Introduction to modern analysis

Paul Garrett

version December 4, 2017

[This document is http://www.math.umn.edu/~garrett/m/real/notes_2017-18/modern_analysis.pdf]

1. Preview, Real Analysis 2016-17

1. What is the issue? What obstacles to overcome?
2. Limits and integrals
3. Measures: an attempt at greater generality

(Everything below admits substantial generalization beyond what is literally asserted. Determining the extent of various possible generalizations is often a task in itself, and is often tangential to the main enterprise.)

1.1 What is the issue? What obstacles to overcome?

For most purposes, up until 1800 and even afterward, *function* meant *formula*. Also, it was often assumed without comment that decent functions could be represented by *power series*.

[1.1.1] Euler and the wave equation Many people had considered the (linear) *wave equation*

$$(\Delta_x - \frac{\partial^2}{\partial t^2})\, u \;=\; 0$$

where the *spatial* variable $x \in \mathbb{R}^n$ (mostly $n = 1, 2, 3$) and *time* $t \in \mathbb{R}$, and *Laplace's operator* is

$$\Delta_x \;=\; \frac{\partial^2}{\partial x_1^2} + \ldots + \frac{\partial^2}{\partial x_n^2}$$

For one-dimensional spatial variable, the wave operator *factors*:

$$\frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial t^2} \;=\; \Big(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\Big) \circ \Big(\frac{\partial}{\partial x} - \frac{\partial}{\partial t}\Big) \;=\; \Big(\frac{\partial}{\partial x} - \frac{\partial}{\partial t}\Big) \circ \Big(\frac{\partial}{\partial x} + \frac{\partial}{\partial t}\Big)$$

Thus, apparently, any function u of the form

$$u(x,t) \;=\; f(x-t) + g(x+t)$$

is a solution. The two pieces are *incoming* and *outgoing* components of the solution.

The cognitive dissonance arises when one imagines, as apparently Euler did, that such a formula *makes sense even when f and g are not differentiable (in a classical sense).*

But there did not seem to be any natural or conceptual way to exclude problemmatical functions f, g from this formula, and this heated up the discussion of *what is a function?*

[1.1.2] A success story: convergent power series By soon after 1800, Abel and others had carefully proven that power series (real or complex) with a positive radius of convergence r really could be differentiated correctly by doing the obvious thing, namely, differentiating term by term:

$$\frac{d}{dz} \sum_{n=0}^{\infty} c_n\,(z-z_o)^n \;=\; \sum_{n=0}^{\infty} \frac{d}{dz} c_n\,(z-z_o)^n \;=\; \sum_{n=0}^{\infty} c_n\, n(z-z_o)^{n-1} \qquad \text{(still convergent in } |z-z_o| < r\text{)}$$

This completely justified what people had been doing all along.

[1.1.3] Fourier 1811-22 and the heat equation The heat equation is

$$(\Delta_x - \frac{\partial}{\partial t})\, u \;=\; 0 \qquad \text{(with } \textit{initial condition} \text{ prescribing } u(x,0)\text{)}$$

Especially in the case of one-dimensional spatial variable x confined to a finite interval such as $[0, 2\pi]$, Fourier had the inspiration to express an alleged solution as a superposition of eigenfunctions for Δ_x on $[0, 2\pi]$, namely, constants and $\sin(nx)$ and $\cos(nx)$ for $n = 1, 2, 3, \ldots$:

$$u(x,t) \;=\; c_o(t) + \sum_{n \geq 1} \Big(a_n(t) \cos(nx) + b_n(t) \sin(nx) \Big)$$

This *separated variables*, and if we imagine we can apply the heat operator *termwise*,

$$0 \;=\; (\Delta_x - \frac{\partial}{\partial t})\, u \;=\; -c_o'(t) + \sum_{n \geq 1} \Big(-n^2 a_n - a_n' \big) \cos(nx) + (-n^2 b_n - b_n') \sin(nx) \Big)$$

If we believe in *uniqueness* of such expressions in x, this gives

$$-c_o' = 0 \qquad -n^2 a_n - a_n' = 0 \qquad -n^2 b_n - b_n' = 0 \qquad (\text{for } n = 1, 2, 3, \ldots)$$

so $c_o(t)$ is a constant, and $a_n(t)$ and $b_n(t)$ are constant multiples of $e^{-n^2 t}$:

$$u(x,t) \;=\; -c_o + \sum_{n \geq 1} e^{-n^2 t} \Big(a_n \cos(nx) + b_n \sin(nx) \Big) \qquad (\text{with constants } c_o, a_n, b_n)$$

The initial condition at time t presumably determines the constants, by

$$u(x,0) \;=\; -c_o + \sum_{n \geq 1} \Big(a_n \cos(nx) + b_n \sin(nx) \Big)$$

The explicit claim that *every* function $x \to u(x,0)$ could be represented by such a *Fourier series* was appealing, since this device then gave a solution to the heat equation, and would prove uniqueness. *But what is a function?*

Soon after his initial epiphany, Fourier also found the correct formulas determining coefficients:

$$c_o \;=\; \frac{1}{2\pi} \int_0^{2\pi} u(x,t)\, dx \qquad a_n \;=\; \frac{1}{2\pi} \int_0^{2\pi} u(x,t) \cdot \cos(nx)\, dx \qquad b_n \;=\; \frac{1}{2\pi} \int_0^{2\pi} u(x,t) \cdot \sin(nx)\, dx$$

Further, under relatively mild hypotheses [1] on the smoothness-or-not of $x \to u(x,0)$, Fourier proved that the series converges *pointwise* to $u(x,0)$ and $u(x,t)$ for $t > 0$. [2]

However, Fourier made much broader claim about the range of *functions* representable by such series, revivifying the argument over *what is a function?*

More technically, there is the issue of the legitimacy of *termwise differentiation.* Indeed, functions meeting the conditions for pointwise convergence could have derivatives *not* meeting the condition, yet termwise differentiation would still make sense. For example, the periodic *sawtooth function* is

$$(\text{sawtooth}) \;=\; \sum_{n \geq 1} \frac{\sin(nx)}{n}$$

[1] For example, if a function is piecewise C^1 except for finitely-many jumps in $[0, 2\pi]$, where left and right derivatives exist, then, away from the jumps, the Fourier series converges pointwise to the function.

[2] Apparently what is often called the *Dirichlet kernel* and used to prove this pointwise convergence was in fact used by Fourier prior to Dirichlet's 1829 paper proving convergence of Fourier series.

This converges (not absolutely) to the sawtooth function's values $x - \pi$ for $0 < x < 2\pi$. The sawtooth is differentiable in $(0, 2\pi)$, but termwise differentiation gives

$$\frac{d}{dx}(\text{sawtooth}) \;=\; \sum_{n\geq 1} \sin(nx) \qquad\qquad (???)$$

For most values of $x \in (0, 2\pi)$, the summands do not go to 0. Differentiating again should give 0 for $0 < x < 2\pi$, but

$$\frac{d^2}{dx^2}(\text{sawtooth}) \;=\; \sum_{n\geq 1} n \cdot \sin(nx) \qquad\qquad (???)$$

and so on. These expressions do not converge *pointwise*, and cast reasonable doubt on the legitimacy of this approach.

However, in fact, although these infinite sums of functions do not converge pointwise, they do converge perfectly well in certain topological vector spaces of (generalized) functions, namely, the Sobolev spaces discussed below. But this development would have to wait until the 1930s and 1940s.

Another tension arose when people subsequently discovered that the Fourier series of typical continuous functions would fail to converge pointwise at infinitely many points. (For example, we will prove this via *Baire's Theorem*.)

Yet there is *Parseval's theorem*, that for f such that $\int_0^{2\pi} |f|^2 < \infty$, there is a nice relation between the this integral of f and its Fourier coefficients:

$$\int_0^{2\pi} |f|^2 \;=\; |c_o|^2 + \sum_{n\geq 1} |a_n|^2 + |b_n|^2$$

This implies that, even if the partial sums of the Fourier series of such a function do not converge to the function *pointwise*, they *do* converge to the function in the *mean-square* or L^2 metric

$$d_{L^2}(f,g) \;=\; |f-g|_{L^2[0,2\pi]} \;=\; \Big(\int_0^{2\pi} |f(x)-g(x)|^2\,dx\Big)^{1/2}$$

That this is a metric on $C^o[0,2\pi]$ uses the integral form of the *Cauchy-Schwarz-Bunyakowsky inequality*, due to Bunyakowsky. But convergence of a sequence of continuous functions in this L^2-metric does *not* imply pointwise convergence, since the pointwise evaluation maps maps $f \to f(x_o)$ are *not continuous*: there are sequences $\{f_n\}$ of continuous functions that are Cauchy sequences in the L^2 topology, but so that $\{f_n(x_o)\}$ is *not* a Cauchy sequence of real or complex numbers.

Also, simple pointwise convergence does not imply L^2 convergence in general, and simple pointwise convergence does not imply convergence in the sup-norm topology on $C^o[0,2\pi]$, either.

The seemingly natural notion of pointwise convergence is not all that we had hoped it would be. As a corollary, there are problems if we exclusive think of *functions* as producing pointwise values: there are L^2 limits of Cauchy sequences of continuous functions that lack well-defined pointwise limits.

[1.1.4] Sturm and Liouville 1830s eigenfunction expansions On the heels of Fourier's ideas, Sturm and Liouville had a similar idea about expressing functions f on $[0,2\pi]$ in terms of *eigenfunctions* for differential operators of the form

$$Lu \;=\; -(pu')' + q \qquad\qquad (\text{with } p(x) > 0 \text{ on } [0,2\pi], \text{ real-valued } q)$$

with various possible *boundary conditions* at 0 and 2π. For example, we might require $u(0) = u(2\pi)$ and $u'(0) = u'(2\pi)$ (the periodic case), or we might require $u(0) = 0 = u(2\pi)$ (the Dirichlet condition).

That is, they argued first toward the conclusion that the *eigenfunction equation*

$$Lu = \lambda \cdot u \qquad \text{(with the boundary conditions)}$$

should have a list $0 \leq \lambda_1 \leq \lambda_2 \leq \lambda_3, \ldots$ of non-negative real numbers such that there would be non-trivial (real-valued) solutions u_n to the equation $Lu = \lambda_n \cdot u$ and meeting the boundary conditions. Then, when normalized so that $\int_0^2 \pi |u_n(x)|^2 \, dx = 1$, an *arbitrary* (real-valued) function f on $[0, 2\pi]$ should be expressible as

$$f(x) = \sum_{n \geq 1} \Big(\int_0^{2\pi} f(t) \cdot u_n(t) \, dx \Big) \cdot u_n(x)$$

Their difficulty at the time was that various notions of convergence were still unsettled, and the *linear algebra* needed to express things this clearly had not yet been invented. Heuristics were not made into proofs (of some assertions) until Steklov 1898-9, and Bocher 1895-6.

By now we know that for $\int_0^{2\pi} |f(x)|^2 \, dx < +\infty$, that expansion *does* converge in the L^2-metric, and we think of the coefficients as being given by *inner products* of f with the exponentials in the space $L^2[0, 2\pi]$:

$$\langle f, g \rangle_{L^2[0,2\pi]} = \frac{1}{2\pi} \int_0^{2\pi} f(x) \cdot \overline{g(x)} \, dx$$

(with complex conjugation for complex-valued functions). One characterization of the whole space $L^2[0, 2\pi]$ is as the *completion* of $C^o[0, 2\pi]$ with respect to the metric obtained from the L^2-norm.

But pointwise convergence is potentially confusing: with the Dirichlet condition, the eigenfunctions are $u_n(x) = \sin(nx/2)/\sqrt{2\pi}$. But there are many reasonable functions meeting the condition $\int_0^{2\pi} |f(x)|^2 \, dx < +\infty$ that do not vanish at 0 and 2π, for example, the constant function 1. So, in an L^2 (mean-square) sense,

$$1 = \frac{1}{2\pi} \sum_{n \geq 1} \Big(\int_0^{2\pi} 1 \cdot \sin(nt/2) \, dx \Big) \cdot \sin(nx/2) = \frac{1}{2\pi} \sum_{n=1,3,5,\ldots} \frac{\sin(nx/2)}{2n}$$

but this certainly cannot converge pointwise as the endpoints. It *does* provably converge pointwise in the interior.

[1.1.5] Green's functions 1828

Another approach to solving linear differential equations $Lu = f$ on $\mathbb{R}^n$, not only in one dimension like the Sturm-Liouville equations, was conceived by Green about 1828, and has similar applications to partial differential equations like the heat equation and wave equation.

One way to talk about the method is to refer to a *fundamental solution* or *Green's function* $G(x, y)$ for the given differential operator L, characterized by solving the differential equation $Lu = f$ by

$$u(x) = \int_{\mathbb{R}^n} G(x, y) \, f(y) \, dy$$

Green's original idea and subsequent applications arose in physically meaningful situations, problems, so the sensibility of solutions to problems obtained by such ideas could be confirmed to some degree by direct observation of physical phenomena.

But, from a mathematical viewpoint, *why* should any such thing exist?

If we already believe from Sturm-Liouville that there is an *orthonormal basis* $\{u_n\}$ for $L^2[0, 2\pi]$ consisting of eigenfunctions u_n for L, in an equation $Lu = f$ expand both u and f in terms of eigenfunctions, computing coefficients by inner products, as in Fourier's case:

$$L\Big(\sum_n \langle u, u_n \rangle \cdot u_n \Big) = Lu = f = \sum_n \langle f, u_n \rangle \cdot u_n$$

Of course, we assume that we can apply L termwise (!), so this gives

$$\sum_n \langle f, u_n \rangle \cdot u_n \;=\; \sum_n \langle u, u_n \rangle \cdot Lu_n \;=\; \sum_n \langle u, u_n \rangle \cdot \lambda_n \cdot u_n$$

Presumably these expansions are unique, so $\langle u, u_n \rangle \cdot \lambda_n = \langle f, u_n \rangle \cdot u_n$ for all n. That is, apparently

$$u(x) \;=\; \sum_n \frac{\langle f, u_n \rangle}{\lambda_n} \cdot u_n(x) \;=\; \Big\langle f(y),\, \sum_n \frac{u_n(y)}{\lambda_n} \cdot u_n(x) \Big\rangle$$

That is, apparently,

$$G(x,y) \;=\; \sum_n \frac{1}{\lambda_n} u_n(x) \cdot u_n(y)$$

For that matter, a more *scandalous* description of $G(x,y)$, but which makes considerable sense in a physical context where Dirac's δ idealizes a *point-mass*, is

$$L_x\, G(x,y) \;=\; \delta(x-y) \qquad\qquad \text{(with a Dirac } \delta\text{-function)}$$

which would have been essentially impossible to make mathematically rigorous until well into the 20th century. Nevertheless, if we apply L_x to the eigenfunction expansion, apparently

$$\delta(x-y) \;=\; L_x G(x,y) \;=\; L_x \sum_n \frac{1}{\lambda_n} u_n(x) \cdot u_n(y) \;=\; \sum_n \frac{1}{\lambda_n} L_x u_n(x) \cdot u_n(y) \;=\; \sum_n u_n(x) \cdot u_n(y)$$

If true, this would be very convenient. But *pointwise* it cannot make sense.

Still, in one dimension, reasonable second-order differential operators L on finite intervals have Green's functions obtained in a straightforward way from two linearly independent solutions, based on the idea that $\frac{d^2}{dx^2}|x| = 2\delta$, as follows. Find one solution u to $Lu = 0$ with $u(0) = 0$, and a solution v to $Lv = 0$ with $v(2\pi) = 0$, and splice them together so that their *values* match at a point $x \in [0, 2\pi]$, but their *derivatives* differ suitably, creating a *corner*.

For example, for the equation $-u'' = f$ on $[0, 2\pi]$ with boundary conditions $u(0) = 0 = u(2\pi)$, solutions of $u'' = 0$ are just linear functions, the solution vanishing at the left edge is x, and the solution vanishing at the right edge is $2\pi - x$. To find the linear combination agreeing at y and derivatives differing by 1, solve for coefficients a, b in

$$\begin{cases} a \cdot y & = & b \cdot (2\pi - y) \\[2ex] a + 1 & = & -b \end{cases}$$

and obtain

$$G(x,y) \;=\; \begin{cases} \Big(\frac{y}{2\pi} - 1\Big) \cdot x & \text{(for } 0 \le x \le y) \\[2ex] -\frac{y}{2\pi} \cdot (2\pi - x) & \text{(for } y \le x \le 2\pi) \end{cases}$$

In two dimensions or higher, the geometry is more complicated. Nevertheless, it has been appreciated for a long time, in one way or another, that

$$\begin{cases} \Delta \log|x| & = & \text{(constant)} \cdot \delta \quad (\text{in } \mathbb{R}^2) \\[2ex] \Delta \frac{1}{|x|^{n-2}} & = & \text{(constant)} \cdot \delta \quad (\text{in } \mathbb{R}^n,\ n \ge 3) \end{cases}$$

with elementary constants. An elementary computation certainly shows that the Laplacian annihilates those functions *away from* 0, but we are lacking a *persuasive* or *conceptual* argument that *at* 0 we get δ.

Looking at that one-dimensional situation further, apparently

$$\begin{cases} \left(\frac{y}{2\pi} - 1\right) \cdot x & (\text{for } 0 \le x \le y) \\ -\frac{y}{2\pi} \cdot (2\pi - x) & (\text{for } y \le x \le 2\pi) \end{cases} \;=\; G(x,y) \;=\; \frac{1}{2\pi} \sum_{n \ge 1} \sin\frac{nx}{2} \cdot \frac{\sin\frac{ny}{2}}{-n^2/4}$$

and applying Δ gives

$$\delta(x-y) \;=\; \frac{1}{2\pi} \sum_{n\ge 1} \sin\frac{nx}{2} \cdot \frac{\sin ny}{2} \qquad (???)$$

But the other eigenfunction expansion similarly apparently gives

$$\delta(x-y) \;=\; \frac{1}{2\pi} \cdot \Big(1 + \sum_{n\ge 1} \sin nx + \cos nx\Big) \qquad (???)$$

and the two expressions are not easily comparable. The heuristic is attractive and useful, but a more refined viewpoint is obviously needed to avoid seeming paradoxes.

[1.1.6] Heaviside 1880s Also used δ as an idealization of an *impulse* in electrical circuits and similar, with great success. Despite his successes in predicting observable phenomena, mathematicians at the time were apparently disdainful of the mathematics itself, which was unrigorizable at the time.

[1.1.7] Dirac 1928-9 In nascent quantum physics, Dirac not only used point-masses and point-charges, but geometrically more complicated generalized functions, and did subtle computations that correctly predicted physical phenomena. In contrast to Hilbert's and Schmidt's conversion of *differential* operators to *integral* operators with better continuity properties, Dirac directly manipulated differential operators without apparent concern for their not being everywhere defined or continuous.

Partly in reaction to Dirac's physics success, careful rigorization of unbounded/discontinuous operators, modelling differential operators, was accomplished by Stone and von Neumann by 1930, and more simply in important special situations by Friedrichs in 1934. In 1934 and thereafter, Sobolev created a basic framework adequate to deal with certain generalized functions.

[1.1.8] Kronig-Penney 1931, Bethe-Peierls 1935 ... but Dirac's success prompted even-more-audacious mathematics: idealizing δ as a very-short-range-acting *potential*, to model nuclear foces (as opposed to electromagnetism or gravity), physicists considered *singular potential* equations

$$(-\Delta + \delta)\, u \;=\; f$$

The intention is fairly clear, but it is not obvious how to be sure one is manipulating such a thing correctly from a mathematical viewpoint. Still, testable physical conclusions were correctly reached, and Nobel prizes were won.

[1.1.9] Fourier transforms, Plancherel 1910, Wiener 1933, Bochner 1932 In 1910, Plancherel proved the basic fact that Fourier transform on reasonable functions f with $\int_{\mathbb{R}} |f| < \infty$ gave an $L^2(\mathbb{R})$-isometry. That is, with

$$\widehat{f}(\xi) \;=\; \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-i\xi x}\, f(x)\, dx$$

the L^2 norm of $\widehat{f}$ is equal to that of f. This allows the Fourier transform to be *extended by continuity* to give a map of $L^2(\mathbb{R})$ to itself, although the literal integral does not converge well for general functions in L^2 but not in L^1. Part of the lesson is that maps given by integrals *cannot* be taken literally, but, happily, *need not* be taken literally.

Fourier inversion is that f can be reconstructed from its Fourier transform:

$$f(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x}\; \widehat{f}(\xi)\; d\xi$$

There is a non-trivial issue of the sense of convergence of the integral! A naive but reasonable attempt to prove Fourier inversion is the obvious interchange of the order of integration:

$$\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x}\; \widehat{f}(\xi)\; d\xi \;=\; f(x) \;=\; \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\xi x}\Big(\int_{\mathbb{R}} e^{-i\xi u} f(v)\; dv\Big)\; d\xi \;=\; \frac{1}{2\pi} \int_{\mathbb{R}} f(v)\Big(\int_{\mathbb{R}} e^{i\xi(x-v)}\; d\xi\Big)\; dv$$

If we could believe various heuristics that the inner integral is $2\pi\delta(x-v)$, we'd be done. Indeed, this can be justified later, but Fourier inversion is prior.

These examples and others raise basic questions:

What kind of functions can be integrated?

What kind of infinite-sum expansions of functions are legitimate?

What kind of convergence do infinite-sum expansions have?

1.2 Limits and integrals

Archetypical issue: integrating on a finite interval $[a, b]$ on the real line,

$$\text{when is } \lim_n \int_a^b f_n \;=\; \int_a^b \lim_n f_n \;\; ???$$

And *limit* in what sense? And what kind of *functions* can be integrated?

As a positive example, if the functions f_n are *continuous*, and if the limit is *uniformly pointwise*, meaning that for every $\varepsilon > 0$ there is n_o such that for every $m, n \geq n_o$ and for every $x \in [a, b]$, the limit $\lim_n f_n$ is itself a continuous function, and, indeed, the integral of the limit is the limit of the integrals. For continuous functions on finite intervals, the Riemann integral behaves well with *uniformly* pointwise limits, and gives us a description of *integral* that allows us to prove the previous assertion.

However, even when the functions f_n are very nice, if the limit is merely *pointwise*, but not *uniformly* so, then the limit function need not be continuous, and the limit of the integrals need not be the integral of the limit.

Also, a pointwise limit of continuous functions need not be continuous! But we can salvage a little, even though the issue will not go away:

[1.2.1] Theorem: *(Dini)* For a pointwise *monotone* (increasing or decreasing) sequence of real-valued continuous functions f_n on a finite interval $[a, b]$, if the limit is continuous, then the limit is *uniform* pointwise.

The classic example of failure of the integral of the (pointwise) limit to be the limit of the integrals is the sequence of *tent functions* f_n just to the right of 0: $f_n(x) = 0$ on $[\frac{2}{n}, 1]$, and on $[0, \frac{2}{n}]$ is a triangular tent of height n, to make the area under it be 1:

$$f_n(x) \;=\; \begin{cases} 0 & (\text{for } x \leq 0) \\ n^2 \cdot x & (\text{for } 0 \leq x \leq \frac{1}{n}) \\ n - n^2 \cdot (x - \frac{1}{n}) & (\text{for } \frac{1}{n} \leq x \leq \frac{2}{n}) \\ 0 & (\text{for } x \geq \frac{2}{n}) \end{cases}$$

For every individual $x \in \mathbb{R}$, the pointwise limit is $\lim_n f_n(x) = 0$, but the integral of the zero function is not 1.

On the other hand, for $g \in C^o(\mathbb{R})$, while the pointwise limit of these tent functions f_n is 0 everywhere,

$$\lim_n \int_{\mathbb{R}} f_n(x)\, g(x)\, dx \;=\; g(0) \;=\; \delta(g)$$

That is, in a very tangible sense, $f_n \longrightarrow \delta$, where δ is the Dirac delta function at 0, which we imagine produces $g(0)$ when integrated against a continuous function g.

Measure theory can accommodate the Dirac delta, because it is a kind of *measure*. But its *derivative*[3] is not a measure. Nevertheless, using tent-functions, we can make a sequence of continuous functions h_n that go to 0 everywhere pointwise, but so that

$$\lim_n \int_{\mathbb{R}} h_n(x)\, g(x)\, dx \;=\; g'(0)$$

for differentiable g with continuous derivative g'. Specifically, let h_n be a downward-pointing tent to the left together with an upward-pointing tent to the right, with each tent having area $n/2$ (rather than 1):

$$h_n(x) \;=\; \begin{cases} 0 & (\text{for } x \le -\frac{1}{n}) \\ -2n^3 \cdot (x + \frac{1}{n}) & (\text{for } -\frac{1}{n} \le x \le -\frac{1}{2n}) \\ 2n^3 \cdot x & (\text{for } -\frac{1}{2n} \le x \le \frac{1}{2n}) \\ n^2 - 2n^3 \cdot (x - \frac{1}{2n}) & (\text{for } \frac{1}{2n} \le x \le \frac{1}{n}) \\ 0 & (\text{for } x \ge \frac{1}{n}) \end{cases}$$

Among other things, such examples are further evidence for the unfortunate limitations of the notion of *pointwise* values and limits.

1.3 Measures: one attempt at greater generality

Motivated by 19th century difficulties related to Fourier series, eigenfunction expansions, and related matters, soon after 1900 several people developed ideas to deal with *pointwise limits* of sequences of somewhat larger classes of functions.

The *Borel subsets* of $\mathbb{R}$ is the smallest collection of subsets of $\mathbb{R}$ closed under taking *countable unions*, under *countable intersections*, under *complements*, and containing all open and closed subsets of $\mathbb{R}$. This is also called the Borel σ-*algebra* in $\mathbb{R}$.

There is traditional terminology for certain simple types of Borel sets. For example a G_δ is a *countable intersection of open sets*, while an F_σ is a *countable union of closed sets*. The notation can be iterated: a $G_{\delta\sigma}$ is a countable union of countable intersections of opens, and so on. We will not need this.

A *Borel measure* μ is a way of assigning (often *positive*) real numbers (measures) to Borel sets, in a fashion that is *countably additive* for disjoint unions:

$$\mu(E_1 \cup E_2 \cup E_3 \cup \ldots) \;=\; \mu(E_1) + \mu(E_2) + \mu(E_3) + \ldots \qquad (\text{for } \textit{disjoint} \text{ Borel sets } E_1, E_2, E_3, \ldots)$$

[3] And we do not mean *derivative* of Dirac delta in the measure-theory context of *Radon-Nikodym* derivative, either.

A prototype is *Lebesgue (outer) measure* of a Borel set $E \subset \mathbb{R}$, described by

$$\mu(E) \;=\; \inf\{\sum_{n=1}^{\infty} |b_n - a_n| \;:\; E \subset \bigcup_{n=1}^{\infty} (a_n, b_n)\}$$

That is, it is the *inf* of the sums of lengths of the intervals in a countable cover of E by open intervals. For example, any countable set has (Lebesgue) measure 0.

We can consider larger classes of real-valued or complex-valued functions than just continuous ones, for example, various classes of *measurable* functions. The simplest useful choice is: A real-valued or complex-valued function f on $\mathbb{R}$ is *Borel-measureable* when the inverse image $f^{-1}(U)$ is a Borel set for every open set U in the target space.

It is occasionally useful to also allow the target space for functions to be the *two-point compactification* $Y = \{-\infty\} \cup \mathbb{R} \cup +\infty$ of the real line, with neighborhood basis $-\infty \cup (-\infty, a)$ at $-\infty$ and $(a, +\infty) \cup \{+\infty\}$ at $+\infty$ when we need to allow functions to blow up in some fashion.

A positive indicator:

[1.3.1] Theorem: Every pointwise limit of Borel-measurable functions f_n is Borel-measurable.

Verifying that we have not inadvertently needlessly included functions wildly unrelated to continuous functions:

[1.3.2] Theorem: *(Lusin)* Continuous functions approximate Borel-measurable functions well: given Borel-measurable real-valued or complex-valued f on $\mathbb{R}$, for every $\varepsilon > 0$ and for every Borel subset $\Omega \subset \mathbb{R}$ of finite Lebesgue measure, there is a relative closed $E \subset \Omega$ such that $\mu(\Omega - E) < \varepsilon$, and $f|_E$ is *continuous*.

Not much better can be done than Lusin's theorem says: for example, continuous approximations to the Heaviside step function

$$H(x) \;=\; \begin{cases} 0 & \text{for } x < 0 \\ \\ 1 & \text{for } x \geq 0 \end{cases}$$

have to go from 0 to 1 *somewhere*, by the Intermediate Value Theorem, so will be in $(\frac{1}{4}, \frac{3}{4})$ on an open set of strictly positive measure.

[1.3.3] Remark: It turns out that the everyday use of measure theory, measurable functions, and so on, does *not* proceed by way of Lusin's theorem or similar direct connections with continuous functions, but, rather, by direct interaction with the more general ideas.

A sequence $\{f_n\}$ of Borel-measurable functions on $\mathbb{R}$ converges (pointwise) *almost everywhere* when there is a Borel set $N \subset \mathbb{R}$ of measure 0 such that $\{f_n\}$ converges pointwise on $\mathbb{R} - N$.

[1.3.4] Theorem: *(Severini, Egoroff)* Pointwise convergence of sequences of Borel-measurable functions is approximately *uniform* convergence: given a almost-everywhere pointwise-convergent sequence $\{f_n\}$ of Borel-measurable functions on $\mathbb{R}$, for every $\varepsilon > 0$ and for every Borel subset $\Omega \subset \mathbb{R}$ of finite Lebesgue measure, there is a Borel subset $E \subset \Omega$ such that $\{f_n\}$ converges *uniformly* pointwise on E.

[1.3.5] Remark: Again, despite the connection that the Severini-Egoroff theorem makes between pointwise and *uniform* pointwise convergence, this idea turns out *not* to be the way to understand convergence of measurable functions. Instead, the game becomes ascertaining additional conditions that guarantee convergence of integrals, as just below.

With such notion of *measure*, there is a corresponding *integrability* and *integral*, due to Lebesgue. It amounts to replacing the literal rectangles used in Riemann integration by more general rectangles, with bases not just intervals, but measurable sets, as follows.

The *characteristic function* or *indicator function* ch_E or χ_E of a measurable subset $E \subset \mathbb{R}$ is 1 on E and 0 off. A *simple function* is a finite, positive-coefficiented, linear combination of characteristic functions of bounded measurable sets, that is, is of the form

$$\text{(simple function) } s \;=\; \sum_{i=1}^{n} c_i \cdot \mathrm{ch}_{E_i} \qquad\qquad \text{(with } c_i \geq 0\text{)}$$

The *integral* of s is what one would expect:

$$\int s\, d\mu \;=\; \int \Big(\sum_{i=1}^{n} c_i \cdot \mathrm{ch}_{E_i}\Big)\, d\mu \;=\; \sum_i c_i \cdot \mu(E_i)$$

Next, the measure of a *non-negative* function f is the *sup* of the integrals of all simple functions between f and 0:

$$\int f\, d\mu \;=\; \sup_{0\leq s\leq f} \int s\, d\mu \qquad\qquad \text{(sup over simple } s \text{ with } 0 \leq s(x) \leq f(x) \text{ for all } x\text{)}$$

After proving that the positive and negative parts f_+ and f_- of Borel measurable real-valued f are again Borel measurable,

$$\int f\, d\mu \;=\; \int f_+\, d\mu - \int (-f_-)\, d\mu$$

Similarly, for complex-valued f, break f into real and imaginary parts.

There are details to be checked:

[1.3.6] Theorem: Borel-measurable functions f, g taking values in $[0, +\infty]$ are *integrable*, in the sense that the previous prescription yields an assignment $f \to \int_{\mathbb{R}} f \in [0, +\infty]$ such that for positive constants a, b

$$\int_{\mathbb{R}} (af + bg) \;=\; a\int_{\mathbb{R}} f + b\int_{\mathbb{R}} g \qquad\qquad \text{(for all } a, b \geq 0\text{)}$$

For complex-valued Borel-measurable f, g, the absolute values $|f|$ and $|g|$ are Borel-measurable. Assuming $\int_{\mathbb{R}} |f| < \infty$ and $\int_{\mathbb{R}} |g| < \infty$, for any complex a, b

$$\int_{\mathbb{R}} (af + bg) \;=\; a\int_{\mathbb{R}} f + b\int_{\mathbb{R}} g$$

Now we have practical criteria for the integral of a pointwise sequence to be the limit of the integrals:

[1.3.7] Theorem: *(Lebesgue's dominated convergence)* For Borel-measurable f_n with pointwise limit f, if there is non-negative Borel-measurable real-valued g such that $|f_n(x)| \leq g(x)$ for all x, and if g is *integrable* in the sense that $\int_{\mathbb{R}} g < +\infty$, then the pointwise limit is integrable, and

$$\lim_n \int_{\mathbb{R}} f_n \;=\; \int_{\mathbb{R}} \lim_n f_n$$

[1.3.8] Theorem: *(Monotone convergence)* For measurable extended-real-valued f_n with (extended-real) pointwise limit f, if $f_n(x) \leq f_{n+1}(x)$ for all x and for all indices n, then

$$\lim_n \int_{\mathbb{R}} f_n \;=\; \int_{\mathbb{R}} \lim_n f_n$$

(although the limit may be $+\infty$).

Less decisive-appearing, but unconditional, is

[1.3.9] Theorem: *(Fatou's lemma)* For Borel-measurable f_n with values in $[0, +\infty]$, the pointwise $f(x) = \liminf_n f_n(x)$ is Borel-measurable, and

$$\int \liminf_n f_n(x)\, dx \;\le\; \liminf_n \int f_n$$

More interesting, and more useful: after figuring out how to characterize measure on product spaces,

[1.3.10] Theorem: *(Fubini-Tonelli)* For complex-valued measurable f, g, if any one of $\int_{\mathbb{R}} \int_{\mathbb{R}} |f(x,y)|\, dx\, dy$, $\int_{\mathbb{R}} \int_{\mathbb{R}} |f(x,y)|\, dy\, dx$, or $\int_{\mathbb{R}\times\mathbb{R}} |f(x,y)|\, d\mathrm{vol}$ is finite, then the all are finite, and are equal. For $[0, +\infty]$-valued functions f, we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\, dx\, dy \;=\; \int_{\mathbb{R}} \int_{\mathbb{R}} f(x,y)\, dy\, dx \;=\; \int_{\mathbb{R}\times\mathbb{R}} f(x,y)\, d\mathrm{vol}$$

although the values may be $+\infty$.

2. Review of metric spaces and point-set topology

2.1 Euclidean spaces

Let $\mathbb{R}^n$ be the usual Euclidean n-space, that is, ordered n-tuples $x = (x_1, \ldots, x_n)$ of real numbers. In addition to vector addition (termwise) and scalar multiplication, we have the usual *distance function* on $\mathbb{R}^n$, in coordinates $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, defined by

$$d(x, y) \;=\; \sqrt{(x_1 - y_1)^2 + \ldots + (x_n - y_n)^2}$$

Of course there is visible *symmetry* $d(x, y) = d(y, x)$, and *positivity*: $d(x, y) = 0$ only for $x = y$. The *triangle inequality*

$$d(x, z) \;\le\; d(x, y) + d(y, z)$$

is not trivial to prove. In the one-dimensional case, the triangle inequality is an inequality on absolute values, and can be proven case-by-case. In $\mathbb{R}^n$, it is best to use the following set-up. The usual *inner product* (or *dot-product*) on $\mathbb{R}^n$ is

$$x \cdot y \;=\; \langle x, y\rangle \;=\; \langle (x_1, \ldots, x_n), (y_1, \ldots, y_n)\rangle \;=\; x_1 y_1 + \ldots + x_n y_n$$

and $|x|^2 = \langle x, x\rangle$. *Context* distinguishes the *norm* $|x|$ of $x \in \mathbb{R}^n$ from the usual absolute value $|c|$ on real or complex numbers c. The distance is expressible as

$$d(x, y) \;=\; |x - y|$$

The inner product $\langle x, y\rangle$ is *linear* in both arguments: in the first argument

$$\langle x + x', y\rangle \;=\; \langle x, y\rangle + \langle x', y\rangle \qquad\qquad \langle cx, y\rangle \;=\; c \cdot \langle x, y\rangle \qquad\qquad (\text{for } x, x', y \in \mathbb{R}^n \text{ and scalar } c)$$

and similarly for the second argument. The triangle inequality will be a corollary of the following universally-useful inequality:

[2.1.1] Claim: *(Cauchy-Schwarz-Bunyakowsky inequality)* For $x, y \in \mathbb{R}^n$,

$$|\langle x, y\rangle| \;\le\; |x| \cdot |y|$$

Assuming that neither x nor y is 0, *strict* inequality holds unless x and y are scalar multiples of each other.

Proof: If $|y| = 0$, the assertions are trivially true. Thus, take $y \neq 0$. With real t, consider the quadratic polynomial function

$$f(t) \;=\; |x - ty|^2 \;=\; |x|^2 - 2t\langle x, y\rangle + t^2|y|^2$$

Certainly $f(t) \ge 0$ for all $t \in \mathbb{R}$, since $|x - ty| \ge 0$. Its minimum occurs where $f'(t) = 0$, namely, where $-2\langle x, y\rangle + 2t|y|^2 = 0$. This is where $t = \langle x, y\rangle/|y|^2$. Thus,

$$0 \;\le\; (\text{minimum}) \;\le\; f(\langle x, y\rangle/|y|^2) \;=\; |x|^2 - 2\frac{\langle x, y\rangle}{|y|^2}\langle x, y\rangle + \Big(\frac{\langle x, y\rangle}{|y|^2}\Big)^2 \cdot |y|^2 \;=\; |x|^2 - \Big(\frac{\langle x, y\rangle}{|y|^2}\Big)^2 \cdot |y|^2$$

Multiplying out by $|y|^2$,

$$0 \;\le\; |x|^2 \cdot |y|^2 - \langle x, y\rangle^2$$

which gives the inequality. Further, for the inequality to be an *equality*, it must be that $|x - ty| = 0$, so x is a multiple of y. ///

[2.1.2] Remark: We did not use properties of $\mathbb{R}^n$, only of the inner product!

[2.1.3] Corollary: *(Triangle inequality)* For $x, y, z \in \mathbb{R}^n$,

$$|x + y| \;\le\; |x| + |y|$$

Therefore,

$$d(x, z) \;=\; |x - z| \;=\; |(x - y) - (z - y)| \;\le\; |x - y| + |z - y| \;=\; d(x, y) + d(y, z)$$

Proof: With the Cauchy-Schwarz-Bunyakowsky inequality in hand, this is a direct computation:

$$|x + y|^2 \;=\; \langle x + y, x + y\rangle \;=\; \langle x, x\rangle + \langle x, y\rangle + \langle y, x\rangle + \langle y, y\rangle \;=\; |x|^2 + 2\langle x, y\rangle + |y|^2 \;\le\; |x|^2 + 2|\langle x, y|\rangle + |y|^2$$

$$\le\; |x|^2 + 2|x| \cdot |y| + |y|^2 \;=\; (|x| + |y|)^2$$

Taking positive square roots gives the result. ///

The *open ball* B of radius $r > 0$ centered at a point y is

$$B \;=\; \{x \in \mathbb{R}^n \;:\; d(x, y) < r\}$$

The *closed ball* $\overline{B}$ of radius $r > 0$ centered at a point y is

$$\overline{B} \;=\; \{x \in \mathbb{R}^n \;:\; d(x, y) \le r\}$$

Obviously in many regards the two are barely different from each other. However, the fact that the *closed* ball includes its *boundary* (in both an intuitive an technical sense as below) the sphere

$$S^{n-1} \;=\; \{x \in \mathbb{R}^n \;:\; d(x, y) = r\}$$

while the *open* ball does *not*. A different distinction is what we'll exploit most directly:

[2.1.4] Corollary: For any point x in an open ball B in $\mathbb{R}^n$, for sufficiently small radius $\varepsilon > 0$ the open ball of radius ε centered at x is contained in B.

Proof: This is essentially the triangle inequality. Let B be the open ball of radius r centered at y. Then $x \in B$ if and only if $|x - y| < r$. Thus, we can take $\varepsilon > 0$ such that $|x - y| + \varepsilon < r$. For $|z - x| < \varepsilon$, by the triangle inequality

$$|z - y| \;\le\; |z - x| + |x - y| \;<\; \varepsilon + |x - y| \;<\; r$$

That is, the open ball of radius ε at x is inside B. ///

An *open set* in $\mathbb{R}^n$ is any set with the property observed in the latter corollary, namely a set U in $\mathbb{R}^n$ is *open* if for every x in U there is an open ball centered at x contained in U.

This definition allows us to rewrite the epsilon-delta definition of *continuity* in a useful form:

[2.1.5] Claim: A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is *continuous* if and only if the inverse image

$$f^{-1}(U) = \{x \in \mathbb{R}^m \;:\; f(x) \in U\}$$

of every *open* set U in $\mathbb{R}^n$ is *open* in $\mathbb{R}^m$. *(We prove this below for general metric spaces.)* ///

Some properties of open sets in $\mathbb{R}^n$ that will be abstracted:

[2.1.6] Claim: The union of an *arbitrary* set of open subsets of $\mathbb{R}^n$ is open. The intersection of a *finite* set of open subsets of $\mathbb{R}^n$ is open.

Proof: A point $x \in \mathbb{R}^n$ is in the union U of an arbitrary set $\{U_\alpha : \alpha \in A\}$ of open subsets of $\mathbb{R}^n$ exactly when there is some U_α so that $x \in U_\alpha$. Then a small-enough open ball B centered at x is inside U_α, so $B \subset U_\alpha \subset U$.

For x in the intersection $I = U_1 \cap \ldots \cap U_m$ of a finite number of opens, let $\varepsilon_j > 0$ such that the open ε_j-ball at x is contained in U_j. Let ε be the minimum of the ε_j. The minimum of a *finite* set of (strictly) positive real numbers is still (strictly) positive, so $\varepsilon > 0$, and the ε-ball at x is contained inside every ε_j-ball at x, so is contained in the intersection. ///

One of many equivalent ways to say that a set E in $\mathbb{R}^n$ is *bounded* is that it is contained in some (sufficiently large) *ball.* [4] At various technical points in advanced calculus, we find ourselves caring about *closed and bounded* sets, and perhaps proving the *Heine-Borel property* or *Bolzano-Weierstraß property* [5]

[2.1.7] Theorem: A set E in $\mathbb{R}^n$ is closed and bounded *if and only if* every sequence of points in E has a *convergent subsequence.* ///

2.2 Metric spaces

By design, the previous discussion of Euclidean spaces made minimal use of particular features of Euclidean space. This allows *abstraction* of some relevant features in a manner that uses our intuition about Euclidean spaces to suggest things about less familiar spaces. The process of abstraction has several different stopping places, and this section looks at one of the first.

We can abstract the *distance function* on $\mathbb{R}^n$ usefully, as follows. For a set X be a set, a non-negative-real-valued function

$$d : X \times X \longrightarrow \mathbb{R}$$

is a *distance function* if it satisfies the conditions

$$\begin{cases} d(x,y) \geq 0 & \text{(with equality only for } x = y\text{)} \quad \textit{(positivity)} \\ d(x,y) = d(y,x) & \textit{(symmetry)} \\ d(x,z) \leq d(x,y) + d(y,z) & \textit{(triangle inequality)} \end{cases}$$

for all points $x, y, z \in X$. Such a distance function is also called a *metric.* The set X with the metric d is a *metric space.*

In analogy with the situation for $\mathbb{R}$ and $\mathbb{R}^n$, a sequence $\{x_n\}$ in a metric space X is *convergent* to $x \in X$ when, for every $\varepsilon > 0$, there is n_o such that, for all $n \geq n_o$, $|x_n - x| < \varepsilon$. Likewise, a sequence $\{x_n\}$ in X is a *Cauchy* sequence when, for all $\varepsilon > 0$, there is n_o such that for all $m, n \geq n_o$, $|x_m - x_n| < \varepsilon$. A metric space is *complete* if every Cauchy sequence is convergent.

---

[4] A few moments' thought show that it does not matter where the ball is centered, nor whether the ball is closed or open.

[5] This property is not at all trivial to prove, especially from an elementary viewpoint.

The following standard lemma makes a bit of intuition explicit:

[2.2.1] Lemma: Let $\{x_i\}$ be a Cauchy sequence in a metric space X, d converging to x in X. Given $\varepsilon > 0$, let N be sufficiently large such $d(x_i, x_j) < \varepsilon$ for $i, j \geq N$. Then $d(x_i, x) \leq \varepsilon$ for $i \geq N$.

Proof: Let $\delta > 0$ and take $j \geq N$ also large enough such that $d(x_j, x) < \delta$. Then for $i \geq N$ by the triangle inequality

$$d(x_i, x) \;\leq\; d(x_i, x_j) + d(x_j, x) \;<\; \varepsilon + \delta$$

Since this holds for every $\delta > 0$ we have the result. ///

[2.2.2] Example: Variants of the usual Euclidean metric on $\mathbb{R}^n$ also make sense:

$$d_1(x, y) \;=\; |x_1 - y_1| + \ldots + |x_n - y_n| \qquad\qquad d_\infty(x, y) \;=\; \max_i |x_i - y_i|$$

In fact, the triangle inequality for these metrics are easy to prove, needing just the triangle inequality for the absolute value on $\mathbb{R}$. Later, we will see [6] that

$$d_p(x, y) \;=\; \Big(|x_1 - y_1|^p + \ldots + |x_n - y_n|^p\Big)^{1/p} \qquad\qquad (\text{for } 1 \leq p < \infty)$$

also gives a metric.

[2.2.3] Example: A *discrete set* or *discrete metric space* X is one in which (roughly) no two distinct points are close to each other. That is, for each $x \in X$ there should be a bound $\delta_x > 0$ such that $d(x, y) \geq \delta_x$ for all $y \neq x$ in X. For example, the set $\mathbb{Z}$ of integers, with the natural distance

$$d(x, y) \;=\; |x - y| \qquad\qquad (\text{with usual absolute value})$$

has the property that $|x - y| \geq 1$ for distinct integers. Every discrete metric space is complete.

[2.2.4] Example: *Any* set X can be made into a *discrete* metric space by defining

$$d(x, y) \;=\; \begin{cases} 1 & (\text{for } x \neq y) \\ 0 & (\text{for } x = y) \end{cases}$$

This is obviously positive and symmetric, and satisfies the triangle inequality condition for silly reasons. Little is learned from this example except that it is possible to do such things.

[2.2.5] Example: The collection $C^o[a, b]$ of continuous functions [7]) on an interval $[a, b]$ on the real line can be given the metric

$$d(f, g) \;=\; \sup_{x \in [a,b]} |f(x) - g(x)|$$

Positivity and symmetry are easy, and the triangle inequality is not hard, either. This metric space is *complete*, because a Cauchy sequence is a *uniformly pointwise convergent* sequence of continuous functions.

[2.2.6] Example: The collection $C^o(\mathbb{R})$ of continuous functions [8] on the *whole* real line does *not* have an obvious candidate for a metric, since the *sup* metric of the previous example may give infinite values. Yet there is the metric

$$d(f, g) \;=\; \sum_{n=1}^{\infty} 2^{-n} \frac{\sup_{|x| \leq n} |f(x) - g(x)|}{1 + \sup_{|x| \leq n} |f(x) - g(x)|}$$

---

[6] The triangle inequality for such metrics is an instance of the *Hölder inequality.*

[7] Throughout discussion of these examples, it doesn't matter much whether we think of real-valued functions or complex-valued functions.

[8] Real-valued or complex-valued, for example.

This metric space is complete, for similar reasons as $C^o[a,b]$.

[2.2.7] Example: A sort of infinite-dimensional analogue of the standard metric on $\mathbb{R}^n$ is the space ℓ^2, the collection of all sequences $\alpha = (\alpha_1, \alpha_2, \ldots)$ of complex numbers such that $\sum_{n\geq 1} |\alpha_n|^2 < +\infty$. The metric is

$$d(\alpha, \beta) \;=\; \sqrt{\sum_{n\geq 1} |\alpha_n - \beta_n|^2}$$

In fact, ℓ^2 is a vector space, being closed under addition and under scalar multiplication, with inner product

$$\langle \alpha, \beta \rangle \;=\; \sum_{n\geq 1} \alpha_n \cdot \overline{\beta}_n$$

The associated *norm* is $|\alpha| = \langle \alpha, \alpha \rangle^{\frac{1}{2}}$, and $d(\alpha, \beta) = |\alpha - \beta|$. The Cauchy-Schwarz-Bunyakowsky holds for ℓ^2, by the same proof as given earlier, and proves the triangle inequality. This metric space is complete.

[2.2.8] Example: For $1 \leq p < \infty$, the sequence spaces ℓ^p is

$$\ell^p \;=\; \{x = (x_1, x_2, \ldots) \;:\; \sum_{i=1}^{\infty} |x_i|^p < \infty\}$$

with metric

$$d_p(x, y) \;=\; \Big(\sum_{i=1}^{\infty} |x_i - y_i|^p\Big)^{1/p}$$

Proof of the triangle inequality needs *Hölder's inequality*. These metric spaces are complete. Unlike the case of varying metrics on $\mathbb{R}^n$, the underlying sets ℓ^p are not the same. For example, ℓ^2 is strictly larger than ℓ^1.

[2.2.9] Example: Even before having a modern notion of measure and integral, a partial analogue of ℓ^2 can be formulated: on $C^o[a,b]$, form an inner product

$$\langle f, g \rangle \;=\; \int_a^b f(x)\, \overline{g(x)}\, dx$$

It is easy to check that this does give a hermitian inner product. The L^2 norm is $|f|_{L^2} = \langle f, f \rangle^{\frac{1}{2}}$, and the distance function is $d(f,g) = |f - g|$. The basic properties of a metric are immediate, except that the triangle inequality needs the integral form of the Cauchy-Schwarz-Bunyakowsky inequality, whose proof is the same as that given earlier. This metric space is *not* complete, because there are sequences of continuous functions that are Cauchy in this L^2 metric (but not in the $C^o[a,b]$ metric) and do not converge to a continuous function. For example, we can make piecewise-linear continuous functions approaching the discontinuous function that is 0 on $[a, \frac{a+b}{2}]$ and 1 on $[\frac{a+b}{2}, b]$, by

$$f_n(x) \;=\; \begin{cases} 0 & (\text{for } a \leq x \leq \frac{a+b}{2} - \frac{1}{n}) \\ \frac{n}{2} \cdot (x - (\frac{a+b}{2} - \frac{1}{n})) & (\text{for } \frac{a+b}{2} - \frac{1}{n} \leq x \leq \frac{a+b}{2} + \frac{1}{n}) \\ 1 & (\text{for } \frac{a+b}{2} + \frac{1}{n} \leq x \leq b) \end{cases}$$

(Draw a picture.) The pointwise limit is 0 to the left of the midpoint, and 1 to the right. Despite the fact that the pointwise limit does not exist at the midpoint,

$$d_2(f_i, f_j)^2 \;\leq\; \int_{\frac{a+b}{2}-\frac{1}{n}}^{\frac{a+b}{2}+\frac{1}{n}} 1\, dx \;\leq\; \frac{2}{n} \qquad (\text{for } i, j \geq n)$$

which goes to 0 as $n \to \infty$. That is, $\{f_n\}$ is Cauchy in the L^2 metric, but does not converge to a continuous function.

2.3 Completions of metric spaces

Again, a metric space X, d is *complete* when every Cauchy sequence is convergent. Completeness is a convenient feature, because then we can take limits without leaving the space. As in the example of $C^o[a,b]$ with the L^2 inner product, we might want to imbed a non-complete metric space in a complete one in an optimal and universal way.

A traditional notion of the *completion* of a metric space X is a *construction* of a complete metric space $\widetilde{X}$ with a distance-preserving injection $j : X \to \widetilde{X}$ so that $j(X)$ is *dense* in $\widetilde{X}$, in the sense that every point of $\widetilde{X}$ is the limit of a Cauchy sequence in $j(X)$.

The intention is that every Cauchy sequence has a limit, so we should (somehow!) *adjoin* points as needed for these limits. However, different Cauchy sequences may happen to have the same limit.

Thus, we want an equivalence relation on Cauchy sequences that says they should have the same limit, even without knowing the limit exists or having somehow constructed or adjoined the limit point.

Define an equivalence relation $\sim$ on the set C of Cauchy sequences in X, by

$$\{x_s\} \sim \{y_t\} \iff \lim_s d(x_s, y_s) \;=\; 0$$

Attempt to define a metric on the set $C/\sim$ of equivalence classes by

$$d(\{x_s\}, \{y_t\}) \;=\; \lim_s \, d(x_s, y_s)$$

We must verify that this is well-defined on the quotient $C/\sim$ and gives a metric. We have an injection $j : X \to C/\sim$ by

$$x \to \{x, x, x, \ldots\} \bmod \;\sim$$

[2.3.1] Claim: $j : X \to C/\sim$ is a completion of X.

Proof: Grant for the moment that the distance function on $\widetilde{X} = C/\sim$ is well-defined, and is complete, and show that it has the property of a completion of X. To this end, let $f : X \to Y$ be a *uniformly* continuous map to a complete metric space Y.

Given $z \in \widetilde{X}$, choose a Cauchy sequence x_k in X with $j(x_k)$ converging to z, and *try* to define $F : \widetilde{X} \to Y$ in the natural way, by

$$F(z) \;=\; \lim_k \; f(x_k)$$

Since f is *uniformly* continuous, $f(x_k)$ is Cauchy in Y, and by completeness of Y has a limit, so $F(z)$ *exists*, at least if well-defined.

For well-definedness of $F(z)$, for x_k and x'_k two Cauchy sequences whose images $j(x_k)$ and $j(x'_k)$ approach z, since j is an isometry eventually x_k is close to x'_k, so $f(x_k)$ is eventually close to $f(x'_k)$ in Y, showing $F(z)$ is well-defined.

We saw that every element of $\widetilde{X}$ is a limit of a Cauchy sequence $j(x_k)$ for x_k in X, and *any* continuous $\widetilde{X} \to Y$ respects limits, so F is the only possible extension of f to $\widetilde{X}$.

The obvious argument will show that F is continuous. Namely, let $z, z' \in \widetilde{X}$, with Cauchy sequences x_t and x'_t approaching z and z'. Given $\varepsilon > 0$, by uniform continuity of F, there is N large enough such that

$d_Y(F(j(x_r)), F(j(x_s))) < \varepsilon$ and $d_Y(F(j(x'_r)), F(j(x'_s))) < \varepsilon$ for $r, s \geq N$. From the lemma above (!), for such r even in the limit the strict inequalities are at worst non-strict inequalities:

$$d_Y(f(x_r), F(z)) \;\leq\; \varepsilon \qquad \text{and} \qquad d_Y(f(x'_r), F(z')) \;\leq\; \varepsilon$$

By the triangle inequality, since $f : X \to Y$ is continuous, we can increase r to have $d_X(x_r, x'_r)$ small enough so that $d_Y(f(x_r), f(x'_r)) < \varepsilon$, and then

$$d_Y(F(z), F(z')) \;\leq\; d_Y(F(z), f(x_r)) + d_Y(f(x_r), f(x'_r)) + d_Y(f(x'_r), F(z')) \;\leq\; \varepsilon + \varepsilon + \varepsilon$$

Since $j : X \to \widetilde{X}$ is an isometry,

$$d_X(x_r, x'_r) \;=\; d_{\widetilde{X}}(j(x_r), j(x'_r)) \;\leq\; d_{\widetilde{X}}(j(x_r), z) + d_{\widetilde{X}}(z, z') + d_{\widetilde{X}}(j(x'_r), z')$$

so

$$d_X(x_r, x'_r) \;\leq\; d_{\widetilde{X}}(z, z') \;+\; 2\varepsilon$$

Thus,

$$d_Y(F(z), F(z')) \;\leq\; d_{\widetilde{X}}(z, z') + 4\varepsilon \qquad\qquad (\text{for all } \varepsilon > 0)$$

Thus, F is continuous. Granting that $\widetilde{X} = C/\!\sim$ is complete, etc., it is a completion of X.

It remains to prove that the apparent metric on $\widetilde{X}$ truly is a metric, and that $\widetilde{X}$ is complete.

First, the limit in attempted definition

$$d(\{x_s\}, \{y_t\}) \;=\; \lim_s d(x_s, y_s)$$

does exist: given $\varepsilon > 0$, take N large enough so that $d(x_i, x_j) < \varepsilon$ and $d(y_i, y_j) < \varepsilon$ for $i, j \geq N$. By the triangle inequality,

$$d(x_i, y_i) \;\leq\; d(x_i, x_N) + d(x_N, y_N) + d(y_N, y_i) \;<\; \varepsilon + d(x_N, y_N) + \varepsilon$$

Similarly,

$$d(x_i, y_i) \;\geq\; -d(x_i, x_N) + d(x_N, y_N) - d(y_N, y_i) \;>\; -\varepsilon + d(x_N, y_N) - \varepsilon$$

Thus, unsurprisingly,

$$\Big| d(x_i, y_i) - d(x_N, y_N) \Big| \;<\; 2\varepsilon$$

and the sequence of real numbers $d(x_i, y_i)$ is Cauchy, so convergent.

Similarly, when $\lim_i d(x_i, y_i) = 0$, then $\lim_i d(x_i, z_i) = \lim_i d(y_i, z_i)$ for any other Cauchy sequence z_i, so the distance function is *well-defined* on $C/\!\sim$.

The positivity and symmetry for the alleged metric on $C/\!\sim$ are immediate. For triangle inequality, given x_i, y_i, z_i and $\varepsilon > 0$, let N be large enough so that $d(x_i, x_j) < \varepsilon$, $d(y_i, y_j) < \varepsilon$, and $d(z_i, z_j) < \varepsilon$ for $i, j \geq N$. As just above,

$$\Big| d(\{x_s\}, \{y_s\}) - d(x_i, y_i) \Big| \;<\; 2\varepsilon$$

Thus,

$$d(\{x_s\}, \{y_s\}) \;\leq\; 2\varepsilon + d(x_N, y_N) \;\leq\; 2\varepsilon + d(x_N, z_N) + d(z_N, y_N) \;\leq\; 2\varepsilon + d(\{x_s\}, \{z_s\}) + 2\varepsilon + d(\{z_s\}, \{y_s\}) + 2\varepsilon$$

This holds for all $\varepsilon > 0$, so we have the triangle inequality.

Finally, perhaps anticlimactically, the completeness. Given Cauchy sequences $c_s = \{x_{sj}\}$ in X such that $\{c_s\}$ is Cauchy in $C/\!\sim$, for each s we will choose large-enough $j(s)$ such that the diagonal-ish sequence $y_\ell = x_{\ell, j(\ell)}$ is a Cauchy sequence in X to which $\{c_s\}$ converges.

Given $\varepsilon > 0$, take i large enough so that $d(c_s, c_t) < \varepsilon$ for all $s, t \geq i$. For each i, choose $j(i)$ large enough so that $d(x_{ij}, x_{ij'}) < \varepsilon$ for all $j, j' \geq j(i)$. Let $c = \{x_{i,j(i)} : i = 1, 2, \ldots\}$. For $s \geq i$,

$$d(c_s, c) \;=\; \lim_\ell d(x_{s\ell}, x_{\ell,j(\ell)}) \;\leq\; \sup_{\ell \geq i} d(x_{s\ell}, x_{\ell,j(\ell)}) \;\leq\; \sup_{\ell \geq i} \Big(d(x_{s\ell}, x_{s,j(\ell)}) + d(x_{s,j(\ell)}, x_{\ell,j(\ell)}) \Big) \;\leq\; 2\varepsilon$$

This holds for all $\varepsilon > 0$, so $\lim_s c_s = c$, and $C/\!\sim$ is complete. ///

Many natural metric spaces are complete without any need to complete them. The historically notable exception was $\mathbb{Q}$ itself, completed to $\mathbb{R}$. A slightly more recent example:

[2.3.2] Example: One description of the space $L^2[a, b]$ is as the completion of $C^o[a, b]$ with respect to the L^2 norm above. The more common description depends on notions of *measurable function* and *Lebesgue integral*, and presents the space as equivalence classes of functions, having somewhat ambiguous pointwise values.

2.4 Topologies of metric spaces

The notion of metric space allows a useful generalization of the notion of *continuous function* via the obvious analogue of the epsilon-delta definition:

A function or map $f : X \to Y$ from one metric space (X, d_X) to another metric space (Y, d_Y) is *continuous* at a point $x_o \in X$ when, for every $\varepsilon > 0$ there is $\delta > 0$ such that

$$d_X(x, x_o) < \delta \quad \Longrightarrow \quad d_Y(f(x), f(x_o)) < \varepsilon$$

In a metric space (X, d), the *open ball* of radius $r > 0$ centered at a point y is

$$\{x \in X \,:\, d(x, y) < r\}$$

The *closed ball* of radius $r > 0$ centered at a point y is

$$\{x \in X \,:\, d(x, y) \leq r\}$$

As in $\mathbb{R}^n$, in many regards the two are barely different from each other. However, the *closed* ball includes the *sphere*

$$\{x \in X \,:\, d(x, y) = r\}$$

while the *open* ball does *not*. A different distinction is what we'll exploit most directly:

[2.4.1] Claim: For any point x in an open ball B in X, for sufficiently small radius $\varepsilon > 0$ the open ball of radius ε centered at x is contained in B. *(As for $\mathbb{R}^n$, this follows immediately by use of the triangle inequality.* ///

An *open set* in X is any set with the property observed in this proposition. That is, a set U in X is *open* if for every x in U there is an open ball centered at x contained in U.

This definition allows us to rewrite the epsilon-delta definition of *continuity* in a form that will apply in more general topological spaces:

[2.4.2] Claim: A function $f : X \to Y$ from one metric space to another is *continuous* in the ε-δ sense if and only if the inverse image

$$f^{-1}(U) \;=\; \{x \in \mathbb{R}^m \,:\, f(x) \in U\}$$

of every *open* set U in Y is *open* in X.

Proof: On one hand, suppose f is continuous in the ε-δ sense. For U open in Y and $x \in f^{-1}(U)$, with $f(x) = y$, let $\varepsilon > 0$ be small enough so that the ε-ball at y is inside U. Take $\delta > 0$ small enough so that, by the ε-δ definition of continuity, the δ-ball B at x has image $f(B)$ inside the ε-ball at y. Then $x \in B \subset f^{-1}(U)$. This holds for every $x \in f^{-1}(U)$, so $f^{-1}(U)$ is open.

On the other hand, suppose $f^{-1}(U)$ is open for every open $U \subset Y$. Given $x \in X$ and $\varepsilon > 0$, let U be the ε-ball at $f(x)$. Since $f^{-1}(U)$ is open, there is an open ball B at x contained in $f^{-1}(U)$. Let $\delta > 0$ be the radius of B. ///

A set E in a metric space X is *closed* if and only its *complement*

$$E^c = X - E = \{x \in X : x \notin E\}$$

is *open*.

A set E in a metric space X is *bounded* when it is contained in some (sufficiently large) *ball*. This makes sense in general metric spaces, but does not have the same implications.

2.5 General topological spaces

Many of the ideas and bits of terminology for metric spaces make sense and usefully extend to more general situations. Some do not.

[2.5.1] A topology on a set X is a collection τ of subsets of X, called the *open sets*, such that X itself and the empty set ϕ are in τ, *arbitrary unions* of elements of τ are in τ, and *finite intersections* of elements of τ are in τ. A set X with an explicitly or implicitly specified topology is a *topological space*.

[2.5.2] A continuous map $f : X \to Y$ for topological spaces X, Y is a set-map so that inverse images $f^{-1}(U)$ of opens U in Y are open in X.

Uniform continuity of functions or maps has no natural formulation in general topological spaces, in effect because we have no device by which to compare the topology at varying points, unlike the case of metric spaces, where there is a common notion of distance that does allow such comparisons.

[2.5.3] Closed sets in a topological space are exactly the complements of open sets. Arbitrary intersections of closed sets are closed, and finite unions of closed sets are closed.

[2.5.4] A basis for a topology is a collection of (open) subsets so that any open set is a union of the (open) sets in the basis. In a metric space, the open balls of all possible sizes, at all points, are a natural basis.

[2.5.5] A neighborhood of a point is any set containing an open set containing the point. Often, one considers only *open* neighborhoods, to avoid irrelevant misunderstandings.

[2.5.6] A local basis at a point x in a space X is a collection of open neighborhoods of x such that every neighborhood of x *contains* a neighborhood from the collection. In a metric space, the collection of open balls at a given point with *rational radius* is a countable local basis at that point.

[2.5.7] The closure of a set E (in a topological space X), sometimes denoted $\overline{E}$, is the intersection of all closed sets containing E. It is a closed set. Equivalently, it is the set of $x \in X$ such that every neighborhood of x meets[9] E. The closure of E contains E.

[9] A set X *meets* another set Y if $X \cap Y \neq \phi$.

[2.5.8] The interior of a set E (in a topological space X) is the union of all open sets contained in it. It is open. Equivalently, it is the set of $x \in X$ such that there is a neighborhood of x inside E. The interior of E is a subset of E.

[2.5.9] The boundary of a set E (in a topological space X), often denoted ∂E, is the intersection of the closure of E and the closure of the complement of E. Equivalently, it is the set of $x \in X$ such that every neighborhood of x meets both E and the complement of E.

[2.5.10] A Hausdorff topology is one in which any two points x, y have neighborhoods $U \ni x$ and $V \ni y$ which are disjoint: $U \cap V = \phi$. This is a reasonable condition to impose on a space on which functions should live.

[2.5.11] Claim: Metric spaces are Hausdorff.

Proof: Given $x \neq y$ in a metric space, let B_1 be the open ball of radius $d(x,y)/2$, and let B_2 the open ball of radius $d(x,y)/2$ at y. For any $z \in B_1 \cap B_2$, by the triangle inequality,

$$d(x,y) \;\le\; d(x,z) + d(z,y) \;<\; \frac{d(x,y)}{2} + \frac{d(x,y)}{2} \;=\; d(x,y)$$

which is impossible. Thus, there is no z in the intersection of these two open neighborhoods of x and y. ///

[2.5.12] Claim: In Hausdorff spaces, singleton sets $\{x\}$ are closed.

Proof: Fixing x, for $y \neq x$ let U_y be an open neighborhood of y not containing x. (We do not use the open neighborhood of x not meeting U_y.) Then $E = \bigcup_{y \neq x} U_y$ is open, does not contain x, and contains every other point in the space. Thus, E is the complement of the singleton set $\{x\}$ and is open, so $\{x\}$ is closed. ///

[2.5.13] Convergence of sequences: In a topological space X, a sequence $x_1, x_2, \ldots$ converges to $x_\infty \in X$, written $\lim_n x_n = x_\infty$, if, for every neighborhood U of x_∞, there is an index m such that for all $n \geq m$, $x_n \in U$.

In more general, non-Hausdorff spaces, it is easily possible to have a sequence converge to more than one point, which is fairly contrary to our intention for the notion of *convergence*.

In a *metric* space, the notion of *Cauchy* sequence has a sense, and in a *complete* metric space, the notions of Cauchy sequence and convergent sequence are identical, and there is a unique limit to which such a sequence converges.

In more general, non-Hausdorff spaces, and not-locally-countably-based spaces, things can go haywire in several different ways, which are mostly irrelevant to the situations we care about. Still, one should be aware that not all spaces are Haudorff, and may fail to be countably locally based.

[2.5.14] Sequentially compact sets E in a topological space X are those such that every sequence has a convergent subsequence (with limit in E).

Although the definition of *convergent* does not directly mention potential difficulties and ambiguities, there are indeed problems in non-Haudorff spaces, and in spaces that fail to have countable local bases.

[2.5.15] Accumulation points of a subset E of a topological space X are points $x \in X$ such that every neighborhood of x contains infinitely-many elements of E. Every accumulation point of E lies in the *closure* of E, but not vice-versa.

[2.5.16] Claim: A closed set E is sequentially compact if and only if every sequence in E either has an accumulation point in E, or contains only finitely-many distinct points.

Proof: First, the technicality: if a sequence contains only finitely-many distinct points, it cannot have any accumulation points, but certainly contains convergent subsequences. For a sequence $x_1, x_2, \ldots$ including infinitely-many distinct points, drop any repeated points, so that $x_i \neq x_j$ for all $i \neq j$. For E sequentially compact, there is a subsequence with limit x_∞ in E. Relabel if necessary so that the subsequence is still denoted $x_1, x_2, \ldots$. The subsequence still consists of mutually distinct points. Since $\lim_n x_n = x_\infty$, given a neighborhood U of x_∞, there is m such that $x_n \in U$ for all $n \geq m$. Since $x_m, x_{m+1}, \ldots$ is an infinite set of distinct points, x_∞ is an accumulation point of the subsequence, hence, of the original sequence.

