

Maths pour physiciens

Cours de Frédéric van Wijland. Notes de Martin Teuscher.

Introduction

Cours de Frédéric van Wijland, enseignant à l'Université de Paris (Paris VII) (bureau 724A bâtiment Condorcet), laboratoire « matière et systèmes complexes » ; responsable du parcours de physique théorique M2/ICFP.

Adresse mail : fvw@univ-paris-diderot.fr

Les mathématiques présentées dans ce cours seront des mathématiques utiles à court terme :

- Physique quantique : algèbre linéaire, équations différentielles ou aux dérivées partielles, analyse complexe
- Physique statistique : probabilités, statistiques, analyse complexe
- Mécanique analytique : fonctions de plusieurs variables, extrema liés, analyse complexe
- Expérimentations : théorie de Fourier

Les groupes et la représentation des groupes ne seront pas abordés dans ce cours : on pourra mettre à profit les cours d'algèbre I et II dispensés par le département de mathématiques de l'ENS.

En outre, on pourra trouver un plan du cours et des références bibliographiques sur la page web de Frédéric van Wyland.

Bibliographie courte

- Walter Appel : *Mathématiques pour la physique et les physiciens*
- Jean Dieudonné : *Calcul infinitésimal*
- Laurent Schwartz
- A. Alastery, M. Magro, P. Pujol

Chapitre 1

Intégrations et probabilités

1.1 Espaces probabilisables

Le but de cette section est de modéliser le résultat d'une expérience nous donnant un nombre « aléatoire ». On introduit pour cela un ensemble Ω appelé l'**univers**, représentant les différents résultats possibles de l'expérience aléatoire.

Exemple 1.1 : Dans un jeu de pile ou face, on s'intéresse au nombre de lancers jusqu'au premier pile : $\Omega = \mathbb{N} \cup \{+\infty\}$.

Définition 1.1 : Soit Ω un ensemble. On appelle **tribu** sur Ω toute partie $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ vérifiant :

- (i) $\emptyset \in \mathcal{A}, \Omega \in \mathcal{A}$
- (ii) $\forall A \in \mathcal{A}, A^C = \Omega \setminus A \in \mathcal{A}$ (stabilité par passage au complémentaire)
- (iii) $\forall (A_i)_{i \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}, \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ (stabilité par union dénombrable)

Les éléments de la tribu sont appelés les **événements**.

La donnée du couple (Ω, \mathcal{A}) est appelé **espace probabilisable** ou **mesurable**.

Remarques 1.1 :

- À une propriété que peut vérifier le résultat d'une expérience, on associe une partie de l'univers (les $\omega \in \Omega$ pour lesquels elle est vérifiée) : c'est un événement. Une probabilité sera une application qui à un événement associera un réel de $[0; 1]$.
- Lorsque Ω est fini ou dénombrable, on choisira toujours $\mathcal{P}(\Omega)$ comme tribu. Lorsque Ω est indénombrable, l'ensemble des événements sera une sous-partie de $\mathcal{P}(\Omega)$.
- Les points (ii) et (iii) impliquent que si \mathcal{A} est une tribu, on a également :

$$\forall (A_i)_{i \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}, \bigcap_{i \in \mathbb{N}} A_i \in \mathcal{A}$$

Définition 1.2 : Si $X \subseteq \mathcal{P}(\Omega)$, on appelle **tribu engendrée par X** , notée $\sigma(X)$, la plus petite tribu sur Ω contenant X . Or, on peut montrer aisément qu'une intersection quelconque

de tribus est une tribu, et donc :

$$\sigma(X) = \bigcap_{\substack{\mathcal{A} \text{ tribu} \\ \mathcal{A} \supseteq X}} \mathcal{A}$$

Définition 1.3 : Tribu borélienne.

La tribu engendrée par les ouverts de \mathbb{R} est appelée tribu des boréliens ou tribu borélienne. Elle est notée $\mathcal{B}(\mathbb{R})$ et ses éléments sont les **boréliens**.

Remarques 1.2 :

- On peut montrer à titre d'exercice que cette tribu est engendrée par $\{]-\infty ; a] \}_{a \in \mathbb{R}}$ et même par $\{]-\infty ; q] \}_{q \in \mathbb{Q}}$.
- $\mathcal{B}(\mathbb{R})$ contient tous les ouverts, tous les fermés, ainsi que les unions et intersections dénombrables d'ouverts et de fermés.

1.2 Mesure de Lebesgue

Définition 1.4 : Soit (Ω, \mathcal{A}) un espace mesurable. On appelle **mesure** sur la tribu \mathcal{A} une application $\mu : \mathcal{A} \rightarrow \mathbb{R}_+ \cup \{+\infty\}$ vérifiant :

- (i) $\mu(\emptyset) = 0$
- (ii) $\forall (A_i)_{i \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ famille dénombrable d'événements **2 à 2 disjoints** :

$$\mu \left(\bigcup_{i \in \mathbb{N}} A_i \right) = \sum_{i \in \mathbb{N}} \mu(A_i) \quad (\sigma\text{-additivité})$$

La donnée d'un triplet $(\Omega, \mathcal{A}, \mu)$ s'appelle **espace mesuré**.

Théorème 1.1 : Théorème d'extension de Carathéodory.

L'ensemble $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ est mesurable, et il existe une unique mesure μ sur celui-ci telle que :

$$\forall (a, b) \in \mathbb{R}^2, a < b, \quad \boxed{\mu([a; b]) = b - a}$$

Cette mesure est appelée **mesure de Lebesgue** : elle représente donc la longueur au sens usuel du terme.

Définition 1.5 : Un ensemble $X \subseteq \mathbb{R}$ est dit **négligeable** s'il est de mesure nulle (ou, ce qui est équivalent, inclus dans un ensemble de mesure nulle).

Remarque 1.3 : La σ -additivité d'une mesure (point (ii) de la définition) permet d'affirmer que $\mu(\mathbb{N}) = \mu(\mathbb{Z}) = \mu(\mathbb{Q}) = 0$: \mathbb{Q} est donc négligeable pour la mesure de Lebesgue.

Définition 1.6 : Une propriété est dite vraie **presque partout** si elle est vérifiée partout sauf sur un ensemble négligeable.

Exemple 1.2 : **Fonction de Heaviside**

$$\Theta : \mathbb{R} \rightarrow \mathbb{R}, \quad \Theta(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{2} & \text{si } x = 0 \\ 1 & \text{si } x > 0 \end{cases} \quad \text{est continue presque partout.}$$

1.3 Intégrale de Lebesgue

Définition 1.7 : Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est dite **mesurable** si :

$$\forall A \in \mathcal{B}(\mathbb{R}), \quad f^{-1}(A) \in \mathcal{B}(\mathbb{R})$$

Autrement dit si l'image réciproque de tout ensemble mesurable est un ensemble mesurable.

Définition 1.8 : $f : \mathbb{R} \rightarrow \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ est dite **étagée** s'il existe un nombre fini d'ensembles mesurables A_1, \dots, A_n et $(\alpha_1, \dots, \alpha_n) \in \overline{\mathbb{R}}^n$ tels que :

$$\forall x \in \mathbb{R}, \quad f(x) = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}(x) \quad \text{où} \quad \mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{sinon} \end{cases}$$

avec la convention de notation $\infty \times 0 = 0$ ici.

Remarque 1.4 : Les fonctions en escalier sont étagées.

Définition 1.9 : **Intégrale de Lebesgue.**

(i) Si f est une fonction étagée à valeurs positives, on pose par définition :

$$\int f d\mu = \sum_{i=1}^n \alpha_i \mu(A_i) \in [0; +\infty]$$

(ii) Si f est quelconque à valeurs positives, on pose par définition :

$$\int f d\mu = \sup \left\{ \int g d\mu, \quad g \text{ étagée et } 0 \leq g \leq f \right\} \in [0; +\infty]$$

(iii) Si f est quelconque et intégrable (c.f. ci-dessous), on pose par définition :

$$\int f d\mu = \int f^+ d\mu - \int f^- d\mu \quad \text{où} \quad f^+(x) = \max(0, f(x)) \quad \text{et} \quad f^-(x) = -\min(0, f(x))$$

On dit en outre que f est **intégrable** (ou sommable) lorsque $\int |f| d\mu$ est un nombre fini.

Remarque 1.5 : Toutes les propriétés usuelles de l'intégrale de Riemann sont vérifiées.

1.4 Calcul intégral

On utilise dans toute la suite la mesure de Lebesgue associée à la tribu des boréliens :

$$\forall a < b \in \mathbb{R}, \mu([a; b]) = \mu([a; b[) = \mu(]a; b]) = \mu(]a; b[) = b - a.$$

Théorème 1.2 : Théorème de convergence dominée.

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de \mathbb{R} dans \mathbb{R} convergeant simplement presque partout vers $f : \mathbb{R} \rightarrow \mathbb{R}$ mesurable.

On suppose qu'il existe $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ **intégrable** telle $\forall n \in \mathbb{N}, |f_n| \leq \phi$ presque partout. Alors :

- (i) f est intégrable (ce qui découle immédiatement du fait que les f_n soient majorées par une fonction intégrable)

(ii) Quelque soit A mesurable,
$$\lim_{n \rightarrow +\infty} \int_A f_n d\mu = \int_A f d\mu = \int_A \lim_{n \rightarrow +\infty} f_n d\mu$$

Théorème 1.3 : Théorème d'intégration terme à terme.

Soit $(f_n)_{n \in \mathbb{N}}$ une suite de fonctions mesurables de \mathbb{R} dans \mathbb{R} telle que $\sum_{n \in \mathbb{N}} \int |f_n| d\mu < +\infty$.

Alors :

- (i) $\sum_{n \in \mathbb{N}} f_n$ converge absolument presque partout, et sa somme est intégrable

(ii) Quelque soit A mesurable,
$$\int_A \sum_{n \in \mathbb{N}} f_n d\mu = \sum_{n \in \mathbb{N}} \int_A f_n d\mu$$

Théorème 1.4 : Analogie avec des indices continus.

