Chapitre 13

Calcul différentiel et optimisation

Dans tout ce chapitre, E et F sont deux $\mathbb{R}$-espaces vectoriels normés de dimensions finies n et p. Toutes les normes étant équivalentes sur chacun de ces deux espaces, on notera de façon générique $\|\cdot\|$ une norme sur E et sur F, le contexte devant permettre de savoir sur quel espace on se trouve. De même qu'on déjà généralisé la notion de continuité aux fonctions de E vers F, on va généraliser ici la notion de dérivabilité, et son utilisation en *optimisation* (typiquement la recherche d'extrema).

1 Différentielle d'une fonction

Dans toute cette section, $\mathcal{U}$ est un ouvert de E, $f : \mathcal{U} \to F$ et $a \in \mathcal{U}$.

1.1 Dérivation partielle

Définition 1. Soit $v \in E$ un vecteur non nul. On dit que f est dérivable en a selon le vecteur $v \in E$ lorsque l'application $t \mapsto f(a + tv)$ est dérivable en 0. La dérivée en 0 de cette application est alors appelée *dérivée* de f en a selon v, et notée $D_v f(a)$.

Remarques :

- C'est la limite quand $t \to 0$ du taux d'accroissement $\dfrac{1}{t}\big(f(a + tv) - f(a)\big)$.
- On peut l'interpréter comme la dérivée en a de la restriction de f à la droite affine $D = a + \text{vect}(v)$, pour une variable qui "parcourerait" cette droite à une "vitesse" uniforme v.
- Si on modifie cette "vitesse", on modifie la dérivée de façon proportionnelle : pour $\lambda \in \mathbb{R}$ non nul, on a :

$$D_{\lambda v} f(a) = \lambda D_v f(a)$$

Exercice 1. Calculer la dérivée de $f : (x, y) \mapsto x^2 - y^2$ au point $a = (1, 2)$ suivant le vecteur $v = (3, 5)$.

Définition 2. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. On dit que f admet des dérivées partielles en a dans $\mathcal{B}$ lorsque f est dérivable en a selon e_i, pour tout $i \in [\![1, n]\!]$. La dérivée $D_{e_i} f(a)$ est alors notée $\partial_i f(a)$ ou $\dfrac{\partial f}{\partial x_i}(a)$, et s'appelle i-ème dérivée partielle en a (dans $\mathcal{B}$).

Remarques :

- Dans ces notations, l'utilisation de la base $\mathcal{B}$ est implicite. La notation $\dfrac{\partial f}{\partial x_i}(a)$ fait explicitement référence au fait que la variable vectorielle est notée x et que ses cordonnées dans $\mathcal{B}$ sont $(x_1, \dots, x_n)$.
- La base $\mathcal{B}$ étant fixée, on peut identifier la variable x à ses coordonnées $(x_1, \dots, x_n)$ et supposer que f est en fait une fonction à n variables réelles (définie sur un ouvert de $\mathbb{R}^n$). La dérivée partielle $\dfrac{\partial f}{\partial x_i}(a)$ correspond alors précisément à la dérivée de f en $(a_1, \dots, a_n)$ (coordonnées de a) par rapport à la i-ème variable x_i.
- Si f admet des dérivées partielles en tout point de $\mathcal{U}$, on peut alors considérer les n applications *dérivées partielles* $\partial_i f : \mathcal{U} \to F$, qu'on peut aussi noter $\dfrac{\partial f}{\partial x_i}$. On dit que f admet des dérivées partielles sur $\mathcal{U}$.

- On note généralement (x, y) et (x, y, z) plutôt que (x_1, x_2) et (x_1, x_2, x_3) les points de $\mathbb{R}^2$ ou $\mathbb{R}^3$ respectivement, de sorte que la dérivée partielle par rapport à la première variable (par exemple) sera plutôt notée $\dfrac{\partial f}{\partial x}$ ou $\partial_x f$. De façon générale, la notation doit être cohérente avec le nom des coordonnées variables utilisées pour représenter un vecteur dans une base donnée.

Exercice 2. Étudier l'existence et éventuellement la valeur des dérivées partielles en tout point de $\mathbb{R}^2$ des fonctions définies par :

$$f_1(x, y) = \mathrm{e}^x \cos(y), \qquad f_2(x, y) = \sqrt{1 + x^2y^2}, \qquad f_3(x, y) = x^y$$

1.2 Différentiabilité en un point

On rappelle que pour $f : I \to \mathbb{R}$, avec I intervalle de $\mathbb{R}$ et $a \in I$, f est dérivable en a si, et seulement si, f admet un développement limité d'ordre 1 en a. On a alors :

$$f(a + h) = f(a) + f'(a)h + \mathrm{o}(h)$$

L'application $h \mapsto f'(a)h$ donne une approximation linéaire de $f(a + h) - f(a)$ au voisinage de 0. C'est la généralisation de cette idée d'approximation linéaire qui donne lieu à la notion de différentiabilité.

Définition 3. On dit que $f : \mathcal{U} \to F$ est *différentiable* en $a \in \mathcal{U}$ lorsqu'il existe une application linéaire $L : E \to F$ telle que pour $h \in E$ au voisinage de 0 :

$$f(a + h) = f(a) + L(h) + \mathrm{o}(\|h\|)$$

Remarques :
- la notation $\mathrm{o}(\|h\|)$ signifie que l'application $\varphi : h \mapsto f(a + h) - f(a) - L(h)$, définie au voisinage de 0, vérifie $\|\varphi(h)\| \leqslant \varepsilon\|h\|$ pour tout $\varepsilon > 0$, pour $\|h\|$ assez proche de 0, autrement dit que $\dfrac{\|\varphi(h)\|}{\|h\|} \underset{h\to 0}{\longrightarrow} 0$.
- L'écriture de la différentiabilité de f en a ci-dessus signifie exactement que f admet un développement limité en a à l'ordre 1. L'application $L : E \to F$ donne une approximation linéaire de $h \mapsto f(a+h)-f(a)$ au voisinage de 0.

