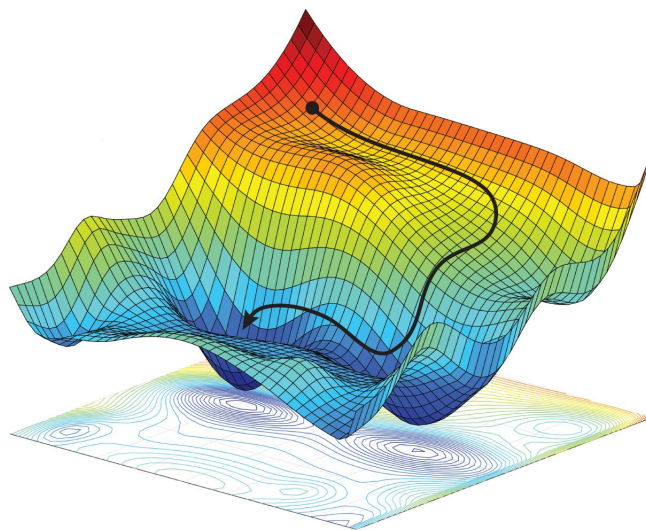

Calcul Différentiel et Optimisation

Cours de L2, 2023-2024
Université Paris-Dauphine

David Gontier & Isabelle Catto

(version du 3 mars 2024).



Calcul Différentiel et Optimisation de **David Gontier & Isabelle Catto** est mis à disposition selon les termes de la **licence Creative Commons Attribution - Pas d'Utilisation Commerciale - Partage dans les Mêmes Conditions 4.0 International**.

1	Topologie dans les espaces vectoriels normés	6
1.1	Premiers rappels et définitions	6
1.1.1	Rappels sur les espaces vectoriels normés	6
1.1.2	Rappels sur les suites	7
1.2	Topologie dans les espaces vectoriels normés	9
1.2.1	Définitions	9
1.2.2	Convergence de suites avec des ouverts/fermés	10
1.3	L'espace \mathbb{R}^d	11
1.3.1	Convergence composante par composante	11
1.3.2	Théorèmes de Bolzano Weierstrass	12
2	Optimisation : existence et unicité	13
2.1	Fonctions continues	13
2.1.1	Continuité des fonctions usuelles	14
2.1.2	Prolongement par continuité	14
2.1.3	Un exemple important	14
2.1.4	Fonctions à m composantes	15
2.2	Optimisation, premières définitions	16
2.2.1	Premières définitions	16
2.2.2	Toutes les normes de \mathbb{R}^d sont équivalentes	18
2.3	Optimisation de fonctions continues, existence	18
2.3.1	Existence de minimiseurs dans le cas fermé borné	19
2.3.2	Existence de minimiseurs pour les fonctions continues coercives	19
2.3.3	Unicité des minimiseurs pour les fonctions strictement convexes	20
3	Fonctions différentiables	22
3.1	Applications linéaires	22
3.1.1	L'espace vectoriel des applications linéaires bornées	22
3.1.2	Le cas de la dimension finie	23
3.2	Fonctions différentiables	25
3.2.1	Retour en dimension $d = 1$	25
3.2.2	Définition de la différentiabilité	25
3.2.3	Dérivées directionnelles et dérivées partielles	27
3.2.4	Matrice Jacobienne	27
3.2.5	Le gradient	29
3.2.6	Règle de la chaîne	29
3.3	Fonctions continûment différentiables	30

3.3.1	Théorème «fondamental» de l'analyse	30
3.3.2	Théorème des accroissements finis	31
3.3.3	Caractérisation des fonctions de classe C^1	32
4	Fonctions de classe C^2	33
4.1	Fonctions deux fois différentiables	33
4.1.1	Premières définitions	33
4.1.2	Dérivées croisées	34
4.1.3	Fonctions de classe C^2	34
4.1.4	La Hessienne	35
4.1.5	Formule de Taylor-Young à l'ordre 2	35
4.2	Application à l'optimisation de fonctions	36
4.2.1	Rappels sur les matrices symétriques	36
4.2.2	Caractérisation des points critiques non dégénérés	38
4.2.3	Un exemple de recherche de minimiseur	39
4.2.4	Un autre exemple type	40

Ces notes de cours ont été écrites dans le cadre du cours *Calcul différentiel et optimisation*¹ (L2, Université Paris-Dauphine).

Le but de ce cours est l'étude de fonctions à plusieurs variables, de type $f(x_1, \dots, x_d)$, et en particulier leur optimisation. Nous définissons la notion de continuité et de différentiabilité pour ce genre de fonctions, et montrons que les optimiseurs de f sont des points critiques. Nous utilisons cette propriété pour trouver les optimiseurs de f .

L'espace \mathbb{R}^d

Dans la suite, nous nous intéressons aux fonctions à plusieurs variables. Nous noterons d le nombre de variables (d pour *dimension*). Une fonction $f(x_1, \dots, x_d)$ peut alors être vue comme une fonction $f(\mathbf{x})$ avec $\mathbf{x} \in \mathbb{R}^d$, avec la convention

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d.$$

Nous notons toujours un vecteur $\mathbf{x} \in \mathbb{R}^d$ comme un vecteur colonne (c'est la convention). On notera aussi $\mathbf{x} = (x_1, \dots, x_d)^T \in \mathbb{R}^d$, où le T signifie "transpose".

1. Page web du cours [ici](#).

CHAPITRE 1

TOPOLOGIE DANS LES ESPACES VECTORIELS NORMÉS

1.1 Premiers rappels et définitions

Dans ce chapitre, nous introduisons des concepts de base de topologie, définissons les fonctions continues, et donnons des critères simples d'existence d'optimiseurs pour les fonctions continues.

1.1.1 Rappels sur les espaces vectoriels normés

Pour commencer, on se place dans un cadre général, et on s'intéresse aux fonctions définies de E dans F , où E et F sont des **espaces vectoriels normés**. Dans la suite, on prendra $E = \mathbb{R}^d$ et $F = \mathbb{R}^m$, mais nous préférons énoncer les premières propriétés et définitions dans un cadre plus abstrait.

On rappelle qu'un ensemble E est un \mathbb{R} -espace vectoriel s'il est muni :

- d'un **élément neutre**, noté $\mathbf{0} \in E$,
- d'une **addition** de $E \times E$ dans E , notée $+$: $\mathbf{x}, \mathbf{y} \in E \mapsto \mathbf{x} + \mathbf{y}$,
- d'une **multiplication par un scalaire** $\mathbb{R} \times E$ dans E , notée \cdot : $\lambda, \mathbf{x} \mapsto \lambda \cdot \mathbf{x}$ (ou $\lambda \mathbf{x}$)

qui vérifient les règles usuelles (que nous ne rappelons pas).

Définition 1.1: Normes

Une **norme** sur E est une application $\mathcal{N} : E \rightarrow \mathbb{R}_+$ telle que

- Pour tout $\mathbf{x} \in E$, on a $\mathcal{N}(\mathbf{x}) = 0$ si et seulement si $\mathbf{x} = \mathbf{0}$,
- (homogénéité) Pour tout $\mathbf{x} \in E$ et tout $\lambda \in \mathbb{R}$, on a $\mathcal{N}(\lambda \mathbf{x}) = |\lambda| \mathcal{N}(\mathbf{x})$,
- (inégalité triangulaire) $\mathcal{N}(\mathbf{x} + \mathbf{y}) \leq \mathcal{N}(\mathbf{x}) + \mathcal{N}(\mathbf{y})$.

En prenant $\mathbf{z} = \mathbf{x} + \mathbf{y}$ dans l'inégalité triangulaire, on a $\mathcal{N}(\mathbf{z}) \leq \mathcal{N}(\mathbf{x}) + \mathcal{N}(\mathbf{z} - \mathbf{x})$, donc $\mathcal{N}(\mathbf{z}) - \mathcal{N}(\mathbf{x}) \leq \mathcal{N}(\mathbf{z} - \mathbf{x})$. Quitte à changer $\mathbf{x} \longleftrightarrow \mathbf{z}$, on a aussi $\mathcal{N}(\mathbf{x}) - \mathcal{N}(\mathbf{z}) \leq \mathcal{N}(\mathbf{z} - \mathbf{x})$, et donc

$$\forall \mathbf{x}, \mathbf{y} \in E, \quad |\mathcal{N}(\mathbf{x}) - \mathcal{N}(\mathbf{y})| \leq \mathcal{N}(\mathbf{x} - \mathbf{y}). \quad (\text{inégalité triangulaire inverse}).$$

Un espace vectoriel E muni d'une norme \mathcal{N} est appelé un **espace vectoriel normé**. On le note (E, \mathcal{N}) pour préciser avec quelle norme nous travaillons. Des normes différentes pourraient induire des résultats différents, comme nous le verrons. Un concept important est celui de *normes équivalentes*.

Définition 1.2: Normes équivalentes

On dit que deux normes \mathcal{N}_1 et \mathcal{N}_2 sur E sont **équivalentes** s'il existe $\alpha > 0$ telle que

$$\forall \mathbf{x} \in E, \quad \mathcal{N}_1(\mathbf{x}) \leq \alpha \mathcal{N}_2(\mathbf{x}), \quad \text{et} \quad \mathcal{N}_2(\mathbf{x}) \leq \alpha \mathcal{N}_1(\mathbf{x}).$$

Proposition 1.3. *Si \mathcal{N}_1 et \mathcal{N}_2 sont équivalentes, et si \mathcal{N}_2 et \mathcal{N}_3 sont équivalentes, alors \mathcal{N}_1 et \mathcal{N}_3 sont équivalentes.*

On dit que être équivalent est une relation d'équivalence...

Démonstration. Soit $\alpha > 0$ et $\beta > 0$ tel que, pour tout $\mathbf{x} \in E$, on a

$$\mathcal{N}_1(\mathbf{x}) \leq \alpha \mathcal{N}_2(\mathbf{x}), \quad \mathcal{N}_2(\mathbf{x}) \leq \alpha \mathcal{N}_1(\mathbf{x}), \quad \text{et} \quad \mathcal{N}_2(\mathbf{x}) \leq \beta \mathcal{N}_3(\mathbf{x}), \quad \mathcal{N}_3(\mathbf{x}) \leq \beta \mathcal{N}_2(\mathbf{x}).$$

Alors, on a, pour tout $\mathbf{x} \in E$,

$$\mathcal{N}_1(\mathbf{x}) \leq \alpha \mathcal{N}_2(\mathbf{x}) \leq \alpha \beta \mathcal{N}_3(\mathbf{x}), \quad \text{et} \quad \mathcal{N}_3(\mathbf{x}) \leq \beta \mathcal{N}_2(\mathbf{x}) \leq \beta \alpha \mathcal{N}_1(\mathbf{x}),$$

ce qui montre que \mathcal{N}_1 et \mathcal{N}_3 sont équivalentes, en prenant la constante $\gamma := \alpha\beta > 0$. \square

Il est d'usage de noter $\|\cdot\| =: \mathcal{N}$ les normes. Nous adopterons les deux notations dans la suite du cours.

1.1.2 Rappels sur les suites

Soit $(E, \|\cdot\|)$ un \mathbb{R} -espace vectoriel normé. Une suite de E est une application $\mathbb{N} \ni n \mapsto \mathbf{x}_n \in E$. Dans la suite, on écrira $(\mathbf{x}_n)_{n \in \mathbb{N}} \in E$ ou simplement $(\mathbf{x}_n) \in E$ pour parler d'une suite de E .

Convergence de suites

Définition 1.4: Convergence

Soit (\mathbf{x}_n) une suite de E , et soit $\mathbf{x}_* \in E$. On dit qu'une suite (\mathbf{x}_n) converge vers \mathbf{x}_* si

$$\lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{x}_*\| = 0.$$

Cela signifie que pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ telle que

$$\forall n \geq N, \quad \|\mathbf{x}_n - \mathbf{x}_*\| < \varepsilon.$$

On écrira parfois simplement $\mathbf{x}_n \rightarrow \mathbf{x}_*$ ou $\lim_{n \rightarrow \infty} \mathbf{x}_n = \mathbf{x}_*$. On fera attention cependant que ces écritures sont trompeuses, car elle n'indique pas avec quelle norme nous travaillons !

Exercice 1.5

Soit $E := C^0([0, 1])$ l'ensemble des fonctions continues de $[0, 1]$ dans \mathbb{R} . Soit $f_n(x) := x^n$.

On note $\|f\|_1 := \int_0^1 f(s) ds$, et $\|f\|_\infty := \max\{f(x), x \in [0, 1]\}$.

a/ Montrer que $\|\cdot\|_1$ et $\|\cdot\|_\infty$ sont des normes sur E .

b/ Montrer que (f_n) converge vers $f_*(x) = 0$ pour la norme $\|\cdot\|_1$, mais pas pour la norme $\|\cdot\|_\infty$.

Cependant, lorsque les normes sont équivalentes, les convergences sont les mêmes, comme le montre le résultat suivant.

Théorème 1.6

Soit \mathcal{N}_1 et \mathcal{N}_2 deux normes équivalentes sur E . Alors, pour toute suite (\mathbf{x}_n) de E et tout $\mathbf{x}_* \in E$, la suite (\mathbf{x}_n) converge vers \mathbf{x}_* pour la norme \mathcal{N}_1 ssi elle converge vers \mathbf{x}_* pour la norme \mathcal{N}_2 .

Démonstration. Si (\mathbf{x}_n) converge vers \mathbf{x}_* pour la norme \mathcal{N}_1 , on a $\mathcal{N}_1(\mathbf{x}_n - \mathbf{x}_*) \rightarrow 0$. Donc

$$0 \leq \mathcal{N}_2(\mathbf{x}_n - \mathbf{x}_*) \leq \alpha \mathcal{N}_1(\mathbf{x}_n - \mathbf{x}_*) \xrightarrow[n \rightarrow \infty]{} 0,$$

donc $\mathcal{N}_2(\mathbf{x}_n - \mathbf{x}_*) \rightarrow 0$, et (\mathbf{x}_n) converge vers \mathbf{x}_* pour la norme \mathcal{N}_2 . La réciproque est obtenue en échangeant les rôles de \mathcal{N}_1 et \mathcal{N}_2 . \square

On dit que les normes équivalentes définissent la même **topologie**, car elles induisent les mêmes notions de convergence.

Suites bornées

Définition 1.7

On dit qu'une suite (\mathbf{x}_n) de E est bornée s'il existe $M \in \mathbb{R}_+$ tel que, pour tout $n \in \mathbb{N}$, on a $\|\mathbf{x}_n\| \leq M$.

On rappelle le résultat classique suivant.

Théorème 1.8: Les suites convergentes sont bornées

Si (\mathbf{x}_n) est une suite convergente, alors (\mathbf{x}_n) est une suite bornée.

Démonstration. Soit \mathbf{x}_* la limite de la suite. Pour $\varepsilon = 1$, il existe $N \in \mathbb{N}$ tel que $\|\mathbf{x}_n - \mathbf{x}_*\| < 1$ pour tout $n \geq N$. En particulier, pour $n \geq N$, on a $\|\mathbf{x}_n\| \leq \|\mathbf{x}_*\| + \|\mathbf{x}_* - \mathbf{x}_n\| \leq \|\mathbf{x}_*\| + 1 =: R_1$. Les termes \mathbf{x}_n avec $n \geq N$ sont donc bornés par R_1 . Par ailleurs, on pose $R_2 := \max_{1 \leq i \leq N} \|\mathbf{x}_i\|$. C'est le maximum d'un nombre fini de valeurs, donc $R_2 < \infty$. Les N premiers termes de la suite sont bornés par R_2 . Au final, tous les termes de la suite sont bornés par $R := \max\{R_1, R_2\}$. \square

Sous-suites

Définition 1.9: Sous-suite

Une **extraction** est une fonction $\phi : \mathbb{N} \rightarrow \mathbb{N}$ qui est strictement croissante (donc injective). Une suite (\mathbf{y}_n) est une **sous suite** de (\mathbf{x}_n) si il existe une extraction ϕ telle que $\mathbf{y}_n = \mathbf{x}_{\phi(n)}$.

Exercice 1.10

Soit ϕ une extraction. Montrer que $\phi(n) \geq n$.
Soit ϕ_1 et ϕ_2 deux extractions. Montrer que $\phi_1 \circ \phi_2$ est encore une extraction.

Exemple 1.11. *Considérons la suite $x_n = (-1)^n$. Alors en prenant l'extraction $\phi(n) := 2n$, on voit que la suite constante $y_n = x_{2n} = 1$ est une sous-suite de x_n . De même en prenant l'extraction $\phi(n) := 2n + 1$, la suite constante $y_n = -1$ est une sous-suite de x_n .*

Définition 1.12: Valeur d'adhérence

Soit (\mathbf{x}_n) une suite de E , et soit $\mathbf{x}_* \in E$. On dit que \mathbf{x}_* est une **valeur d'adhérence** de la suite (\mathbf{x}_n) s'il existe une sous-suite de (\mathbf{x}_n) qui converge de \mathbf{x}_* . Autrement dit, il existe une extraction $\phi : \mathbb{N} \rightarrow \mathbb{N}$ telle que $\lim_{n \rightarrow \infty} \mathbf{x}_{\phi(n)} = \mathbf{x}_*$.

Dans l'exemple précédent, les points -1 et 1 sont des valeurs d'adhérence de la suite $(-1)^n$. Une suite peut avoir une infinité de valeurs d'adhérence, comme le montre l'exercice suivant.

Exercice 1.13

Soit (x_n) une suite de variables aléatoires indépendantes identiquement distribuées selon la loi uniforme de $[0, 1]$. Montrer que pour tout $a \in [0, 1]$, a est une valeur d'adhérence de la suite (x_n) , presque sûrement.

