ANALYSIS IV FOR PHYSICS

JUHAN ARU

Section 0

Introduction and motivation

As motivation, let us consider the mathematical description of heat transmission on a homogeneous circular rod: the heat equation.

The heat equation on an interval $[0, 1)$ (describing the rod) is given by describing the evolution of the temperature profile

$$\frac{\partial u(t, x)}{\partial t} = D\Delta u(t, x)$$

together with some initial condition $u_0(x) = u(0, x)$ and the boundary condition $u(t, 0) = u(t, 1)$ for all $t \geq 0$ to express that the ends of the rod are connected. Recall that in the 1D case $\Delta f := \frac{\partial^2 f}{\partial x^2}$ and $D > 0$ is the diffusion coefficient.

The revolutionary idea of Fourier was as follows. He noticed empirically that the heat profile over time shows spatially oscillatory behaviour, and thus also motivated by the solution of the wave equation using waves, he proposed to write any solution using spatially oscillating functions like $f_n(x) = \sin(2\pi x)$ and $g_n(x) = \cos(2\pi nx)$. More precisely, one could try to find a solution of the form

$$u(t, x) = \sum_{n\geq 1} s_n(t)\sin(2\pi nx) + \sum_{n\geq 0} c_n(t)\cos(2\pi nx).$$

But now notice that $\Delta f_n = -4D\pi^2 n^2 f_n$ and thus if we try a solution of the form $u_n(x, t) = f_n(x)s_n(t)$ with f_n as above, we obtain an equation

$$\frac{\partial s_n(t)}{\partial t} = -4D\pi^2 n^2 s_n(t).$$

This is a well-known ODE that is easily solved: $s_n(t) = \exp(-4D\pi^2 n^2 t)s_n(0)$. Similarly for the cos terms we get $c_n(t) = \exp(-4D\pi^2 n^2 t)c_n(0)$.

We conclude that it would make sense to propose a solution of the form

$$u(t, x) = \sum_{n\geq 1} s_n(0)\exp(-D4\pi n^2 t)\sin(2\pi nx) + \sum_{n\geq 0} c_n(0)\exp(-D4\pi n^2 t)\cos(2\pi nx).$$

Notice that the initial condition then translates to the condition:

$$u_0 = \sum_{n\geq 1} s_n(0)\sin(2\pi nx) + \sum_{n\geq 0} c_n(0)\cos(2\pi nx).$$

If we do find such $(s_n(0), c_n(0))_{n\geq 0}$, then we may have found at least one solution to the heat equation on the circular rod.

Now, this may sound very convincing, but on a closer look there are several questions here:

1. (1) We have infinite sums - do they even converge? When do they converge and in which sense?
2. (2) For which functions u_0 does the above-given expansion hold? In other words for which initial conditions can we find a solution by this method?
3. (3) Are such expansions unique? Are the solutions we find unique?

(4) Can one approximate solutions? For example this is relevant when trying to numerically solve the equation. This is a question about convergence - and further, how does the notion of convergence relate to the coefficients s_n, c_n?
(5) More generally, how should one measure closesness of different initial conditions, different solutions?
(6) What happens for non-circular rods, e.g. rods with endpoints in heat-baths? Or in higher dimensions?
(7) What about more non-homogeneous case where D is no longer a constant in space? Or when we replace Δ with more general (linear) operators, including for example also some outside influences?

The aim of this course is to study the right mathematical framework to ask and answer such questions. This will bring us to study function spaces, the Lebesgue integral and spectral theory of linear operators. To see why some of those aspects might enter let us further consider a simplified model.

0.1 A discrete model

To understand what we may hope to achieve, let us consider the same problem of heat diffusion but on a discretised space. For example we think that the rod instead is decomposed of n small containers which can exchange heat between its neighbours.

The temperature profile is now given by $u(x,t) : \{0, 1, \dots, n\} \to \mathbb{R}$, with the periodicity condition $u(0,t) = u(n,t)$ for all $t \geq 0$.

The evolution is still given by

$$\frac{\partial u(t,x)}{\partial t} = K\Delta_d u(t,x)$$

together with some initial condition $u_0(x) = u(0,x)$, only instead of the real Laplacian, we have the discrete Laplacian $\Delta_d f(x) := \frac{1}{d_x}\sum_{y\sim x} f(y) - f(x)$, where $y \sim x$ means that y, x are neighbours in the underlying discrete graph and d_x is the number of neighbours of the vertex x. In our concrete case we have a circular graph with n vertices and thus $\Delta_d f(x) := \frac{f(y)+f(z)-2f(x)}{2}$, where y, z denote the neighbouring vertices.

Now notice that now the problem is really a system of n coupled ordinary differential equations of second degree and Δ_d is just a linear operator on $\mathbb{R}^n \to \mathbb{R}^n$. So how do we solve it?

Let use the same steps as above but see that they have a very simple and concrete meaning here:

- Notice that each u_t can be seen as a vector in $\mathbb{R}^n$ with coordinates and Δ_d can be seen as a symmetric linear operator on $\mathbb{R}^n$(check it!)
- As such Δ_d can be diagonalized: there is an orthonormal basis $\phi_1, \dots, \phi_n$ and eigenvalues $\lambda_1, \dots, \lambda_n$ such that $\Delta_d \phi_i = \lambda_i \phi_i$. In particular any function $u : \mathbb{R}^n \to \mathbb{R}^n$ can be uniqueley written as $\sum_{i=1}^n c_i \phi_i$.
- But now if we write $u_i(t) := c_i(t)\phi_i$, then again each $c_i(t)$ satisfies now a decoupled ODE
$$\frac{\partial c_i(t)}{\partial t} = K\lambda_i s_n(t)$$
and thus has a solution $c_i(0)\exp(K\lambda_i t)$.

- We conclude a solution by finding $c_i(0)$ by determining the unique expansion $u_0 := \sum_{i=1}^n c_i(0)\phi_i$.
- Given the uniqueness of the expansion, this solution is also unique.
- And finally, we can easily compare solutions just using for example the Euclidean norm. For example conclude that if the initial conditions are close, then so will be the solutions at all times $t > 0$. We also know that this distance is equivalently measured using the distances between two sets of coefficients $(c_i)_{i=1\dots n}, (\tilde{c}_i)_{i=1\dots n}$ - and here using the Euclidean norm instead of some other norm is important.

Hence in this set-up all works super well and would work equally well as long as we have a symmetric linear operator L instead of Δ_d.

What did we use here?

- We used the fact that $\mathbb{R}^n$ is finite-dimensional and thus there exist basis that give unique expansions for each vector
- We used the fact that Δ_d is linear and symmetric and by the spectral theorem can be diagonalised and we can find a basis of eigenvectors
- We used implicitly the linearity of the equation

None of these facts are clear in our original set-up as the space of functions from $[0, 1]$ to $\mathbb{R}$ is no longer finite-dimensional!

To address those we will have to look at spaces of functions and try to first see which such spaces have a nice structure. For example, which spaces of functions satisfy linearity? Which can be define a norm and talk about orthonormality? For which spaces do we have orthonormal expansions? Looking for such nice properties brings us for example to also introduce the Lebesgue integral to construct nice basis of functions.

After that, having spent some time understanding function spaces, we briefly at the study of linear operators on such spaces and in particular find some set-ups where there are similar orthonormal decompositions using eigenfunctions. We then put all this together to rigorously explain solving the inhomogeneous heat equation and other similar problems.

But this is already enough of introduction, let us get going!

Section 1

The space of continuous functions

Let us start with maybe the most intuitive of function spaces - the space of continuous functions. This is partly a recap, as you have been working with continuous functions in Analysis I-III, and we are just putting things in a wider context.

To start off the functions will be taking values on closed boxes $D \subseteq \mathbb{R}^n$, i.e. rectangles $[a_1, b_1] \times \cdots \times [a_n, b_n]$) and taking values in $\mathbb{R}$. At the end of the section we will discuss to what extent we can (and may want to) generalize both of these choices. You may safely just suppose $D = [0, 1]$, as no actual extra difficulty comes from going to higher dimensions.

The set of all continuous functions from $D \to R$ will be denoted by $C(D, \mathbb{R})$:

$$C(D, \mathbb{R}) := \{f : D \to R, f \text{ continuous}\}.$$

In what follows we will try to understand the structure of this space.

1.1 Vector space structure of $C(D, \mathbb{R})$

The first observation we can make about the space $C(D, R)$ is that it has a linear structure like for example the vector space $(\mathbb{R}^n, +)$: if $f, g \in C(D, \mathbb{R})$, then also the function $h(x) := f(x) + g(x)$ is in $C(D, \mathbb{R})$, as is $\lambda f(x)$ where $\lambda \in \mathbb{R}$.

Let us quickly check this for the first statement: for every $x \in D$, by continuity of f, g we can choose δ_f, δ_g such that if $y \in D, \|x - y\| < \delta_f$ then $|f(x) - f(y)| \leq \frac{\epsilon}{2}$ and if $y \in D, \|x - y\| < \delta_g$, then $|g(x) - g(y)| \leq \frac{\epsilon}{2}$. But this means that if $\|x - y\| < \min(\delta_f, \delta_g)$, we have that $|h(x) - h(y)| < \epsilon$ by the triangle inequality.

Exercise 1.1. *Show that in fact $C(D, \mathbb{R})$ has also multiplicative structure: i.e. if $f, g \in C(D, \mathbb{R})$, then also the product $h(x) := f(x)g(x)$ is in $C(D, \mathbb{R})$. What about the function $\max(f, g)$?*

In fact, the space $C(D, \mathbb{R})$ with addition satisfies all the axioms of a vector space! Indeed, the identity element would be just the constant zero function, the inverse element of f the function $-f$ and all conditions are nicely met, as you can easily and patiently check.

Exercise 1.2. *Recall the axioms of a vector space and verify them in the case of $(C(D, \mathbb{R}), +)$.*

In what follows we will often also call the vector space just $C(D, \mathbb{R})$.