Conversely, if a sequence has an accumulation point, it has a subsequence converging to that accumulation point. ///

[2.5.17] Compact sets in topological spaces are subsets such that *every open cover has a finite subcover*. That is, K is compact when, for any collection of open sets $\{U_\alpha : \alpha \in A\}$ such that $K \subset \bigcup_{\alpha \in A} U_\alpha$, there is a finite collection $U_{\alpha_1}, \ldots, U_{\alpha_n}$ such that $K \subset U_{\alpha_1} \cup \ldots \cup U_{\alpha_n}$.

[2.5.18] Claim: For $f : X \to Y$ continuous and K compact in X, the image $f(K)$ is compact in Y.

Proof: Given an open cover $\{U_\alpha : \alpha \in A\}$ of $f(K)$, the inverse images $f^{-1}(U_\alpha)$ give an open cover of K. Thus, there is a finite subcover $f^{-1}(U_{\alpha_1}), \ldots, f^{-1}(U_{\alpha_n})$. Then $U_{\alpha_1}, \ldots, U_{\alpha_n}$ is a (finite) cover of $f(K)$. ///

Since singleton sets $\{x\}$ are certainly compact, the following generalizes the earlier claim about closedness of singleton sets in Hausdorff spaces:

[2.5.19] Claim: In Hausdorff spaces, compact sets are *closed*.

Proof: Let E be a compact subset of X. For $y \notin E$, for each $x \in E$, let $U_x \ni y$ be open and $V_x \ni x$ open so that $U_x \cap V_x = \phi$. Then $\{V_x : x \in E\}$ is an open cover of E, with finite subcover $E \subset V_{x_1} \cup \ldots \cup V_{x_n}$. The finite intersection $W_y = U_{x_1} \cap \ldots \cap U_{x_n}$ is open, and disjoint from $V_{x_1} \cup \ldots \cup V_{x_n}$, so is disjoint from E. Thus, W_y is open and contains y. The union $W = \bigcup_{y \notin E} W_y$ is open, and contains every $y \notin E$. Thus E is the complement of an open set, so is closed. ///

[2.5.20] Claim: In Hausdorff spaces, a *nested* collection of compact sets has non-empty intersection.

Proof: Let X be the ambient space, and K_α the compacts, with index set A *totally ordered*, in the sense A has an order relation $<$ such that for every distinct $\alpha, \beta \in A$, either $\alpha < \beta$ or $\beta < \alpha$. The *nested* condition is that if $\alpha < \beta$ then $K_\alpha \supset K_\beta$. (It can equally well be the opposite direction of containment.) We claim that $\bigcap_\alpha K_\alpha$ is compact.

From above, each K_α is closed, so the complements $U_\alpha = X - K_\alpha$ are open. If $\bigcap_\alpha K_\alpha = \phi$, then $\bigcup_\alpha U_\alpha = X$. In particular, $\bigcup_\alpha U_\alpha \supset K_\beta$ for all indices β. For fixed index α_o, let $U_{\alpha_1} \cup \ldots \cup U_{\alpha_n}$ be a finite subcover of K_{α_o}, so certainly a cover of $K_{\alpha'}$ for all $\alpha' > \alpha$. Because of the nested-ness, for $\beta = \max\{\alpha_1, \ldots, \alpha_n\}$, $U_\beta = U_{\alpha_1} \cup \ldots \cup U_{\alpha_n}$. But U_β is the complement of K_β, so certainly cannot cover it, contradiction. ///

[2.5.21] A locally compact topology is one in which every point has a neighborhood with compact closure. This is a reasonable condition to impose on a space on which functions will live. $\mathbb{R}^n$ is locally compact, but the metric space ℓ^2 is *not*. Later, we will see that *no* infinite-dimensional Hilbert space or Banach space is locally compact. That is, *natural spaces of functions* are not usually locally compact, but the physical spaces on which the functions live usually *are* locally compact.

[2.5.22] Separable topological spaces are those with countable dense subsets. For example, the countable set $\mathbb{Q}^n$ is dense in $\mathbb{R}^n$. Nearly all topological spaces arising in practice are separable, but most basic results do not directly use this property.

[2.5.23] Countably-based topological spaces are those with a countable basis. Sometimes such spaces are called *second-countable*. Perhaps counter-intuitively, *first-countable* spaces are those in which every point has a countable *local* basis. Many topological spaces arising in practice are countably-based, but most basic results do not directly use this property.

[2.5.24] Claim: Separable metric spaces are countably-based. Specifically, for countable dense subset S of metric space X, open balls of *rational* radius centered at points of S form a basis.

Proof: Since there are only countably-many $s \in S$ and only countably many rational radiuses, the set of such open balls is indeed countable.

Fix an open $U \subset X$. Given $x \in U$, let $r > 0$ be sufficiently small so that the open ball at x of radius r is inside U. Let $s_x \in S$ be such that $d(x, s) < r/2$. By density of rational numbers in $\mathbb{R}$, there is a rational number q_x such that $d(x, s) < q_x < r/2$. Thus, by the triangle inequality, the ball B_x at s_x of radius q_x contains x and lies inside the open ball at x of radius r, so $B_x \subset U$.

The union of all B_x over $x \in U$ is a subset of U containing all $x \in U$, so is U itself. ///

2.6 Compactness versus sequential compactness

In general topological spaces, *compactness* is a stronger condition than *sequential compactness*. First, without any further hypotheses on the spaces, however noting the point that sequential compactness easily fails to be what we anticipate in topological spaces that are not necessarily Hausdorff or locally countably-based:

[2.6.1] Claim: Compact sets are sequentially compact.

Proof: Given a sequence, if some $y \in E$ is an accumulation point, then there is a subsequence converging to y, and we are done. If no $y \in E$ is an accumulation point of the given sequence, then each $y \in E$ has an open neighborhood U_y such that U_y meets the sequence in only finitely-many points. The sets U_y cover E. For E compact, there is a finite subcover $U_{y_1}, \ldots, U_{y_n}$. Each U_{y_i} contains only finitely-many points of the sequence, so the sequence contains only finitely-many distinct points, so certainly has a convergent subsequence. ///

[2.6.2] Claim: In a countably-based topological space X, sequentially compact sets are compact.

Proof: Let $E \subset X$ be sequentially compact. The opens in an arbitrary cover of E are (necessarily countable) unions of some of the countably-many opens in the countable basis for X. Thus, it suffices to show that a *countable* cover $E \subset U_1 \cup U_2 \cup \ldots$ admits a finite subcover.

If *no* finite collection of the U_n covers E, then for each $n = 1, 2, \ldots$ there is $e_n \in E$ such that $e_n \notin U_1 \cup \ldots \cup U_n$. Since every e_n does lie in *some* U_i, we can replace $\{e_n\}$ by a subsequence so that $e_i \neq e_j$ for all $i \neq j$, and still $e_n \notin U_1 \cup \ldots \cup U_n$.

By sequential compactness, $e_1, e_2, \ldots$ has a convergent subsequence, with limit $e_\infty \in E$. The point e_∞ lies in some U_m. Thus, there would be infinitely-many indices n such that $e_n \in U_m$. This is impossible, since $e_n \notin U_1 \cup \ldots \cup U_n$. Thus, there must be a finite subcover. ///

The argument for the previous claim can be improved, to show

[2.6.3] Claim: In complete metric spaces, sequentially compact sets are compact.

[2.6.4] Remark: Again,

Proof: Let $\{U_\alpha : \alpha \in A\}$ be an open cover of a subset E of a complete metric space X, admitting no finite subcover. Using an equivalent of the Axiom of Choice, we can arrange to have a *minimal* subcover, that is, so that no U_β can be removed an still cover E. We do this at the end of the argument.

Granting this, without loss of generality the open cover is *minimal*, and not finite. Using the minimality (and again using the Axiom of Choice), for each index $\beta \in A$, let x_β be a point in E that is *not* in $\bigcup_{\alpha\neq\beta} U_\alpha$. Since the cover is minimal, these x_β's must be *distinct*. Since the cover is not finite, there are infinitely-many (distinct) x_β's. Since the are distinct, any countable subset of $\{x_\beta : \beta \in A\}$ gives a sequence $y_1, y_2, \ldots$ of distinct points. By sequential compactness, this sequence has at least one accumulation point $y_\infty \in E$.

Let U_{α_o} be an open in the cover containing y_∞. Since $\lim_n y_n = y_\infty$, there is n_o such that for all $n \geq n_o$ we have $y_n \in U_{\alpha_o}$. All those y_n's are among the x_β's, but the only x_β in U_{α_o} is x_{α_o}. That is, there cannot be infinitely-many distinct x_β's in U_{n_o}. Thus, assuming that a minimal cover is infinite leads to a contradiction.

To obtain a minimal subcover from a given cover $\{U_\alpha : \alpha \in A\}$, *well-order* the index set A. We choose a minimal subcover by transfinite induction, as follows. The idea is to ask, in the order chosen for A, cumulatively, whether or not U_α can be removed from the current subcover while still having a cover of the given set. That is, we inductively define a subset B of the index set A by transfinite induction: initially, $B = A$. At the α^{th} stage, remove α from B if U_α is unnecessary for maintaining the cover property. That is, remove α if

$$E \;\subset\; \bigcup_{\beta<\alpha,\ \beta\in B} U_\beta \;\cup\; \bigcup_{\beta>\alpha} U_\beta$$

otherwise keep α in B. By transfinite induction, B is an index set for a subcover of $\{U_\alpha : \alpha \in A\}$, and that subcover is *minimal* in the sense that no open can be removed without the result failing to be a cover.
///

2.7 Total-boundedness criterion for compact closure

In general metric spaces, closed and bounded sets need not be compact (nor sequentially compact). More is required, as follows.

A set E in a metric space is *totally bounded* if, given $\varepsilon > 0$, there are finitely-many open balls of radius ε covering E. The property of *total boundedness* in a metric space is generally stronger than mere *boundedness*. It is immediate that any subset of a totally bounded set is totally bounded.

[2.7.1] Theorem: A set E in a metric space X has compact closure *if and only if* it is totally bounded.

[2.7.2] Remark: Sometimes a set with compact closure is said to be *pre-compact*.

Proof: Certainly if a set has compact closure then it admits a finite covering by open balls of arbitrarily small (positive) radius, by the compactness.

On the other hand, suppose that a set E is totally bounded in a complete metric space X. To show that E has compact closure it suffices to show *sequential* compactness, namely, that any sequence $\{x_i\}$ in E has a convergent subsequence.

We choose such a subsequence as follows. Cover E by finitely-many open balls of radius 1, invoking the total boundedness. In at least one of these balls there are infinitely-many elements from the sequence. Pick such a ball B_1, and let i_1 be the smallest index so that x_{i_1} lies in this ball.

The set $E \cap B_1$ is still totally bounded (and contains infinitely-many elements from the sequence). Cover it by finitely-many open balls of radius $1/2$, and choose a ball B_2 with infinitely-many elements of the sequence lying in $E \cap B_1 \cap B_2$. Choose the index i_2 to be the smallest one so that both $i_2 > i_1$ and so that x_{i_2} lies inside $E \cap B_1 \cap B_2$.

Proceeding inductively, suppose that indices $i_1 < \ldots < i_n$ have been chosen, and balls B_i of radius $1/i$, so that

$$x_i \in E \cap B_1 \cap B_2 \cap \ldots \cap B_i$$

Then cover $E \cap B_1 \cap \ldots \cap B_n$ by finitely-many balls of radius $1/(n+1)$ and choose one, call it B_{n+1}, containing infinitely-many elements of the sequence. Let i_{n+1} be the first index so that $i_{n+1} > i_n$ and so that

$$x_{n+1} \in E \cap B_1 \cap \ldots \cap B_{n+1}$$

Then for $m < n$ we have $d(x_{i_m}, x_{i_n}) \leq \frac{1}{m}$ so this subsequence is Cauchy. ///

2.8 Baire's theorem

This standard result is both indispensable and mysterious.

A set E in a topological space X is *nowhere dense* if its closure $\bar{E}$ contains no non-empty open set. A *countable union* of nowhere dense sets is said to be *of first category*, while every other subset (if any) is *of second category*. The idea (not at all clear from this traditional terminology) is that first category sets are *small*, while second category sets are *large*. In this terminology, the theorem's assertion is equivalent to the assertion that (non-empty) *complete metric* spaces and *locally compact Hausdorff* spaces are *of second category*.

A G_δ set is a countable intersection of open sets. Concommitantly, an F_σ set is a countable union of closed sets. Again, the following theorem can be paraphrased as asserting that, in a complete metric space, *a countable intersection of dense G_δ's is still a dense G_δ.*

[2.8.1] Theorem: *(Baire)* Let X be either a complete metric space or a locally compact Hausdorff topological space. The intersection of a *countable* collection $U_1, U_2, \ldots$ of *dense open subsets* U_i of X is still *dense* in X.

Proof: Let B_o be a non-empty open set in X, and show that $\bigcap_i U_i$ meets B_o. Suppose that we have inductively chosen an open ball B_{n-1}. By the denseness of U_n, there is an open ball B_n whose closure $\overline{B_n}$ satisfies

$$\overline{B_n} \subset B_{n-1} \cap U_n$$

Further, for complete metric spaces, take B_n to have radius less than $1/n$ (or any other sequence of reals going to 0), and in the locally compact Hausdorff case take B_n to have compact closure.

Let

$$K = \bigcap_{n \geq 1} \overline{B_n} \subset B_o \cap \bigcap_{n \geq 1} U_n$$

For complete metric spaces, the centers of the nested balls B_n form a Cauchy sequence (since they are nested and the radii go to 0). By completeness, this Cauchy sequence *converges*, and the limit point lies inside each *closure* $\overline{B_n}$, so lies in the intersection. In particular, K is non-empty. For locally compact Hausdorff spaces, the intersection of a nested family of non-empty compact sets is non-empty, so K is non-empty, and B_o necessarily meets the intersection of the U_n. ///

2.9 Appendix: mapping-property characterization of completion

Our *intention* is that, when a metric space X is not complete, there should be a *complete* metric space $\widetilde{X}$ and an *isometry* (distance-preserving) $j : X \to \widetilde{X}$, such that every isometry $f : X \to Y$ to *complete* metric space Y *factors through* j uniquely. That is, there are *commutative diagrams*[10] of continuous maps

$$\begin{tikzcd} \widetilde{X} \arrow[dr, dotted, "\exists!"] & \\ X \arrow[u, "j"] \arrow[r, dashed, "\forall"'] & Y \end{tikzcd} \qquad \text{(for every isometry } X \to Y\text{)}$$

Without describing any *constructions* of completions, we can prove some things about the behavior of *any possible* completion. In particular, we prove that any two completions are *naturally isometrically isomorphic* to each other. *Thus, the outcome will be independent of construction.*

[2.9.1] Claim: *(Uniqueness)* Let $i : X \to Y$ and $j : X \to Z$ be two completions of a metric space X. Then there is a unique isometric homeomorphism $h : Y \to Z$ such that $j = h \circ i$. That is, we have a commutative diagram

$$\begin{tikzcd} Y \arrow[rr, dotted, "\exists!"] & & Z \\ & X \arrow[ul, "i"] \arrow[ur, "j"'] & \end{tikzcd}$$

Proof: First, take $Y = Z$ and $f : X \to Y$ to be the inclusion i, in the characterization of $i : X \to Y$. The characterization of $i : X \to Y$ shows that there is *unique* isometry $f : Y \to Y$ fitting into a commutative diagram

$$\begin{tikzcd} Y \arrow[dr, dotted, "\exists!\, f"] & \\ X \arrow[u, "i"] \arrow[r, "i"'] & Y \end{tikzcd}$$

Since the *identity* map $Y \to Y$ certainly fits into this diagram, the *only* map f fitting into the diagram is the identity on Y.

Next, applying the characterizations of both $i : X \to Y$ and $j : X \to Z$, we have unique $f : Y \to Z$ and $g : Z \to Y$ fitting into

$$\begin{tikzcd} Y \arrow[dr, dotted, "\exists!\, f"] & \\ X \arrow[u, "i"] \arrow[r, "j"'] & Z \end{tikzcd} \qquad \begin{tikzcd} Z \arrow[dr, dotted, "\exists!\, g"] & \\ X \arrow[u, "j"] \arrow[r, "i"'] & Y \end{tikzcd}$$

Then $f \circ g : Y \to Y$ and $g \circ f : Z \to Z$ fit into

$$\begin{tikzcd} Y \arrow[dr, dotted, "f \circ g"] & \\ X \arrow[u, "i"] \arrow[r, "i"'] & Y \end{tikzcd} \qquad \begin{tikzcd} Z \arrow[dr, dotted, "g \circ f"] & \\ X \arrow[u, "j"] \arrow[r, "j"'] & Z \end{tikzcd}$$

[10] A diagram of maps is *commutative* when the composite map from one object to another within the diagram does not depend on the route taken within the diagram.

By the first observation, this means that $f \circ g$ is the identity on Y, and $g \circ f$ is the identity on Z, so f and g are mutual inverses, and Y and Z are *homeomorphic*. ///

[2.9.2] Remark: A virtue of the characterization of completion is that it does not refer to the *internals* of any completion.

Next, we see that the mapping-property characterization of a completion does not introduce more points than absolutely necessary:

[2.9.3] Claim: Every point in a completion $\widetilde{X}$ of X is the limit of a Cauchy sequence in X. That is, X is *dense* in $\widetilde{X}$.

Proof: Write $d(,)$ for both the metric on X and its extension to $\widetilde{X}$. Let $Y \subset \widetilde{X}$ be the collection of limits of Cauchy sequences of points in X. We claim that Y itself is *complete*. Indeed, given a Cauchy sequence $\{y_i\}$ in Y with limit $z \in \widetilde{X}$, let $x_i \in X$ such that $d(x_i, y_i) < 2^{-i}$. It will suffice to show that $\{x_i\}$ is Cauchy with limit z. Indeed, given $\varepsilon > 0$, take N large enough so that $d(y_i, z) < \varepsilon/2$ for all $i \geq N$, and increase N if necessary so that $2^{-i} < \varepsilon/2$. Then, by the triangle inequality, $d(x_i, z) < \varepsilon$ for all $i \geq N$. Thus, Y is complete.

By the defining property of $\widetilde{X}$, every isometry $f : X \to Z$ to complete Z has a unique extension to an isometry $F : \widetilde{X} \to Z$ fitting into

$$\begin{tikzcd}
Y \arrow[r, "\subset"] & \widetilde{X} \arrow[dr, dotted, "F"] & \\
& X \arrow[ul] \arrow[u, "j"] \arrow[r, "f"] & Z
\end{tikzcd}$$

Since Y is already complete and $j(X) \subset Y$, the restriction of F to Y gives a diagram

$$\begin{tikzcd}
Y \arrow[dr, dotted, "F"] & \\
X \arrow[u, "j"] \arrow[r, "f"] & Z
\end{tikzcd}$$

That is, Y fits the characterization of a completion of X. By uniqueness, $Y \subset \widetilde{X}$ is a homeomorphism, so $Y = \widetilde{X}$. ///

3. Review examples discussion

[3.1] (There is not much hope in making sense of the outcome of an uncountable number of non-zero operations:) Let Ω be an *uncountable* collection of positive real numbers. Letting F range over all finite subsets of Ω, show that $\sup_F \sum_{\alpha\in F} \alpha = +\infty$.

Discussion: Let $\Omega_1 = \{\omega \in \Omega : \omega > 1\}$, and for $n = 2, 3, \ldots$, let $\Omega_n = \{\omega \in \Omega : \frac{1}{n} < \omega \le \frac{1}{n-1}\}$. There are countably many such sets, so in (at least) one of them Ω_{n_o} there must be infinitely-many elements of Ω (or else Ω would be a countable union of countable sets, hence countable). Then

$$\sup_F \sum_{\alpha\in F} \;\ge\; \sup_{F\subset\Omega_{n_o}} \sum_{\alpha\in F} \;\ge\; \sup_{F\subset\Omega_{n_o}} \#F\cdot\frac{1}{n_o} \;=\; \frac{1}{n_o}\sup_{F\subset\Omega_{n_o}} \#F \;=\; +\infty$$

because Ω_{n_o} is infinite. ///

[3.2] Prove (or review the proof) that a continuous real-valued function f on a finite closed interval $[a,b] \subset \mathbb{R}$ is *uniformly* continuous: for all $\varepsilon > 0$ there is $\delta > 0$ such that, for all $x, y \in [a,b]$, $|x-y| < \delta$ implies $|f(x) - f(y)| < \varepsilon$.

Discussion: Given $\varepsilon > 0$ and $x \in [a,b]$, take $\delta_x > 0$ such that $|x'-x| < 2\delta_x$ implies $|f(x')-f(x)| < \varepsilon/2$. The open intervals $(x-\delta_x, x+\delta_x)$ cover the compact set $[a,b]$, so there is a finite subcover $\{(x_j-\delta_{x_j}, x_j+\delta_{x_j}) : j = 1,\ldots,N\}$. The minimum $\delta = \min_{j=1,\ldots,N}\delta_j$ is positive (see above). For given $x \in [a,b]$, $x \in (x_j-\delta_{x_j}, x_j+\delta_{x_j})$ for some j.

For x' such that $|x'-x| < \delta$, we have $|x'-x_j| \le |x'-x| + |x-x_j| \le \delta + \delta_j \le 2\delta_j$, so $|f(x') - f(x_j)| < \varepsilon/2$, and

$$|f(x') - f(x)| \;\le\; |f(x') - f(x_j)| + |f(x_j) - f(x)| \;<\; \frac{\varepsilon}{2} + \frac{\varepsilon}{2} \;=\; \varepsilon$$

which is the uniform continuity. ///

[3.3] Prove (or review the proof) that a *uniform* pointwise limit of continuous, real-valued functions on $[a,b]$ is continuous.

Discussion: This is the archetype of a three-ε argument. Let the sequence by $\{f_n\}$, and the pointwise limit $f(x) = \lim_n f_n(x)$. Given $\varepsilon > 0$, by the *uniform* pointwise approach to the limit, take n_o large enough so that for all $m, n \ge n_o$, for all $x \in [a,b]$, $|f_m(x) - f_n(x)| < \varepsilon$. Then $|f(x) - f_n(x)| \le \varepsilon$ for all $x \in [a,b]$, for all $n \ge n_o$. By the uniform continuity of f_{n_o} on $[a,b]$, let $\delta > 0$ so that $|f_{n_o}(x) - f_{n_o}(y)| < \varepsilon$ for all $|x-y| < \delta$. Then

$$|f(x) - f(y)| \;\le\; |f(x) - f_{n_o}(x)| + |f_{n_o}(x) - f_{n_o}(y)| + |f_{n_o}(y) - f(y)| \;<\; \varepsilon + \varepsilon + \varepsilon$$

as desired. ///

Note: In the latter situation, there is no compulsion to go back and replace ε by $\varepsilon/3$, since it is obviously possible to do so.

[3.4] Prove (or review the proof) of the *Fundamental Theorem of Calculus*: for a *continuous* function f on $[a,b]$, the function $F(x) = \int_a^x f(t)\,dt$ is *continuously differentiable*, and has derivative f. (Use Riemann's integral.)

Discussion: We use the finite additivity property

$$\int_a^c f(x)\,dx \;=\; \int_a^v f(x)\,dx + \int_v^c f(x)\,dx \qquad \text{(for all } v < c \text{ between } a \text{ and } b)$$

Thus,

$$\frac{F(x+\delta)-F(x)}{\delta} - f(x) \;=\; \frac{\int_x^{x+\delta} f(t)\,dt}{\delta} - f(x)$$

By continuity of f, given $\varepsilon > 0$, take $\delta_o > 0$ sufficiently small so that

$$\sup_{y:x\le y\le x+\delta_o} |f(y)-f(x)| \;<\; \varepsilon$$

Then

$$\frac{\int_x^{x+\delta} f(t)\,dt}{\delta} - f(x) \;<\; \frac{(f(x)+\varepsilon)\cdot\delta}{\delta} - f(x) \;=\; \varepsilon$$

and, similarly,

$$\frac{\int_x^{x+\delta} f(t)\,dt}{\delta} - f(x) \;>\; \frac{(f(x)-\varepsilon)\cdot\delta}{\delta} - f(x) \;=\; -\varepsilon$$

Thus, given $\varepsilon > 0$, there is $\delta_o > 0$ such that for every $0 < \delta \le \delta_o$

$$\Big|\frac{F(x+\delta)-F(x)}{\delta} - f(x)\Big| \;<\; \varepsilon$$

(Finding $\delta_o < 0$ for the same inequality is similar.) ///

[The following is a discussion of the question I **meant** *to ask!!!]*

[3.5] Prove (or review the proof) that for a sequence of real-valued functions f_n on $[0,1]$ approaching f *uniformly* pointwise, $\lim_n \int_0^1 f_n(x)\,dx = \int_0^1 \lim_n f_n(x)\,dx$. (Use Riemann's integral.)

Discussion: Given $\varepsilon > 0$, let n_o be large enough so that for all $n \ge n_o$, for all $x \in [a,b]$, $|f_n(x) - f(x)| < \varepsilon$. Using *linearity* of integrals,

$$\int_a^b f(x)\,dx \;=\; \int_a^b f(x) - f_{n_o}(x)\,dx + \int_a^b f_{n_o}(x)\,dx$$

Upper and lower bounds are obtained from any upper and lower Riemann sums, for any partition $a = x_1 < \ldots < x_n = b$ of the interval:

$$\int_a^b f(x) - f_{n_o}(x)\,dx \;<\; \sum_{j=1}^n (x_{j+1}-x_j)\cdot\varepsilon \;=\; (b-a)\cdot\varepsilon$$

and similarly for a lower bound. ///

[3.6] Show that every open subset of $\mathbb{R}$ is a *countable* union of open intervals.

Discussion: Let S be the set. For $s \in S$, since S is open, there is $0 < \delta_s \in \mathbb{Q}$ such that $(s-2\delta_s, s+2\delta_s) \subset S$. By density of $\mathbb{Q}$ in $\mathbb{R}$ there is q_s in the smaller interval $(s-\delta_s, s+\delta_s)$. Certainly $s \in (q_s - \delta_s, q_s + \delta_s)$, and $(q_s - \delta_s, q_s + \delta_s) \subset S$, because for $|t - q_s| < \delta_s$

$$|s-t| \;\le\; |s-q_s| + |q_s - t| \;<\; \delta + \delta$$

The collection of *all* pairs $(q,\delta) \in \mathbb{Q}\times\mathbb{Q}$ of rationals q,δ is countable, so the subset of (distinct) pairs occuring as q_s, δ_s for $s \in S$ is countable. (Apparently many of the pairs (q,δ) appear as (q_s,δ_s) for many different $s \in S$.) ///

[3.7] Define *Lebesgue (outer) measure* $\mu(E)$ of subsets E of $\mathbb{R}$ given by

$$\mu(E) \;=\; \inf\{\sum_{n=1}^{\infty} |b_n - a_n| \;:\; E \subset \bigcup_{n=1}^{\infty} (a_n, b_n)\}$$

Show that $\mu(\mathbb{Q}) = 0$. Show that $\mu(M) = 0$, where M is Cantor's middle-thirds set.

Discussion: Enumerate the rationals as $r_1, r_2, \ldots$. Given $\varepsilon > 0$, let $U_{n,\varepsilon}$ be the interval $(r_n - \frac{\varepsilon}{2^n}, r + \frac{\varepsilon}{2^n})$. The union of these intervals contains $\mathbb{Q}$, and the sum of lengths is $\varepsilon \cdot (\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots) = \varepsilon$.

The Cantor middle-thirds set can be described in terms of base-three expansions, as follows. All real numbers r in $[0,1]$ have (ternary) expansion $r = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$ with all coefficients a_n in the set $\{0,1,2\}$. The expansion is unambiguous except for the possibility of coefficients all 2 beyond a certain point, which we exclude by using

$$\frac{2}{3^n} + \frac{2}{3^{n+1}} + \frac{2}{3^{n+2}} + \ldots \;=\; 2 \cdot \frac{3^{-n}}{1 - \frac{1}{3}} \;=\; 2 \cdot \frac{3^{1-n}}{3-1} \;=\; 3^{1-n}$$

Then the middle-thirds set C is the set of reals $r = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$ with all coefficients a_n in the set $\{0,2\}$ (with the convention excluding endlessly repeating 2's).

Alternatively, the middle-thirds set C is formed as a *nested intersection*, as follows. Let C_1 be $[0,1]$ with the middle third $(\frac{1}{3}, \frac{2}{3})$ removed. Let C_2 be C_1 with the middle third thirds $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ removed, and so on. At each step, the sum of lengths of the remaining intervals is multiplied by $(1 - \frac{1}{3}) = \frac{2}{3}$, and the number of intervals is multiplied by 2. After n middle-third removals, the result C_n is a union of 2^n intervals each of length 3^{-n}. The Cantor middle-thirds set is $C = \bigcap_n C_n$.

Given $\varepsilon > 0$, choose n large enough so that $2^n/3^n < \varepsilon/2$. Cover each of the 2^n intervals of length 3^{-n} making up C_n by an open interval of length $2 \cdot 3^{-n}$. The sum of the lengths of these 2^n open intervals is

$$2^n \cdot (2 \cdot 3^{-n}) \;=\; 2 \cdot (2/3)^n \;<\; 2 \cdot \frac{\varepsilon}{2} \;=\; \varepsilon$$

This exhibits an open cover of C_n with sum of lengths less than ε. Since $C \subset C_n$, this gives such a cover of C itself, as desired. ///

4. Measure and integral

4.1 Borel-measurable functions and pointwise limits

Pointwise limits of continuous functions on $\mathbb{R}$ or on intervals $[a, b]$ need not be continuous. We want a class of functions closed under taking pointwise limits of sequences. The following is the simplest form of a general discussion.

The collection of *Borel subsets* of $\mathbb{R}$ is the smallest collection of subsets of $\mathbb{R}$ closed under taking *countable unions*, under *countable intersections*, under *complements*, and containing all open and closed subsets of $\mathbb{R}$. This is also called the Borel *σ-algebra* in $\mathbb{R}$. To be sure that this description makes sense, we prove:

[4.1.1] Claim: Intersections of σ-algebras of subsets of $\mathbb{R}$ are σ-algebras. Thus, the *smallest* σ-algebra containing a given set of sets is the intersection of all σ-algebras containing it.

Proof: Let S be a set of subsets of a set X, and $\{A_i : i \in I\}$ a collection of σ-algebras containing S. Let A be the intersection $\bigcap_i A_i$. Given a countable collection $E_1, E_2, \ldots$ of sets in A, for every $i \in I$ the set E_j are in A_i, so their intersection and union are in A_i. Since this holds for every $i \in I$, that intersection and union are in A. The argument for complements is even simpler. ///

There is traditional terminology for certain simple types of Borel sets. For example a *countable intersection of open sets* is a G_δ set, while a *countable union of closed sets* is an F_σ. The notation can be iterated: a $G_{\delta\sigma}$ is a countable union of countable intersections of opens, and so on. We will not need this.

A simple useful choice of larger class of functions than continuous is: a real-valued or complex-valued function f on $\mathbb{R}$ is *Borel-measurable* when the inverse image $f^{-1}(U)$ is a Borel set for every open set U in the target space.

First, we verify some immediate desirable properties:

[4.1.2] Claim: The sum and product of two Borel-measurable functions are Borel-measurable. For non-vanishing Borel-measurable f, $1/f$ is Borel-measurable.

Proof: As a warm-up to this argument, it is useful to rewrite the $\varepsilon - \delta$ proof, that the sum of two continuous functions is continuous, in terms of the condition that inverse images of opens are open.

For Borel-measurable f, g on $\mathbb{R}$, let $f \oplus g$ be the $\mathbb{R} \times \mathbb{R}$-valued function on $\mathbb{R} \times \mathbb{R}$ defined by $(f \oplus g)(x, y) = (f(x), g(y))$. Let $s : \mathbb{R} \times \mathbb{R}$ be the sum map, $s(x, y) = x + y$. Let $\Delta : \mathbb{R} \to \mathbb{R} \times \mathbb{R}$ be the diagonal map $\Delta(x) = (x, x)$. Both s and Δ are continuous, and

$$(f + g)^{-1} \;=\; \Delta^{-1} \circ (f \oplus g)^{-1} \circ s^{-1}$$

Since s is continuous, for open $U \subset \mathbb{R}$, $s^{-1}(U)$ is open in $\mathbb{R} \times \mathbb{R}$, and is a countable union of open rectangles $(a_i, b_i) \times (c_i, d_i)$. Then

$$(f \oplus g)^{-1}(s^{-1}(U)) \;=\; \bigcup_i (f \oplus g)^{-1}((a_i, b_i) \times (c_i, d_i)) \;=\; \bigcup_i f^{-1}(a_i, b_i) \times g^{-1}(c_i, d_i)$$

and every inverse image $f^{-1}(a_i, b_i)$ and $g^{-1}(c_i, d_i)$ is *Borel measurable.* Then

$$\Delta^{-1}\Big(f^{-1}(a_i, b_i) \times g^{-1}(c_i, d_i)\Big) \;=\; f^{-1}(a_i, b_i) \cap g^{-1}(c_i, d_i)\Big) \;=\; \text{(Borel measurable)}$$

The countable union indexed by i is still Borel-measurable, so $(f+g)^{-1}(U)$ is measurable. The arguments for product and inverse are nearly identical, since product and inverse (away from 0) are continuous. ///

It is sometimes useful to allow the target space for functions to be the *two-point compactification* $Y = \{-\infty\} \cup \mathbb{R} \cup +\infty$ of the real line, with neighborhood basis $-\infty \cup (-\infty, a)$ at $-\infty$ and $(a, +\infty) \cup \{+\infty\}$ at $+\infty$ when we need to allow functions to blow up in some fashion. But $\pm\infty$ are not numbers, and do not admit consistent manipulation as though they were.

A more serious positive indicator of the reasonable-ness of Borel-measurable functions as a larger class containing continuous functions:

[4.1.3] Theorem: Every pointwise limit of Borel-measurable functions is Borel-measurable. More generally, every countable *inf* and countable *sup* of Borel-measurable functions is Borel-measurable, as is every countable *liminf* and *limsup*.

Proof: We prove that a countable $f(x) = \inf_n f_n(x)$ is measurable. Observe that $f(x) < b$ if and only if there is some n such that $f_n(x) < b$. Thus,

$$f^{-1}(-\infty, b) \;=\; \bigcup_n f_n^{-1}(-\infty, b) \;=\; \text{(countable union of measurables)} \;=\; \text{(measurable)}$$

Further,

$$f^{-1}(-\infty, a] \;=\; \bigcap_n f^{-1}(-\infty, a + \frac{1}{n}) \;=\; \text{(countable intersection of measurables)} \;=\; \text{(measurable)}$$

and then

$$f^{-1}(a, b) \;=\; f^{-1}(-\infty, b) - f^{-1}(-\infty, a] \;=\; f^{-1}(-\infty, b) \cap (\mathbb{R} - f^{-1}(-\infty, a])$$

$$= \text{(intersection of measurable with complement of measurable)} \;=\; \text{(measurable)}$$

A nearly identical argument proves measurability of countable *sup*s of measurable functions.

A slight enhancement of this argument treats *liminfs* and *limsups*: $\limsup_n f_n(x) < b$ if and only if, for all n_o, there is $n \ge n_o$ such that $f_n(x) < b$:

$$\{x : \liminf_n f_n(x) < b\} \;=\; \bigcap_{n \ge 1} \Big(\bigcup_{n \ge n_o} f_n^{-1}(-\infty, b) \Big)$$

$$= \text{(countable intersection of countable unions of measurables)} \;=\; \text{(measurable)}$$

The rest of the argument for measurability of pointwise *liminfs* is identical to that for *infs*, and also for *limsups*. When pointwise $\lim_n f_n(x)$ exists, it is $\liminf_n f_n(x)$, showing that countable limits of measurable are measurable. ///

4.2 Lebesgue-measurable functions and almost-everywhere pointwise limits

A sequence $\{f_n\}$ of Borel-measurable functions on $\mathbb{R}$ converges (pointwise) *almost everywhere* when there is a Borel set $N \subset \mathbb{R}$ of measure 0 such that $\{f_n\}$ converges pointwise on $\mathbb{R} - N$. One of Lebesgue's discoveries was that ignoring what may happen on sets of measure zero was an essential simplifying point in many situations.

However, there are sets of Lebesgue measure 0 that are not Borel sets. Thus, *almost-everywhere* pointwise limits of Borel-measurable functions may fall into a larger class. That is, there is a larger σ-algebra than that of Borel sets. Indeed, the description of the Lebesgue (outer) measure suggests that *any subset F of a Borel set E of measure zero should itself be measurable, with measure zero.*

The smallest σ-algebra containing all Borel sets in $\mathbb{R}$ and containing all subsets of Lebesgue-measure-zero Borel sets is the σ-algebra of *Lebesgue-measurable* sets in $\mathbb{R}$.

[4.2.1] Claim: Finite sums, finite products, and inverses (of non-zero) Lebesgue-measurable functions are Lebesgue-measurable.

Proof: The proofs in the previous section did not use any specifics of the σ-algebra of Borel-measurable functions, so the same proofs succeed. ///

[4.2.2] Theorem: Every pointwise-almost-everywhere limit of Lebesgue-measurable functions f_n is Lebesgue-measurable.

Proof: Again, the proofs in the previous section did not use any specifics of the σ-algebra of Borel-measurable functions. ///

4.3 Borel measures

A *Borel measure* μ is an assignment of (often *non-negative*) real numbers $\mu(E)$ (measures) to Borel sets E, in a fashion that is *countably additive* for disjoint unions:

$$\mu(E_1 \cup E_2 \cup E_3 \cup \ldots) \;=\; \mu(E_1) + \mu(E_2) + \mu(E_3) + \ldots \qquad \text{(for } \textit{disjoint} \text{ Borel sets } E_1, E_2, E_3, \ldots)$$

The most important prototype of a Borel measure is *Lebesgue (outer) measure* of a Borel set $E \subset \mathbb{R}$, described by

$$\mu(E) \;=\; \inf\{\sum_{n=1}^{\infty} |b_n - a_n| \;:\; E \subset \bigcup_{n=1}^{\infty} (a_n, b_n)\}$$

That is, it is the *inf* of the sums of lengths of the intervals in a countable cover of E by open intervals. For example, any countable set has (Lebesgue) measure 0.

That is, there is a σ-algebra A including Borel sets (equivalently, including open sets), and μ is a (often non-negative real-valued) function on A with the countable additivity above.

[... iou ...]

[4.3.1] Remark: Assuming the Axiom of Choice, one can prove that there is no Borel measure μ with σ-algebra containing *all* subsets of $\mathbb{R}$. So our ambitions for assigning measures should be more modest.

4.4 Lebesgue integrals

With such notion of *measure*, there is a corresponding *integrability* and *integral*, due to Lebesgue. It amounts to replacing the literal rectangles used in Riemann integration by more general rectangles, with bases not just intervals, but measurable sets, as follows.

The *characteristic function* or *indicator function* ch_E or χ_E of a measurable subset $E \subset \mathbb{R}$ is 1 on E and 0 off. A *simple function* is a finite, positive-coefficiented, linear combination of characteristic functions of bounded measurable sets, that is, is of the form

$$\text{(simple function) } s \;=\; \sum_{i=1}^{n} c_i \cdot \mathrm{ch}_{E_i} \qquad\qquad (\text{with } c_i \geq 0)$$

The *integral* of s is what one would expect:

$$\int s\, d\mu \;=\; \int \Big(\sum_{i=1}^{n} c_i \cdot \mathrm{ch}_{E_i}\Big)\, d\mu \;=\; \sum_i c_i \cdot \mu(E_i)$$

Next, the measure of a *non-negative* function f is the *sup* of the integrals of all simple functions between f and 0:

$$\int f\, d\mu \;=\; \sup_{0 \leq s \leq f} \int s\, d\mu \qquad\qquad (\text{sup over simple } s \text{ with } 0 \leq s(x) \leq f(x) \text{ for all } x)$$

After proving that the positive and negative parts f_+ and f_- of Borel measurable real-valued f are again Borel measurable,

$$\int f\, d\mu \;=\; \int f_+\, d\mu - \int (-f_-)\, d\mu$$

Similarly, for complex-valued f, break f into real and imaginary parts.

There are details to be checked:

[4.4.1] Theorem: Borel-measurable functions f, g taking values in $[0, +\infty]$ are *integrable*, in the sense that the previous prescription yields an assignment $f \to \int_{\mathbb{R}} f \in [0, +\infty]$ such that for positive constants a, b

$$\int_{\mathbb{R}} (af + bg) \;=\; a \int_{\mathbb{R}} f + b \int_{\mathbb{R}} g \qquad\qquad (\text{for all } a, b \geq 0)$$

For complex-valued Borel-measurable f, g, the absolute values $|f|$ and $|g|$ are Borel-measurable. Assuming $\int_{\mathbb{R}} |f| < \infty$ and $\int_{\mathbb{R}} |g| < \infty$, for any complex a, b

$$\int_{\mathbb{R}} (af + bg) \;=\; a \int_{\mathbb{R}} f + b \int_{\mathbb{R}} g$$

Proof: [... iou ...] ///

For a Borel-measurable function f on $\mathbb{R}$ and Borel-measurable set $E \subset \mathbb{R}$, the *integral of f over E* is

$$\int_E f \;=\; \int_{\mathbb{R}} \mathrm{ch}_E \cdot f$$

where ch_E is the characteristic function of f.

4.5 Convergence theorems: monotone, dominated

Easy, natural examples show that *pointwise* limits $f = \lim_n f_n$ of measurable functions f_n, while still measurable, need *not* satisfy $\int f = \lim \int f_n$. That is, this failure is not a pathology, but, rather, is completely reasonable. Hence additional conditions are essential to know that the integral of a pointwise limit is the limit of the integrals.

First, a relatively simple initial step:

[4.5.1] Theorem: *(Fatou's lemma)* For Borel-measurable f_n with values in $[0, +\infty]$, the pointwise $f(x) = \liminf_n f_n(x)$ is Borel-measurable, and

$$\int \liminf_n f_n(x)\, dx \;\le\; \liminf_n \int f_n$$

Proof: [... iou ...] ///

[4.5.2] Theorem: *(Lebesgue: monotone convergence)* Let $f_1, f_2, \ldots$ be a sequence of non-negative real-valued Lebesgue-measurable functions on $[a, b]$, with $f_1(x) \le f_2(x) \le \ldots$ for all x. Then $\int_a^b \lim_n f_n(x)\, dx = \lim_n \int_a^b f_n(x)\, dx$. This includes the possibility that some of the limits of the pointwise values are $+\infty$, and that the integral of the limit is $+\infty$.

Proof: [... iou ...] ///

[4.5.3] Theorem: *(Lebesgue: dominated convergence)* Let $f_1, f_2, \ldots$ be a sequence of complex-valued Lebesgue-measurable functions on $[a, b]$, with $|f_n(x)| \le g(x)$ for all x, for some measurable g with $\int_a^b g(x)\, dx < +\infty$. Then $\int_a^b \lim_n f_n(x)\, dx = \lim_n \int_a^b f_n(x)\, dx$.

Proof: [... iou ...] ///

4.6 Urysohn's lemma

Urysohn's lemma proves existence of sufficiently many functions on reasonable topological spaces.

[4.6.1] Theorem: *(Urysohn)* In a locally compact Hausdorff topological space X, given a compact subset K contained in an open set U, there is a continuous function $0 \le f \le 1$ which is 1 on K and 0 off U.

Proof: First, we prove that there is an open set V such that

$$K \subset V \subset \overline{V} \subset U$$

For each $x \in K$ let V_x be an open neighborhood of x with compact closure. By compactness of K, some finite subcollection $V_{x_1}, \ldots, V_{x_n}$ of these V_x cover K, so K is contained in the open set $W = \bigcup_i V_{x_i}$ which has compact closure $\bigcup_i \overline{V}_{x_i}$ since the union is *finite*.

Using the compactness again in a similar fashion, for each x in the closed set $X - U$ there is an open W_x containing K and a neighborhood U_x of x such that $W_x \cap U_x = \phi$.

Then

$$\bigcap_{x \in X-U} (X - U) \cap \overline{W} \cap \overline{W}_x \;=\; \phi$$

These are compact subsets in a Hausdorff space, so (again from compactness) some *finite* subcollection has empty intersection, say

$$(X - U) \cap \left(\overline{W} \cap \overline{W}_{x_1} \cap \ldots \cap \overline{W}_{x_n}\right) \;=\; \phi$$

That is,

$$\overline{W} \cap \overline{W}_{x_1} \cap \ldots \cap \overline{W}_{x_n} \;\subset\; U$$

Thus, the open set

$$V = W \cap W_{x_1} \cap \ldots \cap W_{x_n}$$

meets the requirements.

Using the possibility of inserting an open subset and its closure between any $K \subset U$ with K compact and U open, we inductively create opens V_r (with compact closures) indexed by rational numbers r in the interval $0 \leq r \leq 1$ such that, for $r > s$,

$$K \subset V_r \subset \overline{V}_r \subset V_s \subset \overline{V}_s \subset U$$

From any such configuration of opens we construct the desired continuous function f by

$$f(x) = \sup\{r \text{ rational in } [0,1] : x \in V_r, \} = \inf\{r \text{ rational in } [0,1] : x \in \overline{V}_r, \}$$

It is not immediate that this sup and inf are the same, but if we *grant* their equality then we can prove the *continuity* of this function $f(x)$. Indeed, the sup description expresses f as the supremum of characteristic functions of open sets, so f is at least *lower semi-continuous*. [11] The inf description expresses f as an infimum of characteristic functions of closed sets so is *upper* semi-continuous. Thus, f would be continuous.

To finish the argument, we must construct the sets V_r and prove equality of the inf and sup descriptions of the function f.

To construct the sets V_i, start by finding V_0 and V_1 such that

$$K \subset V_1 \subset \overline{V}_1 \subset V_0 \subset \overline{V}_0 \subset U$$

Fix a well-ordering $r_1, r_2, \ldots$ of the rationals in the open interval $(0,1)$. Supposing that $V_{r_1}, \ldots, v_{r_n}$ have been chosen. let i, j be indices in the range $1, \ldots, n$ such that

$$r_j > r_{n+1} > r_i$$

and r_j is the *smallest* among $r_1, \ldots, r_n$ *above* r_{n+1}, while r_i is the *largest* among $r_1, \ldots, r_n$ *below* r_{n+1}. Using the first observation of this argument, find $V_{r_{n+1}}$ such that

$$V_{r_j} \subset \overline{V}_{r_j} \subset V_{r_{n+1}} \subset \overline{V}_{r_{n+1}} \subset V_{r_i} \subset \overline{V}_{r_i}$$

This constructs the nested family of opens.

Let $f(x)$ be the sup and $g(x)$ the inf of the characteristic functions above. If $f(x) > g(x)$ then there are $r > s$ such that $x \in V_r$ and $x \notin \overline{V}_s$. But $r > s$ implies that $V_r \subset \overline{V}_s$, so this cannot happen. If $g(x) > f(x)$, then there are rationals $r > s$ such that

$$g(x) > r > s > f(x)$$

Then $s > f(x)$ implies that $x \notin V_s$, and $r < g(x)$ implies $x \in \overline{V}_r$. But $V_r \subset \overline{V}_s$, contradiction. Thus, $f(x) = g(x)$. ///

4.7 Comparison to continuous functions: Lusin's theorem

One aspect of the following theorem is that we have not inadvertently needlessly included functions wildly unrelated to continuous functions:

[11] A (real-valued) function f is *lower* semi-continuous when for all bounds B the set $\{x : f(x) > B\}$ is open. The function f is *upper* semi-continuous when for all bounds B the set $\{x : f(x) < B\}$ is open. It is easy to show that a sup of lower semi-continuous functions is lower semi-continuous, and an inf of upper semi-continuous functions is upper semi-continuous. As expected, a function both upper and lower semi-continuous is continuous.

[4.7.1] Theorem: *(Lusin)* Continuous functions approximate Borel-measurable functions well: given Borel-measurable real-valued or complex-valued f on $\mathbb{R}$, for every $\varepsilon > 0$ and for every Borel subset $\Omega \subset \mathbb{R}$ of finite Lebesgue measure, there is a relative closed $E \subset \Omega$ such that $\mu(\Omega - E) < \varepsilon$, and $f|_E$ is *continuous.*

Proof: [... iou ...] ///

Not much better can be done than Lusin's theorem says: for example, continuous approximations to the Heaviside step function

$$H(x) \;=\; \begin{cases} 0 & \text{for } x < 0 \\ \\ 1 & \text{for } x \geq 0 \end{cases}$$

have to go from 0 to 1 *somewhere*, by the Intermediate Value Theorem, so will be in $(\frac{1}{4}, \frac{3}{4})$ on an open set of strictly positive measure.

[4.7.2] Remark: It turns out that the everyday use of measure theory, measurable functions, and so on, does *not* proceed by way of Lusin's theorem or similar direct connections with continuous functions, but, rather, by direct interaction with the more general ideas.

4.8 Comparison to uniform pointwise convergence: Severini-Egoroff

[4.8.1] Theorem: *(Severini, Egoroff)* Pointwise convergence of sequences of Borel-measurable functions is approximately *uniform* convergence: given a almost-everywhere pointwise-convergent sequence $\{f_n\}$ of Borel-measurable functions on $\mathbb{R}$, for every $\varepsilon > 0$ and for every Borel subset $\Omega \subset \mathbb{R}$ of finite Lebesgue measure, there is a Borel subset $E \subset \Omega$ such that $\{f_n\}$ converges *uniformly* pointwise on E.

Proof: [... iou ...] ///

[4.8.2] Remark: Despite the connection that the Severini-Egoroff theorem makes between pointwise and *uniform* pointwise convergence, this idea turns out *not* to be the way to understand convergence of measurable functions. Instead, the game becomes ascertaining additional conditions that guarantee convergence of integrals, as earlier.

4.9 Abstract integration on measure spaces

An elementary but fundamental result is

[4.9.1] Proposition: Let f be a $[0, +\infty]$-valued measurable function on X. Then there are simple functions $s_1, s_2, s_3, \ldots$ *with non-negative real coefficients* so that for all $x \in X$, $s_1(x) \leq s_2(x) \leq s_3(x) \leq \ldots \leq f(x)$, and for all $x \in X$, $\lim_n s_n(x) = f(x)$.

Note: Some authors distinguish between *positive* measures and *complex* measures, where the distinction is meant to be that the former are $[0, \infty]$-valued, while the latter are constrained to assume only *'finite'* complex values.

The *integral of a characteristic function* χ_E is taken to be simply

$$\int_X \chi_E \; d\mu = \mu(E)$$

Then the *integral of a simple function*

$$s(x) = \sum_{1 \leq i \leq n} c_i \chi_{E_i}$$

(with $c_i \geq 0$) is defined to be

$$\int_X \sum_{1\leq i\leq n} c_i \chi_{E_i} = \sum_{1\leq i\leq n} c_i \int_X \chi_{E_i}\, d\mu = \sum_{1\leq i\leq n} c_i \int_X \mu E_i$$

For a $[0, +\infty]$-valued function f, we write

$$0 \leq s \leq f$$

for a *simple* function s if s has *non-negative real* coefficients, and if for all $x \in X$

$$0 \leq s(x) \leq f(x)$$

Then the *Lebesgue integral* of f is defined to be

$$\int_X f\, d\mu = \sup_{s:0\leq s\leq f} \int_X s\, d\mu$$

Note that at this point we can only integrate *non-negative real-valued* functions.

The standard space

$$L^1(X, \mu) = \{\text{complex-valued measurable } f \text{ so that } \textstyle\int_X |f|\, d\mu < \infty\}$$

Since $|f|$ is non-negative real-valued, we can indeed make sense of this. This is the collection of *integrable* functions f. Then write

$$f(x) = u(x) + iv(x)$$

where both u, v are real-valued, and write

$$u = u_+ - u_- \qquad v = v_+ - v_-$$

where u_+, v_+ are the 'positive parts' and where u_-, v_- are the 'negative parts' of these functions. Define the *Lebesgue integral*

$$\int_X f\, d\mu = \int_X u_+\, d\mu - \int_X u_-\, d\mu + i\int_X v_+\, d\mu - i\int_X v_-\, d\mu$$

Then we have to check that this definition, in terms of integrals of non-negative functions, really has the presumed properties. It is in proving such that we need the *integrability*.

For brevity, when there is no chance of confusion we will often simply write

$$\int_X f$$

rather than either of

$$\int_X f\, d\mu, \qquad \int_X f(x)\, d\mu(x)$$

for the integral of f on the measure space X with respect to the measure μ.

4.10 Lebesgue-Radon-Nikodym theorem

Let μ, ν be two positive measures on a common sigma algebra $\mathcal{A}$ on a set X. Say that ν is *absolutely continuous* with respect to μ if $\mu(E) = 0$ implies $\nu(E) = 0$ for all measurable sets E. This is often written $\nu < \mu$. The measure μ is *supported on* or *concentrated on* a subset X_o of X if, for all measurable E,

$$\mu(E) = \mu(E \cap X_o)$$

The two measures μ, ν are *mutually singular* if μ is supported on X_1 and ν is supported on X_2 and $X_1 \cap X_2 = \emptyset$. This is often written $\mu \perp \nu$.

[4.10.1] Theorem: Theorem. Let μ, ν be positive measures on a common sigma-algebra $\mathcal{A}$ on a set X. There is a unique pair of positive measures ν_a and ν_s so that

$$\nu_a < \mu \qquad \nu_s \perp \mu$$

Further, there is $\varphi \in L^1(X, \mu)$ so that for any measurable set E

$$\nu_a(E) = \int_X \varphi \, d\mu$$

The function φ is the *Radon-Nikodym derivative* of ν_a with respect to μ, and is often written as

$$\varphi = \frac{d\nu_a}{d\mu}$$

The pair (ν_a, ν_s) is the *Lebesgue decomposition* of ν with respect to μ.

5. Examples discussion

[5.8] Show that ℓ^2 is *complete* as a metric space.

Discussion: We can do this directly, although it is also a special case of the general fact that $L^2(X,\mu)$ is complete. Indeed, the argument will be a somewhat simpler version of the more general proof.

Let $f_1, f_2, \ldots$ be a Cauchy sequence in ℓ^2. Let $f(n)$ be the n^{th} component of $f \in \ell^2$, for $n = 1, 2, \ldots$. For any $f \in \ell^2$, certainly $|f(n)| \le |f|_{\ell^2}$, so for each n the scalar sequence $f_1(n), f_2(n), f_3(n), \ldots$ must be Cauchy, thus has a limit $f(n)$. We claim that $f = (f(1), f(2), f(3), \ldots)$ is in ℓ^2, and is the ℓ^2 limit of the f_i.

Given $\varepsilon > 0$, there is N sufficiently large so that $|f_i - f_j|_{\ell^2} < \varepsilon$ for all $i, j \ge N$. By a discrete version of Fatou's lemma, for $i \ge N$,

$$\sum_n |f(n) - f_i(n)|^2 \;=\; \sum_n \lim_j |f_j(n) - f_i(n)|^2 \;=\; \sum_n \liminf_j |f_j(n) - f_i(n)|^2 \;\le\; \liminf_j \sum_n |f_j(n) - f_i(n)|^2$$

$$\le\; \liminf_j |f_j - f_i|^2_{\ell^2} \;\le\; \liminf_j \varepsilon^2 \;=\; \varepsilon^2$$

Thus, $f - f_i \in \ell^2$, so $f = (f - f_i) + f_i \in \ell^2$. Then the previous computation shows that for given ε for $i \ge N$ we have $|f - f_i| \le \varepsilon$. Thus, $f_i \to f$ in ℓ^2. ///

Discrete version of Fatou's Lemma: We claim that for $[0, +\infty]$-valued functions f_j on $\{1, 2, 3, \ldots\}$,

$$\sum_{n=1}^{\infty} \liminf_j f_j(n) \;\le\; \liminf_j \sum_{n=1}^{\infty} f_j(n)$$

Proof: Letting $g_j(n) = \inf_{i \ge j} f_j(n)$, certainly $g_j(n) \le f_j(n)$ for all n, and $\sum_n g_j(n) \le \sum_n f_j(n)$. Also, $g_1(n) \le g_2(n) \le \ldots$ for all n, and $\lim_j g_j(n) = \liminf_j f_j(n)$. A discrete form of the Monotone Convergence Theorem, proven just below, is

$$\sum_n \lim_j g_j(n) \;=\; \lim_j \sum_n g_j(n)$$

Thus,

$$\sum_n \liminf_j f_j(n) \;=\; \sum_n \lim_j g_j(n) \;=\; \lim_j \sum_n g_j(n) \;=\; \liminf_j \sum_n g_j(n) \;\le\; \liminf_j \sum_n f_j(n)$$

as claimed. ///

Similarly, we have

Discrete version of Lebesgue's Monotone Convergence Theorem: For $[0, +\infty]$-valued functions f_j on $\{1, 2, 3, \ldots\}$, with $f_1(n) \le f_2(n) \le \ldots$ for all n,

$$\lim_j \sum_{n=1}^{\infty} f_j(n) \;=\; \sum_{n=1}^{\infty} \lim_j f_j(n) \qquad\qquad \text{(allowing value } +\infty)$$

Proof: Each non-decreasing sequence $f_1(n) \le f_2(n) \le \ldots$ has a limit $f(n) \in [0, +\infty]$. Similarly, since $\sum_n f_j(n) \le \sum_n f_{j+1}(n)$, the non-decreasing sequence of these sums has a limit $S = \lim_j \sum_n f_j(n)$. Since $f_j(n) \le f(n)$, certainly $\sum_n f_j(n) \le \sum_n f(n)$, and $S \le \sum_n f(n)$.

Fix N, and put $g(n) = f(n)$ for $n \le N$ and $g(n) = 0$ for $n > N$. For $\varepsilon > 0$, let

$$E_j \;=\; \{n : \sum_n f_j(n) \ge (1-\varepsilon) \cdot \sum_n g(n)\} \qquad \text{(for } j = 1, 2, \ldots)$$

Certainly $E_1 \subset E_2 \subset \ldots$, since $f_{j+1}(n) \ge f_j(n)$ for all n. We claim that $\bigcup E_j = \{1, 2, \ldots\}$: for $f(n) > 0$,

$$\lim_j f_j(n) \;=\; f(n) \;>\; (1-\varepsilon) \cdot f(n) \;\ge\; (1-\varepsilon) \cdot g(n) \qquad \text{(for all } n)$$

and for $f(n) = 0$, also $g(n) = 0$, and

$$f_1(n) \;\ge\; 0 \;\ge\; (1-\varepsilon) \cdot g(n)$$

Then

$$\sum_n f_j(n) \;\ge\; \sum_{n \in E_j} f_j(n) \;\ge\; (1-\varepsilon) \cdot \sum_{n \in E_j} g(n)$$

The set of n for which $g(n)$ is non-zero is finite, so there is j_o such that for $j \ge j_o$

$$\sum_{n \in E_j} g(n) \;=\; \sum_n g(n) \qquad \text{(for all } j \ge j_o)$$

That is, $\lim_j \sum_n f_j(n) \ge (1-\varepsilon) \sum_n g(n)$. Then

$$S \;=\; \lim_j \sum_n f_j(n) \;\ge\; (1-\varepsilon) \cdot \lim_j \sum_{n \in E_j} g(n) \;=\; (1-\varepsilon) \cdot \sum_n g(n)$$

This holds for every $\varepsilon > 0$, so $S \ge \sum_n g(n) = \sum_{n \le N} f(n)$. This holds for every N, so $S \ge \sum_n f(n)$. ///

[5.9] Show that the characteristic function χ_E of a measurable set E is measurable.

Discussion: For non-empty open $U \subset \mathbb{R}$, $\chi_E^{-1}(U)$ is the measurable set ϕ if U does not contain either 0 or 1. If $U \ni 1$ but $U \not\ni 0$, then $\chi_E^{-1}(U) = E$, which is measurable. If $U \ni 0$ but $U \not\ni 1$, then $\chi_E^{-1}(U) = E^c$, the complement of E, which is measurable. If U contains both 0 and 1, then $\chi_E^{-1}(U)$ is the whole domain space, which is measurable. ///

[5.10] Show that the product of two $\mathbb{R}$-valued measurable functions on $\mathbb{R}$ is measurable.

Discussion: Let f, g be measurable functions. Let $\Delta : \mathbb{R} \to \mathbb{R}^2$ by $\Delta(x) = (x, x)$, $s : \mathbb{R}^2 \to \mathbb{R}$ by $m(x, y) = x \cdot y$, and $f \oplus g : \mathbb{R}^2 \to \mathbb{R}^2$ by $(f \oplus g)(x, y) = (f(x), g(y))$. Clearly $m \circ (f \oplus g) \circ \Delta = f \cdot g$, and $(f \cdot g)^{-1} = \Delta^{-1} \circ (f \oplus g)^{-1} \circ m^{-1}$.

For open $U \subset \mathbb{R}$, $m^{-1}(U) \subset \mathbb{R}^2$ is open, because m is continuous. Since $\mathbb{R}^2$ is countably based, and in fact has a countable basis consisting of rectangles with rational endpoints, so $m^{-1}(U)$ is a countable unions of rectangles $(a_i, b_i) \times (c_i, d_i)$. Then

$$(f \oplus g)^{-1} \circ m^{-1}(U) \;=\; (f \oplus g)^{-1}(\bigcup_i (a_i, b_i) \times (c_i, d_i))$$

$$= \bigcup_i (f \oplus g)^{-1}((a_i, b_i) \times (c_i, d_i)) \;=\; \bigcup_i f^{-1}(a_i, b_i) \times g^{-1}(c_i, d_i)$$

The sets $f^{-1}(a_i, b_i) \subset \mathbb{R}$ and $g^{-1}(c_i, d_i) \subset \mathbb{R}$ are Borel sets, so their product is a Borel set in $\mathbb{R}^2$. Then

$$\Delta^{-1}(E_1 \times E_2) \;=\; E_1 \cap E_2 \qquad \text{(for } E_1, E_2 \text{ measurable in } \mathbb{R})$$

is measurable. ///

[5.11] Use Urysohn's lemma to prove that $C^o[a,b]$ is dense in $L^1[a,b]$.

Discussion: By the Lebesgue definition of integrals, *simple* functions are dense in $L^1[a,b]$, so it suffices to show that *simple* functions can be well approximated by continuous functions. Granting ourselves the *(outer and inner) regularity* of Lebesgue measure μ, for measurable E there are open U and compact K such that $K \subset E \subset U$, and $\mathfrak{m}(U) - \mu(K) < \varepsilon$. Invoke Urysohn to make a continuous function f taking values in $[0,1]$ and $f|_K = 1$ and $f = 0$ off U. Then

$$\int_a^b |f - \mathrm{ch}_E| \;=\; \int_K |f - \mathrm{ch}_E| + \int_{E-K} |f - \mathrm{ch}_E| + \int_{U-E} |f - \mathrm{ch}_E| \;\le\; \int_K |1-1| + \int_{E-K} 1 + \int_{U-E} 1$$
$$=\; \mu(E-K) + \mu(U-E) \;=\; \mu(U-K) \;<\; \varepsilon$$

as desired. ///

[5.12] Comparing L^p spaces. Let $1 \le p, p' < \infty$. When is $L^p[a,b] \subset L^{p'}[a,b]$ for finite intervals $[a,b]$ and Lebesgue measure? When is $L^p(\mathbb{R}) \subset L^{p'}(\mathbb{R})$? When is $\ell^p \subset \ell^{p'}$?

Discussion: Take $p < p'$. We claim that $L^p[a,b] \supset L^{p'}[a,b]$, with proper containment. The function f that is $(x-a)^{-\frac{1}{p'}}$ on $(a,b]$ and 0 off that interval is *not* in $L^{p'}$, but is in L^p. Given $f \in L^{p'}[a,b]$, let E be the set of $x \in [a,b]$ where $|f(x)| \ge 1$. Then $\int_a^b |f|^{p'} < \infty$ if and only if $\int_E |f|^{p'} < \infty$. On E, $|f|^p < |f|^{p'}$, so $\int_E |f|^p < \infty$, and then also $\int_a^b |f|^p < \infty$, so $f \in L^p[a,b]$. ///

We claim that $L^p(\mathbb{R})$ and $L^{p'}(\mathbb{R})$ are not comparable for $p \ne p'$. Take $1 \le p < p'$. On one hand, $1/(1+|x|)^{1/p'+\varepsilon}$ is in $L^{p'}$ for all $\varepsilon > 0$, but not in L^p for ε small enough so that $\frac{1}{p'} + \varepsilon < \frac{1}{p}$. On the other hand, the function f that is $x^{-\frac{1}{p'}}$ on $(0,1]$ and 0 off that interval is *not* in $L^{p'}$, but is in L^p.

We claim that for $1 \le p < p' < \infty$, $\ell^p \subset \ell^{p'}$, with strict containment. Indeed, $f(n) = 1/n^p$ is not in ℓ^p, but is in $\ell^{p'}$. Let $E = \{n \in \{1,2,\ldots\} : |f(n)| < 1\}$. Then $f \in \ell^p$ if and only if the *complement* of E is finite, and if $\sum_{n\in E} |f(n)|^p < \infty$. Certainly $|f(n)|^p > |f(n)|^{p'}$ for $n \in E$, and the complement of E is finite, so $\sum_{n\in E} |f(n)|^{p'} < \sum_{n\in E} |f(n)|^p$, and $f \in \ell^{p'}$. ///

[5.13] For positive real numbers $w_1, \ldots, w_n$ such that $\sum_i w_i = 1$, and for positive real numbers $a_1, \ldots, a_n$, show that

$$a_1^{w_1} \ldots a_n^{w_n} \;\le\; w_1 a_1 + \ldots + w_n a_n$$

Discussion: This is a corollary of Jensen's inequality, similar to the arithmetic-geometric mean, but with unequal weights. Namely, let $X = \{1, 2, \ldots, n\}$ with measure $\mu(i) = w_i$, and function $f(i) = \log a_i$. Then Jensen's inequality is

$$\Big(\sum_{i=1}^n w_i \cdot \log a_i\Big) \;=\; \sum_{i=1}^n w_i \cdot e^{\log a_i}$$

which simplifies to the assertion. ///

[5.14] In ℓ^2, show that the point in the closed unit ball closest to a point v *not* inside that ball is $v/|v|_{\ell^2}$.

Discussion: The minimum principle assures that there is a *unique* closest point w in the closed unit ball B to v, because B is convex, closed, non-empty, and v is not in B.

Suppose w is closer than $v/v|$. Then

$$|v|^2 - 2|v| + 1 \;=\; |v - \frac{v}{|v|}|^2 \;>\; |v-w|^2 \;=\; |v|^2 - \langle v,w\rangle - \langle w,v\rangle + |w|^2 \;=\; |v|^2 - \langle v,w\rangle - \langle w,v\rangle + 1$$

Thus,

$$2|v| \; < \; \langle v, w\rangle + \langle w, v\rangle$$

Thus, the sum of the two inner products is *positive*, and by Cauchy-Schwarz-Bunyakowsky:

$$2|v| \; < \; \langle v, w\rangle + \langle w, v\rangle \; = \; |\langle v, w\rangle + \langle w, v\rangle| \; \le \; 2|v| \cdot |w|$$

Thus, $1 < |w|$, which is impossible. ///

[5.15] For a measurable set $E \subset [0, 2\pi]$, show that

$$\lim_{n\to\infty} \int_E \cos nx \; dx \; = \; 0 \; = \; \lim_{n\to\infty} \int_E \sin nx \; dx$$

Discussion: This is an instance of a *Riemann-Lebesgue lemma*, namely, that Fourier coefficients of an L^2 function on $[0, 2\pi]$ go to 0. Here, the L^2 function is the characteristic function of E, and we use sines and cosines instead of exponentials. ///

[5.16] One form of the *sawtooth* function is $f(x) = x - \pi$ on $[0, 2\pi]$. Compute the Fourier coefficients $\widehat{f}(n)$. Write out the conclusion of Parseval's theorem for this function.