Proposition 1. *En notant $(f_1, \dots, f_p)$ les composantes de f dans une base de F, f est différentiable en a si, et seulement si, toutes les f_i le sont.*

Remarque : Le choix d'une base de F permet ainsi de se ramener au cas des fonctions à valeurs réelles. On sera maintes fois amené à profiter de cette simplification pour écrire divers résultats théoriques.

Proposition 2. *Si f est différentiable en a, alors f est continue en a.*

Remarque : La réciproque est fausse (voir exercice plus loin).

1.3 Différentielle en un point

Proposition 3. *Si f est différentiable en a, avec $f(a + h) = f(a) + L(h) + \mathrm{o}(\|h\|)$, l'application linéaire L est unique.*

Définition 4. Cette application linéaire est appelée *différentielle* de f en a, et est notée $\mathrm{d}f(a)$.

Remarques :
- On notera bien la nature de $\mathrm{d}f(a)$: un élément de $\mathcal{L}(E, F)$. On l'appelle aussi *application linéaire tangente* à f en a.
- En cas de différentiabilité de f en a, il y a donc unicité du développement limité d'ordre 1 en a qui s'écrit, pour h au voisinage de 0 :
$$f(a + h) = f(a) + \mathrm{d}f(a) \cdot h + \mathrm{o}(\|h\|),$$
avec la notation $\mathrm{d}f(a) \cdot h$ pour $\mathrm{d}f(a)(h)$. En posant $x = a + h$, cela revient à écrire pour x au voisinage de a :
$$f(x) = f(a) + \mathrm{d}f(a) \cdot (x - a) + \mathrm{o}(\|x - a\|)$$

- Dans le cas d'une fonction à une variable réelle $f : I \to F$ avec I un intervalle de $\mathbb{R}$, l'écriture d'un développement limité d'ordre 1 en a montre que la différentiabilité de f en a est équivalente à la dérivabilité en a avec $\mathrm{d}f(a) : h \mapsto f'(a)h$ et donc $f'(a) = \mathrm{d}f(a) \cdot 1$.

Exercice 3. Montrer que l'application $f : (x, y) \mapsto x^2 - y^2$ est différentiable en $(a, b) \in \mathbb{R}^2$ et expliciter sa différentielle. Que vaut $\mathrm{d}f(1, 2) \cdot (3, 5)$? Comparer avec la dérivée en $(1, 2)$ selon le vecteur $(3, 5)$.

L'exercice précédent illustre le fait que la différentiabilité en a implique la dérivabilité selon n'importe quel vecteur et qu'une relation simple apparait entre les deux concepts :

Proposition 4. *Si f est différentiable en a, alors pour tout $v \in E$, f est dérivable en a selon v, avec*

$$D_v f(a) = \mathrm{d}f(a) \cdot v$$

Remarque : Attention, La réciproque fausse ! f peut très bien admettre des dérivées en a selon n'importe quel vecteur sans pour autant être différentiable en a.

Exercice 4. Soit $f : (x, y) \mapsto y \sin\left(\dfrac{x}{y}\right)$ définie sur $\mathbb{R}^2 \backslash \mathrm{vect}\{(1, 0)\}$.

a) Montrer que l'on peut prolonger f à $\mathbb{R}^2$ par continuité.

b) Montrer que f admet des dérivées partielles en 0 selon n'importe quel vecteur.

c) f est-elle différentiable en 0 ?

1.4 Différentielle sur un ouvert

Définition 5. On dit que $f : \mathcal{U} \to F$ est *différentiable* sur $\mathcal{U}$ lorsque f est différentiable en tout point de $\mathcal{U}$. L'application $\mathrm{d}f : x \mapsto \mathrm{d}f(x)$ s'appelle alors *différentielle* de f.

Remarques :

- Attention, alors que f est définie sur E à valeurs dans F, $\mathrm{d}f$ est définie sur E mais à valeurs dans $\mathcal{L}(E, F)$.
- Pour une fonction de la variable réelle, on a bien l'équivalence f dérivable si, et seulement si, f différentiable et f' peut s'identifier à $\mathrm{d}f$ car $F \simeq \mathcal{L}(\mathbb{R}, F)$: pour $x \in \mathbb{R}$ on a $f'(x) = \mathrm{d}f(x) \cdot 1$ et $\mathrm{d}f(x) : h \mapsto hf'(x)$.

Exemples :

- Si f est constante, $\mathrm{d}f = 0$
- Si f est linéaire, $\mathrm{d}f$ est constante avec $\mathrm{d}f : x \mapsto f$.

Remarque : Ces exemples généralisent donc une situation bien connue pour les fonctions à une variable !

2 Opérations sur les différentielles et les dérivées partielles

2.1 Différentielle et dérivées partielles, matrice jacobienne

Proposition 5. *On suppose E muni d'une base $\mathcal{B} = (e_1, \ldots, e_n)$. Si f est différentiable sur $\mathcal{U}$, alors f admet des dérivées partielles sur $\mathcal{U}$ dans $\mathcal{B}$. Pour tout $j \in [\![1, n]\!]$, on a :*

$$\partial_j f : x \mapsto \mathrm{d}f(x) \cdot e_j$$

Remarques :

- En notant $(h_1, \ldots, h_n)$ les coordonnées de h dans $\mathcal{B}$, le développement limité d'ordre 1 en a s'écrit :

$$f(a + h) = f(a) + \sum_{j=1}^{n} h_j \partial_j f(a) + \mathrm{o}(\|h\|)$$

- Si on munit également F d'une base, et en notant $(f_1, \ldots, f_p)$ les composantes de f, la matrice colonne $\begin{pmatrix} \partial_j f_1(a) \\ \vdots \\ \partial_j f_p(a) \end{pmatrix}$ représente pour tout $j \in [\![1, n]\!]$ le vecteur $\mathrm{d}f(a) \cdot e_j$ dans cette base de F. La matrice $\big(\partial_j f_i(a)\big)_{i,j}$ représente donc la différentielle $\mathrm{d}f(x)$ en a.

