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 :

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

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