Measure and Integration (under construction)

Gustav Holzegel*

April 24, 2017

Abstract

These notes accompany the lecture course "Measure and Integration" at Imperial College London (Autumn 2016). They follow very closely the text "Real-Analysis" by Stein-Shakarchi, in fact most proofs are simple rephrasings of the proofs presented in the aforementioned book.

Contents

1 Motivation **3**
1.1 Quick review of the Riemann integral . . . 3
1.2 Drawbacks of the class $\mathfrak{R}$, motivation of the Lebesgue theory . . 4
1.2.1 Limits of functions . . . 5
1.2.2 Length of curves . . . 5
1.2.3 The Fundamental Theorem of Calculus . . . 5
1.3 Measures of sets in $\mathbb{R}$ . . . 6
1.4 Literature and Further Reading . . . 6

2 Measure Theory: Lebesgue Measure in $\mathbb{R}^d$ **7**
2.1 Preliminaries and Notation . . . 7
2.2 Volume of Rectangles and Cubes . . . 7
2.3 The exterior measure . . . 9
2.3.1 Examples . . . 9
2.3.2 Properties of the exterior measure . . . 10
2.4 The class of (Lebesgue) measurable sets . . . 12
2.4.1 The property of countable additivity . . . 14
2.4.2 Regularity properties of the Lebesgue measure . . . 15
2.4.3 Invariance properties of the Lebesgue measure . . . 16
2.4.4 σ-algebras and Borel sets . . . 16
2.5 Construction of a non-measurable set . . . 17
2.6 Measurable Functions . . . 19
2.6.1 Some abstract preliminary remarks . . . 19
2.6.2 Definitions and equivalent formulations of measurability . 19
2.6.3 Properties 1: Behaviour under compositions . . . 21
2.6.4 Properties 2: Behaviour under limits . . . 21
2.6.5 Properties 3: Behaviour of sums and products . . . 22

*Imperial College London, Department of Mathematics, South Kensington Campus, London SW7 2AZ, United Kingdom.

2.6.6 The notion of “almost everywhere” . . . 22
2.7 Building blocks of integration theory . . . 23
2.7.1 Simple functions . . . 23
2.7.2 Step functions . . . 23
2.8 Approximation Theorems . . . 23
2.8.1 Approximating a measurable function by simple functions 23
2.8.2 Approximating a mesurable function by step functions . . 25
2.9 Littlewoods Three Principles . . . 26

3 Integration Theory: The Lebesgue Integral 28
3.1 Simple Functions . . . 28
3.2 Bounded functions on sets of finite measure . . . 29
3.2.1 Riemann integrable functions are Lebesgue integrable . . 32
3.3 Non-negative functions . . . 33
3.4 The General case and the notion of integrable . . . 35
3.5 Aside: Complex-valued functions . . . 37
3.6 The space of integrable functions as a normed vector space . . . 37
3.7 Dense families in $L^1\left(\mathbb{R}^d\right)$ . . . 39
3.8 Fubini's Theorem . . . 39
3.8.1 Slices of measurable sets and functions . . . 40
3.8.2 Statement and Discussion of Fubini's and Tonelli's Theorem 40
3.8.3 The proof of Tonelli's Theorem (using Fubini) . . . 42
3.8.4 The proof of Fubini's Theorem . . . 42

4 Differentiation and Integration 46
4.1 Differentiation of the Integral . . . 46
4.1.1 Proof of the Lebesgue Differentiation Theorem . . . 47
4.1.2 Proof of Proposition 4.1 . . . 49
4.1.3 Final Remarks . . . 50
4.2 Differentiation of functions . . . 50
4.2.1 Functions of bounded variation . . . 50
4.2.2 Examples of functions of bounded variation . . . 51
4.2.3 Characterisation of functions of bounded variation . . . 52
4.2.4 Bounded variation implies differentiable a.e. . . . 54
4.2.5 Absolute Continuity and the Fundamental Theorem of the Lebesgue integral . . . 55
4.2.6 The proof of Theorem 4.8 . . . 56
4.2.7 The proof of Theorem 4.6 (non examinable) . . . 59

5 Abstract Measure Theory 60
5.1 Measure Spaces: Definition and basic examples . . . 60
5.2 Exterior measure and Carathéodory's theorem . . . 62
5.3 Premeasures and the extension theorem . . . 64
5.3.1 Construction of a measure from a premeasure . . . 65
5.3.2 The proof of Proposition 5.1 . . . 66
5.4 A further example: Hausdorff measure . . . 67
5.5 Integration on a general measure space . . . 68
5.6 Construction of product measures . . . 69
5.7 General Fubini theorem . . . 70

6 The change of variables formula **72**
6.1 An example illustrating Theorem 6.1 72
6.2 A reformulation of Theorem 6.1 73
6.3 Proof of Theorem 6.2 73

7 Mastery Material: L^p-spaces **76**
7.1 Definition 76
7.2 The Hölder inequality 76
7.3 Minkowski's inequality 77

1 Motivation

1.1 Quick review of the Riemann integral

In your second year analysis course you defined the Riemann integral. Let us remind ourselves of the simplest situation and consider a bounded real-valued function f defined on the interval $[a, b]$.

- A *partition* P of $[a, b]$ is a finite set of points $x_0, x_1, ..., x_n$ with

$$a = x_0 \leq x_1 \leq x_2 \leq ... \leq x_{n-1} \leq x_n = b\,.$$

- Given a partition P, we define

$$M_i = \sup_{x_{i-1} \leq x \leq x_i} f(x) \qquad \text{and} \qquad m_i = \inf_{x_{i-1} \leq x \leq x_i} f(x)$$

and

$$U(P, f) = \sum_{i=1}^{n} M_i (x_i - x_{i-1}) \quad \text{the upper sum}$$

$$L(P, f) = \sum_{i=1}^{n} m_i (x_i - x_{i-1}) \quad \text{the lower sum}\,.$$

- Finally, we define the *upper* and *lower Riemann integrals* of f on $[a, b]$ as

$$\overline{\int_a^b} f dx := \inf_P U(P, f) \qquad \text{and} \qquad \underline{\int_a^b} f dx := \sup_P L(P, f) \tag{1}$$

Note that since f is bounded these objects are well-defined.

Definition 1.1. *We say that f is Riemann-integrable over $[a, b]$ if the upper and the lower Riemann integrals agree. In this case we write $f \in \mathfrak{R}$ and denote the common value by $\int_a^b f dx$.*

The theory then proceeds along the following lines

Definition 1.2. *We say that the partition $P^\star$ is a refinement of P if $P^\star \supset P$. Given two partitions P_1 and P_2 we define their common refinement as $P^\star = P_1 \cup P_2$.*

One easily shows

Theorem 1.1. *If $P^\star$ is a refinement of P then*

$$L(P,f) \leq L(P^\star,f) \quad \textit{and} \quad U(P,f) \geq U(P^\star,f) \, .$$

Theorem 1.2. *We have*

$$\underline{\int_a^b} f dx \leq \overline{\int_a^b} f dx$$

Proof. Let $P^\star$ be the common refinement of two arbitrary partitions P_1 and P_2. We have

$$L(P_1,f) \leq L(P^\star,f) \leq U(P^\star,f) \leq U(P_2,f)$$

by the previous theorem. Fixing P_2 and taking the sup over all partitions P_1 we obtain

$$\underline{\int_a^b} f dx \leq U(P_2,f)$$

for any partition P_2. Taking now the infimum over all partitions P_2, we are done. □

Theorem 1.3. *The function f is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition P such that*

$$U(P,f) - L(P,f) < \epsilon \, .$$

Proof. The function f being Riemann integrable means that, for $\epsilon > 0$ prescribed, there exist partitions P_1, P_2 (with common refinement $P^\star$) such that

$$U(P^\star,f) - \int_a^b f dx \leq U(P_1,f) - \int_a^b f dx < \frac{\epsilon}{2} \, ,$$

$$\int_a^b f dx - L(P^\star,f) \leq \int_a^b f dx - L(P_1,f) < \frac{\epsilon}{2} \, .$$

Adding the two inequalities proves the first direction. Conversely, to check whether f is Riemann integrable, it suffices to show $\inf_P U(P,f) - \sup_P L(P,f) < \epsilon$ for any $\epsilon > 0$. With $\epsilon > 0$ prescribed we take the partition $\hat{P}$ promised by the assumption to obtain

$$\inf_P U(P,f) - \sup_P L(P,f) \leq U\left(\hat{P},f\right) - L\left(\hat{P},f\right) < \epsilon \, .$$

□

One can then use the criterion of Theorem 1.3 to prove

Theorem 1.4. *If f is continuous on $[a,b]$ then it is Riemann integrable.*

Proof. Exercise. Hint: Use that a continuous function on a compact interval is uniformly continuous. □

1.2 Drawbacks of the class $\mathfrak{R}$, motivation of the Lebesgue theory

In this course you will learn about a new integral, the Lebesgue integral, which will allow us to integrate a much larger class of functions. Why would one want to do that?

1.2.1 Limits of functions

Well, one main drawbacks of the class of Riemann integrable functions $\mathfrak{R}$ is that it does not behave well under taking limits.

To see this, consider the function

$$f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0,1], \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

This function is not Riemann integrable (why?). On the other hand, for (x_n) an enumeration of the rational numbers in $[0,1]$, the function

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{x_1, x_2, ..., x_n\}, \\ 0 & \text{otherwise.} \end{cases}$$

is Riemann integrable for every n (why? with what value?) and we have $f_n \to f$ pointwise. We conclude that the limit of a sequence of Riemann integrable functions does not have to be Riemann integrable.

A perhaps less academic example is provided by the following sequence of functions that you will construct on the first example sheet: (f_n) is a sequence of *continuous* functions $f_n : [0,1] \to \mathbb{R}$ with $f_n \to f$ pointwise such that

- $0 \leq f_n \leq 1$
- (f_n) decreases monotonically as $n \to \infty$
- f is not Riemann integrable

The above implies that $s_n = \int_0^1 f_n dx$ is a decreasing sequence of positive numbers and hence converges. It is very tempting to define $\int_0^1 f dx$ to be that limit. The Lebesgue integral will allow us to do this in this particular situation ("monotone convergence theorem") and in much more general ones.

1.2.2 Length of curves

Let $\Gamma(t) = (x(t), y(t))$ for $a \leq t \leq b$ be a curve in the plane with x and y continuous functions. We can define the length L of the curve as the supremum of the length of polygonal approximations (picture). We call Γ a rectifiable curve if $L < \infty$. You may recall that if x and y are continuously differentiable, then the above limiting procedure leads to the formula

$$L = \int_a^b \sqrt{\dot{x}^2(t) + \dot{y}^2(t)} dt .$$

A natural question is whether weaker conditions on x and y suffice to guarantee rectifiability of Γ and whether we can make sense of the formula above in this case. This will lead us to *functions of bounded variation.*

1.2.3 The Fundamental Theorem of Calculus

For F a differentiable function whose derivative is Riemann integrable on $[a,b]$ we have the FTC

$$\int_a^b F'(x) dx = F(b) - F(a) . \tag{3}$$

Now there are functions whose derivative is not Riemann integrable. Can we still make sense of the formula above? What is the class of functions for which an identity as above holds?

1.3 Measures of sets in $\mathbb{R}$

As we shall see, the key to answering the above questions lies in understanding the size or "measure" of sets in $\mathbb{R}^d$. In particular, discussing the case $d = 1$ for simplicity, we will construct a function ("measure") defined on a certain class of subsets of $\mathbb{R}$ (denoted by ℓ)

$$m : \ell \to \mathbb{R}_0^+ \cup \{\infty\}$$

with the following properties

1. $m(E) = b - a$ if E is the interval $[a, b]$, $a \le b$.
2. $m(E) = \sum_{n=1}^{\infty} m(E_n)$ whenever $E = \cup_{n=1}^{\infty} E_n$ is a disjoint union (countable additivity).
   In particular, $m(E_1 \cup E_2) = m(E_1) + m(E_2)$ if E_1, E_2 are disjoint. (finite additivity)
3. $m(E + h) = m(E)$ for every $h \in \mathbb{R}$.
   (translation invariance)

We will show existence and uniqueness of such a measure, *the Lebesgue measure*, provided one restricts ℓ to the class of "measurable sets". This class will be closed under countable unions and intersections as well as taking complements and contain moreover the open sets. But it won't comprise *all* subsets of $\mathbb{R}$! There are non-measurable sets. In other words there is no function m with the above properties that is defined on all sets of $\mathbb{R}$.[1]

Once we have identified the class of measurable sets, we shall define measurable functions and then construct the Lebesgue integral. Armed with this, we shall be able to answer some of the questions raised in Section 1.2.

This will probably take three quarters of the course. In the last part we will discuss what is called "abstract measure theory", which will be important, for instance, in applications to probability.

1.4 Literature and Further Reading

As mentioned, Sections 2-5 of these notes will follow very closely the book "Real Analysis" by Stein-Shakarchi (Princeton University Press) covering Chapters 1-3 and the first half of Chapter 6 of that book. For Section 5.6 I also used "Measure Theory and Integration" by M. Taylor (Graduate Studies in Mathematics). The proof of the change of variables formula in Section 6 is taken from "Analysis II" by Theodor Bröcker (Spektrum Lehrbuch; in German). I also recommend the "classics": Folland's "Real Analysis" (Wiley) and Rudin's "Real and Complex Analysis" (Mc Graw Hill). These books take a more advanced point of view than these notes but will be nice and accessible as complementary reading.

[1] The construction of such non-measurable sets involves the axiom of choice. We will construct such a set later. In higher dimensions, one has the Banach-Tarski paradox, which proves (using again the axiom of choice) that the unit ball in $\mathbb{R}^3$ can be decomposed into a finite number of disjoint sets A_i ($i \geq 5$) which can reassembled – using only translations and rotations applied to the A_i – into two copies of the unit ball.

2 Measure Theory: Lebesgue Measure in $\mathbb{R}^d$

2.1 Preliminaries and Notation

- $x = (x_1, .., x_d)$ a point in $\mathbb{R}^d$
- $|x| = \sqrt{(x_1)^2 + ...(x_d)^2}$ the Euclidean norm
- $d(x, y) = |x - y|$ the distance of two points $x, y \in \mathbb{R}^d$
- $d(E, F) = \inf_{x \in E; y \in F} |x - y|$ distance between two sets $E, F \subset \mathbb{R}^d$.
- $B_r(x) := \{y \in \mathbb{R}^d \mid |x - y| < r\}$ open ball around x of radius r in $\mathbb{R}^d$.
- recall definitions of open, closed, bounded
- $E \subset \mathbb{R}^d$ is compact if for any open cover $E \subset \bigcup_{\alpha \in A} U_\alpha$ (U_α open) there exists a finite subcover, i.e. $E \subset \bigcup_{j \in J} U_j$ for J a finite subset of A.
- recall in $\mathbb{R}^d$ a subset $E \in \mathbb{R}^d$ is compact if and only if it is closed and bounded (Heine-Borel)
- $x \in \mathbb{R}^d$ is a *limit point* of $E \subset \mathbb{R}^d$ if for every $r > 0$ the ball $B_r(x)$ contains points of E.
- $x \in \mathbb{R}^d$ is an *isolated point* of $E \subset \mathbb{R}^d$ if for some $r > 0$ the ball $B_r(x)$ satisfies $B_r(x) \cap E = \{x\}$.
- A closed set is *perfect* if it does not contain any isolated points.
- The *closure* of $E \subset \mathbb{R}^d$, denoted $\overline{E}$ is the union of E and all its limit points.
- The *boundary* of $E \subset \mathbb{R}^d$, denoted ∂E is the closure of E without the interior of E, i.e. $\partial E = \overline{E} \setminus int(E)$.

2.2 Volume of Rectangles and Cubes

Our first basic tool to measure the size of sets will be *closed rectangles*, defined as follows

$$R = [a_1, b_1] \times [a_2, b_2] \times ... \times [a_d, b_d]$$

with $a_i \leq b_i$ for all $i = 1, ..., d$. Note that R is closed and has axes parallel to the coordinate axes of $\mathbb{R}^d$. We define the length of the sides of R as $b_i - a_i$ for $i = 1, ..., d$ and the *volume* of R as

$$|R| = (b_1 - a_1) \cdot ... \cdot (b_d - a_d) \tag{4}$$

A closed rectangle with $b_i - a_i = c$ for all $i = 1, ..., d$ is called a cube.

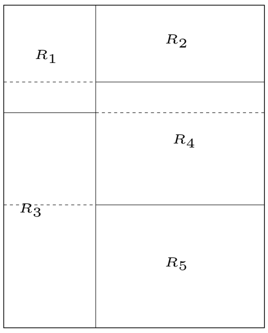
We make the obvious definition for an open rectangle.

A union of closed rectangles is said to be *almost disjoint* if the interiors of the rectangles are disjoint.

Lemma 2.1. *If a rectangle is the almost disjoint union of finitely many rectangles , say $R = \bigcup_{k=1}^{N} R_k$, then*

$$|R| = \sum_{k=1}^{N} |R_k|$$

Proof. One first extends the sides of the rectangles R_k to obtain new rectangles $\tilde{R}_1, ..., \tilde{R}_M$ as shown in the figure below.



This way one has the almost disjoint unions

$$R = \bigcup_{j=1}^{M} \tilde{R}_j \quad \text{and} \quad R_k = \bigcup_{j \in J_k} \tilde{R}_j \quad \text{for } k = 1, ..., N$$

where J_k is a partition of the integers from 1 to M. We claim that

$$|R| = \sum_{j=1}^{M} |\tilde{R}_j| \,.$$

This follows from writing out the volumes on the left and on the right and applying the distributive law. The same argument can be applied to the rectangles R_k, so

$$|R_k| = \sum_{j \in J_k} |\tilde{R}_j| \,.$$

Combining the above, we obtain

$$|R| = \sum_{j=1}^{M} |\tilde{R}_j| = \sum_{k=1}^{N} \sum_{j \in J_k} |\tilde{R}_j| = \sum_{j=1}^{N} |R_k| \,.$$

$\square$

Lemma 2.2. *If $R \subset \bigcup_{k=1}^{N} R_k$ with R and R_k (closed) rectangles, then $|R| \leq \sum_{k=1}^{N} |R_k|$.*

Proof. Exercise. Repeat the previous proof. $\square$

Our idea will now be to approximate an arbitrary set in $\mathbb{R}^d$ by rectangles. We first note the following

Theorem 2.1. *Every open subset $U \subset \mathbb{R}^d$, $d \geq 1$ can be written as a countable union of almost disjoint closed cubes.*

Proof. We construct the union as a (countably infinite) sequence of steps. We start with the grid of mesh 1 on $\mathbb{R}^d$ with lines parallel to the coordinate axes. A cube Q of the grid is accepted if $Q \subset U$, recjected if $Q \subset U^c$ and tentatively accepted otherwise. In the second step we bisect the tentatively accepted cubes to cubes of length 2^{-1} and again accept, reject or tentatively accept the subcubes. Continuing this procedure indefinitely, we obtain a countable union of accepted cubes and we claim that this union is U. To see this, note first that taking the grid of mesh size 2^{-N} of $\mathbb{R}^d$ we know that any cube contained in U has either been accepted or is contained in a cube that has been accepted in a previous step. Therefore, it suffices to show that $x \in U$ is contained in a cube of size 2^{-N} contained in U for large enough N. But this is easily deduced from the fact that U is open. $\square$

2.3 The exterior measure

Theorem 2.1 suggests to define the measure of an arbitrary open set U as

$$m(U) = \sum_{j=1}^{\infty} |R_j| \quad \text{where} \quad U = \bigcup_{j=1}^{\infty} R_j \ \text{ with } R_j \text{ almost disjoint.}$$

However, we would still have to show independence of this quantity from the decomposition. We're now going to achieve this and much more.

Definition 2.1. *For $E \subset \mathbb{R}^d$ any subset of $\mathbb{R}^d$ we define the* **exterior measure** *of E by*

$$m_\star(E) = \inf_{E \subset \bigcup_{j=1}^{\infty} Q_j} \sum_{j=1}^{\infty} |Q_j|$$

where we are taking the infimum over all countable coverings of E by closed cubes.

Note that $0 \leq m_\star \leq \infty$. Remarkably, replacing countable by finite in the above definition would yield a different quantity (see Example Sheet 1).

2.3.1 Examples

Let us compute the exterior measure for some elementary sets to see whether it agrees with our intuitive definition of volume above.

1. The exterior measure of a point is zero. (Why?)
2. The exterior measure of a closed cube Q is equal to its volume.
   To see this, note first that since Q covers itself we have $m_\star(Q) \leq |Q|$. To show the reverse direction we need to show that any covering Q_j satisfies

$$\sum_{j=1}^{\infty} |Q_j| \geq |Q| \, .$$

In fact, it suffices to show that any covering Q_j satisfies

$$\sum_{j=1}^{\infty} |Q_j| \geq |Q| - \epsilon\,.$$

for any $\epsilon > 0$. To show the latter, given $\epsilon > 0$ we choose for each j an open cube S_j with $Q_j \subset S_j$ and $|S_j| \leq |Q_j| + \frac{\epsilon}{2^j}$. Then $\bigcup_{j=1}^{\infty} S_j$ is an open cover of Q. Since Q is compact, there is a finite open subcover $\bigcup_{j=1}^{N} S_{k_j}$ of Q and clearly also $Q \subset \bigcup_{j=1}^{N} \overline{S_{k_j}}$. Now we can apply Lemma 2.2 to conclude $|Q| \leq \sum_{j=1}^{N} |S_{k_j}|$ and therefore

$$|Q| \leq \sum_{j=1}^{N} |S_{k_j}| \leq \sum_{j=1}^{N} \left(|Q_{k_j}| + \frac{\epsilon}{2^{k_j}} \right) \leq \sum_{j=1}^{\infty} |Q_j| + \epsilon$$

as desired.

3. The exterior measure of an open cube Q is equal to its volume.
Again, we have $m_\star(Q) \leq |\overline{Q}| = |Q|$ since the closed cube $\overline{Q}$ covers Q. In the reverse direction observe that we can find, for any $\epsilon > 0$, a closed cube Q_{in} contained in Q such that $|Q_{in}| \geq |Q| - \epsilon$. Therefore, we have $m_\star(Q) \geq m_\star(Q_{in}) \geq |Q| - \epsilon$, the first inequality holding because any covering of Q is also one of Q_{in}.

4. The exterior measure of a rectangle R is equal to its volume.
We sketch the argument. First of all, arguing as for the cube in 2. above, we obtain $|R| \leq m_\star(R)$.
For the reverse direction consider a grid of cubes of length $1/k$ on $\mathbb{R}^d$ and denote by $\tilde{Q}$ the cubes entirely contained in R and by $\tilde{Q}'$ those cubes intersecting both R and the complement R^c. Clearly for fixed k there are only finitely many cubes in $\tilde{Q}$ and $\tilde{Q}'$. In fact, it is easy to see that the number of cubes in $\tilde{Q}'$ is smaller than Ck^{d-1} for some uniform constant C depending only on the side lengths of R. We finally note that

$$R \subset \bigcup_{Q \in \tilde{Q} \cup \tilde{Q}'} Q$$

where the right hand side is a rectangle expressed as the union of finitely many almost disjoint union of cubes. By monotonicity and Lemma 2.1 we have

$$m_\star(R) \leq \sum_{Q \in \tilde{Q}} |Q| + \sum_{Q \in \tilde{Q}'} |Q| \leq |R| + \frac{C}{k}$$

and choosing k large we obtain the desired inequality for any $\epsilon > 0$.

5. The exterior measure of $\mathbb{R}^d$ is infinite, $m_\star\left(\mathbb{R}^d\right) = \infty$, as any covering of $\mathbb{R}^d$ must in particular cover arbitrarily large cubes.

2.3.2 Properties of the exterior measure

In the following we denote by E_i and E subsets of $\mathbb{R}^d$.

Proposition 2.1. *The exterior measure $m_\star$ satisfies the following properties:*

1. *If $E_1 \subset E_2$ then $m_\star(E_1) \leq m_\star(E_2)$ (monotonicity)*
2. *If $E = \bigcup_{j=1}^\infty E_j$, then $m_\star(E) \leq \sum_{j=1}^\infty m_\star(E_j)$ (countable subadditivity)*
3. *If $E \subset \mathbb{R}^d$, then*
$$m_\star(E) = \inf_{E \subset \mathcal{U}} m_\star(\mathcal{U})$$
with the infimum taken over all open sets that contain E.
4. *If $E = E_1 \bigcup E_2$ and $d(E_1, E_2) > 0$, then $m_\star(E) = m_\star(E_1) + m_\star(E_2)$ (finite additivity for sets with positive distance)*
5. *If $E = \bigcup_{j=1}^\infty Q_j$ is a union of almost disjoint cubes, then $m_\star(E) = \sum_{j=1}^\infty |Q_j|$ (countable additivity for almost disjoint cubes)*

Before we prove this, let us make some remarks. We note first that 5. gives us (in view of Theorem 2.1) a notion of volume of an arbitrary open set which is independent of the decomposition into cubes.

We also remark that one *cannot* conclude in general that if E_1 and E_2 are disjoint sets in $\mathbb{R}^d$, then $m(E_1) + m(E_2) = m(E_1 \cup E_2)$ holds.[2] However, for the class of sets ("Lebesgue measurable sets") that we are going to define in the next section, this property does hold, in fact it does so for countably many disjoint sets (countable additivity)!

Proof. The first property follows since any covering of E_2 of closed cubes is also a covering of E_1. For the second property we use the $\frac{\epsilon}{2^j}$-trick: With $\epsilon > 0$ arbitrary and fixed, we choose for any E_j a covering by closed subes $Q_{j,n}$ with
$$m_\star(E_j) \geq \sum_{n=1}^\infty |Q_{j,n}| - \frac{\epsilon}{2^j}\,.$$
Since $E \subset \bigcup_{j,n} Q_{j,n}$ is a covering by closed cubes we have
$$m_\star(E) \leq \sum_{j,n} |Q_{j,n}| = \sum_{j=1}^\infty \sum_{n=1}^\infty |Q_{j,n}| \leq \sum_{j=1}^\infty \left(m_\star(E_j) + \frac{\epsilon}{2^j} \right) \leq \sum_{j=1}^\infty m_\star(E_j) + \epsilon$$
where we inserted the previous estimate in the third step. Since this inequality holds for any $\epsilon > 0$ we are done.

To prove 3. we note that by 1. we clearly have $m_\star(E) \leq \inf m_\star(\mathcal{U})$, so we only need to show the $\geq$-direction. For this we can assume in addition that $m_\star(E) < \infty$ as otherwise the inequality holds trivially. To prove the inequality, it clearly suffices to construct for any $\epsilon > 0$ a set $\mathcal{U}$ with
$$m_\star(\mathcal{U}) \leq m_\star(E) + \epsilon\,. \tag{5}$$
To construct this $\mathcal{U}$ for $\epsilon > 0$ given, we use the $\frac{\epsilon}{2^j}$-trick: We first cover E with closed cubes such that
$$m_\star(E) \geq \sum_{j=1}^\infty |Q_j| - \frac{\epsilon}{2} \tag{6}$$

[2]See Example Sheet 2 after having read Section 2.5.

and then, for each cube Q_j choose an open cube Q_j^0 with $Q_j \subset Q_j^0$ and $|Q_j| \geq |Q_j^0| - \frac{\epsilon}{2}\frac{1}{2^j}$. We then define the union $\mathcal{U} := \bigcup_{j=1}^{\infty} Q_j^0$ and claim it satisfies the desired inequality. To check this we note

$$m_\star(\mathcal{U}) \leq \sum_{j=1}^{\infty} |Q_j^0| \leq \sum_{j=1}^{\infty} \left(|Q_j| + \frac{\epsilon}{2}\frac{1}{2^j} \right) \leq m_\star(E) + \epsilon$$

where the first step follows from monotonicity and the last from inserting (6).

