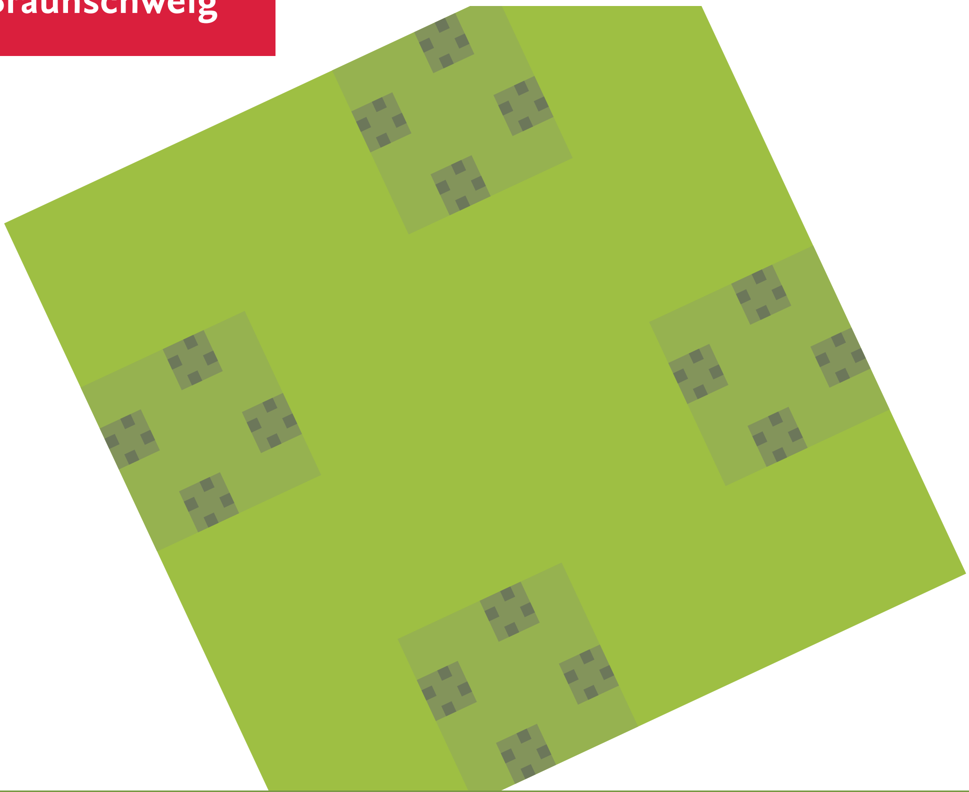


Geometric Measure Theory

July 23, 2014

Dirk Lorenz

Lecture notes for the course "Geometric Measure Theory" held in the summer term 2014 at TU Braunschweig. Mainly uncorrected - not proofread. Expect typos and unbalanced presentation. Handle with care. Please report errors to d.lorenz@tu-braunschweig.de.

Contents

1 Introduction

1.1 What shall "measure" mean?

The term "measure" can be understood as a generalization of the terms "length", "area" or "volume" and for these words one has a firm and intuitive understanding of what they shall mean: The length of some line is what we get, when we measure it with a ruler, the volume of some body can be obtained by measuring how much space it occupies, e.g. by putting it under water.

To make the notion of length or volume mathematically precise we would like to have a notion of length for subsets of $\mathbb{R}$, a notion of area for subsets of $\mathbb{R}^2$, and a notion of volume for subsets of $\mathbb{R}^3$. However, it turns out that this is not that simple. The simple intuition that the volume of a body should be the sum of the volume of its "atoms" leads to a problem here: A typical body consists of an infinite number of infinitely small points. Naturally, each point has volume zeros and hence the total volume would be of the kind "$0 \times \infty$" which is indeterminate in general. Another problem occurs as two bodies such as $[0,1]^3$ and $[0,2]^3$ have the same "number of points" in the sense that they are in bijection and consequently, should have the same volume although one is eight times as large as the other. Hence, one can break the set $[0,1]^3$ into a number of pieces (its points) and reassemble them in a different way to obtain the larger set $[0,2]^3$. One suspects that the infinite number of pieces that are used here is to blame for this pathological behavior.

This is not totally true: The celebrated **Banach-Tarski paradox** states that it is possible to disassemble the three dimensional unit ball into just five pieces that can be reassembled by translating and rotating them to form two disjoint copies of the unit ball. Here, the problem is not the infinite number of pieces but the fact that one can build incredibly complicated sets for which a reasonable definition of volume may fail.

However, mathematicians have found a workaround and we will present a short path towards a notion of measure in Chapter 2. In a nutshell, we define a notion of measure for all subsets, but we will abandon some strict requirements such as the additivity of measure for infinitely many disjoint subsets. However, we will see that the resulting measure always induces a fairly large number of sets for which the desired countable additivity holds and these sets will be called measurable sets.

Measures, as used above in the one, two or three dimensional case, already have a geometric flavor. They shall give the size of sets. But there are more geometric questions that are related to measures, e.g.: What is the "length of a curve" in two- or three-dimensional space"? What is the "area of a surface" in three-dimensional space? You may already know answers to these questions from a course on Anaylsis where one uses integrals in the case that the curves or surfaces are smooth. Here we are going to answer these questions in a pure "measure theoretic way" that does not assume any smoothness of the objects.

1.2 Basic notation

This section will be updated during the course...

We are going to work mainly with measures on n-dimensional euclidean space $\mathbb{R}^n$ (although a considerable amount of the theory would also work on general sets and an even larger part could be done for metric spaces). We call the **euclidean norm** in $\mathbb{R}^n$ simply by the name **absolute value** and denote it by $|\cdot|$. For $x \in \mathbb{R}^n$ and $r > 0$ we will denote the **open ball** of radius r around x by $B_r(x) = \{y : |x - y| < r\}$. We also use the **inner product** on $\mathbb{R}^n$ and denote $x \cdot y = \sum_{i=1}^n x_i y_i$.

For some set A we denote by $A^\complement$ its **complement**, by $\overline{A}$ its **closure** and by A° its **interior**. From these topological operations we can build the **topological boundary** $\partial A = \overline{A} \setminus A^\circ$.

This book has an index—however, this is not made with great care. You may use it, if you find it useful, but you should not expect to find every notion with precise pointers there.

1.3 Reading hints and literature

Although, the lectures will be in German, the lecture notes will be in English. On the one hand, this is due to the fact that the literature that I use to prepare the course is English. On the other hand, being able to read mathematics in English is a tremendously helpful skill, as there are many resources on the internet that are only available in English. If you have trouble to translate some words or phrase I suggest to do one of the following:

1. If it is just some usual English word: Use `http://dict.leo.org/`. I use this online dictionary all the time, and hence, you should arrive at the translation that I also had in mind.

2. If it is a technical term and Leo is not helpful: Check out the English Wikipedia page `http://en.wikipedia.org/` and then switch to German. Quite often this gives the correct hint.

Finally: No fear! You'll get used to English mathematics faster than you think.

I recommend the following list of books on geometric (and "standard non-geometric") measure theory for the course:

[AFP00] Luigi Ambrosio, Nicolo Fusco, and Diego Pallara. **Functions of Bounded Variation and Free Discontinuity Problems**. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. **A course in metric geometry**, volume 33. American Mathematical Society Providence, 2001.

[BMA06] Giuseppe Buttazzo, Gérard Michaille, and Hedy Attouch. **Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization**, volume 6. SIAM, 2006.

[EG92] Lawrence C. Evans and Ronald F. Gariepy. **Measure theory and fine properties of functions**. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[Fed69] Herbert Federer. **Geometric measure theory**. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[FL07] Irene Fonseca and Giovanni Leoni. **Modern methods in the calculus of variations: L^p spaces**. Springer Monographs in Mathematics. Springer, New York, 2007.

[KP08] Steven G. Krantz and Harold R. Parks. **Geometric integration theory**. Cornerstones. Birkhäuser Boston Inc., Boston, MA, 2008.

[Mat95] Pertti Mattila. **Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability**. The Cambridge studies in advanced mathematics. Cambridge University Press, 1995.

[Mor09] Frank Morgan. **Geometric measure theory**. Elsevier/Academic Press, Amsterdam, fourth edition, 2009. A beginner's guide.

[Rud87] Walter Rudin. **Real and complex analysis**. Tata McGraw-Hill Education, 1987.

[Tao11] Terence Tao. **An introduction to measure theory**, volume 126. American Mathematical Soc., 2011. available (with Errata) at http://terrytao.wordpress.com/books/an-introduction-to-measure-theory/.

Remark 1.3.1. Although we haven't really started yet, I should drop a warning here: The books I mention above follow (basically) two different roads into measure theory.

1. One road is to define measures in a way such that all sets have a measure assigned to them. Then, as a consequence, the measure does not have some desired additivity property. However, one can show that one always has a large enough set of subsets (the so-called measurable sets) such that these desired properties always hold on this set of subsets.

   This road may be seen as the fast lane. One should take this road if one is already confident that important measures, like the Lebesgue measure, exists and has a good understanding of the associated measurable sets.

2. The second road is by defining measures such that the set of measurable sets is part of their definition and also demanding an additivity property. Now one can work with additivity right away but the cost is that not all sets have a measure. To repair this small defect, one can construct an "outer measure" from the measure which then assigns a measure to all sets. However, additivity is lost for outer measures. The outer measures are then precisely the objects that have been called measured on the fast lane.

   This may be seen as a bike lane into measure theory. It seems that taking this lane takes a bit more effort, but it may well be worth it.

Books that follow the fast lane are [Mat95, Mor09, Fed69, EG92, AFP00, KP08].
Books that follow the second lane are [Tao11, FL07], and also [BMA06, BBI01] (although they do not give a detailed introduction into measure theory)

Note that there are many other resources to learn measure theory but be aware of the different routes. For example, the standard and excellent German books "Maß und Integral" by Brokate and Kersting, "Maß- und Integrationstheorie" by Elstrodt or the book with the same name by Bauer also take the "bike lane".

2 Measures

2.1 Carathéodory's construction

We head off to our trip into measure theory and start by defining what an outer measure shall be and postpone the definition of measure to Definition 2.1.6.

Definition 2.1.1. Let Ω be a set. A mapping $\mu : \mathfrak{P}(\Omega) \to [0, \infty]$ is called an **outer measure** if it is **countably subadditive**, i.e. for sets A_i, $i \in \mathbb{N}$ and A such that $A \subset \bigcup_{i\in\mathbb{N}} A_i$ it holds that

$$\mu(A) \leq \sum_{i\in\mathbb{N}} \mu(A_i).$$

A set A is called μ**-measurable** (or simply **measurable**) if for all $E \subset \Omega$ it holds that

$$\mu(E) = \mu(E \cap A) + \mu(E \cap A^{\complement}). \tag{2.1}$$

Remark 2.1.2.

1. Note that we used arithmetic in the set $[0, \infty]$. That is not a problem here because everything is non-negative and then we use the convention that $a + \infty = \infty$. Also, if a series $\sum_{i\in\mathbb{N}} a_i$ with $a_i \geq 0$ diverges, we set its value to ∞.

2. Condition (2.1) for measurability is called **Carathéodory's condition**. It may seem obscure at first glance, but will become clear later.

3. Observe that $E \cap A^{\complement} = E \setminus A$, hence we often write Carathéodory's condition as $\mu(E) = \mu(E \cap A) + \mu(E \setminus A)$.

4. Since the empty union is empty and the empty sum is zero, countable subadditivity implies $\mu(\emptyset) = 0$.

5. Subadditivity implies that outer measures are **monotone**, i.e. that for $A \subset B$ it holds that $\mu(A) \leq \mu(B)$.

Example 2.1.3 (Counting measure)**.** For every set Ω we define the **counting measure**, that is for $A \subset \Omega$ we set

$$|A| = \begin{cases} \text{number of elements in } A & A \text{ finite} \\ \infty & \text{else.} \end{cases}$$

As it is defined, an outer measure assigns a number to all subsets of the ground set Ω. However, in general not all subsets will be measurable. This may seem unsatisfying, but it is somehow unavoidable as we will see below. Another unfortunate thing is, that we only have countable **sub**additivity for an outer measure and not additivity, i.e. the measure of the disjoint union of sets may have a measure that is actually **smaller** that the sum of the measures of the parts. We want more: Something like (countable) **additivity** for disjoint sets seems desiriable. In the light of this concerns, the following theorem is a real blessing.

Theorem 2.1.4 (Carathéodory's Theorem)**.** *Denote by $\mathcal{A}_\mu$ the set of all measurable sets. Then $\mathcal{A}$ is a σ-**algebra**, i.e. for $A_i \in \mathcal{A}_\mu$, $i \in \mathbb{N}$ it holds that $A_i^{\complement}, \bigcup_{i\in\mathbb{N}} A_i, \bigcap_{i\in\mathbb{N}} A_i \in \mathcal{A}_\mu$. Moreover, μ is **countably additive** on $\mathcal{A}_\mu$, i.e. for disjoint $A_i \in \mathcal{A}_\mu$, $i \in \mathbb{N}$ it holds that*

$$\mu(\bigcup_{i\in\mathbb{N}} A_i) = \sum_{i\in\mathbb{N}} \mu(A_i).$$

Remark 2.1.5. Put differently, Carathéodory's Theorem says the following: An outer measure always defines a set of measurable sets that is closed under complements and countable unions and intersections (i.e. is a σ-algebra) and the outer measure behaves well (i.e. countable additive) on these sets. Some books use the term "measure" always in conjunction with an underlying σ-algebra and demand countable additivity on that σ-algebra right away. Note that we now use "let $A \in \mathcal{A}_\mu$" and "let A be measurable" as synonyms.

Proof. We proceed in several steps:

1. First we show that $\mathcal{A}_\mu$ is closed under complements and finite unions and intersections. If A is measurable, then obviously $\mu(E) = \mu(E \cap A^{\complement}) + \mu(E \cap A)$ which shows measurability of $A^{\complement}$. Now let A and B be measurable. We use (2.1) multiple times to see that for E it holds that

$$\begin{aligned} \mu(E) &= \mu(E \cap A) + \mu(E \setminus A), \\ \mu(E \cap A) &= \mu\big((E \cap A) \cap B\big) + \mu\big((E \cap A) \setminus B\big) \\ \mu\big(E \setminus (A \cap B)\big) &= \mu\big((E \setminus (A \cap B)) \cap A\big) + \mu\big((E \setminus (A \cap B) \setminus A\big) \\ &= \mu\big((E \cap A) \setminus B\big) + \mu(E \setminus A). \end{aligned}$$

   Plugging the two last equalities into the first one, we obtain

$$\mu(E) = \mu\big(E \cap (A \cap B)\big) + \mu\big(E \setminus (A \cap B)\big)$$

   as desired. Measurability of $A \cup B$ follows from $A \cup B = (A^{\complement} \cap B^{\complement})^{\complement}$, since we already proved that measurability is preserved under complements and intersections. Inductively, we obtain that $\mathcal{A}_\mu$ is closed under finite unions and intersections.

2. Second, we show that μ is finitely additive on $\mathcal{A}_\mu$. Let A and B be disjoint and measurable. Then it holds for all E that

$$\begin{aligned}\mu(E \cap (A \cup B)) &= \mu\Big(\underbrace{(E \cap (A \cup B)) \cap B}_{=E\cap B}\Big) + \mu\Big(\underbrace{(E \cap (A \cup B)) \cap B^{\complement}}_{=E\cap A,\ \text{since } A\cap B=\emptyset}\Big)\\ &= \mu(E \cap B) + \mu(E \cap A).\end{aligned}$$

   Setting $E = \Omega$ shows $\mu(A \cup B) = \mu(A) + \mu(B)$ and induction shows finite additivity.

3. Now we show countable additivity. Let A_i, $i \in \mathbb{N}$ be disjoint measurable sets and let $A = \bigcup_{i\in\mathbb{N}} A_i$. Then, of course, for any N: $A \supset \bigcup_{i=1}^N A_i$. By monotonicity and finite additivity we get

$$\mu(A) \geq \mu(\bigcup_{i=1}^N A_i) = \bigcup_{i=1}^N \mu(A_i).$$

   Since this holds for any N it also holds in the limit, i.e. $\mu(A) \geq \sum_{i\in\mathbb{N}} \mu(A_i)$. In other words: We have show that μ is countably super-additive. But since countable subadditivity is assumed, we conclude countable additivity.

4. Last we show that $\mathcal{A}_\mu$ is a σ-algebra and for that it remains to show that it is closed under countable unions. Let A_i, $i \in \mathbb{N}$ be measurable and set $A = \bigcup_{i\in\mathbb{N}} A_i$. Without loss of generality we assume that the A_i are pairwise disjoint*. We aim to show that for any E it holds that

$$\mu(E) \geq \mu(E \cap A) + \mu(E \cap A^{\complement})$$

   (since the reverse inequality follows from subadditivity). We have $E \cap A = \bigcup_{i\in\mathbb{N}}(E \cap A_i)$ and by countable subadditivity we get

$$\mu(E \cap A) + \mu(E \cap A^{\complement}) \leq \sum_{i\in\mathbb{N}} \mu(E \cap A_i) + \mu(E \cap A^{\complement}).$$

   Now set $B_i = A_1 \cup \cdots \cup A_i$ and since B_i is measurable (which we know from the first step), we get by finite additivity

$$\mu(E \cap B_n) = \sum_{i=1}^n \mu(E \cap A_i).$$

   Observe that $B_n \subset A$ and $A^{\complement} \subset B_n^{\complement}$. By monotonicity of μ and measurability of B_n we get

$$\begin{aligned}\sum_{i=1}^n \mu(E \cap A_i) + \mu(E \cap A^{\complement}) &= \mu(E \cap B_n) + \mu(E \cap A^{\complement})\\ &\leq \mu(E \cap B_n) + \mu(E \cap B_n^{\complement}) = \mu(E).\end{aligned}$$

*If they were not, we would take $\tilde{A}_i = A_i \setminus (A_1 \cup \cdots \cup A_{i-1})$ instead.

$\square$

Carathéodory's Theorem shows that any outer measure induces a σ-algebra of measurable sets on which the measure really behaves countably additive. As is sounds natural that a measure shall only be able to measure measurable sets, the definition of the underlying σ-algebra is part of the definition of a measure:

Definition 2.1.6. Let $\mathcal{A}$ be a σ-algebra. Then $\mu : \mathcal{A} \to [0, \infty]$ is called a **measure** if μ is countable additive. We call the triple $(\Omega, \mathcal{A}, \mu)$ a measure space.

Hence, we can rewrite the statement of Carathéodory's Theorem as follows:

> Every outer measure μ is a measure on the σ-algebra of μ-measurable sets.

Remark 2.1.7. Let me reiterate Remark 1.3.1: Some of the books I recommended do not use the term "outer measure" and use the term measure instead. Hence, in these books all sets have a measure but measures are not always countably additive. It is a matter of taste, which approach is used. We will mainly work with measures, however, we will not go into detail of constructions of underlying σ-algebras, nor spend time on more of their properties.

Remark 2.1.8. In some sense the converse of Carathéodory's Theorem is also true: If we have a measure μ on a σ-algebra $\mathcal{A}$ on a set Ω we can construct

$$\mu^*(E) = \inf\{\sum_{n=1}^{\infty} \mu(A_n) \ : \ E \subset \bigcup_{n \in \mathbb{N}} A_n, \ A_n \in \mathcal{A}\}.$$

One can verify that μ^* is indeed an outer measure that has at least $\mathcal{A}$ as measurable sets. We will do this very construction in the next section to construct Lebesgue measure.

In the following we will usually only speak of "measures" even if an outer measure is considered, but sometimes we add this term for clarity. If outer measure are applied to measurable sets only, we usually omit the term. We also do not make use of the notion μ^* for the outer measure associated to μ and usually denote both objects with μ.

2.2 Lebesgue measure

One fundamental motivation for measure theory was to define a notion of measure or "length" for subset of the real line $\mathbb{R}$. In this section we illustrate what Carathéodory's construction gives in this situation and we will end up with the so-called Lebesgue measure.

We start by defining the length of intervals: There shall be no doubt that the unit interval $[0,1]$ shall have length $1^{\dagger}$. Similarly, for $a > b$ there shall be no doubt that the length of $[a,b]$ shall be $b-a$—this is just how we measure the length of a rod with a rule. Since single points $\{a\}$ should have length 0 we should also assign $]a,b[= b-a$.

Now, let us build an outer measure (in the sense of Defintion 2.1.1) λ out of this. To do this, we need to define $\lambda(A)$ for any set $A \subset \mathbb{R}$ in a way that intervals keep their length and that λ will be countably subadditive. A simple idea is the following: cover a given set by a countable number of intervals and take the "minimum-total-length" that is needed. We will call the construction "Lebesgue measure" although it will be an outer measure only. Anyway, Carathéodory's Theorem shows that it is a measure on some σ-algebra.

Definition 2.2.1 (One dimensional Lebesgue measure). For $A \subset \mathbb{R}$ we define the **Lebesgue measure** as

$$\lambda(A) = \inf\{\sum_{j=1}^{\infty}(b_j - a_j) \ : \ A \subset \bigcup_{j=1}^{\infty}]a_j, b_j[\}.$$

This is a valid definition and $\lambda(A)$ is indeed well defined for any set $A \subset \mathbb{R}$; but note that $\lambda(A)$ may well be ∞. Of course, we always have $\lambda(A) \geq 0$.

Lemma 2.2.2. *The one dimensional Lebesgue measure is countably subadditive.*

Proof. Let A and A_i such that $A \subset \bigcup_{i=1}^{\infty} A_i$. Now take arbitrary countable covers of the A_i by intervals $I_j^i =]a_j^i, b_j^i[$, i.e. $A_i \subset \bigcup_{j=1}^{\infty} I_j^i$. Consequently A is covered by all I_j^i, i.e. $A \subset \bigcup_{i=1}^{\infty}\bigcup_{j=1}^{\infty} I_j^i$. Hence

$$\lambda(A) \leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\lambda(I_j^i) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} b_j^i - a_j^i.$$

Since $(I_j^i)_j$ was an arbitrary cover of A_i we can take the infimum over all covers of the A_i's on the right hand side and arrive at

$$\lambda(A) \leq \sum_{i=1}^{\infty}\lambda(A_i)$$

as desired. □

Example 2.2.3. 1. It holds that $\lambda([a,b]) = b-a$: For any $\epsilon > 0$ we have $[a,b] \subset]a-\epsilon, b+\epsilon[$, i.e. the interval $]a-\epsilon, b+\epsilon[$ is an open cover of $[a,b]$ and hence,

$$\lambda([a,b]) \leq b-a+\epsilon.$$

† Well, this is basically just our convention. We could assign a different positive number here and this could be seen similar to the choice between meters and yards.

Consequently, $\lambda([a,b]) \leq b-a$.

To show the opposite inequality, note that we only need to consider finite open covers (as $[a,b]$ is compact we can always reduce the value inside the infimum by considering a finite subcover). Now assume that $[a,b]$ is covered by the intervals $]a_1,b_1[$, ..., $]a_n,b_n[$. We argue by induction that then it holds that $b-a \leq \sum_{i=1}^n b_i - a_i$: For $n=1$ the result is clear - assume that it holds for all n up to some N. Consider $[a,b] \subset \bigcup_{i=1}^{N+1}]a_i,b_i[$. At least one interval contains a and hence, we can (up to renumbering) suppose that $a \in]a_{N+1},b_{N+1}[$. Also we can suppose that $b_{N+1} < b$ (otherwise we would have $b-a \leq b_{N+1}-a_{N+1}$ already). We then have $[b_{N+1},b] \subset \bigcup_{i=1}^{N}]a_i,b_i[$ and by the induction hypothesis $b - b_{N+1} \leq \sum_{i=1}^N (b_i - a_i)$. Consequently, it holds

$$b-a \leq (b-b_{N+1}) + (b_{N+1}-a) \leq \sum_{i=1}^{N} (b_i-a_i) + (b_{N+1}-a_{N+1})$$

as desired.

2. As an immediate consequence we get that $\lambda(]a,b[) = b-a$: for all $\epsilon > 0$ we have $[a+\epsilon, b-\epsilon] \subset]a,b[\subset [a,b]$.

3. Another immediate consequence is the following: If we decompose an interval $I=[a,b]$ into n subintervals such that $I_1 = [a,a_1]$, $I_2 = [a_1,a_2]$, ...,$I_n = [a_{n-1},b]$ then clearly $\lambda(I) = b-a = (b-a_{n-1}) + \cdots + (a_2 - a_1) + (a_1 - a) = \lambda(I_n) + \cdots \lambda(I_1)$.

4. The Lebesgue measure of $\mathbb{R}$ is infinite: By monotonicity and the previous points we have for all $r>0$ that$\lambda(\mathbb{R}) \geq \lambda([-r,r]) = 2r$.

5. Since $\emptyset \subset [-\epsilon,\epsilon]$ we have $\lambda(\emptyset) = 0$ and similarly we see that $\lambda(\{x\}) = 0$ for every $x \in \mathbb{R}$.

6. By countable subadditivity of λ we get that all countable sets have Lebesgue measure zero, especially $\lambda(\mathbb{Q}) = 0$.

Exercise 1. Calculate the measure of $\mathbb{Q}$ directly from Definition 2.2.1 – compare with Example 2.2.3, 6..

Hint: You have to construct open intervals that cover $\mathbb{Q}$ such that the sum of their length is arbitrarily small! You can conclude that $\mathbb{Q}$ has open neighborhoods of arbitrarily small measure. This may seem surprising in view of the fact that $\mathbb{Q}$ is dense in $\mathbb{R}$, i.e. the topological closure of $\mathbb{Q}$ is $\mathbb{R}$. But this merely says that topological and measure theoretical notions do not always mix properly.

Exercise 2. Show that the Lebesgue measure is **translation invariant**, i.e. for every $x \in \mathbb{R}$ and $E \subset \mathbb{R}$ it holds that $\lambda(E) = \lambda(x+E)$.

The next classical example shows that the one-dimensional Lebesgue measure from Definition 2.2.1 is not countable additive. An immediate consequence is, that not all subsets of $\mathbb{R}$ are Lebesgue measurable.

Example 2.2.4 (Non countable additivity)**.** We work in the unit interval $[0,1]$ and consider "cosets of $\mathbb{Q}$" i.e. sets of the form $A_x = (x+\mathbb{Q}) \cap [0,1]$. Note that these coset may be equal for different x's, as e.g. for $q \in \mathbb{Q}$: $A_q = q + \mathbb{Q} = \mathbb{Q} = A_0$. Also note that two cosets either coincide or are disjunct: If two cosets A_x and A_y share a point z, then $x - y = (x-z) - (y-z)$ is rational and hence $A_x = A_y$. Now we define an uncountable subset of $[0,1]$ as follows: Consider the set of all different cosets and for each pick one representative. Let E be the union of all these representatives[‡].

Now define

$$X = \bigcup_{q\in\mathbb{Q},\ q\in[-1,1]} (q+E).$$

Of course $[0,1] \subset X \subset [-1,2]$ (the first inclusion is true since each $y \in [0,1]$ is in some coset A_x but then $x - y$ is some rational number $q \in [-1,1]$ and consequently $y = q + x$, i.e. $y \in q + E \subset X$).

Our aim is, to show that

$$\lambda(X) \neq \sum_{q\in\mathbb{Q},\ q\in[-1,1]} \lambda(q+E).$$

By monotonicity of λ we know that $1 \leq \lambda(X) \leq 3$. By translation invariance of λ we obtain that

$$\sum_{q\in\mathbb{Q},\ q\in[-1,1]} \lambda(q+E) = \sum_{q\in\mathbb{Q},\ q\in[-1,1]} \lambda(E).$$

We don't know the value of $\lambda(E)$ but either way: if it would be zero, then the whole sum would be zero and if it would not be zero that the sum would be infinite—in any case it would not be between 1 and 3[§].

Since we now know that not all sets are measurable for the Lebesgue measure, the next natural question to ask is: "What sets are Lebesgue measurable?".

Lemma 2.2.5. *Every closed interval is Lebesgue measurable.*

[‡]A different way to construct E is as follows: Consider the equivalence relation $x \sim y$ if $x - y \in \mathbb{Q}$ on $[0,1]$. Build E by collecting one representative of each equivalence class. The construction of the set involves the set theoretic "Axiom of Choice" that is subject to some debate among mathematicians—on the one hand, it can be stated in a totally intuitive way but on the other hand, it has some puzzling consequences.

[§]The first case would also violate the countable subadditivity of λ and hence we know that $\lambda(E) > 0$.

Proof. Let $I = [a, b]$, $E \subset \mathbb{R}$. We need to show Carathéodory's criterion $\lambda(E) = \lambda(E \cap I) + \lambda(E \cap I^\complement)$. It is clear by subadditivity that "$\leq$" holds and hence, we only need to show "$\geq$".

Let $\epsilon > 0$. There is a cover of E by countably many intervals $I_i = [a_i, b_i]$ such that

$$\sum_{i=1}^{\infty} b_i - a_i \leq \lambda(E) + \epsilon.$$

Now we cut from interval I_i the parts away that lay outside of I, i.e. we write $I_i = I_{1,i} + I_{2,i} + I_{3,i}$ with $I_{1,i} \subset I$ and $I_{2/3,i} \subset I^\complement$[¶]. It holds (cf. Example 2.2.3, 3.) $\lambda(I_i) = \lambda(I_{1,i}) + \lambda(I_{2,i}) + \lambda(I_{3,i})$. We obtain

$$\lambda(E) + \epsilon \geq \sum_{i=1}^{\infty} \lambda(I_{1,i}) + \sum_{i=1}^{\infty} \lambda(I_{2,i}) + \sum_{i=1}^{\infty} \lambda(I_{3,i})$$

The intervals $\{I_{1,i} \,:\, i \in \mathbb{N}\}$ cover $E \cap I$ while the rectangles $\{I_{j,i} \,:\, j = 2, 3,\ i \in \mathbb{N}\}$ cover $E \cap I^\complement$. Hence, we have

$$\lambda(E \cap I) \leq \sum_{i=1}^{\infty} \lambda(I_{1,i}), \qquad \lambda(E \cap I^\complement) \leq \sum_{i=1}^{\infty} \lambda(I_{2,i}) + \sum_{i=1}^{\infty} \lambda(I_{3,i})$$

and this gives

$$\lambda(E) + \epsilon \geq \lambda(E \cap I) + \lambda(E \cap I^\complement).$$

Since this holds for any $\epsilon > 0$ we also have $\lambda(E) \geq \lambda(E \cap I) + \lambda(E \cap I^\complement)$ as desired. $\square$

Since all closed intervals are Lebesgue measurable, but also the set of all measurable sets forms a σ-algebra, we have a whole lot more measurable sets:

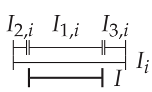
- Countable unions and intersections of closed intervals are Lebesgue measurable.
- Open intervals are Lebesgue measurable, as it holds that

$$]a, b[= \bigcup_{n \in \mathbb{N}} [a + \tfrac{1}{n}, b - \tfrac{1}{n}].$$

In view of our findings, the next lemma is of importance:

Lemma 2.2.6. *Let Ω be a set and $\mathcal{A}$ be a subset of $\mathfrak{P}(\Omega)$. Then there exists a smallest σ-algebra $\sigma(\mathcal{A})$ such that $\mathcal{A} \subset \sigma(\mathcal{A})$.*

¶ A picture of the situation looks like this:



Note that it may well happen that one or more of the intervals $I_{1/2/3,i}$ is empty.

Proof. Let $\mathcal{M}$ be the family of all σ-algebras in Ω that contain $\mathcal{A}$. Since $\mathfrak{P}(\Omega)$ is such a σ-algebra, $\mathcal{M}$ is not empty. Now let $\sigma(\mathcal{A})$ be the intersection of all elements of $\mathcal{M}$. Clearly $\mathcal{A} \subset \sigma(\mathcal{A})$ and also that $\sigma(\mathcal{A})$ lies in every σ-algebra that contains $\mathcal{A}$. It remains to show that $\sigma(\mathcal{A})$ is a σ-algebra itself:

Let $A_i \in \sigma(\mathcal{A})$, $i \in \mathbb{N}$. Take any σ-algebra $\mathcal{A}^* \in \mathcal{M}$. Then $A_i \in \mathcal{A}^*$ and hence $\bigcup A_i \in \mathcal{A}^*$. Since this is true for any $\mathcal{A}^* \in \mathcal{M}$, we conclude that $\bigcup A_i \in \mathcal{A}$. The other requirements for being a σ-algebra are show similarly. □

The σ-algebra $\sigma(\mathcal{A})$ is called the σ-algebra generated by $\mathcal{A}$.

We conclude: The Lebesgue measurable sets include the whole σ-algebra that is generated by the open and closed intervals. This σ-algebra is called the **Borel σ-algebra** on $\mathbb{R}$ and its elements are called **Borel sets.** In other words: We know that all Borel sets are Lebesgue measurable. Borel sets are many. Countable unions and intersections of open and closed sets. Also countable unions and intersection of countable unions and intersection of open and closed set. And even countable unions and intersection of countable unions and intersections of countable unions and intersections of open and closed sets. You see where this is going. In view of this huge abundance of Borel sets it seems bizarre that there still exists sets that are not Lebesgue measurable, but we have constructed such a beast in Example 2.2.4.

