# Chapitre III

# Mesure et intégration

**Sommaire**



## III.1 Motivations, vue d'ensemble

Cette première section précise au travers d'exemples la nature des objets abstraits construits dans les sections suivantes, et les difficultés associées à cette construction. La notion centrale est celle de *mesure*. Comme cadre conceptuel d'appréhension du réel, cette notion unique de mesure répond à deux enjeux, qu'il nous paraît important de distinguer malgré le fait qu'ils correspondent à la même notion mathématique.

En premier lieu, une mesure permet de structurer le fond d'un espace destiné à accueillir de la matière. Par espace nous entendons par exemple l'espace euclidien usuel, qui est en dimension 3 un modèle de l'espace physique dans lequel nous vivons, sur lequel il peut être pertinent de définir des *champs* (champ de densité, de concentration d'un polluant, de température, de densité de population, ...). Considérons par exemple un milieu occupant une certaine zone de l'espace euclidien, milieu dont on connait la densité. Si l'on suppose la densité constante sur une zone $A$, la masse portée par $A$ est le produit entre cette valeur de densité et le volume de la zone. Il est donc essentiel de savoir estimer le volume des zones susceptibles d'accueillir de la matière, pour pouvoir estimer la masse correspondante. Définir une mesure consiste précisément à concevoir une procédure pour associer à une zone son volume. Même s'il n'est pas dans les usages d'affecter une unité physique aux grandeurs mathématiques, on pourra concevoir cette mesure comme s'exprimant en unité de volume (ou d'aire s'il s'agit de l'espace bi-dimensionnel, ou de longueur s'il s'agit d'un espace à une dimension). Il s'agit d'une donnée *statique* associée à l'espace considéré. Dans le cas de l'espace euclidien, ce volume est canoniquement défini dans le cas de formes simples : longueur d'un intervalle, aire d'un rectangle, volume d'un parallélogramme. La notion d'intégration d'une fonction constante sur de tels ensembles est basée sur le simple produit de la valeur à intégrer par le volume. Si, suivant l'intuition associée à la notion de volume, on décrète que le volume de la réunion de deux zones disjointes est la somme des volumes des zones élémentaires, on peut estimer le volume de toutes les zones qui peuvent se construire comme réunion disjointe finie de ces formes simples. Définir le volume de n'importe quel ensemble est plus délicat et même, d'une certaine manière, impossible, comme nous le verrons. La construction de la mesure de Lebesgue, qui est un point essentiel des sections qui suivent, permettra de définir un tel volume pour une classe très générale de zones de l'espace euclidien, et permettra de construire un cadre définissant la notion d'intégrales pour des fonctions très générales.

Les mesures ont également vocation à représenter des quantités absolues de matière (fluides, matériau solide, cellules, individus, ...), distributions d'une certaine substance susceptible d'évoluer en temps, d'être transportée, supprimée, développée. L'objet mathématique associé est le même, mais la nature de la réalité qu'il a vocation à représenter est différente. Il sera ici naturel de penser la mesure associée comme exprimée en kg, en moles ...., qui mesurent des quantités de matières associées à des principes de conservation.

Nous proposons dans les paragraphes qui suivent quelques exemples de situations réelles qui illustrent les deux types de mesures évoqués ci-dessus et les liens qu'elles entretiennent : mesures de type *volume*, qui formalisent la capacité de parties de l'espace sous-jacent à accueillir de la matière, et mesures de type *masse*, qui représentent des quantités de choses réelles. Nous nous restreignons dans ces exemples à des ensembles finis, de telle sorte que les objets mathématiques sont très simples à définir. Nous évoquerons dans la suite de cette introduction les difficultés posées par la construction de telles mesures pour des ensembles infinis.

**Superficies, densités, et nombre d'habitants.**
Considérons l'ensemble $X = [\![1, N]\!]$ des grandes villes françaises, numérotées de 1 à $N$. On note $\mu_i > 0$ la superficie de la ville $i$. À la collection des $\mu_i$ est naturellement associée une

application $\mu$ de l'ensemble $\mathcal{P}(X)$ des parties de $X$ dans $\mathbb{R}_+$ :

$$\mu \ : \ A \in \mathcal{P}(X) \longmapsto \mu(A) = \sum_{i \in A} \mu_i \in \mathbb{R}_+. \tag{III.1.1}$$

Cette application est *additive* au sens où, si $A$ et $B$ sont disjoints, $\mu(A \cup B) = \mu(A) + \mu(B)$. Pour reprendre une terminologie physique, cette application définit une variable *extensive*.

Il s'agit d'une mesure au sens volumique évoqué ci-dessus, qui structure l'ensemble des villes en termes de capacité d'accueil. Notons maintenant $\rho_i$ la densité d'habitants dans la ville $i$. Il s'agit là d'une variable *intensive*[1]. Le produit $m_i = \rho_i \mu_i$ est le nombre d'habitants dans cette ville $i$. On peut, comme précédemment pour les $\mu_i$, associer à la collection des villes une application $m$ de $\mathcal{P}(X)$ dans $\mathbb{R}$, additive par construction. Cette application est une nouvelle mesure sur $X$, de type "masse". Le nombre total d'habitants dans le sous-ensemble $A \subset X$ de villes peut s'écrire comme un produit de dualité[2] noté $\langle \rho \, , \mu \rangle_A$ entre les collections de superficies et de densités

$$m_A = \langle \rho \, , \mu \rangle_A = \sum_{i \in A} \rho_i \mu_i.$$

Il s'agit là de la version discrète d'une intégrale, construite par mise en dualité d'une mesure volume ($\mu$, version discrète de la mesure de Lebesgue construite plus loin) et d'une variable intensive (densité $\rho$, qui joue le rôle d'une fonction à intégrer sur un domaine). La mesure masse $m$ est la variable sommable, produit de la variable extensive $\mu$ et la variable intensive $\rho$.

On peut aussi définir, dans le cas présent d'une collection finie de villes, des mesures qui correspondent à des probabilités. Prenons l'exemple d'un crime commis à Paris à l'heure H d'un jour J. Vingt-quatre heures après, l'assassin court toujours, et les enquêteurs cherchent à estimer dans quelle ville il pourrait être. L'état de leur opinion concernant la position du fugitif peut être encodé par une mesure $m = (m_i)$. Si l'on sait qu'il ne dispose pas de véhicule et que l'on considère que prendre le train était risqué pour lui, on considèrera que la probabilité associée à Paris est de 0.75. Si l'on sait qu'il a des contacts à Lyon, on évaluera à 0.15 la probabilité qu'il y soit, le complément étant distribué sur le reste du pays en fonction des informations que l'on peut avoir. On a ici l'exemple typique d'une mesure (ici de probabilité, c'est à dire normalisée à 1) qui évolue au cours du temps, en fonction des informations reçues.

L'intérêt d'introduire la notion de *mesure* pour les exemples ci-dessus, alors que les objets manipulés se ramènent à des tableaux de nombres réels, n'est pas immédiat. Nous verrons qu'il est néanmoins fécond de considérer par exemple la collection $\mu = (\mu_i)$ des superficies comme une application qui, à un ensemble de villes $I \subset [\![1, N]\!]$, associe la population totale des villes concernées, selon l'expression (III.1.1). Cette application attribue 0 à l'ensemble vide, et vérifie par construction la règle de sommation suivante : l'image de la réunion de deux ensembles disjoints est la somme des images (on dira que l'application est *additive*), ce qui peut s'écrire

$$A \cap B = \emptyset \Longrightarrow \mu\,(A \cup B) = \mu\,(A) + \mu\,(B)\,.$$

1. La densité associée à la réunion de deux villes de même densité $\rho$ est $\rho$, et pas $2\rho$.

2. Un produit de dualité entre deux espaces vectoriels $E$ et $F$ de même dimension est simplement une application bilinéaire de $E \times F$ dans $\mathbb{R}$. On dit que cette application met les espaces en dualité. L'exemple le plus simple est le cas d'un espace euclidien, qui est en dualité avec lui même par le biais de son produit scalaire.

Nous définirons une mesure comme une application qui à une partie associe un réel positif, et qui vérifie des conditions du type de celles qui précèdent.

**Aérosols.**
On considère maintenant une collection de $N$ micro-gouttelettes sphériques flottant dans l'air. Si l'on note $\mu_i$ le volume de la gouttelette $i$, on peut définir une application de l'ensemble des parties de $X = [\![1, N]\!]$ dans $\mathbb{R}^+$ associant à une sous-collection de gouttelettes son volume total. Si l'on note $\rho$ la densité du fluide considéré, on peut associer à la collection une nouvelle mesure, de type masse, simplement définie par ses valeurs en chaque entité, $m_i = \rho\mu_i$, la mesure associée, définie comme application de $\mathcal{P}(X)$ dans $\mathbb{R}_+$, s'en déduisant simplement par additivité. On a ainsi construit une nouvelle mesure exprimant une variable extensive, construite comme produit d'une première mesure volume avec une variable intensive. On peut dans ce contexte continuer l'empilement des mesures en considérant que chaque particule est animée d'une vitesse $u_i$. Cette collection de vitesses peut être vue comme une fonction sur $X$. Cette variable vectorielle intensive peut être adossée avec la mesure $m$ (extensive) pour former une nouvelle variable extensive (la quantité de mouvement), construite selon $p_i = m_i u_i$. Il s'agit de la version discrète de ce que l'on appellera une *mesure vectorielle*. C'est une variable extensive (la quantité de mouvement d'un système est la somme des quantités de mouvement de ses constituants). Dans ce contexte, on dira que la vitesse est mesurable $m$-presque partout. Ici, l'ensemble étant fini, cela signifie simplement que cela n'a pas de sens de définir la vitesse d'un objet qui n'a pas de masse, puisque cette vitesse sans masse ne pourrait intervenir d'aucune manière dans un modèle mécanique cohérent.

On remarquera que la variable intensive vitesse peut être intégrée selon cette nouvelle mesure vectorielle, pour former une quantité scalaire qui représente l'énergie cinétique

$$E_A = \langle u\,, p\rangle_A = \frac{1}{2}\sum_{i\in A} m_i u_i^2.$$

**Vers l'infini : le cas de l'intervalle** $]0, 1[$
Les cadres présentés ci-dessus peuvent être étendu assez naturellement à des ensembles dénombrables, on remplace alors les sommes finies par des sommes infinies, des séries, de nombres positifs, en acceptant éventuellement que la série puisse prendre la valeur $+\infty$. On remarquera néanmoins que, s'il est possible d'affecter une masse à chaque point d'une collection dénombrable de façon à ce que la masse totale soit *finie*, la distribution est forcément inégalitaire, ou identiquement nulle. En effet, si chaque point de notre ensemble dénombrable a une masse $m$, on a l'alternative suivante : si $m > 0$ la masse totale est infinie, et si $m = 0$ la masse totale est nulle. Une version temporelle de cet énoncé, qui évoque le paradoxe d'Achille et de la tortue, pourrait être : disposant d'un temps fini, on peut faire une infinité de choses qui chacune prend un certain temps, mais c'est impossible en attribuant un temps identique à chacune des tâches. On retrouvera cet argument très simple au cœur de la construction d'un des ensembles pathologiques évoqués ci-après.