We leave the proof of 4. as an exercise. [Outline: Note that $\leq$-direction follows from monotonicity. For $\geq$ choose $0 < \delta < d(E_1, E_2)$ and cover $E = E_1 \cup E_2$ by cubes such that $m_\star(E) \geq \sum_{j=1}^{\infty} |Q_j| - \epsilon$. Refine the cubes such that they all have length smaller than $\delta/2$. Note that each cube can only intersect either E_1 or E_2. Conclude.]

For 5. we note that the $\leq$ direction follows from monotonicity. To show $\geq$, we give ourselves ϵ of room. Let $\epsilon > 0$ be fixed. For each Q_j choose a closed cube $\tilde{Q}_j$ strictly contained in Q_j with

$$|\tilde{Q}_j| \geq |Q_j| - \frac{\epsilon}{2^j}\,.$$

For fixed N, the cubes $\tilde{Q}_j$ are disjoint and compact, hence (by an Exercise on Example Sheet 2) finite distance apart. We can hence apply 4. and conclude

$$m_\star(E) \geq m_\star\left(\bigcup_{j=1}^{N} \tilde{Q}_j \right) = \sum_{j=1}^{N} m_\star\left(\tilde{Q}_j\right) = \sum_{j=1}^{N} |\tilde{Q}_j| \geq \sum_{j=1}^{N} \left(Q_j| + \frac{\epsilon}{2^j} \right)$$

and hence after taking the limit $N \to \infty$ that

$$m_\star(E) \geq \sum_{j=1}^{\infty} |Q_j| + \epsilon$$

for any $\epsilon > 0$. The desired inequality follows. $\square$

2.4 The class of (Lebesgue) measurable sets

The motivation to define a class of measurable subsets on $\mathbb{R}^d$ that does not comprise *all* subsets of $\mathbb{R}^d$ is to have a class of sets for which countable additivity holds (cf. Theorem 2.2 below).

Definition 2.2. *A subset $E \subset \mathbb{R}^d$ is* **(Lebesgue) measurable** *if for any $\epsilon > 0$ there exists an open set $\mathcal{U}$ with $E \subset \mathcal{U}$ and*

$$m_\star(\mathcal{U} \setminus E) \leq \epsilon\,.$$

If E is (Lebesgue) measurable we define its **(Lebesgue) measure** *$m(E)$ to be $m(E) = m_\star(E)$.*

How big is the class of Lebesgue measurable sets? An answer is given by the following Proposition, which shows that this class of sets is closed under taking countable intersections, countable unions and complements (something we will call a σ-algebra of sets below, see Section 2.4.4), and contains the open and closed sets.

Proposition 2.2. *The following sets are (Lebesgue) measurable:*

1. *Every open set is measurable.*
2. *Sets of exterior measure* 0 *are measurable.*
3. *A countable union of measurable sets is measurable.*
4. *Closed sets are measurable.*
5. *The complement of a measurable set is measurable.*
6. *A countable intersection of measurable sets is measurable.*

Proof. Item (1) follows from the definition and (2) is a simple consequence of the third property of the exterior measure. Namely, if E has measure zero, there exists a $\mathcal{U}$ open with $E \subset \mathcal{U}$ and $m_\star(\mathcal{U}) \leq 0 + \epsilon$. Since $\mathcal{U} \setminus E \subset \mathcal{U}$ monotonicity implies $m_\star(\mathcal{U} \setminus E) \leq \epsilon$. For item (3) we use the standard $\frac{\epsilon}{2^n}$-trick. Let $A_1, A_2, ...$ be measurable sets. This means that we can pick $\mathcal{U}_i$ open such that $A_i \subset \mathcal{U}_i$ and $m_\star(\mathcal{U}_i \setminus A_i) \leq \frac{\epsilon}{2^i}$. Now $\cup_i A_i \subset \cup_i \mathcal{U}_i$ and hence by monotonicity

$$m_\star\left(\bigcup_i \mathcal{U}_i \setminus \bigcup_i A_i\right) \leq m_\star\left(\bigcup_i (\mathcal{U}_i \setminus \mathcal{A}_i)\right) \leq \sum_{i=1}^{\infty} m_\star(\mathcal{U}_i \setminus A_i) \leq \epsilon\,.$$

Turning to (4) we first observe that it suffices to prove the claim for closed and *bounded* (hence compact sets). This is because we can write an arbitrary closed set F as a countable union of compact sets $F = \cup_{n=1}^{\infty}\left(\overline{B_n}(0) \cap F\right)$. We can then use item (3). In particular we will assume now that F is compact, so in particular $m_\star(F) < \infty$. Let $\epsilon > 0$ be prescribed. By property (3) of the exterior measure we find a $\mathcal{U}$ open such that $F \subset \mathcal{U}$ and $\boxed{m_\star(U) \leq m_\star(F) + \epsilon}$. Since F is closed, $\mathcal{U} \setminus F$ is open and by Theorem 2.1 we can express it as a countable union of almost disjoint closed cubes, $\mathcal{U} \setminus F = \sum_{j=1}^{\infty} Q_j$. Clearly it suffices to show the measure of this is ϵ-small. To do this we observe that for any N, the union $K = \cup_{i=1}^{N} Q_j$ is compact. Since K and F are disjoint, they are a positive distance apart and we conclude

$$m_\star(\mathcal{U}) \geq m_\star(K \cup F) = m_\star(K) + m_\star(F) = \sum_{j=1}^{N} m_\star(Q_j) + m_\star(F)$$

Since $m_\star(F) < \infty$ we can subtract it an combine the above with the boxed inequality above to obtain for any N

$$\sum_{j=1}^{N} m_\star(Q_j) \leq \epsilon\,.$$

Taking the limit $N \to \infty$ we obtain the desired result as

$$m_\star(\mathcal{U} \setminus F) = m_\star(\cup_j Q_j) = \sum_{j=1}^{\infty} m_\star(Q_j) \leq \epsilon\,.$$

To show (5) we proceed as follows. For every $n \geq 1$ we choose an $\mathcal{U}_n$ open with $E \subset \mathcal{U}_n$ and $m_\star(\mathcal{U}_n \setminus E) \leq \frac{1}{n}$, which we can do since E is measurable. We now note that $(\mathcal{U}_n)^c$ is closed (hence measurable by (4)) and that $S = \cup_{n=1}^{\infty} (\mathcal{U}_n)^c$ is also measurable by (3). We now easily check the inclusions

$$S \subset E^c \quad \text{and} \quad E^c \setminus S \subset \mathcal{U}_n \setminus E \quad \text{for any } n.$$

to conclude

$$m_\star (E^c \setminus S) \leq m_\star (\mathcal{U}_n \setminus E) \leq \frac{1}{n}$$

and hence that $m_\star (E^c \setminus S) = 0$. Since $E^c = S \cup (E^c \setminus S)$ is a union of two measurable sets it is measurable.

For (6) it suffices to note that by de Morgan's laws

$$\bigcap_{j=1}^{\infty} E_j = \left(\bigcup_{j=1}^{\infty} (E_j)^c \right)^c$$

and use (5) and (3). $\square$

2.4.1 The property of countable additivity

The Lebesgue measurable sets satisfy the following crucial property which is called countable additivity:

Theorem 2.2. *If $E_1, E_2, ...$ are disjoint measurable sets and $E = \cup_{j=1}^{\infty} E_j$, then*

$$m(E) = \sum_{j=1}^{\infty} m(E_j) \, .$$

Proof. We first claim that it suffices to prove this for the E_j being bounded. Why? Suppose we had this result and we are trying to prove the general case. We take $(Q_k)_{k=1}^{\infty}$ the sequence of cubes of length k. We have $Q_k \subset Q_{k+1}$ for all $k \geq 1$ and we define $S_1 = Q_1$ and $S_k = Q_k \setminus Q_{k-1}$ for all $k \geq 2$. We define

$$E_{j,k} = E_j \cap S_k$$

which are measurable and bounded sets, disjoint for all j and k. We have

$$E = \bigcup_{j,k} E_{j,k} \quad \text{and} \quad E_j = \bigcup_{k=1}^{\infty} E_{j,k}$$

are both disjoint unions of bounded measurable sets. Since we are assuming we have the result in this case, we conclude

$$m(E) = m\left(\bigcup_{j,k}^{\infty} E_{j,k} \right) = \sum_{j,k} m(E_{j,k}) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} m(E_{j,k}) = \sum_{j=1}^{\infty} m(E_j) \, .$$

So now let's prove it assuming that the E_j are bounded. Clearly "$\leq$" holds by monotonicity, so we only need to prove "$\geq$" (and we can assume $m(E) < \infty$). Recall the idea of how we proved this in the case of cubes: We found strictly smaller closed cubes, then worked for finite N. The Lebesgue measurability gives us the following analogue:

Lemma 2.3. *Let $E \subset \mathbb{R}^d$ be measurable. Then for every $\epsilon > 0$ there exists a closed set $F \subset E$ with $m(E \setminus F) \leq \epsilon$.*

Proof. Apply the definition of measurability to the complement: There exists an open $\mathcal{U}$ with $E^c \subset \mathcal{U}$ and $m(\mathcal{U} \setminus E^c) \leq \epsilon$. If we define $F = \mathcal{U}^c$, then F is closed and in view of $E \setminus F = F^c \setminus E^c$ we have $m(E \setminus F) \leq \epsilon$. ☐

In particular, we can find a closed set F_j in each E_j with $m(E_j \setminus F_j) \leq \frac{\epsilon}{2^j}$. Now for fixed N the $F_1, ..., F_N$ are closed, bounded (hence compact) and disjoint, hence positive distance apart and we can apply the properties of the exterior measure to conclude for all N

$$m(E) \geq m\left(\bigcup_{j=1}^N F_j\right) = \sum_{j=1}^N m(F_j) \geq \sum_{j=1}^N \left(m(E_j) - \frac{\epsilon}{2^j}\right) \geq \sum_{j=1}^N m(E_j) - \epsilon\,.$$

The step in the middle follows from monotonicity applied to $E_j = F_j \cup (E_j \setminus F_j)$. Taking the limit as $N \to \infty$ and observing that this inequality holds for any $\epsilon > 0$ we have shown the desired inequality. ☐

2.4.2 Regularity properties of the Lebesgue measure

The following Proposition establishes certain "continuity" properties (from above and below) of the Lebesgue measure

Proposition 2.3. *Let $E_1, E_2, ...$ be a countable collection of measurable subsets of $\mathbb{R}^d$.*

1. *If the E_k are increasing to E in that $E_k \subset E_{k+1}$ for all k and $E = \cup_{i=1}^\infty E_i$. Then*
$$m(E) = \lim_{N\to\infty} m(E_N)\ .$$

2. *If the E_k are decreasing to E in that $E_k \supset E_{k+1}$ for all k and $E = \cap_{i=1}^\infty E_i$* **and $m(E_k) < \infty$ for some k**, *then*
$$m(E) = \lim_{N\to\infty} m(E_N)$$

Remark 2.1. *The bold emphasises the additional condition in the decreasing case. To see that the conclusion is not generally valid without this condition consider the case $E_n = [n, \infty)$.*

Proof. For the first part, set $G_1 = E_1$ and $G_k = E_k \setminus E_{k-1}$. The G_k are then disjoint and measurable and $\cup_k G_k = \cup_k E_k$. Now apply countable additivity for disjoint measurable sets (Theorem 2.2) to obtain

$$m(\cup_{i=1}^\infty E_i) = m(\cup_{i=1}^\infty G_i) = \sum_{i=1}^\infty m(G_i) = \lim_{N\to\infty} \sum_{i=1}^N m(G_i) = \lim_{N\to\infty} m\left(\cup_{i=1}^N G_i\right)$$

and noting that $E_N = \cup_{i=1}^N G_i$ the proof is complete.

For the second part we first note that wlog $m(E_1) < \infty$ as we can always forget about finitely many elements in the sequence. We now reduce it to case 1 by defining $F_j = E_1 \setminus E_j$ which is clearly an increasing sequence $F_1 \subset F_2 \subset$

$F_3 \subset ....$ and also $m(F_j) = m(E_1) - m(E_j)$ since $E_1 = F_j \cup E_j$ is a disjoint union of measurable sets. We also observe that $\cup_{j=1}^{\infty} F_j = E_1 \setminus E$. Combining these things and using the first part we find

$$\lim_{j\to\infty} (m(E_1) - m(E_j)) = \lim_{j\to\infty} m(F_j) = m\left(\cup_{j=1}^{\infty} F_j\right) = m(E_1 \setminus E) = m(E_1) - m(E) \ .$$

As $m(E_1) < \infty$, we can subtract it from both sides and obtain the result. □

2.4.3 Invariance properties of the Lebesgue measure

The Lebesgue measure is invariant under translations rotations and reflections. In this section we discuss the translation invariance.

For $E \subset \mathbb{R}^d$ measurable we define the h-translated set

$$E_h = E + h := \{x + h \mid x \in E\} \quad \text{for } h \in \mathbb{R}^d \text{ fixed.}$$

Theorem 2.3. *If E is measurable, then the h-translated set E_h is also measurable and $m(E) = m(E_h)$.*

Proof. The conclusion clearly holds if E is a cube. This observation allows us to conclude that $m_\star(E) = m_\star(E_h)$ holds for an arbitrary set E, since given any covering Q_j of E by cubes, the h-translated cubes will cover E_h. Finally, if E is measurable and $\mathcal{U}$ open with $E \subset \mathcal{U}$ and $m_\star(\mathcal{U} \setminus E) \leq \epsilon$ is given, then the translated set $\mathcal{U}_h$ is also open, satisfies $E_h \subset \mathcal{U}_h$ and $m_\star(\mathcal{U}_h \setminus E_h) \leq \epsilon$. □

In the same way we can prove the invariance under reflexions and rotations.

2.4.4 σ-algebras and Borel sets

Definition 2.3. *A σ-algebra is a non-empty collection of subsets of $\mathbb{R}^d$ that is closed under countable unions, countable intersections and complements.*

Note that the empty set and the set $E = \mathbb{R}^d$ are contained in any σ-algebra (why?). To give some examples, we note that clearly the collection of all subsets of $\mathbb{R}^d$ forms a σ-algebra, $\mathcal{M}_{all}$. We have also seen the σ-algebra of Lebesgue measurable sets (cf. Proposition 2.2), $\mathcal{M}_{Lebesgue}$. It turns out that (assuming the axiom of choice) $\mathcal{M}_{Lebesgue}$ is properly contained in $\mathcal{M}_{all}$. In other words, there do exist non-measurable subsets on $\mathbb{R}^d$ (cf. Section 2.5). As another example we will construct the Borel σ-algebra, $\mathcal{B}_{\mathbb{R}}$ below, which will turn out to be properly contained in $\mathcal{M}_{Lebesgue}$.

We note that it is easy to show that the intersection of two σ-algebras $\mathcal{M}$ and $\mathcal{N}$ is again a σ algebra. (Clearly, if $A \in \mathcal{M} \cap \mathcal{N}$ then $A \in \mathcal{M}$ and $A \in \mathcal{N}$. Hence $A^c \in \mathcal{M}$ and $A^c \in \mathcal{N}$ and therefore $A^c \in \mathcal{M} \cap \mathcal{N}$. Similarly for countable unions and intersections.) In fact, using the argument sketched in the bracket one shows that the intersection of an arbitrary collection (not necessarily countable) of σ-algebras is again a σ-algebra. This observation allows us to prove the following Theorem:

Theorem 2.4. *If $\mathcal{F}$ is an arbitrary collection of subsets of $\mathbb{R}^d$, there exists a unique smallest σ-algebra $\mathcal{M}$ which contains $\mathcal{F}$.*

We call the σ-algebra promised in the theorem *the σ-algebra generated by $\mathcal{F}$* and denote it by $\mathcal{M}(\mathcal{F})$.

Proof. Take the intersection of all σ-algebras which contain $\mathcal{F}$. Note this intersection is non-empty as $\mathcal{M}_{all}$ is a σ-algebra containing $\mathcal{F}$. The intersection is the smallest σ algebra containing $\mathcal{F}$ in the sense that $\mathcal{M}(\mathcal{F})$ is contained in any σ-algebra which includes the sets from $\mathcal{F}$. This also gives the uniqueness. ☐

Definition 2.4. *If $\mathfrak{U}$ denotes the collection of all open sets in $\mathbb{R}^d$, then $\mathcal{M}(\mathfrak{U})$ is called the* **Borel σ-algebra**, *denoted $\mathcal{B}_{\mathbb{R}}$ (containing the* **Borel sets***).*

Observation 2.1. *The Borel σ-algebra can also be generated by closed cubes, i.e. if $\mathfrak{Q}$ denotes the collection of all closed cubes in $\mathbb{R}^d$, then $\mathcal{M}(\mathfrak{Q}) = \mathcal{B}_{\mathbb{R}}$.*

To verify the Observation note first that any open set lies in the σ-algebra of closed cubes by Theorem 2.1 and conversely any cube lies in the Borel σ-algebra. We combine this with the following general fact: If $\mathcal{E}$ is any collection of subsets of $\mathbb{R}^d$ satisfying $\mathcal{E} \subset \mathcal{M}(\mathcal{F})$, then $\mathcal{M}(\mathcal{E}) \subset \mathcal{M}(\mathcal{F})$. (Indeed, $\mathcal{M}(\mathcal{F})$ is a σ-algebra containing $\mathcal{E}$, so the smallest σ-algebra containing $\mathcal{E}$ must be contained in it.)

It is of course natural to ask whether the Borel sets are properly contained in the Lebesgue measurable sets. The answer is yes and the following Proposition (as well as the exercises of Example Sheet 3) clarifies the relation between the two. First we need a definition

Definition 2.5. *A* G_δ**-set** *is a countable intersection of open sets. An* F_σ**-set** *is countable union of closed sets.*

Proposition 2.4. *Let $E \subset \mathbb{R}^d$. Then E is measurable*

1. *if and only if E is G_δ with a set of measure zero removed,*
2. *if and only if E is the union of an F_σ and a set of measure zero.*

In particular, any Lebesgue measurable set can be obtained from a Borel set by adjoining a set of a measure zero to the latter.

Proof. If E satisfies the conditions in (1) or (2) then E is measurable since G_δ, F_σ and sets of measure zero are measurable.

For the converse of (1) let now E be measurable. We choose for any $n \geq 1$ an open $\mathcal{U}_n$ with $m(\mathcal{U}_n \setminus E) \leq \frac{1}{n}$. Then clearly $S = \cap_{n=1}^{\infty} \mathcal{U}_n$ is a G_δ containing E and the measure of the complement $S \setminus E$ is zero in view of $m(S \setminus E) \leq m(\mathcal{U}_n \setminus E) \leq \frac{1}{n}$ being true for all $n \geq 1$. Write $E = S \setminus (S \setminus E)$.

For the converse of (2) let E be measurable and choose for any $n \geq 1$ a closed F_n with $m(E \setminus F_n) \leq \frac{1}{n}$. Then clearly $S = \cup_{n=1}^{\infty} F_n$ is an F_σ which is contained in E and the complement $E \setminus S$ has measure zero in view of $m(E \setminus S) \leq m(E \setminus F_n) \leq \frac{1}{n}$ being true for all $n \geq 1$. Write $E = S \cup (E \setminus S)$. ☐

2.5 Construction of a non-measurable set

We construct a non-measurable set in $d = 1$. We start with $X = [0, 1]$ and define the following equivalence relation on X.

$$x \sim y \quad \text{if } x - y \in \mathbb{Q}. \qquad \text{(Note that } x - y \in [-1, 1].\text{)}$$

It is easy to check this is indeed an equivalence relation. By the fundamental theorem of equivalence relations X can be partitioned as the union of the (disjoint) equivalence classes

$$X = [0,1] = \bigcup_{\alpha} E_\alpha$$

To construct the non-measurable set $\mathcal{N}$ we choose from every equivalence class E_α precisely one element x_α (this uses the axiom of choice). The set $\mathcal{N}$ is the collection of these chosen x_α:

$$\mathcal{N} := \{x_\alpha\}$$

We claim that $\mathcal{N}$ is not measurable. Suppose it was. Take $(r_k)_{k=1}^{\infty}$ an enumeration of the rationals in $[-1,1]$ and consider the translates

$$\mathcal{N}_k := \mathcal{N} + r_k$$

- We first note that the $\mathcal{N}_k$ are disjoint. Indeed, if $u \in \mathcal{N}_k \cap \mathcal{N}_{k'}$ for $k \neq k'$, then

  $$u = x_\alpha + r_k = x_\beta + r_{k'} \quad \text{and hence} \quad x_\alpha - x_\beta = r_{k'} - r_k \neq 0$$

  from which we conclude $\alpha \neq \beta$ and that x_α and x_β are in the same equivalence class, which is impossible, since we selected precisely one element from each equivalence class.

- We have $\mathcal{N}_k \subset [-1,2]$ for any k and in particular $\cup_{k=1}^{\infty} \mathcal{N}_k \subset [-1,2]$. This follows easily from the fact that $\mathcal{N} \subset [0,1]$ and $r_k \in [-1,1]$.

- We have $[0,1] \subset \bigcup_k \mathcal{N}_k$. To see this, let $x \in [0,1]$. Since x sits in some equivalence class, we have $x = x_\alpha + r_k$ for some x_α and some $r_k \in [-1,1]$. It follows that $x \in \mathcal{N}_k$ for some k.

- We have $m(\mathcal{N}) = m(\mathcal{N}_k)$ by translation invariance.

Combining the facts above we have

$$[0,1] \subset \bigcup_k \mathcal{N}_k \subset [-1,2]$$

with the union being a disjoint union. Monotonicity and the countable additivity for measurable sets imply that

$$1 \leq \sum_{k=1}^{\infty} m(\mathcal{N}_k) \leq 3$$

This is a contradiction both in the case where $m(\mathcal{N}) = m(\mathcal{N}_k) = 0$ and when $m(\mathcal{N}) = m(\mathcal{N}_k) > 0$.

Remark 2.2. *We can use the non measurable set $\mathcal{N}$ to show that finite additivity generally fails for the exterior measure as was claimed in Section 2.3.2.*

2.6 Measurable Functions

From measurable sets we now turn to measurable functions. To motivate the definition, it is actually worth taking a step back viewing things from a slightly more abstract point of view.

2.6.1 Some abstract preliminary remarks

Remember that a topological space (X, τ) is a set X together with a collection τ of subsets of X (which contains the empty set and X itself) which are declared to be open and which are closed under arbitrary unions and *finite* intersections. If X and Y are topological spaces, then $f : X \to Y$ is continuous is $f^{-1}(\mathcal{V})$ is open for any open set $\mathcal{V}$ in Y.

We can define a measure space $(X, \mathcal{M})$ as a set X together with a collection of subsets $\mathcal{M}$ (which contains the empty set and X itself) which are declared to be measurable and which are closed under countable unions, countable intersections and taking complements.

If X is a measure space and Y is a topological space, then we say that $f : X \to Y$ is measurable provided that $f^{-1}(\mathcal{V})$ is measurable for all open sets $\mathcal{V}$ in Y.

In what we do in the next 4-5 weeks, X will always be $\mathbb{R}^d$ and Y will either be $\mathbb{R}$ or the extended reals $\overline{\mathbb{R}}$. Recall the latter is a topological space whose open sets are the segments (a, b), $[-\infty, a)$, $(b, \infty]$ for $a, b \in \mathbb{R}$ and unions thereof.

Because for us $Y = \mathbb{R}$ or $Y = \overline{\mathbb{R}}$, you can in principle forget about the abstract point of view just described. However, it is useful to have it at the back of your mind.

2.6.2 Definitions and equivalent formulations of measurability

For E a measurable subset of $\mathbb{R}^d$ we will consider functions $f : \mathbb{R}^d \supset E \to \mathbb{R}$, called finite (real)-valued functions and $f : \mathbb{R}^d \supset E \to \overline{\mathbb{R}}$ called extended real-valued functions. We'll be most interested in functions defined on all of $\mathbb{R}^d$, i.e. the case $E = \mathbb{R}^d$.

Definition 2.6. *Let Y be a topological space (for us $\mathbb{R}$ or $\overline{\mathbb{R}}$). A function $f : \mathbb{R}^d \supset E \to Y$ is measurable if $f^{-1}(\mathcal{V})$ is measurable for all $\mathcal{V}$ open in Y.*

Proposition 2.5. *Let $Y = \mathbb{R}$. Then the following are equivalent*

1. *f is measurable*
2. *$f^{-1}((a, \infty))$ is measurable for all $a \in \mathbb{R}$*
3. *$f^{-1}([a, \infty))$ is measurable for all $a \in \mathbb{R}$*
4. *$f^{-1}((-\infty, a))$ is measurable for all $a \in \mathbb{R}$*
5. *$f^{-1}(a, b)$ is measurable for all $a, b \in \mathbb{R}$.*
6. *...*

Proof. (1) $\implies$ (2) is trivial. To get (3) from (2) we observe

$$f^{-1}([a, \infty)) = f^{-1}\left(\bigcap_{n=1}^{\infty}\left(a - \frac{1}{n}, \infty\right)\right) = \bigcap_{n=1}^{\infty} f^{-1}\left(\left(a - \frac{1}{n}, \infty\right)\right)$$

where we noted that the inverse image commutes with arbitrary intersections. To get (4) from (3) we note $f^{-1}((-\infty,a)) = f^{-1}([a,\infty)^c) = \left(f^{-1}([a,\infty))\right)^c$ using that the inverse image commutes with taking the complement. I leave the implication (4) to (5) to you and we conclude (5) $\implies$ (1) by noting that any open set $\mathcal{U}$ can be written as $\cup_{n=1}^{\infty}(a_n,b_n)$ (why?). $\square$

Using the same arguments one proves

Proposition 2.6. *Let $Y = \overline{\mathbb{R}}$. Then the following are equivalent*

1. *f is measurable*
2. *$f^{-1}((a,\infty])$ is measurable for all $a \in \mathbb{R}$*
3. *$f^{-1}([a,\infty])$ is measurable for all $a \in \mathbb{R}$*
4. *$f^{-1}([\infty,a))$ is measurable for all $a \in \mathbb{R}$*
5. *$f^{-1}(a,b)$ is measurable for all $a,b \in \mathbb{R}$ and $f^{-1}(\infty)$ and $f^{-1}(-\infty)$ are measurable*
6. *...*

Proof. Exercise. $\square$

To sum up this discussion from a practical point of view, in order to check whether a given function $f : \mathbb{R}^d \supset E \to \mathbb{R}$ is measurable it suffices to check whether the sets $\{x \in E \mid f(x) > a\}$ are measurable for all a. Note the latter sets are equal to $f^{-1}((a,\infty])$ if $Y = \overline{\mathbb{R}}$ and equal to $f^{-1}((a,\infty))$ if $Y = \mathbb{R}$. Alternatively it suffices to check that the sets $\{x \in E \mid f(x) < a\}$ are measurable for all a. Etc.

The key ingredient in the above propositions was that the inverse image commutes with arbitrary unions, intersections and complements combined with the σ-algebra structure on the measurable sets. We can formulate this a bit more abstractly as follows:

Proposition 2.7. *The following are equivalent:*

1. *The function $f : \mathbb{R}^d \supset E \to Y$ is measurable*
2. *The set $f^{-1}(F)$ is measurable for any $F \in \mathcal{B}_Y$, the Borel σ-algebra of Y.*
3. *The set $f^{-1}(F)$ is measurable for any $F \in \mathcal{F}$ where $\mathcal{F}$ is any collection of sets generating $\mathcal{B}_Y$, the Borel σ-algebra of Y.*

Proof. The implications (2) $\implies$ (3) and (2) $\implies$ (1) are immediate.

