(*April 26, 2012*)

Functions on circles

Paul Garrett garrett@math.umn.edu http://www.math.umn.edu/~garrett/

The simplest physical object with an interesting function theory is the circle, $S^1 = \mathbb{R}/\mathbb{Z}$, which inherits its group structure and *translation-invariant* differential operator d/dx from the real line $\mathbb{R}$. The exponential functions $x \to e^{2\pi i n x}$ (for $n \in \mathbb{Z}$) are group homomorphisms $\mathbb{R}/\mathbb{Z} \to \mathbb{C}^\times$ *and* are eigenfunctions for d/dx on $\mathbb{R}/\mathbb{Z}$. Finite or infinite linear combinations

$$\sum_{n\in\mathbb{Z}} c_n\, e^{2\pi i n x}$$

are **Fourier series**. [1] A function expressed as such a sum is a linear combination of eigenvectors for d/dx. That is, *if* every function has a Fourier expansion, [2] then the linear operator of differentiation is *diagonalized*. However, infinite-dimensional linear algebra is subtler than finite-dimensional. The fundamental questions are [3]

In what sense(s) can a function be expressed as a Fourier series?

Can we differentiate a Fourier series term-by-term?

How cautious must we be in differentiating functions that are only piecewise differentiable?

What will derivatives of discontinuous functions be?

---

[1] In the early 19^{th} century, J. Fourier was an impassioned advocate of the use of such sums, naturally (in the viewpoint of the time) writing sines and cosines rather than complex exponentials. Although Euler, the Bernouillis, and others had used sines and cosines in similar fashions and for similar ends, Fourier made too extravagant a claim, namely that *all* functions could be expressed in such terms. Unfortunately, in those days there were not clear descriptions of classes of functions, nor perhaps even what it would mean to *represent* a function by such a series.

[2] The notion of *has a Fourier expansion* would need to clarify what *has* means. Must it mean that pointwise values can be retrieved from the Fourier series? Less? More?

[3] 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 was more consonant with the expectations of the time, though creating unreasonable expectations for the differentiability of Fourier series.

Several issues are *implicit* in these questions, and the *best* answers need viewpoints created only in the 1930's and 1940's by Sobolev [4] and Schwartz, enabling us to speak legitimately of *generalized functions* (a.k.a., *distributions*). And there are natural technical questions, like

Why define generalized functions as **dual spaces***?*

In brief, Schwartz' 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 functions with jump discontinuitues don't genuinely have, we can instead declare that functions are things we can *integrate against.* For given φ, the map that integrates against it

$$f \longrightarrow \int f(x)\,\varphi(x)\,dx$$

is a *functional* ($\mathbb{C}$-valued linear map), and is probably *continuous* in any 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 we prove that every distribution is a high-order derivative of a continuous function (possibly plus a constant). Thus, since we do want to be able to take derivatives indefinitely, there is no waste in the enlargement of the concept.

Further, in any of the several natural topologies on distributions, 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* yields no excess.

This discussion is easiest on the circle, and making use of Fourier series. [5] This story is a prototype for later, more complicated, examples.

There is an important auxiliary technical lesson here. 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 *Sobolev imbeddings* or *Sobolev inequalities.*

We assume only a modest prior acquaintance with Hilbert and Banach spaces.

---

1. *A confusing example*

Let $s(x)$ be the **sawtooth function** [6]

$$s(x) \;=\; x - \frac{1}{2} \qquad (\text{for } 0 \le x < 1)$$

[4] K. Friedrichs' discussion of unbounded operators on Hilbert spaces used norms defined in terms of derivatives, but only internal to proofs, while for Sobolev these norms were major objects themselves.

[5] The classic reference on trigonometric series is A. Zygmund, *Trigonometric Series, I, II*, first published in Warsaw in 1935, reprinted several times, including a 1959 Cambridge University Press edition. Our viewpoint neglects many interesting details of this situation *but* is readily adaptible to more complicated situations, so necessarily our treatment is different from Zygmund's.

[6] One may also take $s(x) = x$ for $-\frac{1}{2} < x < \frac{1}{2}$ and extend by periodicity. This definition avoids the subtraction of $1/2$, and has the same operational features. In the end, it doesn't matter much which choice of sawtooth function we take.

and made *periodic* by demanding $s(x+n) = s(x)$ for all $n \in \mathbb{Z}$. In other words,

$$s(x) \;=\; x - [\![x]\!] - \frac{1}{2} \qquad (\text{for } x \in \mathbb{R})$$

where as usual $[\![x]\!]$ is the greatest integer less than or equal x. Away from integers, this function is infinitely differentiable, with derivative 1. At integers it jumps down from value to $1/2$ to value $-1/2$. There is no reason to worry about defining a value *at* integers.