Les véritables difficultés commencent lorsque l'on s'intéresse à des ensembles qui ont ce que l'on appelle la *puissance du continu*, comme la droite réelle, ou l'espace physique $\mathbb{R}^3$. Considérons pour fixer les idées le cas de l'intervalle réel $X = ]0, 1[$. On cherche à définir sur cet ensemble une notion de volume (il s'agit plutôt en l'occurrence d'une notion de longueur, que nous verrons ici comme un volume monodimensionnel). Plus précisément, on cherche à construire une *mesure*, c'est-à-dire une application $\mu$ qui à une partie $A$ de $]0, 1[$ associe un nombre réel positif ou nul, et qui généralise à des ensembles quelconques la notion de longueur. On souhaite donc en particulier que $\mu(]a,b[) = b - a$. Le caractère extensif de la notion de longueur impose une propriété d'additivité. On demande donc que la mesure d'une union d'ensembles disjoints soit égale à la somme des mesures des ensembles. Comme nous le verrons plus loin, il est nécessaire pour aboutir à une notion "utilisable" que cette propriété s'étende à des collections dénombrables de parties, on parlera de *$\sigma-$additivité*. L'intervalle fermé $[a,b]$ étant l'intersection des intervalles$]a - 1/n, b + 1/n[$, sa longueur est la même que celle de l'intervalle ouvert. On en déduit que la mesure des singletons (comme les extrémités de l'intervalle) est nulle. On peut étendre immédiatement cette mesure à des réunions dénombrables d'intervalles, mais on se heurte ensuite à un mur. Pour des raisons assez profondes qui tiennent à la nature même de la droite réelle, et malgré l'apparente simplicité du problème, il est *impossible* de définir une telle application, qui affecterait aux intervalles leurs longueurs, qui serait $\sigma$-additive (manière distinguée de dire que cela correspond à une variable extensive), qui affecterait à une partie quelconque de l'intervalle [3] $]0,1[$ ce qu'il conviendrait alors d'appeler sa longueur. On peut contourner le problème par le haut en suivant un principe inhérent à la notion intuitive de volume : si un ensemble est inclus dans un autre, ce dernier a un plus gros volume. Si l'on se donne $A \subset ]0,1[$, on peut considérer l'ensemble des collections dénombrables d'intervalles (on s'affranchit du caractère disjoint des collections) qui recouvrent $A$. Si l'on était capable de définir une mesure pour $A$, cette mesure serait inférieure où égale à la mesure de toute collection qui recouvre $A$, qui est elle-même inférieure à la somme des longueurs des intervalles. Il est ainsi naturel de considérer la quantité $\mu^\star(A)$ définie comme l'infimum de la somme des longueurs des intervalles, infimum sur l'ensemble des collections qui recouvrent $A$. On appellera cette quantité la *mesure extérieure* de Lebesgue de $A$. Cette démarche conduit néanmoins à un problème : il apparaît qu'il existe des parties de $X$ qui vérifient des propriétés que nous qualifierons de *bizarres*. Il existe en effet des ensembles $B$, dont le complémentaire dans $X$ est noté $B^c$, qui conduisent à une violation de la propriété d'additivité que l'on souhaite voir vérifier par la mesure. Plus précisément, il existe certaines parties $B$ telles que, pour certaines parties $A$, l'identité

$$\mu^\star(A) = \mu^\star(A \cap B) + \mu^\star(A \cap B^c)$$

n'est *pas vérifiée*. Plus précisément $\mu^\star(A)$ est strictement inférieur à la somme des mesures des parties disjointes $A \cap B$ et $A \cap B^c$ qui le constituent. Le mathématicien se retrouve dans la position d'un arpenteur étudiant une région $A$, composée exclusivement de 2 propriétés $A_1$ et $A_2$ sans recouvrement, imbriquées l'une dans l'autre de façon extrêmement complexe, et telle que l'aire estimée de $A$ selon la méthode évoquée ci-dessus est *strictement inférieure* à la somme des aires de $A_1$ et $A_2$.

Il n'existe pas de manière complètement satisfaisante de régler ce nouveau problème. La démarche conduisant à des "monstres", on choisit simplement de les exclure de l'approche, et de se concentrer sur les parties $B$ pour lesquelles l'identité ci-dessus est vérifiée pour toute partie $A$ (parties appelées *mesurables*, et dont la collection s'appelle une *tribu* comme on le verra) pour définir une mesure. Cette mesure, qui est la restriction de la mesure extérieure ci-dessus à la collection $\mathcal{A}$ des ensembles mesurables, vérifie alors de bonnes propriétés, au prix de *l'exclusion de certains ensembles pathologiques*, qu'il est d'ailleurs impossible de décrire

3. On peut aussi formuler ce problème dans le plan $\mathbb{R}^2$ en considérant à la place des intervalles des rectangles, dont on sait calculer l'aire, ou dans l'espace physique $\mathbb{R}^3$ en considérant des pavés (i.e. parallélépipèdes), dont on sait calculer le volume.

explicitement [4]. Une fois cette construction réalisée, la définition de la notion d'intégrale s'ensuit naturellement. L'intégrale d'une fonction constante égale à $\rho$ (que l'on peut voir ici comme une densité) sur une partie $A$ est simplement le produit $\rho \times \mu(A)$, qui est alors la masse de la matière contenue dans $A$. On peut étendre facilement cette définition aux fonctions qui prennent un nombre fini de valeurs (fonctions dites *simples*, ou *étagées* dans le cadre de la théorie de la mesure) sur des parties mesurables, en sommant simplement les différentes contributions, comme pour calculer la masse d'un objet composite à partir des densités de ses constituants, et des volumes des différentes zones qu'ils occupent. On peut alors étendre cette notion d'intégrale à une classe très générale de fonctions (nous ne considérons pour l'instant que des fonctions positives), appelées *mesurables*, en définissant l'intégrale comme le supremum des intégrales des fonctions étagées qui sont partout inférieures ou égales à la fonction considérée.

**Cadre général.**
La démarche décrite précédemment s'inscrit dans un cadre général qui dépasse le cas particulier de la droite des réels, et qui constitue les bases de la théorie des probabilités. Les sections qui suivent présentent ce cadre abstrait, et en parallèle la construction progressive de l'intégrale de Lebesgue. Le point de départ est la notion de tribu déjà évoquée ci-dessous : une tribu sera définie comme une famille de parties d'un ensemble $X$ qui vérifient un certain nombre de propriétés, essentiellement de stabilité (par complémentarité et par union dénombrable). On définira ensuite la notion de mesure sur une tribu, qui est une application à valeurs dans $\mathbb{R}^+$, et a vocation à affecter à une partie de $X$ son volume. On demandera assez naturellement à ce que "rien" ne prenne pas de place ($\mu(\emptyset) = 0$), et l'on exigera par ailleurs, pour respecter le caractère extensif de la notion que l'on souhaite définir, une propriété d'additivité : la mesure d'une union disjointe (dénombrable) de parties est la somme des mesures de ces parties. On donnera un sens très général à la notion de mesure extérieure, déjà évoquée plus haut dans le cas de l'intervalle $]0, 1[$, définie sur l'ensemble des parties d'un ensemble, en relaxant la propriété d'additivité (remplacée par une propriété de sous-additivité), et en imposant la monotonie (qui n'est plus garantie sinon, du fait que l'on a relaxé l'additivité). On qualifie alors de mesurable une partie qui vérifie la propriété d'additivité évoquée précédemment, et l'on peut montrer une propriété très générale : la familles des parties mesurables est une tribu, et la mesure extérieure restreinte à cette tribu est une *mesure*. C'est ce résultat qui permettra de définir la mesure de Lebesgue à partir de la mesure de Lebesgue extérieure introduite dans le paragraphe précédent.

Terminons cette longue introduction par quelques mots sur la théorie des probabilités, qui constitue une motivation important à l'étude détaillée des notions de tribu, classe monotone, mesure, …, même si elle n'est pas centrale dans ce cours. Dans ce contexte l'ensemble $X$ est vu comme un ensemble d'*éventualités* (on parle de l'univers des possibles, issues possibles d'une expérience). Définir une tribu consiste à définir un sous-ensemble de parties que l'on souhaite considérer comme des *événements*, c'est-à-dire comme des propriétés vérifiées par le résultat de l'expérience, exprimées au travers de l'appartenance à une des sous-parties de la tribu. Par exemple si l'on sait qu'une météorite est tombée en Europe, on concevra $X$ comme l'ensemble des positions géographiques de cette zone (que l'on peut identifier à une carte au sens usuel du terme). On peut imaginer une tribu comme un ensemble d'assertions potentiellement pertinentes. Par exemple : 'La météorite est tombée en Alsace', 'La météorité n'est pas tombée

4. La construction de ces contre-exemples nécessite l'*axiome du choix*, ce qui confère un caractère très abstrait à ces contre-exemples.

en Alsace', 'La météorite est tombée à l'ouest de Berlin ou au nord d'Helsinki', 'La météorite est tombée en zone urbaine', 'La métérorite n'est pas tombée en Europe' ....
Noter que l'on peut choisir de structurer l'ensemble des assertions potentiellement pertinentes de façon plus ou moins détaillée (ce qui correspondra à la notion de tribu plus ou moins *fine*). Si l'on ne s'intéresse qu'au pays atteint, on se limitera à des assertions du type : 'La météorite est tombée dans un pays d'Europe du nord', ce qui correspond à une tribu finie (un élément de la tribu est un sous-ensemble de pays, éventuellement vide). A l'autre extrême, on peut envisager l'ensemble des assertions possibles correspondant à une localisation exacte du point d'impact. Comme nous l'avons évoqué plus haut, c'est cette volonté de structurer un espace ayant la puissance du continu qui soulève des difficultés profondes, qui sont abordées dans les sections qui suivent. Dans le contexte des probabilités, la définition d'une mesure de masse totale 1 sur la tribu choisie permettra d'affecter à chacun de ses membres $A$ un nombre quantifiant la probabilité que l'assertion $x \in A$ soit vérifiée. L'événement $X$, de probabilité 1, est certain, et l'événement vide, de probabilité nulle, est impossible (il correspond à la dernière assertion ci-dessus). Les opérations sur les parties - événements sont associées à des opérations logiques : $A \cup B$ correspond à l'événement $x \in A$ *ou* $x \in B$, $A \cap B$ correspond à l'événement $x \in A$ *et* $x \in B$.

## III.2 Tribus, espaces mesurables

### III.2.1 Tribus

**Définition III.2.1.** (Tribu / $\sigma$-algèbre ($\bullet$))
Soit $X$ un ensemble. On appelle tribu (ou $\sigma$-algèbre) sur $X$ un ensemble $\mathcal{A}$ de parties de $X$ qui vérifie les propriétés suivantes :

(i) $\emptyset \in \mathcal{A}$,
(ii) $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$,
(iii) Si $(A_n)$ est une collection dénombrable d'éléments de $\mathcal{A}$, alors

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}.$$

Si la propriété (*iii*) est restreinte aux collections finies, on dira que $\mathcal{A}$ est une *algèbre*.

On appelle $(X, \mathcal{A})$ (ensemble muni de sa tribu) un *espace mesurable*.

**Proposition III.2.2.** ($\bullet$) Toute tribu est stable par intersection dénombrable.

*Démonstration.* Soit $(A_n)$ une collection dénombrable d'éléments d'une tribu $\mathcal{A}$. On a

$$\bigcap_{n \in \mathbb{N}} A_n = \left( \bigcup_{n \in \mathbb{N}} A_n^c \right)^c,$$

qui appartient à $\mathcal{A}$ par complémentarité et union dénombrable. □

**Exemples III.2.1.** Nous donnons ici quelques exemples de tribus associées à un ensemble $X$ quelconque.

1. (Tribu discrète). Pour tout ensemble $X$, l'ensemble $\mathcal{P}(X)$ des parties de $X$ est une tribu. Comme on le verra, dès que $X$ est non dénombrable, par exemple sur $\mathbb{R}$, cette tribu est essentiellement *inutilisable*, car il est impossible de lui associer une mesure non triviale qui possède de bonnes propriétés.
2. (Tribu grossière). Pour tout ensemble $X$, $\{\emptyset, X\}$ est une tribu à 2 membres.
3. Pour tout ensemble $X$ l'ensemble des parties $A$ telles que $A$ ou $A^c$ est dénombrable est une tribu. Si $X$ est fini ou dénombrable, on retrouve la tribu discrète, mais si $X$ est non dénombrable, on obtient une tribu strictement moins fine que la tribu discrète.

