LECTURE NOTES FOR THE COURSE

MATH 245A: GRADUATE ANALYSIS

FALL QUARTER 2014

BY MAREK BISKUP

Note: Preliminary version, comments welcome!
# CONTENTS

1. A brief history of integral ................................................................. 1  
   1.1. Newton’s integral ................................................................. 1  
   1.2. Cauchy’s integral ................................................................. 2  
   1.3. Riemann’s integral ............................................................... 3  
   1.4. Stieltjes’ integral ............................................................... 5  
   1.5. Stieltjes and beyond: Itô-Stratonovich and Young integrals .......... 7  
2. Peano-Jordan content ................................................................. 10  
   2.1. Elementary sets and (pre)content ........................................... 10  
   2.2. Outer/inner content, measurability ....................................... 13  
   2.3. Non-measurable sets, relation to topology ............................... 17  
   2.4. Connection with Riemann integral ........................................ 18  
3. Lebesgue measure ..................................................................... 20  
   3.1. Lebesgue outer measure ........................................................ 20  
   3.2. Lebesgue measurable sets .................................................... 24  
   3.3. Lebesgue measure ............................................................... 27  
   3.4. Convergence Theorems ........................................................ 29  
   3.5. Nonmeasurable sets ............................................................ 31  
4. Lebesgue integral: bounded case .................................................. 33  
   4.1. Bounded functions on finite measure spaces ............................. 33  
   4.2. Integrability vs measurability .............................................. 37  
   4.3. Bounded Convergence Theorem .............................................. 41  
   4.4. Characterization of Riemann integrability .............................. 43  
5. Lebesgue integral: general case ................................................... 47  
   5.1. Unsigned integral .............................................................. 47  
   5.2. Signed integral and convergence theorems .............................. 50  
   5.3. Absolute integrability and $L^1$-space ................................. 53  
   5.4. Uniform integrability ........................................................ 55  
6. Abstract measure theory ............................................................ 59  
   6.1. From semialgebras to outer measures .................................... 59  
   6.2. Carathéodory measurability and extension theorem ................... 62  
   6.3. Uniqueness of extension ..................................................... 66  
   6.4. Lebesgue-Stieltjes measures .............................................. 69  
7. More on outer measures ............................................................. 73  
   7.1. Regular outer measures ...................................................... 73  
   7.2. Borel regular and Radon measures ....................................... 75  
   7.3. Metric outer measures ...................................................... 77  
   7.4. Hausdorff measure and dimension ....................................... 80  
   7.5. Frostman’s lemma ............................................................. 85
1. A BRIEF HISTORY OF INTEGRAL

The MATH 245A analysis course is by and large devoted to Lebesgue’s theory of measure and integration. This theory stands on the foundations laid over 200 years by many great minds working in mathematics. In order to appreciate these better, we begin by a brief recount of the history of integral and the rather winding road that led to Lebesgue’s ultimate formulation.

In what follows, I will freely invoke various technical terms such as limit, continuity, derivative, etc, pertinent to differential calculus. These are introduced and discussed in detail in various undergraduate analysis courses (e.g., MATH 131).

1.1 Newton’s integral.

At the time of Newton (and Leibniz), integral was understood in the sense of an “inverse” to derivative. The interpretation in terms of “area under graph of function” was known but was regarded as secondary. We may thus take Newton’s definition of integral to be:

Definition 1.1 (Newton’s integral) Let \( f: [a,b] \rightarrow \mathbb{R} \) and let \( F \) be such that \( F'(x) = f(x) \) for all \( x \in [a,b] \). Then

\[
\int_a^b f(x) \, dx := F(b) - F(a)
\] (1.1)

Newton’s integral thus required the function to have an antiderivative (a.k.a. the primitive function). The definition needs to be supplied by the following observation that, we note, uses only differential calculus:

Lemma 1.2 If \( F: [a,b] \rightarrow \mathbb{R} \) is such that \( F'(x) = 0 \) for all \( x \in [a,b] \), then \( F \) is constant.

Indeed, if there are two antiderivatives for \( f(x) \), then their difference has zero derivative and so must be constant. The lemma is proved by way of:

Theorem 1.3 (Mean-Value Theorem) Let \( f: [a,b] \rightarrow \mathbb{R} \) be continuous on \([a,b]\) and such that \( f'(x) \) exists for all \( x \in (a,b) \). Then there is \( c \in (a,b) \) such that

\[
\frac{f(b) - f(a)}{b - a} = f'(c).
\] (1.2)

It is worth noting that, since we require no regularity of \( f' \), the Mean-Value Theorem requires some facts from point-set topology that were certainly not available at the time of Newton. Indeed, a linear transformation reduces the statement to the case \( f(b) = f(a) \), which goes by the name Rolles’ Theorem. This is in turn proved by using that a continuous function on a compact interval achieves its maximum and minimum in the interior for which we need the Bolzano-Weierstrass Theorem (stating that every bounded sequence of reals has a converging subsequence). A simple argument underlying the first-derivative criterion then gives the result.

An attractive feature of Newton’s definition is that, once the integral is well defined, we immediately get both basic formulations of the Fundamental Theorem of Calculus:

Fundamental Theorem of Calculus I: The derivative of the integral with respect to the upper limit is the integrand,

\[
\frac{d}{dx} \int_a^x f(t) \, dt = f(x)
\] (1.3)
Fundamental Theorem of Calculus II: The integral of the derivative of a function equals the function itself,

$$\int_{a}^{x} F'(t) \, dt = F(x) \quad (1.4)$$

A major shortcoming of Newton’s integral is that its definition is non-constructive. In particular, there is no direct way to decide whether a given function admit antiderivatives. This is, in a sense, the key issue in all definitions that followed Newton’s.

1.2 Cauchy’s integral.

In his 1821 monograph, Cauchy put forward a definition of integral that is directly based on the interpretation of “area under graph of function.” To state this precisely, we need some notation. (For the impatient, this notation will carry through a bunch of forthcoming subsections.) Given an interval \([a, b]\), a marked partition \(\Pi\) of an interval \([a, b]\) is a collection of intervals \([[x_{i-1}, x_i] : i = 1, \ldots, n]\) and points \(z_1, \ldots, z_n\) such that

\[ a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b \quad (1.5) \]

and

\[ z_i \in [x_{i-1}, x_i], \quad i = 1, \ldots, n. \quad (1.6) \]

The norm of the partition \(\Pi\) is defined by

\[ \|\Pi\| := \max_{i=1, \ldots, n} |x_i - x_{i-1}|. \quad (1.7) \]

Given a function \(f : [a, b] \rightarrow \mathbb{R}\) and a marked partition \(\Pi\) of \([a, b]\), we then define

\[ R(f, \Pi) := \sum_{i=1}^{n} f(z_i)(x_i - x_{i-1}). \quad (1.8) \]

(Here “\(R\)” is for Riemann sum.) Cauchy then more or less proved:

**Theorem 1.4 (Cauchy’s integral)**  Let \(f : [a, b] \rightarrow \mathbb{R}\) be continuous. Then the limit

$$\int_{a}^{b} f(x) \, dx := \lim_{\|\Pi\| \to 0} R(f, \Pi) \quad (1.9)$$

exists (finitely) in the sense that

$$\lim_{\delta \to 0} \inf_{\Pi : \|\Pi\| < \delta} R(f, \Pi) = \lim_{\delta \to 0} \sup_{\Pi : \|\Pi\| < \delta} R(f, \Pi) \quad (1.10)$$

where the supremum/infimum is over marked partitions \(\Pi\) of \([a, b]\).

For the proof, it suffices to note that since \(f\) is continuous, it is in fact uniformly continuous (again by the aforementioned Bolzano-Weierstrass Theorem). Explicitly, given any \(\varepsilon > 0\) there is \(\delta > 0\) such that

\[ x, y \in [a, b] \quad \& \quad |x - y| < \delta \quad \Rightarrow \quad |f(y) - f(x)| < \varepsilon. \quad (1.11) \]

It follows that, if \(\Pi\) and \(\Pi'\) are two marked partitions with same \(x_1, \ldots, x_n\) but (possibly) different \(z_1, \ldots, z_n\),

\[ |R(f, \Pi) - R(f, \Pi')| \leq \sum_{i=1}^{n} |f(z_i) - f(z'_i)| (x_i - x_{i-1}) < \varepsilon (b - a). \quad (1.12) \]
The right-hand side tends to zero as $\varepsilon \downarrow 0$ and so the choice of the marked points thus plays no role for the limit and so we may as well choose one that makes $R(f, \Pi)$ maximal. This corresponds to $z_i$ being the maximizer of $f$ over $[x_{i-1}, x_i]$. But for such partition the limit exists by monotonicity because refining the partition decreases (maximal) $R(f, \Pi)$. (We will discuss this more in the context of upper and lower integrals in the next subsection.)

Cauchy completed his theory by providing proofs of both versions of the Fundamental Theorem of Calculus. Thanks to him, there was a first constructive definition of the definite integral for which the classic Newton-Leibnitz theory applied. It seemed that harmony was achieved; unfortunately, not for too long.

1.3 Riemann’s integral.

The first half of the 19th century was marked by gradual but steady increase in precision with which mathematicians approached various questions surrounding the theory of functions. Initially, there was not even a consensus about what it meant for a function to be continuous — before Cauchy, continuity meant that the same expression defining $f(x)$ applied to the entire domain of allowed $x$ values. Numerous examples of discontinuous functions (in present-day sense of the word) were being considered and argued about. This permitted to extend Cauchy’s definition to even functions with a finite number of points of discontinuity, in fact, even such points where the integrand is unbounded.

For instance, when $f : [a, c) \cup (c, b) \to \mathbb{R}$ is given with $f$ continuous on both $[a, c)$ and $(c, b]$, we may define

$$\int_a^b f(x) \, dx := \lim_{\varepsilon, \varepsilon' \downarrow 0} \left( \int_a^{c-\varepsilon} f(x) \, dx + \int_{c+\varepsilon'}^b f(x) \, dx \right),$$

provided the double limit exists. The resulting object is referred to as an improper integral. (It is easy to check that, if $f$ is in fact continuous on all of $[a, b]$, then this coincides with the original notion.) There are even more stringent recipes, e.g., given $a < 0 < b$ we set

$$PV \int_a^b \frac{1}{x} \, dx := \lim_{\varepsilon \downarrow 0} \left( \int_a^{-\varepsilon} \frac{1}{x} \, dx + \int_{\varepsilon}^b \frac{1}{x} \, dx \right)$$

where “PV” stands for principal value. (It is easy to check that the integral of $\frac{1}{x}$ fails to exists as an improper integral, yet it does exists in the sense of principal value.) Notwithstanding, all of these attempts carry additional choices that make the concept difficult to use. In particular, there are properties that naturally hold for the ordinary integral but fail for these extensions.

In his habilitation work, Riemann made a significant philosophical leap compared to all earlier attempts. Here is the definition he used:

**Definition 1.5** (Riemann’s integral) Given $f : [a, b] \to \mathbb{R}$, we say that the Riemann integral $\int_a^b f(x) \, dx$ exists if the limit

$$\int_a^b f(x) \, dx := \lim_{\|\Pi\| \to 0} R(f, \Pi),$$

defined in the sense of equality (1.10), exists (finitely). The function $f$ is then said to be Riemann integrable on $[a, b]$. 


The most notable fact is that the existence of an integral is no longer derived from assumptions on regularity of \( f \); instead, it becomes a regularity property itself. To make the work with his concept easier, Riemann supplied a condition for \( f \) to be integrable in his sense:

**Lemma 1.6** A function \( f : [a, b] \to \mathbb{R} \) is Riemann integrable on \([a, b] \) if and only if

\[
\lim_{\|\Pi\| \to 0} \sum_{i=1}^{n} \operatorname{osc}(f, [x_{i-1}, x_i]) (x_i - x_{i-1}) = 0. \tag{1.16}
\]

Here \( \operatorname{osc}(f, A) \) is the oscillation of \( f \) on set \( A \) which is defined by

\[
\operatorname{osc}(f, A) := \sup\{ |f(y) - f(x)| : x, y \in A \} \tag{1.17}
\]

and the limit is over partitions of \([a, b] \) without marked points.

We leave the proof of this fact to homework. Note that this criterion shows almost instantly that all continuous \( f \) are Riemann integrable. Riemann’s definition thus subsumes Cauchy’s completely. We will later see this does not quite apply to Newton’s definition.

Another way how to define the integral is by way already touched upon in the section on Cauchy’s integral. This method is generally attributed to Darboux — and, in the French literature, Darboux’s name is occasionally used instead of Riemann’s in connection with this integral — although there are certainly predecessors to this. The idea is to introduce the upper Darboux sum,

\[
U(f, \Pi) := \sum_{i=1}^{n} \left( \sup_{t \in [x_{i-1}, x_i]} f(t) \right) (x_i - x_{i-1}) \tag{1.18}
\]

and the lower Darboux sum,

\[
L(f, \Pi) := \sum_{i=1}^{n} \left( \inf_{t \in [x_{i-1}, x_i]} f(t) \right) (x_i - x_{i-1}). \tag{1.19}
\]

These sums now depend only on the partition \( \Pi \); no additional marked points are necessary. An encouraging feature of these is the behavior under refinements of the partition. Indeed, consider a partition \( \Pi' \) obtained by inserting a point (or points) into \( \Pi \). Then

\[
U(f, \Pi) \geq U(f, \Pi') \quad \text{while} \quad L(f, \Pi) \leq L(f, \Pi'). \tag{1.20}
\]

It follows that the limit \( \|\Pi\| \to 0 \) can be simulated by taking infimum, resp., supremum of these quantities. This leads to the definition of upper and lower Riemann/Darboux integrals,

\[
\int_{a}^{b} f(x) \, dx := \sup_{\Pi} L(f, \Pi) \quad \text{and} \quad \int_{a}^{b} f(x) \, dx := \inf_{\Pi} U(f, \Pi). \tag{1.21}
\]

A key point to note is then:

**Lemma 1.7** \( f : [a, b] \to \mathbb{R} \) is Riemann integrable on \([a, b] \) if and only if

\[
\int_{a}^{b} f(x) \, dx = \int_{a}^{b} f(x) \, dx \tag{1.22}
\]

The common value of these integrals then coincides with \( \int_{a}^{b} f(x) \, dx \).
Given its generality, it is no surprise that Riemann’s definition includes functions that had heretofore seemed inaccessible. To give an example, pick a non-increasing sequence \( \{ \alpha_n \} \) with \( a_n \geq 0 \) and consider an enumeration of rationals into a sequence, \( \mathbb{Q} = \{ q_i : i \geq 1 \} \). Define

\[
f(x) := \begin{cases} 
\alpha_n, & \text{if } x = q_n \text{ for some } n \geq 1, \\
0, & \text{otherwise.}
\end{cases}
\]  

(1.23)

Since \( \{ \alpha_n \} \) is non-increasing, \( \text{osc}(f, [x_{i-1}, x_i]) > \alpha_n \) only for at most \( n \) intervals in the partition. Hence, the sum in (1.16) is less than \( n \alpha_n + \alpha_n (b - a) \). Thus, if \( n \alpha_n \to 0 \), this function is Riemann integrable. But once \( a_n > 0 \) for all \( n \), this function is also discontinuous at all rationals! Moreover, it is not hard to check that

\[
\int_a^b f(x) \, dx = 0
\]

(1.24)

for all \( a < b \). In particular, the Fundamental Theorem of Calculus I fails at rational \( x \).

The previous example shows that Riemann’s integral is strictly more general than Cauchy’s. To see that it does not even subsume Newton’s integral, we state:

**Lemma 1.8 (Volterra’s example)** There exists a function \( F : [a, b] \to \mathbb{R} \) with \( F'(x) \) defined and bounded at all \( x \in [a, b] \) and yet \( F' \) not Riemann integrable on \([a, b]\).

Thus, although Riemann’s definition of integral is more general than Cauchy’s, it is perhaps too general to make the FTC I work at all points and yet not general enough to make the FTC II work. This is a quandary that only Lebegue’s notion of measurable function (and identification of those functions that differ on a set of measure zero) was able to resolve.

There are numerous other shortcomings of the Riemann integral that complicate its use. We name just a few:

1. The integral requires the integrated function to be bounded.
2. The integral behaves poorly under pointwise limits: There is a sequence of continuous functions converging pointwise to a bounded function that is not Riemann integrable.

Further shortcomings concern generalizations of the Riemann integral to higher dimensions which should be quite obvious: Partition integration domain into squares and take a limit of the corresponding Riemann sum. Notwithstanding, we then run into:

3. Generalizations to two (or higher) dimensions require additional conditions on the underlying domain of integration. Or, taking this problem back to one dimension, the integral is defined more or less only for intervals.
4. Fubini’s theorem, stating that an integral over a rectangle in \( \mathbb{R}^2 \) can be computed as two successive one-dimensional integrals requires unwieldy conditions. Indeed, even if \( f : \mathbb{R}^2 \to \mathbb{R} \) is Riemann integrable, \( g(x) := f(x, y) \) may not be.

Again, all of these are resolved very elegantly in Lebesgue’s theory.

### 1.4 Stieltjes’ integral.

Before we move to discussing Lebesgue’s theory, it is worth pointing out a generalization of Riemann’ integral, due to Stieltjes. This involves two functions, \( f, g : [a, b] \to \mathbb{R} \) for which we
consider the sum
\[ S(f, dg, \Pi) := \sum_{i=1}^{n} f(z_i) \left( g(x_i) - g(x_{i-1}) \right). \] (1.25)

Obviously, when \( g(x) := x \), this degenerates to \( R(f, \Pi) \).

**Definition 1.9** (Stieltjes’ integral) A function \( f \) is said to be Riemann-Stieltjes integrable with respect to \( g \) on \([a, b]\) if
\[ \int_{a}^{b} f dg := \lim_{\|\Pi\| \to 0} S(f, dg, \Pi) \] (1.26)
exists (in the above sense). The resulting value defines the Riemann-Stieltjes integral.

We note that, although the Stieltjes integral appears quite asymmetric in \( f \) and \( g \), there is actually a lot of symmetry. In fact, we have:

**Lemma 1.10** If \( \int_{a}^{b} f dg \) exists, then so does \( \int_{a}^{b} g df \) and
\[ \int_{a}^{b} f dg + \int_{a}^{b} g df = f(b)g(b) - f(a)g(a). \] (1.27)

Note that this boils down to the integration by parts formula when \( f \) and \( g \) are continuously differentiable. Indeed, we have:

**Lemma 1.11** If \( g \) is continuously differentiable,
\[ \int_{a}^{b} f dg = \int_{a}^{b} f(x)g'(x)dx \] (1.28)
whenever the Riemann integral on the right exists.

We leave both lemmas as an exercise in the homework. The integral has a lot of “standard” properties although the additivity property with respect to domain of integration,
\[ \int_{a}^{b} f dg = \int_{a}^{c} f dg + \int_{c}^{b} f dg \] (1.29)
fails in general. Again, we leave this as an exercise to the reader.

There is a criterion for the existence which is formulated using the notion of (first) variation of \( f \) on \([a, b]\). This is the quantity
\[ V(f, [a, b]) := \sup_{\Pi} \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})|. \] (1.30)
(Again, thanks to the triangle inequality, insertion of points into the partition increases the sum so we take the supremum instead of \( \|\Pi\| \to 0 \).) A function \( f \) is said to be of finite variation on \([a, b]\) if \( V(f, [a, b]) < \infty \). By analyzing the trivial decomposition
\[ f(x) = V(f, [a, x]) - V(f, [a, x]) - f(x), \quad x \in [a, b], \] (1.31)
we obtain the classic theorem due to Jordan that a function is of finite variation on a finite closed interval if and only if it can be written as the difference of two bounded non-decreasing functions. It thus often suffices to develop the integral \( \int_{a}^{b} f dg \) for monotone \( g \) only.

The promised criterion for existence is then:
Lemma 1.12 Suppose $f, g \colon [a, b] \to \mathbb{R}$ with $f$ continuous and $g$ of finite variation. Then the Stieltjes integral $\int_{a}^{b} f \, dg$ exists and, in fact,

$$\left| \int_{a}^{b} f \, dg \right| \leq \left( \sup_{x \in [a, b]} |f(x)| \right) V(g, [a, b])$$

(1.32)

Proof. Similar to Riemann’s integral, a sufficient condition for the limit underpinning the definition of Stieltjes integral to exist is

$$\lim_{\|\Pi\| \to 0} \sum_{i=1}^{n} \text{osc}(f, [x_{i-1}, x_{i}]) |g(x_{i}) - g(x_{i-1})| = 0.$$  

(1.33)

Under the assumption of continuity, and thus uniform continuity, of $f$, given $\varepsilon > 0$ there is $\delta > 0$ such that once $\|\Pi\| < \delta$, we have $\text{osc}(f, [x_{i-1}, x_{i}]) < \varepsilon$. For such partitions, the sum is at most $\varepsilon V(g, [a, b])$. Taking $\varepsilon \downarrow 0$ proves (1.33).

To get the bound on the integral, we note that $S(f, dg, \Pi)$ is itself bounded by the right-hand side of (1.32). This, naturally, survives taking the limit. \qed

1.5 Stieltjes and beyond: Itô-Stratonovich and Young integrals.

The restriction to Stieltjes integrals with respect to functions of finite variation seemed fine initially but, as more and more irregular functions become considered by mathematicians, it became clear that other criteria had to be developed as well. One example when this is necessary is the stochastic or Itô integral. Here $g$ is taken to be a path of a random process called Brownian motion, which is continuous, but nowhere differentiable and of infinite variation. Notwithstanding, if we define the $p$-variation of $g$ by

$$V^{(p)}(g, [a, b]) := \lim_{\|\Pi\| \to 0} \sum_{i=1}^{n} |f(x_{i}) - f(x_{i-1})|^{p},$$

(1.34)

where for $p \geq 1$ the existence of the limit needs to be proved by a separate argument, the Brownian motion has the $p = 2$ (i.e., the second) variation (well-defined) finite and positive over any interval of positive length.

Notwithstanding, in this case the integral cannot be defined in the usual sense. This is seen rather simply from the following observation. Let $L(f, dg, \Pi)$ denote the left endpoint approximation of the integral,

$$L(f, dg, \Pi) = \sum_{i=1}^{n} f(x_{i-1})(g(x_{i}) - g(x_{i-1}))$$

(1.35)

and let

$$R(f, dg, \Pi) = \sum_{i=1}^{n} f(x_{i})(g(x_{i}) - g(x_{i-1}))$$

(1.36)

be the corresponding right endpoint approximation. Then

$$R(f, dg, \Pi) - L(f, dg, \Pi) = \sum_{i=1}^{n} (f(x_{i}) - f(x_{i-1}))(g(x_{i}) - g(x_{i-1}))$$

(1.37)
and so if \( f = g \), then
\[
R(f, df, \Pi) - L(f, df, \Pi) = \sum_{i=1}^{n} (f(x_i) - f(x_{i-1}))^2
\] (1.38)

Assuming that the \( \|\Pi\| \to 0 \) limit of the right-hand side exists, we get that the limits of the right and left endpoint approximations differ exactly by the second variation of \( f \). Under such conditions the Stieltjes integral \( \int_{a}^{b} f df \) cannot exist in the usual sense.

It turns out that the integral using the left-endpoint approximation does exist in the sense of \( L^2 \)-convergence and bears the name \( \text{Itô integral} \). If mid-point approximation is used instead, it is called the \( \text{Stratonovich integral} \). Each of these has its own merits, although the Fundamental Theorem of Calculus works only for the Stratonovich integral.

Let us turn the whole story in the positive direction by noting a simple but useful observation made by Young that gives an easy-to-check criterion for existence of the Stieltjes integral. The main idea is that one can trade an increase in regularity of \( f \) against a decrease in regularity of \( g \). The resulting integral is sometimes called Young’s integral. The set of all \( \alpha \)-Hölder functions on a set \( A \) will be denoted by \( C^\alpha(A) \). Then we have:

**Theorem 1.13** (Young’s integral) Suppose that \( f : [a, b] \to \mathbb{R} \) are such that \( f \in C^\alpha([a, b]) \) and \( g \in C^\beta([a, b]) \) for some \( \alpha, \beta > 0 \) with \( \alpha + \beta > 1 \). Then \( f \) is Riemann-Stieltjes integrable with respect to \( g \) on \([a, b]\) (and vice versa).

The key point of the proof is:

**Lemma 1.14** (Hölder’s inequality) Let \( p, q \in (1, \infty) \) be such that \( \frac{1}{p} + \frac{1}{q} = 1 \). Then for all real numbers \( a_1, \ldots, a_n, b_1, \ldots, b_n \geq 0 \)
\[
\sum_{i=1}^{n} a_i b_i \leq \left( \sum_{i=1}^{n} a_i^p \right)^{1/p} \left( \sum_{i=1}^{n} b_i^q \right)^{1/q}
\] (1.39)

**Proof:** This is an arithmetic-geometric inequality in disguise. A quick way to make the argument is to consider the function \( \varphi(x,y) := xy - x^p/y^q \). It is straightforward to check that, once \( \frac{1}{p} + \frac{1}{q} = 1 \), we have \( \varphi(x,y) \geq 0 \) for all \( x, y \geq 0 \). Assuming that not all \( a_i \)'s and \( b_i \)'s are zero (otherwise the inequality is trivial), we may set
\[
\tilde{a}_i := \frac{a_i}{\left( \sum_{i=1}^{n} a_i^p \right)^{1/p}} \quad \text{and} \quad \tilde{b}_i := \frac{b_i}{\left( \sum_{i=1}^{n} b_i^q \right)^{1/q}}
\] (1.40)

and use the non-negativity of \( \varphi \) to get
\[
\tilde{a}_i \tilde{b}_i \leq \tilde{a}_i^p + \tilde{b}_i^q.
\] (1.41)

Summing this over \( i \) then yields
\[
\sum_{i=1}^{n} \tilde{a}_i \tilde{b}_i \leq \frac{1}{p} \sum_{i=1}^{n} \tilde{a}_i^p + \frac{1}{q} \sum_{i=1}^{n} \tilde{b}_i^q.
\] (1.42)
But the sums on the right are both equal to one and, since $\frac{1}{p} + \frac{1}{q} = 1$, the sum on the left is thus bounded by one. It is now easy to check that this is the desired bound.  

**Proof of Theorem 1.13.** Our goal is to show (1.33). To treat $f$ and $g$ on the same footing, we first bound the sum as

$$
\sum_{i=1}^{n} \text{osc}(f, [x_{i-1}, x_i]) |g(x_i) - x_{i-1}| \leq \sum_{i=1}^{n} \text{osc}(f, [x_{i-1}, x_i]) \text{osc}(g, [x_{i-1}, x_i]) \tag{1.43}
$$

Next we apply Hölder’s inequality to estimate this by

$$
\left( \sum_{i=1}^{n} \text{osc}(f, [x_{i-1}, x_i])^p \right)^{1/p} \left( \sum_{i=1}^{n} \text{osc}(g, [x_{i-1}, x_i])^q \right)^{1/q} \tag{1.44}
$$

for any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$. Thanks to $f \in C^\alpha$ we have $|f(x) - f(y)| \leq K_f |x - y|^\alpha$ and thus

$$
\text{osc}(f, [x_{i-1}, x_i]) \leq K_f |x_i - x_{i-1}|^\alpha \tag{1.45}
$$

and similarly

$$
\text{osc}(g, [x_{i-1}, x_i]) \leq K_g |x_i - x_{i-1}|^\beta \tag{1.46}
$$

The fact that $\alpha, \beta > 0$ with $\alpha + \beta > 1$ permits us to find $p$ and $q$ such that $\frac{1}{p} + \frac{1}{q} = 1$ and yet $\alpha p \geq 1$ and $\beta q \geq 1$. This implies

$$
|x_i - x_{i-1}|^{\alpha p} = |x_i - x_{i-1}| |x_i - x_{i-1}|^{\alpha p - 1} \leq |x_i - x_{i-1}| \|\Pi\|^\alpha \tag{1.47}
$$

and, since $\sum_{i=1}^{n} |x_i - x_{i-1}| = (b - a)$, also

$$
\left( \sum_{i=1}^{n} \text{osc}(f, [x_{i-1}, x_i])^p \right)^{1/p} \leq K_f \left( \sum_{i=1}^{n} |x_i - x_{i-1}|^{\alpha p} \right)^{1/p} \tag{1.48}
$$

$$
\leq K_f \|\Pi\|^{\alpha - 1/p} (b - a)^{1/p}
$$

An analogous estimate (with $\alpha$ and $p$ replaced by $\beta$ and $q$) applies to $g$. It follows that

$$
\sum_{i=1}^{n} \text{osc}(f, [x_{i-1}, x_i]) |g(x_i) - x_{i-1}| \leq K_f K_g \|\Pi\|^{\alpha + \beta - 1} (b - a) \tag{1.49}
$$

and, since $\alpha + \beta > 1$, the criterion (1.33) for existence of the Stieltjes integral is satisfied. (Note that we even get a rate of convergence in terms of $\|\Pi\|$ which can be useful for numerics.)  

Note that the Young criterion for existence of $\int_{a}^{b} f \, df$ requires $f$ to be Hölder with an exponent strictly larger than $1/2$. This is not consistent with the second variation of $f$ being positive and so the Itô/Stratonovitch integrals are not included in Young’s class. This is no surprise since we showed that different choices of the “rule” can lead to different limits.  

We also remark that there are numerous other integrals that further generalize those above (and that even includes the Lebesgue integral that we have not discussed yet). We will not develop these here as they are a bit too special.
2. Peano-Jordan Content

The somewhat dissatisfactory situation with the Riemann integral led Cantor, Peano and particularly Jordan develop a theory of (what we now call) the Peano-Jordan content (a.k.a. just as Jordan content). The underlying motivation is that, since the Riemann integral is after all a means to define the area of a region in $\mathbb{R}^2$ — namely that bounded between one of the axes and the graph of a function — we may as well try to address the notion of area directly from the outset. The usefulness of this philosophical turn of thought becomes even more apparent when higher-dimensional Riemann integrals are considered; there one often wants to integrate $f$ only over a bounded region in, say, $\mathbb{R}^2$ or $\mathbb{R}^3$ and then even integrating function $f = 1$ over such regions leads to problem with defining areas and volumes.

Compared to the Lebesgue measure, the Peano-Jordan content has numerous shortcomings (essentially identical to those of the Riemann integral) and so, for present day mathematics, its relevance rests mostly in the historical role it played at its time. However, the development of Peano-Jordan content will demonstrate many technical aspects of a full-fledged theory of measure and so we may start with it just as well.

2.1 Elementary sets and (pre)content.

The development of any theory of content or measure has essentially two parts:

1. Identify a class of elementary sets for which we agree what their content/measure is.
2. Extend this to a larger class, called measurable sets, by way of approximations from within and without by elementary sets.

Of course, there are numerous aspects one should be concerned about in implementing such a program. For instance, if the class of elementary sets is chosen too large, or the specific meaning of content we agree on is poorly behaved, the approximations considered in (2) might lead to inconsistent answers even for some elementary set. On the other hand, if the class of elementary sets is too small, too few additional sets could be reached by approximations. As we shall see later, these are exactly the concerns that took quite a while to tune out correctly.

For the Peano-Jordan content, our choice of elementary set will be as follows:

**Definition 2.1** (Elementary sets) A half-open box in $\mathbb{R}^d$ is a set of the form

$$I := (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_d, b_d)$$

where $-\infty < a_i < b_i < \infty$ for each $i = 1, \ldots, d$. An elementary set is then any union of the form $\bigcup_{i=1}^n I_i$ where $I_i \in \mathcal{I}$ for all $i = 1, \ldots, n$. We will write $\mathcal{I}$ to denote the set of all half-open boxes including the empty set and $\mathcal{E}$ for the class of elementary sets.

**Lemma 2.2** The class $\mathcal{E}$ is closed under unions, intersections and set differences.

**Proof.** The closure under unions is trivial. For intersections we use that if $I, J$ are non-disjoint half-open boxes, then so is $I \cap J$. For set differences we use that if $I, J$ are half-open boxes, then $J \setminus I$ is either empty or a finite union of half-open boxes. To extend this to elementary sets, we just apply basic set operations. \(\square\)
For a half-open box $I$ as in (2.1) we then define its (pre)content by

$$|I| := \prod_{i=1}^{d} (b_i - a_i).$$

(2.2)

with the proviso $|\emptyset| = 0$. When a set is a union of a finite number of disjoint half-open boxes $\{I_i\}$, our intuition would dictate to set its (pre)content to the sum $\sum_{i=1}^{n} |I_i|$. However, this harbors a consistency problem: We do not know that a different representation as a sum of half-open boxes would result in the same value. This, and a few other issues that are related to this problem, is addressed in:

**Proposition 2.3** (Consistency, subadditivity, additivity) Let $\{I_i: i = 1, \ldots, n\} \subset \mathcal{I}$ and $\{J_j: j = 1, \ldots, m\} \subset \mathcal{I}$ with $I_i \cap I_{i'} = \emptyset$ whenever $i \neq i'$. Then

$$\bigcup_{i=1}^{n} I_i \subseteq \bigcup_{j=1}^{m} J_j \Rightarrow \sum_{i=1}^{n} |I_i| \leq \sum_{j=1}^{m} |J_j|.$$  

(2.3)

In particular, if the unions are equal and $\{J_j: j = 1, \ldots, m\}$ are disjoint, then equality holds.

We begin with an elementary version of the claim:

**Lemma 2.4** Let $\{I_i: i = 1, \ldots, n\} \subset \mathcal{I}$ be disjoint with $I := \bigcup_{i=1}^{n} I_i \in \mathcal{I}$. Then

$$|I| = \sum_{i=1}^{n} |I_i|.$$  

(2.4)

**Proof in $d = 1$**. Suppose $I$ is a half-open interval and $\{I_i: i = 1, \ldots, n\}$ are disjoint half-open intervals such that $I = \bigcup_{i=1}^{n} I_i$. We can always relabel the intervals such that $I_i = (a_i, b_i]$ for

$$-\infty < a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_n < b_n < \infty.$$  

(2.5)

But then the fact that the union of $I_i$’s must be an interval forces $b_i = a_{i+1}$ for all $i = 1, \ldots, n - 1$ along with $a_1 = a$ and $b_n = b$. This and the specific form

$$|(a_i, b_i]| = (b_i - a_i)$$  

(2.6)

implies

$$\sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n} (b_i - a_i) = \sum_{i=1}^{n} b_i - \sum_{i=1}^{n} a_i$$  

$$= b + \sum_{i=1}^{n-1} b_i - \sum_{i=2}^{n} a_i = b - a$$  

(2.7)

and so we get $|I| = \sum_{i=1}^{n} |I_i|$ as desired. \hfill $\Box$

**Proof in $d \geq 2$**. The proof has two steps. First we address the case of (what we will call) product partitions. Then we will use that to handle general partitions as well.

**Step 1** (Product partitions): By a product partition of a half-open box $I$ we will mean any collection of sets of the form $J_{i_1} \times \cdots \times J_{i_d}$ such that

1. $J_{i_j}$ is a half-open interval in $\mathbb{R}$ for all $i = 1, \ldots, d$ and all $j = 1, \ldots, m_i$,
2. $J_{i_j} \cap J_{i_{j'}} = \emptyset$ unless $j = k$, and
(3) their union is all of $I$, i.e.,

$$I = \bigcup_{j_1=1}^{m_1} \cdots \bigcup_{j_d=1}^{m_d} (J_{1,j_1} \times \cdots \times J_{d,j_d}). \quad (2.8)$$

Naturally, the union in (2.8) is disjoint thanks to (2). Observe also that

$$I$$ is necessarily a half-open interval and

$$d$$

conclude that

$$K_i := \bigcup_{j=1}^{m_i} J_{i,j} \quad (2.9)$$

is necessarily an half-open interval and $I = K_1 \times \cdots \times K_d$.

