Class notes, 6211 and 6212

Ovidiu Costin

Contents

Part I: 6211

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 $\sum\limits_{k=1}^{n} x_k y_k$ in $\mathbb{R}^n$ and $\sum\limits_{k=1}^{n} x_k \bar{y}_k$ 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

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

where D_N is the *Dirichlet kernel*,

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

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

Remark 1.0.1. *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 1.0.2. *Note that for any $f \in C(\mathbb{T})$ and any $a \in \mathbb{R}$, we have*

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

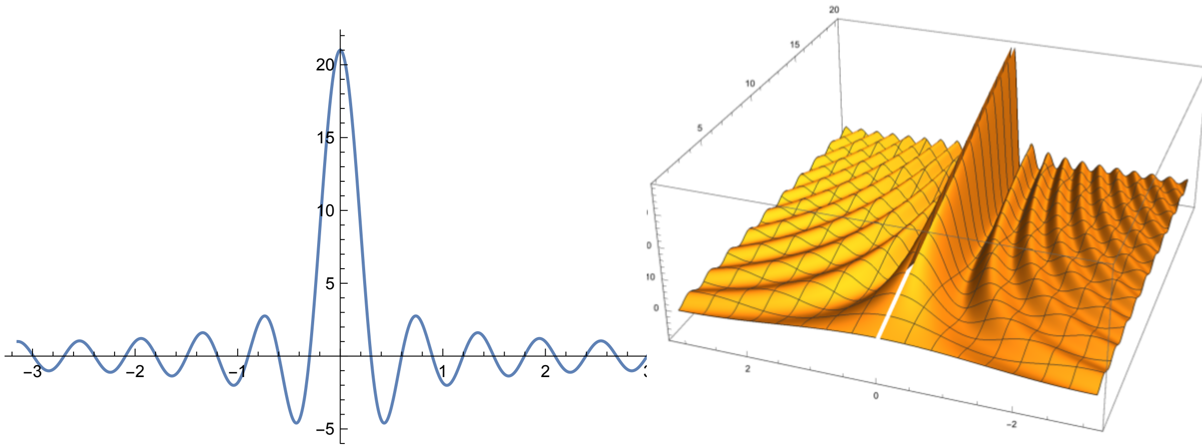

Figure 1: The Dirichlet kernel for $n = 10$, D_{10} (left) D_n on I for $n = 1, ..., 20$. The peak grows like n, with a width $1/n$ and oscillations of frequency n away from it (right)

Lemma 1.0.3. *Let $(a, b) \subset [-\frac{1}{2}, \frac{1}{2})$. Then*

$$\lim_{n\to\infty} \int_a^b D_n(x)dx = \begin{cases} 0 \text{ if } 0 \notin [a, b] \\ 1 \text{ if } 0 \in (a, b) \\ \frac{1}{2} \text{ if } 0 \in \{a, b\} \end{cases} \tag{3}$$

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

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} D_n(s)ds = 1; \quad \text{and, since } D_n \text{ is even, } \int_0^{\frac{1}{2}} D_n(s)ds = \tfrac{1}{2} \tag{4}$$

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

$$\begin{aligned} &\left| \int_a^b \frac{\sin((2n+1)\pi s)}{\sin \pi s} ds \right| \\ &\quad = \left| -\frac{\cos(\pi b(2n+1))}{\sin(\pi b)\pi(2n+1)} + \frac{\cos(\pi a(2n+1))}{\sin(\pi a)\pi(2n+1)} - \int_a^b \frac{\cos(\pi s)\cos((2n+1)\pi s)}{\sin^2(\pi s)(2n+1)} ds \right| \\ &\quad \leqslant \frac{4}{\pi(2n+1)} \left(|a|^{-1} + |b|^{-1} \right) \end{aligned} \tag{5}$$

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 (4), the result follows. $\square$

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 1.0.3 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

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

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 1.0.4. *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 1.0.5. *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 1.0.3

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

$$= \frac{1}{2}\left(f(\xi^+) + f(\xi^-) - 2L\right) + \int_{-\frac{1}{2}}^{0} f'(\xi - s) \int_{-\frac{1}{2}}^{s} D_n(t) dt ds + \int_{0}^{\frac{1}{2}} f'(\xi - s) \int_{\frac{1}{2}}^{s} D_n(t) dt ds \quad (6)$$

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 $\varepsilon > 0$ we write

$$\int_{0}^{\frac{1}{2}} f'(\xi - s) \int_{\frac{1}{2}}^{s} D_n(t) dt ds = \int_{0}^{\varepsilon} f'(\xi - s) \int_{\frac{1}{2}}^{s} D_n(t) dt ds + \int_{\varepsilon}^{\frac{1}{2}} f'(\xi - s) \int_{\frac{1}{2}}^{s} D_n(t) dt ds \quad (7)$$

Now

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

by Lemma 1.0.3 and (5). On the other hand, since $\int_s^{1/2} D_n(t) dt$ is bounded by a constant c independent of n [1]

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

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

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 1.0.3.*

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

$$\left| L - \sum_{k=-n}^{n} f_k e^{2\pi i k x} \right| = O(n^{-1} \ln n) \quad (9)$$

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

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

(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)*

$$1 = x^0 \tag{10}$$

$$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{\pi k} \sin(2\pi kx) = x^1 \quad (\text{if } |x| \neq \tfrac{1}{2}) \tag{11}$$

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

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

$$\frac{3}{\pi^4}\sum_{k=1}^{\infty} \frac{(-1)^{1+k}\cos(2\pi kx)}{k^4} = -\frac{1}{2}x^2 + x^4 + \frac{7}{240} \tag{14}$$

Convergence of these series follows from Theorem 1.0.4. Note that convergence in (11) 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 (12) 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 1.0.6. *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 1.0.7. *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 1.0.8. *The f above is an example of a continuous but nowhere differentiable function. Try your hand in proving this.*

Note 1.0.9. *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

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

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 1.1.1. *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. $\displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} F_n(s)ds = 1, \forall n \in \mathbb{N}$.

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

Indeed, 1. is obvious, 2. is clear from the definition because for any $k \in \mathbb{N}$, $\displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} D_k(s)ds = 1$, 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 $f(x) = \displaystyle\int_{-\frac{1}{2}}^{\frac{1}{2}} f(x)F_n(s)ds$ 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 $\varepsilon > 0$. Using uniform continuity, choose δ so that wenever $|s - s'| \leqslant \delta$ we have $|f(s) - f(s')| < \varepsilon/2$, and choose n_0 s.t. for all $n \geqslant n_0$ we have $\int_{\frac{1}{2}\geqslant|y|>\delta} F_n(s)ds \leqslant \frac{1}{8}\varepsilon m^{-1}$. With this, we see that for all $n \geqslant n_0$ and all x

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

□

Corollary 1.1.2. *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 1.1.3. *Trigonometric polynomials are dense in $C(\mathbb{T})$.*

Proof. This follows immediately from Theorem 1.1.1: let $f \in C(\mathbb{T})$ and $\varepsilon > 0$ be arbitrary; let n_0 be s.t. $\|f - \sigma_{n_0} f\|_u \leqslant \varepsilon$; note that $\sigma_{n_0} f$ is a trig polynomial. □

Note 1.1.4. *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

$$\frac{1}{n}\sum_{k=1}^{n} f(x_j) \to \int_0^1 f(x)dx \tag{15}$$

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* frac y *denote the fractional part of y. Show that the points $\{\text{frac}\, x_j : j \in \mathbb{N}\}$ are dense in $(0, 1)$.*

Theorem 1.1.5 (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 1.1.6 (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 1.1.5. 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 $f(x) = e^{2\pi i k x}, k \in \mathbb{Z}$, and thus for all trig polynomials. Use the density of trig polynomials to complete the proof.

□

Exercise 8. *Check that (15) extends to piecewise continuous functions. Monotone bounded functions are Riemann integrable. Does (15) 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 1.2.1. *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 1.2.2. 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

[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.

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 1.2.3 (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

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

Note 1.2.4. *(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,

$$0 \leqslant \|f - \sum_{k=-n}^{m} f_k e^{2\pi ikx}\| = \langle f - \sum_{k=-n}^{m} f_k e^{2\pi ikx}, f - \sum_{k=-n}^{m} f_k e^{2\pi ikx}\rangle = \langle f, f\rangle - \sum_{k=-n}^{m} |f_k|^2 \tag{16}$$

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.

$$\sum_{k=-\infty}^{\infty} |f_k|^2 \leqslant \|f\|_2^2 \tag{17}$$

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

Proposition 1.2.5. *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 1.2.6. *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 1.2.4 a fortiori in $L^2_{\mathcal{R}}$, and the properties above trivially hold for trig polynomials. □

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

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

and

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

Definition 1.2.8. *A sequence $(s_k)_{k\in\mathbb{N}}$ in a normed space is Cauchy if for any $\varepsilon > 0$ there is an n_0 s.t. for all $n_1 > n_0$ and $n_2 > n_0$ we have*

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

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

Proposition 1.2.9. *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. $\square$

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 $\{[f] : f \text{ Riemann integrable on } [a, b]\}$, where

$$[f] = \{g : \|f - g\|_1 = 0\} \tag{18}$$

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

$$\varphi_{a,b} f = \int_a^b f(s) ds \tag{19}$$

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 $\varphi_{a,b}$ 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 2.0.1. *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 2.0.2. *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 2.0.3. *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 2.0.4. *$\mathcal{M}(\mathcal{E})$, 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 2.0.5. *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 2.1.1. *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*

$$\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j) \tag{20}$$

Definition 2.1.2. *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 2.1.3 (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 $\mu(S^1) = 1$. $\square$

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

Note 2.1.4. 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 2.2.1. *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 2.2.2. *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 $E \in \mathcal{E}$.*

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

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, $X = \prod_\alpha X_\alpha$. On X there is a naturally induced σ-algebra, namely the smallest σ-algebra that makes all canonical projections π_α measurable:

Definition 2.3.1. *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 $\bigotimes_{\alpha \in A} \mathcal{M}_\alpha$.*

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

Proposition 2.3.2. *If the index set A is countable, then $\bigotimes_{\alpha\in A} \mathcal{M}_\alpha$ 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 2.3.3. *Assume $\mathcal{M}_\alpha$ is generated by $\mathcal{E}_\alpha$. $\bigotimes_{\alpha\in A} \mathcal{M}_\alpha$ is generated by $\mathcal{F}_1 = \{\pi^{-1}(E_\alpha) : E_\alpha \in \mathcal{E}_\alpha, \alpha \in A\}$.*

Proof. By Proposition 2.2.2 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 2.3.4. *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 2.3.3 $\bigotimes_1^n \mathcal{B}_{X_j}$ 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 2.3.5. $\mathcal{B}_{\mathbb{R}^n} = \bigotimes_1^n \mathcal{B}_\mathbb{R}$.

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 2.3.6. *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

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

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

□

2.4 More about measures

Theorem 2.4.1. *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 2.4.2. *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 2.1.1 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 2.4.3. *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 $\varepsilon > 0$ the sequence $(x_n)_{n\in\mathbb{N}}$ [5] is contained in the union of the intervals $J_n = (x_n - \varepsilon 2^{-n}, x_n + \varepsilon 2^{-n})$. 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 $\mathcal{O}_\varepsilon \supset A$ and a closed set $\mathcal{C}_\varepsilon \subset A$ such that $\lambda(\mathcal{O}_\varepsilon \setminus \mathcal{C}_\varepsilon) < \varepsilon$ it is natural to try $\lambda(A) := \lim_{\varepsilon\to 0} \lambda(\mathcal{O}_\varepsilon)$. (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 3.1.1. *Let X be a set. A function $\mu^* : \mathcal{P}(X) \to [0,\infty]$ is an outer measure on X if*

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

$$\lambda^*(A) = \inf\left\{\sum_{n\in\mathbb{N}} \lambda(\mathcal{O}_i) : \mathcal{O}_i \text{ open interval}, A \subset \bigcup_{i\in\mathbb{N}} O_i\right\} \tag{21}$$

More generally, we have the following result.

Proposition 3.1.2. *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*

$$\mu^*(A) = \inf\left\{\sum_{n\in\mathbb{N}} \rho(E_i) : E_i \in \mathcal{E}, A \subset \bigcup_{i\in\mathbb{N}} E_i\right\} \tag{22}$$

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 $B \subset \cup_j E_j \Rightarrow A \subset \cup_j E_j$.

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

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

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

It follows that

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

and

$$\mu^*\left(\bigcup_{j\in\mathbb{N}} A_i\right) \leqslant \sum_{(i,j)\in\mathbb{N}^2} \rho(E_{i,j}) \leqslant \sum_{i\in\mathbb{N}} \mu^*(A_i) + \varepsilon \tag{23}$$

(Justify the use of double indices.) Since (23) holds for any positive ε, subadditivity follows. $\square$

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 $A \in \mathcal{B}_{\mathbb{R}}$

$$\lambda^*(B) = \lambda^*(B \cap A) + \lambda^*(B \cap A^c), \forall B \in \mathcal{B}_{\mathbb{R}} \tag{24}$$

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 (24) for all $B \in \mathcal{P}(\mathbb{R})$, when A is an open set.*

This suggests the following.

Definition 3.1.3. *Let μ^* be an outer measure on X. A set $A \subset X$ is called μ^*-measurable if*

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c), \forall B \in \mathcal{P}(X) \tag{25}$$

Note that, by subadditivity of outer measures (24) holds whenever the left side is no less than the right side.

Theorem 3.1.4 (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

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

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,

$$\mu^*(E \cap S) \geqslant \sum_{j\in\mathbb{N}} \mu^*(E \cap A_j) \geqslant \mu^*(\cup_j E \cap A_j) = \mu^*(E \cap S) \Rightarrow \mu^*(E \cap S) = \sum_{j\in\mathbb{N}} \mu^*(E \cap A_j) \tag{26}$$

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 (26).

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: $\mu^*(E \cap M^c) \geqslant \mu^*(E \cap N^c) = \mu^*(E)$. $\square$

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 3.2.1. *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

$$\mu^*(E) = \inf\left\{ \sum_{j\in\mathbb{N}} \mu_0(A_j) : A_j \in \mathcal{A}, E \subset \bigcup_{j\in\mathbb{N}} A_j \right\} \tag{27}$$

Note that, by monotonicity and the fact that $\mathcal{A}$ is an algebra, the unions in (27) can be assumed disjoint.

Theorem 3.2.2. *(a) Let $\mathcal{A}$ be an algebra on X and μ_0 a premeasure on $\mathcal{A}$. Then $\mu := \mu^*$ defined by (27) 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 3.2.3. *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 (27), 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 $\varepsilon > 0$. There is a disjoint sequence $(A_j)_{j \in \mathbb{N}}$ in $\mathcal{A}$ s.t. $E \subset \cup_j A_j$ and $\mu^*(E) + \varepsilon \geqslant \sum_j \mu_0(A_j)$. Thus,

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

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

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

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 (29). For $A \in \mathcal{A}$ define

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

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

$$F(x) = \begin{cases} \mu((0, x] \text{ if } x > 0 \\ 0 \text{ if } x = 0 \\ -\mu((x, 0]) \text{ if } x < 0 \end{cases} \tag{28}$$

Find all the points of discontinuity of F.

[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 $\int_{-\infty}^{x} \delta(s) ds$.

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 2.3.6 the algebra

$$\mathcal{A} = \left\{ \bigcup_{j=1}^{n} I_j : I_j \in \mathcal{E}, n \in \mathbb{N} \right\} \tag{29}$$

Definition 4.0.1. *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 4.0.2. *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

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

where all sets above are in $\mathcal{A}$. Thus $\sigma-$additivity reduces to continuity of μ_0 from above (see Theorem 2.4.1 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 $\varepsilon > 0$. 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 $\mu_0(A_k \setminus \hat{A}_k) \leqslant \varepsilon 2^{-k}$ for all k, where we denoted $\hat{A}_k = \cup_j (a'_{kj}, b_{kj}]$. Note that [7]

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

[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, $x \in A_{j_0}$.

Hence

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

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

Theorem 4.0.4. *1. For any Borel measure μ, the function F in (28) 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 4.0.2 shows that μ_F is a premeasure on $\mathcal{A}$. Since $\mathcal{A}$ generates $\mathcal{B}_{\mathbb{R}}$, Theorems 3.2.2 and 3.1.4 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 4.0.5. *Since $\mu_F = \mu_F^*$ on $\mathcal{M}_\mu$ we have, for $E \in \mathcal{M}_\mu$,*

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

Since for any $\varepsilon > 0$ 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. $\mu_F(J_j \setminus I_j) \leqslant 2^{-j}\varepsilon$ (check), it follows that for $E \in \mathcal{M}_\mu$,

$$\mu_F(E) = \inf\left\{ \sum_{i \in \mathbb{N}} \mu((a_i, b_i)) : E \subset \bigcup_{i \in \mathbb{N}} (a_i, b_i) \right\} \tag{31}$$

Definition 4.0.6. *A Borel measure on a topological space X is regular if for any $E \in \mathcal{B}_X$ we have*

$$\inf\{\mu(\mathcal{O}) : E \subset \mathcal{O}, \mathcal{O} \text{ open}\} = \mu(E) = \sup\{\mu(F) : E \supset K, K \text{ compact}\} \tag{32}$$

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

Lemma 4.0.7. *For any $\varepsilon > 0$ and any $E \in \mathcal{B}$ there is an $\mathcal{O} \supset E$ open s.t. $\mu(\mathcal{O} \setminus E) < \varepsilon$.*

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

Theorem 4.0.8. *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 4.0.7 shows that for any E there is an $\mathcal{O}$ with measure arbitrarily close to $\mu(E)$.

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

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

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

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

□

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 4.0.9. *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 $\varepsilon_n \to 0$ and for each n pick $\mathcal{O}_n$ open and K_n compact s.t.

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

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

Definition 4.0.11. *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 4.0.12. *If $E \in \mathcal{L}$ then $E + x_0$ and rE are in $\mathcal{L}$ and*

$$m(E + x_0) = m(E); \quad m(rE) = |r|m(E) \tag{33}$$

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 (33) holds for intervals. It follows that $\mathcal{B}_{\mathbb{R}}$ is also invariant under translations and dilations, and$m$ satisfies (33) on $\mathcal{B}_{\mathbb{R}}$. Since the translation and dilation of a null set is a null set (why?), the result follows from Theorem 4.0.9. □

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 4.1.1. *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 definied 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 $0.a_1a_2\cdots$ 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 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 5.1.1. *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 2.2.1. □

Proposition 5.1.2. *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 $\mathcal{B}_Y$. □

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

Definition 5.1.3. 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 5.1.4. 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 5.1.5 (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 5.1.6. 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 5.1.7 (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 5.1.8–5.1.13 are straightforward and left as an exercise.

Proposition 5.1.8. *Let $(X, \mathcal{M})$ be a measurable space and $f : X \to \mathbb{R}$. Then the following are equivalent:*

1. *f is measurable.*
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 5.1.9. 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 5.1.10. *Let $(X\mathcal{M})$ be a measurable space, and $Y, \mathcal{M}, Y_\alpha, \mathcal{M}_\alpha$ be as in Definition 5.1.9. 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).*

[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.

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 5.1.11. *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 5.1.12. *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 5.1.13. *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 5.1.14. *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 5.2.1. *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*

$$\sum_{j=1}^{n} z_j \chi_{A_j} \tag{34}$$

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 (34).

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 5.2.2. *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 $(J_{nk})_{k=1,\dots,2^{2n}}$ 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 5.2.3. *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 5.2.4. *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 2.4.3, and it extends by linearity $\mathfrak{S}$. Let $(\varphi_n)_{n\in\mathbb{N}}$ 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 $(\varphi_n)_{n\in\mathbb{N}}$ 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 5.3.1. *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 $\varphi \in L^+ \cap \mathfrak{S}$. Then, for some $n \in \mathbb{N}$, $\operatorname{ran}(\varphi) = \{a_1, ..., a_n\} \subset [0, \infty)$ and

$$\varphi = \sum_{j=1}^{n} a_j \chi_{A_j}; \quad A_j := f^{-1}(\{a_j\}) \tag{35}$$

It is natural to define the integral of φ by

$$\int \varphi d\mu = \sum_{j=1}^{n} a_j \mu(A_j) \tag{36}$$

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

$$\int_A \varphi d\mu = \int \chi_A \varphi d\mu \tag{37}$$

Proposition 5.3.2. *Let $\varphi = \sum_{i=1}^{n} a_i \chi_{A_i}, \psi = \sum_{i=1}^{m} b_j \chi_{B_i} \in L^+ \cap \mathfrak{S}$. Then*

1. *(Compatibility with the cone structure) $\int \varphi d\mu \geqslant 0$, $\forall c \geqslant 0 : \int c\varphi = c \int \varphi$ and $\int(\varphi + \psi) = \int \varphi + \int \psi$.*
2. *$A \mapsto \int_A \varphi$ 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 $\varphi + \psi$ 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 $\varphi = \chi_B$ for some measurable B, $\int \varphi = \mu(A \cap B)$ which is a measure on $\mathcal{M}$. The rest follows from Lemma 5.3.1. □

Note that 1. implies

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

Definition 5.3.3. *If $f \in L^+$, we define*

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

Proposition 5.3.4. *Def. 5.3.3 coincides with (37) for $f \in \mathfrak{S} \cap L^+$. The integral is nonnegative, commutes with the cone operations (cf. Proposition 5.3.2, 1.), addition and multiplication by nonnegative numbers.*

Proof. For additivity, see Theorem 5.3.6 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 5.3.5 (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 $\varphi \leqslant f$ in $\mathfrak{S} \cap L^+$ s.t. $\alpha \int f d\mu \leqslant \int \varphi d\mu$. By monotonicity, the sets $A_n = \{x \in X : f_n(x) \geqslant \alpha\varphi\}$ 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

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

□

Theorem 5.3.6. 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 $(\varphi_n)_{n\in\mathbb{N}}, (\psi_n)_{n\in\mathbb{N}}$ increase to f and g respectively as in Theorem 5.2.2, then $\varphi_n + \psi_n \nearrow f + g$, and by dominated convergence

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

2. An application of the monotone convergence theorem.

□

Theorem 5.3.7. *For $f \in L^+$, $\int f d\mu = 0$ iff $f = 0$ a.e.*

If $f = 0$ a.e. and $0 \leqslant \varphi \leqslant f$ then clearly $\varphi = 0$ a.e. implying (check) $\int \varphi d\mu = 0$. 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 5.3.8. *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 5.3.7. □

The second important result is the following.

Theorem 5.3.9 (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

$$f_n = \begin{cases} \chi_{(1,2]}, & \text{if } n \text{ is even} \\ \chi_{[0,1]}, & \text{if } n \text{ is odd} \end{cases} \tag{38}$$

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.

Proposition 5.3.10. *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 5.3.11. *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 6.0.1. *If $f : X \to \mathbb{C}$ define*

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

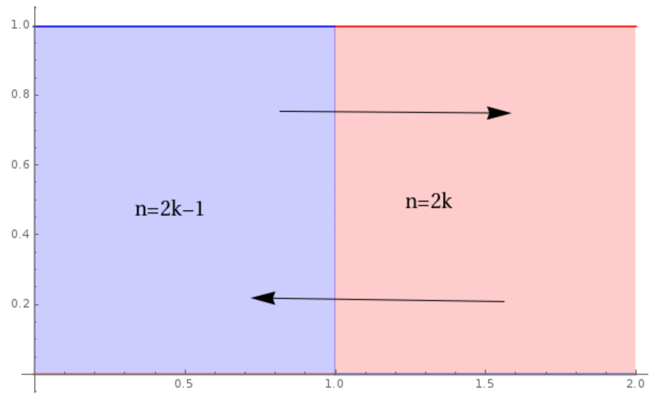

Figure 2: The sequence in (38).

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 6.0.2. *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 $\varphi(v) = \varphi(v^+) - \varphi(v^-)$. Additivity on C now translates into additivity on V [9]. □

Proposition 6.0.3. *If f is integrable, then*

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

Proof. Let $\alpha = \text{csgn}(\int f)$, $\beta = \overline{\alpha}$ 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 $\int |f| = 0$.*

[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 $\varphi(v^+) + \varphi(w^-) = \varphi(v^+ + w^-) = \varphi(w^+ + v^-) = \varphi(w^+) + \varphi(v^-)$ immediately implying consistency

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 6.0.4. *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 5.2.4 any equivalence class contains a Borel measurable function (still nonunique). Another advantage of working with equivalence classes is the following:

Proposition 6.0.5. *L^1 is a normed vector space with*

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

Proof. This is an easy exercise. ☐

Definition 6.0.6. *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 6.0.7. *1. If $f \in L^1$, then the set $\{x : F(x) \neq 0\}$ is σ-finite for any $F \in f$.*

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 5.3.11.

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 6.0.8 (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 6.0.9. *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 6.0.10. 1. *(Density of simple functions in L^1) For any $f\in L^1$ and any $\varepsilon>0$ there is an L^1-simple function φ s.t.*

$$\int|f-\varphi|d\mu<\varepsilon$$

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.*

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

Proof. 1. Let $(\varphi_n)_{n\in\mathbb{N}}$ be a sequence of simple functions converging to f, as in Theorem 5.2.2. Then $|\varphi_n|\leqslant f$ and dominated convergence implies $\lim_{n\to\infty}\int|\varphi_n-f|d\mu=0$. Thus, for any $\varepsilon>0$ there is an n s.t., for $\varphi=\varphi_n$, $\int|\varphi-f|d\mu<\varepsilon$

2. Let φ be as above, $\varepsilon>0$, and write $\varphi=\sum a_j\chi_{A_j}$. The statement follows from the fact that, by Proposition 4.0.10 for any $\varepsilon>0$ and any j there is an open set $\mathcal{O}$ which is a finite union of intervals s. t. $\int|\chi_{A_j}-\chi_{\mathcal{O}}|d\mu=\mu(\mathcal{O}\Delta A_j)<\varepsilon/j$.

3. For each interval J and any $\varepsilon>0$ there is a continuous function g s.t. $\int|g-\chi_J|<\varepsilon$ (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 7.0.1. *A bounded function f on an interval $[a,b]$ is Riemann integrable if*

$$\sup_{\varphi\leqslant f;\ \varphi\in\mathfrak{S}_R}\int_a^b\varphi dx=\inf_{\varphi\geqslant f;\ \psi\in\mathfrak{S}_R}\int_a^b\psi dx \tag{39}$$

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

Note 7.0.2. *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 7.0.3. 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 $(\varphi_n)_{n\in\mathbb{N}}$ 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 $\psi - \varphi \geqslant 0$. Since for all n, $|\varphi_n| \leqslant |f| + |\varphi_1|$ and $|\psi_n| \leqslant |f| + |\psi_1|$, dominated convergence applies and

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

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

2. Take the φ_k, ψ_k as in Note 7.0.2. Note that there must exist a sequence of partitions P_k of $[a,b]$ such that, as $n \to \infty$ we have $\sup_{I_{nk}} |\varphi_n - \psi_n| \to 0$ 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 7.0.4. 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 8.0.1. *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*

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

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 6.0.10.)*

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

Definition 8.0.2 (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 $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ implies that $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Lemma 8.0.3 (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 $x = \rho e^{i\varphi}$ and $e^{i\varphi}$ is in the right half plane.

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

$$\lim_{x\to\infty} x^{\beta+1} \int_0^\infty e^{-xs} G(s) ds = \lim_{x\to\infty} \int_0^\infty \frac{G(t/x)}{(t/x)^\beta} e^{-t} t^\beta dt = \Gamma(\beta+1) \tag{40}$$

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

Note 8.0.4. 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 8.0.5. 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

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

$$\bigcap_{n_2\in\mathbb{N}} \bigcup_{n_1\in\mathbb{N}} A_{\mathbf{n}} = (\forall n_2)(\exists n_1)P(n_1, n_2, ..n_k)(x)\} \tag{42}$$

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 (41),(42), we will sometimes use the shorthand

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

We also see that

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

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

$$\mu(P_1) < \infty \Rightarrow \mu(\forall n : P_n) \leqslant \inf_n \mu(P_n) \tag{43}$$

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

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

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

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

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

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

is one such metric.

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

Theorem 9.0.2 (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,

$$\mu\left(|F_{n+1}(x) - F_n(x)| \geqslant 2^{-n}\right) \leqslant 2^{-n} \text{ and} \tag{44}$$

$$\mu\left((\exists m \geqslant n)|F_n(x) - F_m(x)| \geqslant 2^{-n}\right) \leqslant 2 \cdot 2^{-n} \tag{45}$$

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 (45) and (43), $\mu(N) = 0$. 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 9.0.3. *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

$$\underbrace{\mu(|f - f_n| \geqslant y) \leqslant y^{-1} \int |f_n - f| d\mu}_{\text{this is called Markov's inequality}} \leqslant \|f_n - f\|_1 \tag{46}$$

□

Theorem 9.0.4 (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 $\varepsilon > 0$ there is an A s.t. $\mu(X \setminus A_\varepsilon) < \varepsilon$ and $(f_n)_{n\in\mathbb{N}}$ converges uniformly on A_ε.*

Proof. Let $\varepsilon > 0$. 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 $\mu(A_k) > \mu(X) - \frac{\varepsilon}{2^k}$ 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 9.0.5 (Lusin's theorem). *Let $f : [a, b] \to \mathbb{C}$ be Lebesgue measurable. Then for any $\varepsilon > 0$ there is a set $A_\varepsilon \subset [a, b]$ of measure $> b - a - \varepsilon$ s.t. $f|_{A_\varepsilon}$ is continuous.*

Proof. This follows easily from Egoroff's theorem and Theorem 6.0.10. 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. $\mu(\mathcal{O} \setminus K) < \varepsilon$ 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 9.0.6. *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 9.0.7. 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 9.0.8. *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 10.0.1. 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 10.0.2.

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 2.3.2.

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 9.0.8. □

Definition 10.0.3 (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 10.0.4 (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

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

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

3. $(\cup E_j)_x = \cup (E_j)_x$, 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 5.2.2. □

Theorem 10.0.5. *Let $(X, \mathcal{S}, \mu)$, $(Y, \mathcal{T}, \lambda)$ be $\sigma-$ finite measure spaces. Let $Q \in \mathcal{S} \times \mathcal{T}$, and define*

$$\varphi = x \mapsto \lambda(Q_x); \ \psi = y \mapsto \mu(Q^y) \tag{47}$$

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

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

Proof. We see from Theorem 10.0.4 that the definitions (47) and (48) make sense. Note also that $\lambda(Q_x) = \int_Y \chi_Q(x,y) d\lambda(y)$, and thus we can write (48) as

$$\int_X d\mu \int_Y \chi_Q d\lambda = \int_Y d\lambda \int_X \chi_Q d\mu \tag{49}$$

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:

1. $\mathcal{M}'$ contains all measurable rectangles;
2. $\mathcal{M}'$ is closed under countable monotone unions, $\cup E_i$, $E_i \subset E_{i+1}$;
3. $\mathcal{M}'$ is closed under countable disjoint unions;
4. $\mathcal{M}'$ 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 $\varphi_i = \lambda((E_i)_x), \varphi = \lambda(E_x)$, $\psi_i = \mu((E_i)^y), \varphi = \mu(E^y)$. Continuity from below of λ and μ implies that $\varphi_i \nearrow \varphi$, $\psi_i \nearrow \psi$, and (48) 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 10.0.6. *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 6.0.9 shows that $\mu \times \lambda$ is $\sigma-$ additive on $\mathcal{S} \times \mathcal{T}$. Check that $\mu \times \lambda$ is $\sigma-$finite.

Theorem 10.0.7 (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,

$$\int_{X\times Y} f d(\mu\times\lambda) = \int_X d\mu(x) \int_Y f(x,y) d\lambda(y) = \int_Y d\lambda(y) \int_X f(x,y) d\mu(x) \tag{50}$$

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 (50) holds.*

Proof. (a) If $Q \in \mathcal{S}\times\mathcal{T}$ and $f = \chi_Q$, then this follows from Theorem 10.0.4. 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 $\text{ran}(f) = \mathbb{R}$, 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 10.0.5, 10.0.7 cannot be omitted.

Note 10.0.8. *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 10.0.7 to the completion of the measures $\mu,\lambda,\mu\times\lambda$ is left as an exercise:

Theorem 10.0.9. *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 10.0.7 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* $\dfrac{1}{x^2} = \int_0^\infty te^{-xt}dt$ *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 11.1.1. *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 $\varepsilon > 0$, $m(Q \Delta \cup_{j=1}^n R_j) < \varepsilon$ for some disjoint rectangles R_j whose sides are intervals.*

Proof. By theorem 3.2.2 m^n is the extension of m^n, as restricted to the algebra $\mathcal{E}$ of elementary sets. In particular, for any $\varepsilon > 0$ there is a disjoint family of rectangles R_k containing Q s.t. $\mu(\cup_1^\infty R_k \cap Q^c) < \varepsilon$. With this, the proof of Theorem 4.0.8 translates with little change to a proof of 1; the proof of 2. is the same, up to notations to that of Theorem 4.0.9. 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. $\mu(\mathcal{O}) < \mu(A) + \varepsilon/2$. Now $\mathcal{O} = \cup_{j=1}^\infty I_j$ for some open intervals I_j, and thus there is an N s.t. $\mu(\mathcal{O} \setminus \cup_{j=1}^N I_J) < \varepsilon/2$. ☐

Theorem 11.1.2. *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 11.1.3. *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 11.1.4 (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 11.1.5. *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 11.1.6. *If A is measurable, then $m(TA) = |\det T|m(A)$.*

Note 11.1.7. *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 11.1.4, the rest is a corollary of that theorem. $\square$

Theorem 11.1.8 (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 11.1.9. *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 11.1.7. By density and theorem 11.1.1 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 11.1.10. *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*

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

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

Lemma 11.1.11. *For $0 < \varepsilon < d$ and $x_0 \in R$ let $R_0 = x_0 + \varepsilon R$. We have*

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

uniformly in x_0.

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

$$\lim_{\varepsilon \to 0} \varepsilon^{-n} \int_{R_0} f(G(x)) J_x dx \ = \ \lim_{\varepsilon \to 0} \int_R f(z_0 + \varepsilon M_{x_0} y + \varepsilon \varphi(x_0, y, \varepsilon)) J_{x_0 + \varepsilon y} dy \ = \ m(R) f(z_0) J_{x_0} \quad (51)$$

uniformly in x_0.

Next, define ψ for G^{-1} as in Note 11.1.10. Note that $z_0 + \varepsilon u \in G(R_0)$ means $x_0 + \varepsilon M_{x_0}^{-1} u + \varepsilon \psi(z_0, u, \varepsilon) \in R_0$ or $u + M_{x_0} \psi(z_0, u, \varepsilon) \in M_{x_0} R$ which means

$$\lim_{\varepsilon \to 0} \varepsilon^{-n} \int_{G(R_0)} f dm = \lim_{\varepsilon \to 0} \int f(z_0 + \varepsilon u + \varepsilon M_{x_0} \psi(z_0, u, \varepsilon)) \chi_{M_{x_0} R}(u + M_{x_0} \psi(z_0, u, \varepsilon)) du = J_{x_0} m(R) f(z_0) \quad (52)$$

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

To end the proof of the theorem, take $\varepsilon = 1/N$ 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 12.0.1. *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 12.0.2. *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,

$$\int_{\mathbb{R}^n} \exp\left(-a\sum_{k=1}^n x_i^2\right) dm = \left(\int_{\mathbb{R}} e^{-ax^2} dx\right)^n \tag{53}$$

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

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

which, using (53) implies the result.

2. Now we write the left side of (53) in polar coordinates in $\mathbb{R}^n$. Let $\mathcal{S} = \sigma(S^{n-1})$.

$$\pi^{n/2} = \int_{\mathbb{R}^n} \exp\left(-\sum_{k=1}^{n} x_i^2\right) dm = \mathcal{S}\int_0^\infty e^{-r^2} r^{n-1} dr = \tfrac{1}{2}\mathcal{S}\int_0^\infty e^{-u} u^{n/2} du = \tfrac{1}{2}\Gamma(\tfrac{n}{2})\mathcal{S} \tag{54}$$

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 13.0.1. *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 13.0.2. *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 13.0.3. *In the setting of Definition 14.1.1, 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 13.0.4. *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 13.0.5. *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 13.0.6. *Any subset of a positive set is a positive set.*

Proposition 13.0.7. *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 13.1.1 (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 13.1.2 (Quasi-positive sets). *Let A be a set of finite measure. Then, for any $\varepsilon > 0$ there is an $A_\varepsilon \subset A$. s.t. all its subsets have measure $\geqslant -\varepsilon$.*

Proof. By contradiction. Let $B_1 \subset A, \nu(B_1) < -\varepsilon$. 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. $\nu(B_2) < -\varepsilon$. Inductively, we construct a set of B_k contained in $A \setminus \cup_{j<k} B_j$ with $\nu(B_k) < -\varepsilon$. But then $B = \cup B_k$ has measure $-\infty$ and $\nu(A) = \nu(B) + \nu(A \setminus B) = -\infty$, contradiction. □

Lemma 13.1.3. *If A is of finite measure, then A contains a positive set $P, \nu(P) \geqslant \nu(A)$.*

Take $\varepsilon = 1/n$ 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 13.1.4.

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 13.1.5 (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 13.1.6 (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 13.1.7. *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 χ_A is the characteristic function of a null set, then $\chi_A f = 0$ a.e. □

14 The Lebesgue-Radon-Nikodym theorem

This theorem is, in a sense, a converse of Lemma 13.1.7.

Theorem 14.0.1 (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 $A \in \mathcal{M}$,*

$$\nu(A) = \nu(A \cap N) + \int_A f d\mu \tag{55}$$

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 14.0.2. *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

$$k_1 = \min\left\{k \in \mathbb{N} : \exists E_1 \in \mathcal{E} \cap \mathcal{P}(X \setminus E_0) \text{ with } \mu(E_1) > k^{-1}\right\} \tag{56}$$

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 14.0.3. *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 14.0.2. 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 14.0.4. *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 14.0.5. *If $\nu \ll \mu$ and f is as in Theorem 14.1.6, then we write $f = \dfrac{d\nu}{d\mu}$.*

Corollary 14.0.6. *1. Assume ν, μ are $\sigma-$finite measures, μ is positive, $\nu \ll \mu$ and $\varphi \in L^1(\mu)$. If $f = d\nu/d\mu$, then $f\varphi \in L^1(\mu)$ and*

$$\int g d\nu = \int \varphi f d\mu$$

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 14.0.7. *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 14.0.8 (Existence of an upper bound). *If $(\mu_j)_{j=1,\dots,n}$ are measures, then $\mu_k \leqslant \sum_j \mu_j$ for all $k \leqslant n$.*

$\boxed{\textbf{HW 11/13 (Recitation day) : 4,5,6,7 p. 88 in Folland; turn in: Ex 41,42 in the notes.}}$

Lemma 14.0.9. *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 14.1.1. *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 14.1.2.

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 14.1.3. *The variation of a complex measure ν is the set function*

$$|\nu|(E) = \sup_{\uplus A_i = E} \sum_{i\in\mathbb{N}} |\nu(A_i)| \qquad \forall E \in \mathcal{M} \tag{57}$$

The total variation of ν is defined as $|\nu| = |\nu|(X)$.

Note 14.1.4.

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 14.1.5. *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 $\varepsilon > 0$ and let n be s.t. $|\mu|(|f| > 2n) < \varepsilon$. 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 $N \to \infty, \varepsilon \to 0$ and noting that $\sum_k \alpha_k \chi_{E_k} + n\chi_{|f|>n}$ converge pointwise to $|f|$. $\square$

The following generalization is immediate.

Theorem 14.1.6 (Lebesgue-Radon-Nikodym, L-R-N)**.** *Let ν be a complex measure and μ a σ−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 14.1.7. *We have $d\nu = f d|\nu|$ where $f \in L^1$, and $|f| = 1$ a.e.*

Proof. Note 14.1.4 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. $\square$

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 15.0.1. *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 $\varepsilon > 0$ there is a $\delta > 0$ s.t.*
$$\left| \frac{\mu(I)}{m(I)} - A \right| < \varepsilon$$
for any open interval of length $< \delta$ containing x.

Proof. Straight from the definition of differentiation. $\square$

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

$$(Q_r\mu)(x) = \frac{\mu(B_{x,r})}{m(B_{x,r})} \tag{58}$$

where $m = m^k$ is the Lebesgue measure.

Definition 15.0.2. *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 15.0.3 (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. $\square$

Definition 15.0.4 (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 15.0.5. *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 15.0.6. *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 15.0.7. *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 15.0.8 (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 15.0.9 (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 15.0.8. 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 15.1.1. *Let $f \in L^1(\mathbb{R}^k)$. The point $x \in \mathbb{R}^k$ is a Lebesgue point of f if*

$$\lim_{r \to 0} \frac{1}{m(B_{x,r})} \int_{B_{x,r}} |f - f(x)| dm = 0 \tag{59}$$

Note 15.1.2. *Clearly, if f is continuous, then all points are Lebesgue points.*

The following is a fundamental result in the theory of Lebesgue differentiation.

Theorem 15.1.3. *If $f \in L^1(\mathbb{R}^k)$, then almost every $x \in \mathbb{R}^k$ is a Lebesgue point of f.*

Proof. Let

$$(T_r f)(x) = \frac{1}{m(B_{x,r})} \int_{B_{x,r}} |f - f(x)| dm \text{ and } (Tf)(x) = \limsup_{r \to 0} (T_r f)(x)$$

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 15.1.4. *Assume μ is a complex Borel measure on $\mathbb{R}^k$ and that $\mu \ll m$. Then $D\mu$ (cf. Definition 15.0.2) exists a.e. and equals $d\mu/dm$.*

Proof. Let $f = d\mu/dm$. Then,

$$f(x) = \lim_{r \to 0} \frac{1}{m(B_{x,r})} \int_{B_{x,r}} f \, dm = \lim_{r \to 0} \frac{\mu(B_{x,r})}{m(B_{x,r})} \quad a.e.[m]$$

Thus, $(D\mu)(x)$ exists and equals $f(x)$ at every Lebesgue point of f. □

15.1.2 Nicely shrinking sets

Definition 15.1.5. *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 15.1.6. *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

$$\lim_{n\to\infty} \frac{1}{m(E_n(x))} \int_{E_n(x)} |f - f(x)| dm = 0 \tag{60}$$

If the (E_n) are balls, then (60) 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 15.1.7. *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

$$(\overline{D}\mu)(x) := \lim_{n\to\infty}(M_n\mu)(x) \tag{61}$$

is measurable. Note also that $M_n\mu \leqslant M\mu$.

Choose $\lambda > 0, \varepsilon > 0$ and a compact set K s.t., by the regularity of Borel measures, $\mu(K) > |\mu| - \varepsilon$. Let μ_1 be the restriction of μ to K, and $\mu_2 = \mu - \mu_1$. We see that $|\mu_2| < \varepsilon$, and if $x \in K^c$ we have

$$(\overline{D}\mu)(x) = (\overline{D}\mu_2)(x) \leqslant (M\mu_2)(x)$$

hence

$$m(\overline{D}\mu > \lambda) \leqslant m(K) + m(M\mu_2 > \lambda) \leqslant 3^k\lambda^{-1}|\mu_2| \leqslant 3^k\lambda^{-1}\varepsilon \tag{62}$$

Since (62) holds for arbitrary $\varepsilon > 0, \lambda > 0$, the result follows. □

Corollary 15.1.8. *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*

$$\lim_{k\to\infty} \frac{\mu(E_k(x))}{m(E_k(x))} = 0 \ \ a.e.[m] \tag{63}$$

As another corollary, we have the following strengthening of Theorem 15.1.4.

Theorem 15.1.9. *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 15.2.1. *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 15.2.2. *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 $\{f \le a\}$ has density 1 at x_0;
2. The supremum of $a \in \{-\infty\}\cup\mathbb{R}$ such that the set $\{f \ge a\}$ has density 1 at x_0

They are usually denoted by

$$\text{ap}\limsup_{x\to x_0} f(x) \qquad \text{and} \qquad \text{ap}\liminf_{x\to x_0} f(x)$$

(some authors use also the notation $\overline{\lim}\,\text{ap}$ and $\underline{\lim}\,\text{ap}$). It follows from the definition that ap lim inf $\le$ ap lim sup: if the two numbers coincide then the result is called approximate limit of f at x_0 and it is denoted by

$$\text{ap}\lim_{x\to x_0} f(x)\,.$$

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 15.2.3. *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 15.2.3.

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 15.2.4. *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 w.l.o.g. that μ itself is positive. (The same applies to complex measurable functions.)

Recall the definitions of the distribution function of a measure (p. 28) and of the variation of a complex measure (p. 62).

Exercise 49. *Show that the following is an equivalent definition of the variation of a measure:*

$$|\nu|(E) = \sup\left\{\sum_{i=1}^{n} |\nu(A_i)| : n \in \mathbb{N}, \biguplus_{1}^{n} A_i = E\right\} \qquad \forall E \in \mathcal{M} \tag{64}$$

Definition 16.0.1. *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*

$$T_F(x) = \sup\left\{\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, -\infty < x_0 < ... < x_n = x\right\} \tag{65}$$

2. *Using the fact that $|\mu|$ is a Borel measure, show that T_F is increasing and right-continuous.*

Definition 16.0.2. *If $F : \mathbb{R} \to \mathbb{C}$, we define the total its variation function T_F by (65).*

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

$$T_F([a,b]) := \sup\left\{\sum_{j=1}^{n} |F(x_j) - F(x_{j-1})| : n \in \mathbb{N}, a = x_0 < ... < x_n = b\right\} = T_F(b) - T_F(a)$$

Note 16.0.3 (Geometrical interpretation)**.** Since all norms on $\mathbb{R}^k$ are equivalent, a real-valued function is in $BV[a, b]$ iff

$$\sup\left\{\sum_{j=1}^{n} \sqrt{(x_j - x_{j-1})^2 + (F(x_j) - F(x_{j-1}))^2} : n \in \mathbb{N}, a = x_0 < ... < x_n = b\right\} < \infty$$

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 16.0.4. 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 $\sum_{j\geqslant 0} |F(x_j) - F(x_{j-1})| < \varepsilon$ for any decreasing unbounded sequence $x_0, x_1, ...$, hence $T_F(-\infty) = 0$.

