Class notes

1 Fourier Series

Two landmark discoveries are typically credited in the development of analysis: Calculus (circ. 1665) and Fourier series, introduced by Joseph Fourier (1822). The latter mark the passage from finite-dimensional to infinite-dimensional mathematics.

A choice of an orthonormal basis $\{e_k\}_{k=1,\dots,n}$ in $\mathbb{R}^n$ or $\mathbb{C}^n$ allows for a representation of vectors as strings of scalar components $x = (x_k)_{k=1,\dots,n}$ and the inner product $\langle x, y\rangle$ as $\sum_{k=1}^{n} x_k y_k$ in $\mathbb{R}^n$ and $\sum_{k=1}^{n} x_k \overline{y}_k$ in $\mathbb{C}^n$ where $\overline{a+ib} = a - ib$. The natural generalization of the inner product and of the norm in the "continuum limit", for two, say continuous, functions $f, g : [a, b] \to \mathbb{C}$ are

$$\langle f, g\rangle = \int_a^b f(t)\overline{g(t)}dt; \quad \|f\|_2^2 = \langle f, f\rangle$$

which are the Hilbert inner product and the Hilbert norm.

With this generalization we may wonder, for a given orthonormal basis $(e_k)_{k\in\mathbb{N}}$, which functions can be represented by their, now infinite, set of components $(f_k)_{k\in\mathbb{N}}$ where $f_k = \langle f, e_k\rangle$, $k \in \mathbb{N}$ (sometimes $\mathbb{Z}$ is a better choice than $\mathbb{N}$). A possible choice of a basis are the monomials $\left(x^{k-1}\right)_{k\in\mathbb{N}}$ which can be recombined to become an orthonormal set. If $[a, b] = [-1, 1]$, then the $e_k s$ are the Legendre polynomials $(P_k)_{k+1\in\mathbb{N}}$:

$$P_0(x) = 1,\ P_1(x) = x,\ P_2(x) = \frac{1}{2}\left(3x^2 - 1\right),\ P_3(x) = \frac{1}{2}\left(5x^3 - 3x\right),\ P_4(x) = \frac{1}{8}\left(35x^4 - 30x^2 + 3\right), \cdots$$

and in general,

$$P_n(x) = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}^2 (x-1)^{n-k}(x+1)^k$$

These polynomials satisfy the orthogonality condition

$$\int_{-1}^{1} P_m(x)P_n(x)\,\mathrm{d}x = \frac{2}{2n+1}\delta_{mn}$$

In infinite dimensions the question of which functions can be written as $\sum_{k=0}^{\infty} c_k P_k$ is, in this naive formulation, **not** well posed. By an infinite sum we must mean some form of a limit. This could be a uniform limit, a pointwise limit, a limit in the sense of Cèsaro averages, or, with $\mathbb{C}^n$ in mind, a limit in the "distance" sense, i.e. in the sense of integral means of order two:

$$\lim_{N\to\infty} \|f - \sum_{k=0}^{N} f_k P_k\|_2 = 0$$

Each of these definitions leads to quite different answers, as we shall see in due course. The last one can only be satisfactorily answered after replacing Riemann integrals with the much more general and well-behaved Lebesgue integration, which in turn requires measure theory that we will study in Chapter 2.

You probably noted that in $\mathbb{C}^n$ a good choice of the basis often simplifies the analysis. This is even more so in infinite dimensional (Hilbert) spaces. A very important orthonormal set (in the Hilbert space L^2) on $[0,1]$ is $\left(e^{2\pi ikx}\right)_{k\in\mathbb{Z}}$; (finite) linear combinations of $e^{2\pi ikx}$ are called trig polynomials. Series of the form

$$\sum_{k\in\mathbb{Z}} f_k e^{2\pi ikx}$$

are called Fourier series.

Exercise: Check orthonormality of this set.

If f is represented convergently by such a series and if the series converges in a suitable sense, then $f' = \sum_{k\in\mathbb{Z}} f_k(2\pi ik)e^{2\pi ikx}$. In other words, in the "basis"

$$(e_k)_{k\in\mathbb{Z}} = \left(e^{2\pi ikx}\right)_{k\in\mathbb{Z}}$$

if a function is represented by the sequence of coefficients $(f_k)_k$, then its derivative is represented by the sequence $(2\pi ikf_k)_k$. Differentiation is transformed into multiplication, and hence a differential equation $P(d/dx)f = g$ (where P is some polynomial) becomes an algebraic equation, $P(2\pi ik)f_k = g_k$. This property makes Fourier series particularly useful, if not even crucial, in the analysis of differential, partial differential or difference equations. Indeed, it was the discovery of Fourier that they provide the general solution of the heat equation (a very imprecise statement at that time), solution that was previously unknown, that triggered many important developments in modern analysis. One needed to understand in what sense are these series convergent, to which functions, and in what sense the solution is the most general one.

Note that, because of orthonormality, again assuming suitable convergence, we have

$$\langle \sum_{k\in\mathbb{Z}} f_k e^{2\pi ikx}, e^{2\pi ijx}\rangle = f_j$$

which leads to the definition of the Fourier coefficients

$$f_k := \langle f, e^{2\pi ikx}\rangle = \int_0^1 f(t)e^{-2\pi ikt}dt$$

Note that if we aim at a good form of pointwise convergence the represented function should have the property $f(0) = f(1)$, and that, if convergence is uniform then $f \in C(\mathbb{T})$, the continuous functions on the torus $\mathbb{T}$, which in one dimension is S^1.

To study convergence of Fourier series, note that

$$\sum_{-N}^{N} f_k e^{2\pi i k x} = \int_0^1 f(s) D_N(x-s) ds = \int_0^1 f(x-s) D_N(s) ds = D_N * f \tag{1}$$

where D_N is the *Dirichlet kernel*,

$$D_n(x) = \sum_{k=-n}^{n} e^{2\pi i k x} = \frac{\sin(2n+1)\pi x}{\sin \pi x} \quad (x \in \mathbb{C} \setminus \mathbb{Z})$$

Exercise: Prove the identity above. One way is to factor out $e^{-2\pi i n x}$ and note that the remaining sum is a geometric progression.

Remark 1. *We have $e_0(x) = 1$ and thus $\int_0^1 e_0(s)ds = 1$. For any $k \neq 0$ however, $\int_0^1 e_k(s)ds = 0$. Thus, for all $\mathbb{Z} \ni n \geqslant 0$,*

$$\int_0^1 D_n(s)ds = 1$$

Remark 2. *Note that for any $f \in C(\mathbb{T})$ and any $a \in \mathbb{R}$, we have*

$$\int_0^1 f(s)ds = \int_a^{1+a} f(s)ds$$

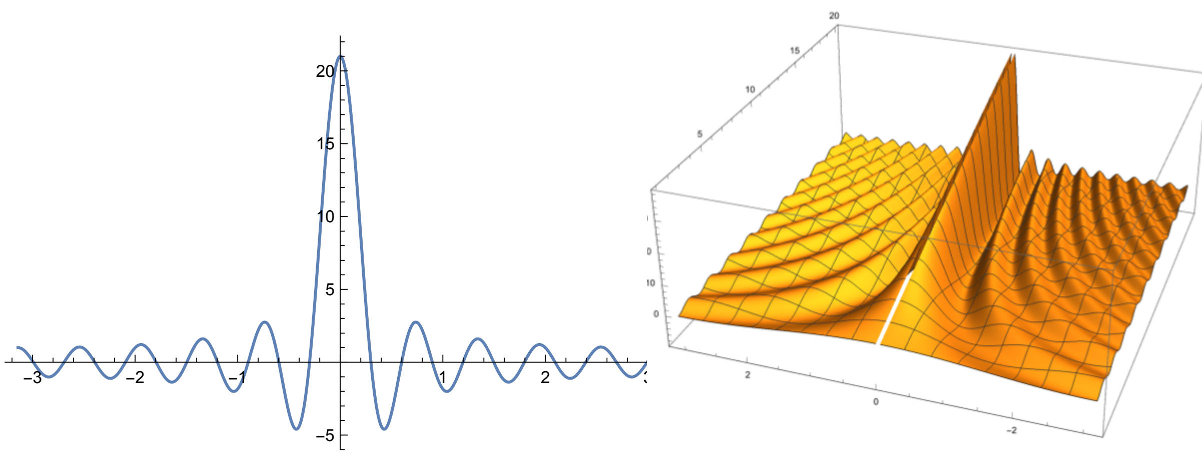

Figure 1: The Dirichlet kernel for $n = 10$, D_{10} (left) D_n on I for $n = 1, ..., 20$. The peak grows like n, with a width $1/n$ and oscillations of frequency n away from it (right)

Lemma 3. *Let $(a, b) \subset [-\frac{1}{2}, \frac{1}{2})$. Then*

$$\lim_{n\to\infty} \int_a^b D_n(x)dx = \begin{cases} 0 \text{ if } 0 \notin [a,b] \\ 1 \text{ if } 0 \in (a,b) \\ \frac{1}{2} \text{ if } 0 \in \{a,b\} \end{cases} \tag{2}$$

If $0 \notin [A, b]$, then the limit is uniform with respect to $a \in [A, b]$.

Proof. First, from the definition it follows that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} D_n(s)ds = 1; \quad \text{and, since } D_n \text{ is even, } \int_0^{\frac{1}{2}} D_n(s)ds = \tfrac{1}{2} \tag{3}$$

Assume now $0 \notin [a, b]$; by integration by parts,

$$\begin{aligned} &\left| \int_a^b \frac{\sin((2n+1)\pi s)}{\sin \pi s} ds \right| \\ &\quad = \left| -\frac{\cos(\pi b(2n+1))}{\sin(\pi b)\pi(2n+1)} + \frac{\cos(\pi a(2n+1))}{\sin(\pi a)\pi(2n+1)} - \int_a^b \frac{\cos(\pi s)\cos((2n+1)\pi s)}{\sin^2(\pi s)(2n+1)} ds \right| \\ &\qquad\qquad \leqslant \frac{4}{\pi(2n+1)} \left(|a|^{-1} + |b|^{-1} \right) \end{aligned} \tag{4}$$

where we used the fact that $|\sin \pi x| \geqslant x/2$ for x in $[-\frac{1}{2}, \frac{1}{2}]$ (justify this!) and

$$\int \frac{1}{\sin^2(\pi s)} ds = -\frac{1}{\pi} \cot(\pi s) + C$$

Combining with (3), the result follows. □

The local behavior of the Dirichlet kernel, the Lemma above and (1) *might suggest* the conjecture that for any continuous function f $D_N * f$ converges to f as $N \to \infty$, in turn entailing that the Fourier series of a continuous function converges to the function itself. One may indeed be tempted to think of taking a fine enough partition of $[-\frac{1}{2}, \frac{1}{2}]$, so that f is "basically constant" on each subinterval, and apply Lemma 3 to derive that the only nonvanishing contribution in (1) comes from the interval around $s = 0$ which "converges to $f(x)$". Not only is this argument **wrong**, but the whole **conjecture is wrong**. However, this fact has only been discovered towards the end of the 19^{th} century, and it came as a surprise. To understand what the "correct results" really are necessitated an integration theory better than Riemann's and many other modern developments in analysis leading to a final answer, a deep result whose proof is very difficult of Lennart Carleson in 1966. In the subsequent sections we will clarify these issues (except for proving Carleson's theorem!) while developing appropriate mathematical tools, the tools of mathematical analysis.

Exercise 1. *(a) Show that $|D_n|$ is bounded by $2n+1$ for any nonnegative integer n, by using the expression of D_n as a sum (or the fact that $|\sin x| \leqslant |x|$ for all $x \in \mathbb{R}$). Let $x_j = \dfrac{(2j+1)}{2(2n+1)}$. Show that there are positive constants $c_1, c_2 > 0$ s.t. $|D_n(x)| > c_1 n j^{-1}$ on each interval $\{x : |x - x_j| < c_2 n^{-1}\}$ and all integers j with $0 < |j| < n/2$. Show that this implies that*

$$\lim_{n\to\infty} \int_0^1 |D_n(s)|ds = \lim_{n\to\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} |D_n(s)|ds = \infty$$

(b) Show that for any x there is a sequence of functions $(f_n)_{n\in\mathbb{N}}$ in $C(\mathbb{T})$ such that $\sup_{s\in\mathbb{T}, n\in\mathbb{N}} |f_n(s)| =$

1 *and*

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} D_n(s) f_n(x-s)ds \to \infty \quad as \quad n \to \infty$$

This and the uniform boundedness theorem that we'll see later implies that there are continuous functions for which the Fourier series diverge at least one point.

Theorem 4. *Assume f and f' are in $C(\mathbb{T})$. Then the symmetric Fourier sums of f (the first sum in (1)) converges in the uniform norm to f. If f is only piecewise continuously differentiable, with bounded derivative, then the sums converge in uniform norm in any compact set that does not contain a discontinuity. At points of discontinuity, if the lateral limits of the function exist, then the symmetric sums converge to the half sum of these lateral limits.*

Note 5. *Note that we do not claim absolute convergence which* **cannot** *hold if f is discontinuous (why?)*

Proof. Let f_k as usual be the Fourier coefficients of f. As we will see, we can reduce the analysis of that of a function with at most one exceptional point where the lateral limits exist. If the function is smooth everywhere, let ξ be any point; otherwise choose ξ to be the discontinuity point. Call the left (right) limit of f at ξ $f(\xi^-)$ ($f(\xi^+)$ resp.). We seek to see whether the Fourier sums of f converge to a limit, call it L. We have, by integration by parts, and Lemma 3

$$\sum_{k=-n}^{n} f_k e^{2\pi i k\xi} - L = \int_{-\frac{1}{2}}^{\frac{1}{2}} (f(\xi-s)-L)D_n(s)ds = \int_{-\frac{1}{2}}^{0} (f(\xi-s)-L)D_n(s)ds + \int_{0}^{\frac{1}{2}} (f(\xi-s)-L)D_n(s)ds$$

$$= \frac{1}{2}\left(f(\xi^+)+f(\xi^-)-2L\right) + \int_{-\frac{1}{2}}^{0} f'(\xi-s)\int_{-\frac{1}{2}}^{s} D_n(t)dtds + \int_{0}^{\frac{1}{2}} f'(\xi-s)\int_{\frac{1}{2}}^{s} D_n(t)dtds \quad (5)$$

Let's take the second integral; the first one is dealt with similarly. Let $m = \|f'\|_u = \sup_{x\in[-\frac{1}{2},\frac{1}{2}]} |f'(x)|$. For a small $\epsilon > 0$ we write

$$\int_{0}^{\frac{1}{2}} f'(\xi-s)\int_{\frac{1}{2}}^{s} D_n(t)dtds = \int_{0}^{\epsilon} f'(\xi-s)\int_{\frac{1}{2}}^{s} D_n(t)dtds + \int_{\epsilon}^{\frac{1}{2}} f'(\xi-s)\int_{\frac{1}{2}}^{s} D_n(t)dtds \quad (6)$$

Now

$$\left|\int_{\epsilon}^{\frac{1}{2}} f'(\xi-s)\int_{\frac{1}{2}}^{s} D_n(t)dtds\right| \leqslant m\int_{\epsilon}^{\frac{1}{2}}\left|\int_{\frac{1}{2}}^{s} D_n(t)dt\right|ds \leqslant m\frac{8}{\pi(2n+1)}\int_{\epsilon}^{\frac{1}{2}}\frac{dt}{t} \leqslant \frac{16m}{(2n+1)\pi}\ln\epsilon^{-1} \quad (7)$$

by Lemma 3 and (4). On the other hand, since $\int_s^{1/2} D_n(t)dt$ is bounded by a constant c independent of n [1]

$$\left|\int_{0}^{\epsilon} f'(\xi-s)\int_{\frac{1}{2}}^{s} D_n(t)dtds\right| \leqslant cm\epsilon$$

Clearly, there is a limit L, namely $L = \frac{1}{2}(f(\xi^-)+f(\xi^+))$. If f is C^1 throughout, $L = f(\xi)$. □

[1] Let $m = 2n+1$. Again using $\sin\pi x \geqslant x/2$, first note that $2\int_0^{\frac{\pi}{m}} \frac{\sin mx}{x} = 2\int_0^{\pi}\frac{\sin x}{x}$. Now, for $x \in [(2j+1)/m, (2j+2)/m), j \geqslant 0$, $\sin mx < 0$ and we have $\frac{\sin mx}{x} \leqslant \sin mx/(2j+2)$; on $[(2j+2)/m,(2j+3)/m), j \geqslant 0$ $\sin mx > 0$ and we have $\frac{\sin mx}{x} \leqslant \sin mx/(2j+2)$ implying $0 < \int_{1/2}^{s} D_n(t)dt < c + O(1/n)$ where $c = 1.0598....$

Exercise 2. *Let f be as in the theorem, and assume it is discontinuous at $\{x_1, ..., x_n\} \subset (-1/2, 1/2)$, where lateral limits exist. Let $F(x) = \int_{-\frac{1}{2}}^{x} f'(s)ds$. Show that F is continuous and piecewise differentiable on $[-1/2, 1/2)$. Show that its periodic extension to the whole of $\mathbb{R}$ has at most one discontinuity per period, at the points $x = \frac{1}{2}+j, j \in \mathbb{Z}$. Show that F has lateral limits everywhere. Thus the proof in the theorem applies to F. Let $\theta(x)$ be the Heaviside function, equal zero for $x < 0$ and one for $x > 0$. Then, a piecewise continuous function f with piecewise continuous derivative and points of discontinuity $\{x_1, ..., x_m\}$ equals $F(x) + \sum_{i=1}^{m} \theta(x - x_i)(f(x_i^+) - f(x_i^-))$. Complete the proof of the theorem by reducing the analysis to the θ function, for which you can apply the approach in the proof of Lemma 3.*

Exercise 3. *We can of course choose a different ϵ for each n. Show that with the choice $\epsilon = n^{-1}$ we get, for large enough n,*

$$\left| L - \sum_{k=-n}^{n} f_k e^{2\pi i k x} \right| = O(n^{-1} \ln n) \tag{8}$$

Exercise 4. *(a) Check the recurrence relation ($n \in \mathbb{N}, |k| \in \mathbb{N}$)*

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} s^n e^{-2i\pi ks}\, ds = \frac{2^{-n-1}(-1)^k(1-(-1)^n)}{\pi k} i - \frac{i\,n}{2k\pi} \int_{-\frac{1}{2}}^{\frac{1}{2}} s^{n-1} e^{-2i\pi ks}\, ds$$

(b) Check that the symmetric Fourier series on the interval $[-\frac{1}{2}, \frac{1}{2}]$ of the monomials $x^k, k = 0, 1, 2, 3$ are (the exponentials were re-expressed as trig functions to simplify the formulas)

$$1 = x^0 \tag{9}$$

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{\pi k} \sin(2\pi kx) = x^1 \quad (\text{if } |x| \neq \tfrac{1}{2}) \tag{10}$$

$$\frac{1}{12} + \frac{1}{\pi^2} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^2} \cos(2\pi kx) = x^2 \tag{11}$$

$$\frac{3}{2\pi^3} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^3} \sin(2\pi kx) = -\frac{1}{4}x + x^3 \tag{12}$$

$$\frac{3}{\pi^4} \sum_{k=1}^{\infty} \frac{(-1)^{1+k} \cos(2\pi kx)}{k^4} = -\frac{1}{2}x^2 + x^4 + \frac{7}{240} \tag{13}$$

Convergence of these series follows from Theorem 4. Note that convergence in (10) is **not** *absolute (why?) (all others are).*

(c) Assume that f is continuously differentiable on $\mathbb{T}$ except for one point x_0 where f is discontinuous. Assume that f and f' have lateral limits at x_0. Mapping $\mathbb{T}$ to $[-1/2, 1/2)$, place the discontinuity of the mapped function (keep the notation f) at the right endpoint. Show that there is an α s.t. $f + \alpha x$ extends to a continuous periodic function on R with piecewise continuous derivative.

(d) Use (11) to show that $\sum_{k\in\mathbb{N}} \frac{1}{k^2} = \frac{\pi^2}{6}$. Rely on the previous parts of this exercise to calculate $\sum_{k\in\mathbb{N}} \frac{1}{k^4}$.

The connections between the behavior of the Fourier coefficients and the regularity (differentiability, Hölder continuity etc.) of a function are also very interesting and important. Here is a starting point:

Theorem 6. *Let $f \in C^n(\mathbb{T})$ (i.e., f is continuous together with its first n derivatives) and let $(f_k)_{k\in\mathbb{Z}}$ be its Fourier coefficients. Then $f_k = O(|k|^{-n})$ as $|k| \to \infty$.*

In the opposite direction, if $|f_k| = O(|k|^{-m})$ for some $m > n+1$ for large $|k|$, then $f \in C^n(\mathbb{T})$ (we'll be able to find stronger statements for this "converse" in due course).

Proof. The proof is by simple integration by parts, n times:

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} f(s)e^{-2\pi iks}ds = (-2\pi ik)^{-n}\int_{-\frac{1}{2}}^{\frac{1}{2}} f^{(n)}(s)e^{-2\pi iks}ds \quad \text{and thus } |f_k| \leqslant (2\pi k)^{-n}\|f^{(n)}\|_u$$

if $|k| \in \mathbb{N}$, where $\|\cdot\|_u$ is the uniform norm. The opposite direction statement is left as an exercise of differentiation of suitably convergent function series. □

Note 7. *This shows that divergence of the Fourier series of general continuous functions is due to their lack of sufficient smoothness.*

Exercise 5. *In the class of continuous functions whose Fourier series converge, the rate of convergence is arbitrarily slow. Consider the lacunary Fourier series*

$$f(x) = \sum_{k=1}^{\infty} k^{-\alpha}\cos(2^k\pi x)$$

where $\alpha > 1$. Show that this series converges absolutely and uniformly and (thus) f is continuous. Show that the Fourier series of f is just the sum in right hand side. Since $|f_m|$ is zero if $|m| \neq 2^{k-1}$ for some $k \in \mathbb{N}$ and equals $\frac{1}{2}(\log_2|2m|)^{-\alpha}$ if $|m| = 2^k$, we see that the Fourier coefficients $|f_m|$ decay slower than any power of m. Adapt this argument to find functions for which the Fourier series converge, but the coefficients have arbitrarily slow decay (and think of some rigorous definition of the concept of "arbitrarily slow"). See Fourier sums of f with $\alpha = 3/2$ and $1, ..., 20$ terms.

Note 8. *The f above is an example of a continuous but nowhere differentiable function. Try your hand in proving this.*

Note 9. *A refinement of the construction above gives Fejér's example of a continuous function whose Fourier sums blow up at $x = 0$. Fejér's function is (in our notation and conventions)*

$$f(x) = \sum_{k=1}^{\infty} \frac{1}{k^2}\sin\left[(2^{k^3}+1)(x-\tfrac{1}{2})\pi\right]$$

The analysis of the convergence vs. divergence of the Fourier sums of f is quite elementary; if you are curious, click on this link: Fejér's counterexample link.

1.1 Fejér's theorem

In various weaker senses, Fourier series of continuous functions do converge to their associated functions. For $f \in C(\mathbb{T})$ and $n \in \mathbb{N}$ let

$$s_n(x) = \sum_{k=-n}^{n} f_k e^{2\pi ikx}$$

and take the Césaro means of s_n,

$$\sigma_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} s_k(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x-t)F_n(t)dt$$

where

$$F_n(x) = n^{-1}\sum_{k=0}^{n-1} D_k(x) = n^{-1}\left(\frac{\sin(n\pi x)}{\sin \pi x}\right)^2$$

(check the explicit expression of F_n)

Theorem 10. *If f is in $C(\mathbb{T})$, then the sequence $(\sigma_n)_{n\in\mathbb{N}}$ of Cesàro means of the sequence $(s_n)_{n\in\mathbb{N}}$ of partial sums of the Fourier series of f converges uniformly to f on $\mathbb{T}$.*

Proof. We first claim that F_n is an *approximation of the identity*, by which it is meant that

1. $F_n \geqslant 0, \forall n \in \mathbb{N}$.
2. $\int_{-\frac{1}{2}}^{\frac{1}{2}} F_n(s)ds = 1, \forall n \in \mathbb{N}$.
3. For any $\delta \in (0, \frac{1}{2}]$, $\lim_{n\to\infty} \int_{|x|\in[\delta,\frac{1}{2}]} F_n(s)ds = 0$.

Indeed, 1. is obvious, 2. is clear from the definition because for any $k \in \mathbb{N}$, $\int_{-\frac{1}{2}}^{\frac{1}{2}} D_k(s)ds = 1$, while 3. follows from the fact that for $|x| > \delta$, since $|\sin x| \geqslant x/2$, we have $F_n(x) \leqslant 4n^{-1}$.

The proof follows from these three basic properties of the Fejér kernel and from the uniform continuity of f. Let $m = \|f\|_u$. We have $m > 0$ unless $f = 0$ in which case the proof is immediate. Note that $f(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)F_n(s)ds$ by 2. above. We now see that

$$(\sigma_n f)(x) - f(x) = \int_{0\leqslant|s|\leqslant\delta} F_n(s)(f(x-s) - f(x))ds + \int_{\delta\leqslant|s|\leqslant\frac{1}{2}} F_n(s)(f(x-s) - f(x))ds$$

Take some $\epsilon > 0$. Using uniform continuity, choose δ so that wenever $|s - s'| \leqslant \delta$ we have $|f(s) - f(s')| < \epsilon/2$, and choose n_0 s.t. for all $n \geqslant n_0$ we have $\int_{\frac{1}{2}\geqslant|y|>\delta} F_n(s)ds \leqslant \frac{1}{8}\epsilon m^{-1}$. With this, we see that for all $n \geqslant n_0$ and all x

$$|(\sigma_n f)(x) - f(x)| < \epsilon$$

□

Corollary 11. *If f and g are continuous and have the same Fourier coefficients, then $f = g$.*

Proof. The Césaro sums of the Fourier series of f converge to f, and also to g. □

Corollary 12. *Trigonometric polynomials are dense in $C(\mathbb{T})$.*

Proof. This follows immediately from Theorem 10: let $f \in C(\mathbb{T})$ and $\epsilon > 0$ be arbitrary; let n_0 be s.t. $\|f - \sigma_{n_0} f\|_u \leqslant \epsilon$; note that $\sigma_{n_0} f$ is a trig polynomial. □

Note 13. *This density does* **not** *imply that the Fourier sums of continuous functions converge. Make sure you understand the distinction.*

An important consequence of these results is Weyl's equidistribution theorem. A sequence of real numbers $(x_j)_{j\in\mathbb{N}}$ is equidistributed modulo one if, by definition, for any $f \in C(\mathbb{T})$ we have

$$\frac{1}{n}\sum_{k=1}^{n} f(x_j) \to \int_0^1 f(x)dx \tag{14}$$

Note that this also means that, in the sense of Césaro means, $(f(x_j))_{j\in\mathbb{N}}$ converges to the integral of f.

Exercise 6. *Let $(x_j)_{j\in\mathbb{N}}$ be equidistributed mod 1 and let* $\operatorname{frac} y$ *denote the fractional part of y. Show that the points $\{\operatorname{frac} x_j : j \in \mathbb{N}\}$ are dense in $(0,1)$.*

Theorem 14 (Weyl)**.** *Let x_0 and α be real numbers. Then, the sequence $(x_0 + k\alpha)$ is equidistributed mod 1 iff α is irrational.*

Note 15 (Rotation on the circle)**.** *We can visualize the points x_k above as the points on the unit circle obtained by starting at an angle $x_0 \mod 2\pi$ and successively rotating by an angle $\alpha \mod 2\pi$. See Irrational rotation animation.*

Exercise 7. *Show that the sequence $(x_0 + k\alpha)$ is equidistributed mod one* **iff** *the empirical probability of finding a point in any arc-interval on the circle (in the sense of the Note) approaches the arclength mod 2π as the number of rotations increases without bound. We recall that the empirical probability is the ratio between the number of favorable events divided by the total number of events. The term "equidistributed" is suggested by this interpretation.*

Proof of Theorem 14. We leave it as an easy exercise to show that irrationality of α is necessary. Verify that irrationality is sufficient for Césaro-convergence to the integral of f for all trig monomials $f(x) = e^{2\pi ikx}, k \in \mathbb{Z}$, and thus for all trig polynomials. Use the density of trig polynomials to complete the proof.

□

Exercise 8. *Check that* (14) *extends to piecewise continuous functions. Monotone bounded functions are Riemann integrable. Does* (14) *extend to them?*

1.2 Introduction to normed spaces and Hilbert spaces

In the following, F is the field of scalars, and it is either $\mathbb{R}$ or $\mathbb{C}$. Complex conjugation is denoted by overline, as usual.

Definition 16. *An inner product space is a vector space $\mathcal{V}$ over the field F together with an inner product, i.e., with a map*

$$\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \to F$$

which satisfies the following axioms: for all vectors $x, y, z \in \mathcal{V}$ we have

1. *Conjugate symmetry:* $\langle x, y \rangle = \overline{\langle y, x \rangle}$

2. *Linearity in the first argument:*

$$\langle ax, y\rangle = a\langle x, y\rangle$$
$$\langle x+y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$$

3. *Positive-definiteness:*

$$\langle x, x\rangle \geq 0$$
$$\langle x, x\rangle = 0 \Leftrightarrow x = 0\,.$$

Note 17. We write $\|x\|^2 = \langle x, x\rangle$; $\|\cdot\|$ is then a norm.

The map $\langle\cdot,\cdot\rangle : \mathcal{V}\times\mathcal{V}$ is a positive definite *sesquilinear form*, in this case a map which is linear in the first variable and *conjugate-linear* in the second [2]. In some constructions it is convenient to allow, more generally, *semi-definite* sesquilinear forms, ones that have degenerate kernel, that is $\|x\| = 0$ for some nonzero vectors. Such forms are also called weak inner products.

Theorem 18 (Cauchy-Schwarz)**.** *Let $\mathcal{V}$ be an inner product space and x, y be any two elements of $\mathcal{V}$. We have $|\langle x, y\rangle| \leqslant \|x\|\|y\|$.*

Proof. Note for any $a \in \mathbb{C}$ we have

$$0 \leqslant \|x - ay\|^2 = \langle x, x\rangle + |a|^2\langle y, y\rangle - \langle x, ay\rangle - \langle ay, x\rangle = \langle x, x\rangle + |a|^2\langle y, y\rangle - 2\Re(a\langle x, y\rangle)$$

Write the polar decomposition $\langle x, y\rangle = |\langle x, y\rangle|e^{i\alpha}$ (if $\langle x, y\rangle = 0$ any α works). By replacing a by $|a|e^{-i\alpha}$ we see that $f(|a|) = \langle x, x\rangle + |a|^2\langle y, y\rangle - 2|a||\langle x, y\rangle| \geqslant 0$. The trick is now to note that $f(|a|)$ is a quadratic polynomial in $|a|$ which is nonnegative, and thus it has nonpositive discriminant: $4|\langle x, y\rangle|^2 - 4\langle x, x\rangle\langle y, y\rangle \leqslant 0$, which is what we intended to prove. □

$\mathbb{R}^n$ and $\mathbb{C}^n$ with the usual dot products are clearly inner product spaces. Define now

$$\ell^2(\mathbb{N}) = \left\{ x = (x_i)_{i\in\mathbb{N}} \,\middle|\, \sum_{i\in\mathbb{N}} |x_i|^2 < \infty \right\};\ \ \ell^2(\mathbb{Z}) = \left\{ x = (x_i)_{i\in\mathbb{Z}} \,\middle|\, \sum_{i\in\mathbb{Z}} |x_i|^2 < \infty \right\}$$

These are inner-product space, with the inner product

$$\langle x, y\rangle = \sum_{i\in\mathbb{N}} x_i\overline{y_i} \ \text{ and } \langle x, y\rangle = \sum_{i\in\mathbb{Z}} x_i\overline{y_i}$$

respectively. So is the space

$$L^2_{\mathcal{R}}((a,b)) = \left\{ f : (a,b) \to \mathbb{C} \,\middle|\, f \text{ Riemann integrable, } \int_a^b |f(s)|^2 ds := \|f\|_2^2 < \infty \right\}$$

with the inner product

$$\langle f, g\rangle = \int_a^b f(s)\overline{g(s)}ds$$

[2] In physics the convention is a bit different, the form is conjugate-linear in the first entry and linear in the second. Each convention has its own merits but in the end of course it does not make any real difference which convention we choose.

Note 19. *(a) If the interval is finite, then $L^2_{\mathcal{R}}((a,b))$ is the same as the space of all Riemann integrable functions.*

(b) If (a,b) is a finite interval, then the sup convergence is stronger than $L^2_{\mathcal{R}}$ convergence. Indeed, $\|\cdot\|_2^2 \leqslant \|\cdot\|_u^2(b-a)$. In fact, it is strictly stronger. For instance, the sequence of characteristic functions of any family of intervals of total length $1/n$ converges to zero in $L^2_{\mathcal{R}}$, but not pointwise in general, let alone uniformly.

The conditions for Riemann integrability will prove to be too strong for a number of important purposes, and the remedy is a more general integral, the Lebesgue integral. We see that, if (a,b) is a finite interval, then **any** Riemann integrable function is in $L^2_{\mathcal{R}}((a,b))$; this is an easy exercise.

Using Cauchy-Schwarz we see that the inner product is well defined on $\ell^2(\mathbb{Z})$ and $L^2_{\mathcal{R}}((a,b))$.

In fact $\ell^2(\mathbb{Z})$ and $L^2_{\mathcal{R}}((a,b))$ have interesting connections. Let, for simplicity $a=-1/2, b=1/2$. Note that, if f is Riemann integrable so is $e^{2\pi ikx}f(x)$ and the Fourier coefficients of f

$$f_k=\int_{-1/2}^{1/2} f(s)e^{-2\pi iks}ds,\ k\in\mathbb{Z}$$

are well-defined.

Furthermore,

$$0\leqslant \|f-\sum_{k=-n}^{m} f_k e^{2\pi ikx}\| = \langle f-\sum_{k=-n}^{m} f_k e^{2\pi ikx}, f-\sum_{k=-n}^{m} f_k e^{2\pi ikx}\rangle = \langle f,f\rangle - \sum_{k=-n}^{m}|f_k|^2 \tag{15}$$

and we see that

$$\sum_{k=-n}^{m}|f_k|^2\leqslant \|f\|_2^2$$

and (because of positiveness of the terms) $\sum_{k=-\infty}^{\infty}|f_k|^2$ converges and.

$$\sum_{k=-\infty}^{\infty}|f_k|^2\leqslant \|f\|_2^2 \tag{16}$$

(16) is called *Bessel's inequality*. We have also proved the following.