Given such a product partition, we can use the product form of (2.2) and the distributive law for multiplication around summation to get

$$\sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} |J_{1,j_1} \times \cdots \times J_{d,j_d}| = \sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} \left( \prod_{i=1}^{d} |J_{i,j_i}| \right) = \prod_{i=1}^{d} \left( \sum_{j=1}^{m_i} |J_{i,j}| \right) \quad (2.10)$$

From the (already proved) claim in $d = 1$ we know that $\sum_{j=1}^{m_i} |J_{i,j}| = |K_i|$. Relabeling things we conclude that

$$\sum_{j_1=1}^{m_1} \cdots \sum_{j_d=1}^{m_d} |J_{1,j_1} \times \cdots \times J_{d,j_d}| = |I|. \quad (2.11)$$

This verifies the claim in the case of product partitions.

**STEP 2** (General partitions): Let $I \in \mathcal{S}$ and suppose that $I = \bigcup_{i=1}^{n} I_i$ for some $\{I_i\} \subset \mathcal{S}$ with $I_i \cap I_j = \emptyset$ whenever $i \neq j$. Since $I_i$ is itself a product of half-open intervals, for each $k = 1, \ldots, d$ we can enumerate all endpoints of all intervals in the $k$-th coordinate directions that appear in the product constituting $I_i$ for some $i$. Labeling these endpoints increasingly, for each $k = 1, \ldots, d$ we get a sequence

$$a_{k,1} \leq a_{k,2} \leq \cdots \leq a_{k,2n-1} \leq a_{k,2n}. \quad (2.12)$$

We can now use these to define a product partition $\bigcup_{i=1}^{m} J_i$ of $I$, where $m = (2n)^d$ and each $J_j$ is a half-open box of the form

$$(a_{1,i_1}, a_{1,i_1+1}) \times \cdots \times (a_{d,i_d}, a_{d,i_d+1}) \quad (2.13)$$

(which is empty whenever $a_{k,i_k+1} = a_{k,i_k}$ for any $k = 1, \ldots, d$). Then

$$|I| = \sum_{j=1}^{m} |J_j| \quad (2.14)$$

by **STEP 1**.

Thanks to disjointness of $\{I_i\}$, each $J_j$ has a non-empty intersection with at most one $I_i$, and since both $\{I_i\}$ and $\{J_j\}$ are partitions of $I$, we have

$$J_j \cap I_i \neq \emptyset \quad \Rightarrow \quad J_j \subset I_i. \quad (2.15)$$

Setting $S_i := \{ j : J_j \cap I_i \neq \emptyset \}$, it then follows that $\{J_j : j \in S_i\}$ constitutes a product partition of $I_i$, for each $i = 1, \ldots, n$. By **STEP 1** again, we thus have

$$\sum_{j \in S_i} |J_j| = |I_i|, \quad i = 1, \ldots, d, \quad (2.16)$$
and summing this over $i$ with the help of (2.14) then yields
\[
|I| = \sum_{j=1}^{m} |J_j| = \sum_{i=1}^{n} \sum_{j \in S_i} |J_j| = \sum_{i=1}^{n} |I_i| \tag{2.17}
\]
where we used that $S_i \cap S_{i'} = \emptyset$ when $i \neq i'$ and $\bigcup_{i=1}^{n} S_i = \{1, \ldots, m\}$. □

By small modifications of the proof, we also get:

**Corollary 2.5** Let $\{I_i: i = 1, \ldots, n\} \subset \mathcal{I}$ and $I \in \mathcal{I}$. Then
\[
I \subseteq \bigcup_{i=1}^{n} I_i \implies |I| \leq \sum_{i=1}^{n} |I_i|. \tag{2.18}
\]

**Proof.** Repeating the argument in STEP 2 of the previous proof, we can write $\bigcup_{i=1}^{n} I_i$ as $\bigcup_{j=1}^{m} J_j$ where $J_j \in \mathcal{I}$ and $\{J_j\}$ is a product partition arising from collecting all endpoints of all intervals constituting $\{I_i\}$. Given $i \in \{1, \ldots, n\}$, let $S_i$ be as there. Then $I_i = \bigcup_{j \in S_i} J_j$ and thus $|I_i| = \sum_{j \in S_i} |J_j|$. Note also that $I = \bigcup_{j=1}^{m} (J_j \cap I)$ and $\{J_j \cap I\} \subset \mathcal{I}$ is disjoint. The only part of that proof that needs a chance is thus (2.17) which becomes
\[
|I| = \sum_{j=1}^{m} |J_j \cap I| \leq \sum_{j=1}^{m} |J_j| \leq \sum_{i=1}^{n} \sum_{j \in S_i} |J_j| = \sum_{i=1}^{n} |I_i|, \tag{2.19}
\]

where the first inequality comes from the fact that $|J \cap I| \leq |J|$ whenever $I, J \in \mathcal{I}$ — the side-lengths of $I \cap J$ are at most those of $I$ — while the second inequality follows from the fact that for each $j$ there is at least one $i$ such that $j \in S_i$. □

**Proof of Proposition 2.3.** Let us first assume that both $\{I_i\}$ and $\{J_j\}$ are disjoint collections with equal unions. Then $\{I_i \cap J_j: i = 1, \ldots, n, j = 1, \ldots, m\}$ is also a disjoint collection of half-open boxes with the same union. Then
\[
|I_i| = \sum_{j=1}^{m} |I_i \cap J_j| \quad \text{and} \quad |J_j| = \sum_{i=1}^{n} |I_i \cap J_j| \tag{2.20}
\]
by (2.4). Hence
\[
\sum_{i=1}^{n} |I_i| = \sum_{i=1}^{n} \sum_{j=1}^{m} |I_i \cap J_j| = \sum_{j=1}^{m} \sum_{i=1}^{n} |I_i \cap J_j| = \sum_{j=1}^{m} |J_j|. \tag{2.21}
\]
This proves the claim in the case when $\{J_j\}$ are disjoint and the unions are equal. For the case when $\{J_j\}$ are not necessarily disjoint and the union of $\{I_i\}$ is only included in the union of $\{J_j\}$, we use (2.18) instead of (2.4). □

### 2.2 Outer/inner content, measurability.

We are now ready to proclaim
\[
\{I_i\} \subset \mathcal{I} \text{ disjoint} \implies c\left(\bigcup_{i=1}^{n} I_i\right) := \sum_{i=1}^{n} |I_i|. \tag{2.22}
\]
to be the content of any elementary set. (The independence of representation follows from Proposition 2.3.) A trivial consequence of the definition is
\[ E, F \in \mathcal{E} \text{ disjoint} \implies c(E \cup F) = c(E) + c(F). \] (2.23)
In order to extend the notion of content to more general sets, we invoke approximations:

**Definition 2.6** (Inner/outer content, measurability) Let \( A \subset \mathbb{R} \). Then we define:

1. the outer Peano-Jordan content \( c^*(A) \) of \( A \) by
\[
c^*(A) := \inf \left\{ \sum_{i=1}^{n} |I_i| : \{I_i\} \subset \mathcal{I} \text{ AND } \bigcup_{i=1}^{n} I_i \supset A \right\};
\] (2.24)

2. the inner Peano-Jordan content \( c_*(A) \) of \( A \) by
\[
c_*(A) := \sup \left\{ \sum_{i=1}^{n} |I_i| : \{I_i\} \subset \mathcal{I} \text{ disjoint AND } \bigcup_{i=1}^{n} I_i \subset A \right\};
\] (2.25)

3. We say that \( A \) is Peano-Jordan measurable if \( c^*(A) = c_*(A) \). In such a case we call
\[
c(A) := c^*(A) (= c_*(A)) \quad \text{(2.26)}
\]
the Peano-Jordan content of \( A \). Let \( \mathcal{J} \) denote the class of Peano-Jordan measurable sets.

First we observe that the apparent discrepancy (use of “disjoint” collections) between the definition of outer and inner content plays little role:

**Lemma 2.7** Requiring that the families \( \{I_i\} \) are disjoint in the definition of \( c^*(A) \) leads to the same value of the infimum. In particular, for any \( A \subset \mathbb{R}^d \) we have
\[
c_*(A) \leq c^*(A). \quad \text{(2.27)}
\]

**Proof.** By Proposition 2.3, representing the union \( \bigcup_{j=1}^{m} J_j \) as a disjoint union \( \bigcup_{i=1}^{n} I_i \) does not increase the value of the sum of (pre)contents. Once “disjoint” is required also in the definition of \( c^*(A) \), Proposition 2.3 ensures that we get a value no smaller than \( c_*(A) \). \hfill \Box

We are then able to conclude:

**Lemma 2.8** (Elementary sets are measurable) Every \( A \in \mathcal{E} \) is measurable and \( c(A) \) agrees with the expression in (2.22).

**Proof.** Let \( A \) be elementary. Using \( A \) itself in the approximations leading to \( c^*(A) \) and \( c_*(A) \) shows, with the help of the lemma before, that \( c_*(A) = c^*(A) = \text{the value in (2.22)} \). \hfill \Box

Hence, the outer content of \( A \) is thus the least value of content seen in approximations of \( A \) by elementary sets from outside while the inner content of \( A \) is the largest value of content seen in approximations of \( A \) by elementary sets from within, i.e.,
\[
c^*(A) = \inf \{ c(E) : E \in \mathcal{E}, A \subseteq E \}; \quad \text{(2.28)}
\]
and
\[
c_*(A) = \sup \{ c(E) : E \in \mathcal{E}, E \subseteq A \}. \quad \text{(2.29)}
\]
Just as for the concept of Riemann integrability, being Peano-Jordan measurable then signifies that these approximations result in the same value.

Let us now collect some basic consequences of the definitions of outer/inner content. The first one of these is:

**Lemma 2.9 (Positivity and monotonicity)** For any \( A \), both \( c^+(A) \geq 0 \) and \( c_+(A) \geq 0 \). In fact,

\[
A \subset B \implies c^+(A) \leq c^+(B) \quad \text{and} \quad c_+(A) \leq c_+(B).
\]

**(2.30)**

**Proof.** The monotonicity follows immediately. The positivity is then a consequence of \( \emptyset \in \mathcal{I} \), \( c_+(\emptyset) = 0 \) and (2.27). \( \square \)

This has a trivial corollary that does not even need a proof:

**Corollary 2.10 (Null sets are measurable)** If \( A \) is a null set in the sense that \( c_+(A) = 0 \) then \( A \in \mathcal{J} \) and, in fact, \( B \in \mathcal{J} \) for all \( B \subseteq A \).

Another, albeit somewhat less direct, consequence of the definitions is:

**Lemma 2.11 (Sub/superadditivity)** For any sets \( A_1, \ldots, A_n \),

\[
c^\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n c^+(A_i) \quad \text{(subadditivity of} \ c^+)\]

**(2.31)**

Similarly, if \( A_i \cap A_j = \emptyset \) whenever \( i \neq j \), then also

\[
c_\left(\bigcup_{i=1}^n A_i\right) \geq \sum_{i=1}^n c_+(A_i). \quad \text{(superadditivity of} \ c_+)\]

**(2.32)**

**Proof.** We start with (2.31). In light of positivity, we may assume that \( c^+(A_i) < \infty \) for all \( i \); otherwise there is nothing to prove. Let \( \varepsilon > 0 \) and, for each \( i = 1, \ldots, n \) find \( E_i \in \mathcal{E} \) such that

\[
A_i \subset E_i \quad \text{and} \quad c^+(A_i) + \varepsilon/n \geq c(E_i), \quad i = 1, \ldots, n.
\]

**(2.33)**

Since \( \bigcup_{i=1}^n E_i \supseteq \bigcup_{i=1}^n A_i \), from Lemma 2.2, (2.28) and (2.23) we thus get

\[
c^\left(\bigcup_{i=1}^n A_i\right) \leq c\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n c(E_i) \leq \sum_{i=1}^n \left(c^+(A_i) + \varepsilon/n\right) \leq \varepsilon + \sum_{i=1}^n c^+(A_i).
\]

**(2.34)**

But \( \varepsilon \) is arbitrary and so (2.31) follows. The proof of (2.32) is completely analogous. \( \square \)

It turns out that \( A \subseteq \mathbb{R}^d \) is unbounded if and only if \( c^+(A) = \infty \). (It is instructive to write a proof of this fact.) This shows that, rather than for subsets of all of \( \mathbb{R} \), the Peano-Jordan content should be considered for subsets of a (measurable) bounded set only. Here we will find another consequence of the above definitions quite handy:

**Lemma 2.12** For any bounded \( B \in \mathcal{J} \) and any \( A \subseteq B \),

\[
c_+(A) = c(B) - c^+(B \setminus A).
\]

**(2.35)**

In particular,

\[
c^+(A) - c_+(A) = c^+(B \setminus A) - c_+(B \setminus A).
\]

**(2.36)**
Proof. Fix \( \varepsilon > 0 \) and let \( E, F \in \mathcal{E} \) be such that \( E \supseteq B \) (this is where we need that \( B \) is bounded) and \( F \subseteq A \) obey \( c(F) \geq c_*(A) - \varepsilon \). As \( E \setminus F \in \mathcal{E} \) by Lemma 2.2 and \( E \setminus F \supseteq B \setminus A \), from (2.23) we have

\[
c(E) = c(F) + c(E \setminus F) \geq c_*(A) - \varepsilon + c^*(B \setminus A).
\]

(2.37)

Varying \( E \) along a sequence such that \( c(E) \) decreases to \( c^*(B) = c(B) \), we get that \( \leq \) in (2.35).

For the opposite inequality, given \( \varepsilon > 0 \), we pick \( E, F \in \mathcal{E} \) such that \( E \subseteq B \) and \( F \supseteq B \setminus A \) (again using the boundedness of \( B \)) obeys \( c(F) \leq c^*(B \setminus A) + \varepsilon \). Then \( E \setminus F \subseteq A \) and, since \( E \setminus F \in \mathcal{E} \), from (2.23) we get

\[
c(E) = c(F) + c(E \setminus F) \leq c^*(B \setminus A) + \varepsilon + c_*(A).
\]

(2.38)

Taking \( E \) along a sequence such that \( c(E) \) tends to \( c_*(B) = c(B) \), we get \( \geq \) in (2.35). The second conclusion then follows by applying the first one to \( A \) and \( B \setminus A \). \( \square \)

Note that going via complements could be considered as an alternative way to define inner content. (We will see that this is exactly what is done for the Lebesgue measure.) With these properties established, we now get:

**Theorem 2.13** The following holds:

1. If \( A_1, \ldots, A_n \in \mathcal{F} \) are disjoint, then also \( \bigcup_{i=1}^{n} A_i \in \mathcal{F} \) and

\[
c\left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} c(A_i).
\]

(2.39)

2. If \( A, B \) are bounded and measurable, then so is \( A \cap B \) and \( B \setminus A \).

In particular, for any bounded \( D \in \mathcal{F} \), the class \( \{ A \in \mathcal{F} : A \subseteq D \} \) of Peano-Jordan measurable subsets of \( D \) is an algebra (of subsets of \( D \)) and \( c \) is an additive set function on it.

**Proof.** To get (1) we note that, by Lemmas 2.7 and 2.11,

\[
\sum_{i=1}^{n} c_*(A_i) \leq c_*\left( \bigcup_{i=1}^{n} A_i \right) \leq c^*\left( \bigcup_{i=1}^{n} A_i \right) \leq \sum_{i=1}^{n} c^*(A_i).
\]

(2.40)

The left and right-hand sides agree whenever \( A_1, \ldots, A_n \in \mathcal{F} \). The claim then follows from the definition of measurability.

For (2) let \( D \in \mathcal{F} \) be bounded and note that, by Lemma 2.12, \( A \subseteq D \) is measurable if and only if \( D \setminus A \) is measurable. Now, for \( A, B \subseteq D \),

\[
A \cap B = D \setminus \left( (D \setminus A) \cup (D \setminus B) \right)
\]

\[
A \setminus B = A \cap (D \setminus B)
\]

(2.41)

thus show that if \( A, B \) are measurable, then so are \( A \cap B \) and \( A \setminus B \). \( \square \)

**Remark 2.14** We note that all results in this section use only some very basic facts about elementary sets. Namely, we need that \( \mathcal{E} \) is closed under unions, intersections and set differences and the content is well-defined, finite and finitely additive on \( \mathcal{E} \).
2.3 Non-measurable sets, relation to topology.

A natural question to ask is whether there are non-measurable sets — i.e., those that fail to be measurable. The answer to this is quite easy:

**Lemma 2.15** The set \( A := \mathbb{Q} \cap [0, 1] \) is not Peano-Jordan measurable.

*Proof.* The set \( A := \mathbb{Q} \cap [0, 1] \) is totally disconnected and so it contains no non-trivial intervals. Hence, we get \( c_*(A) = 0 \). On the other hand, any elementary set that contains \( A \) must contain \([0, 1]\) and so \( c^*(A) \geq c^*([0, 1]) = 1 \).

A slightly more subtle question is what properties of sets characterize Peano-Jordan measurability. A simple exercise shows:

**Lemma 2.16** Let \( A \subset \mathbb{R}^d \) be bounded. Then \( A \in \mathcal{F} \) if and only if for each \( \varepsilon > 0 \) there are \( E, F \in \mathcal{E} \) with \( F \subseteq A \subseteq E \) and \( c^*(E \setminus F) < \varepsilon \).

Unfortunately, this is still very much in the spirit of the original definition. A more interesting characterization arises when we try to relate measurability to topology. Recall that, for any set \( A \subset \mathbb{R}^d \), \( \overline{A} \) denotes the closure of \( A \) while \( A^o \) denotes the interior of \( A \) with respect to the usual Euclidean metric (and thus topology) on \( \mathbb{R}^d \). The boundary \( \partial A \) of \( A \) is then given by \( \partial A := \overline{A} \setminus A^o \).

**Proposition 2.17** (Topological characterization of PJ-measurability) For any bounded \( A \subset \mathbb{R}^d \),

\[
c^*(A) = c^*(\overline{A}) \quad \text{and} \quad c_*(A) = c_*(A^o)
\]

and so \( A \) is measurable if and only if \( c^*(\overline{A}) = c_*(A^o) \). Hence, if \( A \) is bounded and measurable then so is any \( B \) such that \( A^o \subseteq B \subseteq \overline{A} \) and \( c(B) = c(A) \) for all of these. In particular,

\[
A \in \mathcal{F} \quad \Rightarrow \quad c(A) = c(\overline{A}) = c(A^o).
\]

Finally, for \( A \subset \mathbb{R}^d \) bounded, Peano-Jordan measurability is characterized by

\[
A \in \mathcal{F} \quad \iff \quad c^*(\partial A) = 0.
\]

*Proof.* We begin by proving (2.42). Let \( E \in \mathcal{E} \) be an elementary sets such that \( A \subseteq E \) and \( c^*(A) + \varepsilon \geq c(E) \). Writing \( E \) as the disjoint union \( \bigcup_{i=1}^n I_i \) of half-open boxes \( \{I_i\} \), we now find (by enlarging \( I_i \) slightly in all directions) half-open boxes \( I'_i \) such that

\[
\mathcal{T}_i \subseteq I'_i \quad \text{and} \quad |I'_i| \leq |I_i| + \varepsilon / n.
\]

The fact that \( \overline{A} \) is the smallest closed set containing \( A \) implies

\[
\overline{A} \subseteq \bigcup_{i=1}^n \mathcal{T}_i \subseteq \bigcup_{i=1}^n I'_i = E'
\]

and we thus get

\[
c^*(\overline{A}) \leq \sum_{i=1}^n |I'_i| \leq \sum_{i=1}^n (|I_i| + \varepsilon / n) = \varepsilon + c(E) \leq c^*(A) + 2\varepsilon.
\]

The monotonicity with respect to inclusion then yields \( c^*(\overline{A}) = c^*(A) \). The second part of (2.42) is proved analogously: We find an elementary \( F \subset A \) with \( c_*(A) - \varepsilon \leq c(F) \), write it as a disjoint union of half-open boxes \( \{I_i\} \), shrink these to half-open boxes \( I'_i \subset (I_i)^0 \) with close value of
content, use that \( A^\circ \) is the largest open set contained in \( A \) and then apply a similar calculation as in (2.47) to conclude \( c_*(A^\circ) \geq c_*(A) \). Monotonicity then finishes the claim.

From (2.42) we then immediately get that \( A \in \mathcal{J} \) if and only if \( c^*(\bar{A}) = c_*(A^\circ) \). Then (2.43) follows from the monotonicity of inner/outer content with respect to inclusion. Similarly we get that, once \( A \in \mathcal{J} \), \( A^\circ \subseteq B \subseteq \bar{A} \) implies \( B \in \mathcal{J} \) as well. It thus remains to prove the equivalence (2.44). Lemma 2.16 gives the implication \( \Rightarrow \): If \( E, F \) are as in the lemma, then \( \partial A \subseteq E \setminus F \) and so \( c^*(\partial A) \leq c(E \setminus F) \), which can be made as small as desired. For the opposite implication, \( \Leftarrow \), we use \( c^*(\partial A) = 0 \) to find a non-empty \( F \in \mathcal{E} \) such that \( \partial A \subseteq F^\circ \) and \( c(F) < \varepsilon \). Then \( \bar{A} \setminus F^\circ \subseteq A^\circ \).

But \( \bar{A} \setminus F^\circ \) is closed and so one can find \( E \in \mathcal{E} \) such that

\[
\bar{A} \setminus F^\circ \subseteq E \subseteq A^\circ \quad \text{and} \quad c_*(A) - \varepsilon \geq c(E). \tag{2.48}
\]

But \( E \cup F \in \mathcal{E} \) and \( E \cup F \supseteq \bar{A} \). Therefore,

\[
c^*(\bar{A}) \leq c(E \cup F) \leq c(E) + c(F) < c_*(A^\circ) + 2\varepsilon. \tag{2.49}
\]

Hence, \( c^*(\bar{A}) = c_*(A^\circ) \) and, by the first part of the claim, \( A \in \mathcal{J} \).

\[\Box\]

### 2.4 Connection with Riemann integral.

The fact that measurability of \( A \) implies measurability of both \( \bar{A} \) and \( A^\circ \) seems quite natural: indeed, we usually expect that the closure and interior of a set are more regular than the set itself. However, this conclusion has its limitations:

**Lemma 2.18** There exists a closed set \( A \subset [0, 1] \) which is not Peano-Jordan measurable.

**Proof (sketch).** One example for this would be any fat Cantor set in \([0, 1]\). (The existence of such sets is what underlies Lemma 1.8.) Even more simple is an example constructed as follows: Let \( \{q_n\} \) enumerate \( \mathbb{Q} \) and, given \( \varepsilon > 0 \), let

\[
A := [0, 1] \setminus \bigcup_{n \geq 1} (q_n - \varepsilon 2^{-n}, q_n + \varepsilon 2^{-n}). \tag{2.50}
\]

It is obvious that \( A \) is closed (possibly empty). By truncating the union at a finite \( n \) and replacing open intervals by half-open ones, we readily find \( c^*(A) \geq 1 - 2\varepsilon \). On the other hand, \( A \) is totally disconnected and so \( c_*(A) = 0 \). So for \( \varepsilon < \frac{1}{2} \), \( A \) is a non-empty closed set which is not Peano-Jordan measurable. \( \Box \)

The conclusion that the contents of \( A, \bar{A} \) and \( A^\circ \) are all equal was, at the time of conception of this theory, considered as completely natural. However, later it was found too restrictive and, in fact, a limitation of the theory. Here we will demonstrate this further by stating without proof two lemmas that show that measurability is essentially equivalent to Riemann integrability.

**Lemma 2.19** Let \( A \subset \mathbb{R}^d \) be a bounded set. Then \( A \in \mathcal{J} \) if and only if \( 1_A \), the characteristic function of \( A \), is Riemann integrable.

**Lemma 2.20** Let \( a, b \in \mathbb{R} \) obey \( a < b \) and let \( f : [a, b] \to [0, \infty) \) be a function. Then

\[
f \text{ is Riemann integrable } \iff \{ (x, t) \in [a, b] \times [0, \infty) : f(x) \leq t \} \in \mathcal{J} \tag{2.51}
\]
Underlying these is the fact that the Peano-Jordan content is not countably additive. The obstruction is explained in the next lemma, whose proof we also leave to the reader:

**Lemma 2.21** There are sets $A_1, A_2, \cdots \in \mathcal{F}$ with $A_i \in [0, 1]$ for all $i$ such that $\bigcup_{i=1}^\infty A_i \notin \mathcal{F}$.

(Note, however, that if the union is measurable, and the sets $A_i$ are disjoint, the content is countably additive on this sequence.) The punchline is that, despite all of the conceptual developments, we still have not left the paradigm (and thus have to face also the associated shortcomings) of Riemann integrability. However, as we will see in the next chapter, we are already on a good way to do so.
3. LEBESGUE MEASURE

We can now finally start developing the theory of Lebesgue measure. There are two key improvements compared to Peano-Jordan content: We define the outer measure using countable (as opposed to finite) covers and, abandoning the concept of inner approximations, we define measurability directly from outer measure.

Since we will often talk about infinite series throughout, it is useful to recall the following basic facts: For any sequence \( \{a_i : i \in \mathbb{N}\} \) satisfying \( a_i \geq 0 \) for each \( i \),

\[
\sum_{i=1}^\infty a_i := \lim_{n \to \infty} \sum_{i=1}^n a_i
\]

(3.1)

coinsides with

\[
\sum_{i \in \mathbb{N}} a_i := \sup \left\{ \sum_{i \in A} a_i : A \subset \mathbb{N}, \text{ finite} \right\}.
\]

(3.2)

In particular, the value of the infinite sum is the same for all reorderings of the sequence \( \{a_i\} \). For such \textit{absolutely converging} sums we pretty much get all of the standard properties of finite sums. For instance, if \( a_i, b_i \geq 0 \) and \( \varepsilon \geq 0 \), then

\[
\sum_{i \in \mathbb{N}} (a_i + \varepsilon b_i) = \sum_{i \in \mathbb{N}} a_i + \varepsilon \sum_{i \in \mathbb{N}} b_i.
\]

(3.3)

Similarly,

\[
0 \leq a_i \leq b_i \quad \forall i \in \mathbb{N} \quad \Rightarrow \quad \sum_{i \in \mathbb{N}} a_i \leq \sum_{i \in \mathbb{N}} b_i.
\]

(3.4)

Finally, if \( \{a_{ij} : i, j \in \mathbb{N}\} \) obey \( a_{ij} \geq 0 \), then

\[
\sum_{i, j \in \mathbb{N}} a_{ij} = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} a_{ij}
\]

(3.5)

(This will be later called the discrete Tonelli Theorem.) We leave (easy) verification of these claims to the reader.

3.1 Lebesgue outer measure.

Let us now move to the development of Lebesgue’s theory. Recall that \( \mathcal{I} \) denotes the class of nonempty half-open boxes in \( \mathbb{R}^d \) and that the outer content was defined by

\[
c^*(A) := \inf \left\{ \sum_{i=1}^n |I_i| : \{I_i\} \subset \mathcal{I} \text{ AND } \bigcup_{i=1}^n I_i \supset A \right\}.
\]

(3.6)

The first important insight provided by Lebesgue is to enlarge the covers to countably infinite ones. This leads to:

Definition 3.1 (Lebesgue outer measure) For any \( A \subset \mathbb{R}^d \), the \textit{Lebesgue outer measure} \( \lambda^*(A) \) of \( A \) is defined by

\[
\lambda^*(A) := \inf \left\{ \sum_{i=1}^\infty |I_i| : \{I_i\} \subset \mathcal{I} \text{ AND } \bigcup_{i=1}^\infty I_i \supset A \right\},
\]

(3.7)

The outer measure has the following properties (which, as we will see, can be used to defined outer measures abstractly):
Proposition 3.2 (Outer measure properties) We have:

1. (positivity) $\lambda^*(A) \geq 0$ and $\lambda^*(\emptyset) = 0$.
2. (monotonicity) if $A \subseteq B$, then $\lambda^*(A) \leq \lambda^*(B)$.
3. (countable subadditivity) for any sets $\{A_i : i \in \mathbb{N}\}$,
   \[
   \lambda^* \left( \bigcup_{i \in \mathbb{N}} A_i \right) \leq \sum_{i \in \mathbb{N}} \lambda^*(A_i). \tag{3.8}
   \]

Proof. (1) The real line can be covered by a countable union of unit intervals and so $\lambda^*(A) \geq 0$ for all sets $A$. The claim for $A := \emptyset$ would be trivial if $\emptyset$ were included $\mathcal{F}$. However, since it is not, for any $\varepsilon > 0$ we pick a collection $\{I_i \} \subset \mathcal{F}$ such that $|I_i| \leq \varepsilon 2^{-i}$. Then $\mu^*(\emptyset) \leq \sum_{i=1}^{\infty} |I_i| \leq \varepsilon$ and so we must have $\mu(\emptyset) = 0$.

For (2) we just observe that if $\{I_i \} \subset \mathcal{F}$ covers $A$ then it covers $B$. So the set of covers of $A$ is at least as large as the set of covers of $B$.

To get (3), we may as well assume that $\mu^*(A_i) < \infty$ for all $i$ because otherwise the claim holds trivially. Given $\varepsilon > 0$, for each $i$ we can then find $\{I_{ij} : j \in \mathbb{N}\} \subset \mathcal{F}$ such that

\[
A_i \subseteq \bigcup_{j \in \mathbb{N}} I_{ij} \quad \text{and} \quad \lambda^*(A_i) + \varepsilon 2^{-i} \geq \sum_{j \in \mathbb{N}} |I_{ij}| \tag{3.9}
\]

hold for each $i$. But then

\[
\bigcup_{i \in \mathbb{N}} A_i \subseteq \bigcup_{i,j \in \mathbb{N}} I_{ij} \tag{3.10}
\]

and so, by the definition of $\lambda^*$,

\[
\lambda^* \left( \bigcup_{i \in \mathbb{N}} A_i \right) \leq \sum_{i,j \in \mathbb{N}} |I_{ij}| = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} |I_{ij}|
   \leq \sum_{i \in \mathbb{N}} \left( \lambda^*(A_i) + \varepsilon 2^{-i} \right) = \varepsilon + \sum_{i \in \mathbb{N}} \lambda^*(A_i) \tag{3.11}
\]

where we used (3.3–3.5) and $\sum_{i \in \mathbb{N}} 2^{-i} = 1$ in the intermediate steps. Since this holds for all $\varepsilon > 0$, the claim follows. \qed

A natural question is whether the outer measure in fact gives a better approximation of the “volume” of $A$ from without than the outer content does. This is settled in:

Lemma 3.3 The infimum defining $\lambda^*(A)$ is unchanged if we permit $I_i$ to be empty. In particular, for each $A \subset \mathbb{R}^d$,

\[
\lambda_*(A) \leq c^*(A). \tag{3.12}
\]

Proof. Let $\{I_i : i = 1, \ldots, n\}$ be a cover of $A$ by $I_i \in \mathcal{F} \cup \{\emptyset\}$. Given $\varepsilon > 0$, define $I_i' := I_i$ if $I_i \neq \emptyset$ and $I_i' := (0, \varepsilon 2^{-i})^d$ otherwise. Then $\{I_i' \} \subset \mathcal{F}$ is a cover of $A$ and $\sum_{i \in \mathbb{N}} |I_i'| \leq \varepsilon d + \sum_{i \in \mathbb{N}} |I_i|$. Hence, allowing $I_i$ to be empty does not affect the value of the infimum. The conclusion $\lambda_*(A) \leq c^*(A)$ then follows by comparing (3.6) with (3.7). \qed

The comparison with the inner content is somewhat harder. We will need a lemma:
Lemma 3.4  Let $I \in \mathcal{I}$ and let $\{I_i\} \subset \mathcal{I}$ be such that $I \subset \bigcup_{i \in \mathbb{N}} I_i$. Then
\[ |I| \leq \sum_{i \in \mathbb{N}} |I_i|. \tag{3.13} \]

Proof. In $d = 1$ this may perhaps be proved by reordering the sets monotonically using transfinite induction but an argument based on topology is much easier. We will use that each half-open box contains a closed box, and is contained in an open box, with both of these of nearly the same volume. More precisely, fix $\varepsilon > 0$. Then there is $J \in \mathcal{I}$ such that $J \subset I$ and $|J| \geq |I| - \varepsilon$. Similarly, for each $i \in \mathbb{N}$, there is $I'_i \in \mathcal{I}$ be such that $I_i \subset (I'_i)^K$ and $|I'_i| \leq |I_i| + \varepsilon 2^{-i}$ for each $i$.

A key fact is that $\{(I'_i)^\circ : i \in \mathbb{N}\}$ form an open cover of the compact set $J$ and so, by the Heine-Borel Theorem, there is $n \in \mathbb{N}$ such that
\[ J \subset J \subset \bigcup_{i=1}^n (I'_i)^\circ \subset \bigcup_{i=1}^n I'_i. \tag{3.14} \]

By Corollary 2.5 we thus get
\[ |I| - \varepsilon \leq |J| \leq \sum_{i=1}^n |I'_i| \leq \varepsilon + \sum_{i=1}^n |I_i| \leq \varepsilon + \sum_{i \in \mathbb{N}} |I_i|. \tag{3.15} \]

Since $\varepsilon$ was arbitrary, we are done. \qed

This implies:

Lemma 3.5  For any $A \subset \mathbb{R}^d$,
\[ \lambda^*(A) \geq c_*(A). \tag{3.16} \]

In particular, if $A \in \mathcal{I}$, then $\lambda^*(A) = c(A)$.

Proof. We start by proving the ultimate conclusion for elementary sets: If $E \in \mathcal{E}$, then $\lambda^*(E) = c(E)$. Clearly, $\lambda^*(E) < \infty$ for any $E \in \mathcal{E}$. By Lemma 3.3, for any $E \in \mathcal{E}$ we have $\lambda^*(E) \leq c(E)$.

Now, given $\varepsilon > 0$, let $\{I_i\}$ be a cover of $E$ such that $\lambda^*(E) \geq \sum_{i \in \mathbb{N}} |I_i| - \varepsilon$. Write $E = \bigcup_{j=1}^n J_j$ with with $\{J_j : j \in \mathbb{N}\}$ a cover of $J$ and so, by Lemma 3.4,
\[ |J_j| \leq \sum_{i \in \mathbb{N}} |I_i \cap J_j|. \tag{3.17} \]

On the other hand, $\{I_i \cap J_j : j = 1, \ldots, n\}$ is a disjoint partition of $I_i$ and so, by Lemma 2.4,
\[ |I_i| = \sum_{j=1}^n |I_i \cap J_j|. \tag{3.18} \]

Combining these we get
\[ \lambda^*(E) + \varepsilon \geq \sum_{i \in \mathbb{N}} |I_i| = \sum_{i \in \mathbb{N}} \sum_{j=1}^n |I_i \cap J_j| \]
\[ \quad = \sum_{j=1}^n \sum_{i \in \mathbb{N}} |I_i \cap J_j| \geq \sum_{j=1}^n |J_j| = c(E) \tag{3.19} \]

where we again used (3.5) to write the two sums in a convenient order. Since $\varepsilon$ was arbitrary, we get $\lambda^*(E) = c(E)$ as claimed.
Now recall that \( c_\ast(A) \) is the supremum of \( c(E) \) over all \( E \in \mathcal{E} \) with \( E \subset A \). Thus we get
\[
c_\ast(A) = \sup \{ \lambda^\ast(E) : E \in \mathcal{E}, E \subset A \}.
\] (3.20)
But Proposition 3.2(2) gives \( \lambda^\ast(E) \leq \lambda^\ast(A) \) for all such \( E \) and so \( c_\ast(A) \leq \lambda^\ast(A) \) as well. If \( A \in \mathcal{J} \), then \( c^\ast(A) = c_\ast(A) \) and so they also equal \( \lambda^\ast(A) \).