Théorème 1.14

Si (\mathbf{x}_n) converge vers \mathbf{x}_* , alors toute sous-suite de (\mathbf{x}_n) converge aussi vers \mathbf{x}_* . En particulier, \mathbf{x}_* est la seule valeur d'adhérence de la suite (\mathbf{x}_n) .

Démonstration. Soit $\varepsilon > 0$, et soit $N \in \mathbb{N}$ tel que, pour tout $n \geq N$, on a $\|\mathbf{x}_n - \mathbf{x}_*\| < \varepsilon$. Soit ϕ une extraction quelconque, et soit $\mathbf{y}_n = \mathbf{x}_{\phi(n)}$ une sous-suite de (\mathbf{x}_n) . Alors, pour tout $n \geq N$, on a $\phi(n) \geq n \geq N$, donc $\|\mathbf{y}_n - \mathbf{x}_*\| = \|\mathbf{x}_{\phi(n)} - \mathbf{x}_*\| < \varepsilon$. \square

1.2 Topologie dans les espaces vectoriels normés

Dans la suite, on fixe une norme $\|\cdot\|$ dans l'espace vectoriel E .

1.2.1 Définitions

Nous commençons avec des définitions.

Définition 1.15: Boule ouverte, boule fermée

La **boule ouverte** de centre $\mathbf{x}_0 \in E$ et de rayon $r > 0$ est le sous-ensemble de E défini par

$$\mathcal{B}(\mathbf{x}_0, r) := \{\mathbf{x} \in E, \|\mathbf{x} - \mathbf{x}_0\| < r\}.$$

La **boule fermée** de centre $\mathbf{x}_0 \in E$ et de rayon $r > 0$ est le sous-ensemble de E défini par

$$\overline{\mathcal{B}(\mathbf{x}_0, r)} := \{\mathbf{x} \in E, \|\mathbf{x} - \mathbf{x}_0\| \leq r\}.$$

La seule différence entre la boule ouverte et fermée est que la *sphère* $\|\mathbf{x} - \mathbf{x}_0\| = r$ est incluse dans la boule fermée, mais pas dans la boule ouverte. Par exemple, lorsque $E = \mathbb{R}$ avec $\|x\| := |x|$ (valeur absolue), on a

$$\mathcal{B}(0, 1) =] - 1, 1[, \quad \text{et} \quad \overline{\mathcal{B}(0, 1)} = [-1, 1].$$

Dans ce cours, nous adoptons la convention anglo-saxonne, et notons $(-1, 1)$ pour $] - 1, 1[$, $(-1, 0]$ pour $] - 1, 0]$, etc.

Définition 1.16: Ouverts, Fermés

Un sous-ensemble $A \subset E$ est **ouvert** si, pour tout $\mathbf{x} \in A$, il existe $r > 0$ telle que $\mathcal{B}(\mathbf{x}, r) \subset A$.

Un sous-ensemble $A \subset E$ est **fermé** si son complémentaire est ouvert, *i.e.* si $\mathbb{R}^d \setminus A$ est ouvert.

Par exemple, l'ensemble \emptyset est ouvert (on rappelle que toute proposition commençant par "*pour tout* $\mathbf{x} \in \emptyset$, ..." est vraie). L'ensemble E est ouvert (prendre $r = 1$ par exemple). En prenant les complémentaires, on obtient que E et \emptyset sont aussi des ensembles fermés (en fait, on peut montrer que ce sont les seuls dans E).

Théorème 1.17

Pour tout $\mathbf{x}_0 \in E$ et pour tout $r > 0$, la boule ouverte $\mathcal{B}(\mathbf{x}_0, r)$ est un ensemble ouvert.

Pour tout $\mathbf{x}_0 \in E$ et pour tout $r > 0$, la boule fermée $\overline{\mathcal{B}(\mathbf{x}_0, r)}$ est un ensemble fermé.

Démonstration. Soit $\mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r)$ et soit $r' := \|\mathbf{x} - \mathbf{x}_0\|$. Par définition de $\mathcal{B}(\mathbf{x}_0, r)$, on a $r' < r$. Soit $\varepsilon > 0$ suffisamment petit pour que $r' + \varepsilon < r$. Nous prétendons que $\mathcal{B}(\mathbf{x}, \varepsilon) \subset \mathcal{B}(\mathbf{x}_0, r)$.

En effet, soit $\mathbf{y} \in \mathcal{B}(\mathbf{x}, \varepsilon)$, de sorte que $\|\mathbf{x} - \mathbf{y}\| < \varepsilon$. En utilisant l'inégalité triangulaire, on a

$$\|\mathbf{x}_0 - \mathbf{y}\| \leq \|\mathbf{x}_0 - \mathbf{x}\| + \|\mathbf{x} - \mathbf{y}\| \leq r' + \varepsilon < r.$$

Donc $\mathbf{y} \in \mathcal{B}(\mathbf{x}_0, r)$. Ceci étant vrai pour tout $\mathbf{y} \in \mathcal{B}(\mathbf{x}, \varepsilon)$, on a $\mathcal{B}(\mathbf{x}, \varepsilon) \subset \mathcal{B}(\mathbf{x}_0, r)$.

Pour la deuxième partie, nous remarquons que le complémentaire de $\overline{\mathcal{B}(\mathbf{x}_0, r)}$ est l'ensemble

$$\mathbb{R}^d \setminus \overline{\mathcal{B}(\mathbf{x}_0, r)} = \{\mathbf{x} \in \mathbb{R}^d, \|\mathbf{x} - \mathbf{x}_0\| > r\}.$$

Soit $\mathbf{x} \in \mathbb{R}^d \setminus \overline{\mathcal{B}(\mathbf{x}_0, r)}$. On pose comme avant $r' := \|\mathbf{x} - \mathbf{x}_0\| > r$ et $\varepsilon > 0$ tel que $r' - \varepsilon > r$. On répète les arguments précédents en utilisant cette fois l'inégalité triangulaire inverse :

$$\|\mathbf{x}_0 - \mathbf{y}\| \geq \|\mathbf{x}_0 - \mathbf{x}\| - \|\mathbf{x} - \mathbf{y}\| \geq r' - \varepsilon > r.$$

□

Exercice 1.18

Dans la preuve précédente, montrer qu'on peut prendre $\varepsilon = \frac{1}{2}(r - r') > 0$.
Soit $\mathbf{x} \neq \mathbf{y}$ dans \mathbb{R}^d . Montrer qu'il existe $\varepsilon > 0$ tel que $\mathcal{B}(\mathbf{x}, \varepsilon) \cap \mathcal{B}(\mathbf{y}, \varepsilon) = \emptyset$.

En pratique, on sera intéressé à des ouverts autour d'un point \mathbf{x} donné. Cette notion est tellement importante qu'on lui donne une définition.

Définition 1.19: Voisinage de \mathbf{x}

On dit que $A \subset E$ est un **voisinage** de $\mathbf{x} \in E$ si $\mathbf{x} \in A$ et si A est un ensemble ouvert.

Exercice 1.20

Montrer que si A est un voisinage de \mathbf{x} , alors il existe $r > 0$ tel que $\mathcal{B}(\mathbf{x}, r) \subset A$.

1.2.2 Convergence de suites avec des ouverts/fermés

Grâce à la notion de boule ouverte, on peut traduire la notion de convergence en terme "topologique", de la manière suivante.

Théorème 1.21: Caractérisation de la limite

Un point \mathbf{x}_* est la limite de la suite (\mathbf{x}_n) ssi pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$ on a $\mathbf{x}_n \in \mathcal{B}(\mathbf{x}_*, \varepsilon)$.

Autrement dit, si \mathbf{x}_* est la limite de la suite \mathbf{x}_n (pour la norme $\|\cdot\|$), alors quelque soit le voisinage de \mathbf{x}_* , toute la suite est dans ce voisinage à partir d'un certain rang.

Théorème 1.22: Caractérisation des valeurs d'adhérence

Un point \mathbf{x}_* est valeur d'adhérence de la suite (\mathbf{x}_n) ssi pour tout $\varepsilon > 0$ et pour $N \in \mathbb{N}$, il existe $n \geq N$ tel que $\mathbf{x}_n \in \mathcal{B}(\mathbf{x}_*, \varepsilon)$.

Autrement dit, \mathbf{x}_* est valeur d'adhérence si pour tout voisinage de \mathbf{x}_* , il y a une infinité d'éléments de la suite (\mathbf{x}_n) qui appartiennent à ce voisinage.

Démonstration. Prenons une suite $\varepsilon_n > 0$ de nombres positifs qui tend vers 0 (par exemple $\varepsilon_n = \frac{1}{n}$). On construit l'extraction par induction. On prend $N_0 = 0$, et on considère un entier n_1 tel que $\mathbf{x}_{n_1} \in \mathcal{B}(\mathbf{x}_*, \varepsilon_1)$. On pose $\phi(1) = n_1$. Ensuite, on pose $N_1 = n_1 + 1$, et on considère un entier $n_2 \geq N_1 > n_1$ tel que $\mathbf{x}_{n_2} \in \mathcal{B}(\mathbf{x}_*, \varepsilon_2)$. On pose $\phi(2) = n_2$. Et ainsi de suite.

Par construction, on a $n_1 < n_2 < \dots$, donc la fonction ϕ est strictement croissante. C'est bien une extraction. De plus, on a $\mathbf{x}_{\phi(n)} \in \mathcal{B}(\mathbf{x}_*, \varepsilon_n)$, donc $0 \leq \|\mathbf{x}_{\phi(n)} - \mathbf{x}_*\| < \varepsilon_n$. La suite ε_n tend vers 0, donc $\|\mathbf{x}_{\phi(n)} - \mathbf{x}_*\|$ aussi, ce qui prouve le théorème. □

Nous terminons cette section avec un résultat qui fait le lien entre les ensembles fermés, et les convergences de suites.

Théorème 1.23: Caractérisation des fermés

Soit A est un ensemble fermé de E , et soit $(\mathbf{x}_n) \subset A$ une suite de A qui converge dans \mathbb{R}^d vers un certain \mathbf{x}_* . Alors $\mathbf{x}_* \in A$.

Démonstration. Supposons par l'absurde que $\mathbf{x}_* \notin A$. Alors $\mathbf{x}_* \in A^c$ qui est un ensemble ouvert. En particulier, il existe $\varepsilon > 0$ tel que $\mathcal{B}(\mathbf{x}_*, \varepsilon) \cap A = \emptyset$. Cela implique que pour tout $\mathbf{x} \in A$, on a $\|\mathbf{x} - \mathbf{x}_*\| \geq \varepsilon > 0$. En particulier, on a $\|\mathbf{x}_n - \mathbf{x}_*\| \geq \varepsilon$, qui ne tend pas vers 0. Contradiction. \square

1.3 L'espace \mathbb{R}^d

Nous nous intéressons maintenant au cas où $E = \mathbb{R}^d$. Dans la suite, on note $\|\cdot\|$ la *norme euclidienne* de \mathbb{R}^d . On rappelle que celle-ci est définie par

$$\forall \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d, \quad \|\mathbf{x}\| := \|\mathbf{x}\|_{\mathbb{R}^d} = \sqrt{x_1^2 + \cdots + x_d^2}.$$

Cette norme est aussi parfois notée $\|x\|_2$. On peut munir \mathbb{R}^d d'autres normes, et les définitions suivantes font encore sens. On démontrera dans la suite qu'en dimension finie (notre cas ici, d est la dimension), toutes les normes sont équivalentes.

1.3.1 Convergence composante par composante

Dans notre cas de l'espace \mathbb{R}^d , on peut vérifier la convergence dans \mathbb{R}^d (convergence de vecteurs), en vérifiant la convergence de chaque composante (convergence de nombres réels).

Théorème 1.24: Convergence composante par composante

Soit (\mathbf{x}_n) une suite de \mathbb{R}^d , et soit $\mathbf{x}_* \in \mathbb{R}^d$. On note $\mathbf{x}_n = (x_{1,n}, \dots, x_{d,n})^T$ et $\mathbf{x}_* = (x_{1,*}, \dots, x_{d,*})^T$. Alors (\mathbf{x}_n) converge vers \mathbf{x}_* ssi pour tout $1 \leq i \leq d$, la suite **réelle** $(x_{j,n})$ converge vers $x_{j,*}$.

Démonstration. Supposons pour commencer que (\mathbf{x}_n) converge vers \mathbf{x}_* . Soit $1 \leq i \leq d$. On a

$$0 \leq |x_{i,n} - x_{i,*}| \leq \left(\sum_{j=1}^d |x_{j,n} - x_{j,*}|^2 \right)^{1/2} = \|\mathbf{x}_n - \mathbf{x}_*\|.$$

Comme (\mathbf{x}_n) converge vers \mathbf{x}_* , le terme de droite converge vers 0 lorsque $n \rightarrow \infty$. Donc $|x_{i,n} - x_{i,*}|$ aussi, ce qui montre que $(x_{i,n})$ converge vers $x_{i,*}$ (dans \mathbb{R}).

Supposons maintenant que $x_{j,n} - x_{j,*}$ converge vers 0 pour tout $1 \leq j \leq d$. Alors $\lim_{n \rightarrow \infty} |x_{j,n} - x_{j,*}| = 0$. En prenant le carré, en sommant, et en utilisant les propriétés des limites, on obtient

$$\lim_{n \rightarrow \infty} \|\mathbf{x}_n - \mathbf{x}_*\|^2 = \lim_{n \rightarrow \infty} \sum_{j=1}^d |x_{j,n} - x_{j,*}|^2 = \sum_{j=1}^d \lim_{n \rightarrow \infty} |x_{j,n} - x_{j,*}|^2 = 0.$$

\square

1.3.2 Théorèmes de Bolzano Weierstrass

On rappelle le théorème de Bolzano–Weierstrass classique, dans \mathbb{R} .

Théorème 1.25: Bolzano–Weierstrass dans \mathbb{R} .

Si $(x_n) \subset \mathbb{R}$ est une suite **réelle** bornée, alors (x_n) admet une valeur d'adhérence.

Démonstration. Comme (x_n) est bornée, il existe $R > 0$ telle que $|x_n| < R$. Soit $\nu = [-R, R] \rightarrow \mathbb{N} \cup \{\infty\}$ la fonction définie par

$$\forall t \in [-R, R], \quad \nu(t) := \text{Card} \{n \in \mathbb{N}, \quad x_n < t\}.$$

Le nombre $\nu(t)$ compte le nombre d'éléments de la suite (x_n) qui sont plus petits que t . On s'autorise à prendre la valeur ∞ lorsqu'il y a une infinité d'éléments de la suite plus petits que t .

La suite $\nu(t)$ est croissante, et vérifie $\nu(-R) = 0$ et $\nu(R) = \infty$. On pose

$$t_* := \inf \{t \in [-R, R], \quad \forall s > t, \quad \nu(s) = \infty\}.$$

Montrons que t_* est une valeur d'adhérence de (x_n) . Par définition de t_* , pour tout $\varepsilon > 0$, on a $\nu(t_* + \varepsilon) = \infty$ et $\nu(t_* - \varepsilon) < \infty$. En particulier, il y a une infinité d'éléments de la suite dans l'intervalle $(t_* - \varepsilon, t_* + \varepsilon)$, et on conclut avec le Théorème 1.22. \square

La preuve précédente utilise fortement le caractère **ordonné** de \mathbb{R} . Dans le cas \mathbb{R}^d , on peut répéter l'argument “coordonnée par coordonnée”. On obtient le théorème suivant.

Théorème 1.26: Bolzano–Weierstrass dans \mathbb{R}^d .

Si $(\mathbf{x}_n) \subset \mathbb{R}^d$ est une suite (vectorielle) bornée, alors (\mathbf{x}_n) admet une valeur d'adhérence.

Démonstration. Si (\mathbf{x}_n) est bornée dans \mathbb{R}^d , alors la suite réelle $(x_{1,n})$ (première composante) est bornée dans \mathbb{R} . D'après le théorème de Bolzano–Weierstrass dans \mathbb{R} , on peut trouver une extraction $\phi_1 : \mathbb{N} \rightarrow \mathbb{N}$ et une limite $y_1 \in \mathbb{R}$ tel que $(x_{1,\phi_1(n)}) \rightarrow y_1$.

On considère maintenant la suite extraite réelle $\mathbf{x}_{2,\phi_1(n)}$ (deuxième composante). Cette suite est bornée dans \mathbb{R} . D'après le théorème de Bolzano–Weierstrass dans \mathbb{R} , on peut trouver une extraction $\phi_2 : \mathbb{N} \rightarrow \mathbb{N}$ et une limite $y_2 \in \mathbb{R}$ tel que $x_{2,\phi_1(\phi_2(n))} \rightarrow y_2$ (attention à l'ordre des extractions). De plus, par le Théorème 1.14 (toute sous-suite d'une suite convergente converge vers la même limite), on a encore $x_{1,\phi_1(\phi_2(n))} \rightarrow y_1$. Après cette seconde extraction, les deux premières composantes convergent.

On continue le raisonnement, et on construit les extractions $\phi_1, \phi_2, \dots, \phi_d$. On pose enfin

$$\phi(n) := \phi_1 \circ \phi_2 \circ \dots \circ \phi_d(n) = \phi_1(\phi_2(\dots(\phi_d(n)))).$$

Alors, par construction, la j -ème coordonnée de $\mathbf{x}_{\phi(n)}$ converge vers y_j . Donc la suite $\mathbf{x}_{\phi(n)}$ converge vers $\mathbf{y} := (y_1, \dots, y_d)^T$. \square

Ce résultat est faux en général en dimension infinie (que nous n'étudierons pas dans ce cours). Une raison est que la composée d'une infinité d'extraction n'est pas forcément bien définie. Par exemple, si $\phi(n) := n + 1$, on a $\phi^{(k)}(n) = n + k$ (composée k fois), et $\phi^{(\infty)}$ n'a aucun sens.