Proposition 20. *If f is Riemann integrable on $[-1/2,1/2]$, then the sequence of its Fourier coefficients $(f_k)_{k\in\mathbb{Z}}$ is in $\ell^2(\mathbb{Z})$, and $\|(f_k)_{k\in\mathbb{Z}}\|_{\ell^2}\leqslant\|f\|_2$.*

Corollary 21. *If $f\in C(\mathbb{T})$, then*

$$\|\sum_{k=-n}^{m} f_k e^{2\pi ikx}-f\|_2\to 0 \text{ as } n,m\to\infty$$

and

$$\sum_{k=-\infty}^{\infty}|f_k|^2 = \|f\|_2^2$$

Proof. This is straightforward, since trig polynomials are dense in $C(\mathbb{T})$ in the uniform norm, and by Note 19 a fortiori in $L^2_{\mathcal{R}}$, and the properties above trivially hold for trig polynomials. $\square$

Exercise 9. *Show that continuous functions are dense in the space of Riemann integrable functions in the sense of $L^2_{\mathcal{R}}((a,b))$.*

It follows that

Corollary 22. *If f is Riemann integrable on $[-1/2, 1/2]$ then*

$$\|\sum_{k=-n}^{m} f_k e^{2\pi i k x} - f\|_2 \to 0 \text{ as } n, m \to \infty$$

and

$$\sum_{k=-\infty}^{\infty} |f_k|^2 = \|f\|_2^2$$

Definition 23. *A sequence $(s_k)_{k\in\mathbb{N}}$ in a normed space is Cauchy if for any $\epsilon > 0$ there is an n_0 s.t. for all $n_1 > n_0$ and $n_2 > n_0$ we have*

$$\|s_{n_1} - s_{n_2}\| < \epsilon$$

A normed space in which every Cauchy sequence is convergent is complete.

Proposition 24. *The spaces $\ell^2(\mathbb{N})$ and $\ell^2(\mathbb{Z})$ are complete.*

Proof. We show this for $\ell^2(\mathbb{N})$; the proof for $\ell^2(\mathbb{Z})$ is similar (it even follows from it).

If $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\ell^2(\mathbb{N})$, then for every $i \in \mathbb{N}$ the number sequence $\{(x_n)_i\}_{n\in\mathbb{N}}$ is Cauchy (indeed $|(x_n)_i - (x_m)_i|^2 \leqslant \|x_n - x_m\|^2$). Let $y_i = \lim_n (x_n)_i$. Let n_0 be s.t. $(\forall n, m \geqslant n_0), (\|x_n - x_m\| < 1)$. The triangle inequality implies that $\forall n \geqslant n_0, \|x_n\| \leqslant C$ where $C = 1 + \|x_{n_0}\|$. It follows that, for all n, $\sum_{i=1}^{n} |y_i|^2 = \lim_{k\to\infty} \sum_{i=1}^{n} |(x_k)_i|^2 \leqslant C$ and since $|y_i|$ are positive and the sums are bounded, the sum converges to $\|y\|^2 \leqslant C$, that is $y \in \ell^2(\mathbb{N})$. Similarly, since $\lim_{k\to\infty} \sum_{i=0}^{n} |(x_k)_i - y_i|^2 = 0$ for any n, we can use the triangle inequality to complete the proof. □

Exercise 10. *(a) Show that there is no Riemann integrable function whose Fourier coefficients are $S = (|k|^{-1})_{k\in\mathbb{Z}\setminus\{0\}}$.*

(b) Clearly $S \in \ell^2(\mathbb{Z})$, and S is the limit of Fourier coefficients of trig polynomials. Check that these trig polynomials form a Cauchy sequence in $L^2_{\mathcal{R}}((-1/2, 1/2))$, but it is not convergent in $L^2_{\mathcal{R}}((-1/2, 1/2))$.

(c) Check that the symmetric Fourier sums corresponding to S converge uniformly on any compact set in $(-1/2, 0) \cup (0, 1/2)$.

(Bonus, 3p) In fact the the symmetric Fourier sums in part (c) above converge uniformly, on such compact sets not containing zero to $-2\log(2|\sin(\pi x)|)$.

2 Measure theory

Here is a way to extend Riemann integration enough so that the issues we encountered would be resolved.

To start with, take a finite interval $[a, b] \subset \mathbb{R}$. Define a "norm" on functions that relates to the value of the Riemann integral:

$$\|f\|_1 := \int_a^b |f(s)| ds$$

The problem is that this is only semidefinite: any Riemann integrable function that is nonzero only on a countable set has norm zero. To upgrade a semi-definite-form space to an actual normed space, we mod-out the elements of zero norm, and we end up with a set of equivalence classes $\{[f] : f$ Riemann integrable on $[a,b]\}$, where

$$[f] = \{g : \|f-g\|_1 = 0\} \tag{17}$$

Check that the space of equivalence classes above is a linear space V. $\|\cdot\|_1$ is now a norm on V.

Exercise 11. *If $(a,b) \subset \mathbb{R}, 0 < b-a < \infty$, then $\|f\|_1 \leqslant \|f\|_2$ (where the norms are those of $L^1((a,b))$ and $L^2((a,b))$ resp.). Adapt the example in Exercise 10 to find a sequence of Riemann integrable functions which is Cauchy in $\|\cdot\|$ but does not converge to a Riemann integrable function.*

Note that for any a,b the functionals given on the Riemann integrable functions by

$$\phi_{a,b} f = \int_a^b f(s)ds \tag{18}$$

are bounded w.r.t. $\|\cdot\|_1$. Now define $L^1([a,b])$ to be the completion of V under $\|\cdot\|$; the extension by continuity of the functionals $\phi_{a,b}$ is an integral on L^1. We are left with questions about what exactly we achieved. Can the elements of L^1 be interpreted as classes of equivalence of functions? This is not very straightforward since the characteristic function of an interval of size $1/n$ on $\mathbb{T}$ carried by an irrational rotation will tend in L^1 to zero but pointwise it converges nowhere. What is the equivalence relation? What are the properties of integration? We will return to this approach later.

A more systematic and motivated approach is to start from the geometrical interpretation of the Riemann integral of a nonnegative function: it represents the area under the graph of that function. With this in mind, we ask more generally: which sets can have an area (volume in $\mathbb{R}^3$ etc.), and for those, how do we define an area?

It turns out that not every set can have a volume; call the good sets "measurable". The class of measurable sets however should be closed under intersection, union, and complement. Furthermore, the union of a countable family of disjoint sets should also be measurable, with measure equal to the sum of individual measures. Indeed this is well defined, as a sum of positive terms. The sum could be infinity (thus, we should allow $+\infty$ as a possible volume). Eliminating conditions that follow from each-other we define:

Definition 25. *Let X be any nonempty set. An algebra $\mathcal{A}$ of sets on X is a nonempty collection of subsets of X, closed under finite unions and complements. A σ-algebra on X is an algebra which is closed under countable unions.*

Note 26. *Algebras are closed under finite intersections and σ-algebras are closed under countable intersections, since $\cap_j A_j = (\cup_j A_j^c)^c$. The empty set and X are in $\mathcal{A}$ as $\emptyset = A \cap A^c$ and $X = \emptyset^c$. Closure under unions is implied by closure under disjoint unions. Indeed, we can inductively remove the pairwise intersections if nonempty. Namely, in the sequence $(A_j)_{j\in\mathbb{N}}$ we replace A_j by $\tilde{A}_j := A_j \cap (\cup_{i<j} A_i)^c$; then (check) $\cup_j A_j = \cup_j \tilde{A}_j$. Check also that we have only used operations permitted in algebras/σ-algebras.*

Let X be a space and $\mathcal{M}$ a σ-algebra on X.

Definition 27. *The pair $(X, \mathcal{M})$ is called a measurable space.*

Some simple examples are, at one extreme, $\mathcal{A} = \{\emptyset, X\}$ and $\mathcal{A} = \{A : A \subset X\} = \mathcal{P}(X)$ at the other.

An important concept is that of a σ-algebra generated by a family $\mathcal{E}$ of sets:

Definition 28. *$\mathcal{M}(\mathcal{E})$, the σ-algebra generated by $\mathcal{E}$ is the intersection of all σ-algebras containing $\mathcal{E}$ ($\mathcal{P}(X)$ is one of those).*

Check that the intersection of a family of σ-algebras in a σ-algebra.

In a topological set obviously open sets play a special role. A σ-algebra compatible with the topology should contain the open sets.

Definition 29. *The **Borel σ-algebra** on a topological space X, $\mathcal{B}_X$, is the σ-algebra generated by the open sets in X.*

Clearly closed sets, countable intersections of open sets (called G_δ sets) countable unions of closed sets (called F_σ sets), and many more that we will uncover, are in $\mathcal{B}_X$.

2.1 Measures

The definition below generalizes some of the properties we would expect from volumes in $\mathbb{R}^n$.

Definition 30. *Let $\mathcal{M}$ be a σ-algebra on the set X. A function $\mu : \mathcal{M} \to [0, \infty]$ is a measure on $\mathcal{M}$ if*

1. $\mu(\emptyset) = 0$
2. *(σ-additivity) If $(A_j)_{j\in\mathbb{N}}$ is a family of mutually disjoint sets, then*

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j) \tag{19}$$

Definition 31. *The triple $(X, \mathcal{M}, \mu)$ where $\mathcal{M}$ is a σ-algebra on X and μ is a measure on $\mathcal{M}$ is called a measure space.*

Exercise 12 (The counting measure)**.** *Let X be any nonempty set and take $\mathcal{M}$ to be any σ-algebra on X (including the maximal one, $\mathcal{P}(X)$).*

1. *For any $A \in \mathcal{M}$, let $\mu(A)$ be the number of points in A (understood to be zero if $A = \emptyset$, n if there is a bijection between A and $\{1, ..., n\}$ and $+\infty$ otherwise). Show that μ is a measure on $A \in \mathcal{M}$.*
2. *(The Dirac mass at x_0) Let x_0 be a point in X, and for any $A \in \mathcal{M}$ let $\delta_{x_0}(A)$ be one if $x_0 \in A$ and zero otherwise. Show that δ_{x_0} is a measure on $\mathcal{M}$.*

For a measure μ to agree with our intuition about volumes, we would require more properties from it: invariance under Euclidean transformations (these are the isometries of $\mathbb{R}^n$) and normalization, namely the measure of a (hyper)cube of side a in $\mathbb{R}^n$ should equal a^n (in one dimension $\mu((a, b)) = b - a$). In particular, the underlying σ-algebra should be at least as large as the Borel σ-algebra on $\mathbb{R}^n$.

$\mathcal{M}$, however, cannot be too large; for instance, we cannot have $\mathcal{M} = \mathcal{P}(\mathbb{R}^n)$.

Proposition 32 (Existence of non-measurable sets). *Let $\mathcal{M}$ be any σ-algebra on $\mathbb{R}$ such that there is a measure on $\mathcal{M}$ that is invariant under Euclidean transformations and normalized. Then there are sets N in $\mathbb{R}$, $N \notin \mathcal{M}$.*

Proof. The construction is simpler if we work mod 1, and then translation becomes rotation on S^1, the circle of circumference 1. Assume the contrary. Consider the equivalence relation on $[0,1)$ mod 1 $x \sim y$ iff $x - y \in \mathbb{Q}$. Let $\mathcal{C}$ be the collection of equivalence classes modulo $\sim$. Using the axiom of choice (AC) [3], let E be a set which contains exactly one element from each class. (By the AC there is a choice function $F : \mathcal{C} \to S^1$ s.t. $\forall C \in \mathcal{C}, F(C) \in C$; then $E = F(\mathcal{C})$.) For each $r \in \mathbb{Q}$ let $E_r = \{x + r : x \in E\}$. By definition, if $r \neq r', E_r \cap E_{r'} = \emptyset$, and E_r is obtained from E by translation by r, and thus $\forall r \in \mathbb{Q}$, $\mu(E_r) = \mu(E)$. Clearly, $\cup_{r\in\mathbb{Q}} E_r = S^1$ $(*)$. Therefore, if $\mu(E) = 0$, then $\mu(S^1) = 0$ and if $\mu(E) > 0$, then $\mu(S^1) = +\infty$ which contradict the normalization $\mu(S^1) = 1$. □

Note 33. The AC is crucial to the proof. The existence of a set E as above is independent of ZF, the axioms of mathematics without the AC. Furthermore, one can show that there is no definition even in ZFC (ZF+AC) that, provably and uniquely , defines such an E. That is, these cannot be of the form $\{x \in \mathbb{R} : P(x)\}$ where P is some predicate; in particular, no "specific example" can be "constructed". If you are "given" such an E you can't check it really is one. Nor can one define a σ-algebra with the properties in the Proposition. [4]. (A more detailed and careful formulation of these impossibility statements is needed to make them really rigorous and correct; that's beyond the scope of these notes though; see Non-measurable sets and the AC)

More strikingly, using the AC the Banach-Tarsky paradox produces a finite partition of the unit cube in $\mathbb{R}^n, n \geqslant 3$ in subsets which can be rearranged by Euclidean transformations (by cut and paste!) to become two unit cubes (or any other number of them of any size, for that matter) obviously violating the normalization condition. This precludes even the existence of a finitely additive such measure on $\mathbb{R}^n$. (The use of the AC means however that you definitely cannot do this at home with Play Doh.)

2.2 Measurable functions

Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets and $f : \mathcal{X} \to \mathcal{Y}$ be a map between them. The inverse image through f: $f^{-1}(\mathtt{Y}) := \{x \in \mathcal{X} : f(x) \in \mathtt{Y}\}$ is a map between $\mathcal{P}(\mathcal{Y})$ and $\mathcal{P}(\mathcal{X})$ which commutes with $\cup, \cap$ and complements; that is, we have $f^{-1}(\mathtt{Y}_1 \cup \mathtt{Y}_2) = f^{-1}(\mathtt{Y}_1) \cup f^{-1}(\mathtt{Y}_2)$ etc.

Exercise 13. *Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets, let $\mathcal{N}$ be a σ-algebra on $\mathcal{Y}$ and $f : \mathcal{X} \to \mathcal{Y}$ be a map between these sets. Show that $\mathcal{M} := \{f^{-1}(\mathtt{Y}) : \mathtt{Y} \in \mathcal{N}\}$ is a σ-algebra on $\mathcal{X}$.*

Definition 34. *Let $(\mathcal{X}, \mathcal{M})$ and $(\mathcal{Y}, \mathcal{N})$ be measurable spaces. A function $f : \mathcal{X} \to \mathcal{Y}$ is called measurable (more precisely $(\mathcal{M}, \mathcal{N})$-measurable) if the preimage through f of any set in $\mathcal{N}$ is in $\mathcal{M}$, that is: $f^{-1}(\mathcal{N}) \subset \mathcal{M}$.*

Proposition 35. *Assume $(\mathcal{Y}, \mathcal{N})$ is a measurable space where $\mathcal{N}$ is generated by $\mathcal{E} \subset \mathcal{P}(\mathcal{Y})$. Let $\mathcal{X}$ be a set, and $f : \mathcal{X} \to \mathcal{Y}$. Then, the σ-algebra $f^{-1}(\mathcal{N})$ is generated by $f^{-1}(\mathcal{E})$. In particular, if $(\mathcal{X}, \mathcal{M})$ and $(\mathcal{Y}, \mathcal{N})$ are measurable spaces and $\mathcal{N}$ is generated by $\mathcal{E}$, then f is $(\mathcal{M}, \mathcal{N})$-measurable iff $f^{-1}(E) \in \mathcal{M}$ for all $E \in \mathcal{E}$.*

[3]Formally, this states: $\forall X \left[\emptyset \notin X \implies \exists f\colon X \to \bigcup X \quad \forall A \in X\,(f(A) \in A)\right]$.

[4]More generally, this applies to any set which is not Lebesgue measurable, a notion that we'll discuss later.

Proof. Necessity is obvious. For sufficiency note that the collection $\{Y \subset \mathcal{Y} : f^{-1}(Y) \in \mathcal{M}\}$ is a σ-algebra which contains $\mathcal{E}$, thus it contains the σ-algebra generated by $\mathcal{E}$. □

2.3 Product σ-algebras

Let A be an index set, $(\mathcal{X}_\alpha, \mathcal{M}_\alpha)_{\alpha \in A}$, a collection of measurable spaces, and X their Cartesian product, $X = \prod\limits_\alpha X_\alpha$. On X there is a naturally induced σ-algebra, namely the smallest σ-algebra that makes all canonical projections π_α measurable:

Definition 36. *Let $(\mathcal{X}_\alpha, \mathcal{M}_\alpha)$ and X be as above. The product σ-algebra on X is the σ-algebra generated by the collection of sets $\{\pi_\alpha^{-1}(E_\alpha) : E_\alpha \in \mathcal{M}_\alpha, \alpha \in A\}$. The product σ-algebra is denoted by $\bigotimes\limits_{\alpha \in A} \mathcal{M}_\alpha$.*

Proposition 37. *If the index set A is countable, then $\bigotimes\limits_{\alpha \in A} \mathcal{M}_\alpha$ is generated by the collection of all $\prod_\alpha E_\alpha$ where $E_\alpha \in \mathcal{M}_\alpha$.*

Proof. Simply note that $\prod_\alpha E_\alpha = \cap_\alpha \pi^{-1}(E_\alpha)$ are measurable, and that $\pi^{-1}(E_\alpha) = \cap_{\beta \in A} \pi^{-1}(E_\beta)$ for a suitable choice of the $(E_\beta)_{\beta \in A}$ (which?) □

Proposition 38. *Assume $\mathcal{M}_\alpha$ is generated by $\mathcal{E}_\alpha$. $\bigotimes\limits_{\alpha \in A} \mathcal{M}_\alpha$ is generated by $\mathcal{F}_1 = \{\pi^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha, \alpha \in A\}$.*

Proof. By Proposition 35 the $(\pi_\alpha)_{\alpha \in A}$ are measurable iff $\pi_\alpha^{-1}(E_{\beta,\alpha})$ are measurable for all $E_{\beta,\alpha} \in \mathcal{E}_\alpha$ and all α. □

Proposition 39. *Let $X_1, ..., X_n$ be metric spaces and let $X = \prod_1^n X_j$ be equipped with the product metric. Then the product Borel σ-algebra, $\bigotimes_1^n \mathcal{B}_{X_j}$ is contained in the Borel σ-algebra on X, $\mathcal{B}_X$, and the two coincide if X_j are separable.*

Proof. By Proposition 38 $\bigotimes_1^n \mathcal{B}_{X_j}$ is generated by $\pi_j^{-1}(O_j)$ where O_j are open in X_j; by definition of the product metric, these sets are open in X, thus elements of $\mathcal{B}_X$. For the second part we will find a countable base of the topology of X of the form $(\prod_{k=1}^n O_{jk})_{j \in \mathbb{N}}$ where for each k, $(O_{jk})_{j \in \mathbb{N}}$ form a basis in the topology of X_j. Take a countable dense set D_j in each X_j, and the countable collection of all balls $\mathcal{E}_j = (B_{j,n})_{n \in \mathbb{N}}$ of rational radii centered at some point in D_j. Clearly, for $j = 1...n$, $\mathcal{B}_{X_j}$ is generated by $\mathcal{E}_j$. Now, the set of points $x \in X$ such that for any j, x_j is in some B_{jn} is dense in X. A ball of radius r in X is by definition the product of balls of radius r in each X_j and the result follows. □

Corollary 40. $\mathcal{B}_{\mathbb{R}^n} = \bigotimes_1^n \mathcal{B}_{\mathbb{R}}$.

To reduce some proofs in the sequel to simpler cases, we introduce elementary families. These are collections $\mathcal{E} \subset \mathcal{P}(X)$ such that

1. $\emptyset \in \mathcal{E}$.
2. If $E_1, E_2 \in \mathcal{E}$ then $E_1 \cap E_2 \in \mathcal{E}$.
3. If $E \in \mathcal{E}$, then E^c is a finite disjoint union of elements of $\mathcal{E}$

Proposition 41. *If $\mathcal{E}$ is an elementary family, then the collection $\mathcal{A}$ of finite disjoint unions of elements of $\mathcal{E}$ is an algebra.*

Proof. If $A, B \in \mathcal{A}$ then $A = \cup_j E_j$ and $B = \cup_k F_k$ where the finitely many E_j, as well as the F_k, are mutually disjoint in $\mathcal{E}$. Then

$$A \bigcap B = \bigcup_{j,k} (E_j \bigcap F_k)$$

where it is easy to check that the sets in the collection $(E_j \cap F_k)_{j,k}$ are mutually disjoint elements of $\mathcal{E}$. Then, for disjoint sets E_j and E_{j,k_j} we have

$$A^c = (\bigcup_{j=1}^{n} E_j)^c = \bigcap_{j=1}^{n} E_j^c = \bigcap_{j=1}^{n} (\bigcup_{k_j} E_{j,k_j}) = \bigcup_{k_1,\dots,k_n} \bigcap_{j=1}^{n} E_{j,k_j}$$

again a disjoint union of elements of $\mathcal{E}$.

$\square$

2.4 More about measures

Theorem 42. *Let $(X, \mathcal{M}, \mu)$ be a measure space and $A, B, (E_j)_{j\in\mathbb{N}}$ measurable sets.. Then μ is*

1. *Monotonic: $A \subset B \Rightarrow \mu(A) < \mu(B)$.*
2. *Subadditive: $\mu(\bigcup_{j\in\mathbb{N}} E_j) \leqslant \sum_{k\in\mathbb{N}} \mu(E_j)$*
3. *Continuous from below. If $E_1 \subset E_2 \cdots$, then $\mu(\bigcup_{j\in\mathbb{N}} E_j) = \lim_{j\to\infty} \mu(E_j)$.*
4. *Continuous from above. If $E_1 \supset E_2 \cdots$, and $\mu(E_1) < \infty$, then $\mu(\bigcap_{j\in\mathbb{N}} E_j) = \lim_{j\to\infty} \mu(E_j)$.*

Proof. 1. We have $B = A \bigcup (B \setminus A)$ and thus $\mu(B) = \mu(A) + \mu(B \setminus A) \geqslant \mu(A)$.

2. We replace the union by an equivalent disjoint union:

$$\mu(\bigcup_{k\in\mathbb{N}} E_j) = \mu\left(\bigcup_{k\in\mathbb{N}}\left(E_k \bigcap_{j<k} E_j^c\right)\right) = \sum_{k\in\mathbb{N}} \mu(E_k \bigcap_{j<k} E_j^c) \leqslant \sum_{k\in\mathbb{N}} \mu(E_j)$$

by 1.

3. Similarly, setting $E_0 = \emptyset$,

$$\mu(\bigcup_{k\in\mathbb{N}} E_k) = \mu\left(\bigcup_{k\in\mathbb{N}}\left(E_k \bigcap_{j<k} E_j^c\right)\right) = \sum_{k\in\mathbb{N}} \mu(E_k \bigcap E_{k-1}^c) = \lim_{n\to\infty} \sum_{k=1}^{n} \mu(E_k \bigcap E_{k-1}^c) = \lim_{n\to\infty} \mu(E_n)$$

4. Note that $E_1 = (E_1 \setminus E_2) \cup E_2 = (E_1 \setminus E_2) \cup (E_2 \setminus E_3) \cup E_3 = \cdots = \bigcap_{i\in\mathbb{N}} E_i \cup \bigcup_{j\in\mathbb{N}} (E_j \setminus E_{j+1})$ where all the unions are disjoint. Hence,

$$\mu(E) = \sum_{j\leqslant n} \mu(E_j \setminus E_{j+1}) + \mu(E_{n+1}) = \sum_{j\in\mathbb{N}} \mu(E_j \setminus E_{j+1}) + \mu(\bigcap_{i\in\mathbb{N}} E_i)$$

easily (how?) completing the proof.

Exercise 14 (Suggested by one of you). *Let X be an infinite set and let κ be its cardinal. Let Y be an infinite set of cardinality $\kappa' < \kappa$. Let $\mathcal{M}$ an infinite σ-algebra on Y, and let its cardinality be κ''. Show that there is a σ-algebra of cardinality κ'' in X. As a hint, Ex.1 p. 24 in Folland could help.*

Notes about this exercise:

The order among cardinal numbers $|\cdot|$, is defined as follows: $|Y| \leqslant |X|$ if there exists an injective function $f : Y \to X$. The axiom of choice implies (and in fact is equivalent to) the statement that given two sets X and Y we have $|Y| \leqslant |X|$ or $|Y| \leqslant |X|$.

Exercise 15. *Let $X = \mathbb{Q} \cap [0,1]$, let $\mathcal{E}$ be the family of all intervals of the form $\{q \in \mathbb{Q} : a < q \leqslant b\}$ where $a, b \in X$, and $\mathcal{A}$ be the algebra generated by $\mathcal{E}$.*

(1) What is the σ-algebra $\mathcal{M}$ generated by $\mathcal{E}$?

(2) Define a set-function on $\mathcal{E}$ by $\mu((a,b]) = b - a$. Show that it extends to a finitely-additive measure on $\mathcal{A}$.

(3) Does μ extend to a σ-additive measure to $\mathcal{M}$? (In other words: is there a measure on $\mathcal{M}$ which agrees with μ on $\mathcal{A}$?)

□

Note 43. *The condition $\mu(E_1) < \infty$ can clearly be relaxed to $\mu(E_n) < \infty$ for some n since any finite subfamily of E_k can be removed from the intersection. However the condition $\mu(E_n) < \infty$ for some n is needed. Indeed, let μ be the counting measure on $\mathcal{P}(\mathbb{N})$ and let $E_n = \{n, n+1, ...\}$. Clearly $\cap_n E_n = \emptyset$ while $\mu(E_k) = +\infty$ for all k.*

Property 2 in Definition 30 is called σ-additivity of the measure. A function μ which is only additive for finite families of disjoint sets is called finitely additive.

A measure on $(X, \mathcal{M})$ is semifinite if any $E \in \mathcal{M}$ with $\mu(E) \neq 0$ has a subset of finite positive measure. It is finite if $\mu(X) < \infty$, which, by the previous theorem, implies $\mu(E) < \infty$ for all $E \in \mathcal{M}$. An important notion is that of *σ-finite* measures, meaning that there is a countable partition of X in disjoint sets E_j, $\cup_j E_j = X$ s.t. $\mu(E_j) < \infty$ for any j. More generally E is σ-finite in $(X, \mathcal{M}, \mu)$ if there is a countable partition of E in disjoint sets E_j, $\cup_j E_j = E$ s.t. $\mu(E_j) < \infty$ for any j. Clearly, the counting measure on $\mathcal{P}(\mathbb{N})$ is σ-finite. Check that the counting measure is σ-finite on $\mathcal{P}(X)$ iff X is finite or countable.

Measure zero sets. A set $E \in \mathcal{M}$ is of measure zero w.r.t. $(X, \mathcal{M}, \mu)$ if $\mu(E) = 0$. Clearly a countable union of measure zero sets has measure zero. A property holds ***μ*-almost everywhere** if it holds except on a set of measure zero. We simply say that the property holds almost everywhere, abbreviated a.e., when the μ used is clear from the context.

By monotonicity, if $M, N \in \mathcal{M}$ with $M \subset N$, then $\mu(N) = 0$ entails $\mu(M) = 0$. It is natural to extend $\mathcal{M}$ and μ so that all subsets of a set of measure zero are measurable, with measure zero. The resulting measure is called complete. Such an extension is always possible.

Theorem 44. *Let $(X, \mathcal{M}, \mu)$ be a measure space, $\mathcal{N} = \{N \in \mathcal{M} : \mu(N) = 0\}$ and $\overline{\mathcal{N}} = \{M \subset N : N \in \mathcal{N}\}$.*

1. *Let $\overline{\mathcal{M}} = \{A \cup M : A \in \mathcal{M}, M \in \overline{\mathcal{N}}\}$. Then $\overline{\mathcal{M}}$ is a σ-algebra.*
2. *Define $\overline{\mu}$ on $\overline{\mu}(A \cup M) = \mu(A)$. Then $(X, \overline{\mathcal{M}}, \overline{\mu})$ is a measure space and $\overline{\mu}$ extends μ.*

Proof. Note that $\overline{\mathcal{N}}$ is closed under countable unions and intersections. Closure under countable unions of $\overline{\mathcal{M}}$ is clear: $\cup_i(A_i \cup M_i) = (\cup_i A_i) \cup (\cup_i M_i)$. Now,

$$(A \cup M)^c = A^c \cap M^c = A^c \cap (N^c \cup (N \setminus M)) = (A^c \cap N^c) \cup (A^c \cap (N \setminus M))$$

which proves 1 noting that $A^c \cap (N \setminus M) \in \overline{\mathcal{N}}$.

2. The only part that may not be straightforward is the consistency of the definition: If $A \cup M = A' \cup M'$, then we should have $\mu(A) = \mu(A')$. For some $N' \in \mathcal{N}$ we have

$$A \setminus A' \subset (A \cup M) \setminus A' = (A' \cup M') \setminus A' = M' \cap (A')^c \subset N' \cap (A')^c \in \mathcal{N}$$

and similarly $A' \setminus A \in \mathcal{N}$ implying $\mu(A \Delta A') = 0$ and the result follows. □

3 Construction of measures

We start with an informal discussion on defining a measure of length λ on $\mathcal{B}_{\mathbb{R}}$. As noted, the measure should be translation-invariant, and such that $\mu((a,b)) = b - a$ ($= +\infty$ for unbounded intervals). Countable sets would have measure zero, since they can be covered by a union of open intervals of arbitrarily small total length. Indeed, for any $\epsilon > 0$ the sequence $(x_n)_{n\in\mathbb{N}}$ [5] is contained in the union of the intervals $J_n = (x_n - \epsilon 2^{-n}, x_n + \epsilon 2^{-n})$. In particular if J is an interval with endpoints a, b $\lambda(J) = b - a$ regardless of whether the interval is open, closed, or half-open. Any open set in $\mathbb{R}$ is a countable union of open intervals, and, by the usual trick of making the union disjoint, it is a countable union of disjoint intervals. This allows us to define $\lambda(\mathcal{O})$ for any open set $\mathcal{O}$, and from it $\lambda(\mathcal{C})$ for any closed set $\mathcal{C}$. What else can we define? If $A \in \mathcal{B}_{\mathbb{R}}$ has the property that for any ϵ there exist an open set $\mathcal{O}_\epsilon \supset A$ and a closed set $\mathcal{C}_\epsilon \subset A$ such that $\lambda(\mathcal{O}_\epsilon \setminus \mathcal{C}_\epsilon) < \epsilon$ it is natural to try $\lambda(A) := \lim_{\epsilon\to 0} \lambda(\mathcal{O}_\epsilon)$. (Think why it would be a bad idea to try to approximate sets with open sets from inside, or with closed sets from the outside). Proceeding this way, it's a pretty steep climb, where we would have to check all sorts of consistencies, whether any $A \subset \mathcal{B}_{\mathbb{R}}$ has a measure, etc. The concept of outer measure is a nice way to minimize this work.

3.1 Outer measures

Definition 45. *Let X be a set. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is an outer measure on X if*

1. *$\mu^*(\varnothing) = 0$.*
2. *(Monotonicity) If $A \subset B$, then $\mu^*(A) \leqslant \mu^*(B)$.*

[5] Do we need the axiom of choice to present a countable set as a sequence?

3. *(Countable subadditivity)* $\mu^*(\bigcup_{i\in\mathbb{N}} A_i) \leqslant \sum_{i\in\mathbb{N}} \mu^*(A_i)$.

Note that, unlike in the *σ-additive* case, $3 \not\Rightarrow 2$. For example, on $\mathbb{R}$ an outer measure is

$$\lambda^*(A) = \inf\left\{\sum_{n\in\mathbb{N}} \lambda(\mathcal{O}_i) : \mathcal{O}_i \text{ open interval}, A \subset \bigcup_{i\in\mathbb{N}} O_i\right\} \tag{20}$$

More generally, we have the following result.

Proposition 46. *Let $\mathcal{E} \subset \mathcal{P}(X)$ and $\rho : \mathcal{E} \to [0,\infty]$ be such that $\emptyset$ and X are in $\mathcal{E}$ and $\rho(\emptyset) = 0$. For $A \in \mathcal{P}(X)$ let*

$$\mu^*(A) = \inf\left\{\sum_{n\in\mathbb{N}} \rho(E_i) : E_i \in \mathcal{E}, A \subset \bigcup_{i\in\mathbb{N}} E_i\right\} \tag{21}$$

Then μ^* is an outer measure on X.

Proof. Note that (1) μ^* is well-defined since $A \in \mathcal{P}(X) \Rightarrow A \subset \cup_j X$ and (2) μ^* is nonnegative. Furthermore, since $\emptyset \subset \cup_j \emptyset$, we have $\mu(\emptyset) = 0$. Monotonicity is also easy, since $A \subset B$ and $B \subset \cup_j E_j \Rightarrow A \subset \cup_j E_j$.

To show subadditivity, let $A_i \in \mathcal{P}(X), i \in \mathbb{N}$ and $\epsilon > 0$. By definition, for each i there are sets $E_{ij} \subset \mathcal{E}$ such that $A_i \subset \cup_j E_{ij}$ and

$$\sum_{j\in\mathbb{N}} \rho(E_{ij}) - \epsilon 2^{-j} \leqslant \mu^*(A_i) \leqslant \sum_{j\in\mathbb{N}} \rho(E_{ij})$$

It follows that

$$\bigcup_{j\in\mathbb{N}} A_i \subset \bigcup_{(i,j)\in\mathbb{N}^2} E_{ij}$$

and

$$\mu^*\left(\bigcup_{j\in\mathbb{N}} A_i\right) \leqslant \sum_{(i,j)\in\mathbb{N}^2} \rho(E_{i,j}) \leqslant \sum_{i\in\mathbb{N}} \mu^*(A_i) + \epsilon \tag{22}$$

(Justify the use of double indices.) Since (22) holds for any positive ϵ, subadditivity follows. □

We could similarly define an inner measure on $\mathbb{R}$ by taking sup over compact sets contained in a given $A \in \mathcal{P}(\mathbb{R})$. Then, measurable sets should be those for which the inner and outer measure coincide. However, another clever trick allows us to save half of the effort, and rely solely on outer measures. Returning to the length measure, we expect to have $\lambda(A) = \lambda^*(A)$ for any $A \in \mathcal{B}_{\mathbb{R}}$. This implies that, for $A \in \mathcal{B}_{\mathbb{R}}$

$$\lambda^*(B) = \lambda^*(B \cap A) + \lambda^*(B \cap A^c), \forall B \in \mathcal{B}_{\mathbb{R}} \tag{23}$$

The key observation is that the equality above is a property of A rather than of B (it reflects the way A splits other sets.)