Discussion: We have the orthonormal basis $e_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}$ with $n \in \mathbb{Z}$ for the Hilbert space $L^2[0, 2\pi]$. The Fourier coefficients are determined by Fourier's formula

$$\widehat{f}(n) \; = \; \int_0^{2\pi} f(x) \; \frac{e^{-inx}}{\sqrt{2\pi}} \; dx$$

For $n = 0$, this is 0. For $n \neq 0$, integrate by parts, to get

$$\widehat{f}(n) \; = \; \Big[f(x) \cdot \frac{e^{-inx}}{\sqrt{2\pi} \cdot (-in)} \Big]_0^{2\pi} - \int_0^{2\pi} 1 \cdot \; \frac{e^{-inx}}{\sqrt{2\pi} \cdot (-in)} \; dx$$

$$= \; \Big((\pi \cdot \frac{1}{\sqrt{2\pi} \cdot (-in)}) - (-\pi \cdot \frac{1}{\sqrt{2\pi} \cdot (-in)}) \Big) - 0 \; = \; \frac{2\pi}{\sqrt{2\pi} \cdot (-in)} \; = \; \frac{\sqrt{2\pi}}{-in}$$

The L^2 norm of f is

$$\int_0^{2\pi} (x - \pi)^2 \; dx \; = \; \Big[\frac{(x-\pi)^3}{3} \Big]_0^{2\pi} \; = \; \frac{\pi^3 - (-\pi)^3}{3} \; = \; \frac{2\pi^3}{3}$$

Thus, by Parseval,

$$\sum_{n \neq 0} \Big| \frac{\sqrt{2\pi}}{-in} \Big|^2 \; = \; \frac{2\pi^3}{3}$$

This simplifies first to

$$2 \sum_{n \ge 1} \frac{2\pi}{n^2} \; = \; \frac{2\pi^3}{3}$$

and then to

$$\sum_{n \ge 1} \frac{1}{n^2} \; = \; \frac{\pi^2}{6}$$

That is, Parseval applied to the sawtooth function evaluates $\zeta(2)$. ///

[5.17] For fixed $y \in [0,1]$, show that there is *no* $f_y \in L^2[0,1]$ so that $\langle g, f_y \rangle = g(y)$ for all $g \in L^2[0,1]$.

Discussion: Part of the issue here is whether L^2 functions truly have meaningful pointwise values at all, and we generally imagine that they do *not*, although such a negative fact may be hard to express formulaically.

Among many approaches, one is to suppose such f exists. Choose an orthonormal basis for $L^2[0,1]$ consisting of the continuous functions $\psi_n(x) = e^{2\pi i n x}$, and see what the condition $\langle f_y, \psi_n \rangle = \psi_n(y)$ imposes on the alleged f_y. Indeed, this condition completely determines the Fourier coefficients of the alleged f_y: since $\psi_n \in L^2[0,1]$, $\langle \psi_n, f_y \rangle = \psi_n(y)$, and then

$$\overline{\widehat{f}_y(n)} \;=\; \int_0^1 \overline{f_y(x)\,\overline{\psi}_n(x)}\,dx \;=\; \langle \psi_n, f_y \rangle \;=\; \psi_n(y)$$

so

$$f_y \;=\; \sum_{n\in\mathbb{Z}} \overline{\psi}_n(y)\cdot\psi_n \qquad\qquad \text{(with equality in an } L^2 \text{ sense)}$$

By Parseval,

$$|f_y|^2_{L^2} \;=\; \sum_n |\psi_n(y)|^2 \;=\; +\infty$$

since $|\psi_n(y)| = 1$ for all n. Thus, there can be no such f_y in L^2. ///

In contrast to the previous example's outcome: Let V be the complex vector space of power series $f(z) = \sum_{n\ge 0} c_n\, z^n$ convergent on the open unit disk D in $\mathbb{C}$, having finite norm

$$|f| \;=\; \Big(\int_D |f(x+iy)|^2\;dx\,dy\Big)^{\frac{1}{2}}$$

with hermitian inner product

$$\langle f, g\rangle \;=\; \int_D f(x+iy)\cdot\overline{g(x+iy)}\;dx\,dy$$

It is not hard to show that $\langle z^m, z^n\rangle = 0$ unless $m = n$, in which case it is $\frac{2\pi}{2n+1}$, and that $\psi_n(z) = z^n \cdot \frac{\sqrt{2n+1}}{\sqrt{2\pi}}$ is an orthonormal basis for V. The sum $f_w(z) = \sum_{n\ge 0}\psi_n(z)\,\overline{\psi_n(w)}$ converges absolutely for $z, w \in D$, and

$$\langle g(-),\, f_w\rangle \;=\; g(w) \qquad\qquad \text{(for } w \text{ in the disk)}$$

For each fixed $w \in D$, pointwise evaluation $g \to g(w)$ is a continuous linear functional on V.

6. Riesz-Markov-Kakutani theorem, Lebesgue measure

6.1 Riesz-Markov-Kakutani theorem and regularity

Let X be a locally compact, Hausdorff, topological space. A map $f \to \lambda(f)$ of continuous, compactly supported functions $C_c^o(X)$ to scalars is *positive* when $\lambda(f) \geq 0$ for $f \in C_c^o(X)$ taking values in $[0, +\infty)$.

[6.1.1] Theorem: *(Riesz, Markov, Kakutani, independently)* Given a positive functional λ on $C_c^o(X)$, there is a σ-algebra A containing all Borel sets, and a positive measure μ on A, such that

$$\lambda(f) \;=\; \int_X f(x)\, d\mu(x) \qquad\qquad (\text{for all } f \in C_c^o(X))$$

• *Outer regularity* holds unconditionally, namely, that for $E \in A$, $\mu(E) \inf_{U \supset E} \mu(U)$ where U ranges over *open* sets containing E.
• *Inner regularity* is conditional: for open E, and for $\mu(E) < \infty$, $\mu(E) = \sup_{K \subset E} \mu(K)$ where K ranges over *compact* sets contained in E.
• μ is *complete*, in the sense that $E' \subset E \in A$ and $\mu(E) = 0$ implies that $E' \in A$.

Proof: (Standard... *[... iou ...]*) ///

With a further mild assumption on the physical space X, including familiar spaces such as $\mathbb{R}^n$, in fact we have unconditional *regularity*:

[6.1.2] Theorem: Suppose further that X is *σ-compact*, meaning that it is a countable union of compact subsets. Then, in the situation of the previous theorem, μ is unconditionally *inner regular*: $\mu(E) = \sup_{K \subset E} \mu(K)$ as K ranges over compacts contained in E. Thus, the measure μ is a *positive, regular, Borel measure.*

Proof: (Standard... *[... iou ...]*) ///

6.2 Lebesgue measure

As a corollary of the Riesz-Markov-Kakutani theorem we have a different description of the Lebesgue measure and integral, as an extension of the Riemann integral, with the very useful side effect of proving inner and outer regularity.

In the Riesz-Markov-Kakutani theorem, take $X = \mathbb{R}^n$, and $\lambda(f)$ to be the usual Riemann integral for $f \in C_c^o(\mathbb{R}^n)$, and let Lebesgue measure be the associated *positive, regular, Borel* measure. With this description of Lebesgue measure, as opposed to the more tangible (but also more awkward) Lebesgue outer measure, we must verify that all the expected properties do hold.

[6.2.1] Corollary: Let μ be Lebesgue measure, induced by the Riesz-Markov-Kakutani theorem from the Riemann integral on $C_c^o(\mathbb{R}^n)$.
• μ is *translation-invariant* in the sense that $\mu(E+x) = \mu(E)$ for all $x \in \mathbb{R}^n$.
• The Lebesgue measure of a cube $(a_1, b_1) \times \ldots \times (a_n, b_n)$ is the product $\prod_i |b_i - a_i|$, and similarly for closed and half-open intervals and their products.

Proof: (Standard... *[... iou ...]*) ///

7. Product measures and Fubini-Tonelli theorem

1. Product measures
2. Fubini-Tonelli theorem(s)
3. Completions of measures

7.1 Product measures, completions of measures

Let X, μ and Y, ν be measure spaces with corresponding σ-algebras A, B. Assume X and Y are σ*-finite*, in the sense that they are countable unions of finite-measure sets.

First, the *product* σ-algebra is the σ-algebra in $X \times Y$ generated by all products $E \times F$ with $E \in A$ and $F \in B$.

For iterated integrals to make sense, we need to check a few things. For $E \in A \times B$, for $x \in X$ and $y \in Y$, let

$$E_x \;=\; \{y \in Y : (x,y) \in E\} \qquad \text{and} \qquad E^y \;=\; \{x \in X : (x,y) \in E\}$$

As a consistency check, we have

[7.1.1] Theorem: For $E \in A \times B$, for $x \in X$ and $y \in Y$, $E_x \in A$ and $E^y \in B$. The function $x \to \nu(E_x)$ is μ-measurable, $y \to \mu(E^y)$ is ν-measurable, and

$$\int_X \nu(E_x)\, d\mu(x) \;=\; \int_Y \mu(E^y)\, d\nu(y)$$

Proof: [... iou ...] ///

Then the *product measure* $\mu \times \nu$ can be defined in the expected fashion: for $E \in A \times B$,

$$(\mu \times \nu)(E) \;=\; \int_X \nu(E_x)\, d\mu(x) \;=\; \int_Y \mu(E^y)\, d\nu(y)$$

7.2 Fubini-Tonelli theorem(s)

Let X, μ and Y, ν be measure spaces with corresponding σ-algebras A, B. Assume X and Y are σ*-finite*.

[7.2.1] Theorem: *(Fubini-Tonelli)* For complex-valued measurable f, g, if any one of

$$\int_X \int_Y |f(x,y)|\, d\mu(x)\, d\nu(y) \qquad \int_Y \int_X |f(x,y)|\, d\nu(y)\, d\mu(x) \qquad \int_{X\times Y} |f(x,y)|\, d\mu{\times}\nu$$

is finite, then they *all* are finite, and are equal. For $[0, +\infty]$-valued functions f,

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

although the values may be $+\infty$.

Proof: [... iou ...] ///

To explain what the *product measure* $\mu \times \nu$ should be, and also for a proof of the theorem, we need the notion of *monotone class*. A monotone class in a set X is a set $\mathcal{M}$ of subsets of X closed under countable ascending unions and under countable descending intersections. That is, if

$$M_1 \subset M_2 \subset M_3 \subset \ldots$$

$$N_1 \supset N_2 \supset N_3 \supset \ldots$$

are collections of sets in $\mathcal{M}$, then

$$\bigcup_i M_i \qquad \bigcap_i N_i$$

both lie in $\mathcal{M}$, as well. Another characterization of $\mathcal{A} \times \mathcal{B}$ is that it is the smallest monotone class containing all products $E \times F$ with $E \in \mathcal{A}$ and $F \in \mathcal{B}$.

Let f be a $\mathcal{A} \times \mathcal{B}$-measurable function on $X \times Y$. (Note that this does not depend upon having a 'product measure', but only upon the sigma-algebra!) Then all the functions

$$x \to f(x,y) \qquad \text{(for fixed } y \in Y$$

$$y \to f(x,y) \qquad \text{(for fixed } x \in X$$

are measurable (in appropriate senses). In particular, we could apply this to the *characteristic function* of a set $G \in \mathcal{A} \times \mathcal{B}$.

Now we come to the point where the sigma-finiteness of X and Y is necessary. For $G \in \mathcal{A} \times \mathcal{B}$, let

$$f(x) = \nu(G_x) \qquad g(y) = \mu(G_y)$$

where G_x, G_y are as above. We have already noted that f, g are *measurable*. Further,

$$\int_X f(x)\; d\mu(x) = \int_Y g(y)\; d\nu(y)$$

This is proven by showing that the collection of G for which the conclusion is true is a *monotone class* containing all products $E \times F$.

In light of the latter equality, we can define the *product measure* $\mu \times \nu$ on $G \in \mathcal{A} \times \mathcal{B}$ by

$$(\mu \times \nu)(G) = \int_X f(x)\; d\mu(x) = \int_Y g(y)\; d\nu(y)$$

with notation as just above. The *countable additivity* follows from a preliminary version of Fubini's theorem, namely that if f_i are countably-many $[0, +\infty]$-valued functions on a measure space Ω, then

$$\int_\Omega \sum_i f_i = \sum_i \int_\Omega f_i$$

which itself is a little corollary of the monotone convergence theorem.

sectionCompletions of measures

For example, a reasonable measure on $\mathbb{R}^m \times \mathbb{R}^n$ should include many sets not expressible as countable unions of products $E \times F$ where $E \subset \mathbb{R}^m$ and $F \subset \mathbb{R}^n$. For example, diagonal subsets of the form $D = \{(x,x) : 0 \le x \le 1\} \subset \mathbb{R}^2$ are not countable unions of products, but should surely be measurable.

One way to accomplish this is by *completion* of the product measure.

Then the *completion* of $\mu \times \nu$ further assigns measure 0 to *any* subset S of $T \in A \times B$ with $(\mu \times \nu)(T) = 0$, and adjoins all such sets to the σ-algebra $A \times B$.

[7.2.2] Claim: Lebesgue measure on $\mathbb{R}^m \times \mathbb{R}^n$ is the completion of the product of Lebesgue measures on $\mathbb{R}^m$ and $\mathbb{R}^n$.

Proof: [... iou ...] ///

Completing a product measure is usually what we want, but it slightly complicates the statement of the corresponding Fubini-Tonelli theorem:

[7.2.3] Theorem: Let X, A, μ and Y, B, ν be *complete* measure spaces, with X, Y σ-finite. Let f be a function on $X \times Y$ measurable with respect to the *completion* of the product measure. Then $x \to f(x, y)$ and $y \to f(x, y)$ are μ-measurable and ν-measurable (only) *almost everywhere*.

Proof: [... iou ...] ///

[7.2.4] Remark: To be precise, *completeness* is a property of σ-algebras, not of measures.

8. Hilbert spaces

Hilbert spaces are possibly-infinite-dimensional analogues of the familiar finite-dimensional Euclidean spaces. In particular, Hilbert spaces have *inner products*, so notions of *perpendicularity* (or *orthogonality*), and *orthogonal projection* are available. Reasonably enough, in the infinite-dimensional case we must be careful not to extrapolate too far based only on the finite-dimensional case.

Unfortunately, few naturally-occurring spaces of functions are Hilbert spaces. Given the intuitive geometry of Hilbert spaces, this is disappointing, suggesting that physical intuition is a little distant from the behavior of natural spaces of functions. However, a little later we will see that suitable *families* of Hilbert spaces *can* capture what we want. Such ideas were developed by Beppo Levi (1906), Frobenius (1907), and Sobolev (1930's). These ideas do fit into Schwartz' (c. 1950) formulation of his notion of *distributions*, but it seems that they were not explicitly incorporated, or perhaps were viewed as completely obvious at that point. We will see that Levi-Sobolev ideas offer some useful specifics in addition to Schwartz' over-arching ideas.

Most of the geometric results on Hilbert spaces are corollaries of the *minimum principle*.

Most of what is done here applies to vector spaces over either $\mathbb{R}$ or $\mathbb{C}$.

8.1 Cauchy-Schwarz-Bunyakowsky inequality

A complex vector space V with a complex-valued function

$$\langle , \rangle : V \times V \to\to \mathbb{C}$$

of two variables on V is a *(hermitian) inner product space* or *pre-Hilbert space*, and $\langle , \rangle$ is a *(hermitian) inner product*, when we have the usual conditions

$$\begin{cases} \langle x, y\rangle & = & \overline{\langle y, x\rangle} & \text{(the } \textit{hermitian-symmetric} \text{ property)} \\ \langle x+x', y\rangle & = & \langle x, y\rangle + \langle x', y\rangle & (\textit{additivity in} \text{ first argument}) \\ \langle x, y+y'\rangle & = & \langle x, y\rangle + \langle x, y'\rangle & (\textit{additivity} \text{ in second argument}) \\ \langle x, x\rangle & \geq & 0 & \text{(and equality only for } x=0\text{: } \textit{positivity}) \\ \langle \alpha x, y\rangle & = & \alpha\langle x, y\rangle & (\textit{linearity} \text{ in first argument}) \\ \langle x, \alpha y\rangle & = & \bar{\alpha}\langle x, y\rangle & (\textit{conjugate-linearity} \text{ in second argument}) \end{cases}$$

Among other easy consequences of these requirements, for all $x, y \in V$

$$\langle x, 0\rangle \;=\; \langle 0, y\rangle \;=\; 0$$

where inside the angle-brackets the 0 is the zero-*vector*, and outside it is the zero-*scalar*.

The *associated norm* $|\,|$ on V is defined by

$$|x| \;=\; \langle x, x\rangle^{1/2}$$

with the non-negative square-root. Even though we use the same notation for the norm on V as for the usual complex value $||$, context will make clear which is meant. The *metric* on a Hilbert space is $d(v,w) = |v-w|$: the triangle inequality follows from the *Cauchy-Schwarz-Bunyakowsky inequality* just below.

For two vectors v, w in a pre-Hilbert space, if $\langle v, w\rangle = 0$ then v, w are *orthogonal* or *perpendicular*, sometimes written $v \perp w$. A vector v is a *unit vector* if $|v| = 1$.

There are several essential algebraic identities, variously and ambiguously called *polarization identities*. First, there is

$$|x+y|^2 + |x-y|^2 \;=\; 2|x|^2 + 2|y|^2$$

which is obtained simply by expanding the left-hand side and cancelling where opposite signs appear. In a similar vein,

$$|x+y|^2 - |x-y|^2 \;=\; 2\langle x,y\rangle + 2\langle y,x\rangle \;=\; 4\mathrm{Re}\,\langle x,y\rangle$$

Therefore,

$$(|x+y|^2 - |x-y|^2) + i(|x+iy|^2 - |x-iy|^2) \;=\; 4\langle x,y\rangle$$

These and closely-related identites are of frequent use.

[8.1.1] Theorem: *(Cauchy-Schwarz-Bunyakowsky inequality)*

$$|\langle x,y\rangle| \;\leq\; |x|\cdot|y|$$

with strict inequality unless x, y are collinear, i.e., unless one of x, y is a multiple of the other.

Proof: Suppose that x is not a scalar multiple of y, and that neither x nor y is 0. Then $x - \alpha y$ is not 0 for any complex α. Consider

$$0 \;<\; |x-\alpha y|^2$$

We know that the inequality is indeed *strict* for all α since x is not a multiple of y. Expanding this,

$$0 \;<\; |x|^2 - \alpha\langle x,y\rangle - \bar{\alpha}\langle y,x\rangle + \alpha\bar{\alpha}|y|^2$$

Let

$$\alpha \;=\; \mu t$$

with real t and with $|\mu| = 1$ so that

$$\mu\langle x,y\rangle \;=\; |\langle x,y\rangle|$$

Then

$$0 \;<\; |x|^2 - 2t|\langle x,y\rangle| + t^2|y|^2$$

The *minimum* of the right-hand side, viewed as a function of the real variable t, occurs when the derivative vanishes, i.e., when

$$0 \;=\; -2|\langle x,y\rangle| + 2t|y|^2$$

Using this minimization as a *mnemonic* for the value of t to substitute, we indeed substitute

$$t \;=\; \frac{|\langle x,y\rangle|}{|y|^2}$$

into the inequality to obtain

$$0 \;<\; |x|^2 + \left(\frac{|\langle x,y\rangle|}{|y|^2}\right)^2 \cdot |y|^2 - 2\frac{|\langle x,y\rangle|}{|y|^2}\cdot|\langle x,y\rangle|$$

which simplifies to

$$|\langle x,y\rangle|^2 \;<\; |x|^2\cdot|y|^2$$

as desired. ///

[8.1.2] Corollary: *(Triangle inequality)* For v, w in a Hilbert space V, we have $|v+w| \le |v| + |w|$. Thus, with distance function $d(v,w) = |v-w|$, we have the triangle inequality

$$d(x,z) \;=\; |x-z| \;=\; |(x-y)+(y-z)| \;\le\; |x-y| + |y-z| \;=\; d(x,y) + d(y,z)$$

Proof: Squaring and expanding, noting that $\langle v,w\rangle + \langle w,v\rangle = 2\mathrm{Re}\,\langle v,w\rangle$,

$$(|v|+|w|)^2 - |v+w|^2 \;=\; \Big(|v|^2 + 2|v|\cdot|w| + |w|^2\Big) - \Big(|v|^2 + \langle v,w\rangle + \langle w,v\rangle + |w|^2\Big) \;\ge\; 2|v|\cdot|w| - 2|\langle v,w\rangle| \;\ge\; 0$$

giving the asserted inequality. ///

An inner product space *complete* with respect to the metric arising from its inner product (and norm) is a *Hilbert space*.

[8.1.3] Continuity issues

The map

$$\langle,\rangle : V \times V \longrightarrow \mathbb{C}$$

is *continuous* as a function of two variables. Indeed, suppose that $|x-x'| < \varepsilon$ and $|y-y'| < \varepsilon$ for $x, x', y, y' \in V$. Then

$$\langle x,y\rangle - \langle x',y'\rangle \;=\; \langle x-x',y\rangle + \langle x',y\rangle - \langle x',y'\rangle \;=\; \langle x-x',y\rangle + \langle x',y-y'\rangle$$

Using the triangle inequality for the ordinary absolute value, and then the Cauchy-Schwarz-Bunyakowsky inequality, we obtain

$$|\langle x,y\rangle - \langle x',y'\rangle| \;\le\; |\langle x-x',y\rangle| + |\langle x',y-y'\rangle| \;\le\; |x-x'||y| + |x'||y-y'|$$

$$<\; \varepsilon(|y| + |x'|)$$

This proves the continuity of the inner product.

Further, scalar multiplication and vector addition are readily seen to be continuous. In particular, it is easy to check that for any fixed $y \in V$ and for any fixed $\lambda \in \mathbb{C}^\times$ both maps

$$x \to x + y$$

$$x \to \lambda x$$

are *homeomorphisms* of V to itself.

8.2 Example: ℓ^2

Before further abstract discussion, we note that, up to isomorphism, there is essentially just one *infinite-dimensional* Hilbert space occurring in practice, namely the space ℓ^2 constructed as follows. Most infinite-dimensional Hilbert spaces occurring in practice have a countable dense subset, because the Hilbert spaces are completions of spaces of continuous functions on topological spaces with a countably-based topology.

Lest anyone be fooled, often subtlety is in the description of the isomorphisms and *mappings* among such Hilbert spaces.

Let ℓ^2 be the collection of sequences $f = \{f(i) : 1 \le i < \infty\}$ of complex numbers meeting the constraint

$$\sum_{i=1}^{\infty} |f(i)|^2 < +\infty$$

For two such sequences f and g, the *inner product* is

$$\langle f, g\rangle = \sum_i f(i)\overline{g(i)}$$

[8.2.1] Claim: ℓ^2 is a vector space. The sum defining the inner product on ℓ^2 is absolutely convergent.

Proof: That ℓ^2 is closed under scalar multiplication is clear. For $f, g \in \ell^2$, by Cauchy-Schwarz-Bunyakowsky,

$$\Big|\sum_{n\le N} f(i)\cdot\overline{g(i)}\Big| \le \sum_{n\le N} |f(i)|\cdot|g(i)| \le \Big|\sum_{n\le N}|f(i)|^2\Big|^{\frac{1}{2}}\cdot\Big|\sum_{n\le N}|g(i)|^2\Big|^{\frac{1}{2}} \le |f|_{\ell^2}\cdot|g|_{\ell^2}$$

giving the absolute convergence of the infinite sum for $\langle f, g\rangle$. Then, expanding,

$$\sum_{n\le N}|f(i)+g(i)|^2 \le \sum_{n\le N}|f(i)|^2 + 2|f(i)|\cdot|g(i)| + |g(i)|^2 < +\infty$$

by the previous. ///

[8.2.2] Claim: ℓ^2 is *complete*.

Proof: Let $\{f_n\}$ be a Cauchy sequence of elements in ℓ^2. For every $i \in \{1, 2, 3, \ldots\}$,

$$|f_m(i) - f_n(i)|^2 \le \sum_{i\ge1}|f_m(i) - f_n(i)|^2 = |f_m - f_n|^2_{\ell^2}$$

so $f(i) = \lim_n f_n(i)$ exists for every i, and is the obvious candidate for the limit in ℓ^2. It remains to see that this limit is indeed in ℓ^2. This will follow from an easy case of Fatou's lemma:

$$\sum_i |f(i)|^2 = \sum_i |\lim_n f_n(i)| = \sum_i |\liminf_n f_n(i)| \le \liminf_n \sum_i |f_n(i)| = \liminf_n |f_n|^2_{\ell^2}$$

Since $\{f_n\}$ is a Cauchy sequence, certainly $\lim_n |f_n|^2_{\ell^2}$ exists. ///

[8.2.3] Remark: A similar result holds for $L^2(X, \mu)$ for general measure spaces X, μ, but requires more preparation.

8.3 Completions, infinite sums

An arbitrary pre-Hilbert space can be *completed* as metric space, giving a *Hilbert space*. Since metric spaces have *countable local bases* for their topology (e.g., open balls of radii $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \ldots$) all points in the completion are limits of Cauchy *sequences* (rather than being limits of more complicated Cauchy *nets*). The completion inherits an inner product defined by a limiting process

$$\langle \lim_m x_m, \lim_n y_n \rangle \;=\; \lim_{m,n} \langle x_m, y_n \rangle$$

It is not hard to verify that the indicated limit exists (for Cauchy sequences $\{x_m\}, \{y_n\}$), and gives a hermitian inner product on the completion. The completion process *does nothing* to a space which is already complete.

In a Hilbert space, we can consider *infinite* sums

$$\sum_{\alpha \in A} v_\alpha$$

for sets $\{v_\alpha : \alpha \in A\}$ of vectors in V. Not wishing to have a notation that only treats sums indexed by $1, 2, 3, \ldots$, we can consider the *directed system* $\mathcal{A}$ of all finite subsets of A. Consider the *net* of *finite partial sums* of $\sum v_\alpha$ indexed by $\mathcal{A}$ by

$$s(A_o) \;=\; \sum_{\alpha \in A_o} v_\alpha$$

where $A_o \in \mathcal{A}$. This is a *Cauchy net* if, given $\varepsilon > 0$, there is a finite subset A_o of A so that for any two finite subsets A_1, A_2 of A both containing A_o we have

$$|s(A_1) - s(A_2)| \;<\; \varepsilon$$

If the net is Cauchy, then by the *completeness* there is a unique $v \in V$, the *limit of the Cauchy net*, so that for all $\varepsilon > 0$ there is a finite subset A_o of A so that for any finite subset A_1 of A containing A_o we have

$$|s(A_1) - v| \;<\; \varepsilon$$

8.4 Minimum principle, orthogonality

This fundamental minimum principle, that *a non-empty closed convex*[12] *set in a Hilbert space has a unique element of least norm*, is essential in the sequel. It generally *fails* in more general types of topological vector spaces.

[8.4.1] Theorem: A non-empty closed convex subset of a HIlbert space has a unique element of least norm.

Proof: For two elements x, y in a closed convex set C inside a Hilbert space with both $|x|$ and $|y|$ within $\varepsilon > 0$ of the infimum μ of the norms of elements of C,

$$|x-y|^2 \;=\; 2|x|^2+2|y|^2-|x+y|^2 \;=\; 2|x|^2+2|y|^2-4\Big(\frac{|x+y|}{2}\Big)^2 \;\le\; 2(\mu+\varepsilon)^2+2(\mu+\varepsilon)^2-4\mu^2 \;=\; \varepsilon\cdot(8\mu+4\varepsilon)$$

since $\frac{x+y}{2} \in C$ by convexity of C. Thus, any sequence (or *net*) in C with norms approaching the inf is a Cauchy sequence (*net*). Since C is closed, such a sequence converges to an element of C. Further, the

[12] Recall that a set C in a vector space is *convex* when $tx + (1-t)y \in C$ for all $x, y \in C$ and for all $0 \le t \le 1$.

inequality shows that any two Cauchy sequences (or *nets*) converging to points minimizing the norm on C have the same limit. Thus, the minimizing point is *unique*. ///

[8.4.2] Corollary: Given a closed, convex, non-empty subset E of a Hilbert space V, and a point $v \in V$ not in E, there is a unique point $w \in E$ closest to v.

Proof: Since $v \notin E$, the set $E - v$ does not contain 0. The map $x \to x - v$ is a homeomorphism, because non-zero scalar multiplication and vector addition are continuous, and have continuous inverses. Thus, $E - v$ is closed. It is also still convex. Thus, there is a unique element $x_o - v \in E - v$ of smallest norm. That is, $|x_o - v| < |x - v|$ for all $x \neq x_o$ in E. That is, the distance from x_o is the minimum. ///

[8.4.3] Orthogonal projections to closed subspaces Existence of orthogonal projections makes essential use of the minimization principle. Let W be a complex vector subspace of a Hilbert space V. If W is *closed* in the topology on V then, reasonably enough, we say that W is a *closed subspace*. For an arbitrary complex vector subspace W of a Hilbert space V, the topological closure $\bar{W}$ is readily checked to be a complex vector subspace of V, so is a *closed subspace*. Because it is necessarily complete, *a closed subspace of a Hilbert space is a Hilbert space in its own right.*

Let W be a closed subspace of a Hilbert space V. Let $v \in V$. From the corollary just above, the closed convex subset W contains a unique element w_o closest to v.

[8.4.4] Claim: The element w_o is the *orthogonal projection* of v to W, in the sense that $w_o \in W$ is the unique element in W such that $\langle v - w_o, w\rangle = 0$ for all $w \in W$.

Proof: For two vectors $w_1, w_2 \in W$ so that

$$\langle v - w_i, w\rangle \;=\; 0 \qquad\qquad (\text{for both } i = 1, 2 \text{ and for all } w \in W)$$

by subtracting, we would have

$$\langle w_1 - w_2, w\rangle \;=\; 0$$

for all $w \in W$. In particular, with $w = w_1 - w_2$, necessarily $w_1 - w_2 = 0$, proving *uniqueness* of the orthogonal projection.

With w_o the unique element of W closest to v, for any $w \in W$, since $w_o + w$ is still in W,

$$|v - w_o|^2 < |v - (w_o + w)|^2$$

Expanded slightly, this is

$$|v - w_o|^2 \;\le\; |v - w_o|^2 - \langle v - w_o, w\rangle - \langle w, v - w_o\rangle + |w|^2$$

which gives

$$\langle v - w_o, w\rangle + \langle w, v - w_o\rangle \;\le\; |w|^2$$

Replacing w by μw with μ a complex number with $|\mu| = 1$ and

$$\langle v - w_o, \mu w\rangle \;=\; |\langle v - w_o, w\rangle|$$

this gives

$$2|\langle v - w_o, w\rangle| \;\le\; |w|^2 \qquad\qquad (\text{for } w \neq 0)$$

Replacing w by tw with $t > 0$ gives

$$2t|\langle v - w_o, w\rangle| \;\le\; t^2|w|^2$$

Dividing by t and letting $t \to 0^+$, this gives

$$|\langle v - w_o, w\rangle| \;\le\; 0$$

as required. ///

[8.4.5] Orthogonal complements $W^\perp$

Let W be a vector subspace of a Hilbert space V. The *orthogonal complement* $W^\perp$ of W is

$$W^\perp = \{v \in V : \langle v, w\rangle = 0, \quad \forall w \in W\}$$

It is easy to check that $W^\perp$ is a complex vector subspace of V. Since for each $w \in W$ the set

$$w^\perp = \{v \in V : \langle v, w\rangle = 0\}$$

is the inverse image of the closed set $\{0\}$ of $\mathbb{C}$ under the continuous map

$$v \to \langle v, w\rangle$$

it is *closed*. Thus, the orthogonal complement $W^\perp$ is the intersection of a family of closed sets, so is *closed*.

One point here is that *if the topological closure $\bar{W}$ of W in a Hilbert space V is properly smaller than V then $W^\perp \neq \{0\}$*. Indeed, if $\bar{W} \neq V$ then we can find $y \notin \bar{W}$. Let py be the orthogonal projection of y to $\bar{W}$. Then $y_o = y - py$ is non-zero and is orthogonal to W, so is orthogonal to $\bar{W}$, by continuity of the inner product. Thus, as claimed, $W^\perp \neq \{0\}$.

As a corollary, for any complex vector subspace W of the Hilbert space V, *the topological closure of W in V is the subspace*

$$\bar{W} = W^{\perp\perp}$$

One direction of containment, namely that

$$\bar{W} \subset W^{\perp\perp}$$

is easy: it is immediate that $W \subset W^{\perp\perp}$, and then since the latter is closed we get the asserted containment. If $W^{\perp\perp}$ were strictly larger than $\bar{W}$, then there would be y in it not lying in $\bar{W}$. Now $W^{\perp\perp}$ is a Hilbert space in its own right, in which $\bar{W}$ is a closed subspace, so the orthogonal complement of $\bar{W}$ in $W^{\perp\perp}$ contains a non-zero element z, from above. But then $z \in W^\perp$, and certainly

$$W^\perp \cap (W^\perp)^\perp = \{0\}$$

contradiction. ///

[8.4.6] Orthonormal sets, separability

A set $\{e_\alpha : \alpha \in A\}$ in a pre-Hilbert space V is *orthogonal* if

$$\langle e_\alpha, e_\beta\rangle = 0 \qquad \text{(for all } \alpha \neq \beta)$$

When

$$|e_\alpha| = 1$$

for all indices the set is *orthonormal.* An orthogonal set of non-zero vectors is turned into an ortho*normal* one by replacing each e_α by $e_\alpha/|e_\alpha|$.

We claim that not only are the elements e_α in an orthonormal set *linearly independent* in the usual purely algebraic sense, but, further, in a convergent infinite sum $\sum_{\alpha\in A} c_\alpha e_\alpha$ with complex c_α with

$$\sum_\alpha c_\alpha e_\alpha = 0$$

then all coefficients c_α are 0. Indeed, given $\varepsilon > 0$ take a large-enough finite subset A_o of A so that for any finite subset $A_1 \supset A_o$

$$|\sum_{\alpha\in A_1} c_\alpha e_\alpha| < \varepsilon$$

For any particular index β we may as well enlarge A_1 to include β, and by Cauchy-Schwarz-Bunyakowsky.

$$\Big|\langle \sum_{\alpha\in A_1} c_\alpha e_\alpha, e_\beta\rangle\Big| \;\le\; \Big|\sum_{\alpha\in A_1} c_\alpha e_\alpha\Big|\cdot|e_\beta| \;<\; \varepsilon\cdot|e_\beta| \;=\; \varepsilon$$

On the other hand, using the orthonormality,

$$\Big|\langle \sum_{\alpha\in A_1} c_\alpha e_\alpha, e_\beta\rangle\Big| \;=\; |c_\beta|\cdot|e_\beta|^2 \;=\; |c_\beta|$$

Together, $|c_\beta| \;<\; \varepsilon$. This holds for all $\varepsilon > 0$, so $c_\beta = 0$. This holds for all indices β. ///

A *maximal* orthonormal set in a pre-Hilbert space is called an *orthonormal basis.* The property of maximality of an orthonormal set $\{e_\alpha : \alpha\in A\}$ is the natural one, that there be no *other* unit vector e perpendicular to all the e_α.

Let $\{e_\alpha : \alpha\in A\}$ be an orthonormal set in a Hilbert space V. Let W_o be the complex vector space of all finite linear combinations of vectors in $\{e_\alpha : \alpha\in A\}$. Then we claim that $\{e_\alpha : \alpha\in A\}$ *is an orthonormal basis if and only if* W_o *is dense in* V. Indeed, if the closure W of W_o were a proper subspace of V, then it would have a non-trivial orthogonal complement, so we could make a further unit vector, so $\{e_\alpha : \alpha\in A\}$ could not have been maximal. On the other hand, if $\{e_\alpha : \alpha\in A\}$ is not maximal, let e be a unit vector orthogonal to all the e_α. Then e is orthogonal to all finite linear combinations of the e_α, so is orthogonal to W_o, and thus to W by continuity. That is, W_o cannot be dense. ///

Next, we show that *any orthonormal set can be enlarged to be an orthonormal basis.* To prove this requires invocation of an equivalent of the Axiom of Choice. Specifically, we want to *order* the collection X of orthonormal sets (containing the given one) by inclusion, and note that any *totally ordered* collection of orthonormal sets in X has a supremum, namely the union of all. Thus, we are entitled to conclude that there *are* maximal orthonormal sets containing the given one. If such a maximal one were not an orthonormal basis, then (as observed just above) we could find a further unit vector orthogonal to all vectors in the orthonormal set, contradicting the maximality within X. ///

If a Hilbert space has a *countable* orthonormal basis, then it is called *separable.* Most Hilbert spaces of practical interest are separable, but at the same time most elementary results do not make any essential use of separability so there is no compulsion to worry about this at the moment.

8.5 Bessel inequality, Parseval isomorphism

Let $\{e_\alpha : \alpha\in A\}$ be an orthonormal basis in a Hilbert space V. *Granting* for the moment that $v\in V$ has an expression

$$v = \sum_\alpha c_\alpha e_\alpha$$

we can determine the coefficients c_α, as follows. By the continuity of the inner product, this equality yields

$$\langle v, e_\beta\rangle \;=\; \langle \sum_\alpha c_\alpha e_\alpha, e_\beta\rangle \;=\; \sum_\alpha c_\alpha \langle e_\alpha, e_\beta\rangle \;=\; c_\beta$$

An expression

$$v \;=\; \sum_\alpha c_\alpha\, e_\alpha \;=\; \sum_\alpha \langle v, e_\alpha\rangle\, e_\alpha$$

is an *abstract Fourier expansion.* The coefficients $c_\alpha = \langle v, e_\alpha\rangle$ are the (abstract) *Fourier coefficients* in terms of the orthonormal basis. When the orthonormal basis $\{e_\alpha : \alpha\in A\}$ is understood, we may write $\widehat{v}(\alpha)$ for $\langle v, e_\alpha\rangle$.

[8.5.1] Remark: We have not quite proven that every vector *has* such an expression. We do so after proving a necessary preparatory result.

[8.5.2] Claim: *(Bessel's inequality)* Let $\{e_\beta : \beta \in B\}$ be an orthonormal set in a Hilbert space V. Then

$$|v|^2 \;\geq\; \sum_{\beta \in B} |\langle v, e_\beta\rangle|^2$$

Proof: Just using the positivity (and continuity) and orthonormality

$$0 \;\leq\; |v - \sum_{\beta \in B} \langle v, e_\beta\rangle\, e_\beta|^2 \;=\; |v|^2 - \sum_{\beta \in B} \langle v, e_\beta\rangle \overline{\langle v, e_\beta\rangle} - \sum_{\beta \in B} \overline{\langle v, e_\beta\rangle}\langle v, e_\beta\rangle + \sum_{\beta \in B} |\langle v, e_\beta\rangle|^2 \;=\; |v|^2 - \sum_{\beta \in B} |\langle v, e_\beta\rangle|^2$$

This gives the desired inequality. ///

[8.5.3] Claim: Every vector $v \in V$ has a unique expression as

$$v = \sum_{\alpha \in A} c_\alpha e_\alpha$$

More precisely, for $v \in V$ and for each finite subset B of A let

$$v_B = \text{projection of } v \text{ to } \textstyle\sum_{\alpha \in B} \mathbb{C} \cdot e_\alpha \;=\; \displaystyle\sum_{\alpha \in B} \langle v, e_\alpha\rangle\, e_\alpha$$

Then the net

$$\{v_B : B \text{ finite}, B \subset A\}$$

is Cauchy and has limit v.

Proof: Uniqueness follows from the previous discussion of the density of the subspace V_o of finite linear combinations of the e_α.

Bessel's inequality

$$|v|^2 \geq \sum_{\alpha \in B} |\langle v, e_\alpha\rangle|^2$$

implies that the net is Cauchy, since the tails of a convergent sum must go to 0. Let w be the limit of this net. Given $\varepsilon > 0$, let B be a large enough finite subset of A such that for finite subset $C \supset B$ $|w - v_C| < \varepsilon$. Given $\alpha \in A$ enlarge B if necessary so that $\alpha \in B$. Then

$$|\langle v - w, e_\alpha\rangle| \;\leq\; |\langle v - v_B, e_\alpha\rangle| + |\langle w - v_B, e_\alpha\rangle| \;\leq\; 0 + |w - v_B| \;<\; \varepsilon$$

since $\langle v - v_B, e_\alpha\rangle = 0$ for $\alpha \in B$. Thus, if $v \neq w$, we can construct a further vector of length 1 orthogonal to all the e_α, namely a unit vector in the direction of $v - w$. This would contradict the maximality of the collection of e_α. ///

[8.5.4] Remark: If V were only a pre-Hilbert space, that is, were not complete, then a maximal collection of mutually orthogonal vectors of length 1 may not have the property of the theorem. That is, the collection of (finite) linear combinations may fail to be dense. This is visible in the proof above, wherein we needed to be able to take the limit that yielded the auxiliary vector w. For example, inside the standard ℓ^2 let $e_1, e_2, \ldots$ be the usual $e_1 = (1, 0, 0, \ldots)$, $e_1 = (0, 1, 0, \ldots)$, (etc.) and let

$$v_1 = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \ldots)$$

Let V be pre-Hilbert space which is the (algebraic) span of $v_1, e_2, e_3, \ldots$ Certainly

$$B = \{e_2, e_3, \ldots\}$$

is an orthonormal set. In fact, this collection is a *maximal* orthonormal set in V, but v_1 is *not* in the closure of the span of B.

For $v \in V$, write

$$\hat{v} = \langle v, e_\alpha \rangle$$

[8.5.5] Corollary: *(Parseval isomorphism)* With orthonormal basis $\{e_\alpha : \alpha \in A\}$, the map $v \to \hat{v}$ is *an isomorphism of Hilbert spaces* $V \to \ell^2(A)$. That is, the map is an isomorphism of complex vector spaces, is a homeomorphism of topological spaces, and

$$\langle v, w \rangle = \langle \hat{v}, \hat{w} \rangle \qquad |v|^2 = |\hat{v}|^2_{\ell^2(A)}$$

where the inner product on the left is that in V and the inner product on the right is that in $\ell^2(A)$.) That is,

$$|v|^2 = \sum_{\alpha \in A} |\langle v, e_\alpha \rangle|^2$$

Proof: Expand any vector v in terms of the given orthonormal basis as

$$v = \sum_\alpha \hat{v}(\alpha)\, e_\alpha = \sum_\alpha \langle v, e_\alpha \rangle\, e_\alpha$$

The assertion that $\langle v, w \rangle = \langle \hat{v}, \hat{w} \rangle$ is a consequence of the expansion in terms of the orthonormal basis, together with continuity. That $\hat{v}$ lies in $\ell^2(A)$, and in fact has norm equal to that of v, is the assertion of Parseval.

The only thing of any note is the point that any $\{c_\alpha\} \in \ell^2(A)$ can actually occur as the (abstract) Fourier coefficients of some vector in V. That is, for $f \in \ell^2(A)$, we want to show that the net of finite sums

$$\sum_{\alpha \in A_o} f(\alpha) e_\alpha$$

(for A_o a finite subset of A) is *Cauchy.* Since $f \in \ell^2(A)$, for given $\varepsilon > 0$ there is large-enough finite A_o so that

$$\left(\sum_{\alpha \in A - A_o} |f(\alpha)|^2 \right)^{1/2} = \Big| \sum_{\alpha \in A - A_o} f(\alpha) e_\alpha \Big| < \varepsilon$$

(using the orthonormality). Then for A_1, A_2 both containing A_o,

$$\Big| \sum_{\alpha \in A_1} f(\alpha) e_\alpha - \sum_{\alpha \in A_2} f(\alpha) e_\alpha \Big|^2 = \sum_{\alpha \in (A_1 \cup A_2) - A_o} |f(\alpha) e_\alpha|^2 \le \sum_{\alpha \in A - A_o} |f(\alpha)|^2 < \varepsilon^2$$

From this the Cauchy property follows. ///

8.6 Riemann-Lebesgue lemma

The result of this section is an essentially trivial consequence of previous observations, and is certainly much simpler to prove than the genuine Riemann-Lebesgue lemma for Fourier *transforms.*

Let $\{e_\alpha : \alpha \in A\}$ be an orthonormal basis for a Hilbert space V. For $v \in V$, write

$$\widehat{v}(\alpha) \ = \ \langle v, e_\alpha \rangle$$

The *Riemann-Lebesgue lemma* relevant here is

$$\lim_\alpha |\widehat{v}(\alpha)| \ = \ 0$$

More explicitly, this means that for given $\varepsilon > 0$ there is a finite subset A_o of A so that for $\alpha \notin A_o$ we have

$$|\widehat{v}(\alpha)| \ < \ \varepsilon$$

This follows from the fact that the infinite sum

$$\sum_\alpha |\hat{v}(\alpha)|^2$$

is *convergent.*

8.7 Gram-Schmidt process

Let $S = \{v_n : n = 1, 2, 3, \ldots\}$ be a well-ordered set of vectors in a pre-Hilbert space V. For simplicity, we are also assuming that S is *countable.* Let V_o be the collection of all finite linear combinations of S, and suppose that V_o is *dense* in V. Then we can obtain an *orthonormal basis* from S by the following procedure, called the *Gram-Schmidt process:*

Let v_{n_1} be the first of the v_i which is non-zero, and put

$$e_1 \ = \ \frac{v_{n_1}}{|v_{n_1}|}$$

Let v_{n_2} be the first of the v_i which is *not* a multiple of e_1. Put

$$f_2 \ = \ v_{n_2} - \langle v_{n_2}, e_1 \rangle e_1$$

and

$$e_2 \ = \ \frac{f_2}{|f_2|}$$

Inductively, suppose we have chosen $e_1, \ldots, e_k$ which form an orthonormal set. Let $v_{n_{k+1}}$ be the first of the v_i not expressible as a linear combination of $e_1, \ldots, e_k$. Put

$$f_{k+1} \ = \ v_{n_{k+1}} - \sum_{1 \le i \le k} \langle v_{n_{k+1}}, e_i \rangle e_i$$

and

$$e_{k+1} \ = \ \frac{f_{k+1}}{|f_{k+1}|}$$

Then induction on k proves that the collection of all finite linear combinations of $e_1, \ldots, e_k$ is the same as the collection of all finite linear combinations of $v_{n_1}, v_{n_2}, v_{n_3}, \ldots, v_{n_k}$. Thus, the collection of all finite linear combinations of the orthonormal set $e_1, e_2, \ldots$ is *dense* in V, so this is an orthonormal basis.

8.8 Linear maps, linear functionals, Riesz-Fréchet theorem

We consider maps $T : V \to W$ from one Hilbert space to another which are not only *linear*, but also *continuous*. The linearity is

$$T(av + bw) \;=\; a \cdot Tv + b \cdot Tw \qquad\qquad \text{(for scalars } a, b \text{ and } v, w \in V\text{)}$$

and the continuity is as expected for a map from one metric space to another: given $v \in V$ and given $\varepsilon > 0$, there is small enough $\delta > 0$ such that for $v' \in V$ with $|v' - v|_V < \delta$, we have $|Tv - Tv'|_W < \varepsilon$.

A *(continuous, linear) functional* λ on a Hilbert space V is a continuous linear map $\lambda : V \to \mathbb{C}$.

The *kernel* or *nullspace* of a linear map T is

$$\ker T \;=\; \{v \in V : Tv = 0\}$$

A linear map $T : V \to W$ is *bounded* when there is a finite real constant C so that, for all $v \in V$,

$$|Tv|_W \;<\; C|v|_V \qquad\qquad \text{(for all } v \in V\text{)}$$

The collection of all continuous linear functionals on a Hilbert space V is denoted by V^*.

[8.8.1] Claim: *Continuity* of a linear map $T : V \to W$ is equivalent to *boundedness*.

Proof: Continuity *at zero* is the assertion that for all $\varepsilon > 0$ there is an open ball $B = \{v \in V : |v|_V < \delta\}$ (with $\delta > 0$) such that $|Tv|_W < \varepsilon$ for $v \in B$. In particular, take $\delta > 0$ so that for $|v| < \delta$ we have

$$|Tv| \;<\; 1$$

For arbitrary $0 \neq x \in V$ we have

$$\left|\frac{\delta}{2|x|} \cdot x\right| \;<\; \delta$$

Therefore,

$$\left|T\Big(\frac{\delta}{2|x|} \cdot x\Big)\right|_W \;<\; 1$$

By the linearity of T,

$$|Tx|_W \;<\; \frac{2}{\delta} \cdot |x|_V$$

That is, continuity implies boundedness.

On the other hand, suppose that there is a finite real constant C so that, for all $x \in V$,

$$|Tx| \;<\; C|x|$$

For $|x - y| < \varepsilon/C$

$$|Tx - Ty|_W \;=\; |T(x-y)|_W \;<\; C|x - y|_V \;<\; C \cdot \frac{\varepsilon}{C} \;=\; \varepsilon$$

showing that boundedness implies continuity. Thus, *boundedness and continuity are equivalent.* ///

For a pre-Hilbert space V with completion $\bar{V}$, *a continuous linear functional* λ *on* V *has a unique extension to a continuous linear functional on* $\bar{V}$, defined by

$$\bar{\lambda}(\lim_n x_n) \;=\; \lim_n \lambda(x_n)$$

It is not difficult to check that this formula gives a well-defined function (due to the continuity of the original λ), and is additive and linear.

The dual V^* has a natural *norm*

$$|\lambda|_{V^*} \;=\; \sup_{v \in V : |v| \le 1} |\lambda(v)| \qquad\qquad (\text{for } \lambda \in V^*)$$

By the *minimum principle*, the sup is attained.

[8.8.2] Theorem: *(Riesz-Fréchet)* Every continuous linear functional λ on a Hilbert space V is of the form

$$\lambda(x) \;=\; \langle x, y_\lambda \rangle$$

for a uniquely-determined y_λ in V. Further, $|y_\lambda|_V = |\lambda|_{V^*}$. Thus, the map $V \to V^*$ by $v \to \lambda_v$ defined by $\lambda_v(w) = \langle w, v\rangle_V$ is a *conjugate-linear* isomorphism $V \to V^*$: it preserves vector addition and preserves the metric, but scalar multiplication is conjugated: $y_{a\lambda} = \overline{a} \cdot y_\lambda$ for $a \in \mathbb{C}$.

Proof: The kernel $\ker \lambda$ of a non-zero continuous linear functional λ is a proper closed subspace. From above, there is a non-zero element $z \in (\ker \lambda)^\perp$. Replace z by $z/\lambda(z)$ so that $\lambda(z) = 1$ without loss of generality. For any $v \in V$,

$$\lambda(v - \lambda(v)z) \;=\; \lambda(v) - \lambda(v) \cdot 1 \;=\; 0$$

so $v - \lambda(v)z \in \ker \lambda$. Therefore,

$$0 \;=\; \langle v - \lambda(v)z, z\rangle$$

Thus,

$$\langle v, z\rangle \;=\; \lambda(v) \cdot \langle z, z\rangle$$

so that

$$\langle v, \frac{z}{\langle z, z\rangle}\rangle \;=\; \lambda(v)$$

proving *existence*. For *uniqueness*, when $\langle x, z\rangle = \langle x, z'\rangle$ for specific z, z' and for all x, then $\langle x, z - z'\rangle = 0$ for all x gives $z = z'$, giving uniqueness.

Of course, every $y \in V$ gives a continuous linear functional by $x \to \langle x, y\rangle_V$. This is the inverse map to $\lambda \to y_\lambda$, so both are bijections. Addition is preserved:

$$\langle v, y_{\lambda+\mu}\rangle_V \;=\; (\lambda + \mu)(v) \;=\; \lambda v + \mu v \;=\; \langle v, y_\lambda\rangle_V + \langle v, y_\mu\rangle_V \;=\; \langle v, y_\lambda + y_\mu\rangle_V \qquad (\text{for all } v)$$

The conjugation of scalars follows similarly, from the hermitian-ness of $\langle,\rangle_V$:

$$\langle v, y_{a\lambda}\rangle_V \;=\; (a\lambda)(v) \;=\; a \cdot \lambda v \;=\; a\langle v, y_\lambda\rangle_V \;=\; \langle v, \overline{a} \cdot y_\lambda\rangle_V$$

as claimed. ///

[8.8.3] Corollary: The dual V^* has a natural Hilbert space structure, given by

$$\langle \lambda, \mu\rangle_{V^*} \;=\; \overline{\langle y_\lambda, y_\mu\rangle_V} \qquad (\text{where } \lambda(v) = \langle v, y_\lambda\rangle_V \text{ and } \mu(v) = \langle v, y_\mu\rangle_V, \text{ for all } v \in V)$$

Proof: Checking the pre-Hilbert space properties is straightforward. Completeness follows from the property $|y_\lambda|_V = |\lambda|_{V^*}$. ///

[8.8.4] Corollary: $V \approx (V^*)^*$ as Hilbert spaces, given by map $\varphi : V \to V^{**}$ by $\varphi(v)(\lambda) = \lambda v$.

Proof: One checks directly that φ gives a continuous, injective, complex-linear map $V \to V^{**}$. We claim that it is composite of the two conjugate-linear isomorphisms $V \to V^*$ and $V^* \to (V^*)^*$. Let $v \to \lambda_v = \langle -, v\rangle$ be the map $V \to V^*$, and $\mu \to \Lambda_m u = \langle -, \mu\rangle_{V^*}$ the map $V^* \to (V^*)^*$. For $v, w \in V$,

$$\varphi(v)(\lambda_w) \;=\; \lambda_w(v) \;=\; \langle v, w\rangle_V \;=\; \langle v, w\rangle_V \;=\; \langle \lambda_w, \lambda_v\rangle_{V^*} \;=\; \Lambda_{\lambda_v}(\lambda_w)$$

By Riesz-Fréchet, the vectors λ_w fill out V^* for $w \in V$. Thus, $\varphi(v) = \Lambda_{\lambda_v}$, as claimed. ///

8.9 Adjoints

[8.9.1] Claim: Given a continuous linear map $T : V \to W$ of Hilbert spaces, there is a unique continuous linear $T^* : W^* \to V^*$ characterized by

$$(T^*\mu)(v) \;=\; \mu(Tv) \qquad\qquad (\text{for } \mu \in W^*, \text{ for all } v \in V)$$

Proof: The map $V \to \mathbb{C}$ by $v \to Tv \to \mu(Tv)$ is a composite of continuous functions, so is continuous. It is linear for the same reason. Call it $T^*\mu \in V^*$. To show that $\mu \to T^*\mu$ is continuous, it is convenient to look at *bounds*: since T is continuous, it is bounded, so there is C such that $|Tv|_W \le C \cdot |v|_V$, and then

$$|(T^*\mu)(v)| \;=\; |\mu(Tv)| \;\le\; |\mu|_{W^*} \cdot |Tv|_W \;\le\; |\mu|_{W^*} \cdot C \cdot |v|_V$$

Thus, $|T^*\mu|_{V^*} \le |\mu|_{W^*} \cdot C < \infty$, so T^* is continuous. ///

[8.9.2] Remark: Somewhat surprisingly, for most continuous linear maps $T : V \to W$ of Hilbert spaces, the Riesz-Fréchet conjugate-linear isomorphisms $\alpha_V : V \to V^*$ and $\alpha_W : W \to W^*$ are not compatible with adjoints. That is, it is rare that the following square *commutes*:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "\alpha_V"'] & W \arrow[d, "\alpha_W"] \\ V^* & W^* \arrow[l, "T^*"'] \end{tikzcd}$$

In fact, the only situation in which such a square commutes is when T is an isometry to its image.

9. Introduction to Fourier series

1. Pointwise convergence
2. Fourier-Dirichlet kernel versus approximate identities
3. Fejer kernel
4. Density of trigonometric polynomials in $C^o(\mathbb{T})$
5. Exponentials form a Hilbert-space basis for $L^2(\mathbb{T})$

In his 1822 treatise on *heat*, J. Fourier espoused the wonderful idea that any function on an interval, say $[0,1]$, could be represented as an superposition of functions $\sin 2\pi nx$ and $\cos 2\pi nx$. Equivalently, as superposition of $e^{2\pi inx}$ with $n \in \mathbb{Z}$. Since these functions are eigenfunctions for the operator d/dx, such expressions facilitated solution of differential equations.

The issue of *convergence*, which in those days could only have meant *pointwise convergence*, was recognized. The first publication proving pointwise convergence was by P. Dirichlet in 1829, although the device in the proof had appeared in an earlier manuscript of Fourier whose publication was delayed.

B. Riemann's 1854 *Habilitationschrift* concerned the representability of functions by trigonometric series.

In 1915 N. Luzin conjectured that Fourier series of functions in $L^2(\mathbb{T})$ converge almost everywhere pointwise. Decades later, in 1966, L. Carlson proved Luzin's conjecture. In 1968, R. Hunt generalized this to $L^p(\mathbb{T})$ functions for $p > 1$.

From the other side, in 1876 P. du Bois-Reymond found a *continuous* function whose Fourier series diverges at a single point. Via the *uniform boundedness theorem*, we will show later that there are continuous functions whose Fourier series diverges at any given countable collection of points. A. Kolmogorov (1923/26) gave an example of an $L^1(\mathbb{T})$ function whose Fourier series diverges pointwise *everywhere*.

The *density* of trigonometric polynomials (*finite* Fourier series) in the space of continuous function $C^o(\mathbb{T})$ can be made to follow from Weierstraß' aproximation theorem. We give a somewhat different proof of density of trigonometric polynomials in $C^o(\mathbb{T})$, introducing and using the *Fejér kernel*. In 1904, L. Fejér gave an even more direct proof of the density of trigonometric polynomials in $L^2(\mathbb{T})$, in effect using an *approximate identity* made directly in terms of trigonometric polynomials. We reproduce this proof.

Urysohn's lemma implies the density of $C^o(\mathbb{T})$ in $L^2(\mathbb{T})$, so we have the *completeness* of Fourier series in $L^2(\mathbb{T})$. Thus, Fourier series of $L^2(\mathbb{T})$ functions converge to them *in the* $L^2(\mathbb{T})$ *topology*.

A. Zygmund's *Trigonometric Series, I, II* contains much more bibliographic and historical information.

9.1 Pointwise convergence

On $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ with coordinates $x \to e^{2\pi ix}$, the *Fourier series* of $f \in L^2(\mathbb{T})$ is [13]

$$f \sim \sum_{n\in\mathbb{Z}} \widehat{f}(n) \cdot \psi_n \qquad\qquad (\text{with } \psi_n(x) = e^{2\pi inx} \text{ and } \widehat{f}(n) = \int_{\mathbb{T}} f \cdot \overline{\psi}_n)$$

We do not write *equality* of the function and its Fourier series, since the question of possible *senses* of equality is significant. After all, the right-hand side is an infinite sum, possibly *numerical*, but also possibly

[13] Temporarily, to integrate a function F on the quotient $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, it suffices to let the quotient map be $q : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$, and integrate $F \circ q$ on a convenient set of representatives for the quotient, such as $[0,1]$ or any other interval $[x_o, x_o+1]$. In fact, we will suppress reference to the quotient map and identify functions on $\mathbb{T}$ with periodic functions on $\mathbb{R}$. A more systematic approach to integration on quotients, that does not require determination of nice sets of representatives, will be discussed later.

of *functions*, and the latter offers several potential interpretations. It is completely natural to ask for *pointwise* convergence of a Fourier series, and, implicitly, convergence to the function of which it is the Fourier series. We address this first.

We can prove pointwise convergence even before proving that the exponentials give an orthonormal *basis* for $L^2[0,1]$. The hypotheses of the convergence claim below are not optimal, but are sufficient for some purposes, and are tangible.

We also need:

[9.1.1] Claim: *(Riemann-Lebesgue)* For $f \in L^2(\mathbb{T})$, the Fourier coefficients $\widehat{f}(n)$ of f go to 0.

Proof: The $L^2[0,1]$ norm of $\psi_n(x) = e^{2\pi i n x}$ is 1. *Bessel's inequality*

$$|f|_{L^2}^2 \quad \geq \quad \sum_n \Big| \langle f, \psi_n \rangle \Big|^2$$

from abstract Hilbert-space theory applies to an orthonormal *set*, whether or not it is an orthonormal *basis*. Thus, the sum on the right converges, so by Cauchy's criterion the summands go to 0. ///

A function f on $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ is *(finitely) piecewise* C^o when there are finitely many real numbers $a_0 \leq a_1 \leq \ldots \leq a_{n-1} \leq a_n = a_0 + 1$ and C^0 functions f_i on $[a_i, a_{i+1}]$ such that

$$f_i(x) \;=\; f(x) \quad \text{on } [a_i, a_{i+1}] \quad \text{(except possibly at the endpoints)}$$

Thus, while $f_i(a_{i+1})$ may differ from $f_{i+1}(a_{i+1})$, and $f(a_{i+1})$ may be different from both of these, the function f is continuous in the interiors of the intervals, and behaves well *near* the endpoints, if not *at* the endpoints.

Write

$$\langle f, F \rangle \;=\; \int_{\mathbb{T}} f \cdot \overline{F} \;=\; \int_0^1 f(x)\, \overline{F(x)}\, dx$$

and

$$\widehat{f}(n) \;=\; \langle f, \psi_n \rangle \;=\; \int_{\mathbb{T}} f \cdot \overline{\psi}_n \;=\; \int_0^1 f(x)\, \overline{\psi}_n(x)\, dx$$

[9.1.2] Claim: Let f be finitely piecewise C^o on $\mathbb{T}$. Let x_o be a point at which f has both *left and right derivatives* (even if they do not agree), and is *continuous*. Then the Fourier series of f evaluated at x_o converges to $f(x_o)$. That is,

$$f(x_o) \;=\; \sum_{n \in \mathbb{Z}} \widehat{f}(n)\, \psi_n(x_o) \qquad \text{(a convergent sum)}$$

Proof: First, make reductions to unclutter the notation. By considering $f(x) - f(x_o)$, and observing that *constants* are represented pointwise by their Fourier expansions, we can assume that $f(x_o) = 0$. The Fourier coeffients of *translates* of a function f are expressible in terms of the Fourier coefficients of f itself, using the periodicity of f as a function on $\mathbb{R}$:

$$\int_0^1 f(x + x_o)\, \overline{\psi}_n(x)\, dx \;=\; \int_{x_o}^{1+x_o} f(x)\, \overline{\psi}_n(x - x_o)\, dx \;=\; \int_{x_o}^1 f(x)\, \overline{\psi}_n(x - x_o)\, dx + \int_1^{1+x_o} f(x)\, \overline{\psi}_n(x - x_o)\, dx$$

$$= \int_{x_o}^1 f(x)\, \overline{\psi}_n(x - x_o)\, dx + \int_0^{x_o} f(x)\, \overline{\psi}_n(x - x_o)\, dx \;=\; \int_0^1 f(x) \overline{\psi}_n(x - x_o)\, dx$$

$$= \psi_n(x_o) \int_0^1 f(x)\, \overline{\psi}_n(x)\, dx \;=\; \psi_n(x_o) \cdot \widehat{f}(n)$$

The left-hand side is the n^{th} Fourier coefficient of the translate $x \to f(x+x_o)$, that is, the n^{th} Fourier *term* of $x \to f(x+x_o)$ evaluated at 0, while the right-hand side is 2π times the n^{th} Fourier *term* of $f(x)$ evaluated at x_o. Thus, we can simplify further by taking $x_o = 0$, without loss of generality. [14]

A partial sum of the Fourier expansion evaluated at 0 is

$$\sum_{-M\le n<N} \int_0^2 f(x)\overline{\psi}_n(x)\,dx \;=\; \int_0^1 f(x) \sum_{-M\le n<N} \overline{\psi}_n(x)\,dx$$

$$= \int_0^1 f(x)\frac{\overline{\psi}_N(x) - \overline{\psi}_{-M}(x)}{\psi_{-1}(x)-1}\,dx$$

by summing the geometric series. This is

$$\int_0^1 \frac{f(x)}{\psi_{-1}(x)-1}(\overline{\psi}_N(x) - \overline{\psi}_{-M}(x))\,dx \;=\; \Big\langle \frac{f}{\psi_{-1}-1},\, \psi_N \Big\rangle - \Big\langle \frac{f}{\psi_{-1}-1},\, \psi_{-M} \Big\rangle$$

The latter two terms are Fourier coefficients of $f/(\psi_{-1}-1)$, so go to 0 by the Riemann-Lebesgue lemma for $f(x)/(\psi_{-1}(x)-1)$ in $L^2(\mathbb{T})$. Since $x_o = 0$ and $f(x_o) = 0$

$$\frac{f(x)}{\psi_{-1}(x)-1} \;=\; \frac{f(x)}{x}\cdot\frac{x}{\psi_{-1}(x)-1} \;=\; \frac{f(x)-f(x_o)}{x-x_o}\cdot\frac{x-x_o}{e^{-2\pi i x} - e^{-2\pi i x_o}}$$

The existence of left and right derivatives of f at $x_o = 0$ is exactly the hypothesis that this expression has left and right limits at x_o, even if they do not agree.

At all other points the division by $\psi_{-1}(x)-1$ does not disturb the continuity. Thus, $f/(\psi_{-1}-1)$ is still at least *continuous* on each interval $[a_i, a_{i+1}]$ on which f was essentially a C^o function. Therefore, $f/(\psi_{-1}-1)$ is continuous on a finite set of closed (finite) intervals, so bounded on each one. Thus, $f/(\psi_{-1}-1)$ is indeed L^2, and we can invoke Riemann-Lebesgue to see that the integral goes to $0 = f(x_o)$. ///

[9.1.3] Corollary: The Fourier series of $f \in C^1(\mathbb{T})$ converges pointwise to f everywhere. ///

[9.1.4] Remark: Pointwise convergence does not give L^2 convergence, and we have *not* yet proven that the exponentials are an orthonormal *basis* for the Hilbert space $L^2(\mathbb{T})$. The pointwise result just proven is suggestive, but not decisive.

[14] The rearrangement of the integral would have been simpler if we integrated directly on $\mathbb{T}$, a *group*, rather than on a set of representatives in $\mathbb{R}$ which had to be rearranged. That is, the change of variables that replaces x by $x - x_o$ is an automorphism of $\mathbb{T}$, and the measure is invariant under translation, so for F on $\mathbb{T}$ we can write simply

$$\int_{\mathbb{T}} F(x)\,dx \;=\; \int_{\mathbb{T}} F(x-x_o)\,dx$$

9.2 Fourier-Dirichlet kernel versus approximate identities

Under suitable hypotheses on f, in the above proof of pointwise convergence, rewriting a little, we have

$$f(x) \;=\; \lim_N \int_{\mathbb{T}} f(x+\xi)\cdot\Big(\sum_{-N\le n\le N} \overline{\psi}_n(\xi)\Big)\,d\xi \;=\; \lim_N \int_{\mathbb{T}} f(x+\xi)\cdot \frac{e^{2\pi i N\xi}-e^{-2\pi i(N+1)\xi}}{1-e^{-2\pi i\xi}}\,d\xi$$

$$=\; \lim_N \int_{\mathbb{T}} f(x+\xi)\cdot \frac{e^{2\pi i(N+\frac{1}{2})\xi}-e^{-2\pi i(N+\frac{1}{2})\xi}}{e^{\pi i\xi)}-e^{-\pi i\xi}} d\xi \;=\; \lim_N \int_{\mathbb{T}} f(x+\xi)\cdot \frac{\sin(2\pi(N+\frac{1}{2})\xi)}{\sin(\pi\xi)} d\xi$$

The sequence of functions

$$K_N(\xi) \;=\; \frac{\sin(2\pi(N+\frac{1}{2})\xi)}{\sin(\pi\xi)} d\xi$$

are often called the *Dirichlet kernel(s)*, although these functions did appear earlier in work of Fourier himself, whose publication was delayed.

Unlike the *Fejér kernel* in the following section, the Fourier-Dirichlet kernel does *not* have properties that would make it an *approximate identity.* An *approximate identity* on $\mathbb{T}$ is a sequence $\{\varphi_1, \varphi_2, \ldots\}$ of continuous functions such that

$$\int_{\mathbb{T}} \varphi_n \;=\; 1 \qquad\qquad (\text{for all } n)$$

and such that the *masses* bunch up near $1 \in \mathbb{T}$, in the sense that for every neighborhood U of 1 in $\mathbb{T}$,

$$\lim_n \int_U \varphi_n \;\to\; 1$$

The virtue of an approximate identity, not possessed by the Fourier-Dirichlet kernel, is

[9.2.1] Claim: For an approximate identity $\{\varphi_n\}$ on $\mathbb{T}$ and for $f \in C^o(\mathbb{T})$,

$$\lim_n \int_{\mathbb{T}} f(x+\xi)\,\varphi_n(\xi)\,d\xi \;\;=\;\; f(x) \qquad\qquad (\text{uniformly in } x \in \mathbb{T})$$

Proof: By the uniform continuity of f on compact $\mathbb{T}$, given $\varepsilon > 0$, there is $\delta > 0$ such that $|f(x) - f(y)| < \varepsilon$ for all x, y with $|x - y| < \delta$, in terms of the parametrization $\mathbb{R} \to \mathbb{T}$.