Des bases de E et F étant fixées, on peut en fait se ramener au cas d'une application de $\mathbb{R}^n$ vers $\mathbb{R}^p$.

Définition 6. On suppose ici que $\mathcal{U}$ est un ouvert de $\mathbb{R}^n$ et soit $f : \mathcal{U} \to \mathbb{R}^p$. différentiable en $a \in \mathcal{U}$. On appelle *matrice jacobienne* de f en a la matrice $J_f(a)$ représentative de $\mathrm{d}f(a)$ dans les bases canonique. On a explicitement

$$J_f(a) = \left(\frac{\partial f_i}{\partial x_j}(a)\right)_{i,j} \in \mathcal{M}_{p,n}(\mathbb{R})$$

Remarque : Le développement limité d'ordre 1 en a s'écrit alors matriciellement $f(a+h) = f(a) + J_f(a)h + \mathrm{o}(\|h\|)$, c'est-à-dire :

$$f\left(\begin{pmatrix} a_1 + h_1 \\ \vdots \\ a_n + h_n \end{pmatrix}\right) = f\left(\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}\right) + \begin{pmatrix} \partial_1 f_1(a) & \cdots & \partial_n f_1(a) \\ \vdots & \ddots & \vdots \\ \partial_1 f_p(a) & \cdots & \partial_n f_p(a) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} + \mathrm{o}(\|h\|)$$

Exercice 5. Écrire la jacobienne de l'application $f : (x, y, z) \mapsto (xyz, x^2y + y)$ en tout point de $\mathbb{R}^3$.

2.2 Gradient

Remarque : Pour une fonction $f : \mathcal{U} \to \mathbb{R}$ différentiable en $a \in \mathcal{U}$, avec $\mathcal{U}$ un ouvert de E, la différentielle $\mathrm{d}f(a)$ est une forme linéaire sur E. Elle est représentée par une matrice ligne dans une base $\mathcal{B}$ de E, ce qui donne pour $h \in E$ de composantes $(h_1, \ldots, h_n)$ dans $\mathcal{B}$:

$$\mathrm{d}f(a) \cdot h = \begin{pmatrix} \partial_1 f(a) & \cdots & \partial_n f(a) \end{pmatrix} \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix} = \sum_{j=1}^{n} \partial_j f(a) h_j$$

On a bien envie d'interpréter cela comme un produit scalaire ….

Définition 7. Si E est euclidien et $f : \mathcal{U} \to \mathbb{R}$ est différentiable en $a \in \mathcal{U}$, on appelle *gradient* de f en a l'unique vecteur de E, noté $\nabla f(a)$ tel que

$$\forall h \in E, \quad \mathrm{d}f(a) \cdot h = \left\langle \nabla f(a), h \right\rangle$$

Remarque : Le développement limité de f en a peut alors s'écrire :

$$f(x) = f(a) + \left\langle \nabla f(a), x - a \right\rangle + \mathrm{o}(\|x - a\|)$$

Proposition 6. *Si $\mathcal{B}$ est une base orthonormée de E, les coordonnées de $\nabla f(a)$ dans $\mathcal{B}$ sont les dérivées partielles $\big(\partial_1 f(a), \ldots, \partial_n f(a)\big)$.*

Remarques :

- Comme on a pu le voir dans le chapitre sur les endomorphismes d'un espace euclidien, le résultat précédent ne s'applique plus si la base n'est pas orthonormée.
- Dans le cas où $E = \mathbb{R}^n$, on introduit naturellement la structure euclidienne canonique et le gradient de f en a s'identifie donc directement au vecteur des dérivées partielles $\left(\frac{\partial f}{\partial x_1}(a), \ldots, \frac{\partial f}{\partial x_n}(a)\right)$.

Exercice 6. Montrer que $\nabla f(a)$ est positivement colinéaire au vecteur unitaire selon laquelle la dérivée partielle de f en a est maximale.

2.3 Différentielle et linéarité

Proposition 7. *Si $f_1, \dots, f_q : \mathcal{U} \to F$ sont différentiables sur $\mathcal{U}$, alors pour tous $\lambda_1, \dots, \lambda_q \in \mathbb{R}$, l'application $f = \sum_{k=1}^{q} \lambda_k f_k$ est différentiable sur $\mathcal{U}$ et*

$$\mathrm{d}f = \sum_{k=1}^{q} \lambda_k \mathrm{d}f_k$$

Proposition 8. *Soient $G, F_1, \dots, F_q$ des $\mathbb{R}$ espaces vectoriels normés et $M : F_1 \times \cdots \times F_q \to G$ une application multinéaire. Si $f_1 : \mathcal{U} \to F_1, \dots, f_q : \mathcal{U} \to F_q$ sont différentiables sur $\mathcal{U}$, alors l'application $f = M(f_1, \dots, f_q)$ aussi et*

$$\mathrm{d}f(x) : h \mapsto \sum_{k=1}^{q} M\big(f_1(x), \dots, \mathrm{d}f_k(x) \cdot h, \dots, f_q(x)\big)$$

Exercice 7. Ecrire une formule donnant le gradient $\nabla(fg)$ de deux fonctions $f, g : \mathcal{U} \to \mathbb{R}$ différentiables sur un ouvert $\mathcal{U}$ d'un espace euclidien E.

2.4 Différentielle et composition : règle de la chaîne

Rappel : Pour des fonctions f et g de la variable réelle dérivables on a la formule $(g \circ f)' = f' \times g' \circ f$.