Exercice 1.27

Soit E l'ensemble des suites réelles bornées. Pour $\mathbf{u} = (u_n) \in E$, on pose $\|\mathbf{u}\|_E := \sup_{n \in \mathbb{N}} |u_n|$.

a/ Montrer que $(E, \|\cdot\|_E)$ est un \mathbb{R} -espace vectoriel normé (appelé souvent $\ell^\infty(\mathbb{N}, \mathbb{R})$).

b/ Pour $k \in \mathbb{N}$, soit $\mathbf{u}_k \in E$ la suite dont les k premiers termes valent 0, et tous les suivants valent 1. Montrer que pour tout $k, k' \in \mathbb{N}$, $\|\mathbf{u}_k - \mathbf{u}_{k'}\| = 1$.

c/ En déduire que la suite (\mathbf{u}_k) de E n'admet pas de sous-suite convergente.

2.1 Fonctions continues

Dans la suite, on se fixe deux espaces vectoriels normés, $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$, et on s'intéresse aux fonctions continues de $(E, \|\cdot\|_E)$ dans $(F, \|\cdot\|_F)$.

Nous rappelons qu'une fonction de E dans F devrait être notée (f, \mathcal{D}_f) , où $\mathcal{D}_f \subset E$ est le **domaine de définition** de f . C'est un sous-ensemble de E qui vérifie que $f(\mathbf{x})$ est bien défini pour tout $\mathbf{x} \in \mathcal{D}_f$. Cependant, on utilise en pratique la notation f (sans expliciter son domaine). Cela peut parfois (mais rarement) porter à confusion, voir le paragraphe suivant la définition.

Définition 2.1: Fonction continue, première définition

Soit (f, \mathcal{D}_f) une fonction de E dans F . Soit $\mathbf{x}_* \in \mathcal{D}_f$. On dit que f est continue en \mathbf{x}_* si :

$$\forall \varepsilon > 0, \quad \exists \delta > 0, \quad \forall \mathbf{y} \in \mathcal{D}_f, \quad \|\mathbf{y} - \mathbf{x}_*\|_E < \delta \implies \|f(\mathbf{x}_*) - f(\mathbf{y})\|_F < \varepsilon.$$

On dit que f est continue sur $K \subset \mathcal{D}_f$ si f est continue en tout point de K .

Voici une deuxième définition possible, et équivalente (le fait que ces deux définitions soient équivalentes est laissé en exercice).

Définition 2.2: Fonction continue, deuxième définition

Soit (f, \mathcal{D}_f) une fonction de E dans F . Soit $\mathbf{x}_* \in \mathcal{D}_f$. On dit que f est continue en \mathbf{x}_* si, pour toute suite $(\mathbf{x}_n)_n$ de \mathcal{D}_f qui converge vers \mathbf{x}_* (dans E), la suite $(f(\mathbf{x}_n))_n$ converge vers $f(\mathbf{x}_*)$ (dans F).

Remarque 2.3. *La notion de continuité dépend a priori des normes qu'on choisit sur E et sur F .*

Avant de donner des exemples explicites, nous faisons une petite digression. Avec cette définition, une fonction f de \mathbb{Z} dans \mathbb{R} est toujours continue (pourquoi?)... Considérons les fonctions (f_1, \mathbb{R}) et (f_2, \mathbb{Q}) définies par $f_1(x) = f_2(x) = \mathbf{1}(x \in \mathbb{Q})$. Alors, quand bien même f_1 et f_2 ont la même définition, la fonction f_2 est continue, mais pas la fonction f_1 ... Cela vient du fait qu'on a restreint le domaine de définition de f_1 . La notion de continuité dépend donc du choix du domaine. Dire «*f est continue en \mathbf{x}_0* » est un abus de langage. On devrait plutôt dire «*(f, \mathcal{D}_f) est continue en \mathbf{x}_0* ». Cependant, cela alourdit les notations, et on supposera dans la suite qu'on travaille toujours avec le **domaine maximal** de f .

Voici quelques exemples.

- (**addition**) La fonction $A : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $A(x_1, x_2) := x_1 + x_2$ est continue (car la somme des limites est égale à la limite des sommes...)
- (**multiplication**) La fonction $M : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $P(x_1, x_2) := x_1 x_2$ est continue (car le produit des limites est égale à la limite des produits...)
- (**projection**) Pour tout $1 \leq j \leq d$, la fonction $P_j : (x_1, \dots, x_d) \mapsto x_j$ est continue (cf. Théorème 1.24).
- (**norme**) La norme $\mathcal{N}(\cdot) := \|\cdot\|_E : E \rightarrow \mathbb{R}$ est continue. En effet, si \mathbf{x}_n converge vers \mathbf{x}_* , on a, avec l'inégalité triangulaire,

$$0 \leq |\mathcal{N}(\mathbf{x}_n) - \mathcal{N}(\mathbf{x}_*)| \leq \mathcal{N}(\mathbf{x}_n - \mathbf{x}_*) \xrightarrow[n \rightarrow \infty]{} 0.$$

Théorème 2.4: La composition de fonctions continues est continue

Soit (f, \mathcal{D}_f) une fonction de $(E, \|\cdot\|_E)$ dans $(F, \|\cdot\|_F)$, et soit (g, \mathcal{D}_g) une fonction de $(F, \|\cdot\|_F)$ dans $(G, \|\cdot\|_G)$. On suppose que $f(\mathcal{D}_f) \subset \mathcal{D}_g$. Alors $(g \circ f, \mathcal{D}_f)$ est continue de \mathcal{D}_f dans G .

Démonstration. Soit $\mathbf{x}_* \in \mathcal{D}_f$ fixé, et soit (\mathbf{x}_n) une suite de \mathcal{D}_f qui converge vers \mathbf{x}_* . Comme f est continue, la suite $\mathbf{y}_n := f(\mathbf{x}_n)$ converge vers $\mathbf{y}_* := f(\mathbf{x}_*)$. Comme g est continue, la suite $g(\mathbf{y}_n)$ converge vers $g(\mathbf{y}_*)$. Donc $g(f(\mathbf{x}_n)) \rightarrow g(f(\mathbf{x}_*))$. On en déduit que $g \circ f$ est continue en \mathbf{x}_* . Ceci étant vrai pour tout $\mathbf{x}_* \in \mathcal{D}_f$, $g \circ f$ est continue sur \mathcal{D}_f . \square

2.1.1 Continuité des fonctions usuelles

Le Théorème 2.4 permet de composer les fonctions continues. Par applications successives de ce théorème, on en déduit que (presque) toutes les fonctions usuelles, et leur composition, sont continues sur leur domaine (maximal) de définition.

En pratique, lorsqu'on voudra démontrer qu'une fonction f donnée est continue, on pourra se contenter de calculer son domaine de définition, et d'écrire

La fonction f est continue sur son domaine de définition comme composée de fonctions usuelles continues sur leur domaine de définition.

Il faudra tout de même préciser le domaine (maximal) de définition de f .

2.1.2 Prolongement par continuité

Définition 2.5: Prolongement par continuité

Soit (f, \mathcal{D}_f) une fonction de E dans F , et soit $\mathbf{x}_* \in E \setminus \mathcal{D}_f$. On dit que f admet un prolongement par continuité en \mathbf{x}_* s'il existe une limite $\ell \in F$ telle que, pour toute suite (\mathbf{x}_n) de $\mathcal{D}_f \cup \{\mathbf{x}_*\}$ qui converge vers \mathbf{x}_* , on a $f(\mathbf{x}_n) \rightarrow \ell$.

Dans ce cas, l'extension de f par continuité en \mathbf{x}_* est la fonction $(\tilde{f}, \mathcal{D}_{\tilde{f}})$ avec $\mathcal{D}_{\tilde{f}} := \mathcal{D}_f \cup \{\mathbf{x}_*\}$ définie par

$$\forall \mathbf{x} \in \mathcal{D}_{\tilde{f}}, \quad \tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \in \mathcal{D}_f \\ \ell & \text{if } \mathbf{x} = \mathbf{x}_*. \end{cases}$$

Si f était une fonction continue sur \mathcal{D}_f , alors la nouvelle fonction \tilde{f} est continue sur $\mathcal{D}_{\tilde{f}}$.

2.1.3 Un exemple important

Considérons l'exercice (classique) suivant.

Exercice 2.6

La fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, définie sur $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ par

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

admet-elle un prolongement par continuité en $\mathbf{0}$?

On remarque que f est continue sur son domaine de définition $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ comme composée de fonctions usuelles. Supposons que f admette un prolongement par continuité en $\mathbf{0}$, avec limite $\ell \in \mathbb{R}$. Alors, en prenant la suite $\mathbf{x}_n := (\frac{1}{n}, 0)$, on obtient $f(\mathbf{x}_n) \rightarrow 0$, donc $\ell = 0$. Cependant, en prenant la suite $\mathbf{y}_n := (\frac{1}{n}, \frac{1}{n})$, on a $f(\mathbf{y}_n) \rightarrow \frac{1}{2}$, donc $\ell = \frac{1}{2}$, contradiction. Ceci montre que f n'admet pas de prolongement par continuité en $\mathbf{0}$.

Cet exemple est assez surprenant, car les applications partielles $x \mapsto f(x, 0)$ et $y \mapsto f(0, y)$ admettent des prolongements par continuité sur \mathbb{R} (ce sont les fonctions nulles), mais pas la fonction f . Cela vient du fait que dans \mathbb{R}^2 , on peut approcher le point $\mathbf{0}$ de plein de façons différentes, par seulement selon les axes $x = 0$ et $y = 0$.

Par exemple, pour tout $\theta \in \mathbb{R}$, la fonction $g_\theta : \mathbb{R}_* \rightarrow \mathbb{R}$ définie par $g_\theta(r) := f(r \cos(\theta), r \sin(\theta))$ admet un prolongement par continuité en $r = 0$ avec $g_\theta(0) = \cos(\theta) \sin(\theta)$. Donc la fonction f initiale peut se prolonger par continuité le long de toute droite passant par $\mathbf{0}$. En revanche, elle admet des limites différentes selon la droite qu'on considère...

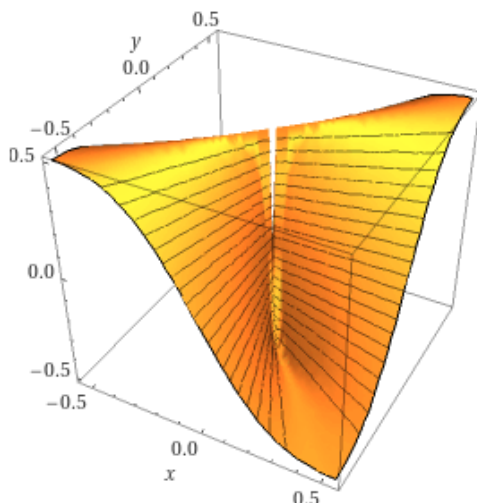


FIGURE 2.1 – La surface de f de l'Exercice 2.6. Il faut imaginer une *nappe pincée*.

2.1.4 Fonctions à m composantes

Dans la suite, on étudiera parfois des fonctions f de \mathbb{R}^d à valeurs dans \mathbb{R}^m . Ici, d est le nombre de **variables**, et m est le nombre de **composantes** de la fonction f . Une telle fonction est de la forme

$$f(\mathbf{x}) = f(x_1, \dots, x_d) = \begin{pmatrix} f_1(\mathbf{x}) \\ f_2(\mathbf{x}) \\ \vdots \\ f_m(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} f_1(x_1, \dots, x_d) \\ f_2(x_1, \dots, x_d) \\ \vdots \\ f_m(x_1, \dots, x_d) \end{pmatrix}.$$

Il est important de noter que les composantes de f sont *indépendantes*. On peut imaginer une situation par exemple où f_1 est une température, f_2 une pression, f_3 un nombre de pommes, etc. En appliquant le Théorème 1.24, on obtient de résultat suivant

Théorème 2.7

Soit (f, \mathcal{D}_f) une fonction de \mathbb{R}^d dans \mathbb{R}^m . Alors f est continue en $\mathbf{x}_* \in \mathcal{D}_f$ ssi toutes ses composantes sont continues en \mathbf{x}_* . En particulier, f est continue sur $K \subset \mathcal{D}_f$ ssi toutes ses composantes sont continues sur K .

Il n'y a donc pas de difficultés supplémentaires à étudier des fonctions à plusieurs composantes : on se ramène au cas des fonctions à valeurs réelles.

2.2 Optimisation, premières définitions

Dans cette section, nous commençons notre étude sur l'optimisation de fonctions. Notre but sera par exemple de trouver la plus petite valeur que prend une fonction f , et où elle attend cette valeur.

Pour commencer, nous remarquons qu'on ne peut optimiser qu'une fonction à valeurs réelles ($m = 1$ composante)! En effet, on doit pouvoir comparer les valeurs $f(\mathbf{x})$ et $f(\mathbf{y})$, et pouvoir écrire $f(\mathbf{x}) \leq f(\mathbf{y})$ ou $f(\mathbf{x}) > f(\mathbf{y})$. Or nous n'avons une notion d'ordre que pour des nombres réelles. Dans la suite, on s'intéresse donc exclusivement aux fonctions à valeurs dans \mathbb{R} .

2.2.1 Premières définitions

Nous commençons avec quelques définitions.

Infimum et minimum d'un sous-ensemble de \mathbb{R} .

Définition 2.8: Infimum d'un ensemble

Soit $I \subset \mathbb{R}$ un sous ensemble *quelconque* de \mathbb{R} .

Un nombre $m \in \mathbb{R}$ est un **minorant** de I si pour tout $x \in I$, on a $m < x$.

L'ensemble I est **borné inférieurement** si I admet au moins un minorant.

L'**infimum** de I est le plus grand des minorants. C'est le nombre $m_* \in \mathbb{R}$, noté $\inf I$, tel que, si m est un minorant de I , alors $m_* \geq m$.

Par convention, si I n'est pas borné inférieurement, on note $\inf I := -\infty$.

On définit de même la notion de **majorant**, **borné supérieurement** et de **supremum** de I .

Exemples :

- Si $I = \mathbb{R}$, on a $\inf I = -\infty$ et $\sup I = +\infty$.
- Si $I = (a, b)$, on a $\inf I = a$ et $\sup I = b$.
- Si $I = [a, b]$, on a $\inf I = a$, et $\sup I = b$.

Dans le premier cas, l'infimum et le supremum de I ne sont pas dans I , et dans le second cas, ils sont dans I . Ceci motive la définition suivante.

Définition 2.9

Soit $m_* := \inf I$ l'infimum de I . On dit que m_* est un **minimum** de I si $m_* \in I$.

Ainsi, a n'est pas un minimum de (a, b) , mais est un minimum de $[a, b]$. Par convention, si $m_* = -\infty$, m_* n'est pas un minimum. On définit de même la notion de **maximum** de I .

Théorème 2.10: Caractérisation de l'infimum avec des suites

Soit $I \subset \mathbb{R}$ un ensemble quelconque.

Si I est non borné inférieurement, alors il existe une suite (x_n) de I qui diverge vers $-\infty$.

Si I est borné inférieurement, et si $m \in \mathbb{R}$ est un minorant de I , alors m est l'infimum de I ssi il existe une suite $(x_n) \in I$ qui converge vers m_* .

Démonstration. Nous prouvons le deuxième point seulement. Soit m_* l'infimum de K . Comme m est un minorant, on a $m \leq m_*$. Si $m < m_*$, alors pour toute suite (x_n) de I , on a $m < m_* \leq x_n$, donc $|x_n - m| > |m_* - m|$, et la suite (x_n) ne peut pas converger vers m . Au contraire, si $m = m_*$ est l'infimum, alors, pour tout $n \in \mathbb{N}^*$, le nombre $m_* + \frac{1}{n}$ n'est pas un minorant de K . Donc il existe $x_n \in I$ avec $m_* \leq x_n \leq m_* + \frac{1}{n}$. Ceci construit une suite (x_n) de I qui converge vers m_* . \square

Exercice 2.11

Montrer que si I est fermé et borné inférieurement, alors I admet un minimum.

Infimum et minimum d'une fonction réelle.

Les notions précédentes s'appliquent pour des fonctions. Soit $K \subset \mathbb{R}^d$ un sous-ensemble *quelconque* de \mathbb{R}^d , et soit $f : K \rightarrow \mathbb{R}$ une fonction réelle. On note

$$I := f(K) = \{f(\mathbf{x}), \mathbf{x} \in K\} \subset \mathbb{R}.$$

Avec ces notations, on appelle l'infimum de f sur K le nombre $\inf I$. On le note parfois

$$\inf_K f \quad \text{ou} \quad \inf \{f(\mathbf{x}), \mathbf{x} \in K\}.$$

On note de même le supremum de f par $\sup_K f$.

Définition 2.12: Minimum et minimiseur d'une fonction

Soit $m_* := \inf \{f(\mathbf{x}), \mathbf{x} \in K\}$ l'infimum de f . On dit que m_* est le **minimum** de f si m_* est le minimum de I . Dans ce cas, il existe $\mathbf{x}_* \in K$ tel que $f(\mathbf{x}_*) = m_*$. Tout point $\mathbf{x}_* \in K$ vérifiant $f(\mathbf{x}_*) = \inf_K f$ est appelé un **minimiseur** de f sur K .

