Analysis III
Lecture notes

Prof. Dr. Christoph Thiele*
Winter term 2015/16
Universität Bonn

July 5, 2016

Contents

*Notes by Polona Durcik and Michał Warchalski.

1 Measure theory

In Analysis II we used an approach to integration in $\mathbb{R}^n$ defining integrals of continuous, compactly supported functions in $\mathbb{R}^n$. Certain technical steps were then taken, to extend this notion to integrals over certain bounded domains, say boxes Q, in $\mathbb{R}^n$.

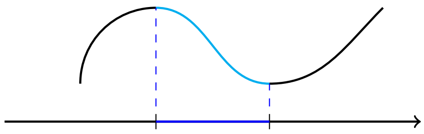


Figure 1: Restriction of a continuous function to an interval $Q \subset \mathbb{R}$.

A natural way to integrate such functions we would be to make sense of the expression

$$\int_{\mathbb{R}^n} f\mathbf{1}_Q$$

where $\mathbf{1}_Q$ equals 1 if $x \in Q$ and it equals 0 otherwise. However, the function $f1_Q$ is in general not continuous on $\mathbb{R}^n$, so a number of technical steps were needed to circumvent this issue. This example shows the potential usefulness of more general notions of integrals.
The function $\mathbf{1}_Q$ is one of the most basic examples of discontinuous functions. It is called the *characteristic function* of Q and can be defined for any set E by setting

$$\mathbf{1}_E(x) = \begin{cases} 1; & x \in E \\ 0; & x \notin E \end{cases}$$

If we knew how to integrate such functions, we would define

$$\int_{\mathbb{R}^n} \mathbf{1}_E(x)dx$$

to be the *volume* (or *measure*) of E.
On the other hand, we can see (exercise) that we can approximate any bounded compactly supported function with finite linear combinations of characteristic functions, all supported on the same compact set, uniformly. This will then allow by linearity of the integral and approximation arguments

to extend the definition of the integral from characteristic functions to such functions as well.

Question: Can one assign a volume to every set $E \subset \mathbb{R}^n$?

Of course, the volume should have certain properties which agree with our intuitive notion of volume. For instance, the volume of a set should be invariant under translations of that set. In higher dimensions it should also be invariant under its rotations. The unit ball should have a finite volume. Moreover, the volume of a union of disjoint sets should equal the sum of the volumes of these sets, which is a natural condition related to linearity of the integral. Namely, if $E \cap F = \emptyset$, then $\mathbf{1}_{E\cup F} = \mathbf{1}_E + \mathbf{1}_F$, which implies

$$\int_{\mathbb{R}^n} \mathbf{1}_{E\cup F} = \int_{\mathbb{R}^n} \mathbf{1}_E + \int_{\mathbb{R}^n} \mathbf{1}_F$$

and hence volume$(E \cup F)$ = volume(E) + volume(F).
The following theorem says that under these conditions, it is not possible to assign volume to every subset of $\mathbb{R}$.
By $\mathcal{P}(\mathbb{R})$ we denote the power set of $\mathbb{R}$, i.e. the set of all subsets of $\mathbb{R}$.

Theorem 1.1. *There does not exist $\mu : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ such that*

1. *(Translation invariance) $\mu(E + y) = \mu(y)$ for all $E \in \mathcal{P}(\mathbb{R}), y \in \mathbb{R}$, where $E + y := \{x + y : x \in E\}$*
2. *(Countable additivity) If A_j, $j \in \mathbb{N}$ are pairwise disjoint, i.e. $A_j \cap A_k$ for $j \neq k$, then*
$$\mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j)$$
3. $0 < \mu([0, 1)) < \infty$

Note that since the function μ takes vaues in $[0, \infty]$, it is very natural to consider countable sums as in the last item.

Proof. Assuming existence of such a map we are going to derive a contradiction. We start with some preliminary observations.

- $\mu(\emptyset) = 0$

  To see this set $A_1 = [0, 1)$, $A_j = \emptyset$ for $j > 1$. These sets are pairwise disjoint. Then
$$\mu([0, 1)) = \mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) \overset{2.}{=} \sum_{j=1}^{\infty} \mu(A_j) = \mu([0, 1)) + \sum_{j=2}^{\infty} \mu(A_j).$$

By 3. $\mu([0,1)) \neq \infty$, so we have $\sum_{j=2}^{\infty} \mu(A_j) = 0$. Since $\mu(A_j) \geq 0$ for each j, we must have $\mu(A_2) = 0$. That is, $\mu(\emptyset) = 0$.

- (Finite additivity) Property 2. holds also for finite families A_j, $j = 1, \ldots, N$.

  This follows by setting $A_j = \emptyset$ for $j > N$ and using $\mu(\emptyset) = 0$.

- (Monotonicity) If $A \subset B$, then $\mu(A) \leq \mu(B)$.

  To see this we write $B = A \cup (B \setminus A)$, which is a disjoint union. Then we use finite additivity of μ to conclude

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

  Each of the terms on the right hand-side is non-negative, so we must have $\mu(A) \leq \mu(B)$.

- (Countable subadditivity) For any (not necessarily pairwise disjoint) sets A_j, $j \in \mathbb{N}$, we have

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

  To see this for two sets A, B write $A \cup B = A \cup (B \setminus A)$ and conclude

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

  the last inequality following by monotonicity (since $B \setminus A \subset B$). For countably many sets induct, the details are left as an exercise.

We return to the proof of the theorem. Define

$$R := \{x + \mathbb{Q} : x \in \mathbb{R}\}$$

If $M, N \in R$, then $M = N$ or $M \cap N = \emptyset$. Indeed, suppose that $M = x + Q$, $N = x' + Q$ for some $x, x' \in \mathbb{R}$. If $M \cap N \neq \emptyset$, then there exists $y \in M \cap N$ and therefore $q, q' \in \mathbb{Q}$ such that $x + q = y = x' + q'$. Then for each $p \in Q$ we have

$$x + p = x' + \underbrace{q' - q + p}_{\in \mathbb{Q}} \in N,$$

which implies $M \subset N$. By symmetry it follows $N \subset M$ and hence $M = N$.

Next we claim that for each $M \in R$ there exists $y \in M$ such that $y \in [0, 1)$. That is, $M \cap [0, 1) \neq \emptyset$ for each $M \in R$. To see this we write as before $M = x + \mathbb{Q}$ for some $x \in \mathbb{R}$. Since $\mathbb{Q}$ is dense[1] in $\mathbb{R}$ there is $q \in \mathbb{Q}$ with $|x - 1/2 - q| \leq 1/4$, i.e. $1/4 \leq x - q \leq 3/4$. So set $y := x - q \in M$.
By the axiom of choice there exists a function $\varphi : R \to [0, 1)$ such that for each $M \in R : \varphi(M) \in M$. The range of this map ϕ is a subset of $[0, 1)$ which contains exactly one element of each $M \in \mathbb{R}$. We denote this subset by $A := \varphi(R)$. Now we claim that

$$\bigcup_{q \in \mathbb{Q}} A + q = \mathbb{R},$$

i.e. the union of all possible translates of A by rational numbers is the whole real line. Indeed, note that for $z \in \mathbb{R}$ we have $z \in z + \mathbb{Q} \in R$. Then $z - \varphi(z + \mathbb{Q}) \in \mathbb{Q}$ and so

$$z \in \underbrace{\varphi(z + \mathbb{Q})}_{\in A} + \mathbb{Q}.$$

This implies $\mathbb{R} \subset \cup_q A + q$, the other inclusion being trivial.
Now we can conclude

$$0 < \mu([0, 1)) \overset{(*)}{\leq} \mu(\mathbb{R}) \leq \sum_{q \in \mathbb{Q}} \mu(A + q) \overset{(**)}{=} \sum_{q \in \mathbb{Q}} \mu(A)$$

where $(*)$ follows by monotonicity and $(**)$ by 1. This implies $\mu(A) \neq 0$.
Let us now consider all possible translates of A by rational number smaller than 10. We must have

$$\Big(\bigcup_{q \in \mathbb{Q}, |q| < 10} A + q \Big) \subset [-11, 11).$$

Since $[-11, 11]$ can be viewed as the union of translates of 22 copies of $[0, 1)$, by translation invariance and additivity we have

$$\mu([-11, 11)) = 22\mu([0, 1)) < \infty.$$

But then

$$\infty > \mu([-11, 11)) \overset{(*)}{\geq} \sum_{q \in \mathbb{Q}, |q| < 10} \mu(A + q) \overset{(**)}{\geq} \sum_{q \in \mathbb{Q}, |q| < 10} \mu(A) \overset{(**)}{\geq} \sum_{q \in \mathbb{Q}, |q| < 10} \varepsilon = \infty$$

which is a contradiction. We used: $(*)$ monotonicity and additivity, $(**)$ translation invariance, $(***)$ $\mu(A) \neq 0$ and thus $\mu(A) \geq \varepsilon$ for some $\varepsilon > 0$. $\square$

[1] At this point we observe that the same proof would work with $\mathbb{Q}$ replaced by the dyadic numbers.

In $\mathbb{R}^3$ it is natural to require that the volume of a set is invariant under rotations of this set. Our next goal is to show that in $\mathbb{R}^3$, a map satisfying rotation invariance and properties 1.-3. from Theorem (1.1) does not exist even if we replace 2. by the weaker notion of finite additivity. The interval $[0, 1)$ is now replaced by the unit ball around the origin $B_1(0)$ (the exact choice of this set is not so important).

Theorem 1.2. *There does not exist $\mu : \mathcal{P}(\mathbb{R}^3) \to [0, \infty]$ such that*

1. *(Translation invariance) $\mu(E + y) = \mu(y)$ for all $E \in \mathcal{P}(\mathbb{R}), y \in \mathbb{R}$, where $E + y := \{x + y : x \in E\}$*
2. *(Finite additivity) For any natural number $N \geq 2$, If A_j, $j = 1, \dots, N$ are pairwise disjoint, then*
$$\mu\Big(\bigcup_{j=1}^{N} A_j\Big) = \sum_{j=1}^{N} \mu(A_j)$$
3. $0 < \mu(B_1(0)) < \infty$
4. *(Rotation invariance) $\mu(E) = \mu(TE)$ for all rotations $T \in SO(3)$ and for all $E \in \mathcal{P}(\mathbb{R}^3)$*

We remark that the theorem holds more generally for $\mathbb{R}^n$, $n \geq 3$.

Proof (sketch). Consider the rotations

$$\varphi = \frac{1}{5}\begin{pmatrix} 3 & -4 & 0 \\ 4 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \quad \psi = \frac{1}{5}\begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 4 \\ 0 & -4 & 3 \end{pmatrix}.$$

One can compute

$$\varphi^{-1} = \frac{1}{5}\begin{pmatrix} 3 & 4 & 0 \\ -4 & 3 & 0 \\ 0 & 0 & 5 \end{pmatrix}, \quad \psi^{-1} = \frac{1}{5}\begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & -4 \\ 0 & 4 & 3 \end{pmatrix}.$$

We first note that these are indeed rtotations (exercise) and these rotations are by angles $2\pi\alpha$ with irrational α, in other words positive integer powers of these matrices can not be the identity matrix. By symmetry it suffices to prove this for φ. The entries of 5φ modulo 5 are

$$\begin{pmatrix} 3 & -4 & 0 \\ 4 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and the square of this matrix is congruent itself modulo 5. Hence any positive integer power of this matrix is congruent to itself. Hence $5^n\phi^n$ has some entries not divisible by 5, and hence ϕ^n does not have integer entries and cannot be the identity matrix.
Let G be the set of all products of the form

$$g = g_1^{n_1} g_2^{n_2} \dots g_N^{n_N} \ , \tag{1}$$

where $N \in \mathbb{N}$ (the case $N = 0$ is the empty product, which is equal to the identity matrix), where each $n_j \geq 1$ and each $g_j \in S = \{\varphi, \varphi^{-1}, \psi, \psi^{-1}\}$ and we have the property that if $g_j \in \{\varphi, \varphi^{-1}\}$ then $g_{j+1} \in \{\psi, \psi^{-1}\}$ and conversely if $g_j \in \{\psi, \psi^{-1}\}$ then $g_{j+1} \in \{\varphi, \varphi^{-1}\}$ for all $j < N$. Then G forms a multiplicative group, that is the product of two such matrices is again in G, and the inverse matrix of an element in G is also in G (exercise).
Note that (2) can be written as

$$g = \frac{1}{5^k} \begin{pmatrix} a & u & x \\ b & v & y \\ c & w & z \end{pmatrix},$$

with integers $a, b, c, u, v, w, x, y, z$ and $k = n_1 + n_2 + \cdots + n_N$ We claim that the central entry v is not divisible by 5.
If $N = 0$ this claim is obvious. If $N = 1$ then the claim follows from the previous remarks on powers of elements of S. We then proceed by induction on N. Assume the claim holds for the matrix g as above with certain $N \geq 1$ and let $g_{N+1} \in \{\varphi, \varphi^{-1}\}$ if $g_N \in \{\psi, \psi^{-1}\}$ and $g_{N+1} \in \{\psi, \psi^{-1}\}$ if $g_N \in \{\varphi, \varphi^{-1}\}$. Of the various cases we consider by symmetry the case $g_{N+1} = \varphi$. Then since the last factor of g is in $\{\psi, \psi^{-1}\}$ we note that b is divisible by 5. By the discussion of powers of ϕ we see that the middle entry of $5^{k+1} g g_{N+1}^{n_{N+1}}$ is congruent $-4b + 3v$ modulo 5, which is congruent $3v$ modulo 5 and since v by induction is not divisible by 5 the middle entry of $5^{k+1} g g_{N+1}^{n_{N+1}}$ is not divisible by 5 neither. The other cases follow similarly.
We call k the length of g as in (2) and the above argument shows that the length of g can be read off the middle entry of g as the power of 3 in the denominator of the reduced fraction. We next claim that if some other element

$$h = h_1^{m_1} h_2^{m_2} \dots h_M^{m_M} \ , \tag{2}$$

of analoguous form as in (2) is equal g, then $N = M$ and $n_j = m_j$ for all j and $g_j = h_j$ for all j. To see the claim we may induct by the length of g, which has to be the length of h as well since the length is determined by the two central entries which by assumption are the same. If the length is

zero then necessarily $N = M = 0$ and the claim is clear. If teh length is positive, then by induction it suffices to show that $g_1 = h_1$, since one can then apply induction hypothesis on the elements $g_1^{-1}g$ and $h_1^{-1}h$ which after cancelling the first two factors are seen to be of shorter length. However, g_1 is the unique element in S so that $g_1^{-1}g$ has smaller length than g (exercise) and thus g_1 is determined by g and therefore $g_1 = h_1$.
We can now partition G as

$$G = \{I\} \cup G_\varphi \cup G_{\varphi^{-1}} \cup G_\psi \cup G_{\psi^{-1}} \tag{3}$$

where G_φ contains those elements g from G for which the representation (2) starts with $g_1 = \varphi$ etc.
However, we claim that we can also write

$$G = G_{\varphi^{-1}} \cup \varphi^{-1}G_\varphi \tag{4}$$

and symmetrically

$$G = G_{\psi^{-1}} \cup \psi^{-1}G_\psi.$$

To see (4) we note $\varphi^{-1}G_\varphi \subset \{I\} \cup G_\varphi \cup G_\psi \cup G_{\psi^{-1}}$ since for $g \in G_\varphi$ of the form (2) we can express $\varphi^{-1}g$ in the corresponding form by erasing one g_1 from the left in the product (2) and the remaining product is of the form (2) starting with an element not equal to φ^{-1}. Conversely, $\varphi^{-1}G_\varphi \supset \{I\} \cup G_\varphi \cup G_\psi \cup G_{\psi^{-1}}$ since any element g' in the set on the right hand side can be written as $\varphi^{-1}\varphi g'$ where $\varphi g' \in G_\varphi$.
Now we proceed similarly as in the proof Theorem (1.1). We consider G acting on the set

$$K := B_1(0) \setminus \{x \in B_1(0) : x = gx \text{ for some } g \in G, g \neq I\}\ .$$

One can show (exercise) that $\mu(K) = \mu(B_1(0))$ and thus this set can be seen as "almost all" of the unit ball. If $x \in K$ then $gx \in K$ for $g \in G$ (exercise). We denote

$$R := \{Gx : x \in K\}$$

and show as before that any $M, N \in R$ which are not disjoint must coincide. Then we use the axiom of choice to find a set $A \subset B_1(0)$ which contains exactly one element of each $M \in R$. By (3) we can write K as the union

$$K = AG = A \cup G_\varphi A \cup G_{\varphi^{-1}}A \cup G_\psi A \cup G_{\psi^{-1}}A.$$

Since K contains no fixed points under rotations in G, one can see that the union on the right hand side is disjoint. For example if $ga \in G_\varphi A$ and

$hb \in G_\psi A$ and $ga = hb$ then $a = g^{-1}hb$ and by definition of A we have $a = b$ and since K is fixed point free we have $g^{-1}h = I$, and hence $g = h$, a contradiction to $g \in G_\varphi$ and $h \in G_\psi$.
However, by (4) we can also partition

$$K = G_{\varphi^{-1}}A \cup \varphi^{-1}G_\varphi A = G_{\psi^{-1}}A \cup \psi^{-1}G_\psi A.$$

By rotation invariance of μ we have $\mu(G_\varphi A) = \mu(\varphi^{-1}G_\varphi)$ and $\mu(G_\psi A) = \mu(\psi^{-1}G_\psi)$. Finite additivity of μ then implies

$$\begin{aligned}\mu(K) = \mu(GA) &= \mu(A) + \mu(G_{\varphi^{-1}}A) + \mu(G_\varphi A) + \mu(G_{\psi^{-1}}A) + \mu(G_\psi A)\\ &= \mu(G_{\varphi^{-1}}A) + \mu(G_\varphi A)\\ &= \mu(G_{\psi^{-1}}A) + \mu(G_\psi A)\end{aligned}$$

and hence $2\mu(K) \le \mu(K)$. This contradicts $0 < \mu(K) < \infty$. □

End of lecture 1. October 20, 2015

Theorem 1.1 from the previous lecture states that one cannot assign a volume to every subset of $\mathbb{R}$ if we require that it is non-zero and finite for the unit ball, translation invariant and countably additive on disjoint sets. We remark that it generalizes to higher dimensions:

Theorem 1.3. *There does not exist $\mu : \mathcal{P}(\mathbb{R}^d) \to [0, \infty]$ such that the following holds*

1. *(Translation invariance) $\mu(E + y) = \mu(y)$ for all $E \in \mathcal{P}(\mathbb{R}), y \in \mathbb{R}$, where $E + y := \{x + y : x \in E\}$*
2. $0 < \mu(B_1(0)) < \infty$
3. *(Countable additivity) If $E_j, j \in \mathbb{N}$ are pairwise disjoint, then*

$$\mu\Big(\bigcup_{i=1}^{\infty} E_i\Big) = \sum_{j=1}^{\infty} \mu(E_i)$$

Therefore we would like to relax some of the assumptions. Recall that we would like to have additivity of the volume since this would be related to linearity of the integral. Replacing countable with finite additivity is not sufficient if the volume should also be rotation invariant, as seen in the last lecture.
It seems likely that one would have to restrict the domain of μ and define the volume just for certain subsets of $\mathbb{R}^d$. However, our strategy it to first consider only countably subadditive maps on $\mathcal{P}(\mathbb{R}^d)$ which are called *outer measures.* Later we shall state a condition on the subsets of $\mathbb{R}^d$ on which these subadditive maps are actually additive.

1.1 Outer measure

Definition 1.4. Let X be a set. A map $\mu : \mathcal{P}(X) \to [0, \infty]$ is called an *outer measure*, if

1. $\mu(\emptyset) = 0$
2. (Monotonicity) If $E_1, E_2 \subset X$, then $\mu(E_1) \leq \mu(E_2)$.
3. (Countable subadditivity) For any $E_j \subset X$, $j \in \mathbb{N}$ we have

$$\mu(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu(E_i).$$

In our applications, X will typically be $\mathbb{R}^d$. In the previous lecture we showed that 1.-3. follow provided μ is countably additive on disjoint sets. Since now we assume only countable subadditivity of μ, 1. and 2. do not follow. To avoid possible pathological examples we now additionally assume these two properties.
How to define an outer measure for any subset of $\mathbb{R}^d$? The power set $\mathcal{P}(\mathbb{R}^d)$ has the cardinality larger than continuum! The idea is to assign a certain concrete quantity to the sets in some smaller subcollection of $\mathcal{P}(\mathbb{R}^d)$. We call these sets *generating sets*. Then one abstractly defines an outer measure of any set in $\mathbb{R}^d$ by covering this set with generating sets. Let us make this construction precise.

Theorem 1.5. *Let X be a set. Let $\mathcal{T} \subset \mathcal{P}(X)$ and $\tau : \mathcal{T} \to [0, \infty]$. The map $\mu : \mathcal{P}(X) \to [0, \infty]$ given by*

$$\mu(E) = \inf_{\substack{\mathcal{T}' \subset \mathcal{T} \\ \cup_{T \in \mathcal{T}'} T \supset E}} \sum_{\mathcal{T}'} \tau(T)$$

is an outer measure.

If $X = \mathbb{R}^d$, we may take $\mathcal{T}$ to be the collection of all balls B with rational radii r and rational centers. We assign them the quantity $\tau(B) = r^d$, which up to a constant coincides with our naive notion of the volume of a ball. The collection of all such balls is countable, which is much less than the cardinality of $\mathcal{P}(\mathbb{R})$. Then we generate the outer measure of any set by covering it with balls. Since the collection of balls with rational radii is translation and rotation invariant, this implies translation and rotation invariance of the generated outer measure.

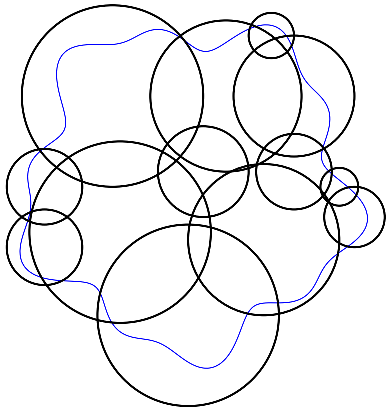


Figure 2: Covering a set with balls.

Proof. We need to check the three defining properties of outer measures.

1. The empty collection $\mathcal{T} = \emptyset$ covers the empty set and $\sum_{T\in\emptyset} \tau(T) = 0$, so $\mu(\emptyset) = 0$.

2. Let $E_1 \subset E_2$ and assume that $\mu(E_2) < \mu(E_1)$. Then there exists $\mathcal{T}'$ such that $E_2 \subset \bigcup_{T\in\mathcal{T}'} T$ and $\sum_{T\in\mathcal{T}'} \tau(T) < \mu(E_1)$. This contradicts $E_1 \subset E_2 \subset \bigcup_{T\in\mathcal{T}'} T$.

3. We need to show that for any $\varepsilon > 0$ we have
$$\mu(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \mu(E_i) + \varepsilon.$$
Let $\varepsilon > 0$. By definition of inf we may choose a collection $\mathcal{T}_i$ such that $\bigcup_{T\in\mathcal{T}_i} T \supset E_i$ and
$$\sum_{T\in\mathcal{T}_i} \tau(T) \leq \mu(E_i) + \frac{\varepsilon}{2^i}$$
Then $\bigcup_{i=1}^{\infty} E_i \subset \bigcup_{i=1}^{\infty} \bigcup_{T\in\mathcal{T}_i} T$ and we have
$$\mu(\bigcup_{i=1}^{\infty} E_i) \leq \sum_{i=1}^{\infty} \bigcup_{\mathcal{T}\in T_i} \tau(T) \leq \sum_{i=1}^{\infty} \mu(E_i) + \varepsilon \sum_{i=1}^{\infty} \frac{1}{2^i} \leq \sum_{i=1}^{\infty} \mu(E_i) + \varepsilon.$$
Observe that having only countably many sets E_i is crucial, as otherwise we would not get a summable geometric series $\sum_{i=1}^{\infty} 2^{-i}$.

$\square$

Note that in this theorem we may choose τ freely. However, it is not clear whether the generated μ is then a reasonable quantity. Note that for $E \in \mathcal{T}$ we know that $\mu(E) \leq \tau(E)$ since the set is its own cover. It seems reasonable that this should be the best open cover and that we should have $\mu(E) = \tau(E)$ for $E \in \mathcal{T}$. However, this is not true in general. In principle it could happen that $\mu(E) = 0$ for all $E \in \mathcal{T}$. But in many examples in practice, μ and τ agree on generating sets.

Example. Let $X = \mathbb{R}^d$ and let $\mathcal{T}$ be the set of all dyadic cubes in $\mathbb{R}^d$, i.e.

$$\mathcal{T} = \{Q = [2^k n_1, 2^k(n_1+1) \times \cdots \times [2^k n_d, 2^k(n_d+1)) : k, n_1, \ldots, n_d \in \mathbb{Z}\}.$$

The number k is called the *scale* of Q. An example of a dyadic cube of scale 0 is the unit cube $[0,1) \times \cdots \times [0,1)$. Since the sides of a dyadic cube are half open intervals, any two dyadic cubes are either disjoint or one is contained in the other. Each dyadic cube of scale k partitions into 2^d dyadic cubes of scale $k-1$, called the *children* of the cube.

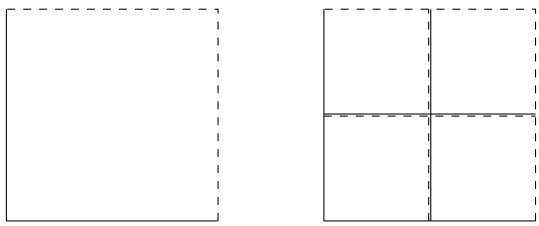


Figure 3: A dyadic square in $\mathbb{R}^2$ and its children.

For $Q \in \mathcal{T}$ we set

$$\tau(Q) = 2^{kd}$$

which coincides with our naive interpretation of the volume of Q. The outer measure μ generated by τ via coverings as is called the *Lebesgue outer measure* on $\mathbb{R}^d$.

Theorem 1.6. *Let everything be as in the previous example. Then for each dyadic cube $Q \in \mathcal{T}$ we have $\mu(Q) = \tau(Q)$.*

Therefore the Lebesgue outer measure is indeed a "reasonable" quantity. This is a very important result and showing it requires some work. If we generate the outer measure by arbitrary cubes or balls and define $\tau(Q)$ analogously (i.e. as their "volume"), then μ also agrees with τ on generating sets. However, the proof of this fact is more technical than in the dyadic case.

Proof. We have $\mu(Q) \leq \tau(Q)$ since Q covers itself. So it remains to show that if $Q \in \mathcal{T}$ and $\mathcal{T}' \subset \mathcal{T}$ with $\bigcup_{Q' \in \mathcal{T}'} Q' \supset Q$, then

$$\tau(Q) \leq \sum_{Q' \in \mathcal{T}'} \tau(Q').$$

First we consider the special case when $\mathcal{T}'$ is finite. Let $Q \in \mathcal{T}$ be of scale k. Without loss of generality we may remove the cubes from $\mathcal{T}'$ which do not intersect Q. That is, we may assume that for each $Q \in \mathcal{T}'$ we have $Q \cap Q' \neq \emptyset$. Denote by $k_{\min}$ be the smallest scale of the cubes in $\mathcal{T}'$. If $k_{\min} > k$ we have

$$\tau(Q) \leq \tau(Q_{\min}) \leq \sum_{Q' \in \mathcal{T}'} \tau(Q')$$

for any cube $Q_{\min}$ of scale $k_{\min}$, which shows the claim. So we may assume that $k_{\min} \leq k$.

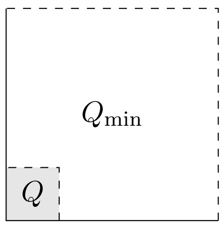


Figure 4: $k < k_{\min}$

Now we induct on $k - k_{\min} \geq 0$. If $k - k_{\min} = 0$, i.e. $k = k_{\min}$, we are done by the same reasoning as above. Assume now that $k - k_{\min} > 0$. Denote by $k_{\max}$ the largest scale of the cubes in $\mathcal{T}'$. Without loss of generality we may assume $k_{\max} < k$, otherwise we reason as above and get $\tau(Q) \leq \tau(Q_{\max}) \leq \sum_{Q' \in \mathcal{T}'} \tau(Q')$. In other words, we may assume that the scale of any cube in $\mathcal{T}'$ is strictly smaller than k.
The cube Q partitions into 2^d dyadic children Q_j of scale $k-1$, i.e.

$$Q = \bigcup_{j=1}^{2^d} Q_j.$$

Since any two dyadic cubes are either disjoint or one is contained in the other, every $Q' \in \mathcal{T}'$ is contained exactly one of the children Q_j.[2]

[2]Here we crucially use the dyadic structure. If we had an arbitrary collection of cubes, we would not be able to draw the same conclusion.

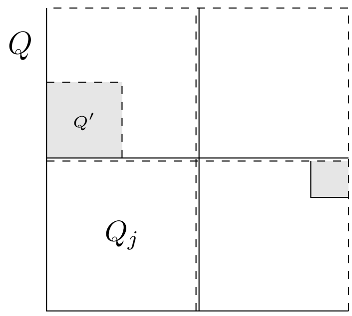


Figure 5: Covering Q with $Q' \in \mathcal{T}'$.

Set now

$$\mathcal{T}_j := \{Q' \in \mathcal{T} : Q' \subset Q_j\}.$$

We have $Q_j \subset \bigcup_{Q' \in \mathcal{T}_j} Q'$, which follows by a similar dyadic argument. The cubes Q_j are of scale $k-1$, so by induction

$$\tau(Q_j) \leq \sum_{Q' \in \mathcal{T}_j} \tau(Q').$$

Together with $\tau(Q) = 2^{dk} = 2^d 2^{d(k-1)} = \sum_{j=1}^{2^d} \tau(Q_j)$ we may then estimate

$$\tau(Q) = \sum_{j=1}^{2^d} \tau(Q_j) \leq \sum_{j=1}^{2^d} \sum_{Q' \in \mathcal{T}_j} \tau(Q') \leq \sum_{Q' \in \mathcal{T}} \tau(Q).$$

For the last inequality we used disjointness of the collections $\mathcal{T}_j$ for different j. This shows the claim for finite $\mathcal{T}'$.

For a general collection $\mathcal{T}'$ we argue by compactness. If Q were compact and the cubes in $\mathcal{T}'$ open, then we would have an open covering of a compact set. So there would exist a finite subcovering of Q and we could apply the argument for finite collections. Dyadic cubes are half open, so this is not the case. However, we can approximate dyadic cubes in question by open/closed sets, respectively, and obtain the desired bound within a factor $(1+\varepsilon)^2$. More precisely, we show the following: for every $\varepsilon > 0$ we have

$$\tau(Q) \leq (1+\varepsilon)^2 \sum_{Q' \in \mathcal{T}'} \tau(Q').$$

To show this pick a number L which is large enough. We consider the $L-$th generation of Q consisting of cubes of scale $k-L$. They partitions Q, so

$$Q = \bigcup_{j=1}^{2^{dL}} Q_j.$$

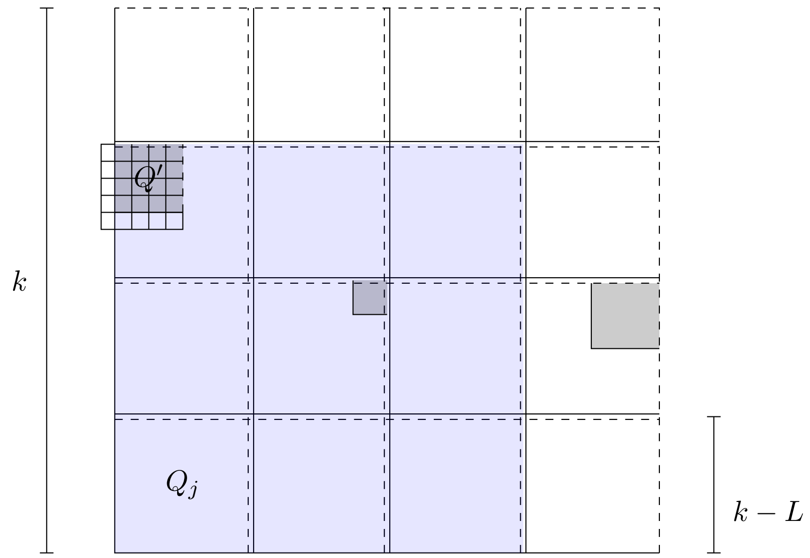


Figure 6: Partitioning of Q and covering it by open sets.

Denote by $\widetilde{\mathcal{T}}$ the set of all cubes Q_j with $\overline{Q_j} \subset Q$. (In Figure 6 this would be the dyadic cubes contained in the light blue cube.) We have

$$\sum_{Q_j \in \widetilde{\mathcal{T}}} \tau(Q_j) = 2^{dk} - d2^{(d-1)L} 2^{d(k-L)} \geq \tau(Q) \frac{1}{1+\varepsilon}$$

where the last inequality holds provided L is sufficiently large.[3]
For the sets in our covering we use a similar argument to pass to open sets. For every $Q' \in \mathcal{T}'$ of scale k' define $\mathcal{T}_{Q'}$ to be the set of all Q'' of scale $k' - L$ such that $\overline{Q''} \cap Q' \neq \emptyset$. (In Figure 6 these would be the small cubes in and around Q'.) Then

$$\sum_{Q'' \in \mathcal{T}_{Q'}} \tau(Q'') \leq \tau(Q')(1+\varepsilon)$$

provided L is large enough. The proof of this is similar as before using the definition of τ and we leave it to the reader. Now we have

$$K := \overline{\bigcup_{\widetilde{Q} \in \widetilde{\mathcal{T}}} \widetilde{Q}} \subset Q \subset \bigcup_{Q' \in \mathcal{T}'} Q' \subset \bigcup_{Q' \in \mathcal{T}'} \operatorname{int}\Big(\bigcup_{Q'' \in \mathcal{T}_{Q'}} Q'' \Big)$$

The set K is closed and bounded and hence compact, while on the right hand-side we have an open covering of K indexed by the cubes in $\mathcal{T}'$. By

[3]Note that for the argument we used only the definition of τ and no additivity of volume.

compactness there exists a finite subcollection $\mathcal{T}'' \subset \mathcal{T}'$ which already covers K. Hence

$$\bigcup_{\widetilde{Q}\in\widetilde{\mathcal{T}}} \widetilde{Q} \subset \bigcup_{Q'\in\mathcal{T}''} \bigcup_{Q''\in\mathcal{T}_{Q'}} Q''$$

Putting everything together and using the claim for finite collections $\mathcal{T}''$ we obtain

$$\frac{\tau(Q)}{1+\varepsilon} \leq \sum_{\widetilde{Q}\in\widetilde{\mathcal{T}}} \tau(\widetilde{Q}) \leq \sum_{Q'\in\mathcal{T}''} \sum_{Q''\in\mathcal{T}_{Q'}} \tau(Q'') \leq \sum_{Q'\in\mathcal{T}'} \tau(Q')(1+\varepsilon).$$

Therefore, for every $\varepsilon > 0$ we have

$$\tau(Q) \leq \sum_{Q'\in\mathcal{T}'} \tau(Q')(1+\varepsilon)^2.$$

This establishes the claim. $\square$

It is important to note that the fact $\mu(Q) = \tau(Q)$ for dyadic cubes Q relies on our particular choice of τ. If we defined $\tau(Q) = 2^{k\alpha}$ with $\alpha > d$, then it would be more efficient to cover Q with smaller cubes. One can show that in this case $\mu(Q) = 0$ for all dyadic cubes Q.

The Lebesgue outer measure $\mu : \mathcal{P}(\mathbb{R}^d) \to [0,\infty]$ satisfies

1. Translation invariance
2. $0 < \mu(B_1(0)) < \infty$
3. Countable subadditivity

The property 2. holds by monotonicity of μ, since the unit ball contains a dyadic cube and is also contained in a dyadic cube. Translation invariance will be proved later. For the outer measure generated by dyadic cubes translation invariance is not obvious, since the collection of dyadic cubes is not invariant under arbitrary translations.

1.2 Measurable sets

To build a linear integration theory we would like to have countable additivity of μ rather than subadditivity. For this we restrict ourselves to a certain class of sets for which an outer measure is countably additive.

Definition 1.7 (Caratheodory)**.** Let X be a set and μ an outer measure on X. A set $E \subset X$ is called *measurable*, if for all $F \subset X$

$$\mu(F) = \mu(F \cap E) + \mu(F \cap E^c). \tag{5}$$

The intuition behind this definition is that inequalities " $\leq$ " for E translate into inequalities " $\geq$ " for E^c, and both together can be used to imply equalities. To work out this intuition in detail, one needs the Caratheodory condition.
Note that

$$\mu(F) \leq \mu(F \cap E) + \mu(F \cap E^c)$$

is always true by subadditivity of μ (applied to $F = (F \cap E) \cup (F \cap E^c)$). Thus, a set is measurable, if we have additivity of μ on these special sets.
In general outer measures need not have many measurable sets. However, given some measurable sets, the following theorem shows how to construct new measurable sets using finitary operations.

Theorem 1.8. *Let E_1, E_2 be measurable. Then*

1. *$\emptyset$ is measurable.*
2. *E_1^c is measurable.*
3. *$E_1 \cup E_2$ is measurable.*
4. *$E_1 \cap E_2$ is measurable.*
5. *$E_1 \setminus E_2$ is measurable.*

Proof. By subadditivity we only need to show the inequality " $\geq$ " in (5).

1. We have $\mu(F) = 0 + \mu(F) = \mu(\emptyset) + \mu(F)$.
2. Clear by symmetry.
3. Since E_1 is measurable, we have

   $$\mu(F) = \mu(F \cap E_1) + \mu(F \cap E_1^c)$$

   Since E_2 is measurable, this equals

   $$\mu(F \cap E_1 \cap E_2) + \mu(F \cap E_1 \cap E_2^c) + \mu(F \cap E_1^c \cap E_2) + \mu(F \cap E_1^c \cap E_2^c)$$

   Note that $E_1 \cup E_2 = (E_1 \cap E_2) \cup (E_1 \cap E_2^c) \cup (E_1^c \cap E_2)$.

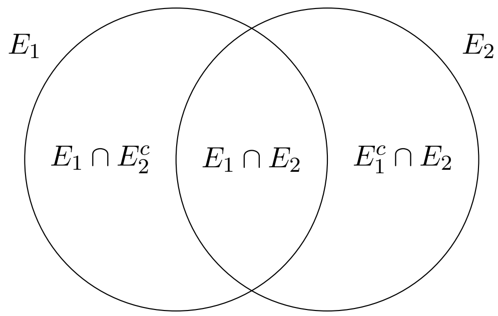


Figure 7: Partitioning of $E_1 \cup E_2$.

By subadditivity of μ we then have

$$\mu(F \cap E_1 \cap E_2) + \mu(F \cap E_1 \cap E_2^c) + \mu(F \cap E_1^c \cap E_2) \geq \mu(E_1 \cup E_2).$$

Together with $F \cap E_1^c \cap E_2^c = F \cap (E_1 \cup E_2)^c$ this shows

$$\mu(F) \geq \mu(F \cap (E_1 \cup E_2)) + \mu(F \cap (E_1 \cup E_2)^c)$$

for any set $F \subset X$.

4. We write
$$E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$$
and use 2., 3. and again 2.

5. We write $E_1 \setminus E_2 = E_1 \cap E_2^c$ and use 2. and 4.

□

By induction, 3. and 4. easily generalize to finite unions and intersections. Our next theorem states a countable version of 3. and shows that (5) is indeed sufficient for additivity on any countable collection of pairwise disjoint measurable sets.

Theorem 1.9. *Let E_j be pairwise disjoint measurable sets. Then $\bigcup_{j=1}^{\infty} E_j$ is measurable and*

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

Proof. Special case 1: Two sets E_1, E_2. We have already seen that $E_1 \cup E_2$ is measurable. Furthermore we have

$$\mu(E_1 \cup E_2) \overset{(*)}{=} \mu(E_1 \cap (E_1 \cup E_2)) + \mu(E_1^c \cap (E_1 \cup E_2)) \overset{(**)}{=} \mu(E_1) + \mu(E_2)$$

which shows the claim in this case. We used $(*)$: E_1 measurable $(**)$: $E_2 \subset E_1^c$.
Special case 2: Finitely many sets $E_1, \ldots, E_n$. By induction we show that $G_n = \bigcup_{j=1}^n E_j$ is measurable and that

$$\mu(G_n) = \mu\Big(\bigcup_{j=1}^{\infty} E_j\Big) = \sum_{j=1}^{n} \mu(E_j)$$

(exercise, write $G_{n+1} = G_n \cup E_{n+1}$).
Case 3. Countably many sets E_j, $j \in \mathbb{N}$. Denote $G = \cup_{n=1}^{\infty} G_n$. We have

$$\begin{aligned}
\mu(F) &\overset{(*)}{=} \mu(F \cap G_n) + \mu(F \cap G_n^c) \\
&\overset{(**)}{=} \sum_{j=1}^{n} \mu(F \cap E_j) + \mu(F \cap G_n^c) \\
&\overset{(***)}{\geq} \sum_{j=1}^{n} \mu(F \cap E_j) + \mu(F \cap G^c)
\end{aligned}$$

where we used $(*)$: measurability of G_n. $(**)$: finite additivity. $(***)$: $G_n \subset G$ and monotonicity. In the limit we obtain

$$\mu(F) \geq \sum_{j=1}^{\infty} \mu(F \cap E_j) + \mu(F \cap G^c)$$

By subadditivity of μ we estimate this further as

$$\mu(F) \geq \sum_{j=1}^{\infty} \mu(F \cap E_j) + \mu(F \cap G^c) \overset{1.}{\geq} \mu(F \cap G) + \mu(F \cap G^c) \overset{2.}{\geq} \mu(F).$$

Since the left and the right hand-side of the last display coincide, all inequalities must be equalities. In particular, equality in 2. shows that G is measurable. Equality in 1. then implies that (setting $F = G$ and using $G \cap E_j = E_j$, $G \cap G = G$, $G \cap G^c = \emptyset$)

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

This shows countable additivity of μ. $\square$

The last two theorems described how to construct measurable sets from existing measurable sets. However, we still do not know if there are *any* measurable sets at all. If $X = \mathbb{R}^d$ and the measure is generated by the dyadic cubes (*the Lebesgue measure*), the answer is affirmative.[4]

Theorem 1.10. *Dyadic cubes are Lebesgue measurable.*

This theorem is a consequence of the following lemma, which states that it suffices to verify the Caratheodory condition (5) on generating sets.

Lemma 1.11. *Let μ on $\mathcal{P}(X)$ be generated by $\mathcal{T}, \tau$. A set E is measurable if and only if for each $T \in \mathcal{T}$ we have*

$$\mu(T) = \mu(T \cap E) + \mu(T \cap E^c).$$

Proof. We only need to show ($\Leftarrow$). Assume that for each $T \in \mathcal{T}$ we have

$$\mu(T) = \mu(T \cap E) + \mu(T \cap E^c).$$

Let $F \subset X$. We need to show that for any $\varepsilon > 0$ we have

$$\varepsilon + \mu(F) \geq \mu(F \cap E) + \mu(F \cap E^c).$$

If $\mu(F) = \infty$ there is nothing to show. So assume that $\mu(F) \leq \infty$. Let $\varepsilon > 0$ and pick $\mathcal{T}' \subset T$ with $F \subset \bigcup_{T \in \mathcal{T}'} T$ and

$$\varepsilon + \mu(F) \geq \sum_{T \in \mathcal{T}'} T$$

Since $F \cap E \subset \bigcup_{T \in \mathcal{T}'} E \cap T$, by monotonicity and subadditivity we have

$$\mu(F \cap E) \leq \sum_{T \in \mathcal{T}'} \mu(E \cap T)$$

In the same way we show

$$\mu(F \cap E^c) \leq \sum_{T \in \mathcal{T}'} \mu(E^c \cap T)$$

Summing the last two displays we obtain

$$\begin{aligned} \mu(F \cap E) + \mu(F \cap E^c) &\leq \sum_{T \in \mathcal{T}'} \mu(E \cap T) + \mu(E^c \cap T) \\ &= \sum_{T \in \mathcal{T}'} \mu(T) \leq \mu(F) + \varepsilon \end{aligned}$$

as desired. $\square$

[4] This is also true if the generating sets are balls or arbitrary cubes, but the proof is the easiest in the dyadic case.

Since any two dyadic cubes are either disjoint or one is contained in the other, measurability of dyadic cubes follows from this characterization.

To see Theorem 1.10, it suffcies to show for any two dyadic cubes T amd T'

$$\mu(T) = \mu(T \cap T') + \mu(T \cap (T')^c) \ .$$

This is immediate if $\mu(T') \geq \mu(T)$ since then one of the sets on the right hand side is empty. If $\mu(T') < \mu(T)$, then one may argue by induction on the difference in scales between T and T', applying the inductive hypothesis to the children of T, at most one of which may intersect T'.
Observe that Theorem 1.10 implies that all open sets in $\mathbb{R}^d$ are measurable since we can write them as a countable union of dyadic cubes. Closed sets are also measurable, since they are complements of open sets.