En notant u la variable de f et x celle de g, on peut décider d'identifier x à f, ce qui permet d'interpréter x à la fois comme une variable pour g et comme une fonction de u. La dérivée de g par rapport à x désigne alors g', tandis que la dérivée de g par rapport à u désigne $(g \circ f)'$. La règle de composition ci-dessus peut s'écrire alors :

$$\frac{\mathrm{d}g}{\mathrm{d}u} = \frac{\mathrm{d}g}{\mathrm{d}x}\,\frac{\mathrm{d}x}{\mathrm{d}u}$$

C'est ce qu'on appelle la *règle de la chaîne*, qu'on va généraliser aux fonctions à plusieurs variables.

Noter que dans la règle ci-dessus, si $\frac{\mathrm{d}g}{\mathrm{d}u}$ s'évalue en u_0, alors $\frac{\mathrm{d}x}{\mathrm{d}u}$ aussi, mais $\frac{\mathrm{d}g}{\mathrm{d}x}$ s'évalue en $x(u_0) = f(u_0)$.

Proposition 9. *E, F et G étant trois espaces vectoriels normés de dimension finie, $\mathcal{U}$ un ouvert de E et $\mathcal{V}$ un ouvert de F, on considère $f : \mathcal{U} \to F$ tel que $\mathrm{Im}(f) \subset \mathcal{V}$ et $g : \mathcal{V} \to G$. Si f est différentiable sur $\mathcal{U}$ et g différentiable sur $\mathcal{V}$, alors $g \circ f$ est différentiable sur $\mathcal{U}$ avec pour tout $x \in \mathcal{U}$:*

$$\mathrm{d}(g \circ f)(x) = \mathrm{d}g\big(f(x)\big) \circ \mathrm{d}f(x)$$

Remarques :
- On peut retenir qu'en un point x "la différentielle d'une composée est la composée des différentielles".
- Attention à ne pas écrire $\mathrm{d}(g \circ f) = \mathrm{d}g \circ \mathrm{d}f$, ce qui n'aurait pas de sens car $\mathrm{d}g$ est définie sur $\mathcal{V} \subset F$ tandis que $\mathrm{d}f$ est à valeurs dans $\mathcal{L}(E, F)$. Si on veut "enlever x", on peut écrire $\mathrm{d}(g \circ f) = \big(\mathrm{d}g \circ f(\cdot)\big) \circ \mathrm{d}f$ (où "$\cdot$" est à remplacer par la variable).

Exercice 8.

a) Retrouver la formule de dérivation d'une composée rappelée plus haut.

b) On suppose $\varphi : I \to E$ (avec I intervalle de $\mathbb{R}$) et $f \in \mathcal{L}(E, F)$. Retrouver un résultat déjà vu à propos de la dérivation de $f \circ \varphi$.

c) Plus généralement, pour $\gamma : I \to E$ dérivable à valeurs dans $\mathcal{U}$ et $f : \mathcal{U} \to F$ différentiable, écrire la dérivée de $f \circ \gamma$ à l'aide de γ' et de $\mathrm{d}f$.

Des bases étant fixées, la composition de deux différentielles se traduit par un produit matriciel (de jacobiennes). Cela donne des formules générales concernant les dérivées partielles d'une composée d'applications différentielles.

Proposition 10. *On reprend les notations de la proposition ci-dessus avec $G = \mathbb{R}$, et on suppose fixées des bases de E et de F, de dimensions respectives n et p. Si f et g sont différentiables, on a, pour tout $j \in [\![1, n]\!]$:*

$$\partial_j(g \circ f)(u) = \sum_{i=1}^{p} \partial_i g\big(f(u)\big) \partial_j f_i(u)$$

Remarques :

- Comme dans le rappel plus haut, il est parlant d'écrire ce genre de formules en procédant à une "identification" entre variables et fonctions "intermédiaires". Il s'agit d'un point de vue "à la physicienne" où de façon générale une fonction g désignerait une *quantité* qui dépend de n variables $(x_1, \dots, x_n)$, elles-mêmes dépendant chacune de m variables $(u_1, \dots, u_m)$: g peut donc aussi être vue comme une fonction de $(u_1, \dots, u_m)$:
$$(u_1, \dots, u_m) \mapsto g\big(x_1(u_1, \dots, u_m), \dots, x_n(u_1, \dots, u_m)\big)$$
Lorsque tout est différentiable, on peut écrire, pour tout $j \in [\![1, m]\!]$, et en omettant "l'évaluation" :
$$\frac{\partial g}{\partial u_j} = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} \frac{\partial x_i}{\partial u_j}$$
C'est cela qu'on appelle généralement la *règle de la chaine*, et qui généralise la formule $\dfrac{\mathrm{d}g}{\mathrm{d}u} = \dfrac{\mathrm{d}g}{\mathrm{d}x} \dfrac{\mathrm{d}x}{\mathrm{d}u}$ traduisant la dérivée d'une composition.
- Dans le cas où les n variables $(x_1, \dots, x_n)$ ne dépendent chacune que d'une seule variable t, on retrouve une formule de dérivation pour l'application $t \mapsto g\big(x_1(t), \dots, x_n(t)\big)$ que l'on peut écrire $g \circ \gamma$, avec $\gamma : I \to \mathbb{R}^n$, à savoir :
$$g \circ \gamma)' = \frac{\mathrm{d}g}{\mathrm{d}t} = \sum_{i=1}^{n} \frac{\partial g}{\partial x_i} x_i'$$

Exercice 9. Appliquer la formule des dérivées partielles d'une composition au cas de la transformation dans le plan de coordonnées cartésiennes à polaires, pour une fonction $(x, y) \mapsto g(x, y)$, avec $x : (r, \theta) \mapsto r\cos(\theta)$ et $y : (r, \theta) \mapsto r\sin(\theta)$.