Exercise 16. *Check that λ^* satisfies (23) for all $B \in \mathcal{P}(\mathbb{R})$, when A is an open set.*

This suggests the following.

Definition 47. *Let μ^* be an outer measure on X. A set $A \subset X$ is called μ^*-measurable if*

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c), \forall B \in \mathcal{P}(X) \tag{24}$$

Note that, by subadditivity of outer measures (23) holds whenever the left side is no less than the right side.

Theorem 48 (Carathéodory's theorem)**.** *If μ^* is an outer measure on X, then the collection $\mathcal{M}$ of all μ^*-measurable sets is a σ-algebra and μ^* (restricted to $\mathcal{M}$) is a complete measure on $\mathcal{M}$.*

Proof. I. $\mathcal{M}$ is closed under complements. This is obvious.

II. Closure under finite unions. Note that if A and B are measurable and E is any set, we first split it by A and then by B to get

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B) + \underbrace{\mu^*(E \cap A^c \cap B^c)}_{\mu^*(E \cap (A \cup B)^c)}$$
$$\geqslant \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$$

since

$$A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$$

Thus $\mathcal{M}$ is an algebra.

III. Closure under countable unions follows now if we show closure under countable disjoint unions.

Let $(A_j)_{j \in \mathbb{N}}$ be disjoint, $S_n = \cup_{j=1}^n A_j$ and $S = \cup_{j=1}^\infty A_j$. For $E \subset X$, since the A_j and S_j are measurable, we have

$$\mu^*(E \cap S_n) = \mu^*(E \cap S_n \cap A_n) + \mu^*(E \cap S_n \cap A_n^c) = \mu^*(E \cap A_n) + \mu^*(E \cap S_{n-1}) \overset{\text{induction}}{=} \cdots = \sum_{j=1}^n \mu^*(E \cap A_j)$$

Since $S \supset S_n$ and $E \cap S = \cup_j E \cap A_j$, we get, by subadditivity and monotonicity,

$$\mu^*(E \cap S) \geqslant \sum_{j \in \mathbb{N}} \mu^*(E \cap A_j) \geqslant \mu^*(\cup_j E \cap A_j) = \mu^*(E \cap S) \Rightarrow \mu^*(E \cap S) = \sum_{j \in \mathbb{N}} \mu^*(E \cap A_j) \tag{25}$$

Since $E \cap S_n^c \supset E \cap S^c$ and S_n are measurable, we now get

$$\mu^*(E) = \mu^*(E \cap S_n) + \mu^*(E \cap S_n^c) \geqslant \sum_{j=1}^n \mu^*(E \cap A_j) + \mu^*(E \cap S^c) \underset{n \to \infty}{\to} \mu^*(E \cap S) + \mu^*(E \cap S^c)$$

implying $S \in \mathcal{M}$. σ-additivity follows by taking $E = S$ in (25).

IV. Completeness: Let $N \in \mathcal{M}$ be s.t. $\mu^*(N) = 0$. By monotonicity, $\mu^*(E \cap N) = 0$ for any $E \subset X$, and since N is measurable, $\mu^*(E) = \mu^*(E \cap N^c)$. Let $M \subset N$. Again using monotonicity, $\mu^*(E \cap M) = 0$. Thus, we only need to show $\mu^*(E \cap M^c) = \mu^*(E)$ which follows from monotonicity: $\mu^*(E \cap M^c) \geqslant \mu^*(E \cap N^c) = \mu^*(E)$. □

HW for 09/17 : Problems 1–5 on p. 24 in Folland, and turn in: Ex 10,14 and 15 in these notes.

3.2 Measures from pre-measures

Definition 49. *Let $\mathcal{A}$ be an algebra in X. A function $\mu_0 : \mathcal{A} \to [0, \infty]$ is called* premeasure *if*

1. $\mu_0(\emptyset) = 0$.
2. *If $(A_j)_{j\in\mathbb{N}}$ is a sequence of disjoint sets in $\mathcal{A}$ s.t. $\bigcup_{j\in\mathbb{N}} A_j \in \mathcal{A}$, then $\mu_0(\bigcup_{j\in\mathbb{N}} A_j) = \sum_{j\in\mathbb{N}} \mu_0(A_j)$.*

The outer measure induced by μ_0 is

$$\mu^*(E) = \inf\left\{ \sum_{j\in\mathbb{N}} \mu_0(A_j) : A_j \in \mathcal{A}, E \subset \bigcup_{j\in\mathbb{N}} A_j \right\} \tag{26}$$

Note that, by monotonicity and the fact that $\mathcal{A}$ is an algebra, the unions in (26) can be assumed disjoint.

Theorem 50. *(a) Let $\mathcal{A}$ be an algebra on X and μ_0 a premeasure on $\mathcal{A}$. Then $\mu := \mu^*$ defined by (26) is a measure on $\mathcal{M}$, the σ-algebra generated by $\mathcal{A}$ and coincides with μ_0 on $\mathcal{A}$.*

(b) If μ_0 is σ-finite, then μ is the unique measure with this property. Otherwise, any other measure ν as above has the property that $\mu(A) \geqslant \nu(A)$ for all $A \in \mathcal{M}$, and $\mu - \nu = 0$ on all sets of finite μ measure.

For the proof we need the following result.

Lemma 51. *Under the conditions of the theorem,*

1. $\mu^*|_{\mathcal{A}} = \mu_0$.
2. *The sets in $\mathcal{A}$ are μ^*-measurable.*

Proof. 1. Note first that for any $E \in \mathcal{A}$ we have $\mu_0(E) \geqslant \mu^*(E)$. To prove the opposite inequality, let $E \in \mathcal{A}$ and A_j as in (26), assumed w.l.o.g. to be disjoint. Then $E = E \cap \cup_j A_j = \cup_j (E \cap A_j)$ and, since μ_0 is a premeasure,

$$\mu_0(E) = \sum_{j\in\mathbb{N}} \mu_0(E \cap A_j) \leqslant \sum_{j\in\mathbb{N}} \mu_0(A_j)$$

implying $\mu_0(E) \leqslant \mu^*(E)$.

2. Let $A \in \mathcal{A}$, $E \subset X$ and $\epsilon > 0$. There is a disjoint sequence $(A_j)_{j\in\mathbb{N}}$ in $\mathcal{A}$ s.t. $E \subset \cup_j A_j$ and $\mu^*(E) + \epsilon \geqslant \sum_j \mu_0(A_j)$. Thus,

$$\mu^*(E) + \epsilon \geqslant \sum_{j\in\mathbb{N}} \mu_0(A_j) = \sum_{j\in\mathbb{N}} \mu_0(A_j \cap A) + \sum_{j\in\mathbb{N}} \mu_0(A_j \cap A^c) \geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

implying, since ϵ is arbitrary, that A is measurable.

□

Proof of the Theorem. (a) follows from the Lemma and Carathéodory's theorem.

(b) We first prove that any measure μ as in the theorem has the property $\nu(A) \leqslant \mu(A)$ on $\mathcal{M}$. If $E \in \mathcal{M}$ and A_j are disjoint sets whose union contains E, by monotonicity of ν we must have

$$\nu(E) \leqslant \sum_{j\in\mathbb{N}} \nu(A_j) = \sum_{j\in\mathbb{N}} \mu_0(A_j)$$

and thus $\nu(E) \leqslant \mu(E)$.

We claim that, if $A_j \in \mathcal{A}$ are disjoint and $A = \cup_j A_j$, then $\mu(A) = \nu(A)$. Indeed, we have

$$\nu(A) = \sum_{j\in\mathbb{N}} \nu(A_j) = \sum_{j\in\mathbb{N}} \mu_0(A_j) = \sum_{j\in\mathbb{N}} \mu(A_j) = \mu(A)$$

If $\mu(E) < \infty$, then, for any $\epsilon > 0$ there is a disjoint family of $A_j \in \mathcal{A}$ whose union A contains E, s.t. $\mu(A) = \sum_j \mu_0(A_j) \leqslant \mu(E) + \epsilon$ and hence $\nu(A \setminus E) \leqslant \mu(A \setminus E) \leqslant \epsilon$. Now

$$\mu(A) = \nu(A) = \nu(E) + \nu(A \setminus E) \leqslant \nu(E) + \epsilon$$

and thus $\mu(E) = \nu(E)$.

If μ is σ-finite, then $X = \cup_j A_j$ where A_j are disjoint and $\mu(A_j) < \infty$. Then,

$$\nu(E) = \sum_{j\in\mathbb{N}} \nu(E \cap A_j) = \sum_{j\in\mathbb{N}} \mu(E \cap A_j) = \mu(E)$$

□

Exercise 17. *[Done in class (*) Show that the function μ in Exercise 15 is not a premeasure.]*

1. Use the function μ in Exercise 15 to define an outer measure on $\mathbb{Q}$. What is the measure on $\mathbb{Q}$ induced by this outer measure?

2. Describe all translation-invariant measures on $\mathcal{P}(\mathbb{Q})$.

3. Describe all finite measures on $\mathcal{P}(\mathbb{Q})$.

4. Let ρ be a finite measure *on $\mathbb{Q}$ s.t. any singleton has positive measure and define the function f on $\mathbb{C} \setminus \mathbb{R}$ by*

$$f(z) = \sum_{r\in\mathbb{Q}} \frac{\rho(r)}{z - r}$$

Show that the series above converges absolutely and uniformly on compact sets in the open and lower upper half-planes, and that for any $r \in \mathbb{Q}$ the limit of $|f|$ when $z \to r$ along a vertical line is $+\infty$.

Remark. *For those who took Complex Analysis, this shows that f is analytic in the open and lower upper half-planes, and that $\mathbb{R}$ is a natural boundary for f. Think why there must exist points $\xi \in \mathbb{R}$ where the limit as $z \to \xi$ from the upper half plane either does not exist or it is not infinite.*

If μ is a finite measure on $\mathcal{B}_{\mathbb{R}}$, then its distribution function is $F = x \mapsto \mu(-\infty, x]$. For instance, for the Dirac mass at 0, F is the Heaviside function θ, *extended by* $\theta(0) = 1$. [6] Distribution functions are increasing (meaning: nondecreasing) and right continuous since $\mu((-\infty, x]) = \lim_{x_n \to x+0} \mu((-\infty, x_n])$. (What is different if we take $\lim_{x_n \to x-0}$ instead?)

Exercise 18. *(i) Let F be increasing and right-continuous on $\mathbb{R}$. Show that F has at most countably many discontinuities.*

(ii) Let $C = \{x_j : j \in \mathbb{N}\} \subset \mathbb{R}$, $(\rho_j)_{j\in\mathbb{N}}$ be a sequence of positive numbers s.t. $\sum_{j=1}^{\infty} \rho_j < \infty$, and $\mathcal{A}$ as in (28). For $A \in \mathcal{A}$ define

$$\mu_0(A) = \sum_{x_j \in A} \rho_j$$

[6] The Dirac mass will be seen to correspond to a distribution (in a different sense, that of distributions) while θ is a function, is the integral, in the sense of distributions $\int_{-\infty}^{x} \delta(s)ds$.

Show that μ_0 is a premeasure on $\mathcal{A}$. Show that there is a unique measure μ on $\mathcal{B}_{\mathbb{R}}$ which extends μ_0, and that μ is a finite measure.

(iii) Show that the distribution function of μ is discontinuous at any point in C.

Exercise 19. *Define $\rho : \mathbb{Q} \to \mathbb{Q}$ by $\rho(r) = 1$ if $r \in \mathbb{Z}$ and $\rho(r) = 1/|q|^3$ if $\rho = p/q$, p, q coprime. Let $\mathcal{A}$ the algebra generated by the right-closed left-open intervals on $\mathbb{R}$. Define μ_0 on $\mathcal{A}$ by $\mu(A) = \sum_{r \in A \cap \mathbb{Q}} \rho(r)$.*

(a) Show that μ_0 extends uniquely to a ($\sigma-$finite) measure on μ on $\mathbb{R}$ which is invariant under shift by one. Are there other shifts under which it is invariant?

(b) Show that $\mu(\{x\}) \neq 0$ iff $x \in \mathbb{Q}$.

(c) Let

$$F(x) = \begin{cases} \mu((0,x] \text{ if } x > 0 \\ 0 \text{ if } x = 0 \\ -\mu((x,0]) \text{ if } x < 0 \end{cases} \tag{27}$$

Find all the points of discontinuity of F.

4 Borel measures on the real line

In this section we will classify all Borel measures on $\mathbb{R}$, defined as measures on $\mathcal{B}_{\mathbb{R}}$ and find their properties. We will show that any Borel measure on $\mathbb{R}$ arises from some increasing, right-continuous function F.

We take the elementary family $\mathcal{E}$ of half-open intervals of the form $(a, b]$, $-\infty \leqslant a \leqslant b < \infty$ and define, using Proposition 41 the algebra

$$\mathcal{A} = \left\{ \bigcup_{j=1}^{n} I_j : I_j \in \mathcal{E}, n \in \mathbb{N} \right\} \tag{28}$$

Definition 52. *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right-continuous. Let $J = (a, b] \in \mathcal{E}$. We define $\mu_0(J) = F(b) - F(a)$ (where we let $F(-\infty) = -\infty$ if F is unbounded below) and extend it to $\mathcal{A}$ by $\mu_0(\cup_{k=1}^{n} I_k) = \sum_{k=1}^{n} \mu_0(I_k)$ whenever I_k are disjoint intervals.*

Proposition 53. *The function μ_0 is a premeasure on $\mathcal{A}$.*

Proof. I. μ_0 is well-defined. It is easy to see that for any finite disjoint partition of $I = (a, b]$ in subintervals $J_i = (a_i, b_i]$ we have $\mu_0(I) = \sum_i \mu_0(J_i)$.

Assume $\cup_{k=1}^{n} I_k = \cup_{l=1}^{m} J_l$, where the sets $\{I_k\}_k$, as well as the sets $\{J_l\}_l$ are disjoint in $\mathcal{E}$. The previous reasoning shows that

$$\sum_{k=1}^{n} \mu(I_k) = \sum_{k,l} \mu(I_k \cap J_l) = \sum_{l=1}^{m} \mu(J_l)$$

(there is an equivalent common subpartition, in other words).

The hard part is to show σ-additivity; let $\{I_k\}_{k\in\mathbb{N}}$ be disjoint sets in $\mathcal{A}$ such that $A = \bigcup_{j\in\mathbb{N}} I_j \in \mathcal{A}$. We leave it as an exercise that it is enough to show σ-additivity when $A \subset [-N, N]$ for some N, in which case $\mu_0(A) < \infty$. Note that

$$\mu_0\Big(\bigcup_{j\in\mathbb{N}} I_j\Big) = \sum_{j=1}^{n} \mu_0(I_j) + \mu_0\Big(\bigcup_{j>n} I_j\Big)$$

where all sets above are in $\mathcal{A}$. Thus $\sigma-$additivity reduces to continuity of μ_0 from above (see Theorem 42 for the definition).

Let $A_k = \cup_{j=1}^{n_k}(a_{kj}, b_{kj}]$ in $\mathcal{A}$ be a decreasing family such that $\cap_{k\in\mathbb{N}} A_k = \emptyset$, denote $c = \lim_{n\to\infty} \mu_0(A_k)$, and let $\epsilon > 0$. Since $\lim_{n\to\infty} F(a_{kj} + 1/n) = F(a_{kj})$, there are points $a'_{kj} \in (a_{kj}, b_{kj})$ such that $\mu_0(A_k \setminus \hat{A}_k) \leqslant \epsilon 2^{-k}$ for all k, where we denoted $\hat{A}_k = \cup_j (a'_{kj}, b_{kj}]$. Note that [7]

$$\bigcap_{j=1}^{n} A_j \subset \left(\bigcap_{j=1}^{n} \hat{A}_j\right) \cup \left(\bigcup_{j=1}^{n} (A_j \setminus \hat{A}_j)\right)$$

Hence

$$c \leqslant \mu_0\Big(\bigcap_{j=1}^{n} A_j\Big) \leqslant \mu_0\Big(\bigcap_{j=1}^{n} \hat{A}_j\Big) + \sum_{j=1}^{n} \mu_0(A_j \setminus \hat{A}_j) \leqslant \mu_0\Big(\bigcap_{j=1}^{n} \hat{A}_j\Big) + \epsilon \Rightarrow \mu_0\Big(\bigcap_{j=1}^{n} \hat{A}_j\Big) \geqslant c - \epsilon$$

The sequence of nested compact sets $K_n = \cap_1^n \overline{\hat{A}_k} \subset \cap_1^n A_k$ have empty intersection. Since $K_n \supset \bigcap_{j=1}^{n} \hat{A}_j$, for small enough ϵ, all K_n's are nonempty unless $c = 0$. □

Note 54. *The proof in Folland uses the Heine-Borel theorem, which was discovered exactly for this purpose!*

Theorem 55. *1. For any Borel measure μ, the function F in (27) is increasing and right-continuous.*

2. Conversely, for any increasing, right-continuous $F : \mathbb{R} \to \mathbb{R}$ there is a unique measure μ_F on $\mathcal{B}_{\mathbb{R}}$ s.t. for all a, b $\mu_F((a, b]) = F(b) - F(a)$. If G is a function as above s.t. or all a, b $\mu_F((a, b]) = G(b) - G(a)$, then $G - F$ is constant. The measure μ is complete on a σ-algebra containing $\mathcal{B}_{\mathbb{R}}$.

Proof. 1. See Exercise 18.

2. Proposition 53 shows that μ_F is a premeasure on $\mathcal{A}$. Since $\mathcal{A}$ generates $\mathcal{B}_{\mathbb{R}}$, Theorems 50 and 48 show that μ_F extends to a complete measure on a σ-algebra $\mathcal{M}$ containing $\mathcal{B}_{\mathbb{R}}$. Clearly, if G has the same properties, then $(F - G)(b) = (F - g)(a)$ for any finite a, b implying the result. □

The measure μ_F is called the Lebesgue-Stieltjes measure associated to F.

Note 56. *Since $\mu_F = \mu_F^*$ on $\mathcal{M}_\mu$ we have, for $E \in \mathcal{M}_\mu$,*

$$\mu_F(E) = \inf\left\{\sum_{i\in\mathbb{N}} [F(b_i) - F(a_i)] : E \subset \bigcup_{i\in\mathbb{N}} (a_i, b_i]\right\} = \inf\left\{\sum_{i\in\mathbb{N}} \mu((a_i, b_i]) : E \subset \bigcup_{i\in\mathbb{N}} (a_i, b_i]\right\} \tag{29}$$

Since for any $\epsilon > 0$ and any interval $I_j = (a_j, b_j]$ there is an open interval $J_j = (a'_j, b'_j) \supset I_j$ s.t.

[7] In words: if x is in all A_j, then either x is in all $\hat{A}_j$ or there is a j_0, $x \notin \hat{A}_{j_0}$ but since x is in all A_j, $x \in A_{j_0}$.

$\mu_F(J_j \setminus I_j) \leqslant 2^{-j}\epsilon$ *(check), it follows that for* $E \in \mathcal{M}_\mu$,

$$\mu_F(E) = \inf\left\{ \sum_{i\in\mathbb{N}} \mu((a_i, b_i)) : E \subset \bigcup_{i\in\mathbb{N}} (a_i, b_i) \right\} \tag{30}$$

Definition 57. *A Borel measure on a topological space* X *is regular if for any* $E \in \mathcal{B}_X$ *we have*

$$\inf\{\mu(\mathcal{O}) : E \subset \mathcal{O}, \mathcal{O} \text{ open}\} = \mu(E) = \sup\{\mu(F) : E \supset K, K \text{ compact}\} \tag{31}$$

It is outer regular if the first equality holds, and inner regular if the second one holds.

Lemma 58. *For any* $\epsilon > 0$ *and any* $E \in \mathcal{B}$ *there is an* $\mathcal{O} \supset E$ *open s.t.* $\mu(\mathcal{O} \setminus E) < \epsilon$.

Proof. We write $E = \cup_{n\in\mathbb{N}}(E \cap [-n, n])$ and let $\epsilon > 0$. Since $\mu(E \cap [-n, n])$ is finite, we can find an $\mathcal{O}_n$ s.t. $\mu(\mathcal{O}_n) \geqslant \mu(E) \geqslant \mu(\mathcal{O}_n) - \epsilon 2^{-n}$. The rest is straightforward. □

Theorem 59. *All Borel measures on* $\mathbb{R}$ *are regular.*

Proof. For outer regularity, we see that $E \subset \mathcal{O}$ implies $\mu(\mathcal{O}) \geqslant \mu(E)$ whereas Lemma 58 shows that for any E there is an $\mathcal{O}$ with measure arbitrarily close to $\mu(E)$.

Using σ-finiteness and an $\epsilon 2^{-n}$ argument, it is enough to show inner-regularity on bounded sets, E, for which clearly the measure is finite. For a given $\epsilon > 0$, find $\mathcal{O} \supset E^c$ s.t. $\mu(\mathcal{O} \setminus E^c) = \mu(\mathcal{O} \cap E) \leqslant \epsilon$. Now,

$$\mu(E) = \mu(\mathcal{O} \cap E) + \mu(\mathcal{O}^c \cap E) \leqslant \epsilon + \mu(\mathcal{O}^c)$$

Now $K \subset \mathcal{O}^c \subset E$ is compact, and

$$\mu(K) \leqslant \mu(E) \leqslant \epsilon + \mu(K)$$

□

Recall that an F_σ set is a countable union of closed sets; in $\mathbb{R}$ (and in $\sigma-$compact spaces) this is the same as a countable union of compact sets. A G_δ set is a countable intersection of open sets. In $\mathbb{R}$, F_σ sets are complements of G_δ sets.

Theorem 60. *Let* μ *be a Borel measure on* $\mathbb{R}$ *and* $\mathcal{M}_\mu$ *its associated* $\sigma-$*algebra and* $E \subset \mathbb{R}$. *The following properties are equivalent:*

1. $E \in \mathcal{M}_\mu$.
2. *There is an* F_σ *set* F *s.t.* $F \subset E$ *and* $\mu(E \setminus F) = 0$.
3. *There is a* G_δ *set* G *s.t.* $G \supset E$ *and* $\mu(G \setminus E) = 0$.

Proof. $2 \Rightarrow 1$ and $3 \Rightarrow 1$ follow from the completeness of the measure.

$1 \Rightarrow 2$ and $1 \Rightarrow 3$ follow from regularity: take a sequence $\epsilon_n \to 0$ and for each n pick $\mathcal{O}_n$ open and K_n compact s.t.

$$\mathcal{O}_n \supset E \supset K_n \quad \text{and} \quad \mu(O_n \setminus K_n) \leqslant \epsilon_n$$

Then the sets $G = \cap_n \mathcal{O}_n$ and $F = \cup_n K_n$ have the required properties. □

Set-theoretically, $\mathcal{B}_{\mathbb{R}}$ is of course much richer than the collection of F_σ and G_δ sets. Measures, as we see, cannot give justice to all these extra riches. The following is left as an easy exercise:

Proposition 61. *If E, μ and $\mathcal{M}_\mu$ are as above, $\mu(E) < \infty$ and $\epsilon > 0$, then there is a finite union of open intervals A s.t. $\mu(E\Delta A) < \epsilon$.*

Definition 62. *The Lebesgue measure on $\mathcal{B}_{\mathbb{R}}$ is the measure m induced by $F(x) = x$. The sets in the σ-algebra of m, $\mathcal{L}$, are called Lebesgue measurable. The translation of a set E by x_0, $\{x + x_0 : x \in E\}$, is denoted by $E + x_0$. The dilation of E by r, $\{rx : x \in E\}$ is denoted by rE.*

Since m is generated by the interval length, it is translation-invariant as the theorem below shows.

Theorem 63. *If $E \in \mathcal{L}$ then $E + x_0$ and rE are in $\mathcal{L}$ and*

$$m(E + x_0) = m(E); \;\; m(rE) = |r|m(E) \tag{32}$$

Proof. Translations and dilations commute with countable unions and complements (check). The algebra $\mathcal{A}$ of unions of half-open sets is invariant under translations and dilations, and (32) holds for intervals. It follows that $\mathcal{B}_{\mathbb{R}}$ is also invariant under translations and dilations, and$m$ satisfies (32) on $\mathcal{B}_{\mathbb{R}}$. Since the translation and dilation of a null set is a null set (why?), the result follows from Theorem 60. ☐

Clearly, countable sets have zero Lebesgue measure. There are many uncountable ones with measure zero, however. let's first look at the Lebesgue measure from a very different perspective.

4.1 Push-forward of a measure

Definition 64. *Let $(X_1, \mathcal{M}_1, \mu_1)$ be a measure space, $(X_2, \mathcal{M}_2)$ a measurable space and $f : X_1 \to X_2$ a measurable function. The pushforward measure $f_*(\mu)$ is defined as*

$$(f_*(\mu))(A) = \mu(f^{-1}(A)), \;\; A \in \mathcal{M}_2$$

Exercise 20. *Check that $(X_2, \mathcal{M}_2, (f_*(\mu))$ is a measure space.*

4.2 Coin tosses and the Lebesgue measure

A measure space $(X, \mathcal{M}, P)$ is called a probability space if $P(X) = 1$. The space X is called sample space, $\mathcal{M}$ is called the σ-algebra of events and P is the probability measure. $A \cup B$ is the event "A **or** B" and $A \cap B$ is the event "A **and** B". Two events, A and B are called independent if $P(A \cap B) = P(A)P(B)$.

If $(Y_\alpha, \mathcal{M}_\alpha, P_\alpha)$ are probability spaces, the product space $\otimes_\alpha Y_\alpha$ is endowed with the σ-algebra $\mathcal{M} = \otimes_\alpha \mathcal{M}_\alpha$ generated by the canonical projections. Finite intersections of sets of the form $C_\beta(A_\beta) = \pi_\beta^{-1}(A_\beta), A_\beta \in \mathcal{M}_\beta$ are called cylinder sets. Clearly, the family of cylinder sets generates $\mathcal{M}$. The product measure is generated by $P(C_\beta(A_\beta) \cap C_\gamma(A_\gamma)) = P_\beta(A_\beta)P_\gamma(A_\gamma)$- making events in different spaces independent of each-other. We will go through the details of the general construction later in the course. Here we focus on a particular case, relevant to the Lebesgue measure.

Coin tosses. In a single coin toss there are two possible outcomes, H or T, where H is head and T is tail. We let $X = \{H, T\}$. The σ-algebra of events is simply $\mathcal{M} = \mathcal{P}(X)$. The probability measure describing a fair coin is given by $P(\{H\}) = P(\{T\}) = \frac{1}{2}$.

(a) From now on we denote $H = 1, T = 0$. For n tosses of the coin, the underlying space is X^n, the set of all length-n sequences $(x_i)_{i=1...n}$ where $x_i \in \{0, 1\}$. The σ-algebra on X^n is $\mathcal{M}_n = \otimes_1^n \mathcal{M} = \mathcal{P}(X^n)$. The probability measure on $\mathcal{P}(X^n)$ describing independent coin tosses is the uniform measure $P(\{x\}) = 2^{-n}$ for any $x \in X$. Check that the probability that a sequence starts with $x_1 = 1$, "$P(x_1 = 1)$" is $1/2$, $P(x_1 = x_2) = 1/2$ and that the events $x_i = a, x_j = b$ are independent for $x \neq j$.

(b) For $n > m$, $\mathcal{M}_m$ is embedded in $\mathcal{M}_n$ as the σ-algebra generated by the cylinders $C_1, ..., C_m$. Check that the definition of P is consistent w.r.t. this embedding.

(c) The space of infinitely many coin tosses is $\Omega = \{0, 1\}^{\mathbb{N}} = \prod_{i \in \mathbb{N}} X_i$ where $\forall i, X_i = X$. The σ-algebra $\mathcal{M}$ on Ω is, as we know, generated by the canonical projections π_i. As before, $\mathcal{M}_n$ is embedded in Ω as the σ-algebra $\mathcal{M}_n'$ generated by $\pi_1, ..., \pi_n$. Check that $\mathcal{A} = \cup_n \mathcal{M}_n'$ is an algebra generating $\mathcal{M}$.

(d) Define the measure μ_0 on $\mathcal{A}$ as follows. If $A \in \mathcal{A}$, then $A \in \mathcal{M}_n'$ for some n (not unique), identified with an $A \in \mathcal{M}_n$. Let $\mu_0(A) = 2^{-n}\#(A)$ where $\#(A)$ is the counting measure. Check that the definition is compatible with the embeddings.

(e) Let $f : [0, 1) \to \Omega$ be defined as follows. If $0.a_1 a_2 \cdots$ is the binary representation of $x \in [0, 1)$, then

$$f(x) = (a_1, a_2, \cdots) \in \Omega$$

Check that f is measurable. Furthermore, if C is the cylinder defined by $x_0 = a_0, ..., x_k = a_k$, then $f^{-1}(C)$ is an interval of Lebesgue measure 2^{-k}. Show that $f_*(m)$ is the extension of P from $\mathcal{A}$ to $\mathcal{M}$. This f is injective but not surjective; the set $\Omega \setminus f((0, 1])$ is the set of sequences that end in an infinite string of zeros or of ones, a set of probability 0 (check).

(f) With this construction the Lebesgue measure on $[0, 1)$ is becomes probability measure on binary digits, treated as being independent. The measure of $\mathbb{Q} \cap [0, 1)$ is the probability of a sequence which becomes eventually periodic, zero (check).

HW for 09/28 : Problems 18–22 on p. 32 in Folland, and turn in: Ex 17–19 in these notes.

4.3 The Cantor set

The Cantor ternary set $\mathcal{C}$ is obtained by removing the open middle third from $[0, 1]$ and then successively removing the open middle from the remaining set of intervals. The Cantor ternary set consists of all remaining points in $[0, 1]$, those that are not removed at any step. Check that the Cantor set consists of all $x \in [0, 1]$ whose base 3 expansion consists of 0 and 2 only. Clearly, there is a surjection f from $\mathcal{C}$ to $[0, 1]$, by associating $x \in \mathcal{C}$ the number $f(x) \in [0, 1]$ whose binary expansion is obtained from the ternary expansion of x substituting a 1 for each 2. This shows that $\text{card}(\mathcal{C}) = \mathfrak{c}$. Check that $m(\mathcal{C}) = 0$. Using the probabilistic interpretation of m and the arithmetic interpretation of $\mathcal{C}$, this is obvious: the probability that 1 is missing from the first n ternary digits is $(2/3)^n$. The function f described above is known as Cantor's function.

The Cantor set, therefore, has empty interior: it cannot contain any interval of non-zero length. It may seem that only endpoints of intervals are left, but this is not the case. $0.020202\cdots = \frac{1}{4}$ is clearly in $\mathcal{C}$ yet it is not an endpoint of any middle segment, because it is not a multiple of any power of $1/3$. Of course, this follows from cardinality too, since the set of endpoints of removed intervals is countable.

Exercise 21. *In this exercise, $\mathcal{C}$ is the Cantor set and f is Cantor's function.*

1. *In (a) and (b): True or false? Explain.*

   (a) If F is an increasing, continuously differentiable function and μ_F is the Borel measure induced by F, then $\mu_F(\mathcal{C}) = 0$.

   (b) If F is an increasing function and there are $C > 0$ and $\alpha \in (0,1)$ s.t. $\forall x, y : |F(x) - F(y)| \leqslant C|x-y|^\alpha$ and μ_F is the Borel measure induced by F, then $\mu_F(\mathcal{C}) = 0$.

2. *Show that the interior of $\mathcal{C}$ is empty. What is the boundary of $\mathcal{C}$?*

3. *Let $F = f$ and μ_F the Borel induced measure. Find $\mu_F([0,1] \setminus \mathcal{C})$.*

4.4 Cantor's function (a.k.a Devil's staircase)

The Cantor function has a good number of surprising properties. It is clearly increasing, and $f(x_1) = f(x_2)$ iff $x_1 = .0...0_n222\cdots$, $x_2 = 0.0...0_{n-1}2$. We extend f by a constant on $[x_1, x_2]$, and the extended f is defined on $[0,1]$ with values in $[0,1]$. Note that $f([0,1]) = [0,1]$ and f is continuous. Cantor's function was presented as a counterexample to an (incorrect) extension of the fundamental theorem of calculus claimed by Harnack. Indeed, f is differentiable almost everywhere with zero derivative (check). f is flat almost everywhere, yet somehow manages to continuously increase from zero to one. If we take $F = f$ in our construction of Borel measures, it gives rise to a continuous measure that is singular with respect to m (definitions will come later).

Exercise 22. *Is there any Borel measure on $\mathbb{R}$ (a measure on the Borel sets of $\mathbb{R}$) which is finite on compact sets) for which the Borel sets of measure zero are exactly the countable sets? (One possibility is the following. For $x = 0.a_1a_2... \in (0,1)$ let $A_x = \{x = 0.b_1a_1b_2a_2, ... : 0.b_1b_2... \in (0,1)\}$. These sets are uncountably many disjoint sets, their union is $(0,1)$, and each of them is uncountable.)*

5 Integration

The starting point will be the functions for which we already have a good candidate for the integral: characteristic functions (whose integral should equal the measure of the set) and from here, of course, linear combinations of characteristic functions of bounded sets.

5.1 Measurable functions (cont.)

Proposition 65. *If $X_i, \mathcal{M}_i, i = 1, ..., n+1$ are measurable spaces and $f_i : X_i \to X_{i+1}, i = 1, ..., n$ are measurable, then so is the composition $f_n \circ \cdots \circ f_1$.*

Proof. Straightforward verification of Definition 34. □

Proposition 66. *Let X, Y be topological spaces with the Borel σ-algebras. Any continuous function from X to Y is measurable.*

Proof. By definition, the inverse image of open sets is open, and open sets generate $\mathcal{B}_X$ and $\mathcal{B}_Y$. $\square$

Exercise 23. *Show that $A \in \mathbb{R}$ is Borel measurable* **iff** *χ_A is Borel measurable.*

Definition 67. Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{R}$. f is called measurable if $f^{-1}(\mathcal{B}_{\mathbb{R}}) \subset \mathcal{M}$. An important particular case is $(X, \mathcal{M}) = (\mathbb{R}, \mathcal{L})$, in which case f is called Lebesgue measurable.