Now we might be also interested in summing infinitely many functions, i.e. looking at sums $\sum_{n \geq 1} f_n$. But in what sense can we talk about it? More generally, given a sequence of $(g_n)_{n \geq 1}$ in which sense can we talk about its convergence and limit?

The first idea might be to define limits pointwise: for each $x \in D$ the sequence $(g_n(x))_{n \geq 1}$ is just a sequence of real numbers and thus we know what its convergence means. Thus we may want to define the convergence of $(g_n)_{n \geq 1}$ as functions to mean the convergence of $(g_n(x))_{n \geq 1}$ for all $x \in D$. This is called pointwise convergence and as you have already seen it suffers a small drawback:

Exercise 1.3. *For D a closed box in $\mathbb{R}^n$, find a sequence of functions in $C(D, \mathbb{R})$ that converges pointwise to a function that is not continuous.*

It is a good idea to start from the case $D = [0, 1]$ (which we discussed in class), but then think how to do it in general.

We will look for other notions of convergence and to do this will introduce a norm on the set $C(D, \mathbb{R})$.

1.2 The uniform norm on $C(D, \mathbb{R})$

Recall that the vector space $\mathbb{R}^n$ comes also with several natural norms that give a notion of length of a vector and give us a way to measure distances between vectors. It comes out that one can also endow $C(D, \mathbb{R})$ with a natural norm.

Definition 1.1 (The supremum (or uniform) norm)**.** *For $f \in C(D, \mathbb{R})$ we define its supremum (or uniform) norm $\|f\|_\infty := \sup_{x \in D} |f(x)|$.*

In this definition we use the fact that D is closed and bounded - otherwise the supremum might not be finite.

Exercise 1.4. *Find an example of D that is not closed or not bounded, and $f \in C(D, \mathbb{R})$ such that $\|f\|_\infty$ as defined above is infinite.*

We called the expression above a norm, but recall that a norm on a vector space has again a precise mathematical definition and its conditions need to be checked:

Proposition 1.2. *$\|f\|_\infty$ indeed defines a norm on the vector space $C(D, \mathbb{R})$.*

Proof. We need to check the conditions for a norm.

(1) $\|f\|_\infty \geq 0$ with equality if and only if f is equal to the constant zero function. This is clear.
(2) $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty$ is also clear.
(3) Finally, we need to check the triangle inequality $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$. We have

$$\|f + g\|_\infty = \sup_{x \in D} |f(x) + g(x)| \leq \sup_{x \in D} (|f(x)| + |g(x)|)$$

by the triangle inequality. But now

$$\sup_{x \in D} (|f(x)| + |g(x)|) \leq \sup_{x \in D} |f(x)| + \sup_{x \in D} |g(x)|$$

and we conclude.

□

Thus $(C(D, \mathbb{R}), +, \|\cdot\|_\infty)$ is a normed vector space pretty much like $\mathbb{R}^n$ with any of these norms. This gives us a way to talk about convergence that is much more natural:

Proposition 1.3. *Let $(f_n)_{n \geq 1}$ be a sequence of $C(D, \mathbb{R})$ functions converging to some function $f : D \to \mathbb{R}$ w.r.t. the uniform norm. Then in fact f is continuous.*

This is a restatement of a result from Analysis I which says that pointwise limits of continuous functions are not continuous.

The proof technique is called the 3ϵ or $\epsilon/3$ argument and you have again seen it already in Analysis I. Let us give the proof just to understand what is now different from the earlier situation

Proof. It suffices to show that for every $x \in D$, we can find $\delta > 0$ such that $|f(x)-f(y)| < 3\epsilon$ whenever $\|x-y\| < \delta$.

We can first choose a fixed $n \in \mathbb{N}$ large enough so that $\|f_n - f\|_\infty < \epsilon$, and in particular $|f_n(x) - f(x)| < \epsilon$ for every $x \in D$ by the definition (these are the first two epsilons).

Further, by continuity of f_n we can choose $\delta > 0$ such that for every $y \in D$ with $\|x-y\| < \delta$, we have that $|f_n(x) - f_n(y)| < \epsilon$ (this is the third epsilon). Putting things together using triangle inequality we obtain:

$$
\begin{aligned}
(1.1) \quad |f(x) - f(y)| = |f(x) - f_n(x) + f_n(x) - f_n(y) + f_n(y) - f(y)| \le \\
\le |f(x) - f_n(x)| + |f_n(x) - f_n(y)| + |f_n(y) - f(y)| < 3\epsilon.
\end{aligned}
$$

□

Notice that for pointwise convergence the first step fails: we can not necessarily choose an n such that $\sup_{x\in D} |f_n(x) - f(x)| < 3\epsilon$.

Thus using this norm the set $C(D, \mathbb{R})$ is also closed under taking convergent sequences. In fact, it is even nicer than that and there are no gaps at all in the space, e.g. the space is complete - a notion you have met for $\mathbb{R}^n$ and that we recall here.

Definition 1.4 (Completeness of a normed space). *A normed space $(X, \|\cdot\|)$ is called complete if every Cauchy sequence $(x_n)_{n\ge1}$ (i.e. every sequence such that for every $\epsilon > 0$, there is an n_ϵ with $\|x_n - x_m\| \le \epsilon$ for all $n, m \ge n_\epsilon$) converges to an element $x \in X$.*

Theorem 1.5. *The space $(C(D, \mathbb{R}), +, \|\cdot\|_\infty)$ is a complete normed vector space.*

The idea is to us completeness of $\mathbb{R}$ to define a potential limiting function and then to verify that it really is that function.

Proof. We only need to check the completeness. So let $(f_n)_{n\ge1}$ be a Cauchy sequence in $C(D, \mathbb{R})$. As for every $x \in D$, $(f_n(x))_{n\ge1}$ is Cauchy and $\mathbb{R}$ is complete, we now a limit exists and we can denote this limit by $f(x)$. It remains to see that $f_n \to f$ in the uniform norm and that f is continuous. The latter claim follows from the proposition above, so we need to just prove the convergence w.r.t. the uniform norm. This is left as an exercise on the exercise sheet.

□

Remark 1.6. *Mathematicians call any normed vector space that is complete a Banach space. Such spaces are quite important in setting up quantum field theory.*

The completeness of the space has important application, one of them is finding solutions to ODEs via approximation. The tool used there is the Banach contraction mapping theorem that you have already met in Analysis II according to the course sheets and that is just recalled here:

Theorem 1.7 (Banach contraction mapping theorem). *Let $F : C(D, \mathbb{R}) \to C(D, \mathbb{R})$ be contractive w.r.t. the uniform norm: $\|F(f) - F(g)\|_\infty < C\|f - g\|_\infty$ with $C < 1$. Then there is a unique solution to $F(f) = f$ that can be obtained from the limit $\lim_{n\to\infty} F^{(n)}(f)$.*

1.3 Fourier series for continuous functions

The expansion of a function f on $[0,1]$ to a series of the form

$$f(x) = \sum_{n\geq 1} s_n \sin(2\pi n x) + \sum_{n\geq 0} c_n \cos(2\pi n x) \tag{1.2}$$

is called the Fourier expansion or Fourier series. We saw in the introduction that it could be quite useful, but we didn't see any results on its existence / uniqueness. So let us look at this in the context of continuous functions f now.

In fact we will see that these questions resolve themselves very smoothly once we find the "right functional space", but it is instructive to consider the questions already.

The first question is how should we go about finding the coefficients s_n, c_n? There the key is the following lemma.

Lemma 1.8. *The following orthogonality relations hold for integers $m, n \geq 0$:*

1. *Cosine-cosine Orthogonality:*

$$\int_0^1 \cos(2\pi n x)\cos(2\pi m x)\,dx = \begin{cases} 1, & \text{if } n = m = 0, \\ \frac{1}{2}, & \text{if } n = m \neq 0, \\ 0, & \text{if } n \neq m. \end{cases}$$

2. *Sine-Sine Orthogonality:*

$$\int_0^1 \sin(2\pi n x)\sin(2\pi m x)\,dx = \begin{cases} 0, & \text{if } n = 0 \text{ or } m = 0, \\ \frac{1}{2}, & \text{if } n = m \neq 0, \\ 0, & \text{if } n \neq m. \end{cases}$$

3. *Sine-Cosine Orthogonality:*

$$\int_0^1 \sin(2\pi n x)\cos(2\pi m x)\,dx = 0 \quad \forall n, m.$$

Proof. The proof is a simple consequence of trigonometric identities and their integrals and is left for the exercise sheet. □

The consequence of this observation is that if we expect the representation (1.2) to hold in any nice sense, then the coefficients s_n, c_n should be given by:

- Cosine Coefficients c_n:

$$c_n = 2\int_0^1 f(x)\cos(2\pi n x)\,dx, \quad \text{for } n \geq 1.$$

  For the constant term c_0, we have:

$$c_0 = \int_0^1 f(x)\,dx.$$

- Sine Coefficients s_n:

$$s_n = 2\int_0^1 f(x)\sin(2\pi n x)\,dx, \quad \text{for } n \geq 1.$$

Further notice that if we want it to hold at the endpoints, then we better have $f(0) = f(1)$ as this also holds for every function in the series.

Maybe a bit surprisingly both the existence and uniqueness are really not clear even for continuous functions!

Indeed, the understanding of counterexamples has evolved with time. The first observation is as follows

- There exists a continuous function f satisfying $f(0) = f(1)$ whose Fourier series converges pointwise everywhere but does not converge uniformly.

It is not easy to come up with such a function but once given, it is easy to check (probably on the exercise sheet).

A more stunning claim comes from the second half of 19th century from Du Bois-Reymond:

- There exist continuous functions $f \in C([0,1],\mathbb{R})$ with $f(0) = f(1)$ such that the Fourier series diverges at a point $x \in [0,1]$.