2.5 Fonctions de classe $\mathcal{C}^1$

Définition 8. On dit que $f : \mathcal{U} \to F$ est de classe $\mathcal{C}^1$ lorsque f est différentiable et $\mathrm{d}f$ continue sur $\mathcal{U}$.

Remarques :

- Rappelons une fois de plus que $\mathrm{d}f$ est une application de E à valeurs dans $\mathcal{L}(E, F)$.
- Une base de E étant fixée, f est de classe $\mathcal{C}^1$ si, et seulement si, f est différentiable et toutes les fonctions dérivées partielles $\partial_i f$ sont continues.
- Pour $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$ (ou en se ramenant à cette situation via des bases), f est de classe $\mathcal{C}^1$ si, et seulement si, f est différentiable et la fonction jacobienne $J_f : \mathcal{U} \to \mathcal{M}_{p,n}(\mathbb{R})$ est continue.
- Si E euclidien et pour $f : E \to \mathbb{R}$, f est de classe $\mathcal{C}^1$ si, et seulement si, f est différentiable et la fonction gradient $\nabla f : \mathcal{U} \to E$ est continue.

En fait, il se trouve que l'existence et la continuité des dérivées partielles suffit : cela implique automatiquement la différentiabilité :

Théorème 1. *Une base quelconque de E étant fixée, $f : \mathcal{U} \to F$ est de classe $\mathcal{C}^1$ si, et seulement si, f admet des dérivées partielles continues sur $\mathcal{U}$.*

Remarque : C'est cette caractérisation qu'on utilise en pratique pour montrer qu'une fonction définie via des fonctions usuelles est différentiable !

Pour rendre plus précis l'énoncé suivant, on reprendra les propositions 7, 8 et 9 et on remplacera "différentiable" par "de classe $\mathcal{C}^1$".

Proposition 11.

- *Une combinaison linéaire d'applications de classe $\mathcal{C}^1$ est de classe $\mathcal{C}^1$;*
- *Un produit (ou plus généralement une composition multilinéaire) d'applications de classe $\mathcal{C}^1$ est de classe $\mathcal{C}^1$;*
- *La composition de deux applications de classe $\mathcal{C}^1$ est de classe $\mathcal{C}^1$.*

3 Applications géométriques et optimisation au premier ordre

3.1 Dérivation le long d'un arc

Proposition 12. *Soit I un intervalle de $\mathbb{R}$ et $\gamma : I \to E$ une application. Si γ est dérivable en $t \in I$ et si $f : \mathcal{U} \to F$ est différentiable en $\gamma(t)$, alors $f \circ \gamma$ est dérivable en t et :*

$$(f \circ \gamma)'(t) = \mathrm{d}f\big(\gamma(t)\big) \cdot \gamma'(t)$$

On dit qu'il s'agit de la dérivation de f en $\gamma(t)$ le long de l'arc γ.

Remarques :

- En prenant en particulier $\gamma : t \mapsto a + tv$ avec $a \in \mathcal{U}$ et $v \in E$ non nul, on a $\gamma'(t) = v$ pour tout t et on retrouve notamment en $t = 0$:
$$(f \circ \gamma)'(0) = \mathrm{d}f(a) \cdot v = D_v f(a)$$
- Dans le résultat ci-dessus, $\gamma'(t)$ s'interprète comme un vecteur *tangent* à l'arc γ au point $\gamma(t)$. Dériver f en $\gamma(t)$ le long de l'arc revient à dériver selon ce vecteur tangent. La norme de $\gamma'(t)$ s'interprète comme la vitesse en t le long de ce parcours : les variations de f le long de cet arc seront d'autant plus grandes que $\|\gamma'(t)\|$ est grand.

Lorsque γ est un chemin de E de classe $\mathcal{C}^1$, la formule de la proposition précédente permet d'obtenir la variation globale d'une fonction différentiable entre deux points via l'intégration de la différentielle de f le long de ce chemin, et de généraliser ainsi d'une certaine façon la formule $f(b) - f(a) = \displaystyle\int_a^b f'(x)\mathrm{d}x$.

Proposition 13. *Soit $f : \mathcal{U} \to F$ de classe $\mathcal{C}^1$ et $\gamma : [0,1] \to \mathcal{U}$ de classe $\mathcal{C}^1$ également. Si $\gamma(0) = a$ et $\gamma(1) = b$, on a :*

$$f(b) - f(a) = \int_0^1 \mathrm{d}f\big(\gamma(t)\big) \cdot \gamma'(t)\mathrm{d}t.$$

Remarques :

- Cela s'applique en particulier dans le cas d'un segment $[a, b] \subset \mathcal{U}$: en notant $v = b - a$ et $\gamma : t \mapsto a + tv$, on obtient
$$f(b) - f(a) = \int_0^1 \mathrm{d}f(a + tv) \cdot v = \int_0^1 D_v f\big(a(1-t) + tb\big)\mathrm{d}t$$
- Si de plus f est à valeurs dans $\mathbb{R}$ et E euclidien, on peut réécrire cela :
$$f(b) - f(a) = \left\langle \int_0^1 \nabla f(a(1-t) + tb)\mathrm{d}t \ , \ b - a \right\rangle$$

3.2 Caractérisation des fonctions constantes

Il est clair que la différentielle d'une application constante est nulle. La réciproque est vraie pour des fonctions dérivables sur un intervalle de $\mathbb{R}$. Cela se généralise dans le cas de fonctions définies sur un ouvert connexe par arcs.