On considère : $f : [a; b] \times \mathbb{R} \rightarrow \mathbb{R} \text{ ou } \mathbb{C}$ et $g(x) = \int f(x, t) dt$.
 $(x, t) \mapsto f(x, t)$

On suppose que $x \mapsto f(x, t)$ est continue en $x_0 \in [a; b]$ pour presque tout t , et qu'il existe $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ **intégrable** telle que $\forall x \in \mathcal{V}(x_0)$ un voisinage de x_0 , $|f(x, t)| \leq \phi(t)$ pour presque tout t .

Alors $x \mapsto g(x) = \int f(x, t) dt$ est continue en x_0 .

Théorème 1.5 : Théorème de Leibniz ou dérivation sous le signe intégral.

Soit $x_0 \in [a; b] \subseteq \mathbb{R}$ et $x \mapsto f(x, t)$ dérivable dans un voisinage de x_0 pour presque tout t . On suppose qu'il existe $\phi : \mathbb{R} \rightarrow \mathbb{R}_+$ **intégrable** telle que $\forall x \in [a; b], \left| \frac{\partial f}{\partial x}(x, t) \right| \leq \phi(t)$ presque partout. Alors :

- (i) $x \mapsto \int f(x, t) dt$ est dérivable

(ii)
$$\frac{d}{dx} \int f(x, t) dt = \int \frac{\partial f}{\partial x}(x, t) dt$$

Remarque 1.6 : Ce dernier théorème se généralise pour x appartenant à n'importe quel intervalle I de \mathbb{R} , et où la condition de domination est remplacée par « $\forall x \in K$ compact de I ».

1.5 Fonctions de plusieurs variables

On construit de la même façon une mesure sur \mathbb{R}^2 , avec :

$$\int f(x, y) d\mu(x, y) = \int \left[\int f(x, y) d\mu(y) \right] d\mu(x) = \int \left[\int f(x, y) d\mu(x) \right] d\mu(y)$$

et à condition que ces intégrales existent.

Remarque 1.7 :

- On devrait commencer par s'intéresser à l'unicité d'une mesure sur \mathbb{R}^n et montrer que la tribu engendrée par $\{A \times B \in \mathcal{B}(\mathbb{R})^2\}$ forme bien les boréliens de \mathbb{R}^2 .
- Cette mesure coïncide avec la mesure du volume d'un pavé droit sur \mathbb{R}^n .

Théorème 1.6 : Théorème de Fubini.

Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ mesurable telle que $\int \left[\int |f(x, y)| dy \right] dx < +\infty$. Alors f est intégrable sur \mathbb{R}^2 , et :

- (i) $\int \left[\int |f(x, y)| dx \right] dy < +\infty$
- (ii) $x \mapsto \int f(x, y) dy$ est intégrable pour presque tout x
- (iii) $\int f(x, y) dx dy = \int \int f(x, y) dy dx$

Pour manipuler les intégrales de \mathbb{R}^n , il est utile de savoir changer de variable.

Définition 1.10 : Soit $\varphi : U \rightarrow V$ de classe \mathcal{C}^1 où U et V sont des ouverts de \mathbb{R}^n . On appelle **matrice jacobienne** de φ la matrice d'éléments $\boxed{\partial_j \varphi_i}$. On la note J_φ . (Elle dépend du point x_0 de \mathbb{R}^n où on calcule les dérivées de φ , c'est donc en réalité $J_\varphi(x_0)$.)

Théorème 1.7 : Théorème de changement de variable.

Soit $\varphi : U \rightarrow V$ où U et V sont des ouverts de \mathbb{R}^n de classe \mathcal{C}^1 , bijective, et telle que $\det J_\varphi = \det (\partial_j \varphi_i)_{i,j} \neq 0$. Alors $\forall f : V \rightarrow \mathbb{R}^p$:

$$f \text{ intégrable sur } V \iff f \circ \varphi \text{ intégrable sur } U$$

et :

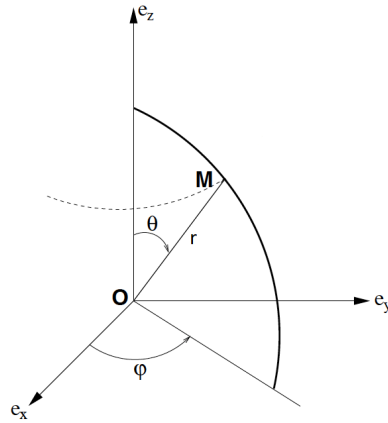
$$\int_V f(v_1, \dots, v_n) dv_1 \dots dv_n = \int_U f \circ \varphi(u_1, \dots, u_n) |\det J_\varphi| du_1 \dots du_n$$

Mnémotechniquement : $\boxed{dv_1 \dots dv_n = du_1 \dots du_n \left| \det \left(\frac{\partial v_i}{\partial u_j} \right)_{i,j} \right|}$

Remarques 1.8 :

- Interprétation géométrique : dans le plan, l'aire du parallélogramme porté par deux vecteurs \vec{u}, \vec{v} vaut $|\det(\vec{u}|\vec{v})|$. Dans l'espace, le volume du prisme droit porté par trois vecteurs $\vec{u}, \vec{v}, \vec{w}$ vaut $|\det(\vec{u}|\vec{v}|\vec{w})|$.
- Si φ était une application linéaire (invertible), alors J serait la matrice des n vecteurs de base.

Exemple 1.3 : Coordonnées sphériques.



Soit $\varphi : \mathbb{R}^3 \rightarrow [0; +\infty[\times [0; \pi] \times [0; 2\pi[$ telle que
$$\begin{cases} \varphi_1(x, y, z) = \sqrt{x^2 + y^2 + z^2} = r \\ \varphi_2(x, y, z) = \theta \\ \varphi_3(x, y, z) = \phi \end{cases}$$

$$\text{On a : } \begin{cases} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta \end{cases}$$

Si l'on veut effectuer le changement de variable $dx dy dz = [?] dr d\theta d\phi$, alors :

$$[?] = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} & \frac{\partial x}{\partial \phi} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} & \frac{\partial y}{\partial \phi} \\ \frac{\partial z}{\partial r} & \frac{\partial z}{\partial \theta} & \frac{\partial z}{\partial \phi} \end{vmatrix} = \dots = r^2 \sin \theta \implies dx dy dz = r^2 \sin \theta dr d\theta d\phi$$

1.6 Probabilités

Définition 1.11 : Soit (Ω, \mathcal{T}) un espace probablisable. On appelle **mesure de probabilité** ou **probabilité** $\mathbb{P} : \mathcal{T} \rightarrow \mathbb{R}_+$ toute mesure \mathbb{P} sur \mathcal{T} telle que $\boxed{\mathbb{P}(\Omega) = 1}$.

La donnée du triplet $(\Omega, \mathcal{T}, \mathbb{P})$ s'appelle **espace probablisé**.

Remarque 1.9 : L'exemple qui nous intéresse le plus est celui de $p : \mathbb{R}^n \rightarrow \mathbb{R}$, intégrable et telle que $\int p d\mu = 1$. Ceci définit bien une mesure de probabilité sur \mathbb{R}^n .

Définition 1.12 : On appelle **loi normale** $\mathcal{N}(m, \sigma^2)$, **centrée** en $m \in \mathbb{R}$, de **variance** $\sigma^2 \neq 0$, la loi (c.f. définition 1.16) correspondant à la densité de probabilité (c.f. remarque 1.15) :

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

En physique, on parlera de **loi gaussienne de moyenne m et de variance σ^2** .

Définition 1.13 : Soient A et B deux événements tels que $\mathbb{P}(A) \neq 0$. On appelle **probabilité de B sachant A** :

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$$

Définition 1.14 : Deux événements A et B sont dits **indépendants** si $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Si $\mathbb{P}(A) \neq 0$, cela est équivalent à $\mathbb{P}(B|A) = \mathbb{P}(B)$.

1.7 Variables aléatoires

1.7.1 Point de vue mathématique

Définition 1.15 : On appelle **variable aléatoire** toute fonction $X : (\Omega, \mathcal{T}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ définie sur Ω telle que :

$$\forall A \in \mathcal{B}(\mathbb{R}), X^{-1}(A) \in \mathcal{T}$$

Autrement dit l'image réciproque de tout événement est un événement.

Remarque 1.10 : Une fonction mesurable est donc une variable aléatoire de $\mathcal{B}(\mathbb{R})$ dans $\mathcal{B}(\mathbb{R})$.

Définition 1.16 : Si $X : \Omega \rightarrow \mathbb{R}$ est une variable aléatoire, on appelle **loi de probabilité** P_X la probabilité image de \mathbb{P} par X : $P_X = \mathbb{P} \circ X^{-1}$.

1.7.2 Moments *et al.*

Définition 1.17 : Soient X une variable aléatoire, P_X sa loi et $k \in \mathbb{N}$. On définit le **moment d'ordre k de X** par :

$$m_k = \mathbb{E}(X^k) = \langle X^k \rangle = \int x^k dP_X(x)$$

Remarque 1.11 : Il existe des lois pour lesquelles le moment d'ordre k n'existe pas, par exemple : $dP_X(x) = p(x)dx$ avec $p(x) = \frac{1}{\pi} \frac{1}{1+x^2}$ (**loi de Cauchy ou lorentzienne**).

Définition 1.18 : Soit X une variable aléatoire. On appelle **fonction génératrice des moments** de X :

$$\begin{aligned} G : \mathbb{C} &\rightarrow \mathbb{C} \\ t &\mapsto \mathbb{E}(e^{tX}) = \langle e^{tX} \rangle = \int e^{tx} dP_X(x) \end{aligned}$$

lorsque celle-ci est définie.