This was then extended by several people, including Kolmogorov to show that

- There are continuous functions $f \in C([0,1],\mathbb{R})$ with $f(0) = f(1)$ where the Fourier series diverges at infinitely many or even dense set of points.

Finally, Katznelson showed in 1970s that in fact

- For every continuous function f and every $\epsilon > 0$, there is some continuous function g with $\|g - f\| < \epsilon$ and the Fourier series of g diverges at some point.

This means that these unpleasant functions are really everywhere!

There are two ways out of this. First, one could just try to restrict the set of functions that one is considering. Second, one could try to weaken further the notion of convergence and maybe give up having pointwise convergence. We will mainly concentrate on the second direction, as the first is too restrictive. But to finish this section let us still show how the first direction can give us some nice results:

Proposition 1.9. *Let $f \in C^2([0,1])$ be twice continuously differentiable and satisfying $f(0) = f(1)$ and $f'(0) = f'(1)$. Then its Fourier series*

$$f(x) = \lim_{N\to\infty} \sum_{n\leq N} (s_n \sin(2\pi nx) + c_n \cos(2\pi nx),$$

converges w.r.t. $\|\cdot\|_\infty$.

Remark 1.10. *In fact the result holds under much less stringent conditions, e.g. when the functions are so-called Holder continuous, i.e. satisfying $|f(x) - f(y)| < |x-y|^a$ for some $a > 0$. Just the proof then needs a bit more care and is out of the scope for us.*

The key ingredient is the following lemma, which we observed when guessing the solution to the heat equation and that really explains why Fourier series are so useful:

Lemma 1.11. *Suppose that $f \in C([0,1],\mathbb{R})$ is k times continuously differentiable and satisfies $f^j(0) = f^j(1)$ for all $j = 0 \dots k-1$* [1] *Then there is some $C > 0$ such that for all $n \geq 1$ $|c_n| \leq Cn^{-k}$ and $|s_n| \leq Cn^{-k}$.*

The full proof is on the exercise sheet, but let's see the case $k = 1$.

[1] Here by $f^j(x)$ we mean the $j-th$ derivative of f at x, the $0-$th derivative being the function itself.

- We have by integration by parts that

$$\int_{[0,1]} \sin(2\pi nx) f(x)dx = \frac{1}{2\pi n}\int_{[0,1]} \cos(2\pi nx)f'(x) \leq \frac{1}{2\pi n}\|f'\|_\infty.$$

Let us proceed to the proof of the proposition.

Proof of Proposition 1.9. By the lemma we have that $|c_n|, |s_n| \leq Cn^{-2}$. Hence the Fourier series is Cauchy in the uniform norm. Indeed, if we denote by S_M the partial series

$$S_M(f) = \sum_{n\geq 1}^{N} (s_n \sin(2\pi nx) + c_n \cos(2\pi nx)),$$

then by the triangle inequality for all $M > N$:

$$\|S_M(f) - S_N(f)\|_\infty \leq \sum_{N<n\leq M} (\|s_n \sin(2\pi nx)\|_\infty + \|c_n \cos(2\pi nx)\|_\infty),$$

but $\|\sin(2\pi nx)\|_\infty = \|\cos(2\pi nx)\|_\infty = 1$ and hence we can bound the sum by

$$2C \sum_{N<n\leq M} n^{-2} \leq 2CN^{-1},$$

which goes to 0 as $N \to \infty$. Hence as $C([0,1], \mathbb{R})$ is complete for the uniform norm, we obtain the convergence to some continuous function g.

To conclude the theorem, we still need to argue that $f = g$. To do this we observe first (this is on the exercise sheet) that for all $n \geq 0$

$$\int_0^1 (f-g)\sin(2\pi nx)dx = \int_0^1 (f-g)\cos(2\pi nx)dx = 0.$$

It then follows from the next proposition that $g = f$. □

Proposition 1.12. *Suppose f is a continuous function on $[0,1]$. Then $s_n = 0, c_n = 0$ for all $n \geq 0$ if and only if $f(x) = 0$ for all $x \in [0,1]$.*

In particular, if the Fourier series of a function converges uniformly, then it is equal to the function itself and each function has at most one expansion in Fourier series.

Before proceeding further, you should pause and think why this is not immediate.

In fact proving this proposition giving the means we have is not completely straightforward. We will later see how it becomes slick and swift once we have found the right functional space for the Fourier series, where each function has a unique series converging exactly w.r.t. to the norm of the space.

We will prove here the proposition modulo a key construction, that is given on the example sheet.

Proof. We want to show that $s_n = 0, c_n = 0$ for all $n \geq 0$ gives $f = 0$. We will argue by contradiction and show that if for some $x_0 \in (0,1)$ it holds that $f(x_0) \neq 0$, then there is a contradiction with the hypothesis of the proposition.

The idea is to construct an approximations T_{m,x_0} of the identity, or if you wish an approximations of the Dirac delta function δ_{x_0} using finite sums of sines and cosines and to argue that 1) on the one hand $\int_0^1 f(x)T_{m,x_0}dx \approx f(x_0)$ for m large and 2) on the other hand by hypothesis $\int_0^1 f(x)T_{m,x_0}dx = 0$ for all $m \geq 1$. The construction is recorded in the following lemma.

Lemma 1.13. *For each $x_0 \in (0,1)$ one can construct a series of functions $T_{m,x_0}(x)$ as a linear sum of $\sin(2\pi xn), \cos(2\pi xm)$ with $m, n \leq N$, i.e. by setting*

$$T_{m,x_0}(x) = \sum_{n \leq N} (a_{n,x_0} \sin(2\pi nx) + b_{n,x_0} \cos(2\pi nx)$$

such that the following points hold.

(1) For every $m \geq 1$, $x \in [0,1]$ we have $T_{m,x_0}(x) \geq 0$
(2) For every $m \geq 1$ we have $\int_0^1 T_{m,x_0}(x) = 1$.
(3) For all $\delta > 0$: $\int_0^1 1_{|x-x_0|>\delta} T_{m,x_0}(x)dx \to 0$ as $N \to \infty$.

Given such a sequence of $(T_{m,x_0})_{m\geq 1}$, we obtain the contradiction as follows.
On the one hand by the hypothesis for all $m \geq 1$ we have

$$\int_0^1 f(x)T_{m,x_0}(x)dx = \sum_{n\leq m} \left(a_{n,x_0} \int_0^1 f(x)\sin(2\pi nx)dx + b_{n,x_0} \int_0^1 f(x)\cos(2\pi nx)dx \right) = 0.$$

On the other hand suppose $f(x_0) \neq 0$, say WLOG $f(x_0) > 0$. Then there is some $\delta > 0$ such that $f(x) > f(x_0)/2$ in some region $[-\delta + x_0, x_0 + \delta]$. Write

$$\int_0^1 f(x)T_{m,x_0}(x)dx = \int_{-\delta+x_0}^{\delta+x_0} f(x)T_{m,x_0}dx + \int_0^1 1_{|x-x_0|>\delta} f(x)T_{m,x_0}dx.$$

We can bound the second term in absolute value by

$$\|f\|_\infty \int \int_0^1 1_{|x-x_0|>\delta} T_{m,x_0}(x)dx,$$

which goes to zero by the lemma. The first term however can be bounded from below by $f(x_0)/2 \int_{-\delta+x_0}^{\delta+x_0} T_{m,x_0}(x)dx$. Combining the conditions of Lemma, we see that for m large enough this integral is larger than say $1/2$ and thus the whole term is larger than $f(x_0)/4$ for m large enough. And in particular we conclude that $\int_0^1 f(x)T_{m,x_0}(x)dx \neq 0$ for m large enough! This gives a contradiction. But our assumption was that $f(x_0) \neq 0$, so this cannot hold and we conclude the proposition. □

This was in the end not hard, but quite a fiddly proof and moreover also the existence of Fourier series for continuous functions had several delicate points. We would prefer if the existence and uniqueness would be simple consequences of a good set-up, like in the case of $\mathbb{R}^n$. With this in mind, we will go towards larger function spaces.

Section 2

Lebesgue measure and Lebesgue integral on $\mathbb{R}^n$

We will continue our aim of constructing a convenient / appropriate function space for the Fourier expansions. Motivated by the finite-dimensional example, we would want to construct a space of functions with an inner product of the type $\int f(x)g(x)dx$ and then see the Fourier series as an orthogonormal basis of this space.

To do this, we will have to make a detour and renew our understanding of two intimately linked notions: 1) the integral of a function 2) the size / measure of subsets of $\mathbb{R}^n$.

But let us start off by discussing why the Riemann integral does not suffice.

2.1 An issue with the Riemann integral

One way to define the Riemann integral of a function $f : [0,1] \to \mathbb{R}$ is as follows.

(1) We subdivide $[0,1]$ into 2^n equal disjoint intervals $D_{n,i} = [i2^{-n}, (i+1)2^{-n}]$ each of size 2^{-n};
(2) We call a function Riemann integrable if $U_n := 2^{-n}\sum_{i=0}^{2^n}\sup_{x\in D_{n,i}} f(x)$ (which is decreasing) and $L_n := 2^{-n}\sum_{i=0}^{2^n}\inf_{x\in D_{n,i}} f(x)$ (which is increasing) both converge to the same limit.
(3) We define the Riemann integral of f, that from now on we denote for clarity by $\circledR\int_0^1 f(x)dx$ to be equal to this limit.