Proposition 14. *Si $\mathcal{U}$ est un ouvert connexe par arcs de E et si $f : \mathcal{U} \to F$ est de classe $\mathcal{C}^1$, alors f est constante si, et seulement si, $\mathrm{d}f$ est nulle.*

Remarques :

- On prouve facilement ce résultat à l'aide de celui du paragraphe précédent pour une connexité par arcs de classe $\mathcal{C}^1$ ou même $\mathcal{C}^1$ par morceaux, et en particulier dans le cas où $\mathcal{U}$ est convexe. La démonstration dans le cas général est admise.
- Si $\mathcal{U}$ est consituée de plusieurs composantes connexes par arcs, $\mathrm{d}f = 0$ implique que f est constante sur chacune de ces composantes, mais il n'y a aucune raison que ces constantes soient les mêmes d'une composante à l'autre.

3.3 Points critiques et extrema

Définition 9. Soit $f : \mathcal{U} \to F$ différentiable. On dit que $a \in \mathcal{U}$ est un *point critique* de f lorsque $\mathrm{d}f(a) = 0$.

Remarques :
- Si on travaille dans $\mathbb{R}^n$ ou si une base est fixée, un point critique est donc un point pour lequel toutes les dérivées partielles s'annulent.
- Si a est un point critique de f, les variations de $f(x)$ pour x au voisinage de a sont négligeables devant $x - a$:
$$f(x) = f(a) + \mathrm{o}(\|x - a)\|$$
On peut dire que f est localement constante à l'ordre 1.

On peut maintenant généraliser une propriété bien connue des fonctions à une variable réelle.

Proposition 15. *Si $f : \mathcal{U} \to \mathbb{R}$ est différentiable et admet un extremum local en $a \in \mathcal{U}$, alors a est un point critique de f.*

Remarques :
- On savait déjà la réciproque fausse pour une fonction à une seule variable (penser à $x \mapsto x^3$ en 0). Mais il s'agit de cas "pathologiques" correspondant à une dérivée qui s'annule sans changer de signe.
- Dans le cas $\dim(E) > 1$, la fausseté de la réciproque devient "courante", et ce d'autant plus que $\dim(E)$ est grand : la restriction locale de f à deux droites passant par a peut admettre un minimum pour l'une et un maximum pour l'autre (point de type *selle*).

Exemples :
- $f : (x, y) \mapsto x^2 + y^2$ admet un minimum local en $(0, 0)$, donc un point critique ;
- $g : (x, y) \mapsto x^2 - y^2$ n'admet pas d'extremum local en $(0, 0)$, qui est néanmoins un point critique.

Remarques :
- Le résultat précédent reste tout de même utile pour déterminer les extrema locaux ou globaux sur un ouvert, s'il en existe : lorsque f est différentiable, on sait qu'il est inutile de chercher en dehors des points critiques !
- Dans le cas où $f : \mathcal{D} \to \mathbb{R}$ n'est pas définie sur un ouvert, le résultat précédent ne s'applique que pour les points intérieurs à $\mathcal{D}$ (tant que f est différentiable sur $\mathcal{U} = \mathring{\mathcal{D}}$) et il faut donc examiner à part les points de la frontière.
- En particulier, il peut être utile de se restreindre à un compact K pour pouvoir justifier l'existence d'extremums globaux à l'aide du théorème des bornes atteintes. On pourra alors les chercher parmi les points critiques de l'intérieur de K ou sur sa frontière.

Exercice 10. Dans chacun des cas suivants, démontrer que f admet un maximum global sur K et le déterminer.

a) $f : (x, y) \mapsto xy(1 - x - y)$ et $K = \{(x, y) \in (\mathbb{R}^+)^2 \mid x + y \leqslant 1\}$.

b) $f : (x, y) \mapsto x - y + x^3 + y^3$ et $K = [0, 1]^2$;

c) $f : (x, y) \mapsto \sin(x) \sin(y) \sin(x + y)$ et $K = [0, \frac{\pi}{2}]^2$.

4 Classe $\mathcal{C}^k$ et optimisation au second ordre

Dans toute cette section, on se place dans le cas $E = \mathbb{R}^n$, muni de sa base canonique, et $F = \mathbb{R}^p$ (on peut toujours se ramener au cas $F = \mathbb{R}$ en raisonnant composante par composante). Comme précédemment, $\mathcal{U}$ est un ouvert de E.

4.1 Dérivées partielles d'ordre supérieur et classe $\mathcal{C}^k$

Si $f : \mathcal{U} \to F$ admet des dérivées partielles en tout point de $\mathcal{U}$, on dispose de n nouvelles fonctions $\partial_1 f, \ldots, \partial_n f$, définies sur $\mathcal{U}$, qui admettent peut-être elles-même des dérivées partielles : ce sont dans ce cas des dérivées partielles secondes de f. Ce procédé peut être réitéré.

Définition 10. Soit $f : \mathcal{U} \to F$, $a \in \mathcal{U}$ et $(j_1, \ldots, j_k) \in [\![1, n]\!]^n$. On dit que f admet une *dérivée partielle k-ième* en a par rapport aux variables $x_{j_1}, \ldots, x_{j_k}$ successivement lorsque :

- $\dfrac{\partial f}{\partial x_{j_1}}, \dfrac{\partial}{\partial x_{j_2}}\left(\dfrac{\partial f}{\partial x_{j_1}}\right), \ldots, \dfrac{\partial}{\partial x_{j_{k-1}}}\left(\dfrac{\partial}{\partial x_{j_{k-2}}}\left(\cdots\left(\dfrac{\partial f}{\partial x_{j_1}}\right)\cdots\right)\right)$ existent sur un voisinage de a ;
- $\dfrac{\partial}{\partial x_{j_k}}\left(\dfrac{\partial}{\partial x_{j_{k-1}}}\left(\cdots\left(\dfrac{\partial f}{\partial x_{j_1}}\right)\cdots\right)\right)(a)$ existe

Ce dernier élément est appelé *dérivée partielle k-ième* en a, par rapport aux variables $x_{j_1}, \ldots, x_{j_k}$ successivement.