Let $\mathcal{N}$ be the collection of all sets on Y such that $f^{-1}(F)$ is measurable for all $F \in \mathcal{N}$. One easily checks that this is a σ-algebra on Y using that the inverse image commutes with unions interections and complements (do it!).

(1) $\implies$ (2): If f is measurable, then the σ-algebra $\mathcal{N}$ must contain the open sets in Y and since the Borel σ-algebra is by definition the smallest such σ-algebra we have $\mathcal{B}_Y \subset \mathcal{N}$ and in particular the inverse image of any Borel set must be measurable.

(3) $\implies$ (2): If (3) holds, then $\mathcal{F} \subset \mathcal{N}$. Hence $\mathcal{M}(\mathcal{F}) \subset \mathcal{N}$ and since $\mathcal{M}(\mathcal{F}) = \mathcal{B}_Y$ we are done. $\square$

Note that in view of Proposition 2.7, Proposition 2.5 follows immediately from the observation that the intervals appearing in it (individually) generate $\mathcal{B}_{\mathbb{R}}$ and similarly for Proposition 2.6.

2.6.3 Properties 1: Behaviour under compositions

Proposition 2.8. *Let Y be a topological space.*

1. *If $f : \mathbb{R}^d \to Y$ is continuous, then it is measurable.*
2. *If Z is a topological space and $\Phi : Y \to Z$ is continuous, then the composition $\Phi \circ f$ is measurable if $f : \mathbb{R}^d \to Y$ is measurable.*

Proof. For (1) we simply note that by continuity $f^{-1}(\mathcal{V})$ is open (hence measurable) for any $\mathcal{V} \subset Y$ open. For (2) we observe that $(\Phi \circ f)^{-1}(\mathcal{W}) = f^{-1}\left(\Phi^{-1}(\mathcal{W})\right)$ and since $\Phi^{-1}(\mathcal{W})$ is open for any open $\mathcal{W} \subset Z$ we conclude that the composition is measurable if f is. $\square$

Warning: The composition of two measurable functions does not have to be measurable, even if the Φ in the composition $f \circ \Phi$ is continuous. However, the composition of two Borel measurable functions is Borel measurable.

2.6.4 Properties 2: Behaviour under limits

We recall that if (a_n) is a sequence in $\overline{\mathbb{R}}$ we can consider the sequence

$$b_k = \sup(a_k, a_{k+1}, ...)$$

which is clearly a non-increasing sequence in k: $b_1 \geq b_2 \geq ....$ In particular we can define

$$B = \inf(b_1, b_2, ...)$$

We call B the upper limit of the (a_n) and use the notation

$$B = \limsup_{n\to\infty} a_n = \inf_{k\geq 1} \sup_{n\geq k} a_n. \tag{7}$$

It can be proved that there exists a subsequence of (a_n) that converges to B and that B is the largest number with this property.

The $\liminf a_n$ is defined analogously replacing sup by inf and conversely in the above definition.

Recall that $\liminf a_n = \limsup a_n$ if and only if the limit $\lim a_n$ exists.

Suppose now (f_n) is a sequence of measurable functions $f_n : \mathbb{R}^d \to \overline{\mathbb{R}}$. Then we can define pointwise the functions

$$\left(\sup_n f_n\right)(x) := \sup_n (f_n(x)) \quad \text{and} \quad \left(\inf_n f_n\right)(x) := \inf_n (f_n(x))$$

$$\left(\limsup_{n\to\infty}\right)(x) := \limsup_{n\to\infty} (f_n(x)) \quad \text{and} \quad \left(\liminf_{n\to\infty}\right)(x) := \liminf_{n\to\infty} (f_n(x))$$

Theorem 2.5. *If (f_n) is a sequence of measurable functions as above, then the four functions above are all measurable.*

Proof. We observe that $\sup_n f_n$ is measurable if the set $\{x|\sup_n(f_n(x)) > a\}$ is measurable for all $a \in \mathbb{R}$. But the latter set can be written as $\cup_n\{x|f_n(x) > a\} = \cup_n f_n^{-1}((a,\infty])$ and this set is clearly measurable. The inf is done analogously and for the lim sup it follows from the definition (7). $\square$

Corollary 2.1. *If the sequence in the theorem converges to f, i.e. $\lim_{n\to\infty}(x) = f(x)$ for every x, then f is measurable.*

2.6.5 Properties 3: Behaviour of sums and products

Theorem 2.6. *Let $Y = \mathbb{R}$ or $Y = \mathbb{R}$. If $f : \mathbb{R}^d \supset E \to Y$ and $g : \mathbb{R}^d \supset E \to Y$ are measurable, then*

1. *$-f$ is measurable*
2. *f^2 is measurable*
3. *$k \cdot f$ for $k \in \mathbb{R}$ is measurable*
   $f+g$ and $f \cdot g$ are measurable if both f and g are finite valued (i.e. $Y = \mathbb{R}$).
4. *The functions $\max(f,g)$, $\min(f,g)$ and $|f|$ are measurable*

Proof. The first follows from $\{x \mid -f(x) > a\} = \{x \mid f(x) < -a\}$ for every $a \in \mathbb{R}$. For the second, we have for $a \geq 0$ the identity

$$\{x \mid f^2(x) > a\} = \{x \mid f(x) > a^{\frac{1}{2}}\} \bigcup \{x \mid f(x) < -a^{\frac{1}{2}}\}$$

and for $a < 0$ that $\{x \mid f^2(x) > a\} = \{x \mid f^2(x) > 0\} \cup \{x \mid f(x) = 0\}$. For the third assertion, observe that

$$\Big\{x \mid k \cdot f(x) > a\Big\} = \Big\{x \mid f(x) > \frac{a}{k}\Big\},$$

$$\{x \mid (f+g)(x) > a\} = \bigcup_{r\in\mathbb{Q}} \{x \mid f(x) > a - r\} \cap \{x \mid g(x) > r\},$$

and the formula

$$fg = \frac{1}{4}\left((f+g)^2 - (f-g)^2\right) .$$

For the fourth we note that

$$\{x \mid \max(f,g)(x) > a\} = \{x \mid f(x) > a\} \cap \{x \mid g(x) > a\}$$

$$\{x \mid \min(f,g)(x) > a\} = \{x \mid f(x) > a\} \cup \{x \mid g(x) > a\}$$

and the formula $|f| = \max(f,0) - \min(f,0)$ for the absolute value. $\square$

2.6.6 The notion of "almost everywhere"

Let $f : E \to Y$ and $g : E \to Y$ be two functions. We say that f **is equal to** g **almost everywhere** if the set $\{x \mid f(x) \neq g(x)\}$ has measure zero. More generally we define a statement to hold almost everywhere, if it holds except for a measure zero set.

We conclude that if $f = g$ for almost every $x \in E$ as above, then f is measurable if g is. This follows as the sets $\{x \mid f(x) > a\}$ and $\{x \mid g(x) > a\}$

differ by a set of measure zero (which is measurable). Indeed, if N is the set where the two sets differ, then

$$\{x \mid f(x) > a\} = (\{x \mid g(x) > a\} \cap N^c) \cup (\{x \mid f(x) > a\} \cap N)$$

and the first set is measurable by the measurability of g and the second because it is a subset of a set of measure 0.

2.7 Building blocks of integration theory

Given a set $E \subset \mathbb{R}^d$ we define the characteristic function of E as

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases}$$

It is easy to see that the function χ_E is measurable if and only if the set E is measurable.

2.7.1 Simple functions

A simple function is defined as a finite linear combination (with real coefficients) of characteristic functions, i.e.

$$f(x) = \sum_{k=1}^{N} a_k \chi_{E_k}(x)$$

for some $N \in \mathbb{N}$, where $a_k \in \mathbb{R}$ and the E_k are measurable sets. Equivalently, we may define a simple function as a measurable function $f : \mathbb{R}^d \to \mathbb{R}$, which assumes only finitely many distinct values. Note that a simple function has a canonical expression where all the a_k are distinct and the E_k are disjoint measurable sets (why?). The simple functions will constitute the building blocks for the Lebesgue integral that we are about to define.

2.7.2 Step functions

Step functions are a narrower class of functions and form the building blocks for the Riemann integral that you already know. They are defined as finite linear combinations of characteristic functions of *rectangles*, i.e.

$$f(x) = \sum_{k=1}^{N} a_k \chi_{R_k}(x)$$

with the R_k being rectangles.

2.8 Approximation Theorems

2.8.1 Approximating a measurable function by simple functions

Our first approximation theorem concerns approximating *non-negative* measurable functions by simple functions. (It works in the same way whether the range is $[0, \infty]$ or $[0, \infty)$.)

Theorem 2.7. *Let $f : \mathbb{R}^d \to [0, \infty]$ be measurable and non-negative. Then there exists an increasing sequence of non-negative simple functions $(\phi_k)_{k=1}^{\infty}$ that converges pointwise to f:*

$$\phi_k(x) \leq \phi_{k+1}(x) \quad \textit{for all } k \textit{ and } x \quad \textit{and} \quad \lim_{k\to\infty} \phi_k(x) = f(x) \quad \textit{for all } x$$

Proof. We first truncate f as follows: We let Q_N denote the cube of length N centred at the origin and define

$$F_N(x) = \begin{cases} f(x) & \text{if } x \in Q_N \text{ and } f(x) \leq N \\ N & \text{if } x \in Q_N \text{ and } f(x) > N \\ 0 & \text{otherwise} \end{cases}$$

It is easy to see that F_N is still measurable and that $F_N(x) \to f(x)$ for every x as $N \to \infty$. We now approximate F_N by a sequence of simple functions by partitioning the range as follows: We decompose the range $[0, N]$ into $N \cdot M$ intervals of length $1/M$ and define

$$E_{\ell,M} = \left\{ x \in Q_N \mid \frac{\ell}{M} < F_N(x) \leq \frac{\ell+1}{M} \right\} \quad \text{for } 0 \leq \ell < NM\,.$$

These sets are all measurable and we can thus define the simple function

$$F_{N,M}(x) = \sum_{\ell=0}^{NM-1} \frac{\ell}{M} \chi_{E_{\ell,M}}(x)\,.$$

Note that by construction we have $F_N(x) - F_{N,M}(x) \leq \frac{1}{M}$ for all x. We finally choose $M = N = 2^k$ and define a sequence of simple functions via

$$\phi_k(x) := F_{2^k,2^k}(x)\,.$$

By construction we have $F_{2^k}(x) - \phi_k(x) \leq \frac{1}{2^k}$ for all x and $\phi_k(x) \to f(x)$ for all x as $k \to \infty$. Finally, ϕ_k is also increasing (why?). □

We next remove the assumption that f should be non-negative:

Theorem 2.8. *Let $f : \mathbb{R}^d \to [-\infty, \infty]$ be measurable. Then there exists a sequence of simple functions $(\phi_k)_{k=1}^{\infty}$ that satisfies*

$$|\phi_k(x)| \leq |\phi_{k+1}(x)| \quad \textit{for all } k \textit{ and } x \quad \textit{and} \quad \lim_{k\to\infty} \phi_k(x) = f(x) \quad \textit{for all } x.$$

In particular, $|\phi_k(x)| \leq |f(x)|$ holds for all x and all k.

Proof. We decompose $f(x) = f^+(x) - f^-(x)$ where $f^+(x) := \max(f(x), 0)$ and $f^-(x) = \max(-f(x), 0)$ are the positive and the negative part of f respectively. Recall that the latter are measurable by Theorem 2.6. Since both f^+ and f^- are non-negative, we can apply the previous Theorem providing us with simple functions $\phi_k \to f^+$ and $\psi_k \to f^-$. Hence $\Phi_k := \phi_k - \psi_k$ satisfies $\Phi_k(x) \to f(x)$ for all x as $k \to \infty$. Finally, note that $|\Phi_k(x)| = \phi_k(x) + \psi_k(x)$ (since if $\phi_k(x) \neq 0$ for some x then $\psi_k(x) = 0$ for this x and conversely!), so $|\Phi_k|$ is indeed increasing. The last claim is straightforward. □

2.8.2 Approximating a mesurable function by step functions

We can also approximate a measurable function by the simpler step functions. The price we pay is that the convergence is only *almost everywhere.* Recall the measurable function (2) from the introduction to illustrate that one cannot expect the convergence to hold everywhere when approximating with step functions.

We first isolate the following proposition:

Proposition 2.9. *Let $h = \sum_{k=1}^{N} a_k \chi_{E_k}$ be a simple function with $m(E_k) < \infty$ for all k. Then for any $\epsilon > 0$ there exists a step function φ such that*

$$m(\{x \mid \varphi(x) \neq h(x)\}) \leq \epsilon \quad \textit{for all } k.$$

In other words, a step function approximates the simple function up to a set of arbitrary small measure.

Proof. One first convinces oneself that it suffices to prove the result for $h = \chi_E$ and E measurable with $m(E) < \infty$ (why?). To prove the latter, recall from Example Sheet 2 that one can approximate E by finitely many closed cubes such that $m\left(E\Delta \bigcup_{j=1}^{N} Q_j\right) \leq \frac{\epsilon}{2}$. Moreover one can (using the usual procedure of extending the sides of the cubes to form almost disjoint rectangles) find disjoint rectangles with $\bigcup_{j=1}^{N} Q_j = \bigcup_{j=1}^{M} R_j$. We then define $\varphi(x) = \sum_{k=1}^{M} \chi_{R_k}(x)$. $\square$

Theorem 2.9. *Let $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ be measurable. Then there exists a sequence of step functions $(\varphi_k)_{k=1}^{\infty}$ that convergence pointwise for almost every x.*

Proof. We first note that f can be approximated by a sequence (ψ_k) of simple functions by Theorem 2.7 (with all E_j appearing in each of the ψ_k having finite measure). For each ψ_k we apply Proposition 2.9 find a step function φ_k approximating ψ_k such that

$$m(\{x \mid \varphi_k(x) \neq \psi_k(x)\}) \leq \frac{1}{2^k} \quad \text{for all } k.$$

We claim that $\varphi_k \to f$ except on a set of measure 0. To see this, we use the Borel-Cantelli Lemma (Example Sheet 2). Indeed, the fact that

$$\sum_{k=1}^{\infty} m(\{x \mid \varphi_k(x) \neq \psi_k(x)\}) < \infty$$

allows us to conclude that the set

$$M = \{x \mid \varphi_k(x) \neq \psi_k(x) \quad \text{for infinitely many } k\}$$

has measure zero. We have $\mathbb{R}^d = M \cup M^c$ and for any $x \in M^c$ we have $\varphi_k(x) = \psi_k(x)$ for all $k \geq K$ for sufficiently large K. Hence $\lim_{k\to\infty} \varphi_k(x) \to f(x)$ for any $x \in M^c$. $\square$

2.9 Littlewoods Three Principles

Before we turn to the integration theory let us introduce three heuristic principles which nicely summarise the relation of the new notions of "measurable set" and "measurable function" that we introduced above to the more familiar notions of rectangles and continuous functions:

1. Every measurable set is nearly a finite union of closed cubes. (See Example Sheet 2, Exercise 3c for the precise statement.)
2. Every measurable function is nearly continuous.
3. Every convergent sequence of measurable functions is nearly uniformly convergent.

Let us first make item 3 more precise (Egoroff's Theorem) and then use it to formulate and prove item 2 precisely (Lusin's Theorem).

Theorem 2.10 (Egoroff). *Suppose (f_k) is a sequence of measurable functions $f_k : E \to \mathbb{R}$ with E measurable and $m(E) < \infty$. Assume that*

$$f_k \to f \quad \text{a.e. on } E.$$

Given $\epsilon > 0$ we can find a closed set $A_\epsilon \subset E$ such that $m(A \setminus A_\epsilon) \le \epsilon$ and $f_k \to f$ uniformly on A_ϵ.

Remark 2.3. *Here is an example illustrating the theorem. Let $f_n : [-1,1] \to \mathbb{R}$*

$$f_n(x) = \begin{cases} \frac{1}{nx} & \text{if } x > 0 \\ 0 & \text{if } x \le 0 \end{cases}$$

Then $f_n(x) \to 0$ for all x but not uniformly near zero.

Remark 2.4. *If $m(E) = \infty$ the result may not hold as the example of the moving bump $f_n = \chi_{[n,n+1]}$ illustrates. The f_n converge to zero but not uniformly, even after removing a set of large measure.*

Remark 2.5. *The theorem remains true for the target space being the extended reals $\overline{\mathbb{R}}$. (How do we define uniform convergence in this case?)*

Proof. We first observe that is suffices to prove the result for $f_k \to f$ *everywhere* on E.[3] For later purposes we also define an N (depending only on $\epsilon > 0$) by $\sum_{n=N}^{\infty} \frac{1}{2^n} < \frac{\epsilon}{2}$. Now let us start the proof. We define the set

$$E_k^n = \left\{ x \in E \;\middle|\; |f_j(x) - f(x)| < \frac{1}{n} \text{ for all } j > k \right\}$$

The set E_k^n contains all x in the domain for which $f_j(x)$ is already $1/n$-close to the value of the limit function $f(x)$ for all j bigger than k. Note that for fixed n, any x is eventually contained in some E_k^n by the assumption that $f_j \to f$

[3] Indeed, if we prove the result for this case, then for general f we change f on a set of measure 0 (say N) to $\tilde{f}$ such that $f_k \to \tilde{f}$ everywhere. Applying the result we obtain a closed set A_ϵ with $f_k \to \tilde{f}$ uniformly on A_ϵ and $m(E \setminus A_\epsilon) < \epsilon$. We then choose an open set $\mathcal{U}$ with $N \subset \mathcal{U}$ and $m(\mathcal{U}) \le \epsilon$. We now have that $f_k \to \tilde{f} = f$ uniformly on the closed set $\mathcal{A}_\epsilon \cap \mathcal{U}^c$ and also that $m(E \setminus (A_\epsilon \cap \mathcal{U})) \le m(E \setminus A_\epsilon) + m(\mathcal{U}) \le \epsilon + \epsilon = 2\epsilon$.

everywhere, so in particular $E = \bigcup_k E_k^n$ for any n. We also have $E_k^n \subset E_{k+1}^n$ and hence Proposition 2.3 applies, giving $\lim_{k\to\infty} m\left(E_k^n\right) = m\left(E\right)$. In view of $m\left(E_k^n\right) + m\left(E \setminus E_k^n\right) = m\left(E\right)$ we conclude that for any n there exists a k_n such that $m\left(E \setminus E_k^n\right) \leq \frac{1}{2^n}$ holds for all $k > k_n$. We now define

$$\tilde{A}_\epsilon = \bigcap_{n\geq N} E_{k_n}^n$$

The point is that with this definition, f_j is uniformly continuous on $\tilde{A}_\epsilon$. To see this, recall that what we have to show is given $\delta > 0$ there exists a J such that $|f_j\left(x\right) - f\left(x\right)| < \delta$ holds for all $j > J$ and all $x \in \tilde{A}_\epsilon$. Now indeed if we choose $n \geq N$ with $\frac{1}{n} < \delta$ we have for any $x \in \tilde{A}_\epsilon$ (hence $x \in E_{k_n}^n$ for any $n \geq N$) the inequality

$$|f_j\left(x\right) - f\left(x\right)| < \frac{1}{n} < \delta \text{ for all } j > k_n.$$

Note that k_n just depends on δ (and ϵ).

Finally, we claim that $m\left(E \setminus \tilde{A}_\epsilon\right) < \frac{\epsilon}{2}$. This simply follows from observing that

$$E \setminus \bigcap_{n\geq N} E_{k_n}^n = \bigcup_{n\geq N} \left(E \setminus E_{k_n}^n\right)$$

and using subadditivity of the measure together with our choice of N. To finish the proof, we choose a closed subset $A_\epsilon \subset \tilde{A}_\epsilon$ with $m\left(\tilde{A}_\epsilon \setminus A_\epsilon\right) < \frac{\epsilon}{2}$ (cf. Lemma 2.3). As a result we have $m\left(E \setminus A_\epsilon\right) \leq m\left(E \setminus \tilde{A}_\epsilon\right) + m\left(A_\epsilon \setminus \tilde{A}_\epsilon\right) < \epsilon$ and of course f_j converge uniformly on the subset $A_\epsilon \subset \tilde{A}_\epsilon$. $\square$

We are now ready to formulate and prove the second of Littlewood's principle.

Theorem 2.11 (Lusin)**.** *Let $f : E \to \mathbb{R}$ be measurable and finite valued, with $m\left(E\right) < \infty$. Then for any $\epsilon > 0$ there exists a closed set F_ϵ with $F_\epsilon \subset E$ and $m\left(E \setminus F_\epsilon\right) \leq \epsilon$ such that $f\big|_{F_\epsilon}$ is continuous.*

Remark 2.6. *Note that the theorem does not make the (stronger) statement that f is continuous on E at the points of F_ϵ. Again the example of (2) is useful here. That function is discontinuous at all points of $[0,1]$. What is F_ϵ?*

Proof. We use the result from Example Sheet 3 that every measurable function is the pointwise limit almost everywhere of a sequence of continuous functions together with Egoroff's Theorem.

Given f we find (f_n) continuous with $f_n \to f$ almost everywhere. Using Egoroff's Theorem we find a closed set $F_\epsilon \subset E$ where the convergence is uniform and such that $m\left(E \setminus F_\epsilon\right) \leq \epsilon$. But the limit of a uniformly convergent sequence of continuous functions is continuous proving that $f|_{F_\epsilon}$ $\square$

3 Integration Theory: The Lebesgue Integral

From now on all functions that appear are assumed to be measurable. Our goal is to define the Lebesgue integral of a measurable function. To achieve this, we proceed in stages. We first define the integral for simple function, the bounded functions supported on a set of finite measure, then non-negative functions and finally the general case.

3.1 Simple Functions

Given a simple function in canonical form (a_k distinct and non-zero, E_k disjoint)

$$\varphi(x) = \sum_{k=1}^{N} a_k \chi_{E_k}(x)$$

(recall by definition $m(E_k) < \infty$) we define the **Lebesgue integral** of φ as

$$\int_{\mathbb{R}^d} \varphi(x)\, dx := \sum_{k=1}^{N} a_k m(E_k) \ . \tag{8}$$

We can also define the integral of φ over a measurable subset $E \subset \mathbb{R}^d$ with $m(E) < \infty$ by observing that $\varphi(x) \cdot \chi_E(x)$ is still a simple function (why?) and setting

$$\int_E \varphi(x)\, dx := \int_{\mathbb{R}^d} \varphi(x) \cdot \chi_E(x)\, dx \, . \tag{9}$$

Note that with this definition we can already integrate the function (2) from the introduction since the rationals in $[0, 1]$ are measurable with measure zero.

Proposition 3.1. *Let φ, ψ be simple functions. The Lebesgue integral defined as in (8) satisfies*

1. *Independence of the representation: If $\varphi = \sum_{k=1}^{N} a_k \chi_{E_k}$ is any representation of φ (not necessarily canonical), then*
$$\int_{\mathbb{R}^d} \varphi(x)\, dx = \sum_{k=1}^{N} a_k m(E_k) \ .$$
2. *Linearity: For $\lambda, \mu \in \mathbb{R}$ we have*
$$\int (\lambda\varphi + \mu\psi) = \lambda \int \varphi + \mu \int \psi \, .$$
3. *Additivity: If E and F are disjoint subsets of finite measure, then*
$$\int_{E \cup F} \varphi = \int_E \varphi + \int_F \varphi \, .$$
4. *Monotonicity:*
$$\varphi \le \psi \quad \textit{on } \mathbb{R}^d \qquad \Longrightarrow \qquad \int_{\mathbb{R}^d} \varphi \le \int_{\mathbb{R}^d} \psi \, .$$

5. *Triangle inequality: If φ is simple, then so is $|\varphi|$ and*

$$\Big|\int \varphi\Big| \leq \int |\varphi| \, .$$

We have allowed ourselves to write the shorthand $\int$ instead of $\int_{\mathbb{R}^d}$ above.

Proof. The proof of (1) is a bit fiddly and is relegated to Example Sheet 4. For (2) we note that

$$\lambda\varphi + \mu\psi = \lambda \sum_{k=1}^{N} a_k \chi_{E_k} + \mu \sum_{\ell=1}^{M} b_\ell \chi_{F_\ell}$$

and by the independence of the representation proven in (1) we have

$$\int (\lambda\varphi + \mu\psi) = \lambda \sum_{k=1}^{N} a_k m(E_k) + \mu \sum_{\ell=1}^{M} b_k m(E_\ell) = \lambda \int \varphi + \mu \int \psi \, .$$

For (3) we simply note $\chi_{E\cup F} = \chi_E + \chi_F$ and use the linearity established in (2). For (4) we note that if $\eta \geq 0$ then the canonical form is everywhere non-negative and hence $\int \eta \geq 0$ by definition. Setting $\eta = \varphi - \psi$ and using linearity the result follows. For (5), we put φ in canonical form, $\varphi = \sum_{k=1}^{N} a_k \chi_{E_k}$, hence $|\varphi| = \sum_{k=1}^{N} |a_k| \chi_{E_k}$ is simple and observe

$$\Big|\int \varphi\Big| = \Big|\sum_{k=1}^{N} a_k m(E_k)\Big| \leq \sum_{k=1}^{N} |a_k| m(E_k) = \int |\varphi| \, .$$

□

Observation 3.1. *If $f = g$ almost everywhere for f, g simple, then $\int f = \int g$.*

Indeed, if $h = 0$ almost everywhere, then its canonical representation must look like $h = \sum_{k=1}^{N} a_k \chi_{E_k}$ with $m(E_k) = 0$, which implies $\int h = 0$.

3.2 Bounded functions on sets of finite measure

We now extend our definition of the Lebesgue integral from simple functions to bounded functions supported on sets of finite measure.

We define the support of $f : \mathbb{R}^d \to \mathbb{R}$ to be

$$\textit{supp } f := \{x \mid f(x) \neq 0\}$$

Note that the set *supp* f is measurable if f is measurable. We say that f is supported on E if $f(x) = 0$ whenever $x \notin E$.

Lemma 3.1. *Let f be a bounded function supported on a set E of finite measure. If (φ_n) is any sequence of simple functions bounded by M, supported on E, and such that $\varphi_n(x) \to f(x)$ for almost every x, then*

1. *The limit $\lim_{n\to\infty} \int \varphi_n$ exists*

2. *Moreover,*

$$f = 0 \textit{ almost everywhere} \quad \Longrightarrow \quad \lim_{n\to\infty} \int \varphi_n = 0.$$

Remark 3.1. *Note that by Theorem 2.8, given an f as in the Lemma, there always exists a (φ_n) satisfying the assumptions in the Lemma.*

Proof. Defining $I_n = \int \varphi_n$ we set out to prove that I_n is a Cauchy sequence. Let us fix $\epsilon > 0$ arbitrary. Given the sequence (φ_n) as in the Lemma, we first use Egoroff's Theorem to find a closed subset F of E where the convergence of φ_n is uniform and such that $m(E \setminus F) \leq \epsilon$. Given that the convergence is uniform on F, we can find for the prescribed $\epsilon > 0$ an N such that $|\varphi_m(x) - \varphi_n(x)| \leq \epsilon$ holds for all $m, n \geq N$ and all $x \in F$. But then for $m, n \geq N$ we have

$$|I_m - I_n| \leq \int_E |\varphi_m - \varphi_n| = \int_F |\varphi_m - \varphi_n| + \int_{E \setminus F} |\varphi_m - \varphi_n| \leq \epsilon m(F) + \epsilon 2M$$

where we used the triangle inequality for the second integral in the last step. Since $m(F) \leq m(E) < \infty$ and ϵ is arbitrary we conclude that I_m is Cauchy and (1) is proven. For (2) one simply repeats the above argument now showing that for any $\epsilon > 0$ one can find N such that $|I_n| \leq \epsilon m(E) + \epsilon M$ for $n \geq N$. ☐

Using the Lemma we can define the Lebesgue integral for $f : \mathbb{R}^d \to \mathbb{R}$ a bounded function supported on a set of finite measure:

Definition 3.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded function supported on a set E of finite measure. We define its Lebesgue integral as*

$$\int_{\mathbb{R}^d} f(x)\, dx := \lim_{n \to \infty} \int_{\mathbb{R}^d} \varphi_n(x)\, dx \tag{10}$$

where φ_n is any sequence of simple functions satisfying $|\varphi_n| \leq M$ for some constant M, each φ_n supported on E and $\varphi_n(x) \to f(x)$ for a.e. x as $n \to \infty$.