Note 16.0.5.

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

$$F(x) \in [F(x^-), F(x^+)] \tag{66}$$

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 15.1.9 to show that $f = G'$ exists a.e., is in L^1 and $f = d\mu/dm$.

Proposition 16.0.6. *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 w.l.o.g. 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

$$|H(y) - H(x)| \leqslant \lambda([x, y]) \tag{67}$$

and note that (67) 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 16.1.1. *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 16.1.2. 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 16.1.3. *$F \in NBV$ 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 w.l.o.g. that F is real-valued. Now Proposition 16.1.1 implies $T_F \in NBV$ and thus the two increasing functions in the canonical decomposition of F are also NBV, by Note 16.0.5 3. The rest is immediate. ☐

Definition 16.1.4. *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 16.1.2.*

Note 16.1.5. $F \in AC([a,b]) \Rightarrow F \in BV([a,b])$ 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 $n\varepsilon$.

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 16.1.6. *If $F \in BV([a,b])$ is right-continuous, then $\mu_F \ll m$ iff for any $\varepsilon > 0$ there is a $\delta > 0$ s.t. for any* **finite disjoint collection of intervals** $(a_j, b_j) \subset [a,b]$, $j = 1, ..., n$

$$\sum_{j=1}^{n}(b_j - a_j) < \delta \Rightarrow \sum_{j=1}^{n} |F(b_j) - F(a_j)| < \varepsilon \tag{68}$$

Proof. If $\mu_F \ll \mu$, then (68) is a direct consequence of the definition of absolute continuity. Conversely, fix $\varepsilon > 0$, let δ (68) be s.t. the last inequality in (68) holds with $\varepsilon/2$ 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. $\sum_n^\infty |\mu|(J_k) < \varepsilon/2$. Then $\sum_1^\infty |\mu|(J_k) < \varepsilon$. ☐

Proposition 16.1.7. *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 13.1.7 and Theorem 15.1.6. ☐

We have proved the following important result:

Theorem 16.1.8 (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. $F(x) = F(a) + \displaystyle\int_a^x f(s)ds$ *for some* $f \in L^1([a,b], m)$;
3. F' *exists a.e.,* $F' \in L^1([a,b], m)$ *and* $F(x) = F(a) + \displaystyle\int_a^x F'(s)ds$

An interesting result (see Rudin) is

Theorem 16.1.9. *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 16.2.1. *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 16.3.1. *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 16.3.2. *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 17.0.1. *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 17.1.1. *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...)

Theorem 17.1.2 (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)."

[10] In the sense of the one point compactification of $\mathbb{R}$.

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 17.2.1. *A Hausdorff space $\mathbb{X}$ is locally compact (LCH) if every point has a compact neighborhood.*

Here $\mathbb{X}$ is always an LCH space, $C, K, \mathcal{O}$ are a closed, compact and open resp. sets in $\mathbb{X}$.

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 17.2.2. *$E \subset \mathbb{X}$ is closed* **iff** *$E \cap K$ is closed for any compact K.*

Proof. Exercise. □

Proposition 17.2.3. *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, w.l.o.g., we assume $\mathcal{O}$ is precompact. Then $\partial\mathcal{O}$ and x are closed and Note 17.1.1 completes the proof. □

Proposition 17.2.4. *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 17.2.3 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 17.2.5 (Urysohn's Lemma, LCH version)**.** *Let $K \subset \mathcal{O}$ as in Proposition 17.2.4. 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 17.2.6 (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 17.2.7. *A space is $\sigma-$compact if it is the countable union of compact sets.*

Proposition 17.2.8. *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 17.2.9. *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 17.2.8 above; then $\mathcal{O}'_n = \cup_1^n \mathcal{O}_j$ is such a family. □

17.3 Support of a function

Definition 17.3.1. 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 $C \prec f$.

17.4 Partitions of unity

Definition 17.4.1.

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 17.4.2. *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 17.5.1. • 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 17.5.2. $C_c(\mathbb{X}) \subset C_0(\mathbb{X})$ 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 17.5.3. 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 17.5.4. *$\mathbb{X}^*$ 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 17.5.5. 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 17.6.1. *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 17.6.2. *$C(\{0,1\})$ 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 17.6.2, $\mathcal{A} = C(\{0,1\})$. 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 17.6.3. *$\mathcal{A} \subset C(\mathbb{X})$ is called a lattice if $f, g \in \mathcal{A}$ implies $f \wedge g$ and $f \vee g$ are also in $\mathcal{A}$.*

Note 17.6.4. *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 17.6.5. *A subset $\mathcal{A}$ of $C(\mathbb{X}, \mathbb{R})$ is said to separate points if $x \neq y \in \mathbb{X} \Rightarrow \exists f \in \mathcal{A}, f(x) \neq f(y)$.*

Here $\mathbb{X}$ is always an compact Hausdorf space.

Theorem 17.6.6 (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 17.6.7. *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 17.6.8. *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 $\varepsilon > 0$ we construct a $g \in \mathcal{A}$ s.t. $\|f - g\|_u < \varepsilon$ 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 $U_{xy} = \{z \in \mathbb{X} : f(z) < g_{xy}(z) + \varepsilon\}$ and $L_{xy} = \{z \in \mathbb{X} : f(z) > g_{xy}(z) - \varepsilon\}$. 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 $f < \Gamma_y + \varepsilon$ on $\mathbb{X}$ and $f > \Gamma_y - \varepsilon$ 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 $\|f - g\|_u < \varepsilon$ completing the proof. ☐

Proof of Theorem 17.6.10. Clearly, for any $x, y \in \mathbb{X}$, the restriction $\mathcal{A}_{xy} = \{g$ 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 17.6.8, $\mathcal{A} = C(\mathbb{X}, \mathbb{R})$. 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 17.6.9. *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 1.1.1) 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 17.6.10 (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 17.6.10 $u(x) + iv(x) \in \mathcal{A}$ for any u, v in $C(\mathbb{X}, \mathbb{R})$. $\square$

Note 17.6.11. *By Urysohn's lemma, in any normal space, continuous functions separate points.*

Exercise 53.

1. Use Stone-Weierstrass to show that $\{e^{2\pi i k x} : k \in \mathbb{Z}\}$ 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. $\forall 0 \leqslant n \in \mathbb{Z}$, $\displaystyle\int_0^1 s^n f(s) ds = 0$. 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 18.0.1. 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 18.0.2. 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 18.0.3. 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 18.1.1. 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 18.1.2. 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 19.0.1. 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 19.0.2 (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 F. By construction, $F \subset z_\lambda$ for some λ, $z_0 = z_\lambda$ on F, and $(z_\lambda)|_F$ is a partial cluster point of (x_α). Then z_0 is an upper bound for $\mathcal{L}$.

Let z be a maximal element of $\mathcal{P}$. If the dom(z)= 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 . □

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.

Theorem 19.0.3. *If I is countable and $\{X_i\}_{i\in I}$ are second countable and compact, then ZF+DC (the axiom of dependent choice) imply $X = \prod_{i\in I} X_i$ is compact.*

Proof. It is easy to check that X is also second countable.

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. □

20 Arzelá-Ascoli's theorem

Definition 20.0.1.
1. The family $\mathcal{F}$ of complex-valued functions on X is pointwise equibounded if $\forall x \in X \sup_{F\in\mathcal{F}} |F(x)| = M(x) < \infty$.
2. The family $\mathcal{F}$ is equicontinuous if $\forall\varepsilon\exists\delta$ s.t. $\forall x, y \in X$ and $F \in \mathcal{F}$, $d(x,y) < \delta \Rightarrow |F(x) - F(y)| < \varepsilon$.

Theorem 20.0.2.
1. *Let X be a separable metric space. If X is compact, then every sequence $\{F_n\}_{n\in\mathbb{N}} : X \to \mathbb{C}$ of equicontinuous and pointwise equibounded function has a subsequence which converges uniformly in X.*
2. *More generally, if X is a separable metric space, then 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 in X.*

Proof. Let d be a metric on X and let E be a countable dense subset of X.

1. $C(X)$ with the uniform norm is a metric space, thus sequential. The space $Y = \prod_{e\in E}\{z \in \mathbb{C} : |z| \leqslant M(e)\}$ is compact. Let $(F_{n,E})_{n\in\mathbb{N}}$ be the restriction of $(F_n)_{n\in\mathbb{N}}$ to E. This is a sequence in Y, and there is a sub*sequence* defined by a $g : \mathbb{N} \to \mathbb{N}$ s.t. $(F_{g(n,E)})_{n\in\mathbb{N}}$ converges. Take $\varepsilon > 0$ and let $\delta > 0$ be s.t. if $d(x,y) < \delta$, then $(\forall n)$ $(|F_n(x) - F_n(y)| < \varepsilon/3)$. Since E is dense, $\mathcal{O}_e := \{x : d(x,e) < \delta\}, e \in E$ cover K, and by compactness, there is a *finite* set $E_n = \{e_1, ...e_n\}$ s.t. for all $x \in K$, $d(x, E_n) < \delta$. For $e_k \in E_n$, let m_k be s.t. $\forall n, n' > m_k$, $|F_{g(n),E}(e_k) - F_{g(n'),E}(e_k)| \leqslant \varepsilon/3$, and $m = \max\{m_k\}$. By the triangle inequality, $|F_{g(n)}(x) - F_{g(n')}(x)| < \varepsilon$ for all $n, n' > m$ and $x \in X$, and the result follows.

2. Since X is a separable metric space, it is $\sigma-$compact. Let K_j be an increasing sequence of compact sets that cover X. Let (F_{g_1}) be a subsequence of (F_n) uniformly convergent on K_1, and inductively for $j \geqslant 2$, (F_{g_j}) be a subsequence of $(F_{g_{j-1}})$ uniformly convergent on K_j. Then, the diagonal sequence $F_{g_1(1)}, ..., F_{g_j(j)}, ...$ converges uniformly on any K_m. ☐

Note 20.0.3. 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 $|F(e) - F(y)| < \varepsilon$, which can be seen using the compact cover formulation of compactness.

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*.

Part II: 6212

Hilbert spaces are defined by abstracting the structure needed for the properties above to hold.

21 Norms, seminorms and inner products

Definitions

Given a vector space V over a subfield $\mathbb{K}$ of the complex numbers, a **norm** on V is a nonnegative-valued scalar function $p : V \to [0, +\infty)$ with the following properties: for all $a \in \mathbb{K}$ and all $u, v \in V$,

1.
$$p(u + v) \leqslant p(u) + p(v)$$
(p is subadditive, or: p satisfies the triangle inequality).

2. $p(av) = |a|p(v)$ (p is absolutely homogeneous, or absolutely scalable).

3. If $p(v) = 0$ then $v = 0$ is the zero vector (p is positive definite).

A **seminorm** on V is a function $p : V \to \mathbb{R}$ with the properties 1 and 2 above.

Every vector space V with seminorm p induces a normed space V/W, called the quotient space, where W is the subspace of V consisting of all vectors v in V with $p(v) = 0$ (check that W is a subspace). The induced norm on V/W is defined by:

$$p(W + v) = p(v)$$

Two norms (or seminorms) p and q on a vector space V are equivalent if there exist two positive constants c_1 and c_2 such that $c_1 q(v) \leqslant p(v) \leqslant_2 q(v)$ for every vector v in V.

A topological vector space is called **normable** (seminormable) if the topology of the space can be induced by a norm (seminorm).

An inner product $\langle x, y\rangle$ over a vector space is a complex-valued function that satisfies the following properties:

1. The inner product of a pair of elements is equal to the complex conjugate of the inner product of the swapped elements:
$$\langle y, x\rangle = \overline{\langle x, y\rangle}\,. \tag{1}$$

2. The inner product is linear in its first argument. For all complex numbers a and b,

$$\langle ax_1 + bx_2, y\rangle = a\langle x_1, y\rangle + b\langle x_2, y\rangle\,. \tag{2}$$

3. The inner product of an element with itself is positive definite:

$$\langle x, x\rangle \geq 0 \tag{3}$$

where the case of equality holds precisely when $x = 0$. It follows from properties (1) and (2) that a complex inner product is antilinear in its second argument, meaning that

$$\langle x, ay_1 + by_2\rangle = \bar{a}\langle x, y_1\rangle + \bar{b}\langle x, y_2\rangle\,. \tag{1+2}$$

It is easily checked that the quantity $\|x\| := \sqrt{\langle x, x\rangle}$ is a norm on $\mathcal{H}$.

22 Hilbert spaces

A Hilbert space $\mathcal{H}$ is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product (that is, the distance between $x, y \in \mathcal{H}$ is $\|x - y\|$).

Theorem 22.0.1 (Cauchy-Schwarz)**.** *For any $x, y \in \mathcal{H}$ we have $|\langle x, y\rangle| \leqslant \|x\|\|y\|$.*

We have equality iff x, y are linearly dependent.

Proof. There is nothing to prove if $x = 0$ or $y = 0$, so we assume this is not the case. Note now that for any z, $\|z\| \geqslant 0$. In particular, for any $a \in \mathbb{C}$ we have

$$0 \leqslant \|x - ay\|^2 = \langle x, x\rangle + |a|^2\langle y, y\rangle - 2\Re(a\langle y, x\rangle) = f(a) \tag{69}$$

We write $\langle y, x\rangle = |\langle x, y\rangle|e^{i\alpha}$ (if $\langle x, y\rangle = 0$ any α works). For $t \in \mathbb{R}$,

$$f(te^{-i\alpha}) = \langle x, x\rangle + t^2\langle y, y\rangle - 2t|\langle x, y\rangle| \geqslant 0$$

is a nonnegative quadratic polynomial in t 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. □

Proposition 22.0.2. *The function $x \to \|x\| = \sqrt{\langle x, x\rangle}$ is a norm.*

Proof. First of all, by the definition of the inner product and norm, $\|x\| = 0$ iff $x = 0$ and $\|\lambda x\| = |\lambda|\|x\|$. To prove the triangle inequality, we note that

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

□

22.1 Example: ℓ^2

Definition 22.1.1. *Let*

$$\ell^2 = \{x : \mathbb{N} \to \mathbb{C} \,|\, \|x\|^2 = \sum_{i\in\mathbb{N}} |x(i)|^2 =: \sum_{i\in\mathbb{N}} |x_i|^2 < \infty\}$$

and define

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

which, by Cauchy-Schwarz is well-defined on ℓ^2.

Proposition 22.1.2. *ℓ^2 is complete thus it is a Hilbert space.*

Proof. If $\{x_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in ℓ^2, then for every $i \in \mathbb{N}$ the number sequence of the ith component $\{(x_n)_i\}_{n\in\mathbb{N}}$ is Cauchy (indeed $|(x_n)_i - (x_m)_i|^2 \leqslant \|x_n - x_m\|^2$). Let $y_i = \lim_n (x_n)_i$.

We need to show that $y \in \ell^2$, and y is the limit of x_n. 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 $y \in \ell^2$. Similarly, since $\lim_{k\to\infty} \sum_{i=0}^{n} |(x_k)_i - y_i|^2 = 0$ for any n, by the above we can use dominated convergence to show that $\|x_k - y\| \to 0$.

□

Proposition 22.1.3. *The inner product is a continuous function from $\mathcal{H} \times \mathcal{H}$ to $\mathbb{C}$. In particular, if $\{(x_n, y_n)\}_{n\in\mathbb{N}}$ converges to (x, y) then $\langle x_n, y_n\rangle \to \langle x, y\rangle$.*

Proof. By Cauchy-Schwarz, as $\|h_1\|, \|h_2\| \to 0$ we have,

$$|\langle x + h_1, y + h_2\rangle - \langle x, y\rangle| = |\langle x, h_2\rangle + \langle h_1, y\rangle + \langle h_1, h_2\rangle| \leqslant \|x\|\|h_1\| + \|y\|\|h_2\| + \|h_1\|\|h_2\| \to 0$$

□

Proposition 22.1.4 (The parallelogram law)**.**

$$\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2) \tag{P.L.}$$

Proof. A straightforward calculation, see (69) above. □

We see that a Hilbert space is a complete normed space where the norm comes from an inner product. A natural and important question arises: given a norm, can we always define an inner product that induces the norm? The answer is no and, remarkably, (P.L.) is the necessary and sufficient condition for the norm to come from an inner product.

Proposition 22.1.5. *Let $\mathcal{S}$ be a complete normed space, with norm $\|\cdot\|$. Then the norm comes from an inner product iff it satisfies the parallelogram law.*

Proof. We have already shown that an inner-product-induced norm satisfies the parallelogram law. In the opposite direction, a calculation assuming the existence of an inner product leads the following explicit formula for the inner product, called *the polarization identity*:

$$\langle x,\ y\rangle = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2 + i\|x + iy\|^2 - i\|x - iy\|^2\right)\ \forall\, x, y \in \mathcal{H}$$

(for Hilbert spaces over $\mathbb{R}$ it has the form $\langle x,\ y\rangle = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right)$).

It remains to check that *assuming the parallelogram law* the formula above defines an inner product (meaning: with properties (1)...(3) above). This is elementary, but by no means trivial! See original proof by P. Jordan & J. von Neumann, Annals, 1935. A geometric argument based on Euclid's three line theorem is N. Falkner, MAA 100,3, (1993). □

Corollary 22.1.6. *The inner product is continuous.*

22.2 Orthogonality

The notion of orthogonality, $x \perp y$ if by definition $\langle x, y\rangle = 0$ obviously extends to general Hilbert spaces. So does the following

Proposition 22.2.1 (Pythagorean equality)**.** *If $x_1, ..., x_n$ are pairwise orthogonal, then*

$$\left\|\sum_{i=1}^{n} x_i\right\|^2 = \sum_{i=1}^{n} \|x_i\|^2$$

Proof.

$$\left\| \sum_{i=1}^{n} x_i \right\|^2 = \left\langle \sum_{i=1}^{n} x_i \middle| \sum_{i=1}^{n} x_i \right\rangle = \sum_{i,j \leqslant n} \langle x_i, x_j \rangle = \sum_{i=1}^{n} \langle x_i, x_i \rangle = \sum_{i=1}^{n} \|x_i\|^2$$

□

Definition 22.2.2. *The linear span (linear hull, or simply span) of a set of vectors $S = \{v_\alpha : \alpha \in A\}$ over the scalar field $\mathbb{K}$ is*

$$\operatorname{span}(S) = \left\{ \sum_{i=1}^{k} \lambda_i v_i \middle| k \in \mathbb{N}, v_i \in S, \lambda_i \in \mathbb{K} \right\}$$

22.3 The Gram-Schmidt process

Given a family $\{x_i\}_{i\in\mathbb{N}}$ of linearly independent vectors, we can construct, from them, an orthonormal family $\{e_i\}_{i\in\mathbb{N}}$, inductively: start with $v_1 = x_1$; let c be s.t. $v_2 = cv_1 + x_2 \perp v_1$ (which gives $c = -\langle x_2, v_1\rangle / \|v_1\|^2$). Having constructed $v_1, ..., v_n$ pairwise orthogonal, choose $c_{n1}, ..., c_{nn}$ s.t. $v_{n+1} = c_{n1}x_1 + ... + c_{nn}x_n + x_{n+1}$ is orthogonal on $v_1, ..., v_n$ (this is a linear system with nonzero determinant). Then $\{v_i\}_{i\in\mathbb{N}}$ is an orthogonal family with the property that span($\{x_1, ..., x_n\}$=span($\{v_1, ..., v_n\}$ for all n. All these v_i are nonzero vectors, and an orthonormal family is simply given by $e_i = v_i / \|v_i\|$.

Proposition 22.3.1. *Let $\{x_i\}_{i\in\mathbb{N}}$ be a set of vectors in $\mathcal{H}$ and let $\mathcal{V}$ be the closure of span($\{x_i\}_{i\in\mathbb{N}}$). We assume that $\mathcal{V}$ is infinite dimensional (the finite dimensional case is similar, and simpler). Then there exists an orthonormal set $\{e_i\}_{i\in\mathbb{N}}$ such that $\mathcal{V}$ is the closure of span($\{e_i\}_{i\in\mathbb{N}}$).*

Proof. We can assume w.l.o.g. that $\{x_i\}_{i\in\mathbb{N}}$ are linearly independent, since we can inductively eliminate the dependent vectors without affecting the span. With $\pi_v x = \dfrac{\langle v, x\rangle}{\langle v, v\rangle} v$, the Gram-Schmidt procedure is:

$$\begin{aligned}
v_1 &= x_1, & e_1 &= \frac{v_1}{\|v_1\|} \\
v_2 &= x_2 - \pi_{v_1} x_2, & e_2 &= \frac{v_2}{\|v_2\|} \\
&\vdots & &\vdots \\
v_k &= x_k - \sum_{j=1}^{k-1} \pi_{v_j} x_k, & e_k &= \frac{v_k}{\|v_k\|}. \\
&\ldots
\end{aligned}$$

Note that, for all $k \in \mathbb{N}$, we have span$\{x_1, ..., x_k\}$ =span$\{v_1, ..., v_k\}$, implying that span$\{x_i : i \in \mathbb{N}\}$ = span$\{v_i : i \in \mathbb{N}\}$ =span$\{e_i : i \in \mathbb{N}\}$, hence the closures of these spans also coincide. □

22.4 A very short proof of Cauchy-Schwarz

Proof. In case x, y are linearly dependent the inequality is an equality. Otherwise w.l.o.g., we may assume $\|x\| = \|y\| = 1$. Define $e_1 = x$ and let e_2 be obtained from e_1 and y by Gram-Schmidt.

Then $y = y_1 e_1 + y_2 e_2$ and
$$|\langle x, y\rangle|^2 = |y_1|^2 \leqslant |y_1^2| + |y_2^2| = 1$$

$\square$

22.5 Orthogonal projections

In the following, $\mathcal{H}$ is a Hilbert space.

Definition 22.5.1 (The orthogonal complement of a space)**.** *If $\mathcal{S}$ is a subspace of $\mathcal{H}$, then its orthogonal complement, $\mathcal{S}^\perp$ is the* ***closed linear subspace*** *(check these properties!) of $\mathcal{H}$ defined by*

$$\mathcal{S}^\perp = \{x \in \mathcal{H} : (\forall y \in \mathcal{S})(\langle x, y\rangle = 0)\}$$

The sum of two subspaces $\mathcal{V}_1, \mathcal{V}_2$ is defined by

$$\mathcal{V}_1 + \mathcal{V}_2 = \{v_1 + v_2 : v_1 \in \mathcal{V}_1, v_2 \in \mathcal{V}_2\}$$

If $\mathcal{V}_1 \cap \mathcal{V}_2 = \{0\}$, then the sum is **direct***, written $\mathcal{V}_1 \oplus \mathcal{V}_2$ and for any $x \in \mathcal{V}_1 \oplus \mathcal{V}_2$ there is a unique pair v_1, v_2, $v_i \in \mathcal{V}_i$ s.t. $x = v_1 + v_2$ (check!).*

Lemma 22.5.2 (Orthogonality and an extremal property)**.** *If $\mathcal{M}$ is a* **closed** *subspace of $\mathcal{H}$, then*

1. *there is a unique $\mu \in \mathcal{M}$ s.t. $\forall m \in \mathcal{M}$, $m \neq \mu$, $\|x - m\| > \|x - \mu\|$.*
2. *If μ is as in 1., then $x - \mu \in \mathcal{M}^\perp$. Conversely, if $y \in \mathcal{M}$ is s.t $x - y \in \mathcal{M}^\perp$, then $y = \mu$.*

Proof. 1. Let $d = \inf_{y \in \mathcal{M}} \|x - y\|$. Since $0 \in \mathcal{M}$, $d \leqslant \|x\|$. Thus there is a sequence $\{y_m\}_{m \in \mathbb{N}}$ in $\mathcal{M}$ s.t. $d - \|x - y_m\| \to 0$. We show that this sequence is convergent to some $\mu \in \mathcal{M}$. Note that this proves both existence and uniqueness of a $\mu \in \mathcal{M}$ s.t. $\|x - \mu\|$ is minimal.

Since $\mathcal{M}$ is a closed subspace of the complete Hilbert space $\mathcal{H}$, it suffices to show that $\{y_m\}_{m \in \mathbb{N}}$ is Cauchy. Here we use the parallelogram law:

$$\|y_m - y_n\|^2 = \|(x - y_n) - (x - y_m)\|^2 = 2\|x - y_n\|^2 + 2\|x - y_m\|^2 - \|2x - y_n - y_m\|^2$$
$$= 2\|x - y_n\|^2 + 2\|x - y_m\|^2 - 4\|x - \tfrac{1}{2}(y_n + y_m)\|^2 \leqslant 2\|x - y_n\|^2 + 2\|x - y_m\|^2 - 4d^2 \to 0 \text{ as } n, m \to \infty$$

2. Next we show that $x - \mu \in \mathcal{M}^\perp$. Let $y \in \mathcal{M}$ be arbitrary and define $m = \mu - \alpha y$. Then

$$\|x - m\|^2 = \|x - \mu\|^2 + |\alpha|^2\|y\|^2 + 2\Re(\overline{\alpha}\langle x - \mu, y\rangle$$

Assume $\langle x - \mu, y\rangle \neq 0$, write $\langle x - \mu, y\rangle = |\langle x - \mu, y\rangle| e^{i\varphi}$ and choose $\alpha = -|\alpha| e^{i\varphi}$. We get

$$\|x - m\|^2 = \|x - \mu\|^2 + |\alpha|^2\|y\|^2 - 2|\alpha||\langle x - \mu, y\rangle| < d^2$$

if $|\alpha| < 2|\langle x - \mu, y\rangle|\|y\|^{-2}$, a contradiction.

Finally, if $y \in \mathcal{M}$ is s.t. $x - y \in \mathcal{M}^\perp$, then in particular $x - y \perp y - \mu$, hence

$$\|x - \mu\|^2 = \|x - y\|^2 + \|y - \mu\|^2 = d^2 + \|y - \mu\|^2 \Rightarrow y = \mu$$

$\square$

Corollary 22.5.3. *If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then any $x \in \mathcal{H}$ can be uniquely written as $x = m + m^\perp$ with $m \in \mathcal{M}$ and $m^\perp \in \mathcal{M}^\perp$. Hence $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$.*

Definition 22.5.4 (Orthogonal projections)**.** *Let $\mathcal{M}$ be a closed subspace of $\mathcal{H}$ and $\mathcal{M}^\perp$ its orthogonal complement. Let $x = m + m^\perp$ with $m \in \mathcal{M}$ and $m^\perp$ in $\mathcal{M}^\perp$and define*

$$\pi_{\mathcal{M}} x = m; \ \pi_{\mathcal{M}^\perp} x = m^\perp$$

The operator $\pi_{\mathcal{M}}$ is called the orthogonal projection on $\mathcal{M}$.

Proposition 22.5.5. *1. The operator $\pi_{\mathcal{M}}$ is the identity on $\mathcal{M}$, and is idempotent: $(\pi_{\mathcal{M}})^2 = \pi_{\mathcal{M}}$.*
2. Furthermore, $\pi_{\mathcal{M}^\perp}$ is the orthogonal projection on $\mathcal{M}^\perp$ and $(\pi_{\mathcal{M}^\perp})^2 = \pi_{\mathcal{M}^\perp}$.
3. We have $(\mathcal{M}^\perp)^\perp = \mathcal{M}$.

Proof. 1. If $t \in \mathcal{M}$, then the unique decomposition of t in $\mathcal{M} \oplus \mathcal{M}^\perp$ is $t = t + 0$ and thus $\pi_{\mathcal{M}} t = t$. Since, by definition, $\pi_{\mathcal{M}} x \in \mathcal{M}$ for any $x \in \mathcal{H}$, we have $(\pi_{\mathcal{M}})^2 = \pi_{\mathcal{M}}$.

2. The space $\mathcal{M}^\perp$ is also linear and closed, because of the continuity of the scalar product. Now, by the uniqueness of the decomposition $x = m + m^\perp$ and the fact that $m \perp \mathcal{M}^\perp$, Lemma 22.5.2 implies that $m^\perp = \pi_{\mathcal{M}^\perp} x$.

3. Clearly any vector in $\mathcal{M}$ is in $(\mathcal{M}^\perp)^\perp$. Conversely, $x \in (\mathcal{M}^\perp)^\perp \Rightarrow \pi_{\mathcal{M}^\perp} x = 0 \Rightarrow x = \pi_{\mathcal{M}} x \in \mathcal{M}$. □

Corollary 22.5.6. *1. The closure of a subspace $\mathcal{M} \subset \mathcal{H}$ is $\overline{\mathcal{M}} = (\mathcal{M}^\perp)^\perp$.*
If $\mathcal{M}$ is a closed subspace of $\mathcal{H}$, then

$$\pi_{\mathcal{M}} + \pi_{\mathcal{M}^\perp} = I$$

where I is the identity on $\mathcal{H}$.

22.6 Bessel's inequality, Parseval's equality, orthonormal bases

Theorem 22.6.1 (Bessel's inequality)**.** *Let $\{e_i\}_{i \in \mathbb{N}}$ be an orthonormal sequence in $\mathcal{H}$. Then*

$$\sum_{i=1}^{\infty} |\langle x, e_i \rangle|^2 \leqslant \|x\|^2$$

Proof. Let $\mathcal{H}_n = \operatorname{span}(\{e_1, ..., e_n\}) := \{c_1 e_1 + \cdots c_n e_n | c_i \in \mathbb{C}\}$. Clearly, $\mathcal{H}_n$ is a closed subspace of $\mathcal{H}$. We can then write

$$x = \pi_{\mathcal{H}_n} x + x^\perp = \sum_{i=1}^{n} \langle x, e_i \rangle e_i + x^\perp$$

and, by the Pythagorean equality,

$$\|x\|^2 = \|\pi_{\mathcal{H}_n} x\|^2 + \|x^\perp\|^2 \geqslant \|\pi_{\mathcal{H}_n} x\|^2 = \sum_{i=1}^{n} |\langle x, e_i \rangle|^2$$

Since this holds for any n, taking $n \to \infty$, the result follows. □

Corollary 22.6.2. *Let $\{e_\alpha\}_{\alpha\in A}$ be an orthonormal family in $\mathcal{H}$. Then, for any $x \in \mathcal{H}$*

$$\sum_{\alpha\in A} |\langle x, e_\alpha\rangle|^2 := \sup_{\alpha_1,\dots,\alpha_n\in A, n\in\mathbb{N}} \sum_{i=1}^{n} |\langle x, e_{\alpha_n}\rangle|^2 \leqslant \|x\|^2$$

and the set $\{\alpha \in A : \langle x, e_{\alpha_n}\rangle \neq 0\}$ is countable.

Proof. Only countability needs to be shown. It is well known however that an uncountable sum of strictly positive numbers is infinity. □

Definition 22.6.3. *An orthonormal set $\{e_\alpha\}_{\alpha\in A}$ is called* **an orthonormal basis (Hilbert space basis)** *in the Hilbert space $\mathcal{H}$ if any $x \in \mathcal{H}$ can be written as a finite or countable infinite linear combination*

$$x = \sum_{k=1}^{\infty} c_k e_{\alpha_k}$$

Note 22.6.4. 1. *An orthonormal basis is* not *a vector space basis (unless $\mathcal{H}$ is finite-dimensional).*

2. *Using Bessel's inequality, Cauchy-Schwarz and dominated convergence, we see that $c_k = \langle x, e_k\rangle$, hence*
$$x = \sum_{k=1}^{\infty} \langle x, e_{\alpha_k}\rangle e_{\alpha_k} \tag{70}$$

3. *If $\{e_\alpha\}_{\alpha\in A}$ is an orthonormal basis and $\langle x, e_\alpha\rangle = 0$ for all $\alpha \in A$, then $x = 0$.*

Proposition 22.6.5. *Any separable Hilbert space $\mathcal{H}$ has a countable orthonormal basis.*

Proof. Let $\{v_i\}_{i\in\mathbb{N}}$ be a countable dense set in $\mathcal{H}$. The closure of the span of $\{v_i\}_{i\in\mathbb{N}}$ is, of course, $\mathcal{H}$, and so is the span of $\{e_i\}_{i\in\mathbb{N}}$, constructed by Gram-Schmidt. Note that, by Bessel's inequality,

$$\sum_{k=1}^{\infty} |\langle x, e_k\rangle|^2 \leqslant \|x\|^2 \Rightarrow \sum_{k=1}^{\infty} \langle x, e_k\rangle e_k \in \mathcal{H}$$

The difference $x - \sum_{k=1}^{\infty} \langle x, e_k\rangle e_k$ is orthogonal to all the e_k, $k \in \mathbb{N}$, thus, by Note 22.6.4 3, is zero. □

Theorem 22.6.6. *If $\{e_i\}_{i\in\mathbb{N}}$ is an orthonormal set in a separable Hilbert space $\mathcal{H}$, then the following are equivalent:*

a. *(**Completeness**) If $\forall j$, $\langle x, e_j\rangle = 0$, then $x = 0$.*

b. *(**Parseval's identity**) $\forall x \in \mathcal{H}$, $\|x\|^2 = \sum_{k=1}^{\infty} |\langle x, e_k\rangle|^2$.*

c. *$\{e_i\}_{i\in\mathbb{N}}$ is an orthonormal basis for $\mathcal{H}$.*

Proof. (b.⇒ a.) is clear.
(a.⇒ c.) We see that $x - \sum_{k\in\mathbb{N}} \langle x, e_k\rangle e_k$ is orthogonal to all e_j, $j \in \mathbb{N}$, and thus it is zero.
(c.⇒ b.) This is simply the Pythagorean theorem plus the continuity of the norm. □

Exercise 58. *Let $\mathcal{H}$ be a Hilbert space, separable or not, and let $\{e_\alpha\}_{\alpha\in A}$ be an orthonormal set in $\mathcal{H}$. Then, the following statements are equivalent.*

1. (**Completeness condition**) *$\forall \alpha$, $\langle x, e_\alpha \rangle = 0$ holds iff $x = 0$*
2. (**Density condition**) *The span of $\{e_\alpha\}_{\alpha \in A}$ is dense in $\mathcal{H}$.*
3. (**Orthonormal basis condition**) *For any $x \in \mathcal{H}$, $\langle x, e_\alpha \rangle = 0$ except for a countable set $(e_{\alpha_k})_{k \in \mathbb{N}}$ and*
$$x = \sum_{k \in \mathbb{N}} \langle x, e_{\alpha_k} \rangle e_{\alpha_k}$$
4. (**Maximality condition**) *If $\{e'_\beta\}_{\beta \in B}$ is an orthonormal set in $\mathcal{H}$ which contains $\{e_\alpha\}_{\alpha \in A}$, then $\{e'_\beta\}_{\beta \in B} = \{e_\alpha\}_{\alpha \in A}$.*
5. (**Parseval's identity condition**) *$\forall x \in \mathcal{H}$, $\|x\|^2 = \sum_{\alpha \in A} |\langle x, e_\alpha \rangle|^2$.*

Exercise 59. Show that the $S = \{1, x, x^2, ...\}$ is a linearly independent set in $\mathcal{H} = L^2([-1,1])$ whose span is dense in $\mathcal{H}$. Thus Gram-Schmidt produces an orthonormal system of polynomials P_n out of S. ($\sqrt{\frac{2}{2n+1}} P_n$ are the Legendre polynomials.) Thus, any $f \in \mathcal{H}$ can be written as $f = \sum_{k \in \mathbb{N}} c_k P_k$. Show that, although as mentioned, the span of S is dense in $\mathcal{H}$, the set $\{f \in \mathcal{H} : f = \sum_{k \in \mathbb{N}} c_k x^k\}$ is a strict subspace of $\mathcal{H}$. Is it closed? Can you identify it?

Note 22.6.7. Nonseparable Hilbert spaces rarely occur in applications. A prototypical example is

$$\ell^2(A) := \left\{ f : A \to \mathbb{C} \,\middle|\, \sum_{\alpha \in A} |f(\alpha)|^2 < \infty \right\}$$

when A is not countable.

Also, Corollary 22.6.2 shows that *even in non-separable Hilbert spaces* we only need a countable family *at a time*.

Theorem 22.6.8. *In a Hilbert space $\mathcal{H}$, any orthonormal set S is contained in an orthonormal basis for $\mathcal{H}$.*

Proof. Let $\mathcal{E}$ be the family of all orthonormal sets containing S ordered by inclusion. If $\mathcal{C}$ is a chain in $\mathcal{E}$, then it has a maximal element, namely the union of the sets in $\mathcal{C}$ as it is easily verified. Now, Zorn's Lemma implies that $\mathcal{E}$ has a maximal element, which by Exercise 58 4, is a basis for $\mathcal{H}$. $\square$

An example of a Hilbert basis in ℓ^2 is the set $e_k = (0, .., 1, 0...)$, with 1 in the kth position.

Definition 22.6.9. *Let $\mathcal{H}_1, \mathcal{H}_2$ be Hilbert spaces and $U : \mathcal{H}_1 \to \mathcal{H}_2$ be linear and norm preserving, that is $\|Ux\|_2 = \|x\|_1$ for all $x \in \mathcal{H}_1$. Then U is called an isometry.*

Let $\mathcal{H}_1, \mathcal{H}_2$ be Hilbert spaces and $U : \mathcal{H}_1 \to \mathcal{H}_2$ be linear, inner product preserving, $\langle Ux, Uy \rangle = \langle x, y \rangle$, and onto. Then U is called unitary.

Proposition 22.6.10. *U is unitary iff it is an isometry and onto.*

Note 22.6.11. *Unitary maps are isomorphisms, w.r.t the structure of a Hilbert space.*

Proof. If U is unitary, then $\|Ux\|^2 = \langle Ux, Ux \rangle = \langle x, x \rangle = \|x\|^2$. Conversely, the polarization identity shows that any isometry preserves the inner product. $\square$

Proposition 22.6.12 (Any two separable Hilbert spaces are isomorphic)**.** *Any separable Hilbert space $\mathcal{H}$ is isomorphic to ℓ^2.*

Proof. Let $\{e_n\}_{n\in\mathbb{N}}$ be an orthonormal basis in $\mathcal{H}$. Define $U : \mathcal{H} \to \ell^2$ by $U(x) = (\langle x, e_1\rangle, ..., \langle x, e_n\rangle, ...)$ and check that this is an isometry. □

For a nonseparable Hilbert space with a Hilbert basis $\{e_\alpha\}_{\alpha\in A}$, a similar statement holds, except $\ell^2 = \ell^2(\mathbb{N})$ is replaced by the more general $\ell^2(A)$, for an adequate set A.

23 Normed vector spaces

Definition 23.0.1.

1. *A vector space V endowed with a norm $\|\cdot\|$ is called a* **normed vector space***.*
2. *Two norms $\|\|_1, \|\|_2$ on the vector space V are* **equivalent** *if there exist two positive constants c_1 c_2 s.t. $\forall v \in V$, $c_1\|v\|_1 \leqslant \|v\|_2 \leqslant c_2\|v\|_1$.*
3. *A* **Banach space** *is a normed space which is complete w.r.t the norm topology, that is the distance between x, y is $\|x - y\|$.*
4. *A series $\sum_{n\in\mathbb{N}} v_n$ of vectors in a normed space is* **absolutely convergent** *if $\sum_{n\in\mathbb{N}} \|v_n\|$ converges.*

Proposition 23.0.2. *An absolutely convergent series $\sum_{n\in\mathbb{N}} v_n$ is Cauchy. In the opposite direction, if $\sum_{n\in\mathbb{N}} v_n$ is Cauchy, then there exists a strictly increasing sequence $(n_i)_{i\in\mathbb{N}}$ in $\mathbb{N}$ s.t. $n_1 = 1$ and s.t., with $w_i = \sum\limits_{n_i\leqslant j<n_{i+1}} v_j$, the series $\sum\limits_{i\in\mathbb{N}} w_i$ is absolutely convergent.*

Proof. Assume $\sum\limits_{n\in\mathbb{N}} v_n$ is absolutely convergent, and let $\varepsilon > 0$. Then, the series of norms $\sum\limits_{n\in\mathbb{N}} \|v_n\|$ is Cauchy and there is an n_0 s.t. for all $m \geqslant n \geqslant n_0$ we have

$$\left\|\sum_{j=n}^{m} v_j\right\| \leqslant \sum_{j=n}^{m} \|v_j\| < \varepsilon$$

Hence $\sum\limits_{n\in\mathbb{N}} v_n$ is Cauchy.

In the opposite direction, assume $\sum_{n\in\mathbb{N}} v_n$ is Cauchy. Choose $\varepsilon_i = 2^{-i}, i \in \mathbb{N}$, let $n_1 = 1$ and, inductively for $i > 1$, define $n_i > n_{i-1}$ so that $\forall n \geqslant m \geqslant n_i$ we have $\left|\sum\limits_{m\leqslant j\leqslant n} v_i\right| \leqslant \varepsilon_i$. Defining $w_i = \sum\limits_{n_i\leqslant j<n_{i+1}} v_j$, the result follows. □

Theorem 23.0.3. *A normed vector space V is complete iff every absolutely convergent series in V converges.*

Proof. Note first that, in a linear space, every Cauchy *sequence* converges iff every Cauchy *series* converges. Let $\sum_{n\in\mathbb{N}} v_n$ be Cauchy in V. With the construction of Proposition 23.0.2, the series $w_i = \sum\limits_{n_i\leqslant j<n_{i+1}} v_j$ is absolutely convergent, thus convergent, to some $v \in V$. Then, for any integer $m \in [n_i, n_{i+1})$, $\left\|v - \sum\limits_{j\geqslant m} v_j\right\| \leqslant \left\|\sum\limits_{m\leqslant j<n_i} v_j\right\| + \left\|v - \sum\limits_{k\geqslant i} w_i\right\|$ hence $\sum\limits_{n\in\mathbb{N}} v_n$ also converges to v. □

Proposition 23.0.4. *1. If V_1, V_2 are normed vector spaces, then the product space $V_1 \times V_2$ is a normed vector space under the product norm defined, for $v_i \in V_i$, by $\|(v_1, v_2)\| := \|v_1\|_1 + \|v_2\|_2$.*

2. If V' is a closed linear subspace of V, then the quotient space V/V' is a normed vector space under the quotient norm

$$\|v + V'\| := \inf_{v' \in V'} \{\|v + v'\|\} \tag{71}$$

Proof. This is easy to check. □

Note 23.0.5. *Recall that all norms in $\mathbb{C}^n$ are equivalent. Therefore, the product norm is equivalent to many other choices, s.a.* $\max\{\|\|_1, \|\|_2\}$.

23.1 Functionals and linear operators

Definition 23.1.1. *1. A* **linear operator** *(or map) between two vector spaces V_1, V_2 over the same scalar field is a function $L : V_1 \to V_2$ which satisfies $L(ax + by) = aLx + bLy$ for all $x, y \in V_1$ and all scalars a, b.*

2. A linear operator having the scalar field as the target space is called **linear functional**.

3. An operator $L : V_1 \to V_2$ between two normed vector spaces V_1, V_2 is called **bounded** *if there exists a constant $C \in [0, \infty)$ s.t., for all $v \in V_1$ we have*

$$\|Lv\|_2 \leqslant C\|v\|_1 \tag{72}$$

4. If V_1, V_2 are normed vector spaces, then $\mathcal{L}(V_1, V_2)$ denotes the space of linear bounded operators from V_1 to V_2.

5. A **Banach algebra** *is a Banach space which is an algebra for which the norm of the product is bounded by the product of the norms, that is, $\|xy\| \leq \|x\| \, \|y\|$.*

6. If X is a normed vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, then the space of its bounded linear functionals $X^* := \mathcal{L}(X, \mathbb{K})$ is the very important **dual of** *X.*

Note 23.1.2. *1. $L \in \mathcal{L}(V_1, V_2)$ iff L is linear from V_1 to V_2 and*

$$\|L\| := \sup_{\|v\|_1 = 1} \|Lv\|_2 < \infty \tag{73}$$

The quantity $\|L\|$ is called the **norm of the linear map** *L.*

2. $\mathcal{L}(V_1, V_2)$ is a normed space with the operator norm.

Proposition 23.1.3. *Let Y be a complete normed space and X a normed space. Then:*
a) $\mathcal{L}(X, Y)$ is a complete normed space, and
b) $\mathcal{L}(Y, Y)$ with the operator norm is a Banach algebra.

Proof. a) If $(T_n)_n$ is a Cauchy sequence in $\mathcal{L}(X, Y)$, then for any $x \in X$ $(T_n x)_n$ is Cauchy in Y, thus convergent. Now, $Tx = \lim_n T_n x$ defines a linear operator $T \in \mathcal{L}(X, Y)$, since it is easy to check that $\|T - T_n\| \to 0$ and $\|T_n\| \to \|T\|$ as $n \to \infty$.

b) Let $T_1, T_2 \in \mathcal{L}(Y, Y)$. We have

$$\|T_1T_2y\| = \|T_1(T_2y)\| \leqslant \|T_1\|\|T_2y\| \leqslant \|T_1\|\|T_2\|\|y\|; \text{ thus } \|T_1T_2\| \leqslant \|T_1\|\|T_2\|$$

The result follows from a). $\square$

HOMEWORK, DUE JAN 22, 2019 (TUE)

Folland: 7,9,10,16 (pp.155–156) and turn in:

Exercise 60 (Recitation exercise). Let X be Hausdorff. Prove that X^* (the one-point compactification of X) is Hausdorff iff X is locally compact (LCH).

Exercise 61.

1. Let $\mathcal{H}$ be a Hilbert space, M a closed subspace of $\mathcal{H}$ and $P = \pi_M$ the orthogonal projection on M. Show that P is bounded and that for all x, y in $\mathcal{H}$ we have

$$\langle Px, y\rangle = \langle x, Py\rangle \quad (*)$$

For a bounded operator, the symmetry property above is called self-adjointness.

2. Recalling that orthogonal projections are also idempotent ($P^2 = P$), prove the following converse: Let P be a bounded operator from $\mathcal{H}$ to itself which is self-adjoint (that is satisfies (*)) and idempotent. Show that there is a closed subspace M of $\mathcal{H}$ such that $P = \pi_M$.

23.2 The Hahn-Banach theorem

This is a fundamental theorem that guarantees the existence of extensions of bounded linear functionals defined on subspaces of a given normed linear space.

