INTEGRATION THEORY

TOMAS SJÖDIN

Contents

Date: December 9, 2022.

1. Notation and Prerequisites

We denote by $\mathbb{R}$ and $\overline{\mathbb{R}}$ the real numbers and extended real numbers (i.e. including $\pm\infty$) respectively. We also let $\mathbb{N} = \{1, 2, 3, \ldots\}$ denote the natural numbers and $\mathbb{Q}$ the rational numbers.

Let X denote a non-empty set. For a function $f : X \to \overline{\mathbb{R}}$ the word positive means in its non-strict sense, so f is positive if $f \geq 0$ on X. Likewise the words increasing and decreasing is in the non-strict sense.

By definition we put

$$0 \cdot (\pm\infty) = 0,$$

but of-course care always has to be taken with algebraic operations involving infinity.

We will use the following notation for set operations. Suppose A, B are sets, then $A \cup B$ denotes the union of A and B, $A \cap B$ denotes their intersection and $A \setminus B$ the difference, i.e. the set of all points in A which does not belong to B. In case we work with subsets of some fixed space X we also use the notation A^c for the set $X \setminus A$. For families of sets $\{A_i\}_{i \in I}$ we write $\cup_{i \in I} A_i$ for their union, and $\cap_{i \in I} A_i$ for the intersection. In case $I = \mathbb{N}$ we also use the notation $\cup_{i=1}^{\infty} A_i$,$\cap_{i=1}^{\infty} A_i$. We denote the set of all subsets of X by $\mathcal{P}(X)$. When it comes to sequences of points (sets/functions) we will often just write that "x_n is a sequence" rather than "$\{x_n\}_{n \in \mathbb{N}}$ is a sequence".

Sums of real numbers $\{a_i\}_{i \in I}$, where I is at most countable will be denoted by

$$\sum_{i \in I} a_i.$$

In case $I = \{1, 2, \ldots, n\}$ or $I = \mathbb{N}$ we also use the notation

$$\sum_{i=1}^{n} a_i \text{ and } \sum_{i=1}^{\infty} a_i \text{ respectively.}$$

For two functions $f, g : X \to \overline{\mathbb{R}}$ we write $f \leq g$ if $f(x) \leq g(x)$ for all $x \in X$. If $A \subset \overline{\mathbb{R}}$ then $\sup A$ and $\inf A$ denotes the supremum and infimum of A respectively. Furthermore we introduce the following notation:

- $(f \wedge g)(x) = \min\{f(x), g(x)\}$, $(f \vee g)(x) = \max\{f(x), g(x)\}$,
- We more generally also define for a family of functions $\{f_i\}_{i \in I}$

$$(\bigvee_{i \in I} f_i)(x) = \sup\{f_i(x) : i \in I\}, \quad (\bigwedge_{i \in I} f_i)(x) = \inf\{f_i(x) : i \in I\},$$

  and in case $I = \mathbb{N}$ we also write

$$(\bigvee_{n=1}^{\infty} f_n)(x) = \sup\{f_n(x) : n \in \mathbb{N}\}, \quad (\bigwedge_{n=1}^{\infty} f_n)(x) = \inf\{f_n(x) : n \in \mathbb{N}\},$$

- $f^+ = f \vee 0$, $f^- = -f \vee 0$ so that $f = f^+ - f^-$ and $|f| = f^+ + f^-$,
- if a sequence f_n of functions from X to $\overline{\mathbb{R}}$ converges pointwise to the function f, and if f_n is increasing (in its non-strict sense), then we write $f_n \nearrow f$. Similarly we write $f_n \searrow f$ for decreasing convergence.

(Note that inequalities involving $\vee, \wedge$ by definition reduces directly to pointwise statements about $\min, \max$ of real numbers.)

The choice of notation $\vee, \wedge$ similar to $\cup, \cap$ is of-course no coincidence. Indeed if we by χ_A denote the characteristic function of A, then

$$\chi_A \vee \chi_B = \chi_{A \cup B} \text{ and } \chi_A \wedge \chi_B = \chi_{A \cap B}.$$

Some further simple properties of $\vee, \wedge$, which are direct consequences of the corresponding statements for real numbers since they are pointwise statements, are as follows

Lemma 1.1. *Suppose f, g, h are real-valued functions on some fixed set X. Then the following identities holds:*

(a) $f^- = \frac{1}{2}(|f| - f)$,
(b) $f^+ = \frac{1}{2}(|f| + f)$,
(c) $f \vee (g \wedge h) = (f \vee g) \wedge (f \vee h)$,
(d) $f \wedge (g \vee h) = (f \wedge g) \vee (f \wedge h)$,
(e) $f \vee g = (f + h) \vee (g + h) - h$,
(f) $f \wedge g = (f + h) \wedge (g + h) - h$

Exercise 1.1. Prove Lemma 1.1.

1.1. **Metric Spaces.** We assume that the reader is familiar with metric spaces. A pair (X, ρ) where X is a non-empty set and $\rho : X \times X \to [0, \infty)$ such that for all $x, y, z \in X$

- $\rho(x, y) = \rho(y, x)$,
- $\rho(x, y) \le \rho(x, z) + \rho(z, y)$,
- $\rho(x, y) = 0$ if and only if $x = y$,

is called a metric space, and ρ is called a metric on X. We denote the open ball with radius $r > 0$ and center x by

$$B(x, r) = \{y \in X : \rho(x, y) < r\},$$

and the sphere with radius $r > 0$ and center x by

$$S(x, r) = \{y \in X : \rho(x, y) = r\}.$$

Note that we always have that $B(x, r)$ is open, $S(x, r)$ is closed and $\partial B(x, r) \subset S(x, r)$ but there are situations when these are not the same (e.g. in $X = \mathbb{N}$ we have $B(1, 1) = \{1\}$ so $\partial B(1, 1) = \emptyset$, but $S(1, 1) = \{2\}$).

The space (X, ρ) is said to be **complete** if every Cauchy sequence $x_1, x_2, x_3, \ldots$ in X, i.e. such that given $\varepsilon > 0$ there is N such that $\rho(x_n, x_m) < \varepsilon$ for all $n, m \ge N$, is convergent to some x (i.e. $\rho(x_n, x) < \varepsilon$ for all $n \ge N$).

The space (X, ρ) is said to be **compact** if there for any **open cover** of X by open sets $\{O_i\}_{i \in I}$ (i.e. such that $X = \cup_{i \in I} O_i$) is a finite subcover (that is a finite subset $J \subset I$ such that $X = \cup_{i \in J} O_i$). This is equivalent to the condition that any sequence x_n of points in X has a convergent subsequence.

Given two metric spaces (X, ρ_X) and (Y, ρ_Y) then we say that $f : X \to Y$ is continuous at $x \in X$ if for every $\varepsilon > 0$ there is $\delta_x > 0$ such that

$$\rho_Y(f(x), f(y)) \le \varepsilon \text{ for all } y \in X \text{ such that } \rho_X(x, y) < \delta_x.$$

In case f is continuous at every point of X, then we say that f is continuous on X. If furthermore we can choose δ_x independently of x, then we say that f is uniformly continuous on X.

1.2. **Normed spaces, Banach spaces.** A pair $(V, \|\cdot\|)$ where V is a real vector space and $\|x\| \in [0, \infty)$ for all $x \in V$ such that

- $\|kx\| = |k|\|x\|$ for all $k \in \mathbb{R}$ and $x \in V$,
- $\|x + y\| \le \|x\| + \|y\|$ for all $x, y \in V$,
- $\|x\| = 0$ if and only if $x = 0$,

is called a normed linear (or vector) space, and $\|\cdot\|$ is called a norm on V. A norm induces a metric by

$$\rho(x, y) = \|x - y\|.$$

If (V, ρ) is complete, then $(V, \|\cdot\|)$ (or simply V) is called a Banach space.

A continuous (or bounded) linear functional on a normed real vector space V is a real-valued linear function $F : V \to \mathbb{R}$ such that there is a constant C such that

$$|F(v)| \le C||v|| \quad \text{for all } v \in V.$$

One usually introduces a norm on these functionals as

$$||F|| = \inf\{C : |F(v)| \le C||v|| \quad \text{for all } v \in V\}.$$

It is not hard to verify that these themselves forms a vector space called the **dual space** of V.

This space is furthermore always a Banach space.

2. Motivation

- The concept of area goes back a long time, and this is in some sense the starting point of integration theory.
- The principles of integration were formulated independently by Isaac Newton and Gottfried Leibniz in the late 17th century, through the fundamental theorem of calculus.
- Probably the first definition that a modern mathematician would say comes close to a rigorous definition seems to go back to Cauchy (Lécons sur le calcul infinitesmal, p.81), who defined an integral which he "proved" to be well defined for continuous functions on closed bounded intervals of the real line.
- In the 19:th century Fourier wrote his famous book Théorie analytique de la chaleur (The Analytic Theory of Heat), where the representation of very general functions using infinite trigonometric series were discussed.
- Now the natural question arose. If we can represent a function using Fourier series, then we could also consider any convergent Fourier series to be a function. But as it turned out such a function may not always be so well behaved, and naive definitions of integrals would not do!
- Riemann introduced the first rigorous definition of an integral, and it is usually good enough in situations when we do explicit calculations.
- But it is unfortunately not enough when we need to deal with limiting processes.

E.g. Suppose $f(x) = \sum_{n=0}^{\infty}(a_n \cos(nx) + b_n \sin(nx))$ (Fourier series):
when do we have that $f(x)$ is integrable and when can we say that

$$\int_a^b f(x)dx = \sum_{n=0}^{\infty} \int_a^b (a_n \cos(nx) + b_n \sin(nx))dx?$$

Note that if we define

$$f_k(x) = \sum_{n=0}^{k}(a_n \cos(nx) + b_n \sin(nx)),$$

then clearly (if the Fourier series converges) $f_k(x) \to f(x)$ pointwise, and also clearly

$$\int_a^b f_k(x) = \sum_{n=1}^{k} \int_a^b (a_n \cos(nx) + b_n \sin(nx))dx$$

for any reasonable definition of the integral since the sum is finite.

Hence the question is equivalent to whether

$$\int_a^b f(x)dx = \lim_{k\to\infty} \int_a^b f_k(x)dx.$$

For Riemanns definition it is hard to prove such convergence results. A simple example to explain the problem is given by the Dirichlet function f, which is the characteristic function of the set of rational points in $[0, 1]$. Since the rationals are countable we see that there is an increasing sequence f_n such that f_n is zero apart from n points where it is 1, and $f_n \nearrow f$ everywhere on $[0, 1]$. Clearly $\int_0^1 f_n(x)dx = 0$ in Riemann's sense, but f is not Riemann integrable.

2.1. **Ways to define a more general integral.** There are several ways to define a generalization of the Riemann integral, but there are two that are mainstream (and they are more or less equivalent):

- Lebesgue's approach, starting with generalizing length which gives rise to measure theory, and then define the integral using this.
- Daniell's approach starting with the Riemann integral for say the continuous functions and extending this directly to a larger class of functions.

Here we will begin by developing measure theory, but the integration theory section is developed for the Daniell approach, mainly because this offers no essential extra difficulties and is a bit more general. As it turns out it is not actually much more general, but as a byproduct one get for instance the Riesz representation theorem for linear functionals on the space of continuous functions almost for free.

We do however not only want one-dimensional integrals, but want a theory which can be used to cover situations such as higher dimensional integrals, integrals on curves and surfaces. Furthermore measure theory is the foundation of modern probability theory as introduced by Kolmogorov. Therefore we will use an axiomatic approach, which is relevant in all the above cases.

3. Some general advice

Integration theory is a technical subject. Indeed nothing else should be expected since we are dealing with a subject where simpler natural definitions such as the one by Riemann fails. There are a few general tips that I find worthwhile to write down already here. It is probably a good idea to go through them quickly now and come back to them later.

We will later define measure spaces $(X, \mathcal{M}, \mu)$ where X is a set, $\mathcal{M}$ a suitable collection of subsets of X (called a σ-algebra) and μ a measure which is a positive function on $\mathcal{M}$ with suitable properties.

Very often when we want to prove a statement regarding measures a good approach is the following: Let Φ denote the set of all subsets of $\mathcal{M}$ which satisfies the statement. Maybe a large class is easily seen to belong to Φ (for instance maybe all intervals if we work on the real line). Show that the class Φ is closed under suitable algebraic operations (unions and complementation typically), and finally that it is closed under monotone limits of sequences of sets. Then this typically forces Φ to be the whole set $\mathcal{M}$.

Later when we define integrals a similar approach is often reasonable: Let Φ denote the set of all functions satisfying the statement. Prove that it contains many elements (often so called elementary/simple functions but it could be continuous functions or something else). Often the last step makes it sufficient to prove that Φ is closed under monotone limits, which can often be done through one of the limit theorems we will treat later, but on occasion one could also in analogy with the above need to prove that Φ is closed under suitable algebraic operations (typically that Φ is a vector lattice, or at-least a vector space). Again this will typically force Φ to contain all integrable functions

Finally it is as always important to think about special cases and examples. However, as stated above, the Lebesgue theory is all about generalizing the earlier definitions to include cases where these fails. Therefore it is important not to think of to simple cases. Often it is enough to consider sets and functions on the real line, but choose somewhat nasty such. One example being for instance the intersection between the rational numbers and intervals for instance, and the associated Dirichlet function which is zero in all irrational numbers and one in the rational ones. Other important sets to keep in mind are Cantor type sets (see example 5.23 below).

4. Algebras of sets and lattices of functions

4.1. Vector lattices.

Definition 4.1 (Lattice/Vector Lattice)**.** Let X be a set.

- We say that a collection $\mathcal{L}$ of functions from X to $\overline{\mathbb{R}}$ is a lattice (for $\min, \max$) if for all $f, g \in \mathcal{L}$ we have $f \wedge g \in \mathcal{L}$ and $f \vee g \in \mathcal{L}$,
- We say that a collection of functions $\mathcal{H}$ from X to $\mathbb{R}$ is a vector lattice if it is a lattice and a linear space (i.e. $af + bg \in \mathcal{H}$ if $a, b \in \mathbb{R}$ and $f, g \in \mathcal{H}$).

In these notes lattices/vector lattices will always be with respect to pointwise max/min. Notice that any vector lattice satisfies $0 \in \mathcal{H}$, so for any $h \in \mathcal{H}$ we have in particular $h^+, h^-, |h| \in \mathcal{H}$. We actually have a converse to this.

Proposition 4.2. *If $\mathcal{H}$ is a vector space of real-valued functions then the following are equivalent*

(a) *$\mathcal{H}$ is a vector lattice,*
(b) *$|h| \in \mathcal{H}$ for every $h \in \mathcal{H}$,*
(c) *$h^+ \in \mathcal{H}$ for every $h \in \mathcal{H}$,*
(d) *$h^- \in \mathcal{H}$ for every $h \in \mathcal{H}$.*

Proof. This is a simple consequence of Lemma 1.1, because if for instance (b) holds, then by Lemma 1.1 (a) and (b) we see that h^+, h^- also belongs to $\mathcal{H}$, due to that $\mathcal{H}$ is a vector space. Furthermore $f \vee g = (f - g)^+ + g$ according to Lemma 1.1 part (e) (applied to $h = -g$), and hence for any f, g in $\mathcal{H}$ also $f \vee g$ in $\mathcal{H}$, and likewise we may treat $f \wedge g$. □

Example 4.3. If K is a compact metric space and $\mathcal{H} = C(K)$ denotes the set of all continuous real-valued functions, then $\mathcal{H}$ is a vector lattice of bounded functions.

4.2. **Algebras, σ−algebras and Monotone Classes.**

Definition 4.4. Let X be a non-empty set. If $\mathcal{A} \subset \mathcal{P}(X)$ is a non-empty family which is closed under finite unions and complements, i.e.

($\mathcal{A}1$) $A_1, A_2, \ldots, A_k \in \mathcal{A} \Rightarrow \cup_{j=1}^k A_j \in \mathcal{A}$,

($\mathcal{A}2$) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$,

then $\mathcal{A}$ is called an algebra of sets.

- Note that if $\mathcal{A}$ is an algebra then $X = A \cup A^c$ and $\emptyset = X^c$ belongs to $\mathcal{A}$.
- Since $\cap_{j=1}^k A_j = \left(\cup_{j=1}^k A_j^c\right)^c$ we see that algebras are also closed under finite intersections. Indeed one could equally well have replaced unions by intersections in the definition of algebras.
- Since $A \setminus B = A \cap B^c$ we see that $A \setminus B \in \mathcal{A}$ if $A, B \in \mathcal{A}$

Definition 4.5. If $\mathcal{M} \subset \mathcal{P}(X)$ is an algebra which is closed under countable unions, then $\mathcal{M}$ is called a σ−algebra.

Since $\cap_{j=1}^\infty A_j = \left(\cup_{j=1}^\infty A_j^c\right)^c$ we see that σ−algebras are also closed under countable intersections.

Definition 4.6. A family $\Phi \subset \mathcal{P}(X)$ which is closed under countable increasing unions and countable decreasing intersections is called a monotone class in X.

I.e. Φ is a monotone class if $E_1 \subset E_2 \subset E_3 \subset \ldots$, where each $E_i \in \Phi$, implies that $\cup_{i=1}^\infty E_i \in \Phi$ and $D_1 \supset D_2 \supset D_3 \supset \ldots$, where each D_i in Φ, implies that $\cap_{i=1}^\infty D_i \in \Phi$.

The set $\mathcal{P}(X)$ itself is of-course an example of an algebra, a σ−algebra and a monotone class. Algebras and σ-algebras are very fundamental concepts for measure theory, and a good understanding of them is crucial for the future understanding of the material below.

Exercise 4.1. Prove that an algebra $\mathcal{A}$ is a σ-algebra if and only if it is closed under monotone increasing limits. In particular $\mathcal{A}$ is a σ-algebra if and only if it is a monotone class.

Exercise 4.2. Let X be a set and let $\mathcal{M}$ denote the set of all subsets E of X such that either E or E^c is at most countable. Prove that $\mathcal{M}$ is a σ-algebra.

Exercise 4.3. Suppose $\mathcal{A}$ is an algebra of sets in X, and let $E \subset X$ be any set. Show that the collection

$$\mathcal{A}_E = \{A \cap E : A \in \mathcal{A}\}$$

is an algebra (called the induced algebra) of sets in E. Prove that in case $\mathcal{A}$ is also a σ-algebra then so is $\mathcal{A}_E$.

Exercise 4.4. Let $\{\mathcal{A}_i\}_{i \in I}$ be any collection of algebras on the set X. Show that the intersection $\mathcal{A} = \cap_{i \in I} \mathcal{A}_i$ is also an algebra. Also prove the corresponding result for σ-algebras and monotone classes.

Exercise 4.5. Suppose $\mathcal{E}$ is any family of subsets of the set X.

(a) Show that there is a smallest algebra $\mathcal{A}(\mathcal{E})$ containing all sets in $\mathcal{E}$. This is called the **algebra generated by** $\mathcal{E}$.

(b) Show that there is a smallest σ-algebra $\mathcal{M}(\mathcal{E})$ containing all sets in $\mathcal{E}$. This is called the **σ-algebra generated by** $\mathcal{E}$.

(c) Show that there is a smallest monotone class $\Phi(\mathcal{E})$ containing all sets in $\mathcal{E}$. This is called the **monotone class generated by** $\mathcal{E}$.

(*Hint: $\mathcal{P}(X)$ is an algebra/σ-algebra/monotone class containing $\mathcal{E}$. Now use the previous exercise.*)

Exercise 4.6. Prove that if $\mathcal{A}$ is an algebra which is closed under countable disjoint unions, then $\mathcal{A}$ is a σ-algebra. (*Hint: Given a sequence of sets* $E_1, E_2, E_3, \dots$ *in* $\mathcal{A}$ *look at* $A_1 = E_1, A_2 = E_2 \setminus E_1$, $\dots$, $A_n = E_n \setminus \cup_{j=1}^{n-1} E_j, \dots$)

Definition 4.7 (Borel sets)**.** Given a metric space (X, ρ) the Borel σ-algebra $\mathcal{B}_X$ is the smallest σ-algebra containing all open sets.

Lemma 4.8 (Monotone Class Lemma)**.** *If* $\mathcal{A}$ *is an algebra, then* $\mathcal{M}(\mathcal{A}) = \Phi(\mathcal{A})$.

Proof. Clearly $\Phi(\mathcal{A}) \subset \mathcal{M}(\mathcal{A})$ since σ-algebras are monotone classes. So we only need to prove that $\Phi(\mathcal{A})$ is a σ-algebra. To do so fix $E \in \Phi(\mathcal{A})$ and define

$$\Phi_E = \{F \in \Phi(\mathcal{A}) : F \setminus E, E \setminus F \text{ and } E \cap F \text{ are in } \Phi(\mathcal{A})\}.$$

Φ_E is for each E a monotone class, and $F \in \Phi_E$ if and only if $E \in \Phi_F$. Furthermore if $E \in \mathcal{A}$, then $\mathcal{A} \subset \Phi_E$ since $\mathcal{A}$ is an algebra. Hence we see that $\Phi(\mathcal{A}) \subset \Phi_E$ for every $E \in \mathcal{A}$. Since this implies that any $F \in \Phi(\mathcal{A})$ also belongs to Φ_E in case $E \in \mathcal{A}$ we see that any $E \in \mathcal{A}$ belongs to Φ_F, and hence $\Phi(\mathcal{A}) \subset \Phi_F$ for any $F \in \Phi(\mathcal{A})$. Hence for any $E, F \in \Phi(\mathcal{A})$ $E \setminus F$, $F \setminus E$ and $E \cap F$ belongs to $\Phi(\mathcal{A})$. So $\Phi(\mathcal{A})$ is an algebra and a monotone class, which proves that it is a σ-algebra. □

Lemma 4.9. *Suppose* $\mathcal{A}$ *is an algebra on* X *and* $A_i \in \mathcal{A}$ *for all* $i \in I = \{1, 2, \dots, n\}$. *If we for each subset* $J \subset I$ *define*

$$C_J = \left(\bigcap_{i \in J} A_i \right) \setminus \left(\bigcup_{i \notin J} A_i \right),$$

then we have that the sets C_J *all belong to* $\mathcal{A}$, *they are disjoint and*

$$A_i = \bigcup_{\{J : i \in J\}} C_J.$$

Proof. If $J_1 \neq J_2$, then there must be at least one i which belongs to one but not the other. Without loss assume that $i \in J_1$. Then $C_{J_1} \subset A_i$ and $C_{J_2} \subset A_i^c$, and hence they are disjoint. That they all belong to $\mathcal{A}$ is obvious since there are only a finite number of operations and all sets A_i belongs to $\mathcal{A}$ by assumption.

Finally if $x \in A_i$ then let J_x denote the set of all j such that $x \in A_j$. Clearly $i \in J_x$ and $x \in \cap_{j \in J_x} A_j$ but $x \notin \cup_{j \notin J_x} A_j$ so $x \in C_{J_x}$. □

Proposition 4.10. *If* $\mathcal{E}$ *is a collection of subsets of* X *such that*

- $\emptyset \in \mathcal{E}$,
- *if* $A, B \in \mathcal{E}$ *then* $A \cap B \in \mathcal{E}$,
- *if* $A \in \mathcal{E}$ *then* A^c *is a finite disjoint union of elements in* $\mathcal{E}$.