Si X est une variable aléatoire vectorielle dans \mathbb{R}^n et $t \in \mathbb{R}^n$, la fonction génératrice est définie par :

$$G(t) = \mathbb{E}(e^{\langle t|X \rangle}) = \int e^{\langle t|x \rangle} dP_X(x) \quad \text{où } \langle | \rangle \text{ est le produit scalaire}$$

Remarques 1.12 :

- G s'appelle fonction génératrice car $G^{(k)}(0) = m_k$. Calculer G donne donc accès à tous les moments de la variable aléatoire associée.
- On définit parfois $G(t) = \mathbb{E}(e^{itX})$ qu'on appelle **fonction caractéristique** de X . Si X est une variable aléatoire discrète à valeurs dans \mathbb{N} , G est définie par :

$$G(z) = \sum_{n \in \mathbb{N}} z^n \mathbb{P}(X = n) \quad \text{et alors} \quad \left(z \frac{d}{dz} \right)^k G \Big|_{z=1} = \langle X^k \rangle$$

Définition 1.19 : La quantité qui va se révéler importante physiquement est la **fonction génératrice des cumulants** définie par $W(t) = \ln G(t) = \ln \mathbb{E}(e^{tX})$.

Définition 1.20 : On définit le **cumulant d'ordre k** K_k ou moment connexe d'ordre k par :

$$K_k = W^{(k)}(0)$$

Propriété 1.1 : Les fonctions G et W étant développables en série entière, on a donc :

$$G(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} m_k$$

$$W(z) = \sum_{k=0}^{\infty} \frac{z^k}{k!} K_k$$

Définition 1.21 : En outre, $\mu_k = \langle (X - m)^k \rangle$ est appelé **moment centré d'ordre k** .

Remarque 1.13 : Liste des premiers cumulants.

- $K_1 = m_1 = \langle X \rangle$
- $K_2 = \mu_2^2 = m_2 - m_1^2 = \langle (X - \langle X \rangle)^2 \rangle = \text{Var}(X)$
- $K_3 = \mu_3 = m_3 - 3m_1m_2 + 2m_1^3$
- $K_4 = m_4 - 4m_1m_3 - 3m_2^2 + 12m_1^2m_2 - 6m_1^4 \neq \mu_4$

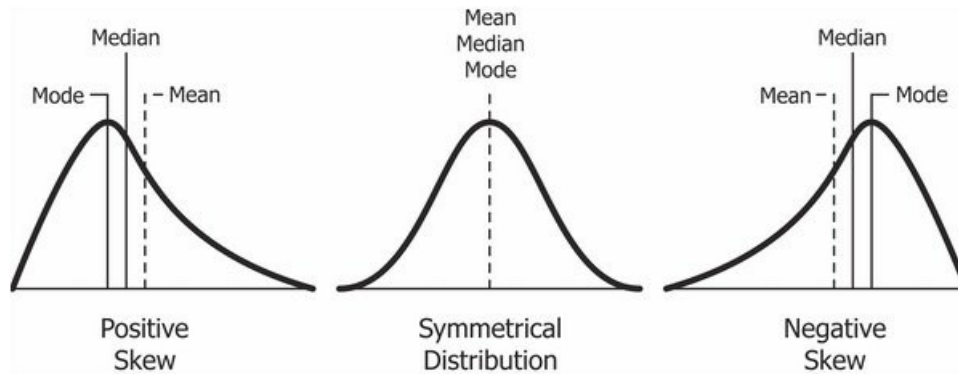
Remarque 1.14 : Si $z \in \mathbb{R}$, $W''(z) = \frac{G''(z) - (G'(z))^2}{G(z)^2} \geq 0$ (par Cauchy-Schwarz ou parce que $\text{Var}(X) \geq 0$). W est donc convexe.

Définition 1.22 : On appelle **skewness** ou **degré d'asymétrie** d'une variable aléatoire X la quantité :

$$\frac{K_3}{\sigma^3}$$

Comme son nom l'indique, il quantifie le degré d'asymétrie d'une loi de probabilité.

Exemple 1.4 :



Définition 1.23 : Il est fréquent en physique de travailler avec des signaux pairs, on définit alors le **paramètre de non-gaussianité ou NGP ou paramètre de Binder** par :

$$\text{NGP} = \frac{K_4}{3K_2^2}$$

Si la loi est paire, $\text{NGP} = \frac{\langle X^4 \rangle}{3\sigma^4} - 1$ (et est nul pour une loi gaussienne).

Théorème 1.8 : **Théorème de Marcinkiewicz.**

W ne peut pas être un polynôme de degré supérieur ou égal à 3.

\implies Tronquer W à un ordre fini peut être dangereux.

Théorème 1.9 : Théorème de transfert.

Si X est une variable aléatoire et $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction, l'espérance de la variable aléatoire $f(X)$ vaut :

$$\mathbb{E}(f(X)) = \langle f(X) \rangle = \int_{\Omega} f(X) dp = \int f(x) dP_X(x)$$

Remarque 1.15 : Deux situations vont donc se présenter :

- (i) X est une variable aléatoire discrète : la loi P_X est donnée par les probabilités $p_i = \mathbb{P}(X = x_i)$, et $\int dP_X(x)(\dots) = \sum_i p_i(\dots)$.

On écrira $P_X = \sum_i p_i \delta_{X, x_i}$ où δ est le symbole de Kronecker

- (ii) X est une variable aléatoire continue : alors $P_X(x) = p(x)dx$, et $p(x)$ s'appelle une **densité de probabilité**.

1.7.3 Variables aléatoires continues

Définition 1.24 : Soit X une variable aléatoire, on appelle **fonction de répartition** la fonction $F(x) = \mathbb{P}(X \leq x)$. C'est donc la probabilité image par X de $] -\infty ; x]$.

On a donc :

- $F(x) = \int_{-\infty}^x dP_X(x')$
- Puisque $\mathbb{P}(\mathbb{R}) = 1$, $\mathbb{P}(X > x) = 1 - F(x)$
- $\mathbb{P}(a < X \leq b) = F(b) - F(a)$

On peut également, si X et Y sont deux variables aléatoires sur le même espace, définir une fonction de répartition pour plusieurs variables par : $F(x, y) = \mathbb{P}([X \leq x] \cap [Y \leq y])$.

Définition 1.25 : Si $dP_{X,Y}(x, y) = p(x, y)dx dy$ où $p(x, y)$ densité de probabilité, on appelle **loi marginale de X** : $p(x) = \int p(x, y)dy$.

Propriété 1.2 : Si X est une variable aléatoire et $Y = f(X)$ une variable aléatoire construite à partir de X , la densité de probabilité de Y est donnée par :

$$p(y) = \frac{p(x)}{f'(x)}$$

Ceci est cohérent avec le fait que :

$$dP = dP_X(x) = dP_Y(y) \iff p_X(x)dx = p_Y(y)dy \iff p_Y(y) = \frac{dx}{dy}p_X(x) = \frac{p_X(x)}{f'(x)}$$

Remarque 1.16 : Attention ! Cela est plus compliqué avec n variables : si $(y_1, \dots, y_n) = \varphi(x_1, \dots, x_n)$ avec $\varphi : \mathbb{R}^n \rightarrow \mathbb{R}^n \mathcal{C}^1$, alors :

$$p(x_1, \dots, x_n) dx_1 \dots dx_n = p(y_1, \dots, y_n) dy_1 \dots dy_n \left| \det \left(\frac{\partial y_i}{\partial x_j} \right)_{i,j} \right|$$

1.8 Une bibliothèque d'exemples

► c.f. Garin Crooks.

1.8.1 Loi normale

Définition 1.26 : La **loi normale** $\mathcal{N}(m, \sigma^2)$ est une loi de densité :

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(x-m)^2}{2\sigma^2} \right)$$

En faisant le calcul, on trouve que cette loi a pour espérance m et pour variance σ^2 . On constate également que :

$$\mu_{2n} = \int (x-m)^{2n} p(x) dx = (2n-1)!! \sigma^{2n}$$

où $(2n-1)!! = 1 \times 3 \times 5 \times \dots \times (2n-1)$.

La fonction génératrice de la loi normale vaut :

$$G(z) = \int e^{zx} p(x) dx = e^{mz - \frac{\sigma^2 z^2}{2}}$$

d'où l'on déduit la fonction génératrice des cumulants :

$$W(z) = \ln G(z) = mz - \frac{\sigma^2 z^2}{2}$$

Remarque 1.17 : On constate que W est un polynôme de degré 2 : tous les cumulants d'ordre $k \geq 3$ sont nuls.

On peut en fait montrer que ceci caractérise la loi gaussienne : si W a la forme ci-dessus, alors la densité de probabilité est celle d'une gaussienne.

Propriété 1.3 : La fonction de répartition de la loi normale vaut :

$$F(x) = \int_{-\infty}^x p(x') dx' = \frac{1}{2} \left(1 - \operatorname{erf} \left(\frac{m-x}{\sqrt{2}\sigma} \right) \right)$$

où $\operatorname{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-u^2} du$ est la fonction **erreur**.

De plus, la probabilité $\mathbb{P}(m - x \leq X \leq m + x)$ est donnée par $\operatorname{erf}\left(\frac{x}{2\sigma}\right)$.

Remarque 1.18 : Pour déterminer à quel point une densité est similaire à une gaussienne, on regarde le skewness $\frac{K_3}{\mu_2^{3/2}} = \frac{K_3}{\sigma^3}$, et si la loi est symétrique alors on regarde $\frac{K_4}{3\mu_2^2} = \frac{\langle X^4 \rangle}{3\langle X^2 \rangle} - 1$. Si ce paramètre de non-gaussianité est nul, alors on a affaire à une gaussienne.