Anticipating that Fourier coefficients of $\mathbb{Z}$-periodic functions are computed by integrating against $\psi_n(x) = e^{2\pi i n x}$ (conjugated), by integrating by parts

$$\int_0^1 s(x) \cdot e^{-2\pi i n x}\, dx \;=\; \begin{cases} \frac{1}{-2\pi i n} & (\text{for } n \neq 0) \\ 0 & (\text{for } n = 0) \end{cases}$$

Thus, in whatever sense a function *is* its Fourier expansion, we would have

$$s(x) \;\sim\; \frac{1}{-2\pi i} \sum_{n \neq 0} \frac{1}{n} \cdot e^{2\pi i n x}$$

Even though this series does not converge absolutely for any value of x, as we will see below, [7] it *does* converge to the value of $s(x)$ for x *not* an integer. Since $s(x)$ has discontinuities at integers *anyway*, this is hardly surprising. Nothing disturbing has happened yet.

Now differentiate. The sawtooth function *is* differentiable away from $\mathbb{Z}$, with value 1, and with uncertain value at integers. At this point we must differentiate the Fourier series term-by-term, whether or not we feel confident in doing so. The blatant differentiability of $s(x)$ away from integers suggests that it is not entirely ridiculous to differentiate. Then

$$s'(x) \;=\; \begin{cases} 1 & (\text{for } x \notin \mathbb{Z}) \\ ? & (\text{for } x \in \mathbb{Z}) \end{cases} \;\sim\; -\sum_{n \neq 0} e^{2\pi i n x}$$

The right-hand side is hard to interpret as having values. On the other hand, reasonably interpreted, it is still ok to integrate against this sum: letting c_n be the n^{th} Fourier coefficient of a (smooth) function f,

$$\int_{S^1} f(x)\, (-\sum_{n \neq 0} e^{2\pi i n x}) dx \;=\; -\sum_{n \neq 0} \int_{S^1} f(x)\, e^{2\pi i n x}\, dx$$

$$= \; -\sum_{n \neq 0} c_{-n} \;=\; c_0 - \sum_{n \in \mathbb{Z}} c_n\, e^{2\pi i n \cdot 0} \;=\; \int_{S^1} f(x)\, dx - f(0)$$

The map

$$f \longrightarrow \int_{S^1} f(x)\, dx - f(0)$$

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 S^1. Further, the $-f(0)$ term forcefully suggests that *the derivative of the discontinuity of the sawtooth function is the (periodic) evaluation-at-*0 *functional times* (-1). [8]

---

[7] We will prove that at points where *left and right derivatives* exist, *piecewise continuous* functions' Fourier series do converge to their pointwise values.

[8] The *jump* is downward rather than upward.

[1.0.1] 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, the left-hand side *is* differentiable away from $\mathbb{Z}$, and by repeated differentiation

$$s^{(k+1)}(x) \;=\; \begin{cases} 0 & (\text{for } x \notin \mathbb{Z}) \\ ? & (\text{for } x \in \mathbb{Z}) \end{cases} \;\sim\; -(2\pi i)^k \sum_{n\neq 0} n^k \cdot e^{2\pi i n x}$$

By now the right-hand sides are vividly not convergent. The summands do not go to zero, amd are not bounded.

One can continue differentiating in this symbolic sense, but it is not clear what meaning to attach.

One reaction is simply to 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. 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*. Further, we will see that termwise differentiation of Fourier series does capture (or reflect) this extended notion of function and derivative.

---

2. *Natural function spaces*

Let's review families of functions of natural interest. In all cases, the object is to give the vector space of functions a metric (if possible) which makes it *complete*, so that we can *take limits* and be sure to stay in the same class of functions. For example, *pointwise* limits of continuous functions easily fail to be continuous. [9]

First, we care about *continuous* complex-valued functions. Although we have in mind continuous functions on the circle $S^1 = \mathbb{R}/\mathbb{Z}$, the basic result depends only upon the *compactness* of $\mathbb{R}/\mathbb{Z}$.

We give the set $C^o(K)$ of (complex-valued) continuous functions on a compact topological space K the metric [10]

$$d(f,g) \;=\; \sup_{x\in K} |f(x) - g(x)|$$

The sup is *finite* because K is compact and $f - g$ is continuous. [11] The right-hand side of this last equation

[9] Becoming aware of such possibilities and figuring out how to form graceful hypotheses to avoid them was the fruit of several embarrassing errors and much experimentation throughout the 19^{th} century. Unifying abstract notions such as *metric space* and general *topological space* were only available in the early 20^{th} century, with the work of Hausdorff, Fréchet, and others.

[10] Convergence of a sequence of functions in this topology is sometimes called *uniform convergence*.