On définit pareillement la notion de **maximum** et de **maximiseur** de f sur K .

Exemples

- Soit $K := \mathbb{R}^d \setminus \{\mathbf{0}\}$, et soit $f : K \rightarrow \mathbb{R}$ définie par $f(\mathbf{x}) = \|\mathbf{x}\|$. Alors 0 est l'infimum de f sur K , mais 0 n'est pas un minimiseur, car $\mathbf{0} \notin K$.
- Soit $K = \mathbb{R}$, et $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = \sin^2(x)$. Alors 0 est l'infimum de f . C'est un minimum de f , et l'ensemble des minimiseurs de f est l'ensemble $\pi\mathbb{Z}$.

Grâce au Théorème 2.10, on a une caractérisation de l'infimum de f sur K .

Théorème 2.13

Soit m un minorant de f sur K . Alors m est l'infimum de f sur K ssi il existe une suite (\mathbf{x}_n) de K telle que $f(\mathbf{x}_n) \rightarrow m$.

Définition 2.14: Suite minimisante

Soit $m_* := \inf_K f$ l'infimum de f sur K . Une **suite minimisante** de f sur K est une suite (\mathbf{x}_n) de K telle que $f(\mathbf{x}_n) \rightarrow m_*$.

On remarquera que la suite (\mathbf{x}_n) ne converge pas forcément. Seules les images $f(\mathbf{x}_n)$ convergent...

2.2.2 Toutes les normes de \mathbb{R}^d sont équivalentes

Théorème 2.15: En dimension finie, toutes les normes sont équivalentes

Dans le cas $E = \mathbb{R}^d$, toutes les normes de E sont équivalentes.

Démonstration. Soit \mathcal{N} une norme sur \mathbb{R}^d . Il suffit de montrer que \mathcal{N} est équivalente à la norme $\|\cdot\|$ euclidienne. Pour commencer, en notant $(\mathbf{e}_i)_{1 \leq i \leq d}$ la base canonique de \mathbb{R}^d , et en notant $\mathbb{R}^d \ni \mathbf{x} = \sum_{i=1}^d x_i \mathbf{e}_i$ avec $x_i \in \mathbb{R}$, on a

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \mathcal{N}(\mathbf{x}) = \mathcal{N}\left(\sum_{i=1}^d x_i \mathbf{e}_i\right) \leq \sum_{i=1}^d \mathcal{N}(x_i \mathbf{e}_i) = \sum_{i=1}^d |x_i| \mathcal{N}(\mathbf{e}_i) \leq \left(\sum_{i=1}^d |x_i|^2\right)^{1/2} \left(\sum_{i=1}^d \mathcal{N}(\mathbf{e}_i)^2\right)^{1/2}.$$

On a utilisé l'inégalité de Cauchy-Schwarz pour la dernière inégalité. Ceci montre que

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \mathcal{N}(\mathbf{x}) \leq \alpha_1 \|\mathbf{x}\|, \quad \text{avec} \quad \alpha_1 := \left(\sum_{i=1}^d \mathcal{N}(\mathbf{e}_i)^2\right)^{1/2}.$$

Montrons l'autre inégalité. Soit

$$m := \inf \left\{ \mathcal{N}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \quad \|\mathbf{x}\| = 1 \right\}, \quad (2.1)$$

et soit (\mathbf{x}_n) une suite minimisante pour ce problème, c'est à dire qui vérifie $\|\mathbf{x}_n\| = 1$ et $\mathcal{N}(\mathbf{x}_n) \rightarrow m$. La suite (\mathbf{x}_n) est bornée dans \mathbb{R}^d . D'après le théorème de Bolzano-Weierstrass, il existe une sous-suite $(\mathbf{x}_{\phi(n)})$ et un élément $\mathbf{x}_* \in \mathbb{R}^d$ telle que \mathbf{x}_n converge vers \mathbf{x}_* pour la norme $\|\cdot\|$. Calculons $\|\mathbf{x}_*\|$ et $\mathcal{N}(\mathbf{x}_*)$. Pour commencer, on a, par l'inégalité triangulaire inverse,

$$0 \leq \left| \|\mathbf{x}_{\phi(n)}\| - \|\mathbf{x}_*\| \right| \leq \|\mathbf{x}_{\phi(n)} - \mathbf{x}_*\| \xrightarrow[n \rightarrow \infty]{} 0.$$

Donc $\|\mathbf{x}_{\phi(n)}\|$ converge vers $\|\mathbf{x}_*\|$ dans \mathbb{R} , et comme $\|\mathbf{x}_{\phi(n)}\| = 1$, on en déduit que $\|\mathbf{x}_*\| = 1$. De même, on a

$$0 \leq \left| \mathcal{N}(\mathbf{x}_{\phi(n)}) - \mathcal{N}(\mathbf{x}_*) \right| \leq \mathcal{N}(\mathbf{x}_{\phi(n)} - \mathbf{x}_*) \leq \alpha_1 \|\mathbf{x}_{\phi(n)} - \mathbf{x}_*\| \xrightarrow[n \rightarrow \infty]{} 0.$$

Donc $\mathcal{N}(\mathbf{x}_n)$ converge vers $\mathcal{N}(\mathbf{x}_*)$ dans \mathbb{R} , et comme $\mathcal{N}(\mathbf{x}_{\phi(n)})$ converge vers m , on en déduit que $\mathcal{N}(\mathbf{x}_*) = m$. Autrement dit, on a trouvé $\mathbf{x}_* \in \mathbb{R}^d$ avec $\|\mathbf{x}_*\| = 1$ et $\mathcal{N}(\mathbf{x}_*) = m$. Ainsi, m est un minimum du problème d'optimisation (2.1), et \mathbf{x}_* est un minimiseur. Comme $\|\mathbf{x}_*\| = 1$, on en déduit que $\mathbf{x}_* \neq \mathbf{0}$, et donc $m = \mathcal{N}(\mathbf{x}_*) > 0$.

Pour conclure, on remarque que pour tout $\mathbf{x} \in \mathbb{R}^d \setminus \{\mathbf{0}\}$, on a

$$\mathcal{N}(\mathbf{x}) = \mathcal{N}\left(\|\mathbf{x}\| \frac{\mathbf{x}}{\|\mathbf{x}\|}\right) = \|\mathbf{x}\| \mathcal{N}\left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right) \geq \|\mathbf{x}\| m.$$

Cette inégalité est encore vraie avec $\mathbf{x} = \mathbf{0}$. Ceci montre que

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \|\mathbf{x}\| \leq \alpha_2 \mathcal{N}(\mathbf{x}), \quad \text{avec} \quad \alpha_2 := \frac{1}{m} > 0.$$

En prenant $\alpha := \max\{\alpha_1, \alpha_2\}$, on trouve que \mathcal{N} et $\|\cdot\|$ sont équivalentes. □

2.3 Optimisation de fonctions continues, existence

Dans cette section, nous donnons des critères simples pour qu'un problème d'optimisation ait une solution.

2.3.1 Existence de minimiseurs dans le cas fermé borné

Nous commençons avec l'important théorème de Weierstrass.

Théorème 2.16: Théorème de Weierstrass

Soit $K \subset \mathbb{R}^d$ un ensemble **fermé** et **borné**, et soit $f : K \rightarrow \mathbb{R}$ une fonction **continue** sur K . Alors f atteint son minimum sur K , c'est à dire qu'il existe $\mathbf{x}_* \in K$ tel que $f(\mathbf{x}_*) = \inf_K f$.

Démonstration. Soit $m_* := \inf_K f$, et soit (\mathbf{x}_n) une suite minimisante de f .

La suite (\mathbf{x}_n) vit dans K , qui est borné. D'après le théorème de Bolzano–Weierstrass, on peut en extraire une sous-suite $\mathbf{x}_{\phi(n)}$ qui converge vers un certain $\mathbf{x}_* \in \mathbb{R}^d$.

La fonction f est continue, donc $f(\mathbf{x}_{\phi(n)}) \rightarrow f(\mathbf{x}_*)$. Par ailleurs, comme (\mathbf{x}_n) est une suite minimisante, on a aussi $f(\mathbf{x}_{\phi(n)}) \rightarrow m_*$. Par unicité de la limite, on en déduit que $f(\mathbf{x}_*) = m_*$.

De plus, l'ensemble K est fermé, et $\mathbf{x}_{\phi(n)}$ est une suite de K qui converge vers \mathbf{x}_* , donc $\mathbf{x}_* \in K$. Ainsi, \mathbf{x}_* est un minimiseur de f sur K . □

En particulier, comme f est définie sur K , $f(\mathbf{x}_*) \in \mathbb{R}$ ne prend pas la valeur $-\infty$, donc $m_* \in \mathbb{R}$ non plus. On en déduit que, sous les mêmes hypothèses, la fonction f est bornée inférieurement.

Évidemment, le théorème reste vrai si on remplace “minimum” par “maximum”.

2.3.2 Existence de minimiseurs pour les fonctions continues coercives

Dans la pratique, on étudie parfois des fonctions $f : \mathbb{R}^d \rightarrow \mathbb{R}$ définie sur tout l'espace \mathbb{R}^d . L'espace \mathbb{R}^d n'est pas borné, et on ne peut pas utiliser le théorème de Weierstrass. On utilise plutôt la notion de **coercivité**.

Définition 2.17: Fonctions coercives

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$. On dit que f est **coercive**, si, pour tout $M \in \mathbb{R}$, il existe $R > 0$ tel que

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \|\mathbf{x}\| > R \implies f(\mathbf{x}) > M.$$

Cela signifie qu'en dehors d'une grande boule $\mathcal{B}(\mathbf{0}, R)$, la fonction f prend des valeurs très grandes, plus grandes qu'une valeur M arbitraire. On note parfois $\lim_{\|\mathbf{x}\| \rightarrow \infty} f(\mathbf{x}) = \infty$ lorsque f est coercive.

Exercice 2.18

Soit $\alpha > 0$. Montrer que la fonction $f : \mathbb{R}^d \rightarrow \mathbb{R}$ définie par $f(\mathbf{x}) = \|\mathbf{x}\|^\alpha$ est coercive.

Théorème 2.19: Existence de minimiseurs pour les fonctions continues coercives

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction **continue** et **coercive**. Alors f atteint son minimum : il existe $\mathbf{x}_* \in \mathbb{R}^d$ tel que $f(\mathbf{x}_*) = \inf_{\mathbb{R}^d} f$.

Démonstration. Notons $M := f(\mathbf{0}) + 1$. Comme f est coercive, il existe $R > 0$ tel que pour tout $\mathbf{x} \in \mathbb{R}^d \setminus \overline{\mathcal{B}(\mathbf{0}, R)}$, on a $f(\mathbf{x}) > M$. On divise l'espace \mathbb{R}^d en deux, et on note

$$K_1 := \overline{\mathcal{B}(\mathbf{0}, R)}, \quad \text{et} \quad K_2 := \mathbb{R}^d \setminus K_1.$$

Pour commencer, on remarque que

$$\inf_{\mathbb{R}^d} f = \min \left\{ \inf_{K_1} f, \inf_{K_2} f \right\}.$$

Par ailleurs, pour $\mathbf{x} \in K_2$, on a $\|\mathbf{x}\| > R$, et donc $f(\mathbf{x}) > M > f(\mathbf{0})$. En particulier, comme $\mathbf{0} \in K_1$ on a

$$\forall \mathbf{x} \in K_2, \quad \inf_{K_1} f \leq f(\mathbf{0}) = M - 1 < f(\mathbf{x}) - 1.$$

On en déduit que $\inf_{K_1} f < \inf_{K_2} f$, et donc $\inf_{\mathbb{R}^d} f = \inf_{K_1} f$. L'ensemble K_1 est la boule fermée de centre $\mathbf{0}$ et de rayon R . C'est donc un ensemble fermé et borné. De plus, f est continue sur cet ensemble. D'après le théorème de Bolzano–Weierstrass, f atteint son minimum sur K_1 . Il existe $\mathbf{x}_* \in K_1 \subset \mathbb{R}^d$ tel que $f(\mathbf{x}_*) = \inf_{K_1} f = \inf_{\mathbb{R}^d} f$. On en déduit que \mathbf{x}_* est un minimiseur de f sur tout l'espace \mathbb{R}^d . \square

2.3.3 Unicité des minimiseurs pour les fonctions strictement convexes

Dans les sections précédentes, nous avons trouvé des critères simples pour qu'une fonction admette des minimiseurs. Nous cherchons maintenant un critère simple pour que ce minimiseur soit **unique**. Une bonne notion est celle de **convexité**.

Définition 2.20: Ensemble convexe

Soit $K \subset \mathbb{R}^d$ un ensemble quelconque. On dit que K est **convexe**, si, pour tous $\mathbf{x}, \mathbf{y} \in K$ et tout $0 \leq t \leq 1$, le point $t\mathbf{x} + (1-t)\mathbf{y}$ est dans K .

On remarquera que le point $t\mathbf{x} + (1-t)\mathbf{y}$ est une combinaison barycentrique des points \mathbf{x} et \mathbf{y} . En particulier, il parcourt le segment $[\mathbf{x}, \mathbf{y}]$ lorsque t parcourt l'intervalle $[0, 1]$. Autrement dit, K est convexe si, pour toute paire de points \mathbf{x}, \mathbf{y} dans K , le segment $[\mathbf{x}, \mathbf{y}]$ est inclus dans K .

Définition 2.21: Fonction convexe

Soit $K \subset \mathbb{R}^d$ un ensemble convexe, et soit $f : K \rightarrow \mathbb{R}$ une fonction définie sur K . On dit que f est **convexe** si

$$\forall \mathbf{x}, \mathbf{y} \in K, \quad \forall t \in [0, 1], \quad f(t\mathbf{x} + (1-t)\mathbf{y}) \leq tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$

On dit que f est **strictement convexe** si

$$\forall \mathbf{x} \neq \mathbf{y} \in K, \quad \forall t \in (0, 1), \quad f(t\mathbf{x} + (1-t)\mathbf{y}) < tf(\mathbf{x}) + (1-t)f(\mathbf{y}).$$

Lorsque la fonction $-f$ est convexe, on dit que la fonction est **concave**. Pour les fonctions concaves ou strictement concaves, les inégalités sont dans l'autre sens. Nous n'utiliserons cependant pas cette notion dans ce cours, car nous sommes intéressés surtout par la minimisation de fonctions.

Exercice 2.22

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction convexe, et soit $E \in \mathbb{R}$. Montrer que l'ensemble $K_E := \{\mathbf{x} \in \mathbb{R}^d, f(\mathbf{x}) \leq E\}$ est un ensemble convexe de \mathbb{R}^d .

Exemples

- L'ensemble \mathbb{R}^d est convexe.
- Si $\mathcal{N} : \mathbb{R}^d \rightarrow \mathbb{R}$ est une norme, alors \mathcal{N} est convexe. En effet, on a, avec l'inégalité triangulaire,

$$\mathcal{N}(t\mathbf{x} + (1-t)\mathbf{y}) \leq \mathcal{N}(t\mathbf{x}) + \mathcal{N}((1-t)\mathbf{y}) = t\mathcal{N}(\mathbf{x}) + (1-t)\mathcal{N}(\mathbf{y}).$$

- Les boules ouvertes et fermées sont des ensembles convexes. En effet, ces ensembles sont de la forme $\overline{\mathcal{B}}(0, E) = \{\mathbf{x} \in \mathbb{R}^d, \mathcal{N}(\mathbf{x}) \leq E\}$ (cf exercice précédent).

Exercice 2.23

Montrer que si K_1 et K_2 sont deux ensembles convexes, alors l'intersection $K_1 \cap K_2$ est encore convexe. Est-ce que l'union $K_1 \cup K_2$ est convexe ?

Théorème 2.24: Unicité du minimiseur pour les fonctions strictement convexes

Soit $K \subset \mathbb{R}^d$ un ensemble convexe, et soit $f : K \rightarrow \mathbb{R}$ une fonction strictement convexe sur K . Alors, si f admet un minimiseur, celui-ci est unique.

Démonstration. Supposons que f admette deux minimiseurs $\mathbf{x}_0 \neq \mathbf{x}_1$ dans K . Donc $f(\mathbf{x}_0) = f(\mathbf{x}_1) = m_* := \inf_K f$. On considère le point $\mathbf{x}_* := \frac{1}{2}\mathbf{x}_0 + \frac{1}{2}\mathbf{x}_1$. Comme K est convexe, on a $\mathbf{x}_* \in K$, et comme f est strictement convexe, on a

$$f(\mathbf{x}_*) = f\left(\frac{1}{2}\mathbf{x}_0 + \frac{1}{2}\mathbf{x}_1\right) < \frac{1}{2}f(\mathbf{x}_0) + \frac{1}{2}f(\mathbf{x}_1) = m_*.$$

Le point \mathbf{x}_* de K vérifie $f(\mathbf{x}_*) < m_*$. Ceci contredit la définition de m_* comme minimum. Donc $\mathbf{x}_0 = \mathbf{x}_1$, et le minimiseur est unique. \square

3.1 Applications linéaires

3.1.1 L'espace vectoriel des applications linéaires bornées

Avant de définir la dérivée de fonctions à plusieurs variables, nous faisons un rappel sur les applications linéaires. Dans la suite, on considère $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ deux \mathbb{R} -espaces vectoriels normés.

Définition 3.1: Application linéaire

Soit $L : E \rightarrow F$. On dit que L est une **application linéaire** de E dans F si :

- pour tous $\mathbf{x}, \mathbf{y} \in E$, on a $L(\mathbf{x} + \mathbf{y}) = L(\mathbf{x}) + L(\mathbf{y})$,
- pour tout $\lambda \in \mathbb{R}$ et tout $\mathbf{x} \in E$, on a $L(\lambda\mathbf{x}) = \lambda L(\mathbf{x})$.