Let X be a normed space over $\mathbb{K} = \mathbb{R}$ *or* $\mathbb{C}$. Then, since $\mathbb{K}$ is complete, X^* is a Banach space. Let $\varphi \in X^*$.

Assume first $\mathbb{K} = \mathbb{C}$. If we write $\varphi(x) = u(x) + iv(x)$ where u, v are real-valued, then u, v are linear. Note that

$$i\varphi(x) = \varphi(ix) = u(ix) + iv(ix) = iu(x) - v(x)$$

hence

$$u(ix) = -v(x); \quad v(ix) = u(x); \quad \varphi(x) = u(x) - iu(ix) \tag{74}$$

Proposition 23.2.1. *Let X be a normed space and $\varphi \in \mathcal{L}(X, \mathbb{C})$. If $\varphi = u + iv$, where u, v are real-valued, then $u, v \in \mathcal{L}(X, \mathbb{R})$ with $\|u\| = \|v\| = \|\varphi\|$.*

Proof. For any $x \in X$ we have $u(x) = \Re\varphi(x)$ and thus $|u(x)| \leqslant |\varphi(x)|$ implying $\|u\| \leqslant \|\varphi\|$. In the opposite direction, let $\varepsilon > 0$. Choose x s.t. $\|\varphi(x)\| \geqslant (1-\varepsilon)\|\varphi\|\|x\|$, and write $\varphi(x) = |\varphi(x)|e^{i\theta}$. Then, $\varphi(e^{-i\theta}x) = |\varphi(x)| = u(xe^{-i\theta})$. Since ε is arbitrary, it follows that $\|u\| \geqslant \|\varphi\|$. $\square$

Definition 23.2.2. *A* **sublinear functional** *p on X is a map from X to $\mathbb{R}$ s.t. for all $x, y \in X$ and $\lambda \geqslant 0$,*

$$p(x+y) \leqslant p(x) + p(y) \text{ and } p(\lambda x) = \lambda p(x)$$

Norms and seminorms are examples of sublinear functionals.

Theorem 23.2.3 (The Hahn-Banach theorem)**.** *Let X be a vector space over $\mathbb{R}$ and $M \subset X$ a subspace. Let φ be a linear functional on M s.t. $\varphi(m) \leqslant p(m)\ \forall m \in M$, for some sublinear functional p defined on X. Then, there exists a linear functional $\Phi : X \to \mathbb{R}$ which extends φ s.t. $\Phi(x) \leqslant p(x)\ \forall x \in X$.*

The extension is highly non-unique, in general.

Proof. We first show that, given an $x \in X \setminus M$, there is an extension Φ of φ to $\mathrm{span}(M \cup \{x\})$. For this we define a $\Phi(x)$ and, for $t \in \mathbb{R}$ and $m \in M$, let $\Phi(m + tx) = \varphi(m) + t\Phi(x)$ where we need to arrange

$$\Phi(m + tx) \leqslant p(m + tx)$$

for all real t. Definitely this holds when $t = 0$. For $t \neq 0$, writing $\sigma = \mathrm{sgn}(t)$, we have

$$\Phi(m + tx) = |t| \left(\varphi(m') + \sigma\Phi(x) \right)$$

(where $m' = m/|t|$) and we need to arrange

$$\sup_{s>0, m \in M} \frac{\varphi(m) - p(m - sx)}{s} \leqslant \Phi(x) \leqslant \inf_{t>0, m' \in M} \frac{p(m' + tx) - \varphi(m')}{t}$$

Clearly, this is possible iff the inf on the right side is no less than the sup on the left side, which in turn holds if, for all $m \in M, s, t > 0$, we have

$$\frac{\varphi(m) - p(m - sx)}{s} \leqslant \frac{p(m + tx) - \varphi(m)}{t}$$

which can be rewritten as

$$\varphi((s + t)m) \leqslant p(sm + stx) + p(tm - stx)$$

which holds since

$$p((s + t)m) = p(sm + stx + tm - stx) \leqslant p(sm + stx) + p(tm - stx)$$

The rest of the proof is just a straightforward application of Zorn's lemma: order the functions Φ with the required properties (thought of as sets pairs of points) by set inclusion, and note that the union of a chain of functions Φ is again a function with the required properties. □

Theorem 23.2.4 (The Hahn-Banach theorem, complex version). *Let X be a vector space over $\mathbb{C}$ and $M \subset X$ a subspace. Let φ be a linear functional on M s.t. $|\varphi(m)| \leqslant p(m)\ \forall m \in M$, for some* **seminorm** *on X. Then, there exists a linear functional $\Phi : X \to \mathbb{R}$ which extends φ s.t. $|\Phi(x)| \leqslant p(x)\ \forall x \in X$.*

Proof. A simple exercise, using the third equality in (74). □

Some important consequences of the complex Hahn-Banach theorem are in characterizing the duals of normed vector spaces.

Theorem 23.2.5. *Let X be a normed vector space.*

1. *If M is a closed subspace of X and $x \notin M$, then there exists a functional $\varphi \in X^*$ s.t. $\|\varphi\| = 1$ and $\varphi(x) = dist(x, M) := \inf_{m \in M} \|x - m\|$.*

2. *For any $x \neq 0$ in X there is a $\varphi \in X^*$ s.t. $\|\varphi\| = 1$ and $\varphi(x) = \|x\|$.*

3. *The functionals in X^* separate the points of X.*

4. *(The double dual, X^{**}.) The points $x \in X$ induce linear functionals $\hat{x}$ on X^* by*

$$\hat{x}(\varphi) = \varphi(x)$$

The map $x \mapsto \hat{x}$ is a linear isometry from X into (possibly a subspace of) X^{**}.

Proof. 1. Let M' be the space generated by M and x and define, for $y = m + tx \in M'$, $\varphi(y) = t\text{dist}(x, M)$. Now we simply check that Hahn-Banach applies to extend φ from M' to the whole of X.

2. Follows from 1, taking $M = \{0\}$.

3. If $x \neq y$, then $z = x - y \neq 0$ and the result follows from 2.

4. It is clear that the functional $\hat{x}$ is linear. Now,

$$|\hat{x}(\varphi)| = |\varphi(x)| \leqslant \|\varphi\| \|x\|$$

thus $\|\hat{x}\| \leqslant \|x\|$, while 2. above implies $\|\hat{x}\| \geqslant \|x\|$.

□

Definition 23.2.6. 1. *(**the weak topology**) Let X be a normed vector space. The weak topology on X is the topology induced by X^* on X, defined as the coarsest topology s.t. all elements of X^* are continuous. Equivalently, a net $(x_\alpha)_{\alpha \in A}$ in X converges weakly to x iff $\forall \varphi \in X^*$, the net of complex numbers $(\varphi(x_\alpha))_{\alpha \in A}$ converges to $\varphi(x)$.*

2. *(**the weak* topology**) This is a topology on X^* which is weaker than the weak topology on X^*. This is the topology of pointwise convergence: $(\varphi_\alpha)_{\alpha \in A}$ converges to φ iff $(\varphi_\alpha(x))_{\alpha \in A}$ converges to $\varphi(x)$ for **for any x** in X.*

Three topologies play an important role on bounded linear operators between two normed spaces. The finest is the operator norm topology, $T_n \to T$ iff $\|T_n - T\| \to 0$, and the following two.

Definition 23.2.7. *Let X, Y be Banach spaces and let $(T_\alpha)_{\alpha \in A}$ be a generalized sequence of operators.*

1. *$(T_\alpha)_{\alpha \in A}$ converges to T in the **strong operator topology** iff, for all $x \in X$, $(T_\alpha x)_{\alpha \in A}$ converges to Tx in the norm of Y.*

2. *$(T_\alpha)_{\alpha \in A}$ converges to T in the **the weak operator topology** iff, for all $x \in X$, $(T_\alpha x)_{\alpha \in A}$ converges to Tx in the weak topology on Y.*

An important result about weak* topology is the weak* compactness of the closed unit ball in X^*:

Theorem 23.2.8 (Banach-Alaoglu). *If X is a normed vector space, the closed unit ball $B^* = \{\varphi \in X^* | \|\varphi\| \leqslant 1$ is compact in the weak* topology.*

Proof. This is a quite straightforward consequence of Tychonoff's theorem. The set of all complex-valued functions on X s.t. for all $x \in X$ $|f(x)| \leqslant \|x\|$ is the set

$$F = \prod_{x \in X} F_x; F_x := \{z \in \mathbb{C} | |z| \leqslant \|x\|\}$$

By Tychonoff's theorem, F is compact in the topology of pointwise convergence, the same as the weak* topology on the subset B^* of F consisting of *linear* functionals. Thus the statement is equivalent to saying that B^* is a closed subset of F, which in turn is the same as saying that linearity is preserved when taking limits of convergent nets, which is immediate. $\square$

Recall that dual spaces are always complete. Given the natural embedding $\hat{X}$ of X in X^{**} given by 4., the closure of $\hat{X}$ is identified with the completion of X.

Definition 23.2.9. *X is* **reflexive** *if $X^{**} = \hat{X}$.*

An important example of a reflexive space is a Hilbert space as follows from the following theorem.

23.3 The Riesz representation theorem

Let $\mathcal{H}$ be a Hilbert space, say over $\mathbb{C}$, and $y \in \mathcal{H}$. The function $x \mapsto \langle x, y\rangle$ from $\mathcal{H}$ to $\mathbb{C}$ is a continuous linear functional on $\mathcal{H}$. The converse is an important result.

Proposition 23.3.1 (The Riesz representation theorem)**.** *If Λ is a continuous linear functional from $\mathcal{H}$ to $\mathbb{C}$, then there is a unique $y \in \mathcal{H}$ s.t.*

$$\forall x \in \mathcal{H}, \ \Lambda x = \langle x, y\rangle \tag{75}$$

In particular, $\mathcal{H}$ is **isomorphic to its dual**, $\mathcal{H}^*$

Proof. Uniqueness follows from the fact that $\langle x, y\rangle = \langle x, y'\rangle$ for all $x \in \mathcal{H}$ iff $y = y'$.

Existence of a y: Let $\mathcal{M} = \{x \in \mathcal{H} : \Lambda x = 0\}$. Clearly $\mathcal{M}$ is a closed linear subspace of $\mathcal{H}$. Now if $\mathcal{M} = \mathcal{H}$ then 0 is the only y s.t. (75) holds, and we are done. Otherwise, we claim that $\mathcal{M}^\perp$ is one dimensional. Indeed, let $0 \neq e \in \mathcal{M}^\perp$; note that this implies $\Lambda e \neq 0$. We rescale e so that $\Lambda e = 1$. Let $0 \neq x \in \mathcal{M}^\perp$ and let $\Lambda x = b$ (again, necessarily $b \neq 0$). Then

$$x - be \in \mathcal{M}^\perp \text{ and } x - be \in \mathcal{M} \ \ (\text{since } \Lambda(x - be) = 0) \ \Rightarrow x - be = 0$$

This means x is linearly dependent on e, and $\mathcal{M}^\perp$ is one-dimensional. Let $y = \frac{e}{\|e\|^2}$. For $x \in \mathcal{H}$ we have

$$x = \pi_{\mathcal{M}} x + \pi_e x = \pi_{\mathcal{M}} x + \frac{\langle x, e\rangle}{\|e\|^2} e; \text{ hence } \Lambda x = \langle x, y\rangle$$

and it follows that, for all $x \in \mathcal{H}$, $|\Lambda x| \leqslant \dfrac{\|x\|}{\|e\|}$ hence

$$\|\Lambda\| \leqslant \frac{1}{\|e\|}$$

For the norm, we note the inequality above and the fact that by definition $\dfrac{|\Lambda e|}{\|e\|} = \dfrac{1}{\|e\|}$, and thus $\|\Lambda\| = \frac{1}{\|e\|} = \|y\|$ $\square$

23.4 Adjoints

Let $A : \mathcal{H} \to \mathcal{H}$ be a bounded operator. Its adjoint is defined as the operator A^* with the property

$$\forall x, y \in \mathcal{H}, \langle Ax, y\rangle = \langle x, A^* y\rangle$$

Exercise 62. *A. Use the Riesz representation theorem to show that A^* defined above exists and is unique. Check that $A^{**} = A$.*

B. For fixed y, $\langle x, A^* y\rangle$ is a linear functional, and by the Riesz representation theorem

$$\|A^* y\| = \sup_{\|u\|=1} |\langle u, A^* y\rangle| = \sup_{\|u\|=1} |\langle Au, y\rangle| \leqslant \|A\|\|y\|$$

This implies $\|A^*\| \leqslant \|A\|$. Then $\|A\| = \|A^{**}\| \leqslant \|A^*\|$, hence $\|A\| = \|A^*\|$. Thus A^* is a bounded operator with the same norm as A. Check that $\|A^* A\| = \|AA^*\| = \|A\|^2$.

Note 23.4.1.
1. *Let $\mathcal{H}$ be a Hilbert space. Let $(e_n)_{n\in\mathbb{N}}$ be a countably infinite orthonormal basis in $\mathcal{H}$. Check that the sequence $(e_n)_{n\in\mathbb{N}}$ converges weakly to zero, but is has no norm-convergent subsequence. Check that the weak topology on a Hilbert space $\mathcal{H}$ is the same as the norm topology iff $\mathcal{H}$ is finite-dimensional.*

2. *It can be shown that if $\mathcal{B}$ is a Banach space (and more generally, in fact) and the weak topology on X^* coincides with the weak* topology on X^*, then $\mathcal{B}$ is reflexive.*

24 Consequences of the Baire category theorem

Definition 24.0.1. *A Baire space topological space such that every intersection of a countable collection of open dense sets in the space is also dense.*

As a reminder, the Baire category theorem states

Theorem 24.0.2 (Baire category theorem, BCT).
1. Every complete metric space is a Baire space. Equivalently, a non-empty complete metric space is **not** a countable union of nowhere-dense sets [equivalently, nowhere-dense closed sets].

2. Every locally compact Hausdorff space is a Baire space.

This theorem has a number of fundamental consequences in analysis. In the following, we use the notations for open balls in a normed space:

$$B_a(x) = \{y \in X : \|y - x\| < a\}; \quad B_a(0) \equiv B_a$$

Theorem 24.0.3 (Uniform boundedness principle).
1. *Assume X is a Banach space, Y is a normed space, and $A \subset \mathcal{L}(X, Y)$. Then*

$$\left(\forall x \in X, \sup_{T\in A} \|Tx\| < \infty\right) \Leftrightarrow \sup_{T\in A} \|T\| < \infty$$

2. *(Generalization) If X and Y are normed spaces and there is a non-meager set $X_1 \subset X$ such that $(\forall x \in X_1, \sup\limits_{T\in A} \|Tx\| < \infty)$, then $\sup_{T\in A} \|T\| < \infty$.*

Proof. 1. ($\Leftarrow$) is trivial. ($\Rightarrow$) For $n \in \mathbb{N}$ let

$$E_n = \{x \in X : \sup_{T\in A} \|Tx\| \leqslant n\} = \cap_{T\in A}\{x \in X : \forall T \in A, \|Tx\| \leqslant n\}$$

Clearly, E_n are closed and $X = \bigcup\limits_{n\in\mathbb{N}} E_n$. Then, there is an m s.t. E_m has nonempty interior: $\exists a, x_0$ s.t. $\overline{B_a}(x_0) \subset E_m$. Take any $u \in X$ with $\|u\| = 1$. Then both x_0 and $x_0 + au$ are in $\overline{B_a}(x_0)$ and

$$Tu = \tfrac{1}{a}T(au) = \tfrac{1}{a}T(x_0 + au - x_0)$$

$$\text{hence } \|Tu\| \leqslant \tfrac{1}{a}\|T(x_0 + au)\| + \tfrac{1}{a}\|Tx_0\| \leqslant \frac{2m}{a} \text{ and thus } \sup_{\substack{\|u\|=1\\ T\in A}} \|Tu\| \leqslant \frac{2m}{a}$$

2. Copy the proof above, basically. □

Theorem 24.0.4 (The open mapping theorem). *Let X, Y be Banach spaces and $T \in \mathcal{L}(X, Y)$ be* **surjective**. *Then if $\mathcal{O}$ is open in X, $T(\mathcal{O})$ is open in Y.*

Proof *(*an adaptation of Reed-Simon p. 82*).* We start with some straightforward preparatory steps to reduce the complexity of the more difficult part of the proof.

a) It is enough to prove that for any x and $\mathcal{N}_x$ a neighborhood of it, $T(\mathcal{N}_x)$ is a neighborhood of $T(x)$.

b) Since, by linearity, $\forall y, \mathcal{N}_y$ we have

$$T(y + \mathcal{N}_y) = T(y) + T(\mathcal{N}_y)$$

it suffices to prove that $\exists\mathcal{N}_0$ a neighborhood of $0 \in X$ s.t. $T(\mathcal{N}_0)$ is a neighborhood of $0 \in Y$. Note also the scaling property

$$T(B_r) = rT(B_1)$$

c) Clearly b) holds if there exist $r, r' > 0$ s.t.

$$T(B_r^X) \supset B_{r'}^Y$$

From now on we will omit the superscripts X but keep the superscripts Y.

d) Again by linearity it is enough to show that for some r, $T(B_r)$ contains *some ball*, not necessarily centered at zero, that is, $T(B_r)$ has nonempty interior.

Now, since $(\forall y \in Y)(\exists x \in X)(y = Tx)$ (and clearly $x \in T(B_n)$ for some n) we must have

$$Y = \bigcup_{n=1}^{\infty} T(B_n) \subseteq \bigcup_{n=1}^{\infty} \overline{T(B_n)}$$

Hence, by BCT $(\exists n \in \mathbb{N})(\overline{T(B_n)}^{\circ} \neq \emptyset)$ [12]. By linearity, this happens for all n: there exist

[12] A° is as usual the interior of A.

$\varepsilon > 0, y \in Y$ s.t.

$$\overline{T(B_{1/2})} \supset B^Y_{2\varepsilon}(y_0) \tag{76}$$

Finally, for any $y \in B^Y_\varepsilon$, y_0 and $y_0 + y \in B^Y_{2\varepsilon}(y_0)$, so $y \in \overline{T(B_{1/2}) + T(B_{1/2})} \subset \overline{T(B_1)}$. Hence

$$\overline{T(B_1)} \supset B^Y_\varepsilon \tag{77}$$

What we really need however is nontrivially stronger [13], namely that,

$$\exists n, T(B_n)^\circ \neq \emptyset$$

which follows from the lemma below.

Lemma 24.0.5. *Let $T \in \mathcal{L}(X,Y)$. If $\overline{T(B_1)}$ contains a ball B^Y_ε, then $\overline{T(B_1)} \subset T(B_2)$.*

(In fact $\overline{T(B_1)} \subset T(B_{1+\delta})$ for any $\delta > 0$.)

Proof. Let $y \in \overline{T(B_1)}$ and ε as above. There are points x in B_1 s.t. $T(x)$ is arbitrarily close to y. Let $x_1 \in B_1$ be s.t. $y - T(x_1) \in B^Y_{\varepsilon/2} \subset \overline{T(B_{1/2})}$ (by scaling). Now let $x_2 \in B_{1/2}$ be s.t. $(y - T(x_1)) - T(x_2) \in B^Y_{\varepsilon/4}$ and, inductively, let $x_{n+1} \in B_{1/2^n}$ be s.t. $y - T(x_1) - \cdots - T(x_{n+1}) \in B^Y_{\varepsilon/2^n}$. But you see that $x = \sum_n x_n$ converges to an element in $\overline{B_{1/2+1/4+\ldots}} \subset B_2$, and by continuity $y = Tx$, thus $y \in B^Y_2$. (By modifying the selection of $\{x_n\}$, you can prove the result above with $1+\delta$ instead of 2.) □

□

Theorem 24.0.6 (Inverse mapping theorem). *If $T \in \mathcal{L}(X,Y)$ is one to one onto, then T^{-1} is also continuous, $T^{-1} \in \mathcal{L}(Y,X)$ (continuous linear bijections are bicontinuous).*

Proof. T is one-to-one, thus onto, thus open, implying by definition continuity of T^{-1}. □

Definition 24.0.7. *If $T \in \mathcal{L}(X,Y)$, its* **graph** *is*

$$\Gamma(T) = \{(x,y) \in X \times Y : y = Tx\} = \{(x,Tx) : x \in X\}$$

Theorem 24.0.8 (Closed graph theorem). *Let X, Y be Banach spaces and $T : X \to Y$ be linear. Then $T \in \mathcal{L}(X,Y)$* **iff** *$\Gamma(T)$ is a closed subset of $X \times Y$ w.r.t the product norm.*

(Note that here we do not assume injectivity or surjectivity.)

Proof. Assume T is continuous. If $\{(x_n, T(x_n))\}_{n\in\mathbb{N}}$ converges to (x,y) then $x_n \to x$ and $Tx_n \to y$. By continuity however, $T(x_n) \to Tx$, thus $y = Tx$ and $(x,y) \in \Gamma(T)$.

In the opposite direction, since $\Gamma(T)$ is a linear closed subspace of $X \times Y$, it is itself a Banach space. Recall that the canonical projections $\pi_1 : \Gamma(T) \to X, \pi_2 : \Gamma(X) \to Y$ are continuous. Note that $\pi_1(x, Tx) = x$ is a linear bijection between X and $\Gamma(T)$, thus, by the inverse mapping theorem, its inverse is continuous too. To finish the proof we simply note that $T = \pi_2(\pi_1^{-1})$ is a composition of continuous functions. □

[13] Note that $T(\overline{B_1}) \subset \overline{T(B_1)}$, but the inclusion can be strict as Exercise 64 below shows. BCT also implies that not all $T(B_n)$ are nowhere dense, but this means by definition that the closure of some $T(B_n)$ has nonempty interior, which is the same as above.

Note 24.0.9. *We see that a bounded operator between Banach spaces can fail to have a bounded inverse only for the trivial reason that it does not have an inverse* **at all** *(that is, if it is not surjective or not injective).*

Definition 24.0.10. 1. More generally, if X, Y are Banach spaces and $D(A) \subset X$, then an operator $A : D(A) \to Y$ is called **closed** if $\Gamma(A)$ is closed.

2. If, in 1. above, $X = Y = \mathcal{H}$ is a Hilbert space, $A : D(A) \to \mathcal{H}$ is called **symmetric, or formally self-adjoint**, if for all $x, y \in D(A)$ we have $\langle Ax, y\rangle = \langle x, Ay\rangle$.

3. If $D(A)$ is dense in $\mathcal{H}$, the **domain of the adjoint** is defined as

$$D(A^*) = \{y \in \mathcal{H} : \exists z \in \mathcal{H} \; s.t \; \forall x \in D(A), \langle Ax, y\rangle = \langle x, z\rangle\}$$

and we write $z = A^*y$ (this z is unique) and $A^* : D(A^*) \to \mathcal{H}$ is called the **adjoint** of A.

4. An operator as in 3. above is called **self-adjoint** if $A = A^*$ (meaning A is symmetric, and $D(A) = D(A^*)$.)

Exercise 63. (A) Show that the definition of A^* in 3. above is correct (i.e. z is indeed unique), and that $A^* : D(A^*) \to \mathcal{H}$ is linear.

(B) Show that differentiation, $\partial := f \mapsto f'$ defined on

$$D(\partial) = \{g \in AC([0, \infty)) \cap L^2(\mathbb{R}^+) : g(0) = 0\}$$

is closed, but not bounded.

(C) Show that $p := -i\partial$ defined on $D(\partial)$ above is symmetric but not self-adjoint.

Exercise 64. a) Let $\mathcal{H} = L^2[0, 1]$. Then the operator A defined on $\mathcal{H}$ by

$$(Af)(x) = \int_0^x f(s)ds$$

is bounded (check).

b) Show that $M =$ran$(A) \neq \mathcal{H}$.

c) Show that (the linear space) ran(A) is **not** closed. (This implies A is not invertible from $\mathcal{H}$ to $\mathcal{H}$.) Note that $\Gamma(A)$ is **closed** nevertheless. ($\Gamma(A)$ is a closed subspace of $\mathcal{H} \times \mathcal{H}$ which does not mean that the *direct images* of the projections $\pi_{1,2}$ of $\Gamma(A)$ are closed!)

d) What is $M^{\perp}$?

e) Show that $A(\mathcal{H})^\circ = \emptyset$ while $\overline{A(\mathcal{H})} = \mathcal{H}$.

Corollary 24.0.11. *Corollary 1: Let $\|\cdot\|_1$ and $\|\cdot\|_2$ be two norms defined on X. Assume X is a Banach space in both norms, and that furthermore, for some $C_1 > 0$ and all $x \in X$ we have $\|x\|_1 \leqslant C_1\|x\|_2$. Then the two norms are equivalent, that is, there is a $C_2 > 0$ s.t. for all $x \in X$, $\|x\|_2 \leqslant C_2\|x\|_1$.*

Proof. Exercise. (Hint: take $T = I$, that is, $Tx = x$ for all x.) $\square$

Homework, due Jan 30, 2019 (Wed)
Folland: 28,30,34,37 (pp.164–165) and turn in Exercises 63 and 64 in the notes.

25 L^p spaces

$L^p, p \in [1, \infty]$ spaces play a central role in all branches of analysis.

Definition 25.0.1. *Let $(X, \mathcal{M}, \mu)$ be a measure space. For f measurable on X and $p \in \mathbb{R}^+$ define*

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{1/p}$$

We will show shortly that $\| \|_p$ is a norm iff $p \geqslant 1$. We first check that the triangle inequality *fails* for $p \in (0,1)$

Proposition 25.0.2. *[Some elementary identities]*

1. *If $a, b > 0$ and $p \in (0, \infty)$, then*

$$(a^p + b^p)^{1/p} \leqslant a + b \quad \text{if} \quad p \geqslant 1 \quad \text{and} \quad (a^p + b^p)^{1/p} > a + b \text{ if } \quad p < 1 \tag{78}$$

2. *If $a, b > 0$, $a \neq b$ and $\lambda \in (0,1)$, then*

$$a^\lambda b^{1-\lambda} < \lambda a + (1 - \lambda) b \tag{79}$$

and we have equality above if $a, b \geqslant 0$ and $a = b$ or $ab = 0$.

Proof. 1. With $x = b/a$ the inequality is equivalent to $1 + x^p \leqslant (1 + x)^p$ for $x > 0$; let $f(x) = 1 + x^p - (1 + x)^p$. We have $f(0) = 0$ and $f'(x) = p[x^{p-1} - (1 + x)^{p-1}] \leqslant 0$ if $p \geqslant 1$ and $f'(x) > 0$ otherwise and (78) follows.

2. With x as in 1., the proof is very similar: the function is now $f(x) = x^\lambda - \lambda x - (1 - \lambda)$ which has a unique maximum at $x = 1$.

□

Corollary 25.0.3. *For $p \in (0,1)$, the triangle inequality fails in any $(X, \mathcal{M}, \mu)$ which has disjoint sets of positive measure.*

Proof. If $p \in (0,1)$, $\mu(E_1) > 0, \mu(E_2) > 0$ and $E_1 \cap E_2 = \emptyset$ it follows that

$$\| \chi_{E_1} + \chi_{E_2} \|_p > \| \chi_{E_1} \|_p + \| \chi_{E_2} \|_p$$

□

Note 25.0.4. 1. Failure of the triangle inequality and other oddities when $p \in (0,1)$ make L^p for $p \in (0,1)$ spaces quite pathological and not very useful.

2. For $p \geqslant 1$, induction shows that for any $n \in \mathbb{N}$ and positive numbers $a_1, ..., a_n$ we have

$$\sum_{i=1}^{n} a_i^p \leqslant \left(\sum_{i=1}^{n} a_i\right)^p \tag{80}$$

and, relying as usual on approximation by simple functions, we could easily show that $\| \|_p$ is a norm for $p \geqslant 1$. However, we get this inequality as a byproduct of another important inequality, Hölder's inequality.

Theorem 25.0.5 (Hölder's inequality)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space. Let $p \in (1, \infty)$ and let q be the* **conjugate exponent, or conjugate index** *satisfying $q^{-1} + p^{-1} = 1$. Then, for any two measurable functions f, g we have*

$$\|fg\|_1 \leqslant \|f\|_p \|g\|_q$$

If $f \in L^p$ and $g \in L^q$, then $f, g \in L^1$ and in this case we have equality iff for some $a, b \geqslant 0, a + b > 0$ we have $\alpha|f|^p = \beta|g|^q$.

Proof. W.l.o.g. we may assume $\|f\|_p, \|g\|_q$ are not zero or infinity, and furthermore, that $\|f\|_p = \|g\|_q = 1$, and, by replacing f, g by their absolute values, we assume $f \geqslant 0, g \geqslant 0$. By Proposition 25.0.2 2., and since $p^{-1} + q^{-1} = 1$, we have (pointwise)

$$fg \leqslant p^{-1} f^p + q^{-1} g^q \tag{81}$$

which by integration gives

$$\|fg\|_1 \leqslant p^{-1}\|f\|_p^p + q^{-1}\|g\|_q^q = 1 = \|f\|_p \|g\|_q$$

Clearly, equality holds if we have equality a.e. in (81), and the result follows, again by Proposition 25.0.2 2.

□

Theorem 25.0.6 (Minkowski's inequality)**.** *Let $p \in [1, \infty)$ and $f, g \in L^p$. Then,*

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p \tag{82}$$

Proof. For $p = 1$ or if $\|f + g\|_p = 0$ this is clear. If $p > 1$ and $\|f + g\|_p \neq 0$ we have

$$\begin{aligned} \|f+g\|_p^p &= \int |f+g|^p \,\mathrm{d}\mu = \int |f+g| \cdot |f+g|^{p-1} \,\mathrm{d}\mu \leq \int (|f| + |g|)|f+g|^{p-1} \,\mathrm{d}\mu \\ &= \int |f||f+g|^{p-1} \,\mathrm{d}\mu + \int |g||f+g|^{p-1} \,\mathrm{d}\mu \leqslant (\|f\|_p + \|g\|_p) \left(\int |f+g|^{(p-1)\left(\frac{p}{p-1}\right)} \,\mathrm{d}\mu\right)^{1-\frac{1}{p}} \\ &= (\|f\|_p + \|g\|_p)\, \|f+g\|_p^{p-1} \end{aligned} \tag{83}$$

by Hölder's inequality if $p \neq 1$ (since $q = p/(p-1)$). The conclusion is now straightforward. □

Exercise 65. Use measure-theoretic arguments to prove Minkowski's inequality directly from (80).

Theorem 25.0.7. *For any $p \in [1, \infty)$, L^p is a Banach space w.r.t. $\| \|_p$.*

Proof. By Minkowski's inequality $\|\|_p$ satisfies the triangle inequality and the rest of the properties of a norm are immediate.

Next, recalling Theorem 23.0.3 we show that a series of L^p functions which is absolutely convergent converges in L^p. Let $(f_k)_{k\in\mathbb{N}}$ be a sequence in L^p s.t.

$$\sum_{k\in\mathbb{N}} \|f_k\|_p = S < \infty \tag{84}$$

The trick is now to let $A(x) = \sum_{k=1}^{\infty} |f_k|$ and $A_N(x) = \sum_{k=1}^{N} |f_k|$. The sum of the infinite series A, exists pointwise as a function with values in $[0, \infty]$. Next we check that $A \in L^p$, which follows by monotone convergence from the fact that $A_N \nearrow A$, and

$$\forall N \in \mathbb{N},\ \|A_N\|_p \leqslant \sum_{k=1}^{N} \|f_k\|_p \leqslant S \tag{85}$$

which implies that A^p thus A are finite a.e. This, in turn, implies that $F = \sum_{k=1}^{\infty} f_k$ is convergent a.e. Now, $|F| \leqslant A$ pointwise shows that $F \in L^p$. It remains to show that $F - \sum_{k=1}^{N} f_k$ converges to zero in L^p which follows from pointwise convergence to zero, (84), (85) and dominated convergence. □

The following result is shown in a similar way using Theorem 5.2.2 2, and the proof is left as an exercise.

Proposition 25.0.8. *The set of simple functions of the form $\sum_{k\leqslant n} c_k \chi_{E_k}$ with $\mu(E_k) < \infty$ is dense in L^p, $p \in [1, \infty)$.*

25.0.1 The space L^∞

This space is the limit, in a precise sense, of L^p as $p \to \infty$. The norm in L^∞ is similar to a sup norm, now allowing for the functions to be defined a.e.

Definition 25.0.9 (essup norm)**.** *We let $(X, \mathcal{M}, \mu)$ be a measure space and f measurable on X. Define the essential supremum of f by*

$$\text{essup}(f) = \|f\|_\infty = \inf\{C \geq 0 : |f(x)| \leq C \ a.e.[\mu]\}$$

Equivalently,

$$\text{essup}(f) = \|f\|_\infty = \inf\{C \geq 0 : \mu(|f(x)| > C) = 0\}$$

$L^\infty(X, \mathcal{M}, \mu)$ is a set of equivalence classes f modulo null sets

$$L^\infty(X, \mathcal{M}, \mu) = \{f : X \to \mathbb{C} : f \text{ measurable }, \|f\|_\infty < \infty\}$$

Note that in each equivalence class in L^∞ there are functions bounded everywhere by their norm, and that the dependence of $L^\infty(\mathcal{M}, \mu)$ on the measure is relatively weak: If $\mu \ll \nu$ and $\nu \ll \mu$, then $L^\infty(\mathcal{M}, \mu) = L^\infty(\mathcal{M}, \nu)$

Proposition 25.0.10. *Assume $(X, \mathcal{M}, \mu)$ is a measure space and $f \in L^\infty(\mathcal{M}, \mu) \cap L^{p_0}(\mathcal{M}, \mu)$ for some*

$p_0 \geqslant 1$. Then $f \in L^p(\mathcal{M}, \mu)$ for all $p \geqslant p_0$ and

$$\|f\|_\infty = \lim_{p\to\infty} \|f\|_p$$

Proof. We can assume w.l.o.g. that $\|f\|_\infty \neq 0$ and by homogeneity, that $\|f\|_\infty = 1$. Then, for $p \geqslant p_0$, we have $|f|^p \leqslant |f|^{p_0}$ a.e., and thus $\|f\|_p$ is decreasing in p, implying $f \in L_p(\mathcal{M}, \mu)$ for all $p \geqslant p_0$ and that the limit exists. We write

$$\int |f|^p d\mu = \int |f|^{p-p_0} |f|^{p_0} d\mu = \int |f|^{p-p_0} d\nu; \quad d\nu := |f|^{p_0} d\mu$$

and noting that $(p - p_0)/p \to 1$ as $p \to \infty$ it is enough to prove the property for finite measures. We have

$$\|f\|_p \leqslant \nu(X)^{1/p} \to 1 \text{ as } p \to \infty$$

Let now $\varepsilon > 0$ and E be the set of positive measure where $|f| \geqslant 1 - \varepsilon$. We have

$$\|f\|_p \geqslant (1-\varepsilon)\nu(E)^{1/p} \to 1 \text{ as } p \to \infty$$

□

The proof of the following theorem is an easy exercise.

Theorem 25.0.11. *1. If f and g are measurable on X then $\|fg\|_1 \leqslant \|f\|_1 \|g\|_\infty$.*

2. $(f_n)_{n\in\mathbb{N}}$ is Cauchy in $\|\|_\infty$ iff there is a set $X' \subset X$ of full measure s.t. f_n converges pointwise uniformly on X'.

3. $\|\|_\infty$ is a Banach space.

4. Simple functions are dense in L^∞.

Proposition 25.0.12. *If $\mu(X) = M < \infty$, $0 < p < q \leqslant \infty$ and f is measurable, then $\|f\|_p \leqslant \|f\|_q M^\alpha$, where $\alpha = p^{-1} - q^{-1}$, and, in particular, $L^p(X, \mu) \supset L^q(\mu)$.*

Proof. The case $q = \infty$ is immediate, so assume q is finite. Replacing f by $|f|$, we may assume $f \geqslant 0$. We have

$$\|f\|_p^p = \|f^p \cdot 1\|_1 \leqslant \|f^p\|_{q/p} \|1\|_{q/(q-p)} = \|f\|_q M^\alpha$$

□

The notation $\ell^p(X)$ stands for $L^p(X)$ when the measurable space is $(X, \mathcal{P}(X))$ and the measure is the counting measure.

Proposition 25.0.13. *For any set A and $0 < p < q \leqslant \infty$, $\ell^p(A) \subset \ell^q(A)$, and $\|\|_p \geqslant \|\|_q$.*

Proof. We assume A is an infinite set, since otherwise the proof is trivial. Note first that for a sum $\sum_{\alpha\in A} |x_\alpha|^p$ to be finite we must have $x_\alpha = 0$ for all but countably many α, and for $\sum_{\alpha_n, n\in\mathbb{N}} |x_{\alpha_n}|^p$ to be finite we must have $|x_{\alpha_n}|^p \to 0$ as $n \to \infty$, in particular $|x_{\alpha_n}| < 1$ for all large n. But then $|x_{\alpha_n}|^q \leqslant |x_{\alpha_n}|^p$ for all large n implying the statement. □

Note 25.0.14. 1. If any open set in X has nonzero measure, then the uniform norm on continuous functions is the same as the $\|\|_\infty$.

2. If $\mathcal{M} = [0, \infty)$ with the Lebesgue measure, and $p \geqslant 1$, then $x^{-a}\chi_{[0,1]}$ is in L^p if $a < 1/p$ and $x^{-a}\chi_{[1,\infty)}$ is in L^p if $a > 1/p$. See also Proposition 25.0.15 2. below.

The last example shows there is no nontrivial inclusion between L^p spaces (except in special cases as in the propositions above). However, the collection of L^p spaces possesses important interpolation properties, of which we first note some elementary ones.

Proposition 25.0.15. Let $0 < p < r \leqslant \infty$.

1. For any $q \in (p, r)$ and $f \in L^q$ there exist $f_1 \in L^p$ and $f_2 \in L^r$ s.t. $f = f_1 + f_2$, that is $L^q \subset L^p + L^r$.
2. For $q \in (p, r)$, write q^{-1} as the a convex decomposition, $q^{-1} = \lambda p^{-1} + (1-\lambda)r^{-1}$, $\lambda \in (0, 1)$. We have $\|\cdot\|_q \leqslant \|\cdot\|_p^\lambda \, \|\cdot\|_r^{1-\lambda}$ and in particular $L^p \cap L^r \subset L^q$.

Proof. 1. Let $f \in L^q$ and write $f = f\chi_{|f|>1} + f\chi_{|f|\leqslant 1}$. Clearly the first function is in L^p and the second one in L^r.

2. We may assume $r < \infty$, since the case $r = \infty$ is straightforward, and as before we take a nonnegative measurable function f. By Hölder's inequality, we have

$$\|f\|_q^q = \|f^q\|_1 = \|f^{\lambda q} f^{(1-\lambda)q}\|_1 \leqslant \|f^{q\lambda}\|_{p/(q\lambda)} \|f^{q(1-\lambda)}\|_{r/(q(1-\lambda))} = \|f\|_p^{\lambda q} \|f\|_r^{(1-\lambda)q}$$

The case $r = \infty$ is straightforward (and also follows as a limit from the above). $\square$

We now prove a useful, measure-theoretic lemma.

Lemma 25.0.16. *Let $(X, \mathcal{M}, \mu)$ be a finite measure space, $g \in L^1$ and $\mathcal{C} \subset \mathbb{C}$ a closed set. If for all $E \in \mathcal{M}$ with $\mu(E) > 0$ the averages of g are in $\mathcal{C}$, i.e.*

$$\mu(E)^{-1} \int_E g \, d\mu \in \mathcal{C}$$

then $\mu(\{x : g(x) \notin \mathcal{C}\}) = 0$.

Proof. The complement $\mathcal{C}^c$ is a countable union of closed disks of the form $\overline{B_{r_n}(x_n)}$. We show that for any of them, say $\overline{B_r(x)}$ we have $\mu(\{x : g(x) \in \overline{B_r(x)}\}) \neq 0$. Indeed, otherwise

$$\left| x - \mu(E)^{-1} \int_E g \, d\mu \right| = \left| \mu(E)^{-1} \int_E (g - x) \, d\mu \right| \leqslant \mu(E)^{-1} \int_E |g - x| \, d\mu \leqslant r$$

a contradiction, since $\mathcal{C}^c$ is open and $\operatorname{dist}(\overline{B_r(x)}, \mathcal{C}) > 0$.

$\square$

Lemma 25.0.17.

1. Let $(X, \mathcal{M}, \mu)$ be a measure space and assume μ is σ-finite. Then, there exists a $w \in L^1(\mu)$ s.t. $0 < w(x) < 1$ everywhere in X. In particular, $d\tilde{\mu} = w \, d\mu$ is finite, and μ and $\tilde{\mu}$ are mutually absolutely continuous.
2. With w as above, the map $f \mapsto w^{1/p} f$ is an isometric isomorphism between $L^p(\tilde{\mu})$ and $L^p(\mu)$.

Proof. 1. Let $(E_n)_{n\in\mathbb{N}}$ be sets of finite measure s.t. $X = \cup_n E_n$. Such a w is given by

$$w = \sum_{n\in\mathbb{N}} w_n \qquad \text{where} \quad w_n(x) = \begin{cases} \dfrac{2^{-n}}{1+\mu(E_n)} & x \in E_n \\ \\ 0 & x \notin E_n \end{cases}$$

2. This follows easily from the fact that $0 < w < 1$. □

Exercise 66. Assume $f : \mathbb{R} \to \mathbb{C}$ is periodic with period 2π and absolutely continuous on any compact set in $\mathbb{R}$. Check that the Fourier coefficients of f' are well defined. Assume the sequence of Fourier coefficients of f' is in $\ell^p(\mathbb{Z})$ for some $p \in [1, \infty)$.

(a) Show that the Fourier series of f converges uniformly and absolutely to some continuous function, $\tilde{f}$.

(b) Show that $\tilde{f} = f$.

26 The dual of L^p

The following summarizes the picture when the measure is σ-finite (more generally, semifinite). For $1 \leqslant p < \infty$, the dual of L^p is isometrically isomorphic to L^q where q is the conjugate exponent to p; because of the isomorphism the dual of $L^p, p \in [1, \infty)$ is identified with L^q. It follows that L^p is reflexive for all $p \in (1, \infty)$. If $p = 1$, then $q = L^\infty$ and if the measure is sigma-finite then the dual of L^1 is L^∞. In general, the dual of L^∞ is much larger than L^1 [14].

Theorem 26.0.1. In the following $(X, \mathcal{M}, \mu)$ is a measure space, $L^p = L^p(X, \mathcal{M}, \mu)$ and q is the conjugate exponent to p. ($(1, \infty)$ are dual exponents.)

1. Let $1 \leqslant p \leqslant \infty$ and $g \in L^q$. Then, Φ_g defined by

$$\Phi_g(f) = \int_X fg d\mu$$

   is a bounded linear functional on L^p and

$$\|\Phi_g\| \leqslant \|g\|_q \tag{86}$$

2. Assume μ is σ-finite. Let $1 \leqslant p < \infty$ and let Φ be a bounded linear functional on L^p. Then, there exists a unique $g \in L^q$ such that

$$\Phi(f) = \int_X fg d\mu \tag{87}$$

[14] This statement uses a strong form of the axiom of choice, and there are models of ZF where the dual of L^∞ is L^1; see M. Väth, Indag. Mathem., N.S. 9 (4), pp. 619–625 (1998).

Moreover, if Φ and g are as in (87), then

$$\|\Phi\| = \|g\|_q \tag{88}$$

Note 26.0.2. *Theorem 26.0.1,1 above implies Theorem 6.14 in Folland.*

Exercise 67. Assume $(X, \mathcal{M}, \mu)$ is a $\sigma-$finite measure space. Based on Theorem 26.0.1 show that the dual of L^p is indeed isometrically isomorphic to L^q for all $1 \leqslant p < \infty$ (as usual, we assume $1/p + 1/q = 1$).

Proof. **(adapted from Rudin)**. 1. This follows immediately from Hölder's inequality.

2. We first prove uniqueness. By linearity, it suffices to show that $\Phi = 0 \Rightarrow g = 0$. Let $E \subset X$ have finite measure. Then, $0 = \Phi(\chi_E) = \int_E g$, and thus $g = 0$ a.e. (check!)

For existence, we show that Φ induces a complex measure λ, that $\lambda \ll \mu$ and $\frac{d\lambda}{d\mu} \in L^q$.

(A) It is helpful to analyze first the case when μ *is finite*. Note that in this case any measurable characteristic function is in all $L^p, 1 \leqslant p \leqslant \infty$. Let E be a measurable subset of X and define

$$\lambda(E) = \Phi(\chi_E) \tag{89}$$

If $E_1, ..., E_n$ are mutually disjoint and their union is E, then $\chi_E = \sum_1^n \chi_{E_i}$, implying finite additivity of λ. Now, if $(E_i)_{i \in \mathbb{N}}$ are mutually disjoint and their union is E, then $\chi_E - \chi_{\cup_1^n E_i} = \chi_{E \setminus \cup_1^n E_i}$, hence

$$\|\chi_E - \chi_{\cup_1^n E_i}\|_p^p = \mu(E \setminus \cup_1^n E_i) \to 0 \ \text{ as } \ n \to \infty \tag{90}$$

The continuity of Φ now implies

$$\lambda(E \setminus \cup_1^n E_i) \to 0 \ \text{ as } \ n \to \infty \tag{91}$$

Thus λ is a complex σ-additive measure, which furthermore is absolutely continuous w.r.t. μ, because $\mu(E) = 0 \Rightarrow \|\chi_E\|_p = 0$. Now, the Radon-Nikodym theorem implies that there is a $g \in L^1(\mu)$ s.t.

$$\Phi(\chi_E) = \int_E g d\mu = \int_X \chi_E g d\mu, \qquad \forall E \in \mathcal{M} \tag{92}$$

By linearity, the density of simple functions in L^p and continuity of Φ, we have

$$\Phi(f) = \int_X f g d\mu \tag{93}$$

for any $f \in L^p$.