Remark 2.2.7. A final minor pathology is that there exists Lebesgue measurable sets that are not Borel measurable but their construction is even more cumbersome than the construction of a non-Lebesgue-measurable set and also necessarily use the Axiom of Choice, see this link.

Another important class of Lebesgue measurable sets are Null sets:

Lemma 2.2.8. *In $N \subset \mathbb{R}$ fulfills $\lambda(N) = 0$ then N is Lebesgue measurable and we call N a (Lebesgue)* ***Null set****.*

Proof. For any E we have by monotonicity of λ that $0 \leq \lambda(E \cap N) \leq \lambda(N) = 0$. Hence, we have $\lambda(E) \geq \lambda(E \cap N^{\complement}) = \lambda(E \cap N^{\complement}) + \lambda(E \cap N)$. Since λ is subadditive we also have the reversed inequality and Carathéodory's condition is fulfilled. □

Remark 2.2.9. Note that for the above lemma the word "Lebesgue" does not play any role. Null sets for general outer measures are always measurable.

Remark 2.2.10. Within measure theory Null sets are "negligible sets" as they are too small to play a role. As we will see in the chapter about integration, it is often possible and helpful to discard Null sets altogether. This is captured by the notion of "almost everywhere". More precisely: If μ is an (outer) measure on Ω, then we say that some property holds μ **almost everywhere** in Ω (μ a.e) if there exists a Null set N such that the property holds on $\Omega \setminus N$.

Exercise 3. We define the famous **Cantor set** as follows: Define $C_1 = [0, 1/3] \cup [2/3, 1]$ (i.e. delete the middle third from the unit interval). Now define, recursively, $C_{n+1} = \frac{C_n}{3} \cup \left(\frac{2}{3} + \frac{C_n}{3}\right)$, see the following sketch:

In other words, in each iteration we remove the middle third of each of the remaining intervals. The limiting set is denoted by C; an explicit formula for this set is

$$C = \bigcap_{m=1}^{\infty} \bigcap_{k=0}^{3^{m-1}-1} \left([0, \tfrac{3k+1}{3^m}] \cup [\tfrac{3k+2}{3^m}, 1]\right).$$

1. Show that the Lebesgue measure of C is zero.
2. Show that the set C has uncountably many elements, i.e. it has as many elements as $[0, 1]$.

   Hint: Show that C contains precisely these numbers that have a ternary description that does not use the digit 1. Use this to construct a bijection of C onto $[0, 1]$.

Furthermore one can deduce at least finite additivity of Lebesgue measure for sets that are is a sense "far enough apart from each other".

Lemma 2.2.11. *Let $A, B \subset \mathbb{R}$ such their **distance***

$$\operatorname{dist}(A, B) = \inf\{|x - y| \ : \ x \in A, \ y \in B\}$$

is positive. Then it holds that $\lambda(A \cup B) = \lambda(A) + \lambda(B)$.

Note that A and B need not to be measurable here!

Proof. Subadditivity implies that $\lambda(A \cup B) \leq \lambda(A) + \lambda(B)$ holds and we only need to proof the converse inequality. The inequality is trivial if $\lambda(A \cup B)$ is infinite and hence, as assume that is has finite Lebesgue measure (and by monotonicity the same holds for A and B). Let $\epsilon > 0$ and cover $A \cup B$ by a countable number of open intervals $I_n =]a_n, b_n[$ such that

$$\sum_{i=1}^{\infty} b_i - a_i \leq \lambda(A \cup B) + \epsilon.$$

Without loss of generality, we could assume that that all intervals have a length smaller than $\operatorname{dist}(A, B)$ (if this would not be the case, we could cut every interval in half until they are small enough—the sums would stay the same and we would still cover $A \cup B$).

Since the intervals are so small, each interval only intersects either A or B. Hence, we can divide the family I_n into two parts I_n^A and I_n^B where the I_n^A cover A and I_n^B cover B. Then we have $\lambda(A) \leq \sum_n b_n^A - a_n^A$ and $\lambda(B) \leq \sum_n b_n^B - a_n^B$. Summing gives $\lambda(A) + \lambda(B) \leq \sum_n b_n - a_n$ and thus, $\lambda(A) + \lambda(B) \leq \lambda(A \cup B) + \epsilon$ which shows the assertion. □

Exercise 4. In this exercise we show (in a non-constructive way) that there are in fact many non-Borel sets that are Lebesgue measurable.

Denote by $\mathfrak{c}$ the cardinality of the real numbers (which is the same as the cardinality of the powerset of the natural numbers). It holds that the Borel σ-algebra is also generated by the set of all sets of the form $]a - r, a + r[$ for rational a and r. One can infer from this that the cardinality of the Borel σ-algebra is also $\mathfrak{c}$, i.e. there are $\mathfrak{c}$ Borel sets.

Show that there are at least as many Lebesgue measurable set as the powerset of $\mathbb{R}$ has elements, i.e. at least $2^{\mathfrak{c}}$ elements.

Hint: Use the Cantor set and Lemma 2.2.8.

We close this section with two remarks:

Remark 2.2.12 (n-dimensional Lebesgue measure)**.** One can make the whole construction of the Lebesgue measure on $\mathbb{R}$ that we did here also for $\mathbb{R}^n$. One does basically the same things as we did, but now works with "boxes" instead of intervals, i.e. one considers sets of the form $[a_1, b_1] \times \cdots \times [a_n, b_n]$ and assigns them a measure of $(b_1 - a_1) \cdots (b_n - a_n)$. Basically only the proof of Lemma 2.2.5 becomes it bit more cumbersome but structurally the same. We denote the Lebesgue-measure on $\mathbb{R}^n$ by λ^n or also by λ.

Remark 2.2.13 (Borel σ-algebra)**.** The Borel σ-algebra can be defined for any topological space, since the only thing one needs is a notion of open and closed sets. Consequently, the Borel σ-algebra also exists for all metric spaces. We now see that geometry (defined in terms of distances, i.e. by the metric) somehow naturally touches the notion of measures.

2.3 General measures

Let us recall the achievement of Carathéodory's construction: If we have a set Ω and a subadditive set function μ (an outer measure) that assigns a non-negative value (possibly ∞) to every subset of Ω, there always is σ-algebra of measurable sets on which the outer measure is countably additive, i.e. indeed a measure.

Example 2.3.1 (Dirac measures)**.** An important example of a measure is the so-called Dirac measure: For any set Ω, a point $a \in \Omega$ and any subset $A \subset \Omega$ define

$$\delta_a(A) = \begin{cases} 1 & a \in A \\ 0 & a \notin A. \end{cases}$$

Example 2.3.2. Assuming for a moment that you already know the Lebesgue integral and consider a Lebesgue measurable set $\Omega \subset \mathbb{R}^n$ and a Lebesgue integrable function $f : \Omega \to [0, \infty[$ such that $\int_\Omega f(x)\mathrm{d}x < \infty$. Then $\lambda_f(A) = \int_A f(x)\mathrm{d}x$ is also a measure.

One can construct new measures out of old ones by several constructions:

Lemma 2.3.3. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and let $E \subset \Omega$. Then $\mathcal{A}\llcorner E = \{A \cap E \,:\, A \in \mathcal{A}\}$ is a σ-algebra. Also*

$$\mu\llcorner E(A) = \mu(E \cap A)$$

is a measure (on both $(\Omega, \mathcal{A})$ and $(E, \mathcal{A} \cap E)$).

*The σ-algebra $\mathcal{A}\llcorner E$ is the **restriction of** $\mathcal{A}$ to E and the measure $\mu\llcorner E$ is the **restriction of** μ to E.*

Proof. If $B \in \mathcal{A}\llcorner E$, then $B = A \cap E$ for some $A \in \mathcal{A}$. Since $A^\complement \in \mathcal{A}$, also $E \setminus B = (\Omega \setminus B) \cap E \in \mathcal{A}\llcorner E$. If $B_1, \cdots \in \mathcal{A}\llcorner E$, then $B_k = A_k \cap E$ for some $A_k \in \mathcal{A}$ and hence $\cup B_k = \cup A_k \cap E \in \mathcal{A}\llcorner E$.

The proof that $\mu\llcorner E$ is a measure is a good **Exercise**. $\square$

Example 2.3.4 (Restriction of Lebesgue measure)**.** On every subset $\Omega \subset \mathbb{R}^n$ we can define the Lebesgue measure by restriction $\lambda\llcorner\Omega$. Usually, the set is clear from the context and we denote the resulting measure again by λ.

If we have measure μ on some set Ω_1 and some mapping $f : \Omega_1 \to \Omega_2$ can define a measure on Ω_2:

Definition 2.3.5. For a measure μ on Ω_1 and a mapping $f : \Omega_1 \to \Omega_2$ we define the **push-forward** of μ by f as: For $B \subset \Omega_2$

$$f\#\mu(B) = \mu(f^{-1}(B)).$$

For further use, we define some finer notions of measures: Recall that on any metric space (Ω, d) (i.e. on every set on which one can especially talk about open and closed sets) we have the Borel σ-algebra (cf. Remark 2.2.13).

Definition 2.3.6. A measure μ on a metric space (Ω, d) is called

- **finite** if $\mu(\Omega) < \infty$,
- **locally finite** if for every $x \in \Omega$ there is $r > 0$ such that $\mu(B_r(x)) < \infty$.
- a **Borel measure**, if all Borel sets are measurable,
- a **regular Borel measure** if it is a Borel measure and if for every A there exists a Borel set B such that $A \subset B$ and $\mu(A) = \mu(B)$ and
- a **Radon measure** if it is a Borel measure and

1. $\mu(K) < \infty$ for compact K,
2. $\mu(V) = \sup\{\mu(K) \, : \, K \subset V$ is compact$\}$ for open V and
3. $\mu(A) = \inf\{\mu(V) \, : \, A \subset V, \, V$ open$\}$ for all A.

Example 2.3.7. 1. The Lebesgue measure is a Radon measure on every set $\Omega \subset \mathbb{R}^n$.

2. The Dirac measure from Example 2.3.1 is a Radon measure on every metric space Ω.

3. The counting measure from Example 2.1.3 is a regular Borel measure on every metric space but it is a Radon measure only if every compact set is finite.

Lemma 2.3.8. *If μ is a regular Borel measure and A is measurable with $\mu(A) < \infty$, then $\mu \llcorner A$ is also a regular Borel measure.*

Proof. Let B be a Borel set with $A \subset B$ and $\mu(A) = \mu(B)$. Then $B = A \cup (B \setminus A)$ and the union is disjoint. Since μ is additive on measurable sets, we have $\mu(B) = \mu(A) + \mu(B \setminus A)$ and hence, $\mu(A \setminus B) = 0$. Let $C \subset \Omega$ and let D be a Borel set with $B \cap C \subset D$ and $\mu(B \cap C) = \mu(D)$. Then $C \subset D \cup (\Omega \setminus B) =: E$. Also

$$\begin{aligned}(\mu \llcorner A)(E) &\leq \mu(B \cap E) = \mu(B \cap D) \leq \mu(D) \\ &= \mu(B \cap C) = \mu(A \cap C) = (\mu \llcorner A)(C).\end{aligned}$$

Thus, $(\mu \llcorner A)(E) = (\mu \llcorner A)(C)$ and hence μ is a regular Borel measure. □

The following approximation theorem holds for regular Borel measures:

Theorem 2.3.9. *Let μ be a regular Borel measure on Ω, A be measurable and $\epsilon > 0$. Then:*

1. *If $\mu(A) < \infty$, then there is a closed set $C \subset A$ such that $\mu(A \setminus C) < \epsilon$.*

2. *If there are open sets $V_1, V_2, \ldots$ such that $A \subset \bigcup_{i \in \mathbb{N}} V_i$ and $\mu(V_i) < \infty$, then there is an open set V such that $A \subset V$ and $\mu(V \setminus A) < \epsilon$.*

Proof. 1. We may assume that $\mu(\Omega) < \infty$ (otherwise, Lemma 2.3.8 would allow to consider $\mu \llcorner A$ instead of μ). First consider, that A is a Borel set. Denote by $\mathcal{A}$ the family of all subsets of Ω such that for every $\epsilon > 0$ there exists a closed set C and an open set V such that $C \subset A \subset V$ and $\mu(V \setminus C) < \epsilon$. One can show that $\mathcal{A}$ is a σ-algebra which also contains the Borel sets. Consequently, we have established the first claim for Borel sets.

If A is now measurable with $\mu(A) < \infty$ but no necessarily Borel, then there is a Borel set B such that $\mu(A) = \mu(B)$. Then $\mu(B \setminus A) = 0$ and

$B \setminus A$ is contained in some Borel set D with $\mu(D) = 0$. Thus, $E = B \setminus D$ is also a Borel set with $E \subset A$ and $\mu(A \setminus E) = 0$. Since the claim holds for E, it also holds for A.

2. We apply the first claim to the sets $V_i \setminus A$ to obtain closed sets $C_i \subset V_i \setminus A$ such that $\mu((V_i \setminus A) \setminus C_i) \leq \epsilon/2^i$. Then $A \subset V = \bigcup_{i\in\mathbb{N}}(V_i \setminus C_i)$ and V is open and fulfills $\mu(V \setminus A) \leq \sum_{i\in\mathbb{N}} \epsilon/2^i = \epsilon$.

□

The following corollary can be proved as an **Exercise**.

Corollary 2.3.10. *A measure μ on $\mathbb{R}^n$ is a Radon measure if and only it is a locally finite, regular Borel measure.*

A measure may put no measure on certain subsets, or, the other way round, the mass of a measure may be concentrated within a certain set. This is formalized by the notion of the **support** of a measure.

Definition 2.3.11. Let μ be a Borel measure on a metric space Ω. The **support** of μ is

$$\mathrm{supp}\mu = \{x \in \Omega \,:\, r > 0 \implies \mu(B_r(x)) > 0\}.$$

The following exercise shows that the support is always a closed set.

Exercise 5. Show that the support of a Borel measure μ is the largest closed set C such that for all open sets U with $C \cap U \neq \emptyset$ it holds that $\mu(C \cap U) > 0$.

Exercise 6. 1. Let $f : X \to Y$ be a Borel function between two measurable spaces (i.e. inverse images of Borel sets are Borel sets). Show that $f\#\mu$ is a Borel measure if μ is a Borel measure.

2. Let $f : X \to Y$ be a continuous function between two metric spaces and μ be a Radon measure on X that has compact support. Show that $f\#\mu$ is a Radon measure and that $\mathrm{supp}(f\#\mu) = f(\mathrm{supp}\mu)$.

Example 2.3.12. 1. The support of the Lebesgue measure λ on $\mathbb{R}$ is $\mathbb{R}$.

2. The support of a Dirac measure δ_a is $\{a\}$.

3. For the measure λ_f from Example 2.3.2 it holds that $\mathrm{supp}\lambda_f = \mathrm{supp} f$.

3 Integration, Lebesgue spaces

Second stop of our trip: integration. Arguably one of the most prominent uses of measures. Recall that Riemann's definition of the integral of a function $f : [a, b] \to \mathbb{R}$ relies on the idea of partitioning the domain of f into subintervals and then approximate the function by simpler functions that are constant on these subintervals. One drawback of this approach is, that it gets much more complicated if the function is defined on a more complicated domain in $\mathbb{R}^n$, say. The measure theoretic approach to integration, developed by Lebesgue, is, to partition the range of $f : \Omega \to \mathbb{R}$ into subintervals (which is always possible, regardless of the structure of Ω). By taking inverse images of these partitions one obtains a partition of Ω which can be used to approximate the function. But to make this approach work, one would need that these inverse images form reasonable sets, i.e., measurable sets.

3.1 Measurable functions

Definition 3.1.1. Let μ be a measure on a set Ω. A **μ-measurable realvalued function** is a realvalued function that is defined μ a.e. on Ω such that the every upper level sets of f are measurable, i.e. there is a set $D \subset \Omega$ such that

1. $f : D \to \mathbb{R}$
2. $\mu(\Omega \setminus D) = 0$
3. $\{x \ : \ f(x) > t\} = f^{-1}(]t, \infty[)$ is μ measurable for every t.

In the following it will be useful that realvalued functions f are allowed to take the values $+\infty$ and $-\infty$.

Definition 3.1.2. We define the **extended real numbers** as

$$\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}.$$

The ordering on $\overline{\mathbb{R}}$ is an extension of the ordering on $\mathbb{R}$ by adding that for all $r \in \mathbb{R}$ it holds that $-\infty \leq r \leq \infty$.

For the definition of **extended real valued measurable functions** we also demand the inverse images of $\pm\infty$ to be measurable.

We also extend the addition and multiplication but we make some exceptions to avoid contradictions. The rules are captured in these tables:

$+$	$-\infty$	$x \in \mathbb{R}$	$+\infty$
$+\infty$	undefined	$+\infty$	$+\infty$
$-\infty$	$-\infty$	$-\infty$	undefined

$\cdot$	$-\infty \leq x < 0$	0	$0 < x < +\infty$
$+\infty$	$-\infty$	undefined	$+\infty$
$-\infty$	$+\infty$	undefined	$-\infty$

Remark 3.1.3. 1. We often say **measurable function** instead of μ-measurable realvalued function.

2. For the purpose of integration, two functions that agree a.e. are considered equivalent—this indeed defines an equivalence relation.

3. Without loss of generality, one can assume that a measurable function is defined for every $x \in \Omega$. In fact, let $f : D \to \mathbb{R}$ and $y_0 \in \mathbb{R}$. We can define $\tilde{f} : \Omega \to \mathbb{R}$ by setting $\tilde{f} = f$ on D and $\tilde{f}(x) = y_0$ for all $x \in \Omega \setminus D$. Then $\tilde{f}$ is also μ measurable and is equal to f a.e.

   However, usually it is not very helpful to think of measurable functions to be defined everywhere.

4. Often, a measurable function is defined between two measurable spaces and one demands that the inverse images of measurable sets under the mapping are measurable. In our case, this reads that a function is measurable if the inverse image of every Borel set in $\mathbb{R}$ is measurable, which is equivalent to our definition, see below.

5. **Complex valued measurable functions** or $\mathbb{R}^n$**-valued measurable functions** are defined component-wise, is via measurability of the real and imaginary parts or the components, respectively.

Exercise 7. Show that the definition of measurable functions does not change if one replaces the upper level sets by $\{x : f(x) \geq t\}$ or the respective lower level sets $\{x : f(x) < t\}$, $\{x : f(x) \leq t\}$.

Remark 3.1.4. Since one often considers sets of the form $\{x : f(x) > t\}$ and the like, people use the abbreviation $\{f > t\}$ and similar notations. As of now it is not clear if these notes will also use this convention. I tend to use the long form but may get tired of doing so at some point.

Exercise 8. Show that upper or lower continuous as well as continuous functions $f : \mathbb{R}^n \to \mathbb{R}$ are measurable.

It is hard to construct functions that are not measurable since many operations preserve measurability.

Lemma 3.1.5. *If f and g are measurable functions then, λf ($\lambda \in \mathbb{R}$) and $f + g$ are measurable, $f \cdot g$ is measurable,* $\max(f, g)$ *and* $\min(f, g)$ *are measurable. Also, for* *a sequence* (f^n) *of measurable functions* $\limsup_n f^n$ *and* $\liminf_n f^n$ *(understood pointwise) are measurable.*

Proof. Measurability of λf is clear. Now observe that f is measurable, if the sets $\{x \,:\, f(x) > a\}$ are measurable for $a \in \mathbb{Q}$ and for the measurablity of $f + g$ observe that

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

Now note that also f^2 is measurable, since $\{x \,:\, f^2(x) > t\} = \{x \,:\, f(x) > \sqrt{t}\} \cup \{x \,:\, f(x) < -\sqrt{t}\}$. Consequently, fg is measurable, since $fg = \frac{1}{2}((f+g)^2 - f^2 - g^2)$.

The proof that $\max(f, g)$ and $\min(f, g)$ are measurable is a simple **Exercise**. To see that $\liminf f^n$ and $\limsup f^n$ are measurable, note that $\liminf_{n\to\infty} f^n = \sup_{n>1} \inf_{r>n} f^n$ and $\limsup_{n\to\infty} f^n = \inf_{n>1} \sup_{r>n} f^n$ (and you may fill in the details as an **Exercise**). □

Interestingly, the set of all measurable functions, as complicated as it may be, is easily characterized by the following principle. We call in **principle of monotonicity**, sometimes it is also called **monotone class theorem**.

Theorem 3.1.6 (Principle of monotonicity). *Let* $(\Omega, \mathcal{A})$ *be a measurable space (i.e. a set* Ω *with a* σ*-algebra* $\mathcal{A}$*) and let* $\mathcal{K}$ *a set of functions* $f : \Omega \to [0, \infty]$ *that obeys the following properties:*

1. $f, g \in \mathcal{K}$, $\alpha, \beta > 0$, *then* $\alpha f + \beta g \in \mathcal{K}$
2. $f_1, f_2, \dots \in \mathcal{K}$, $f_1 \leq f_2 \leq \cdots$, *then* $\sup f_n \in \mathcal{K}$,
3. *for* $A \in \mathcal{A}$ *it holds* $\mathbb{1}_A \in \mathcal{K}$.*

Then $\mathcal{K}$ *contains all non-negative measurable functions on* Ω.

Proof. We take some measurable $f : \Omega \to [0, \infty]$ and show that it belongs to $\mathcal{K}$. Since f is measurable, the sets

$$A_{n,k} = \{x \,:\, k/2^n < f(x) \leq (k+1)/2^n\}$$

are in $\mathcal{A}$. Hence, by 1. and 3. we conclude that the functions

$$h_n = \sum_{k=1}^{n2^n} \frac{k}{2^n} \mathbb{1}_{A_{n,k}} + n\mathbb{1}_{\{x \,:\, f(x)=\infty\}}$$

belong to $\mathcal{K}$. By construction we have that $f_1 \leq f_2 \leq \cdots$ and $\sup f_n = f$ and by 2. we conclude $f \in \mathcal{K}$. □

*By $\mathbb{1}_A$ be denote the **characteristic function** of A with domain Ω i.e.

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}.$$

Now we state a fundamental theorem about measurable functions that is one of the three basic principles of analysis as formulated by Littlewood. These principles are, is rough form,

1. Every measurable set in $\mathbb{R}$ is almost a finite union of intervals.
2. Every measurable function is almost continuous.
3. Every pointwise convergent sequence of measurable functions is almost uniformly convergent.

The first statement is related to the regularity of measure and we not treat it here. The second and third statement go under the names of Egoroff's and Luzin's Theorem.

Theorem 3.1.7 (Egoroff's Theorem). *Let μ be a finite measure on Ω (i.e. $\mu(\Omega) < \infty$ and let (f_n) be a sequence of measurable functions such that*

$$f^n(x) \to f(x) \text{ for almost every } x \in \Omega.$$

Then, for every $\epsilon > 0$ there exists a measurable set $A_\epsilon \subset \Omega$ with $\mu(A_\epsilon) > \mu(\Omega) - \epsilon$ such that f^n converges to f uniformly on A_ϵ.

Proof. Let $\delta > 0$. Pointwise convergence at some x means that there is an integer $M(\delta, x)$ such that $|f^n(x) - f(x)| < \delta$ for $j > M(\delta, x)$. For some $N \in \mathbb{N}$ we define $S(\delta, N) = \{x : M(\delta, x) \le N\}$. Clearly, $S(\delta, N)$ is nondecreasing with increasing N and δ. Also $S(\delta, N)$ is measurable, since $S(\delta, N) = \bigcup_{M=1}^{N} \bigcap_{j>M} B_j$ with $B_j = \{x : |f^j(x) - f(x)| < \delta\}$. Now, we define $S(\delta) = \bigcup_{N\in\mathbb{N}} S(\delta, N)$ and observe that almost every x is in some $S(\delta, N)$ and consequently $\mu(S(\delta)) = \mu(\Omega)$.

Thus, for every $\delta > 0$ and $\tau > 0$ there is an N such that $\mu(S(\delta, N)) > \mu(\Omega) - \tau$. Now let $\delta_1 > \delta_2 > \cdots$ be a decreasing Null sequence and let N_j be such that $\mu(S(\delta_j, N_j)) > \mu(\Omega) - \epsilon/2^j$. Set $A_\epsilon = \bigcap_{j\in\mathbb{N}} S(\delta_j, N_j)$. By construction, $f^n \to f$ uniformly on A_ϵ.

Finally, by de Morgan's law $\left(\bigcap_{j\in\mathbb{N}} S(\delta_j, N_j)\right)^{\complement} = \bigcup_{j\in\mathbb{N}} S(\delta_j, N_j)^{\complement}$ and the measure of the right hand side is less the ϵ. □

Luzin's theorem may be proved later in the course or could be found in [Tao11, Theorem 1.3.28].

3.2 Integration

It is simpler to define integration for non-negative functions (as one avoids to treat situations of $\infty - \infty$). Hence, the positive and negative part of a function will be helpful:

Definition 3.2.1. For $f : \Omega \to \overline{\mathbb{R}}$ we define the **positive part** to be the function $f^+ : \Omega \to [0, \infty]$ defined by

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) > 0 \\ 0 & \text{else} \end{cases}$$

and the **negative part** to be the function $f^- : \Omega \to [0, \infty]$ defined by

$$f^-(x) = \begin{cases} -f(x) & \text{if } f(x) < 0 \\ 0 & \text{else.} \end{cases}$$

Obviously we have $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

Now we are close to defining the integral of measurable functions. The idea is basically the same as for Riemann integration. but there is a slight twist. For Riemann integration you break up the **domain** of the function into simple parts, approximate the function you want to integrate by simpler functions that are constant of the simple parts and the perform some limit. In the Lebesgue theory of integration you break up the **range** of the function into simple parts—the rest is basically similar to Riemann integration. Note that a partition of the range of function also implies a partition of the domain of the function by taking inverse images. However, these inverse images are usually not very simple anymore. But here we are only interested to the size (resp. measure) of the inverse images and this is the point where the notion of measure we developed really pays off.

Let's get down to business. We approximate measurable functions by **simple functions** in the following sense:

Definition 3.2.2. A **simple function** is a function that is a linear combination of characteristic functions, i.e. it can be written as

$$f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$$

for some numbers $a_i \in \mathbb{R}$ and sets A_i.

Clearly, a simple function is measurable, if the sets A_i are measurable.

For a non-negative and simple function it is pretty obvious, what the integral should be: Each summand of the for $a_i \mathbb{1}_{A_i}$ brings an additional area under the graph of the form $a_i \mu(A_i)$ and the total area under the graph should be the sum of these contributions. The next lemma states is a reformulation of the principle of monotonicity (Theorem 3.1.6) and states that it is possible to approximate measurable functions with simple functions which then allows us to extend the notion of the integral to non-negative measurable functions.

Lemma 3.2.3. *Let μ be a measure on Ω and let $f : \Omega \to [0, \infty]$ be measurable. Then there exists a sequence of measurable and simple functions $h_n : \Omega \to [0, \infty]$ such that*

1. *the sequence (h_n) is non-decreasing, i.e. $0 \le h_1 \le h_2 \le \cdots$, and*
2. *converges pointwise to f, i.e. $\lim_{n\to\infty} h_n(x) = f(x)$ for all $x \in \Omega$.*

Proof. We can use exactly the same construction as in the proof of the principle of monotonicity (Theorem 3.1.6). □

The integral of a measurable function is now defined via approximation by simple functions:

Definition 3.2.4. Let μ be a measure on a set Ω and let $f : \Omega \to \overline{\mathbb{R}}$ be measurable. Then, the **integral** of f with respect to μ is denoted by

$$\int f \mathrm{d}\mu \quad \text{or} \quad \int_\Omega f(x)\mathrm{d}\mu(x) \quad \text{(or variations thereof)}$$

and defined as follows:

1. If f is non-negative and simple then we set (with the notation from Defintion 3.2.2)
$$\int f \mathrm{d}\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$
2. If f is non-negative, we set
$$\int f \mathrm{d}\mu = \sup\{\int h \mathrm{d}\mu \ : \ 0 \le h \le f,\ h \text{ simple and measurable}\}.$$
3. In case at least one of $\int f^+ \mathrm{d}\mu$ and $\int f^- \mathrm{d}\mu$ is finite, we set
$$\int f \mathrm{d}\mu = \int f^+ \mathrm{d}\mu - \int f^- \mathrm{d}\mu.$$
4. In case both $\int f^+ \mathrm{d}\mu$ and $\int f^- \mathrm{d}\mu$ are infinite, the value $\int f \mathrm{d}\mu$ is undefined.

Definition 3.2.5.
1. To integrate a function f over a subset A of its domain we multiply f by the characteristic function of A, i.e.
$$\int_A f \mathrm{d}\mu = \int f \mathbb{1}_A \mathrm{d}\mu.$$
2. If f is complex valued or has values in $\mathbb{R}^n$ we integrate real and complex parts, respectively the components, independently.
3. If $\int |f| \mathrm{d}\mu < \infty$ we say that f is (μ-)**absolutely integrable** or μ**-summable**.

Exercise 9. Show that $\int_A f \mathrm{d}\mu = \int f \mathrm{d}(\mu \llcorner A)$.

Example 3.2.6. 1. Taking the Lebesgue measure λ^n we obtain the celebrated Lebesgue integral. We often use the familiar notation

$$\int f \mathrm{d}\lambda^n = \int_{\mathbb{R}^n} f(x)\mathrm{d}x \text{ and } \int_A f \mathrm{d}\lambda^n = \int_A f(x)\mathrm{d}x.$$

2. Series are a special case of integrals: Take the counting measure, which we denote by #, on $\mathbb{N}$ and a function $f : \mathbb{N} \to \mathbb{R}$ (which is nothing else, then a sequence $f_n = f(n)$). Then $\int f \mathrm{d}\# = \sum_{n\in\mathbb{N}} f_n$ (note that the value of the series is demanded to be independent on the ordering of summation).

Some simple facts about the integral are collected in the following theorem.

Theorem 3.2.7. *Let μ be a measure on a set Ω and let $f, g : \Omega \to [0, \infty]$ measurable.*

1. *If $A \subset \Omega$ is measurable and $f(x) = 0$ for almost every $x \in A$, then*

$$\int_A f \mathrm{d}\mu = 0.$$

2. *If $A \subset \Omega$ is a Null set, then*

$$\int_A f \mathrm{d}\mu = 0.$$

3. *If $0 \le c < \infty$, then*

$$\int (cf)\mathrm{d}\mu = c \int f \mathrm{d}\mu.$$

4. *If $f \le g$, then*

$$\int f \mathrm{d}\mu \le \int g \mathrm{d}\mu.$$

5. *If $A \subset B \subset \Omega$ are measurable, then*

$$\int_A f \mathrm{d}\mu \le \int_B g \mathrm{d}\mu$$

The **Proof** is basically evident from the definition of the integral and you may want to sketch the details for yourself.

Note that the expected additivity $\int (f + g)\mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu$ is not part of the above list. The reason is, that it is not a straightforward consequence of the definition, but requires some preparation:

Lemma 3.2.8. *Let μ be a measure on Ω and let $f : \Omega \to [0, \infty]$ be measurable. If $0 \le h_1 \le h_2 \le \cdots \le f$ is a sequence of simple and measurable functions that converges pointwise to f, then it holds that*

$$\lim_{n\to\infty} \int h_n \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

Proof. The inequality $\lim_{n\to\infty} \int h_n \mathrm{d}\mu \leq \int f \mathrm{d}\mu$ is immediate from the definition. To obtain the reverse inequality, let g be a simple and measurable function with $0 \leq g \leq f$ and write $g = \sum_i a_i \mathbb{1}_{A_i}$. Since g may not be below some h_n, we make it a bit smaller by multiplying it with some $0 < c < 1$ and set for each $m \in \mathbb{N}$ set

$$E_m = \{x \,:\, c\,g(x) \leq h_m(x)\} \text{ and } g_m = c\,g\,\mathbb{1}_{E_m}.$$

For $m \leq n$ we have $g_m \leq h_n$ and by Theorem 3.2.7, 4. we obtain

$$\int g_m \mathrm{d}\mu \leq \lim_{n\to\infty} \int h_n \mathrm{d}\mu.$$

Finally, we note that for each $i = 1, \ldots, k$ the sets $A_i \cap E_m$ increase to A_i as $m \to \infty$ and hence $\mu(A_i) = \lim_{m\to\infty} \mu(A_i \cap E_m)$. Thus, it holds that

$$c \int g \mathrm{d}\mu = \int c\,g \mathrm{d}\mu = \lim_{m\to\infty} \int g_m \mathrm{d}\mu \leq \lim_{n\to\infty} h_n \mathrm{d}\mu.$$

Since, c and g are arbitrary, the result follows. □

Now we get additivity of the integral quite easily:

Theorem 3.2.9. *Let μ be a measure on a set Ω an the $f, g : \Omega \to [0, \infty]$ be measurable. Then it holds that*

$$\int (f+g) \mathrm{d}\mu = \int f \mathrm{d}\mu + \int g \mathrm{d}\mu.$$

Proof. The result clearly holds for simple functions. The general case follows by approximating f and g with a non-decreasing sequence of simple functions using Lemma 3.2.3 and Lemma 3.2.8. □

3.3 Convergence theorems and consequences

Now we state the celebrated theorems that allow to interchange limits with integrals that made the abstract theory of integration so successful:

Theorem 3.3.1 (Convergence theorems for integrals)**.** *Let μ be a measure on a set Ω.*

1. ***Monotone convergence:*** *If $0 \leq f_1 \leq f_2 \leq \cdots$ are measurable, then*

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int \lim_{n\to\infty} f_n \mathrm{d}\mu.$$

2. ***Fatou's Lemma:*** *If $f_1, f_2, \ldots$ are measurable and non-negative, then*

$$\liminf_{n\to\infty} \int f_n \mathrm{d}\mu \geq \int \liminf_{n\to\infty} f_n \mathrm{d}\mu.$$

3. ***Reversed Fatou's Lemma:*** *If $f_1, f_2, \ldots$ are measurable and there exists a μ-summable function $g \geq f_n$, then*

$$\limsup_{n\to\infty} \int f_n \mathrm{d}\mu \leq \int \limsup_{n\to\infty} f_n \mathrm{d}\mu.$$

4. ***Dominated convergence:*** *Let $f_1, f_2, \ldots$ be measurable functions such that $f_n \to f$ μ-a.e. and there exists a μ-summable function g such that $|f_n| \leq g$ (i.e. g dominates the sequence (f_n)), then*

$$\lim_{n\to\infty} \int |f_n - f| \mathrm{d}\mu = 0 \quad \textit{and} \quad \lim_{n\to\infty} \int f_n \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

Proof. 1. The proof of the monotone convergence result can be done along the lines of the proof of Lemma 3.2.8.