On dit que cette application linéaire est **bornée** s'il existe $M > 0$ tel que pour tout $\mathbf{x} \in E$ avec $\|\mathbf{x}\|_E = 1$, on a $\|L(\mathbf{x})\|_F \leq M$.

On note $\mathcal{L}(E, F)$ l'ensemble des applications linéaires bornées de E dans F .

Exemples :

- L'application nulle $L(\mathbf{x}) = \mathbf{0}_F$ est linéaire. Elle est bornée, avec $M = 0$.
- L'identité $L : E \rightarrow E$ définie par $L(\mathbf{x}) = \mathbf{x}$ est linéaire, bornée, avec $M = 1$.
- Si $E = \mathbb{R}^n$ et $F = \mathbb{R}^m$. Alors pour toute matrice $A \in \mathcal{M}_{m,n}(\mathbb{R})$, l'application $L_A : \mathbf{x} \mapsto A\mathbf{x}$ est linéaire. Elle est bornée (cf plus tard).

Si L est bornée, on a que, pour tout $\mathbf{x} \in E \setminus \{\mathbf{0}\}$, en remarquant $\frac{\mathbf{x}}{\|\mathbf{x}\|_E}$ est de norme 1,

$$\|L(\mathbf{x})\|_F = \left\| L \left(\|\mathbf{x}\|_E \frac{\mathbf{x}}{\|\mathbf{x}\|_E} \right) \right\|_F = \|\mathbf{x}\|_E \left\| L \left(\frac{\mathbf{x}}{\|\mathbf{x}\|_E} \right) \right\|_F \leq M \|\mathbf{x}\|_E.$$

Ceci est vrai aussi pour $\mathbf{x} = \mathbf{0}$. Le raisonnement précédent qui consiste à remonter une inégalité sur la **sphère** de E à tout l'espace E s'appelle un argument **d'homogénéité**. Le plus petit $M > 0$ qui vérifie cette inégalité pour tout $\mathbf{x} \in E$ s'appelle la **norme d'opérateur** de L , notée $\|L\|_{\text{op}}$. Autrement dit, si $L \in \mathcal{L}(E, F)$,

$$\|L\|_{\text{op}} = \inf \{ \|L(\mathbf{x})\|_F, \quad \mathbf{x} \in E, \quad \|\mathbf{x}\|_E = 1 \}.$$

On peut écrire l'inégalité utile suivante :

$$\forall L \in \mathcal{L}(E, F), \quad \forall \mathbf{x} \in E, \quad \|L(\mathbf{x})\|_F \leq \|L\|_{\text{op}} \|\mathbf{x}\|_E. \quad (3.1)$$

Théorème 3.2

L'espace $\mathcal{L}(E, F)$ muni de la norme $\|\cdot\|_{\text{op}}$ est un espace vectoriel.

Démonstration. On vérifie directement que $\mathcal{L}(E, F)$ est un espace vectoriel, avec $(L_1 + L_2)(\mathbf{x}) := L_1(\mathbf{x}) + L_2(\mathbf{x})$, et $(\lambda L)(\mathbf{x}) := \lambda L(\mathbf{x})$. Montrons que $\|\cdot\|_{\text{op}}$ est une norme. Pour commencer, si $\|L\|_{\text{op}} = 0$, alors l'inégalité (3.1) implique que $L(\mathbf{x}) = \mathbf{0}$ pour tout $\mathbf{x} \in E$, donc L est l'application nulle. Ensuite, on vérifie directement que $\|\lambda L\|_{\text{op}} = |\lambda| \|L\|_{\text{op}}$. Il reste à vérifier l'inégalité triangulaire. On a, pour tout $\mathbf{x} \in E$ avec $\|\mathbf{x}\|_E = 1$,

$$\|(L_1 + L_2)(\mathbf{x})\|_F \leq \|L_1(\mathbf{x})\|_F + \|L_2(\mathbf{x})\|_F \leq \|L_1\|_{\text{op}} + \|L_2\|_{\text{op}}.$$

En prenant l'infimum sur $\mathbf{x} \in E$ avec $\|\mathbf{x}\|_E = 1$, on obtient bien $\|L_1 + L_2\|_{\text{op}} \leq \|L_1\|_{\text{op}} + \|L_2\|_{\text{op}}$. \square

Théorème 3.3

Une application linéaire $L : E \rightarrow F$ est continue ssi elle est bornée.

Attention, ce théorème n'est vrai que pour des applications linéaires!

Démonstration. Supposons que L soit bornée. Soit (\mathbf{x}_n) une suite de E qui converge vers \mathbf{x}_* dans E . On a

$$\|L(\mathbf{x}_n) - L(\mathbf{x}_*)\|_F = \|L(\mathbf{x}_n - \mathbf{x}_*)\|_F \leq \|L\|_{\text{op}} \|\mathbf{x}_n - \mathbf{x}_*\| \xrightarrow{n \rightarrow \infty} 0.$$

Ceci montre que L est continue, en tant que fonction de E dans F .

Supposons au contraire que L ne soit pas bornée. Dans ce cas, il existe une suite $(\mathbf{x}_n) \in E$ avec $\|\mathbf{x}_n\|_E = 1$ et $\|L(\mathbf{x}_n)\|_F \rightarrow \infty$. On pose $\mathbf{y}_n := \frac{\mathbf{x}_n}{\|L(\mathbf{x}_n)\|_F}$. On a $\|\mathbf{y}_n\|_E = \frac{1}{\|L(\mathbf{x}_n)\|_F} \rightarrow 0$, donc la suite (\mathbf{y}_n) tend vers $\mathbf{0}_E$ dans E . En revanche, on a

$$\|L(\mathbf{y}_n)\|_F = \left\| L \left(\frac{\mathbf{x}_n}{\|L(\mathbf{x}_n)\|_F} \right) \right\|_F = \frac{1}{\|L(\mathbf{x}_n)\|_F} \|L(\mathbf{x}_n)\|_F = 1.$$

Or $\mathbf{y}_n \rightarrow \mathbf{0}_E$. On en déduit que L n'est pas continue en $\mathbf{0}_E$, donc a fortiori n'est pas continue. \square

Remarque 3.4. Le raisonnement précédent montre que si L est continue en $\mathbf{0}$, alors L est bornée, donc L est continue sur tout E . Pour les applications linéaire, être continue en $\mathbf{0}$ ou être continue partout est équivalent.

3.1.2 Le cas de la dimension finie

Dans le cas où $E = \mathbb{R}^d$ et $F = \mathbb{R}^m$ sont de dimension finie, on peut dire plus de choses.

Théorème 3.5

Une application linéaire $L : \mathbb{R}^d \rightarrow \mathbb{R}^m$ est automatiquement bornée, donc continue.

Démonstration. Notons $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ la base canonique de \mathbb{R}^d . On a, pour $\mathbf{x} = \sum_{j=1}^d x_j \mathbf{e}_j \in \mathbb{R}^d$,

$$\|L(\mathbf{x})\|_{\mathbb{R}^m} = \left\| \sum_{j=1}^d x_j L(\mathbf{e}_j) \right\|_{\mathbb{R}^m} \leq \sum_{j=1}^d |x_j| \|L(\mathbf{e}_j)\|_{\mathbb{R}^m} \leq \left(\sum_{j=1}^d |x_j|^2 \right)^{1/2} \left(\sum_{j=1}^d \|L(\mathbf{e}_j)\|_{\mathbb{R}^m}^2 \right)^{1/2}.$$

On a utilisé l'inégalité de Cauchy-Schwarz pour la dernière inégalité. On reconnaît la norme de \mathbf{x} dans le membre de droite. Ainsi,

$$\forall \mathbf{x} \in \mathbb{R}^d, \|\mathbf{x}\|_{\mathbb{R}^d} = 1, \quad \|L(\mathbf{x})\|_{\mathbb{R}^m} \leq M, \quad \text{avec} \quad M = \left(\sum_{j=1}^d \|L(\mathbf{e}_j)\|_{\mathbb{R}^m}^2 \right)^{1/2} < \infty.$$

□

Représentation matricielle

Soit $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ la base canonique de $E = \mathbb{R}^d$, et soit $(\mathbf{f}_1, \dots, \mathbf{f}_m)$ celle de $F = \mathbb{R}^m$ (en fait, le raisonnement suivant marche avec n'importe quelle base).

Pour tout $1 \leq j \leq d$, le vecteur $L(\mathbf{e}_j)$ appartient à F , et peut être décomposé dans la base $(\mathbf{f}_1, \dots, \mathbf{f}_m)$. On note a_{ij} les coefficients correspondants, c'est à dire

$$L(\mathbf{e}_j) = \sum_{i=1}^m a_{ij} \mathbf{f}_i = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in F = \mathbb{R}^m.$$

Soit $A = (a_{ij})_{1 \leq i \leq m, 1 \leq j \leq d}$ la matrice dont les vecteurs colonnes sont les $L(\mathbf{e}_i)$ pour $1 \leq i \leq d$, c'est à dire

$$A = (L(\mathbf{e}_1), L(\mathbf{e}_2), \dots, L(\mathbf{e}_d)) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{md} \end{pmatrix} \in \mathcal{M}_{m \times d}(\mathbb{R}).$$

On dit que A est la **matrice représentative de L dans les bases $(\mathbf{e}_1, \dots, \mathbf{e}_d)$ et $(\mathbf{f}_1, \dots, \mathbf{f}_m)$** . On notera en effet que la matrice A dépend du choix de la base choisie (alors que l'application L n'en dépend pas).

L'importance de cette matrice, et du calcul matriciel en général, vient du fait que, pour tout $\mathbf{x} = \sum x_j \mathbf{e}_j$, on a

$$L(\mathbf{x}) = L\left(\sum_{j=1}^d x_j \mathbf{e}_j\right) = \sum_{j=1}^d x_j L(\mathbf{e}_j) = \sum_{j=1}^d \sum_{i=1}^m a_{ij} x_j \mathbf{f}_i.$$

Autrement dit, si $\mathbf{y} = L(\mathbf{x})$, alors les coefficients y_i de \mathbf{y} (dans la base canonique de F) sont obtenus à partir des coefficients x_i de \mathbf{x} (dans la base canonique de E) via la formule

$$y_i = \sum_{j=1}^d a_{ij} x_j, \quad \text{ou encore,} \quad \mathbf{y} = A\mathbf{x}.$$

où la dernière égalité est la multiplication matrice-vecteur de A par \mathbf{x} . On remarquera qu'il y a un léger abus de notation ici. En effet, l'égalité $\mathbf{y} = L(\mathbf{x})$ ne dépend pas de la base choisie ($\mathbf{x} \in E$ et $\mathbf{y} \in F$ sont des vecteurs abstraits), alors que dans l'égalité $\mathbf{y} = A\mathbf{x}$, on a posé

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \in \mathbb{R}^d, \quad \text{et} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m,$$

et cette notation vectorielle dépend du choix de la base de $E = \mathbb{R}^d$ et de $F = \mathbb{R}^m$. Dans la suite, on fera cet abus de notation, avec la convention qu'on travaille toujours dans la base canonique.

On en déduit le résultat suivant.

Théorème 3.6

L'espace vectoriel $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ est en bijection avec l'espace vectoriel $\mathcal{M}_{m \times d}(\mathbb{R})$, qui est en bijection avec $\mathbb{R}^{m \times d}$. En particulier, $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$ est de dimension finie, de dimension $d \times m$, et toutes ses normes sont équivalentes.

3.2 Fonctions différentiables

3.2.1 Retour en dimension $d = 1$

En dimension $d = 1$, on rappelle la définition de la dérivabilité d'une fonction. On dit qu'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est **dérivable** en $x_0 \in \mathbb{R}$ si la limite

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}$$

existe. Si c'est le cas, on note $f'(x_0) \in \mathbb{R}$ cette limite. Malheureusement, cette définition ne peut pas s'étendre au cas des dimensions supérieures. En effet, on aurait naturellement $\mathbf{x}_0 \in \mathbb{R}^d$ et $\mathbf{h} \in \mathbb{R}^d$, or on ne peut pas diviser par un vecteur \mathbf{h} ...

Pour contourner cette difficulté, on réécrit la définition précédente d'une autre manière. Posons

$$\mathcal{E}(h) := \frac{f(x_0 + h) - f(x_0)}{h} - f'(x_0).$$

Dire que le ratio converge vers $f'(x)$ lorsque $h \rightarrow 0$, c'est dire que $\lim_{h \rightarrow 0} \mathcal{E}(h) = 0$. Ainsi, on a, en multipliant par h et en ré-ordonnant les termes,

$$f(x_0 + h) = f(x_0) + f'(x_0)h + h\mathcal{E}(h).$$

On remarque que le dernier terme est un $o(h)$ (cf plus tard pour la définition de $o(\cdot)$). Ainsi, la définition de la dérivabilité en dimension 1 est équivalente à dire que f admet un **développement limité**, ou **développement de Taylor**, à l'ordre 1, de la forme

$$f(x_0 + h) = f(x_0) + f'(x_0)h + o(h).$$

En remarquant que la fonction $h \mapsto f'(x_0)h$ est une fonction linéaire de \mathbb{R} dans \mathbb{R} , on comprend comment étendre la définition aux dimensions supérieures.

3.2.2 Définition de la différentiabilité

On se place dans le cadre général des espaces vectoriels normés $(E, \|\cdot\|_E)$ et $(F, \|\cdot\|_F)$ pour énoncer la définition. Dans la suite, on écrira $f : U \subset E \rightarrow F$ pour indiquer que U est un ouvert de E et que f est définie sur U (donc $U \subset \mathcal{D}_f \subset E$).

Définition 3.7: Différentiabilité en un point.

Soit $f : U \subset E \rightarrow F$, et soit $\mathbf{x}_0 \in U$. On dit que f est **différentiable** en \mathbf{x}_0 s'il existe une application linéaire bornée $\mathcal{L}(E, F)$ telle que

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + L(\mathbf{h}) + o(\mathbf{h}).$$

Si c'est le cas, L est unique, est notée $df_{\mathbf{x}_0}$, et est appelée **différentielle** de f en \mathbf{x}_0 . Si f est différentiable en tout point de U , on dit que que f est différentiable sur U .

La formule

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + df_{\mathbf{x}_0}(\mathbf{h}) + o(\mathbf{h})$$

s'appelle parfois la **formule de Taylor** (à l'ordre 1).

Définition 3.8: Notation de Landau

Une fonction $o(\mathbf{h})$ est une fonction de la forme $\|\mathbf{h}\|\mathcal{E}(\mathbf{h})$ avec $\mathcal{E}(\mathbf{h}) \rightarrow \mathbf{0}$ lorsque $\mathbf{h} \rightarrow \mathbf{0}$.

Une fonction \mathcal{E} vérifie $\mathcal{E}(\mathbf{h}) \rightarrow \mathbf{0}$ lorsque $h \rightarrow \mathbf{0}$ ssi elle est continue en $\mathbf{0}$ avec $\mathcal{E}(\mathbf{0}) = \mathbf{0}$.

On insiste sur le fait que $df_{\mathbf{x}_0}$ est une application linéaire.

★ **Exemple (fonctions de \mathbb{R} dans \mathbb{R}).** Dans l'exemple précédent $d = 1$, la différentielle de $f : \mathbb{R} \rightarrow \mathbb{R}$ au point $x_0 \in \mathbb{R}$ est l'application linéaire $df_{x_0} \in \mathcal{L}(\mathbb{R}, \mathbb{R})$ définie par

$$\forall h \in \mathbb{R}, \quad df_{x_0}(h) := f'(x_0)h.$$

En quelque sorte, la **dérivée** $f'(x_0)$ est le coefficient réel qui représente la **différentielle** df_{x_0} .