We already know that the limit exists but of course we need to show it is independent of the particular approximation (φ_n). Suppose therefore we have two sequences (φ_n) and (ψ_n) with the above properties. Then $(\varphi_n - \psi_n)$ converges to zero almost everywhere, is bounded by $2M$ and is supported on a set of finite measure. By part (2) of Lemma 3.1 we conclude $\lim \int_{\mathbb{R}^d} \varphi_n = \lim \int_{\mathbb{R}^d} \psi_n$ (the individual limits existing by part (1)).

As in (9), we define the Lebesgue integral of a bounded function with $m(supp(f)) < \infty$ over a measurable subset $E \subset \mathbb{R}^d$ of finite measure by

$$\int_E f(x)\, dx := \int_{\mathbb{R}^d} f(x) \cdot \chi_E(x)\, dx\,. \tag{11}$$

Proposition 3.2. *The Lebesgue integral on bounded functions supported on sets of finite measure defined in (10) and (11) is linear, additive, monotone and satisfies the triangle inequality (cf. Proposition 3.1 for the precise formulation)*

Proof. Follows by approximating with simple functions and taking limits. ☐

We can now prove our first convergence theorem for the Lebesgue integral which is a statement about interchanging the limit with the integral.

Theorem 3.1 (Bounded Convergence Theorem)**.** *Let (f_n) be a sequence of measurable functions $f_n : \mathbb{R}^d \to \mathbb{R}$ with*

- *$|f_n(x)| \leq M$ for all n,*

- *each f_n supported on E with $m(E) < \infty$,*
- *$f_n(x) \to f(x)$ for almost every x as $n \to \infty$.*

Then f is measurable, bounded a.e., supported on E for a.e. x and

$$\int_E |f_n - f| \to 0 \quad \text{as } n \to 0,$$

which by the triangle inequality immediately implies

$$\int_E f_n \to \int_E f \quad \text{as } n \to \infty. \tag{12}$$

Remark 3.2. *We will only prove the theorem with the additional assumption that $|f(x)| \leq M$ holds for* <u>*all*</u> *x (it is easy to see the assumptions already imply this a.e.). Otherwise $\int_E |f_n - f|$ appearing in the conclusion is not (yet) defined! In Section 3.3 we will define the integral for unbounded functions and also see that the integrand can always be changed on a set of measure 0 without affecting the integral. Hence the conclusion of the Theorem remains true in this case.*

Proof. The limiting function f is measurable combining Corollary 2.1 and the remark in Section 2.6.6. We also have from combining the first and the third item that $|f(x)| \leq M$ for a.e. x and combining the the second and the third that f is supported on E except for a set of measure 0. It remains to show the estimate. The proof is almost identical to the one of Lemma 3.1. Given $\epsilon > 0$ we find (by Egoroff's theorem) $F \subset E$ with $f_n \to f$ uniformly on F and $m(E \setminus F) \leq \epsilon$. By the uniformity on F we can find N such that for $n \geq N$ we have $|f_n(x) - f(x)| \leq \epsilon$ for all $n \geq N$. But then

$$\int_E |f_n - f| dx = \int_F |f_n - f| dx + \int_{E \setminus F} |f_n - f| \leq \epsilon m(E) + 2M m(E \setminus F)$$

holds for all $n \geq N$ which proves the claim. (Note that we have used Remark 3.2 here.) ☐

Remark 3.3. *Note that the conclusion of the theorem can be phrased as the interchange of the limit and the integral in this particular situation:*

$$\lim_{n \to \infty} \int_E f_n = \int_E \lim_{n \to \infty} f_n .$$

For the Riemann integral we needed uniform convergence to draw this conclusion. Here we obtain uniform convergence up to an arbitrary small set from Egoroff's theorem, which together with the boundedness *of the functions involved allows us to draw the above conclusion.*

Remark 3.4. *Note that boundedness is indeed essential as the example of $f_n(x) = n\chi_{\left(0, \frac{1}{n}\right]}$ shows.*

3.2.1 Riemann integrable functions are Lebesgue integrable

Using the bounded convergence theorem, we can show that a Riemann integrable function is measurable and that the Riemann integral agrees with the Lebesgue integral in this case.

Theorem 3.2. *Suppose f is Riemann integrable on the closed interval $[a,b]$. Then f is measurable and*

$$\int_{[a,b]}^{\mathcal{R}} f(x)\,dx = \int_{[a,b]}^{\mathcal{L}} f(x)\,dx\,.$$

Proof. Since f is Riemann integrable it is bounded, say $|f(x)| \leq M$. Moreover, there exists sequences of step functions (φ_k) and (ψ_k) with $|\varphi_k(x)| \leq M$ and $|\psi_k(x)| \leq M$ for all k and $x \in [a,b]$, such that

$$\varphi_1(x) \leq \varphi_2(x) \leq ... \leq f(x) \leq ... \leq \psi_2(x) \leq \psi_1(x)$$

(cf. Section 1.1, refining the partitions of the upper and lower Riemann sums) and with

$$\lim_{k\to\infty} \int_{[a,b]}^{\mathcal{R}} \varphi_k(x)\,dx = \int_{[a,b]}^{\mathcal{R}} f(x)\,dx = \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{R}} \psi_k(x)\,dx\,. \tag{13}$$

Now on step functions the Riemann and the Lebesgue integral agree by definition, so for all k we have

$$\int_{[a,b]}^{\mathcal{R}} \varphi_k(x)\,dx = \int_{[a,b]}^{\mathcal{L}} \varphi_k(x)\,dx \quad , \quad \int_{[a,b]}^{\mathcal{R}} \psi_k(x)\,dx = \int_{[a,b]}^{\mathcal{L}} \psi_k(x)\,dx\,.$$

We now define pointwise $\tilde{\varphi}(x) = \lim_{k\to\infty} \varphi_k(x)$ and $\tilde{\psi}(x) = \lim_{k\to\infty} \psi_k(x)$, these limits existing by monotonicity and boundedness of $f(x)$. The functions $\tilde{\varphi}$ and $\tilde{\psi}$ are measurable (Corollary 2.1), bounded and supported on $[a,b]$. Hence the Bounded Convergence Theorem yields

$$\lim_{k\to\infty} \int_{[a,b]}^{\mathcal{L}} \varphi_k(x)\,dx = \int_{[a,b]}^{\mathcal{L}} \tilde{\varphi}(x)\,dx \quad \text{and} \quad \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{L}} \psi_k(x)\,dx = \int_{[a,b]}^{\mathcal{L}} \tilde{\psi}(x)\,dx.$$

Combining this with the previous equalities we easily obtain

$$\int_{[a,b]}^{\mathcal{L}} \left(\tilde{\varphi}(x) - \tilde{\psi}(x)\right) dx = 0 \quad \text{where also } \tilde{\varphi} - \tilde{\psi} \geq 0$$

from which we claim we can conclude $\tilde{\varphi} = \tilde{\psi}$ almost everywhere (Exercise, or see item (6) of Proposition 3.3 below). But then we must have $\tilde{\varphi}(x) \leq f(x) \leq \tilde{\psi}(x) = \tilde{\varphi}(x)$ almost everywhere and hence $f = \tilde{\varphi}$ and $f = \tilde{\psi}$ almost everywhere. Finally, since $\varphi_k \to f$ a.e. by construction we have by definition of the Lebesgue integral

$$\int_{[a,b]}^{\mathcal{L}} f(x)\,dx = \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{L}} \varphi_k(x)\,dx = \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{R}} \varphi_k(x)\,dx = \lim_{k\to\infty} \int_{[a,b]}^{\mathcal{R}} f(x)\,dx$$

□

3.3 Non-negative functions

We proceed further in enlarging our class of functions that we can integrate. We now consider measurable *extended valued* functions

$$f : \mathbb{R}^d \to [0, \infty]$$

Not only are these functions potentially unbounded, they can also take the values $\pm\infty$ on a measurable set and they may also be supported on a set of infinite measure, for instance all of $\mathbb{R}^d$. We define the **extended Lebesgue integral** for these functions by

$$\int_{\mathbb{R}^d} f(x)\, dx := \sup_g \int g(x)\, dx \tag{14}$$

where we take the supremum over all measurable functions g such that $0 \leq g \leq f$, such that g is bounded and supported on a set of finite measure.

Now the expression on the left is either finite or infinite. In the first case we shall say that f is **Lebesgue integrable**. As usual we define for $E \subset \mathbb{R}^d$ measurable the integral

$$\int_E f(x)\, dx := \int_{\mathbb{R}^d} f(x)\, \chi_E(x)\, dx \tag{15}$$

Example 3.1. *The function $F_a(x) = \left(1 + |x|^2\right)^{-a/2}$ is integrable for $a > d$.*

The extended Lebesgue integral has the familiar properties of the integral:

Proposition 3.3. *The integral defined in (14) and (15) satisfies the following:*

1. *Linearity: If $f, g \geq 0$ and $\lambda, \mu \in \mathbb{R}$ are both positive, then*
$$\int (\lambda f + \mu g) = \lambda \int f + \mu \int g$$
2. *Additivity: If E and F are disjoint subsets of $\mathbb{R}^d$ and $f \geq 0$, then*
$$\int_{E \cup F} f = \int_E f + \int_F f$$
3. *Monotonicity: If $0 \leq f \leq g$ then $\int f \leq \int g$.*
4. *If g is integrable and $0 \leq f \leq g$ then f is integrable.*
5. *If f is integrable then $f(x) < \infty$ for a.e. x.*
6. *If $\int f = 0$, then $f(x) = 0$ for a.e. x.*

Proof. The third is immediate from the definition and so is the fourth. For the first, note that it suffices to show the statement with $\lambda = \mu = 1$ since $\int \lambda f = \lambda \int f$ for $\lambda > 0$ follows again immediately from the definition. To show it for $\lambda = \mu = 1$ we first show the $\geq$ direction. Let $\varphi_1 \leq f$ and $\varphi_2 \leq g$ be two bounded functions supported on a set of finite measure. Then we have $\varphi_1 + \varphi_2 \leq f + g$ and hence

$$\int f + g \geq \int (\varphi_1 + \varphi_2) = \int \varphi_1 + \int \varphi_2$$

and taking the sup over all φ_1 and φ_2 we obtain the first direction. For the second, we let $\eta \geq 0$ be bounded, supported on a set of finite measure with $\eta \leq f + g$. We then define $\eta_1(x) = \min(f(x), \eta(x))$ and $\eta_2 = \eta - \eta_1$. We note $\eta_1 \leq f$ and $\eta_2 \leq g$ and hence

$$\int \eta = \int (\eta_1 + \eta_2) = \int \eta_1 + \int \eta_2 \leq \int f + \int g\,.$$

Taking the sup over all η we obtain the $\leq$ direction and item 1 is proven. The second item then immediately follows using the linearity.

It remains to prove (5) and (6). For (5) we let $E_k = \{x \mid f(x) \geq k\}$ and $E_\infty = \{x \mid f(x) = \infty\}$. We then have $E_k \supset E_{k+1}$ and $E_\infty = \bigcap_k E_k$. The integrability tells us that $\infty > \int f \geq \int f\chi_{E_k} \geq k \cdot m(E_k)$ and hence $m(E_k) \to 0$ as $k \to \infty$. Combining this with Proposition 2.3 we obtain $m(E_\infty) = 0$.

For (6) we let $E_k = \{x \mid f(x) > \frac{1}{k}\}$. Since $\{x \mid f(x) > 0\} = \bigcup_{k=1}^{\infty} E_k$ it suffices to show that $m(E_k) = 0$ for all k to conclude $m(\{x \mid f(x) > 0\}) = 0$. We can infer this statement from $0 = \int f \geq \int f\chi_{E_k} \geq \frac{1}{k} m(E_k)$ for all $k \geq 1$. □

Note that the converse of (6) in also true. If $\eta \geq 0$ and $\eta = 0$ a.e. then $\int \eta = \sup_g \int g(x)\,dx = 0$. (Otherwise there would exist a bounded function supported on a set of measure zero with non-zero integral – a contradiction.) This shows that the extra assumption in Remark 3.2 is unnecessary for the conclusion of Theorem 3.1 to be valid.

Next we will try to prove convergence results for the extended Lebesgue integral. In particular, we can revisit the example of Remark 3.4 and ask about a general statement regarding the exchange of the limit and the integral if the functions under consideration are not uniformly bounded.

Lemma 3.2 (Fatou's Lemma)**.** *Let (f_n) be a sequence of measurable functions with $f_n \geq 0$. If $\lim_{n\to\infty} f_n(x) = f(x)$ for almost every x, then*

$$\int f \leq \liminf_{n\to\infty} \int f_n\,.$$

Note that both the left hand side and the right hand side may be $+\infty$.

Proof. We first note that it suffices to prove

$$\int g \leq \liminf_{n\to\infty} \int f_n$$

for any g with $0 \leq g \leq f$ bounded and supported on a set of finite measure because we can then take the supremum on the left to obtain the result. This observation allows us to use the Bounded Convergence Theorem: We first define $g_n(x) := \min(g(x), f_n(x))$ and observe that $g_n \geq 0$ is measurable, bounded and supported on a set of finite measure. Moreover, $g_n(x) \to g(x)$ for a.e. x as follows from $g_n(x) - g(x) = \min(0, f_n(x) - f(x) + f(x) - g(x))$. Monotonicity of the integral and the bounded convergence theorem yields

$$\int f_n \geq \int g_n \to \int g\,,$$

so taking the lim inf on the left already produces the desired result. □

For a general sequence of measurable non-negative functions the inequality given by Fatou's Lemma is the best one can do. However, under additional assumption the interchange of the limit and the integral (and hence equality) can be inferred. The following theorem provides such a setting and is one of the cornerstones of the Lebesgue-theory of integration. It will be used many many times in the sequel so it is important to know this statement well!

Theorem 3.3 (Monotone Convergence Theorem, MCT). *Let (f_n) be a sequence of (extended real-valued) measurable functions with $f_n \geq 0$ and $f_n \to f$ a.e. Suppose in addition $f_n(x) \leq f_{n+1}(x)$ holds a.e. in x for any n. Then*

$$\lim_{n\to\infty} \int f_n = \int f \,. \tag{16}$$

Proof. We have $\int f_n \leq \int f$ for any n by monotonicity of the integral. Hence

$$\limsup_{n\to\infty} \int f_n \leq \int f \leq \liminf_{n\to\infty} \int f_n$$

with the second inequality being Fatou's Lemma. ☐

Remark 3.5. *We write the shorthand $f_n \nearrow f$ a.e. if $f_n \to f$ a.e. and $f_n(x) \leq f_{n+1}(x)$ a.e. in x for any n.*

Remark 3.6. *Note that both sides of the equality in (16) can be $+\infty$ that is (16) holds in the extended sense.*

Remark 3.7. *Note that the MCT provides a useful tool to compute $\int f$ via simple functions and the approximation of Theorem 2.7.*

Before we turn to the other cornerstone of the Lebesgue theory, the dominated convergence theorem (DCT), we enlarge our class of functions for the integral one more time.

3.4 The General case and the notion of integrable

Let $f : \mathbb{R}^d \to \mathbb{R}$ be measurable.[4]

Definition 3.2. *We say that f above is integrable if $|f|$ is Lebesgue integrable in the sense of the previous section. For such f we define*

$$\int f = \int f^+ - \int f^-$$

where $f^+ = \max(f(x), 0)$ and $f^- = \max(-f(x), 0)$ are the positive and negative part of f respectively.

Remark 3.8. *Note that $f = f^+ - f^-$ and $|f| = f^+ + f^-$ and hence $f^\pm \leq |f|$, so $f^\pm$ are indeed integrable. Note also that given two different decompositions of f into a difference of non-negative functions, i.e. $f = f^+ - f^- = g^+ - g^-$, one has $f^+ + g^- = f^- + g^+$ and hence $\int f^+ + \int g^- = \int f^- + \int g^+$ by linearity of the integral on non-negative functions. It follows that $\int f^+ - \int f^- = \int g^+ - \int g^-$ and hence that the value of the integral of f is independent of the decomposition.*

[4] We could allow $\overline{\mathbb{R}}$ here but we will see below that the notion of integrability requires the set of points where $f(x) = \pm\infty$ to have measure zero.

Remark 3.9. *We can always modify f on a set of measure zero without affecting the integrability of f or the value of the integral (cf. the comments after the proof of Proposition 3.3). Therefore, we may adopt the convention that a function f can be undefined on a set of measure zero. Cf. also Footnote 4.*

Proposition 3.4. *The integral defined above is linear, additive, monotone and satisfies the triangle inequality.*

Proof. Exercise. □

We next prove two interesting regularity properties of integrable functions:

Proposition 3.5. *Suppose f is integrable on $\mathbb{R}^d$. Then for every $\epsilon > 0$*

1. *There exists a set of finite measure B (a large ball, for instance) such that*

$$\int_{B^c} |f| < \epsilon \qquad \text{“vanishing at infinity”}$$

2. *There is a $\delta > 0$ such that*

$$\int_E |f| < \epsilon \quad \text{for all } E \text{ with } m(E) < \delta \qquad \text{“absolute continuity”}$$

The first statement expresses the intuitive fact that the function f has to go to zero at infinity in a suitable sense in order to be integrable. The second property says that for fixed f, if one integrates over sufficiently small sets, the integral is small as well. The name absolute continuity will become clearer to us later (see Section 4.2.5 and also Example Sheet 5).

Proof. Wlog $f \geq 0$ since otherwise we look at $|f|$.

For the first statement let B_n be the ball of radius n centred at the origin and $f_n(x) = f(x)\chi_{B_n}(x)$. Note that $f_n \nearrow f$ and hence by the MCT (Theorem 3.3), $\int f_n \to \int f < \infty$. This means there exists an N such that $\int f - \int f_N < \epsilon$ which is equivalent to $\int f(x)\chi_{(B_N)^c} < \epsilon$ and hence the desired result.

For the second statement we set $E_n = \{x \mid f(x) \leq n\}$ and $f_n(x) := f(x)\chi_{E_n}(x)$. Noting that $f_n \nearrow f$ and hence by the MCT $\int f_n \to \int f$, we conclude the existence of an N with $\int f - \int f_N < \frac{\epsilon}{2}$. But this means that

$$\int_E f = \int_E f - f_N + \int_E f_N < \frac{\epsilon}{2} + m(E)\, N\, .$$

Now if we choose $\delta < \frac{\epsilon}{2N}$, then $m(E) < \delta$ implies that $\int_E f < \epsilon$ as desired. □

We are ready to prove the other cornerstone of the Lebesgue Theory, the dominated convergence theorem (DCT):

Theorem 3.4 (Dominated Convergence Theorem (DCT))**.** *Let (f_n) be a sequence of measurable functions with $f_n \to f$ a.e. If $|f_n(x)| \leq g(x)$ where g is integrable, then f is integrable and*

$$\int |f_n - f| \to 0 \quad \text{as } n \to \infty.$$

Hence by the triangle inequality

$$\lim_{n\to\infty} \int f_n = \int f.$$

Proof. We provide two proofs. One is via Fatou's Lemma. Starting from $|f_n - f| \leq 2g$ a.e., which holds by the triangle inequality, we apply Fatou's Lemma to the sequence of non-negative (after a change on a set of measure zero) functions $2g - |f - f_n|$ to obtain

$$\int 2g \leq \int 2g + \liminf_{n \to \infty} \left(-|f_n - f| \right) .$$

Since $\int g$ is finite by assumption we obtain $\limsup_{n\to\infty} \int |f_n - f| \leq 0$ which proves the result.

The other proof uses Proposition 3.5. Given $\epsilon > 0$ we first choose a large ball B_M such that $\int_{(B_M)^c} g < \epsilon$ by (1) of Proposition 3.5. We next invoke Egoroff's theorem to choose $X \subset B_M$ with $m(X) < \delta$ such that $f_n \to f$ uniformly on $B_M \setminus X$. Here $\delta > 0$ is chosen as in (2) of Proposition 3.5, i.e. in particular so that $\int_X g < \epsilon$. Using the uniform convergence on $B_M \setminus X$ we choose N large such that $|f_n(x) - f(x)| < \frac{\epsilon}{m(B_M)}$ holds for all $n \geq N$ and all $x \in B_M \setminus X$. Combining everything we obtain for $n \geq N$

$$\int |f_n - f| = \int_{(B_M)^c} |f_n - f| + \int_{B_M \setminus X} |f_n - f| + \int_X |f_n - f|$$

$$\int |f_n - f| \leq \int_{(B_M)^c} 2g + \epsilon + \int_X 2g \leq 2\epsilon + \epsilon + 2\epsilon < 5\epsilon$$

which is what we needed to prove. □

3.5 Aside: Complex-valued functions

So far we have considered real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$. The theory easily extends to complex-valued functions $f : \mathbb{R}^d \to \mathbb{C}$. Indeed, we say that $f(x) = u(x) + iv(x)$ is Lebesgue integrable if $|f(x)| = \sqrt{u(x)^2 + v(x)^2}$ is integrable. It is easy to see that f is Lebesgue integrable if and only if both u and v are integrable. We define the Lebesgue integral as

$$\int f(x)\, dx := \int u(x)\, dx + i \int v(x)\, dx$$

As considering complex valued functions does not really add anything conceptually new to the theory we will continue to develop it for real valued functions.

3.6 The space of integrable functions as a normed vector space

The class of integrable functions form a vector space, which we shall denote $L^1(\mathbb{R}^d)$. It is equipped with a (semi)norm

$$\|f\|_{L^1(\mathbb{R}^d)} := \int_{\mathbb{R}^d} |f(x)| dx$$

It is a semi-norm precisely because $\|f\|_{L^1(\mathbb{R}^d)} = 0$ only implies $f = 0$ a.e. However, it is not hard to show the following:

Exercise 3.1. *Show that $L^1(\mathbb{R}^d)$ is a normed vectorspace if we define the elements to be equivalence classes of functions agreeing almost everywhere.*

Recall that a norm induces a metric, here $d(x,y) = \|x-y\|_{L^1(\mathbb{R}^d)}$, which makes L^1 a metric space. Recall also that a metric space is called complete if every Cauchy sequence (x_n) in X converges to a limit $x \in X$, i.e. $d(x_n, x) \to 0$ as $n \to \infty$.

Theorem 3.5 (Riesz-Fischer)**.** *The vectorspace $L^1(\mathbb{R}^d)$ is complete with respect to the metric induced by the norm.*

Remark 3.10. *The theorem expresses how the Lebesgue integral can be understood as the completion of the Riemann integral. See also .... below.*

Proof. Let (f_n) be a Cauchy sequence. We first find a candidate limit f with $f \in L^1$ and then prove $f_n \to f$ in the L^1-norm.

The key idea is to extract from (f_n) a subsequence which converges *pointwise* to some f (using the completeness of $\mathbb{R}$).[5] To do this, we first construct a subsequence (f_{n_j}) with

$$\|f_{n_{j+1}} - f_{n_j}\|_{L^1} \leq \frac{1}{2^j} \quad \text{for all } j \geq 1. \tag{17}$$

We the define

$$g_K(x) = |f_{n_1}(x)| + \sum_{j=1}^{k} |f_{n_{j+1}}(x) - f_{n_j}(x)|$$

$$g(x) = |f_{n_1}(x)| + \sum_{j=1}^{\infty} |f_{n_{j+1}}(x) - f_{n_j}(x)|$$

Note that g makes sense as an extended valued function. Since $g_K \nearrow g$ almost everywhere as $k \to \infty$ we have that g is measurable and by the MCT that

$$\int g = \lim_{k\to\infty} \int g_k = \int |f_{n_1}(x)| + \lim_{k\to\infty} \sum_{j=1}^{k} \|f_{n_{j+1}}(x) - f_{n_j}(x)\|_{L^1} < \infty$$

the last inequality following from the property (17) and that $f_{n_1} \in L^1$. We conclude that g is integrable and hence $g(x) < \infty$ for a.e. x. This implies that the sum converges absolutely for a.e. x which means that the right hand side of

$$f_{n_{j+1}}(x) = f_{n_1}(x) + \sum_{j=1}^{k} f_{n_{j+1}}(x) - f_{n_j}(x)$$

converges for a.e. x, hence so does the left hand side. We conclude that the subsequence $f_{n_{j+1}}$ converges pointwise for a.e. x to some limiting function which we call f. Since $|f(x)| \leq g(x)$ for a.e. x we conclude that f is integrable, so $f \in L^1(\mathbb{R}^d)$. It remains to show that $f_{n_j} \to f$ in L^1. But this is immediate from $|f(x) - f_{n_k}(x)| \leq 2g$ for a.e. x and the dominant convergence theorem.

[5] Going to a subsequence is necessary as we can have $\|f_n - f\|_{L^1} \to 0$ for some (f_n) and f such that $f_n(x) \to f(x)$ for no x! See Example Sheet 6.

To finish the proof we recall that if a subsequence of a Cauchy sequence converges to a limit f, then so must the entire sequence.[6] Hence $f_n \to f$ in L^1 and the theorem is proven. $\square$

Corollary 3.1. *Let $f_n \to f$ in L^1. Then there exists a subsequence (f_{n_k}) with $f_{n_k} \to f$ a.e. pointwise.*

Proof. The assumption implies that (f_n) is Cauchy in L^1. Then repeat the construction in the proof of Theorem 3.5. $\square$

3.7 Dense families in $L^1\left(\mathbb{R}^d\right)$

We next consider certain families of simple (both in the colloquial and the precise sense) functions which are dense in L^1. Recall the definition of dense:

Definition 3.3. *A family $\mathcal{G}$ of integrable functions is dense in L^1 if for any $f \in L^1$ we can find a $g \in \mathcal{G}$ with $\|f-g\|_{L^1} < \epsilon$.*

Why are dense families useful? In a typical application one wants to establish an identity for integrable functions which involves the L^1-norm. To prove the identity, it may be simpler to prove it for a dense family of functions in L^1 because a (say) continuous function is much easier to manipulate than a general element of L^1. Finally, a density argument allows one to extend the identity to all L^1-functions. Example Sheet 5 provides an example.

Theorem 3.6. *The following families of functions are dense in $L^1\left(\mathbb{R}^d\right)$*

1. *simple functions*
2. *step functions*
3. *continuous functions of compact support*

Proof. Exercise. Outline: For the first note that one may assume $f \geq 0$ as one can approximate separately for f^+ and f^-. Then approximate f with (φ_k) an increasing sequence of simple functions converging pointwise to f and apply the MCT to show convergence in L^1. For the second part it suffices to approximate the characteristic function of a set of finite measure by a step function (why). For this Problem 3 from Example Sheet 2 will be handy. Finally for the third conclusion one needs to smooth the edges of a step function. $\square$

3.8 Fubini's Theorem

We now turn to an important analytical tool in the integration theory. It allows one to convert a d-dimensional integral into a d_1-dimensional and a d_2-dimensional one ($d_1 + d_2 = d$).