It is easy to see that Riemann integral satisfies some nice properties:

Exercise 2.1. *Show that the Riemann integral satisfies some desirable properties:*

- *All continuous functions on* $[0,1]$ *are Riemann integrable*
- *Every function* f *that changes value finitely many times is Riemann integrable*
- *Linearity: If* f, g *are Riemann integrable on* $[0,1]$*, then so is their sum and the integral is equal to the sums.*

However, the Riemann integrability does not behave well under limits or infinite sums. Indeed, consider a enumeration $q_1, q_2, \dots$ of all rational numbers in $[0,1]$ (can you give a concrete one?) and define $f_n(x) = 1$ if $x \in \{q_1, \dots, q_n\}$ and $f_n(x) = 0$ otherwise. Then each f_n is Riemann-integrable (with $\circledR\int_0^1 f_n(x) = 0$) by the exercise above, but the limit is not Riemann integrable as in every interval the sup is equal to 1 and inf is equal to 0 and thus $U_n = 1$ for all $n \geq 1$ and $L_n = 0$ for all $n \geq 1$.

We will see how this is remedied with the notion of Lebesgue integral.

2.2 The Lebesgue measure

We start however by revisiting the notion of size / volume / measure of subsets of $\mathbb{R}^n$. This is directly related to the integral as even in the case of Riemann integral, if the set $A \subseteq \mathbb{R}$ is nice enough then $\int_{\mathbb{R}} 1_A(x)dx = \text{size}(A)$. What should nice enough be is one of the main questions.

As said, the Lebesgue measure on $\mathbb{R}$ generalizes the notion of length and assigns each permissible subset of $\mathbb{R}$ a size. More formally, the Lebesgue measure is a function $\mu : \mathcal{F} \to$

$[0, \infty)$, where $\mathcal{F}$ is some collection of subsets of $\mathbb{R}$ satisfying some collection of properties. What should such natural properties be?

(1) First, in the case of $\mathbb{R}$, we would like the length / measure of each interval $[a, b]$, (a, b), $[a, b)$ or $(a, b]$ to be just $b - a$. In particular each point $\{x\}$ should have length 0
(2) We certainly would want also $L(\emptyset) = 0$ and $L(A) \geq 0$ for all $A \geq 0$.
(3) Second, we would like measure to satisfy some additivity properties. For example the size of the union of two disjoint sets should clearly be just the sum of their sizes: i.e. in symbols $L(A_1 \cup A_2) = L(A_1) + L(A_2)$. By induction this should hold for any finite number of disjoint intervals: $L(A_1 \cup \cdots \cup A_n) = L(A_1) + \ldots L(A_n)$.
(4) Further, it might make sense for this additivity to hold also if we have countably many disjoint sets? But attention! We cannot ask it for all infinite unions: indeed, for example $[0, 1]$ can be seen as a disjoint union of all points $\{x\}$ in $[0, 1]$, but the sum of their lengths would be 0 whereas the measure of $[0, 1]$ has to be clearly 1!

Observe that only the first property has something specific to do with $\mathbb{R}$, all the others are of very abstract nature. A big breakthrough by Lebesgue was to understand that combining these properties gives the right mathematical framework to talk of size / measure on any set! This is encapsulated in the following general definition:

Definition 2.1 (Measure space, Borel 1898, Lebesgue 1901-1903)**.** *A measure space is a triple* $(\Omega, \mathcal{F}, \mu)$*, where*

- Ω *is a set, called the sample space or the universe.*
- $\mathcal{F}$ *is a set of subsets of* Ω*, satisfying:*
  - $\emptyset \in \mathcal{F}$*;*
  - *if* $A \in \mathcal{F}$*, then also* $A^c \in \mathcal{F}$*;*
  - *If* $A_1, A_2, \dots \in \mathcal{F}$*, then also* $\bigcup_{n \geq 1} A_n \in \mathcal{F}$*.*

  $\mathcal{F}$ *is called a* σ*-algebra and any* $A \in \mathcal{F}$ *is called a measurable set.*
- *And finally, we have a function* $\mu : \mathcal{F} \to [0, \infty]$ *satisfying* $\mu(\emptyset) = 0$ *and countable additivity for disjoint sets: if* $A_1, A_2, \dots \in \mathcal{F}$ *are pairwise disjoint,*

$$\mu(\bigcup_{n \geq 1} A_n) = \sum_{n \geq 1} \mu(A_n).$$

  This function μ *is called a measure. If* $\mu(\Omega) < \infty$*, we call* μ *a finite measure.*

Geometrically we interpret:

- Ω as our space of points
- $\mathcal{F}$ as the collection of subsets for which our notion of volume can be defined
- μ our notion of volume: it gives each measurable set its volume.

We can define a measure on any set of points, finite or infinite. Some telling examples are:

Example 2.2 (Counting measure)**.** *On any set* Ω *one can define the counting measure* μ_c*: we set* $\mathcal{F} := \mathcal{P}(\Omega)$ *(the set of all subsets), and* $\mu_c(\{\omega\}) := 1$ *for any* $\omega \in \Omega$*. For any finite set* E*,* $\mu_c(E)$ *gives its number of elements. If* E *is infinite, then so is* $\mu_c(E)$*. In particular, if* Ω *is an infinite set, then* $\mu_c(\Omega) = \infty$*, so this is a measure, but not a finite measure.*

Notice that on a space with finite number of points it gives the natural uniform measure - each point is treated in the same way. However, it is not the natural measure of size on say

$\mathbb{R}$ as the size of say $[0,1]$ would be infinite. The natural uniform measure on $[0,1]$ or $\mathbb{R}$ will be called the Lebesgue measure, but its existence is already mathematically non-trivial - we will come to this in a bit.

Example 2.3 (Delta measure)**.** *The (Dirac) delta function that you have seen mentioned in the courses, is actually a measure, not a function and can be defined on any space and for any $\sigma-$algebra that contains points. On any set Ω one can define the Dirac delta measure μ_x at the point x as follows: suppose $\mathcal{F}$ contains points and we set $\mu_x(\{x\}) = 1$ and more generally $\mu_x(F) = 1$ if $x \in F$ and $\mu_x(F) = 0$ otherwise, for every $F \in \mathcal{F}$.*

We will come back to this and its connection to the 'Dirac delta function' you have seen before later on.

Finally, a both nice and important aspect of the framework of measure spaces is that it also gives the mathematical basis for probability theory - this was observed by A. Kolmogorov some 30 years after the introduction of measure spaces! A probability space is a measure space with total mass equal to 1, i.e. $\mu(\Omega) = 1$. In that case we often use the notation of $\mathbb{P}$ for the measure μ. The framework of probability is used for observing / measuring what's going on in the world:

- Ω as the space of all microstates / all possible outcomes; e.g. the states of the atmosphere
- $\mathcal{F}$ is the collection of observable events / outcomes: i.e. subsets of microstates, whose happening or not happening can be observed; for example we can maybe only measure macroscopic parameters like temperature, or the amount of rain over an hour
- The measure $\mathbb{P}$ will assign a number in $[0,1]$, called probability, to each observable event. Those events that surely happen, get probability 1.

Example 2.4. *The probability space for describing a fair coin toss would be*

$$(\{H,T\}, \{\emptyset, \{H\}, \{T\}, \{H,T\}\}, \mathbb{P}),$$

where $\mathbb{P}(\{H\}) = \mathbb{P}(\{F\}) = 1/2$.

The probability space for describing a fair dice would be

$$(\{1,2,3,4,5,6\}, \mathcal{P}(\{1,2,3,4,5,6\}), \mathbb{P}),$$

where we define $\mathbb{P}(F) = |F|/6$. If instead we paint all the faces 1,2,3,4,5 black so they become indistinguishable, we can modify our model by taking $\mathcal{F} = \{\emptyset, \{1,2,3,4,5,6\}, \{1,2,3,4,5\}, \{6\}\}$ and using the probability measure $\widetilde{\mathbb{P}}$ defined only on these subsets, still with the same formula as above.

Exercise 2.2. *Find a measure space to describe two unrelated fair coin tosses. What assumptions are you making in giving the description? Define a sigma-algebra suitable for studying the situation where one can only ask if the two coins have the same side up, or different sides up.*

Finally, as mentioned not all natural measure spaces are simple to define. We already mentioned that the natural "uniform" measure on $[0,1]$ or $\mathbb{R}$ needs some work. But one would actually also want to define natural measures on more complicated structures like the space of all continuous functions - indeed, this gives one way to formalize path-integrals in quantum mechanics. This was achieved by Wiener in the beginning of 20th century; the similar task for string theory, i.e. defining probability measures over surfaces with different metric structures has been partially resolved only in the recent years.

2.2.1 Basic properties of measure spaces

Before discussing the Lebesgue measure let us play aroud a bit with the notion of a measure space.

First, the following lemma helps to see which other sets would be measurable:

Lemma 2.5 (Constructing more measurable sets)**.** *Consider a set Ω with a σ-algebra $\mathcal{F}$.*

(1) If $A_1, A_2, \dots, \in \mathcal{F}$, then also $\bigcap_{n\geq 1} A_n \in \mathcal{F}$.

(2) Then also $\Omega \in \mathcal{F}$ and if $A, B \in \mathcal{F}$, then also $A \setminus B \in \mathcal{F}$.

(3) For any $n \geq 1$, if $A_1, \dots, A_n \in \mathcal{F}$, then also $A_1 \cup \dots \cup A_n \in \mathcal{F}$ and $A_1 \cap \dots \cap A_n \in \mathcal{F}$.

Proof of Lemma 2.5. By de Morgan's laws for any sets $(A_i)_{i\in I}$, we have that

$$\bigcap_{i\in I} A_i = (\bigcup_{i\in I} A_i^c)^c.$$

Property (1) follows from this, as if $A_1, A_2, \dots \in \mathcal{F}$, then by the definition of a σ-algebra also $A_1^c, A_2^c, \dots \in \mathcal{F}$ and hence

$$(\bigcup_{i\geq 1} A_i^c)^c \in \mathcal{F}.$$

For (3), again by de Morgan laws, it suffices to show that $A_1 \cup \dots \cup A_n \in \mathcal{F}$. But this follows from the definition of a σ-algebra, as $A_1 \cup \dots \cup A_n = \bigcup_{i\geq 1} A_i$ with $A_k = \emptyset$ for $k \geq n+1$. Finally, for (2) we can just write $\Omega = \emptyset^c$.