Note 68. If $A \in \mathcal{L}$, then $A = B \cup N$ where B is a Borel set and $m(N) = 0$. If $f : \mathbb{R} \to \mathbb{R}$, is Borel measurable, then $f^{-1}(A) \in \mathcal{L}$ for any $A \in \mathcal{L}$ **iff** $f^{-1}(N)$ is measurable for every null set (set of Lebesgue measure zero) N. This is not necessarily the case even if f is continuous, as the next note shows. There we construct such a function which bijectively and bicontinuously maps an uncountable null set to a set of measure zero. Then a nonmeasurable set is bijectively and bicontinuously mapped into a set of measure zero.

This means that a composition of Lebesgue measurable functions need not be Lebesgue measurable. Examine carefully all these definitions.

Exercise 24. *Let f be continuous and strictly increasing from $\mathbb{R}$ to $\mathbb{R}$. Then f maps Borel sets to Borel sets.*

Note 69 (Relation to the axiom of choice)**.** ZF is consistent with the statement "$\mathbb{R}$ is a countable union of countable sets". Therefore, there are models of ZF where the Lebesgue measurable sets are exactly the Borel sets. Consequently also, in such models the theory of Lebesgue measure can fail totally. A weak form of the AC guarantees that a countable union of countable sets is countable, and rules out the quoted statement. This is the axiom of countable choice, stating that there is a choice function for any countable family of sets. It is weaker than the axiom of dependent choice. [8] The axiom of dependent choice is considered more benign than the full AC, in that no spectacularly counterintuitive result (such as the Banach-Tarski paradox) exists based on it.

Note 70. Here we construct a continuous bijection from $[0, 2]$ to $[0, 1]$ such that *$h^{-1}(\mathcal{C})$ has positive measure*. We start from the Cantor function f. It is not a bijection, but $g := x \mapsto f(x) + x$ applies bijectively $[0, 1]$ to $[0, 2]$. The forward image of $\mathcal{C}$ is $C = \mathcal{C} + [0, 1]$, a set of measure 1. The function $h = g^{-1}$ has the emphasized property above. Let E now be a nonmeasurable set in C (how do we know it must exist?). Then $h : C \to \mathcal{C}$. Any subset of $\mathcal{C}$ has measure zero, and one of these, say N_1, must have the property $h^{-1}(N_1) = E$.

Definition 71 (Measurability on a set)**.** *Let $E \in \mathcal{M}$ and $f : E \to (Y, \mathcal{N})$. f is called measurable on E if it is measurable from $(E, \mathcal{M}_E)$ to $(Y, \mathcal{N})$, where $\mathcal{M}_E = \{E \cap A : A \in \mathcal{M}\}$.*

The proofs of Propositions 72–77 are straightforward and left as an exercise.

Proposition 72. *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{R}$. Then the following are equivalent:*

1. *f is measurable.*

[8] The axiom of dependent choice states the following: Let $\mathcal{R}$ be a binary relation on a non-empty set S. Suppose that $\forall a \in S \exists b \in S : a\mathcal{R}b$. Then there exists a sequence in S, $(x_n)_{n\in\mathbb{N}}$ s.t $\forall n \in \mathbb{N} : x_n \mathcal{R} x_{n+1}$. This axiom is equivalent to the Baire category theorem for complete metric spaces.

2. *For any* $a \in \mathbb{R}$, $f^{-1}((a,\infty))$ *is measurable.*

3. *For any* $a \in \mathbb{R}$, $f^{-1}([a,\infty))$ *is measurable.*

4. *For any* $a \in \mathbb{R}$, $f^{-1}((-\infty,a))$ *is measurable.*

5. *For any* $a \in \mathbb{R}$, $f^{-1}((-\infty,a])$ *is measurable.*

Exercise 25. *Choose a convenient characterization from the list above and show that any increasing function from* $\mathbb{R}$ *to* $\mathbb{R}$ *is measurable.*

Definition 73. If X is a set, $(Y_\alpha, \mathcal{M}_\alpha)_{\alpha\in A}$ are measurable spaces and $(f_\alpha)_{\alpha\in A}$ are functions from X to Y_α, then the σ-algebra generated by $(f_\alpha)_{\alpha\in A}$ is the smallest σ-algebra in X s.t. all f_α, $\alpha \in A$ are measurable. An example is the product space $Y = \otimes_\alpha Y_\alpha$ and the canonical projections π_α: they generate the product σ-algebra $\mathcal{M}$.

Proposition 74. *Let* $(X\mathcal{M})$ *be a measurable space, and* $Y, \mathcal{M}, Y_\alpha, \mathcal{M}_\alpha$ *be as in Definition 73. Then* $f : X \to Y$ *is measurable iff* $\pi_\alpha \circ f$ *is measurable for any* α *(i.e.,* f *is measurable iff it is componentwise measurable).*

For some purposes it is convenient to consider functions with values in $[-\infty,\infty]$. This is equivalent to letting $g = \tanh \circ f$ and allowing for the range of g to be $[-1,1]$. The arithmetic disallows for $\infty - \infty$ but allows for $0\cdot\infty$ defined to be zero.

Proposition 75. *The functions "+":* $(x,y) \mapsto x+y$ *and "·":* $(x,y) \mapsto xy$ *are measurable. If* $f,g : \mathbb{R} \to \mathbb{R}$ *are measurable so are* $(f,g) : \mathbb{R} \to \mathbb{R}^2$, $f+g = {}''+''((f,g))$ *and* $fg = {}''\cdot''((f,g))$.

Proposition 76. *If* $f : X \to \mathbb{R}$ *is measurable, then so is* $|f|$. *If* $f,g : X \to \mathbb{R}$ *are measurable, then so are* $f \vee g = \max\{f,g\} = \frac{1}{2}|f-g| + \frac{1}{2}(f-g)$, $f \wedge g = \min\{f,g\}$, $f^+ = f \vee 0$ *and* $f^- = f \wedge 0$. *The functions sgn=* $\chi_{[0,\infty)} - \chi_{(-\infty,0]}$ *and csgn=* $z/|z|\chi_{|z|>0}$, *are measurable.*

Let $(g_i)_{i\in\mathbb{N}}$ from X to $\mathbb{R}$ be measurable. Let $\inf_i g_i =: g$, $\liminf_{n\in\mathbb{N}} g_n = h$. Then

$$\{x : g(x) \geqslant a\} = \cap_{i\in\mathbb{N}}\{x : g_i(x) \geqslant a\} \text{ and } \{x : h(x) \geqslant a\} = \underbrace{\cup_n \cap_{i\geqslant n}}_{\exists n:\forall i\geqslant n}\{x : g_i(x) \geqslant a\}$$

Proposition 77. *Let* $(g_i)_{i\in\mathbb{N}}$ *from* X *to* $\mathbb{R}$ *be measurable. Then so are* $\inf_i g_i, \sup_i g_i, \liminf_i g_i$ *and* $\limsup_i g_i$. *If* $G(x) = \lim_{i\to\infty} g_i(x)$ *exists for all* x, *then* G *is measurable.*

(for the last statement note that the limit, when it exists, coincides with limsup).

Exercise 26. *Extend, where possible, these results to functions defined on* X *with values in* $\mathbb{C}$.

Note 78. *A measurable function* f *between a probability space* $(X, \mathcal{M}, P)$ *and a measure space* $(Y, \mathcal{N}, \mu)$ *is called a random variable. If* $Y = \mathbb{R}$, *then* $F_f(x) := P(f \leqslant x)$ *is the cumulative distribution function.*

Here is a probabilistic interpretation of the Cantor function. In base 3, start with the initial string "0.". At each $n \in \mathbb{N}$ flip a coin. If the result is H, then append a 2 to the previous string, otherwise append a zero. The probability that the resulting number is $\leqslant x$ is $f(x)$. This is made precise in the following exercise.

Exercise 27. *In §4.2 replace "1" by "2" in all sequences and sequence spaces.*

(a) With this interpretation, show that there is a bijection between Ω and the Cantor set. (The image through this bijection of the measure P that we constructed on $\mathcal{C}$ is a uniform measure on $\mathcal{C}$.)

(b) The identity map restricted to $\mathcal{C}$, J, is measurable relative to $\mathcal{C}$, thus a random variable. Show that the cumulative distribution function for J is the Cantor function.

There is an equivalent jump process (with discrete time $n \in \mathbb{N}$). A particle sits in the center of the middle third interval. Right before the interval is removed, it randomly jumps away with equal probability to the middle of the right or middle of the left interval. And so on. The probability of its eventual location point being $\leqslant x$ is $f(x)$.

5.2 Simple functions

Definition 79. *Let $(X, \mathcal{M})$ be a measurable set. A measurable function from X to $\mathbb{C}$ which has discrete range, $\{z_1, ..., z_n\} \subset \mathbb{C}$ is called a simple function. Let $A_1, ..., A_n$ be measurable sets in X and $z_1, ..., z_n$ complex numbers. Then the linear combination*

$$\sum_{j=1}^{n} z_j \chi_{A_j} \tag{33}$$

is clearly a simple function if $\cup A_j = X$. We convene that, if one of the z_i happens to be zero, we keep a term $0 \cdot \chi_{A_i}$ in (33).

We denote the space of simple functions by $\mathfrak{S}$.

An analogy with counting the money in a jar with coins is often used to illustrate the fundamental difference between Riemann integration and Lebesgue integration. One method is to take the coins out one by one and add the values as we go. The second one is to take out all the coins, sort them by value, count the number of coins in each pile, multiply by the value and then calculate the total. The first method corresponds to Riemann integration, while the second one to Lebesgue. Mathematically the difference is partitioning the domain or the range of a function.

Theorem 80.

1. *Let $f : X \to [0, \infty]$ be measurable. There is an increasing sequence $(f_i)_{i\in\mathbb{N}}$ in $\mathfrak{S}$ pointwise convergent to f, uniformly so on any set where f is bounded.*
2. *Let $f : X \to \mathbb{C}$ be measurable. There is a sequence $(f_i)_{i\in\mathbb{N}}$ in $\mathfrak{S}$, such that $(|f_i|)_{i\in\mathbb{N}}$ is an increasing sequence, and $f_n \to f$ pointwise everywhere, and uniformly on any set where f is bounded.*

Proof. 1. For each $n \in \mathbb{N}$ partition the interval $[0, 2^n]$ in the range of f in 2^{2n} left-open-right-closed intervals $(J_{nk})_{k=1,...,2^{2n}}$ of length 2^{-n}. Let $v_{n,k}$ be the left end of J_{nk}, $A_{nk} = f^{-1}(J_{nk}), B_n = f^{-1}((2^n, \infty])$ and define

$$f_n = \sum_{k\leqslant 2^n} v_{nk}\chi_{A_{nk}} + 2^n \chi_{B_n} \in \mathfrak{S}$$

Pointwise convergence is immediate. Let A be a set where f is bounded. Then, for some n_0 and all $n > n_0$ we have $A \subset B_n^c$. By construction, on $B_{n_0}^c$, $|f - f_n| \leqslant 2^{-n}$.

2. We write $f = (\Re f)^+ - \Re(f)^- + i(\Im f)^+ - (\Im f)^-$. The result follows by applying 1. to each term above. □

Proposition 81. *Assume $(X, \mathcal{M}, \mu)$ is a measure space and μ is complete. Assume $g, (f_n)_{n\in\mathbb{N}}$ are measurable from X to $\mathbb{R}$. Then*

1. *If $f : X \to \mathbb{R}$ and $f = g$ a.e., then f is measurable.*

2. *If $f_n \to f$ pointwise a.e., then f is measurable.*

Proof. Straightforward. □

Proposition 82. *Let $(X, \overline{\mathcal{M}}, \overline{\mu})$ be the completion of $(X, \mathcal{M}, \mu)$ and assume f is $\overline{\mathcal{M}}$-measurable. Then there exists an $\mathcal{M}$-measurable g which coincides with f a.e.*

Proof. For characteristic functions this property is clear from Theorem 44, and it extends by linearity $\mathfrak{S}$. Let $(\phi_n)_{n\in\mathbb{N}}$ in $\mathfrak{S}$ be a sequence converging pointwise to f. Choose a sequence of $\mathcal{M}$-measurable functions $(\psi_n)_{n\in\mathbb{N}}$ which coincide with $(\phi_n)_{n\in\mathbb{N}}$ except on some null sets $(N_n)_{n\in\mathbb{N}}$. Let $N = \cup_{n\in\mathbb{N}} N_n$. Then the sequence $(\chi_{X\setminus N}\psi_n)_{n\in\mathbb{N}}$ converges pointwise everywhere, thus to a measurable function, and the limit equals f on $X \setminus N$. □

HW 08/08 : 32,33 on p. 40, 8,10 on pp. 48,49 in Folland; turn in: Ex 21,22 in the notes.

5.3 Integration of positive functions

Lemma 83. *Let $(X, \mathcal{M})$ be a measurable space and μ, ν measures on $(X, \mathcal{M})$. Then $\mu + \nu$ and $c\mu$ are measures on $(X, \mathcal{M})$ for any $c \geqslant 0$.*

Proof. Straightforward verification. □

In this section the space $(X, \mathcal{M}, \mu)$ is fixed. Let L^+ be the convex cone of nonnegative measurable functions:

$$L^+ = \{f \in \mathcal{M} : f \geqslant 0\}$$

Let $\phi \in L^+ \cap \mathfrak{S}$. Then, for some $n \in \mathbb{N}$, $\operatorname{ran}(\phi) = \{a_1, ..., a_n\} \subset [0, \infty)$ and

$$\phi = \sum_{j=1}^{n} a_j \chi_{A_j}; \quad A_j := f^{-1}(\{a_j\}) \tag{34}$$

It is natural to define the integral of ϕ by

$$\int \phi d\mu = \sum_{j=1}^{n} a_j \mu(A_j) \tag{35}$$

where, as usual $0 \cdot \infty = 0$. Other notations are $\int \phi(x) d\mu(x)$, $\int \phi(x)\mu(dx)$ or simply $\int \phi$ when the context is clear. Likewise, when $A \in \mathcal{M}$ we define

$$\int_A \phi d\mu = \int \chi_A \phi d\mu \tag{36}$$

Proposition 84. *Let $\phi = \sum_{i=1}^{n} a_i \chi_{A_i}, \psi = \sum_{i=1}^{m} b_j \chi_{B_i} \in L^+ \cap \mathfrak{S}$. Then*

1. *(Compatibility with the cone structure) $\int \phi d\mu \geqslant 0$, $\forall c \geqslant 0 : \int c\phi = c \int \phi$ and $\int(\phi + \psi) = \int \phi + \int \psi$.*

2. *$A \mapsto \int_A \phi$ is a measure on $\mathcal{M}$.*

Proof. 1. Nonnegativity and multiplicativity by constants are clear. Linearity follows easily if we note that the range of $\phi + \psi$ is $\{a_i + b_j : 1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m\}$ ($a_i + b_j$ are not necessarily distinct), and that these values are taken on the disjoint sets $C_{ij} = A_i \cap B_j$, $1 \leqslant i \leqslant n, 1 \leqslant j \leqslant m$.

2. When $\phi = \chi_B$ for some measurable B, $\int \phi = \mu(A \cap B)$ which is a measure on $\mathcal{M}$. The rest follows from Lemma 83. □

Note that 1. implies

$$\phi \leqslant \psi \Rightarrow \int \phi d\mu \leqslant \int \psi d\mu$$

Definition 85. *If $f \in L^+$, we define*

$$\int f d\mu = \sup_{\substack{\phi \in \mathfrak{S} \cap L^+ \\ \phi \leqslant f}} \int \phi d\mu$$

Proposition 86. *Def. 85 coincides with (36) for $f \in \mathfrak{S} \cap L^+$. The integral is nonnegative, commutes with the cone operations (cf. Proposition 84, 1.), addition and multiplication by nonnegative numbers.*

Proof. For additivity, see Theorem 88 below. The rest is straightforward. □

Exercise 28. *Show multiplicativity with a constant when $c = +\infty$.*

The first important theorem about the properties of the integral is

Theorem 87 (The monotone convergence theorem)**.** *If $(f_n)_{n \in \mathbb{N}}$ is an increasing sequence in L^+, then*

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

Proof. The limit (possibly $+\infty$) $\lim_{n \to \infty} \int f_n(x) = f(x)$ clearly exists for any $x \in X$, and since $\forall n : f \geqslant f_n$, we have

$$\lim_{n \to \infty} \int f_n d\mu \leqslant \int f d\mu$$

For the opposite inequality choose $\alpha \in (0,1)$ and a $\phi \leqslant f$ in $\mathfrak{S} \cap L^+$ s.t. $\alpha \int f d\mu \leqslant \int \phi d\mu$. By monotonicity, the sets $A_n = \{x \in X : f_n(x) \geqslant \alpha\phi\}$ are measurable and increasing, and since $f_n \to f$, $A_n \nearrow X$. Since $\alpha < 1$ is arbitrary, using monotonicity of the integral and sequence, the result follows from

$$\lim_{n \to \infty} \int_X f_n d\mu \geqslant \lim_{n \to \infty} \int_{A_n} f_n d\mu \geqslant \alpha \lim_{n \to \infty} \int_{A_n} \phi d\mu = \alpha \int_X \phi d\mu \geqslant \alpha^2 \int f d\mu$$

□

Theorem 88. 1. *The integral is additive on L^+.*

2. *If $(f_n)_{n \in \mathbb{N}}$ is a sequence in L^+, then*

$$\int \sum_{n \in \mathbb{N}} f_n \, d\mu = \sum_{n \in \mathbb{N}} \int f_n \, d\mu$$

Proof. 1. We have already shown linearity on $\mathfrak{S} \cap L^+$. We can use approximation by simple functions and the monotone convergence theorem to prove the rest. If $(\phi_n)_{n\in\mathbb{N}}, (\psi_n)_{n\in\mathbb{N}}$ increase to f and g respectively as in Theorem 80, then $\phi_n + \psi_n \nearrow f + g$, and by dominated convergence

$$\int f d\mu + \int g d\mu = \lim_{n\to\infty} \left(\int \phi_n d\mu + \int \psi_n d\mu \right) = \lim_{n\to\infty} \int (\phi_n + \psi_n) d\mu = \int (f+g) d\mu$$

2. An application of the monotone convergence theorem.

□

Theorem 89. *For $f \in L^+$, $\int f d\mu = 0$ iff $f = 0$ a.e.*

If $f = 0$ a.e. and $0 \leqslant \phi \leqslant f$ then clearly $\phi = 0$ a.e. implying (check) $\int \phi d\mu = 0$. If $\int f d\mu = 0$, consider the disjoint sets $A_0 = f^{-1}(\{0\})$ and $A_n = f^{-1}((n^{-1}, (n-1)^{-1}]), n \in \mathbb{N}$. We have $\sum_{n+1\in\mathbb{N}} \chi_{A_n} = 1$ and, by monotone convergence,

$$0 = \int f d\mu = \sum_{n\in\mathbb{N}} \int f \chi_{A_n} d\mu \geqslant \sum_{n\in\mathbb{N}} n^{-1} \mu(A_n)$$

implying that $\mu(A_n) = 0$ for all n and thus $f = 0$ a.e.

Corollary 90. *Assume $(f_n)_{n\in\mathbb{N}}$ are in L^+ and increase a.e. to f. Then, by monotone convergence,*

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

Proof. Let $A = \{x \in X : \lim_n f_n(x) = f(x)\}$. We have $\mu(A^c) = 0$ (in particular A is measurable). Then

$$\int f(1-\chi_A) d\mu = \lim_{n\to\infty} \int f_n (1-\chi_A) d\mu = 0$$

and the result follows from Theorem 89. □

The second important result is the following.

Theorem 91 (Fatou's lemma). *If $(f_n)_{n\in\mathbb{N}}$ is a sequence in L^+, then*

$$\liminf_{n\in\mathbb{N}} \int f_n d\mu \geqslant \int \liminf_{n\in\mathbb{N}} f_n d\mu$$

Proof. Let $f = \liminf_n f_n$ and $g_n = \inf_{k\geqslant n} f_k$. We have $g_n \nearrow f$ and $g_n \leqslant f_n$ and thus, for all n we have, by monotone convergence,

$$\int f d\mu = \lim_n \int g_n d\mu \leqslant \int f_n d\mu \Rightarrow \liminf_{n\in\mathbb{N}} \int f_n d\mu \geqslant \int f d\mu$$

□

Here is a useful illustration of what may go wrong to make the inequality strict (Rudin) Let $X = [0,2]$ and $E = (1,2]$, and for $n \in \mathbb{N}$ let

$$f_n = \begin{cases} \chi_{(1,2]}, & \text{if } n \text{ is even} \\ \chi_{[0,1]}, & \text{if } n \text{ is odd} \end{cases} \tag{37}$$

Note that $\liminf_n f_n = 0$. However, for all n we have

$$\int_{[0,2]} f_n = 1 > \int_{[0,2]} \liminf_n f_n = 0$$

The following two results are left as simple exercises.

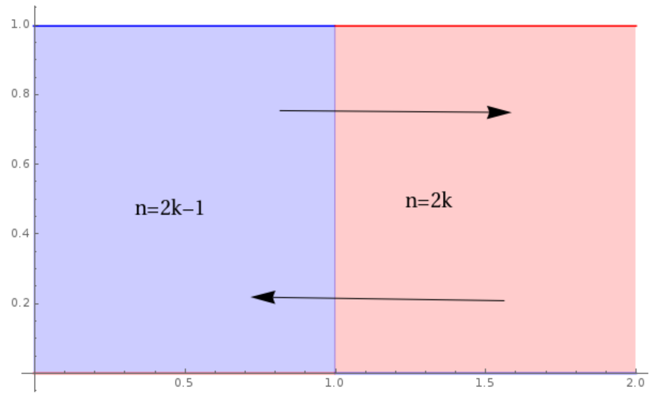

Figure 2: The sequence in (37).

Proposition 92. *If $(f_n)_{n\in\mathbb{N}}$ are functions in L^+ and $f_n \to f$ a.e., then $\int f d\mu \leqslant \liminf_n \int f_n d\mu$.*

Proposition 93. *If $f \in L^+$ and $\int f d\mu < \infty$, then $\{x : f(x) = \infty\}$ is a null set and $\{x : f(x) > 0\}$ is sigma-finite.*

6 Integration of complex-valued functions

As before, we fix $(X, \mathcal{M}, \mu)$. Consider now functions $f : X \to \mathbb{C}$ (in this setting $\infty \notin \mathrm{ran}(f)$).

Definition 94. *If $f : X \to \mathbb{C}$ define*

$$\int f d\mu = \left(\int (\Re f)^+ d\mu - \int (\Re f)^- d\mu\right) + i\left(\int (\Im f)^+ d\mu - \int (\Im f)^- d\mu\right)$$

The function f is said to be integrable if all four integrals above are finite. Equivalently, f is integrable if $\int |f| d\mu < \infty$. More generally, if $A \subset X$, then f is integrable on A if $f\chi_A$ is integrable, that is, $\int_A |f| < \infty$.

Proposition 95. *The set of integrable functions is a vector space and the integral is a linear, complex-valued functional on it.*

Proof. Multiplicativity by scalars is straightforward. Assume f, g are integrable and let $h = f + g$. Since $|h| \leqslant |f| + |g|$, h is integrable. To show linearity, it is enough to show linearity of the real part and imaginary part separately, and clearly the same argument applies for both, reducing the question to that of real-valued functions. Here we use a simple useful trick to obtain linearity linearity from cone additivity. Let C be a convex cone over a vector space V with the property that any $v \in V$ can be written uniquely as $v^+ - v^-$, $v^+, v^- \in C$. Let ϕ be compatible with the structure of C. In the setting at hand, $v^+ = f^+\chi_{f\geqslant 0} - f^-\chi_{f\leqslant 0}$ (other decompositions amount to the same since $f^+(x) = f^-(x) \Rightarrow f(x) = 0$). Extend ϕ to V by $\phi(v) = \phi(v^+) - \phi(v^-)$. Additivity on C now translates into additivity on V [9]. □

Proposition 96. *If f is integrable, then*

$$\left|\int f d\mu\right| \leqslant \int |f| d\mu$$

Proof. Let $\alpha = \operatorname{csgn}(\int f)$, $\beta = \overline{\alpha}$ and $g = \Re(\beta f)$. Then,

$$\left|\int f\right| = \beta \int f = \Re\left(\beta \int f\right) = \int \Re(\beta f) = \int g = \int g^+ - \int g^- \leqslant \int g^+ + \int g^- = \int |g| \leqslant \int |f|$$

□

Exercise 29. *Check that $f = 0$ a.e. iff $\int |f| = 0$.*

We see that, insofar as the theory of integration goes, two functions that differ on a null set are indistinguishable. It is then natural to work with the equivalence classes of integrable functions modulo values on null sets, rather than with individual functions.

Definition 97. *Denote $f \sim g$ iff $f - g = 0$ a.e. We define $L^1 = L^1(\mu) = L^1(\mu, X)$ to be the vector space of equivalence classes of integrable functions from X to $\mathbb{C}$.*

Clearly, an equivalence class of functions is not a function. However, it is standard practice to still call the elements of L^1 functions, and make the distinction explicitly when (rarely) this is needed. If we are dealing with the equivalence class of a continuous (or monotonic, or smooth etc.) function, there is a natural representative of that class and working with the classes or with the representatives is the same. If there is no way to naturally pick an element of the class, it doesn't matter much which one is referred to anyway. Note that, by Proposition 82 any equivalence class contains a Borel measurable function (still nonunique). Another advantage of working with equivalence classes is the following:

Proposition 98. *L^1 is a normed vector space with*

$$\|f\|_1 = \int |f|$$

Proof. This is an easy exercise. □

Definition 99. *If $(f_n)_{n\in\mathbb{N}}$ is an L^1 sequence, we say that $f_n \to f$ a.e., if for some representatives of f_n and f, the sequence of functions $(f_n)_n$ converges to f a.e.*

Exercise 30. *Check that this definition implies that convergence a.e. holds regardless of the choices of representatives.*

Proposition 100. 1. *If $f \in L^1$, then the set $\{x : F(x) \neq 0\}$ is σ-finite for any $F \in f$.*

[9] Uniqueness of the decomposition is not needed. Instead, one can check consistency of the definition: If $v^+ - v^- = w^+ - w^-$, then $v^+ + w^- = w^+ + v^-$, hence $\phi(v^+) + \phi(w^-) = \phi(v^+ + w^-) = \phi(w^+ + v^-) = \phi(w^+) + \phi(v^-)$ immediately implying consistency

2. *If $f \in L^1$, then $\forall A : \int_A f d\mu = 0$ iff $f = 0$ a.e.*

Proof. 1. Follows immediately from Proposition 93.

2. $f = 0$ a.e. implies $\chi_A f = 0$ a.e. for all measurable A. Conversely, if $\int_A f = 0$ for all A, then $g = \Re f$ and $h = \Im f$ have the same property. Define again the disjoint sets $A_n = \{x : g^+(x) \in (n^{-1}, (n-1)^{-1}]\}$. Since $g^- = 0$ on $A = \cup A_n$, $f = f^+$ on A. Then,

$$\int_A f(x) = 0 \geqslant \sum_{n \in \mathbb{N}} n^{-1} \mu(A_n)$$

We thus have $\mu(A) = 0$ and $g^+ = 0$ a.e.; similarly $g^- = 0$ a.e. □

Theorem 101 (The dominated convergence theorem)**.** *Assume the L^1 sequence $(f_n)_{n \in \mathbb{N}}$ converges a.e. to f and there is a $g \in L^1$ such that $\forall n : |f_n| \leqslant g$. Then $f \in L^1$ and*

$$\int |f_n - f| d\mu \to 0 \text{ and thus } \int (f_n - f) \to 0 \Leftrightarrow \int f_n \to \int f$$

Proof. Since $f_n(x) \to f(x)$, we have $|f(x)| \leqslant g(x)$ implying $|f - f_n| \leqslant 2g$ a.e. Since $\limsup_n |f(x) - f_n(x)| = 0$ a.e., Fatou's Lemma implies

$$\int 2g d\mu = \int \liminf_{n \in \mathbb{N}} (2g - |f - f_n|) d\mu \leqslant \int 2g d\mu - \limsup_{n \in \mathbb{N}} \int |f - f_n| d\mu$$

implying the result. □

Proposition 102. *Assume $(f_j)_{n \in \mathbb{N}}$ is an L^1 sequence s.t. $\sum_{n=1}^{\infty} \int |f_n| d\mu < \infty$. Then, $\sum_{n=1}^{\infty} f_j$ converges a.e. to an L^1 function f, and $\int \sum_{n \in \mathbb{N}} f_j = \int f$.*

Proof. Take $g = \sum_{n \in \mathbb{N}} |f_n|$. The rest is an easy exercise. □

Theorem 103. 1. *(Density of simple functions in L^1) For any $f \in L^1$ and any $\epsilon > 0$ there is an L^1-simple function ϕ s.t.*

$$\int |f - \phi| d\mu < \epsilon$$

2. *If μ is a Borel measure on $\mathbb{R}$, then ϕ can be chosen of the form $\sum a_n \chi_{J_n}$, a finite sum, where the J_n are finite unions of open intervals.*

3. *(Density of $C_c(\mathbb{R})$ in $L^1(\mathbb{R})$) If μ is a Borel measure on $\mathbb{R}$, then, for any $f \in L^1$, there is a continuous function g with compact support s.t.*

$$\int |f - g| d\mu < \epsilon$$

Proof. 1. Let $(\phi_n)_{n \in \mathbb{N}}$ be a sequence of simple functions converging to f, as in Theorem 80. Then $|\phi_n| \leqslant f$ and dominated convergence implies $\lim_{n \to \infty} \int |\phi_n - f| d\mu = 0$. Thus, for any $\epsilon > 0$ there is an n s.t., for $\phi = \phi_n$, $\int |\phi - f| d\mu < \epsilon$

2. Let ϕ be as above, $\epsilon > 0$, and write $\phi = \sum a_j \chi_{A_j}$. The statement follows from the fact that, by Proposition 61 for any $\epsilon > 0$ and any j there is an open set $\mathcal{O}$ which is a finite union of intervals s. t. $\int |\chi_{A_j} - \chi_{\mathcal{O}}| d\mu = \mu(\mathcal{O} \Delta A_j) < \epsilon / j$.

3. For each interval J and any $\epsilon > 0$ there is a continuous function g s.t. $\int |g - \chi_J| < \epsilon$ (construct such a function). □

Exercise 31. *Derive the monotone convergence theorem from the dominated convergence theorem.*

7 The link with the Riemann integral

Riemann integration can be recast in terms of the Jordan content (or Jordan measure; however, it is only finitely additive). Consider as "simple sets" **finite** unions of intervals. For the purpose of Riemann integration, the intervals, $\mathcal{J}_n = (I_k)_{\mathbb{N}\ni k\leqslant n}$ will constitute a partition of some fixed interval, $[a, b]$. Consider the family of simple functions

$$\mathfrak{S}_R := \left\{ \sum_{k=1}^{n} a_k \chi_{I_k} : I_k \in \mathcal{J}_n, n \in \mathbb{N} \right\}$$

Definition 104. *A bounded function f on an interval $[a, b]$ is Riemann integrable if*

$$\sup_{\phi\leqslant f;\ \phi\in\mathfrak{S}_R} \int_a^b \phi dx = \inf_{\phi\geqslant f;\ \psi\in\mathfrak{S}_R} \int_a^b \psi dx \tag{38}$$

The common limit, when it exists, is the Riemann integral $\int_a^b f(x)dx$. Here $\int_a^b \phi dx = \sum_{k\leqslant n} a_k(x_{k-1} - x_k)$ where the x_ks are the endpoints of the intervals J_k.

Note 105. *Letting $a_k = \inf_{I_k} f$ in the decomposition of the functions ϕ and $b_k = \sup_{I_k} f$ in the decomposition of the functions ψ, we recognize the usual definition of the Riemann integral.*

Theorem 106. *1. If f is Riemann integrable on $[a, b]$ then it is in L^1 and $\int_a^b f(x)dx = \int_{[a,b]} f dm$.*

2. f is Riemann integrable iff *it is continuous except on a null set.*

Proof. 1. As usual, we can construct an increasing sequence $(\phi_n)_{n\in\mathbb{N}}$ and a decreasing sequence $(\psi_n)_{n\in\mathbb{N}}$ for which the integrals converge to the same limit. By monotonicity and boundedness, these two sequences are pointwise convergent on $[a, b]$, say to ϕ and ψ resp, and $\psi - \phi \geqslant 0$. Since for all n, $|\phi_n| \leqslant |f| + |\phi_1|$ and $|\psi_n| \leqslant |f| + |\psi_1|$, dominated convergence applies and

$$\int_{[a,b]} \phi dm = \int_{[a,b]} \psi dm = \int_a^b f dx$$

Since $\int_{[a,b]} |\psi - \phi| dm = \int_{[a,b]}(\psi - \phi)dm = 0$, we have $\phi = f = \psi$ a.e., f is measurable, and $\int_{[a,b]} f dm = \int_{[a,b]} \phi dm = \int_a^b f dx$.

2. Take the ϕ_k, ψ_k as in Note 105. Note that there must exist a sequence of partitions P_k of $[a, b]$ such that, as $n \to \infty$ we have $\sup_{I_{nk}} |\phi_n - \psi_n| \to 0$ a.e., which implies continuity a.e. (work out the details of this and its converse; see also Exercise 23 in Folland). □

Exercise 32. (Dominated and monotone convergence failure for Riemann integration) *Find a monotone sequence of Riemann integrable functions converging to $\chi_{\mathbb{Q}}$. Can such a sequence consist of continuous functions?*

Remark 107. 1. *The Lebesgue integral is a proper extension of the Riemann integral. Hence the often used notation $\int_a^b f(x)dx$ for $\int_{[a,b]} f dm$.*

2. *Whenever $f \in L^1$ is Riemann integrable, substitutions, integration by parts etc. can be applied to the Lebesgue integral, as long as the functions remain Riemann integrable and in L^1 all along.*

8 Some applications of the convergence theorems

Theorem 108. *Let $[a,b] \subset \mathbb{R}$, $f : X \times [a,b] \to \mathbb{C}$ be s.t. $\forall t \in [a,b], f(\cdot,t) \in L^1(X,\mu)$. Let $F = t \mapsto \int_X f(x,t)d\mu$.*

1. *Assume there is a $g \in L^1(X,\mu)$ s.t. $\sup_{t\in[a,b]} f(x,t) \leqslant g(x)$ and $\forall x : f(x,t)$ is continuous in t at $t = t_0$. Then F is continuous at t_0.*

2. *Assume f is continuous in t for $t \in [c,d] \subset [a,b]$, $\frac{\partial f}{\partial t}$ exists for $t \in (a,b)$ and $\sup_{t\in(c,d)} |\frac{\partial f}{\partial t}| \leqslant g \in L^1(X,\mu)$. Then F is differentiable on (c,d) and $F'(t) = \int_X \frac{\partial f(x,t)}{\partial t} d\mu(x)$.*

Proof. Both continuity and differentiability can be stated in terms of limits of sequences.