To see that this is not entirely trivial, let us start by trying to integrate the function $f(x,y) = \frac{x^2-y^2}{(x^2+y^2)^2}$ over the unit square $[0,1]\times[0,1]$. Naively we might do this in two ways. First integrate in y and then in x

$$\int_{[0,1]\times[0,1]} \frac{x^2-y^2}{(x^2+y^2)^2} dydx =^? \int_0^1 dx \int_0^1 dy \frac{x^2-y^2}{(x^2+y^2)^2} = \frac{\pi}{4},$$

[6] Let $\epsilon > 0$ be given. Choose N such that $\|f_n - f_m\| < \frac{\epsilon}{2}$ for $m, n \geq N$. Choose $n_k \geq N$ such that $\|f - f_{n_k}\| < \frac{\epsilon}{2}$, so in particular $\|f_n - f_{n_k}\| < \frac{\epsilon}{2}$ for all $n \geq N$. The triangle inequality now implies $\|f_n - f\| < \epsilon$ for all $n \geq N$ as desired.

or first in x and then in y

$$\int_{[0,1]\times[0,1]} \frac{x^2-y^2}{(x^2+y^2)^2} dydx =^? \int_0^1 dy \int_0^1 dx \frac{x^2-y^2}{(x^2+y^2)^2} = -\frac{\pi}{4}\,.$$

The fact that the result depends on the order in which the integration is carried out tells us that some care is needed to state assumptions when a d-dimensional integral can be computed in terms of iterated ones.

3.8.1 Slices of measurable sets and functions

We begin by setting up some notation.

- We let $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ with $d = d_1 + d_2$ and $d_1, d_2 \geq 1$.
- For $E \subset \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$ a subset we define the slices of E

  $$E_x = \{y \in \mathbb{R}^{d_2} \mid (x,y) \in E\} \quad \text{and} \quad E_y = \{x \in \mathbb{R}^{d_1} \mid (x,y) \in E\}$$

- For a measurable function $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ we define

  the slice corresponding to $y \in \mathbb{R}^{d_2}$ as the function $f^y(x) := f(x,y)$ with y fixed

  the slice corresponding to $x \in \mathbb{R}^{d_1}$ as the function $f^x(y) := f(x,y)$ with x fixed

The big question now is:

- If E is a measurable set, are E_x and E_y also measurable?
- If f is a measurable function, are the slices f^x and f^y also measurable?

It is not hard to see that the answer is generally no. Take $\mathbb{R}^2$ and the set $E := \mathcal{N} \times \{0\} \subset \mathbb{R} \times \{0\} \subset \mathbb{R}^2$, with $\mathcal{N}$ a non-measurable set in $\mathbb{R}$. As a subset of a measure zero set in $\mathbb{R}^2$, E is measurable. However, the slice $E^{y=0} = \mathcal{N}$ is not measurable in $\mathbb{R}$.

What rescues as is that *measurability holds for almost every slice.*

Let us first state the two fundamental theorems of this section and then discuss their content in a sequence of remarks.

3.8.2 Statement and Discussion of Fubini's and Tonelli's Theorem

Theorem 3.7 (Fubini)**.** *Let $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$ be integrable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then for almost every $y \in \mathbb{R}^{d_2}$*

1. *The slice f^y is integrable in $\mathbb{R}^{d_1}$*
2. *The function defined by $y \mapsto \int_{\mathbb{R}^{d_1}} f^y(x)\,dx$ is integrable in $\mathbb{R}^{d_2}$.*

Moreover, the integral of f can be computed iteratively

3. $$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f(x,y)\,dx = \int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f^y(x)\,dx = \int_{\mathbb{R}^d} f\,.$$

Remark 3.11. *Recall by definition all integrable functions are in particular measurable.*

Theorem 3.8 (Tonelli)**.** *Let $f : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to [0, \infty]$ be measurable and non-negative on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. Then for almost every $y \in \mathbb{R}^{d_2}$*

1. *The slice f^y is measurable on $\mathbb{R}^{d_1}$*

2. *The function defined by $y \mapsto \int_{\mathbb{R}^{d_1}} f^y(x)\, dx$ is measurable on $\mathbb{R}^{d_2}$.*

Moreover, the integral of f can be computed iteratively

3. $$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f(x, y)\, dx = \int_{\mathbb{R}^d} f \qquad \text{in the extended sense.}$$

Remark 3.12. *Both theorems are symmetric in x and y, i.e. the conclusion (in say Fubini) is also that f^x is integrable a.e. in $\mathbb{R}^{d_2}$, that $x \mapsto \int_{\mathbb{R}^{d_2}} f^x(y)\, dy$ is integrable in $\mathbb{R}^{d_1}$ and that*

$$\int_{\mathbb{R}^{d_1}} dx \int_{\mathbb{R}^{d_2}} f(x, y)\, dy = \int_{\mathbb{R}^d} f\,.$$

Remark 3.13. *Note that the function in (2) of Fubini, $y \mapsto \int_{\mathbb{R}^{d_1}} f^y(x)\, dx$, is defined for almost every y. This is consistent with our earlier convention that an integrable function can be undefined on a set of measure zero, cf. Remark 3.9. Similarly for (2) in Tonelli's Theorem, where $y \mapsto \int_{\mathbb{R}^{d_1}} f^y(x)\, dx$ is a measurable function on $\mathbb{R}^{d_2}$ minus a set of measure 0 and hence agrees a.e. with a measurable function on $\mathbb{R}^{d_2}$.*

Remark 3.14. *Note that in Fubini's Theorem we are assuming that f is integrable. In Tonelli's theorem this is not assumed and in particular both sides of (3) in Tonelli's Theorem can be infinite. The point is that if that is the case, both, the iterated integrals and the d-dimensional one have to yield $+\infty$! This provides a useful strategy to compute $\int_{\mathbb{R}^d} f$ for an arbitrary measurable function: First compute $\int_{\mathbb{R}^d} |f|$. To compute this, one can by Tonelli's theorem use ANY convenient iterated integration:*

- *If one of them yields $+\infty$ then all of them have to and we can conclude by (3) of Tonelli that f is not integrable.*

- *If one of them yields a number smaller than $+\infty$, then any iteration of integrals has to yield that number and (3) of Tonelli implies that f is integrable. Now the assumptions of Fubini's theorem hold for this f and we are allowed to compute $\int_{\mathbb{R}^d} f$ using any version of iterated integrals.*

Examples of this will be seen on Example Sheet 6.

Remark 3.15. *Revisiting the example in the beginning we conclude that this f cannot be integrable over the unit square. (Exercise: Show this directly.)*

Remark 3.16. *Even if both of the iterated integrals exist and agree one cannot infer that f in integrable over $\mathbb{R}^d$. Try the function defined by $f(x, y) = \frac{xy}{(x^2+y^2)^2}$ for $x^2 + y^2 \neq 0$ and $f(x, y) = 0$ for $x = y = 0$. [The proof is easiest in polar coordinates which we have not introduced rigorously yet but give you an immediate intuition of how things fail here.]*

3.8.3 The proof of Tonelli's Theorem (using Fubini)

Since we want to apply Fubini's theorem, we start with the truncation

$$f_k(x,y) = \begin{cases} f(x,y) & \text{if } |(x,y)| \leq k \text{ and } f(x,y) < k \\ 0 & \text{otherwise} \end{cases}$$

Clearly each f_k is measurable and in fact integrable. We clearly have $f_k \nearrow f$ and by the MCT also the limit (in the extended sense)

$$\int_{\mathbb{R}^d} f_k(x,y) \to \int_{\mathbb{R}^d} f(x,y) \ . \tag{18}$$

By Fubini's theorem applied to f_k, there exists an E_k with $m(E_k) = 0$ such that f_k^y is integrable (in particular measurable) for all $y \in (E_k)^c$. Set $E = \bigcup_k E_k$. Then $m(E) = 0$ and f_k^y is integrable (in particular measurable) for all $y \in E^c$ and *all* k. Since $f_k^y \nearrow f^y$, the function f^y (being the limit of a sequence of measurable functions) is also measurable, proving (1) of Tonelli. Further, the MCT implies that

$$\int_{\mathbb{R}^{d_1}} f_k^y = \int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx \nearrow \int_{\mathbb{R}^{d_1}} f(x,y)\,dx = \int_{\mathbb{R}^{d_1}} f^y \qquad \text{for } y \in E^c. \tag{19}$$

Note that the right hand side may well be $+\infty$ for some $y \in E^c$!

Applying Fubini again, we know that $y \mapsto \int_{\mathbb{R}^{d_1}} f_k(x,y)\,dx$ is a sequence of integrable (hence measurable) functions defined almost everywhere on $\mathbb{R}^{d_2}$. By (19) this sequence increases to the function $\int_{\mathbb{R}^{d_1}} f(x,y)$, hence the latter is a measurable function, proving (2) of Tonelli.

By the remarks in the previous paragraph, we can apply the MCT again to (19) obtaining

$$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} dx \ f_k(x,y) \to \int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} dx \ f(x,y) \ . \tag{20}$$

Finally, we also know, by part (3) of Fubini applied to f_k, the equality

$$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} dx \ f_k(x,y) = \int_{\mathbb{R}^d} f_k(x,y) \ . \tag{21}$$

Combining (20), (21) and (18) yields the statement (3) of Tonelli's theorem.

3.8.4 The proof of Fubini's Theorem

We let $\mathcal{F} \subset L^1(\mathbb{R}^d)$ be the set of integrable functions satisfying all three conclusions of Fubini's theorem and prove $L^1(\mathbb{R}^d) \subset \mathcal{F}$. The proof has four steps:

(1) Prove that $\mathcal{F}$ is closed under finite linear combinations

(2) Prove that $\mathcal{F}$ is closed under limits

(3) Prove that $f = \chi_E$ with E a measurable set of finite measure is in $\mathcal{F}$. This will be proven along the following lines:

   (a) Prove it for E an open cube.

(b) Prove it for E the boundary of a closed cube.

(c) Prove it for E a finite union of closed cubes.

(d) Prove it for E open and of finite measure

(e) Prove it for E a G_δ ofr finite measure

(f) Prove it for E having measure zero.

(g) Prove it for E an arbitrary finite measure set.

(4) Conclude that any $f \in L^1\left(\mathbb{R}^d\right)$ is in $\mathcal{F}$ by approximating f with simple functions and using the previous steps.

Step 1. Let $(f_n)_{k=1}^N \subset \mathcal{F}$. Then for each k we have a set A_k with $m\left(A_k\right)=0$ and f_n^y being integrable on $\mathbb{R}^{d_1}$ for all $y \in (A_k)^c$. Defining $A=\bigcup_{k=1}^N A_k$ we have $m\left(A\right)=0$ and that f_n^y is measurable and integrable for any n and any $y \in A^c$. Clearly any linear combination of the f_n^y satisfies the same statement proving conclusion (1). Moreover the linearity of the integral immediately implies conclusion (2) and (3).

Step 2. We prove that if (f_k) is a sequence in $\mathcal{F}$ with $f_k \nearrow f$ for some $f \in L^1\left(\mathbb{R}^d\right)$, then $f \in \mathcal{F}$. Note this immediately implies the same statement for $f_k \searrow f$ since in this case $-f_k \nearrow -f$.

To prove this statement, note that we can restrict ourselves to $f_k \geq 0$ as otherwise we can consider $f_k - f_1 \geq 0$. We then have immediately

$$\lim_{k\to\infty} \int_{\mathbb{R}^d} f_k\left(x,y\right) dxdy = \int_{\mathbb{R}^d} f\left(x,y\right) \tag{22}$$

from the MCT. Furthermore, by the assumption that $f_k \in \mathcal{F}$, we have for each k a set A_k with $m\left(A_k\right)=0$ and f_k^y being (measurable and) integrable on $\mathbb{R}^{d_1}$ for $y \in (A_k)^c$. Setting as usual $A=\bigcup_{k=1}^\infty A_k$ we have $m\left(A\right)=0$ and f_k^y being integrable for all k and all $y \in A^c$.

Now from the fact that $f_k^y \nearrow f^y$, it follows that f^y is measurable and the MCT produces

$$\int_{\mathbb{R}^{d_1}} f_k^y\left(x\right) dx \nearrow \int_{\mathbb{R}^{d_1}} f^y\left(x\right) dx \tag{23}$$

with the left hand side being integrable (hence measurable) by assumption. It follows that the right hand side is measurable and applying the MCT again yields

$$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f_k^y\left(x\right) dx \to \int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f^y\left(x\right) dx\,. \tag{24}$$

By (3) of Fubini applied to $f_k \in \mathcal{F}$ we know that the left hand side satisfies

$$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f_k^y\left(x\right) dx = \int_{\mathbb{R}^d} f_k\left(x,y\right) \to \int_{\mathbb{R}^d} f\left(x,y\right) \tag{25}$$

with the second (limit) statement being simply (22) from above. Combining (24) and (25) yields the conclusion (3) of Fubini for f, namely

$$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} f^y\left(x\right) dx = \int_{\mathbb{R}^d} f\left(x,y\right) < \infty\,. \tag{26}$$

Here the $< \infty$ follows from the assumption that $f \in L^1$. We now see that (26) implies that the measurable function $\int_{\mathbb{R}^{d_1}} f^y(x)$ is integrable for almost every y (which is conclusion (2) of Fubini) and from $\int_{\mathbb{R}^{d_1}} f^y(x) < \infty$ for almost every y we conclude that f^y is integrable (which is conclusion (1) of Fubini).

Step 3.

(a) Let E a bounded open cube in $\mathbb{R}^d$, $E = Q_1 \times Q_2$ with Q_i and open cube in $\mathbb{R}^{d_i}$. For each fixed y, the characteristic function $\chi_E(x,y)$ is measurable in x and integrable with (recall the notation (4))

$$g(y) = \int_{\mathbb{R}^{d_1}} \chi_E(x,y)\,dx = |Q_1|\chi_{Q_2}$$

as it gives the volume of Q_1 if $y \in Q_2$ and zero otherwise (draw a picture!). Now $g(y)$ is measurable and integrable with

$$\int_{\mathbb{R}^{d_2}} g(y)\,dy = |Q_1||Q_2|\,.$$

Since also $\int_{\mathbb{R}^d} \chi_E(x,y) = |E| = |Q_1||Q_2|$ we have established all three conclusions of Fubini's theorem and hence that $\chi_E \in \mathcal{F}$.

(b) Let E be the boundary of a closed cube. We have $\int_{\mathbb{R}^d} \chi_E(x,y) = 0$ since the boundary is a measure zero set in $\mathbb{R}^d$. On the other hand, we observe that for almost every y, the slice E^y has measure 0 (what are the exceptions? draw a picture!). Hence $g(y) = \int_{\mathbb{R}^{d_1}} \chi_E(x,y)\;dx = 0$ for a.e. x. Since $g(y)$ is zero almost everywhere, it is integrable with $\int_{\mathbb{R}^{d_2}} g(y)\,dy = 0$. This establishes all three conclusions of Fubini and hence $\chi_E \in \mathcal{F}$.

(c) Let E be a finite union of almost disjoint closed cubes, $E = \bigcup_{k=1}^{K} Q_k$. If we let $\tilde{Q}_k$ denote the interior of Q_k we can write

$$\chi_E = \sum_k \chi_{\tilde{Q}_k} + \sum_\ell \chi_{A_\ell}$$

where A_ℓ denotes the various (finitely many!) boundary components of the finite union. Step 1 immediately gives $\chi_E \in \mathcal{F}$.

(d) Let E be open and of finite measure. By Theorem 2.1 we can write $E = \sum_{j=1}^{\infty} Q_j$. We define the sequence (f_n) of integrable functions $\sum_{k=1}^{k} \chi_{Q_j}$ which by the previous step is a sequence in $\mathcal{F}$. Clearly also $f_n \nearrow \chi_E$ and $\chi_E \in L^1\left(\mathbb{R}^d\right)$ since E has finite measure. Step 2 implies $\chi_E \in \mathcal{F}$.

(e) Let E be a G_δ of finite measure, i.e. $E = \bigcap_{j=1}^{\infty} \tilde{U}_j$ with $\tilde{U}_j$ open. Since E has finite measure we can find a $\tilde{U}_0$ open with $E \subset \tilde{U}_0$ and $m\left(\tilde{U}_0\right) < \infty$. Then the sequence

$$U_k = \bigcap_{j=1}^{k} \tilde{U}_j \cap \tilde{U}_0$$

is a sequence of open sets which decreases to E, hence $f_k = \chi_{U_k} \searrow \chi_E$ and by Step 2 we conclude $\chi_E \in \mathcal{F}$.

(f) Let E be a set of measure zero. There is a G_δ-set G with $E \subset G$ and $m(G) = 0$ (why?). We know that $\chi_G \in \mathcal{F}$ by the previous step and from

$$\int_{\mathbb{R}^{d_2}} dy \int_{\mathbb{R}^{d_1}} dx \ \chi_G(x,y) = \int_{\mathbb{R}^d} \chi_G = 0$$

we infer $\int_{\mathbb{R}^{d_1}} \chi_G(x,y) = 0$ for a.e. y. Since $0 \leq \chi_E \leq \chi_G$, the same statement holds for χ_E. The three conclusions of Fubini are now immediate and we conclude $\chi_E \in \mathcal{F}$.

(g) Let E be an arbitrary measurable set of finite measure. By Proposition 2.4 we can write $E = G \setminus N$ for G a G_δ-set and N as set of measure zero contained in G. Therefore $\chi_E = \chi_G - \chi_N$ and since this is a finite linear combination of functions belonging to $\mathcal{F}$ we conclude by Step 1 that $\chi_E \in \mathcal{F}$.

Step 4. We now conclude the proof. If $f \in L^1(\mathcal{R}^d)$ we have $f = f^+ - f^-$ and we will show $f^+ \in \mathcal{F}$ and $f^- \in \mathcal{F}$ separately. To show $f^+ \in \mathcal{F}$ pick (by Theorem 2.7) an increasing sequence of simple functions (φ_k) with $\varphi_k \nearrow f^+$. Each φ_k is in $\mathcal{F}$ by Steps 1, 2 and 3 and the definition of a simple function. By Step 2 we conclude $f^+ \in \mathcal{F}$. Of course $f^- \in \mathcal{F}$ is proven analogously.

4 Differentiation and Integration

Now that we have defined a new integral, the Lebesgue integral, we shall investigate its relation with differentiation. In your first year analysis courses you met this relation as the Fundamental Theorem of Calculus (involving the Riemann integral).

4.1 Differentiation of the Integral

We first would like to investigate whether the following theorem is true:

Theorem 4.1. *Given $f : \mathbb{R} \to \mathbb{R}$ integrable on $[a, b]$, the function*

$$F(x) = \int_a^x f(y)\, dy \qquad a \le x \le b \tag{27}$$

is differentiable for almost every $x \in [a, b]$ and $F'(x) = f(x)$ holds for a.e. $x \in [a, b]$.

Note that if f is continuous this statement holds by the fundamental theorem of calculus. The above theorem indeed turns out to be true as stated. To prove it, we will be lead to the averaging problem.

It is easy to see that Theorem 4.1 follows if we can show that for almost every x we have

$$\lim_{h \to 0^+} \frac{1}{h} \int_x^{x+h} f(y)\, dy = \lim_{h \to 0^+} \frac{1}{h} \int_{x-h}^{x} f(y)\, dy = f(x)\ . \tag{28}$$

We reformulate this as the averaging problem:

$$\text{(AVP)} \qquad \text{Does } \lim_{\substack{|I| \to 0 \\ x \in I}} \frac{1}{|I|} \int_I f(y)\, dy = f(x) \text{ hold for a.e. } x,$$

where I denotes a (say open) interval containing x. Below we will study the averaging problem in dimension d. More precisely, we will prove the following

Theorem 4.2. *(Lebesgue-Differentiation-Theorem) Suppose f is integrable on $\mathbb{R}^d$. Then*

$$\lim_{\substack{m(B) \to 0 \\ x \in B}} \frac{1}{m(B)} \int_B f(y)\, dy = f(x) \quad \textit{holds for a.e. } x,$$

where B denotes an open ball containing x.

By repeating the proof of Theorem 4.2 with the limits in (28) replacing the expressions in Theorem 4.2, we also obtain

Theorem 4.3. *Suppose f is integrable on $\mathbb{R}$. Then (28) holds for a.e. x hence proving Theorem 4.1.*

4.1.1 Proof of the Lebesgue Differentiation Theorem

To prove Theorem 4.2 we shall use the observation that the Theorem is true for f being continuous and that the continuous functions are dense in L^1. To estimate the error-terms that arise in the approximation by continuous functions we shall need the important Hardy-Littlewood maximal function.

Definition 4.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be integrable. The maximal function $f^\star$ is defined by*

$$f^\star(x) = \sup_{B \ni x} \frac{1}{m(B)} \int_B |f(y)| dy$$

where the sup *is taken over all balls containing x.*

Compared with the averages on the left hand side of the equality Theorem 4.2 we replace f by its absolute value and instead of taking the limit for small balls we take the sup over *all* balls which contain x.

Remark 4.1. *There are other variants of the maximal function. One can replace the balls by balls centred around x or the balls by cubes (or more general sets of "bounded exccentricity"). Furthermore, in the 1-dimensional case one can define $f_R^\star(x) = \sup_{h>0} \frac{1}{h} \int_x^{x+h} |f(y)| dy$ and $f_L^\star(x) = \sup_{h>0} \frac{1}{h} \int_{x-h}^{x} |f(y)| dy$. All results about the maximal function proven in Proposition 4.1 below also hold for these variants (the proof being exactly the same).*

Proposition 4.1. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be integrable. Then the maximal function $f^\star$ satisfies:*

(1) $f^\star$ is measurable

(2) $f^\star(x) < \infty$ for a.e. x.

(3) $f^\star$ satisfies the estimate

$$m\left(\{x \in \mathbb{R}^d \mid f^\star(x) > \alpha\}\right) \le \frac{3^d}{\alpha} \|f\|_{L^1(\mathbb{R}^d)} .$$

As we shall see, $f^\star$ is in general not integrable, i.e. not in $L^1\left(\mathbb{R}^d\right)$. The conclusion (3) serves as a substitute: It does not control the L^1-norm of $f^\star$ but the measure of the set on the left. To understand this better, recall that by Chebychev's inequality (Example Sheet 2) we have for arbitrary $g \in L^1$ the inequality

$$m\left(\{x \in \mathbb{R}^d \mid g(x) > \alpha\}\right) \le \frac{1}{\alpha} \|g\|_{L^1(\mathbb{R}^d)} .$$

Hence controlling left hand side for $f^\star$ in (3) is indeed weaker than controlling the L^1-norm of $f^\star$ (which, as mentioned, is generally impossible).

Proof. In order not to disturb the proof of Theorem 4.2 we postpone the proof to Section 4.1.2. ☐

Proof of Theorem 4.2. It suffices to show that for each $\alpha > 0$ the set

$$E_\alpha = \left\{ x \;\Big|\; \limsup_{\substack{m(B) \to 0 \\ x \in B}} \left| \frac{1}{m(B)} \int_B f(y)\, dy - f(x) \right| > 2\alpha \right\}$$

has measure zero.[7] To do this, we fix α and show for any $\epsilon > 0$ we have $m(E_\alpha) < \epsilon$, hence $m(E_\alpha) = 0$.

Fix α and let $\epsilon > 0$. We choose a continuous function of compact support g with

$$\|f - g\|_{L^1(\mathbb{R}^d)} < \epsilon\,.$$

Since g is continuous, we have for all x that (why?)

$$\lim_{\substack{m(B)\to 0\\ x\in B}} \frac{1}{m(B)} \int_B g(y)\,dy = g(x)\,.$$

We write the difference of the expression defining E_α as

$$\begin{aligned} \Big|\frac{1}{m(B)} \int_B f(y)\,dy - f(x)\Big| \leq \Big|\frac{1}{m(B)} \int_B (f(y) - g(y))\,dy\Big| \\ + \Big|\frac{1}{m(B)} \int_B g(y)\,dy - g(x)\Big| + |g(x) - f(x)|. \end{aligned} \tag{29}$$

Taking the lim sup we observe that the second term on the right goes to zero while the first term can be estimated by the maximal function as clearly the lim sup is dominated by the sup over all balls. Hence

$$\limsup_{\substack{m(B)\to 0\\ x\in B}} \Big|\frac{1}{m(B)} \int_B f(y)\,dy - f(x)\Big| \leq (f^\star - g^\star)(x) + |g(x) - f(x)|. \tag{30}$$

We we now define the sets

$$F_\alpha = \{x \mid (f-g)^\star(x) > \alpha\} \quad \text{and} \quad G_\alpha = \{x \mid |f(x) - g(x)| > \alpha\}\,,$$

then $E_\alpha \subset F_\alpha \cup G_\alpha$ as clearly at least one summand in (30) has to be be bigger than α in order in order for the sum to be potentially bigger than 2α, as is required for the definition of E_α. But then

$$m(G_\alpha) \leq \frac{1}{\alpha}\|f - g\|_{L^1(\mathbb{R}^d)} \quad \text{by Chebychev,} \tag{31}$$

$$m(F_\alpha) \leq \frac{3^d}{\alpha}\|f - g\|_{L^1(\mathbb{R}^d)} \quad \text{by (3) of Proposition 4.1.} \tag{32}$$

We conclude

$$m(E_\alpha) \leq \frac{3^d}{\alpha}\epsilon + \frac{1}{\alpha}\epsilon$$

and since we have shown this for arbitrary $\epsilon > 0$, $m(E_\alpha) = 0$. $\square$

[7] Here

$$\limsup_{\substack{m(B)\to 0\\ x\in B}} |...| := \inf_{\delta>0} \sup_{B_\delta \ni x} |...|$$

If the E_α have measure zero then $\bigcup_{n=1}^\infty E_{\frac{1}{n}}$ has measure zero and hence $\limsup |...| = 0$ for almost every x. This implies that the limit in Theorem 4.2 exists and is equal to $f(x)$ for a.e. x.

4.1.2 Proof of Proposition 4.1

The first assertion follows from observing that $E_\alpha = \{x \in \mathbb{R}^d \mid f^\star(x) > \alpha\}$ is open. Indeed, if $x \in E_\alpha$, there is an open ball B with $x \in B$ and $\frac{1}{m(B)}\int_B |f(y)|dy > \alpha$. But then, since a small neighborhood of x is also contained in B we have $f(\tilde{x}) > \alpha$ for all points in that neighborhood.

Assertion (3) follows from (2) by observing that

$$\{x \mid f^\star(x) = \infty\} = \bigcap_{n=1}^{\infty} \{x \in \mathbb{R}^d \mid f(x) > n\}.$$

and hence by monotonicity and (2), $m(\{x \mid f^\star(x) = \infty\}) \leq \frac{3^d}{n}\|f\|_{L^1}$ for any n, which proves $m(\{x \mid f^\star(x) = \infty\}) = 0$ as desired.

To prove (2) we use the following version of the Vitali Covering Lemma

Lemma 4.1. *Suppose $\mathcal{B} = \{B_1, B_2, ..., B_N\}$ is a finite collection of open balls in $\mathbb{R}^d$. Then there exists a disjoint subcollection $B_{i_1}, ..., B_{i_k}$ of $\mathcal{B}$ that satisfies*

$$m\left(\bigcup_{n=1}^{N} B_n\right) \leq 3^d \sum_{j=1}^{k} m\left(B_{i_j}\right) .$$

In other words, we can always pick a disjoint subcollection which covers a fixed fraction of the original collection, the fraction not depending on the number of balls in the collection.