End of lectures 2 and 3. October 27 and 29, 2015

So far we have been working with the space $X = \mathbb{R}^n$ and the space of all dyadic cubes $\mathcal{T}$. Then, on the set of dyadic cubes, we defined

$$\tau\colon \mathcal{T} \to [0,\infty], \ \ \tau(Q) = 2^{dk},$$

and used the above to generate the Lebesgue outer measure μ on the power set $\mathcal{P}(X)$ via

$$\mu(E) = \inf_{\substack{\mathcal{T}'\subset\mathcal{T} \\ E\subset\bigcup_{\mathcal{T}'} Q}} \sum_{\mathcal{T}'} \tau(Q).$$

We could have as well taken all cubes or balls instead of dyadic cubes, however this would make the combinatorics more difficult. This time we are going to state an abstract theorem that lets us to validate that the outer measures generated in a number of ways are again the Lebesgue outer measure μ.

Theorem 1.12. *Let X be a set, $\mathcal{T}_i \subset \mathcal{P}(X)$, $\tau_i\colon \mathcal{T}_i \to [0,\infty]$ and let μ_i be the outer measure generated from τ_i for $i = 1, 2$. If for all $T \in \mathcal{T}_1$, $\tau_1(T) \geq \mu_2(T)$ and for all $T \in T_2$, $\tau_2(T) \geq \mu_1(T)$, then we have $\mu_1 = \mu_2$.*

Proof. We show $\mu_1 \geq \mu_2$ and $\mu_1 \leq \mu_2$.
Let $E \subset X$. We obtain

$$\mu_1(E) \overset{\text{Definition}}{=} \inf_{\substack{\mathcal{T}_1'\subset\mathcal{T} \\ E\subset\bigcup_{\mathcal{T}'} Q}} \sum_{\mathcal{T}_1} \tau_1(T) \overset{\text{Assumption}}{\geq} \inf_{\substack{\mathcal{T}_1'\subset\mathcal{T} \\ E\subset\bigcup_{\mathcal{T}'} Q}} \sum_{\mathcal{T}_1} \mu_2(T) \overset{\text{Subadditivity}}{\geq} \mu_2(E),$$

where in the first equality we used the definition of μ_1, in the first inequality we used the assumption $\tau_1 \geq \mu_2$ and the subadditivity of μ_2 in the second. Symmetrically we get that $\mu_2(E) \leq \mu_1(E)$. $\square$

Now we apply the theorem we have just proven to show that the outer measures generated by the set of all axe parallel rectangular boxes and the dyadic cubes are equal.
Let

$$\mathcal{T}_1 = \text{the family of dyadic cubes}, \quad \tau_1(Q) = 2^{dk}, \text{ where } k \text{ is the order of Q}$$

$$\mathcal{T}_2 = \{[a_1, b_1) \times ... \times [a_d, b_d) \colon a_i < b_i \text{ for } 1 \leq j \leq d\}, \quad \tau_2(Q) = \prod_{j=1}^{d}(b_j - a_j).$$

First of all, note that $T_1 \subset T_2$ and essentially "for free" we obtain that

$$\tau_1(Q) = 2^{dk} = \tau_2(Q) \overset{Q \in \mathcal{T}_2}{\geq} \mu_2(Q).$$

We are left with proving the reverse inequality, which follows from the next proposition.

Proposition 1.13. *For all $\epsilon > 0$ and for all $Q \in \mathcal{T}_2$*

$$(1+\epsilon)^d \tau_2(Q) \geq \mu_1(Q).$$

Proof. The idea of the proof is to effectively approximate any box Q by dyadic cubes, one can see the two dimensional situation in the picture 1.2. Choose k small enough so that

$$\frac{2}{\epsilon} 2^k < \min_j (b_j - a_j).$$

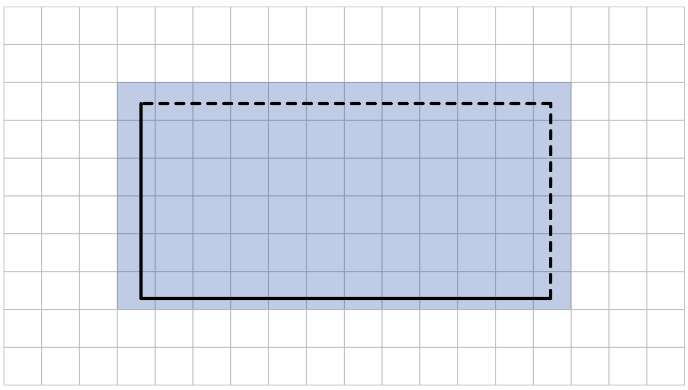


Figure 8: Covering of Q by small dyadic cubes in $\mathbb{R}^2$

Define

$$N_j = \{n \in \mathbb{Z}\colon [2^k n, 2^k(n+1)) \cap [a_j, b_j) \neq \emptyset\},$$

so the cardinality $\#N_j$ is the number of dyadic intervals of order k, which cut Q in the j-th coordinate. Moreover, observe that

$$Q \subset \bigcup_{n_1 \in N_1, \ldots, n_d \in N_d} [2^k n_1, 2^k(n+1)) \times \ldots \times [2^k n_d, 2^k(n_d+1)),$$

so

$$\mu_1(Q) \leq 2^{kd} \prod_{j=1}^{d} \#N_j.$$

Denote by l_j and m_j the minimal number and the maximal number in N_j respectively. Note that

$$2^k(l_j+1) \geq a_j,$$
$$2^k m_j \leq b_j.$$

Substracting the second inequality from the first we obtain

$$2^k(m_j - l_j - 1) \geq b_j - a_j \implies 2^k(\#N_j - 2) \leq b_j - a_j.$$

Recall, we defined ϵ in such a way that

$$2^k \#N_j \leq (b_j - a_j)(1+\epsilon)$$

Putting the facts together we have

$$\mu_1(Q) \leq 2^{kd} \prod_{j=1}^{d} \#N_j \leq 2^{kd} 2^{-kd} \prod_{j=1}^{d} (b_j - a_j)(1+\epsilon)^d = \tau_2(Q)(1+\epsilon)^d.$$

This finishes the proof of the proposition. □

Because ϵ in the proposition is arbitrarily small we have the following inequality $\tau_2(Q) \geq \mu_1(Q)$, what implies the equality of outer measures generated by τ_1 and τ_2, i.e. $\mu_1 = \mu_2$.
Now we shall prove that the Lebesgue outer measure is invariant under dilations. Let $\Lambda \in \mathbb{R}^{d \times d}$ be a fixed diagonal matrix with positive entries, in particular we have $\det(\Lambda) > 0$. For a set $E \subset \mathbb{R}^d$ define

$$\Lambda E = \{\Lambda x\colon x \in E\}.$$

Let $\mathcal{T}_2$ and τ_2 be as before and let

$$\mathcal{T}_3 = \{\Lambda Q\colon Q \in \mathcal{T}_2\}.$$

Note that the map $Q \mapsto \Lambda Q$ is a bijection between $\mathcal{T}_2$ and $\mathcal{T}_3$, where the inverse is given by $Q \mapsto \Lambda^{-1}Q$. This means that any element of $\mathcal{T}_3$ can be uniquely represented as ΛQ. For any $Q \in \mathcal{T}_2$ put

$$\tau_3(\Lambda Q) = \det(\Lambda)\tau_2(Q)$$

and observe that if $\mathcal{T}_2'$ covers E, then $\mathcal{T}_3' = \{\Lambda Q\colon Q \in \mathcal{T}_2'\}$ covers ΛE and $\mathcal{T}_2 = \mathcal{T}_3$.

Proposition 1.14. *For $E \subset \mathbb{R}^d$, $\mu_3(\Lambda E) = \det(\Lambda)\mu_2(E)$.*

Proof. For any $Q = [a_1, b_1) \times ... \times [a_d, b_d) \in \mathcal{T}_2 = \mathcal{T}_3$

$$\tau_2(\Lambda Q) = \prod_{j=1}^{d}(\lambda_j b_j - \lambda_j a_j) = \det(\Lambda)\prod_{j=1}^{d}(b_j - a_j) = \det(\Lambda)\tau_2(Q) = \tau_3(\Lambda Q),$$

so $\mu_2(\Lambda\cdot) = \mu_3(\Lambda\cdot) = \det(\Lambda)\mu_2(\cdot)$. $\square$

Observe, we can similarly argue that the Lebesgue measure is invariant under translation, i.e. that for $E \subset \mathbb{R}^n$ and $y \in \mathbb{R}^n$ it satisfies the relation

$$\mu(E + y) = \mu(E).$$

Let us proceed with showing that the outer measure generated by balls and the Lebesgue measure are in fact the same. In the following we denote by τ the measure generating function defined on all boxes and by μ the Lebesgue outer measure. First of all we shall make sure that μ of the unit (and as a consequence any) ball, denoted by $B_1(0)$, is strictly positive and finite. Note that $[0, 1/d)^d \subset B_1(0) \subset [-1, 1)^d$, so by the monotonicity of the outer measure

$$0 < \mu([0, 1/d)^d \leq \mu(B_1(0)) \leq \mu([-1, 1)^d) < \infty$$

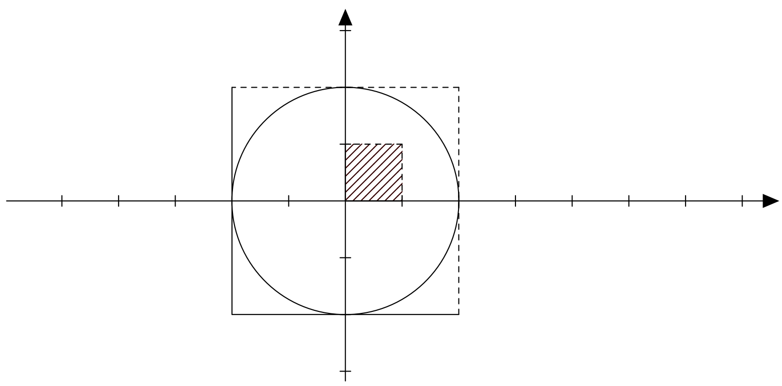


Figure 9: $[0, 1/d)^d \subset B_1(0) \subset [-1, 1)^d$ for $d = 2$

Put $c := \mu(B_1(0))$ and let

$$\mathcal{T}_4 = \{B_r(x) \colon x \in \mathbb{R}^d,\ r \in (0,\infty)\},\quad \tau_4(B_r(x)) = cr^d.$$

Proposition 1.15. *We have $\mu = \mu_4$, where μ is the Lebesgue outer measure and μ_4 is the outer measure generated by τ_4.*

Proof. We shall prove that $\tau \geq \mu_4$ on $\mathcal{T}$ and $\mu \leq \tau_4$. Because of the translation and dilation invariance of the Lebesgue outer measure we obtain

$$\mu(B_r(x)) = cr^d = \tau_4(B_r(x)).$$

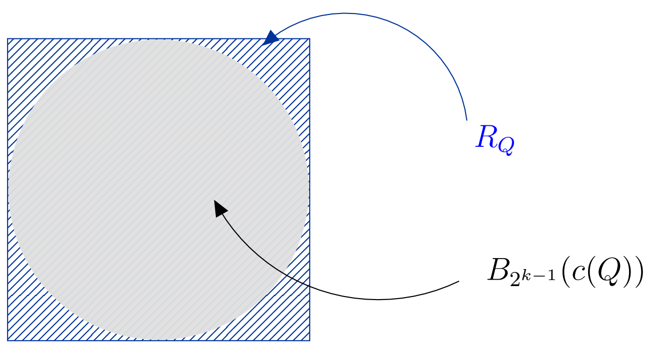
This gives the "$\geq$" inequality. The inequality in the other direction is a bit more difficult and will follow from the next proposition.

Proposition 1.16. *For all $\epsilon > 0$ we have*

$$\mu_4([0,1)^d) \geq (1+\epsilon)\tau([0,1)^d).$$

Proof. The point is to cover $[0,1)^d$ by balls efficiently. For a dyadic cube $Q \in \mathcal{T}$ of order k define the "frames" (where $c(Q)$ denotes the center of Q)

$$R_Q = Q \setminus B_{2^{k-1}}(c(Q))$$



We will now show that for any dyadic cube Q there exists a disjoint decomposition such that

$$R_Q = R_Q^1 \cup R_Q^2,$$

with $\mu(R_Q^1) < \delta\mu(Q)$ and R_Q^2 is a disjoint union of dyadic cubes. Let k be the order of Q and let $k' < k$ be small enough, we will specify it later. Define

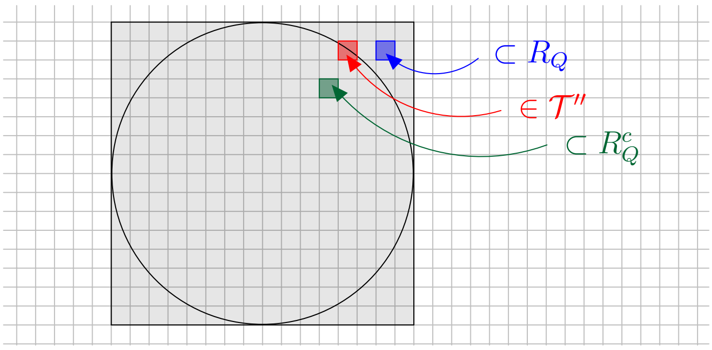
$$\mathcal{T}' = \{Q' \subset Q\colon \text{ the order of } Q' = k'\}.$$

Observe that we can decompose Q into disjoint subsets as follows

$$Q = \bigcup_{Q' \in \mathcal{T}', Q' \subset Q} Q' = \overbrace{\bigcup_{\substack{Q' \in \mathcal{T}', Q' \subset Q \\ Q' \subset R_Q}} Q'}^{R_Q^2} \cup \bigcup_{\substack{Q' \in \mathcal{T}', Q' \subset Q \\ Q' \subset R_Q^c}} Q' \cup \bigcup_{\mathcal{T}''} Q',$$

where

$$\mathcal{T}'' = \{Q' \in \mathcal{T}'\colon Q' \subset Q,\ Q' \not\subset R_Q,\ Q' \not\subset R_Q^c\}.$$



Lemma 1.17. *Let Q be the unit cube. For each $Q' \in \mathcal{T}''$ there exists $1 \leq j \leq d$ such that there are at most $8d^2$ elements of $\mathcal{T}''$ which differ from Q' only in the j-th coordinate.*

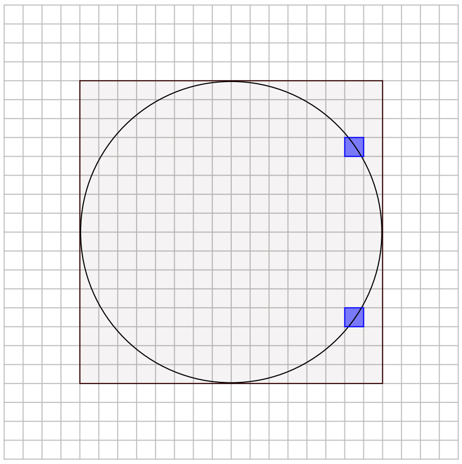


Figure 10: The situation in $\mathbb{R}^2$ - the two blue cubes are elements of $\mathcal{T}''$ which differ from each other only in the second coordinate.

Proof. Let $x' \in Q'$ with $\sum_{j=1}^{d}(x'_j)^2 = 1$. Choose $1 \leq j \leq d$ such that $|x'_j| > 1/d$. Let $Q'' \in \mathcal{T}''$ differ from Q' only in the j-th coordinate. There also exists $x'' \in Q''$ with $\sum_{j=1}^{d}(x''_j)^2 = 1$. Now observe that we have

$$|x'_j - x''_j| \leq (2d^2-1)2^{k'} \text{ or } |x'_j + x''_j| \leq (2d^2-1)2^{k'}.$$

Assume on the contrary that, for example, the left hand side of the alternative above does not hold. We have the equality

$$|\sum_{i \neq j}(x'_i)^2 - (x''_i)^2| = |(x'_j)^2 - (x''_j)^2|,$$

which would give by our assumption

$$(d-1)2^{k'} \cdot 2 \cdot 2^k \geq |\sum_{i \neq j}(x'_i - x''_i)(x'_i + x''_i)| = |(x'_j - x''_j)(x'_j + x''_j)| \geq (2d^2-1)2^{k'}\frac{1}{d}.$$

This would mean that $(d-1)2^{k'} \cdot 2 \cdot 2^k \geq (2d^2-1)2^{k'} \cdot (1/d)$, what is a contradiction.

□

□

With a use of the fact that we have just proven we can decompose $\mathcal{T}''$ into pairwise disjoint collections $\mathcal{T}_j$ of the cubes which differ from each other only in the j-th coordinate. Hence

$$\mathcal{T}'' = \bigcup_{j=1}^{d} \mathcal{T}_j$$

and for each j we can bound the cardinality of $\mathcal{T}_j$ by $8d^2$. This gives at most $8d^3$ cubes in $\mathcal{T}''$ and each cube has measure equal to $2^{dk'}$. Therefore the cubes in $\mathcal{T}''$ have total measure not greater than $2^{dk'}8d^3$, what is $\leq \delta$ if k' is small enough. Now we can summarize what have done so far in the proof as follows.

Lemma 1.18. *Let E be a disjoint union of frames. Then we can write it as a disjoint sum*

$$E = A \cup B \cup C,$$

where $\mu(A) < \delta\mu(E)$, B is a disjoint union of balls and C is a disjoint union of frame with the property $\mu(C) \leq (1-(1/d)^d)\mu(E)$.

End of lecture 4. November 3, 2015

We use the lemma to define recursively

$$E_1 = R_Q,$$

$$E_{j+1} = C_j, \text{ where } E_j = A_j \cup B_j \cup C_j \text{ as in the lemma },$$

$$\mu(A_j) \leq \frac{\delta}{2^j}\mu(E_j), \;\; \mu(C_j) \leq (1-(1/d)^d)^j\mu(E).$$

Thus, we rewrite Q as

$$Q = \bigcup_{j=1}^{N} A_j \cup \bigcup_{j=1}^{N} B_j \cup C_N$$

and

$$\mu(\bigcup_{j=1}^{N} A_j) \leq \sum_{j=1}^{\infty} \frac{\delta}{2^j} \leq \delta,$$

$$\bigcup_{j=1}^{N} B_j = \bigcup_{i=1}^{M} B_{r_i}(x_i) \text{ is a disjoint union of balls,}$$

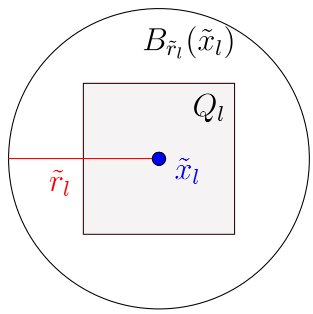
$$\mu(C_N) \leq (1-(1/d)^d)^N \leq \delta \text{ for } N \text{ big enough.}$$

Note that

$$\mu(B_{2^{k-1}}(c(Q))) + \sum_{i=1}^{M} \mu(B_{r_i}(x_i)) \leq \mu(Q)$$

The set $F = \bigcup_{j=1}^{N} A_j \cup C_N$ has small outer measure, namely $\mu(F) \leq \delta\mu(Q)$, what means that there exist cubes Q_l such that $\bigcup_{l=1}^{\infty} Q_l$ cover F and

$$\sum_{l=1}^{\infty} \mu(Q_l) \leq 3\delta\mu(Q).$$



Let $\tilde{r}_l = 2^{k_l}$, where k_l is the order of the cube Q_l and $\tilde{x}_l$ be the center of Q_l. Then of course the union of $B_{\tilde{r}_l}(\tilde{x}_l)$ covers F and because of the monotonicity and the dilation invariance its measure is small (the constant c below stands for the measure of the unit ball)

$$\mu(\bigcup_{l=1}^{\infty} B_{\tilde{r}_l}(\tilde{x}_l)) \leq \sum_{l=1}^{\infty} \mu(B_{\tilde{r}_l}(\tilde{x}_l)) \leq 3c2^d\delta\mu(Q).$$

Putting everything together, Q is covered by a union of balls

$$Q \subset B_{2^{k-1}}(c(Q)) \cup \bigcup_i B_{r_i}(x_i) \cup \bigcup_l B_{\tilde{r}_l}(\tilde{x}_l),$$

so

$$\mu_4(Q) \leq \mu(Q) + 3c\delta 2^d \mu(Q).$$

This means that for any dyadic cube Q and any $\epsilon > 0$, $\mu_4(Q) \leq (1+\epsilon)\tau(Q)$ and finishes the proof of the proposition. $\square$

As a corollary we obtain the invariance of the Lebesgue measure under orthogonal transformations (rotations and reflections).

Corollary 1.19. *Let $U \in \mathbb{R}^{d\times d}$ be a matrix satisfying $UU^T = Id$. Then we have $\mu(UE) = \mu(E)$ for all $E \subset \mathbb{R}^d$.*

Proof. We use that one can generate the Lebesgue measure via balls. This collection of sets as well as the corresponding map τ is invariant under orthogonal transformations, hence Lebesgue outer measure is invariant under orthogonal transformations. $\square$

Corollary 1.20. *Let A be a regular linear transformation of $\mathbb{R}^d$. Then it holds that $\mu(AE) = |\det A|\mu(E)$.*

Remark. Recall, one can represent a regular linear transformation A as $A = U\Lambda^{1/2}V$, where $UU^T = Id$, $VV^T = Id$ and Λ is a diagonal matrix with positive diagonal entries. This is a way to see it: notice that AA^T is a symmetric positive-definite matrix, so it can be diagonalized - let Λ be its diagonalization, it clearly has positive entries. Hence there exists an orthogonal matrix U with $AA^T = U\Lambda U^T$ and

$$\underbrace{\Lambda^{-1/2}U^T A}_{=:V}\underbrace{A^T U\Lambda^{-1/2}}_{=:V^T} = Id \implies A = U\Lambda^{1/2}V.$$

Proof. Following the remark, write $A = U\Lambda^{1/2}V$.

$$\mu(AE)\mu(U\Lambda^{1/2}VE) \overset{\text{orthogonal invariance}}{=} \mu(\Lambda^{1/2}VE)$$
$$\overset{\text{dilation invariance}}{=} \det(\Lambda^{1/2})\mu(VE) \overset{\text{orthogonal invariance}}{=} \det(\Lambda^{1/2})\mu(E).$$

We conclude the proof by observing

$$|\det(A)| = |\det(U\Lambda^{1/2}V)| = |det(U)|det(\Lambda^{1/2})|\det(V)| = \det(\Lambda^{1/2}),$$

since the determinant of an orthogonal matrix is equal to ± 1. $\square$

Now we define the notion of σ-algebra. In a nutshell, σ-algebra is a family subsets that is closed under countable unions and complements, containing the empty set.

Definition 1.21. A subset $\mathcal{A}$ of $\mathcal{P}(\mathbb{R}^d)$ is called a σ-algebra if the following three conditions hold

1. $\emptyset \in \mathcal{A}$,
2. $E \in \mathcal{A} \implies E^c \in \mathcal{A}$
3. $E_i \in \mathcal{A}$ for $i \in \mathbb{N} \implies \bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$.

We have already seen the collection of measurable sets $\mathcal{M}(\mathbb{R}^d)$ is a sigma algebra. We also saw in the first lecture that $\mathcal{M}(\mathbb{R}^d)$ is strictly contained in the power set $\mathcal{P}(\mathbb{R}^d)$. The Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ is the smallest σ-algebra containing all the open sets in $\mathbb{R}^d$. We are going to see that also $\mathcal{B}(\mathbb{R}^d)$ is strictly contained $\mathcal{M}(\mathbb{R}^d)$.

Theorem 1.22. $\mathcal{B}(\mathbb{R}^d) \subsetneq \mathcal{M}(\mathbb{R}^d)$.

Before we proceed with the proof we state and prove severals helpful facts.

Lemma 1.23. *Let $f\colon \mathbb{R}^d \to \mathbb{R}^d$ be a continuous function. Let*

$$\mathcal{A} = \{E \subset \mathbb{R}^d \colon E = f^{-1}(V) \text{ for some } V \in \mathcal{B}(\mathbb{R}^d)\},$$

where $f^{-1}(U)$ is the preimage of V. Then $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$.

Remark. Note that $\mathcal{A}$ is a σ-algebra.

Proof. Let $\mathcal{B}'$ be the collection of all $V \in \mathcal{B}(\mathbb{R}^d)$ with $f^{-1}(V) \in \mathcal{B}(\mathbb{R}^d)$.
$\mathcal{B}'$ contains all open subsets, because if V is open, then also $f^{-1}(V)$ (by continuity). $\mathcal{B}'$ is a *sigma*-algebra as well, let us validate the three conditions:

1. $f^{-1}(\emptyset) = \emptyset \in \mathcal{B}(\mathbb{R}^d)$, so $\emptyset \in \mathcal{B}'$.
2. $V \in \mathcal{B}' \implies f^{-1}(V) \in \mathcal{B}(\mathbb{R}^d) \implies f^{-1}(V^c) = (f^{-1}(V))^c \in \mathcal{B}(\mathbb{R}^d) \implies V^c \in \mathcal{B}'$.
3. Similarly as 2.

That means, $\mathcal{B}'$ is a σ-algebra which contains all open sets, so by the definition of the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d) \subset \mathcal{B}'$ and consequently $\mathcal{A} \subset \mathcal{B}(\mathbb{R}^d)$. $\square$

Now let us define the Cantor ternary set:

$$C = \left\{ x = \sum_{i=1}^{\infty} \frac{a_i}{3^i} \colon a_i \in \{0, 2\} \text{ for all } i \right\}.$$

Note that clearly $C \subset [0, 1]$ with $0, 1 \in C$.

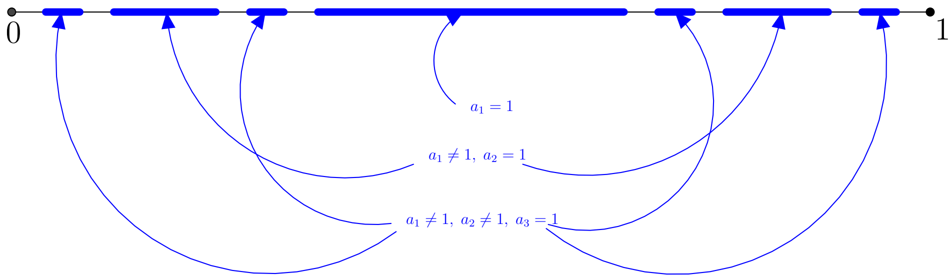


Figure 11: Construction of the Cantor set: at each step we remove the blue open intervals - the middle "one third" from each of the intervals that were not removed before. Here we present only three steps, the full construction is repeating continuing the procedure infinitely many times.

Let us calculate the outer measure of the Cantor set. Notice that at each the i-th step of the construction, which we presented in the picture, we remove an interval of measure $2/3^i$. This gives for every $N > 0$

$$\mu(C) \leq 1 - \sum_{i=1}^{N} \frac{2}{3^i}.$$

Since this holds for all $N > 0$, we conclude

$$\mu(C) \leq 1 - \sum_{i=1}^{\infty} \frac{2}{3^i} = 0.$$

so the measure the Cantor set is a nonempty set whose outer measure is equal zero. Moreover the cardinality of C is equality to the cardinality of the set of real numbers. This is because C has at eleast as many elements as the set of infinite sequences taking values in $\{0, 2\}$ and taking both values infinitely often.

Remark. Take $E \subset \mathbb{R}^d$ with $\mu(E) = 0$. Then for any $F \subset \mathbb{R}^d$

$$\mu(F) \geq \underbrace{\mu(F \cap E)}_{\leq \mu(E) = 0} + \mu(F \cap E^c).$$

This means that E is Lebesgue measurable.

Hence, the Cantor set is also Lebesgue measurable.

Example (Devil's staircase)**.** Note that any number $x \in [0,1]$ we can write as

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}, \quad \text{with } a_i \in \{0,1,2\}.$$

Define $f\colon [0,1] \to [0,1]$

$$f(x) = \begin{cases} \sum_{i=1}^{\infty} \frac{1}{2}\frac{a_i}{2^i} \text{ if all } a_i \in \{0,2\}, \\ \sum_{i=1}^{N-1} \frac{1}{2}\frac{a_i}{2^i} + \frac{1}{2^N} \text{ if } \min_{a_i=1} i = N \end{cases}.$$

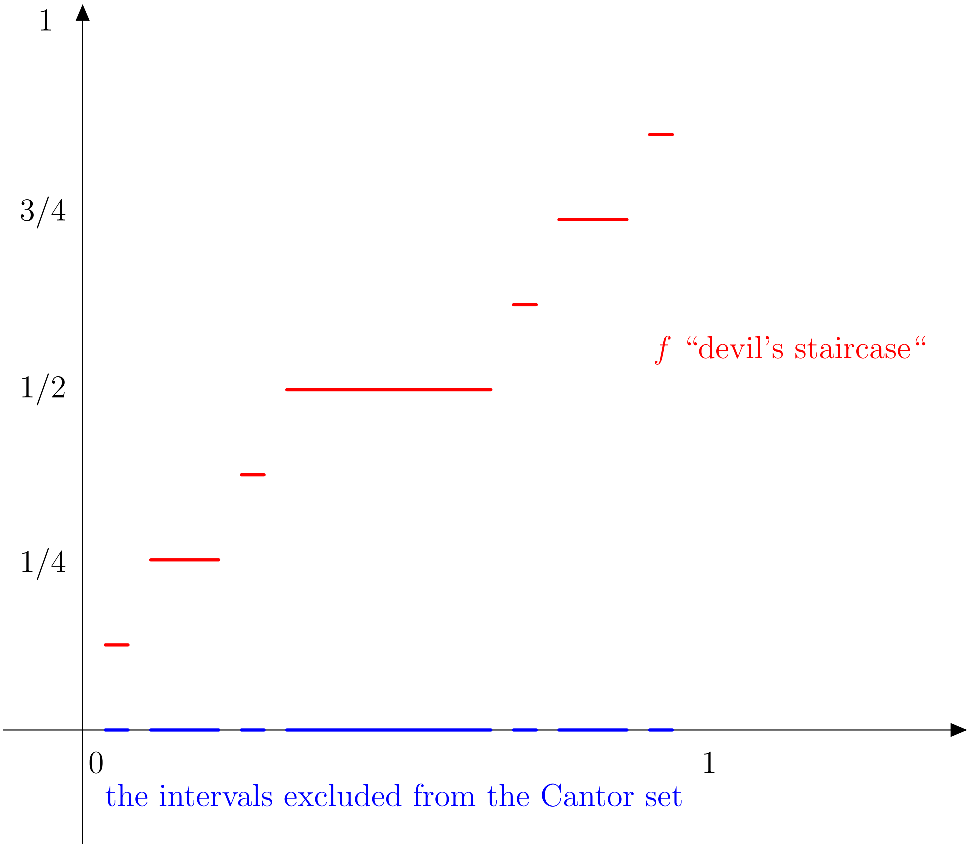


Figure 12: The first three steps of the construction of the devil's staircase.

f is well defined: let

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i} = \sum_{i=1}^{\infty} \frac{b_i}{3^i}$$

be two different expansions of x. Then, without loss of generality, there exists a natural number N such that for all $i < N$, $a_i = b_i$, and $a_N = 0$, $b_N = 0$, and for $i > N$ $a_i = 2$, $b_i = 0$. Notice that our definition of f gives the same value for both expansions. Similarly we obtain that f is monotone. We may then also conclude that f is continuous: we know from methods from Analysis I that if a monotone function g defined on $[0, 1]$ has the property that $g([0, 1])$ contains all dyadic numbers in $[0, 1]$, then it is continuous (exercise). One can easily check that the range of our function f certainly contains all dyadic numbers between 0 and 1.

Remark. Let E be a countable subset of $\mathbb{R}^d$. Then $\mu(E) = 0$. That is because $E = \{x_i\}_{i\in\mathbb{N}}$ and we have the following covering

$$E \subset \bigcup_{i=1}^{\infty} \left(x_i - \frac{\delta}{2^i}, x_i + \frac{\delta}{2^i} \right),$$

where $\delta > 0$ is arbitrary. Thus, by monotonicity and subadditivity we can bound the outer measure of E by the sum of measures of the small intervals

$$\mu(E) \leq \sum_{i=1}^{\infty} \mu \left(\left((x_i - \frac{\delta}{2^i}, x_i + \frac{\delta}{2^i} \right) \right) = 2\delta.$$

We can take δ as small as we wish, so $\mu(E) = 0$.

End of lecture 5. November 5, 2015

Recall that we want to show Theorem 1.22 which states that the inclusion $\mathcal{B} \subset \mathcal{M}$ is strict. That is, that there exists Lebesgue measurable sets which are not Borel sets.

For this we need some more preparatory statements.

Definition 1.24. Let X be a space with a σ-algebra $\mathcal{A}$. Let Y be a metric space. A function $f : X \to Y$ is called $\mathcal{A}$-measurable if for every open ball $B \subset Y$, $f^{-1}(B) \in \mathcal{A}$.

Thus a function is measurable if the preimage of any open ball is measurable.

Lemma 1.25. *If $Y = \mathbb{R}$, then the following are equivalent*[5]

1. *For every $x \in \mathbb{Y} \cup -\mathbb{Y}$ we have $f^{-1}((x, \infty)) \in \mathcal{A}$.*
2. *For every $x \in \mathbb{Y} \cup -\mathbb{Y}$ we have $f^{-1}([x, \infty)) \in \mathcal{A}$.*

[5] $\mathbb{Y}$ denotes the dyadic numbers

3. *f is $\mathcal{A}$-measurable*

Proof. 3. $\Rightarrow$ 1. The open set (x,∞) is a union of open balls $(x,\infty)=\cup_i B_i(x_i)$ with $x_i\in\mathbb{Y}$. Then

$$f^{-1}((x,\infty))=\bigcup_i f^{-1}(B_i)$$

Since B_i are measurable, so are $f^{-1}(B_i)$ and hence the countable union.
1. $\Rightarrow$ 2. We write

$$[x,\infty)=\bigcap_{y\in\mathbb{Y},\, y<x}(y,\infty)$$

so that

$$f^{-1}([x,\infty))=\bigcap_{y\in\mathbb{Y},\, y<x} f^{-1}(y,\infty),$$

and argue argue analogously as in 1.
2. $\Rightarrow$ 3. For any interval (a,b) write

$$(a,b)=\bigcup_{x\in\mathbb{Y},x>a}[x,\infty)\setminus\bigcap_{y\in\mathbb{Y},y<b}[y,\infty).$$

$\square$

Theorem 1.26. *If f is $\mathcal{A}$-measurable, then the preimage of every Borel set is $\mathcal{A}$-measurable.*

The proof is analogous to the proof of Lemma 1.23 from the previous lecture.

Now we are ready to prove existence of Lebesgue measurable sets which are not Borel sets.

Proof of Theorem 1.22. Denote by f be the devil's staircase function. Define its "inverse function"

$$g(x)=\inf_{f(y)=x} y.$$

This is not really the inverse function since for a given x there is no unique y which would be mapped to x with f. But we may define the value of g at x to be the infimum over all such y. Since f is continuous, the infimum $g(x)$ also satisfies

$$f(g(x))=x$$

for all $x\in[0,1]$. This implies that g is injective. Since g is monotone, g is Borel measurable (exercise). Observe that $g([0,1])\subset C$ where C denotes the Cantor set.

Let now $A \subset [0, 1]$ be a set which is not Lebesgue measurable. In particular, A is not a Borel set. By injectivity of g we have

$$A = g^{-1}(g(A))$$

(Recall that in general $A \subset g^{-1}(g(A))$, but for injective function there is an equality.) We have $g(A) \subset C$ and hence $g(A)$ is Lebesgue measurable. However, $g(A)$ is not Borel. If it were, $g^{-1}(g(A))$ would have to be Borel by the previous theorem, which is a contradiction. □

The next theorem states that any Lebesgue measurable set can be approximated, up to a set of measure zero, from the outside with open and from the inside with closed sets, respectively.

Theorem 1.27. *If $E \subset \mathbb{R}^d$ is Lebesgue measurable, then*

1. *there exist open sets F_i, $i = 1 \ldots, \infty$, such that*

$$F_i \supset E \quad \text{and} \quad \mu\Big(\bigcap_{i=1}^{\infty} F_i \setminus E\Big) = 0.$$

2. *there exist closed sets G_i, $i = 1, \ldots, \infty$, such that*

$$G_i \subset E \quad \text{and} \quad \mu(E \setminus \bigcup_{i=1}^{\infty} G_i) = 0.$$

Proof. Note that 2. follows from 1. by taking complements since closed sets are complements of open sets. So we only need to show 1.
Case 1. E is bounded, i.e $E \subset B_N(0)$ for some $N > 0$.
Then $\mu(E) < \infty$. For $n \geq 1$ we find an open covering

$$E \subset \underbrace{\bigcup_{i=1}^{\infty} B_{n_i}(x_i)}_{\tilde{F_n}}$$

with $x_i \in E, n_i \geq 0$, such that

$$\mu(E) + 2^{-n} \geq \sum_{i=1}^{\infty} \mu(B_{r_i}(x_i)).$$

Define

$$F_n := \bigcap_{j=1}^{n} \tilde{F}_{n_j}$$

which is a open set. We have $F_{n+1} \subset F_n$. Since $E \subset \cap_{n=1}^{\infty} F_n$ and $\mu(F_n \setminus E) < 2^{-n}$ it follows

$$\mu\Big(\bigcap_{n=1}^{\infty} F_n \setminus E\Big) = 0.$$

Case 2. E is unbounded.
We intersect E with a sequence of annuli and for $m = 1, \ldots, \infty$, inductively define

$$\begin{aligned} E_1 &:= E \cap B_1(0) \\ E_m &:= E \cap B_m(0) \setminus B_{m-1}(0) \end{aligned}$$

Each E_m is bounded. So we can apply the theorem to each of these bounded sets and take the union over m. Since each bounded piece can be approximated up to an error of measure zero, the error of the union amounts to a countable union of measure zero sets, which has measure zero. The details are left as an exercise. $\square$

1.3 Littlewood's three principles

Littlewood gives an intuitive guide for understanding measurable sets and functions.

1. Every measurable set in $\mathbb{R}^d$ is nearly a disjoint union of dyadic cubes.
2. Every measurable function is nearly continuous.
3. Every convergent sequence of measurable functions is nearly uniformly convergent.

Now we will explain how to understand the word "nearly" in each of the statements and make them precise.

Principle 1.

Theorem 1.28. *Let $E \subset \mathbb{R}^d$ be Lebesgue measurable and $\mu(E) < \infty$. For every $\varepsilon > 0$ there exists F which is a finite disjoint union of dyadic cubes such that $\mu(E \Delta F) < \varepsilon$.*

Recall that for two sets E and F, the symmetric difference $E\Delta F$ is defined as $E\Delta F = (E \setminus F) \cup (F \setminus E) = (E \cap F^c) \cup (F \cap E^c)$.

Proof. Let $\mathcal{T}'$ be a collection of dyadic cubes with

$$E \subset \bigcup_{Q\in\mathcal{T}'} Q, \qquad \mu(E) + \frac{\varepsilon}{2} \geq \sum_{Q\in\mathcal{T}'} \mu(Q).$$

Note that without loss of generality we may assume that $\mathcal{T}'$ consists of pairwise disjoint cubes. Choose a finite subcollection $\mathcal{T}'' \subset \mathcal{T}'$ such that

$$\sum_{Q\in\mathcal{T}'} \mu(Q) - \frac{\varepsilon}{2} \leq \sum_{Q\in\mathcal{T}''} \mu(Q).$$

Set

$$F := \bigcup_{Q\in\mathcal{T}''} Q.$$

Then we have

$$\begin{aligned}
\mu(E\Delta F) &\leq \mu(E\setminus F) + \mu(F\setminus E)\\
&\leq \mu\Big(\bigcup_{\mathcal{T}'\setminus\mathcal{T}''} Q\Big) + \mu\Big(\Big(\bigcup_{\mathcal{T}'} Q\Big)\setminus E\Big)\\
&\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon
\end{aligned}$$

where we used that $\mu((\bigcup_{\mathcal{T}'} Q)\setminus E) = \mu(\bigcup_{\mathcal{T}'} Q) - \mu(E)$ since E is measurable. $\square$

<u>Principle 3.</u>

Theorem 1.29 (Egorov). *Let $E \subset \mathbb{R}^d$ be Lebesgue measurable and $\mu(E) < \infty$. Let $f, f_k : E \to \mathbb{R}$ be measurable and for each $x \in E$, $\lim_{k\to\infty} f_k(x) = f(x)$. Then for every $\varepsilon > 0$ there is a closed set $F \subset E$ with $\mu(E\setminus F) < \varepsilon$ such that $f_k|_F$ converges to $f|_F$ uniformly.*

Proof. For $k, n \in \mathbb{N}$ define

$$E_k^n := \{x \in E : |f_j(x) - f(x)| < 2^{-n} \text{ for all } j > k\}$$

Exercise: show that E_k^n are Lebesgue measurable. We have $E_k^n \subset E_{k+1}^n$ and

$$\bigcup_{k=1}^{\infty} E_k^n = E.$$

(The inclusion " $\subseteq$ " is obvious, while " $\supseteq$ " holds due to pointwise convergence.) We claim that

$$\lim_{k\to\infty} \mu(E_k^n) = \mu(E).$$

To see this we split E_k^n into disjoint annuli, i.e.

$$E_k^n = E_1^n \cup \bigcup_{j=2}^{k} E_j^n \setminus E_{j-1}^n.$$

By additivity of μ we have

$$\mu(E_k^n) = \mu(E_1^n) + \sum_{j=2}^{k} \mu(E_j^n \setminus E_{j-1}^n)$$

Taking the limit on both sides we establish the claim. Now we choose k_n large enough such that

$$\mu(E \setminus E_{k_n}^n) < 2^{-n}.$$

We also choose N large enough such that

$$\mu\big(E \setminus \underbrace{\bigcap_{n=N}^{\infty} E_{k_n}^n}_{\tilde{F}}\big) < \sum_{n=N}^{\infty} 2^{-n} < \varepsilon/2.$$

On $\tilde{F}$, f_k converges to f uniformly. Indeed, for every $\delta > 0$ we find $n > N$ such that $2^{-n} < \delta$. Then for each $x \in \tilde{F} \subset E_{k_n}^n$ and for all $k \geq k_n$ we have $|f_k(x) - f(x)| < \delta$. By Theorem 1.27 we find a closed set $F \subset \tilde{F}$ with $\mu(\tilde{F} \setminus F) \leq \varepsilon/2$. Then the convergence of $f_k|_F \to f|_F$ is uniform on F. $\square$

Before proceeding with principle 2. we state a result on pointwise approximation of measurable functions with finite linear combination of characteristic functions.

Theorem 1.30. *Let $E \subset \mathbb{R}^d$ be measurable with $\mu(E) < \infty$. Let $f : E \to \mathbb{R}$ be measurable. Then there exists a sequence of functions $f_k : \mathbb{R}^d \to \mathbb{R}$ which are finite linear combinations of characteristic functions of dyadic cubes and a set $E' \subset E$ with $\mu(E \setminus E') = 0$ such that for all $x \in E'$ we have $\lim_{k\to\infty} f_k(x) = f(x)$.*

One also says that f_k converges to f *almost everywhere* on E, abbreviated *a.e. on E*. In general we say that some property holds almost everywhere if the set of elements on which the property does not hold has measure zero.

The proof of this theorem is left as an exercise. One first considers functions $f = \mathbf{1}_F$ which are characteristic function of measurable sets in $\mathbb{R}^n$ and one approximates F with dyadic cubes. For a general f one has to pass to finite linear combinations of characteristic function of measurable sets and finally to finite linear combinations of characteristic functions of dyadic cubes.

Now we have everything prepared to state a precise formulation of the second principle.

Principle 2.

Theorem 1.31 (Lusin). *Let E be Lebesgue measurable with $\mu(E) < \infty$. Let $f : E \to \mathbb{R}$ be measurable and bounded. Let $\varepsilon > 0$. Then there exists a closed set $F \subset E$ with $\mu(E \setminus F) < \varepsilon$ such that $f|_F$ is continuous.*

Proof. By the previous theorem we find a sequence f_k of finite linear combinations of characteristic function of dyadic cubes such that

$$\lim_{k\to\infty} f_k(x) = f(x) \quad \text{a.e. on } E$$

Note that each f_k is continuous a.e., since each characteristic function has only two points of discontinuity. Since countable unions of sets with measure zero have measure zero, almost everywhere are all f_k continuous and converge to f. By Egorov's theorem, there is a closed set $F \subset E$ with $\mu(E \setminus F) < \varepsilon$ such that $f_k \to f$ uniformly on F. Therefore, $f|_F$ is continuous. $\square$

End of lecture 6. November 11, 2015

2 Integration theory

This lecture we pass from the measure therory to integration theory.
Reminder: In Analysis I we have defined the integral for montone increasing or decreasing functions. Let $\mathbb{X} := [0, \infty)$ and denote the dyadic numbers of order $-k$ by $\mathbb{Y}_k$. Let $f\colon \mathbb{X} \to \mathbb{X}$ be a montone decreasing function. For $a, b \in \mathbb{Y}_{k_0}$ we defined

$$\int_a^b f(x)dx = \sup_{k>k_0} \underbrace{\sum_{\substack{y\in\mathbb{Y}_k \\ a<y\leq b}} 2^{-k} f(y)}_{=:L_k(f)\text{ - lower Riemann sum}} = \inf_{k>k_0} \underbrace{\sum_{\substack{y\in\mathbb{Y}_k \\ a\leq y<b}} 2^{-k} f(y)}_{=:U_k(f)\text{ - upper Riemann sum}}$$

We used that $U_k(f) \leq L_k(f) + 2^{-k} f(a)$ and monotonicity of the lower and the upper sums in k , which together gives equality of the supremum of lower Riemann sums with infimum of the upper Riemann sumds, the value of which we call the Newton integral. The same argument applies if $b = \infty$, the value of the newton integral may be finit or infinte in this case. Now, let $f : \mathbb{X} \to \mathbb{X} \cup \{\infty\}$. We the define

$$\int_0^\infty f(x)dx = \sup_{k>0} \sum_{y \in \mathbb{Y}_k \setminus \{0\}} 2^{-k} f(y).$$

This has a chance to be finite if f takes the value ∞ at most at the point 0. If f takes the value ∞ at 0, the upper Riemann sums will all be infinte, which is why we base the theory on the lower Riemann sums in this case.
Our aim will be to reduce the integral on a general outer measure space to the integral of monotone decreasing functions $\mathbb{X} \to \mathbb{X} \cup \{\infty\}$. The idea is to consider dyadsic intervals mnot one the domain of the function to be integrated, which will be an arbitrary outer measure space, but on the range of functions, whjich will still be (contained in) $\mathbb{X}$, see the nearby figures.