[11] A continuous function f on a compact set K is *bounded*: let U_n be $\{x \in K : |f(x)| < n\}$, for $n = 1, 2, \ldots$. The union of the these open U_n's is all of K. By compactness, there is a finite subcover, proving the boundedness.

has a simpler incarnation, the *(sup) norm*[12]

$$|f|_\infty \;=\; |f|_{C^o} = \sup_{x\in K} |f(x)|$$

Thus, $d(f,g) = |f-g|_{C^o}$. A main feature of continuous functions is that they have *pointwise values.* [13] We have the unsurprising but important result:

[2.0.1] Claim: With the sup-norm topology, for $x \in K$ the **evaluation functional**[14] $C^o(K) \to \mathbb{C}$ by

$$f \;\longrightarrow\; f(x)$$

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. ///

[2.0.2] Theorem: The space $C^o(K)$ of (complex-valued) continuous functions on a compact topological space K is *complete.*

[2.0.3] Remark: While there are many *pointwise* limits of continuous functions that are *not* continuous, none of these is *uniform.* Indeed, in the early 19^{th} century the distinction between the two notions was not clearly established, precipitating technical confusion.

[2.0.4] 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 $i \ge N$

$$|f_i(x) - f(x)| \;\le\; |f_i(x) - f_j(x)| + |f_j(x) - f(x)| \;<\; \varepsilon + \varepsilon'$$

Since this is true for every positive ε'

$$|f_i(x) - f(x)| \;\le\; \varepsilon \qquad\qquad (\text{for all } i \ge N)$$

---

[12] Recall that a *norm* $v \to |v|$ on a complex vector space V is a non-negative real-valued function $v \to |v|$ which is *positive* (meaning that $|v| = 0$ only for $v = 0$), *homogeneous* (meaning that $|\alpha\cdot v| = |\alpha|_{\mathbb{C}} \cdot |v|$ for complex α, where $|\alpha|_{\mathbb{C}}$ is the usual complex absolute value), and satisfies the *triangle inequality* (that $|v+w| \le |v| + |w|$). The first two properties are readily verified for the sup norm, and the triangle inequality follows from the readily verifiable fact that the *sup of the sum* is less than or equal the *sum of the sups.*

[13] As innocuous as the property of having well-defined pointwise values is, it is *not* shared by L^2-functions, for example, and this causes many misunderstandings.

[14] As usual, a *(continuous) functional* is a (continuous) linear map to $\mathbb{C}$.

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 $i, j \geq N$ we have $|f_i - f_j| < \varepsilon$. From the previous paragraph

$$|f_i(x) - f(x)| \;\leq\; \varepsilon \qquad\qquad \text{(for every } x \text{ and for } i \geq N)$$

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

$$|f(x) - f(y)| \;\leq\; |f(x) - f_i(x)| + |f_i(x) - f_i(y)| + |f(y) - f_i(y)| \;<\; \varepsilon + \varepsilon + \varepsilon$$

Thus, the pointwise limit f is continuous at every x in U. ///

To talk about *differentiability* return to the concrete situation of $\mathbb{R}$ and its quotient $S^1 = \mathbb{R}/\mathbb{Z}$.

The quotient map $q : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$ yields continuous functions under composition $f \circ q$ for $f \in C^o(S^1) = C^o(\mathbb{R}/\mathbb{Z})$. A stronger assertion 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 $F(x + n) = F(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. Indeed, the periodicity gives a *well-defined* function f on $\mathbb{R}/\mathbb{Z}$. Then the continuity of f follows immediately from the definition of the quotient topology on $S^1 = \mathbb{R}/\mathbb{Z}$.

As usual, a real-valued or complex-valued function f on $\mathbb{R}$ is **continuously differentiable** if it has a derivative which is 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 recursively define

$$f^{(i)} \;=\; \Big(f^{(i-1)}\Big)' \qquad\qquad \text{(for } i > 1)$$

when the corresponding limits exist.

We must make explicit our expectation that *differentiation* on the circle $S^1 = \mathbb{R}/\mathbb{Z}$ is descended from differentiation on the real line. That is, we should *characterize* differentiation on $S^1 = \mathbb{R}/\mathbb{Z}$ in terms of such a compatibility relation. Thus, for $f \in C^k(S^1)$, require that the differentiation D on S^1 be related to the differentiation on $\mathbb{R}$ by

$$(Df) \circ q \;=\; \frac{d}{dx}(f \circ q)$$

Via the quotient map $q : \mathbb{R} \to \mathbb{R}/\mathbb{Z}$, make a *preliminary* definition of the collection of k-times continuously differentiable functions on S^1, with a topology, by

$$C^k(S^1) \;=\; \{f \;\text{ on } S^1 : \; f \circ q \in C^k(\mathbb{R})\}$$

with the C^k**-norm** [15]

$$|f|_{C^k} \;=\; \sum_{0 \leq i \leq k} |(f \circ q)^{(i)}|_\infty \;=\; \sum_{0 \leq i \leq k} \sup_x |(f \circ q)^{(i)}(x)|$$

where $F^{(i)}$ is the (continuous!) i^{th} derivative of F on $\mathbb{R}$. The **associated metric** on $C^k(S^1)$ is

$$d(f, g) \;=\; |f - g|_{C^k}$$

[15] Granting that the sup norm on *continuous* functions is a norm, verification that the C^k-norm is a norm is straightforward.