Proof. The proof relies on the simple observation that if two balls B and B' intersect and $\rho(B) \geq \rho(B')$ (with ρ denoting the radius), then B' is contained in a ball $\tilde{B}$ which is concentric with B and has 3 times the radius of B (draw a picture!).

We then construct the subcollection as follows. In the first step we pick B_{i_1} to be the largest ball in the collection. We then delete from the collection $\mathcal{B}$ the ball B_{i_1} and all balls intersecting it.

In the second step, we pick B_{i_2} to be the largest ball in the remaining collection and delete from the latter B_{i_2} and all balls intersecting B_{i_2}. And so on.

After finitely many steps we obtain a disjoint subcollection of balls B_{i_1}, ..., B_{i_k} which is such that the union of the $\tilde{B}_{i_k}$ (the balls with three times the radius) contains the union of the original B_i. Therefore

$$m\left(\bigcup_{n=1}^{N} B_n\right) \leq m\left(\bigcup_{j=1}^{k} \tilde{B}_{i_j}\right) \leq \sum_{j=1}^{k} m\left(\tilde{B}_{i_j}\right) = 3^d \sum_{j=1}^{k} m\left(B_{i_j}\right) ,$$

where we have used the dilation invariance of the Lebesgue measure in the last step. ∎

With the Lemma we can prove (3) of the Proposition. We let

$$E_\alpha = \{x \mid f^\star(x) > \alpha\} .$$

Given $x \in E_\alpha$ there exists a ball B_x containing x with

$$\frac{1}{m(B_x)}\int_{B_x} |f(y)|dy > \alpha \quad \text{or equivalently} \quad m(B_x) \leq \frac{1}{\alpha}\int_{B_x} |f(y)|dy. \tag{33}$$

We fix an arbitrary compact set $K \subset E_\alpha$. We have $K \subset \bigcup_{x \in E_\alpha} B_x$ and by compactness, $K \subset \bigcup_{n=1}^N B_n$ for a finite subcollection. Now

$$m(K) \le m\left(\bigcup_{n=1}^{N} B_n\right) \le 3^d \sum_{j=1}^{k} m\left(B_{i_j}\right) \le \frac{3^d}{\alpha} \sum_{j=1}^{k} \int_{B_{i_j}} |f(y)| dy$$

where we have used the monotonicity in the first, the covering Lemma in the second and (33) in the third step. Now it is clear that we have

$$m(K) \le \frac{3^d}{\alpha} \int_{\bigcup_j B_{i_j}} |f(y)| dy \le \frac{3^d}{\alpha} \int_{\mathbb{R}^d} |f(y)| dy \, .$$

Since this holds for all compact subsets K of E_α the estimate holds for E_α itself (Exhaust E_α by increasing compact sets and take the limit.)

4.1.3 Final Remarks

Note that applying the Lebesgue Differentiation Theorem to $|f|$ one obtains

Corollary 4.1. *We have $f^\star(x) \ge |f(x)|$ for almost every x.*

Secondly, note that the Lebesgue Differentiation Theorem holds under the weaker hypothesis that $f \in L^1_{loc}\left(\mathbb{R}^d\right)$ since differentiability is a local property. Here $f \in L^1_{loc}\left(\mathbb{R}^d\right)$ if for every ball B the function $f(x)\chi_B(x)$ is integrable. For instance $f(x) = e^{x^2}$ is in $L^1_{loc}\left(\mathbb{R}^d\right)$ but not integrable.

4.2 Differentiation of functions

We now want to ask the following question: What conditions on $F : \mathbb{R} \to \mathbb{R}$ guarantee that a) F is differentiable almost everywhere and b) the identity

$$F(b) - F(a) = \int_a^b F'(x) dx \tag{34}$$

holds. We know that if F is continuously differentiable, then the above identity holds by the fundamental theorem of calculus.[8] We also know, by what we did in the previous section that if F is given as the indefinite integral of an integrable function $F(x) = \int_a^x f(y)dy$, then a) and b) are true. But how do we characterise such functions? These circle of questions leads us to the study of functions of bounded variation, which itself is intimately connected to the study of rectifiability of curves in the plane.

4.2.1 Functions of bounded variation

Let γ be a curve in the plane $\gamma : [a, b] \to \mathbb{R}^2$ with $\gamma(t) = (x(t), y(t))$, where we assume x and y to be continuous real valued functions on the interval $[a, b]$.

We say that γ is **rectifiable** if there exists an $M < \infty$ such that for *any* partition $a = t_0 < t_1 < ... < t_N = b$ of $[a, b]$ we have

$$\sum_{j=1}^{N} |\gamma(t_j) - \gamma(t_{j-1})| \le M \, .$$

[8] In fact, F differentiable almost everywhere and F' Riemann integrable is sufficient.

The **length of a recitifable curve** is defined as the smallest such M or, equivalently, as the sup over all partitions, i.e.

$$L(\gamma) = \sup_{a=t_0<t_1<...<t_N=b} \sum_{j=1}^{N} |\gamma(t_j) - \gamma(t_{j-1})|$$

The question we now ask is: What conditions on $x(t)$ and $y(t)$ for a given curve guarantee that the curve is rectifiable? If x and y are continuously differentiable, we know that this is the case and we can even establish the formula $L(\gamma) = \int_a^b dt|\dot{\gamma}|dt = \int_a^b dt\sqrt{\dot{x}^2(t) + \dot{y}^2(t)}dt$. But what about weaker conditions?

Let $F : [a,b] \to \mathbb{R}$ be a function, not necessarily continuous. Then given a partition $\mathcal{P}$ of $[a,b]$, say $a = t_0 < t_1 < ... < t_N = b$, we define

$$V_{F,\mathcal{P}} = \sum_{i=1}^{N} |F(t_j) - F(t_{j-1})| \quad \text{to be the variation of } F \text{ on } \mathcal{P}.$$

Definition 4.2. *The function $F : [a,b] \to \mathbb{R}$ is* **of bounded variation** *if there exists an $M < \infty$ such that $V_{F,\mathcal{P}} \leq M$ holds for all partitions $\mathcal{P}$ of $[a,b]$.*

Observation 4.1. *If $\mathcal{P}_1$ is a refinement of $\mathcal{P}_2$ (meaning that every point of $\mathcal{P}_2$ is also a point in $\mathcal{P}_1$), then $V_{F,\mathcal{P}_1} \geq V_{F,\mathcal{P}_2}$.*

This is easily seen by considering a finite sequence of one-point refinements and the triangle inequality.

Theorem 4.4. *A curve $\gamma : [a,b] \to \mathbb{R}^2$, $\gamma(t) = (x(t), y(t))$ is rectifiable if and only if both the functions x and y are of founded variation.*

Proof. This follows by observing that for any partition $\mathcal{P}$ we have

$$\begin{aligned}
&\sum_{j=1}^{N} \sqrt{(x(t_j) - x(t_{j-1}))^2 + (y(t_j) - y(t_{j-1}))^2} \\
&\leq \sum_{j=1}^{N} \left(|x(t_j) - x(t_{j-1})| + |y(t_j) - y(t_{j-1})|\right) \\
\leq 2 &\sum_{j=1}^{N} \sqrt{(x(t_j) - x(t_{j-1}))^2 + (y(t_j) - y(t_{j-1}))^2}
\end{aligned} \tag{35}$$

□

4.2.2 Examples of functions of bounded variation

1. If $F : [a,b] \to \mathbb{R}$ is monotone and bounded, then F is of bounded variation. To see this, say that F is increasing (otherwise consider $-F$) and $|F(x)| \leq M$. Then we have for any partition with $a = t_0 < t_1 < ... < t_N = b$ that $\sum_{j=1}^{N} |F(t_j) - F(t_{j+1})| = \sum_{j=1}^{N} F(t_j) - F(t_{j+1}) = F(b) - F(a) \leq 2M$.

2. If $F : [a, b] \to \mathbb{R}$ is Lipschitz on $[a, b]$ then F is of bounded variation. To see this, let $|F(y) - F(x)| \leq L|y - x|$ for all $x, y \in [a, b]$.[9] But then

$$\sum_{j=1}^{N} |F(t_j) - F(t_{j+1})| = \sum_{j=1}^{N} L|t_{j+1} - t_j| \leq L(b - a)$$

holds for any partition.

3. Let $a, b > 0$. The function $F : [0, 1] \to \mathbb{R}$ given by

$$F(x) = \begin{cases} x^a \sin(x^{-b}) & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}$$

is of bounded variation on $[0, 1]$ if and only if $a > b$. See Exercise Sheet 7.

Note that (as the last example shows) F continuous does not imply F of bounded variation (and neither the other way around, as a simple jump function shows)!

4.2.3 Characterisation of functions of bounded variation

The next theorem shows that the first example above in some sense captures all functions of bounded variation:

Theorem 4.5. *A real valued function $F : [a, b] \to \mathbb{R}$ is of bounded variation if and only if F is the difference of two increasing functions.*

Proof. We first define the following functions:

$$T_F(a, x) = \sup_{\mathcal{P} \text{ of } [a,x]} \sum_{j=1}^{N} |F(t_j) - F(t_{j-1})|$$

the total variation of F on $[a, x]$ for $a \leq x \leq b$,

$$P_F(a, x) = \sup_{\mathcal{P} \text{ of } [a,x]} \sum_{(+)} F(t_j) - F(t_{j-1})$$

the positive variation of F (where the sum is over those j for which $F(t_j) - F(t_{j-1}) \geq 0$ and

$$N_F(a, x) = \sup_{\mathcal{P} \text{ of } [a,x]} \sum_{(-)} - (F(t_j) - F(t_{j-1}))$$

the negative variation of F (where the sum is over those j for which $F(t_j) - F(t_{j-1}) \leq 0$. Observe that T_F and also P_F and N_F are increasing and bounded, hence functions of bounded variation. The functions are related as follows:

Lemma 4.2. *Suppose $F : [a, b] \to \mathbb{R}$ is of bounded variation on $[a, b]$. Then for all $a \leq x \leq b$ we have the relations:*

$$F(x) - F(a) = P_F(a, x) - N_F(a, x)$$

$$T_F(a, x) = P_F(a, x) + N_F(a, x)$$

[9]Note this is in particular satisfied if F is differentiable everywhere and $|F'| \leq L$ (use the mean value theorem).

Proof. Note that the above relations would clearly hold for any partition if there was no sup in the definition of N_F, P_F and T_F. The idea of the proof is therefore to borrow an ϵ. Let $\epsilon > 0$ be given. We pick partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ such that

$$0 \leq P_F - \sum_{(+)\in\mathcal{P}_1} F(t_j) - F(t_{j-1}) < \epsilon \ , \ 0 \leq N_F - \sum_{(-)\in\mathcal{P}_2} -(F(t_j) - F(t_{j-1})) < \epsilon$$

Refining to a common partition $\mathcal{P}$ we have (why?)

$$0 \leq P_F - \sum_{(+)\in\mathcal{P}} F(t_j) - F(t_{j-1}) < \epsilon \ , \ 0 \leq N_F - \sum_{(-)\in\mathcal{P}} -(F(t_j) - F(t_{j-1})) < \epsilon$$

Now since for the partition $\mathcal{P}$ we have

$$F(x) - F(a) = \sum_{(+)\in\mathcal{P}} F(t_j) - F(t_{j-1}) + \sum_{(-)\in\mathcal{P}} F(t_j) - F(t_{j-1})$$

we can add $N_F - P_F$ on both sides and use the above estimates to obtain

$$|F(x) - F(a) - P_F + N_F| \leq 2\epsilon.$$

Since ϵ was arbitrary the first relation is established. For the second estimate we note that for any partition $\mathcal{P}$ of $[a, x]$, $a = t_0 < t_1 < ... < t_N = x$ we have

$$\sum_{j=1}^{N} |F(t_j) - F(t_{j-1})| = \sum_{(+)\in\mathcal{P}} F(t_j) - F(t_{j-1}) + \sum_{(-)\in\mathcal{P}} - (F(t_j) - F(t_{j-1})) \, .$$

It follows that $T_F \leq P_F + N_F$ (why?). For the other direction, start from

$$\sum_{(+)\in\mathcal{P}} F(t_j) - F(t_{j-1}) + \sum_{(-)\in\mathcal{P}} - (F(t_j) - F(t_{j-1})) \leq T_F \tag{36}$$

which holds for any partition. Given $\epsilon > 0$ find partitions $\mathcal{P}_1$ and $\mathcal{P}_2$ with

$$0 \leq P_F - \sum_{(+)\in\mathcal{P}_1} F(t_j) - F(t_{j-1}) < \epsilon \ , \ 0 \leq N_F - \sum_{(+)\in\mathcal{P}_2} F(t_j) - F(t_{j-1}) < \epsilon$$

In particular, refining to a common partition $\mathcal{P}$ the two estimates hold replacing $\mathcal{P}_1$ and $\mathcal{P}_2$ by $\mathcal{P}$. Combining this with (36) yields $P_F + N_F \leq T_F + 2\epsilon$ and since ϵ was arbitrary $P_F + N_F \leq T_F$ as desired. $\square$

Using the Lemma, we easily complete the proof of Theorem 4.5. Note that if F_1 and F_2 are increasing and bounded, then they both are of bounded variation and hence their difference is of bounded variation. Conversely if F is of bounded variation, we can set

$$F_1(x) = P_F(a, x) + F(a) \quad \text{and} \quad F_2(x) = N_F(a, x)$$

Both F_1 and F_2 are increasing and bounded and their difference is $F(x)$ by the Lemma. $\square$

4.2.4 Bounded variation implies differentiable a.e.

Now that we have introduced the class of functions of bounded variation we can answer the question a) posed at the beginning of Section 4.2 by the following

Theorem 4.6. *If $F : [a, b] \to \mathbb{R}$ is of bounded variation, then F is differentiable a.e., i.e. the limit*

$$\lim_{h \to 0} \frac{F(x+h) - F(x)}{h}$$

exists for almost every $x \in [a, b]$.

The proof of this theorem is quite intricate and we will postpone it to Section 4.2.7, where we prove it under the additional assumption that F is also continuous. For the general case, see Stein-Shakarchi.

Even the weaker statement (assuming that F is also continuous) leads to the following corollaries, the first one following from the earlier observation that Lipschitz functions are of bounded variation:

Corollary 4.2 (Rademacher's theorem in 1 dimension)**.** *If $F : [a, b] \to \mathbb{R}$ is Lipschitz, then it is differentiable a.e.*

Corollary 4.3. *If $F : [a, b] \to \mathbb{R}$ is increasing and continuous, then F' exists almost everywhere. Moreover F' is integrable and*

$$\int_a^b F'(x)dx \leq F(b) - F(a)\,. \tag{37}$$

Remark 4.2. *Recall that ideally we would like to show the equality (34) instead of the inequality in the corollary. However, the example of the Cantor-Lebesgue function, studied in detail on the example Sheets 1, 3 and 8 shows that one cannot expect equality to hold without additional assumptions (beyond bounded variation) on F. The additional condition guaranteeing (34) will be that of absolute continuity. See Section 4.2.5.*

Proof. Note that F is certainly in BV and hence the derivative F' exists a.e. by Theorem 4.6. In particular, for $n \geq 1$ the difference quotients

$$G_n(x) = \frac{F(x + \frac{1}{n}) - F(x)}{\frac{1}{n}} \geq 0$$

form a sequence of non-negative measurable functions converging to F', which is hence a measurable function. Fatou's Lemma (Lemma 3.2) applies and yields

$$\int_a^b F'(x)dx = \int_a^b \lim_{n\to\infty} G_n(x)dx \leq \liminf_{n\to\infty} \int_a^b G_n(x)dx\,.$$

We finally extend F to a continuous function on all of $\mathbb{R}$ and observe that the integral on the right hand side can be written as

$$\begin{aligned}
\int_a^b G_n(x) &= \frac{1}{1/n}\int_a^b F\left(x + 1/n\right)dx - \frac{1}{1/n}\int_a^b F(x)dx \\
&= \frac{1}{1/n}\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} F\left(y\right)dy - \frac{1}{1/n}\int_a^b F(y)dy \\
&= \frac{1}{1/n}\int_{b}^{b+\frac{1}{n}} F\left(y\right)dy - \frac{1}{1/n}\int_a^{a+\frac{1}{n}} F(y)dy
\end{aligned} \tag{38}$$

from which it is manifest that the right hand side converges (independently of the extension) to $F(b) - F(a)$ as desired, in view of the continuity of F. $\square$

4.2.5 Absolute Continuity and the Fundamental Theorem of the Lebesgue integral

We will now tackle the problem of finding the necessary and sufficient conditions on F that guarantees the validity of the identity (34). It turns out that the correct notion is that of absolute continuity:

Definition 4.3. *The function $F : [a, b] \to \mathbb{R}$ is absolutely continuous if for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any collection of disjoint intervals (a_k, b_k), $k = 1, ..., N$ of $[a, b]$ we have*

$$\sum_{k=1}^{N} (b_k - a_k) < \delta \qquad \Longrightarrow \qquad \sum_{k=1}^{N} |F(b_k) - F(a_k)| < \epsilon$$

One immediately deduces that absolute continuity implies continuity, in fact uniform continuity (why?).

Exercise 4.1. *Show that if F is Lipschitz, then it is absolutely continuous. Give an example of a function which is not Lipschitz but absolutely continuous.*

Lemma 4.3. *If $F : [a, b] \to \mathbb{R}$ is absolutely continuous, then it is of bounded variation on $[a, b]$.*

Proof. Let δ be the δ associated with the choice $\epsilon = 1$ in the definition of absolute continuity. Fix once and for all a partition $\mathcal{P}$ with mesh-size smaller than δ, say $a = t_0 < t_1 < ... < t_N = b$. In each interval (t_{i-1}, t_i), the total variation is $T_F\left(t_{i-1}, t_i\right) < 1$ by the definition of absolute continuity. But since $T_F\left(a, b\right) = \sum_{i=1}^{N} T_F\left(t_{i-1}, t_i\right) < N$ the lemma is proven. $\square$

Recall also the following fact, proven on Example Sheet 7:

Lemma 4.4. *If $F : [a, b] \to \mathbb{R}$ is continuous and of bounded variation, then its total variation is continuous. As a consequence, the F_1 and F_2 in the decomposition $F = F_1 - F_2$ of Theorem 4.5 are both continuous (in addition to being increasing and bounded).*

Why is the assumption of absolute continuity the correct notion to establish the identity (34)? To see this, note first that if $F : [a, b] \to \mathbb{R}$ is given by

$$F(x) = \int_a^x f(y)dy$$

with f integrable on $[a, b]$, then F is absolutely continuous. This follows from Proposition 3.5 and hence justifies the name introduced there. From this observation it is immediate that absolute continuity of F is a necessary condition if F is to satisfy the identity (34): Indeed, if (34) holds, then F' is integrable and therefore the right hand side of (34) is absolutely continuous. Hence so is the right hand side.

The next theorem shows that absolute continuity is also a sufficient condition.

Theorem 4.7 (Fundamental Theorem of Lebesgue integration)**.**

1. *If $F : [a,b] \to \mathbb{R}$ is absolutely continuous on $[a,b]$, then F' exists almost everywhere and is integrable. Moreover*

$$F(x) - F(a) = \int_a^x F'(y)dy \quad \textit{holds for all } a \le x \le b. \tag{39}$$

2. *If f is integrable on $[a,b]$, there exists an absolutely continuous function F such that $F'(x) = f(x)$ a.e., for instance*

$$F(x) = \int_a^x f(y)dy\,.$$

Note that we already proved the second part of the theorem: Indeed we have observed that the expression for F is absolutely integrable and by the Lebesgue differentiation theorem we know that $F'(x) = f(x)$ holds almost everywhere. Hence the difficulty is proving the first part. I claim the first part will follow if we can prove

Theorem 4.8. *If $F : [a,b] \to \mathbb{R}$ is absolutely continuous on $[a,b]$, then F' exists a.e. Moreover, if $F'(x) = 0$ for a.e. x, then F is constant.*

Indeed, assuming Theorem 4.8 for the moment, the proof of Theorem 4.7 becomes rather short. We first observe that F absolutely continuous implies that F is continuous and of bounded variation (Lemma 4.3) and hence that the F_1 and F_2 in the decomposition $F = F_1 - F_2$ are increasing and continuous (Lemma 4.4). Corollary 4.3 then implies that (F_1 and F_2 hence) F is differentiable a.e. and also that (F_1' and F_2' hence) F' is integrable on $[a,b]$.

We now define $G(x) = \int_a^x F'(y)dy$. Clearly G is absolutely continuous and hence so is $G(x) - F(x)$. Theorem 4.1 implies that $G'(x) = F'(x)$ a.e. We conclude that the function $G - F$ is absolutely continuous an has derivative zero a.e. and applying Theorem 4.8 that $G - F$ is constant. Observing $(G-F)(x) = (G-F)(a) = -F(a)$ produces the identity (39).

It remains to prove Theorem 4.8. Just like the proof of Theorem 4.6 (which we postponed to Section 4.2.7), the proof is quite intricate and isolated in the following subsection 4.2.6. While you don't have to remember the details of the proof of these theorems you should realised that they (together with the Lebesgue differentiation theorem) are at the heart of the Fundamental Theorem of Lebesgue integration, Theorem 4.7.

4.2.6 The proof of Theorem 4.8

The proof of Theorem 4.8 relies on a more elaborate version of the Vitali covering lemma than the one we have already seen in Lemma 4.1. You may wonder at this point why covering Lemmas appear at all in this context. In a typical situation we would like to determine the measure of some set which we happen to know is covered by balls. Determining the measure is of course much easier if we can select a finite disjoint subcollection of balls which covers the set (or at least a large fraction of it).

Definition 4.4. *A collection $\mathcal{B}$ of balls $\{B\}$ is a* **Vitali covering** *of a set E if for every $x \in E$ and $\eta > 0$ there is a ball $B \in \mathcal{B}$ such that $x \in \mathcal{B}$ and $m(B) < \eta$.*

In other words, in a Vitali covering of E every point is covered by arbitrary small balls. The next lemma asserts that give a Vitali covering of a set of finite measure, we can pick *finitely* many balls which cover the set E up to an arbitrary prescribed $\delta > 0$:

Lemma 4.5. *Suppose $E \subset \mathbb{R}^d$ is a set of finite measure, $m(E) < \infty$, and $\mathcal{B}$ is a Vitali covering of E. Then, for any $\delta > 0$ we can find finitely many balls $B_1, B_2, ..., B_N$ in $\mathcal{B}$ which are disjoint and so that*

$$\sum_{j=1}^{N} m(B_j) \geq m(E) - \delta .$$

Moreover, we can select the balls such that also

$$m\left(E \setminus \bigcup_{j=1}^{N} B_j\right) < 2\delta .$$

Note that the first estimate alone does still allow a large fraction of E not to be covered by balls since the positivity on the left hand side of the inequality could come from a large ball which lies mostly outside of E. The second estimate excludes that possibility stating that the set that is not covered by balls of the finite subcollection is also small.(Draw some pictures!)

Proof. The idea of the proof of the Lemma is to approximate the set E from inside using compact sets, and then use the elementary covering Lemma 4.1 to extract a finite disjoint subcollection covering at least a part of E. One then looks at the part not yet covered and – in case it is still too large – approximates it again from inside by a compact set, applies the old covering lemma and so on. This procedure will eventually lead to the δ approximation claimed.

The details are as follows. It clearly suffices to prove the estimates for $\delta < m(E)$. Fix such a δ. Let us also pick an open set $\mathcal{U}$ with $E \subset \mathcal{U}$ and $m(\mathcal{U}) < m(E) + \delta$.

Step 1: We pick a compact set $E_1' \subset E$ with $m(E_1') \geq m(E) - \epsilon > \delta - \epsilon \geq \delta$ (why can we do this? Example Sheet 3!). We cover E_1' by balls from $\mathcal{B}$ such that every ball of the covering also lies in $\mathcal{U}$ (this is possible because E and hence E_1' are covered by balls of arbitrarily small radius). Using compactness we choose a finite sub-collection of balls covering E_1' (and contained in $\mathcal{U}$) and using Lemma 4.1 a finite disjoint subcollection $B_1, ...B_{N_1}$ such that

$$\sum_{i=1}^{N_1} m(B_i) \geq \frac{1}{3^d} m(E_1') \geq \frac{\delta}{3^d} .$$

Step 2: If

$$\sum_{i=1}^{N_1} m(B_i) \geq m(E) - \delta$$

then the first estimate is already proven. Otherwise, we have

$$\sum_{i=1}^{N_1} m(B_i) < m(E) - \delta$$

and hence

$$E_2 = E \setminus \bigcup_{i=1}^{N_1} \overline{B_i}$$

has measure $m(E_2) > \delta$ (why?). We then repeat procedure of Step 1, i.e. we find a compact subset E_2' with $m(E_2') \geq \delta$. We can cover the set E_2' by finitely many balls contained in $\mathcal{U}$ and disjoint from $\bigcup_{i=1}^{N_1} \overline{B_i}$ (why? – note that any point in E_2' has finite distance from $\bigcup_{i=1}^{N_1} \overline{B_i}$). Using the old covering Lemma 4.1, we select a finite disjoint collection of these balls $B_{N_1+1}, ..., B_{N_2}$ such that

$$\sum_{i=N_1+1}^{N_2} m(B_i) \geq \frac{1}{3^d} m(E_2') \geq \frac{\delta}{3^d}\,.$$

Overall, we now have a subcollection of *disjoint* balls $B_1, ..., B_{N_2}$ with

$$\sum_{i=1}^{N_2} m(B_i) \geq 2\frac{\delta}{3^d}$$

Step 3: We repeat Step 2, i.e. we ask whether

$$\sum_{i=1}^{N_2} m(B_i) \geq m(E) - \delta$$

in which case we are done or

$$\sum_{i=1}^{N_2} m(B_i) < m(E) - \delta$$

in which case we repeat the procedure in Step 2. If the procedure has not terminated after k steps we have the estimate

$$\sum_{i=1}^{N_k} m(B_i) \geq k\frac{\delta}{3^d}\,.$$

This yields a contradiction if $k \geq \frac{m(E)-\delta}{\delta} 3^d$, because then the right hand side is $\geq m(E) - \delta$ in contradiction with the procedure having not terminated.

To prove the second estimate, we observe that since all balls in the iteration above are contained in $\mathcal{U}$ we have the disjoint union

$$\left(E \setminus \bigcup_{n=1}^{N} B_i\right) \bigcup \left(\bigcup_{n=1}^{N} B_i\right) \subset \mathcal{U}\,.$$

From this we deduce

$$m\left(E \setminus \bigcup_{n=1}^{N} B_i\right) \leq m(\mathcal{U}) - m\left(\bigcup_{n=1}^{N} B_i\right) \leq m(E) + \delta - (m(E) - \delta) = 2\delta.$$

□

Using the Lemma (in dimension $d = 1$), we can now complete the proof of Theorem 4.8. Note that the difficulty is to prove that $F' = 0$ implies F is constant, since the existence of F' follows from the fact that F is of bounded variation and Theorem 4.6.