**Proposition III.2.3.** (Une intersection de tribus est une tribu ($\bullet$))
Soit $X$ un ensemble. Toute intersection de tribus sur $X$ est une tribu.

*Démonstration.* C'est une conséquence directe de la définition. □

Comme pour toute propriété stable par intersection [5], on peut définir la notion de plus petite tribu contenant une collection de parties de $X$.

**Définition III.2.4.** (Tribu engendrée ($\bullet$))
Soit $\mathcal{C} \subset \mathcal{P}(X)$ une collection de parties de $X$. On appelle tribu engendrée par $\mathcal{C}$, et l'on note $\sigma(\mathcal{C})$, la plus petite tribu contenant $\mathcal{C}$. Elle est définie comme l'intersection de toutes les tribus contenant $\mathcal{C}$.

**Exercice III.2.1.** Soit $X$ un ensemble infini. Décrire la tribu engendrée par la collection des singletons, selon que $X$ soit dénombrable ou pas.

**Définition III.2.5.** (Tribu borélienne sur $\mathbb{R}$($\bullet$))
On appelle tribu borélienne sur $\mathbb{R}$ la tribu engendrée par les ouverts de $\mathbb{R}$. On la note $\mathcal{B}(\mathbb{R})$.

**Proposition III.2.6.** ($\bullet$) La tribu $\mathcal{B}(\mathbb{R})$ des boréliens de $\mathbb{R}$ est engendrée par les intervalles de la forme $]-\infty, a]$, avec $a \in \mathbb{R}$.

*Démonstration.* Notons en premier lieu que le fermé $]-\infty, a]$ est le complémentaire d'un ouvert, tous ces intervalles sont donc dans $\mathcal{B}(\{\mathbb{R}\})$, la tribu engendrée est donc contenue dans la tribu des boréliens. Pour montrer l'inclusion réciproque, tout ouvert de $\mathbb{R}$ étant réunion dénombrables d'intervalles ouverts (voir proposition I.3.12), il suffit de montrer que la tribu engendrée par les $]-\infty, a]$ contient les intervalles ouverts. Par complémentarité, cette tribu contient les intervalles du type $]a, +\infty[$. Par ailleurs, l'union des $]-\infty, b-1/n]$ est l'intervalle $]-\infty, b[$. La tribu contient donc (d'après la stabilité par intersection assurée par la proposition III.2.2), pour tous $a < b$, l'intervalle $]a, +\infty[\, \cap \,]-\infty, b[\, = \,]a, b[$ ce qui termine la démonstration. □

Il sera utile lors de la construction de l'intégrale de considérer des fonctions réelles qui peuvent prendre des valeurs infinies ($+\infty$ ou $-\infty$), c'est-à-dire à valeurs dans la droite réelle

5. On pourra penser par exemple au fait, pour une partie de l'espace $\mathbb{R}^d$, d'être un sous-espace vectoriel, un sous-espace affine, d'être convexe, d'être fermée, d'être conique, ... On parle en général d'*enveloppe* linéaire, affine, convexe, fermée, conique. Le terme d'enveloppe n'est pas utilisé dans le cas des tribus, mais le principe de construction est le même.

achevée $\overline{\mathbb{R}}$ (voir définition I.3.13, page 14). Rappelons que les ouverts de cette droite réelle achevée sont les ensembles de type $U$, $U\cup]a,+\infty]$, $U\cup[-\infty,b[$, ou $U\cup]a,+\infty]\cup[-\infty,b[$, où $U$ est un ouvert de $\mathbb{R}$ (voir proposition I.3.14).

**Proposition III.2.7.** (•) La tribu $\mathcal{B}(\overline{\mathbb{R}})$ des boréliens de $\overline{\mathbb{R}}$ est engendrée par les intervalles de la forme $[-\infty,b]$.

*Démonstration.* Par union dénombrable, $\mathcal{B}(\overline{\mathbb{R}})$ contient les intervalles du type $[-\infty,b[$, donc les intervalles ouverts de $\mathbb{R}$ $]a,b[=[-\infty,b[\setminus[-\infty,a]$. Cette tribu contient également les $]b,+\infty]=]-\infty,b]^c$. On a montré que la tribu engendrée par les $[-\infty,b]$ contient les intervalles du type $[-\infty,a[$, $]a,c[$, et $]c,+\infty]$, elle contient donc tous les ouverts de $\overline{\mathbb{R}}$ par union dénombrable. □

### III.2.2 Applications mesurables

Nous commençons par montrer que l'image réciproque d'une tribu par une application (quelconque) est une tribu sur l'espace de départ.

**Proposition III.2.8.** (Image réciproque d'une tribu(•))
Soit $f$ une application d'un ensemble $X$ vers un ensemble $X'$ muni d'une tribu $\mathcal{A}'$. L'image réciproque de $\mathcal{A}'$ par $f$, c'est-à-dire la famille $\mathcal{A}$ des parties $A$ de $X$ qui s'écrivent

$$A = f^{-1}(A') = \{x\in X\,,\ f(x)\in A'\}\,,$$

avec $A'\in\mathcal{A}'$, est une tribu sur $X$.

*Démonstration.* On a $\emptyset = f^{-1}(\emptyset)$. Par ailleurs, pour tout $A'\in\mathcal{A}'$,

$$f^{-1}(A')^c = \{x\in X\,,\ f(x)\notin A'\} = \{x\in X\,,\ f(x)\in (A')^c\} = f^{-1}((A')^c)\in\mathcal{A} \text{ car } (A')^c\in\mathcal{A}'$$

Enfin, pour toute collection $(A'_n)$ de $\mathcal{A}'$, on a

$$\bigcup_{n\in\mathbb{N}} f^{-1}(A'_n) = f^{-1}\left(\bigcup_{n\in\mathbb{N}} A'_n\right) \text{ avec } \bigcup_{n\in\mathbb{N}} A'_n\in\mathcal{A}',$$

qui appartient bien à $\mathcal{A}$. □

L'image directe d'une tribu par une application n'est en général pas une tribu, comme on peut s'en convaincre en considérant par exemple une application constante vers un ensemble de cardinal $\geq 2$. On peut en revanche pousser en avant une tribu par une application pour obtenir une tribu, alors appelée *tribu image*, comme exprimé par la proposition suivante.

**Proposition III.2.9.** (Tribu image)
Soit $f$ une application de $X$ dans $X'$, et $\mathcal{A}$ une tribu sur $X$. La collection de parties

$$\mathcal{A}' = f_\sharp\mathcal{A} = \left\{A'\subset X'\,,\ f^{-1}(A')\in\mathcal{A}\right\}$$

est une tribu sur $X'$.

*Démonstration.* On a $f^{-1}(\emptyset) = \emptyset \in \mathcal{A}$ et $f^{-1}(X') = X \in \mathcal{A}$. Par ailleurs, si $A' \in \mathcal{A}'$,

$$f^{-1}(A'^c) = \left(f^{-1}(A')\right)^c \in \mathcal{A},$$

d'où $A'^c \in \mathcal{A}'$. Enfin, pour toute famille $(A'_n)$ de $\mathcal{A}'$,

$$f^{-1}\left(\bigcup_{n\in\mathbb{N}} (A'_n)\right) = \bigcup_{n\in\mathbb{N}} f^{-1}(A'_n) \in \mathcal{A},$$

d'où l'on déduit que l'union est dans $\mathcal{A}'$. □

**Exercice III.2.2.** Soit $f : : X \longrightarrow X'$ une application constante et $\mathcal{A}$ une tribu sur $X$. Identifier

$$f_\sharp\mathcal{A} = \left\{A' \subset X',\ f^{-1}(A') \in \mathcal{A}\right\}.$$

Si maintenant $\mathcal{A}'$ est une tribu sur $X'$, identifier $f^{-1}(\mathcal{A}')$.

**Définition III.2.10.** (Application mesurable (•))
Soit $f$ une application d'un espace ensemble $X$ vers un ensemble $X'$. On suppose $X$ et $X'$ munis de tribus $\mathcal{A}$ et $\mathcal{A}'$, respectivement. On dit que $f$ est mesurable de $(X, \mathcal{A})$ vers $(X', \mathcal{A}')$ si l'image réciproque de toute partie de $\mathcal{A}'$ est dans $\mathcal{A}$ :

$$\forall A' \in \mathcal{A}',\ f^{-1}(A') \in \mathcal{A}.$$

On notera que, par définition, une application $f$ de $X$ dans $(X', \mathcal{A}')$ est toujours mesurable, si l'on munit l'espace de départ de la tribu $f^{-1}(\mathcal{A}')$. Cette propriété est d'un intérêt limité du fait que la tribu sur l'ensemble de départ dépend de l'application.

**Exercice III.2.3.** Soit $f$ une application de $X$ dans $X'$.

Si l'on se donne une tribu $\mathcal{A}'$ sur $X$, montrer que $f^{-1}(\mathcal{A})$ est la plus petite tribu sur $X$ telle que $f$ soit mesurable, c'est-à-dire que, si $f$ est $\mathcal{A}-\mathcal{A}'$ mesurable, alors $\mathcal{A}$ contient $f^{-1}(\mathcal{A})$.

Si l'on se donne maintenant une tribu $\mathcal{A}$ sur $X$, montrer que $f_\sharp\mathcal{A}$ est la plus grande tribu sur $X'$ telle que $f$ soit mesurable, c'est-à-dire que, si $f$ est $\mathcal{A} - \mathcal{A}'$ mesurable, alors $\mathcal{A}'$ est contenue dans $f_\sharp(\mathcal{A})$.

**Exercice III.2.4.** Montrer qu'une application constante (qui envoie tous les éléments de l'espace de départ vers un même point de l'espace d'arrivée), est toujours mesurable.

**Exercice III.2.5.** Soit Id l'application identité de $(X, \mathcal{A})$ vers $(X, \mathcal{A}')$. A quelle condition cette application est-elle mesurable ?

**Proposition III.2.11.** Soit $f$ une application de $X$ vers $X'$. On suppose l'espace d'arrivée $X'$ muni d'une tribu $\mathcal{A}'$, engendrée par $\mathcal{C}' \subset \mathcal{P}(X')$. La tribu engendrée par l'image réciproque de $\mathcal{C}'$ est $\mathcal{A} = f^{-1}(\mathcal{A}')$, image réciproque de la tribu $\mathcal{A}'$ engendrée par $\mathcal{C}'$ :

$$\sigma\left(f^{-1}\left(\mathcal{C}'\right)\right) = \mathcal{A} = f^{-1}\left(\sigma(\mathcal{C}')\right) = f^{-1}\left(\mathcal{A}'\right).$$

*Démonstration.* On note $\mathcal{B} = \sigma\left(f^{-1}\left(\mathcal{C}'\right)\right)$. Comme $\mathcal{C}' \subset \mathcal{A}'$, on a $f^{-1}(\mathcal{C}') \subset f^{-1}(\mathcal{A}')$. Comme par ailleurs $f^{-1}\left(\mathcal{A}'\right)$ est une tribu, on a $\mathcal{B} \subset \mathcal{A} = f^{-1}\left(\mathcal{A}'\right)$. Pour montrer l'inclusion réciproque, on introduit

$$\mathcal{B}' = \left\{B' \subset X',\ f^{-1}(B') \in \mathcal{B}\right\} = f_\sharp \mathcal{B}.$$

C'est une tribu d'après la proposition III.2.9. ci-dessus. Comme $f^{-1}(\mathcal{C}') \subset \mathcal{B}$, cette tribu contient $\mathcal{C}'$, et donc aussi la la tribu engendrée par $\mathcal{C}'$, qui est $\mathcal{A}'$. On a finalement

$$\mathcal{A}' \subset \mathcal{B}' \Longrightarrow \underbrace{f^{-1}(\mathcal{A}')}_{\mathcal{A}} \subset f^{-1}(\mathcal{B}') \subset \mathcal{B},$$

la dernière inclusion venant simplement de la définition de $\mathcal{B}'$, qui est composée de parties dont l'image réciproque est dans $\mathcal{B}$. □

**Proposition III.2.12.** (Critère de mesurabilité d'une application (• • •))
Soit $f$ une application de $X$ (muni d'une tribu $\mathcal{A}$) vers $X'$ (muni d'une tribu $\mathcal{A}'$). On suppose que la tribu $\mathcal{A}'$ de l'espace d'arrivée est engendrée par $\mathcal{C}' \subset \mathcal{P}(X')$.