Notation :

- $\dfrac{\partial}{\partial x_{j_k}}\left(\dfrac{\partial}{\partial x_{j_{k-1}}}\left(\cdots\left(\dfrac{\partial f}{\partial x_{j_1}}\right)\cdots\right)\right)(a)$ sera plutôt notée $\dfrac{\partial^k f}{\partial x_{j_k}\partial x_{j_{k-1}}\cdots\partial x_{j_1}}(a)$, ou encore $\partial_{j_k}\cdots\partial_{j_1} f(a)$ (on n'explicite ainsi pas le nom des variables réelles) ou même encore plus simplement $\partial_{j_k,\ldots,j_1} f(a)$.
- Lorsqu'il y a répétion d'une même variable pour des dérivations successsives, on peut utiliser un exposant pour la première notation. On pourra noter par exemple $\dfrac{\partial^3 f}{\partial x_k \partial x_j^2}(a)$ au lieu de $\dfrac{\partial^3 f}{\partial x_k \partial x_j \partial x_j}(a)$

Définition 11. On dit que $f : \mathcal{U} \to F$ est de classe $\mathcal{C}^k$ lorsqu'elle admet des dérivées partielles successives en tout point de $\mathcal{U}$, jusqu'à l'ordre k inclus et par rapport à toutes les variables, et lorsque toutes ces dérivées partielles sont continues sur $\mathcal{U}$.

Remarque : On peut démontrer que f est de classe $\mathcal{C}^k$ si, et seulement si, f est de classe $\mathcal{C}^1$ et les fonctions dérivées partielles de classe $\mathcal{C}^{k-1}$ sur $\mathcal{U}$. Cela donne une définition alternative récursive.

4.2 Théorème de Schwarz

Théorème 2. *Si $f : \mathcal{U} \to F$ est de classe $\mathcal{C}^2$, alors pour tout $a \in \mathcal{U}$ on a :*

$$\forall (j,k) \in [\![1, n]\!]^n, \quad \frac{\partial^2 f}{\partial x_j \partial x_k}(a) = \frac{\partial^2 f}{\partial x_k \partial x_j}(a)$$

Remarque : En particulier pour une fonction de classe $\mathcal{C}^2$ à deux variables (x, y) (cas $E = \mathbb{R}^2$), on a

$$\frac{\partial^2 f}{\partial x \partial y}(a) = \frac{\partial^2 f}{\partial y \partial x}(a)$$

Corollaire 1. *Si $f : \mathcal{U} \to F$ est de classe $\mathcal{C}^k$, les dérivées partielles de f jusqu'à l'ordre k ne dépendent pas de l'ordre de dérivation. Plus précisément, pour tout $m \in [\![1, k]\!]$ et pour toute permutation $\sigma \in \mathcal{S}_m$:*

$$\frac{\partial^m f}{\partial x_{j_1}\partial x_{j_2}\cdots\partial x_{j_m}} = \frac{\partial^m f}{\partial x_{j_{\sigma(1)}}\partial x_{j_{\sigma(2)}}\cdots\partial x_{j_{\sigma(m)}}}$$

4.3 Classe $\mathcal{C}^k$ et opérations

Proposition 16. *Si $f, g : \mathcal{U} \to F$ sont de classe $\mathcal{C}^k$, alors :*

- *Pour tout $\lambda, \mu \in \mathbb{R}$, la fonction $\lambda f + \mu g$ est de classe $\mathcal{C}^k$.*
- *Si $F = \mathbb{R}$, la fonction fg est de classe $\mathcal{C}^k$.*

Remarques :

- L'ensemble $\mathcal{C}^k(\mathcal{U}, \mathbb{R})$ des fonctions de classe $\mathcal{C}^k$ sur $\mathcal{U}$ et à valeurs dans $\mathbb{R}$ est donc une $\mathbb{R}$-algèbre commutative.
- Comme dans le cas de la différentiabilité, on a plus généralement qu'une composition multilinéaire d'applications de classe $\mathcal{C}^k$ est de classe $\mathcal{C}^k$.

Proposition 17. *Si $f : \mathcal{U} \to F$ et $g : U \to G$, avec U ouvert de F contenant $f(\mathcal{U})$, sont de classe $\mathcal{C}^k$, alors $g \circ f$ est de classe $\mathcal{C}^k$.*

4.4 Exemples d'équations aux dérivées partielles

Voir TD MPI n°5

4.5 Matrice hessienne d'une application de classe $\mathcal{C}^2$

Définition 12. Soit $f : \mathcal{U} \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in \mathcal{U}$. On appelle *matrice hessienne* de f en a la matrice

$$H_f(a) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a)\right)_{1 \leqslant i,j \leqslant n}$$

Remarque : Ne pas confondre matrice jacobienne et matrice hessienne : la matrice jacobienne donne les dérivées partielles d'ordre 1 de $f : \mathcal{U} \to F$. Pour $E = \mathbb{R}^n$ et $F = \mathbb{R}^p$, c'est une matrice de $\mathcal{M}_{p,n}(\mathbb{R})$. La matrice hessienne donne les dérivées partielles d'ordre 2 de $f : \mathcal{U} \to \mathbb{R}$. Pour $E = \mathbb{R}^n$ c'est une matrice carrée de $M_n(\mathbb{R})$. Si $F = \mathbb{R}^p$, on peut introduire p matrices hessiennes, pour chacune des p composantes de f.

Le résultat suivant est une conséquence immédiate du théorème de Schwarz.