Then every element in the algebra $\mathcal{A}(\mathcal{E})$ *can be written as a finite disjoint union of elements in* $\mathcal{E}$.

Proof. Let $\mathcal{A}$ denote the collection of all finite disjoint unions of elements in $\mathcal{E}$. We need to prove that $\mathcal{A}$ is an algebra, i.e. closed under taking finite unions and complements.

To prove that $\mathcal{A}$ is closed under taking unions it is clearly enough to prove that the union of any finite family of sets $A_1, A_2, \dots, A_n$ in $\mathcal{E}$ also lies in $\mathcal{A}$ (even if they are not disjoint). We do this by induction over n. For $n = 1$ there is of-course nothing to prove. Suppose now that

$$A_1 \cup A_2 \cup \cdots \cup A_{n-1} = \bigcup_{j=1}^{m} B_j,$$

where the family $B_1, B_2, \dots, B_m$ is disjoint in $\mathcal{E}$. Then let $A_n^c = \cup_{k=1}^r C_k$ where the C_k are disjoint in $\mathcal{E}$. Then we have

$$A_1 \cup A_2 \cup \cdots \cup A_{n-1} \cup A_n = A_n \cup \bigcup_{j=1}^{m} (B_j \cap A_n^c) = A_n \cup \bigcup_{j=1}^{m} \bigcup_{k=1}^{r} (B_j \cap C_k),$$

and the last is a disjoint union of sets in $\mathcal{E}$.

To prove that $\mathcal{A}$ is closed under taking complements, if $A_1, A_2, \dots, A_n \in \mathcal{E}$ with $A_m^c = \cup_{j=1}^{J_m} B_m^j$, where the B_m^j are disjoint elements in $\mathcal{E}$, then

$$\left(\bigcup_{m=1}^{n} A_m \right)^c = \bigcap_{m=1}^{n} \left(\bigcup_{j=1}^{J_m} B_m^j \right) = \bigcup_{\substack{1 \le j_m \le J_m \\ 1 \le m \le n}} \left(B_1^{j_1} \cap B_2^{j_2} \cap \cdots \cap B_n^{j_n} \right),$$

which belongs to $\mathcal{A}$.

□

4.3. **Algebras versus vector lattices.** Later when we define integrals we start with a vector lattice $\mathcal{H}$ that we will call **elementary functions**. The most important case is the case of so called **simple functions** related to an algebra of sets $\mathcal{A}$. We need to understand these simple functions in detail.

Definition 4.11. Let $\mathcal{A}$ be an algebra of subsets of X, and let $\mathcal{H}_{\mathcal{A}}$ denote the linear span of all characteristic functions χ_A for $A \in \mathcal{A}$. I.e.

$$\mathcal{H}_{\mathcal{A}} = \left\{ \sum_{i=1}^{n} a_i \chi_{A_i} : a_1, a_2, \dots, a_n \in \mathbb{R}, A_1, A_2, \dots, A_n \in \mathcal{A} \right\}.$$

Then the functions in $\mathcal{H}_{\mathcal{A}}$ are called simple functions over $\mathcal{A}$.

First of all note that if

$$\phi = \sum_{i=1}^{n} a_i \chi_{A_i}$$

is a simple function, then of-course the representation on the right hand side is not unique. When we wish to define an integral of such later we of-course only want this to depend on ϕ and not the particular representation of ϕ.

If we apply Lemma 4.9 (using the notation from that lemma) we have with $c_J = \sum_{j \in J} a_j$ that

$$\phi = \sum_{J \subset I} c_J \chi_{C_J},$$

so we can always make a refinement to get a case where the sets C_J are disjoint. Also note that ϕ is a simple function if and only if ϕ takes a finite number of different non-zero values $b_1, b_2, \dots, b_k$ and $B_{b_i} = \phi^{-1}(b_i) \in \mathcal{A}$ for each b_i as well as $\phi^{-1}(0) \in \mathcal{A}$. The set $\phi^{-1}(0)$ may be empty, i.e. the value 0 is never attained, or $\phi^{-1}(0) = X$ which corresponds to that ϕ is identically 0. Then

$$\phi = \sum_{i=1}^{k} b_i \chi_{B_{b_i}}.$$

Indeed for all b_j we have

$$B_{b_j} = \bigcup_{\{J : c_J = b_j\}} C_J.$$

This is called the **canonical representation** of ϕ.

Proposition 4.12. *The set $\mathcal{H}_{\mathcal{A}}$ is a vector lattice.*

Proof. Obviously $\mathcal{H}$ is a vector space, hence it is, according to Proposition 4.2, enough to prove that $|\phi| \in \mathcal{H}$ for each $\phi \in \mathcal{H}$. But using the notation from above, if

$$\phi = \sum_{i=1}^{k} b_i \chi_{B_{b_i}}$$

is the canonical representation of ϕ, then

$$|\phi| = \sum_{i=1}^{k} |b_i| \chi_{B_{b_i}}$$

which clearly belong to $\mathcal{H}$ by definition. $\square$

5. Measures

5.1. Measures.

Definition 5.1. Let X be a non-empty set and let $\mathcal{M}$ be a $\sigma-$algebra on X. A measure on $\mathcal{M}$ (or X when $\mathcal{M}$ is understood from the context) is a function $\mu : \mathcal{M} \to [0, \infty]$ such that

($\mu 1$) $\mu(\emptyset) = 0$,
($\mu 2$) For any (pairwise) disjoint collection of sets $\{E_j\}_{j=1}^{\infty}$ in $\mathcal{M}$ we have

$$\mu(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} \mu(E_j).$$

The property ($\mu 2$) is called **countable additivity**, and it essentially captures both the linearity and limiting properties of measures. The triple $(X, \mathcal{M}, \mu)$ is called a **measure space**. Note that countable additivity implies **finite additivity**, since $\emptyset \in \mathcal{M}$.

Some standard terminology concerning $(X, \mathcal{M}, \mu)$:

- The sets in $\mathcal{M}$ are called $\mu-$ measurable or simply measurable,
- μ is called finite if $\mu(X) < \infty$,
- μ is called $\sigma-$finite if $X = \cup_{j=1}^{\infty} E_j$ where $\mu(E_j) < \infty$ for each j.
- A property which holds apart from a set of $\mu-$measure 0 is said to hold $\mu-$almost everywhere ($\mu-$a.e.).

Example 5.2. Most examples of measure spaces of interest requires work to "construct". Here are a few that are important and simple however.
Let $\mathcal{M} = \mathcal{P}(X)$ and let x be a point in X. Define $\mu(E) = 1$ if $x \in E$ and 0 otherwise. Then μ is a measure called the dirac measure at x and is often denoted by δ_x.
Let $\mathcal{M} = \mathcal{P}(X)$ and define $\mu(E)$ =number of points in E (infinite if there are infinitely many). This is the so called counting measure on E.
A more general example which contains the two above as special cases is given if we let $\mathcal{M} = \mathcal{P}(X)$, let $f(x)$ be a positive function on X and define

$$\mu(E) = \sup\{\sum_{j=1}^{k} f(x_j) : x_1, x_2, \ldots x_k \in E\}.$$

Exercise 5.1. Show that if $\mathcal{M}$ is a σ-algebra on X, μ_i is a measure on $\mathcal{M}$ and $c_i \in [0, \infty)$ for each $i = 1, 2, \ldots, n$, then $\sum_{i=1}^{n} c_i \mu_i$ is also a measure on $\mathcal{M}$.

Exercise 5.2. Suppose $(X, \mathcal{M}, \mu)$ is a measure space, and let $E \in \mathcal{M}$. Define $\mu|_E : \mathcal{M} \to [0, \infty]$ by

$$\mu|_E(A) = \mu(A \cap E) \text{ for all } A \in \mathcal{M}.$$

Prove that $\mu|_E$ is a measure. This is called the **restriction** of μ to E. (Note that one could also consider this as a measure on the induced σ-algebra $\mathcal{M}_E$ in a natural way. Formally this is of-course not the same, but the choice makes little difference in practice.)

Exercise 5.3. Suppose $(X, \mathcal{M}, \mu)$ is a measure space. Show that for any $E, F \in \mathcal{M}$ we have $\mu(E) + \mu(F) = \mu(E \cup F) + \mu(E \cap F)$.

Theorem 5.3. *Let $(X, \mathcal{M}, \mu)$ be a measure space.*

(a) *If $A, B \in \mathcal{M}$ and $A \subset B$ then $\mu(A) \leq \mu(B)$,*
(b) *If $E_j \in \mathcal{M}$ for $j = 1, 2, \ldots$, then $\mu(\cup_{j=1}^{\infty} E_j) \leq \sum_{j=1}^{\infty} \mu(E_j)$,*
(c) *If $E_j \in \mathcal{M}$ for $j = 1, 2, \ldots$ and $E_1 \subset E_2 \subset \ldots$, then $\mu(\cup_{j=1}^{\infty} E_j) = \lim_{j\to\infty} \mu(E_j)$,*
(d) *If $E_j \in \mathcal{M}$ for $j = 1, 2, \ldots$ and $E_1 \supset E_2 \supset \ldots$ and $\mu(E_1) < \infty$, then $\mu(\cap_{j=1}^{\infty} E_j) = \lim_{j\to\infty} \mu(E_j)$.*

Proof. We prove (a),(b) and (c) and leave (d) as an exercise. (a) follows since $B = A \cup (B \setminus A)$ and the union is disjoint, so $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$.

To prove (b) define $A_1 = E_1$ and for $n \geq 2$ $A_n = E_n \setminus \cup_{j=1}^{n-1} E_j$. Then the sets $A_n \subset E_n$ are disjoint and $\cup_{j=1}^{n} A_j = \cup_{j=1}^{n} E_j$ for each n (including $n = \infty$). Hence by part (a) we get

$$\mu(\cup_{j=1}^{\infty} E_j) = \mu(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j) \leq \sum_{j=1}^{\infty} \mu(E_j).$$

To prove (c) note that with the above notation

$$\begin{aligned}\mu(\cup_{j=1}^{\infty} E_j) = \mu(\cup_{j=1}^{\infty} A_j) &= \sum_{j=1}^{\infty} \mu(A_j) \\ &= \lim_{n\to\infty} \sum_{j=1}^{n} \mu(A_j) \\ &= \lim_{n\to\infty} \mu(\cup_{j=1}^{n} A_j) = \lim_{n\to\infty} \mu(E_n).\end{aligned}$$

□

The first property is called **monotonicity**, the second **subadditivity** and the two last continuity from below and above respectively.

Note that the assumption $\mu(E_1) < \infty$ in (d) can not be dropped. For instance if we on $\mathbb{N}$ let μ be the counting measure and $E_j = \{n \geq j\}$, then each $\mu(E_j) = \infty$, but $\cap_{j=1}^{\infty} E_j = \emptyset$.

Exercise 5.4. Prove Theorem 5.3 (d).

Exercise 5.5. Suppose $\mathcal{E}$ satisfies the assumptions of Proposition 4.10. Assume that μ and η are two finite measures on $\mathcal{M}(\mathcal{E})$. Prove that if $\mu(A) = \eta(A)$ for all $A \in \mathcal{E}$, then this also holds for all $A \in \mathcal{M}(\mathcal{E})$ (i.e. $\mu = \eta$). *(Hint: Use the result of Proposition 4.10 together with the monotone class lemma and continuity of measures.)*

5.2. **Complete measures.** A measure μ is said to be complete if for every set A such that there is a measurable set B with the property that $A \subset B$ and $\mu(B) = 0$, then A is itself measurable.

Theorem 5.4. *If $(X, \mathcal{M}, \mu)$ is a measure space and we let $\overline{\mathcal{M}}$ consist of all sets $A \in \mathcal{P}(X)$ such that there are $E, F \in \mathcal{M}$ with $E \subset A$, $A \setminus E \subset F$ and $\mu(F) = 0$, then $\overline{\mathcal{M}}$ is a $\sigma-$algebra which contains $\mathcal{M}$.*
Furthermore if we define

$$\overline{\mu}(A) = \mu(E),$$

where E is as in the definition above, then $\overline{\mu}$ is a measure on $\overline{\mathcal{M}}$ such that $\mu(A) = \overline{\mu}(A)$ for all $A \in \mathcal{M}$.

We call the measure space $(X, \overline{\mathcal{M}}, \overline{\mu})$ the completion of $(X, \mathcal{M}, \mu)$.

Exercise 5.6. Prove Theorem 5.4.

5.3. **Outer measures.**

Outer measures are mainly important as a tool to construct measures, but they do have some interest in themselves as-well.

Definition 5.5. Let X be a non-empty set. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ such that

(μ^*1) $\mu^*(\emptyset) = 0$,

(μ^*2) $\mu^*(A) \le \mu^*(B)$ if $A \subset B$,

(μ^*3) $\mu^*(\cup_{j=1}^{\infty} E_j) \le \sum_{j=1}^{\infty} \mu^*(E_j)$,

is called an outer measure on X.

In particular μ^* is **monotone** ((μ^*2)) and **countably subadditive** ((μ^*3)).

Definition 5.6. If μ^* is an outer measure on X, then $A \subset X$ is called μ^*-measurable if

$$\mu^*(E) = \mu^*(A \cap E) + \mu^*(A^c \cap E) \tag{1}$$

holds for all sets $E \subset X$.

Remark 5.7. *Note that*

$$\mu^*(E) \le \mu^*(A \cap E) + \mu^*(A^c \cap E)$$

always holds by subadditivity, so we only need to check the reverse inequality

$$\mu^*(E) \ge \mu^*(A \cap E) + \mu^*(A^c \cap E).$$

Also note that in case $\mu^*(E) = \infty$ this holds trivially.

It is not so easy to motivate why this is the right definition other than that it gives a good theory. Exercise 5.8 will later explain it a bit through the concept of inner measure.

What we can say is that in case there is some σ-algebra $\mathcal{M}$ such that μ^* restricted to $\mathcal{M}$ is a measure, then if $E \in \mathcal{M}$ equation (1) must of-course hold for every set $A \in \mathcal{M}$ by additivity.

In case of Dirac and counting measures (which are actually outer measures since they are measures on all sets) all sets are measurable, but this is VERY RARE.

The following theorem due to Caratheodory is one of the most fundamental results regarding outer measures.

Theorem 5.8. *If μ^* is an outer measure on X then the set $\mathcal{M}$ of all μ^*-measurable sets forms a $\sigma-$algebra. Furthermore the restriction $\mu = \mu^*|_{\mathcal{M}}$ is a complete measure on $\mathcal{M}$.*

Proof. Below $E \subset X$ is an arbitrary set throughout. First of all note that in case $A \subset X$ satisfies $\mu^*(A) = 0$, then we have

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) \le \mu^*(A) + \mu^*(E) = \mu^*(E)$$

by monotonicity, and hence $A \in \mathcal{M}$. In particular $\emptyset \in \mathcal{M}$. It is also clear that $\mathcal{M}$ is closed under complementation due to the symmetry of the definition.

Suppose now that $A, B \in \mathcal{M}$, then if we use equation (1) first for A and then for B we get

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c) \\ &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c).\end{aligned}$$

But $A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$, so by subadditivity we get

$$\mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) \ge \mu^*(E \cap (A \cup B)),$$

and hence (since $A^c \cap B^c = (A \cup B)^c$)

$$\mu^*(E) \ge \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).$$

So $\mathcal{M}$ is an algebra. Furthermore μ^* is finitely additive on $\mathcal{M}$, because if $A \cap B = \emptyset$, then

$$\mu^*(A \cup B) = \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^c) = \mu^*(A) + \mu^*(B).$$

Finally we only need to prove that $\mathcal{M}$ is closed under countable disjoint unions, and that μ^* is countably additive on such to complete the proof. So assume now that $A_1, A_2, A_3, \ldots$ is a disjoint sequence of sets in $\mathcal{M}$, and let $B_n = \cup_{j=1}^n A_j$, $B = \cup_{j=1}^\infty A_j$. Then

$$\begin{aligned}\mu^*(E \cap B_n) &= \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) \\ &= \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}),\end{aligned}$$

so by induction we have $\mu^*(E \cap B_n) = \sum_{j=1}^n \mu^*(E \cap A_j)$. Therefore

$$\begin{aligned}\mu^*(E) &\geq \sum_{j=1}^\infty \mu^*(E \cap A_j) + \mu^*(E \cap B^c) \\ &\geq \mu^*\left(\bigcup_{j=1}^\infty (E \cap A_j)\right) + \mu^*(E \cap B^c) \\ &= \mu^*(E \cap B) + \mu^*(E \cap B^c) \geq \mu^*(E).\end{aligned}$$

Hence all the inequalities above must be equalities, which proves that $B \in \mathcal{M}$. If we let $E = B$ above we also see that μ^* is countably additive on $\mathcal{M}$. □

5.4. **Construction of outer measures.**

Definition 5.9. A class of subsets $\mathcal{K}$ of a set X is called a sequential covering class if $\emptyset \in \mathcal{K}$ and for every set $A \subset X$ there is a sequence E_n of sets in $\mathcal{K}$ such that

$$A \subset \bigcup_{n=1}^\infty E_n.$$

Of-course it is necessary and sufficient that $X \subset \cup_{n=1}^\infty E_n$ above for some sequence E_n in $\mathcal{K}$ for this to be a sequential covering class.

Theorem 5.10. *Let $\mathcal{K}$ be a sequential covering class on X, and let $\lambda : \mathcal{K} \to [0, \infty]$ be such that $\lambda(\emptyset) = 0$. For every set $A \subset X$ define:*

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^\infty \lambda(E_n) : E_n \in \mathcal{K}, \bigcup_{n=1}^\infty E_n \supset A \right\}.$$

Then μ^* is an outer measure on X.

The assumption that $\mathcal{K}$ is a sequential covering class guarantees that there actually is something to take the above infimum over for any subset A of X, so μ^* is well defined.

Proof. The only thing that is not trivial to check is the countable subadditivity. To do so let A_n be a sequence of sets. We need to prove that

$$\mu^*\left(\bigcup_{n=1}^\infty A_n\right) \leq \sum_{n=1}^\infty \mu^*(A_n).$$

To do so notice that by definition, for a given $\varepsilon > 0$ we may for each n choose a sequence E_k^n of sets in $\mathcal{K}$ such that

$$A_n \subset \bigcup_{k=1}^\infty E_k^n,$$

and

$$\sum_{k=1}^\infty \lambda(E_k^n) \leq \mu^*(A_n) + \frac{\varepsilon}{2^n}.$$

Hence

$$\mu^*(A) \leq \sum_{n,k=1}^\infty \lambda(E_k^n) \leq \sum_{n=1}^\infty \mu^*(A_n) + \varepsilon.$$

□

When constructing outer measures one usually starts with some sequential covering class on which one has decided what the measure should be. For instance if we want to generalize length on the real line, then the class of all intervals (a, b) would be a natural choice, and we would then put $\lambda((a, b)) = b - a$. But for the above construction to be helpful one would need to make sure of two things:

- the value is preserved in the sense that $\mu^*(E) = \lambda(E)$ for any $E \in \mathcal{K}$,
- the sets in $\mathcal{K}$ are μ^*-measurable.

To deal with the first question many books starts from the assumption that λ is a pre-measure on an algebra $\mathcal{K}$:

Definition 5.11. If $\mathcal{K}$ is an algebra and $\lambda : \mathcal{K} \to [0, \infty]$ is such that

- $\lambda(\emptyset) = 0$,
- for any countable disjoint sequence of sets $E_1, E_2, E_3, \ldots$ in $\mathcal{K}$ such that also $\cup_{n=1}^{\infty} E_n \in \mathcal{K}$ we have
$$\lambda(\cup_{n=1}^{\infty} E_n) = \sum_{n=1}^{\infty} \lambda(E_n),$$

then λ is called a pre-measure on $\mathcal{K}$

One downside of this approach is that one needs quite a bit of control on both the set $\mathcal{K}$ and the function λ. For instance the open intervals does not form an algebra on $\mathbb{R}$ (one can build an algebra from half-open intervals however). The following definition of a pre-outer measure is more flexible.

Definition 5.12. Suppose $\mathcal{K}$ is a sequential covering class, and that $\lambda : \mathcal{K} \to [0, \infty]$. Then we call λ a pre-outer measure on $\mathcal{K}$ if $\lambda(\emptyset) = 0$ and if for every $E \in \mathcal{K}$ and every covering of E by a countable union $\cup_{j=1}^{\infty} E_j$ of sets in $\mathcal{K}$ we have
$$\lambda(E) \leq \sum_{j=1}^{\infty} \lambda(E_j).$$

The following theorem is an immediate consequence of the definition of μ^* from λ.

Theorem 5.13. *If λ is a pre-outer measure on $\mathcal{K}$, then the outer measure μ^* constructed from $\lambda, \mathcal{K}$ satisfies that*
$$\mu^*(E) = \lambda(E) \text{ for all } E \in \mathcal{K}.$$

It is not difficult to show in the case of length on $\mathbb{R}$ with $\mathcal{K}$ consisting of open intervals (including the empty set) that this gives us a pre-outer measure.

Theorem 5.14. *Suppose that $\mathcal{K}$ is an algebra, and λ is a finitely additive pre-outer measure on $\mathcal{K}$. If we define μ^* as above, then all sets in $\mathcal{K}$ are μ^*-measurable with*
$$\mu^*(A) = \lambda(A) \text{ for all } A \in \mathcal{K}.$$

Remark 5.15. *The conditions in the theorem above are automatically fulfilled if λ is a pre-measure on $\mathcal{K}$.*

Proof. The stated equality is a special case of the preceding theorem, so we only need to prove that all sets in $\mathcal{K}$ are measurable. Suppose now that $A \in \mathcal{K}$ and $E \subset X$ is arbitrary. Choose a covering $\cup_{i=1}^{\infty} A_i$ of E by sets in $\mathcal{K}$. Then since $A_i \cap A$ and $A_i \cap A^c$ all belong to $\mathcal{K}$ (since $\mathcal{K}$ is an algebra) we see that
$$\mu^*(E \cap A) \leq \sum_{i=1}^{\infty} \lambda(A_i \cap A),$$
$$\mu^*(E \cap A^c) \leq \sum_{i=1}^{\infty} \lambda(A_i \cap A^c).$$

Since λ is assumed to be additive we get
$$\lambda(A_i \cap A) + \lambda(A_i \cap A^c) = \lambda(A_i),$$

and hence

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) \leq \sum_{i=1}^{\infty} \lambda(A_i).$$

Since $\mu^*(E)$ by definition is the infimum of sums of the form on the right hand side we see that

$$\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \cap A^c).$$

□

Exercise 5.7. In case $\mathcal{K} = \mathcal{P}(X)$, then λ is a pre-outer measure if and only if it is an outer measure.

Exercise 5.8. Suppose that μ_0 is a finite pre-measure on some algebra $\mathcal{A}$ on X (i.e. $\mu_0(X) < \infty$). Let μ^* denote the outer measure constructed from μ_0 and $\mathcal{A}$ as above, and define the inner measure μ_* on X by

$$\mu_*(A) = \mu_0(X) - \mu^*(A^c).$$

Prove that a set E is measurable (in the sense of equation (1)) if and only if $\mu^*(E) = \mu_*(E)$.

5.5. **Metric outer measures.**

Suppose now that (X, ρ) is a metric space. The $\sigma-$algebra generated by all closed sets (or equivalently all open sets) is called the Borel σ-algebra, and we will denote it by $\mathcal{B} = \mathcal{B}_X$.