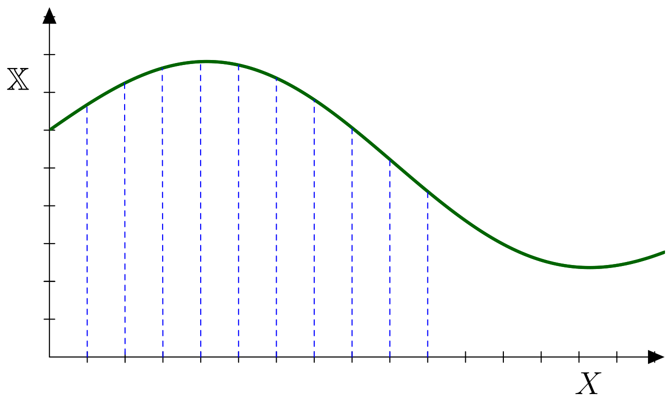


Figure 13: The idea of integration by decomposing the domain into dyadic intervals (works e.g. if the domain X is $\mathbb{X}$)

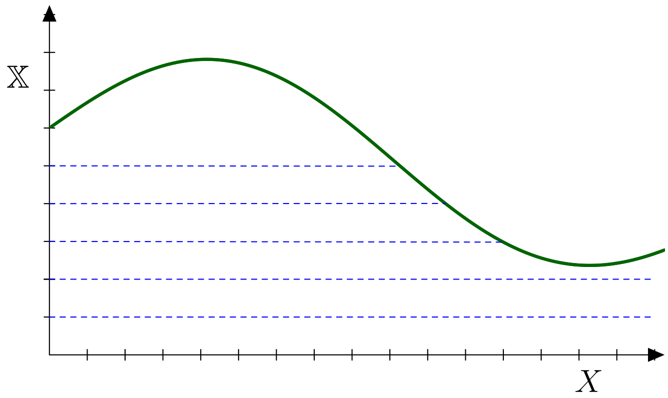


Figure 14: The idea of integration by decompositng the range into dyadic intervals

Definition 2.1. Let X be a set with an outer measure μ. Let $f\colon X \to \mathbb{X}$. We define

$$\int_X f d\mu := \int_0^\infty \mu(\{x\colon f(x) > \lambda\})d\lambda.$$

Remark. Note that the function $g(\lambda) = \mu(\{x\colon f(x) > \lambda\})$ is monotone decreasing, so the Newton integral on right hand side is well defined! Indeed, if $\lambda_1 < \lambda_2$, then

$$\{x\colon f(x) > \lambda_1\} \supset \{x\colon f(x) > \lambda_2\},$$

so by the monotonicity

$$\mu(\{x\colon f(x) > \lambda_1\}) \geq \mu(\{x\colon f(x) > \lambda_2\}).$$

Remark. We also have

$$\int_X f d\mu = \int_0^\infty \mu(\{x\colon f(x) \geq \lambda\})d\lambda,$$

because the functions $g_1(\lambda) = \mu(\{x\colon f(x) > \lambda\})$, $g_2(\lambda) = \mu(\{x\colon f(x) \geq \lambda\})$ have the same right-sided limits, so the respective Newton integrals are equal.

Definition 2.2.

$$\mu(f > \lambda) := \mu(\{x\colon f(x) > \lambda\}).$$

Remark. If $f = g$ almost everywhere in X, then

$$\int_X f d\mu = \int_X g d\mu.$$

Let $E = \{x\colon f(x) \neq g(x)\}$. We have that $\mu(E) = 0$ and

$$\mu(f > \lambda) \leq \mu(g > \lambda) + \mu(E) = \mu(g > \lambda).$$

Exactly the same argument gives $\mu(g > \lambda) \leq \mu(f > \lambda)$, so $\mu(f > \lambda) = \mu(g > \lambda)$.

Remark. Note that if $f \leq g$, then for each λ,

$$\mu(f > \lambda) \leq \mu(g > \lambda),$$

so

$$\int_X f d\mu \leq \int_X g d\mu.$$

We will prove two important theorems that let one to pull out a pointwise limit of functions outside of the integral.

Theorem 2.3 (Monotone convergence). *Let μ be an outer measure on a space X with the family of Caratheodory measurable sets $\mathcal{M}(X)$. Let $f_n\colon X \to \mathbb{X}$ be a monotone increasing sequence of $\mathcal{M}(X)$-measurable functions, i.e. for all $x \in X$ and $n \in \mathbb{N}$, $f_n(x) \leq f_{n+1}(x)$. Then*

$$\sup_n \int_X f_n d\mu = \int_X (\sup_n f_n) d\mu.$$

Remark. Both supremums above are actually limits.

Proof. Let $E_n := \{x\colon f_n(x) > \lambda\}$. First of all, we notice that

$$\{x\colon \sup_n f_n(x) > \lambda\} = \bigcup_n E_n.$$

We show "$\supset$" and "$\subset$". If there exists k such that $f_k(x) > \lambda$, then the supremum at point x is clearly bigger than λ. This shows "$\supset$". On the other hand, if $\sup_n f_n(x) > \lambda$, then there exists k such that $f_k(x) > \lambda$, we have "$\subset$".
The sequence of functions is increasing, so $E_n \subset E_{n+1}$. Hence, we can write E_n as the following disjoint sum

$$E_n = E_1 \cup \bigcup_{k=2}^{\infty} E_k \setminus E_{k-1},$$

where all the sets are measurable. Using this we can express the sum of E_n's a disjoint sum of measurable sets

$$\bigcup_n E_n = E_1 \cup \bigcup_{k=2}^{\infty} E_k \setminus E_{k-1},$$

what gives

$$\mu(\bigcup_n E_n) = \mu(E_1) + \sum_{k=2}^{\infty} \mu(E_k \setminus E_{k-1}).$$

The considerations above show that

$$\mu(\sup_n f_n > \lambda) = \mu(\bigcup_n E_n) = \sup_n \mu(E_n).$$

Using this we can rewrite $\int_X (\sup_n f_n) d\mu$ as

$$\sup_{k>0} \sup_{\substack{A \subset \mathbb{Y}_k \\ A \text{ finite}}} \sum_{\lambda \in A} \mu(\sup_n f_n > \lambda) 2^{-k} = \sup_{k>0} \sup_{\substack{A \subset \mathbb{Y}_k \\ A \text{ finite}}} \sum_{\lambda \in A} \sup_n \mu(f_n > \lambda) 2^{-k}.$$

Now we can simply pull the innermost supremum, which is a supremum of a monotone sequence and thus a limit, first out of the finite sum over A and then outside of the two remaining supremums, to obtain

$$\sup_n \sup_{k>0} \sup_{\substack{A \subset \mathbb{Y}_k \\ A \text{ finite}}} \sum_{\lambda \in A} \mu(f_n > \lambda) 2^{-k} = \sup_n \int_X f_n d\mu.$$

□

Theorem 2.4 (Fatou)**.** *Let μ be an outer measure on a space X with the family of Caratheodory measurable sets $\mathcal{M}(X)$. Let $f_n \colon X \to \mathbb{X}$ be a sequence of $\mathcal{M}(X)$-measurable functions. Then the following inequality holds*

$$\int_X \liminf_n f_n d\mu \leq \liminf_n \int_X f_n d\mu.$$

Proof. Rewrite the right hand side of the inequality similarly as we did in the proof of the monotone convergence

$$\sup_{n_0} \inf_{n>n_0} \sup_{k>0} \sup_{\substack{A \subset \mathbb{Y}_k \setminus \{0\} \\ A \text{ finite}}} \sum_{\lambda \in A} \mu(f_n > \lambda) 2^{-k}.$$

Notice that if we move the infimum above inside the inner sum we obtain a quantity that is smaller or equal.

$$\geq \sup_{n_0} \sup_{k>0} \sup_{\substack{A \subset \mathbb{Y}_k \setminus \{0\} \\ A \text{ finite}}} \sum_{\lambda \in A} \mu(\inf_{n>n_0} f_n > \lambda) 2^{-k}.$$

Moreover, we can move the most outer supremum inside the sum exactly the same way as we did in the previous proof (or one can say that we are simply applying the monotone convergence here) and obtain that the display above is

$$\geq \sup_{k>0} \sup_{\substack{A\subset \mathbb{Y}_k\setminus\{0\} \\ A \text{ finite}}} \sum_{\lambda\in A} \mu(\sup_{n_0} \inf_{n>n_0} f_n > \lambda) 2^{-k} = \int_X \liminf_n f_n d\mu.$$

□

Remark. Note that for a measurable set E we have

$$\int_X 1_E d\mu = \int_0^\infty \mu(1_E > \lambda) d\mu \overset{\mu(1_E>\lambda)=0 \text{ for } \lambda\geq 1}{=} \int_0^\infty \mu(1_E > \lambda) d\mu$$
$$\overset{\mu(1_E>\lambda)=\mu(E) \text{ for } \lambda<1}{=} \int_0^1 \mu(E) = \mu(E).$$

Theorem 2.5. *Let $\lambda_j \in \mathbb{X}$ and E_j for $1 \leq j \leq N$ be measurable and pairwise disjoint sets with $\mu(E_j) < \infty$. Then*

$$\int_X \sum_{j=0}^N \lambda_j 1_{E_j} d\mu = \sum_{j=1}^N \lambda_j \mu(E_j).$$

Remark. A function $\sum_{j=0}^N \lambda_j 1_{E_j}$ with pairwise disjoint measurable sets E_j we call simple.

Proof. Without loss of generality let $\lambda_j \leq \lambda_{j+1}$ and $\lambda_0 = 0$ (if not $\lambda_j = 0$, then we can simply add the empty set to the collection with coefficient 0). We compute, starting from the left hand side

$$\int_X \sum_{j=1}^N \lambda_j 1_{E_j} d\mu \overset{\text{Definition}}{=} \int_0^\infty \mu(\sum_{j=0}^N \lambda_j 1_{E_j} > \lambda) d\lambda$$
$$\overset{\text{measurability of } E_j}{=} \int_0^\infty \sum_{j:\,\lambda_j>\lambda} \mu(E_j) d\lambda = \sum_{k=1}^N \int_{\lambda_{k-1}}^{\lambda_k} \sum_{j:\,j\geq k} \mu(E_j) d\lambda$$
$$= \sum_{k=1}^N (\lambda_k - \lambda_{k-1}) \sum_{j\geq k} \mu(E_j) \overset{\text{reordering of the sum}}{=} \sum_{j=0}^N \sum_{j\geq k} (\lambda_k - \lambda_{k-1}) \mu(E_j) = \sum_{j=0}^N \lambda_j \mu(E_j).$$

□

End of lecture 7. November 12, 2015

In the previous lecture we showed linearity of the integral for characteristic functions of pairwise disjoint measurable sets (Theorem 2.5. Note that measurability of the sets is crucial for that.). Using this we can now show additivity of the integral for simple functions, i.e. for finite linear combinations of characteristic functions of pairwise disjoint measurable sets.

Lemma 2.6. *Let f, g be simple functions. Then*

$$\int_X f + g \, d\mu = \int_X f d\mu + \int_X g d\mu.$$

Proof. Let $f = \sum_{i=1}^N \lambda_i \mathbf{1}_{E_i}$ and $g = \sum_{j=1}^M \nu_j \mathbf{1}_{E_j}$. We may assume that $\cup E_i = \cup E_j$, as otherwise we add $0 \cdot \mathbf{1}_{\cup E_i \setminus \cup E_j}$ to the sum for g and $0 \cdot \mathbf{1}_{\cup E_j \setminus \cup E_i}$ to the sum for f. Then

$$f + g = \sum_{i=1}^N \sum_{j=1}^M (\lambda_i + \nu_j) \mathbf{1}_{(E_i \cap E_j)}$$

which is a simple function. We have

$$\begin{aligned}
\int_X f + g \, d\mu &\overset{(*)}{=} \sum_i \sum_j (\lambda_i + \nu_j) \mu(E_i \cap E_j) \\
&= \sum_i \sum_j \lambda_i \mu(E_i \cap E_j) + \sum_j \sum_i \nu_j \mu(E_i \cap E_j) \\
&\overset{(**)}{=} \sum_i \lambda_i \mu(E_i) + \sum_j \lambda_j \mu(E_j) \\
&= \int_X f d\mu + \int_X g d\mu
\end{aligned}$$

where we used $(*)$: linearity for characteristic functions of meas. sets, $(**)$: measurability of $E_i \cap E_j$. $\square$

By approximating measurable functions from below by simple functions, we can now show additivity of the integral for measurable functions.

Theorem 2.7. *Let $f, g : X \to \mathbb{X}$ be measurable. Then*

$$\int_X f + g \, d\mu = \int_X f d\mu + \int_X g d\mu$$

Proof. Define

$$f_k = \sum_{n=0}^{2^{2k}} 2^{-k} n \mathbf{1}_{\{x: 2^{-k} n < f(x) \leq 2^{-k}(n+1)\}}$$

(see Figure 15). Then for each x, $f_k(x) \le f(x) \le f_k(x) + 2^{-k}$, so we have $\sup_k f_k = f$. Observe that the sequence f_k is monotonously increasing and that f_k are measurable (since f is measurable). We have

$$\begin{aligned}
\int_X f + g\,d\mu &= \int_X \sup_k f_k + \sup_k g_k\,d\mu \\
&= \int_X \sup_k (f_k + g_k) \\
&\overset{(1)}{=} \sup_k \Big(\int_X f_k + g_k\,d\mu \Big) \\
&\overset{(2)}{=} \sup_k \Big(\int_X f_k d\mu + \int_X g_k d\mu \Big) \\
&= \sup_k \int f_k d\mu + \sup_k \int g_k d\mu \\
&\overset{(3)}{=} \int_X f d\mu + \int_X g d\mu
\end{aligned}$$

We used (2): additivity for simple functions. (1), (3): monotone convergence.

$\square$

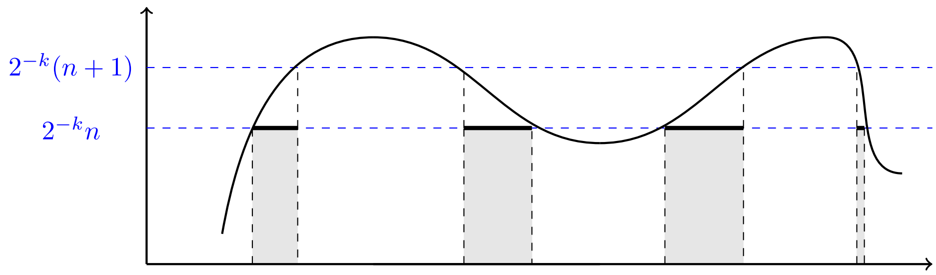


Figure 15: Function $2^{-k}n\mathbf{1}_{\{x:2^{-k}n<f(x)\le 2^{-k}(n+1)\}}$.

To establish linearity of the integral it remains to show homogeneity.

Theorem 2.8. *Let $c \in \mathbb{X}$, $f : X \to \mathbb{X}$ measurable. Then*

$$c\int_X f d\mu = \int_X cf d\mu$$

Proof (sketch).

1. $c = 2^k$, $k \ge 0$: Induct on k. (Write $2f = f + f$ etc.)
2. $c = 2^{-k}, k \ge 0$: Write $f = 2^k 2^{-k} f$ and use homogeneity for 2^k.
3. $c = 2^k n$, $k \in \mathbb{Z}, n \in \mathbb{N}$: Induct on n.

4. $c \in \mathbb{X}$: Approximate c with an increasing sequence of numbers of the form $2^k n$ and use monotone convergence theorem.

$\square$

The third convergence theorem (the first two being monotone convergence theorem and Fatou's lemma) is the following.

Theorem 2.9 (Dominated converge theorem). *Let $f_n : X \to \mathbb{X}$ be measurable. Assume that for each $x \in X$ $\lim_{n\to\infty} f_n(x) = f(x)$ exists. Assume that $f_n \leq h$ for some measurable function h with $\int_X h < \infty$. Then*

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X \lim_{n\to\infty} f_n d\mu = \int_X f d\mu$$

Proof. Define g_n via $f_n + g_n = h$. Then g_n is measurable (exercise), $g(x) = \lim_{n\to\infty} g_n(x)$ exists and $f + g = h$. We have

$$\begin{aligned}
\int_X h d\mu &= \int_X \liminf_{n\to\infty}(f_n + g_n) d\mu \\
&\stackrel{(1)}{=} \int_X \liminf_{n\to\infty} f_n d\mu + \int_X \liminf_{n\to\infty} g_n d\mu \\
&\stackrel{(2)}{\leq} \liminf_{n\to\infty} \int f_n d\mu + \liminf_{n\to\infty} \int g_n d\mu \\
&\leq \liminf_{n\to\infty} \Big(\int_X f_n d\mu + \int_X g_n d\mu \Big) \\
&\stackrel{(3)}{=} \liminf_{n\to\infty} \int_X f_n + g_n \, d\mu \\
&= \liminf_{n\to\infty} \int_X h d\mu = \int_X h d\mu
\end{aligned}$$

(1): Linearity for measurable functions $\liminf f_n$ and $\liminf g_n$ (Exercise: they are indeed measurable.) (2): Fatou. (3): Additivity of the integral.

The above sequence of inequalities shows that all inequalities must be equalities. In particular, equality in (2) implies that

$$\liminf_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu \tag{6}$$

$$\liminf_{n\to\infty} \int_X g_n d\mu = \int_X g d\mu \tag{7}$$

Since $\liminf(\int g_n) = h + \liminf(-\int f_n) = h - \limsup(\int f_n)$, we have

$$\liminf_{n\to\infty} \int_X g_n d\mu + \limsup_{n\to\infty} \int_X f_n d\mu = \int_X h d\mu = \int_X g d\mu + \int_X f d\mu.$$

The equality (7) then implies that

$$\limsup_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu.$$

Together with (6) this implies that $\lim_{n\to\infty} \int_X f_n d\mu$ exists and

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu.$$

□

The following examples show that the inequality in Fatou's lemma can in general not be turned into an equality. (See also Figure 16.)

Example. (Escape to vertical infinity.) Let $f_k = 2^k \mathbf{1}_{(0,2^{-k}]}$. Then $\int_X f_k d\mu = 1$ and hence

$$\liminf_{k\to\infty} \int_X f_k d\mu = 1.$$

But $\liminf_{k\to\infty} f_k = 0$ and so

$$\int_X \liminf_{k\to\infty} f_k = 0.$$

Example. (Escape to horizontal infinity.) Let $f_k = 2^{-k}\mathbf{1}_{(0,2^k]}$. Then $\int_X f_k d\mu = 1$, but $\liminf_{k\to\infty} f_k(x) \le \liminf_{k\to\infty} 2^{-k} = 0$.

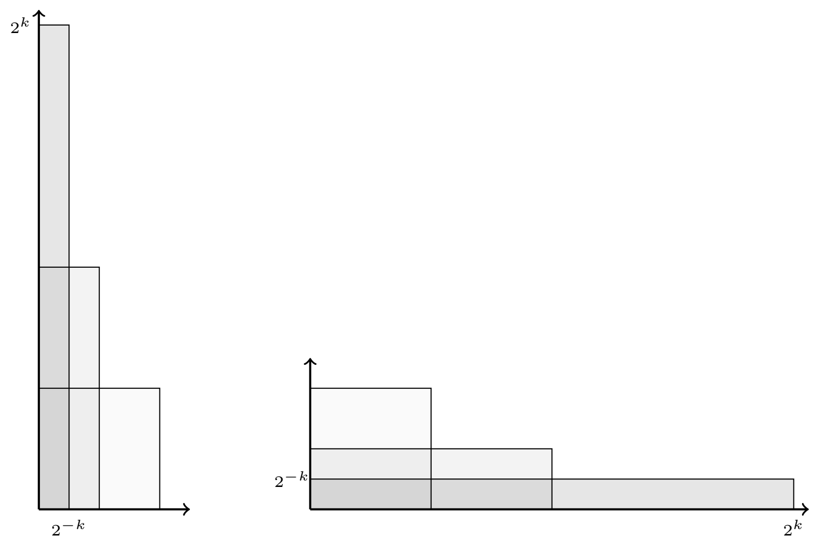


Figure 16: Escape to vertical and horizontal infinity.

Our next goal is to show that if $f : [a,b] \to \mathbb{X}$ is monotone, its Lebesgue and Newton integral coincide.

Theorem 2.10. *Let $f : [a, b] \to \mathbb{X}$ be monotone increasing. Then*

$$\int_a^b f(x)dx = \int_X f d\mu$$

The integral on the left hand-side should be understood in the Newton sense, while the integral on the right hand-side in the Lebesgue sense.

Proof. The idea is to prove the theorem for simple functions and use monotone convergence theorem to extend it to general monotone functions. Suppose f is a positive monotone simple function of the form

$$\sum_{j=1}^{n} \lambda_j \mathbf{1}_{E_j}$$

with $\lambda_j < \lambda_{j+1}$, $\cup E_j = [a, b]$. Since f is monotone, E_j lies left of E_{j+1}. The intervals E_j are of the form (a_j, b_j) or $(a_j, b_j]$ if f is left continuous and $[a_j, b_j)$ or $[a_j, b_j]$ if f is right continuous. Depending on that we also define $\tilde{E}_j$ as $(a_j, b]$ or $[a_j, b]$, respectively.
We have

$$\begin{aligned}
\int_a^b \sum_{j=1}^{n} \lambda_j \mathbf{1}_{E_j} &= \int_a^b \sum_{j=1}^{n} (\lambda_j - \lambda_{j-1}) \mathbf{1}_{\tilde{E}_j} dx \\
&= \sum_{j=1}^{n} (\lambda_j - \lambda_{j-1}) \underbrace{\mu(\tilde{E}_j)}_{b-a_j} \\
&= \sum_{j=1}^{n} \int_{\lambda_{j-1}}^{\lambda_j} \mu(\tilde{E}_j) d\lambda \\
&= \int_0^{\lambda_n} \mu(\tilde{E}_j) d\lambda \\
&= \int_0^{\infty} \mu(\{x : f(x) > \lambda\}) d\lambda \\
&= \int_X f d\mu
\end{aligned}$$

To prove the theorem for general monotone functions, approximate them with an increasing sequence of monotone simple function and use monotone convergence theorem. The details are left as an exercise. $\square$

Now we extend the notion of integrability to functions mapping into $\mathbb{R}$.

Definition 2.11. Let (X, μ) be an outer measure space. A function $f : X \to \mathbb{R}$ is called *integrable*, if we can write it as

$$f = f_1 - f_2$$

with

$$\begin{aligned} f_1 : X \to \mathbb{X} \quad \text{measurable,} \quad \int_X f_1 d\mu < \infty \\ f_2 : X \to \mathbb{X} \quad \text{measurable,} \quad \int_X f_2 d\mu < \infty \end{aligned}$$

Definition 2.12. The integral of an integrable function $f : X \to \mathbb{R}$ is defined as

$$\int_X f d\mu = \int_X f_1 d\mu - \int_X f_2 d\mu.$$

Since f, f_1, f_2 are measurable, the integral is well defined, i.e. it is independent of the choice of f_1, f_2. Indeed, assume we have two representations

$$f_1 - f_2 = \tilde{f}_1 - \tilde{f}_2$$

Then

$$f_1 + \tilde{f}_2 = \tilde{f}_1 + f_2$$

Since the functions are measurable, by additivity of the integral

$$\int f_1 + \int \tilde{f}_2 = \int \tilde{f}_1 + \int f_2$$

and hence

$$\int f_1 - \int f_2 = \int \tilde{f}_1 - \int \tilde{f}_2.$$

Remark. The function $1/t$ is integrable as a function $\mathbb{X} \to \mathbb{X}$ with an infinite integral

$$\int_0^\infty \frac{1}{t} dt = \infty$$

However, it is not integrable as a function $\mathbb{X} \to \mathbb{R}$ (since this would contradict the above integral being infinite).

Further remarks.

Let X be a set and $\mathcal{A} \subset \mathcal{P}(X)$ a $\sigma-$algebra. Let μ be a σ-additive measure on $\mathcal{A}$. It is a natural question whether $\mathcal{A}$ consists of Caratheodory measurable sets with respect to some outer measure.
More precisely, set $\mathcal{T} = \mathcal{A}$ and $\tau(E) = \mu(E)$. Denote by μ^* the outer measure generated by $\mathcal{T}, \tau$ and denote by $\mathcal{M}(X)$ the set of all Caratheodory measurable sets with respect to μ^*.

Question 1: For $E \in \mathcal{T}$, do we have $\mu^*(E) = \tau(E)$ $(= \mu(E))$?
Answer: Yes. suppose that $E \subset \cup E_j$. Since μ is a σ-additive measure,

$$\tau(E) = \mu(E) \leq \sum_j \mu(E_j).$$

Question 2: Is $\mathcal{A} = \mathcal{M}$, i.e. are all sets in $\mathcal{A}$ Caratheodory measurable with respect to μ^*?
Answer: No. The inclusion $\mathcal{A} \subset \mathcal{M}$ holds, i.e. if $E \in \mathcal{A}$, then it is Caratheodory measurable. To see that we need to check that for each $F \in \mathcal{T}\,(= \mathcal{A})$ we have

$$\mu^*(F) = \mu^*(F \cap E) + \mu^*(F \cap E^c).$$

But from the answer on the first question it follows

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

where the last equality holds since μ is σ-additive. However, we do not have $\mathcal{M} \subset \mathcal{A}$. This can be seen by taking $\mathcal{A} = \mathcal{B}$ the Borel σ-algebra and μ the Lebesgue measure restricted to $\mathcal{B}$. One can check that the generated outer measure μ^* is the Lebesgue outer measure and $\mathcal{M}$ are the Lebesgue measurable sets. But we already know that $\mathcal{B} \subsetneq \mathcal{M}$.

The inclusion $\mathcal{A} \subset \mathcal{M}$ is strict in general. $\mathcal{M}$ is the largest σ-algebra on which one can define a measure extending the measure μ on $\mathcal{A}$. $\mathcal{M}$ is called the *completion* of $\mathcal{A}$. Note that our theory is built in such a way that we always work with the complete σ-algebra.

End of lecture 8. November 17, 2015

2.1 Product measures

We consider products of two measure spaces, for example one can think of $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. The theory we discuss in this section can be then be extended to any finite product of measure spaces by induction. In certain situations, for example in probability theory, one deals with infinite products of measure spaces, we will also discuss such spaces in the future.
Let (X_1, μ_1), (X_2, μ_2) be outer measure spaces. Assume that the outer measure μ_i is generated by the collection $\mathcal{T}_i$ of Caratheodory measurable sets with map $\mu_i(E) = \tau_i(E)$ for $E_i \in \mathcal{T}_i$. Let $X = X_1 \times X_2$ be the Cartesian product of the respective sets. Consider the outer measure μ on X generated by the family of subsets

$$\mathcal{T} = \{E_1 \times E_2 \colon E_1 \subset X_1,\ E_2 \subset X_2 \ \text{ Caratheodory measurable subsets}\}$$

and the function

$$\tau(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2).$$

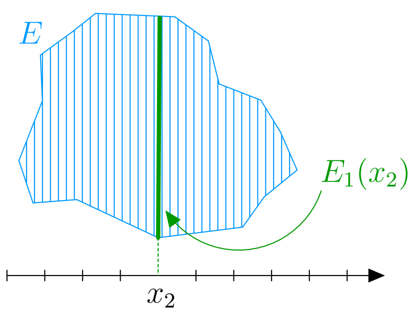
Theorem 2.13. *Consider the setup as above and assume $\mu_1(X_1), \mu_2(X_2) < \infty$. If $E \subset X$ is measurable, then for almost all $x_2 \in X_2$ the set*

$$E_1(x_2) := \{x_1 \in X_1 \colon (x_1, x_2) \in E\}$$

is measurable and $\mu_1(E_1(x_2))$ is a measurable function of x_2. Moreover, we have

$$\mu(E) = \int_{X_2} \mu_1(E_1(.)) d\mu_2.$$

The same statement holds with $i = 1, 2$ interchanged.



Proof. Our goal is to show that for any $\epsilon, \eta > 0$, the μ_2-measure of the set

$$\begin{aligned} S_\epsilon := \{x_2\colon\ &\nexists F_1, G_1 \subset X_1 \text{ measurable such that} \\ &E_1(x_2) \subset F_1,\ X_1 \setminus E_1(x_2) \subset G_1,\ \mu_1(F_1) + \mu_1(G_1) \leq \mu_1(X_1) + \epsilon\}. \end{aligned}$$

is smaller than η. Let us see how this implies the first statement of our theorem.
First, note that if the above holds for any $\eta > 0$, then we have $\mu_2(S_\epsilon) = 0$ for any $\epsilon > 0$. Hence $\mu_2(\bigcup_{k=1}^\infty S_{2^{-k}}) = 0$. It suffices to show measurability of $E_1(x_2)$ for all $x_2 \in X_2 \setminus \bigcup_{k=1}^\infty S_{2^{-k}}$. In this situation for all natural $k > 0$ there exist measurable subsets $F_1^{(k)}$, $G_1^{(k)}$ of X_1 such that

$$E_1(x_2) \subset F_1^{(k)},\ X_1 \setminus E_1(x_2) \subset G_1^{(k)},\ \mu_1(F_1^{(k)}) + \mu_1(G_1^{(k)}) \leq \mu_1(X_1) + 2^{-k}.$$

Next, note that the sets

$$F_1 := \bigcap_k F_1^{(k)},\ G_1 := \bigcap_k G_1^{(k)}$$

are both measurable and $E_1 \subset F_1$, $X \setminus E_1 \subset G_1$ and

$$\mu_1(F_1) + \mu_1(G_1) \leq \mu_1(X_1).$$

$$\mu_1(E_1(x_2)) + \mu_1(E_1(x_2)^c) = \mu_1(X_1).$$

We then have for all Caratheodory measurable sets $Y_1 \subset X_1$

$$\mu_1(E_1(x_2) \cap Y_1) + \mu_1(E_1(x_2)^c \cap Y_1) = \mu_1(Y_1).$$

Sinxce the Caratheodory measurable sets generate μ_1, we see that $E_1(x_2)$ is measurable.
We saw how the first statement of the theorem follows if we show our goal. Now let us see how to achieve the goal.
Let $E \subset X$ be measurable. Choose coverings of E and $X \setminus E$

$$E \subset \bigcup_{n=1}^\infty F_1^n \times F_2^n,$$

$$X \setminus E \subset \bigcup_{n=1}^\infty G_1^n \times G_2^n,$$

with F_i^n's measurable and such that

$$\eta\epsilon + \mu(E) \geq \sum_n \mu_1(F_1^n)\mu_2(F_2^n),$$

$$\eta\epsilon + \mu(X \setminus E) \geq \sum_n \mu_1(G_1^n)\mu_2(G_2^n).$$

For $x_2 \in X_2$ define

$$F(x_2) = \bigcup_{n\colon x_2 \in F_2^n} F_1^n$$

and similarly

$$G(x_2) = \bigcup_{n\colon x_2 \in G_2^n} G_1^n.$$

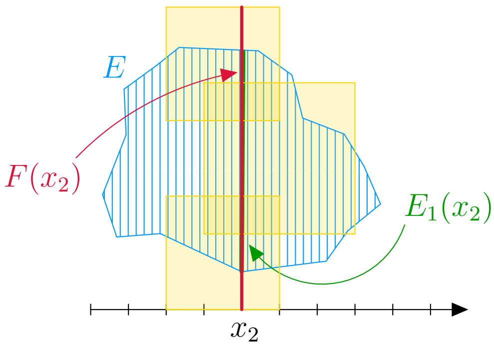


Figure 17: Defining set $F(x_2)$: the yellow squares represent the elements of the covering of E that intersect $E_1(x_2)$. $F(x_2)$ is the intersection of these squares with the line $y = x_2$.

Integrating $\mu_1(F(x_2))$ we obtain

$$\int_{X_2} \mu_1(F(x_2))d\mu_2 \leq \int_{X_2} \sum_{n\colon x_2 \in F_2^n} \mu_1(F_2^n)d\mu_2 = \int_{X_2} \sum_n \mu_1(F_1^n)1_{F_2^n}d\mu_2$$
$$\overset{\text{additivity + mon. conv.}}{=} \sum_n \mu_1(F_1^n)\mu_2(F_2^n) \leq \mu(E) + \epsilon\eta.$$

Analogously we get the inequality

$$\int_{X_2} \mu_1(G(x_2))d\mu_2 \leq \mu(X \setminus E) + \epsilon\eta.$$

Note the following simple identity for measurable sets

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

In the following we will use it several times. Notice that adding up the last two inequalities and using the last identity we obtain

$$\int_{X_2} \mu_1(X_1) d\mu_2 + \int_{X_2} \mu_1(F(x_2) \cap G(x_2)) d\mu_2 \leq \mu(X) + 2\epsilon\eta.$$

This, however, implies that

$$\mu_2(\{x_2 \colon \mu_1(F(x_2) \cap G(x_2)) > \epsilon\}) < 2\eta.$$

The last thing we need to notice is that if for some x_2 it holds that

$$\mu_1(F(x_2) \cap G(x_2)) > \epsilon,$$

then, again, by the identity above

$$\mu_1(F(x_2)) + \mu_1(G(x_2)) \leq \mu(X) + \epsilon.$$

This proves the inequality we stated at the very beginning of this proof.
We still need to argue that $\mu(E)$ can be recovered integrating $\mu_1(E_1(x_2))$. Since $E_1(x_2) \subset F_1(x_2)$ and $X_1 \setminus E_1(x_2) \subset G_1(x_2)$, by the previous step of the proof we also have

$$\int_{X_2} \mu_1(E_1(x_2)) d\mu_2 \leq \mu(X \setminus E) + \epsilon\eta.$$

$$\int_{X_2} \mu_1(X_1 \setminus E_1(x_2)) d\mu_2 \leq \mu(X \setminus E) + \epsilon\eta.$$

Given four numbers a, b, A, B, such that $a \leq A$ and $b \leq B$

$$a + b = A + B \implies a = A, \ b = B.$$

Letting $\epsilon, \eta \to 0$, this is exactly the case here, hence in particular

$$\mu(E) = \int_{X_2} \mu_1(E_1(x_2)) d\mu_2.$$

Taking better and better coverings $F^{(n)}$ of E we obtain a decreasing sequence of nonnegative functions $\mu_1(F^{(n)}(\cdot))$ convergent to $\mu_1(E_1(x_2))$, so the measurability follows. $\square$

The above theorem holds also if both spaces X_1 and X_2 are *σ-finite*.

Definition 2.14. A measurable space X is called σ-finite if there exists a countable family of subsets $\{F_n \colon n = 1, 2, ...\}$, such that for all n, $\mu(F_n) < \infty$ and $X = \bigcup_n F_n$.

Note that in the above definition we can assume that $F_n \subset F_{n+1}$. In this case, the previous theorem holds for $X \cap F_n$, for any n. A suitable (exercise) limiting argument $n \to \infty$ shows the theorem for for the whole space X.

Proposition 2.15. *If an outer measure μ_i on X_i is generated via sets $E_i \in \mathcal{T}_i$ and a nonnegative function τ_i for $i = 1, 2$, then the function*

$$\tau(E_1 \times E_2) = \tau_1(E_1)\tau_2(E_2),$$

defined for $E_1 \times E_2 \in \mathcal{T} = \mathcal{T}_1 \times \mathcal{T}_2$, generates a product measure on $X = X_1 \times X_2$ and for $E_i \in \mathcal{T}_i$ we have

$$\mu(E_1 \times E_2) = \mu_1(E_1)\mu_2(E_2).$$

Now we will prove the theorems of Tonelli and Fubini, which let us to write an integral over a product measure space as an iterated integral over the respective spaces as well as change the order of integration under certain assumptions.

Theorem 2.16 (Tonelli)**.** *Let (X_1, μ_1), (X_2, μ_2) be σ-finite outer measure spaces. Let $f \colon X_1 \times X_2 \to \mathbb{X}$ be a measurable function. Then the function $f_{x_2} \colon X_1 \to \mathbb{X}$, defined as $f_{x_2}(x_1) = f(x_1, x_2)$ is measurable. Moreover, the function*

$$\int_{X_1} f_{x_2} d\mu_1 \colon X_2 \to \mathbb{X}$$

is measurable and the following equality holds

$$\int_{X_2} \left[\int_{X_1} f_{x_2} d\mu_1 \right] d\mu_2 = \int_X f d\mu.$$

Proof. If $f = 1_E$ is the characteristic function of a measurable set, the statement follows from the previous theorem. Otherwise we can approximate f by a monotone sequence of simple functions, for which the statement is true by linearity of the integral, and use the monotone convergence theorem, which gives the measurability statements and the equality of the integrals. $\square$

Remark. One has to be a bit careful concerning the notion of measurability in product spaces. Take a subset E of $\mathbb{R}$ that is not measurable. Note that the set $\{0\} \times E$ has measure zero in $\mathbb{R}^2$, so it is measurable!

Recall the definition of integrable functions: a function $f \colon X \to \mathbb{R}$ *integrable* if there exists a decomposition $f = f_1 - f_2$, with $f_1, f_2 \colon X \to \mathbb{X}$ measurable and $\int f_1 < \infty$, $\int f_2 < \infty$. Then we also have that

$$\int f = \int f_1 - \int f_2.$$

Theorem 2.17 (Fubini)**.** *Let X_1, X_2 be as in the statement of Tonelli's theorem. Let $f\colon X \to \mathbb{R}$ be an integrable function. Then the function $f_{x_2}\colon X \to \mathbb{R}$ is integrable for almost all $x_2 \in X_2$ as well as the function*

$$\int_{X_1} f_{x_2} d\mu_1 \colon X_2 \to \mathbb{R}$$

is integrable. Moreover, the following equality holds

$$\int_{X_2} \left[\int_{X_1} f_{x_2} d\mu_1 \right] d\mu_2 = \int_X f d\mu.$$

Example. Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be defined as follows.

$$f(x) = \begin{cases} 2^{2(k-1)} \text{ if } x \in (2^{-k}, 2^{-k+1}] \times (2^{-k}, 2^{-k+1}] \text{ for } k \in \mathbb{N}, \\ -2^{2(k-1)+1} \text{ if } x \in (2^{-k-1}, 2^{-k}] \times (2^{-k}, 2^{-k+1}] \text{ for } k \in \mathbb{N}, \\ 0 \text{ otherwise} \end{cases}$$

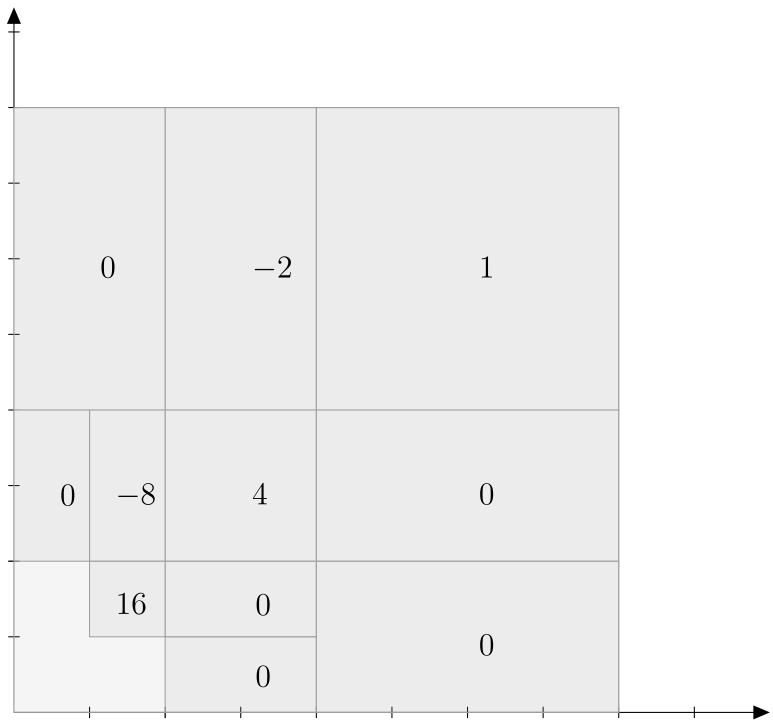


Figure 18: First steps of the construction of function f.

Notice that all f_{x_2} and all f_{x_1} are integrable, and $\int_{X_1} f_{x_2} d\mu_1$ is integrable in x_2 and $\int_{X_2} f_{x_1} d\mu_2$ is integrable in x_1, but f is not integrable!

End of lecture 9. November 19, 2015

2.2 Outer Radon measures

Recall the construction of the Lebesgue measure on $\mathbb{R}^d$. We define $\mathcal{T}$ to be the set of all dyadic cubes in $\mathbb{R}^d$, which are of the form $Q = \prod_{i=1}^d [a_i, b_i)$ where $a_i = 2^k n, b_i = 2^k(n+1)$ and $n, k \in \mathbb{Z}$. Their side-length is $|a_i - b_i| = 2^k$. We define $\tau : \mathcal{T} \to [0, \infty)$ by setting $\tau(Q) = 2^{kd}$, which is the product of the side-lengths of Q. Then we generate the outer measure μ of any set $E \subset \mathbb{R}^d$ by covering it with dyadic cubes. We have shown with some work that the generated outer measure satisfies $\mu(Q) = \tau(Q)$, which got us started for the development of the theory.

In this chapter we generalize this construction by considering different maps τ than above. Since we want that the generated μ satisfies $\mu(Q) = \tau(Q)$ for all $Q \in \mathcal{T}$, we need to impose certain conditions on τ.

Definition 2.18. Let $\mathcal{T}$ be the set of all dyadic cubes in $\mathbb{R}^d$ and $\tau : \mathcal{T} \to [0, \infty)$ a map satisfying

(1) (Martingale condition) For every $Q \in \mathcal{T}$

$$\tau(Q) = \sum_{\substack{Q' \subset Q \\ k'+1=k}} \tau(Q')$$

(2) (Regularity condition) For every $Q \in \mathcal{T}$ and every $\varepsilon > 0$ there exists k_0 such that

$$\sum_{\substack{Q' \subset Q \\ k'+k_0=k \\ \overline{Q'} \not\subset Q}} \tau(Q') \leq \varepsilon \tau(Q).$$

The outer measure generated by $\mathcal{T}, \tau$ is called an *outer Radon measure.*

Note that $\tau(Q) = 2^{kd}$ which generates the Lebesgue outer measure satisfies both conditions. The first one is immediate, while the second one can be obtained by counting the cubes Q' at the boundary of Q.
Recall that these properties were crucial to show that $\mu(Q) = \tau(Q)$ for the Lebesgue measure. Indeed, suppose that Q is covered by dyadic cubes in some collection $\mathcal{T}'$. For finite $\mathcal{T}'$, induction and the martingale condition are

used to show that the best covering of Q is itself[6]. If $\mathcal{T}'$ is countably infinite, then we argue by compactness to reduce to the finite case. For that we need to make sure that we can approximate Q from inside by a compact sets and from the outside by open sets. This used to the condition (2).

Theorem 2.19. *If $\mathcal{T}, \tau$ generate an outer Radon measure μ on $\mathbb{R}^d$, then $\mu(Q) = \tau(Q)$ for all $Q \in \mathcal{T}$.*

Proof. The proof proceeds as in the Lebesgue case, using (1) and (2) at appropriate places. Here we only briefly sketch the argument. We need to show that if $Q \subset \bigcup_{\mathcal{T}'} Q'$, then

$$\tau(Q) \leq \sum_{\mathcal{T}'} \tau(Q')$$

If $\mathcal{T}'$ is finite, this can be shown using the martingale property (1) and induction of the scale of cubes. If $\mathcal{T}'$ is is countably infinite, we use a compactness and ε-argument together with the regularity condition (2). $\square$

Restricting the outer measure to Caratheodory measurable sets gives rise to a Radon measure μ. The dyadic cubes are μ-measurable. This can be seen using $\tau(Q) = \mu(Q)$ and the martingale condition (1).

Example. Now we discuss some examples of Radon measures.