Exemple 1.5 : La première rencontre de la gaussienne en physique est en théorie cinétique des gaz :

$$\mathbb{P}(\vec{v} \leq \vec{V} \leq \vec{v} + d\vec{v}) = p(\vec{v}) dv_x dv_y dv_z$$

avec :
$$p(\vec{v}) = \frac{1}{\left(\frac{2\pi k_B T}{m}\right)^{3/2}} \exp\left(\frac{-m \vec{v}^2}{2k_B T}\right)$$

Ainsi, la gaussienne peut se généraliser à plusieurs dimensions. La densité de probabilité s'écrit alors :

$$p(x_1, \dots, x_n) = C \exp\left(-\frac{1}{2} {}^t x \Gamma x\right) = C \exp\left(-\frac{1}{2} \sum_{i,j} x_i \Gamma_{i,j} x_j\right) \quad (\text{notée } \exp\left(-\frac{1}{2} x_i \Gamma_{i,j} x_j\right))$$

où Γ est une **matrice symétrique réelle à spectre strictement positif** et $C = \frac{\sqrt{\det \Gamma}}{(2\pi)^{N/2}}$.

1.8.2 Loi binomiale

La loi binomiale est une loi discrète de paramètres $(p \in [0; 1], n \in \mathbb{N})$ telle que $p_k = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

Son espérance est np et sa variance $np(1-p)$.

Propriété 1.4 :

$$\lim_{\substack{n \rightarrow +\infty \\ pn \text{ fixe}}} p_k = e^{-\lambda} \frac{\lambda^k}{k!}$$

\implies La loi binomiale converge vers une loi de Poisson.

Remarque 1.19 : On peut voir que la distribution de p_k en fonction de k pour de grandes valeurs de n présente un profil de largeur caractéristique $\frac{1}{\sqrt{n}}$. Il faut donc regarder p_k à la bonne échelle pour étudier son comportement. Pour cela, on centre et on réduit la variable aléatoire en posant $\xi = \frac{X_n - pn}{\sqrt{p(1-p)n}}$. Si $n \rightarrow +\infty$ avec ξ restant d'ordre 1 « fixé », on montre que $\frac{\ln p_k}{n} = -\frac{\xi^2}{2} + o(1)$ et donc un profil de probabilité en $e^{-\xi^2/2}$. C'est l'objet de la partie suivante.

1.9 Théorème central limite

1.9.1 Plusieurs types de convergence

Définition 1.27 : Soit $(X_n)_n$ une suite de variables aléatoires sur un même univers Ω . On dit que :

- (i) X_n **converge presque sûrement** vers une variable aléatoire X si :

$$\mathbb{P}(X_n \xrightarrow[n \rightarrow +\infty]{} X) = 1$$

i.e. il existe un sous-ensemble $\Omega_0 \subseteq \Omega$ de mesure nulle tel que :

$$\forall \omega \in \Omega \setminus \Omega_0, \forall \varepsilon > 0, \exists N \in \mathbb{N}, \forall n \geq N, |X_n(\omega) - X| < \varepsilon$$

- (ii) X_n **converge en probabilité** vers une variable aléatoire X si :

$$\forall \varepsilon > 0, \mathbb{P}(|X_n - X| \geq \varepsilon) \xrightarrow[n \rightarrow +\infty]{} 0$$

- (iii) X_n **converge en loi** vers une variable aléatoire X si :

$$\forall x \in \mathbb{R}, F_{X_n}(x) = \mathbb{P}(X_n \leq x) \xrightarrow[n \rightarrow +\infty]{} F_X(x) = \mathbb{P}(X \leq x)$$

On a : $(i) \implies (ii) \implies (iii)$.

1.9.2 Loi des grands nombres

Théorème 1.10 : Loi faible des grands nombres.

Soient $(X_n)_n$ des **v.a.i.i.d.** (variables aléatoires indépendantes identiquement distribuées i.e. de même loi) d'espérance finie m et de variance finie. Alors :

$$\frac{S_n}{n} = \frac{X_1 + \dots + X_n}{n} \quad \text{converge en probabilité vers } m \text{ lorsque } n \rightarrow +\infty.$$

Démonstration : connue.

Remarque 1.20 : On a envie d'étendre cette propriété de S_n aux grandeurs physiques intensives : si $T_N = \frac{2}{3N} \sum_{i=1}^N \frac{1}{2} m \vec{v}_i^2$, cela permet de définir une température moyenne.

1.9.3 Théorème central limite

Théorème 1.11 : Théorème Central Limite.

Soient $(X_n)_n$ des v.a.i.i.d. d'espérance finie m et de variance finie σ^2 . Alors :

$$s_N = \frac{\sum_{i=1}^N X_i - Nm}{\sqrt{N}\sigma} \quad \text{converge en loi vers une loi normale centrée réduite } \mathcal{N}(0, 1).$$

Démonstration 1.1 :

On introduit :

$$\begin{aligned} G(z) &= \mathbb{E} \left(\exp \left[z \frac{\sum_{i=1}^N X_i - Nm}{\sqrt{N}\sigma} \right] \right) \\ &= \mathbb{E} \left(\exp \left[\frac{z}{\sigma\sqrt{N}} \left(\sum_{i=1}^N X_i - m \right) \right] \right) \\ &= \prod_{i=1}^N \mathbb{E} \left(\exp \left(\frac{z}{\sigma\sqrt{N}} (X_i - m) \right) \right) \quad (\text{indépendance des } X_i) \\ &= \mathbb{E} \left(\exp \left(\frac{z}{\sigma\sqrt{N}} (X_1 - m) \right) \right)^N \quad (\text{identique distribution}) \\ &= \mathbb{E} \left(1 + \frac{z}{\sigma\sqrt{N}} (X_1 - m) + \frac{z^2}{2\sigma^2 N} (X_1 - m)^2 + o \left(\frac{1}{n} \right) \right)^N \\ &= \left(1 + \frac{z^2}{2N} + o \left(\frac{1}{n} \right) \right)^N \\ &\xrightarrow{n \rightarrow +\infty} e^{\frac{z^2}{2}} = \int \frac{1}{\sqrt{2\pi}} e^{tz} e^{-\frac{t^2}{2}} dt \end{aligned}$$

Ainsi S_N converge bien en loi vers une loi normale $\mathcal{N}(0, 1)$.

\Rightarrow Il faut retenir que la somme d'un grand nombre de v.a.i.i.d. suit une loi gaussienne.

1.9.4 Grandes déviations

Définition 1.28 : Si $(X_n)_n$ sont des variables aléatoires et $s_n = \frac{1}{n} \sum_{i=1}^n X_i$, on note :

$$I(s) = \lim_{n \rightarrow +\infty} \frac{1}{n} \ln \mathbb{P}(S_n = s)$$

lorsque cette limite existe. I s'appelle la **fonction de grande déviation** de s_n .

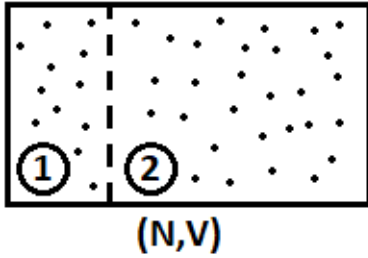
Dans les conditions du théorème central limite, on sait que $I(s) = -\frac{(s - m)^2}{2\sigma^2}$.

Remarque 1.21 : $I(s)$ est également appelée fonction de taux.

$I(s)$ renseigne sur l'occurrence d'une valeur s de s_n éloignée de la moyenne m :

$$\mathbb{P}(s_n = s) \sim e^{nI(s)}$$

Exemple 1.6 : On reprend le dispositif suivant :



On a N particules dans un volume V , on note p la probabilité d'être dans le volume (1) ($p = \frac{V_1}{V}$) et $1 - p$ celle d'être dans le volume (2).

On pose x_i qui vaut 1 si la particule n° i est dans (1) et 0 sinon. Soit $n = \sum_{i=1}^N X_i$ = nombre de particules dans (1) et $\nu_n = \frac{n}{N}$ = moyenne instantanée du nombre de particules dans (1). Alors :

$$\mathbb{P}(\nu_n = \nu) = \binom{N}{n} p^n (1-p)^{N-n} = e^{NI(\nu)}$$

où $n = \nu N$ et $I(\nu) = -\nu \ln \nu - (1 - \nu) \ln(1 - \nu) + \nu \ln p + (1 - \nu) \ln(1 - p)$.

Remarque 1.22 : Lorsque I présente une singularité, on dit que le système subit une **transition de phase**. En physique, I est une sortie d'« énergie libre ».

1.10 Information

Définition 1.29 : Si X est une variable aléatoire discrète pouvant prendre les valeurs x_i avec probabilité p_i , on définit l'**entropie de Shannon** associée à X par :

$$H_b(X) = - \sum_i p_i \log_b p_i \quad \text{où } \log_b(x) = \frac{\ln(x)}{\ln(b)}$$

Cette entropie est maximale si les p_i sont toutes égales, et minimale si l'une des p_i vaut 1. Elle est une mesure du défaut d'information que l'on détient sur une variable aléatoire.

Propriété 1.5 : Si X et Y sont deux variables aléatoires indépendantes, alors $H_b(X + Y) = H_b(X) + H_b(Y)$. L'entropie de Shannon est extensive.

Définition 1.30 : Si l'on veut comparer deux lois entre elles, on introduit l'**entropie / la distance / la divergence de Kullback-Leibler** :

$$\mathcal{D}_{KL}(\mathbb{P} \parallel \mathbb{Q}) = - \sum_x \mathbb{P}(x) \ln \frac{\mathbb{Q}(x)}{\mathbb{P}(x)}$$

où \mathbb{P} et \mathbb{Q} sont deux lois sur le même espace.

Remarque 1.23 :

- $\Delta \mathcal{D}_{KL}$ est toujours positive, mais elle n'est pas symétrique : le terme « distance » est mal adapté.
- On peut également définir \mathcal{D}_{KL} pour des densités.
- \mathcal{D}_{KL} représente l'**information perdue lorsqu'on approche \mathbb{P} par \mathbb{Q}** .