The fact that $A \setminus B \in \mathcal{F}$ is left as an exercise. □

The statements are also very intuitive at least in the context of probability: e.g. the first one says that if we can observe if some events $A_1, A_2, \dots$ happen, then we can observe if they all happen at once; the second property says that if two events can be observed, then we can always also observe if one of them happened but not the other one.

In a similar vein, the basic conditions on the measure, give rise to several natural properties too:

Proposition 2.6 (Basic properties of a measure and a probability measure)**.** *Consider a measure space $(\Omega, \mathcal{F}, \mu)$. Let $A_1, A_2, \dots \in \mathcal{F}$. Then*

(1) For any $n \geq 1$, and $A_1, \dots, A_n$ disjoint, we have finite additivity

$$\mu(A_1) + \dots + \mu(A_n) = \mu(A_1 \cup \dots \cup A_n).$$

In particular if $A_1 \subseteq A_2$ then $\mu(A_1) \leq \mu(A_2)$.

(2) If for all $n \geq 1$, we have $A_n \subseteq A_{n+1}$, then as $n \to \infty$, it holds that $\mu(A_n) \to \mu(\bigcup_{k\geq 1} A_k)$.

(3) We have countable subadditivity (also called the union bound): $\mu(\bigcup_{n\geq 1} A_n) \leq \sum_{n\geq 1} \mu(A_n)$.

If in fact $\mu(\Omega)$ is finite (e.g. a probability measure), we further also have the following properties:

(4) For any $A \in \mathcal{F}$, we have that $\mu(A^c) = \mu(\Omega) - \mu(A)$.

(5) If for all $n \geq 1$, we have $A_n \supseteq A_{n+1}$, then as $n \to \infty$, it holds that $\mu(A_n) \to \mu(\bigcap_{k\geq 1} A_k)$.

Again, please do check that all these properties also make sense intuitively!

Proof of Proposition 2.6. Finite additivity follows from countable additivity by taking $A_k = \emptyset$ for $k \geq n+1$.

(2), (3) are left as exercises.

For (4), we just notice that A and A^c are disjoint and $A \cup A^c = \Omega$. Thus by disjoint additivity $\mathbb{P}(A) + \mathbb{P}(A^c) = 1$. Finally, for (5), define $B_n = A_n^c$. Then $\mathbb{P}(A_n) = \mathbb{P}(B_n^c) = 1 - \mathbb{P}(B_n)$. Similarly $\mathbb{P}(\bigcap_{k\geq 1} A_k) = 1 - \mathbb{P}(\bigcup_{k\geq 1} B_k)$. Thus the result follows from (2). The rest is left as an exercise ☐

2.2.2 The Lebesgue measure

The Lebesgue measure is the right notion uniform measure on the spaces $\mathbb{R}^n$ (or say a unit cube $[0,1]^n$ or a ball). This measure is called uniform because it is isotropic, i.e. it treats all the points in the set equally. More formally, it is up to a multiplicative constant the measure μ such that $\mu(A) = \mu(\lambda + A)$, where A is some measurable set and $\lambda + A := \{a + \lambda : a \in A\}$.

To define the uniform measure on $\mathbb{R}^n$, we should first pick the right $\sigma-$algebra. First, it certainly has to be big enough to contain at least all the boxes. Now, it is an interesting fact that in the standard axiomatization of mathematics [2] one cannot take the $\sigma-$-algebra to be equal to $\mathcal{P}(\mathbb{R}^n)$ - otherwise one runs into contradictions as explained in the non-examinable part of the example sheet. However, there are some $\sigma-$algebras that are big enough to contain all sets we might be interested in and small enough at the same time to create no contradiction.

Definition 2.7 (Borel σ-algebra)**.** *The Borel σ-algebra $\mathcal{F}_B$ on $\mathbb{R}^n$ is defined as the smallest $\sigma-$algebra containing all boxes, i.e. all sets of the form $\Pi_{i=1}^n [a_i, b_i]$ with real numbers $a_i < b_i$.*

This definition hides a claim: the fact that such a smallest $\sigma-$algebra exists. However, it is a simple but not that illuminating exercise to show that an arbitrary intersection of $\sigma-$algebras is a $\sigma-$algebra and thus the smallest has a well-defined meaning. It is maybe more interesting to see what it contains, i.e. what we can measure [3]:

Example 2.8. *The Borel $\sigma-$algebra contains for example all points, i.e. sets of the form $\{x\}$: indeed, we can write*

$$\{x\} = \bigcap_{m\geq 1} (\{x\} + [-m^{-1}, m^{-1}]^n).$$

Exercise 2.3. *Show that the Borel $\sigma-$algebra on $\mathbb{R}^n$ also contains all products of half-lines $\Pi_{i=1}^n (-\infty, a_i]$, all open balls $B(x,r)$ and in fact all open sets of $\mathbb{R}^n$*

The main theorem of this section is then the following result, that we assume without proof:

Theorem 2.9 (Existence and uniqueness of Lebesgue measure)**.** *There is a unique measure λ defined on $(\mathbb{R}^n, \mathcal{F}_B)$ such that the measure of each box $\Pi_{i=1}^n [a_i, b_i]$ is given by $\Pi_{i=1}^n (b_i - a_i)$*

[2] Meaning that we assume the axiom of choice

[3] It is maybe as interesting to see that there are sets in the power-set of $\mathbb{R}^n$ that do not belong to the Borel $\sigma-$algebra. However, describing them explicitly is not that easy - if interested, see the for fun section on the example sheet.

Some other nice properties of the Lebesgue measure follow from this theorem:

- It is translation invariant: for every set $A \in \mathcal{F}_B$, if we denote by $A + b$ the set $\{a + b : a \in A\}$, then the Lebesgue measure λ satisfies $\lambda(A) = \lambda(A + b)$. Indeed, denote by $\widetilde{\lambda}(A) := \lambda(A + b)$. This defines another measure on $(\mathbb{R}^n, \mathcal{F}_B)$ such that $\widetilde{\lambda}(\text{box})$ equals the volume of the box. Thus by uniqueness part of the theorem we obtain $\widetilde{\lambda} = \lambda$ and hence $\lambda(A + b) = \lambda(A)$ for all Borel sets A.
- It can be also proved that the Lebesgue measure is rotation invariant: for every set $A \in \mathcal{F}_B$, if we denote by $R(A)$ the set rotated by the rotation matrix R, then the Lebesgue measure λ satisfies $\lambda(A) = \lambda(R(A))$.

Maybe somewhat surprisingly the proof of this natural theorem is not immediate. The problem is the following: it is simple to assign measure to each box, or each finite union of disjoint boxes etc...however, the Borel σ-algebra is much richer than that. Indeed, there are sets in the Borel σ-algebra that one cannot obtain in a finite number of steps by starting with boxes and taking iteratively unions, intersections and complements in any order. Hence the fact that one can assign a measure to all Borel sets in a way that the axioms are satisfied and boxes have the right size is not immediate. Also the statement of uniqueness is non-evident for the same reason - why should equality for all boxes imply it for all Borel sets?

The proof goes beyond the scope of this course, but here is the sketch for one of the possible approaches for those interested (not examinable).

$\star$ *Start of non-examinable section* $\star$

For any rectangle $R = \Pi_{i=1}^n [a_i, b_i]$, let's denote by $|R|$ its natural volume $\Pi_{i=1}^n (b_i - a_i)$.

(1) First, we define for any set $A \subseteq \mathbb{R}^d$ a notion of size called the exterior measure: $m^*(E) := \inf \sum_{i=1}^\infty |R_i|$, where the infimum is over all coverings of the set E using rectangles - this gives a certain approximation of size from above.

Notice that from this definition it is not immediate that even $m^*(R) = |R|$, but that can be argued for both closed and open rectangles. Also, it is important that we allow for countably many rectangles - see exercise sheet.

(2) It comes out that showing all the axioms of the measure for all subsets of $\mathbb{R}^d$ is impossible [4]. So now comes the key idea of choosing a subclass of sets which is large enough to contain Borel sets, but small enough to be able to make everything work: we call a set measurable if for every $\epsilon > 0$, there is some countable collection of rectangles $(R_i)_{i \geq 1}$ such that $E \subseteq \bigcup_{i \geq 1} R_i$ and $m^*(E \triangle (\bigcup_{i \geq 1} R_i)) < \epsilon$. This means that our earlier approximation from above can be chosen to fit well.

(3) It then remains to argue that these sets actually form a $\sigma-$algebra and that all axioms are satisfied for $((\mathbb{R}^d, \mathcal{F}_L, m^*)$. In fact they form a $\sigma-$algebra, called the Lebesgue $\sigma-$algebra $\mathcal{F}_L$, that is even larger than $\mathcal{F}_B$!

This final step doesn't require any big theorems or inputs, but it does require quite a bit of care in setting up the order of the argument. It is then an easy conclusion that $\mathcal{F}_B \subseteq \mathcal{F}_L$, as $\mathcal{F}_B$ can be generated from just rectangles and we can conclude.

$\star$ *End of non-examinable section* $\star$

[4] as long as one assumes the Axiom of Choice

Example 2.10. *The Lebesgue measure of a point is zero: indeed for every $\epsilon > 0$, we have that $\lambda(\{x\}) \leq \lambda(\{x\} + [-\epsilon, \epsilon]^n) = (2\epsilon)^n$, which can be made arbitrarily small.*

Hence also the measure of all rational numbers is zero: we have by countably additivity $\mu(\mathbb{Q}) = \sum_{q \in \mathbb{Q}} \mu(\{q\}) = 0$.

Exercise 2.4. *Show that the Lebesgue measure of $\mathbb{R}^n$ is infinite and that the Lebesgue measure of the line segment $[0, 1] \times \{0\} \cdots \times \{0\} \subseteq R^n$ is zero.*

Now consider the Lebesgue measure on $\mathbb{R}$. Prove that the measure of irrational numbers contained in $[0, R]$ is equal to R; prove also that the Lebesgue measure of the Cantor set is zero.