(i) If $p = 1$ we are nearly done. Indeed, for any E

$$|\lambda(E)| = \left| \int_E g d\mu \right| = |\Phi(\chi_E)| \leqslant \|\Phi\| \|\chi_E\|_1 = \|\Phi\| \mu(E) \tag{94}$$

which implies that the total variation of λ satisfies $|\lambda|(X) \leqslant \|\Phi\| \mu(X)$ and, by Lemma 25.0.16,

$$\|g\|_\infty \leqslant \|\Phi\| \tag{95}$$

Combining with (86), we get that the dual of L^1 is L^∞ since

$$\|g\|_\infty = \|\Phi\| \tag{96}$$

(ii) Let now $p > 1$. Since the measure is finite, we have $L^\infty \subset L^p \subset L^1$ for all $p \geqslant 1$, and L^∞ is dense in L^p. We need to show that $g \in L^q$. Let $\alpha = csgn(g)$ $(= \overline{sgn}(g))$, $E_n = \{x : |g(x)| \leqslant n\}$ and define $G = \chi_{E_n}\alpha|g|^{q-1}$. Clearly, $G \in L^\infty$, $|G|^p = |g|^q$ on E_n, and (93) gives

$$\int_{E_n} |g|^q d\mu = \int_{E_n} Ggd\mu = |\Phi(G\chi_{E_n})| \leqslant \|\Phi\| \left(\int_{E_n} |g|^q d\mu\right)^{1/p}$$

hence, by solving the inequality for the first integral above,

$$\int_X \chi_{E_n}|g|^q d\mu \leqslant \|\Phi\|^q, \quad \forall n \in \mathbb{N} \tag{97}$$

and, by monotone convergence, $\|g\|_q \leqslant \|\Phi\|$. Continuity of Φ and of the right side of (93) now implies that (93) holds on L^p.

Let now $\mu(X) = \infty$ but assume μ is σ-finite. Here we use the isomorphism provided by Lemma 25.0.17. Let $\Phi \in (L^p(\mu))^*$ and define Ψ on $(L^p(\tilde{\mu}))^*$ be given by

$$\Psi(f) = \Phi(w^{1/p}f) \tag{98}$$

The isomorphism implies $\|\Psi\|_{(L^p(\tilde{\mu}))^*} = \|\Phi\|_{(L^p(\mu))^*}$. Since $\tilde{\mu}$ is finite, there is a $G \in L^q(\tilde{\mu})$ s.t.

$$\Psi(f) = \int_X fGd\tilde{\mu} \quad \text{for all} \quad f \in L^p(\tilde{\mu}) \tag{99}$$

For $p = 1$ let $g = G$, and for $p > 1$ let $g = w^{1/q}G$. If $p = 1$ we have $\|g\|_\infty = \|G\|_\infty = \|\Psi\|_{(L^1(\tilde{\mu}))^*} = \|\Phi\|_{(L^1(\mu))^*}$, while for $p > 1$,

$$\int_X |g|^q d\mu = \int_X |G|^q d\tilde{\mu} = \|\Psi\|^q_{(L^p(\tilde{\mu}))^*} = \|\Phi\|^q_{(L^p(\mu))^*} \tag{100}$$

this implies (88) and, since $Gd\tilde{\mu} = w^{1/p}gd\mu$ we get, for all $f \in L^p(\mu)$,

$$\Phi(f) = \Psi(w^{-1/p}f) = \int_X w^{-1/p}fGd\tilde{\mu} = \int_X fgd\mu \tag{101}$$

□

Corollary 26.0.3. *L^p is reflexive for $p \in (1, \infty)$.*

Exercise 68. $C[0,1]$ is dense in $L^p[0,1]$ for all $1 \leqslant p < \infty$ (note the inequalities, L^∞ is not included!) and $\Phi = f \mapsto f(0)$ is a linear functional on $C[0,1]$. However, Φ does not extend to a bounded functional on L^p. Why not?

26.1 The dual of L^∞ is not L^1

An "example" of a continuous functional on L^∞ is obtained using Hahn-Banach (which uses the axiom of choice, hence the quotation marks). Let $\Phi = f \mapsto f(0)$ defined on $C[-1,1]$. By Hahn-Banach this extends (nonuniquely, of course) to L^∞. But Φ cannot be given by an L^1 element (for f in $C[-1,1]$ write Φ as $\int f dm$ where m is the Dirac measure at zero). Note also the effect of using the axiom of choice: the extended functionals associate some generalized value at a point to functions in L^∞.

Homework: 4,9,10,12,13 pp. 186–7 in Folland, and turn in Exercises 66 and 67 in these notes.

26.2 Inequalities in L^p spaces

Proposition 26.2.1 (Chebyshev's inequality). Let $0 < p < \infty$, $\alpha > 0$ and $f \in L^p$. Then,

$$\mu(\{x : |f(x)| > \alpha\}) \leqslant \alpha^{-p}\|f\|_p^p$$

Proof. This follows immediately from Markov's inequality, (46). □

Definition 26.2.2. An operator T of the form

$$(T\varphi)(t) = \int_X K(t,x)\varphi(t)d\mu(t) \tag{102}$$

(under suitable assumptions on K and φ) is called an integral operator (more precisely, a linear integral operator), and K is called the kernel of T.

Proposition 26.2.3. Assume $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ are $\sigma-$ finite measure spaces and $K : X \times Y \to \mathbb{C}$ a kernel which is uniformly L^1 w.r.t. μ and ν, that is, there is a $C > 0$ s.t.

$$\|K(\cdot, y)\|_{L^1(\mu)} \leqslant C \text{ for a.e. } y\ [\nu]; \quad \text{and} \quad \|K(x, \cdot)\|_{L^1(\nu)} \leqslant C \text{ for a.e. } x[\mu] \tag{103}$$

Then, for any $1 \leqslant p < \infty$ T is a bounded operator from L^p to L^p with norm $\|T\|_{L^p \to L^p} \leqslant C$ and the integral in (102) converges absolutely.

Proof. If $p = 1$ this follows directly from Fubini-Tonelli, while for $p = \infty$ it follows from majorizing $|f|$ by its norm.

Assume now $1 < p < \infty$, write $|K| = |K|^{1/q}|K|^{1/p}$ and apply Hölder's inequality:

$$\int |K(x,y)f(y)|d\nu \leqslant \|K(x,\cdot)\|_{L^1(\nu)}^{1/q} \left(\int |K(x,y)||f(y)|^p d\nu\right)^{\frac{1}{p}} \leqslant C^{\frac{1}{q}} \left(\int |K(x,y)||f(y)|^p d\nu\right)^{\frac{1}{p}} \tag{104}$$

By Fubini-Tonelli,

$$\int \left(\int |K(x,y)||f(y)|d\nu\right)^p d\mu \leqslant C^{\frac{p}{q}} \int\int |K(x,y)||f(y)|^p d\nu d\mu \leqslant C^{1+\frac{p}{q}} \|f\|_p \tag{105}$$

and the result follows by taking the p-th root. □

Writing Minkowski's inequality for the nonnegative functions f_i as

$$\left(\int \left(\sum_i f_i\right)^p\right)^{\frac{1}{p}} \leqslant \sum_i \left(\int f_i^p\right)^{1/p}$$

suggests a generalization, in which the sum is replaced by an integral:

Theorem 26.2.4 (Minkowski's inequality for integrals). Assume $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ are $\sigma-$ finite measure spaces and $f : X \times Y$ a nonnegative $\mathcal{M} \otimes \mathcal{N}$-measurable function. Then,

$$\left(\int_X \left(\int_Y f(x,y)d\nu\right)^p d\mu\right)^{\frac{1}{p}} \leqslant \int_Y \left(\int_X f(x,y)^p d\mu\right)^{\frac{1}{p}} d\nu$$

Proof. If $p = 1$ this is simply Fubini-Tonelli. If now $1 < p < \infty$ we use the $L^p - L^q$ duality to estimate the integrals via (88). Take a nonnegative $g \in L^q$, and note that, by Hölder and Fubini-Tonelli,

$$\int_X \left(\int_Y f(x,y)d\nu\right) g(x)d\mu = \int_X \int_Y f(x,y)g(x)d\mu d\nu \leqslant \|g\|_q \int_Y \left(\int_X f(x,y)^p d\mu\right)^{\frac{1}{p}} d\nu \tag{106}$$

□

Corollary 26.2.5. Let $1 \leqslant p \leqslant \infty$, $f(\cdot,y) \in L^p(\mu)$ a.e.$[d\nu]$ and assume $y \mapsto \|f(\cdot,y)\|_p \in L^1(\nu)$. Then, $f(x,\cdot) \in L^1(\nu)$, $x \mapsto \int f(x,y)d\nu \in L^p(\mu)$ and

$$\left\|\int f(\cdot,y)d\nu\right\|_p \leqslant \int \|f(\cdot,y)\|_p d\nu$$

Proof. This is a straightforward consequence of the previous theorem, except for the case $p = \infty$, which is a result of the nonnegativity of integrals. □

Let now K be a measurable kernel on $(\mathbb{R}^+)^2$ which, for all $\lambda > 0$, satisfies $\lambda K(\lambda x, \lambda y) = K(x, y)$ and

$$\int_0^\infty |K(x,1)| x^{-1/p} dx = C < \infty$$

for some $1 \leqslant p \leqslant \infty$. Let $f \in L^p$ and $g \in L^q$ where q is the dual exponent to p, and define the integral operators

$$Tf(y) = \int_0^\infty K(x,y) f(x) dx, \quad Sg(x) = \int_0^\infty K(x,y) g(y) dy$$

Proposition 26.2.6. Under the hypotheses above T is a bounded operator from L^p to L^p, S is a bounded operator from L^q to L^q and we have $\|T\|_p = \|S\|_q = C$.

Proof. Assume as before that f, g are nonnegative. Let $f(z\cdot) = y \mapsto f(zy)$. We first note the following scaling properties.

$$\|f(z\cdot)\|_p^p = \int_0^\infty f(yz)^p dy = z^{-1} \int_0^\infty f(u)^p du$$

$$\int_0^\infty |K(x,y) f(x)| dx = \int_0^\infty |yK(yz,y) f(yz)| dz = \int_0^\infty |K(z,1) f(zy)| dz$$

$$\int_0^\infty |K(1,y)| y^{-1/q} dy = \int_0^\infty |K(y^{-1},1)| y^{-1-\frac{1}{q}} dy = \int_0^\infty |K(u,1)| u^{-\frac{1}{p}} du = C \tag{107}$$

Using Proposition 26.2.3 we get

$$\|Tf\|_p \leqslant \int_0^\infty |K(z,1)| \|f(z\cdot)\|_p dz = \|f\|_p \int_0^\infty |K(z,1)| z^{-1/p} dz = C$$

The statement about S follows in the same way, now using (107). $\square$

Corollary 26.2.7. *Consider the following integral operators, with kernel $K(x,y) = y^{-1} \chi_{\{(x,y): x<y\}}$:*

$$Tf(y) = \frac{1}{y} \int_0^y f(x) dx; \quad Sg(x) = \int_x^\infty y^{-1} g(y) dy$$

Then, for $1 \leqslant p < \infty$, $T : L^p \to L^p$ and $S : L^q \to L^q$ are bounded with norm $\frac{p}{p-1} = q$.

Proof. This follows from Proposition 26.2.6, noting that $\int_0^\infty |K(x,1)| x^{-1/p} dp = \int_0^1 x^{-1/p} dp = \dfrac{p}{p-1} = q$. $\square$

26.3 Weak L^p

Let f be measurable on a measure space $(X, \mathcal{M}, \mu)$. The distribution function of f, $\lambda_f : (0, \infty) \to [0, \infty]$ (compare to the definition following Exercise 17) is given by

$$\lambda_f(\alpha) = \mu(\{x : |f(x)| > \alpha\})$$

Proposition 26.3.1.

1. λ_f is decreasing and right continuous.
2. $|f| \leqslant |g| \Rightarrow \lambda_f \leqslant \lambda_g$.
3. $|f_n| \nearrow |f| \Rightarrow \lambda_{f_n} \nearrow \lambda_f$.
4. $f = g + h \Rightarrow \lambda_f(\alpha) \leqslant \lambda_g(\alpha/2) + \lambda_h(\alpha/2)$.

Proof. Continuity follows from the continuity from below of μ, since $\{x : |f(x)| > \alpha\} = \cup_j \{x : |f(x)| > \alpha + 1/j\}$. The rest is straightforward. $\square$

Assume now that $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$. By the usual construction of measures from distribution functions, ν given by $\nu((a, b]) = \lambda_f(b) - \lambda_f(a)$ defines a measure on $\mathbb{R}^+$.

Proposition 26.3.2. *If $\lambda_f(\alpha) < \infty$ for all $\alpha > 0$ and $\varphi : \mathbb{R}^+$ is measurable and nonnegative, then*

$$\int_X \varphi(|f|) d\mu = -\int_0^\infty \varphi(\alpha) d\lambda_f(\alpha)$$

Note 26.3.3. We can think of this formula as a representation of the integral as one in terms of possible values of the function against the "probability density" of a value to occur.

Another way to view it is as a generalized change of variable. Indeed, let f be continuous with sufficient decay and g a diffeomorphism. Then, check that

$$\int_{\mathbb{R}} f(g(x)) dx = \int_{g(\mathbb{R})} f(u) d\mu(u); \quad \mu = m(g^{-1})$$

Proof. We first prove the result when φ is a characteristic function of an interval.

$$\int_X \chi_{(a,b]} \circ f = -\mu(\{x : a < |f(x)| \leqslant b\}) = -(\lambda_f(b) - \lambda_f(a)) = -\int_0^\infty \chi_{(a,b]} d\lambda_f \tag{108}$$

From here the result can be extended to characteristic functions of general measurable sets, then to simple functions and finally to any nonnegative measurable function, as usual. $\square$

In particular, we have

$$\|f\|_p^p = \int_X |f|^p d\mu = -\int_0^\infty \alpha^p d\lambda_f(\alpha) \tag{109}$$

Proposition 26.3.4. *Let f be measurable on the measure space $(X, \mathcal{M}, \mu)$. We have*

$$\int_X |f|^p d\mu = p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha \tag{110}$$

Proof. If $\|f\|_p = \infty$ then the right side is $+\infty$ as well. Otherwise, let first f be a simple function. Then, both f and λ_f vanish for large values of the argument. Noting that α^p is continuous, we have, by Proposition 16.2.1,

$$p\int_0^\infty \alpha^{p-1}\lambda_f(\alpha)d\alpha = -\int_0^\infty \alpha^p d\lambda_f(\alpha)$$

For a general f, there is a sequence of simple functions that converge monotonically to $|f|$, and the rest follows from monotone convergence.

□

Exercise 69. Check that, if $\|f\|_p < \infty$, then

$$[f]_p := \left(\sup_{\alpha>0} \alpha^p \lambda_f(\alpha)\right)^{1/p} < \infty$$

but the converse is not true.

Show that

$$[f+g]_p \leqslant 2\left([f]_p^p + [g]_p^p\right)^{1/p}; \quad \text{and} \quad [cf]_p = |c|[f]_p$$

Definition 26.3.5. Let $(X, \mathcal{M}, \mu)$ be a measure space. For $p \in (0, \infty)$, $L^{p,w}(X, \mu)$, or **weak** L^p, is defined as

$$L^{p,w}(X,\mu) = \{f \text{ measurable} : [f]_p < \infty\}$$

Proposition 26.3.6. Weak L^p is a topological space, $\|f\|_p \leqslant [f]_p$, and thus $L^p \subset L^{p,w}$. The inclusion can be strict.

Proof. Chebyshev's inequality implies $\|f\|_p \leqslant [f]_p$. An example of a strict inclusion is $L^p(\mathbb{R}^+)$ where $x^{-1/p} \in L^{p,w}$ but not in L^p. The rest follows from Exercise 69 above. □

The following decomposition result is sometimes useful; its proof is a simple exercise.

Proposition 26.3.7. Let f be measurable and real-valued. Take $A > 0$, define $E_A = \{x : |f(x)| > A\}$ and

$$h_A = f\chi_{E_A^c} + A\,\mathrm{sgn}(f)\chi_{E_A}; \quad g_A = f - h_A$$

Then,

$$\lambda_{h_A}(\alpha) = \lambda_f(\alpha)\chi_{\alpha<A}; \quad \lambda_{g_A}(\alpha) = \lambda_f(A+\alpha)$$

26.4 L^p interpolation theorems

An essential ingredient in the interpolation theorems in this section is the following consequence of the maximum principle in complex analysis.

Theorem 26.4.1. *[***The Hadamard three-lines theorem***]* Let f be holomorphic and bounded in the vertical strip $\{z = x + iy | a \leq x \leq b\}$ and continuous up to its boundary. If

$$M(x) = \sup_y |f(x+iy)|$$

then $\log M$ is a convex function on $[a, b]$.

Equivalently, if $x = ta + (1-t)b$ with $0 \leq t \leq 1$, then

$$M(x) \leq M(a)^t M(b)^{1-t}$$

Proof. After an affine transformation of the variable z it can be assumed that $a = 0$ and $b = 1$. The function

$$F_n(z) = f(z)M(0)^{z-1}M(1)^{-z}e^{z^2/n}e^{-1/n}$$

is entire. We see that, for any n, $|F_n(z)| \leqslant 1$ on the boundary of the strip and, with $z = x + iy$, $|F_n(z)| \to 0$ uniformly in $x, 0 \leq x \leq 1$ as $|y| \to \infty$. Hence $|F_n| \leqslant 1$ on the boundary of the rectangle $\{z = x + iy | 0 \leq x \leq 1, |y| = m\}$ if m is large enough, and by the maximum principle $|F_n| \leqslant 1$ in the strip. The result follows by letting $n \to \infty$. □

Exercise 70. Prove Hölder's inequality as a corollary of the three-lines theorem as follows. Let p, q be conjugate exponents in $(1, \infty)$, assume $f \in L^p, g \in L^q$ and define

$$h(z) = \int |f|^{pz}|g|^{q(1-z)}$$

Check that h satisfies the hypotheses of Theorem 26.4.1 with $a = 0, b = 1$ and that this implies Hölder's inequality.

Exercise 71. Let $\theta_1 \in (-\pi, \pi)$ and let $\theta_2 \in \mathbb{R}$ be s.t. $0 < \theta_2 - \theta_1 < 2\pi$. Define the open sector $S = \left\{ re^{i\theta} : r > 0 \text{ and } \theta_1 < \theta < \theta_2 \right\}$. Let f be holomorphic in S, continuous and bounded on $\overline{S}$. For $\theta \in [\theta_1, \theta_2]$, set

$$M(\theta) = \sup_{r>0} |f(re^{i\theta})|$$

Prove that

$$M(\theta) \leqslant M(\theta_1)^{\frac{\theta_2-\theta}{\theta_2-\theta_1}} M(\theta_2)^{\frac{\theta-\theta_1}{\theta_2-\theta_1}}$$

Theorem 26.4.2 (Riesz-Thorin interpolation theorem). Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be σ-finite measure spaces and let $1 \leqslant p_0, p_1, q_0, q_1 \leqslant \infty$. Denote

$$\frac{1}{p_t} = \frac{1-t}{p_0} + \frac{t}{p_1}; \quad \frac{1}{q_t} = \frac{1-t}{q_0} + \frac{t}{q_1} \quad (t \in (0,1)) \tag{111}$$

Assume T is a linear map from $L^{p_0}(\mu) + L^{p_1}(\mu)$ to $L^{q_0}(\nu) + L^{q_1}(\nu)$ s.t.

$$\|T\|_{L^{p_0} \to L^{q_0}} \leqslant M_0 \text{ and } \|T\|_{L^{p_1} \to L^{q_1}} \leqslant M_1$$

Then

$$\|T\|_{L^{p_t} \to L^{q_t}} \leqslant M_t := M_0^{1-t} M_1^t, \quad \forall t \in (0,1)$$

Furthermore, $\log M_t$ is a convex function of $t \in (0,1)$.

Proof. Note that the case $p_0 = p_1$ follows from Proposition 25.0.1, replacing f by Tf and taking $\lambda = t, p_0 = r, p_1 = p$.

We now assume $p_0 < p_1$. We first prove the result for simple functions, by constructing interpolating expressions to which the three-line theorem applies. The duality expressed in (88) comes in handy at this stage. We then use the density of simple functions in L^p to complete the proof.

Using the form (73) of the norm of T, take $f = \sum_{j=1}^n |a_j| e^{i\theta_j} \chi_{E_j}$ and $g = \sum_{k=1}^n |b_k| e^{i\theta_k} \chi_{E_k}$ to be simple functions with $\|f\|_{p_t} = 1, \|g\|_{q_t'} = 1$, where q_t' is the conjugate exponent to q_t. Extending (115) to the strip $S = \{z = x + iy | x \in (0,1), y \in \mathbb{R}\}$, define

$$\frac{1}{p_z} = \alpha(z) = \frac{1-z}{p_0} + \frac{z}{p_1}; \quad \frac{1}{q_z} = \beta(z) = \frac{1-z}{q_0} + \frac{z}{q_1} \quad (z \in S) \tag{112}$$

We also extend f, g to S, as follows

$$f_z := \sum_{j=1}^n |a_j|^{\frac{\alpha(z)}{\alpha(t)}} e^{i\theta_j} \chi_{E_j}; \quad g_z = \sum_{k=1}^n |b_k|^{\frac{1-\beta(z)}{1-\beta(t)}} e^{i\theta_k} \chi_{E_k}; \quad t \in (0,1); \quad z \in S$$

if $\beta(t) \neq 1$ and $g_z = g$ otherwise.

Now we define the functional whose norm will provide us with the result. Let

$$\Phi(z) = \int_Y g_z \cdot (Tf_z)\, d\nu \tag{113}$$

Expanding out we get, if $\beta(t) \neq 1$,

$$\Phi(z) = \sum_{j,k} |a_j|^{\frac{\alpha(z)}{\alpha(t)}} |b_k|^{\frac{1-\beta(z)}{1-\beta(t)}} \chi_{E_k} e^{i\theta_j + i\theta_k} \int_Y \chi_{F_k} \left(T\chi_{E_j}\right) d\nu$$

and, if $\beta = 1$,

$$\Phi(z) = \sum_{j,k} |a_j|^{\frac{\alpha(z)}{\alpha(t)}} |b_k| \chi_{E_k} e^{i\theta_j + i\theta_k} \int_Y \chi_{F_k} \left(T\chi_{E_j}\right) d\nu$$

Since $z \mapsto \exp(bz)$ is entire and, for real b, bounded in any vertical strip of finite width, the functional Φ satisfies the analyticity and boundedness assumptions in S.

Note also that for any simple function h, with $h(x) = \sum_{j=1}^n |c_j| e^{i\varphi_j} \chi_{E_j}(x)$, and any $x \in X$, at most one of the terms in the sum is nonzero. Hence, $h^p = \sum_{j=1}^n |c_j|^p e^{ip\varphi_j} \chi_{E_j}$. In particular,

$$|f_z|^a = \sum_{j=1}^n |a_j|^{\frac{a\Re\alpha(z)}{\alpha(t)}} \chi_{E_j}; \quad |g_z|^a = \sum_{k=1}^n |b_k|^{\frac{a - a\Re\beta(z)}{1-\beta(t)}} \chi_{E_k} \tag{114}$$

Let's check the bounds for f_z, g_z on $\Re z = 0$. Using (114) we have, for $s \in \mathbb{R}$,

$$|f_{is}| = |f|^{\frac{p_t}{p_0}}; \quad |g_{is}| = |g|^{\frac{q_t'}{q_0'}}$$

We now use Hölder in the definition (113), the bound $\|T\|_{L^{p_0} \to L^{q_0}} \leqslant M_0$, and the fact that $f \in L^{p_t}$ iff $|f_{is}| \in L^{p_0}$, to get

$$|\Phi(is)| \leqslant \|Tf_{is}\|_{q_0} \|g_{is}\|_{q_0'} \leqslant M_0 \|f_{is}\|_{p_0} \|g_{is}\|_{q_0'} = M_0 \|f\|_{p_t} \|g\|_{q_t'} = M_0$$

Similarly,

$$|\Phi(1+is)| \leqslant M_1$$

and the result in the theorem follows, for simple functions. Extending this to general L^{p_t} requires another interesting step.

Let $f \in L^{p_t}$ and f_n a sequence of simple functions s.t., for all n, $|f_n| \leqslant f$ and $f_n \to f$ pointwise. Now we use the decomposition in Proposition 25.0.15: with $E = \{x \,|\, |f(x)| > 1\}$, let $g = f\chi_E$ and $h = f - g = f\chi_{E^c}$. Then, $f \in L^{p_t}$ implies $g \in L^{p_0}$ and $h \in L^{p_1}$. Define also $g_n = \chi_E f_n$, $h_n = \chi_{E^c} f_n$. By dominated convergence,

$$g_n \to g \ \text{ in } L^{p_0}, \ h_n \to h \ \text{ in } L^{p_1}$$

By the assumptions on T,

$$Tg_n \to Tg \ \text{ in } L^{q_0} \ \text{ and } Th_n \to Th \ \text{ in } L^{q_1}$$

Therefore, there is a subsequence s.t. $Tg_{n_k} \to Tg$ and $Th_{n_k} \to Th$ pointwise a.e., hence $Tf_{n_k} \to Tf$

pointwise a.e. Now, the Fatou Lemma gives

$$\|Tf\|_{q_t} \leqslant \liminf_{n\to\infty} \|Tf_n\|_{q_t} \leqslant \liminf_{n\to\infty} M_0^{1-t} M_1^t \|f_n\|_{p_t} = M_0^{1-t} M_1^t \|f\|_{p_t}$$

Convexity follows by applying this result for different pairs of indices p between p_0 and p_1. □

Another powerful interpolation theorem is due to Marcinkiewicz, which uses weak L^p assumptions instead.

Exercise 72. (A) On $L^1(\mathbb{T}) + L^2(\mathbb{T})$ define the Fourier transform operator $f \mapsto \hat{f}$ by

$$\hat{f}(k) = \int_0^1 f(t) e^{-2\pi i k t} dt$$

Use L^1 and L^2 estimates to prove that $f \mapsto \hat{f}$ is bounded from $L^p(\mathbb{T})$ to $\ell^q(\mathbb{Z})$ for $p \in [1, 2]$, where $p^{-1} + q^{-1} = 1$.

(B) Let $(\varphi_n)_{n\in\mathbb{N}}$ be a sequence of functions in $L^p([0,1])$. For which $p \in [1, \infty]$, if any, is it true that weak convergence of the sequence (meaning as functionals on L^q) implies strong convergence? (Prove or provide corresponding counterexamples).

Exercise 73. For $f, g \in L^1(\mathbb{R}^n)$ define the convolution

$$(f * g)(t) := \int_{-\infty}^{\infty} f(\tau) g(t - \tau)\, d\tau$$

Prove the following theorem, known as Young's inequality for convolution.

Theorem 26.4.3. Let $p, q, r \in [1, \infty]$ satisfy $p^{-1} + q^{-1} = 1 + r^{-1}$. For $f \in L^p(\mathbb{R}^n) + L^1(\mathbb{R}^n)$ and $g \in L^q(\mathbb{R}^n)$, we have $f * g \in L^r(\mathbb{R}^n)$ and

$$\|f * g\|_r \leqslant \|f\|_p \|g\|_q$$

Homework: Problems 20,22,35,41 in Folland, Chap. 6 and turn in Ex. 72 above.

Definition 26.4.4. Let T be now a map from a vector subspace $\mathcal{D}$ of $(X, \mathcal{M}, \mu)$ to the **measurable** functions on $(Y, \mathcal{N}, \nu)$. Let $p, q \in [1, \infty]$.

1. T is called sublinear if $|T(f+g)| \leqslant |Tf| + |Tg|$ and $|T(cf)| = c|Tf|$ for all $c > 0$.

2. A sublinear map T is of strong type (p, q) if $\mathcal{D} \supset L^p(\mu)$ and $\|Tf\|_q \leqslant \|f\|_p \leqslant C$ for some $C \in \mathbb{R}^+$ and all $f \in L^q$. We will abbreviate this by $\|T\|_{L^p \to L^q} \leqslant C$.

3. The sublinear map T is of weak type (p,∞) if it is of strong type (p,∞). If $q<\infty$ the sublinear map T is of weak type (p,q) if $\mathcal{D}\supset L^p(\mu)$, T maps $L^p(\mu)$ to $L^{q,w}(\nu)$ and $[Tf]_q \leqslant C\|f\|_p$ for some $C\in\mathbb{R}^+$ and all $f\in L^p$.

Theorem 26.4.5 (The Marcinkiewicz interpolation theorem). Let $(X,\mathcal{M},\mu)$ and $(Y,\mathcal{N},\nu)$ be σ-finite measure spaces and let $1\leqslant p_0\leqslant p_1\leqslant\infty$, $1\leqslant q_0\leqslant q_1\leqslant\infty$ and assume further that $p_0\leqslant q_0, p_1\leqslant q_1, q_0\neq q_1$. Let $\mathcal{D}=L^{p_0}(\mu)+L^{p_1}(\mu)$ and T be sublinear from $\mathcal{D}$ to Y be of weak types (p_0,q_0) and (p_1,q_1). Denote

$$\frac{1}{p_t}=:\frac{1}{p}=\frac{1-t}{p_0}+\frac{t}{p_1};\quad \frac{1}{q_t}=:\frac{1}{q}=\frac{1-t}{q_0}+\frac{t}{q_1}\qquad (t\in(0,1)) \tag{115}$$

Then T is strong type (p,q). More precisely, if $[Tf]_{q_0}\leqslant C_0\|f\|_{p_0}$ and $[T]_{q_1}\leqslant C_1\|f\|_{p_1}$, then $\|T\|_{L^p\to L^q}\leqslant B$ where B depends on $p,p_0,p_1,p_1,q_1,q_0,q_1$. As $p\to p_j$, $B=O\left(t^{-\frac{1}{q}}(1-t)^{-\frac{1}{q}}\right)$, $j=0,1$.

Proof. The case $p_0=p_1$ is an easy version of the proof for $p_0<p_1$, that we assume. We also take $q_0,q_1<\infty$ for the moment. With p,q as in (115) and $f\in L^p(\mu)$, we estimate $\|Tf\|_q$ by decomposing first f as in Proposition 26.3.7: $f=g_A+h_A$, and use distribution functions to link $L^{p,w}$ and L^p estimates.

We write for the norm $\|T\|_q^q$,

$$\|T\|_q^q=q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(\alpha)d\alpha=2^q q\int_0^\infty \alpha^{q-1}\lambda_{Tf}(2\alpha)d\alpha \tag{116}$$

where we wrote 2α to use 4. in Proposition 26.3.1:

$$\lambda_{Tf}(2\alpha)\leqslant\lambda_{Tg_A}(\alpha)+\lambda_{Th_A}(\alpha) \tag{117}$$

We now link the p_0 norm of g_A,h_A to λ_f, with the aim at ultimately finding a bound in terms of $\|f\|_p$.

We have

$$\|h_A\|_{p_1}^{p_1}=p_1\int_0^\infty \beta^{p_1-1}\lambda_{h_A}(\beta)d\beta=p_1\int_0^A\beta^{p_1-1}\lambda_f(\beta)d\beta \tag{118}$$

and similarly,

$$\|g_A\|_{p_0}^{p_0}=p_0\int_0^\infty \beta^{p_0-1}\lambda_f(\beta+\alpha)d\beta=p_0\int_A^\infty(\beta-A)^{p_0-1}\lambda_f(\beta)d\beta\leqslant p_0\int_A^\infty\beta^{p_0-1}\lambda_f(\beta)d\beta \tag{119}$$

We now estimate the contribution of λ_{Th_A}, via (117), to $\|T\|_q^q$ and the weak norm estimate. For any $\alpha>0$ we have

$$C_1\|h_A\|_{p_1}\geqslant[Th_A]_{p_1}\geqslant\lambda_{Th_A}^{1/p_1}\alpha$$

hence

$$\lambda_{Th_A}(\alpha)\leqslant\alpha^{-q_1}C_1^{q_1}\|h_A\|_{p_1}^{q_1}\leqslant\alpha^{-q_1}C_1^{q_1}p_1^{q_1}\left(\int_0^A\beta^{p_1-1}\lambda_f(\beta)d\beta\right)^{q_1/p_1} \tag{120}$$

hence

$$2^q q\int_0^\infty \alpha^{q-1}\lambda_{Th_A}(\alpha)d\alpha \leqslant 2^q q C_1^{q_1} p_1^{q_1/p_1}\int_0^\infty \alpha^{q-q_1-1}\left(\int_0^A \beta^{p_1-1}\lambda_f(\beta)d\beta\right)^{q_1/p_1} d\alpha \tag{121}$$

This is true for any $A>0$, and the optimal choice turns out to be $A=\alpha^\sigma$ where

$$\sigma := \frac{p_0(q_0-q)}{q_0(p_0-p)} = \frac{p^{-1}(q^{-1}-q_0^{-1})}{q^{-1}(p^{-1}-p_0^{-1})} = \frac{p^{-1}(q^{-1}-q_1^{-1})}{q^{-1}(p^{-1}-p_1^{-1})} = \frac{p_1(q_1-q)}{q_1(p_1-p)} \tag{122}$$

Now we apply Minkowski's inequality to switch the order of integration in the estimates:

$$\int_0^\infty \alpha^{q-q_1-1}\left(\int_0^\infty \chi_{\beta<\alpha^\sigma}\beta^{p_1-1}\lambda_f(\beta)d\beta\right)^{\frac{q_1}{p_1}} d\alpha \leqslant \left[\int_0^\infty\left(\int_0^\infty \chi_{\alpha^\sigma>\beta}\alpha^{\frac{q_1}{p_1}(q-q_1-1)}d\alpha\right)^{\frac{p_1}{q_1}}\beta^{p_1-1}\lambda_f(\beta)d\beta\right]^{\frac{q_1}{p_1}} \tag{123}$$

Take first $q_1>q_0$. In this case $q-q_0>0, \sigma>0$ and $\alpha^\sigma>\beta \Leftrightarrow \alpha>\beta^{\frac{1}{\sigma}}$. Hence,

$$\int_0^\infty\left(\int_0^\infty \chi_{\alpha^\sigma>\beta}\alpha^{\frac{q_1}{p_1}(q-q_1-1)}d\alpha\right)^{\frac{p_1}{q_1}}\beta^{p_1-1}\lambda_f(\beta)d\beta = \int_0^\infty\left(\int_{\beta^{\frac{1}{\sigma}}}^\infty \alpha^{\frac{q_1}{p_1}(q-q_1-1)}d\alpha\right)^{\frac{p_1}{q_1}}\beta^{p_1-1}\lambda_f(\beta)d\beta$$
$$= (q-q_1)^{-p_1/q_1}\int_0^\infty \beta^{p_1-1+p_1(q-q_1)/q_0\sigma}\lambda_f(\beta)d\beta = |q-q_1|^{-p_1/q_1}\int_0^\infty \beta^{p-1}\lambda_f(\beta)d\beta = \frac{\|f\|_p^p}{p|q-q_1|^{p_1/q_1}} \tag{124}$$

Note that this is now a norm estimate. Similar calculations show that the inequality above holds when $q_1<q_0$ and that the counterpart integral for g_A is bounded by

$$\frac{\|f\|_p^p}{p|q-q_0|^{p_0/q_0}} \tag{125}$$

Combining the estimates, we get

$$\|Tf\|_q \leqslant B\|f\|_p;\quad B = 2q^{1/q}\left(\frac{C_0^{q_0}(p_0/p)^{q_0/p_0}}{|q-q_0|}+\frac{C_1^{q_1}(p_1/p)^{q_1/p_1}}{|q-q_1|}\right)^{1/q} \tag{126}$$

The remaining range of p,q only requires small modifications, basically in the choice of A, which is set to solve the equation $\lambda_{Th_A}(\alpha)=0$ in the cases $p_1=q_1=\infty$ and $q_0<q_1=\infty, p_0<p_1<\infty$, and from the equation $\lambda_{Tg_A}(\alpha)=0$ if $q_1<q_0=\infty, p_0<p_1<\infty$. See details in Folland. □

26.5 Some applications

1. The Hilbert transform This is an important operator in a number of areas of mathematics.

Definition 26.5.1. The Hilbert transform of a function f is defined to be

$$(\mathcal{H}f)(x) = \frac{1}{\pi}\,\mathrm{p.\,v.}\int_{-\infty}^{+\infty}\frac{f(\tau)}{x-\tau}\,d\tau := \frac{1}{\pi}\lim_{\varepsilon\to 0}\int_{|t|>\varepsilon} f(x-t)\frac{dt}{t} \tag{127}$$

whenever the limit exists.

It is easy to see that the limit does exist if f is a smooth function, but it is far from clear if $\mathcal{H}$ makes sense on L^p spaces.

The Calderon-Zygmund decomposition, a fundamental technique in harmonic analysis, is used to show that $\mathcal{H}$ is bounded from L^1 into $L^{1,w}$. The approach is somewhat similar to that used for the weak estimates of the Hardy-Littlewood maximal operator. Using the Fourier transform we will show that $\mathcal{H}$ is bounded from L^2 into L^2. The Marcinkiewicz interpolation theorem entails that $\mathcal{H}$ is bounded in all $L^p, p \in (1,2]$. Now,

$$\int_X g\mathcal{H}f d\mu = -\int_X f\mathcal{H}g d\mu$$

shows boundedness from L^p into L^p $p \in (2,\infty)$.

2. The Hardy-Littlewood maximal operator, M. (See Definition 15.0.6.) We have

Proposition 26.5.2. *M is bounded from from $L^1(\mathbb{R}^d)$ into $L^{1,w}(\mathbb{R}^d)$ and from $L^p(\mathbb{R}^d)$ into $L^p(\mathbb{R}^d)$, for $d \geqslant 1$ and $1 < p \leqslant \infty$.*

Proof. Theorem 15.0.8 exactly states that M is bounded from L^1 into $L^{1,w}$, and it is obvious that M is bounded between L^∞ and L^∞. The rest is a straightforward application of the Marcinkiewicz interpolation theorem. □

27 Radon measures

In order to better understand properties of various mathematical objects it is often very useful to analyze the natural functions (ones compatible with the structure) defined on them. These would be linear functionals on topological vector spaces, representations in the case abstract algebraic structures and in the case of topological spaces, the space of *continuous* functions defined on them (in fact specifying the continuous functions determines the topology). We can go one step further, look at continuous functions as a topological space (in the topologies mentioned in the previous section) and analyze its dual.

We will focus on $C_c(\mathbb{X})$ and $C_0(\mathbb{X})$, see Definition 17.5.1, where $\mathbb{X}$ is LCH. Roughly, it turns out that the continuous functionals on C_0 are given by finite measures with nice regularity properties (finite Radon measures), and that any finite Borel measure on such spaces $\mathbb{X}$ is Radon, thus regular.

Notation. In this section $\mathbb{X}$ will denote a locally compact Hausdorff space, $\mathcal{O}$ will denote open sets and K will denote compact sets. Two important subspaces of continuous functions on $\mathbb{X}$ are $C_c(\mathbb{X})$ and $C_0(\mathbb{X})$; we start with $C_c(\mathbb{X})$, see Definition 17.5.1. Recall Urysohn's lemmma, partitions of unity and the symbol $\prec$ indicating the support of a function. When we write $K \prec \varphi$ or $\varphi \prec \mathcal{O}$ it will be understood that $\varphi \in C_c(\mathbb{X}, [0,1])$, $\varphi = 1$ on K and zero outside $\mathcal{O}$.

Definition 27.0.1.

1. A measure μ is called locally finite if $\mu(K) < \infty$ for any compact K.
2. Recall that, if $(\mathbb{X}, \mathcal{M}, \mu)$ is a measure space, with $\mathbb{X}$ LCH, then μ is called inner regular if for all $E \in \mathcal{M}$ we have $\mu(E) = \sup_{K \subset E} \mu(K)$, outer regular if $\mu(E) = \inf_{\mathcal{O} \supset E} \mu(\mathcal{O})$ and regular if it is bouth inner and outer regular.
3. Recall also that, if μ^* is an outer measure on $\mathcal{M}$, then E is called μ^* measurable if
$$\forall A \subset \mathbb{X},\ \mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c) \tag{128}$$
4. A linear functional Λ on a space of functions $\mathcal{D}$ is a positive linear functional if $\Lambda f \geqslant 0$ for any $f \geqslant 0$ in $\mathcal{D}$. Of course, this is the same as requiring
$$f \leqslant g \Rightarrow \Lambda f \leqslant \Lambda g, \quad \forall f, g \in \mathcal{D}$$

The following continuity property is automatic from positivity.

Proposition 27.0.2. Let Λ be a positive linear functional on $C_c(\mathbb{X})$ and let $K \subset \mathbb{X}$ be compact. There is a $C_K \geqslant 0$ s.t., for all f with support in K we have
$$|\Lambda f| \leqslant C_K \|f\|_K$$
where $\|f\|_K$ is the sup norm on K.

Proof. Since we can write $f = u^+ - u^- + i(v^+ - v^-)$ where the functions in the decomposition are continuous and nonnegative, it is enough to prove the result when f itself is nonnegative. Fix a $\varphi \in C_c(\mathbb{X}, [0,1])$ s.t. $\varphi(K) = \{1\}$, i.e. $K \prec \varphi$. Then, with $n = \|f\|_K$, we have
$$f = \varphi f \leqslant n\varphi \ \text{ hence } \ \Lambda f \leqslant n \Lambda \varphi = C_K \|f\|_K; \ \ C_K = \Lambda \varphi$$
□

Candidates for positive functionals on $C_c(\mathbb{X})$ are integrals with respect to positive Borel measures,
$$\Lambda f = \int_{\mathbb{X}} f d\mu \tag{129}$$

where μ must be locally finite:

Proposition 27.0.3. If (129) is a positive linear functional on $C_c(\mathbb{X})$, then $\mu(K) < \infty$ for any K.

Proof. Using the density of $C(K)$, thus of $C_c(\mathbb{X})$, in $L^1(\mu, K)$, we extend by continuity Λ to $L^1(\mu, \mathbb{X})$; clearly positivity is preserved. Take a compact K and an $f \in C_c(\mathbb{X}, [0,1])$ s.t. $K \prec f$. Then $\chi_K \leqslant f$ and

$$0 \leqslant \mu(K) = \int_{\mathbb{X}} \chi_K d\mu \leqslant \int_{\mathbb{X}} f d\mu < \infty$$

□

Relying on the fact that $\mathbb{X}$ is an LCH, the functionals are given by (129) with μ a *Radon measure* defined below.

In fact, we prove in the sequel that a positive linear functional naturally generates an **outer Radon measure** μ^*, defined as follows:

Definition 27.0.4. An outer measure μ^* is Radon if

1. For any compact K, $\mu^*(K) < \infty$[15](μ^* is locally finite).
2. any open set is μ^*-measurable. Thus Borel sets are μ^*-measurable.
3. $\forall E \subset \mathbb{X}, \mu^*(E) = \inf\{\mu^*(\mathcal{O}) : \mathcal{O} \supset E\}$ (outer regularity; as usual $\mathcal{O}$ denotes open sets)
4. $\forall \mathcal{O},\ \mu^*(\mathcal{O}) = \sup\{\mu^*(K) : K \subset \mathcal{O}\}$ (inner regularity on open sets).

By the Caratheodory theorem, μ defined as the restriction of μ^* to the σ-algebra $\mathfrak{M}$ of μ^*-measurable sets is a measure on $\mathfrak{M}$, and with μ^* a Radon outer measure, μ is called a **Radon measure**. Caratheodory's construction shows that Radon measures are complete.

Lemma 27.0.5. A Radon measure is inner regular on all measurable sets of finite measure, and more generally on all measurable σ-finite sets.

Proof. Indeed, 1) assume first $m = \mu(E) < \infty$ and let $\mathcal{O} \supset E, \mu(\mathcal{O} \setminus E) < \varepsilon/2$, $\mathcal{O}' \supset \mathcal{O} \setminus E$, $\mu(\mathcal{O}') < \varepsilon$. Let $K \subset \mathcal{O}$, $\mu(K) > m - \varepsilon$. Then $K' = K \cap (\mathcal{O}')^c \subset E$ is compact and $\mu(K') > m - 2\varepsilon$.

2. Take now an E with $\mu(E) = \infty$. By assumption $E = \cup_{j\in\mathbb{N}} E_j$ where $\mu(\cup_{j\leqslant n} E_j) \to \infty$. By 1) above, there is a family $K_j \subset \cup_{j\in\mathbb{N}} E_j$ with $\mu(K_j) \to \infty$. □

[15]In more general spaces one requires that every point has a neighborhood of finite measure; for LCH this is equivalent to the given condition.

Exercise 74. *[When are sets outer-Radon measurable?]* If μ^* is a Radon outer measure, show that $E \subset \mathbb{X}$ is μ^*-measurable iff $E \cap K$ is measurable for every K. (Hint: Reduce the problem to measurability of all $E \cap \mathcal{O}$, $\mu(\mathcal{O}) < \infty$.)

Homework: Problems 1,2,3,4 in Folland, Chap. 7 and turn in Ex. 73, 74 above, due Fri March 8.

Theorem 27.0.6 (Riesz representation theorem)**.** Let Λ be a positive linear functional on $C_c(\mathbb{X})$. Then, there exists a unique Radon measure on a σ-algebra $\mathfrak{M} \supset \mathcal{B}(\mathbb{X})$ s.t. (129) holds.