Let N be the image in $\mathbb{T}$ of $(-\delta, \delta) \subset \mathbb{R}$. Invoking the approximate identity property, let n_o be large enough so that $|\int_N \varphi_n - 1| < \varepsilon$ for all $n \ge n_o$. Since $\int_{\mathbb{T}} \varphi_n = 1$ and $\varphi_n(\xi) \ge 0$ for all ξ, this also implies $|\int_{\mathbb{T}-N} \varphi_n| < \varepsilon$. Then

$$\int_{\mathbb{T}} f(x+\xi)\;\varphi_n(\xi)\;d\xi \;=\; \int_N f(x+\xi)\;\varphi_n(\xi)\;d\xi + \int_{\mathbb{T}-N} f(x+\xi)\;\varphi_n(\xi)\;d\xi$$

The first integral, over U, is

$$\int_U f(x)\;\varphi_n(\xi)\;d\xi + \int_U (f(x+\xi)-f(x))\varphi_n(\xi)\;d\xi \;=\; f(x)\cdot\int_U \varphi_n(\xi)\;d\xi + \int_U (f(x+\xi)-f(x))\varphi_n(\xi)\;d\xi$$

As $n \to +\infty$, the first summand goes to $f(x)\cdot 1$ uniformly in x. The second summand is small:

$$\Big|\int_U (f(x+\xi)-f(x))\varphi_n(\xi)\;d\xi\Big| \;=\; \int_U |f(x+\xi)-f(x)|\cdot\varphi_n(\xi)\;d\xi \;<\; \int_U \varepsilon\cdot\varphi_n(\xi)\;d\xi \;\le\; \varepsilon\int_{\mathbb{T}} \varphi_n(\xi)\;d\xi \;=\; \varepsilon$$

Similarly, the integral over $\mathbb{T} - U$ is small, uniformly in x:

$$\Big|\int_{\mathbb{T}-U} f(x+\xi)\,\varphi_n(\xi)\;d\xi\Big| \;\le\; \sup_{y\in\mathbb{T}}|f(y)|\cdot\int_{\mathbb{T}-U} \varphi_n(\xi)\;d\xi \;<\; \sup_{y\in\mathbb{T}}|f(y)|\cdot\varepsilon$$

giving the assertion of the claim. ///

9.3 Fejer kernel

We give Fejér's approximate identity consisting of trigonometric polynomials (finite Fourier series), whose property rearranges to prove sup-norm convergence of a sequence of trigonometric polynomials to given $f \in C^o(\mathbb{T})$.

However, the trigonometric polynomials converging uniformly pointwise to $f \in C^o(\mathbb{T})$ are *not* the finite partial sums of the Fourier series of f, but, rather the *Cesaro-summed* version of these partial sums. That is, given a sequence $b_1, b_2, \ldots$, the Cesaro-summed sequence is

$$s_1 = \frac{b_1}{1} \qquad s_2 = \frac{b_1+b_2}{2} \qquad s_3 = \frac{b_1+b_2+b_3}{3} \qquad s_4 = \frac{b_1+b_2+b_3+b_4}{4} \qquad \ldots$$

On one hand, if the original sequence converges, then the Cesaro-summed sequence also converges, with the same limit. On the other hand, the Cesaro-summed sequence may converge though the original does not.

As it happens, Cesaro-summing the sequence of Fourier-Dirichlet kernels $K_N(x) = \sin(2\pi(N+\frac{1}{2})x)/\sin(\pi x)$ produces an approximate identity: the Fejér kernel is

$$F_n(x) \;=\; \frac{1}{n}\sum_{j=1}^{n} K_{j-1}(x) \;=\; \frac{1}{n}\sum_{j=0}^{n-1}\Big(\sum_{-j\le\ell\le j} e^{2\pi\ell x}\Big) \;=\; \sum_{-(n-1)\le\ell\le n-1} \frac{n-|\ell|}{n} e^{2\pi\ell x} \;=\; \sum_{-n\le\ell\le n} \frac{n-|\ell|}{n} e^{2\pi\ell x}$$

Visibly, $F_n(x)$ is a finite Fourier series. As with the Fourier-Dirichlet kernel, we can sum geometric series to simplify:

[9.3.1] Claim:

$$F_n(x) \;=\; \frac{1}{n}\cdot\frac{1-\cos 2\pi nx}{1-\cos 2\pi x}$$

In particular, $F_n(x) \ge 0$ for all x.

Proof: From

$$K_n(x) \;=\; \frac{\sin(2\pi(n+\frac{1}{2})x)}{\sin(\pi x)}$$

computing directly,

$$\sum_{j=1}^{n} \frac{\sin(2\pi(j-\frac{1}{2})x)}{\sin(\pi x)} \;=\; \frac{1}{2i\sin\pi x}\sum_{j=1}^{n}\Big(e^{2\pi i(j-\frac{1}{2})x} - e^{-2\pi i(j-\frac{1}{2})x}\Big)$$

$$= \frac{1}{2i\sin\pi x}\Big(\frac{e^{\pi ix} - e^{2\pi i(n+\frac{1}{2})x}}{1-e^{2\pi ix}} - \frac{e^{-\pi ix} - e^{-2\pi i(n+\frac{1}{2})x}}{1-e^{-2\pi ix}}\Big)$$

$$= \frac{1}{2i\sin\pi x}\Big(\frac{1-e^{2\pi inx}}{e^{-\pi ix}-e^{\pi ix}} - \frac{1-e^{-2\pi inx}}{e^{\pi ix}-e^{-\pi ix}}\Big) \;=\; \frac{1}{2i\sin\pi x}\,\frac{e^{2\pi inx}-2+e^{-2\pi inx}}{e^{\pi ix}-e^{-\pi ix}}$$

$$= \frac{1}{2i\sin\pi x}\,\frac{2(\cos 2\pi nx - 1)}{2i\sin\pi x} \;=\; \frac{1-\cos 2\pi nx}{2(\sin\pi x)^2} \;=\; \frac{1-\cos 2\pi nx}{1-\cos 2\pi x}$$

as asserted. ///

We check the other properties for $\{F_n\}$ to be an approximate identity:

[9.3.2] Claim: $\int_{\mathbb{T}} F_n(x)\,dx = 1$, and

$$\int_{|x|\le\frac{1}{\sqrt{n}}} F_n(x)\,dx \longrightarrow 0 \qquad\qquad (\text{as } n\to\infty)$$

Proof: First,

$$\int_{\mathbb{T}} F_n(x)\,dx \;=\; \sum_{-n\le\ell\le n} \frac{n-|\ell|}{n}\int_{\mathbb{T}} e^{2\pi i\ell x}\,dx \;=\; \sum_{-n\le\ell\le n} \frac{n-|\ell|}{n}\begin{cases} 0 & (\text{for } \ell\neq 0)\\ \\ 1 & (\text{for } \ell = 0)\end{cases} \;=\; 1$$

To show that the masses bunch up at 0, note that on the interval $[-\frac{1}{2},\frac{1}{2}]$,

$$F_n(x) \;=\; \frac{1}{n}\cdot\frac{1-\cos 2\pi nx}{1-\cos 2\pi x} \;=\; \frac{1}{n}\cdot\frac{(1-\cos 2\pi nx)/x^2}{(1-\cos 2\pi x)/x^2}$$

The denominator $(1-\cos 2\pi x)/x^2$ is non-vanishing and continuous on that interval, so is uniformly bounded away from 0. Thus, it suffices to show that the integral of $(1-\cos 2\pi nx)/nx^2$ on $[-\frac{1}{2},\frac{1}{2}]$ outside $[-\frac{1}{\sqrt{n}},\frac{1}{\sqrt{n}}]$ goes to 0. Indeed,

$$\frac{1}{n}\int_{\frac{1}{\sqrt{n}}}^{1} \frac{1-\cos 2\pi nx}{x^2}\,dx \;=\; n\cdot\int_{\frac{1}{\sqrt{n}}}^{1} \frac{1-\cos 2\pi nx}{(nx)^2}\,dx \;=\; \int_{\sqrt{n}}^{n} \frac{1-\cos 2\pi x}{x^2}\,dx$$

by replacing x by x/n. This is dominated by

$$\int_{\sqrt{n}}^{\infty} \frac{dx}{x^2}\,dx \;=\; \frac{1}{\sqrt{n}} \;\longrightarrow\; 0$$

This proves that $F_n(x)$ forms an approximate identity. ///

9.4 Completeness of Fourier series in $L^2(\mathbb{T})$

Using the approximate identity property of the Fejér kernels F_n, we can prove

[9.4.1] Corollary: The vector space of finite trigonometric polynomials is dense in $C^o(\mathbb{T})$, and, hence, in $L^2(\mathbb{T})$.

Proof: On one hand, from the discussion of approximate identities,

$$\int_{\mathbb{T}} F_n(\xi)\,f(x+\xi)\,d\xi \;\longrightarrow\; f(x) \qquad\qquad (\text{in sup-norm})$$

On the other hand, by rearranging and changing variables [15]

$$\int_{\mathbb{T}} F_n(\xi)\,f(x+\xi)\,d\xi \;=\; \sum_{|\ell|\le n}\frac{n-|\ell|}{n}\int_{\mathbb{T}} e^{2\pi in\xi}\;f(x+\xi)\,d\xi \;=\; \sum_{|\ell|\le n}\frac{n-|\ell|}{n}\int_{T} e^{2\pi in(\xi-x)}\;f(\xi)\,d\xi$$

$$=\; \sum_{|\ell|\le n}\Big(\frac{n-|\ell|}{n}\int_{T} e^{2\pi in\xi)}\;f(\xi)\,d\xi\Big)\cdot e^{-2\pi inx}$$

[15] As earlier, if we imagine we are integrating on an interval, then a change of variables entails breaking the interval into two pieces and rearranging. This necessity is avoided if we know how to integrate on groups $\mathbb{T}$, whether or not expressible as quotients $\mathbb{R}/\mathbb{Z}$.

That is, these trigonometric polynomials approach $f \in C^o(\mathbb{T})$ in sup-norm. ///

9.5 Exponentials form a Hilbert-space basis for $L^2(\mathbb{T})$

[9.5.1] Corollary: The exponentials $x \to e^{2\pi i n x}$ for $n \in \mathbb{Z}$ are an orthonormal basis, that is, a Hilbert space basis, for $L^2(\mathbb{T})$.

Proof: That they are mutually orthogonal, and have L^2-norms all 1, is a direct computation. The density of trigonometric polynomials in $L^2(\mathbb{T})$ is the assertion that the vector space of finite linear combinations of these exponentials is *dense* in $L^2(\mathbb{T})$. That is, there are no (non-zero) vectors orthogonal to all exponentials. ///

[9.5.2] Corollary: The Fourier series $\sum_n \langle f, \psi_n \rangle \cdot \psi_n$ of $f \in L^2(\mathbb{T})$ converges to f in the L^2 topology. ///

10. L^p spaces, convexity, basic inequalities

10.1 Examples: spaces L^p

Given a measure space X, for $1 \le p < \infty$ the usual L^p spaces are

$$L^p(X) = \{\text{measurable } f : |f|_{L^p} < \infty\} \text{ modulo } \sim$$

with the usual L^p norm

$$|f|_{L^p} \;=\; \left(\int_X |f|^p\right)^{1/p}$$

and associated metric

$$d(f,g) \;=\; |f-g|_{L^p}$$

taking the quotient by the equivalence relation

$$f \sim g \text{ if } f - g = 0 \text{ off a set of measure } 0$$

[10.1.1] Remark: These L^p functions have inevitably ambiguous pointwise values, in conflict with the naive formal definition of *function*.

A simple instance of this construction is

$$\ell^p \;=\; \{\text{complex sequences } \{c_i\} \text{ with } \sum_i |c_i|^p < \infty\}$$

with norm $|(c_1, c_2, \ldots)|_{\ell^p} = \left(\sum_i |c_i|^p\right)^{1/p}$. The analogue of the following theorem for ℓ^p is more elementary.

[10.1.2] Theorem: The space $L^p(X)$ is a complete metric space.

[10.1.3] Remark: In fact, as used in the proof, a Cauchy sequence f_i in $L^p(X)$ has a subsequence converging *pointwise* off a set of measure 0 in X.

Proof: The triangle inequality here is *Minkowski's inequality.* To prove completeness, choose a subsequence f_{n_i} such that

$$|f_{n_i} - f_{n_{i+1}}|_p \;<\; 2^{-i}$$

and put

$$g_n(x) \;=\; \sum_{1 \le i \le n} |f_{n_{i+1}}(x) - f_{n_i}(x)|$$

and

$$g(x) \;=\; \sum_{1 \le i < \infty} |f_{n_{i+1}}(x) - f_{n_i}(x)|$$

The infinite sum is not necessarily claimed to converge to a finite value for every x. The triangle inequality shows that $|g_n|_p \le 1$. *Fatou's Lemma* asserts that for $[0,\infty]$-valued measurable functions h_i

$$\int_X \left(\liminf_i h_i\right) \;\le\; \liminf_i \int_X h_i$$

Thus, $|g|_p \leq 1$, so is finite. Thus,

$$f_{n_1}(x) + \sum_{i\geq 1} (f_{n_{i+1}}(x) - f_{n_i}(x))$$

converges for almost all $x \in X$. Let $f(x)$ be the sum at points x where the series converges, and on the measure-zero set where the series does not converge put $f(x) = 0$. Certainly

$$f(x) \;=\; \lim_i f_{n_i}(x) \quad \text{(for almost all } x\text{)}$$

Now prove that this almost-everywhere pointwise limit is the L^p-limit of the original sequence. For $\varepsilon > 0$ take N such that $|f_m - f_n|_p < \varepsilon$ for $m, n \geq N$. Fatou's lemma gives

$$\int |f - f_n|^p \leq \liminf_i \int |f_{n_i} - f_n|^p \leq \varepsilon^p$$

Thus $f - f_n$ is in L^p and hence f is in L^p. And $|f - f_n|_p \to 0$. ///

[10.1.4] Theorem: For a locally compact Hausdorff topological space X with positive regular Borel measure μ, the space $C_c^0(X)$ of compactly-supported continuous functions is *dense* in $L^p(X,\mu)$.

Proof: From the definition of *integral* attached to a measure, an L^p function is approximable in L^p metric by a *simple* function, that is, a measurable function assuming only finitely-many values. That is, a simple function is a *finite* linear combination of characteristic functions of measurable sets E. Thus, it suffices to approximate characteristic functions of measurable sets by continuous functions. The assumed *regularity* of the measure gives compact K and open U such that $K \subset E \subset U$ and $\mu(U - E) < \varepsilon$, for given $\varepsilon > 0$. Urysohn's lemma says that there is continuous f identically 1 on K and identically 0 off U. Thus, f approximates the characteristic function of E. ///

[10.1.5] Corollary: For locally compact Hausdorff X with regular Borel measure μ, $L^p(X,\mu)$ is the L^p-metric completion of $C_c^o(X)$, the compactly-supported continuous functions. ///

[10.1.6] Remark: *Defining* $L^p(X,\mu)$ to be the L^p completion of $C_c^o(X)$ avoids discussion of ambiguous values on sets of measure zero.

10.2 Convexity and inequalities

A function f on an interval $(a,b) \subset \mathbb{R}$ is *convex* when its graph bends upward, in the sense that a line segment connecting two points on the graph lies *above* the graph. That is,

$$f(tx + (1-t)y) \;\geq\; tf(x) + (1-t)f(y) \qquad \text{(for } 0 \leq t \leq 1 \text{ and } a < x < y < b\text{)}$$

The prototype is the exponential function $x \to e^x$.

[10.2.1] Claim: Convex $\mathbb{R}$-valued functions on an open interval (a,b) (allowing $a = -\infty$ and/or $b = +\infty$) are *continuous.*

Proof: Let g be continuous on (a,b) and take $x \in (a,b)$. Fix any s,t such that $a < s < x < t < b$. For y in the range $x < y < t$, the point $(y, g(y))$ is on or above the line through $(s, g(s))$ and $(x, g(x))$, and is below the line through $(x, g(x))$ and $(t, g(t))$, so $g(y) \to g(x)$ as $y \to x^+$. For $s < y < x$, the same argument gives *left*-continuity. ///

[10.2.2] Theorem: *(Jensen's inequality)* Let X be a measure space with positive measure of total measure 1. Let $f \in L^1(X)$ be an $\mathbb{R}$-valued function on X with $a < f(x) < b$ for all $x \in X$, where a, b can also be $-\infty$ and $+\infty$. For convex g on (a,b),

$$g\Big(\int_X f\Big) \;\leq\; \int_X g \cdot f$$

Proof: First, $a < f(x) < b$ gives $a < \int_X f < b$. The convexity condition can be rewritten as the condition that slopes of secants increase from left to right. Thus, for example,

$$\frac{g(y)-g(x)}{y-x} \;\le\; \frac{g(z)-g(y)}{z-y} \qquad\qquad \text{(for } x<y<z \text{ inside } (a,b)\text{)}$$

Applying this with $y = \int_X f$,

$$\frac{g(\int f)-g(x)}{\int f - x} \;\le\; \frac{g(z)-g(\int f)}{z-\int f} \qquad\qquad \text{(for all } a<x<\textstyle\int_X f \text{ and for all } \int_X f < z < b\text{)}$$

With

$$S \;=\; \sup_{a<x<\int f} \frac{g(\int f)-g(x)}{\int f - x}$$

we have

$$\frac{g(\int f)-g(x)}{\int f - x} \;\le\; S \;\le\; \frac{g(z)-g(\int f)}{z-\int f} \qquad\qquad \text{(for all } a<x<\textstyle\int_X f \text{ and for all } \int_X f < z < b\text{)}$$

Thus, from the left half of the latter inequality,

$$g(x) \;\ge\; g(\int_X f) + S\cdot(x - \int_X f) \qquad\qquad \text{(for } a<x<\textstyle\int_X f\text{)}$$

and from the right half

$$g(z) \;\ge\; g(\int_X f) + S\cdot(z - \int_X f) \qquad\qquad \text{(for } \textstyle\int_X f < z < b\text{)}$$

Thus,

$$g(w) \;\ge\; g(\int_X f) + S\cdot(w - \int_X f) \qquad\qquad \text{(for } \textit{all } w \text{ in the range } a<w<b\text{)}$$

In particular, letting $w = f(x)$ now with $x \in X$,

$$g(f(x)) \;\ge\; g(\int_X f) + S\cdot(f(x) - \int_X f) \qquad\qquad \text{(for } \textit{all } w \text{ in the range } a<w<b\text{)}$$

Since the convex function g is *continuous*, $g \circ f$ is measurable. Integrating in $x \in X$, using the fact that the total measure is 1,

$$\int_X g\circ f \;\ge\; g(\int_X f) + S\cdot(\int_X f - \int_X f) \;=\; g(\int_X f) + S\cdot 0$$

as claimed. ///

[10.2.3] Corollary: *(Arithmetic-geometric mean inequality)* For positive real numbers $a_1, \ldots, a_n$,

$$(a_1 a_2 \ldots a_n)^{1/n} \;\le\; \frac{a_1+a_2+\ldots+a_n}{n}$$

Proof: In Jensen's inequality, take $g(x) = e^x$, take X a finite set with n (distinct) elements $\{x_1, \ldots, x_n\}$, with each point having measure $1/n$, and $f(x_i) = \log a_i$. Jensen's inequality gives

$$\exp\Big(\frac{\log a_1 + \ldots + \log a_n}{n}\Big) \;\le\; \frac{e^{\log a_1} + \ldots + e^{\log a_n}}{n}$$

which gives the assertion. ///

Conjugate exponents are numbers $p, q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} \;=\; 1$$

For example, p and $\frac{p}{p-1}$ are conjugate exponents.

Generalizing the Cauchy-Schwarz-Bunyakowsky inequality,

[10.2.4] Corollary: *(Hölder)* For conjugate exponents p, q and $[0, +\infty]$-valued measurable functions f, g,

$$\int_X f \cdot g \;\leq\; \Big(\int_X f^p\Big)^{\frac{1}{p}} \cdot \Big(\int_X g^q\Big)^{\frac{1}{q}}$$

Proof: The assertion is trivial if either integral on the right-hand side is $+\infty$ or 0, so suppose the two quantities

$$I \;=\; \Big(\int_X f^p\Big)^{\frac{1}{p}} \qquad\qquad J \;=\; \Big(\int_X g^q\Big)^{\frac{1}{q}}$$

are finite and non-zero. Renormalize by taking $\varphi = f/I$ and $\psi = g/J$, so that $\int \varphi^p = 1 = \int \psi^q$. For $x \in X$ with $0 < \varphi(x) < \infty$ and $0 < \psi(x) < \infty$, there are real numbers u, v such that $e^{u/p} = \varphi(x)$ and $e^{v/q} = \psi(x)$. Invoking Jensen's inequality on a measure space with just two points with measures $\frac{1}{p}$ and $\frac{1}{q}$, using the convexity of the exponential function,

$$\varphi(x)\psi(x) \;=\; e^{\frac{u}{p}+\frac{v}{q}} \;\leq\; \frac{e^u}{p} + \frac{e^v}{q} \;=\; \frac{\varphi(x)^p}{p} + \frac{\psi(x)^q}{q}$$

Integrating,

$$\int_X \varphi \cdot \psi \;\leq\; \int_X \frac{\varphi(x)^p}{p} + \frac{\psi(x)^q}{q} \;=\; \frac{1}{p} + \frac{1}{q} \;=\; 1$$

From the renormalization, we are done. ///

For the triangle inequality in L^p spaces for general p, we need

[10.2.5] Corollary: *(Minkowski)* For $1 < p < +\infty$ and $[0, +\infty]$-valued measurable functions f, g,

$$\Big(\int_X (f+g)^p\Big)^{\frac{1}{p}} \;\leq\; \Big(\int_X f^p\Big)^{\frac{1}{p}} + \Big(\int_X g^p\Big)^{\frac{1}{p}}$$

Proof: We prove Minkowski's inequality from Hölder's, using the conjugate exponents p and $q = \frac{p}{p-1}$.

$$\int (f+g)^p \;=\; \int f \cdot (f+g)^{p-1} + \int g \cdot (f+g)^{p-1}$$

$$\leq\; \Big(\int f^p\Big)^{\frac{1}{p}} \cdot \Big(\int (f+g)^{(p-1)q}\Big)^{\frac{1}{q}} \;+\; \Big(\int g^p\Big)^{\frac{1}{p}} \cdot \Big(\int (f+g)^{(p-1)q}\Big)^{\frac{1}{q}}$$

$$=\; \Big[\Big(\int f^p\Big)^{\frac{1}{p}} + \Big(\int g^p\Big)^{\frac{1}{p}}\Big] \cdot \Big(\int (f+g)^p\Big)^{\frac{p-1}{p}}$$

Dividing through by $\big(\int (f+g)^p\big)^{\frac{p-1}{p}}$ gives Minkowski's inequality. ///

11. Examples discussion

[11.18] Show that every vector subspace of $\mathbb{R}^n$ and/or $\mathbb{C}^n$ is (topologically) *closed*.

Discussion: Let $v_1, \ldots, v_m$ be an orthonormal basis for the given vector subspace W. For a Cauchy sequence $\{w_n\}$ in W, we claim that for each j the sequence $\langle w_n, v_j\rangle$ is Cauchy: by Cauchy-Schwarz-Bunyakowsky,

$$|\langle w_n, v_j\rangle - \langle w_{n'}, v_j\rangle| \;=\; |\langle w_n - w_{n'}, v_j\rangle| \;\le\; |w_n - w_{n'}| \cdot |v_j| \;=\; |w_n - w_{n'}|$$

Thus, by completeness of $\mathbb{R}$ and/or $\mathbb{C}$, that sequence has a limit c_j. As expected, we claim that $\lim_n w_n = \sum_{j=1}^m c_j \cdot v_j$. Indeed, using the orthonormality of the v_j's,

$$\Big| w_n - \sum_{j=1}^m c_j \cdot v_j \Big|^2 \;=\; \Big| w_n - \sum_{j=1}^m \lim_i \langle w_i, v_j\rangle \cdot v_j \Big|^2 \;=\; \Big| \sum_{j=1}^m \lim_i \langle w_n - w_i, v_j\rangle \cdot v_j \Big|^2$$

$$\le\; \sum_{j=1}^m |\lim_i \langle w_n - w_i, v_j\rangle|^2 \;=\; \lim_i \sum_{j=1}^m |\langle w_n - w_i, v_j\rangle|^2 \;\le\; \lim_i \sum_{j=1}^m |w_n - w_i| \cdot |v_j| \;=\; \lim_i m \cdot |w_n - w_i|$$

Take n_o large enough so that $|w_n - w_i| < \varepsilon$ for $i, n \ge n_o$. Then the latter expression is at most $m \cdot \varepsilon$. This holds for all $\varepsilon > 0$, so the limit is 0. ///

[11.19] For a subspace W of a Hilbert space V, show that $(W^\perp)^\perp$ is the closure of the subspace W in V.

Discussion: Let $\lambda_x(v) = \langle v, x\rangle$ for $x, v \in V$. Then $W^\perp = \bigcap_{w\in W} \ker \lambda_w$. Similarly, $(W^\perp)^\perp = \bigcap_{x \in W^\perp} \ker \lambda_x$. From the discussion in the Riesz-Fréchet theorem, or directly via Cauchy-Schwarz-Bunyakowsky, each λ_x is continuous, so $\ker \lambda_x = \lambda_x^{-1}(\{0\})$ is closed, since $\{0\}$ is closed. (One might check that the kernel of a linear map is a vector subspace.) An arbitrary intersection of closed sets is closed, so $(W^\perp)^\perp$ is closed.

Certainly $(W^\perp)^\perp \supset W$, because for each $w \in W$, $\langle x, w\rangle = 0$ for all $x \in W^\perp$. Thus, $(W^\perp)^\perp$ is a closed subspace, containing W. Being a closed subspace of a Hilbert space, $(W^\perp)^\perp$ is a Hilbert space itself. If $(W^\perp)^\perp$ were strictly larger than the topological closure $\overline{W}$ of W, then there would be $0 \ne y \in (W^\perp)^\perp$ orthogonal to $\overline{W}$. Then y would be orthogonal to W itself, so $0 \ne y \in W^\perp$, contradicting $0 \ne y \in (W^\perp)^\perp$. ///

[11.20] Show that for $0 < x < 1$

$$\sum_{n\ge1} \frac{\sin 2\pi n x}{n} \;=\; \pi(\tfrac{1}{2} - x)$$

Discussion: The Fourier series of the right-hand side is computed to be that given on the left-hand side. By the Fourier-Dirichlet result on pointwise convergence, since $\pi(\frac{1}{2} - x)$ is finitely-piecewise C^o, and has left and right derivatives in $(0, 1)$, its Fourier series converges to it pointwise there. ///

[11.21] Let $c_1, c_2, \ldots$ be positive real, converging monotonically to 0. For $0 < x < 1$, prove that $\sum_{n\ge0} c_n\, e^{2\pi i n x}$ *converges pointwise.*

Discussion: The expression as a Fourier series should not distract us from seeing an instance of the generalized alternating-decreasing criterion again, sometimes called *Dirichlet's criterion*: for a positive real sequence $c_1, c_2, \ldots$ monotone-decreasing to 0, and for a (possibly complex) sequence $b_1, b_2, \ldots$ with *bounded partial sums* $B_n = b_1 + \ldots + b_n$, the sum $\sum_n b_n c_n$ converges. The partial sums $\sum_{n\le N} e^{2\pi i n x}$ are bounded for $0 < x < 1$, by summing geometric series, so this criterion applies here.

The proof of the criterion itself is by *summation by parts*, a discrete analogue of integration by parts. That is, rewrite the tails of the sum as

$$\sum_{M\le n\le N} b_n c_n \;=\; \sum_{M\le n\le N} (B_n - B_{n-1})c_n \;=\; -B_{M-1}c_M + \sum_{M\le n\le N} B_n(c_n - c_{n+1}) + B_N c_{N+1}$$

Since the partial sums are bounded, the first and last summand go to 0. Letting β be a bound for all the $|B_n|$, the summation is

$$\Big|\sum_{M\le n\le N} B_n(c_n - c_{n+1})\Big| \;\le\; \sum_{M\le n\le N} |B_n|\cdot|c_n - c_{n+1}| \;=\; \sum_{M\le n\le N} |B_n|\cdot(c_n - c_{n+1}) \;\le\; \sum_{M\le n\le N} \beta\cdot(c_n - c_{n+1})$$

$$=\; \beta\cdot\sum_{M\le n\le N}(c_n - c_{n+1}) \;=\; \beta\cdot(c_M - c_{N+1})$$

by telescoping the series. Again, c_M and c_{N+1} go to 0. ///

[11.22] Show that the sup-norm completion of the space $C^o_c(\mathbb{R})$ of compactly-supported continuous functions is the space $C^o_o(\mathbb{R})$ of continuous functions going to 0 at infinity. An analogous assertion and argument should hold for any topological space in place of $\mathbb{R}$.

Discussion: The argument for this is general enough that we can replace $\mathbb{R}$ by a more general topological space X, probably locally compact and Hausdorff so that Urysohn's lemma assures us a good supply of continuous functions for auxiliary purposes. Then $C^o_o(X)$ is defined to be the collection of continuous functions f such that, given $\varepsilon > 0$, there is a compact $K\subset X$ such that $|f(x)| < \varepsilon$ for $x\notin K$.

First, show that any $f\in C^o_o(\mathbb{R})$ is a sup-norm limit of functions from $C^o_c(\mathbb{R})$. Given $\varepsilon > 0$, let K be sufficiently large so that $|f(x)| < \varepsilon$ for $x\notin K$. We claim that there is an open $U\supset K$ with compact closure $\overline{U}$ (which would be obvious on $\mathbb{R}$ or $\mathbb{R}^n$). For each $x\in K$, let $U_x\ni x$ be an open set with compact closure (using the local compactness). By compactness of K, there is a finite subcover $K\subset U_{x_1}\cup\ldots\cup U_{x_n}$. Then the closure of $U = U_{x_1}\cup\ldots\cup U_{x_n}$ is compact, as claimed. Then, invoking Urysohn's Lemma, let φ be a continuous function on X taking values in the interval $[0,1]$, that is 1 on K, and 0 off U, so φ has compact support. Then $\varphi\cdot f$ is continuous and has compact support, and

$$\sup_{x\in X}|f(x) - \varphi(x)\cdot f(x)| \;\le\; \sup_{x\in K}|f(x) - \varphi(x)\cdot f(x)| + \sup_{x\notin K}|f(x) - \varphi(x)\cdot f(x)| \;=\; 0 + \sup_{x\notin K}|f(x) - \varphi(x)\cdot f(x)|$$

$$\le\; \sup|1-\varphi|\cdot\sup_{x\notin K}|f(x)| \;<\; 1\cdot\varepsilon$$

That is, we can approximate f to within ε, as claimed.

On the other hand, now show that any sup-norm Cauchy sequence of $f_n\in C^o_c(X)$ has a pointwise limit f in $C^o_o(X)$. First, on any compact, the limit of the f_n's is *uniform* pointwise, so is continuous on compacts. Since every point $x\in X$ has a neighborhood U_x with compact closure, the pointwise limit is continuous on U_x. Thus, the pointwise limit is continuous at every point, hence continuous. Given $\varepsilon > 0$, take n_o sufficiently large so that $\sup_{x\in X}|f_m(x) - f_n(x)| < \varepsilon$ for all $m,n\ge n_o$. Let K be the support of f_{n_o}. Then

$$\sup_{x\notin K}|f(x)| \;=\; \sup_{x\notin K}|f(x) - f_{n_o}(x)| \;\le\; \sup_{x\in X}|f(x) - f_{n_o}| \;\le\; \varepsilon$$

Thus, the pointwise limit goes to 0 at infinity. ///

[11.23] Compute $\displaystyle\int_{\mathbb{R}}\Big(\frac{\sin x}{x}\Big)^2\,dx$. (*Hint:* use Plancherel.)

Discussion: From a standard stock of easy Fourier transforms, the Fourier transform of a characteristic function of a symmetrical interval is very close to the given function:

$$\widehat{\mathrm{ch}_{[-1,1]}}(\xi) \;=\; \int_{-1}^{1} e^{-2\pi i\xi x}\, dx \;=\; \frac{e^{-2\pi i\xi} - e^{2\pi i\xi}}{-2\pi i\xi} \;=\; \frac{\sin 2\pi\xi}{\pi\xi}$$

Applying Plancherel, we have

$$2 \;=\; \int_{\mathbb{R}} |\mathrm{ch}_{[-1,1]}|^2 \;=\; \int_{\mathbb{R}} \Big(\frac{\sin 2\pi\xi}{\pi\xi}\Big)^2\, d\xi$$

The change of variables replacing ξ by $\xi/2\pi$ gives

$$2 \;=\; \int_{\mathbb{R}} \Big(\frac{\sin\xi}{\xi/2}\Big)^2\, \frac{d\xi}{2\pi} \;=\; \frac{2}{\pi}\int_{\mathbb{R}} \Big(\frac{\sin\xi}{\xi}\Big)^2\, d\xi$$

Thus, the desired integral is π. ///

[11.24] For $f \in L^2(\mathbb{R})$ and $t \in \mathbb{R}$, show that there is a constant C (depending on f) such that

$$\Big|\int_{t-\delta}^{t+\delta} f(x)\, dx\Big| \;<\; C\cdot\sqrt{\delta}$$

Formulate and prove the corresponding assertion for L^p with $1 < p < \infty$.

Discussion: Let h_δ be the characteristic function of $[t-\delta, t+\delta]$. By Cauchy-Schwarz-Bunyakowsky

$$\Big|\int_{t-\delta}^{t+\delta} f\Big| \;=\; |\langle f, h_\delta\rangle_{L^2}| \;\le\; |f|_{L^2}\cdot|h_\delta|_{L^2} \;=\; |f|_{L^2}\cdot\sqrt{2\delta}$$

The case of conjugate exponents $\frac{1}{p}+\frac{1}{q} = 1$ is the same, using Hölder's inequality rather than Cauchy-Schwarz-Bunyakowsky. There is no immediate analogue for L^1, although a weaker result is possible, as in the next example. ///

[11.25] For $f \in L^1(\mathbb{R})$ and $t \in \mathbb{R}$, show that, given $\varepsilon > 0$, there is $\delta > 0$ such that

$$\Big|\int_{t-\delta}^{t+\delta} f(x)\, dx\Big| \;<\; \varepsilon$$

Sharpen the first example to show that

$$\int_{t-\delta}^{t+\delta} f(x)\, dx \;=\; o(\sqrt{\delta}) \qquad\qquad (\text{as } \delta\to 0^+)$$

where Landau's little-o notation is that $f(x) = o(g(x))$ as $x \to a$ when $\lim_{x\to a} f(x)/g(x) = 0$.

Discussion: Let $S_n = \{x : \frac{1}{n+1} \le |x-t| < \frac{1}{n}\}$. Then

$$\Big|\sum_{n\ge 1}\int_{S_n} f\Big| \;\le\; \sum_{n\ge 1}\int_{S_n} |f| \;\le\; |f|_{L^1}$$

Thus, the sum of non-negative terms $\sum_{n\ge 1}\int_{S_n}|f|$ is convergent, so the tails $\sum_{n\ge N}\int_{S_n}|f|$ go to 0 as $N \to +\infty$. Thus,

$$\Big|\int_{|x-t|\le 1/N} f\Big| \;\le\; \int_{|x-t|\le 1/N} |f| \;=\; \sum_{n\ge N}\int_{S_n} |f|$$

goes to 0 as $N \to +\infty$. Then this idea can be applied to $\int_{|x-t|<\delta} |f|^p$ in the previous example. ///

12. Examples discussion

[12.26] Fix $x_o \in [a,b]$. Show that $\lambda(f) = f(x_o)$ is a continuous linear functional on $C^o[a,b]$.

Discussion: Recall that, for linear maps, continuity is equivalent to continuity at 0, which is equivalent to being *bounded*, in the sense that there exists a constant C such that $|\lambda(f)| \le C \cdot |f|_{C^o}$ for all f. Here,

$$|\lambda(f)| \;=\; |f(x_o)| \;\le\; \sup_{x\in[a,b]} |f(x)| \;=\; |f|_{C^o}$$

so the constant $C = 1$ succeeds. ///

[12.27] Prove that Cesaro summation

$$b_1 = \frac{a_1}{1},\;\; b_2 = \frac{a_1+a_2}{2},\;\; b_3 = \frac{a_1+a_2+a_3}{3}, \ldots$$

converts every convergent sequence $a_1, a_2, \ldots$ to a convergent sequence $b_1, b_2, \ldots$ with the same limit.

Discussion: Let $\{a_n\}$ converge to A. Thus, given $\varepsilon > 0$, there is n_o be such for $n > n_o$ we have $|a_n - A| < \varepsilon$. Let $M = \max_{n\le n_o} |a_n|$. For $n \ge n_o$, by the triangle inequality,

$$|b_n - A| \;=\; \frac{|(a_1 - A) + \ldots + (a_n - A)|}{n} \;\le\; \frac{|a_1| + \ldots + |a_{n_o}|}{n} + \frac{n_o \cdot |A|}{n} \;+\; \frac{|a_{n_o+1} - A| + \ldots + |a_n - A|}{n}$$

$$< \;\frac{n_o \cdot M}{n} + \frac{n_o \cdot |A|}{n} + \varepsilon$$

For n sufficiently large, depending on A, n_o and M, the sum of the first two terms can be made smaller than ε. Replace ε by $\varepsilon/2$ throughout, if desired. ///

[12.28] (Collecting Fourier transform pairs...) Compute the Fourier transforms of

$$\chi_{[a,b]} \qquad\qquad e^{-\pi x^2} \qquad\qquad f(x) = \begin{cases} e^{-x} & (\text{for } x > 0) \\ 0 & (\text{for } x \le 0) \end{cases}$$

Discussion: The first of these is direct:

$$\widehat{\chi_{[a,b]}}(\xi) \;=\; \int_{\mathbb{R}} e^{-2\pi i\xi x} \chi_{[a,b]}(x)\,dx \;=\; \int_a^b e^{-2\pi i \xi x}\,dx \;=\; \begin{cases} \dfrac{e^{-2\pi i\xi b} - e^{2\pi i \xi a}}{-2\pi i \xi} & (\text{for } \xi \ne 0) \\ b - a & (\text{for } \xi = 0) \end{cases}$$

Since the latter function is *not* in $L^1(\mathbb{R})$, but *is* in $L^2(\mathbb{R})$, we define its Fourier transform (or inverse Fourier transform) *indirectly*, via either the inversion theorem, or by extending-by-continuity via Plancherel, expressing the function as an L^2 limite of L^1 functions.

The third is similarly direct:

$$\widehat{f}(\xi) \;=\; \int_0^\infty e^{-2\pi i \xi x}\, e^{-x}\,dx \;=\; \int_0^\infty e^{-(2\pi i\xi + 1)x}\,dx \;=\; \Big[\frac{e^{-(2\pi i \xi+1)x}}{-(2\pi i\xi + 1)}\Big]_0^\infty \;=\; \frac{1}{2\pi i \xi + 1}$$

Again, the latter function is not in L^1, but is in L^2, so its Fourier transform is most conveniently defined indirectly.

The Gaussian's Fourier transform is less trivial to evaluate, but is a very important example to have in hand, with many different applications throughout mathematics. One approach is as follows. Letting $f(x) = e^{-\pi x^2}$,

$$\widehat{f}(\xi) \;=\; \int_{\mathbb{R}} e^{-2\pi i\xi x}\, e^{-\pi x^2}\; dx \;=\; \int_{\mathbb{R}} e^{-\pi(x^2+2i\xi x)}\; dx \;=\; \int_{\mathbb{R}} e^{-\pi(x^2+i\xi)^2-\pi\xi^2}\; dx \;=\; e^{-\pi\xi^2}\int_{\mathbb{R}} e^{-\pi(x+i\xi)^2}\; dx$$

by completing the square. The unobvious claim is that the integral does not depend on ξ, and, in fact, has value 1. Perhaps the optimal approach here is to observe that the integral is equal to a complex contour integral:

$$\int_{\mathbb{R}} e^{-\pi(x^2+i\xi)^2}\; dx \;=\; \int_{i\xi-\infty}^{i\xi+\infty} e^{-\pi z^2}\; dz$$

along the line $\mathrm{Im}\,(z) = i\xi$. Given the good decay of the integrand as $|\mathrm{Re}\,(z)| \to \infty$, by Cauchy-Goursat theory, the contour can be *moved* to integration along the real line, giving

$$\int_{\mathbb{R}} e^{-\pi(x^2+i\xi)^2}\; dx \;=\; \int_{i\xi-\infty}^{i\xi+\infty} e^{-\pi z^2}\; dz \;=\; \int_{-\infty}^{\infty} e^{-\pi x^2}\; dx \;=\; 1$$

The fact that the latter integral has value 1 comes from the usual trick involving polar coordinates:

$$\Big(\int_{-\infty}^{\infty} e^{-\pi x^2}\; dx\Big)^2 \;=\; \int_{\mathbb{R}^2} e^{-\pi(x^2+y^2)}\; dx\,dy \;=\; \int_0^{2\pi}\int_0^{\infty} e^{-\pi r^2}\; r\,dr\,d\theta \;=\; 2\pi\int_0^{\infty} e^{-\pi r^2}\; r\,dr$$

Replacing r by $\sqrt{t}$, this is

$$\pi\int_0^{\infty} e^{\pi t}\; dt \;=\; \pi\cdot\frac{1}{\pi} \;=\; 1$$

Thus, with the present normalization of Fourier transform and corresponding normalization of Gaussian, the Gaussian is its own Fourier transform. ///

[12.29] Show that $\chi_{[a,b]} * \chi_{[c,d]}$ is a piecewise-linear function, and express it explicitly.

Discussion: Once enunciated, this fact (and the explicit expression) should be just a matter of book-keeping. We do assume that $a \le b$ and $c \le d$. Also, by symmetry, without loss of generality we can suppose that $|b-a| \ge |d-c|$. This is used in the treatment of cases below.

$$(\chi_{[a,b]} * \chi_{[c,d]})(x) \;=\; \int_{\mathbb{R}} \chi_{[a,b]}(x-y)\cdot\chi_{[c,d]}(y)\; dy \;=\; \int_c^d \chi_{[a,b]}(x-y)\; dy$$

$$=\; \int_c^d \chi_{[a-x,b-x]}(-y)\; dy \;=\; \int_{-d}^{-c} \chi_{[x-b,x-a]}(y)\; dy \;=\; \text{meas}\Big([-d,-c]\cap[x-b,x-a]\Big)$$

Looking at the cases of overlap, using $b-a \ge d-c$, this is

$$\begin{cases} 0 & (\text{for } x-a \le -d, \text{ that is, } [x-b,x-a] \text{ is to the left of } [-d,-c]) \\ (x-a)-(-d) & (\text{for } x-b \le -d \le x-a \le -c) \\ (-c)-(-d) & (\text{for } x-b \le -d \le -c \le x-a, \text{ that is, } [-d,-c] \subset [x-b,x-a]) \\ (-c)-(x-b) & (\text{for } -d \le x-b \le -c \le x-a) \\ 0 & (\text{for } x-b \ge -c, \text{ that is, } [x-b,x-a] \text{ is to the right of } [-d,-c]) \end{cases}$$

$$= \begin{cases} 0 & (\text{for } x \le a-d) \\ x-a+d & (\text{for } a-d \le x \le a-c) \\ d-c & (\text{for } a-c \le x \le b-d) \\ b-c-x & (\text{for } b-d \le x \le b-c) \\ 0 & (\text{for } x \ge b-c) \end{cases}$$

We used the fact that $b-a \ge d-c$ implies $a-c \le b-d$. It is useful to consider the special configuration $[a,b]=[-A,A]$ and $[c,d]=[-B,B]$ with $A \ge B \ge 0$: the convolution is

$$\begin{cases} 0 & (\text{for } x \le -A-B) \\ x+A+B & (\text{for } -A-B \le x \le -A+B) \\ 2B & (\text{for } -A+B \le x \le A-B) \\ A+B-x & (\text{for } A-B \le x \le A+B) \\ 0 & (\text{for } x \ge A+B) \end{cases}$$

In particular, the convolution is supported inside $[-A-B, A+B]$. Similarly, for f and g supported in $[-a,a]$ and $[-b,b]$, the convolution is supported in $[-a-b,\, a+b]$. ///

[12.30] Evaluate the *Borwein integral*

$$\int_{\mathbb{R}} \frac{\sin x}{x} \cdot \frac{\sin x/3}{x/3} \cdot \frac{\sin x/5}{x/5}\, dx$$

Discussion: View this as an inner product and invoke Plancherel:

$$\int_{\mathbb{R}} \frac{\sin x}{x} \cdot \frac{\sin x/3}{x/3} \cdot \frac{\sin x/5}{x/5}\, dx \;=\; \Big\langle \frac{\sin x}{x}, \frac{\sin x/3}{x/3} \cdot \frac{\sin x/5}{x/5} \Big\rangle \;=\; \Big\langle \Big(\frac{\sin x}{x}\Big)^{\frown}, \Big(\frac{\sin x/3}{x/3} \cdot \frac{\sin x/5}{x/5}\Big)^{\frown} \Big\rangle$$

Since Fourier transform converts pointwise multiplication to convolution, this is

$$\Big\langle \Big(\frac{\sin x}{x}\Big)^{\frown}, \Big(\frac{\sin x/3}{x/3}\Big)^{\frown} * \Big(\frac{\sin x/5}{x/5}\Big)^{\frown} \Big\rangle$$

We have computed that

$$\widehat{\chi_{[-a,a]}}(\xi) \;=\; \frac{\sin 2\pi a \xi}{\pi \xi} \;=\; 2a \cdot \frac{\sin 2\pi a\xi}{2\pi a \xi}$$

That is, by linearity of Fourier transform,

$$\Big(\frac{1}{2a}\chi_{[-a,a]}\Big)^{\frown}(\xi) \;=\; \frac{\sin(2\pi a)\xi}{(2\pi a)\xi}$$

By Fourier inversion, noting that $\frac{\sin x}{x}$ is not in L^1, only in L^2, so the inverse transform is not necessarily the literal integral,

$$\Big(\frac{\sin(2\pi a)\xi}{(2\pi a)\xi}\Big)^{\frown}(x) \;=\; \frac{1}{2a}\chi_{[-a,a]}(x)$$

Replacing a by $a/2\pi$ gives

$$\Big(\frac{\sin a\xi}{a\xi}\Big)^{\frown}(x) \;=\; \frac{\pi}{a}\,\chi_{[-\frac{a}{2\pi},\,\frac{a}{2\pi}]}(x)$$

We will use $a = 1, \frac{1}{3}$, and $\frac{1}{5}$. The relevant convolution was also computed above, but all we need is the fact that the support of

$$3\pi\,\chi_{[-\frac{1}{6\pi},\,\frac{1}{6\pi}]} * 5\pi\,\chi_{[-\frac{1}{10\pi},\,\frac{1}{10\pi}]}$$

is inside the interval $[-\frac{1}{6\pi} - \frac{1}{10\pi},\ \frac{1}{6\pi} + \frac{1}{10\pi}]$. Thus, the integral of three *sinc* functions is equal to

$$\int_{\mathbb{R}} \pi\chi_{[\frac{-1}{2\pi},\frac{1}{2\pi}]}(x)\cdot\Big(3\pi\chi_{[-\frac{1}{6\pi},\,\frac{1}{6\pi}]} * 5\pi\,\chi_{[-\frac{1}{10\pi},\,\frac{1}{10\pi}]}\Big)(x)\;dx \;=\; \pi\cdot 3\pi\cdot 5\pi\int_{-1/\pi}^{1/\pi}\Big(\chi_{[-\frac{1}{6\pi},\,\frac{1}{6\pi}]} * \chi_{[-\frac{1}{10\pi},\,\frac{1}{10\pi}]}\Big)(x)\;dx$$

$$=\; \pi\cdot 3\pi\cdot 5\pi\int_{\mathbb{R}}\Big(\chi_{[-\frac{1}{6\pi},\,\frac{1}{6\pi}]} * \chi_{[-\frac{1}{10\pi},\,\frac{1}{10\pi}]}\Big)(x)\;dx$$

since $[-1/2\pi,\ 1/2\pi]$ contains the support of the convolution. Observing that (invoking Fubini-Tonelli as necessary),

$$\int_{\mathbb{R}}(f*g)(x)\;dx \;=\; \int_{\mathbb{R}}\int_{\mathbb{R}} f(x-y)g(y)\;dx\,dy \;=\; \int_{\mathbb{R}}\int_{\mathbb{R}} f(x)g(y)\;dx\,dy \;=\; \int_{\mathbb{R}} f(x)\;dx\cdot\int_{\mathbb{R}} g(x)\;dy$$

the integral of the convolution is

$$\int_{\mathbb{R}}\chi_{[-\frac{1}{6\pi},\,\frac{1}{6\pi}]}\cdot\int_{\mathbb{R}}\chi_{[-\frac{1}{10\pi},\,\frac{1}{10\pi}]} \;=\; \frac{1}{3\pi}\cdot\frac{1}{5\pi}$$

Thus, the whole is

$$\pi\cdot 3\pi\cdot 5\pi\cdot\frac{1}{3\pi}\cdot\frac{1}{5\pi} \;=\; \pi$$

Similarly, the integral of $f_1 * \ldots f_n$ is the product of the integrals $\int f_i$. With the support of f_i inside $[-a_i, a_i]$, the support of the convolution is inside $[-a_1 - \ldots - a_n,\ a_1 + \ldots + a_n]$. Thus, since $\frac{1}{3} + \frac{1}{5} + \ldots + \frac{1}{13} < 1$, the same argument shows that

$$\int_{\mathbb{R}}\frac{\sin x}{x}\cdot\frac{\sin x/3}{x/3}\cdot\frac{\sin x/5}{x/2n+1}\;dx \;=\; \pi \qquad\qquad (\text{for } 2n+1 = 3,5,7,9,11,13)$$

but for $2n+1 = 15$, the support of the Fourier transform of $\frac{\sin x}{x}$ no longer contains the support of the convolution. ///

[12.31] Compute $e^{-\pi x^2} * e^{-\pi x^2}$ and $\dfrac{\sin x}{x} * \dfrac{\sin x}{x}$. (Be careful what you assert: $\frac{\sin x}{x}$ is not in $L^1(\mathbb{R})$.)

Discussion: The idea is to invoke $f * g = (\widehat{f}\cdot\widehat{g})^{\frown}$ for *even* functions $f, g \in L^1$, since for even functions the inverse Fourier transform is the same as the forward Fourier transform. Conveniently, Gaussians are in $L^1 \cap L^2$, and, from above, have Fourier transforms which are again Gaussians:

$$\widehat{e^{-\pi a x^2}}(\xi) \;=\; \frac{1}{\sqrt{a}}e^{-\pi\xi^2/a} \qquad\qquad (\text{for } a > 0)$$

so

$$e^{-\pi x^2} * e^{-\pi x^2}(\xi) \;=\; \widehat{e^{-\pi x^2} \cdot e^{-\pi x^2}}(\xi) \;=\; \widehat{e^{-2\pi x^2}}(\xi) \;=\; \frac{1}{\sqrt{2}} e^{-\pi \xi^2/2}$$

For the other example, the bound $|f * g|_{L^1} \le |f|_{L^p} \cdot |g|_{L^q}$ for conjugate exponents p, q shows that $f * g \in L^1$ for $f, g \in L^2$. Thus, the same identity holds for $f, g \in L^2$, with the Plancherel extension of Fourier transform. That is, $\widehat{f}$ and $\widehat{g}$ need not be the literal integrals for the Fourier transform, but its extension by continuity to L^2. Above, we computed the Fourier transform of characteristic functions of intervals:

$$\chi_{\widehat{[-a,a]}} a(\xi) \;=\; \frac{\sin 2\pi a \xi}{\pi \xi}$$

Thus,

$$\left(\pi \cdot \chi_{[-1/2\pi,\, 1/2\pi]}\right)\widehat{\ }\,(\xi) \;=\; \frac{\sin \xi}{\xi}$$

Then

$$\Big(\frac{\sin x}{x} * \frac{\sin x}{x}\Big)(\xi) \;=\; \Big((\pi \cdot \chi_{[-1/2\pi,\, 1/2\pi]}) \cdot (\pi \cdot \chi_{[-1/2\pi,\, 1/2\pi]})\Big)\widehat{\ }(\xi)$$

$$=\; \pi \cdot (\pi \cdot \chi_{[-1/2\pi,\, 1/2\pi]})\widehat{\ }(\xi) \;=\; \pi \cdot \frac{\sin \xi}{\xi}$$

///

[12.32] Prove that every $f \in C^o_c(\mathbb{R})$ can be uniformly approximated (in sup norm) arbitrarily well as superpositions of Gaussians: given $\varepsilon > 0$, there is $g \in C^o_c(\mathbb{R})$ and sufficiently large n such that

$$\sup_{x \in \mathbb{R}} \Big| f(x) - \int_{\mathbb{R}} g(\xi) \cdot n e^{-\pi n^2 (\xi - x)^2} \, d\xi \Big| \;<\; \varepsilon$$

Discussion: This is an instance of an *approximate identity* and the basic property of such. Namely, for an approximate identity $\{\varphi_n\}$ on $\mathbb{R}$ and $f \in C^o_c(\mathbb{R})$, we have

$$\sup_{x \in \mathbb{R}} \Big| f(x) - \int_{\mathbb{R}} \varphi_n(\xi) \cdot f(x + \xi) \, d\xi \Big| \;\longrightarrow\; 0 \qquad \text{(as } n \to +\infty)$$

By replacing ξ by $\xi - x$ in the integral, we have

$$\sup_{x \in \mathbb{R}} \Big| f(x) - \int_{\mathbb{R}} f(\xi) \cdot \varphi_n(\xi - x) \, d\xi \Big| \;\longrightarrow\; 0 \qquad \text{(as } n \to +\infty)$$

Rather than reproving this general assertion in the example at hand, we simply clarify the interpretation in terms of approximate identities. That is, with $\varphi_1(x) = e^{-\pi x^2}$, we that the sequence $\varphi_n(x) = n \cdot \varphi_1(nx)$ is an approximate identity. More generally, we prove

[12.0.6] Claim: Let $\varphi \in C^o(\mathbb{R})$ be a non-negative $\mathbb{R}$-valued function, with $\int_{\mathbb{R}} \varphi = 1$. Then $\varphi_n(x) = n \cdot \varphi(n \cdot x)$ is an approximate identity.

Proof: The non-negative real-valued-ness is of course immediate. The integral of φ_n is

$$\int_{\mathbb{R}} \varphi_n(x) \, dx \;=\; \int_{\mathbb{R}} n \cdot \varphi(n \cdot x) \, dx \;=\; \int_{\mathbb{R}} n \cdot \varphi(x) \, \frac{dx}{n} \;=\; \int_{\mathbb{R}} \varphi(x) \, dx \;=\; 1$$

by replacing x by x/n in the integral. Finally, to see that the masses of the φ_n bunch up near 0: Since $\varphi \ge 0$ and

$$\lim_n \int_{-\sqrt{n}}^{\sqrt{n}} \varphi(x) \, dx \;=\; \int_{\mathbb{R}} \varphi(x) \, dx \;=\; 1$$

given $\varepsilon > 0$ there is sufficiently large n_o such that for all $n \geq n_o$

$$1 \;\leq\; \lim_n \int_{-\sqrt{n}}^{\sqrt{n}} \varphi(x)\, dx \;>\; 1-\varepsilon$$

Then, by replacing x by x/n in the integral,

$$\int_{-\frac{1}{\sqrt{n}}}^{\frac{1}{\sqrt{n}}} \varphi_n(x)\, dx \;=\; \int_{-\frac{1}{\sqrt{n}}}^{\frac{1}{\sqrt{n}}} n\cdot\varphi(n\cdot x)\, dx \;=\; \int_{-\sqrt{n}}^{\sqrt{n}} \varphi(x)\, dx \;>\; 1-\varepsilon$$

The verifies the bunching-up property. ///

[12.33] Without worrying too much about identifying the finite, positive constant $\int_{\mathbb{R}} \frac{(\sin x)^2}{x^2}\, dx$, prove that, for given $f \in C_c^o(\mathbb{R})$, given $\varepsilon > 0$, there is sufficiently large n and a function $g \in C_c^o(\mathbb{R})$ such that

$$\sup_{x\in\mathbb{R}} \Big| f(x) - \int_{\mathbb{R}} g(\xi)\cdot \frac{(\sin n(x-\xi))^2}{(x-\xi)^2}\, d\xi \Big| \;<\; \varepsilon$$

Discussion: After the more general discussion of the previous example, this is just another such. ///

[12.34] Show that the *principal value* functional

$$f \;\longrightarrow\; PV\int_{\mathbb{R}} \frac{f(x)}{x}\, dx \;=\; \lim_{\varepsilon\to 0}\Big(\int_{-\infty}^{-\varepsilon} \frac{f(x)}{x}\, dx + \int_{\varepsilon}^{\infty} \frac{f(x)}{x}\, dx\Big)$$

is equal to

$$-\int_{\mathbb{R}} f'(x)\cdot \log|x|\, dx$$

for f *continuously* differentiable on $\mathbb{R}$, with hypotheses on the decay of f and f' at infinity.

Discussion: This is an exercise in careful integration by parts, in the course of which we discover reasonable hypotheses on f and f' so that the natural heuristic is a proof.

For fixed small $\varepsilon > 0$ and large $M > 0$, integration by parts gives

$$\int_{-M}^{-\varepsilon} \frac{f(x)}{x}\, dx + \int_{\varepsilon}^{M} \frac{f(x)}{x}\, dx$$

$$= \Big[f(x)\cdot\log|x|\Big]_{-M}^{-\varepsilon} + \Big[f(x)\cdot\log x\Big]_{\varepsilon}^{M} - \int_{-M}^{-\varepsilon} f'(x)\cdot\log|x|\, dx - \int_{\varepsilon}^{M} f'(x)\cdot\log x\, dx$$

The simplest way to make the boundary terms near $\pm\infty$ vanish is that they vanish *individually.* For example, it does not suffice that $f \in L^1(\mathbb{R})$ or $L^2(\mathbb{R})$, because such a hypothesis by itself does not assure that $f(x)\cdot\log|x|$ goes to 0 at $\pm\infty$, since f could have narrower-and-narrower spikes parading out to infinity. It is true that an additional condition on the derivative might promise this asymptotic behavior of f, but let's not be toooo clever. So, exactly require that $f(x)\cdot\log|x|$ goes to 0 at $\pm\infty$. In contrast, making

$$\lim_{\varepsilon\to 0^+} \Big(f(\varepsilon)\cdot\log|\varepsilon| - f(-\varepsilon)\cdot\log|\varepsilon|\Big) \;=\; 0$$

is the most subtle issue here. It cannot reasonably accomplished by having the individuals go to 0, unless we require some sort of vanishing of f at 0, which would be undesirable here. Here is where *differentiability* of

f at 0 can be used: a Taylor-Maclaurin expansion with remainder ensures that $f(\varepsilon) - f(-\varepsilon) = O(\varepsilon)$. Since $\lim_{\varepsilon \to 0^+} \varepsilon \log \varepsilon = 0$, the combination of these boundary terms does go to zero.

For the limit as $M \to +\infty$ of the individual integrals of $f'(x) \cdot \log |x|$ to exist in a simple fashion, it suffices that $\lim_{M\to\infty} \int_M^\infty f'(x) \cdot \log |x| \, dx = 0$ and $\lim_{M\to\infty} \int_{-\infty}^{-M} f'(x) \cdot \log |x| \, dx = 0$. The last question is about what it takes to make

$$\lim_{\varepsilon \to 0^+} \Big(\int_\varepsilon^\infty f'(x) \cdot \log |x| \, dx + \int_{-\infty}^{-\varepsilon} f'(x) \cdot \log |x| \, dx \Big) \;=\; \int_{\mathbb{R}} f'(x) \cdot \log |x| \, dx$$

Since $\log |x|$ is locally integrable, for f' being merely *essentially bounded* on some interval $[-\varepsilon_o, \varepsilon_o]$, e.g., *continuous*, the two individual integrals $\int_{\varepsilon'}^{\varepsilon} f'(x) \cdot \log |x| \, dx$ and $\int_{-\varepsilon}^{-\varepsilon'} f'(x) \cdot \log |x| \, dx$ for $0 < \varepsilon' < \varepsilon$ go to zero as $\varepsilon \to 0^+$. Thus, with these natural sufficient constraints, we have the indicated identity. ///

[12.35] Let $\psi_n(x) = e^{2\pi i n x}$. Let $\delta_{\mathbb{Z}}$ be the *Dirac comb*, that is, a periodic version of Dirac's δ, describable as having Fourier series

$$\delta_{\mathbb{Z}} \;=\; \sum_{n\in\mathbb{Z}} 1 \cdot \psi_n \qquad\qquad \text{(converging in } H^{-1}(\mathbb{T}) \text{ or even } H^{-\frac{1}{2}-\varepsilon}(\mathbb{T}) \text{ for all } \varepsilon > 0)$$

With $\lambda \notin \mathbb{R}$, show that the differential equation

$$u'' - \lambda \cdot u \;=\; \delta_{\mathbb{Z}}$$

has a periodic solution $u \in H^{\frac{3}{2}-\varepsilon}(\mathbb{T}) \subset C^o(\mathbb{T})$, using Fourier series, *by division.* Show that the equation $v'' - \lambda v = f$ is solved by

$$v(x) \;=\; \int_{\mathbb{T}} u(x-t)\, f(t)\, dt \;=\; \int_0^1 u(x-t)\, f(t)\, dt$$

Discussion: Using the *spectral* characterization of the $H^{-\frac{1}{2}-\varepsilon}(\mathbb{T})$ norm,

$$\Big| \sum_{n\in\mathbb{Z}} 1 \cdot \psi_n \Big|^2_{H^{-\frac{1}{2}-\varepsilon}} \;=\; \sum_{n\in\mathbb{Z}} |1|^2 \cdot (1+n^2)^{-\frac{1}{2}-\varepsilon}$$

which is convergent for all $\varepsilon > 0$, by comparison to $\sum_{n\neq 0} 1/n^2$. So that Fourier series converges in $H^{-\frac{1}{2}-\varepsilon}(\mathbb{T})$ and produces a *generalized function* there.

The extension by continuity of d/dx from $C^\infty(\mathbb{T}) \to C^\infty(\mathbb{T})$ to $\widetilde{\frac{d}{dx}} : H^s(\mathbb{T}) \to H^{s-1}(\mathbb{T})$ is continuous, by design. Similarly, $\widetilde{\frac{d^2}{dx^2}} : H^s(\mathbb{T}) \to H^{s-2}(\mathbb{T})$ is continuous. That is, since infinite sums are the corresponding limits of finite partial sums, this continuity means that termwise differentiation is correct. Let $u = \sum_n c_n \psi_n$, and solve, dropping the tilde from the notation,

$$\sum_{n\in\mathbb{Z}} 1 \cdot \psi_n \;=\; \delta_{\mathbb{Z}} \;=\; u'' - \lambda u \;=\; \sum_n c_n \Big(\frac{d^2}{dx^2} - \lambda \Big) \psi_n \;=\; \sum_n c_n (-4\pi^2 n^2 - \lambda) \cdot \psi_n$$

Equating coefficients, $c_n = 1/(-4\pi^2 n^2 - \lambda)$, for λ not equal to $-4\pi^2 n^2$ for integer n. Another easy estimate shows that this u has gained 2 Sobolev indices, so is in $H^{\frac{3}{2}-\varepsilon}(\mathbb{T})$.

By Sobolev imbedding/inequality, $H^s(\mathbb{T}) \subset C^o(\mathbb{T})$ for all $s > \frac{1}{2}$, so the solution is continuous (and, in fact, satisfies a further Lipschitz condition).

To see that $v'' - \lambda v = f$ is solved by

$$v(x) \;=\; \int_{\mathbb{T}} u(x-t)\, f(t)\; dt \;=\; \int_0^1 u(x-t)\, f(t)\; dt$$

take f such that $\widehat{f} \in \ell^1_{\mathbb{Z}}(\mathbb{T})$, meaning that $\sum_n |\widehat{f}(n)| < \infty$. A somewhat stronger, more intuitive assumption is that $f \in C^2(\mathbb{T})$, and then by integration by parts

$$\widehat{f''}(n) \;=\; \int_{\mathbb{T}} e^{-2\pi i n x}\, f''(x)\; dx \;=\; \int_{\mathbb{T}} (-2\pi i n)^2\, e^{-2\pi i n x}\, f(x)\; dx \;=\; (2\pi i n)^2 \cdot \widehat{f}(n)$$

(On the circle $\mathbb{T}$, and/or for $\mathbb{Z}$-periodic functions, there are no boundary terms in integration by parts.) We do not even to invoke Riemann-Lebesgue, since $|\widehat{f''}(n)|$ is *bounded*, so there is a constant C such that $|\widehat{f}(n)| \le C/n^2$, so $\widehat{f} \in \ell^1_{\mathbb{Z}}$.

Then Fubini-Tonelli assures the legitimacy of interchanging sum and limit: [16]

$$\int_{\mathbb{T}} u(x-t)\, f(t)\; dt \;=\; \int_{\mathbb{T}} \sum_m \frac{1}{-4\pi^2 m^2 - \lambda} e^{2\pi i m (x-t)} \cdot \sum_n \widehat{f}(n) e^{2\pi i n t}\; dt$$

$$=\; \sum_{m,n} \frac{\widehat{f}(n)}{-4\pi^2 m^2 - \lambda} \int_{\mathbb{T}} e^{2\pi i m (x-t)}\, e^{2\pi i n t}\; dt \;=\; \sum_n \frac{\widehat{f}(n)}{-4\pi^2 n^2 - \lambda} e^{2\pi i n x}$$

by mutual orthogonality of distinct exponentials (in every Sobolev space). By Riemann-Lebesgue, $\widehat{f}(n) \to 0$, so

$$\sum_n |\widehat{f}(n)|^2 \cdot (1+n^2)^s \;<\; \infty \qquad\qquad (\text{for any } s < -\tfrac{1}{2})$$

so $f \in H^{-1}(\mathbb{T})$, for example. Application of the (extension of) $\frac{d^2}{dx^2} - \lambda$ *termwise* (again, justified by continuity of the extension) produces the Fourier expansion of f. ///

[12.0.7] Remark: The latter example illustrates the utility of using *generalized functions* even in a discussion that seems not to refer to them: there was no need to *guess* the function $u(x-t)$ (sometimes called a *Green's function*) solving the differential equation, since we *solved* for it using the Fourier expansion of $\delta_{\mathbb{Z}}$ that only converges in $H^{-\frac{1}{2}-\varepsilon}(\mathbb{T})$.

[16] In fact, we will see later that for u continuous and f in any Sobolev space, the interchange is justified.

13. Banach Spaces

1. Basic definitions
2. Riesz' Lemma
3. Counter-examples for unique norm-minimizing element
4. Normed spaces of continuous linear maps
5. Dual spaces of normed spaces
6. Baire's theorem
7. Banach-Steinhaus/uniform-boundedness theorem
8. Open mapping theorem
9. Closed graph theorem
10. Hahn-Banach theorem

Many natural spaces of functions, such as $C^o(K)$ for K compact, and $C^k[a,b]$, have natural structures of Banach spaces.

Abstractly, Banach spaces are less convenient than Hilbert spaces, but still sufficiently simple so many important properties hold. Several standard results true in greater generality have simpler proofs for Banach spaces.

Riesz' lemma is an elementary result often an adequate substitute in Banach spaces for the lack of sharper Hilbert-space properties. We include natural counter-examples to the *minimum principle* valid in Hilbert spaces, but not generally valid in Banach spaces.

The Banach-Steinhaus/uniform-boundedness theorem, open mapping theorem, and closed graph theorem are not elementary, since they invoke the Baire category theorem. The Hahn-Banach theorem is non-trivial, but does *not* use completeness.

13.1 Basic Definitions

A real or complex [17] vectorspace V with a real-valued function, the *norm*,

$$| \, | : V \longrightarrow \mathbb{R}$$

with properties

$$|x+y| \le |x| + |y| \qquad \text{(triangle inequality)}$$

$$|\alpha x| = |\alpha| \cdot |x| \qquad (\alpha \text{ complex}, \ x \in V)$$

$$|x| = 0 \ \Rightarrow \ x = 0 \qquad \text{(positivity)}$$

is a *normed complex vectorspace*, or simply *normed space*. Because of the triangle inequality, the function

$$d(x,y) \;=\; |x-y|$$

is a *metric*. The *symmetry* comes from

$$d(y,x) \;=\; |y-x| \;=\; |(-1)\cdot(x-y)| \;=\; |-1|\cdot|x-y| \;=\; |x-y| \;=\; d(x,y)$$

When V is *complete* with respect to this metric, V is a *Banach space.*

---

[17] In fact, for many purposes, the scalars need not be $\mathbb{R}$ or $\mathbb{C}$, need not be locally compact, and need not even be commutative. The basic results hold for Banach spaces over non-discrete, complete, normed division rings. This allows scalars like the p-adic field $\mathbb{Q}_p$, or Hamiltonian quaternions $\mathbb{H}$, and so on.

Hilbert spaces are Banach spaces, but many natural Banach spaces are *not* Hilbert spaces, and may fail to enjoy useful properties of Hilbert spaces. *Riesz' lemma* below is sometimes a sufficient substitute.