The above argument has a simple, albeit quite non-trivial, consequence:

**Corollary 3.6** If \( \{ I_i \} \subset \mathcal{J} \) are disjoint, then
\[
\lambda^\ast\left( \bigcup_{i \in \mathbb{N}} I_i \right) = \sum_{i \in \mathbb{N}} |I_i|.
\] (3.21)
In particular, for any pair of disjoint collections \( \{ I_i \} \subset \mathcal{J} \) and \( \{ J_j \} \subset \mathcal{J} \),
\[
\bigcup_{i \in \mathbb{N}} I_i = \bigcup_{j \in \mathbb{N}} J_j \quad \Rightarrow \quad \sum_{i \in \mathbb{N}} |I_i| = \sum_{j \in \mathbb{N}} |J_j|.
\] (3.22)

**Proof.** By subadditivity (Proposition 3.2(3)) we have “\( \leq \)” For the other direction we use that \( \bigcup_{i \in \mathbb{N}} I_i \supset E \) for \( E := \bigcup_{i=1}^n I_i \). This is an elementary set so the previous lemma gives us
\[
\lambda^\ast\left( \bigcup_{i \in \mathbb{N}} I_i \right) \geq \lambda^\ast\left( \bigcup_{i=1}^n I_i \right) = c_\ast\left( \bigcup_{i=1}^n I_i \right) = \sum_{i=1}^n |I_i|
\] (3.23)
by Proposition 3.2(2) and Theorem 2.13(1). Taking \( n \to \infty \), we get “\( \geq \)” as well. The second part follows trivially from the first. \( \square \)

We remark that, although (3.22) is quite intuitive, a direct proof may be quite challenging.

Lemmas 3.3 and 3.5 can be further strengthened for bounded sets. Given a bounded \( A \subset \mathbb{R}^d \) and any \( I \in \mathcal{J} \) such that \( A \subset I \), define
\[
\lambda_\ast(A) := |I| - \lambda^\ast(I \setminus A).
\] (3.24)
This is the *inner Lebesgue measure*. As is easy to check (using, e.g., Corollary 3.6), this definition is independent of the choice of \( I \). We then have:

**Lemma 3.7** Let \( A \subset \mathbb{R}^d \) be bounded. Then
\[
c_\ast(A) \leq \lambda_\ast(A) \leq \lambda^\ast(A) \leq c^\ast(A).
\] (3.25)
In particular, if \( A \in \mathcal{J} \), then \( \lambda^\ast(A) = \lambda_\ast(A) \).

**Proof.** This is a direct consequence of Lemmas 3.3, 3.5 and 2.12. \( \square \)

This property suggest calling a bounded set \( A \subset \mathbb{R}^d \) to be Lebesgue measurable whenever \( \lambda^\ast(A) = \lambda_\ast(A) \). The conclusion of the lemma is then that every Peano-Jordan measurable set is automatically Lebesgue measurable. Unfortunately, defining Lebesgue measurability via this route would make the whole concept restricted to bounded sets. So another, more general, definition will need to be employed that subsumes the inner/outer measure definition for bounded sets.
3.2 Lebesgue measurable sets.

Our definition of measurability will be tied to topology. Recall that a set \( O \subset \mathbb{R}^d \) is open if, with each point \( x \), it contains a ball of a positive radius centered at \( x \). The class of open sets defines a topology. The following property ties topology to the Lebesgue outer measure:

**Lemma 3.8** For each \( A \subset \mathbb{R}^d \),

\[
\lambda^*(A) = \inf \{ \lambda^*(O) : O \subset A, \text{ open} \}.
\]  \hspace{1cm} (3.26)

**Proof:** Whenever \( A \subset O \), we have \( \lambda^*(A) \leq \lambda^*(O) \) by Proposition 3.2(2). Hence, \( \lambda^*(A) \) is at most the infimum. For the opposite inequality assume, without loss of generality, that \( \lambda^*(A) < \infty \). Then for each \( \epsilon > 0 \) there are \( \{I_i\} \subset \mathcal{F} \) such that \( A \subset \bigcup_{i \in \mathbb{N}} I_i \) and \( \lambda^*(A) + \epsilon \geq \sum_{i \in \mathbb{N}} |I_i| \). Now enlarge each \( I_i \) to get \( I'_i \) open with \( I'_i \supset I_i \) and \( |I'_i| \leq |I_i| + \epsilon 2^{-i} \). Define \( O := \bigcup_{I \in \mathcal{F}} I'_i \). Then \( O \) is open and \( O \supset A \). Moreover,

\[
\lambda^*(O) \leq \sum_{i \in \mathbb{N}} |I'_i| \leq \epsilon + \sum_{i \in \mathbb{N}} |I_i| \leq 2 \epsilon + \lambda^*(A).
\]  \hspace{1cm} (3.27)

Since \( \epsilon \) was arbitrary, the claim follows. \( \Box \)

When an open set \( O \supset A \) is gradually decreased along a sequence \( O_n \) so that \( \lambda^*(O_n) \downarrow \lambda^*(A) \), we may ask whether \( \lambda^*(O_n \setminus A) \) tends to zero as well. We may even consider the intersection \( B := \bigcap_{n \in \mathbb{N}} O_n \) and ask about the value of \( \lambda^*(B \setminus A) \). None of these seem to hold generally and so we single out a class of sets for which these properties can be proved:

**Definition 3.9** (Lebesgue measurable sets) A set \( A \subset \mathbb{R}^d \) is Lebesgue measurable if for every \( \epsilon > 0 \) there is an open set \( O \subset A \) such that \( \lambda^*(O \setminus A) < \epsilon \).

We will write \( \mathcal{L} \) or, when explicit mention of dimension is necessary, \( \mathcal{L}(\mathbb{R}^d) \) to denote the class of Lebesgue measurable sets in \( \mathbb{R}^d \). This class has the following properties:

**Proposition 3.10** We have:

1. If \( O \subset \mathbb{R}^d \) is open, then \( O \in \mathcal{L}(\mathbb{R}^d) \).
2. If \( \{A_n\} \subset \mathcal{L}(\mathbb{R}^d) \), then also \( \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{L}(\mathbb{R}^d) \).
3. If \( C \subset \mathbb{R}^d \) is closed, then \( C \in \mathcal{L}(\mathbb{R}^d) \).
4. If \( \lambda^*(A) = 0 \) then \( A \in \mathcal{L}(\mathbb{R}^d) \).
5. If \( A \in \mathcal{L}(\mathbb{R}^d) \) then also \( A^c \in \mathcal{L}(\mathbb{R}^d) \).

**Proof of (1) and (2).** (1) is trivial as for \( A \) open one takes \( O := A \). For (2), let \( \{A_n\} \subset \mathcal{L}(\mathbb{R}^d) \). Then, given \( \epsilon > 0 \), we can find open sets \( O_n \supset A_n \) such that \( \lambda^*(O_n \setminus A_n) < \epsilon 2^{-n} \). Then \( A := \bigcup_{n \in \mathbb{N}} A_n \subset O := \bigcup_{n \in \mathbb{N}} O_n \) with \( O \) open and, since

\[
\bigcup_{n \in \mathbb{N}} O_n \setminus \bigcup_{n \in \mathbb{N}} A_n \subset \bigcup_{n \in \mathbb{N}} (O_n \setminus A_n),
\]  \hspace{1cm} (3.28)

the monotonicity and countable subadditivity of \( \lambda^* \) show

\[
\lambda^*(A_n \setminus O_n) \leq \sum_{n \in \mathbb{N}} \lambda^*(O_n \setminus A_n) \leq \sum_{n \in \mathbb{N}} \epsilon 2^{-n} = \epsilon.
\]  \hspace{1cm} (3.29)

Hence \( A \in \mathcal{L}(\mathbb{R}^d) \) as well. \( \Box \)
For the proof of the remaining parts, we need some lemmas:

**Lemma 3.11** If \( O \subset \mathbb{R}^d \) is open then there are \( \{I_i\} \subset \mathcal{I} \) disjoint such that \( O = \bigcup_{i \in \mathbb{N}} I_i \).

**Proof.** This is based on the use of dyadic partitions of \( \mathbb{R}^n \). Let

\[
D_n := \left\{ (k_1 2^{-n}, (k_1 + 1)2^{-n}] \times \cdots \times (k_d 2^{-n}, (k_d + 1)2^{-n}] : k_1, \ldots, k_d \in \mathbb{Z} \right\},
\]

(3.30)

We call elements of \( D_n \) the dyadic (half-open) cubes of base \( 2^{-n} \). These are rather special because, for any \( m, n \in \mathbb{N} \) with \( m \leq n \) we have

\[
I \in D_m, J \in D_n \quad \Rightarrow \quad J \subseteq I \text{ OR } I \cap J = \emptyset.
\]

(3.31)

This is because the partitions induced by \( D_n \) and \( D_m \) are nested.

Now pick \( O \) open and define a sequence of sets \( A_n \) inductively as follows: Set \( A_1 := O \) and, given \( A_n \), let \( A_{n+1} \) be the set of all \( x \in A_n \) such that if \( I \in D_{n+1} \) obeys \( x \in I \), then \( A_n \setminus I \neq \emptyset \). (In other words, \( A_{n+1} \) is \( A_n \) with all \( I \in D_{n+1} \) that fit entirely into \( A_n \) removed.) Then

\[
A_n \setminus A_{n-1} = \bigcup_{i=1}^{m(n)} J_{n,i}
\]

(3.32)

for some \( J_{n,i} \in D_n \) with \( \{J_{n,i} : n \in \mathbb{N}, i=1, \ldots, m(n)\} \) disjoint. Obviously, \( A_n \) decreases with \( n \); we claim that \( A_n \downarrow \emptyset \). Indeed, with each \( x \in O \), the set \( O \) contains an open Euclidean ball containing \( x \) and thus also a dyadic cube \( I \in D_{n+1} \) with \( x \in I \), for some \( n \in \mathbb{N} \). By (3.31), either this \( I \) is either entirely contained in one of dyadic cubes constituting \( O \setminus A_n \), or is disjoint from all of them. In the former case \( x \in O \setminus A_n \), in the latter case \( x \in A \setminus A_{n+1} \). Hence,

\[
O = A_1 = \bigcup_{n \geq 2} (A_n \setminus A_{n-1}) = \bigcup_{n \geq 2} \bigcup_{i=1}^{m(n)} J_{n,i}.
\]

(3.33)

Since the dyadic cubes on the right are by construction disjoint, the claim follows. \( \square \)

The other lemma gives a statement concerning additivity of the outer measure \( \lambda^* \). We note that outer measures are generally not additive — i.e., we may have \( \lambda^* (A \cup B) < \lambda^* (A) + \lambda^* (B) \) even if \( A, B \subset \mathbb{R}^d \) obey \( A \cap B = \emptyset \). However, if we assume that \( A \) and \( B \) are separated by a positive distance, then additivity holds:

**Lemma 3.12** Let \( \rho(A, B) := \inf \{|x-y|_\infty : x \in A, y \in B\} \) where \( |x-y|_\infty := \max_{i=1, \ldots, d} |x_i - y_i| \). Then for all \( A, B \subset \mathbb{R}^d \),

\[
\rho(A, B) > 0 \quad \Rightarrow \quad \lambda^* (A \cup B) = \lambda^* (A) + \lambda^* (B).
\]

(3.34)

**Proof.** In light of subadditivity of \( \lambda^* \), we only need to show “\( \geq \)” which in turn permits us to assume \( \lambda^* (A \cup B) < \infty \). We will use the fact that, by Lemma 2.4, given any \( \delta > 0 \), if every cover \( \{I_i\} \) of a set by half-open boxes can be refined into a cover by boxes whose every side is less than \( \delta \) without changing \( \sum_{i \in \mathbb{N}} |I_i| \).

Thus, pick \( \delta := 1/2 \rho(A, B) \) and let \( \{I_i\} \) be a cover of \( A \cup B \) by half-open boxes (or empty sets) with the largest side at most \( \delta \) so that

\[
\lambda^* (A \cup B) + \varepsilon \geq \sum_{i \in \mathbb{N}} |I_i|.
\]

(3.35)
Define $S := \{i \in \mathbb{N}: I_i \cap A \neq \emptyset\}$ and note that then $A \subset \bigcup_{i \in S} I_i$. By our choice of $\delta$, $i \in S$ implies $I_i \cap B = \emptyset$ and so also $B \subset \bigcup_{i \in \mathbb{N} \setminus S} I_i$. Hence, by the definition of $\lambda^*$ and (3.35),

$$\lambda^*(A) + \lambda^*(B) \leq \sum_{i \in S} |I_i| + \sum_{i \in \mathbb{N} \setminus S} |I_i| = \sum_{i \in \mathbb{N}} |I_i| \leq \lambda^*(A \cup B) + \varepsilon. \quad (3.36)$$

Since $\varepsilon$ was arbitrary, the claim thus follows. \hfill \square

We may now finish the proof of Proposition 3.10:

**Proof of (3).** Pick $C \subset \mathbb{R}^d$ closed and assume, for the start, that $C$ is also bounded (and thus compact). If $O$ is an open set such that $O \supseteq C$, then also $O \setminus C$ is open. By Lemma 3.11 there are disjoint $\{I_i\} \subset \mathcal{F}$ so that

$$O \setminus C = \bigcup_{i \in \mathbb{N}} I_i. \quad (3.37)$$

By Corollary 3.6,

$$\lambda^*(O \setminus C) = \sum_{i \in \mathbb{N}} |I_i|. \quad (3.38)$$

Next let $\varepsilon > 0$ and let $I'_i \in \mathcal{F}$ be such that $\overline{I_i} \subset I'_i$ and $|I'_i| + \varepsilon 2^{-i} \leq |I_i|$. By the Bolzano-Weierstrass theorem, two compact sets $A, B \subset \mathbb{R}^d$ with $A \cap B = \emptyset$ necessarily obey $\rho(A, B) > 0$, and since a finite union of compact sets is compact,

$$\rho\left(C, \bigcup_{i=1}^n I'_i\right) > 0, \quad n \in \mathbb{N}. \quad (3.39)$$

Corollary 3.6 gives $\lambda^*(\bigcup_{i=1}^n I'_i) = \sum_{i=1}^n |I'_i|$ and, by (3.37), the monotonicity of $\lambda^*$, Lemma 3.12 and the construction of $\{I'_i\}$ we thus get

$$\lambda^*(O) \geq \lambda^*\left(C \cup \bigcup_{i=1}^n I'_i\right) = \lambda^*(C) + \lambda^*\left(\bigcup_{i=1}^n I'_i\right) = \lambda^*(C) + \sum_{i=1}^n |I'_i| \quad (3.40)$$

$$\geq \lambda^*(C) - \varepsilon + \sum_{i=1}^n |I_i|. \quad (3.41)$$

With the help of (3.38), the limits $n \to \infty$ and $\varepsilon \downarrow 0$ yield

$$\lambda^*(O \setminus C) = \sum_{i \in \mathbb{N}} |I_i| \leq \lambda^*(O) - \lambda^*(C). \quad (3.42)$$

Since $C$ is bounded, $\lambda^*(C) < \infty$. By Lemma 3.8 for each $\varepsilon > 0$ there is $O \supseteq C$ open so that $\lambda^*(O) \leq \lambda^*(C) + \varepsilon$. The above then shows $\lambda^*(O \setminus C) < \varepsilon$ and so $C \in \mathcal{L}(\mathbb{R}^d)$. This proves (3) for every $C$ compact. As every closed set in $\mathbb{R}^d$ can be written as a countable union of compact sets, (2) extends this to general closed sets as well. \hfill \square

**Proof of (4).** Let $A \subset \mathbb{R}^d$ be such that $\lambda^*(A) = 0$. Since $\mathbb{R}^d$ is open, the infimum in Lemma 3.8 is not over an empty set. Hence, for each $\varepsilon > 0$ there is an $O \supset A$ open such that $\lambda^*(O) < \varepsilon$. But $\lambda^*(O \setminus A) \leq \lambda^*(O)$ and so $A \in \mathcal{L}(\mathbb{R}^d)$ as claimed. \hfill \square
Proof of (5). Let \( A \in \mathcal{L}(\mathbb{R}^d) \). Then there are \( O_n \supset A \) open with \( \lambda^*(O_n \setminus A) \to 0 \). Define \( B := \bigcup_{n \in \mathbb{N}} O_n \) and note that \( A^c \setminus B \subseteq A^c \setminus O_n^c \) for each \( n \in \mathbb{N} \). But \( A \subset O_n \) implies \( A^c \setminus O_n^c = O_n \setminus A \) and so
\[
\lambda^*(A^c \setminus B) \leq \lambda^*(A^c \setminus O_n^c) = \lambda^*(O_n \setminus A) \quad \text{n} \to \infty.
\]
Hence \( \lambda^*(A^c \setminus B) = 0 \) and, by (4), \( A^c \setminus B \in \mathcal{L}(\mathbb{R}^d) \). But \( B \in \mathcal{L}(\mathbb{R}^d) \) as well by (1,2,3) and, since \( B \subset A^c \), also \( A^c = B \cup A^c \in \mathcal{L}(\mathbb{R}^d) \) by (2).

Since \( \emptyset \) and \( \mathbb{R}^d \) are automatically open, Proposition 3.10 shows that the structure of \( \mathcal{L}(\mathbb{R}^d) \) fits into the following important concept:

**Definition 3.13** A class \( \mathcal{M} \) of subsets of a set \( X \) is a \( \sigma \)-algebra if

1. \( / 0, X \in \mathcal{M} \),
2. \( A \in \mathcal{M} \Rightarrow A^c \in \mathcal{M} \),
3. \( \{A_n\} \in \mathcal{M} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{M} \).

Note that this specializes the concept of an algebra, introduced earlier in these notes, by requiring that also countable unions of sets in \( \mathcal{M} \) are contained in \( \mathcal{M} \). Obviously, \( \mathcal{M} \) is then closed also under countable intersection and set differences.

An important question is whether the characterization we used to define \( \mathcal{L}(\mathbb{R}^d) \) is unique. Consider the following classes of sets:

\[
\mathcal{F}_\sigma(\mathbb{R}^d) := \left\{ \bigcup_{n \in \mathbb{N}} K_n : K_n \subset \mathbb{R}^d, \text{closed} \right\} \quad (3.43)
\]
and
\[
\mathcal{G}_\delta(\mathbb{R}^d) := \left\{ \bigcap_{n \in \mathbb{N}} O_n : O_n \subset \mathbb{R}^d, \text{open} \right\} \quad (3.44)
\]

As a simple consequence of Proposition 3.10, we get:

**Corollary 3.14** (Alternative definitions of Lebesgue measurable sets) \( A \in \mathcal{L}(\mathbb{R}^d) \) is equivalent to any of these three statements:

1. For each \( \varepsilon > 0 \) there is \( K \subset A \) closed so that \( \lambda^*(A \setminus K) < \varepsilon \).
2. There is \( B \in \mathcal{G}_\delta(\mathbb{R}^d) \) such that \( B \supset A \) and \( \lambda^*(B \setminus A) = 0 \).
3. There is \( C \in \mathcal{F}_\sigma(\mathbb{R}^d) \) such that \( C \subset A \) and \( \lambda^*(A \setminus C) = 0 \).

**Proof.** (1) If \( O \supset A^c \) is open then \( K := O^c \) is closed with \( K \subset A \) and \( A \setminus K = O \setminus A^c \). For (2) use the argument from the proof (5) of Proposition 3.10; (3) then follows by complementation. \( \Box \)

### 3.3 Lebesgue measure.

We now move on to defining the Lebesgue measure. Here is a general definition:

**Definition 3.15** (General measure) Let \( \mathcal{M} \) be a \( \sigma \)-algebra of subsets of a set \( X \). The map \( \mu : \mathcal{M} \to [0, \infty] \) is a measure if

1. \( \mu(\emptyset) = 0 \), and
Condition (2) is referred to as countable subadditivity of $\mu$.

The main purpose of condition (1) is to ensure that $\mu$ is not identically infinity on $\mathcal{M}$. (Note that, by (2), $\mu(\emptyset)$ can only take values 0 or $\infty$.) Going back to our main line of reasoning, we now claim:

**Proposition 3.16**  The restriction of $\lambda^*$ to $\mathcal{L}(\mathbb{R}^d)$ is a measure.

**Proof.** We have $\lambda^*(\emptyset) = 0$ so we only need to show that $\lambda^*$ acts additively on countable unions of disjoint sets from $\mathcal{L}(\mathbb{R}^d)$. First we prove finite additivity for bounded sets. Let $A, B \in \mathcal{L}(\mathbb{R}^d)$ be bounded. By Corollary 3.14, for each $\varepsilon > 0$ there are closed sets $K_1 \subset A$ and $K_2 \subset B$ (thus compact) such that $\lambda^*(A \setminus K_1), \lambda^*(B \setminus K_2) < \varepsilon$. If also $A \cap B = \emptyset$, the fact that $K_1$ and $K_2$ are compact ensures that $\rho(K_1, K_2) > 0$. Hence

$$\lambda^*(K_1 \cup K_2) = \lambda^*(K_1) + \lambda^*(K_2)$$

by Lemma 3.12. But $\lambda^*(A) \leq \lambda^*(K_1) + \varepsilon$ and $\lambda^*(B) \leq \lambda^*(K_2) + \varepsilon$ by subadditivity of $\lambda^*$ and so

$$\lambda^*(A \cup B) \geq \lambda^*(K_1 \cup K_2) \geq \lambda^*(A) + \lambda^*(B) - 2\varepsilon.$$  \hspace{1cm} (3.47)

By subadditivity again, $\lambda^*$ is additive on finite unions of bounded disjoint sets.

Now consider a countable collection $\{A_n\} \subset \mathcal{L}(\mathbb{R}^d)$ of bounded disjoint sets. Then

$$\lambda^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) \geq \lambda^*\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \lambda^*(A_i) \rightarrow \sum_{n \in \mathbb{N}} \lambda^*(A_n)$$

where we used finite additivity to get the middle equality. Hence, $\lambda^*$ is additive on countable unions of disjoint bounded sets. To remove the boundedness restriction, let $\mathcal{D}$ be a partition of $\mathbb{R}^d$ into shifts of $(0,1]^d$ by integers in each lattice direction. If $\{D_m\}$ enumerates $\mathcal{D}$ and $\{A_n\} \subset \mathcal{L}(\mathbb{R}^d)$ is a disjoint collection, then

$$\lambda^*\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \lambda^*\left(\bigcup_{m,n \in \mathbb{N}} (A_n \cap D_m)\right)$$

$$= \sum_{m,n \in \mathbb{N}} \lambda^*(A_n \cap D_m) = \sum_{n \in \mathbb{N}} \sum_{m \in \mathbb{N}} \lambda^*(A_n \cap D_m)$$

$$= \sum_{n \in \mathbb{N}} \lambda^*(\bigcup_{m \in \mathbb{N}} (A_n \cap D_m)) = \sum_{n \in \mathbb{N}} \lambda^*(A_n)$$

where the second and fourth equality follow because $\{(A_n \cap B_m) : m,n \in \mathbb{N}\}$ is a countable collection of disjoint bounded sets, the third equality is a consequence of (3.5) and the last equality follows because $\{D_m : m \in \mathbb{N}\}$ form a partition of $\mathbb{R}^d$. \qed

**Definition 3.17** (Lebesgue measure)  Let us write $\lambda(A)$ for $\lambda^*(A)$ whenever $A \in \mathcal{L}(\mathbb{R}^d)$. We will call $\lambda(A)$ the Lebesgue measure of $A$.

Returning to the concept of inner measure for bounded sets, we now observe:
Proposition 3.18  Let $A \subset \mathbb{R}^d$ be bounded. Then $A \in \mathcal{L}(\mathbb{R}^d)$ if and only if $\lambda^*(A) = \lambda_*(A)$.

Proof. Homework exercise. $\square$

The Lebesgue measure has a lot of symmetries. In particular, we have:

Proposition 3.19 (Translation invariance)  For any $A \subset \mathbb{R}^d$ and $z \in \mathbb{R}^d$, define

$$z + A := \{z + x : x \in A\} \quad (3.50)$$

Then for all $A \in \mathcal{L}(\mathbb{R}^d)$ and all $z \in \mathbb{R}^d$,

$$z + A \in \mathcal{L}(\mathbb{R}^d) \quad \text{and} \quad \lambda(z + A) = \lambda(A). \quad (3.51)$$

Proof. Homework exercise (easy). $\square$

The Lebesgue measure is further invariant under rotations (i.e., linear maps induced by orthogonal matrices) and transforms nicely under dilations. We will revisit this question later.

3.4 Convergence Theorems.

As a corollary to the definition of measure, we get useful convergence theorems. (We state these for a general measure as no other properties are needed.) The first theorem deals with monotone sequences. Recall that we write $A_n \uparrow A$ whenever $n \mapsto A_n$ is increasing and $A = \bigcup_{n \in \mathbb{N}} A_n$. Similarly, we write $A_n \downarrow A$ whenever $n \mapsto A_n$ is decreasing and $A = \bigcap_{n \in \mathbb{N}} A_n$.

Theorem 3.20 (Monotone Convergence Theorem for sets)  Let $\mu$ be a measure on $\sigma$-algebra $\mathcal{M}$. Let $\{A_n\} \in \mathcal{M}$. Then we have:

1. If $A_n \uparrow A$, then $\mu(A_n) \uparrow \mu(A)$.
2. If $A_n \downarrow A$ and $\mu(A_1) < \infty$, then $\mu(A_n) \downarrow \mu(A)$.

Proof. If $A_n \uparrow A$ then $\{A_n \setminus A_{n-1} : n \in \mathbb{N}\}$, where $A_0 := \emptyset$, are disjoint with

$$A_n = \bigcup_{k=1}^{n} (A_k \setminus A_{k-1}) \quad \text{and} \quad A = \bigcup_{k \in \mathbb{N}} (A_k \setminus A_{k-1}) \quad (3.52)$$

Therefore,

$$\mu(A_n) = \sum_{k=1}^{n} \mu(A_k \setminus A_{k-1}) \xrightarrow{n \to \infty} \sum_{k \in \mathbb{N}} \mu(A_k \setminus A_{k-1}) = \mu(A). \quad (3.53)$$

This proves (1).

For (2), let $A_n \downarrow A$. Then $A_1 \setminus A_n \uparrow A_1 \setminus A$ and so, by $\mu(A_1) < \infty$,

$$\mu(A_n) = \mu(A_1) - \mu(A_1 \setminus A_n) \xrightarrow{n \to \infty} \mu(A_1) - \mu(A_1 \setminus A) = \mu(A). \quad (3.54)$$

where the limit follows by part (1). $\square$

Our next convergence result addresses the situation when the sequence of sets is not necessarily monotone. Given a collection of sets $\{A_n\}$, we define

$$\limsup_{n \to \infty} A_n := \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k \quad (3.55)$$

and
\[
\liminf_{n \to \infty} A_n := \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k
\] (3.56)

Obviously, \(\liminf_{n \to \infty} A_n \subset \limsup_{n \to \infty} A_n\) and so we say
\[
\lim A_n \text{ exists if } \limsup_{n \to \infty} A_n = \liminf_{n \to \infty} A_n \tag{3.57}
\]

The limit is the common value of the \textit{limes inferior} and the \textit{limes superior}. As is easy to check, this subsumes the aforementioned case when \(\{A_n\}\) is either increasing or decreasing.

\textbf{Lemma 3.21} (Fatou’s lemma for sets) \textit{Let \(\mu\) be a measure on a \(\sigma\)-algebra \(\mathcal{M}\). Let \(\{A_n\} \subset \mathcal{M}\). Then \(\liminf_{n \to \infty} A_n \subset \mathcal{M}\) and}
\[
\mu \left( \liminf_{n \to \infty} A_n \right) \leq \liminf_{n \to \infty} \mu(A_n). \tag{3.58}
\]

\textit{Proof.} For each \(m \geq n\) we have \(A_m \supset \bigcap_{k \geq n} A_k\). Hence,
\[
\inf_{m \geq n} \mu(A_m) \geq \mu \left( \bigcap_{k \geq n} A_k \right). \tag{3.59}
\]

Since \(\bigcap_{k \geq n} A_k \uparrow \liminf_{n \to \infty} A_n\), the claim follows from the Upward Monotone Convergence Theorem for sets. \(\square\)

Due to the asymmetry between the Upward and Downward version of the Monotone Convergence Theorem for sets, the corresponding theorem for \textit{limes superior} requires bounded measures:

\textbf{Lemma 3.22} (Reverse Fatou’s lemma for sets) \textit{Let \(\mu\) be a measure on a \(\sigma\)-algebra \(\mathcal{M}\). Let \(\{A_n\} \subset \mathcal{M}\). Then \(\limsup_{n \to \infty} A_n \subset \mathcal{M}\) and, if there is \(B \in \mathcal{M}\) with \(\mu(B) < \infty\) and \(B \supset A_n\) for each \(n\), then}
\[
\mu \left( \limsup_{n \to \infty} A_n \right) \leq \limsup_{n \to \infty} \mu(A_n). \tag{3.60}
\]

\textit{Proof.} We have
\[
\limsup_{n \to \infty} A_n = B \setminus \liminf_{n \to \infty} (B \setminus A_n) \tag{3.61}
\]

Once \(\mu(B) < \infty\), the result follows by complementation from Fatou’s lemma. \(\square\)

When the limit of \(A_n\) exists, these simplify into:

\textbf{Theorem 3.23} (Dominated Convergence Theorem for sets) \textit{Let \(\mu\) be a measure on \(\sigma\)-algebra \(\mathcal{M}\). Let \(\{A_n\} \subset \mathcal{M}\) be such that \(\lim_{n \to \infty} A_n\) exists. Then \(\lim_{n \to \infty} A_n \in \mathcal{M}\) and if there is a \(B \in \mathcal{M}\) with \(\mu(B) < \infty\) and \(B \supset A_n\) for all \(n\), we have}
\[
\mu \left( \lim_{n \to \infty} A_n \right) = \lim_{n \to \infty} \mu(A_n). \tag{3.62}
\]

\textit{(In particular, the limit on the right exists, albeit only in \([0, \infty]\).)}

\textit{Proof.} Apply Fatou and reverse Fatou to get
\[
\limsup_{n \to \infty} \mu(A_n) \leq \mu \left( \liminf_{n \to \infty} A_n \right) \leq \liminf_{n \to \infty} \mu(A_n). \tag{3.63}
\]

Hence the limit of \(\mu(A_n)\) exists and equals the measure of the limit of \(A_n\). \(\square\)
We remark that, perhaps with the exception of the Monotone Convergence Theorem, the above results are far more frequently used in their versions for integrals of measurable functions.

### 3.5 Nonmeasurable sets.

We have seen that all open and closed sets are Lebesgue measurable, and so are the sets in $\mathcal{F}_\sigma(\mathbb{R}^d)$ and $\mathcal{G}_\delta(\mathbb{R}^d)$ — and in particular, also the counterexample to Peano-Jordan measurability from the proof of Lemma 2.18. In fact, for any set $\mathcal{A} \subset \mathcal{L}(\mathbb{R}^d)$, if we denote by

$$\mathcal{A}_\sigma := \left\{ \bigcup_{n \in \mathbb{N}} A_n : A_n \in \mathcal{A} \right\}$$

and

$$\mathcal{A}_\delta := \left\{ \bigcap_{n \in \mathbb{N}} O_n : O_n \in \mathcal{A} \right\}$$

the sets in $A_\delta, A_\sigma$, $A_\delta \sigma := (A_\delta)_\sigma$, $A_\sigma \delta := (A_\sigma)_\delta$, $A_\sigma \delta \sigma := (A_\sigma \delta)_\sigma$, etc., all lie in $\mathcal{L}(\mathbb{R}^d)$. A question thus arises whether there are in fact any sets that are not Lebesgue measurable. This was resolved rather soon after Lebesgue set up his theory:

**Theorem 3.24 (Vitali)** Assuming the Axiom of Choice, there is $A \subset [0, 1]$ such that $A \notin \mathcal{L}(\mathbb{R}^d)$.

Here we recall:

**Axiom of Choice** For any set $X$ there is a map $f : 2^X \to X$ so that $f(A) \in A$ for every $A \in 2^X$.

In words, the Axiom of Choice guarantees that from each subset $A$ of $X$ we can choose a representative element $f(x)$ that, naturally, belongs to $A$. This axiom of set theory (not the part of the standard ZF system) is often invoked in analysis where a proof without it is not known (or known not to be possible, as is the example of Vitali’s theorem).

**Proof of Theorem 3.24.** Define an equivalence relation $\sim$ on $\mathbb{R}$ by saying $x \sim y$ if and only if $x - y \in \mathbb{Q}$. Denoting by $[x] := \{ y \in \mathbb{R} : y \sim x \}$ the equivalence class containing $x$, we have $\mathbb{R} = \bigcup_{x \in \mathbb{R}} [x]$. The Axiom of Choice guarantees the existence of a map $f : \{ [x] : x \in \mathbb{R} \} \to \mathbb{R}$ so that $f([x]) \in [x]$. By taking this map modulo integers, we may (and will) assume that $f([x]) \in [0, 1)$ for all $x \in \mathbb{R}$.

Let $A := \{ f(x) : x \in \mathbb{R} \}$. Then $A \subset [0, 1]$. Writing $q + A := \{ q + x : x \in A \}$, we now claim

$$\forall q, q' \in \mathbb{Q}: \quad q + A \cap q' + A \neq \emptyset \iff q = q'$$

(3.66)

Indeed, if $x, y \in A$ are such that $x + q = y + q'$ for $q, q' \in \mathbb{Q}$, then $y \in [x]$. But $A$ contains at most one representative of $[x]$ and so $y = x$. Then also $q = q'$. (The other direction is trivial.)

From the above we know that $\{ q + A : q \in \mathbb{Q} \}$ are disjoint. Next we observe

$$\bigcup_{q \in \mathbb{Q}} (q + A) = \mathbb{R}.$$  (3.67)

This is because each $y \in \mathbb{R}$ belongs to $[x]$ for some $x \in A$. Finally, we note that, since $A \subset [0, 1]$,

$$\bigcup_{q \in [0,1] \cap \mathbb{Q}} (q + A) \subset [0, 2].$$  (3.68)
Now let us assume that $A \in \mathcal{L}^e(\mathbb{R})$ and so $\lambda(A)$ is meaningful. Then Proposition 3.19 tells us that also $q + A \in \mathcal{L}^e(\mathbb{R})$ and

$$\lambda(q + A) = \lambda(A)$$

(3.69)

for all $q \in \mathbb{Q}$. So, in light of (3.68), (3.66) and

$$2 = \lambda([0,2]) \geq \sum_{q \in [0,1] \cap \mathbb{Q}} \lambda(q + A) = \sum_{q \in [0,1] \cap \mathbb{Q}} \lambda(A),$$

(3.70)

we must have $\lambda(A) = 0$. However, (3.67) gives

$$\lambda(\mathbb{R}) = \sum_{q \in \mathbb{Q}} \lambda(q + A) = \sum_{q \in \mathbb{Q}} \lambda(A) = 0$$

(3.71)

a contradiction with $\lambda(\mathbb{R}) = \infty$. Hence $A \notin \mathcal{L}^e(\mathbb{R})$ after all. $\square$

It is known that the Axiom of Choice cannot be left out from the assumptions. In fact, there are set theories without the Axiom of Choice where all subsets of $[0,1]$ are measurable.
4. Lebesgue integral: bounded case

Having developed the concept of the Lebesgue measure, we will now move to defining the Lebesgue integral. It would be a waste to restrict attention to integrals with respect to the Lebesgue measure only; indeed, working with general measure adds only little extra work. (We will still call the integral the Lebesgue integral, though.) The theory is best appreciated if started from integrating bounded functions on sets of finite measure.

4.1 Bounded functions on finite measure spaces.

Recall the definitions of the \( \sigma \)-algebra and general measure from Definitions 3.13 and 3.15. We start with some standard notions:

**Definition 4.1**  A pair \((X,F)\), where \(F\) is a \( \sigma \)-algebra of subsets of \(X\), is called a measurable space. If \(\mu\) is a measure on \((X,F)\), we call \((X,F,\mu)\) a measure space. The measure space \((X,F,\mu)\) is finite if \(\mu(X) < \infty\). It is a probability space if \(\mu(X) = 1\).