We also define for a point $x \in X$ and sets $A, B \subset X$

$$\rho(x, B) = \inf\{\rho(x, y) : y \in B\},$$

$$\rho(A, B) = \inf\{\rho(x, y) : x \in A, y \in B\}.$$

Definition 5.16. An outer measure μ^* on a metric space (X, ρ) is called a metric outer measure if for any sets $A, B \subset X$ we have

$$\rho(A, B) > 0 \Rightarrow \mu^*(A \cup B) = \mu^*(A) + \mu^*(B).$$

Theorem 5.17. *If μ^* is a metric outer measure on (X, ρ), then every Borel set is μ^*-measurable.*

Proof. Since the set of μ^*-measurable sets is a $\sigma-$algebra it is enough to prove that all the closed sets are measurable. So let F be a closed set, and A any set with $\mu^*(A) < \infty$. We need to prove that

$$\mu^*(A) \geq \mu^*(A \cap F) + \mu^*(A \cap F^c).$$

(Note that if $\mu^*(A) = \infty$ this inequality is trivially satisfied, and the reverse inequality is also always satisfied by subadditivity.)

The first step is to prove that there is a sequence of sets $E_n \subset A \cap F^c$ such that

$$\rho(E_n, F) \geq 1/n,$$

and

$$\lim_{n\to\infty} \mu^*(E_n) = \mu^*(A \cap F^c).$$

Once we have this it is easy to complete the proof using the definition of metric outer measures.

To prove the existence of E_n define

$$E_n = \{x \in A \cap F^c : \rho(x, F) \geq 1/n\}.$$

We will prove that these E_n satisfies the above. (Note that $A \cap F^c \subset F^c$ which is open). Clearly E_n is an increasing sequence of sets, so $\mu^*(E_n)$ is an increasing sequence of real numbers which all are bounded from above by $\mu^*(A \cap F^c)$. It is also easy to see that $A \cap F^c = \cup_{n=1}^{\infty} E_n$ (using that F is closed). It is now enough to show that $\mu^*(E_{2n}) \to \mu^*(A \cap F^c)$ as $n \to \infty$. Define the sets

$$G_n = E_{n+1} \setminus E_n$$

and note that

$$A \cap F^c = E_{2n} \cup \left(\bigcup_{k=2n}^{\infty} G_k \right) = E_{2n} \cup \left(\bigcup_{k=n}^{\infty} G_{2k} \right) \cup \left(\bigcup_{k=n}^{\infty} G_{2k+1} \right).$$

Hence

$$\mu^*(E_{2n}) \leq \mu^*(A \cap F^c) \leq \mu^*(E_{2n}) + \sum_{k=n}^{\infty} \mu^*(G_{2k}) + \sum_{k=n}^{\infty} \mu^*(G_{2k+1}).$$

Now we use (which is the main point of the construction)

$$\rho(G_{2k}, G_{2k+2}) \geq \frac{1}{2k+1} - \frac{1}{2k+2} = \frac{1}{(2k+1)(2k+2)} > 0.$$

Since $E_{2n} \supset \cup_{k=1}^{n-1} G_{2k}$ we get (using that μ^* is a metric outer measure)

$$\sum_{k=1}^{n} \mu^*(G_{2k}) = \mu^*(\cup_{k=1}^{n} G_{2k}) \leq \mu^*(E_{2n}).$$

Hence

$$\sum_{k=1}^{\infty} \mu^*(G_{2k})$$

is convergent. A similar argument will show that

$$\sum_{k=1}^{\infty} \mu^*(G_{2k+1})$$

is convergent. Therefore, since this implies that $\sum_{k=n}^{\infty} \mu^*(G_{2k}) + \sum_{k=n}^{\infty} \mu^*(G_{2k+1}) \to 0$ as $n \to \infty$, together with the above this proves the statement. $\square$

Now we come to the construction of metric outer measures, and we will need some assumptions on our sequential covering class $\mathcal{K}$. We define the diameter of the set A

$$d(A) = \sup\{\rho(x, y) : x, y \in A\}, \quad d(\emptyset) = 0.$$

Theorem 5.18. *Assume that $\mathcal{K}$ is a sequential covering class on X such that for each n the set*

$$\mathcal{K}_n = \{A \in \mathcal{K} : d(A) \leq 1/n\}$$

is also a sequential covering class. Furthermore assume that $\lambda : \mathcal{K} \to [0, \infty]$ with $\lambda(\emptyset) = 0$. For each n we define (in accordance with the previous section) the outer measures

$$\mu_n^*(A) = \inf \left\{ \sum_{k=1}^{\infty} \lambda(E_k) : E_k \in \mathcal{K}_n, \bigcup_{k=1}^{\infty} E_k \supset A \right\}.$$

Then

$$\mu_n^*(A) \leq \mu_{n+1}^*(A)$$

holds for any subset A of X. If we furthermore define

$$\mu_0^*(A) = \lim_{n \to \infty} \mu_n^*(A),$$

then μ_0^* is a metric outer measure.

Proof. Since $\mathcal{K}_n \supset \mathcal{K}_{n+1}$ we see that $\mu_n^*(A) \leq \mu_{n+1}^*(A)$ holds for any $A \subset X$.

To prove that μ_0^* is an outer measure is easy using that all μ_n^* are.

To prove that it is a metric outer measure, assume that $\rho(A, B) > 1/n$ for all $n \geq n_0$ (if $\rho(A, B) > 0$, then there clearly is such n_0). Let $\varepsilon > 0$. For each n we may then cover $A \cup B$ with sets $E_k^n \in \mathcal{K}_n$ such that

$$\mu_n^*(A \cup B) + \varepsilon \geq \sum_{k=1}^{\infty} \lambda(E_k^n).$$

Since $d(E_k^n) \leq 1/n$ for each k we see that each E_k^n can intersect at most one of A and B. From this it follows that

$$\mu_0^*(A \cup B) = \mu_0^*(A) + \mu_0^*(B).$$

(Indeed this follows also for μ_n^* for each $n \geq n_0$). $\square$

One may ask when the definition of μ^* given in section 1:

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \lambda(E_n) : E_n \in \mathcal{K}, \bigcup_{n=1}^{\infty} E_n \supset A \right\}$$

gives the same measure as μ_0^*. In general this is not true. For this to hold one also needs assumptions on λ:

Theorem 5.19. *Assume that for each set $E \in \mathcal{K}$, $\varepsilon > 0$ and $n \in \mathbb{N}$ there is a sequence $E_k \in \mathcal{K}_n$ such that*

$$E \subset \bigcup_{k=1}^{\infty} E_k,$$

and

$$\sum_{k=1}^{\infty} \lambda(E_k) \leq \lambda(E) + \varepsilon.$$

Then $\mu^* = \mu_0^*$.

Proof. Clearly $\mu^*(A) \leq \mu_0^*(A)$ for all sets A by construction. By assumption, for any $A \subset X$ there is a sequence E_j in $\mathcal{K}$ such that

$$\mu^*(A) + \varepsilon/2 \geq \sum_{j=1}^{\infty} \lambda(E_j).$$

Also by assumption each E_j can be covered by sets E_k^j from $\mathcal{K}_n$ such that

$$\sum_{k=1}^{\infty} \lambda(E_k^j) \leq \lambda(E_j) + \frac{\varepsilon}{2^{j+1}}.$$

Hence

$$\begin{aligned} \mu^*(A) &\geq \sum_{j=1}^{\infty} \lambda(E_j) - \frac{\varepsilon}{2} \geq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \left(\lambda(E_k^j) - \frac{\varepsilon}{2^{j+1}} \right) - \frac{\varepsilon}{2} \\ &= \sum_{j,k=1}^{\infty} \lambda(E_k^j) - \varepsilon. \end{aligned}$$

Hence

$$\mu^*(A) \geq \mu_n^*(A) - \varepsilon.$$

Since ε and n are arbitrary the proof is done. □

Exercise 5.9 (Lebesgue measure on $\mathbb{R}^N$)**.** Call a set R a rectangle in $\mathbb{R}^N$ if there are numbers $a_i \leq b_i$, $i = 1, 2, \ldots, N$, such that

$$R = \left\{ x \in \mathbb{R}^N : a_i \leq x_i \leq b_i \text{ for all } i = 1, 2, \ldots, N \right\}.$$

Let $\mathcal{K}$ be the set of all rectangles including $\emptyset$, and define

$$\lambda(R) = (b_1 - a_1)(b_2 - a_2) \cdots (b_N - a_N).$$

Also put $\lambda(\emptyset) = 0$. Show that with the notation from above

$$\mu^* = \mu_0^*,$$

and it is a metric outer measure. The restriction of μ^* to the μ^*-measurable sets, which we denote by $\mathcal{M}_{m_N}$, is called the $N-$dimensional Lebesgue measure, and we denote it by m_N (or simply m when N is understood from the context).
In particular every Borel set is m_N-measurable, and (using the notation from above)

$$m_N(R) = (b_1 - a_1)(b_2 - a_2) \cdots (b_N - a_N)$$

for every rectangle $R \subset \mathbb{R}^N$.

Example 5.20 (Hausdorff measure)**.** This time we let $\mathcal{K}$ consist of all balls (including the empty set)

$$B(a,r) = \{x \in \mathbb{R}^N : |x-a| < r\}.$$

Define, for $s \in (0,\infty)$

$$\lambda_s(B(a,r)) = r^s.$$

The pair $\lambda, \mathcal{K}$ satisfies the hypothesis of theorem 5.18, but not 5.19 in general. The measure μ_0^* constructed above using these is called the $s-$dimensional Hausdorff measure and is often denoted $\mathcal{H}^s$. (It is instructive to look at the case $N = 2$ and $s = 1$ and try to visualize what this measure is for a curve in the plane.) There is an important point here as-well. In metric spaces balls have a central role, and unless we are in one dimension it is not so easy to build algebras around these as it is for rectangles. Therefore it is often better not to rely to heavily on the concept of pre-measures.

5.6. **Regularity and Invariance of Lebesgue measure.**

Theorem 5.21 (Regularity of m_N). *Let $E \in \mathcal{M}_{m_N}$. Then*

$$\begin{aligned} m_N(E) &= \inf\{m_N(U) : E \subset U, \quad U \text{ open}\} \\ &= \sup\{m_N(K) : K \subset E, \quad K \text{ compact}\}. \end{aligned}$$

Theorem 5.22 (Invariance of m_N). *If $E \in \mathcal{M}_{m_N}$, then for every $s \in \mathbb{R}^N$*

$$m_N(\{x + s : x \in E\}) = m_N(E),$$

and for every $t \in \mathbb{R}$

$$m_N(\{tx : x \in E\}) = |t|^N m_N(E).$$

Example 5.23. Before looking at the next exercise here are two situations to look out for, even on the real line. Suppose we let A denote the set of all rational numbers in $[0, 1]$. Let furthermore m denote Lebesgue measure restricted to $[0, 1]$. Since A is countable we clearly have $m(A) = 0$. Indeed, suppose $\varepsilon > 0$ and we enumerate the points in A as $q_0, q_1, q_2, \ldots$, then we may put

$$O = \bigcup_{i=0}^{\infty} \left(q_i - \frac{\varepsilon}{2^{i+1}}, q_i + \frac{\varepsilon}{2^{i+1}}\right).$$

Then $A \subset O$ where O is open, and we have

$$m(O) \leq \sum_{i=0}^{\infty} \frac{\varepsilon}{2^i} = \varepsilon.$$

Also $K = [0, 1] \setminus O$ is compact, and $m(K) \geq 1 - \varepsilon$.
Note that an open subset on the real line is always a finite or countable union of disjoint intervals, whereas nothing of the sort is true for compact sets.
Another case to consider is the Smith–Volterra–Cantor set (similar to the classical Cantor middle third set) which is constructed as follows. (The below material is taken from the Wikipedia page "Smith-Volterra-Cantor set" which contains some nice graphics as-well.)
In the first step we remove the middle $1/4$ from $[0, 1]$ so the remaining set is

$$\left[0, \frac{3}{8}\right] \cup \left[\frac{5}{8}, 1\right].$$

The following steps consist of removing subintervals of width $1/2^{2n}$ from the middle of each of the $2n - 1$ remaining intervals. So for the second step the intervals $(5/32, 7/32)$ and $(25/32, 27/32)$ are removed, leaving

$$\left[0, \frac{5}{32}\right] \cup \left[\frac{7}{32}, \frac{3}{8}\right] \cup \left[\frac{5}{8}, \frac{25}{32}\right] \cup \left[\frac{27}{32}, 1\right].$$

Continuing indefinitely with this removal, the Smith–Volterra–Cantor set is then the set of points that are never removed. Since a set of measure

$$\sum_{n=1}^{\infty} \frac{2^n}{2^{2n+2}} = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = \frac{1}{2}$$

is removed from $[0, 1]$ we see that this set has measure $1/2$.

Exercise 5.10 (*). Prove Theorems 5.21 and 5.22.

Example 5.24 (Non-measurable set)**.** Here we will give an example of a set $N \subset \mathbb{R}$ which is not Lebesgue measurable. More precisely we will prove that there is no map $\mu : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ such that

(1) $\mu(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$ for all disjoint sequences $E_i \subset \mathbb{R}$,
(2) $\mu(\{y : y - x \in E\}) = \mu(E)$ for all $E \subset \mathbb{R}$ and all $x \in \mathbb{R}$,
(3) $\mu([0, 1)) = 1$.

I.e. there is no countably additive translation invariant function μ generalizing length to all subsets of $\mathbb{R}$. Note that the construction below depends on the axiom of choice (indeed if one removes the axiom of choice it is consistent with the other axioms of set theory to assume that all subsets of $\mathbb{R}$ are Lebesgue measurable).
To prove this we define the equivalence relation on $[0, 1)$ by

$$x \sim y \text{ if } x - y \in \mathbb{Q},$$

where $\mathbb{Q}$ denotes the rational numbers. Let $N \subset [0, 1)$ contain exactly one element from each equivalence class (axiom of choice). Also let $R = \mathbb{Q} \cap [0, 1]$ and put

$$N_r = \{x + r : x \in N \cap [0, 1 - r)\} \cup \{x + r - 1 : x \in N \cap [1 - r, 1)\} \text{ for } r \in R.$$

It is easy to see using (2) and (3) above that $\mu(N_r) = \mu(N)$ for each r, and by construction these forms a countable family of pairwise disjoint sets. Hence

$$1 = \mu([0, 1)) = \sum_{r \in R} \mu(N_r) = \sum_{r \in R} \mu(N).$$

However there is only two possibilities for the last sum. Either $\mu(N) = 0$, in which case we get the sum 0, or $\mu(N) > 0$ in which case we get the sum ∞. In either case this gives a contradiction.

Exercise 5.11 (**)**.** Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. Let

$$\mathcal{K} = \{(a, b] : -\infty < a < b < \infty\} \cup \{\emptyset\}.$$

Let

$$\lambda_F((a, b]) = F(b) - F(a) \text{ and } \lambda_F(\emptyset) = 0.$$

Prove that λ_F is a pre-outer measure, that it satisfies the assumptions of Theorem 5.19 and hence that the outer measure μ_F^* constructed from λ_F and $\mathcal{K}$ satisfies

$$\mu_F^*((a, b])) = F(b) - F(a),$$

for all $(a, b] \in \mathcal{K}$. The measure μ_F, which is the restriction of μ_F^* to the μ_F^*-measurable sets (which then includes all the Borel sets) is called the Lebesgue-Steltjes measure. Prove also that

$$\mu_F((a, b)) = F(b^-) - F(a),$$

where $F(b^-) = \lim_{x \to b, x < b} F(x)$.

6. Integrals

In this section we will outline the fundamentals of the definitions and properties of the integral from the Daniell point of view. The case when we start from a measure space $(X, \mathcal{M}, \mu)$ is however the most important, so we start with some basic notions about simple functions and measurability of functions defined on such a space.

We will here restrict attention to extended real-valued functions, i.e. functions

$$f : X \to \overline{\mathbb{R}} = [-\infty, \infty].$$

$\overline{\mathbb{R}}$ is given the usual topology, i.e. we say that a set O is open if:

- for every $x \in O \cap \mathbb{R}$ there is some $r > 0$ such that the ball $B(x, r)$ is contained in O,
- if $\infty \in O$ then there is some $T \in \mathbb{R}$ such that $(T, \infty] \subset O$,
- if $-\infty \in O$ then there is some $T \in \mathbb{R}$ such that $[-\infty, T) \subset O$.

6.1. Measurable functions in the measure space context.

We say that $f : X \to \overline{\mathbb{R}}$ is ($\mathcal{M}$-)measurable if for every open set $O \subset \overline{\mathbb{R}}$ the set $f^{-1}(O) \in \mathcal{M}$.

Exercise 6.1. Show that for any function $f : X \to \overline{\mathbb{R}}$ the set

$$\{A \subset \overline{\mathbb{R}} : f^{-1}(A) \in \mathcal{M}\}$$

forms a σ-algebra in $\overline{\mathbb{R}}$.

Hence f is measurable if and only if $f^{-1}(O) \in \mathcal{M}$ for every Borel set $O \subset \overline{\mathbb{R}}$. It also follows that in case the family of sets $\mathcal{A}$ generates the Borel $\sigma-$algebra (i.e. all sets in $\mathcal{A}$ are Borel sets, and the smallest $\sigma-$algebra containing all of them is the Borel $\sigma-$algebra) then f is measurable iff $f^{-1}(O) \in \mathcal{M}$ for every Borel set $O \in \mathcal{A}$.

In particular f is measurable iff $f^{-1}((a, \infty])$ is measurable for every $a \in \mathbb{R}$, or similarly iff $f^{-1}([-\infty, a))$ is measurable for every $a \in \mathbb{R}$.

The set of measurable functions can quite easily be proved to be closed under many operations:

Theorem 6.1. (a) *If $f, g : X \to \mathbb{R}$ are measurable then*

$$f + g \text{ is measurable.}$$

(b) *If $f, g : X \to \overline{\mathbb{R}}$ are measurable and $a \in \mathbb{R}$ then*

$$af,\ f \vee g \text{ and } f \wedge g$$

are measurable,

(c) *If $f_n : X \to \overline{\mathbb{R}}$ are measurable for all $n \in \mathbb{N}$ then so are*

$$\sup_n f_n,\ \inf_n f_n,\ \limsup_{n\to\infty} f_n,\ \liminf_{n\to\infty} f_n.$$

The reason for the assumption of real values in (a) is simply because the sum $f + g$ is not defined at points where one of the functions is ∞ and the other $-\infty$. In case there are no such points then the result still holds for extended real-valued functions.

Proof. Let us prove two of these and leave the rest as an exercise. Suppose f, g are measurable, then

$$(f+g)^{-1}((\alpha,\infty]) = \{x \in X : f(x) + g(x) > \alpha\} = \bigcup_{q\in\mathbb{Q}} (\{x \in X : f(x) > \alpha - q\} \cap \{x \in X : g(x) > q\}).$$

The right hand side of this is measurable, since each set in the countable union is measurable by assumption.

If we look at a sequence g_n of measurable functions we have instead

$$(\sup_n g_n)^{-1}((\alpha,\infty]) = \{x \in X : \sup_n g_n(x) > \alpha\} = \bigcup_{n=1}^{\infty} \{x \in X : g_n(x) > \alpha\}.$$

Again the right hand side is measurable since it is a countable union of measurable sets. □

Exercise 6.2. Prove the remaining cases of Theorem 6.1.

Exercise 6.3. Suppose $f, g : X \to \overline{\mathbb{R}}$ are measurable and define the set

$$A = \{x \in X : f(x) = \infty, g(x) = -\infty\} \cup \{x \in X : f(x) = -\infty, g(x) = \infty\}.$$

If we let $h(x) = f(x) + g(x)$ for all $x \notin A$ and $h(x) = 0$ otherwise prove that h is measurable.

Exercise 6.4. Give an example of an uncountable family $\{f_i\}_{i\in I}$ of measurable functions for which the supremum is not measurable. *(Hint: Look at $X = \mathbb{R}$ with Lebesgue measure m. You may use the result that there is a non-measurable set $E \subset \mathbb{R}$ without proof.)*

6.2. **Simple functions.**

A function which can be written on the form

$$\phi = \sum_{i=1}^{n} a_i \chi_{A_i}$$

where each $a_i \in \mathbb{R}$ and $A_i \in \mathcal{M}$ is called a **measurable simple function** (so in particular it has values in $\mathbb{R}$). We denote the set of all such by $\mathcal{H}_{\mathcal{M}}$.

Recall that according to Proposition 4.12 the set of measurable simple functions is a vector lattice (w.r.t. max/min).

If we furthermore can choose the sets A_i such that $\mu(A_i) < \infty$ for every i then we say that ϕ is an **integrable simple function**. We denote the set of all such by $\mathcal{H}_\mu$.

Exercise 6.5. Prove that the set $\mathcal{H}_\mu$ of all integrable simple functions is a vector lattice.

Proposition 6.2. *A function $f : X \to \overline{\mathbb{R}}$ is measurable if and only if it is the pointwise limit of a sequence of simple functions. This sequence can be chosen to be increasing in case $f \geq 0$.*
In case the measure space $(X, \mathcal{M}, \mu)$ is σ-finite, then these can even be chosen to be integrable simple functions.

Proof. Suppose $f \geq 0$, then for each n define the functions

$$f_n(x) = \begin{cases} \frac{i-1}{2^n} & \text{if } \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n}, \quad i = 1, 2, \ldots, n2^n \\ n & \text{if } f(x) \geq n. \end{cases}$$

In case f is not positive we may simply choose $f_n^+ \to f^+$ and $f_n^- \to f^-$ pointwise, then $f_n^+ - f_n^-$ converges to f pointwise.

Finally, in case X is σ-finite we can always choose an increasing sequence of sets E_n of measurable sets with finite measure such that $\cup_{n=1}^{\infty} E_n = X$ and multiply our given sequence f_n with χ_{E_n}. □

6.3. **The integral of simple functions in the measure space context.** For an integrable simple functions we know what we want the integral to be:

$$I_\mu(\phi) = \int \phi d\mu := \sum_{i=1}^{n} a_i \mu(A_i).$$

Exercise 6.6. Prove that the value $I_\mu(\phi)$ does not depend on the particular representation of ϕ of the form $\phi = \sum_{i=1}^{n} a_i \chi_{A_i}$, and that $I_\mu : \mathcal{H}_\mu \to \mathbb{R}$ is linear and monotone in the sense that if $\phi \leq \psi$ then $\int \phi d\mu \leq \int \psi d\mu$. *(Hint: given two representations use Lemma 4.9 to get a representation which is a refinement of both in the natural sense.)*

Exercise 6.7. Let $X = \mathbb{N}$, $\mathcal{M} = \mathcal{P}(\mathbb{N})$ and μ be counting measure. Describe the integrable simple functions and the integral I_μ for such.