L'application $f$ est mesurable si et seulement si $f^{-1}(C') \in \mathcal{A}$ pour tout $C' \in \mathcal{C}'$.

*Démonstration.* Si $f$ est mesurable, alors $\mathcal{C}' \subset \mathcal{A}' \Longrightarrow f^{-1}(\mathcal{C}') \subset \mathcal{A}$. Réciproquement, si $f^{-1}(\mathcal{C}') \subset \mathcal{A}$, alors la tribu $\mathcal{A}$ contient la tribu engendrée par $f^{-1}(\mathcal{C}')$, qui n'est autre que $f^{-1}(\mathcal{A}')$ d'après la proposition III.2.11. On a donc $f^{-1}(\mathcal{A}') \subset \mathcal{A}$, qui exprime la mesurabilité de $f$. □

**Exercice III.2.6.** Soit $(X, \mathcal{A})$ et $(X', \mathcal{A}')$ deux espaces mesurables. Décrire l'ensemble des applications mesurables de $X$ vers $X'$, dans le cas où $X$ est muni de la tribu grossière $\mathcal{A} = \{\emptyset, X\}$. Même question si $X$ est muni de la tribu discrète $\mathcal{A} = \mathcal{P}(X)$. Que peut-on dire si $X'$ est muni de la tribu grossière ? De la tribu discrète ?

### III.2.3 Classes monotones

**Définition III.2.13.** (Classe monotone (•))
Soit $X$ un ensemble. On appelle classe monotone (ou $\lambda$-système) sur $X$ un ensemble $\mathcal{D}$ de parties de $X$ qui vérifie les propriétés suivantes :

(i) $X \in \mathcal{D}$,
(ii) $A,\ B \in \mathcal{D},\ A \subset B \Longrightarrow B \setminus A \in \mathcal{D}$,
(iii) si $(A_n)$ est une suite croissante d'éléments de $\mathcal{D}$ ($A_n \subset A_{n+1}$ pour tout $n$), alors

$$\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{D}.$$

**Proposition III.2.14.** Toute tribu sur $X$ est une classe monotone.

*Démonstration.* Soit $\mathcal{A}$ une tribu sur $X$. On a par définition $X = \emptyset^c \in \mathcal{A}$. Pour tous $A$, $B$ dans $\mathcal{D}$, avec $A \subset B$, on a $B \setminus A = B \cap A^c \in \mathcal{A}$. Enfin toute réunion dénombrable d'éléments de $\mathcal{A}$ est dans $\mathcal{A}$. □

**Proposition III.2.15.** Toute intersection de classes monotones est une classe monotone.

**Définition III.2.16.** (Classe monotone engendrée par un ensemble de parties (•))
La propriété étant stable par intersection, et l'ensemble $\mathcal{P}(X)$ de toutes les parties étant une classe monotone, on peut définir la notion de classe monotone engendrée par un ensemble $\mathcal{C}$ de parties, définie comme l'intersection de toutes les classes monotones qui contiennent $\mathcal{C}$.

**Définition III.2.17.** ($\pi$-système (•))
On appelle $\pi$–système sur un ensemble $X$ un sous-ensemble $\mathcal{C}$ de parties de $X$ stable par intersection finie :

$$A \in \mathcal{C}\,,\ B \in \mathcal{C} \Longrightarrow A \bigcap B \in \mathcal{C}.$$

**Remarque III.2.18.** Certains auteurs ajoutent la condition qu'un $\pi$–système soit être non vide, et / ou qu'il doit contenir $X$.

**Exemples III.2.2.** Les ensembles de parties suivants sont des $\pi$–systèmes :

1. L'ensemble $\mathcal{P}(X)$ des parties de $X$.
2. L'ensemble des ouverts d'un espace topologique.
3. L'ensemble des fermés d'un espace topologique.
4. La famille $\{]-\infty, c]\,,\ c \in \mathbb{R}\}$ de parties de $\mathbb{R}$.
5. L'ensemble d'intervalles ouverts $\{]a, b[\,,\ -\infty \leq a_i \leq b_i \leq +\infty\}$.
6. Les rectangles ouverts de $\mathbb{R}^2$, de type $]a_1, a_2[\times]b_1, b_2[$, avec $-\infty \leq a_i \leq b_i \leq +\infty$, $i = 1, 2$.
7. Les rectangles fermés.
8. Les rectangles semi-ouverts semi-fermé de $\mathbb{R}^2$, de type $[a_1, a_2[\times[b_1, b_2[$, avec $-\infty \leq a_i \leq b_i \leq +\infty$, $i = 1, 2$.
9. Tout ensemble de singletons : pour $A \subset X$, $\{\{x\}\ ,\ x \in A\}$ auquel on rajoute la partie vide est un $\pi$-système.

**Proposition III.2.19.** (•) Soit $\mathcal{D}$ une classe monotone stable par intersection finie (i.e. $\mathcal{D}$ est aussi un $\pi$ – système). Alors $\mathcal{D}$ est une tribu.

*Démonstration.* On a $\emptyset = X \setminus X \in \mathcal{D}$ et, pour tout $A \in \mathcal{D}$, $A^c = X \setminus A \in \mathcal{D}$. Considérons maintenant une famille $(A_n)$ d'éléments de $\mathcal{D}$. Il s'agit de montrer que l'union des $A_n$ est dans $\mathcal{D}$. On pose $B_0 = A_0$ et, considérant que $B_n$ est construit, et qu'il appartient à $\mathcal{D}$, on définit $B_{n+1}$ comme $B_n \cup A_{n+1} = A_0 \cup A_1 \cup \ldots \cup A_{n+1}$. Montrons par récurrence que $B_{n+1} \in \mathcal{D}$. Supposons $B_n$ dans $\mathcal{D}$. On a

$$B_{n+1} = B_n \cup A_{n+1} = \left(\underbrace{B_n^c \cap A_{n+1}^c}_{\in \mathcal{D}}\right)^c$$

qui est dans $\mathcal{D}$ comme complémentaire d'un élément de $\mathcal{D}$.

La suite $(B_n)$ d'éléments de $\mathcal{D}$, est croissante par construction, d'où $\cup B_n \in \mathcal{D}$, et cette union s'identifie par construction à l'union des $A_n$. La famille $\mathcal{D}$ est donc bien une tribu. $\square$

**Proposition III.2.20.** (Lemme de classe monotone)
(••) Soit $\mathcal{C}$ un $\pi$ – système sur l'ensemble $X$. La classe monotone $\mathcal{D}$ engendrée par $\mathcal{C}$ est égale à la tribu $\mathcal{A} = \sigma(\mathcal{C})$ engendrée par $\mathcal{C}$.

*Démonstration.* La tribu $\mathcal{A}$ contient $\mathcal{C}$, et c'est une classe monotone (proposition III.2.14), elle contient donc $\mathcal{D}$ qui est la plus petite classe monotone contenant $\mathcal{C}$. Pour montrer l'inclusion inverse, nous allons montrer que $\mathcal{D}$ est une tribu (qui alors contient nécessairement $\mathcal{A}$, qui est la plus petite). D'après la proposition III.2.19, il suffit de montrer que $\mathcal{D}$ est un $\pi$ – système, i.e. qu'elle est stable par intersection (l'appartenance de $X$ à $\mathcal{D}$ est acquise, par définition d'une classe monotone). On considère dans un premier temps

$$\mathcal{D}' = \{A \in \mathcal{D}\,,\ A \cap C \in \mathcal{D} \quad \forall C \in \mathcal{C}\}\,.$$

Cet ensemble de parties contient le $\pi$ – système $\mathcal{C}$, et en particulier $X$. Par ailleurs, pour tous $A$, $B \in \mathcal{D}'$, $A \subset B$, on a

$$(B \setminus A) \cap C = (\underbrace{B \cap C}_{\in \mathcal{D}}) \setminus (\underbrace{A \cap C}_{\in \mathcal{D}}) \in \mathcal{D}.$$

Montrons que $\mathcal{D}'$ est également stable par union croissante. Pour toute suite croissante $(A_n)$ dans $\mathcal{D}'$, on a

$$\left(\bigcup A_n\right) \cap C = \bigcup \left(\underbrace{A_n \cap C}_{\in \mathcal{D}}\right) \in \mathcal{D}.$$

L'ensemble $\mathcal{D}'$ est donc une classe monotone, qui contient $\mathcal{C}$, et qui est contenue dans la classe monotone $\mathcal{D}$ engendrée par $\mathcal{C}$, elle s'identifie donc à $\mathcal{D}$. On a ainsi montré que, pour tout $A \in \mathcal{D}$, $C \in \mathcal{C}$, $A \cap C \in \mathcal{D}$. Il reste à montrer la stabilité par intersection finie, et pas seulement avec les éléments de $\mathcal{C}$. Pour cela on introduit

$$\mathcal{D}'' = \left\{A \in \mathcal{D}\,,\ A \cap A' \in \mathcal{D} \quad \forall A' \in \mathcal{D}\right\}.$$

Cet ensemble contient $\mathcal{C}$ comme on vient de le montrer, et c'est une classe monotone (la démonstration est la même que précédemment) contenue dans $\mathcal{D}$, elle s'identifie donc à $\mathcal{D}$. On a ainsi montré que $\mathcal{D}$ est un $\pi$ – système. Comme c'est aussi une classe monotone, c'est une tribu d'après la proposition III.2.14, qui contient $\mathcal{C}$, elle contient donc $\mathcal{A} = \sigma(\mathcal{C})$, ce qui termine la preuve. □

## III.3 Mesures

**Définition III.3.1.** (•) Soit $X$ un ensemble, et $\mathcal{A}$ une tribu sur $X$. On appelle mesure une application de $\mathcal{A}$ dans $[0, +\infty]$ telle que

(i) $\mu(\emptyset) = 0$,

(ii) si $(A_n)$ est une collection dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints, alors

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n=0}^{+\infty} \mu\left(A_n\right).$$

Le triplet $(X, \mathcal{A}, \mu)$ (ensemble $X$ muni d'une tribu $\mathcal{A}$ et d'une mesure associée) est appelé *espace mesuré.*
On dit que la mesure est *finie* si $\mu(X) < +\infty$.
On dit que la mesure est $\sigma-$ *finie* si $X$ est réunion dénombrable d'ensembles de mesure finie.

**Exemples III.3.1.** (•) Nous donnons ici quelques exemples de mesures associées à un ensemble $X$ quelconque.

1. (Mesure de comptage). Soit $X$ un ensemble et $\mathcal{A}$ une tribu sur $X$. On définit $\mu(A)$ comme le cardinal de $A$.
2. (Masse ponctuelle) Soit $X$ un ensemble, $\mathcal{A}$ une tribu sur $X$, et $x \in X$. L'application $\delta_x$ qui à $A \in \mathcal{A}$ associe 1 si $x \in A$, 0 sinon, est une mesure appelée masse ponctuelle en $x$.
3. Toute combinaison positive de masses ponctuelles
$$\mu = \sum_{i=1}^{n} \alpha_i \delta_{x_i}\,,\ \ \alpha_i \geq 0 \quad \forall i = 1, \ldots, n$$
est également une mesure.
4. (Mesure grossière) Soit $X$ un ensemble de cardinal $\geq 2$, et $\mathcal{A}$ une tribu sur $X$. On pose $\mu(\emptyset) = 0$ et $\mu(A) = +\infty$ dès que $A \neq \emptyset$. Il s'agit d'une manière quelque peu grossière (d'où le nom), binaire, d'appréhender le monde, en distinguant deux types d'ensembles (disons des zones de l'espace physique pour fixer les idées) : l'ensemble vide, de volume nul, et tout ensemble non vide, auquel on attribuerait par convention une quantité infinie. Cette mesure extérieure ne distingue en quelque sorte que le “rien” et le ”quelque chose”. On peut associer à cette mesure extérieure une addition [6] primitive basée sur :

rien + rien = rien, rien + quelque chose = quelque chose, et
quelque chose + quelque chose = quelque chose.