Let \((X,F)\) be a measurable space and consider a bounded function \(f : X \to \mathbb{R}\). The definition of the Lebesgue integral we will give is similar to Darboux’ version of Riemann’s integral: We will approximate \(f\) from below and above by “piece-wise” constant functions — to be called simple functions — for which the integral is canonically defined. If the resulting upper and lower (Lebesgue) integrals coincide, we proclaim the common value the (Lebesgue) integral of \(f\). The simple functions are defined as follows:

**Definition 4.2**  Let \((X,F)\) be a measurable space. A map \(\phi : X \to \mathbb{R}\) is a simple function if there are disjoint sets \(\{A_1,\ldots,A_n\} \subset F\) and values \(a_1,\ldots,a_n \in \mathbb{R}\) such that

\[
\phi = \sum_{i=1}^{n} a_i 1_{A_i}. \tag{4.1}
\]

Obviously, we may even require that the values \(\{a_i\}\) are distinct and non-zero. Under this restriction, the representation (4.1) of \(\phi\) is unique. The integral of simple functions with respect to a measure \(\mu\) on \((X,F)\) is then defined canonically by

\[
\int \phi \, d\mu := \sum_{i=1}^{n} a_i \mu(A_i). \tag{4.2}
\]

Here we see the reason for requiring the \(A_i\)'s to be measurable.

The distinctness restriction on \(\{a_i\}\) are sometimes inconvenient to impose. However, without this restriction the representation of the simple function is no longer unique and one then needs to check that the integral does not depend on the representation. It suffices to prove:

**Lemma 4.3** (Independence of representation) Suppose \(\mu(X) < \infty\) and consider a collection of numbers \(a_1,\ldots,a_n, b_1,\ldots,b_m \in \mathbb{R}\) and sets \(A_1,\ldots,A_n, B_1,\ldots,B_m \in F\) be such that both families \(\{A_i\}\) and \(\{B_j\}\) are disjoint and \(\{b_j\}\) are distinct. If

\[
\sum_{i=1}^{n} a_i 1_{A_i}(x) = \sum_{j=1}^{m} b_j 1_{B_j}(x), \quad x \in X, \tag{4.3}
\]
then also
\[ \sum_{i=1}^{n} a_i \mu(A_i) = \sum_{j=1}^{m} b_j \mu(B_j). \]  \hspace{1cm} (4.4)

In particular, (4.2) applies whenever (4.1) holds, regardless of distinctness restriction on \{a_i\}.

**Proof.** A term with \(a_i = 0\) or \(b_j = 0\) can be left out of the expression so let us also assume that \(a_1, \ldots, a_n, b_1, \ldots, b_m \neq 0\). Then (4.3) shows \( \bigcup_{i=1}^{n} A_i = \bigcup_{j=1}^{m} B_j \). Hence \{\(A_i\)\} and \{\(B_j\)\} are both disjoint partitions of this union and so
\[ A_i = \bigcup_{j=1}^{m} (A_i \cap B_j) \quad \text{and} \quad B_j = \bigcup_{i=1}^{n} (A_i \cap B_j). \]  \hspace{1cm} (4.5)

Since also \(A_i \cap B_j \neq \emptyset\) \(\Rightarrow\) \(a_i = b_j\) (4.6) we thus have
\[ \sum_{i=1}^{n} a_i \mu(A_i) = \sum_{j=1}^{m} \sum_{i=1}^{n} a_i \mu(A_i \cap B_j) \]
\[ = \sum_{j=1}^{m} \sum_{i=1}^{n} b_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} b_j \mu(B_j) \]  \hspace{1cm} (4.7)

where we also used finite additivity of \(\mu\). \(\square\)

The integral in (4.2) has the following immediate properties:

**Lemma 4.4** Suppose \(\mu(X) < \infty\). Then we have:

1. If \(\phi_1, \phi_2\) are simple then so is \(\phi_1 + \phi_2\) and

\[ \int (\phi_1 + \phi_2) \, d\mu = \int \phi_1 \, d\mu + \int \phi_2 \, d\mu. \]  \hspace{1cm} (4.8)

2. If \(\phi\) is simple and \(c \in \mathbb{R}\), then \(c\phi\) is simple and

\[ \int (c\phi) \, d\mu = c \int \phi \, d\mu. \]  \hspace{1cm} (4.9)

3. If \(\phi_1, \phi_2\) are simple with \(\phi_1 \leq \phi_2\), then

\[ \int \phi_1 \, d\mu \leq \int \phi_2 \, d\mu. \]  \hspace{1cm} (4.10)

4. If \(\phi\) is simple, then

\[ \left| \int \phi \, d\mu \right| \leq (\sup_{x \in X} |\phi(x)|) \mu(X). \]  \hspace{1cm} (4.11)

**Proof.** Property (2) is immediate from the definition of the integral. An important technical point for the proof of (1) and (3) is that if \(\phi_1\) and \(\phi_2\) are simple functions then there are disjoint sets \{\(A_i\)\} and (not necessarily distinct) collections of numbers \(a_1, \ldots, a_n\) and \(b_1, \ldots, b_m\) such that
\[ \phi_1 = \sum_{i=1}^{n} a_i 1_{A_i} \quad \text{and} \quad \phi_2 = \sum_{i=1}^{m} b_i 1_{A_i}. \]  \hspace{1cm} (4.12)
Using this representation of both simple functions, to get (1) it then suffices to note that \( \phi_1 + \phi_2 = \sum_{i=1}^{n}(a_i + b_i)1_{A_i} \) and so
\[
\int (\phi_1 + \phi_2) \, d\mu = \sum_{i=1}^{n}(a_i + b_i) \mu(A_i)
\]
\[
= \sum_{i=1}^{n}a_i \mu(A_i) + \sum_{i=1}^{n}b_i \mu(A_i) = \int \phi_1 \, d\mu + \int \phi_2 \, d\mu,
\]
where in the first and third inequality we used Lemma 4.3.

For (3) we in turn note that, for the representation as in (4.12), \( \phi_1 \leq \phi_2 \) implies \( a_i \leq b_i \) whenever \( A_i \neq \emptyset \) and so (again with the help of Lemma 4.3),
\[
\int \phi_1 \, d\mu = \sum_{i=1}^{n}a_i \mu(A_i) \leq \sum_{i=1}^{n}b_i \mu(A_i) = \int \phi_2 \, d\mu.
\]
Finally, for property (4) we use the fact that \( \phi_1 := \left( \sup_{x \in X} |\phi(x)| \right) \) is simple and \( -\phi_1 \leq \phi \leq \phi_1 \). So, by (2) and (3) we have
\[
-(\sup_{x \in X} |\phi(x)|) \mu(X) = \int (-\phi_1) \, d\mu \leq \int \phi \, d\mu \leq \int \phi_1 \, d\mu = (\sup_{x \in X} |\phi(x)|) \mu(X).
\]

The claim in (4) now follows. \( \square \)

Now let us assume that \( f : X \to \mathbb{R} \) is a bounded function and that \( \mu(X) < \infty \). Similarly to the Darboux approach to Riemann integration, we define the lower Lebesgue integral of \( f \) with respect to \( \mu \) by
\[
\int f \, d\mu := \sup \left\{ \int \phi \, d\mu : \phi \, \text{simple}, \phi \leq f \right\}
\]
and the upper Lebesgue integral of \( f \) with respect to \( \mu \) by
\[
\int f \, d\mu := \inf \left\{ \int \phi \, d\mu : \phi \, \text{simple}, \phi \geq f \right\}.
\]
By Lemma 4.4 \( \phi_1 \leq f \leq \phi_2 \) implies \( \int \phi_1 \, d\mu \leq \int \phi_2 \, d\mu \) and so
\[
\int f \, d\mu \leq \int f \, d\mu.
\]
Lemma 4.4(4) also shows that both integrals are finite. We then define:

**Definition 4.5** (Lebesgue integrability) **Assuming \( \mu(X) < \infty \), a bounded function \( f : X \to \mathbb{R} \) is said to be Lebesgue integrable with respect to \( \mu \), or just integrable, if**
\[
\int f \, d\mu = \int f \, d\mu.
\]
**For integrable \( f \) we then define its** Lebesgue integral with respect to \( \mu \) by
\[
\int f \, d\mu := \int f \, d\mu := \int f \, d\mu.
\]

The integral thus defined naturally extends, and inherits many properties we already showed for, the integral of simple functions:
Lemma 4.6 Suppose $\mu(X) < \infty$ and let $f, g: X \to \mathbb{R}$ below be bounded functions. Then:
(0) If $f$ is simple then $f$ is integrable and its Lebesgue integral coincides with that in (4.2).
(1) If $f, g$ are integrable then so is $f + g$ and
\[
\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu.
\] (4.21)
(2) If $f$ is integrable and $c \in \mathbb{R}$, then $cf$ is simple and
\[
\int (cf) \, d\mu = c \int f \, d\mu.
\] (4.22)
(3) If $f, g$ are integrable with $f \leq g$, then
\[
\int f \, d\mu \leq \int g \, d\mu.
\] (4.23)
(4) If $f$ is integrable, then
\[
\left| \int f \, d\mu \right| \leq (\sup_{x \in X} |f(x)|) \mu(X).
\] (4.24)

Proof. (0) If $f$ is simple and $\phi \leq f \leq \phi'$, then $\int \phi \, d\mu \leq \int f \, d\mu \leq \int \phi' \, d\mu$. This shows that the upper integral of $f$ is at least $\int f \, d\mu$ while the lower integral is at most $\int f \, d\mu$. But $f$ is an upper/lower bound on itself, so the two integrals actually coincide and equal $\int f \, d\mu$ as claimed.

(1) If $\phi_1, \phi_2$ are simple functions such that $\phi_1 \leq f \phi_2$ and $\phi'_1, \phi'_2$ are simple functions such that $\phi'_1 \leq g \leq \phi'_2$, then $\phi_1 + \phi_2 \leq f + g \leq \phi'_1 + \phi'_2$. Hence we get
\[
\int f \, d\mu + \int g \, d\mu \leq \int (f + g) \, d\mu \leq \int f \, d\mu + \int g \, d\mu,
\] (4.25)
where all integrals are finite. So if both $f$ and $g$ are integrable, the extreme left and right hand sides coincide and the claim follows.

(2) If $c \geq 0$ and $\phi_1 \leq f \leq \phi_2$ then $c\phi_1 \leq cf \leq c\phi_2$. This shows
\[
c \int f \, d\mu \leq \int (cf) \, d\mu \leq \int (cf) \, d\mu \leq c \int f \, d\mu.
\] (4.26)

If $f$ is integrable, we right and left hand sides are equal. For $c < 0$ the proof is completely analogous except that $\phi_1 \leq f \leq \phi_2$ now implies $c\phi_2 \leq cf \leq c\phi_1$. The same argument follows with the roles of upper/lower integrals interchanged in appropriate places.

(3) Let $f \leq g$. If $\phi$ is simple with $\phi \leq f$ then also $\phi \leq g$. It follows that
\[
\int f \, d\mu \leq \int g \, d\mu.
\] (4.27)

For integrable $f$ and $g$, this induces a similar comparison of the Lebesgue integrals.

(4) If $\phi := \sup_{x \in X} |f(x)|$, then $\phi$ is simple with $-\phi \leq f \leq \phi$. It follows that
\[
-(\sup_{x \in X} |f(x)|) \mu(X) = \int (-\phi) \, d\mu \leq \int f \, d\mu \leq \int f \, d\mu \leq \int \phi \, d\mu = (\sup_{x \in X} |f(x)|) \mu(X).
\] (4.28)

For $f$ integrable, this wraps up into the claimed inequality. \hfill \Box

The properties in the above lemma can be summarized by saying that the map $f \mapsto \int f \, d\mu$ is a positive bounded linear functional on the linear space of bounded integrable functions.
In integration theory one sometimes wants to integrate \( f \) only on a subset of the underlying space. For this we note:

**Lemma 4.7** Let \( \mu(X) < \infty \) and let \( f : X \to \mathbb{R} \) be a bounded integrable function. Then \( f1_A \) is integrable for all \( A \in \mathcal{F} \) as well.

**Proof.** Let \( A \in \mathcal{F} \). Since \( f \) is integrable, for each \( \varepsilon > 0 \) there are simple functions \( \phi, \phi' \) such that \( \phi \leq f \leq \phi' \) and \( \int \phi' d\mu - \int \phi d\mu < \varepsilon \). But then both \( \phi 1_A \) and \( \phi' 1_A \) are simple functions with

\[
\int (\phi' 1_A) d\mu - \int (\phi 1_A) d\mu = \int ((\phi' - \phi) 1_A) d\mu \leq \int (\phi' - \phi) d\mu < \varepsilon.
\]

(4.29)

So \( f1_A \) is integrable as well. \( \Box \)

In light of this observation, we put forward:

**Definition 4.8** Given any \( A \in \mathcal{F} \), we will say that \( f : X \to \mathbb{R} \) is integrable on \( A \) whenever \( f1_A \) is integrable. In this case we write

\[
\int_A f d\mu := \int (f1_A) d\mu.
\]

(4.30)

It is a straightforward exercise, albeit somewhat lengthy, to prove:

**Lemma 4.9** Let \( (X, \mathcal{F}, \mu) \) be a measure space and let \( A \in \mathcal{F} \). Define

\[
\mathcal{F}_A := \{ B \cap A : B \in \mathcal{F} \} \quad \text{and} \quad \mu_A(B) := \mu(A \cap B).
\]

(4.31)

Then \( (X, \mathcal{F}_A, \mu_A) \) is a measure space. Moreover, if \( \mu(X) < \infty \) and \( f : X \to \mathbb{R} \) is bounded and integrable on \( A \) (with respect to \( \mu \)), then the restriction \( f_A \) of \( f \) to \( A \) is integrable with respect to \( \mu_A \) and

\[
\int_A f d\mu = \int f_A d\mu_A.
\]

(4.32)

We leave the proof of this lemma to the reader.

### 4.2 Integrability vs measurability.

We note that, in all of the above statements, it would suffice to have \( \mu \) only finitely additive on \( \mathcal{F} \) with \( \mu(X) < \infty \). (Reformulating the finiteness condition appropriately, one can even use finitely additive set functions taking both positive and negative values; cf the forthcoming chapters.) However, even if we use that \( \mu \) is actually countably additive, it is not clear that the integral behaves naturally under point-wise limits. Explicitly, if uniformly bounded integrable functions \( f_n \) converge pointwise to a bounded function \( f \), it is not apparent whether or not \( f \) is integrable and whether or not the limit of integrals is the integral of the limit.

Another natural question is whether manageable regularity conditions can at all be produced that guarantee integrability. (Remember that this was a serious issue surrounding the otherwise very elegant definition of the Riemann integral.) As it turns out, a proper condition that takes care of both problems is *measurability*. One can arrive at this concept quite naturally by recalling one of the main innovation that Lebesgue brought to integration theory: Instead of defining the
integral by partitioning the domain of $f$ (that is, the set $X$) we partition the range of values of $f$ and use the pre-image map,

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

(4.33)

to induce a partition of $X$. Naturally, if one is then to assign a meaning to the resulting “piece-wise” constant approximation of $f$, all sets in the induced partition need to be measurable. This in turn forces $f$ to belong to the the class of functions described by:

**Definition 4.10** Let $(X, \mathcal{F})$ be a measurable space and $f: X \to \mathbb{R}$ a function. We say that $f$ is measurable (or $\mathcal{F}$-measurable, if the underlying $\sigma$-algebra needs to be emphasized) if

$$\forall a < b: \quad f^{-1}((a,b]) \in \mathcal{F}.$$  

(4.34)

If $f$ takes values in the generalized real numbers, $\mathbb{R} \cup \{ \pm \infty \}$, then we require this for $a < b$ including $a = -\infty$ and $b = \infty$.

The restriction to (preimages of) half-open intervals is quite arbitrary; indeed, we just need a class of sets that generates enough sets by countable unions and complements. Indeed, we have:

**Lemma 4.11** Let $(X, \mathcal{F})$ be a measurable space and $f: X \to \mathbb{R}$. Then

$$\{ B \subset \mathbb{R} : f^{-1}(B) \in \mathcal{F} \}$$

(4.35)

is a $\sigma$-algebra.

**Proof.** We have $f^{-1}(A^c) = f^{-1}(A)^c$ and $f^{-1}(\bigcup_{i \in \mathbb{N}} A_i) = \bigcup_{i \in \mathbb{N}} f^{-1}(A_i)$. Hence the stated collection is closed under complements and countable unions. Since also $f^{-1}(\emptyset) = \emptyset$ and $f^{-1}(\mathbb{R}) = X$, this collection is a $\sigma$-algebra. \hfill \Box

As a consequence, the same concept would be defined if we just required

$$f^{-1}((-\infty,a]) \in \mathcal{F}, \quad a \in \mathbb{R}.$$  

(4.36)

or

$$f^{-1}((-\infty,a)) \in \mathcal{F}, \quad a \in \mathbb{R}.$$  

(4.37)

(In fact, requiring this just for $a \in \mathbb{Q}$ would suffice.) Notice that, in particular, every level set $f^{-1}(\{a\})$ of a measurable function $f$ is measurable.

Obviously, all simple functions are automatically measurable. Many natural operations on functions preserve measurability. Indeed, we have:

**Lemma 4.12** We have:

1. If $f, g$ are measurable then so is $f + g$.
2. If $f$ is measurable and $c \in \mathbb{R}$, then $cf$ is also measurable.
3. If $f, g$ are measurable and $A \in \mathcal{F}$, then also $f1_A + g1_A$ is measurable.

**Proof.** (1) Let us write $f^{-1}(A) = \{ f \in A \}$. Then

$$\{ f + g < a \} = \bigcup_{p,q \in \mathbb{Q}} \{ f < p \} \cap \{ g < q \}.$$  

(4.38)

The sets $\{ f < p \}, \{ g < q \} \in \mathcal{F}$ when $f, g$ are measurable, and since the union is countable, $\{ f + g < a \} \in \mathcal{F}$ for any $a \in \mathbb{R}$ as well. It follows that $f + g$ is measurable.
(2) Here we just note that, for $c > 0$, we have $\{ cf < a \} = \{ f < a/c \}$, while for $c < 0$ we have $\{ cf < a \} = \{ f > a/c \}$. Since $f := 0$ is also measurable, we are done.

(3) Let $h := f1_A + g1_A^c$. Then

$$\{ h < a \} = (A \cap \{ f < a \}) \cup (A^c \cap \{ g < a \})$$

and all sets on the right are clearly measurable.

The class of measurable functions is a vector space. Perhaps more importantly, measurability behaves quite well under pointwise limits. This includes the cases when the limit is plus/minus infinity, but for this we need to use the proviso in the the above definition that deals with functions taking values in generalized reals.

**Lemma 4.13** If $\{ f_n \}$ are measurable, then so are the functions

$$\sup_{n \in \mathbb{N}} f_n, \quad \inf_{n \in \mathbb{N}} f_n, \quad \limsup_{n \to \infty} f_n, \quad \liminf_{n \to \infty} f_n.$$  \hspace{1cm} (4.40)

In particular, if $f := \lim_{n \to \infty} f_n$ exists, then $f$ is also measurable.

**Proof.** For any $a \in \mathbb{R} \cup \{ \pm \infty \}$ we have

$$\{ \sup_{n \in \mathbb{N}} f_n > a \} = \bigcup_{n \in \mathbb{N}} \{ f_n > a \}$$

and so the supremum of $f_n$ is measurable (as a function taking values in $\mathbb{R} \cup \{ \pm \infty \}$) if $\{ f_n \}$ are measurable. Similarly, we get also that the infimum is measurable as well. As

$$\limsup_{n \to \infty} f_n = \inf_{n \geq 1, m \geq n} \sup_{m \geq n} f_m$$

and

$$\liminf_{n \to \infty} f_n = \sup_{n \geq 1, m \geq n} \inf_{m \geq n} f_m$$

these are measurable as well. If limit $f := \lim_{n \to \infty} f_n$ exists, then it equals the *limes superior* (as well as the *limes inferior*) and so it is measurable as well.

We are now ready to tie measurability with the concept of integrability introduced earlier. One direction is quite immediate:

**Proposition 4.14** (Measurability implies integrability) Assume $\mu(X) < \infty$ and let $f : X \to \mathbb{R}$ be a bounded function. Then $f$ measurable $\Rightarrow$ $f$ integrable. \hspace{1cm} (4.43)

**Proof.** Let $f : X \to \mathbb{R}$ be a bounded measurable function. Pick $n \in \mathbb{N}$ and consider the functions

$$\phi_n := \sum_{i \in \mathbb{Z}} \frac{i}{n} f^{-1}\left(\left[\frac{i}{n}, \frac{i+1}{n}\right]\right) \quad \text{and} \quad \phi'_n := \sum_{i \in \mathbb{Z}} \frac{i+1}{n} f^{-1}\left(\left[\frac{i}{n}, \frac{i+1}{n}\right]\right)$$

(4.44)

We write these as infinite sums but, since $f$ is bounded, the preimage set is non-empty only for a finite number if $i$'s. Since $f$ is measurable, the preimages lie in $\mathcal{F}$ and so both $\phi_n$ and $\phi'_n$ are simple. Now $\phi_n \leq f \leq \phi'_n$ and so, by the definition of the upper/lower Lebesgue integrals,

$$\int \phi_n \, d\mu \leq \int f \, d\mu \leq \int f \, d\mu \leq \int \phi'_n \, d\mu.$$ 

(4.45)
Moreover, \( 0 \leq \phi'_n - \phi_n \leq \frac{1}{n} \) and Lemma 4.6(1,2) thus shows
\[
0 \leq \int \phi'_n \, d\mu - \int \phi_n \, d\mu \leq \frac{1}{n} \mu(X) \quad (4.46)
\]
Putting these together, we conclude
\[
0 \leq \int f \, d\mu - \int f \, d\mu \leq \frac{1}{n} \mu(X). \quad (4.47)
\]
Taking \( n \to \infty \) shows that \( f \) is integrable. \( \square \)

Proposition 4.14 now permits us to state an inequality that will be useful in the following:

**Lemma 4.15** (Markov’s inequality) \( \) Suppose \( \mu(X) < \infty \) and let \( f : X \to [0, \infty) \) be bounded and measurable. Then for all \( \lambda > 0 \),
\[
\mu(\{x : f(x) \geq \lambda\}) \leq \frac{1}{\lambda} \int f \, d\mu. \quad (4.48)
\]

**Proof.** Let \( \lambda > 0 \). The function \( \phi := \lambda 1_{\{f \geq \lambda\}} \) is simple because \( f \) is measurable and obeys \( \phi \leq f \). Hence
\[
\lambda \mu(\{x : f(x) \geq \lambda\}) = \int \phi \, d\mu \leq \int f \, d\mu. \quad (4.49)
\]
The claim follows by dividing both sides by \( \lambda \). \( \square \)

As it turns out, a converse to Proposition 4.14 holds as well assuming one additional property of the underlying measure space. This condition is given in:

**Definition 4.16** (Completeness) \( \) A measure space \( (X, \mathcal{F}, \mu) \) — or just the measure \( \mu \) or the \( \sigma \)-algebra \( \mathcal{F} \) — is said to be complete if, whenever \( N \in \mathcal{F} \) obeys \( \mu(N) = 0 \), we have \( A \in \mathcal{F} \) for every \( A \subset N \). A set \( N \in \mathcal{F} \) with \( \mu(N) = 0 \) is called a null set.

We remark that the Lebesgue measure we constructed earlier is automatically complete. As we will see, this is common for all constructions involving outer measures.

**Proposition 4.17** (Integrability and completeness imply measurability) \( \) Assume \( \mu(X) < \infty \) and let \( f : X \to \mathbb{R} \) be a bounded function. If \( (X, \mathcal{F}, \mu) \) is complete, then \( f \) integrable \( \Rightarrow \) \( f \) measurable. \( (4.50) \)

**Proof.** Suppose that \( f \) is integrable. Then, for each \( n \in \mathbb{N} \), there are simple functions \( \phi_n \) and \( \phi'_n \) (not necessarily those above, of course) with
\[
\phi_n \leq f \leq \phi'_n \quad \text{and} \quad 0 \leq \int \phi'_n \, d\mu - \int \phi_n \, d\mu \leq \frac{1}{n}. \quad (4.51)
\]
The maximum of any finite number of simple functions is simple and so we may replace \( \phi_n \) by \( \max\{\phi_1, \ldots, \phi_n\} \) and get that \( \phi_n \) is increasing. Similarly, we may assume that \( \phi'_n \) is decreasing and the bound (4.51) still holds. The limits
\[
g := \lim_{n \to \infty} \phi_n \quad \text{and} \quad h := \lim_{n \to \infty} \phi'_n \quad (4.52)
\]
then exist and obey \( g \leq f \leq h \).
Since simple functions are automatically measurable, Lemma 4.13 implies that both \( g \) and \( h \) are measurable. As \( g \) and \( h \) squeeze \( f \) in between, and were obtained by monotone limits of simple functions, both are also bounded. Proposition 4.14 ensures that they are integrable. But then \( \phi_n \leq g \leq h \leq \phi'_n \) and Lemma 4.6(3) imply
\[
\int \phi_n \, d\mu \leq \int g \, d\mu \leq \int h \, d\mu \leq \int \phi'_n \, d\mu \tag{4.53}
\]
and so, by (4.51),
\[
\int g \, d\mu = \int h \, d\mu. \tag{4.54}
\]
But \( h - g \) is non-negative and measurable by Lemma 4.12, so Markov’s inequality (Lemma 4.15) and Lemma 4.6 yield
\[
\mu(h - g \geq \varepsilon) \leq \frac{1}{\varepsilon} \int (h - g) \, d\mu = \frac{1}{\varepsilon} \int h \, d\mu - \frac{1}{\varepsilon} \int g \, d\mu = 0, \quad \varepsilon > 0. \tag{4.55}
\]
It follows that \( \mu(h - g \geq \varepsilon) = 0 \) for any \( \varepsilon > 0 \) and, by taking \( \varepsilon \downarrow 0 \) along a sequence and applying the Upward Monotone Convergence Theorem for sets,
\[
\mu(h - g > 0) = 0. \tag{4.56}
\]
But \( f \) is squeezed between \( g \) and \( h \) and so \( f \) equals \( g \) (and \( h \)) whenever \( g = h \). Hence \( f \) equals \( g \) (and \( h \)) on a set of full measure. To prove that \( f \) is measurable, pick \( a \in \mathbb{R} \) and observe
\[
\{f < a\} = \{g < a\} \cap \{h - g = 0\} \cup \{f < a\} \cap \{h - g > 0\}. \tag{4.57}
\]
The first set in the union is measurable because \( g \) and \( h - g \) are measurable. The second set in the union is a subset of \( \{h - g > 0\} \), which is a set of zero \( \mu \)-measure. Since \((X, \mathcal{F}, \mu)\) is complete, any subset of \( \{h - g > 0\} \) is measurable as well. It follows that \( f \) is measurable. \( \square \)

In complete measure spaces, integrability is thus equivalent to measurability. In non-complete spaces this is not true in general, although the proof does imply that every integrable function can be modified on a null set to produce a measurable function. As we will later often regard functions that differ only on subsets of null sets as equivalent, the notions of integrability and measurability blend together as well.

4.3 Bounded Convergence Theorem.

With the above characterization in hand, we can now examine the behavior of the bounded integral under pointwise limits. We in fact do not need sequences of functions to converge everywhere, it suffices to have weaker versions of this convergence. We first introduce a general concept of a property being true \textit{almost everywhere}:

**Definition 4.18** Let \( P(x) \) be a property — i.e., a boolean variable — assigned to a point \( x \) and let \( \mu \) be a measure on a measurable space \((X, \mathcal{F})\). We say that a property \( P \) holds \( \mu \)-almost everywhere, or just \( \mu \)-a.e., if the set \( \{x \in X : P(x) \text{ fails}\} \) is a subset of a \( \mu \)-null set.

Note that in complete spaces we may require that \( \{x \in X : P(x) \text{ fails}\} \) is a null set. We can now apply this general notion to the concept of convergence:
Definition 4.19 We say that \( f_n \) converges to \( f \) almost everywhere, or “\( f_n \to f \) a.e.,” if the set
\[
\{ x \in X : \limsup_{n \to \infty} |f_n(x) - f(x)| > 0 \}
\]
is a subset of a null set.

A yet weaker version of convergence is:

Definition 4.20 Let \( \{f_n\} \) and \( f \) be measurable functions. We say that \( f_n \to f \) in measure if
\[
\forall \epsilon > 0 : \quad \mu(\{|f_n - f| > \epsilon \}) \to 0, \quad n \to \infty.
\]

Here we note that \( \{|f_n - f| > \epsilon \} \) is a measurable set because (by Lemma 4.12) if \( f \) is measurable then so is \( |f| = f^+1_{\{f \geq 0\}} - f^-1_{\{f < 0\}} \). We now note:

Lemma 4.21 Let \( \{f_n\} \) and \( f \) be measurable functions on a measure space \((X, \mathcal{F}, \mu)\). Then:

1. If \( f_n \to f \) pointwise, then \( f_n \to f \) almost everywhere.
2. If \( f_n \to f \) almost everywhere, then \( f_n \to f \) in measure.
3. If \( f_n \to f \) in measure, then there is an increasing sequence \( \{n_k\} \subset \mathbb{N} \) such that \( f_{n_k} \to f \) almost everywhere.

Proof. (1) is true trivially. For (2) we note that
\[
\mu(\{|f_n - f| \geq \epsilon \}) \leq \mu(\sup_{m \geq n} |f_m - f| \geq \epsilon) \to \mu(\limsup_{n \to \infty} |f_n - f| \geq \epsilon),
\]
where we used the Monotone Convergence Theorem for sets to get the last step. If \( f_n \to f \) a.e., then the right-hand side is zero for every \( \epsilon > 0 \). For (3) we define \( \{n_k\} \) inductively by \( n_0 := 1 \) and
\[
n_k := \inf \left\{ n > n_{k-1} : \mu(\{|f_n - f| > 2^{-k} \}) < 2^{-k} \right\}, \quad k \in \mathbb{N},
\]
where we used that \( f_n \to f \) in measure to observe that the set in infimum is non-empty (in fact, it is co-finite in \( \mathbb{N} \)). We thus have \( \mu(\{|f_n - f| > 2^{-k} \}) < 2^{-k} \) for all \( k \in \mathbb{N} \) and since
\[
\limsup_{k \to \infty} |f_{n_k} - f| \geq \epsilon \subseteq \limsup_{k \to \infty} \bigcup_{i \geq k} \{|f_n - f| > \epsilon \},
\]
we get
\[
P(\limsup_{k \to \infty} |f_{n_k} - f| > \epsilon) \leq \lim_{k \to \infty} P\left(\bigcup_{i \geq k} \{|f_n - f| > \epsilon \}\right)
\leq \lim_{k \to \infty} \sum_{i \geq k} P(|f_n - f| > 2^{-i}) = \lim_{k \to \infty} \sum_{i \geq k} 2^{-i} = 0.
\]
Here we first used the definition of limes superior for sets, then applied a union bound along with the assumption that \( P(|f_n - f| > \epsilon) \leq P(|f_n - f| > 2^{-i}) \) for all \( i \geq k \) once \( \epsilon \geq 2^{-k} \) and then used the defining property of \( n_k \). Taking \( \epsilon \downarrow 0 \) along a sequence and applying the Upward Monotone Convergence for sets, we get that \( f_{n_k} \to f \) a.e. \( \square \)

It is not hard to check that all three types of convergence discussed above are distinct from each other. We leave the construction of requisite counterexamples to the reader. A key point is that our next convergence theorem applies to the weakest of these three.
Theorem 4.22 (Bounded Convergence Theorem)

Suppose \( \mu(X) < \infty \) and let \( \{f_n\} \) and \( f \) be bounded measurable functions with \( \sup_{n \in \mathbb{N}} \sup_{x \in X} |f_n(x)| < \infty \). Then

\[
f_n \to f \text{ in measure } \implies \int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu. \tag{4.64}
\]

Proof. Let \( K \) be a constant such that \( |f_n| \leq K \) and \( |f| \leq K \). Then for any \( \varepsilon > 0 \),

\[
\left| \int f_n \, d\mu - \int f \, d\mu \right| = \left| \int (f_n - f) \, d\mu \right| \\
\leq \left| \int (f_n - f)^1_{|f_n - f| > \varepsilon} \, d\mu \right| + \left| \int (f_n - f)^1_{|f_n - f| \leq \varepsilon} \, d\mu \right| \tag{4.65}
\]

where we first used additivity of the integral, then wrote \( 1 = 1_{|f_n - f| > \varepsilon} + 1_{|f_n - f| \leq \varepsilon} \) and applied additivity of integral and subadditivity of absolute value. Then, in the first absolute value, we bounded \( (f_n - f)^1_{|f_n - f| > \varepsilon} \) between \( \phi := 2K1_{|f_n - f| > \varepsilon} \) and \( -\phi \), while in the second part we applied Lemma 4.6(4) along with the fact that the integrated function is at most \( \varepsilon \). The claim now follows by taking \( n \to \infty \) followed by \( \varepsilon \downarrow 0 \).

Even if the limit function \( f \) is bounded, the uniform boundedness of the sequence \( \{f_n\} \) cannot be omitted. Indeed, in the measure space \( ([0, 1], \mathcal{L}'([0, 1]), \lambda) \), the functions \( f_n(x) := n1_{[0,1/n]}(x) \) converge to zero in Lebesgue measure and yet \( \int f_n \, d\lambda = 1 \).

### 4.4 Characterization of Riemann integrability.

A natural question to ask at this point is to what extent does the Lebesgue integral generalize the Riemann integral. This will be answered quite elegantly by an argument similar to that used in the proof of Proposition 4.17. We will henceforth work with functions on a finite interval \( [a, b] \subset \mathbb{R} \) which we naturally endow with the \( \sigma \)-algebra \( \mathcal{L}'([a, b]) \) of Lebesgue-measurable subsets thereof and the Lebesgue measure \( \lambda \). To reduce clutter of notation, we write the resulting Lebesgue integral as \( \int_{[a,b]} f \, d\lambda \); see (4.30).

We begin by stating the main theorem. Recall that Definition 4.18 assigns a unique meaning to the phrase “\( f \) is continuous a.e.” Using this concept, we have:

Theorem 4.23 (Lebesgue’s characterization of Riemann integrability)

Let \( f: [a, b] \to \mathbb{R} \) be a bounded function. Then

\[
f \text{ is Riemann integrable } \iff f \text{ is continuous Lebesgue a.e.} \tag{4.66}
\]

The proof will consist of several observations that we state as lemmas. Consider a bounded function \( f: [a, b] \to \mathbb{R} \). We will have to develop a good handle of the set \( \{x: f \text{ is continuous at } x\} \).

For this we define

\[
m_\delta(x) := \inf \{ f(y) : y \in [a, b], |y-x| < \delta \} \\
M_\delta(x) := \sup \{ f(y) : y \in [a, b], |y-x| < \delta \} \tag{4.67}
\]

and set

\[
m(x) := \lim_{\delta \downarrow 0} m_\delta(x) \quad \text{and} \quad M(x) := \lim_{\delta \downarrow 0} M_\delta(x). \tag{4.68}
\]

245A lecture notes (c 2017 M. Biskup) Preliminary version (typeset: September 22, 2017)
These limits exist because (as is easy to check) $\delta \mapsto m_\delta(x)$ is non-decreasing while $\delta \mapsto M_\delta(x)$ is non-increasing and

$$m_\delta(x) \leq f(x) \leq M_\delta(x), \quad \delta > 0. \quad (4.69)$$

The connection with continuity of $f$ is now provided by:

**Lemma 4.24** Let $f: [a,b] \to \mathbb{R}$ be a bounded function and let $m(x)$ and $M(x)$ be defined as above. Then we have:

1. $m(x) \leq f(x) \leq M(x)$ for all $x \in [a,b]$.
2. $f$ is continuous at $x \in [a,b]$ if and only if $m(x) = M(x)$.
3. $M$ is upper-semicontinuous and $m$ is lower-semicontinuous on $[a,b]$, and thus
4. both $m$ and $M$ are Lebesgue measurable.