2. Fatou's Lemma follows from monotone convergence, by setting $g_n = \inf_{k\geq n} f_k$ and noting that g_n is monotonically increasing to $\liminf_{n\to\infty} f_n$. By the monotonicity of the integral (Lemma 3.2.7, 4.) we get

$$\int g_n \mathrm{d}\mu \leq \int f_n \mathrm{d}\mu$$

and by monotone convergence, we have

$$\int \liminf_{n\to\infty} f_n \mathrm{d}\mu = \lim_{n\to\infty} \int g_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu.$$

3. The reverse Fatou follows by applying Fatou to the sequence $g - f_n$ which is integrable and non-negative by assumption.

4. For dominated convergence start by noting that also f is measurable and dominated by g. Also,by assumption $|f - f_n| \to 0$ a.e.

Now observe that by monotonicity and linearity of the integral

$$|\int f \mathrm{d}\mu - \int f_n \mathrm{d}\mu| = |\int f - f_n \mathrm{d}\mu| \leq \int |f - f_n| \mathrm{d}\mu.$$

Now, the reversed Fatou is applicable to $|f - f_n|$ since $|f - f_n| \leq |f| + |f_n| \leq 2g$ and leads to

$$\limsup_{n\to\infty} \int |f - f_n| \mathrm{d}\mu \leq \int \limsup_{n\to\infty} |f - f_n| \mathrm{d}\mu = 0.$$

This shows that $\lim_{n\to\infty} \int |f - f_n| \mathrm{d}\mu$ exists and equals zero as desired.

□

A direct consequence of monotone convergence, is the following corollary on the interchange of the integral with series for non-negative functions:

Corollary 3.3.2. *Let $f_n \geq 0$ be measurable. Then $\int \sum_{n=1}^{\infty} f_n \mathrm{d}\mu = \sum_{n=1}^{\infty} \int f_n \mathrm{d}\mu$.*

Example 3.3.3. The assumptions in the above theorems can not be weakened in general. Here are some counterexamples for the Lebesgue measure on $\mathbb{R}$:

1. Monotonically decreasing non-negative functions do not work: Consider the $f_n = \frac{1}{n}\mathbb{1}_{[0,n]}$. This sequence converges pointwise (also uniformly) to $f \equiv 0$, their integrals are $\int f_n \mathrm{d}\lambda = 1$ and do not converge to zero. Note that this example also shows that the inequality in Fatou's lemma can be strict, i.e. we may loose some mass in the limit.

2. Basically the same counterexample shows that Fatou's lemma does not work without the non-negativity assumption: Take $f_n = -\frac{1}{n}\mathbb{1}_{[0,n]}$. Again, we have uniform convergence to $f \equiv 0$ (and hence, also $\liminf f_n \equiv 0$) with zero integral but $\int f_n \mathrm{d}\lambda = -1 \leq 0$.

3. That domination can not be omitted for the reverse can be shown with the same sequence $f_n = \frac{1}{n}\mathbb{1}_{[0,n]}$ as in the first point. We have also $\limsup f_n \equiv 0$ but $\limsup \int f_n \mathrm{d}\lambda = 1 \geq 0$. Note that the smallest g that dominates all f_n is $g(x) = \sum_{n=1}^{\infty} \frac{1}{n}\mathbb{1}_{[n-1,n]}$ which is, however, not summable as its integral in $+\infty$. This is also shows that domination can not be omitted for to dominated convergence.

Remark 3.3.4. Note that the convergence theorems for integrals also apply to the case of the counting measure # on $\mathbb{N}$. Hence, they also allow to interchange limits with series if the conditions are fulfilled.

With the help of the convergence theorems one can not only prove many things but also compute the integral of some strange functions:

Exercise 10. This exercise shows that Riemann integrals do not go as well with limits as the Lebesgue integral does.

1. Show that the function $l(x) = -\log(|x|)$ is integrable over any compact interval (both Riemann and Lebesgue).

2. Let r_k we a countable dense subset of $[0,1]$ (e.g. an enumeration of the rationals in $[0,1]$. Consider the functions $f_n(x) = \sum_{k=1}^{n} 2^{-k} l(x - r_k)$ and the pointwise limit f (for $n \to \infty$). Show that the integrals of f_n are bounded independently of n.

3. Show that f is Lebesgue integrable (by using a limit theorem for the Lebesgue integral) with a finite integral.

4. Show that f is unbounded on every open subset of $[0,1]$ and hence, is not Riemann integrable (in the sense that it does not have a finite integral).

One may also integrate functions according to the “layer cake formula”. The proof uses monotone convergence.

Theorem 3.3.5. *Let μ be a regular Borel measure and $f \geq 0$ be measurable. Then it holds that*

$$\int f \mathrm{d}\mu = \int_0^\infty \mu(\{x \,:\, f(x) > t\})\mathrm{d}t.$$

Proof. We use the approximating sequence h_n similar to the one in the proof of the principle of monotonicity (Theorem 3.1.6) but now we write

$$h_n = 2^{-n} \sum_{k=1}^\infty \mathbb{1}_{\{x \,:\, f(x) > k/2^n\}}.$$

With Corollary 3.3.2 we get with the help of the **ceiling function** $\lceil x \rceil$ (which is the smallest integer not less than x) that

$$\int h_n \mathrm{d}\mu = 2^{-n} \sum_{k=1}^\infty \mu(\{x \,:\, f(x) > k/2^n\}) = \int_0^\infty \mu(\{x \,:\, f(x) > \lceil t2^n \rceil / 2^n\})\mathrm{d}t.$$

We have, by construction that $0 \leq h_1 \leq h_2 \leq \cdots$ and $h_n \to f$ and for the right hand side it holds that $\lceil t2^n \rceil / 2^n \to t\,(n \to \infty)$ and $\{x \,:\, f(x) > \lceil t2^n \rceil / 2^n\}) \to \{x \,:\, f(x) > t\}$ (from the inside). The assertion follows from the regularity of the measure and monotone convergence. □

As another application, we prove a transformation formula for integrals. We use the push-forward of a measure to express the formula.

Theorem 3.3.6 (Transformation of integrals)**.** *Let μ be a measure on Ω and let $\phi : \Omega \to \Omega'$ be measurable (with respect to some σ-algebra on Ω'). Then it holds for measurable $f : \Omega' \to [0, \infty]$ that*

$$\int f \mathrm{d}(\phi\#\mu) = \int f \circ \phi \mathrm{d}\mu.$$

Proof. Consider the set $\mathcal{K} = \{f \geq 0 \,:\, f \text{ measurable}, \int f(\phi\#\mu) = \int f \circ \phi \mathrm{d}\mu\}$. The set $\mathcal{K}$ contains all characteristic functions of measurable sets: Indeed for measurable $A' \in \Omega'$ it holds that

$$\int \mathbb{1}_{A'} \mathrm{d}(\phi\#\mu) = (\phi\#\mu)(A') = \mu(\phi^{-1}(A'))$$

and since $\mathbb{1}_{A'} \circ \phi = \mathbb{1}_{\phi^{-1}(A')}$ we get

$$\int \mathbb{1}_{A'} \mathrm{d}(\phi\#\mu) = \int \mathbb{1}_{\phi^{-1}(A')} \mathrm{d}\mu.$$

Additivity of the integral (Theorem 3.2.9) and monotone convergence show that $\mathcal{K}$ fulfills all requirements for the principle of monotonicity (Theorem 3.1.6) and hence $\mathcal{K}$ contains all non-negative measurable functions which shows the claim. □

Remark 3.3.7. The transformation formula is related to the well known formula

$$\int_{\phi(\Omega)} f(y)\mathrm{d}y = \int_\Omega f(\phi(x))|\det D\phi(x)|\mathrm{d}x$$

for diffeomorphisms ϕ by the following result: For $\Omega, \Omega' \subset \mathbb{R}^n$, a diffeomorphism $\phi : \Omega \to \Omega'$ and a Borel set $A \subset \Omega$ it holds that

$$\lambda(\phi(A)) = \int_A |\det D\phi(x)|\mathrm{d}x.^{\dagger}$$

Setting $A' = \phi(A)$, this is equivalent to

$$\int_{\phi(\Omega)} \mathbb{1}_{A'}(y)\mathrm{d}y = \int_\Omega (\mathbb{1}_{A'} \circ \phi)(x)|\det D\phi(x)|\mathrm{d}x.$$

By the monotonicity principle (Theorem 3.1.6) this implies the transformation formula for diffeomorphisms.

Exercise 11. Let $f : \Omega \to [0, \infty]$ be measurable and $p > 0$. Show that

$$\int f^p \mathrm{d}\mu = p \int_0^\infty t^{p-1}\mu(\{x \,:\, f(x) > t\})\mathrm{d}t.$$

3.4 Lebesgue spaces

Measure theory and integration is nicely related to functional analysis. We will explore this in more detail later. Now we turn to the fact that one can gather measurable functions in vector spaces that can even be normed and moreover, turn out to be complete. These spaces go under the name **Lebesgue spaces** and are a very convenient tool in various fields such as partial differential equations or the calculus of variations. We will be brief about Lebesgue spaces as they are usually treated in more detail in a course on functional analysis.

The set of all measurable functions is indeed a vector space (cf. Lemma 3.1.5). That's nice, but it would be nicer if there were some additional structure to exploit, e.g. norms, metrics or topologies.

We already defined summable functions as functions such that $\int |f|\mathrm{d}\mu < \infty$ (we will use the same notion for complex valued or real valued functions). Similarly, we may define for $0 < p < \infty$ pth-power summable functions as functions f such that $\int |f|^p \mathrm{d}\mu < \infty$.

Definition 3.4.1 (Lebesgue space)**.** Let μ be a measure on Ω and let $0 < p < \infty$. Then the **Lebesgue space** of pth-power summable functions is

$$L^p(\Omega) = \{f \,:\, f \text{ measurable and } \int |f|^p \mathrm{d}\mu < \infty\}.$$

To emphasize the respective measure one may also write $L^p(\Omega, \mu)$ or, if both Ω and μ are clear from the context or do not play any role , even only L^p. If $\Omega = \mathbb{N}$ (or $\Omega = \mathbb{Z}$) with the counting measure, then one writes $L^p(\mathbb{N}, \#) = \ell^p$ for the space of **pth-power summable sequences**.

† A proof can be found in Brokate and Kersting's "Maß und Integral" (Satz X.1, p. 103).

From here on we always use the convention that functions agreeing almost everywhere are equivalent. Note that this convention is unnecessary in the case of ℓ^p spaces.

The L^p spaces would be uninteresting if they merely were sets, but in fact they are vector spaces: If $f, g \in L^p$, then by the rather crude estimate

$$|f+g|^p \leq (2\max(|f|,|g|))^p = 2^p \max(|f|^p, |g|^p) \leq 2^p(|f|^p + |g^p|)$$

we conclude by linearity of the integral that $f+g \in L^p$.

One may go one step further and introduce a (quasi-)norm on the L^p-spaces:

Theorem 3.4.2. *The functional*

$$\|f\|_{L^p} = \Big(\int |f|^p \mathrm{d}\mu\Big)^{1/p}$$

has the following properties:

1. $\|f\|_{L^p} = 0$ *if and only if* $f = 0$ *a.e.*
2. *for* $c \in \mathbb{C}$ *it holds that* $\|cf\|_{L^p} = |c|\|f\|_{L^p}$
3. *for any* $0 < p < \infty$ *it holds that*

$$\|f+g\|_{L^p} \leq 2^{1+\frac{1}{p}}(\|f\|_{L^p} + \|g\|_{L^p})$$

and if $p \geq 1$ *it even holds*

$$\|f+g\|_{L^p} \leq \|f\|_{L^p} + \|g\|_{L^p}.$$

Proof. The first two points are obvious and for the first statement of the third point, use the crude estimate from above twice.

To show that last statement we argue as follows: Consider f and g such that $\|f\|_{L^p} + \|g\|_{L^p} = 1$ and hence, we can write $f = \theta F$ and $g = (1-\theta)G$ with $0 \leq \theta \leq 1$ and $F, G \in L^p$ with $\|F\|_{L^p} = \|G\|_{L^p} = 1$. Consequently, by convexity of $x \mapsto |x|^p$, and the normalizations

$$\int |f+g|^p \mathrm{d}\mu = \int |\theta F + (1-\theta)G|^p \mathrm{d}\mu \leq \int \theta|F|^p + (1-\theta)|G|^p \mathrm{d}\mu \leq 1.$$

Hence, the claim holds under the normalization $\|f\|_{L^p} + \|g\|_{L^p} = 1$. In the other cases we can normalize the desired inequality by dividing f and g by $\|f\|_{L^p} + \|g\|_{L^p}$ to obtain

$$\frac{\|f+g\|_{L^p}}{\|f\|_{L^p} + \|g\|_{L^p}} \leq 1.$$

□

We can express the content of the above theorem in a more elaborate way as: the functional $\|f\|_{L^p}$ defines a norm on the L^p-space for $p \geq 1$ and a quasi-norm in the case $0 < p < 1$.

The triangle inequality for the L^p-norm is also called **Minkowski inequality**.

The limit case $p = \infty$ can also be treated: We call a measurable function f **essentially bounded**, if there is an M such that $|f(x)| \leq M$ a.e. and the space $L^\infty(\Omega)$ is the set of all essentially bounded functions. The least bound M with this property is called **essential supremum** of f and denoted by $\|f\|_{L^\infty}$. This is a norm on $L^\infty(\Omega)$.

Exercise 12. Let $f \in L^{p_0} \cap L^\infty$ for some $p_0 < \infty$. Show that $\|f\|_p \to \|f\|_{L^\infty}$ for $p \to \infty$.

In some cases the L^p spaces are ordered by their exponent, but not in all cases:

Exercise 13.

1. Show that $L^p(\Omega) \subset L^q(\Omega)$ for $p \leq q$ and $\mu(\Omega) < \infty$.

   Hint: You may use Hölder's inequality without proof.

2. Show that $\ell^p \subset \ell^q$ for $p \geq q$.
3. Show that for $\Omega = \mathbb{R}^n$ neither of $L^p(\Omega)$ and $L^q(\Omega)$ is contained in the other.

   Hint: The point here is not special to $\Omega = \mathbb{R}^n$, but it relies on the fact that here both $\mu(\Omega) = \infty$ and μ does not have **atoms**, i.e. no sets "of smallest measure".

The notion of a norm implies a notion of convergence and we write

$$f^n \to f \text{ in } L^p \quad \text{if} \quad \|f^n - f\|_{L^p} \to 0.$$

In fact, these notions of convergence are in general all different for different values of p.

Exercise 14. Let $1 \leq p < q \leq \infty$. Give an example of a sequence f_n that converges in $L^p(\mathbb{R})$ and not in $L^q(\mathbb{R})$ and vice versa.

Exercise 15. Give an example of a sequence f_n that converges to 0 in $L^p([0,1])$ (with the Lebesgue measure) but does not converge to 0 pointwise almost everywhere.

To make things even more complicated, we introduce another type of convergence of sequences of functions:

Definition 3.4.3. A sequence f_n of measurable functions is said to **converge in measure** to f, if for any $\epsilon > 0$ it holds that

$$\mu(\{x \,:\, |f_n(x) - f(x)| > \epsilon\}) \xrightarrow{n\to\infty} 0$$

and we write

$$f_n \to f \text{ in measure.}$$

Lemma 3.4.4. *If f_n converges in measure to f then there exists a subsequence of f_n that converges to f pointwise almost everywhere.*

Proof. Since $\mu(\{x : |f_n(x) - f(x)| > \epsilon\})$ converges to zero for every ϵ, we can find indices n_k such that for $m > n_k$ we have

$$\mu(\{x : |f_m(x) - f(x)| > \tfrac{1}{k}\}) \leq 2^{-k}.$$

Hence,

$$\sum_{k=1}^{\infty} \mu(\{x : |f_m(x) - f(x)| > \tfrac{1}{k}\}) \leq 1.$$

Now denote by D the set of all x such that $f_{n_k}(x)$ does not converge to $f(x)$. Since for any N it holds that

$$D \subset \bigcup_{k \geq N} \{x : |f_{n_k}(x) - f(x)| > \tfrac{1}{k}\}$$

we have that

$$\mu(D) \leq \mu(\bigcup_{k \geq N} \{x : |f_{n_k}(x) - f(x)| > \tfrac{1}{k}\}) \leq \sum_{k > N} \mu(\{x : |f_{n_k}(x) - f(x)| > \tfrac{1}{k}\}).$$

Since the series on the right hand side converges, the right hand side goes to zero for $N \to \infty$, showing that D is a Null set, as desired. □

Lemma 3.4.5. *If f_n converges to f in L^p for some $0 < p < \infty$, then it also converges to f in measure. As a consequence, f_n has a subsequence that converges to f pointwise a.e.*

Proof. This follows from the basic but important inequality $\epsilon \mathbb{1}_{\{x : f(x) > \epsilon\}} \leq f$, leading to the **Markov inequality**

$$\mu(\{x : |f(x)| > \epsilon\}) \leq \tfrac{1}{\epsilon} \int |f| \mathrm{d}\mu$$

by observing

$$\begin{aligned} \mu(\{x : |f_n(x) - f(x)| > \epsilon\}) &= \mu(\{x : |f_n(x) - f(x)|^p > \epsilon^p\}) \\ &\leq \tfrac{1}{\epsilon^p} \int |f_n - f|^p \mathrm{d}\mu \end{aligned}$$

□

What makes the L^p spaces particularly useful is that they are indeed **complete**, i.e. closed with respect to the convergence induced by their norms. Since complete normed spaces are called Banach spaces, we state:

Theorem 3.4.6. *The space L^p is a Banach space for $1 \leq p \leq \infty$.*

Proof. The only thing that remains to prove is the completeness, i.e. to show that Cauchy sequences converge. Let f_n be a Cauchy sequence in L^p, that is for every $\epsilon > 0$ there is N such that for $n, m > N$ it holds that $\|f_n - f_m\|_{L^p} \leq \epsilon$. Similar to the proof of Lemma 3.4.4 we show that a subsequence of f_n has a pointwise a.e. limit: For every $\delta > 0$ it holds that $\mu(\{x : |f_n(x) - f_m(x)| > \delta\}) \leq \frac{\|f_n - f_m\|_{L^p}^p}{\delta^p}$, hence f_n is a Cauchy sequence in measure, i.e. for every $\epsilon > 0$ and $\delta > 0$ there is an N such that for $n, m > N$ it holds that

$$\mu(\{x : |f_n(x) - f_m(x)| > \delta\}) \leq \epsilon.$$

Hence, we can find a subsequence f_{n_k} such that

$$\mu(\{x : |f_{n_k}(x) - f_{n_{k+1}}(x)| > 2^{-k}\}) \leq 2^{-k}.$$

We define by M the set of x such that for infinitely many k we have $|f_{n_k}(x) - f_{n_{k+1}}(x)| > 2^{-k}$ and conclude, analogously to the proof of Lemma 3.4.4, that M is a Null set. Consequently, for every $x \notin M$ there is an N such that for $k > N$ we have $|f_{n_k}(x) - f_{n_{k+1}}(x)| \leq 2^{-k}$, i.e. $f_n(x)$ is a Cauchy sequence for every $x \notin M$ and we set $f(x)$ as the limit (which exists due to completeness of $\mathbb{R}$).

Using Fatou's Lemma we get

$$\begin{aligned}
\|f - f_n\|_{L^p}^p &= \int |\lim_{k\to\infty} f_{n_k} - f_n|^p \mathrm{d}\mu \\
&= \int \lim_{k\to\infty} |f_{n_k} - f_n|^p \mathrm{d}\mu \\
&\leq \liminf_{k\to\infty} \int |f_{n_k} - f_n|^p \mathrm{d}\mu.
\end{aligned}$$

Since we started with a Cauchy sequence f_n, the right hand side becomes small for $n \to \infty$ and this shows that $f_n \to f$ in L^p as desired. □

Exercise 16. Consider the set

$$R^p([0,1]) = \{f : [0,1] \to \mathbb{R} : |f|^p \text{ is Riemann integrable}\}.$$

Could this be turned into a Banach space with norm $\|f\|_{R^p} = (\int_0^1 |f(x)| \mathrm{d}x)^{1/p}$? Comment on the obstructions and ways to circumvent them (if possible).

When working with functions in Lebesgue spaces it is often useful, to know that these functions can be approximated by continuous functions:

Theorem 3.4.7. *Let $f \in L^1(\mathbb{R}^n)$ and $\epsilon > 0$. Then there exists a continuous function with compact support $g : \mathbb{R}^n \to \mathbb{R}$ such that $\|f - g\|_{L^1} \leq \epsilon$.*

We do not give a formal **proof** but argue in a handwaving way: First note that we can approximate f by an L^1 function with compact support (simply restrict f to a large enough compact set). Then note that we know that we can approximate f by simple and measurable functions (i.e. one which is a finite linear combination of indicator function). Now note that we can approximate the indicator function of one measurable set as a finite sum of indicator functions of indicator functions of boxes, i.e. by functions $\mathbb{1}_{[a_1,b_1]\times\cdots\times[a_n,b_n]}$. For these functions one can construct an approximating continuous function by hand (e.g. in a piecewise linear way). The claim follows by putting pieces together and keeping track of all ϵs.

Theorem 3.4.8 (Lusin's Theorem). *Let $f \in L^1(\mathbb{R}^n)$ and $\epsilon > 0$. Then there exists a measurable set A such that $\lambda^n(A) < \epsilon$ and f restricted to $\mathbb{R}^n \setminus A$ is continuous.*

Proof. Take f_n continuous with compact support such that $\|f - f_n\|_{L^1} \leq \epsilon/4^n$. Since for all $\delta > 0$ it holds by the Markov inequality

$$\lambda^n(\{x \;:\; |f(x) - f_n(x)| \geq \delta\} \leq \tfrac{1}{\delta}\int |f - f_n|\mathrm{d}\lambda^n,$$

we conclude that there exists measurable sets A_n with $\lambda^n(A_n) \leq \epsilon/2^{n+1}$ and $|f(x) - f_n(x)| \leq 1/2^{n-1}$ outside of A_n. Now set $A = \bigcup A_n$ and observe that $\lambda^n(A) \leq \epsilon/2$. Moreover, $f_n \to f$ uniformly outside of A, thus f is continuous outside of A.

□

3.5 Product measures and interchanging integrals

To integrate functions that are defined on Cartesian products, we first define the product of measures:

Definition 3.5.1 (Product of measures). Let $\mu_{1/2}$ be a measures on $\Omega_{1/2}$, respectively. Then the **Cartesian product** of μ_1 and μ_2 is denoted by $\mu_1 \times \mu_2$ and defined by

$$\begin{aligned}(\mu_1 \times \mu_2)(E) = \inf\{\sum_{n\in\mathbb{N}} \mu_1(A_n)\mu_2(B_n) \;:\; E \subset \bigcup_{n\in N} A_n \times B_n,\\ A_n\ \mu_1\text{-measurable},\ B_n\ \mu_2\text{-measurable}\}\end{aligned}$$

One can show that $\mu_1 \times \mu_2$ is indeed an outer measure. The question how the respective σ-algebra looks like is not treated here. In fact, there are other definitions of the product measure, e.g. one could only demand that a product measure should fulfill $(\mu_1 \times \mu_2)(A \times B) = \mu_1(A)\mu_2(B)$, but this does not give a uniquely defined product measure in general. In fact the uniqueness fails if one of the measure spaces in "too large" in that it is not **σ-finite**, i.e. there is no sequence of sets A_n such that $\mu(A_n) < \infty$ and $\bigcup_{n\in\mathbb{N}} A_n = \Omega$. We don't go into details of these issues here and only state the important theorem about interchanging integrals:

Theorem 3.5.2 (Fubini). *Let $\mu_{1/2}$ be two σ-finite measures on the sets $\Omega_{1/2}$, respectively. If f is a $(\mu_1 \times \mu_2)$-summable function, then it holds that*

1. *$x \mapsto f(x,y)$ is μ_1 summable for μ_2-almost all y,*
2. *$y \mapsto f(x,y)$ is μ_2 summable for μ_1-almost all y,*
3. *$g(x) = \int f(x,y) \mathrm{d}\mu_2(y)$ is μ_1-summable,*
4. *$h(y) = \int f(x,y) \mathrm{d}\mu_1(x)$ is μ_2-summable, and*
5. *it holds that*

$$\begin{aligned}\int f \mathrm{d}(\mu_1 \times \mu_2) &= \int \Big(\int f(x,y) \mathrm{d}\mu_1(x) \Big) \mathrm{d}\mu_2(y) \\ &= \int \Big(\int f(x,y) \mathrm{d}\mu_2(y) \Big) \mathrm{d}\mu_1(x).\end{aligned} \tag{*}$$

Another theorem with similar conclusion but different assumption is:

Theorem 3.5.3 (Tonelli). *If f is a non-negative $(\mu_1 \times \mu_2)$-measurable function, then $g(x) = \int f(x,y) \mathrm{d}\mu_2(y)$ is μ_1-measurable, $h(y) = \int f(x,y) \mathrm{d}\mu_1(x)$ is μ_2-measurable and equation (*) holds.*

Both **proofs** can be found in [Rud87, Chapter 8].

Remark 3.5.4. Both Fubini and Tonelli allow to interchange two series, i.e. to use $\sum_{m\in\mathbb{N}} \sum_{n\in\mathbb{N}} a_{nm} = \sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} a_{nm}$, the former in the case when $\sum_{n,m} |a_{n,m}| < \infty$, the latter in the case where $a_{nm} \geq 0$.

Example 3.5.5. Without the σ-finiteness the statement may be false: Consider $\Omega_1 = \Omega_2 = [0,1]$, $\mu_1 = \lambda$ the Lebesgue measure and $\mu_2 = \#$ the counting measure. Now consider $f : [0,1]^2 \to \mathbb{R}$ defined by

$$f(x,y) = \begin{cases} 1 & x = y \\ 0 & \text{else.} \end{cases}$$

Then

$$\int \Big(\int f(x,y) \mathrm{d}\lambda(x) \Big) \mathrm{d}\#(y) = \int 0 \mathrm{d}\# = 0$$

and

$$\int \Big(\int f(x,y) \mathrm{d}\#(y) \Big) \mathrm{d}\lambda(x) = \int 1 \mathrm{d}\lambda = 1$$

Example 3.5.6. As an application of Fubini's theorem, we rederive the layer-cake formula from Theorem 3.3.5: For a measurable $f \geq 0$ we define the following subset of $\Omega \times [0,\infty[$

$$A = \{(x,t) \,:\, t < f(x)\}$$

(also called the **subgraph**). We calculate the integral of $\mathbb{1}_A$ with respect to $\mu \times \lambda^1$ in two different ways, using Fubini: First

$$\begin{aligned}\int \mathbb{1}_A \mathrm{d}(\mu \times \lambda^1) &= \int\int \mathbb{1}_A(x,t)\mathrm{d}\mu(x)\mathrm{d}\lambda^1(t) = \int \mu(\{x \,:\, f(x) > t\})\mathrm{d}\lambda^1(t)\\ &= \int_0^\infty \mu(\{x \,:\, f(x) > t\})\mathrm{d}t\end{aligned}$$

and then the other way round

$$\int \mathbb{1}_A \mathrm{d}(\mu \times \lambda^1) = \int\int \mathbb{1}_A(x,t)\mathrm{d}\lambda^1(t)\mathrm{d}\mu(x) = \int f(x)\mathrm{d}\mu(x).$$

Exercise 17. Show that Fubini's Theorem may not be true without the summability condition by the following example:Let $\Omega_1 = \Omega_2 = \mathbb{N}$ and $\mu_1 = \mu_2 = \#$ and consider

$$f_{n,m} = \begin{cases} 1 & m = n \\ -1 & m = n+1 \\ 0 & \text{else} \end{cases}.$$

4 Vector measures, decomposition, covering, representation

We now start to get more geometrical. First we treat measures with values that are not non-negative anymore but may be vector valued. Then, using delicate and technical covering techniques, we are going to introduce the differentiation of measures which leads to a simple proof of two important structure theorems for measures: the Radon-Nikodym theorem and the Lebesgue decomposition theorem.

4.1 Vector measures, decomposition

Up to now, measures have been non-negative and the value ∞ was allowed. Here we now allow measures that have their values in $\mathbb{R}^m$ (we could also treat $\mathbb{C}$). However, we can not allow the value ∞ anymore, as this could lead to undefined situations of the form $\infty - \infty$.

Definition 4.1.1. Let $(\Omega, \mathcal{A})$ be a measurable space.

1. We say that $\mu : \mathcal{A} \to \mathbb{R}^m$ is a **vector measure** or a $\mathbb{R}^m$**-valued measure** if $\mu(\emptyset) = 0$ and for countably many disjoint sets A_n it holds that

$$\mu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mu(A_n).$$

   In the case $m = 1$ we also say **real measure** or **signed measure**.

2. The **variation measure** $|\mu|$ of a vector measure μ is defined by

$$|\mu|(A) = \sup\{\sum_{n=0}^{\infty} |\mu(A_n)| \ : \ A_n \in \mathcal{A} \ \text{disjoint} \ A = \bigcup_{n\in\mathbb{N}} A_n\}.$$

3. For a real measure we define its **positive part** and **negative part**, respectively, as

$$\mu^+ = \frac{|\mu| + \mu}{2}, \qquad \mu^- = \frac{|\mu| - \mu}{2}.$$

We are going to show that the variation measure (and consequently also the positive and negative part) are indeed measures. Before we do so, we have a look at some remarks on vector measures:

Remark 4.1.2. 1. Note that the absolute convergence of the series in the first point is implicitly required as the value of the series can not depend on the order of summation, since the union does not depend on it.

2. Vector measures form a vector space: We simply define addition and scalar multiplication "pointwise", i.e. $(\mu + \nu)(A) = \mu(A) + \nu(A)$ and $(\lambda\mu)(A) = \lambda\mu(A)$.

3. Some books deal with measures with values in Banach spaces, but we do not do this here.

4. Integration with respect signed measure is defined as
$$\int f \mathrm{d}\mu = \int f \mathrm{d}\mu^+ - \int f \mathrm{d}\mu^-.$$
and with respect to $|\mu|$ analogously as
$$\int f \mathrm{d}|\mu| = \int f \mathrm{d}\mu^+ + \int f \mathrm{d}\mu^-.$$

5. We integrate with respect to vector measure $\mu = (\mu_1, \ldots, \mu_m)$ either componentwise, i.e. for measurable $f : \Omega \to \mathbb{R}$
$$\int f \mathrm{d}\mu = (\int f \mathrm{d}\mu_1, \ldots, \int f \mathrm{d}\mu_m)$$
or for $f : \Omega \to \mathbb{R}^m$ by
$$\int f \cdot \mathrm{d}\mu = \sum_{i=1}^{m} \int f_i \mathrm{d}\mu_i.$$

Example 4.1.3. 1. The counting measure # is a real measure only on finite sets.

2. We can build a simple vector measure from the Dirac measures δ_{x_n} for $x_n \in \Omega$ with the help of a sequence $c_n \in \mathbb{R}^m$ such that $\sum_n |c_n| < \infty$ by setting
$$\mu(A) = (\sum_n c_n \delta_{x_n})(A) = \sum_{\{n \,:\, x_n \in A\}} c_n.$$

3. In the case of the Lebesgue measure λ on $\Omega \subset \mathbb{R}^n$ (for instance) and a λ-summable function $f : \Omega \to \mathbb{R}^m$ we define a vector measure as
$$(f\lambda)(A) = \int_A f \mathrm{d}\lambda = (\int_A f_1 \mathrm{d}\lambda, \ldots, \int_A f_m \mathrm{d}\lambda).$$

Exercise 18. Show that $|f\lambda| = |f|\lambda$, $(f\lambda)^+ = f^+\lambda$ and $(f\lambda)^- = f^-\lambda$.