Most norms on Banach spaces do *not* arise from inner products. Norms arising from inner products recover the inner product via the *polarization* identities

$$4\langle x,y\rangle \;=\; |x+y|^2 \;-\; |x-y|^2 \qquad \text{(real vector space)}$$

$$4\langle x,y\rangle \;=\; |x+y|^2 \;-\; |x-y|^2 \;+\; i|x+iy|^2 \;-\; i|x-iy|^2 \qquad \text{(complex vector space)}$$

Given a norm on a vector space, *if* the polarization expression gives an inner product, *then* the norm is produced by that inner product. However, checking whether the polarization expression is bilinar or hermitian, may be awkward or non-intuitive.

13.2 Riesz' Lemma

The following essentially elementary inequality is sometimes an adequate substitute for corollaries of the Hilbert-space minimum principle and its corollaries. Once one sees the proof, it is not surprising, but,

[13.2.1] Lemma: *(Riesz)* For a non-dense subspace X of a Banach space Y, given $r < 1$, there is $y \in Y$ with $|y| = 1$ and $\inf_{x\in X} |x-y| \geq r$.

Proof: Take y_1 not in the closure of X, and put $R = \inf_{x\in X} |x-y_1|$. Thus, $R > 0$. For $\varepsilon > 0$, let $x_1 \in X$ be such that $|x_1 - y_1| < R + \varepsilon$. Put $y = (y_1 - x_1)/|x_1 - y_1|$, so $|y| = 1$. And

$$\inf_{x\in X} |x-y| \;=\; \inf_{x\in X} \Big| x + \frac{x_1}{|x_1-y_1|} - \frac{y_1}{|x_1-y_1|} \Big| \;=\; \inf_{x\in X} \Big| \frac{x}{|x_1-y_1|} + \frac{x_1}{|x_1-y_1|} - \frac{y_1}{|x_1-y_1|} \Big|$$

$$=\; \frac{\inf_{x\in X} |x-y_1|}{|x_1-y_1|} \;=\; \frac{R}{R+\varepsilon}$$

By choosing $\varepsilon > 0$ small, $R/(R+\varepsilon)$ can be made arbitrarily close to 1. ///

13.3 Counter-examples to unique norm-minimizing element

The (true) *minimum principle* for Hilbert spaces is that a closed, convex subset has a unique element of minimum norm. This has many important elementary corollaries special to Hilbert spaces, such as existence of orthogonal complements to subspaces, and often fails for Banach spaces.

An important historical example of failure of functionals to attain their infs on closed, convex subsets of Banach spaces is the falsity of the *Dirichlet principle* as originally naively proposed. [18]

The (true) minimizing principle in a Hilbert space V is that, in a closed, convex, non-empty subset $E \subset V$, there is a unique element of least norm. As an example corollary, for non-dense subspace W of a Hilbert

---

[18] The *Dirichlet principle*, invoked by Riemann but observed by Weierstraß to be false as stated, would assert that a solution of $\Delta u = f$ on an open set Ω in $\mathbb{R}^n$, with boundary condition $u|_{\partial\Omega} = g$ on $\partial\Omega$, is a minimizer of the *energy integral*

$$E(u) \;=\; \int_\Omega \tfrac{1}{2}|\nabla u|^2 + |u|^2$$

on the Banach subspace of $C^2(\Omega)$ functions u satisfying $u|_{\partial\Omega} = g$. However, the infimum need not be attained in that Banach space. Hilbert justified Dirichlet's principle in certain circumstances. Beppo Levi (1906) observed that using energy integrals to form the norm (squared) of a pre-Hilbert space in $C^2(\Omega)$, and *completing* to a Hilbert space, *does* guarantee existence of a solution *in that Hilbert space.*

space V, there is $v \in V$ with $|v| = 1$ and $\inf_{w \in W} |v - w| = 1$, by taking v to be a unit-length vector in the *orthogonal complement* to W. This minimization property typically *fails* in Banach spaces, as follows.

[13.3.1] Example: *Many* minimizing elements can exist: in the Banach space $L^1[a,b]$, in the closed, convex subset $E = \{f : \int_a^b f = 1\}$, there are infinitely-many norm-minimizing elements.

[13.3.2] Example: In the Banach space $Y = C^o[0,2]$, with closed convex subset

$$E \;=\; \{f \in C^o[0,2] \;:\; \int_0^1 f(x)\,dx - \int_1^2 f(x)\,dx = 1\}$$

there is *no* norm-minimizing element. To this end, let

$$s(x) \;=\; \begin{cases} 1 & (\text{for } 0 \le x \le 1) \\ \\ -1 & (\text{for } 1 \le x \le 2) \end{cases}$$

and

$$\lambda(f) \;=\; \int_0^2 f(x) \cdot s(x)\,dx$$

Certainly $C^o[0,2] \subset L^2[0,2]$, so by Cauchy-Schwarz-Bunyakowsky,

$$|\lambda(f)| \;=\; |\langle f, s\rangle| \;\le\; |f|_{L^2[0,2]} \cdot |s|_{L^2[0,2]} \;=\; |f|_{L^2[0,2]} \cdot \sqrt{2}$$

with equality only for f a scalar multiple of s. Also, certainly

$$|f|_{L^2[0,2]} \;\le\; \Big(\int_0^2 |f|_Y^2\Big)^{\frac{1}{2}} \;=\; |f|_Y \cdot \sqrt{2} \qquad\qquad (\text{for } f \in C^o[0,2])$$

Since s is not continuous, non-zero $f \in Y$ is never a constant multiple of s, so Cauchy-Schwarz-Bunyakowsky gives a *strict* inequality

$$|\lambda(f)| \;<\; |f|_{L^2[0,2]} \cdot \sqrt{2} \;\le\; |f|_Y \cdot 2 \qquad\qquad (\text{for all } 0 \ne f \in Y)$$

Thus,

$$\frac{1}{2} \;<\; |f|_Y \qquad\qquad (\text{for } f \in E)$$

Yet it is easy to arrange continuous functions f with $\lambda(f) = 1$ and sup-norm $|f|_Y$ *approaching* $1/2$ from above, by approximating $\frac{1}{2}s(x)$ by continuous functions. For example, form a continuous, piecewise-linear function

$$g(x) \;=\; \begin{cases} \frac{1}{2} & (\text{for } 0 \le x \le 1-\varepsilon) \\ \\ \frac{1}{2} - \dfrac{x-(1-\varepsilon)}{2\varepsilon} & (\text{for } 1-\varepsilon \le x \le 1+\varepsilon) \\ \\ -\frac{1}{2} & (\text{for } 1+\varepsilon \le x \le 2) \end{cases}$$

The sup-norm of g is obviously $\frac{1}{2}$, and $\lambda(g) = 1 - \frac{1}{2}\varepsilon$. Thus, functions $f = g/(1-\frac{1}{2}\varepsilon)$ have $\lambda(f) = 1$ and sup norms approaching $\frac{1}{2}$ from above. This proves the claimed failure. ///

13.4 Normed spaces of linear maps

There is a *natural norm* on the set of continuous linear maps $T : X \to Y$ from one normed space X to another normed space Y. Even when X, Y are Hilbert spaces, the set of continuous linear maps $X \to Y$ is generally only a *Banach* space.

Let $\mathrm{Hom}^o(X, Y)$ denote[19] the collection of continuous linear maps from the normed vectorspace X to the normed vectorspace Y. Use the same notation $|\,|$ for the norms on both X and Y, since context will make clear which is meant.

A linear (not necessarily continuous) map $T : X \to Y$ from one normed space to another has *uniform operator norm*

$$|T| \;=\; |T|_{\text{uniform}} \;=\; \sup_{|x| \le 1} |Tx|$$

where we allow the value $+\infty$. Such T is called *bounded* if $|T| < +\infty$. There are several obvious variants of the expression for the uniform norm:

$$|T| \;=\; \sup_{|x| \le 1} |Tx| \;=\; \sup_{|x| < 1} |Tx| = \sup_{|x| \neq 0} \frac{|Tx|}{|x|}$$

[13.4.1] Proposition: For a linear map $T : X \to Y$ from one normed space to another, the following conditions are equivalent:
- T is *continuous.*
- T is *continuous at* 0.
- T is *bounded.*

Proof: First, show that continuity *at a point* x_o implies continuity *everywhere*. For another point x_1, given $\varepsilon > 0$, take $\delta > 0$ so that $|x - x_o| < \delta$ implies $|Tx - Tx_o| < \varepsilon$. Then for $|x' - x_1| < \delta$

$$|(x' + x_o - x_1) - x_o| \;<\; \delta$$

By linearity of T,

$$|Tx' - Tx_1| \;=\; |T(x' + x_o - x_1) - Tx_o| \;<\; \varepsilon$$

which is the desired continuity at x_1.

Now suppose that T is continuous at 0. For $\varepsilon > 0$ there is $\delta > 0$ so that $|x| < \delta$ implies $|Tx| < \varepsilon$. For $x \neq 0$,

$$\Big|\frac{\delta}{2|x|} x\Big| \;<\; \delta$$

so

$$\Big|T\frac{\delta}{2|x|} \cdot x\Big| \;<\; \varepsilon$$

Multiplying out and using the linearity, boundedness is obtained:

$$|Tx| \;<\; \frac{2\varepsilon}{\delta} \cdot |x|$$

[19] Another traditional notation for the collection of continuous linear maps from X to Y is $B(X, Y)$, where B stands for *bounded.*

Finally, prove that boundedness implies continuity at 0. Suppose there is C such that $|Tx| < C|x|$ for all x. Then, given $\varepsilon > 0$, for $|x| < \varepsilon/C$

$$|Tx| < C|x| < C \cdot \frac{\varepsilon}{C} \; = \; \varepsilon$$

which is continuity at 0. ///

The space $\mathrm{Hom}^o(X, Y)$ of continuous linear maps from one normed space X to another normed space Y has a natural structure of vectorspace by

$$(\alpha T)(x) \; = \; \alpha \cdot (Tx) \qquad \text{and} \qquad (S+T)x \; = \; Sx + Tx$$

for $\alpha \in \mathbb{C}$, $S, T \in \mathrm{Hom}^o(X, Y)$, and $x \in X$.

[13.4.2] Proposition: With the uniform operator norm, the space $\mathrm{Hom}^o(X, Y)$ of continuous linear operators from a normed space X to a *Banach* space Y is *complete*, whether or not X itself is complete. Thus, $\mathrm{Hom}^o(X, Y)$ is a Banach space.

Proof: Let $\{T_i\}$ be a Cauchy sequence of continuous linear maps $T : X \to Y$. Try defining the limit operator T in the natural fashion, by

$$Tx \; = \; \lim_i Tx_i$$

First, check that this limit exists. Given $\varepsilon > 0$, take i_o large enough so that $|T_i - T_j| < \varepsilon$ for $i, j > i_o$. By the definition of the uniform operator norm,

$$|T_i x - T_j x| \; < \; |x|\varepsilon$$

Thus, the sequence of values $T_i x$ is Cauchy in Y, so has a limit in Y. Call the limit Tx.

We need to prove that the map $x \to Tx$ is *continuous* and *linear*. The arguments are inevitable. Given $c \in \mathbb{C}$ and $x \in X$, for given $\varepsilon > 0$ choose index i so that for $j > i$ both $|Tx - T_j x| < \varepsilon$ and $|Tcx - T_j cx| < \varepsilon$. Then

$$|Tcx - cTx| \leq |Tcx - T_j cx| + |cT_j x - cTx| = |Tcx - T_j cx| + |c| \cdot |T_j x - Tx| < (1 + |c|)\varepsilon$$

This is true for every ε, so $Tcx = cTx$. Similarly, given $x, x' \in X$, for $\varepsilon > 0$ choose an index i so that for $j > i$ $|Tx - T_j x| < \varepsilon$ and $|Ty - T_j y| < \varepsilon$ and $|T(x+y) - T_j(x+y)| < \varepsilon$. Then

$$|T(x+y) - Tx - Ty| \leq |T(x+y) - T_j(x+y)| + |T_j x - Tx| + |T_j y - Ty| < 3\varepsilon$$

This holds for every ε, so $T(x+y) = Tx + Ty$.

For continuity, show that T is *bounded*. Choose an index i_o so that for $i, j \geq i_o$

$$|T_i - T_j| \leq 1$$

This is possible since the sequence of operators is Cauchy. For such i, j

$$|T_i - T_j x| \; \leq \; |x|$$

for all x. Thus, for $i \geq i_o$

$$|T_i x| \; \leq \; |(T_i - T_{i_o})x| + |T_{i_o} x| \; \leq \; |x|(1 + |T_{i_o}|)$$

Taking a limsup,

$$\limsup_i |T_i x| \; \leq \; |x|(1 + |T_{i_o}|)$$

This implies that T is bounded, and so is continuous.

Finally, we should see that $Tx = \lim_i T_i x$ is the operator-norm limit of the T_i. Given $\varepsilon > 0$, let i_o be sufficiently large so that $|T_i x - T_j x| < \varepsilon$ for all $i, j \geq i_o$ and for all $|x| \leq 1$. Then $|Tx - Tx_i| \leq \varepsilon$ and

$$\sup_{|x|\leq 1} |Tx - T_i x| \;\leq\; \sup_{|x|\leq 1} \varepsilon \;=\; \varepsilon$$

giving the desired outcome. ///

13.5 Dual spaces of normed spaces

This section considers an important special case of continuous linear maps between normed spaces, namely continuous linear maps from Banach spaces to *scalars*. All assertions are special cases of those for continuous linear maps to general Banach spaces, but deserve special attention.

For X a normed vectorspace with norm $|\,|$, a *continuous* linear map $\lambda : X \to \mathbb{C}$ is a *(continuous linear) functional* on X. Let

$$X^* \;=\; \mathrm{Hom}^o(X, \mathbb{C})$$

denote the collection of all such (continuous) functionals.

As more generally, for *any* linear map $\lambda : X \to \mathbb{C}$ of a normed vectorspace to $\mathbb{C}$, the *norm* $|\lambda|$ is

$$|\lambda| \;=\; \sup_{|x|\leq 1} |\lambda x|$$

where $|\lambda x|$ is the absolute value of the value $\lambda x \in \mathbb{C}$. We allow the value $+\infty$. Such a linear map λ is *bounded* if $|\lambda| < +\infty$.

As a special case of the corresponding general result:

[13.5.1] Corollary: For a k-linear map $\lambda : X \to k$ from a normed space X to k, the following conditions are equivalent:
- The map λ is *continuous.*
- The map λ is *continuous at one point.*
- The map λ is *bounded.*

Proof: These are special cases of the earlier proposition where the target was a general Banach space.
///

The *dual space*

$$X^* \;=\; \mathrm{Hom}^o(X, \mathbb{C})$$

of X is the collection of *continuous* linear functionals on X. This dual space has a natural structure of vectorspace by

$$(\alpha\lambda)(x) \;=\; \alpha \cdot (\lambda x) \qquad \text{and} \qquad (\lambda + \mu)x \;=\; \lambda x + \mu x$$

for $\alpha \in \mathbb{C}$, $\lambda, \mu \in X^*$, and $x \in X$. It is easy to check that the norm

$$|\lambda| \;=\; \sup_{|x|\leq 1} |\lambda x|$$

really is a norm on X^*, in that it meets the conditions

- *Positivity:* $|\lambda| \geq 0$ with equality only if $\lambda = 0$.
- *Homogeneity:* $|\alpha\lambda| = |\alpha| \cdot |\lambda|$ for $\alpha \in k$ and $\lambda \in X^*$. As a special case of the discussion of the uniform norm on linear maps, we have

[13.5.2] Corollary: The dual space X^* of a normed space X, with the natural norm, is a Banach space. That is, with respect to the natural norm on continuous functionals, it is *complete*. ///

13.6 Baire's theorem

Baire's theorem is not specifically about Banach spaces, as it applies more generally to complete metric spaces (and locally compact Hausdorff spaces), but it is the foundation for the subsequent basic non-trivial results on Banach spaces: uniform boundedness, open mapping, and closed graph theorems.

A set E in a topological space X is *nowhere dense* if its closure $\bar{E}$ contains no non-empty open set. A *countable union* of nowhere dense sets is said to be *of first category*, while every other subset (if any) is *of second category*. The idea (not at all clear from this traditional terminology) is that first category sets are *small*, while second category sets are *large*. In this terminology, the theorem's assertion is equivalent to the assertion that (non-empty) *complete metric* spaces and *locally compact Hausdorff* spaces are *of second category*.

A G_δ set is a countable intersection of open sets. Concommitantly, an F_σ set is a countable union of closed sets. Again, the following theorem can be paraphrased as asserting that, in a complete metric space, *a countable intersection of dense G_δ's is still a dense G_δ.*

[13.6.1] Theorem: *(Baire)* Let X be either a complete metric space or a locally compact Hausdorff topological space. The intersection of a *countable* collection $U_1, U_2, \ldots$ of *dense open subsets* U_i of X is still *dense* in X.

Proof: Let B_o be a non-empty open set in X, and show that $\bigcap_i U_i$ meets B_o. Suppose that we have inductively chosen an open ball B_{n-1}. By the denseness of U_n, there is an open ball B_n whose closure $\overline{B_n}$ satisfies

$$\overline{B_n} \subset B_{n-1} \cap U_n$$

Further, for complete metric spaces, take B_n to have radius less than $1/n$ (or any other sequence of reals going to 0), and in the locally compact Hausdorff case take B_n to have compact closure.

Let

$$K = \bigcap_{n \geq 1} \overline{B_n} \subset B_o \cap \bigcap_{n \geq 1} U_n$$

For complete metric spaces, the centers of the nested balls B_n form a Cauchy sequence (since they are nested and the radii go to 0). By completeness, this Cauchy sequence *converges*, and the limit point lies inside each *closure* $\overline{B_n}$, so lies in the intersection. In particular, K is non-empty. For locally compact Hausdorff spaces, the intersection of a nested family of non-empty compact sets is non-empty, so K is non-empty, and B_o necessarily meets the intersection of the U_n. ///

13.7 Banach-Steinhaus/uniform-boundedness theorem

This result is *non-trivial* in the sense that it uses the *Baire category theorem.*

[13.7.1] Theorem: *(Banach-Steinhaus/uniform boundedness)* For a family of continuous linear maps $T_\alpha : X \to Y$ from a Banach space X to a normed space Y, *either* there is a uniform bound $M < \infty$ so that $|T_\alpha| \leq M$ for all α, *or* there is $x \in X$ such that

$$\sup_\alpha \frac{|T_\alpha x|}{|x|} = +\infty$$

In the latter case, in fact, there is a dense G_δ of such x.

Proof: Let $p(x) = \sup_\alpha |T_\alpha x|$. We allow the possibility that $p(x) = +\infty$. Being the sup of continuous functions, p is *lower semi-continuous*: for each integer n, the set $U_n = \{x : p(x) > n\}$ is *open*.

On one hand, if every U_n is dense in X, by Baire category the intersection is dense, so is *non-empty*. By definition, it is a dense G_δ. On that set p is $+\infty$.

On the other hand, if one of the U_n is *not* dense, then there is a ball B of radius $r > 0$ about a point x_o which does not meet U_n. For $|x - x_o| < r$ and for all α

$$|T_\alpha(x - x_o)| \;\leq\; |T_\alpha x| + |T_\alpha x_o| \;\leq\; 2n$$

As $x - x_o$ varies over the open ball of radius r the vector $x' = (x - x_o)/r$ varies over the open ball of radius 1, and

$$|T_\alpha x'| \;=\; \left|T_\alpha \frac{(x - x_o)}{r}\right| \;\leq\; 2n/r$$

Thus, $|T_\alpha| \leq 2n/r$, which is the uniform boundedness. ///

13.8 Open mapping theorem

The open mapping theorem is non-trivial, since it invokes the Baire category theorem.

[13.8.1] Theorem: *(open mapping)* For a continuous linear *surjection* $T : X \to Y$ of Banach spaces, there is $\delta > 0$ such that for all $y \in Y$ with $|y| < \delta$ there is $x \in X$ with $|x| \leq 1$ such that $Tx = y$. In particular, T is an *open map*.

[13.8.2] Corollary: A *bijective* continuous linear map of Banach spaces is an *isomorphism*. ///

Proof: In the corollary the non-trivial point is that T is *open*, which is the point of the theorem. The linearity of the inverse is easy.

For every $y \in Y$ there is $x \in X$ so that $Tx = y$. For some integer n we have $n > |x|$, so Y is the union of the sets $TB(n)$, with usual open balls

$$B(n) \;=\; \{x \in X : |x| < n\}$$

By Baire category, the *closure* of some one of the sets $TB(n)$ contains a non-empty open ball

$$V \;=\; \{y \in Y : |y - y_o| < r\}$$

for some $r > 0$ and $y_o \in Y$. Since we are in a metric space, the conclusion is that every point of V occurs as the limit of a Cauchy sequence consisting of elements from $TB(n)$.

Certainly

$$\{y \in Y : |y| < r\} \;\subset\; \{y_1 - y_2 : y_1, y_2 \in V\}$$

Thus, every point in the ball B'_r of radius r centered at 0 in Y is the sum of two limits of Cauchy sequences from $TB(n)$. Thus, surely every point in B'_r is the limit of a single Cauchy sequence from the image $TB(2n)$ of the open ball $B(2n)$ of twice the radius. That is, the *closure* of $TB(2n)$ contains the ball $B'(r)$.

Using the linearity of T, the *closure* of $TB(\rho)$ contains the ball $B'(r\rho/2n)$ in Y.

Given $|y| < 1$, choose $x_1 \in B(2n/r)$ so that $|y - Tx_1| < \varepsilon$. Choose $x_2 \in B(\varepsilon \cdot \frac{2n}{r})$ so that

$$|(y - Tx_1) - Tx_2| \;<\; \varepsilon/2$$

Choose $x_3 \in B(\frac{\varepsilon}{2} \cdot \frac{2n}{r})$ so that

$$|(y - Tx_1 - Tx_2) - Tx_3| \ < \ \varepsilon/2^2$$

Choose $x_4 \in B(\frac{\varepsilon}{2^2} \cdot \frac{2n}{r})$ so that

$$|(y - Tx_1 - Tx_2 - Tx_3) - Tx_4| \ < \ \varepsilon/2^3$$

and so on. The sequence

$$x_1, \ x_1 + x_2, \ x_1 + x_2 + x_3, \ \ldots$$

is Cauchy in X. Since X is complete, the limit x of this sequence exists in X, and $Tx = y$. We find that

$$x \ \in \ B(\frac{2n}{r}) + B(\varepsilon\frac{2n}{r}) + B(\frac{\varepsilon}{2} \cdot \frac{2n}{r}) + B(\frac{\varepsilon}{2^2} \cdot \frac{2n}{r}) + \ldots \subset B((1 + 2\varepsilon)\frac{2n}{r})$$

Thus,

$$TB((1 + \varepsilon)\frac{2n}{r}) \ \supset \ \{y \in Y : |y| < 1\}$$

This proves open-ness at 0. ///

13.9 Closed graph theorem

The closed graph theorem uses the open mapping theorem, so invokes Baire category, so is non-trivial.

It is straightforward to show [20] that a *continuous* map $f : X \to Y$ of *Hausdorff* topological spaces has *closed graph*

$$\Gamma_f \ = \ \{(x, y) \ : \ f(x) = y\} \ \ \subset \ \ X \times Y$$

Similarly, a topological space X is Hausdorff if and only if the diagonal $X^\Delta = \{(x, x) : x \in X\}$ is closed in $X \times X$. [21]

[13.9.1] Theorem: A linear map $T : V \to W$ of Banach spaces is continuous if it has closed graph

$$\Gamma \ = \ \Gamma_T \ = \ \{(v, w) \ : \ Tv = w\}$$

Proof: It is routine to check that $V \times W$ with norm $|v \times w| = |v| \cdot |w|$ is a Banach space. Since Γ is a closed subspace of $V \times W$, it is a Banach space itself with the restriction of this norm.

The projection $\pi_V : V \times W \to V$ is a continuous linear map. The restriction $\pi_V|_\Gamma$ of π_V to Γ is still continuous, and still *surjective*, because it T is an everywhere-defined function on V. By the open mapping theorem, $\pi_V|_\Gamma$ is *open*. Thus, the bijection $\pi_V|_\Gamma$ is a *homeomorphism*. Letting $\pi_W : V \times W \to W$ be the projection to W,

$$T \ = \ \pi_W \circ \big(\pi_V|\Gamma\big)^{-1} \ \ : \ \ V \longrightarrow \ W$$

expresses T as a composite of continuous functions. ///

[20] To show that a continuous map $f : X \to Y$ of topological spaces with Y Hausdorff has closed graph Γ_f, show the complement is open. Take $(x, y) \notin \Gamma_f$. Let V_1 be a neighborhood of $f(x)$ and V_2 a neighborhood of y such that $V_1 \cap V_2 = \phi$, using Hausdorff-ness. By continuity of f, for x' in a suitable neighborhood U of x, the image $f(x')$ is inside V_1. Thus, the neighborhood $U \times V_2$ of (x, y) does not meet Γ_f.

[21] To show that closed-ness of the diagonal X^Δ in $X \times X$ implies X is Hausdorff, let $x_1 \neq x_2$ be points in X. Then there is a neighborhood $U_1 \times U_2$ of (x_1, x_2), with U_i a neighborhood of x_i, not meeting the diagonal. That is, $(x, x') \in U_1 \times U_2$ implies $x \neq x'$. That is, $U_1 \cap U_2 = \phi$.

[13.9.2] Remark: The proof introduced two readily verifiable, useful ideas: a product of Banach spaces is a Banach space, and a closed vector subspace of a Banach space is a Banach space.

13.10 Hahn-Banach Theorem

Hahn-Banach does *not* use completeness, much less Baire category. The salient feature is *convexity*, and the scalars must be $\mathbb{R}$ or $\mathbb{C}$. Indeed, the Hahn-Banach theorem seems to be a result about *real* vectorspaces. Note that a $\mathbb{C}$-vectorspace may immediately be considered as a $\mathbb{R}$-vectorspace simply by forgetting some of the structure.

For Y a vector subspace of X, and for $S : Y \to Z$ a linear map to another vectorspace Z, a linear map $T : X \to Z$ is an *extension* of S to X when the restriction $T|_Y$ of T to Y is S.

[13.10.1] Theorem: *(Hahn-Banach)* Let X be a normed vectorspace with scalars $\mathbb{R}$ or $\mathbb{C}$, Y be a subspace, and λ be a continuous linear functional on Y. Then there is an extension Λ of λ to X such that

$$|\Lambda| \;=\; |\lambda|$$

[13.10.2] Corollary: Given $x \neq y$ in a normed space X, neither a scalar multiple of the other, there is a continuous linear functional λ on X so that $\lambda x = 1$ while $\lambda y = 0$. ///

[13.10.3] Corollary: Let Y be a closed subspace of a normed space X, and $x_o \notin Y$. Then there is a continuous linear functional λ on X which is 0 on Y, has $|\lambda| = 1$, and $\lambda(x_o) = |x_o|$. ///

Proof: We treat the case that the scalars are $\mathbb{R}$, and reduce the complex case to this.

The critical part is to extend a linear functional by just one dimension. That is, for given $x_o \notin Y$ make an extension λ' of λ to $Y' = Y + \mathbb{R}x_o$. Every vector in Y' has a unique expression as $y + cx_o$ with $c \in \mathbb{R}$, so define functionals by

$$\mu(y + cx_o) \;=\; \lambda y + c\ell \qquad \text{(for arbitrary } \ell \in \mathbb{R})$$

The issue is to choose ℓ so that $|\mu| = |\lambda|$.

Certainly $\lambda = 0$ is extendable by $\Lambda = 0$, so we consider the case that $|\lambda| \neq 0$. We can divide by $|\lambda|$ to suppose that $|\lambda| = 1$.

The condition $|\mu| = |\lambda|$ is a condition on ℓ:

$$|\lambda y + c\ell| \;\le\; |y + cx_o| \qquad \text{(for every } y \in Y)$$

We have simplified to the situation that we know this *does* hold for $c = 0$. So for $c \neq 0$, divide through by $|c|$ and replace $y \in Y$ by cy, so that the condition becomes

$$|\lambda y + \ell| \;\le\; |y + x_o| \qquad \text{(for every } y \in Y)$$

Replacing y by $-y$, the condition on ℓ is that

$$|\ell - \lambda y| \;\le\; |y - x_o| \qquad \text{(for every } y \in Y)$$

For a single $y \in Y$, the condition on ℓ is that

$$\lambda y - |y - x_o| \;\le\; \ell \le \lambda y + |y - x_o|$$

To have a common solution ℓ, it is exactly necessary that every *lower* bound be less than every *upper* bound. To see that this is so, start from

$$\lambda y_1 - \lambda y_2 \ = \ \lambda(y_1 - y_2) \ \leq \ |\lambda(y_1 - y_2)| \ \leq \ |y_1 - y_2| \ \leq \ |y_1 - x_o| + |y_2 - x_o|$$

by the triangle inequality. Subtracting $|y_1 - x_o|$ from both sides and adding λy_2 to both sides,

$$\lambda y_1 - |y_1 - x_o| \ \leq \ \lambda y_2 + |y_2 - x_o|$$

as desired. That is, we have proven the existence of at least one extension from Y to $Y' = Y + \mathbb{R}x_o$ with the same norm.

An equivalent of the Axiom of Choice will extend to the *whole* space while preserving the norm, as follows. Consider the set of pairs (Z, ζ) where Z is a subspace containing Y and ζ is a continuous linear functional on Z extending λ and with $|\zeta| \leq 1$. *Order* these by

$$(Z, \zeta) \ \leq \ (Z', \zeta')$$

when $Z \subset Z'$ and ζ' extends ζ. For a totally ordered collection (Z_α, ζ_α) of such,

$$Z' \ = \ \bigcup_\alpha Z_\alpha$$

is a subspace of X. In general, of course, the union of a family of subspaces would not be a subspace, but these are *nested*.

We obtain a continuous linear functional ζ' on this union Z', extending λ and with $|\zeta'| \leq 1$, as follows. Any *finite* batch of elements already occur inside some Z_α. Given $z \in Z'$, let α be any index large enough so that $z \in Z_\alpha$, and put

$$\zeta'(z) \ = \ \zeta_\alpha(z)$$

The family is totally ordered, so the choice of α does not matter so long as it is sufficiently large. Certainly for $c \in R$

$$\zeta'(cz) \ = \ \zeta_\alpha(cz) \ = \ c\zeta_\alpha(z) \ = \ c\zeta'(z)$$

For z_1 and z_2 and α large enough so that *both* z_1 and z_2 are in Z_α,

$$\zeta'(z_1 + z_2) \ = \ \zeta_\alpha(z_1 + z_2) \ = \ \zeta_\alpha(z_1) + \zeta_\alpha(z_2) \ = \ \zeta'(z_1) + \zeta'(z_2)$$

proving linearity. Thus, there is a *maximal* pair (Z', ζ'). The earlier argument shows that Z' must be all of X, since otherwise we could construct a further extension, contradicting the maximality. *This completes the proof for the case that the scalars are the real numbers.*

To reduce the complex case to the real case, the main trick is that, for λ_o a *real*-linear *real*-valued functional, the functional

$$\lambda x \ = \ \lambda_o(x) - i\lambda(ix)$$

is *complex*-linear, and has the *same norm* as λ_o. In particular, when

$$\lambda_o(x) \ = \ \mathrm{Re}\,\lambda(x) \ = \ \frac{\lambda x + \overline{\lambda x}}{2}$$

is the real part of λ we recover λ itself by this formula.

Granting this, given λ on a complex subspace, take its real part λ_o, a real-linear functional, and extend λ_o to a real-linear functional Λ_o with the same norm. Then the desired extension of λ is

$$\Lambda x \ = \ \Lambda_o(x) - i\Lambda(ix)$$

proving the theorem in the complex case.

Consider the construction

$$\lambda x \;=\; \lambda_o(x) - i\lambda(ix)$$

Since $\lambda_o(x+y) = \lambda_o x + \lambda_o y$ it follows that λ also has this additivity property. For a, b real,

$$\lambda((a+bi)x) \;=\; \lambda_o((a+bi)x) - i\lambda_o(i(a+bi)x) \;=\; \lambda_o(ax) + \lambda_o(ibx) - i\lambda_o(iax) - i\lambda_o(-bx)$$

$$=\; a\lambda_o x + b\lambda_o(ix) - ia\lambda_o(ix) + ib\lambda_o x = (a+bi)\lambda_o x - i(a+bi)\lambda_o(ix) = (a+bi)\lambda(x)$$

This gives the linearity.

Regarding the norm: since λ_o is real-valued, always

$$|\lambda_o(x)| \;\leq\; \sqrt{\lambda_o(x)^2 + \lambda_o(ix)^2} \;=\; |\lambda x|$$

On the other hand, given x there is a complex number μ of absolute value 1 so that $\mu\lambda(x) = |\lambda x|$. And

$$\lambda_o(x) \;=\; \lambda(x) + \overline{\lambda(x)}$$

Then

$$|\lambda(x)| \;=\; \mu\lambda(x) \;=\; \lambda(\mu x) \;=\; \lambda_o(\mu x) - i\lambda_o(i\mu x)$$

Since the left-hand side is real, and since λ_o is real-valued, $\lambda_o(\mu x) = 0$. Thus,

$$|\lambda(x)| \;=\; \lambda_o(\mu x)$$

Since $|\mu x| = |x|$, we have equality of norms of the functionals λ_o and λ. This completes the justification of the reduction of the complex case to the real case. ///

14. Basic applications of Banach space ideas

14.1 A good trick using uniform boundedness

The following sort of claim may seem nearly obviously true, but there is a missing key ingredient:

[14.1.1] Claim: Let $b = (b_1, b_2, \ldots)$ be a sequence of complex numbers such that $\sum_n b_n c_n$ is convergent for every $c = (c_1, c_2, \ldots) \in \ell^2$. Then $b \in \ell^2$.

Proof: Notably, the assumumption that the indicated sums are finite (convergent) does not directly give enough information to conclude that the map $\lambda(c) = \sum_n b_n c_n$ is a *continuous* linear functional on ℓ^2. The uniform boundedness theorem is needed to reach this conclusion.

Namely, let $\lambda_N(c) = \sum_{n \le N} b_n c_n$. These functionals *are* continuous on ℓ^2. By uniform boundedness, *either* there is a uniform bound $\beta < +\infty$ such that $\sup_N |\lambda_N(c)| \le \beta \cdot |c|$ for all $c \in \ell^2$, *or* there is a dense (hence, non-empty) G_δ such that $\sup_N |\lambda_N(c)|/|c| = +\infty$. But the assumption is that all the latter sups are finite. Thus, there must be a *uniform* bound, so $\lambda(c) = \sum_n b_n c_n$ *is* a continuous linear functional. By Riesz-Fréchet, it is given by an element of ℓ^2. ///

[14.1.2] Remark: If we know that the dual of L^p is L^q for σ-finite measure spaces X, then the same sort of argument applies.

14.2 Fourier series of C^o functions can diverge

The *density* of finite Fourier series in $C^o(\mathbb{T})$ makes no claim about *which* finite Fourier series approach a given $f \in C^o(\mathbb{T})$. Indeed, the density proof given via the Féjer kernel uses finite Fourier series quite distinct from the finite partial sums of the Fourier series of f itself, namely,

$$N^{th} \text{ Féjer sum } = \frac{1}{N} \sum_{|n| \le N} (N - |n|) \cdot \widehat{f}(n) \cdot e^{2\pi i n x}$$

The Banach-Steinhaus/uniform-boundedness theorem has a decisive corollary about convergence failure of Fourier series of $C^o(\mathbb{T})$ functions:

[14.2.1] Corollary: There is $f \in C^o(\mathbb{T})$ whose Fourier series

$$\sum_{n \in \mathbb{Z}} \widehat{f}(n)\, e^{inx} \qquad\qquad (\text{with } \widehat{f}(n) = \tfrac{1}{2\pi} \int_0^{2\pi} e^{-inx}\, f(x)\, dx)$$

diverges at 0. In fact, the divergence can be arranged for a dense G_δ of continuous functions, and at any given countable set of points on $\mathbb{T}$.

Proof: To invoke Banach-Steinhaus, consider the functionals given by partial sums of the Fourier series of f, evaluated at 0:

$$\lambda_N(f) = \sum_{|n| \le N} \hat{f}(n) = \sum_{|n| \le N} \hat{f}(n) \cdot e^{2\pi i n \cdot 0}$$

There is an easy upper bound

$$|\lambda_N(f)| \;\le\; \int_0^1 \Big|\sum_{|n|\le N} e^{-2\pi inx}\Big| \cdot \big|f(x)\big|\,dx \;\le\; |f|_{C^o} \cdot \int_0^1 \Big|\sum_{|n|\le N} e^{-2\pi inx}\Big|\,dx \;=\; |f|_{C^o} \cdot \Big|\sum_{|n|\le N} e^{-2\pi inx}\Big|_{L^1(\mathbb{T})}$$

We will show that equality holds, namely, that

$$|\lambda_N| \;=\; \Big|\sum_{|n|\le N} e^{-2\pi inx}\Big|_{L^1}$$

and show that the latter L^1-norms go to ∞ as $N \to \infty$.

Summing the finite geometric series and rearranging:

$$\sum_{|n|\le N} e^{-2\pi inx} \;=\; \frac{e^{-2\pi iNx} - e^{-2\pi i(-N-1)x}}{e^{-2\pi ix} - 1} \;=\; \frac{e^{2\pi i(N+\frac{1}{2})x} - e^{-2\pi i(N+\frac{1}{2})x}}{e^{\pi ix} - e^{-\pi ix}} \;=\; \frac{\sin 2\pi(N+\frac{1}{2})x}{\sin\frac{2\pi x}{2}}$$

The elementary inequality $|\sin t| \le |t|$ gives a lower bound

$$\int_0^1 \left|\frac{\sin 2\pi(N+\frac{1}{2})x}{\sin\frac{2\pi x}{2}}\right| dx \;\ge\; \int_0^1 \Big|\sin 2\pi(N+\tfrac{1}{2})x\Big| \cdot \frac{2}{2\pi x}\,dx \;=\; \int_0^{2\pi(N+\frac{1}{2})} |\sin x| \cdot \frac{2}{2\pi x}\,dx$$

$$\ge\; \sum_{\ell=1}^N \frac{1}{\pi\ell} \int_{2\pi(\ell-1)}^{2\pi\ell} |\sin x|\,dx \;\ge\; \sum_{\ell=1}^N \frac{1}{\pi\ell} \;\longrightarrow\; +\infty \qquad\qquad (\text{as } N \to \infty)$$

Thus, the L^1-norms do go to ∞.

We claim that the norm of the *functional* is the L^1-norm of the *kernel:* let $g(x)$ be the *sign* of the Dirichlet kernel

$$\sum_{|n|\le N} e^{-2\pi inx} \;=\; \frac{\sin 2\pi(N+\frac{1}{2})x}{\sin\frac{2\pi x}{2}}$$

Let g_j be a sequence of periodic continuous functions with $|g_j| \le 1$ and going to g pointwise. By *dominated convergence*

$$\lim_j \lambda_N(g_j) \;=\; \lim_j \int_0^1 g_j(x) \sum_{|n|\le N} e^{-2\pi inx}\,dx \;=\; \int_0^1 g(x) \sum_{|n|\le N} e^{-2\pi inx}\,dx \;=\; \int_0^1 |\sum_{|n|\le N} e^{-2\pi inx}|\,dx$$

By Banach-Steinhaus for the Banach space $C^o(\mathbb{T})$, since (as demonstrated above) there is *no* uniform bound $|\lambda_N| \le M$ for all N, there *exists* f in the unit ball of $C^o(\mathbb{T})$ such that

$$\sup_N |\lambda_N\, v| \;=\; +\infty$$

In fact, the collection of such v is *dense* in the unit ball, and is an intersection of a *countable* collection of dense open sets (a G_δ). That is, the Fourier series of f does not converge at 0.

The result can be strengthened by using Baire's theorem again. For a dense countable set of points x_j in the interval, let $\lambda_{j,N}$ be the continuous linear functionals on $C^o(\mathbb{T})$ defined by evaluation of finite partial sums of the Fourier series at x_j's:

$$\lambda_{j,N}(f) \;=\; \sum_{|n|\le N} \hat{f}(n)\, e^{2\pi inx_j}$$

As in the previous, the set E_j of functions f where

$$\sup_N |\lambda_{j,N} f| \;=\; +\infty$$

is a dense G_δ, so the intersection $E = \bigcap_j E_j$ is a dense G_δ, and, in particular, not empty. ///

14.3 Riemann-Lebesgue for $f \to \widehat{f}$ on $L^1(\mathbb{T})$ and $L^1(\mathbb{R})$

The space c_o of two-sided sequences *vanishing at infinity* is

$$c_o \;=\; \{\{a_n : n \in \mathbb{Z}\} \;:\; \lim_{|n|\to\infty} a_n = 0\}$$

The space c_o is a Banach space with norm $|\{a_n\}|_{c_o} = \sup_n |a_n|$. Parametrizing the circle $\mathbb{T}$ by the interval $[0,1]$ by the exponential map $x \to e^{2\pi i x}$, the Banach space $L^1(\mathbb{T}) = L^1[0,1]$ is measurable functions f on $[0,1]$ with finite integrals $\int_0^1 |f|$ (modulo the equivalence relation of equality almost everywhere). The space $L^1[0,1]$ contains and is strictly larger than $L^2[0,1]$. On $L^2[0,1]$, Fourier transform is an isometry to $\ell^2(\mathbb{Z})$, by Parseval's theorem, and a relatively trivial form of a Riemann-Lebesgue lemma is that $\widehat{f} \in c_o$ for $f \in L^2[0,1]$. The version for L^1 is less trivial:

[14.3.1] Lemma: *(Riemann-Lebesgue)* $\widehat{f} \in c_o$ for $f \in L^1(\mathbb{T})$.

Proof: Finite linear combinations of exponentials are dense in $C^o(\mathbb{T})$, for example by Féjer's argument, and $C^o(\mathbb{T})$ is dense in $L^1(\mathbb{T})$, essentially by the definition of integral and Urysohn's lemma. Thus, given $f \in L^1$ there is $g \in C^o(\mathbb{T})$ such that $|f - g|_{L^1} < \varepsilon$ and a finite linear combination h of exponentials such that $|g - h|_{C^o} < \varepsilon$. Then $|f - h|_{L^1} < 2\pi \cdot 2\varepsilon$.

Given such h, for large-enough n the Fourier coefficients are 0, by orthogonality of distinct exponentials. Thus,

$$|\widehat{f}(n)| \;=\; \frac{1}{2\pi}\Big| \int_0^{2\pi} (f(x) - h(x))\, e^{-inx}\, dx \Big| \;\le\; \frac{|f - h|_{L^1}}{2\pi} \;<\; 2\varepsilon \qquad \text{(for } n \text{ large, depending on } f\text{)}$$

This proves this Riemann-Lebesgue Lemma. ///

14.4 Non-surjection of $L^1[0,1] \to c_o$ by $f \to \widehat{f}$

Baire theorem and open mapping prove this.

[14.4.1] Corollary: *(of Baire and Open Mapping) Not* every sequence in c_o is the collection of Fourier coefficients of an $L^1(\mathbb{T})$ function.

Proof: The Fourier-coefficient map

$$Tf \;=\; \{\widehat{f}(n) : n \in \mathbb{Z}\} \;\in\; c_0$$

does map $L^1[0,1] \to c_o$, by Riemann-Lebesgue. The obvious inequality

$$|\widehat{f}(n)| \;=\; \Big| \int_0^1 f(x)\, e^{-2\pi i n x}\, dx \Big| \;\le\; \int_0^1 |f(x)|\, dx \;=\; |f|_{L^1}$$

shows $|T| \le 1$, so T is continuous. Taking $f(x) = 1$ shows $|T| = 1$.

The density of finite Fourier series in C^o and density of C^o in L^1, as in the proof of the Riemann-Lebesgue lemma, shows that T is injective. If T were also *surjective*, then the open mapping theorem would guarantee $\delta > 0$ such that for every L^1 function f

$$|\widehat{f}|_{\sup} \;\geq\; \delta \cdot |f|_{L^1}$$

However, this is impossible: with

$$f_N(x) \;=\; \sum_{|n|\leq N} e^{-2\pi i n x}$$

the sup norm of $\widehat{f}_N$ is certainly 1, yet the computation about divergence of Fourier series above shows that the L^1 norm of f_N goes to ∞ like $\log N$ as $N \to +\infty$. Thus, there is no such $\delta > 0$. Thus, T cannot be surjective. ///

14.5 $C^\infty(\mathbb{T})$ is dense in $C^o(\mathbb{T})$

Féjer's argument proves that the Cesaro-summed finite partial sums of Fourier series of a continuous function converge to that function in the $C^o(\mathbb{T})$ topology (that is, uniformly poinwise). These finite partial sums, as well as their Cesaro-summed forms, are in $C^\infty(\mathbb{T})$. Thus,

[14.5.1] Corollary: $C^\infty(\mathbb{T})$ is dense in $C^o(\mathbb{T})$. ///

14.6 Typical C^o functions are nowhere differentiable

[14.6.1] Claim: In $C^o[a,b]$, there is (at least) a dense G_δ of functions which at every point fail to be differentiable.

Proof: Anticipating the application of Baire's theorem, we present everywhere-not-differentiable functions as a countable intersection of dense opens. First, for fixed large $n > 0$ and small $h \neq 0$, let

$$X_{n,h} \;=\; \{f \in C^o[a,b] : |f(x+h) - f(x)| > n \cdot |h|, \text{ for all } x \in [a,b] \text{ such that } x+h \in [a,b]\}$$

To show that $X_{n,h}$ is open, we observe that for a given $f \in X_{n,h}$, the function $|f(x+h) - f(x)| - n \cdot |h|$ is continuous in x, and is positive. Thus, since the function is continuous on the compact interval $[a,b]$, its inf is strictly positive. Thus, for g with $|g - f|_{C^o}$ sufficiently small, $|g(x+h) - g(x)| - n \cdot |h|$ is still positive. That is, $g \in X_{n,h}$.

Next, each union

$$Y_{n,h} \;=\; \bigcup_{h' \neq 0,\; |h'|<|h|} X_{n,h'}$$

$$=\; \{f \in C^o[a,b] : \text{for every } x \in [a,b], \text{ there is } 0 < h' < h \text{ such that } |f(x+h') - f(x)| > n \cdot |h'|\}$$

(where implicitly $x + h' \in [a,b]$) is a union of opens, so is open.

Density of $Y_{n,h}$ in $C^o[a,b]$ is that, for given $f \in C^o[a,b]$, there is $g \in Y_{n,h}$ near f. To prove this, first approximate f to within $\varepsilon > 0$ in sup norm by $g \in C^1[a,b]$. Among the several possible ways to do this, we choose the following. First, adjust f by subtracting a polynomial to make $f(a) = f(b)$. Extending f by periodicity, Féjer's Cesaro-summed version of the finite partial sums of its Fourier series converge to it in sup norm. These finite approximations are all C^∞, in fact, proving that we can approximate f to within $\varepsilon > 0$ in sup norm by a C^1 function g.

In particular, the derivative of g is a continuous function on $[a,b]$, so is *bounded* in absolute value, say by β.

Next, we use auxiliary piecewise-C^1 functions $\varphi_{N,\varepsilon}$ in $C^o[a,b]$ with sup norms less than a given $\varepsilon > 0$, but with absolute values of derivatives strictly greater than a given N, for any pair ε, N. For example, we can easily make piecewise-*linear* continuous functions $\varphi_{N,\varepsilon}$ with slopes $\pm(N+1)$, changing sign so often that they stay strictly between $\pm\varepsilon$. For $N > \beta$, $g + \varphi_{2N,\varepsilon}$ is in $Y_{N,h}$ for all $h > 0$, and

$$|f - (g + \varphi_{2N,\varepsilon})|_{C^o} \ \leq\ |f - g|_{C^o} + |\varphi_{2N,\varepsilon}|_{C^o} \ <\ \varepsilon + \varepsilon$$

This proves the density of every open $Y_{N,h}$ in $C^o[a,b]$.

By Baire's theorem, the countable intersection $\bigcap_{n=1,2,\ldots} Y_{n,\frac{1}{n}}$ of dense compacts is still dense. ///

15. Banach spaces $C^k[a,b]$

1. Banach spaces $C^k[a,b]$
2. Non-Banach limit $C^\infty[a,b]$ of Banach spaces $C^k[a,b]$

We specify natural topologies, in which differentiation or other natural operators are *continuous*, and so that the space is *complete*.

Many familiar and useful spaces of continuous or differentiable functions, such as $C^k[a,b]$, have natural metric structures, and are *complete*. In these cases, the metric $d(,)$ comes from a *norm* $|\cdot|$, on the functions, giving Banach spaces.

Other natural function spaces, such as $C^\infty[a,b]$, are *not* Banach, but still do have a metric topology and are complete: these are *Fréchet spaces*, appearing as (projective) *limits* of Banach spaces, as below. These lack some of the conveniences of Banach spaces, but their expressions as *limits* of Banach spaces is often sufficient.

15.1 Banach spaces $C^k[a,b]$

We give the vector space $C^k[a,b]$ of k-times continuously differentiable functions on an interval $[a,b]$ a metric which makes it *complete*. Mere *pointwise* limits of continuous functions easily fail to be continuous. First recall the standard

[15.1.1] Claim: The set $C^o(K)$ of complex-valued continuous functions on a compact set K is *complete* with the metric $|f-g|_{C^o}$, with the C^o-*norm* $|f|_{C^o} = \sup_{x\in K} |f(x)|$.

Proof: This is a typical three-epsilon argument. To show that a Cauchy sequence $\{f_i\}$ of continuous functions has a *pointwise* limit which is a continuous function, first argue that f_i has a pointwise limit at every $x \in K$. Given $\varepsilon > 0$, choose N large enough such that $|f_i - f_j| < \varepsilon$ for all $i, j \geq N$. Then $|f_i(x) - f_j(x)| < \varepsilon$ for any x in K. Thus, the sequence of values $f_i(x)$ is a Cauchy sequence of complex numbers, so has a limit $f(x)$. Further, given $\varepsilon' > 0$ choose $j \geq N$ sufficiently large such that $|f_j(x) - f(x)| < \varepsilon'$. For $i \geq N$

$$|f_i(x) - f(x)| \;\leq\; |f_i(x) - f_j(x)| + |f_j(x) - f(x)| \;<\; \varepsilon + \varepsilon'$$

This is true for every positive ε', so $|f_i(x) - f(x)| \leq \varepsilon$ for every x in K. That is, the pointwise limit is approached uniformly in $x \in [a,b]$.

To prove that $f(x)$ is continuous, for $\varepsilon > 0$, take N be large enough so that $|f_i - f_j| < \varepsilon$ for all $i, j \geq N$. From the previous paragraph $|f_i(x) - f(x)| \leq \varepsilon$ for every x and for $i \geq N$. Fix $i \geq N$ and $x \in K$, and choose a small enough neigborhood U of x such that $|f_i(x) - f_i(y)| < \varepsilon$ for any y in U. Then

$$|f(x) - f(y)| \;\leq\; |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f(y) - f_i(y)| \;\leq\; \varepsilon + |f_i(x) - f_i(y)| + \varepsilon \;<\; \varepsilon + \varepsilon + \varepsilon$$

Thus, the pointwise limit f is continuous at every x in U. ///

Unsurprisingly, but significantly:

[15.1.2] Claim: For $x \in [a,b]$, the *evaluation* map $f \to f(x)$ is a continuous linear functional on $C^o[a,b]$.

Proof: For $|f-g|_{C^o} < \varepsilon$, we have

$$|f(x) - g(x)| \;\leq\; |f-g|_{C^o} \;<\; \varepsilon$$

proving the continuity. ///

As usual, a real-valued or complex-valued function f on a closed interval $[a, b] \subset \mathbb{R}$ is *continuously differentiable* when it has a derivative which is itself a continuous function. That is, the limit

$$f'(x) \;=\; \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

exists for all $x \in [a, b]$, and the function $f'(x)$ is in $C^o[a, b]$. Let $C^k[a, b]$ be the collection of k-times continuously differentiable functions on $[a, b]$, with the C^k-*norm*

$$|f|_{C^k} \;=\; \sum_{0 \le i \le k} \sup_{x \in [a,b]} |f^{(i)}(x)| \;=\; \sum_{0 \le i \le k} |f^{(i)}|_\infty$$

where $f^{(i)}$ is the i^{th} derivative of f. The *associated metric* on $C^k[a, b]$ is $|f - g|_{C^k}$.

Similar to the assertion about evaluation on $C^o[a, b]$,

[15.1.3] Claim: For $x \in [a, b]$ and $0 \le j \le k$, the *evaluation* map $f \to f^{(j)}(x)$ is a continuous linear functional on $C^k[a, b]$.

Proof: For $|f - g|_{C^k} < \varepsilon$,

$$|f^{(j)}(x) - g^{(j)}(x)| \;\le\; |f - g|_{C^k} \;<\; \varepsilon$$

proving the continuity. ///

We see that $C^k[a, b]$ is a Banach space:

[15.1.4] Theorem: The normed metric space $C^k[a, b]$ is complete.

Proof: For a Cauchy sequence $\{f_i\}$ in $C^k[a, b]$, all the pointwise limits $\lim_i f_i^{(j)}(x)$ of j-fold derivatives exist for $0 \le j \le k$, and are uniformly continuous. The issue is to show that $\lim_i f^{(j)}$ is differentiable, with derivative $\lim_i f^{(j+1)}$. It suffices to show that, for a Cauchy sequence f_n in $C^1[a, b]$, with pointwise limits $f(x) = \lim_n f_n(x)$ and $g(x) = \lim_n f_n'(x)$ we have $g = f'$. By the fundamental theorem of calculus, for any index i,

$$f_i(x) - f_i(a) \;=\; \int_a^x f_i'(t)\, dt$$

Since the f_i' *uniformly* approach g, given $\varepsilon > 0$ there is i_o such that $|f_i'(t) - g(t)| < \varepsilon$ for $i \ge i_o$ and for *all* t in the interval, so for such i

$$\Big| \int_a^x f_i'(t)\, dt - \int_a^x g(t)\, dt \Big| \;\le\; \int_a^x |f_i'(t) - g(t)|\, dt \;\le\; \varepsilon \cdot |x - a| \;\longrightarrow\; 0$$

Thus,

$$\lim_i f_i(x) - f_i(a) \;=\; \lim_i \int_a^x f_i'(t)\, dt \;=\; \int_a^x g(t)\, dt$$

from which $f' = g$. ///

By design, we have

[15.1.5] Theorem: The map $\frac{d}{dx} : C^k[a, b] \to C^{k-1}[a, b]$ is continuous.

Proof: As usual, for a linear map $T : V \to W$, by linearity $Tv - Tv' = T(v - v')$ it suffices to check continuity at 0. For Banach spaces the homogeneity $|\sigma \cdot v|_V = |\alpha| \cdot |v|_V$ shows that continuity is equivalent to existence of a constant B such that $|Tv|_W \le B \cdot |v|_V$ for $v \in V$. Then

$$|\frac{d}{dx} f|_{C^{k-1}} \;=\; \sum_{0 \le i \le k-1} \sup_{x \in [a,b]} |(\frac{df}{dx})^{(i)}(x)| \;=\; \sum_{1 \le i \le k} \sup_{x \in [a,b]} |f^{(i)}(x)| \;\le\; 1 \cdot |f|_{C^k}$$

as desired. ///

15.2 Non-Banach limit $C^\infty[a,b]$ of Banach spaces $C^k[a,b]$

The space $C^\infty[a,b]$ of infinitely differentiable complex-valued functions on a (finite) interval $[a,b]$ in $\mathbb{R}$ is not a Banach space. [22] Nevertheless, the topology is *completely determined* by its relation to the Banach spaces $C^k[a,b]$. That is, there is a *unique* reasonable topology on $C^\infty[a,b]$. After explaining and proving this uniqueness, we also show that this topology is *complete metric.*

This function space can be presented as

$$C^\infty[a,b] \;=\; \bigcap_{k\geq 0} C^k[a,b]$$

and we reasonably require that whatever topology $C^\infty[a,b]$ should have, each inclusion $C^\infty[a,b] \longrightarrow C^k[a,b]$ is continuous.

At the same time, given a family of *continuous linear* maps $Z \to C^k[a,b]$ from a vector space Z in some reasonable class, with the *compatibility* condition of giving commutative diagrams

$$\begin{tikzcd}
C^k[a,b] \arrow[r, "\subset"] & C^{k-1}[a,b] \\
 & Z \arrow[ul] \arrow[u]
\end{tikzcd}$$

the image of Z actually lies in the intersection $C^\infty[a,b]$. Thus, diagrammatically, for every family of compatible maps $Z \to C^k[a,b]$, there is a *unique* $Z \to C^\infty[a,b]$ fitting into a commutative diagram

$$\begin{tikzcd}
C^\infty[a,b] \arrow[rr, bend left] \arrow[rrr, bend left] & \ldots \arrow[r] & C^1[a,b] \arrow[r] & C^o[a,b] \\
 & Z \arrow[ul, dotted, "\exists!"] \arrow[ur, dashed] \arrow[urr, dashed, "\forall"'] & &
\end{tikzcd}$$

We require that this induced map $Z \to C^\infty[a,b]$ is *continuous.*

When we know that these conditions are met, we would say that $C^\infty[a,b]$ is the (projective) *limit* of the spaces $C^k[a,b]$, written

$$C^\infty[a,b] \;=\; \lim_k C^k[a,b]$$

with implicit reference to the inclusions $C^{k+1}[a,b] \to C^k[a,b]$ and $C^\infty[a,b] \to C^k[a,b]$.

[15.2.1] Claim: Up to unique isomorphism, there exists at most one topology on $C^\infty[a,b]$ such that to every compatible family of continuous linear maps $Z \to C^k[a,b]$ from a topological vector space Z there is a unique continuous linear $Z \to C^\infty[a,b]$ fitting into a commutative diagram as just above.

Proof: Let X, Y be $C^\infty[a,b]$ with two topologies fitting into such diagrams, and show $X \approx Y$, and for a unique isomorphism. First, claim that the identity map $\mathrm{id}_X : X \to X$ is the only map $\varphi : X \to X$ fitting into a commutative diagram

[22] It is not essential to prove that there is no reasonable Banach space structure on $C^\infty[a,b]$, but this can be readily proven in a suitable context.

$$\begin{tikzcd}
X \arrow[rr, bend left] \arrow[rrr, bend left] & \cdots \arrow[r] & C^1[a,b] \arrow[r] \arrow[d, equal] & C^o[a,b] \arrow[d, equal] \\
X \arrow[u, "\varphi"] \arrow[rr, bend right] \arrow[rrr, bend right] & \cdots \arrow[r] & C^1[a,b] \arrow[r] & C^o[a,b]
\end{tikzcd}$$

Indeed, given a compatible family of maps $X \to C^k[a,b]$, there is *unique* φ fitting into

$$\begin{tikzcd}
X \arrow[rr, bend left] \arrow[rrr, bend left] & \cdots \arrow[r] & C^1[a,b] \arrow[r] & C^o[a,b] \\
& X \arrow[ul, dotted, "\varphi"] \arrow[ur, dashed] \arrow[urr, dashed, "\forall"'] & &
\end{tikzcd}$$

Since the identity map id_X fits, necessarily $\varphi = \mathrm{id}_X$. Similarly, given the compatible family of inclusions $Y \to C^k[a,b]$, there is unique $f : Y \to X$ fitting into

$$\begin{tikzcd}
X \arrow[rr, bend left] \arrow[rrr, bend left] & \cdots \arrow[r] & C^1[a,b] \arrow[r] & C^o[a,b] \\
& Y \arrow[ul, dotted, "f"] \arrow[ur] \arrow[urr] & &
\end{tikzcd}$$

Similarly, given the compatible family of inclusions $X \to C^k[a,b]$, there is unique $g : X \to Y$ fitting into

$$\begin{tikzcd}
Y \arrow[rr, bend left] \arrow[rrr, bend left] & \cdots \arrow[r] & C^1[a,b] \arrow[r] & C^o[a,b] \\
& X \arrow[ul, dotted, "g"] \arrow[ur] \arrow[urr] & &
\end{tikzcd}$$

Then $f \circ g : X \to X$ fits into a diagram

$$\begin{tikzcd}
X \arrow[rr, bend left] \arrow[rrr, bend left] & \cdots \arrow[r] & C^1[a,b] \arrow[r] & C^o[a,b] \\
& X \arrow[ul, dotted, "f\circ g"] \arrow[ur] \arrow[urr] & &
\end{tikzcd}$$

Therefore, $f \circ g = \mathrm{id}_X$. Similarly, $g \circ f = \mathrm{id}_Y$. That is, f, g are mutual inverses, so are isomorphisms of topological vector spaces. ///

Existence of a topology on $C^\infty[a,b]$ satisfying the condition above will be proven by identifying $C^\infty[a,b]$ as the obvious diagonal *closed subspace* of the *topological product* of the *limitands* $C^k[a,b]$:

$$C^\infty[a,b] \;=\; \{\{f_k : f_k \in C^k[a,b]\} : f_k = f_{k+1} \text{ for all } k\}$$

An arbitrary *product* of topological spaces X_α for α in an index set A is a topological space X with (*projections*) $p_\alpha : X \to X_\alpha$, such that every family $f_\alpha : Z \to X_\alpha$ of maps from any other topological space Z *factors through* the p_α *uniquely*, in the sense that there is a unique $f : Z \to X$ such that $f_\alpha = p_\alpha \circ f$ for all α. Pictorially, *all triangles commute* in the diagram

$$\begin{tikzcd}
Z \arrow[rrr, dotted, "f"] \arrow[drr, dashed, "f_\alpha"'] \arrow[drrrr, dashed, "f_\beta"] & & & X \arrow[dl, "p_\alpha"'] \arrow[dr, "p_\beta"] & & \\
& \cdots & X_\alpha & \cdots & X_\beta & \cdots
\end{tikzcd}$$

A similar argument to that for uniqueness of limits proves *uniqueness* of products up to unique isomorphism. *Construction* of products is by putting the usual product topology with basis consisting of products $\prod_\alpha Y_\alpha$ with $Y_\alpha = X_\alpha$ for all but finitely-many indices, on the Cartesian product of the *sets* X_α, whose existence we grant ourselves. Proof that this usual is *a* product amounts to unwinding the definitions. By uniqueness, in particular, despite the plausibility of the *box topology* on the product, it cannot function as a product topology since it differs from the standard product topology in general.

[15.2.2] Claim: Giving the diagonal copy of $C^\infty[a,b]$ inside $\prod_k C^k[a,b]$ the subspace topology yields a (projective) limit topology.

Proof: The projection maps $p_k : \prod_j C^j[a,b] \to C^k[a,b]$ from the whole product to the factors $C^k[a,b]$ are continuous, so their restrictions to the diagonally imbedded $C^\infty[a,b]$ are continuous. Further, letting $i_k : C^k[a,b] \to C^{k-1}[a,b]$ be the inclusion, on that diagonal copy of $C^\infty[a,b]$ we have $i_k \circ p_k = p_{k-1}$ as required.

On the other hand, *any* family of maps $\varphi_k : Z \to C^k[a,b]$ induces a map $\widetilde{\varphi} : Z \to \prod C^k[a,b]$ such that $p_k \circ \widetilde{\varphi} = \varphi_k$, by the property of the product. *Compatibility* $i_k \circ \varphi_k = \varphi_{k-1}$ implies that the image of $\widetilde{\varphi}$ is inside the diagonal, that is, inside the copy of $C^\infty[a,b]$. ///

A *countable* product of *metric* spaces X_k with metrics d_k has no canonical single metric, but is *metrizable*. One of many topologically equivalent metrics is the usual

$$d(\{x_k\},\{y_k\}) \;=\; \sum_{k=0}^{\infty} 2^{-k} \frac{d_k(x_k - y_k)}{d_k(x_k - y_k) + 1}$$

When the metric spaces X_k are *complete*, the product is complete. A closed subspace of a complete metrizable space is complete metrizable, so we have

[15.2.3] Corollary: $C^\infty[a,b]$ is complete metrizable. ///

Abstracting the above, for a (not necessarily countable) family

$$\ldots \xrightarrow{\varphi_2} B_1 \xrightarrow{\varphi_1} B_o$$

of Banach spaces with continuous linear transition maps as indicated, *not* recessarily requiring the continuous linear maps to be injective (or surjective), a *(projective) limit* $\lim_i B_i$ is a topological vector space with continuous linear maps $\lim_i B_i \to B_j$ such that, for every compatible family of continuous linear maps $Z \to B_i$ there is unique continuous linear $Z \to \lim_i B_i$ fitting into

$$\begin{tikzcd} \lim_i B_i \arrow[rr, bend left] \arrow[rrr, bend left] & \ldots \arrow[r, "\varphi_2"] & B_1 \arrow[r, "\varphi_1"] & B_o \\ & Z \arrow[ul, dotted] \arrow[ur, dashed] \arrow[urr, dashed] & & \end{tikzcd}$$

The same *uniqueness* proof as above shows that there is at most one topological vector space $\lim_i B_i$. For *existence* by *construction*, the earlier argument needs only minor adjustment. The conclusion of complete metrizability would hold when the family is countable.

Before declaring $C^\infty[a,b]$ to be a *Fréchet* space, we must certify that it is *locally convex*, in the sense that every point has a local basis of *convex* opens. Normed spaces are immediately locally convex, because open balls are convex: for $0 \le t \le 1$ and x, y in the ε-ball at 0 in a normed space,

$$|tx + (1-t)y| \;\le\; |tx| + |(1-t)y| \;\le\; t|x| + (1-t)|y| \;<\; t\cdot\varepsilon + (1-t)\cdot\varepsilon \;=\; \varepsilon$$

Product topologies of locally convex vectorspaces are locally convex, from the *construction* of the product. The construction of the limit as the diagonal in the product, with the subspace topology, shows that it is locally convex. In particular, *countable limits of Banach spaces are locally convex, hence, are Fréchet.* All spaces of practical interest are locally convex for simple reasons, so demonstrating local convexity is rarely interesting.

[15.2.4] Theorem: $\frac{d}{dx} : C^\infty[a,b] \to C^\infty[a,b]$ is continuous.

Proof: In fact, the differentiation operator is characterized via the expression of $C^\infty[a,b]$ as a limit. We already know that differentiation d/dx gives a continuous map $C^k[a,b] \to C^{k-1}[a,b]$. Differentiation is compatible with the inclusions among the $C^k[a,b]$. Thus, we have a commutative diagram

$$\begin{tikzcd}
C^\infty[a,b] \arrow[rr, bend left] \arrow[rrr, bend left] & \ldots & C^k[a,b] \arrow[r] & C^{k-1}[a,b] \arrow[r] & \ldots \\
C^\infty[a,b] \arrow[rr, bend right] \arrow[rrr, bend right] & \ldots & C^k[a,b] \arrow[r] \arrow[ur, "\frac{d}{dx}"] & C^{k-1}[a,b] \arrow[r] \arrow[ur, "\frac{d}{dx}"] & \ldots
\end{tikzcd}$$

Composing the projections with d/dx gives (dashed) induced maps from $C^\infty[a,b]$ to the limitands, inducing a unique (dotted) continuous linear map to the limit, as in

$$\begin{tikzcd}
C^\infty[a,b] \arrow[rr, bend left] \arrow[rrr, bend left] & \ldots & C^k[a,b] \arrow[r] & C^{k-1}[a,b] \arrow[r] & \ldots \\
C^\infty[a,b] \arrow[u, dotted, "\frac{d}{dx}"] \arrow[rr, bend right] \arrow[rrr, bend right] \arrow[urr, dashed] \arrow[urrr, dashed] \arrow[urrrr, dashed] & \ldots & C^k[a,b] \arrow[r] \arrow[ur] & C^{k-1}[a,b] \arrow[r] \arrow[ur] & \ldots
\end{tikzcd}$$

This proves the continuity of differentiation in the limit topology. ///

In a slightly different vein, we have

[15.2.5] Claim: For all $x \in [a,b]$ and for all non-negative integers k, the evaluation map $f \to f^{(k)}(x)$ is a continuous linear map $C^\infty[a,b] \to \mathbb{C}$.

Proof: The inclusion $C^\infty[a,b] \to C^k[a,b]$ is continuous, and the evaluation of the k^{th} derivative is continuous. ///

16. $C^\infty(\mathbb{T})$ is not normable

1. Countable limits of Banach spaces
2. Maps from limits of Banach spaces to normed spaces factor through limitands
3. $C^\infty(\mathbb{T})$ is not normable

Many natural function spaces, such as $C^\infty[a,b]$ and $C^\infty(\mathbb{T})$, are *not* Banach, nor even *norm-able* but still do have a metric topology and are complete: these are *Fréchet spaces*, appearing as countable (projective) *limits* of Banach spaces. It is reasonable to ask *why* these spaces are not Banach, and in fact not even *normable*, that is, their topologies cannot be given by a any norm, regardless of metric completeness.