It clearly suffices to show $F(b) = F(a)$ as we can then replace $[a, b]$ by an arbitrarily small subinterval. We let

$$E := \{x \in (a, b) \mid F'(x) \text{ exists and is zero}\}$$

and we know $m(E) = b - a$ by assumption. We have for each $x \in E$ that

$$\lim_{h \to 0} \left| \frac{F(x+h) - F(x)}{h} \right| = 0 .$$

Fix now $\epsilon > 0$. In view of the existence of the above limit, we can find for each $\eta > 0$ an open interval around each $x \in E$, $I_x = (a_x, b_x) \subset [a, b]$ with length smaller than η, i.e. $b_x - a_x < \eta$, such that

$$|F(b_x) - F(a_x)| \leq \epsilon(b_x - a_x) .$$

The collection of these intervals forms a Vitali covering of the set E. Hence we can apply Lemma 4.5: For any $\delta > 0$ (which we will choose momentarily dependent on ϵ) we can select finitely many *disjoint* intervals $I_i = I_{x_i} = (a_{x_i}, b_{x_i})$ with $1 \leq i \leq N$ such that

$$\sum_{i=1}^{N} m(I_i) \geq m(E) - \delta = (b-a) - \delta$$

On each interval I_i we have $|F(b_i) - F(a_i)| \leq \epsilon (b_i - a_i)$ as this is a property of all "balls" in the Vitali covering. Summing this inequality yields

$$\sum_{i=1}^{N} |F(b_i) - F(a_i)| \leq \epsilon(b-a)$$

because the intervals (a_i, b_i) are disjoint and contained in $[a, b]$.

We now consider the complement of $\bigcup_{j=1}^{N} I_j$ in $[a, b]$, denoted A. It consists of finitely many closed intervals with $m(A) < \delta$, so $A = \bigcup_{k=1}^{M} [\alpha_k, \beta_k]$.

The idea is to use the absolute continuity on these disjoint intervals. Indeed, we choose δ sufficiently small (depending only on ϵ) such that $m(A) < \delta$ implies

$$\sum_{j=1}^{M} |F(\beta_k) - F(\alpha_k)| \leq \epsilon .$$

As a consequence we then have from using the triangle inequality repeatedly

$$F(b) - F(a) \leq \sum_{j=1}^{N} |F(b_j) - F(a_j)| + \sum_{j=1}^{M} |F(\beta_j) - F(\alpha_j)| \leq \epsilon(b-a) + \epsilon .$$

Since this holds for any $\epsilon > 0$ we conclude $F(b) = F(a)$ as desired.

4.2.7 The proof of Theorem 4.6 (non examinable)

I will add this non-examinable material to the notes at some point. It will not be lectured in class.

5 Abstract Measure Theory

So far we have successfully dealt with the problem of defining a measure for sets on $\mathbb{R}^n$. We recall that the main steps of the analysis were

1. an elementary notion of measure for the simplest sets (rectangles or cubes)
2. the introduction of an exterior measure (defined on *all* subsets of $\mathbb{R}^d$) which assigned a "measure" as the infimum of countable coverings by cubes and was consistent with the elementary measure on the rectangles
3. the introduction of the class of Lebesgue measurable sets which satisfied the desired property of countable additivity

Given the class of Lebesgue measurable sets, we then defined measurable functions $f : \mathbb{R}^d \to \mathbb{R}$ and developed an integration theory which allowed us to integrate a much larger class of functions than the class of Riemann integrable functions.

Our goal in this section is to develop the abstract framework that will allow us to construct general measure spaces. The above "pedestrian" construction of the Lebesgue measure on $\mathbb{R}^d$ can then be viewed as a particular example of the abstract construction. More interestingly perhaps, the general construction allows to construct many interesting measure spaces which appear in probability and geometric measure theory.

5.1 Measure Spaces: Definition and basic examples

Definition 5.1. *A* **measure space** *consists of a set X equipped with two fundamental objects*

1. *A* σ**-algebra** $\mathcal{M}$ *of sets (i.e. a non-empty collection of subsets of X which is closed under complements and countable unions and intersections), which are called measurable sets*
2. *A* **measure**
$$\mu : \mathcal{M} \to [0, \infty]$$
i.e. a function with the property of being countably additive, i.e. if $E_1, E_2, ...$ is a countable collection of disjoint sets in $\mathcal{M}$, then
$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n) \ .$$

A measure space is denoted $(X, \mathcal{M}, \mu)$. Sometimes (X, μ) or even just X is used, if the σ-algebra and the measure that are being used are clear from the context.

Exercise 5.1. *Let $(X, \mathcal{M}, \mu)$ be a measure space. Show that the measure is monotone (i.e. $A \subset B$ implies $\mu(A) \leq \mu(B)$ for $A, B \in \mathcal{M}$) and that for any countable collection of sets (E_n) one has $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n)$.*

We make three more definitions of properties that a measure space might have or not have. Their meaning will become clearer in the discussion of the examples that follow immediately.

Definition 5.2.

1. *We say $(X, \mathcal{M}, \mu)$ is* **finite** *if $\mu(X) < \infty$.*
2. *We say $(X, \mathcal{M}, \mu)$ is* **σ-finite** *if there exists a countable collection $(E_i)_{i=1}^{\infty}$ of sets of finite measure ($\mu(E_i) < \infty$) such that $X = \bigcup_{i=1}^{\infty} E_i$.*
3. *We say that $(X, \mathcal{M}, \mu)$ is* **complete** *if the following statement is true: Given any $E \subset \mathcal{M}$ with $\mu(E) = 0$, any $F \subset E$ is also in $\mathcal{M}$ (and has necessarily $\mu(F) = 0$).*

We give a couple of examples of general measure spaces:

1. If X is a non-empty set and $\mathcal{M} = (X, \emptyset)$ we can define $\mu(X)$ arbitrarily and obtain a (rather trivial) measure space.
2. If $X = \mathbb{R}^d$ and $\mathcal{M}$ is the collection of Lebesgue measurable sets, then for any measurable non-negative function $f : \mathbb{R}^d \to \mathbb{R}$, the function $\mu : \mathcal{M} \to [0, \infty]$

$$\mu(E) = \int_E f dx$$

defines a measure and hence $(\mathbb{R}^d, \mathcal{M}, \mu)$ is a measure space. [You should verify this, i.e. check countable additivity using the properties of the integral (additivity, MCT) proven earlier.] The choice $f = 1$ leads to the familiar Lebesgue measure.

A variant of this example is produced by replacing the σ-algebra $\mathcal{M}$ with the (smaller) σ-algebra of Borel-sets on $\mathbb{R}^d$, denoted $\tilde{\mathcal{M}}$. Then $(\mathbb{R}^d, \tilde{\mathcal{M}}, \mu)$ is also a measure space. However, unlike $(\mathbb{R}^d, \mathcal{M}, \mu)$ it is not complete. To see this recall that there are Borel sets of measure zero which contain sets which are only Lebesgue- but not Borel measurable.[10]

3. Let X be a non-empty set and $\mathcal{M} = \mathcal{P}(X)$ be the σ-algebra of all subsets of X. Fix $x_0 \in X$. We can define the measure

$$\mu(E) = \begin{cases} 1 & \text{if } x_0 \in E \\ 0 & \text{if otherwise} \end{cases}$$

for $E \subset \mathcal{M}$. This is the so-called Dirac measure.

4. Let $X = (x_n)_{n=1}^{\infty}$ be a countable set and again $\mathcal{M} = \mathcal{P}(X)$ be the σ-algebra of all subsets of X. If $(\mu_n)_{n=1}^{\infty}$ is a sequence in $[0, \infty]$ we can define the measure μ by

$$\mu(x_n) = \mu_n .$$

By countable additivity we have $\mu(E) = \sum_{x_n \in E} \mu_n$ for $E \subset \mathcal{M}$. If $\mu_n = 1$ for all n the measure μ is called the counting measure as it counts the elements of a set.

[10] This can be seen from Exercise 5 on Example Sheet 3 where you constructed an injective, strictly increasing function $g : [0, 1] \to [0, 1]$ whose image was contained in $\mathfrak{C}$. Since g is monotone it is Borel measurable (Exercise 3 on Example Sheet 3), i.e. it pulls back Borel sets to Borel sets. Taking N a non-measurable subset of $[0, 1]$ we know that $F = g(N)$ is Lebesgue measurable with measure zero because it is a subset of $\mathfrak{C}$, which has measure 0. However, $F = g(N)$ cannot be Borel measurable because if it was, $g^{-1}(F) = N$ would have to be a Borel set.

5.2 Exterior measure and Carathéodory's theorem

While we already gave a few examples of general measure spaces, a natural question is how to construct more interesting examples. Here the notion of an exterior or outer measure is key.

Definition 5.3. *Let X be a non-empty set. An exterior measure (or "outer measure") $\mu_\star$ on X is a function $\mu_\star : \mathcal{P}(X) \to [0, \infty]$ defined on all subsets of X satisfying*

1. $\mu_\star(\emptyset) = 0$.
2. *Monotonicity: If $E_1 \subset E_2$ then $\mu_\star(E_1) \leq \mu_\star(E_2)$.*
3. *Subadditivity: If $E_1, E_2, \ldots$ is a countable family of sets, then*

$$\mu_\star\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu_\star(E_n) .$$

We will give examples of exterior measures below (see Section ... for the exterior Hausdorff measure), here we only note that the exterior measure we defined on sets in $\mathbb{R}^d$ in Section 2.3 satisfies the definition.

We have now reached a critical point. The key step to define the Lebesgue measure from the exterior measure was to give up on measuring *all* subsets of $\mathbb{R}^d$ and instead define a class of measurable sets on which the measure was countably additive. However, Definition 2.2 explicitly used the topology of $\mathbb{R}^d$, i.e. the notion of an open set, which a general measure space does not come equipped with. Here Carathéodory found a very clever criterion which works in the general case (and reduces to our old criterion in the Lebesgue case, cf. Example Sheet 9):

Definition 5.4. *A set $E \subset X$ is (Carathéodory) measurable if for all sets $A \subset X$ one has*

$$\mu_\star(A) = \mu_\star(E \cap A) + \mu_\star(E^c \cap A) . \tag{40}$$

In other words, a measurable set separates any set into two parts which behave well with respect to the exterior measure. As mentioned, the condition can be seen to be equivalent to the condition of being Lebesgue measurable in the case of $X = \mathbb{R}^d$ and the exterior measure of Section 2.3 (Example Sheet 9).

Observation 5.1. *To show that a set $E \subset X$ is measurable, it suffices to check whether the inequality*

$$\mu_\star(A) \geq \mu_\star(E \cap A) + \mu_\star(E^c \cap A)$$

holds for all $A \subset X$, as the reverse inequality holds by the subadditivity of the exterior measure.

The observation immediately implies that sets of exterior measure 0 are measurable since $\mu_\star(A) \geq \mu_\star(E^c \cap A)$ holds by monotonicity.

Theorem 5.1. *Given an exterior measure $\mu_\star$ on a set X, the collection $\mathcal{M}$ of (Carathéodory) measurable sets forms a σ-algebra. Moreover, $\mu_\star$ restricted to $\mathcal{M}$ is a measure.*

Proof. In view of the symmetry of the condition (40), we clearly have that $E \in \mathcal{M}$ implies $E^c \in \mathcal{M}$. It is also easily checked that $\emptyset \in \mathcal{M}$ and hence $X \in \mathcal{M}$.

Having shown non-emptyness and closure under complements, we note that it suffices to show that the class $\mathcal{M}$ is closed under *disjoint* countable unions and that we have countable additivity on $\mathcal{M}$.[11] To establish this, we first show that $\mathcal{M}$ is closed under *finite* unions and *finitely* additive on $\mathcal{M}$ (Step 1) and then move to the countable disjoint case (Step 2).

Step 1: Let $E_1, E_2 \in \mathcal{M}$ and $A \subset X$ be arbitrary. We first use that the condition (40) holds for E_1 and E_2 to produce the inequality

$$\begin{aligned} \mu_\star(A) &= \mu_\star(E_2 \cap A) + \mu_\star(E_2^c \cap A) \\ &= \mu_\star(E_1 \cap E_2 \cap A) + \mu_\star(E_1^c \cap E_2 \cap A) \\ &\quad + \mu_\star(E_1 \cap E_2^c \cap A) + \mu_\star(E_1^c \cap E_2^c \cap A) \end{aligned} \tag{41}$$

The last term can be written as $\mu_\star((E_1 \cup E_2)^c \cap A)$. For the other three terms on the right hand side note that

$$E_1 \cup E_2 = (E_1 \cap E_2) \cup (E_1^c \cap E_2) \cup (E_1 \cap E_2^c)$$

and hence by the subadditivity of $\mu_\star$ we obtain

$$\mu_\star(A) \geq \mu_\star(E_1 \cup E_2 \cap A) + \mu_\star((E_1 \cup E_2)^c \cap A)$$

which proves $E_1 \cup E_2 \in \mathcal{M}$ in view of the above observation. Note that this also implies that $E_1 \cap E_2 \in \mathcal{M}$ since the latter can be written as the complement of the union of two sets in $\mathcal{M}$. To show that $\mu_\star$ is finitely additive assume that E_1 and E_2 are disjoint and observe that

$$\mu_\star(E_1 \cup E_2) = \mu_\star(E_1 \cap (E_1 \cup E_2)) + \mu_\star(E_1^c \cap (E_1 \cup E_2)) = \mu_\star(E_1) + \mu_\star(E_2)$$

where the first equality follows from the fact that E_1 is measurable and the second equality exploits the assumption that $E_1 \cap E_2 = \emptyset$.

Step 2: Now let $E_1, E_2, ...$ be a countable collection of *disjoint* sets in $\mathcal{M}$. Define

$$G_n = \bigcup_{j=1}^{n} E_j \quad \text{and} \quad G = \bigcup_{j=1}^{\infty} E_j$$

We clearly have $G_n \in \mathcal{M}$ by Step 1 and $G^c \subset (G_n)^c$ by definition. For $A \subset X$ arbitrary we start from

$$\mu_\star(A) = \mu_\star(G_n \cap A) + \mu_\star((G_n)^c \cap A) \geq \mu_\star(G_n \cap A) + \mu_\star(G^c \cap A) \tag{42}$$

and try to deal with the first term on the right hand side. We have

$$\begin{aligned} \mu_\star(G_n \cap A) &= \mu_\star(E_n \cap (G_n \cap A)) + \mu_\star((E_n)^c \cap (G_n \cap A)) \\ &= \mu_\star(E_n \cap A) + \mu_\star(G_{n-1} \cap A) \end{aligned} \tag{43}$$

[11] You should verify this. Note that any countable union can be written as a disjoint countable union (how?) and that closure under countable intersection follows using closure under complements and countable unions via de Morgan's laws.

where the second step exploits the disjointness of the E_j. An easy induction of the formula (43) yields

$$\mu_\star (G_n \cap A) = \sum_{j=1}^{n} \mu_\star (E_j \cap A) \ .$$

Now (42) becomes

$$\mu_\star (A) \geq \sum_{j=1}^{n} \mu_\star (E_j \cap A) + \mu_\star (G^c \cap A)$$

and taking the limit as $n \to \infty$ yields

$$\mu_\star (A) \geq \sum_{j=1}^{\infty} \mu_\star (E_j \cap A) + \mu_\star (G^c \cap A) \geq \mu_\star (G \cap A) + \mu_\star (G^c \cap A) \geq \mu_\star (A) \ , \tag{44}$$

where the last two inequalities both follow from the subadditivity of the exterior measure. This shows (again in view of Observation 5.1) that G is measurable. Moreover all inequalities in (44) are actually equalities and hence taking $A = G$ in (44) also yields countable additivity of $\mu_\star$ on $\mathcal{M}$. □

As a final remark we note that the measure space obtained in the above theorem is complete. Indeed, if E has measure zero and $F \subset E$ then the exterior measure of F is zero by monotonicity (recall the exterior measure is defined on all subsets). Since we observed earlier that sets of exterior measure zero are measurable, the claim follows.

5.3 Premeasures and the extension theorem

So far we have seen how to construct a general measure from an exterior measure using Carathéodory's criterion (40). This of course shifts the problem to constructing an exterior measure for subsets of a general X. This is typically done via a premeasure, which is a notion of measure on a smaller, more elementary class of sets (similar to the rectangles we used in the case of the Lebesgue measure).

Definition 5.5. *Let X be a set. An algebra (of sets) in X is a non-empty collection of subsets of X that is closed under complements, finite unions and finite intersections.*

Example 5.1. *The collection of sets arising as finite disjoint unions of sets of the form $(a, b]$, (a, ∞) and $\emptyset$ with $-\infty \leq a < b < \infty$, forms an algebra on $\mathbb{R}$.*

Definition 5.6. *Let $\mathcal{A}$ be an algebra of sets in X. A premeasure on $\mathcal{A}$ is a function $\mu_0 : \mathcal{A} \to [0, \infty]$ that satisfies*

1. $\mu_0 (\emptyset) = 0$
2. *If $E_1, E_2, ...$ is a countable collection of disjoint sets in $\mathcal{A}$ with $\bigcup_{k=1}^{\infty} E_k \in \mathcal{A}$, then*

$$\mu_0 \left(\bigcup_{k=1}^{\infty} E_k \right) = \sum_{k=1}^{\infty} \mu_0 (E_k) \ .$$

In particular, μ_0 is finitely additive on $\mathcal{A}$.

Note that $\bigcup_{k=1}^{\infty} E_k \in \mathcal{A}$ in the second item is an *assumption* unless the union happens to be finite. Note also that a premeasure in monotone (why?)

5.3.1 Construction of a measure from a premeasure

We now show how to construct a general measure from a premeasure. This gives an alternative construction of the Lebesgue measure, which we will describe below. The key in the following

Proposition 5.1. *Let X be a set and μ_0 be a premeasure on an algebra of sets $\mathcal{A}$ in X. Define $\mu_\star : \mathcal{P}(X) \to [0, \infty]$ by*

$$\mu_\star(E) = \inf \left\{ \sum_{j=1}^{\infty} \mu_0(E_j) \mid E \subset \bigcup_{j=1}^{\infty} E_j \text{ where } E_j \in \mathcal{A} \text{ for all } j \right\}.$$

Then

1. *$\mu_\star$ is an exterior measure on X*
2. *$\mu_\star(E) = \mu_0(E)$ for all $E \in \mathcal{A}$*
3. *All sets in $\mathcal{A}$ are (Carathéodory) measurable (i.e. (40) holds)*

The above proposition generates an exterior measure $\mu_\star$ from a premeasure μ_0. We can then apply Carathéodory's theorem (Theorem 5.1) to construct from $\mu_\star$ a measure μ on the σ-algebra of Carathéodory measurable sets $\mathcal{M}_C$. Now since by the above Proposition $\mathcal{A} \subset \mathcal{M}_C$, we have that the σ-algebra generated by $\mathcal{A}$,[12] denoted $\mathcal{M}(\mathcal{A})$, is contained in $\mathcal{M}_C$ and hence in particular μ restricts to a measure on $\mathcal{M}(\mathcal{A})$. (Of course $\mathcal{M}_C$ can be strictly larger than $\mathcal{M}(\mathcal{A})$!) These considerations therefore establish the following

Theorem 5.2 (Hahn-extension)**.** *Let X be a set and μ_0 be a premeasure on an algebra of sets $\mathcal{A}$ in X. Denote by $\mathcal{M}$ the σ-algebra generated by $\mathcal{A}$. Then there exists a measure μ on $\mathcal{M}$ that extends μ_0.*

We make an important remark about the uniqueness. If the premeasure μ_0 is σ-finite (i.e. if X can be written as $X = \bigcup_{i=1}^{\infty} E_i$ for a countable collection (E_i) of sets in $\mathcal{A}$ with $\mu_0(\mathcal{A}) < \infty$) then the measure μ whose existence is promised in the theorem is unique (see Question 4 on Example Sheet 9).

Example 5.2. *Combining Theorem 5.2 and Example 5.1 we outline another construction of the Lebesgue measure on the Borel sets of $\mathbb{R}$ (cf. the third example below Definition 5.2). One starts with the algebra $\mathcal{A}$ of intervals in Example 5.1 and defines the premeasure $\mu_0(I) = |I|$ on the intervals in $\mathcal{A}$. Since $\mathcal{A}$ generates the Borel σ-algebra on $\mathbb{R}$ and since μ_0 is σ-finite, Theorem 5.2 generates a unique measure on the Borel σ-algebra on $\mathbb{R}$. The completion of this measure is precisely the Lebesgue measure defined on the σ-algebra of Lebesgue measurable sets. This last step (completion) will be carried out in Question 2 of Sheet 9.*

[12] Recall this is the smallest σ-algebra containing the sets of $\mathcal{A}$.

5.3.2 The proof of Proposition 5.1

To prove the first part note first that $\mu_\star$ is well-defined since we can choose $E_j = X$ for all j. We also easily see $\mu_\star(\emptyset) = 0$ and $E_1 \subset E_2$ implies $\mu_\star(E_1) \leq \mu_\star(E_2)$. To establish the subadditivity property we repeat the proof of Proposition 2.1. We fix $\epsilon > 0$ and given $E_1, E_2, ...$ in X we choose for each E_i a collection $(E_{i,j})$ in $\mathcal{A}$ with $E_i \subset \bigcup_{j=1}^\infty E_{i,j}$ with $\mu_\star(E_i) + \frac{\epsilon}{2^i} \geq \sum_{j=1}^\infty \mu_0(E_{i,j})$. Then $\bigcup_{i,j} E_{i,j}$ is a countable collection of sets in $\mathcal{A}$ which covers $\bigcup_i E_i$ and hence $\mu_\star(\bigcup_i E_i) \leq \sum_{i,j} \mu_0(E_{i,j}) \leq \sum_i \mu_\star(E_i) + \epsilon$. Since this holds for any $\epsilon > 0$ we are done.

To prove the second part (restriction of $\mu_\star$ to $\mathcal{A}$ coincides with μ_0) we suppose $E \in \mathcal{A}$. Clearly $\mu_\star(E) \leq \mu_0(E)$ since E covers itself. To prove the reverse inequality we let $E \subset \bigcup_{j=1}^\infty E_j$ with $E_j \in \mathcal{A}$ for all j be any covering of E. We then define the sets

$$E_k' = E \cap \left(E_k \setminus \bigcup_{j=1}^{k-1} E_j \right)$$

and note that the E_k' are disjoint elements of $\mathcal{A}$, that $E_k' \subset E_k$ and that $E = \bigcup_{k=1}^\infty E_k'$ (check this!). By the countable additivity of the premeasure we then have

$$\mu_0(E) = \sum_{k=1}^\infty \mu_0(E_k') \leq \sum_{k=1}^\infty \mu_0(E_k)$$

and taking the infimum over all coverings of E by (E_k) in $\mathcal{A}$ yields the claim as this turns the right hand side into $\mu_\star(E)$.

To prove the third part (all sets in $\mathcal{A}$ are measurable for $\mu_\star$) we let A be an arbitrary subset of X, $E \in \mathcal{A}$ and $\epsilon > 0$. It suffices to show

$$\epsilon + \mu_\star(A) \geq \mu_\star(E \cap A) + \mu_\star(E^c \cap A) \,. \tag{45}$$

To prove this, we find a countable collection $E_1, E_2, ...$ in $\mathcal{A}$ with $A \subset \bigcup_{j=1}^\infty E_j$ and

$$\sum_{j=1}^\infty \mu_0(E_j) \leq \mu_\star(A) + \epsilon \,. \tag{46}$$

Since μ_0 is a premeasure, it is finitely additive and we have

$$\sum_{j=1}^n \mu_0(E_j) = \sum_{j=1}^n \mu_0(E \cap E_j) + \sum_{j=1}^n \mu_0(E^c \cap E_j) \,.$$

Taking the limit $n \to \infty$ (note that all terms are increasing in n) we finally find

$$\sum_{j=1}^\infty \mu_0(E_j) = \sum_{j=1}^\infty \mu_0(E \cap E_j) + \sum_{j=1}^\infty \mu_0(E^c \cap E_j) \geq \mu_\star(E \cap A) + \mu_\star(E^c \cap A) \,,$$

with the last inequality following since $\bigcup_{j=1}^\infty E \cap E_j$ is a countable union of sets in $\mathcal{A}$ which covers $E \cap A$. Combining this with (46) yields (45) as desired.

5.4 A further example: Hausdorff measure

In this section we present an application of Carathéodory's construction to construct the α-dimensional Hausdorff measure for sets in $\mathbb{R}^d$. The discussion will be very informal and should merely illustrate that the abstract construction that we went through has interesting applications.

The heuristic idea for Hausdorff measure is to measure the α-dimensional volume of sets in $\mathbb{R}^d$ for $\alpha < d$. For instance a sphere in $\mathbb{R}^3$ should have non-trivial 2-dimensional Hausdorff-measure (namely it's area) while its Lebesgue measure is of course zero. Similarly a interval of length 2 on the x-axis in $\mathbb{R}^3$ should have 1-dimensional Hausdorff-measure equal to 2 etc.

The key idea to construct a measure with these properties lies in the scaling properties of a set. Given a subset $E \subset \mathbb{R}^d$, suppose that scaling the set E by n can be written as adjoining m almost disjoint copies of the original set, i.e.

$$nE = E_1 \cup E_2 \cup ... \cup E_m$$

where the E_i are disjoint congruent copies of E. For instance, if you scale the unit interval in $\mathbb{R}^3$ on the x-axis by n the resulting set is

$$[0, n] \times \{0\} \times \{0\} = \bigcup_{j=1}^{n} [j-1, j] \times \{0\} \times \{0\}$$

so the above holds with $m = n$. The same example with a rectangle yields $m = n^2$. It is intuitively clear that the exponent in the relation $m = n^\alpha$ is what we would call the dimension of the set under consideration.

For a more non-trivial example, consider the Cantor set. It is easy to see that scaling the Cantor set $\mathfrak{C}$ by a factor of 3, we obtain to disjoint copies of the Cantor set, so in this case we have $2 = 3^\alpha$ and it would be tempting to say that the Cantor set has fractional dimension $\frac{\log 2}{\log 3}$.

We now give the definition which formalises the above considerations. For any $E \subset \mathbb{R}^d$ we define the exterior α-dimensional Hausdorff-measure of E as

$$m^\star_\alpha(E) := \lim_{\delta \to 0} \mathcal{H}^\delta_\alpha(E)$$

where

$$\mathcal{H}^\delta_\alpha(E) = \inf \Big\{ \sum_k (diam F_k)^\alpha \ \mid E \subset \bigcup_{k=1}^{\infty} F_k \quad , \quad diam F_k \leq \delta \ \text{ for all } k. \Big\}$$

Here the diameter of a set A is defined as $diam A = \sup\{|x-y|, x, y \in A\}$. Note that $\mathcal{H}^\delta_\alpha(E)$ is well defined because countably many balls of diameter δ cover all of $\mathbb{R}^d$. Note also that as δ decreases, $\mathcal{H}^\delta_\alpha(E)$ increases because we are taking the infimum over fewer sets (the elements F_k in the covering are restricted to be smaller in diameter). Hence the limit is actually defined.

We remark that the coverings by F_k in the definition cannot be replaced by coverings of ball of diameter smaller than δ. This would yield a different quantity. This makes the Hausdorff-measure of a set hard to compute in general.

One can check that $m^\star_\alpha$ is monotone and sub-additive and hence indeed an exterior measure. It moreover satisfies that if the distance of two sets E_1 and E_2 is strictly positive then we have $m^\star_\alpha(E_1 \cap E_2) = m^\star_\alpha(E_1) + m^\star_\alpha(E_2)$, i.e. additivity (this makes $m_\alpha^\star$ a so-called metric exterior measure). The first two facts allow us to apply Carathéodory's theorem (Theorem 5.1) to construct from $m_\alpha^\star$ a measure m_α on the σ-algebra of Carathéodory measurable sets. The third fact allows one to prove that the σ-algebra of Carathéodory measurable sets contains the closed subsets and hence in particular the σ-algebra of Borel sets. The measure m_α restricted to the Borel sets is commonly known as the α-dimensional Hausdorff measure on $\mathbb{R}^d$. More on this in Stein-Shakarchi.