The variation measure is indeed a measure:

Theorem 4.1.4. *The variation $|\mu|$ of a vector measure μ is a positive and finite measure.*

Proof. First we show that $|\mu|$ is countable subadditive: Let $A_n \in \mathcal{A}$ such that $A \subset \cup A_n$. Set $A'_0 = A_0$ and $A'_n = A_n \setminus \bigcup_{j=0}^{n-1} A_j$ for $n \geq 1$. Moreover, let B_j be a countable partition of A and, noting that $A'_n \cap B_j$ is a partition of B_j, we get from countable additivity of μ that

$$\begin{aligned}\sum_{j\in\mathbb{N}} |\mu(B_j)| = \sum_{j\in\mathbb{N}} |\sum_{n\in\mathbb{N}} \mu(A'_n \cap B_j)| &\leq \sum_{j\in\mathbb{N}} \sum_{n\in\mathbb{N}} |\mu(A'_n \cap B_j)| \leq \sum_{n\in\mathbb{N}} |\mu|(A'_n)\\ &\leq \sum_{n\in\mathbb{N}} |\mu|(A_n).\end{aligned}$$

Taking the supremum over all partitions B_j we obtain $|\mu|(A) \leq \sum_{n\in\mathbb{N}} |\mu|(A_n)$.

Now we prove finite additivity: Let A^1 and A^2 be disjoint and let $\epsilon > 0$. Then there exists partitions A^i_n of A^i (for $i = 1, 2$, respectively) such that for $i = 1, 2$:

$$|\mu|(A^i) \leq \sum_{n\in\mathbb{N}} |\mu(A^i_n)| + \epsilon.$$

Then

$$|\mu|(A^1 \cup A^2) \geq \sum_{i=1,2} \sum_{n\in\mathbb{N}} |\mu(A^i_n)| \geq |\mu|(A^1) + |\mu|(A^2) - 2\epsilon$$

and we conclude $|\mu|(A^1 \cup A^2) \geq |\mu|(A^1) + |\mu|(A^2)$ (and the reverse inequality is true by subadditivity). Note that the above reasoning also shows countable additivity (by considering A_i $i \in \mathbb{N}$ disjoint and approximating up to $\epsilon/2^i$). However, countable additivity follows from countable subadditivity and finite additivity in general: To see this consider disjoint A_n and countable subadditive and additive ν and observe

$$\nu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) \leq \sum_{n\in\mathbb{N}} \nu(A_n) = \lim_{N\to\infty} \sum_{n=0}^{N} \nu(A_n) = \lim_{N\to\infty} \nu\Big(\bigcup_{n=0}^{N} A_n\Big) \leq \nu\Big(\bigcup_{n\in\mathbb{N}} A_n\Big).$$

To prove finiteness, observe that we only need to consider the case of a real measure (since the vector case follows from the fact that $\mu = (\mu_1, \ldots, \mu_m)$ fulfills $|\mu|(A) \leq \sum_{n=1}^{m} |\mu_n|(A)$).

Assume by contradiction that $|\mu|(\Omega) = \infty$. Then there exists a countable partition of Ω into A_n and $N \in \mathbb{N}$ such that $\sum_{n=0}^{N} |\mu(A_n)| > 2(|\mu(\Omega)| + 1)$. We can find a set A (as a union of some A_n's) such that $|\mu(A)| > |\mu(\Omega)| + 1$. Setting $B = \Omega \setminus A$, we compute $|\mu(B)| = |\mu(\Omega) - \mu(A)| \geq |\mu(A)| - |\mu(\Omega)| > 1$. By additivity of μ we have $|\mu|(A) = \infty$ or $|\mu|(B) = \infty$. In the latter case, set $A_1 = A$ and repeat the above argument in B to obtain a splitting of B into disjoint A_2 and B_1 with $|\mu(A_2)| > 1$ and $|\mu|(B_1) = \infty$ (otherwise $A_1 = B$). Iterating leads to a sequence of disjoint sets A_n such that $|\mu(A_n)| > 1$ and hence, the series $\sum \mu(A_n)$ can not be convergent. By contradiction, this proves that $|\mu|$ is a finite measure. $\square$

We already have seen that vector valued measures form a vector space, but with the help of the variation we can even turn the space into a normed space:

Theorem 4.1.5. *The space $\mathfrak{M}(\Omega, \mathbb{R}^m)$ of all $\mathbb{R}^m$-valued measures is a normed space when endowed with the norm*

$$\|\mu\|_{\mathfrak{M}} = |\mu|(\Omega).$$

Proof. If $\|\mu\|_{\mathfrak{M}} = 0$ then $|\mu|(\Omega) = 0$, hence, all sets are null sets for $|\mu|$ and hence, $|\mu|$ is the "null measure". Consequently, μ^+ is null, and (by considering $|-\mu|$) also μ^- is null.

Obviously for $c \in \mathbb{C}$:

$$\|c\mu\|_{\mathfrak{M}} = |c\mu|\Omega = |c||\mu|(\Omega) = |c|\|\mu\|_{\mathfrak{M}}.$$

For the triangle inequality, note the subadditivity of the supremum to conclude

$$\begin{aligned}
\|\mu+\nu\|_{\mathfrak{M}} &= \sup\{\textstyle\sum |\mu(A_n)+\nu(A_n)| \ : \ A_n \in \mathcal{A} \ \text{disjoint} \ \Omega = \bigcup_{n\in\mathbb{N}} A_n\} \\
&\leq \sup\{\textstyle\sum |\mu(A_n)| + \sum |\nu(A_n)| \ : \ A_n \in \mathcal{A} \ \text{disjoint} \ \Omega = \bigcup_{n\in\mathbb{N}} A_n\} \\
&\leq \sup\{\textstyle\sum |\mu(A_n)| \ : \ A_n \in \mathcal{A} \ \text{disjoint} \ \Omega = \bigcup_{n\in\mathbb{N}} A_n\} \\
&\qquad + \sup\{\textstyle\sum |\nu(A_n)| \ : \ A_n \in \mathcal{A} \ \text{disjoint} \ \Omega = \bigcup_{n\in\mathbb{N}} A_n\} \\
&= \|\mu\|_{\mathfrak{M}} + \|\nu\|_{\mathfrak{M}}.
\end{aligned}$$

□

Later we will see that the space $\mathfrak{M}(\Omega, \mathbb{R}^m)$ is even a Banach space.

We already encountered the decomposition of a real measure μ into its positive and negative part μ^+ and μ^-, respectively; this decompostion of a signed measure as the sum of two (positive) measures is called **Jordan decompostion**. This decomposition also gives a decomposition of the underlying set Ω:

Definition 4.1.6 (Hahn decomposition)**.** Let μ be a signed measure on a set Ω. A **Hahn decompostion** of Ω (w.r.t. μ) is a disjoint decomposition $\Omega = \Omega^+ \cup \Omega^-$ such that for any A it holds that $\mu\llcorner\Omega^+(A) = \mu(\Omega^+ \cap A) \geq 0$ and $\mu\llcorner\Omega^-(A) = \mu(\Omega^- \cap A) \leq 0$.

Remark 4.1.7. A Hahn decomposition always exists but is in general not unique: Consider the sets $\operatorname{supp}\mu^+, \operatorname{supp}\mu^- \subset \Omega$. Obviously $\mu\llcorner\operatorname{supp}\mu^+ \geq 0$ and $\mu\llcorner\operatorname{supp}\mu^- \leq 0$. However, $\operatorname{supp}\mu^+ \cup \operatorname{supp}\mu^-$ may not be the whole Ω. The remaining set (for which all subsets are Null sets) may be added to $\operatorname{supp}\mu^+$ and $\operatorname{supp}\mu^-$ in arbitrary parts to form a valid Hahn decomposition.

4.2 Covering theorems

Now we introduce a technical tool in measure theory—the so-called covering theorems. Actually, covering theorems are not necessarily measure theoretic concepts but are more related to the geometry of underlying space. Here we are going to work in $\mathbb{R}^n$ only, i.e all sets will be subsets of $\mathbb{R}^n$ with its usual topology induced by the euclidean norm.

We are going to work with closed balls B, i.e. sets of the form $\{y : |y-x| \leq r\}$. For any set $A \subset \mathbb{R}^n$ we denote by $\operatorname{diam} A$ its diameter, i.e. $\operatorname{diam} A = \sup\{|x-y| : x, y \in A\}$. Note that for $c > 0$ and closed ball B we will write, with a slight abuse of notation, cB for the closed ball with radius c times the radius of B but the same center, i.e. for $B = \{y : |y-x| \leq r\}$ we mean $cB = \{y : |y-x| \leq cr\}$.

Definition 4.2.1. 1. A collection $\mathcal{F}$ of closed balls in $\mathbb{R}^n$ is a **cover** of a set $A \subset \mathbb{R}^n$ if

$$A \subset \bigcup_{B \in \mathcal{F}} B$$

and

2. $\mathcal{F}$ is a **fine cover** of A if, in addition, for each $x \in A$ it holds that

$$\inf\{\operatorname{diam} B : x \in B,\ B \in \mathcal{F}\} = 0.$$

Theorem 4.2.2 (Vitali's covering theorem)**.** *Let $\mathcal{F}$ be a collection of closed balls in $\mathbb{R}^n$ with positive radius and*

$$D = \sup\{\operatorname{diam} B : B \in \mathcal{F}\} < \infty.$$

Then there exists a countable subcollection $\mathcal{G}$ of $\mathcal{F}$ of disjoint closed balls such that

$$\bigcup_{B \in \mathcal{F}} B \subset \bigcup_{B \in \mathcal{G}} 5B.$$

Proof. We set $\mathcal{F}_j = \{B \in \mathcal{F} : D/2^{j+1} < \operatorname{diam} B \leq D/2^j\}$ and define $\mathcal{G}_j$ as a subcollection of $\mathcal{F}_j$ as follows:

1. Let $H_0 = \mathcal{F}_0$ and $\mathcal{G}_0$ by any maximal disjoint collection of balls in $\mathcal{H}_0$,

2. Recursively,

$$\mathcal{H}_{k+1} = \{B \in \mathcal{F}_{k+1} : B \cap B' = \emptyset \text{ for all } B' \in \bigcup_{j=0}^{k} \mathcal{G}_j\}.$$

and let $\mathcal{G}_{k+1}$ be any maximal disjoint subcollection of $\mathcal{H}_{k+1}$.

Then we define $\mathcal{G} = \bigcup_{j=0}^{\infty} \mathcal{G}_j$. By construction, $\mathcal{G}$ is a subcollection of $\mathcal{F}$ of disjoint balls.

Now we aim to show that for each $B \in \mathcal{F}$ there exists a ball $B' \in \mathcal{G}$ such that $B \cap B' \neq \emptyset$ and $B \subset 5B'$. To that end, choose some $B \in \mathcal{F}$. Then there exists some k such that $B \in \mathcal{F}_k$. Either B does not belong to $\mathcal{H}_k$, which implies $k > 0$ and that B intersects some ball $C \in \bigcup_{j=0}^{k-1} \mathcal{G}_j$, or B belongs to $\mathcal{H}_k$ and then, by maximality of $\mathcal{G}_k$, there is some ball in $C \in \mathcal{G}_k$. In any case, there is a $C \in \bigcup_{j=0}^{k} \mathcal{G}_k$ such that $B \cap C \neq \emptyset$. Such a ball has radius larger that $D/2^{k+1}$ but B has radius smaller than $D/2^k$. I.e. the radius of B is less than twice the radius of C. Since B and C intersect, it holds that $B \subset 5C$. $\square$

We give two (even more) technical corollaries that will be used in the following:

Corollary 4.2.3. *Let $\mathcal{F}$ be a fine cover of A with closed balls with positive radius and*

$$\sup\{\mathrm{diam} B \ : \ B \in \mathcal{F}\} < \infty.$$

Then there exists a countable disjoint subcollection $\mathcal{G}$ of $\mathcal{F}$ such that for any finite subset $\{B_1, \dots, B_m\} \subset \mathcal{F}$ it holds that

$$A \setminus \bigcup_{k=1}^{m} B_k \subset \bigcup_{B \in \mathcal{G} \setminus \{B_1,\dots,B_m\}} 5B.$$

Proof. Build $\mathcal{G}$ as in the proof of the Vitali covering theorem and select $\{B_1, \dots, B_m\} \subset \mathcal{F}$. If $A \subset \bigcup_{k=1}^{m} B_k$ we are done. Otherwise, let $x \in A \setminus \bigcup_{k=1}^{m} B_k$. Since the balls are closed and $\mathcal{F}$ is a fine cover, there exists $B \in \mathcal{F}$ with $x \in B$ and $B \cap B_k = \emptyset$ for $k = 1, \dots, m$. But then, similarly to the proof above, there exists a ball $B' \in \mathcal{G} \setminus \{B_1, \dots, B_m\}$ such that $B \cap B' \neq \emptyset$ and $B \subset 5B'$. $\square$

The second corollary of the Vitali covering theorem shows that we can "fill up" (in a measure theoretic sense) any open set with countably many disjoint closed balls. One should note that the only assumption on the set U there is, that is it open and no further regularity is demanded.

Corollary 4.2.4. *Let $U \subset \mathbb{R}^n$ be open and $\delta > 0$. Then there exists a countable collection $\mathcal{G}$ of disjoint closed balls in U such that* $\mathrm{diam} B \leq \delta$ *for all $B \in \mathcal{G}$ and*

$$\lambda^n(U \setminus \bigcup_{B \in \mathcal{G}} B) = 0.$$

Proof. Choose $\theta \in]1 - \frac{1}{5^n}, 1[$ and without loss of generality, we assume that U has finite measure (if it has not, we consider the not more that countably many sets $U_m = \{x \in U \ : \ m < |x| < m+1\}$ separately).

Now we claim that there exists a finite collection $\{B_1,\dots,B_{M_1}\}$ of disjoint closed balls in U such that $\operatorname{diam} B_i < \delta$ and

$$\lambda^n(U\setminus\bigcup_{i=1}^{M_1}B_i)\leq\theta\lambda^n(U).$$

To see this, let $\mathcal{F}_1=\{B\ :\ B \text{ closed ball }, B\subset U, \operatorname{diam} B<\delta\}$. By the Vitali covering theorem, there exists a countable, disjoint subcollection $\mathcal{G}_1\subset\mathcal{F}_1$ such that $U\subset\bigcup_{B\in\mathcal{G}_1}5B$. It follows that

$$\begin{aligned}\lambda^n(U)&\leq\sum_{B\in\mathcal{G}_1}\lambda^n(5B)=5^n\sum_{B\in\mathcal{G}_1}\lambda^n(B)\\&=5^n\lambda^n(\bigcup_{B\in\mathcal{G}_1}B).\end{aligned}$$

(Note that we used that λ^n is the Lebesgue measure in this step.) Hence,

$$\lambda^n(\bigcup_{B\in\mathcal{G}_1}B)\geq\tfrac{1}{5^n}\lambda^n(U)$$

and consequently

$$\lambda^n(U\setminus\bigcup_{B\in\mathcal{G}_1}B)\leq\Big(1-\tfrac{1}{5^n}\Big)\lambda^n(U).$$

Since $\mathcal{G}_1$ is countable, there exists $B_1,\dots,B_{M_1}$ in $\mathcal{G}_1$ with the desired property, i.e. the claim is proven.

Now we set

$$\begin{aligned}U_2&=U\setminus\bigcup_{i=1}^{M_1}B_i\\\mathcal{F}_2&=\{B\ :\ B \text{ closed ball }, B\subset U_2, \operatorname{diam} B<\delta\},\end{aligned}$$

and, as above, find finitely many $B_{M_1+1},\dots,B_{M_2}$ in $\mathcal{F}_2$ such that

$$\begin{aligned}\lambda^n(U\setminus\bigcup_{i=1}^{M_2}B_i)&=\lambda^n(U_2\setminus\bigcup_{i=M_1+1}^{M_2}B_i)\leq\theta\lambda^n(U_2)\\&\leq\theta^2\lambda^n(U).\end{aligned}$$

We continue this process to obtain a countable collection of disjoint balls such that

$$\lambda^n(U\setminus\bigcup_{i=1}^{M_k}B_i)\leq\theta^k\lambda^n(U)$$

and since $\theta^k\to 0$ for $k\to\infty$ the corollary is proven. $\square$

Now we state Besicovitch covering theorem and we derive a corollary similar to Corollary 4.2.4. Since Besicovitch covering theorem does not work with enlarged balls, it will be applicable to general Radon measures μ on $\mathbb{R}^n$ and not only to λ^n (in this case one can not, in general, control $\mu(cB)$ in terms of $\mu(B)$ for $c > 0$).

Theorem 4.2.5 (Besicovitch covering theorem). *Let $A \subset \mathbb{R}^n$ and $\mathcal{F}$ be a collection of closed balls with positive radius such that each $x \in A$ is a center of some ball in $\mathcal{F}$ and*

$$D = \sup\{\mathrm{diam} B \,:\, B \in \mathcal{F}\} < \infty.$$

Then there exists a constant N_n, depending only on n (not on A or $\mathcal{F}$), such that there exist $\mathcal{G}_1, \dots, \mathcal{G}_{N_n} \subset \mathcal{F}$ such that each $\mathcal{G}_i$ is a countable collection of disjoint balls in $\mathcal{F}$ and

$$A \subset \bigcup_{i=1}^{N_n} \bigcup_{B \in \mathcal{G}_i} B.$$

Proof. The proof is lengthy and we present it in ten steps:

1. We start with the special case of bounded A. We choose a ball $B_1 = B_{r_1}(a_1) \in \mathcal{F}$ such that $r_1 \geq \frac{3}{4}\frac{D}{2}$ and inductively choose B_j as follows: Set $A_j = A \setminus \bigcup_{i=1}^{j-1} B_i$ and if $A_j = \emptyset$ stop and set $J = j-1$, otherwise choose $B_j = B_{r_j}(a_j) \in \mathcal{F}$ such that $a_j \in A_j$ and

$$r_j \geq \tfrac{3}{4} \sup\{r \,:\, B_r(a) \in \mathcal{F},\ a \in A_j\}. \tag{*}$$

   If $A_j \neq \emptyset$ for all j, set $J = \infty$.

2. Our construction ensures that for $j > i$ it holds that $\frac{3}{4} r_j \leq r_i$. To see this note that for $j > i$ we have $a_j \in A_i$ and so by (*) we get

$$r_i \geq \tfrac{3}{4} \sup\{r \,:\, B_r(a) \in \mathcal{F},\ a \in A_i\} \geq \tfrac{3}{4} r_j$$

   (since $A_j \subset A_i$ for $j > i$).

3. Also, our construction ensures that the balls $\{B_{r_j/3}(a_j)\}_{j=1,\dots,J}$ are disjoint: To see this note that for $j > i$ we have $a_j \notin B_i$ and hence

$$|a_i - a_j| > r_i = \tfrac{r_i}{3} + \tfrac{2r_i}{3} \geq \tfrac{r_i}{3} + \tfrac{2}{3}\tfrac{3}{4} r_j > \tfrac{r_i}{3} + \tfrac{r_j}{3}.$$

4. Since the balls $\{B_{r_j/3}(a_j)\}_{j=1,\dots,J}$ are disjoint and all $a_j \in A$ with A bounded we have that $r_j \to 0$ if $J = \infty$.

5. We now have that $A \subset \bigcup_{j=1}^{J} B_j$: If $J < \infty$ this is trivially true, hence, suppose $J = \infty$. For $a \in A$ there exists $r > 0$ such that $B_r(a) \in \mathcal{F}$. Since the radii r_j tend to zero, we have some j such that $r_j < \frac{3}{4} r$ which is a contradiction to the choice of r_j if $a \notin \bigcup_{i=1}^{j-1} B_i$.

6. Now fix some $k > 1$ and consider $I = \{j \,:\, 1 \leq j < k,\ B_j \cap B_k \neq \emptyset\}$ and set $K = I \cap \{j \,:\, r_j \leq 3r_k\}$. Our next goal is to estimate the cardinality of I. We start with the claim that $\#(K) \leq 20^n$: Let $j \in K$. Then $B_j \cap B_k \neq \emptyset$ and $r_j \leq 3r_k$. Choose $x \in B_{r_j/3}(a_j)$. Then

$$\begin{aligned} |x - a_k| &\leq |x - a_j| + |a_j - a_k| \leq \tfrac{r_j}{3} + r_j + r_k \\ &= \tfrac{4}{3} r_j + r_k \leq 5r_k \end{aligned}$$

and consequently, $B_{r_j/3}(a_j) \subset B_{5r_k}(a_k)$ for all $j \in K$. Let us denote by $V_n = \lambda^n(B_1(0)$ the volume of the unit ball. Since the balls $B_{r_i/3}(a_i)$ are disjoint (see step 3), we have

$$\begin{aligned} V_n 5^n r_k^n &= \lambda^n(B_{5r_k}(a_k)) \\ &\geq \sum_{j\in K} \lambda^n(B_{r_j/3}(a_j)) \\ &= V_n \sum_{j\in K} \left(\tfrac{r_j}{3}\right)^n \\ &\geq V_n \sum_{j\in K} \left(\tfrac{r_k}{4}\right)^n = V_n \#(K) \tfrac{r_k^n}{4^n} \end{aligned}$$

and we conclude

$$5^n \geq \#(K) \tfrac{1}{4^n}.$$

Now estimate $\#(I \setminus K)$: Let $i, j \in I \setminus K$ with $i \neq j$. Then $1 \leq i, j < k$, $B_i \cap B_k \neq \emptyset$ $B_j \cap B_k \neq \emptyset$, $r_i > 3r_k$, $r_j > 3r_k$. For simplicity we take $a_k = 0$ (which could be achieved by translation). Denote by $\theta \in [0, \pi]$ the angle between the vectors a_i and a_j and we aim to the find a lower bound on θ. But this will take some substeps:

a) We first collect some facts: Since $i, j < k$ and $0 = a_k \notin B_i \cup B_j$ we get that $r_i < |a_i|$ and $r_j < |a_j|$. Since B_k intersects both B_i and B_j, we have $|a_i| \leq r_i + r_k$ and $|a_j| \leq r_j + r_k$. Finally, we assume (without loss of generality) that $|a_i| \leq |a_j|$. We remember for further use:

$$\begin{aligned} &3r_k < r_i < |a_i| \leq r_i + r_k \\ &3r_k < r_j < |a_j| \leq r_j + r_k \\ &|a_i| \leq |a_j|. \end{aligned}$$

b) We show that for $\cos(\theta) > \frac{5}{6}$ we have $a_i \in B_j$: Suppose first that $|a_i - a_j| \geq |a_j|$ (especially $a_i \notin B_j$). Then, by the Law of Cosines:

$$\begin{aligned} \cos(\theta) &= \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} \\ &\leq \frac{|a_i|^2}{2|a_i||a_j|} = \frac{|a_i|}{2|a_j|} \leq \frac{1}{2} < \frac{5}{6}. \end{aligned}$$

Second, suppose $|a_i - a_j| \le |a_j|$ and $a_i \notin B_j$. Then $r_j < |a_i - a_j|$ and

$$\begin{aligned}
\cos(\theta) &= \frac{|a_i|^2 + |a_j|^2 - |a_i - a_j|^2}{2|a_i||a_j|} \\
&= \frac{|a_i|}{2|a_j|} + \frac{(|a_j| - |a_i - a_j|)(|a_j| + |a_i - a_j|)}{2|a_i||a_j|} \\
&\le \frac{1}{2} + \frac{(|a_j| - |a_i - a_j|)(2|a_j|)}{2|a_i||a_j|} \\
&\le \frac{1}{2} + \frac{r_j + r_k - r_j}{r_i} = \frac{1}{2} + \frac{r_k}{r_i} \le \frac{5}{6}.
\end{aligned}$$

c) Now we show that if $a_i \in B_j$, then with $\epsilon(\theta) = \frac{8}{3}(1 - \cos(\theta))$ it holds that

$$0 \le |a_i - a_j| + |a_i| - |a_j| \le |a_j|\epsilon(\theta).$$

To see this start by noting that since $a_i \in B_j$, we have $i < j$ and hence, $a_j \notin B_i$ and thus $|a_i - a_j| > r_i$. In consequence, we get

$$\begin{aligned}
0 &\le \frac{|a_i - a_j| + |a_i| - |a_j|}{|a_j|} \\
&\le \frac{|a_i - a_j| + |a_i| - |a_j|}{|a_j|} \cdot \frac{|a_i - a_j| - |a_i| + |a_j|}{|a_i - a_j|} \\
&\le \frac{|a_i - a_j|^2 - (|a_j| - |a_i|)^2}{|a_j||a_i - a_j|} \\
&= \frac{|a_i|^2 + |a_j|^2 - 2|a_i||a_j|\cos(\theta) - |a_i|^2 - |a_j|^2 + 2|a_i||a_j|}{|a_j||a_i - a_j|} \\
&= \frac{2|a_i|(1 - \cos(\theta))}{|a_i - a_j|} \\
&\le \frac{2(r_i + r_k)(1 - \cos(\theta))}{r_i} \\
&\le \frac{2(1 + \frac{1}{3})r_i(1 - \cos(\theta))}{r_i} = \epsilon(\theta).
\end{aligned}$$

d) Now we show that for $a_i \in B_j$ we have $\cos(\theta) \le \frac{61}{64}$: Since $a_i \in B_j$ we have $a_j \notin B_i$ and hence $r_i < |a_i - a_j| \le r_j$. Since $i < j$, $r_j \le \frac{4}{3} r_i$ and thus:

$$\begin{aligned}
|a_i - a_j| + |a_i| - |a_j| &\ge r_i + r_i - r_j - r_k \\
&\ge \tfrac{3}{2} r_j - r_j - r_k \\
&= \tfrac{1}{2} r_j - r_k \ge \tfrac{1}{6} r_j \\
&= \tfrac{1}{6}\tfrac{3}{4}\left(r_j + \tfrac{1}{3} r_j\right) \\
&\ge \tfrac{1}{8}(r_j + r_k) \ge \tfrac{1}{8}|a_j|.
\end{aligned}$$

Be the previous result, we get

$$\tfrac{1}{8}|a_j| \leq |a_i - a_j| + |a_i| - |a_j| \leq |a_j|\epsilon(\theta)$$

ans this can be deformed into the claim $\cos(\theta) \leq \frac{61}{64}$.

We collect the previous results and get: For all $i, j \in I \setminus K$ the angle θ between $a_i - a_k$ and $a_j - a_k$ fulfills $\theta \geq \arccos\left(\frac{61}{64}\right) = \theta_0 \approx 0.307 > 0$.

7. Now we show there exists a constant L_n (only dependent on the dimension n) such that $\#(I \setminus K) \leq L_n$. To see this start by fixing $r_0 > 0$ such that if $|x| = 1$ and $y, z \in B_{r_0}(x)$, then the angle between y and z is less than θ_0. Then choose L_n such that $\partial B_1(0)$ can be covered by L_n balls with radius r_n and centers on $\partial B_1(0)$ (but cannot be covered by $L_n - 1$ such balls).

   Then ∂B_k can be covered by L_n balls with radius $r_0 r_k$ and centers on ∂B_k. By the previous point, we have for $i, j \in I \setminus K$ with $i \neq j$ then, the angle between $a_i - a_k$ and $a_j - a_k$ is larger than θ_0. Thus, by construction of r_0, the rays $a_j - a_k$ and $a_i - a_k$ can not both go through the same ball on ∂B_k. Consequently, $\#(I \setminus K) \leq L_n$.

8. We set $M_n = 20^n + L_n + 1$ and collect the previous results to get

$$\#(I) = \#(K) + \#(I \setminus K) \leq 20^n + L_n < M_n.$$

9. Now, finally, we define the sets $\mathcal{G}_1, \ldots, \mathcal{G}_{M_n}$:

   We start with a mapping $\sigma : \{1, 2, \ldots\} \to \{1, 2, \ldots, M_n\}$ defined by

   - $\sigma(i) = i$ for $1 \leq i \leq M_n$,
   - for $k \geq M_n$ define $\sigma(k+1)$ inductively as follows: By the above we know that $\#\{j \ : \ 1 \leq j \leq k,\ B_j \cap B_{k+1} \neq \emptyset\} < M_n$ and hence, there exists $l \in \{1, \ldots, M_n\}$ such that $B_{k+1} \cap B_j = \emptyset$ such that $\sigma(j) = l$ $(1 \leq j \leq k)$. Set $\sigma(k+1) = l$

   Now let $\mathcal{G}_j = \{B_i \ : \ \sigma(i) = j\}$ for $1 \leq j \leq M_n$.

   By construction of $\sigma(i)$, each G_j consists of disjoint balls from $\mathcal{F}$. Moreover, each B_i is in some $\mathcal{G}_j$, so

$$A \subset \bigcup_{i=1}^{J} B_i = \bigcup_{i=1}^{M_n} \bigcup_{B \in G_i} B,$$

   and the proof is done for the case of bounded A (with $N_n = M_n$).

10. Now let's move on to unbounded A: For $l \geq 1$ define $A_l = \{x \in A \ : \ 3D(l-1) \leq |x| \leq 3Dl\}$ and set $\mathcal{F}^l = \{B_r(a) \in \mathcal{F} \ : \ a \in A_l\}$. Then, be

the previous step, there exist countable collections $\mathcal{G}_1^l, \ldots, \mathcal{G}_{M_n}^l$ of disjoint closed balls in $\mathcal{F}^l$ such that

$$A_l \subset \bigcup_{i=1}^{M_n} \bigcup_{B \in \mathcal{G}_i^l} B.$$

Let for $1 \leq j \leq M_n$

$$\mathcal{G}_j = \bigcup_{l=1}^{\infty} \mathcal{G}_j^{2l-1}$$
$$\mathcal{G}_{j+M_n} = \bigcup_{l=1}^{\infty} \mathcal{G}_j^{2l}.$$

By definition, the balls in each $\mathcal{G}_j$ are disjoint and we have proven the theorem with $N_n = 2M_n$.

□

This, indeed pretty lengthy proof is taken from [EG92, p.30]. There are shorter proofs of Besicovitch's covering theorem. You'll find one in [KP08, p.103] that is only one and a half pages (although it contains a figure). But alas, it builds on three lemmas and their proofs need about three pages as well...

The following variant of Corollary 4.2.4 on filling up sets with balls is a consequence of the Besicovitch covering theorem:

Corollary 4.2.6. *Let μ be a regular Borel measure on $\mathbb{R}^n$ and $\mathcal{F}$ any collection of closed ball with positive radius. Let A denote the set of centers of the balls in $\mathcal{F}$ (no measurability of A is assumed) and assume that $\mu(A) < \infty$ and also that for each $a \in A$ it holds that $\inf\{r : B_r(a) \in \mathcal{F}\} = 0$. Then for each open set $U \subset \mathbb{R}^n$ there exists a countable subcollection $\mathcal{G}$ of disjoint balls in $\mathcal{F}$ such that*

$$\bigcup_{B \in \mathcal{G}} B \subset U$$

and

$$\mu\Big((A \cap U) \setminus \bigcup_{B \in \mathcal{G}} B\Big) = 0.$$

Proof. Choose $\theta \in]1 - \frac{1}{N_n}, 1[$.

First we claim that there exists a finite subcollection $B_1, \ldots B_{M_1}$ of disjoint closed balls in U such that

$$\mu\Big((A \cap U) \setminus \bigcup_{i=1}^{M_1} B_i\Big) \leq \theta\, \mu(A \cap U).$$

To see this, let $\mathcal{F}_1 = \{B \ : \ B \in \mathcal{F},\ \operatorname{diam} B \leq 1,\ B \subset U\}$. By the Besicovitch covering theorem, there exists subcollections $\mathcal{G}_1, \dots, \mathcal{G}_{N_n}$ of disjoint balls in $\mathcal{F}_1$ such that

$$A \cap U \subset \bigcup_{i=1}^{N_n} \bigcup_{B \in \mathcal{G}_i} B.$$

Thus

$$\mu(A \cap U) \leq \sum_{i=1}^{N_n} \mu\Big(A \cap U \cap \bigcup_{B \in \mathcal{G}_i} B\Big)$$

and consequently, there exists and integer $j \in \{1, \dots, N_n\}$ such that

$$\mu\Big(A \cap U \cap \bigcup_{B \in \mathcal{G}_j} B\Big) \geq \tfrac{1}{N_n} \mu(A \cap U).$$

By the regularity of μ, there exists balls $B_1, \dots B_{M_1} \in \mathcal{G}_j$ such that

$$\mu\Big(A \cap U \cap \bigcup_{i=1}^{M_1} B_i\Big) \geq (1 - \theta)\mu(A \cap U).$$

But since $\bigcup_{i=1}^{M_1} B_i$ is μ-measurable, it holds that

$$\mu(A \cap U) = \mu\Big(A \cap U \cap \bigcup_{i=1}^{M_1} B_i\Big) + \mu\Big((A \cap U) \setminus \bigcup_{i=1}^{M_1} B_i\Big)$$

and hence, the above claim holds.

