\frac{n^n y_1 y_2 \cdots y_n}{(\sum y_j)^n}} \leq 1$$

$$\frac{n \sqrt[n]{y_1 y_2 \cdots y_n}}{\sum_{j=1}^n y_j} \leq 1$$

Ce qui équivaut à :

$$\sqrt[n]{y_1 y_2 \cdots y_n} \leq \frac{\sum_{j=1}^n y_j}{n}$$

Étude du cas d'égalité. Analysons maintenant pour quels $y_i > 0$ on obtient l'égalité :

$$\sqrt[n]{y_1 y_2 \cdots y_n} = \frac{y_1 + \cdots + y_n}{n}$$

L'égalité dans l'inégalité finale provient de l'égalité dans l'inégalité intermédiaire $\sqrt[n]{x_1 \cdots x_n} \leq 1$ appliquée aux $x_i = \frac{ny_i}{\sum y_j}$. En effet, on a :

$$\frac{n \sqrt[n]{y_1 \cdots y_n}}{\sum y_j} = 1 \iff \sqrt[n]{x_1 \cdots x_n} = 1$$

Or, d'après l'étude du problème de maximisation, la valeur maximale 1 n'est atteinte que lorsque tous les x_i sont égaux entre eux, c'est-à-dire lorsque $x_1 = x_2 = \cdots = x_n$. En effet, la condition $\nabla f = \lambda \nabla g$ nous a conduit à l'égalité des x_i, et le point $(1, \ldots, 1)$ est l'unique point vérifiant à la fois cette condition et la contrainte.
Ainsi, l'égalité $\sqrt[n]{x_1 \cdots x_n} = 1$ équivaut à $x_1 = x_2 = \cdots = x_n$. En revenant à la définition des x_i en fonction des y_i :

$$\frac{ny_1}{\sum_{j=1}^n y_j} = \frac{ny_2}{\sum_{j=1}^n y_j} = \cdots = \frac{ny_n}{\sum_{j=1}^n y_j}$$

En multipliant chaque égalité par $\frac{\sum y_j}{n} > 0$, on obtient :

$$y_1 = y_2 = \cdots = y_n$$

Réciproquement, si tous les y_i sont égaux à une même valeur $y > 0$, alors :

$$\frac{y_1 + \cdots + y_n}{n} = \frac{ny}{n} = y \quad \text{et} \quad \sqrt[n]{y_1 \cdots y_n} = \sqrt[n]{y^n} = y$$

L'égalité est bien vérifiée.
On peut donc conclure : l'égalité dans l'inégalité arithmético-géométrique a lieu si et seulement si tous les nombres $y_1, \ldots, y_n$ sont égaux.

IV.4 Étude à l'ordre 2

Si $f : I \to \mathbb{R}$ est une fonction définie sur un intervalle ouvert de $\mathbb{R}$, et si $a \in I$ est un point critique de f, on peut déterminer s'il s'agit effectivement d'un extremum par un argument de convexité locale, ce qui, en cas de régularité suffisante, se ramène à l'étude du signe de $f''(a)$. Nous recherchons une condition similaire lorsque f est une fonction définie sur un ouvert de $\mathbb{R}^n$, à l'aide de la matrice hessienne $H_f(a)$.

Proposition IV.15 (CN du second ordre)**.** *Soit $f : \Omega \to \mathbb{R}$ (où Ω est un ouvert de $\mathbb{R}^n$) de classe C^2. Si f admet un minimum local (resp. maximum local) en $a \in \Omega$, alors a est un point critique de f et $H_f(a) \in S_+(\mathbb{R})$ (resp. $-H_f(a) \in S_+(\mathbb{R})$).*

Éléments de preuve. D'après la formule de Taylor-Young, le signe de $f(a+h) - f(a)$ est celui de $h^T H_f(a) h$. □

Théorème IV.16 (CS du second ordre)**.** *Soit $f : \Omega \to \mathbb{R}$ (où Ω est un ouvert de $\mathbb{R}^n$) de classe C^2, et soit $a \in \Omega$. Si a est un point critique de f et si $H_f(a) \in S_{++}(\mathbb{R})$ (resp. $-H_f(a) \in S_{++}(\mathbb{R})$), alors f admet un minimum local strict (resp. maximum local strict) en a.*

Éléments de preuve. Toujours le contrôle du signe par Taylor-Young, en travaillant avec une norme adaptée à $H_f(a)$. □

Remarque IV.17 (Avertissement)**.** *La CN du second ordre n'est pas une CS.*

- *La CS du second ordre n'est pas une CN.*

Exemple IV.18. *Étudier les CN et CS du second ordre pour les fonctions suivantes, en 0.*

1. $f : (x, y) \mapsto (x^2 + y^2)^2$.
2. $g : (x, y) \mapsto x^3 y^3$.

Méthode IV.19 (Étude d'un point critique)**.** *Pour savoir si un point critique correspond à un extremum local, on étudiera le signe de la hessienne,*

- *soit directement par mises sous forme canonique successives de la forme quadratique $h \mapsto h^T H_f(a) h$,*
- *soit par étude des valeurs propres de $H_f(a)$.*

On peut alors conclure :

- *Si $H_f(a)$ n'est pas positive (i.e. s'il existe une valeur propre strictement négative, ou s'il existe h tel que $h^T H_f(a) h < 0$), alors f n'admet pas d'extremum local en a.*
- *Si $H_f(a)$ est strictement positive (si $\mathrm{Spec}(H_f(a)) \subset \mathbb{R}_+^*$), alors f admet un minimum local.*

On ne peut pas conclure de cette façon si ces conditions ou les conditions symétriques (pour un maximum) ne sont pas satisfaites, à savoir dans le cas où $0 \in \mathrm{Spec}(H_f(a))$.