**Proposition III.3.2.** Soit $\mu$ une mesure $\sigma-$finie. Il existe une partition $(D_n)$ de $X$ consituée d'ensembles mesurables et de mesure finies.

*Démonstration.* Par définition $X$ s'écrit comme réunion d'ensembles $A_n$ de mesures finies. On prend $D_0 = A_0$, et l'on définit $D_n$ comme $A_n \setminus (A_{n-1} \cup \ldots \cup A_0)$. On note $I$ l'ensemble des indices pour lesquels cet ensemble est non vide [7] Par construction, $D_n$ est une partition de $X$, et $\mu(D_n) \leq \mu(A_n) - \mu(D_{n-1} \cap A_n) \leq \mu(A_n) < +\infty$. □

**Proposition III.3.3.** (Monotonie (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, $A$ et $B$ dans $\mathcal{A}$. La mesure $\mu$ est *monotone*, c'est-à-dire que si $A \in \mathcal{A}$ est inclus dans $B \in \mathcal{A}$, alors la mesure de $A$ est plus petite que celle de $B$ :
$$\forall A\,,\ B \in \mathcal{A}\,,\ A \subset B \Longrightarrow \mu(A) \leq \mu(B).$$

6. Il s'agit bien d'une loi de composition interne sur l'ensemble à deux éléments {'rien', 'quelque chose'}, que l'on peut obtenir en quotientant $X$ par la relation d'équivalence basée sur la notion d'être vide ou pas. Pour les lecteurs sensibles au plaisir de désigner des choses simples par des termes abscons, rajoutons que cette loi est associative, commutative, possède un élément neutre ('rien'), et munit donc notre petit univers à deux éléments d'une structure de *magma associatif unifère abélien* (sic).
7. Par définition, une partition ne doit contenir que des parties non vides.

Si la mesure de $A$ est finie, on a de plus $\mu(B \setminus A) = \mu(B) - \mu(A)$.

*Démonstration.* On écrit simplement $B$ comme réunion disjointe $B = A \cup (B \setminus A)$, qui implique, d'après la définition,

$$\mu(B) = \mu(A) + \mu(B \setminus A),$$

d'où les propriétés annoncées. □

**Proposition III.3.4.** (Sous-additivité (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)$ une suite dans $\mathcal{A}$. On a

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) \leq \sum_{n=0}^{+\infty} \mu(A_n).$$

*Démonstration.* La démonstration est basée sur la construction d'une suite $(B_k)$ d'éléments disjoints de $\mathcal{A}$, telle que l'union des $n$ premiers termes s'identifie à l'union des $n$ premiers $A_k$. On pose $B_0 = A_0$ et l'on construit par récurrence

$$B_n = A_n \setminus \left(\bigcup_{k=0}^{n-1} A_k\right) = A_n \cap A_1^c \cap \ldots A_{n-1}^c,$$

qui appartient bien à la tribu $\mathcal{A}$, et qui est tel que $\mu(B_n) \leq \mu(A_n)$. On a donc

$$\mu\left(\bigcup_{n=0}^{+\infty} A_n\right) = \mu\left(\bigcup_{n=0}^{+\infty} B_n\right) = \sum_{n=0}^{+\infty} \mu(B_n) \leq \sum_{n=0}^{+\infty} \mu(A_n),$$

qui est l'inégalité annoncée. □

**Proposition III.3.5.** Soit $(X, \mathcal{A}, \mu)$ un espace mesuré, et $(A_n)$ une suite de parties de $\mathcal{A}$, croissante pour l'inclusion. On a alors

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mu(A_n) \in [0, +\infty].$$

*Démonstration.* On construit une suite auxiliaire de parties $B_k$ en posant $B_0 = A_0$, et par récurrence $B_n = A_n \setminus A_{n-1}$. Les $B_n$ sont disjoints par construction, et $A_n$ est l'union des $B_j$ pour $j \leq n$. On a donc

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \sum_{j=0}^{+\infty} \mu(B_j) = \lim_{n\to+\infty} \sum_{j=0}^{n} \mu(B_j)$$

$$= \lim_{n\to+\infty} \mu\left(\bigcup_{n=0}^{n} B_j\right) = \lim_{n\to+\infty} \mu(A_n).$$

qui termine la preuve. □

**Définition III.3.6.** (Ensemble négligeable, mesure complète (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit que l'ensemble $N \subset X$ est *négligeable* s'il est inclus dans une partie $A \in \mathcal{A}$ de mesure nulle. On dit qu'une propriété est vérifiée *$\mu$-presque partout*, ou qu'elle est vérifiée pour $\mu$-presque tout $x$, si elle est vérifiée en dehors d'un ensemble négligeable. On dit que la mesure $\mu$ est *complète* si tous les ensembles négligeables sont dans $\mathcal{A}$.

**Définition III.3.7.** (Ensembles de mesure pleine (•))
Soit $(X, \mathcal{A}, \mu)$ un espace mesuré. On dit que $B \in \mathcal{A}$ est de mesure pleine si, pour tout $A \in \mathcal{A}$, $\mu(A \cap B) = \mu(A)$.

**Définition III.3.8.** (Absolue continuité d'une mesure par rapport à une autre (••))
Soit $(X, \mathcal{A})$ un espace mesurable, $\lambda$ et $\mu$ deux mesures sur $\mathcal{A}$. On dit que $\mu$ est *absolument continue* par rapport à $\lambda$, et l'on écrit $\mu \ll \lambda$, si pour tout $A \in \mathcal{A}$ tel que $\lambda(A) = 0$, on a aussi $\mu(A) = 0$.

**Remarque III.3.9.** Cette propriété qui caractérise d'une certaine manière le positionnement relatif de deux mesures joue un rôle essentiel dans les applications. La situation typique est la suivante : on a une mesure définie sur un espace mesurable, notée $\lambda$ (il s'agira en général de la mesure de Lebesgue sur $\mathbb{R}$ ou $\mathbb{R}^d$ qui va être définie dans la section suivante) qui, pour reprendre l'esprit des remarques introductives à cette partie, *tapisse* en quelque sorte l'espace sous-jacent, en permettant d'affecter à une zone son volume (que l'on peut concevoir comme une capacité à accueilir de la matière). Cette mesure sera en général *statique*, au sens où elle est définie une fois pour toutes. On fera vivre sur ce même espace d'autres mesures, que nous appellerons $\mu$, qui représentent typiquement une distribution de matière, susceptible d'évoluer au cours du temps. Dire que l'on a absolue continuité de $\mu$ par rapport à $\lambda$ signifie que l'on n'a pas de concentration : une zone de volume nul ne peut contenir qu'une masse elle-même nulle. Dans cette situation, le théorème de Radon Nykodim (qui dépasse le cadre de ce cours sous sa forme présente) assurera que l'on peut représenter la mesure $\mu$ par une densité adossée à la mesure $\lambda$, ce qui permet de représenter une distribution de matière comme une *fonction*.

La proposition suivante donne un critère très utile en pratique d'égalité entre deux mesures, elle jouera un rôle essentiel dans la construction des mesures-produits.

**Proposition III.3.10.** (••) Soit $(X, \mathcal{A})$ un espace mesurable, et $\mathcal{C}$ un $\pi$ − système sur $X$ (définition III.2.17), qui engendre $\mathcal{A}$. Soient $\mu$ et $\nu$ deux mesures $\sigma$ − finies sur $\mathcal{A}$, telles que $\mu(X) = \nu(X)$ et $\mu(C) = \nu(C)$ pour tout $C$ dans $\mathcal{C}$. Alors $\mu = \nu$.

*Démonstration.* On considère l'ensemble $\mathcal{D}$ de parties défini par

$$\mathcal{D} = \{A \in \mathcal{A}\,,\ \mu(A) = \nu(A)\}\ .$$

On souhaite montrer que $\mathcal{D} = \mathcal{A}$. Il suffit pour cela de montrer que $\mathcal{D}$ est une classe monotone qui contient $\mathcal{C}$. En effet, cela impliquera que $\mathcal{D}$ contient la classe monotone engendrée par $\mathcal{C}$ qui, d'après la proposition III.2.20, est la tribu engendrée par $\mathcal{C}$.
On a par hypothèse $X \in \mathcal{D}$. Pour tous $A$, $B$ dans $\mathcal{D}$, avec $A \subset B$, on a

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)\,.$$

Pour toute suite croissante dans $\mathcal{D}$, on a, d'après la proposition III.3.5 ;

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \lim_{n\to+\infty} \mu\left(A_n\right) = \lim_{n\to+\infty} \nu\left(A_n\right) = \nu\left(\bigcup_{n\in\mathbb{N}} A_n\right).$$

Nous avons donc montré que $\mathcal{D}$ est une classe monotone. Comme évoqué ci-dessus, contenant $\mathcal{C}$, elle contient la classe monotone engendrée par $\mathcal{C}$, qui est la tribu engendrée par $\mathcal{C}$, c'est-à-dire $\mathcal{A}$. On a donc $\mu(A) = \nu(A)$ pour tout $A \in \mathcal{A}$, les mesures sont donc les mêmes. □

**Corollaire III.3.11.** On se place dans les hypothèses de la proposition précédente, en ne supposant plus les mesures finies. On suppose en revanche qu'il existe une suite croissante $(C_n)$ d'éléments de $\mathcal{C}$, dont l'union est égale à $X$, et telle que $\mu(C_n) < +\infty$ et $\nu(C_n) < +\infty$ pour tout $n$. Alors $\mu = \nu$.

*Démonstration.* Pour tout $n$ on définit

$$\mu_n(A) = \mu(A \cap C_n)\,,\ \nu_n(A) = \nu(A \cap C_n),$$

D'après la proposition III.3.10, on a $\mu_n = \nu_n$ pour tout $n$, d'où, pour tout $A \in \mathcal{A}$,

$$\mu(A) = \lim_n \mu_n(A) = \lim_n \nu_n(A) = \nu(A),$$

qui exprime l'identité des mesures $\mu$ et $\nu$. □

## III.4 Mesures extérieures

Cette section présente la notion de mesure extérieure, qui correspond à une mesure à la laquelle on aurait ôté la condition de $\sigma$–additivité, remplacée par une notion affaiblie de $\sigma$–*sous*-additivité. Cette notion constituera une étape essentielle dans la construction de la mesure de Lebesgue. Comme on le verra, il est assez facile de construire explicitement une telle mesure extérieure définie sur l'ensemble des parties de $\mathbb{R}$ ou $\mathbb{R}^d$, et c'est en dégrossissant cette mesure extérieure que nous aboutirons à la mesure de Lebesgue.