Chapitre 2

Fonctions de plusieurs variables

2.1 Différentielle

2.1.1 Dérivées partielles du premier ordre

⚠ Les définitions et propriétés présentes ici sont parfois incomplètes, imprécises ou à hypothèses trop fortes. Un cours de MP* pourra être utile si l'on recherche plus de généralité.

Définition 2.1 : Soient E, F des \mathbb{R} -ev, U un ouvert de E et $f : U \rightarrow F$. Soit $v \in E \setminus \{0\}$. On dit que f est **dérivable en $a \in E$ selon v** si $t \in \mathbb{R} \mapsto f(a + tv)$ est dérivable en 0, et on note par définition :

$$\frac{\partial f}{\partial v}(a) = \left. \frac{d}{dt} f(a + tv) \right|_{t=0}$$

Si on choisit pour v un vecteur d'une base de E , $\frac{\partial f}{\partial e_i}$ s'appelle **dérivée partielle de f par rapport à x_i** et est notée $\frac{\partial f}{\partial x_i}$ ou encore $\partial_i f(a)$.

Définition 2.2 : Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ (ou plus généralement $f : U \rightarrow \mathbb{R}^m$ où U ouvert de \mathbb{R}^n). On dit que f est **différentiable** en $a \in \mathbb{R}^n$ si :

$$\exists L \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m), \quad f(a + h) = f(a) + L(h) + o(\|h\|) \text{ au voisinage de } a$$

L est **unique** : on l'appelle **différentielle de f en a** , notée $df(a)$. (On a donc $df(a) : h \mapsto df(a)(h)$ qui est application linéaire.)

Théorème 2.1 : Lien entre différentielle et dérivées partielles.

La différentielle de f en a est égale à l'application linéaire :

$$df(a) : h = {}^t(h_1, \dots, h_n) \in \mathbb{R}^n \mapsto df(a)(h) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} h_i$$

Définition 2.3 : La matrice dans la base canonique de $df(x)$ est appelée **Jacobienne de f en x** et vaut :

$$J_f(x) = \left(\frac{\partial f_i}{\partial x_j} \right)_{i \in \llbracket 1; m \rrbracket, j \in \llbracket 1; n \rrbracket}$$

où f_i est la i -ème coordonnée de f .

Définition 2.4 : On note $\nabla f = {}^t J_f$. Dans le cas particulier où f est à valeurs dans \mathbb{R} (i.e. $m = 1$), $\nabla f = {}^t(\partial_1 f, \dots, \partial_n f)$ est appelé le **gradient** de f .

La Jacobienne est utile dans la composition d'applications différentielles, comme le montre le théorème suivant.

Théorème 2.2 : Règle de la chaîne.

Si $f : \mathbb{R}^m \rightarrow \mathbb{R}^n$ et $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, alors :

$$\forall i \in \llbracket 1; m \rrbracket, \quad \frac{\partial (g \circ f)}{\partial x_i}(x) = \sum_{j=1}^n \frac{\partial g}{\partial y_j}(f(x)) \frac{\partial f_j}{\partial x_i}(x)$$

Ou encore :

$$J_{g \circ f}(x) = J_g(f(x)) J_f(x)$$

2.1.2 Ordres supérieurs

Théorème 2.3 : Théorème de Schwarz.

Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$ de classe \mathcal{C}^2 . Alors :

$$\forall (i, j) \in \llbracket 1; n \rrbracket^2, \quad \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

Remarque 2.1 : Plus généralement, si f est de classe \mathcal{C}^p , alors toutes les dérivées partielles d'ordre p commutent entre elles.

Théorème 2.4 :

Formule de Taylor-Young Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}$ de classe \mathcal{C}^{p+1} . Alors $\forall a = (a_1, \dots, a_n)$:

$$f(x) = f(a) + \sum_{k=1}^p \frac{1}{k!} \sum_{i_1 + \dots + i_n = k} \frac{k!}{i_1! \dots i_n!} (x_1 - a_1)^{i_1} \dots (x_n - a_n)^{i_n} \frac{\partial^k f}{\partial^{i_1} x_1 \dots \partial^{i_n} x_n} \Big|_a + O(\|x - a\|^{p+1})$$

Cette formule généralise le développement limité sur \mathbb{R}^n .

Exemple 2.1 :

$$f(x, y) = f(0, 0) + x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} + \frac{1}{2} x^2 \frac{\partial^2 f}{\partial x^2} + \frac{1}{2} y^2 \frac{\partial^2 f}{\partial y^2} + xy \frac{\partial^2 f}{\partial x \partial y} + O(\|(x, y)^3\|)$$

où l'on a utilisé le théorème de Schwarz.

Définition 2.5 : L'ordre particulièrement intéressant en physique est l'ordre 2. On appelle **matrice Hessienne de f en a** la matrice :

$$\text{Hess}(f)(a) = \nabla^2 f(a) = \left(\frac{\partial^2 f(a)}{\partial x_i \partial x_j} \right)_{i,j}$$

Le développement limité de f à l'ordre 2 s'écrit alors :

$$f(x) = f(a) + \langle \nabla f | x - a \rangle + {}^t(x - a) \nabla^2 f(a) (x - a) + O(\|(x, y)^3\|)$$

Remarque 2.2 :

- Si f est \mathcal{C}^2 , le théorème de Schwarz assure que la **Hessienne est une matrice symétrique**.
- La Hessienne de f en a est la matrice de la **forme quadratique** approchant f en a : elle va donner l'allure locale de f .

2.1.3 Calcul vectoriel

Définition 2.6 :

- Si $\vec{F} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ est un champ de vecteurs, on définit la **divergence de \vec{F}** par :

$$\text{div} \vec{F} = \vec{\nabla} \cdot \vec{F} = \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}$$

- Si $n = 3$, on définit également le **rotationnel de \vec{F}** par :

$$\vec{\text{rot}} \vec{F} = \vec{\nabla} \wedge \vec{F} = \begin{vmatrix} \frac{\partial F_2}{\partial x_3} - \frac{\partial F_3}{\partial x_2} \\ \frac{\partial F_3}{\partial x_1} - \frac{\partial F_1}{\partial x_3} \\ \frac{\partial F_1}{\partial x_2} - \frac{\partial F_2}{\partial x_1} \end{vmatrix}$$

Remarque 2.3 : Si $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ est un difféomorphisme, on peut utiliser la jacobienne pour réexprimer ces opérateurs vectoriels à l'aide des $y_i = \varphi_i(x)$ et de $\tilde{F}(y) = F(\varphi(x))$ (par exemple pour les exprimer en coordonnées cylindriques ou sphériques).

2.2 Convexité

2.2.1 Définitions

Définition 2.7 : On dit que $U \subseteq \mathbb{R}^n$ est **convexe** si :

$$\forall (x, y) \in U^2, \quad [x; y] \subseteq U, \text{ i.e. } \forall (x, y) \in U^2, \forall t \in [0; 1], \quad tx + (1 - t)y \in U$$

Définition 2.8 : Soit $U \subseteq \mathbb{R}^n$ convexe et $f : U \rightarrow \mathbb{R}$. On dit que f est :

— **convexe** si :

$$\boxed{\forall (x, y) \in U^2, \forall t \in [0; 1], \quad f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y)}$$

— **strictement convexe** si pour $x \neq y$ on a l'inégalité stricte dans l'inégalité précédente

— **fortement convexe** si :

$$\exists \alpha > 0, \forall (x, y) \in U^2, \forall t \in [0; 1], \quad f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) - \alpha t(1 - t)\|x - y\|^2$$

— **concave** si $-f$ est convexe

Théorème 2.5 :

Soient $\Omega \subseteq \mathbb{R}^n$ ouvert, $f : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 et $U \subseteq \Omega$ convexe. Les assertions suivantes sont équivalentes :

- (i) f est convexe sur U
- (ii) $\forall (x, y) \in U^2, \quad f(y) - f(x) \geq \langle \nabla f(x) \mid y - x \rangle$
- (iii) ∇f est monotone sur U , i.e. $\forall (x, y) \in U^2, \quad \langle \nabla f(y) - \nabla f(x) \mid y - x \rangle \geq 0$

Théorème 2.6 :

Soit $\Omega \subseteq \mathbb{R}^n$ ouvert, $f : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^2 et $U \subseteq \Omega$ convexe.

$$\boxed{f \text{ est convexe sur } U \text{ si et seulement si } \text{Hess}(f) \in \mathcal{S}_n^+(\mathbb{R})}$$

(la Hessienne de f est une matrice symétrique positive en tout point de U)

Remarque 2.4 : Localement, cela signifie que f a toujours l'allure d'une cuvette.

Définition 2.9 : Soit $f : \Omega \rightarrow \mathbb{R}$ où $\Omega \subseteq \mathbb{R}^n$ est non borné. On dit que f est **coercive** si :

$$\boxed{\lim_{\substack{\|x\| \rightarrow +\infty \\ x \in \Omega}} f(x) = +\infty}$$

Théorème 2.7 :

Soit $\Omega \subseteq \mathbb{R}^n$ ouvert, $U \subseteq \Omega$ convexe non borné et $f : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 **convexe**. Alors f est coercive sur U .

2.2.2 Minima

Définition 2.10 : Soient $U \subseteq \mathbb{R}^n$, $u^* \in U$, $f : U \rightarrow \mathbb{R}$. On dit que u^* est un :

— **minimum (resp. maximum) global ou absolu** de f sur U si :

$$\forall u \in U, \quad f(u) \geq f(u^*) \text{ (resp. } f(u) \leq f(u^*))$$

— **minimum (resp. maximum) local ou relatif** de f sur U si :

$$\exists \mathcal{V} \subseteq U \text{ voisinage de } u^*, \quad \forall u \in \mathcal{V}, \quad f(u) \geq f(u^*) \text{ (resp. } f(u) \leq f(u^*))$$

— **extremum (global ou local)** si c'est un minimum ou un maximum (global ou local)

Remarque 2.5 : Un extremum local sur U se situe nécessairement dans l'intérieur de U (noté $\overset{\circ}{U}$), car un de ses voisinages est inclus dans U .