In brief, in tangible terms, the root cause of this impossibility is that no estimates on the first k derivatives of a function on $\mathbb{T}$ give an estimate on the $(k+1)^{th}$ derivative, for any k. This is discussed precisely below, and abstracted somewhat.

16.1 Countable limits of Banach spaces

We could take *countable limit of Banach spaces* as the definition of *Fréchet space*.

As earlier, $C^\infty(\mathbb{T})$ is a countable nested intersection, which is a countable (projective) *limit*:

$$C^\infty(\mathbb{T}) \;=\; \bigcap_{k\geq 0} C^k(\mathbb{T}) \;=\; \lim_k C^k(\mathbb{T})$$

From very general category-theory arguments, *there is at most one projective-limit topology* on $C^\infty(\mathbb{T})$, up to unique isomorphism. Existence of the topology on X satisfying the limit condition can be proven by identifying X as the diagonal *closed subspace* of the *topological product* of the *limitands* X_k: letting $p_{k,k-1} : X_k \to X_{k-1}$ be the transition maps,

$$X \;=\; \{\{x_k : x_k \in C^k[a,b]\} : p_{k,k-1}(x_k) = x_{k-1} \text{ for all } k\}$$

The subspace topology on X is the limit topology, seen as follows. The projection maps $p_k : \prod_j X_j \to X_k$ from the whole product to the factors X_k are continuous, so their restrictions to the diagonally imbedded X are continuous. Further, letting $i_k : X_k \to X_{k-1}$ be the transition map, on that diagonal copy of X we have $i_k \circ p_k = p_{k-1}$ as required.

On the other hand, *any* family of maps $\varphi_k : Z \to X_k$ induces a map $\widetilde{\varphi} : Z \to \prod X_k$ such that $p_k \circ \widetilde{\varphi} = \varphi_k$, by the property of the product. *Compatibility* $i_k \circ \varphi_k = \varphi_{k-1}$ implies that the image of $\widetilde{\varphi}$ is inside the diagonal, that is, inside the copy of X. Thus, this construction does produce a limit.

A *countable* product of *metric* spaces X_k with metrics d_k has no canonical single metric, but is *metrizable*. One of many topologically equivalent metrics is the usual

$$d(\{x_k\},\{y_k\}) \;=\; \sum_{k=0}^{\infty} 2^{-k} \frac{d_k(x_k - y_k)}{d_k(x_k - y_k) + 1}$$

When the metric spaces X_k are *complete*, the product is complete. A closed subspace of a complete metrizable space is complete metrizable, so the diagonal X is complete metric.

Even in general, the topologies on vector spaces V are required to be *translation invariant*, meaning that for an open neighborhood U of 0, for any $x \in V$, the set $x + U = \{x + u : u \in U\}$ is an open neighborhood of x, and vice-versa. [23] Thus, to specify the topology on a limit X of Banach spaces X_k, we need only give a

[23] For Hilbert and Banach spaces, this translation-invariance is clear, since the topology is metric, and comes from a norm.

local basis at 0. From the construction above, a local basis is given by all sets

$$U_{k,\delta} \;=\; \{x \in X : |p_k(x)|_{X_k} < \delta\} \qquad\qquad \text{(for } \delta > 0 \text{ and index } k\text{)}$$

16.2 Maps from limits of Banach spaces to normed spaces factor through limitands

[16.2.1] Lemma: Given a continuous linear map T from $C^\infty(\mathbb{T})$ to a *normed space* Y, there is an index k such that when $C^\infty(\mathbb{T})$ is given the (weaker) C^k topology, $T : C^\infty(\mathbb{T}) \to Y$ is still continuous.

[16.2.2] Corollary: Every continuous linear map T from $C^\infty(\mathbb{T})$ to a Banach space Y factors through some limitand $C^k(\mathbb{T})$. That is, there is $T_k : C^k(\mathbb{T}) \to Y$ such that $T = T_k \circ i_k$, where $i_k : C^\infty(\mathbb{T}) \to C^k(\mathbb{T})$ is the inclusion.

Proof: (of Corollary) After applying the lemma, since the target space of T is *complete*, we can extend $T : C^\infty(\mathbb{T}) \to Y$ by continuity (in the C^k topology) to the C^k-completion of C^∞, which is C^k. ///

The lemma is a special case of the analogous lemma that has nothing to do with spaces of functions, but, rather, is true for more general reasons:

[16.2.3] Lemma: Let $X = \lim_k X_k$ be a limit of Banach spaces X_k, with projection maps $p_k : X \to X_k$. Suppose that $p_k(X)$ is *dense* in X_k. Then every continuous linear map $T : X \to Y$ to a *normed space* Y factors through some limitand X_k. That is, there is $T_k : X_k \to Y$ such that $T = T_k \circ p_k$.

Proof: Given $\varepsilon > 0$, by the description above of the topology on the limit, there are $\delta > 0$ and index k such that $T(U_{k,\delta})$ is inside the ε-ball at 0 in Y.

Then, given any other $\varepsilon' > 0$, we claim that T maps

$$\frac{\varepsilon'}{\varepsilon} \cdot U_{k,\delta} \;=\; U_{k,\delta\varepsilon'/\varepsilon}$$

to the open ε'-ball in Y. Indeed,

$$|T\big(\frac{\varepsilon'}{\varepsilon} \cdot U_{k,\delta}\big)|_Y \;=\; \frac{\varepsilon'}{\varepsilon} \cdot |T(U_{k,\delta})|_Y \;<\; \frac{\varepsilon'}{\varepsilon} \cdot \varepsilon \;=\; \varepsilon'$$

as claimed. Thus, $T : X \to Y$ is continuous when X is given the X_k topology, for the index k that makes this work. Thus, T extends by continuity to the $|\cdot|_{X_k}$-completion of X. By the density assumption, this is X_i. ///

[16.2.4] Remark: Finite Fourier series, which are in $C^\infty(\mathbb{T})$, are dense in every $C^k(\mathbb{T})$, so $C^\infty(\mathbb{T})$ is dense in every $C^k(\mathbb{T})$.

[16.2.5] Remark: In the case that $Y = \mathbb{C}$, the density assumption is unnecessary, since Hahn-Banach gives an extension. But for general Banach Y, without the density assumption, we can only conclude that T factors through the $|\cdot|_{X_k}$-completion of X, since not all closed subspaces of Banach spaces are *complemented*.

16.3 $C^\infty(\mathbb{T})$ is not normable

If $C^\infty(\mathbb{T})$ were *normable*, then the identity map $j : C^\infty(\mathbb{T}) \to C^\infty(\mathbb{T})$ would be continuous when the source is given the C^k topology. In particular, for every $\varepsilon > 0$, there would be a sufficiently small $|\cdot|_{X_k}$-ball B whose image in $C^\infty(\mathbb{T})$ under the inclusion is inside the ε-ball in the $C^{k+1}(\mathbb{T})$ topology on $C^\infty(\mathbb{T})$. Specifically, for

$\varepsilon = 1$, there should be a sufficiently small $\delta > 0$ such that the δ-ball in the C^k topology is inside the unit ball in the C^{k+1} topology.

However, it is easy-enough to construct C^∞ functions whose C^k norms are arbitrarily small, but whose C^{k+1} norm is 1, for example, e^{iNx}/N^{k+1}. Thus, we achieve a contradiction. ///

17. Introduction to Levi-Sobolev spaces

The simplest case of a Levi-Sobolev *imbedding theorem* asserts that the +1-index Levi-Sobolev Hilbert space $H^1[a,b]$ described below is inside $C^o[a,b]$. This is a corollary of a Levi-Sobolev *inequality* asserting that the $C^o[a,b]$ norm is *dominated* by the $H^1[a,b]$ norm. All that is used is the fundamental theorem of calculus and the Cauchy-Schwarz-Bunyakowsky inequality. The point is that there is a large *Hilbert space* $H^1[a,b]$ inside the *Banach* space $C^o[a,b]$.

We will do much more with this idea subsequently.

We can think of $L^2[a,b]$ as

$$L^2[a,b] \;=\; \text{completion of }\; C^o[a,b] \text{ with respect to } |f|_{L^2} = \Big(\int_a^b |f(t)|^2\,dt\Big)^{1/2}$$

In fact, by this point we have shown that every $C^k[a,b]$ is dense in $L^2[a,b]$.

The +1-index *Levi-Sobolev* space [24] $H^1[a,b]$ is

$$H^1[a,b] \;=\; \text{completion of }\; C^1[a,b] \text{ with respect to } |f|_{H^1} = \Big(|f|^2_{L^2[a,b]} + |f'|^2_{L^2[a,b]}\Big)^{1/2}$$

[17.0.1] Theorem:

(Levi-Sobolev inequality) On $C^1[a,b]$, the $H^1[a,b]$-norm *dominates* the $C^o[a,b]$-norm. That is, there is a constant C depending only on a,b such that $|f|_{C^o[a,b]} \le C\cdot|f|_{H^1[a,b]}$ for every $f\in C^1[a,b]$.

Proof: For $a\le x\le y\le b$, for $f\in C^1[a,b]$, the fundamental theorem of calculus gives

$$|f(y)-f(x)| \;=\; \Big|\int_x^y f'(t)\,dt\Big| \;\le\; \int_x^y |f'(t)|\,dt \;\le\; \Big(\int_x^y |f'(t)|^2\,dt\Big)^{1/2}\cdot\Big(\int_x^y 1\,dt\Big)^{1/2}$$
$$\le\; |f'|_{L^2}\cdot|x-y|^{\frac12} \;\le\; |f'|_{L^2}\cdot|a-b|^{\frac12}$$

Using the continuity of $f\in C^1[a,b]$, let $y\in[a,b]$ be such that $|f(y)| = \min_x|f(x)|$. Using the previous inequality,

$$|f(x)| \;\le\; |f(y)|+|f(x)-f(y)| \;\le\; \frac{\int_a^b |f(t)|\,dt}{|a-b|} \;+\; |f(x)-f(y)| \;\le\; \frac{\int_a^b |f|\cdot 1}{|a-b|} \;+\; |f'|_{L^2}\cdot|a-b|^{\frac12}$$

$$\le\; \frac{|f|^{\frac12}_{L^2}\cdot|a-b|^{\frac12}}{|a-b|} + |f'|_{L^2}\cdot|a-b|^{\frac12} \;=\; \frac{|f|^{\frac12}_{L^2}}{|a-b|^{\frac12}} + |f'|_{L^2}\cdot|a-b|^{\frac12} \;\le\; \Big(|f|_{L^2}+|f'|_{L^2}\Big)\cdot\Big(|a-b|^{-\frac12}+|a-b|^{\frac12}\Big)$$

$$\le\; 2(|f|^2+|f'|^2)^{1/2}\cdot\Big(|a-b|^{-\frac12}+|a-b|^{\frac12}\Big) \;=\; |f|_{H^1}\cdot 2\Big(|a-b|^{-\frac12}+|a-b|^{\frac12}\Big)$$

Thus, on $C^1[a,b]$ the H^1 norm dominates the C^o-norm. ///

[17.0.2] Corollary:

(Levi-Sobolev imbedding) $H^1[a,b]\subset C^o[a,b]$.

Proof: Since $H^1[a,b]$ is the H^1-norm completion of $C^1[a,b]$, every $f\in H^1[a,b]$ is an H^1-limit of functions $f_n\in C^1[a,b]$. That is, $|f-f_n|_{H^1[a,b]}\to 0$. Since the H^1-norm dominates the C^o-norm, $|f-f_n|_{C^o[a,b]}\to 0$. A C^o limit of continuous functions is continuous, so f is continuous. ///

[24] ... also denoted $W^{1,2}[a,b]$, where the superscript 2 refers to L^2, rather than L^p. Beppo Levi noted the importance of taking Hilbert space completion with respect to this norm in 1906, giving a correct formulation of *Dirichlet's principle*. Sobolev's systematic development of these ideas was in the mid-1930's.

In fact, we have a stronger conclusion than continuity, namely, a *Lipschitz condition* with exponent $\frac{1}{2}$:

[17.0.3] Corollary: (of proof of theorem) $|f(x) - f(y)| \le |f'|_{L^2} \cdot |x-y|^{\frac{1}{2}}$ for $f \in H^1[a,b]$. ///

18. Generalized functions (distributions) on circles

The simplest physical object with an interesting function theory is the circle, $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, which inherits *group structure* and *translation-invariant* differential operator d/dx from the real line $\mathbb{R}$. Equivalently, we can consider *periodic* functions on $\mathbb{R}$. This is *not* quite the same as considering functions f on the interval $[0, 2\pi]$, unless we also explicitly require matching of function values and derivatives' values (when they exist) at the endpoints: $f(0) = f(2\pi)$, $f'(0) = f'(2\pi)$, and so on, to the extent applicable.

The exponential functions $x \to e^{inx}$ for $n \in \mathbb{Z}$ are all the continuous group homomorphisms $\mathbb{R}/2\pi\mathbb{Z} \to \mathbb{C}^\times$ *and* are all the eigenfunctions for d/dx on $\mathbb{R}/2\pi\mathbb{Z}$. Finite or infinite linear combinations

$$\sum_{n\in\mathbb{Z}} c_n\, e^{inx}$$

are *Fourier series.* [25] Conveniently, a function so expressed is a linear combination, although an *infinite* linear combination, of eigenvectors for d/dx. That is, on functions *with* Fourier expansions [26] the linear operator of differentiation is *diagonalized.* However, infinite-dimensional linear algebra is subtler than finite-dimensional. Some fundamental questions are [27]

> *In what sense(s) can a function be expressed as a Fourier series?*
>
> *Can a Fourier series be differentiated term-by-term?*
>
> *How cautious must we be in differentiating functions that are only piecewise differentiable?*
>
> *What will derivatives of discontinuous functions be?*

Several further issues are *implicit*, and the *best* answers need viewpoints created first in 1906 by Beppo Levi, 1907 by G. Frobenius, in the 1930's by Sobolev, and Schwartz post-1949, enabling legitimate discussion of

---

[25] In the early 19^{th} century, J. Fourier was an impassioned advocate of the use of such sums, of course writing sines and cosines rather than complex exponentials. Euler, the Bernouillis, and others had used such sums in similar fashions and for similar ends, but Fourier made a claim extravagant for the time, namely that *all functions* could be expressed in such terms. Unfortunately, in those days there was no clear idea of what a *function* was, no vocabulary to specificy *classes* of functions, and no specification of what it would mean to *represent* a function by such a series.

[26] The notion of *has a Fourier expansion* would need to clarify what *has such an expansion* means. Must it mean that pointwise values can be retrieved from the Fourier series? Less? More?

[27] At about the time Fourier was promoting Fourier series, Abel proved that convergent power series *can* be differentiated term-by-term in the interior of their interval (on $\mathbb{R}$) or disk (in $\mathbb{C}$) of convergence, and *are* infinitely-differentiable functions. Abel's result fit the optimistic expectations of the time, but created unreasonable expectations for the behavior of Fourier series.

generalized functions (a.k.a., *distributions*).[28] There are natural technical questions, like

Why define generalized functions as dual spaces?

In brief, Schwartz' 1940's insight to define generalized functions as *dual spaces* is a natural consequence of one natural *relaxation* of the notion of *function*. Rather than demand that functions produce *pointwise values*, which precipitated endless classical discussion of what to do with jump discontinuities, instead declare that *functions* in the broadest sense are merely things that can be *integrated against*. For given φ, the map that integrates against φ,

$$f \longrightarrow \int f(x)\,\varphi(x)\,dx$$

is a *functional* (a $\mathbb{C}$-valued linear map), and is, or ought to be, probably *continuous* in a reasonable topology. To consider the collection of *all* continuous linear functionals is a reasonable way to enlarge the collection of functions, as *things to be integrated against*.

From the other side, it might have been that this generalization of *function* is needlessly extravagant, but it turns out that every distribution on the circle $\mathbb{T}$ is a high-order derivative of a continuous function. Thus, since we *do* want to be able to take derivatives indefinitely, there is no waste.

Further, in any of the several natural topologies on distributions, very nice ordinary functions are *dense*, and the space of distributions is *complete* in a sense subsuming that for metric spaces. Thus, taking limits *yields* all distributions, *and* produces no excess.

This discussion is easiest on the circle $\mathbb{T}$, or products $\mathbb{T}^n$ of circles, making use of Fourier series, and clarifying many technical questions about Fourier series.[29] This story is a prototype for more complicated examples.

There is an important auxiliary technical point. Natural spaces of functions *do not* have structures of Hilbert spaces, but typically, of Banach spaces. Nevertheless, the simplicity of Hilbert spaces motivates comparisons of natural function spaces with related Hilbert spaces. Such comparisons are *Levi-Sobolev imbeddings* or *Levi-Sobolev inequalities*.

The present discussion presumes acquaintance with the basics of Fourier series, namely, the Fourier-Dirichlet kernel, the theorem of Fourier-Dirichlet on pointwise convergence for finitely-piecewise-continuous at points with left derivative and right derivative, Féjer's kernel, Féjer's theorem on the density of finite Fourier series in $C^o(\mathbb{T})$, and completeness of exponentials in $L^2(\mathbb{T})$.

We also presume that the notion of *(projective) limit* of Banach spaces is appreciated to some degree, at least in examples such as the *nested intersection*

$$C^\infty(\mathbb{T}) \;=\; \bigcap_k C^k(\mathbb{T}) \;=\; \lim_k C^k(\mathbb{T})$$

We recall this, and introduce *colimits*, especially in the case of *ascending unions* of spaces of *duals* of limits.

---

[28] K. Friedrichs' important 1934-5 discussions of semi-bounded unbounded operators on Hilbert spaces used norms defined in terms of derivatives, but only internally in proofs, while for Levi, Frobenius, and Sobolev these norms were significant objects themselves.

[29] The classic reference is A. Zygmund, *Trigonometric Series, I, II*, first published in Warsaw in 1935, reprinted several times, including a 1959 Cambridge University Press edition. The present discussion neglects many interesting details, *but* is readily adaptible to more complicated situations, so necessarily our treatment is different from Zygmund's.

18.1 Provocative example

Let $s(x)$ be the *sawtooth function*[30]

$$s(x) \;=\; x - \pi \qquad (\text{for } 0 \le x < 2\pi)$$

and made *periodic* by demanding $s(x + 2\pi n) = s(x)$ for all $n \in \mathbb{Z}$. In other words, letting $[\![x/2\pi]\!]$ be the greatest integer less than or equal $x/2\pi$,

$$s(x) \quad = \quad x \;-\; 2\pi \cdot \left[\!\!\left[\frac{x}{2\pi}\right]\!\!\right] \;-\; \pi \qquad\qquad (\text{for } x \in \mathbb{R})$$

Away from $2\pi\mathbb{Z}$, the sawtooth function is infinitely differentiable, with derivative 1. At $x \in 2\pi\mathbb{Z}$ the sawtooth jumps down from value to π to value $-\pi$. There is no reason to worry about defining a value at $x \in 2\pi\mathbb{Z}$.

The exponential functions $\psi_n(x) = e^{inx}$ are not quite an *orthonormal* basis for the Hilbert space $L^2[0, 2\pi]$, but are *orthogonal*:

$$\int_0^{2\pi} \psi_m(x) \cdot \overline{\psi}_n(x)\, dx \;\;=\;\; \begin{cases} 0 & (\text{for } m \neq n) \\ \\ 2\pi & (\text{for } m = n) \end{cases}$$

Anticipating that Fourier coefficients $\widehat{f}(n)$ of $2\pi\mathbb{Z}$-periodic functions f are computed [31] by integrating against $\psi_n(x) = e^{inx}$ (conjugated):

$$\widehat{f}(n) \;\;=\;\; \frac{1}{2\pi} \int_0^{2\pi} f(x)\, e^{-inx}\, dx$$

integration by parts gives

$$\widehat{s}(n) \;\;=\;\; \frac{1}{2\pi} \int_0^{2\pi} s(x) \cdot e^{-inx}\, dx \;=\; \begin{cases} \dfrac{1}{-in} & (\text{for } n \neq 0) \\ \\ 0 & (\text{for } n = 0) \end{cases}$$

Thus, in whatever sense a function *is* its Fourier expansion, we anticipate that

$$s(x) \;\;\sim\;\; \sum_{n \in \mathbb{Z}} \widehat{s}(n) \cdot e^{inx} \;\;=\;\; \sum_{n \neq 0} \frac{1}{-in} \cdot e^{inx}$$

Even though this series does not converge absolutely for *any* value of x, we already know (by Fourier-Dirichlet) that it *does* converge to the value of $s(x)$ for $x \notin 2\pi\mathbb{Z}$. Since $s(x)$ has discontinuities at $2\pi\mathbb{Z}$ *anyway*, this is hardly surprising. Nothing disturbing has happened.

Now differentiate. The sawtooth function *is* differentiable away from $2\pi\mathbb{Z}$, with value 1, and with uncertain value at $2\pi\mathbb{Z}$. With exogenous reasons to differentiate the Fourier series term-by-term, with or without confidence in doing so, and the blatant differentiability of $s(x)$ away from $2\pi\mathbb{Z}$ suggests it's not entirely ridiculous to differentiate term-by-term. Then

$$s'(x) \;=\; \begin{cases} 1 & (\text{for } x \notin 2\pi\mathbb{Z}) \\ \\ ? & (\text{for } x \in 2\pi\mathbb{Z}) \end{cases} \qquad \sim \quad -\sum_{n \neq 0} e^{inx}$$

[30] One may also take $s(x) = x$ for $-\pi < x < \pi$ and extend by periodicity. This definition avoids the subtraction of π, and has the same operational features. In the end, it doesn't matter.

[31] Apparently at first Fourier did not have this expression for the Fourier coefficients!

The right-hand side is hard to interpret, certainly as having pointwise values. On the other hand, reasonably interpreted, it is still ok to integrate against this sum: letting $\widehat{f}(n)$ be the n^{th} Fourier coefficient of a *smooth* function f, and not worrying about justifications,

$$\int_0^{2\pi} f(x) \Big(-\sum_{n\neq 0} e^{inx} \Big)\, dx \;=\; -\sum_{n\neq 0} \int_0^{2\pi} f(x)\, e^{inx}\; dx \;=\; -2\pi \sum_{n\neq 0} \widehat{f}(-n)$$

$$=\; 2\pi\widehat{f}(0) - 2\pi \sum_{n\in\mathbb{Z}} \widehat{f}(n)\, e^{in\cdot 0} \;=\; 2\pi\widehat{f}(0) - 2\pi f(0) \;=\; \int_0^{2\pi} f(x)\, dx \;-\; 2\pi \cdot f(0)$$

The map

$$f \;\longrightarrow\; \int_0^{2\pi} f(x)\, dx \;-\; 2\pi \cdot f(0)$$

has a sense for continuous f, and gives a *functional*. That the derivative of the sawtooth is *mostly* 1 gives the integral of f (against 1) over $[0, 2\pi]$. Further, the $-2\pi f(0)$ term forcefully suggests that *the derivative of the discontinuity of the sawtooth function is the (periodic) evaluation-at-*0 *functional* $f \to f(0)$ *multiplied by* -2π). [32]

[18.1.1] Remark: A truly disastrous choice at this point would be to think that since $s'(x)$ is *almost everywhere* 1 (in a measure-theoretic sense) that its singularities are somehow *removable*, and thus pretend that $s'(x) = 1$. This would give $s''(x) = 0$, and make the following worse than it is, and impossible to explain.

Still, $s'(x)$ *is* differentiable away from $2\pi\mathbb{Z}$, and by repeated differentiation

$$s^{(k+1)}(x) \;=\; \begin{cases} 0 & (\text{for } x \notin 2\pi\mathbb{Z}) \\ \\ ? & (\text{for } x \in 2\pi\mathbb{Z}) \end{cases} \quad\sim\quad -(i)^k \sum_{n\neq 0} n^k \cdot e^{inx}$$

By now the right-hand sides are vividly not convergent. The summands do not go to zero, in fact, are *unbounded*.

One can continue differentiating in this symbolic sense, but the meaning is unclear.

One reaction is to simply object to differentiating a non-differentiable function, even if its discontinuities are mild. This is not productive.

Another unproductive viewpoint is to deny that Fourier series reliably represent the functions that produced their coefficients.

A happier and more useful response is to suspect that the above computation is *correct*, though the question mark needs explanation, *and* that the right-hand side is correct and meaningful, *despite* its divergence in classical senses. The question is *what* meaning to attach. This requires preparation.

We will establish a context in which the derivatives of the sawtooth, and derivatives of other discontinuous functions, are *things to integrate against*, rather than *things to evaluate pointwise*, and see that termwise differentiation of Fourier series *does* capture an extended notion of function and derivative.

[32] The *jump* is downward rather than upward.

18.2 Natural function spaces on the circle $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$

We review natural families of functions. In all cases, the object is to give the vector space of functions a metric (if possible) which makes it *complete*, to allow *taking limits* inside the same class of functions. For example, *pointwise* limits of continuous functions easily fail to be continuous, but *uniform* pointwise limits of continuous functions *are* continuous. [33]

[18.2.1] Continuous functions and sup-norm

First, we care about *continuous* complex-valued functions. Although we have in mind continuous functions on the circle $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, the basic result depends only upon the *compactness* of $\mathbb{R}/2\pi\mathbb{Z}$.

As usual, we give the set $C^o(K)$ of (complex-valued) continuous functions on a compact topological space K the metric

$$d(f,g) \;=\; \sup_{x\in K} |f(x) - g(x)|$$

The sup is *finite* because K is compact and $f - g$ is continuous. The right-hand side of this last equation arises from the *(sup) norm*

$$|f|_\infty \;=\; |f|_{C^o} = \sup_{x\in K} |f(x)|$$

and $d(f,g) = |f-g|_{C^o}$. A main feature of continuous functions is that they have *pointwise values*. Recall the unsurprising but important

[18.2.2] Claim: With the $C^o(K)$ topology, for $x \in K$ the *evaluation functional*[34] $C^o(K) \to \mathbb{C}$ by $f \to f(x)$ is *continuous.*

Proof: The inequality

$$|f(x) - g(x)| \;\le\; \sup_{y\in K} |f(y) - g(y)| \qquad\qquad (\text{for } f,g \in C^o(K))$$

proves the continuity of evaluation. ///

Also, recall, yet again, the iconic

[18.2.3] Theorem: The space $C^o(K)$ of (complex-valued) continuous functions on a compact topological space K is *complete.*

[18.2.4] Remark: Thus, being complete with respect to the metric arising in this fashion from a *norm*, by definition $C^o(K)$ is a *Banach space.*

Proof: This is a typical three-epsilon argument. The point is the *completeness*, namely that a Cauchy sequence of continuous functions has a *pointwise* limit which is a continuous function. First we observe that a Cauchy sequence f_i does have a pointwise limit. Given $\varepsilon > 0$, choose N large enough such that for $i, j \ge N$ we have $|f_i - f_j| < \varepsilon$. Then, for any x in K, $|f_i(x) - f_j(x)| < \varepsilon$. Thus, the sequence of values $f_i(x)$ is a Cauchy sequence of complex numbers, so has a limit $f(x)$. Further, given $\varepsilon' > 0$, choose $j \ge N$ sufficiently large such that $|f_j(x) - f(x)| < \varepsilon'$. Then for all $i \ge N$

$$|f_i(x) - f(x)| \;\le\; |f_i(x) - f_j(x)| + |f_j(x) - f(x)| \;<\; \varepsilon + \varepsilon'$$

[33] Awareness of such possibilities and figuring out how to avoid them was the fruit of embarrassing errors and experimentation throughout the 19^{th} century. Unifying abstract notions such as *metric space* and general *topological space* only became available in the early 20^{th} century, with the work of Hausdorff, Fréchet, and others.

[34] As usual, a *(continuous) functional* is a (continuous) linear map to $\mathbb{C}$.

Since this is true for every positive ε'

$$|f_i(x) - f(x)| \le \varepsilon \qquad \text{(for all } i \ge N)$$

This holds for every x in K, so the pointwise limit is uniform in x.

Now prove that $f(x)$ is continuous. Given $\varepsilon > 0$, let N be large enough so that for $i, j \ge N$ we have $|f_i - f_j| < \varepsilon$. From the previous paragraph

$$|f_i(x) - f(x)| \le \varepsilon \qquad \text{(for every } x \text{ and for } i \ge N)$$

Fix $i \ge N$ and $x \in K$, and choose a small enough neigborhood U of x such that $|f_i(x) - f_i(y)| < \varepsilon$ for any y in U. Then

$$|f(x) - f(y)| \le |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f(y) - f_i(y)| < \varepsilon + \varepsilon + \varepsilon$$

Thus, the pointwise limit f is continuous at every x in U. ///

[18.2.5] Differentiation on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$

To talk about *differentiability* return to the concrete situation of $\mathbb{R}$ and its quotient $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$.

The continuous quotient map $q : \mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z}$ yields continuous functions under composition $f \circ q$ for $f \in C^o(\mathbb{T}) = C^o(\mathbb{R}/2\pi\mathbb{Z})$. More is true, namely, that a continuous function F on $\mathbb{R}$ is of the form $f \circ q$ if and only if F is *periodic* in the sense that $F(x + 2\pi n) = F(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. Indeed, the periodicity gives a *well-defined* function f on $\mathbb{R}/2\pi\mathbb{Z}$. Then the continuity of f follows immediately from the definition of the quotient topology on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$.

As usual, a real-valued or complex-valued function f on $\mathbb{R}$ is *continuously differentiable* if it has a derivative itself a continuous function. That is, the limit

$$\frac{df}{dx}(x) = f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

is required to exist for all x, *and* the function f' is in $C^o(\mathbb{R})$. Let $f^{(1)} = f'$, and inductively define

$$f^{(i)} = \Big(f^{(i-1)}\Big)' \qquad \text{(for } i > 1)$$

when the corresponding limits exist.

We can make explicit the expectation that *differentiation* on the circle $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ is descended from differentiation on the real line. That is, *characterize* differentiation on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ in terms of such a compatibility relation. Thus, for $f \in C^k(\mathbb{T})$, require that the differentiation D on $\mathbb{T}$ be related to the differentiation on $\mathbb{R}$ by

$$(Df) \circ q = \frac{d}{dx}(f \circ q)$$

Via the quotient map $q : \mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z}$, make a *preliminary* definition of the collection of k-times continuously differentiable functions on $\mathbb{T}$, with a topology, by

$$C^k(\mathbb{T}) = \{f \text{ on } \mathbb{T} : f \circ q \in C^k(\mathbb{R})\}$$

with the C^k-*norm*[35]

$$|f|_{C^k} = \sum_{0 \le i \le k} |(f \circ q)^{(i)}|_\infty = \sum_{0 \le i \le k} \sup_x |(f \circ q)^{(i)}(x)|$$

[35] Granting that the sup norm on *continuous* functions is a norm, verification that the C^k-norm is a norm is straightforward.

where $F^{(i)}$ is the (continuous!) i^{th} derivative of F on $\mathbb{R}$. The *associated metric* on $C^k(\mathbb{T})$ is

$$d(f, g) \;=\; |f - g|_{C^k}$$

[18.2.6] Remark: Among other features, the norm on the spaces C^k makes continuity of the *differentiation* map $C^k \to C^{k-1}$ clear.

[18.2.7] Remark: Implicit in this definition is that, viewed as functions on $[0, 2\pi]$, the values and derivatives *must agree at the endpoints*: $f(0) = f(2\pi)$ for f continuous on $\mathbb{T}$, $f'(0) = f'(2\pi)$ for $f \in C^1(\mathbb{T})$, and so on. This is not whimsical, but is intrinsic to the structure of $\mathbb{T}$.

An often-seen equivalent version of the norm is

$$|f|^{\text{var}}_{C^k} \;=\; \sup_{0 \le i \le k} |(f \circ q)^{(i)}|_\infty \;=\; \sup_{0 \le i \le k} \sup_x |(f \circ q)^{(i)}(x)|$$

These two norms give the same topology, since for complex numbers $a_0, \ldots, a_k$

$$\sup_{0 \le i \le k} |a_i| \;\le\; \sum_{0 \le i \le k} |a_i| \;\le\; (k+1) \cdot \sup_{0 \le i \le k} |a_i|$$

[18.2.8] Claim: There is a unique, well-defined, continuous (differentiation) map $D : C^k(\mathbb{T}) \to C^{k-1}(\mathbb{T})$ giving a commutative diagram

$$\begin{array}{ccc} C^k(\mathbb{R}) & \xrightarrow{d/dx} & C^{k-1}(\mathbb{R}) \\ {\scriptstyle -\circ q}\uparrow & & \uparrow{\scriptstyle -\circ q} \\ C^k(\mathbb{T}) & \xrightarrow{D} & C^{k-1}(\mathbb{T}) \end{array}$$

[18.2.9] Remark: One might feel that this proof is needlessly complicated. However, it is worthwhile to do it this way. This approach applies broadly, *and* is as terse as possible without ignoring important details.

Proof: The point is that differentiation of periodic functions yields periodic functions. That is, we claim that, for $f \in C^k(\mathbb{T})$, the pullback $f \circ q$ has derivative $\frac{d}{dx}(f \circ q)$ which is the pullback $g \circ q$ of a unique function $g \in \mathrm{C}^{k-1}(\mathbb{T})$. To see this, first recall that, by definition of the quotient topology, a continuous function F on $\mathbb{R}$ descends to a continuous function on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ if and only if it is $2\pi\mathbb{Z}$-invariant, that is $F(x + 2\pi n) = F(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. Then, from our definition of $C^k(\mathbb{T})$, a function $F \in C^k(\mathbb{R})$ is a pullback via q from $C^k(\mathbb{R}/2\pi\mathbb{Z})$ exactly when $F^{(i)}(x + 2\pi n) = F^{(i)}(x)$ for all $x \in \mathbb{R}$, $n \in \mathbb{Z}$, and $0 \le i \le k$, since then these continuous functions descend to the circle. Let

$$(T_y F)(x) \;=\; F(x + y) \qquad\qquad (\text{for } x, y \in \mathbb{R})$$

Since $\frac{d}{dx}$ is a linear, constant-coefficient differential operator, the operations T_y and $\frac{d}{dx}$ commute, that is, $\frac{\partial F}{\partial x}(x + y) = \frac{\partial}{\partial x}(F(x + y))$, which is to say

$$T_y \circ \frac{d}{dx} \;=\; \frac{d}{dx} \circ T_y$$

In particular, for $n \in \mathbb{Z}$,

$$T_{2\pi n}(\frac{d}{dx}(f \circ q)) \;=\; \frac{d}{dx}(T_{2\pi n}(f \circ q)) \;=\; \frac{d}{dx}(f \circ q)$$

This shows that a (continuous) derivative is periodic when the (continuously differentiable) function is periodic.

From the definition of the C^k-norm,

$$|Df|_{C^{k-1}} \;\le\; |f|_{C^k}$$

so differentiation is continuous. ///

[18.2.10] Remark: In light of the uniqueness of differentiation on $\mathbb{T}$, from now on write d/dx for the differentiation D on $\mathbb{T}$, and $f^{(k)}$ for $D^k f$, and rewrite the description of $C^k(\mathbb{T})$ more simply, as

$$C^k(\mathbb{T}) \;=\; \{f \;\text{ on } \mathbb{T} :\; f\circ q \in C^k(\mathbb{R})\}$$

with the C^k-*norm*

$$|f|_{C^k} \;=\; \sum_{0\le i\le k} |f^{(i)}|_\infty \;=\; \sum_{0\le i\le k} \sup_x |f^{(i)}(x)|$$

where $f^{(i)}$ is the (continuous!) i^{th} derivative of f. The *associated metric* on $C^k(\mathbb{T})$ still is

$$d(f,g) \;=\; |f-g|_{C^k}$$

There is the alternative norm

$$|f|^{\text{var}}_{C^k} \;=\; \sup_{0\le i\le k}\,\sup_x |f^{(i)}(x)| \;=\; \sup_{0\le i\le k} |f^{(i)}|_\infty$$

Again, these two norms give the same topology, for the same reason as before.

[18.2.11] Claim: With the topology above, the space $C^k(\mathbb{T})$ is *complete*, so is a *Banach* space.

Proof: The case $k=1$ illustrates all the points. For a Cauchy sequence $\{f_n\}$ in $C^1(\mathbb{T})$, both $\{f_n\}$ and $\{f_n'\}$ are Cauchy in $C^o(\mathbb{T})$, so converge uniformly pointwise: let

$$f(x) \;=\; \lim_n f_n(x) \qquad\qquad g(x) \;=\; \lim_n f_n'(x)$$

The convergence is uniformly pointwise, so f and g are C^o. *If* we knew that f were pointwise differentiable, then the demonstrated continuity of $\frac{d}{dx} : C^1(\mathbb{T}) \to C^o(\mathbb{T})$ gives the expected conclusion, that $f' = g$.

What could go wrong? One issue is whether f *is differentiable* at all, and why its derivative is g.

By the fundamental theorem of calculus, for any index i, since f_i is continuous,[36]

$$f_i(x) - f_i(a) \;=\; \int_a^x f_i'(t)\,dt$$

Interchanging limit and integral[37] shows that the limit of the right-hand side is

$$\lim_i \int_a^x f_i'(t)\,dt \;=\; \int_a^x \lim_i f_i'(t)\,dt \;=\; \int_a^x g(t)\,dt$$

Thus, the limit of the left-hand side is

$$f(x) - f(a) \;=\; \int_a^x g(t)\,dt$$

[36] The fundamental theorem of calculus for integrals of *continuous* functions needs only the simplest notion of an integral, for example, Riemann integrals.

[37] For example, interchange of limit and integral is justified by the simplest form of Lebesgue's Dominated Convergence Theorem. Also, for uniform pointwise limits of continuous functions, this can be proven directly.

from which $f' = g$. That the derivative f' of the limit f is the limit of the derivatives is not a surprise, since if f is differentiable, what else could its derivative be? The point is that f *is* differentiable, ascertained by computing its derivative, which happens to be g. ///

[18.2.12] Remark: Again, the differentiation map $C^1(\mathbb{T}) \to C^o(\mathbb{T})$ is continuous *by design*. Thus, *if* a limit of C^1 functions f_n *is* differentiable, its derivative must be the obvious thing, namely, the limit of the derivatives f_n'. The issue was whether the limit of the f_n is differentiable. The proof shows that it is differentiable by computing its derivative via the Mean Value Theorem.

By construction, and from the corresponding result for C^o,

[18.2.13] Claim: With the C^k-topology, for $x \in \mathbb{T}$ and integer $0 \le i \le k$, the *evaluation functional* $C^k(\mathbb{T}) \to \mathbb{C}$ by

$$f \longrightarrow f^{(i)}(x)$$

is *continuous*. ///

This applies to Fourier series, without any claim about what functions are representable as Fourier series. With $\psi_n(x) = e^{inx}$,

[18.2.14] Claim: For complex numbers c_n, when

$$\sum_n |c_n| \cdot |n|^k \ < \ +\infty$$

the Fourier series $\sum c_n\, \psi_n$ converges to a function in $C^k(\mathbb{T})$, and its derivative is computed by termwise differentiation

$$\frac{d}{dx} \sum c_n\, \psi_n \ = \ \sum (in)\, c_n\, \psi_n \in C^{k-1}(\mathbb{T})$$

Proof: The $C^o(\mathbb{T})$ norm of a Fourier series is easily estimated, by

$$\Big| \sum_{|n| \le N} c_n\, \psi_n(x) \Big| \ \le \ \sum_{|n| \le N} |c_n| \qquad \text{(for all } x \in \mathbb{T})$$

The right-hand side is independent of $x \in \mathbb{T}$, so bounds the sup over $x \in \mathbb{T}$. Similarly, estimate derivatives (of partial sums) by

$$\Big| \Big(\sum_{|n| \le N} c_n\, \psi_n \Big)^{(k)} \Big| \ \le \ \sum_{|n| \le N} |c_n|\, n^k$$

Thus, the hypothesis of the claim implies that the partial sums form a Cauchy sequence in C^k. The partial sums of a Fourier series are *finite* sums, so can be differentiated term-by-term. Thus, we have a Cauchy sequence of C^k functions, which converges to a C^k function, by the completeness of C^k. That is, the given estimate assures that the Fourier series converges to a C^k function.

Further, since differentiation is a continuous map $C^k \to C^{k-1}$, it maps Cauchy sequences to Cauchy sequences. In particular, the Cauchy sequence of derivatives of partial sums converges to the derivative of the limit of the original Cauchy sequence. ///

We *want* the following to hold. Unsurprisingly, it does:

[18.2.15] Claim: The inclusion $C^k(\mathbb{T}) \subset C^{k-1}(\mathbb{T})$ is continuous. [38]

[38] In fact, the image of C^k in C^{k-1} is *dense*, but, we will prove this later as a side-effect of sharper results.

Proof: The point is that, for $f \in C^k(\mathbb{T})$ the obvious inequality

$$|f|_{C^{k-1}} \;\le\; |f|_{C^k}$$

gives an explicit estimate for the continuity. ///

18.3 Topology on $C^\infty(\mathbb{T})$

Next, we care about infinitely differentiable [39] functions, that is, *smooth* functions, denoted $C^\infty(\mathbb{T})$. At least as *sets* (or vector spaces),

$$C^\infty(\mathbb{T}) \;=\; \bigcap_k C^k(\mathbb{T})$$

However, this space $C^\infty(\mathbb{T})$ of smooth functions provably *does not* have a structure of Banach space. Observing that a descending intersection is a *(projective) limit* we should declare that

$$C^\infty(\mathbb{T}) \;=\; \lim_k C^k(\mathbb{T})$$

That is, for every topological vector space V and compatible [40] family of continuous linear maps $f_k : V \to C^k(\mathbb{T})$, there is a unique $f : V \to C^\infty(\mathbb{T})$ such that all triangles commute in the diagram

$$\begin{tikzcd}
C^\infty(\mathbb{T}) \arrow[rr, bend left] \arrow[rrr, bend left] & \ldots \arrow[r] & C^1(\mathbb{T}) \arrow[r] & C^o(\mathbb{T}) \\
& V \arrow[ul, dotted, "f"] \arrow[ur, dashed, "f_1"] \arrow[urr, dashed, "f_o"'] & &
\end{tikzcd}$$

Unfortunately, we may be temporarily insufficiently sophisticated about what kind of object this limit might be. In particular, we do not know what kind of auxiliary objects to use in the very definition of *limit*.

Too optimistic speculation about what the limit might be leads to trouble: as it happens, this limit is provably *not* a Banach space (nor Hilbert space). [41] As we have seen, a limit of topological spaces has a unique topology, whatever it may be, by the categorical characterization of this topology.

[18.3.1] Remark: There is also the disquieting question of what *test objects* V we should consider in the diagrammatic characterization, with compatible mappings $V \to C^k(\mathbb{T})$ to *characterize* the limit.

The broadest necessary class of vector spaces with topologies is the following. A *topological vector space* is what one would reasonably imagine, namely, a (complex) vector space V with a topology such that

$$V \times V \to V \quad \text{by} \quad v \times w \to v + w \quad \text{is continuous}$$

---

[39] Use of *infinitely* here is potentially misleading, but is standard. Sometimes the phrase *indefinitely differentiable* is used, but this also offers its own potential for confusion. A better (and standard) contemporary usage is *smooth*.

[40] As earlier, for the maps f_k to be *compatible* means that, naming the inclusion $i_k : C^k(\mathbb{R}) \to C^{k-1}(\mathbb{R})$, $i_k \circ f_k = f_{k-1}$. That is, all the triangles in the relevant diagram commute.

[41] The non-Banach-ness of $C^\infty(\mathbb{T})$ is not the main point, but it is reasonable to wonder how this is proven. Briefly, with a definition of *topological vector space*, we will prove that a topological vector space is *normable* if and only if there is a local basis at 0 consisting of *bounded* opens. This is independent of *completeness*. The relevant sense of bounded *cannot* be the usual metric sense. Instead, a set E in a topological vector space is *bounded* when, for every open neighborhood U of 0, there is $t > 0$ such that $E \subset z \cdot U$ for all complex z with $|z| \ge t$. That is, sufficiently large *dilates* of opens eventually contain E. But we will eventually that open balls in $C^k(\mathbb{T})$ are *not* contained in *any* dilate of any open ball in $C^{k+1}(\mathbb{T})$. The definition of the limit topology then shows that $C^\infty(\mathbb{T})$ is not normable. A more detailed discussion will be given later.

and such that

$$\mathbb{C} \times V \to V \quad \text{by} \quad \alpha \times v \to \alpha \cdot v \quad \text{is continuous}$$

and such that the topology is *Hausdorff*. [42] We require that the topological vector spaces be *locally convex* in the sense that there is a local basis at 0 consisting of *convex* sets. [43] It is easy to prove that Hilbert and Banach spaces are locally convex, which is why the issue is invisible in that context. Dismayingly, there are easily constructed complete (invariantly) metrized topological vector spaces which are *not* locally convex. [44]

Returning to the discussion of limits of topological vector spaces: since the continuity requirements for a topological vector space are of the form $A \times B \to C$ (rather than having the arrow going the other direction), there is a *diagrammatic* argument that the continuous algebraic operations on the limitands induce continuous algebraic operations on the limit, in the limit topology (as limit of topological spaces).

[18.3.2] Claim: Products and limits of topological vector spaces exist. Products and limits of locally convex spaces are locally convex. (Proof in appendix.)

[18.3.3] Remark: As usual, if they exist at all, then products and limits are unique up to unique isomorphism.

Thus, $C^\infty(\mathbb{T})$ has a (limit) topology for general reasons. As proven earlier for such spaces on intervals $[a, b]$,

[18.3.4] Claim: Differentiation $f \to f'$ is a *continuous* map $C^\infty(\mathbb{T}) \to C^\infty(\mathbb{T})$.

[18.3.5] Remark: *Of course* differentiation maps the smooth functions to themselves. Continuity of differentiation in the *limit* topology is less clear.

Proof: Differentiation d/dx gives a continuous map $C^k(\mathbb{T}) \to C^{k-1}(\mathbb{T})$. Differentiation is compatible with the inclusions among the $C^k(\mathbb{T})$. Thus, we have a commutative diagram

$$\begin{tikzcd}
C^\infty(\mathbb{T}) \arrow[rr, bend left] \arrow[rrr, bend left] & \dots & C^k(\mathbb{T}) \arrow[r] & C^{k-1}(\mathbb{T}) \arrow[r] & \dots \\
C^\infty(\mathbb{T}) \arrow[rr, bend right] \arrow[rrr, bend right] & \dots & C^k(\mathbb{T}) \arrow[r] \arrow[ur, "\frac{d}{dx}"] & C^{k-1}(\mathbb{T}) \arrow[r] \arrow[ur, "\frac{d}{dx}"] & \dots
\end{tikzcd}$$

Composing the projections with d/dx gives (dashed) induced maps from $C^\infty(\mathbb{T})$ to the limitands, inducing a unique (dotted) map to the limit, as in

$$\begin{tikzcd}
C^\infty(\mathbb{T}) \arrow[rr, bend left] \arrow[rrr, bend left] & \dots & C^k(\mathbb{T}) \arrow[r] & C^{k-1}(\mathbb{T}) \arrow[r] & \dots \\
C^\infty(\mathbb{T}) \arrow[u, dotted, "\frac{d}{dx}"] \arrow[urr, dashed] \arrow[urrr, dashed] \arrow[urrrr, dashed] \arrow[rr, bend right] \arrow[rrr, bend right] & \dots & C^k(\mathbb{T}) \arrow[r] \arrow[ur] & C^{k-1}(\mathbb{T}) \arrow[r] \arrow[ur] & \dots
\end{tikzcd}$$

[42] In fact, soon after giving the definition, one can show that the weaker condition that *points are closed*, implies the Hausdorff condition in topological spaces which are vector spaces with continuous vector addition and scalar multiplication. Indeed, the inverse image of $\{0\}$ under $x \times y \to x - y$ is the diagonal.

[43] This sense of convexity is the usual: a set X in a vector space is convex when, for all tuples $x_1, \ldots, x_n$ of points in X and all tuples $t_1, \ldots, t_n$ of non-negative reals with $\Sigma_i t_i = 1$, the sum $\Sigma_i t_i x_i$ is again in X.

[44] The simplest examples of complete metric topological vector spaces which are *not* locally convex are spaces ℓ^p with $0 < p < 1$. The metric comes from a norm-like function which is *not* a norm: $|\{c_n\}|_p = \sum_n |c_n|^p$. No, there is no p^{th} root taken, unlike the spaces ℓ^p with $p \geq 1$, and this causes the function $|\ |_p$ to *lose* the homogeneity it would need to be a norm. Nevertheless, such a space is *complete*. It is an amusing exercise to prove that it is not locally convex.

This proves the continuity of differentiation, in the limit topology. ///

[18.3.6] Corollary: When a Fourier series $\sum_n c_n \psi_n$ satisfies

$$\sum_m |c_n|\,|n|^N \;<\; +\infty \qquad\qquad \text{(for every } N\text{)}$$

the series is a smooth function, which can be differentiated term-by-term, and its derivative is

$$\sum_m c_n \cdot in \cdot \psi_n$$

Proof: The hypothesis assures that the Fourier series lies in C^k for every k. Differentiation is continuous in the limit topology on C^∞. ///

[18.3.7] Remark: This continuity is necessary to define differentiation of *distributions* below.

18.4 Distributions: generalized functions

Although much amplification is needed, having an appropriate topology on $C^\infty(\mathbb{T})$ allows the bare definition: a *distribution* or *generalized function*[45] on $\mathbb{T}$ is a *continuous linear functional*[46]

$$u : C^\infty(\mathbb{T}) \longrightarrow \mathbb{C}$$

Why a dual space? Unsurprisingly, especially with a precise *intrinsic* notion of *integral* on $\mathbb{T}$ in the next section, a function $\varphi \in C^o(\mathbb{T})$ gives rise to a distribution u_φ by *integration against* φ,

$$u_\varphi(f) \;=\; \int_{\mathbb{T}} f(x)\,\varphi(x)\,dx \qquad\qquad (\text{for } f \in C^\infty(\mathbb{T}))$$

Thus, we relax our notion of *function*, no longer requiring *pointwise values*, but only that a function can be *integrated against*. Then it may make sense to declare functionals in a dual space to be *generalized* functions. The vector space of distributions is denoted

$$\text{distributions} \;=\; \text{continuous dual of } C^\infty(\mathbb{T}) \;=\; \mathrm{Hom}^o_{\mathbb{C}}(C^\infty(\mathbb{T}),\mathbb{C}) \;=\; C^\infty(\mathbb{T})^*$$

That is, given a reasonable notion of integral, we have a continuous imbedding

$$C^o(\mathbb{T}) \;\subset\; C^\infty(\mathbb{T})^* \quad \text{by} \quad \varphi \longrightarrow u_\varphi \quad \text{where (again)} \quad u_\varphi(f) \;=\; \int_{\mathbb{T}} f(x)\,\varphi(x)\,dx \qquad (f \in C^\infty(\mathbb{T}))$$

Typically, the dual of a limit of topological vector spaces is not the colimit of the duals of the limitands. Duals of *colimits do* behave well, in the sense that in reasonable situations

$$\mathrm{Hom}(\mathrm{colim}_i X_i, Z) \;\approx\; \mathrm{lim}_i \mathrm{Hom}(X_i, Z)$$

---

[45] What's in a name? In this case, *generalized function* expresses the *intention* to think of distributions as extensions of ordinary functions, not as abstract things in a dual space.

[46] The standard usage is that a *functional* on a complex vector space V is a $\mathbb{C}$-linear map from V to $\mathbb{C}$. Continuity may or may not be required, and the topology in which continuity is required may vary. It is in this sense that there is a subject *functional analysis*.

But $C^\infty(\mathbb{T})$ is a *limit*, not a colimit. Luckily, the dual of a limit of *Banach spaces* is the colimit of the duals:

[18.4.1] Theorem: Let $X = \lim_i B_i$ be a limit of Banach spaces B_i with projections $p_i : X \to B_i$. Any $\lambda \in X^* = \mathrm{Hom}^o_{\mathbb{C}}(X, \mathbb{C})$ *factors through* some B_i. That is, there is $\lambda_j : B_j \to \mathbb{C}$ such that

$$\lambda \;=\; \lambda_j \circ p_j : X \to \mathbb{C}$$

Therefore,

$$(\lim_i B_i)^* \;\approx\; \mathrm{colim}_i B_i^*$$

Proof: Without loss of generality, each B_i is the closure of the image of X, since otherwise replace of each B_i by that closure.

Let U be an open neighborhood of 0 in $X = \lim_i B_i$ such that $\lambda(U)$ is inside the open unit ball at 0 in $\mathbb{C}$, by the continuity at 0. By properties of the limit topology [47] there are finitely-many indices $i_1, \ldots, i_n$ and open neighborhoods V_{i_t} of 0 in B_{i_t} such that

$$\bigcap_{t=1}^{n} p_{i_t}^{-1} V_{i_t} \;\subset\; U \qquad\qquad \text{(projections } p_i \text{ from the } \textit{limit } X)$$

To have λ factor (continuously) through a limitand B_j, we need a *single* condition to replace the conditions from $i_1, \ldots, i_n$. Let j be *any* index [48] with $j \geq i_t$ for all t, and

$$V_j' \;=\; \bigcap_{t=1}^{n} p_{i_t,j}^{-1} V_{i_t} \subset B_j$$

By the compatibility

$$p_{i_t}^{-1} \;=\; p_j^{-1} \circ p_{i_t,j}^{-1}$$

we have a single sufficient condition, namely $p_j^{-1} V_j' \subset U$. By the linearity of λ, for $\varepsilon > 0$

$$\lambda(\varepsilon \cdot p_j^{-1} V_j) \;=\; \varepsilon \cdot \lambda(p_j^{-1} V_j) \;\subset\; \varepsilon\text{-ball in } \mathbb{C}$$

By continuity [49] of scalar multiplication on B_j, $\varepsilon \cdot V_j'$ is an open containing 0 in B_j.

We claim that λ factors through $p_j X$ with the subspace topology from B_j. This makes $p_j X$ a *normed* space, if not Banach. [50] Simplifying notation, let $\lambda : X \to \mathbb{C}$ and $p : X \to N$ be continuous linear to a normed space N, with

$$\lambda(p^{-1} V) \;\subset\; \text{unit ball in } \mathbb{C} \qquad\qquad \text{(for some neighborhood } V \text{ of 0 in } N)$$

[47] Recall that $X = \lim_i B_i$ is the closed subspace (with the subspace topology) of the product $Y = \Pi_i B_i$ of all tuples $\{b_i\}$ in which $p_{ij} : b_i \to b_j$ for $i > j$ under the transition maps $p_{ij} : B_i \to B_j$. A local basis at 0 in the product consists of products $V = \Pi_i V_i$ of opens V_i in B_i with $V_i = B_i$ for all but finitely-many i, say $i_1, \ldots, i_n$.

[48] The index set need not be the positive integers, but must be a *poset* (partially ordered set), *directed*, in the sense that for any two indices i, j there is an index k such that $k > i$ and $k > j$.

[49] Multiplication by a non-zero scalar is a *homeomorphism*: scalar multiplication by $\varepsilon \neq 0$ is continuous, scalar multiplication by ε^{-1} is continuous, and these are mutual inverses, so these scalar multiplications are *homeomorphisms*.

[50] Recall that a *normed* space is a topological vector with topology given by a *norm* $|\ \ |$ as in a Banach space, but *without* the requirement that the space is *complete* with respect to the metric $d(x, y) = |x - y|$. This slightly complicated assertion is correct: in most useful situations $p_j X$ is rarely all of B_j, even when B_j is a *completion* of $p_j X$.

We claim that λ factors through $p : X \to N$ as a (continuous) linear map. Indeed, by the linearity of λ,

$$\lambda(\frac{1}{n} \cdot p^{-1}V) \subset \frac{1}{n}\text{-ball in } \mathbb{C}$$

so

$$\lambda\left(\bigcap_n \frac{1}{n} \cdot p^{-1}V\right) \subset \frac{1}{m}\text{-ball} \qquad\qquad (\text{for all } m)$$

Then

$$\lambda\left(\bigcap_n \frac{1}{n} \cdot p^{-1}V\right) \subset \bigcap_m \frac{1}{m}\text{-ball} = \{0\}$$

Thus,

$$\bigcap_n p^{-1}(\frac{1}{n} \cdot V) = \bigcap_n \frac{1}{n} \cdot p^{-1}V \subset \ker\lambda$$

For x, x' in X with $px = px'$, certainly $px - px' \in \frac{1}{n}V$ for all $n = 1, 2, \ldots$. Therefore,

$$x - x' \in \bigcap_n p^{-1}(\frac{1}{n}V) \subset \ker\lambda$$

and $\lambda x = \lambda x'$. This proves the subordinate claim that λ factors through $p : X \to N$ via a (not necessarily continuous) linear map $\mu : N \to \mathbb{C}$. For the continuity of μ, by its linearity

$$\mu(\varepsilon \cdot V) = \varepsilon \cdot \mu V \subset \varepsilon\text{-ball in } \mathbb{C}$$

proving the continuity of $\mu : N \to \mathbb{C}$. [51] This proves the claim.

The claim gives continuous linear $\lambda_j : p_j X \to \mathbb{C}$ through which λ factors.

Then $\lambda_j : p_j X \to \mathbb{C}$ extends by continuity [52] to the closure of $p_j X$ in B_j, which is B_j, giving the desired map. ///

[18.4.2] Remark: The same proof shows that a continuous linear map from a limit of Banach spaces to a *normed* space factors through a limitand, when the images of projections are dense in the limitands.

[18.4.3] Corollary: The space of distributions on $\mathbb{T}$ is the ascending union (colimit)

$$C^\infty(\mathbb{T})^* = \left(\lim{}_k C^k(\mathbb{T})\right)^* = \text{colim}_k C^k(\mathbb{T})^* = \bigcup{}_k C^k(\mathbb{T})^*$$

of duals of the Banach spaces $C^k(\mathbb{T})$. ///

[51] Here we need V to be *open*, not merely a *set* containing 0. Continuity at 0 is all that is needed for continuity of *linear* maps, since $|\lambda(x)| < \varepsilon$ for $|x| < \delta$ gives $|\lambda(x - x')| < \varepsilon$ for $|x - x'| < \delta$.

[52] The extension by continuity is unambiguous, since λ_j is *linear*. In more detail: for λ a continuous linear function on a dense subspace Y of a topological vector space X, given $\varepsilon > 0$, take *convex* neighborhood U of 0 in X such that $|\lambda y| < \varepsilon$ for $y \in U$. We may suppose $U = -U$ by replacing U by $-U \cap U$. Let y_i be a Cauchy net approaching $x \in X$. For y_i and y_j inside $x + \frac{1}{2}U$, $|\lambda y_i - \lambda y_j| = |\lambda(y_i - y_j)|$, using the linearity. By the symmetry $U = -U$, since $y_i - y_j \in \frac{1}{2} \cdot 2U = U$, this gives $|\lambda y_i - \lambda y_j| < \varepsilon$. Then unambiguously define λx to be the limit of the λy_i.

The *order* of a distribution u is the smallest k such that $u \in C^k(\mathbb{T})^*$. Since for the circle the space of all distributions is exactly this colimit, the order of a distribution is well-defined. [53]

Distributions as generalized *functions* should be *differentiable*, compatibly with the differentiation of *functions*. The idea is that differentiation of distributions should be compatible with *integration by parts* for distributions given by integration against C^1 functions. Assuming an integral on $\mathbb{T}$ as in the next section, for *functions* f, g, by integration by parts,

$$\int_{\mathbb{T}} f(x)\, g'(x)\, dx \;=\; -\int_{\mathbb{T}} f'(x)\, g(x)\, dx$$

with no boundary terms because $\mathbb{T}$ has empty boundary. Note the negative sign. Motivated by this, define the *distributional derivative* u' of $u \in C^\infty(\mathbb{T})^*$ to be another distribution defined by

$$u'(f) \;=\; -u(f') \qquad\qquad (\text{for any } f \in C^\infty(\mathbb{T}))$$

The continuity of differentiation $\frac{d}{dx} : C^\infty(\mathbb{T}) \to C^\infty(\mathbb{T})$ assures that u' is a distribution, since

$$u' = -(u \circ \frac{d}{dx}) \;:\; C^\infty(\mathbb{T}) \to \mathbb{C}$$

18.5 Invariant integration, periodicization

We an *(invariant) integral* on the circle $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. The main property required is *translation invariance*, meaning that, for a (for example) continuous function f on $\mathbb{T}$,

$$\int_{\mathbb{T}} f(x+y)\, dx \;=\; \int_{\mathbb{T}} f(x)\, dx \qquad\qquad (\text{for all } y \in \mathbb{T})$$

This invariance is sufficient to prove that various important integrals *vanish*.

For example, let $\psi_m(x) = e^{imx}$. As an instance of an important idea, without explicit calculus-like computations,

[18.5.1] Claim: *(Cancellation Lemma)* For $m \neq n$, for *any* reasonable translation-invariant integral on $\mathbb{T}$

$$\int_{\mathbb{T}} \psi_m(x)\, \overline{\psi}_n(x)\, dx \;=\; 0$$

Proof: For $m \neq n$, the function $f(x) = \psi_m(x)\overline{\psi}_n(x)$ is a *non-trivial* (not identically 1) continuous group homomorphism $\mathbb{T} \to \mathbb{C}^\times$, meaning that there is $y \in \mathbb{T}$ such that $f(y) \neq 1$. The *change of variables* $x \to x+y$ in the integral does not change the overall value of the integral, so

$$\int_{\mathbb{T}} f(x)\, dx \;=\; \int_{\mathbb{T}} f(x+y)\, dx \;=\; \int_{\mathbb{T}} f(x)\cdot f(y)\, dx \;=\; f(y)\int_{\mathbb{T}} f(x)\, dx$$

[53] The *Riesz representation theorem* asserts that the dual of $C^o(\mathbb{T})$ is *Borel measures* on $\mathbb{T}$, so order-zero distributions are *Borel measures*. For example, elements η of $L^2(\mathbb{T})$ are Borel measures, by giving integrals $f \to \int_{\mathbb{T}} f(x)\, \eta(x)\, dx$ for $f \in C^o(\mathbb{T})$. Thus, integrating continuous functions against Borel measures is a semi-classical instance of *generalizing functions* in our present style, integrating against *measures*. However, the duals of the higher $C^k(\mathbb{T})$'s don't have such a classical interpretation. The fact that $C^1(\mathbb{T})$ can be construed as *distributional derivatives of Borel measures* is not strongly related to Radon-Nikodym derivatives of measures, because, for example, the distributional derivative of a point-mass measure is not a measure.

Thus, the integral I has the property that $I = t \cdot I$ where $t \neq 1$. This gives $(1-t) \cdot I = 0$, so $I = 0$ since $t \neq 1$. ///

[18.5.2] Remark: This *vanishing trick* is impressive, since nothing specific about the continuous group homomorphism f or topological group ($\mathbb{T}$ here) is used, apart from the finiteness of the total measure of the group, which comes from its *compactness*. That is, the same proof would show that *integrals over compact groups of non-trivial group homomorphisms are* 0. However, a notion of *invariant measure* [54] for general groups requires effort. Nevertheless, with an invariant measure, the same argument succeeds.

Less critically than the invariance, we want a *normalization* [55]

$$\int_{\mathbb{T}} 1\, dx \;=\; \mathrm{vol}\,(\mathbb{T}) \;=\; \mathrm{vol}\,(\mathbb{R}/2\pi\mathbb{Z}) \;=\; 2\pi$$

Then

$$\int_{\mathbb{T}} |\psi_n(x)|^2\, dx \;=\; \int_{\mathbb{T}} 1\, dx \;=\; 2\pi$$

Thus, without any explicit presentation of the integral or measure, we have proven that the distinct exponentials are an *orthogonal set* with norms $\sqrt{2\pi}$ with respect to the inner product

$$\langle f, g\rangle \;=\; \int_{\mathbb{T}} f(x)\, \overline{g}(x)\, dx$$

An *integration by parts* formula should be expected, with no *boundary terms* since $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ has empty boundary. Indeed, without *constructing* the invariant integral, we prove what we want from its *properties*:

[18.5.3] Claim: Let $f \to \int_{\mathbb{T}} f(x)\, dx$ be an invariant integral on $\mathbb{T}$, for $f \in C^o(\mathbb{T})$. Then for $f \in C^1(\mathbb{T})$

$$\int_{\mathbb{T}} f'(x)\, dx \;=\; 0$$

and we have the *integration by parts* formula for $f, g \in C^1(\mathbb{T})$

$$\int_{\mathbb{T}} f(x)\, g'(x)\, dx \;=\; -\int_{\mathbb{T}} f(x)'\, g(x)\, dx$$

[18.5.4] Remark: Vanishing of integrals of derivatives does *not* depend on the particulars of the situation. The same argument succeeds on an arbitrary group possessing (translation) invariant differentiation(s) and an invariant integral. Thus, the specific geometry of the circle is *not* needed to argue that $\int_{\mathbb{T}} f'(x)dx = \int_0^{2\pi} f(x)\, dx = f(2\pi) - f(0) = 0$ because f is periodic. The latter classical argument is valid, but fails to show a generally applicable mechanism. The same independence of particulars applies to the integration by parts rule.

Proof: The translation invariance of the integral makes the integral of a derivative 0, by direct computation, as follows. We interchange a *differentiation* and an *integral*. [56]

$$\int_{\mathbb{T}} f'(x)\, dx \;=\; \int_{\mathbb{T}} \frac{\partial}{\partial t}\Big|_{t=0}\; f(x+t)\, dx \;=\; \frac{d}{dt}\Big|_{t=0} \int_{\mathbb{T}} f(x+t)\, dx \;=\; \frac{d}{dt}\Big|_{t=0} \int_{\mathbb{T}} f(x)\, dx \;=\; 0$$

[54] Translation-invariant measures on topological groups are called *Haar measures*. General proof of their *existence* takes a little work, and invokes the Riesz representation theorem. *Uniqueness* can be made to be an example of a more general argument about uniqueness of invariant functionals.

[55] The measure of the circle need not be normalized to be 2π, but this is natural when presenting it as $\mathbb{R}/2\pi\mathbb{Z}$.

[56] The argument bluntly demands this interchange of limit and differentiation, so *justification* of it is secondary to the act itself. In the near future this and many other necessary interchanges are definitively justified via *Gelfand-Pettis* (also called *weak*) integrals. In the present concrete situation elementary (but opaque) arguments could be invoked, but we do not do this.

by changing variables in the integral. Then apply this to the function $(f \cdot g)' = f'g + fg'$ to obtain

$$\int_{\mathbb{T}} f'(x)\, g(x)\, dx + \int_{\mathbb{T}} f(x)\, g'(x)\, dx \;=\; 0$$

which gives the integration by parts formula. ///

The usual (Lebesgue) integral on the uniformizing $\mathbb{R}$ has the corresponding property of translation invariance. Since we present the circle as a quotient $\mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z} = \mathbb{T}$ of $\mathbb{R}$ we expect a *compatibility*[57]

$$\int_{\mathbb{R}} F(x)\, dx \;=\; \int_{\mathbb{R}/2\pi\mathbb{Z}} \left(\sum_{n \in \mathbb{Z}} F(x + 2\pi n) \right) dx$$

for at least *compactly-supported* continuous functions F on $\mathbb{R}$.

Indeed, we can *define* integrals of functions on $\mathbb{T}$ by this compatibility relation, by expressing a continuous function f on $\mathbb{T}$ as a *periodicization* (or *automorphization*)

$$f(x) \;=\; \sum_{n \in \mathbb{Z}} F(x + 2\pi n)$$

of a compactly supported continuous function F on $\mathbb{R}$, and *define*

$$\int_{\mathbb{T}} f(x)\, dx \;=\; \int_{\mathbb{R}} F(x)\, dx$$

We still need to prove that this value is independent of the choice of F for given f.

The properties required of an integral on $\mathbb{T}$ are clear. Sadly, we are not in a good position (yet) either to prove *uniqueness* or to give a *construction* as gracefully as these ideas deserve.

Postponing a systematic approach, we neglect any proof of uniqueness, and for a construction revert to an ugly-but-tangible reduction of the problem to integration on an interval. That is, note that in the quotient $q : \mathbb{R} \to \mathbb{R}/2\pi\mathbb{Z} = \mathbb{T}$ the interval $[0, 2\pi]$ maps surjectively, with the endpoints being identified (and no other points identified). In traditional terminology, $[0, 2\pi]$ is a *fundamental domain*[58] for the action of $\mathbb{Z}$ on $\mathbb{R}$. Then define the integral of f on $\mathbb{T}$ by

$$\int_{\mathbb{T}} f(x)\, dx \;=\; \int_0^{2\pi} (f \circ q)(x)\, dx$$

with usual (Lebesgue) measure on the unit interval. Verification of the *compatibility* with integration on $\mathbb{R}$ is silly, from this viewpoint.