Lemma 6.3. *In case ϕ_n is a decreasing sequence of integrable simple functions such that $\phi_n \searrow 0$ then $I_\mu(\phi_n) \to 0$.*

Proof. Let $\varepsilon > 0$. Put $A_n = \{x \in X : \phi_n(x) > \varepsilon\}$. By assumption we have that $\cap_{n=1}^{\infty} A_n = \emptyset$ and $\mu(A_1) < \infty$. Let $M = \sup\{\phi_1(x) : x \in X\}$ and $E = \{x \in X : \phi_1(x) > 0\}$. The fact that ϕ_1 is a positive integrable simple function implies that $M \in [0, \infty)$ and that $\mu(E) < \infty$. Since the sequence is decreasing we get

$$\phi_n \leq \varepsilon \chi_E + M \chi_{A_n}.$$

Hence by monotonicity and linearity we get

$$0 \leq I_\mu \phi_n \leq \varepsilon I_\mu \chi_E + M I_\mu \chi_{A_n} = \varepsilon \mu(E) + M \mu(A_n).$$

Since $\varepsilon > 0$ is arbitrary and $\mu(A_n) \to 0$ as $n \to 0$ by continuity from above for measures we get the result. $\square$

6.4. **The Daniell integral.** Here we will work with the Daniell integral, which is a bit more general than the Lebesgue approach. This choice is due to that it gives some definite advantages later without actually offering any extra difficulties in the proofs.

For the rest of this section we assume that X is a set and $\mathcal{H}$ a fixed vector lattice of bounded real valued functions on X.

Definition 6.4. A function $I : \mathcal{H} \to \mathbb{R}$ which is:

- **Linear:** $I(af + bg) = aI(f) + bI(g)$ for all $a, b \in \mathbb{R}$, $f, g \in \mathcal{H}$,
- **Positive:** For all $h \in \mathcal{H}$ with $h \geq 0$ we have $I(h) \geq 0$,
- **Order Continuous:** For all sequences $h_n \in \mathcal{H}$ with $h_n \searrow 0$ we have $I(h_n) \to 0$,

will be called an elementary integral (or Daniell integral) on $\mathcal{H}$. We will usually write $Ih = I(h)$. We call the triple $(X, \mathcal{H}, I)$ an integration space.

Example 6.5. The most important case for us is when we start with a measure space $(X, \mathcal{M}, \mu)$. Then we let $\mathcal{H} = \mathcal{H}_\mu$ denote the integrable simple functions, which we know forms a vector lattice, and for a function $f \in \mathcal{H}$ we let

$$If = I_\mu(f) = \int f d\mu.$$

Then it follows from our earlier results that $(X, \mathcal{H}_\mu, I_\mu)$ is an integration space.

Example 6.6. Let R denote a closed and bounded rectangle in $\mathbb{R}^N$, and let $\mathcal{H} = C(R)$ denote the set of all continuous real valued functions on R. Then $\mathcal{H}$ is a vector lattice, and if we let If denote the Riemann integral of f, then I is an elementary integral on $\mathcal{H}$. (This is so because if a sequence f_n of continuous functions decreases to 0, then the convergence is automatically uniform.)
This example offers a different approach to define the Lebesgue integral in $\mathbb{R}^N$.

Note that we have the following:

- $h \leq k \Rightarrow Ih \leq Ik$, because $0 \leq I(k - h) = Ik - Ih$.
- $-Ih^- \leq Ih \leq Ih^+ \leq I|h|$ and $-Ih^+ \leq -Ih \leq Ih^- \leq I|h|$.
  In particular we have the **triangle inequality:**
- $|Ih| \leq I|h|$.

For the rest of this section I is an elementary integral on the set of elementary functions $\mathcal{H}$. We will now extend the definition of the integral in two steps.

Definition 6.7.

$$\mathcal{L}^+ = \left\{ f : X \to \overline{\mathbb{R}} : \text{ there is a sequence } h_n \in \mathcal{H} \text{ such that } h_n \nearrow f \text{ and } \sup_n Ih_n < \infty \right\}.$$

For $f \in \mathcal{L}^+$ we define

$$If = \sup\{Ih : h \in \mathcal{H}, h \leq f\}.$$

Clearly $\mathcal{H} \subset \mathcal{L}^+$ since we may for $f \in \mathcal{H}$ choose $h_n = f$ for each n in the definition above, and also clearly the definition of I still is consistent with the original value of I on $\mathcal{H}$ by monotonicity. Note that a function in $\mathcal{L}^+$ does not need to be positive, but they are always bounded from below by some function in $\mathcal{H}$, and hence bounded from below by some constant.

Remark 6.8. *Note that in case we work in the context of a measure space $(X, \mathcal{M}, \mu)$ with the integrable simple functions as our elementary functions, then every function in $\mathcal{L}^+$ is measurable, since it is a pointwise limit of a sequence of measurable functions.*

Lemma 6.9. *If h_n is a sequence in $\mathcal{H}$, $h_n \nearrow f$ and $k \in \mathcal{H}$ satisfies $k \leq f$ then $Ik \leq \lim_{n\to\infty} Ih_n$. In particular $Ih_n \nearrow If$.*

Proof. Note that $(k - h_n)^+ \searrow 0$ and $(k - h_n)^+ \in \mathcal{H}$. Hence

$$I(k - h_n) \leq I(k - h_n)^+ \searrow 0.$$

For the last part note that by definition for any $\varepsilon > 0$ there is $k \in \mathcal{H}$ with $k \leq f$ such that $Ik \geq If - \varepsilon$. Now $Ih_n \nearrow \alpha$ such that $\alpha \geq Ik$ gives the result. $\square$

Theorem 6.10. *$\mathcal{L}^+$ is a positive cone (i.e. $af + bg \in \mathcal{L}^+$ for all $a, b \in [0, \infty)$ and $f, g \in \mathcal{L}^+$) and a lattice (but not a vector lattice).*
Furthermore:

(a) *$I(af + bg) = aIf + bIg$ holds for all $f, g \in \mathcal{L}^+$ and $a, b \in [0, \infty)$.*
(b) *$f \in \mathcal{L}^+$, $f \geq 0 \Rightarrow If \geq 0$.*
(c) *If f_n is a sequence in $\mathcal{L}^+$ such that $f_n \nearrow f$, then either $If_n \nearrow \infty$ or $f \in \mathcal{L}^+$ and $If_n \nearrow If$.*
(d) *If f_n is a sequence in $\mathcal{L}^+$ such that $f_n \searrow 0$, then $If_n \searrow 0$.*

Proof. That $\mathcal{L}^+$ is a positive cone and a lattice is easy to verify (just look at $ah_n + bk_n, h_n \vee k_n, h_n \wedge k_n$ where $h_n \nearrow f$ and $k_n \nearrow g$), and also (a) and (b) follows more or less immediately from the definition.

To prove (c), for each n we choose a sequence $h_m^n \in \mathcal{H}$ such that $h_m^n \nearrow f_n$ as $m \to \infty$. Now define $g_n = h_n^1 \vee h_n^2 \vee \cdots \vee h_n^n \in \mathcal{H}$. Clearly g_n is increasing, and since $h_n^i \leq g_n \leq f_n$ for all $i \leq n$ it follows that $g_n \nearrow f$, and also, by Lemma 6.9, we have $Ig_n \nearrow If$. Since $Ig_n \leq If_n \leq If$ we get the desired conclusion.

To prove (d) assume that $f_n \searrow 0$, let $\varepsilon > 0$ be fixed and choose $h_1 \in \mathcal{H}$ such that $h_1 \leq f_1$ and $If_1 \leq Ih_1 + \varepsilon/2$ and $h_2 \leq h_1 \wedge f_2$ such that $I(h_1 \wedge f_2) \leq Ih_2 + \varepsilon/4$. Now note that $I(h_1 \wedge f_2) + I(h_1 \vee f_2) = Ih_1 + If_2$ and $h_1 \vee f_2 \leq f_1$, so

$$Ih_2 + \varepsilon/4 \geq I(h_1 \wedge f_2) = Ih_1 + If_2 - I(h_1 \vee f_2) \geq Ih_1 + If_2 - If_1 \geq If_2 - \varepsilon/2.$$

So

$$Ih_2 \geq If_2 - (\varepsilon/2 + \varepsilon/4).$$

Inductively choose $h_n \leq h_{n-1} \wedge f_n$ and $Ih_n \geq I(h_{n-1} \wedge f_n) - \varepsilon/2^n$. Then one gets

$$Ih_n \geq If_n - \sum_{j=1}^{n} \varepsilon/2^j \geq If_n - \varepsilon.$$

Now since $h_n \searrow 0$ we have by definition that $Ih_n \searrow 0$ and hence $If_n \searrow 0$ since ε is arbitrary.

(We should remark here that some care has to be taken when proving part (d). It would be tempting to try to use (c) on $f_1 - f_n$ for instance, but note that $\mathcal{L}^+$ is not a vector space, so we can not use subtraction.) $\square$

We now further will extend the definition of I.

Definition 6.11. We let $\mathcal{L}$ denote the set of all extended real-valued functions f such that for every $\varepsilon > 0$ there are $h, g \in \mathcal{L}^+$ such that $-h \leq f \leq g$ and $I(h + g) \leq \varepsilon$.

Proposition 6.12. *A function f belongs to $\mathcal{L}$ if and only if there are decreasing sequences $h_n, g_n \in \mathcal{L}^+$ and $\alpha \in \mathbb{R}$ such that*

- *$-h_n \leq f \leq g_n$ for all n,*
- *$-Ih_n \nearrow \alpha$ and $Ig_n \searrow \alpha$.*

Proof. If h_n, g_n are as in the statement, then we have $Ig_n + Ih_n = Ig_n - (-Ih_n) \searrow 0$, and hence we see the sufficiency.

Conversely, if we for each n may choose g_n', h_n' such that $-h_n' \leq f \leq g_n'$ and $I(h_n' + g_n') \leq 1/n$, then we may define $h_n = \wedge_{j=1}^n h_j'$ and $g_n = \wedge_{j=1}^n g_j'$. It is easy to see that $-h_n' \leq -h_n \leq f \leq g_n \leq g_n'$

for each n, so in particular $0 \leq I(g_n + h_n) \leq 1/n$. Clearly Ig_n is decreasing to some α, and since $-Ih_n = Ig_n - I(g_n + h_n)$ we get that $-Ih_n$ (which is increasing) also converges to α. $\square$

We now define for any $f \in \mathcal{L}$

$$If = \alpha,$$

where α is as in Proposition 6.12.

We have $\mathcal{L}^+ \subset \mathcal{L}$ because if $f \in \mathcal{L}^+$, then we may choose $g_n = f$ for all n, and by definition there is a sequence $h_n \in \mathcal{H}$ such that $-h_n \nearrow f$. Note that I is well defined on $\mathcal{L}$ since if $h, g \in \mathcal{L}^+$ are such that $-h \leq f \leq g$ and h_n, g_n are sequences as in the proposition, then $g + h_n \geq 0$, and hence $I(g + h_n) = Ig + Ih_n \geq 0$, or equivalently $-Ih_n \leq Ig$. Similarly $-Ih \leq Ig_n$. Taking limits we hence see that indeed $-Ih \leq \alpha \leq Ig$, so there can be at most one such value α. Finally we note that if $f \in \mathcal{L}$ then

$$If = \sup\{-Ih : h \in \mathcal{L}^+, -h \leq f\} = \inf\{Ig : g \in \mathcal{L}^+, f \leq g\}.$$

The set of integrable functions $\mathcal{L}$ forms a vector space as long as we handle the points where the functions might be $\pm\infty$ in an appropriate way. What we need is the concept of a null set:

Definition 6.13. We say that a set $A \subset X$ is a $(I-)$ null set if there is a function $f \in \mathcal{L}$ such that $f \geq 0$ on X, $f \geq 1$ on A and $If = 0$. A property which holds apart from a null set is said to hold (I-)almost everywhere (a.e.).

Clearly the above definition is equivalent to that $\chi_A \in \mathcal{L}$ with $I\chi_A = 0$, since $0 \leq \chi_A \leq f$. The definition is equivalent to that for each $\varepsilon > 0$ there is $0 \leq g \in \mathcal{L}^+$ such that $g \geq 1$ on A and $Ig \leq \varepsilon$. It is furthermore easy to verify that **a countable union of null sets is itself a null set** (look at $f_n/2^n$).

Exercise 6.8. In case we work in the context of a measure space $(X, \mathcal{M}, \mu)$, show that A is I_μ-null if and only if there is a set $B \supset A$ in $\mathcal{M}$ such that $\mu(B) = 0$. (In general A need not be measurable, but if μ is complete, then it is measurable with measure zero.)

We also have the following characterization.

Lemma 6.14. *A set $A \subset X$ is a null set if and only if for every $\varepsilon > 0$ there is an increasing sequence $h_n \in \mathcal{H}$ with $h_n \geq 0$ on X, $\lim_{n\to\infty} h_n \geq 1$ on A and $Ih_n \leq \varepsilon$ for all n.*

Proof. By definition, if A is a null set, then there is $f \in \mathcal{L}$ such that $f \geq 0$ on X, $f \geq 1$ on A and $If = 0$. Now by definition there is $g \in \mathcal{L}^+$ such that $f \leq g$ and $Ig < If + \varepsilon = \varepsilon$. Again by definition there is $h_n \in \mathcal{H}$ such that $h_n \nearrow g$, and (by looking at $h_n \vee 0$ if necessary instead) we may assume $h_n \geq 0$. This gives the desired sequence.

Conversely assume that we for each m choose an increasing sequence h_n^m with $h_n^m \geq 0$ on X, $\lim_{n\to\infty} h_n^m \geq 1$ on A and $Ih_n^m \leq 1/m$. This sequence $h_n^m \nearrow g_m$ as $n \to \infty$ where by definition now $g_m \in \mathcal{L}^+$ and $g_m \geq 0$ on X, $g_m \geq 1$ on A and $Ig_m \leq 1/m$. Now look at $h_n = g_1 \wedge g_2 \wedge \ldots \wedge g_n$. This sequence of functions in $\mathcal{L}^+$ satisfies $h_n \geq 0$ on X, $h_n \geq 1$ on A and $Ih_n \leq Ig_n \leq 1/n$. Furthermore h_n is decreasing to some function f. Since $Ih_n \searrow 0$ and $0 \leq f \leq h_n$ we see that $f \in \mathcal{L}$ and $If = 0$ as we wanted. $\square$

Remark 6.15. *In case we work in the context of a measure space $(X, \mathcal{M}, \mu)$ with the set of integrable simple functions as our elementary functions, then we see that for any element $f \in \mathcal{L}$ there are elements $h, g \in \mathcal{L}$ such that $h \leq f \leq g$, both h and g are measurable and $h = g$ μ-a.e. To see this note that if we choose h_n, g_n as in Proposition 6.12, then $h_n \searrow -h$ and $g_n \searrow g$ with the stated properties, because $g_n + h_n$ is decreasing with $I(g_n + h_n) \searrow 0$.*

In case μ is complete then in particular f is measurable. In this context we will also write

$$\mathcal{L}^1(\mu) = \{f \in \mathcal{L} : f \text{ is } \mathcal{M}\text{-measurable}\}.$$

In particular $\mathcal{L}^1(\mu)$ and $\mathcal{L}$ are equal if μ is complete, and in either case any function $f \in \mathcal{L}$ can be modified on a set of measure zero to become an element in $\mathcal{L}^1(\mu)$.

We also define

$$\int f d\mu = I_\mu f \text{ for all } f \in \mathcal{L}^1(\mu).$$

Lemma 6.16. (a) *If $f \in \mathcal{L}$ then the set $\{x \in X : |f(x)| = \infty\}$ is a null set. Furthermore if $f \in \mathcal{L}$ and $u : X \to \overline{\mathbb{R}}$ differs from f on at most a null set, then $u \in \mathcal{L}$ and $If = Iu$.*
(b) *If $f \geq 0$ belongs to $\mathcal{L}$ and $If = 0$, then $\{x \in X : f(x) > 0\}$ is a null set.*
(c) *If $f_n \geq 0$ is a decreasing sequence in $\mathcal{L}$ and $If_n \searrow 0$, then $f_n \searrow 0$ a.e.*

Proof. (a): There are $h, g \in \mathcal{L}^+$ such that $-h \leq f \leq g$, and hence

$$\{x \in X : |f(x)| = \infty\} \subset \{x \in X : h^+(x) = \infty\} \cup \{x \in X : g^+(x) = \infty\}.$$

Therefore it is enough to prove that $A = \{x \in X : k(x) = \infty\}$ is a null set for every positive function $k \in \mathcal{L}^+$. But now the function $k/n \in \mathcal{L}^+$ is infinite on A for each n, and since $I(k/n) = \frac{1}{n} Ik \searrow 0$ this shows that A is a null set, because we have $0 \leq \chi_A \leq k/n$ for each n.

Since for the null set $A = \{x \in X : u(x) \neq f(x)\}$ we have $\pm\infty\chi_A \in \mathcal{L}$ we see that (with the notation from above) $-h - \infty\chi_A \leq u \leq g + \infty\chi_A$, and hence we easily get from the definition that $u \in \mathcal{L}$ with the same integral as f.

(b): It is enough to prove that $A_j = \{x \in X : f(x) > 1/j\}$ is a null set for each j. Now by definition of the integral there is a sequence of functions $g_n \in \mathcal{L}^+$ such that $f \leq g_n$ and Ig_n decreases to zero. But for each j we have $\chi_{A_j} \leq jg_n$, and $I(jg_n)$ decreases to zero. Therefore $I(\chi_{A_j}) = 0$ by definition.

(c): By definition of $\mathcal{L}$ we may for each n choose $g_n \in \mathcal{L}^+$ with $f_n \leq g_n$ and $Ig_n \leq If_n + 1/n$. Let $g_n' = \wedge_{j=1}^n g_j$. Then $f_n \leq g_n' \leq g_n$ and g_n' is a decreasing sequence in $\mathcal{L}^+$. Since g_n' decreases to a function g with $0 \leq g \leq g_n'$ we see that $g \in \mathcal{L}$ and $Ig = 0$. Therefore $\{x \in X : g(x) \neq 0\}$ is a null set, and by construction it contains the set of points where the limit of f_n is non-zero. $\square$

What the above lemma says in particular is that how we interpret for instance a sum of two functions $f + g$ in $\mathcal{L}$ on the set where the functions are $\pm\infty$ does not make any difference for the integral. For this reason one can allow integrable functions to be defined only on a subset of X whose complement is a null set. With this at hand we can now formulate the following theorem, where we now use the above in the interpretation of the statement about linear combinations of functions in $\mathcal{L}$.

Theorem 6.17. *$\mathcal{L}$ is a vector lattice. Furthermore $I : \mathcal{L} \to \mathbb{R}$ satisfies:*
(a) *For all $a, b \in \mathbb{R}$ and $f, g \in \mathcal{L}$ we have $I(af + bg) = aIf + bIg$.*
(b) *$f \in \mathcal{L}, f \geq 0 \Rightarrow If \geq 0$.*
(c) *For any sequence $f_n \searrow 0$ in $\mathcal{L}$ we have $If_n \searrow 0$.*

Proof. We note that if $f_1, f_2 \in \mathcal{L}$, and $h_i, g_i \in \mathcal{L}^+$ satisfies $-h_i \leq f_i \leq g_i$ and $I(h_i + g_i) \leq \varepsilon/2$, then $-(h_1 + h_2) \leq f_1 + f_2 \leq g_1 + g_2$ and $I(h_1 + h_2 + g_1 + g_2) \leq \varepsilon$, and hence $f_1 + f_2 \in \mathcal{L}$. Also $-(h_1 \wedge h_2) \leq f_1 \vee f_2 \leq g_1 \vee g_2$, and $I((h_1 \wedge h_2) + (g_1 \vee g_2)) \leq I((h_1 + g_1) + (h_2 + g_2)) \leq \varepsilon$, because $(h_1 \wedge h_2) + (g_1 \vee g_2) \leq h_1 + g_1 + h_2 + g_2$ (since $h_i + g_i \geq 0$). If on the other hand $f \in \mathcal{L}$ and $-h \leq f \leq g$ with $I(h + g) \leq \varepsilon$, then for $a \geq 0$ we have $-ah \leq af \leq ag$ and $I(ah + ag) = aI(h + g) \leq a\varepsilon$. For $a < 0$ we get $ag \leq af \leq -ah$ instead, and $I(-ag - ah) = -aI(h + g) \leq -a\varepsilon$. From these statements it indeed follows easily that $\mathcal{L}$ is a vector lattice and that I as defined on $\mathcal{L}$ is linear, so (a) is proved. (Note that all inequalities and sums are meant a.e. above).

(b) is obvious and the proof of (c) is similar to the proof of part (d) of Theorem 6.10, where we instead of choosing functions $h_n \in \mathcal{H}$ choose functions h_n such that $-h_n \in \mathcal{L}^+$. $\square$

Exercise 6.9. Let $X = \mathbb{N}$, $\mathcal{M} = \mathcal{P}(\mathbb{N})$ and μ be counting measure. Describe $\mathcal{L}^1(\mu)$ and the integral I_μ in this context.

Exercise 6.10 (Integration over subsets)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space. Given $E \in \mathcal{M}$ then we may define integrals with respect to μ over E by the formula

$$\int_E f d\mu = \int f \chi_E d\mu,$$

where we interpret $f\chi_E$ as 0 on E^c regardless of the value of f there. We could alternatively let (see Exercise 5.2 for the definition of $\mu|_E$)

$$\int_E f d\mu = \int f d\mu|_E.$$

Show that these two definitions gives the same value for any function $f \in \mathcal{L}^1(\mu)$.