Définition 2.11 : Soient $U \subseteq \mathbb{R}^n$ et $u^* \in U$. On dit que $\omega \in \mathbb{R}^n$ est une **direction admissible pour U en u^*** si $\exists t_0 > 0, \forall t \in [0; t_0], u^* + t\omega \in U$.

Théorème 2.8 :

Soient $\Omega \subseteq \mathbb{R}^n$ ouvert, $U \subseteq \Omega$, $f : \Omega \rightarrow \mathbb{R}$ \mathcal{C}^1 admettant un minimum u^* sur U . Alors :

- (i) Pour tout direction admissible $\omega \in \mathbb{R}^n$ en u^* , $\langle \nabla f(u^*) | \omega \rangle \geq 0$
- (ii) En particulier, si $u \in U$ est tel que $u - u^*$ est admissible, alors :

$$\langle \nabla f(u^*) | u - u^* \rangle \geq 0 \quad (\text{Inéquation d'Euler})$$

Propriété 2.1 : Si $U \subseteq \mathbb{R}^n$, $u^* \in \overset{\circ}{U}$ et $b \in \mathbb{R}^n$, alors :

$$(\forall u \in U, \langle b | u - u^* \rangle \geq 0) \iff b = 0$$

Définition 2.12 : Soient $\Omega \subseteq \mathbb{R}^n$ ouvert, $f : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 et $U \subseteq \Omega$ convexe. On dit que $u \in U$ est un **point critique de f** si $\nabla f(u) = 0$.

Théorème 2.9 : Extrema et points critiques.

Soient $\Omega \subseteq \mathbb{R}^n$ ouvert, $f : \Omega \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 et $U \subseteq \Omega$ convexe. Soit u^* un extremum local de f (appartenant donc à $\overset{\circ}{U}$). Alors **u^* est un point critique de f** :

$$\boxed{\nabla f(u^*) = 0} \quad (\text{Équation d'Euler})$$

Théorème 2.10 :

Sous les mêmes hypothèses que le théorème 2.9, on suppose de plus que f est convexe. Alors pour $u^* \in U$:

u^* minimum local de $f \iff u^*$ minimum global de $f \iff \forall u \in U, \langle \nabla f(u^*) | u - u^* \rangle \geq 0$

Théorème 2.11 :

Sous les mêmes hypothèses que le théorème 2.9, on suppose de plus que f est de classe \mathcal{C}^2 . Soit $u^* \in \overset{\circ}{U}$ tel que $\nabla f(u^*) = 0$. Alors :

- (i) **(Condition nécessaire)** Si u^* est un minimum local, alors la matrice Hessienne $\nabla^2 f(u^*)$ est symétrique positive.
- (ii) **(Condition suffisante)** Si $\nabla^2 f(u^*)$ est symétrique définie positive, alors u^* est un minimum local.

2.2.3 Existence et unicité

Théorème 2.12 : Théorème des bornes atteintes.

Soit $U \subseteq \mathbb{R}^n$ non vide. Soit $f : U \rightarrow \mathbb{R}$ continue.

- (i) Si U est compact, alors il existe un maximum et un minimum de f sur U .
- (ii) Si U est seulement fermé (et donc dans le cas de la dimension finie non borné), et que f est coercive sur U , alors il existe un minimum de f sur U .

Théorème 2.13 : Condition d'unicité du minimum.

Soit $U \subseteq \mathbb{R}^n$ convexe. Soit $f : U \rightarrow \mathbb{R}$ strictement convexe. Alors il existe au plus un minimum de f sur U .

Théorème 2.14 : Condition d'existence et unicité du minimum.

Soit $U \subseteq \mathbb{R}^n$ fermé et convexe. Soit $f : U \rightarrow \mathbb{R}$ \mathcal{C}^1 fortement convexe sur U . Alors il existe un et un seul minimum de f sur U .

2.3 Optimisation sans contraintes

► **Non traité.** Il s'agit de rechercher le minimum d'une fonction : c'est un problème algorithmique.

2.4 Optimisation avec contraintes

► L'objectif est le même qu'au 2.3 : on a $U \subseteq \mathbb{R}^n$, $J : \mathbb{R}^n \rightarrow \mathbb{R}$ et on cherche $\min_{u \in U} J(u)$ où U est un ensemble qui dépend des contraintes. Dans la suite, on choisira pour U un ensemble donné par :

$$U = \{x \in \mathbb{R}^n \mid \forall i \in \llbracket 1; m \rrbracket, \theta_i(x) \leq 0\}$$

où $\forall i \in \llbracket 1; m \rrbracket$, $\theta_i : \mathbb{R}^n \rightarrow \mathbb{R}$ est une contrainte.

Ceci inclut la contrainte égalité, car : $\theta_i = \bar{0} \iff \theta_i \leq \bar{0} \text{ et } \theta_i \geq \bar{0}$.

Remarque 2.6 : Si tous les θ_i sont convexes alors U l'est aussi.

2.4.1 Multiplicateurs de Lagrange

On se concentre ici sur le cas des contraintes égalités. On note :

$$\tilde{O} = \{x \in \mathbb{R}^n \mid \forall i \in \llbracket 1; m \rrbracket, \theta_i(x) = 0\}$$

où m est le nombre de contraintes avec $m \leq n$.

Définition 2.13 : On dit que $x \in \mathbb{R}^n$ est un **point régulier** si la famille des $\{\nabla \theta_i(x)\}_i$ est libre.

Théorème 2.15 : Multiplicateurs de Lagrange.

Soit $x^* \in \tilde{O}$ un point régulier qui soit également un extremum de J sur \tilde{O} .

Alors il existe m réels $\lambda_1, \dots, \lambda_m$ appelés **multiplicateurs de Lagrange**, tels que :

$$\nabla \left(J - \sum_{i=1}^m \lambda_i \theta_i \right) (x^*) = 0$$

La quantité $\mathcal{L}(x) = J(x) - \sum_{i=1}^m \lambda_i \theta_i(x)$ est appelé **Lagrangien**.

Remarque 2.7 : C'est donc une condition nécessaire : tout extremum point régulier est un point critique de $J - \sum_{i=1}^m \lambda_i \theta_i$ pour les bons λ_i . Cela permet ainsi de calculer explicitement cet extremum.

Démonstration 2.1 :

On va démontrer ce théorème en utilisant le théorème d'inversion locale.

On a $J : \mathbb{R}^n \rightarrow \mathbb{R}$ et m contraintes $\theta_i : \mathbb{R}^n \rightarrow \mathbb{R}$ pour $i \in \llbracket n - m + 1; n \rrbracket$. On imagine que l'on a pu exhiber m fonctions $X_{n-m+1}, \dots, X_n : \mathbb{R}^{n-m} \rightarrow \mathbb{R}$ telles que :

$$\theta_i(x_1, \dots, x_{n-m}, X_{n-m+1}(x_1, \dots, x_{n-m}), \dots, X_n(x_1, \dots, x_{n-m})) = 0 \quad \forall i \in \llbracket n - m + 1; n \rrbracket$$

(c'est-à-dire qu'on a réussi à exprimer sur \tilde{O} les m dernières coordonnées en fonction des $n - m$ premières.)

On pose :

$$\tilde{\theta}_i(x_1, \dots, x_{n-m}) = \theta_i(x_1, \dots, x_{n-m}, X_{n-m+1}(x_1, \dots, x_{n-m}), \dots, X_n(x_1, \dots, x_{n-m}))$$

et

$$\tilde{J}(x_1, \dots, x_{n-m}) = J(x_1, \dots, x_{n-m}, X_{n-m+1}(x_1, \dots, x_{n-m}), \dots, X_n(x_1, \dots, x_{n-m}))$$

des fonctions de \mathbb{R}^{n-m} dans \mathbb{R} . Le but est de trouver l'extremum de \tilde{J} sous contraintes.

Soit x^* un point critique de \tilde{J} :

$$\forall i \in \llbracket 1; n - m \rrbracket, \quad \frac{\partial \tilde{J}}{\partial x_i}(x^*) = 0 = \frac{\partial J}{\partial x_i} + \sum_{j=n-m+1}^n \frac{\partial J}{\partial x_j} \frac{\partial X_j}{\partial x_i}$$

Ces équations fournissent x_1^*, \dots, x_{n-m}^* et par extension $x_{n-m+1}^* = X_{n-m+1}(x_1^*, \dots, x_{n-m}^*), \dots, x_n^* = X_n(x_1^*, \dots, x_{n-m}^*)$.

D'autre part les θ_i sont nulles, ce qui donne :

$$\forall i \in \llbracket n-m+1; n \rrbracket, \forall j \in \llbracket 1; n-m \rrbracket, \quad \frac{\partial \tilde{\theta}_i}{\partial x_j}(x^*) = 0 = \frac{\partial \theta_i}{\partial x_j} + \sum_{l=n-m+1}^n \frac{\partial \theta_i}{\partial x_l} \frac{\partial X_l}{\partial x_j} \quad (2.1)$$

On a $m(n-m)$ quantités $\frac{\partial X_l}{\partial x_j}$ que l'on souhaite exprimer en fonction des $\frac{\partial \theta_i}{\partial x_j}$, $i \in \llbracket n-m+1; n \rrbracket, \forall j \in \llbracket 1; n-m \rrbracket$ (également $m(n-m)$ quantités).