2.3 Lebesgue integral

Recall our grand plan was to construct a function space which has a nice inner product of the form $\int f(x)g(x)dx$ and all the nice properties of a function space like linearity, closedness under limits and completeness. With Riemann integral this would never be possible, as we saw it does not behave that well under taking limits. Hence let us see another notion of integral, called the Lebesgue integral. To start off, let's see that defining a measure always gives us a natural space of functions and those will be the candidates for defining the integral for.

2.3.1 Measurable functions

Each measure space comes with a class of natural functions, called measurable functions. These will also form the class of functions for which we aim to define the Lebesgue integral.

We will constrain ourselves to working with functions from $\mathbb{R}^n \to \mathbb{R}$, although the notion of a "measurable" function is quite a bit larger, applying to maps between any two sets together with $\sigma-$algebras; in our case these would be the pairs $(\mathbb{R}^n, \mathcal{F}_B)$ and $(\mathbb{R}, \mathcal{F}_B)$, where in both cases we consider the Borel $\sigma-$algebra.

The simplest measurable functions (on $(\mathbb{R}^n, \mathcal{F}_B)$) are those given by characteristic functions 1_E for some Borel-measurable set $E \in \mathcal{F}_B$, i.e. functions that tell us whether x is in a set - then $1_E(x) = 1$ - or not, in which case $1_E(x) = 0$. Their countable linear combinations are called simple functions:

Definition 2.11 (Simple functions). *Let $E_1, E_2, \dots$ be disjoint Borel sets in $\mathbb{R}^n$ and $c_1, c_2, \dots$ real numbers. Then a function of the form $f(x) = \sum_{i \geq 1} c_i 1_{x \in E_i}$ is called a simple function.*

We can then define

Definition 2.12 (Measurable function). *We call a function $f : \mathbb{R}^n \to \mathbb{R}$ measurable if it is a pointwise limit of simple functions.*

This definition is natural, however it is not so easy to work with. So let us start by proving an equivalence with another rather nice definition.

Proposition 2.13. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is measurable if and only if for every $a < b$ the preimage $f^{-1}([a, b)))$ is Borel measurable.*

We will sometimes call this condition the preimage condition.

The proof consists of two lemmas, one for each direction both teaching us something about measurable functions:

Lemma 2.14. *Suppose that the sequence of functions $(f_i)_{i\geq 1}$ from $\mathbb{R}^n$ to $\mathbb{R}$ is such that for every $a < b$ the preimage $f_i^{-1}([a,b)))$ is Borel measurable. Suppose also that f_i converge pointwise to f.*

Then f also satisfies the same property, i.e. the preimages $f^{-1}([a,b))$ are Borel measurable.

Lemma 2.15. *Suppose f is such that for every $a < b$ the preimage $f^{-1}([a,b)))$ is Borel measurable. Then f is a pointwise limit of simple functions f_n.*

Further, a sequence can be chosen to be pointwise increasing and to converge uniformly.

The proof of proposition follows from these two lemmas.

Proof of Proposition. Lemma 2.15 tells us directly that if f satisfies the preimage condition, then it is measurable.

Let us now show conversely that each measurable function satisfies the preimage condition. Using Lemma 2.14 and the definition of measurable functions it satisfies to show that each simple function satisfies the preimage property.

So, consider a simple function $g = \sum_i c_i 1_{E_i}$ with $c_i \in \mathbb{R}$ and $E_i \in \mathcal{F}_B$. Then $f^{-1}([a,b)) = \cup_{i:c_i\in[a,b)]} E_i$ is a countable union of Borel measurable sets and thus Borel measurable as desired, finishing the proof. □

Let us now prove the two lemmas.

Proof of Lemma 2.14. Our aim is to show that $f^{-1}([a,b))$ is a Borel set and this follows from:

$$f^{-1}([a,b)) = \bigcap_{j\geq 1}\bigcup_{k\geq 1}\bigcup_{n\geq 1}\bigcap_{m\geq n} f_m^{-1}((a-1/j, b-1/k)).$$

The verification of this equality is on the exercise sheet

□

Proof of Lemma 2.15. Consider $f_n : \mathbb{R}^n \to \mathbb{R}$ defined by

$$f_n(x) := 2^{-n}\lfloor 2^n f(x)\rfloor.$$

Each f_n is a simple function as we can write

$$f_n(x) = \sum_{k\in\mathbb{Z}} k2^{-n} 1_{\{f(x)\in[k2^{-n},(k+1)2^{-n})\}}$$

and by assumption the sets $\{f(x) \in [k2^{-n}, (k+1)2^{-n})\}$ are measurable. Further we notice that

$$f_n(x) = 2^{-n}\lfloor 2^n f(x)\rfloor = 2^{-m}2^{m-n}\lfloor 2^{-n} f(x)\rfloor \geq 2^{-m}\lfloor 2^{-} f(x)\rfloor = f_m(x)$$

proving monotonicity. As also

$$f_n(x) \geq 2^{-n}2^n(f(x) - 2^{-n}) = f(x) - 2^{-n}$$

and thus $\|f(x) - f_n(x)\| \leq 2^{-n}$ and we obtain uniform convergence. □

Several nice properties of the space of measurable functions can be now verified. First, the space of measurable functions again has a linear structure:

Lemma 2.16. *If f, g are measurable, then so are λf for $\lambda \in \mathbb{R}$ and $f + g$.*

Proof. This is on the exercise sheet. □

Second, the space of measurable functions is closed under pointwise limits.

Lemma 2.17. *Let $(f_n)_{n\geq 1}$ be a sequence of measurable functions converging pointwise to a function f. Then f is also measurable.*

Motice that this lemma follows directly from Lemma 2.14 under the equivalence of definitions given by Proposition 2.13. Finally, the space contains all continuous functions.

Lemma 2.18. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be continuous, then f is also measurable.*

Proof. This is also on the exercise sheet

□

To finish this section we remark again that in everything we did above we didn't use at all that the domain of our functions was $\mathbb{R}^n$! We could have equally well worked on any other measure space, laying groundwork for defining integration in a very large generality!

2.3.2 The idea behind Lebesgue integral

Recall that if a function $f : \mathbb{R} \to \mathbb{R}$ is Riemann-integrable then we can calculate its Riemann integral on $[0,1]$ using the following approximation procedure:

- we subdivide $[0,1]$ into 2^n equal disjoint intervals D_i each of size 2^{-n};
- we calculate the approximated integral $2^{-n}\sum_{i=0}^{2n-1} f(i2^{-n})$;
- we take the limit $n \to \infty$.

To calculate the Lebesgue integral (that we will shortly define) for a Lebesgue-integrable function on $[0,1]$ we can also proceed via an approximation, but rather in the image of the function:

- we take the dyadic approximations from the previous subsection: $f_n := 2^{-n}\lfloor 2^n f(x)\rfloor$;
- we calculate $\sum_{i\in\mathbb{Z}} i2^{-n}\lambda(x \in [0,1] : f_n(x) = i2^{-n})$;
- and take the limit $n \to \infty$.

So in some sense the difference w.r.t. to the Riemann integral is that we group the values not according to the vicinity in the domain $[0,1]$, but rather based on the vicinity of the function values. So if you wish, you can think that the Lebesgue integral treats each function in a more personal way, the approximations are based on its behaviour.

Let us now move to the formal definition of the Lebesgue integral, which we do using a slightly wider class of approximations.

2.3.3 Definition of the Lebesgue integral via simple functions

There are several ways to define the Riemann integral.[5] Similarly, there are multiple equivalent approaches to constructing the Lebesgue integral (e.g., Stein–Shakarchi, Kolmogorov–Fomin, and Boccarini all present slightly different versions). Last year we picked a definition via dyadic approximations that is maybe simplest to state and intuitive to grasp; this year we go for a definition that is simplest to work with mathematically.

[5] For instance, one can define it using upper and lower (Darboux) sums with arbitrary partitions or just dyadic ones; or even avoid these altogether and define integrability via convergence of approximating Riemann sums in a suitable sense.

Although the definition we will give works for measurable functions on any measure space $(\Omega, \mathcal{F}, \mu)$, we will focus on the case $(\Omega, \mathcal{F}, \mu) = (\mathbb{R}^n, \mathcal{F}_B, \lambda)$, i.e., $\mathbb{R}^n$ with its Borel sigma-algebra and Lebesgue measure.

For simple functions, i.e. step functions of the form $f(x) = \sum_i c_i 1_{E_i}$, with E_i are disjoint Borel sets and $c_i \in \mathbb{R}$ the Lebesgue integral is simple to define:

Lemma 2.19 (Lebesgue integral for simple functions)**.** *Let $f(x)$ be a simple function given e.g. by $f(x) = \sum_i c_i 1_{E_i}$. We call f Lebesgue integrable if $\sum_i |c_i| \lambda(E_i) < \infty$ and define its Lebesgue integral by*

$$\int_{\mathbb{R}^n} f(x) \lambda(dx) := \sum_i c_i \lambda(E_i).$$

Further, being integrable and the value of the integral are independent of the chosen representation of f as a simple function.

This is called a lemma and not a definition because of the final part. For example the function $f(x) = 1_{[0,1]}$ could be equivalently written as $f(x) = 1_{[0,1/2)} + 1_{[1/2,1]}$ or even as an infinite sum $f(x) = \sum_i 1_{E_i}$ where $(E_i)_{i \geq 1}$ is any partition of $[0,1]$ into disjoint Borel sets (can you find one?). Thus, one does need to verify that integrability and the integral do not depend on the choice of the representation. Luckily, this is a simple check.

Proof. Denote by S the set the image of f, i.e. the set $\{f(x) : x \in \mathbb{R}^n\}$. Notice that for a simple function it is always countable.