★ **Exemple (fonctions linéaires) :** Si $L \in \mathcal{L}(E, F)$ est une application linéaire bornée/continue, alors L est différentiable sur E , avec, pour tout $\mathbf{x}_0 \in E$, $dL_{\mathbf{x}_0} = L$. En effet, on a

$$L(\mathbf{x}_0 + \mathbf{h}) = L(\mathbf{x}_0) + L(\mathbf{h}) + o(\mathbf{h}), \quad \text{avec } o(\mathbf{h}) = \mathbf{0}.$$

★ **Exemple (forme quadratique) :** Soit $A \in \mathcal{M}_{d,d}(\mathbb{R})$, et $f : \mathbb{R}^d \rightarrow \mathbb{R}$ définie par $f(\mathbf{x}) := \langle \mathbf{x}, A\mathbf{x} \rangle$. Calculons la différentielle de f . On a

$$f(\mathbf{x} + \mathbf{h}) = \langle \mathbf{x} + \mathbf{h}, A(\mathbf{x} + \mathbf{h}) \rangle = \langle \mathbf{x}, A\mathbf{x} \rangle + \langle \mathbf{x}, A\mathbf{h} \rangle + \langle \mathbf{h}, A\mathbf{x} \rangle + \langle \mathbf{h}, A\mathbf{h} \rangle.$$

Le premier terme est $f(\mathbf{x})$. Le second et le troisième sont linéaires en \mathbf{h} (donc compose la différentielle). Montrons que le dernier terme est un $o(\mathbf{h})$. On pose

$$\mathcal{E}(\mathbf{h}) := \begin{cases} \frac{\langle \mathbf{h}, A\mathbf{h} \rangle}{\|\mathbf{h}\|} & \text{si } \mathbf{h} \neq \mathbf{0} \\ 0 & \text{si } \mathbf{h} = \mathbf{0}. \end{cases}$$

de sorte que $\langle \mathbf{h}, A\mathbf{h} \rangle = \|\mathbf{h}\|\mathcal{E}(\mathbf{h})$. Montrons que \mathcal{E} est continue en $\mathbf{0}$ avec $\mathcal{E}(\mathbf{0}) = 0$. Pour cela, on utilise l'inégalité de Cauchy-Schwarz et l'inégalité de la norme d'opérateur de A , et on a

$$|\langle \mathbf{h}, A\mathbf{h} \rangle| \leq \|\mathbf{h}\| \cdot \|A\mathbf{h}\| \leq \|\mathbf{h}\| \cdot \|A\|_{\text{op}} \cdot \|\mathbf{h}\| = \|A\|_{\text{op}} \|\mathbf{h}\|^2.$$

Ceci montre que $|\mathcal{E}(\mathbf{h})| \leq \|A\|_{\text{op}} \|\mathbf{h}\|$, donc \mathcal{E} est bien continue en $\mathbf{0}$ avec $\mathcal{E}(\mathbf{0}) = 0$. Ainsi, notre développement précédent est de la forme $f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + df_{\mathbf{x}}(\mathbf{h}) + o(\mathbf{h})$ avec, *par identification*,

$$df_{\mathbf{x}} : \mathbb{R}^d \rightarrow \mathbb{R}, \quad \text{définie par } df_{\mathbf{x}}(\mathbf{h}) := \langle \mathbf{x}, A\mathbf{h} \rangle + \langle \mathbf{h}, A\mathbf{x} \rangle.$$

Théorème 3.9: Différentiabilité implique continuité

Si $f : U \subset E \rightarrow F$ est différentiable en $\mathbf{x}_0 \in E$, alors f est continue en \mathbf{x}_0 .

Démonstration. Soit (\mathbf{x}_n) une suite qui converge vers \mathbf{x}_0 , et posons $\mathbf{h}_n := \mathbf{x}_n - \mathbf{x}_0$, de sorte que $\mathbf{x}_n = \mathbf{x}_0 + \mathbf{h}_n$. La suite \mathbf{h}_n tend vers $\mathbf{0}$ dans E . On a

$$\|f(\mathbf{x}_0 + \mathbf{h}_n) - f(\mathbf{x}_0)\|_F = \|df_{\mathbf{x}_0}(\mathbf{h}_n) + \|\mathbf{h}_n\|_E \mathcal{E}(\mathbf{h}_n)\|_F \leq \|df_{\mathbf{x}_0}\|_{\text{op}} \|\mathbf{h}_n\|_E + \|\mathbf{h}_n\|_E \|\mathcal{E}(\mathbf{h}_n)\|_F.$$

Comme \mathcal{E} est continue en $\mathbf{0}$ avec $\mathcal{E}(\mathbf{0}_E) = \mathbf{0}_F$, et que $\mathbf{h}_n \rightarrow \mathbf{0}$, le terme de droite converge vers 0 lorsque $n \rightarrow \infty$. Ainsi, on a $f(\mathbf{x}_n) \rightarrow f(\mathbf{x}_0)$, comme souhaité. \square

3.2.3 Dérivées directionnelles et dérivées partielles

Soit $f : U \subset E \rightarrow F$, soit $\mathbf{x}_0 \in U$ et soit $\mathbf{h} \in E$ une **direction** quelconque de E . On s'intéresse à la fonction réelle, $g_{\mathbf{x}_0, \mathbf{h}} : \mathbb{R} \rightarrow F$ définie par

$$g_{\mathbf{x}_0, \mathbf{h}}(t) := f(\mathbf{x}_0 + t\mathbf{h}).$$

Cette fonction est bien définie dans un voisinage de $t = 0$. On peut dériver g , en tant que fonction ayant une seule variable. Cela motive la définition suivante.

Définition 3.10

Si $g_{\mathbf{x}_0, \mathbf{h}}$ est dérivable en $t = 0$, on dit que f admet une **dérivée directionnelle** au point $\mathbf{x}_0 \in U$, dans la direction $\mathbf{h} \in E$. On la note

$$\partial_{\mathbf{h}} f(\mathbf{x}_0) := \frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}_0) := g'_{\mathbf{x}_0, \mathbf{h}}(0) = \lim_{t \rightarrow 0} \frac{f(\mathbf{x}_0 + t\mathbf{h}) - f(\mathbf{x}_0)}{t} \in F.$$

Dans le cas important où $E = \mathbb{R}^d$ et $\mathbf{h} = \mathbf{e}_i$ est le i -ème vecteur de la base canonique de \mathbb{R}^d , on parle plutôt de **dérivée partielle**. On écrit dans ce cas

$$\partial_i f(\mathbf{x}_0) := \frac{\partial f}{\partial x_i}(\mathbf{x}_0) := \frac{\partial f}{\partial \mathbf{e}_i}(\mathbf{x}_0) := \lim_{t \rightarrow 0} \frac{f(x_1, \dots, x_{i-1}, x_i + t, x_{i+1}, \dots, x_d) - f(x_1, \dots, x_d)}{t}.$$

En pratique, pour calculer cette dérivée, on voit les variables $x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_d$ comme fixées, et on dérive "normalement" dans la variable x_i .

Théorème 3.11

Si f est différentiable en \mathbf{x}_0 , alors, pour tout $\mathbf{h} \in E$, on a

$$\frac{\partial f}{\partial \mathbf{h}}(\mathbf{x}_0) = df_{\mathbf{x}_0}(\mathbf{h}) \in F.$$

Ce résultat découle directement de la formule de Taylor. En revanche, il se peut que pour tout $\mathbf{h} \in E$, les dérivées directionnelles $\partial_{\mathbf{h}} f$ existent, mais que f ne soit pas différentiable!

Exemple 3.12. Reprenons la fonction $f(x, y) := \frac{xy}{x^2 + y^2}$ vue dans l'Exercice 2.6. Pour tout $x \in \mathbb{R}$, on a

$$\frac{\partial f}{\partial x}(x, 0) = \lim_{h \rightarrow 0} \frac{f(x + h, 0) - f(x, 0)}{h} = 0,$$

et de même $\frac{\partial f}{\partial y}(0, y) = 0$. En particulier, on a $\frac{\partial f}{\partial x}(0, 0) = \frac{\partial f}{\partial y}(0, 0) = 0$, mais f n'est pas différentiable en $(0, 0)$. Elle n'est même pas continue en $(0, 0)$ (cf Exercice 2.6).

3.2.4 Matrice Jacobienne

On se place maintenant dans le cas où $E = \mathbb{R}^d$ et $F = \mathbb{R}^m$ (le cas utile en pratique). Dans ce cas, l'application $df_{\mathbf{x}_0} \in \mathcal{L}(E, F)$ peut être représentée par une matrice de $\mathcal{M}_{m \times d}(\mathbb{R})$, dans les bases canoniques.

Définition 3.13: Jacobienne

La matrice de $\mathcal{M}_{m \times d}(\mathbb{R})$ qui représente $df_{\mathbf{x}_0}(\mathbf{h})$ dans les bases canoniques s'appellent **matrice Jacobienne**, et est notée

$$J_f(\mathbf{x}_0) \in \mathcal{M}_{m \times d}(\mathbb{R}).$$

Par définition de la matrice représentative, c'est la matrice telle que

$$\forall \mathbf{h} \in \mathbb{R}^d, \quad df_{\mathbf{x}_0}(\mathbf{h}) = J_f(\mathbf{x}_0)\mathbf{h}.$$

Dans le membre de gauche, on a une application linéaire $df_{\mathbf{x}_0}$ qui agit sur le vecteur $\mathbf{h} \in E$, et dans le membre de droite, on a la multiplication de la matrice $J_f(\mathbf{x}_0)$ et du vecteur $\mathbf{h} \in \mathbb{R}^d$.

Dans la suite on note

$$f(\mathbf{x}) = \begin{pmatrix} f_1(x_1, \dots, x_d) \\ f_2(x_1, \dots, x_d) \\ \vdots \\ f_m(x_1, \dots, x_d) \end{pmatrix},$$

les composantes de f dans la base canonique de $F = \mathbb{R}^m$.

Théorème 3.14: Identification de la Jacobienne

Si $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$ est différentiable en $\mathbf{x}_0 \in U$, alors les coefficients de la matrice Jacobienne $J_f(\mathbf{x}_0)$ sont les dérivées partielles des f_i . Plus exactement,

$$J_f(\mathbf{x}_0) = (\partial_{x_1} f, \dots, \partial_{x_d} f) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_d} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \dots & \frac{\partial f_m}{\partial x_d} \end{pmatrix}(\mathbf{x}_0) \in \mathcal{M}_{m \times d}(\mathbb{R}).$$

La preuve du théorème découle directement de la définition de $\partial_{x_i} f$.

Remarque 3.15. *Les composantes de f ne se parlent pas. Chaque composante "vit" dans sa ligne ! On raisonne composante par composante.*

Une autre façon d'écrire les choses est la suivante.

Définition 3.16: Base duale

Pour $1 \leq i \leq d$, on pose $dx_i : \mathbb{R}^d \rightarrow \mathbb{R}$ l'application linéaire définie par

$$dx_i(\mathbf{x}) = x_i.$$

La famille (dx_1, \dots, dx_d) , qui est une famille de $\mathcal{L}(\mathbb{R}^d, \mathbb{R})$, s'appelle parfois la **base duale** de $(\mathbf{e}_1, \dots, \mathbf{e}_d)$.

Autrement dit, dx_i est l'application linéaire qui prend un vecteur $\mathbf{x} \in \mathbb{R}^d$ en entrée, et retourne sa i -ème composante (réelle).

Avec la notation de la base duale, on peut écrire

$$df_{\mathbf{x}_0} = \frac{\partial f}{\partial x_1} dx_1 + \dots + \frac{\partial f}{\partial x_d} dx_d.$$

Il s'agit d'une égalité au sens de $\mathcal{L}(\mathbb{R}^d, \mathbb{R}^m)$, c'est à dire au sens des applications linéaires :

$$\forall \mathbf{h} \in \mathbb{R}^m, \quad df_{\mathbf{x}_0}(\mathbf{h}) = \frac{\partial f}{\partial x_1} h_1 + \dots + \frac{\partial f}{\partial x_d} h_d \in \mathbb{R}^m.$$

Remarque 3.17. *En une dimension $d = 1$, on a $df = f'(x)dx$. Le dx qui apparaît est celui dans les intégrales. En fait, la notation $\int f'(x)dx$ signifie $\int df_{\mathbf{x}}$: on intègre des **formes différentielles**.*

3.2.5 Le gradient

On se place enfin dans le cas où $F = \mathbb{R}$, c'est à dire dans le cas où $f = f(x_1, \dots, x_d)$ est une fonction à valeurs réelles. On rappelle que ce sont les fonctions qu'on peut optimiser.

Dans ce cas, f n'a qu'une seule composante, et on a

$$J_f(\mathbf{x}_0) = (\partial_1 f, \dots, \partial_d f) \in \mathcal{M}_{1,d}(\mathbb{R}).$$

La Jacobienne est donc un **vecteur ligne**.

Définition 3.18: Gradient

Soit $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction réelle différentiable en $\mathbf{x}_0 \in U$. Le **gradient** de f en \mathbf{x}_0 est le **vecteur** de \mathbb{R}^d , noté $\nabla f(\mathbf{x}_0)$ dont les composantes sont

$$\nabla f(\mathbf{x}_0) := \begin{pmatrix} \partial_1 f \\ \partial_2 f \\ \vdots \\ \partial_d f \end{pmatrix}(\mathbf{x}_0).$$

L'importance de ce vecteur vient de la constatation suivante. On rappelle que le **produit scalaire** de \mathbb{R}^d est défini par

$$\forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^d, \quad \langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^d x_i y_i = (x_1 \ x_2 \ \dots \ x_d) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_d \end{pmatrix} = \mathbf{x}^T \mathbf{y} \in \mathbb{R}.$$

Alors on peut écrire

$$df_{\mathbf{x}_0}(\mathbf{h}) = J_f(\mathbf{x}_0)\mathbf{h} = \langle \nabla f(\mathbf{x}_0), \mathbf{h} \rangle.$$

L'avantage de la dernière expression est qu'elle est de nouveau indépendante du choix de la base de \mathbb{R}^d . Le vecteur gradient est un vecteur de \mathbb{R}^d bien défini (qui se trouve avoir les $\partial_i f$ comme coefficients dans la base canonique). En fait, on peut retenir que

Le gradient de f en \mathbf{x}_0 est le vecteur qui pointe vers la plus grande pente montante de f .

Cela vient de l'inégalité de Cauchy-Schwarz :

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{h} \rangle + o(\mathbf{h}) \leq f(\mathbf{x}_0) + \|\nabla f(\mathbf{x}_0)\|_{\mathbb{R}^d} \cdot \|\mathbf{h}\|_{\mathbb{R}^d} + o(\mathbf{h}),$$

avec égalité ssi $\mathbf{h} = \alpha \nabla f(\mathbf{x}_0)$, avec $\alpha > 0$.

3.2.6 Règle de la chaîne

Théorème 3.19: Règle de la chaîne

Soient $f : U \subset E \rightarrow F$ et $g : V \subset F \rightarrow G$ deux fonctions, soit $\mathbf{x}_0 \in U$ telle que f est différentiable en \mathbf{x}_0 , et g est différentiable en $f(\mathbf{x}_0) \in V$. Alors $g \circ f : U \subset E \rightarrow G$ est différentiable en \mathbf{x}_0 , et

$$d(g \circ f)(\mathbf{x}_0) = dg_{f(\mathbf{x}_0)} \circ df_{\mathbf{x}_0}.$$

Dans le cas où $E = \mathbb{R}^d$, $F = \mathbb{R}^m$ et $G = \mathbb{R}^p$, on peut écrire de manière équivalente le produit matriciel des Jacobiennes

$$J_{g \circ f}(\mathbf{x}_0) = J_g(f(\mathbf{x}_0)) \cdot J_f(\mathbf{x}_0).$$

La composée à droite est la composée d'applications linéaires. On note que $df_{\mathbf{x}_0} \in \mathcal{L}(E, F)$ et que $dg_{f(\mathbf{x}_0)} \in \mathcal{L}(F, G)$, donc la composée $dg_{f(\mathbf{x}_0)} \circ df_{\mathbf{x}_0}$ est bien une application linéaire bornée de $\mathcal{L}(E, G)$.

Remarque 3.20. *Les matrices ne commutent pas (ni les applications linéaires). L'ordre est important dans ce théorème. Lorsque f et g sont des fonctions de \mathbb{R} dans \mathbb{R} , on a*

$$(g \circ f)'(x) = g'(f(x)) \cdot f'(x) = f'(x) \cdot g'(f(x)),$$

mais dans le cas général, seule la première expression est valide.

La démonstration de ce résultat est immédiate, il suffit d'écrire les développements de Taylor.

3.3 Fonctions continûment différentiables

Nous avons vu la notion de fonctions différentiables sur un ouvert $U \subset E$. Nous nous demandons maintenant quand ces différentielles sont continues.

Définition 3.21

Soit $f : U \subset E \rightarrow F$ une fonction différentiable sur u , et soit $U \ni \mathbf{x} \mapsto d\mathbf{f}_{\mathbf{x}} \in \mathcal{L}(E, F)$ sa différentielle. On dit que f est **continûment différentiable**, ou **de classe C^1** , sur U si l'application $\mathbf{x} \mapsto d\mathbf{f}_{\mathbf{x}}$ est continue, en tant qu'application de $U \subset E$ dans $\mathcal{L}(E, F)$. On note $C^1(U, F)$ l'ensemble des fonctions continûment différentiables sur U .

Remarquons que l'expression $d\mathbf{f}_{\mathbf{x}}(\mathbf{h})$ dépend à la fois de \mathbf{x} et de \mathbf{h} . Dans les sections précédentes, nous avons vu que à \mathbf{x} fixé, l'application $\mathbf{h} \mapsto d\mathbf{f}_{\mathbf{x}}(\mathbf{h})$ était continue (elle était même linéaire!). Maintenant, on regarde plutôt l'application $\mathbf{x} \mapsto d\mathbf{f}_{\mathbf{x}}(\mathbf{h})$, qui n'a *a priori* aucune raison d'être continue (et encore moins linéaire).

★ **Exemple (fonctions de \mathbb{R} dans \mathbb{R}).** Dans le cas des fonctions $f : \mathbb{R} \rightarrow \mathbb{R}$, on a vu que la différentielle était $d\mathbf{f}_x(h) = f'(x)h$. Cette expression est continue et linéaire en h . Cependant, elle est continue en x ssi $x \mapsto f'(x)$ est continue. On retrouve la notion "classique".

3.3.1 Théorème «fondamental» de l'analyse

On rappelle le résultat suivant, valide pour une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ de classe C^1 .

Théorème 3.22: Théorème fondamental de l'analyse, première forme

Soit $f \in C^1(\mathbb{R}, \mathbb{R})$. Pour tout $a < b$, on a

$$f(b) - f(a) = \int_a^b f'(s) ds.$$

Notons que comme f est de classe C^1 , on a f' continue, et l'intégrale est bien définie (intégrale de Riemann pour les fonctions continues).

Nous en donnons une deuxième forme, équivalente, et très pratique aussi.