Or (2.1) donne : $\sum_{l=n-m+1}^n \underbrace{\left(-\frac{\partial \theta_i}{\partial x_l}\right)}_{(a)} \underbrace{\frac{\partial X_l}{\partial x_j}}_{(b)} = \underbrace{\frac{\partial \theta_i}{\partial x_j}}_{(c)}$, soit matriciellement : $\underbrace{M}_{(a)} \underbrace{\frac{\partial X}{\partial x_j}}_{(b)} = \underbrace{\frac{\partial \theta}{\partial x_j}}_{(c)}$

Exemple 2.2 : Écrivons explicitement le cas $m = 2$:

$$M = \begin{bmatrix} -\frac{\partial \theta_{n-1}}{\partial x_{n-1}} & -\frac{\partial \theta_{n-1}}{\partial x_n} \\ -\frac{\partial \theta_n}{\partial x_{n-1}} & -\frac{\partial \theta_n}{\partial x_n} \end{bmatrix} \quad \frac{\partial X}{\partial x_j} = \begin{pmatrix} \frac{\partial X_{n-1}}{\partial x_j} \\ \frac{\partial X_n}{\partial x_j} \end{pmatrix} \quad \frac{\partial \theta}{\partial x_j} = \begin{pmatrix} \frac{\partial \theta_{n-1}}{\partial x_j} \\ \frac{\partial \theta_n}{\partial x_j} \end{pmatrix} \quad \forall j \in \llbracket 1; n-m \rrbracket$$

Revenons à la condition $\frac{\partial \tilde{J}}{\partial x_i} = 0$, $i \in \llbracket 1; n-m \rrbracket$. Pour $m = 2$ on a :

$$\frac{\partial J}{\partial x_i} + \frac{\partial J}{\partial x_{n-1}} \frac{\partial X_{n-1}}{\partial x_i} + \frac{\partial J}{\partial x_n} \frac{\partial X_n}{\partial x_i} = 0$$

D'où en utilisant $\frac{\partial X}{\partial x_i} = M^{-1} \frac{\partial \theta}{\partial x_i}$:

$$\frac{\partial J}{\partial x_i} + (M^{-1})_{11} \frac{\partial \theta_{n-1}}{\partial x_i} \frac{\partial J}{\partial x_{n-1}} + (M^{-1})_{12} \frac{\partial \theta_n}{\partial x_i} \frac{\partial J}{\partial x_{n-1}} + (M^{-1})_{21} \frac{\partial \theta_{n-1}}{\partial x_i} \frac{\partial J}{\partial x_n} + (M^{-1})_{22} \frac{\partial \theta_n}{\partial x_i} \frac{\partial J}{\partial x_n} = 0$$

Et donc en posant $-\lambda_1 = (M^{-1})_{11} \frac{\partial J}{\partial x_{n-1}} + (M^{-1})_{21} \frac{\partial J}{\partial x_n}$ et $-\lambda_2 = (M^{-1})_{12} \frac{\partial J}{\partial x_{n-1}} + (M^{-1})_{22} \frac{\partial J}{\partial x_n}$:

$$\frac{\partial J}{\partial x_i} - \lambda_1 \frac{\partial \theta_{n-1}}{\partial x_i} - \lambda_2 \frac{\partial \theta_n}{\partial x_i} = 0$$

On constate donc que pour $i \in \llbracket 1; n-m \rrbracket$, x^* vérifie :

$$\frac{\partial}{\partial x_i} \left(J - \sum_{j=n-m+1}^n \lambda_j \theta_j \right) (x^*) = 0$$

Mais cette égalité est-elle également vraie pour $i \in \llbracket n-m+1; n \rrbracket$?

Vérifions dans le cas $m = 2$ en substituant ($i = n-1$ dans ce qui suit) :

$$\begin{aligned}
 & \frac{\partial J}{\partial x_{n-1}} - \lambda_1 \frac{\partial \theta_{n-1}}{\partial x_{n-1}} - \lambda_2 \frac{\partial \theta_n}{\partial x_{n-1}} \\
 = & \frac{\partial J}{\partial x_{n-1}} + \left[(M^{-1})_{11} \frac{\partial J}{\partial x_{n-1}} + (M^{-1})_{21} \frac{\partial J}{\partial x_n} \right] (-M_{11}) + \left[(M^{-1})_{12} \frac{\partial J}{\partial x_{n-1}} + (M^{-1})_{22} \frac{\partial J}{\partial x_n} \right] (-M_{21}) \\
 = & \frac{\partial J}{\partial x_{n-1}} - ((M^{-1})_{11} M_{11} + (M^{-1})_{12} M_{21}) \frac{\partial J}{\partial x_{n-1}} - ((M^{-1})_{21} M_{11} + (M^{-1})_{22} M_{21}) \frac{\partial J}{\partial x_n} \\
 = & \frac{\partial J}{\partial x_{n-1}} - (M^{-1} M)_{11} \frac{\partial J}{\partial x_{n-1}} - (M^{-1} M)_{21} \frac{\partial J}{\partial x_n} \\
 = & \frac{\partial J}{\partial x_{n-1}} - 1 \times \frac{\partial J}{\partial x_{n-1}} - 0 \times \frac{\partial J}{\partial x_n} \\
 = & 0
 \end{aligned}$$

Conclusion : il existe $\lambda_1, \dots, \lambda_{n-m+1}$ tels que les points critiques de J vérifiant les contraintes soient points critiques de $\mathcal{L} = J - \sum \lambda_j \theta_j$.

Remarque 2.8 : Cette méthode est extrêmement puissante. Elle permet notamment de retrouver facilement des résultats classiques comme l'inégalité arithmético-géométrique :

$$\forall (x_1, \dots, x_n) \in (\mathbb{R}^+)^n, \quad \frac{1}{n} \sum_{i=1}^n x_i \geq \sqrt[n]{x_1 \dots x_n}$$

Exercice : On introduit $J(x_1, \dots, x_n) = \prod_{i=1}^n x_i$ et $\theta(x_1, \dots, x_n) = \sum_{i=1}^n x_i - s$. Montrer l'inégalité précédente à l'aide des multiplicateurs de Lagrange.

2.4.2 Optimisation avec contraintes inégalités

On revient à $O = \{x \in \mathbb{R}^n \mid \forall i \in \llbracket 1; m \rrbracket, \theta_i(x) \leq 0\}$ et on cherche $\min_{x \in O} J(x)$.

Définition 2.14 : On dit que $\theta_i \leq \bar{0}$ est une contrainte **active** en $v \in O$ si $\theta_i(v) = 0$. On pose $I(v) = \{i \in \llbracket 1; m \rrbracket \mid \theta_i(v) = 0\}$.

Théorème 2.16 : Condition d'optimalité de Krush-Kuhn-Tucker.

Si x^* est un point régulier et un minimum local de J sur O , il existe m réels $p_1, \dots, p_m \geq 0$ (multiplicateurs de KKT-Lagrange) tels que :

$$\nabla \left(J + \sum_{i=1}^m p_i \theta_i \right) (x^*) = 0 \quad \text{et} : \quad \forall i \in \llbracket 1; m \rrbracket, \quad p_i \theta_i(x^*) = 0$$

Remarque 2.9 : C'est une condition nécessaire qui devient suffisante si J et les $(\theta_i)_i$ sont convexes.

Chapitre 3

Distributions

3.1 Motivations

3.1.1 Loi piquée

Si X_1, \dots, X_n sont des v.a.i.i.d centrées et de variance finie σ^2 , on sait que la loi de $u = \frac{1}{n} \sum x_i$ est proche de $p_n(u) = \frac{1}{\sqrt{2\pi\sigma^2/n}} e^{-\frac{nu^2}{2\sigma^2}}$. Cependant, la représentation de p_n possède toujours une largeur σ , que l'on pourrait vouloir faire tendre vers 0 pour modéliser une certitude que nos v.a. valent 0. De plus, en général on ne s'intéresse pas à p_n mais à :

$$\begin{aligned} \langle f(u) \rangle &= \int p_n(u) f(u) du \\ &= \int \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{u'^2}{2\sigma^2}\right) f\left(\frac{u'}{\sqrt{n}}\right) du' \\ &\xrightarrow{n \rightarrow +\infty} f(0) \end{aligned}$$

► On aimerait donc définir une forme de loi de probabilité infiniment piquée en 0 et en lien avec $f(0)$.

3.1.2 Électromagnétisme

Le champ créé par une charge ponctuelle q située à l'origine s'écrit $\vec{E} = \frac{1}{4\pi\epsilon_0} \frac{\vec{r}}{r^3}$.

D'après les lois de Maxwell, $\text{div} \vec{E} = 0$ en tout $r \neq 0$. En supposant que le théorème d'Ostrogradsky-Green s'applique (ce qui n'est pas le cas puisque $\mathbb{R}^3 \setminus \{0\}$ n'est pas un ouvert étoilé, mais l'on cherche des motivations à ce qui va suivre), on aurait alors :

$$\iiint \text{div} \vec{E} dV = \oiint \vec{E} \cdot d\vec{S} = \frac{Q_{\text{int}}}{\epsilon_0} = \frac{q}{\epsilon_0} \neq 0, \quad \text{mais alors } \text{div} \vec{E} \text{ n'est pas nulle partout...}$$

► On voudrait un moyen de pouvoir écrire $\text{div} \vec{E} = \frac{\rho}{\epsilon_0}$ et d'intégrer cette relation même là où elle est mal définie (i.e. en 0).

3.1.3 Élastique

Considérons une force s'exerçant entre deux sphères de rayon σ repérées par \vec{r}_1 et \vec{r}_2 : le potentiel s'écrit $V(\vec{r}_1 - \vec{r}_2)$, $\vec{F}_{2/1} = -\vec{\nabla}_{\vec{r}_1} V$. Mais ces objets ne peuvent s'interpénétrer donc V est infini en $r = \sigma$!

► Peut-on définir cette discontinuité proprement ?

3.1.4 Problèmes linéaires

— Si A est une matrice inversible, $Y = AX \implies X = A^{-1}Y$

— Si $\dot{x} + ax = y$, on aimerait écrire $y = Ax \implies x = A^{-1}y$

► A-t-on $AA^{-1} = Id$? Comment définir Id ?