7. Convergence theorems

Theorem 7.1 (Monotone Convergence Theorem)**.** *If $f_n \in \mathcal{L}$, $f_n \geq 0$ and $f_n \nearrow f$ a.e., then either $If_n \to \infty$ or $f \in \mathcal{L}$ with $If_n \nearrow If$.*

Proof. Without loss of generality we may assume that the convergence is everywhere. Assume that $If_n \nearrow C < \infty$. Let

$$k_1 = f_1, k_2 = f_2 - f_1, \ldots, k_n = f_n - f_{n-1}$$

so that each $k_i \geq 0$ and

$$k_1 + k_2 + \ldots + k_n = f_n.$$

Hence $\sum_{n=1}^{\infty} k_n = f$, and each $k_n \in \mathcal{L}$, $k_n \geq 0$. For each n we may now choose $h'_n, g'_n \in \mathcal{L}^+$ such that

$$0 \leq -h'_n \leq k_n \leq g'_n$$

and

$$I(h'_n + g'_n) \leq \varepsilon/2^n.$$

(Because if $h''_n \in \mathcal{L}^+$ satisfies $-h''_n \leq k_n$ then $h'_n = h''_n \wedge 0 \in \mathcal{L}^+$ also satisfies $-h'_n = (-h''_n)^+ \leq k_n$ since $k_n \geq 0$.) Then

$$h_n = \sum_{j=1}^{n} h'_j \in \mathcal{L}^+ \text{ and } g_n = \sum_{j=1}^{n} g'_j \in \mathcal{L}^+.$$

Furthermore

$$-h_n \leq f_n \leq g_n, \quad I(h_n + g_n) \leq \varepsilon \text{ and } -Ih_n \leq If_n \leq Ig_n.$$

Now $g_n \nearrow g \in \mathcal{L}^+$ for some $g \in \mathcal{L}^+$, and $-h_n$ is also increasing. By theorem 6.10 we get that $Ig_n \nearrow Ig$, where by construction $C \leq Ig \leq C + \varepsilon$. But also $-Ih_n$ is increasing to some real number α such that $\alpha \geq C - \varepsilon$, and hence there is some N such that $-Ih_N > C - 2\varepsilon$. But now $-h_N \leq f \leq g$ and $I(g + h_N) \leq 3\varepsilon$. This gives the result. □

Theorem 7.2 (Beppo-Levi)**.** *Suppose $f_i \in \mathcal{L}$ satisfies $f_i \geq 0$ a.e. for each $i \in \mathbb{N}$. Then either $\lim_{n\to\infty} I(\sum_{i=1}^{n} f_i) = \infty$ or $f = \sum_{i=1}^{\infty} f_i \in \mathcal{L}$ and*

$$I(f) = \sum_{i=1}^{\infty} I(f_i).$$

Proof. Apply the monotone convergence theorem to the sequence $g_n = \sum_{i=1}^{n} f_i$. □

Theorem 7.3 (Fatou's Lemma)**.** *If $f_n \in \mathcal{L}$, $f_n \geq 0$ and $f_n \to f$ pointwise a.e. on X, then either $\liminf_{n\to\infty} If_n = \infty$ or $f \in \mathcal{L}$ with $\liminf_{n\to\infty} If_n \geq If$.*

Proof. Without loss assume that the convergence is everywhere. Assume $\liminf_{n\to\infty} If_n < \infty$. Let $g_n = \wedge_{j=n}^{\infty} f_j$. Then $g_n \le f_n$, g_n is increasing and $g_n \nearrow f$ pointwise on X. Furthermore note that since $0 \le f_n - \wedge_{j=n}^{N} f_j \in \mathcal{L}$ and $I(f_n - \wedge_{j=n}^{N} f_j) \le I(f_n) < \infty$ we see by monotone convergence that $f_n - g_n$ belongs to $\mathcal{L}$, and hence by linearity so does g_n. If we again apply the monotone convergence theorem, then we have $Ig_n \nearrow If$, and hence since $Ig_n \le If_n$ for every n we get the result. □

Theorem 7.4 (Dominated Convergence Theorem). *If $f_n \in \mathcal{L}$, $f_n \to f$ pointwise a.e. on X and $|f_n| \le g \in \mathcal{L}$ a.e. for all n, then $f \in \mathcal{L}$ and $If_n \to If$.*

Proof. Without loss we may assume that the convergence and inequalities holds everywhere. Since $g + f_n \ge 0$ we get by Fatou's lemma

$$\liminf_{n\to\infty} I(g + f_n) = Ig + \liminf_{n\to\infty} If_n \ge I(g + f) = Ig + If,$$

and hence

$$\liminf_{n\to\infty} If_n \ge If.$$

But we also have $g - f_n \ge 0$, and if we apply Fatou's lemma again we get in a similar manner

$$\liminf_{n\to\infty} I(g - f_n) = Ig - \limsup_{n\to\infty} If_n \ge Ig - If,$$

and hence

$$\limsup_{n\to\infty} If_n \le If.$$

□

Exercise 7.1. Suppose $1 \in \mathcal{L}$ and f_n is a sequence in $\mathcal{L}$ which converges uniformly to f. Prove that also $f \in \mathcal{L}$ and $If_n \to If$. Give an example for $\mathbb{R}$ with Lebesgue measure $m = m_1$ of a sequence of positive functions f_n in $\mathcal{L}^1(m)$ which converges uniformly to 0 but $\int f_n d\mu = 1$ for each n. (The assumption $1 \in \mathcal{L}$ corresponds to that the set X has finite measure, so in particular $1 \notin \mathcal{L}^1(m)$.)

Exercise 7.2. On $(0, 1)$ with Lebesgue measure m, show that the sequence

$$f_n(x) = \begin{cases} n & 0 < x < 1/n, \\ 0 & 1/n \le x < 1 \end{cases}$$

satisfies that $f_n \to 0$ pointwise but

$$\int f_n dm = 1.$$

(So the boundedness assumption $|f_n| \le g$ can not be dropped in the dominated convergence theorem.)

Exercise 7.3. Suppose we have two integration spaces $(X, \mathcal{H}_1, I_1)$ and $(X, \mathcal{H}_2, I_2)$ over the same space X, and use $\mathcal{L}_i$, $\mathcal{L}_i^+$ to denote the corresponding classes of integrable functions.
Prove that in case $\mathcal{H}_1 \subset \mathcal{L}_2^+$ with $I_1 h = I_2 h$ for all $h \in \mathcal{H}_1$, then $\mathcal{L}_1 \subset \mathcal{L}_2$ with $I_1 f = I_2 f$ for all $f \in \mathcal{L}_1$, by completing the following steps:

(1) If $h \in \mathcal{L}_1^+$ then there is $h_n \nearrow h$, where $h_n \in \mathcal{H}_1$ and $I_1 h_n = I_2 h_n \le C$ for all n. Hence $h \in \mathcal{L}_2^+$ and $I_1 h_n = I_2 h_n \nearrow I_2 h = I_1 h$ by Theorem 6.10 (c).
(2) If $f \in \mathcal{L}_1$ then there is by assumption non-increasing sequences $h_n, g_n \in \mathcal{L}_1^+$ with $-h_n \le f \le g_n$ as in the definition of the space $\mathcal{L}_1$. Since these functions belongs to $\mathcal{L}_2^+$ with the same value of the integrals, we see that by definition $f \in \mathcal{L}_2$ with $I_1 f = I_2 f$.

(*In particular in case $\mathcal{H}_1 \subset \mathcal{L}_2^+$ and $\mathcal{H}_2 \subset \mathcal{L}_1^+$ with equality of the integrals, then $\mathcal{L}_1 = \mathcal{L}_2$ and the integrals coincide, so in particular we could start from different lattices and still end up with the same integration spaces in the end.*)

8. Two theorems of Stone

8.1. **Stone's theorem.** Here we will look at a special case of Stone's theorem which relates the general Daniell integral to the Lebesgue approach. We will only treat the case when the constant function $1 \in \mathcal{L}$, which corresponds to that X has finite measure. This can be very much generalized, but some assumption about "measurability" of X is needed for the result to hold. (Indeed the assumption can be replaced by that $(1 \wedge h) \vee k \in \mathcal{L}$ for all $h, k \in \mathcal{L}$, but we will not enter into further discussion on this topic here.)

So assume now that $1 \in \mathcal{L}$. We define the following class of sets on X

$$\mathcal{M} = \{A \subset X : \chi_A \in \mathcal{L}\}.$$

Then $\mathcal{M}$ is a $\sigma-$algebra. For instance $X \in \mathcal{M}$ is the same as the assumption that $1 \in \mathcal{L}$, and $\emptyset \in \mathcal{M}$ is the same as $0 \in \mathcal{L}$. That it is closed under complementation and countable unions follows easily from the fact that $\mathcal{L}$ is a vector lattice and monotone convergence respectively. We now also define the measure μ for $A \in \mathcal{M}$ as

$$\mu(A) = I\chi_A.$$

It again follows more or less immediately by definition that this indeed is a measure on $\mathcal{M}$. Also note that this measure is complete.

Theorem 8.1 (Stone). *$\mathcal{L} = \mathcal{L}^1(\mu)$ and*

$$If = I_\mu f = \int f d\mu \text{ for all } f \in \mathcal{L}.$$

Proof. As noted above $(X, \mathcal{M}, \mu)$ is a measure space, and furthermore any $\mu-$integrable simple function f is in $\mathcal{L}$, because the latter space is a vector space. By linearity it also follows immediately that $If = \int f d\mu$ holds for any such function. Furthermore we know that any function in $\mathcal{L}^1(\mu)$ may be written as a difference of monotone limits of such functions, and hence it is clear that $\mathcal{L}^1(\mu) \subset \mathcal{L}$ and that $If = \int f d\mu$ holds for all $f \in \mathcal{L}^1(\mu)$.

The more difficult direction is to prove that $\mathcal{L} \subset \mathcal{L}^1(\mu)$. To prove this inclusion we note that it is enough to prove that any $f \in \mathcal{L}$ with $f \geq 0$ belongs to $\mathcal{L}^1(\mu)$, since any function in $\mathcal{L}$ can be written as the difference of two such functions and $\mathcal{L}^1(\mu)$ is a vector space. So we fix such f. First of all we note that for any $a \in \mathbb{R}$ the set

$$E = \{x \in X : f(x) > a\}$$

is $\mu-$measurable, i.e. $\chi_E \in \mathcal{L}$. To see this note that the functions $f - a \in \mathcal{L}$ and hence all the functions $g_n = 1 \wedge n(f-a)^+ \in \mathcal{L}$. But it is clear that $g_n \nearrow \chi_E$ and hence we get the statement by the dominated convergence theorem. Now for any pair of integers k, n we define the sets

$$E_{k,n} = \{x \in X : f(x) > k2^{-n}\},$$

and the functions

$$f_n = 2^{-n} \sum_{k=1}^{2^{2n}} \chi_{E_{k,n}}.$$

Then we have

$$f_n(x) = \begin{cases} 0 & \text{if } f(x) = 0 \\ 2^{2j}2^{-n} & \text{if } 2^{2j}2^{-n} \leq f(x) < 2^{2(j+1)}2^{-n} \\ 2^n & \text{if } f(x) \geq 2^n \end{cases}.$$

Hence we see that all $f_n \in \mathcal{L}^1(\mu)$, since these are $\mu-$integrable simple functions. Furthermore we see that $f_n \nearrow f$. Since we have

$$\int f_n d\mu = If_n \nearrow If < \infty$$

we may apply the monotone convergence theorem (for $\mathcal{L}^1(\mu)$) to see that $f \in \mathcal{L}^1(\mu)$ and that

$$\int f d\mu = If.$$

This proves the theorem. $\square$

Exercise 8.1. Let R be a closed bounded rectangle in $\mathbb{R}^N$ and $\mathcal{H} = C(R)$ as before and let If be the Riemann integral of f for $f \in \mathcal{H}$. Show that the measure μ given from Stone's theorem above coincides with the Lebesgue measure restricted to R. *(Hint: Use exercise 7.3).*

8.2. **Iterated integrals and the Fubini-Stone theorem.** For this subsection we let $(X, I_X, \mathcal{H}_X)$, $(Y, I_Y, \mathcal{H}_Y)$ and $(W, I_W, \mathcal{H}_W)$ be three integration spaces where furthermore $W = X \times Y$. We will write $\mathcal{L}_X, \mathcal{L}_Y, \mathcal{L}_W$ for the corresponding spaces of integrable functions. The Fubini-Stone theorem states conditions to ensure that we can compute I_W as an iterated integral

$$I_W f(x,y) = I_Y(I_X f(x,y)),$$

in the sense that the function $I_X f(x,y)$ is defined for a.e. fixed y, and hence is a function of y which we may possibly integrate with respect to I_Y. (Of-course the roles of X and Y can be interchanged typically.) Obviously for anything like this to be true I_W must be appropriately connected to I_X and I_Y.

Theorem 8.2 (Fubini-Stone)**.** *With the notation from above, suppose we for every $h \in \mathcal{H}_W$ have*

(a) *$h(x,y) \in \mathcal{L}_X$ for a.e. $y \in Y$,*
(b) *$I_X h(x,y) \in \mathcal{L}_Y$,*
(c) *$I_W h(x,y) = I_Y(I_X h(x,y))$.*

In that case (a),(b) and (c) also holds for all $h \in \mathcal{L}_W$.

Proof. Let Φ denote the set of all functions in $\mathcal{L}_W$ for which $(a), (b)$ and (c) holds. By assumption $\mathcal{H}_W \subset \Phi$ and we want to prove that $\Phi = \mathcal{L}_W$. We first note that due to linearity in all arguments Φ is a vector space. Second of all, suppose ϕ_n is an increasing sequence of functions in Φ such that $\phi_n \nearrow \phi \in \mathcal{L}_W$. If we put $g_n(y) = I_X \phi_n(x,y)$, then g_n is an increasing sequence in $\mathcal{L}_Y$ by assumption, which converges to some element $g \in \mathcal{L}_Y$ by the monotone convergence theorem. Applying the monotone convergence theorem we get

$$\begin{aligned}
&I_W \phi_n \to I_W \phi, \\
&I_X \phi_n(x,y) \to I_X \phi(x,y) \text{ for a.e. } y \in Y \\
&I_Y(I_X \phi_n(x,y)) = I_Y g_n \to I_Y g.
\end{aligned}$$

(The a.e. part in the second line above is apart form the set of all $y \in Y$ for which (a) fails to hold for some n, which is by assumption a countable union of null sets, and hence itself a null set.) In particular we see that $\mathcal{L}_W^+ \subset \Phi$. Using linearity we see that also Φ is closed under bounded decreasing limits.

Suppose finally that f is a function in $\mathcal{L}_W$. Then there are decreasing sequences $h_n, g_n \in \mathcal{L}_W^+$ such that $-h_n \leq f \leq g_n$ and $I_W(h_n + g_n) < 1/n$. First of all this implies that $h_n + g_n$ has a limit which belongs to Φ and

$$I_Y(I_X(h_n(x,y) + g_n(x,y))) \leq 1/n.$$

However this implies that we must have $I_X(h_n(x,y) + g_n(x,y)) \to 0$ for a.e. $y \in Y$. Hence for all y apart from these we must have that $f(x,y) \in \mathcal{L}_X$, since $-I_X h_n(x,y) \leq I_X f(x,y) \leq I_X g_n(x,y)$ holds for all y. This also shows that $I_X f(x,y)$ belongs to $\mathcal{L}_Y$, and the theorem follows. □

Exercise 8.2. Let X and Y be two closed and bounded rectangles in $\mathbb{R}^N$ and $\mathbb{R}^M$ respectively. Furthermore let $W = X \times Y$, which then can be regarded as a rectangle in $\mathbb{R}^{N+M}$ in a natural way. Also let $\mathcal{H}_X, \mathcal{H}_Y$ and $\mathcal{H}_W$ denote the set of continuous functions on X, Y and W respectively. Show that the assumptions in the Fubini-Stone theorem is satisfied if I_X, I_Y and I_W are the Riemann integral of the functions in $\mathcal{H}_X, \mathcal{H}_y$ and $\mathcal{H}_W$ respectively. (Results about the Riemann integral in higher dimensions from advanced calculus can be used without proof.) In this example condition (a) in the Fubini-Stone theorem is satisfied everywhere for the elements in $\mathcal{H}_W$, but is the same true for all elements in $\mathcal{L}_W$?

9. Different types of convergence

In this section we let $(X, \mathcal{M}, \mu)$ be a measure space.

Let f_n be a sequence of measurable functions on X. We know what it means that f_n converges to f pointwise a.e. on X, but there are a few other important concepts of convergence as-well. Indeed just having a.e. convergence says very little in regards to integrals as the following example explains.

Example 9.1. If we define on $[0,1]$ with Lebesgue measure $f_n = n\chi_{(0,1/n]}$, then f_n converges to 0 everywhere pointwise, but we still have $\int_0^1 f_n dm = 1$ for each n.
On the other hand if we put

$$f_n = \frac{n}{\ln(n)}\chi_{(0,1/n]}$$

we have that f_n converges to 0 pointwise and $\int_0^1 f_n dm = 1/\ln(n)$, which converges to 0. However note that if $f_n \leq g$ for each n, then we would have $g(x)$ at-least as big as something comparable to $-1/(x\ln(x))$, which is not integrable on $[0,1]$, and hence we can not have dominated convergence here.

Definition 9.2. We say that a sequence of a.e. real valued functions f_n converges to f

- in $\mathcal{L}^1(\mu)$ if
$$\lim_{n\to\infty}\int |f - f_n| d\mu = 0,$$
- in measure if for every $\varepsilon > 0$ we have
$$\lim_{n\to\infty}\mu(\{x : |f_n(x) - f(x)| \geq \varepsilon\}) = 0,$$
- almost uniformly if there for every $\varepsilon > 0$ is a set $E \subset X$ such that $\mu(E^c) < \varepsilon$ and f_n converges to f uniformly on E.

Remark 9.3. *It is implicitly assumed in the definition for $\mathcal{L}^1(\mu)$ convergence that the functions f_n and f belongs to $\mathcal{L}^1(\mu)$. Also note that in either of the cases f is automatically also a.e. real valued.*

Example 9.4. If we have a sequence f_n which converges to f a.e. on X and we have dominated convergence in the sense that $|f_n| \leq g$, where $g \in \mathcal{L}^1(\mu)$, then according to the dominated convergence theorem, since $|f_n - f|$ converges to 0 a.e. and we have $|f_n - f| \leq 2g$ a.e.,

$$\int |f_n - f| d\mu \to 0 \text{ as } n \to \infty,$$

i.e. $f_n \to f$ in $\mathcal{L}^1(\mu)$. Note however that, as we gave an example of above, it is possible to have convergence in $\mathcal{L}^1$ even if no such g exists.

Exercise 9.1. Prove that in case $f_n \to f$ in measure and also $f_n \to g$ in measure then we must have $f = g$ μ-a.e. (i.e. the limit is unique up to null sets).

Remark 9.5. *We will prove in theorem 9.6 below that in case $f_n \to f$ either in $\mathcal{L}^1(\mu)$ or almost uniformly, then it also converges in measure. So in particular if a sequence for instance converges to say f in measure and to g almost uniformly, then $f = g$ a.e.*

Exercise 9.2. Prove that $\|f\|_1 = \int |f| d\mu$ satisfies the following for all $k \in \mathbb{R}$ and $f, g \in \mathcal{L}^1(\mu)$:

- $\|kf\|_1 = |k|\|f\|_1$,
- $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$,
- $\|f\|_1 = 0$ if and only if $f = 0$ a.e.

Theorem 9.6. *Suppose the functions f_n, f are μ-a.e. real valued*

(a) *If f_n converges to f in measure, then there is a subsequence which converges to f almost uniformly.*
(b) *If f_n converges to f in $\mathcal{L}^1(\mu)$, then f_n converges to f in measure.*
(c) *If f_n converges to f almost uniformly, then it converges also a.e. and in measure.*

Proof. (a): Since

$$\{x \in X : |f_n(x) - f_m(x)| \geq \varepsilon\} \subset \{x \in X : |f_n(x) - f(x)| \geq \varepsilon/2\} \cup \{x \in X : |f_m(x) - f(x)| \geq \varepsilon/2\}$$

we see that we for each positive integer k we may choose n_k so that

$$\mu\left(\left\{x \in X : |f_{n_k}(x) - f_m(x)| \geq \frac{1}{2^k}\right\}\right) < \frac{1}{2^k} \text{ for all } m \geq n_k.$$

We may furthermore choose these so that $n_1 < n_2 < n_3 < \cdots$. If we let

$$E_k = \left\{x \in X : |f_{n_k}(x) - f_{n_{k+1}}(x)| \geq \frac{1}{2^k}\right\},$$

and

$$F_m = \cup_{k=m}^{\infty} E_k,$$

then for all $k \geq j \geq m$ and all $x \in X \setminus F_m$ we have

$$|f_{n_j}(x) - f_{n_k}(x)| = \left|\sum_{l=j}^{k} (f_{n_l}(x) - f_{n_{l+1}}(x))\right| \leq \sum_{l=j}^{k} \left|(f_{n_l}(x) - f_{n_{l+1}}(x))\right| < \sum_{l=j}^{k} \frac{1}{2^l} \leq \frac{1}{2^{m-1}}.$$

Hence f_{n_k} converges to some function $\tilde{f}$ uniformly on $X \setminus F_m$ for each m. By exercise 9.1 it is clear that we must indeed have $\tilde{f} = f$ on $X \setminus F_m$ for each m (because if a sequence converges in measure on X, then so it does on any measurable subset of X trivially). But we also have

$$\mu(F_m) \leq \sum_{k=m}^{\infty} \mu(E_k) \leq \frac{1}{2^{m-1}} \to 0 \text{ as } m \to \infty.$$

(b): Since

$$\int |f_n - f| d\mu \geq \varepsilon \mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\}),$$

we see that $\mu(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\})$ must go to zero as $n \to \infty$ for each fixed $\varepsilon > 0$, since the integral on the left hand side does so.

(c): Let for each $n \in \mathbb{N}$ F_n be a set such that $\mu(F_n) \leq 1/n$ and such that f_n converges uniformly to f in $X \setminus F_n$. Then in particular f_n converges pointwise to f in $X \setminus \cap_{n=1}^{\infty} F_n$, and $\cap_{n=1}^{\infty} F_n$ has measure zero, so f_n converges to f a.e. Also for $\varepsilon > 0$ choose n such that $\varepsilon > 1/n$ and N such that $|f_m - f| \leq 1/n$ on $X \setminus F_n$ for all $m \geq N$ (which is possible due to the uniform convergence). Then $\{x \in X : |f_m(x) - f(x)| \geq 1/n\} \subset F_n$ for all $m \geq N$, and from this we easily get convergence in measure. ∎

Theorem 9.7 (Egoroff)**.** *If $\mu(X) < \infty$ and $f_n \to f$ a.e., where the functions f_n, f are μ-a.e. real valued, then f_n converges to f almost uniformly.*

Proof. We may without loss assume that f and f_n are everywhere real-valued. For $n, k \in \mathbb{N}$ let

$$E_n^k = \bigcap_{m=n}^{\infty} \left\{x \in X : |f_m(x) - f(x)| < \frac{1}{k}\right\} = \left\{x \in X : |f_m(x) - f(x)| < \frac{1}{k} \text{ for all } m \geq n\right\}.$$

Then $E_n^k \subset E_{n+1}^k$. Since f_n converges to f a.e. their union $E^k = \cup_{n=1}^{\infty} E_n^k$ satisfies $\mu(X \setminus E^k) = 0$. Since this implies that $\lim_{n\to\infty} \mu(X \setminus E_n^k) = 0$ (by continuity from above for measures) there are for a given $\varepsilon > 0$ integers n_k such that

$$\mu(X \setminus E_n^k) < \frac{\varepsilon}{2^k} \text{ if } n \geq n_k.$$

If we let $F = \cap_{k=1}^{\infty} E_{n_k}^k$ then

$$\mu(X \setminus F) = \mu\left(\cup_{k=1}^{\infty}(X \setminus E_{n_k}^k)\right) \leq \sum_{k=1}^{\infty} \mu(X \setminus E_{n_k}^k) < \varepsilon.$$

Since $F \subset E_{n_k}^k$ it follows that $|f_n(x) - f(x)| < 1/k$ for any $n \geq n_k$, and hence the series converges uniformly on F. $\square$

Example 9.8. Let for each $n \in \mathbb{N}$ and $k \in \mathbb{N}$ such that $1 \leq k \leq 2^n$

$$\psi_{n,k} = \chi_{[(k-1)/2^n, k/2^n]}.$$

The set of pairs (n,k) of this form is a subset of $\mathbb{N}^2$, and hence we may enumerate the above double sequence as $\phi_1, \phi_2, \phi_3, \ldots$ It is easy to see that this sequence converges to 0 in measure and in $\mathcal{L}^1$ on $[0,1]$ (with Lebesgue measure). However note that for any point $x \in [0,1]$ there are arbitrarily large i such that $\phi_i(x) = 1$, and hence it does not converge to 0 a.e. But we may easily choose a subsequence, e.g. $\psi_{n,1}$, which converges to 0 a.e. This sequence also converges to 0 almost uniformly.

Exercise 9.3. On $\mathbb{R}$ with Lebesgue measure, give an example of a sequence $f_n \geq 0$ which converges to 0 everywhere, but which does not converge to 0 in measure.