For 1., dominated convergence implies that $\lim_{n\to\infty} F(t_n) = F(t_0)$ for any sequence $(t_n)_{n\in\mathbb{N}}$ converging to t_0.

For 2., note first that, by the MVT, the function

$$h = (x,s,t) \mapsto \frac{f(x,s) - f(x,t)}{s-t}\chi_{s\neq t} + \frac{\partial f}{\partial t}\chi_{s=t}$$

is bounded in absolute value by g and differentiability of F at t_0 is equivalent to sequential continuity of $h(\cdot,s,t_0)$ at $s = t_0$. □

Exercise 33. *Let $f \in L^1(\mathbb{R})$. The Fourier transform of f is defined as*

$$\hat{f} = k \mapsto \int_{\mathbb{R}} f(x)e^{-2\pi ikx}dx$$

1. *Show that $\hat{f}$ is continuous.*

2. *(**The Riemann-Lebesgue Lemma**). Show that $\lim_{k\to\pm\infty} \hat{f}(k) = 0$. (Hint: Prove this when $f = \chi_{[a,b]}$ and use Theorem 103.)*

Exercise 34. *Let $f \in L^1(\mathbb{R},m)$ and let $F(x) = \int_{-\infty}^x f dm$. Note that $F(x) = \displaystyle\int_{\mathbb{R}} f\ \chi_{(-\infty,x)}dm$. Show that F is continuous.*

Definition 109 (Definition of the Gamma function)**.** *For z in the right half plane $\{z : \Re z > 0\}$ define the Gamma function by*

$$\Gamma(z) = \int_{\mathbb{R}^+} t^{z-1}e^{-t}dt$$

Integration by parts shows that $\Gamma(x+1) = x\Gamma(x)$ (the recurrence formula). The recurrence formula shows that Γ is analytic in $\mathbb{C}$, except for simple poles at $\mathbb{Z} \setminus \mathbb{N}$. Induction shows that $\Gamma(n) = (n-1)!, \forall n \in \mathbb{N}$. Show that the Euler-Poisson integral $\displaystyle\int_{-\infty}^{\infty} e^{-x^2}\,dx = \sqrt{\pi}$ implies that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Lemma 110 (Watson's Lemma). *Let $F \in L^1(\mathbb{R}^+)$, and assume*

$$\lim_{s\to 0^+} s^{-\beta}F(s) = 1$$

where $\Re(\beta) > -1$. Then

$$\lim_{x\to\infty} x^{\beta+1}\int_0^\infty F(s)e^{-sx}ds = \Gamma(\beta+1)$$

The same is true in the limit $\rho \to \infty$ if $x = \rho e^{i\phi}$ and $e^{i\phi}$ is in the right half plane.

Proof. It suffices to prove the result for $G = F\chi_{s\leqslant\epsilon}$ for any choice of $\epsilon > 0$ since, by dominated convergence, $\lim\limits_{x\to\infty}\int_{x_0}^\infty F(s)(e^{-xs}x^{\beta+1})ds = 0$. Choose ϵ s.t. $\sup_{0\leqslant s\leqslant\epsilon}|G(s)| < 2$. We have

$$\lim_{x\to\infty} x^{\beta+1}\int_0^\infty e^{-xs}G(s)ds = \lim_{x\to\infty}\int_0^\infty \frac{G(t/x)}{(t/x)^\beta}e^{-t}t^\beta dt = \Gamma(\beta+1) \tag{39}$$

where we used dominated convergence. Fill in the details and extend to the complex case. □

Note 111. Often, Watson's lemma is stated as follows: if $F(s) \sim s^\beta$ for small s, then, for large x, $\int_0^\infty e^{-xs}F(s)ds \sim \frac{\Gamma(\beta+1)}{x^{\beta+1}}$.

Exercise 35. 1. *Let $f(x) = \int_0^\infty (1+s)^{-1}e^{-xs}ds$. Use Watson's lemma and induction to show that, for any n,*

$$\lim_{x\to\infty}\frac{x^{n+2}(-1)^{n+1}}{(n+1)!}\left(f(x) - \sum_{j=0}^n j!\,\frac{(-1)^j}{x^{j+1}})\right) = 1$$

2. *Show that $z \mapsto f(1/z)$ in 1. extended by $f(0) = 0$ is infinitely differentiable at zero from the right, and it has the right-sided Taylor series*

$$\sum_{n=0}^\infty n!(-1)^n z^{n+1}$$

Exercise 36. *Let $g(s) = s - \ln s$. Check that $x^{-x-1}\Gamma(x+1) = \int_0^1 e^{-xg(s)}ds + \int_1^\infty e^{-xg(s)}ds$. Note that g is monotonic and differentiable on $(0,1)$ and $(1,\infty)$, and that*

$$\lim_{s\to 1}\frac{g(s)-1}{(s-1)^2} = 2$$

Change variable to $u = g(s)$ and apply Watson's lemma to prove Stirling's formula

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n(1+o(1)) \quad \text{as } n\to\infty$$

Exercise 37. *Define $\ell^2(\mathbb{N}) = \left\{f : \mathbb{N}\to\mathbb{C} \,|\, \|f\|_2 := \sum_{n\in\mathbb{N}}|f(n)|^2 < \infty\right\}$. Let $(f_k)_{k\in\mathbb{N}}$ be a Cauchy sequence in $\ell^2(\mathbb{N})$. Show that the limit $\lim_{n\to\infty} f_n(k) =: f(k)$ exists for all k. For each k, choose $N(k)$ so* *that* $\|f_{N(k+1)} - f_{N(k)}\|_2 \leqslant 2^{-k}$ *and use dominated convergence to show that*

$$f_{N(1)} + \sum_{k\in\mathbb{N}} \left(f_{N(k+1)} - f_{N(k)})\right)$$

converges in ℓ^2 *to conclude that* ℓ^2 *is a complete normed space.*

Note 112. The following observation may help in dealing with the operations needed in measure theory proofs. If $A_{\mathbf{n}}$ are sets given by $\{x \in X : P(n_1, n_2, ..n_k)(x))\}$ where P is some property ("predicate") with k parameters, say integer-valued, then

$$\bigcup_{n_1\in\mathbb{N}} A_{\mathbf{n}} = \{x \in X : (\exists n_1)(P(n_1, n_2, ..n_k)(x))\} \tag{40}$$

$$\bigcap_{n_2\in\mathbb{N}} \bigcup_{n_1\in\mathbb{N}} A_{\mathbf{n}} = (\forall n_2)(\exists n_1)P(n_1, n_2, ..n_k)(x)\} \tag{41}$$

and so on, a dictionary that you can refine yourselves. This dictionary also suggests why one needs the AC for proving existence of Borel or Lebesgue non-measurable sets in $\mathbb{R}$.

In view of (40),(41), we will sometimes use the shorthand

$$\mu(P(x)) := \mu\left(\{x : P(x)\}\right)$$

We also see that

$$\text{If } \; P \Rightarrow Q \text{ then } \mu(P) \leqslant \mu(Q)$$

$$\mu(\exists n : P_n) \leqslant \sum_n \mu(P_n)$$

$$\mu(P_1) < \infty \Rightarrow \mu(\forall n : P_n) \leqslant \inf_n \mu(P_n) \tag{42}$$

9 Topologies on spaces of measurable functions

Among the important types of convergence are pointwise convergence, convergence in L^p, $p \in [1, \infty]$ (defined so far for $p = 1, 2, \infty$, the latter being uniform convergence) and convergence in measure introduced next.

Definition 113. *A sequence of measurable functions* $(f_n)_{n\in\mathbb{N}}$ *converges in measure to* f *if*

$$\sup_{\epsilon>0} \lim_{n\to\infty} \mu(\{x : |f_n(x) - f(x)| \geqslant \epsilon\} = 0$$

A sequence $(f_n)_{n\in\mathbb{N}}$ *is Cauchy in measure if*

$$\sup_{\epsilon>0} \lim_{m,n\to\infty} \mu(\{x : |f_n(x) - f_m(x)| \geqslant \epsilon\} = 0$$

Exercise 38. *1. The topology of convergence in measure is metrizable. Check that*

$$\rho(f,g) = \inf_{\epsilon>0} \left[\epsilon + \mu(|f-g| > \epsilon)\right]$$

is one such metric.

2. Let $X = \mathbb{R}$. Is the topology of pointwise convergence metrizable?

Theorem 114 (Completeness)**.** *Assume $(f_n)_{n\in\mathbb{N}}$ is Cauchy in measure. Then $(f_n)_{n\in\mathbb{N}}$ converges in measure to a measurable f, and a subsequence $(f_{n_j})_{j\in\mathbb{N}}$ converges pointwise a.e. to f. The limit is unique modulo values on null sets.*

Proof. We first find a subsequence F_j which converges pointwise a.e to f. For each n let $j(n)$ be s.t. for all $j' > j(n)$ we have

$$\mu\left(|f_{j'}(x) - f_{j(n)}(x)| \geqslant 2^{-n}\right) \leqslant 2^{-n}$$

Let $F_n = f_{j(n)}$. It follows that, for all n,

$$\mu\left(|F_{n+1}(x) - F_n(x)| \geqslant 2^{-n}\right) \leqslant 2^{-n} \text{ and} \tag{43}$$

$$\mu\left((\exists m \geqslant n)|F_n(x) - F_m(x)| \geqslant 2^{-n}\right) \leqslant 2 \cdot 2^{-n} \tag{44}$$

Let N be the set where (F_n) *does not* converge. For $x \in N$,

$$(\exists k)(\forall n)(\exists m \geqslant n) : |F_n(x) - F_m(x)| \geqslant k^{-1} \Rightarrow (\forall n)(\exists m \geqslant n) : |F_n(x) - F_m(x)| \geqslant 2^{-n}$$

and thus, by (44) and (42), $\mu(N) = 0$. Thus $(F_j)_{j\in\mathbb{N}}$ converges pointwise a.e. to some f, implying in particular that f is measurable. Since $\mu(|F_j - f| \geqslant 2^{-j}) \leqslant \sum \mu(|F_j - F_{j+1}| \geqslant 2^{-j} = 2 \cdot 2^{-j}$, we have $F_j \to f$ in measure as well. Returning to the definition of the F_j, we have $f_n \to f$ in measure. □

Proposition 115. *L^1 convergence implies convergence in measure (and in particular the existence of a pointwise a.e. convergent subsequence).*

Proof. Assume $(f_n)_{n\in\mathbb{N}}$ are in L^1 and $\|f_n - f\|_1 \to 0$. Then for any $y > 0$ we have

$$\underbrace{\mu(|f - f_n| \geqslant y) \leqslant y^{-1} \int |f_n - f| d\mu}_{\text{this is called Markov's inequality}} \leqslant \|f_n - f\|_1 \tag{45}$$

□

Theorem 116 (Egoroff)**.** *Assume $\mu(X) < \infty$ and that the sequence of measurable functions $(f_n)_{n\in\mathbb{N}}$ converges pointwise a.e. to f. Then, for any $\epsilon > 0$ there is an A s.t. $\mu(X \setminus A_\epsilon) < \epsilon$ and $(f_n)_{n\in\mathbb{N}}$ converges uniformly on A_ϵ.*

Proof. Let $\epsilon > 0$. For any $k \in \mathbb{N}$ we have $\mu((\forall n)(\exists m \geqslant n)|f_m(x) - f(x)| \geqslant 1/k\} = 0$. For each $k \in \mathbb{N}$ choose A_k with $\mu(A_k) > \mu(X) - \frac{\epsilon}{2^k}$ and $\exists N(k)$ s.t. $\sup_{x\in A_k, m\geqslant N(k)} |f_m(x) - f(x)| \leqslant k^{-1}$. The sought-for set is $\cap_k A_k$. □

Corollary 117 (Lusin's theorem)**.** *Let $f : [a, b] \to \mathbb{C}$ be Lebesgue measurable. Then for any $\epsilon > 0$ there is a set $A_\epsilon \subset [a, b]$ of measure $> b - a - \epsilon$ s.t. $f|_{A_\epsilon}$ is continuous.*

Proof. This follows easily from Egoroff's theorem and Theorem 103. See also the problem set of Prof. Falkner, p. 48 for a direct proof. □

Exercise 39. *Prove Lusin's theorem by showing first that it holds for characteristic functions. If $A \subset [a, b]$ then there exist $K \subset A \subset \mathcal{O}$ s.t. $\mu(\mathcal{O} \setminus K) < \epsilon$ and χ_A is continuous on $K \cup \mathcal{O}^c$.*

If $\mu(X) < \infty$, then uniform convergence of L^1 functions implies L^1 convergence, which implies convergence in measure, which in turn implies pointwise convergence a.e. of a subsequence. In general, these implications cannot be reversed. When $\mu(X) = \infty$, aside from the results above, there is basically a sea of counterexamples.

Exercise 40. *Consider the sequence $f_n = \chi_{J_n}$ where J_n is the interval where $|x - (\log_2 n \mod 1)| \leqslant n^{-1}$. Show that $\|f_n\|_1 \to 0$ but f_n is pointwise everywhere divergent.*

HW 10/22 : 20,21,26,28,34,42 in Folland; turn in: Ex 33,35 in the notes. We end this section with a useful general result about constructing σ-algebras.

Definition 118. *Let X be a set. A monotone class $\mathcal{S} \subset \mathcal{P}(X)$ is a collection of sets with the following properties:*

1. *If $A_n \subset X$ and $A_n \subset A_{n+1}$ $\forall n \in \mathbb{N}$, then $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{S}$.*
2. *If $A_n \subset X$ and $A_n \supset A_{n+1}$ $\forall n \in \mathbb{N}$, then $\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{S}$.*

Note 119. Clearly any σ-algebra is a monotone class, and the intersection of a family of monotone classes is a monotone class. Thus, given a collection of sets $\mathcal{G}$, there is always a smallest monotone class containing $\mathcal{G}$, called the monotone class *generated* by $\mathcal{G}$.

Theorem 120. *Let X be a set and $\mathcal{A}$ an algebra in X. The monotone class generated by $\mathcal{A}$ coincides with the σ-algebra generated by $\mathcal{A}$.*

Proof. Let $\mathcal{S}$ be the monotone class generated by $\mathcal{A}$. We first note that is suffices to show that $\mathcal{S}$ is closed under finite unions and complements. Indeed, it then follows that $\mathcal{S}$ is closed under countable unions (since $\bigcup_{j\leqslant n} A_j$ is increasing).

1. (Closure under finite unions) We fix an $A \in \mathcal{A}$ stay $\mathcal{S}$: let $\mathcal{C}(A) = \{B \in \mathcal{S} : B \cup A \in \mathcal{S}\}$. Clearly, $\mathcal{A} \subset \mathcal{C}(A)$. If $(B_j)_j$ is an increasing sequence in $\mathcal{C}(A)$, then $A \cup \bigcup_1^n B_j = A \cup B_n$, and

$$A \cup \bigcup_{n\in\mathbb{N}} B_n = \bigcup_{n\in\mathbb{N}} A \cup B_n \in \mathcal{S}$$

since $\mathcal{S}$ is a monotone class, and thus $\mathcal{C}(A)$ is closed under countable monotone unions. A very similar argument shows that $\mathcal{C}(A)$ is closed under countable monotone intersections.

Therefore $\mathcal{C}(A)$ is a monotone class containing $\mathcal{A}$, hence $\mathcal{C}(A) = \mathcal{S}$. Repeating this argument, but now with $A \in \mathcal{S}$, closure under finite unions follows.

2. (Closure under complements). The proof is similar: let $\mathcal{C} = \{A \in \mathcal{S} : A^c \in \mathcal{S}$. Clearly, $\mathcal{A} \subset \mathcal{C}$. Now, the complement of a monotone union is a monotone intersection and vice-versa, and thus $\mathcal{C} = \mathcal{S}$. □

10 Product measures and integration on product spaces

Let $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \lambda)$ be measure spaces. We will define the product measure and integral on $X \times Y$ via iterated integrals

$$\int_{X\times Y} f(x,y)d(\mu\times\lambda) := \int_X d\mu \int_Y f(x,y)d\lambda = \int_Y d\lambda \int_X f(x,y)d\mu;\ (\mu\times\lambda)(A) = \int_{X\times Y} \chi_A d(\mu\times\lambda)$$

whose consistency needs some work.

Definition 121. Rectangles *are sets of the form* $A \times B$*,* $A \in \mathcal{S}, B \in \mathcal{T}$*. The family* $\mathcal{E}$ *of elementary sets is the set of all finite disjoint unions of rectangles.*

Let $\mathcal{M}$ be the σ-algebra generated by $\mathcal{E}$.

Proposition 122. *1.* $\mathcal{M} = \mathcal{S} \times \mathcal{T}$*.*

2. $\mathcal{E}$ *is an algebra.*

3. $\mathcal{M}$ *is the monotone class generated by* $\mathcal{E}$*.*

Figure 3: $A_2 \times B_2$ is the orange rectangle and the multicolored one is $A_1 \times B_1$.

Proof. 1. is simply Proposition 37.

2. Clearly $X \times Y$ and $\emptyset$ are in $\mathcal{E}$. Check that

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2)$$

(the magenta region in Fig. 3) and

$$(A_1 \times B_1) \setminus (A_2 \times B_2) = [(A_1 \cap A_2) \times (B_1 \setminus B_2)] \cup [(A_1 \setminus A_2) \times B_1]$$

(the dark yellow rectangle union the green one). It is now straightforward to show that $\mathcal{E}$ is closed under intersections and set differences.

3. This now follows from Theorem 120. □

Definition 123 (Sections). *We will denote $E_x = \{y : (x,y) \in E\}$ and $E^y = \{x : (x,y) \in E\}$. If f is $\mathcal{M}$ measurable, then we write $f_x = y \mapsto f(x,y)$ and $f^y = x \mapsto f(x,y)$.*

Theorem 124 (Sections are measurable). *1. If $E \in \mathcal{M}$, then $E_x \in \mathcal{T}$ for any $x \in X$, and $E^y \in \mathcal{S}$ for any $y \in Y$.*

2. If f is $\mathcal{M}$ measurable, then f_x is $\mathcal{T}-$measurable and f^y is $\mathcal{S}-$measurable.

Proof. 1. As usual, we let $\mathcal{M}'$ be the family of sets in $X \times Y$ s.t. Rectangles are in $\mathcal{M}'$ since $(A \times B)_x = B$ if $x \in A$ and $\emptyset$ otherwise. Using the fact that $\mathcal{T}$is a σ-algebra we see that

a. $X \times Y \in \mathcal{M}'$;

b. $(E^c)_x = (E_x)^c$, entailing that $\mathcal{M}'$ is closed under complements;

c. $(\cup E_j)_x = \cup(E_j)_x$, hence $\mathcal{M}'$ is closed under countable unions.

Thus $\mathcal{M}'$ is a σ-algebra containing $\mathcal{E}$.

2. This is clearly the case for characteristic functions of sets in $\mathcal{M}$. Since $(f + ag)_x = f_x + ag_x$, all simple functions have this property, and the result follow from Theorem 80. □

Theorem 125. *Let $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \lambda)$ be $\sigma-$ finite measure spaces. Let $Q \in \mathcal{S} \times \mathcal{T}$, and define*

$$\phi = x \mapsto \lambda(Q_x);\ \psi = y \mapsto \mu(Q^y) \tag{46}$$

Then, ϕ is $\mathcal{S}$-measurable, ψ is $\mathcal{T}$-measurable, and

$$\int_X \phi d\mu = \int_Y \psi d\lambda \tag{47}$$

Proof. We see from Theorem 124 that the definitions (46) and (47) make sense. Note also that $\lambda(Q_x) = \int_Y \chi_Q(x,y) d\lambda(y)$, and thus we can write (47) as

$$\int_X d\mu \int_Y \chi_Q d\lambda = \int_Y d\lambda \int_X \chi_Q d\mu \tag{48}$$

Let $(X_n)_{n\in\mathbb{N}}, (Y_m)_{n\in\mathbb{N}}$ be disjoint, of finite measure, and s.t. $X = \cup_{n\in\mathbb{N}} X_n$ and $Y = \cup_{m\in\mathbb{N}} Y_m$. Let $\mathcal{M}'$ be the family of all sets in $\mathcal{M}$ for which the statement in the theorem holds. We list some of the properties of $\mathcal{M}'$ that we will subsequently verify:

a. $\mathcal{M}'$ contains all measurable rectangles;

b. $\mathcal{M}'$ is closed under countable monotone unions, $\cup E_i$, $E_i \subset E_{i+1}$;

c. $\mathcal{M}'$ is closed under countable disjoint unions;

d. $\mathcal{M}'$ is closed under countable monotone intersections. Since any $E \in \mathcal{S} \times \mathcal{T}$ equals $\cup_{m,n}[E \cap (X_n \times Y_m)]$ it is enough to check this when $(E_i)_{i\in\mathbb{N}}$ is a decreasing family of sets in $\mathcal{S} \times \mathcal{T}$ s.t. $E_1 \subset A \times B$ where $\mu(A) + \mu(B) < \infty$.

For a. note that, if $E = A \times B$, then $\lambda(Q_x) = \lambda(B)\chi_A(x)$ and $\mu(Q^y) = \mu(A)\chi_B(y)$.

For b. let $\phi_i = \lambda((E_i)_x), \phi = \lambda(E_x)$, $\psi_i = \mu((E_i)^y), \phi = \mu(E^y)$. Continuity from below of λ and μ implies that $\phi_i \nearrow \phi$, $\psi_i \nearrow \psi$, and (47) follows from monotone convergence.

c.: For finite unions this is clear, since the characteristic function of a disjoint union is the sum of the characteristic functions of the individual sets. For countable ones, this now follows from b.

d. Same as b., using continuity from above and dominated convergence.

Let $\mathcal{M}''$ be the class of all $Q \in \mathcal{S} \times \mathcal{T}$ s.t., for all m and n, $Q \cap (X_n \times Y_m) \in \mathcal{M}'$. (b.&d.) show that $\mathcal{M}''$ is a monotone class containing $\mathcal{E}$, and thus $\mathcal{M}'' = \mathcal{S} \times \mathcal{T}$. Therefore $Q \cap (X_n \times Y_m) \in \mathcal{M}'$ for all m, n, and since these sets are disjoint, c. implies that their union is in $\mathcal{M}'$ completing the proof.

□

Definition 126. *Let $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \lambda)$ be $\sigma-$ finite measure spaces. For $Q \in \mathcal{S} \times \mathcal{T}$ define*

$$(\mu \times \lambda)(Q) = \int_X \lambda(Q_x) d\mu(x) = \int_Y \mu(Q^y) d\lambda(y)$$

Proposition 102 shows that $\mu \times \lambda$ is $\sigma-$ additive on $\mathcal{S} \times \mathcal{T}$. Check that $\mu \times \lambda$ is $\sigma-$finite.

Theorem 127 (Fubini). *Let $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \lambda)$ be $\sigma-$ finite measure spaces and f measurable on $X \times Y$.*

1. *If $ran(f) \subset [0, \infty]$ and*

$$(\nu_x f)(x) = \int_Y f_x d\lambda; \ \ (\nu^y f)(y) = \int_X f^y d\mu$$

then $\nu_x f$ is $\mathcal{S}-$measurable, $\nu^y f$ is $\mathcal{T}-$measurable, and

$$\int_{X \times Y} f d(\mu \times \lambda) = \int_X \nu_x f d\mu = \int_Y \nu^y f d\lambda$$

or, spelled out,

$$\int_{X \times Y} f d(\mu \times \lambda) = \int_X d\mu(x) \int_Y f(x, y) d\lambda(y) = \int_Y d\lambda(y) \int_X f(x, y) d\mu(x) \tag{49}$$

2. *If $ran(f) \subset \mathbb{C}$ and if*

$$\int_X \nu_x |f| d\mu = \int_X d\mu \int_Y |f|_x d\lambda < \infty$$

then $f \in L^1(\mu \times \lambda)$.

3. *If $f \in L^1(\mu \times \lambda)$, then $f^y \in L^1(\mu)$ a.e., and $f_x \in L^1(\lambda)$ a.e. Furthermore, $\nu_x f$ and $\nu^y f$ are in L^1 and (49) holds.*

Proof. (a) If $Q \in \mathcal{S} \times \mathcal{T}$ and $f = \chi_Q$, then this follows from Theorem 124. Hence, the property holds for all simple functions. Consequently, if $0 \leqslant s_1 \leqslant s_2 \leqslant \cdots$ is a sequence of simple functions s.t. $s_n \nearrow f$ pointwise in $X \times Y$, then, for all n,

$$\int_X \nu_x s_n d\mu = \int_{X \times Y} s_n d(\mu \times \lambda)$$

Now, as $n \to \infty$, monotone convergence implies $\nu_x s_n \nearrow f_x$ and $\int_{X \times Y} s_n d(\mu \times \lambda) \nearrow \int_{X \times Y} f d(\mu \times \lambda)$.

(b) This is simply (a) applied to $|f|$.

(c) Clearly it is enough to show this when $\mathrm{ran}(f) = \mathbb{R}$, in which case we write $f = f^+ - f^-$ and we note that (a) separately applies to f^+ and to f^-. Since f^+ and f^- are bounded by $|f|$, $\nu_x f^+ \in L^1(\mu)$ and $\nu_x f^- \in L^1(\lambda)$. Thus, except for a null set, both $\nu_x f^+$ and $\nu_x f^-$ are finite and on this set $\nu_x f = \nu_x f^+ - \nu_x f^-$ and the result follows. □

In Real and Complex Analysis, pp. 166-167, Rudin shows that the various hypotheses in Theorems 125, 127 cannot be omitted.

Note 128. *Even if μ, λ are complete, $\mu \times \lambda$ need not be. Indeed, any straight line is a null set w.r.t. the two-dimensional Lebesgue measure. The set $\{0\} \times V \subset \mathbb{R}^2$, where V is a nonmeasurable Vitaly set, is contained in a null Borel set, $\{0\} \times \mathbb{R}$, but it is not measurable (why?).*

The following extension of Theorem 127 to the completion of the measures $\mu, \lambda, \mu \times \lambda$ is left as an exercise:

Theorem 129. *Let $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \lambda)$ be complete $\sigma-$ finite measure spaces. Let $(\mu \times \lambda)^*$ be the completion of the product measure, and $(\mathcal{S} \times \mathcal{T})^*$ be the associated σ-algebra on $X \times Y$. Then Theorem 127 applies with one difference: the measurablility of f_x, f^y is guaranteed only a.e., and thus $\nu_x f$, $\nu^y f$ are only defined a.e.*

Exercise 41. *1. Use the relation $\frac{1}{x^2} = \int_0^\infty t e^{-xt} dt$ and Fubini to show that*

$$\int_0^\infty \frac{\sin^2 x}{x^2} dx = \frac{\pi}{2}$$

2. Show that

$$\int_0^\infty \frac{\sin x}{x} dx := \lim_{N \to \infty} \int_0^N \frac{\sin x}{x} dx = \frac{\pi}{2}$$

(Note that $x^{-1} \sin x$ in not in $L^1(\mathbb{R}^+)$, and thus the first integral above is improper, and it is defined as a limit.)

Exercise 42. *Use any of the theorems developed so far to solve the following problems.*

1. *Assume $f(x) = \sum_{k \geqslant 0} a_k x^k$ and $g(x) = \sum_{k \geqslant 0} b_k x^k$ converge for all x in the open unit disk. Then $f(x) g(x) = \sum_{k=0}^\infty x^k \sum_{j=0}^k a_j b_{k-j}$ where the series converges in the open unit disk.*
2. *Assume $\sum_{k \geqslant 0} |a_k| < \infty$. Then $\sum_{k \geqslant 0} a_k$ is convergent, and all rearrangements of the series are convergent to the same value. That is, if $f : \mathbb{N} \to \mathbb{N}$ is any bijection, then $\sum_{k \geqslant 0} a_k = \sum_{k \geqslant 0} a_{f(k)}$.*
3. *Assume $F, G \in L^1([0, \infty)$ and that, for $\Re x > 0$, $f(x) = \int_{\mathbb{R}^+} F(p) e^{-px} dp$ and $g(x) = \int_{\mathbb{R}^+} G(p) e^{-px} dp$. Then*
$$f(x) g(x) = \int_{\mathbb{R}^+} \left(\int_{[0,p]} F(s) G(p-s) ds \right) e^{-xp} dp$$
4. *Justify Archimedes' method of approximating π by showing that the area of the unit disk $\mathbb{D}$ is the limit as $n \to \infty$ of the areas of regular polygons with n sides inscribed in $\mathbb{D}$. How many sides do you need to guarantee that the value you get is within at most 10^{-10} away from π?*

11 The $n-$dimensional Lebesgue integral

The Lebesgue measure m^n on $\mathbb{R}^n$ is the completion of the product measure on $(\mathbb{R}^n, \otimes_1^n \mathcal{B}_{\mathbb{R}}, \times_1^n m)$ where m is the Lebesgue measure on $\mathbb{R}$. The completion of the σ-algebra $\otimes_1^n \mathcal{B}_{\mathbb{R}}$ is denoted by $\mathcal{L}^n$ (remember, this completion is not $\otimes_1^n \mathcal{L}$!) Common notations for the integral with respect to this measure are

$$\int_{\mathbb{R}^n} f dm; \ \int_{\mathbb{R}^n} f(x) m(dx); \ \int_{\mathbb{R}^n} f(x) d^n x; \ \int f(\mathbf{x}) d\mathbf{x}$$

while the measure m^n is often written simply m.

11.1 Extensions of results from 1d

Theorem 130. *If $Q \in \mathcal{L}^n$, then*

1. $m(Q) = \inf_{\mathcal{O} \supset Q} \mu(\mathcal{O}) = \sup_{K \subset Q} \mu(K)$.

2. *There exist an F_σ set F and a G_δ set G s.t. $F \subset Q \subset G$ and $\mu(G \setminus F) = 0$.*

3. *If $m(Q) < \infty$ then, for any $\epsilon > 0$, $m(Q \Delta \cup_{j=1}^n R_j) < \epsilon$ for some disjoint rectangles R_j whose sides are intervals.*

Proof. By theorem 50 m^n is the extension of m^n, as restricted to the algebra $\mathcal{E}$ of elementary sets. In particular, for any $\epsilon > 0$ there is a disjoint family of rectangles R_k containing Q s.t. $\mu(\cup_1^\infty R_k \cap Q^c) < \epsilon$. With this, the proof of Theorem 59 translates with little change to a proof of 1; the proof of 2. is the same, up to notations to that of Theorem 60. Finally, for 3, by the usual 2^{-n} argument, it is enough to prove the result for a rectangle, and thus for a side of a rectangle. The latter follows in the usual way: If $A \subset \mathbb{R}$ has finite measure, then there is an $\mathcal{O} \supset A$ s.t. $\mu(\mathcal{O}) < \mu(A) + \epsilon/2$. Now $\mathcal{O} = \cup_{j=1}^\infty I_j$ for some open intervals I_j, and thus there is an N s.t. $\mu(\mathcal{O} \setminus \cup_{j=1}^N I_J) < \epsilon/2$. □

Theorem 131. *Continuous functions are dense in $L^1(\mathbb{R}^n)$; so are simple functions, $\sum_1^N a_n \chi_{R_n}$, where R_n are products of intervals.*

Proof. The second statement follows easily from the previous theorem. If $R_n = \prod_1^n \chi_{I_j}$ for some intervals $I_j \subset \mathbb{R}$, then $\chi_{R_n} = \prod_1^n \chi_{I_j}$, which can be approximated by a product of continuous functions of one variable. Then the result holds for $\sum_1^N a_n \chi_{R_n}$, where R_n are products of intervals, and density takes care of the rest. □

Theorem 132. *Let $R \in \mathbb{R}^n$ be a product of closed intervals and f bounded on R.*

1. *If f is Riemann integrable on R, then f is Lebesgue measurable and the Riemann integral of f on R equals $\int_R f dm$.*

2. *f is Riemann integrable on R iff the set of discontinuities of f has measure zero.*

Proof. Again, basically a copy of the 1-d proof. □

The theory of Jordan content in $\mathbb{R}^n$ is very similar to that in $\mathbb{R}$.

Theorem 133 (Behavior of set-measure w.r.t. linear-affine transformations)**.**

1. $\int \chi_A(x+a)dx = \int \chi_A(x)dx$.
2. *If* $c \neq 0$, *then* $\int \chi_A[(c^{-1}x_1, ..., x_n)]dx = |c| \int \chi_A(x)dx$
3. $\int \chi_A[(x_1, .., x_k, x_{k+1}, ..., x_n)]dx = \int \chi_A[(x_1, .., x_{k+1}, x_k, ..., x_n)]dx$.
4. $\int \chi_A[(x_1, x_2, ..., x_n)]dx = \int \chi_A[(x_1 + x_2, x_2, ..., x_n)]dx$.

Proof. For 1–3, it suffices to show the result for products $\prod_{j=1}^n \chi(I_j)$, where I_j are intervals. But, by Fubini, the integral is the product of one-dimensional integrals and the proof is immediate.

4. By the above, it suffices to show this in $\mathbb{R}^2$. We have, by Fubini,

$$\int \chi(x_1 + x_2, x_2)dx = \int dx_2 \int \chi_1(x_1 + x_2)dx_1 = \int dx_2 \int \chi_1(x_1)dx_1$$

by 1. □

Exercise 43. *Assume* R *is a product of intervals,* Ω *is some open set in* $\mathbb{R}^m$ *and* $g : R \times \Omega \to \mathbb{C}$ *is continuous. Then*

1. $$y \mapsto \int \chi_R(x + g(x,y))dm(x)$$
   is continuous in Ω.
2. *Let* $T \in GL(\mathbb{R}^n)$ *and assume* g *is continuous on* $TR \times \Omega$. *Then*
   $$y \mapsto \int \chi_{TR}(x + g(x,y))dm(x)$$
   is continuous in $y \in \Omega$.