We note that a function $g$ is lower-semicontinuous if $\{g > a\}$ is open for all $a$, and is upper-semicontinuous if $\{g < a\}$ is open for all $a$. Alternatively, $g$ is lower-semicontinuous if $x_n \to x$ implies $g(x) \leq \liminf_{n \to \infty} g(x_n)$, etc.

**Proof of Lemma 4.24.** (1) is immediate from (4.69). For (2) we note that $f$ is continuous at $x$ if and only if for each $\epsilon > 0$ there is $\delta > 0$ such that (the oscillation of $f$ obeys)

$$\text{osc}_{A_\delta(x)}(f) < \epsilon \quad \text{for} \quad A_\delta(x) := \{y \in [a,b]: |y - x| < \delta\}. \quad (4.70)$$

But $\text{osc}_{A_\delta(x)}(f) = M_\delta(x) - m_\delta(x)$ and so the claim follows.

For (3) we note that, $M(x) < a$ implies $M_\delta(x) < a$ for some $\delta > 0$ small. But then also $M_\delta(y) < a$ for $y$ such that $|y - x| < \delta$. Hence $\{M < a\}$ is open and so $M$ is upper-semicontinuous. The lower-semicontinuity of $m$ is proved analogously.

Finally, to get (4) we note that a lower, resp., upper-semicontinuous function preimages intervals $(a, \infty)$, resp., $(-\infty, a)$ into open sets which are ex definitio Lebesgue measurable. Invoking Lemma 4.11 — and also the alternative conditions (4.36–4.37) for measurability afterwards — we conclude that both $M$ and $m$ are Lebesgue measurable.

We now define another concept:

**Definition 4.25** A step function $\phi$ on $[a,b]$ is a function of the form

$$\phi(x) := \sum_{i=1}^{n} a_i \mathbb{1}_{I_i} \quad (4.71)$$

where $a_1, \ldots, a_n \in \mathbb{R}$ and $\{I_i\}$ are disjoint intervals such that $\bigcup_{i=1}^{n} I_i = [a,b]$.

Note that every step function is automatically simple. A key point is that, for $f: [a,b] \to \mathbb{R}$ bounded, the Darboux-Riemann lower/upper integrals are given by

$$\int_{a}^{b} f(x) \, dx = \sup \left\{ \int_{[a,b]} \phi \, d\lambda : \phi \text{ step function, } \phi \leq f \right\} \quad (4.72)$$

and

$$\int_{a}^{b} f(x) \, dx = \inf \left\{ \int_{[a,b]} \phi \, d\lambda : \phi \text{ step function, } \phi \geq f \right\}, \quad (4.73)$$

where we notice that a step function is Lebesgue measurable and the Lebesgue integral $\int_{[a,b]} \phi \, d\lambda$ is exactly the value of the Riemann sum associated with $\phi$. From this we immediately conclude:
Lemma 4.26 (Riemann integrability implies Lebesgue integrability) For \( f : [a, b] \to \mathbb{R} \) bounded,

\[
\int_a^b f(x) \, dx \leq \int (f1_{[a,b]}) \, d\lambda \leq \int (f1_{[a,b]}) \, d\lambda \leq \int_a^b f(x) \, dx. \tag{4.74}
\]

In particular, if \( f \) is Riemann integrable on \([a, b]\), then it is Lebesgue integrable on \([a, b]\).

Proof. Every step function is a simple function and so we get (4.74) by comparing (4.72–4.73) with (4.16–4.17). If \( f \) is Riemann integrable on \([a, b]\), the extreme ends of (4.74) coincide and so all inequalities are in fact equalities. In particular, \( f1_{[a,b]} \) is Lebesgue integrable.

The connection to the above functions \( m \) and \( M \) is now provided by:

Lemma 4.27 Let \( f : [a, b] \to \mathbb{R} \) be bounded. If \( \phi, \phi' \) are step functions such that \( \phi \leq f \leq \phi' \), then also \( \phi \leq m \leq M \leq \phi' \) at the continuity points of both \( \phi \) and \( \phi' \). In particular, we have

\[
\int_a^b f(x) \, dx = \int_{[a,b]} m \, d\lambda \quad \text{and} \quad \int_a^b f(x) \, dx = \int_{[a,b]} M \, d\lambda \quad \tag{4.75}
\]

Proof. If \( \phi \) is a step function such that \( \phi \leq f \) and \( x \) is a continuity point of \( \phi \), then \( \phi \) is constant on a \( \delta \)-neighborhood of \( x \). But then \( \phi \) lies below the infimum of \( f \) on this neighborhood and so \( \phi(x) \leq m_\delta(x) \leq m(x) \). This implies the first clause of the lemma.

Since \( \lambda \) assigns zero mass to singletons, \( \phi \leq f \leq \phi' \) thus implies

\[
\int_{[a,b]} \phi \, d\lambda \leq \int_{[a,b]} m \, d\lambda \quad \text{and} \quad \int_{[a,b]} \phi' \, d\lambda \geq \int_{[a,b]} M \, d\lambda. \quad \tag{4.76}
\]

Then we have

\[
\int_a^b f(x) \, dx \leq \int_{[a,b]} m \, d\lambda \quad \text{and} \quad \int_a^b f(x) \, dx \geq \int_{[a,b]} M \, d\lambda. \quad \tag{4.77}
\]

by a direct application of (4.72–4.73).

To prove that equalities take place, consider the collection \( \{I_i : i = 1, \ldots, 2^n\} \) of the dyadic intervals of the form

\[
[a,a+2^{-n}(b-a)], (a+2^{-n}(b-a),a+2^{1-n}(b-a)], \\
\ldots, (a+k2^{-n}(b-a),a+(k+1)2^{-n}(b-a)], \ldots, (b-2^{-n}(b-a),b]. \tag{4.78}
\]

(Notice that the first interval is closed; all others half-open.). The endpoints of these intervals lie in the set \( \mathbb{D}_n := \{k2^{-n} : k = 0, \ldots, 2^n\} \). Note that \( \bigcup_{i=1}^{2^n} I_i = [a,b] \) and so

\[
\phi_n := \sum_{i=1}^{2^n} (\inf_{x \in I_i} f(x))1_{I_i} \quad \text{and} \quad \phi_n' := \sum_{i=1}^{2^n} (\sup_{x \in I_i} f(x))1_{I_i} \quad \tag{4.79}
\]

are step functions. Let \( \mathbb{D} := \bigcup_{n \in \mathbb{N}} \mathbb{D}_n \). Since \( \phi_n \leq f \leq \phi_n' \) are continuous on \( \mathbb{D} \), the first clause in the lemma and the definition of \( m_\delta \) and \( M_\delta \) show

\[
\phi_n \leq m \quad \text{and} \quad \phi_n \to m \quad \text{on} \ \mathbb{D}.
\]

\[
\phi_n' \geq M \quad \text{and} \quad \phi_n' \to M \quad \text{on} \ \mathbb{D}. \tag{4.80}
\]
Since $f$ is bounded, the functions $\{\phi_n\}, \{\phi'_n\}, m, M$ are uniformly bounded. Moreover, $\mathbb{D}$ is countable and so the convergence $\phi_n \to m$ and $\phi'_n \to M$ takes place Lebesgue almost everywhere, and thus in measure. Invoking again (4.72–4.73) we get

$$\int_a^b f(x) \geq \limsup_{n \to \infty} \int_{[a,b]} \phi_n \, d\lambda = \int_{[a,b]} m \, d\lambda$$

(4.81)

and similarly

$$\int_a^b f(x) \leq \liminf_{n \to \infty} \int_{[a,b]} \phi_n \, d\lambda = \int_{[a,b]} M \, d\lambda,$$

(4.82)

where the limit is a consequence of the Bounded Convergence Theorem. The claim follows. □

We are now finally ready to prove the desired result:

**Proof of Theorem 4.23.** Suppose $f$ is Riemann integrable. By Lemma 4.27, the Lebesgue integrals of $M$ and $m$ with respect to the Lebesgue measure on $[a, b]$ are equal. Hence, by Markov’s inequality (Lemma 4.15), for each $\varepsilon > 0$ we have

$$\mu(M - m > \varepsilon) \leq \frac{1}{\varepsilon} \int_{[a,b]} (M - m) \, d\lambda = 0.$$  

(4.83)

Taking $\varepsilon \downarrow 0$, the Upward Monotone Convergence Theorem for sets shows $\mu(M - m > 0) = 0$. Lemma 4.24 implies that $f$ is continuous a.e.

On the other hand, if $f$ is bounded and continuous a.e., then also $M = m$ a.e. by Lemma 4.24. Then the Lebesgue integrals of (bounded functions) $m$ and $M$ with respect to (finite) Lebesgue measure on $[a, b]$ are equal. By Lemma 4.27 again, $f$ is Riemann integrable. □
5. **Lebesgue Integral: General Case**

Although the Lebesgue integral defined in the previous chapter is in many ways much better behaved than the Riemann integral, it shares its restriction to bounded functions and bounded measure spaces only. To remove this restriction one might proceed similarly as for the improper Riemann integral; e.g., by taking limits of

\[
\int_{A_n} (f \land M \lor M') \, d\mu
\]  

as \( M \to \infty, M' \to -\infty \) and \( n \to \infty \) for some increasing sequence \( \{A_n\} \) of measurable sets with \( \mu(A_n) < \infty \) and \( \bigcup_{n \in \mathbb{N}} A_n = X \). Unfortunately, the existence of such a limit is not guaranteed without further conditions on the integrand, and we may not even be able to find a sequence \( \{A_n\} \) with the stated properties.

The way around this that has generally been adopted is to first define the integral of non-negative functions on arbitrary measure spaces. This is the **unsigned integral** below. A key difference between the “bounded integral” from the previous chapter is that here we abandon integrability as a primary concept and rather put measurability first. Notwithstanding, the bounded integral and approximations of the form (5.1) are key to proving properties such as additivity etc.

Once the unsigned integral is defined and its properties established, a more or less formal extension leads to the **signed integral** that covers also functions that can take both positive and negative values.

### 5.1 Unsigned integral.

Let \( (X, \mathcal{F}, \mu) \) be a measure space with \( \mu(X) \) not necessarily finite. We now put forward:

**Definition 5.1** (Lower unsigned integral) *Let \( f : X \to [0, \infty] \). Then we set:*

\[
\underline{\int f \, d\mu} := \sup \left\{ \int \phi \, d\mu : \phi \text{ simple, } 0 \leq \phi \leq f \right\}.
\]  

(5.2)

This is the analogue of what we called the lower Lebesgue integral in the bounded case. The corresponding upper integral will not be defined at all. Observe the following properties (shared, generally, by all lower integrals):

**Lemma 5.2** *Let \( f, g \) be functions on \( X \). Then we have:*

1. *if \( 0 \leq f \leq g \) then*

\[
\underline{\int f \, d\mu} \leq \underline{\int g \, d\mu},
\]  

(5.3)

2. *if \( f, g \geq 0 \) then*

\[
\underline{\int (f + g) \, d\mu} \geq \underline{\int f \, d\mu} + \underline{\int g \, d\mu}
\]  

(5.4)

3. *if \( c \in [0, \infty) \) and \( f \geq 0 \) then (using the convention \( 0 \cdot \infty = 0 \))*

\[
\underline{\int (cf) \, d\mu} = c \underline{\int f \, d\mu}.
\]  

(5.5)
Proof. Let $\phi$ be a simple function with $0 \leq \phi \leq f$. For (1) we note that $f \leq g$ then implies $0 \leq \phi \leq g$. For (2) we note that if also $0 \leq \phi' \leq g$ is simple, then $\phi + \phi' \leq f + g$. Finally, for (3) we observe that $c\phi$ is simple with $c\phi \leq f$. □

It is not hard to check that, although the lower integral is superadditive, it is generally not additive unless further conditions on $f$ and $g$ are imposed. It turns out that, as before, it is enough to require measurability. (Since we permit functions to take value $+\infty$, we mean the notion of measurability for functions taking extended-real values.) What needs to be proved is:

**Proposition 5.3** (Superadditivity on measurable functions) Let $f, g: X \to [0, \infty]$ be measurable. Then we have:

$$\int (f + g) \, d\mu \leq \int f \, d\mu + \int g \, d\mu$$ (5.6)

First we observe that if we reduce the unsigned lower integral to bounded functions with bounded measure of their support,

$$\text{supp}(f) := \{x \in X : f(x) \neq 0\},$$ (5.7)

then additivity can be proved directly. This is because we have:

**Lemma 5.4** Let $f: X \to [0, \infty]$ be bounded, measurable and assume there is $A \in \mathcal{F}$ such that $\text{supp}(f) \subset A$ and $\mu(A) < \infty$. Then

$$\int f \, d\mu = \int_A f \, d\mu,$$ (5.8)

where the integral on the right is in the sense of Definition 4.8.

Proof. As $0 \leq \phi \leq f$ forces $\phi = 0$ on $A^c$, we may restrict all definitions of integrals to the measure space $(A, \mathcal{F}_A, \mu_A)$ (see Lemma 4.9) and get exactly the same values. But $\mu_A(A) < \infty$ and $f$ is bounded, and so the lower unsigned integral coincides with the lower “bounded” integral. Now $f_A^{-1}(B \setminus \{0\}) = f^{-1}(B \setminus \{0\})$ for any set $B \subset X$ shows that the function $f_A$ (see again Lemma 4.9) is $\mathcal{F}_A$-measurable. Since $f_A$ is also bounded, it is integrable with respect to $\mu_A$ and so the lower “bounded” integral on measure space $(A, \mathcal{F}_A, \mu_A)$ is equal to the integral $\int f_A \, d\mu_A$. By Lemma 4.9 again, this is the integral on the right of (5.8). □

The passage from the “bounded” integral to the unsigned integral will be provided by methods not dissimilar to those mentioned very early in this chapter:

**Lemma 5.5** Let $f: X \to [0, \infty]$ be measurable. Then

1. We have:

$$\int f \wedge M \, d\mu \xrightarrow{M \to \infty} \int f \, d\mu,$$ (5.9)

2. if $\{A_n\} \subset \mathcal{F}$ obey $A_n \uparrow A \supseteq \{f > 0\}$, then

$$\int (f 1_{A_n}) \, d\mu \xrightarrow{n \to \infty} \int f \, d\mu.$$ (5.10)

Proof. (1) We have $f \wedge M \leq f$ and so the limit (which exists by monotonicity) is at most the integral on the right. For the opposite inequality, note that if $\phi$ is simple with $0 \leq \phi \leq f$ then
\( \phi \leq (f \land M) \) as soon as \( M \geq \max \phi \). (Simple functions are necessarily bounded.) Hence

\[
\int \phi \, d\mu \leq \lim_{M \to \infty} \int (f \land M) \, d\mu, \quad \forall \phi \text{ simple, } 0 \leq \phi \leq f. \tag{5.11}
\]

Taking supremum over such \( \phi \) we get \( \int f \, d\mu \) on the left. Part (1) is proved.

For (2) the limit again exists by monotonicity and it is at most the integral on the right. For the opposite inequality let \( \phi = \sum_{i=1}^k a_i 1_{B_i} \) be simple with \( 0 \leq \phi \leq f \). Then \( \phi 1_{A_n} \) is also simple with \( \phi 1_{A_n} \leq \phi 1_A \). But \( \lambda \supseteq \{ f > 0 \} \supseteq \{ \phi > 0 \} \) and so, in fact, \( \phi 1_{A_n} \to \phi \). Hence, \( B_i \cap A_i \cap B_i \) whenever \( a_i > 0 \) and so, by the Monotone Convergence Theorem for sets,

\[
\int (\phi 1_{A_n}) \, d\mu \geq \int (\phi 1_{A_n}) \, d\mu = \sum_{i=1}^k a_i \mu(A_n \cap B_i) \to \sum_{i=1}^k a_i \mu(B_i) = \int \phi \, d\mu. \tag{5.12}
\]

Taking supremum over \( \phi \), we get that the limit of lower integrals of \( f1_{A_n} \) dominates the lower integral of \( f \).

Before we use these in the proof of Proposition 5.3, we also generalize the Markov inequality to the unsigned integral:

**Lemma 5.6** (Markov inequality again) *For any measurable \( f : X \to [0, \infty] \) and any \( \lambda > 0 \),*

\[
\mu(\{ f \geq \lambda \}) \leq \frac{1}{\lambda} \int f \, d\mu. \tag{5.13}
\]

*Proof.* We have \( \lambda 1_{\{ f \geq \lambda \}} \leq f \) whenever \( \lambda \geq 0 \). The claim follows (as before) by integrating these with respect to \( \mu \) and using Lemma 5.2(1). \( \Box \)

*Proof of Proposition 5.3.* Without loss of generality, assume that both lower integrals of \( f \) and \( g \) are finite. (Otherwise the claim holds by Lemma 5.2(2).) Define

\[
B_n := \{ f > 1/n \} \quad \text{and} \quad C_n := \{ g > 1/n \}. \tag{5.14}
\]

By Markov’s inequality, \( \mu(B_n) < \infty \) and \( \mu(C_n) < \infty \) for all \( n \in \mathbb{N} \). Hence also \( A_n := B_n \cap C_n \) obeys \( \mu(A_n) < \infty \). Now

\[
(f + g) \land M \leq f \land M + g \land M \tag{5.15}
\]

for any \( M \geq 0 \) and so

\[
\int_{A_n} ((f + g) \land M) \, d\mu \leq \int_{A_n} (f \land M + g \land M) \, d\mu
\]

\[
= \int_{A_n} (f \land M) \, d\mu + \int_{A_n} (g \land M) \, d\mu
\]

\[
= \int (f \land M) 1_{A_n} \, d\mu + \int (g \land M) 1_{A_n} \, d\mu \leq \int f \, d\mu + \int g \, d\mu,
\]

where we used the additivity of the “bounded” integral to get the middle equality and then applied Lemma 5.4 and Lemma 5.2(1) to get the conclusion. However,

\[
A_n \uparrow \{ f + g > 0 \} \tag{5.17}
\]

and so, by Lemma 5.4 and 5.5,

\[
\int_{A_n} ((f + g) \land M) \, d\mu \to \int (f + g) \land M \, d\mu \to \int (f + g) \, d\mu. \tag{5.18}
\]
The desired conclusion follows.

In light of this, it now makes sense to put forward:

**Definition 5.7 (Unsigned integral)** For any \( f : X \to [0, \infty] \) measurable, we define the unsigned Lebesgue integral of \( f \) with respect to \( \mu \) by

\[
\int f \, d\mu := \int f \, d\mu.
\] (5.19)

For non-measurable \( f \) the unsigned integral is left undefined, and so the concept of integrability is now completely tied to measurability. Naturally, once \( f \) is measurable, all properties of unsigned lower integral proved above are shared by its unsigned integral.

**5.2 Signed integral and convergence theorems.**

This unsigned integral shares the properties (1) and (3) from Lemma 5.2 and is additive in the integrand thanks to Lemma 5.2(2) and Proposition 5.3. The unsigned integral is actually well behaved even under monotone pointwise limits. Indeed, we have:

**Theorem 5.8 (Monotone Convergence Theorem)** Let \( f_n : X \to [0, \infty] \) be measurable with \( f_n \uparrow f \). Then also \( f \) is measurable and

\[
\int f_n \, d\mu \uparrow \int f \, d\mu.
\] (5.20)

**Proof.** The limit of the integrals of \( f_n \) exists and, since \( f_n \leq f \), is at most the integral of \( f \). For the opposite direction we may assume that the limit is actually finite. Let \( A_n := \{ f_n > 1/n \} \). By Markov’s inequality we have \( \mu(A_n) < \infty \) and so \( (f_n \wedge M)1_{A_k} \) are bounded functions with support of finite measure. The Bounded Convergence Theorem and Lemma 5.4 then show

\[
\int f_n \, d\mu \geq \int (f_n \wedge M)1_{A_k} \, d\mu \rightarrow_{n \to \infty} \int (f \wedge M)1_{A_k} \, d\mu
\] (5.21)

and so the limit of the integrals is at most the integral on the right. But \( A_K \uparrow \{ f > 0 \} \) and so taking \( K \to \infty \) followed by \( M \to \infty \) makes the right hand side tend to the integral of \( f \). \( \square \)

Two simple but useful consequences are in order:

**Corollary 5.9** Let \( \{ f_n \} \) be measurable functions \( f_n : X \to [0, \infty] \). Then

\[
\int \left( \sum_{n \in \mathbb{N}} f_n \right) \, d\mu = \sum_{n \in \mathbb{N}} \int f_n \, d\mu.
\] (5.22)

**Proof.** Define \( g_n := \sum_{k=1}^{n} f_k \) and \( g := \sum_{k \in \mathbb{N}} f_k \). By finite additivity of the integral,

\[
\int g_n \, d\mu = \sum_{k=1}^{n} \int f_k \, d\mu
\] (5.23)

The left hand side tends to the integral of \( g \) by the Monotone Convergence Theorem, while the right-hand side converges to the infinite sum by (3.1). \( \square \)
Corollary 5.10  Let \( f : X \mapsto [0, \infty] \) be measurable. For each \( A \in \mathcal{F} \) define
\[
\nu(A) := \int_A f \, d\mu := \int (f1_A) \, d\mu.
\] (5.24)
Then \( \nu \) is a measure on \( (X, \mathcal{F}) \).

Proof. Obviously, \( \nu(\emptyset) = 0 \). If \( \{A_n\} \subset \mathcal{F} \) are disjoint, then
\[
1_{\bigcup_{n \in N} A_n} = \sum_{n \in N} 1_{A_n}
\] (5.25)
and so by (3.3) and the previous corollary
\[
\nu\left( \bigcup_{n \in N} A_n \right) = \int f \left( \sum_{n \in N} 1_{A_n} \right) \, d\mu = \sum_{n \in N} \int 1_{A_n} \, d\mu = \sum_{n \in N} \nu(A_n).
\] (5.26)
Hence, \( \nu \) is countably additive and so it is a measure. □

An interesting question is when is a measure \( \nu \) on \( (X, \mathcal{F}) \) of the form (5.24). We will answer this later in these notes. Another useful consequence of the Monotone Convergence Theorem is:

Lemma 5.11 (Fatou’s lemma)  Let \( f_n : X \rightarrow [0, \infty] \) be measurable. Then
\[
\int \liminf_{n \rightarrow \infty} f_n \, d\mu \leq \liminf_{n \rightarrow \infty} \int f_n \, d\mu.
\] (5.27)

Proof. For each \( m \geq n \) we have \( f_m \geq \inf_{k \geq n} f_k \) and so
\[
\inf_{m \geq n} \int f_m \, d\mu \geq \int \inf_{m \geq n} f_m \, d\mu.
\] (5.28)
Since \( \inf_{m \geq n} f_m \) increases to the limes inferior as \( n \rightarrow \infty \), the claim follows by the Monotone Convergence Theorem. □

Having discussed the unsigned integral enough, we can move to the signed integral. Given a function \( f \), let
\[
f^+ := f1_{\{f \geq 0\}} \quad \text{and} \quad f^- = -f1_{\{f < 0\}}
\] (5.29)
be its positive and negative parts. Obviously, \( f^+, f^- \) are measurable whenever \( f \) is and
\[
f = f^+ - f^-.
\] (5.30)

We now put forward:

Definition 5.12 (Signed integral)  For any \( f : X \rightarrow \mathbb{R} \) measurable such that the unsigned integrals of \( f^+ \) and \( f^- \) obey
\[
\min\left\{ \int f^+ \, d\mu, \int f^- \, d\mu \right\} < \infty
\] (5.31)
we define the (signed) Lebesgue integral of \( f \) with respect to \( \mu \) by
\[
\int f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu.
\] (5.32)
We say that such an \( f \) is integrable (with respect to \( \mu \)).
Notice that the above does permit the signed integral to be equal to \( \pm \infty \). As we will see that the signed integral is considerably weaker than the unsigned one. Notwithstanding, the expected properties still hold, at least with correct provisos:

**Lemma 5.13** If \( f : X \to \mathbb{R} \) is integrable and \( c \in \mathbb{R} \), then

\[
\int (cf) \, d\mu = c \int f \, d\mu \tag{5.33}
\]

whenever the right-hand side is not of the form \( 0 \times \infty \). In particular, \( cf \) is integrable whenever \( f \) is. Similarly, if \( f, g : X \to \mathbb{R} \) are integrable, then

\[
f + g \text{ is integrable and } \int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu, \tag{5.34}
\]

provided the right-hand side is not a difference of two infinities of same sign.

**Proof.** For the first property we note that if \( c > 0 \), then \( (cf)^\pm = cf^\pm \) while for \( c < 0 \) we get \( (cf)^\pm = cf^\mp \). The claim then follows from Lemma 5.2(2) and the definition of unsigned integral. For \( c = 0 \) we assume that both \( \int f^\pm \, d\mu < \infty \) and so the claim again follows by Lemma 5.2(2). That integrability of \( f \) implies integrability of \( cf \) is a consequence of these.

For the second property, observe that if the expression on the right of (5.34) is defined and meaningful, then at least one of \( f^+ + g^+ \) and \( f^- + g^- \) has finite signed integral. Since

\[
(f + g)^+ = (f^+ + g^+) \mathbb{1}_{\{f + g \geq 0\}} - (f^- + g^-) \mathbb{1}_{\{f + g < 0\}},
\]

\[
(f + g)^- = -(f^+ + g^+) \mathbb{1}_{\{f + g < 0\}} + (f^- + g^-) \mathbb{1}_{\{f + g < 0\}},
\]

we have \((f + g)^+ \leq f^+ + g^+ \) and \((f + g)^- \leq f^- + g^- \) and so \( f + g \) is integrable. Moreover, by the additivity of the unsigned integral we also have

\[
\int (f + g)^+ \, d\mu = \int (f^+ + g^+) \mathbb{1}_{\{f + g \geq 0\}} \, d\mu - \int (f^- + g^-) \mathbb{1}_{\{f + g < 0\}} \, d\mu,
\]

\[
\int (f + g)^- \, d\mu = \int (f^- + g^-) \mathbb{1}_{\{f + g < 0\}} \, d\mu - \int (f^+ + g^+) \mathbb{1}_{\{f + g < 0\}} \, d\mu, \tag{5.36}
\]

with both expressions on the right meaningful. Thus

\[
\int (f + g) \, d\mu := \int (f + g)^+ \, d\mu - \int (f + g)^- \, d\mu
\]

\[
= \int (f^+ + g^+) \mathbb{1}_{\{f + g \geq 0\}} \, d\mu - \int (f^- + g^-) \mathbb{1}_{\{f + g \geq 0\}} \, d\mu
\]

\[
- \int (f^- + g^-) \mathbb{1}_{\{f + g < 0\}} \, d\mu + \int (f^+ + g^+) \mathbb{1}_{\{f + g < 0\}} \, d\mu \tag{5.37}
\]

\[
= \int (f^+ + g^+) \, d\mu - \int (f^- + g^-) \, d\mu.
\]

Breaking this expression into a sum of four integrals, and noting that only one type of infinity can occur among these, we can reassemble this into the right-hand side of (5.34). \[ \square \]

Thus, also the signed integral is additive provided the resulting expressions make sense. Concerning behavior under limits, here we get:
Theorem 5.14 (Dominated Convergence Theorem) Let \( \{f_n\} \) and \( g \) be measurable functions such that \( \int g \, d\mu < \infty \) and \( |f_n| \leq g \) for all \( n \in \mathbb{N} \). Then

\[ f_n \rightarrow f \text{ a.e.} \Rightarrow \int f_n \, d\mu \rightarrow \int f \, d\mu \]  

(5.38)

and, in particular, \( f \) is integrable.

Proof. We have \( g + f_n \geq 0 \) and so, by Fatou’s lemma,

\[ \int \left( g + \liminf_{n \to \infty} f_n \right) \, d\mu \leq \liminf_{n \to \infty} \int (g + f_n) \, d\mu \]  

(5.39)

Since \( g \geq 0 \) and \( \int g \, d\mu < \infty \), we can separate the integral of \( g \) on both sides and get

\[ \int \liminf_{n \to \infty} f_n \, d\mu \leq \liminf_{n \to \infty} \int f_n \, d\mu. \]  

(5.40)

Similarly we get \( g - f_n \geq 0 \) and, since \( \liminf_{n \to \infty} (g - f_n) = g - \limsup_{n \to \infty} f_n \), Fatou’s lemma tells us

\[ \int \left( g - \limsup_{n \to \infty} f_n \right) \, d\mu \leq \liminf_{n \to \infty} \int (g - f_n) \, d\mu \]  

(5.41)

Separating the integral of \( g \) again, this can be rewritten as

\[ -\int \limsup_{n \to \infty} f_n \, d\mu \leq -\limsup_{n \to \infty} \int f_n \, d\mu. \]  

(5.42)

Combining (5.40) and (5.42), we obtain

\[ \int \liminf_{n \to \infty} f_n \, d\mu \leq \liminf_{n \to \infty} \int f_n \, d\mu \leq \limsup_{n \to \infty} \int f_n \, d\mu \leq \int \limsup_{n \to \infty} f_n \, d\mu. \]  

(5.43)

When \( f_n \rightarrow f \text{ a.e.} \), both \( \text{limes superior and limes inferior of } f_n \) are equal to \( f \text{ a.e.} \). The extreme ends of the inequality thus coincide, the limit of integrals of \( f_n \) exists and equals the integral of the a.e.-limit. \( \Box \)

It is instructive to note that without the existence of the \textit{dominating function} \( g \), the theorem falls. This can be for (one or both of) the following reasons:

1. the functions \( f_n \) increase to infinity locally, e.g., \( f_n := n1_{[0,1/n]} \) which tends to zero Lebesgue a.e. and yet its integral equals one, or
2. (in infinite measure spaces) the functions \( f_n \) spread over sets of larger and larger measure, e.g., \( f_n := n^{-1}1_{[0,n]} \) which tends to zero pointwise (in fact uniformly) and yet its integral with respect to the Lebesgue measure equals one.

As is easy to check, the existence of a dominating function makes these impossible.

5.3 Absolute integrability and \( L^1 \)-space.

The situation with the signed integral improves if we assume that both the positive and negative part of the function integrates to a finite number.

Definition 5.15 A measurable function \( f : X \rightarrow \mathbb{R} \) is said to be \textit{absolutely integrable} if the signed integral of \( |f| \) obeys

\[ \int |f| \, d\mu < \infty. \]  

(5.44)
We will denote the set of absolutely integrable functions by \( L^1 \) (or \( L^1(X, \mathcal{F}, \mu) \) or \( L^1(\mu) \), depending on context).

Since \( |f| = f^+ + f^- \), absolute integrability implies integrability and we have the bound
\[
\left| \int f \, d\mu \right| \leq \int |f| \, d\mu.
\] (5.45)

As a consequence of the properties of the signed integral, \( L^1 \) is closed under (pointwise) addition and scalar multiplication and is thus a vector space. Note, however, that \( L^1 \) is not closed under pointwise limits (even in a.e. sense) as \( f_n \in L^1 \) and \( f_n \to f \) does not imply \( f \in L^1 \). (Take, e.g., \( f_n := 1_{[0,n]} \).) The notion of convergence that ensures this is given by:

**Definition 5.16** Let \( \{f_n\}, f \) be measurable functions. We say that
\[
f_n \to f \text{ in } L^1(\mu) \quad \text{if} \quad \int |f_n - f| \, d\mu \to 0
\] (5.46)

To assess the strength of this convergence, we note:

**Lemma 5.17** Let \( \{f_n\} \subset L^1 \). If \( f_n \to f \) in \( L^1 \), then \( f \in L^1 \), \( f_n \to f \) in measure and
\[
\lim_{n \to \infty} \int f_n \, d\mu = \int f \, d\mu \quad \text{and} \quad \lim_{n \to \infty} \int |f_n| \, d\mu = \int |f| \, d\mu
\] (5.47)

**Proof.** Assume without loss of generality that \( \int |f_n - f| \, d\mu < \infty \) for all \( n \). (Otherwise drop those terms from the sequence.) Then we claim that
\[
\sup_{n \in \mathbb{N}} \int |f_n| \, d\mu < \infty.
\] (5.48)

Indeed, by the triangle inequality for the absolute value and finite additivity of the unsigned integral,
\[
\int |f_n| \, d\mu \leq \int |f_1| \, d\mu + \int |f_1 - f| \, d\mu + \int |f_n - f| \, d\mu.
\] (5.49)

The first two integrals are finite while the last integral converges to zero as \( n \to \infty \) and so the supremum over \( n \) of the right-hand side is finite indeed.

Next we pick \( \varepsilon > 0 \) and use Markov’s inequality to show
\[
\mu(\{|f_n - f| > \varepsilon\}) \leq \frac{1}{\varepsilon} \int |f_n - f| \, d\mu \to 0.
\] (5.50)

So \( f_n \to f \) in measure. By Lemma 4.21 there is \( \{n_k\} \) such that \( f_{n_k} \to f \) a.e. and Fatou’s lemma then ensures
\[
\int |f| \, d\mu \leq \lim_{k \to \infty} \int |f_{n_k}| \, d\mu \leq \sup_{n \in \mathbb{N}} \int |f_n| \, d\mu < \infty.
\] (5.51)

So \( f \in L^1 \) as claimed. But then we use (5.45) to write
\[
\left| \int f_n \, d\mu - \int f \, d\mu \right| = \left| \int (f_n - f) \, d\mu \right| \leq \int |f_n - f| \, d\mu \to 0
\] (5.52)

and so the integral of \( f_n \) tends to that of \( f \). Replacing \( f_n \) by \( |f_n| \) and \( f \) by \( |f| \) and using that \( |a| + |b| \leq |a + b| \), the same holds for the integrals of absolute values.
As we just saw, from an $L^1$-converging sequence (or even a sequence converging in measure) one can always choose an a.e.-converging subsequence. However, $L^1$-convergence does not imply a.e.-convergence or even just convergence in measure. Indeed, consider a sequence $\{f_n\}$ defined by

$$f_{2^n+k} := 1_{[k2^{-n},(k+1)2^{-n}]}, \quad k = 0, \ldots, 2^n - 1.$$  

(5.53)

Then $f_n \to 0$ in $L^1$ and measure and yet $\limsup_{n \to \infty} f_n = 1_{[0,1]}$ while $\liminf_{n \to \infty} f_n = 0$. The example $f_n := n1_{[0,1/n]}$ shows that convergence in measure does not imply $L^1$ convergence. Nonetheless, for bounded functions on finite measure spaces, convergence in measure is equivalent to $L^1$ convergence. Indeed:

**Lemma 5.18** Suppose $\{f_n\}$ obey $|f_n| \leq K$ a.e. for some constant $K \in (0, \infty)$ and all $n \in \mathbb{N}$. If $\mu(X) < \infty$, then

$$f_n \to f \text{ in measure} \iff f_n \to f \text{ in } L^1.$$  

(5.54)

**Proof.** We already proved $\iff$, so let us focus on $\Rightarrow$. Suppose $f_n \to f$ in measure. Then

$$\mu(|f| \geq K + \varepsilon) \leq \mu(|f_n - f| \geq \varepsilon) + \mu(|f_n| \geq K)$$

(5.55)

so $|f_n| \leq K$ a.e. implies also $|f| \leq K$ a.e. Therefore,

$$\int |f_n - f| \, d\mu = \int |f_n - f| 1_{\{|f_n - f| > \varepsilon\}} \, d\mu + \int |f_n - f| 1_{\{|f_n - f| \leq \varepsilon\}} \, d\mu \leq 2K\mu(|f_n - f| > \varepsilon) + \varepsilon\mu(X).$$

(5.56)

But this tends to zero as $n \to \infty$ followed by $\varepsilon \downarrow 0$ and so $f_n \to f$ in $L^1$ as well. \hfill $\square$

### 5.4 Uniform integrability.

A natural question to ask is what conditions on a pointwise a.e. convergent sequence of measurable functions $\{f_n\}$ imply convergence in $L^1$. For this we first collect the necessary conditions:

**Proposition 5.19** (Necessary conditions for $L^1$-convergence) Suppose that $\{f_n\}$ and $f$ are measurable functions with $f_n \to f$ in $L^1$. Then

1. $f \in L^1$ and

$$\int |f_n| \, d\mu \to \int |f| \, d\mu,$$

(5.57)

2. $\lim_{M \to \infty} \limsup_{n \to \infty} \int |f_n| 1_{\{|f_n| \geq M\}} \, d\mu = 0,$$

(5.58)

3. $\lim_{\delta \downarrow 0} \limsup_{n \geq 1} \int |f_n| 1_{\{|f_n| \leq \delta\}} \, d\mu = 0.$$

(5.59)

Note that the condition (3) is trivial in finite measure spaces (and, in particular, in probability spaces). The proof will need the following lemma:

**Lemma 5.20** Let $f \in L^1$. Then

$$\forall \varepsilon > 0 \exists \delta > 0 \forall A \in \mathcal{F}: \quad \mu(A) < \delta \quad \Rightarrow \quad \int_A |f| \, d\mu < \varepsilon.$$  

(5.60)
Proof. Suppose the claim fails. Then there is $\varepsilon > 0$ such that for each $n \geq 1$ there is $A_n \in \mathcal{F}$ with $\mu(A_n) < 2^{-n}$ such that $\int_{A_n} |f| \, d\mu \geq \varepsilon$. Define $B_n := \bigcup_{k \geq n} A_k$. Then $\mu(B_n) \leq 2^{-n+1}$ and $B_n \downarrow B := \limsup_{n \to \infty} A_n$. By the Downward Monotone Convergence Theorem for sets, $\mu(B) \leq \limsup_{n \to \infty} \mu(A_n) = 0$. However, the Dominated Convergence Theorem (with $|f|$ being the dominating function for functions $|f|1_{B_n}$) ensures

$$\varepsilon \leq \int_{A_n} |f| \, d\mu \leq \int_{B_n} |f| \, d\mu \xrightarrow{n \to \infty} \int_{B} |f| \, d\mu = 0,$$

(5.61)
a contradiction. So the claim holds after all. \qed

As we will see later, the statement of this lemma is related to the properties of absolutely continuous measures. With this lemma in hand, we can now proceed to give:

Proof of Proposition 5.19. (1) is a restatement of part of Lemma 5.17.