Furthermore, for all $\mathcal{O}$

$$\mu(\mathcal{O}) = \sup_{\varphi \prec \mathcal{O}} \Lambda\varphi \tag{130}$$

and for all K

$$\mu(K) = \inf_{K \prec f} \Lambda\varphi \tag{131}$$

Proof. **1. Uniqueness**

Assume we have two measures $\mu_{1,2}$ with the properties above. Using outer regularity and inner regularity on open sets, it is enough to show they coincide on compact sets. Let K be arbitrary and $\mathcal{O} \supset K$ be s.t. $\mu_2(\mathcal{O}) < \mu_2(K) + \varepsilon$. Let $\varphi \in C_c(\mathbb{X}, [0,1])$ be s.t. $K \prec \varphi \prec \mathcal{O}$; reasoning as in Proposition 27.0.3, we have

$$\mu_1(K) = \int_{\mathbb{X}} \chi_K d\mu_1 \leqslant \int_{\mathbb{X}} \varphi d\mu_1 = \Lambda\varphi = \int_{\mathbb{X}} \varphi d\mu_2 \leqslant \int_{\mathbb{X}} \chi_{\mathcal{O}} d\mu_2 = \mu_2(\mathcal{O}) \leqslant \mu_2 + \varepsilon$$

and interchanging $1 \leftrightarrow 2$ we have $|\mu_1(K) - \mu_2(K)| < \varepsilon$. $\square$

Construction of μ and $\mathfrak{M}$

It is natural to define the following set function on open sets:

$$\mu(\mathcal{O}) = \sup\{\Lambda\varphi : \varphi \prec \mathcal{O}\} \tag{132}$$

Now we note that

$$\mu(\mathcal{O}) \leqslant \sum_{j \in \mathbb{N}} \mu(\mathcal{O}_j) \text{ if } \mathcal{O} \subset \cup_{j \in \mathbb{N}} O_j \tag{133}$$

Indeed, for any $\varphi \prec \mathcal{O}$, $K = \text{supp}\varphi \subset \cup_1^n \mathcal{O}_i$ for some n. With $\rho_i \prec \mathcal{O}_i$ a partition of unity, we see that $\varphi = \sum_1^n \varphi\rho_i$ and

$$\Lambda\varphi = \sum_{i=1}^n \Lambda(\varphi\rho_i) \leqslant \sum_{i=1}^n \mu(\mathcal{O}_i) \text{ since } \varphi\rho_i \prec \mathcal{O}_i$$

Therefore, by Proposition 3.1.2, the set function

$$\mu^*(E) = \inf\{\mu(\mathcal{O}) : E \subset \mathcal{O}\} = \inf\{\sum_j \mu(\mathcal{O}_j) : E \subset \cup_j \mathcal{O}_j\} \text{ (by (133))}; \quad E \subset \mathbb{X} \tag{134}$$

is an outer measure on $\mathcal{P}(\mathbb{X})$.

Note 27.0.7. It is clear that $\mu^*(\mathcal{O}) = \mu(\mathcal{O})$ for any open set $\mathcal{O}$. Remember also that, to check measurability, it is enough to show that the left side of (128) is $\geqslant$ its right side when $\mu^*(A) < \infty$.

◀

▶ Now we show that **open sets are μ^*-measurable**.
▹ We start by showing that for any open sets $\mathcal{O}, \mathcal{O}'$ we have

$$\mu^*(\mathcal{O}') = \mu^*(\mathcal{O}' \cap \mathcal{O}) + \mu^*(\mathcal{O}' \setminus \mathcal{O}) \text{ or equivalently } \mu(\mathcal{O}') = \mu(\mathcal{O}' \cap \mathcal{O}) + \mu^*(\mathcal{O}' \setminus \mathcal{O}) \tag{135}$$

Take $K \prec \varphi \prec \mathcal{O}' \cap \mathcal{O}$ s.t. $\Lambda\varphi > \mu(\mathcal{O}' \cap \mathcal{O}) - \varepsilon$ and $\psi \prec \mathcal{O}' \setminus K$ s.t. $\Lambda\psi > \mu(\mathcal{O}' \setminus K) - \varepsilon$. Clearly, $\varphi + \psi \prec \mathcal{O}'$. Hence,

$$\mu(\mathcal{O}') \geqslant \Lambda\varphi + \Lambda\psi \geqslant \mu(\mathcal{O}' \cap \mathcal{O}) + \mu(\mathcal{O}' \setminus K) - 2\varepsilon$$

Noting now that $\mathcal{O}' \setminus \mathcal{O} = \mathcal{O}' \setminus (\mathcal{O}' \cap \mathcal{O}) \subset \mathcal{O}' \setminus K$ we see that $\mu^*(\mathcal{O}' \setminus \mathcal{O}) \leqslant \mu^*(\mathcal{O}' \setminus K)$ and (135) follows.
▹ Assume now $\mu^*(A) < \infty$ and take $\mathcal{O}' \supset A$ s.t. $\mu^*(A) \geqslant \mu(\mathcal{O}') - \varepsilon$. Then,

$$\mu^*(A) \geqslant \mu(\mathcal{O}') - \varepsilon = \mu^*(\mathcal{O}' \cap \mathcal{O}) + \mu^*(\mathcal{O}' \cap \mathcal{O}^c) - \varepsilon \geqslant \mu^*(A \cap \mathcal{O}) + \mu^*(A \cap \mathcal{O}^c) - \varepsilon \tag{136}$$

Note 27.0.8. At this stage, applying the Caratheodory theorem, we see that μ **is a measure** on a $\sigma-$ algebra which contains the open sets, and hence it contains $\mathcal{B}(\mathbb{X})$.

▶ μ **satisfies** (131) Take $\varepsilon \in (0,1)$, φ s.t. $K \prec \varphi$, and define $\mathcal{O}_\varepsilon = \{x : \varphi(x) > 1 - \varepsilon\}$. For any $\psi \prec \mathcal{O}_\varepsilon$ we have $\psi \leqslant (1-\varepsilon)^{-1}\varphi$, implying

$$\mu(K) \leqslant \mu(\mathcal{O}_\varepsilon) = \sup_{\psi \prec \mathcal{O}_\varepsilon} \Lambda\psi \leqslant (1-\varepsilon)^{-1}\Lambda\varphi$$

(in particular $\mu(K) < \infty$). In the opposite direction, we want to find an φ, $K \prec \varphi$ s.t. $\mu(K) \geqslant \Lambda\varphi - \varepsilon$. Let $\mathcal{O} \supset K$ be s.t. $\mu(\mathcal{O}) \leqslant \mu(K) + \varepsilon$ and take $K \prec \varphi \prec \mathcal{O}$. Then $\Lambda\varphi \leqslant \mu(\mathcal{O}) \leqslant \mu(K) + \varepsilon$ as desired. ◀

▶ μ **is inner regular on open sets**. Let $m < \mu(\mathcal{O})$; choose $\varphi \prec \mathcal{O}$ s.t. $\Lambda\varphi > m$, and let $K = \text{supp}(\varphi)$. For any $\mathcal{O}' \supset K$ we have $\varphi \prec \mathcal{O}'$, hence $\mu(\mathcal{O}') \geqslant \Lambda\varphi$ entailing $\mu(K) \geqslant \Lambda\varphi > m$. ◀

▶ $\boxed{(\forall\varphi \in C_c(\mathbb{X}))(\Lambda\varphi = \int_{\mathbb{X}} \varphi d\mu)}$ We can assume $\varphi \in C_c(\mathbb{X}, [0,1])$. Let $K = \text{supp}(\varphi)$, $\varepsilon > 0$ and

$$y_1 < 0 < y_2 < \cdots < 1 < y_n \text{ be s.t } \forall i \max_i\{y_i - y_{i-1}\} < \varepsilon \text{ and } \mu(\varphi^{-1}(\{y_i\})) = 0 \text{ [16]} \tag{137}$$

Let $\mathcal{O} \supset K$, $\mu(\mathcal{O}) < \infty$. Then $\varphi^{-1}((y_{i-1}, y_i)) \cap \mathcal{O} := \mathcal{O}_i$ are open, mutually disjoint and $\mu(K \setminus \cup_i \mathcal{O}_i) = 0$, by (137). For $i = 1, ..., n$ choose $K_i \subset \mathcal{O}_i$ so that $\mu(\mathcal{O}_i \setminus K_i) < \frac{\varepsilon}{n}$ and ψ_i s.t. $K_i \prec \psi_i \prec \mathcal{O}_i$. If $\frac{\varepsilon'}{n} := \mu(O_i) - \Lambda\psi_i$, then $\varepsilon' < \varepsilon$. By the mean value theorem, $\forall i\ \exists v_i \in [y_{i-1}, y_i]$ s.t

$$\int_{\mathbb{X}} \varphi d\mu = \sum_i \int_{\mathcal{O}_i} \varphi d\mu = \sum_{i=1}^n v_i \mu(\mathcal{O}_i) = \Lambda\left(\sum_{i=1}^n v_i\psi_i\right) - \varepsilon' \Rightarrow \left|\int_{\mathbb{X}} \varphi d\mu - \Lambda \sum_{i=1}^n \psi_i v_i\right| < \varepsilon \tag{138}$$

Write $\varphi - \sum_i v_i\psi_i = \varphi_1 + \varphi_2$ with $\varphi_1 = \sum_i(\varphi - v_i)\psi_i$ and $\varphi_2 = \varphi - \varphi\sum_i \psi_i$. By (138), $\|\varphi_1\|_u < \varepsilon$, hence $|\Lambda\varphi_1| < \varepsilon$. Now $\|\varphi_2\|_u \leqslant 1$ and $\varphi_2 \prec \cup_i(\mathcal{O}_i \setminus K_i)$; hence, by (131), $|\Lambda\varphi_2| < \varepsilon$. The triangle inequality and (138) now give

$$\left|\int_{\mathbb{X}} \varphi d\mu - \Lambda\varphi\right| < 3\varepsilon \text{ ◀}$$

This completes the proof of Theorem 27.0.6. □

Proposition 27.0.9. Assume $\mathbb{X}$ is σ-compact. Let μ be a Radon measure and $\mathfrak{M}$ be the σ-algebra of μ-measurable sets. Then

(a) For any $E \in \mathfrak{M}$ and $\varepsilon > 0$ there is a closed set $\mathcal{C}$ and an open $\mathcal{O}$ s.t. $\mathcal{C} \subset E \subset \mathcal{O}$ and $\mu(\mathcal{O} \setminus \mathcal{C}) < \varepsilon$.

(b) μ is a regular Borel measure.

(c) If $E \in \mathfrak{M}$, then there is a pair (F, G) of F_σ, G_δ sets s.t. $F \subset E \subset \mathcal{O}$ and $\mu(\mathcal{O} \setminus F) = 0$.

Proof. Let $X = \cup_n K_n$ where K_n are compact. Let $E \in \mathfrak{M}$. Clearly, $\mu(E \cap K_n) < \infty$ and thus, by outer regularity, for any $\varepsilon > 0$, there are $\mathcal{O}_n \supset E \cap K_n$ with $\mu(\mathcal{O}_n \setminus [E \cap K_n]) < \varepsilon 2^{-n-1}$. With $\mathcal{O} = \cup_n \mathcal{O}_n$, we have $\mathcal{O} \setminus E \subset \cup_n(\mathcal{O}_n \setminus [E \cap K_n])$ and thus

$$\mu(\mathcal{O} \setminus E) < \varepsilon/2$$

The same is true for E^c, and thus there is an open set $\mathcal{O}' \supset E^c$ s.t. $\mu(\mathcal{O}' \setminus E^c) < \varepsilon/2$. If $C = (\mathcal{O}')^c$, then C is closed and $E \setminus C = E \cap \mathcal{O}' = \mathcal{O}' \setminus E^c$ implying the result.

Note that every closed set $\mathcal{C}$ is σ-compact, since $C = \cup(C \cap K_n)$. By continuity from below, $\mu(C) = \lim_n \mu(\cup_{j=1}^n [C \cap K_j])$ proving inner regularity of closed sets, thus by (a), of all sets.

(c) Apply (a) with $\varepsilon = j^{-1}, j \in \mathbb{N}$: there exist $C_j \subset E \subset \mathcal{O}_j$ s.t. $\mu(\mathcal{O}_j \setminus C_j) < \varepsilon$. Now $F = \cup F_j \subset E \subset G = \cap \mathcal{O}_j$ and the result follows. □

27.1 The Baire σ-algebra

Another natural σ-algebra when studying $C_c(\mathbb{X})$ is the Baire σ-algebra $\mathcal{B}_0(\mathbb{X})$, the smallest σ-algebra with respect to which all functions in $C_c(\mathbb{X})$ are measurable. The elements of $\mathcal{B}_0(\mathbb{X})$ are

[16] This is possible, since otherwise $\mu(K) = \infty$ in contradiction with (131).

called Baire sets. Clearly $\mathcal{B}_0(\mathbb{X}) \subset \mathcal{B}(\mathbb{X})$; the two coincide if $\mathbb{X}$ is second countable (see Exercise 5/p. 216 in Folland).

27.2 Regularity of Borel measures

In this section we assume that $\mathbb{X}$ has the additional property that

$$\text{every open set } \mathcal{O} \subset \mathbb{X} \text{ is } \sigma\text{-compact} \tag{139}$$

This is the case if $\mathbb{X}$ is second countable.

Theorem 27.2.1. Assume $\mathbb{X}$ satisfies (139). Then, every locally finite Borel measure λ on $\mathbb{X}$ is regular (and thus Radon).

Proof. The functional $\Lambda f = \int_{\mathbb{X}} f d\lambda$ is well-defined on $C_c(\mathbb{X})$ (since continuous functions are measurable, and $f = 0$ outside K implies $|f| \leqslant \|f\| \mathcal{X}(K) \Rightarrow \Lambda|f| \leqslant \|f\| \lambda(K)$). Then, there is a regular Radon measure μ s.t.

$$\int_{\mathbb{X}} f d\lambda = \int_{\mathbb{X}} f d\mu$$

We now show that $\lambda = \mu$.

Take an open set $\mathcal{O}$, and, recalling that $\mathbb{X}$ is an LCH, let $\mathcal{O} = \cup_{j \in \mathbb{N}} K_j$, as in Proposition 17.2.9, where the compact sets K_j can be arranged to be increasing, and then $K_i \nearrow \mathcal{O}$. For each i, let $K_i \prec \varphi_i \prec \mathcal{O}$. Now, since $\chi_{K_i} \leqslant \varphi_i \leqslant \chi_O$ we have $\varphi_i \to \chi(\mathcal{O})$; defining $g_k = \max_{j \leqslant k} \varphi_j$, we have $g_k \nearrow \chi_{\mathcal{O}}$ as $k \to \infty$, and by the monotone convergence theorem,

$$\lambda(\mathcal{O}) = \lim_{k \to \infty} \int_{\mathbb{X}} g_k d\lambda = \lim_{k \to \infty} \int_{\mathbb{X}} g_k d\mu = \mu(\mathcal{O}) \tag{140}$$

Now, with $E \in \mathcal{B}(\mathbb{X})$ arbitrary, by the regularity of the measure μ, for any $\varepsilon > 0$, there is a pair $C \subset E \subset \mathcal{O}$ with $\varepsilon \geqslant \mu(\mathcal{O} \setminus C) = \lambda(\mathcal{O} \setminus C)$ (since $\mathcal{O} \setminus C$ is open). If $\mu(\mathcal{O}) = \infty$ then $\mu(E) = \lambda(E) = \infty$. Otherwise, $\lambda(\mathcal{O} \setminus E) \leqslant \lambda(\mathcal{O} \setminus C) = \mu(\mathcal{O} \setminus C) \leqslant \varepsilon$, hence

$$|\mu(\mathcal{O}) - \mu(E)| \leqslant \varepsilon \text{ and } |\lambda(\mathcal{O}) - \lambda(E)| \leqslant \varepsilon \Rightarrow |\mu(E) - \lambda(E)| < 2\varepsilon$$

□

Corollary 27.2.2. Locally finite Borel measures on $\mathbb{R}^n$ are regular.

Proposition 27.2.3. If μ is a Radon measure on $\mathbb{X}$, then $C_c(\mathbb{X})$ is dense in $L^p, 1 \leqslant p < \infty$.

Proof. Given the density of simple functions, it suffices to show that χ_E can be approached arbitrarily in p norm, when $\mu(E) < \infty$. Take then $K \subset E \subset \mathcal{O}$ with $\mu(\mathcal{O} \setminus K) < \varepsilon$ and let $K \prec \varphi \prec \mathcal{O}$.

Then, φ and $\mathcal{X}_E$ are equal outside of $\mathcal{O} \setminus K$, and their difference is at most one. This means that $\|\varphi - \chi_E\|_p < \varepsilon^{1/p}$. ☐

Note 27.2.4. Recall that, if $\mathbb{X}$ is LCH, then $C_0(\mathbb{X})$ is a closed subspace of $\mathrm{BC}(\mathbb{X})$ w.r.t. $\|\cdot\|_u$, thus it is a Banach space w.r.t. $\|\cdot\|_u$ space, and that $C_c(\mathbb{X})$ is dense in $C_0(\mathbb{X})$.

28 The dual of $C_0(\mathbb{X})$

Let's first determine what are the **positive**, continuous linear functionals on $C_0(\mathbb{X})$. Let Λ be such a functional; clearly its restriction to $\varphi \in C_c(\mathbb{X})$ is a positive linear functional and thus

$$\Lambda\varphi = \int_{\mathbb{X}} \varphi d\mu; \quad (\forall \varphi \in C_c(\mathbb{X})) \tag{141}$$

where μ is a Radon measure. Since $\mathbb{X}$ is LCH, $C_c(\mathbb{X})$ is dense in $C_0(\mathbb{X})$, so the question is which Λ as (141) extend to continuously $C_0(\mathbb{X})$. Assume Λ does indeed extend continuously and let $\varphi \prec \mathbb{X}$. Clearly, $\|\varphi\|_u \leqslant 1$ and, by (130)

$$\mu(\mathbb{X}) = \sup_{\varphi \prec \mathbb{X}} \Lambda\varphi \leqslant \|\Lambda\| \|\varphi\|_u \leqslant \|\Lambda\| < \infty \tag{142}$$

Conversely, if $\mu(\mathbb{X}) < \infty$, then Λ in (141) has norm at most $\mu(\mathbb{X})$.

Definition 28.0.1. A measure s.t. (142) holds is called a finite Radon measure.

We found:

Proposition 28.0.2. Λ is a continuous positive linear functional on $C_0(\mathbb{X})$ **iff** it is given by (141) for a finite Radon measure μ.

We now turn to general, complex, continuous linear functionals on $C_0(\mathbb{X})$, that is, we want to find $(C_0(\mathbb{X}))^*$. Since the real and imaginary part of a continuous linear functional are real-valued continuous linear functionals, it suffices to determine these. We will see that, by an appropriate decomposition of the functional, the real-valued continuous linear functionals are still of the form (130), for a signed measure μ s.t. $|\mu|(\mathbb{X}) < \infty$.

Definition 28.0.3. A subset C of a vector space V is a reproducing positive cone if

1. $x, y \in C$ and $a, b \geqslant 0$ imply $ax + by \in C$
2. $C \cap (-C) = 0$

3.$\forall z \in V \exists x, y \in C$ s.t. $z = x - y$

It is easy to check that $C_0^+ = C_0(\mathbb{X}, [0, \infty))$ is a reproducing positive cone in $C_0(\mathbb{X})$. For, if $f, g \in C_0^+$, $\max\{f, g\} \in C_0^+$, $\min\{f, g\} \in C_0^+$.

Lemma 28.0.4. Let C be a reproducing positive cone in V. Any additive function $L : C \to [0, \infty)$, i.e. an L s.t., for all $a, b \geqslant 0$ and $x, y \in C$, $L(ax + by) = aLx + bLy$, extends as a linear functional on V.

Proof. Note that if x, x', y, y' are in C and $x - y = x' - y'$ then $Lx - Ly = Lx' - Ly'$ (apply L to $x + y' = x' + y$). If, for $z \in V$, we set $Lz = Lx - Ly$ where $z = x - y$, then L is well defined and linear, and extends L from C to V as it is easy to check. □

The following constructions and proofs are motivated by the expectation, that, in analogy with the Riesz representation theorem above, continuous functionals on $C_0(\mathbb{X})$ should be in bijection with complex measures on $\mathbb{X}$.

Lemma 28.0.5. If $\Lambda \in C_0^*(\mathbb{X})$ is real-valued, then there exist positive functionals $\Lambda^\pm \in C_0^*(\mathbb{X})$ s.t. $\Lambda = \Lambda^+ - \Lambda^-$.

Proof. Define first Λ^+ on the cone C_0^+ by

$$\Lambda^+ f = \sup_{g \in C_0^+, g \leqslant f} \Lambda g, \qquad \text{for } f \in C_0^+ \tag{143}$$

Check that $f \in C_0^+ \Rightarrow \Lambda^+ f \geqslant 0$. We show that

$$f, g \in C_0^+ \text{ and } a, b \geqslant 0 \Rightarrow \Lambda^+(af + bg) = a\Lambda^+ f + b\Lambda^+ g$$

The fact that $\Lambda^+(|a|f) = |a|\Lambda^+ f$ follows from (143). It remains to check that $\Lambda^+(f_1 + f_2) = \Lambda^+ f_1 + \Lambda^+ f_2$ on C_0^+. The key observation here is that $g \leqslant f_1 + f_2$ in C_0^+ iff $\exists g_1, g_2 \in C_0^+$ s.t. $g = g_1 + g_2$ and $g_i \leqslant f_i$ [17].) Extend Λ^+ as Lemma 28.0.4. Now, $\Lambda^- := \Lambda^+ - \Lambda$ is evidently linear and positive, and thus Λ is the difference of two positive functionals.

Clearly, $\|g\| \leqslant \|f\|$ whenever $0 \leqslant g \leqslant f$. Since since $|\Lambda^+ f| \leqslant \sup_{0 \leqslant g \leqslant f} |\Lambda g| \leqslant \|\Lambda\| \|f\|$, we have $\|\Lambda^\pm\| \leqslant \|\Lambda\|$. □

Exercise 75. Let μ be a finite signed Radon measure and $\Lambda f = \int_{\mathbb{X}} f d\mu$. Let $\mu = \mu^+ - \mu^-$ be the Hahn-Jordan decomposition of μ. Show that the linear functional Λ^+ obtained in Lemma 28.0.5 is given by $\Lambda^+ f = \int_{\mathbb{X}} f d\mu^+$.

[17] In the less obvious direction, define $g_1 = \min\{g, f_1\}$, $g_2 = g - g_1$, and check the condition.

Note 28.0.6 (Reminder: complex measures). 1. If μ is a complex measure on a sigma-algebra $\mathcal{M}$ on $\mathbb{X}$, then its total variation measure $|\mu|$ is the positive finite measure given by

$$|\mu|(E) = \sup_{\oplus_i E_i = E} \sum_{i=1}^{\infty} |\mu(E_i)|; \quad E, E_1, E_2, \ldots \in \mathcal{M} \tag{144}$$

and we have

$$d\mu = e^{i\theta(x)} d|\mu| \text{ for some measurable } \theta : X \to [-\pi, \pi) \tag{145}$$

Definition 28.0.7. μ is a signed Radon measure if $\mu = \mu_1 - \mu_2$ and μ_1, μ_2 are Radon measures. μ is a complex Radon measure if μ is a complex measure (finite, in particular) and $\mu = \mu_1 + i\mu_2$ where μ_1, μ_2 are signed Radon measures.

Corollary 28.0.8. $\Lambda \in C_0^*(\mathbb{X})$ **iff** $\Lambda f = \int_{\mathbb{X}} f d\mu$ where μ is a complex Radon measure.

Proposition 28.0.9. $\|\mu\| = |\mu|(X)$ is a norm on the linear space $M(X)$ of complex Radon measures.

Proof.

$$\|\mu + \nu\| = |\mu + \nu|(X) = \sup_{\oplus_i E_i = X} \sum_i |\mu(E_i) + \nu(E_i)| \leqslant \sup_{\oplus_i E_i = X} \sum_i |\mu(E_i)| + \sup_{\oplus_i E_i = X} |\nu(E_i)| = \|\mu\| + \|\nu\|$$

and the rest is straightforward. □

Lemma 28.0.10. If μ is a complex Radon measure and $\Lambda = f \mapsto \int_{\mathbb{X}} f d\mu$, $\Lambda : C_0(\mathbb{X}) \to \mathbb{C}$, then

$$\|\mu\| = \|\Lambda\|$$

Proof. In one direction, with $\|f\| = 1$, $|\Lambda f| \leqslant \int_{\mathbb{X}} |f| d|\mu| \leqslant |\mu|(\mathbb{X}) = \|\mu\|$. In the opposite direction, by (145) $d|\mu| = v d\mu$ with $|v| = 1$. Let K be s.t $|\mu|(\mathbb{X} \setminus K) = \varepsilon$ and take a φ s.t. $K \prec \varphi$. Then,

$$\|\mu\| = \int_{\mathbb{X}} d|\mu| = \int_K d|\mu| + \varepsilon = \left| \int_K \varphi v d\mu \right| + \varepsilon \leqslant \int_{\mathbb{X}} |\varphi v| d|\mu| + \varepsilon \leqslant \|\Lambda\| + \varepsilon$$

□

Theorem 28.0.11 (The Riesz representation theorem)**.** The map $\mu \to \Lambda_\mu$ is an isometric isomorphism of $M(\mathbb{X})$ to $C_0^*(\mathbb{X})$.

Proof. The bijection was shown in Corollary , and Lemma 28.0.10 completes the proof. □

Locally finite Borel measure in $\mathbb{R}^n$ are Radon, as discussed. By Theorem 1.16 (Folland) μ is a locally finite Borel measure on $\mathbb{R}^+$ iff it is a Lebesgue-Stieltjes measure, that is it is given by $\mu((a,b]) = F(b) - F(a)$ for some right-continuous, increasing, bounded F. Thus $\Lambda \in C_0^*(\mathbb{R})$ iff $\Lambda f = \int_{\mathbb{R}} f dF$ for some $F = F_1 - F_2 + i(F_3 - F_4)$ with F_i as above.

Note that $C_0([0,1]) \subset L^2([0,1])$, and the continuous functionals on $L^2[0,1])$ are given by the Riesz representation theorem, $f \mapsto \Lambda_\varphi f = \int_{[0,1]} f\overline{\varphi} dm$ where m is the Lebesgue measure and $\varphi \in L^2$. Now $L^2([0,1]) \subset L^1([0,1])$ (by Cauchy-Schwarz) and thus the subclass of continuous functionals on $C_c([0,1])$ that extend to L^2 are generated by a subclass of measures μ s.t. $d\mu = \overline{\varphi}dm, \varphi \in L^2$. We may view then dF as a **generalization** of the differential of F. We'll make more sense of all this in distribution theory.

Definition 28.0.12. *The weak* topology on $M(\mathbb{X})$ is called* **the vague** *topology. It means $\mu_n \to \mu$ if $\int f d\mu_n \to \int f d\mu$ for all f.*

Exercise 76. *(a) Is $X = \mathbb{N}$ with the discrete topology a LCH space?*

(b) What is $C_0(X)^*$, if X is as in a)?

29 Fourier series, cont.

Note 29.0.1. Recall that

$$\mathcal{F} = f \mapsto \sum_{k \in \mathbb{Z}} c_k e^{2\pi i k x}, \; c_k = \int_{\mathbb{T}} f(s) e^{-2\pi i k s} ds = \langle f, e_k \rangle$$

is an isomorphism between $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$. Recall also that the Fourier series of a characteristic function $\mathcal{X}$ of an interval on $\mathbb{T}$ converges to $\mathcal{X}$ at any point of continuity of $\mathcal{X}$ and to $1/2$ otherwise. Finally, we know that, if $f \in C^n(\mathbb{T}), n \geqslant 1$, then the Fourier series of f converges pointwise uniformly to f, together with $n-1$ derivatives. We keep the notation $S_n(f)$ for the nth symmetric partial Fourier sum of (f).

Theorem 29.0.2 (The Riemann-Lebesgue Lemma, first iteration...)**.** *Assume $f \in L^1([-\pi, \pi])$. Then,*

$$\lim_{|n| \to \infty} \int_{-1/2}^{1/2} f(s) e^{2\pi i n s} ds = 0 \tag{146}$$

Proof. Take first $f \in L^2(\mathbb{T})$. The integral above equals $c_n = \langle f, e_n \rangle$. By Bessel's inequality, $\sum_{n \in \mathbb{Z}} |c_n|^2 < \infty$, in particular $c_n \to 0$ as $|n| \to \infty$. Since L^2 is dense in L^1, the result follows by an $\varepsilon/3$ argument. □

29.1 Pointwise convergence

Recall that the Dirichlet kernel is defined as

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

Proposition 29.1.1. For all $n \geqslant 1$, $\|D_n\|_1 \geqslant \frac{4}{\pi^2} \log n$.

By modifying slightly the proof below, you can show that

$$\lim_{n\to\infty} \frac{\|D_n\|_1}{\log n} = \frac{4}{\pi^2}$$

Proof. Let $m = 2n+1$, and make the change of variable $x = (2\pi)^{-1}s$. Since $|\sin s| \leqslant |s|$ we get

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{\sin(ms)}{\sin(s/2)}\right| ds = \frac{1}{\pi}\int_0^{\pi} \frac{|\sin(ms)|}{\sin(\frac{s}{2})} \geqslant \frac{2}{\pi}\int_0^{\pi} \frac{|\sin(ms)|}{s} = \frac{2}{\pi}\int_0^{m\pi} \frac{|\sin s|}{s} ds$$
$$\geqslant \frac{2}{\pi}\sum_{k=0}^{m-1} \frac{(-1)^k}{k+1}\int_{k\pi}^{(k+1)\pi} \sin s ds = \frac{4}{\pi^2}\sum_{j=1} \frac{1}{j} \geqslant \frac{4}{\pi^2}(\log m + \gamma) \quad (147)$$

where γ is the Euler constant. □

What this shows is that, for any fixed a, the family $\Lambda_{a;n} = f \mapsto S_n(f;a)$, $n \in \mathbb{N}$ is not norm-bounded over the Banach space $C(\mathbb{T})$. From this and the uniform boundedness principle we see that, for any a, there is at least one continuous function for which the Fourier series diverges at a, and, in fact, the family of continuous functions whose Fourier series converges at a is of first Baire category in $C(\mathbb{T})$.

Note 29.1.2. It is a deep theorem (Carleson, 1966) that, for a fixed function in L^p, $p \in (1,\infty)$ (in particular, continuous), the set of points where the symmetric Fourier series converges pointwise is of full measure. In the opposite direction, for any set of zero measure there is a continuous function whose Fourier series diverges on that set.

Proposition 29.1.3. If $f \in AC(\mathbb{T})$ and $f' \in L^2(\mathbb{T})$ (e.g. $f \in C^1(\mathbb{T})$), then

$$\lim_{n\to\infty} \|S_n(f,x) - f(x)\|_u = 0$$

Proof. Note first that, under these assumptions for f,

$$\int_{\mathbb{T}} f'(s)e^{-iks}ds = ik\int_{\mathbb{T}} f(s)e^{-iks}ds \Rightarrow S_n(f') = S_n(f)'$$

Since $f' \in L^2$, we have $\sum_{k\in\mathbb{Z}} |kc_k|^2 < \infty$. This implies that

$$\sum_{k\neq 0} |c_k| = \sum_{k\neq 0} |c_k|k(k^{-1}) \leqslant \left(\sum_{k\neq 0} |kc_k|^2\right)^{1/2} \left(\sum_{k\neq 0} k^{-2}\right)^{1/2} < \infty$$

Thus the Fourier series converges absolutely, and then uniformly by the Weierstrass M test. Since the S_n converge in L^2 to f, the pointwise limit is f as well. $\square$

Exercise 77. *Assume $\{g_n\}_{n\in\mathbb{N}}$ is a sequence of functions in $AC([-\pi,\pi])$ with L^2 derivatives. Assume further that $\lim_{n\to\infty} \|g_n\|_2 + \|g_n'\|_2 = 0$. Show that $\lim_{n\to\infty} \|g_n\|_\infty = 0$.*

30 The heat equation

The heat equation is a *parabolic* partial differential equation that describes the time variation of the temperature distribution $u(x,t)$ in a given region Ω:

$$\frac{\partial u}{\partial t} = \Delta u, \quad u(0,x) = u_0(x),\ x \in \Omega;\ \ u(t,\cdot)_{|\partial\Omega} = f \tag{148}$$

where Δ is the Laplacian and the spacial variables run over some domain $\Omega \subset \mathbb{R}^n$. Here u_0 is the initial condition, the temperature distribution at $t = 0$, and f is the boundary condition, the temperature distribution on $\partial\Omega$. The function u is assumed C^2 with continuous partial derivatives up to $\partial\Omega$.

Equilibrium distributions are time-independent solutions of (148), in the sense

$$\Delta u = 0, \quad x \in \Omega;\ \ u_{|\partial\Omega} = f \tag{149}$$

Proposition 30.0.1 (Uniqueness)**.** *If u_1, u_2 solve (148) or (149), then $u_1 = u_2$.*

Proof. If u_1, u_2 are solutions, then $u_1 - u_2 = v$ is a solution of the PDE with $v(0,x) = v_{|\partial\Omega} = 0$. We show that the only such solution is zero. The proof is based on the *energy method*. Start with (148), $u_0 = f = 0$, multiply by v and integrate over Ω:

$$\int_\Omega v\frac{\partial v}{\partial t}dV = \frac{1}{2}\frac{d}{dt}\int_\Omega v^2 dV = \int_\Omega v\Delta v dV = \int_\Omega [\nabla\cdot(v\nabla v) - (\nabla v)^2]dV \tag{150}$$

where we used the identity $\nabla\cdot(v\nabla v) = (\nabla v)^2 + v\Delta v$. Now, since $v = 0$ on $\partial\Omega$ the divergence theorem implies

$$\int_\Omega \nabla\cdot(v\nabla v)dS = \int_{\partial\Omega} v\nabla v\cdot dS = 0$$

and thus

$$\frac{d}{dt}\underbrace{\int_\Omega v^2 dV}_{\geqslant 0} = -2\int_\Omega (\nabla v)^2 dV \leqslant 0 \tag{151}$$

Since $\int_\Omega v^2 dV \geqslant 0$, is nonincreasing and vanishes at $t = 0$, it means $\int_\Omega v^2 dV = 0$ and thus $v = 0$ for all x, t. For (149), the left side of (151) is simply zero, giving $\nabla v = 0 \Rightarrow v = const = 0$. $\square$

With uniqueness settled, we now aim at finding solutions of the PDE. Begin with (148) in two dimensions, $\Omega = \mathbb{D}$, the unit disk. The equation becomes

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \quad u_{|\mathbb{T}} = f(\theta) \ \ (f \in C^2) \tag{152}$$

This equation also describes the electric potential $u(x, y)$ in a disk where charges are placed on $\mathbb{T}$ only, with a density f.

In polar coordinates we get

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r}\frac{\partial u}{\partial r} + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} = 0, \quad u_{|\mathbb{T}} = f(\theta) \tag{153}$$

A method of solving simple PDEs such as (153) is by separation of variables. Inserting $u(r, \theta) = R(r)T(\theta)$ in (153) and dividing by RT we get

$$r^2\frac{R''}{R} + r\frac{R'}{R} = -\frac{T''}{T} \tag{154}$$

Now we note that the left side of the equation above does not depend on θ and the right side does dot depend on r, and thus they are independent of both variables, hence constant, say λ

$$r^2\frac{R''}{R} + r\frac{R'}{R} = \lambda = -\frac{T''}{T} \tag{155}$$

The ODE $T'' = -\lambda T$ has the general solution $C_1 e^{i\sqrt{\lambda}t} + C_2 e^{-i\sqrt{\lambda}t}$. There are constraints on λ: T must be periodic of period 2π, and this means $\lambda = m^2, m \in \mathbb{Z}$, and then

$$T(\theta) = a_m e^{im\theta} + a_{-m}e^{-im\theta} \tag{156}$$

The R equation

$$r^2 R'' + rR' = \lambda R \tag{157}$$

is of Euler type, with solutions $R(r) = Ar^m + Br^{-m}$ if $m \neq 0$ and $R(r) = A + b\ln r$ for $m = 0$. We note that $\ln r$ and r^{-m} for $m > 0$ as well as r^m for $m < 0$ are not C^2. Retaining only the solutions that are C^2, we get the general separated-variables solutions

$$u_m(r, \theta) = a_m r^{|m|} e^{im\theta}, m \in \mathbb{Z}; \ \ a_m \in \mathbb{C} \tag{158}$$

Now, (148) is linear, and thus if U and V are solutions, then so is $aU + bV$. The most general solution that we can obtain from (158) is the closure of the span of such solutions,

$$u(r, \theta) = \sum_{m\in\mathbb{Z}} a_m r^{|m|} e^{im\theta} \tag{159}$$

and with (159) we have at $r = 1$ (we'll check that the limit when $r \to 1$ exists),

$$\sum_{m\in\mathbb{Z}} a_m e^{im\theta} = f(\theta) \tag{160}$$

that is, the left side is the Fourier series of f. Since $f \in C^2$, $|a_m| \leqslant const/m^2$ for large m and, by the Weierstrass M test the series in (159) converges absolutely and uniformly for all $r \leqslant 1$, to an analytic function in the open unit disk. We have thus proved:

Theorem 30.0.2. *The heat equation in a disk* (152) *has a unique solution,* (159), (160).

Exercise 78. *Separate variables in the time-dependent heat equation in a disk. The radial ODE has solutions as Bessel functions, $J_m(\lambda r)$; stop here if you are not familiar with them.*

30.1 Examples

(1) Take a disk where the temperature on the boundary is given by $f(\theta) = \sin\theta$. Then, the (unique) solution is simply $r\sin\theta = y$. (2) Similarly, for any trig polynomial, the series represent-

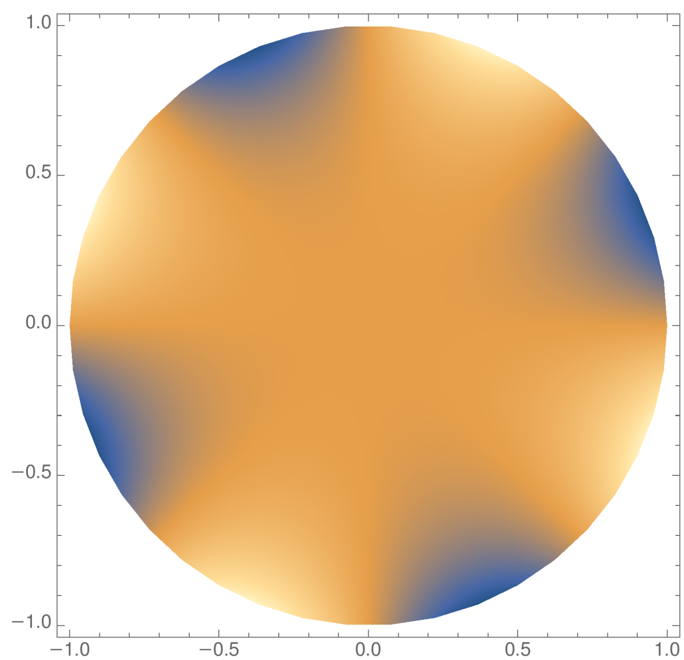

Figure 4: Solution of the heat equation in the disk with condition $\sin(4\theta)$ on $\mathbb{T}$.

ing u is finite. It is interesting to see what happens if the temperature has many changes on the boundary, say $u = \sin(4\theta)$. Write the solution in closed form, as a function of x, y.

Exercise 79. *Show that the heat equation on $\mathbb{T}$,*

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial \theta^2}; \quad u(0,x) = u_0(x) \in C^2(\mathbb{T}) \tag{161}$$

has the unique solution

$$u(t,\theta) = \sum_{m\in\mathbb{Z}} a_m e^{-m^2 t + imx} \text{ where } u_0(x) = \sum_{k\in\mathbb{Z}} a_m e^{imx} \tag{162}$$

In a few steps from here Fourier analysis intersects another major topic in analysis, complex function theory.

Lemma 30.1.1. *The Fourier coefficients of a real-valued function come in complex-conjugate pairs:* $a_{-m} = \overline{a_m}$.

Proof. Check this. □

Thus we can write

$$u(r,\theta) = 2\Re \sum_{m\geqslant 0} a_m r^m e^{im\theta} = 2\Re \sum_{m\geqslant 0} a_m (re^{i\theta})^m = 2\Re \sum_{m\geqslant 0} a_m z^m; \quad z = re^{i\theta} \tag{163}$$

By the Weierstrass M test, the series

$$U(z) = \sum_{m\geqslant 0} a_m z^m \tag{164}$$

converges absolutely and uniformly and absolutely in the open unit disk, and thus U is analytic there.

Let's look again at the definition of the Fourier coefficients:

$$a_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\theta)e^{-im\theta}d\theta = \frac{1}{2\pi}\int_{-\pi}^{\pi}(2\Re U(e^{i\theta}))e^{-im\theta}d\theta =^{(e^{i\theta}=\zeta)} \; = \frac{1}{2\pi i}\int_{\mathbb{T}} 2(\Re U(\zeta))\zeta^{-m-1}d\zeta \tag{165}$$

Substituting in (164) we get, for $|z|<1$,

$$\Re U(z) = \sum_{m\geqslant 0}\frac{1}{2\pi i}\int_{\mathbb{T}}(\Re U(\zeta))\zeta^{-m-1}z^m d\zeta = \frac{1}{2\pi i}\int_{\mathbb{T}}(\Re U(z))\sum_{m\geqslant 0} z^m\zeta^{-m-1}d\zeta = \frac{1}{2\pi i}\int_{\mathbb{T}}\frac{\Re U(\zeta)}{\zeta - z}d\zeta \tag{166}$$

A similar results holds with $\Re$ replaced by $\Im$. Indeed, $\Im U(re^{i\theta})$ satisfies the heat equation with boundary condition $\Im U(e^{i\theta})$. Adding up these two, we obtain the celebrated Cauchy formula

$$U(z) = \frac{1}{2\pi i}\int_{\mathbb{T}}\frac{U(\zeta)}{\zeta - z}d\zeta, \quad z\in\mathbb{D} \tag{167}$$

(for the unit disk, and under C^2 assumptions–a result weaker than the one in complex analysis). This is simply meant to illustrate deeper links between various branches of analysis. It is not necessarily a particularly natural way to build complex analysis, nor is it the path that led Cauchy to it in the early nineteenth century.

Note 30.1.2. *(a) We did not* **not** *prove that the heat equation extended to* $\mathbb{C}$ *with a given complex boundary condition has a solution. It generally doesn't! See what the conditions are needed to have* $\Re\sum_{m\geqslant 0} a_m z^m + i\Im\sum_{m\geqslant 0} b_m z^m = \sum_{m\geqslant 0} c_m z^m$.

(b) Functions u *that satisfy* $\Delta u = 0$ *in a domain in* $\mathbb{R}^n$ *are called harmonic. We see that, in 2d, they are the real or imaginary part of analytic functions.*

30.2 The vibrating string

The equation for a vibrating string is the one-dimensional *wave equation*

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0 \tag{168}$$

We change variables to place the *fixed* endpoints of the string at $-\pi, \pi$. Let the initial shape of the string be given by u_0. The problem becomes

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0; \quad u(0,x) = u_0(x); \quad u(t, \pm\pi) = 0 \tag{169}$$

Exercise 80. *Assume that $u_0(x)$ is C^2. Solve (169) by separation of variables and show that*

$$u(x,t) = \sum_{m=1}^{\infty} (a_m \cos(mt) + b_m \sin(mt)) \sin(mx);$$

$$u(0,x) = \sum_{m=1}^{\infty} a_m \sin(mx), \quad u_t(0,x) = \sum_{m=1}^{\infty} m b_m \sin(mx) \tag{170}$$

Notice that the time dependence is a superposition of cosines of integer multiples of a fundamental frequency, generated by the *fundamental mode* $\sin x$. If we normalize again the units so that the fundamental mode is $440Hz$ (A 440) the next frequency is A 880, one octave up, and the third one is E 1320 *"a perfect fifth"*. The theory of harmony originates in the understanding of string vibrations, which goes back to ancient Greece (harmonikos = "skilled in music"). "Harmonic Analysis" takes its name from this.

30.3 The Poincaré-Wirtinger inequality

We now only prove a special case of the Poincaré-Wirtinger inequality, whose general form is better stated after we introduce Sobolev spaces.

Proposition 30.3.1. If $f \in C^1(\mathbb{T})$ and $\int_{\mathbb{T}} f = 0$, then

$$\|f\|_2 \leqslant \frac{1}{2\pi} \|f'\|_2 \tag{171}$$

The constant $(2\pi)^{-1}$ is optimal, and equality holds iff $f(x) = ae^{2\pi i x} + be^{-2\pi i x}$, $a, b \in \mathbb{C}$.

Proof. Let the Fourier coefficients of f be $\{c_n\}_{n\in\mathbb{Z}}$, and note that under the assumptions in the Proposition, $c_0 = 0$. We have

$$\|f\|_2 = \sum_{n\in\mathbb{Z}} |c_n|^2 \leqslant \sum_{n\in\mathbb{Z}} |nc_n|^2 = \frac{1}{4\pi^2} \sum_{n\in\mathbb{Z}} |2\pi n c_n|^2 = \frac{1}{4\pi^2} \|f'\|_2^2$$

The last statement is an easy exercise. $\square$

Corollary 30.3.2. If $f \in C^1([a,b])$ and $\int_a^b f = 0$, then

$$\|f\|_2 \leqslant \frac{b-a}{\pi} \|f'\|_2$$

(where $\|\cdot\|_2$ denotes the L^2 norm on $[a,b]$). The constant $(b-a)/\pi$ is optimal. Equality occurs iff $f(x) = ae^{\frac{\pi i(x-a)}{b-a}} + be^{-\frac{\pi i(x-a)}{b-a}}; a, b \in \mathbb{C}$.

Exercise 81. *(a) Fourier series can be defined of course for functions of more general period T. If we are interested in functions f periodic on $a, a+T$, then $f(\omega x + \beta)$ is periodic on $[-\pi,\pi]$, if $\omega = 2\pi/T$ and $\beta = -\pi - a\omega$. Carry out the changes of variables and write the Fourier series of f in terms of the exponentials $\{e^{ik\omega x}\}_{k\in\mathbb{Z}}$.*

(b) If f is as in the statement, extend it to a function on $[a-T, a+T]$ which is odd with respect to a, and then apply (a) and the result in the proof above.

30.4 The Riemann-Lebesgue lemma (for $L^1(\mathbb{R})$)

Proposition 30.4.1. If $f \in L^1(\mathbb{R})$, then $\hat{f} = x \mapsto \int_{\mathbb{R}} f(s)e^{ixs}ds \in C_0(\mathbb{R})$

Proof. First, $|e^{i(x+\varepsilon)s} - e^{ixs}||f(s)| \leqslant 2|f(s)|$ and continuity follows by dominated convergence. For the second part note that if $f = \chi_{[a,b]}$ then, for $x \neq 0$, $|\hat{f}| \leqslant 2/|x|$. By Theorem 11.1.2, simple functions of the form $\sum_1^n a_k \chi_{J_k}$ where the J_k are bounded intervals, is dense in L^1. The rest follows from the triangle inequality. □

Exercise 82. Extend this result to $\mathbb{R}^n$: if $f \in L^1(\mathbb{R}^n)$, then $\mathbf{x} \mapsto \int_{\mathbb{R}^n} f(\mathbf{s})e^{i\mathbf{x}\cdot\mathbf{s}}d^n\mathbf{s} \in C_0(\mathbb{R}^n)$.