Now let $U_2 = U \setminus \bigcup_{i=1}^{M_1} B_i$, $\mathcal{F} = \{B \ : \ B \in \mathcal{F},\ \operatorname{diam} B \leq 1,\ B \subset U_2\}$ and, as above, find finitely many disjoint balls $B_{M_1+1}, \dots, B_{M_2}$ in $\mathcal{F}_2$ such that

$$\begin{aligned}
\mu\Big((A \cap U) \setminus \bigcup_{i=1}^{M_2} B_i\Big) &= \mu\Big((A \cap U_2) \setminus \bigcup_{i=M_1+1}^{M_2} B_i\Big) \\
&\leq \theta \mu(A \cap U_2) \\
&\leq \theta^2 \mu(A \cap U).
\end{aligned}$$

Repeat this process to obtain a countable collection of disjoint balls from $\mathcal{F}$ and within U such that

$$\mu\Big((A \cap U) \setminus \bigcup_{i=1}^{M_k} B_i\Big) \leq \theta^k \mu(A \cap U)$$

and since $\theta^k \to 0$ for $k \to \infty$ and $\mu(A) < \infty$, the corollary in proved. $\square$

4.3 Differentiation of measures

The next thing we aim at, is a notion of the derivative of a Radon measure with respect to another one. This may sound strange at first sight, but will turn out to be fairly natural. We start with the definition:

Definition 4.3.1. Let μ and ν be Radon measures on $\mathbb{R}^n$ and $x \in \mathbb{R}^n$. We define the **upper and lower derivative** of ν with respect to μ, respectively, at x as

$$\overline{D}_\mu \nu(x) = \limsup_{r \to 0} \frac{\nu(B_r(x))}{\mu(B_r(x))}$$
$$\underline{D}_\mu \nu(x) = \liminf_{r \to 0} \frac{\nu(B_r(x))}{\mu(B_r(x))}$$

(where we adopt the convention that $\frac{0}{0} = 0$). If $\overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x) < \infty$, then we say that ν is **differentiable** with respect to μ and write

$$D_\mu \nu(x) = \overline{D}_\mu \nu(x) = \underline{D}_\mu \nu(x).$$

We say that $D_\mu \nu$ is the **Besicovitch derivative** of ν at x with respect to μ or also that $D_\mu \nu$ is the **density** of ν with respect to μ.

Example 4.3.2. Consider $\mu = \lambda^n$ be Lebesgue measure on $\mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$ be continuous. We set $\nu = f\,\mu$, i.e. $\nu(A) = \int_A f \mathrm{d}\mu$. Then it holds that

$$\frac{\nu(B_r(x))}{\mu(B_r(x))} = \frac{\int_{B_r(x)} f \mathrm{d}\mu}{\lambda^n(B_r(x))} \to f(x).$$

I.e. $D_\mu \nu(x) = D_\mu(f\,\mu)(x) = f(x)$ which explain the word "density".

Our next lemma looks quite innocent but will be a frequently used tool. It is also the place in where we employ the Besicovitch covering theorem.

Lemma 4.3.3. *Let $0 < \alpha < \infty$. Then it holds that*

1. $A \subset \{x : \underline{D}_\mu \nu(x) \le \alpha\}$ *implies* $\nu(A) \le \alpha \mu(A)$.
2. $A \subset \{x : \overline{D}_\mu \nu(x) \ge \alpha\}$ *implies* $\nu(A) \ge \alpha \mu(A)$.

(No measurability of A assumed.)

Proof. We may assume $\mu(\mathbb{R}^n), \nu(\mathbb{R}^n) < \infty$, since otherwise we could consider μ and ν restricted to a compact subset of $\mathbb{R}^n$.

Let $\epsilon > 0$, U open, such that $A \subset U$ with A from 1. Set

$$\mathcal{F} = \{B : B = B_r(a),\ a \in A, B \subset U, \nu(B) \le (\alpha + \epsilon)\mu(B)\}.$$

Then $\inf\{r : B_r(a) \in \mathcal{F}\} = 0$ for each $a \in A$ and by Corollary 4.2.6 there is a countable collection $\mathcal{G}$ of disjoint balls in $\mathcal{F}$ such that

$$\nu(A \setminus \bigcup_{B \in \mathcal{G}} B) = 0.$$

Then

$$\nu(A) \leq \sum_{B\in\mathcal{G}} \nu(B) \leq (\alpha+\epsilon) \sum_{B\in\mathcal{G}} \mu(B) \leq (\alpha+\epsilon)\mu(U).$$

This holds for any $U \supset A$ and by regularity of μ this leads to $\nu(A) \leq (\alpha + \epsilon)\mu(A)$. Since this holds for any $\epsilon > 0$ we are done with 1. The proof of 2. is similar. □

This lemma allows us to prove that the derivative of one Radon measure with respect to another one always exists:

Theorem 4.3.4. *Let μ and ν be Radon measures on $\mathbb{R}^n$. The $D_\mu\nu$ exists and is finite μ-a.e. and $D_\mu\nu$ is μ-measurable.*

Proof. Again we can concentrate on $\nu(\mathbb{R}^n), \mu(\mathbb{R}^n) < \infty$.

Set $I = \{x : \overline{D}_\mu\nu(x) = \infty\}$ and for $0 < a < b$ define $R(a,b) = \{x : \underline{D}_\mu\nu(x) < a < b < \overline{D}_\mu\nu(x) < \infty\}$. Note that for each $\alpha > 0$, $I \subset \{x : \overline{D}_\mu\nu(x) \geq \alpha\}$ and thus, by Lemma 4.3.3

$$\mu(I) \leq \tfrac{1}{\alpha}\nu(I).$$

Sending $\alpha \to \infty$ we conclude $\mu(I) = 0$ and hence, $\overline{D}_\mu\nu$ is finite μ-a.e.

Evoking Lemma 4.3.3 again, we get

$$b\,\mu(R(a,b)) \leq \nu(R(a,b)) \leq a\,\mu(R(a,b)),$$

but since $b > a$ we obtain $\mu(R(a,b)) = 0$. Furthermore,

$$\{x : \underline{D}_\mu\nu(x) < \overline{D}_\mu\nu(x)\} = \bigcup_{\substack{0<a<b\\ a,b\in\mathbb{Q}}} R(a,b)$$

and consequently, $D_\mu\nu$ exists and is finite μ-a.e.

We prove the auxiliary statement that for each $x \in \mathbb{R}^n$ and $r > 0$ we have $\limsup_{y\to x} \mu(B_r(y)) \leq \mu(B_r(x))$ (similar for ν): Choose a sequence $y_k \to x$ and set $f_k = \mathbb{1}_{B_r(y_k)}$ and $f = \mathbb{1}_{B_r(x)}$. Then $\limsup f_k \leq f$ and hence, by reversed Fatou's Lemma, $\int_{B_{2r}(x)} f\mathrm{d}\mu \geq \int_{B_{2r}(x)} \limsup f_k \mathrm{d}\mu \geq \limsup \int_{B_{2r}(x)} f_k\mathrm{d}\mu$. That is $\mu(B_r(x)) \geq \limsup \mu(B_r(y_k))$ as claimed.

Next we show that $D_\mu\nu$ is μ-measurable. We know from the previous step, that the functions $x \mapsto \mu(B_r(x))$ and $x \mapsto \nu(B_r(x))$ are upper semi-continuous and hence, Borel measurable. Consequently, for every $r > 0$

$$f_r(x) = \frac{\nu(B_r(x))}{\mu(B_r(x))}$$

is μ-measurable. Since $D_\mu\nu = \lim_{r\to 0} f_r$ it follows that $D_\mu\nu$ is μ-measurable as limit of μ-measurable functions. □

Once we have a notion for derivatives of measure, the question arises, if a measure can be reconstructed from its derivative (in the spirit of the fundamental theorem of calculus). It turns out, that the answer to this question depends on the following notion:

Definition 4.3.5. A measure ν is said to be **absolutely continuous** with respect to a measure μ, written as $\mu \ll \nu$, if for $A \subset \mathbb{R}^n$ it holds that $\mu(A) = 0$ implies $\nu(A) = 0$.

The measure μ and ν are said to be **mutually singular**, written as $\mu \perp \nu$, if there exists a Borel set $B \subset \mathbb{R}^n$ such that $\mu(B^{\complement}) = 0$ and $\nu(B) = 0$.

Remark 4.3.6. Note that if $\nu \ll \mu$ and A μ-measurable, then there exists a Borel set B with $A \subset B$ and $\mu(B \setminus A) = 0$. Hence $\nu(B \setminus A) = 0$ which shows that A is also ν-measurable.

Theorem 4.3.7 (Differentiation theorem for Radon measures)**.** *Let μ, ν be Radon measures on $\mathbb{R}^n$ such that $\nu \ll \mu$. Then it holds for all μ-measurable set A that*

$$\nu(A) = \int_A D_\mu \nu \, \mathrm{d}\mu.$$

Remark 4.3.8. The theorem could be called "fundamental theorem for the calculus of measures" as it roughly says "differentiation is inverted by integration". To make the analogy more clear, consider a monotone and continuously differentiable function $f : \mathbb{R} \to \mathbb{R}$ with $f(x) \to 0$ for $x \to -\infty$ and define the measure $\nu([a, b]) = f(b) - f(a)$. Let's calculate the derivative of this ν with respect to the Lebesgue measure λ:

$$D_\lambda \nu(x) = \lim_{r\to 0} \frac{\nu([x-r, x+r])}{\lambda([x-r, x+r])} = \lim_{r\to 0} \frac{f(x+r) - f(x-r)}{2r} = f'(x).$$

The differentiation theorem for Radon measures in the case of $A =]-\infty, x]$ then reads as

$$f(x) = \nu(]-\infty, x]) = \int_{]-\infty,x]} D_\lambda \nu(x) \mathrm{d}\lambda = \int_{-\infty}^{x} f'(x) \mathrm{d}x.$$

Moreover, it is also a formulation of the so-called **Radon-Nikodym theorem** which is often stated as an existence theorem as follows: If $\nu \ll \mu$, then there exists a function f such that $\nu = f\mu$. Here we do not have to provide a prove of existence, but have the machinery at hand, to calculate the related function f directly as $D_\mu \nu$.

Proof. By Remark 4.3.6, all μ-measurable sets are also ν-measurable.

We define the μ-measurable sets

$$Z = \{x \,:\, D_\mu \nu(x) = 0\}$$
$$I = \{x \,:\, D_\mu \nu(x) = \infty\}.$$

By Theorem 4.3.4, we have $\mu(I) = 0$ and hence $\nu(I) = 0$. Also Lemma 4.3.3 implies that $\nu(Z) \le \alpha\mu(Z)$ for all $\alpha > 0$ and hence $\nu(Z) = 0$. We conclude that

$$\nu(Z) = 0 = \int_Z D_\mu \nu \mathrm{d}\mu$$

and

$$\nu(I) = 0 = \int_I D_\mu \nu \mathrm{d}\mu.$$

Now let A be μ-measurable and for some $t > 1$. For each $m \in \mathbb{Z}$ define

$$A_m = \{x \in A \;:\; t^m \le D_\mu\nu(x) < t^{m+1}\}.$$

Then the A_m are μ- and hence, ν-measurable. Moreover

$$A \setminus \bigcup_{m\in\mathbb{Z}} A_m \subset Z \cup I \cup \{x \;:\; \overline{D}_\mu\nu(x) \neq \underline{D}_\mu\nu(x)\}.$$

Thus,

$$\mu(A \setminus \bigcup_{m\in\mathbb{Z}} A_m) = \nu(A \setminus \bigcup_{m\in\mathbb{Z}} A_m) = 0.$$

Consequently, by Lemma 4.3.3

$$\begin{aligned}
\nu(A) &= \sum_{m\in\mathbb{Z}} \nu(A_m) \le \sum_{m\in\mathbb{Z}} t^{m+1}\mu(A^m) \\
&= t\sum_{m\in\mathbb{Z}} t^m\mu(A_m) \le t\sum_{m\in\mathbb{Z}} \int_{A_m} D_\mu\nu \,\mathrm{d}\mu \\
&= t\int_A D_\mu\nu \,\mathrm{d}\mu.
\end{aligned}$$

Similarly, again by Lemma 4.3.3

$$\begin{aligned}
\nu(A) &= \sum_{m\in\mathbb{Z}} \nu(A_m) \ge \sum_{m\in\mathbb{Z}} t^{m}\mu(A^m) \\
&= \tfrac{1}{t}\sum_{m\in\mathbb{Z}} t^{m+1}\mu(A_m) \ge \tfrac{1}{t}\sum_{m\in\mathbb{Z}} \int_{A_m} D_\mu\nu \,\mathrm{d}\mu \\
&= \tfrac{1}{t}\int_A D_\mu\nu \,\mathrm{d}\mu.
\end{aligned}$$

Combining both estimates, we obtain for any $t > 1$

$$\tfrac{1}{t}\int_A D_\mu\nu \,\mathrm{d}\mu \le \nu(A) \le t\int_A D_\mu\nu \,\mathrm{d}\mu.$$

The conclusion follows by sending $t \to 1$. □

The differentiation theorem allows us, to derive the following fundamental decomposition theorem for Radon measures:

Theorem 4.3.9 (Lebesgue decomposition). *Let μ and ν be Radon measure on $\mathbb{R}^n$. The ν can be decomposed as $\nu = \nu_{\mathrm{ac}} + \nu_{\mathrm{s}}$ with Radon measures ν_{ac} and μ_{s} such that*

$$\nu_{\mathrm{ac}} \ll \mu \ \text{ and } \ \nu_s \perp \mu,$$

i.e., ν_{ac} is absolutely continuous w.r.t. μ and ν_{s} is singular to μ.

Furthermore it holds that

$$D_\mu \nu = D_\mu \nu_{\mathrm{ac}} \ \text{ and } \ D_\mu \nu_{\mathrm{s}} = 0 \ \mu\text{-a.e.}$$

and consequently, for each Borel set A

$$\nu(A) = \int_A D_\mu \nu \, \mathrm{d}\mu + \nu_{\mathrm{s}}(A).$$

Proof. As before we consider only finite measures μ and ν.

Define

$$\mathcal{M} = \{A \subset \mathbb{R}^n \ : \ A \ \text{Borel}\,, \ \mu(A^{\complement}) = 0\}$$

and choose $B_k \in \mathcal{M}$ such that for $k = 1, \ldots$

$$\nu(B_k) \leq \inf_{A \in \mathcal{M}} \nu(A) + \tfrac{1}{k}.$$

and set $B = \bigcap_{k=1}^{\infty} B_k$. Since

$$\mu(B^{\complement}) \leq \sum_{k=1}^{\infty} \mu(B_k^{\complement}) = 0$$

we know that $B \in \mathcal{M}$ and hence,

$$\nu(B) = \inf_{A \in \mathcal{M}} \nu(A). \tag{*}$$

Now we define

$$\nu_{\mathrm{ac}} = \nu \llcorner B, \ \text{ and } \ \nu_{\mathrm{s}} = \nu \llcorner B^{\complement}.$$

From Lemma 2.3.8 and Corollary 2.3.10 we conclude that ν_{ac} and ν_{s} are Radon measures.

Now assume that $A \subset B$ is a Borel set, $\mu(A) = 0$ but $\nu(A) > 0$. Then $B \setminus A \in \mathcal{M}$ and $\nu(B \setminus A) < \nu(B)$ but this contradicts (*). Consequently $\nu(A) = 0$ for $A \subset B$ and hence $\nu_{\mathrm{ac}} \ll \mu$. On the other hand $\mu(B^{\complement}) = 0$ which shows $\nu_s \perp \mu$.

Finally, fix $\alpha > 0$ and set $C = \{x \in B \ : \ D_\mu \nu_{\mathrm{s}}(x) \geq \alpha\}$. By Lemma 4.3.3 we obtain $\alpha \mu(C) \leq \nu_s(C) = 0$ and therefore, $D_\mu \nu_s = 0$ μ-a.e. This implies $D_\mu \nu_{\mathrm{ac}} = D_\mu \nu$ μ-a.e. $\square$

We now turn to something that may be called an application of differentiation of measures: The notion of Lebesgue points.

In the following it will be helpful to denote the average of a function f over some set E (w.r.t. a measure μ) by

$$\fint_E f \mathrm{d}\mu = \frac{1}{\mu(E)} \int_E f \mathrm{d}\mu$$

whenever $0 < \mu(E) < \infty$ and the integral on the right hand side is defined. We also use the notion of a **locally integrable function** and write $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mu)$ if $f : \mathbb{R}^n \to \mathbb{R}$ is μ-measurable and $\int_K |f| \mathrm{d}\mu$ exists for every compact set K. Similarly, $f \in L^p_{\mathrm{loc}}(\mathbb{R}^n, \mu)$ for $p > 1$ if $\int_K |f|^p \mathrm{d}\mu < \infty$ for every compact K.

Theorem 4.3.10 (Lebesgue differentiation theorem)**.** *Let μ be a Radon measure on $\mathbb{R}^n$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}^n, \mu)$. Then it holds for μ-a.e. $x \in \mathbb{R}^n$ that*

$$\lim_{r\to 0} \fint_{B_r(x)} f \mathrm{d}\mu = f(x).$$

You shall do the proof as an **exercise**.

In a spirit similar to the Lebesgue differentiation theorem, we define the following:

Definition 4.3.11. Let μ be a Radon measure on $\mathbb{R}^n$ and $1 \le p < \infty$ and $f \in L^p_{\mathrm{loc}}(\mathbb{R}^n, \mu)$. Then $x \in \mathbb{R}^n$ is called a **Lebesgue point** of f if

$$\lim_{r\to 0} \fint_{B_r(x)} |f - f(x)|^p \mathrm{d}\mu = 0.$$

Theorem 4.3.12. *For a Radon measure μ, $1 \le p < \infty$ and $f \in L^p_{\mathrm{loc}}(\mathbb{R}, \mu)$, μ-a.e. points are Lebesgue points.*

Again, work out the proof as an **exercise**.

For the case of the Lebesgue measure, one can replace the balls centered at x by arbitrary balls that contain x:

Corollary 4.3.13. *Let $f \in {}^p_{\mathrm{loc}}(\mathbb{R}^n, \lambda^n)$ for some $1 \le p < \infty$. Then it holds for λ^n-a.e. x that*

$$\lim_{\mathrm{diam}(B)\to 0} \fint_B |f(y) - f(x)| \mathrm{d}y = 0$$

where the limit is taken over all balls B containing x.

Proof. Let B_k be a sequence of closed balls that contain x and $d_k = \mathrm{diam}(B_k) \to 0$. Then $B_k \subset B_{d_k}(x)$, thus $2^n \lambda^n(B_k) \ge \lambda^n(B_{d_k}(x))$, and consequently, for every Lebesgue point (w.r.t. λ^n)

$$\begin{aligned}\fint_{B_k} |f(y) - f(x)|^p \mathrm{d}y &\le \frac{1}{\lambda^n(B_k)} \int_{B_{d_k}(x)} |f(y) - f(x)|^p \mathrm{d}y \\ &\le \frac{2^n}{\lambda^n(B_{d_k}(x))} \int_{B_{d_k}(x)} |f(y) - f(x)|^p \mathrm{d}y \\ &= 2^n \fint_{B_{d_k}(x)} |f(y) - f(x)|^p \mathrm{d}y \to 0.\end{aligned}$$

$\square$

Finally we remark on a generalization of the Lebesgue decomposition theorem to vector measures: For a positive Radon measure μ and a signed or vector measure ν we say that ν is **absolutely continuous** continuous w.r.t. μ if $|\nu|$ is absolutely continuous w.r.t μ, i.e. $\nu \ll \mu$ if $|\nu| \ll \mu$. Similarly, we say that two signed or vector measure μ and ν are **mutually singular** if $|\mu|$ and $|\nu|$ are so, i.e. $\mu \perp \nu$ if $|\mu| \perp |\nu|$. We have the following "vector form" of the Lebesgue decomposition theorem:

Corollary 4.3.14. *Let μ be a Radon measure and ν be a $\mathbb{R}^m$-valued vector measure, both on $\mathbb{R}^n$. Then there exist vector measures ν_{ac} and ν_s such that $\nu = \nu_{\mathrm{ac}} + \nu_s$, $\nu_{\mathrm{ac}} \ll \mu$ and $\nu_s \perp \mu$. Moreover, there exists a function $f \in L^1(\mathbb{R}^n, \mu)^m$ (uniquely determined μ-a.e.) such that $\nu_{\mathrm{ac}} = f\mu$.*

The **proof** simply consists of the coordinate-wise application of the Lebesgue decomposition theorem to ν_i and μ. The function $f = (f_1, \ldots, f_m)$ is given by $f_i = D_\mu \nu_i$.

When we apply the above theorem in the special case of $\mu = |\nu|$ (obviously $\nu \ll |\nu|$) we obtain:

Corollary 4.3.15 (Polar decomposition)**.** *Let ν by a $\mathbb{R}^m$-valued vector measure on $\mathbb{R}^n$. Then there exists a function $\sigma \in L^1(\mathbb{R}^n, |\mu|)^m$ with $|\sigma| = 1$ $|\nu|$-a.e. such that $\nu = \sigma|\nu|$.*

Proof. By the above theorem we get the existence of a function σ such that $\nu = \sigma|\nu|$ and it remains to show that $|\sigma| = 1$. This follows from the following general statement (that can be found, e.g., in [AFP00, Proposition 1.23]): For a positive measure μ and $g \in L^1(\mathbb{R}^n, \mu)^m$ it holds that $|g\mu| = |g|\mu$ (since then $|\nu| = |\sigma|\nu|| = |\sigma||\nu|$). □

5 Riesz representation and weak convergence of measures

The next part of the trip goes through a realm that is too lovely to not make a stop: functional analysis. We will see that measure theory can be paired in an intricate way with the spaces of continuous functions in a functional analytic manner. In fact, we will collect measures into their own Banach spaces and these spaces are in fact pretty rich in structure.

5.1 Riesz representation

Definition 5.1.1. For $\Omega \subset \mathbb{R}^n$ we denote by $C(\Omega, \mathbb{R}^m)$ the set of continuous functions from $\Omega \to \mathbb{R}^m$ and by $C_c(\Omega, \mathbb{R}^m)$ the set of continuous functions from $\Omega \to \mathbb{R}^m$ that have compact support.

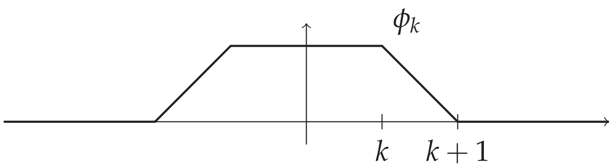
The sets $C(\Omega, \mathbb{R}^m)$ and $C_c(\Omega, \mathbb{R}^m)$ are vector spaces. Note that the spaces behave extremely differently if Ω is open or compact: for open Ω, the space $C(\Omega, \mathbb{R}^m)$ contains some weird functions since they may grow arbitrarily fast towards the boundary of Ω (consider $\exp(x^{2n})$ or $\exp(\exp(\exp(x^2)))$ or the like on $\Omega = \mathbb{R}$). However, for compact Ω, all functions in $C(\Omega, \mathbb{R}^m)$ are bounded and attain their maximum. Hence, the space $C_c(\Omega, \mathbb{R}^m)$ is in some sense more natural for general Ω.

We equip $C_c(\Omega, \mathbb{R}^m)$ with the sup norm and since $f \in C_c(\Omega, \mathbb{R}^m)$ attains its supremum (as a continuous function on a compact set) we set

$$\|f\|_\infty = \max\{|f(x)| \;:\; x \in \Omega\}.$$

(Note that it does not make sense to define the sup-norm $C(\Omega, \mathbb{R}^m)$ if Ω is not compact.)

Remark 5.1.2. The space $C_c(\Omega, \mathbb{R}^m)$ is not complete if Ω is not compact: In the case $\Omega = \mathbb{R}$ consider the function $f(x) = \exp(-x^2)$. This function f does not lie in $C_c(\Omega, \mathbb{R}^m)$ but it can be approximated: Consider continuous cutoff-functions ϕ_k as follows



and note that the sequence $f\phi_k$ is a Cauchy sequence in $C_c(\mathbb{R},\mathbb{R})$.

For general sets Ω (say, open), a similar construction is possible: Exhaust Ω by a sequence of increasing compact sets K_k and define cutoff function ϕ_k such that $\phi_k \equiv 1$ on K_k, $\phi_k \equiv 0$ outside of K_{k+1}, and continuous*. Then, for any $f : \Omega \to \mathbb{R}$ for which $\|f|_{K_k^\complement}\|_\infty \to 0$ it holds that $f\phi_k$ is Cauchy in $C_c(\Omega,\mathbb{R})$.

In view of this remark, it makes sense, to consider the closure w.r.t. the sup norm:

Lemma 5.1.3. *We define by $C_0(\Omega,\mathbb{R}^m)$ the closure of $C_c(\Omega,\mathbb{R}^m)$ with respect to the sup norm. A function $f : \Omega \to \mathbb{R}^m$ is in $C_0(\Omega,\mathbb{R}^m)$ if it is continuous and for every $\epsilon > 0$ there exists a compact set $K \subset \Omega$ such that* $\sup\{|f(x)| \;:\; x \notin K\} < \epsilon$.

Proof. If f satisfies the stated conditions, then it can be approximated by functions in $C_c(\Omega,\mathbb{R}^m)$ (similarly as we have seen in the above remark). On the other hand, let f be a uniform limit of functions (f_k) in $C_c(\Omega,\mathbb{R}^m)$. Then $\|f - f_k\|_\infty \to 0$ and especially f has to be arbitrarily small outside the support of f_k. $\square$

Usually one says that $C_0(\Omega,\mathbb{R}^m)$ is the space of functions that "vanish at the open boundary of Ω". By definition, $C_0(\Omega,\mathbb{R}^m)$ is a Banach space.

In the case $m = 1$ abbreviate to $C(\Omega)$, $C_c(\Omega)$, and $C_0(\Omega)$.

Exercise 19. Additionally to the spaces $C(\Omega,\mathbb{R}^m)$, $C_c(\Omega,\mathbb{R}^m)$ and $C_0(\Omega,\mathbb{R}^m)$ one considers $C_b(\Omega,\mathbb{R}^m)$, the space of **bounded** continuous functions (also equipped with the sup norm).

1. Is $C_b(\Omega,\mathbb{R}^m)$ a Banach space?

2. For open Ω, investigate the mutual inclusion between the four spaces. Also investigate if the inclusions are strict or if one has equality.

3. Treat the same problem for compact Ω.

Now consider a non-negative Radon measure μ on Ω and a μ-measurable function $\sigma : \Omega \to \mathbb{R}^m$ with $|\sigma| \leq 1$. Since functions f in $C_c(\Omega,\mathbb{R}^m)$ are bounded and Borel, we have

$$\int f \cdot \sigma \mathrm{d}\mu \leq \|f\|_\infty |\mu|(\mathrm{supp} f).$$

In other words: The mapping $f \mapsto \int f \cdot \sigma \mathrm{d}\mu$ defines a linear mapping from $C_c(\Omega,\mathbb{R}^m)$ to $\mathbb{R}$ and more, this mapping is continuous when restricted to $C_c(K,\mathbb{R}^m)$ for any compact set $K \subset \Omega$. Hence, every Radon measure is an element of the **dual space** of $C_c(\Omega,\mathbb{R}^m)$. The Riesz representation theorem is basically the fact that indeed all such functionals on $C_c(\Omega,\mathbb{R}^m)$ are given

*The existence of such function is ensured by Urysohn's Lemma.

in this way. This theorem is tremendously helpful when dealing with Radon measures and opens the way to treat Radon measures by duality.

For convenience, we formulate the following theorem for functions on the whole space $\Omega = \mathbb{R}^n$ and first state the version of the theorem for functionals on the space $C_c(\mathbb{R}^n, \mathbb{R}^m)$.

Theorem 5.1.4 (Riesz representation theorem for $C_c(\mathbb{R}^n, \mathbb{R}^m)$)**.** *Let L be a linear functional on $C_c(\mathbb{R}^n, \mathbb{R}^m)$ that is continuous when restricted to $C_c(K, \mathbb{R}^m)$ for any compact $K \subset \mathbb{R}^n$. Then there exists a non-negative Radon measure μ and a μ-measurable function $\sigma : \mathbb{R}^n \to \mathbb{R}^m$ such that $|\sigma| = 1$ (μ-a.e.) and for all $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ it holds that*

$$L(f) = \int_{\mathbb{R}^n} f \cdot \sigma \mathrm{d}\mu.$$

Remark 5.1.5. The condition that L is continuous when restricted to $C_c(K, \mathbb{R}^m)$ for any compact K can be written as

$$\sup\{L(f) \;:\; f \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ \|f\|_\infty \leq 1,\ \operatorname{supp}(f) \subset K\} < \infty.$$

Proof. This is another lengthy and technical proof.

1. We start by defining for every open set $V \subset \mathbb{R}^n$

   $$\mu(V) = \sup\{L(f) \;:\; f \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ \|f\|_\infty \leq 1,\ \operatorname{supp}(f) \subset V\}$$

   and then for arbitrary $A \subset \mathbb{R}^n$

   $$\mu(A) = \inf\{\mu(V) \;:\; A \subset V \text{ open}\}.$$

   Note that this construction is very similar to the construction of the Lebesgue measure on $\mathbb{R}$ in Definition 2.2.1. However, we do not cover general sets A by "simpler ones" but take the infimum over all larger open sets. Hence, we prove by hand that μ is indeed a measure:

   Let V and V_i, $i \in \mathbb{N}$, be open sets in $\mathbb{R}^n$ with $V \subset \bigcup_{i \in \mathbb{N}} V_i$. Choose some $g \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ with $\|g\|_\infty \leq 1$ and $\operatorname{supp}(g) \subset V$. Since $\operatorname{supp}(g)$ is compact and covered by the V_i's, there exists a finite subcover which we take to be $\operatorname{supp}(g) \subset \bigcup_{j=0}^{k} V_j$. Now we need a so called "smooth partition of unity" on the V_j, i.e. smooth functions ζ_j such that $\operatorname{supp}(\zeta_j) \subset V_j$ for $0 \leq j \leq k$ and $\sum_{j=0}^{k} \zeta_j(x) = 1$ for $x \in \operatorname{supp}(g)$.[†] Then we can write $g = \sum_{j=0}^{k} g\zeta_j$ and it follows

   $$|L(g)| = |\sum_{j=0}^{k} L(g\zeta_j)| \leq \sum_{j=0}^{k} |L(g\zeta_j)| \leq \sum_{j=0}^{k} \mu(V_j).$$

[†] Partitions of unity exists under a huge variety of circumstances (and especially, with any desired smoothness). Their construction is often cumbersome and we do not go into the details here.

Taking the supremum over the respective g we obtain $\mu(V) \leq \sum_{j=0}^{k} \mu(V_j)$. Now let A_j, $j \in \mathbb{N}$, be arbitrary sets with $A \subset \bigcup_{j\in\mathbb{N}} A_j$. We fix $\epsilon > 0$ and choose open sets V_j such that $A_j \subset V_j$ and $\mu(A_j) + \epsilon/2^j \geq \mu(V_j)$. Then

$$\mu(A) \leq \mu(\bigcup_{j\in\mathbb{N}} V_j) \leq \sum_{j\in\mathbb{N}} \mu(V_j) \leq \sum_{j\in\mathbb{N}} \mu(A_j) + \epsilon$$

which shows that μ is countably subadditive and hence, an (outer) measure.

2. Now we aim to prove that μ is a Radon measure:
a) μ is Borel: Let U_1 and U_2 be open with $\mathrm{dist}(U_1, U_2) > 0$. By the definition of μ we get $\mu(U_1 \cup U_2) = \mu(U_1) + \mu(U_2)$, i.e. additivity of the sets are a bit apart (cf. a similar fact for the Lebesgue measure, Lemma 2.2.11). Consequently $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$ for any $A_{1/2} \subset \mathbb{R}^n$ with $\mathrm{dist}(A_1, A_2) > 0$. Now it is a general fact that (outer) measures with the property that they are additive on "separated sets" are Borel measures (see, e.g.[EG92, Section 1.1.1, Theorem 5]).
b) μ is Borel regular: Zhe very definition of μ shows that μ is indeed Borel regular: To see this consider $A \subset \mathbb{R}^n$ and V_k such that $A \subset V_k$ and $\mu(V_k) \leq \mu(A) + \frac{1}{k}$. Then

$$\mu(\bigcap_j V_j) \leq \mu(V_k) \leq \mu(A) + \tfrac{1}{k} \leq \mu(\bigcap_j V_j) + \tfrac{1}{k}$$

and thus, $\mu(A) = \mu(\bigcap_k V_k)$ and $\bigcap_k V_k$ is obviously Borel.
c) μ is Radon: By the assumed continuity of L we have that $\mu(B_r(x))$ is finite, since $\overline{B_r(x)}$ is compact. Corollary 2.3.10 shows that μ is a Radon measure.

3. As next step we consider $f \in C_c(\mathbb{R}^n)$ with $f \geq 0$ and set

$$\Lambda(f) = \sup\{|L(g)| \ : \ g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \leq f\}.$$

We aim to show that Λ is a continuous, linear functional (on the cone of non-negative, continuous functions with compact support).

First observe that $\Lambda(f_1) \leq \Lambda(f_2)$ for $f_1 \leq f_2$ and also $\Lambda(cf) = c\Lambda(f)$ for $c \geq 0$ (i.e. Λ is monotone and positively homogeneous). To see additivity, i.e. that $\Lambda(f_1 + f_2) = \Lambda(f_1) + \Lambda(f_2)$ consider $g_{1/2} \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ with $|g_i| \leq f_i$. Then $|g_1 + g_2| \leq |g_1| + |g_2| \leq f_1 + f_2$. Without loss of generality, we can assume that $L(g_i) \geq 0$ (otherwise take $-g_i$ instead). Therefore,

$$|L(g_1)| + |L(g_2)| = L(g_1 + g_2) = |L(g_1 + g_2)| \leq \Lambda(f_1 + f_2).$$

Taking the suprema over $g_1, g_2 \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ we obtain

$$\Lambda(f_1) + \Lambda(f_2) \leq \Lambda(f_1 + f_2).$$

Now fix $g \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ with $|g| \le f_1 + f_2$ and set for $i = 1, 2$

$$g_i = \begin{cases} \frac{f_i g}{f_1 + f_2} & \text{if } f_1 + f_2 > 0 \\ 0 & \text{else.} \end{cases}$$

Obviously, $g_{1/2} \in C_c(\mathbb{R}^n, \mathbb{R}^m)$, $g = g_1 + g_2$ and $|g_i| \le f_i$, so that

$$|L(g)| \le |L(g_1)| + |L(g_2)| \le \Lambda(f_1) + \Lambda(f_2)$$

and consequently $\Lambda(f_1 + f_2) \le \Lambda(f_1) + \Lambda(f_2)$.