3.2 L'objet mathématique

3.2.1 L'espace vectoriel des fonctions tests

Définition 3.1 :

$$\mathcal{D} = \{\varphi : \mathbb{R}^n \rightarrow \mathbb{R} \text{ ou } \mathbb{C} \mid \varphi \in \mathcal{C}^\infty \text{ et } \exists K \subseteq \mathbb{R}^n \text{ compact, } \forall x \notin K, \varphi(x) = 0\}$$

\mathcal{D} est l'espace vectoriel des **fonctions \mathcal{C}^∞ à support borné**. On appelle cet espace l'**ensemble des fonctions tests**.

Exemple 3.1 : La fonction suivante appartient à \mathcal{D} :

$$\rho : \mathbb{R} \rightarrow \mathbb{R}, \quad \rho(x) = \begin{cases} \exp\left(-\frac{1}{1-x^2}\right) & \text{si } x \in [-1; 1] \\ 0 & \text{sinon} \end{cases}$$

Théorème 3.1 : Un théorème d'approximation.

Toute fonction continue f à support borné peut être approchée uniformément par une suite de fonctions de \mathcal{D} , et on peut astreindre les supports de ces fonctions à être contenues dans un voisinage arbitraire de celui de f .

Définition : Convergence sur l'espace des fonctions tests.

On définit le **mode de convergence** suivant sur l'ensemble des fonctions tests :

$$\forall (\varphi_n)_{n \in \mathbb{N}} \in \mathcal{D}^{\mathbb{N}}, \varphi_n \xrightarrow{n \rightarrow +\infty} \varphi \in \mathcal{D} \text{ ssi } \begin{cases} \text{i) } \exists K \text{ compact tel que } \forall n \in \mathbb{N}, \text{Supp } \varphi_n \subseteq K \text{ et } \text{Supp } \varphi \subseteq K \\ \text{ii) } \forall k \in \mathbb{N}, \varphi_n^{(k)} \xrightarrow{CVU} \varphi^{(k)} \text{ i.e. } \|\varphi_n^{(k)} - \varphi^{(k)}\|_\infty \xrightarrow{n \rightarrow +\infty} 0 \end{cases}$$

3.2.2 Distributions

Définition 3.2 : On appelle **distribution** toute **forme linéaire continue sur l'espace des fonctions tests** (au sens séquentiel, avec la convergence définie précédemment).

On ne notera pas $T(\varphi)$ mais :

$$\begin{aligned} T : \mathcal{D} &\rightarrow \mathbb{R} \text{ ou } \mathbb{C} \\ \varphi &\mapsto \langle T, \varphi \rangle \end{aligned}$$

$T : \mathcal{D} \rightarrow \mathbb{R} \text{ ou } \mathbb{C}$ est donc une distribution si :

- (i) $\forall (\varphi, \psi) \in \mathcal{D}^2, \forall \lambda \in \mathbb{R} \text{ ou } \mathbb{C}, \quad \langle T, \varphi + \lambda\psi \rangle = \langle T, \varphi \rangle + \lambda \langle T, \psi \rangle$
- (ii) $\forall (\varphi_n)_{n \in \mathbb{N}} \in \mathcal{D}^{\mathbb{N}}, \varphi_n \xrightarrow{n \rightarrow +\infty} \varphi \implies \langle T, \varphi_n \rangle \xrightarrow{n \rightarrow +\infty} \langle T, \varphi \rangle$

Notation : On constate que les distributions forment un sous-espace vectoriel du dual \mathcal{D}^* de \mathcal{D} , que l'on note \mathcal{D}' . $(T, \varphi) \in \mathcal{D}' \times \mathcal{D} \mapsto \langle T, \varphi \rangle$ peut ainsi être vue comme une forme bilinéaire.

Définition 3.3 : On note $L_{\text{loc}}^1(\mathbb{R})$ l'ensemble des **fonctions localement sommables**, c'est-à-dire telles que $\forall K$ compact, $\int_K |f| < +\infty$.

Propriété 3.1 : Soit $f \in L_{\text{loc}}^1(\mathbb{R})$, alors $T_f : \mathcal{D} \rightarrow \mathbb{R} \text{ ou } \mathbb{C}$ définie par :

$$\boxed{\langle T_f, \varphi \rangle = \int_{\mathbb{R}} f(x) \varphi(x) dx}$$

constitue une distribution. (L'intégrale n'est jamais sur \mathbb{R} entier puisque φ à support borné.)

Théorème 3.2 :

Si f et g sont deux fonctions égales presque partout, alors $T_f = T_g$. Réciproquement, si $T_f = T_g$, alors f et g sont presque partout égales.

Exemples 3.2 :

1. Si $f \in L_{\text{loc}}^1(\mathbb{R})$, on voit que $\langle T_f, \varphi \rangle = \int f \varphi = \langle f, \varphi \rangle$: par abus de notation on confondra $T_f = f$ dans l'écriture \langle, \rangle .
2. $\langle 1, \varphi \rangle = \int \varphi$
3. Si D est l'endomorphisme de dérivation dans \mathcal{D} , $\langle D, \varphi \rangle = \int D(\varphi)$ définit également une distribution.

Définition 3.4 : Distribution de Dirac.

Si $a \in \mathbb{R}^n$, on appelle distribution de Dirac la distribution suivante :

$$\boxed{\langle \delta_a, \varphi \rangle = \varphi(a)}$$

Elle correspond donc à l'évaluation en a , qui est bien une forme linéaire continue. Pour $a = 0$, on notera simplement $\delta_0 = \delta$.

3.2.3 Caractéristiques des distributions

Définition 3.5 : On dit que $T \in \mathcal{D}'$ est une distribution **régulière** s'il existe $f \in L^1_{\text{loc}}(\mathbb{R})$ et K compact telle que $T = T_f \mathbb{1}_K$ (c.f. propriété 3.1). Dans le cas contraire, on dit que T est une distribution **singulière**.

Exemple 3.3 : La distribution de Dirac δ est singulière. (Pour autant, on verra que toutes les propriétés naturelles des distributions singulières s'étendent bien aux distributions singulières, ainsi on ne se privera pas d'écrire $\int \delta(x)f(x)dx = f(0)$ de façon courante en physique bien que cet objet n'existe pas.)

Définition 3.6 : Si T_1 et T_2 sont deux distributions, on dit que $T_1 = T_2$ sur Ω ouvert de \mathbb{R}^n si :

$$\forall \varphi \in \mathcal{D} \text{ telle que } \text{Supp } \varphi \subseteq \Omega, \langle T_1, \varphi \rangle = \langle T_2, \varphi \rangle$$

Définition 3.7 : On définit le **support** d'une distribution T comme le complémentaire de l'union des ouverts sur lesquels elle est nulle (au sens de la définition précédente), cette union étant toujours un ouvert.

$$\text{Supp } T = \overline{\bigcup_{\substack{\Omega \text{ ouvert} \\ T=0 \text{ sur } \Omega}} \Omega}$$

Exemple 3.4 : $\text{Supp } \delta_\alpha = \{\alpha\}$.

3.3 Dérivation

3.3.1 Définitions

Supposons que l'on se donne une fonction f de classe \mathcal{C}^1 sur \mathbb{R}^n . On a alors :

$$\begin{aligned} \left\langle \frac{\partial f}{\partial x_1}, \varphi \right\rangle &= \left\langle T_{\partial_1 f}, \varphi \right\rangle = \int \frac{\partial f}{\partial x_1} \varphi d^n x \\ &= \int dx_2, \dots, dx_n \left([f(x)\varphi(x)]_{x_1=-\infty}^{x_1=+\infty} - \int dx_1 f(x) \frac{\partial \varphi}{\partial x_1} \right) \\ &= - \int f \frac{\partial \varphi}{\partial x_1} d^n x \\ &= - \left\langle f, \frac{\partial \varphi}{\partial x_1} \right\rangle \end{aligned}$$

Ceci donne donc l'idée de définir une dérivation sur l'espace des distributions compatibles avec la dérivation classique.

Définition 3.8 : Si $T \in \mathcal{D}'$, on définit la **dérivée de T** par :

$$\begin{aligned} T' : \mathcal{D} &\rightarrow \mathbb{R} \text{ ou } \mathbb{C} \\ \varphi &\mapsto \boxed{\langle T', \varphi \rangle = -\langle T, \varphi' \rangle} \end{aligned}$$

- Si φ est à plusieurs variables : $\left\langle \frac{\partial T}{\partial x_i}, \varphi \right\rangle = -\left\langle T, \frac{\partial \varphi}{\partial x_i} \right\rangle$
- Pour les dérivées d'ordre supérieur :

$$\left\langle \frac{\partial^2 T}{\partial x \partial y}, \varphi \right\rangle = -\left\langle \frac{\partial T}{\partial y}, \frac{\partial \varphi}{\partial x} \right\rangle = +\left\langle T, \frac{\partial^2 \varphi}{\partial y \partial x} \right\rangle \underset{\text{Schwarz, } \varphi \in C^\infty}{=} +\left\langle T, \frac{\partial^2 \varphi}{\partial x \partial y} \right\rangle$$

- De façon générale :

$$\boxed{\langle \partial_{x_1}^{p_1} \dots \partial_{x_n}^{p_n} T, \varphi \rangle = (-1)^{\sum p_i} \langle T, \partial_{x_1}^{p_1} \dots \partial_{x_n}^{p_n} \varphi \rangle}$$

Exemple 3.5 : $\langle \Delta T, \varphi \rangle = \langle T, \Delta \varphi \rangle$ où Δ est l'opérateur Laplacien.

Remarque 3.1 : On peut, comme pour les fonctions, voir que :

$$T' = \lim_{h \rightarrow 0} \frac{\tau_h T - T}{h} \quad \text{où :} \quad \langle \tau_h T, \varphi \rangle = \langle T, x \mapsto \varphi(x - h) \rangle$$