### III.4.1 Définitions, premières propriétés

**Définition III.4.1.** (•) Soit $X$ un ensemble. On appelle *mesure extérieure* sur $X$ une application $\mu^\star$ de $\mathcal{P}(X)$ dans $[0, +\infty]$ telle que

(i) $\mu^\star(\emptyset) = 0$,
(ii) si $A \subset B$ alors $\mu^\star(A) \leq \mu^\star(B)$,
(iii) si $(A_n)$ est une collection de parties de $X$, alors

$$\mu^\star\left(\bigcup_{n\in\mathbb{N}} A_n\right) \leq \sum_{n=0}^{+\infty} \mu^\star\left(A_n\right).$$

**Exemples III.4.1.** Nous donnons ici quelques exemples de mesures extérieures associées à un ensemble $X$ quelconque. Précisons en premier lieu que toute mesure définie sur la tribu constituée de toutes les parties d'un ensemble est une mesure extérieure, du fait que la monotonie est une propriété des mesures (voir proposition III.3.3), ainsi que la sous-additivité (voir proposition III.3.4). La mesure de comptage, ou toute masse ponctuelle (voir exemples III.3.1), si on les définit sur la tribu discrète $\mathcal{P}(X)$, sont donc des mesures extérieures.
Autres exemples :

1. Soit $X$ un ensemble. On pose $\mu^\star(\emptyset) = 0$ et $\mu^\star(A) = 1$ pour tout $A \in \mathcal{P}(X)$, $A \neq \emptyset$ (voir exercice III.4.1).
2. Soit $X$ un ensemble. Pour tout $A \in \mathcal{P}(X)$, on pose $\mu^\star(A) = 0$ si $A$ est dénombrable, et $\mu^\star(A) = 1$ dès que $A$ est non dénombrable.

Une mesure extérieure n'est pas en général une mesure sur $\mathcal{P}(X)$. Nous verrons néanmoins que sa restriction à une sous-partie de $\mathcal{P}(X)$ est bien une mesure. Cette sous-partie est constituée des ensembles appelés $\mu^\star-$ mesurables. Il s'agit d'ensemble qui, avec leur complémentaire, constituent une partition de l'espace complet vis-à-vis de laquelle la mesure extérieure est *additive*, au sens précisé ci-dessous.

**Définition III.4.2.** (Ensembles mesurables pour une mesure extérieure (•))
Soit $X$ un ensemble, et $\mu^\star$ une mesure extérieure sur $X$. On dit que $B$ est $\mu^\star-$mesurable si, pour tout $A \subset X$, on a

$$\mu^\star(A) = \mu^\star(A \cap B) + \mu^\star(A \cap B^c). \tag{III.4.1}$$

Du fait de la sous-additivité de $\mu^\star$, l'inégalité

$$\mu^\star(A) \geq \mu^\star(A \cap B) + \mu^\star(A \cap B^c)$$

pour tout $A$ suffit pour assurer la $\mu^\star-$ mesurablité de $B$.

**Exercice III.4.1.** Soit $X$ un ensemble non vide et $\mu^\star$ l'application de $\mathcal{P}(X)$ dans $\mathbb{R}_+$ définie par $\mu^\star(\emptyset) = 0$ et $\mu^\star(A) = 1$ pour tout $A \in \mathcal{P}(X)$, différent de $\emptyset$. Montrer que $\mu^\star$ est une mesure extérieure. À quelles conditions $\mu^\star$ est-elle une mesure ? Quels sont les parties de $X$ mesurables pour $\mu^\star$ ?

Nous verrons que, si l'on s'en tient à la définition, il se peut qu'une mesure extérieure admette très peu d'ensembles mesurables (voir exercice III.4.1 ci-après). La proposition suivante établit que, dans tous les cas, les ensembles *très petits* (de mesure extérieure nulle), ou *très gros* (i.e. de complémentaire très petit), sont mesurables au sens de la définition précédente.

**Proposition III.4.3.** (•) Soit $X$ un ensemble et $\mu^\star$ une mesure extérieure sur $X$ (définition III.4.2). Tout $B \subset X$ tel que $\mu^\star(B) = 0$ ou $\mu^\star(B^c) = 0$ est $\mu^\star-$ mesurable.

*Démonstration.* Soit $B$ une telle partie. Comme indiqué dans la définition, il suffit de vérifier que, pour tout $A \subset X$, $\mu^\star(A) \geq \mu^\star(A \cap B) + \mu^\star(A \cap B^c)$. Or, si $\mu^\star(B) = 0$, alors $\mu^\star(A \cap B) \leq \mu^\star(B) = 0$ par monotonie, et $\mu^\star(A \cap B^c) \leq \mu^\star(A)$ par monotonie également. Le cas $\mu^\star(B^c) = 0$ se traite de la même manière. ☐

### III.4.2 D'une mesure extérieure à une mesure

Le théorème suivant constitue un outil très général pour construire des couples tribus - mesures à partir de la donnée d'une mesure extérieure. Nous l'utiliserons pour construire la mesure de Lebesgue à partir de sa version extérieure (définie par la proposition III.4.6 ci-après).

**Théorème III.4.4.** (●●) Soit $X$ un ensemble et $\mu^\star$ une mesure extérieure sur $X$ (définition III.4.2). On note $\mathcal{A}_{\mu^\star}$ la collection des parties $\mu^\star-$ mesurables (selon la définition III.4.2).

Alors $\mathcal{A}_{\mu^\star}$ est une tribu, et la restriction de $\mu^\star$ à $\mathcal{A}_{\mu^\star}$ est une mesure.

*Démonstration.* (●●●) Montrons dans un premier temps que $\mathcal{A}_{\mu^\star}$ est une algèbre, c'est-à-dire qu'elle vérifie les conditions de la définition III.2.13, avec condition $(iii)$ limitée aux collections finies. La proposition III.4.3 assure que $\emptyset$ et $X$ sont dans $\mathcal{A}_{\mu^\star}$. Par ailleurs l'identité (III.4.1) est inchangée si l'on échange les rôles de $B$ et $B^c$, $\mathcal{A}_{\mu^\star}$ est donc stable par complémentarité.

Montrons maintenant la stabilité par union. Soient $B_1$, $B_2$ deux parties de $\mathcal{A}_{\mu^\star}$ (la démarche ci-dessous est illustrée par la figure III.4.1). Comme $B_1$ est $\mu^\star$–mesurable, on a, pour tout $A \subset X$,

$$\begin{aligned}\mu^\star(A\cap(B_1\cup B_2)) &= \mu^\star(A\cap(B_1\cup B_2)\cap B_1)+\mu^\star(A\cap(B_1\cup B_2)\cap B_1^c)\\ &= \mu^\star(A\cap B_1)+\mu^\star(A\cap B_1^c\cap B_2).\end{aligned}$$

On a donc, en utilisant $(B_1\cup B_2)^c = B_1^c\cap B_2^c$ et la $\mu^\star-$ mesurabilité de $B_2$,

$$\begin{aligned} & \mu^\star(A\cap(B_1\cup B_2))+\mu^\star(A\cap(B_1\cup B_2)^c)\\ = & \mu^\star(A\cap B_1)+\mu^\star((A\cap B_1^c)\cap B_2)+\mu^\star((A\cap B_1^c)\cap B_2^c)\\ = & \mu^\star(A\cap B_1)+\mu^\star(A\cap B_1^c)\end{aligned}$$

qui est égal à $\mu^\star(A)$ du fait que $B_1\in\mathcal{A}_{\mu^\star}$. L'ensemble $\mathcal{A}_{\mu^\star}$ est donc stable par unions finies, il s'agit bien d'une algèbre.

Pour montrer que c'est une tribu, il reste à montrer qu'elle est stable par union dénombrable. Considérons une suite $(B_k)$ d'éléments de $\mathcal{A}_{\mu^\star}$, supposés deux à deux *disjoints.* Nous allons montrer par récurrence que, pour tout $n$, la partition finie (recouvrement de $X$ par $n+1$ parties disjointes)

$$B_1\,,\ B_2\,,\ \ldots\,,\ B_n\,,\ \left(\bigcup_{k=1}^n B_k\right)^c$$

respecte $\mu^\star$, au sens où, pour tout $A$

$$\mu^\star(A)=\sum_{k=1}^n\mu^\star(A\cap B_k)+\mu^\star\left(A\cap\left(\bigcup_{k=1}^n B_k\right)^c\right).$$

Nous allons en fait démontrer par récurrence l'identité équivalente

$$\mu^\star(A)=\sum_{k=1}^n\mu^\star(A\cap B_k)+\mu^\star\left(A\cap\left(\bigcap_{k=1}^n B_k^c\right)\right). \tag{III.4.2}$$

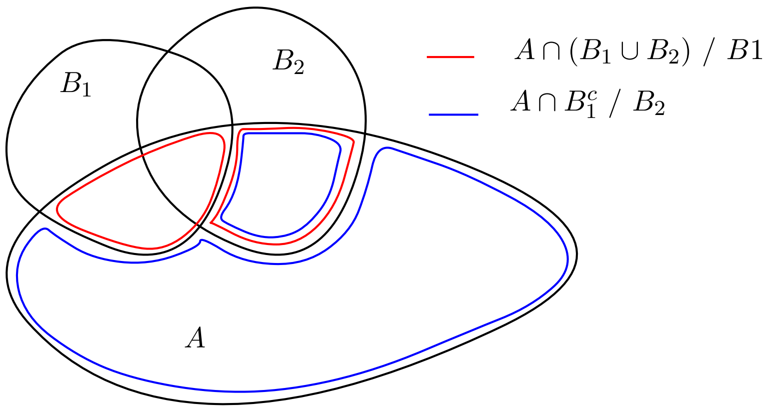


FIGURE III.4.1 – Mesurabilité de $B_1 \cup B_2$. La notation $C$ / $D$ de la légende indique que l'on écrit $C$ comme l'union de $C \cap D$ et de $C \cap D^c$, où $D$ est un ensemble mesurable.

L'identité pour $n = 1$ exprime simplement la $\mu^\star-$ mesurablité de $B_1$. Supposons maintenant qu'elle est vraie jusqu'au rang $n$. L'appartenance de $B_{n+1}$ à $\mathcal{A}_{\mu^\star}$ implique

$$\begin{aligned}
&\mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right)\right) \\
&= \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right) \cap B_{n+1}\right) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n} B_k^c\right) \cap B_{n+1}^c\right) \\
&= \mu^\star\left(A \cap B_{n+1}\right) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{n+1} B_k^c\right)\right)
\end{aligned}$$

qui établit (III.4.2). On remarque maintenant que, par monotonie, le second terme de (III.4.2) n'augmente pas si l'on remplace l'intersection finie par l'intersection de tous les $B_k^c$. On a donc, en faisant tendre $n$ vers $+\infty$,

$$\mu^\star(A) \geq \sum_{k=1}^{+\infty} \mu^\star(A \cap B_i) + \mu^\star\left(A \cap \left(\bigcap_{k=1}^{+\infty} B_k^c\right)\right), \tag{III.4.3}$$

d'où, par $\sigma$-sous-additivité de $\mu^\star$,

$$\mu^\star(A) \geq \mu^\star\left(A \cap \left(\bigcup_{k=1}^{+\infty} B_k\right)\right) + \mu^\star\left(A \cap \left(\bigcup_{k=1}^{+\infty} B_k\right)^c\right),$$

qui est supérieur ou égal à $\mu^\star(A)$ par sous-additivité. On a donc identité entre $\mu^\star(A)$ et l'expression ci-dessus, pour tout $A$, ce qui prouve que l'union des $B_k$ est dans $\mathcal{A}_{\mu^\star}$. On en déduit immédiatement la stabilité par union dénombrable générale (sans le caractère disjoint deux à deux), en notant que l'union des $B_k$ dans $\mathcal{A}_{\mu^\star}$ peut s'écrire comme union d'ensembles disjoints

$$B_1\,,\ B_2 \cap B_1^c,\ B_3 \cap B_2^c \cap B_1^c\,,\ \ldots\,,\ B_n \cap B_{n-1}^c \cap B_{n-2}^c \cap \ldots \cap B_1^c,\ \ldots$$

Nous avons ainsi démontré que $\mathcal{A}_{\mu^\star}$ est une tribu.