4. Now we show that Λ can be written as an integral, namely $\Lambda(f) = \int f \mathrm{d}\mu$ for $f \in C_c^+(\mathbb{R}^n)$: Let $\epsilon > 0$ and choose $0 = t_0 < t_1 < \cdots < t_N$ such that $t_N = 2\|f\|_\infty$, $0 < t_i - t_{i-1} < \epsilon$ and $\mu(f^{-1}(t_i)) = 0$ for $i = 1, \ldots, N$. Set $U_j = f^{-1}(]t_{j-1}, t_j[)$. The sets U_j are open and $\mu(U_j) < \infty$. Since μ is a Radon measure, there exist compact sets K_j such that $K_j \subset U_j$ and $\mu(U_j \setminus K_j) < \epsilon/N$. Also there exist functions $g_j \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ with $|g_j| \le 1$, $\operatorname{supp}(g_j) \subset U_j$ and $L(g_j) \ge \mu(U) - \epsilon/N$. Note also that there exist functions $h_j \in C_c^+(\mathbb{R}^n)$ such that $\operatorname{supp}(h_j) \subset U_j$, $0 \le h_j \le 1$ and $h_j = 1$ on $K_j \cup \operatorname{supp}(g_j)$ (which is a compact set). Then $h_j \ge |g_j|$ and hence

$$\Lambda(h_j) \ge |L(g_j)| \ge \mu(U_j) - \epsilon/N$$

and on the other hand

$$\begin{aligned} \Lambda(h_j) &= \sup\{|L(g)| \;:\; g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \le h_j\} \\ &\le \sup\{|L(g)| \;:\; g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \le 1,\ \operatorname{supp}(g) \subset U_j\} \\ &= \mu(U_j) \end{aligned}$$

and consequently, $\mu(U_j) - \epsilon/N \le \Lambda(h_j) \le \mu(U_j)$.

Now define

$$A = \{x \;:\; f(x)(1 - \sum_{j=1}^N h_j(x)) > 0\}$$

which is, by continuity of f and the h_j, an open set. We compute

$$\begin{aligned} \Lambda(f - f\sum_{j=1}^N h_j) &= \sup\{|L(g)| \;:\; g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \le f(1 - \sum_{j=1}^N h_j)\} \\ &\le \sup\{|L(g)| \;:\; g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \le \|f\|_\infty \mathbb{1}_A\} \\ &\le \|f\|_\infty \sup\{L(g) \;:\; g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \le \mathbb{1}_A\} \\ &= \|f\|_\infty \mu(A) \\ &= \|f\|_\infty \mu(\bigcup_{j=1}^N (U_j \setminus \{h_j = 1\})) \\ &\le \|f\|_\infty \sum_{j=1}^N \mu(U_j \setminus K_j) \le \epsilon \|f\|_\infty. \end{aligned}$$

It follows that

$$\begin{aligned}\Lambda(f) &= \Lambda(f - f\sum_{j=1}^N h_j) + \Lambda(f\sum_{j=1}^N h_j) \leq \epsilon\|f\|_\infty + \sum_{j=1}^N \Lambda(fh_j) \\ &\leq \epsilon\|f\|_\infty + \sum_{j=1}^N t_j\mu(U_j)\end{aligned}$$

and on the other hand

$$\begin{aligned}\Lambda(f) &\geq \sum_{j=1}^N \Lambda(fh_j) \\ &\geq \sum_{j=1}^N t_{j-1}(\mu(U_j) - \epsilon/N) \\ &\geq \sum_{j=1}^N t_{j-1}\mu(U_j) - t_N\epsilon.\end{aligned}$$

We obtain the inequality

$$\sum_{j=1}^N t_{j-1}\mu(U_j) - t_N\epsilon \leq \Lambda(f) \leq \epsilon\|f\|_\infty + \sum_{j=1}^N t_j\mu(U_j)$$

which we combine with

$$\sum_{j=1}^N t_{j-1}\mu(U_j) \leq \int_{\mathbb{R}^n} f\mathrm{d}\mu \leq \sum_{j=1}^N t_j\mu(U_j)$$

to obtain

$$\begin{aligned}|\Lambda(f) - \int f\mathrm{d}\mu| &\leq \sum_{j=1}^N (t_j - t_{j-1})\mu(U_j) + \epsilon\|f\|_\infty + \epsilon t_N \\ &\leq \epsilon\mu(\mathrm{supp}(f)) + 3\epsilon\|f\|_\infty\end{aligned}$$

as desired.

5. Now we prove the existence of the desired function σ: For any direction $e \in \mathbb{R}^m$ (i.e. $|e| = 1$) we define $\Lambda_e(f) = L(ef)$ (defined for $f \in C_c(\mathbb{R}^n)$). This Λ_e is linear and it holds that

$$\begin{aligned}|\Lambda_e(f)| = |L(ef)| &\leq \sup\{|L(g)| \ : \ g \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ |g| \leq |f|\} \\ &= \Lambda(|f|) = \int |f|\mathrm{d}\mu.\end{aligned}$$

In other words: $|\Lambda_e(f)| \leq \|f\|_{L^1(\mathbb{R}^n,\mu)}$. By the Hahn-Banach theorem, we can extend Λ_e to a linear functional on the space $L^1(\mathbb{R}^n, \mu)$, i.e. it can be represented by a function $\sigma_e \in (L^1(\mathbb{R}^n, \mu))^* = L^\infty(\mathbb{R}^n, \mu)$:

$$\Lambda_e(f) = \int f\sigma_e\mathrm{d}\mu.$$

Now we take the standard basis $e_1, \dots, e_m$ of $\mathbb{R}^m$, define $\sigma = \sum_{j=1}^m \sigma_{e_j} e_j$. For any $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ we have $f = \sum_{j=1}^m (f \cdot e_j), e_j$ and hence,

$$L(f) = \sum_{j=1}^m L((f \cdot e_j)e_j) = \sum_{j=1}^m \int (f \cdot e_j)\sigma_j \mathrm{d}\mu = \int f \cdot \sigma \mathrm{d}\mu.$$

6. In the last step we show that $|\sigma| = 1$ μ-a.e.: Let $U \subset \mathbb{R}^n$ be open with $\mu(U) < \infty$. By definition

$$\mu(U) = \sup\{\int f \cdot \sigma \mathrm{d}\mu \;:\; f \in C_c(\mathbb{R}^n, \mathbb{R}^m),\ \|f\|_\infty \leq 1,\ \mathrm{supp}(f) \subset U\}.$$

Now we approximate the "direction $\sigma/|\sigma|$" by a continuous function, i.e. $f_k \in C_c(\mathbb{R}^n, \mathbb{R}^m)$, $|f_k| \leq 1$, $\mathrm{supp}(f_k) \subset U$ with $f_k \cdot \sigma \to |\sigma|$ μ-a.e. Then be dominated convergence and the above

$$\int_U |\sigma| \mathrm{d}\mu = \lim_{k\to\infty} \int_U f_k \cdot \sigma \mathrm{d}\mu \leq \mu(U).$$

On the other hand, if $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$ with $|f| \leq 1$ and $\mathrm{supp}(f) \subset U$, then

$$\int_U f \cdot \sigma \mathrm{d}\mu \leq \int_U |\sigma| \mathrm{d}\mu$$

i.e. $\mu(U) = \int_U |\sigma| \mathrm{d}\mu$ for all open sets U. It follows $|\sigma| = 1$ μ-a.e. as desired.

□

Corollary 5.1.6 (Riesz representation theorem for $C_0(\mathbb{R}^n, \mathbb{R}^m)$). *Every linear and continuous functional L on $C_0(\mathbb{R}^n, \mathbb{R}^m)$ can be represented by a $\mathbb{R}^m$-valued Radon measure ν, i.e. $L(f) = \int f \cdot \mathrm{d}\nu$.*

Proof. Observe that L does fulfill the condition in the Riesz representation theorem and hence there exists a non-negative measure μ and a μ-measurable function σ with $|\sigma| = 1$ μ-a.e. such that for all $f \in C_c(\mathbb{R}^n, \mathbb{R}^m)$

$$L(f) = \int f \cdot \sigma \mathrm{d}\mu.$$

Since the representation is valid on $C_c(\mathbb{R}^n, \mathbb{R}^m)$ it extends also to the closure $C_0(\mathbb{R}^n, \mathbb{R}^m)$. We set $\nu = \sigma\mu$ and need to show, that this is a vector valued Radon measure. Countable additivity on Borel sets is clear since μ is a Radon measure. It remains to show that ν is finite. But since L is continuous on $C_0(\mathbb{R}^n, \mathbb{R}^m)$ it holds that

$$\sup\{|L(f)| \;:\; f \in C_0(\mathbb{R}^n, \mathbb{R}^m),\ \|f\|_\infty \leq 1\} < \infty. \tag{*}$$

By the definition of μ in step 1 of the proof of Theorem 5.1.4, $\mu(\mathbb{R}^n)$ is precisely the supremum on the left hand side and hence, finite. □

Exercise 20. We define three linear functionals on spaces of continuous functions:

1. For $f \in C_c(\mathbb{R}, \mathbb{R})$ define the functional
$$L(f) = f(0) + \int_0^\infty f(x)\mathrm{d}x - \int_{-1}^0 f(x)\mathrm{d}x.$$
2. For $f \in C_c(\mathbb{R}^2, \mathbb{R})$ define the functional
$$L(f) = \int_{-\infty}^\infty f(x,0)\mathrm{d}x.$$
3. For $f \in C_c(\mathbb{R}^2, \mathbb{R}^2)$ define the functional
$$L(f) = \int_0^{2\pi} f(\cos(t), \sin(t)) \cdot \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \mathrm{d}x.$$

In each case: Show that L fulfills the conditions of the Riesz representation theorem and derive the measure ν (or, equivalently, the measure μ and the function σ).

Exercise 21. The functional $L(f) = f'(0)$ is defined on $C_c^1(\mathbb{R})$ which is a dense in $C_c(\mathbb{R})$. Is it possible to extend it to a functional on $C_c(\mathbb{R})$ that fulfills the conditions of the Riesz representation theorem?

Remark 5.1.7. In view of the definition of the space $\mathfrak{M}(\mathbb{R}^n, \mathbb{R}^m)$ (cf. Theorem 4.1.5) the Riesz representation theorem says that $\mathfrak{M}(\mathbb{R}^n, \mathbb{R}^m)$ (equipped with the variation norm) is the dual space of and $C_0(\mathbb{R}^n, \mathbb{R}^m)$ i.e.
$$C_0(\mathbb{R}^n, \mathbb{R}^m)^* = \mathfrak{M}(\mathbb{R}^n, \mathbb{R}^m).$$
Also the Riesz representation theorem stays true for the spaces $C_0(\Omega, \mathbb{R}^m)$ for open or compact sets $\Omega \subset \mathbb{R}^n$. In the following we will use this result also for this case.

A direct consequence of the duality of C_0 and $\mathfrak{M}$ is a new representation of the variation norm and the fact that $\mathfrak{M}$ is a Banach space.

Corollary 5.1.8. *The variation norm on $\mathfrak{M}(\Omega, \mathbb{R}^m)$ is given by*
$$\|\nu\|_{\mathfrak{M}} = \sup\{\int f \cdot \mathrm{d}\nu \ : \ f \in C_0(\Omega, \mathbb{R}^m), \ \|f\|_\infty \leq 1\}.$$
As we have identified $\mathfrak{M}(\Omega, \mathbb{R}^m)$ as a dual space, we can directly infer that it is complete and hence, a Banach space.

5.2 Convergence of sequences of measures

A norm on a space defines a notion of convergence. In the case of the variation norm on $\mathfrak{M}(\Omega, \mathbb{R}^m)$ this is

$$\mu_k \to \mu \iff \|\mu_k - \mu\|_{\mathfrak{M}} \to 0.$$

In the context of duality of Banach space, this convergence w.r.t. the norm is usually called **strong convergence**.

Example 5.2.1 (Strong convergence generalizes L^1 convergence)**.** Let ν be a Radon measure, f_k a sequence in $L^1(\Omega, \nu)$ and $\mu_k = f_k\nu$. Observe that

$$\begin{aligned}\|\mu_k\|_{\mathfrak{M}} &= \sup\{\int h f_k \mathrm{d}\nu \ : \ h \in C_0(\Omega),\ \|h\|_\infty \leq 1\}\\ &\leq \sup\{\int h f_k \mathrm{d}\nu \ : \ h \in L^\infty(\Omega),\ \|h\|_\infty \leq 1\}\\ &= \|f_k\|_1.\end{aligned}$$

If now $f_k \to f$ in $L^1(\Omega, \nu)$ we conclude that for $\mu = f\nu$ it holds that

$$\|\mu_k - \mu\|_{\mathfrak{M}} \leq \|f_k - f\|_1 \to 0.$$

Example 5.2.2 (Not strongly converging)**.** Consider a sequence x_k of points in Ω such that $x_k \to x \in \Omega$ (usual convergence in $\mathbb{R}^n$). For the Dirac measures δ_{x_k} and δ_x it holds that

$$\begin{aligned}\|\delta_{x_k} - \delta_x\|_{\mathfrak{M}} &= \sup\{\int h \mathrm{d}(\delta_{x_k} - \delta_x) \ : \ h \in C_0(\Omega),\ \|h\|_\infty \leq 1\}\\ &= \sup\{h(x_k) - h(x) \ : \ h \in C_0(\Omega),\ \|h\|_\infty \leq 1\}.\end{aligned}$$

If $x_k \neq x$ there is always a continuous function h such that $h(x_k) = 1$ and $h(x) = -1$ and we conclude that

$$\|\delta_{x_k} - \delta_x\|_{\mathfrak{M}} = 2.$$

In other words: δ_{x_k} does (in general) not strongly converge to δ_x if $x_k \to x$. Hence, the notion of strong convergence in $\mathfrak{M}(\Omega)$ does not seem to reflect the geometry of the underlying set Ω.

The duality of $C_0(\mathbb{R}^n, \mathbb{R}^m)$ and $\mathfrak{M}(\mathbb{R}^n, \mathbb{R}^m)$ naturally implies another notion of convergence on $\mathfrak{M}(\mathbb{R}^n, \mathbb{R}^m)$, namely the so-called weak* convergence:

Definition 5.2.3 (Weak* convergence in $\mathfrak{M}$)**.** A sequence μ_k in $\mathfrak{M}(\Omega, \mathbb{R}^m)$ **converges weakly*** to $\mu \in \mathfrak{M}(\Omega, \mathbb{R}^m)$ and we write $\mu_k \overset{*}{\rightharpoonup} \mu$ if for every $f \in C_0(\Omega, \mathbb{R}^m)$ it holds that

$$\int f \mathrm{d}\mu_k \to \int f \mathrm{d}\mu.$$

Note that there is also the **weak* topology** on $\mathfrak{M}$ (as well as on any other dual space of a normed space). However, usually the weak* topology can not be described in total by its convergent sequences, e.g. in general the closure of a set w.r.t the weak* topology is in general not the "sequential closure", i.e. the set of all limit points of convergent sequences. In this notes we only deal with weak* convergence and do not use the weak* topology in any way.

Example 5.2.4 (Weakly* converging δs)**.** Let us recheck Example 5.2.2: We observe that for any $f \in C_0(\Omega)$:

$$\int f \mathrm{d}\delta_{x_k} = f(x_k) \to f(x) = \int f \mathrm{d}\delta_x$$

and voilà:

$$\delta_{x_k} \stackrel{*}{\rightharpoonup} \delta_x.$$

Exercise 22.

1. Show that $\delta_n \stackrel{*}{\rightharpoonup} 0$ in $\mathfrak{M}(\mathbb{R})$ for $n \to \infty$.
2. Define $\mu_k = \frac{1}{k}\sum_{i=1}^{k} \delta_{i/k}$. Show that $\mu_k \stackrel{*}{\rightharpoonup} \lambda^1 \llcorner [0,1]$ in $\mathfrak{M}(\mathbb{R})$ for $k \to \infty$.

Exercise 23. Does the sequence $\mu_k = k\lambda^1 \llcorner [0, 1/k]$ converge weakly* or strongly? (In case of convergence, also provide the limit.)

Remark 5.2.5. Although the name "weak* convergence" is the proper terminology in terms of functional analysis, it is often called "weak convergence of measures". This is mainly due to the fact that it is indeed "weaker" than strong convergence and that the functional analytic notion of "weak convergence of measures" uses the dual space of $\mathfrak{M}(\Omega, \mathbb{R}^m)$ which is hard to get a grip on (especially, it can not be identified with a space of functions or measures).

The weak* convergence has some properties that follow directly from general fact from functional analysis.

Theorem 5.2.6. *Any bounded sequence μ_k in $\mathfrak{M}(\Omega, \mathbb{R}^m)$ has a weakly* convergent subsequence, i.e. there exists $\mu \in \mathfrak{M}(\Omega, \mathbb{R}^m)$ such that $\mu_{n_k} \stackrel{*}{\rightharpoonup} \mu$.*

Proof. This follows directly from the sequential Banach-Alaoglu theorem[‡]. Hence, the only thing to show is, that $C_0(\Omega, \mathbb{R}^m)$ is separable, and for this it is enough to consider the case of $C_0(\Omega)$: For compact Ω separability follows from the Weierstrass approximation theorem. Every continuous function of a compact set can be approximated uniformly by polynomials and hence, the polynomials with rational coefficients (adjusted with a continuous cutoff function to be zero on the boundary of Ω) are dense in $C_0(\Omega)$. If Ω is open exhaust Ω with compact sets K_n and take the union of all the countable dense sets in $C_0(K_n)$ which forms a countable dense set in $C_0(\Omega)$. □

[‡]The theorem states that any bounded sequence in the dual space of a separable Banach space has a weakly* convergent subsequence.

Weak* convergence of a measure also tells something how the measure of sets evolves with the measure:

Theorem 5.2.7. *Let μ_k be a sequence in $\mathfrak{M}(\Omega)$ of non-negative measures such that $\mu_k \overset{*}{\rightharpoonup} \mu$. Then:*

1. *if A is open, then*
$$\mu(A) \le \liminf_{k\to\infty} \mu_k(A).$$
2. *if K is compact, then*
$$\mu(K) \ge \limsup_{k\to\infty} \mu_k(K).$$
3. *if A is open, $\overline{A}$ is compact and $\mu(\partial A) = 0$, then*
$$\mu(A) = \lim_{k\to\infty} \mu_k(A).$$

Proof. 1. Let $K \subset A$ be compact. There is a compactly supported continuous function ϕ on Ω such that $0 \le \phi \le 1$, $\phi \equiv 1$ on K and $\phi \equiv 0$ on $\Omega \setminus A$. Then, be weak* convergence of μ_k and since $\mu_k(A) \ge \int \phi \mathrm{d}\mu_k$ we get
$$\mu(K) \le \int \phi \mathrm{d}\mu = \lim_{k\to\infty} \int \phi \mathrm{d}\mu_k \le \liminf_{k\to\infty} \mu_k(A).$$
Since μ is a Radon measure we get
$$\mu(A) = \sup\{\mu(K) \;:\; K \subset A,\ \text{compact}\} \le \liminf_{k\to\infty} \mu_k(A).$$

2. Now let $A \supset K$ be open. We take a similar cutoff function ϕ such that ϕ is continuous, compactly supported, $0 \le \phi \le 1$, $\phi \equiv 1$ on K and $\phi \equiv 0$ in $\Omega \setminus A$. Then, similarly to the above part,
$$\limsup_{k\to\infty} \mu_k(K) \le \lim_{k\to\infty} \int \phi \mathrm{d}\mu_k = \int \phi \mathrm{d}\mu \le \mu(A).$$
Taking the supremum over such A we obtain
$$\limsup_{k\to\infty} \mu_k(K) \le \inf\{\mu(A) \;:\; K \subset A,\ \text{open}\} = \mu(K).$$

3. If $\mu(\partial A) = 0$, we get by the previous two points, that
$$\limsup_{k\to\infty} \mu_k(A) \le \limsup_{k\to\infty} \mu_k(\overline{A}) \le \mu(\overline{A}) = \mu(A) \le \liminf_{k\to\infty} \mu(A).$$

□

Exercise 24. Provide examples to show that the inequalities in Theorem 5.2.7 1. and 2. can be strict, i.e. find sequences (μ_k) of measures and sets A open and K compact such that $\liminf \mu_k(A) > \mu(A)$ and $\limsup \mu_k(K) < \mu(K)$.

For signed measures, the situation is a little bit different. Basically, one additionally needs to ensure weak* convergence of the variation measures as this is not implied by weak* convergence:

Exercise 25. Construct a sequence μ_k such that $\mu_k \overset{*}{\rightharpoonup} \mu$ but $|\mu_k| \not\overset{*}{\rightharpoonup} |\mu|$.

Theorem 5.2.8. *Let (ν_k) be a sequence of signed measures in $\mathfrak{M}(\Omega)$ such that $\nu_k \overset{*}{\rightharpoonup} \nu$ in $\mathfrak{M}(\Omega)$ and $|\nu_k| \overset{*}{\rightharpoonup} \mu$ in $\mathfrak{M}(\Omega)$. If $A \subset \Omega$ is open such that $\overline{A}$ is compact and $\mu(\partial A) = 0$, then it holds for every bounded $f \in C(\Omega)$ that*

$$\int_A f \mathrm{d}\nu_k \to \int_A f \mathrm{d}\nu$$

and in particular

$$\nu_k(A) \to \nu(A).$$

Proof. We take some function $f \in C(\Omega)$. Since μ (the weak* limit of $|\nu_k|$) is Radon and $\mu(\partial A) = 0$ we can find, for any $\epsilon > 0$, an open superset $U \supset \partial A$ such that $\mu(U) \leq \epsilon$.

Observe that $U \cup A \supset \overline{A}$ and hence, there is a continuous cutoff function ϕ such that $0 \leq \phi \leq 1$, $\phi \equiv 1$ on $\overline{A}$ and $\phi \equiv 0$ on $(U \cup A)^{\complement}$.

We get

$$\int_A f \mathrm{d}\nu_k - \int_A f \mathrm{d}\nu = \int \phi f \mathrm{d}\nu_k - \int_{U\setminus A} \phi f \mathrm{d}\nu_k - \int \phi f \mathrm{d}\nu + \int_{U\setminus A} \phi f \mathrm{d}\nu$$

and consequently

$$\begin{aligned} \Big|\int_A f \mathrm{d}\nu_k - \int_A f \mathrm{d}\nu\Big| \leq \Big|\int \phi f \mathrm{d}\nu_k - \int \phi f \mathrm{d}\nu\Big| \\ + \|f\|_\infty(|\nu_k|(\operatorname{supp}\phi \setminus A) + |\nu|(\operatorname{supp}\phi \setminus A)). \end{aligned}$$

The first term on the right goes to zero since $\nu_k \overset{*}{\rightharpoonup} \nu$ and $\phi f \in C_c(\Omega)$ and hence, since $|\nu| \leq \mu$, it follows

$$\limsup_{k\to\infty} \Big|\int_A f \mathrm{d}\nu_k - \int_A f \mathrm{d}\nu\Big| \leq 2\|f\|_\infty \mu(\operatorname{supp}\phi \setminus A) \leq 2\|f\|_\infty \mu(U) \leq 2\|f\|_\infty \epsilon.$$

The second assertion follows by setting $f \equiv 1$. □

5.3 Metrization of weak* convergence

Example 5.2.4 showed that the weak* convergence indeed covers something of the underlying geometry of the set Ω. While the observation that $x_k \to x$ in Ω implies $\delta_{x_k} \to \delta_x$ in $\mathfrak{M}(\Omega)$ is nice, one could hope for a little bit more: The convergence of x_k to x can be *quantified* by the distance $|x_k - x|$. However, weak* convergence (in general and also in the particular case here), does not allow for a quantitative expression for the convergence. We have also seen in

Example 5.2.2, that the variation norm (which does quantify strong convergence) does not work in the case of weak* convergence. However, the particular duality between continuous functions and Radon measure allows for a **metrization** of the notion of weak convergence (on bounded sets) that reflects the geometry. Even a little bit more is true: weak convergence can be quantified by a norm (although, on bounded sets only). To define the respective norm, we first recall the definition of the Lipschitz constant:

Definition 5.3.1 (Lipschitz constant)**.** Let $\Omega \subset \mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}^m$ is called **Lipschitz continuous** if there is a constant $L > 0$ such that for all $x, y \in \Omega$ it holds that

$$|f(x) - f(y)| \leq L|x - y|.$$

The smallest such constant L is called **Lipschitz constant** of f and denoted by $\mathrm{Lip}(f)$.

Remark 5.3.2. In the following we will always work with **compact and convex** sets Ω. Compactness is needed as a technical assumption. The reason that we also assume convexity of the domain Ω is only that one may have a better intuition about Lipschitz functions in these sets. In a convex set we can always connect two points in Ω with a straight line which lies inside of Ω. In consequence, when we have two different points x and y in Ω and we want to find a function f with Lipschitz constant L, such that the difference $f(x) - f(y)$ is as large as possible, we just go from x to y with constant slope L and extend the resulting function to be Lipschitz continuous everywhere and obtain a function such that $f(x) - f(y) = L|x - y|$. If the domain would not be convex, this strategy does not work in general: If we have two points x and y such that the connecting line does not lie in Ω entirely, but we still follow the same approach by going with constant slope L along the shortest connection of the two point we will obtain a function with $|f(x) - f(y)| > L|x - y|$, and hence, does not fulfill the Lipschitz condition.

However, one could develop the following theory also for domains Ω that are not convex.

Definition 5.3.3 (Kantorovich-Rubinstein norm)**.** For Ω compact and convex we define the **Kantorovich-Rubinstein norm** (KR norm) on $\mathfrak{M}(\Omega)$ as

$$\|\mu\|_{\mathrm{KR}} = \sup\{\int f \mathrm{d}\mu \ : \ \|f\|_\infty \leq 1,\ \mathrm{Lip}(f) \leq 1\}$$

Exercise 26. Prove that the Kantorovich-Rubinstein norm has the properties of a norm (positive definite, positively homogeneous, triangle inequality).

As any other norm, the Kantorovich-Rubinstein norm implies a metric:

$$d_{\mathrm{KR}}(\mu, \nu) = \|\mu - \nu\|_{\mathrm{KR}}.$$

Remark 5.3.4 (Other norms and metrics for measures)**.** There are other norms that are similar to the Kantorovich-Rubinstein norm. E.g. if one only cares to measure the distance between non-negative measures of equal mass, i.e. $\mu(\Omega) = \nu(\Omega)$, then one may define

$$d(\mu,\nu) = \sup\{\int f\mathrm{d}(\mu-\nu) \ : \ \mathrm{Lip}(f) \leq 1\}$$

(since $\int \mathrm{d}(\mu-\nu) = 0$, the addition of constants to f does not change the value of the integral).

Another popular metric that is defined for non-negative measures of equal mass is the Prokhorov metric. To define it, we first define the "ϵ-enlargement" of a set A as $A_\epsilon = \{x \ : \ \inf\{|x-y| \ : \ y \in A\} \leq \epsilon\}$. Then, the Prokhorov metric is

$$d_P(\mu,\nu) = \inf\{\epsilon > 0 \ : \ \mu(A) \leq \nu(A_\epsilon) + \epsilon,\ \nu(A) \leq \mu(A_\epsilon) + \epsilon\}.$$

For signed measures there is also the bounded-Lipschitz norm

$$\|\mu\|_{BL} = \sup\{\int f\mathrm{d}\mu \ : \ \|f\|_\infty + \mathrm{Lip}(f) \leq 1\}$$

(which is indeed equivalent to the Kantorovich-Rubinstein norm).

Example 5.3.5. Let $\mu \in \mathfrak{M}(\Omega)$ be non-negative. Then the supremum is attained if f is the constant function 1 and hence,

$$\|\mu\|_{\mathrm{KR}} = \int 1\mathrm{d}\mu = \mu(\Omega) = \|\mu\|_{\mathfrak{M}}.$$

In other words, for non-negative measures the KR norm and the variation norm coincide.

Exercise 27.

1. Show that $\|\mu\|_{\mathrm{KR}} \leq \|\mu\|_{\mathfrak{M}}$.
2. Show that for measures μ with $\int \mathrm{d}\mu = 0$ it holds that

$$\|\mu\|_{\mathrm{KR}} \leq \frac{\mathrm{diam}(\Omega)}{2}\|\mu\|_{\mathfrak{M}}.$$

Example 5.3.6 (KR-distance of Diracs)**.** For two Dirac measure δ_x and δ_y we get the following: If x and y are far enough apart, namely $|x-y| \geq 2$, then there is a function with Lipschitz constant 1 such that $f(x) = 1$ and $f(y) = -1$ and hence the Kantorovich-Rubinstein distance is 2. If x and y are closer together, we can not get the values $f(x)$ and $f(y)$ that far apart, at most we can get $|f(x) - f(y)| \leq |x-y|$ and since this can be achieved, we obtain

$$\|\delta_x - \delta_y\|_{\mathrm{KR}} = \begin{cases} 2, & \text{if } |x-y| \geq 2 \\ |x-y|, & \text{if } |x-y| \leq 2. \end{cases}$$

Especially, we see that the Kantorovich-Rubinstein norm indeed quantifies the weak* convergence of δ_{x_k} to δ_x if $x_k \to x$.

Before we come to the main result in this section, which states that weak* convergence of measures is indeed described by the Kantorovich-Rubinstein norm, we recall the theorem of Arzelà and Ascoli. We give a form for Lipschitz continuous functions.

Theorem 5.3.7 (Arzelà-Ascoli, Lipschitz version)**.** *Let f_n be a sequence of functions such that $\|f_n\|_\infty \le C$ and $\operatorname{Lip}(f_n) \le L$. Then there is a subsequence f_{n_k} and a function f with $\|f\|_\infty \le C$ and $\operatorname{Lip}(f) \le L$ such that f_{n_k} converges uniformly to f.*

Theorem 5.3.8. *Let μ_k be a sequence of signed Radon measures such that $\|\mu_k\|_{\mathfrak{M}}$ is bounded. Then it holds that $\mu_k \overset{*}{\rightharpoonup} 0$ if and only if $\|\mu_k\|_{\mathrm{KR}} \to 0$.*

Proof. Let $\mu_k \overset{*}{\rightharpoonup} 0$, i.e. for all $f \in C_0(\Omega)$ it holds that $\int f \mathrm{d}\mu_k \to 0$.

We claim that for the KR-norm it holds that there exists functions f_k (bounded and Lipschitz) such that

$$\|\mu_k\|_{\mathrm{KR}} = \int f_k \mathrm{d}\mu_k$$

i.e. the supremum in the definition of the norm is actually attained. This can be seen from the Arzelà-Ascoli theorem (Theorem 5.3.7): A maximizing sequence of continuous functions has a uniformly convergent subsequence the limit of which is indeed a maximizer and also fulfills the same Lipschitz constant.

Again by Arzelà-Ascoli, we have a subsequence f_{n_k} which convergences uniformly to some $\bar{f}$. Hence we have

$$\begin{aligned}\|\mu_{n_k}\|_{\mathrm{KR}} = \int f_{n_k} \mathrm{d}\mu_{n_k} &= \int (f_{n_k} - \bar{f}) \mathrm{d}\mu_{n_k} + \int \bar{f} \mathrm{d}\mu_{n_k}\\ &\le \|f_{n_k} - \bar{f}\|_\infty \|\mu_{n_k}\|_{\mathfrak{M}} + \int \bar{f} \mathrm{d}\mu_{n_k}\end{aligned}$$

and the right hand side converges to zero. The same reasoning can be applied to all subsequences of μ_k and hence, the subsequence-subsequence lemma shows that $\|\mu_k\|_{\mathrm{KR}} \to 0$ for the whole sequence.

Now assume that $\|\mu_k\|_{\mathrm{KR}} \to 0$ and let $g \in C_0(\Omega)$. We need to show that $\int g \mathrm{d}\mu_k \to 0$.

For $C_n > 0$ we define a continuous function by

$$f_n = \operatorname{argmin}\{\|h - g\|_\infty \;:\; \operatorname{Lip}(h) \le C_n\}.$$

and obviously, f_n is Lipschitz continuous with constant bounded by C_n. Also it holds that $f_n \to g$ uniformly, if $C_n \to \infty$.