Corollaire IV.20 (Explicitation pour $n = 2$)**.** *Soit Ω un ouvert de $\mathbb{R}^2$, $f : \Omega \to \mathbb{R}$ de classe C^2, et a un point critique de f. Alors :*

- *Si $\det(H_f(a)) > 0$ et $\mathrm{tr}(H_f(a)) > 0$, alors f admet un minimum local strict en a.*
- *Si $\det(H_f(a)) > 0$ et $\mathrm{tr}(H_f(a)) < 0$, alors f admet un maximum local strict en a.*
- *si $\det(H_f(a)) < 0$, alors f n'admet pas d'extremum local en a.*

Exemple IV.21. *Étudier la nature du point critique* $(0,0)$ *dans les 3 cas suivants :*

1. $f_1(x,y) = y^4 + x^2y + x^2 + 2xy + 2y^2$
2. $f_2(x,y) = x^3y + x^2 - 4xy + y^2$
3. $f_3(x,y) = xy^2 + x^3 - x^2 + 2xy - 2y^2$.

Solution Étudions la nature du point critique $(0,0)$ pour les trois fonctions proposées.
1. Étude de $f_1(x,y) = y^4 + x^2y + x^2 + 2xy + 2y^2$
Conditions nécessaires du premier ordre :

$$\frac{\partial f_1}{\partial x} = 2xy + 2x + 2y, \qquad \frac{\partial f_1}{\partial y} = 4y^3 + x^2 + 2x + 4y.$$

En $(0,0)$: $\frac{\partial f_1}{\partial x}(0,0) = 0$, $\frac{\partial f_1}{\partial y}(0,0) = 0$. $(0,0)$ est bien un point critique.
Conditions du second ordre :

$$\frac{\partial^2 f_1}{\partial x^2} = 2y + 2, \quad \frac{\partial^2 f_1}{\partial x \partial y} = 2x + 2, \quad \frac{\partial^2 f_1}{\partial y^2} = 12y^2 + 4.$$

En $(0,0)$:

$$H_{f_1}(0,0) = \begin{pmatrix} 2 & 2 \\ 2 & 4 \end{pmatrix}.$$

$\det(H) = 2 \times 4 - 2 \times 2 = 8 - 4 = 4 > 0$, $\mathrm{tr}(H) = 2 + 4 = 6 > 0$.
D'après le corollaire, f_1 admet un **minimum local strict** en $(0,0)$. La condition suffisante du second ordre est vérifiée.

2. Étude de $f_2(x,y) = x^3y + x^2 - 4xy + y^2$
Conditions nécessaires du premier ordre :

$$\frac{\partial f_2}{\partial x} = 3x^2y + 2x - 4y, \qquad \frac{\partial f_2}{\partial y} = x^3 - 4x + 2y.$$

En $(0,0)$: $\frac{\partial f_2}{\partial x}(0,0) = 0$, $\frac{\partial f_2}{\partial y}(0,0) = 0$. $(0,0)$ est un point critique.
Conditions du second ordre :

$$\frac{\partial^2 f_2}{\partial x^2} = 6xy + 2, \quad \frac{\partial^2 f_2}{\partial x \partial y} = 3x^2 - 4, \quad \frac{\partial^2 f_2}{\partial y^2} = 2.$$

En $(0,0)$:

$$H_{f_2}(0,0) = \begin{pmatrix} 2 & -4 \\ -4 & 2 \end{pmatrix}.$$

$\det(H) = 2 \times 2 - (-4) \times (-4) = 4 - 16 = -12 < 0$.
D'après le corollaire, f_2 **n'admet pas d'extremum local** en $(0,0)$: c'est un point selle (ou col). La condition nécessaire du second ordre (hessienne positive ou négative) n'est pas satisfaite.

3. Étude de $f_3(x,y) = xy^2 + x^3 - x^2 + 2xy - 2y^2$
Conditions nécessaires du premier ordre :

$$\frac{\partial f_3}{\partial x} = y^2 + 3x^2 - 2x + 2y, \qquad \frac{\partial f_3}{\partial y} = 2xy + 2x - 4y.$$

En $(0,0)$: $\frac{\partial f_3}{\partial x}(0,0) = 0$, $\frac{\partial f_3}{\partial y}(0,0) = 0$. $(0,0)$ est un point critique.
Conditions du second ordre :

$$\frac{\partial^2 f_3}{\partial x^2} = 6x - 2, \quad \frac{\partial^2 f_3}{\partial x \partial y} = 2y + 2, \quad \frac{\partial^2 f_3}{\partial y^2} = 2x - 4.$$

En $(0,0)$:

$$H_{f_3}(0,0) = \begin{pmatrix} -2 & 2 \\ 2 & -4 \end{pmatrix}.$$

$\det(H) = (-2) \times (-4) - 2 \times 2 = 8 - 4 = 4 > 0$, $\mathrm{tr}(H) = -2 - 4 = -6 < 0$.
D'après le corollaire, f_3 admet un **maximum local strict** en $(0,0)$. La condition suffisante du second ordre pour un maximum est vérifiée.

- f_1 : minimum local strict en $(0,0)$.
- f_2 : point selle en $(0,0)$ (pas d'extremum).
- f_3 : maximum local strict en $(0,0)$.