Then observe that for every $s \in S$, we can define $F_s := \{x : f(x) = s\}$ that depend only on the function f. Further, for any representation $f(x) = \sum_i c_i 1_{E_i}$ we have $F_s = \cup_{i: c_i = s} E_i$ and in particular all F_s are Borel and disjoint for different $s \in S$.

As $\sum_i |c_i| \lambda(E_i) = \sum_{s \in S} |s| \lambda(F_s)$ and $\sum_i c_i \lambda(E_i) = \sum_{s \in S} s \lambda(F_s)$ we conclude that both integrability and the integral are well-defined and independent of the representation. ☐

Example 2.20. *For example, in contrast to the Riemann integral $f(x) = 1_{\mathbb{Q}}(x)$ is integrable with integral equal to 0. Similarly, and $f(x) = 1_{[0,1] \setminus \mathbb{Q}}(x)$ is integrable with integral equal to 1 - both are themselves simple functions!*

For general measurable functions we will proceed in two steps: first we define the Lebesgue integral for non-negative functions, and then generalise it to all measurable functions by separating into non-negative and positive parts.

Definition 2.21 (Lebesgue integral)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be non-negative and measurable. Then we define*

$$\int_{\mathbb{R}^n} f(x) \, \lambda(dx) := \sup \left\{ \int_{\mathbb{R}^n} g(x) \, \lambda(dx) \;\middle|\; 0 \leq g \leq f, \, g \text{ simple} \right\}.$$

We say that f is integrable if this supremum is finite.

For a general measurable function f, we write $f = f_+ - f_-$, where f_+, f_- are the positive and negative parts of f, given by

$$f_+ = \max(f, 0), \quad f_- = \max(-f, 0).$$

We say that f is integrable if both f_+ and f_- are integrable, and define

$$\int_{\mathbb{R}^n} f(x) \, \lambda(dx) := \int_{\mathbb{R}^n} f_+(x) \, \lambda(dx) - \int_{\mathbb{R}^n} f_-(x) \, \lambda(dx).$$

We will sometimes for the sake of brevity also use the shortcut $\int f\lambda(dx) := \int_{\mathbb{R}^n} f(x)\,\lambda(dx)$ or even $\int f d\lambda$.

Remark 2.22. *One could also try to alternatively define the integral as a limit of integrals of any sequence of uniformly approximating simple functions. This works well when integrating over sets of finite measure (like say* $[0,1]$*); however, as you see on the example sheet, it would require care when integrating over sets of infinite measure, like* $\mathbb{R}$ *or* $\mathbb{R}^n$*.*

Exercise 2.5. *Verify from the definitions that* $f(x) = x1_{[0,1]}$ *is measurable and integrable. Calculate its integral also from the definition. What about* $f(x) = x^{-1}1_{(0,1]}$

Remark 2.23. *We can further define the integral over any Borel integrable set* E*, which we denote by* $\int_E f(x)\lambda(dx)$ *by just considering the integral of* $1_E(x)f(x)$*, which as a product of measurable functions is nicely measurable.*

Remark 2.24. *We can similarly define an integral over complex-valued functions by just separating the real and imaginary parts, i.e. if* $f(x) = r(x) + iq(x)$ *we call it integrable if the real functions* r, q *are and just set* $\int f d\lambda = \int r d\lambda + i\int g d\lambda$*.*

Whereas it is natural to define the Lebesgue integral via simple functions, as countable collections go well with the measure-theoretic framework, it is technically convenient to observe that one can actually work with simple functions that are given by just finite sums.

Lemma 2.25. *Let us call a simple functions* f *simple and finite, if it can be written as* $f(x) = \sum_{i=1}^n c_i 1_{E_i}(x)$ *for some finite disjoint sets* $E_1, \dots, E_n$ *and some real numbers* $c_1, \dots, c_n$*. Then we have that*

$$\sup\left\{\int_{\mathbb{R}^n} g(x)\,\lambda(dx) \;\middle|\; 0 \le g \le f,\ g \text{ simple}\right\} = \sup\left\{\int_{\mathbb{R}^n} g(x)\,\lambda(dx) \;\middle|\; 0 \le g \le f,\ g \text{ simple and finite}\right\}$$

and in particular one can equivalently define the Lebesgue integral by just considering simple functions that are given by finite sums.

Proof. It is clear that the LHS is larger than the RHS. So it just remains to show that RHS is at least as big as the LHS. To do this notice that for any non-negative integrable simple function $g(x) = \sum_{i\ge1} c_i 1_{E_i}(x)$, i.e. a function for which $\sum_{i\ge1} c_i\lambda(E_i) < \infty$, we can associate a simple finite function $g_\epsilon(x) = \sum_{i=1}^{N_\epsilon} c_i 1_{E_i}$, where N_ϵ is chosen such that $\sum_{i>N_\epsilon} c_i\lambda(E_i) < \epsilon$. By definition this guarantees that

$$|\int_{\mathbb{R}^n} g(x)\lambda(dx) - \int_{\mathbb{R}^n} g_\epsilon(x)\lambda(dx)| < \epsilon.$$

Now denote by S_L the supremum on the LHS and by S_R the supremum on the RHS of the equality in the lemma. By definition we can choose g such that $\int_{\mathbb{R}^n} g(x)\lambda(dx) \ge S_L - \epsilon$. But then by construction $\int_{\mathbb{R}^n} g_\epsilon(x)\lambda(dx) \ge S_L - 2\epsilon$ and by taking $\epsilon \to 0$ we see that RHS is also at least as large as the LHS.

We conclude that two suprema agree, and the integral could equivalently have been defined using only finite simple functions. □

2.3.4 Basic properties of the Lebesgue integral

We begin by examining some basic properties of the Lebesgue integral. Notice that several of these natural properties do not hold for the Riemann integral!

Proposition 2.26 (Basic properties of the Lebesgue integral). *Let $f : \mathbb{R}^n \to \mathbb{R}$ be measurable. Then*

(1) if $f \geq 0$ and f is integrable, then $\int f d\lambda \geq 0$
(2) if $|f(x)| \leq C$ for all $x \in \mathbb{R}^n$, then it is integrable over any finite box $[a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$
(3) if $\lambda(f \neq 0) := \lambda(\{x : f(x) \neq 0\}) = 0$, then f is integrable and $\int f d\lambda = 0$
(4) if $f \geq 0$ and $\int f d\lambda = 0$, then $\lambda(f \neq 0) = 0$.

Notice that even property 2 is rather interesting: it somehow says that for a measurable function only unboundedness can prevent it from being integrable! We have separated the question of "regularity" of the function (carried by the notion of measurability) from that of its size (which governs integrability).

Proof. The first property comes directly from the definition. The others are on the exercise sheet.

□

In particular, applying property (3) to the difference $f - g$ of two measurable f, g we should intuitively obtain:

Corollary 2.27. *Let f, g be two measurable functions such that $\lambda(f \neq g) := \lambda(\{x : f(x) \neq g(x)\}) = 0$. Then f is integrable iff g is integrable and $\int f d\lambda = \int g d\lambda$.*

However, writing down the proof we will see that we are still lacking a tool to show this nicely [6]

Proof. Define $h = g - f$. Then h is measurable and $\lambda(h \neq 0) = 0$. Thus by the proposition h is integrable and $\int h d\lambda = 0$.

We would like to know say that $g = f + h$ and it is integrable because f, h are and further

$$\int g d\lambda = \int (f + h) d\lambda = \int f d\lambda + \int h d\lambda.$$

But here we are already using the linearity of the Lebesgue integral, something we still need to prove and that is stated in the next proposition. □

Proposition 2.28. *Let f, g be Lebesgue integrable functions. We have the following linearity statement. For $a, b \in \mathbb{R}$, then $af + bg$ is integrable and*

$$\int (af + bg) d\lambda = a \int f d\lambda + b \int g d\lambda.$$

It comes out this linearity is not as straightforward to prove as one hopes!

Indeed, it is straight-forward to check that for finite simple functions f, g it holds that $\int (f + g) d\lambda = \int f d\lambda + \int g d\lambda$. To see this set $f = \sum_{i=1}^{n} c_i 1_{E_i}$ and $g = \sum_{i=1}^{m} d_j 1_{F_j}$, then

$$g + f = \sum_{i=0\ldots n} \sum_{j=0\ldots m} (c_i + d_j) 1_{E_i \cap F_j},$$

where we define $c_0 = d_0 := 0$ and $E_0 := \mathbb{R}^n \setminus \cup_{i=1}^{n} E_i$ and $F_0 = \mathbb{R}^n \setminus \cup_{j=1}^{m} F_j$. Here we have introduced c_0, d_0 because the function f takes value 0 in the complement of $\cup_{i=1}^{N} E_i$ and g takes value 0 in the complement of $\cup_{j=1}^{M} F_j$.

[6] One can show this particular case also by hand, but it is a bit tiring and not worthwhile.

It can be then checked from the definition that $\int (f+g)d\lambda = \int f d\lambda + \int g d\lambda$ for these finite simple functions (on the exercise sheet).

It is also easy to see that for non-negative measurable f, g we have that $\int (f+g)d\lambda \geq \int f d\lambda + \int g d\lambda$. Indeed, whenever $h \geq 0, j \geq 0$ are simple finite functions bounded from above by f, g respectively, the function $k := h + j$ is a simple finite function bounded from above by $f + g$. Thus

$$\sup \left\{ \int_{\mathbb{R}^n} k(x)\, \lambda(dx) \;\middle|\; 0 \leq k \leq f+g,\, k \text{ simple and finite} \right\}$$

is larger than the sum of

$$\sup \left\{ \int_{\mathbb{R}^n} h(x)\, \lambda(dx) \;\middle|\; 0 \leq h \leq f,\, h \text{ simple and finite} \right\}$$

and

$$\sup \left\{ \int_{\mathbb{R}^n} j(x)\, \lambda(dx) \;\middle|\; 0 \leq j \leq g,\, j \text{ simple and finite} \right\}$$

The other inequality, however, requires a few tools. The issue is the following:

- given a simple function k below $f + g$, it is not straight-forward to construct two simple functions, h below f, and j below g with $h + j = k$.