Note: since χ_R *is Borel measurable and* $x \mapsto x + g(x,y)$ *is continuous, the composition is measurable.*

Theorem 134. *If* $M \in GL(\mathbb{R}^n)$ *and* $f \in L^1$ *or* $f \geqslant 0$ *is* $\mathcal{L}^n$*−measurable, then*

$$\int f(x)dx = |\det T| \int f(Tx)dx$$

Corollary 135. *If* A *is measurable, then* $m(TA) = |\det T| m(A)$.

Note 136. *Note that the corollary implies that* $m(T^{-1}(N)) = 0$ *for every null set in* $\mathcal{B}_{\mathbb{R}^n}$. *Then, if* B *is a Borel set, then* $f^{-1}(B) = B_1 \cup N_1$ *where* N_1 *is a null set in* $\mathcal{L}^n$. *We have* $T^{-1}(B_1 \cup N_1) = T^{-1}(B_1) \cup T^{-1}(N_1)$, *and if* N *is a null Borel set containing* N_1, *then* $T^{-1}(N_1) \subset T^{-1}(N)$ *is of measure zero, and thus measurability of* $f \circ G$ *follows.*

Proof. Writing an open set as a countable union of boxes, we see that $m(H(\mathcal{O})) \leqslant \alpha m(\mathcal{O})$ and the result follows. By density, it is enough to show the equality above for linear combinations of χ_R where R are products of intervals, thus for just one such χ_R. Recalling that GL is generated by the simple transformations 2-4 in Theorem 133, the rest is a corollary of that theorem. □

Theorem 137 (Change of variables)**.** *Let Ω be an open set in $\mathbb{R}^n$ and $G : \Omega \to G(\Omega)$ be an $\mathbb{R}^n$ diffeomorphism. If f is Lebesgue measurable on $G(\Omega)$ then $f \circ G$ is measurable on Ω. If $f \geqslant 0$ or $f \in L^1(G(\Omega))$ then*

$$\int_{G(\Omega)} f\,dm = \int_{\Omega} (f \circ G)\,|\det D_x G|\;dm$$

Corollary 138. *If $Q \in \Omega$ is $\mathcal{L}^n-$measurable, then $G(Q)$ is measurable and*

$$m(G(Q)) = \int_Q |\det D_x G| dm$$

Proof. Measurability follows from the Corollary, as in Note 136. By density and theorem 130 it suffices to prove this when f is continuous and $Q = R$, a product of closed intervals. Let $M_x = D_x G$ and $J_x = |\det M_x|$. We first prove the following.

Note 139. *Let Ω be an open set in $\mathbb{R}^n$ and $G : \Omega \to G(\Omega)$ be an $\mathbb{R}^n$ diffeomorphism. Let $K \in \Omega$ be compact and $2d \leqslant$dist$(K, \partial\Omega)$. From the Taylor series with remainder theorem we see that the function*

$$\phi := (x, y, \epsilon) \mapsto \begin{cases} \epsilon^{-1}\left(G(x) - G(x + \epsilon y) + \epsilon M_x y\right);\ \epsilon \neq 0 \\ 0; \qquad \epsilon = 0 \end{cases}$$

is uniformly continuous in the compact set $K_1 = \{(x, y, \epsilon) : 0 \leqslant \epsilon \leqslant d, x \in K, x + \epsilon y \in K\}$. Indeed, continuity follows from the fact that $G \in C^1$ and uniform continuity follows from the fact that K is compact.

Lemma 140. *For $0 < \epsilon < d$ and $x_0 \in R$ let $R_0 = x_0 + \epsilon R$. We have*

$$\lim_{\epsilon \to 0} \epsilon^{-n} \left(\int_{R_0} f(G(x)) J_x dx - \int_{G(R_0)} f dm \right) = 0$$

uniformly in x_0.

Proof. Let $x = x_0 + \epsilon y$ and $z_0 = G(x_0)$. Then $x \in R_0 \Leftrightarrow y \in R$ and

$$\lim_{\epsilon \to 0} \epsilon^{-n} \int_{R_0} f(G(x)) J_x dx \;=\; \lim_{\epsilon \to 0} \int_R f(z_0 + \epsilon M_{x_0} y + \epsilon \phi(x_0, y, \epsilon)) J_{x_0 + \epsilon y} dy \;=\; m(R) f(z_0) J_{x_0} \quad (50)$$

uniformly in x_0.

Next, define ψ for G^{-1} as in Note 139. Note that $z_0 + \epsilon u \in G(R_0)$ means $x_0 + \epsilon M_{x_0}^{-1} u + \epsilon \psi(z_0, u, \epsilon) \in R_0$ or $u + M_{x_0} \psi(z_0, u, \epsilon) \in M_{x_0} R$ which means

$$\lim_{\epsilon \to 0} \epsilon^{-n} \int_{G(R_0)} f dm = \lim_{\epsilon \to 0} \int f(z_0 + \epsilon u + \epsilon M_{x_0} \psi(z_0, u, \epsilon)) \chi_{M_{x_0} R}(u + M_{x_0} \psi(z_0, u, \epsilon)) du = J_{x_0} m(R) f(z_0) \quad (51)$$

uniformly in the parameters, by Exercise 43, implying the result.

□

To end the proof of the theorem, take $\epsilon = 1/N$ and a partition of R in N^n boxes, $B_k =$

$x_k + N^{-1}R$ and check that

$$\int_R f(G(x))J_x dm = \lim_{N\to\infty}\sum_{k=1}^{N^n} f(G(x_k))J_{x_k} m(R_k) = \lim_{N\to\infty}\sum_{k=1}^{N^n}\int_{G(R_k)} f dm = \int_{G(R)} f(u)du$$

□

12 Polar coordinates

This is an important set of coordinates adapted to $SO(n)$ symmetry. Let S^{n-1} be the unit sphere in n dimensions, and, for $x \in \mathbb{R}^n \setminus \{0\}$ let

$$\Phi(x) = \left(|x|, \frac{x}{|x|}\right) := (r, x')$$

which is a diffeomorphism between $\mathbb{R}^n \setminus \{0\}$ and $\mathbb{R}^+ \times S^{n-1}$. On $\mathbb{R}^+ \times S^{n-1}$, the natural measure is m_*, the push-forward of Φ.

Next, we are are looking at a simple example of the inverse problem of constructing a product measure, the *disintegration of a measure*: we want to write m_* as a product measure. It is easy to see what the first component of the product should be. Taking as a measurable set a ball of radius R centered at zero, we see that the measure induced by Φ on $\mathbb{R}^+$ is

$$m\left(\Phi^{-1}(\{(r,s) : r \leqslant R, s \in S^{n-1})\right) = C_n R^n$$

where C_n is a constant (unimportant at this stage, which will be determined shortly). This implies that the measure on $\mathbb{R}^+$ should be (up to an irrelevant constant) $C_n n r^{n-1} dr$. Absorbing the constant in the measure of the sphere, we simply take $d\rho = r^{n-1}dr$.

Theorem 141. *There is a unique measure σ on S^{n-1} s.t. $m_* = \rho \times \sigma$. Furthermore, if $f \in \Lambda^1(\mathbb{R}^n)$, then*

$$\int f dm = \int_{\mathbb{R}^+}\int_{S^{n-1}} f(rx')d\sigma(x')r^{n-1}dr$$

Proof. We know that the last equation holds as soon as we find a σ s.t. $m_* = \rho \times \sigma$. To see what σ should be we now concentrate on the x' component. Let $A \in \mathcal{B}_{S^{n-1}}$ and define

$$A_r = \{r'x' : r' \leqslant r, x' \in A\} = \Phi^{-1}((0, r'] \times E)$$

We need to have

$$m(A_1) = \int_0^1 \int_E d\sigma(x')r^{n-1}dr = n^{-1}\sigma(E)$$

which implies that we should have $\sigma(E) = n\, m(A_1)$ which we take as a definition. By the behavior of the Lebesgue measure under dilations, we have $m(A_r) = r^n m(A_1)$. Take now a rectangle $R = J \times B$, $J = (r_1, r_2]$ an interval in $\mathbb{R}^+$ and B measurable in S^{n-1}. Then $R = A_{r_2} \setminus A_{r_1}$ implying

$$\mu_*(R) = m(A_{r_2}) - m(A_{r_1}) = \rho(J)\sigma(A)$$

From this point on, it is standard to construct from this a measure on the σ-algebra on $\mathcal{B}_{\mathbb{R}^n}$. It agrees with m on rectangles, which completes the proof (try to complete it yourself, then look in Folland). □

The following is a neat trick to $\sigma(S^{n-1})$, by calculating an integral in two ways.

Proposition 142. *1. For a>0*

$$\int_{\mathbb{R}^n} \exp\left(-a\sum_{k=1}^n x_i^2\right) dm = \left(\frac{\pi}{a}\right)^{n/2}$$

2.

$$\sigma(S^{n-1}) = \frac{2\pi^{n/2}}{\Gamma(n/2)}$$

Proof. 1. By Fubini,

$$\int_{\mathbb{R}^n} \exp\left(-a\sum_{k=1}^n x_i^2\right) dm = \left(\int_{\mathbb{R}} e^{-ax^2}dx\right)^n \tag{52}$$

and thus, using polar coordinates in $\mathbb{R}^2$ we get

$$\int_{\mathbb{R}} e^{-ax^2}dx = \left(2\pi\int_0^\infty e^{-ar^2}rdr\right)^{1/2} = \frac{\pi}{a}$$

which, using (52) implies the result.

2. Now we write the left side of (52) in polar coordinates in $\mathbb{R}^n$. Let $\mathcal{S} = \sigma(S^{n-1})$.

$$\pi^{n/2} = \int_{\mathbb{R}^n} \exp\left(-\sum_{k=1}^n x_i^2\right) dm = \mathcal{S}\int_0^\infty e^{-r^2}r^{n-1}dr = \tfrac{1}{2}\mathcal{S}\int_0^\infty e^{-u}u^{n/2}du = \tfrac{1}{2}\Gamma(\tfrac{n}{2})\mathcal{S} \tag{53}$$

and the result follows. □

Exercise 44. *Show that, for $n \in \mathbb{N}$,*

$$\int_{\mathbb{R}} x^n e^{-\beta x^2}dx = \frac{1}{2}\left((-1)^{2n}+1\right)\beta^{-n-\frac{1}{2}}\Gamma\left(n+\frac{1}{2}\right)$$

13 Signed measures

Definition 143. *A signed measure on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to [-\infty, \infty]$ s.t.*

1. *$\nu(\emptyset) = 0$.*
2. *at least one of the values $+\infty$, $-\infty$ is not in $\operatorname{ran}(\nu)$.*
3. *If $(A_j)_{j\in\mathbb{N}}$ are disjoint and measurable, then $\nu(\bigcup_{j\in\mathbb{N}} A_j) = \sum_{j=1}^\infty \nu(A_j)$ and $\nu(A_j) < \infty$ for all j or else $\nu(A_j) > -\infty$ for all j.*

Note 144. *The second condition is needed since if we had two sets $A_{\pm}$ s.t. $\nu(A_{\pm}) = \pm\infty$, then additivity would imply the nonsensical statement $\nu(A_+ \cup A_-) = \nu(A_+) + \nu(A_+) - \nu(A_- \cap A_-)$.*

Proposition 145. *In the setting of Definition 166, if A_j are measurable and $|\nu(\cup_{j\in\mathbb{N}} A_j)| < \infty$, then the series 2) converges absolutely.*

Proof. The definition implies that all rearrangements of the series converge, hence the series converges absolutely. □

Definition 146. *f is called extended integrable if $f^+ \in L^1$ or $f^- \in L^1$.*

Exercise 45.

1. *Show that, if μ is a measure and f is extended integrable, then $\nu := A \mapsto \int_A f d\mu$ is a signed measure.*
2. *Let ν be a signed measure. Show that:*
   (a) *if $(A_j)_{j\in\mathbb{N}}$ are increasing sets, then $\nu(\cup_j A_j) = \lim_{j\to\infty} \nu(A_j)$;*
   (b) *if $(A_j)_{j\in\mathbb{N}}$ are decreasing sets and $|\nu(A_1)| < \infty$, then $\nu(\cap_j A_j) = \lim_{j\to\infty} \nu(A_j)$.*

Definition 147. *If ν is a signed measure and A is a measurable set s.t.* **all** *of its measurable subsets have nonnegative measure, then A is called a positive set for ν. A negative set for ν is a positive set for $-\nu$ and a null set for ν is a set which is both positive and negative.*

Note 148. *Any subset of a positive set is a positive set.*

Proposition 149. *If $(A_j)_{j\in\mathbb{N}}$ are positive sets, then so is their union.*

Proof. Define as usual the disjoint sets $B_j = A_j \setminus \cup_{k\leqslant j} A_k$, whose union equals $\cup_j A_j$. Since $B_j \subset A_j$ for all j, the B_js are also positive sets. Now, if $E \subset \cup A_j$, then

$$\nu(E) = \sum_{j\in\mathbb{N}} \nu(E \cap B_j) \geqslant 0$$

□

13.1 Two decomposition theorems

Theorem 150 (The Hahn decomposition theorem). *Let ν be a signed measure on $(X, \mathcal{M})$. Then there exists a* disjoint measurable decomposition $X = X_+ \cup X_-$, *unique up to null sets, and s.t. $\pm\nu$ is positive on $X_\pm$.*

Proof (R. Doss, PAMS 80,2,(1980)). Assume w.l.o.g. that $+\infty$ is the excluded value of ν.

Lemma 151 (Quasi-positive sets)**.** *Let A be a set of finite measure. Then, for any $\epsilon > 0$ there is an $A_\epsilon \subset A$. s.t. all its subsets have measure $\geqslant -\epsilon$.*

Proof. By contradiction. Let $B_1 \subset A, \nu(B_1) < -\epsilon$. Since $\nu(A) = \nu(B_1) + \nu(A')$, $A' = A \setminus B$ we have $\nu(A') > \nu(A)$ and it therefore A' contains a set B_2 (clearly disjoint from B_1) s.t. $\nu(B_2) < -\epsilon$. Inductively, we construct a set of B_k contained in $A \setminus \cup_{j<k} B_j$ with $\nu(B_k) < -\epsilon$. But then $B = \cup B_k$ has measure $-\infty$ and $\nu(A) = \nu(B) + \nu(A \setminus B) = -\infty$, contradiction. □

Lemma 152. *If A is of finite measure, then A contains a positive set $P, \nu(P) \geqslant \nu(A)$.*

Take $\epsilon = 1/n$ and $P = \cap A_{1/n}$, a decreasing intersection of sets of finite measure $\geqslant \nu(A)$. Check that if $B \subset P$, then $\nu(B) \geqslant 0$.

To complete the proof of the theorem, we find a set of maximal measure and its corresponding P will be the positive set of X. Let

$$M = \sup_{P \in \mathcal{M}, \nu(P) \geqslant 0} \nu(P)$$

If P_n are s.t. $\nu(P_n) \to M$, then $X_+ = \cup P_n$ is clearly a positive set. Then $X_- = X_+^c$ is a negative set, for if $A \subset X \setminus P$ and $\nu(A) > 0$, then $\nu(A \cup P) = \nu(A) + \nu(P) > M$. Uniqueness up to null sets is a simple exercise. □

Definition 153. *1. The measure λ is absolutely continuous w.r.t. ν, written $\boxed{\lambda \ll \nu}$, if every null set of ν is a null set for λ.*

2. ν is concentrated on $X_1 \in \mathcal{M}$ if any measurable set $E \subset X_1^c$ is a null set.

3. ν_1 and ν_2 are mutually singular, $\boxed{\nu_1 \perp \nu_2}$, if ν_1 and ν_2 are concentrated on disjoint sets, X_1, X_2.

Exercise 46. *Check that*

1. *The relation $\ll$ is transitive.*
2. *$\perp$ is symmetric.*
3. *$(\nu \ll \lambda$ and $\nu \perp \lambda) \Rightarrow \nu = 0$.*
4. *$(\nu_1 \perp \nu$ and $\nu_2 \perp \nu) \Rightarrow \nu_1 + \nu_2 \perp \nu$.*

Theorem 154 (The Jordan decomposition theorem). *Any signed measure ν can be uniquely written as the difference of two mutually singular positive measures: $\nu = \nu^+ - \nu^-$.*

Proof. Take a Hahn decomposition $X = X_+ \cup X_-$, and define $\nu^+(A) = \nu(A \cap X_+)$ and $\nu^-(A) = -\nu(A \cap X_-)$. The rest is a simple exercise. □

Definition 155 (Total variation)**.** *If ν is a signed measure, its total variation $|\nu|$ is the positive measure $\nu^+ + \nu^-$.*

Exercise 47. *Check the following. N is a null set for ν iff it is null for $|\nu|$; thus ν and $|\nu|$ are mutually absolutely continuous, and $\nu \ll \mu \Leftrightarrow |\nu| \ll \mu$. We have $\nu \perp \lambda$ iff $|\nu| \perp \lambda$ iff both ν^+, ν^- are $\perp \lambda$. Also, $\nu \ll \mu \Leftrightarrow |\nu| \ll \mu$*

Lemma 156. *If $(X, \mathcal{M}, \mu)$ is a measure space and $f \in L^1(X)$, then $\nu := A \mapsto \int_A f dm$ is a measure on $\mathcal{M}$ and $\nu \ll \mu$.*

Proof. We have proved already that ν is a measure on $\mathcal{L}$. If $\mathcal{X}_A$ is the characteristic function of a null set, then $\mathcal{X}_A f = 0$ a.e. □

14 The Lebesgue-Radon-Nikodym theorem

This theorem is, in a sense, a converse of Lemma 156.

Theorem 157 (Lebesgue-Radon-Nikodym)**.** *1. Let μ and ν be finite measures on $X, \mathcal{M}$. Then there exists a $\mu-$null set N and an $f \in L^1(\mu)$ s.t. for every $A \in \mathcal{M}$,*

$$\nu(A) = \nu(A \cap N) + \int_A f d\mu \tag{54}$$

With $\lambda = A \mapsto \nu(A \cap N)$ we write $d\nu = d\lambda + f d\mu$.

2. *(Generalization) Let now ν be a signed $\sigma-$finite measure and μ a $\sigma-$finite positive measure on $X, \mathcal{M}$. Then there exists a unique decomposition $\nu = \lambda + \rho$ into $\sigma-$finite signed measures λ, ρ s.t. $\lambda \perp \mu$ and $\rho \ll \mu$. Furthermore, there is an f as above s.t. $d\rho = f d\mu$, uniquely defined a.e.*

Proof: G. Koumoulis, AMM, V115,6 (2008). The proof is based on a general strategy to construct such objects, by constructing f as the supremum of functions s.t. $\forall A$, $\int_A f d\mu \leqslant \nu(A)$.

If $\mathcal{F}$ is a countable family in $\mathcal{M}$ we let $\cup\mathcal{F} = \cup_{F\in\mathcal{F}} F$. We first show the following.

Lemma 158. *Let $X, \mathcal{M}, \mu)$ be a finite measure space. Then, for any family of measurable sets $\mathcal{E}$ there is a countable disjoint subfamily $\mathcal{F} \subset \mathcal{E}$ s.t. if $E \in \mathcal{E} \cap \mathcal{P}(X \setminus \cup\mathcal{F})$, then $\mu(E) = 0$.*

Proof 1, using the AC. (A proof without using the full AC is given below.) Let Z be the collection of subfamilies $\mathcal{G}$ of $\mathcal{E}$ consisting of disjoint, non-null sets. The partial order on Z is inclusion. Since μ is finite, any $\mathcal{G}$ as above is countable. Then $\mathcal{F}$ is any maximal element of Z. □

Proof 2, without the full AC. Let $\mathcal{G}$ be the collection of non-null sets in $\mathcal{E}$. We construct $\mathcal{F}$ as follows. Let $E_0 \in \mathcal{E}$. If $E_0 \in X \setminus E_0 \Rightarrow E_0 \notin \mathcal{P}$, then we are done. If not, let

$$k_1 = \min\left\{k \in \mathbb{N} : \exists E_1 \in \mathcal{E} \cap \mathcal{P}(X \setminus E_0) \text{ with } \mu(E_1) > k^{-1}\right\} \tag{55}$$

and choose an E_1 as above. Unless the construction ends in a finite number of steps, construct E_n similarly, replacing k_1 by k_n, E_1 by E_n, E_0 by $\cup_0^{n-1} E_j$. Note that the values k_i can repeat only finitely many times. Therefore, if $E \in \mathcal{E} \cap \mathcal{P}(X \setminus \cup_0^\infty E_j)$, then $\mu(E) = 0$.

□

We now prove 1. Let $H = \{h : X \to [0, \infty] : h \text{ measurable and } \forall A \in \mathcal{M}, \int_A h d\mu \leqslant \nu(A)\}$.

Clearly H is nonempty since $0 \in H$. Also, H is closed under taking the maximum of two functions, $h_1 \wedge h_2$. Indeed, if $X_1 = \{x : h_1(x) \geqslant h_2(x)\}$ and $X_2 = \{x : h_1(x) < h_2(x)\}$ then $X = X_1 \uplus X_2$, hence

$$\int_A h_1 \wedge h_2 d\mu = \int_{A\cap X_1} h_1 d\mu + \int_{A\cap X_2} h_2 d\mu \leqslant \nu(A \cap X_1) + \nu(A \cap X_2) = \nu(A)$$

Let $\alpha = \sup_H \|h\|_1$. Then $\alpha \leqslant \nu(X)$ and there is a sequence, which we can assume is increasing, of h_n s.t. $\int h_n d\mu \to \alpha$. By the monotone convergence theorem, $h_n \to f \in H$, $\int f d\mu = \alpha$. Redefining f on a null set we may assume $f : X \to [0, \infty)$.

Let $\lambda = A \mapsto \nu(A) - \int_A f d\mu$, a positive measure.

Lemma 159. *For any non-null $A \in \mathcal{M}$ and $n \in \mathbb{N}$, there is an $E \subset A$ s.t. $\mu(E) > n\lambda(E)$.*

Proof. For any $n \in \mathbb{N}$ and any $A \in \mathcal{M}$, $\int (f + n^{-1}\chi_A)d\mu > \alpha$, hence $f + n^{-1}\chi_A \notin H$. Thus there is a $B \in \mathcal{M}$ s.t. $\int_B (f + n^{-1}\chi_A)d\mu > \nu(B)$. Hence, $\mu(A \cap B) > n\left(\nu(B) - \int_B f d\mu\right) = \lambda(B) \geqslant n\lambda(A \cap B)$. □

For each n, define $\mathcal{E}_n = \{E \in \mathcal{M} : \mu(E) > n\lambda(E)\}$, and note that $\mathcal{E}_n$ are closed under countable unions. Clearly, there are no null sets in $\mathcal{E}_n$. For each $\mathcal{E}_n$ let $\mathcal{F}_n$ be as in Lemma 158. Defining $E_n = \cup \mathcal{F}_n$, we have $E_n \in \mathcal{E}_n$. Now we must have $\mu(X \setminus E_n) = 0$, or else, by the Lemma above, we would find an $E \subset X \setminus E_n$ in $\mathcal{E}_n$. Let $N = X \cup_j E_j$, a $\mu-$null set. Since $X \setminus N \subset \cap E_n$, we have $\lambda(X \setminus N) = 0$. Thus λ is concentrated on N and μ on $X \setminus N$, and

$$\nu(A) - \int_A f d\mu = \lambda(A) = \lambda(A \cap N) = \nu(A \cap N) - \int_{A \cap N} f d\mu = \nu(A \cap N)$$

.

2. If μ, ν are $\sigma-$finite positive measures, by taking intersections we can write $X = \cup A_j$ where A_j are disjoint and $\mu-$ and ν-finite. On each A_j we let $\mu_j = \mu \cap A_j, \nu_j = \nu \cap A_j, \lambda_j = \lambda \cap A_j$ and $f_j = f\chi_{A_j}$ as in 1. Then $\mu = \sum \mu_j$, $\nu = \sum \nu_j$ etc. is the desired decomposition. The signed measure case is an easy exercise. If we have two such functions f_1, f_2 then $\int_A (f_1 - f_2)d\mu = 0$ for all A implying uniqueness. □

Corollary 160. *Let ν and μ be measures. Then $\nu \ll \mu$ iff $\lim_n \mu(E_n) = 0 \Rightarrow \lim_n \nu(E_n) = 0$.*

Definition 161. *If $\nu \ll \mu$ and f is as in Theorem 171, then we write $f = \dfrac{d\nu}{d\mu}$.*

Corollary 162. *1. Assume ν, μ are $\sigma-$finite measures, μ is positive, $\nu \ll \mu$ and $\phi \in L^1(\mu)$. If $f = d\nu/d\mu$, then $f\phi \in L^1(\mu)$ and*

$$\int g d\nu = \int \phi f d\mu$$

2. If λ is a positive measure, $\mu \ll \lambda$ and $d\mu = g d\lambda$, then $d\mu/d\lambda = fg$.

Proof. By density of simple functions, since 1 and 2 hold for characteristic functions. □

Corollary 163. *If μ and ν are mutually absolutely continuous, then $(d\nu/d\mu) \neq 0$ a.e., and $d\mu/d\nu = 1/(d\nu/d\mu)$ a.e.*

The following result, whose proof is immediate, will be useful.

Proposition 164 (Existence of an upper bound)**.** *If $(\mu_j)_{j=1,\ldots,n}$ are measures, then $\mu_k \leqslant \sum_j \mu_j$ for all $k \leqslant n$.*

HW 11/13 (Recitation day) : 4,5,6,7 p. 88 in Folland; turn in: Ex 41,42 in the notes.

Lemma 165. *Let μ be a measure and $\nu \ll \mu$ a signed measure, both assumed $\sigma-$finite, and let f be s.t. $\nu = f d\mu$. Then $d|\nu| = |f| d\mu$.*

Proof. Let X_+ and X_- be the Hahn decomposition for ν. If $A_\pm \subset X_\pm$, then $\nu(A) = \int_{A_\pm} f d\mu$, which implies $\pm f$ are positive when restricted to $X_\pm$. Then, $f\chi_{X^+} = f^+$ and $f\chi_{X^-} = f^-$, and the rest is straightforward. □

14.1 Complex measures

Definition 166. *A complex measure on $(X, \mathcal{M})$ is a function $\nu : \mathcal{M} \to \mathbb{C}$ s.t.*

1. $\nu(\emptyset) = 0$.
2. *If $(A_j)_{j\in\mathbb{N}}$ are disjoint and measurable, then* $\nu(\bigcup_{j\in\mathbb{N}} A_j) = \sum_{j=1}^{\infty} \nu(A_j)$.

Note 167. 1. The range of a complex measure does not include the point at infinity: as we know, in the special case of signed measures, allowing for $\pm\infty$ leads to contradictions.

2. Convergence of the infinite sum in Condition 2. implies absolute convergence.
3. Writing $\nu = \nu_r + i\nu_i$, we see that ν_r and ν_i are signed measures with values in $\mathbb{R}$, hence $|\nu_r|(X), |\nu_i|(X)$ are both finite, and the range of ν is a bounded set in $\mathbb{C}$.

Definition 168. *The variation of a complex measure ν is the set function*

$$|\nu|(E) = \sup_{\uplus A_i = E} \sum_{i\in\mathbb{N}} |\nu(A_i)| \qquad \forall E \in \mathcal{M} \tag{56}$$

The total variation of ν is defined as $|\nu| = |\nu|(X)$.

Note 169. 1. Observe that $A \subset B$ implies $|\nu|(A) \leqslant |\nu|(B)$ and $|\nu|(X) \leqslant |\nu_r|(X) + |\nu_i|(X)$, and thus the set function $|\nu|$ is bounded. Clearly, $|\nu(A)| \leqslant |\nu|(A)$ for any measurable A.

2. The definitions of $\ll$, $\perp$ and their properties are the same as for signed measures.

Exercise 48. *1. Show that $|\nu|$ is finitely additive and continuous from below, and is thus a positive measure on $\mathcal{M}$.*

Lemma 170. *Let μ be a measure on $(X, \mathcal{M})$, and $f \in L^1(\mu)$ and define $\nu = A \mapsto \int_A f d\mu$ where $A \in \mathcal{M}$. Then, $|\nu|(A) = A \mapsto \int_A |f| d\mu$ for all $A \in \mathcal{M}$.*

Proof. Since $f \in L^1$, $\lim_{n\to\infty} |\mu|(|f| > n) = 0$. Since the measures and the σ-algebra can be restricted to any set, it is enough to prove this when $A = X$.

Choose $\epsilon > 0$ and let n be s.t. $|\mu|(|f| > 2n) < \epsilon$. Partition the box $B = \{z : |\Re(z)| \leqslant n, |\Im(z)| \leqslant n\}$ into N^2 congruent sub-boxes B_k. If $E_k = f^{-1}(B_k)$ and $E \subset E_k$, we have $\nu(E) = \alpha_{E,k}|\mu(E)|, \alpha_{k,E} \in B_k$ and thus $|\nu|(E) = |\alpha_{k,E}||\mu|(E)$. Since $|\nu|(X) = \sum_k |\nu|(E_k) = \sum_{k=1}^{N^2} |\alpha_k||\mu|(E_k)$, the result follows by taking $N \to \infty, \epsilon \to 0$ and noting that $\sum_k \alpha_k \chi_{E_k} + n\chi_{|f|>n}$ converge pointwise to $|f|$. ☐

The following generalization is immediate.

Theorem 171 (Lebesgue-Radon-Nikodym, L-R-N)**.** *Let ν be a complex measure and μ a $\sigma-$finite positive measure on $X, \mathcal{M}$. Then there exists a unique decomposition $\nu = \lambda + \rho$ into complex measures λ, ρ s.t. $\lambda \perp \mu$ and $\rho \ll \mu$. Furthermore, there is an f s.t. $d\rho = f d\mu$, uniquely defined a.e.*

Corollary 172. *We have $d\nu = f d|\nu|$ where $f \in L^1$, and $|f| = 1$ a.e.*

Proof. Note 169 shows that $\nu \ll |\nu|$. By Exercise 48, we have $d|\nu| = |f| d|\nu|$, and using uniqueness of the L-R-N derivative, $|f| = 1$ a.e. ☐

15 Differentiation

One of the new major ideas of calculus was the discovery of duality between areas to tangents expressed by the fundamental theorem of calculus. The extension to Lebesgue integrals in $\mathbb{R}^n$ requires significant technical machinery and in the process we will encounter two important objects in analysis. We start with the following elementary theorem.

Theorem 173. *Let μ be a complex Borel measure on $\mathbb{R}^1$ and let F be its distribution function, $F(x) = \mu((-\infty, x])$. Then the following statements are equivalent:*

1. *F is differentiable at x and $f'(x) = A$.*
2. *For every $\epsilon > 0$ there is a $\delta > 0$ s.t.*
$$\left| \frac{\mu(I)}{m(I)} - A \right| < \epsilon$$
for any open interval of length $< \delta$ containing x.

Proof. Straight from the definition of differentiation. □

We want to extend this type of result to $\mathbb{R}^k$, where k will be the same throughout this section. We also denote $B_{x,r} = \{x' \in \mathbb{R}^k : |x' - x| < r\}$. Let μ be a complex Borel measure on $\mathbb{R}^k$. Consider the quotients
$$(Q_r\mu)(x) = \frac{\mu(B_{x,r})}{m(B_{x,r})} \tag{57}$$
where $m = m^k$ is the Lebesgue measure.

Definition 174. *The symmetric derivative $D\mu$ at x is defined as*
$$(D\mu)(x) = \lim_{r \to 0} (Q_r \mu(x))$$
at those points where the limit exists.

Theorem 175 (The Vitali covering theorem, finite version)**.** *If $\mathcal{O}$ is the union of a finite collection of balls B_{x_i,r_i}, $1 \leqslant i \leqslant N$, then there exists a set $S \subset \{1, ..., N\}$ so that*

1. *The balls B_{x_i,r_i} with $i \in S$ are disjoint*
2. *$\mathcal{O} \subset \cup_{i \in S} B_{x_i,3r_i}$.*
3. *$m(\mathcal{O}) \leqslant 3^k \sum_{i \in S} m(B_{x_i,3r_i})$.*

Proof. A key (elementary) property here, that you should check, is:

Claim. If $r' \leqslant r$ and $B_{x',r'} \cap B_{x,r} \neq \emptyset$, then $B_{x',r'} \subset B_{x,3r}$.

Re-index the set so that $r_1 \geqslant r_2 \geqslant \cdots \geqslant r_N$. Let B_1 be the first one and discard all other balls that intersect B_1. If there is any left, choose the first and call it B_2, and so on until the process terminates with some B_n. The collection is clearly disjoint, and by the claim, $\mathcal{O} \subset \cup_j B_j$ proving 2., and by the scaling properties of the Lebesgue measure, 3. follows. □

Definition 176 (Weak L^1). *Weak L^1 is defined as*

$$WL^1 = \{f \text{ measurable } : \|f\|_{WL^1} := \sup_{\lambda>0} \lambda m(|f| > \lambda) < \infty\}$$

Note 177. *We have $L^1 \subsetneq WL^1$: Markov's inequality shows the inclusion and $x \mapsto 1/x$ in $\mathbb{R}$ shows that it is strict.*

The **Hardy-Littlewood maximal operator** takes a locally integrable function $f : \mathbb{R}^k \to \mathbb{C}$ and returns another function Mf that, at each point $x \in \mathbb{R}^k$, gives the maximum average value that $|f|$ can have on balls centered at that point.

Definition 178. *The Hardy-Littlewood maximal operator of Mf is given by*

$$Mf(x) = \sup_{r>0} \frac{1}{m(B_{x,r})} \int_{B_{x,r}} |f(y)|\, dy$$

The maximal function of a positive *measure μ is defined by*

$$(M\mu)(x) = \sup_{r>0} (Q_r\mu)(x)$$

The maximal function of a complex measure μ is $M|\mu|$.