Exercise 9.4 (*)**.** We say that a sequence of a.e. real-valued functions f_n on a measure space $(X, \mathcal{M}, \mu)$ is **Cauchy in measure** if

$$\lim_{n,m\to\infty} \mu(\{x \in X : |f_n(x) - f_m(x)| \geq \varepsilon\}) = 0$$

for all $\varepsilon > 0$. Prove that if f_n is Cauchy in measure, then there is a function f which is a.e. real-valued and such that $f_n \to f$ in measure. *(Hint: adapt the proof of part (a) of Theorem 9.6.)*

Exercise 9.5 (**)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space. A family $\{f_i\}_{i\in I} \subset \mathcal{L}^1(\mu)$ is called **uniformly integrable** if for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$\left|\int_E f_i d\mu\right| < \varepsilon \text{ for all } E \in \mathcal{M} \text{ with } \mu(E) < \delta.$$

Prove that:

(a) If I is finite, then any family $\{f_i\}_{i\in I} \subset \mathcal{L}^1(\mu)$ is uniformly integrable.
(b) If $f_n \to f$ in $\mathcal{L}^1(\mu)$ then $\{f_n\}_{n\in\mathbb{N}}$ is uniformly integrable.

10. Product measures

Now we fix two measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \eta)$. A set of the form $A \times B$ where $A \in \mathcal{M}$ and $B \in \mathcal{N}$ is called a measurable rectangle in $X \times Y$. Let $\mathcal{K}$ denote the set of all such measurable rectangles, including the empty set. Clearly $\mathcal{K}$ is a sequential covering class, since $X \times Y$ belongs to it. If we are given any families of sets $A_i \subset X$ and $B_i \subset Y$ for $i = 1, 2, \ldots, n$ then

$$\bigcap_{i=1}^{n}(A_i \times B_i) = (\cap_{i=1}^{n} A_i) \times (\cap_{i=1}^{n} B_i).$$

Hence $\mathcal{K}$ is closed under finite intersections. If we let $\mathcal{A}$ denote the collection of all finite disjoint unions of members from $\mathcal{K}$, then $\mathcal{A}$ is an algebra due to Proposition 4.10.

Let $\mathcal{M} \otimes \mathcal{N}$ denote the so called **product $\sigma-$algebra** which is the smallest $\sigma-$algebra containing all rectangles. We now wish to introduce a measure $\mu \times \eta$ on $\mathcal{M} \otimes \mathcal{N}$ such that

$$\mu \times \eta(A \times B) = \mu(A)\eta(B) \quad \text{for all } A \in \mathcal{M},\ B \in \mathcal{N}.$$

Now let $\lambda : \mathcal{K} \to [0, \infty]$ be defined as

$$\lambda(A \times B) = \mu(A)\eta(B)$$

for a rectangle $A \times B \in \mathcal{K}$ ($\lambda(\emptyset) = 0$).

Lemma 10.1. *The function λ is a pre-outer measure on $\mathcal{K}$.*

Proof. The only non-trivial thing we need to prove is that if $A \times B \subset \cup_{i=1}^{\infty} A_i \times B_i$ for $A, A_i \in \mathcal{M}$, $B, B_i \in \mathcal{N}$, then

$$\mu(A)\eta(B) \le \sum_{i=1}^{\infty} \mu(A_i)\eta(B_i).$$

To do so note that

$$\chi_{A\times B}(x,y) = \chi_A(x)\chi_B(y) \le \chi_{\cup_{i=1}^{\infty} A_i\times B_i}(x,y) \le \sum_{i=1}^{\infty} \chi_{A_i}(x)\chi_{B_i}(y).$$

Hence if we first fix $y \in Y$ we get

$$\begin{aligned}\mu(A)\chi_B(y) &= \int \chi_A(x)\chi_B(y)d\mu(x) \le \int \left(\sum_{i=1}^{\infty} \chi_{A_i}(x)\chi_{B_i}(y)\right) d\mu(x) \\ &= \sum_{i=1}^{\infty} \int \chi_{A_i}(x)\chi_{B_i}(y)d\mu(x) = \sum_{i=1}^{\infty} \mu(A_i)\chi_{B_i}(y).\end{aligned}$$

To be precise, Beppo-Levi's theorem that we used (as we have developed the material) might not actually be applicable in case $\sum_{i=1}^{\infty} \chi_{A_i}(x)\chi_{B_i}(y)$ is not a limit of functions in $\mathcal{L}^1(\mu)$ (which may happen in case $(X, \mathcal{M}, \mu)$ is not σ-finite). However if this happens, then by definition this would have to imply that there is A_i such that $\mu(A_i) = \infty$ and $y \in B_i$, in which case the stated inequality holds trivially. Now we can do a similar argument with y:

$$\begin{aligned}\mu(A)\eta(B) &= \int \mu(A)\chi_B(y)d\eta(y) \le \int \left(\sum_{i=1}^{\infty} \mu(A_i)\chi_{B_i}(y)\right) d\eta(y) \\ &\le \sum_{i=1}^{\infty} \int \mu(A_i)\chi_{B_i}(y)d\mu(y) = \sum_{i=1}^{\infty} \mu(A_i)\eta(B_i).\end{aligned}$$

□

We define the outer measure on $X \times Y$

$$\pi^*(E) = \inf \left\{ \sum_{i=1}^{\infty} \lambda(A_i \times B_i) : E \subset \bigcup_{i=1}^{\infty} (A_i \times B_i), A_i \times B_i \in \mathcal{K} \right\}.$$

Since λ is a pre-outer measure we know that

$$\pi^*(A \times B) = \lambda(A \times B) = \mu(A)\eta(B).$$

Lemma 10.2. *The sets in $\mathcal{M} \otimes \mathcal{N}$ are π^*-measurable.*

Proof. It is enough to prove that $A \times B$ is measurable for each measurable rectangle. For a given set $E \subset X \times Y$ with $\pi^*(E) < \infty$ and $\varepsilon > 0$ we may choose a cover $E \subset \cup_{i=1}^{\infty} A_i \times B_i$ of measurable rectangles such that

$$\pi^*(E) + \varepsilon \ge \sum_{i=1}^{\infty} \mu(A_i)\eta(B_i).$$

Now we use that $(A \times B)^c = (A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c)$ to get

$$\begin{aligned}&E \cap (A \times B) \subset \left(\bigcup_{i=1}^{\infty} (A_i \times B_i)\right) \cap (A \times B) = \bigcup_{i=1}^{\infty} ((A_i \cap A) \times (B_i \cap B)), \\ &E \cap (A \times B)^c = E \cap ((A^c \times B) \cup (A \times B^c) \cup (A^c \times B^c)) \\ &= (E \cap (A^c \times B)) \cup (E \cap (A \times B^c)) \cup (E \cap (A^c \times B^c) \\ &\subset \bigcup_{i=1}^{\infty} ((A_i \cap A^c) \times (B_i \cap B)) \cup \bigcup_{i=1}^{\infty} ((A_i \cap A) \times (B_i \cap B^c)) \cup \bigcup_{i=1}^{\infty} ((A_i \cap A^c) \times (B_i \cap B^c)).\end{aligned}$$

Hence

$$
\begin{aligned}
&\pi^*(E \cap (A \times B)) + \pi^*(E \cap (A \times B)^c) \\
&\leq \sum_{i=1}^{\infty} \mu((A_i \cap A)\eta(B_i \cap B)) \\
&+ \sum_{i=1}^{\infty} (\mu(A_i \cap A^c)\eta(B_i \cap B) + \mu(A_i \cap A)\eta(B_i \cap B^c) + \mu(A_i \cap A^c)\eta(B_i \cap B^c)) \\
&= \sum_{i=1}^{\infty} (\mu(A_i)\eta(B_i \cap B) + \mu(A_i)\eta(B_i \cap B^c)) = \sum_{i=1}^{\infty} \mu(A_i)\eta(B_i) \\
&\leq \pi^*(E) + \varepsilon.
\end{aligned}
$$

Since $\varepsilon > 0$ is arbitrary it follows by definition that $A \times B$ is π^*-measurable. □

Definition 10.3. The restriction of π^* to the σ-algebra $\mathcal{M} \otimes \mathcal{N}$, which we denote $\mu \times \eta$, is called the product measure of μ and η:

$$\mu \times \eta = \pi^*|_{\mathcal{M} \otimes \mathcal{N}}.$$

Remark 10.4. *Typically this measure is not complete (i.e. there are typically π^*-measurable sets which does not belong to $\mathcal{M} \otimes \mathcal{N}$). The reason for this choice is that we want sections*

$$E^y = \{x \in X : (x, y) \in E\}, E_x = \{y \in Y : (x, y) \in E\}$$

to be measurable for a measurable set E.

Now we want to get the Fubini theorem for measure spaces. We already have the Fubini-Stone theorem, which takes us quite a bit on the way, but there is a problem to be addressed. For the product measure above, for $\mathcal{K}$ to be a sequential covering class in general we need to allow sets of the form $A \times B$ where A and B might have infinite measure. If we want to define a suitable integration space of elementary functions and apply the Fubini-Stone theorem this of-course does not work, since then we need a finite value of the integral. Therefore we from now on assume that X and Y are σ-finite (indeed, as we will see in Exercise 10.1 the theorem fails to hold otherwise).

Lemma 10.5. *Suppose that X and Y are σ-finite, then the set $\mathcal{K}'$ of all rectangles $A \times B$ with $\mu(A) < \infty$ and $\eta(B) < \infty$ is a sequential covering class. If we define*

$$\pi'(E) = \inf\{\sum_{i=1}^{\infty} \lambda(A_i \times B_i) : E \subset \bigcup_{i=1}^{\infty} A_i \times B_i, A_i \times B_i \in \mathcal{K}'\}.$$

then $\pi' = \pi^*$.

Proof. If we choose increasing sequences X_n and Y_n in $\mathcal{M}$ and $\mathcal{N}$ respectively with finite measures and such that $X = \cup_{n=1}^{\infty} X_n$ and $Y = \cup_{n=1}^{\infty} Y_n$, then $X \times Y = \cup_{n=1}^{\infty} X_n \times Y_n$, and hence $\mathcal{K}'$ is a sequential covering class.

To see that $\pi' = \pi^*$, then note that since the infimum is taken over a smaller set we always have $\pi'(E) \geq \pi^*(E)$. The proof that π' is a pre-outer measure such that all rectangles in $\mathcal{K}'$ are measurable goes trough word by word as for $\mathcal{K}$ above. Hence we see that

$$\pi'(A \times B) = \pi^*(A \times B)$$

holds for all $A \times B \in \mathcal{K}'$.

Suppose now that $E \subset X_n \times Y_n$ for some n. Then for any cover of E by sets in $\mathcal{K}$, say $E \subset \cup_{i=1}^{\infty} A_i \times B_i$ we may look at the cover $E \subset (A_i \cap X_n) \times (B_i \cap Y_n)$, and see that

$$\sum_{i=1}^{\infty} \lambda(A_i \times B_i) \leq \sum_{i=1}^{\infty} \lambda((A_i \cap X_n) \times (B_i \cap Y_n)).$$

Since the sets $(A_i \cap X_n) \times (B_i \cap Y_n)$ belongs to $\mathcal{K}'$ we see by definition that we get $\pi'(E) \leq \pi^*(E)$ in this case.

Finally for general E with $\pi^*(E) < \infty$ and $\varepsilon > 0$ choose $A \times B \in \mathcal{K}$ with $\pi^*(E) \geq \mu(A)\eta(B) - \varepsilon$. Since $A \times B = \cup_{n=1}^{\infty}(A \cap X_n) \times (B \cap Y_n)$ we see that indeed $A \times B$ is π'-measurable. But due to continuity from below for measures we then get, since $\pi'((A \cap X_n) \times (B \cap Y_n)) = \pi^*((A \cap X_n) \times (B \cap Y_n))$ for each n, that

$$\pi'(A \times B) = \pi^*(A \times B).$$

Hence

$$\pi'(E) \leq \pi^*(A \times B) \leq \pi^*(E) + \varepsilon.$$

□

Now let $W = X \times Y$. To apply the Fubini-Stone theorem to this context let $\mathcal{H}_W$ denote the set of all functions of the form

$$h(x, y) = \sum_{i=1}^{n} c_i \chi_{A_i \times B_i}(x, y) \text{ where } c_i \in \mathbb{R}, A_i \times B_i \in \mathcal{K}'.$$

For each fixed $y \in Y$ we have $h(\cdot, y) \in \mathcal{L}^1(\mu)$ since it is then an integrable simple function ($\chi_{A_i \times B_i}(x, y) = \chi_{A_i}(x)\chi_{B_i}(y)$).

Lemma 10.6. *The set $\mathcal{H}_W$ is a vector lattice of bounded functions, and the function*

$$I_W h(x, y) = I_W\left(\sum_{i=1}^{n} c_i \chi_{A_i \times B_i}\right) = \sum_{i=1}^{n} c_i \mu(A_i)\eta(B_i)$$

is a Daniell integral on $\mathcal{H}_W$ (in particular the value does not depend on the particular representation of h). Furthermore for any $h(x, y) \in \mathcal{H}_W$ as above we have

$$\begin{aligned}\int h(x, y) d\mu \times \eta(x, y) &= \sum_{i=1}^{n} c_i \int \chi_{A_i \times B_i}(x, y) d\mu \times \eta \\ &= \sum_{i=1}^{n} c_i \mu(A_i)\eta(B_i) = \int \left(\int \sum_{i=1}^{n} c_i \chi_{A_i}(x)\chi_{B_i}(y) d\mu(x) \right) d\eta(y).\end{aligned}$$

Proof. Note that $\mathcal{H}_W$ precisely consists of the $\mathcal{A}$-simple functions, and according to Proposition 4.12 we know that they form a vector lattice.

The stated equalities are simple consequences of linearity, and since $I_W = I_{\mu \times \eta}$ on $\mathcal{H}_W$ it is also clear that it is a Daniell integral on $\mathcal{H}_W$. □

Due to the above lemma we see that the Fubini-Stone theorem applies and we get:

Theorem 10.7 (Fubini)**.** *Assume that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \eta)$ are σ-finite. If $f \in \mathcal{L}^1(\mu \times \eta)$ then for a.e. $y \in Y$ the function $f(\cdot, y)$ belongs to $\mathcal{L}^1(\mu)$ and $\int f(x, \cdot) d\mu(x)$ belongs to $\mathcal{L}^1(\eta)$. Furthermore*

$$\int f d\mu \times \eta = \int \left(\int f(x, y) d\mu(x) \right) d\eta(y).$$

Similarly

$$\int f d\mu \times \eta = \int \left(\int f(x, y) d\eta(y) \right) d\mu(x).$$

Fubini's student Tonelli also made the following addition to the theorem.

Theorem 10.8 (Tonelli)**.** *Suppose $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \eta)$ are σ-finite measure spaces and $f : X \times Y \to [0, \infty]$ is $\mu \times \eta$-measurable. If*

$$\int \left(\int f(x, y) d\mu(x) \right) d\eta(y) < \infty,$$

then $f(x, y) \in \mathcal{L}^1(\mu \times \eta)$.

(One can of-course interchange the roles of X and Y above.)

Proof. Let $f_n \nearrow f$ where each $f_n \in \mathcal{L}^1(\mu \times \eta)$ is positive (indeed we know that we may even choose f_n to be integrable simple functions). Then

$$\int f_n d\mu \times \eta = \int \left(\int f_n(x,y) d\mu(x) \right) d\eta(y) \nearrow \int \left(\int f(x,y) d\mu(x) \right) d\eta(y) < \infty,$$

which follows by applying the monotone convergence theorem twice. But this implies immediately by the monotone convergence theorem that $f \in \mathcal{L}^1(\mu \times \eta)$, and hence Fubini's theorem is applicable. $\square$

Remark 10.9. *Tonelli's theorem can be used in connection with Fubini's theorem in the following way: given a measurable function $f : X \times Y \to \overline{\mathbb{R}}$, show that*

$$\int \left(\int |f(x,y)| d\mu(x) \right) d\eta(y) < \infty,$$

to draw the conclusion that $|f|$, and hence f, lies in $\mathcal{L}^1(\mu \times \eta)$

It should be remarked that there are a few things to be careful with:

- *The assumption of σ-finiteness is necessary (see Exercise 10.1).*
- *The assumption that $f \geq 0$ in Tonelli's theorem is necessary (see Exercise 10.2).*
- *The assumption that f is $\mu \times \eta$-measurable is necessary. There are cases where a function $f(x,y)$ may be non-measurable on $X \times Y$ but satisfies that for each fixed x $f(x,\cdot)$ is measurable as a function on Y, and for each fixed y $f(\cdot,y)$ is measurable as a function on X.*

Exercise 10.1. Let $X = [0,1]$ with μ Lebesgue measure restricted to X, and $Y = [0,1]$ with counting measure η (so η is not σ-finite). Let $E = \{(x,x) : x \in [0,1]\} \subset X \times Y$. Prove that, using the notation from above, $\pi^*(E) = \infty$ but

$$\int \left(\int \chi_E(x,y) d\mu(x) \right) d\eta(y) = 0,$$

$$\int \left(\int \chi_E(x,y) d\eta(x) \right) d\mu(y) = 1.$$

Exercise 10.2. Let $X = Y = \mathbb{N}$ with $\mu = \eta$ counting measure. Let furthermore $f(m,n) = 1$ if $m = n$, $f(m,n) = -1$ if $m = n+1$ and $f(m,n) = 0$ otherwise. Prove that $\int |f| d\mu \times \eta = \infty$ but $\int (\int f d\mu) d\eta$ and $\int (\int f d\eta) d\mu$ are both finite but unequal.

Exercise 10.3 (*)**.** As stated above the measure $\mu \times \eta$ is typically not complete. Find (in a book) and state a version of Fubini's theorem for the completion of $\mu \times \eta$.

Exercise 10.4 (Cavalieri's Principle **)**.** Show that if $(X, \mathcal{M}, \mu)$ is a measure space, then for any positive measurable function f and any $q \in (0,\infty)$ we have

$$\int f^q d\mu = q \int_0^\infty t^{q-1} \mu(\{x \in X : f(x) > t\}) dt.$$

(The last integral is with respect to Lebesgue measure on $(0,\infty)$. An interesting remark is that in case we take $q = 1$, then the function $\mu(\{x \in X : f(x) > t\})$ is decreasing, and hence the integral would make sense even from the Riemann point of view. This is actually a different approach to define the integral from a measure μ. Sometimes Cavalieri's principle is also called the bathtub principle.)

11.1. **Signed Measures.**

Definition 11.1. If $\mathcal{M}$ is a $\sigma-$algebra on X and $\nu : \mathcal{M} \to (-\infty, \infty]$ or $\nu : \mathcal{M} \to [-\infty, \infty)$ is such that

- $\nu(\emptyset) = 0$,
- if $A_1, A_2, A_3, \ldots \in \mathcal{M}$ are pairwise disjoint, then $\nu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \nu(A_i)$,

then ν is called a signed measure on $\mathcal{M}$.

Example 11.2. Suppose $(X, \mathcal{M}, \mu)$ is a measure space ($\mu \geq 0$) and that $f \in \mathcal{L}^1(\mu)$, then

$$\nu(A) = \int_A f d\mu := \int f \chi_A d\mu \quad \text{for } A \in \mathcal{M}, \tag{2}$$

is a signed measure on $\mathcal{M}$.
If μ_1 and μ_2 are two (positive) measures on $\mathcal{M}$, where at-least one is finite then

$$\nu(A) = \mu_1(A) - \mu_2(A) \quad \text{for } A \in \mathcal{M}, \tag{3}$$

is a signed measure on $\mathcal{M}$.

Exercise 11.1. Prove that ν as defined by equations (2) and (3) are signed measures.

In this section we assume henceforth that ν is a signed measure on $\mathcal{M}$.

Definition 11.3. A set $E \in \mathcal{M}$ is called positive for ν if for any subset $A \subset E$ which belongs to $\mathcal{M}$ we have $\nu(A) \geq 0$. Similarly we define a set E to be negative for ν if $\nu(A) \leq 0$ for all $A \subset E$, and null if $\nu(A) = 0$ for all $A \subset E$.

Exercise 11.2. Prove that if $\{B_j\}_{j\in\mathbb{N}}$ is a sequence of negative sets for ν, then the union $\cup_{j=1}^{\infty} B_j$ is also negative for ν. (The corresponding result for positive sets is of-course also true.)

Exercise 11.3. Prove that if $E \subset F$ and $|\nu(F)| < \infty$, then also $|\nu(E)| < \infty$.*(Hint: Use that $\nu(F) = \nu(E) + \nu(F \setminus E)$.)*

Theorem 11.4 (Hahn Decomposition)**.** *There is a decomposition $X = E^+ \cup E^-$ where $E^+ \cap E^- = \emptyset$ and E^+ and E^- are positive and negative for ν respectively. Furthermore these are unique up to null-sets.*

Proof. The uniqueness up to null sets is easy to prove and left to the reader. We assume without loss that ν takes values in $(-\infty, \infty]$. Define

$$\beta = \inf\{\nu(E) : E \subset X \text{ is negative for } \nu\}.$$

First of all note that $-\infty < \beta < \infty$ (this is why we assume that ν takes values in $(-\infty, \infty]$). Then there is some sequence $\{B_j\}_{j\in\mathbb{N}}$ of negative sets such that $\nu(B_j) \to \beta$ as $n \to \infty$. The union $B = \cup_{j=1}^{\infty} B_j$ is itself also a negative set, and we have $\nu(B) \leq \nu(B_j)$ for each j. Hence $\nu(B) = \beta$.

We shall now prove that $A = B^c$ is a positive set, which will complete the proof, since we may then put $E^+ = A$ and $E^- = B$. If this set is not positive then by definition there is some subset $E_0 \subset A$ with $\nu(E_0) < 0$. But E_0 can not be a negative set, because then $\nu(B \cup E_0) < \beta$ with $B \cup E_0$ a negative set, which contradicts the definition of β. Therefore E_0 contains a set E_1 such that $\nu(E_1) > 0$. We let m_1 be the smallest positive integer such that there is such a set with $\nu(E_1) \geq 1/m_1$. Since $|\nu(E_0)| < \infty$ we have that $|\nu(E_1)| < \infty$ (by the previous exercise). Since

$$\nu(E_0 \setminus E_1) = \nu(E_0) - \nu(E_1) \leq \nu(E_0) - \frac{1}{m_1} < 0,$$

we may apply the same argument to $E_0 \setminus E_1$ and choose a smallest integer m_2 for which there is a set $E_2 \subset E_0 \setminus E_1$ and $\nu(E_2) \geq 1/m_2$. We may continue this construction to get a sequence of numbers m_k and subsets $E_k \subset E_0 \setminus \cup_{j=1}^{k-1} E_j$ with $\nu(E_k) \geq 1/m_k$. By construction these sets are in particular mutually disjoint, and hence

$$\nu(\cup_{j=1}^{\infty} E_j) = \sum_{j=1}^{\infty} \nu(E_j) < \infty.$$

Since $\nu(E_j) \geq 1/m_j$ this in particular means that $1/m_j \to 0$ as $j \to \infty$. But then it follows that for any measurable subset $F \subset F_0 = E_0 \setminus \cup_{j=1}^{\infty} E_j$ we have

$$\nu(F) \leq \frac{1}{m_k - 1} \to 0 \text{ as } k \to \infty.$$

Hence $\nu(F) \leq 0$ for all such F, and therefore F_0 is a negative set with

$$\nu(F_0) = \nu(E_0) - \sum_{j=1}^{\infty} \nu(E_k) < \nu(E_0) < 0.$$

As for E_0 above we see that this gives a contradiction to the definition of β and B. □

Definition 11.5. Suppose that both μ and ν are signed measures on $\mathcal{M}$, then we say that μ and ν are mutually singular, written $\mu \perp \nu$, if $X = E \cup F$ where $E \cap F = \emptyset$ and E and F are null-sets for μ and ν respectively.