Il reste à vérifier que la restriction de $\mu^\star$ à $\mathcal{A}_{\mu^\star}$ est bien une mesure. Il suffit pour cela de vérifier l'additivité dénombrable. Considérons une suite $(B_k)$ d'éléments de $\mathcal{A}_{\mu^\star}$ disjoints deux à deux. On écrit simplement l'inégalité (III.4.3) en prenant pour $A$ l'union des $B_k$. Il vient

$$\mu^\star(A) \geq \sum_{k=1}^{+\infty} \mu^\star(B_k) + 0$$

qui est supérieur à $\mu^\star(A)$ par sous-additivité, d'où l'identité entre les deux expressions. $\square$

**Remarque III.4.5.** Noter que ce théorème peut aboutir dans certains cas à un résultat très "pauvre" pour certaines mesures extérieures. Comme l'illustre l'exercice III.4.1, si la mesure extérieure ne présente pas de bonnes propriétés d'additivité, les seuls ensembles mesurables sont $X$ et $\emptyset$.

### III.4.3 Mesure de Lebesgue

La première étape dans la construction de la mesure de Lebesgue est la définition d'une mesure *extérieure* de Lebesgue, selon le principe décrit dans l'introduction : on sait la valeur que l'on veut à la longueur d'un intervalle, la longueur totale d'une réunion d'intervalles (avec possibles recouvrements) est supérieure à la sommes des longueurs. On considère donc que la mesure d'un ensemble, telle que l'on souhaite la définir, est inférieure à la somme des longueurs des intervalles, pour tout recouvrement de l'ensemble. Ces considérations conduisent à la définition de ce que l'on appelle la mesure extérieure de Lebesgue, donnée par la proposition qui suit (la proposition établit que l'objet défini est bien une mesure extérieure au sens de la définition III.4.2).

**Proposition III.4.6.** (Mesure de Lebesgue extérieure sur $\mathbb{R}$ (••))
Pour tout $A \subset \mathbb{R}$, on note $C_A$ l'ensemble des suites d'intervalles ouverts dont l'union recouvre $A$ :

$$C_A = \left\{ (]a_i, b_i[)_{i\in\mathbb{N}} \ , \ A \subset \bigcup_{\mathbb{N}} ]a_i, b_i[ \right\}.$$

On autorise les intervalles à être vides (i.e. $a_i$ peut être égal à $b_i$), ce qui revient à autoriser les collections finies. On définit alors $\lambda^\star \ : \ \mathcal{P}(X) \longrightarrow [0, +\infty]$ par

$$\lambda^\star(A) = \inf_{C_A} \left( \sum_i (b_i - a_i) \right). \qquad \text{(III.4.4)}$$

Cette application est une mesure extérieure, appelée *mesure extérieure de Lebesgue*, et elle attribue à tout intervalle sa longueur.

*Démonstration.* (• • •) Montrons que $\lambda^\star$ vérifie les trois conditions de la définition III.4.2.

$(i)$ En premier lieu, l'ensemble vide est recouvert par une réunion d'intervalles vides. On a donc $\lambda^\star(\emptyset) = 0$.

$(ii)$ Ensuite, si $A \subset B$, alors toute suite d'intervalles qui recouvre $B$ recouvre aussi $A$, l'infimum de (III.4.4) qui définit $\lambda^\star(A)$, porte donc sur un ensemble plus grand que celui associé à $B$, d'où $\lambda^\star(A) \leq \lambda^\star(B)$.

*(iii)* Il s'agit maintenant de démontrer la $\sigma$–sous-additivité. On considère une suite $(A_n)$ de parties de $X$. Il s'agit de montrer que la mesure extérieure de l'union est inférieure à la somme des mesures. Remarquons tout d'abord que si la somme des mesures est infinie, alors l'inégalité est immédiatement vraie. On peut donc supposer que toutes les mesures sont finies. Pour tout $\varepsilon > 0$, il existe une collection $(]a_i^n, b_i^n[)$ qui réalise (III.4.4) à $\varepsilon/2^n$ près, i.e.

$$\sum (b_i^n - a_i^n) \leq \lambda^\star(A_n) + \frac{\varepsilon}{2^n}.$$

L'union de toutes ces collections d'intervalles est elle-même une collection dénombrable d'intervalles ouverts, dont la longueur totale majore par définition la mesure de l'union des $A_n$. On a donc

$$\lambda^\star\left(\bigcup A_n\right) \leq \sum_{n=0}^{+\infty} \left(\lambda^\star(A_n) + \frac{\varepsilon}{2^n}\right) = \sum_{n=0}^{+\infty} \lambda^\star(A_n) + 2\varepsilon,$$

pour tout $\varepsilon$, d'où la $\sigma$–sous-additivité.

Il reste à montrer que $\lambda^\star$ affecte aux intervalles (ouverts, fermés, ou mixtes) leur longueur. Considérons l'intervalle fermé $[a,b]$. Pour tout $\varepsilon$ on peut recouvrir cet intervalle par $]a-\varepsilon, b+\varepsilon[$ et des intervalles vides, la quantité $\lambda^\star(A)$ est donc majorée par une quantité arbitrairement proche de $b-a$, on a donc $\lambda^\star(]a,b[) \leq b-a$. Montrons l'inégalité inverse. On considère pour cela un recouvrement de $[a,b]$ par des intervalles ouverts. Comme $[a,b]$ est compact, on peut en extraire un recouvrement fini, que l'on note $(]a_i, b_i[)_{1\leq i\leq n}$ (on suppose que l'on ne garde dans ce recouvrement que des intervalles utiles, i.e. qui rencontrent $]a,b[$). Si l'un des intervalles recouvre $]a,b[$, la longueur totale est supérieure ou égale à $b-a$. Dans le cas contraire, il existe nécessairement $i_1$ tel que $a_{i_1} < a$. On a $a < b_{i_1} < b$ et il existe nécessairement $i_2$ tel que $a_{i_2} < b_{i_1}$ (sinon $b_{i_1} \in ]a,b[$ ne serait pas couvert). On construit ainsi une suite $]a_{i_k}, b_{i_k}[$, avec $a_{i_k} < b_{i_{k+1}}$. Comme la collection finie recouvre $]a,b[$, on finit par arriver à $b_{i_k} > b$, on arrête alors la construction et l'on note $n$ le rang atteint. On a alors

$$\begin{aligned} b-a &\leq b_{i_n} - a_{i_1} = b_{i_n} - a_{i_n} + \underbrace{a_{i_n}}_{\leq b_{i_{n-1}}} - \cdots - a_{i_2} + \underbrace{a_{i_2}}_{\leq b_{i_1}} - a_{i_1} \\ &\leq b_{i_n} - a_{i_n} + \cdots + b_{i_2} - a_{i_2} + b_{i_1} - a_{i_1}, \end{aligned}$$

qui est inférieur ou égal à la longueur totale de la collection d'intervalles initiale (car on en a enlevé certains). On a donc $b-a \leq \lambda^\star(A)$. Pour finir si l'on considère l'intervalle $]a,b[$, on peut encadrer (pour l'inclusion) cet intervalle par des intervalles fermés $[a+\varepsilon, b-\varepsilon]$ et $[a-\varepsilon, b+\varepsilon]$, dont la mesure tend vers $b-a$, on a donc également $\lambda^\star(]a,b[) = b-a$. Le raisonnement est analogue pour les intervalles de type $[a,b[$ et $]a,b]$. □

**Définition III.4.7.** (Mesure de Lebesgue extérieure sur $\mathbb{R}^d$ (• • •))
On définit de la même manière une mesure extérieure sur $\mathbb{R}^d$ en remplaçant les collections d'intervalles par des collections de pavés

$$I_1 \times I_2 \times \dots I_d = \left\{(x_1, \dots, x_d) \in \mathbb{R}^d\,,\ x_k \in I_k \text{ pour } \ k = 1, \dots, d\right\}$$

où les $I_k$ sont des intervalles ouverts de $\mathbb{R}$. Le volume du pavé est le produit des longueurs des intervalles qui le définissent. On vérifie de façon analogue que l'application ainsi construite de $\mathcal{P}(\mathbb{R}^d)$ dans $[0, +\infty]$ est bien une mesure extérieure.

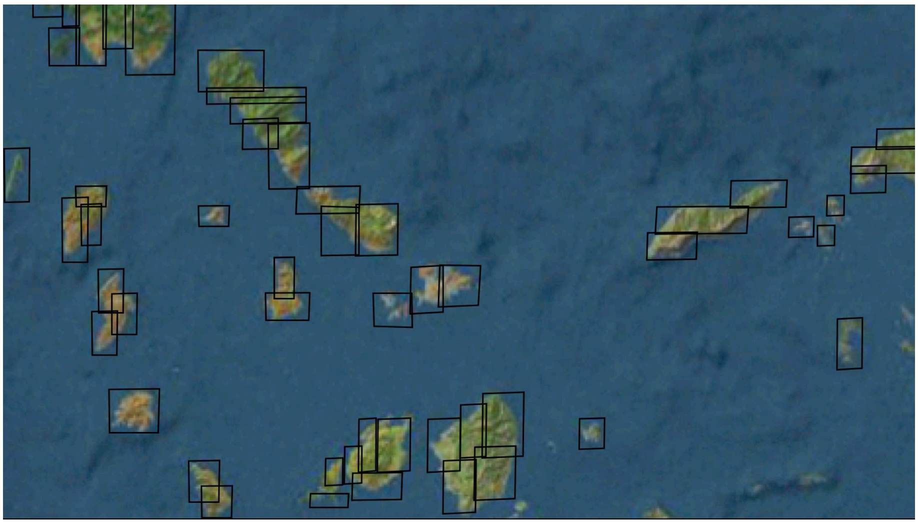

FIGURE III.4.2 – Exemple de recouvrement d'un ensemble (îles grecques) par des pavés (rectangles)

**Proposition III.4.8.** (••) Les parties boréliennes de $\mathbb{R}$ sont mesurables pour la mesure extérieure de Lebesgue $\lambda^\star$ définie ci-dessus.

*Démonstration.* Montrons dans un premier temps que les intervalles de type $]-\infty, b]$ sont mesurables (au sens de la définition III.4.2). Soit $B$ un tel intervalle. Il suffit de montrer que, pour tout $A \subset \mathbb{R}$,

$$\lambda^\star(A) \geq \lambda^\star(A \cap B) + \lambda^\star(A \cap B^c). \tag{III.4.5}$$

Si $\lambda^\star(A) = +\infty$, l'inégalité est automatiquement vérifiée. On se place maintenant dans le cas $\lambda^\star(A) < +\infty$. Soit $\varepsilon > 0$. Par définition de $\lambda^\star$, il existe une collection $(]a_n, b_n[)$ d'intervalles (bornés) dont l'union contient $A$ telle que

$$\lambda^\star(A) \geq \sum (b_n - a_n) - \varepsilon.$$

Dans les lignes qui suivent, nous allons construire, à partir de ce recouvrement, des recouvrements de $A \cap B$ et $A \cap B^c$. Plus précisément, nous allons pour chaque intervalle $]a_n, b_n[$, construire deux intervalles ouverts contenant respectivement $]a_n, b_n[\cap B$ et $]a_n, b_n[\cap B^c$. La seul difficulté est que l'on ne peut pas prendre directement ces intersections comme intervalles ouverts, car certains sont du type $]a_n, b]$. L'idée est alors de recouvrir par un intervalle ouvert du type $]a_n, b + \varepsilon/2^n[$, de façon à ce que que, même en sommant ensuite toutes les longueurs, on ne rajoute d'une quantité contrôlé par $\varepsilon$ arbitrairement petit.