Théorème 3.23: Théorème fondamentale de l'analyse, deuxième forme

Soit $f \in C^1(\mathbb{R}, \mathbb{R})$. Pour tout $x \in \mathbb{R}$ et tout $h \in \mathbb{R}$, on a

$$f(x+h) - f(x) = h \int_0^1 f'(x+th) dt.$$

Démonstration. On applique le théorème précédent avec $a = x$ et $b = x + h$, et on a

$$f(x + h) - f(x) = \int_x^{x+h} f'(s) ds.$$

On fait ensuite le changement de variable $s = s(t) = x + th$, donc $ds = h dt$. \square

On étend maintenant ce théorème aux fonctions à plusieurs variables. Soit $f : C^1(U, F)$ une fonction de classe C^1 sur $U \subset E$. Alors pour $\mathbf{x} \in U$ et $\mathbf{h} \in E$, on introduit la fonction $g : \mathbb{R} \rightarrow F$, définie par

$$g_{\mathbf{x}, \mathbf{h}}(t) := f(\mathbf{x} + t\mathbf{h}).$$

Comme f est de classe C^1 , on a g de classe C^1 . En appliquant le théorème fondamental de l'analyse à la fonction g , on obtient

$$g_{\mathbf{x}, \mathbf{h}}(1) - g_{\mathbf{x}, \mathbf{h}}(0) = \int_0^1 g'_{\mathbf{x}, \mathbf{h}}(\mathbf{x} + t\mathbf{h}) dt = \int_0^1 df_{\mathbf{x}+t\mathbf{h}} \cdot \mathbf{h} dt,$$

où on a utilisé le Théorème 3.11 pour la dernière égalité. On en déduit le théorème suivant.

Théorème 3.24: Théorème fondamental de l'analyse

Soit $f \in C^1(U, F)$ avec $U \subset E$ ouvert convexe. Pour tout $x \in U$ et tout $\mathbf{h} \in E$ tel que $\mathbf{x} + \mathbf{h} \in U$, on a

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \left(\int_0^1 df_{\mathbf{x}+t\mathbf{h}} dt \right) \cdot \mathbf{h}.$$

Dans le cas où $E = \mathbb{R}^d$ et $F = \mathbb{R}^m$, on peut aussi écrire

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \left(\int_0^1 J_f(\mathbf{x} + t\mathbf{h}) dt \right) \times \mathbf{h}.$$

Enfin, dans le cas où $m = 1$ (f est une fonction réelle), on a aussi

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t\mathbf{h}), \mathbf{h} \rangle dt.$$

On rappelle qu'on peut intégrer des matrices et des vecteurs, en intégrant "composante par composant".

3.3.2 Théorème des accroissements finis

On commence par rappeler le théorème de la moyenne.

Lemme 3.25. Si $g \in C^0(\mathbb{R}, \mathbb{R})$ est continue à valeurs réelles, alors pour tout $a < b$, il existe $c \in [a, b]$ telle que

$$\frac{1}{b-a} \int_a^b g(s) ds = g(c).$$

Démonstration. La fonction g est continue sur le compact $[a, b]$, donc est atteint son minimum m et son maximum M en x_m et x_M respectivement, avec $x_m, x_M \in [a, b]$. On a $m \leq g(x) \leq M$ pour tout $x \in [a, b]$, donc

$$m \leq \frac{1}{|b-a|} \int_a^b g(s) ds \leq M.$$

D'après le théorème des valeurs intermédiaires pour les fonctions continues, il existe un point $c \in [x_m, x_M]$ telle que

$$\frac{1}{|b-a|} \int_a^b g(s) ds = g(c),$$

ce qui conclut la preuve. \square

En prenant $g(t) = \langle \nabla f(\mathbf{x} + t\mathbf{h}), \mathbf{h} \rangle$ dans le troisième cas du Théorème 3.24, on obtient le résultat suivant.

Théorème 3.26: Théorème des accroissements finis dans \mathbb{R}^d

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction réelle ($m = 1$) de classe C^1 . Alors, pour tout $\mathbf{x} \in \mathbb{R}^d$ et $\mathbf{h} \in \mathbb{R}^d$, il existe $t_* \in [0, 1]$ telle que

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \langle \nabla f(\mathbf{x} + t_*\mathbf{h}), \mathbf{h} \rangle.$$

Il est important ici que la fonction soit réelle, c'est à dire qu'elle n'a qu'une seule composante. Un résultat similaire pour des fonctions f avec plusieurs composantes ne peut pas être vraie. Si $f = (f_1, f_2)$ est une fonction à deux composantes, on peut appliquer le théorème des accroissements finis à f_1 et f_2 , et trouver des paramètres t_1 et t_2 pour ces deux composantes. En revanche, il n'y a aucune raison a priori pour que $t_1 = t_2$...

3.3.3 Caractérisation des fonctions de classe C^1

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}^m$ une fonction de classe C^1 . Alors toutes les dérivées partielles sont continues, et en particulier, la fonction Jacobienne $\mathbf{x} \mapsto J_f(\mathbf{x})$ est continue (de \mathbb{R}^d dans $\mathcal{M}_{m \times d}(\mathbb{R})$), ou, de manière équivalente, toutes les dérivées partielles sont continues. En fait, la réciproque est vraie.

Théorème 3.27: Caractérisation des fonctions C^1

Soit $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}^m$. La fonction f est de classe C^1 sur U ssi toutes les dérivées partielles $\frac{\partial f}{\partial x_i}$ existent et sont continues sur U .

On rappelle que ce n'est pas parce que les dérivées partielles de f en \mathbf{x}_0 existent que f est différentiable en \mathbf{x}_0 (cf Exemple 3.12).

Démonstration. Il suffit de démontrer le théorème dans le cas $m = 1$ (on travaille composante par composante). On suppose que toutes les dérivées partielles $\frac{\partial f}{\partial x_i}$ existent et sont continues sur U . Alors ∇f est aussi continue sur U .

Soit $\mathbf{x}_0 \in U$, et soit $\varepsilon > 0$ tel que $\mathcal{B}(\mathbf{x}_0, \varepsilon) \subset U$ (on rappelle que U est ouvert). D'après le théorème des valeurs intermédiaires, pour tout $\mathbf{h} \in \mathcal{B}(\mathbf{0}, \varepsilon)$, on a $\mathbf{x}_0 + \mathbf{h} \in U$, et il existe $t_{\mathbf{h}} \in [0, 1]$ tel que

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \langle \nabla f(\mathbf{x}_0 + t_{\mathbf{h}}\mathbf{h}), \mathbf{h} \rangle.$$

On a donc

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + \langle \nabla f(\mathbf{x}_0), \mathbf{h} \rangle + \|\mathbf{h}\| \mathcal{E}(\mathbf{h}), \quad (3.2)$$

avec l'erreur

$$\mathcal{E}(\mathbf{h}) := \frac{1}{\|\mathbf{h}\|} \langle \nabla f(\mathbf{x}_0 + t_{\mathbf{h}}\mathbf{h}) - \nabla f(\mathbf{x}_0), \mathbf{h} \rangle.$$

D'après l'inégalité de Cauchy-Schwarz, on a

$$|\mathcal{E}(\mathbf{h})| \leq \|\nabla f(\mathbf{x}_0 + t_{\mathbf{h}}\mathbf{h}) - \nabla f(\mathbf{x}_0)\|_{\mathbb{R}^m}.$$

Par continuité de ∇f , et comme $t_{\mathbf{h}}\mathbf{h} \rightarrow \mathbf{0}$ lorsque $\mathbf{h} \rightarrow \mathbf{0}$ (on rappelle que $t_{\mathbf{h}} \in [0, 1]$ est borné), on a $\mathcal{E}(\mathbf{h}) \rightarrow 0$ lorsque $\mathbf{h} \rightarrow \mathbf{0}$. Ceci montre que la formule (3.2) est un développement de Taylor de f à l'ordre 1. On en déduit que f est différentiable en \mathbf{x}_0 , et que sa différentielle est l'application linéaire

$$df_{\mathbf{x}_0} : \mathbf{h} \mapsto \langle \nabla f(\mathbf{x}_0), \mathbf{h} \rangle.$$

Ceci étant vrai pour $\mathbf{x}_0 \in U$, on a que f est différentiable sur U . De plus, sa différentielle est continue (en \mathbf{x}), car $\mathbf{x} \mapsto \nabla f(\mathbf{x})$ est continue, donc f est de classe C^1 . \square

On s'intéresse maintenant aux dérivées supérieures de f . Dans le cas de la dérivée première, on a vu que l'objet mathématique qui apparaissait naturellement était les applications linéaires. Pour les dérivées secondes, on fera naturellement apparaître des applications quadratiques.

4.1 Fonctions deux fois différentiables

4.1.1 Premières définitions

Définition 4.1: Différentielles supérieures

Soit $f : U \subset E \rightarrow F$ une fonction de classe C^1 sur U , de différentielle $df = U \rightarrow \mathcal{L}(E, F)$. On dit que f est deux fois différentiable sur U si df est différentiable sur U . On dit que f est de classe C^2 sur U si df est de classe C^1 sur U .

Par induction, on peut dire que f est de classe C^k sur U si df est de classe C^{k-1} sur U .

On remarquera que df est une fonction de E dans $F_1 := \mathcal{L}(E, F)$. Dans le cas où $E = \mathbb{R}^d$ et $F = \mathbb{R}^m$, cet espace F_1 est de dimension $d \times m$. Il faut vérifier que $\mathbf{x} \mapsto df_{\mathbf{x}}$ est de classe C^1 en tant que fonction de d variables et $d \times m$ composantes. On note d^2f la différentielle de df . Si df est différentiable en \mathbf{x} , on a, par définition

$$df_{\mathbf{x}+\mathbf{h}} = df_{\mathbf{x}} + d^2f_{\mathbf{x}}(\mathbf{h}) + \|\mathbf{h}\|\mathcal{E}(\mathbf{h}),$$

qui est une égalité dans $\mathcal{L}(E, F)$. En particulier, $\mathcal{E}(\mathbf{h})$ est ici une fonction linéaire de $\mathcal{L}(E, F)$, avec $\|\mathcal{E}(\mathbf{h})\|_{\text{op}} \rightarrow 0$ lorsque $\mathbf{h} \rightarrow \mathbf{0}$. De plus, l'application $\mathbf{h} \ni \mathbb{R}^d \mapsto d^2f_{\mathbf{x}}(\mathbf{h}) \in \mathcal{L}(E, F)$ est linéaire par définition d'une différentielle. Deux formes linéaires sont égales si elles coïncident sur tous les arguments. Ainsi, on peut écrire de manière équivalente

$$\forall \mathbf{k} \in \mathbb{R}^d, \quad df_{\mathbf{x}+\mathbf{h}}(\mathbf{k}) = df_{\mathbf{x}}(\mathbf{k}) + d^2f_{\mathbf{x}}(\mathbf{h})(\mathbf{k}) + \|\mathbf{h}\|\mathcal{E}(\mathbf{h})(\mathbf{k}).$$

L'application $(\mathbf{h}, \mathbf{k}) \mapsto d^2f_{\mathbf{x}}(\mathbf{h}, \mathbf{k})$ est linéaire en \mathbf{h} et en \mathbf{k} , donc est une application bilinéaire. On écrit souvent $d^2f_{\mathbf{x}}(\mathbf{h}, \mathbf{k})$ cette application bilinéaire de $\mathbb{R}^d \times \mathbb{R}^d$ dans \mathbb{R}^m . En anticipant un peu sur la suite, cette application bilinéaire est moralement

$$d^2f_{\mathbf{x}}(\mathbf{h}, \mathbf{k}) = \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}) h_i k_j.$$

4.1.2 Dérivées croisées

Dans la suite, on s'intéressera principalement au cas où $E = \mathbb{R}^d$ et $F = \mathbb{R}$. Dans le cas où $F = \mathbb{R}^m$, on pourra travailler *composante par composante*.

Définition 4.2: Dérivées partielles d'ordre supérieure

Soit $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction deux fois différentiable. Pour $1 \leq i, j \leq d$ on note

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) := \partial_{x_i x_j}^2 f(\mathbf{x}) := \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right) (\mathbf{x}).$$

Par exemple, si $f(x, y) = \cos(x)e^{y^2}$, on a $\partial_x f = -\sin(x)e^{y^2}$, et $\partial_{yx}^2 f = -2y \sin(x)e^{y^2}$.

4.1.3 Fonctions de classe C^2

On rappelle le Théorème 3.27 qui dit qu'une fonction est C^1 si toutes les dérivées partielles sont continues. En appliquant ce résultat à la fonction df , on obtient la caractérisation suivante.

Théorème 4.3: Caractérisation des fonctions de classe C^2

Soit $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$. La fonction f est de classe C^2 sur U ssi toutes les dérivées secondes existent et sont continues sur U .

On remarquera qu'il y a d^2 fonctions $\partial_{x_i x_j}^2 f$ lorsque i et j varient entre 1 et d . De même, on peut vérifier qu'une fonction f est de classe C^3 si toutes les dérivées troisièmes $\partial_{x_i x_j x_k}^3 f(\mathbf{x})$ sont continues, et il y a d^3 telles fonctions. Et ainsi de suite.

On dit qu'une fonction f est de classe C^∞ si elle est de classe C^k pour tout $k \in \mathbb{N}$. En pratique, toutes les fonctions *usuelles* sont C^∞ sur leur domaine (ouvert) de définition.

Le théorème suivant affirme que l'ordre de dérivation dans les dérivées croisées n'importe pas dès que f est suffisamment régulière.

Théorème 4.4: Théorème de Schwarz (admis)

Si $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$ est de classe C^2 sur U , alors, pour tout $\mathbf{x} \in U$, on a

$$\forall 1 \leq i, j \leq d, \quad \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}).$$

L'Exercice suivant montre qu'il n'y pas toujours égalité lorsque f n'est que deux fois différentiable (au lieu de C^2).

Exercice 4.5

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) := \frac{x^3 y}{x^2 + y^2}, \quad f(0, 0) = 0.$$

Montrer que $\partial_{xy}^2 f \neq \partial_{yx}^2 f$.

4.1.4 La Hessienne

Définition 4.6: Matrice Hessienne

Soit $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction de classe C^2 . La Hessienne de f est la matrice de taille $d \times d$ contenant les dérivées croisées de f . Explicitement,

$$H_f(\mathbf{x}) := \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_d} & \cdots & \cdots & \frac{\partial^2 f}{\partial x_d^2} \end{pmatrix}(\mathbf{x}).$$

D'après le théorème de Schwarz, cette matrice est **symétrique**.

On pourra remarquer que *la Hessienne est la Jacobienne du gradient!* En effet, on rappelle que

$$\nabla f(\mathbf{x}) = \begin{pmatrix} \partial_{x_1} f \\ \vdots \\ \partial_{x_d} f \end{pmatrix}(\mathbf{x})$$

qui peut-être vue comme une fonction de \mathbb{R}^d dans \mathbb{R}^d (le gradient est bien un vecteur colonne). En prenant la Jacobienne de cette application, on retrouve bien la Hessienne de f .

4.1.5 Formule de Taylor-Young à l'ordre 2

Nous avons vu que la formule de Taylor à l'ordre 1 permettait de **définir** la différentielle, puis la notion d'être C^1 etc. Nous démontrons maintenant une formule de Taylor à l'ordre supérieur.

Théorème 4.7: Formule de Taylor à l'ordre 2

Si $f : \mathbb{R}^d \rightarrow \mathbb{R}$ est de classe C^2 , alors

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

On retrouve la partie linéaire en \mathbf{h} dans $\langle \nabla f(\mathbf{x}), \mathbf{h} \rangle$. La partie suivante est une partie quadratique en \mathbf{h} , dans le sens où $\mathbf{h} \mapsto \frac{1}{2} \langle \mathbf{h}, H_f(\mathbf{x}) \mathbf{h} \rangle$ est quadratique en \mathbf{h} . Ici, la notation $o(\mathbf{h}^2)$ est un abus de langage (que signifie un vecteur au carré?) qui signifie une fonction de type

$$o(\mathbf{h}^2) = \|\mathbf{h}\|^2 \mathcal{E}(\mathbf{h}), \quad \text{avec } \mathcal{E}(\mathbf{h}) \xrightarrow{\mathbf{h} \rightarrow 0} 0.$$

Démonstration. On commence par appliquer le théorème fondamental de l'analyse à la fonction f (Théorème 3.24), et on obtient

$$f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) = \int_0^1 \langle \nabla f(\mathbf{x} + t\mathbf{h}), \mathbf{h} \rangle dt = \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle + \int_0^1 \langle \nabla f(\mathbf{x} + t\mathbf{h}) - \nabla f(\mathbf{x}), \mathbf{h} \rangle dt.$$

Par ailleurs, en appliquant le théorème fondamental à la fonction ∇f , et en rappelant que la Hessienne est la Jacobienne du gradient, on a

$$\nabla f(\mathbf{x} + t\mathbf{h}) - \nabla f(\mathbf{x}) = \left(\int_0^1 H_f(\mathbf{x} + st\mathbf{h}) ds \right) \times t\mathbf{h}.$$

On en déduit que

$$\begin{aligned} f(\mathbf{x} + \mathbf{h}) - f(\mathbf{x}) &= \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle + \int_0^1 \int_0^1 \langle \mathbf{h}, H_f(\mathbf{x} + st\mathbf{h})t\mathbf{h} \rangle dsdt \\ &= \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle + \int_0^1 \int_0^1 \langle \mathbf{h}, H_f(\mathbf{x})t\mathbf{h} \rangle dsdt + \int_0^1 \int_0^1 \langle \mathbf{h}, [H_f(\mathbf{x} + st\mathbf{h}) - H_f(\mathbf{x})]t\mathbf{h} \rangle dsdt. \end{aligned}$$

Dans la première intégrale, on a $\int_0^1 t dt = \frac{1}{2}$. Le dernier terme est de la forme

$$\|\mathbf{h}\|^2 \mathcal{E}(\mathbf{h}), \quad \text{avec} \quad \mathcal{E}(\mathbf{h}) = \int_0^1 \int_0^1 \left\langle \frac{\mathbf{h}}{\|\mathbf{h}\|}, [H_f(\mathbf{x} + st\mathbf{h}) - H_f(\mathbf{x})]t \frac{\mathbf{h}}{\|\mathbf{h}\|} \right\rangle dsdt$$

et vérifie, par Cauchy-Schwarz et la propriété des normes opérateurs,

$$|\mathcal{E}(\mathbf{h})| \leq \int_0^1 \int_0^1 \|H_f(\mathbf{x} + st\mathbf{h}) - H_f(\mathbf{x})\|_{\text{op}} t dsdt,$$

qui tend vers 0 lorsque $\mathbf{h} \rightarrow \mathbf{0}$ par continuité de la Hessienne. Au final, on a bien trouvé

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

□

4.2 Application à l'optimisation de fonctions

La formule de Taylor Théorème 4.7 donne la meilleure approximation de f par une forme quadratique. Or il est facile de déterminer les minima des formes quadratiques. Dans la suite, nous rappelons les propriétés des formes quadratiques, et appliquons les résultats pour la recherche de minima.