Lemma 179. *Let μ be a positive Borel measure. The function $M\mu : \mathbb{R}^k \to [0, \infty]$ is lower semicontinuous, hence measurable.*

Proof. Let $E = \{M\mu > \lambda\}$ for some $\lambda > 0$ and $x \in E$. There is an r and a $\lambda' > \lambda$ s.t.

$$\mu(B_{x,r}) = \lambda' m(B_{x,r})$$

Let $\delta > 0$ be s.t.

$$\frac{\lambda'}{\lambda} > \frac{(r+\delta)^k}{r^k}$$

If $|x' - x| < \delta$, then $B_{x',r+\delta} \supset B_{x,r}$ and therefore

$$\mu(B_{x',r+\delta}) \geqslant \lambda' m(B_{x,r}) = \frac{\lambda' r^k}{(r+\delta)^k}\, m(B_{x',r+\delta}) > \lambda m(B_{x',r+\delta})$$

Hence $B_{x,\delta} \subset E$, proving that E is open. □

Theorem 180 (Weak Type Estimate)**.** *If μ is a complex Borel measure on $\mathbb{R}^k$ and $\lambda > 0$, then*

$$m(M\mu > \lambda) \leqslant 3^k \lambda^{-1} |\mu|$$

In particular, for $k \geqslant 1$ and $f \in L^1(\mathbb{R}^k)$ there is a constant $C_k > 0$ s.t. for all $\lambda > 0$, we have:

$$m(Mf > \lambda) < 3^k \lambda^{-1} \|f\|_{L^1(\mathbb{R}^k)}$$

The second statement reads: M is a continuous operator from L^1 to weak L^1 with a bound 3^k.

The following strong-type estimate is an immediate consequence of the Weak Type Estimate and the Marcinkiewicz interpolation theorem (that we'll study in Chapter 5):

Theorem 181 (Strong Type Estimate). *For $k \geqslant 1$ and $f \in L^p(\mathbb{R}^k), 1 < p \leqslant \infty$ there is a constant $C_{pk} > 0$ s.t.*

$$\|Mf\|_{L^p(\mathbb{R}^k)} \leq C_{pk}\|f\|_{L^p(\mathbb{R}^k)}$$

This statement reads: M is a continuous operator from L^p to L^p for any $p > 1$.

Proof of Theorem 180. Fix μ and $\lambda > 0$ Let K be a compact subset of $\{M\mu > \lambda\}$. If $x \in K$, then for some $\delta > 0$

$$|\mu|(B_{x,\delta}) > \lambda m(B_{x,\delta})$$

Extract a finite collection from these $B_{x,\delta}$ which cover K. By the finite Vitali covering theorem it contains a disjoint subcollection $B_j, ..., B_n$ that satisfies

$$m(K) \leqslant 3^k \sum_1^n m(B_i) \leqslant 3^k \lambda^{-1} \sum_1^n |\mu|(B_i) \leqslant 3^k |\mu| \lambda^{-1}$$

where the last inequality uses the disjointness of the balls. The regularity of Borel measures completes the proof. □

15.1 Lebesgue points

Definition 182. *Let $f \in L^1(\mathbb{R}^k)$. The point $x \in \mathbb{R}^k$ is a Lebesgue point of f if*

$$\lim_{r \to 0} \frac{1}{m(B_{x,r})} \int_{B_{x,r}} |f - f(x)| dm = 0 \tag{58}$$

Note 183. *Clearly, if f is continuous, then all points are Lebesgue points.*

The following is a fundamental result in the theory of Lebesgue differentiation.

Theorem 184. *If $f \in L^1(\mathbb{R}^k)$, then almost every $x \in \mathbb{R}^k$ is a Lebesgue point of f.*

Proof. Let

$$(T_r f)(x) = \frac{1}{m(B_{x,r})} \int_{B_{x,r}} |f - f(x)| dm \quad \text{and} \quad (Tf)(x) = \limsup_{r \to 0} (T_r f)(x)$$

We will show that $(Tf)(x) = 0$ a.e. Let $n \in \mathbb{N}$, choose $g \in C(\mathbb{R}^k)$ s.t. $\|f - g\|_1 < 1/n$ and denote $h = f - g$. Since g is continuous, we have $Tg = 0$. Simply writing $|h(y) - h(x)| \leqslant |h(y)| + |h(x)|$, we get, for any x,

$$(Th)(x) \leqslant |h(x)| + \sup_r \frac{1}{m(B_{x,r})} \int_{B_{x,r}} |h| dm = |h(x)| + (Mh)(x)$$

Now, $T_r f \leqslant T_r g + T_r h = T_r h$, implying

$$Tf \leqslant Mh + |h|$$

Let $\lambda > 0$. The set $\{x : (Tf)(x) > 2\lambda\}$ is contained in the measurable set $\{x : (Mh)(x) > \lambda \text{ or } |h(x)| > \lambda\}$ whose measure is

$$\leqslant m(Mh > \lambda) + m(|h| > \lambda) \leqslant \lambda^{-1}(3^k + 1)n^{-1}$$

Since this holds for any n it follows that $\{Tf > 2\lambda\}$ is contained in a null set. Now $\{x : (Tf)(x) > 0\} \subset \{x : \exists m > 0 (Tf)(x) > m^{-1}\}$, also a null set. □

15.1.1 Differentiation of absolutely continuous measures

Theorem 185. *Assume μ is a complex Borel measure on $\mathbb{R}^k$ and that $\mu \ll m$. Then $D\mu$ (cf. Definition 174) exists a.e. and equals $d\mu/dm$.*

Proof. Let $f = d\mu/dm$. Then,

$$f(x) = \lim_{r\to 0} \frac{1}{m(B_{x,r})} \int_{B_{x,r}} f\, dm = \lim_{r\to 0} \frac{\mu(B_{x,r})}{m(B_{x,r})} \ \ a.e.[m]$$

Thus, $(D\mu)(x)$ exists and equals $f(x)$ at every Lebesgue point of f. □

15.1.2 Nicely shrinking sets

Definition 186. *Let $x \in \mathbb{R}^k$. The sequence $(E_n)_{n\in\mathbb{N}}$ of Borel sets is said to shrink nicely to x if there is an $\alpha > 0$ and a sequence of balls $(B_{x,r_n})_{n\in\mathbb{N}}$ s.t. $r_n \to 0$ and for all j $E_j \subset B_{x,r_j}$ and $m(E_j) \geqslant \alpha m(B_{x,r_j})$*

Theorem 187. *Assume for each $x \in \mathbb{R}^k$ the sequence $(E_n(x))_{n\in\mathbb{N}}$ shrinks nicely to x. Let $f \in L^1$. Then, at every Lebesgue point of f we have*

$$\lim_{n\to\infty} \frac{1}{m(E_n(x))} \int_{E_n(x)} f dm = f(x)$$

(local averages of integrable functions converge to their local values.)

Proof. Write the result in the equivalent form

$$\lim_{n\to\infty} \frac{1}{m(E_n(x))} \int_{E_n(x)} |f - f(x)| dm = 0 \tag{59}$$

If the (E_n) are balls, then (59) holds at any Lebesgue point of f. The result now follows by easy estimates since, for some sequence of balls we have $m(B_{x,r_n}) \geqslant m(E_n) \geqslant \alpha m(B_{x,r_n})$. □

Proposition 188. *Let μ be a complex Borel measure s.t. $\mu \perp m$. Then*

$$D\mu = 0 \ \ a.e.\ [m]$$

Proof. Clearly, it is enough to show this for positive measures. Define now $(M_n\mu)(x) = \sup_{0<r<n^{-1}} (Q_r\mu)(x)$. In the same way as for M, we can check that M_n is upper semicontinuous, and thus

$$(\overline{D}\mu)(x) := \lim_{n\to\infty} (M_n\mu)(x) \tag{60}$$

is measurable. Note also that $M_n\mu \leqslant M\mu$.

Choose $\lambda > 0, \epsilon > 0$ and a compact set K s.t., by the regularity of Borel measures, $\mu(K) > |\mu| - \epsilon$. Let μ_1 be the restriction of μ to K, and $\mu_2 = \mu - \mu_1$. We see that $|\mu_2| < \epsilon$, and if $x \in K^c$ we have

$$(\overline{D}\mu)(x) = (\overline{D}\mu_2)(x) \leqslant (M\mu_2)(x)$$

hence

$$m(\overline{D}\mu > \lambda) \leqslant m(K) + m(M\mu_2 > \lambda) \leqslant 3^k\lambda^{-1}|\mu_2| \leqslant 3^k\lambda^{-1}\epsilon \tag{61}$$

Since (61) holds for arbitrary $\epsilon > 0, \lambda > 0$, the result follows. □

Corollary 189. *Assume that for each $x \in \mathbb{R}^k$ the sequence $(E_k(x))_k$ shrinks nicely and μ is a complex Borel measure s.t. $\mu \perp m$. Then*

$$\lim_{k\to\infty} \frac{\mu(E_k(x))}{m(E_k(x))} = 0 \ \ a.e.[m] \tag{62}$$

As another corollary, we have the following strengthening of Theorem 185.

Theorem 190. *Assume for each $x \in \mathbb{R}^k$ the sequence $(E_k(x))_k$ shrinks nicely and μ is a complex Borel measure on $\mathbb{R}^k$. Let $d\mu = d\lambda + f dm$ be the L-R-N decomposition of μ. Then,*

$$\lim_{k\to\infty} \frac{\mu(E_k(x))}{m(E_k(x))} = f(x) \ a.e.\ [m]$$

(in particular, the limit exists a.e.)

15.2 Metric density

Definition 191. *Let E be Lebesgue measurable in $\mathbb{R}^k$. The metric density of E at x is*

$$\lim_{r\to 0} \frac{m(E \cap B_{x,r})}{m(B_{x,r})}$$

when the limit exists.

Proposition 192. *The metric density of a Lebesgue measurable set in $\mathbb{R}^k$ exists a.e., and it is 1 for a.e. for x in E and 0 a.e. in E^c.*

Proof. Write the metric density using the characteristic function of E. □

We see that for $x \in \mathbb{R}$, either most points in tiny neighborhoods of x are in E or most points are in the complement! This property has a topological flavor to it. See Approximate continuity, below.

Approximate continuity, from the Encyclopedia of Mathematics

Consider a (Lebesgue)-measurable set $E \subset \mathbb{R}^n$, a measurable function $f : E \to \mathbb{R}$ and a point $x_0 \in \mathbb{R}^n$ where E has Lebesgue density 1. The approximate upper and lower limits of f at x_0 are defined, respectively, as

1. The infimum of $a \in \mathbb{R} \cup \{\infty\}$ such that the set $\{f \leq a\}$ has density 1 at x_0;
2. The supremum of $a \in \{-\infty\} \cup \mathbb{R}$ such that the set $\{f \geq a\}$ has density 1 at x_0

They are usually denoted by

$$\operatorname{ap}\limsup_{x\to x_0} f(x) \qquad \text{and} \qquad \operatorname{ap}\liminf_{x\to x_0} f(x)$$

(some authors use also the notation $\overline{\lim}\,\mathrm{ap}$ and $\underline{\lim}\,\mathrm{ap}$). It follows from the definition that $\operatorname{ap}\liminf \le \operatorname{ap}\limsup$: if the two numbers coincide then the result is called approximate limit of f at x_0 and it is denoted by

$$\operatorname{ap}\lim_{x\to x_0} f(x)\,.$$

The approximate limit of a function taking values in a finite-dimensional vector space can be defined using its coordinate functions and the definition above.

Observe that the approximate limit of f and g are the same if f and g differ on a set of measure zero. A useful characterization of the approximate limit is given by the following

Proposition 193. *Consider a (Lebesgue)-measurable set $E \subset \mathbb{R}^n$, a measurable function $f : E \to \mathbb{R}$ and a point $x_0 \in \mathbb{R}^n$. f has approximate limit L at x_0 if and only if there is a measurable set $F \subset E$ which has density 1 at x_0 and such that*

$$\lim_{x\in F, x\to x_0} f(x) = L\,.$$

In general, the existence of an ordinary limit does not follow from the existence of an approximate limit. An approximate limit displays the elementary properties of limits –uniqueness, and theorems on the limit of a sum, difference, product and quotient of two functions– these properties follow indeed easily from Proposition 193.

If the domain E of f is a subset of $\mathbb{R}$ we can define one-sided (right and left) approximate upper and lower limits: we just substitute all *density* 1 requirements with the *right-hand* or the *left-hand* density 1 requirement, that are, respectively,

$$\lim_{r\downarrow 0} \frac{\lambda(G\cap]x_0, x_0 + r[)}{r} = 1 \qquad \text{and} \qquad \lim_{r\downarrow 0} \frac{\lambda(G\cap]x_0 - r, x_0[)}{r} = 1$$

for a generic measurable set $G \subset \mathbb{R}$ (here λ denotes the Lebesgue measure on $\mathbb{R}$). For instance, to define the approximate upper limit L at x_0 of a function $f : E \to \mathbb{R}$ we require that the right-hand density of E at x_0 is 1: L is then the infimum of the numbers $a \in \mathbb{R}\cup\{\infty\}$ such that $\{f \le a\}$ has right-hand density 1 at x_0. The corresponding notation is

$$\operatorname{ap}\limsup_{x\to x_0^+} f(x)\,.$$

Approximate limits are used to define approximately continuous and approximate differentiable functions.

Definition 194. *Consider a (Lebesgue) measurable set $E \subset \mathbb{R}^n$, a measurable function $f : E \to \mathbb{R}^k$ and a point $x_0 \in E$ where the Lebesgue density of E is 1. f is approximately continuous at x_0 if and only if the approximate limit of f at x_0 exists and equals $f(x_0)$.*

It follows from Lusin's theorem that a measurable function is approximately continuous at almost every point. Points of approximate continuity are related to Lebesgue points. A Lebesgue point is always a point of approximate continuity. Conversely, if f is essentially bounded, the points of approximate continuity of f are also Lebesgue points.

16 Total variation, absolute continuity

This section is devoted to Borel measures and measurable functions on $\mathbb{R}$. Given that a complex measure μ can be uniquely decomposed into positive measures μ_i: $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$, for many of the results below we can assume wlog that μ itself is positive. (The same applies to complex measurable functions.)

Recall the definitions of the distribution function of a measure (p. 23) and of the variation of a complex measure (p. 58).

Exercise 49. *Show that the following is an equivalent definition of the variation of a measure:*

$$|\nu|(E) = \sup\left\{\sum_{i=1}^{n} |\nu(A_i)| : n \in \mathbb{N}, \biguplus_{1}^{n} A_i = E\right\} \qquad \forall E \in \mathcal{M} \tag{63}$$

Definition 195. *Let μ be a complex Borel measure and take its canonical decomposition into four positive measures μ_i. Let F_i be the distribution functions of μ_i. We define the distribution function of μ as $F_\mu = F_1 - F_2 + i(F_3 - F_4)$. Equivalently, $F_\mu(x) = \mu((-\infty, x])$.*

Let $F = F_\mu$ be the distribution function of the complex Borel measure μ. We define the total variation function of F as $T_F(x) = |\mu|((-\infty, x])$.

Exercise 50. *Let μ and F be as in the definition above.*

1. *Show that*

$$T_F(x) = \sup\left\{\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, -\infty < x_0 < \dots < x_n = x\right\} \tag{64}$$

2. *Using the fact that $|\mu|$ is a Borel measure, show that T_F is increasing and right-continuous.*

Definition 196. *If $F : \mathbb{R} \to \mathbb{C}$, we define the total its variation function T_F by (64).*

We say that F is of bounded variation, $F \in BV$, if $\lim_{x\to\infty} T_F(x) < \infty$. The total variation of F on $[a, b]$ is defined by

$$T_F([a,b]) := \sup\left\{\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, a = x_0 < \dots < x_n = b\right\} = T_F(b) - T_F(a)$$

Note 197 (Geometrical interpretation)**.** Since all norms on $\mathbb{R}^k$ are equivalent, a real-valued function is in $BV[a, b]$ iff

$$\sup\left\{\sum_{j=1}^{n} \sqrt{(x_j - x_{j-1})^2 + (F(x_j) - F(x_{j-1}))^2} : n \in \mathbb{N}, a = x_0 < \dots < x_n = b\right\} < \infty$$

that is, the lengths of the polygonal lines with vertices on the graph of F are bounded. Thus, $F \in BV$ iff the graph $\{(x, F(x)) : x \in [a, b]\}$, completed by vertical lines at the points of discontinuity, is a *rectifiable curve*.

Note 198. Let $F \in BV$. As sees in Exercise 50, T_F is increasing, and thus $T_F(\pm\infty) = \lim_{x\to\pm\infty} T_F(x)$, exist. Then, for any sequence $x_j \searrow -\infty$, $\sum_j |F(x_j) - F(x_{j-1})| < \infty$. This implies that for any ϵ there is an x_0 so that $\sum_{j\geqslant 0} |F(x_j) - F(x_{j-1})| < \epsilon$ for any decreasing unbounded sequence $x_0, x_1, \ldots$, hence $T_F(-\infty) = 0$.

Note 199.

1. If F is real-valued and monotonic and $[a,b] \subset \mathbb{R}$, then the total variation of F on $[a,b]$ is finite. (Indeed, with $x_j > x_{j-1}$ we have $|F(x_j) - F(x_{j-1})| = F(x_j) - F(x_{j-1})$.)
2. If F is monotonic on $\mathbb{R}$, then $F(x^+)$ and $F(x^-)$ exist for all $x \in \mathbb{R}$, and they define a right continuous and left continuous function, resp. Furthermore, for all $x \in \mathbb{R}$ we have
$$F(x) \in [F(x^-), F(x^+)] \tag{65}$$
In particular, the points of discontinuity F are exactly those of $x \mapsto F(x^+)$, and therefore F only has jump discontinuities, and there are at most countably many of them. (Recall Exercise 18.)
3. Linear combinations of BV functions are BV functions.
4. If F is real-valued, then the functions $T_F \pm F$ are increasing. Indeed, if $b > a$ then $T_F(b) - T_F(a) \geqslant |F(b) - F(a)|$, hence $T_F(b) - F(b) \geqslant T_F(a) - F(a)$ and $T_F(b) + F(b) \geqslant T_F(a) + F(a)$.
5. If $F \in BV$ is real-valued, then F is bounded and can be written as the difference of two increasing bounded functions: $F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F)$. This is, in a sense, a converse of 1.+3. The fact that F is bounded follows from $|F(y) - F(x)| \leqslant T_F(+\infty) - T_F(-\infty) = T(+\infty)$.
6. By 5. and 2., if $F \in BV$, then F has lateral limits at any point and has at most countably many discontinuities.
7. If $F \in BV$, then $|F| \in BV$. This follows from the triangle inequality: $||a| - |b|| \leqslant |a - b|$.

Exercise 51.

1. Let μ be the Borel measure with distribution function $G(x)$. Use Theorem 190 to show that $f = G'$ exists a.e., is in L^1 and $f = d\mu/dm$.

Proposition 200. *Let $F \in BV$ and denote $G^\pm(x) = F(x^\pm)$. Then, F' exists and equals $G^{\pm\prime}$ a.e.*

Proof. We can assume wlog that F is increasing. Note that G^+ and and $x \mapsto -G^-(-x)$ are right-continuous and increasing, thus $G^\pm$ are differentiable a.e. So it suffices to prove the result for $G = G^+$. Let $H = G - F$, a positive function and let S be the countable set of singularities of F (which is also the set of singularities of G) and let $\rho(s) = G^+(s) - G^-(s)$. Define a measure on $\mathbb{R}$ by $\lambda(A) = \sum_{s\in S} \rho(s)$ (compare with Exercise 18). Clearly $\lambda \perp m$. Check that for any x, y we have
$$|H(y) - H(x)| \leqslant \lambda([x,y]) \tag{66}$$
and note that (66) implies that $H' = 0$ a.e. □

16.1 NBV, AC

The space of normalized functions of bounded variation is defined as

$$NBV = \{F \in BV : \ F \text{ right continuous and } F(-\infty) = 0\}$$

Proposition 201. *If $F \in BV$ is right continuous, then $T_F \in NBV$.*

Proof. We have already shown that $T_F(-\infty) = 0$. Take a sequence $x_n \searrow x_0$ and, for each n, a sequence of finite partitions $(x_j^{(m)})$ of $[x_0, x_n]$ s.t. $\lim_{m\to\infty} \sum_j |F(x_j^{(m)}) - F(x_{j-1}^{(m)})| = T_F(x_n) - T_F(x_0)$. Noting once more that, for any $a < b$, we have $T_F(b) - T_F(a) \geqslant |F(b) - F(a)|$, the rest follows by dominated convergence. □

Note 202. If $F \in BV$ is right-continuous, then G given by $F - F(a)$ on $[a,b]$, zero for $x < a$ and $F(b)$ for $x > b$ is right-continuous and in NBG. Define μ_F on $[a,b]$ as μ_G restricted to $[a,b]$.

Theorem 203. *$F \in NBV$ iff $F(x)$ defines a Borel measure, $F(x) = \mu((-\infty, x])$.*

Proof. $F \in NBV$ iff both $\Re F$ and $\Im F$ are in NBV, and thus we may assume wlog that F is real-valued. Now Proposition 201 implies $T_F \in NBV$ and thus the two increasing functions in the canonical decomposition of F are also NBV, by Note 199 3. The rest is immediate. □

Definition 204. *If $F \in BV([a,b])$ is right-continuous, we say that F is absolutely continuous if $\mu_F \ll m$ where μ_F is defined on $[a,b]$ as in Note 202.*

Note 205. $F \in AC([a,b]) \Rightarrow F \in BV([a,b])$ and F is continuous. Continuity is clear. To bound the variation of F, take a pair ϵ, δ as in the definition of AC and choose $n > \delta^{-1}$. Partitioning $[a,b]$ into n congruent subintervals, we see that the total variation of F cannot exceed $n\epsilon$.

However, AC is a strictly stronger condition than continuity+BV. Take the Cantor function F: it is continuous and increasing, thus BV. Since F is constant on any excluded interval, with b_j the left endpoint of the intervals excluded up to stage n, and a_j the right endpoint of the preceding interval, we have

$$\sum_{j \leqslant 2^{n+1}-1} |F(b_j) - F(a_j)| = 1$$

while

$$\sum_{j \leqslant 2^{n+1}-1} (b_j - a_j) = \left(\frac{2}{3}\right)^n$$

Proposition 206. *If $F \in BV([a,b])$ is right-continuous, then $\mu_F \ll m$ iff for any $\epsilon > 0$ there is a $\delta > 0$ s.t. for any* **finite disjoint collection of intervals** $(a_j, b_j) \subset [a,b]$, $j = 1, ..., n$

$$\sum_{j=1}^{n} (b_j - a_j) < \delta \Rightarrow \sum_{j=1}^{n} |F(b_j) - F(a_j)| < \epsilon \tag{67}$$

Proof. If $\mu_F \ll \mu$, then (67) is a direct consequence of the definition of absolute continuity. Conversely, fix $\epsilon > 0$, let δ (67) be s.t. the last inequality in (67) holds with $\epsilon/2$ instead of ϵ. Let E be s.t. $m(E) < \delta$. Choose an open set $\mathcal{O} \supset E$ s.t. $m(\mathcal{O}) < \delta$, write $\mathcal{O}$ as a countable union of intervals J_k, and let n be s.t. $\sum_n^\infty |\mu|(J_k) < \epsilon/2$. Then $\sum_1^\infty |\mu|(J_k) < \epsilon$. □

Proposition 207. *If $f \in L^1([a,b], m)$, then $F := x \mapsto \int_a^x f(s)ds \in AC$. Conversely, $F \in AC([a,b])$ implies $f = F'$ exists a.e., and $F(x) = F(a) + \int_a^x f(s)ds$.*

Proof. This is an immediate consequence of Lemma 156 and Theorem 187. □

We have proved the following important result:

Theorem 208 (The Fundamental Theorem of Calculus). *Let $[a,b] \subset \mathbb{R}$ and $F : [a,b] \to \mathbb{C}$. The following are equivalent:*

1. *$F \in AC([a,b])$;*
2. *$F(x) = F(a) + \int_a^x f(s)ds$ for some $f \in L^1([a,b],m)$;*
3. *F' exists a.e., $F' \in L^1([a,b],m)$ and $F(x) = F(a) + \int_a^x F'(s)ds$*

An interesting result (see Rudin) is

Theorem 209. *Let $F : [a,b] \to \mathbb{C}$ be differentiable* **everywhere** *with derivative in L^1. Then, for any $x \in [a,b]$,*

$$F(x) = F(a) + \int_a^x F'(s)ds$$

16.2 Lebesgue-Stieltjes integrals

Let $F \in NBV$ and μ_F the associated measure. Then $\int g d\mu_F$ is also denoted by $\int g dF$, and is called a Lebesgue-Stieltjes integral.

Proposition 210. *If F and G are continuous and in NBG, then FG is continuous and in NBG and $d(FG) = FdG + GdF$. We can write this as a generalization of integration by parts:*

$$\int_{(a,b]} FdG + \int_{(a,b]} GdF = F(b)G(b) - F(a)G(a)$$

Proof. This follows from continuity and the fact that

$$F(y)G(y) - F(x)G(x) = G(y)(F(y) - F(x)) + F(x)(G(y) - G(x))$$

□

16.3 Types of measures on $\mathbb{R}^n$

Definition 211. *Let μ be a complex Borel measure on $\mathbb{R}^n$. Then,*

1. *μ is called* **discrete** *if $\mu = \sum_{j\in\mathbb{N}} \mu(\{x_j\})\delta_{x_j}$ for a discrete set $\{x_j : x \in \mathbb{N}\} \subset \mathbb{R}^k$ (δ_x is the Dirac mass at x);*
2. *μ is called* **continuous** *if $\mu(\{x\}) = 0$ for all $x \in \mathbb{R}$.*
3. *Let μ be continuous, and $d\mu = d\lambda + fdm$ be the L-R-N decomposition of μ. Then $\mu_{sc} = \lambda$ is the* **singular continuous** *part of μ, and $\mu_{ac} = fdm$ is the* **absolutely continuous** *part of μ*

Note 212. *Clearly, if μ is not continuous, then there exists an at most countable set of points $\{x_j : x \in \mathbb{N}\} \subset \mathbb{R}^k$ with $\mu(\{x_j\}) \neq 0$. Let $\mu_d = \sum_{j\in\mathbb{N}} \mu(\{x_j\})\delta_{x_j}$ is continuous. Then $\mu - \mu_d$ is continuous. Therefore, any complex Borel measure on $\mathbb{R}^k$ can be uniquely decomposed as*

$$\mu = \mu_d + \mu_{sc} + \mu_{ac}$$

An example of a measure which is singularly continuous is dF where F is Cantor's function.

Exercise 52. A function F is said to have the Lusin N property on $[a, b]$ if for any null subset N, $m(f(N)) = 0$.

- Let f be continuous and increasing. Show that it has the Lusin property iff it is AC. Some hints: for the if part monotonicity allows you to write the forward image of an interval. For the only if part, (1) define $G(x) = x + F(x)$ and show that G is continuous, increasing and has the Lusin property. (2) Show that $\mu(A) = m(G(A))$ defines a positive bounded measure, and use L-R-N to complete the proof.

HW 11/28: 31,37,42 in Folland; turn in: Ex 52 above.

For a history of some important theorems in topology, see Folland's article A Tale of Topology.

17 Semicontinuouos functions

Definition 213. *Let f be a real-valued (or extended-real valued [10]) function on X, a topological space. Then f is called* **lower semicontinuous** *if for any $\alpha \in \mathbb{R}$ the set*

$$\{x : f(x) > \alpha\}$$

is open, and **upper semicontinuous** *if for any $\alpha \in \mathbb{R}$ the set*

$$\{x : f(x) < \alpha\}$$

is open.

Check that a function $f : X \to \mathbb{R}$ is continuous iff it is both upper and lower semicontinuous. Examples of functions that are only semi-continuous are:

a. Characteristic functions of **open sets**: these are lower semicontinuous.

b. Characteristic functions of **closed sets**: these are upper semicontinuous.

c. The sup of any collection of lower semicontinuous functions is lower semicontinuous. The inf of any collection of upper semicontinuous functions is upper semicontinuous.

Though it's straightforward, it's useful to go through the arguments and check all this.

17.1 Urysohn's lemma

In a normal space, closed sets are separated by open sets. It means, if C_1, C_2 are closed, then there *are disjoint open sets* $\mathcal{O}_1, \mathcal{O}_2$ containing C_1, C_2, respectively. This property is, interestingly, equivalent to an apparently stronger property, that there is a continuous function f which is zero on C_1 and one on C_2. That is, indicator functions can be smoothened in a way that does not alter their functionality.

Note 214. *In a normal space, for any closed set C and open set $\mathcal{O} \supset C$ there is an open sets $\mathcal{O}_1$ s.t.*

$$C \subset \mathcal{O}_1 \subset \overline{\mathcal{O}_1} \subset \mathcal{O}$$

(check this: $C \cap \overline{\mathcal{O}^c} = \emptyset$; thus, we can separate C from $\overline{\mathcal{O}^c}$ by open sets...)

[10] In the sense of the one point compactification of $\mathbb{R}$.

Theorem 215 (Urysohn's lemma)**.** *Let X be normal. For any two nonempty closed disjoint subsets A, B of X, there is an $f \in C(X, [0,1])$ such that $f(A) = \{0\}$ and $f(B) = \{1\}$.*

Equivalently,

"For any $C \subset \mathcal{O}$, C closed and $\mathcal{O}$ open, there is an $f \in C(X, [0,1])$ such that $f(C) = \{1\}$ *and* $f(\mathcal{O} \setminus \mathcal{O}_1) = \{0\}$ ($\mathcal{O}_1$ as above)."

Note that this does not say that f can only be zero on A, or 1 on B, a property which is *stronger*. This theorem is quite deep. The idea is to squeeze a countably infinite family of (distinct) open sets between A and B, order them using the rationals in $[0,1]$, $\{\mathcal{O}_r\}_{r\in\mathbb{Q}}$ in such a way that the order of the rationals is preserved

$$s > r \Rightarrow \overline{\mathcal{O}_s} \subset \mathcal{O}_r \tag{(*)}$$

(meaning also that the sets are densely ordered.) Define $f(x) = r$ if $x \in \mathcal{O}_r$ and extend f by continuity. Basically.

It's not obvious that such a construction is possible and that it yields the right answer; we need more work.

Proof (following Rudin). Let $r_0 = 0$, $r_1 = 1$, and let $Q = (r_2, r_3, r_4, ...)$ be an enumeration of the rationals in $(0,1)$. Let $\mathcal{O}_0, \mathcal{O}_1$ be open sets such that

$$C \subset \mathcal{O}_1 \subset \overline{\mathcal{O}_1} \subset \mathcal{O}_0 \subset \overline{\mathcal{O}_0} \subset \mathcal{O}$$

Inductively, suppose that for all $n \geqslant 1$ we have constructed $\mathcal{O}_{r_1}, ..., \mathcal{O}_{r_n}$ so that for all $i, j \leqslant n$ we have

$$r_j > r_i \Rightarrow \overline{\mathcal{O}_{r_j}} \subset \mathcal{O}_{r_i}$$

Order the $r_i, i \leqslant n : 0 < r'_1 < ... < r'_n < 1$. Take the next rational in r_{n+1} in Q, and find the i so that

$$0 < r'_1 < r'_2 < ... < r'_i < r'_{n+1} \equiv r_{n+1} < r'_{i+1} < ... < r'_n < 1$$

Now choose a $\mathcal{O}_{r_{n+1}}$ so that

$$\overline{\mathcal{O}_{r'_{i+1}}} \subset \mathcal{O}_{r_{n+1}} \subset \overline{\mathcal{O}_{r_{n+1}}} \subset \mathcal{O}_{r'_i}$$

In this way, we get a family $\{\mathcal{O}_r\}_{r\in\mathbb{Q}\cap(0,1)}$ with the property (*) above.

Let now

$$f_r(x) = \begin{cases} r \text{ if } x \in \mathcal{O}_r \\ 0 \ \textit{otherwise} \end{cases} ; \ f = \sup_r f_r; \quad g_s(x) = \begin{cases} 1 \text{ if } x \in \overline{\mathcal{O}_s} \\ s \ \textit{otherwise} \end{cases} ; \ \ g = \inf_s g_s \tag{(**)}$$

□

f is lower semicontinuous, g is upper semicontinuous, $f(X) \subset [0,1]$, $f(C) = \{1\}$, $f(\overline{\mathcal{O}_0}) = \{0\}$.

We show that $f = g$, which implies continuity. Note that $f_r(x) > g_s(x)$ only if $r > s$, $x \in \mathcal{O}_r$ and $x \notin \overline{\mathcal{O}_s}$. But then $\mathcal{O}_r \subset \overline{\mathcal{O}_s}$ which is impossible. This proves $f \leqslant g$.

Assume $f(x) < g(x)$ for some $x \in [0,1]$. Then $f(x) < r < s < g(x)$ for some rationals r, s. Since $f(x) < r$ we have that $f_r(x) = 0$ implying $x \notin \mathcal{O}_r$. Similarly, since $g(x) > s$ we must have $x \in \overline{\mathcal{O}_s}$. This contradicts (*).

17.2 Locally compact Hausdorff spaces

Definition 216. *A Hausdorff space $\mathbb{X}$ is locally compact (LCH) if every point has a compact neighborhood.*

In the following, $\mathbb{X}$ will be a locally compact space (LCH). A set is said to be precompact if its closure is compact.