1. Lebesgue measure

2. Dirac measure. For $Q \in \mathcal{T}$ we set

$$\tau(Q) = \begin{cases} 1 : (0, \dots, 0) \in Q \\ 0 : (0, \dots, 0) \notin Q \end{cases}$$

   Martingale condition (1): If $\tau(Q) = 0$, then $(0, \dots, 0) \notin Q$ and hence $(0, \dots, 0) \notin Q'$ for every $Q' \subset Q$. Then $\tau(Q') = 0$ for every $Q' \subset Q$ and both sides of the equality in (1) are 0. If $\tau(Q) = 1$, then $(0, \dots, 0) \in Q$. There is exactly one cube $Q' \subset Q$, $k' + 1 = k$ which contains $(0, \dots, 0)$. Both sides of the equality in (1) are 1.

   Regularity condition (2): We have (exercise)

$$\bigcap_{k_0 \in \mathbb{N}} \bigcup_{\substack{Q' : \overline{Q'} \subset Q \\ k' + k_0 = k}} Q' = \emptyset$$

[6]For this we need only " $\leq$ " of the martingale condition.

So there exists k_0 such that $(0,\dots,0) \notin \bigcup_{\substack{Q':\overline{Q'}\subset Q\\ k'+k_0=k}}$. This implies

$$\sum_{\substack{Q':\overline{Q'}\subset Q\\ k'+k_0=k}} \tau(Q') = 0 \leq \varepsilon\tau(Q).$$

3. Let $x^{(n)}$ be a sequence in $\mathbb{R}^d$ and let λ_n be a sequence in $[0,\infty)$ with $\sum_n \lambda_n < \infty$. For $Q \in \mathcal{T}$ define

   $$\tau(Q) = \sum_{n:x^{(n)}\in Q} \lambda_n$$

   The conditions (1) and (2) hold. This can be shown similarly as for the Dirac measure. The generated measure can be seen as a finite linear combination of Dirac measures at different points.

4. Let $x^{(n)}$ be a sequence in $\mathbb{R}^d$ and let λ_n be a sequence in $[0,\infty)$ such that for every bounded set $E \subset \mathbb{R}^d$ we have

   $$\sum_{x_n\in E} \lambda_n < \infty$$

   We define τ as before by

   $$\tau(Q) = \sum_{n:x^{(n)}\in Q} \lambda_n.$$

   This is a generalization of the previous example and (1) and (2) can be shown similarly.

5. Let $d = 1$. Let $f : \mathbb{R} \to \mathbb{R}$ be monotone. For $[a,b) \in \mathcal{T}$ define[7]

   $$\tau([a,b)) = \lim_{b'\nearrow b} f(b') - \lim_{a'\nearrow a} f(a')$$

   Both conditions are fulfilled. (1): For $c = (a+b)/2$ we have

   $$\begin{aligned}\tau([a,b)) &= \big(\lim_{c'\nearrow c} f(c') - \lim_{a'\nearrow a} f(a')\big) + \big(\lim_{b'\nearrow c} f(b') - \lim_{c'\nearrow a} f(c')\big)\\ &= \tau([a,c)) + \tau([c,b))\end{aligned}$$

[7] Every monotone function has one-sided limits. Here we consider left-hand limits since our intervals are open on the right. If we had $(a,b]$ we would consider right-hand limits.

(2): We need to show that for k' large enough

$$\tau([b-2^{-k'}, b)) \leq \varepsilon\tau([a,b))$$

This follows from

$$\lim_{k'\to\infty}\big(\lim_{b'\nearrow b} f(b') - \lim_{c'\nearrow b-2^{-k'}} f(c')\big) = 0$$

The abstract measure theory carries over from the Lebesgue case. In particular, we have the following theorem.

Theorem 2.20. *If μ is an outer Radon measure, then every Borel set is μ-measurable.*

Proof. As for the Lebesgue measure. □

Thus, Borel sets are measurable with respect to any Radon measure. However, for a general set in $\mathbb{R}^d$, measurability depends on the considered measure.

Example. Every subset of $\mathbb{R}^d$ is measurable with respect to the Dirac measure.

Thus there exist sets which are measurable with respect to the Dirac measure but are not Lebesgue measurable. On the other hand, there exist Radon measures μ such that there are Lebesgue measurable sets which are not μ-measurable. This can be seen from the following example.

Example. The devil's staircase is a monotone functions. So we can τ as in Example 2.2, 5 with f begin the devil's staircase. This generates an outer measure μ. Denote by $\tilde{C}$ the set of all elements in the Cantor set which do not have a finite tertiary representation. Denote by $\tilde{I}$ the set of all elements in $[0,1)$ with do not have a finite binary representation. Then f restricts to a bijection $f:\tilde{C}\to\tilde{I}$. This map transports μ onto the Lebesgue measure (it can be interpreted as transporting the structure from C to $[0,1)$). More precisely, if $E\subset\tilde{C}$, then

$$\mu(E) = \mu_{\text{Leb}}(f(E))$$

If $E\subset\tilde{I}$ is not Lebesgue measurable, then $f^{-1}(E)$ is not μ-measurable in $\tilde{C}$. But $f^{-1}(E)\subset C$ and hence $\mu_{\text{Leb}}(E)=0$.

Integration with respect to a Radon measure

If μ is a Radon measure, we define the integral

$$\int_{\mathbb{R}^d} f d\mu$$

in the same way as in the Lebesgue case. The abstract theory carries over (except for translation invariance, which can already be seen to fail for the Dirac measure). In particular, the convergence theorems hold etc.

Continuous functions $f : \mathbb{R}^d \to [0, \infty)$ are μ- measurable. If f is also compactly supported, then

$$\int_{\mathbb{R}^d} f d\mu < \infty$$

This can be seen as follows. Denote $K = \mathrm{supp}(f)$. One can cover K with finitely many dyadic cubes $Q' \in \mathcal{T}'$ of scale 0. Denote by C the maximum of f on K (it exists by continuity of f and compactness of K). Then we have

$$f \leq C \sum_{Q' \in \mathcal{T}'} \mathbf{1}_{Q'}$$

and therefore

$$\int_{\mathbb{R}^d} f d\mu \leq C \sum_{\mathcal{T}'} \mu(Q') < \infty.$$

which shows the claim.
Our next goal is to give a characterisation of Radon measures. For a Radon measure μ we can define [8] $\Lambda : C_c(\mathbb{R}^d, [0, \infty)) \to [0, \infty)$ given by

$$\Lambda(f) = \int_{\mathbb{R}^d} f d\mu$$

It is additive, i.e.

$$\Lambda(f, g) = \Lambda(f) + \Lambda(g).$$

Our main point is that the "converse" also holds. Every such additive positive functional Λ is given as an integral with respect to a unique Radon measure. We remark that since Λ acts on positive functions, additivity implies monotonicity, i.e.

$$f \leq g \Rightarrow \Lambda(f) \leq \Lambda(g)$$

We also remark that there is one-to-one correspondence between the Radon measure and its generating data. More precisely, let $\mathcal{T}$ be the collection of

[8] C_c denotes continuous compactly supported functions.

generating sets on X and define τ, τ' which generate Radon measures μ, μ', respectively. If $\tau \neq \tau'$ (on generating sets), then $\mu \neq \mu'$. Indeed, assuming $\mu = \mu'$, for every $Q \in \mathcal{T}$ we obtain

$$\tau(Q) = \mu(Q) = \mu'(Q) = \tau(Q')$$

which is a contradiction.

Theorem 2.21. *Let $\Lambda : C_c(\mathbb{R}^d, [0,\infty)) \to [0,\infty)$ be such that for each $f, g \in C_c(\mathbb{R}^d, [0,\infty))$*

$$\Lambda(f+g) = \Lambda(f) + \Lambda(g)$$

Then there exists exactly one Radon measure μ on $\mathbb{R}^d$ such that

$$\Lambda(f) = \int_{\mathbb{R}^d} f d\mu$$

Proof. We prove the theorem in case $d = 1$, in higher dimensions one argues similarly. We construct τ which generates a unique measure μ. Let $Q \in \mathcal{T}$ and write $Q = [a, a+2^k)$ where $a, k \in \mathbb{Z}$. For $n \geq 1$ we define the functions f_n^Q as shown in Figure 19.[9]

Figure 19: Functions f_Q^n (blue) and f_Q^m (red), $n < m$.

The reason we chose this function and not $\mathbf{1}_{[a,a+2^k)}$ is that the characteristic function is not continuous and hence does not lie in the domain of Λ.
Define g_n^Q and h_n^Q such that

$$f_n^Q = g_n^Q + h_n^Q$$

as shown in Figure 20 (we write $b = a + 2^k$).

[9]Here we only provide a picture corresponding to the construction. One can write the function explicitly as an exercise. In higher dimensions one would take tensor products of such functions.

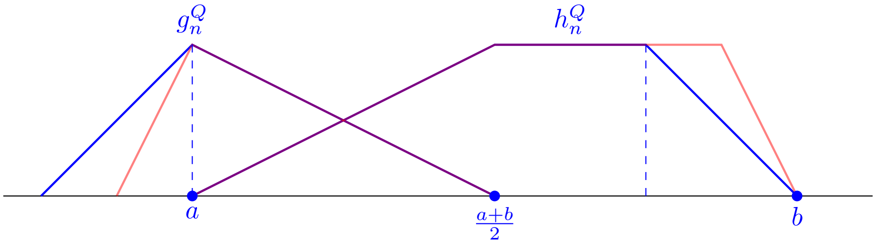


Figure 20: Functions g_Q^n, h_Q^n (blue) and g_Q^m, h_Q^m (red), $n < m$.

Then g_Q^n is monotonously decreasing in n and h_Q^n monotonously increasing. By monotonicity of Λ we have that $\Lambda(g_Q^n)$ is monotonously decreasing and $\Lambda(h_Q^n)$ monotonously increasing. Therefore the limits

$$\lim_{n\to\infty} \Lambda(g_Q^n) \quad \text{and} \quad \lim_{n\to\infty} \Lambda(h_Q^n)$$

exist. Hence also

$$\begin{aligned}\lim_{n\to\infty} (\Lambda(g_Q^n) + \Lambda(h_Q^n)) &= \lim_{n\to\infty} \Lambda(g_Q^n + h_Q^n) \\ &= \lim_{n\to\infty} \Lambda(f_Q^n)\end{aligned}$$

exists. For $Q \in \mathcal{T}$ we set

$$\tau(Q) = \lim_{n\to\infty} \Lambda(f_Q^n).$$

Now we check the martingale and regularity condition. Martingale condition: Observe that

$$f_n^Q = f_{n-1}^{Q_{\text{left}}} + f_{n-1}^{Q_{\text{right}}}$$

where Q_{left} and Q_{right} denote the left and the right child of Q, respectively. See figure 21.

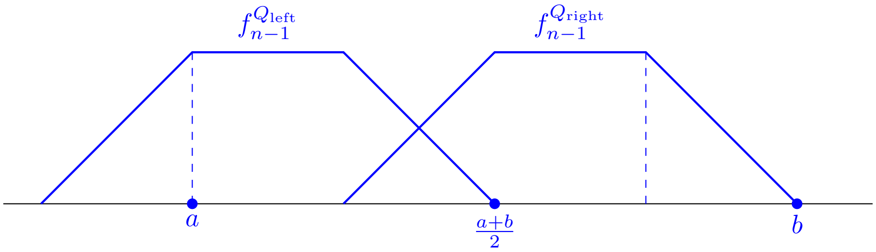


Figure 21: Functions $f_{n-1}^{Q_{\text{left}}}$ and $f_{n-1}^{Q_{\text{right}}}$

Applying Λ to f_n^Q and taking the limit we obtain

$$\tau(Q) = \tau(Q_{\text{left}}) + \tau(Q_{\text{right}})$$

Regularity condition: We show that for every $Q = [a, b)$ and every $\varepsilon > 0$ there exists k_0 such that $\tau([b - 2^{-k-k_0}, b)) \leq \varepsilon$. For $k_1 > k_0$ this then holds by monotonicity of Λ. Pick n_0 so large that for any $n \geq n_0$

$$\Lambda(h_n^Q) - \Lambda(h_{n_0}^Q) \leq \frac{\varepsilon}{2}$$

See Figure 22. Choose k_0 such that $h_{n_0}^Q \leq \frac{\varepsilon}{2}$ on $[b - 2 \cdot 2^{-k-k_0}, b)$. Denote $Q' = [b - 2^{-k-k_0}, b)$. Then we can estimate for m big enough

$$h_{n_0}^Q + (1 - \varepsilon) f_m^{Q'} \leq h_{n_0+k_0}^Q,$$

and in consequence

$$(1 - \varepsilon)\Lambda(f_m^{Q'}) \leq \Lambda(h_{n_0+k_0}^Q - h_{n_0}^Q) \leq \frac{\varepsilon}{2}$$

So passing to the limit as $m \to \infty$ and choosing n_0 and k_0 so that $\varepsilon < 1/2$ we obtain

$$\tau(Q') \leq \varepsilon.$$

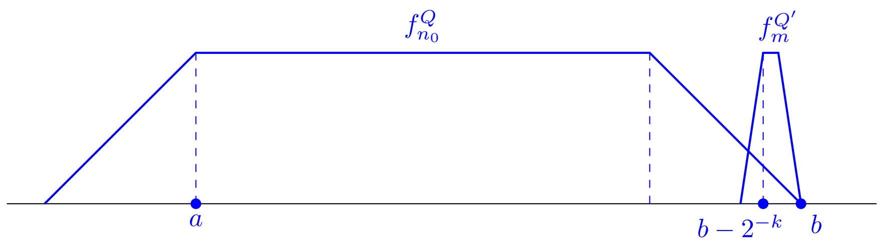


Figure 22: Functions $f_{n_0}^Q$ and $f_m^{Q'}$

End of lecture 10. November 24, 2015

Hence we get that τ generates a Radon measure μ. We still have to verify that for $f \in C_c(\mathbb{R}^d, [0, \infty))$ we have

$$\Lambda(f) = \int_{\mathbb{R}^d} f d\mu.$$

In order to do that, we will show that for any $f \in C_c(\mathbb{R}^d, [0, \infty))$ and any $\varepsilon > 0$ there exists a constant C_f, possibly dependent on f, such that

$$|\Lambda(f) - \int_{\mathbb{R}^d} f d\mu| \leq C_f \varepsilon.$$

Let N be such that f is equal to zero on $[-N, N]^c$. We define h as in Figure 23.

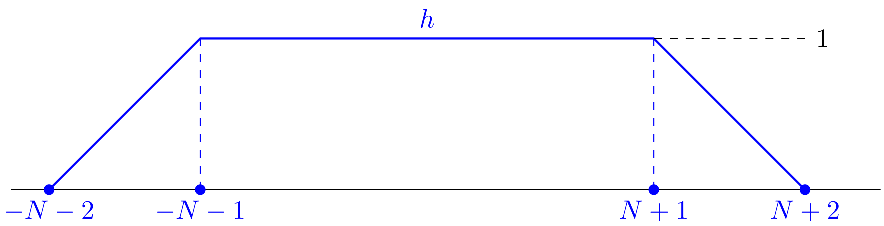


Figure 23: Function h a piecewise linear function supported on $[-N-2, N+2]$ and equal to 1 on $[-N-1, N+1]$.

Let $k > 0$ be such that

$$|x - y| \leq 3 \cdot 2^{-k} \implies |f(x) - f(y)| \leq \varepsilon,$$

we can do this because f is a uniformly continuous function. Define

$$\mathcal{T}' = \{\text{dyadic cubes in } [-N, N] \text{ of the order } k\}.$$

Because of the last the previous display we obtain a good approximation of f via a simple function constant on $Q \in \mathcal{T}'$, in the sense that

$$|\int_{\mathbb{R}^d} f d\mu - \int_{\mathbb{R}^d} \sum_{Q \in \mathcal{T}'} f(c(Q)) 1_Q d\mu| \leq \varepsilon \mu([-N, N]).$$

where $c(Q)$ is, as usual, the center of Q and the measure of $[-N, N]$ appeared, since it is the actual domain of the integration above.

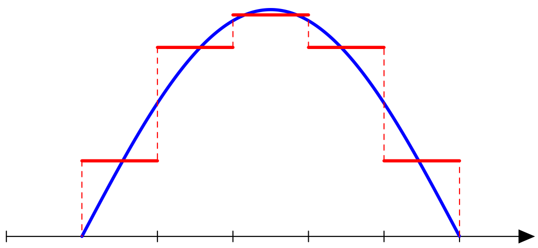


Figure 24: Function f(blue) and its simple approximation $\sum_{Q\in\mathcal{T}'} f(c(Q))1_Q$ (red).

The integral of the simple function is equal to

$$\sum_{Q'\in\mathcal{T}'} f(c(Q))\mu(Q).$$

For n big enough, by the previous part of the proof, we have

$$|\sum_{Q'\in\mathcal{T}'} f(c(Q))\mu(Q) - \sum_{Q'\in\mathcal{T}'} f(c(Q))\Lambda(f_n^Q)| \leq \varepsilon.$$

The above means that we reduced the situation to showing

$$|\Lambda(\sum_{Q'\in\mathcal{T}'} f(c(Q))f_n^Q) - \Lambda(f)| \leq \varepsilon.$$

Note that we used the additivity of Λ to pull it outside of the sum. Now we are going to make a use of the function h we introduced earlier. Notice that

$$\sum_{Q'\in\mathcal{T}'} f(c(Q))f_n^Q \leq f + \varepsilon h$$

and

$$f \leq \sum_{Q'\in\mathcal{T}'} f(c(Q))f_n^Q + \varepsilon h,$$

so we get

$$|\Lambda(\sum_{Q'\in\mathcal{T}'} f(c(Q))f_n^Q) - \Lambda(f)| \leq \varepsilon\Lambda(h).$$

We are left with arguing that μ is unique. Suppose that there exists a Radon measure $\tilde{\mu}$ such that

$$\Lambda(f) = \int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f d\tilde{\mu}.$$

Then in particular for any dyadic cube Q and n

$$\int_{\mathbb{R}^d} f_n^Q d\mu = \int_{\mathbb{R}^d} f_n^Q d\tilde{\mu},$$

so passing to the limit on both sides as $n \to \infty$ and using the Lebesgue dominated convergence

$$\int_{\mathbb{R}^d} 1_Q d\mu = \int_{\mathbb{R}^d} 1_Q d\tilde{\mu}$$

what gives

$$\tau(Q) = \mu(Q) = \tilde{\mu}(Q) = \tilde{\tau}(Q),$$

and we have already seen that if the generating functions of two measures are equal, then the generated measures coincide. $\square$

Let μ be the Lebesgue measure on $\mathbb{R}^d$ and $f\colon \mathbb{R}^d \to [0,\infty)$ a measurable function with the additional condition for $Q \in \mathcal{T}$

$$\int_{\mathbb{R}^d} f 1_Q d\mu < \infty.$$

Proposition 2.22. *The map τ generates a Radon measure ν.*

Proof. We shall check the martingale and the regularity conditions.

(1)
$$\int_{\mathbb{R}^d} f 1_Q d\mu = \sum_{\substack{Q' \subset Q \\ \text{ord of } Q' = \text{ord of } Q+1}} \int_{\mathbb{R}^d} f 1_{Q'} d\mu,$$

by the additivity of the integral.

(2) Note that the condition

$$\limsup_{\varepsilon \to 0} \sup_{\substack{E \subset \mathbb{R}^d \\ \mu(E) \le \varepsilon}} \int_{\mathbb{R}^d} f 1_Q 1_E d\mu$$

implies the regularity. Let us prove that this statement is actually true. Take a measurable E with $\mu(E) \leq \varepsilon$. Indeed, we have

$$\int_{\mathbb{R}^d} f 1_Q 1_E d\mu = \int_0^\infty \mu(f 1_Q 1_E > \lambda) d\lambda$$
$$\leq \int_0^\infty \min(\mu(E), \mu(f 1_Q > \lambda)) \leq \int_0^\infty \min(\varepsilon, \mu(f 1_Q > \lambda)) d\lambda.$$

The function inside the last integral is bounded from above by $\mu(f 1_Q > \lambda)$, which is integrable. Moreover, it is pointwise convergent to 0 as $\varepsilon \to 0$. Hence, we can use the Lebesgue dominated convergence theorem to see that as $\varepsilon \to 0$, the last integral goes to

$$\int_0^\infty 0 d\lambda = 0.$$

□

Example. Define $\tau \colon \mathcal{T} \to [0, \infty)$ as follows

$$\tau_x(Q) = \begin{cases} 1, \ x \in Q \\ 0, x \notin Q \end{cases}.$$

Denote by ν_x the Radon measure generated by τ_x. Let Q_x^k be the dyadic cube of order k that contains x. Here a dyadic cube of order (scale) k has side-length 2^{-k}. Notice that

$$\lim_{k \to \infty} \mu(Q_x^k) = 0$$

but

$$\lim_{k \to \infty} \nu_x(Q_x^k) = 1!$$

The previous example motivates the following definition of a measure *absolutely continuous* with respect to the Lebesgue measure.

Definition 2.23. A Radon measure ν is called absolutely continuous with respect to the Lebesgue measure μ if

$$\limsup_{\varepsilon \to 0} \left[\sup_{E \colon \mu(E) \leq \varepsilon} \nu(E) \right] = 0.$$

Theorem 2.24. *Let $f\colon \mathbb{R}^d \to [0,\infty)$ be a Lebesgue measurable function such that for any $Q \in \mathcal{T}$*

$$\int_{\mathbb{R}^d} f 1_Q d\mu < \infty.$$

Then we have for almost all $x \in \mathbb{R}^d$

$$\lim_{k\to\infty} \frac{\int_{\mathbb{R}^d} f 1_{Q_x^k} d\mu}{\mu(Q_x^k)} = f(x),$$

where Q_x^k is, as in the example before, the cube of order k that contains x.

End of lecture 11. November 26, 2015

Note that one cannot recover f at *every* point. Namely, if we change f on a set of measure zero, then τ remains unchanged. Therefore the above limit cannot equal $f(x)$ for all $x \in \mathbb{R}^d$. We also remark that not all Radon measures are given as ν in the theorem, i.e. by integration with respect to the Lebesgue measure. For instance, if ν is the Dirac measure, then for all $x \neq 0$ the above limit is zero. If it were given as stated in the theorem, then one would recover f at almost every point. This is not possible unless f is zero almost everywhere. But then τ would be zero for all dyadic cubes, which is a contradiction.

Proof. 0) It suffices to show the claim for functions of the form $f\mathbf{1}_{Q_0}$, where Q_0 is a cube of scale 0. That is, we may assume that f is supported on such a cube. Indeed, since cubes of any fixed scale partition $\mathbb{R}^d$ we can write

$$f = \sum_{Q \text{ of scale } 0} f\mathbf{1}_Q$$

and apply the theorem to $f\mathbf{1}_Q$. For each Q we obtain an exceptional set F_Q of measure zero such that for x in its complement the sequence $\nu(Q_x^k)/\mu(Q_x^k)$ converges to $f\mathbf{1}_Q(x)$. Then we define

$$F := \bigcup_{Q \text{ of scale } 0} F_Q$$

which has measure zero and for x in the complement of F, the above sequence converges. It converges to $f(x)$ since as soon as $k > 0$, the sequence is the same in case of f and $f\mathbf{1}_Q$. From now one we thus assume f is supported on a fixed cube Q_0.

1) Next we observe the claim if f is a finite linear combination of the characteristic functions of dyadic cubes. Let k be the highest of orders of these cubes. Then f is clearly constant on each dyadic cube of order k and the limit considered in the statement of the theorem stabilizes at k being equal to $f(x)$.

2) Let now f be the characteristic function of a measurable set $E \subset Q_0$. Claim: for each $\varepsilon > 0$ there exists an exceptional set F with $\mu(F) < \varepsilon$ and k_0 such that for all $k > k_0$ and for all $x \notin F$

$$\Big|\frac{\nu(Q_x^k)}{\mu(Q_x^k)} - f(x)\Big| \le \varepsilon$$

Once the claim is established we define

$$G_n := \bigcup_{n' \ge n} F_{2^{-n'}}$$

where $F_{2^{-n'}}$ is the exceptional set for $\varepsilon = 2^{-n'}$. Summing the geometric series we have $\mu(G_n) \le 2^{-n+1}$. If $x \notin G_n$, then $x \notin F_{2^{-n'}}$ for all $n' \ge n$. Thus for every $n' \ge n$ there exists k_0 such that for all $k \ge k_0$

$$\Big|\frac{\nu(Q_x^k)}{\mu(Q_x^k)} - f(x)\Big| \le 2^{-n'}$$

This shows that if $x \notin G_n$ for some n, then

$$\lim_{k\to\infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)} = f(x)$$

Define now

$$G := \bigcap_n G_n$$

and note that $\mu(G) = 0$. If $x \notin G$, then $x \notin G_n$ for some n and hence

$$\lim_{k\to\infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)} = f(x)$$

This shows the theorem.
Now we prove the claim. Let $< \varepsilon < 1$. By Littlewood's first principle we can choose a finite union of disjoint dyadic cubes $\mathcal{T}'$ such that for

$$F_1 := \Big(\bigcup_{\mathcal{T}'} Q'\Big) \Delta E$$

we have

$$\mu(F_1) \le \frac{\varepsilon^2}{2} \le \frac{\varepsilon}{2}$$

Let $\mathcal{T}''$ be the collection of maximal (with respect to set inclusion) dyadic cubes such that

$$\frac{\mu(F_1 \cap Q)}{\mu(Q)} \ge \varepsilon$$

Define

$$F_2 := \bigcup_{\mathcal{T}''} Q''$$

which is a disjoint union due to maximality of the cubes in $\mathcal{T}''$. We have

$$\mu(F_2) = \sum_{\mathcal{T}''} \mu(Q'') \overset{(1)}{\le} \sum_{\mathcal{T}''} \frac{1}{\varepsilon}\mu(F_1 \cap Q'') \overset{(2)}{\le} \frac{1}{\varepsilon}\mu(F_1) \overset{(3)}{\le} \frac{\varepsilon}{2}$$

where in (1) we use that $Q'' \in \mathcal{T}''$, in (2) disjointness of Q'' and in (3) that $\mu(F_1) \le \varepsilon^2/2$. Define now $F := F_1 \cup F_2$. Then $\mu(F) \le \varepsilon$. Denote by k_0 be the largest scale of the cubes in $\mathcal{T}'$. Let $k \ge k_0$ and $x \notin F$. Writing $\tilde{f} = \sum_{\mathcal{T}'} \mathbf{1}_{Q'}$ we have

$$\Big|\frac{\nu(Q_x^k)}{\mu(Q_x^k)} - f(x)\Big| \le \Big|\frac{\nu(Q_x^k)}{\mu(Q_x^k)} - \frac{\tilde{\nu}(Q_x^k)}{\mu(Q_x^k)}\Big| + \underbrace{\Big|\frac{\tilde{\nu}(Q_x^k)}{\mu(Q_x^k)} - \tilde{f}(x)\Big|}_{=0 \text{ (observe)}} + \underbrace{|\tilde{f}(x) - f(x)|}_{=0 \text{ since } \tilde{f}(x)=f(x) \text{ on } F_1^c}$$

Since $|\nu(Q_x^k) - \tilde{\nu}(Q_x^k)|$ is at most $\mu(F_1 \cap Q_x^k)$, the last display is bounded by

$$\frac{\mu(F_1 \cap Q_x^k)}{\mu(Q_x^k)} \le \frac{\varepsilon}{2}$$

The last inequality holds by maximality of the cubes in F_2. Indeed, if $x \notin F$ then $x \notin F_2$ and hence the above quotient is necessarily less than $\varepsilon/2$. This finishes the proof of the special case 2).

3) Let f be a measurable bounded function $f = f\mathbf{1}_{Q_0}$. Let $\varepsilon > 0$. We can approximate f by a finite linear combination of characteristic functions of dyadic cubes $\tilde{f}$ such that

$$\tilde{f} \le f \le \tilde{f} + \varepsilon \quad \text{on} \quad Q_0$$

Then we use 2) finitely many times. Denote by F the union of the exceptional sets one obtains in this process. We have $\mu(F) = 0$. If $x \notin F$, then

$$\limsup_{k\to\infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)} \le \varepsilon + \limsup_{k\to\infty} \frac{\tilde{\nu}(Q_x^k)}{\mu(Q_x^k)} \le \varepsilon + \tilde{f}(x) \le 2\varepsilon + f(x)$$

Similarly we show that

$$\liminf_{k\to\infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)} \geq f(x) - 2\varepsilon$$

This implies that the limit exists and it equals $f(x)$.

4) Let f be a measurable (not necessarily bounded) function $f = f\mathbf{1}_{Q_0}$. Let $n \in \mathbb{N}$ and let $\mathcal{T}'$ be the collection of maximal dyadic cubes such that

$$\frac{\nu(Q)}{\mu(Q)} \geq 2^n$$

Define the exceptional set $F_n = \bigcup_{\mathcal{T}'} Q'$. Then

$$\mu(F_n) \leq \sum_{\mathcal{T}'} \mu(Q') \leq \sum_{\mathcal{T}'} 2^{-n}\nu(Q') \leq 2^{-n}\nu(Q_0)$$

Choose $\tilde{f}$ such that $\tilde{f}(x) = f(x)$ if $x \notin F_n$ and such that $\tilde{f}$ is constant on $Q' \in \mathcal{T}'$ with $\int \tilde{f}\mathbf{1}_{Q'}d\mu = \int f\mathbf{1}_{Q'}d\mu$ if $x \in Q' \in \mathcal{T}$. Note that one can choose such a function. It is now an exercise that $\tilde{f}$ is bounded almost everywhere on F_n^c. Then one has to deduce that for every $\varepsilon > 0$ there is a set F of measure less than ε and k_0 such that for all $k \geq k_0$ and $x \notin F$

$$\Big|\frac{\nu(Q_x^k)}{\mu(Q_x^k)} - f(x)\Big| \leq \Big|\frac{\nu(Q_x^k)}{\mu(Q_x^k)} - \frac{\tilde{\nu}(Q_x^k)}{\mu(Q_x^k)}\Big| + \Big|\frac{\tilde{\nu}(Q_x^k)}{\mu(Q_x^k)} - \tilde{f}(x)\Big| + |\tilde{f}(x) - f(x)| \leq \varepsilon$$

□

If we replace ν by an arbitrary Radon measure, then we have the following theorem.

Theorem 2.25. *Let ν be a Radon measure on $\mathbb{R}^d$. Then*

$$\lim_{k\to\infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)}$$

exists for μ-almost all $x \in \mathbb{R}^d$.

Note that nothing is said about the value of the limit. It may also happen that it is zero almost everywhere as it is the case if ν is the Dirac measure.

Proof. By the same reasoning as in the previous proof it suffices to show the claim for $\tilde{\nu}$ of the form

$$\tilde{\tau} = \nu(Q \cap Q_0)$$

for some dyadic cube Q_0 of scale 0. So from now on assume $\nu = \tilde{\nu}$. Let $n \in \mathbb{N}$, $k \in \mathbb{Z}$ and we denote $[a, b) = [2^{-k}n, 2^{-k}(n+1))$. Define $F_{a,b}$ to be the set of all $x \in \mathbb{R}^d$ such that there exists a sequence k_n such that for all $n \in \mathbb{N}$

$$\frac{\nu(Q_x^{k_{2n}})}{\mu(Q_x^{k_{2n}})} \leq a \quad \text{and} \quad \frac{\nu(Q_x^{k_{2n+1}})}{\mu(Q_x^{k_{2n+1}})} \geq b$$

In some sense, the interval $[a, b)$ measures the convergence of the sequence as it measures the difference between its lim inf and lim sup. Our goal is to show that $\mu(F_{a,b}) = 0$ for all intervals $[a, b)$. Namely, if the limit does not exist, then necessarily

$$\liminf_{k \to \infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)} < \limsup_{k \to \infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)}$$

and hence there exits a, b of the above form such that

$$\liminf_{k \to \infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)} < a < b < \limsup_{k \to \infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)}$$

This implies $x \in F_{a,b}$. Thus, if this set has measure zero, then the limit exists for almost all $x \in \mathbb{R}^d$.

To prove $\mu(F_{a,b}) = 0$ we perform the following stopping time argument. Define $\mathcal{T}_1$ to be the collection of maximal dyadic cubes $Q \subset Q_0$ such that

$$\frac{\nu(Q)}{\mu(Q)} \geq b$$

Define $\mathcal{T}_2$ be the collection of maximal dyadic cubes Q such that there exists $Q' \in \mathcal{T}_1$ with $Q \subset Q'$ and

$$\frac{\nu(Q)}{\mu(Q)} \leq a$$

For any $n \in \mathbb{N}$ we define $\mathcal{T}_{n+1}$ as follows. If n is even, then we define it as the collection of all dyadic cubes such that there is $Q' \in \mathcal{T}_n$ with $Q \subset Q'$ and

$$\frac{\nu(Q)}{\mu(Q)} \geq b$$

while for odd n we set $\mathcal{T}_{n+1}$ to be the collection of all dyadic cubes such that there exists $Q' \in \mathcal{T}_n$ with $Q \subset Q'$ and

$$\frac{\nu(Q)}{\mu(Q)} \leq a$$

Now we set
$$F_n := \bigcup_{\mathcal{T}_n} Q$$
which is disjoint by maximality of the cubes in $\mathcal{T}_n$. Observe that
$$F_{a,b} \subset \bigcap_n F_n$$
It remains to show that $\limsup_{n\to\infty} \mu(F_n) = 0$. If n is odd, then we have
$$\sum_{Q\in\mathcal{T}_{n+1}} \mu(Q) \le \sum_{Q'\in\mathcal{T}_n} \sum_{Q\subset Q',\, Q\in\mathcal{T}_{n+1}} \mu(Q) \le \sum_{Q'\in\mathcal{T}_n} \mu(Q')$$
If n is even, then
$$\begin{aligned}\sum_{Q\in\mathcal{T}_{n+1}} \mu(Q) &\le \sum_{Q'\in\mathcal{T}_n} \sum_{Q\subset Q',\, Q\in\mathcal{T}_{n+1}} \mu(Q)\\ &\le \sum_{Q'}\sum_{Q} \frac{1}{b}\nu(Q)\\ &\le \sum_{Q'} \frac{a}{b}\mu(Q')\\ &\le (1-\varepsilon)\sum_{Q'}\mu(Q')\end{aligned}$$
By induction we see that
$$\mu(F_n) \le (1-\varepsilon)^{n/2}\mu(Q_0)$$
Letting $n\to\infty$ finishes the proof. □

End of lecture 12. December 1, 2015

2.3 Structure of Radon measures

Let ν be a Radon measure in $\mathbb{R}^d$. Last time we showed that the limit
$$g(x) := \lim_{k\to\infty} \frac{\nu(Q_x^k)}{\mu(Q_x^k)}$$
exists μ-almost everywhere. The set of all points for which the above limit exists we call the Lebesgue points. Let us put $g(x) := 0$ for all x such that the above limit does not exist or is equal to ∞. Observe that if
$$v(Q) = \int f 1_Q d\mu \text{ for all } Q \in \mathcal{T},$$
then one obtains $g(x) = f(x)$ for μ almost all x.

Proposition 2.26. *If ν is a Radon measure in $\mathbb{R}^d$, then we have for all $Q \in \mathcal{T}$*

$$\int g 1_Q d\mu \leq \nu(Q).$$

Proof. Define

$$g_k = \sum_{Q':Ord(Q')=k} \frac{\nu(Q')}{\mu(Q')} 1_{Q'}$$

Clearly g_k is a measurable function. Notice that

$$g \leq \liminf_k g_k$$

and the function on the right hand side is again measurable since it is the liminf of a sequence measurable functions. Applying Fatou's lemma we estimate

$$\begin{aligned}\int g 1_Q d\mu \leq \int \liminf_k g_k 1_Q d\mu &\overset{\text{Fatou}}{\leq} \liminf_k \int g_k 1_Q d\mu \\ = \liminf_k \sum_{\substack{Q' \subset Q \\ Ord(Q')=k}} \int \frac{\nu(Q')}{\mu(Q')} 1_{Q'} d\mu &= \liminf_k \sum_{\substack{Q' \subset Q \\ Ord(Q')=k}} \nu(Q') = \nu(Q).\end{aligned}$$

□

We will now use the function g to decompose our Radon measure ν into two parts of a character quite different from each other. Let ν_1 be the Radon measure generated by g, i.e. satisfying for any $Q \in \mathcal{T}$ the relation

$$\nu_1(Q) = \int g 1_Q d\mu.$$

Then by the previous proposition for any $Q \in \mathcal{T}$ we have the inequality

$$\nu_1(Q) \leq \nu(Q).$$

Example. If $\nu = \delta_0$ is the Dirac delta concentrated at 0, then $g(x) = 0$ for all the points x different from 0 and the limit at 0 is equal to ∞, thus $g(x) = 0$ for all x. Note that in this case with the notation above we have $\nu_1 = 0$ and $\nu - \nu_1 = \delta_0$, what shows that the inequality $\nu_1 \leq \nu$ may be strict!

Notice that ν_1 is absolutely continuous with respect to μ and put $\nu_2 := \nu - \nu_1$. ν_2 is a non-negative measure, because ν dominates ν_1.

Definition 2.27. A Radon measure ν is called singular with respect to μ if there exists a set E with $\mu(E) = 0$ and such that for any $F \subset \mathbb{R}^d$, $\nu(F \cap E) = \nu(F)$.

Remark. In the previous example $\nu = \delta_0$ is singular with respect to the Lebesgue measure and $E = \{0\}$.

Theorem 2.28. *ν_2 is singular with respect to μ.*

Proof. Let E be the set of the Lebesgue points of ν. Let $E_1 \subset E$ be the set of points for which

$$g(x) = \lim_k \frac{\int g 1_{Q_x^k} d\mu}{\mu(Q_x^k)}.$$

Put $E_2 = \mathbb{R}^d \setminus E_1$, $\mu(E_2) = 0$. We know E is a set of full measure and the set of the points for which $g(x)$ is equal to the limit above has also full measure. The intersection of two sets of full measure is still a set of full measure - this shows that $\mu(E_2) = 0$. Thus, we are left with proving that for an arbitrary $F \subset \mathbb{R}^d$, $\nu_2(F \cap E_2) = \nu_2(F)$.
Let $x \in E_1$, we have

$$\liminf_k \frac{\nu_2(Q_x^k)}{\mu(Q_x^k)} = \lim_k \frac{\nu(Q_x^k)}{\mu(Q_x^k)} - \lim_k \frac{\nu_1(Q_x^k)}{\mu(Q_x^k)} = g(x) - g(x) = 0.$$

Fix a cube Q_0 of order 0. Let $\varepsilon > 0$ and $x \in E_1 \cap Q_0$. There exists a cube Q_x containing x with

$$\frac{\nu_2(Q_x)}{\mu(Q_x)} < \varepsilon.$$

Define $\mathcal{T}'$ to be the collection of such Q_x's that are maximal with respect to inclusion. Notice that $Q \cap E_1 \subset \bigcup \mathcal{T}'$ and the cubes in $\mathcal{T}'$ are disjoint, hence

$$\nu_2(Q_0 \cap E_1) \leq \sum_{Q' \in T'} \nu_2(Q') \leq \varepsilon \sum_{Q' \in \mathcal{T}'} \mu(Q') \leq \varepsilon \mu(Q_0).$$

$\varepsilon > 0$ was arbitrary what gives that $\nu_2(Q_0 \cap E_1) = 0$. Moreover, Q_0 was arbitrary, so $\nu_2(E_1) = 0$. This however, implies that for any $F \subset R^d$, $\nu_2(F \cap E_2) = \nu_2(F)$ and finishes the proof. $\square$

Theorem 2.29. *Let μ and ν be Radon measures on $\mathbb{R}^d$. There exist two Radon measures ν_1 adn ν_2 with $\nu = \nu_1 + \nu_2$ such that ν_1 is absolutely continuous with respect to μ and ν_2 is singular with respect to μ. The measures ν_1 and ν_2 in the decomposition are unique.*

Proof. We prove the existence the same way we did for the Lebesuge case. Let us prove the uniqueness. Suppose that

$$\nu = \nu_1 + \nu_2 = \tilde{\nu_1} + \tilde{\nu_2}$$

are two decomposition of ν with the demanded properties. Substracting ν_2 and $\tilde{\nu_1}$ on the both sides we obtain

$$\nu_1 - \tilde{\nu_1} = \tilde{\nu_2} - \nu_2.$$

Observe that the measure on the right hand side is supported on the $E_2 \cup \tilde{E}_2$ which is has μ-measure zero. Thus, for any F

$$(\tilde{\nu_2} - \nu_2)(F) = (\tilde{\nu_2} - \nu_2)(F \cap (E_2 \cup \tilde{E}_2)) = (\tilde{\nu_1} - \nu_1)(F \cap (E_2 \cup \tilde{E}_2)) = 0,$$

because both ν_1, $\tilde{\nu_1}$ are absolutely continuous. That implies $\tilde{\nu_2} = \nu_2$ and $\tilde{\nu_1} = \nu_1$. □

A signed Radon measure is of the form $\nu_1 - \nu_2$ for two non-negative Radon measures ν_1, ν_2 with $\nu_1(\mathbb{R}^d) < \infty$, $\nu_2(\mathbb{R}^d) < \infty$.

Theorem 2.30 (Jordan-Hahn decomposition)**.** *Let ν be a signed Radon measure. Then there exist two disjoint ν-measurable sets E_+ and E_- with $E_- \cup E_+ = \mathbb{R}^d$ and such that for all $F \subset \mathbb{R}^d$*

$$\nu(F \cap E_+) \geq 0 \text{ and } \nu(F \cap E_-) \leq 0.$$

Remark. Note that if ν was generated by a function, then we could simply take the positive and the negative part of the function to obtain the decomposition. That is the idea behind the proof of the theorem.

Proof. Let $\nu = \nu_1 - \nu_2$ as remarked before the statement of the theorem and let $\mu = \nu_1 + \nu_2$. Note that if for a set E, $\mu(E) = 0$, then $\nu_1(E), \nu_2(E) \leq \mu(E) = 0$, so both ν_1 and ν_2 are absolutely continuous with respect to μ. For $i = 1, 2$ choose a function f_i such that

$$\nu_i(Q) = \int f_i 1_Q d\mu.$$

This gives us the possibility of writing ν as

$$\nu(Q) = \int (f_1 - f_2) 1_Q d\mu.$$

Define E_+ and E_- followingly

$$E_+ := \{x \colon (f_1 - f_2)(x) \geq 0\},$$

$$E_- := \{x\colon (f_1 - f_2)(x) < 0\},$$

Both of them are μ-measurable and clearly we have $\nu(F \cap E_+) \geq 0$ and $\nu(F \cap E_+) \leq 0$ for any $F \subset \mathbb{R}^d$. We just need to verify that E_+, E_- are ν-measurable. Let $Q \in \mathcal{T}$:

$$\begin{aligned}\nu(Q \cap E_+) + \nu(Q \cap E_-) &= \int (f_1 - f_2)1_{E_+}1_Q d\mu + \int (f_1 - f_2)1_{E_-}1_Q d\mu \\ &= \int (f_1 - f_2)1_Q d\mu = \nu(Q).\end{aligned}$$

□

We have already shown that there exists the unique decomposition of a Radon measure into the singular and the absolutely continuous part. We are going to decompose it further, namely split the singular part into a *pure point measure* and a *singular continuous measure*.

Definition 2.31. A Radon measure ν is called a pure point measure if it is of the form

$$\nu(Q) = \sum_{x_i \in Q} \lambda_i$$

for a set of points Q and real numbers λ_i.

Let ν be an arbitrary Radon measure. Define

$$X := \{x\colon \overbrace{\lim_k \nu(Q_x^k)}^{\lambda_x} > 0\}.$$

Observe that the limit inside the brackets is well defined since the sequence $\nu(Q_x^k)$ is montonically decreasing. Moreover, $X = \bigcup_n X_n$, where

$$X_n := \{x\colon \lim_k \nu(Q_x^k) > 2^{-n}\}.$$

Lemma 2.32. *For $Q \in T$*

$$\sum_{x \in X_n \cap Q_0} \lambda_x \leq \nu(Q)$$

Proof. Suppose that the statement of the theorem does not hold. Choose a finite subset $\{x_i\}$ such that

$$\sum_i \lambda_{x_i} > \nu(Q)$$

and choose an integer k such that

$$2^{-k} \leq \min_{i,j}(|x_i - x_j|).$$

With such choice of k the sets $Q^k_{x_i}$ are pairwise disjoint, so

$$\sum_i \lambda_{x_i} \leq \sum_i \nu(Q^k_{x_i}) \leq \nu(Q).$$

This is a contradiction. $\square$

In particular, the last proof gives us that the sets X_n and X are countable. Passing to the supremum

$$\tau_{pp}(Q) := \sum_{x \in X \cap Q} \lambda_x = \sup_n \sum_{x \in X_n \cap Q} \lambda_x \leq \nu(Q)$$

For $Q \in \mathcal{T}$. Hence, τ_{pp} generates a Radon measure ν_{pp} such that for all $Q \in T$

$$\nu_{pp}(Q) \leq \nu(Q).$$

Moreover, one can observe that $\nu_{pp} \leq \nu_s$, where ν_s is the singular part of the decomposition singular+absolutely continuous, because (ν_{ac} is the absolutely continuous part)

$$\lim_k \nu(Q^k_x) = \lim_k \nu_s(Q^k_x) + \lim_k \nu_{ac}(Q^k_x),$$

and the rightmost limit is equal to 0. Thus, we can write

$$\nu_s = \nu_{sc} + \nu_{pp},$$

where ν_{pp} is the pure point measure defined above and ν_{sc} is a non-negative Radon measure, which we call singular continuous.