We now use

$$|\int g \mathrm{d}\mu_k| \le |\int (g - f_n) \mathrm{d}\mu_k| + |\int f_n \mathrm{d}\mu_k|.$$

The first term converges to zero if $f_n \to g$ uniformly since $\|\mu_k\|_{\mathfrak{M}}$ is bounded. For the second term, observe that $\tilde{f}_n = \frac{1}{\max(C_n, \|f_n\|_\infty)} f$ is bounded by one and

Lipschitz constant smaller than one and hence,

$$|\int f_n \mathrm{d}\mu_k| = \max(C_n, \|f_n\|_\infty)|\int \tilde{f}_n \mathrm{d}\mu_k| \leq \max(C_n, \|f_n\|_\infty)\|\mu_k\|_{\mathrm{KR}}.$$

To finish the proof we need to choose n in dependence on k such that $n(k) \to \infty$ and $\max(C_{n(k)}, \|f_{n(k)}\|_\infty)\|\mu_k\|_{\mathrm{KR}} \to 0$. Since $\|f_{n(k)}\|_\infty \leq \|f_{n(k)} - g\|_\infty + \|g\|_\infty$, we see that $\|f_{n(k)}\|_\infty$ is bounded when $n(k) \to \infty$ and consequently, the choice $C_{n(k)} = 1/\sqrt{\|\mu_k\|_{\mathrm{KR}}}$ does the trick. □

6 Hausdorff measures

As we are already up to our neck into geometric properties of measures we confidently move on and aim to introduce the concept of a lower-dimensional measure, i.e. a measure on $\mathbb{R}^n$ (or subsets thereof) that can accurately measure what one would perceive as "m-dimensional size". In principle there are different possibilities to achieve this goal. We follow the approach that leads to Hausdorff measures which is due to Hausdorff (Who'd have thunk!) and, again, Carathéodory. Indeed, the construction due to Carathéodory is pretty general and we will start in this generality.

6.1 General construction

We are going to work in $\Omega \subset \mathbb{R}^n$ (however, in this section, a metric space Ω would also work) and start off with an abstract definition that resembles the construction of Lebesgue's measure a bit:

Definition 6.1.1. Let $\Omega \subset \mathbb{R}^n$, $\mathcal{F}$ be a family of subsets of Ω and ζ a non-negative function of $\mathcal{F}$. Assume that

1. For every $\delta > 0$ there are $E_i \in \mathcal{F}$, $i \in \mathbb{N}$ such that $\operatorname{diam}(E_i) \leq \delta$ and $\Omega \subset \bigcup_{i=1}^{\infty} E_i$.

2. For every $\delta > 0$ there is $E \subset \mathcal{F}$ such that $\zeta(E) \leq \delta$ and $\operatorname{diam}(E) \leq \delta$.

Then we define for $0 < \delta \leq \infty$ and $A \subset \Omega$

$$\psi_\delta(A) = \inf\{\sum_{i\in\mathbb{N}} \zeta(E_i) \ : \ A \subset \bigcup_{i\in\mathbb{N}} E_i,\ \operatorname{diam}(E_i) \leq \delta,\ E_i \in \mathcal{F}\}.$$

Conceptually, the family $\mathcal{F}$ contains some model sets for which we know how to measure them (these were the intervals in the case of the Lebesgue measure). The function ζ now gives a "measure" to these model sets (has been the "length" of the intervals for Lebesgue's measure).

One could say that assumption 1. is only needed to ensure that we there exists one such covering that is used in the infimum. Assumption 2. guarantees that $\psi_\delta(\emptyset) = 0$.

Note that the definition is in principle very similar to the definition of Lebesgue measure: Instead of intervals (or boxes) we now have a general family of sets and instead of the side length of the boxes we now use another function ζ.

Similarly to the case of Lebesgue's measure one can prove that the ψ_δ's are measures:

Exercise 28. Show that the ψ_δ from Definition 6.1.1 are (outer) measures, i.e. they are countably subadditive.

However, the measures ψ_δ do not obey good additivity properties and are even not Borel measures even, if the sets $\mathcal{F}$ are nice. We will see a counterexample below in the special case of Hausdorff measure.

We observe that $\epsilon < \delta$ implies $\psi_\delta(A) \leq \psi_\epsilon(A)$. Hence, we may define

$$\psi(A) = \lim_{\delta \searrow 0} \psi_\delta(A) = \sup_{\delta > 0} \psi_\delta(A).$$

It turns out that the function ψ has better properties than the function ψ_δ.

Theorem 6.1.2. *1. ψ is a Borel measure.*

2. If all members of $\mathcal{F}$ are Borel sets, ψ is a regular Borel measure.

Proof. 1. To show that ψ is an (outer) measure observe that, because ψ_δ is a measure and, by definition, $\psi \geq \psi_\delta$, for any A, covered by A_i's

$$\psi(A) = \lim_{\delta \searrow 0} \psi_\delta(A) \leq \lim_{\delta \searrow 0} \sum_i \psi_\delta(A_i) \leq \sum_i \psi(A_i).$$

To show that ψ is a Borel measure, we use the criterion, that a measure is Borel if it is additive for "separated sets" (cf. the proof of the Riesz representation theorem 5.1.4, step 2). Let $A, B \subset \Omega$ with $\operatorname{dist}(A, B) > 0$ and choose $\delta > 0$ such that $\delta < \operatorname{dist}(A, B)/2$. If the sets $E_i \in \mathcal{F}$ cover $A \cup B$ and also satisfy $\operatorname{diam}(E_i) < \delta$, the none of them intersects both A and B. In other words, the covering separates and we have

$$\sum_i \zeta(E_i) \geq \sum_{A \cap E_i \neq \emptyset} \zeta(E_i) + \sum_{B \cap E_i \neq \emptyset} \zeta(E_i) \geq \psi_\delta(A) + \psi_\delta(B).$$

Taking the infimum over all such covering, we obtain $\psi_\delta(A \cup B) \geq \psi_\delta(A) + \psi_\delta(B)$. The opposite inequality holds by subadditivity of ψ_δ and hence, $\psi_\delta(A \cup B) = \psi_\delta(A) + \psi_\delta(B)$. Letting $\delta \to 0$ we get $\psi(A \cup B) = \psi(A) + \psi(B)$ and hence, ψ is a Borel measure.

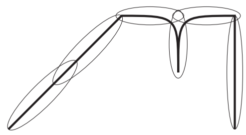
2. Let $A \subset \Omega$ and for every $j = 1, 2, \ldots$ choose sets $E_i^j \in \mathcal{F}$ such that $\operatorname{diam}(E_i^j) \leq 1/j$, $A \subset \bigcup_i E_i^j$, and $\sum_i \zeta(E_i^j) \leq \psi_{1/j}(A) + 1/j$. The set $B = \bigcap_j \bigcup_i E_i^j$ is, by construction, a Borel set with $A \subset B$ and $\psi(A) = \psi(B)$ and hence, ψ is Borel regular.

□

6.2 Hausdorff measures

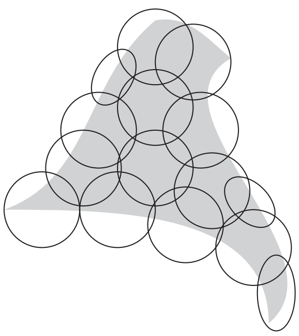
We specialize the general construction from the previous section, i.e. we choose a particular class of sets $\mathcal{F}$ and a particular function ζ. Our goal is, to measure lower dimensional structures and the question is, how should one choose $\mathcal{F}$ and ζ to achieve this. As a motivation, we consider the situation of a one-dimensional structure in two dimensions:

We first leave aside the question of the class of sets $\mathcal{F}$ (for our motivation, we only consider sets that we can draw, hence, they will be not too complicated anyway) and focus on the question of the function ζ. Of course we want that ζ is "smaller for smaller sets" since we would like to catch all details of the sets to be measured and hence, smaller sets should be preferred. Consider the following set A (which is a path) and which is covered by smaller sets E_i:



How should we measure the "size" $\zeta(E_i)$ such that the sum $\sum_i \zeta(E_i)$ will approach the length of the path when we make the sets smaller and smaller? Obviously, the volume, e.g. in the sense of the two dimensional Lebesgue measure, i.e. $\zeta(E) = \lambda^2(E)$, is not a good idea since we could make the ellipses, that we used, thinner and thinner and would obtain covers with an arbitrarily small sum. Consequently the resulting measure would be zero. We need something which relates to the length. An idea is the diameter: $\zeta(E) = \operatorname{diam}(E)$. This works in this case: The ellipses would align with the graph when we make them smaller and smaller and their diameter equals the length of their principle axis. Since the diameter is defined for any set, one may try to choose as $\mathcal{F}$ the class of all sets. Note that the same choice of $\zeta = \operatorname{diam}$ and $\mathcal{F}$ also works if the path would lie in three dimensional space. Maybe your higher dimensional intuition also allows you to imagine that it will also work in higher dimensions.

Let's see, what could work if we have a two-dimensional surface in three-dimensional space (or higher dimensions). Due to drawing limitation, let's draw the situation of a two dimensional subset in two-dimensional space. Consider the following situation:



How should we choose ζ here? Of course the two-dimensional Lebesgue measure would work. But imagine that all the covering circles and ellipses would be three dimensional balls and ellipses. Then the three-dimensional Lebesgue measure would not work for a similar reason that the 2D Lebesgue measure did not work for the graph above (we could make the balls and ellipses as flat as we like). Also the diameter itself would not work: The total sum of diameters would be larger if we cover with smaller balls which is not as intended. However, the *squared* diameter goes in the right direction: Although we would not get the same as the intuitive area, we would get something that is in the right order (and off by a factor).

It turns out that the right scaling of the diameter raised to the right power really does the trick here. The precise definition is as follows:

Definition 6.2.1 (Hausdorff measure)**.** Let $m, n \in \mathbb{N}$ and denote by V_m be the volume of the m-dimensional unit ball. Then the m-dimensional Hausdorff measure in $\mathbb{R}^n$ is obtained by the general construction in Definition 6.1.1 by choosing $\mathcal{F}$ to contain all sets in $\mathbb{R}^n$ and for $E \subset \mathbb{R}^n$ define

$$\zeta(E) = V_m \Big(\frac{\operatorname{diam}(E)}{2}\Big)^m.$$

More explicitly, denote for $\delta > 0$ and $A \subset \mathbb{R}^n$

$$\mathfrak{H}^m_\delta(A) = \inf\{V_m \sum_i \Big(\frac{\operatorname{diam}(E)}{2}\Big)^m \;:\; A \subset \bigcup_i E_i,\ \operatorname{diam}(E_i) \le \delta\}$$

and the m**-dimensional Hausdorff measure** is

$$\mathfrak{H}^m(A) = \lim_{\delta \searrow 0} \mathfrak{H}^m_\delta(A).$$

Example 6.2.2. Let's look at $m = 0$. A measure that measures the "zero-dimensional" content should be counting the "zero-dimensional objects" which are points. The zero-dimensional unit-ball is a point and hence, it makes sense that $V_0 = 1$. Consequently, if we agree that $0^0 = 0$ here,

$$\zeta(E) = \begin{cases} 0, & \text{if } E = \emptyset \\ 1, & \text{else.} \end{cases}$$

We calculate the number $\mathfrak{H}^0_\delta(A)$ as the minimal number of sets of diameter smaller than δ that is needed to cover A. We conclude that $\mathfrak{H}^0 = \#$, i.e. the zero-dimensional Hausdorff measure is the counting measure.

In the one-dimensional case $m = 1$, one obviously has $V_1 = 2$ (the length of the interval $[-1, 1]$. It turns out that the $\mathfrak{H}^1$ equals the usual length for rectifiable curves. For "higher-dimensional objects", $\mathfrak{H}^1$ delivers ∞ while "zero-dimensional objects" get the one-dimensional Hausdorff measure zero.

Exercise 29. 1. Recall that a **rectifiable curve** is a mapping $\gamma : [a, b] \to \mathbb{R}^n$ such that its "length" is finite, i.e.

$$\operatorname{length}(\gamma) = \sup\{\sum_{i=1}^{n} |\gamma(t_i) - \gamma(t_{i-1})| \ : \ a = t_0 < t_1 \cdots < t_n = b\} < \infty.$$

Show that

$$\mathfrak{H}^1(\gamma([a, b])) = \operatorname{length}(\gamma).$$

In other words: The one-dimensional Hausdorff measures measure the length of rectifiable curves.

2. Let x_i be countably many points in $\mathbb{R}^n$ and denote by $C = [0, 1]^n$ the n-dimensional unit cube. Show that $\mathfrak{H}^1(\bigcup_{i\in\mathbb{N}}\{x_i\}) = 0$ and $\mathfrak{H}^1(C) = \infty$.

Let's get a bit more concrete about the volume V_m of the m-dimensional unit ball in the general case (although the exact value of this number does not play a role here). It has an explicit formula in terms of the Γ-function:

$$V_m = \frac{\pi^{m/2}}{\Gamma(\frac{m}{2}+1)}.^{*}$$

However, the Γ-function makes sense for all $s \geq 0$ and we can also speak of the constant

$$V_s = \frac{\pi^{s/2}}{\Gamma(\frac{s}{2}+1)}$$

for $s \geq 0$ (although there is no such thing as s-dimensional unit ball).†

Hence, there is no principal obstruction to define the s-dimensional Hausdorff measure for $s \geq 0$ and we do that:

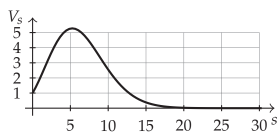
Definition 6.2.3 (Fractional Hausdorff measure)**.** Let $s > 0$ $n = 1, 2, \ldots$. Then the s-dimensional **Hausdorff measure** is defined similar to the m-dimensional one by replacing m by s and it is denoted by $\mathfrak{H}^s$.

Exercise 30. Let U be an open ball in $\mathbb{R}^n$, $n \geq 2$ with $\operatorname{diam}(U) = \delta$ and let $0 \leq s \leq 1$. Show that $\mathfrak{H}^s_\delta(U) = \mathfrak{H}^s_\delta(\overline{U}) = \mathfrak{H}^s_\delta(\partial U)$. Conclude that $\mathfrak{H}^s_\delta$ is not a Borel measure.

---

*Some explicit values are $V_0 = 1$, $V_1 = 2$, $V_2 = \pi$, $V_3 = \frac{4}{3}\pi$, $V_4 = \frac{\pi^2}{2}$, more can be found at the respectve Wikipedia page.

†Curiously, this volume V_s has a maximum of about 5.28 for $s \approx 5.26$ and tends to zero for $s \to \infty$; here's a plot:



Remark 6.2.4 (Normalization and relation to the Lebesgue measure)**.** Some authors (e.g.[Mat95]) do not include the scaling by the volume of the unit ball in the definition of Hausdorff measure. Then the definition looks a bit cleaner:

$$\tilde{\mathfrak{H}}^s_\delta(A) = \inf\{\sum_i \operatorname{diam}(E_i)^s \ : \ A \subset \bigcup_i E_i,\ \operatorname{diam}(E_i) \leq \delta\}$$

and $\tilde{\mathfrak{H}}^s(A) = \lim_{\delta \searrow 0} \tilde{\mathfrak{H}}^s_\delta(A)$. If one is only interested to decide whether the s-dimensional measure is zero, non-zero but finite, or infinite, this simpler form is of course valid. However, for the measure $\tilde{\mathfrak{H}}^s$ one does not have that it reproduces what should be the length, area or volume . Especially, one has for $s = n$ in $\mathbb{R}^n$ that $\tilde{\mathfrak{H}}^n = \frac{2^n}{V_n}\lambda^n$ while with the scaling from Definition 6.2.1 one has $\mathfrak{H}^n = \lambda^n$. The proof of this equality, as natural and desirable it may seem, is not straightforward. We will not prove it yet (maybe later); however, if you are keen, have a look at [EG92, §2.2] or [Fed69, §2.10].

The Hausdorff measures are translation invariant and scale according to their intended dimension:

Lemma 6.2.5. *Let $A \subset \mathbb{R}^n$, $a \in \mathbb{R}^n$ and $0 < s, t < \infty$. Then it holds that $\mathfrak{H}^s(A + a) = \mathfrak{H}^s(A)$ and $\mathfrak{H}^s(tA) = t^s\mathfrak{H}^s(A)$.*

Proof. The results follow directly from the fact that if some E_i's cover A, then the sets $E_i + a$ cover $A + a$ and the sets tE_i cover tA. □

That we allowed us to use all sets to form the model family $\mathcal{F}$ was not particularly important, as the next result shows:

Lemma 6.2.6. *Let $0 \leq s$ and ζ as in Definition 6.2.3. Then the definition of the s-dimensional Hausdorff measure does not change if we set*

1. *$\mathcal{F} = \{U \subset \mathbb{R}^n \ : \ U$ is open$\}$ or*
2. *$\mathcal{F} = \{A \subset \mathbb{R}^n \ : \ A$ is closed$\}$ or*
3. *$\mathcal{F} = \{C \subset \mathbb{R}^n \ : \ C$ is convex$\}$.*

Proof. Let us denote the resulting measures by $\psi^{\mathcal{F}}$ (without capturing the properties of the respective family $\mathcal{F}$).

By the very definition of Hausdorff measure (or also in the general construction) it is clear that the measure gets larger if we restrict the class of sets, i.e. it holds $\psi^{\mathcal{F}} \geq \mathfrak{H}^s$ in all three cases.

To show the reverse direction note that for an arbitrary set $E \subset \mathbb{R}^n$ it holds that the diameter of any set stays the same for the closure and the convex hull[‡]. Hence, we can turn any covering with arbitrary sets into one with closed or convex sets with the same sum inside the infimum. Since the $\psi^{\mathcal{F}}$ is defined

[‡]The convex hull is the smallest convex set that contains the set.

by infima over all such covering, we get $\psi^{\mathcal{F}} \leq \mathfrak{H}^s$ for the case of closed and convex sets. For the case of open sets, note that for any set E and $\epsilon > 0$, the set $E^\epsilon = \{x : \operatorname{dist}(x, E) < \epsilon\}$ is open and fulfills $\operatorname{diam}(E^\epsilon) \leq \operatorname{diam}(E) + 2\epsilon$. Hence, we can turn any arbitrary covering with E_i's into an open one while increasing the sum less than any ϵ (use the sets $E_i^{\epsilon/2^i}$). □

From the above and Theorem 6.1.2 we conclude:

Corollary 6.2.7. *The measure $\mathfrak{H}^s$ are regular Borel measures.*

Remark 6.2.8. Note that $\mathfrak{H}^s$ is in general not a Radon measure since it does not assign finite values to compact sets. However, if one has a $\mathfrak{H}^s$-measurable set A in $\mathbb{R}^n$ with $\mathfrak{H}^s(A) < \infty$, then $\mathfrak{H}^s \llcorner A$ is a Radon measure (by using Lemma 2.3.8 and Corollary 2.3.10). With this construction we have another pretty large class of Radon measures on $\mathbb{R}^n$ with some kind of intuitive understanding.

The Hausdorff measures behave controlled if one maps the sets by Hölder continuous maps:

Theorem 6.2.9. *Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$ be Hölder continuous with exponent $\alpha \in]0, 1]$ and constant C.*[§] *Then it holds for any $s \geq 0$ that*

$$\mathfrak{H}^{s/\alpha}(f(A)) \leq C^{s/\alpha}\mathfrak{H}^s(A).$$

Proof. We cover A by sets E_i with $\operatorname{diam} E_i \leq \delta$. Then it holds

$$\operatorname{diam}(f(A \cap E_i)) \leq \operatorname{diam}(A \cap E_i)^\alpha \leq C\operatorname{diam}(E_i)^\alpha.$$

We conclude that the sets $F_i = f(A \cap E_i)$ covers $f(A)$ and fulfill $\operatorname{diam}(F_i) \leq \epsilon = C\delta^\alpha$ and also

$$\sum_i \operatorname{diam}(f(A \cap E_i))^{s/\alpha} \leq C^{s/\alpha} \sum_i \operatorname{diam}(E_i)^s.$$

It follows

$$\mathfrak{H}_\epsilon^{s/\alpha}(f(A)) \leq C^{s/\alpha}\mathfrak{H}_\delta^s(A).$$

Letting $\delta \to 0$ implies $\epsilon \to 0$ and the claim follows. □

In the following we will be keen to know if some set has zero or non-zero Hausdorff measure for some s. It turns out that one can infer this by simply considering $\mathfrak{H}_\delta^s$ for a single δ (even $\delta = 0$) and even without considering a measure at all:

Lemma 6.2.10. *Let $A \subset \mathbb{R}^n$, $0 \leq s < \infty$ and $0 < \delta \leq \infty$. Then the following conditions are equivalent:*

1. $\mathfrak{H}^s(A) = 0$.

[§] That means that there exists C such that for all $x, y \in A$ it holds that $|f(x) - f(y)| \leq C|x - y|^\alpha$.

2. $\mathfrak{H}^s_\delta(A) = 0$.
3. *For all* $\epsilon > 0$ *there exists sets* E_i, $i \in \mathbb{N}$, *such that*

$$A \subset \bigcup_i E_i, \qquad \text{and} \qquad \sum_i \operatorname{diam}(E_i)^s < \epsilon.$$

Proof. 1. $\implies$ 2. $\implies$ 3. is obvious from the definitions.

Condition 3. clearly implies that $\mathfrak{H}^s_\delta(A)$ is zero for all $\delta > 0$ which implies 1. □

As the next theorem shows, the Hausdorff measure has some monotonicity in the exponent s and even more: If it is finite for some value of s than it is zero for larger ones and if it is non-zero for some exponent it is infinity for at smaller ones.

Theorem 6.2.11. *For any* $0 \leq s < t < \infty$ *and any* $A \subset \mathbb{R}^n$ *it holds that*

1. $\mathfrak{H}^s(A) < \infty$ *implies* $\mathfrak{H}^t(A) = 0$,
2. $\mathfrak{H}^t(A) > 0$ *implies* $\mathfrak{H}^s(A) = \infty$.

Proof.
1. Let $\delta > 0$ and $A \subset \bigcup_i E_i$ with $\operatorname{diam}(E_i) \leq \delta$ and $V_s/2^s \sum_i \operatorname{diam}(E_i)^s \leq \mathfrak{H}^s_\delta(A) + 1$. Then $\operatorname{diam}(E_i)^t = \delta^{t-s}\operatorname{diam}(E_i)^s$ and hence,
$$\mathfrak{H}^t_\delta(A) \leq \tfrac{V_t}{2^t} \sum_i \operatorname{diam}(E_i)^t \leq \tfrac{V_t}{2^t}\delta^{t-s} \sum_i \operatorname{diam}(E_i)^s \leq \tfrac{V_t}{2^t}\tfrac{2^s}{V_s}\delta^{t-s}(\mathfrak{H}^s_\delta(A) + 1). \tag{*}$$
Letting $\delta \searrow 0$ proves the assertion.
2. Diving (*) by δ^{t-s} and letting $\delta \searrow 0$ proves the assertion.

□

As a corollary we get that there are no sets of positive s-dimensional Hausdorff measure if s is larger than the dimension of the euclidean space.

Corollary 6.2.12. *For* $s > n$ *and any* $A \subset \mathbb{R}^n$ *it holds that* $\mathfrak{H}^s(A) = 0$.

Proof. It holds that $\mathbb{R}^n = \bigcup_k B_k(0)$ and $\lambda^n(B_k(0)) < \infty$. From Remark 6.2.4 we remember that $\mathfrak{H}^n = \lambda^n$ and hence $\mathfrak{H}^n(B_k(0)) < \infty$ and by Theorem 6.2.11 $\mathfrak{H}^s(B_k(0)) = 0$. Consequently $\mathfrak{H}^s(\mathbb{R}^n) = 0$. □

6.3 Hausdorff dimension

Theorem 6.2.11 tell us, that the Hausdorff measure $\mathfrak{H}^s$ of a set has a pretty simple behavior as a function of s: There is at most one value for s for which it is nonzero but finite. For s large enough it is zero, and if it becomes nonzero for some value of s it will be infinity for any smaller value of s. Hence, for every set A there is some value s at which the Hausdorff measure jumps from zero to infinity (or drops the other way—depending on whether you decrease or increase s). This point tells us, “in which dimension the set lives”.

Definition 6.3.1 (Hausdorff dimension)**.** The **Hausdorff dimension** of a set A is equivalently expressed as

$$\begin{aligned}\dim A &= \sup\{s \;:\; \mathfrak{H}^s(A) > 0\} = \sup\{s \;:\; \mathfrak{H}^s(A) = \infty\}\\ &= \inf\{t \;:\; \mathfrak{H}^t(A) < \infty\} = \inf\{t \;:\; \mathfrak{H}^t(A) = 0\}.\end{aligned}$$

In the case some set is empty we use the obvious possibility, e.g. if $\mathfrak{H}^s(A) = 0$ for all s, we set $\dim A$.

Another way to phrase the definition is:

> The Hausdorff dimension of a set A is the unique number $\dim A$ for which $s < \dim A$ implies $\mathfrak{H}^s(A) = \infty$ and $t > \dim A$ implies $\mathfrak{H}^t(A) = 0$.

Note that at the border-case $s = \dim A$, all three cases are possible: It may be that $\mathfrak{H}^s(A)$ is zero, infinite or in between. But if there is some s such that $0 < \mathfrak{H}^s(A) < \infty$, then $s = \dim A$.

Example 6.3.2. 1. Consider a set A consisting of finitel many points $x_1, \ldots, x_n$. We have seen in Example 6.2.2 that $\mathfrak{H}^0 = \#$ and hence $\mathfrak{H}^0(A) = n$. Moreover, for any $\epsilon > 0$ we can cover A with the sets $E_j = B_{\epsilon^{1/s}/n^{1/s}}(x_j)$ and obtain $\mathfrak{H}^s(A) \leq V_s \sum i = 1^n \frac{\epsilon}{n}$ and conclude that $\mathfrak{H}^s(A) \leq V_s\epsilon$ for $s > 0$. Thus, $\dim A = 0$.

2. We already know that $\mathfrak{H}^s(\mathbb{R}^n) = 0$ for $s > n$. Moreover, $\mathfrak{H}^n(\mathbb{R}^n) = \infty$, and hence $\dim \mathbb{R}^n = n$.

Some simple fact about the Hausdorff dimension are:

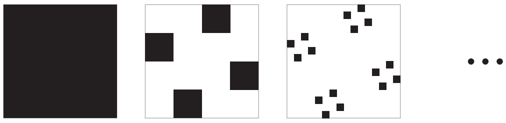
Lemma 6.3.3. *If $A \subset B$, then $\dim A \leq \dim B$ and for $A_i \subset \Omega$ it holds that $\dim \bigcup_i A_i = \sup_i \dim A_i$.*

Proof. The first claim follows since $\mathfrak{H}^s$ is monotone.

Since $\bigcup_i A_i \supset A_i$ we conclude that $\dim \bigcup A_i \geq \sup_i \dim A_i$. However, if $s > \dim A_i$ for all i, then, $\mathfrak{H}^s(A_i) = 0$ for all i. Hence, $\mathfrak{H}^s(\bigcup_i A_i) \leq \sum_i \mathfrak{H}^s(A_i) = 0$ for these s which shows the other inequality. □

One shall not think that a set $A \subset \mathbb{R}^n$ with integer Hausdorff dimension k has to look anything like a k-dimensional surface, as the next example shows:

Example 6.3.4 (Cantor dust in two dimensions, with Hausdorff-dimension 1)**.** We construct a set recursively and start with $A_1 = [0,1]^2 \subset \mathbb{R}^2$. The set A^2 arises by tiling the square A^2 into 16 subsquares of equal size and only keep four of these squares, namely the third on each side. To obtain A^3 we repeat this procedure with every of the four squares. We proceed like this procedure to obtain the next A^k's. The A^k's get thin very quickly as can be seen in a picture:



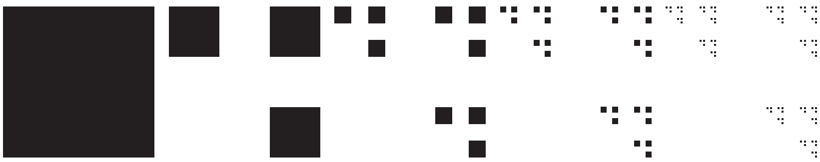
A_1 A_2 A_3

Now we estimate the 1-dimensional Hausdorff measure of the limiting set $A = \bigcap A_i$: For an upper bound, observe that A^k consists of 4^{k-1} squares with side-length $4^{-(k-1)}$, i.e. diameter equal so $\sqrt{2} \cdot 4^{-(k-1)}$. These squares form a cover for A^k and hence we get with $\delta = \sqrt{2} \cdot 4^{-(k-1)}$ that $\mathfrak{H}^1_\delta(A^k) \leq 4^{k-1} \cdot 4^{-(k-1)} \cdot \sqrt{2} = \sqrt{2}$. Sending $k \to \infty$ gives $\delta \to 0$ and we conclude that $\mathfrak{H}^1(A) \leq \sqrt{2}$.

For a lower bound we employ Theorem 6.2.9 and consider the projection P that projects $\mathbb{R}^2$ orthogonally onto $\mathbb{R}$. Clearly, P is Lipschitz continuous with constant $C = 1$, i.e. Hölder-1 continuous with the same constant. Hence, $1 = \mathfrak{H}^1([0,1]) = \mathfrak{H}^1(P(A)) \leq \mathfrak{H}^1(A)$.

We conclude that the $\mathfrak{H}^1(A)$ is somewhere between 1 and $\sqrt{2}$ but certainly not zero and finite. Hence $\dim(A) = 1$.

Exercise 31. Consider the variant of the Cantor dust:



Here, the square is divided into nine equal subsquares, six of which are deleted and only the upper left, upper right and lower right one are kept.

Show that this variant also has Hausdorff dimension 1 and also show that its one-dimensional Hausdorff measure is equal to $\sqrt{2}$.

Let's see if we can find sets with truly fractional Hausdorff dimension.

Example 6.3.5 (Cantor sets in $\mathbb{R}$)**.** Now we look at Cantor sets in one dimension and show that one can indeed construct subsets of $\mathbb{R}$ with Hausdorff dimension s for any $s \in [0,1]$.

Let $0 < \lambda < 1/2$ and denote $I_{0,1} = [0,1]$, $I_{1,1} = [0,\lambda]$ and $I_{1,2} = [1-\lambda, 1]$, i.e. we delete an open interval of length $1-2\lambda$ from the middle of $I_{0,1}$. We continue this process of deleting an interval in the middle of each already given interval and hence, remove $1-2\lambda$ of its length as follows: For given intervals $I_{k-1,1}, \dots, I_{k-1,2^{k-1}}$ we define $I_{k,1}, \dots, I_{k,2^k}$ by deleting in the middle of each $I_{k-1,j}$ an interval of length $(1-2\lambda)\operatorname{diam}(I_{k-1,j}) = (1-2\lambda)\lambda^{k-1}$. Especially, we have $\operatorname{diam}(I_{k,j}) = \lambda^k$.

The limit set of this construction is

$$C_\lambda = \bigcap_{k=0}^{\infty} \bigcup_{j=1}^{2^k} I_{k,j}.$$

For larger λ, the sets are larger, while for smaller λ, the sets are smaller:

$\lambda = 1/3$ $\lambda = 1/4$ $\lambda = 1/5$

An upper bound on the Hausdorff measure can be found by observing that the intervals in the kth level itself give a λ^k-fine cover of C_λ, i.e. for any k: $C_\lambda \subset \bigcup_j I_{k,j}$ and hence,

$$\mathfrak{H}^s_{\lambda^k}(C_\lambda) \leq \frac{V_s}{2^s}\sum_{j=1}^{2^k}\operatorname{diam}(I_{k,j})^s = \frac{V_s}{2^s}2^k\lambda^{ks} = \frac{V_s}{2^s}(2\lambda^s)^k.$$

Letting $k \to \infty$ gives $\lambda^k \to 0$ and the upper bound stays finite if $2\lambda^s \leq 1$, i.e. the smallest s such that this happens is $s_\lambda = -\frac{\log(2)}{\log(\lambda)}$ (note that $s_\lambda \geq 0$ since $\lambda < 1/2 < 1$, i.e. $\log(\lambda) < 0$). For this s we have $2\lambda^s = 1$ and hence

$$\mathfrak{H}^{s_\lambda}(C_\lambda) = \lim_{k\to\infty}\mathfrak{H}^{s_\lambda}_{\lambda^k}(C_\lambda) \leq \frac{V_s}{2^s}.$$

Now we aim to find a lower bound for $\mathfrak{H}^{s_\lambda}(C_\lambda)$ which is (as usual) a harder task. Let U_i be an arbitrary cover of C_λ with open intervals. Since C_λ is compact, finitely many U_i's are enough, say the $U_1, \dots, U_N$. Since the interior of C_λ is empty, we may enlarge the intervals U_i slightly to ensure that their endpoints do not lie in C_λ. Then there is $\delta > 0$ such that the distance of all these endpoints to C_λ is larger than δ. We choose k so large that $\delta > \lambda^k$ and since $\operatorname{diam}(I_{k,j}) = \lambda^k$ it follows that each $I_{k,j}$ is contained in some U_i.

Now we aim to show that for any open interval U and any fixed l it holds that

$$\sum_{I_{l,i}\subset U}\operatorname{diam}(I_{l,i})^{s_\lambda} \leq 5\operatorname{diam}(U)^{s_\lambda}. \tag{*}$$

If we had shown (*) we could conclude that

$$5\sum_j \operatorname{diam}(U_j)^{s_\lambda} \geq \sum_j\sum_{I_{k,i}\subset U_j}\operatorname{diam}(I_{k,i})^{s_\lambda} \geq \sum_{i=1}^{2^k}\operatorname{diam}(I_{k,i})^{s_\lambda} = 2^k\lambda^{ks_\lambda} = 1$$

(using that each $I_{k,i}$ is contained in some U_j). From this we could conclude that for any open cover U_j of C_λ we have

$$\sum_j \operatorname{diam}(U_j)^{s_\lambda} \geq \tfrac{1}{5}$$

which shows that $\mathfrak{H}^s(C_\lambda) \geq \frac{V_{s_\lambda}}{2^{s_\lambda}}\,\frac{1}{5} > 0$.