For (2) we note that $|f_n| \leq |f_n - f| + |f|$ and so

$$|f_n|1_{|f_n| \geq 2M} \leq |f_n - f| + |f|1_{|f| \geq M} + |f|1_{|f - f_n| \geq M}.$$  

(5.62)
The integral of the first term tends to zero as $n \to \infty$ by our assumption of $L^1$ convergence. The second term is dominated by the integrable function $|f|$ and tends to zero a.e. (integrability of $|f|$ implies that $|f| < \infty$ a.e.). The Dominated Convergence Theorem shows

$$\int |f|1_{|f| \geq M} \, d\mu \xrightarrow{M \to \infty} 0$$  

(5.63)
For the last term in (5.62) we recall that $f_n \to f$ in $L^1$ implies that $f_n \to f$ in measure. This yields

$$\limsup_{n \to \infty} \mu(|f - f_n| \geq M) = 0$$  

(5.64)
and so

$$\limsup_{n \to \infty} \int |f|1_{|f - f_n| \geq M} = 0$$  

(5.65)
by Lemma 5.20.

Part (3) is proved analogously. From $|f_n| \leq |f_n - f| + |f|$ and $|f| \leq |f_n - f| + |f_n|$ we have

$$|f_n|1_{|f_n| \leq \delta} \leq |f_n - f| + |f|1_{|f| \leq \delta} + |f|1_{|f-f_n| \geq \delta}.$$  

(5.66)
The integral of the first term again vanishes by assumed $L^1$-convergence. The last integral tends to zero in the limit $n \to \infty$ by Lemma 5.20. The middle function tends to zero in the limit $\delta \downarrow 0$, and so does its integral thanks to the Dominated Convergence Theorem. \qed

As is not hard to check, the limes superior in parts (2) and (3) of Proposition 5.19 can be replaced by supremum. We now come up with a name for families satisfying these sorts of conditions:

Definition 5.21 A family $\{f_\alpha: \alpha \in I\}$ (any cardinality) of measurable functions is said to be uniformly integrable (UI) if the following hold:

1. $\{f_\alpha\} \subset L^1(\mu)$ and

$$\sup_{\alpha \in I} \int |f_\alpha| \, d\mu < \infty,$$

(5.67)
\[
\lim_{M \to \infty} \sup_{\alpha \in I} \int |f_\alpha| 1_{\{|f_\alpha| \geq M\}} \, d\mu = 0, \quad (5.68)
\]
\[
\lim_{\delta \downarrow 0} \sup_{\alpha \in I} \int |f_\alpha| 1_{\{|f_\alpha| \leq \delta\}} \, d\mu = 0. \quad (5.69)
\]

To see some examples, we note:

Lemma 5.22  The following families are UI:

1. \(\{f\}\) for \(f \in L^1\),
2. \(\{f_1, \ldots, f_n\} \subset L^1\),
3. \(\{f_{\alpha,k} : k = 1, \ldots, n\}\) if \(\{f_{\alpha,k}\}\), \(k = 1, \ldots, n\) are UI,
4. \(\{f \in L^1 : |f| \leq g\}\) for any \(g \in L^1\).

Proof. Exercise for the reader. □

We also note that Lemma 5.20 holds uniformly over UI families (we state this in a slightly weaker form):

Lemma 5.23  Let \(\{f_\alpha : \alpha \in I\} \subset L^1\) obey

\[
\lim_{M \to \infty} \sup_{\alpha \in I} \int |f_\alpha| 1_{\{|f_\alpha| \geq M\}} \, d\mu = 0. \quad (5.70)
\]

Then

\[
\forall \varepsilon > 0 \exists \delta > 0 \forall A \in \mathcal{F} : \mu(A) < \delta \implies \sup_{\alpha \in I} \int_A |f_\alpha| \, d\mu < \varepsilon. \quad (5.71)
\]

Proof. We will give a slightly different proof than given for Lemma 5.20. Fix \(\varepsilon > 0\) and suppose that for some \(A \in \mathcal{F}\) we have \(\sup_{\alpha \in I} \int_A |f_\alpha| > 2\varepsilon\). By (5.70), there is \(M \in (0, \infty)\) such that

\[
\sup_{\alpha \in I} \int |f_\alpha| 1_{\{|f_\alpha| \geq M\}} \, d\mu < \varepsilon \quad (5.72)
\]

and then

\[
\sup_{\alpha \in I} \int_A |f_\alpha| \, d\mu \leq \varepsilon + \sup_{\alpha \in I} \int |f_\alpha| 1_{\{|f_\alpha| \leq M\}} \, d\mu \leq \varepsilon + M \mu(A). \quad (5.73)
\]

Thus, if \(\mu(A) < \varepsilon / M\), the right hand side is at most \(2\varepsilon\). □

The main result of this section is a rather far-reaching generalization of Lemma 5.18.

Theorem 5.24  Let \(\{f_n\}\) and \(f\) be measurable functions. If \(\{f_n\}\) is UI, then

\[f_n \to f\; \text{in measure} \iff f_n \to f \; \text{in } L^1.\]  (5.74)

Proof. Thanks to Lemma 5.17, we only need to prove \(\Rightarrow\). Fix \(\delta > 0\), consider the integral of \(|f_n - f|\) and split it according to whether \(|f_n - f| > \delta\) or not. The first part yields

\[
\int |f_n - f| 1_{\{|f_n - f| > \delta\}} \, d\mu \leq \int (|f_n| + |f|) 1_{\{|f_n - f| > \delta\}} \, d\mu \quad (5.75)
\]

Splitting the integral and applying Lemma 5.23 for the UI family \(\{f\} \cup \{f_n : n \in \mathbb{N}\}\) along with \(f_n \to f\) in measure, the right-hand side tends to zero as \(n \to \infty\).
For the second part of the integral, we write
\[ \int |f_n - f| 1_{\{|f_n - f| \leq \delta\}} \, d\mu = \int |f_n - f| 1_{\{|f_n - f| \leq \delta\}} 1_{\{|f| > \delta\}} \, d\mu + \int |f_n - f| 1_{\{|f_n - f| \leq \delta\}} 1_{\{|f| \leq \delta\}} \, d\mu \] (5.76)

Since \( \mu(|f| > \delta) < \infty \) by Markov’s inequality, the first integral tends to zero as \( n \to \infty \) by the Bounded Convergence Theorem. The second integral we further bound as
\[ \int |f_n - f| 1_{\{|f_n - f| \leq \delta\}} 1_{\{|f| \leq \delta\}} \, d\mu \leq \int |f_n| 1_{\{|f_n| \leq 2\delta\}} \, d\mu + \int |f| 1_{\{|f| \leq \delta\}} \, d\mu. \] (5.77)

The supremum over \( n \) of the first term tends to zero as \( \delta \downarrow 0 \) thanks to UI of \( \{f_n\} \). By the Dominated Convergence Theorem, the second integral tends to zero as \( \delta \downarrow 0 \) as well. \( \square \)

Uniform integrability plays very important role in probability, particularly, in martingale convergence theory etc.
6. Abstract Measure Theory

Having developed the Lebesgue integral with respect to the general measures, we now have a general concept with few specific examples to actually test it on. (Indeed, so far we have constructed only one measure; namely, the Lebesgue measure on $\mathbb{R}^d$.) In this chapter we will mend this by providing tools how to construct measures on abstract measurable spaces. As for the Lebesgue measure, the key tool will be outer measures.

6.1 From semialgebras to outer measures.

Consider an arbitrary set $X$. Our goal is to develop structures needed for putting a measure on $X$. For this, let us recall how the construction of the Lebesgue measure went: We began by defining the measure of half-open boxes. Then we extended this to elementary sets, which were finite unions of half-open boxes. Next, we constructed the Lebesgue outer measure by coverings by countable collections of half-open boxes. Finally, we used topology to define measurable sets and checked that the outer measure restricted to these is actually a measure.

We will proceed along this plan even for a general set $X$ except, in the absence of topology, we will define measurability differently. As is easy to check, the class of half-open boxes in $\mathbb{R}^d$ forms the following structure:

**Definition 6.1** A class $\mathcal{S} \subset 2^X$ is a semialgebra on $X$ if it satisfies the following requirements:

(1) $\emptyset, X \in \mathcal{S}$,
(2) $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$,
(3) if $A \in \mathcal{S}$, then there are disjoint $B_1, \ldots, B_n \in \mathcal{S}$ such that

$$A^c = \bigcup_{i=1}^n B_i. \quad (6.1)$$

Recall that an algebra is a subset of $2^X$ that contains $\emptyset$ and $X$ and is closed under finite unions and complements. Moving up to our generalization of elementary sets, we observe:

**Lemma 6.2** If $\mathcal{S}$ is a semialgebra, then

$$\mathcal{S} := \left\{ \bigcup_{i=1}^n A_i : \{A_i\} \subset \mathcal{S} \right\} \quad (6.2)$$

is an algebra. The same object is obtained when the union in required to be disjoint.

**Proof.** Let $A_1, \ldots, A_n \subset \mathcal{S}$. By the property (3) of the semialgebra, for each $i = 1, \ldots, n$ there are \{ $B_{i,j}$: $j = 1, \ldots, m$ $\} \subset \mathcal{S}$ such that

$$A_i^c = \bigcup_{j=1}^m B_{i,j}. \quad (6.3)$$

(We can always add empty sets to make the union the same size for all $i$.) From de Morgan’s laws and the distributive property of union around intersection we then get

$$\left( \bigcup_{i=1}^n A_i \right)^c = \bigcap_{i=1}^n A_i^c = \bigcap_{i=1}^n \bigcup_{j=1}^m B_{i,j} = \bigcup_{j_1=1}^m \bigcup_{j_2=1}^m \cdots \bigcup_{j_n=1}^m \left( \bigcap_{i=1}^n B_{i,j_i} \right). \quad (6.4)$$
But $\bigcap_{i=1}^{n} B_{i,j} \in \mathcal{F}$ by property (2) of the semialgebra and so $\bigcup_{i=1}^{n} A_{i} \in \mathcal{F}$. It follows that $\mathcal{A}$ is closed under complements. As $\mathcal{A}$ is obviously closed under finite unions and contains $\emptyset$ and $X$ (because it includes $\mathcal{F}$ which contains these) it is an algebra.

To see that we can define the algebra using disjoint unions, let $A = \bigcup_{i=1}^{n} A_{i}$ with $\{A_{i}\} \subset \mathcal{F}$ disjoint and let $B \in \mathcal{F}$. By property (3) of the semialgebra, there are disjoint sets $\{B_{j}\} \subset \mathcal{F}$ such that $B_{c} = \bigcup_{j=1}^{m} B_{j}$. Then

$$B \cup A = B \cup (A \cap B_{c}) = B \cup \bigcup_{i=1}^{n} \bigcup_{j=1}^{m} (A_{i} \cap B_{j}). \quad (6.5)$$

Now $\{A_{i} \cap B_{j}\}$ are disjoint and they are (being subsets of $B_{c}$) disjoint from $B$. Hence, if any union of $n$ sets from $\mathcal{F}$ can be disjointified, also any union of $n + 1$ sets from $\mathcal{F}$ can be disjointified. By induction, $\mathcal{A}$ can be defined by disjoint unions as well. \hfill \Box

Now consider a set function $\mu : \mathcal{F} \to \mathbb{R} \cup \{\pm \infty\}$, which we permit, for the time being, to take both positive and negative (and even infinite) values. Our goal is to extend it to a finitely additive set function on $\mathcal{A}$ in accord with the following definition:

**Definition 6.3** A function $\mu : \mathcal{A} \to \mathbb{R} \cup \{\pm \infty\}$ on an algebra $\mathcal{A}$ is finitely additive if for any disjoint $\{A_{i}\} \subset \mathcal{A}$,

$$\mu \left( \bigcup_{i=1}^{n} A_{i} \right) = \sum_{i=1}^{n} \mu (A_{i}) \quad (6.6)$$

and the sum on the right-hand side makes sense.

The only way the sum would not both $+\infty$ and $-\infty$ appears among some of the terms. If this happens for disjoint sets then, indeed, we have a problem indeed. That disjointness does not provide enough protection for both infinities to coexists is the content of:

**Lemma 6.4** If $\mu$ is a finitely additive set function on an algebra, then $\mu$ cannot take both $+\infty$ and $-\infty$ there.

**Proof.** Let $A, B$ be sets in the algebra such that $\mu (A) = +\infty$ and $\mu (B) = -\infty$. Then

$$\mu (A) = \mu (A \cap B) + \mu (A \setminus B) \quad \text{and} \quad \mu (B) = \mu (A \cap B) + \mu (B \setminus A) \quad (6.7)$$

with both sums on the right making sense. This implies that $\mu (A \cap B)$ must be finite while $\mu (A \setminus B) = +\infty$ and $\mu (B \setminus A) = -\infty$. But then $\mu$ is certainly not additive on the disjoint union of $A \setminus B$ and $B \setminus A$ as that would lead to a meaningless expression. \hfill \Box

Thus we may assume from the start that $\mu$ takes at most one of the infinities. There is, however, still another problem for the extension from $\mathcal{F}$ to $\mathcal{A}$: Finite additivity may be violated already on $\mathcal{F}$. That this is the only obstruction is the content of:

**Lemma 6.5** (Extension to algebra) Let $\mathcal{F} \subset 2^{X}$ be a semialgebra and let $\mu : \mathcal{F} \to \mathbb{R} \cup \{\infty\}$ be a set function that is finitely additive on $\mathcal{F}$ in the sense that

$$\forall \{A_{i}\} \subset \mathcal{F} \text{ disjoint}: \quad \bigcup_{i=1}^{n} A_{i} \in \mathcal{F} \implies \mu \left( \bigcup_{i=1}^{n} A_{i} \right) = \sum_{i=1}^{n} \mu (A_{i}) \quad (6.8)$$

\hfill 245A lecture notes (©2017 M. Biskup)

Preliminary version (typeset: September 22, 2017)
Then for any two disjoint families \( \{A_i\} \subset \mathcal{S} \) and \( \{B_j\} \subset \mathcal{S} \),
\[
\bigcup_{i=1}^{m} A_i = \bigcup_{j=1}^{n} B_j \quad \Rightarrow \quad \sum_{i=1}^{m} \mu(A_i) = \sum_{j=1}^{n} \mu(B_j).
\]  
(6.9)

Under such conditions \( \mu \) extends uniquely to a finitely additive set function on \( \mathcal{S} \).

**Proof.** Let \( \{A_i\} \subset \mathcal{S} \) and \( \{B_j\} \subset \mathcal{S} \) be two disjoint families with \( \bigcup_{i=1}^{m} A_i = \bigcup_{j=1}^{n} B_j \). Note that \( \{A_i \cap B_j\} \) are disjoint. Then \( A_i = \bigcup_{j=1}^{n} (A_i \cap B_j) \) and \( B_j = \bigcup_{i=1}^{m} (A_i \cap B_j) \) and so, by the assumption,
\[
\mu(A_i) = \sum_{i=1}^{m} \mu(A_i) \quad \text{and} \quad \mu(B_j) = \sum_{j=1}^{n} \mu(B_j).
\]  
(6.10)

Then
\[
\sum_{i=1}^{m} \mu(A_i) = \sum_{i=1}^{m} \left( \sum_{j=1}^{n} \mu(A_i \cap B_j) \right) = \sum_{j=1}^{n} \left( \sum_{i=1}^{m} \mu(A_i \cap B_j) \right) = \sum_{j=1}^{n} \mu(B_j)
\]  
(6.11)

thus proving the first part of the claim.

By Lemma 5.2, every set \( A \in \mathcal{S} \) can be written as a disjoint union \( \bigcup_{i=1}^{m} A_i \) for \( \{A_i\} \subset \mathcal{S} \). We then define
\[
\mu(A) := \sum_{i=1}^{m} \mu(A_i)
\]  
(6.12)

and note that, by the first part of the claim, this does not depend on the representation of \( A \). If now \( B_1, \ldots, B_m \in \mathcal{S} \) are disjoint sets and \( B_i = \bigcup_{j=1}^{n} B_{ij} \) is their representation as a union of a disjoint family \( \{B_{ij} : j = 1, \ldots, n\} \subset \mathcal{S} \) (again, we use empty sets to make all these unions have the same number of terms), we have
\[
\mu\left( \bigcup_{i=1}^{m} B_i \right) = \mu\left( \bigcup_{i=1}^{m} \bigcup_{j=1}^{n} B_{ij} \right) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mu(B_{ij}) = \sum_{i=1}^{m} \sum_{j=1}^{n} \mu(B_i) = \sum_{i=1}^{m} \mu(B_i).
\]  
(6.13)

Here in the first equality we used that (by disjointness of \( \{B_i\} \)) the whole collection \( \{B_{ij} : i = 1, \ldots, m, j = 1, \ldots, n\} \) consists of pairwise disjoint sets, then (in the second equality) we applied the definition (6.12) and then (in the third equality) we used (6.12) again to contract the second sum into a union. As (6.13) is a statement of finite additivity of \( \mu \) on \( \mathcal{S} \), we are done.

Keeping in mind our plan from the construction of the Lebesgue measure, we now go back to non-negative set functions only and axiomatize the properties from Proposition 3.2:

**Definition 6.6 (Outer measure)** Given a set \( X \), an outer measure on \( X \) is a map \( \mu^* : 2^X \to [0, \infty] \) satisfying the following properties:

1. \( \mu^*(\emptyset) = 0 \),
2. (monotonicity) if \( A \subset B \), then \( \mu^*(A) \leq \mu^*(B) \),
3. (countable subadditivity) for any sets \( \{A_i : i \in \mathbb{N}\} \),
\[
\mu^*\left( \bigcup_{i \in \mathbb{N}} A_i \right) \leq \sum_{i \in \mathbb{N}} \mu^*(A_i).
\]  
(6.14)

An imitation of the construction of the Lebesgue outer measure yields:
Proposition 6.7  Let \( \mathcal{C} \subset 2^X \) obey \( \emptyset \in \mathcal{C} \) and let \( \mu_0: \mathcal{C} \to [0,\infty] \) satisfy \( \mu_0(\mathcal{C}) = 0 \). Then

\[
\mu^*(A) := \inf \left\{ \sum_{i \in \mathbb{N}} \mu_0(A_i) : \{A_i\} \subset \mathcal{C}, \bigcup_{i \in \mathbb{N}} A_i \supset A \right\}
\]

(6.15)
is an outer measure on \( X \).

Proof. The proof is almost identical to that for the Lebesgue measure. First off, \( \mu^* \geq 0 \) because \( \mu_0 \geq 0 \). Hence, \( \mu^*(\emptyset) \leq \mu_0(\emptyset) = 0 \). Next if \( \{B_i\} \subset \mathcal{C} \) covers \( B \) then it also covers any \( A \) with \( A \subset B \) and so \( \mu^*(A) \leq \mu^*(B) \). Finally, if \( \{A_i\} \) is a collection of sets with \( \mu^*(A_i) \leq \infty \) (otherwise there is nothing to prove), then for each \( \varepsilon > 0 \) we may find \( \{B_{ij}\} \) with

\[
A_i \subset \bigcup_{j \in \mathbb{N}} B_{ij} \quad \text{and} \quad \mu^*(A_i) + \varepsilon 2^{-i} \geq \sum_{j \in \mathbb{N}} \mu_0(B_{ij})
\]

(6.16)
valid for all \( i \). But then \( \bigcup_{i,j \in \mathbb{N}} B_{ij} \supset \bigcup_{i \in \mathbb{N}} A_i \) and thus

\[
\mu^*( \bigcup_{i \in \mathbb{N}} A_i ) \leq \sum_{i,j \in \mathbb{N}} \mu_0(B_{ij}) = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \mu(B_{ij}) \\
\leq \sum_{i \in \mathbb{N}} (\mu^*(A_i) + \varepsilon 2^{-i}) = \varepsilon + \sum_{i \in \mathbb{N}} \mu^*(A_i)
\]

(6.17)

Since this holds for all \( \varepsilon > 0 \), \( \mu^* \) is countably subadditive and thus an outer measure. \( \square \)

The class \( \mathcal{C} \) of sets in the above construction will sometimes be referred to as the covering class. (Thus, for the Lebesgue measure, the covering class where half-open boxes.) There are three natural questions that surround this construction:

1. Is \( \mu^*(A) = \mu_0(A) \) for sets in the covering class?
2. Is there a restriction of \( \mu^* \) to a \( \sigma \)-algebra where it becomes a measure?
3. If (2) is answered affirmatively, are the sets in the covering class measurable?

In the forthcoming sections we will answer all these at various level of detail.

6.2 Carathéodory measurability and extension theorem.

When \( \mu^*(X) < \infty \), a natural idea how to define measurable sets goes via the inner measure. Indeed, we could set

\[
\mu_*(A) := \mu^*(X) - \mu^*(X \setminus A)
\]

(6.18)

and try to define \( A \) to be measurable when \( \mu^*(A) = \mu_*(A) \). However, this is quickly found to run into a dead end. Indeed, consider the example of \( X := \{1,2,3\} \) with \( \mu^* \) given by

\[
\mu^*(\emptyset) = 0, \quad \mu^*(X) = 3, \quad \mu(A) = 1 \quad \text{for} \quad A \neq \emptyset,X.
\]

(6.19)

Then \( \mu_*(A) = \mu^*(A) \) for all sets \( A \subset X \) and yet \( \mu^* \) is not a measure on \( 2^X \) — e.g., for each \( B := \{1,2\} \) and \( A := \{1\} \) we have \( \mu^*(B) = 1 \) and \( \mu^*(B \cap A) + \mu^*(B \setminus A) = 2 \).

The moral of this example is that, for \( A \) to be measurable, it has to satisfy more than just the condition \( \mu^*(X) = \mu^*(X \cap A) + \mu^*(X \setminus A) \). As it turns out, we have to require that \( \{A,A^c\} \) partitions not just \( X \) additively, but in fact all subsets thereof:
Definition 6.8 (Carathéodory measurability) Let $\mu^*$ be an outer measure on a set $X$. We call $A \subset X$ (Carathéodory) measurable if
\[
\forall E \subset X: \quad \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \setminus A).
\] (6.20)

We will write $\Sigma(\mu^*)$ to denote the class of all (Carathéodory) measurable sets.

Note that, thanks to the subadditivity of $\mu^*$, the nontrivial condition to check is that
\[
\mu^*(E) \geq \mu^*(E \cap A) + \mu^*(E \setminus A).
\] (6.21)

The principal result of this section is then:

Theorem 6.9 (Carathéodory) Let $\mu^*$ be an outer measure on a set $X$. Then:

1. $\Sigma(\mu^*)$ is a $\sigma$-algebra, and
2. $\mu^*$ is a measure on $\Sigma(\mu^*)$.

Proof. The proof comes in four steps.

STEP 1: $\Sigma(\mu^*)$ is an algebra

Obviously, $\emptyset, X \in \Sigma(\mu^*)$ and if $A \in \Sigma(\mu^*)$ then also $A^c \in \Sigma(\mu^*)$. Now let $A, B \in \Sigma(\mu^*)$ and pick an arbitrary set $E \subset X$. Then
\[
E \setminus (A \cup B) = E \cap A^c \cap B^c \quad \text{and} \quad E \cap (A \cup B) = (E \cap A) \cup (E \cap A^c \cap B)
\] (6.22)

and so by subadditivity
\[
\mu^*(E \cap (A \cup B)) + \mu^*(E \setminus (A \cup B)) \\
\leq \mu^*(E \cap A) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B) \\
= \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E)
\] (6.23)

where the first equality is because $A \in \Sigma(\mu^*)$ and the second is due to $B \in \Sigma(\mu^*)$. This shows (6.21) and thus (6.20) for $A \cup B$. Hence, $\Sigma(\mu^*)$ is an algebra.

STEP 2: For any $A_1, \ldots, A_n \in \Sigma(\mu^*)$ disjoint and any $E \subset X$,
\[
\mu^*(E \cap \bigcup_{i=1}^n A_i) = \sum_{i=1}^n \mu^*(E \cap A_i).
\] (6.24)

In particular, $\mu^*$ is finitely additive on $\Sigma(\mu^*)$.

We observe that, since $A_{n+1} \in \Sigma(\mu^*)$ and $A_{n+1}$ is disjoint from $\bigcup_{i=1}^n A_i$,
\[
\mu^*(E \cap \bigcup_{i=1}^{n+1} A_i) = \mu^*\left((E \cap \bigcup_{i=1}^{n+1} A_i) \cap A_{n+1}\right) + \mu^*\left((E \cap \bigcup_{i=1}^{n+1} A_i) \setminus A_{n+1}\right)
\] (6.25)

\[= \mu^*(E \cap A_{n+1}) + \mu^*(E \cap \bigcup_{i=1}^n A_i)
\]

Since the claim holds trivially for $n = 1$, the identity (6.24) follows by induction on $n$. Setting $E := X$ then shows that $\mu^*$ is finitely additive.

STEP 3: $\Sigma(\mu^*)$ is a $\sigma$-algebra
Let \( \{A_n : n \in \mathbb{N} \} \subset \Sigma(\mu^*) \). Then \( \bigcup_{n \in \mathbb{N}} A_n = \bigcup_{n \in \mathbb{N}} A'_n \) where \( A'_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i \) are now disjoint from each other. But \( \Sigma(\mu^*) \) is an algebra and so \( \{A'_n\} \in \Sigma(\mu^*) \) as well. Hence, we may suppose that \( \{A_n\} \) are disjoint. Then (6.24) gives

\[
\mu^*(E) \overset{\text{STEP 1}}{=} \mu^\ast\left(E \cap \bigcup_{i=1}^{n} A_i\right) + \mu^\ast\left(E \setminus \bigcup_{i=1}^{n} A_i\right)
\]

where we also used monotonicity of \( \mu^* \) to replace the finite union by infinite union in the second term on the right. Taking \( n \to \infty \) and using subadditivity of \( \mu^* \) now shows

\[
\mu^*(E) \geq \sum_{i \in \mathbb{N}} \mu^\ast\left(E \cap A_i\right) + \mu^\ast\left(E \setminus \bigcup_{i \in \mathbb{N}} A_i\right)
\]

and so (6.21) holds for the countable union \( \bigcup_{i \in \mathbb{N}} A_i \). Hence \( \Sigma(\mu^*) \) is closed under countable unions and, being an algebra, it is a \( \sigma \)-algebra.

**STEP 4: \( \mu^* \) is a measure on \( \Sigma(\mu^*) \)**

Let \( \{A_n\} \subset \Sigma(\mu^*) \) be disjoint. Taking \( E := \bigcup_{i \in \mathbb{N}} A_i \) in the previous argument shows

\[
\mu^*(\bigcup_{i \in \mathbb{N}} A_i) \geq \sum_{i \in \mathbb{N}} \mu^*(A_i)
\]

and, by countable subadditivity of \( \mu^* \), equality must in fact hold here. Hence \( \mu^* \) is countably additive on \( \Sigma(\mu^*) \) and, since \( \mu^*(\emptyset) = 0 \), it is a measure. \( \square \)

Note that any \( A \subset X \) with \( \mu^*(A) = 0 \) is automatically Carathéodory measurable. As a consequence, the measure space \( (X, \Sigma(\mu^*), \mu^*) \) is automatically complete (see Definition 4.16).

The Carathéodory theorem completely answers question (2) from the end of the class section. To address the remaining two questions, and to make practical use of Carathéodory extension easier, we state and prove:

**Theorem 6.10 (Hahn-Kolmogorov)** Let \( \mathcal{A} \subset 2^X \) be an algebra and assume that \( \mu_0 : \mathcal{A} \to [0, \infty] \) obeys \( \mu_0(\emptyset) = 0 \) and that the following holds:

1. **(finite additivity)** \( A, B \in \mathcal{A}, A \cap B = \emptyset \) implies \( \mu_0(A \cup B) = \mu_0(A) + \mu_0(B) \).
2. **(countable subadditivity)** \( \{A_i\} \subset \mathcal{A} \) and \( \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A} \) imply \( \mu_0(\bigcup_{i \in \mathbb{N}} A_i) \leq \sum_{i \in \mathbb{N}} \mu_0(A_i) \).

Define the outer measure generated by the covering class \( \mathcal{A} \) and set function \( \mu_0 \) by

\[
\mu^*(A) := \inf\left\{ \sum_{i \in \mathbb{N}} \mu_0(A_i) : \{A_i\} \subset \mathcal{A}, \bigcup_{i \in \mathbb{N}} A_i \supset A \right\}.
\]

Then

\[
\mathcal{A} \subset \Sigma(\mu^*) \quad \text{and} \quad \mu^* = \mu_0 \text{ on } \mathcal{A}.
\]

In particular, \( \mu^* \) extends \( \mu_0 \) to a measure on \( \Sigma(\mu^*) \).

**Proof.** The proof comes again in two steps.

**STEP 1:** \( \mathcal{A} \subset \Sigma(\mu^*) \)
Let $A \in \mathcal{A}$ and pick any $E \subset X$. We need to show (6.21). For this we may assume $\mu^*(E) < \infty$. By (6.29), there are sets $\{B_i\} \subset \mathcal{A}$ such that

$$E \subset \bigcup_{i \in \mathbb{N}} B_i \quad \text{and} \quad \mu^*(E) + \varepsilon \geq \sum_{i \in \mathbb{N}} \mu_0(B_i). \quad (6.31)$$

But $\mu_0$ is finitely additive and so $\mu(B_i) = \mu(B_i \cap A) + \mu_0(B_i \setminus A)$ and so

$$\mu^*(E) + \varepsilon \geq \sum_{i \in \mathbb{N}} \mu_0(B_i) = \sum_{i \in \mathbb{N}} \mu_0(B_i \cap A) + \sum_{i \in \mathbb{N}} \mu_0(B_i \setminus A) \geq \mu^*(\bigcup_{i \in \mathbb{N}} B_i \cap A) + \mu^*(\bigcup_{i \in \mathbb{N}} B_i \setminus A) \geq \mu^*(E \cap A) + \mu^*(E \setminus A), \quad (6.32)$$

where we used $\{B_i \cap A\}$ and $\{B_i \setminus A\}$ cover their own unions and then applied (6.29) and the monotonicity of $\mu^*$. As $\varepsilon$ was arbitrary, we have $A \in \Sigma(\mu^*)$.

**STEP 2: $\mu^* = \mu_0$ on $\mathcal{A}$**

Pick $A \in \mathcal{A}$. We have $\mu^*(A) \leq \mu_0(A)$ because $\emptyset \in \mathcal{A}$ and $\mu_0(\emptyset) = 0$. For the opposite inequality let $\{B_i\} \subset \mathcal{A}$ be any collection of sets such that $A \subset \bigcup_{i \in \mathbb{N}} B_i$. Then $\bigcup_{i \in \mathbb{N}}(B_i \cap A) = A$ and so, by the monotonicity and countable subadditivity of $\mu_0$ on $\mathcal{A}$,

$$\sum_{i \in \mathbb{N}} \mu_0(B_i) \geq \sum_{i \in \mathbb{N}} \mu_0(B_i \cap A) \geq \mu_0(A). \quad (6.33)$$

Taking infimum over all such $\{B_i\}$ gives $\mu^*(A) \geq \mu_0(A)$ as desired. \hfill \square

To demonstrate the power of this theorem, let us consider a couple of examples. The first one will be quite familiar:

$$\mathcal{A} := \{A \subset \mathbb{R}^d : \text{ Peano-Jordan measurable} \} \quad \text{and} \quad \mu_0(A) := c(A). \quad (6.34)$$

Then (6.29) yields $\mu^* = \lambda^*$ and so we have a definition of the Lebesgue measure on the class $\Sigma(\lambda^*)$ of Carathéodory measurable sets. However, the relation between $\Sigma(\lambda^*)$ and $\mathcal{L}(\mathbb{R}^d)$, and the values of $\lambda^*$ thereupon, is unclear at this point (we will return to this later).

As another example, pick any set $X$ and consider the collection of sets

$$\mathcal{A} := \{A \subset X : \text{ finite or co-finite} \}. \quad (6.35)$$

This is easily verified to be an algebra and so we set

$$\mu_0(A) := \begin{cases} 0, & \text{if } A \text{ finite,} \\ 1, & \text{if } A \text{ co-finite.} \end{cases} \quad (6.36)$$

Then $\mu_0$ is additive and countably additive on $\mathcal{A}$. The definition (6.29) yields

$$\mu^*(A) := \begin{cases} 0, & \text{if } A \text{ countably infinite,} \\ 1, & \text{else} \end{cases} \quad (6.37)$$

and, as is again checked easily, $\Sigma(\mu^*) = \{A \subset X : \text{ countable or co-countable} \}$. Further examples will be produced in the sections on Lebesgue-Stieltjes measures and product spaces.

In some cases verifying the conditions of the Hahn-Kolmogorov theorem in an algebra can be quite a nuisance. Invariably (as we did in the context of the Lebesgue measure in Lemma 3.4) this is usually easier at the level of semialgebras. We then need the following general fact:
Lemma 6.11  Let \( \mathcal{S} \) be a semialgebra and \( \mu_0 : \mathcal{S} \to [0, \infty] \) obey \( \mu_0(\emptyset) = 0 \). If \( \mu_0 \) is finitely additive and countably subadditive on \( \mathcal{S} \), then so is also its unique extension (guaranteed by Lemma 6.5) to the algebra \( \mathcal{A} \) generated by taking all finite unions of sets from \( \mathcal{S} \).

**Proof.** That additivity of \( \mu_0 \) on \( \mathcal{S} \) implies additivity on \( \mathcal{A} \) was the subject of Lemma 6.5. To show that countable additivity on \( \mathcal{S} \) implies countable additivity on \( \mathcal{S} \), we note that, since \( \mathcal{A} \) is an algebra, it suffices to check this for disjoint unions.