Definition 2.33. A singular with respect to μ measure ν is called singular continuous with respect to μ if $\lim_k \nu(Q^k_x) = 0$ for all x.

After the considerations above we arrive at the following theorem.

Theorem 2.34. *Let μ be the Lebesgue measure and ν be a Radon measure. Then there exists the unique decomposition*

$$\nu = \nu_{ac} + \nu_{sc} + \nu_{pp},$$

with ν_{ac} absolutely continuous with respect to μ, ν_{sc} singular continuous with respect to μ and ν_{pp} a pure point measure.

Remark. Observe that $L^1(\mathbb{R}^d)$ is equal to the set of signed Radon measures absolutely continuous with respect to μ. Two elements of this space are equivalent if they differ from each other only on a set of measure zero.

End of lecture 13. December 3, 2015

2.4 L^p spaces

The space $L^1_+(\mathbb{R}^d)$ is defined as

the set A of all absolutely continuous (non-negative) Radon measures ν with $\nu(\mathbb{R}^d) < \infty$

or, equivalently,

the set B of all equivalence classes of measurable functions $f : \mathbb{R}^d \to [0, \infty)$ with

$$\int_{\mathbb{R}^d} f d\mu < \infty,$$

where $f \sim g$ if $f(x) = g(x)$ for μ-almost all x.

We restrict our attention to non-negative functions an measures in order to have simpler notation. We already know that A and B are in bijective correspondence and hence the definitions are equivalent. Namely, if $\nu \in A$, then for

$$f_k := \sum_{\mathrm{ord}(Q)=k} \frac{\nu(Q)}{\mu(Q)} \mathbf{1}_Q$$

the limit $\lim_{k\to\infty} f_k(x)$ exists for almost all x. We define $f(x)$ to be this limit, whenever it exists, and 0 otherwise. If ν is absolutely continuous, then $\nu(Q) = \int f \mathbf{1}_Q d\mu$ and $\int f d\mu = \nu(\mathbb{R}^d) < \infty$. On the other hand, if we take an equivalence class in B and a representative f, then we define $\tau(Q) = \int f \mathbf{1}_Q d\mu$ which generates a Radon measure (exercise). Note that τ is independent of the chosen representative.
However, the space L^1 is not suitable for a product theory. To elaborate on this we first need to observe that one can define product on equivalence classes. Indeed, if $f_1 \sim f_2$ and $g_1 \sim g_2$, then $f_1 g_1 \sim f_2 g_2$ since they differ at most on the union of the sets where f_1 differs from f_2 and g_1 differs from g_2. This union has measure zero. Now, if[10] $f, g \in L^1(\mathbb{R}^d_+)$, then the product fg is in general not in L^1. To see this we consider

$$f(x) = \frac{1}{|x|^{2/3}} \mathbf{1}_{[-1,1]}(x)$$

which is integrable, i.e. belongs to $L^1_+(\mathbb{R}^d)$. However,

$$f^2(x) = \frac{1}{|x|^{4/3}} \mathbf{1}_{[-1,1]}(x)$$

[10]From now on we identify a function with its equivalence class

is not integrable. (See Figure 2.4.)

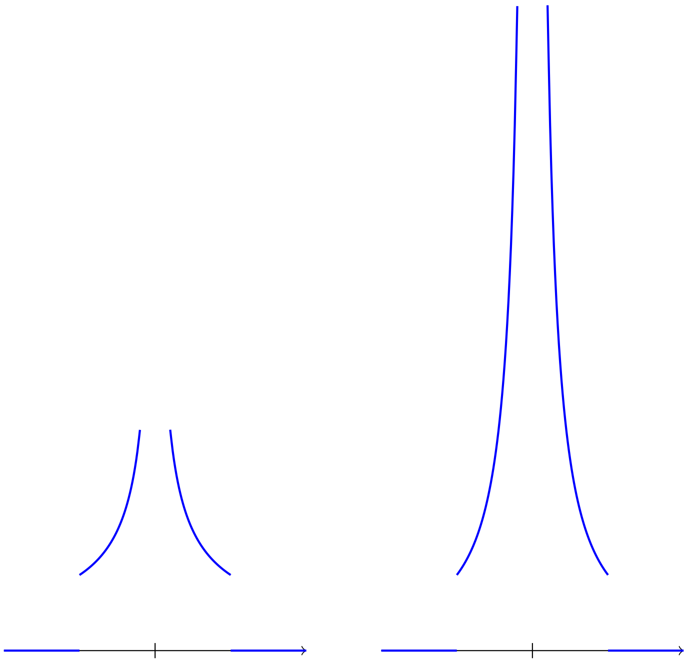


Figure 25: Functions f (left) and f^2 (right).

So the L^1_+ itself does not control products of functions. Because of that we shall introduce spaces L^p_+, $1 \leq p \leq \infty$. Although products of elements in L^p_+ will in general not be in L^p_+, we shall see that if $f \in L^{p_1}_+$ and $g \in L^{p_2}_+$ for some p_1, p_2, then $fg \in L^{p_3}_+$ for some p_3.
The space L^∞_+ is defined as the set of all non-negative essentially bounded measurable functions, i.e. bounded up to a set of measure zero. This is the only L^p_+ space which is closed under multiplication.

For $p \in \mathbb{R}$ with $1 < p < \infty$ we define $L^p_+(\mathbb{R}^d)$ to be

the set A of all (non-negative) Radon measures ν with

$$\sup_k \sum_{\mathrm{ord}(Q)=k} \left(\frac{\nu(Q)}{\mu(Q)}\right)^p \mu(Q) < \infty \tag{8}$$

or, equivalently,

the set B of all equivalence classes of measurable functions $f : \mathbb{R}^d \to [0, \infty)$

with[11]

$$\int_{\mathbb{R}^d} f^p d\mu < \infty$$

Note that in contrast with $p = 1$, in A we do not require absolute continuity of the measure. We shall see that this already follows from the other assumption.
In Analysis II we already considered $p = 2$ when we defined the Hilbert space $L^2(\mathbb{R})$ via A. A construction via B was then not available as we did not have the Lebesgue integration theory.
Before proving equivalence of both definitions we turn our attention to the following inequality.

Hölder's inequality. Let $a_i, b_i \in [0, \infty)$, $i = 1, \ldots, N$, and $\frac{1}{p} + \frac{1}{q} = 1$, $1 < p, q < \infty$. Then

$$\sum_{i=1}^N a_i b_i \le \Big(\sum_{i=1}^N a_i^p\Big)^{1/p} \Big(\sum_{i=1}^N b_i^q\Big)^{1/q}$$

Proof of Hölder's inequality. Without loss of generality we may assume $\sum_{i=1}^N a_i^p = 1$ and $\sum_{i=1}^N b_i^q = 1$, otherwise we replace a_i, b_i by

$$\frac{a_i}{\Big(\sum_{i=1}^N a_i^p\Big)^{1/p}} \quad \text{and} \quad \frac{b_i}{\Big(\sum_{i=1}^N b_i^q\Big)^{1/q}}.$$

(This follows from homogeneity, i.e. $\sum(\lambda a_i) b_i = \lambda \sum a_i b_i$ and $(\sum(\lambda a_i)^p)^{1/p} = \lambda(\sum a_i^p)^{1/p}$ for $\lambda > 0$). For $a, b > 0$ have

$$a^{\frac{1}{b}} b^{\frac{1}{q}} \le \frac{1}{p} a + \frac{1}{q} b$$

For $p = q = 2$ this is just the arithmetic-geometric mean inequality. To see the above inequality we observe that it is equivalent to

$$e^{\frac{1}{p}\ln a + \frac{1}{q}\ln b} \le \frac{1}{p} e^{\ln a} + \frac{1}{q} e^{\ln b}$$

which holds by convexity of the exponential function (see Figure 26). To finish the proof of Hölder's inequality we now write

$$\sum_{i=1}^N a_i b_i = \sum_{i=1}^N (a_i^p)^{\frac{1}{p}} (b_i^q)^{\frac{1}{q}} \le \sum_{i=1}^N \frac{1}{p} a_i^p + \frac{1}{q} b_i^q \overset{(*)}{\le} \frac{1}{p} + \frac{1}{q} = 1$$

where $(*)$ holds by our normalization. $\square$

[11] Recall that $f^p = e^{p \ln f}$

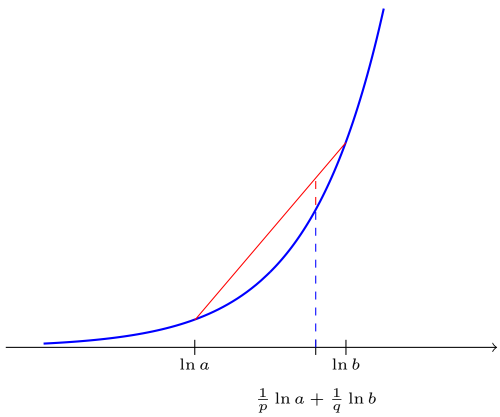


Figure 26: Convexity of e^x.

Hölder's inequality holds for countable sums as well and also for integrals. In this case it takes the following form. If f, g are measurable, then

$$\int fg d\mu \leq \Big(\int f^p d\mu\Big)^{\frac{1}{p}} \Big(\int g^q d\mu\Big)^{\frac{1}{q}}$$

Note that measurability is necessary since in the last step of the proof $(*)$ we need additivity of the integral.

Now we are ready to show equivalence of definitions via A and B. If $\nu \in A$, then it is absolutely continuous. Indeed, let $\mathcal{T}'$ be any disjoint collection of dyadic cubes. For $1 < p, q < \infty$, $1/p + 1/q = 1$ we write

$$\begin{aligned}\sum_{Q\in\mathcal{T}'} \nu(Q) &= \sum_{Q\in\mathcal{T}'} \frac{\nu(Q)}{\mu(Q)} \mu(Q)^{1/p} \mu(Q)^{1/q} \\ &\overset{\text{Hoelder}}{\leq} \Big(\sum_{Q\in\mathcal{T}'} \Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q)\Big)^{\frac{1}{p}} \Big(\sum_{Q\in\mathcal{T}'} \mu(Q)\Big)^{\frac{1}{q}}\end{aligned}$$

The claim is now that

$$\sum_{Q\in\mathcal{T}'} \Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q) \leq \sup_k \sum_{\mathrm{ord}(Q')=k} \Big(\frac{\nu(Q')}{\mu(Q')}\Big)^p \mu(Q') = C \leq \infty$$

where boundedness by a finite constant C follows by the definition of A. Note that on the left hand-side the sum is over any disjoint collection of cubes $\mathcal{T}'$, while on the right hand-side the sums are over collections of cubes of the same order. The claim implies

$$\sum_{Q\in\mathcal{T}'} \nu(Q) \leq C \Big(\sum_{Q\in\mathcal{T}'} \mu(Q)\Big)^{\frac{1}{q}}.$$

which then implies absolute continuity of ν (this is not hard to see).

To see the claim it suffices to consider finite collections $\mathcal{T}'$ since the bound is independent of the collection. Then we may assume that the order of $Q \in \mathcal{T}'$ is at most k. We need to show that if we partition $Q \in \mathcal{T}'$ into its children, the sum on the left hand-side increases. Then we keep partitioning until all cubes in the collection are of the same order. Since at each step the sum does not decrease, we are done.

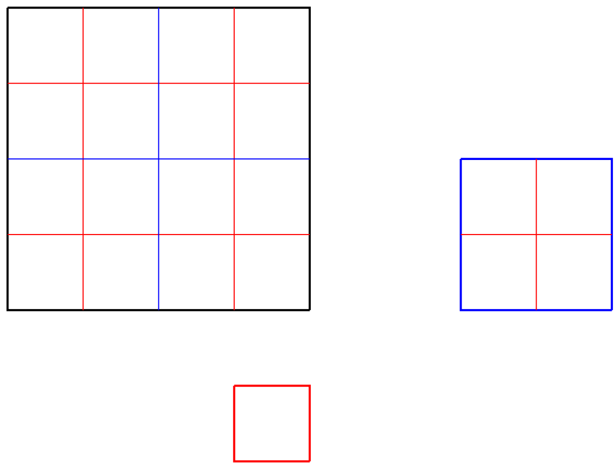


Figure 27: Partitioning of $Q \in \mathcal{T}'$ (all cubes are half open).

To see that at the partitioning the sum increases we write

$$\begin{aligned}\nu(Q) = \sum_{\substack{Q' \subset Q \\ k'=k+1}} \nu(Q') &= \sum_{\substack{Q' \subset Q \\ k'=k+1}} \frac{\nu(Q')}{\mu(Q')} \mu(Q')^{\frac{1}{p}} \mu(Q')^{\frac{1}{q}} \\ &\leq \Big(\sum_{\substack{Q' \subset Q \\ k'=k+1}} \Big(\frac{\nu(Q')}{\mu(Q')}\Big)^p \mu(Q')\Big)^{\frac{1}{p}} \underbrace{\Big(\sum_{\substack{Q' \subset Q \\ k'=k+1}} \mu(Q')\Big)^{\frac{1}{q}}}_{\mu(Q)^{\frac{1}{q}}}\end{aligned}$$

Since $\frac{1}{p} + \frac{1}{q} = 1$, i.e. $1 - p = -\frac{p}{q}$, this shows

$$\Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q) \leq \sum_{\substack{Q' \subset Q \\ k'=k+1}} \Big(\frac{\nu(Q')}{\mu(Q')}\Big)^p \mu(Q')$$

as desired. This in particular show that the sequence in (8) is monotonously increasing and so the supremum equals the limit as $k \to \infty$.

Now we return to showing equivalence of the definitions of L^p_+. We have just seen that if $\nu \in A$, then it is absolutely continuous. Define

$$f_k := \sum_{\mathrm{ord}(Q)=k} \frac{\nu(Q)}{\mu(Q)} \mathbf{1}_Q$$

Then $\lim_{k\to\infty} f_k(x)$ exists a.e. We set $f(x)$ to be that limit, whenever it exists, and 0 otherwise. We also set

$$\nu(Q) = \int f \mathbf{1}_Q d\mu$$

It remains to check that f is p-integrable:

$$\begin{aligned}
\int f^p d\mu = \int \liminf_{k\to\infty} f_k^p &\overset{(1)}{\leq} \liminf_{k\to\infty} \int f_k^p \\
&= \liminf_{k\to\infty} \sum_{\mathrm{ord}(Q)=k} \Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q) \\
&\overset{(2)}{=} \lim_{k\to\infty} \sum_{\mathrm{ord}(Q)=k} \Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q) \\
&\overset{(3)}{\leq} C < \infty
\end{aligned}$$

where (1): Fatou (2) monotonously increasing (3) A.

However, we also observe the reverse inequality. We write

$$\begin{aligned}
\frac{\nu(Q)}{\mu(Q)} = \frac{\int f \mathbf{1}_Q d\mu}{\mu(Q)} &= \frac{\int f \mathbf{1}_Q \mathbf{1}_Q d\mu}{\mu(Q)} \\
&\leq \frac{\Big(\int f^p \mathbf{1}_Q d\mu\Big)^{\frac{1}{p}} \Big(\int \mathbf{1}_Q d\mu\Big)^{\frac{1}{q}}}{\mu(Q)} \\
&= \Big(\frac{\int f^p \mathbf{1}_Q d\mu}{\mu(Q)}\Big)^{\frac{1}{p}}
\end{aligned}$$

Then

$$\Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q) \leq \int f^p \mathbf{1}_Q d\mu$$

Summing over Q of order k we obtain

$$\sum_{\mathrm{ord}(Q)=k} \Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q) \leq \int f^p d\mu$$

This shows that we actually have

$$\int f^p d\mu = \lim_{k\to\infty} \sum_{\mathrm{ord}(Q)=k} \Big(\frac{\nu(Q)}{\mu(Q)}\Big)^p \mu(Q)$$

We denote

$$\|f\|_p := \Big(\int f^p d\mu\Big)^{\frac{1}{p}}$$

Let now $f \in B$. Then

$$\int f\mathbf{1}_Q d\mu \le \Big(\int f^p \mathbf{1}_Q d\mu\Big)^{\frac{1}{p}} \Big(\int \mathbf{1}_Q d\mu\Big)^{\frac{1}{q}} < \infty$$

and we set this integral to be $\tau(Q)$. It is an exercise to check that τ generates a Radon measure.

Note that with the above introduced notation, Hölder's inequality reads

$$\int fg d\mu \le \|f\|_p \|g\|_q$$

Minkowski's inequality Let $1 \le p < \infty$, $f, g \in L^p_+(\mathbb{R}^d)$. Then $f + g \in L^p_+(\mathbb{R}^d)$ and

$$\|f+g\|_p \le \|f\|_p + \|g\|_p$$

Proof. If $p = 1$ we have equality since

$$\int f + g = \int f + \int g$$

Let now $1 < p < \infty$. Then there exists $1 < q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$. We have

$$(f+g)^p = f(f+g)^{p-1} + g(f+g)^{p-1}$$

and hence

$$\begin{aligned}
\|f+g\|_p^p &= \int (f+g)^p d\mu \\
&= \int f(f+g)^{p-1} d\mu + \int g(f+g)^{p-1} d\mu \\
&\le \|f\|_p \|(f+g)^{p-1}\|_q + \|g\|_p \|(f+g)^{p-1}\|_q \\
&= (\|f\|_p + \|g\|_p)\|(f+g)^{p-1}\|_q \\
&= (\|f\|_p + \|g\|_p)\|f+g\|_p^{p-1}
\end{aligned}$$

which shows the claim. The last inequality is deduced by

$$\|(f+g)^{p-1}\|_q = \Big(\int (f+g)^{q(p-1)}\Big)^{\frac{1}{q}} = \|f+g\|_p^{\frac{p}{q}} = \|f+g\|_p^{p-1}$$

since $q(p-1) = p$. $\square$

From now on we always assume $1 \leq p < \infty$.
The spaces $L_+^p(\mathbb{R}^d)$ are metric spaces with the metric

$$d(f,g) = \||f-g|\|_p$$

We show that this is indeed a metric.

1. $d(f,g) = d(g,f)$ is clear
2. $d(f,h) = \||f-h|\|_p \leq \||f-g|+|g-h|\|_p \leq \||f-g|\|_p + \||g-h|\|_p = d(f,g) + d(g,h)$
3. $d(f,f) = \|0\|_p = 0$. If $d(f,g) = 0$, then we need to show that $f - g \sim 0$. Assume that $d(f,g) = 0$. Then $\||f-g|\|_p = 0$ which means that

$$\int_0^\infty \mu(|f-g|^p > \lambda) d\lambda = 0$$

Since $\mu(|f-g|^p > \lambda)$ is a monotonously decreasing function, for all $\lambda > 0$ we must have

$$\mu(|f-g|^p > \lambda) = 0$$

Since $\mu(|f-g|^p > 0) = \cup_{k\in\mathbb{Z}} \mu(|f-g|^p > 2^{-k})$ and the union is countable, we see that

$$\mu(|f-g|^p > 0) = 0$$

But this means that $f(x) = g(x)$ for almost all x and hence f and g belong to the same equivalence class.

The spaces $L_+^p(\mathbb{R}^d)$ are complete. To see this let f_n be a Cauchy sequence in $L_+^p(\mathbb{R}^d)$. Consider a subsequence which converges rapidly, i.e. a subsequence f_{n_m} such that

$$\|f_{n_{m+1}} - f_{n_m}\|_p \leq 2^{-m}$$

for all $m \geq 1$. Set now

$$f_{n_{M+1}} := f_{n_1} + \sum_{m=1}^{M} f_{n_{m+1}} - f_{n_m}$$

$$g_{n_{M+1}} := |f_{n_1}| + \sum_{m=1}^{M} |f_{n_{m+1}} - f_{n_m}|$$

By Minkowski's inequality we have

$$\|g_{n_{M+1}}\|_p \leq \|f_{n_1}\|_p + \sum_{m=1}^{M} \|f_{n_{m+1}} - f_{n_m}\|_p < C < \infty$$

with C independent of m. Since $g_{n_M}^p$ is monotonously increasing, by the monotone convergence theorem

$$\int g^p = \lim_{M\to\infty} \int g_{n_M}^p = \lim_{M\to\infty} \|g_{n_M}\|_p^p < \infty$$

where we set $g := \lim_{k\to\infty} g_{n_M}$. Since $|f_{n_M}|^p \leq g^p$, by the dominated convergence theorem $\lim_{M\to\infty} f_{n_M} =: f$ belongs to L_+^p. Since $|f - f_{n_m}|^p \leq (2g)^p$, by the dominated convergence $\|f - f_{n_M}\|_p \to 0$ as $M \to \infty$. Let now $\varepsilon > 0$. There exists N such that for all $n, n' > N$ we have $\|f_n - f_{n'}\|_p < \varepsilon/2$. Choose n_m such that $\|f_{n_m} - f\|_p < \varepsilon/2$. Then for all $n > N$

$$\|f_n - f\|_p \leq \|f_n - f_{n_m}\|_p + \|f_{n_m} - f\|_p < \varepsilon$$

$\square$

End of lecture 14. December 8, 2015

Now we are going to see that the class of functions constant on the cubes of order k, for some $k \in \mathbb{Z}$, is dense in $L_+^p(\mathbb{R}^d)$ for $1 < p < \infty$. Writing this out in terms of Radon measures we have the following.

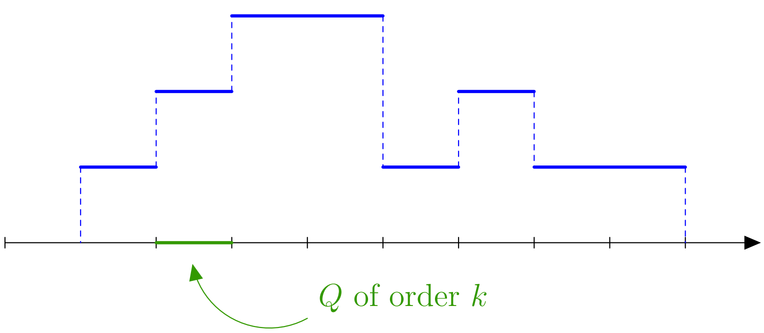


Figure 28: An example of a function constant on dyadic cubes of order k on $\mathbb{R}$.

Theorem 2.35. *Let $1 < p < \infty$. The set of all measures $\nu \in L^p_+(\mathbb{R}^d)$ satisfying the condition:*

$$\textit{There exists a } k \in \mathbb{Z} \textit{ with } \frac{\nu(Q)}{\mu(Q)} = \frac{\nu(Q')}{\mu(Q')}, \quad \textit{if } Q' \subset Q,\ Ord(Q) = k,$$

is dense in $L^p_+(\mathbb{R}^d)$.

Proof. Let $f \in L^p_+$. Define a sequence

$$f_k := \sum_{Q:\, Ord(Q)=k} \frac{\int f 1_Q d\mu}{\mu(Q)} 1_Q.$$

At this point, our goal to show that

$$\limsup_{k\to\infty} d(f, f_k) = 0,$$

i.e. we need to prove

$$\limsup_{k\to\infty} \int |f - f_k|^p d\mu = 0.$$

Notice that f_k converges to f pointwise almost everywhere. Since we wish to apply the Lebesgue dominated convergence theorem, we should validate that

$$\int (\sup_k f_k)^p d\mu < \infty.$$

Because of the bound

$$|f - f_k|^p \leq 2^p(f_k^p + f^p) \leq 2^p(\sup_k (f_k)^p + f^p),$$

this would give that our sequence is dominated by an integrable function and would let us to apply the theorem, since we have $\sup_k (f_k)^p = (\sup_k f_k)^p$. Let us then prove the boundedness of $\sup_k f_k$ in L^p_+.

$$\begin{aligned}\int (\sup_k f_k)^p d\mu &= \int_0^\infty \mu((\sup_k f_k)^p > \lambda) d\lambda \\ &\overset{\text{change of var.}}{=} 2^p \int_0^\infty \mu(\sup_k f_k > 2\lambda^{1/p}) d\lambda.\end{aligned}$$

Given a function g, we decompose it as follows

$$g = g_1 + g_2, \text{ with } g_1 = \min(g, \lambda^{1/p}).$$

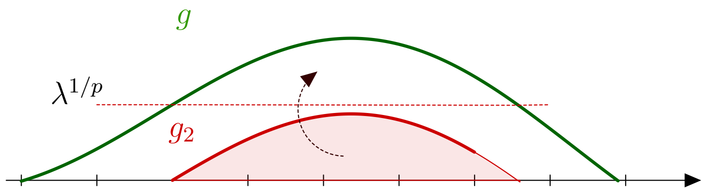


Figure 29: The idea of splitting a function g into g_1+g_2. Function g_1 is equal to function g(green) to the left of the left intersection point of the red dotted line (which denotes the level $\lambda^{1/p}$) with the graph of g and to the right of the right intersection point; it is equal to the red dotted segment between the intersection points.

Notice that if at some point $\sup_k f_k > 2\lambda^{1/p}$, then either $\sup_k (f_k)_1 > \lambda^{1/p}$ or $\sup_k (f_k)_2 > \lambda^{1/p}$ holds. But because of the way we defined $(f_k)_1$, $\sup_k (f_1)_k > \lambda^{1/p}$ is not possible, hence

$$\{\sup_k f_k > 2\lambda^{1/p}\} \subset \{\sup_k (f_k)_2 > \lambda^{1/p}\}.$$

Thus, using the monotonicity of μ we obtain

$$2^p \int_0^\infty \mu(\sup_k f_k > 2\lambda^{1/p}) d\lambda \leq 2^p \int_0^\infty \mu(\sup_k (f_k)_2 > \lambda^{1/p}) d\lambda.$$

Let $\mathcal{T}'$ be the family of maximal dyadic cubes with the property

$$\frac{\int f_2 1_Q d\mu}{\mu(Q)} > \lambda^{1/p}.$$

Observe that $\{\sup_k (f_k)_2 > \lambda^{1/p}\} \subset \bigcup \mathcal{T}'$ and moreover, because the maximal cubes we have chosen are pairwise disjoint

$$\sum_{Q\in\mathcal{T}'} \mu(Q) \leq \lambda^{-1/p} \sum_{Q\in\mathcal{T}'} \int f_2 1_Q d\mu \leq \lambda^{-1/p} \int f_2 d\mu.$$

Plugging this into our computation we get

$$
\begin{aligned}
2^p \int_0^\infty \mu(\sup_k (f_k)_2 > \lambda^{1/p}) d\lambda &\le 2^p \int_0^\infty \lambda^{-1/p} \overbrace{\int_0^\infty \mu(f_2 > t) dt}^{=\int f_2 d\mu} d\lambda \\
&\overset{\text{def. of } f_2}{\le} 2^p \int_0^\infty \lambda^{-1/p} \int_{\lambda^{1/p}}^\infty \mu(f > t) dt d\lambda \\
\overset{t=\lambda^{1/p}s}{=} 2^p \int_0^\infty \int_1^\infty \mu(f > \lambda^{1/p} s) ds d\lambda &\overset{\text{Fubini}}{=} 2^p \int_1^\infty \int_0^\infty \mu(f^p > \lambda s^p) d\lambda ds \\
\overset{\lambda=s^{-p}u}{=} 2^p \int_1^\infty s^{-p} \int_0^\infty \mu(f^p > u) du ds &= \frac{2^p}{p-1} \int_0^\infty \mu(f^p > u) du \\
&= \frac{2^p}{p-1} \|f\|_p^p.
\end{aligned}
$$

Since $f \in L_+^p$ we obtain that $(\sup_k f_k) \in L_+^p$ what lets us to pass to the limit as $k \to \infty$

$$\| |f - f_k| \|_p^p \to 0.$$

$\square$

Remark.

$$Mf := \sup_k f_k,$$

which appears in the previous proof, is called the maximal function of f. We have shown that the following estimate holds

$$\|Mf\|_p \le \frac{2}{(p-1)^{1/p}} \|f\|_p.$$

Proposition 2.36. *Let $1 < p < \infty$, $f \in L_+^p(\mathbb{R}^d)$ and q be such that*

$$1/p + 1/q = 1.$$

Then there exists $h \in L_+^q(\mathbb{R}^d)$ with $\|h\|_q = 1$ and $\|f\|_p = \int fh d\mu$.

Remark. The statement can also be read as follows: for every $f \in L_+^p$ there exists an $h \in L_+^q$ with

$$\int fh d\mu = \|f\|_p \|h\|_q.$$

Proof. Let us put

$$h(x) = \begin{cases} \frac{f(x)^{p-1}}{\|f\|_p^{p-1}}, & \text{if } f(x) \neq 0 \\ 0, & \text{otherwise} \end{cases}$$

and note that

$$\int fhd\mu = \int \frac{ff^{p-1}}{\|f\|_p^{p-1}}d\mu = \|f\|_p.$$

$\square$

The next theorem, called the Riesz representation is a very important fact telling us, that bounded functionals on L_+^p with $1 < p < \infty$ are given by integrating against a function from L_+^q for q satisfying, $1/p + 1/q = 1$.

Theorem 2.37 (Riesz representation theorem)**.** *Let $1 < p < \infty$ and*

$$\Lambda \colon L_+^p(\mathbb{R}^d) \to [0, \infty)$$

be a functional satisfying the additivity property

$$\Lambda(f+g) = \Lambda(f) + \Lambda(g).$$

Then there exists $h \in L_+^q(\mathbb{R}^d)$, where $1/q + 1/p = 1$, such that for any $f \in L_+^p(\mathbb{R}^d)$

$$\Lambda(f) = \int fhd\mu.$$

Proof. The first step of the proof is the following lemma, known as the uniform boundedness principle or the Banach-Steinhaus theorem.

Lemma 2.38. *Let Λ be as in the statement of the theorem. There exists a constant $C < \infty$ such that for all $f \in L_+^p(\mathbb{R}^d)$*

$$\Lambda(f) \leq C\|f\|_p.$$

Proof. Assume the opposite, i.e. that there exists a sequence of functions $\{f_n\}$ with

$$\Lambda(f_n) \geq 2^{2n+1}\|f_n\|_p.$$

Without loss of generality, by linearity of the functional we may assume that

$$2^{-n-1} \leq \|f_n\|_p \leq 2^{-n} \text{ and then } \Lambda(f_n) > 2^n.$$

Define

$$g_N = \sum_{n=1}^{N} f_n.$$

The series $\|g_N\|_p$ is uniformly in N and absolutely bounded by 1, because

$$\|g_N\|_p \leq \sum_{n=1}^{N} \|f_n\|_p \leq 1.$$

Hence the monotone sequence $g_N \to g$ pointwise and in L^p_+ as $N \to \infty$ and $\|g\|_p \le 1$, by the Lebesgue monotone convergence applied for the sequence $\{g_N\}$. Since Λ maps L^p_+ to $[0, \infty)$ we get

$$\Lambda(g) < \infty.$$

On the other hand, for any $N \in \mathbb{N}$

$$\Lambda(g) \ge \Lambda(g - g_N) + \Lambda(g_N) \ge \Lambda(g_N) \ge \Lambda(f_N) \ge 2^N.$$

This is a contradiction. □

The second step of the proof is to show that the function $\tau(Q) = \Lambda(1_Q)$ generates a Radon measure $\nu \in L^q_+$.

By additivity of Λ we obtain the martingale condition for τ. We are thus left with validating the regularity condition. It follows from the following inequality

$$\sup_k \sum_{Q\colon Ord(Q)=k} \left(\frac{\nu(Q)}{\mu(Q)}\right)^q \mu(Q) < \infty,$$

which we should check anyway, since we want $\nu \in L^q_+$. Without loss of generality we may deal here only with the case when $k = 0$. Hence, we shall show that

$$\left(\sum_{Q\colon Ord(Q)=0} \left(\frac{\nu(Q)}{\mu(Q)}\right)^q \mu(Q)\right)^{1/q} \overset{\mu(Q)=1}{=} \left(\sum_{Q\colon Ord(Q)=0} (\nu(Q))^q\right)^{1/q} < \infty.$$

By the last proposition there exists a sequence $\{a_Q\}$ with $\sum_{Ord(Q)=0} a_Q^p = 1$ and

$$\left(\sum_{Q\colon Ord(Q)=0} (\nu(Q))^q\right)^{1/q} = \sum_{Q\colon Ord(Q)=0} a_Q \nu(Q).$$

Arguing similarly as in the proof of the uniform boundedness principle one can show that the series

$$\sum_{Q\colon Ord(Q)=0} a_Q 1_Q$$

converges to a function in L^p_+. Using this and assuming momentarily that the functional Λ is continuous(we will show in a second) we obtain

$$\sum_{Q\colon Ord(Q)=0} a_Q \nu(Q) = \sum_{Q\colon Ord(Q)=0} a_Q \Lambda(1_Q) \overset{\text{continuity}}{=} \Lambda\Big(\sum_{Q\colon Ord(Q)=0} a_Q 1_Q\Big)$$

$$\overset{\text{uni. bdd. principle}}{\le} C\|\sum_{Q\colon Ord(Q)=0} a_Q 1_Q\|_p = C\left(\sum_{Q\colon Ord(Q)=0} a_Q^p\right)^{1/p} = C.$$

Λ is indeed continuous, because

$$\Lambda(f) - \Lambda(g) \leq \Lambda(g) + \Lambda(|f-g|) - \Lambda(g) = \Lambda(|f-g|) \overset{\text{u.b.prin.}}{\leq} C\||f-g|\|_p.$$

Similarly we see that $\Lambda(g) - \Lambda(f) \leq C\||f-g|\|_p$, what gives

$$|\Lambda(g) - \Lambda(f)| \leq C\||f-g|\|_p.$$

This means that Λ is Lipschitz and therefore continuous. We have already seen that

$$f_k = \frac{\int f 1_Q d\mu}{\mu(Q)} 1_Q,$$

converges to f in L^p_+ as $k \to \infty$. Thus, by this fact and continuity of Λ we obtain

$$\begin{aligned}
\Lambda(f) = \lim_{k\to\infty} \Lambda(f_k) &= \lim_{k\to\infty} \sum_{Q\colon Ord(Q)=k} \Lambda\left(\frac{\int f 1_Q d\mu}{\mu(Q)} 1_Q\right) \\
&= \lim_{k\to\infty} \sum_{Q\colon Ord(Q)=k} \frac{\int f 1_Q d\mu}{\mu(Q)} \Lambda(1_Q) = \lim_{k\to\infty} \sum_{Q\colon Ord(Q)=k} \frac{\int f 1_Q d\mu}{\mu(Q)} \nu(Q) \\
&= \lim_{k\to\infty} \sum_{Q\colon Ord(Q)=k} \frac{\int f 1_Q d\mu}{\mu(Q)} \int h 1_Q d\mu = \lim_{k\to\infty} \int h f_k d\mu = \int h f d\mu,
\end{aligned}$$

where the last inequality follows from the Lebesgue dominated convergence, because h is in L^q_+, $f_k h \leq (\sup_k f_k) h$ and by Hölder's inequality

$$\int (\sup_k f_k) h d\mu \leq \|\sup_k f_k\|_p \|h\|_q < \infty.$$

□

End of lecture 15. December 10, 2015

3 An excursion to Probability theory

Definition 3.1. A (nonnegative) Radon measure on $\mathbb{R}$ with $\nu(\mathbb{R}) = 1$ is called a random variable.

The actual word random variable refers to real variable implicit in the above, it is the variable parameterizing $\mathbb{R}$.
A different but closely related definition of random variable in the literature makes the random variable a function (say x) on a measure space Ω with total mass 1. The space Ω is thought of as the space of all possible events of a random experiment, and the function x is a specific observable of that experiment, which can take real values. One can think of "$x(\omega)$" for an $\omega \in \Omega$ as the observation resulting from of an experiment. The aim of probability theory is to answer the question 'what is the probability?' of an event A or equivalently 'what is the measure?' of the set $\{x \in A\}$, that is why we defined a random variable as a Radon measure. Hence, we are interested in the values of "x" rather then the arguments and its domain. Our definition of random variable does not rely on a space Ω, relative to the just mentioned definition it is the "push forward" of the measure Ω to the real line $\mathbb{R}$ under the map x. (A formal definition of "push forward" is postponed, or can be regarded as exercise.)
For a Borel set $B \subset \mathbb{R}$, $\nu(B)$ is called the probability that the random variable takes a value in B, also written $P(x \in B) := \nu(B)$ where x makes an explicit appearance though it is a dummy variable as much as an integration variable is a dummy variable. Note x does not appear in the expression $\nu(B)$.

Example. Consider rolling a 6-sided symmetric dice. The Radon measure/random variable associated with "x" in this case is simply

$$\nu = \sum_{i=1}^{6} \frac{1}{6}\delta_0(x-i).$$

Definition 3.2. If $\int |x| d\nu(x) < \infty$, then we define *the expected value*

$$E(x) := \int x \; d\nu(x).$$

Write abs for the absolute value function and id for the identity function, then we can write without dummy integration variable the condition on ν as $\int \mathrm{abs} d\nu$ and the expectation value as $\int \mathrm{id} d\nu$.
Analogously to a random variable, we define a *random vector.*

Definition 3.3. A random d-vector is a Radon measure ν on $\mathbb{R}^d$ with $\nu(\mathbb{R}^d) = 1$.

Just as it is customary to give a random variable a dummy variable, a random vector is customarily given a tuple of dummy variables "$x = (x_i)_{i=1}^d$".

Given a random vector ν on $\mathbb{R}^d$, we define the component ν_i as the Radon measure generated by

$$\tau_i(Q) = \nu(\mathbb{R} \times ... \times \mathbb{R} \times \underbrace{Q}_{i\text{-th component}} \times \mathbb{R} \times ... \times \mathbb{R}) = \nu(\{x \in \mathbb{R}^d, x_i \in Q\}).$$

We easily obtain that τ satisfies the martingale condition. Regularity follows from the fact that if $E_l \subset \mathbb{R}$ is a sequence with $\bigcap_{l=0}^{\infty} E_l = \emptyset$, then

$$\lim_{L\to\infty} \nu(\mathbb{R} \times ... \times \bigcap_{l=0}^{L} E_l \times ... \times \mathbb{R}) = 0.$$

(Note that all measures involved are finite.) Thus, the generated ν_i's are indeed Radon measures with (exercise:)

$$\nu_i(\mathbb{R}) = \nu(\mathbb{R}^d) = 1 \text{ and } \nu_i(Q) = \tau_i(Q).$$

Note that for a ν_i measurable function f we have

$$\int f d\nu_i = \int f \circ \Pi_i d\nu. \tag{9}$$

where Π_i is the projection $x \to x_i$, also written as $x \to \langle x, e_i \rangle$ with the $i-th$ unit vector e_i. This is clear from the definitions if f is the characteristic function of a dyadic cube. Then one proves the general case by first considering linear combinations of characteristic functions of dyadic cubes and then limits thereof. We skip the details here.
Above we dealt with the i-th component of a random variable ν satisfying

$$\nu_i(Q) = \nu(\mathbb{R} \times ... \times Q \times \mathbb{R} \times ... \times R) = \nu(\{x \in \mathbb{R}^d \colon x_i \in Q\}).$$

One can also consider components of a rotation of a random vector. These leads in general to a component relative to a unit vector. For a unit vector v in $\mathbb{R}^d$, i.e. a vector having the property that

$$\sum_{i=1}^{d} v_i^2 = 1,$$

we can obtain as above the v component of a random variable ν satisfying

$$\nu_v(Q) := \nu(\{x \in \mathbb{R}^d \colon v \cdot x \in Q\}),$$

where "$\cdot$" denotes the scalar product.

Next, let $\nu_1, ..., \nu_d$ be random variables. We define *the independent product* of those random variables as the product measure generated by

$$\tau(Q) = \prod_{i=1}^{d} \nu_i(Q), \quad Q = Q_1 \times ... \times Q_d \subset \mathbb{R}^d.$$

We shall not elaborate the details of this construction, as we have discussed product measures before.
It is importnat to note that not every random vector is the independent product of its components, as elaborated in the example below. This is much like the fact that not every function in two variables can be written as the product of two functions in one variable, $f(x, y) = f_1(x)f_2(y)$. Hence the word "independent", which denotes an additional property.

Example. Let $Q_1 = [0, 1)^2$ and $Q_2 = [1, 2)^2$ and let μ denote Lebesgue measure on $\mathbb{R}^2$. Define a random variable ν with: $\nu(Q) = \frac{1}{2}\mu(Q \cap Q_1) + \frac{1}{2}\mu(Q \cap Q_2)$. Notice that ν it is not the product measure of any ν_1, ν_2. Indeed, suppose it was the product measure Then, the following three pairs of equalities would hold

$$\begin{aligned} 0 &= \nu(Q_1 \times Q_2) = \nu_1(Q_1)\nu_2(Q_2), \\ \frac{1}{2} &= \nu(Q_1 \times Q_1) = \nu_1(Q_1)\nu_2(Q_1), \\ \frac{1}{2} &= \nu(Q_2 \times Q_2) = \nu_1(Q_2)\nu_2(Q_2). \end{aligned}$$

$\nu(Q_1 \times Q_1)$, $\nu(Q_2 \times Q_2)$ are nonzero, what gives that $\nu_1(Q_1)$ and $\nu_2(Q_2)$ are nonzero. This contradicts the first equality.

We proceed with proving a "baby version" of the central limit theorem.
Let ν be a random variable and let ν^d denote the d-dimensional product measure. In probability this setup is called a d-tuple of $i.i.d$ random variable (this means "independent identically distributed"). Consider the unit vector

$$v = \frac{1}{d^{1/2}}(1, ..., 1)$$

and the random variable "$\frac{x_1+x_2+...+x_d}{d^{1/2}}$", given by the v component of ν^d as defined above

$$\mu_d = (\nu^d)_v.$$

Recall the Fourier transform of μ_d at a point ξ is equal to

$$h_d(\xi) = \int e^{2\pi i y\xi} d\mu_d(y).$$

Note that since μ_d is a random variable, the function $e^{2\pi i y\xi}d\mu_d(y)$ is absolutely integrable and we have $|h_d(\xi)| \leq 1$, because

$$\left|\int e^{2\pi i y\xi}d\mu_d(y)\right| \leq \int 1 d\mu_d = 1.$$

Using (9) for the unit vector v we obtain with Fubini

$$h_d(\xi) = \int e^{2\pi i(x_1+...+x_d)\xi}d\nu(x_1)...d\nu(x_d) = \prod_{j=1}^{d}\left(\int e^{2\pi i\frac{1}{d^{1/2}}x_j\xi}d\nu(x_j)\right).$$

Since the random variables are i.i.d, we have in fact shown that

$$h_d(\xi) = \left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d,$$

where

$$f(\xi) := \int e^{2\pi i x\xi}d\nu(x).$$

Let us assume that f is two times continuously differentiable at $\xi = 0$ and $f''(0) \neq 0$. Moreover, we have

$$f(0) = 1, \text{ because } f(0) = \int 1 d\nu = 1,$$

$$f'(0) = 0, \text{ because } |f(\xi)| \leq \int 1 d\nu = 1 = f(0), \text{ so } f \text{ is maximal at } 0.$$

$$f''(0) < 0, \text{ because at a local maximum } \leq 0 \text{ and } f''(0) \neq 0.$$

Let us denote $-\rho := f''(0) < 0$ and consider the Gaussian

$$g(\xi) = e^{-\rho\frac{\xi^2}{2}}.$$

Observe the properties

$$g(0) = 1$$
$$g'(0) = 0,$$
$$g''(0) = -\rho.$$

Fix $\varepsilon > 0$ small. Notice that for ξ small enough we have by considering Taylor's approximation

$$e^{-(\rho+\varepsilon)\frac{\xi^2}{2}} =\leq \left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d \leq= e^{-(\rho-\varepsilon)\frac{\xi^2}{2}}.$$

Now fix ξ and let d be large enough. Then $\xi/d^{1/2}$ is small enough so the above applies and we obtain.

$$e^{-(\rho+\varepsilon)\frac{\xi^2}{2}} = (e^{-(\rho+\varepsilon)\frac{\xi^2}{2d}})^d \leq \left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d \leq \left(e^{-(\rho-\varepsilon)\frac{\xi^2}{2d}}\right)^d = e^{-(\rho-\varepsilon)\frac{\xi^2}{2}}.$$

Here we have used in the first and last identity an invariance property of the Gaussian function under scaling. Letting d tend to infinity we obtain

$$e^{-(\rho+\varepsilon)\frac{\xi^2}{2}} \leq \liminf_{d\to\infty}\left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d \leq \limsup_{d\to\infty}\left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d \leq e^{-(\rho-\varepsilon)\frac{\xi^2}{2}}.$$

Finally, letting $\epsilon \to 0$ we get

$$e^{-\rho\frac{\xi^2}{2}} = \lim_{d\to\infty}\left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d.$$

Written more explicitly:

$$e^{-\rho\frac{\xi^2}{2}} = \lim_{d\to\infty}\left(f\left(\frac{\xi}{d^{1/2}}\right)\right)^d = \lim_{d\to\infty}\int e^{2\pi i y\xi}d\mu_d(y).$$

That is what we have called the "baby version" of the central limit theorem: the Fourier transform of μ_d converges pointwise to a Gaussian with the appropriate scaling factor ρ as determined by ν. The statement of a more elaborate version of the theorem is the following.