Exercise 83. 1. Consider the function f given by $f(x) = x^{-a}\chi_{[1,\infty)}(x)$. Show that $F(k) = \int_{\mathbb{R}} e^{ikx}f(x)dx \in C_0(\mathbb{R})$ if $a > 1$, and $F \in C_0(\mathbb{R}\setminus\{0\})$ if $a \in (0,1]$. Show furthermore that for $a \in (0,1)$, $k^{1-a}F(k)$ is bounded for small k, and, when $a = 1$, $F(k) + \ln k$ is bounded near $k = 0$. (Hint: integration by parts is one way; perhaps an even shorter way is to change variable $u = kx$.)

2. If $f : \mathbb{R} \to \mathbb{C}$ is s.t. $f \in L^1(-1,1)$, and for some $a_1, a_2 > 0$ and $C_1, C_2 \in \mathbb{C}$ we have $f - C_1x^{-a_1} \in L^1(1,\infty)$ and $f - C_2x^{-a_2} \in L^1(-\infty,-1)$, then $x \mapsto \int_{\mathbb{R}} f(s)e^{ixs}ds \to 0$ as $x \to \infty$. (What this says is that the L^1 condition can be replaced by L^1 up to explicit additive negative powers of x which themselves may not be in L^1.)

Homework: Problems 25,27,28 in Folland, p. 262 and turn in Ex. 83 above, due Mon April 1.

We have the following extension to Proposition 30.4.1:

30.5 Hurwitz's proof of the isoperimetric inequality

A curve is rectifiable iff the supremum of the perimeters of polygons built by joining finitely many points on the curve is finite. As we discussed, a parametrized curve $\gamma = t \mapsto (x(t), y(t))$ is rectifiable if the function γ is of bounded variation.

Theorem 30.5.1. Assume Γ is a rectifiable simple closed curve in $\mathbb{R}$ of length 2π. Then the area of the interior of the curve is $\leqslant \pi$ and it equals π iff the curve is a circle.

Hurwitz gave the first rigorous proof of this theorem in 1902. He used Fourier series along the lines of the proof below, where, for simplicity, we assume that γ is a smooth curve.

Proof. We can assume without loss of generality that the length of γ is one. If $\mathcal{D}$ =intγ and A is the area of $\mathcal{D}$, then

$$A = \iint_{\mathcal{D}} dxdy = \frac{1}{2}\int_{\Gamma} xdy - ydx = \left|\frac{1}{2}\int_{-\pi}^{\pi}\left[x(s)y'(s) - y(s)x'(s)\right] ds\right| \tag{172}$$

where we used Green's theorem

$$\int_{\Gamma} Ldx + Mdy = \iint_{\mathrm{int}(\Gamma)} \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}\right) dxdy$$

for the vector field $L = -y, M = x$. The arclength measure is given by $d\ell = \sqrt{(x')^2 + (y')^2}ds$. Parameterizing by arclength $\ell := t$ instead of s, $\gamma = \ell \to (x(\ell), y(\ell))$, $\ell \in [0,1]$, we have $(x')^2 + (y')^2 dt = 1$

$$\int_{\Gamma}[(x')^2 + (y')^2]d\ell = 1 = \|x'\|^2 + \|y'\|^2 \tag{173}$$

By changing the origin, we can arrange $\int_{\gamma} xd\ell = \int_{\gamma} yd\ell = 0$. From (195) we have

$$|A| \leqslant \frac{1}{2}|\langle x, y'\rangle| + \frac{1}{2}|\langle x', y\rangle| \leqslant \frac{1}{2\pi}\|x'\|\|y'\| \leqslant \frac{1}{4\pi}\left(\|x'\|^2 + \|y'\|^2\right) = \frac{1}{4\pi}$$

where we used Proposition 30.3.1. Equality is achieved iff the curve is parametrically given by $\gamma(t) = (\sin(2\pi t), \cos(2\pi t))$. $\square$

31 Some conditions for pointwise convergence

$\Lambda^{\alpha}(\mathbb{T})$ is the class of functions on $\mathbb{T}$ which are Hölder continuous of exponent α: $f \in \Lambda^{\alpha}(\mathbb{T})$ if

$$\lambda_{\alpha}(f) = \sup_{x \neq y \in \mathbb{T}} \frac{|f(x) - f(y)|}{|x - y|^{\alpha}} < \infty$$

Theorem 31.0.1. If $f \in \Lambda^\alpha(\mathbb{T})$, $\alpha \in (0,1]$, then there is a constant $C > 0$ such that $\|S_n - f\|_u \leqslant C \ln(n) n^{-\alpha}$, where C depends on α only.

Proof. This proof shows that convergence is linked to the rapid oscillation of D_n, through $\sin(nx + x/2)$, which triggers, in this class of functions, substantial cancellations.

Again we change variable to the interval $[-\pi, \pi]$. We can assume

$$\|f\|_u \leqslant 1$$

Let $m = n + 1/2$ and write $S_n(x) = (2\pi)^{-1} \int_{-\pi}^{\pi} D_n(s) f(x-s) ds$, and thus

$$2\pi(S_n(x) - f(x)) = \int_{-\pi}^{\pi} D_n(s)(f(x-s) - f(x)) ds = \int_{|s| \leqslant \varepsilon} D_n(s)(f(x-s) - f(x)) ds + \int_{|s| > \varepsilon} D_n(s)(f(x-s) - f(x)) ds \quad (174)$$

Note 31.0.2. In this note C will denote some positive constant whose value can depend on $\lambda_\alpha(f)$ but not on ε, n and whose exact value would not alter the conclusion. This is a notational device, to avoid writing $C_1, C_2, ...$ and so on.

where ε will be chosen suitably small. We start with an estimate of the $|s| \leqslant \varepsilon$ integral. For small ε, $\sin(s/2) > Cs$ and

$$\int_{|s| \leqslant \varepsilon} |D_n(s)(f(x-s) - f(x))| ds \leqslant C \int_{|s| \leqslant \varepsilon} \left| \frac{f(x-s) - f(x)}{s} \right| ds \leqslant C \lambda_\alpha(f) \int_{|s| \leqslant \varepsilon} |s|^{\alpha - 1} ds \leqslant \frac{C\varepsilon^\alpha}{\alpha}$$

Cancellations are responsible for decay in the remaining region; we identify the cancellations and rewrite the integral so that these are singled out: we have $\sin(ms) = -\sin(m(s + \frac{\pi}{m}))$. Let $I_k = \{x : |x| \in [\varepsilon + k\frac{\pi}{m}, \varepsilon + (k+1)\frac{\pi}{m}]\}$, $k_1 \in \mathbb{N}$ be the largest j so that $\varepsilon + (2j-1)\frac{\pi}{m} < \pi$ and

$$h(s, x) = \frac{f(x-s) - f(x)}{\sin(s/2)}$$

We get

$$\int_{|s| > \varepsilon} D_n(s)(f(x-s) - f(x)) ds = \int_{|s| > \varepsilon} h(s, x) \sin(ms) ds = \sum_{k=0}^{k_1} \int_{I_k} h(s, x) \sin(ms) ds + \varepsilon_m \quad (175)$$

where ε_m is the contribution of the endpoint intervals:

$$|\varepsilon_m| \leqslant \left| \int_{|s| \in [\varepsilon + (2k_1 - 1)\frac{\pi}{m}, \pi]} h(s, x) \sin(ms) ds \right| \leqslant \frac{C\pi}{m}$$

We combine successive integrals by shifting the variable by $\mp\pi/m$ ($-$ for $s > 0$ and $+$ for $s < 0$)

in all odd-index intervals:

$$\sum_{k=0}^{k_1}\int_{I_k} h(s,x)\sin(ms)ds = \sum_{j=0}^{k_1-1}\int_{I_{2j}\cup I_{2j+1}} h(s,x)\sin(ms)ds = \sum_{j=0}^{k_1-1}\int_{I_{2j}}(h(s,x)-h(s\mp\tfrac{\pi}{m}))\sin(ms)ds$$

Now, in each interval I_{2j}, $\sin ms$ is positive and the oscillations have been removed. At this stage, we can take absolute values without significant loss in the estimates.

$$\left|\sum_{k=0}^{k_1}\int_{I_k} h(s,x)\sin(ms)ds\right| \leqslant \sum_{j=0}^{k_1-1}\int_{I_{2j}}|h(s,x)-h(s\mp\tfrac{\pi}{m})|ds \tag{176}$$

we note that, if $|s|>\varepsilon$, then $|h(s+\delta,x)-h(s)|\leqslant |s|^{-2+\alpha}|\delta|+2\lambda_\alpha(f)|s|^{-1}|\delta|^\alpha$ and the right side of (176) is bounded by

$$C\left(m^{-1}\int_\varepsilon^\pi s^{-2+\alpha}ds+m^{-\alpha}\int_\varepsilon^\pi s^{-1}ds\right)\leqslant C\left(\frac{1}{m\varepsilon^{1-\alpha}}+\frac{1}{m^\alpha}|\log\varepsilon|\right)$$

We now choose ε to obtain a best estimate (up to constants). Choosing $\varepsilon^\alpha=m^{-1}\varepsilon^{-1+\alpha}$ we get

$$\|S_n-f\|_u\leqslant Cm^{-\alpha}\log m \tag{177}$$

□

Exercise 84. Use a similar approach to show that the Fourier coefficients of a function $f\in\Lambda^\alpha(\mathbb{T})$ decay at least as fast as const.$|n|^{-\alpha}$ as $n\to\infty$.

Exercise 85 (Abel means and Abel summability)**.** If $\{a_n\}_{n\in\mathbb{Z}}$ is a sequence, then the *Abel mean* of the sequence is the *function*

$$A(r,\theta)=\sum_{n=-\infty}^{\infty} r^{|n|}a_n e^{in\theta}$$

Note that, if a_n are the Fourier coefficients of a C^2 function f, then the Abel mean is the solution of the heat equation in the disk with f on the boundary! The sequence is *Abel summable* if

$$\lim_{r\to1}A(r,0)=A$$

exists. What is the Abel sum of $1-2+3-4\cdots$?

Show that (convergent to A) implies (Cesàro summable to A) implies (Abel summable to A).

We can think of these summation methods as extensions of convergent summation: extensions of the functional that associates to a convergent sequence its limit. These functionals have a number of expected properties, see. Both fail to commute with multiplication of sequences.

More powerful summation methods exist: Borel summation of series is an important summation method (it relies on a form of Fourier analysis!).

Timestamp: 04/22/2019, 6:10AM

31.1 Approximation to the identity

The convolution of two functions on $\mathbb{T}$ is defined as the *commutative and distributive* product

$$(f * g)(x) = \int_{-\pi}^{\pi} f(s)g(x-s)ds$$

Theorem 31.1.1 (Young's convolution inequality)**.** If $f \in L^1$, $g \in L^p$, and $1 \leqslant p \leqslant \infty$, then $\|f * g\|_p \leq \|f\|_1 \|g\|_p$.

Proof. Use Minkowski's inequality for integrals. □

The difficulties in establishing pointwise convergence of Fourier series ultimately boils down to the divergence of the L^1 norm of the Dirichlet kernel. A *good kernel, or approximation to the identity, or approximate identity* is one which has most of the features of the Dirichlet kernel, but with finite L^1 norm.

Definition 31.1.2. A family $\{K_n\}_{n \in \mathbb{N}} \subset L^1(\mathbb{T})$ is said to be an approximation to the identity (approximate identity) if

(a) For all $n \geqslant 1$, with $\hat{K}_n := f \mapsto K_n * f$, we have

$$\int_{-\pi}^{\pi} K_n(s)ds = 1 \quad (\text{i.e. } \hat{K}_n 1 = 1) \tag{178}$$

(b)

$$\sup_{n \geqslant 1} \int_{-\pi}^{\pi} |K_n(s)|ds = M < \infty \quad ((\text{i.e. } \forall n,\ \|\hat{K}_n\|_{L^\infty \to L^\infty} = \|K_n\|_1 \leqslant M) \tag{179}$$

(c) For any $\varepsilon > 0$ we have

$$\lim_{n \to \infty} \int_{|x| \in [\varepsilon, \pi]} |K_n(s)|ds = 0 \quad (\text{Approximate identity}) \tag{180}$$

Note 31.1.3. For positive kernels (179) follows from (178).

Theorem 31.1.4. (a) Let $\{K_n\}_{n\in\mathbb{N}}$ be an approximation to the identity family. Then, for any $f \in L^\infty(\mathbb{T})$ and any point x of continuity of f we have

$$\lim_{n\to\infty}(K_n * f)(x) = f(x) \tag{181}$$

If $f \in C(\mathbb{T})$ then $\lim_{n\to\infty} \|K_n * f - f\|_u = 0$.

(b) If $f \in L^p(\mathbb{T}), 1 \leqslant p < \infty$, then $\lim_{n\to\infty} \|\hat{K}_n f - f\|_p = 0$,

(c) For $1 \leqslant p < \infty$ we have $\sup_n \|\hat{K}_n\|_{p\to p} \leqslant M$. If $1 \leqslant p < \infty$, the sequence of operators $\{\hat{K}_n\}_{n\in\mathbb{N}}$ converges weakly to the identity.

Proof. The proof is similar –but simpler– to that of Theorem 31.0.1. Let x be a point of continuity of f. Given ε, let δ be s.t. $|f(x-s) - f(x)| \leqslant \varepsilon$ if $|s| \leqslant \delta$. We decompose the integral $(K_n * f)(x) - f(x)$ as in (174),

$$\int_{-\pi}^{\pi} K_n(s)(f(x-s) - f(x))ds = \int_{|s|\leqslant\delta} K_n(s)(f(x-s) - f(x))ds + \int_{|s|>\delta} K_n(s)(f(x-s) - f(x))ds \tag{182}$$

We bound the first integral by using the sup norm for $f(x-s) - f(x)$ and the L^1 norm for K_n:

$$\left|\int_{|s|\leqslant\delta} K_n(s)(f(x-s) - f(x))ds\right| \leqslant \varepsilon \int_{|s|\leqslant\delta} |K_n(s)|\, ds \leqslant \varepsilon M \tag{183}$$

and we use the assumptions on K_n in the second one

$$\left|\int_{|s|>\delta} K_n(s)(f(x-s) - f(x))ds\right| \leqslant 2\|f\|_u \int_{|s|>\delta} |K_n(s)|ds \to 0 \text{ as } n \to \infty \tag{184}$$

(b) Let $f \in L^p$ and let $g \in C(\mathbb{T})$ be s.t. $\|f - g\|_p < \varepsilon$. Using Young's inequality for convolution, we see that $\sup_n \|\hat{K}_n\|_{p\to p} = \sup\{\|\hat{K}_n u\|_p : n \in \mathbb{N} \text{ and } u \in L^p \text{ with } \|u\|_p = 1\} \leqslant M$ and, for large enough n and some constant $C > 0$,

$$\|\hat{K}_n f - f\|_p \leqslant \|\hat{K}_n g - g\|_p + \|f - g\|_p + \|\hat{K}_n(g - f)\|_p \leqslant 2\pi\|\hat{K}_n g - g\|_u + 2\|f - g\|_p \leqslant C\varepsilon$$

(c) follows immediately from (b). □

Note 31.1.5. The "dictionary" between 1-periodic and 2π-periodic functions is as follows. If f is one-periodic, then $g = x \mapsto f((2\pi)^{-1}x)$ is 2π periodic, and we have

$$f((2\pi)^{-1}x) = \sum_{k\in\mathbb{Z}} e^{ikx} \frac{1}{2\pi} \int_0^{2\pi} f((2\pi)^{-1}s)e^{-iks} ds$$

which implies

$$f(x) = \sum_{k \in \mathbb{Z}} e^{2\pi i k x} \int_0^1 f(s) e^{-2\pi i k s} ds \tag{185}$$

Exercise 86. Let $f \in L^1(\mathbb{T})$. Show $f = 0$ a.e. iff for all the Fourier coefficients

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) e^{-ins} ds = 0 \tag{186}$$

$n \in \mathbb{Z}$, vanish.

32 The Poisson kernel

Let $f \in L^1(\mathbb{T})$. Then, its Fourier coefficients $\{a_n\}_{n \in \mathbb{N}}$ are bounded, and thus the Abel means

$$A_r(f)(t) = \sum_{n \in \mathbb{Z}} a_n r^{|n|} e^{int} \tag{187}$$

converge absolutely and uniformly for $r < 1$, and we can interchange summation and integration in (186) to write

$$A_r(f)(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sum_{n \in \mathbb{Z}} r^{|n|} e^{in(t-s)} ds = (P_r * f)(t) \tag{188}$$

where $P_r(t)$ is the Poisson kernel,

$$P_r(t) = \sum_{n \in \mathbb{Z}} r^{|n|} e^{int} = \frac{1 - r^2}{1 - 2r \cos t + r^2} \tag{189}$$

as you can easily check.

Proposition 32.0.1. P_r are an approximation to the identity. (Here the family is indexed by the continuous variable $r \in [0, 1)$, with definitions similar to those in the discrete case.)

Proof. The fact that $\frac{1}{2\pi} \int_{-\pi}^{\pi} P_r = 1$ (property (a)) follows from integrating the series in (187) term by term and noting that all contributions for $n \neq 0$ vanish. For (b) we note that the kernels are positive. Property (c) follows from the fact that P_r are bounded and go to zero uniformly in any interval of the form in $(\varepsilon, \pi], \varepsilon > 0$. $\square$

As a consequence, we have

Theorem 32.0.2. The Fourier series of an $L^\infty(\mathbb{T})$ function is Abel summable to f at any point of continuity of f. If $f \in C(\mathbb{T})$, then the series is uniformly Abel summable to f.

Returning to the heat equation, we find that

Theorem 32.0.3. The heat eq. (153) with f **continuous**, the uniform limit of $u(r,\theta)$ as $r \to 1$, has a unique solution (159) and (160).

Exercise 87. Check that the map U, $U(f) = \{a_n\}_{n\in\mathbb{Z}}$, where $\{a_n\}_{n\in\mathbb{Z}}$ are the Fourier coefficients of f, is a unitary operator between $L^2([-\pi,\pi])$ and $\ell^2(\mathbb{Z})$. What is the image under U of the functions in $AC(\mathbb{T})$ with derivative in $L^2(\mathbb{T})$? (This is the domain of definition of the self-adjoint operator $i\frac{d}{dx}$ on $\mathbb{T}$.)

32.1 Several variables

Assume $f \in C^1((\mathbb{T})^2)$. Then,

$$f(x,y) = \sum_{k\in\mathbb{Z}} c_k(x)e^{iky} \text{ where } c_k(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x,t)e^{-ikt}dt \tag{190}$$

Now, $c_k \in C^1(\mathbb{T})$ (why?), and hence

$$f(x,y) = \sum_{k\in\mathbb{Z}}\left(\sum_{m\in\mathbb{Z}} c_{k,m}e^{imx}\right)e^{iky} \text{ where } c_{km} = \frac{1}{2\pi}\int_{-\pi}^{\pi} c_k(s)e^{-ims}ds = \iint_{[-\pi,\pi]^2} f(s,t)e^{-i(ms+kt)}dsdt \tag{191}$$

Uniform and absolute convergence (justify!) means that, we can write

$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{Z}^2} c_{\mathbf{n}}e^{i\mathbf{n}\cdot\mathbf{x}} \tag{192}$$

Exercise 88. (a) Under smoothness conditions as above, formulate and prove a theorem about Fourier series in n dimensions.

(b) Write down a formula for the Fourier series of functions which are periodic, but have different periods in the different directions in $\mathbb{R}^n$.

The following exercise illustrated the duality between regularity (smoothness) and decay of the Fourier coefficients for functions that have point singularities. By the latter we mean that for each point at which the function is not smooth, there is an interval centered at that point in which there is no other point of non-smoothness.

Exercise 89. Let

$$f(x) = \sum_{n\in\mathbb{N}} \frac{\sin nx}{n^\alpha} =: \frac{i}{2}\left(\mathrm{Li}_\alpha(e^{-ix}) - \mathrm{Li}_\alpha(e^{ix})\right); \quad (\alpha > 0, x \in [-\pi, \pi]) \tag{193}$$

(1) Show that (193) converges pointwise for all x.

(2) We now use a rudimentary form of Borel summation (see also more about this form of Borel summation) to determine the regularity of f. Using the definition of the Gamma function, show that

$$\frac{1}{n^\alpha} = \frac{1}{\Gamma(\alpha)}\int_0^\infty p^{\alpha-1}e^{-np}dp \tag{194}$$

Show that this implies that for $x \neq 0$ we have

$$f(x) = \frac{1}{\Gamma(\alpha)}\int_0^\infty p^{\alpha-1}\sum_{n\in\mathbb{N}}\sin(nx)e^{-np}dp = \frac{1}{\Gamma(\alpha)}\int_0^\infty p^{\alpha-1}\frac{\sin x}{2(\cosh p - \cos x)}dp \tag{195}$$

For $\alpha = 1$ the last integral is elementary,

$$2f(x) = \begin{cases} -x-\pi, \ x < 0 \\ 0, \ x = 0 \\ -x+\pi, \ x > 0 \end{cases}$$

Prove that $f(x)$ is C^∞ away from zero (actually, it is analytic).

(3) Take now $x > 0$ and small. Write (195) as

$$\frac{1}{x\Gamma(\alpha)}\int_0^\infty p^{\alpha-1}\frac{x^2a(x)}{p^2b(p)+x^2c(x)}dp \tag{196}$$

and show that a, b and c are smooth in a neighborhood of zero, that $a(0) = b(0) = c(0) = 1$ and that $b(p) \geqslant 1$ for $p \geqslant 0$. With the change of variable $p = xq$ we get, for $x > 0$ small,

$$f(x) = \frac{1}{x^{\alpha-1}\Gamma(\alpha)}\int_0^\infty q^{\alpha-1}\frac{a(x)}{q^2b(qx)+c(x)}dq \tag{197}$$

and that, as $x \to 0^+$ we have

$$\lim_{x\to 0}\int_0^\infty q^{\alpha-1}\frac{a(x)}{q^2b(qx)+c(x)}dq = \int_0^\infty \frac{q^{\alpha-1}}{q^2+1}dq = \frac{\pi}{2\sin(\alpha\pi/2)} \tag{198}$$

(The last expression is most easily proved by the residue theorem, but you **don't need to justify it**; this explicit value is not terribly important here.) Use (198) to conclude that

$$\lim_{x\to 0}\frac{f(x)}{\left(\frac{|x|^\alpha}{x}\right)} = \frac{\pi}{2\Gamma(\alpha)\sin(\alpha\pi/2)}; \quad \alpha \in (0,1) \tag{199}$$

Thus f has precisely one point singularity, $x = 0$. Show that, for $\alpha \in (1,2)$, $f \in C^{\alpha-1}(\mathbb{T})$.

(d) Show that for $\alpha \in (0,1)$, $f \in L^p$ for any $p \in [1, 1/(1-\alpha))$. Are the Fourier coefficients of f those implied by the series?

Note 32.1.1. It is useful to sketch this function for some $\alpha \in (0,1)$.

This particular relation, $1/n^\alpha \mapsto (x - x_0)^{\alpha-1}$ between decay and regularity is generally true for point singularities. In the general class Λ^α, the (sharp) correspondence is $1/n^\alpha \leftrightarrow f \in \Lambda^\alpha$ with a proof similar to that of Theorem 31.0.1.

33 The Fourier transform

If f is not periodic, but compactly supported, we can extend it to a periodic function with period, say, the size of its support, and then we can analyze it using Fourier series.

Now if $f \in L^1(\mathbb{R}^n)$ is not periodic, we can still define, for any $k \in \mathbb{R}$,

$$(\mathcal{F}f)(k) = \hat{f}(k) = \int_{\mathbb{R}^n} e^{-2\pi i\langle x,k\rangle} dx, \quad k \in \mathbb{R}^n \tag{200}$$

The function $\hat{f}$ is called the **Fourier transform** of f. The **inverse Fourier transform** (we'll shortly why "inverse") is

$$(\mathcal{F}^{-1}f)(k) = \check{f}(k) = \int_{\mathbb{R}^n} e^{2\pi i\langle x,k\rangle} dx, \quad k \in \mathbb{R}^n \tag{201}$$

Lemma 33.0.1. The translation $\tau_a := f \mapsto f(x+a)$ is continuous in L^p, $1 \leqslant p < \infty$.

Proof. Since $C_c(\mathbb{R}^n)$ is dense in $L^p, 1 \leqslant p < \infty$ it suffices to prove this for $C_c(\mathbb{R}^n)$. Let f be continuous and compactly supported in K. Translation is evidently linear, and thus it suffices to prove continuity at zero. We have

$$\lim_{a\to 0} \|\tau_a f - f\|_p \leqslant m(K)^{1/p} \lim_{a\to 0} \|\tau_a f - f\|_\infty = 0 \tag{202}$$

by uniform continuity. $\square$

It is convenient to first analyze these transforms in a space of smooth, rapidly decreasing functions.

33.1 The Schwartz space $\mathcal{S}$

Definition 33.1.1. A topological vector space X is called a Fréchet space if it satisfies the following three properties:

1. X is a Hausdorff space,
2. Its topology may be induced by a countable family of seminorms $(\|\cdot\|_k), k \in \mathbb{N}$. That is, the sets
$$\{y : \|y - x\|_k < \varepsilon, \forall k \leqslant K\}$$
let where $\varepsilon \in \mathbb{R}^+$ and $K \in \mathbb{N}$, form a base of neighborhoods.
3. X is complete with respect to the family of semi-norms.

Note 33.1.2. If X is a Fréchet space, then a sequence converges in X iff it converges in each seminorm.

The topology induced by a family of seminorms is Hausdorff iff

$$\bigcap_{k \in \mathbb{N}} \{x \in X : \|x\|_k = 0\} = \{0\}$$

It is easy to see that A Fréchet space is a special case of a metrizable space, one in which the metric is translation invariant, $\rho(f, g) = \rho(f - g, 0)$. If the family of semi-norms is $\|\cdot\|_n$, then a metric which induces the same topology is

$$\rho(f, 0) = \sum_{n \in \mathbb{N}} 2^{-n} \frac{\|f\|_n}{1 + \|f\|_n}$$

Conversely, a metric space is Fréchet if it is complete, *locally convex, see below* and the metric is translation-invariant.

Let α, β be multiindices, that is tuples $(m_1, m_2, ..., m_n) \in \mathbb{N}_0^n$, where $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. We use the multidimensional conventions

$$|\alpha| = \sum_{i=1}^n \alpha_i, \quad x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}, \quad \partial^\alpha = \frac{\partial^{|\alpha|}}{(\partial x)^\alpha} \quad |x| = \sqrt{\sum_{i=1}^n x_i^2}; \text{ and } \binom{n}{\alpha} = \frac{n!}{\prod_{i=1}^n \alpha_i!}$$

The Schwartz space $\mathcal{S}$ of rapidly decreasing functions on $\mathbb{R}^n$ is defined as

$$S(\mathbb{R}^n) = \left\{ f \in C^\infty(\mathbb{R}^n) : \|f\|_{N,\beta} < \infty \quad \forall N \in \mathbb{N}_0, \beta \in \mathbb{N}_0^n \right\} \tag{203}$$

where

$$\|f\|_{N,\beta} = \sup_{x \in \mathbb{R}^n} \left| (1 + |x|)^N \partial^\beta f(x) \right|. \tag{204}$$

These are smooth functions that decrease, for large $|x|$, faster than any inverse power of $|x|$.

Note 33.1.3. Recall that , that if $(f_n)_{n\in\mathbb{N}}$ is a sequence of differentiable functions on $\mathbb{R}$ s.t. $\{f_n'\}_{n\in\mathbb{N}}$ converge uniformly to some function h and $\{f_n(x_0)\}_{n\in\mathbb{N}}$ converges for some x_0, then $\{f_n\}_{n\in\mathbb{N}}$ converge uniformly on compact sets to some f and $h = f'$.

Proposition 33.1.4. $\mathcal{S}$ is a Fréchet space.

Proof. Only completeness needs to be checked. Since $C(\mathbb{R}^n)$ is complete, a Cauchy sequence $\{f_k\}_{k\in\mathbb{N}}$ in all $\|\cdot\|_{N,\beta}$ implies that $\{f_k\}_{k\in\mathbb{N}}$ convergences in all $\|\cdot\|_{N,\beta}$ to some functions $g_{N,\beta}$. To identify this limit we can use the property in Note 33.1.3. $\square$

Lemma 33.1.5. The families of seminorms

$$\left\{\||x|^\alpha \partial^\beta f\|_\infty\right\}_{N,\beta\in\mathbb{N}_0^n} \text{ and } \left\{\|(1+|x|)^N \partial^\beta f\|_\infty\right\}_{\alpha\in\mathbb{N}_0^n, N\in\mathbb{N}_0}; \tag{205}$$

induce the same topology on $\mathcal{S}$.

Proof. Indeed,

$$|x|^\alpha < (1+|x|)^{|\alpha|};\quad (1+|x|)^N = \sum_{k=0}^N \binom{N}{k}|x|^k \leqslant \sum_{k=0}^N \binom{N}{k}\left(\sum_{i=1}^n |x_i|\right)^k = \sum_{\beta,|\beta|\leqslant N} a_\beta |x|^\beta$$

for some nonnegative coefficients a_β and thus the distance induced by the first family of seminorms goes to zero iff the distance induced by the second one does. $\square$

Compactly supported smooth functions, $C_c^\infty(\mathbb{R}^n)$ are an important subset of $\mathcal{S}$. A prototypical such function is the function η below, compactly supported in the unit ball and smooth.

Proposition 33.1.6. The function

$$\eta(1-|x|^2) := \begin{cases} e^{-\frac{1}{1-|x|^2}};\ |x|<1 \\ 0;\ |x|\geqslant 1 \end{cases} \tag{206}$$

is in $C_c^\infty(\mathbb{R}^n)$.

Proof. This follows from the chain rule and the fact that the function $t \mapsto e^{-1/t}\chi_{\mathbb{R}^+}(t)$ is in $C^\infty(\mathbb{R})$, see Exercise 3/p.239 in Folland. □

This function can be used as a building block to define other interesting compactly supported functions. For instance, the function

$$\varphi(x) = \begin{cases} 1; |x| \leqslant 1 \\ \frac{\exp\left(\frac{1}{|x|^2-1}+\frac{1}{|x|^2-4}\right)}{1+\exp\left(\frac{1}{|x|^2-1}+\frac{1}{|x|^2-4}\right)};\ |x| \in (1,2) \\ 0; |x| \geqslant 2 \end{cases} \tag{207}$$

is a smooth function, compactly supported in the ball of radius 2 and equals 1 in the closed ball of radius 1: $B_1(0) \prec \varphi \prec B_2(0)^c$.

Proposition 33.1.7. $C_c^\infty(\mathbb{R}^n)$ is dense in $\mathcal{S}(\mathbb{R}^n)$

Homework: Problems 8,9,13,15 from Folland, Chapter 8 and turn in Ex. 85 and 86 from the notes. Due Mon. April 8.

Proof. Let $\varphi_n = x \mapsto \varphi(x/n)$ with φ as in (207). If $f \in \mathcal{S}(\mathbb{R}^n)$,then $\{f\varphi_n\}_{n\in\mathbb{N}}$ is a sequence of compactly supported functions which, we claim, converges to f in the topology of $\mathcal{S}$. Indeed, we have

$$\begin{aligned} |x|^\gamma \partial^\alpha(f(x)\varphi_n(x)) &= |x|^\gamma \sum_{\beta\leqslant\alpha} n^{\beta-\alpha}\binom{\alpha}{\beta}\partial^\beta f(x)(\partial^{\alpha-\beta}\varphi)(x/n) \\ &= |x|^\gamma\varphi(x/n)f(x) + |x|^\gamma \sum_{\beta\leqslant\alpha,\beta\neq\alpha} n^{\beta-\alpha}\binom{\alpha}{\beta}\partial^\beta f(x)(\partial^{\alpha-\beta}\varphi)(x/n) \\ &= |x|^\gamma\partial^\alpha f(x) + |x|^\gamma \sum_{\beta\leqslant\alpha,\beta\neq\alpha} n^{\beta-\alpha}\binom{\alpha}{\beta}\partial^\beta f(x)(\partial^{\alpha-\beta}\varphi)(x/n) - (1-\varphi(x/n))\partial^\alpha f(x) \end{aligned} \tag{208}$$

and note that, for any h, since $\sup_{x\in\mathbb{R}^n}|h(x)| = \sup_{x\in\mathbb{R}^n}|h(x/n)|$, we have

$$x^\gamma \sum_{\beta\leqslant\alpha,\beta\neq\alpha} n^{\beta-\alpha}\binom{\alpha}{\beta}\partial^\beta f(x)(\partial^{\alpha-\beta}\varphi)(x/n) \to 0 \text{ as } n\to\infty$$

Finally, $1-\varphi(x/n) = 0$ if $|x| \leqslant n$. Since $|\partial^\alpha f(x)| \leqslant \|f\|_{\alpha,|\gamma|+1}(1+|x|)^{-|\gamma|-1}$, we have $x^\gamma(1-\varphi(x/n))\partial^\alpha f(x) \to 0$ as well. □

Other important examples of functions in $\mathcal{S}$ are the Gaussians, or polynomials multiplying Gaussians,

$$x_i e^{-ax^2}, \quad (a>0)$$

Lemma 33.1.8. The maps $\mathcal{F}$ and $\mathcal{F}^{-1}$ are continuous linear transformations from $\mathcal{S}$ into itself. Furthermore, $\mathcal{F}$ interchanges multiplication by the variable with differentiation, as follows:

$$\mathcal{F}(\partial^\alpha x^\beta f) = (-1)^\beta (2\pi i)^{\alpha-\beta} \xi^\alpha \partial^\beta \mathcal{F}(f) \tag{209}$$

Proof. We have, by integration by parts,

$$\mathcal{F}(\partial^\alpha x^\beta f) = (-2\pi i \xi)^\alpha (-1)^\alpha \mathcal{F}(x^\beta f) = \frac{(-2\pi i \xi)^\alpha (-1)^\alpha}{(-2\pi i)^\beta} \partial^\beta \mathcal{F}(f) \tag{210}$$

Linearity is clear. Expanding out $\partial^\alpha x^\beta f$, we see that, up to constants independent of f,

$$\|\hat{f}\|_{\alpha,\beta} = C\|\mathcal{F}(\partial^\alpha x^\beta f)\|_\infty \leqslant C' \sum_{\alpha' \leqslant \alpha, \beta' \leqslant |\beta|+n} \|f\|_{\alpha',\beta'} \tag{211}$$

□

Lemma 33.1.9 (Improper Riemann integrals and sums)**.** Assume $f \in C(\mathbb{R}^n)$ and $|x|^{n+3} f$ is bounded. Then,

$$\lim_{\varepsilon \to 0} \varepsilon^n \sum_{k \in \mathbb{Z}^n} f(\varepsilon k) = \int_{\mathbb{R}^n} f(x) dx \tag{212}$$

We note that $n+3$ is suboptimal, but that's all we need, for now.

Proof. We denote by $C_a(x_0)$ the cube of side a centered at x_0 and parallel to the axes. Note first that $|x|^{n+2} f$ is uniformly continuous on $\mathbb{R}^n$. For a $\delta > 0$, let ε' be s.t. $|s| < \varepsilon'$ implies $\sup_{x \in K} |f(x+s) - f(x)| \leqslant \delta(|x|+1)^{-n-2}$. For any $\varepsilon \leqslant \varepsilon'$ we have

$$\int_{\mathbb{R}^n} f dm = \sum_{k \in \mathbb{Z}^n} \int_{C_\varepsilon(k\varepsilon)} f(k\varepsilon + s) ds = \sum_{k \in \mathbb{Z}^n} \int_{C_\varepsilon(k\varepsilon)} f(k\varepsilon) dm + \sum_{k \in \mathbb{Z}^n} \int_{C_\varepsilon(k\varepsilon)} [f(k\varepsilon + s) - f(k\varepsilon)] dm$$
$$= \sum_{k \in \mathbb{Z}^n} f(\varepsilon k) + O(\delta) \tag{213}$$

since, for some $C > 0$ independent of f and δ, we have

$$\int_{C_\varepsilon(k\varepsilon)} |f(k\varepsilon + s) - f(k\varepsilon)| dm \leqslant \frac{\delta \varepsilon^n}{(|k\varepsilon|+1)^{n+2}} \leqslant \delta \sum_{k \in \mathbb{Z}} \frac{1}{(|k|+1)^{n+2}} \leqslant C\delta$$

□

Theorem 33.1.10 (Fourier inversion theorem in $\mathcal{S}$)**.** (i) The Fourier transform is one to one from $\mathcal{S}(\mathbb{R}^n)$ onto itself and $\mathcal{S}(\mathbb{R}^n)$ $\mathcal{F}^{-1}\mathcal{F} = \mathcal{F}\mathcal{F}^{-1} = I$, the identity operator.

(ii) (Plancherel) If $f \in \mathcal{S}(\mathbb{R}^n)$, then $\|f\|_2 = \|\mathcal{F}f\|_2$.

Proof. The fact that $\mathcal{F}$ is one-to-one onto will follow from the inversion formula. Since $\mathcal{F}^{-1}\mathcal{F}$ is continuous, it suffices to show that $\mathcal{F}^{-1}\mathcal{F} = I$ on the dense set $C_c^\infty(\mathbb{R}^n)$. Take $f \in C_c^\infty(\mathbb{R}^n)$ and ε small enough so that f is supported in $K = [-\varepsilon^{-1}/2, \varepsilon^{-1}/2]^n$. Expanding f in Fourier series we get

$$f(x) = \varepsilon^n \sum_{k\in\mathbb{Z}^n} e^{2\pi ikx\varepsilon} \int_K f(s)e^{-2\pi iks\varepsilon}ds = \varepsilon^n \sum_{k\in\mathbb{Z}^n} e^{2\pi ikx\varepsilon} \int_{\mathbb{R}^n} f(s)e^{-2\pi iks\varepsilon}ds$$
$$= \varepsilon^n \sum_{k\in\mathbb{Z}^n} e^{2\pi ikx\varepsilon}(\mathcal{F}f)(k\varepsilon) \quad (214)$$

which, by Lemma 33.1.9, converges to $\mathcal{F}^{-1}\mathcal{F}f$ as $\varepsilon \to 0$.

(ii) Similarly, it is enough to prove this in $C_c^\infty(\mathbb{R}^n)$. Let $f \in C_c^\infty(\mathbb{R}^n)$ and K, ε be as above. By Note 29.0.1 we have

$$\int_{\mathbb{R}^n} |f(s)|^2 ds = \int_K |f(s)|^2 ds = \varepsilon^n \sum_{k\in\mathbb{Z}^n} |(\mathcal{F}f)(k\varepsilon)|^2 \underset{\varepsilon\to 0}{\to} \int_{\mathbb{R}^n} |(\mathcal{F}f)(k)|^2 dk \quad (215)$$

□

Corollary 33.1.11. $\mathcal{F}$ extends to an isomorphism on $L^2(\mathbb{R}^n)$ with $\mathcal{F}^{-1}$ as its inverse.

Lemma 33.1.12. $L^1(\mathbb{R}^n) \cap C_0(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$.

Proof. Let $f \in L^1(\mathbb{R}^n) \cap C_0(\mathbb{R}^n)$ and K be the compact set outside which $|f| \leqslant 1$. Then,

$$\|f\|_2 = \int_K |f|^2 dm + \int_{K^c} |f|^2 dm \leqslant \int_K |f|^2 dm + \int_{K^c} |f| dm \leqslant \int_K |f|^2 dm + \|f\|_1 < \infty$$

□

Lemma 33.1.13 (A formula for the extension of $\mathcal{F}$ to L^2). *If $f \in L^2$ and $\hat{f} = \mathcal{F}f$, then*

$$\lim_{n\to\infty} \left\| \hat{f} - \int_{|x|\leqslant n} e^{-2\pi i\xi s} f(s)dm \right\|_2 = 0$$

If $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$, then the extension of $\mathcal{F}$ to L^2 is the same as $\mathcal{F}$.

Any sequence of compact sets whose union is $\mathbb{R}^n$ would yield the same result.

Proof. This follows from the fact that, by dominated convergence, $\|f - \chi_{|x|\leqslant n} f\|_2 \to 0$ as $R \to \infty$ and the continuity of the Fourier transform in $L^2(\mathbb{R}^n)$. The second part is an easy corollary. □

Note 33.1.14. The result above is sometimes written

$$\hat{f} = \operatorname*{l.i.m}_{R\to\infty} \int_{|x|\leqslant R} e^{-2\pi i\xi s} f(s)dm$$

Theorem 33.1.15. If $f \in L^1(\mathbb{R}^n)$ and $\mathcal{F}f \in L^1(\mathbb{R}^n)$, then $\mathcal{F}^{-1}\mathcal{F}f = f$.

Proof. This follows immediately from Lemmas 33.1.12 and 33.1.13. □

Theorem 33.1.16 (Hausdorff-Young inequality)**.** Assume $1 \leqslant p \leqslant 2$ and $p^{-1} + q^{-1} = 1$. Then, the Fourier transform is a bounded map from L^p to L^q with norm at most one.

Proof. We use interpolation. Note that the Fourier transform is continuous from $L^2(\mathbb{R}^n)$ to $L^2(\mathbb{R}^n)$, and from $L^1(\mathbb{R}^n)$ into $L^\infty(\mathbb{R}^n)$. The result now follows from the Riesz-Thorin interpolation theorem with $p_0 = q_0 = 2, p_1 = 1, q_1 = \infty$. □

33.2 The Fourier inversion theorem, a direct approach

We show the inversion formula in $\mathbb{R}$. Let $f \in \mathcal{S}(\mathbb{R})$. Then, $\mathcal{F}^{-1}\mathcal{F}f$ equals

$$\begin{aligned}
&\lim_{R\to\infty}\int_{-R}^{R} e^{i\xi x}\int_{-\infty}^{\infty} e^{-i\xi y} f(y)dyd\xi = \lim_{R\to\infty}\int_{-\infty}^{\infty} f(y)\int_{-R}^{R} e^{i\xi(x-y)}d\xi\, dy = \lim_{R\to\infty}\int_{-\infty}^{\infty} f(x-u)\int_{-R}^{R} e^{i\xi u}d\xi\, du \\
&= 2\lim_{R\to\infty}\left(\int_{-\infty}^{0} + \int_{0}^{\infty}\right) f(x-u)\frac{\sin Ru}{u}du = 2\lim_{R\to\infty}\left(\int_{-\infty}^{0} + \int_{0}^{\infty}\right)[f(x-u)+f(x+u)]\frac{\sin Ru}{u}du \\
&\qquad = 2\lim_{R\to\infty}\int_0^\infty \frac{f(x+s)+f(x-s)-2f(x)}{s}\sin(Rs)ds + 4f(x)\int_0^\infty \frac{\sin s}{s}ds = 2\pi f(x)
\end{aligned}$$

In the last integral above we changed Ru to s, and the integral before it goes to zero by the Riemann-Lebesgue result in Exercise 83, 2. and the fact that the expression multiplying $\sin \xi u$ is smooth.

Note also the appearance in the process of the kernel $u^{-1}\sin(Ru)$, a continuous analog of the Dirichlet kernel, in concentrating the main contribution of the integral to a vanishing neighborhood of zero.

Proposition 33.2.1. If $f(x) = e^{-\pi\alpha|x|^2}$ with $\Re(\alpha) > 0$, then $\hat{f}(\xi) = \alpha^{-n/2}e^{-\pi|\xi|^2/\alpha}$.

Proof. In one dimension this follows from the fact that

$$\frac{d\hat{f}}{d\xi} = -\frac{2\pi}{\alpha}\xi\hat{f}$$

as it can be checked by integration by parts and that $\hat{f}(0) = \alpha^{-1/2}$. The extension to $\mathbb{R}^d$ is immediate, since the multiple integral is a product of one-dimensional integrals of the type above. □

34 Supplementary material: Some applications of the Fourier transform

34.1 The Schrödinger equation for a free particle in $\mathbb{R}^d$

The wave function $\psi(x,t)$ of a particle has the following interpretation:$|\psi(x,t)|^2 dm$ is the probability density that, as a result of a measurement at time t, the particle will be found at position x. Then clearly we must have $\int_{\mathbb{R}^d} |\psi(x,t)|^2 dm(x) = 1$ for any t, in particular $\psi \in L^2(\mathbb{R}^d)$.

In the case of a single particle of mass m in an external potential $V(x,t)$, ψ satisfies the PDE

$$i\hbar \frac{\partial \psi}{\partial t} = -\tfrac{1}{2m}\Delta\psi + V(x,t)\psi$$

This is an evolution equation which requires an initial condition $\psi(x,t_0) = \psi_0(x)$. Here $E = -\frac{1}{2m}\Delta$ is the *kinetic energy operator* $E = \frac{p^2}{2m} =: \frac{1}{2m}\nabla^2$. In atomic units, $\hbar = 2m = 1$. A particle is free if the external potential is zero,

$$i\frac{\partial \psi}{\partial t} = -\Delta\psi$$

The Laplacian is a symmetric operator,

$$\operatorname*{l.i.m.}_{R\to\infty} \int_{|x|\leqslant R} (\psi\,\Delta\varphi - \varphi\,\Delta\psi)\; dV = \operatorname*{l.i.m.}_{R\to\infty} \oint_{|x|=R} \left(\psi\frac{\partial\varphi}{\partial \mathbf{n}} - \varphi\frac{\partial\psi}{\partial \mathbf{n}}\right) dS = 0$$

Lemma 34.1.1. *If $\|\psi(x,0)\|_2 = 1$, then $\|\psi(x,t)\|_2 = 1$ for all t.*

Such an evolution is called unitary, for obvious reasons.