Proposition 18. *Pour $f : \mathcal{U} \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in \mathcal{U}$, la matrice $H_f(a)$ est symétrique.*

4.6 Formule de Taylor-Young à l'ordre 2

Proposition 19. *Soit $f : \mathcal{U} \to \mathbb{R}$ de classe $\mathcal{C}^2$. Alors pour $a \in \mathcal{U}$, et $h \in E$ au voisinage de 0, on a :*

$$\begin{aligned} f(a+h) &= f(a) + \langle \nabla f(a) \mid h \rangle + \frac{1}{2}\langle H_f(a) \cdot h \mid h \rangle + \mathrm{o}\big(\|h\|^2\big) \\ &= f(a) + \nabla f(a)^T h + \frac{1}{2} h^T H_f(a) h + \mathrm{o}\big(\|h\|^2\big) \end{aligned}$$

Remarques :

- De façon explicite par rapport aux dérivées partielles, et en tenant compte de la symétrie de la hessienne, cela s'écrit :

$$f(a+h) = f(a) + \underbrace{\sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(a) h_i}_{\text{ordre } 1} + \underbrace{\frac{1}{2}\sum_{i=1}^{n} \frac{\partial^2 f}{\partial {x_i}^2}(a) h_i^2 + \sum_{1 \leqslant i < j \leqslant n} \frac{\partial^2 f}{\partial x_i \partial x_j}(a) h_i h_j}_{\text{ordre } 2} + o\big(\|h\|^2\big)$$

- En se plaçant dans une base orthonormée de vecteurs propres, la matrice matrice hessienne (relativement à cette nouvelle base) devient diagonale, les coefficients diagonaux $\lambda_1, \dots, \lambda_n$ étant les valeurs propres. En notant $(x'_1, \dots, x'_n)$ les nouvelles variables associées à ce changement de base, et $(h'_1, \dots, h'_n)$ les composantes de $h \in \mathbb{R}^n$ au voisinage de 0, développement limité d'ordre 2 se réécrit alors :

$$f(a+h) = f(a) + \sum_{i=1}^{n} \frac{\partial f}{\partial x'_i}(a) h'_i + \frac{1}{2}\sum_{i=1}^{n} \lambda_i h_i'^2 + o\big(\|h\|^2\big),$$

avec $\lambda_i = \dfrac{\partial^2 f}{\partial {x'_i}^2}(a)$.

4.7 Étude des points critiques à l'ordre 2

Rappel : Soit $f : I \to \mathbb{R}$ de classe $\mathcal{C}^2$ avec I un intervalle ouvert de $\mathbb{R}$ et soit $a \in I$. Alors :

- si f admet un minimum local en a, on a $f'(a) = 0$ et $f''(a) \geqslant 0$.
- Si f admet un maximum local en a, on a $f'(a) = 0$ et $f''(a) \leqslant 0$.

Les réciproques des deux énoncés ci-dessus sont fausses car tout est possible dans le cas $f''(a) = 0$ (penser à $x \mapsto x^3$ et $x \mapsto \pm x^4$ en 0). Cependant, elle deviennent vraies pour une dérivée seconde non nulle :

- Si $f'(a) = 0$ et $f''(a) > 0$, f admet un minimum local (strict) en a.
- Si $f'(a) = 0$ et $f''(a) < 0$, f admet un maximum local (strict) en a.

Généralisons tout cela ...

Proposition 20. *Soit $f : \mathcal{U} \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in \mathcal{U}$. Si f admet un minimum local en a, alors a est un point critique de f et $H_f(a)$ est une matrice symétrique positive : $H \in \mathcal{S}_n^+(\mathbb{R})$.*

Remarques :

- Ce résultat devient clair lorsqu'on se place dans une base orthonormée de vecteur propres, car le DL d'ordre 2 s'écrit alors (on reprend les notations de la dernière remarque du paragraphe précédent) :
$$f(a+h) - f(a) = \sum_{i=1}^{n} \lambda_i h_i'^2 + o\big(\|h\|^2\big),$$
avec $\lambda_i \geqslant 0$ pour tout $i \in [\![1, n]\!]$.
- Ce résultat s'adapte au cas d'un maximum local : $H_f(a)$ est alors une matrice symétrique *négative*, autrement dit $-H_f(a) \in \mathcal{S}_n^+(\mathbb{R})$ (les valeurs propres sont alors négatives ou nulles).
- La réciproque est fausse : la présence d'une valeur propre nulle rend possible l'existence d'un vecteur $v \in E$ non nul (un vecteur propre associé) tel que l'application $t \mapsto f(a+tv)$, bien que de dérivée nulle en 0, n'admette pas de minimum local en 0 : on pourrait avoir, par exemple, $f(a+tv) = t^3$ ou encore $f(a+tv) = -t^4$.

Comme dans le cas d'une fonction à une seule variable, la réciproque devient vraie lorsque la hessienne est strictement positive :

Proposition 21. *Soit $f : \mathcal{U} \to \mathbb{R}$ de classe $\mathcal{C}^2$ et $a \in \mathcal{U}$. Si a est un point critique de f et si $H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R})$, alors f admet un minimum local (strict) en a.*

Remarques :

- Ce résultat s'adapte dans le cas $-H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R})$ (matrice hessienne *strictement négative*) : f admet alors un maximum local (strict) en a.
- Dans le cas $n = 2$ la nature d'un point critique a pour f peut être explicitée à l'aide de la trace et du déterminant de la hessienne :
  — Si $\det\big(H_f(a)\big) > 0$ et $\operatorname{tr}\big(H_f(a)\big) > 0$, alors f admet un minimum local en a
  — Si $\det\big(H_f(a)\big) > 0$ et $\operatorname{tr}\big(H_f(a)\big) < 0$, alors f admet un maximum local en a ;
  — Si $\det\big(H_f(a)\big) < 0$, alors a est un point de type *selle*.
  Les fonctions $x \mapsto x^2 + y^2$, $x \mapsto -x^2 - y^2$ et $x \mapsto x^2 - y^2$ illustrent respectivement en $(0, 0)$ chacun des trois cas précédents.

Exercice 11. Déterminer les extrema de la fonction $f : (x, y) \mapsto x^3 + 3xy^2 - 15x - 12y$.