Plus précisément, pour tout $n$, l'intersection de $]a_n, b_n[$ avec $B = ]-\infty, b]$ est soit vide (si $a_n \geq b$), soit un intervalle du type $]a_n, b]$ (si $a_n < b < b_n$), soit $]a_n, b_n[$ (si $b_n \leq b$). Dans le premier cas (intersection vide), on pose $a'_n = b'_n = 0$, de telle sorte de $]a'_n, b'_n[= \emptyset$. Dans le cas où il s'agit d'un intervalle du type $]a_n, b]$, ou si c'est l'intervalle $]a_n, b_n[$ lui-même, on pose

$a'_n = a_n$ et $b'_n = b + \varepsilon/2^n$. De la même manière, l'intersection de $]a_n, b_n[$ avec $B^c = ]b, +\infty[$ est soit vide (on pose alors $a''_n = b''_n = 0$), soit l'intervalle $]b, b_n[$, soit $]a_n, b_n[$. Dans ces deux derniers cas, on pose alors simplement $a''_n = \max(a_n, b)$, $b''_n = b_n$. Par construction, on a

$$b'_n - a'_n + b''_n - a''_n \leq b_n - a_n + \varepsilon/2^n.$$

L'union $]a'_n, b'_n[$ recouvre $A \cap B$, et l'union $]a''_n, b''_n[$ recouvre $A \cap B^c$. On a donc

$$\lambda^\star(A \cap B) \leq \sum (b'_n - a'_n) \text{ et } \lambda^\star(A \cap B^c) \leq \sum (b''_n - a''_n),$$

et ainsi

$$\lambda^\star(A \cap B) + \lambda^\star(A \cap B^c) \leq \sum (b'_n - a'_n) + \sum (b''_n - a''_n) \leq \sum (b_n - a_n) + 2\varepsilon \leq \lambda^\star(A) + 3\varepsilon,$$

pour tout $\varepsilon > 0$, ce qui prouve l'inégalité (III.4.5). L'intervalle $]-\infty, b]$ est donc mesurable pour tout $b \in \mathbb{R}$.

D'après le théorème III.4.4, la famille des parties mesurables pour $\lambda^\star$ est une tribu, qui contient les $]-\infty, b]$ d'après ce que l'on vient de voir. Cette famille contient donc la tribu engendrée par ces intervalles, qui est la tribu borélienne d'après la proposition III.2.6. $\square$

**Définition III.4.9.** (Mesure de Lebesgue sur $\mathbb{R}$ et $\mathbb{R}^d$ ($\bullet$))
On définit la mesure de Lebesgue $\lambda$ comme la mesure construite, selon le théorème III.4.4, à partir de la mesure extérieure de Lebesgue $\lambda^\star$ définie par la proposition III.4.6. Cette mesure est définie sur la tribu des parties mesurables pour la mesure extérieure $\lambda^\star$. On appelle *tribu de Lebesgue* cette tribu. Elle contient la tribu des boréliens d'après la proposition III.4.8.

La mesure de Lebesgue sur $\mathbb{R}^d$ est définie de la même manière à partir de la mesure de Lebesgue extérieure sur $\mathbb{R}^d$ (proposition III.4.7).

**Exercice III.4.2.** (Caractère $\sigma$-fini de la mesure de Lebesgue ($\bullet$))
Montrer que la mesure de Lebesgue sur $\mathbb{R}$ n'est pas finie, mais qu'elle est $\sigma$–finie.

**Proposition III.4.10.** (Invariance par translation de $\lambda$)
La mesure de Lebesgue est invariante par translation sur $\mathbb{R}$ (et sur $\mathbb{R}^d$) : pour tout $A$ mesurable, tout $c \in \mathbb{R}$, $A + c$ est mesurable, et

$$\lambda(A + c) = \lambda(A).$$

*Démonstration.* Les collections d'intervalles $(]a_i, b_i[)_{i \in \mathbb{N}}$ qui recouvrent $A + c$ sont obtenues à partir de celles recouvrant $A$ en translatant tous les intervalles de $c$. Comme les translations ne changent pas les longueurs des intervalles, la mesure extérieure $\lambda^\star(A+c)$ est égale à $\lambda^\star(A)$. La mesure de Lebesgue $\lambda$ étant égale à cette mesure extérieure sur un sous-ensemble de $\mathcal{P}(X)$, on en déduit l'invariance par translation pour $\lambda$. $\square$

**Remarque III.4.11.** La propriété précédente implique une certaine forme d'*unicité* de la mesure de Lebesgue : si une mesure définie sur les boréliens affectent aux intervalles leur longueurs, alors il s'agit de la mesure de Lebesgue.

La question de savoir si toutes les parties de $X$ sont mesurables pour $\lambda^\star$ n'a pas encore été abordée. Si c'était le cas, la tribu associée $\mathcal{A}$ (voir théorème III.4.4) serait $\mathcal{P}(X)$ tout entier, et $\lambda^\star$ serait une mesure sans qu'il soit nécessaire d'élaguer la tribu discrète de ses membres non mesurables. Nous allons voir que ça n'est pas le cas : il existe bien des parties de $\mathbb{R}$ qui *ne sont pas mesurables* pour $\lambda^\star$, et qui sont donc exclues de la tribu des parties sur laquelle $\lambda$ est définie. La proposition suivante établit directement l'existence d'un ensemble non mesurable pour $\lambda$. Le principe de cette démonstration reprend une idée simple évoquée dans l'introduction : nous allons en substance décomposer l'intervalle $]0,1[$ en une infinité dénombrable de parties qui, *si elles sont mesurables*, ont nécessairement pour mesure une même valeur. On est alors confronté à une alternative sans issue : si cette valeur est nulle, alors $]0,1[$ est de mesure nulle, et si la valeur est strictement positive, la mesure de $]0,1[$ est infinie.

**Proposition III.4.12.** (Ensemble de Vitali (• • •))
Il existe une partie de $\mathbb{R}$ qui n'est pas $\lambda$–mesurable .

*Démonstration.* On se propose de construire une partie de $I = ]0,1[$ qui n'est pas mesurable. On introduit sur $I$ la relation d'équivalence suivante :

$$x \,\mathcal{R}\, y \Longleftrightarrow y - x \in \mathbb{Q}.$$

On choisit[8] un représentant de chaque classe, et l'on note $C$ l'ensemble des représentants ainsi choisis. On considère maintenant une énumération $(q_n)$ des rationnels de l'intervalle $]-1,1[$, et l'on s'intéresse à la collection des ensembles $q_n + C$, dont nous allons montrer qu'elle vérifie trois propriétés :

$(i)$ Les $q_n + C$ sont disjoints. En effet, si $q_n + x = q_m + y$, avec $x$ et $y$ dans $C$, on a $y - x = q_n - q_m \in \mathbb{Q}$ donc $y$ et $x$ appartiennent à la même classe. Mais comme $C$ est constitué de représentants uniques de chaque classe, cela implique $x = y$, d'où nécessairement $q_n = q_m$.

$(ii)$ L'union des $q_n + C$ est incluse dans l'intervalle $]-1,2[$. C'est une conséquence directe du fait que $C \subset ]0,1[$ et $q_n \in ]-1,1[$ pour tout $n$.

$(iii)$ L'intervalle $]0,1[$ est inclus dans l'union des $q_n + C$. Tout $y \in ]0,1[$ appartient à sa classe $\overline{y}$, qui admet un (unique) représentant $x$ dans $C$, donc $y = x + q$, avec $q$ rationnel de l'intervalle $]-1,1[$, il s'agit donc de l'un des $q_n$ de notre énumération.

Si $C$ est mesurable, alors les $q_n + C$ sont mesurables, avec $\lambda(q_n + C) = \lambda(C)$ (d'après l'invariance par translation établie dans la proposition III.4.10). D'après $(i)$ et l'additivité de la mesure $\lambda$, on a alors

$$\lambda\left(\bigcup_{n\in\mathbb{N}} (q_n + C)\right) = \sum_{n\in\mathbb{N}} \lambda\left(q_n + C\right) = \sum_{n\in\mathbb{N}} \lambda\left(C\right).$$

Si $\lambda(C) = 0$, alors la somme ci-dessus est nulle, ce qui est absurde car, d'après $(iii)$, $\cup(q_n + C)$ contient l'intervalle $]0,1[$, qui est de mesure 1. Si $\lambda(C) > 0$, alors la somme est infinie, ce qui

8. Il s'agit d'un point très délicat de la construction. Lorsque l'on dispose d'une infinité d'ensemble, il existe parfois une manière de *choisir* un élément de chacun des ensembles. Par exemple s'il s'agit d'intervalles fermés bornés, on peut prendre la borne inférieure. Ici il s'agit d'ensembles qui n'admettent pas de plus petit (ni de plus grand) élément, la procédure n'est donc pas applicable. De fait, on peut se convaincre qu'il n'existe pas de procédure systématique pour effectuer ce choix, et la possibilité d'extraire de la collection d'ensemble une collection de représentants repose sur ce qu'on appelle l'*axiome du choix*, qui est une assertion que l'on ne peut pas démontrer à partir des axiomes de base de la théorie des ensembles, que l'on doit donc rajouter aux fondements de la théorie pour disposer de cette propriété.

est absurde aussi car, d'après $(ii)$, l'union des $q_n + C$ est contenue dans $]-1, 2[$, qui est de mesure finie. L'ensemble $C$ n'est donc pas mesurable. $\square$

**Théorème III.4.13.** (Régularité de la mesure de Lebesgue (• • •))
La mesure de Lebesgue sur $\mathbb{R}^d$ est dite *régulière* au sens où, pour tout $A$ mesurable,

$$\begin{aligned}\lambda(A) &= \inf\left(\lambda(U)\,,\ U \text{ ouvert } \,,\ A \subset U\right) \\ &= \sup\left(\lambda(K)\,,\ K \text{ compact } \,,\ K \subset A\right).\end{aligned}$$

*Démonstration.* Notons en premier lieu que, par monotonie, on a

$$\begin{aligned}\lambda(A) &\leq \inf\left(\lambda(U)\,,\ U \text{ ouvert}\,,\ A \subset U\right) \\ \lambda(A) &\geq \sup\left(\lambda(K)\,,\ K \text{ compact}\,,\ K \subset A\right).\end{aligned}$$

Le fait que la première inégalité soit une égalité découle directement de la définition : pour tout $\varepsilon > 0$, il existe une collection de pavés dont la réunion contient $A$, et telle que la somme des volumes est inférieure à $\lambda(A) + \varepsilon$, et cette réunion est un ouvert, l'infimum est donc égal à $\lambda(A)$.

Pour l'approximation intérieure, supposons dans un premier temps que $A$ est borné, et considérons un fermé borné $C$ qui contient $A$. Soit $\varepsilon > 0$. D'après ce qui précède, il existe $U$ ouvert contenant $C \setminus A$ tel que

$$\lambda(U) \leq \lambda(C \setminus A) + \varepsilon.$$

Soit $K = C \setminus U = C \cap U^c$. Il s'agit d'un fermé borné par construction, donc d'un compact, et il est inclus dans $A$. On a $C \subset K \cup U$, donc $\lambda(C) \leq \lambda(K) + \lambda(U)$. On a donc finalement (du fait que $\lambda(C) = \lambda(C \setminus A) + \lambda(A)$),

$$\begin{aligned}\lambda(K) \geq \lambda(C) - \lambda(U) &= \lambda(C \setminus A) + \lambda(A) - \underbrace{\lambda(U)}_{\leq \lambda(C\setminus A)+\varepsilon} \\ &\geq \lambda(C \setminus A) + \lambda(A) - \lambda(C \setminus A) - \varepsilon \\ &= \lambda(A) - \varepsilon.\end{aligned}$$

On a donc bien égalité entre $\lambda(A)$ et le supremum.

Si $A$ n'est pas borné, on l'écrit comme union croissante $\cup_j A_j$, où les $A_j$ sont les intersections de $A$ avec les boules fermée de rayon $j$. La mesure de $A$ est la limite des $\lambda(A_j)$ d'après la proposition III.3.5, page 75. Si $\lambda(A) < +\infty$, il existe donc $j$ tel que $\lambda(A_j) \geq \lambda(A) - \varepsilon$, et la construction précédente appliquée à $A_j$ assure l'existence d'un compact $K$ tel que $\lambda(K) \geq \lambda(A) - 2\varepsilon$. Si $\lambda(A) = +\infty$, alors $\lambda(A_j)$ tend vers $+\infty$, et l'on peut construire une suite de compacts $K_j \subset A$ tels que $\lambda(K_j) \geq \lambda(A_j) - 2\varepsilon$, qui tend vers $+\infty$. $\square$