Proof. By taking the complex conjugate of the Schrödinger equation,

$$-i\hbar\frac{\partial\overline{\psi}}{\partial t} = -\tfrac{1}{2m}\Delta\overline{\psi} + V(x,t)\overline{\psi}$$

Multiplying the first equation by $\overline{\psi}$, the second by ψ and subtracting, we get and subtracting the two equations, and integrating over $\mathbb{R}^d$ we get

$$i\frac{d}{dt}\int_{\mathbb{R}^d} |\psi|^2 dm = \operatorname*{l.i.m.}_{R\to\infty} \int_{|x|\leqslant R} \psi\,\Delta\overline{\psi} - \overline{\psi}\,\Delta\psi = 0$$

□

We now take the Fourier transform in $L^2(\mathbb{R}^d)$,

$$i\hat{\psi}' = 4\pi^2\xi^2\psi \Rightarrow \psi(x,t) = \widehat{\psi_0}(\xi)e^{-4\pi^2 i\xi^2 t}$$

The Fourier transform $\hat{\psi}$ is the probability amplitude of the momentum, ξ. We see that the probability distribuion in ξ is $|\widehat{\psi_0}|^2$, and it is independent of time. The momentum is conserved.

Now,

$$\psi(x,t) = \int_{\mathbb{R}^d} e^{-4\pi^2 i\xi^2 t + 2\pi i\xi x}\widehat{\psi_0}(\xi)d\xi$$

What happens when t becomes large? It is not difficult to see that the Riemann-Lebesgue lemma can be adapted to show that $\psi(x,t) \to 0$ as $t \to \infty$.

Concretely, let's assume that $\psi_0(x) = e^{-\pi\alpha x^2}$. Then, by Proposition 33.2.1, we have $\widehat{\psi_0}(\xi) = \alpha^{-d/2}e^{-\pi\xi^2/\alpha}$ and we get, using again Proposition 33.2.1,

$$\psi(x,t) = \alpha^{-d/2}\int_{\mathbb{R}^d} e^{2\pi i\xi x - \pi\xi^2(4\pi i t + \alpha^{-1})}d\xi = (1+4i\pi\alpha t)^{-d/2}\exp\left(-\frac{\pi\alpha x^2}{16\pi^2\alpha^2 t^2+1} + 4i\frac{\pi^2\alpha^2 t x^2}{16\pi^2\alpha^2 t^2+1}\right)$$

If $d = 3$ we see that the probabilty of finding the particle in a ball of fixed radius decays roughly like t^{-3}, while the shape of the probability distribution is an ever widening Gaussian. The particle disperses out of any finite region.

34.2 The Airy equation

The Airy functions Ai and Bi satisfy the ODE

$$y'' = xy$$

The solutions are entire, since it is a linear ODE with entire coefficients. Taking the Fourier transform (with the normalization $\int_{\mathbb{R}} e^{-i\xi x}y(x)dx$) we get

$$-\xi^2\hat{y} = i\frac{d\hat{y}}{d\xi}$$

with the solution

$$\hat{y} = Ce^{i\xi^3/3}$$

meaning

$$y(x) = \int_{-\infty}^{\infty} e^{i\xi^3/3 + i\xi x}d\xi$$

is (up to a multiplicative constant) one of the two linearly independent solutions of the ODE. With the normalization above, it is indeed, the Airy function $Ai(x)$. Or is it even a solution of the ODE? If we differentiate twice in x under the integral sign, we get an integral that does not converge, even conditionally.

But this does *not* mean that $y''(x)$ does not exist! It simply means that the representation is inadequate for this purpose. Instead, the contour of integration can be homotopically rotated:

$$y(x) = \int_{-\infty e^{-\pi i/6}}^{\infty e^{\pi i/6}} e^{i\xi^3/3 + i\xi x}d\xi$$

In this way, whe $|\xi|$ is large, the integrand decreases roughly like $e^{-|\xi|^3/3}$, and $y(x)$ is now manifestly analytic in $\mathbb{C}$!

35 Convolutions and the Fourier transform

Recall that the convolution of f and g is defined as

$$(f * g)(y) = \int_{\mathbb{R}^n} f(x)g(y-x)dx$$

Theorem 31.1.1 shows in particular that convolution is well defined on $L^1(\mathbb{R}^n) \times L^1(\mathbb{R}^n)$. The following theorem shows, in particular, that multiplication and convolution are Fourier-dual to each other.

Theorem 35.0.1. Suppose $f, g \in L^1$. Then

$$\widehat{f * g} = \hat{f}\hat{g}; \ \widehat{fg} = \widehat{f * g}$$

and, if $a \in \mathbb{R}^n$, ten

$$\widehat{\tau_a f}(\xi) = e^{i\xi a}\hat{f}(\xi)$$

Proof. This is a calculation, relying on Fubini:

$$(\widehat{f * g})(\xi) = \iint f(y-x)g(x)e^{-2\pi i\xi y}dxdy = \iint f(y-x)g(x)e^{-2\pi i\xi x}e^{-2\pi i\xi(y-x)}dxdy$$
$$= \int f(y-x)e^{-2\pi i\xi(y-x)}d(y-x)\int g(x)e^{-2\pi i\xi x}dx = \hat{f}(\xi)\hat{g}(\xi) \quad (216)$$

The equality immediately following it is now obvious by the inversion formula. The last equality is clear from an immediate calculation. □

As a result, we should investigate further the properties of convolution.

Note 35.0.2. Assuming that the integrals are well-defined (e.g., $f, g \in L^1$),

a) $f * g = g * f$. This follows from the density of L^2 and the fact that $\hat{f}\hat{g} = \hat{g}\hat{f}$

b) $(f * g) * h = f * (g * h)$. (By the argument in (a).)

c) For $a \in \mathbb{R}^n$, $\tau_a(f * g) = (\tau_a f) * g = f * (\tau_a g)$. (By Theorem 35.0.1 and the argument in (a).)

d) If $f, g \in \mathcal{S}$, then $f * g \in \mathcal{S}$. (By the argument in (a).)

e) If $A = \{x + y : x \in \operatorname{supp}(f), y \in \operatorname{supp}(g)\}$, then $\operatorname{supp}(f * g) \subset A$. This follows from the fact that for all x, if $z \notin A$ then $f(x)g(z-x) = 0$.

Proposition 35.0.3. Let p, q be conjugate exponents, $f \in L^p, g \in L^q$. Then $f * g$ exists pointwise everywhere, $f * g \in BC(\mathbb{R}^n)$ and $\|f * g\|_\infty \leqslant \|f\|_p \|g\|_q$. Furthermore, if $p \in$

$(1,\infty)$, then $f * g \in C_0$.

Proof. Pointwise existence and the uniform bound follow right away from Hölder's inequality. Noting that

$$\int f(x)g(y-x)dx = \int f(x)(Sg)(x)ds$$

where $S = \tau_y \circ J$, $(Jg)(x) = g(-x)$, continuity follows from the fact that translation is continuous in L^p, Lemma 33.0.1. Finally, we note that $p \in (1,\infty)$ implies $q \in (1,\infty)$ and thus $C_c(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ and in $L^q(\mathbb{R}^n)$. By Proposition 35.0.2 e) $C_c(\mathbb{R}^n)$ is preserved by convolution, and if $f_n \to f$ in $L^p(\mathbb{R}^n)$ and $g_n \to g$ in $L^q(\mathbb{R}^n)$, then, by the first part of the Proposition, $f_n * g_n \to f * g$ uniformly. Since the uniform closure of $C_c(\mathbb{R}^n)$ is $C_0(\mathbb{R}^n)$, the result follows. □

From the theorem of differentiation under the integral sign we obtain the following is a refinement of Proposition 35.0.2, d).

Proposition 35.0.4. If $f \in L^1(\mathbb{R}^n)$ and $g \in C^k(\mathbb{R}^n)$ with $\partial^\alpha g \in BC(\mathbb{R}^n)$ for $|\alpha| \leqslant k$, then $f * g \in C^k(\mathbb{R}^n)$ and for all $\alpha, |\alpha \leqslant k$ we have $\partial^\alpha(f * g) = f * (\partial^\alpha g)$.

36 The Poisson summation formula

Theorem 36.0.1. *Assume $f \in C(\mathbb{R}^n)$, $\||x|^{n+\varepsilon} f(x)\|_\infty < \infty$, and $\||\xi|^{n+\varepsilon}\hat{f}(\xi)\|_\infty < \infty$ for some $\varepsilon > 0$. Then,*

$$\sum_{j\in\mathbb{Z}^n} f(j) = \sum_{j\in\mathbb{Z}^n} \hat{f}(j)$$

and more generally,

$$\sum_{j\in\mathbb{Z}^n} f(x+j) = \sum_{j\in\mathbb{Z}^n} \hat{f}(j)e^{2\pi i j\cdot x} \tag{217}$$

The sum $(Pf)(x) := \sum_{j\in\mathbb{Z}^n} f(x+j)$ is called a *periodization* of f.

Proof. Note first that, under the given assumptions, the sums are uniformly and absolutely convergent. The function $\sum_{k\in\mathbb{Z}^n} f(x+k)$ is in $\mathbb{C}(\mathbb{T}^n) \subset L^2(\mathbb{T}^n)$.

$$\begin{aligned}\hat{f}(j) = \int_{\mathbb{R}^n} e^{-2\pi i j\cdot x} f(x)dx &= \sum_{m\in\mathbb{Z}^n} \int_{\mathbb{T}+m} e^{-2\pi i j\cdot x} f(x)dx \\ &= \sum_{m\in\mathbb{Z}^n} \int_{\mathbb{T}} e^{-2\pi i j\cdot x} f(x+m)dx = \int_{\mathbb{T}} e^{-2\pi i j\cdot x} \sum_{m\in\mathbb{Z}^n} f(x+m)dx \end{aligned} \tag{218}$$

For integer j, $\hat{f}(j)$ is the Fourier coefficient of Pf, hence

$$\sum_{j\in\mathbb{Z}^n} \hat{f}(j)e^{2\pi i j x} = \sum_{j\in\mathbb{Z}^n} f(x+j)$$

□

This theorem has many applications, such as calculating sums in closed form, when the Fourier transform of a function is more easily summed than the function itself.

For instance, if $a \in \mathbb{R}^+$, we have

$$\widehat{\frac{1}{x^2+a^2}} = a^{-1}e^{-a|\xi|}$$

which implies, using Poisson summation, that

$$\sum_{j\in\mathbb{Z}} \frac{1}{j^2+a^2} = \pi a^{-1}\coth(a\pi) \tag{219}$$

which, by a limiting procedure (check!) contains the special case

$$\sum_{j\in\mathbb{N}} \frac{1}{j^2} = \frac{\pi^2}{6}$$

Eq. (219) is an instance of the Mittag-Leffler theorem, which expresses a meromorphic function by a "partial-fraction-like" expansion. In the same way we get

$$\sum_{j\in\mathbb{N}} \frac{1}{j^4+a^4} = \frac{\pi\left(\sinh\sqrt{2}\pi a + \sin\sqrt{2}\pi a\right)}{\sqrt{2}a^3\left(\cosh\sqrt{2}\pi a - \cos\sqrt{2}\pi a\right)}$$

implying (how?)

$$\sum_{j\in\mathbb{N}} \frac{1}{j^4} = \frac{\pi^4}{90}$$

Definition 36.0.2. The Jacobi theta function is defined as

$$\vartheta(z;\tau) = \sum_{n=-\infty}^{\infty} \exp\left(\pi i n^2\tau + 2\pi i n z\right) = 1 + 2\sum_{n=1}^{\infty}\left(e^{\pi i\tau}\right)^{n^2}\cos(2\pi n z) = \sum_{n=-\infty}^{\infty} q^{n^2}\eta^n,\ \Re\tau > 0 \tag{220}$$

Here z is any complex number, τ, confined to the upper half plane, is the *half-period ratio*, and q is the *nome*. In terms of θ we have $H_t(x,t) = \vartheta(x;4\pi i t)$.

Exercise 90. Prove the Jacobi duality formula

$$\sum_{n\in\mathbb{Z}} e^{-\pi n^2 x} = x^{-1/2}\sum_{n\in\mathbb{Z}} e^{-\pi\frac{n^2}{x}}$$

This identity is crucial to understanding the way the theta function transforms under the modular group.

36.1 The Gibbs phenomenon

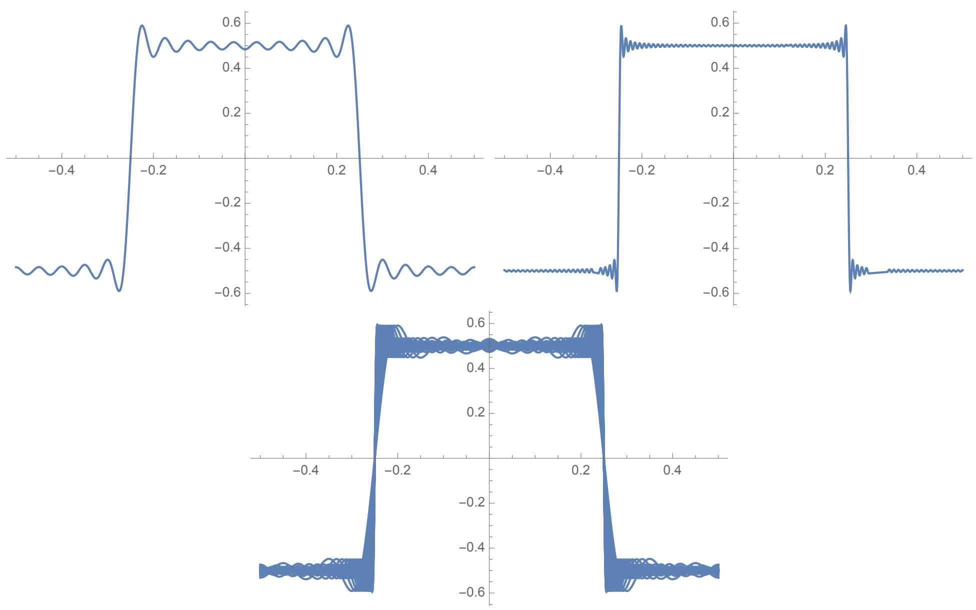

Figure 5: Characteristic function of $[-1/4, 1/4]$: the partial Fourier sum with 20 terms (left), 100 (right) and the graphical superposition of the Fourier sum with $20 - 100$ terms (below).

The Gibbs phenomenon is the remarkable way in which the Fourier series behaves at a jump discontinuity of a piecewise smooth function. The Gibbs phenomenon can be heard as "ringing" near transients, such as sounds from percussion instruments. It roughly results from the fact that we are trying to approximate a discontinuous function by smooth ones. Recalling the duality between smoothness and decay of the Fourier coefficients, a discontinuity will result in their slow decay. Therefore, the Fourier terms in the difference between a partial sum and the limit will have significant amplitude, resulting in fast oscillating defects. This "defect" only occurs in finite sums, since we know that in the limit the Fourier series converges everywhere to the average of the left and right limits of a piecewise-smooth function. This also means that the location of the maximum defect changes with the number of terms, to allow for the limit to exist.

The Fourier sums of the function $f(x) = -1$ if $x \in (-1/2, 0)$ and 1 if $x \in (0, 1/2)$[18] is

$$S_N(x) = \sum_{k=0}^{N} \frac{4\sin(2\pi(2k+1)x)}{\pi(2k+1)} \tag{221}$$

Then,

$$S_N'(x) = 4\frac{\sin(4\pi(N+1)x)}{\sin(2\pi x)} \tag{222}$$

[18]Note that the point values of the function at the discontinuity are irrelevant, as they wash out as a result of the integration involved in calculating Fourier coeffcients.

(we recognize the Dirichlet kernel in (222); is this a coincidence?) An elementary argument shows that the first positive maximum of S_N occurs at $x_0 = \frac{1}{4(N+1)}$. We have

$$S_N(x_0) = \sum_{k=0}^{N} \frac{4\sin\left(\frac{2\pi(2k+1)}{4N+4}\right)}{\pi(2k+1)} \to \frac{1}{\pi}\int_0^{\pi} \frac{\sin x}{x} dx = \pi^{-1}\,\mathrm{Si}(\pi) = 1.1789797444\cdots \text{ as } N \to \infty$$

by observing that $S_N(x_0)$ is a Riemann sum for the integral. We see that the sums converge nonuniformly to f, with an "overshot" of about 18% in uniform norm.

Exercise 91. *Show that the overshot by a factor of $\pi^{-1}\,\mathrm{Si}(\pi)$ of the Fourier sums occurs is the same at any jump discontinuiuty of a piecewise smooth function.*

37 Applications to PDEs

In this chapter we use ξ for the Fourier variable: this is the most frequent convention in PDEs.

37.1 The heat equation on the circle

This is given by

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad u(x,0) = f(x); \quad f \in C^1(\mathbb{T}) \tag{223}$$

We have already shown uniqueness of solutions of (223). For existence we write

$$f(x) = \sum_{j\in\mathbb{Z}} a_j e^{2\pi i j x}$$

By separation of variables we get

$$u(x,t) = \sum_{n\in\mathbb{Z}} f_n e^{-4\pi^2 n^2 t} e^{2\pi i n x} \tag{224}$$

With $\hat{f} = (f_n)_{n\in\mathbb{Z}}$ $\hat{\mathcal{H}}_t = (e^{-4\pi^2 n^2 t})_{n\in\mathbb{Z}}$, the Fourier coefficients of the **heat kernel for the circle**

$$H_t(x) = \sum_{n\in\mathbb{Z}} e^{-4\pi n^2 t} e^{2\pi i n x} \tag{225}$$

we have $\hat{u} = \hat{f}\hat{\mathcal{H}}_t$ and therefore

$$u = f * H_t \tag{226}$$

(where $(f*g)(y) = \int_0^1 f(x)g(y-x)dx$).

37.2 The heat equation on the line; smoothening by convolution

This is the same as (223), except with $x \in \mathbb{R}$. With uniqueness settled in §30, we show existence, and in fact construct the solution, by Fourier transform in x:

$$\frac{\partial \hat{u}}{\partial t} = -4\pi^2 \xi^2 \hat{u} \tag{227}$$

which gives

$$\hat{u}(t,\xi) = \hat{f}(\xi)e^{-4\pi^2\xi^2 t} \tag{228}$$

As before, the convolution theorem implies

$$u = f * \mathcal{H}_t \tag{229}$$

where

$$\mathcal{H}_t(x) = \mathcal{F}^{-1}\left(e^{-4\pi^2\xi^2 t}\right) = (4\pi t)^{-1/2}e^{-x^2/4t} \tag{230}$$

Theorem 37.2.1. If $f \in \mathcal{S}$, then the solution of (223) on $\mathbb{R}$ is (229), and $u(t,\cdot) \in \mathcal{S}$; $\|u(t,\cdot) - f\|_u \to 0$ and $\|u(t,\cdot) - f\|_2 \to 0$ as $t \to 0$

Proof. First, note that $\hat{u}(t,\cdot) \in \mathcal{S}$ for $t > 0$, implying that $u(t,\cdot) \in \mathcal{S}$ for $t > 0$. Next,

$$|u(x,t) - f(x)| = \left|\int_{\mathbb{R}} \hat{f}(\xi)\left(e^{-4\pi^2\xi^2 t} - 1\right)e^{2\pi i\xi x}d\xi\right| \leqslant \int_{\mathbb{R}} |\hat{f}(\xi)|\left|e^{-4\pi^2\xi^2 t} - 1\right| d\xi \to 0 \tag{231}$$

as $t \to 0$ by dominated convergence. For the L^2 norm, by Plancherel,

$$\|u(t,\cdot) - f\|_2^2 = \|\hat{u}(t,\cdot) - \hat{f}\|^2 = \int_{\mathbb{R}} |\hat{f}(\xi)|^2\left|e^{-4\pi^2\xi^2 t} - 1\right|^2 d\xi \to 0 \tag{232}$$

as $t \to 0$ again by dominated convergence. □

Corollary 37.2.2 (Smoothening by convolution)**.** Let $f \in C_c(\mathbb{R})$. Then $g_t = f * H_t \in \mathcal{S}$ (in fact, g_t is *entire*) and $g_t \to f$ uniformly as $t \to 0$.

Proof. Indeed, if $f \in C_c(\mathbb{R})$, then $\hat{f} \in C^\infty(\mathbb{R}) \cap C_0(\mathbb{R})$, hence $\hat{f}(\xi)e^{-4\pi^2\xi^2 t} \in \mathcal{S}$ (using the super-exponential decay in ξ you can show that, in fact, g_t is entire). The rest follows from Theorem 37.2.1. □

Theorem 37.2.3. The heat kernel on the circle is the periodization of the heat kernel on the line:

$$H_t(x) = \sum_{n\in\mathbb{Z}} \mathcal{H}_t(x+n) \tag{233}$$

Proof. This follows immediately from (230), (225) and the general form of Poisson's summation formula. □

Corollary 37.2.4. The heat kernel on the circle is positive, and the family $\{H_t\}_{t\geqslant 0}$ is an approximation to the identity.

Proof. Positivity follows from (233). It is clear from (225) that $\int_{-1/2}^{1/2} H_t(x)dx = 1$ (since only the term with $n = 0$ contributes). We have to show that the integral of H_t over an interval not containing 0, say $J = (\alpha, \beta)$ where $0 < \alpha < \beta < 1/2$ vanishes in the limit $t \to 0$. Note that for $x \in J$ and $0 \neq n \in \mathbb{Z}$ we have $|1 + x/n| \geqslant |1 - x| \geqslant |1 - \beta| := \varepsilon$, implying $|x + n| \geqslant |n|\varepsilon$ and thus

$$\sum_{|n|\geqslant 1} \mathcal{H}_t(x+n) \leqslant \sum_{|n|\geqslant 1} (4\pi t)^{-1/2} e^{-\varepsilon^2 n^2/4t} \to 0 \text{ as } t \to 0 \tag{234}$$

by monotone convergence, which implies, by dominated convergence,

$$\begin{aligned}\int_J H_t(x)dx &\leqslant \frac{e^{-\alpha^2/4t}}{(4\pi t)^{1/2}}(\beta - \alpha) + \int_J \sum_{|n|\geqslant 1} \mathcal{H}_t(x+n)dx \\ &\leqslant \frac{e^{-\alpha^2/4t}}{(4\pi t)^{1/2}}(\beta - \alpha) + \sum_{|n|\geqslant 1} (4\pi t)^{-1/2} e^{-\varepsilon^2 n^2/4t} \to 0 \text{ as } t \to 0 \end{aligned} \tag{235}$$

□

Corollary 37.2.5. For any **continuous** initial condition f, the heat equation on the circle has a unique smooth solution, $u(x,t) = (H_t * f)(x)$.

Proof. Indeed, $H_t * f$ is smooth and solves the heat equation for any $t > 0$ and, by Corollary 37.2.4, $\lim_{t\to 0} H_t * f = f$. □

37.3 General linear PDEs

A differential operator is an operator L of degree m has the general form

$$L = \sum_{|\alpha|\leqslant m} a_\alpha(x) D^\alpha; \text{ giving } Lf = \sum_{|\alpha|\leqslant m} a_\alpha(x)\partial^\alpha f$$

and it is with constant coefficients if $a_\alpha(x) = b_\alpha$ are independent of x. Let $f \in \mathcal{S}$. Then,

$$\mathcal{F}(Lf)(\xi) = \sum_{|\alpha|\leqslant m} (2\pi i)^{-|\alpha|} b_\alpha \xi^\alpha \hat{f}(\xi) \tag{236}$$

and we see that the equation $Lf = g$ in $\mathbb{R}^n$ reduces to a polynomial equation, whenever we can indeed apply the Fourier transform.

The polynomial

$$\sum_{|\alpha|\leqslant m} b_\alpha \eta^\alpha$$

is called **total symbol** of L, or simply **symbol**. The part of the symbol containing the terms of highest degree only,

$$\sum_{|\alpha|=m} b_\alpha \eta^\alpha$$

is called **principal symbol**. For a second order partial differential operator L with principal symbol

$$\sum_{i+j=2} b_{ij}\eta_1^i\eta_2^j$$

the operator is called **elliptic** if the matrix $B = \{b_{ij}\}_{i,j}$ is positive or negative definite, **hyperbolic** if B is not definite but $\det(B) \neq 0$ and **parabolic** if exactly one eigenvalue of B is zero. Thus, the Laplacian Δ is elliptic, the wave operator $\Box = \partial_t^2 - \partial_y^2$ is hyperbolic, and the heat operator $\partial_t - \partial_x^2$ is parabolic. The names of the three types above derive from the form of the symbol: for the Laplacian, the symbol is $\eta_1^2 + \eta_2^2$ whose level lines are ellipses; the level lines are hyperbolas for $\eta_1^2 - \eta_2^2$; the heat equation has total symbol $\eta_1 + \eta_2^2$ whose level lines are parabolas; whether the parabola is concave or convex is also important. Let's examine these four types of equations on the circle, with conditions (initial, boundary, etc) in $\mathcal{S}$ performing (discrete) Fourier transform in one variable only.

For the wave equation, we get

$$[\hat{u}_{tt}]_j = -4\pi^2 j^2 \hat{u}_j$$

with solutions $u_j = a_j e^{-2\pi ijt} + b_j e^{2\pi ijt}$, meaning

$$u(x,t) = \sum_{j\in\mathbb{N}} a_j e^{-2\pi ij(x+t)} + \sum_{j\in\mathbb{N}} b_j e^{2\pi ij(x-t)}$$

and the solution is completely determined if we provide $u(x,0), u_t(x,0)$. We also note that $u(x,t) = f(x+t) + g(x-t)$. Recall also the solutions of the Laplace equation, Theorem 30.0.2 and of the heat equation on the circle, Exercise 78.

The backward heat equation, $u_t = -u_{xx}$ would formally give

$$\sum_{j\in\mathbb{Z}} a_k e^{4\pi^2 j^2 t + ijx}$$

and, for generic initial conditions in $\mathcal{S}$, this is nonsense for any $t > 0$ (the solution, assumed C^2 in x, would have a convergent Fourier series if it existed at all).

The Laplace equation $\Delta u = 0$ is elliptic, and in a given domain it needs **one** boundary condition: either $u_{\partial\Omega} = f$ or the normal derivative $\frac{\partial u}{\partial n}|_{\partial\Omega} = g$.

Note the important role of the *principal symbol*: its nature dictates the growth of the Fourier coefficients, which control the existence and smoothness of solutions.

Note also that if we have a linear PDE with constant coefficients, a Fourier transform converts it to a polynomial equation which can be solved in closed form.

37.4 Operators and symmetries

If $\mathcal{G}$ is a group of transformations on a space of functions $\mathfrak{F}$ and L is a map from $\mathfrak{F}$ to $\mathfrak{F}$, then L is invariant under $\mathcal{G}$ if $L(\gamma f) = \gamma L f$ for all $\gamma \in \mathcal{G}$ and $f \in \mathfrak{f}$.

Another way to write this is to note that $f \mapsto \gamma f$ is itself a linear operator; call it Γ. Then,

L is invariant $\mathcal{G}$ iff, for any $\gamma \in \mathcal{G}$, L and Γ commute, $L\Gamma - \Gamma L =: [L, \Gamma] = 0$. Symmetries often place such restrictions on L that the operator is virtually determined by them. In physics, this is an important way to determine the fundamental laws of various theories.

Let's look at the question of which second order operators commute with the isometries of $\mathbb{R}^n$, the group generated by $\mathcal{T}$ and $O(n)$. Recalling our more general analysis of isometries of Hilbert spaces, all elements of $O(n)$ must be (real-valued) unitary transformations, $R \in O(n) \Rightarrow RR^* = I = RR^t$. In particular, $|\det R| = 1$.

Lemma 37.4.1. The Fourier transform commutes with $O(n) : R \in O(n) \Rightarrow R(\mathcal{F}f(\xi)) = (\mathcal{F}f)(R\xi) = (\mathcal{F}f(R\cdot)(\xi)$.

Proof. Changing variable $Rx = y$,

$$\widehat{f(Rx)} = \int_{\mathbb{R}^n} e^{-2\pi i \langle \xi, x \rangle} f(Rx)dx = \int_{\mathbb{R}^n} e^{-2\pi i \langle \xi, R^t y \rangle} f(y)dy = \int_{\mathbb{R}^n} e^{-2\pi i \langle R\xi, y \rangle} f(y)dy = \hat{f}(R\xi)$$

□

Theorem 37.4.2. A differential operator L commutes with the isometries of $\mathbb{R}^n$ iff it is a polynomial in Δ, $L = Q(\Delta)$.

Proof. It is easy to see, as in the beginning of the paragraph, that L must have constant coefficients. In Fourier space it is a polynomial in $P(\xi)$ which, by Lemma 37.4.1, commutes with $O(n)$. We decompose the polynomial by homogeneous components,

$$P(\xi) = \sum_{m=0}^{M} \sum_{|\alpha|=n} a_\alpha \xi^\alpha = \sum_{m=0}^{M} P_m(\xi)$$

Next, we note that

$$0 = \lambda^{-M}[P(R\lambda\xi) - P(\lambda\xi)] \Rightarrow \lim_{\lambda\to\infty} \lambda^{-M}[P(\lambda R\xi) - P(\lambda\xi)] = P_M(R\xi) - P_M(\xi) = 0$$

This means that the highest order homogeneous polynomial is itself $O(n)$-invariant. Subtracting P_M from P and repeating the argument implies that P_{M-1} commutes with $O(n)$ and inductively, all homogeneous components $P_j(\xi)$ do. Take the unit sphere, $S = \{\xi : |\xi| = 1\}$ and note that $O(n)$ acts transitively on S. This follows from the exercise below. Thus $P_j(R\xi) = P_j(\xi)$ on S implies $P_j = a_j = const$ on S, entailing $P_j(\xi) = a_j|\xi|^j$ which is only possible if j is even, and thus $a_{2k+1} = 0$ and $P_{2k}(\xi) = a_{2j}(\xi_1^2 + ... + \xi_n^2)^j$. □

Exercise 92. Show that $SL(n)$ acts transitively on $\mathbb{R}^n \setminus \{0\}$ and (thus) $O(n)$ acts transitively on the unit sphere in $\mathbb{R}^n$.

37.5 Supplementary material: Adjoints of linear operators

Recall that for a bounded operator B in a Hilbert space $\mathcal{H}$, we can define the adjoint B^* by $\langle Bx, y\rangle = \langle x, B^*y\rangle$, where uniqueness is immediate and existence is guaranteed by the Riesz representation theorem. An operator L which is not necessarily bounded is defined on some domain $\mathrm{dom}(L) = \Omega$ (we may assume that Ω is dense in $\mathcal{H}$, otherwise the natural Hilbert space to work in would be $\mathcal{H}_1 = \overline{\Omega}$). Naturally, the adjoint of L would be an operator L^*, defined on some domain Ω^* with the property

$$\forall (x, y) \in \Omega \times \Omega^*,\ \langle Lx, y\rangle = \langle x, L^*y\rangle$$

Obvious questions are of course existence of such an L^*, and uniqueness. Uniqueness is easy: if we have two operators L_1^* and L_2^* with the property above, then for any y such that $L_{1,2}^*$ are both defined, we have, for any x in the dense set $\mathrm{dom}(L)$,

$$\langle x, (L_1^* - L_2^*)y\rangle = 0 \Rightarrow (L_1^* - L_2^*)y = 0$$

For existence, define

$$\mathrm{dom}(L^*) = \{y \in \mathcal{H} : \exists z \in \mathcal{H}, \langle Lx, y\rangle = \langle x, z\rangle\} \tag{237}$$

and define L^* on $\mathrm{dom}(L^*)$ by

$$L^*y := z \tag{238}$$

Definition 37.5.1. *An operator A on a dense domain $\Omega \subset \mathcal{H}$ is self-adjoint if $A^* = A$. Note that this means that* $\mathrm{dom}(A^*)$ *is no more, and no less than* $\mathrm{dom}(A)$.

Proposition 37.5.2. *Let U be unitary from $\mathcal{H}$ to $\mathcal{H}'$ and $A : \Omega \to \mathcal{H}$ a linear operator with dense domain. Then $U\Omega$ is dense in $\mathcal{H}'$, UA is well defined on $U\Omega$ and its adjoint in UA^*.*

Proof. Since U is an isomorphism, this is a straightforward verification. ☐

Example 37.5.3. Consider the operator $D = i\frac{d}{dx}$ on $\mathbb{T}$. First, we see that for smooth functions, say in $C^\infty(\mathbb{T})$, $\langle Df, g\rangle = \langle f, Dg\rangle$ and D^* exists at least on $C^\infty(\mathbb{T})$, and on it $D^* = D$. From the definition of the adjoint, it is clear that the domain of D^* gets larger if the domain of D shrinks. Suppose we want to determine first the "maximal" set of functions in $L^\infty \subset L^2$ on which we can define differentiation. We keep then $C^\infty(\mathbb{T})$ as an initial domain for D (or choose an even smoother space if it helps), and determine the corresponding domain of D^*.

Let $U = \mathcal{F}$, the discrete Fourier transform, a unitary map between $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$. Then UD is the operator of multiplication by $-2\pi k$ and, to understand what the adjoint of D is, it is enough to determine the adjoint of $-2\pi k$. We have

$$\langle Df, g\rangle = \sum_{k\in\mathbb{Z}} (-2\pi k f_k)\overline{g_k} = \sum_{k\in\mathbb{Z}} f_k \overline{(-2\pi k g_k)} =: \sum_{k\in\mathbb{Z}} f_k z_k$$

which implies $z_k = (-2\pi k g_k), k \in \mathbb{Z}$. Thus, $\mathrm{dom}(D^*) = \Omega^* = \{g \in \mathcal{H} : (kg_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})\}$. Let $(h_k)_{k\in\mathbb{Z}} \in \ell^2(\mathbb{Z})$ be given by $h_k = -2\pi i k g_k, k \in \mathbb{Z}$ and $h = \mathcal{F}^{-1}((h_k)_k)$. For $k \neq 0$ we have

$g_k = i/(2\pi k)h_k$ which means that $g = \int_0^x h + \Lambda(f)$ where $\Lambda(g)$ is an additive constant, which is a **bounded linear functional** on L^2 (why?). Now, $h \in L^2$ implies $g \in AC(\mathbb{T})$ (with derivative in L^2). This is the largest domain of D, with range in L^2. In this simple example, if we extend $\mathrm{dom}(D)$ to Ω^*, the same argument shows that this extended D is self-adjoint.

Example 37.5.3 indicates that if we want to extend D even further, then the extended domain, or range, or both cannot consist of usual functions, even allowing for the generalizations used in the L^p spaces.

Let us first relax the restriction on the range. The dual of $C_c([-a,a])$ is the space of Radon measures on $[-a,a]$. The Heaviside function $\Theta(x)$ is not in AC (it's not even continuous, of course). As an element of the dual of C_c it acts as $\langle\Theta,\varphi\rangle = \int_0^a \varphi(x)dx$. Proceeding as in the previous example, taking φ in the dense set $C^1([-a,a])$, we would define $\Lambda = \frac{d}{dx}\Theta$, as an element of C_c^* by

$$\Lambda\varphi = -\langle\Theta, \frac{d}{dx}\varphi\rangle = -\int_0^a \varphi'(s)ds = \varphi(0) \Rightarrow \Lambda = \delta(x)$$

where $\delta(x)$ is the Dirac mass measure at zero. Thus $\Theta'(x) = \delta(x)$ exists, as a measure, $\delta(x)$. In the same manner, we would get

$$\Theta''(x) = (\varphi \mapsto -\varphi'(0))$$

This is obviously not defined as a bounded functional on $C_c([-a,a])$, but it is in $(C^1([-a,a]))^*$. This logic prompts us to consider the baseline space of **test functions** $\mathcal{D} = C_c^\infty(\mathbb{R}^n)$.

37.6 Supplementary material: the Fourier transform of functions analytic in the lower half plane and the Laplace transform

Let $f \in L^1 \cap C_0(\mathbb{R})$ be s.t. $\hat{f} \in L^1$. Recall that this implies that $\mathcal{F}^{-1}\hat{f} = f$.

Proposition 37.6.1. *(i) Assume that $f \in L^1 \cap C_0(\mathbb{R})$ is s.t. $\hat{f} \in L^1$, and that f is analytic in the upper half plane $\mathcal{H}$, and that $|f(z)| \to 0$ as $|z| \to \infty$ in $\mathcal{H}$. Then $\hat{f}(\xi) = 0$ if $\xi < 0$.*

(ii) Assume $f \in L^1 \cap C_0(\mathbb{R})$ and $f(\xi) = 0$ for $\xi < 0$. Then $\check{f}$ is analytic in the upper half plane and $|f(z)| \to 0$ as $|z| \to \infty$ in $\mathcal{H}$.

Proof. (i) Let $\xi > 0$. Take C_r to be the three upper sides of a box in $\mathbb{C}$: the segment from r to $r - ir$, followed by the segment from $r - ir$ to $-r - ir$ and finally from $-r - ir$ to $-r$. Check that $\int_{C_r} e^{-i\xi x} f(x)dx \to 0$ as $r \to \infty$. Fix an ε and choose r large enough so that $|\int_{|x|>r} e^{-i\xi x} f(x)dx| + |\int_{C_r} e^{-i\xi x} f(x)dx| < \varepsilon$. We then have

$$\left| \int_{\mathbb{R}} e^{-i\xi x} f(x)dx - \int_{[-r,r]\cup C_r} e^{-i\xi x} f(x)dx \right| < \varepsilon$$

where $\int_{[-r,r]\cup C_r}$ means the integral over $[-r,r]$ followed by the integral on C_r discussed above. On the other hand, since f is analytic, Cauchy's theorem implies that $\int_{[-r,r]\cup C_r} e^{-i\xi x} f(x)dx = 0$, and since ε is arbitrary, the result follows.

(ii) Simply use dominated convergence and the Riemann-Lebesgue lemma.

$\square$

Definition 37.6.2. *Let $F \in L^1(\mathbb{R}^+)$. The Laplace transform of $\mathcal{L}$ is defined as*

$$(\mathcal{L}F)(x) = \int_0^\infty e^{-px} F(p) dp, \quad \Re x > 0$$

More generally, if $e^{-ax}F \in L^1$ for some $a > 0$, then $\mathcal{L}F$ is defined by the same formula, for $\Re x > a$.

Theorem 37.6.3. *If $F \in L^1(\mathbb{R}^+)$, then $f(x) = \mathcal{L}F$ is analytic in the right half plane $\mathbb{H}$ and continuous in $\overline{\mathbb{H}}$. If $f(ix) \in L^1(\mathbb{R})$, then, for $p > 0$, F is given by the inverse Laplace transform,*

$$F(p) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ipx} f(ix) dx = \frac{1}{2\pi i} \int_{-i\infty}^{i\infty} f(z) e^{pz} dz$$

If $\sup |z^a f(z)| < \infty$ for some $a > 1$, then we equivalently have

$$F(p) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f(z) e^{pz} dz$$

for any $c \geqslant 0$.

Proof. Analyticity in $\mathbb{H}$ follows from the fact that $F \in L^1$: dominated convergence allows then for differentiation inside the integral. Continuity in $\overline{\mathbb{H}}$ also follows from dominated convergence. In the limit $x \to 2\pi i t$, we get

$$(\mathcal{L}F)(2\pi i t) = \int_0^\infty e^{-2\pi i t p} F(p) dp = \hat{F}(t)$$

The rest follows from the Fourier inversion theorem. □

38 Distribution theory

"Il y a plus de 50 ans que l'ingénieur Heaviside introduisit ses régles de calcul symbolique, dans un mémoire audacieux où des calculs mathématiques fort peu justifiés sont utilisés pour la solution de problèmes de physique. Ce calcul symbolique, où opérationnel, n'a cessé de se développer depuis, et sert de base aux études théoriques des électriciens. Les ingénieurs l'utilisent systématiquement, chacun avec sa conception personnelle, avec la conscience plus ou moins tranquille ; c'est devenu une technique «qui n'est pas rigoureuse mais qui réussit bien». Depuis l'introduction par Dirac de la fameuse fonction $\delta(x)$, qui serait nulle partout sauf pour $x = 0$, de telle sorte que $\int_{-\infty}^{\infty} \delta(x) dx = +1$, les formules du calcul symbolique sont devenues encore plus inacceptables pour la rigueur des mathématiciens. Écrire que la fonction d'Heaviside $Y(x)$ égale 0 pour $x < 0$ et a 1 pour $x \geqslant 0$ a pour dérivée la fonction de Dirac $\delta(x)$ dont la définition même est contradictoire, et parler des dérivées $\delta'(x), \delta''(x)$,... de cette fonction denude d'existence réelle, c'est dépasser les limites qui nous est permises. Comment expliquer le succès de ces méthodes? Quand une telle situation contradictoire se présente, il est bien rare qu'il n'en résulte pas une théorie mathématique nouvelle qui justifie, sous une forme modifiée, le langage des physiciens ; il y a même là une source importante de progrès des mathématiques et de la physique."

"More than 50 years ago the engineer Heaviside introduced his symbolic calculus rules, in an audacious memoir in which mathematical calculations with scant justification were used to solve physical problems. This symbolic calculus, or operational calculus, has not ceased to be developed since, and serves as a foundation for the theoretical studies of electricians. The engineers use it systematically, everyone using his own conception, with a more or less peaceful conscience; it has become a technique "which is not rigorous, but is successful". Ever since Dirac's introduction of the famous function $\delta(x)$, which would be zero everywhere except at $x = 0$, in such a way that $\int_{-\infty}^{\infty} \delta(x)dx = +1$, the formulas of symbolic calculus have become even more unacceptable for the rigor of mathematicians. To write that the Heaviside function $Y(x)$ which equals 0 fo $x < 0$ and 1 for $x \geqslant 1$ has as a derivative the Dirac function $\delta(x)$, whose very definition is contradictory, and then talk about the derivatives $\delta'(x), \delta''(x)$,... of this function devoid of real existence, is to exceed the limits that are permitted to us. How can one explain the success of these methods? When such a contradictory situation presents itself, it is rarely not the case that a new mathematical theory emerges, which justifies, in a modified form, the language of of physicists; there is even, in this, an important source of progress of mathematics and physics."

Laurent Schwartz, Théorie des Distributions

38.1 The space of test functions $\mathcal{D} = C_c^\infty(\mathbb{R}^n)$, and its topology

The topology on $\mathcal{D} = C_c^\infty(\mathbb{R}^n)$ is that of an inductive limit of Fréchet space (called "an LF space". Here we characterize the topolgy by its properties.

(i) A sequence $\{f_n\}_{n\in\mathbb{N}}$ converges to $f \in \mathcal{D}$ iff there is an $n_0 \in \mathbb{N}$ and a compact set K such that **all** f_n with $n \geqslant n_0$ are **supported in** K, and

$$\forall \alpha \in (\mathbb{N} \cup \{0\})^n, \lim_{n\to\infty} \|\partial^\alpha(f_n - f)\|_u = 0$$

(ii) A set $S \subset \mathcal{D}$ is bounded iff there is a compact set K s.t. S is a bounded subset of $C_c^\infty(K)$.

(iii) A sequence is Cauchy if there is a compact set K s.t. all functions are supported in K and the sequence is Cauchy in $C_c^\infty(K)$.

(iv) Let Y be a locally convex topological space. A mapping $A : \mathcal{D} \to Y$ is continuous if it is continuous on every $C_c^\infty(K)$, K a compact set.

(v) A linear functional $\Lambda : \mathcal{D} \to \mathbb{C}$ is continuous iff there is an N and a K s.t.

$$|\Lambda\varphi| \leqslant c_K \sup\{|\partial^\alpha \varphi| : x \in K, |\alpha| \leqslant N\}$$

Definition 38.1.1. 1. $\mathcal{D}'$, the dual of $\mathcal{D}$, is **the space of distributions**. If $F \in \mathcal{D}'$ its value on the function $\varphi \in \mathcal{D}$ is denoted by $\langle F, \varphi\rangle$, or sometimes, when no confusion is possible, $\int F(x)\varphi(x)dx$.

2. The **topology on** $\mathcal{D}'$ is chosen to be the weak-* topology: a net of $(F_\alpha)_{\alpha\in A}$ of distributions converges iff $(\langle F_\alpha, \varphi\rangle)_{\alpha\in A}$ converges for every $\varphi \in \mathcal{D}$.

Note 38.1.2. 1. $\mathcal{D}$ is **not** a sequential space. Likewise, $\mathcal{D}'$ is **not** a sequential space [a] However, convergence along nets does not play any role in the basic construction of distributions.

2. The topology on $\mathcal{D}$ is not metrizable. Indeed, take a sequence of compact sets s.t. $K_j \uparrow \mathbb{R}^n$ as $j \to \infty$. Clearly, $\cup_j C_c^\infty(K_j) = \mathcal{D}$, $C_c^\infty(K_j)$ are closed, with empty interior, since they are proper subspaces of the topological vector space $\mathcal{D}$, as shown in the exercise below.

[a] See R. M. Dudley, Convergence of Sequences of Distributions, Proc. AMS 27, 3 (1971).

Exercise 93. Let V be a topological vector space and S a subspace of V with nonempty interior. Show that $V = S$.

One way is as follows: since the vector space operations, addition and scalar multiplication are continuous, if $\mathcal{O}$ is nonempty and open in S, then for any $s \in S$ we have $\mathcal{O} + s \subset S$. Take $s_0 \in \mathcal{O}$, and note that $\mathcal{O} - s_0$ contains the origin. For any $v \in V$, the map $F_v : \mathbb{C} \to V$ given by $F_v(\lambda) = \lambda v$ is continuous, and thus $F_v^{-1}(\mathcal{O} - s_0)$ is open and nonempty, since it contains 0, and hence there is a nonzero λ so that $\lambda v \in \mathcal{O} - s_0$, thus $v \in S$, hence $S = V$.

Note 38.1.3. $\mathcal{D}$ is an inductive limit of Fréchet spaces (see Appendix A): let $(K_i)_{i\in\mathbb{N}}$ be compact sets in $\mathbb{R}^n$, whose union is $\mathbb{R}^n$ and such that for all i K_i is contained in the interior of K_{i+1} (e.g., the balls of radius i centered at the origin). Then $\mathcal{D}$ is the inductive limit of the sequence of Fréchet spaces $C_c^\infty(K_i)$.