Theorem 3.4 (Central limit theorem)**.** *Let ν be a random variable with*

$$\sigma := \int x^2 d\nu(x) < \infty.$$

Let $\mu_d = (\nu^d)_v$ as before. Then the following convergence in measure holds

- *For all $Q \in \mathcal{T}$: $\lim_{d\to\infty}\mu_d(Q) = \frac{1}{\sqrt{2\pi\sigma}}\int e^{-\frac{x^2}{2\sigma}}1_Q(x)dx$, and*
- $\lim_{d\to\infty}\mu_d(\mathbb{R}^d) = \frac{1}{\sqrt{2\pi\sigma}}\int e^{-\frac{x^2}{2\sigma}}dx$.

Proof. Left as an exercise. The condition of finiteness of σ is used to obtain enough regularity of f as in the above calculations. Convergence is then proved by approximating the characteristic functions from above and below by piecewise linear functions, and use the de la Vallee Poussin kernel to write these piecewise linear functions as superposition of functions $e^{2\pi i x\xi}$ to reduce to the above baby version. The proof of teh second property is a mere observation that both sides of the identity are 1.

$\square$

3.1 Infinite product measures

In the last subsection we saw a remarkable universal limit of averages of d tuples of i.i.d. random variables as $d \to \infty$. Considering large products of measures is of essence in probability theory. It is then only natural to investigate infinite product spaces.
Let ν_i be a sequence of random variables for $i \in \mathbb{N}$. Define the set of infinite sequences taking values in $\mathbb{R}$

$$X = \mathbb{R}^\infty = \{x = (x_i)_{i=1}^\infty \colon x_i \in \mathbb{R}\}$$

and the family $\mathcal{T}$ associated with X to be the set of products of finitely many dyadic intervals and infinitely many copies of $\mathbb{R}$

$$\begin{aligned}\mathcal{T} &= \{Q \times \mathbb{R}^{\infty - d} \colon d \in \mathbb{N}, Q \text{ a dyadic cube in } \mathbb{R}^d\}\\ &= \{(x_i)_{i=1}^\infty \colon (x_i)_{i=1}^d \in Q \text{ for some } d, Q\}.\end{aligned}$$

Moreover, we define the generating function τ as following

$$\tau(Q \times \mathbb{R}^{\infty-d}) = \prod_{i=1}^d \nu_i(Q_i) \cdot \prod_{i=d+1}^\infty \underbrace{\nu_i(\mathbb{R})}_{=1} = \prod_{i=1}^d \nu_i(Q_i).$$

Let ν be the outer measure generated by τ. The following is the infinite dimensional version of an important observation that we made in $\mathbb{R}^n$. It is a baby version of a family of theorems in the literature on infinite product spaces.

Theorem 3.5. *For all $Q \in \mathcal{T}$, $\tau(Q) = \nu(Q)$ holds.*

Proof. We shall show that for any $Q \in \mathcal{T}$ and a sequence $Q^l \in \mathcal{T}$ with $Q \subset \bigcup_l Q^l$ we have

$$\tau(Q) \leq \sum_l \tau(Q^l).$$

Let $\varepsilon > 0$ be small and let $\tilde{Q}_i$ be a sequence of compact intervals that satisfies $\tilde{Q}_i \subset Q_i$ and

$$\begin{aligned}\nu_i(\tilde{Q}_i) &\geq e^{-\varepsilon 2^{-i}} \nu_i(Q_i), \text{ for } i \leq d,\\ \nu_i(\tilde{Q}_i) &\geq e^{-\varepsilon 2^{-i}} \nu_i(\mathbb{R}^d), \text{ for } i > d.\end{aligned}$$

The idea is that we approximate the measure of Q_i (in case $i > d$ we approximate the measure of $\mathbb{R}$) from below by a compact set contained in $\tilde{Q}_i$. Similarly we approximate Q_i^l from above by open intervals. Let $\tilde{Q}_i^l$ be an open interval with $\tilde{Q}_i^l \supset Q_i^l$ and

$$\nu_i(\tilde{Q}_i^l) \leq e^{\varepsilon(2^{-i} + 2^{-l})} \nu(Q_i^l), \text{ for } i \leq d^l,$$

$$\tilde{Q}_i^l = \mathbb{R}, \text{ for } i > d^l.$$

By Cantor's diagonal and a compactness argument (we skip the details) one shows that there exists an $L < \infty$ such that

$$\tilde{Q} \subset \bigcup_{l=1}^{L} \tilde{Q}^l.$$

Then, taking $(\max_l d^l) < \infty$ we can simply restrict ourselves to $\mathbb{R}^{(\max_l d^l)}$ for which the statement is true by the theory of finite product measures, i.e.

$$\tau(\tilde{Q}) \leq \sum_{l=1}^{L} \tau(\tilde{Q}^l).$$

Letting $\varepsilon \to 0$ we finish the proof. □

End of lecture 16. December 15, 2015

We elaborate on some notions introduced in the previous lectures. The product measure of ν_1 on $\mathbb{R}^{d_1}$ and ν_2 on $\mathbb{R}^{d_2}$ is the measure ν on $\mathbb{R}^{d_1+d_2}$ given by

$$\nu(Q) = \nu_1(Q_1 \times \cdots \times Q_{d_1})\nu_2(Q_{d_1+1} \times \cdots \times Q_{d_1+d_2}).$$

Not every measure on $\mathbb{R}^d$ is such a product measure, just as not every function is of the form

$$f(x_1, \ldots, x_d) = f_1(x_1, \ldots, x_{d_1}) f_2(x_{d_1+1}, \ldots, x_{d_1+d_2}).$$

This construction of a product measure corresponds to the partition $\{1, \ldots, d\} = \{1, \ldots, d_1\} \cup \{d_1+1, \ldots, d_2\}$ where $d = d_1+d_2$. Analogously we define product measures with respect to any partition

$$\{1, \ldots, d\} = \bigcup_{\ell=1}^{L} N_l$$

where the union is disjoint.

Let ν be a random vector on $\mathbb{R}^d$ and n an injection from $\{1, \ldots, d'\}$ to $\{1, \ldots, d\}$. We define the projected random vector ν' by setting

$$\begin{aligned}\nu'(Q_1 \times \cdots \times Q_{d'}) &= \nu(\mathbb{R} \times \cdots \times Q_{n(1)} \times \cdots \times \mathbb{R} \times \cdots \times Q_{n(d)}) \\ &= \nu(\{x : x_i \in Q_{n(i)}, 1 \leq i \leq d'\}).\end{aligned}$$

A random vector is called an independent product with respect to a partition $\{1, \ldots, d\} = \bigcup_{\ell=1}^{L} N_l$ if it is the product of the corresponding projections.

Last time we also considered infinite product measures, see Section 3.1. For infinite product measures we analogously define projections of random vectors onto subsets of $\mathbb{N}$ and independent products with respect to partitions of $\mathbb{N}$.

A normalized Gaussian random variable is defined by

$$\nu(Q) = \frac{1}{\sqrt{2\pi}} \int_Q e^{-\frac{x^2}{2}} dx.$$

The function $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the Radon-Nikodym derivative of ν with respect to the Lebesgue measure. The factor $\frac{1}{\sqrt{2\pi}}$ is chosen such that $\nu(\mathbb{R}) = 1$.
Since

$$\prod_{i=1}^{d} \Big(\frac{1}{\sqrt{2\pi}} \int_{Q_i} e^{-\frac{x_i^2}{2}} \Big) = \frac{1}{\sqrt{2\pi}^d} \int_Q e^{-\frac{\sum_{i=1}^d x_i^2}{d}} d\mu = \frac{1}{\sqrt{2\pi}^d} \int_Q e^{-\frac{|x|^2}{2}} d\mu,$$

the independent product of d Gaussian random variables is given by

$$\nu(Q) = \frac{1}{\sqrt{2\pi}^d} \int_Q e^{-\frac{|x|^2}{2}} d\mu.$$

Theorem 3.6. *The independent product of d Gaussian random variables is a rotation invariant Radon measure on $\mathbb{R}^d$.*

Proof. Let $A : \mathbb{R}^d \to \mathbb{R}^d$ be linear and $|Ax| = |x|$ for all $x \in \mathbb{R}^d$. Let $E \subset \mathbb{R}^d$ be Borel. Then

$$\begin{aligned} \nu(AE) = \frac{1}{\sqrt{2\pi}^d} \int_{AE} e^{-\frac{|x|^2}{2}} d\mu &\overset{(1)}{=} \frac{1}{\sqrt{2\pi}^d} \int_E e^{-\frac{|Ax|^2}{2}} d\mu \\ &\overset{(2)}{=} \frac{1}{\sqrt{2\pi}^d} \int_E e^{-\frac{|x|^2}{2}} d\mu = \nu(E). \end{aligned}$$

For the (1) we used rotation invariance of μ and that if $x = Ax' \in AE$, then $x' \in E$. For (2) we used $|Ax| = |x|$. $\square$

For an infinite product of i.i.d. normalized Gaussian random variables we would like to discuss rotation invariance as well. To define rotations we need a Hilbert space structure. An example of an infinite dimensional Hilbert space is $\ell^2(\mathbb{N})$, the space of all square-summable sequences. The infinite product of normalized Gaussians lives on the space $\mathbb{R}^{\mathbb{N}}$ of all sequences of real numbers, such sequences are not necessarily in $\ell^2(\mathbb{N})$. Moreover, for any the space of bounded sequence (i.e., $\ell^\infty(\mathbb{N})$) the following holds.

Lemma 3.7. *The set $\ell^\infty(\mathbb{N})$ has ν-outer measure zero in $\mathbb{R}^{\mathbb{N}}$, where ν is the infinite product of normalized Gaussian random variables.*

Since $\ell^2(\mathbb{N}) \subset \ell^\infty(\mathbb{N})$ then also $\ell^2(\mathbb{N})$ has measure zero. This indicates that developing a suitable Hilbert space theory will not be possible without additional assumptions on the sequences.

Proof. Recall that $x \in \ell^\infty(\mathbb{N})$ is a Banach space with the norm $\|x\|_\infty = \sup_i |x_i|$. We need to show that for every $k > 0$

$$\nu(\underbrace{\{x \in \mathbb{R}^{\mathbb{N}} : \|x\|_\infty \leq 2^k\}}_{\bigcap_i \{x \in \mathbb{R}^{\mathbb{N}} : |x_i| \leq 2^k\}}) = 0.$$

It suffices to show that for every $\varepsilon > 0$ there exists $d \in \mathbb{N}$ such that

$$\nu(\{x \in \mathbb{R}^{\mathbb{N}} : \sup_{i=1,\dots,d} |x_i| \leq 2^k\}) < \varepsilon.$$

Thus we would like to compute

$$\nu\Big(\bigcap_{i=1}^{d} \{x \in \mathbb{R}^{\mathbb{N}} : |x_i| \leq 2^k\}\Big).$$

This is a d-fold product of Gaussian random variables

$$\prod_{i=1}^{d} \Big(\frac{1}{\sqrt{2\pi}} \int_{-2^k}^{2^k} e^{-\frac{x^2}{2}} \, dx\Big) = \prod_{i=1}^{d} \Big(1 - \underbrace{\frac{2}{\sqrt{2\pi}} \int_{2^k}^{\infty} e^{-\frac{x^2}{2}} \, dx}_{=:\delta_k > 0}\Big) = (1 - \delta_k)^d.$$

If d is large enough, the right hand-side is less than ε. $\square$

If we consider the set of sequences with square root logarithmic growth, the result changes.

Theorem 3.8.

$$\nu(\{x \in \mathbb{R}^{\mathbb{N}} : \exists k \forall i : |x_i| \leq 2^k (\ln(e + i))^{1/2}\}) = 1$$

Thus almost all sequences satisfy a square root logarithmic growth estimate.

Proof. We show that the complement has measure zero, i.e.

$$\nu(\underbrace{\{x : \forall k \exists i : |x_i| > 2^k (\ln(e+i))^{1/2}\}}_{\bigcap_k \{x : \exists i : |x_i| > 2^k (\ln(e+i))^{1/2}\}}) = 0.$$

Thus it suffices to show that $\forall \varepsilon > 0 \, \exists k$ such that

$$\nu(\underbrace{\{x : \forall k \exists i : |x_i| > 2^k (\ln(e+i))^{1/2}\}}_{\bigcup_i \{x_i : |x_i| > 2^k (\ln(e+i))^{1/2}\}}) < \varepsilon.$$

So it suffices to show that

$$\sum_i \nu(\{x_i : |x_i| > 2^k (\ln(e+i))^{1/2}\}) < \varepsilon.$$

We have

$$\begin{aligned}
\nu(\{x_i : |x_i| > 2^k (\ln(e+i))^{1/2}\}) &= \frac{2}{\sqrt{2\pi}} \int_{2^k (\ln(e+i))^{1/2}}^{\infty} e^{-\frac{x^2}{2}} \, dx \\
&\leq \frac{2}{\sqrt{2\pi}} \int_{2^k (\ln(e+i))^{1/2}}^{\infty} e^{-\frac{x 2^k (\ln(e+i))^{1/2}}{2}} \, dx \\
&\overset{(1)}{=} \frac{2}{\sqrt{2\pi}} \frac{2}{2^k (\ln(e+i))^{1/2}} e^{-\frac{2^{2k} \ln(e+i)}{2}} \\
&\overset{(2)}{\leq} \frac{4}{\sqrt{2\pi}} \frac{2}{2^k} \Big(\frac{1}{e+i} \Big)^{2^{2k-1}}.
\end{aligned}$$

In (1) we computed the integral, in (2) we used $\ln(e+i) > 1$. If k is large enough, then $(1/e+i)^{2^{2k-1}}$ is summable in i and the term in (2) is less than ε. $\square$

The following are two examples of measurable functions in $\mathbb{R}^{\mathbb{N}}$ with ν as above.

1. Functions which depend only on d variables ($f(x) = f(x')$ if $x_i = x'_i$ for $i = 1, \ldots, d$) and are measurable as functions on $\mathbb{R}^d$.

2. Linear functionals of the form

$$\sum_{i=1}^{\infty} v_i x_i =: f(x) \tag{10}$$

with

$$\sum_{i=1}^{\infty} |v_i| (\ln(e+i))^{1/2} < \infty.$$

The last condition means that a certain weighted ℓ^1-norm of the coefficients is finite. Functionals of the form (10) are defined almost everywhere. Indeed, if x is such that $\sup_i |x_i| \leq 2^k (\ln(e+i))^{1/2}$, then $\sum_{i=1}^{\infty} |v_i||x_i| < \infty$. By the previous theorem almost all x satisfy such a condition. Functionals of the form (10) are measurable since they are pointwise a.e. limits of measurable functions. To see that set

$$f_d := \sum_{i=1}^{d} v_i x_i$$

which is measurable by 1. Since

$$\lim_{d\to\infty} f_d(x) = f(x) \quad \text{a.e.},$$

f is measurable.

Let us call vectors $v := (v_i)$ as in 2. admissible. Note that such admissible vectors are automatically in $\ell^2(\mathbb{N})$.

Theorem 3.9. *If v is admissible, then $v \cdot x$ is defined almost everywhere in $\mathbb{R}^{\mathbb{N}}$ and the random variable ν_v defined by*

$$\nu_v(Q) = \nu(\{x : v \cdot x \in Q\})$$

is normalized Gaussian distributed.

Proof. Consider a unit vector in $\mathbb{R}^d$

$$v^{(d)} = \frac{(v_1, \dots, v_d)}{|(v_1, \dots, v_d)|}$$

where we assume $|(v_1, \dots, v_d)| \neq 0$. By the rotation invariance result in $\mathbb{R}^d$,

$$f^{(d)}(x) = \sum_{i=1}^{d} v_i^{(d)} x_i$$

is normalized Gaussian distributed.
We need to show that for every $\varepsilon > 0$

$$\left| \nu(\{x : v \cdot x \in Q\}) - \frac{1}{\sqrt{2\pi}} \int_Q e^{-\frac{x^2}{2}} dx \right| < \varepsilon.$$

Choose k large enough such that

$$\nu(\underbrace{\{x : \forall i : |x_i| < 2^k(\ln(e+i))^{1/2}\}}_{=:E}) > 1 - \frac{1}{3}\varepsilon.$$

It then suffices to show that

$$\Big|\nu(\{x \in E : v \cdot x \in Q\}) - \frac{1}{\sqrt{2\pi}} \int_Q e^{-\frac{x^2}{2}} dx\Big| < \frac{2}{3}\varepsilon.$$

Note that $v \cdot x$ is well defined for $x \in E$. By the dominated convergence theorem $v^{(d)} \cdot x \to v \cdot x$ a.e. (exercise). So it suffices to show that

$$\Big|\nu(\{x \in E : v^{(d)} \cdot x \in Q\}) - \frac{1}{\sqrt{2\pi}} \int_Q e^{-\frac{x^2}{2}} dx\Big| < \frac{1}{3}\varepsilon.$$

Removing E again, it suffices to show

$$\Big|\nu(\{x : v^{(d)} \cdot x \in Q\}) - \frac{1}{\sqrt{2\pi}} \int_Q e^{-\frac{x^2}{2}} dx\Big| = 0.$$

Observe that this holds by the rotation invariance result in $\mathbb{R}^d$. □

End of lecture 17. December 17, 2015

3.2 Brownian motion

Let us define

$$\ell^\infty(\mathbb{N}, \log^{-1/2}) = \Big\{x \in \mathbb{R}^{\mathbb{N}} : \sup_{i \in \mathbb{N}} \frac{|x_i|}{\log(e+i)^{1/2}} < \infty\Big\}.$$

Last time we showed that $\nu(\ell^\infty(\mathbb{N}, \log^{-1/2})) = 1$. In other words, almost all sequences are contained in $\ell^\infty(\mathbb{N}, \log^{-1/2})$. Thus, if we take $v \in \ell^1(\mathbb{N}, \log^{1/2})$, where

$$\ell^1(\mathbb{N}, \log^{1/2}) = \Big\{x \in \mathbb{R}^{\mathbb{N}} : \sum_{i \in \mathbb{N}} |x_i| \log(e+i)^{1/2} < \infty\Big\},$$

then the sum $x \cdot v = \sum_{i \in \mathbb{N}} x_i v_i$ converges absolutely for almost all $x \in \mathbb{R}^{\mathbb{N}}$ (exactly for all $x \in \ell^\infty(\mathbb{N}, \log^{-1/2})$). In particular, this holds for v with $\|v\|_2^2 = \sum_{i \in \mathbb{N}} |v_i|^2 = 1$. Observe that if $v^{(1)}, \ldots, v^{(n)} \in \ell^1(\mathbb{N}, \log^{1/2})$ and $\|v^{(\ell)}\|_2 = 1$, $\langle v^{(\ell)}, v^{\ell'} \rangle = 0$ für $\ell \neq \ell'$, then the $\nu_{v^{(\ell)}}$ are pairwise independent.

Roughly speaking, our goal is to replace the index set $\mathbb{N}$ by the dyadic numbers $\mathbb{Y}$. Let $\mathcal{I}$ be the set of dyadic intervals $I = \left(\frac{n}{2^k}, \frac{n+1}{2^k}\right], k, n \in \mathbb{N}$. This is a countable set so that it makes sense to consider an infinite product of Gaussian random variables on the space $\mathbb{R}^{\mathcal{I}} = \{(x_I)_{I\in\mathcal{I}} : x_I \in \mathbb{R}\}$. To a dyadic interval $I \in \mathcal{I}$, we associate the *Haar function*

$$h_I = \frac{1}{\sqrt{I}}\Big(\mathbf{1}_{I_\ell} - \mathbf{1}_{I_r}\Big),$$

where I_ℓ, I_r denote the left and right child of I, respectively.

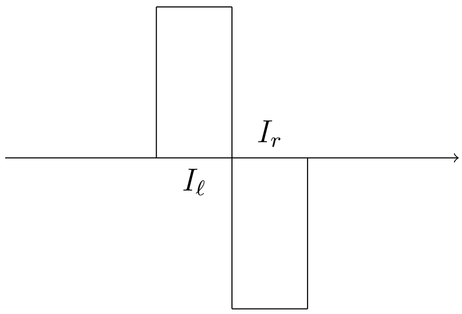


Figure 30: A Haar function h_I.

It is normalized so that $\|h_I\|_2 = 1$. Consider the space

$$\Big\{\sum_{I\in\mathcal{I}} x_I h_I \,:\, x \in \mathbb{R}^{\mathcal{I}}\Big\}.$$

This is the *Gaussian free field*. Naturally the definition needs to be taken with a grain of salt. It is not immediately clear if and in what sense and for which x the sum makes sense. But let us not worry about these issues here. Instead, we will interpret elements of the Gaussian free field as martingales. That is, given x, we would like to look at the averages

$$\int_I \sum_{J\in\mathcal{I}} x_J h_J.$$

Of course this integral doesn't make any sense. However, we can define

$$\rho(I) = \sum_{J\in\mathcal{I}} x_J \int_I h_J.$$

To stress the dependence on x we will also write $\rho_x(I)$. This sum converges absolutely for almost all $x \in \mathbb{R}^{\mathcal{I}}$. This is because for fixed $I \in \mathcal{I}$, the coefficients $\int_I h_J$ exhibit sufficient decay in J; we have

$$\Big|\int_I h_J\Big| = \begin{cases} 0, & \text{falls}\, I\cap J=\emptyset, \\ 0, & \text{falls}\, J\subseteq I, \\ \frac{|I|}{\sqrt{|J|}}, & \text{falls}\, I\subsetneq J. \end{cases}$$

For every k with $2^k > |I|$ there is exactly one $J\in\mathcal{I}$ with $I\subsetneq J$ and we have $|\int_I h_J| = \frac{|I|}{2^{k/2}}$. This exponential decay is certainly more than enough to compete with any sort of logarithmic growth that we need to allow for the admissible set of $(x_I)_{I\in\mathcal{I}}$ to have full measure.
ρ_x is a martingale. Indeed, we have

$$\rho(I) = \rho(I_\ell) + \rho(I_r).$$

Moreover, the following properties hold.

1. $\frac{\rho(I)}{\sqrt{|I|}}$ is normalized Gaussian distributed.
   This follows as in Theorem 3.9. It suffices to check

   $$\sum_{J\in\mathcal{I}}\Big(\frac{1}{\sqrt{|I|}}\int_I h_J\Big)^2 = \sum_{J\in\mathcal{I}}\Big|\Big\langle\frac{\mathbf{1}_I}{\sqrt{|I|}}, h_J\Big\rangle\Big|^2 = \Big\|\frac{\mathbf{1}_I}{\sqrt{|I|}}\Big\|_2^2 = 1.$$

   Here we used that the Haar functions form an orthonormal basis of L^2.

2. If $I, I'\in\mathcal{I}$ are disjoint and $|I| = |I'|$, then $\frac{\rho(I)}{\sqrt{|I|}}, \frac{\rho(I')}{\sqrt{|I'|}}$ are independently distributed.
   This is seen by changing coordinates to the basis

   $$\Big\{\frac{\mathbf{1}_J}{|J|^{1/2}} \,:\, J\in\mathcal{I},\, |J| = |I| = |I'|\Big\}\cup\Big\{h_J \,:\, |J| < |I| = |I'|\Big\}.$$

3. If $I_1,\dots I_n\in\mathcal{I}$ are pairwise disjoint, then $\frac{\rho(I_1)}{\sqrt{|I_1|}},\dots,\frac{\rho(I_n)}{\sqrt{|I_n|}}$ are independently distributed.

A natural question is whether, or rather for which x, ρ_x is a signed Radon measure? We need to check the condition

$$\sup_k \sum_{|I|=2^k, I\subset[0,1]} |\rho(I)| < \infty.$$

Claim.

$$\nu\Big(\Big\{x\in\mathbb{R}^{\mathcal{I}} \,:\, \exists C\,\forall k \,:\, \sum_{|I|=2^k, I\subset[0,1]}|\rho(I)|\le C\Big\}\Big) = 0.$$

Proof. It suffices to show

$$\nu\Big(\Big\{x \in \mathbb{R}^{\mathcal{I}} : \forall k \sum_{|I|=2^k, I \subset [0,1]} |\rho(I)| \leq C\Big\}\Big) = 0$$

for a fixed C. Further, it suffices to show that for all $\varepsilon > 0$ and k we have

$$\nu\Big(\Big\{x \in \mathbb{R}^{\mathcal{I}} : \sum_{|I|=2^k, I \subset [0,1]} |\rho(I)| \leq C\Big\}\Big) \leq \varepsilon.$$

Define

$$\mu = \frac{\frac{|\rho(I)|}{\sqrt{|I|}} - a}{b}.$$

for suitable a, b such that $\int x d\mu = 0$ and $\int x^2 d\mu = 1$. Then by the central limit theorem

$$\sum_{|I|=2^k, I \subset [0,1]} \frac{\frac{|\rho(I)|}{\sqrt{|I|}} - a}{b\sqrt{2^k}}$$

converges to a normalized Gaussian random variable. The rest is left as an exercise. □

That is, $\rho(\cdot, x)$ is almost never a Radon measure. In particular, the limit

$$\lim_{x \in I, |I| \to 0} \frac{\rho(I)}{|I|}.$$

almost never exists.

The *Brownian motion* is defined as a generalization of ρ as

$$w(t) = \sum_{J \in \mathcal{I}} x_J \int_0^t h_J.$$

If $(t_1, t_2]$ is a dyadic interval, then

$$w(t_2) - w(t_1) = \rho((t_1, t_2]).$$

The properties of ρ that we have seen carry over to w and also characterize the Brownian motion in some sense.

1. $\frac{w(t_2)-w(t_1)}{\sqrt{t_2-t_1}}$, $t_1 < t_2$, are normalized Gaussian distributed (*Gaussian increments*).

2. If $t_1 < t_2 < t_3$, then $\frac{w(t_3)-w(t_2)}{\sqrt{t_3-t_2}}, \frac{w(t_2)-w(t_1)}{\sqrt{t_2-t_1}}$ are independent random variables.

Theorem 3.10. *Let $\alpha < \frac{1}{2}$. Then w is almost surely (up to sets of ν-measure 0) a α-*Hölder continuous *function on $[0,1]$. That is, there is C such that for all $t_1 \neq t_2 \in [0,1]$,*

$$\frac{|w(t_2) - w(t_1)|}{|t_2 - t_1|^\alpha} \leq C.$$

It is almost surely not $\frac{1}{2}$-Hölder continuous.

The notion of Hölder continuity is closely related to that of finite variation.

Theorem 3.11. *Let $\alpha > 2$. Then w is almost surely of finite V_α norm, which is defined as*

$$\|w\|_{V_r} = \sup_{N\in\mathbb{N}, t_1 < \cdots < t_N} \left(\sum_{i=1}^{N-1} |w(t_{i+1}) - w(t_i)|^r \right)^{1/r}.$$

It almost surely has infinite V_2 norm.

We don't discuss the proofs.
Variation is related to path integrals. The question is: when can we give meaning to an expression of the form

$$\int_0^T F(\gamma(t))\gamma'(t)dt\,?$$

What are the weakest possible assumptions on γ for this object to have a reasonable meaning?
This is simple if γ has finite 1-variation. For γ having finite V_r, $r < 2$ norm one can define the integral in exchange for some regularity assumptions on F ($F \in C^1$) by means of Taylor series. For $2 \leq r < 3$ we can do no such thing, even when more regularity on F is given (and therefore more terms of the Taylor series are available). This leads to a theory of *rough paths.* Paths given by Brownian motion give examples of such rough paths. The extra information that we require to define the path integral is the value of the integral

$$\int_0^T \gamma(t)\gamma'(t)dt.$$

In the context of the Brownian motion it makes sense to *define*

$$\int_0^T w(t)w'(t)dt := \frac{1}{2}(w(T)^2 - T).$$

This leads to Itō integrals.

End of lecture 18. December 22, 2015

4 Integration on manifolds

We will be discussing integration on submanifolds of $\mathbb{R}^d$. It is also possible to define manifolds intrinsically, without reference to surrounding $\mathbb{R}^d$, but for what we have to say this level of generality is unnecessary and has to be deferred to later.

We first recall and highlight a few facts about Radon measures.

A Radon measure ν on $\mathbb{R}^d$ is an outer measure generated by the family of dyadic cubes $\mathcal{T}$ and a positive function τ thereon,

$$\mathcal{T} = \{Q \text{ a dyadic cube in } \mathbb{R}^d\}$$

$$\tau \colon \mathcal{T} \to [0, \infty),$$

satisfying the martingale and regularity conditions

(1) For all $Q \in \mathcal{T}$
$$\tau(Q) = \sum_{\substack{Q' \subset Q \\ k'+1=k}} \tau(Q')$$

(2) For all $Q \in \mathcal{T}$
$$\lim_{N \to \infty} \sum_{\substack{Q' \subset Q \\ k'+N=k \\ \overline{Q'} \not\subset Q}} \tau(Q') = 0.$$

For an arbitrary subset $E \subset \mathbb{R}^d$ we defined $\nu(E)$ via an infimum taken over coverings of E by dyadic cubes:

$$\nu(E) = \inf_{\substack{\mathcal{T}' \subset \mathcal{T} \\ E \subset \bigcup \mathcal{T}'}} \sum_{Q' \in \mathcal{T}'} \tau(Q').$$

Recall that we proved that for $Q \in \mathcal{T}$, $\tau(Q) = \nu(Q)$. We called a set E measurable with respect to ν if for any $Q \in \mathcal{T}$ the equality

$$\nu(E \cap Q) + \nu(E^c \cap Q) = \nu(Q)$$

holds. Note that in particular every $Q \in \mathcal{T}$ is measurable and therefore all Borel sets are ν-measurable. We called a function $f \colon \mathbb{R}^d \to [0, \infty]$ measurable

if for all $\lambda > 0$ the preimage $f^{-1}((\lambda, \infty])$ is a measurable set. If a function f is Borel measurable, then it is also measurable.
Next, for any function $f\colon \mathbb{R}^d \to [0, \infty]$ we defined the integral as

$$\int f d\nu := \int_0^\infty \nu(x\colon f(x) > \lambda) d\lambda$$

with the right hand side understood as the integral of a monotonically decreasing function, which we introduced as Newton integral in Analysis I. The integral has good properties if one restrict attention to emasurable functions. In particular, for f, g both ν-measurable, we have

$$\int f + g \, d\nu = \int f d\nu + \int g d\nu.$$

Now, let μ and ν be Radon measures and let $Q_{x,k}$ denote the unique dyadic cube of order k that contains x. We proved that for μ-almost all $x \in \mathbb{R}^d$ the limit

$$\lim_{k \to -\infty} \frac{\nu(Q_{x,k})}{\mu(Q_{x,k})}$$

exists. Then we defined $g(x)$ to be equal to the limit above if it exists, if the limit does not exists we put $g(x) = 0$. The function g is called the Radon-Nikodym derivative of ν with respect to μ. It is μ-measurable as a.e. limit of functions constant on dyadic cubes. It is also Borel measurable since the set where a sequence of Borel functions conerges is Borel measurable. The Radon Nikodym derivative is denoted by $g = d\nu/d\mu$.
Recall that ν is absolutely continuous with respect to μ if for all $Q \in \mathcal{T}$

$$\nu(Q) = \int_Q g d\mu \quad (\text{ in general we only have } \nu(Q) \geq \int_Q g d\mu \ ).$$

Note that if ν is absolutely continuous with respect to μ, then for all measurable $E \subset \mathbb{R}^d$

$$\nu(E) = \int_E g d\mu.$$

By a covering argument the inequality $\geq$ holds. Moreover:

(1) The statement holds for sets of μ-measure zero.

(2) One can validate the equality for all Borel subsets.

(3) Putting the previous steps together it holds for all μ-measurable sets.

From this one can easily obtain that actually

$$\int f d\nu = \int f g d\mu$$

holds for all Borel functions f. Let $\varepsilon > 0$ be arbitrary and

$$E_m = \{2^{\varepsilon m} \leq f \leq 2^{\varepsilon(m+1)}\}.$$

Indeed,

$$\begin{aligned}\int f d\nu = \sum_m \int_{E_m} f d\nu \leq 2^{\varepsilon} \sum_m 2^{\varepsilon m} \int_{E_m} 1 d\nu \\ = 2^{\varepsilon} \sum_m 2^{\varepsilon m} \int_{E_m} g d\mu \leq 2^{\varepsilon} \sum_m \int_{E_m} f g d\mu = 2^{\varepsilon} \int f g d\mu.\end{aligned}$$

Since $\epsilon > 0$ was arbitrary we obtain $\leq$.

$$\int f d\nu \leq \int f g d\mu.$$

The reverse follows similarly and we leave it as an exercise.
Suppose that ν is absolutely continuous with respect to μ and μ absolutely continuous with respect to ν. Denote by $g = d\nu/d\mu$ and $\tilde{g} = d\mu/d\nu$ the respective Radon-Nikodym derivatives. Using the last statement we proved, for any Borel function f

$$\int f d\nu = \int f g d\mu = \int f g \tilde{g} d\nu.$$

This, however, means that $g\tilde{g}$ is ν-almost everywhere equal 1: let

$$E_{\varepsilon} = \{x \colon g\tilde{g}(x) > 1 + \varepsilon\};$$

then

$$\int 1_{E_{\varepsilon}} d\nu = \int 1_{E_{\varepsilon}} g\tilde{g} d\nu \geq (1+\varepsilon) \int 1_{E_{\varepsilon}} d\nu,$$

which means that $\nu(E_{\varepsilon}) = 0$. The same argument applies to $\{g\tilde{g} < 1-\varepsilon\}$, so $g\tilde{g} = 1$ ν-almost surely. Note that we have similarly that $g\tilde{g} = \tilde{g}g$ is μ-almost everywhere equal to 1.

Definition 4.1 (Pushforward)**.** Let (X_1, ν_1) be an outer measure space, X_2 be a set and $g \colon X_1 \to X_2$. Define an outer measure on X_2

$$\nu_2(E) = \nu_1(g^{-1}(E)),$$

where $g^{-1}(E)$ denotes the preimage of E. The measure ν_2 as above is called the pushforward of ν_1.

Note that $g^{-1}(E) = \{x : g(x) \in E\}$ is defined even if g is not invertible. The notation $g^{-1}(E)$ should not be mistaken as suggesting that g is invertible, though if it is, then we also have $g^{-1}(E) = \{g^{-1}(x), x \in E\}$.
Let (X_1, ν_1), g and (X_2, ν_2) be as in the definition. Let $f\colon X_2 \to [0, \infty]$. We have

$$\int f\, d\nu_2 = \int_0^\infty \nu_2(\{f > \lambda\}) d\lambda = \int_0^\infty \nu_1(\{f \circ g > \lambda\}) d\lambda = \int f \circ g\, d\nu_1.$$

Example (Polar coordinates)**.** Let ν_1 be the Lebesgue measure on $\mathbb{R}^2$ restricted to $[0, \infty) \times [0, 2\pi]$, i.e.

$$\nu_1(Q) = \mu(Q \cap [0, \infty) \times [0, 2\pi]),$$

where μ is the Lebesgue measure. Define

$$g\colon [0, \infty) \times [0, 2\pi] \to \mathbb{R}^2, \quad g(r, \varphi) = (r \cos \varphi, r \sin \varphi).$$

The pushforward of ν_1 is a measure on $\mathbb{R}^2$. One would like to relate that to Lebesgue measure on $\mathbb{R}^2$.

Theorem 4.2. *Suppose that ν_1 is a Radon measure on $\mathbb{R}^d$ and $g\colon \mathbb{R}^d \to \mathbb{R}^d$. Assume that the push forward ν_2 of ν_1 under g is a Radon measure, and assume that Lebesgue measure μ on $\mathbb{R}^d$ is absolutely continuous with respect to ν_2. Then for any Borel function f and with $h = d\mu/d\nu_2$*

$$\int_{\mathbb{R}^d} f d\mu = \int_{\mathbb{R}^d} f h d\nu_2 = \int_{\mathbb{R}^d} (f \circ g)(h \circ g) d\nu_1.$$

Proof. The first inequality follows, because h is the respective Radon-Nikodym derivative and the second inequality follows from the remarks above. $\square$

End of lecture 19. January 7, 2016

Lemma 4.3. *Let ν_1 be a Radon measure on $\mathbb{R}^d$. Let $g : \mathbb{R}^d \to \mathbb{R}^d$ such that*

1. *$\nu_1(g^{-1}(Q)) < \infty$ for all $Q \in \mathcal{T}$*
2. *$g^{-1}(Q)$ is ν_1 measurable for all $Q \in \mathcal{T}$*
3. *There exists $\tilde{E} \subset \mathbb{R}^d$ such that $\nu_1(\tilde{E}^c) = 0$, g is injective on $\tilde{E}$ and $g(Q \cap \tilde{E})$ is measurable for all $Q \in \mathcal{T}$.*

Then the push forward of ν_1 under g is Radon.

Proof. Define $\tau : \mathcal{T} \to [0, \infty)$ by

$$\tau(Q) := \nu_1(g^{-1}(Q))$$

Note that by 1., $\tau(Q)$ is always finite. From 2. it follows that τ satisfies the martingale condition. Indeed,

$$\sum_{\substack{Q' \subset Q \\ k'+1=k}} \tau(Q') = \sum_{\substack{Q' \subset Q \\ k'+1=k}} \nu_1(g^{-1}(Q')) \overset{2.}{=} \tau(Q)$$

To see that τ satisfies the regularity condition fix Q and define

$$E_N := g^{-1}(\bigcup_{\substack{Q' \subset Q \\ k'+N=k \\ \overline{Q'} \not\subset Q}} Q)$$

Note that $\bigcap_N E_N = \emptyset$. We need to show that $\lim_{N \to \infty} \nu_1(E_N) = 0$. Define

$$F_N := g^{-1}(Q) \setminus E_N, \quad G_N := F_N \setminus F_{N-1}.$$

Note that the sets G_N are measurable. We have $g^{-1}(Q) = \bigcup_N F_N = F_1 \cup \big(\bigcup_{N>1} G_N \big)$. Then by measurability of G_N

$$\nu_1(g^{-1}(Q)) = \nu_1(F_1) + \sum_{N>1} \nu_1(G_n)$$

Since the series on the right hand-side converges it must hold

$$\nu_1(E_N) = \sum_{M>N} \nu_1(G_M) \to 0$$

as $N \to \infty$. So τ generates a Radon outer measure ν. For all $Q \in \mathcal{T}$ we have $\nu(Q) = \tau(Q) = \nu_2(Q)$. It remains to show that $\nu(E) = \nu_2(E)$ for all $E \subset \mathbb{R}^d$. We have

$$\nu_2(E) \leq \inf_{\mathcal{T}', E \subset \cup \mathcal{T}'} \sum_{Q \in \mathcal{T}'} \nu_2(Q) = \nu(E)$$

where $\mathcal{T}'$ goes over all subcollections of $\mathcal{T}$ which cover E. The last equality is by the definition of the outer measure. The reverse inequality holds for measurable sets without further conditions. Indeed, if E is $\nu-$measurable we have for all $Q \in \mathcal{T}$

$$\nu(E \cap Q) + \nu(Q \setminus E) = \nu(Q)$$
$$\nu_2(E \cap Q) + \nu_2(Q \setminus E) \geq \nu_2(Q)$$

Since $\nu_2(E \cap Q) \leq \nu(E \cap Q)$ and $\nu_2(Q \setminus E) \leq \nu(Q \setminus E)$ we must have

$$\nu(E \cap Q) = \nu_2(E \cap Q)$$

for all $Q \in \mathcal{T}$ and hence $\nu(E) = \nu_2(E)$.
However, by the assumption 3. we have $\nu(E) = \nu_2(E)$ for all sets E. To see this, the idea is to restrict so a conull set $\tilde{E}$ on which g is injective and consider a pushforward of ν_2 under g^{-1} to obtain the reverse inequality. To make this precise, let $\tilde{E}$ be the set from 3. Set $F := g(\tilde{E})$ and let $h : \mathbb{R}^d \to \mathbb{R}^d$ be such that

$$h(x) = \begin{cases} g^{-1}(x) & : x \in F \\ 0 & : x \notin F \end{cases}$$

We have $\nu_2(F^c) = 0$ and $g^{-1}(F^c) \subset \tilde{E}^c$. (Use $\nu_1(\tilde{E}^c) = 0$ and $g^{-1}(g(A)) \supset A$, $g^{-1}(A^c) = (g^{-1}(A))^c$.) Define ν_3 to be the push forward of ν_2 under h. Then $\nu_3(Q) = \nu_1(Q)$ for all $Q \in \mathcal{T}$ since

$$\nu_3(Q) = \nu_2(h^{-1}(Q)) \overset{(1)}{=} \nu_2(h^{-1}(Q \cap \tilde{E})) \overset{(2)}{=} \nu_1(Q \cap \tilde{E}) \overset{(3)}{=} \nu_1(Q)$$

We used (1) : $\nu_2(h^{-1}(Q \cap \tilde{E}^c)) = 0$, since $h^{-1}(Q \cap \tilde{E}^c) \subset h^{-1}(\tilde{E}^c) = h^{-1}(\tilde{E})^c = g(\tilde{E})^c = F^c$ (observe that by injectivity of g on $\tilde{E}$ and injectivity (bijectivity) of h on F we have $g(E) = h^{-1}(E)$ for any $E \subset \tilde{E}$), (2): bijectivity of h on F (3): $\nu_1(\tilde{E}^c) = 0$
Moreover, for all $E \subset \tilde{E}$ we have

$$\nu_3(E) \overset{(1)}{=} \nu_2(g(E)) \overset{(2)}{=} \nu_1(E)$$

where (1) holds since ν_3 is a push forward of ν_2 under h and $g(E) = h^{-1}(E)$ by injectivity of g on $\tilde{E}$, while (2) holds since ν_2 is a push forward of ν_1 under g (again use injectivity of g). Thus, $\nu_1(E) = \nu_3(E)$ for all $E \subset \tilde{E}$. We leave it as an exercise to conclude that $\nu = \nu_2$. □

Now we discuss absolute continuity. The following lemma allows to calculate the Radon Nikodym derivative in some cases.

Lemma 4.4. *Let ν_1, μ be Radon measures on $\mathbb{R}^d$. Let $g : \mathbb{R}^d \to \mathbb{R}^d$. Let $E \subset \mathbb{R}^d$ be open such that $g|_E$ is injective, $\nu_1(E^c) = 0$, $\mu(g(E)^c) = 0$. Let g be continuously differentiable on E and assume its derivative is everywhere non-singular. Moreover, let the assumptions of the previous lemma be satisfied. Then the push forward ν_2 of ν_1 under g is a Radon measure, μ is absolutely continuous with respect to ν_2 and for all $x \in E$*

$$\frac{d\mu}{d\nu_2}(g(x)) = \left|\det Dg|_x\right|. \tag{11}$$

The determinant in (11) is called the *Jacobi determinant of g*.

Proof. Applying the inverse function theorem there exists a ball $B(x_1, r)$ such that g is invertible in $B(x_1, r)$ and

$$D(g^{-1})|_{x_2} = (Dg|_{x_1})^{-1}.$$

Without loss of generality we assume that $x_1 = x_2 = 0$. Consider the parallelepiped

$$P_{x_2,k} = (Dg|_{x_1})^{-1}(Q_{x_2,k}),$$

and

$$\tilde{P}_{x_2,k} = (1+\varepsilon)P_{x_2,k},$$

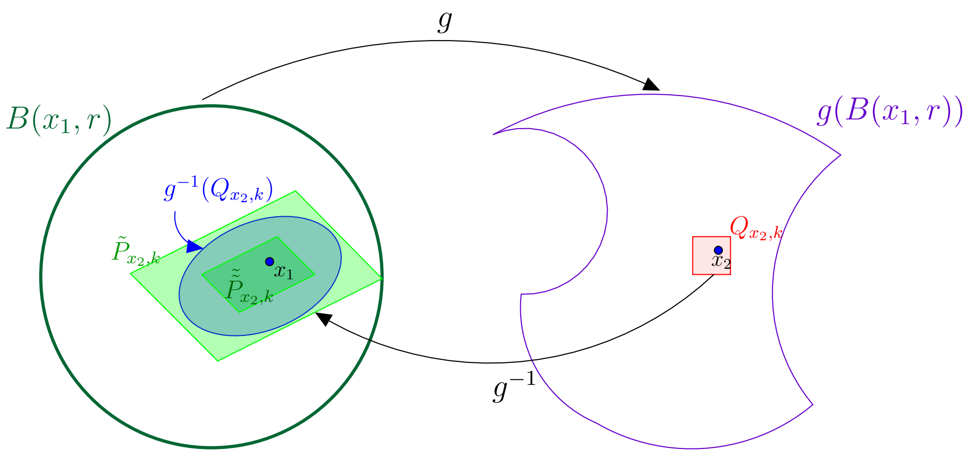
which is a dilated parallelepiped, dilated by a factor $1+\epsilon$ with respect to the center of $P_{x_2,k}$. Notice that for k small enough we have

$$g^{-1}(Q_{x_2,k}) \subset \tilde{P}_{x_2,k}.$$

It holds, because if $y \in Q_{x_2,k}$, then writing the Taylor expansion

$$g^{-1}(y) = \underbrace{(Dg|_{x_1})^{-1}y + \eta|y|}_{\in P_{x_2,k} + \text{ small}\cdot P_{x_2,k} \subset \tilde{P}_{x_2,k}},$$

since $\eta \to 0$ as $k \to -\infty$.