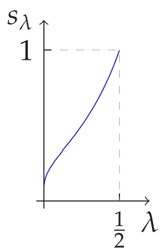
Now we turn to the proof of (*): Assume that some $I_{l,i}$ lies in U (otherwise the inequality is trivial) and let n be the smallest integer such that some $I_{n,i}$ lies in U. Then it holds $n \le l$. Denote by $I_{n,j_1}, \ldots, I_{n,j_p}$ all intervals on level n that intersect U. These have to be less than five (i.e. $p \le 5$) since otherwise there would be some interval $I_{n-1,i}$ on the level $n-1$ which lies in U (remember that U is an interval) and this would contradict the minimality of n. Hence,

$$5\mathrm{diam}(U)^{s_\lambda} \ge \sum_{m=1}^{p} \mathrm{diam}(I_{n,j_m})^{s_\lambda} \ge \sum_{m=1}^{p} \sum_{I_{l,i} \subset I_{n,j_m}} \mathrm{diam}(I_{l,i})^{s_\lambda} \ge \sum_{I_{l,i} \subset U} \mathrm{diam}(I_{l,i})^{s_\lambda}$$

and the claim is proved.

In conclusion:

> The Cantor set C_λ has Hausdorff dimension $\dim C_\lambda = s_\lambda = -\frac{\log(2)}{\log(\lambda)}$. This value varies from zero to one as λ increases from zero to $\frac{1}{2}$:



The famous "middle-third Cantor set" with $\lambda = \frac{1}{3}$ has dimension $\log(2)/\log(3) \approx 0.631$.

6.4 Other lower dimensional measures

In the setting of the general construction of Section 6.1, one could think of different families $\mathcal{F}$ and also different functions ζ. The Hausdorff measure uses all sets for $\mathcal{F}$ and the scaled diameter (raised to an appropriate power) as ζ.

6.4.1 The spherical measure

If we use ζ as for the Hausdorff measure, i.e. $\zeta(E) = c_s \mathrm{diam}(E)^s$ with some constant c_s, but restrict $\mathcal{F}$ to consist of all closed balls, we obtain the so-called s-dimensional **spherical measure** and is denoted by $\mathfrak{S}^s$.

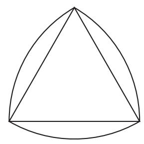
This measure is not very different from the s-dimensional Hausdorff measure as the following lemma shows:

Lemma 6.4.1. *It holds that* $\mathfrak{H}^s \le \mathfrak{S}^s \le 2^s \mathfrak{H}^s$.

Proof. The first inequality is clear since the spherical measure has a more restricted class of sets to use for the infimum in Definition 6.1.1.

For the second inequality notice that if E has the diameter r and x is one of the two points realizing the diameter of E, it follows that $E \subset B_r(x)$ i.e. any bounded set E is contained in a ball with diameter $2\mathrm{diam}(E)$. $\square$

The constant in the upper bound is far from optimal. A better upper bound can be derived by sharper estimates on the diameter of the smallest ball that is needed to cover a set of fixed diameter. In [Fed69, Corollary 2.10.42]) it is shown that in ambient dimension n the inequality $\mathfrak{S}^s \leq \left(\frac{2n}{n+1}\right)^{s/2}\mathfrak{H}^s$ holds. In the case $n = 2$ and $s = 1$ this bound appears to be optimal. If a set has diameter a, the largest line segment it can contain, has length a. However, the largest such set is the so-called Reuleaux triangle:



A basic calculation shows that this set can be covered by a ball of radius as small as $r = \frac{a}{\sqrt{3}}$ which gives the estimate $\mathfrak{S}^s \leq \left(\frac{4}{3}\right)^{1/2}\mathfrak{H}^s$.

The lemma also shows that a notion of "spherical dimension" would not give anything different from the Hausdorff dimension. However, there are sets for which spherical and Hausdorff measure have different values (see[Mat95, Section 5.1] and the reference therein).

6.4.2 The dyadic-net measure

If we restrict the familiy $\mathcal{F}$ to the half-open dyadic cubes, i.e. sets of the form

$$\{x \in \mathbb{R}^n \;:\; k_i 2^{-m} \leq x_i < (k_i + 1)2^{-m}\}$$

for $k \in \mathbb{Z}^n$ and $m \in \mathbb{N}$, we obtain a measure, called the **dyadic-net measure** and we denote it by $\mathfrak{N}^s$. Similar to above, this measure relates to the Hausdorff measure by

$$\mathfrak{H}^s \leq \mathfrak{N}^s \leq 4^s n^{s/2}\mathfrak{H}^s.$$

Hence, the dyadic-net measure also gives the same notion of fractal dimension as the Hausdorff and the spherical measure. In some sense it is simpler to use since each covering with dyadic cubes can be subdivided into a cover with smaller dyadic cubes which may be helpful in some circumstances.

6.4.3 Measures based on projection and λ^m

For an integer m we could reuse the m-dimensional Lebesgue measure to obtain an m-dimensional measure for subsets of $\mathbb{R}^n$ as follows: Denote by $O(n, m)$ the set of all orthonormal projections from $\mathbb{R}^n$ to m-dimensional subspaces, i.e. all $m \times n$ matrices with orthonormal rows. For any $E \subset \mathbb{R}^n$ define

$$\zeta(E) = \sup\{\lambda^m(P(E)) \;:\; P \in O(n, m)\}$$

In other words: For a set E which shall be used for a covering, we count the largest m-dimensional Lebesgue measure that we can get by projecting it onto an m-dimensional subspace.

This gives rise to two measures:

1. If we allow for $\mathcal{F}$ all Borel subsets of $\mathbb{R}^n$ we get the m-dimensional **Gross measure**, denoted by $\mathfrak{G}^m$.
2. If we allow for $\mathcal{F}$ only the closed, convex subsets of $\mathbb{R}^n$, the result is called m-dimensional **Carathéodory measure** and denoted by $\mathfrak{C}^m$.

Obviously $\mathfrak{G}^m \leq \mathfrak{C}^m$ and also for $m = n = 1$: $\mathfrak{G}^1 = \mathfrak{C}^1 = \mathfrak{H}^1 = \lambda^1$.

Moreover it holds that

$$\sup\{\lambda^m(P(E)) \;:\; P \in O(n,m)\} \leq V_m \left(\frac{\operatorname{diam}(E)}{2}\right)^m$$

which shows that Hausdorff and spherical measure are bounded from below by Gross and Carathéodory measure. I could not find neither examples for which the Gross or Carathéodory measure produce something unreasonable in comparison with the Hausdorff measure, nor could I find examples for the application of these measures.

6.4.4 Lebesgue-Stieltjes measure

For $n = m = 1$, a non-decreasing function $g : \mathbb{R} \to \mathbb{R}$ and as $\mathcal{F}$ the non-empty bounded and open intervals we define

$$\zeta(]a,b[) = g(b) - g(a).$$

The resulting measure if the Lebesgue-Stieltjes measure and we denote it by λ_g. The corresponding integral is often written as

$$\int f \mathrm{d}g = \int f \mathrm{d}\lambda_g.$$

If g is continuously differentiable then it holds that

$$\int f \mathrm{d}g = \int f g' \mathrm{d}\lambda.$$

7 Distances between metric measure spaces

We arrived at the last chapter of these notes and in fact, the topic of this chapter is arguable the most geometrical in the whole lecture notes. In this chapter we briefly look at ways to compare objects as sets, metric spaces, and even metric spaces endowed with a measure. Since this is the last chapter and time is too short to cover the topic in depth, I will skip some of the longer proofs. We start with the comparison of sets within a metric set—the concept here is, again, named after Hausdorff.

7.1 Hausdorff distance

Definition 7.1.1 (Hausdorff distance)**.** Let (X, d) be a metric space and let A, B be compact subsets of X. The **Hausdorff distance** between A and B is

$$d_H(A, B) = \max\Big(\sup_{a\in A}\inf_{b\in B} d(a, b), \sup_{b\in B}\inf_{a\in A} d(a, b)\Big).$$

An alternative formulation uses the distance of a point a to a set B, namely with $\operatorname{dist}(a, B) = \inf_{b\in B} d(a, b)$ we have

$$d_H(A, B) = \max\Big(\sup_{a\in A}\operatorname{dist}(a, B), \sup_{b\in B}\operatorname{dist}(b, A)\Big).$$

Another alternative uses the ϵ-enlargements (which already used for the Prokhorov metric in Remark 5.3.4), namely with $A_\epsilon = \{x \;:\; \operatorname{dist}(x, A) \leq \epsilon\}$ we have

$$d_H(A, B) = \inf\{\epsilon > 0 \;:\; B \subset A_\epsilon,\; A \subset B_\epsilon\}.$$

Intuitively, the Hausdorff distance measures, how well the sets A and B cover each other.

Lemma 7.1.2. *The Hausdorff distance satisfies the axioms for a metric.*

Proof. By definition it is symmetric and non-negative. It is also definite since, by the last reformulation, $d_H(A, B) = 0$ implies that every point in A is arbitrarily close to B and vice versa, and by compactness we deduce that $A = B$.

For $a \in A$ and $b \in B$ We have

$$\operatorname{dist}(a, C) \leq d(a, b) + \operatorname{dist}(b, C) \leq d(a, b) + d_H(B, C)$$

and taking the infimum over b we obtain

$$\operatorname{dist}(a, C) \leq \operatorname{dist}(a, B) + d_H(B, C) \leq d_H(A, B) + d_H(B, C).$$

This also holds if we take the supremum over a. Similarly we could show

$$\operatorname{dist}(A, c) \leq d_H(A, B) + d_H(B, C)$$

which concludes the proof of the triangle inequality. $\square$

In other words: The Hausdorff metric turns the set of a compact subsets of a metric space into a metric space itself. We denote this metric space by $(c(X), d_H)$.

Example 7.1.3. Some simple examples in $X = \mathbb{R}$ are:

1. $d_H([0,1],[1,2]) = 1$, $d_H([0,1],[2,3]) = 2$
2. $d_H([0,1], 0) = 1$.
3. $d_H([0,1] \cup [2,3],[1,2]) = 1$.

Theorem 7.1.4. *The metric space $(c(X), d_H)$ is complete if X is so and compact if X is so.*

For the proof we refer to [BBI01, Proposition 7.3.7, Theorem 7.3.8].

One can reformulate the Hausdorff distance with the help of "correspondences":

Definition 7.1.5. Let $A, B \subset X$. A subset $R \subset A \times B$ is a **correspondence** between A and B or a **set coupling** of A and B if

- for any $a \in A$ there exists $b \in B$ such that $(a, b) \in R$ and.
- for any $b \in B$ there exists $a \in A$ such that $(a, b) \in R$.

In other words, $R \subset X \times X$ is a correspondence between A and B if the projection of R onto the first coordinate is A and the projection onto the second one is B.

Lemma 7.1.6. *Let (X, d) be a compact metric space and let $A, B \subset X$ be compact. Then it holds that*

$$d_H(A, B) = \inf\{ \sup_{(a,b)\in R} d(a, b) \ : \ R \text{ set coupling of } A \text{ and } B\}.$$

Proof. Let $\epsilon > 0$ and let R be a set coupling of A and B such that $d(a, b) \leq \epsilon$ for $(a, b) \in R$. It follows that $\inf_{b\in B} d(a, b) \leq \epsilon$ for all $a \in A$ and $\inf_{a\in A} d(a, b) \leq \epsilon$ for all $b \in B$. By the definition of the Hausdorff distance, it follows that $d_H(A, B) \leq \epsilon$.

Now assume that $d_H(A,B) \le \epsilon$. Then, for each $a \in A$ there exists $b \in B$ such that $d(a,b) \le \epsilon$. Hence, we can define $\phi : A \to B$ such that $d(a, \phi(a)) \le \epsilon$ for all $a \in A$. Similarly we obtain $\psi : B \to A$ such that $d(\psi(b), b) \le \epsilon$ for all $b \in B$. Now set $R = \{(a, \phi(a)) \ : \ a \in A\} \cup \{(\psi(b), b) \ : \ b \in B\}$ which is obviously a correspondence between A and B by construction $d(a,b) \le \epsilon$ for $(a,b) \in R$. □

The Hausdorff distance allows to compare different compact sets within a metric space. As a next step, we will compare two different metric spaces.

7.2 The Gromov-Hausdorff distance

If we want to compare two different metric spaces, the Hausdorff distance can not be used directly. However, we may embed any metric space into a larger one and if we embed two metric spaces into the same larger metric space, we may employ the Hausdorff distance there. This idea is due to Gromov:

Definition 7.2.1 (Gromov-Hausdorff distance)**.** Let (X, d_X) and (Y, d_Y) be two compact metric spaces. Then the **Gromov-Hausdorff distance** of X and Y is

$$d_{GH}(X,Y) = \inf\Big\{ d_H(f(X), g(Y)) \ : \ f : (X, d_X) \to (Z, d_Z),\ g : (Y, d_Y) \to (Z, d_Y) \\ \text{isometric embeddings into a metric space } (Z, d_Z) \Big\}.$$

This definition seems pretty complicated as one has to form an infimum over all metric spaces (is this even a proper set?). The following reformulation sounds more mild:

> For two metric spaces X and Y and $r > 0$ it holds that $d_{GH}(X,Y) < r$ if there exists a metric space (Z, d_Z) and isometric embeddings $f : X \to Z$ and $g : Y \to Z$ such that $d_H(f(X), g(Y)) < r$.

As for the Hausdorff distance, there is a different formulation of this distance that uses the notion of coupling. To define these, we use the so-called **disjoint union** $X \sqcup Y$ of two sets X and Y, which can be understood as placing all elements of X and Y into a joint new set but keeping them distinguishable.

Definition 7.2.2 (Metric coupling)**.** Let $(X, d_X), (Y, d_Y)$ be two compact metric spaces. A **metric coupling** of X and Y is a semi-metric* d on the disjoint union $X \sqcup Y$ such that for $x, x' \in X$ it holds $d(x,x') = d_X(x,x')$ and for all $y, y' \in Y$ it holds that $d(y,y') = d_Y(y,y')$.

Example 7.2.3 (Metric coupling of finite spaces)**.** To see that metric couplings may exists consider the situation of two finite metric space X and Y (with

*A semi-metric has all properties of a metric except the definiteness, i.e. it may be zero for non-equal elements. In other contexts semi-metrics are called pseudo-metrics, so be careful. However, semi-metric better fits to semi-norm.

$n_X = \#X$ and $n_Y = \#Y$). Then, the metrics on X and Y can be written as $n_X \times n_X$ and $n_Y \times n_Y$ matrices D_X and D_Y, respectively. Necessarily D_Y and D_X are symmetric, zero on the diagonal and have positive entries elsewhere. Also, the triangle inequalities need to be fulfilled. A metric coupling is then given by an $(n_X + n_Y) \times (n_X + n_Y)$ matrix D of the form

$$D = \begin{bmatrix} D_X & M \\ M^T & D_Y \end{bmatrix}$$

where M is $n_X \times n_Y$ and non-negative (note that we allowed a semi-metric d). The $n_X n_Y$ many entries are also constrained by linear inequalities enforce by the triangle inequality.

Example 7.2.4 (Metric couplings exist)**.** For two compact metric space X and Y set $e = \max(\mathrm{diam}(X), \mathrm{diam}(Y))$. Then we define a metric d on $X \sqcup Y$ by $d(x,y) = e/2$ if $x \in X$ and $y \in Y$ (and, of course, $d(x,x') = d_X(x,x')$ and $d(y,y') = d_Y(y,y')$ if $x,x' \in X$ and $y,y' \in Y$). Indeed, this d is a metric coupling.

The Gromov-Hausdorff distance in terms of set and metric couplings reads as follows:

Theorem 7.2.5. *The Gromov-Hausdorff distance of two compact metric spaces X and Y is equivalently expressed as*

$$d_{GH}(X,Y) = \inf\Big\{ \sup_{(x,y)\in R} d(x,y) \ : \ R \text{ set coupling and } d \text{ metric coupling of } X \text{ and } Y \Big\}.$$

Before we prove this theorem note that we can express the formulation of the theorem with the help of the Hausdorff distance d_H as

$$\begin{aligned} &\inf\Big\{ \sup_{(x,y)\in R} d(x,y) \ : \ R \text{ set coupling and } d \text{ metric coupling of } X \text{ and } Y \Big\} \\ &= \inf\Big\{ d_H(X,Y) \ : \ d_H \text{ Hausdorff metric in } X \sqcup Y \text{ w.r.t any metric coupling } d \text{ of } X \text{ and } Y \Big\} \end{aligned}$$

Proof. Let's denote the value on the right hand side in the formulation of theorem as $d'_{GH}(X,Y)$ for a moment. Note that it holds that $d'_{GH}(X,Y)$ is the infimum over all $r > 0$ such that there exists a metric coupling d of X and Y such that $d_H(X,Y) < r$ (in the Hausdorff metric w.r.t. this coupling).

We identify $X \sqcup Y$ with $f(X) \sqcup g(Y) \subset Z$ for some fixed isometric embeddings f and g, i.e. on $X \sqcup Y$ we define $d(x,y) = d_Z(f(x), g(y))$. Actually this is only a semi-metric as it may be zero for non-equal x and y (if the embedding overlap). However, for this metric d we have $d_H(X,Y) < r$ as demanded. □

Example 7.2.6 (Gromov-Hausdorff distance to a point)**.** Denote by $\boxdot$ the metric space that consists of only one point. Then it holds that

$$d_{GH}(X, \boxdot) = \mathrm{diam}(X)/2.$$

Lemma 7.2.7. *The Gromov-Hausdorff distance is a metric on the space of all isometry classes of compact metric spaces.*

Proof. It is clear that $d_{GH}(X,Y) = d_{GH}(Y,X) \geq 0$.

For the triangle inequality let d_{12} and d_{23} be metric couplings of X_1 and X_2, and X_2 and X_3, respectively. We define a metric coupling d_{13} of X_1 and X_3 by defining for $x_1 \in X_1$ and $x_3 \in X_3$, $d_{13}(x_1,x_3) = \inf_{x_2 \in X_2}[d_{12}(x_1,x_2) + d_{23}(x_2,x_3)]$ (one needs to show that this actually gives a metric coupling). By definition we have $d_H(X_1,X_3) \leq d_H(X_1,X_2) + d_H(X_2,X_3)$ (where the Hausdorff-metrics $d_H(X_i,X_j)$ have to be understood w.r.t. the metric d_{ij}). Taking the infimum over all metric couplings d_{12} and d_{13} we get $d_{GH}(X_1,X_3) \leq d_{GH}(X_1,X_2) + d_{GH}(X_2,X_3)$.

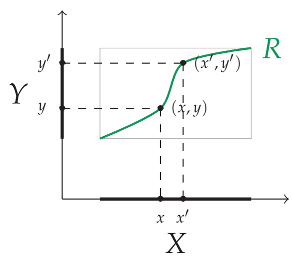
It remains to show that $d_{GH}(X,Y) = 0$ implies that X and Y are isometric and we refer to [BBI01, Theorem 7.3.30] for this fact. □

Since the construction of metric couplings may be cumbersome, one may appreciate the fact that the Gromov-Hausdorff distance can be formulated in terms of set couplings alone. The downside is, that the formulation involves twice as many variables. We start with the notion of "distortion" of a set coupling.

Definition 7.2.8. Let R be a set coupling of two metric spaces (X, d_X) and (Y, d_y). The **distortion** of R is

$$\operatorname{dis} R = \sup\{|d_X(x,x') - d_Y(y,y')| \ : \ (x,y),(x',y') \in R\}.$$

The distortion of a coupling measure how much one may distort the distance between a pair of points and all "coupled" point pairs:



Theorem 7.2.9. *For two compact metric spaces X and Y it holds that*

$$d_{GH}(X,Y) = \tfrac{1}{2}\inf\{\operatorname{dis} R \ : \ R \textit{ set coupling of } X \textit{ and } Y\}.$$

In other words, $d_{GH}(X,Y)$ is the infimum over all $r > 0$ such that there exists a set coupling of X and Y with $\operatorname{dis} R \leq 2r$.

The **proof** can be found in [BBI01, Theorem 7.3.15].

7.3 Gromov-Hausdorff-Wasserstein distances

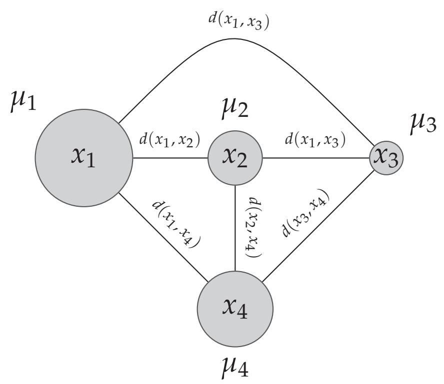
Here we go the last step and give a notion for a distance between metric measure spaces. Basically, these are metric spaces endowed with an additional Borel probability measure. It turns out that there are straight-forward ways to extend the Gromov-Hausdorff distances to mm-spaces, but, somewhat unexpectedly, at least two different extensions are indeed different.

Definition 7.3.1. A **metric measure space** (mm-space for short) is a triple (X, d_X, μ_X) consisting of a compact metric space (X, d_X) and a finite Borel measure μ_X with full support, i.e. $\mathrm{supp}\mu_X = X$.

We call two mm-spaces (X, d_X, μ_X) and (Y, d_Y, μ_Y) **isomorphic** if there exists an isometric $\psi : X \to Y$ such that $\mu_Y = \psi\#\mu_X$.

One obvious example for a metric measure space is given by any Borel probability measure defined on $\mathbb{R}^n$ (equipped with the euclidean metric) and restricted to its support.

Example 7.3.2 (Finite mm-spaces). Finite mm-spaces are particularly simple to visualize. They consist of a finite metric space, i.e. a finite number of points $X = \{x_1, \ldots, x_n\}$, a metric which can be given as a matrix $D_X = (d_X(x_i, x_j))$ and a weight vector $\mu_1, \ldots, \mu_2$. One may visualize this as a graph where the nodes correspond to the points x_i, the edges correspond to the distances $d(x_i, x_j)$ and the nodes carry the weights μ_i.



To define a notion of distance for metric measure spaces, it comes as no surprise that one needs to "couple the measures" somehow.

The notion of measure coupling appears to be straight forward:

Definition 7.3.3. Let (X, d_X, μ_X) and (Y, d_Y, μ_Y) be two mm-spaces. A **measure coupling** of X and Y is a measure γ on the product space $X \times Y$ such that for all measurable sets $A \subset X$ and $B \subset Y$ it holds that

$$\gamma(A \times Y) = \mu_X(A), \quad \text{and} \quad \gamma(X \times B)\mu_Y(B).$$

In other words, the push-forwards of γ onto the coordinates of $X \times Y$ are μ_X and μ_Y, respectively.

Remark 7.3.4. A measure coupling γ of X and Y always induces a set coupling: Since we assumed that μ_X and μ_Y have full support, a set coupling of X and Y is given as suppγ.

To obtain a distance for mm-spaces we start from the Gromov-Hausdorff distance and try to incorporate the measure. We have different possibilities here: For one, we could start from the Gromov-Hausdorff distance in the form of Theorem 7.2.5 which is

$$d_{GH}(X,Y) = \inf\{ \sup_{(x,y)\in R} d(x,y) \ : \ R, d \text{ set and metric coupling, rsp.}\}.$$

Note the we can write this equivalently as

$$d_{GH}(X,Y) = \inf_{d,R} \|d\|_{L^\infty(R)}.$$

This motivates the following distance, which we call **Gromov-Hausdorff-Wasserstein distance**

$$d_{GHW}(X,Y) = \inf_{\gamma,d} \|d\|_{L^\infty(X\times Y,\gamma)}$$

where we take the infimum over all measure couplings γ and all set couplings d.

Note that we consider the ∞-norm on the space $X \times Y$ **equipped with the measure** γ. This is an important point since the measure γ may be supported on a set that is way smaller than $X \times Y$ and everything outside of the support of γ will not contribute to the essential supremum in the ∞-norm. We could make this more obvious by writing

$$d_{GHW}(X,Y) = \inf_{\gamma,R,d} \sup_{(x,y)\in R} d(x,y)$$

where we infimize over all measure, set and metric couplings. Furthermore, we could also define for $p \geq 1$

$$d^p_{GHW}(X,Y) = \inf_{\gamma,d} \|d\|_{L^p(X\times Y,\gamma)}. \tag{7.1}$$

A different possibility comes from the form of Theorem 7.2.9. Recall that we had

$$d_{GH}(X,Y) = \tfrac{1}{2}\inf_R \sup\{|d_X(x,x') - d_Y(y,y')| \ : \ x,x' \in X,\ y,y' \in Y\}$$

where we took the infimum over all set couplings. By defining the function $D : X \times Y \times X \times Y \to \mathbb{R}$ given by $D_{X,Y}(x,y,x'y') = |d_X(x,x') - d_Y(y,y')|$ we can write this as

$$d_{GH}(X,Y) = \tfrac{1}{2}\inf_R \|D_{X,Y}\|_{L^\infty(R\times R)}.$$

This motivates

$$\tilde{d}_{GHW}(X,Y) = \tfrac{1}{2}\inf_{R,\gamma} \|D_{X,Y}\|_{L^\infty(R\times R,\gamma\times\gamma)} = \tfrac{1}{2}\inf_{\gamma} \|D_{X,Y}\|_{L^\infty(X\times Y\times X\times Y,\gamma\times\gamma)}$$

where we infimize over all set and measure couplings (in fact, the set coupling, is not needed here) or its "p-relatives"

$$\begin{aligned}\tilde{d}^p_{GHW}(X,Y) &= \tfrac{1}{2}\inf_{\gamma} \|D_{X,Y}\|_{L^p(\gamma\times\gamma)} \\ &= \tfrac{1}{2}\inf_{\gamma}\Big(\int_{X\times Y}\int_{X\times Y} |d_X(x,x') - d_Y(y,y')|^p \mathrm{d}\gamma(x,y)\mathrm{d}\gamma(x',y')\Big)^{1/p}.\end{aligned} \tag{7.2}$$

The following observation is a direct consequence of the definition.

Lemma 7.3.5. *It holds that*

$$\tilde{d}^p_{GHW}(X,\boxdot) = \tfrac{1}{2}\mathrm{diam}_p(X)$$

where $\mathrm{diam}_p(X) = \Big(\int_X\int_X d_X(x,x')^p \mathrm{d}\mu(x)\mathrm{d}\mu(x')\Big)^{1/p}$ *denotes the p-diameter of* X.

One may note that the "∞ versions" of d^p_{GHW} do not really use that much information of the measures but merely their support. The versions with $p<\infty$, however, really incorporate the structure of the measures.

One the positive side for the ∞ versions is, that they are in fact equal:

Theorem 7.3.6. *It holds that* $d^\infty_{GHW}(X,Y) = \tilde{d}^\infty_{GHW}(X,Y)$.

Proof. Let $\epsilon > 0$ and and $d^\infty_{GHW}(X,Y) < \epsilon$. Further let d and γ be a metric and measure coupling of X and Y such that $d(x,y)\leq\epsilon$ for all $x,y\in\operatorname{supp}\gamma$. Since $d(x,x')\leq d(x,y)+d(y,y')+d(y',x')$ we get $|d_X(x,x')-d_Y(y,y')|\leq d(x,y)+d(x',y')\leq 2\epsilon$ and thus $\tilde{d}^\infty_{GHW}(X,Y)<\epsilon$.

Conversely, let $\epsilon>0$ and γ be a measure coupling of X and Y such that $\|D_{X,Y}\|_{L^\infty(\operatorname{supp}\gamma\times\operatorname{supp}\gamma,\gamma\times\gamma)} < \epsilon$. Then $|d_X(x,x')-d_Y(y,y')|\leq 2\epsilon$ for each $(x,y),(x',y')\in\operatorname{supp}\gamma$. Define d as

$$d(x,y) = \inf_{(x',y')\in\operatorname{supp}\gamma}\big(d_X(x,x')+d_Y(y,y')\big).$$

One can show, that this gives a metric coupling and that shows $d^\infty_{GHW}(X,Y) < \epsilon$. ☐

In contrast to this, the following example shows that both variants of the Gromov-Hausdorff-Wasserstein distance are in fact different for $p<\infty$.

Example 7.3.7. Let Δ_n be the $(n-1)$-simplex, i.e. a set of n distinct point all separated by a distance of 1 and each point carries the mass $1/n$. We denote this measure by ν_n. Further consider the space $\boxdot$ consisting of a single point, equipped with the Dirac measure δ.

Obviously there is only one measure coupling of ν_n and δ on $\Delta_n \times \boxdot$, namely the measure ν_n itself.

The metric couplings are also not that many: They have the form

$$D(\Delta_n, \boxdot) = \left\{ \begin{bmatrix} d_X & f \\ f^T & 0 \end{bmatrix} \ : \ f \in \mathbb{R}^n,\ f_i \geq 0,\ f_i - f_j \leq 1 \leq f_i + f_j,\ i \neq j \right\}$$

The Gromov-Hausdorff-Wasserstein distance from (7.1) with $p = 1$ is

$$d^1_{GHW}(\Delta_n, \boxdot) = \inf_f \sum_i f_i \nu_n(i) = \tfrac{1}{n} \inf \sum_i f_i$$

where the infimum is taken over all f satisfying the above definition. Among the inequalities that define these f, there are $n(n-1)/2$ many of the form $1 \leq f_i + f_j$, $i \neq j$. If we add these $n(n-1)/2$ many inequalities and note that each f_i appears in exactly $n-1$ of them, we obtain $n(n-1)/2 \leq (n-1)\sum_i f_i$. This implies that

$$d^1_{GHW}(\Delta_n, \boxdot) \geq \tfrac{1}{2}.$$

However, the definition from (7.2) gives,

$$\tilde{d}^1_{GWH}(\Delta_n, \boxdot) = \tfrac{1}{2} \int_{\Delta_n} \int_{\Delta_n} d(x_i, x_j) \mathrm{d}\nu_n \mathrm{d}\nu_n = \tfrac{1}{2} \int_{\Delta_n} \tfrac{n-1}{n} \mathrm{d}\nu_n = \tfrac{n-1}{2n}.$$

We observe that $\tilde{d}^1_{GWH}(\Delta_n, \boxdot) < d^1_{GHW}(\Delta_n, \boxdot)$ for $n > 2$.

As far as I see, it is not yet clear, which form of the Gromov-Hausdorff-Wasserstein distances is the most appropriate for different purposes. These purposes may be: Simple to calculate (especially, not NP-hard for finite spaces), good fit to intuition, favorable theoretical properties like (pre-)compactness or completeness. While the form $\tilde{d}^1_{GWH}$ seems to be practically from a numerical point of view, it lacks completeness. Other variants pose better theoretical properties but are hard to handle on the computer… More insight in recent developments on this topic can be found in the article

- Gromov-Wasserstein distances and the metric approach to object matching, F. Mémoli. *Foundations of Computational Mathematics.* 11(4), 2011, 417–487.

With this inconclusive remark, we end these lecture notes and hope that the interested reader will find more mind-food in the Bibliography.

Bibliography

[AFP00] Luigi Ambrosio, Nicolo Fusco, and Diego Pallara. **Functions of Bounded Variation and Free Discontinuity Problems**. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York, 2000.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. **A course in metric geometry**, volume 33. American Mathematical Society Providence, 2001.

[BMA06] Giuseppe Buttazzo, Gérard Michaille, and Hedy Attouch. **Variational analysis in Sobolev and BV spaces: applications to PDEs and optimization**, volume 6. SIAM, 2006.

[EG92] Lawrence C. Evans and Ronald F. Gariepy. **Measure theory and fine properties of functions**. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1992.

[Fed69] Herbert Federer. **Geometric measure theory**. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[FL07] Irene Fonseca and Giovanni Leoni. **Modern methods in the calculus of variations: L^p spaces**. Springer Monographs in Mathematics. Springer, New York, 2007.

[KP08] Steven G. Krantz and Harold R. Parks. **Geometric integration theory**. Cornerstones. Birkhäuser Boston Inc., Boston, MA, 2008.

[Mat95] Pertti Mattila. **Geometry of Sets and Measures in Euclidean Spaces: Fractals and Rectifiability**. The Cambridge studies in advanced mathematics. Cambridge University Press, 1995.

[Mor09] Frank Morgan. **Geometric measure theory**. Elsevier/Academic Press, Amsterdam, fourth edition, 2009. A beginner's guide.

[Rud87] Walter Rudin. **Real and complex analysis**. Tata McGraw-Hill Education, 1987.

[Tao11] Terence Tao. **An introduction to measure theory**, volume 126. American Mathematical Soc., 2011. available (with Errata) at http://terrytao.wordpress.com/books/an-introduction-to-measure-theory/.

Index

Dirk Lorenz | Institute for Analysis and Algebra | TU Braunschweig | 38092 Braunschweig | Germany |
d.lorenz@tu-braunschweig.de