So instead of attempting a direct construction, we will take a detour through some general theorems that will allow us to prove linearity rigorously. In short we will show that under certain conditions, when $f_n \to f$ pointwise, we have $\int f_n d\lambda \to \int f d\lambda$.

More precisely we will use one simple ingredient and one more serious one. The simple ingredient follows directly from the definition of the integral.

Lemma 2.29. *Let $0 \leq g \leq f$ be measurable. Then if f is integrable, so is g and moreover $\int f d\lambda \geq \int g d\lambda$.*

The more substantial theorem is about approximating the integral of f via integrals of approximations f_n.

Theorem 2.30 (Monotone convergence theorem)**.** *Let $0 \leq f_1 \leq f_2 \leq \dots$ be a sequence of integrable functions converging pointwise to some $f = \lim_{n\to\infty} f_n$. Then f is integrable if $\lim_{n\to\infty} \int f_n d\lambda < \infty$ and in this case*

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

Before proving this, let us see how linearity follows.

Proof of Proposition 2.28. Let us just prove the more interesting case: that $\int (f+g)d\lambda = \int f d\lambda + \int g d\lambda$. We saw that this linearity holds for simple functions on the exercise sheet. We will now first show using Monotone convergence theorem that it holds for all integrable non-negative functions, and then use the decomposition into positive and negative parts to argue for the general case.

Pick a sequence of simple functions $f_n, g_n \geq 0$ with $f_n \leq f_{n+1}$ and $g_n \leq g_{n+1}$ and $f_n \to f$, $g_n \to g$ pointwise from below. Then also $f_n + g_n \to f + g$ pointwise from below. On the one hand by the linearity of simple functions for all $n \geq 1$

$$\int (f_n + g_n) d\lambda = \int f_n d\lambda + \int g_n d\lambda.$$

On the other hand by the Monotone convergence theorem

$$\int f_n d\lambda \to \int f d\lambda \; ; \; \int g_n d\lambda \to \int g d\lambda.$$

In particular this means that

$$\lim_{n\to\infty} \int (f_n + g_n) d\lambda = \lim_{n\to\infty} \left(\int f_n d\lambda + \int g_n d\lambda \right) = \int f d\lambda + \int g d\lambda$$

is finite and hence by Monotone convergence theorem $f + g$ is integrable and

$$\lim_{n\to\infty} \int (f_n + g_n) d\lambda = \int (f + g) d\lambda.$$

So we obtain that

$$\int (f + g) d\lambda = \int f d\lambda + \int g d\lambda,$$

as desired.

For general integrable f, g let us write $f = f_+ - f_-$, $g = g_+ - g_-$ and $f + g = (f + g)_+ - (f + g)_-$ and recall that all these positive and negative parts are non-negative measurable functions.

Now, notice that $(f+g)_+ \leq f_+ + g_+$ and $(f+g)_- \leq f_- + g_-$ and thus if f, g are integrable then so are $(f + g)_+, (f + g)_-$ by Lemma 2.29 and hence also $f + g$.

Now we can rewrite the pointwise equality

$$(f + g)_+ - (f + g)_- = f_+ - f_- + g_+ - f_-,$$

as

$$(f + g)_+ + f_- + g_- = (f - g)_- + f_+ + g_+.$$

But to this we can apply the first part of the proof on both sides to conclude that

$$\int (f + g)_+ d\lambda + \int f_- d\lambda + \int g_- d\lambda = \int (f + g)_- d\lambda + \int f_+ d\lambda + \int g_+ d\lambda.$$

It now remains to recombine the terms and to use the definition of the integral to see that

$$\int (f + g) d\lambda = \int f d\lambda + \int g d\lambda$$

as desired.

□

2.3.5 Convergence theorems

Let us now look more closely at the statement in the style

- if $f_n \to f$ pointwise, then $\int f_n d\lambda \to \int f d\lambda$.

We will first see some counterexamples, then prove the Monotone convergence theorem and a few other useful convergence results.

The first failure could be that the limiting f is not integrable. Recall that the pointwise limit of measurable functions is measurable. Thus we at least know that f is regular enough to be potentially integrable. It could fail to be integrable because of "size":

Example 2.31. *Consider the functions* $(f^n)_{n\geq 1}$ *defined on* $\mathbb{R}$ *by* $f^n(x) = 1_{[0,n]} - 1_{[-n,0]}$. *These functions are all finite simple functions and their integral is equal to* 0. *But notice that their pointwise limit* $f = 1_{[0,\infty)} - 1_{(-\infty,0]}$ *is measurable but not integrable.*

But even if the limiting function is integrable, it's integral is not necessarily equal to the limit of integrals.

Example 2.32. *Consider the functions* $(f^n)_{n\geq 1}$ *defined on* $\mathbb{R}$ *by* $f^n(x) = n1_{(0,1/n)}$. *They are finite simple functions and satisfy* $\int f^n d\lambda = 1$ *by definition. But notice that* $f^n(x)$ *converge to the constant* 0 *function pointwise, as for every* $x \in \mathbb{R}$, *there is some* $n_x \in \mathbb{N}$ *such that* $f^n(x) = 0$ *for all* $n \geq n_x$. *But the integral of the constant* 0 *function is just* 0 *and thus the integrals of* f^n *do not converge to the integral of their pointwise limit.*

In this example the functions concentrate the mass on a smaller and smaller region, keeping area under the graph equal to 1. Eventually this tiny vertical box somehow moves out of the interval $(0, 1)$ and disappears. But suppose we ask all of the f_n to be bounded?

Example 2.33. *Consider the functions* $(f^n)_{n\geq 1}$ *defined on* $\mathbb{R}$ *by* $f^n(x) = n^{-1}1_{[0,n}$. *Again they are measurable and bounded, thus integrable with* $\int f^n d\lambda = 1$. *Also* $f^n(x)$ *converge to the constant* 0 *function pointwise too, as for every* $x \in \mathbb{R}$, *there is some* $n_x \in \mathbb{N}$ *such that* $f^n(x) = 0$ *for all* $n \geq n_x$. *But the integral of the constant* 0 *function is just* 0 *and thus the integrals of* f^n *do not converge to the integral of their pointwise limit either.*

In this case the area is kept constant by keeping the box horisontally. But suppose, $\lambda(f_n \neq 0) < C$ for some constant C?

Example 2.34. *Consider the functions* $(f^n)_{n\geq 1}$ *defined on* $\mathbb{R}$ *by* $f^n(x) = 1_{[n,n+1)}$. *Again they are measurable and bounded, thus integrable with* $\int f^n d\lambda = 1$. *Also* $f^n(x)$ *converge to the constant* 0 *function pointwise too, as for every* $x \in \mathbb{R}$, *there is some* $n_x \in \mathbb{N}$ *such that* $f^n(x) = 0$ *for all* $n \geq n_x$. *But the integral of the constant* 0 *function is just* 0 *and thus the integrals of* f^n *do not converge to the integral of their pointwise limit either.*

Now the functions remain bounded but all of the mass moves away to infinity. In some sense the these are the counterexamples to keep in mind and the conditions given in the Monotone convergence theorem rule those cases out. Recall the statement:

Theorem 2.35 (Monotone convergence theorem)**.** *Let* $0 \leq f_1 \leq f_2 \leq \dots$ *be a sequence of integrable functions converging pointwise to some* $f = \lim_{n\to\infty} f_n$. *Then* f *is integrable if* $\lim_{n\to\infty} \int f_n d\lambda < \infty$ *and in this case*

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

And here is the proof.

Proof of the Monotone convergence theorem, Theorem 2.35. First, as $0 \leq f_n \leq f$, it is clear that

$$\lim_{n\to\infty} \int f_n d\lambda \leq \int f d\lambda.$$

The question is why does the other inequality (and thus also the claim on integrability) hold. So we can now assume $\int f d\lambda < \infty$.

Let us start with the case where the limit $f = \lim_{n\to\infty} f_n$ itself is a simple finite function. In particular it can be represented by $f = \sum_{i=1}^m c_i 1_{E_i}$ for some disjoint Borel sets E_i and distinct $c_i > 0$ and its integral equals $\int f d\lambda = \sum_{i=1}^m c_i \lambda(E_i)$. As this is finite by assumption, we have that $\lambda(E_i) < \infty$ for every $i = 1 \dots m$.

For each $\epsilon > 0$ and $n \geq 1$ we can then define the sets $F_n := \{x : f_n(x) \geq (1-\epsilon) f(x)\}$ and further $F_{n,i} = F_n \cap E_i$. These sets are increasing and pointwise convergence of f_n to f guarantees that $\cup_{n\geq 1} F_{n,i} = E_i$ for all $i = 1 \dots m$. But then the properties of the measure λ imply that $\lambda(F_{n,i}) \to \lambda(E_i)$; here we use that $\lambda(E_i) < \infty$. But now by definition $f_n(x) \geq (1-\epsilon) \sum_{i=1}^m c_i 1_{F_{n,i}}$ and thus by Lemma 2.29

$$\int f_n d\lambda \geq (1-\epsilon) \sum_{i=1}^m \lambda(F_{n,i})$$

and we conclude that

$$\lim_{n\to\infty} \int f_n d\lambda \geq (1-\epsilon) \int f d\lambda.$$

As ϵ was arbitrary the claim follows for simple limiting functions f.

But now for general f we can pick any simple function $g \leq f$ and conclude similarly that

$$\lim_{n\to\infty} \int f_n d\lambda \geq \int g d\lambda.$$

As this holds for any simple function g that is bounded above by f, we conclude the claims of the theorem from the definition of the integral: if $\lim_{n\to\infty} \int f_n d\lambda < \infty$, then f is integrable and its integral equals that limit. □