Theorem 11.6 (Jordan Decomposition). *There are unique positive measures ν^+ and ν^- on $\mathcal{M}$ such that $\nu = \nu^+ - \nu^-$ and $\nu^+ \perp \nu^-$.*

Indeed it is easy to see that

$$\nu^+ = \nu|_{E^+} \text{ and } \nu^- = -\nu|_{E^-},$$

where $E^{\pm}$ is a Hahn decomposition for ν. We also introduce the total variation

$$|\nu| = \nu^+ + \nu^-,$$

which is a positive measure on $\mathcal{M}$. Note that $|\nu(E)| \leq |\nu|(E)$, but in general we do not have equality.

Many concepts for a signed measure will by definition be done through the measures ν^+, ν^- and $|\nu|$:

- $\mathcal{L}^1(\nu) = \mathcal{L}^1(|\nu|)$,
- for $f \in \mathcal{L}^1(\nu)$ we define $\int f d\nu = \int f d\nu^+ - \int f d\nu^-$,
- We say that ν is (σ-)finite if $|\nu|$ is (σ-)finite.

Exercise 11.4. Prove that $\nu^{\pm}$ as given above indeed satisfies the statement of the Jordan decomposition theorem. Also prove that in case

$$\nu = \nu_1 - \nu_2,$$

where ν_1, ν_2 are positive measures, then $\nu^+ \leq \nu_1$ and $\nu^- \leq \nu_2$.

Exercise 11.5. Let ν be a signed measure on $(X, \mathcal{M})$. Prove the following for every $B \in \mathcal{M}$:

(a) $\nu^+(B) = \sup\{\nu(A) : A \in \mathcal{M}, A \subset B\}$,
(b) $|\nu|(B) = \sup\{\nu(A_1) - \nu(A_2) : A_1, A_2 \in \mathcal{M}, B = A_1 \cup A_2, A_1 \cap A_2 = \emptyset\}$,
(c) If ν is σ-finite then $|\nu|(B) = \sup\{\int_B f d\nu : f \in \mathcal{L}^1(|\nu|), |f| \leq 1\}$.

Definition 11.7. Suppose that μ and ν are signed measures on $\mathcal{M}$. We say that ν is absolutely continuous with respect to μ, written $\nu \ll \mu$ if $\nu(E) = 0$ for any E such that $|\mu|(E) = 0$.

Exercise 11.6. Prove that $\nu \ll |\nu|$ and that $\nu \ll \mu$ if and only if $|\nu| \ll |\mu|$.

Exercise 11.7 (*)**.** Prove that in case μ and ν are signed measures on $\mathcal{M}$ such that $\nu \ll \mu$ and such that $|\nu|(E) < \infty$ for all sets E with $|\mu|(E) < \infty$, then for any $\varepsilon > 0$ there is $\delta > 0$ such that $|\nu|(E) < \varepsilon$ for any $E \in \mathcal{M}$ with $|\mu|(E) < \delta$. *(Hint: Assume this is not true, then there is $\varepsilon > 0$ and a sequence E_n s.t. $|\mu|(E_n) < 1/2^n$ and $|\nu|(E_n) \geq \varepsilon$. Let $E = \cap_{n=1}^{\infty}(\cup_{j=n}^{\infty} E_j)$, and show that $|\mu|(E) = 0$ but $|\nu|(E) \geq \varepsilon$.)*

Exercise 11.8. Prove that in case $f \in \mathcal{L}^1(\mu)$ where μ is a measure, then the measure ν defined by

$$\nu(E) = \int_E f d\mu = \int f \chi_E d\mu.$$

is a signed measure which is absolutely continuous with respect to μ.

11.2. **The Radon-Nikodym Theorem and Lebesgue Decomposition.**

Theorem 11.8 (Radon-Nikodym)**.** *If μ is a (positive) $\sigma-$finite measure on $\mathcal{M}$ and ν is a signed $\sigma-$finite measure on $\mathcal{M}$ such that $\nu \ll \mu$, then there is a (a.e. unique) measurable function $f = f^+ - f^-$ such that at least one of the functions $f^\pm$ is integrable with respect to μ and*

$$\nu(E) = \int_E f d\mu \quad \textit{for all } E \in \mathcal{M}.$$

Remark 11.9. *The reason for the fact that we can not take $f \in \mathcal{L}^1(\mu)$ is simply because we may have $|\nu|(X) = \infty$. The integral should of-course be interpreted as $\pm\infty$ in those cases that correspond to sets E with $\nu(E) = \pm\infty$. Note also that there is either a lower or an upper bound on ν, since either $\nu^+(E)$ or $\nu^-(E)$ is finite, and we have $-\nu^-(E) \leq \nu(E) \leq \nu^+(E)$ for all E. That is why we can take one of the functions f^+ or f^- integrable.*

Proof. We first reduce the problem to the case when μ, ν are finite. Let X be the union of the disjoint measurable sets X_j with finite μ- and $|\nu|$-measure, which is easily seen to be possible. Suppose the theorem holds for each subset X_n so that there are functions f_n such that

$$\nu(E \cap X_n) = \int_{E \cap X_n} f_n d\mu$$

holds for all measurable sets $E \subset X$. Then if we define $f(x) = f_n(x)$ if $x \in X_n$, which makes sense since the sets are disjoint, then we get for every set E with $|\nu|(E) < \infty$:

$$\sum_{n=1}^{\infty} \int_{E \cap X_n} |f| d\mu = \sum_{n=1}^{\infty} |\nu|(E \cap X_n) = |\nu|(E) < \infty.$$

Hence it is easy to see that f is μ-integrable on E, and that f has the desired properties. We furthermore may reduce to the case that ν is positive, since we may otherwise treat ν^+ and ν^- separately.

As for the uniqueness part, if f, g are two different functions satisfying the theorem, then

$$\int_E g d\mu = \int_E f d\mu$$

holds for every E, and this implies that the functions are equal a.e. (Simply choose for instance $E = \{x \in X : f(x) - g(x) \geq 1/n\}$ to see that it must have measure zero ...).

Now we shall prove the existence in the case of finite positive measures which finishes the proof. To this end let Φ denote the set of all positive functions g for which

$$\int_E g d\mu \leq \nu(E) \quad \text{for all } E \in \mathcal{M}.$$

Let furthermore

$$\alpha = \sup\{\int g d\mu : g \in \Phi\}.$$

Then there is a sequence g_n in Φ such that $\int g_n d\mu \to \alpha$ as $n \to \infty$. Let $h_n = \vee_{j=1}^n g_j$. If we for a fixed n define the sets $E_1 = \{x \in X : h_n(x) = g_1(x)\}$ and then inductively $E_j = \{x \in X : h_n(x) = g_j(x)\} \setminus \cup_{i=1}^{j-1} E_j$, then by construction the sets E_j are disjoint, and for any E we get

$$\int_E h_n d\mu = \sum_{j=1}^n \int_{E \cap E_j} g_j d\mu \leq \sum_{j=1}^n \nu(E \cap E_j) = \nu(E).$$

Therefore we see that also each $h_n \in \Phi$ and this is an increasing sequence. Hence $h_n \nearrow h$, and by the monotone convergence theorem $h \in \Phi$ with

$$\int h d\mu = \alpha.$$

Since $h \in \mathcal{L}^1(\mu)$ we may assume that it is finite-valued everywhere. If we now consider the measure η defined by

$$\eta(E) = \nu(E) - \int_E h d\mu,$$

then by construction this is a positive measure, and the proof will be finished if we prove that $\eta = 0$. Suppose this is not the case. Define the signed measures

$$\eta_n = \eta - \frac{1}{n}\mu.$$

Let $A_n \cup B_n$ be a Hahn decomposition of X w.r.t. η_n, and put

$$A_0 = \cup_{n=1}^\infty A_n, \quad B_0 = \cap_{n=1}^\infty B_n.$$

Since $B_0 \subset B_n$ for each n we have

$$0 \leq \eta(B_0) \leq \frac{1}{n}\mu(B_0) \to 0 \text{ as } n \to \infty.$$

Hence $\eta(B_0) = 0$, and since $\eta \neq 0$ and $A_0 \cap B_0 = \emptyset$, $A_0 \cup B_0 = X$ we must have $\eta(A_0) > 0$. Therefore, since η is absolutely continuous with respect to μ we must also have $\mu(A_0) > 0$, and by continuity from below for measures this implies that $\mu(A_n) > 0$ for some n. Hence

$$\frac{1}{n}\mu(E \cap A_n) \leq \eta(E \cap A_n) = \nu(E \cap A_n) - \int_{E \cap A_n} f d\mu.$$

If we now define $g = h + \chi_{A_n}/n$, then we see that $g \in \Phi$, but this contradicts the definition of α since

$$\int g d\mu = \int h d\mu + \mu(A_n)/n > \alpha.$$

□

The function f is usually called the Radon-Nikodym derivative of ν with respect to μ and written

$$\frac{d\nu}{d\mu} = f.$$

Exercise 11.9. Prove that in case μ and ν are positive measures on $\mathcal{M}$ and $\nu \leq \mu$, then we have

$$0 \leq \frac{d\nu}{d\mu} \leq 1 \text{ a.e.}$$

Theorem 11.10 (Lebesgue decomposition)**.** *Let ν be a $\sigma-$finite signed measure on $\mathcal{M}$ and μ a $\sigma-$finite positive measure on $\mathcal{M}$. Then there are unique $\sigma-$finite signed measures ρ, λ on $\mathcal{M}$ such that*

$$\lambda \perp \mu, \quad \rho \ll \mu \text{ and } \nu = \lambda + \rho.$$

Proof. As in the proof of the Radon-Nikodym theorem this is reduced to the case of finite positive measures, and the uniqueness is easy. Since $\nu \ll \mu + \nu$, and even $\nu \leq \mu + \nu$ there is some measurable function $f : X \to [0, 1]$ such that

$$\nu(E) = \int_E f d(\mu + \nu),$$

for all $E \in \mathcal{M}$.

Let

$$A = \{x \in X : f(x) = 1\}, \quad B = X \setminus A.$$

Then

$$\nu(A) = \int_A d\mu + \int_A d\nu = \mu(A) + \nu(A).$$

Since $\nu(A) < \infty$ it follows that $\mu(A) = 0$. Now define $\lambda(E) = \nu(E \cap A)$, and $\rho(E) = \nu(E) - \lambda(E) = \nu(E \cap B)$. Then it only remains to prove that $\rho \ll \mu$. So suppose $\mu(E) = 0$, then we have

$$\int_{E \cap B} d\nu = \int_{E \cap B} f d\nu,$$

or what amounts to the same thing

$$\int_{E \cap B} (1 - f) d\nu = 0.$$

Since $1 - f > 0$ in B it follows that $\nu(E \cap B) = \rho(E) = 0$. □

11.3. **Differentiation with respect to a doubling measure in a metric space.** In this section we shall work in the context of a metric space (X, ρ). We assume that μ is a (non-zero) Borel measure on X. (The reader who feels uncomfortable with this level of generality may below assume that $X = \mathbb{R}^n$ and that μ is Lebesgue measure.) We will use the notation

$$B(x, r) = \{y \in X : \rho(x, y) < r\}, \quad S(x, r) = \{y \in X : \rho(x, y) = r\}.$$

Note in particular that $B(x, r)$ is an open ball and $S(x, r)$ a closed sphere. Also note that $\partial B(x, r) \subset S(x, r)$, but there are cases where this is not an equality. For instance if we look at $X = \mathbb{N}$ with the usual distance, then $B(1, 1) = \{1\}$ so that the boundary is empty, but $S(1, 1) = \{2\}$. First we need a covering theorem, which is a geometric rather than a measure-theoretic result. The result is true in any separable metric space.

Theorem 11.11 (Vitali covering theorem)**.** *Let (X, ρ) be a separable metric space, and let*

$$\{B(x_j, r_j) : j \in J\}$$

be an arbitrary collection of balls in X such that

$$R = \sup\{r_j : j \in J\} < \infty.$$

Then there exists a (at most) countable subcollection

$$\{B(x_j, r_j) : j \in J'\}, \quad J' \subset J$$

of balls from the original collection which are disjoint and satisfy

$$\bigcup_{j \in J} B(x_j, r_j) \subset \bigcup_{j \in J'} B(x_j, 5r_j).$$

Proof. For $j = 1, 2, 3, \ldots$ let

$$J_i = \left\{ j \in J : \frac{R}{2^i} < 2r_j \leq \frac{R}{2^{i-1}} \right\}.$$

Let J'_1 be an arbitrary maximal subcollection of pairwise disjoint elements from the family

$$\{B(x_j, r_j) : j \in J_1\}.$$

Clearly this must be a countable set since X is separable. Assume now that we have chosen $J'_1, J'_2, \ldots, J'_{i-1}$, then we let J'_i be a maximal subcollection of pairwise disjoint sets in

$$\left\{ j \in J_i : B(x_j, r_j) \cap B(x_k, r_k) = \emptyset \text{ for all } k \in \bigcup_{n=1}^{i-1} J'_n \right\}.$$

By construction there is for each $j \in J_i$ some $k \in \bigcup_{n=1}^{i} J'_n$ such that $B(x_k, r_k) \cap B(x_j, r_j) \neq \emptyset$. Since

$$2r_j \leq \frac{R}{2^{j-1}} = 2\frac{R}{2^j} \leq 4r_k,$$

we see that $B(x_j, r_j) \subset B(x_k, 5r_k)$. Hence if we put $J' = \bigcup_{n=1}^{\infty} J'_n$ the theorem follows. □

Definition 11.12. We say that the measure μ satisfies a doubling condition if there is a constant C_μ such that for all $x \in X$ and all $r > 0$

$$0 < \mu(B(x, 2r)) \leq C_\mu \mu(B(x, r)) < \infty.$$

The constant C_μ is called a doubling constant for μ.

In particular this is satisfied for Lebesgue measure in $\mathbb{R}^N$ with the constant $C_\mu = 2^N$. Since the proofs of the theorems in this section is essentially unchanged (if we add some simplifying assumptions) for the more general case we have decided to treat it here, because it is of importance in many recent developments of analysis in metric spaces, where such doubling conditions are often of crucial importance.

Remark 11.13. *It should be remarked, although outside the scope of these notes, that not every metric space can have a (non-zero) doubling measure on it. Indeed, for complete spaces, such a measure exists if and only if the metric space is doubling, which by definition means that there is a constant C' such that any ball of radius $2r$ in X can be covered by C' balls of radius r.*

We also assume the following for simplicity:

For the rest of this section (X, ρ) is a metric space, and μ a doubling measure with doubling constant C_μ. We will also assume that X is separable, and $\mu(S(x, r)) = 0$ for all x, r.

The first assumption is actually redundant in this situation, because the condition that all balls have strictly positive and finite measure will force X to be separable. The second assumption may however fail even if there is a doubling measure on X, but most interesting examples would satisfy this.

We introduce the notation

$$A_r f(x) = \frac{1}{\mu(B(x, r))} \int_{B(x,r)} f d\mu$$

for averages over balls.

Exercise 11.10 (**)**.** Prove that for any $f \in \mathcal{L}^1(\mu)$ the function $g(x, r) = A_r f(x)$ is (jointly) continuous in (x, r). *(Hint: Prove first that if $x_n \to x$ and $r_n \to r$ then $\mu(B(x_n, r_n)) \to \mu(B(x, r))$. Note also that we use $\mu(S(x, r)) = 0$ here.)*

Exercise 11.11 (**)**.** Prove that if X is a separable metric space, and μ is a Borel measure on X for which every open ball has finite measure, then for any $f \in \mathcal{L}^1(\mu)$ and $\varepsilon > 0$ there is an integrable continuous function g on X such that $\int |f - g| d\mu < \varepsilon$. *(Hint: Consider the class $\mathcal{H} \subset C(X)$ of continuous functions on X with bounded support (i.e. which are zero outside some ball), then these forms a vector lattice. The integral $I_\mu g = \int g d\mu$ for these is an elementary integral. If we start from $(X, \mathcal{H}, I_\mu)$ as our integration space, then we get a space $\mathcal{L}$, which contains $\mathcal{H}$. Prove that for any open ball $B \subset X$ we have $\chi_B \in \mathcal{L}^+$, and use this to prove that indeed $\mathcal{L}^1(\mu) \subset \mathcal{L}$.)*

Definition 11.14. The function

$$Mf(x) = \sup_{r>0} A_r |f|(x)$$

is called the Hardy-Littlewood maximal function.

Theorem 11.15. *There is a constant C_M such that for all $a > 0$*

$$\mu(\{x \in X : Mf(x) > a\}) \leq \frac{C_M}{a} \int |f| d\mu.$$

Proof. Let $E_a = \{x \in X : Mf(x) > a\}$. For each $x \in E_a$ we can choose r_x such that $A_{r_x}|f|(x) > a$ by continuity of the function $A_r|f|(x)$. By the Vitali covering theorem there is a sequence of points $x_1, x_2, x_3, \ldots$ such that the balls $B(x_1, r_{x_1}), B(x_2, r_{x_2}), B(x_3, r_{x_3}), \ldots$ are disjoint, and such that $B(x_1, 5r_{x_1}), B(x_2, 5r_{x_2}), B(x_3, 5r_{x_3}), \dots$ covers E_a. Hence

$$\begin{aligned}\mu(E_a) &\leq \sum_{n=1}^{\infty} \mu(B(x_n, 5r_{x_n})) \\ &\leq \sum_{n=1}^{\infty} \mu(B(x_n, 8r_{x_n})) \\ &\leq C_\mu^3 \sum_{n=1}^{\infty} \mu(B(x_n, r_{x_n})) \\ &\leq C_\mu^3 a^{-1} \sum_{j=1}^{\infty} \int_{B(x_j, r_{x_j})} |f| d\mu \\ &\leq C_\mu^3 a^{-1} \int |f| d\mu.\end{aligned}$$

□

Exercise 11.12. Prove that if g is continuous on X, then $\lim_{r \to 0^+} A_r g(x) = g(x)$.

Theorem 11.16 (Lebesgue differentiation theorem)**.** *Suppose $f \in \mathcal{L}^1(\mu)$, then for almost every x we have*

$$\lim_{r \to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f d\mu = f(x).$$

So the theorem says in particular that the Radon-Nikodym derivative of

$$\nu(E) = \int_E f d\mu$$

with respect to μ is given by

$$\lim_{r \to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} f d\mu.$$

Proof. Let $\varepsilon > 0$. We may choose a continuous function g such that

$$\int |f - g| d\mu < \varepsilon.$$

Then

$$\begin{aligned}&\limsup_{r \to 0} |A_r f(x) - f(x)| \\ &= \limsup_{r \to 0} |A_r(f-g)(x) + (A_r g - g)(x) + (g - f)(x)| \\ &\leq M(f-g)(x) + 0 + |f - g|(x).\end{aligned}$$

So if we put

$$E_a = \{x \in X : \limsup_{r \to 0} |A_r f(x) - f(x)| > a\}, \quad F_a = \{x \in X : |f(x) - g(x)| > a\},$$

then

$$E_a \subset F_{a/2} \cup \{x \in X : M(f-g)(x) > a/2\}.$$

But

$$\frac{a}{2} \mu(F_{a/2}) \leq \int_{F_{a/2}} |f - g| d\mu < \varepsilon,$$

so by Theorem 11.15 we have

$$\mu(E_a) \leq \frac{2\varepsilon}{a} + \frac{2C_M \varepsilon}{a}.$$

Since $\varepsilon > 0$ is arbitrary this implies that $\mu(E_a) = 0$ for all $a > 0$. But for all $x \notin \cup_{n=1}^{\infty} E_{1/n}$ we have $\lim_{r \to 0} A_r f(x) = f(x)$ and hence the theorem is proved. □

Corollary 11.17. *Suppose $f \in \mathcal{L}^1(\mu)$. Then for almost every $x \in X$ we have*
$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| d\mu(y) = 0.$$

The set of points for which $\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f - f(x)| d\mu = 0$ is called the **Lebesgue set** of f.

Proof. For each rational number q put $g_q(y) = |f(y) - q|$. Then according to the above theorem we have
$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} g_q(y) d\mu(y) = g_q(x)$$
on a set D_q such that $\mu(D_q^c) = 0$. This is the same as saying that
$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - q| d\mu(y) = |f(x) - q| \text{ for all } x \in D_q.$$
Let $E = \cap_{q\in\mathbb{Q}} D_q$ (i.e. $E^c = \cup_{q\in\mathbb{Q}} D_q^c$ which has measure zero). Then we may for any $x \in E$ and $\varepsilon > 0$ choose $q \in Q$ such that $|f(x) - q| < \varepsilon$, and hence $|f(y) - f(x)| \le |f(y) - q| + \varepsilon$. Therefore
$$\lim_{r\to 0} \frac{1}{\mu(B(x,r))} \int_{B(x,r)} |f(y) - f(x)| d\mu(y) < |f(x) - q| + \varepsilon < 2\varepsilon.$$
Since $\varepsilon > 0$ is arbitrary the statement follows. $\square$

11.4. **The one-dimensional case.** Here we will simply state without proof some results about the one-dimensional case, when $X \subset \mathbb{R}$ equiped with Lebesgue measure m.

Definition 11.18 (Functions of bounded variation on the real line)**.** Suppose $F : [a,b] \to \mathbb{R}$, then we define the total variation $T_F : [a,b] \to [0,\infty]$ of F over $[a,b]$:
$$T_F(x) = \sup\left\{\sum_{i=1}^{k} |F(x_j) - F(x_{j-1})| : a = x_0 < x_1 < \ldots < x_k = x\right\}.$$
If $T_F(b)$ is finite, then F is said to be of bounded variation on $[a,b]$, written $F \in BV([a,b])$. We also define similarly for $F : \mathbb{R} \to \mathbb{R}$ the total variation $T_F : \mathbb{R} \to \mathbb{R}$ by
$$T_F(x) = \sup\left\{\sum_{i=1}^{k} |F(x_j) - F(x_{j-1})| : -\infty < x_0 < x_1 < \ldots < x_k = x\right\},$$
and if
$$T_F(\infty) := \lim_{x\to\infty} T_F(x) < \infty$$
then we say that $F \in BV(\mathbb{R})$.

Exercise 11.13. If $F \in BV([a,b])$ (or $BV(\mathbb{R})$) then we may write
$$F = (T_F + F)/2 - (T_F - F)/2.$$
Prove that the functions $(T_F + F)/2$ and $(T_F - F)/2$ are both increasing.

Lemma 11.19. *If $F : [a,b] \to \mathbb{R}$ is increasing and we define $G(x) = \lim_{y\to x^+} F(y)$, then G is increasing and right continuous. Furthermore:*
(a) *The set of points where F and G are discontinuous is at most countable,*
(b) *F and G are differentiable a.e. and $F' = G'$ a.e.*

We also introduce the space $NBV(\mathbb{R})$ to denote the set of all functions in $BV(\mathbb{R})$ which are right continuous with $\lim_{x\to-\infty} F(x) = 0$.