5.5 Integration on a general measure space

Our next task is to develop the analogue of the integration theory for the Lebesgue integral to a general measure space $(X, \mathcal{M}, \mu)$. We assume for simplicity that the measure space $(X, \mathcal{M}, \mu)$ is also σ-finite. The punchline is: Everything that we did for the Lebesgue integral generalises and the proofs go through almost word by word. This is why we merely collect rather informally the main points.

Measurable functions are defined as before: A function $f : X \to [0, \infty]$ on a measure space $(X, \mathcal{M}, \mu)$ is measurable if $f^{-1}([-\infty, a)) = \{x \in X \mid f(x) < a\} \in \mathcal{M}$ for all $a \in \mathbb{R}$. Properties of measurable functions (limit of a sequence of measurable functions is measurable etc) continue to hold.

The notion of almost everywhere "a.e." is defined with respect to the measure μ, for instance $f = g$ a.e. means that $\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0$.

We can define simple functions on X as before as finite linear combinations of characteristic functions of measurable sets of finite measure,

$$\sum_{k=1}^{N} a_k \chi_{E_k} .$$

The approximation theorems of Section 2.8 continue to hold true. This actually needs the σ-finite condition.[13]

Egorov's theorem remains true (check this!).

The integral can be define via the same four stage procedure that we carried out for the Lebesgue integral leading to

$$\int_X f(x) d\mu(x)$$

the integral of a measurable function over a general measure space (which is again linear, monotone, etc). We say that f is integrable if

$$\int_X |f(x)| d\mu(x) < \infty .$$

Finally, the important convergence theorems (Fatou's Lemma, the Monotone Convergence Theorem and the Dominant Convergence theorem) all continue to hold. Our final goal, which does not immediately generalise, is to prove a general Fubini theorem for the integral on a general measure space.

[13] Think of the first step in the proof of the approximation theorems where we truncate F to be defined on a set of finite measure (a cube in the Lebesgue case) which in the limit exhausts $\mathbb{R}^d$.

5.6 Construction of product measures

We finally discuss product measures. The idea is the following. Given two measure spaces $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ we would like to construct a σ-algebra "$\mathcal{M} \otimes \mathcal{N}$" of subsets of the Cartesian product $X \times Y$ and a product measure "$\mu \times \nu$" on $\mathcal{M} \otimes \mathcal{N}$.

Why could such a thing be useful? On the one hand, the construction below will provide another way to construct the Lebesgue measure on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ (and more generally, $\mathbb{R}^n$) from the Lebesgue measure on $\mathbb{R}$.[14] On the other hand, think of an application in probability: Given a measure space $X = \{h, t\}$ representing a head-tail-experiment with a measure on $\mathcal{P}(X)$ determined by $\mu(h) = \mu(t) = 1/2$, we would like to consider n experiments (or perhaps even infinitely many), i.e. the space $X \times X \times ... \times X$ equipped with a corresponding product measure.

Given the setting in the first paragraph, we consider the algebra $\mathcal{M} \boxtimes \mathcal{N}$ of finite disjoint unions of rectangles $M \times N \subset X \times Y$ with $M \in \mathcal{M}$ and $N \in \mathcal{N}$.

Exercise 5.2. *Check that $\mathcal{M} \boxtimes \mathcal{N}$ is indeed an algebra.*
Hint: Use Problem 3 of Sheet 9.

We let $\mathcal{M} \otimes \mathcal{N}$ denote the σ-algebra generated by $\mathcal{M} \boxtimes \mathcal{N}$. A particular example is given by the Borel-algebras $\mathcal{M} = \mathcal{B}_{\mathbb{R}}$ and $\mathcal{N} = \mathcal{B}_{\mathbb{R}}$ for which $\mathcal{M} \otimes \mathcal{N} = \mathcal{B}_{\mathbb{R}^2}$ (Exercise).

We now define $\mu_0 : \mathcal{M} \boxtimes \mathcal{N} \to [0, \infty]$ by

$$\mu_0\left(\bigcup_{j=1}^{N} (M_j \times N_j)\right) = \sum_{j=1}^{N} \mu(M_j)\, \nu(N_j)$$

with the convention that $0 \cdot \infty = \infty \cdot 0 = 0$ on the right hand side.

Lemma 5.1. *μ_0 is a premeasure on $\mathcal{M} \boxtimes \mathcal{N}$.*

Proof. It is not hard to see that $\mu_0(\emptyset) = 0$ and that the difficulty is to prove that μ_0 is countably additive.

We prove that if a rectangle $M \times N = \bigcup_{j=1}^{\infty} (M_j \times N_j)$ is a countable union of disjoint rectangles, then we have

$$\mu_0(M \times N) = \sum_{j=1}^{\infty} \mu(M_j)\, \nu(N_j)\ . \tag{47}$$

Note that this statement implies that if a *finite disjoint union* of rectangles is a countable union of disjoint rectangles, then additivity holds (which is what we actually need to prove as any element of $\mathcal{M} \boxtimes \mathcal{N}$ is a finite disjoint union of rectangles).

To prove (47) we first note that

$$\chi_M(x)\, \chi_N(y) = \chi_{M \times N}(x, y) = \sum_{j=1}^{\infty} \chi_{M_j \times N_j}(x, y) = \sum_{j=1}^{\infty} \chi_{M_j} \times \chi_{N_j}(y)$$

[14]Recall that Example 5.2 provided an outline for abstractly constructing Lebesgue measure on $\mathbb{R}$ from a premeasure on the intervals. This construction, followed by the construction of Lebesgue measure on $\mathbb{R}^n$ via product measures is the presentation given in many books on measure theory.

and then integrate with respect to x to obtain – using the MCT – the identity

$$\mu(M)\,\chi_N(y) = \sum_{j=1}^{\infty} \mu(M_j)\,\chi_{N_j}(y)\ .$$

Integrating again, this time in y and using once more the MCT we obtain the desired (47). $\square$

Given Lemma 5.1, we can apply the Hahn extension theorem, Theorem 5.2 above to obtain a (unique if μ and ν are both σ-finite – why?) product measure on $\mathcal{M} \otimes \mathcal{N}$ which extends μ_0, which we denote by $\mu \times \nu$. In the case of Lebesgue measure on $\mathcal{M} = \mathcal{N} = \mathcal{B}_{\mathbb{R}}$ one checks that the product is the Lebesgue measure on $\mathcal{B}_{\mathbb{R}^2}$ that we defined via rectangles.

5.7 General Fubini theorem

Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces and $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$ be the (unique) product measure space defined in the previous section.

Given a measurable function f on $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$, we would like to understand whether the identity

$$\int_X \left[\int_Y f(x,y)\ d\nu(y) \right] d\mu(x) = \int_{X \times Y} f(x,y)\, d(\mu \times \nu)$$

between the iterated integrals and the integral with respect to the product measure holds (and whether the left hand side actually makes sense). This will be the general Fubini theorem for product measures.

To state it, we make the familiar (from the Lebesgue case, cf. Section 3.8.1) definitions of slices (also called sections): We define

- For $E \subset X \times Y$ a subset we define the slices/ sections of E

  $$E_x = \{y \in Y \mid (x,y) \in E\} \quad \text{and} \quad E_y = \{x \in X \mid (x,y) \in E\}.$$

- For a measurable function $f : X \times Y \to \mathbb{R}$ we define

  the slice corresponding to $y \in Y$ fixed as the function $f^y(x) := f(x,y)$,

  the slice corresponding to $x \in X$ fixed as the function $f_x(y) := f(x,y)$.

It is not too hard to show that if f is measurable on $X \times Y$ then f_x is measurable on Y and f_y is measurable on X. (Outline: Start with $\mathcal{F} = \{E \subset X \times Y \mid E_x \in \mathcal{N} \text{ for all } x \quad \text{and} \quad E_y \in \mathcal{M} \text{ for all } y\}$ and prove $\mathcal{F}$ is a σ-algebra containing the rectangles, so $\mathcal{F} \supset \mathcal{M} \otimes \mathcal{N}$. Now observe $(f_x)^{-1}(S) = (f^{-1}(S))_x$.) We are ready to state the general Fubini-Tonelli theorem:

Theorem 5.3 (Fubini-Tonelli). *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces and $(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \times \nu)$ be the (unique) product measure space defined in the previous section. Then*

1. *(Tonelli, i.e. assuming f measurable non-negative)*
   If $f : X \times Y \to [0, \infty]$ is measurable, then the functions

$$g(x) = \int_Y f_x \, d\nu \quad \text{and} \quad h(y) = \int_X f^y \, d\mu \tag{48}$$

   are measurable on X and Y respectively. Moreover the identity

$$\begin{aligned} \int_{X \times Y} f(x,y) \, d(\mu \times \nu) &= \int_X \left[\int_Y f(x,y) \, d\nu(y) \right] d\mu(x) \\ &= \int_Y \left[\int_X f(x,y) \, d\mu(x) \right] d\nu(y) \end{aligned} \tag{49}$$

   holds.

2. *(Fubini, i.e. assuming f integrable)*
   If $f : X \times Y \to \mathbb{R} \in L^1(X \times Y, \mu \times \nu)$), then $f_x \in L^1(Y, \mu)$ for a.e. x, $f^y \in L^1(X, \nu)$ for a.e. y. Moreover, the functions (48) are in $L^1(X, \mu)$ and $L^1(Y, \nu)$ respectively and the formula (49) holds.

We won't prove the general Fubini-Tonelli theorem here since we already went through the proof in the Lebesgue case. You can however easily deduce the second statement from the first.

A nice application of the general Fubini-Tonelli theorem is given in Question 6 on Example sheet 9.

6 The change of variables formula

In this section we prove the famous change of variables formula. The proof exhibits nicely many of the measure theoretic tools that we developed.

Theorem 6.1. *Let $\mathcal{U} \subset \mathbb{R}^n$ be open, $\varphi : \mathcal{U} \to \mathcal{V}$ be a C^1 diffeomorphism*[15] *with an open set $\mathcal{V} \subset \mathbb{R}^n$. Then*

1. *$f : \mathcal{V} \to \mathbb{R}$ is integrable if and only if the function $(f \circ \varphi) |\det D\varphi| : \mathcal{U} \to \mathbb{R}$ is integrable*
2. *The following change of variables formula holds:*

$$\int_{\mathcal{V}=\varphi(\mathcal{U})} f(y)\, dy = \int_{\mathcal{U}} (f \circ \varphi)(x) |\det D\varphi(x)| dx\,. \tag{50}$$

6.1 An example illustrating Theorem 6.1

Before we prove the theorem, let us illustrate it with a concrete example. Suppose we want to integrate $f : \mathbb{R}^2 \to \mathbb{R}$

$$f(y_1, y_2) = e^{-(y_1)^2 - (y_2)^2}$$

over $\mathbb{R}^2$.[16] One way to compute the integral is to go to polar coordinates (x_1, x_2) (which you should think of as $x_1 = r$ and $x_2 = \phi$)

$$\varphi : (x_1, x_2) \mapsto (y_1 = x_1 \cos(x_2), y_2 = x_1 \sin(x_2)) \,.$$

Now φ maps $(0, \infty) \times [0, 2\pi] \to \mathbb{R}^2 \setminus \{0\}$ but not diffeomorphically. However, φ is clearly a diffeomorphism of two open subsets of $\mathbb{R}^2$ when restricted to a map $\mathcal{U} := (0, \infty) \times (0, 2\pi) \to \mathcal{V} := \mathbb{R}^2 \setminus \{(y_1 \geq 0, y_2 = 0)\}$. Its differential is

$$D\varphi(x_1, x_2) = \begin{pmatrix} \cos x_2 & -x_1 \sin x_2 \\ \sin x_2 & x_1 \cos x_2 \end{pmatrix} ,$$

which has Jacobi determinant $|D\varphi(x_1, x_2)| = x_1 > 0$ on the domain considered. Note that $\mathcal{V} = \varphi(\mathcal{U})$ differs from $\mathbb{R}^2$ by a set of measure zero, so the left hand side of (50) is precisely the integral we want to compute, namely

$$\int_{\mathbb{R}^2} e^{-(y_1)^2 - (y_2)^2} dy_1 dy_2 = \int_{\mathcal{V}} e^{-(y_1)^2 - (y_2)^2} dy_1 dy_2 \,. \tag{51}$$

For the right hand side of (50) we note that $f \circ \varphi(x_1, x_2) = e^{-(x_1)^2}$ and hence

$$\int_{\mathcal{U}} e^{-(x_1)^2} x_1 dx_1 dx_2 = \int_0^\infty dx_1\ e^{-(x_1)^2} x_1 \int_0^{2\pi} dx_2 = \pi\,,$$

with the first step following from Fubini and the last step from a simple integration by substitution. We conclude that the desired integral (51) has the value π and hence, as a corollary, that $\int_{-\infty}^{\infty} dy\ e^{-y^2} = \sqrt{\pi}$.

[15] Recall that this means that φ is C^1, bijective and that its inverse is also a C^1 map.

[16] By Fubini we know that the result is $\left(\int_{-\infty}^{\infty} e^{-y^2} dy\right)^2$ and hence in particular that f is integrable but it is not immediate how to compute the one-dimensional integral!

6.2 A reformulation of Theorem 6.1

We next give an equivalent formulation of Theorem 6.1:

Theorem 6.2. *Let $\mathcal{U} \subset \mathbb{R}^n$ be open, $\varphi : \mathcal{U} \to \mathcal{V}$ be a C^1 diffeomorphism with an open set $\mathcal{V} \subset \mathbb{R}^n$. Then we have for any measurable set A in $\mathcal{U}$*

$$\mu_n (\varphi (A)) = \int_A |\det D\varphi (x)| d\mu_n \tag{52}$$

where μ_n denotes the n-dimensional Lebesgue measure.

Theorem 6.3. *Theorem 6.2 and Theorem 6.1 are equivalent, i.e. one can be deduced from the other.*

Proof. It is clear that Theorem 6.1 implies Theorem 6.2. Indeed, apply Theorem 6.1 with $f = \chi_{\varphi(A)}$ for any measurable set A in $\mathcal{U}$ and note $f \circ \varphi = \chi_A$. (If $\varphi (A)$ has infinite Lebesgue measure, then $\chi_{\varphi(A)} \geq 0$ and $\chi_A |\det D\varphi(x)| \geq 0$ both are not integrable by Theorem 6.1 and hence (50) holds as $+\infty = +\infty$.)

Now we claim that Theorem 6.2 also implies Theorem 6.1. To see this, assume Theorem 6.2 is true and let f be integrable (Direction 1). We decompose $f = f_+ + f_-$ and show the integrability of $f \circ \varphi |\det D\varphi|$ and (50) separately for f_+ and f_-.

Pick (f_n) a sequence of simple functions with $f_n \nearrow f_+$ (Theorem 2.7), say $f_n = \sum_{k=1}^{k_n} a_k \chi_{\varphi(A_k)}$ (why can we write f_n like this?). By the linearity of the integral we have that (50) holds for any of the f_n: Indeed, for each f_n (50) is simply a finite sum of identities (52). But since (50) holds for any f_n and since f_n increases to f_+ (which is integrable since f is) and $f_n \circ \varphi$ increases to $f_+ \circ \varphi$, the monotone convergence theorem implies that $f_+ \circ \varphi |\det D\varphi|$ is also integrable and that (50) holds for f_+. Of course f_- is treated entirely analogously.

We finally treat the case where $f_+ \circ \varphi |\det D\varphi|$ (instead of f) is assumed to be integrable (Direction 2). Applying what we have already shown with φ^{-1} we conclude $f \circ \varphi \circ \varphi^{-1} |\det D\varphi| |\det D\varphi^{-1}| = f$ is integrable and that (50) holds. ☐

6.3 Proof of Theorem 6.2

The proof will consist of one preliminary observations followed by 5 steps.

Preliminary Observation: Note that the desired identity (52) is an identity of measures on $\mathcal{U}$. Indeed, both maps

$$A \mapsto \mu (\varphi (A)) \tag{53}$$

and

$$A \mapsto \int_A |\det D\varphi (x)| dx \tag{54}$$

are easily seen to be countably additive.

Step 1. Observe that it suffices to prove the following *local* statement: Every point $p \in \mathcal{U}$ has an open neighbourhood $\mathcal{W}$ such that Theorem 6.2 holds for $\varphi|_{\mathcal{W}} : \mathcal{W} \to \varphi (\mathcal{W})$.

To see this, assume the local statement was proven. Cover $\mathcal{U}$ with such open neighbourhoods $\mathcal{W}_x$ in which the identity holds. Now because $\mathbb{R}^n$ has a countable basis of its topology we can select a *countable* subcover (W_i) (why?). Given this countable subcover $\mathcal{U} \subset \bigcup_{i=1}^{\infty} W_i$, let A be an arbitrary measurable set in $\mathcal{U}$. Define $\tilde{W}_i = A \cap W_i \setminus \bigcup_{j=1}^{i-1} W_i$. Then the $\tilde{W}_i$ are pairwise disjoint and their union is A. Since the identity of measures (52) holds for any $\tilde{W}_i$ and is countably additive, it holds for A itself.

Step 2. Theorem 6.2 (hence Theorem 6.1) holds if φ is a permutation of coordinates.

Step 3. Theorem 6.2 holds for $n = 1$, i.e. $\mathcal{U} \subset \mathbb{R}$.

To see this, we first note that if $A = [a, b]$ is an interval, then $\varphi(A) = [\varphi(a), \varphi(b)]$ and either $\varphi' > 0$ or $\varphi' < 0$ as φ is a C^1-diffeomorphism. In the first case

$$\varphi(b) - \varphi(a) = \int_a^b \varphi' dx$$

while in the second

$$\varphi(a) - \varphi(b) = \int_a^b -\varphi' dx$$

hence verifying $\mu(\varphi(A)) = \int_A |\varphi'(x)| dx$ when A is an interval. This holds for any finite interval (not necessarily closed) and by countable additivity of the identity (52), the latter also holds for intervals of infinite length. Finite disjoint unions of intervals in $\mathcal{U}$ form an algebra $\mathcal{A}$ of sets in $\mathcal{U}$. The two measures (53) and (54) agree on $\mathcal{A}$ hence define a premeasure on $\mathcal{A}$ which is moreover σ-finite (since one can write $\mathcal{U}$ as a countable disjoint union of intervals of finite length). Therefore, since both (53) and (54) extend the same σ-finite premeasure they must agree on the extension (cf. Q4 of Sheet 9).

Step 4. If Theorem 6.2 holds for $\psi : \mathcal{U} \to \mathcal{W}$ and $\rho : \mathcal{W} \to \mathcal{V}$, then it holds for the composition $\rho \circ \psi : \mathcal{U} \to \mathcal{V}$.

Indeed, note

$$\mu(\rho \circ \psi(A)) = \int_{\psi(A)} |\det D\rho(z)| dz$$

$$= \int_A |\det D\rho(\psi(x))| |\det D\psi(x)| dx = \int_A |\det D(\rho \circ \psi)(x)| dx \,, \tag{55}$$

where the first step follows from Theorem 6.2 holding for ρ, the second from Theorem 6.2 (hence Theorem 6.1) holding for ψ and the third from the chain rule and the properties of the determinant.

Step 5. We prove the local statement by induction on the dimension. The case $n = 1$ is Step 3. Our induction assumption is that the local statement (hence the global one) holds for $n - 1$ and we are considering a general diffeomorphism $\varphi : \mathcal{U} \to \mathcal{V} \subset \mathbb{R}^n$ locally near $p \in \mathcal{U}$.

We first claim that wlog we can restrict to $\varphi : (x_1, ..., x_n) \mapsto (\varphi_1(x), ..., \varphi_n(x))$ satisfying $\frac{\partial \varphi_1}{\partial x_1}(p) \neq 0$. Indeed, since for a general φ we have $\frac{\partial \varphi_1}{\partial x_i}(p) \neq 0$ for *some* $i \in \{1, 2, ..., n\}$ (in view Jacobian $\frac{\partial \varphi_i}{\partial x_j}$ having full rank) we can permute the coordinates x_i to achieve $\frac{\partial \varphi_1}{\partial x_1}(p) \neq 0$ (and use Steps 2+4).

We next claim that wlog we can even assume that φ keeps the first coordinate fixed, i.e. that φ has the form $\varphi : (t, x) = (t, \varphi_t(x))$. (Note that with this the map $\varphi_t : \mathcal{U}_t := \mathcal{U} \cap \{x_1 = t\} \to \{t\} \times \mathbb{R}^{n-1}$ is again a diffeomorphism in view of

$$D\varphi(t, x) = \begin{pmatrix} 1 & 0 & ... & 0 \\ ? & & & \\ ? & & D\varphi_t & \\ ? & & & \end{pmatrix}$$

and $\det D\varphi = \det D\varphi_t$.) To verify the claim, suppose one has established (52) for such φ. Then, by Step 2 one has also shown it for any φ which keeps one of the coordinates (not necessarily the first) fixed. Moreover, one can write a general φ near p as the composition of two diffeomorphisms each of which fixes at least one coordinate: Indeed, given general φ with $\frac{\partial \varphi_1}{\partial x_1}(p) \neq 0$, let $\psi : (x_1, ..., x_n) \to (\varphi_1(x), x_2, ..., x_n)$. Then ψ is a local diffeomorphism at p and hence $\rho = \psi \circ \varphi^{-1}$ is a diffeomorphism at $\varphi(p) \in \mathcal{V}$ which keeps the first coordinate fixed.

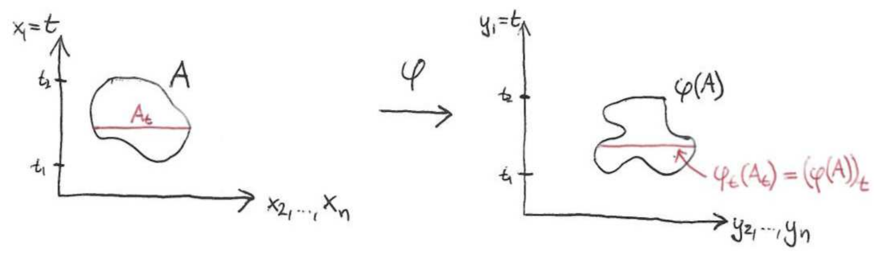
$$\begin{tikzcd} & \mathcal{W} \arrow[dr, "\rho^{-1}"] & \\ \mathcal{U} \arrow[ur, "\psi"] \arrow[rr, "\varphi"] & & \mathcal{V} \end{tikzcd}$$

We have $\varphi = \rho^{-1} \circ \psi$ and since both ρ and ψ fix at least one coordinate, Steps 2 and 4 imply (52) for general φ.

We finally prove the result for φ of the form $\varphi : (t, x) = (t, \varphi_t(x))$ using the induction assumption: First, by Fubini, we have

$$\mu_n(\varphi(A)) = \int_{\mathbb{R}^n} \chi_{\varphi(A)} dt dy_2 ... dy_n = \int_{\mathbb{R}} dt \mu_{n-1}\left((\varphi(A))_t\right) . \qquad (56)$$

Next observe that $(\varphi(A))_t = \varphi_t(A_t)$



and then use the induction assumption and Fubini again (together with $|\det D\varphi| = |\det D\varphi_t|$) to conclude

$$\begin{aligned} \mu_n(\varphi(A)) &= \int_{\mathbb{R}} dt \mu_{n-1}(\varphi_t(A_t)) = \int_{\mathbb{R}} dt \int_{A_t} |\det D\varphi_t| d\mu_{n-1} \\ &= \int_{\mathbb{R}^n} dt \ \chi_{A_t} |\det D\varphi_t| dx_2 ... dx_n = \int_A |\det D\varphi| d\mu_n . \qquad (57) \end{aligned}$$

7 Mastery Material: L^p-spaces

The material is this section is relevant only for the Mastery Question. The main points here are the Hölder and the Minkowski inequality, which you should know and be able to apply.

I leave some gaps in the proofs below which you should fill in on your own. If you need help, a good reference is provided by the first three pages of Section 6 in Folland's book (cf. Section 1.4).

7.1 Definition

Let $(X, \mathcal{M}, \mu)$ be a measure space and f a (say real-valued) measurable function on X. For $1 \leq p < \infty$ we define

$$\|f\|_{L^p} := \left(\int |f|^p d\mu(x) \right)^{1/p}$$

and

$$L^p\left(X, \mathcal{M}, \mu\right) = \{ f : X \to \mathbb{R} \mid f \text{ is measurable and } \|f\|_{L^p} < \infty \},$$

the space of measurable functions whose L^p-norm is finite. We sometimes write $L^p\left(X, \mu\right)$, or simply $L^p\left(X\right)$, or even just L^p for $L^p\left(X, \mathcal{M}, \mu\right)$ to simplify the notation provided no confusion arises. If one identifies two functions which are equal almost everywhere, the space $L^p(X)$ can be shown to be a complete vector space by adapting the proof we gave for $L^1\left(\mathbb{R}^n\right)$ in Section 3.6, Theorem 3.5.

In the following we fix a σ-finite measure space $(X, \mathcal{M}, \mu)$and write L^p for $L^p\left(X, \mathcal{M}, \mu\right)$ below.

7.2 The Hölder inequality

Theorem 7.1. *Let $1 < p < \infty$ and q satisfy $\frac{1}{p} + \frac{1}{q} = 1$. Let $f \in L^p$ and $g \in L^q$. Then the product $fg \in L^1$ with the inequality*

$$\|fg\|_{L^1} \leq \|f\|_{L^p} \|g\|_{L^q}.$$

Remark 7.1. *If the exponents p and q in the space L^p and L^q are related by $\frac{1}{p} + \frac{1}{q} = 1$, one says that the exponents are* **conjugate** *or* **dual** *to one another.*

Proof. Step 1. Prove the following inequality for $A, B \geq 0$ and $0 \leq \theta \leq 1$:

$$A^\theta B^{1-\theta} \leq \theta A + (1-\theta) B$$

Hint: Wlog $B \leq 0$. Setting $A = \tilde{A}B$ it suffices to prove $\tilde{A}^\theta \leq \theta \tilde{A} + (1-\theta)$ for $\tilde{A} \geq 0$, which can be done using elementary calculus.

Step 2. Note that we can assume $\|f\|_{L^p} \neq 0$ and $\|g\|_{L^q} \neq 0$ as otherwise $fg = 0$ a.e. and the inequality is trivially satisfied. Replacing f by $\frac{f}{\|f\|_{L^p}}$ and g by $\frac{g}{\|g\|_{L^q}}$ it suffices to prove $\|fg\|_{L^1} \leq 1$ for f having L^p-norm equal to 1 and g having L^q norm equal to 1.

Step 3. Set $A = |f(x)|^p$ and $B = |g(x)|^q$ and $\theta = \frac{1}{p}$, apply the inequality from Step 1 and integrate it to obtain $\|fg\|_{L^1} \leq 1$ as desired. $\square$

7.3 Minkowski's inequality

Theorem 7.2. *Let* $1 \leq p < \infty$ *and* $f, g \in L^p$*. Then* $f + g \in L^p$ *with the inequality*

$$\|f + g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p} .$$

Proof. The case $p = 1$ is easy (why?) so we let $p > 1$. To verify that $f + g \in L^p$ we first note that (why?)

$$|f(x) + g(x)|^p \leq 2^p \left(|f(x)|^p + |g(x)|^p\right) .$$

To prove the inequality, we observe

$$|f(x) + g(x)|^p \leq |f(x)| \cdot |f(x) + g(x)|^{p-1} + |g(x)| \cdot |f(x) + g(x)|^{p-1} .$$

If q is the conjugate exponent of p, that is $\frac{1}{p} + \frac{1}{q} = 1$, then we see that $|f(x) + g(x)|^{p-1}$ is in L^q (why?). Therefore we can apply Hölder's inequality on the right hand side and this leads to the result after some algebra using $(p-1)q = p$. $\square$