---

[57] In contrast to many sources, this compatibility is *not* about choosing representatives in $[0, 2\pi)$ or anywhere else for $\backslash T$. Rather, this compatibility would be required for a group G (here $\mathbb{R}$), a discrete subgroup Γ (here $2\pi\mathbb{Z}$), and the quotient G/Γ (here $\mathbb{T}$), whether or not that quotient is otherwise identifiable. This compatibility is a sort of *Fubini theorem.* The usual Fubini theorem applies to products $X \times Y$, whose quotients $(X \times Y)/X \approx Y$ are simply the factors, but another version applies to quotients that are not necessarily factors.

[58] The notion of *fundamental domain* for the action of a group Γ on a set X has an obvious appeal, at least that it is more concrete than the notion of *quotient* $\Gamma\backslash X$. However, it is rarely possible to determine an exact fundamental domain, *and* one eventually discovers that the details are seldom useful even if this is possible. Instead, the *quotient* should be treated directly.

This (bad) definition does allow explicit computations, but makes *translation invariance* harder to prove, since the unit interval gets pushed off itself by translation. But we can still manage the verification. [59] Take $y \in \mathbb{R}$, and compute

$$\int_{\mathbb{T}} f(x+y)\,dx \;=\; \int_0^{2\pi} (f\circ q)(x+y)\,dx \;=\; \int_{-y}^{2\pi-y} (f\circ q)(x)\,dx$$

$$= \int_{-y}^{0} (f\circ q)(x)\,dx + \int_0^{2\pi-y} (f\circ q)(x)\,dx \;=\; \int_{-y}^{0} (f\circ q)(x-2\pi)\,dx + \int_0^{2\pi-y} (f\circ q)(x)\,dx$$

since $(f\circ q)(x) = (f\circ q)(x-2\pi)$ by periodicity. Then, replacing x by $x+2\pi$ in the first integral, this is

$$\int_{2\pi-y}^{2\pi} (f\circ q)(x)\,dx + \int_0^{2\pi-y} (f\circ q)(x)\,dx \;=\; \int_0^{2\pi} (f\circ q)(x)\,dx$$

18.6 Levi-Sobolev inequalities, Levi-Sobolev imbeddings

The simplest L^2 theory of Fourier series addresses neither continuity nor differentiability. [60] Yet it *would be* advantageous on general principles to be able to talk about differentiability in the context of Hilbert spaces, since Hilbert spaces have easily understood *dual spaces*. Beppo Levi, Frobenius, and Sobolev made useful comparisons. The idea is to compare C^k norms to norms coming from Hilbert spaces whose inner products refer to derivatives, the Levi-Sobolev spaces.

[18.6.1] Levi-Sobolev inequalities

First, we have an easy estimate for a variant C^k norm:

$$\Big|\sum_{|n|\le N} c_n\, e^{inx}\Big|_{C^k} \;=\; \sup_{0\le j\le k}\,\sup_x \Big|\sum_{|n|\le N} c_n\,(in)^j\, e^{inx}\Big|_{\mathbb{C}} \;\le\; \sum_{|n|\le N} |c_n|\cdot(1+n^2)^{k/2}$$

all for elementary reasons. [61] Perhaps surprisingly, rather try to directly obtain a sup norm estimate on this sum, Cauchy-Schwarz-Bunyakowsky is invoked: for any $s \in \mathbb{R}$

$$\Big|\sum_{|n|\le N} c_n\, e^{inx}\Big|_{C^k} \;\le\; \sum_{|n|\le N} |c_n|\cdot(1+n^2)^{s/2}\cdot\frac{1}{(1+n^2)^{(s-k)/2}}$$

$$\le\; \Big(\sum_{|n|\le N} |c_n|^2\cdot(1+n^2)^s\Big)^{1/2}\cdot\Big(\sum_{|n|\le N}\frac{1}{(1+n^2)^{s-k}}\Big)^{1/2}$$

Convergence of the elementary sum is easy to understand:

$$\sum_{n\in\mathbb{Z}} \frac{1}{(1+n^2)^{s-k}} \;<\; +\infty \qquad\qquad (\text{for } s > k+\tfrac{1}{2})$$

---

[59] While suppressing our disgust.

[60] It was not until the mid-20th century that L. Carleson showed, in L. Carleson, *On convergence and growth of partial sums of Fourier series*, Acta Math. **116** (1966), 135-157, that Fourier series of L^2 functions do converge pointwise almost everywhere. But this is a fragile sort of result.

[61] The awkward expression $(1+n^2)^{1/2}$ is approximately n. However, for $n=0$ we cannot divide by n, and replacing n by $(1+n^2)^{1/2}$ is the traditional device stunt to avoid this annoyance.

Thus, for any $s > k + \frac{1}{2}$ we have a *Levi-Sobolev inequality*

$$\Big|\sum_{|n|\le N} c_n\,\psi_n\Big|_{C^k} \;\le\; \Big(\sum_{n\in\mathbb{Z}} \frac{1}{(1+n^2)^{s-k}}\Big)^{1/2} \cdot \Big(\sum_{|n|\le N} |c_n|^2\cdot(1+n^2)^s\Big)^{1/2}$$

$$\le\; \Big(\sum_{n\in\mathbb{Z}} \frac{1}{(1+n^2)^{s-k}}\Big)^{1/2} \cdot \Big(\sum_{n\in\mathbb{Z}} |c_n|^2\cdot(1+n^2)^s\Big)^{1/2}$$

which is summarized as

$$\Big|\sum_{n\in\mathbb{Z}} c_n\,\psi_n\Big|_{C^k} \;\le\; \Big(\sum_{n\in\mathbb{Z}} \frac{1}{(1+n^2)^{s-k}}\Big)^{1/2} \cdot \Big(\sum_{n\in\mathbb{Z}} |c_n|^2\cdot(1+n^2)^s\Big)^{1/2} \qquad (\text{for } s > k+\tfrac{1}{2})$$

Existence of this comparison makes the right side interesting. Taking away from the right-hand side the uniform constant

$$\omega_{s-k} \;=\; \Big(\sum_{n\in\mathbb{Z}} \frac{1}{(1+n^2)^{s-k}}\Big)^{1/2}$$

gives the s^{th} *Levi-Sobolev norm*

$$s^{th} \text{ Levi-Sobolev norm} \;=\; \Big|\sum_{n\in\mathbb{Z}} c_n\,\psi_n\Big|_{H^s} \;=\; \Big(\sum_{n\in\mathbb{Z}} |c_n|^2\cdot(1+n^2)^s\Big)^{1/2}$$

Paraphrasing, we have proven the dominance relation

$$|\;|_{C^k} \;\le\; \omega_{s-k}\cdot|\;|_{H^s} \qquad (\text{for any } s > k+\tfrac{1}{2})$$

[18.6.2] Levi-Sobolev imbeddings

For $s \ge 0$, the s^{th} *Levi-Sobolev space* is[62]

$$H^s(\mathbb{T}) \;=\; \{f\in L^2(\mathbb{T}) : \sum_n |\widehat{f}(n)|^2\cdot(1+n^2)^s < +\infty\}$$

The inner product on $H^s(\mathbb{T})$ is

$$\Big\langle \sum_n a_n\,\psi_n,\; \sum_n b_n\,\psi_n\Big\rangle \;=\; 2\pi\sum_n a_n\,\overline{b}_n\,(1+n^2)^s$$

[18.6.3] Remark: This definition of $H^s(\mathbb{T})$ defines a useful space of functions or generalized functions only for $s \ge 0$, since for $s < 0$ the constraint $f \in L^2(\mathbb{T})$ is *stronger* (from the Plancherel theorem) than the condition defining $H^s(\mathbb{T})$ in the previous display.

[18.6.4] Remark: The 0^{th} Levi-Sobolev space is just $L^2(\mathbb{T})$.

[62] This definition is fine for $s \ge 0$, but not sufficient for $s < 0$. We will give the broader definition below. Keep in mind that $L^2(\mathbb{T})$ contains $C^o(\mathbb{T})$ and all the $C^k(\mathbb{T})$'s.

[18.6.5] Corollary: For $s > k + \frac{1}{2}$ there is a continuous inclusion

$$H^s(\mathbb{T}) \subset C^k(\mathbb{T})$$

Proof: For $s > k + \frac{1}{2}$, whenever a Fourier series has a finite Levi-Sobolev norm

$$\Big|\sum_n c_n\,\psi_n\Big|_{H^s} = \left(\sum_{n\in\mathbb{Z}} |c_n|^2 \cdot (1+n^2)^s\right)^{1/2} < +\infty$$

the partial sums of the Fourier series are Cauchy in H^s, hence Cauchy in C^k, so converge in the Banach space C^k:

$$\sum_n c_n\,\psi_n = C^k \text{ function on } \mathbb{T}$$

Proof: Apply the Levi-Sobolev inequality $|f|_{C^k} \le \omega \cdot |f|_{H^s}$ to finite linear combinations f of exponentials. Such finite linear combinations are C^k, and the inequality implies that an infinite sum of such, convergent in $H^s(\mathbb{T})$, has sequence of partial sums convergent in $C^k(\mathbb{T})$. That is, by the completeness of $C^k(\mathbb{T})$, the limit is still k times continuously differentiable. Thus, we have the containment. Given the containment, the inequality of norms implies the continuity of the inclusion. ///

[18.6.6] Levi-Sobolev Hilbert spaces

[18.6.7] Claim: The s^{th} Levi-Sobolev space $H^s(\mathbb{T})$ (with $0 \le s \in \mathbb{R}$) is a Hilbert space. In particular, the sequences of Fourier coefficients of functions in $H^s(\mathbb{T})$ are *all* two-sided sequences $\{c_n : n \in \mathbb{Z}\}$ of complex numbers meeting the condition

$$\sum_n |c_n|^2 \cdot (1+n^2)^s < +\infty$$

[18.6.8] Remark: It is clear that the exponentials ψ_n are an *orthogonal* basis for $H^s(\mathbb{T})$, although their norms depend on the index s. In particular, the collection of finite linear combinations of exponentials is *dense* in $H^s(\mathbb{T})$.

[18.6.9] Remark: Again, we do want to *define* these positively-indexed Levi-Sobolev spaces as subspaces of genuine spaces of functions, *not* as sequences of Fourier coefficients meeting the condition, and then *prove* the second assertion of the claim. This does leave open, for the moment, the question of how to define negatively-indexed Levi-Sobolev spaces.

Proof: In effect, this is the space of L^2 functions on which the H^s-norm is finite. If we prove the second assertion of the claim, then invoke the usual proof that L^2 spaces are complete to know that $H^s(\mathbb{T})$ is complete, since it is simply a weighted L^2-space. Given a two-sided sequence $\{c_n\}$ of complex numbers such that

$$\sum_n |c_n|^2 \cdot (1+n^2)^s < +\infty$$

since $s \ge 0$,

$$\sum_n |c_n|^2 < +\infty$$

and, by Plancherel,

$$\sum_n c_n\,\psi_n \in L^2(\mathbb{T})$$

This shows that $H^s(\mathbb{T})$ is a Hilbert space for $s \ge 0$. ///

[18.6.10] Remark: Insisting on viewing $L^2(\mathbb{T})$ as *equivalence classes* of functions may mislead us into making a needlessly complicated assertion about Levi-Sobolev imbeddings $H^s(\mathbb{T}) \subset C^k(\mathbb{T})$ for $s > k + \frac{1}{2}$, by insisting that $H^s(\mathbb{T})$ consists of almost-everywhere equivalence classes of $L^2(\mathbb{T})$ functions, only *one* of which is in $C^k(\mathbb{T})$. This is not a genuine issue.

[18.6.11] Levi-Sobolev norms in terms of derivatives

[18.6.12] Remark: Apart from having the virtue of giving inner-product structures, the expressions appearing in these Levi-Sobolev norms are *natural* because they have meaning in terms of L^2-norms of derivatives. For $f = \sum c_n \psi_n \in C^k(\mathbb{T})$, by Plancherel

$$\text{(norm via derivatives)} \;=\; |f|^2 + |f'|^2 + |f''|^2 + \ldots + |f^{(k)}|^2$$

$$= \sum_n |c_n|^2 \cdot (1 + n^2 + n^4 + \ldots + n^{2k}) \;\le\; \sum_n |c_n|^2 \cdot (1+n^2)^k$$

Conversely,

$$(1+n^2)^k \;\le\; C_k \cdot (1 + n^2 + n^4 + n^6 + \ldots + n^{2k}) \qquad \text{(for some constant } C_k\text{)}$$

so

$$\text{(norm via Fourier coefficients)} \;=\; \sum_n |c_n|^2 \cdot (1+n^2)^k \;\le\; C_k \cdot \Big(|f|^2 + |f'|^2 + |f''|^2 + \ldots + |f^{(k)}|^2\Big)$$

Thus, the two definitions of Levi-Sobolev norms, in terms of weighted L^2 norms of Fourier series, or in terms of L^2 norms of derivatives, give comparable Hilbert space structures. In particular, the *topologies* are identical.

[18.6.13] Corollary: For $k \ge 0$,

$$C^k(\mathbb{T}) \;\subset\; H^k(\mathbb{T})$$

Proof: For $k = 0$, the assertion is that $C^o(\mathbb{T}) \subset L^2(\mathbb{T})$, which holds because $\mathbb{T}$ is compact. Similarly, the relevant derivatives of $f \in C^k(\mathbb{T})$ are in $L^2(\mathbb{T})$, so $f \in H^k(\mathbb{T})$. ///

[18.6.14] Remark: One can work out the corresponding inequalities for Fourier series in several variables, proving that $(k + \frac{n}{2} + \varepsilon)$-fold L^2 differentiability (for any $\varepsilon > 0$) in dimension n is needed to assure k-fold continuous differentiability. This is L^2 Levi-Sobolev theory.

[18.6.15] Uniform pointwise convergence, convergence in $C^k(\mathbb{T})$

At this moment it is very easy to give a *straightforward*, if not *sharp*, result about convergence of C^k functions on $\mathbb{T}$, via the Levi-Sobolev spaces:

[18.6.16] Corollary: The Fourier series of $f \in C^k(\mathbb{T})$ converges to f in $C^{k-1}(\mathbb{T})$.

Proof: A function in $C^k(\mathbb{T})$ is in the Hilbert space $H^k(\mathbb{T})$, meaning that the finite partial sums of the Fourier expansion converge to f in $H^k(\mathbb{T})$. The $H^k(\mathbb{T})$ norm dominates that of $C^{k-1}(\mathbb{T})$, so the Fourier series converges to f in $C^{k-1}(\mathbb{T})$. ///

[18.6.17] Remark: It may seem mildly peculiar that the Fourier series of a C^k function can converge to it only in C^{k-1}.

[18.6.18] L^2-differentiation

[18.6.19] Claim: For every $s \geq 0$, the differentiation map

$$\frac{d}{dx} \;:\; \text{finite Fourier series} \;\longrightarrow\; \text{finite Fourier series}$$

is continuous when the source is given the $H^s(\mathbb{T})$ topology and the target is given the $H^{s-1}(\mathbb{T})$ topology.

Proof: This continuity is by design:

$$\Big|\frac{d}{dx}\sum_{|n|\leq N} c_n\, e^{inx}\Big|^2_{H^{s-1}} \;=\; \Big|\sum_{|n|\leq N} c_n\, in\, e^{inx}\Big|^2_{H^{s-1}} \;\leq\; \sum_{|n|\leq N} |nc_n|^2 \cdot (1+n^2)^{s-1}$$

$$\leq\; \sum_{|n|\leq N} |c_n|^2 \cdot (1+n^2)^s \;=\; \Big|\sum_{|n|\leq N} c_n\, e^{inx}\Big|^2_{H^s}$$

proving the continuity on finite Fourier series. ///

Therefore, we can *extend* $\frac{d}{dx}$ *by continuity* to obtain continuous linear maps

$$(L^2\text{-differentiation}) = (\text{extension by continuity of})\; \frac{d}{dx} \quad : \quad H^s(\mathbb{T}) \;\longrightarrow\; H^{s-1}(\mathbb{T})$$

[18.6.20] Remark: In these terms, *extra* L^2-differentiability is needed to assure comparable classical *continuous* differentiability. Specifically, $(k + \frac{1}{2} + \varepsilon)$-fold L^2-differentiability (for any $\varepsilon > 0$) suffices for k-fold *continuous* differentiability, in this one-dimensional example. *The comparable computations on* $(\mathbb{T})^{\times n}$ *show that the gap widens as the dimension grows.*

18.7 $C^\infty = \lim C^k = \lim H^s = H^\infty$

For larger purposes, the specific comparisons of indices in the containments

$$H^s(\mathbb{T}) \quad\subset\quad C^k(\mathbb{T}) \quad (\text{for } s > k + \tfrac{1}{2})$$

$$C^k(\mathbb{T}) \quad\subset\quad H^s(\mathbb{T}) \quad (\text{for } k \geq s)$$

are secondary, since we are more interested in *smooth functions* $C^\infty(\mathbb{T})$ than functions with *limited* continuous differentiability.

Thus, the point is that the Levi-Sobolev spaces and $C^k(\mathbb{T})$ spaces are *cofinal* under taking *descending intersections*. That is, letting $H^\infty(\mathbb{T})$ be the *intersection* of all the $H^s(\mathbb{T})$, as *sets* we have

$$C^\infty(\mathbb{T}) \;=\; \bigcap_k C^k(\mathbb{T}) \;=\; \bigcap_{s\geq 0} H^s(\mathbb{T}) \;=\; H^\infty(\mathbb{T})$$

Since descending nested intersections are *limits*, the topologies behave well for trivial reasons:

[18.7.1] Theorem: As topological vector spaces

$$C^\infty(\mathbb{T}) \;=\; \lim_k C^k(\mathbb{T}) \;=\; \lim_{s\geq 0} H^s(\mathbb{T}) \;=\; H^\infty(\mathbb{T})$$

Proof: The cofinality of the C^k's and the H^s's gives a natural isomorphism of the two limits, since they can be combined in a larger limit in which each is cofinal. ///

Again, in general duals of limits are not colimits, but we did show earlier that the dual of a limit of *Banach* spaces is the colimit of the duals of the Banach spaces. Thus,

[18.7.2] Corollary: The space of distributions on $\mathbb{T}$ is

$$C^\infty(\mathbb{T})^* \;=\; \mathrm{colim}_k C^k(\mathbb{T})^* \;=\; \mathrm{colim}_{s\ge 0} H^s(\mathbb{T})^* \;=\; H^\infty(\mathbb{T})^*$$

(and the duals $H^s(\mathbb{T})^*$ admit further explication, below). ///

Expressing $C^\infty(\mathbb{T})$ as a limit of the Hilbert spaces $H^s(\mathbb{T})$, as opposed to its more natural expression as a limit of the Banach spaces $C^k(\mathbb{T})$, is convenient when taking *duals*, since by the *Riesz-Fischer theorem*[63] we have explicit expressions for Hilbert space duals. We exploit this possibility below.

18.8 Distributions, generalized functions, again

We will see that distributions on $\mathbb{T}$ have Fourier expansions, greatly facilitating their study. [64]

The exponential functions ψ_n are in $C^\infty(\mathbb{T})$, so for any distribution u we can compute Fourier coefficients of u by

$$(n^{th} \text{ Fourier coefficient of } u) \;=\; \widehat{u}(n) \;=\; \frac{1}{2\pi}\cdot u(\psi_{-n})$$

Write

$$u \;\sim\; \sum_n \widehat{u}(n)\cdot\psi_n$$

even though *pointwise* convergence of the indicated sum is certainly not expected. Define Levi-Sobolev spaces for *all* $s \in \mathbb{R}$ by

$$H^s(\mathbb{T}) \;=\; \{u \in C^\infty(\mathbb{T})^* \;:\; \sum_n |u(\psi_{-n})|^2\cdot(1+n^2)^s < \infty\}$$

and the s^{th} *Levi-Sobolev norm* $|u|_{H^s}$ is

$$|u|^2_{H^s} \;=\; \sum_n |u(\psi_{-n})|^2\cdot(1+n^2)^s$$

For $0 \le s \in \mathbb{Z}$, this definition is visibly compatible with the previous definition via derivatives.

[18.8.1] Remark: The formation of the Levi-Sobolev spaces of both positive and negative indices portrays the classical *functions* of various degrees of (continuous) differentiability together with *distributions* of various orders as fitting together as comparable objects. By contrast, thinking only in terms of the spaces $C^k(\mathbb{T})$ does not immediately suggest a comparison with distributions.

For convenience, define a *weighted* version $\ell^{2,s}$ of (a two-sided version of) the classical Hilbert space ℓ^2 by

$$\ell^{2,s} \;=\; \{\{c_n : n\in\mathbb{Z}\} \;:\; \sum_{n\in\mathbb{Z}} |c_n|^2\cdot(1+n^2)^s < \infty\}$$

---

[63] The Riesz-Fischer theorem asserts that the (continuous) dual V^* of a Hilbert space V is $\mathbb{C}$-conjugate linearly isomorphic to V. The isomorphism from V to V^* attaches the linear functional $v \to \langle v, w\rangle$ to an element $w \in V$. Since our hermitian inner products $\langle,\rangle$ are *conjugate*-linear in the second argument, the map $w \to \langle\;, w\rangle$ is conjugate linear.

[64] In contrast, discussion of distributions on the real line $\mathbb{R}$ is more complicated, due to the non-compactness of $\mathbb{R}$. *Not* every distribution on $\mathbb{R}$ is the Fourier transform of a *function*. Distributions which *admit* Fourier transforms, *tempered* distributions, constitue a proper subset of all distributions on $\mathbb{R}$.

with the weighted version of the usual hermitian inner product, namely,

$$\langle \{c_n\},\ \{d_n\}\rangle \;=\; \sum_{n\in\mathbb{Z}} c_n\,\overline{d_n}\cdot(1+n^2)^s$$

[18.8.2] Claim: The complex bilinear *pairing*

$$\langle,\rangle \;:\; \ell^{2,s}\times\ell^{2,-s} \;\longrightarrow\; \mathbb{C}$$

by

$$\langle \{c_n\},\ \{d_n\}\rangle \;=\; \sum_n c_n\,d_{-n}$$

identifies these two Hilbert spaces as mutual duals, where

$$\ell^{2,-s} \;\longrightarrow\; (\ell^{2,s})^* \quad \text{by} \quad \{d_n\}\to\lambda_{\{d_n\}} \quad \text{where} \quad \lambda_{\{d_n\}}(\{c_n\}) = \sum_n c_n\,d_{-n}$$

[18.8.3] Remark: The minus sign in the subscript in the last formula is not the main point, but is a necessary artifact of our change from a *hermitian* form to a *complex bilinear* form. It is (thus) necessary to maintain compatibility with the Plancherel theorem for ordinary functions.

Proof: The Cauchy-Schwarz-Bunyakowsky inequality gives the continuity of the functional attached to $\{d_n\}$ in $\ell^{2,-s}$ by

$$\Big|\sum_n c_n\cdot d_{-n}\Big| \;\le\; \sum_n |c_n|\,(1+n^2)^{s/2}\cdot|d_{-n}|\,(1+n^2)^{-s/2}$$

$$\le\; \Big(\sum_n |c_n|^2\,(1+n^2)^s\Big)^{1/2}\cdot\Big(\sum_n |d_n|^2\,(1+n^2)^{-s}\Big)^{1/2} \;=\; |\{c_n\}|_{\ell^{2,s}}\cdot|\{d_n\}|_{\ell^{2,-s}}$$

proving the continuity. To prove the surjectivity we adapt the Riesz-Fischer theorem by a renormalization. That is, given a continuous linear functional λ on $\ell^{2,s}$, by Riesz-Fischer there is $\{a_n\}\in\ell^{2,s}$ such that

$$\lambda(\{c_n\}) \;=\; \langle\{c_n\},\,\{a_n\}\rangle_{\ell^{2,s}} \;=\; \sum_n c_n\cdot\overline{a}_n\cdot(1+n^2)^s$$

Take

$$d_n \;=\; \overline{a}_{-n}\cdot(1+n^2)^s$$

Check that this sequence of complex numbers is in $\ell^{2,-s}$, by direct computation, using the fact that $\{a_n\}\in\ell^{2,s}$,

$$\sum_n |d_n|^2\cdot(1+n^2)^{-s} \;=\; \sum_n |\overline{a}_{-n}\cdot(1+n^2)^s|^2\cdot(1+n^2)^{-s} \;=\; \sum_n |a_n|^2\cdot(1+n^2)^s < +\infty$$

Thus, $\ell^{2,-s}$ is (isomorphic to) the dual of $\ell^{2,s}$. ///

[18.8.4] Claim: The map $u\to\{\widehat{u}(n)\}$ on $H^s(\mathbb{T})$ by taking Fourier coefficients is a Hilbert-space isomorphism

$$H^s(\mathbb{T}) \;\approx\; \ell^{2,s}$$

Proof: That the two-sided sequence of Fourier coefficients $u(\psi_{-n})$ is in $\ell^{2,s}$ is part of the definition of $H^s(\mathbb{T})$. The more serious question is *surjectivity*.

Let $\{c_n\} \in \ell^{2,s}$. For $s \geq 0$, the s^{th} Levi-Sobolev norm dominates the 0^{th}, so distributions in $H^s(\mathbb{T})$ are at least $L^2(\mathbb{T})$-functions. The definition of $H^s(\mathbb{T})$ in this case makes $H^s(\mathbb{T})$ a Hilbert space, and we directly invoke the Plancherel theorem, using the orthonormal basis $\frac{\psi_n}{\sqrt{2\pi}} \cdot (1+n^2)^{-s/2}$ for $H^s(\mathbb{T})$. This gives the surjectivity $H^s(\mathbb{T}) \to \ell^{2,s}$ for $s \geq 0$.

For $s < 0$, to prove the surjectivity, for $\{c_n\}$ in $\ell^{2,s}$ we will define a distribution u lying in $H^s(\mathbb{T})$, by

$$u(f) \ = \ \sum_n \widehat{f}(n) \cdot c_{-n} \qquad\qquad (f \in C^\infty(\mathbb{T}))$$

By Cauchy-Schwarz-Bunyakowsky,

$$|\sum_n \widehat{f}(n) \cdot c_{-n}| \ \leq \ \sum_n |\widehat{f}(n)|\,(1+n^2)^{-s/2} \cdot |c_n|\,(1+n^2)^{s/2}$$

$$\leq \ \left(\sum_n |\widehat{f}(n)|^2\,(1+n^2)^{-s}\right)^{1/2} \cdot \left(\sum_n |c_n|^2\,(1+n^2)^{s}\right)^{1/2} \ = \ |f|_{H^{-s}} \cdot |\{c_n\}|_{\ell^{2,s}}$$

This shows that u is a continuous linear functional on $H^{-s}(\mathbb{T})$. For $s < 0$, the test functions $C^\infty(\mathbb{T})$ imbed continuously into $H^{-s}(\mathbb{T})$, so u gives a continuous functional on $C^\infty(\mathbb{T})$, so is a distribution. This proves that the Fourier coefficient map is a surjection to $\ell^{2,s}$ for $s < 0$. ///

[18.8.5] Remark: After this preparation, the remainder of this section is completely unsurprising. The following corollary is the conceptual point of this story.

[18.8.6] Corollary: For any $s \in \mathbb{R}$, the complex bilinear pairing

$$\langle , \rangle : H^s \times H^{-s} \to \mathbb{C} \quad \text{by} \quad f \times u \to \langle f, u\rangle = \sum_n \widehat{f}(n) \cdot \widehat{u}(-n)$$

gives an isomorphism

$$H^{-s} \ \approx \ (H^s)^*$$

by sending $u \in H^{-s}$ to $\lambda_u \in (H^s)^*$ defined by

$$\lambda_u(f) \ = \ \langle f, u\rangle \qquad\qquad (\text{for } f \in H^s(\mathbb{T}))$$

[18.8.7] Remark: The pairing of this last claim is *unsymmetrical*: the left argument is from H^s while the right argument is from H^{-s}.

Proof: This pairing via Fourier coefficients is simply the composition of the maps $H^s(\mathbb{T}) \approx \ell^{2,s}$ and $H^{-s}(\mathbb{T}) \approx \ell^{2,-s}$ with the pairing of $\ell^{2,s}$ and $\ell^{2,-s}$ given just above. ///

[18.8.8] Corollary: The space of all distributions on $\mathbb{T}$ is

$$\text{distributions} \ = C^\infty(\mathbb{T})^* \ = \ \bigcup_{s\geq 0} H^s(\mathbb{T})^* \ = \ \bigcup_{s \geq 0} H^{-s}(\mathbb{T}) \ = \ \text{colim}_{s\geq 0} H^{-s}(\mathbb{T})$$

thus expressing $C^\infty(\mathbb{T})^*$ as an ascending union of Hilbert spaces. ///

[18.8.9] Corollary: A distribution $u \sim \sum_n c_n\, \psi_n$ can be evaluated on $f \in C^\infty(\mathbb{T})$ by

$$u(f) \ = \ \sum_n \widehat{f}(n) \cdot \widehat{u}(-n)$$

Proof: Since u lies in some $H^{-s}(\mathbb{T})$, it gives a continuous functional on $H^s(\mathbb{T})$, which contains $C^\infty(\mathbb{T})$. The Plancherel-like evaluation formula above gives the equality. ///

A collection of Fourier coefficients $\{c_n\}$ is *of moderate growth* when there is a constant C and an exponent N such that

$$|c_n| \;\le\; C \cdot (1+n^2)^N \qquad\qquad (\text{for all } n \in \mathbb{Z})$$

[18.8.10] **Corollary:** Let $\{c_n\}$ be a collection of complex numbers of moderate growth. Then there is a distribution u with those as Fourier coefficients, that is, there is u with

$$u(\psi_{-n}) \;=\; c_n$$

Proof: For constant C and exponent N such that $|c_n| \le C \cdot (1+n^2)^N$,

$$\sum_n |c_n|^2 \cdot (1+n^2)^{-(2N+1)} \;\le\; \sum_n C^2 \cdot (1+n^2)^{2N} \cdot (1+n^2)^{-(2N+1)} \;=\; C^2 \cdot \sum_n (1+n^2)^{-1} \;<\; \infty$$

That is, from the previous discussion, the sequence gives an element of $H^{-(2N+1)}(\mathbb{T}) \subset C^\infty(\mathbb{T})^*$. ///

[18.8.11] **Corollary:** For $u \sim \sum_n c_n\,\psi_n \in H^s(\mathbb{T})$ the derivative (for any $s \in \mathbb{R}$) is

$$u' \;\sim\; \sum_n in \cdot c_n \cdot \psi_n \;\in\; H^{s-1}$$

Proof: Invoke the definition (compatible with integration by parts) of the derivative of distributions, and integrating by parts to see that $\widehat{f'}(n) = in \cdot \widehat{f}(n)$ for $f \in C^\infty(\mathbb{T}) = H^\infty(\mathbb{T})$,

$$u'(f) \;=\; -u(f') \;=\; -\sum_n \widehat{f'}(n) \cdot \widehat{u}(-n) \;=\; -\sum_n in\,\widehat{f}(n) \cdot \widehat{u}(-n) \;=\; \sum_n \widehat{f}(n) \cdot -in\,\widehat{u}(-n)$$

as claimed. The Fourier coefficients $-in \cdot \widehat{u}(n)$ do satisfy

$$\sum_n |in\,\widehat{u}(n)|^2 \cdot (1+n^2)^{s-1} \;\le\; \sum_n (1+n^2)\,|\widehat{u}(n)|^2 \cdot (1+n^2)^{s-1} \;=\; \sum_n |\widehat{u}(n)|^2 \cdot (1+n^2)^s \;=\; |u|^2_{H^s} < \infty$$

which proves that the differentiation maps H^s to H^{s-1} continuously. ///

[18.8.12] **Remark:** In the latter proof the sign in the subscript in the definition of the pairing $\ell^{2,s} \times \ell^{2,-s}$ was essential.

[18.8.13] **Corollary:** The collection of finite linear combinations of exponentials ψ_n is dense in every $H^s(\mathbb{T})$, for $s \in \mathbb{R}$. In particular, $C^\infty(\mathbb{T})$ is dense in every $H^s(\mathbb{T})$, for $s \in \mathbb{R}$.

Proof: The exponentials are an orthogonal basis for every Levi-Sobolev space. ///

[18.8.14] **Remark:** The topology of colimit of Hilbert spaces is the *finest* of several reasonable topologies on distributions. Density in a finer topology is a stronger assertion than density in a coarser topology.

18.9 The provocative example explained

The confusing example of the sawtooth function is clarified in the context we've developed. By now, we know that Fourier series whose coefficients satisfy sufficient decay conditions *are* differentiable. Even when the coefficients do not decay, but only grow *moderately*, the Fourier series is that of a *generalized function*. In other words, *we can (nearly) always differentiate* Fourier series term by term, as long as we can tolerate the outcome being a *generalized* function, rather than necessarily a *classical* function.

Again, $s(x)$ is the *sawtooth function*

$$s(x) \;=\; x - \pi \qquad\qquad (\text{for } 0 \le x < 2\pi)$$

made *periodic* by demanding $s(x + 2\pi n) = s(x)$ for all $n \in \mathbb{Z}$, so

$$s(x) \quad = \quad x - 2\pi \cdot [\![\frac{x}{2\pi}]\!] \;-\; \pi \qquad\qquad (\text{for } x \in \mathbb{R})$$

where $[\![x]\!]$ is the greatest integer less than or equal x. Away from $2\pi\mathbb{Z}$, this function is infinitely differentiable, with derivative 1. At integers it jumps down from value to π to value $-\pi$. We do not attempt to define a value *at* $2\pi\mathbb{Z}$.

We want to *differentiate* this function compatibly with integration by parts, and compatibly with term-by-term differentiation of Fourier series.

The sawtooth function *is* well-enough behaved to give a *distribution* by integrating against it. Therefore, as we saw above, it *can* be differentiated as a distribution, and be correctly differentiated as (as a distribution) by differentiating its Fourier expansion termwise.

A earlier, Fourier coefficients are computed by integrating against e^{-inx}

$$\frac{1}{2\pi}\int_0^1 s(x)\cdot e^{-inx}\,dx \;=\; \begin{cases} \dfrac{1}{-in} & (\text{for } n \neq 0) \\[2ex] 0 & (\text{for } n = 0) \end{cases}$$

Thus, *at least* as a distribution, its Fourier expansion is

$$s(x) \;=\; i\sum_{n\neq 0} \frac{1}{n}\cdot e^{inx}$$

The series *does* converge pointwise to $s(x)$ for x away from (images of) integers, as we proved happens at left and right differentiable points for piecewise C^o functions.

We are entitled to differentiate, at worst within the class of distributions, within which we are assured of a reasonable sense to our computations. *Further*, we are entitled (for any distribution) to differentiate the Fourier series term-by-term. That is, as distributions,

$$\begin{aligned} s'(x) \quad &= \quad -\textstyle\sum_{n\neq 0} e^{inx} \\ s''(x) \quad &= \quad -\textstyle\sum_{n\neq 0} in\, e^{inx} \\ &\ldots \\ s^{(k)}(x) \quad &= \quad -\textstyle\sum_{n\neq 0} (in)^{k-1}\, e^{inx} \end{aligned}$$

and so on, just as successive derivatives of smooth functions $f(x) = \sum_n c_n\, e^{inx}$ are obtained by termwise differentiation

$$f^{(k)}(x) \;=\; \sum_{n\neq 0} (in)^k\, c_n\, e^{inx}$$

The difficulty of interpreting the right-hand side of the Fourier series for $s^{(k)}$ as having pointwise values is irrelevant.

More to the point, these Fourier series are things to integrate smooth functions against, by an extension of the Plancherel formula for inner products of L^2 functions. Namely, for any smooth function $f(x) \sim \sum_n c_n\, e^{inx}$, the imagined integral of f against $s^{(k)}$ should be expressible as the sum of products of Fourier coefficients

$$\text{imagined } \langle f, s^{(k)} \rangle \;=\; \sum_{n\neq 0} c_n \cdot \left(\frac{(in)^k}{-in}\right)^{\text{conj}}$$

(where $\alpha \to \alpha^{\text{conj}}$ is complex conjugation) and the latter expression should behave well when rewritten in a form that refers to the literal function s. Indeed,

$$\sum_{n\neq 0} c_n \cdot \left(\frac{(in)^k}{-in}\right)^{\text{conj}} \;=\; (-1)^k \sum_{n\neq 0} (in)^k\, c_n \cdot \left(\frac{1}{-in}\right)^{\text{conj}} \;=\; (-1)^k \int_{\mathbb{T}} f^{(k)}(x)\, \overline{s}(x)\, dx$$

by the Plancherel theorem applied to the L^2 functions $f^{(k)}$ and s. Let u be the distribution given by integration against s. Then, by the definition of differentiation of distributions, we have computed that

$$(-1)^k \int_{\mathbb{T}} f^{(k)}(x)\, \overline{s}(x)\, dx \;=\; (-1)^k u(f^{(k)}) \;=\; u^{(k)}(f)$$

It is in this sense that the sum $\sum_{n\neq 0} c_n \cdot \frac{(in)^k}{-in}$ is integration of s against f.

Further, for f a smooth function with support away from the discontinuities of s, it is true that $u''(f) = 0$, giving s'' a vague pointwise sense of being 0 away from the discontinuities of s. This was clear at the outset, but now is given precise meaning.

Thus, as claimed at the outset of the discussion of functions on the circle, we can differentiate $s(x)$ legitimately, and the differentiation of the Fourier series of the sawtooth function $s(x)$ correctly represents this differentiation, viewing $s(x)$ and its derivatives as *distributions*.

18.10 Appendix: products and limits of topological vector spaces

Here we carry out the diagrammatical proof that products and limits of topological vector spaces *exist*, and are locally convex when the factors or limitands are locally convex. Nothing surprising happens.

[18.10.1] Claim: Products and limits of topological vector spaces exist. In particular, limits are *closed* (linear) subspaces of the corresponding products. When the factors or limitands are locally convex, so is the product or limit.

[18.10.2] Remark: Part of the point is that products and limits of locally convex topological vector spaces *in the larger category of not-necessarily locally convex topological vector spaces* are nevertheless locally convex. That is, enlarging the category in which we take test objects does not change the outcome, in this case. By contrast, coproducts and colimits in general are sensitive to local convexity of the test objects. [65]

Proof: After we construct products, limits are constructed as closed subspaces of them.

Let V_i be topological vector spaces. We claim that the topological-space product $V = \Pi_i V_i$ (with projections p_i) (with the product topology) is a topological vector space product. Let $\alpha_i : V_i \times V_i \to V_i$ be the addition on V_i. The family of composites $\alpha_i \circ (p_i \times p_i) : V \times V \to V_i$ induces a map $\alpha : V \times V \to V$ as in

$$\begin{tikzcd} V \times V \arrow[r, dashed, "\alpha"] \arrow[d, "p_i \times p_i"'] & V \arrow[d, "p_i"] \\ V_i \times V_i \arrow[r, "\alpha_i"'] & V_i \end{tikzcd}$$

---

[65] For example, uncountable coproducts do not exist among not-necessarily locally convex topological vector spaces, essentially because the not-locally-convex spaces ℓ^p with $0 < p < 1$ exist.

This defines what we will show to be a *vector addition* on V. Similarly, the scalar multiplications $s_i : \mathbb{C} \times V_i \to V_i$ composed with the projections $p_i : V \to V_i$ give a family of maps

$$s_i \circ (1 \times p_i) \;:\; \mathbb{C} \times V \;\longrightarrow\; V_i$$

which induce a map $s : \mathbb{C} \times V \to V$ which we will show to be a *scalar multiplication* on V. That these maps are *continuous* is given us by starting with the topological-space product.

That is, we must prove that vector addition is commutative and associative, that scalar multiplication is associative, and that the two have the usual distributivity. All these proofs are the same in form. For commutativity of vector addition, consider the diagram

$$\begin{tikzcd}
 & V_i \times V_i \arrow[rr, "v\times w\to v+w"] & & V_i \arrow[dd, equal] \\
V \times V \arrow[ur, "p_i\times p_i"] \arrow[rr, dashed, bend left, "v\times w\to v+w"] \arrow[rr, dashed, bend right, "v\times w\to w+v"] \arrow[d, "v\times w\to w\times v"'] & & V \arrow[ur, "p_i"'] \arrow[dr, "p_i"] & \\
V \times V \arrow[r, "p_i\times p_i"] & V_i \times V_i \arrow[rr, "v\times w\to w+v"] & & V_i
\end{tikzcd}$$

The upper half of the diagram is the induced-map definition of vector addition on V, and the lower half is the induced map definition of the reversed-order vector addition. The commutativity of addition on each V_i implies that going around the top of the diagram from $V \times V$ to V_i yields the same as going around the bottom. Thus, the two induced maps $V \times V \to V$ must be the same, since induced maps are *unique*.

The proofs of associativity of vector addition, associativity of scalar multiplication, and distributivity, use the same idea. Thus, *products* of topological vector spaces exist.

We should not forget to prove that the product is *Hausdorff*, since we implicitly require this of topological vector spaces. This is immediate, since a (topological space) product of Hausdorff spaces is readily shown to be Hausdorff.

Consider now the case that each V_i is locally convex. By definition of the product topology, every neighborhood of 0 in the product is of the form $\Pi_i U_i$ where U_i is a neighborhood of 0 in V_i, and all but finitely many of the U_i are the whole V_i. Since V_i is locally convex, we can shrink every U_i that is *not* V_i to be a convex open containing 0, while each *whole* V_i is convex. Thus, the product is locally convex when every factor is.

To construct limits, reduce to the product.

[18.10.3] Claim: Let V_i be topological vector spaces with transition maps $\varphi_i : V_i \to V_{i-1}$. The limit $V = \lim_i V_i$ *exists*, and, in particular, is the closed linear subspace (with subspace topology) of the product $\Pi_i V_i$ (with projections p_i) defined by the (closed) conditions

$$\lim_i V_i \;=\; \{v \in \Pi_i V_i : (\varphi_i \circ p_i)(v) = p_{i-1}(v), \;\; \text{for all } i\}$$

Proof: *(of claim)* Constructing the alleged limit as a closed subspace of the product immediately yields the desired properties of vector addition and scalar multiplication, as well as the Hausdorff-ness. What we must show is that the construction does function as a limit.

Given a compatible family of continuous linear maps $f_i : Z \to V_i$, there is induced a unique continuous linear map $F : Z \to \Pi_i V_i$ to the product, such that $p_i \circ f = f_i$ for all i. The *compatibility* requirement on the f_i exactly asserts that $f(Z)$ sits inside the subspace of $\Pi_i V_i$ defined by the conditions $(\varphi_i \circ p_i)(v) = p_{i-1}(v)$. Thus, f maps to this subspace, as desired.

Further, for all limitands locally convex, we have shown that the product is locally convex. The local convexity of a linear subspace (such as the limit) follows immediately. ///

18.11 Appendix: Fréchet spaces and limits of Banach spaces

A larger class of topological vector spaces arising in practice is the class of *Fréchet spaces*. In the present context, we can give a nice definition: a *Fréchet space* is a *countable* limit of Banach spaces. [66] Thus, for example,

$$C^\infty(\mathbb{T}) \;=\; \bigcap_k C^k(\mathbb{T}) \;=\; \lim_k C^k(\mathbb{T})$$

is a Fréchet space, *by (this) definition.*

Despite its advantages, the present definition is not the usual one. [67] We make a comparison, and elaborate on the features of Fréchet spaces.

A *metric* $d(,)$ on a vector space V is *invariant* (implicitly, under addition), when

$$d(x+z, y+z) \;=\; d(x,y) \qquad \text{(for all } x,y,z \in V)$$

All metrics we'll care about on topological vector spaces will be invariant in this sense.

[18.11.1] Claim: A Fréchet space is locally convex and complete (invariantly) metrizable. [68]

Proof: Let $V = \lim_i B_i$ be a countable limit of Banach spaces B_i, where $\varphi_i : B_i \to B_{i-1}$ are the transition maps and $p_i : V \to B_i$ are the projections. From the appendix, the limit is a closed linear subspace of the product, and the product is the cartesian product with the product topology and component-wise vector addition. Recall that a product of a *countable* collection of metric spaces is metrizable, and is complete if each factor is complete. A closed subspace of a complete metric space is complete metric. Thus, $\lim_i B_i$ is complete metric.

As proven in the previous appendix, *any* product or limit of locally convex spaces is locally convex, whether or not it has a countable cofinal family. Thus, the limit is Fréchet. ///

Addressing the comparison between local convexity and limits of Banach spaces,

[18.11.2] Theorem: Every locally convex topological vector space is a *subspace* of a limit of Banach spaces (and vice-versa).

---

[66] Of course, it suffices that a limit have a countable cofinal subfamily.

[67] A common definition, with superficial appeal, is that a Fréchet space is a complete (invariantly) metrized space that is locally convex. This has the usual disadvantage that there are many different metrics that can give the same topology. This also ignores the manner in which Fréchet spaces usually arise, as countable limits of Banach spaces. There is another common definition that does halfway acknowledge the latter construction, namely, that a Fréchet space is a *complete* topological vector space with topology given by a countable collection of *seminorms*. The latter definition is essentially equivalent to ours, but requires explanation of the suitable notion of *completeness* in a not-necessarily metric situation, as well as explanation of the notion of *seminorm* and how topologies are specified by seminorms. We skirt the latter issues for the moment.

[68] As is necessary to prove the *equivalence* of the various definitions of *Fréchet space*, the converse of this claim is true, namely, that every locally convex and complete (invariantly) metrizable topological vector space is a countable limit of Banach spaces. Proof of the converse requires work, namely, development of ideas about seminorms. Since we don't need this converse at the moment, we do not give the argument.

[18.11.3] Remark: This little theorem encapsulates the construction of *semi-norms* to give a locally convex topology. It can also be used to reduce the general Hahn-Banach theorem for locally convex spaces to the Hahn-Banach theorem for Banach spaces.

Proof: In one direction, we already know that a product or limit of Banach spaces is locally convex, since Banach spaces are locally convex.

In the Banach or normed-space situation, the topology comes from a metric $d(v, w) = |v - w|$ defined in terms of a *single* function $v \to |v|$ with the usual properties

$$|\alpha \cdot v| = |\alpha|_{\mathbb{C}} \cdot |v| \qquad \text{(homogeneity)}$$

$$|v + w| \le |v| + |w| \qquad \text{(triangle inequality)}$$

$$|v| \ge 0, \qquad \text{(equality } \textit{only} \text{ for } v = 0) \qquad \text{(definiteness)}$$

By contrast, for more general (but locally convex) situations, we consider a *family* Φ of functions $p(v)$ for which the definiteness condition is weakened slightly, so we require

$$p(\alpha \cdot v) = |\alpha|_{\mathbb{C}} \cdot p(v) \qquad \text{(homogeneity)}$$

$$p(v + w) \le p(v) + p(w) \qquad \text{(triangle inequality)}$$

$$p(v) \ge 0 \qquad \text{(semi-definiteness)}$$

Such a function $p()$ is a *semi-norm.* For Hausdorff-ness, we further require that the family Φ is *separating* in the sense that, given $v \neq 0$ in V, there is $p \in \Phi$ such that $p(v) > 0$.

A separating family Φ of semi-norms on a complex vector space V gives a *locally convex* topology by taking as local sub-basis [69] at 0 the sets

$$U_{p,\varepsilon} \;=\; \{v \in V : \; p(v) < \varepsilon\} \qquad \text{(for } \varepsilon > 0 \text{ and } p \in \Phi)$$

Each of these is convex, because of the triangle inequality for the semi-norms.

[18.11.4] Remark: The topology obtained from a (separating) family of seminorms may appear to be a random or frivolous generalization of the notion of topology obtained from a *norm.* However, it is the correct extension to encompass *all* locally convex topological vector spaces, as we see now. [70]

For a locally convex topological vector space V, for every open U in a local basis B at 0 of *convex* opens, try to define a *seminorm*

$$p_U(v) \;=\; \inf\{t > 0 : \; t \cdot U \ni v\}$$

We discover some necessary adjustments, and then verify the semi-norm properties.

First, we show that for any $v \in V$ the set over which the inf is taken is non-empty. Since scalar multiplication $\mathbb{C} \times V \to V$ is (jointly!) continuous, for given $v \in V$, given a neighborhood U of $0 \in V$, there are neighborhoods W of $0 \in \mathbb{C}$ and U' of v such that

$$\alpha \cdot w \in U \quad \text{(for all } \alpha \in W \text{ and } w \in U')$$

[69] Again, a *sub-basis* for a topology is a set of opens such that finite intersections form a *basis.* In other words, arbitrary unions of finite intersections give all opens.

[70] The semi-norms we construct here are sometimes called *Minkowski functionals*, even though they are not functionals in the sense of being continuous linear maps.

In particular, since W contains a disk $\{|\alpha| < \varepsilon\}$ for some $\varepsilon > 0$, we have $t \cdot v \in U$ for all $0 < t < \varepsilon$. That is,

$$v \in t \cdot U \qquad \text{(for all } t > \varepsilon^{-1}\text{)}$$

Semi-definiteness of p_U is built into the definition.

To avoid nagging problems, we should verify that, for convex U containing 0, when $v \in t \cdot U$ then $v \in s \cdot U$ for all $s \geq t$. This follows from the convexity, by

$$s^{-1} \cdot v = \frac{t}{s} \cdot (t^{-1} \cdot v) = \frac{t}{s} \cdot (t^{-1} \cdot v) + \frac{s-t}{s} \cdot 0 \in U$$

since $t^{-1} \cdot v$ and 0 are in U.

The homogeneity condition $p(\alpha\, v) = |\alpha|\, p(v)$ already presents a minor issue, since convex sets containing 0 need have no special properties regarding multiplication by complex numbers. That is, the problem is that, given $v \in t \cdot U$, while $\alpha\, v \in \alpha \cdot t \cdot U$, we do *not* know that this implies $\alpha v \in |\alpha| \cdot t \cdot U$. Indeed, in general, it will not. To repair this, to make semi-norms we must use only convex opens U which are *balanced* in the sense that

$$\alpha \cdot U = U \qquad \text{(for } \alpha \in \mathbb{C} \text{ with } |\alpha| = 1\text{)}$$

Then, given $v \in V$, we have $v \in t \cdot U$ if and only if $\alpha\, v \in t \cdot \alpha U$, and now

$$t\, \alpha U = t\, |\alpha|\, \frac{\alpha}{|\alpha|}\, U = t\, |\alpha|\, U$$

by the balanced-ness.

Now we have an obligation to show that there is a local basis (at 0) of convex *balanced* opens. Fortunately, this is easy to see, as follows. Given a convex U containing 0, from the continuity of scalar multiplication, since $0 \cdot v = 0$, there is $\varepsilon > 0$ and a neighborhood W of 0 such that $\alpha \cdot w \in U$ for $|\alpha| < \varepsilon$ and $w \in W$. Let

$$U' = \{\alpha \cdot w : |\alpha| \leq \frac{\varepsilon}{2},\ w \in W\} = \bigcup_{|\alpha| \leq \varepsilon/2} \alpha \cdot W$$

Being a union of the opens $\alpha \cdot W$, this U' is open. It is inside U by arrangement, and is *balanced* by construction. That is, there is indeed a local basis of convex *balanced* opens at 0.

For the *triangle inequality* for p_U, given $v, w \in V$, let t_1, t_2 be such that $v \in t \cdot U$ for $t \geq t_1$ and $w \in t \cdot U$ for $t \geq t_2$. Then, using the convexity,

$$v + w \in t_1 \cdot U + t_2 \cdot U = (t_1 + t_2) \cdot \left(\frac{t_1}{t_1 + t_2} \cdot U + \frac{t_2}{t_1 + t_2} \cdot U \right) \subset (t_1 + t_2) \cdot U$$

This gives the triangle inequality

$$p_U(v + w) \leq p_U(v) + p_U(w)$$

Finally, we check that the semi-norm topology is the original one. This is unsurprising. It suffices to check at 0. On one hand, given an open W containing 0 in V, there is a convex, balanced open U contained in W, and

$$\{v \in V : p_U(v) < 1\} \subset U \subset W$$

Thus, the semi-norm topology is at least as fine as the original topology. On the other hand, given convex balanced open U containing 0, and given $\varepsilon > 0$,

$$\{v \in V : p_U(v) < \varepsilon\} \supset \frac{\varepsilon}{2} \cdot U$$

Thus, each sub-basis open for the semi-norm topology contains an open in the original topology. We conclude that the two topologies are the same.

A summary so far: for a locally convex topological vector space, the semi-norms attached to convex balanced neighborhoods of 0 give a topology identical to the original, and *vice-versa*.

Before completing the proof of the theorem, recall that a *completion* of a set with respect to a *pseudo-metric* can be defined much as the completion with respect to a genuine metric. This is relevant because a semi-norm may only give a pseudo-metric, not a genuine metric.

Let Φ be a (separating) family of seminorms on a vector space V. For a *finite subset* i of Φ, let X_i be the *completion* of V with respect to the semi-norm

$$p_i(v) \;=\; \sum_{p\in i} p(v)$$

with natural map $f_i : V \to X_i$. Order subsets of Φ by $i \geq j$ when $i \supset j$. For $i > j$ we have

$$p_i(v) \;=\; \sum_{p\in i} p(v) \;\geq\; \sum_{p\in j} p(v) = p_j(v)$$

so we have natural continuous (transition) maps

$$\varphi_{ij} : X_i \;\longrightarrow\; X_j \qquad\qquad (\text{for } i > j)$$

We claim that each X_i is a *Banach space*, that V with its semi-norm topology has a natural continuous *inclusion* to the limit $X = \lim_i X_i$, and that V has the topology given by the subspace topology inherited from the limit.

The maps f_i form a compatible family of maps to the X_i, so there is a unique compatible map $f : V \to X$. By the separating property, given $v \neq 0$, there is $p \in \Phi$ such that $p(v) \neq 0$. Then for all i containing p, we have $f_i(v) \neq 0 \in X_i$. The subsets i containing p are *cofinal* in this limit, so $f(v) \neq 0$. Thus, f is an inclusion.

Since the limit is a (closed) subspace of the *product* of the X_i, it suffices to prove that the topology on V (imbedded in $\Pi_i X_i$ via f) is the subspace topology from $\Pi_i X_i$. Since the topology on V is at *least* this fine (since f is continuous), we need only show that the *subspace* topology is at least as fine as the semi-norm topology. To this end, consider a semi-norm-topology sub-basis set

$$\{v \in V : \, p_U(v) < \varepsilon\} \qquad\qquad (\text{for } \varepsilon > 0 \text{ and convex balanced open } U \text{ containing } 0)$$

This is simply the intersection of $f(V)$ with the sub-basis set

$$\Pi_{p\neq\{p_U\}} X_i \;\times\; \{v \in X_{\{p_U\}} : \, p_U(v) < \varepsilon\}$$

with the last factor inside $X_{\{p_U\}}$. Thus, by construction, the map $f : V \to X$ is a homeomorphism of V to its image. ///

19. Fourier transforms

1. Basic classes of functions $\mathcal{D}, \mathscr{S}, \mathcal{E}$ and their duals
2. Standard example computations
3. Riemann-Lebesgue lemma for $L^1(\mathbb{R})$
4. The Schwartz space $\mathscr{S} = \mathscr{S}(\mathbb{R}^n)$
5. Fourier inversion on $\mathscr{S}$
6. L^2-isometry of Fourier transform on $\mathscr{S}$
7. Isometric extension to Plancherel for $L^2(\mathbb{R}^n)$
8. Heisenberg uncertainty principle
9. Tempered distributions
10. Sobolev spaces, Sobolev imbedding

The Fourier transform of $f \in L^1(\mathbb{R})$ is [71]

$$\widehat{f}(\xi) \;=\; \int_{\mathbb{R}} e^{-2\pi i \xi x} \cdot f(x)\, dx$$

Since $f \in L^1(\mathbb{R})$, the integral converges absolutely, and uniformly in $\xi \in \mathbb{R}$. Similarly, on $\mathbb{R}^n$, with the usual inner product $\xi \cdot x = \sum_{j=1}^n \xi_j x_j$,

$$\widehat{f}(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot x} \cdot f(x)\, dx$$

An immediately interesting feature of Fourier transform is that *differentiation* is apparently converted to *multiplication*: at first heuristically, but rigorously proven below, imagining that we can integrate by parts,

$$\frac{\partial f}{\partial x_j}\widehat{\;}(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i \xi\cdot x} \cdot \frac{\partial}{\partial x_j} f(x)\, dx \;=\; \int_{\mathbb{R}^n} \frac{\partial}{\partial x_j} e^{-2\pi i \xi \cdot x} \cdot f(x)\, dx \;=\; \int_{\mathbb{R}^n} (-2\pi i \xi_j) e^{-2\pi i \xi \cdot x} \cdot f(x)\, dx$$

$$=\; (-2\pi i \xi_j) \int_{\mathbb{R}^n} (-2\pi i \xi_j) e^{-2\pi i \xi \cdot x} \cdot f(x)\, dx \;=\; (-2\pi i \xi_j) \widehat{f}(\xi)$$

Thus, the Laplacian $\Delta = \sum_j \frac{\partial^2}{\partial x_j^2}$ is converted to multiplication by $(-2\pi i)^2 \cdot r^2$ where $r^2 = \xi_1^2 + \ldots + \xi_n^2$. Thus, to solve a differential equation such as $(\Delta - \lambda)u = f$, apply Fourier transform to obtain $(-4\pi^2 r^2 - \lambda)\widehat{u} = \widehat{f}$. Divide through by $(-4\pi^2 r^2 - \lambda)$ to obtain

$$\widehat{u} \;=\; \frac{\widehat{f}}{-4\pi^2 r^2 - \lambda}$$

To recover u from $\widehat{u}$, there is *Fourier inversion* (proven below):

$$u(x) \;=\; \int_{\mathbb{R}^n} e^{2\pi i \xi \cdot x}\, \widehat{u}(\xi)\, d\xi$$

There are obvious issues about the integration by parts, the convergence of the relevant integrals, and the inversion formula. In fact, to extend the Fourier transform to $L^2(\mathbb{R}^n)$, the integral definition of the Fourier transform must also be extended to a situation where *the literal integral does not converge*. Similarly, a bit later, the Fourier transform on the *dual* of the Schwartz space $\mathscr{S}(\mathbb{R}^n)$ (below), the *tempered distributions* $\mathscr{S}(\mathbb{R}^n)^*$, is only defined by either an *extension by continuity* or by a *duality*.

19.1 Basic classes of functions $\mathcal{D}, \mathscr{S}, \mathcal{E}$ and their duals

[71] There are other choices of normalizations, that put the 2π in other locations than the exponent, but the differences are inconsequential, so we pick one normalization and use it consistently throughout.

Even though our immediate discussion will be incomplete, it is worthwhile to introduce some basic, standard function spaces. Recall that a function F on $\mathbb{R}^n$ is *of rapid decay* when $\sup_{x\in\mathbb{R}^n} |x|^N f(x)| < +\infty$ for all positive integers N.

$$\begin{cases} \mathcal{D} \;=\; \mathcal{D}(\mathbb{R}^n) \;=\; \text{test functions} \;=\; C_c^\infty(\mathbb{R}^n) \\ \\ \mathscr{S} \;=\; \mathscr{S}(\mathbb{R}^n) \;=\; \text{Schwartz functions} \;=\; \{f \in C^\infty(\mathbb{R}^n) : f \text{ and all its derivatives are of rapid decay}\} \\ \\ \mathcal{E} \;=\; \mathcal{E}(\mathbb{R}^n) \;=\; \text{smooth functions} \;=\; C^\infty(\mathbb{R}^n) \end{cases}$$

The spaces $\mathscr{S}$ and $\mathcal{E}$ will turn out to be Fréchet spaces, while the appropriate topology on test function $\mathcal{D}$ is somewhat more complicated. Without elaborating on these topologies, the *dual spaces*, that is, the vector spaces of continuous linear functionals $\mathcal{D} \to \mathbb{C}$, $\mathscr{S} \to \mathbb{C}$, and $\mathcal{E} \to \mathbb{C}$, are

$$\begin{cases} \mathcal{D}^* \;=\; \mathcal{D}' \;=\; \mathcal{D}(\mathbb{R}^n)^* \;=\; \text{distributions} \\ \\ \mathscr{S}^* \;=\; \mathscr{S}' \;=\; \mathscr{S}(\mathbb{R}^n)^* \;=\; \text{tempered distributions} \\ \\ \mathcal{E}^* \;=\; \mathcal{E}' \;=\; \mathcal{E}(\mathbb{R}^n)^* \;=\; \text{compactly-supported distributions} \end{cases}$$

For the latter name to make better sense, we'd need to describe the *support* of a distribution, and also prove that this naming convention is correct.

The obvious inclusions $\mathcal{D} \subset \mathscr{S} \subset \mathcal{E}$ do turn out to be *continuous* in the relevant topologies. Thus, we have inclusion-reversing containments of duals: $\mathcal{E}^* \subset \mathscr{S}^* \subset \mathcal{D}^*$.

Thus, *tempered* distributions really are a kind of distribution, and *compactly-supported* distributions are a kind of tempered distribution.

Eventually (below), we refine the chain of containments

$$\mathcal{D} \;\subset\; \mathscr{S} \;\subset\; L^2(\mathbb{R}^n) \;\subset\; \mathscr{S}^* \;\subset\; \mathcal{D}^*$$

in various ways. One such refinement is in terms of *Sobolev spaces*.

19.2 Example computations

It is useful and necessary to have a stock of explicitly evaluated Fourier transforms, especially on $\mathbb{R}$. In many cases, it is much less obvious how to go in the opposite direction, so *Fourier inversion* (below) has non-trivial content.

[19.2.1] Characteristic functions of finite intervals It is easy to compute the Fourier transform of the characteristic function $\mathrm{ch}_{[a,b]}$ of an interval $[a, b]$: at least for $\xi \neq 0$, but then extending by continuity (see the *Riemann-Lebesgue Lemma* below),

$$\int_{\mathbb{R}} \mathrm{ch}_{[a,b]}\, e^{-2\pi i\xi x}\, dx \;=\; \int_a^b e^{-2\pi i\xi x}\, dx \;=\; \frac{e^{-2\pi i\xi b} - e^{-2\pi i\xi a}}{-2\pi i\xi}$$

In particular, for a symmetrical interval $[-w, w]$,

$$\int_{\mathbb{R}} \mathrm{ch}_{[-w,w]}\, e^{-2\pi i\xi x}\, dx \;=\; \frac{e^{2\pi i\xi w} - e^{-2\pi i\xi w}}{2\pi i\xi} \;=\; \frac{\sin 2\pi w\xi}{\pi\xi} \;=\; 2w \cdot \frac{\sin 2\pi w\xi}{2\pi w\xi} \;=\; 2w \cdot \mathrm{sinc}(2\pi w\xi)$$

where the (naively-normalized) *sinc* function[72] is $\mathrm{sinc}(x) = \frac{\sin x}{x}$. Anticipating *Fourier inversion* (below), although $\mathrm{sinc}(x)$ is *not* in $L^1(\mathbb{R})$, it is in $L^2(\mathbb{R})$, and its Fourier transform is evidently a characteristic function of an interval. This is not obvious.

[19.2.2] Tent functions Let $f(x)$ be a piecewise-linear, continuous *tent function* of width $2w$ and height h, symmetrically placed about the origin:

$$f(x) \;=\; \begin{cases} 0 & (\text{for } x \le -w) \\ h - \dfrac{h|x|}{w} & (\text{for } |x| \le w) \\ 0 & (\text{for } x \ge w) \end{cases}$$

Breaking the integral into two pieces and integrating by parts twice, for $\xi \neq 0$ but extending by continuity (see below), we find that

$$\widehat{f}(\xi) \;=\; \frac{h}{\pi^2 w}\Big(\frac{\sin \pi w \xi}{\xi}\Big)^2$$

[19.2.3] Gaussians With our normalization of the Fourier transform, the best *Gaussian* is $f(x) = e^{-\pi x^2}$, because

$$\int_{\mathbb{R}} e^{-2\pi i \xi x}\, e^{-\pi x^2}\, dx \;=\; e^{-\pi i \xi^2}$$

The sanest proof of this uses *contour shifting* from complex analysis:

$$\int_{\mathbb{R}} e^{-2\pi i \xi x}\, e^{-\pi x^2}\, dx \;=\; \int_{\mathbb{R}} e^{-\pi(x-i\xi)^2 - \pi\xi^2}\, dx \;=\; e^{-\pi\xi^2}\int_{\mathbb{R}} e^{-\pi(x-i\xi)^2}\, dx \;=\; e^{-\pi\xi^2}\int_{-i\xi-\infty}^{-i\xi+\infty} e^{-\pi x}\, dx$$

$$=\; e^{-\pi\xi^2}\int_{-\infty}^{+\infty} e^{-\pi x}\, dx \;=\; e^{-\pi\xi^2}\cdot 1 \;=\; e^{-\pi\xi^2}$$

because $\int_{-\infty}^{+\infty} e^{-\pi x}\, dx = 1$. Similarly, in $\mathbb{R}^n$, because the Gaussian and the exponentials both factor over coordinates, the same identity holds:

$$\int_{\mathbb{R}^n} e^{-2\pi i \xi\cdot x}\, e^{-\pi|x|^2}\, dx \;=\; e^{-\pi|\xi|^2}$$

[19.2.4] Fourier transforms of rational expressions Often, one-dimensional Fourier transforms of relatively elementary expressions can be evaluated *by residues*, meaning via Cauchy's Residue Theorem from complex analysis. Thus, for example,

$$\int_{\mathbb{R}} e^{-2\pi i \xi x}\, \frac{1}{1+x^2}\, dx \;=\; 2\pi i \frac{e^{-2\pi\xi}}{i+i} \;=\; \pi\, e^{-2\pi\xi}$$

by looking at residues in the upper or lower complex half-plane, depending on the sign of ξ. Thinking of Fourier inversion, it is somewhat less obvious how to go in the other direction, to see that the Fourier transform of $e^{-|\xi|}$ is essentially $1/(1+x^2)$. Similarly, for $2 \le k \in \mathbb{Z}$,

$$\int_{\mathbb{R}} e^{-2\pi i \xi x}\, \frac{1}{(x-i)^k}\, dx \;=\; \begin{cases} (2\pi i)(-2\pi i \xi)^{k-1}\, e^{-2\pi|\xi|} & (\text{for } \xi < 0) \\ 0 & (\text{for } \xi > 0) \end{cases}$$

[72] According to http://en.wikipedia.org/wiki/Sinc_function, the name is a contraction of the Latin name *sinus cardinalis*, bestowed on this function by P. Woodard and I. Davies, *Information theory and inverse probability in telecommunication*, Proc. IEEE-part III: radio and communication engineering **99** (1952), 37-44.

[19.2.5] Translations are converted to multiplications For $f \in L^1(\mathbb{R}^n)$, for $x_o \in \mathbb{R}^n$, certainly $x \to f(x+x_o)$ is still in $L^1(\mathbb{R}^n)$, because Lebesgue measure is translation invariant. Changing variables, replacing x by $x - x_o$,

$$f(*+x_o)\widehat{\ }(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\, f(x+x_o)\, dx \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot(x-x_o)}\, f(x)\, dx$$

$$=\; e^{2\pi i\xi\cdot x_o}\int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\, f(x)\, dx \;=\; e^{2\pi i\xi\cdot x_o}\cdot \widehat{f}(\xi)$$

[19.2.6] Behavior under dilations A similar change of variables applies to dilations $x \to t\cdot x$ with $t > 0$: replacing x by x/t,

$$f(t\cdot *)\widehat{\ }(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\, f(t\cdot x)\, dx \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x/t}\, f(x)\, t^{-n}\, dx$$

$$=\; t^{-n}\int_{\mathbb{R}^n} e^{-2\pi i\frac{\xi}{t}\cdot x}\, f(x)\, dx \;=\; t^{-n}\widehat{f}(t^{-1}\cdot\xi)$$

[19.2.7] Behavior under linear transformations More generally, with an invertible real matrix A, replacing x by $A^{-1}x$,

$$f(A\cdot *)\widehat{\ }(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\, f(Ax)\, dx \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot A^{-1}x}\, f(x)\, (\det A)^{-1}\, dx$$

Since $\xi\cdot A^{-1}x = (A^{-1})^\top\xi\cdot x$, this is

$$(\det A)^{-1}\int_{\mathbb{R}^n} e^{-2\pi i(A^{-1})^\top\xi\cdot x}\, f(x)\, dx \;=\; (\det A)^{-1}\widehat{f}((A^{-1})^\top\xi)$$

[19.2.8] Multiplications are converted to differentiation, and vice-versa For *suitable* f, so that integration by parts succeeds,

$$\widehat{f'}(\xi) \;=\; \int_{\mathbb{R}} e^{-2\pi i\xi x}\,\frac{d}{dx}f(x)\, dx \;=\; -\int_{\mathbb{R}}\frac{d}{dx}e^{-2\pi i\xi x}\, f(x)\, dx$$

$$=\; -2\pi i\xi\int_{\mathbb{R}} e^{-2\pi i\xi x}\, f(x)\, dx \;=\; -2\pi i\xi\,\widehat{f}(\xi)$$

Anticipating Fourier inversion, we would know that, symmetrically, multiplication by x is essentially converted to differentiation. We can also compute this directly, but with a non-trivial issue about moving the differentiation through the integral: [73]

$$\widehat{(xf)}(\xi) \;=\; \int_{\mathbb{R}} e^{-2\pi i\xi x}\, xf(x)\, dx \;=\; \int_{\mathbb{R}}\frac{1}{-2\pi i}\frac{d}{d\xi}e^{-2\pi i\xi x}\, f(x)\, dx$$

$$=\; \frac{1}{-2\pi i}\frac{d}{d\xi}\int_{\mathbb{R}} e^{-2\pi i\xi x}\, f(x)\, dx \;=\; \frac{1}{-2\pi i}\frac{d}{d\xi}\widehat{f}(\xi)$$

[73] For f a *Schwartz function*, that is, smooth and it and all derivatives are of rapid decay (see below), moving the differentiation through the integral is demonstrably legitimate. However, the best proof, which shows that this is a special case of a very general pattern of operators commuting with integrals, is not elementary. It uses Gelfand-Pettis (also called *weak*) vector-valued integrals, which will be discussed later.