Let \( \{A_n : n \in \mathbb{N}\} \subset \mathcal{S} \) be disjoint sets with \( A := \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{S} \). By Lemma 6.2, each \( A_n \) is a finite disjoint union of sets \( \{B_{nj} : j = 1, \ldots, k(n)\} \) and \( A \) is a finite disjoint union of sets \( \{C_i : i = 1, \ldots, m\} \). By the disjointness assumptions, \( C_i \cap B_{nj} \neq \emptyset \Rightarrow B_{nj} \subset C_i \) (6.38) and setting \( S_i := \{(n, j) : n \in \mathbb{N}, j = 1, \ldots, k(n), B_{nj} \subset C_i\} \) (6.39)

we thus have

\[
C_i = \bigcup_{(n, j) \in S_i} B_{nj} \quad \text{and} \quad \mu_0(C_i) \leq \sum_{(n, j) \in S_i} \mu_0(B_{nj}) \tag{6.40}
\]

where the second conclusion follows by countable subadditivity of \( \mu_0 \) on \( \mathcal{S} \). On the other hand, \( A_n \) can be written as the finite disjoint union

\[
A_n = \bigcup_{i=1}^{m} \bigcup_{j : (n, j) \in S_i} B_{nj} \tag{6.41}
\]

and so finite additivity of \( \mu_0 \) implies

\[
\mu(A_n) = \sum_{i=1}^{m} \sum_{j : (n, j) \in S_i} \mu_0(B_{nj}). \tag{6.42}
\]

Putting this together, we conclude

\[
\mu_0(A) = \sum_{i=1}^{m} \mu_0(C_i) \leq \sum_{i=1}^{m} \sum_{(n, j) \in S_i} \mu_0(B_{nj})
= \sum_{n \in \mathbb{N}} \sum_{i=1}^{m} \sum_{j : (n, j) \in S_i} \mu_0(B_{nj}) = \sum_{n \in \mathbb{N}} \mu_0(A_n) \tag{6.43}
\]

as desired. \( \square \)

### 6.3 Uniqueness of extension.

Having constructed extensions of (non-negative) additive set functions on algebras to measures on \( \sigma \)-algebras, we need to address the question of uniqueness. As already noted, at this point we have two (possibly equivalent) definitions of the Lebesgue measure: one on the Lebesgue \( \sigma \)-algebra \( \mathcal{L}(\mathbb{R}^d) \) of sets that can be approximated, in the sense of outer measure, by open sets from outside, and the other by way of Carathéodory measurability. A question is then how to relate these to each other.

That the extension may not be unique is in fact fairly easy to check by the following example. Let \( \mathcal{A} \) be the algebra consisting of finite unions of all half-open intervals in \( \mathbb{R} \) and consider the set...
function $\mu_0$ defined by $\mu_0(A) = \infty$ if $A \neq \emptyset$ and $\mu_0(\emptyset) = 0$. Now consider three measures on $2^\mathbb{R}$:

- **first**, the *counting measure*
  $$\mu_1(A) := \begin{cases} 
  \text{card}(A), & \text{if } A \text{ finite}, \\
  \infty, & \text{else}
  \end{cases} \quad (6.44)$$

- **second**, the “countability-detection” measure
  $$\mu_2(A) := \begin{cases} 
  0, & \text{if } A \text{ finite or countably infinite}, \\
  \infty, & \text{else}
  \end{cases} \quad (6.45)$$

- **and, third**, the identically-infinite measure
  $$\mu_3(A) := \begin{cases} 
  0, & \text{if } A = \emptyset, \\
  \infty, & \text{else}
  \end{cases} \quad (6.46)$$

Since every non-empty set in $\mathcal{A}$ is automatically infinite and uncountable, we have
$$\mu_1 = \mu_2 = \mu_3 = \mu_0 \quad \text{on } \mathcal{A}. \quad (6.47)$$

Yet, clearly, all three measures are different. (Note also that all sets are of course Carathéodory measurable with respect to these measures, regarded as outer measures.)

It is important to note the highlight the issues that underpin the conclusion of this example. First, either the class $\mathcal{A}$ is too small or the class $2^X$ is too large to make all its sets “reachable” by approximations using sets from $\mathcal{A}$. So the relative “size” of the extension $\sigma$-algebra compared to that of the generating algebra must somehow matter. Also, the fact that the measures are infinite must play a role because the examples rely on the fact that removal of a finite-measure set has no effect on the measure of an infinite-measure set.

We will first address the fact that some restrictions on the “size” of the target $\sigma$-algebra must be made should the extension be unique. Noting the easy fact that an intersection of any number of $\sigma$-algebras on $X$ is a $\sigma$-algebra on $X$, we introduce:

**Definition 6.12** For $\mathcal{M} \subset 2^X$, let
$$\sigma(\mathcal{M}) := \bigcap_{\mathcal{F} \text{ : } \sigma\text{-algebra}} \mathcal{F} : \mathcal{F} \supset \mathcal{M} \quad (6.48)$$

be the least $\sigma$-algebra containing $\mathcal{M}$.

As an example, define
$$\mathcal{B}(\mathbb{R}^d) := \sigma\left( \{O \subset \mathbb{R}^d : \text{open}\} \right) \quad (6.49)$$
to be the class of *Borel sets* in $\mathbb{R}^d$. (Here the open sets are defined using the Euclidean metric.)

As is not hard to check, the same object will be obtained if we replace open sets by half-open boxes. Thus,
$$\mathcal{B}(\mathbb{R}^d) \subset \mathcal{L}(\mathbb{R}^d) \cap \Sigma(\lambda^*) \quad (6.50)$$

and, in fact, $\mathcal{B}(\mathbb{R}^d)$ is the least $\sigma$-algebra on which the Lebesgue measure can be defined.

It is obvious that, if we have a measure on a $\sigma$-algebra, then it naturally restricts to a measure on any sub-$\sigma$-algebra thereof. So if there is a hope to prove uniqueness at all, it should be primarily attempted on $\sigma(\mathcal{A})$. This is the conclusion of:
Theorem 6.13 (Uniqueness criterion) Let $\mathcal{A} \subset 2^X$ be an algebra and let $\mu$ and $\nu$ be finite measures on $\sigma(\mathcal{A})$. Then

$$\mu = \nu \quad \text{on } \mathcal{A} \quad \Rightarrow \quad \mu = \nu \quad \text{on } \sigma(\mathcal{A}). \quad (6.51)$$

The same is true even when $\mu$ and $\nu$ are infinite provided that there are sets $\{X_n\} \subset \mathcal{A}$ such that

$$\bigcup_{n \in \mathbb{N}} X_n = X \quad \text{and} \quad \mu(X_n) = \nu(X_n) < \infty, \quad \forall n \in \mathbb{N}. \quad (6.52)$$

The proof is abstract but also quite instructive as its ideas show up in many other reincarnations elsewhere. First we define:

Definition 6.14 (Monotone class) A class of sets $\mathcal{M} \subset 2^X$ is a monotone class if

1. $\forall \{A_n\} \subset \mathcal{M} : \ A_n \uparrow A$ implies $A \in \mathcal{M}$,
2. $\forall \{A_n\} \subset \mathcal{M} : \ A_n \downarrow A$ implies $A \in \mathcal{M}$.

In words, a monotone class is a class that is closed under increasing and decreasing limits.

It is easy to check that any $\sigma$-algebra is a monotone class. The class of all intervals in $\mathbb{R}$ — open, half-open and closed including degenerate ones — is a monotone class but not a $\sigma$-algebra. However, once a monotone class contains an algebra, it contains also the least $\sigma$-algebra generated thereby:

Lemma 6.15 (Monotone Class Lemma) Let $X$ be any set. If $\mathcal{M} \subset 2^X$ is a monotone class and $\mathcal{A} \subset 2^X$ an algebra, then

$$\mathcal{A} \subset \mathcal{M} \quad \Rightarrow \quad \sigma(\mathcal{A}) \subset \mathcal{M}. \quad (6.53)$$

Proof: As is easy to check, the intersection of any number of monotone classes is a monotone class, so it makes sense to define

$$m(\mathcal{A}) := \bigcap_{\mathcal{M} : \ \mathcal{M} \supseteq \mathcal{A}, \ \mathcal{M} \text{ is a monotone class}} \mathcal{M} \quad (6.54)$$

Now $\sigma(\mathcal{A})$ is a monotone class containing $\mathcal{A}$ and so $m(\mathcal{A}) \subseteq \sigma(\mathcal{A})$. Our goal is to show that $\sigma(\mathcal{A}) \subseteq m(\mathcal{A})$ because that, by the minimality of $m(\mathcal{A})$, implies the claim.

First we note that $\emptyset, X \in m(\mathcal{A})$ by their containment in $\mathcal{A}$. Also, since

$$\mathcal{M}' := \{A^c : A \in \mathcal{M}\} \quad (6.55)$$

is a monotone class if $\mathcal{M}$ is a monotone class, we have $A \in m(\mathcal{A})$ if and only if $A^c \in m(\mathcal{A})$. We will now show that $m(\mathcal{A})$ is closed under finite intersection. For this we let

$$\mathcal{M}_A := \{C \in m(\mathcal{A}) : C \cap A \in m(\mathcal{A})\}. \quad (6.56)$$

Now we observe that, since $\mathcal{A}$ is itself closed under intersections and $m(\mathcal{A})$ is a minimal monotone class containing $\mathcal{A}$,

$$\forall A \in \mathcal{A} : \quad \mathcal{A} \subseteq \mathcal{M}_A \quad \text{and so} \quad m(\mathcal{A}) \subseteq \mathcal{M}_A. \quad (6.57)$$

But the conclusion can be rewritten as

$$\forall B \in m(\mathcal{A}) : \quad \mathcal{A} \subseteq \mathcal{M}_B \quad \text{and so} \quad m(\mathcal{A}) \subseteq \mathcal{M}_B. \quad (6.58)$$
This means that if \( A, B \in m(\mathcal{A}) \), then also \( A \cap B \in m(\mathcal{A}) \), i.e., \( m(\mathcal{A}) \) is closed under finite intersections and, in light of the above, it is an algebra. But \( m(\mathcal{A}) \) is also a monotone class and so it is closed under countable intersections as well. It follows that \( m(\mathcal{A}) \) is a \( \sigma \)-algebra containing \( \mathcal{A} \) and so \( m(\mathcal{A}) \supseteq \sigma(\mathcal{A}) \) by the minimality of \( \sigma(\mathcal{A}) \).

**Proof of Theorem 6.13.** Let \( \mu \) and \( \nu \) be finite measures on a \( \sigma \)-algebra \( \mathcal{F} \). Then, by the Monotone Convergence Theorem for sets,

\[
\mathcal{M} := \{ A \in \mathcal{F} : \mu(A) = \nu(A) \}
\]

is a monotone class. So if \( \mathcal{A} \) is an algebra and \( \mu = \nu \) on \( \mathcal{A} \), we have \( \mathcal{A} \subset \mathcal{M} \) and, by the Monotone Class Lemma, \( \sigma(\mathcal{A}) \subset \mathcal{M} \). Hence, \( \mu = \nu \) on \( \sigma(\mathcal{A}) \).

To address the infinite-measure case, consider the sets \( \{ X_n \} \in \mathcal{A} \) such that \( \mu(X_n) = \nu(X_n) < \infty \). Without loss of generality (take finite unions), we may assume that \( X_n \uparrow X \). Define measures

\[
\mu_n(A) := \mu(A \cap X_n) \quad \text{and} \quad \nu_n(A) := \nu(A \cap X_n).
\]

These are finite measures and \( \mu_n = \nu_n \) on \( \mathcal{A} \). Hence, \( \mu_n = \nu_n \) on \( \sigma(\mathcal{A}) \) by the first part of the claim. But \( \mu_n(A) \uparrow \mu(A) \) and \( \nu_n(A) \uparrow \nu(A) \) by the Monotone Convergence Theorem for sets, and so \( \mu = \nu \) on \( \sigma(\mathcal{A}) \) as well. \( \square \)

### 6.4 Lebesgue-Stieltjes measures.

The Hahn-Kolmogorov Theorem gives us a way to construct a large class of measures on the real line. Recall that \( \mathcal{B}(\mathbb{R}) \) denotes the class of Borel sets in \( \mathbb{R} \) (see (6.49)). Then we have:

**Theorem 6.16** Let \( F : \mathbb{R} \to \mathbb{R} \) be a non-decreasing, right-continuous function. Then there exits a unique measure \( \mu_F \) on \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \) such that

\[
\forall a < b : \quad \mu_F((a,b]) = F(b) - F(a).
\]

The key point is to show:

**Lemma 6.17** Let \( \mathcal{I} := \{(a,b] : -\infty \leq a \leq b \leq \infty \} \). Then \( \mathcal{I} \) is a semialgebra on \( \mathbb{R} \) and

\[
\mu_0((a,b]) := F(b) - F(a)
\]

is finitely additive and countably subadditive on \( \mathcal{I} \).

**Proof.** That \( \mathcal{I} \) is a semialgebra is verified easily. It is also immediate that \( \mu_0 \) is finitely additive (see the first part of the proof of Lemma 2.4). To prove countable additivity, thanks to Lemma 6.11, we need to verify that for any disjoint collection \( \{(a_n, b_n] : n \in \mathbb{N}\} \subset \mathcal{I} \) and any \( (a,b) \in \mathcal{I} \),

\[
(a,b) = \bigcup_{n \in \mathbb{N}} (a_n, b_n] \quad \Rightarrow \quad F(b) - F(a) \leq \sum_{n \in \mathbb{N}} (F(b_n) - F(a_n)).
\]

By intersection with a finite interval we may assume that \( (a,b] \) is finite. Then \( F(b) - F(a) < \infty \) and so, by the right-continuity of \( F \), given \( \varepsilon > 0 \) there is \( a' > a \) so that \( F(a') < F(a) + \varepsilon \). Similarly, for each \( n \in \mathbb{N} \), we can find \( b'_n > b_n \) so that \( F(b'_n) < F(b_n) + \varepsilon 2^{-n} \). Since

\[
[a', b] \subset (a,b] = \bigcup_{n \in \mathbb{N}} (a_n, b_n] \subset \bigcup_{n \in \mathbb{N}} (a_n, b'_n]
\]

(6.64)
the open intervals \{ (a_n, b_n') \} constitute a cover of a compact interval \([a', b]\). By the Heine-Borel Theorem, there is a finite subcover.

Let \((\alpha_1, \beta_1'), \ldots, (\alpha_N, \beta_N')\) denote the intervals in a finite subcover so that each interval contains a point that is not contain by the other intervals. We may assume that the intervals are labeled so that \(\alpha_1 < \alpha_2 < \cdots < \alpha_N\), where the strict inequality must hold by disjointness of \{ \{a_n, b_n\} \}. Since each \((\alpha_i, \beta_i')\) contains a point which does not lie in any other interval, we must have

\[
\alpha_i < a' \quad \text{and} \quad \beta_N > b \quad \text{and} \quad \beta_i' > \alpha_{i+1}, \quad i = 1, \ldots, N - 1. \tag{6.65}
\]

By the monotonicity of \(F\), our choices of \(a'\) and \(b_n\) and (6.65) we thus have

\[
F(b) - F(a) - \varepsilon \leq F(b) - F(a') \leq F(\beta_N') - F(\alpha_1) \\
= F(\beta_N') - F(\alpha_N) + \sum_{i=1}^{N-1} (F(\alpha_{i+1}) - F(\alpha_i)) \\
\leq \sum_{i=1}^{N} (F(\beta_i') - F(\alpha_i)) \tag{6.66}
\]

\[
\leq \sum_{n \in \mathbb{N}} (F(b_n') - F(a_n)) \leq \varepsilon + \sum_{n \in \mathbb{N}} (F(b_n) - F(a_n)).
\]

Since \(\varepsilon\) was arbitrary, we get (6.63). \(\square\)

**Proof of Theorem 6.16.** By Lemma 6.11 and the Hahn-Kolmogorov Theorem, \(\mu_0\) from the previous lemma extends to a measure on \(\sigma(\mathcal{A}) = \mathcal{B}(\mathbb{R})\) which agrees with \(\mu_0\) on \(\mathcal{A}\), i.e., (6.61) holds. Since \(\mu_0((n, n+1]) = F(n+1) - F(n) < \infty\) for all \(n \in \mathbb{N}\), the second part of Theorem 6.13 applies and so \(\mu_F\) is unique. \(\square\)

We remark that the construction actually gives an extension to the \(\sigma\)-algebra of Carathéodory measurable sets \(\Sigma(\mu_F^\sigma)\). This \(\sigma\)-algebra may vary quite dramatically depending on \(F\). Here are some natural examples of measures that can be identified via the above construction:

1. if \(F(x) := x\) then \(\mu_F = \lambda\), the Lebesgue measure on \(\mathcal{B}(\mathbb{R})\),
2. if \(F(x) := 1_{\{x \geq a\}}\) then \(\mu_F\) is the Dirac point mass at \(a\), often denoted by \(\delta_a\), such that

\[
\delta_a(A) := \begin{cases} 
1, & \text{if } a \in A, \\
0, & \text{else}.
\end{cases} \tag{6.67}
\]

Note that \(\Sigma(\mu_F^\sigma) = 2^\mathbb{R}\) in this case.

3. if \(F(x) := \sum_{n \in \mathbb{N}} p_n 1_{\{x \geq a_n\}}\) for some \(\{a_i\} \subset \mathbb{R}\) and \(p_i \in (0, \infty)\) with \(\sum_{i \in \mathbb{N}} p_i < \infty\), then

\[
\mu_F(A) = \sum_{i \in \mathbb{N}} p_i \delta_{a_i}(A), \tag{6.68}
\]

i.e., \(\mu_F\) is a collection of point masses with mass at \(a_i\) equal to \(p_i\). (Note that, in general, \(\mu_F\) will charge a singleton \(\{a\}\) if and only if \(F\) is discontinuous at \(a\).) Again \(\Sigma(\mu_F^\sigma) = 2^\mathbb{R}\), i.e., all subsets of \(\mathbb{R}\) are measurable.

4. if \(F\) is continuously differentiable, then \(F(b) - F(a) = \int_a^b F'(x) \, dx = \int_{(a,b]} F' \, d\lambda\) and \(\mu_F\) is then the measure given by

\[
\mu_F(A) = \int_A F' \, d\lambda. \tag{6.69}
\]

The Cantor-Lebesgue function associated with the middle-third Cantor set (marked on the x-axis). This function is sometimes referred to as the “devil’s staircase” in the physics literature.

Note that, reversing this process gives us a way to define a meaning of $f \, d\lambda$, i.e., multiplying the Lebesgue measure by an integrable function.

(5) $F$ is the Cantor-Lebesgue function that is defined as follows: Let $C_n$ be the $n$-th iteration of the process of “removal of middle-third intervals” defining to the middle-third Cantor set in $C_0 := [0, 1]$. Then $C_{n-1} \setminus C_n$ is a collection of $2^n$ open intervals of length $3^{-n}$. Examining these intervals “left to right”, $F$ is constant on each of these and takes values in increasing non-negative odd multiples of $2^{-n}$. See Fig. 1

Let $C := \bigcap_{n \geq 1} C_n$ be the middle-third Cantor set. Then

$$\mu_F([0, 1] \setminus C) = 0$$  \hspace{1cm} (6.70)

because this set is a union of open intervals on which $F$ is piecewise constant. Yet

$$\mu_F([0, 1]) = F(1) - F(0) = 1$$  \hspace{1cm} (6.71)

Since $F$ is also continuous, $\mu_F(\{a\}) = 0$ for all $a$. Note that $\lambda(C) = 0$.

The defining property of the measure $\mu_F$ suggests a natural connection to the Riemann-Stieltjes integration. And, indeed, the Lebesgue integral

$$\int_{[a,b]} f \, d\mu_g$$  \hspace{1cm} (6.72)

bears the same connection to the Riemann-Stieltjes integral

$$\int_a^b f \, dg$$  \hspace{1cm} (6.73)
as the Lebesgue integral (with respect to the Lebesgue measure) bears to the Riemann integral. Thus, for instance, it can be checked that, for a non-decreasing, right-continuous function, the function \( f \) is Riemann-Stieltjes integrable with respect to \( dg \) on \([a, b]\) if and only if \( f \) is \( \mu_g \)-a.e. continuous. Thanks to this connection, we have:

**Definition 6.18** A measure \( \mu \) on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) is a Lebesgue-Stieltjes measure if there is a non-decreasing, right-continuous function \( F \) such that (6.61) holds for \( \mu \). (Then, of course, \( \mu = \mu_F \).)

The class of Lebesgue-Stieltjes measures is quite rich. Indeed, every measure \( \mu \) on \((\mathbb{R}, \mathcal{B}(\mathbb{R}))\) that assigns finite values to compact sets — such Borel measures are called Radon measures — is a Lebesgue-Stieltjes measure. The function \( F \) such that \( \mu = \mu_F \) is then given by

\[
F(x) := \begin{cases} 
\mu([0,x]), & \text{if } x \geq 0, \\
-\mu((x,0)), & \text{if } x < 0.
\end{cases}
\]  
(6.74)

An example of a measure that is not of this kind is the counting measure of the rationals,

\[
\mu(A) := \begin{cases} 
\text{card}(A \cap \mathbb{Q}), & \text{if } A \cap \mathbb{Q} \text{ finite,} \\
\infty, & \text{else,}
\end{cases}
\]  
(6.75)

Obviously, there is no \( F : \mathbb{R} \to \mathbb{R} \) that could make a unit upward jump at every rational.
7. MORE ON OUTER MEASURES

The proof of uniqueness of (the extension of) measures requires restriction to the coarsest \( \sigma \)-algebra containing the generating class of sets. So, in the particular case of the Lebesgue measure on \( \mathbb{R}^d \), we know that the restriction of \( \lambda^* \) to \( \mathcal{B}(\mathbb{R}^d) \) is unique. However, this still leaves open the question whether the extension to the class of Lebesgue measurable sets \( \mathcal{L}(\mathbb{R}^d) \) is unique, and also how that class is related to the Carathéodory measurable sets \( \Sigma(\lambda^*) \). This is what we will answer here, apart from many other things.

7.1 Regular outer measures.

Expounding on the example that we just discussed, recall that, by Corollary 3.14,

\[
\mathcal{L}(\mathbb{R}^d) = \{ F \cup N : F \in \mathcal{F}_\sigma(\mathbb{R}^d), F \cap N = \emptyset, \lambda^*(N) = 0 \},
\]

\[
= \{ G \setminus N : G \in \mathcal{F}_\delta(\mathbb{R}^d), N \subset G, \lambda^*(N) = 0 \},
\]

(7.1)

where \( \mathcal{F}_\sigma(\mathbb{R}^d) \), resp., \( \mathcal{F}_\delta(\mathbb{R}^d) \) are as in (3.43), resp., (3.44). Lebesgue measurable sets can thus be approximated from outside by \( \mathcal{F}_\delta \) sets up to a null set. The following attempts to generalize this approximation property to arbitrary outer measures:

**Definition 7.1** An outer measure \( \mu^* \) on \( X \) is said to be regular if

\[
\forall A \in X \ \exists B \in \Sigma(\mu^*) : \ B \supset A \ \text{and} \ \mu^*(B) = \mu^*(A).
\]

(7.2)

We call such \( B \) a measurable cover of \( A \).

Relying heavily on the setting of the Hahn-Kolmogorov Theorem, we then have:

**Theorem 7.2** If \( \mu^* \) is obtained via (6.29) from an additive (non-negative) set function \( \mu_0 \) on an algebra \( \mathcal{A} \), then \( \mathcal{A} \subset \Sigma(\mu^*) \) and \( \mu^* \) is regular. If in addition \( \mu_0 \) is countably subadditive and there are \( \{ X_n \} \in \mathcal{A} \) such that \( \bigcup_{n \in \mathbb{N}} X_n = X \) and \( \mu_0(X_n) < \infty \), then

\[
\Sigma(\mu^*) = \{ B \cup N : B \in \mathcal{A}_\sigma, B \cap N = \emptyset, \mu^*(N) = 0 \}.
\]

(7.3)

Here, we recall, \( \mathcal{A}_\sigma \) is the class of countable unions of countable intersections of sets from \( \mathcal{A} \).

Before we begin the proof, we note that this implies:

**Corollary 7.3** \( \mathcal{L}(\mathbb{R}^d) = \Sigma(\lambda^*) \).

**Proof.** The outer measure is generated by the algebra \( \mathcal{A} \) of finite unions of half-open boxes. Since \( \mathcal{A}_\sigma \) contains all open sets, \( \mathcal{B}(\mathbb{R}^d) \subset \Sigma(\lambda^*) \cap \mathcal{L}(\mathbb{R}^d) \). Then \( \mathcal{F}_\sigma(\mathbb{R}^d) \subset \mathcal{B}(\mathbb{R}^d) \subset \Sigma(\lambda^*) \) and \( \mathcal{A}_\sigma \subset \mathcal{B}(\mathbb{R}^d) \subset \mathcal{L}(\mathbb{R}^d) \). Now both \( \mathcal{L}(\mathbb{R}^d) \) and \( \Sigma(\lambda^*) \) contain all \( N \subset \mathbb{R}^d \) with \( \lambda^*(N) = 0 \) and so the equality follows by comparing the top line of (7.1) with (7.3).

Moving to the proof of Theorem 7.2, we begin with a lemma:

**Lemma 7.4** Suppose \( \mu^* \) is defined as in Proposition 6.7 from a covering class \( \mathcal{C} \) containing \( X \) and a set function \( \mu_0 \). Then

\[
\mu^*(A) = \inf \{ \mu^*(B) : B \in \mathcal{C}_\sigma, B \supset A \}.
\]

(7.4)
Proof. Let $A \subset X$. If $\mu^*(A) = \infty$ we use $B := X$ so let us suppose $\mu^*(A) < \infty$. By the definition of $\mu^*$, given $\epsilon > 0$ there are $\{A_n\} \subset \mathcal{C}$ with $\bigcup_{n \in \mathbb{N}} A_n \supset A$ and $\sum_{n \in \mathbb{N}} \mu_0(A_n) \leq \mu^*(A) + \epsilon$. Taking $B := \bigcup_{n \in \mathbb{N}} A_n$ we then have $\mu^*(B) \leq \mu^*(A) + \epsilon$. Since $B \in \mathcal{C}_\sigma$, the claim follows. \hfill \Box

Lemma 7.5 If $\mu^*$ is as in Theorem 7.2, then

$$\forall A \subset X \exists B \in \mathcal{A}_{\sigma}\delta : \quad B \supset A \quad \text{and} \quad \mu^*(B) = \mu^*(A).$$ \hfill (7.5)

Proof. Now take a sequence $\{B_n\} \subset \mathcal{A}_\sigma$ with $\mu^*(B_n) \downarrow \mu^*(A)$. Since $\mathcal{A}$ is an algebra, $\mathcal{A}_\sigma$ is closed under finite intersections. Thus, without loss of generality, we may take $B_n$ decreasing, $B_n \downarrow B$. Then $A \subset B \subset B_n$ and so

$$\mu^*(A) \leq \mu^*(B) \leq \mu^*(B_n) \xrightarrow{n \to \infty} \mu^*(A)$$ \hfill (7.6)

i.e., $\mu^*(B) = \mu^*(A)$. Obviously, $B \in \mathcal{A}_{\sigma}\delta$ and so we are done. \hfill \Box

Proof of Theorem 7.2. By checking the proof of Theorem 6.10 we find that $\mathcal{A} \subset \Sigma(\mu^*)$ whenever $\mathcal{A}$ is an algebra and $\mu_0$ is a finitely additive set function on it. But then $\mathcal{A}_{\sigma}\delta \subset \Sigma(\mu^*)$ and so regularity of $\mu^*$ follows from Lemma 7.5.

If $\mu_0$ is countably subadditive, Theorem 6.10 shows $\mu^* = \mu_0$ on $\mathcal{A}$. Let $\{X_n\} \subset \mathcal{A}$ be the sets with the stated properties. Since these lie in an algebra, by taking $X'_n = X_n \setminus (X_1 \cup \ldots \cup X_{n-1})$ instead of $X_n$ we may assume that they are disjoint. Let $A \in \Sigma(\mu^*)$. By Lemma 7.5, for each $n \in \mathbb{N}$ there is $B_n \in \mathcal{A}_{\sigma}\delta$ and a $\mu^*$-null set $N_n$ such that

$$N_n \subset B_n \subset X_n$$ \hfill (7.7)

and

$$A^c \cap X_n = B_n \setminus N \quad \text{with} \quad \mu^*(A^c \cap X_n) = \mu^*(B_n).$$ \hfill (7.8)

Then $A \cap X_n = (X_n \setminus B_n) \cup N_n$ and so

$$A = \bigcup_{n \in \mathbb{N}} (X_n \setminus B_n) \cup \bigcup_{n \in \mathbb{N}} N_n.$$ \hfill (7.9)

Denote the first union by $B$ and the second union by $N$. As $B_n \in \mathcal{A}_{\sigma}\delta$ implies $X_n \setminus B_n \in \mathcal{A}_{\sigma}\delta$ we have $B \in \mathcal{A}_{\sigma}\delta$. On the other hand, (7.7) and the disjointness of $\{X_n\}$ show $B \cap N = \emptyset$. Since $A = B \cup N$, the claim thus follows. \hfill \Box

Before now we already knew that $\Sigma(\mu^*)$ is a $\sigma$-algebra containing $\mathcal{B}(\mathbb{R}^d)$ and all $\mu^*$-null sets. Such $\sigma$-algebras is called a completion of $\mathcal{B}(\mathbb{R}^d)$ with respect to $\mu^*$. The above theorem now shows that $\Sigma(\mu^*)$ is minimal $\sigma$-algebra of this form. Recall that, for Lebesgue-Stieltjes outer measures $\mu^*_F$, this is exactly the reason why $\Sigma(\mu^*_F)$ can be very different for different $F$.

In order to appreciate regularity of outer measures better, we note the following fact:

Theorem 7.6 Let $\mu^*$ be a regular outer measure on $X$ with $\mu^*(X) < \infty$. Then

$$A \in \Sigma(\mu^*) \iff \mu^*(X) = \mu^*(A) + \mu^*(X \setminus A).$$ \hfill (7.10)

In other words, $A$ is $\mu^*$-measurable if and only if the outer and inner measures of $A$ are equal.

Proof. The direction $\Rightarrow$ is immediate from the definition of Carathéodory measurability so we will focus on $\Leftarrow$. Pick $E \subset X$. Then $E$ has a measurable cover $B \supset E$. If $A$ is a set satisfying the
Proof. Suppose first that \( \mu \) is actually quite general: the condition on the right of the claim then, the additivity of \( \mu^* \) on \( \Sigma(\mu^*) \) and subadditivity of \( \mu^* \) in general then yield
\[
\mu^*(E) = \mu^*(B) = \mu^*(X) - \mu^*(B^c) \\
= \mu^*(A) + \mu^*(X \setminus A) - \mu^*(B^c) \\
\geq \mu^*(A) + \mu^*(X \setminus A) - \mu^*(A \cap B^c) - \mu^*(B^c \setminus A) \\
= (\mu^*(A) - \mu^*(A \cap B^c)) + (\mu^*(X \setminus A) - \mu^*(B^c \setminus A)) \\
= \mu^*(A \cap B) + \mu^*(B \setminus A) \geq \mu^*(B) = \mu^*(E).
\]
(7.11)

Since \( E \) was arbitrary, \( A \) is \( \mu^* \)-measurable.

Not every outer measure is regular. Here is an example: Let \( \{a_n : n \geq 0\} \) be a sequence satisfying \( 0 = a_0 < \frac{1}{2} < a_1 < a_2 < \cdots < 1 \) with \( a_n \rightarrow 1 \). Let \( X := \mathbb{N} \) and define
\[
\mu^*(A) := \begin{cases} 
\text{card}(A), & \text{if } A \text{ finite}, \\
1, & \text{else}. 
\end{cases}
\]
(7.12)

Then \( \mu^* \) is an outer measure (thanks to \( a_n > 1/2 \) for \( n \geq 1 \)) but it is not regular as \( \Sigma(\mu^*) = \{\emptyset, X\} \) and so there are no measurable sets that achieve values in \( \{a_i : i \in \mathbb{N}\} \).

### 7.2 Borel regular and Radon measures.

The notion of regularity naturally appears also in the context of Borel measures. These are measures on the least \( \sigma \)-algebra containing all open sets in a metric, or even topological, space. We will denote the Borel \( \sigma \)-algebra on \( X \) by \( \mathcal{B}(X) \).

The following lemma shows that the claim we originally proved for the Lebesgue measure is actually quite general:

**Lemma 7.7** Let \( \mu \) be a measure on \( (X, \mathcal{B}(X)) \). Then for all \( A \in \mathcal{B}(X) \) with \( \mu(A) < \infty \),
\[
\mu(A) = \sup\{\mu(C) : C \subset X, \text{ closed}\}
\]
(7.13)

*Proof.* Suppose first that \( \mu(X) < \infty \) and let
\[
\mathcal{M} := \{A \in \mathcal{B}(X) : (7.13) \text{ holds}\}.
\]
(7.14)

We claim that \( \mathcal{M} \) is a monotone class. Indeed, if \( \{A_n\} \subset \mathcal{M} \) with \( A_n \downarrow A \), then for each \( \epsilon > 0 \) there are \( \{C_n\} \) closed with \( C_n \subset A_n \) and \( \mu(A_n \setminus C_n) < \epsilon 2^{-n} \). Set \( C := \bigcap_{n \in \mathbb{N}} C_n \). Then
\[
A \setminus C = \left( \bigcap_{n \in \mathbb{N}} A_n \right) \setminus \left( \bigcap_{n \in \mathbb{N}} C_n \right) \subset \bigcup_{n \in \mathbb{N}} (A_n \setminus C_n)
\]
(7.15)

and so
\[
\mu(A \setminus C) \leq \sum_{n \in \mathbb{N}} \mu(A_n \setminus C_n) < \sum_{n \in \mathbb{N}} \epsilon 2^{-n} = \epsilon.
\]
(7.16)

Since \( C \) is closed, we have \( A \in \mathcal{M} \).

Now assume instead \( A_n \uparrow A \) and pick \( \{C_n\} \) as before. Again,
\[
A \setminus C = \left( \bigcup_{n \in \mathbb{N}} A_n \right) \setminus \left( \bigcup_{n \in \mathbb{N}} C_n \right) \subset \bigcup_{n \in \mathbb{N}} (A_n \setminus C_n)
\]
(7.17)
and so $\mu(A \setminus C) < \epsilon$. But then also

$$\mu \left( A \setminus \bigcup_{k=1}^{n} C_k \right) < \epsilon$$  \hspace{1cm} (7.18)

because as $n$ increases the left hand side decreases, by the Downward Monotone Convergence Theorem for sets, to $\mu(A \setminus C)$. Since $\bigcup_{k=1}^{n} C_k$ is closed, $A \in \mathcal{M}$ in this case as well.

It follows that $\mathcal{M}$ is a monotone class. Since $\mathcal{M}$ trivially contains all closed sets, it contains also $\mathcal{B}(X)$. This is the stated claim whenever $\mu(X) < \infty$. In the case when $\mu(X) = \infty$, apply this instead to the measure $\mu_A(B) := \mu(A \cap B)$ which is finite by the assumption that $\mu(A) < \infty$. $\Box$

By complementation, we then also have:

**Lemma 7.8** Let $\mu$ be a measure on $(X, \mathcal{B}(X))$ with $\mu(X) < \infty$. Then

$$\forall A \in \mathcal{B}(X): \quad \mu(A) = \inf \{ \mu(O): O \supset A, \text{open} \}.$$ \hspace{1cm} (7.19)

The same is true even if $\mu(X) = \infty$ provided there are open sets $\{O_n\}$ with $X = \bigcup_{n \in \mathbb{N}} O_n$ and $\mu(O_n) < \infty$ for all $n$.

**Proof.** If $\mu(X) < \infty$ pick $A \in \mathcal{B}(X)$ and note that, by the previous lemma, for each $\epsilon > 0$ there is $C \subset A^c$ closed such that $\mu(A^c \setminus C) < \epsilon$. Now define $O := X \setminus C$ and note that $O$ is open. Since $A \setminus C = O \setminus A$ we have

$$\mu(A) \leq \mu(O) = \mu(A) + \mu(O \setminus A) < \mu(A) + \epsilon.$$ \hspace{1cm} (7.20)

This yields (7.19).