Then we have the following characterization of finite signed measures on $\mathbb{R}$.

Theorem 11.20. *If μ_F is a finite signed measure on the Borel sets in $\mathbb{R}$ and we define $F(x) = \mu_F((-\infty, x])$, then $F \in NBV(\mathbb{R})$.*
Conversely if $F \in NBV(\mathbb{R})$, then there is a unique measure μ_F such that
$$F(x) = \mu_F((-\infty, x]).$$

Definition 11.21 (Absolutely continuous functions)**.** A function $f : [a, b] \to \mathbb{R}$ is said to be absolutely continuous if for any $\varepsilon > 0$ there is $\delta > 0$ such that
$$\sum_{i=1}^{k} |f(b_i) - f(a_i)| < \varepsilon$$
whenever $\{(a_i, b_i)\}$ is a finite collection of disjoint subintervals of $[a, b]$ with $\sum_{i=1}^{k} |b_i - a_i| < \delta$.

One can fairly easily prove that any absolutely continuous function is of bounded variation, and that it can be decomposed as a difference of two increasing functions which are also absolutely continuous. Finally we have the following version of the fundamental theorem of calculus for the Lebesgue integral:

Theorem 11.22. *For $F : [a, b] \to \mathbb{R}$ the following are equivalent:*
(a) *F is absolutely continuous on $[a, b]$.*
(b) *$F(x) - F(a) = \int_a^x f(t)dt$ for some $f \in \mathcal{L}^1(m|_{[a,b]})$.*
(c) *F is differentiable a.e.on $[a, b]$, $F' \in \mathcal{L}^1(m|_{[a,b]})$ and $F(x) - F(a) = \int_a^x F'(t)dt$.*

Remark 11.23. *As a side remark not quite related to differentiation is that a bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is continuous m-a.e.*

12. L^p-SPACES(*)

12.1. Measurable functions.

As usual we let $(X, \mathcal{M}, \mu)$ denote a fixed measure space ($\mu \geq 0$). In this section we will by a measurable function f mean an extended real valued function f defined on $X \setminus A$ where $\mu(A) = 0$ and f is measurable as a function on $X \setminus A$. We also will write $f = g$ to mean that they are equal almost everywhere with respect to μ on X (i.e. everything "a.e." in this section). Note in particular that with these conventions the integrable functions forms a vector space.

Definition 12.1 (L^p-norm)**.** For $1 \leq p < \infty$ we define for a measurable function f
$$||f||_p = \left(\int |f|^p d\mu\right)^{1/p}.$$
We also define
$$||f||_\infty = \inf\{\alpha \in [0, \infty] : \mu(\{x : |f(x)| > \alpha\}) = 0\}.$$
This is usually called the **essential supremum** of $|f|$, and it means that a.e. $|f| \leq ||f||_\infty$.

Remark 12.2. *The fact that $\|\cdot\|_\infty$ is not defined through an integral, but the rest of the L^p-norms are indicates that they behave somewhat differently. For instance in a metric space setting one can prove that continuous functions are dense in $L^p(\mu)$ for $p \in [1, \infty)$, but if for instance the measure of every open set is non-zero, then the continuous functions forms a closed subspace of $L^p(\mu)$ and the L^∞-norm restricted to these is the same as the supremum-norm (simply because if we change a continuous function on a set of measure zero it can not be continuous any more).*

Definition 12.3 (L^p-space)**.** We define for $1 \leq p \leq \infty$:
$$\mathcal{L}^p(\mu) = \{f : ||f||_p < \infty\}.$$
The space $L^p(\mu)$ is the set of equivalence classes of functions equal μ-a.e.

The reason for this definition is simply that the "norm" $\|\cdot\|$ is only a semi-norm on $\mathcal{L}^p$ since it does not distinguish between functions which are equal almost everywhere. We however always think of elements in $L^p(\mu)$ as functions defined almost everywhere. We then have that $||\cdot||_p$ is a norm on $L^p(\mu)$ for each p, i.e.:

(a) $||f||_p \geq 0$ with equality if and only if $f = 0$ a.e.,
(b) $||kf||_p = |k|||f||_p$ for all $k \in \mathbb{R}$ and $f \in L^p(\mu)$,
(c) $||f+g||_p \leq ||f||_p + ||g||_p$ for all $f, g \in L^p(\mu)$.

In particular we see that $L^p(\mu)$ is a vector space. The triangle inequality is in this setting usually called Minkowski's inequality, and will be discussed below.

Exercise 12.1 (*)**.** Prove that $\|\cdot\|_p$ satisfies (a) and (b) above.

For a given $p \in [1,\infty]$ we introduce the **dual exponent** $q \in [1,\infty]$ (also commonly denoted p') such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

In case $p = 1$ then we interpret this as that $q = \infty$ and vice versa. Note that p is then the dual exponent of q. **Below p, q will always denote the dual exponents to each other.**

Exercise 12.2 (*)**.** Suppose $1 < p < \infty$. Prove that

$$t^{1/p} \leq \frac{1}{p}t + (1 - 1/p) \text{ for all } t \in (0,\infty).$$

(Hint: look at the derivative of $f(t) = t^{1/p} - t/p - (1 - 1/p)$.)

Theorem 12.4 (Hölder's inequality)**.** *If $1 \leq p \leq \infty$ and $f \in L^p(\mu)$, $g \in L^q(\mu)$ then $fg \in L^1(\mu)$ and*

$$||fg||_1 \leq ||f||_p||g||_q.$$

Proof. The case $p = 1$ and $p = \infty$ are easy and left to the reader. We now assume that $1 < p < \infty$. Also the inequality is trivially satisfied unless $\|f\|_p > 0$ and $\|g\|_q > 0$, which we henceforth assume.

Note that of $a, b \geq 0$ then

$$\|(af)(bg)\|_1 = ab\|fg\|_1, \quad \|af\|_p\|bg\|_q = ab\|f\|_p\|g\|_q$$

and hence we may assume that

$$\|f\|_p = \|g\|_q = 1.$$

Now we re-write the function $|f(x)g(x)|$ as follows:

$$|f(x)g(x)| = (|f(x)|^p)^{1/p}(|g(x)|^q)^{1/q} = (|f(x)|^p)^{1/p}(|g(x)|^q)^{1-1/p} = \left(\frac{|f(x)|^p}{|g(x)|^q}\right)^{1/p}|g(x)|^q.$$

Now we invoke the inequality from the above exercise with $t = |f(x)|^p/|g(x)|^q$ to get

$$\left(\frac{|f(x)|^p}{|g(x)|^q}\right)^{1/p}|g(x)|^q \leq \left(\frac{1}{p}\left(\frac{|f(x)|^p}{|g(x)|^q}\right) + \frac{1}{q}\right)|g(x)|^q = \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q.$$

Hence

$$\int |fg|d\mu \leq \frac{1}{p}\int |f|^p d\mu + \frac{1}{q}\int |g|^q d\mu = \frac{1}{p} + \frac{1}{q} = 1.$$

□

Exercise 12.3. Prove the $p = 1$-case of Hölder's inequality.

Theorem 12.5 (Minkowski's inequality)**.** *If $1 \leq p \leq \infty$ and $f, g \in L^p(\mu)$ then*

$$||f+g||_p \leq ||f||_p + ||g||_p.$$

Proof. The result is trivial if $\|f+g\|_p = 0$, so we assume that it is strictly positive. Note that

$$|f+g|^p \leq (|f|+|g|)|f+g|^{p-1},$$

so by Hölder's inequality and the equality $(p-1)q = p$ we get

$$\begin{aligned}
&\int |f+g|^p d\mu \leq \int |f||f+g|^{p-1} d\mu + \int |g||f+g|^{p-1} d\mu \\
&\leq \|f\|_p \left\||f+g|^{p-1}\right\|_q + \|g\|_p \left\||f+g|^{p-1}\right\|_q \\
&= (\|f\|_p + \|g\|_p) \left(\int |f+g|^p\right)^{1/q}.
\end{aligned}$$

Hence

$$\|f+g\|_p = \left(\int |f+g|^p\right)^{1-1/q} \leq \|f\|_p + \|g\|_p.$$

□

Theorem 12.6 (Riesz-Fischer)**.** *The space $L^p(\mu)$ is a Banach space.*

Recall that this simply means that $L^p(\mu)$ as a metric space is complete. I.e. if f_k is a sequence in $L^p(\mu)$ such that

$$||f_n - f_k||_p \to 0 \text{ as } n, k \to \infty,$$

then there is a function $f \in L^p(\mu)$ such that

$$||f - f_k||_p \to 0 \text{ as } k \to \infty.$$

Proof. Suppose f_k is a Cauchy sequence in $L^p(\mu)$. Suppose we prove that f_k is then also Cauchy in measure, then we know (according to Exercise 9.4) that there is some subsequence f_{k_j} and some function f such that $f_{k_j} \to f$ a.e. We may then apply Fatou's lemma to conclude that

$$\int |f|^p d\mu \leq \liminf_{j\to\infty} \int |f_{k_j}|^p d\mu.$$

The right hand side must be finite since $\|f_{k_j}\|_p \leq \|f_{k_j} - f_l\|_p + \|f_l\|_p$ holds for all l, and since the first expression on the right hand side is assumed to go towards zero as k_j, l goes to infinity there must be a bound on the integrals. Applying Fatou's lemma again gives us

$$\int |f - f_k|^p \leq \liminf_{j\to\infty} \int |f_{k_j} - f_k|^p.$$

Since by asumption there is for any $\varepsilon > 0$ an $N \in \mathbb{N}$ such that

$$\int |f_{k_j} - f_k|^p d\mu < \varepsilon \text{ for all } k_j, k \geq N,$$

we see that indeed

$$\int |f - f_k|^p d\mu \leq \varepsilon \text{ for all } k \geq N.$$

So it only remains to prove that f_k is Cauchy in measure. To do so let for $\varepsilon > 0$

$$E^{\varepsilon}_{j,k} = \{x \in X : |f_k(x) - f_j(x)|^p \geq \varepsilon\} = \{x \in X : |f_k(x) - f_j(x)| \geq \varepsilon^{1/p}\}.$$

Then since $|f_k - f_j|^p \geq \varepsilon \chi_{E^{\varepsilon}_{j,k}}$ we must have $\mu(E^{\varepsilon}_{j,k}) \to 0$ as $k, j \to \infty$, which finishes the proof. □

12.2. **Linear functionals on L^p.** By Hölders inequality we have that for any $g \in L^q(\mu)$ the linear functional

$$F_g(f) := \int fg d\mu$$

satisfies

$$||F_g|| = \inf\{C : |F_g(f)| \leq C||f||_p \quad \text{for all } f \in L^p(\mu)\} \leq ||g||_q.$$

We actually have $||F_g|| = ||g||_q$ (see Exercise 12.5).

Exercise 12.4. Prove that if a measurable function f on X belongs to $L^p(\mu)$ $(1 \leq p < \infty)$ then

$$\|f\|_p = \sup_{g \in L^q(\mu), \|g\|_q = 1} \left| \int f g d\mu \right|.$$

Prove also that in case X is σ-finite, then the result holds also for $p = \infty$. *(Hint: consider the function $g = \operatorname{sgn}(f)|f|^{p-1}\|f\|_p^{-p/q}$ where the sign function $\operatorname{sgn}(x)$ is 0 if $x = 0$ and $x/|x|$ otherwise.)*

In case X is σ-finite then indeed a measurable function f belongs to $L^p(\mu)$ if and only if

$$\sup_{g \in L^q(\mu), \|g\|_q = 1} \left| \int f g d\mu \right| < \infty.$$

Exercise 12.5. Prove that F_g is a continuous linear functional on $L^p(\mu)$ for each $g \in L^q(\mu)$, and prove that $\|F_g\| = \|g\|_q$ (in case $q = \infty$ we assume that X is σ-finite for the last equality).

For σ-finite spaces we also have a converse:

Theorem 12.7. *Suppose $(X, \mathcal{M}, \mu)$ is σ-finite, and $1 \leq p < \infty$. Then every continuous linear functional F on $L^p(\mu)$ is of the form F_g for some unique element $g \in L^q(\mu)$.*

Note that the result above is false for $p = \infty$.

Exercise 12.6. Prove that in case Theorem 12.7 holds for every finite measure space $(X, \mathcal{M}, \mu)$ then it follows that it also holds for every σ-finite X. *(Hint: Write X as a countable increasing union of measurable sets X_n with finite measure. Let g_n be the corresponding function from the theorem defined on X_n. Show that if $n \leq m$ then $g_n = g_m$ in X_n, so that we may define $g = g_n$ on X_n, which gives a measurable function on X. Then apply the monotone convergence theorem a couple of times to draw the necessary conclusions.)*

Proof. The uniqueness is clear, and by the above exercise what remains is to prove the existence of g for a given functional F in case $\mu(X) < \infty$. Let

$$\nu(E) = F(\chi_E).$$

This is then well defined and by linearity and continuity it is easy to see that it indeed is a signed (finite) measure on X (note that if $E_1 \subset E_2 \subset E_3 \subset \cdots$ and $E = \cup_{i=1}^{\infty} E_i$, then $\chi_{E_n} \to \chi_E$ in $L^p(\mu)$ by monotone convergence for any $p \in [1, \infty)$ but typically NOT for $p = \infty$). Furthermore by definition of the space L^p we must have that $\mu(E) = 0$ gives $F(\chi_E) = 0$ (since F must give the same value for all functions equal a.e.). Therefore ν is absolutely continuous with respect to μ, and hence of the form

$$\nu(E) = \int_E g d\mu.$$

Note that by linearity we must have

$$F(h) = \int g h d\mu \tag{4}$$

for all integrable simple functions on X. Suppose now that h is a bounded measurable function. Then there is a sequence of simple functions h_n converging pointwise to h such that $|h_n| \leq \|h\|_\infty$ and therefore we may apply the dominated convergence theorem to conclude that

$$\int h_n g d\mu \to \int h g d\mu,$$

and by our assumptions , since F is continuous, $F(h_n) \to F(h)$ (because again by dominated convergence h_n converges to h in $L^p(\mu)$). Therefore equation (4) holds for all bounded measurable functions h. Now for a given $f \in L^p(\mu)$ let

$$f_n(x) = |f(x)| \operatorname{sgn}(g(x)) \chi_{\{x \in X : |f(x)| \leq n\}}.$$

Then $\|f_n\|_p \leq \|f\|_p$, and therefore

$$|F(f_n)| \leq \|F\|\|f_n\|_p \leq \|F\|\|f\|_p.$$

Since $f_n g \geq 0$ and converges pointwise to $|fg|$ we may apply Fatou's lemma and we get

$$\int |fg| d\mu \leq \liminf_{n\to\infty} \int f_n g d\mu = \liminf_{n\to\infty} F(f_n) \leq \|F\|\|f\|_p.$$

If we use the result of Exercise 12.4 we see that this implies that indeed $g \in L^q(\mu)$. To finish the proof it is enough to note that for any $h \in L^p(\mu)$ there is a sequence of bounded functions $h_n \in L^p(\mu)$ which converges to h in $L^p(\mu)$, and by Hölder's inequality this implies that indeed

$$\|h_n g - hg\|_1 \leq \|h_n - h\|_p \|g\|_q,$$

which shows that indeed we have

$$F(h) = \int hg d\mu \text{ for all } h \in L^p(\mu).$$

□

Exercise 12.7. Let $1 \leq p_1 < p_2 \leq \infty$. Give an example of a function f on $(0, \infty)$ with Lebesgue measure such that $f \in L^p(\mu)$ if and only if $p_1 < p < p_2$. *(Hint: Consider a function of the form $f(x) = x^{-a} |\log x|^b$ for suitable a, b.)*

13. Continuous functions on a compact space(**)

Here we will look at a special case of the Riesz representation theorem. We will only look at continuous functions on compact metric spaces, and just as for Stone's theorem this can be generalized. There are versions for locally compact Hausdorff spaces as-well.

So we will now let (K, ρ) denote a compact metric space (e.g. a closed bounded subset of $\mathbb{R}^N$). Given a function $f : K \to \mathbb{R}$ we define the supremum norm :

$$\|f\|_s = \sup\{|f(x)| : x \in K\}.$$

Exercise 13.1 (**). Prove that $\|\cdot\|_s$ is a norm on $C(K)$ and that $f_n \to f$ in this norm is the same as uniform convergence.

Lemma 13.1. *If (K, ρ) is a compact metric space, then it is separable (i.e. it contains a countable dense subset).*

Proof. To see this, for each n the balls $\{B(x, 1/n)\}_{x \in K}$ obviously is an open cover of K, and hence it has a finite subcover $B(x_1^n, 1/n), \dots, B(x_{m_n}^n, 1/n)$. The collection of all x_j^n is countable, and if we relabel them in a suitable way to $x_1, x_2, \dots$ it is easy to see that we have a countable dense subset of K. □

Lemma 13.2. *If (K, ρ) is a compact metric space, then it is second countable (i.e. there is a countable collection of balls $B(x_n, r_n)$ such that any open set is a union of elements from this collection).*

Proof. If O is open, then by definition this means that for every $x \in O$ there is a ball $B(x, r) \subset O$. Now we may choose one of our points x_n from the previous lemma and a rational number q_n such that $|x - x_n| < q_n < r/3$. Then it follows that $x \in B(x_n, q_n) \subset B(x, r) \subset O$. Hence the countable collection $B(x_n, q)$ with $x_1, x_2, x_3, \dots$ countable and dense and q rational satisfies the statement. □

Theorem 13.3 (Dini)**.** *Suppose f_n is a sequence of functions in $C(K)$ such that $f_n \searrow 0$ pointwise on K. Then the convergence is uniform (i.e. $\|f_n\|_s \searrow 0$).*

Proof. Given $\varepsilon > 0$, what we need to prove is that there is N such that $f_n < \varepsilon$ for all $n \geq N$. To do so look at the sets $B_n = \{x \in K : f_n(x) < \varepsilon\}$. By continuity these sets are open and clearly cover K, since for any $x \in K$ there is an N such that $f_N(x) < \varepsilon$ by assumption. Hence for any $\varepsilon > 0$ there is a finite subcover of K by such sets. However, we also have since f_n is decreasing that the sets B_n are increasing. Hence there must be some N such that $K \subset B_n$ for all $n \geq N$. This is exactly the statement that $f_n < \varepsilon$ for all $n \geq N$ as was to be proved. □

Theorem 13.4 (Riesz)**.** *Suppose $I : C(K) \to \mathbb{R}$ is positive and linear, then I is automatically order-continuous and there is a unique finite Borel measure μ such that*

$$I(f) = \int f d\mu \quad \textit{for all } f \in C(K).$$

What we say in the theorem is that if I is linear and $If \geq 0$ for all $f \geq 0$ in $C(K)$, then we automatically have for any sequence $f_n \searrow 0$ in $C(K)$ that $If_n \to 0$. Hence I is an elementary integral on $C(K)$, which is an elementary lattice as noted earlier.

Proof. To prove that I is order-continuous let $f_n \searrow 0$ in $C(K)$. By Dini's theorem for any $\varepsilon > 0$ there is N such that $f_n \leq \varepsilon$ for all $n \geq N$. Note that positivity and linearity gives that I is monotone in the sense that if $f \leq g$ then $If \leq Ig$ (since we may look at $g - f$ which is positive). Hence

$$If_n \leq I\varepsilon = \varepsilon I1 \quad \text{for all } n \geq N.$$

Since $\varepsilon > 0$ is arbitrary we see that we have $If_n \to 0$.

By Stone's theorem this means that $I = I_\mu$ for some measure μ, which is finite since $\mu(K) = I1$. To prove that this is a Borel measure (i.e. that all Borel sets are measurable), then it is enough to prove that open sets are $\mu-$measurable (because we know that the $\mu-$measurable sets forms a $\sigma-$algebra, and hence if this $\sigma-$algebra contains the open sets it must also by definition contain all Borel sets). But we saw above that any open set O is a countable union of open balls. So it is by the same type of argument enough to prove that all open balls are $\mu-$measurable. But given an open ball $B(x, r)$ then the functions f_n defined below can easily be proved to belong to $C(K)$:

$$f_n(y) := \begin{cases} 1 & |y - x| \leq r - 1/n \\ n(r - |y - x|) & r - 1/n < |y - x| \leq r \\ 0 & |y - x| > r \end{cases}.$$

These functions satisfies $f_n \nearrow \chi_{B(x,r)}$. So we have that $\chi_{B(x,r)} \in \mathcal{L}^+$ as defined before. Hence $B(x, r)$ is $\mu-$measurable by definition and the proof is done. □

Since any continuous linear functional F on $C(K)$ can be written on the form $F = I^+ - I^-$ we can prove the following result using the above:

Theorem 13.5 (Riesz)**.** *Given a continuous linear functional F on $C(K)$ there is a unique finite signed Borel measure μ such that*

$$F(f) = \int f d\mu$$

holds for all $f \in C(K)$. Furthermore μ is positive if and only if $F(f) \geq 0$ for every positive f in $C(K)$. Also

$$\|F\| = \|\mu\| = |\mu|(K).$$

That is, the continuous linear functionals on $C(K)$ and the finite signed Borel measures are in $1 - 1$ correspondence.

Exercise 13.2 (**)**.** Prove Theorem 13.5.

14. Recommended textbooks

The main recommended extra source for the course is:

- R. Bass: Real Analysis for Graduate Students. *A book that is modern and inexpensive (cheap to buy and also downloadable from the authors web page). It however works only from the Lebesgue point of view.*

14.1. Lebesgue-type of treatment.

- G.B. Folland: Real Analysis. *This text has been used as the course text for integration theory at MAI for a long time and is highly recommended. It contains much more than just integration theory, and this course would roughly correspond to chapter 1-3+6.1-6.2. Unfortunately it is unreasonably expensive theses days, and therefore I have decided to switch to the above book, which for this course seems to be just as good.*
- W. Rudin: Real and Complex Analysis. *Again this is a very much used text world-wide, and contains a lot of interesting material apart from integration theory. However it relies very much on the Riesz representation theorem rather than outer measures to show existence of for instance Lebesgue measure.*
- A. Friedman: Foundations of Modern Analysis. *This is a very good, short and inexpensive book, which takes the reader quickly to the main points. Indeed this is the book I learned integration theory from to begin with. It also contains a lot of other material on functional analysis, and is highly recommended.It does not treat the Riesz representation theorem though.*
- T. Tao: An Introduction to Measure Theory. *Although I am not so familiar with this book it seems like a very good text for self-study with much motivation. It is also freely downloadable via the authors home page.*

14.2. Daniell-type treatment (or both).

- G.S. Shilov and B.E. Gurevich: Integral, Measure and Derivative. *A very nice text which works from the Daniell point of view.*
- A.E. Taylor: General Theory of Functions and Integration. *A book which covers both approaches to integration, and perhaps the one that would be most suitable if you want a book that covers essentially everything in these notes.*
- H. Federer: Geometric Measure Theory. *A very comprehensive book, but quite challenging to read. It treats both alternatives to integration theory (in chapter 2). This is the standard reference work also for deeper material on "lower-dimensional" measures, covering theorems, differentiation..., but I would not regard it as very suitable for this course since it is not easy!*