Definition 38.1.4. It is often useful to restrict test functions to smaller sets: If $\mathcal{O}$ is open (K is compact), $\mathcal{D}(\mathcal{O})$ ($\mathcal{D}(K)$, resp.) denote the compactly supported infinitely differentiable functions whose support is contained in $\mathcal{O}$) (K resp.).

38.2 Examples of distributions

Check that the following are examples of elements of $\mathcal{D}'$:

1. (Distributions generalize functions.) Any $f \in L^1(\mathbb{R})$ is a distribution, acting on test functions by $\langle f, \varphi \rangle = \int_{\mathbb{R}^n} f(x)\varphi(x)dx$.
2. Consequently, $\mathcal{D}$ is embedded in $\mathcal{D}'$ by

$$\langle \psi, \varphi \rangle := \int_{\mathbb{R}^n} \psi(s)\varphi(s)ds \tag{239}$$

3. More generally, Radon measures are distributions acting by $\langle \mu, \varphi \rangle = \int_{\mathbb{R}^n} \varphi d\mu$.
4. An important example of a distributions of the form 2 is the Dirac distribution at zero, or the "delta function". This is the functional $\delta(x)$ defined by $\langle \delta, \varphi \rangle = \varphi(0)$. More generally, the delta function at x_0, $\delta_{x_0}(x)$ is the distribution $\langle \delta_{x_0}, \varphi \rangle = \varphi(x_0)$.
5. Derivatives of the delta function at a point, defined by $\langle \partial^\alpha \delta_{x_0}, \varphi \rangle = (-1)^{|\alpha|}(\partial^\alpha \varphi)(x_0)$ are distributions. Check that these derivatives, for $|\alpha| \neq 0$ are not of the form 3.
6. Let $F_N(x, y) = \sum_{k=-N}^{N} e^{-2\pi i k(x-y)}$. Then, with $\varphi \in C_c^\infty([-1/2, 1/2])$, and φ_k the Fourier coefficients of φ, we have

$$\langle F_N(\cdot, y), \varphi \rangle = \sum_{k=-N}^{N} \varphi_k e^{iky} \to \varphi(y) \text{ as } N \to \infty$$

and thus $\sum_{k=-N}^{N} e^{-2\pi i k(x-y)} \to \delta_y(x)$ as $N \to \infty$, in $\mathcal{D}'(-1/2, 1/2)$.

Proposition 38.2.1 (Fundamental sequences)**.** Assume $f \in L^1(\mathbb{R}^n)$ and $\int_{\mathbb{R}^n} f = 1$; for $\varepsilon > 0$ define $f_\varepsilon(x) = \varepsilon^{-n} f(x/\varepsilon)$. Then $f_\varepsilon \to \delta$ as $\varepsilon \to 0$.

Proof. Let $\varphi \in \mathcal{D}$. Then,

$$\langle f_\varepsilon, \varphi \rangle = \varepsilon^{-n} \int_{\mathbb{R}^n} f(x/\varepsilon)\varphi(x)dx = \int_{\mathbb{R}^n} f(x)\varphi(x\varepsilon)dx \to \varphi(0)$$

by dominated convergence. □

38.3 Support of a distribution

General distributions are not functions and we cannot generally speak of the value of a distribution at a point. However, the restriction of a distribution to an open set is a meaningful notion.

Definition 38.3.1. $F \in \mathcal{D}'$ is zero on the open set $\mathcal{O}$ if F restricted to $\mathcal{D}(\mathcal{O})$ is zero. Similarly, if $F, G \in \mathcal{D}'$ we say that F and G agree on $\mathcal{O}$ if $F - G = 0$ on $\mathcal{O}$.

Check that this notion coincides with usual equality of functions (a.e.) if F and G are functions.

Proposition 38.3.2. Let $\mathcal{O}_\alpha$ be open sets with $\cup_\alpha \mathcal{O}_\alpha = \mathcal{O}$. If $F \in \mathcal{D}'(\mathcal{O})$ and $F = 0$ on each $\mathcal{O}_\alpha$, then $F = 0$ on $\mathcal{O}$.

Proof. Let $\varphi \in \mathcal{D}(\mathcal{O})$. Since $\mathrm{supp}(\varphi)$ is a compact set, there exist $m \in \mathbb{N}$ and $\mathcal{O}_1, ..., \mathcal{O}_m$ s.t. $\varphi \prec \cup_1^m \mathcal{O}_j$. Let $\psi_j, j = 1, ..., m$ be a smooth partition of unity on $\mathrm{supp}(\varphi)$ with $\psi_j \prec \mathcal{O}_j$. Then $\langle F, \varphi \rangle = \sum_j \langle F, \psi_j \varphi \rangle = 0$, by assumption. $\square$

Definition 38.3.3. For $F \in \mathcal{D}'$, there is a maximal open set $\mathcal{O}$ in $\mathbb{R}^n$ (possibly empty) on which F is zero. Then, the **support** of F is $\mathbb{R}^n \setminus \mathcal{O}$.

Example 38.3.4. *The delta function at x_0 has $\{x_0\}$ as a support.*

Definition 38.3.5. Let T be a linear continuous operator on $\mathcal{D}$. T has a transpose if there is a linear continuous operator $T^\times$ on $\mathcal{D}$ s.t.

$$\langle T^\times \psi, \varphi \rangle =: \langle \psi, T\varphi \rangle \tag{240}$$

As an example, the transpose of ∂^α is $(-1)^{|\alpha|}\partial^\alpha$. Note that the transpose is uniquely defined by $(T^\times)^\times = T$ and (240). Check that the transposes below exist and satisfy the rules in 1. and 2.

1. $(aT + bS)^\times = aT^\times + bS^\times$.

2. $(TS)^\times = S^\times T^\times$.

38.4 Extension of operators from functions to distributions

Proposition 38.4.1. Assume T is linear and continuous from $\mathcal{D}$ to $\mathcal{D}$. Define $T^\times$ by

$$\langle T^\times F, \varphi \rangle = \langle F, T\varphi \rangle$$

Then $T^\times$ is linear and continuous on $\mathcal{D}'$.

Definition 38.4.2. We define T on $\mathcal{D}'$ by $T = T^\times$, $T^\times$ as above.

Proof. For linearity:

$$\langle T^{\times}(aF_1+bF_2),\varphi\rangle = \langle aF_1+bF_2, T\varphi\rangle = a\langle F_1, T\varphi\rangle + b\langle F_2, T\varphi\rangle = a\langle T^{\times}F_1,\varphi\rangle + b\langle T^{\times}F_2,\varphi\rangle$$

Continuity: By the definition of the topology on $\mathcal{D}'$ (and since for every $\varphi\in\mathcal{D}$ we have $T\varphi\in\mathcal{D}$) if $\{F_\alpha\}_{\alpha\in A}$ is s.t. $F_\alpha\to F$, then $\lim_\alpha\langle T^{\times}F_\alpha,\varphi\rangle = \lim_\alpha\langle F_\alpha, T\varphi\rangle = \langle F, T\varphi\rangle = \langle T^{\times}F,\varphi\rangle$. □

Examples. 1. (Differentiation) Let F be any L^1_{loc} function. Then, F has derivatives of all orders in the sense of distributions, since

$$\langle\partial^\alpha F,\varphi\rangle = (-1)^{|\alpha|}\int_{\mathbb{R}^n}F(x)\partial^\alpha\varphi(x)$$

is a continuous functional on $\mathcal{D}$. That is: if F is a distribution, then $\partial^\alpha F$ is a distribution, defined by $(-1)^{|\alpha|}\langle F,\partial^\alpha\varphi\rangle$.

2. (Multiplication by smooth functions) If $F\in\mathcal{D}'$ and $\psi\in\mathcal{D}$, then $F\psi\in\mathcal{D}'$, since $T^{\times}:=\varphi\mapsto\psi\varphi$ satisfies the hypotheses of the Proposition above (it acts continuously on $\mathcal{D}$), and $F\psi$ is then the distribution $\langle F\psi,\varphi\rangle := \langle F,\psi\varphi\rangle$. Note that smoothness is needed in this definition; if $F=\delta$ and, suppose, ψ is only in L^∞, $(\psi\varphi)(0)$ is undefined, in general. **The product of two distributions is not defined, in general.**

3. (Translation) Since $\int_{\mathbb{R}^n}f(x+a)g(x)dx = \int_{\mathbb{R}^n}f(x)g(x-a)dx$ if $f,g\in\mathcal{D}$, the extension to $\mathcal{D}'$ of translation is $\langle\tau_a F,\varphi\rangle = \langle F,\tau_{-a}\varphi\rangle$, and the proposition applies since $T=\tau_{-a}$ is continuous.

4. (Composition with linear transformations of $\mathbb{R}^n$.) Let M be linear and invertible on $\mathbb{R}^n$. Then $\varphi\to\varphi\circ M^{-1}$ is a continuous linear map on $\mathcal{D}$, and hence $F\circ M$ is well defined:

$$\langle F\circ M,\varphi\rangle = |\det M|^{-1}\langle F,\varphi\circ M^{-1}\rangle$$

In particular, if R is the reflection $\mathsf{R}\varphi(x)=\varphi(-x)$, we have

$$\langle\mathsf{R}F,\varphi\rangle = \langle F,\mathsf{R}\varphi\rangle$$

Exercise 94. 1. Show that the Leibniz rule of differentiation applies to $(F\varphi)^{(n)}$, when $F\in\mathcal{D}'$ and $\varphi\in\mathcal{D}$.

Let $\varphi\in\mathcal{S}$. Consider the following calculations:

$$(\varphi\delta)' = \varphi'\delta+\varphi\delta' = \boxed{\varphi'(0)\delta+\varphi(0)\delta'}$$

$$(\varphi\delta)' = (\varphi(0)\delta)' = \boxed{\varphi(0)\delta'}$$

Both cannot be right. Do a careful calculation and decide which formula is correct and what went wrong in the other.

Exercise 95. 1. Show that there is no sequence nonzero numbers $\{c_k\}_{k\in\mathbb{N}}$ such that $\sum_{k=0}^{\infty} c_k\delta^{(k)}$ converges in the sense of distributions.

2. Show that for any sequence of nonzero numbers $\{c_k\}_{k\in\mathbb{N}}$, there is a smooth function φ such that the sequence $\{c_k\varphi^{(k)}(0)\}_{k\in\mathbb{N}}$ is unbounded.

Exercise 96. Homework: 94,95 from these notes and 2,6,7,13 form Folland, Chapter 9–Not due, I will provide a solution sheet for 94,95

Definition 38.4.3. If K is a compact set,

$$\mathcal{D}(K) = \{\varphi \in \mathcal{D} : \operatorname{supp}(\varphi) \subset K\}$$

Theorem 38.4.4 (Regularity)**.** For any distribution F and compact set $K \subset \mathbb{R}^n$, there is a positive integer $N(K)$ and a positive constant $c(K)$ s.t. for all $\varphi \in \mathcal{D}(K)$,

$$|\langle F, \varphi\rangle| \leqslant c(K)|\varphi|_N, \text{ where } |\varphi|_N = \max_{|\alpha|\leqslant N} \|\partial^\alpha \varphi\|_u \tag{241}$$

In other words, any distribution, restricted to $\mathcal{D}(K)$ is in fact in the dual of $C^N(K)$ for some $N \in \mathbb{N} \cup \{0\}$.

Note that $|\cdot|_N$ is a *norm* in $C^N(K)$.

Proof. By contradiction: assume the inequality in (241) is false for all N. Then, for any n there is a $\varphi_n \in \mathcal{D}(K)$ s.t. $|\langle F, \varphi_n\rangle| = 1$ and $|\varphi_n|_n \leqslant 1/n$. However, the sequence $\{\varphi_n\}_{n\in\mathbb{N}}$ converges to zero in $\mathcal{D}(K)$ which contradicts $|\langle F, \varphi_n\rangle| = 1$ for all n.

□

Definition 38.4.5. For $K \in \mathbb{N}$ consider the space of functions on $\mathbb{T}^n$ which are absolutely continuous together with all derivatives of order up to $K-1$ and derivatives of order K in L^2. Define the norms

$$\|g\|_{2,K}^2 = \sum_{|\alpha|\leqslant K} \|\partial^\alpha g\|_2^2 \tag{242}$$

The space $\mathcal{H}_K = \mathcal{H}_K(\mathbb{T}^n)$ is defined as $\{g : \mathbb{T}^n \to \mathbb{C} |\ \|g\|_{2,K} < \infty\}$

Proposition 38.4.6. (i) $\mathcal{H}_K$ is a Hilbert space.

(ii) $\mathcal{H}_K$ is equivalently characterized by the Fourier coefficient norm

$$\|g\|_{2,K}^2 = \sum_{k\in\mathbb{Z}^n} (1+|k|^2)^K |\hat{g}_k|^2 < \infty \tag{243}$$

(iii) Smooth functions in $\mathbb{T}^n$ are dense in $\mathcal{H}_K$.

(iv) If $u \in \mathcal{H}_K$ and $K > n/2$, then u is continuous and $\|u\|_u < const.\|u\|_K$ where the constant does not depend on u. More generally, if $K > n/2 + M$, then $u \in C^M(\mathbb{T}^n)$ and $|u|_M \leqslant const\|u\|_K$, where the constant does not depend on u. Consequently, iu $K > n/2 + M$, then $\mathcal{H}_K$ is continuously embedded in $C^M(\mathbb{T}^n)$, and is a dense subset of $C^M(\mathbb{T}^n)$.

Proof. (i) Straightforward.

(ii) Parseval.

(iii) Smooth functions in $\mathbb{T}^n$ are those for which all norms above indexed by $K \in \mathbb{N}$ are finite. Density is obvious, as if we simply truncate the series in (243) at $k = k_N$, then the function corresponding to it is smooth for any k_N and in the limit $k_N \to \infty$ we recover the infinite sum.

(iv) Here the argument is similar to that in Exercise 77. We have, by Cauchy-Schwarz

$$(\sup_{\mathbb{T}^n} |u|)^2 \leqslant \left(\sum_{k\in\mathbb{Z}^n} |u_k|\right)^2 \leqslant \left(\sum_{k\in\mathbb{Z}^n} (1+|k|^2)^K |\hat{u}_k|^2\right) \sum_{k\in\mathbb{Z}^n} \frac{1}{(|k|^2+1)^K} \leqslant const.\|u\|_K \tag{244}$$

The case $K > n/2 + M$ is similar and left as an exercise. ☐

Note 38.4.7. $\mathcal{H}_K$ is an instance of a Sobolev space, and (iv) above is an instance of a Sobolev embedding theorem.

Theorem 38.4.8 (A structure theorem). Any distribution with compact support can be written in the form

$$F = \sum_{|\alpha|\leqslant L} \partial^\alpha g_\alpha \tag{245}$$

where g_α are continuous functions and L is some nonnegative integer.

Proof. Let Q be an open cube containing the support of F, and take a smooth ψ s.t. $Q \prec \psi$. Without loss of generality, we may assume that Q is centered at 0 and of side 1. By the definition of the support, we see that $F\psi = F$ ($\psi\varphi$ is zero outside Q for any $\varphi \in \mathcal{D}$). By Theorem 38.4.4, there exist $c = c(Q)$ and $M = M(Q)$ s.t for all $\varphi \in \mathcal{D}$ we have

$$|\langle F, \varphi\rangle| = |\langle F, \psi\varphi\rangle| \leqslant const.|\psi\varphi|_M \leqslant const.|\varphi|_M$$

Let $N > M + n/2$. If $u \in \mathcal{H}_N$ then $u \in C^M(\mathbb{T}^n)$ and F is a linear functional on $\mathcal{H}_N$, through $F(u) =:= F(u\psi)$. F is also continuous in $\|\cdot\|_N$ since $|F(u)| \leqslant const.|u|_M \leqslant const\|u\|_N$. Therefore, F is the inner product with an element $\overline{g} \in \mathcal{H}_N$, and, if $\varphi \in \mathcal{D} \subset \mathcal{H}_N$, we have

$$F(\varphi) = \sum_{|\alpha| \leqslant N} \int_{\mathbb{T}^n} (\partial^\alpha \varphi)(\partial^\alpha g) = (-1)^{|\alpha|} \sum_{|\alpha| \leqslant N} \int_{\mathbb{T}^n} g(\partial^{2\alpha}\varphi) = \sum_{|\alpha| \leqslant N} \langle (-1)^{|\alpha|}\partial^{2\alpha} g, \varphi\rangle$$

□

Note 38.4.9. The functions g_α can be chosen to be compactly supported. Indeed let $\mathcal{X} \in \mathcal{D}$ be s.t $\mathcal{X} = 1$ on the support of F. Then $F = F\mathcal{X}$ and for any $\varphi \in \mathcal{D}$ we have

$$\begin{aligned}\langle F, \varphi\rangle = \langle F, \mathcal{X}\varphi\rangle = \langle \partial^\alpha g, \mathcal{X}\varphi\rangle = \langle g, \partial^\alpha(\mathcal{X}, \varphi)\rangle \\ = \sum_{\beta_1+\beta_2=\alpha} \langle g, c_{\beta_1\beta_2}\partial^{\beta_1}\mathcal{X}\partial^{\beta_2}\varphi\rangle = \sum_{\beta_1+\beta_2=\alpha} \langle c_{\beta_1\beta_2} g\partial^{\beta_1}\mathcal{X}, \partial^{\beta_2}\varphi\rangle \end{aligned} \tag{246}$$

and thus

$$F = \sum_{\beta_1+\beta_2=\alpha} c_{\beta_1\beta_2}\partial^{\beta_2}(g\partial^{\beta_1}\mathcal{X}) = \sum_{|\gamma| \leqslant |\alpha|} \partial^\gamma g_\gamma$$

where $\operatorname{supp} g_\gamma \subset \operatorname{supp} \mathcal{X}$.

Corollary 38.4.10. For any K, $\mathcal{D}$ is embedded densely in $\mathcal{D}'(K)$.

Proof. If $F \in \mathcal{D}'(K)$, then $F = \partial^\beta g$ for some continuous g, by the previous theorem. Let $\{\psi_n\}_{n\in\mathbb{N}}$ be a set of functions in $\mathcal{D}$ converging to g. In the topology of $\mathcal{D}'$, $\partial^\beta \psi_n \to \partial^\beta g$. □

Convolution with elements of $\mathcal{D}$. This is defined, generalizing the convolution in $\mathcal{D}$ by

$$\langle (F * \psi), \varphi\rangle = \langle F, \varphi * \mathsf{R}\psi\rangle$$

For example,

$$\langle (\delta * \psi), \varphi\rangle = \langle \delta, \varphi * \mathsf{R}\psi\rangle = (\varphi * \mathsf{R}\psi)(0) = \int_{\mathbb{R}^n} \varphi(s)\psi(-(0-s))ds = \langle \psi, \varphi\rangle \Rightarrow \delta * \psi = \psi$$

and hence the delta function is the unit for convolution. An alternative formula is obtained as follows. If φ and F are in $\mathcal{D}$ then

$$(F * \varphi)(x) = \int_{\mathbb{R}^n} F(s)\varphi(x-s)ds = \langle F, \tau_x \mathsf{R}\varphi\rangle \tag{247}$$

and simple estimates show that the operation is continuous in $\mathcal{D}$; thus the extension of convolution to $\mathcal{D}'$ is given by (247). The two definitions coincide, by continuity and density of $\mathcal{D}$ in $\mathcal{D}'$.

Proposition 38.4.11 (Smoothing of distributions by convolution). For any $F \in \mathcal{D}'$, $F * \varphi := \langle F, \tau_x \mathsf{R}\varphi\rangle$ is C^∞ and $\partial^\alpha F * \varphi = \langle F, \partial^\alpha \tau_x \mathsf{R}\varphi\rangle = \langle F, \tau_x \partial^\alpha \mathsf{R}\varphi\rangle = (\partial^\alpha F) * \varphi = F * \partial^\alpha \varphi$.

Proof. Note first that the continuity of φ implies that $\lim_{\varepsilon\to 0}(\tau_\varepsilon \mathsf{R}\varphi - \mathsf{R}\varphi) = 0$ in the topology of $\mathcal{D}$. Thus

$$\langle F, \tau_{x+\varepsilon}\mathsf{R}\varphi\rangle \to \langle F, \mathsf{R}\tau_x\varphi\rangle \text{ as } \varepsilon \to 0$$

and thus the (usual) function $g(x) = \langle F, \tau_x \varphi\rangle$ is continuous. Next (take first $n = 1$), we see that $\varepsilon^{-1}(\tau_\varepsilon \mathsf{R}\varphi - \mathsf{R}\varphi) \to \mathsf{R}\varphi'$ in the topology of $\mathcal{D}$, and thus

$$\varepsilon^{-1}(\langle F, \tau_{x+\varepsilon}\mathsf{R}\varphi\rangle - \langle F, \tau_x \mathsf{R}\varphi\rangle) \to \langle F, \tau_x \mathsf{R}\varphi'\rangle = \langle F', \tau_x \mathsf{R}\varphi\rangle > \ \text{ as } \varepsilon \to 0$$

and g defined above is differentiable. Inductively, it is infinitely differentiable. Since proving differentiability involves on variable at a time, the result follows. □

38.5 The Hadamard finite part

Distributions can be used to regularize certain divergent integrals, as first anticipated by Hadamard in the theory of hyperbolic PDEs. I adapt this example from [4]. The integral we want to regularize is

$$\int_0^\infty \varphi(x) x^{-3/2} dx$$

Let $f(x) = 0$ for $x < 0$ and $f(x) = -2x^{-1/2}$ for $x > 0$. Then

$$\begin{aligned}\langle f', \varphi\rangle &= -\langle f, \varphi'\rangle = 2\int_0^\infty \frac{\varphi'(x)}{x^{1/2}} \\ &= 2\lim_{\varepsilon\to 0}\int_\varepsilon^\infty \frac{\varphi'(x)}{x^{1/2}} = \lim_{\varepsilon\to 0}\left(\int_\varepsilon^\infty \frac{\varphi(x)}{x^{3/2}} - 2\varepsilon^{-1/2}\varphi(0)\right) = \int_0^\infty \frac{\varphi(x)-\varphi(0)}{x^{3/2}} dx \end{aligned} \tag{248}$$

The last expression on the first line and the last two on the second line are convenient ways to present this regularization, and clearly they are bounded functionals on the C^1 functions with compact support.

38.6 Green's function

This is a very important method to solve inhomogenneous PDEs (or ODEs). Let L_x be a differential or partial differential operator in some domain with specified boundary conditions. Suppose we solve the non-homogeneous problem $L_x G(x,y) = \delta(x-y)$ (here, we take some licence in the notation, and we agree that $x \in \mathbb{R}^n$ is the variable of the equation $Lg = f$ and $y \in \mathbb{R}^n$ is a parameter). Then,

$$L_x\langle G(x,y), f(y)\rangle = \langle L_x G(x,y), f(y)\rangle = \langle \delta(x-y), f(y)\rangle = f(x)$$

and thus, the solution of the non-homogeneous equation is obtained from a universal kernel for the given equation, the Green function $G(x,y)$ by

$$h(x) = \int G(x,y)f(y)dy \Rightarrow L_x h = f$$

39 The dual of $C^\infty(\mathcal{O})$

The topology on $C^\infty(\mathcal{O})$ is based on uniform convergence on compact sets. Take an increasing sequence of precompact open subsets of $\mathcal{O}$, $\{\mathcal{O}_j\}_{j\in\mathbb{N}}$ with closures $\{K_j\}_{j\in\mathbb{N}}$ whose union is $\mathcal{O}$ and introduce the seminorms

$$\|f\|_{[j,\alpha]} = \sup_{x\in K_j} |\partial^\alpha f(x)| \tag{249}$$

These seminorms define a Fréchet space structure on $C^\infty(\mathcal{O})$.

Proposition 39.0.1. $\mathcal{D}(\mathcal{O})$ is dense in $C^\infty(\mathcal{O})$.

Proof. Indeed, take a sequence of smooth functions ψ_j s.t. $K_j \prec \psi_j \prec \mathcal{O}_{j+1}$. It is clear that $\lim_{j\to\infty} \psi_j = 1$ in the seminorms (249), hence, for any $\varphi \in C^\infty(\mathcal{O})$, $\lim_{j\to\infty} \psi_j\varphi = \varphi$ in these same seminorms. □

Definition 39.0.2. Let $\mathcal{E}'(\mathcal{O})$ be the set of distributions compactly supported in $\mathcal{O}$.

Proposition 39.0.3. Any $F \in \mathcal{E}'(\mathcal{O})$ extends uniquely to a linear continuous functional on $C^\infty(\mathcal{O})$, and conversely, the restriction of a linear continuous functional on $C^\infty(\mathcal{O})$ to $\mathcal{D}(\mathcal{O})$ is in $\mathcal{E}'(\mathcal{O})$.

Note 39.0.4. In the sense above, the dual of $C^\infty(\mathcal{O})$ is $\mathcal{E}'(\mathcal{O})$.

Proof. Let $F \in \mathcal{E}'(\mathcal{O})$. Since $\operatorname{supp} F \subset \cup_j \mathcal{O}_j$, $\exists m \in \mathbb{N}$ s.t. $\operatorname{supp} F \subset \mathcal{O}_{m-1}$. If ψ is s.t. $K_{m-1} \prec \psi \prec \mathcal{O}_m$ then $F\psi = F$, and, by the regularity theorem there exist $C > 0$ and $N \in \mathbb{N}$ s.t. $\forall \varphi \in \mathcal{D}$

$$|\langle F,\varphi\rangle| = |\langle F\psi,\varphi\rangle| = |\langle F,\psi\varphi\rangle| \leqslant C \sum_{|\alpha|\leqslant N} \|\varphi\|_{[m,\alpha]}$$

By continuity, F extends uniquely to $\varphi \in C^\infty(\mathcal{O})$ by $\langle F,g\rangle = \langle F,\psi g\rangle$ with ψ as above.

Conversely, by the same argument as in the regularity theorem, for any continuous functional G on $C^\infty(\mathcal{O})$ there exist $N, m \in \mathbb{N}$ s.t. $\forall \varphi \in C^\infty(\mathcal{O})$ we have

$$|\langle G, \varphi\rangle| \leqslant const \sum_{|\alpha|\leqslant N} \|\varphi_{[m,\alpha]}\| \leqslant const \sum_{|\alpha|\leqslant N} \|\partial^\alpha \varphi\|$$

In particular, G is compactly supported in $\mathcal{O}_m$. Therefore G is a continuous linear functional on $\mathcal{D}(\mathcal{O}) \subset C^\infty(\mathcal{O})$, and thus $G \in \mathcal{D}'(\mathcal{O})$. $\square$

39.0.1 Convolution of distributions

Let $F \in \mathcal{D}'$ and $G \in \mathcal{E}'$. Then, the natural definition of convolution is

$$\langle F * G, \varphi\rangle = \langle F, \mathrm{R}G * \psi\rangle$$

Note 39.0.5. *For this to make sense, we need* $\mathrm{R}G * \psi \in \mathcal{D}$*! Check that this is indeed the case.*

It can be shown that $F * G = G * F$ in a number of ways, e.g. Exercises 20,21 in Folland, or by density!

40 The Fourier transform

We note that $\mathcal{D}$ is not preserved by the Fourier transform. Indeed, the Fourier transform of a compactly supported function (say in $\mathbb{R}$),

$$\hat{\varphi}(\xi) = \int_{\mathbb{R}} e^{-2\pi i \xi x} \varphi(x) dx$$

is entire, and if it vanishes on any open interval, it must be identically zero. Then we need to **enlarge** $\mathcal{D}$. A space containing $\mathcal{D}$ which is invariant under $\mathcal{F}$ is $\mathcal{S}$.

Recall the topolgy of $\mathcal{S}$, and that $C_c^\infty(\mathbb{R}^n) = \mathcal{D}$ is dense in $\mathcal{S}$ in the topology of $\mathcal{S}$.

Definition 40.0.1. $\mathcal{S}'$, the dual of $\mathcal{S}$, is the space of **tempered distributions**.

Functions in $\mathcal{S}$ are required to decay faster than polynomial. The dual objects should have a corresponding growth rate limit.

Examples 40.0.2.

1. Let f be in $L^1_{loc}(\mathbb{R}^n)$ and assume that, for some N, $(1+|x|)^{-N}|f(x)|$ is bounded in $\mathbb{R}^n$. Then $\int_{-\infty}^{\infty} f(x)\varphi(x)dx$ is a continuous functional on $\mathcal{S}$ (check!).

2. $e^{ax} \in \mathcal{S}(\mathbb{R})'$ **iff** $\Re a = 0$. Indeed, if $\Re a = 0$, this follows from the previous example. Otherwise, by symmetry, we can reduce to the case $\Re a > 0$; let φ be compactly supported with $\int \varphi = 1$ and let $\psi_j = \varphi(x-j)e^{-ax}$. We see that $\psi_j \to 0$ as $j \to \infty$ in $\mathcal{S}$, while $\int e^{-ax}\psi_j = 1$ for all j.

Definition 40.0.3. A C^∞ function ψ is **slowly increasing** if, together with its derivatives it does not grow faster than polynomially. More precisely, for any α $\exists N(\alpha) \in \mathbb{N}$ s.t.

$$\left\| \frac{\partial^\alpha \psi}{(1+|x|)^{N(\alpha)}} \right\|_\infty < \infty \tag{250}$$

Proposition 40.0.4. (i) If $F \in \mathcal{S}'$ and ψ is slowly increasing, then $F\psi \in \mathcal{S}'$.
(ii) If $F \in \mathcal{S}'$ and $\psi \in \mathcal{S}$, then $F * \psi$ is slowly increasing, and, for $\varphi \in \mathcal{S}$, we have

$$\langle F, \varphi * \mathsf{R}\psi \rangle = \int_{\mathbb{R}^n} \varphi(x)(F*\psi)(x)dx \tag{251}$$

Proof. (i) $\langle F\psi, \varphi \rangle := \langle F, \psi\varphi \rangle$ is an element of $\mathcal{S}$ since $\psi\varphi$ is in $\mathcal{S}$, as it is easy to check.

(ii) We have already proved that $F * \varphi \in C^\infty$. As in the proof of the regularity theorem, for a given F to be in $\mathcal{S}'$, F must be bounded with respect to a finite number of seminorms that define the Fréchet space $\mathcal{S}$, that is, $\exists m, N \in \mathbb{N}$ and $C > 0$ s.t. $\forall \varphi \in \mathcal{S}$

$$|\langle F, \varphi \rangle| \leqslant C \max\{\|\varphi\|_{m,\alpha} : |\alpha| \leqslant N\}$$

Note also that for any $x, y \in \mathbb{R}^n$, $1 + |x| \leqslant 1 + |x-y| + |y| \leqslant (1+|x-y|)(1+|y|)$. Since $\partial^\beta F * \varphi = F * \partial^\beta \varphi = \langle F, \tau_x \mathsf{R} \partial^\beta \varphi \rangle$ we have

$$|\partial^\beta F * \varphi|(x) \leqslant \max_{|\alpha| \leqslant N} \sup_s (1+|s|)^m |\partial^{\alpha+\beta}\varphi(x-s)| \leqslant (1+|x|)^m \max_{|\alpha| \leqslant N} \sup_s (1+|x-s|)^m |\partial^{\alpha+\beta}\varphi(x-s)|$$
$$\leqslant C(1+|x|)^m \max_{|\alpha| \leqslant N+|\beta|} \|\varphi\|_{m,\alpha} \tag{252}$$

Since $\mathcal{D}$ is dense in $\mathcal{S}$, its embedding in $\mathcal{D}'$ is dense in $\mathcal{D}' \supset \mathcal{S}'$ we can check that $\mathcal{D}$ is dense in $\mathcal{S}'$. (251) is obvious if $F \in \mathcal{D}$, and the rest follows by continuity and dominated convergence. $\square$

We note that for $f, g \in \mathcal{S}$ we have

$$\langle \hat{f}, g \rangle = \iint f(x) g(y) e^{-2\pi i x y} dx dy = \langle f, \hat{g} \rangle \tag{253}$$

the definition of the Fourier transform of a distribution should be: for $F \in \mathcal{S}'$ and $g \in \mathcal{S}$,

$$\langle \hat{F}, g \rangle := \langle F, \hat{g} \rangle \tag{254}$$

It follows by duality that the basic properties of the Fourier transform that we established for functions in $\mathcal{S}$ hold for functions in $\mathcal{S}'$. Check also the following Fourier transforms:

$$\mathcal{F}\delta(x - x_0) = e^{-2\pi i x_0 k} \Rightarrow \mathcal{F}e^{2\pi i k_0 x} = \delta(k - k_0) \tag{255}$$

Note that the last equality can be interpreted as a generalized orthonormality relation of $e^{2\pi i k_1 x}$ and $e^{2\pi i k_2 x}$.

Definition 40.0.5. The Cauchy principal value distribution is defined, for $\varphi \in \mathcal{D}$ as well as for $\varphi \in \mathcal{S}$ by

$$\left[\mathrm{PV}\left(\frac{1}{x}\right)\right](u) = \lim_{\varepsilon \to 0^+} \int_{\mathbb{R}\setminus[-\varepsilon;\varepsilon]} \frac{u(x)}{x}\,\mathrm{d}x = \int_0^{+\infty} \frac{u(x) - u(-x)}{x}$$

Exercise 97. Show that the Cauchy principal value functional is indeed continuous, both on $\mathcal{D}$ and $\mathcal{S}$.

Exercise 98. Show that:

1. $$\mathcal{F}\left[\mathrm{PV}\left(\frac{1}{x}\right)\right] = -\pi i \,\mathrm{sgn}\,(k)$$

   (One way is by approximation by functions in $\mathcal{D}$: show first that $\mathcal{F}\left(\mathrm{sgn}(k)e^{-\varepsilon|k|}\right) = \dfrac{4i\pi x}{4\pi^2 x^2 + \varepsilon^2}$.)

2. $$\mathcal{F}\chi_{[0,\infty)} = \frac{1}{2}\delta(k) + \frac{1}{2\pi i}\mathrm{PV}\left(\frac{1}{k}\right)$$

3. $$\mathcal{F}\left(\delta(x)^{(n)}\right) = (-2\pi i)^n k^n$$

   and that the Fourier transform of linear combinations of the delta function and its derivatives are precisely the polynomials (and vice-versa).

4. Let $\{a_n\}_{n\in\mathbb{Z}}$ be an ℓ^1 sequence. Then

   $$\mathcal{F}\left(\sum_{n\in\mathbb{Z}} a_n e^{2\pi i n x}\right) = \sum_{n\in\mathbb{Z}} a_n \delta(k-n)$$

   which can be seen as an extension to distributions of the Poisson summation formula.

Proposition 40.0.6. Let $F \in \mathcal{E}'$. Then $\hat{F} = \langle F, e^{-2\pi i x\cdot\xi}\rangle$. $\hat{F}$ is an entire function of slow growth.

Proof. We use the decomposition in Note 38.4.9. Clearly, it is enough to prove the result for one distribution of the form $\partial^\gamma g$ where g is continuous and compactly supported. Let $\{\varphi_j\}_{j\in\mathbb{N}} \subset \mathcal{D}$ be a sequence compactly supported in some K converging to g in $\mathcal{D}'$. We then have

$$\widehat{\partial^\gamma \varphi_j} = \int_K \partial^\gamma \varphi_j(x) e^{-2\pi i\xi\cdot x} dx = \langle \partial^\gamma \varphi_j, e^{-2\pi i\xi\cdot x}\rangle = \langle \varphi_j, \partial^\gamma e^{-2\pi i\xi\cdot x}\rangle$$
$$\underset{j\to\infty}{\to} \langle g, \partial^\gamma e^{-2\pi i\xi\cdot x}\rangle = (-2\pi i\xi)^\gamma \int_K g(x))e^{-2\pi i\xi\cdot x}dx \quad (256)$$

and the rest is straightforward. □

41 Sobolev Spaces

The Fourier transform has the important feature of transforming smoothness properties into decay ones (and vice-versa). Furthermore, the Fourier transform is a unitary operator between L^2 spaces. In many applications (PDEs notably) it is convenient to bring together these features: we can introduce L^2 spaces whose norms enforce a given degree of smoothness. We have already noted that the norm

$$\|f\| = \sum_{|\alpha|\leqslant k} \|\partial^\alpha f\|_2^2 \quad (257)$$

comes from an inner product $(\cdot,\cdot)$, and then the space of functions

$$\{f : \|f\| < \infty\} \quad (258)$$

is a Hilbert space, the **Sobolev space** H_k. Taking the Fourier transform of H_k, we obtain the following dual (Fourier) norm

$$\|\hat f\|^2 = \sum_{|\alpha|\leqslant k} \langle \xi^\alpha \hat f, \xi^\alpha \hat f\rangle_2 = \sum_{|\alpha|\leqslant k} \langle (|\xi|^2)^\alpha \hat f, \hat f\rangle_2 \leqslant const\|(1+|\xi|^2)^{k/2}\hat f\|^2$$

where const only depends on k. Noting that $(1+|z|)^m \leqslant const.(1+|z|^m)$, we see that the norm above is *equivalent* to

$$\|\hat f\| = \|(1+|\xi|^2)^{k/2}\hat f\|_2$$

In Fourier space we can immediately generalize the norms from $k\in\mathbb{N}$ to any $s\in\mathbb{R}$, which can be interpreted as a norm weighted by $(1+\Delta)^{s/2}$. In fact, we have the following map:

$$\Lambda_s f = \mathcal{F}^{-1}((1+|\xi|^2)^{s/2}\mathcal{F}f) \quad (259)$$

We are now in position to define the Sobolev space $H_s = W^{s,2}$ by

$$H_s = \{F \in \mathcal{S}' : \Lambda_s f \in L^2\} \quad (260)$$

The spaces $W^{s,p}$ generalize H_s by using L^p norms,

$$W^{k,p}(\mathbb{R}^n) := \{f : \Lambda_s f \in L^p(\mathbb{R}^n)\}$$

Note that the elements of Sobolev spaces are distributions. Nonetheless, we have the following:

Proposition 41.0.1. *If $f \in H_s$, then $\hat{f}$ and $\check{f}$ are tempered functions.*

Proof. Since $\check{f} = \mathsf{R}\hat{f}$, we only check the statements about $\hat{f}$. The fact that $\Lambda_s f$ is a function (an element of L^2, more precisely), means $(1+|\xi|^2)^{s/2}\hat{f}$, and therefore $\hat{f}$, are functions.

Now, treating f as an element of $\mathcal{S}'$, and using the fact that $\hat{f}$ is a function, we have

$$\langle f, \varphi \rangle = \langle \check{f}, \hat{\varphi} \rangle = \int \hat{\varphi} \mathsf{R} \hat{f}$$

which means that $\hat{f}$ is a tempered distribution and thus a tempered function. $\square$

The inner product that we get by polarization is clearly

$$\langle f, g \rangle_{(s)} = \int_{\mathbb{R}^n} \hat{f}(\xi)(1+|\xi|)^s \overline{\hat{g}(\xi)} d^n \xi$$

The following properties follow easily from the definition

Proposition 41.0.2.

1. *$H_0 = L^2$ with $\| \cdot \|_{(0)} = \| \cdot \|_2$.*
2. *The Fourier transform is an isomorphism between H_s and $L^2(\mathbb{R}^n, \mu)$ where $d\mu = (1+|\xi|^2)^s d\xi$.*
3. *$\mathcal{S}$ is dense in H_s for all s (this is most easily seen based on 1. above).*
4. *If $s > t$, then $\| \cdot \|_{(t)} \leqslant \| \cdot \|_{(s)}$ and H_s is dense in H_t in $\| \cdot \|_{(t)}$.*
5. *Λ_t is a unitary isomorphism between H_s and H_{s-t} for all $s, t \in \mathbb{R}$.*
6. *Since $|\xi^\alpha| \leqslant (1+|\xi|^2)^{|\alpha|/2}$, ∂^α is a bounded map between H_s and $H_{s-|\alpha|}$.*

In one dimension $\delta(x)$ is in H_s if $s < -\frac{1}{2}$, and in n dimensions if $s < -\frac{n}{2}$. We see that regularity is measured more finely in this way.

Which Sobolev spaces consist of functions? The following theorem answers this important question.

Theorem 41.0.3 (The Sobolev embedding theorem)**.** *If $s > k + n/2$, then*
(i) $H_s \subset C_0^k$.
(ii) $f \in H_s$ implies $\mathcal{F}(\partial^\alpha f) \in L^1$ and $\|\mathcal{F}(\partial^\alpha f)\|_1 \leqslant C\|f\|_{(s)}$ where C only depends on $k - s$.

Proof. We prove (ii) first as (i) follows from it and the Riemann-Lebesgue lemma. We apply Cauchy-Schwarz:

$$\frac{1}{(2\pi)^{|\alpha|}} \int |\mathcal{F}\partial^\alpha f| = \int |\xi^\alpha f| \leqslant \int (1+|\xi|^2)^{\frac{k}{2}} \hat{f} \leqslant \|(1+|\xi|^2)^s \hat{f}\|_2 \; \|\frac{1}{(1+|\xi|^2)^{s-k}}\|_2$$

$\square$

Theorem 41.0.4. *If $f \in H_{-s}$, then the functional $\varphi \mapsto \langle f, \varphi \rangle$ extends continuously to a functional on H_s with norm $\|f\|_{(-s)}$, and any element in the dual of H_s is of this form.*

(Does this mean that the Hilbert space H_{-s} "is the dual of" H_s?)

Proof. By Proposition 41.0.1 $\check{f}$ is a tempered function. Cauchy-Schwarz implies

$$|\langle f, \varphi\rangle = \int \hat{f}\hat{\varphi} \leqslant \|(1+|\xi|^2)^{-s/2}\hat{f}\| \ \ \|(1+|\xi|^2)^{s/2}\hat{\varphi}\| = \|f\|_{(-s)}\|\varphi\|_{(s)} \tag{261}$$

Conversely, we can start in Fourier space with $\hat{f}$, an element of $\mathcal{F}H_{-s}$ and let it act on an element of H_s by

$$\langle f, \varphi\rangle = \int_{\mathbb{R}^n} \hat{f}(\xi)\hat{\varphi}(\xi)d\xi \left('' = \int_{\mathbb{R}^n} f(x)\varphi(x)dx''\right) = \int_{\mathbb{R}^n} \hat{f}(\xi)(1+|\xi|^2)^{-\frac{s}{2}}\hat{\varphi}(\xi)(1+|\xi|^2)^{\frac{s}{2}}d\xi \tag{262}$$

which, again by Cauchy-Schwarz shows that f is an element of $\mathcal{S}'$ which is also in H_s.

□

42 Appendix A: inductive limits of Fréchet spaces

Let V be a topological vector space over $\mathbb{R}$ or $\mathbb{C}$.

Definition 42.0.1. *The set $A \subset V$ is*

1. *Convex if $a_1, a_2 \in A$ implies $ta_1 + (1-t)a_2 \in A$ for any $t \in [0,1]$.*
2. *Balanced if $a \in A$ implies $\lambda a \in A$ if $|\lambda| \leqslant 1$;*
3. *Bounded if for any neighborhood $\mathcal{V}$ of 0 there is a $\gamma > 0$ s.t. $\gamma\mathcal{V} \supset A$.*
4. *Absorbent or absorbing if $\{\cup_{t>0} tA\} = \mathcal{V}$. (The set A can be scaled out to absorb every point in the space.)*

Definition 42.0.2.
1. *A family of seminorms on a vector space $\mathcal{V}$ is called separating if for any $0 \neq v \in \mathcal{V}$ there is a seminorm $\|\cdot\|_\alpha$ s.t. $\|v\|_\alpha > 0$.*
2. *$\mathcal{V}$ is called locally convex if the origin has a local base of absolutely convex absorbent sets.*

Proposition 42.0.3. *The topological vector space $\mathcal{V}$ is a locally convex space* **iff** *the topology is given by a family of seminorms.*

Proof. For the "if" part, the proof is immediate; the converse requires Minkovky's functionals and the Hahn-Banach separation theorem, see [3]. □

Theorem 42.0.4. *Let $\mathcal{V}$ be a topological vector space whose topology is given by a family of seminorms. Then $\mathcal{V}$ is metrizable, and a translation-invariant metric is determined by*

$$\rho(x, 0) = \sum_{k=1}^{\infty} 2^{-k}\frac{\|x\|_k}{1+\|x\|_k} \tag{263}$$

The balls $B(0,r) := \{x : \rho(x,0) < r\}$ are balanced. If $\mathcal{V}$ is complete with respect to ρ, then it is a Fréchet space.

Proof. Largely a straightforward verification, see [3] , p. 437 and on. □

Definition 42.0.5. *Let*

$$\mathcal{V}_1 \subset \mathcal{V}_2 \subset \cdots \subset \mathcal{V}_j \cdots$$

be a sequence of Fréchet spaces. Thw inductive limit topology on $\mathcal{V}$ is the strongest locally convex topology, in which the injections $\mathcal{V}_i \to \mathcal{V}$ are continuous.

Theorem 42.0.6. *Let $\mathcal{V}$ be an inductive limit of Fréchet spaces as in Definition 42.0.5.*

1. *The open, convex, balanced neighborhoods of zero are the sets W s.t. $W \cap \mathcal{V}_j$ are open, convex, balanced neighborhoods of zero for all j, and these sets uniquely determine the topology of $\mathcal{V}$.*

2. *$A \subset \mathcal{V}$ is bounded* **iff** *A is a bounded subset of some fixed $\mathcal{V}_{n_0}$.*

3. *A sequence is Cauchy in $\mathcal{V}$* **iff** *there is some n_0 a.t. the sequence is contained $\mathcal{V}_{n_0}$ and is Cauchy there.*

4. *Let $\mathcal{X}$ be a locally convex topological vector space. The linear map $T : \mathcal{X} \to \mathcal{V}$ is continuous* **iff** *the restriction of T to $\mathcal{V}_j$ is continuous for every j.*

References

[1] Reed, M. and Simon, B., Methods of Modern Mathematical Physics, vol II, Academic Press (1980).

[2] , Convergence of sequences of distributions, Proceedings Of The American Mathematical Society, v. 27, no 3, (1971).

[3] Gerd Grubb, Distributions and Operators, Springer-Verlag New York (2009).

[4] , L. Schwartz, Théorie des Distributions, Hermann, Paris, 1966.