4.2.1 Rappels sur les matrices symétriques

On a vu que la Hessienne d'une fonction de classe C^2 était une matrice symétrique. On rappelle qu'une matrice symétrique $A \in \mathcal{M}_{d \times d}(\mathbb{R})$ est diagonalisable. Cela signifie qu'il existe une matrice orthogonale $U \in O(d)$ et une matrice diagonale $D = \text{diag}(\lambda_1, \dots, \lambda_d)$ telle que

$$A = UDU^T.$$

Comme $U^T U = U U^T = \mathbb{I}_d$, on a $AU = UD$. En écrivant $U = (\mathbf{u}_1, \dots, \mathbf{u}_d)$ les colonnes de U , la relation $AU = UD$ montre que

$$\forall 1 \leq i \leq d, \quad A\mathbf{u}_i = \lambda_i \mathbf{u}_i,$$

autrement dit, les vecteurs \mathbf{u}_i sont des vecteurs propres de A pour les valeurs propres λ_i . De plus, la relation $U^T U = \mathbb{I}_d$ montre que les vecteurs $(\mathbf{u}_1, \dots, \mathbf{u}_d)$ forment une base orthonormale. Autrement dit, on a trouvé une base orthonormale composée de vecteurs propres de A .

Définition 4.8: Matrices positives, négatives, ...

On note $\mathcal{S}_d(\mathbb{R})$ l'ensemble des matrices symétrique de taille $d \times d$. Pour $A \in \mathcal{S}_d(\mathbb{R})$, on note $\lambda_1(A) \leq \lambda_2(A) \leq \dots \leq \lambda_d(A)$ les valeurs propres de A , rangées dans l'ordre croissant.

On dit que A est **dégénérée** si $\lambda_j(A) = 0$ pour un certain $1 \leq j \leq d$.

On dit que A est **positive** si $\lambda_j(A) \geq 0$ pour tout $1 \leq j \leq d$.

On dit que A est **définie positive** si $\lambda_j(A) > 0$ pour tout $1 \leq j \leq d$.

On note $\mathcal{S}_d^+(\mathbb{R})$ l'ensemble des matrices positives, et $\mathcal{S}_d^{++}(\mathbb{R})$ l'ensemble des matrices définies positives.

On définit de même les matrices négatives et définies négatives. Une matrice A est définie positive ssi elle est positive et non dégénérée. On remarquera que A est dégénérée ssi $\text{Ker}(A) \neq \{0\}$ (car 0 est une valeur propre de A), ssi A est non inversible.

Théorème 4.9: Caractérisation des matrices positives

Soit $A \in \mathcal{S}_d(\mathbb{R})$. On a toujours l'inégalité

$$\forall \mathbf{x} \in \mathbb{R}^d, \quad \langle \mathbf{x}, A\mathbf{x} \rangle \geq \lambda_1(A) \|\mathbf{x}\|^2,$$

avec égalité ssi $\mathbf{x} \in \text{Ker}(A - \lambda_1(A))$.

En particulier, A est positive ssi $\langle \mathbf{x}, A\mathbf{x} \rangle \geq 0$ pour tout $\mathbf{x} \in \mathbb{R}^d$, et A est définie positive ssi $\langle \mathbf{x}, A\mathbf{x} \rangle > 0$ pour tout $\mathbf{x} \in \mathbb{R}^d \setminus \{0\}$.

Démonstration. Soit $\mathbf{x} \in \mathbb{R}^d$. On décompose \mathbf{x} dans la base des (\mathbf{u}_i) , et on écrit

$$\mathbf{x} = x_1 \mathbf{u}_1 + x_2 \mathbf{u}_2 + \cdots + x_d \mathbf{u}_d.$$

En appliquant A , et en utilisant le fait que les \mathbf{u}_i sont des vecteurs propres, on obtient

$$A\mathbf{x} = x_1 \lambda_1 \mathbf{u}_1 + x_2 \lambda_2 \mathbf{u}_2 + \cdots + x_d \lambda_d \mathbf{u}_d.$$

Ainsi, en utilisant que les (\mathbf{u}_i) sont orthonormés, on obtient

$$\langle \mathbf{x}, A\mathbf{x} \rangle = \sum_{i=1}^d |x_i|^2 \lambda_i \geq \lambda_1 \sum_{i=1}^d |x_i|^2 = \lambda_1 \|\mathbf{x}\|^2.$$

□

En règle générale, pour déterminer si une matrice A est positive, il faut la diagonaliser pour trouver ses valeurs propres, ce qui peut être fastidieux. Dans le cas des matrices 2×2 , on a un critère simple pour savoir si une matrice est positive.

Théorème 4.10: Matrices positives de taille 2×2

Une matrice $A \in \mathcal{S}_2(\mathbb{R})$ est positive ssi sa trace et son déterminant sont positifs. Elle est non dégénérée si, de plus, son déterminant est non nul.

Démonstration. On a $\text{Tr}(A) = \lambda_1 + \lambda_2$ (somme des valeurs propres) et $\det(A) = \lambda_1 \lambda_2$ (produit des valeurs propres). Or deux nombres sont positifs ssi la somme et le produit de ces nombres sont positifs. Le résultat suit. □

Attention : le résultat est faux en dimension $d \geq 3$. Par exemple, la matrice

$$A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

vérifie $\text{Tr}(A) = 1 \geq 0$ et $\det(A) = 3 > 0$, mais a deux valeurs propres négatives.

4.2.2 Caractérisation des points critiques non dégénérés

Nous sommes enfin prêts pour trouver des critères “simple” pour déterminer les minima des fonctions.

Théorème 4.11

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction de classe C^2 . Si \mathbf{x}_* est un minimiseur local de f , alors

$$\nabla f(\mathbf{x}_*) = \mathbf{0}, \quad \text{et} \quad H_f(\mathbf{x}_*) \geq 0.$$

Démonstration. On a déjà montré que si \mathbf{x}_* est un minimiseur local, alors \mathbf{x}_* est un point critique, donc $\nabla f(\mathbf{x}_*) = \mathbf{0}$. Par ailleurs, soit $\varepsilon > 0$ tel que \mathbf{x}_* est un minimiseur de f dans $\mathcal{B}(\mathbf{x}_*, \varepsilon)$. Soit $\mathbf{h} \in \mathbb{R}^d$ quelconque. Pour tout $0 \leq t < \varepsilon \|\mathbf{h}\|^{-1}$, on a $\|\mathbf{th}\| \leq t\|\mathbf{h}\| \leq \varepsilon$, et donc $\mathbf{th} \in \mathcal{B}(\mathbf{0}, \varepsilon)$. Ainsi

$$f(\mathbf{x}_*) \leq f(\mathbf{x}_* + \mathbf{th}) = f(\mathbf{x}_*) + \frac{1}{2}t^2 \langle \mathbf{h}, H_f(\mathbf{x}_*)\mathbf{h} \rangle + t^2 \mathbf{h}^2 \mathcal{E}(\mathbf{th}).$$

avec $\mathcal{E}(\mathbf{th}) \rightarrow 0$ lorsque $t \rightarrow 0$. En simplifiant par $f(\mathbf{x}_*)$ et en divisant par t^2 , on obtient

$$\forall 0 \leq t < \varepsilon \|\mathbf{h}\|^{-1}, \quad \frac{1}{2} \langle \mathbf{h}, H_f(\mathbf{x}_*)\mathbf{h} \rangle + \mathbf{h}^2 \mathcal{E}(\mathbf{th}) \geq 0.$$

On prend la limite $t \rightarrow 0$, et on trouve $\langle \mathbf{h}, H_f(\mathbf{x}_*)\mathbf{h} \rangle \geq 0$. Ceci étant vrai pour tout $\mathbf{h} \in \mathbb{R}^d$, on conclut que $H_f(\mathbf{x}_*)$ est une matrice symétrique positive. \square

Définition 4.12: Points critiques non dégénérés

On dit que \mathbf{x}_* est un point critique non dégénéré de f si $\nabla f(\mathbf{x}_*) = \mathbf{0}$ et si $H_f(\mathbf{x}_*)$ est non dégénérée.

Si \mathbf{x}_* est un point critique non dégénéré qui est ni un minimum local, ni un maximum local, on dit que \mathbf{x}_* est un **point selle**.

En particulier, si \mathbf{x}_* est un minimiseur non dégénéré, on a $H_f(\mathbf{x}_*) > 0$. On s'intéresse maintenant à la *réciproque* du résultat précédent.

Théorème 4.13

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction de classe C^2 . Si $\mathbf{x}_* \in \mathbb{R}^d$ est tel que

$$\nabla F(\mathbf{x}_*) = \mathbf{0}, \quad \text{et} \quad H_f(\mathbf{x}_*) > 0,$$

alors \mathbf{x}_* est un minimiseur local non dégénéré de F .

On remarquera qu'on demande $H_f(\mathbf{x}_*)$ d'être définie positive dans le deuxième cas pour pouvoir conclure. Dans le cas $d = 1$ par exemple, les fonctions

$$f_1(x) = x^3, \quad \text{et} \quad f_2 = x^4$$

vérifient toutes les deux $f'(0) = f''(0) = 0$. Dans le premier cas, le point 0 n'est pas un minimiseur de f_1 , mais 0 est un minimiseur de f_2 .

Démonstration. Soit $\lambda_1 > 0$ la plus petite valeur propre de $H_f(\mathbf{x}_*)$. On a

$$f(\mathbf{x}_* + \mathbf{h}) \geq f(\mathbf{x}_*) + \frac{1}{2}\lambda_1\|\mathbf{h}\|^2 + \|\mathbf{h}\|^2 E(\mathbf{h}).$$

On rappelle que E est continue en $\mathbf{h} = \mathbf{0}$ avec $E(\mathbf{0}) = 0$. Soit $\varepsilon > 0$ suffisamment petit pour que $E(\mathbf{h}) > -\frac{1}{2}\lambda_1$ pour tout $\|\mathbf{h}\| < \varepsilon$. Alors, pour tout $\mathbf{h} \in \mathcal{B}(\mathbf{0}, \varepsilon) \setminus \{\mathbf{0}\}$, on a

$$f(\mathbf{x}_* + \mathbf{h}) \geq f(\mathbf{x}_*) + \frac{1}{4}\lambda_1\|\mathbf{h}\|^2 > f(\mathbf{x}_*),$$

ce qui prouve que \mathbf{x}_* est un minimiseur strict dans $\mathcal{B}(\mathbf{x}_*, \varepsilon)$. \square

4.2.3 Un exemple de recherche de minimiseur

On conclut cette section en expliquant comment étudier une fonction à plusieurs variables en pratique. On s'intéressera principalement aux fonctions à deux variables. Un exercice typique est le suivant.

Exercice 4.14: Exemple type

Étudier les points critiques de la fonction

$$f(x, y) := (x^2 + 2y^2)e^{-(x^2+y^2)}.$$

Étude de la régularité

La fonction f est de classe C^∞ sur \mathbb{R}^2 comme composée de fonctions usuelles C^∞ sur leur domaine de définition.

Calculs des dérivées premières

On a

$$\frac{\partial f}{\partial x}(x, y) = e^{-(x^2+y^2)} [2x - 2x(x^2 + 2y^2)] = 2xe^{-(x^2+y^2)} [1 - x^2 - 2y^2].$$

De même,

$$\frac{\partial f}{\partial y}(x, y) = e^{-(x^2+y^2)} [4y - 2y(x^2 + 2y^2)] = 2ye^{-(x^2+y^2)} [2 - x^2 - 2y^2].$$

Recherche des points critiques

Pour trouver les points critiques, on résout le système

$$\begin{cases} \partial_x f(x, y) = 0 \\ \partial_y f(x, y) = 0. \end{cases} \iff \begin{cases} x [1 - x^2 - 2y^2] = 0 \\ y [2 - x^2 - 2y^2] = 0, \end{cases} \iff \begin{cases} x = 0 & \text{ou} & x^2 + 2y^2 = 1 \\ y = 0 & \text{ou} & x^2 + 2y^2 = 2. \end{cases}$$

On en déduit qu'il y a 5 points critiques, à savoir les points

$$(0, 0), \quad (1, 0), \quad (-1, 0), \quad (0, 1) \quad \text{et} \quad (0, -1).$$

Calcul de la Hessienne

On a

$$\begin{aligned} \frac{\partial^2 f}{\partial x^2} &= e^{-(x^2+y^2)} [2 - 6x^2 - 4y^2 - 2x(2x - 2x^3 - 4xy^2)], \\ \frac{\partial^2 f}{\partial y^2} &= e^{-(x^2+y^2)} [4 - 2x^2 - 12y^2 - 2y(4y - 2yx^2 - 4y^3)], \\ \frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial x \partial y} = e^{-(x^2+y^2)} [-8xy - 2y(x^2 - 2x^3 - 4xy^2)], \end{aligned}$$

où on a utilisé le Lemme de Schwarz pour l'égalité $\partial_{xy}^2 f = \partial_{yx}^2 f$. L'expression de la hessienne est donc assez complexe. Cependant, on veut l'évaluer uniquement aux points critiques.

Étude des points critiques

On trouve

$$H_f(0,0) = \begin{pmatrix} 2 & 0 \\ 0 & 4 \end{pmatrix}, \quad H_f(1,0) = H_f(-1,0) = e^{-1} \begin{pmatrix} -4 & 0 \\ 0 & 2 \end{pmatrix}, \quad H_f(0,-1) = H_f(0,1) = e^{-1} \begin{pmatrix} -2 & 0 \\ 0 & -8 \end{pmatrix}.$$

Dans cet exemple, les hessiennes sont diagonales, et il est facile de déterminer la nature des points critiques.

- La hessienne au point $(0,0)$ est positive, donc $(0,0)$ est un **minimum local**. De plus, il est facile de voir que $f > 0$ et que $f(0,0) = 0$. Donc $(0,0)$ est un minimum global!
- La hessienne aux points $(\pm 1,0)$ ont une valeur propre positive et une valeur propre négative. Ces points sont donc des **points selles**.
- La hessienne aux points $(0,\pm 1)$ est négative, donc ces points sont des **maxima locaux**. En fait, on peut montrer que ce sont des maxima globaux, mais c'est plus difficile. Cela vient du fait que la fonction f tend vers 0 en l'infini.

4.2.4 Un autre exemple type

Exercice 4.15: Exemple type (bis)

Étudier les points critiques de la fonction

$$f(x,y) = x^3 + 3xy^2 - 15x - 12y.$$

Étude de la régularité

La fonction f est de classe C^∞ sur \mathbb{R}^2 , car f est un polynôme.

Calculs des dérivées premières

On a

$$\frac{\partial f}{\partial x}(x,y) = 3x^2 + 3y^2 - 15, \quad \text{et} \quad \frac{\partial f}{\partial y}(x,y) = 6xy - 12.$$

Recherche des points critiques

Pour trouver les points critiques, on résout le système

$$\begin{cases} \partial_x f(x,y) = 0 \\ \partial_y f(x,y) = 0. \end{cases} \iff \begin{cases} x^2 + y^2 = 5 \\ xy = 2 \end{cases} \iff \begin{cases} (x+y)^2 = 9 \\ (x-y)^2 = 1 \end{cases} \iff \begin{cases} x+y = \pm 3 \\ x-y = \pm 1 \end{cases}$$

Il y a donc 4 points critiques possibles, suivant les signes qu'on met. On trouve les 4 points critiques

$$(1,2) \quad (-1,-2) \quad (2,1) \quad (-2,-1).$$

Calcul de la Hessienne

On trouve directement

$$H_f(x,y) = 6 \begin{pmatrix} x & y \\ y & x \end{pmatrix}.$$

On remarque que la trace $\text{Tr}(H_f) = 12x$ est du signe de x , et que $\det(H_f) = 36(x^2 - y^2)$ est positif si $|x| > |y|$ et négatif si $|x| < |y|$. On trouve donc

- Pour le point $(1,2)$, la hessienne a une valeur propre positive, et une valeur propre négative. Donc le point $(1,2)$ est un point selle.

- Pour le point $(-1, -2)$, idem.
- Pour le point $(2, 1)$, la hessienne est positive, donc $(2, 1)$ est un minimum local de f .
- Pour le point $(-2, -1)$, la hessienne est négative, donc $(-2, -1)$ est un maximum local de f .