Lemma 217. *$E \subset \mathbb{X}$ is closed* **iff** *$E \cap K$ is closed for any compact K.*

Proof. Exercise. □

Proposition 218. *For any x and any open set $\mathcal{O}$ containing x there is a precompact open set $\mathcal{O}' \ni x$ with $\overline{\mathcal{O}'} \subset \mathcal{O}$.*

Proof. Let $\mathcal{O}''$ be any precompact neighborhood of x. We can replace $\mathcal{O}$ with $\mathcal{O} \cap \mathcal{O}''$; thus, wlog, we assume $\mathcal{O}$ is precompact. Then $\partial\mathcal{O}$ and x are closed and Note 214 completes the proof. □

Proposition 219. *Let K be compact and $\mathcal{O} \supset K$ open in $\mathbb{X}$. Then there exists a precompact $\mathcal{O}'$ s.t. $K \subset \mathcal{O}' \subset \overline{\mathcal{O}'} \subset \mathcal{O}$.*

Proof. By Proposition 218 K can be covered with precompact open sets $\{\mathcal{O}_\alpha\}$ with closure in $\mathcal{O}$ and thus by a finite subset of them $\{\mathcal{O}_i\}_{i \leqslant n}$. □

Theorem 220 (Urysohn's Lemma, LCH version)**.** *Let $K \subset \mathcal{O}$ as in Proposition 219. Then there is an $f \in C([0,1], \mathbb{X})$ and a precompact $\mathcal{O}', \overline{\mathcal{O}'} \subset \mathcal{O}$ s.t. $f(K) = \{1\}$ and $f(\overline{\mathcal{O}'}^c) = \{0\}$.*

Proof. Straightforward application of Urysohn and of the previous results. □

Also with a similar proof we have the following

Theorem 221 (Tietze Extension Theorem)**.** *Let K be compact and $f \in C(K)$. Then there exists $g \in C(\mathbb{X})$ s.t. $g|_K = f$.*

Definition 222. *A space is $\sigma-$compact if it is the countable union of compact sets.*

Proposition 223. *If $\mathbb{X}$ is second countable, then $\mathbb{X}$ is $\sigma-$compact.*

Proof. Let $\mathcal{T} = \{\mathcal{O}_i\}_{i \in \mathbb{N}}$ be a countable base. Each $x \in \mathbb{X}$ has, by assumption, a precompact neighborhood $\mathcal{O}'_x$. Since $\mathcal{T}$ is a base, there is an $i(x)$ and an $\mathcal{O}_{i(x)} \subset \mathcal{O}'_x$ s.t. $x \in \mathcal{O}_{i(x)}$. Then, $\overline{\mathcal{O}_{i(x)}} \subset \overline{\mathcal{O}'_x}$ is compact and $\mathbb{X} = \bigcup_{i(x), x \in \mathbb{X}} \overline{\mathcal{O}_{i(x)}}$, a countable union since it is a subfamily of $\mathcal{T}$. □

Proposition 224. *If $\mathbb{X}$ is $\sigma-$ compact, then there is a countable family of precompact open sets $\{\mathcal{O}_n\}_{n \in \mathbb{N}}$ such that $\overline{\mathcal{O}}_n \subset \mathcal{O}_{n+1}$ for all n and $\mathbb{X} = \cup_{n \in \mathbb{N}} \mathcal{O}_n$.*

Proof. Let $\mathcal{O}_n$ be as in Proposition 223 above; then $\mathcal{O}'_n = \cup_1^n \mathcal{O}_j$ is such a family. □

17.3 Support of a function

Definition 225. If f is a complex-valued function on X, then the support of f is defined as $\text{supp}(f) = \overline{\{x : f(x) \neq 0\}}$.

We say f is supported in $\mathcal{O}$ if $\text{supp}(f) \subset \mathcal{O}$, and we write $f \prec \mathcal{O}$. If $f \in C(X, [0,1])$, C is closed and $f(C) = \{1\}$, then we write $C \prec f$.

Here $\mathbb{X}$ is always an LCH space, $C, K, \mathcal{O}$ are a closed, compact and open resp. sets in $\mathbb{X}$.

17.4 Partitions of unity

Definition 226.

1. A partition of unity on a set E in a topological space is a collection of continuous functions $\{\rho_\alpha\}_{\alpha\in A}$ with values in $[0,1]$ with the property that
   - for any x there is a neighborhood of x where only finitely many ρ_α are nonzero.
   - $\sum_\alpha \rho_\alpha = 1$ on E.
2. A partition is subordinate to an open cover $\mathcal{O}_\alpha$ if $\forall\alpha\ \rho_\alpha \prec \mathcal{O}_\alpha$.

Partitions of unity have many uses in mathematics. An interesting application is in defining integrals on manifolds (with respect to some form). One relies on coordinates to define the integral on a coordinate patch and then uses a partition of unity subordinate to the coordinate patch covering to extend the integral to the whole manifold.

Theorem 227. *Let $K \subset \mathbb{X}$ be compact. For any open cover $\{\mathcal{O}_j\}_{j\leqslant n}$ of K there exists a partition of unity on K, $\{\rho_j\}_{j\leqslant n}$ with $\rho_j \prec \mathcal{O}_j, j \leqslant n$.*

Proof. (adapted from Rudin) For each $x \in K$, x is in some $\mathcal{O}_j$, and there is an $\mathcal{O}_x \subset \mathcal{O}_j$ precompact containing x. By compactness $\exists\{x_1,...,x_n\}$ s.t $K \subset \cup_k \mathcal{O}_{x_k}$. For each j let $\mathcal{O}'_j = \cup\{\mathcal{O}_{x_k} : \mathcal{O}_{x_k} \subset O_j\}$. By Urysohn's lemma, define for each j a continuous function g_j s.t. $\overline{\mathcal{O}'_j} \prec g_j \prec \mathcal{O}_j$. Let

$$\rho_1 = g_1; \quad \rho_2 = g_2(1-g_1); \ \ ...; \quad \rho_n = g_n(1-g_{n-1})\cdots(1-g_1)$$

Clearly $\rho_j \prec \mathcal{O}_j$. By induction we check that

$$\rho_1 + \cdots + \rho_n = 1 - (1-g_1)\cdots(1-g_n)$$

Now, for $x \in K$ at least one g_i is 1, and thus the sum above is 1 on K. □

17.5 Continuous functions

Let $\mathbb{X}$ be CH.

Definition 228. • The topology of uniform convergence on the functions from $\mathbb{X}$ to $\mathbb{R}$ or $\mathbb{C}$ is given by $f_n \to f$ if $\|f_n - f\|_u \to 0$ where $\|\cdot\|_u \equiv \|\cdot\|_\infty$ is the usual sup norm on $\mathbb{X}$.
- $BC(\mathbb{X})$ is the space of bounded continuous functions on $\mathbb{X}$.
- The space $C_c(\mathbb{X})$ of functions with compact support is $\{f \in C(\mathbb{X}) : \text{supp}(f) \text{ is compact}\}$.
- $C_0(\mathbb{X})$ is the space of continuous functions vanishing at infinity:

$$C_0(\mathbb{X}) = \{f \in C(\mathbb{X}) : \forall n \in \mathbb{N}, |f|^{-1}([n^{-1},\infty)) \text{ is compact}\}$$

- $f_n \to f$ **uniformly on compact sets** if $\|f_n - f\|_K = \sup_{x\in K}\|f_n - f\| \to 0$ for all K compact.

Note 229. $C_c(\mathbb{X}) \subset C_0(\mathbb{X})$ and $C_0(\mathbb{X}) \subset BC(\mathbb{X})$. The space $C(\mathbb{X})$ is closed in the space of real or complex functions on $\mathbb{X}$.

Definition 230. The one-point (or Alexandroff) compactification of $\mathbb{X}$ is defined as $\mathbb{X}^* = \mathbb{X} \cup \{\infty\}$ where $\infty \notin \mathbb{X}$. The topology on $\mathbb{X}^*$ is given as follows: $\mathcal{O}$ is open in $\mathbb{X}^*$ if it is open in $\mathbb{X}$ or if $\mathcal{O} \not\subset \mathbb{X}$ and $\mathcal{O}^c$ is compact in $\mathbb{X}$.

Proposition 231. *$\mathbb{X}^*$ is CH. The inclusion $X \to X^*$ is an embedding. f is continuous on $\mathbb{X}^*$ iff $\exists c \in \mathbb{C}$ $(f-c)|_{\mathbb{X}} \in C_0(\mathbb{X})$ and $f(\infty) = c$.*

Proof. Exercise. □

Note 232. We can identify the continuous functions on an LCH that vanish at infinity with the continuous functions on a CH that vanish at a point.

Sup-norm convergence on $\mathbb{X}$ is stronger than uniform convergence on compact sets. The closure of $C_c(\mathbb{X})$ w.r.t the sup norm on on $\mathbb{X}$ is $C_0(\mathbb{X})$.

17.6 The Stone-Weierstrass theorem

This is the sweeping generalization of the theorem of approximation by continuous functions by polynomials. Now $\mathbb{X}$ will be a compact space, and $C(\mathbb{X})$ is the space of continuous functions with the sup norm.

Algebras. Let K be a field, and let A be a vector space over K equipped with an additional binary operation "$\cdot$", called multiplication from $A \times A$ to A. Then A is an algebra over K if the following identities hold for all elements x, y, and z of A, and all elements (often called scalars) a and b of K :

1. Right distributivity: $(x + y) \cdot z = x \cdot z + y \cdot z$
2. Left distributivity: $x \cdot (y + z) = x \cdot y + x \cdot z$
3. Compatibility with scalars: $(ax) \cdot (by) = (ab)(x \cdot y)$

In the following, we will work with algebras in $C(\mathbb{X}, \mathbb{R})$ or $C(\mathbb{X}, \mathbb{C})$, where $\cdot$ is usual multiplication. These two algebras are clearly **associative** and **commutative (abelian)**, and they are closed in the sup norm, or in the norm of uniform convergence on compact sets.

Lemma 233. *Let $\mathbb{X} = \{0, 1\}$. The only subalgebras of $C(\{0,1\}, \mathbb{R})$ are $C(\{0,1\}, \mathbb{R})$, $\{0\}$ and the one-dimensional ones $\{f : f(0) = 0\}$, $\{f : f(1) = 0\}$, $\{f : f = const.\}$.*

Note 234. *$C(\{0,1\})$ is isomorphic to $\mathbb{R}^2$ with componentwise multiplication.*

Proof. It is easy to check that the subsets mentioned are algebras. Conversely, let $\mathcal{A}$ be a subalgebra of $C(\{0,1\})$. Assume there is an $f \in \mathcal{A}$ s.t. $f(0)f(1) \neq 0$ and $f(0) \neq f(1)$. Then, as you can check by taking the determinant, f^2 is linearly independent from f, and, by Note 234, $\mathcal{A} = C(\{0,1\})$. If $f(0) = f(1) \neq 0$ for all $f \in \mathcal{A}$, then $\mathcal{A}$ is the algebra of constants. If $f(0) = 0$ or $f(1) = 0$ but not both, then $\mathcal{A}$ is $\{f : f(0) = 0\}$ or $\{f : f(1) = 0\}$. If $\forall f \in \mathcal{A}$ $f(0) = f(1) = 0$, then clearly $\mathcal{A} = \{0\}$. □

Definition 235. *$\mathcal{A} \subset C(\mathbb{X})$ is called a lattice if $f, g \in \mathcal{A}$ implies $f \wedge g$ and $f \vee g$ are also in $\mathcal{A}$.*

Note 236. *If $\mathcal{A}$ is a linear subspace of $C(\mathbb{X})$, then it is a lattice if $f \in \mathcal{A} \Rightarrow |f| \in \mathcal{A}$.*

Definition 237. *A subset $\mathcal{A}$ of $C(\mathbb{X}, \mathbb{R})$ is said to separate points if $x \neq y \in \mathbb{X} \Rightarrow \exists f \in \mathcal{A}, f(x) \neq f(y)$.*

Theorem 238 (The Stone-Weierstrass theorem). *Let $\mathbb{X}$ be a CH space and $\mathcal{A} \subset C(\mathbb{X}, \mathbb{R})$ a closed subalgebra that separates points. If $1 \in \mathcal{A}$, then $\mathcal{A} = C(\mathbb{X}, \mathbb{R})$; otherwise, there is an $x_0 \in \mathbb{X}$ s.t. $\mathcal{A} = \{f \in C(\mathbb{X}, \mathbb{R}) : f(x_0) = 0\}$.*

Lemma 239. *1. In $C(\mathbb{X}, \mathbb{R})$, $x \mapsto |x|$ is in the closure of polynomials that vanish at zero, in the sup norm on compact sets.*

If $\mathcal{A}$ is a closed subalgebra of $C(\mathbb{X}, \mathbb{R})$, then $\mathcal{A}$ is a lattice.

Proof. 1. For $|t| < 1/2$, by the Taylor series with remainder theorem, the Maclaurin series of $g = t \mapsto \sqrt{1-t}$,

$$S(t) = 1 - \sum_{n \geqslant 1} c_n t^n, \quad c_n = \frac{\frac{1}{2}(1-\frac{1}{2})(2-\frac{1}{2})\cdots(n-1-\frac{1}{2})}{n!} = \frac{\frac{1}{2}\Gamma(n-\frac{1}{2})}{n!\Gamma(\frac{1}{2})} > 0$$

converges to g. Using Stirling's formula we see that $2\sqrt{\pi}c_n = n^{-3/2}(1 + o(1))$ for large n. The Weierstrass M test shows that S converges uniformly to a continuous function f on $[-1, 1]$ and since $f - g = 0$ on $[-\frac{1}{2}, \frac{1}{2}]$, we have $f(t) = \sqrt{1-t}$ on $[-1, 1]$. Note that $P_n(x) = 1 - \sum_{k=1}^{n} c_k (1 - x^2)^k - \sum_{n+1}^{\infty} c_k$ is a sequence of polynomials with $P_n(0) = 0$, converging uniformly to $|x|$ on $[-1, 1]$. If $a \neq 0$, then $aP_n(x/a)$ converge to $|x|$ uniformly on $[-a, a]$.

2. If $f \in \mathcal{A}$ and $\|f\| = a \neq 0$, then $\||f| - aP_n(a^{-1}f)\|_u \to 0$ as $n \to \infty$. □

Lemma 240. *Let $\mathcal{A}$ be a lattice in $C(\mathbb{X}, \mathbb{R})$ and $f \in C(\mathbb{X}, \mathbb{R})$. If for any couple of points $\{x, y\}$ there is a $g \in \mathcal{A}$ s.t. $f = g$ on $\{x, y\}$, then $f \in \mathcal{A}$.*

Proof. Using the stated property and compactenss, for each $\epsilon > 0$ we construct a $g \in \mathcal{A}$ s.t. $\|f - g\|_u < \epsilon$ as follows. For $x, y \in \mathbb{X}$ we let $g_{xy} \in \mathcal{A}$ be the function that coincides with f on $\{x, y\}$, and define the open sets $U_{xy} = \{z \in \mathbb{X} : f(z) < g_{xy}(z) + \epsilon\}$ and $L_{xy} = \{z \in \mathbb{X} : f(z) > g_{xy}(z) - \epsilon\}$. For fixed y, $\{U_{xy}, x \in \mathbb{X}\}$ cover $\mathbb{X}$ (since $x \in U_{xy}$) and thus, by compactness, $\mathbb{X} = \cup_{j \in \mathbb{N}} U_{x_j, y}$ for some finite set $\{x_1, ..., x_n\}$. With $\Gamma_y = \vee_1^n g_{x_j y}$, we have $f < \Gamma_y + \epsilon$ on $\mathbb{X}$ and $f > \Gamma_y - \epsilon$ on $\cap_{j=1}^n L_{x_j y}$ which is an open set containing y. Now, $\{\cap_{j=1}^n L_{x_j y}, y \in \mathbb{X}\}$ cover $\mathbb{X}$, and thus $\mathbb{X} = \cup_{k=1}^m \cap_{j=1}^n L_{x_j y_k}$ for some finite set $\{y_1, ..., y_m\}$. Then $g = \wedge_1^m \Gamma_{y_k}$ has the property $\|f - g\|_u < \epsilon$ completing the proof. □

Proof of Theorem 242. Clearly, for any $x, y \in \mathbb{X}$, the restriction $\mathcal{A}_{xy} = \{g \text{ restricted to } \{x, y\} : g \in \mathcal{A}\}$ is also an algebra, a subalgebra of $C(\{x, y\}, \mathbb{R})$. If for any $\{a, c, x, y\} \in \mathbb{X} \times \mathbb{R}$ there is a $g \in \mathcal{A}$ s.t. $g(x) = a, g(y) = b$, then , by Lemma 240, $\mathcal{A} = C(\mathbb{X}, \mathbb{R})$. Otherwise, there is a pair $\{x, y\}$ s.t. $\mathcal{A}_{xy}$ is a proper subalgebra of $C(\{x, y\}, \mathbb{R})$. Since $\mathcal{A}$ separates points, there are only two possibilities $\mathcal{A} = \{f : f(x) = 0\}$ or $\mathcal{A} = \{f : f(y) = 0\}$. Neither of these cases is possible if $1 \in \mathcal{A}$. □

Corollary 241. *Polynomials are dense in $\mathbb{R}^n$.*

The complex-valued version of Stone-Weierstrass needs stronger conditions. Clearly, $\mathcal{E}_+ = \{e^{2\pi i k x} : k \in \mathbb{N}\}$ is a family in $C([0, 1], \mathbb{C})$ that separates points. Let $\mathcal{E}_- = \{e^{2\pi i m x} : 0 \leqslant m \in \mathbb{Z}\}$. Note that $\int_0^1 e_k e_m dx = 0$ for any $e_k \in \mathcal{E}_+, e_m \in \mathcal{E}_-$, see §1. Since convergence in $\|\|_u$ implies convergence in $\|\|_2$ (why?), the algebra $\mathcal{A}$ generated by $\mathcal{E}_+$ is orthogonal to $\mathcal{E}_-$, and in particular cannot be dense in $C([0, 1], \mathbb{C})$. In fact, the elements of $\mathcal{A}$ can be identified with the boundary values on S^1 of the functions analytic in $\mathbb{D}$, vanishing at zero, and continuous up to the boundary.

However, we know already (cf. Theorem 10) that the algebra generated by $\mathcal{E}_+ \cup \mathcal{E}_-$ is $C(\mathbb{T})$, so what we have to do (at least in this example) is simply require that $f \in \mathcal{A} \Rightarrow \overline{f} \in \mathcal{A}$. Indeed, this is sufficient in general:

Theorem 242 (Complex Stone-Weierstrass theorem). *Let $\mathbb{X}$ be a CH space and $\mathcal{A} \subset C(\mathbb{X}, \mathbb{C})$ a closed subalgebra that separates points and is closed under complex conjugation. If $1 \in \mathcal{A}$, then $\mathcal{A} = C(\mathbb{X}, \mathbb{C})$; otherwise, there is an $x_0 \in \mathbb{X}$ s.t. $\mathcal{A} = \{f \in C(\mathbb{X}, \mathbb{C}) : f(x_0) = 0\}$.*

Proof. Note that $f \in \mathcal{A}$ implies $\Re f$ and $\Im f$ are in $\mathcal{A}$. By Theorem 242 $u(x) + iv(x) \in \mathcal{A}$ for any u, v in $C(\mathbb{X}, \mathbb{R})$. □

Note 243. *By Urysohn's lemma, in any normal space, continuous functions separate points.*

Exercise 53.

1. Use Stone-Weierstrass to show that $\{e^{2\pi ikx} : k \in \mathbb{Z}\}$ form a complete orthonormal set in $L^2([0, 1])$ (meaning that any $f \in L^2([0, 1])$ is an L^2 limit of trig polynomials).

2. Assume $f \in C([0, 1])$ is s.t. $\forall 0 \leqslant n \in \mathbb{Z}$, $\int_0^1 s^n f(s) ds = 0$. Show that $f = 0$.

3. **(The moment problem)** The moments of a Borel measure μ are defined as $\mu_n = \int_0^1 s^n d\mu, 0 \leqslant n \in \mathbb{Z}$, provided the integrals exist. The measure μ is determinate if the moments $\{\mu_n, n \geqslant 0\}$ are unique to μ. Show that compactly supported measures, say on $[0, 1]$, are determinate.

4. Let $\mathbb{X}_1, \mathbb{X}_2$ be compact metric spaces. Show that the algebra generated by continuous functions of one variable is dense in $C(\mathbb{X}_1 \times \mathbb{X}_2, \mathbb{R})$: more precisely the family
$$\left\{ (x, y) \mapsto \sum_{j=1}^{n} f_j(x) g_j(y) : n \in \mathbb{N}, f_j \in C(\mathbb{X}_1), g_j \in C(\mathbb{X}_2), 1 \leqslant j \leqslant n \right\}$$
is dense in $C(K_1 \times K_2)$.

5. If $\mathbb{X}$ is a compact metric space (thus separable) with metric ρ, then $C(\mathbb{X})$ is separable. (Hint: if $\{x_m, m \in \mathbb{N}\}$ is a dense set in $\mathbb{X}$, then $F_{mn} = \wedge\{n^{-1}, \rho(x, x_m)\}$, $(m, n \in \mathbb{N}^2)$, is a family of continuous functions that separates points.)

18 Sequences and nets

A sequence in a topological space X is a function whose domain is an interval of integers with values in X.

Definition 244. Let X be a topological space.

1. $\mathcal{O} \subset X$ is sequentially open if each sequence $(x_n)_{n \in \mathbb{N}}$ in X converging to a point of $\mathcal{O}$ is eventually in $\mathcal{O}$ (i.e. there exists N s.t. $\forall n \geqslant N$ $x_n \in \mathcal{O}$).

2. $C \subset X$ is sequentially closed if the limit of any convergent sequence $(x_n)_{n \in \mathbb{N}}$ in C is also in C.

3. A space X is **sequential** if every sequentially open subset of X is open, or equivalently, every sequentially closed subset of X is closed.

Exercise 54.

1. Every first-countable space is sequential. In particular, second countable, metric, and discrete spaces are sequential.

2. The cocountable topology on X consists of $\emptyset$ and all cocountable subsets of X, that is all sets whose complement is countable. Check that the only closed sets are X and its countable subsets. Show that if X is uncountable, the cocountable topology on X is not sequential.

Definition 245. 1. A directed set is a nonempty set A together with a reflexive and transitive binary relation $\preccurlyeq$ s.t. every pair of elements has an upper bound, i.e. $\forall a, b \in A\ \exists c \in A$ s.t. $a \preccurlyeq c$ and $b \preccurlyeq c$.

2. Let A be a directed set and X be a topological space with topology $\mathcal{T}$. A function $f : A \to X$ is called a net. We write $f = (x_\alpha)_{\alpha\in A}$.

3. We say that (x_α) is eventually in $Y \subset X$ if $\exists\alpha \in A$ s.t. $\forall\beta \in A$, $\beta \succcurlyeq \alpha \Rightarrow x_\beta \in Y$.

4. (x_α) is said to converge to x if for every neighborhood $\mathcal{O}$ of x, (x_α) is eventually in $\mathcal{O}$.

Exercise 55. Show that the neighborhood system of a point x in a topological space with $\subset$ for $\preccurlyeq$ is a directed system.

Definition 246. 1. Let $E \subset X$. The net (x_α) is frequently in E if $\forall\alpha \in A\ \exists\beta \succcurlyeq \alpha$ in A s.t. $x_\beta \in E$.

2. A point $x \in X$ is an accumulation point or cluster point of a net if for every neighborhood $\mathcal{O}$ of x, the net is frequently in $\mathcal{O}$.

18.1 Subnets

Definition 247. A function $h : B \to A$ is **monotone** if $\beta_1 \preccurlyeq \beta_2 \Rightarrow h(\beta_1) \preccurlyeq h(\beta_2)$. B is a **cofinal** subset of A if for every $\alpha \in A$ there is a $\beta \in B$ s.t. $\beta \succcurlyeq \alpha$. The function h is final if $h(B)$ is a cofinal subset of A.

If A and $B \subset A$ are directed sets and $(x_\alpha)_{\alpha\in A}$ and $(y_\beta)_{\beta\in B}$ are nets in A and B resp., then $(y_\beta)_{\beta\in B}$ is a **subnet** of $(x_\alpha)_{\alpha\in A}$ if there is a monotone final function h s.t. for all $\beta \in B$, $y_\beta = x_{h(\beta)}$.

Note 248. A subnet of a sequence is not necessarily a subsequence! See Ex. 57 below.

Exercise 56. Show that

1. A function f between two topological spaces is continuous at x iff for any net (x_α) converging to x we have $\lim_{\alpha\in A} f(x_\alpha) = f(x)$.

2. A net has a limit if and only if all of its subnets have limits. In a Hausdorff space, the limit of a net is unique, and every subnet converges to this limit.

3. A space X is compact if and only if every net $(x_\alpha) \in X$ has a subnet with a limit in X.

4. A net in the product space has a limit if and only if each projection has a limit.

5. A point x in X is a cluster point of a net if and only if there is a subnet which converges to x.

6. limsup and liminf along a net are defined in complete analogy with their counterpart on sequences. Show that $\limsup(x_\alpha + y_\alpha) \leqslant \limsup x_\alpha + \limsup y_\alpha$.

19 Tychonoff's theorem

If $\{X_i\}_{i\in I}$ are topological spaces, then the product space is defined as $X = \prod_{i\in I} X_i = \left\{ f : I \to \bigcup_{i\in I} X_i \;\middle|\; (\forall i)(f(i) \in X_i) \right\}$. The fact that X is nonempty for general nonempty X_i is equivalent to the axiom of choice, AC.

The product topology is defined to be the coarsest topology (i.e. the topology with the fewest open sets) for which all the projections π_i ($\pi_i(x) = x_i$) are continuous. That is, the sets $\bigcap_{j=1}^{n} \pi_{i_j}^{-1}(N_{i_j})$ where N_i are open sets in X_i form a base for the topology on X. In this topology a net $(f_\alpha)_{\alpha\in A}$ converges iff $\forall i \in I, f_\alpha(i)$ converges, that is, the topology is that of pointwise convergence of functions.

Definition 249. 1. A basic neighborhood N of $f \in X$ is determined by a finite subset F of I together with all the neighborhoods $\mathcal{O}_j$ of $f(j) =: f_j$ in X_j, $j \in F$. N consists of all $h \in X$ s.t. $\forall j \in F, h(j) \in \mathcal{O}_j$. We say that N is supported by F, $N = N(\{\mathcal{O}_j : j \in F\})$. Note that basic neighborhoods generate the topology on X.

2. A partially defined member of X is a function g defined on some $J \subset I$, i.e. $g \in \prod_{j\in J} X_j$.

3. If $(x_\alpha)_{\alpha\in A}$ is a net in X, partial cluster point z is a partially defined member of X with domain $J \subset I$ s.t. z is a cluster point of $(x_\alpha|_J)$.

Theorem 250 (Tychonoff). [11] *Assume $\{X_i\}_{i\in I}$ are compact for all $i \in I$. Then $X = \prod_{i\in I} X_i$ is compact in the product topology.*

Proof 1, based on nets, adaptation of Chernoff, (1992) .

We may assume that the spaces X_i are nonempty. Using Zorn's lemma, given a net (x_α) we show that there is a cluster point z of (x_α) with domain I.

Let $\mathcal{P}$ be the set of all partial cluster points of $(x_\alpha)_{\alpha\in A}$. Since by assumption $(x_\alpha)|_{X_1}$ has a cluster point, $\mathcal{P}$ is nonempty. Order $\mathcal{P}$ by function extension. A function being a set of pairs, this is the same as inclusion. That is, $g_1 \subset g_2$ if the domain of g_1 is contained in the domain of g_2 and $g_2 = g_1$ on the domain of g_1.

[11] Adaptation of a Bourbaki proof, see also Loomis, see p.11

Let $\mathcal{L} = \{z_\lambda : \lambda \in \Lambda\}$ be a chain in $\mathcal{P}$, and let $z_0 = \cup_{\lambda\in\Lambda} z_\lambda$. Since any two members of $\mathcal{L}$ agree on their common domain, z_0 is a partially defined member of X. Moreover, z_0 is a partial cluster point of (x_α) as well. Indeed, every basic neighborhood of z_0 has finite support and thus F is contained in the domain of z_λ for some λ, where by definition $z_0 = z_\lambda$ (and z_λ is a partial cluster point). Note now that $\mathcal{L}$ has an upper bound, z_0.

Let z be a maximal element of $\mathcal{P}$. If the domain of z is I, the conclusion follows. Assuming it is not, let $k \in I \setminus J$. Since z must be a cluster point of $(x_\alpha|_J)_{\alpha\in A}$, there is a subnet (x_β) s.t. $(x_\beta|_J)$ converges to z. Now, since X_k is nonempty and compact, the net $(x_\beta|_{X_k})$ has a cluster point $x \in X_k$. Define the function h on $J \cup \{k\}$ by $h = g$ on J and $h|_{X_k} = x$. Then h is a partial cluster point of (x_α), and thus $h \in \mathcal{P}$ extends g strictly.

□

Proof 2, similar to Folland's. Let $\mathcal{F}$ be a family of closed sets in X with the finite intersection property (f.i.p.). We want to show $\bigcap_{F\in\mathcal{F}} F \neq \emptyset$. Clearly this is the case if the same holds for any larger family $\mathcal{F}'$. A subtle point in the proof is to take the largest such set. Note that any chain of families (not necessarily of closed sets) with the f.i.p. $\mathcal{F}_\alpha \subset \mathcal{F}_{\alpha'} \subset \cdots$ has an upper bound, with the f.i.p, namely their union. By Zorn's lemma, there is a maximal family with the f.i.p., $\mathcal{M} \supset \mathcal{F}$. In the following "construct", "choose" etc. are just ways of speaking, as we rely on the AC.

We now construct a point in X which should be in all $F \in \mathcal{F}$ (and, in fact, all $M \in \mathcal{M}$). For any i, $\{\overline{\pi_i(M)} | M \in \mathcal{M}\}$ is a family of closed sets in X_i with the f.i.p. Then, for each i there is an $m_i \in \bigcap_{M\in\mathcal{M}} \overline{\pi_i(M)}$. Choose an m_i for each i and let $m = (m_i)_{i\in I}$.

If we show that $\bigcap_{j=1}^{n} \pi_{i_j}^{-1}(O_{i_j})$ (O_i open nbd of m_i) intersect nontrivially each F, this will imply that $m \in F$ for all our F. This is because each F is closed and for each F it follows that any open nbd of m intersects nontrivially F, implying, by elementary topology, $m \in F$.

The property above is implied by the following: for any O_i as above, $\pi_i^{-1}(O_i) \in \mathcal{M}$.

Now, for any $M \in \mathcal{M}$ we have, by construction, $\overline{\pi_i(M)} \cap O_i \neq \emptyset$. Thus $\overline{\pi_i(M) \cap O_i} \neq \emptyset$ implying $\pi_i(M) \cap O_i \neq \emptyset$ which in turn means $M \cap \pi_i^{-1}(O_i) \neq \emptyset$. Then, adjoining any single set $\pi_i^{-1}(O_i)$ to $\mathcal{M}$, the f.i.p. is preserved. But, then by the maximality of $\mathcal{M}$, $\pi_i^{-1}(O_i) \in \mathcal{M}$, and this holds for any i ending the proof . □

19.1 Further discussions

Assume $\{X_i\}_{i\in\mathbb{N}}$ are second countable. Check that the product space X is also second countable. (The proof, in general, relies on a weak version of the AC (to select a countable open base for each X_i), but does not require it if the X_i are linearly ordered which is the case if $X_i = \mathbb{R}$, or, after simple modifications, $X_i = \mathbb{R}^n$ or $\mathbb{C}$.)

Recall that a sequence is a function $f : \mathbb{N} \to X$ and a subsequence $f \circ g$ is defined by a $g : \mathbb{N} \to \mathbb{N}$ s.t. $\lim_{n\to\infty} g(n) = \infty$.

Without further use of the AC we can prove compactness of the product X of compact, second countable spaces.

Proof. Let $f : \mathbb{N} \to X$ be a sequence. Since X_1 is compact, there is a subsequence defined by an f_1 s.t. $(f \circ f_1)_1 : \mathbb{N} \to X_1$ is convergent. Inductively, there is a subsequence defined by an f_n s.t.

all $(f \circ f_n)_i, i = 1, 2, ...n$ are convergent. Define g by $g(k) = f_k(k)$. Then, as you can easily check, $(f \circ g)_i, i \in \mathbb{N}$ are all convergent, implying that $(f \circ g) : \mathbb{N} \to X$ is convergent. $\square$

Exercise 57. Check that the space $[0,1]^{\mathbb{R}}$ is compact. Show that there is a directed set A and a net $x : A \to \mathbb{N}$, which is a subnet of $1, 2, ...$ along which **any** sequence $a_1, ..., a_n, ...$ in $[0,1]$ converges.

20 Arzelá-Ascoli's theorem

Let X be a separable metric space with metric d; let E be a countable dense subset of X. Let $\{f_n\}_{n\in\mathbb{N}} : X \to \mathbb{C}$ be a sequence of equicontinuous and pointwise equibounded functions. A pointwise equibounded family $\mathcal{F}$ is one s.t. $\forall x \in X \sup_{F\in\mathcal{F}} |F(x)| = M(x) < \infty$ and equicontinuous means that $\forall \epsilon \exists \delta$ s.t. $\forall x, y \in X$ and $F \in \mathcal{F}$, $d(x,y) < \delta \Rightarrow |F(x) - F(y)| < \epsilon$. For the purpose of Arzelá-Ascoli's theorem below, equicontinuity can be replaced with the weaker condition $\forall e \in E$ there is an r s.t. for all y with $d(y,e) < r$ and all F we have $|F(e) - F(y)| < \epsilon$, which can be seen using the compact cover formulation of compactness.

Theorem 251. *Every sequence $\{F_n\}_{n\in\mathbb{N}} : X \to \mathbb{C}$ of equicontinuous and pointwise equibounded function has a subsequence which converges uniformly on compact sets.*

Note 252.

Proof. Let K be a compact set in X. Let $M(x)$ be as above. Then the space $Y = \prod_{e\in E}\{z \in \mathbb{C} : |z| \leqslant M(e)\}$ is compact.[12] This means there is a subsequence defined by a $g : \mathbb{N} \to \mathbb{N}$ s.t. $\{F_{g(n)}\}_{n\in\mathbb{N}} : X \to \mathbb{C}$ restricted to E converges. This subsequence converges on X. Indeed, for any $\epsilon > 0$ there is an $e \in E$ close enough to x, $d(x,e) < \delta$, s.t. $(\forall n)$ $(|F_n(x) - F_n(e)| < \epsilon/3)$. Now by the triangle inequality, $|F_n(x) - F_m(x)| < \epsilon$ for all n, m large enough.

If K is compact, there is a *finite* set $E_n = \{e_1, ...e_n\}$ s.t. for all $x \in K$, $d(x, E_n) < \delta$, δ as above. Check that this implies uniform convergence in K. $\square$

Uniform convergence implies that the limit F of the subsequence is also continuous, and in fact adjoining F to the sequence, the new sequence is also equicontinuous and pointwise equibounded.

An important example of an equibounded, equicontinuous family is the following. Consider the ball B_1 of radius one in $L^1((a,b))$ and the linear map $K : B_1 \to B_{|b-a|}$ given by $KF = \int_a^x F$. Check that $K(B_1)$ is an equibounded, equicontinuous family.

Such a linear map is called *compact operator*.

[12]By the remarks at the beginning of this section, the AC is not needed in this setting.