Let

$$\tilde{\tilde{P}}_{x_2,k} = (1-\varepsilon) P_{x_2,k},$$

scaled with respect to the center of $P_{x_2,k}$. Similarly as above, using the Taylor expansion, one can see that

$$\tilde{\tilde{P}}_{x_2,k} \subset g^{-1}(Q_{x_2,k}).$$

The two inclusions that we have just shown give that

$$(1-\varepsilon)^d \mu(P_{x_2,k}) = \mu(\tilde{\tilde{P}}_{x_2,k}) \leq \mu(g^{-1}(Q_{x_2,k})) \leq \mu(\tilde{P}_{x_2,k}) = (1+\varepsilon)^d \mu(P_{x_2,k}).$$

One can easily finish the argument now, since

$$\mu(P_{x_2,k}) = |\det Dg|_{x_1}|^{-1} \mu(Q_{x_2,k})$$

and $\epsilon > 0$ is arbitrarily small. We leave the details as an exercise. □

Polar coordinates in $\mathbb{R}^d$

There are several assumption in the last two lemmata, but they are easily satisfied in practice. To illustrate this we now discuss their application to polar coordinates.

For $d = 2$ this has already been discussed in the previous lecture. However, in this section we instead of $(r, \varphi) \in [0, \infty) \times [-\pi, \pi]$ consider $(r, \varphi) \in \mathbb{R} \times [0, \pi]$ since this is be more convenient for certain arguments. That is, we consider also negative radii but only half of the possible angles. Let ν_1 be the Lebesgue measure on $\mathbb{R}^d$ restricted to $\mathbb{R} \times [0, \pi]^{n-1}$. Define

$$\begin{aligned} g(r, \varphi_1, \ldots, \varphi_{d-1}) = (&r \cos \varphi, \\ &r \sin \varphi_1 \cos \varphi_2, \\ &r \sin \varphi_1 \sin \varphi_2 \cos \varphi_3, \\ &\ldots, \\ &r \Big(\prod_{i<j} \sin \varphi_i \Big) \cos \varphi_j), \\ &\ldots, \\ &r \prod_{j=1}^{d-1} \sin \varphi_i \,) \end{aligned}$$

Let $E := \mathbb{R} \setminus \{0\} \times (0, \pi)^{d-1}$. We claim that g is injective on E. To see that assume that $g(r, \varphi_1, \ldots, \varphi_{d-1}) = (x_1, \ldots, x_n)$. First, by induction on d one can show that

$$r^2 = \sum_{i=1}^{d} x_i^2.$$

For instance, if $d = 2$, then

$$(r\cos\varphi)^2 + (r\sin\varphi)^2 = r^2$$

and the induction step proceeds similarly (exercise). Therefore, $|r|$ is determined by $(x_1, \ldots, x_d)$. Moreover, observe that $\operatorname{sign}(r) = \operatorname{sign}(x_d)$ since sin is positive on $(0, \pi)$. Since $x_1 = r\cos\varphi_1$ and cos is invertible on $[0, \pi]$, we see that φ_1 is determined by $x_1, \ldots, x_d$ and so

$$\varphi_1 = \cos^{-1}\left(\frac{x}{r}\right).$$

Inductively we proceed with φ_i. This shows that g is injective on E. To see that $\mu(g(E)^c) = 0$ it suffices to show that $g(E)$ contains the set

$$\{(x_1, \ldots, x_d) : x_i \neq 0 \ \forall i\}.$$

This can again be seen by an induction argument (exercise). We have

$$\left|\det(Dg|_{r,\varphi_1,\ldots,\varphi_n})\right| = r^{d-1} \prod_{j=1}^{d-1} (\sin\varphi_i)^{d-1}$$

(exercise). Thus, μ is of the form

$$\begin{aligned}
d = 2 : &\quad r dr d\varphi \\
d = 3 : &\quad r^2 \sin\varphi_1 dr d\varphi_1 d\varphi_2 \\
d = 4 : &\quad r^3 \sin^3\varphi_1 \sin\varphi_2 dr d\varphi_1 d\varphi_2 d\varphi_3 \\
&\vdots
\end{aligned}$$

End of lecture 20. January 12, 2016

Recall, for a space with outer measure (X_1, ν_1) and a map $g\colon X_1 \to X_2$ we defined the pushforward measure on X_2 as

$$\nu_2(E) := \nu_1(g^{-1}(E)).$$

The last two times we were discussing change of coordinates in $\mathbb{R}^d$, in particular the polar coordinates, and saw how it behaves under integration -

we computed the pushforward for a continuosly differentiable, almost everywhere injective map g. Today we will compute the pushforward of a from $\mathbb{R}^k$ to $\mathbb{R}^d$ with $k < d$, what will let us to define integration on submanifolds in $\mathbb{R}^d$.
We have the following result, similar to the case $k = d$, which we discussed the last time.

Lemma 4.5. *Let ν_1 be a Radon measure on $\mathbb{R}^k$, $k < d$, and let $g\colon \mathbb{R}^k \to \mathbb{R}^d$. If the following conditions are satisfied*

(1) $\nu_1(g^{-1}(Q)) < \infty$ for every $Q \in \mathcal{T}$

(2) $g^{-1}(Q)$ is ν_1-measurable for all $Q \in \mathcal{T}$

(3) There exists $E \subset \mathbb{R}^k$ with $\nu_1(E^c) = 0$, g is injective on E and $g(Q \cap E)$ is Borel for all $Q \in \mathcal{T}$,

then the pushforward ν_2 is a Radon measure. Additionally $\nu_2(\mathbb{R}^d \backslash im(g)) = 0$.

Remark. We would like to integrate on manifolds with a help of the above lemma. So far it is not clear how to do it. We shall understand the things locally and in this situation unlike $\mathbb{R}^d$ we prefer to use balls instead of cubes, because of their rotational invariance - this way we do not have to worry how 'tilted' is a manifold at a certain point. Another thing is that for a k-dimensional manifold S and $y \in S$ we naturally expect the measure of $S \cap B(y, r)$ to behave like $r^k c_k$ for small r, where c_k is the Lebesgue measure of the unit ball in $\mathbb{R}^k$.

Definition 4.6. A measure in $\mathbb{R}^d$ is called k-rectifiable if for ν-almost all $y \in \mathbb{R}^d$

$$\lim_{r\to 0} \frac{\nu(B(y,r))}{r^k} \text{ exists, is strictly positive and finite.}$$

It is k-rectifiable and normalised if for ν-almost all $y \in \mathbb{R}^d$

$$\lim_{r\to 0} \frac{\nu(B(y,r))}{r^k} = c_k,$$

where c_k is the measure of the unit ball in $\mathbb{R}^k$.

Exercise 4.7. *Let ν, μ be k-recitfiable and normalised Radon measures on $\mathbb{R}^d$. If ν is absolutely continuous with respect to μ and μ is absolutely continuous with respect to ν, then $\nu = \mu$.*

Let us consider an example of a Radon measure in $\mathbb{R}^2$ for which the above limit with $k = 1$ does not exists, although the quotient is uniformly bounded.

Example. Let

$$\mu_0 = \mu \cdot 1_{[0,1]^2},$$

be the Lebesgue measure μ in $\mathbb{R}^2$ restricted to the unit square. Next, divide the square into 16 smaller squares of the same area. Let Q_i^1 for $i = 1, 2, 3, 4$ be the ones that are in the corners of the big square, just as in the picture. Define μ_1 as follows

$$\mu_1 = 4 \cdot \mu \cdot \sum_{i=1}^{4} 1_{Q_i^1}.$$

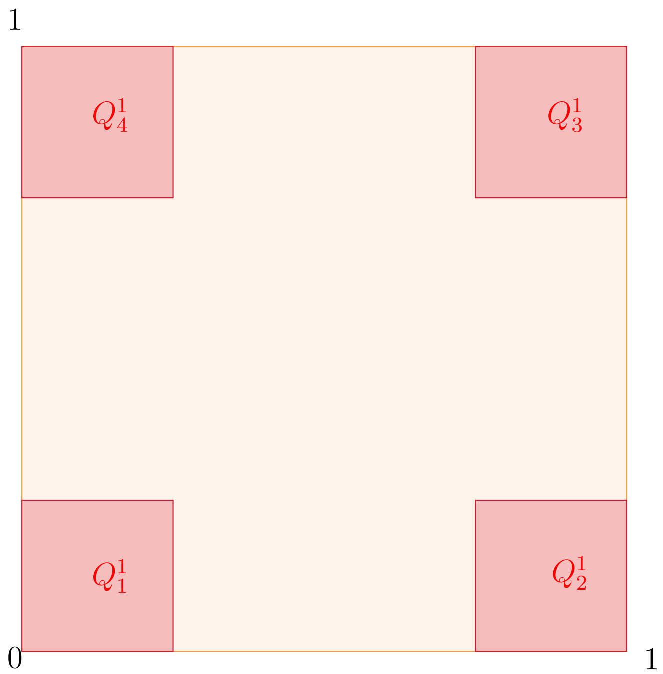


Figure 31: The first step of the construction. μ_1 is supported on the red squares above.

Next, we take each of the corner squares, split it into 16 smaller ones and take the little ones in the corners, just as we did in the first step of the construction (picture). This way we obtain 16 little squares Q_i^2. Then we can define μ_2 similarly

$$\mu_2 = 16 \cdot \mu \cdot \sum_{i=1}^{16} 1_{Q_i^2}.$$

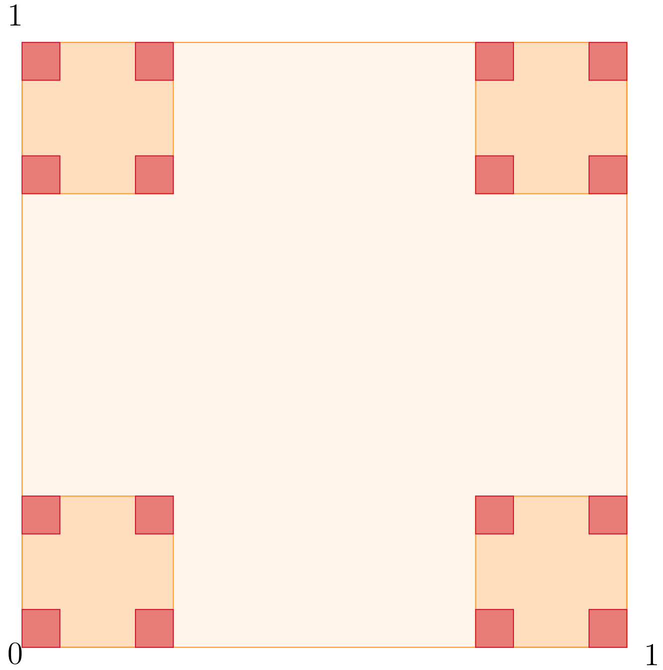


Figure 32: The second step of the construction. Q_i^2 for $i = 1, 2, ..., 16$ are the red squares above. μ_2 is supported on their union.

Contiuining in this manner we obtain the sequence of measures μ_k for $k \in \mathbb{N}$ and for each $k \in \mathbb{N}$, $\mu_k(\mathbb{R}^2) = 1$. One can show that the limit τ defined for $Q \in \mathcal{T}$

$$\tau(Q) = \lim_{k \to \infty} \mu_k(Q)$$

generates a Radon measure μ. We leave it as an exercise. Note that μ is supported on the intersection of the unions of the squares we chose at each step, i.e. the Cantor set C. Next, we observe that $\nu(B(y,r))/r$ is uniformly bounded.

Lemma 4.8. *There exist* $0 < \gamma_1, \gamma_2 < \infty$ *such that for* $y \in C$ *we have*

$$\gamma_1 \leq \frac{\nu(B(y,r))}{r} \leq \gamma_2.$$

Sketch. This is not so difficult to see. Roughly, if $y \in C$, then $\nu(B(y,r))$ behaves up to uniform constants γ_1 and γ_2 like $\nu_n(Q_i^n)$ where 2^{-2n} is the order of Q_i^n - the biggest cube contained in $B(y,r)$. Next, notice that $r \sim 2^{-2n}$ and $\nu_n(Q_i^n) \sim 2^{-2n}$, although the Lebesgue measure of Q_i^n is equal to 2^{-4n}, that is where the scaling according to r^1 comes from. $\square$

Moreover, the $\liminf$ and $\limsup$ are different. Take, for example, a point $y \in C$ to be one of the corners of $[0,1]^2$. There exist (picture) arbitrarily small r_1, r_2 with $(1+\varepsilon)r_1 < r_2$ such that

$$B(y,r_1) \cap C = B(y,r_2) \cap C.$$

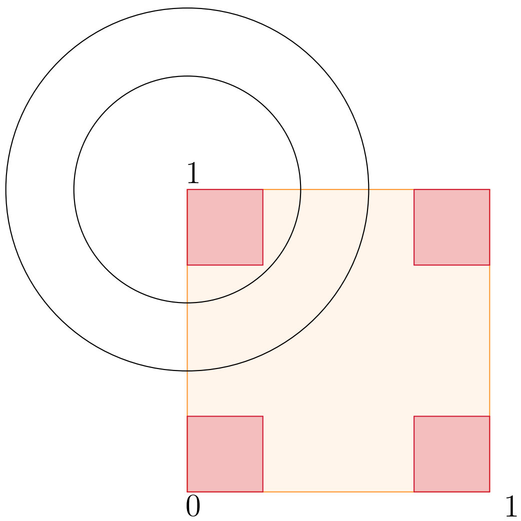


Figure 33: An example of $B(y,r_1)$, $B(y,r_2)$ with $(1+\epsilon)r_1 < r_2$ and $B(y,r_1) \cap C = B(y,r_2) \cap C$. Clearly this construction can be done for arbitrarily small scales, preserving the ratio r_1/r_2.

Theorem 4.9. *Let $0 < k < d$, μ the Lebesgue measure in $\mathbb{R}^k$, $g : \mathbb{R}^k \to \mathbb{R}^d$ and $E \subset \mathbb{R}^k$ open such that $g|_E$ is injective. Assume that g is continuously differentiable on E and that $\mathrm{rg}(Dg|_x) = k$ for all $x \in E$. Then the push forward ν_2 of $\mu \mathbf{1}_E$ is a k-rectifiable Radon measure and the push forward of*

$$\mu \mathbf{1}_E \sqrt{\det(Dg|_x^T \cdot Dg|_x)} \tag{12}$$

is k-rectifiable normalized.

Remark. Note that since $Dg|_x \colon \mathbb{R}^k \to \mathbb{R}^d$, $Dg|_x^T \circ Dg|_x \colon \mathbb{R}^k \to \mathbb{R}^k$ and is nonsingular, so the determinant under the square root makes sense.

Observe that the theorem lets us to define the surface integral, since we have

$$\int_{g(E)} f d\nu_2 = \int_{\mathbb{R}^k} (f \circ g) \sqrt{|\det(Dg|_x^T \cdot Dg|_x)|} d\mu.$$

Sketch. The proof follows from approximation by linear functions. We leave the details as an exercise and consider here only the case when g is linear. Let Γ be the image of g in $\mathbb{R}^d$. Denote by e_i the i-th unit vector in $\mathbb{R}^k$. Hence

$$e_i \overset{g}{\mapsto} v_i = \sum_{j=1}^{d} \langle v_i, e_j \rangle e_j.$$

First case. Assume initially that $v_1, v_2, ..., v_k$ form an orthonormal basis. Let $y \in \Gamma$, we have that

$$g^{-1}(B(y, r)) = B(g^{-1}(y), r),$$

because with this assumptions g is an isometry. Hence,

$$\mu(g^{-1}(B(y, r))) = c_k r^k.$$

Second case. $v_1, v_2, ..., v_k$ linearly independent. Choose an orthonormal basis $w_1, w_2, ..., w_k$ of Γ. Then we have

$$v_i = \sum_{j=1}^{k} \langle v_i, w_j \rangle w_j.$$

Let f be defined as

$$f(y) = (\langle y, w_1 \rangle, \langle y, w_2 \rangle, ..., \langle y, w_k \rangle).$$

For $y \in \Gamma$

$$\nu_2(B(y,r)) = \mu(g^{-1}(B(y,r))) = \mu(g^{-1}f^{-1}(B(z,r))) = |\det(f \circ g)|^{-1} c_k r^k,$$

for $z = f(y)$, by the previous lecture, since $f \circ g \colon \mathbb{R}^k \to \mathbb{R}^k$. This means that we are just left with calculating the determinant in terms of g

$$\begin{aligned}\det(f \circ g)^2 &= \det(\langle v_i, w_j \rangle)^2 = \det(\langle v_i, w_j \rangle) \det(\langle w_j, v_i \rangle) \\ &= \det(\sum_j \langle v_i, w_j \rangle \langle w_j, v_{i'} \rangle) = \det(\sum_j \langle v_i, w_j \rangle \langle w_j, w_j \rangle \langle w_j, v_{i'} \rangle) \\ &\qquad\qquad\qquad\qquad\qquad = \det(\langle v_i, v_{i'} \rangle) = \det(g^T \circ g).\end{aligned}$$

□

End of lecture 21. January 14, 2016

In the last lecture we discussed the following theorem regarding the surface measure:

Let $0 < k < d$, μ the Lebesgue measure in $\mathbb{R}^k$, $g : \mathbb{R}^k \to \mathbb{R}^d$ and $E \subset \mathbb{R}^k$ open such that $g|_E$ is injective. Assume that g is continuously differentiable on E and that $\operatorname{rg}(Dg|_x) = k$ for all $x \in E$. Then the push forward ν_2 of $\mu \mathbf{1}_E$ is a k-rectifiable Radon measure and the push forward of

$$\mu \mathbf{1}_E \sqrt{\det(Dg|_x^T \cdot Dg|_x)} \tag{13}$$

is k-rectifiable normalized.

Our goal is to calculate the determinant in (13). Fix a point $x \in E$ and set

$$v_i := \partial_i g|_x \quad \text{for} \quad i = 1, \dots, k.$$

Since g maps into $\mathbb{R}^d$ it is of the form $g = (g_1, \dots, g_d)$. Set

$$v_{ij} := \partial_i g_j \quad \text{for} \quad i = 1, \dots, k,\ j = 1, \dots, d$$

Then $(v_{ij})_{ij}$ is the matrix $Dg|_x$. [12] Observe that

$$\det(Dg|_x^T \cdot Dg|_x) = \det(A)$$

[12] v_{ij} is in the $i-th$ column and $j-th$ row

where the $k \times k$ matrix A is given by

$$A_{il} = \sum_{j=1}^{d} v_{ij} v_{lj} = \langle v_i, v_l \rangle.$$

The determinant in question is the volume of the parallelepiped spanned by $v_1, \ldots, v_k$. To compute this volume we are interested in the length of the spanning vectors and the angles between them. Note that the scalar product encodes exactly these quantities - the length of the vectors and the angle between them.

To compute the determinant in (13) we first recall some facts from linear algebra. Write $\underline{k} = \{1, \ldots, k\}$. The determinant of a $k \times k$ matrix A is computed by the formula

$$\det(A) = \sum_{\substack{\sigma:\underline{k}\to\underline{k} \\ \text{bijective}}} \varepsilon(\sigma) \prod_{j=1}^{k} A_{i\sigma(i)}$$

where $\varepsilon(\sigma) \in \{-1, 1\}$ is the sign of the permutation σ. The sign is determined by the properties $\varepsilon(\sigma\tilde{\sigma}) = \varepsilon(\sigma)\varepsilon(\tilde{\sigma})$ and $\varepsilon(\sigma) = -1$ if σ is a transposition (a permutation, which exchanges two elements and keeps all others fixed).
Then we can write

$$\begin{aligned}
\det(Dg|_x^T \cdot Dg|_x) &= \sum_{\substack{\sigma:\underline{k}\to\underline{k} \\ \text{bij.}}} \varepsilon(\sigma) \prod_{i=1}^{k} \sum_{j=1}^{d} v_{ij} v_{\sigma(i)j} \\
&\overset{(1)}{=} \sum_{\substack{\sigma:\underline{k}\to\underline{k} \\ \text{bij.}}} \varepsilon(\sigma) \sum_{\rho:\underline{k}\to\underline{d}} \prod_{i=1}^{k} v_{i\rho(i)} v_{\sigma(i)\rho(i)} \\
&\overset{(2)}{=} \sum_{\substack{\sigma:\underline{k}\to\underline{k} \\ \text{bij.}}} \varepsilon(\sigma) \sum_{\substack{\rho:\underline{k}\to\underline{d} \\ \text{injective}}} \prod_{i=1}^{k} v_{i\rho(i)} v_{\sigma(i)\rho(i)} \\
&\overset{(3)}{=} \sum_{\substack{\sigma:\underline{k}\to\underline{k} \\ \text{bij.}}} \varepsilon(\sigma) \sum_{\substack{\tilde{\rho}:\underline{k}\to\underline{d} \\ \text{monotone}}} \sum_{\substack{\tilde{\sigma}:\underline{k}\to\underline{k} \\ \text{bij.}}} \prod_{i=1}^{k} v_{i\tilde{\rho}(\tilde{\sigma}(i))} v_{\sigma(i)\tilde{\rho}(\tilde{\sigma}(i))} \\
&\overset{(4)}{=} \sum_{\substack{\rho:\underline{k}\to\underline{d} \\ \text{mon.}}} \sum_{\substack{\tilde{\sigma}:\underline{k}\to\underline{k} \\ \text{bij.}}} \sum_{\substack{\sigma:\underline{k}\to\underline{k} \\ \text{bij.}}} \varepsilon(\tilde{\sigma})\varepsilon(\tilde{\sigma}\sigma^{-1}) \prod_{i=1}^{k} v_{i\rho(\tilde{\sigma}(i))} \prod_{j=1}^{k} v_j \rho(\tilde{\sigma}(\sigma^{-1}(j)))
\end{aligned}$$

$$\overset{(5)}{=} \sum_{\substack{\rho:\underline{k}\to\underline{d}\\ \text{mon.}}} \sum_{\substack{\tilde{\sigma}:\underline{k}\to\underline{k}\\ \text{bij.}}} \sum_{\substack{\tilde{\tilde{\sigma}}:\underline{k}\to\underline{k}\\ \text{bij.}}} \varepsilon(\tilde{\sigma})\varepsilon(\tilde{\tilde{\sigma}}) \prod_{i=1}^{k} v_{i\rho(\tilde{\sigma}(i))} \prod_{j=1}^{k} v_{j\rho(\tilde{\tilde{\sigma}}(j))}$$

$$\overset{(6)}{=} \sum_{\substack{\rho:\underline{k}\to\underline{d}\\ \text{mon.}}} (\det(v_{i\rho(j)}))^2$$

In this calculation we have argued as follows.
(1): Distributive law. In each of the factors (parametrized by i) we pick the element with the index $v_{i,\rho(i)}v_{\sigma(i)\rho(i)}$ (i.e. $j = \rho(i)$) and sum over all possible choices.
(2): If ρ is not injective, i.e. if $\rho(i) = \rho(i') = j$ for $i \neq i'$, then for a fixed σ, the product contains a factor $v_{\sigma(i)j}v_{\sigma(i')j}$. Now consider a permutation $\tilde{\sigma}$ for which $(\tilde{\sigma}(i), \tilde{\sigma}(i')) = (\sigma(i), \sigma(i'))$ and agrees with σ on all other elements. Then the product corresponding to $\tilde{\sigma}$ contains the same factor, but the sign of the permutation is different. So the terms corresponding to σ and $\tilde{\sigma}$ cancel each other.
(3): Sorting. We write $\rho = \tilde{\rho}\sigma$ where $\tilde{\rho}$ is the monotone map which has the same image as ρ.
(4): We write $i = \sigma^{-1}(j)$ and use $\varepsilon(\tilde{\sigma})^2 = 1$, $\varepsilon(\sigma) = \varepsilon(\sigma^{-1})$.
(5): We denote $\tilde{\tilde{\sigma}} = \tilde{\sigma} \circ \sigma^{-1}$.
(6): The expressions corresponding to $\tilde{\sigma}$ and $\tilde{\tilde{\sigma}}$ split and are equal.

By this calculation we have

$$\sqrt{\det(Dg|_x^T \cdot Dg|_x)} = \sqrt{\sum_{\substack{\rho:\underline{k}\to\underline{d}\\ \text{mon.}}} (\det(v_{i\rho(j)}))^2}.$$

The expression on the right hand side means that from the $d \times k$ matrix $(v_{il})_{il}$ we choose k rows (corresponding to the image of ρ), get a $k \times k$ matrix and compute its determinant. Then we sum the squares of these determinants over all possible choices (monotone maps ρ) and finally take the square root. The expression on the right hand side is the length of the vector

$$\varphi^{(\rho)} := \det(v_{i\rho(j)})$$

in $\mathbb{R}^{\binom{d}{k}}$ (paramterized by ρ; dependence on x is understood despite not being present in the notation). Therefore, the area of the parallelepiped spanned by $v_1, \ldots, v_k$ is computed as

$$\sqrt{\det((\langle v_i, v_l\rangle)_{il})} = \sqrt{\sum_{\substack{\rho:\underline{k}\to\underline{d}\\ \text{mon.}}} (\varphi^{(\rho)})^2}.$$

By the above discussion, the right hand-side should be understood that we project the parallelepiped onto k-dimensional subspaces, compute the volumes of the projections and then take the Euclidean norm (in $\mathbb{R}^{\binom{d}{k}}$) of the vector of these volumes.

We claim that the direction of the vector $\varphi^{(\rho)}$ is, up to a sign, independent of the parametrization g; it depends only on the *tangent space* at $g(x)$ (i.e. the $k-$dimensional space spanned by the vectors $v_1, \ldots, v_k$). To see that, assume we change the coordinates such that

$$\tilde{v}_i = \sum_{i=1}^{k} B_{il} v_l$$

where B is a change of basis matrix. Then $\tilde{v}_1, \ldots, \tilde{v}_k$ span the same linear space. We have

$$\begin{aligned} \tilde{\varphi}^{(\rho)} &= \det(\tilde{v}_{i\rho(j)}) \\ &= \det\Big(\sum_{l=1}^{k} B_{il} v_{l\rho(j)}\Big) \\ &= \det(B)\det(v_{l\rho(j)}) \\ &= \det(B)\varphi^{(\rho)} \end{aligned}$$

Thus, one vector is a multiple of the other (and the one-dimensional space spanned by the vectors $\varphi^{(\rho)}$ is independent of the choice of the basis.) Therefore, only the length of the vector $\varphi^{(\rho)}$ depends on the parametrization g. Its direction does not (modulo a sign).

Now let $e^{(\rho)}_{g(x)}$ be the unit vector in $\mathbb{R}^{\binom{d}{k}}$ (depending on $g(x)$) which is parallel to $\varphi^{(\rho)}$ (pointing in the same direction such that their inner product is positive). Then

$$\sqrt{\det(Dg|_x^T \cdot Dg|_x)} = \|\varphi^{(\rho)}\| = \sum_{\rho} \varphi^{(\rho)} e^{(\rho)}_{g(x)} = \langle \varphi^{(\rho)}, e^{(\rho)}_{g(x)} \rangle \tag{14}$$

Example. Recall that in Analysis II we calculated lengths of C^1-curves $\gamma : [0,1] \to U \subset \mathbb{R}^d$ by the integral

$$\int_0^1 \|\gamma'(t)\| dt \tag{15}$$

More generally, we introduced path integrals along γ of (continuous) vector fields $F : U \to \mathbb{R}^d$ as

$$\int_\gamma F := \int_0^1 F(\gamma(t)) \cdot \gamma'(t) dt$$

If the vector field is a gradient field, and $\gamma(0) = x$, $\gamma(1) = y$, then

$$\int_\gamma F = f(y) - f(x) = \int_0^1 \nabla f(\gamma(t)) \cdot \gamma'(t) dt \tag{16}$$

These constructions should be compared to what we have done in today's lecture. The expression $\|\gamma'\|$ in (15) should be compared to (14). Generalizing (14), for $F : \mathbb{R}^d \to \mathbb{R}^{\binom{d}{k}}$ we can define the integral

$$\int \mathbf{1}_E \langle Dg|_x^{(\rho)}, F_{g(x)}^{(\rho)} \rangle d\mu$$

which should be compared with the right hand-side of (16) (g corresponds to the parametrization of γ and we integrate in the parametrization parameter). Such expressions lead to the theory of differential forms.

Example. Let $k = d - 1$. Then $e_{g(x)}^{(\rho)}$ is the unit vector orthogonal to the tangent space at $g(x)$.
To see that, first recall that $\binom{d}{d-1} = d$. Let e be a unit vector orthogonal to the tangent vectors $v_1, \ldots, v_{d-1}$. Now compute $\det(e, v_1, \ldots, v_{d-1})$ by the Laplace expansion, i.e. formula for the determinant of a matrix in terms of its minors (exercise).

End of lecture 22. January 19, 2016

In the last two lectures we intoduced the surface integral via computing the pushforward $\mu\sqrt{\det(Dg|_x^T \circ Dg|_x)}$ for a continuosly differentiable and almost everywhere injective function $g \colon \mathbb{R}^k \to \mathbb{R}^d$ with $k < d$. This time we are going to follow a different approach. It will be particularly useful in the near future while proving Stokes' theorem. The idea is that we want make sense of the following expression

$$\int_{R^d} \delta(f(x)) dx,$$

where δ is the Dirac delta function. Heuristically it makes sense - we would restrict ourselves to integration over the zero set of a function f, which defines a subsurface in $\mathbb{R}^d$. How to make this idea precise?

The plan is to approximate δ by a family of 'good kernels', note that the definition is slightly different than the one we considered before. Later we correct this definition of the delta integration a little bit (namely, we normalize) in order to agree with our previous surface integral.

Definition 4.10 (Good kernels)**.** We call a family of functions $\{\phi_\varepsilon\}_{\varepsilon>0}$ a family of good kernels if the following conditions are satisfied for each $\varepsilon > 0$ and a constant $C > 0$

1. $\phi_\varepsilon \colon \mathbb{R} \to \mathbb{R}_{\geq 0}$ is continuously differentiable
2. $supp(\phi_\varepsilon) \subset [-\epsilon, \epsilon]$
3. $\int \phi_\epsilon(x)dx = 1$
4. $\|\phi_\varepsilon'\|_\infty \leq \frac{C}{\epsilon^2}$

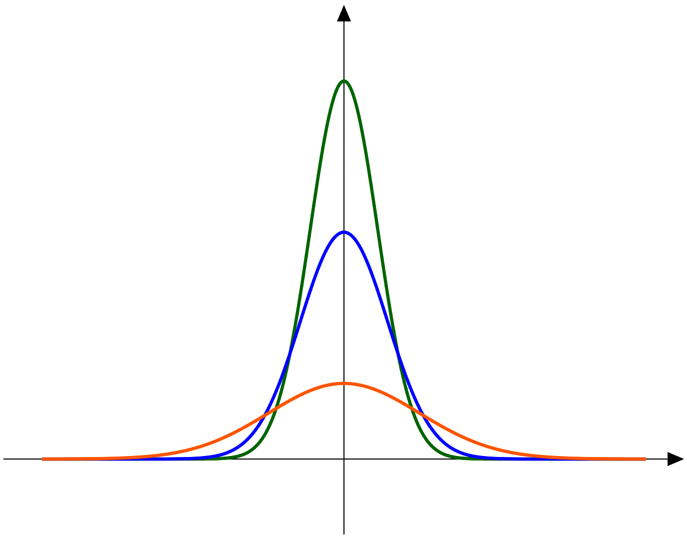


Figure 34: Behaviour of good kernels as $\varepsilon \to 0$. One can see that they approach the Dirac measure, which can be imagined as a single point at infinity for $x = 0$.

Let μ be the Dirac measure at 0, i.e. satisfying

$$\mu(E) = \begin{cases} 1 & , 0 \in E \\ 0 & , 0 \notin E \end{cases}.$$

The first key property of good kernels is that ϕ_ε converge to μ in the following sense.

Proposition 4.11. *Let $g\colon \mathbb{R} \to \mathbb{R}$ be continuous. We have that*

$$\lim_{\varepsilon \to 0} \int g(x)\phi_\varepsilon(x)dx = g(0) = \int g d\mu.$$

Proof.

$$\begin{aligned}\int g(x)\phi_\varepsilon(x)dx &\overset{supp(\phi_\varepsilon)\subset[-\epsilon,\epsilon]}{=} \int_{-\varepsilon}^{\varepsilon} g(x)\phi_\varepsilon(x)dx \\ &= \int_{-\varepsilon}^{\varepsilon} g(0)\phi_\varepsilon(x)dx + \int_{-\varepsilon}^{\varepsilon}(g(x)-g(0))\phi_\varepsilon(x)dx \\ &\overset{\int \phi_\epsilon(x)dx=1}{=} g(0) + \int_{-\varepsilon}^{\varepsilon}(g(x)-g(0))\phi_\varepsilon(x)dx\end{aligned}$$

We are just left with estimating the second term on the right hand side

$$|\int_{-\varepsilon}^{\varepsilon}(g(x)-g(0))\phi_\varepsilon(x)dx| \leq \int_{-\varepsilon}^{\varepsilon}|g(x)-g(0)|\phi_\varepsilon(x)dx \leq \sup_{x\in[-\varepsilon,\varepsilon]}|g(x)-g(0)|.$$

The supremum on the right hand side tends to 0 as $\varepsilon \to 0$ by continuity of g. □

The next proposition describes the behaviour of ϕ_ε when composed with a function f.

Proposition 4.12. *Let $f\colon \mathbb{R} \to \mathbb{R}$ be a continuously differentiable function with finitely many roots $x_1, x_2, ..., x_n$ and $\lim_{|x|\to\infty}|f(x)| = \infty$. Then it holds that*

$$\lim_{\varepsilon\to 0}\int \phi_\varepsilon(f(x))dx = \sum_{i=1}^{n}\frac{1}{|f'(x_i)|}.$$

Proof. Choose ρ small enough so that f is well approximated by its Taylor approximation on $[x_i-\rho, x_i+\rho]$. We will make this choice precise later in the proof depending on ε. Let β and $\beta_1, \beta_2, ..., \beta_n$ be nonnegative functions with

$$\beta + \sum_{i=1}^{n}\beta_i = 1$$

satisfied pointwise, such that β_i is supported in $[x_i-\rho, x_i+\rho]$ and equal 1 in $[x_i-\rho/2, x_i+\rho/2]$. This can be done for ρ small enough since f has finitely many roots, hence they are well separated. Moreover taking ε small enough we can assume that *supp*$(\phi_\varepsilon \circ f) \subset$ *supp*$(\beta)^c$. We have

$$\int_{\mathbb{R}}\phi_\varepsilon(f(x))dx = \sum_{i=1}^{n}\int\phi_\varepsilon(f(x))\beta_i(x)dx + \underbrace{\int\phi_\varepsilon(f(x))\beta(x)dx}_{=0 \text{ for small } \varepsilon}.$$

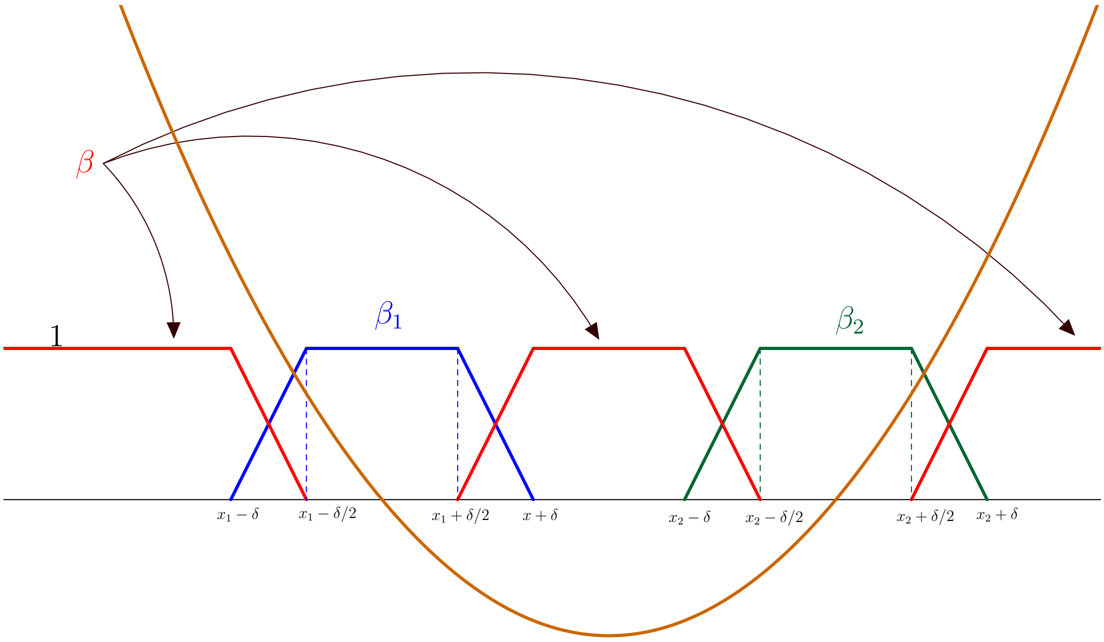


Figure 35: Example of β_1, β_2 and β in the case when f has two roots.

The construction of β's might seem a bit confusing at first, but the general thing to keep in mind is that we want the above equality to hold, so we can localize around each root, consider them separately and 'forget' about the β part that is away from the roots. The last display, for ε small enough so that $\phi_\varepsilon \circ f$ is supported in the union of $[x_i - \rho/2, x_i + \rho/2]$, is equal to

$$\sum_i \int_{x_i-\rho/2}^{x_i+\rho/2} \phi_\varepsilon(f(x))dx$$
$$= \sum_i \int_{x_i-\rho/2}^{x_i+\rho/2} \phi_\varepsilon(f'(x_i)(x-x_i))dx$$
$$+ \sum_i \int_{x_i-\rho/2}^{x_i+\rho/2} \phi_\varepsilon(f(x)) - \phi_\varepsilon(f'(x_i)(x-x_i))dx$$
$$= \underbrace{\sum_i \int_{x_i-\rho/2}^{x_i+\rho/2} \phi_\varepsilon(f'(x_i)(x-x_i))dx}_{\text{by change of variables equals } \sum_i \frac{1}{|f'(x_i)|}} + \sum_i R_i$$

The last thing we need to do is to show that the first order Taylor expansion is good enough to make $\sum_i R_i$ tend to 0 as $\epsilon \to 0$ (together with a right choice of ρ). Let us estimate

$$|\phi_\varphi(f(x)) - \phi_\varphi(f'(x_i)(x-x_i))| = |\int_{f'(x_i)(x-x_i)}^{f(x)} \phi'_\varepsilon(t)dt|$$
$$\leq \|\phi'_\varepsilon\|_\infty |f(x) - f'(x_i)(x-x_i)| \leq \frac{1}{\varepsilon^2} \cdot |x - x_i| \cdot \eta.$$

The last inequality holds due to our assumption on the supremum norm of ϕ'_ε. Note also that we made a use of the fact that $f(x_i) = 0$. Thus, making sure that $|x - x_i| \leq C\varepsilon$ and by Taylor's theorem we have that $\eta \to 0$ as $\varepsilon \to 0$. We can bound the last display by

$$\frac{1}{\varepsilon^2} \cdot C\varepsilon \cdot \eta = C\frac{\eta}{\epsilon}.$$

This gives

$$|R_i| \leq C\eta,$$

for each i. As we mentioned earlier $\eta \to 0$ as $\varepsilon \to 0$, so we are done. □

Example. By the previous proposition (one can also check it easily performing a direct change of variables) we have in particular

$$\lim_{\varepsilon \to 0} \int \phi_\varepsilon(ax)dx = \frac{1}{a}.$$

Let us recall the implicit function theorem, it will be useful for us in the proof of the main theorem of this lecture.

Theorem 4.13. *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable with $f(0,0) = 0$ and $Df|_{(0,0)} = (0,\lambda)$, $\lambda \neq 0$. If $\epsilon > 0$ is small enough, then there exists a continuously differentiable $\gamma\colon [-\varepsilon,\varepsilon] \to \mathbb{R}$ with $\gamma(0) = 0$ and having the following property: if $(x,y) \in B((0,0),\varepsilon)$, then $f(x,y) = 0$ if and only if $(x,y) = (x,\gamma(x))$.*

In other words, the theorem gives a parametrization of the set $\{f(x,y) = 0\}$ in the neighbourhood of $(0,0)$.
Now, let $\alpha \in C_c(\mathbb{R}^2)$ be supported in $B((0,0),\varepsilon)$ for a small ε. Define $g\colon (-\varepsilon,\varepsilon) \to \mathbb{R}^2$ as $\tilde{g}(t) = (t,\gamma(t))$. Then $D\tilde{g} = (1,\gamma')$ and our former surface integral is of the form

$$\int \alpha(t,\gamma(t))\sqrt{1+\gamma'(t)^2}dt.$$

We are going to it compare with the quantity

$$\lim_{\varepsilon\to 0}\int\int \phi_\varepsilon(f(x_1,x_2))\alpha(x_1,x_2)dx_2dx_1.$$

Passing to the limit under the innermost integral sign we obtain using the last proposition that

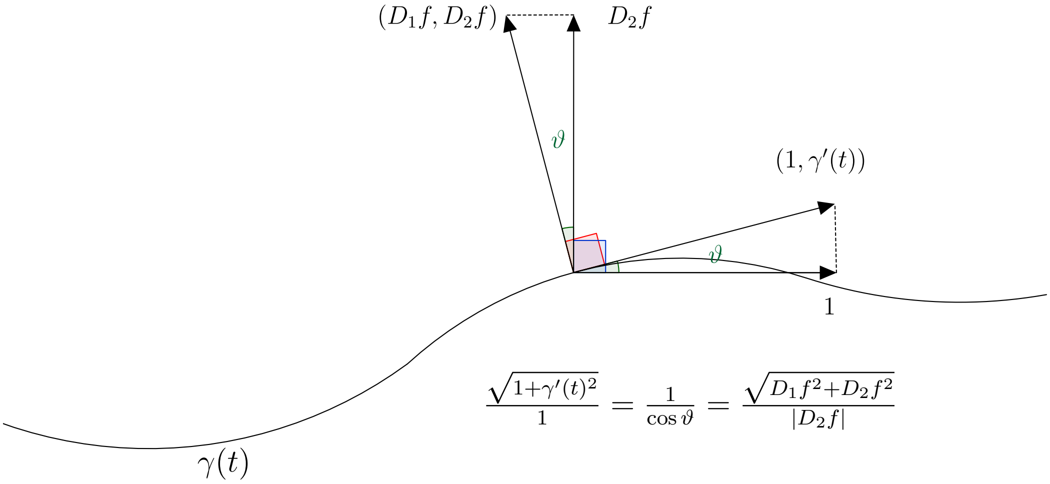
$$\int \frac{1}{D_2f|_{(x_1,\gamma(x_1))}}\alpha(x_1,\gamma(x_1))dx_1,$$

where D_2 is the second component of the gradient. We leave as an exercise to work out the details. Now the question is: what is the factor A that makes the equality

$$\int \alpha(t,\gamma(t))\sqrt{1+\gamma(t)'^2}dt = \lim_{\varepsilon\to 0}\int\int \phi_\varepsilon(f(x_1,x_2))\alpha(x_1,x_2)\cdot A(x_1,x_2)dx_1dx_2$$

hold? A short computation based on the picture below shows that

$$\sqrt{1+\gamma'^2} = \frac{\sqrt{D_1f^2+D_2f^2}}{|D_2f|}$$



Using this we obtain

$$\int \alpha(t,\gamma(t))\sqrt{1+\gamma(t)'}dt$$
$$= \lim_{\varepsilon\to 0}\int\int \phi_\varepsilon(f(x_1,x_2))\alpha(x_1,x_2)\sqrt{D_1 f|^2_{(x_1,x_2)} + D_2 f^2_{(x_1,x_2)}}dx_1dx_2.$$

Finally, we can summarize our computation with the following theorem.

Theorem 4.14. *Let $f\colon \mathbb{R}^2 \to \mathbb{R}$ be continuously differentiable. Let $\alpha \in C_c(\mathbb{R}^2)$. Let $Df|_{(x,y)} \neq (0,0)$ for (x,y)'s with $f(x,y) = 0$, $(x,y) \in supp(\alpha)$. Then the surface integral of α over the curve $\{f = 0\}$ is given by*

$$\lim_{\varepsilon\to 0}\int\int \phi_\varepsilon(f(x_1,x_2))\alpha(x_1,x_2)\sqrt{D_1 f|^2_{(x_1,x_2)} + D_2 f^2_{(x_1,x_2)}}dx_1dx_2.$$

Remark. For the limit above we also use the following notation

$$\int\int \delta(f(x_1,x_2))\alpha(x_1,x_2)\sqrt{D_1 f|^2_{(x_1,x_2)} + D_2 f^2_{(x_1,x_2)}}dx_1dx_2.$$

End of lecture 23. January 26, 2016

Our plan now is to prove a counterpart of the previous theorem in more dimensions for both functions and differential forms. Notice that the actual statement we are after is the equivalence of the surface integral for a parametrised (in the last lecture given by a curve) submanifold and for an implicitly defined (via the zeroes of a function) submanifold. The latter was defined using the delta function and, after a careful computation, properly normalised. We are going to perform similar computations in $\mathbb{R}^d$. Let us start with the sequence of definitions.

Definition 4.15. A parametrised submanifold of dimension k in $\mathbb{R}^d$ is a map $g\colon U \subset \mathbb{R}^k \to \mathbb{R}^d$, where U is open and bounded and g is injective and continuously differentiable; moreover Dg is continuous in the closure of U and $Dg|_x$ has rank k for all $x \in U$.