The issue of moving the differential operator through the integral also arises below in proving that Fourier transform maps the space $\mathscr{S}$ of Schwartz functions to itself.

19.3 Riemann-Lebesgue lemma for $L^1(\mathbb{R})$

Just to be sure that this result is not overlooked, we recall it:

[19.3.1] Theorem: *(Riemann-Lebesgue)* For $f \in L^1(\mathbb{R})$, the Fourier transform $\widehat{f}$ is in the space $C^o_o(\mathbb{R})$ of continuous functions going to 0 at infinity. In fact, the map $f \to \widehat{f}$ is a continuous linear map from the Banach space $L^1(\mathbb{R})$ to the Banach space $C^o_o(\mathbb{R})$, the latter being the sup-norm completion of $C^o_c(\mathbb{R})$.

Proof: First, for $f \in L^1(\mathbb{R})$,

$$|\widehat{f}(\xi)| \;=\; \Big| \int_{\mathbb{R}} e^{-2\pi i\xi x}\, f(x)\, dx \Big| \;\le\; \int_{\mathbb{R}} |e^{-2\pi i\xi x}| \cdot |f(x)|\, dx \;=\; \int_{\mathbb{R}} |f(x)|\, dx \;=\; |f|_{L^1}$$

Thus, for $|f-g|_{L^1} < \varepsilon$, for all $\xi \in \mathbb{R}$, $|\widehat{f}(\xi) - \widehat{g}(\xi)| < \varepsilon$. Thus, Fourier transform is a continuous map of $L^1(\mathbb{R})$ to the Banach space $C^o_{\mathrm{bdd}}(\mathbb{R})$ of bounded continuous functions with sup norm.

For f the characteristic function of a finite interval, the explicit computation above gives $|\widehat{f}(\xi)| \le 1/\pi|\xi|$ for large $|\xi|$, which certainly goes to 0 at infinity.

The theory of the Riemann integral shows that the space of finite linear combinations of characteristic functions of intervals is L^1-dense in the space $C^o_c(\mathbb{R})$ of compactly-supported continuous functions, which is L^1-dense in $L^1(\mathbb{R})$ itself, by Urysohn's lemma and the definition of integral. That is, every $f \in L^1(\mathbb{R})$ is an L^1-limit of finite linear combinations of characteristic functions of finite intervals. The continuity of the Fourier transform as a map $L^1(\mathbb{R}) \to C^o_{\mathrm{bdd}}(\mathbb{R})$ shows that $\widehat{f}$ is the sup-norm limit of Fourier transforms of finite linear combinations of characteristic functions of finite intervals, which are in $C^o_c(\mathbb{R})$. The sup-norm completion of the latter is $C^o_o(\mathbb{R})$, so $\widehat{f} \in C^o_c(\mathbb{R})$. ///

19.4 Schwartz space $\mathscr{S} = \mathscr{S}(\mathbb{R}^n)$

The Schwartz space on $\mathbb{R}^n$ consists of all $f \in C^\infty(\mathbb{R}^n)$ such that

$$\sup_{x\in\mathbb{R}^n} (|x|^2)^N \cdot |f^{(\alpha)}(x)| \;<\; \infty \qquad\qquad \text{(for all } N\text{, and for all multi-indices } \alpha\text{)}$$

where as usual, for a multi-index $\alpha = (\alpha_1, \ldots, \alpha_n)$ with non-negative integer components,

$$f^{(\alpha)} \;=\; \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \frac{\partial^{\alpha_2}}{\partial x_2^{\alpha_2}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f$$

Those supremums

$$\nu_{N,\alpha}(f) \;=\; \sup_{x\in\mathbb{R}^n} (|x|^2)^N \cdot |f^{(\alpha)}(x)|$$

required to be finite for Schwartz functions, are *semi-norms*, in the sense that they are non-negative real-valued functions with properties

$$\begin{cases} \nu_{N,\alpha}(f+g) & \le & \nu_{N,\alpha}(f) + \nu_{N,\alpha}(g) & \text{(triangle inequality)} \\ \nu_{N,\alpha}(c\cdot f) & = & |c| \cdot \nu_{N,\alpha}(f) & \text{(homogeneity)} \end{cases}$$

In the present context, in fact, these seminorms are genuine *norms*, insofar as no one of them is 0 except for the identically-0 function. This family of seminorms is *separating* in the reasonable sense that, if $\nu_{N,\alpha}(f-g) = 0$ for all N, α, then $f = g$.

The natural *topology* on $\mathscr{S}$ associated to this (separating) family of seminorms can be specified by giving a *sub-basis*[74] *at* $0 \in \mathscr{S}$: in a vector space V, we want a topology to be *translation-invariant* in the sense that vector addition $v \to v + v_o$ is a homeomorphism of V to itself. In particular, for every open neighborhood N of 0, $N + v_o$ is an open neighborhood of v_o, and vice-versa.

Here, take a sub-basis at 0 indexed by N, α, and $\varepsilon > 0$:

$$U_{N,\alpha,\varepsilon} \;=\; \{f \in \mathscr{S} : \nu_{N,\alpha}(f) < \varepsilon\}$$

[19.4.1] Theorem: With the latter topology, $\mathscr{S}$ is a complete metrizable space. *[... iou ...]*

[19.4.2] Remark: Since the topology of $\mathscr{S}$ is given by seminorms, the topology is also *locally convex*, meaning that every point has a basis of neighborhoods consisting of *convex* sets. This follows from the convexity of the sub-basis sets, and the fact that an intersection of convex sets is convex. Complete metrizable, locally convex topological vector spaces (with translation-invariant topology, as expected) are *Fréchet spaces.* This is a more general class including Banach spaces. In summary, $\mathscr{S}$ is a Fréchet space.

[19.4.3] Claim: For $f \in \mathscr{S}$,

$$\Big(\frac{\partial}{\partial x_j} f\Big)^\wedge(\xi) \;=\; (-2\pi i) \cdot \xi_j \cdot \widehat{f}(\xi)$$

Proof: We've already sketched the integration by parts argument for this, so now we should check in detail that $f \in \mathscr{S}$ is sufficient for that argument to succeed. For notational simplicity, we treat just the one-dimensional case:

$$\widehat{(f')}(\xi) \;=\; \int_{\mathbb{R}} e^{-2\pi i\xi x} \frac{\partial}{\partial x} f(x)\,dx \;=\; \lim_{N\to+\infty} \int_{|x|\le N} e^{-2\pi i\xi x} \frac{\partial}{\partial x} f(x)\,dx$$

Integrating by parts, the integral is

$$\Big[e^{-2\pi i\xi x}\, f(x)\Big]_{-N}^{N} \;-\; \int_{|x|\le N} \frac{\partial}{\partial x} e^{-2\pi i\xi x} \cdot f(x)\,dx$$

$$=\; e^{-2\pi i\xi N}\, f(N) - e^{2\pi i\xi N}\, f(-N) \;-\; \int_{|x|\le N} (-2\pi i\xi)\, e^{-2\pi i\xi x} \cdot f(x)\,dx$$

The boundary terms go to 0 as $N \to +\infty$, the factor of $-2\pi i\xi$ comes out of the integral, and the limit as $N \to +\infty$ of the integral over $|x| \le N$ becomes the integral over $\mathbb{R}$, as claimed. ///

The following claim, essentially the dual or opposite to the previous, sketched earlier, has a more difficult proof, a part of which we postpone.

[19.4.4] Claim: For $f \in \mathscr{S}(\mathbb{R}^n)$,

$$\frac{\partial}{\partial \xi_j} \widehat{f}(\xi) \;=\; (-2\pi i) \cdot (x_j \cdot f)^\wedge(\xi)$$

Proof: The point is that for Schwartz functions, the differentiation in ξ can pass inside the integral:

$$\frac{\partial}{\partial \xi_j} \widehat{f}(\xi) \;=\; \frac{\partial}{\partial \xi_j} \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\, f(x)\,dx \;=\; \int_{\mathbb{R}^n} \frac{\partial}{\partial \xi_j} e^{-2\pi i\xi\cdot x}\, f(x)\,dx$$

[74] Recall that a set S of sets $U \ni x_o$ is a *sub-basis* at x_o when every neighborhood of x contains a finite intersection of sets from S.

$$= \int_{\mathbb{R}^n} (-2\pi i x_j) e^{-2\pi i \xi \cdot x} f(x)\, dx \;=\; (-2\pi i) \int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot x} x_j f(x)\, dx \;=\; (-2\pi i) \cdot (x_j \cdot f)\widehat{\ }(\xi)$$

As remarked earlier, passing the differential operator inside the integral is best justified in a more sophisticated context, so we will not give any elementary-but-unenlightening argument here. ///

Let *translation* by x on $\mathscr{S}$ be written $T_x f$, where

$$(T_x f)(y) \;=\; f(y+x)$$

[19.4.5] Claim: For each $x \in \mathbb{R}^n$, translation by x is a continuous map $\mathscr{S} \to \mathscr{S}$.

Proof: [... iou ...] ///

19.5 Fourier inversion on $\mathscr{S}$

In our normalization, the inverse Fourier transform is

$$f^\vee(x) \;=\; \int_{\mathbb{R}^n} e^{2\pi i \xi \cdot x} f(\xi)\, d\xi$$

Of course, this is only slightly different from the *forward* Fourier transform, and sources sometimes do not invent a separate symbol for the inverse transform

[19.5.1] Theorem: (*Fourier inversion*) $(\widehat{f})^\vee = f$ for $f \in \mathscr{S}$.

Proof: [... iou ...] ///

[19.5.2] Corollary: Fourier transform is a topological vector space isomorphism $\mathscr{S} \to \mathscr{S}$. [... iou ...]

19.6 L^2-isometry of Fourier transform on $\mathscr{S}$

[19.6.1] Theorem: (*recast by Schwartz, c. 1950*) For $f, g \in \mathscr{S}$, $\langle f, g\rangle_{L^2} = \langle \widehat{f}, \widehat{g}\rangle_{L^2}$. In particular, $|\widehat{f}|_{L^2} = |f|_{L^2}$.

Proof: [... iou ...] ///

19.7 Isometric extension and Plancherel on $L^2(\mathbb{R}^n)$

[19.7.1] Theorem: (*Plancherel, 1910*) There is a unique continuous extension of Fourier transform to an isometry $L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$. That is, the Fourier transform $\mathscr{S} \to \mathscr{S}$ extends by continuity to a map $\mathscr{F} : L^2 \to L^2$, with isometry property

$$\langle \mathscr{F} f,\, \mathscr{F} g\rangle_{L^2} \;=\; \langle f, g\rangle_{L^2} \qquad\qquad (\text{for all } f, g \in L^2(\mathbb{R}^n))$$

Proof: The L^2 Plancherel theorem on $\mathscr{S}$, and the density of $\mathscr{S}$ in L^2, give the result. ///

[19.7.2] Remark: Even though the literal integral for the Fourier transform of typical $f \in L^2$ (but not in L^1) need not converge, it is standard to write the Fourier transform as an integral, with the understanding that it is not the *literal* integral, but is an extension-by-continuity of the literal integral, via Plancherel.

19.8 Heisenberg uncertainty principle

This is a theorem about Fourier transforms, once we grant a certain model of quantum mechanics. That is, there is a mathematical mechanism that yields an inequality, which has an interpretation in physics. [75]

For suitable f on $\mathbb{R}$,

$$|f|^2_{L^2} \;=\; \int_{\mathbb{R}} |f|^2 \;=\; -\int_{\mathbb{R}} x(f\cdot\overline{f})' \;=\; -2\,\mathrm{Re}\int_{\mathbb{R}} xf\overline{f}' \qquad \text{(integrating by parts)}$$

That is,

$$|f|^2_{L^2} \;=\; \big||f|^2_{L^2}\big| \;=\; \left|\int_{\mathbb{R}} |f|^2\right| \;=\; \left|-2\,\mathrm{Re}\int_{\mathbb{R}} xf\overline{f}'\right| \;\le\; 2\int_{\mathbb{R}} |xf\overline{f}'|$$

Next,

$$2\int_{\mathbb{R}} |xf\cdot\overline{f}'| \;\le\; 2\cdot|xf|_{L^2}\cdot|f'|_{L^2} \qquad \text{(Cauchy-Schwarz-Bunyakowsky)}$$

Since Fourier transform is an L^2-isometry, and since Fourier transform converts derivatives to multiplications,

$$|f'|_{L^2} \;=\; |\widehat{f'}|_{L^2} \;=\; 2\pi|\xi\widehat{f}|_{L^2}$$

Thus, we obtain the *Heisenberg inequality*

$$|f|^2_{L^2} \;\le\; 4\pi\cdot|xf|_{L^2}\cdot|\xi\widehat{f}|_{L^2}$$

More generally, a similar argument gives, for any $x_o\in\mathbb{R}$ and any $\xi_o\in\mathbb{R}$,

$$|f|^2_{L^2} \;\le\; 4\pi\cdot|(x-x_o)f|_{L^2}\cdot|(\xi-\xi_o)\widehat{f}|_{L^2}$$

Imagining that $f(x)$ is the probability that a particle's *position* is x, and $\widehat{f}(\xi)$ is the probability that its momentum is ξ, Heisenberg's inequality gives a lower bound on how *spread out* these two probability distributions must be. The physical assumption is that position and momentum *are* related by Fourier transform.

19.9 Tempered distributions

Tempered distributions can be first described as the space $\mathscr{S}^*$ of continuous linear function(al)s $\lambda:\mathscr{S}\to\mathbb{C}$.

[19.9.1] Claim: The Dirac δ, given by $\delta(\varphi)=\varphi(0)$ for $\varphi\in\mathscr{S}$, is a tempered distribution.

Proof: To prove continuity of $\varphi\to\varphi(0)$, it suffices to prove continuity at 0. The easy inequality

$$|f(0)| \;\le\; \sup_{x\in\mathbb{R}^n} |x|^0\cdot|f^{(0)}(x)| \;=\; \nu_{0,0}(f)$$

shows that $|f(0)|$ can be made as small as desired by making $\nu_{0,0}(f)$ sufficiently small, proving continuity. ///

The duality approach *does* allow an easy definition of Fourier transform $\widehat{u}$ of $u\in\mathscr{S}^*$, not by an integral, but by

$$\widehat{u}(\varphi) \;=\; u(\widehat{\varphi}) \qquad (\text{for } \varphi\in\mathscr{S})$$

[75] I think I first saw Heisenberg's Uncertainty Principle presented as a theorem about Fourier transforms in Folland's 1983 Tata Lectures on PDE.

Similarly for inverse Fourier transform, which we've shown truly is an inverse to the Fourier transform on $\mathscr{S}$. It remains to be shown that it is truly an inverse on $\mathscr{S}^*$. Prior to that, we have a basic example:

[19.9.2] Claim: $\widehat{\delta} = 1$. That is, the Fourier transform of the Dirac δ is integrate-against 1.

Proof: From the definition of Fourier transform on $\mathscr{S}^*$ via duality, for $\varphi \in \mathscr{S}$,

$$\widehat{\delta}(\varphi) \;=\; \delta(\widehat{\varphi}) \;=\; \widehat{\varphi}(0) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\, 0\cdot x}\, \varphi(x)\, dx \;=\; \int_{\mathbb{R}^n} 1 \cdot \varphi(x)\, dx \;=\; 1(\varphi)$$

by the literal integral definition of Fourier transform on $\mathscr{S}$. ///

We can give $\mathscr{S}^*$ the *weak dual topology*, also called the *weak *-topology*, by seminorms ν_φ attached to $\varphi \in \mathscr{S}$:

$$\nu_\varphi(u) \;=\; |u(\varphi)| \qquad\qquad (\text{for } u \in \mathscr{S}^* \text{ and } \varphi \in \mathscr{S})$$

This is *not* a topology given by a metric, but is obviously a type of topology that can be given to any dual space. This characterization of tempered distributions by *duality* does not explain their usefulness.

[19.9.3] Theorem: The definition of Fourier transform on $\mathscr{S}^*$ by duality does map $\mathscr{S}^*$ to itself, and is an isomorphism. Fourier inversion for the extended Fourier transform holds on $\mathscr{S}^*$.

Proof: From above, Fourier transform is a continuous linear map of $\mathscr{S}$ to itself. Thus, $\varphi \to \widehat{\varphi} \to u(\widehat{\varphi})$ is a continuous linear functional on $\mathscr{S}$, for any $u \in \mathscr{S}^*$. To prove continuity of $u \to \widehat{u}$ in the weak dual topology, take $\varphi \in \mathscr{S}$, with associated semi-norm ν_φ as above, and compute

$$\nu_\varphi(\widehat{u}) \;=\; |\widehat{u}(\varphi)| \;=\; |u(\widehat{\varphi})| \;=\; \nu_{\widehat{\varphi}}(u)$$

Thus, making $\nu_{\widehat{\varphi}}(u)$ small makes $\nu_\varphi(\widehat{u})$ small, proving continuity of $u \to \widehat{u}$ in the weak dual topology.

To prove Fourier inversion on $\mathscr{S}^*$, let $\mathscr{F}$ be the extended Fourier transform, and $\mathscr{F}'$ the extension of the inverse transform, *not* denoted $\mathscr{F}^{-1}$, to avoid inadvertently begging the question. Then for $\varphi \in \mathscr{S}$,

$$(\mathscr{F}'(\mathscr{F}u))(\varphi) \;=\; (\mathscr{F}u)(\mathscr{F}'\varphi) \;=\; u(\mathscr{F}(\mathscr{F}'\varphi)) \;=\; u(\varphi)$$

by Fourier inversion on $\mathscr{S}$. Since both the transform and its inverse are continuous, both are isomorphisms. ///

There is also a characterization of $\mathscr{S}^*$ as an *extension* of $\mathscr{S}$. First, there is a inclusion $\mathscr{S} \to \mathscr{S}^*$ by taking $\varphi \in \mathscr{S}$ to the *integrate-against* functional u_φ:

$$u_\varphi(f) \;=\; \int_{\mathbb{R}^n} \varphi \cdot f \;=\; \int_{\mathbb{R}^n} \varphi(x) \cdot f(x)\, dx \qquad\qquad (\text{for } f \in \mathscr{S})$$

For most topological vector spaces V, there is no natural inclusion $V \to V^*$, so such inclusions for *spaces of functions* V distinguishes them from the general abstract scenario.

[19.9.4] Claim: The inclusion $\mathscr{S} \to \mathscr{S}^*$ is continuous, and has dense image.

Proof: [... iou ...] ///

That is, we can think of $\mathscr{S}^*$ as a sort of *completion* or *extension* of $\mathscr{S}$, in the weak dual topology on $\mathscr{S}$ itself. From this viewpoint, $\mathscr{S}^*$ consists of *generalized functions.* Thus, the definition of Fourier transform on $\mathscr{S}^*$ should be compatible with that defined by the literal integral on $\mathscr{S}$:

[19.9.5] Claim: The Fourier transform on $\mathscr{S}^*$ defined via duality agrees with the integral definition on $\mathscr{S} \subset \mathscr{S}^*$. That is, with u_φ the integrate-against functional attached to $\varphi \in \mathscr{S}$,

$$\widehat{u_\varphi} \;=\; u_{\widehat{\varphi}}$$

Proof: This compatibility is an easy preliminary form of Plancherel: for $f \in \mathscr{S}$,

$$\widehat{u_\varphi}(f) \;=\; u_\varphi(\widehat{f}) \;=\; \int_{\mathbb{R}^n} \varphi \cdot \widehat{f} \;=\; \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{-2\pi i \xi \cdot x}\varphi(x)\, f(\xi)\, dx\, d\xi \;=\; \int_{\mathbb{R}^n} \widehat{\varphi} \cdot f \;=\; u_{\widehat{\varphi}}(f)$$

by Fubini-Tonelli. ///

We can compute Fourier transforms of tempered distributions by taking weak dual topology limits of Schwartz functions and the literal integral form of the Fourier transform:

[19.9.6] Claim: For a sequence of Schwartz functions φ_i approaching a tempered distribution u in the weak dual topology,

$$(\mathscr{S}^*\text{-}) \lim_i \widehat{\varphi}_i \;=\; \widehat{u}$$

Proof: [... iou ...] ///

We define *derivatives* of tempered distributions in a fashion compatible with the integrate-against inclusion $\mathscr{S} \to \mathscr{S}^*$, specifically, to be compatible with *integration by parts*. That is, for $\varphi, f \in \mathscr{S}$, and integration-by-parts distribution u_φ, in one variable,

$$u_{\varphi'}(f) \;=\; \int_{\mathbb{R}^n} \varphi' \cdot f \;=\; -\int_{\mathbb{R}^n} \varphi \cdot f' \;=\; -u_\varphi(f')$$

Note the sign. Thus, on $\mathbb{R}^n$, for $u \in \mathscr{S}^*$, define u' by

$$\frac{\partial}{\partial x_i} u(f) \;=\; -u\Big(\frac{\partial}{\partial x_i} f\Big) \qquad\qquad (\text{for } f \in \mathscr{S})$$

Similarly, *multiplication* by polynomials can be defined by duality, also in a fashion compatible with $\mathscr{S} \subset \mathscr{S}^*$:

$$(x_i \cdot u)(f) \;=\; u(x_i f) \qquad\qquad (\text{for } f \in \mathscr{S})$$

[19.9.7] Corollary: Differentiation and multiplication by polynomials are continuous maps $\mathscr{S}^* \to \mathscr{S}^*$, with the weak dual topology.

Proof: Again, continuity of a linear map is equivalent to continuity at 0. Given $\varphi \in \mathscr{S}$ and $u \in \mathscr{S}^*$,

$$\nu_\varphi\Big(\frac{\partial}{\partial x_i} u\Big) \;=\; \Big|\frac{\partial}{\partial x_i} u(\varphi)\Big| \;=\; \Big| - u\Big(\frac{\partial}{\partial x_i}\varphi\Big)\Big| \;=\; \nu_{\frac{\partial}{\partial x_i}\varphi}(u)$$

as desired. Similarly,

$$\nu_\varphi(x_i u) \;=\; |(x_i u)(\varphi)| \;=\; |u(x_i \varphi)| \;=\; \nu_{x_i \varphi}(u)$$

///

[19.9.8] Corollary: As for Schwartz functions, Fourier transform *intertwines* differentiation and multiplication on $\mathscr{S}^*$.

Proof: For notational simplicity, let's do this just on $\mathbb{R}$. For $u \in \mathscr{S}^*$, the Fourier transform of the derivative is described, for $\varphi \in \mathscr{S}$, as

$$\widehat{u'}(\varphi) \;=\; u'(\widehat{\varphi}) \;=\; -u(\frac{d}{dx}\widehat{\varphi}) \;=\; -u\Big(-2\pi i\widehat{\xi\varphi}\Big) \;=\; 2\pi i \cdot \widehat{u}(\xi\varphi) \;=\; 2\pi i\,\xi \cdot \widehat{u}(\varphi)$$

That is, $\widehat{u'} = 2\pi i\xi \cdot \widehat{u}$. The same sort of computation proves the reverse. ///

[19.9.9] Remark: Also, this intertwining property can be proven by extending by continuity from $\mathscr{S} \subset \mathscr{S}^*$.

[19.9.10] Polynomials and derivatives of δ From $\widehat{\delta} = 1$ and the intertwining of differentiation and multiplication by powers of $x_1, \ldots, x_n$, for a multi-index α,

$$\widehat{\delta^{(\alpha)}} \;=\; (2\pi i)^{|\alpha|}\, x^\alpha \cdot \widehat{\delta}(x) \;=\; (2\pi i)^{|\alpha|}\, x^\alpha \cdot 1 \;=\; (2\pi i)^{|\alpha|}\, x^\alpha$$

where $x^\alpha = x_1^{\alpha_1} \ldots x_n^{\alpha_n}$, and $|\alpha| = \alpha_1 + \ldots + \alpha_n$. By Fourier inversion,

$$\widehat{x^\alpha} \;=\; \frac{1}{(2\pi i)^{|\alpha|}} \cdot \delta^{(\alpha)}$$

As with differentiation, multiplication by polynomials, and Fourier transform, *translation* of $u \in \mathscr{S}^*$ is defined either by duality or by extension-by-continuity from $\mathscr{S} \subset \mathscr{S}^*$. Just as the possibly unexpected -1 in the derivative, to be compatible with integration by parts, we should see how translation behaves for integrate-against distributions. Let the translate $T_x f$ of $f \in \mathscr{S}$ be defined by $T_x f(y) = f(y+x)$. For $\varphi, f \in \mathscr{S}$,

$$u_{T_x\varphi}(f) \;=\; \int_{\mathbb{R}_n} T_x\varphi \cdot f \;=\; \int_{\mathbb{R}_n} T_x\varphi(y) \cdot f(y)\; dy \;=\; \int_{\mathbb{R}_n} \varphi(y+x) \cdot f(y)\; dy$$
$$=\; \int_{\mathbb{R}_n} \varphi(y) \cdot f(y-x)\; dy \;=\; u_\varphi(T_{-x} f)$$

by replacing y by $y - x$. Thus, again, a sign should enter in the definition of translation of a tempered distribution u:

$$(T_x u)(f) \;=\; u(T_{-x} f) \qquad\qquad (\text{for } f \in \mathscr{S})$$

19.10 Sobolev spaces, Sobolev imbedding

[... iou ...]

20. Distributions supported at 0

[20.0.1] Theorem: A distribution u with *support* $\{0\}$ is a (finite) linear combination of Dirac's δ and its derivatives.

Recall the notion of *support* of a distribution: the *support* of a distribution u is the *complement* of the *union* of all open sets $U \in \mathbb{R}^n$ such that

$$u(f) = 0 \qquad \text{(for } f \in \Delta_K \text{ with compact } K \subset U)$$

Proof: The space Δ of test functions on $\mathbb{R}^n$ is $\Delta = \bigcup_K \Delta_K$, where Δ_K is test functions supported on compact K. The latter is a Fréchet space, with *norms*

$$\nu_{k,K}(f) = \sup_{i \le k,\ x \in K} |f^{(i)}(x)|$$

Thus, it suffices to classify u in Δ_K^* with support $\{0\}$.

We have seen that a continuous linear map T from a *limit* of Banach spaces (such as Δ_K) to $\mathbb{C}$ factors through a limitand. Thus, there is an *order* $k \ge 0$ such that u factors through

$$C_K^k = \{f \in C^k(K) :\ f^{(\alpha)} \text{ vanishes on } \partial K \text{ for all } \alpha \text{ with } |\alpha| \le k\}$$

We need an auxiliary gadget. Fix a smooth compactly-supported function ψ identically 1 on a neighborhood of 0, bounded between 0 and 1, and (necessarily) identically 0 outside some (larger) neighborhood of 0. For $\varepsilon > 0$ let

$$\psi_\varepsilon(x) = \psi(\varepsilon^{-1}x)$$

Since the support of u is just $\{0\}$, for all $\varepsilon > 0$ and for all $f \in \mathcal{D}(\mathbb{R}^n)$ the support of $f - \psi_\varepsilon \cdot f$ does not include 0, so

$$u(\psi_\varepsilon \cdot f) = u(f)$$

Thus, for some constant C (depending on k and K, but not on f)

$$|\psi_\varepsilon f|_k = \sup_{x \in K} \sup_{|\alpha| \le k} |(\psi_\varepsilon f)^{(\alpha)}(x)| \le C \cdot \sup_{|i| \le k} \sup_{x} \sup_{0 \le j \le i} \varepsilon^{-|j|} \left| \psi^{(j)}(\varepsilon^{-1}x)\, f^{(i-j)}(x) \right|$$

For f vanishing to order k at 0, that is, $f^{(\alpha)}(0) = 0$ for all multi-indices α with $|\alpha| \le k$, on a fixed neighborhood of 0, by a Taylor-Maclaurin expansion, for some constant C

$$|f(x)| \le C \cdot |x|^{k+1}$$

and, generally, for α^{th} derivatives with $|\alpha| \le k$,

$$|f^{(\alpha)}(x)| \le C \cdot |x|^{k+1-|\alpha|}$$

For some constant C

$$|\psi_\varepsilon f|_k \le C \cdot \sup_{|i| \le k} \sup_{0 \le j \le i} \varepsilon^{-|j|} \cdot \varepsilon^{k+1-|i|+|j|} \le C \cdot \varepsilon^{k+1-|i|} \le C \cdot \varepsilon^{k+1-k} = C \cdot \varepsilon$$

Thus, for all $\varepsilon > 0$, for smooth f vanishing to order k at 0,

$$|u(f)| = |u(\psi_\varepsilon f)| \le C \cdot \varepsilon$$

Thus, $u(f) = 0$ for such f.

That is, u is 0 on the intersection of the kernels of δ and its derivatives $\delta^{(\alpha)}$ for $|\alpha| \le k$. Generally,

[20.0.2] Proposition: A continuous linear function $\lambda \in V^*$ vanishing on the intersection of the kernels of a finite collection $\lambda_1, \ldots, \lambda_n$ of continuous linear functionals on V is a linear combination of the λ_i.

Proof: The linear map

$$q : V \longrightarrow \mathbb{C}^n \quad \text{by} \quad v \longrightarrow (\lambda_1 v, \ldots, \lambda_n v)$$

is *continuous* since each λ_i is continuous, and λ factors through q, as $\lambda = L \circ q$ for some linear functional L on $\mathbb{C}^n$. We know all the linear functionals on $\mathbb{C}^n$, namely, L is of the form

$$L(z_1, \ldots, z_n) \;=\; c_1 z_1 + \ldots + c_n z_n \qquad \text{(for some constants } c_i\text{)}$$

Thus,

$$\lambda(v) \;=\; (L \circ q)(v) \;=\; L(\lambda_1 v, \ldots, \lambda_n v) \;=\; c_1 \lambda_1(v) + \ldots + c_n \lambda_n(v)$$

expressing λ as a linear combination of the λ_i. ///

[20.0.3] Remark: The following lemma resolves a potential confusion.

[20.0.4] Lemma: For compact K inside the *complement* of the support of a distribution u,

$$u(f) \;=\; 0 \qquad \text{(for } f \in \Delta_K\text{)}$$

Proof: This is plausible, but not utterly trivial. Let $\{U_i : i \in I\}$ be open sets such that for compact K' inside any single U_i and $f \in \Delta_{K'}$ we have $u(f) = 0$. Let $\{\psi_i : i \in I\}$ be a smooth locally finite *partition of unity*[76] subordinate to $\{U_i : i \in I\}$. Take $f \in \Delta_{K'}$ for K' compact inside $U = \bigcup_i U_i$. Then

$$f = f \cdot 1 = \sum_i f \cdot \psi_i$$

and the sum is *finite*. Then

$$u(f) = u(\sum_i f \cdot \psi_i) = \sum_i u(f \cdot \psi_i) = \sum_i 0 = 0$$

(The fact that the sum is finite allows interchange of summation and evaluation.) ///

[76] That is, the functions ψ_i are smooth, take values between 0 and 1, sum to 1 at all points, and on any compact there are only finitely-many which are non-zero. The existence of such partitions of unity is not completely trivial to prove.

21. Examples discussion

[21.36] Give a *persuasive* proof that the function

$$f(x) \;=\; \begin{cases} 0 & (\text{for } x \le 0) \\ e^{-1/x} & (\text{for } x > 0) \end{cases}$$

is infinitely differentiable at 0. Use this kind of construction to make a *smooth step function*: 0 for $x \le 0$ and 1 for $x \ge 1$, and goes monotonically from 0 to 1 in the interval $[0,1]$. Use this to construct a *family of smooth cut-off functions* $\{f_n : n = 1,2,3,\ldots\}$: for each n, $f_n(x) = 1$ for $x \in [-n,n]$, $f_n(x) = 0$ for $x \notin [-(n+1), n+1]$, and f_n goes monotonically from 0 to 1 in $[-(n+1), -n]$ and monotonically from 1 to 0 in $[n, n+1]$.

Discussion: In $x > 0$, by induction, the derivatives are finite linear linear combinations of functions of the form $x^{-n}e^{-1/x}$. It suffices to show that $\lim_{x\to 0^+} x^{-n}e^{-1/x} = 0$. Equivalently, that $\lim_{x\to+\infty} x^n e^{-x} = 0$, which follows from $e^{-x} = 1/e^x$, and

$$x^{-n}e^{-1/x} \;=\; \frac{x^n}{e^x} \;=\; \frac{x^n}{\sum_{m\ge 0}\frac{x^m}{m!}} \;\le\; \frac{x^n}{\frac{x^{n+1}}{(n+1)!}} \;\longrightarrow\; 0 \qquad\qquad (\text{as } x \to +\infty)$$

(This is perhaps a little better than appeals to L'Hospital's Rule.) Thus, f is smooth at 0, with all derivatives 0 there. ///

Next, we make a *smooth bump function* by

$$b(x) \;=\; \begin{cases} 0 & (\text{for } x \le -1) \\ e^{\frac{1}{1-x^2}} & (\text{for } -1 < x < 1) \\ 0 & (\text{for } x \ge 1) \end{cases}$$

A similar argument to the previous shows that this is smooth. Renormalize it to have integral 1 by

$$\beta(x) \;=\; \frac{b(x)}{\int_{-1}^{1} b(t)\;dt}$$

Then $\int_{-1}^{x} \beta(t)\;dt$ is a smooth (monotone) step function that goes from 0 at -1 to 1 at 1. The minor modification $s(x) = 2\int_{-1}^{x} \beta(2t-1)\;dt$ gives a smooth (monotone) step function going from 0 at 0 to 1 at 1. ///

Then $s(x+n+1)$ is a smooth, monotone step function going up from 0 to 1 in $[-n-1, -n]$, and $s(n+1-x)$ for $n \in \mathbb{Z}$ is a smooth, monotone step function going *down* from 1 to 0 in $[n, n+1]$. Thus, the product $f_n(x) = s(x+n+1)\cdot s(n+1-x)$ is the desired smooth cut-off function. ///

[21.37] With $g(x) = f(x + x_o)$, express $\widehat{g}$ in terms of $\widehat{f}$, first for $f \in \mathscr{S}(\mathbb{R}^n)$, then for $f \in \mathscr{S}(\mathbb{R}^n)^*$.

Discussion: For $f \in \mathscr{S}(\mathbb{R}^n)$, the literal integral computes the Fourier transform:

$$\widehat{g}(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\; g(x)\; dxn \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\; f(x+x_o)\; dx$$

Replacing x by $x - x_o$ in the integral gives

$$\widehat{g}(\xi) \;=\; \int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot(x-x_o)}\; f(x)\; dx \;=\; e^{2\pi i\xi\cdot x_o}\int_{\mathbb{R}^n} e^{-2\pi i\xi\cdot x}\; f(x)\; dx \;=\; e^{2\pi i\xi\cdot x_o}\cdot \widehat{f}(\xi)$$

The precise corresponding statement for tempered distributions cannot refer to pointwise values. Write ψ_{x_o} for the function $\xi \to e^{2\pi i \xi \cdot x_o}$. Since ψ_{x_o} is bounded, for a tempered distribution u, $\psi_{x_o} \cdot u$ is the tempered distribution described by

$$(\psi_{x_o} \cdot u)(\varphi) \;=\; u(\psi_{x_o}\,\varphi) \qquad\qquad (\text{for } \varphi \in \mathscr{S})$$

This is compatible with multiplication of (integrate-against-) functions $\mathscr{S} \subset \mathscr{S}^*$. Also, let translation $u \to T_{x_o} u$ be defined by $(T_{x_o} u)(\varphi) = u(T_{-x_o}\varphi)$, again compatibly with integration against Schwartz functions. In these terms, the above argument shows that

$$(T_{x_o} f)\widehat{\;} \;=\; \psi_{x_o} \cdot \widehat{f} \qquad\qquad (\text{for } f \in \mathscr{S})$$

This formulation avoids reference to pointwise values, and thus could make sense for tempered distributions.

One argument is *extension by continuity*: Fourier transform is a continuous map $\mathscr{S}^* \to \mathscr{S}^*$, as is translation $u \to T_{x_o} u$, so the identity extends by continuity to all tempered distributions. ///

Another argument is by *duality*: first,

$$(T_{x_o} u)\widehat{\;}(\varphi) \;=\; (T_{x_o} u)(\widehat{\varphi}) \;=\; u(T_{-x_o}\,\widehat{\varphi}) \;=\; u\Big((\psi_{x_o} \cdot \varphi)\widehat{\;}\Big)$$

by applying the identity to $\varphi, \widehat{\varphi} \in \mathscr{S}$. Going back, this is

$$\widehat{u}(\psi_{x_o} \cdot \varphi) \;=\; (\psi_{x_o} \cdot \widehat{u})(\varphi) \qquad\qquad (\text{for all } \varphi \in \mathscr{S})$$

Altogether, $(T_{x_o} u)\widehat{\;} = \psi_{x_o} \cdot \widehat{u}$.

[21.38] Let V be a vector space, with norms $|\cdot|_1$ and $|\cdot|_2$. Suppose that $|v|_2 \ge |v|_1$ for all $v \in V$. Show that the identity map $i : V \to V$ is continuous, where the source is given the $|\cdot|_2$ topology and the target is given the $|\cdot|_1$ topology. Show that if a sequence $\{v_n\}$ in V is $|\cdot|_2$ Cauchy, then it is $|\cdot|_1$-Cauchy. Let V_j be the completion of V with respect to the metric $|v - v'|_j$. Show that we can *extend i by continuity* to a continuous linear map $I : V_2 \to V_1$, that is, by

$$I(V_2\text{-limit of } V_2\text{-Cauchy sequence } \{v_n\}) \;=\; V_1\text{-limit of } \{v_n\}$$

Discussion: First, it suffices to show that the identity map $i : V \to V$ with indicated topologies is *bounded*, and, indeed,

$$|j(v)|_1 \;=\; |v|_1 \;\le\; |v|_2 \qquad\qquad (\text{for all } v \in V, \text{ by hypothesis})$$

For $\{v_n\}$ Cauchy in the $|\cdot|_2$ topology, given $\varepsilon > 0$, take n_o such that $|v_m - v_n|_2 < \varepsilon$ for all $m, n \ge n_o$. Then the same inequality holds (with the same n_o and ε) for $|\cdot|_2$, so $\{v_n\}$ is Cauchy in the $|\cdot|_1$ topology.

A useful characterization of the completion $\widetilde{X}$ of a metric space X is that there is an isometry $j : X \to \widetilde{X}$, and any non-expanding[77] map $f : X \to Y$ to a complete metric space Y extends uniquely to continuous map $\widetilde{f} : \widetilde{X} \to Y$, with $\widetilde{f} \circ j = f$. In particular,

$$\widetilde{f}(X - \lim_n x_n) \;=\; Y - \lim_n f(x_n)$$

This is well-defined because f is continuous on X. Thus, with $X = V$, $Y = V^2$, and $f : V \to V_2$ given by inclusion, we have the assertion. ///

[21.39] Solve $-u'' + u = \delta$ on $\mathbb{R}$. (*Hint:* use Fourier transform, and grant that $\widehat{\delta} = 1$.)

[77] This sense of *non-expanding* is the reasonable one: $d_Y(f(x), f(x')) \le d_X(x, x')$ for all $x, x' \in X$.

Discussion: Let's assume that we are asking for a solution u that is at worst a tempered distribution. Thus, we can take Fourier transform, obtaining

$$(4\pi^2\xi^2 + 1)\widehat{u} \;=\; \widehat{\delta} \;=\; 1$$

Obviously we want to *divide* by $4\pi^2\xi^2 + 1$. Unlike some other examples, where division was not quite legitimate, here, we can achieve the effect by *multiplication* by the smooth, bounded function $1/(4\pi^2\xi^2 + 1)$, since $4\pi^2\xi^2 + 1$ does not vanish on $\mathbb{R}$. Thus,

$$\widehat{u} \;=\; \frac{1}{4\pi^2\xi^2 + 1}$$

Since the right-hand side is luckily in $L^1(\mathbb{R})$, we can compute its image under Fourier inversion by the literal integral, its inverse Fourier transform will be a continuous function (by Riemann-Lebesgue), so has meaningful pointwise values:

$$u(x) \;=\; \int_{\mathbb{R}} \frac{e^{2\pi i\xi x}}{4\pi^2\xi^2 + 1}\,d\xi$$

The integral can be evaluated by *residues*: depending on the sign of x, we use an auxiliary arc in the upper (for $x > 0$) or lower (for $x < 0$) half-plane, so that $\xi \to e^{2\pi i\xi x}$ is *bounded* in the corresponding half-plane. Thus, we pick up either $2\pi i$ times the residue at $\xi = 1/2\pi i$, or the negative (because the orientation is negative) of the residue at $\xi = -1/2\pi i$. That is, respectively,

$$2\pi i \cdot \frac{e^{2\pi i\cdot(1/2\pi i)\cdot x}}{4\pi^2 \cdot (\frac{1}{2\pi i} - \frac{-1}{2\pi i})} \;=\; -e^{-x} \;=\; -e^{-|x|} \qquad \text{(for } x \geq 0\text{)}$$

and

$$-2\pi i \cdot \frac{e^{2\pi i\cdot(-1/2\pi i)\cdot x}}{4\pi^2 \cdot (\frac{-1}{2\pi i} - \frac{1}{2\pi i})} \;=\; -e^{x} \;=\; -e^{-|x|} \qquad \text{(for } x \leq 0\text{)}$$

[21.40] Show that $u'' = \delta_{\mathbb{Z}}$ has no solution on the circle $\mathbb{T}$. (*Hint:* Use Fourier series, granting the Fourier expansion of $\delta_{\mathbb{Z}}$.) Show that $u'' = \delta_{\mathbb{Z}} - 1$ *does* have a solution.

Discussion: In Fourier series converging in $H^{-\frac{1}{2}-\varepsilon}(\mathbb{T})$ for all $\varepsilon > 0$, $\delta_{\mathbb{Z}} = \sum_{n\in\mathbb{Z}} 1\cdot\psi_n$, where $\psi_n(x) = e^{2\pi inx}$. A function u in the relatively large-yet-tractable space $H^{-\infty}(\mathbb{T})$ has a Fourier expansion $u = \sum_n \widehat{u}(n)\cdot\psi_n$. Application of the (extended-sense) second derivative operator can be done termwise (by design), and annihilates the $n = 0$ term. That is, no u'' can have 0^{th} Fourier coefficient 1, as does $\delta_{\mathbb{Z}}$, so that equation is not solvable. ///

In contrast, $\delta_{\mathbb{Z}}-1$ has exactly lost that difficult Fourier component, and, in terms of Fourier series, $u'' = \delta_{\mathbb{Z}}-1$ is

$$\sum_{n\in\mathbb{Z}} (2\pi in)^2) \cdot \widehat{u}(n) \cdot \psi_n \;=\; \sum_{n\neq 0} 1 \cdot \psi_n$$

has the solution *by division*

$$u \;=\; \sum_{n\neq 0} \frac{1}{(2\pi in)^2}\psi_n$$

[21.41] On the circle $\mathbb{T}$, show that $u'' = f$ has a unique solution for all $f \in L^2(\mathbb{T})$ orthogonal to the constant function 1. (And reflect on the Fredholm alternative?)

Discussion: The orthogonality to 1 means that the 0^{th} Fourier coefficient of f is 0. Thus, on the Fourier series side, for any $u \in H^{-\infty}(\mathbb{T})$, $u'' = f$ is

$$\sum_{n\in\mathbb{Z}} (2\pi in)^2 \cdot \widehat{u}(n) \cdot \psi_n \;=\; \sum_{n\neq 0} \widehat{f}(n) \cdot \psi_n$$

gives

$$u \;=\; \sum_{n\neq 0} \frac{\widehat{f}(n)}{(2\pi i n)^2}\cdot\psi_n$$

and there is no other solution in $H^{-\infty}(\mathbb{T})$. ///

[21.42] The sawtooth function is first defined on $[0,1)$ by $\sigma(x) = x - \frac{1}{2}$, and then extended to $\mathbb{R}$ by periodicity so that $\sigma(x+n) = \sigma(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. After recalling its Fourier expansion, describe the derivatives σ' and σ'' of σ.

Discussion: The 0^{th} Fourier coefficient is 0. For $n \neq 0$, integrating by parts once, the n^{th} Fourier is $-1/2\pi i n$. That is, at least converging in L^2,

$$\sigma(x) \;=\; \sum_{n\neq 0} \frac{1}{-2\pi i n}\, e^{2\pi i n x}$$

In fact, from the Fourier-coefficient criterion for Sobolev spaces, $\sigma \in H^{\frac{1}{2}-\varepsilon}$ for all $\varepsilon > 0$. Differentiating termwise (in an extended sense),

$$\sigma' \;=\; -\sum_{n\neq 0} e^{2\pi i n x} \qquad\qquad (\text{convergent in } H^{-\frac{1}{2}-\varepsilon} \text{ for all } \varepsilon > 0)$$

We might recognize this as being closely related to the Dirac comb

$$\delta_{\mathbb{Z}} \;=\; \sum_{n\in\mathbb{Z}} e^{2\pi i n x} \qquad\qquad (\text{convergent in } H^{-\frac{1}{2}-\varepsilon})$$

Specifically, $\sigma' = 1 - \delta_{\mathbb{Z}}$. Also, looking at the description of σ directly, its derivative is (locally) 1 away from $\mathbb{Z}$, and has a $-\delta_n$ for all $n \in \mathbb{Z}$. That is, yet again,

$$\sigma' \;=\; 1 - \sum_{n\in\mathbb{Z}} \delta_n \;=\; 1 - \delta_{\mathbb{Z}}$$

Similarly, differentiating term-wise once more,

$$\sigma'' \;=\; -\sum_{n\neq 0} 2\pi i n\cdot e^{2\pi i n x} \qquad\qquad (\text{convergent in } H^{-\frac{3}{2}-\varepsilon} \text{ for all } \varepsilon > 0)$$

$$=\; -\sum_{n\in\mathbb{Z}} \delta'_n \;=\; -\delta'_{\mathbb{Z}}$$

///

[21.43] Show that $e^{-\varepsilon\pi x^2} \to 1$ as $\varepsilon \to 0^+$ in the $\mathscr{S}^*$ topology. Compute the Fourier transforms of the functions $e^{-\varepsilon\pi x^2}$, and show that they go to δ in the $\mathscr{S}^*$ topology. Obtain, again, as a corollary, the fact that $\widehat{1} = \delta$ (extended Fourier transform).

Discussion: We must show that, for each $\varphi \in \mathscr{S}$,

$$\lim_{\varepsilon\to 0^+} \int_{\mathbb{R}} (e^{-\varepsilon\pi x^2} - 1)\,\varphi(x)\,dx \;=\; 0$$

Since we are accustomed to other uses of ε, let's rewrite this as

$$\lim_{\eta\to 0^+} \int_{\mathbb{R}} (e^{-\eta\pi x^2} - 1)\,\varphi(x)\,dx \;=\; 0$$

Given $\varepsilon > 0$, for given φ, let N be sufficiently large so that $|\varphi(x)| < \varepsilon/|x|^2$ for $|x| \geq N$. Then

$$\Big|\int_{\mathbb{R}} (e^{-\eta\pi x^2} - 1)\,\varphi(x)\,dx\Big| \;\leq\; \int_{|x|\leq N} |e^{-\eta\pi x^2} - 1| \cdot |\varphi(x)|\,dx + \int_{|x|\geq N} |e^{-\eta\pi x^2} - 1| \cdot |\varphi(x)|\,dx$$

Estimate the second integral:

$$\int_{|x|\geq N} |e^{-\eta\pi x^2} - 1| \cdot |\varphi(x)|\,dx \;\leq\; \int_{|x|\geq N} 2 \cdot \frac{\varepsilon}{|x|^2}\,dx \;\leq\; \frac{4}{N} \cdot \varepsilon$$

For the first integral, given N and ε, for sufficiently small $\eta > 0$, we have $|e^{-\eta\pi x^2} - 1| < \varepsilon$ for all $|x| \leq N$. Thus, $e^{-\eta\pi x^2} \to 1$. ///

Next, the usual trick computes the Fourier transform

$$\widehat{e^{-\eta\pi x^2}}(\xi) \;=\; \frac{1}{\sqrt{\eta}} \cdot e^{-\frac{1}{\eta}\pi\xi^2}$$

We want to show that these go to δ. The continuity of (extended) Fourier transform $\mathscr{S}^* \longrightarrow \mathscr{S}^*$ assures that

$$(\mathscr{S}^*\text{–}\lim_{\eta} e^{-\eta\pi x^2})\widehat{\;} \;=\; \mathscr{S}^*\text{–}\lim_{\eta} \widehat{e^{-\eta\pi x^2}} \;=\; \mathscr{S}^*\text{–}\lim_{\eta} \frac{1}{\sqrt{\eta}} \widehat{e^{-\frac{1}{\eta}\pi x^2}}$$

For each $\varphi \in \mathscr{S}$,

$$\int_{\mathbb{R}} \varphi(x) \cdot \frac{1}{\sqrt{\eta}} \cdot e^{-\frac{1}{\eta}\pi x^2}\,dx \;=\; \int_{\mathbb{R}} \varphi(\sqrt{\eta} \cdot x) \cdot e^{-\pi x^2}\,dx$$

by replacing x by $\sqrt{\eta} \cdot x$. Given $\varepsilon > 0$, let N be large enough so that $|e^{-\pi x^2}| < \varepsilon$ for $|x| \geq N$. Let $\delta > 0$ be small enough so that $|\varphi(x) - \varphi(0)| < \varepsilon$ for $|x| < \delta$. Take $\eta > 0$ sufficiently small so that $N \cdot \sqrt{\eta} < \delta$. Using $\int_{\mathbb{R}} e^{-\pi x^2}\,dx = 1$,

$$\Big|\varphi(0) - \int_{\mathbb{R}} \varphi(\sqrt{\eta} \cdot x) \cdot e^{-\pi x^2}\,dx\Big| \;=\; \Big|\int_{\mathbb{R}} (\varphi(0) - \varphi(\sqrt{\eta} \cdot x)) \cdot e^{-\pi x^2}\,dx\Big|$$

$$\leq\; \int_{|x|\leq N} |\varphi(0) - \varphi(\sqrt{\eta} \cdot x)| \cdot e^{-\pi x^2}\,dx + \int_{|x|\geq N} |\varphi(0) - \varphi(\sqrt{\eta} \cdot x)| \cdot e^{-\pi x^2}\,dx$$

$$<\; \int_{|x|\leq N} \varepsilon \cdot e^{-\pi x^2}\,dx + \int_{|x|\geq N} 2\sup|\varphi| \cdot e^{-\pi x^2}\,dx \;<\; \varepsilon + 2\sup|\varphi| \cdot \varepsilon$$

That is, $\frac{1}{\sqrt{\eta}} \cdot e^{-\frac{1}{\eta}\pi x^2} \longrightarrow \delta$ in the $\mathscr{S}^*$ topology. ///

[21.44] Compute $\widehat{\cos x}$.

Discussion: Start from $\widehat{\delta} = 1$. Using the previous example's identity,

$$(T_{x_o}\delta)\widehat{\;} \;=\; \psi_{x_o} \cdot 1 \;=\; \psi_{x_o}$$

By Fourier inversion, $\widehat{\psi_{x_o}} = T_{-x_o}\delta$. Thus,

$$\widehat{\cos x} \;=\; \tfrac{1}{2}(\psi_{1/2\pi} + \psi_{-1/2\pi})\widehat{\;} \;=\; \tfrac{1}{2}(T_{-1/2\pi}\delta + T_{1/2\pi}\delta)$$

Written in terms of mock-pointwise-values, this is $\widehat{\cos}(\xi) = \dfrac{\delta(\xi - \frac{1}{2\pi}) + \delta(\xi + \frac{1}{2\pi})}{2}$. ///

[21.45] Smooth functions $f \in \mathcal{E}$ act on distributions $u \in \mathcal{D}(\mathbb{R})^*$ by a dualized form of pointwise multiplication: $(f \cdot u)(\varphi) = u(f\varphi)$ for $\varphi \in \mathcal{D}(\mathbb{R})$. Show that if $x \cdot u = 0$, then u is *supported at* 0, in the sense that for $\varphi \in \mathcal{D}$ with $\operatorname{spt}\varphi \not\ni 0$, necessarily $u(\varphi) = 0$. Thus, by the theorem classifying such distributions, u is a linear combination of δ and its derivatives. Show that in fact $x \cdot u = 0$ implies that u is a multiple of δ itself.

Discussion: For $\varphi \in \mathcal{D}$ whose support does *not* include 0, the function $1/x$ is defined and smooth on $\operatorname{spt}\varphi$. Thus, $x \to \varphi(x)/x$ is in $\mathcal{D}$. For such φ,

$$u(\varphi) \;=\; u(x \cdot \frac{\varphi}{x}) \;=\; 0$$

Thus, $\operatorname{spt} u = \{0\}$, so by the theorem is a finite linear combination $u = \sum_{i=0}^n c_i\, \delta^{(i)}$ with scalars c_i. To see that in fact only δ itself can appear, we use the idea that $1, x, \frac{x^2}{2!}, \frac{x^3}{3!}, \ldots, \frac{x^n}{n!}$ are essentially a *dual basis* to $\delta, \delta', \delta'', \ldots, \delta^{(n)}$. One way to make this completely precise is to use a smooth cut-off function $\eta \in \mathcal{D}$ around 0, namely, identically 1 on a neighborhood of 0. Then $\eta \cdot x^i \in \mathcal{D}$, and

$$\delta^{(i)}(\eta \cdot \frac{x^j}{j!}) \;=\; \begin{cases} 1 & (\text{for } i = j) \\ \\ 0 & (\text{for } i \neq j) \end{cases}$$

In particular, this shows that the derivatives of δ are *linearly independent*. For $0 \le j \in \mathbb{Z}$,

$$0 \;=\; (x \cdot u)(x^j) \;=\; (x \cdot \sum_i c_i\, \delta^{(i)})(x^j) \;=\; \sum_i c_i\, \delta^{(i)}(x \cdot x^j) \;=\; \sum_i c_i\, \delta^{(i)}(x^{j+1}) \;=\; (j+1)! \cdot c_{j+1}$$

Thus, $c_j = 0$ for $j \ge 1$, and u is a multiple of δ itself. ///

22. Examples discussion

[22.46] Given f in the Schwartz space $\mathscr{S}$, show that there is $F \in \mathscr{S}$ with $F' = f$ if and only if $\int_{\mathbb{R}} f = 0$.

Discussion: On one hand, if $f = F'$ for $F \in \mathscr{S}$, then $\int_{-\infty}^x f(y)\,dy = F(x)$. Since $\lim_{x\to+\infty} F(x) = 0$, $\int_{\mathbb{R}} f = 0$.

On the other hand, if $\int_{\mathbb{R}} f = 0$, let $F(x) = \int_{-\infty}^x f$, and show that $F \in \mathscr{S}$. Since $F' = f$ by the fundamental theorem of calculus, the (higher) derivatives of F are those of f, so all that needs to be shown is that F itself is of rapid decay. For $x \to -\infty$,

$$|F(x)| \;\le\; \int_{-\infty}^x |f| \;\le\; \int_{-\infty}^x |1+y^2|^{-N} \cdot \sup_{t\in\mathbb{R}} |(1+t^2)^N \cdot f(t)|\,dy \;\le\; \sup_{t\in\mathbb{R}} |(1+t^2)^N \cdot f(t)| \cdot \int_{-\infty}^x |1+y^2|^{-N}\,dy$$

and the latter integral is finite. Using the condition $\int_{\mathbb{R}} f = 0$,

$$F(x) \;=\; \int_{-\infty}^x f \;=\; \int_{\mathbb{R}} f - \int_x^\infty f \;=\; 0 - \int_x^\infty f$$

so for $x \to +\infty$ it suffices to similarly estimate

$$\Big|\int_x^\infty f\Big| \;\le\; \int_x^\infty (1+y^2)^{-N} \cdot \sup_{t\in\mathbb{R}} |(1+t^2)^N \cdot f(t)|\,dy \;\le\; \sup_{t\in\mathbb{R}} |(1+t^2)^N \cdot f(t)| \cdot \int_x^\infty (1+y^2)^{-N}\,dy$$

giving the rapid decay. ///

[22.47] Let $u(x) = e^x \cdot \sin(e^x)$. Explain in what sense the integral $\displaystyle\int_{\mathbb{R}} f(x)\,u(x)\,dx$ converges for every $f \in \mathscr{S}$.

Discussion: The idea is to integrate by parts, noting that $u = v'$ with $v(x) = \cos(e^x)$. We must be careful with the boundary terms:

$$\int_{\mathbb{R}} f(x)\,u(x)\,dx \;=\; \int_{\mathbb{R}} f(x)\,v'(x)\,dx \;=\; \lim_{M,N\to+\infty} \int_{-M}^N f(x)\,v'(x)\,dx$$

$$=\; \lim_{M,N\to+\infty} \Big(\Big[f(x)\,v(x)\Big]_{-M}^N - \int_{-M}^N f'(x)\,v(x)\,dx\Big)$$

Since $v(x)$ is bounded and f' is of rapid decay, the limit *exists*, so the original integral is convergent. Further, the value is correctly determined by integration by parts, namely

$$-\int_{-\infty}^\infty f'(x)\,v(x)\,dx \;=\; -\int_{-\infty}^\infty f'(x)\,\cos(e^x)\,dx$$

That is, for $f \in \mathscr{S}$ and functions such as u obtained by differentiating bounded smooth functions, integration by parts is completely justifiable via the natural estimates. ///

[22.48] Show that $\sin(nx) \to 0$ in the $\mathscr{S}^*$-topology as $n \to +\infty$. (Since $\mathscr{S}$ is strictly larger than $\mathcal{D}$, this implies that $\sin(nx) \to 0$ in the $\mathcal{D}^*$-topology.)

Discussion: We must show that, for each $\varphi \in \mathscr{S}$,

$$\lim_n \int_{\mathbb{R}} \sin(nx)\,\varphi(x)\,dx \;=\; 0$$

On one hand, since Schwartz functions are L^1, we could invoke Riemann-Lebesgue, since (up to normalizations) the indicated integral is $(\widehat{\varphi}(n) - \widehat{\varphi}(-n))/2i$.

On another hand, we also know that $\widehat{\varphi}$ is again a Schwartz function, so $(\widehat{\varphi}(n) - \widehat{\varphi}(-n))/2i \to 0$. (Further, if we know that $\mathscr{S}$ is dense in L^1, then this gives a slightly different proof of Riemann-Lebesgue.) ///

[22.49] Let $-\infty < a < b < c < +\infty$, and

$$f(x) \;=\; \begin{cases} 0 & (\text{for } x < a) \\ A & (\text{for } a < x < b) \\ B & (\text{for } b < x < c) \\ 0 & (\text{for } c < x) \end{cases}$$

Show that (extended) $\frac{d}{dx} f = A\delta_a + (B - A)\delta_b - B\delta_c$.

Discussion: This example asks for *proof* of the plausible intuitive idea that a piecewise constant function has derivative 0 along the intervals where it is constant, and multiples of Dirac deltas where jumps occur. There are at least two approaches to the proof, depending whether one characterizes distributions as elements of a dual space, or as $\mathcal{D}^*$-limits of test functions. Granting the theorem that these two characterizations are equivalent, the operational question is which allows an easier approach to the present question.

Perhaps the characterization by duality is more convenient here. Thus, $f' \in \mathcal{D}^*$ is a linear functional on $\mathcal{D}$ characterized by the extension of integration by parts:

$$(\text{as functional})\; f'(\varphi) \;=\; -f(\varphi') \;=\; -\int_{\mathbb{R}} f(x)\,\varphi'(x)\,dx \qquad (\text{for all } \varphi \in \mathcal{D})$$

Yes, the notation is slightly inconsistent, since on the left f' is a functional on $\mathcal{D}$, in the middle f is a functional on $\mathcal{D}$, while in the integral on the right f is a pointwise-valued function. From the definition of the pointwise-valued function f, integrating by parts or invoking the fundamental theorem of calculus, this is

$$-A\cdot\int_a^b \varphi'(x)\,dx - B\cdot\int_b^c \varphi'(x)\,dx \;=\; -A\cdot(\varphi(b)-\varphi(a)) - B(\varphi(c)-\varphi(b)) \;=\; (A\cdot\delta_a + (B-A)\cdot\delta_b - B\cdot\delta_c)(\varphi)$$

as claimed. ///

[22.50] Show that the principal value functional $u(\varphi) = P.V. \int_{\mathbb{R}} \frac{\varphi(x)}{x}\,dx$ satisfies $x\cdot u = 1$.

Discussion: For $\varphi \in \mathcal{D}$,

$$u(\varphi) \;=\; \lim_{\varepsilon\to 0^+} \int_{|x|\ge\varepsilon} \frac{x\cdot\varphi(x)}{x}\,dx \;=\; \lim_{\varepsilon\to 0^+} \int_{|x|\ge\varepsilon} \varphi(x)\,dx \;=\; \int_{\mathbb{R}} \varphi(x)\,dx \;=\; \int_{\mathbb{R}} 1\cdot\varphi(x)\,dx \;=\; 1(\varphi)$$

since φ is continuous at 0. Thus, $x\cdot u = 1$. ///

[22.51] Compute the Fourier transform of the sign function

$$\mathrm{sgn}(x) \;=\; \begin{cases} 1 & (\text{for } x > 0) \\ -1 & (\text{for } x < 0) \end{cases}$$

Hint: $\frac{d}{dx}\mathrm{sgn} = 2\delta$. Since Fourier transform converts d/dx to multiplication by $2\pi i x$, this implies that $(2\pi i)x\cdot\widehat{\mathrm{sgn}} = 2\widehat{\delta} = 2$. Thus, $(\pi i)x\cdot\widehat{\mathrm{sgn}} = 1$.

Discussion: From the hint, $x \cdot (\pi i \,\widehat{\text{sgn}}) = 1$. Also, the principal-value functional u from the previous example satisfies $x \cdot u = 1$. Thus,

$$x \cdot (u - \pi i \,\widehat{\text{sgn}}) \;=\; 0$$

By another earlier example, this implies that $u - \pi i \,\widehat{\text{sgn}}$ is a multiple of δ. In fact, the multiple is 0, because δ is *even*, while u, sgn, and thus $\widehat{\text{sgn}}$, are all *odd*.[78] That is, $\widehat{\text{sgn}} \;=\; \frac{1}{\pi i} u$. ///

[22.0.5] Remark: In particular, it is *not quite* that $\widehat{\text{sgn}}(\xi) = 1/\pi i\xi$. Indeed, $1/\xi$ is *not* locally integrable, so does not directly describe a distribution. This example shows that, yes, $\xi \cdot \widehat{\text{sgn}} = 1/\pi i$, but apparently we cannot just *divide* (pointwise values). Indeed, we have proven that the principal-value integral is the Fourier transform (up to constants), and it is not quite just an integral.

[22.52] Show that $x\delta' = \delta$ on $\mathbb{R}$. Similarly, on $\mathbb{R}^n$, show that $x_i\delta = 0$.

Discussion: These are direct computations, using the characterizations of multiplication and of derivative by duality. For the first assertion, for $\varphi \mathscr{S}$,

$$(x\delta')(\varphi) \;=\; \delta'(x \cdot \varphi) \;=\; -\delta((x\varphi)') \;=\; -\delta(\varphi + x\varphi') \;=\; -\delta(\varphi) + 0 \cdot \varphi'(0) \;=\; -\delta(\varphi)$$

as claimed. On $\mathbb{R}^n$, for $\varphi \in \mathscr{S}$,

$$(x_i\delta)(\varphi) \;=\; \delta(x_i\varphi) \;=\; 0 \cdot \varphi(0) \;=\; 0$$

as claimed. ///

[22.53] On $\mathbb{R}^n$, show that $\Delta\delta = 2n \cdot \delta$.

Discussion: Another direction computation, using the duality characterization: for $\varphi \in \mathscr{S}$,

$$(r^2\Delta\delta)(\varphi) \;=\; (\Delta\delta)(r^2\varphi) \;=\; = \; (-1)^2\delta(\Delta(r^2\varphi))$$

Compute

$$\Delta(r^2\varphi) \;=\; \sum_i \frac{\partial^2}{\partial x_i^2}(r^2\varphi) \;=\; \sum_i \frac{\partial}{\partial x_i}(2x_i\varphi + r^2\frac{\partial\varphi}{\partial x_i})$$

$$= \; \sum_i 2\varphi + 2x_i\frac{\partial\varphi}{\partial x_i} + r^2\frac{\partial^2\varphi}{\partial x_i^2} \;=\; 2n\varphi + \sum_i 2x_i\frac{\partial\varphi}{\partial x_i} + nr^2\Delta\varphi$$

Applying δ to this gives

$$2n\varphi(0) + \sum_i 2 \cdot 0 \cdot \frac{\partial\varphi}{\partial x_i}(0) + n \cdot 0 \cdot (\Delta\varphi)(0) \;=\; 2n\varphi(0) \;=\; 2n\delta(\varphi)$$

as claimed. ///

[22.54] On $\mathbb{R}^2$, compute the Fourier transform of $(x \pm iy)^n \cdot e^{-\pi(x^2+y^2)}$ for $n = 0, 1, 2, \ldots$. (*Hint:* Re-express things, including Fourier transform, in terms of $z = x + iy$ and $\overline{z} = x - iy$, $w = u + iv$, and $\overline{w} = u - iv$.)

Discussion: Using z and w, the functions are $z^n e^{-\pi z\overline{z}}$ and $\overline{z}^n e^{-\pi z\overline{z}}$, and Fourier transform is

$$\int_{\mathbb{R}^2} e^{-\pi i(z\overline{w}+\overline{z}w)}\, z^n e^{-\pi z\overline{z}}\, dx\, dy \;=\; \int_{\mathbb{R}^2} e^{-\pi i(z\overline{w}+\overline{z}w)}\, \frac{1}{(-\pi)^n}\Big(\frac{\partial}{\partial\overline{z}}\Big)^n e^{-\pi z\overline{z}}\, dx\, dy$$

[78] This notion of parity can be defined for distributions from the obvious notion for functions $(\theta \cdot f)(x) = f(-x)$, and then $(\theta \cdot v)(f) = v(\theta \cdot f)$ for distributions v.

Imagining that we can integrate by parts, this is

$$(-1)^n \frac{1}{(-\pi)^n} \int_{\mathbb{R}^2} \Big(\frac{\partial}{\partial \overline{z}}\Big)^n e^{-\pi i(z\overline{w}+\overline{z}w)}\, e^{-\pi z\overline{z}}\, dx\, dy \;=\; \frac{1}{\pi^n} \int_{\mathbb{R}^2} (-\pi i w)^n e^{-\pi i(z\overline{w}+\overline{z}w)}\, e^{-\pi z\overline{z}}\, dx\, dy$$

$$=\; (-i)^n\, w^n \int_{\mathbb{R}^2} e^{-\pi i(z\overline{w}+\overline{z}w)}\, e^{-\pi z\overline{z}}\, dx\, dy \;=\; i^{-n}\, w^n\, e^{-\pi(w\overline{w})}$$

since we know the Fourier transform of a Gaussian. A similar computation with roles of $z, \overline{z}$ reversed accomplishes the other computation. That is, $(x \pm iy)^n e^{-\pi(x^2+y^2)}$ is an eigenfunction for Fourier transform, with eigenvalue $i^{-|n|}$. ///