In the case when $\mu(X) = \infty$ let $\{O_n\}$ be open, $O_n \uparrow X$ with $\mu(O_n) < \infty$. Considering the measure $\mu_n(A) := \mu(O_n \cap A)$, the finite version of the claim shows that, for each $A \in \mathcal{B}(X)$ and each $\epsilon > 0$, there is an open set $O_n' \supset A$ such that

$$\mu(O_n \cap O_n' \setminus (O_n \cap A)) = \mu_n(O_n' \setminus A) < \epsilon 2^{-n}.$$ \hspace{1cm} (7.21)

Now set $O := \bigcup_{n \in \mathbb{N}} (O_n' \cap O_n)$. Then $A \cap O_n \subset O_n' \cap O_n$ implies $A \subset O$ and

$$O \setminus A = \left( \bigcup_{n \in \mathbb{N}} (O_n \cap O_n') \right) \setminus \left( \bigcup_{n \in \mathbb{N}} (O_n \cap A) \right) \subset \bigcup_{n \in \mathbb{N}} (O_n \cap O_n' \setminus (O_n \cap A)).$$ \hspace{1cm} (7.22)

Hence we get $\mu(O \setminus A) < \epsilon$ as desired. $\Box$

We now compare the above with the properties in the following definition:

**Definition 7.9** (Borel regularity and Radon measures) Let $\mu$ be a measure on the $\sigma$-algebra of Borel sets $\mathcal{B}(X)$. We say that

1. $\mu$ is outer regular if

$$\forall A \in \mathcal{B}(X): \quad \mu(A) = \inf \{ \mu(O): O \supset A, \text{open} \}.$$ \hspace{1cm} (7.23)

2. $\mu$ is inner regular if

$$\forall A \in \mathcal{B}(X): \quad \mu(A) = \sup \{ \mu(K): K \subset A, \text{compact} \}.$$ \hspace{1cm} (7.24)
An outer and inner regular Borel measure is called regular. An inner regular Borel measure obeying
\[ \forall K \subset X \text{ compact: } \mu(K) < \infty \] (7.25)
is called a Radon measure.

For finite measures the above lemmas imply that every finite Borel measure is automatically outer regular. For inner regularity one in addition needs that $X$ can be written as an increasing union of a sequence of compact sets. Not surprisingly, Borel regular measures play an important role in various technical aspects of probability theory. As is not hard to check, the Lebesgue measure on $\mathcal{B}(\mathbb{R})$ and, in fact, all Lebesgue-Stieltjes measures, are Borel regular and, in fact, are Radon measures.

We remark that the notion of regular Borel measure, or even Borel measure, has a slightly fuzzy meaning in the literature. For instance, the book *Measure and Integration* by S.K. Berberian, calls a measure on $(X, \mathcal{B}(X))$ Borel only if it assigns finite values to all compact sets. (In this reference, the Borel sets are those generated by compact sets only while elements in the above class $\mathcal{B}(X)$ are referred to as weakly Borel sets.) We will revisit these subtleties when we talk about Baire sets and Baire measures.

### 7.3 Metric outer measures.

An important property of the Lebesgue measure — used heavily in the proof Proposition 3.10 — is the fact that $\rho(A, B) > 0$ implies that $\lambda^*$ acts additively on $A \cup B$; cf Lemma 3.12. This property is so important that it is given a name:

**Definition 7.10** Let $(X, \rho)$ be a metric space and $\mu^*$ an outer measure on $X$. We say that $\mu^*$ is a metric outer measure if
\[ \forall A, B \subset X: \quad \rho(A, B) > 0 \Rightarrow \mu^*(A \cup B) = \mu^*(A) + \mu^*(B). \] (7.26)

The property of being a metric outer measure is quite strong. Indeed, we have:

**Theorem 7.11** Let $\mu^*$ be an outer measure on a metric space $(X, \rho)$. Then
\[ \mathcal{B}(X) \subset \Sigma(\mu^*) \iff \mu^* \text{ is metric.} \] (7.27)

One direction of proof is quite straightforward:

**Proof of $\Rightarrow$.** Let $A, B \subset X$ be two sets such that $\rho(A, B) > 0$. Define
\[ O := \bigcup_{x \in A} \{ y \in X : \rho(x, y) < \frac{1}{2} \rho(A, B) \} \] (7.28)
Then $O$ is open with $O \supseteq A$ and $O \cap B = \emptyset$. Now $O \in \mathcal{B}(X)$ so $\mathcal{B}(X) \subset \Sigma(\mu^*)$ implies that $O \in \Sigma(\mu^*)$. By the definition of Carathéodory measurability,
\[ \mu^*(A \cup B) = \mu^*((A \cup B) \cap O) + \mu^*((A \cup B) \setminus O) \]
\[ = \mu^*(A) + \mu^*(B). \] (7.29)

Hence, $\mu^*$ is a metric outer measure. \[
\square
\]

For the other direction, we have to work harder. First we prove a lemma:
Lemma 7.12  Let $\mu^*$ be metric outer on $(X, \rho)$. Let $A \subset X$ and suppose there is a nonempty and closed set $K$ such that $K \cap A = \emptyset$. Then

$$A_n := \left\{ x \in A : \rho(x, K) > \frac{1}{n} \right\}$$

(7.30)

obeys

$$A_n \uparrow A \quad \text{and} \quad \mu^*(A_n) \uparrow \mu^*(A).$$

(7.31)

Proof. Obviously, $A_n \uparrow A' := \bigcup_{n \in \mathbb{N}} A_n$ with $A' \subset A$. To see that $A' = A$ note that if $x \in A$ then $\rho(x, K) > 0$. So $x \in A_n$ as soon as $n$ is so large that $\rho(x, K) > \frac{1}{n}$.

To get that also the outer measures converge, we may assume that $\sup_{n \in \mathbb{N}} \mu^*(A_n) < \infty$. Define

$$B_n := A_{n+1} \setminus A_n$$

(7.32)

and note that $\rho(B_n, B_k) > 0$ as soon as $|m-n| > 1$. Since $\mu^*$ is metric, this shows

$$\sum_{m=1}^n \mu^*(B_{2m}) = \mu^*\left( \bigcup_{m=1}^n B_{2m} \right) \leq \mu^*(A_{2n}) \leq \sup_{n \in \mathbb{N}} \mu^*(A_n) < \infty$$

(7.33)

and so

$$\sum_{m \in \mathbb{N}} \mu^*(B_{2m}) < \infty. \quad (7.34)$$

Similarly, doing this for odd-numbered $B_m$’s, we also get

$$\sum_{m \in \mathbb{N}} \mu^*(B_{2m-1}) < \infty. \quad (7.35)$$

Now

$$A \subset A_{2n} \cup \bigcup_{m \geq n} B_{2m} \cup \bigcup_{m \geq n} B_{2m+1}$$

(7.36)

and so, by countable subaditivity of $\mu^*$,

$$\mu^*(A) \leq \mu^*(A_{2n}) + \sum_{m \geq n} \mu^*(B_{2m}) + \sum_{m \geq n} \mu^*(B_{2m+1}).$$

(7.37)

Due to (7.34–7.35), as $n \to \infty$, the last two sums tend to zero and so

$$\mu^*(A) \leq \lim_{n \to \infty} \mu^*(A_n), \quad (7.38)$$

using again the monotonicity of $n \mapsto \mu^*(A_n)$. The claim follows.

Proof of Theorem 7.11. $\Leftarrow$. Assume $\mu^*$ is metric. To get $\mathcal{B}(X) \subset \Sigma(\mu^*)$ it suffices to show that every closed set $K \subset X$ is Carathéodory measurable. Let thus $K \subset X$ be closed and let $E \subset X$. If $E \cap K = \emptyset$ or $E \cap K = \emptyset$, then $\mu^*(E) = \mu^*(E \cap K) + \mu^*(E \setminus K)$ trivially, so let us assume that $A := E \cap K \neq \emptyset$ and $E \cap K \neq \emptyset$. Let $A_n$ be as in the above lemma and assume $n$ is so large that also $A_n \neq \emptyset$. Since $\rho(E \cap K, A_n) > \frac{1}{n}$, the fact that $\mu^*$ is metric yields

$$\mu^*(E) \geq \mu^*(E \setminus K) \cup A_n = \mu^*(E \cap K) + \mu^*(A_n).$$

(7.39)

By the above lemma,

$$\mu^*(A_n) \uparrow \mu^*(A) = \mu^*(E \setminus K). \quad (7.40)$$

Thanks to subadditivity of $\mu^*$, we then have $K \in \Sigma(\mu^*)$. Thus $\mathcal{B}(X) \subset \Sigma(\mu^*)$ as claimed. \qed
Should we desire to have all Borel sets Carathéodory measurable, by Theorem 7.11 we work with a metric outer measure from the outset. Obviously, $\lambda^*$ is a metric outer measure (by Lemma 3.12) and, by Theorem 7.11, all Lebesgue-Stieltjes measures in fact arise from metric outer measures. Notwithstanding, as the next examples shows, this is not automatic for all outer measures; not even the ones obtained by the construction in Proposition 6.7.

Let $\mathcal{C}$ be the set of all closed squares in $\mathbb{R}^2$ and let

$$
\mu_0(C) := \text{side length of } C, \quad C \in \mathcal{C}.
$$

(7.41)

Define $\mu^*$ from $\mu_0$ by the standard procedure in Proposition 6.7 using $\mathcal{C}$ as the covering class. Now pick $\varepsilon \in (0, 1/2)$ small, consider the closed intervals

$$
I_1 := [0, 1/2 - \varepsilon] \quad \text{and} \quad I_2 := [1/2 + \varepsilon, 1]
$$

(7.42)

and define

$$
A := I_1 \times I_1 \cup I_1 \times I_2 \cup I_2 \times I_1 \cup I_2 \times I_2.
$$

(7.43)

Consider a covering of $A$ by countably many squares from $\mathcal{C}$. If $A$ is not covered by a single square, then it has to be covered by at least three squares and (this requires some thinking) the sum of the side-lengths of the covering squares then exceeds 1. Hence, the best way to cover $A$ is by a single square $[0, 1] \times [0, 1]$. It follows that $\mu^*(A) = 1$. On the other hand, by a similar argument, $\mu^*(I_i \times I_j) = 1/2 - \varepsilon$ for all $i, j = 1, 2$. But then, as soon as $4\varepsilon < 1$,

$$
\mu^*(A) = 1 < 4\left(\frac{1}{2} - \varepsilon\right) = \sum_{i,j} \mu^*(I_i \times I_j).
$$

(7.44)

As the four squares $\{I_i \times I_j\}$ are separated from each other by positive distances, $\mu^*$ cannot be a metric outer measure.

It is quite apparent where the obstruction to metric property in the above example comes from. Thanks to the “strange” choice of $\mu_0$ we are not forced to approximate $A$ by “small” sets. (The choice is actually not so strange as we will see in the next section.) As it turns out, the failure of $\mu_0$ to force only “small” sets into the cover of $A$ is in fact the only obstruction for getting a metric outer measure from the construction in Proposition 6.7. Indeed, we have:

**Proposition 7.13** Let $(X, \rho)$ be a metric space. Given $\mathcal{C} \subset 2^X$ and a function $\mu_0 : \mathcal{C} \to [0, \infty]$ with $\mu_0(\emptyset) = 0$, for each $\varepsilon > 0$ and each $A \subset X$ define

$$
\mu^*_\varepsilon(A) := \inf\left\{ \sum_{i \in \mathbb{N}} \mu_0(A_i) : \{A_i\} \subset \mathcal{C}, \bigcup_{i \in \mathbb{N}} A_i \supset A, \text{diam}(A) < \varepsilon \right\},
$$

(7.45)

where

$$
\text{diam}(A) := \sup\{\rho(x,y) : x, y \in A\}
$$

(7.46)

with $\text{diam}(\emptyset) := 0$. Then

$$
\mu^*(A) := \lim_{\varepsilon \downarrow 0} \mu^*_\varepsilon(A)
$$

(7.47)

exists and defines a metric outer measure on $X$.

**Proof.** By our proviso $\text{diam}(\emptyset) := 0$, each $\mathcal{C}_\varepsilon := \{A \in \mathcal{C} : \text{diam}(A) < \varepsilon\}$ is a covering class for each $\varepsilon > 0$ and so, by Proposition 6.7, $\mu^*_\varepsilon$ is an outer measure. Since also $\varepsilon \mapsto \mathcal{C}_\varepsilon$ is decreasing, $\varepsilon \mapsto \mu^*_\varepsilon(A)$ is non-decreasing and so the limit in (7.47) exists (possibly infinite, of course).
To see that $\mu^*$ is an outer measure, observe that both $\mu^e_\varepsilon(\emptyset) = 0$ and $\mu^e_\varepsilon(A) \leq \mu^e_\varepsilon(B)$ for $A \subset B$ survive the limit. The same applies also to countable subadditivity because

$$\sum_{n \in \mathbb{N}} \mu^e_\varepsilon(A_n) \uparrow \sum_{n \in \mathbb{N}} \mu^+(A_n) \quad \text{as } \varepsilon \downarrow 0$$

(7.48)

by the Monotone Convergence Theorem for the counting measure on $\mathbb{N}$ and a non-decreasing family of functions $f_\varepsilon : \mathbb{N} \to [0, \infty]$ given by $f_\varepsilon(n) := \mu^e_\varepsilon(A_n)$. (A direct proof is also possible.)

It remains to show that $\mu^*$ is metric. For this let $A, B \subset X$ obey $\rho(A, B) > 0$ and let $\varepsilon$ be such that $0 < \varepsilon < \frac{1}{2}\rho(A, B)$. Assume that $\mu^*(A \cup B) < \infty$ because otherwise there is nothing to prove (by subadditivity of $\mu^*$). Then there is $\{C_i\} \subset \mathcal{C}_\varepsilon$ such that

$$A \cup B \subset \bigcup_{i \in \mathbb{N}} C_i \quad \text{and} \quad \mu^e_\varepsilon(A \cup B) + \varepsilon \geq \sum_{i \in \mathbb{N}} \mu_0(C_i).$$

(7.49)

Let

$$S := \{i \in \mathbb{N} : C_i \cap A \neq \emptyset\}.$$  

(7.50)

By the restriction on the diameter of the $C_i$’s and the choice of $\varepsilon$ relative to $\rho(A, B)$, if $C_i$ intersects $A$ then it cannot intersect $B$. Therefore,

$$B \cap \bigcup_{i \in S} C_i = \emptyset \quad \text{and so} \quad A \subset \bigcup_{i \in S} C_i \quad \text{and} \quad B \subset \bigcup_{i \in \mathbb{N} \setminus S} C_i.$$

(7.51)

Invoking the additivity of infinite sums of non-negative terms, we then get

$$\mu^e_\varepsilon(A \cup B) + \varepsilon \geq \sum_{i \in S} \mu_0(C_i) + \sum_{i \in \mathbb{N} \setminus S} \mu_0(C_i) \geq \mu^*\varepsilon(A) + \mu^*\varepsilon(B).$$

(7.52)

Taking limit $\varepsilon \downarrow 0$ we find out that $\mu^*$ is superadditive on $A \cup B$; subadditivity of $\mu^*$ proved earlier then shows that $\mu^*$ is in fact additive on $A \cup B$. Hence $\mu^*$ is metric as claimed. \hfill \Box

### 7.4 Hausdorff measure and dimension.

The above construction of metric outer measure permits us to introduce the following object:

**Definition 7.14** (Hausdorff measure) Let $(X, \rho)$ be a metric space and let $\text{diam}(A)$ be as defined above. For any $s \geq 0$,

$$\mathcal{H}^s(A) := \sup_{\varepsilon > 0} \inf \left\{ \sum_{i \in \mathbb{N}} (\text{diam}(A_i))^s \right\} : \{A_i\} \subset 2^X, \bigcup_{i \in \mathbb{N}} A_i \supset A, \text{diam}(A_i) < \varepsilon \}$$

(7.53)

defines the $s$-dimensional Hausdorff (outer) measure of $A \subset X$.

By Proposition 7.13, $\mathcal{H}^s$ is a metric outer measure and thus all Borel sets are $\mathcal{H}^s$-measurable. Variations on the above definition are often considered that differ by the specific choice of the covering family. Some of the alternative choices lead to the same object; for instance, since we always have

$$\text{diam}(\overline{B}) = \text{diam}(B)$$

(7.54)

taking only closed sets in the covering family leads to the same $\mathcal{H}^s(A)$. Also, for any $\varepsilon > 0$, every set $B$ of finite diameter is contained in an open set $O$ such that

$$\text{diam}(O) - \varepsilon \leq \text{diam}(B) \leq \text{diam}(O)$$

(7.55)
and so, at least for \( s > 0 \), choosing open sets for the covering family also gives the same \( \mathcal{H}^s(A) \).

For \( s = 0 \) this depends on the interpretation of the number \( 0^0 \). In this case, we get:

**Lemma 7.15** If \( 0^0 := 0 \) while \( r^0 = 1 \) for \( r > 0 \), then \( \mathcal{H}^0(A) \) is the counting measure of \( A \).

**Proof.** If \( A \) is a set of \( n \) points, then there is \( \varepsilon > 0 \) such that any countable cover of \( A \) by sets of diameter less than \( \varepsilon \) will have at least \( n \) non-empty sets. As there is a covering by \( n \) non-empty sets of diameter less than \( \varepsilon \), the interpretation of \( 0^0 \) implies that \( \mathcal{H}^0(A) = n \). Employing the monotonicity of the outer measure, this proves the claim.

Note that this will be different if we require — as is sometimes done (e.g., in Federer’s book cited below) — that all \( A_i \)'s in the covering of \( A \) in (7.53) be non-empty. In this case \( \mathcal{H}^0(A) = \infty \) for all \( A \) (for \( A = \emptyset \) a separate proviso has to be made anyway). As the reader easily checks, once \( s > 0 \), the fact whether or not we require all sets in the cover to be non-empty plays no role.

In normed vector spaces, i.e., when the metric (and thus diameter) arises from a norm, every bounded set can be enclosed in a closed convex set — namely, its closed convex hull — of equal diameter. So, although restricting the infimum to just closed convex sets generally leads to a larger measure (called the \( d \)-dimensional Carathéodory measure when \( X = \mathbb{R}^d \), in normed vector spaces the two measures coincide. We refer to the book by Federer (Geometric Measure Theory, Springer-Verlag 1969) for further discussion of these subtleties.

We have seen that the \( s := 0 \) Hausdorff measure coincides with the counting measure. For the \( s := d \) case, a similar connection can be made:

**Lemma 7.16** Suppose \( \mathbb{R}^d \) is endowed with the \( \ell_\infty \)-metric, \( p(x,y) := \max_{i=1,...,d} |x_i - y_i| \). Then \( \mathcal{H}^d(A) = \lambda^\star(A) \) holds for every bounded set \( A \subset \mathbb{R}^d \).

**Proof.** Let \( \rho \) be as above and let \( A \subset \mathbb{R}^d \) be a bounded set. Let \( \mathcal{H}^d(A) \) denote the associated Hausdorff (outer) measure. Clearly, \( \mathcal{H}^d(A) < \infty \), so given \( \varepsilon > 0 \) we can find bounded sets \( \{A_i\} \) such that

\[
A \subset \bigcup_{i \in \mathbb{N}} A_i \quad \text{and} \quad \sum_{i \in \mathbb{N}} \text{diam}(A_i)^d < \mathcal{H}^d(A) + \varepsilon. \tag{7.56}
\]

Now each \( A_i \) is contained in a unique minimal rectangular square box (with sides perpendicular to the principal directions in \( \mathbb{R}^d \)) of the same \( \rho \)-diameter, and so we may assume that \( A_i \) are such boxes. Increasing the size slightly, we may even assume that each \( A_i \) is half-open. But \( \text{diam}(A_i)^d = \lambda(A_i) \), and so

\[
\mathcal{H}^d(A) + \varepsilon > \sum_{i \in \mathbb{N}} \text{diam}(A_i)^d = \sum_{i \in \mathbb{N}} \lambda(A_i) \geq \lambda^\star(A). \tag{7.57}
\]

Letting \( \varepsilon \downarrow 0 \), we thus get

\[
\mathcal{H}^d(A) \geq \lambda^\star(A). \tag{7.58}
\]

For the opposite inequality, consider a covering of \( A \) by half-open rectangular boxes \( \{B_i\} \) such that \( \sum_{i \in \mathbb{N}} \lambda(B_i) < \lambda^\star(A) + \varepsilon \). As is now easy to check, each \( B_i \) can be further partitioned into disjoint half-open square boxes \( \{B_{ij} : j \in \mathbb{N}\} \) (some possibly empty) so that

\[
\lambda(B_i) = \sum_{j \in \mathbb{N}} \lambda(B_{ij}), \quad i \in \mathbb{N}. \tag{7.59}
\]
But \( \{ B_{i,j} : i, j \in \mathbb{N} \} \) is a covering of \( A \) and so
\[
\mathcal{H}^d(A) \leq \sum_{i,j \in \mathbb{N}} \text{diam}(B_{i,j})^d = \sum_{i,j \in \mathbb{N}} \lambda(B_{i,j}) = \sum_{i \in \mathbb{N}} \lambda(B_i) < \lambda^*(A) + \varepsilon.
\] (7.60)
Hence
\[
\mathcal{H}^d(A) \leq \lambda^*(A)
\] (7.61)
and so the claim follows for all bounded \( A \subset \mathbb{R}^d \). \( \square \)

For other norm-induces metrics on \( \mathbb{R}^d \) we at least get the following comparison:

**Corollary 7.17**  Let \( \rho \) be any norm-metric on \( \mathbb{R}^d \). Then there are \( c_1, c_2 \in (0, \infty) \) such that
\[
c_1 \lambda^*(A) \leq \mathcal{H}^d(A) \leq c_2 \lambda^*(A)
\] (7.62)
holds for each bounded \( A \subset \mathbb{R}^d \).

**Proof.** This is a direct consequence of Lemma 7.16 and the fact that any two metrics on \( \mathbb{R}^d \) are comparable. \( \square \)

As it turns out, the Hausdorff measure \( \mathcal{H}^s(A) \) is for all but one value of \( s \) either zero or infinity. Indeed, we have:

**Lemma 7.18**  Let \( s' > s \geq 0 \). Then
\[
\begin{align*}
(1) & \quad \mathcal{H}^s(A) < \infty \text{ implies } \mathcal{H}^{s'}(A) = 0 \text{ and, equivalently,} \\
(2) & \quad \mathcal{H}^s(A) > 0 \text{ implies } \mathcal{H}^{s'}(A) = \infty.
\end{align*}
\]

**Proof.** If \( \text{diam}(A_i) < \varepsilon \) then \( \text{diam}(A_i)^{s'} \leq \varepsilon^{s'-s} \text{diam}(A_i)^s \). So \( \mathcal{H}^{s'}(A) \leq \varepsilon^{s'-s} \mathcal{H}^s(A) \) for every \( \varepsilon > 0 \). The two (equivalent) conclusions are now easily checked. \( \square \)

This naturally leads to:

**Definition 7.19**  (Hausdorff dimension) Let \( A \subset X \), where \((X, \rho)\) is a metric space. Then
\[
\text{dim}_H(A) := \inf \{ s \geq 0 : \mathcal{H}^s(A) = 0 \}
\] (7.63)
defines the Hausdorff dimension of \( A \).

Here are some simple consequences of this definition:

**Lemma 7.20**  We have:
\[
\begin{align*}
(1) & \quad \text{if } A \subset B \text{ then } \text{dim}_H(A) \leq \text{dim}_H(B), \\
(2) & \quad \text{if } A \text{ is countable, then } \text{dim}_H(A) = 0,
\end{align*}
\]
In \( \mathbb{R}^d \) endowed with any norm-metric,
\[
\begin{align*}
(3) & \quad \text{if } A \subset \mathbb{R}^d \text{ then } \text{dim}_H(A) \leq d, \\
(4) & \quad \text{if } A \subset \mathbb{R}^d \text{ obeys } \lambda^*(A) > 0, \text{ then } \text{dim}_H(A) = d.
\end{align*}
\]

**Proof.** (1) is immediate from the fact that \( A \subset B \) implies \( \mathcal{H}^s(A) \leq \mathcal{H}^s(B) \). For (2) assume \( A \) is countable and write it as \( A = \{ x_i : i \in \mathbb{N} \} \). Now take \( A_i := \{ y \in X : \rho(x_i, y) \leq \varepsilon 2^{-i} \} \). Then \( A \subset \bigcup_{i \in \mathbb{N}} A_i \) and so
\[
\mathcal{H}^d(A) \leq \sum_{i \in \mathbb{N}} \text{diam}(A_i)^s = \sum_{i \in \mathbb{N}} \varepsilon i 2^{-is} < \infty, \quad s > 0.
\] (7.64)
Taking $\varepsilon \downarrow 0$ (or invoking Lemma 7.18) we get $\mathcal{H}^s(A) = 0$ for all $s > 0$ and so $\dim_{\mathcal{H}}(A) = 0$.

To get (3), let $s > d$. Consider a partition of $[0, 1] = \{I_n: \ v \in \mathbb{N}\}$. Now partition $I_n$ into $2^k$ cubes $\{I_{ni}: i = 1, \ldots, 2^k\}$ of side $2^{-n}$. Let $\rho$ be any norm-metric. Since $\rho$ is comparable with $\ell^\infty$-distance, we have $\text{diam}(I_{ni}) \leq c2^{-n}$. Hence

$$\mathcal{H}^s(A) \leq \mathcal{H}^s([0, 1]) \leq \sum_{n \in \mathbb{N}} \sum_{i=1}^{2^k} \text{diam}(I_{ni})^s \leq c^s \sum_{n \in \mathbb{N}} \sum_{i=1}^{2^k} 2^{-ns} = c^s \sum_{n \in \mathbb{N}} 2^{-n(s-d)} < \infty. \quad (7.65)$$

It follows that $\mathcal{H}^s(A) = 0$ as soon as $s > d$ and thus $\dim_{\mathcal{H}}(A) \leq d$.

To get (4) note that, if $A \subset \mathbb{R}^d$ obeys $\lambda^*(A) > 0$, then (by countable subadditivity of $\lambda^*$) there is a bounded set $A' \subset A$ with $\lambda^*(A') > 0$. Lemma 7.16 then ensures $\mathcal{H}^d(A) \geq \mathcal{H}^d(A') > 0$ and so $\dim_{\mathcal{H}}(A) \geq d$. \hfill \Box

Lemma 7.16 and Corollary 7.17 show that the $d$-dimensional Hausdorff measure is closely related to the Lebesgue measure. However, the lower-dimensional Hausdorff measures have good interpretations as well. We will demonstrate this for the case of 1-dimensional measure. Let $\gamma: [0, 1] \to X$ be a continuous map — we think of this as a curve in the metric space $(X, \rho)$. The arc-length of $\gamma$ is then defined by

$$\text{length}_\rho(\gamma) := \sup_{\eta=0<h_1<\cdots<h_n=1} \sum_{i=1}^{n} \rho(\gamma(t_i), \gamma(t_{i-1})). \quad (7.66)$$

(Obviously, the sum on the right is the length of “piece-wise linear” approximation to the curve. Thanks to the triangle inequality, refinements of the partition increase the sum on the right, so taking the supremum is intuitively a correct thing to do.)

**Lemma 7.21** Let $\gamma: [0, 1] \to X$ be continuous and non-constant. Then

$$\text{length}_\rho(\gamma) \geq \mathcal{H}^1([\gamma([0, 1])] > 0 \quad (7.67)$$

If $\gamma$ is also one-to-one, then

$$\text{length}_\rho(\gamma) = \mathcal{H}^1([\gamma([0, 1])]. \quad (7.68)$$

**Proof.** Left to reader. \hfill \Box

The previous lemma shows that a rectifiable (non-constant, continuous) curve will have positive and finite 1-dimensional Hausdorff measure. The same applies, in particular, to the graph of a function of finite variation. It turns out that weakening the regularity to Hölder continuity yields curves of (possibly) non-trivial Hausdorff dimension:

**Lemma 7.22** Let $f: [0, 1] \to \mathbb{R}$ be a function and consider its graph

$$G_f := \{ (t, f(t)) : t \in [0, 1] \}. \quad (7.69)$$

Let $\alpha \geq 0$. If $f$ is $\alpha$-Hölder in the sense that $|f(t) - f(s)| \leq C|t-s|^\alpha$ for some $C < \infty$, then (with respect to any norm-metric on $\mathbb{R}^2$),

$$\dim_{\mathcal{H}}(G_f) \leq 2 - \alpha. \quad (7.70)$$

**Proof.** Fix $s > 2 - \alpha$, pick $N \in \mathbb{N}$ and consider the points $t_i := i/N$, $i = 0, \ldots, N$. The portion of $G_f$ between $t_i$ and $t_{i+1}$ fits into a rectangle of base $1/N$ and height $CN^{-\alpha}$ and so it can be covered by
C^1N^{1-\alpha} squares of side length 1/N. It follows that
\[ \mathcal{H}^s(G_f) \leq N(C^1N^{1-\alpha})N^{-s} = C^1N^{2-\alpha-s} \] (7.71)
Taking \( N \to \infty \) we get \( \mathcal{H}^s(G_f) = 0 \), i.e., \( \text{dim}_H(G_f) \leq s \) for any \( s > 2 - \alpha \). The claim follows. \( \square \)

The previous lemma admits the following partial converse:

Lemma 7.23 Let \( f : [0,1] \to \mathbb{R} \) be a function such that, for some \( \alpha \geq 0 \) and \( C > 0 \), the following holds: For each \([s,t] \subset [0,1]\) there is \([s',t'] \subset [s,t]\) such that
\[ |f(t') - f(s')| \geq C|t-s|^\alpha. \] (7.72)
Then
\[ \text{dim}_H(G_f) \geq 2 - \alpha. \] (7.73)
Proof. Left to reader. \( \square \)

The above two lemmas are quite useful when one studies the regularity of graphs of random functions such as the Brownian motion; see Fig. 2. The monograph of Môters and Peres gives a detailed account of this application (the above lemmas where taken from there).

Naturally, if we would consider maps \( \gamma : [0,1]^d \to \mathbb{R}^d \), we should obtain images that have Hausdorff dimension \( d' \) (or larger). For instance, the Hausdorff dimension of a unit sphere in \( \mathbb{R}^d \) or the boundary of a unit cube in \( \mathbb{R}^d \) is \( d-1 \), etc. In fact, we have:

Lemma 7.24 Let \( S^{d-1} := \{ x \in \mathbb{R}^d : |x| = 1 \} \) be the unit sphere in \( \mathbb{R}^d \), where \( |x| \) denotes the Euclidean norm of \( x \). Then \( \mathcal{H}^{d-1} \), defined using the Euclidean norm, is a uniform measure on \( S^{d-1} \) of total mass equal to the (ordinary) area of \( S^{d-1} \).
We will possibly revisit these questions later, should it be decided to delve deeper into geometric measure theory.

7.5 Frostman’s lemma.

The computation of Hausdorff dimension is generally an arduous task. Usually an upper bound is considerably easier; indeed, it suffices to produce a covering \( \{A_i\} \) of \( A \) with finite \( \sum_{i \in \mathbb{N}} \text{diam}(A_i)^s \).

The lower bound is typically harder. Here one is often aided by the following fact, proved originally by Frostman in his doctoral thesis:

**Theorem 7.25** (Frostman’s lemma) Let \( A \subset \mathbb{R}^d \) be a closed set and let \( s \geq 0 \). Then \( \mathcal{H}^s(A) > 0 \) holds if and only if there is a Borel probability measure \( \mu \) with \( \mu(A^c) = 0 \) such that, for some \( c \in (0, \infty) \) and all closed sets \( B \subset \mathbb{R}^d \),

\[
\mu(B) \leq c \text{diam}(B)^s.
\] (7.74)

If such measure exists then \( \mathcal{H}^s(A) \geq 1/c \).

One direction of proof is quite easy:

**Proof of sufficiency.** Suppose there is a Borel measure \( \mu \) with \( \mu(A^c) = 0 \) and \( \mu(A) = 1 \) such that (7.74) holds for all closed sets \( B \). Let \( \{A_i\} \) be a cover of \( A \) by closed sets. Then

\[
\sum_{i \in \mathbb{N}} \text{diam}(A_i)^s \geq \frac{1}{c} \sum_{i \in \mathbb{N}} \mu(A_i) \geq \frac{1}{c} \mu(A) = \frac{1}{c}.
\] (7.75)

Hence \( \mathcal{H}^s(A) \geq 1/c > 0 \) as desired. \( \square \)

Let us note a simple consequence of the previous argument:

**Lemma 7.26** Let \( C \subset [0, 1] \) be the Cantor middle-third set. Then \( \dim_H(C) = \frac{\log 2}{\log 3} \).

**Proof.** Let \( s := \frac{\log 2}{\log 3} \). We will show that \( \frac{1}{2} \leq \mathcal{H}^s(C) \leq 1 \) which implies the claim. The upper bound is easy: \( C \) is a subset of \( C_n \), where \( C_n \) is a union of \( 2^n \) intervals of length \( 3^{-n} \). Hence,

\[
\mathcal{H}^s(C) \leq \mathcal{H}^s(C_n) \leq 2^n 3^{-ns} = 1.
\] (7.76)

For the lower bound, consider the Cantor-Lebesgue function introduced in Section 6.4 (see, in particular, Fig 1 for a graph) and let \( \mu \) be the corresponding Lebesgue-Stieltjes measure. As is easy to check, \( \mu \) assigns equals mass (namely, \( 2^{-n} \)) to each interval constituting \( C_n \). If \( B \subset [0, 1] \) is a closed interval, we can thus find \( n \in \mathbb{N} \) so that

\[
3^{-n} \leq \text{diam}(B) \leq 3^{-n+1}.
\] (7.77)

Thanks to the structure of \( C_n \), two “adjacent” intervals are distance \( 3^{-n} \) from each other and any other interval is at least three times further away. Hence, any set of diameter less than \( 3^{-n+1} \) intersects at most two of the intervals constituting \( C_n \) and so

\[
\mu(B) \leq 2^{-n+1} = 2(3^{-n})^s \leq 2 \text{diam}(B)^s.
\] (7.78)

By the sufficiency part of the Frostman lemma, \( \mathcal{H}^s(C) \geq \frac{1}{2} \). \( \square \)

We leave without a proof the fact that the Hausdorff measure on \( C \) of dimension of \( C \) actually equals the uniform measure \( \mu \) constructed in the previous proof. The fact that \( \dim_H(C) = \frac{\log 2}{\log 3} \)
can also be “explained” by self-similarity of \( C \): In order to get two copies of \( C \) one just blows \( C \) up by factor 3. The dimension is exactly the number so that \( 3^{\dim_H(C)} = 2 \). Similar reasoning applies to other (regular) fractal objects, such as Sierpiński carpet (see Fig. 3).

It should be noted that the sufficiency direction in Frostman’s lemma does not require the underlying space to be \( \mathbb{R}^d \). A practical verification of this condition aided by the fact that (7.74) is implied by an integrability criterion. (Strictly speaking, this criterion requires the knowledge of product measure but that will be mended in the next section.)

**Lemma 7.27** (Frostman’s energy condition) Let \((X, \rho)\) be a metric space \( \mu \) is a Borel measure with \( \mu(A) \in (0, \infty) \) and \( \mu(A^c) = 0 \) such that
\[
\int_{A \times A} \frac{1}{\rho(x,y)^d} \mu(dx)\mu(dy) < \infty.
\]

(7.79) Then \( H^\nu(A) > 0 \).

We will prove this lemma once we have Fubini-Tonelli Theorem available. The reason why (7.79) is referred to as **energy condition** is that the integral shows up in harmonic analysis (and electrostatic) as electrostatic energy of charge distribution represented by \( \mu \).