Definition 4.16. The surface integral of a continuous function $h\colon \mathbb{R}^d \to \mathbb{R}$ over a parametrised submanifold in $\mathbb{R}^d$ is defined as

$$\int_U h(g(x))\left(\sum_\rho (\det Dg|_x^\rho)^2\right)^{1/2} dx.$$

The summation goes over all strictly monotone $\rho\colon \underline{k} \to \underline{d}$ and $Dg|_x^\rho$ is the $k\times k$ matrix $D_i g_{\rho(j)}|_x$.

Definition 4.17. A differential form in $\mathbb{R}^d$ is a continuous map $\omega\colon \mathbb{R}^d \to \mathbb{R}^{\binom{d}{k}}$. We denote the components of ω by ω^ρ.

Definition 4.18. The surface integral of a differential form over a parametrised submanifold in $\mathbb{R}^d$ is defined as

$$\int_U \sum_\rho \det Dg|_x^\rho \omega^\rho(g(x))dx.$$

The summation goes over all strictly monotone $\rho\colon \underline{k} \to \underline{d}$ and $Dg|_x^\rho$ is the $k \times k$ matrix $D_i g_{\rho(j)}|_x$.

Note that the last time we considered a parametrised curve, which of course is a special case of a parametrised submanifold.

Definition 4.19. Let $k < d$. An implicitly defined submanifold of dimension k in $\mathbb{R}^d$ (defined implicitly due to the implicit function theorem) is a continuously differentiable map $f\colon \mathbb{R}^d \to \mathbb{R}^{d-k}$ with $\lim_{|x|\to\infty} |f(x)| = \infty$ and the rank of $Df|_x$ is equal $d-k$ if $f(x) = 0$.

One can define an implicitly defined submanifold as a function like we did above, but it should kept in mind that we are actually interested in the set $\{x\colon f(x) = 0\}$. The assumption $\lim_{|x|\to\infty} |f(x)| = \infty$ is then needed to make sure that this set is bounded. Observe that the implicit function theorem automatically gives us regularity of the set $\{x\colon f(x) = 0\}$.

Definition 4.20. The surface integral of a continuous function $h\colon \mathbb{R}^d \to \mathbb{R}$ over an implicitly defined submanifold is defined as

$$\int_{\mathbb{R}^d} \delta f(x) h(x) \left(\sum_\rho (\det Df|_x^{\tilde\rho})^2 \right)^{1/2} dx,$$

where $\rho\colon \underline{k} \to \underline{d}$ is strictly monotone as before and $\tilde\rho\colon \underline{d-k} \to \underline{d}$ is a strictly monotone function determined by ρ, i.e. $im(\rho) \cap im(\tilde\rho) = \emptyset$.

Definition 4.21. The surface integral of a differential form $\omega\colon \mathbb{R}^d \to \mathbb{R}^{\binom{d}{k}}$ over an implicitly defined submanifold in $\mathbb{R}^d$ is defined as[13]

$$\int_U \delta(f(x)) \sum_\rho \varepsilon(\rho) \det Df|_x^{\tilde\rho} \omega^\rho(x)dx,$$

where $\varepsilon(\rho) := \varepsilon(\sigma)$, $\sigma\colon \underline{d} \to \underline{d}$ is a bijection with $\sigma|_{\underline{k}} = \rho$, $\sigma|_{\underline{d-k}} = \tilde\rho$ monotone.

[13] In this context, in the literature one also meets the "wedge" notation $dx_1 \wedge dx_2 \wedge \cdots \wedge dx_k$ or Einstein's notation $\varepsilon_{ijk} f_i g_j$ (tensor calculus).

Theorem 4.22. *Let g be a parametrised submanifold and f an implicitly defined submanifold with*

$$\{x\colon f(x)=0\}=\overline{\{g(y)\colon y\in U\}}.$$

(the left hand side is a closed set, hence the closure on the right hand side) Let $h\colon \mathbb{R}^d\to\mathbb{R}$ continuous and ω be a differential form. Then we have the following.

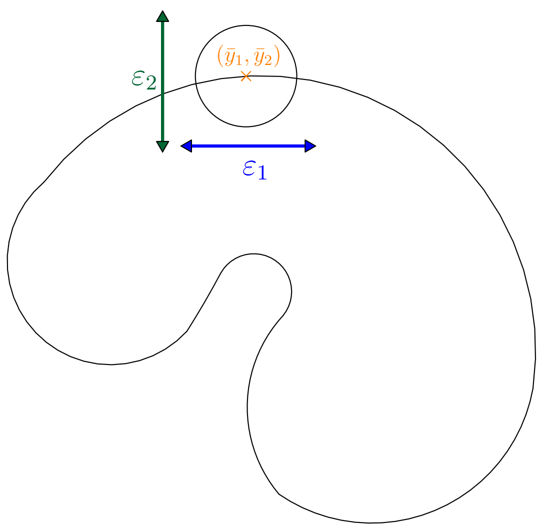
1. *The surface integrals of h with respect to f and g agree.*

2. *The surface integrals of ω with respect to f and g agree.*

We prove the second part of the theorem. Taking $k=1$, we have $\binom{d}{k}=d$, so $\omega\colon\mathbb{R}^d\to\mathbb{R}^d$. Observe that assuming the second statement and adjusting ω properly it implies the first one.
The full statement follows by decomposition of the submanifold into localized pieces and the following lemma.

Lemma 4.23. *For every regular point $\bar{x}\in U$ ($g(\bar{x})\neq g(y)$ for all $y\in\overline{U}\setminus U$) there exists $\varepsilon>0$ so that for each ω supported in $B(g(x),\varepsilon)$ the theorem holds.*

Proof. $Dg|_x$ is of rank k so there exists ρ with $\det(Dg|_x^{\rho})\neq 0$. By permuting the coordinates without loss of generality we can assume that $\det(Dg|_x^{(id)})\neq 0$. Let $g(\bar{x})=(\bar{y}_1,\bar{y}_2)$. The implicit function theorem implies that there exist ε_1, ε_2 such that for all $y_1\in B(\bar{y}_1,\varepsilon_1)$ there exists exactly one $x\in U$ and exactly one $y_2\in B(\bar{y}_2,\varepsilon_2)$ with $g(x)=(y_1,y_2)$.



Let ω be supported in $B(\bar{y}_1, \varepsilon_1) \times B(\bar{y}_2, \varepsilon_2)$ and by the above discussion we can write $V = g^{-1}(B(\bar{y}_1, \varepsilon_1) \times B(\bar{y}_2, \varepsilon_2))$. We compute

$$\int_V \sum_\rho \det Dg|_x^\rho \omega^\rho(g(x))dx$$
$$\overset{\text{var. change}}{=} \int_{B(\bar{y}_1,\varepsilon_1)} \frac{1}{|\det Dg|^{id}_{g^{-1}(y_1,y_2(y_1))}|} \sum_\rho \det Dg|^\rho_{g^{-1}(y_1,y_2(y_1))} \omega^\rho(y_1, y_2(y_1))dy_1.$$

On the other hand

$$\lim_{\eta\to 0} \int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^{d-k}} \phi_\eta(f(y_1,y_2)) \sum_\rho \varepsilon(\rho) \det Df|^{\tilde{\rho}}_{(y_1,y_2)} \omega^\rho(y_1,y_2)dy_2 \right] dy_1$$
$$= \int_{B(\bar{y}_1,\varepsilon_1)} \frac{1}{|\det Df|^{\tilde{id}}_{(y_1,y_2(y_1))}|} \sum_\rho \varepsilon(\rho) \det Df|^{\tilde{\rho}}_{(y_1,y_2)} \omega^\rho(y_1, y_2(y_1))dy_1,$$

similarly as in the last lecture. Two last computations show that we just need to prove that

$$\det Df|^{\tilde{\rho}}_{y_1,y_2(y_1)} \quad \text{and} \quad \det Dg|^\rho_{g^{-1}(y_1,y_2(y_1))}$$

(indexed by ρ and $\tilde{\rho}$) are parallel vectors. Here is a short argument why: first of all, inside both integrals we take the inner product with the same vector $\omega(y_1, y_2(y_1))^\rho$. Moreover, inside both integrals we divide by the respective determinat so the vectors in the last display get normalized having the same first component, at least up to a sign. Now, the signs agree choosing the same orientation of integration on the submanifolds.
Notice that $f \circ g(x) = 0$ for all x, so by the chain rule

$$Df|_{g(x)} \circ Dg|_x = 0$$

for all x. The 0 on the right hand side denotes the zero $k \times (d-k)$ matrix. This means that the vectors

$$v_i = D_i g|_x, \text{ for } i = 1, ..., k \ \text{ and } \ w_j = D_j g|_x, \text{ for } j = 1, ..., d-k$$

are orthogonal, in the sense that for all i, j, $\langle v_i, w_j\rangle = 0$. Hence

$$\det \begin{pmatrix} v_1 \\ \vdots \\ v_k \\ w_1 \\ \vdots \\ w_{d-k} \end{pmatrix} = \sqrt{\sum_{\substack{\rho\colon \underline{k}\to\underline{d} \\ \rho \text{ str. monotone}}} (\det v_{i,\rho(j)})^2} \times \sqrt{\sum_{\substack{\rho\colon \underline{k}\to\underline{d} \\ \rho \text{ str. monotone}}} (\det w_{i,\tilde{\rho}(j)})^2}.$$

However, using another definition of the determinant we also obtain

$$\det \begin{pmatrix} v_1 \\ \vdots \\ v_k \\ w_1 \\ \vdots \\ w_{d-k} \end{pmatrix} = \sum_{\sigma\colon \underline{d}\to\underline{d}} \varepsilon(\sigma) \prod_{i=1}^{k} v_{i\sigma(i)} \prod_{j=k+1}^{d} w_{j\sigma(j)}$$

$$= \sum_{\substack{\rho\colon \underline{k}\to\underline{d} \\ \rho \text{ str. monotone}}} \sum_{\tilde{\sigma}\colon \underline{k}\to\underline{k}} \sum_{\tilde{\tilde{\sigma}}\colon \underline{d-k}\to\underline{d-k}} \varepsilon(\rho)\varepsilon(\tilde{\sigma})\varepsilon(\tilde{\tilde{\sigma}}) \prod_{i=1}^{k} v_{i\rho(\tilde{\sigma}(i))} \prod_{j=k+1}^{d} w_{j\tilde{\rho}(\sigma(\tilde{\tilde{j}}))}$$

Notice that the right hand side equals

$$\sum_{\substack{\rho\colon \underline{k}\to\underline{d} \\ \rho \text{ str. monotone}}} \det(v_{i\rho(j)}) \det(w_{l\tilde{\rho}(m)}).$$

Put $a := \det(v_{i\rho(j)})$, $b := \det(w_{l\tilde{\rho}(m)})$. We just showed that

$$\langle a, b\rangle = \sqrt{\langle a, a\rangle}\sqrt{\langle b, b\rangle}$$

which means precisely that a and b are parallel (equality in the Cauch-Schwarz inequality occurs only in this case). $\square$

End of lecture 24. January 28, 2016

4.1 Surface area of the unit sphere and volume of the unit ball

The surface area of the unit sphere in $\mathbb{R}^d$

$$S^{d-1} = \{x \in \mathbb{R}^d : \sum_{i=1}^{d} x_i^2 - 1 = 0\}$$

is the surface integral of the constant function 1 over S^{d-1}. In terms of the delta notation introduced in the last lecture, it can be written as

$$\int_{\mathbb{R}^d} \delta\Big(\sum_{i=1}^{d} x_i^2 - 1\Big) 2\Big(\sum_{i=1}^{d} x_i^2\Big)^{1/2} dx.$$

Here we have computed $D_i f = 2x_i$ for $f(x) := \sum_{i=1}^d x_i^2 - 1$ and thus $(\sum_i |D_i f|^2)^{1/2} = 2(\sum_{i=1}^d x_i^2)^{1/2}$ (observe that since $k = d-1$, the matrix Df has only one column and the summation over ρ translates into the summation over $i = 1, \dots, d$). Since the integral is non-zero only if $\sum_{i=1}^d x_i^2 = 1$ (exercise), it equals

$$\int_{\mathbb{R}^d} \delta\big(\sum_{i=1}^d x_i^2 - 1\big) 2dx =: c_d.$$

Let us denote

$$I := \int_{\mathbb{R}} e^{-x^2} dx.$$

Then

$$\begin{aligned}
I^d &= \big(\int_{\mathbb{R}} e^{-x^2} dx\big)^d \\
&= \int_{\mathbb{R}^d} e^{-\sum_i x_i^2} dx \\
&\overset{(1)}{=} \int_{\mathbb{R}^d} \int_{\mathbb{R}} e^{-\sum_i x_i^2} \delta\big(\sum_i x_i^2 - r\big) dr dx \\
&= \int_{\mathbb{R}^d} \int_{\mathbb{R}} e^{-r} \delta\big(\sum_i x_i^2 - r\big) dr dx \\
&= \int_0^\infty e^{-r} \int_{\mathbb{R}^d} \delta\big(\sum_i x_i^2 - r\big) dx dr \\
&\overset{(2)}{=} \int_0^\infty e^{-r} \int_{\mathbb{R}^d} \delta\big(r \sum_i y_i^2 - r\big) r^{d/2} dy dr \\
&\overset{(3)}{=} \int_0^\infty e^{-r} r^{d/2-1} \delta\big(\sum_i y_i^2 - 1\big) dy dr \\
&= \frac{c_d}{2} \int_0^\infty e^{-r} r^{d/2-1} dr \\
&= \frac{c_d}{2} \Gamma(\frac{d}{2}),
\end{aligned}$$

where

$$\Gamma(t) = \int_0^\infty e^{-r} r^{t-1} dr$$

is the *Gamma function*.
In the computation we used the following rules of "δ-calculus", which need to be carefully justified using the definition of δ, i.e. by considering a sequence

of good kernels and passing to the limit (exercise). (1) : $\int \delta(c-r)dr = 1$ (+ Fubini) (2) : substitution $y = \sqrt{r}y$ (3) : $\delta(t) = r\delta(rt)$ for $r > 0, t \in \mathbb{R}$.
If $d = 2$ we know that $c_d = 2\pi$ since the sphere is then just a circle with radius 1 (this relation is also often used as the definition of π). Since $\Gamma(1) = 1$, we have $I = \sqrt{\pi}$ and therefore

$$c_d = \frac{2\sqrt{\pi}^d}{\Gamma(\frac{d}{2})}.$$

The volume of the unit ball in $\mathbb{R}^d$ is the integral

$$\begin{aligned}
\int_{|x|<1} dx &= \int_{|x|<1} \int_0^\infty \delta(|x|^2 - r) dr dx \\
&= \int_0^1 \int_{\mathbb{R}^d} \delta(|x| - r) dx dr \\
&= \int_0^1 \int_{\mathbb{R}^d} r^{d/2-1} \delta(|y|^2 - 1) dy dr \\
&= \frac{c_d}{2} \int_0^1 r^{d/2-1} dr \\
&= \frac{c_d}{2} \frac{2}{d} = \frac{c_d}{d}.
\end{aligned}$$

Note that we have first integrated over the spheres of radius $r < 1$ and then over all radii. This is exactly what one does when integrating with respect to polar coordinates. It is also possible to compute the surface area of the unit sphere and the volume of the unit ball via polar coordinates. However, the computation is more involved in this case.
The computation via delta calculus combined with the trick with the Gaussian function e^{-x^2} is very elegant. One could also try to compute c_d directly by definition of δ

$$c_d = \int_{\mathbb{R}^d} \delta(\sum_i x_i^2 - 1) 2dx = \int_{\mathbb{R}^d} \lim_{\varepsilon \to 0} \phi_\varepsilon(\sum_i x_i^2 - 1) 2dx,$$

without referring to the Gaussian integral, however, this is much less convenient.

4.2 Gauss' theorem

The *heaviside function* is given by

$$H(x) = \begin{cases} 1 : x < 0 \\ 0 : x \geq 0 \end{cases}$$

Note that $H' = -\delta$, which should be understood in the limiting sense, i.e. by considering a sequence of good kernels φ_ε and observing that $\phi_\varepsilon = -\int_{-\infty}^{x} \varphi_\varepsilon(t)dt + 1$ converges to H.[14]

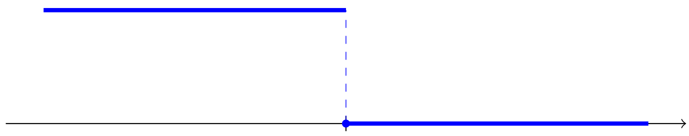


Figure 36: Heaviside function

Write $\sum_i x_i^2 = |x|^2$. Then, the volume of the unit ball can be expressed as

$$\begin{aligned}&\int_{\mathbb{R}^d} H\big(|x|^2 - 1\big)dx\\ =&\frac{1}{d}\sum_{i=1}^{d}\int_{\mathbb{R}^d} H\big(|x|^2-1\big)D_i x_i dx.\end{aligned}$$

To pass from the first to the second line we have inserted $\frac{1}{d}\sum_{i=1}^{d} D_i x_i = 1$. Integrating by parts we have (this step should be justified as described above...)

$$\frac{1}{d}\sum_{i=1}^{d}\int_{\mathbb{R}^d} H\big(|x|^2-1\big)D_i x_i dx \tag{17}$$

$$=\frac{1}{d}\sum_{i=1}^{d}\int_{\mathbb{R}^d} \delta\big(|x|^2-1\big)2x_i x_i dx = \frac{c_d}{d}. \tag{18}$$

Observe that (18) is the integral of a differential form over the sphere $|x| = 1$, with $f(x) = |x|^2 - 1$, $D_i f = 2x_i$ and x_i corresponds to $\varepsilon(\rho)\omega^\rho$. On the other hand, (17) is the integral of the derivative of a differential form over the unit ball. This should be compared with the fundamental theorem of calculus

$$f(b) - f(a) = \int_a^b f'(t)dt$$

where f is a function on an interval $[a, b]$. The left hand-side is the delta integral over the boundary of the interval, while the right hand side is the integral of f' over the interior of the interval.

[14] One should also recall the fact that the derivative of a monotone function is a Radon measure. In our case, the derivative of the monotonously decreasing function H is (the negative of) the Dirac measure.

Theorem 4.24. *Let $f : \mathbb{R}^d \to \mathbb{R}$ implicitly define a submanifold of $\mathbb{R}^d$ with $f(x) > 0$ for x large enough. Let $(F_i)_{i=1}^d$ be a C^1-vector field $\mathbb{R}^d \to \mathbb{R}^d$. Then*

$$\int_{\mathbb{R}^d} \delta(f(x)) \sum_{i=1}^{d} D_i f|_x F_i(x) dx = \int_{\mathbb{R}^d} H(f(x)) \sum_{i=1}^{d} D_i F_i(x) dx$$

The quantity $\sum_{i=1}^d D_i F_i$ is called the *divergence* of F, also denoted div$F$ or ∇F. This theorem is called the *Gauss' theorem* or the *divergence theorem.*
It follows by integration by parts in the language of delta calculus in the same way as in the special case $f(x) = |x|^2 - 1$, $F_i(x) = x_i$, which has been discussed above. The condition $f(x) > 0$ guarantees that $H(f(x)) = 0$ for large x and hence the boundary terms vanish. The term $D_i f$ is due to the chain rule when deriving $H(f(x))$.
Observe that just as in the special case above, the left hand-side is the integral of a differential form over the closed[15] surface given implicitly by $f = 0$, while the right hand-side is an integral of the derivative of the form over the volume enclosed by the surface.

Example (Graviational field of the earth)**.** For $x \neq 0$ define the vector field

$$F_i(x) = \frac{x_i}{|x|^d} = \frac{x_i}{(\sum_i x_i^2)^{d/2}}$$

We have

$$\text{div}F = \sum_{i=1}^{d} D_i \Big(\frac{x_i}{|x|^d} \Big) = \sum_{i=1}^{d} \frac{1}{|x|^d} - \frac{d}{2} \frac{2x_i^2}{(\sum_i x_i^2)^{d/2+1}} = 0$$

Let $f(x) = 2 - 3|x|^2 + |x|^4$. Note that it vanies iff $x \in [1, \sqrt{2}]$ and so

$$H(f(x)) = \begin{cases} 1 & : x \in [1, \sqrt{2}] \\ 0 & : \text{otherwise} \end{cases}$$

Let us check the Gauss' theorem on the region enclosed by the surface $f = 0$. Since div$F$=0, we have

$$\int_{\mathbb{R}^d} H(f(x)) \text{div} F dx = 0$$

Now we compute $D_i f = -6x_i + 4x_i \sum_i x_i^2$ On the boundary of the area in question this equals $-2x_i$, if $|x|^2 = 1$ and $2x_i$ if $|x|^2 = 2$. We compute the

[15]"Closed" means compact and witout boundary.

integral over the boundary on each of the pieces separately. Pick any point in $(1, \sqrt{2})$, say, $\frac{1+\sqrt{2}}{2}$. We have (exercise)

$$\int_{\{x>\frac{1+\sqrt{2}}{2}\}} \delta(f(x))\big(\sum_i |D_i f|^2\big)^{1/2} dx = c_d \sqrt{2}^{d-1}$$

and

$$\int_{\{x<\frac{1+\sqrt{2}}{2}\}} \delta(f(x))\big(\sum_i |D_i f|^2\big)^{1/2} dx = c_d$$

The left hand-side of the Gauss' theorem equals

$$\begin{aligned}
&\int_{\{x>\frac{1+\sqrt{2}}{2}\}} \delta(f(x)) \sum_i D_i f \frac{x_i}{|x|^d} dx \\
= &\int_{\{x>\frac{1+\sqrt{2}}{2}\}} \delta(f(x))\big(\sum_i |D_i f|^2\big)^{1/2} \frac{1}{|x|^{d-1}} dx = c_d \sqrt{2}^{d-1} \frac{1}{\sqrt{2}^{d-1}} = c_d
\end{aligned}$$

For the first equality we used that the vectors $(D_i f)_i$ and $(x_i)_i$ are parallel and pointing in the same direction, hence their scalar product is the product of their lengths. Similarly we compute

$$\begin{aligned}
&\int_{\{x<\frac{1+\sqrt{2}}{2}\}} \delta(f(x)) \sum_i D_i f \frac{x_i}{|x|^d} dx \\
= &\int_{\{x>\frac{1+\sqrt{2}}{2}\}} \delta(f(x))\big(\sum_i |D_i f|^2\big)^{1/2} \frac{-1}{|x|^{d-1}} dx = -c_d.
\end{aligned}$$

Thus, the left hand-side of the Gauss' theorem equals to zero and coincides with the right hand side.

If $d = 3$, the vector field

$$F_i(x) = \begin{cases} x_i & : |x| < 1 \\ \frac{x_i}{|x|^d} & : |x| \geq 1 \end{cases}$$

models the gravitational field of the Earth (up to a sign). Its maximum is on earth's surface and it decreases with altitude as one rises above the Earth's surface. Note that

$$\operatorname{div} F = \begin{cases} d & : |x| < 1 \\ 0 & : |x| \geq 1 \end{cases}$$

Gauss' law for gravity states that

$$\mathrm{div}F = \gamma\rho$$

where γ is a certain constant and ρ mass density at each point. By the Gauss' theorem, this can also be written in the form

$$\int_{\mathbb{R}^3} H(f(x))\gamma\rho dx = \int_{\mathbb{R}^3} \delta(f(x)) D_i f F_i dx \tag{19}$$

where $f(x) = 0$ determines a closed surface in $\mathbb{R}^3$. The left hand-side is is up to a constant $\int_{\{f(x)<0\}} \rho$ which is the total mass enclosed within the surface. Therefore, Gauss' law says that the gravitational flux through any closed surface is proportional to the enclosed mass, where the *gravitational flux* is a surface integral of the gravitational field over the surface (i.e. the right hand side of (19)). In this example we calculated the flux over the region enclosed by $|x|^2 = 1$ and $|x|^2 = 2$, which turned out to be zero. Note that no mass was enclosed.

See `https://en.wikipedia.org/wiki/Gauss's_law_for_gravity` for more details. There is also an analogous result for the electric field.

4.3 Stokes' theorem

In the last chapter we related an integral of a differential form over a closed surface to the integral of its derivative over the volume inside the surface. Suppose we want to integrate over the upper hemisphere rather than over the whole sphere. Analogously to the previous section, we would like to relate the integral over the upper hemisphere to an integral over its boundary (which is the equator).

We first consider $d = 3$. Let $F : \mathbb{R}^3 \to \mathbb{R}^3$ be a vector field. Denote by $\mathrm{curl}(F)$ the vector[16]

$$\mathrm{curl}(F) = -(D_2F_3 - D_3F_2, D_3F_1 - D_1F_3, D_1F_2 - D_2F_1)$$

Let $h(x) = -x_3$ so that $H(h(x)) = 1$ if $x_3 > 0$ and zero otherwise, i.e. we are on the upper hemisphere. Integration by parts (in terms of the delta calculus) yields

$$\int_{\mathbb{R}^3} \delta(f(x)) H(h(x)) \sum_{i=1}^{3} D_i f \,(\mathrm{curl}(F))_i dx$$

[16] Curl is usually defined without the minus sign in front of the bracket, but this definition will be more convenient for us.

$$= \int_{\mathbb{R}^3} \delta(f(x))\delta(h(x)) \sum_{i=1}^{3} \big(D_{i+1} f D_{i-1} h - D_{i-1} f D_{i+1} h\big) F_i dx$$

Observe that the integral on the left hand-side is over the hemisphere, while the one on the right hand-side is over its boundary. These integrals can be expressed in the language of differential forms similarly as in the previous section. The derivatives of h are due to the chain rule. This identity is a special case of the *Stokes' theorem.*

End of lecture 25. February 2, 2016

In this lecture we discuss Stokes' theorem in full generality. First we recall the Laplace expansion for the determinant of a matrix. Let $k \le n$ and let A be a $k \times n$ matrix, B an $n - k \times n$ matrix. Denote by (A, B) the $n \times n$ matrix which is obtained by adding the rows of B to the matrix A. That is, $(A, B)_{ij}$ equals A_{ij} for $i \le k$ and $B_{(i-k)j}$ for $i > k$. Then

$$\det(A, B) = \sum_{\substack{\rho:(1,\dots,k)\to(1,\dots,n) \\ \text{mon., inj.}}} \varepsilon(\rho) \det A^{\rho} \det B^{\tilde{\rho}}$$

where A^{ρ} is a $k \times k$ submatrix of A whose columns are determined by ρ (i.e. the i-th column of A^{ρ} is the $\rho(i)$-th column of A), $\tilde{\rho} : (1, \dots, n - k) \to (1, \dots, n)$ is a monotone injective map with $im(\rho) \cap im(\tilde{\rho})$ (i.e. the "complement" of ρ, note that it is determined by ρ) and $\varepsilon(\rho) := \varepsilon(\sigma)$ where $\sigma : (1, \dots, n) \to (1, \dots, n)$ is such that $\sigma|_{(1,\dots,k)} = \rho$ and $\sigma|_{(k+1,\dots,n)}(i) = \tilde{\rho}(i - k)$.

Let everything be as above and let h be a $1 \times n$ matrix (i.e. "a row") and $\rho' : (1, \dots, k + 1) \to (1, \dots, n)$ a monotone injective map. By the Laplace expansion applied to $(A, h)^{\rho'}$ we have

$$\det(A, h)^{\rho'} = \sum_{\substack{\tau:(1,\dots,k)\to(1,\dots,k+1) \\ \text{mon., inj.}}} \varepsilon(\tau) \det A^{\rho'\circ\tau} h^{\rho'\circ\tilde{\tau}(1)}$$

where $\tilde{\tau} : (1) \to (1, \dots, n)$ is related to τ in the same way as $\tilde{\rho}$ to ρ.

4.3.1 Exterior derivative.

Let ω^{ρ} be a k-differential form in $\mathbb{R}^n$ (an $\binom{n}{k}$ map on $\mathbb{R}^n$). Define the $k + 1$ form

$$(d\omega)^{\rho'} = \sum_{\tau:(1,\dots,k)\to(1,\dots,k+1)} \varepsilon(\tau) D_{\rho'\circ\tilde{\tau}(1)} \omega^{\rho'\circ\tau}$$

where ρ', τ are as above. The form $d\omega$ is called the *exterior derivative of* ω. If ω^ρ is the determinant of a $k \times k$ submatrix A^ρ of a $k \times n$ matrix A (as it was the case in Lecture 22), the exterior derivative can be formally seen as adding the row of partial differential operators $h = (D_1, D_2, \dots, D_n)$ to A and using the Laplace expansion on (A, h).

Note that $dd\omega = 0$ since

$$(dd\omega)^{\rho''} = \sum_{\substack{\vartheta:(1,\dots,k+1)\to(1,\dots,k+2)\\ \text{mon. inj.}}} \sum_{\substack{\tau:(1,\dots,k)\to(1,\dots,k+1)\\ \text{mon. inj.}}} \varepsilon(\vartheta)\varepsilon(\tau) D_{\rho''\circ\tilde{\vartheta}(1)} D_{\rho''\circ\vartheta\circ\tilde{\tau}(1)} \omega^{\rho''\circ\vartheta\circ\tau} = 0$$

The details are left as an exercise, one has to use $D_i D_j = D_j D_i$ and that the corresponding terms appear with a different sign.

Stokes' theorem.

Let $f : \mathbb{R}^n \to \mathbb{R}^{n-k-1}$ and $h : \mathbb{R}^n \to \mathbb{R}$ be smooth maps such that the rank of $D(f,h)(x)$ is maximal whenever $(f,h)(x) = 0$. Let Ω be a surface implicitly defined by $f(x) = 0$, $h(x) < 0$ and let $\partial\Omega$ be its boundary (defined by $(f,h)(x) = 0$). Let ω be a k-form. Then

$$\int_\Omega d\omega = \int_{\partial\Omega} \omega$$

where $d\omega$ is the exterior derivative of ω.

Note that Gauss' theorem and the example discussed at the end of the previous lecture are just a special cases of Stokes' theorem.

Proof. (Sketch) If δ and H are defined as in the previous lecture (Dirac delta and the Heaviside function), the left hand-side can be written as

$$\int_{\mathbb{R}^n} \delta(f)H(h) \sum_{\rho':(1,\dots,k+1)\to(1,\dots,n)} \varepsilon(\rho') \det Df|_x^{\tilde{\rho}'} (d\omega)^{\rho'} dx$$

$$\stackrel{(1)}{=} \int_{\mathbb{R}^n} \delta(f)H(h) \sum_{\rho':(1,\dots,k+1)\to(1,\dots,n)} \varepsilon(\rho') \det Df|_x^{\tilde{\rho}'} \sum_{\tau:(1,\dots,k)\to(1,\dots,k+1)} \varepsilon(\tau) D_{\rho'\circ\tilde{\tau}(1)} \omega^{\rho'\circ\tau} dx$$

$$\stackrel{(2)}{=} \int_{\mathbb{R}^n} \delta(f)\delta(h) \sum_{\rho'} \sum_{\tau} \varepsilon(\rho')\varepsilon(\tau) D_{\rho'\circ\tilde{\tau}(1)} h \det Df|_x^{\tilde{\rho}'} \omega^{\rho'\circ\tau} dx$$

$$\stackrel{(3)}{=} \int_{\mathbb{R}^n} \delta(f,h) \sum_{\rho:(1,\dots,k)\to(1,\dots,n)} \sum_{\sigma:(1,\dots,n-k-1)\to(1,\dots,n-k)} \varepsilon(\rho)\varepsilon(\sigma) D_{\tilde{\rho}\circ\tilde{\sigma}(1)} h \det Df|_x^{\tilde{\rho}\circ\sigma} \omega^\rho dx$$

$$\overset{(4)}{=} \int_{\mathbb{R}^n} \delta(f,h) \sum_{\rho} \varepsilon(\rho) \det D(f,h)|_x^{\tilde{\rho}} \omega^\rho dx$$

We argue as follows.
(1) : The definition of the exterior derivative.
(2) : Partial integration in the language of delta calculus, which is left as an exercise. One derives by the Leibnitz rule and notices that the only non-zero term is when the derivative falls on $H(h)$, which gives $\delta(h)$ (and the derivatives of h). A similar argument used to show $dd\omega = 0$ yields that differentiating $\det Df|_x^{\tilde{\rho}'}$ yields a zero term.
(3) : Define $\rho : (1, \ldots, k) \to (1, \ldots, n)$ by $\rho = \rho' \circ \tau$ and reparametrize the "complement" of ρ', τ by σ. Exercise: check that the signs match.
(4) : Laplace expansion. □

Let us discuss $\mathbb{R}^3$ more thoroughly. If $n = 3$, then 0-forms map into $\mathbb{R}^{\binom{3}{0}} = \mathbb{R}^1$, 1-forms into $\mathbb{R}^{\binom{3}{1}} = \mathbb{R}^3$, 2-forms into $\mathbb{R}^{\binom{3}{2}} = \mathbb{R}^3$ and 4-forms into $\mathbb{R}^{\binom{3}{3}} = \mathbb{R}^1$. Observe that all target spaces have dimension either 1 or $3(= n)$ and that 1- and 2- forms are vector fields. Only in three dimensions k- and $k+1$-forms can be vector fields. We compute the exterior derivative of a k-form for $k = 0, 1, 2$.
<u>$0 \to 1$.</u> Let ω be a 0-form. Then for $i = 1, 2, 3$

$$(d\omega)^i = \sum_{\tau:\emptyset\to(1)} \varepsilon(\tau) D_{\rho\circ\tilde{\tau}(1)} \omega^{\rho\circ\tau} = D_i\omega,$$

so $d\omega$ is the gradient of ω.
<u>$1 \to 2$.</u> If $\omega = (\omega^1, \omega^2, \omega^3)$ is a 1-form, then

$$(d\omega)^{ij} = \sum_{\tau:(1)\to(1,2)} \varepsilon(\tau) D_{\rho\circ\tilde{\tau}(1)} \omega^{\rho\circ\tau} = D_j\omega_i - D_i\omega_j$$

Note that $(\omega^{23}, -\omega^{13}, \omega^{12}) = \mathrm{curl}(\omega)$ and that "d" maps vector fields to vector fields. If we identify 2-forms with vector fields $F = (F_1, F_2, F_3)$ as

$$(\omega^{12}, \omega^{13}, \omega^{23}) \leftrightarrow (F_3, -F_2, F_1),$$

then $d\omega = \mathrm{curl}F$. Observe that if ω is a 0-form, by $dd\omega = 0$ we have

$$\mathrm{curl}\,\mathrm{grad}\,\omega = 0.$$

<u>$2 \to 3$.</u> If $(\omega^{12}, \omega^{13}, \omega^{23})$ is a 2-form, then

$$(d\omega)^{123} = D_3\omega^{12} - D_2\omega^{13} + D_1\omega^{23}$$

which is, using the above identification, the divergence of F (recall $\mathrm{div}F = D_1F_1 + D_2F_2 + D_3F_3$). We have

$$\mathrm{div}\,\mathrm{curl}\,F = 0.$$

4.3.2 Maxwell's equations.

Maxwell's equations describe how electric and magnetic fields are generated and altered by each other and by charges and currents.

- In the last lecture we discussed Gauss' law for gravity. There is an analogous result for electric fields. Denote by ρ the electric charge density and E the electric field. Then

  $$\varepsilon_0 \mathrm{div} E = \rho$$

  where ε_0 is a certain constant which can be determined experimentally. By Stoke's theorem, this is equivalent to

  $$\varepsilon_0 \int_{\partial\Omega} E = \int_{\Omega} \rho$$

  The electric field leaving a volume is proportional to the charge inside.

- Gauss' law for magnetism:

  $$\mathrm{div} B = 0$$

  or, equivalently,

  $$\int_{\partial\Omega} B = 0$$

  (by Stokes). Here B stands for the magnetic field. This means that there are no magnetic monopoles.

- Maxwell-Faraday equation (law for induction):

  $$\mathrm{curl} E = -\frac{\partial B}{\partial t}$$

  or, equivalently,

  $$\int_{\partial\Omega} E = -\frac{\partial}{\partial t} \int_{\Omega} B$$

  The voltage accumulated around a closed circuit is proportional to the time rate of change of the magnetic flux it encloses.

- Ampere's circuital law:

$$\mathrm{curl} B = \mu_0 j + \varepsilon_0 \mu_0 \frac{\partial E}{\partial t}$$

  where μ_0 is a certain constant and j the electric current density. We omit writing the equivalent integral form which is derived using Stokes' theorem. This law says that electric currents and changes in electric fields are proportional to the magnetic field circulating about the area they pierce.

Electromagnetic wave equation in a vacuum. In a vacuum we may assume the above equations read

$$\mathrm{div} E = 0, \quad \mathrm{div} B = 0, \quad \mathrm{curl} E = -\frac{\partial B}{\partial t}, \quad \mathrm{curl} B = \varepsilon_0 \mu_0 \frac{\partial E}{\partial t}$$

Then

$$\begin{aligned}
\mathrm{curl}\,\mathrm{curl} E &= -\,\mathrm{curl} \frac{\partial B}{\partial t} \\
&= -\,\frac{\partial}{\partial t}(\mathrm{curl} B) \\
&= -\,\frac{\partial}{\partial t}(\varepsilon_0 \mu_0 \frac{\partial E}{\partial t}) \\
&= -\,\varepsilon_0 \mu_0 \frac{\partial^2 E}{\partial t^2}
\end{aligned}$$

We also compute

$$\begin{aligned}
(\mathrm{curl}\,\mathrm{curl} E)_1 &= -\,D_2^2 E_1 - D_3^2 E_1 + D_2 D_1 E_2 + D_3 D_1 E_3 \\
&= -\,\sum_{i=1}^{3} D_i^2 E_1 + D_1(\mathrm{div} E) \\
&= -\,\Delta E_1
\end{aligned}$$

where we have denoted $\Delta = \sum_{i=1}^{3} D_i^2$ (the Laplace operator) and used that $D_1(\mathrm{div} E) = 0$ by our assumption. Performing analogous calculation for the other two components of $\mathrm{curl}\,\mathrm{curl} E$ we obtain

$$-\mathrm{curl}\,\mathrm{curl} E = (\Delta E_1, \Delta E_2, \Delta E_3)$$

and hence by the above calculation

$$\Delta E - \mu_0 \varepsilon_0 \frac{\partial^2 E}{\partial t^2} = 0.$$

This is the so-called *wave equation.* (To gain more insight, in one dimension it takes the form

$$\partial_x^2 f - \frac{1}{c^2}\partial_t^2 f = 0$$

where $c = \frac{1}{\sqrt{\mu_0\varepsilon_0}}$. Note that $f(x-ct)$ and $f(x+ct)$ solve the equation. These are travelling waves, propagating with the speed c (the propagation direction depends on the sign). It turns out that any other solution is a linear combination of these two.) The constant $\mu_0\varepsilon_0$ has been determined experimentally and it turns out to be the square of the reciprocal of the speed of light, i.e.

$$\sqrt{\mu_0\varepsilon_0} \approx \frac{1}{3\cdot 10^8 \frac{m}{s^2}}.$$

The electromagnetic waves propagate with the speed of light. See `https://en.wikipedia.org/wiki/A_Dynamical_Theory_of_the_Electromagnetic_Field` for more on this topic.

End of lecture 26. February 9, 2016

5 Appendix

5.1 Riesz representation theorem

This is an appendix to Lecture 15.
Let us define

$$L^p(\mathbb{R}^d) = \{f - g\colon f, g \in L^p_+(\mathbb{R}^d)\}.$$

Let us also define the set of nonnegative linear functionals defined on $L^p_+(\mathbb{R}^d)$

$$D^p_+(\mathbb{R}^d) = \{\Lambda\colon L^p_+(\mathbb{R}^d) \to [0,\infty),\ \Lambda(f+g) = \Lambda(f)+\Lambda(g)\ \text{ for } f, g \in L^p_+(\mathbb{R}^d)\}$$

and the set of linear functionals on $L^p_+(\mathbb{R}^d)$

$$D^p(\mathbb{R}^d) = \{\Lambda\colon L^p_+(\mathbb{R}^d) \to \mathbb{R},\ \text{there exist } \Lambda_1, \Lambda_2 \in D^p_+(\mathbb{R}^d) \text{ with } \Lambda = \Lambda_1 - \Lambda_2\}.$$

Recall that we proved the following.

Theorem 5.1 (Riesz representation theorem for nonnegative functionals)**.** *Let $1 < p < \infty$ and let $\Lambda \in D^p_+(\mathbb{R}^d)$. Then there exists exactly one $h \in L^{p'}_+(\mathbb{R}^d)$ with $\Lambda(f) = \int f \cdot h d\mu$ for all $f \in L^p_+(\mathbb{R}^d)$, where $1/p + 1/p' = 1$.*

The goal of this subsection is prove the counterpart of the above theorem for the set of all functionals (not necessarily nonnegative) $D^p(\mathbb{R}^d)$.

Theorem 5.2 (Riesz representation theorem)**.** *Let $1 < p < \infty$ and let $\Lambda \in D^p(\mathbb{R}^d)$. Then there exists exactly one $h \in L^{p'}(\mathbb{R}^d)$ with $\Lambda(f) = \int f \cdot h d\mu$ for all $f \in L^p_+(\mathbb{R}^d)$, where $1/p + 1/p' = 1$.*

Proof. Let $\Lambda \in D^p(\mathbb{R}^d)$. There exist $\Lambda_1, \Lambda_2 \in D^p_+(\mathbb{R}^d)$ with $\Lambda = \Lambda_1 - \Lambda_2$. By the Riesz theorem for nonnegative functionals we obtain $h_1, h_2 \in L^{p'}(\mathbb{R}^d)$ such that

$$\Lambda_i(f) = \int f \cdot h_i d\mu$$

for $i = 1, 2$ and all $f \in L^p_+(\mathbb{R}^d)$. We conclude putting $h = h_1 - h_2$. □

Now suppose that we have a linear functional $\Lambda \in D^p(\mathbb{R}^d)$ that is additionally *bounded*, i.e. satisfies $\Lambda(f) \le C\|f\|_p$ for a constant C and all $f \in L^p_+(\mathbb{R}^d)$. Does there also exists a unique $h \in L^{p'}(\mathbb{R}^d)$ such that Λ is given by integrating against h? Can we possibly "extract" nonnegative Λ_1, Λ_2 with $\Lambda = \Lambda_1 - \Lambda_2$, so it all boils down to the previous theorem? The following proposition gives a positive answer to these questions.

Proposition 5.3. *$D^p(\mathbb{R}^d)$ is equal to the set*

$$\{\Lambda\colon L^p_+(\mathbb{R}^d) \to \mathbb{R},\ \exists C \text{ with } \Lambda(f) \le C\|f\|_p,\ \Lambda(f+g) = \Lambda(f) + \Lambda(g)\}.$$

Proof. ($\subset$) Let $\Lambda \in D^p(\mathbb{R}^d)$. There exist nonnegative functionals Λ_1, Λ_2 with $\Lambda = \Lambda_1 - \Lambda_2$. Hence by Hölder's inequality

$$|\Lambda(f)| \le \Lambda_1(f) + \Lambda_2(f) = \int f h_1 d\mu + \int f h_2 d\mu \le (\|h_1\|_{p'} + \|h_2\|_{p'})\|f\|_p$$

This proves the first inclusion with $C = \|h_1\|_{p'} + \|h_2\|_{p'}$.
($\supset$) Define Λ_1 as follows

$$\Lambda_1(f) = \sup_{\substack{0 \le \varphi \le f \\ \varphi \in L^p_+(\mathbb{R}^d)}} \Lambda(\varphi).$$

Note that $\Lambda(0) = 0$, so $\Lambda_1(f) \ge 0$. Moreover for $0 \le \varphi \le f$

$$\Lambda(\varphi) \le C\|\varphi\|_p \le C\|f\|_p < \infty.$$

The penultimate bound on the right hand side is uniform in φ, hence taking the supremum we obtain $\Lambda_1(f) < \infty$. Observe that if we show that Λ_1 is additive, then we are done, because

$$\Lambda_2(f) = \Lambda_1(f) - \Lambda(f),$$

defines a functional in $D^p_+(\mathbb{R}^d)$, since $\Lambda_1(f) \geq \Lambda(f)$ and both Λ_1, Λ are additive. Let us then prove that Λ_1 is additive.
($\geq$)

$$\Lambda_1(f+g) = \sup_{0\leq\varphi\leq f+g} \Lambda(\varphi) \geq \sup_{0\leq\varphi_1\leq f} \sup_{0\leq\varphi_2\leq g} \Lambda(\varphi_1+\varphi_2).$$

By additivity this is equal to

$$\sup_{0\leq\varphi\leq f} \Lambda(\varphi_1) + \sup_{0\leq\varphi\leq g} \Lambda(\varphi_2) = \Lambda_1(f) + \Lambda_1(g).$$

($\leq$) Let $0 \leq \varphi \leq f+g$. Define $\varphi_1 = \min(\varphi, f)$, $\varphi_2 = \varphi - \varphi_1$. We have

$$\Lambda(\varphi) = \Lambda(\varphi_1+\varphi_2) \leq \sup_{0\leq\tilde{\varphi}_1\leq f} \Lambda(\tilde{\varphi}_1) + \sup_{0\leq\tilde{\varphi}_2\leq g} \Lambda(\tilde{\varphi}_2) \leq \Lambda_1(f) + \Lambda_2(g).$$

We finish the proof taking the supremum on the left hand side. □

End of lecture 27. February 9, 2016