[2.0.5] Remark: Among other features, the norm on the spaces C^k makes continuity of the *differentiation* map $C^k \to C^{k-1}$ completely obvious.

There is another often-seen version of the norm, namely

$$|f|_{C^k}^{\mathrm{var}} \;=\; \sup_{0\le i\le k} |(f\circ q)^{(i)}|_\infty \;=\; \sup_{0\le i\le k}\, \sup_x\, |(f\circ q)^{(i)}(x)|$$

These two norms give the same topology, for the simple reason that for complex numbers $a_0, \ldots, a_k$

$$\sup_{0\le i\le k} |a_i| \;\le\; \sum_{0\le i\le k} |a_i| \;\le\; (k+1)\cdot \sup_{0\le i\le k} |a_i|$$

[2.0.6] Claim: There is a unique, well-defined, continuous (differentiation) map $D : C^k(S^1) \to C^{k-1}(S^1)$ giving a commutative diagram

$$\begin{array}{ccc} C^k(\mathbb{R}) & \xrightarrow{d/dx} & C^{k-1}(\mathbb{R}) \\ {\scriptstyle -\circ q}\uparrow & & \uparrow{\scriptstyle -\circ q} \\ C^k(S^1) & \xrightarrow{D} & C^{k-1}(S^1) \end{array}$$

[2.0.7] Remark: One might speculate that proof given is needlessly complicated. However, it is worthwhile to do it this way. This approach applies broadly, *and* is as terse as possible without ignoring volatile details.

Proof: The point is that differentiation of periodic functions yields periodic functions. That is, we claim that, for $f \in C^k(S^1)$, the pullback $f\circ q$ has derivative $\frac{d}{dx}(f\circ q)$ which is the pullback $g\circ q$ of a unique function $g \in \mathrm{C}^{k-1}(S^1)$. 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 $S^1 = \mathbb{R}/\mathbb{Z}$ if and only if it is $\mathbb{Z}$-invariant, that is $F(x+n) = F(x)$ for all $x \in \mathbb{R}$ and $n \in \mathbb{Z}$. Then, from our definition of $C^k(S^1)$, a function $F \in C^k(\mathbb{R})$ is a pullback via q from $C^k(\mathbb{R}/\mathbb{Z})$ exactly when $F^{(i)}(x+n) = F^{(i)}(x)$ 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

$$(T_yF)(x) \;=\; F(x+y) \qquad\qquad (\text{for } x, y \in \mathbb{R})$$

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 $n \in \mathbb{Z}$,

$$T_n\Big(\frac{d}{dx}(f\circ q)\Big) \;=\; \frac{d}{dx}(T_n(f\circ q)) \;=\; \frac{d}{dx}(f\circ q)$$

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. ///

[2.0.8] Remark: In light of the uniqueness of differentiation on S^1, from now on write d/dx for the differentiation D on S^1, and $f^{(k)}$ for $D^k f$. Now rewrite the description of $C^k(S^1)$ more simply, as

$$C^k(S^1) \;=\; \{f \;\text{ on } S^1 :\; f\circ q \in C^k(\mathbb{R})\}$$

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 $C^k(S^1)$ 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.

[2.0.9] Claim: With the topology above, the space $C^k(S^1)$ is *complete*, so is a *Banach* space.

Proof: The case $k=1$ illustrates all the points. For a Cauchy sequence $\{f_n\}$ in $C^1(S^1)$, both $\{f_n\}$ and $\{f'_n\}$ are Cauchy in $C^o(S^1)$, 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 $\frac{d}{dx} : C^1(S^1) \to C^o(S^1)$ gives the expected conclusion, that $f' = g$.

What could go wrong? The issue is whether f *is differentiable at all*.

By the fundamental theorem of calculus, for any index i, since f_i is continuous,[16]

$$f_i(x) - f_i(a) = \int_a^x f'_i(t)\,dt$$

Interchanging limit and integral[17] 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

$$f(x) - f(a) = \int_a^x g(t)\,dt$$

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. ///

[2.0.10] Remark: Again, the differentiation map $C^1(S^1) \to C^o(S^1)$ 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

[16] The fundamental theorem of calculus for integrals of *continuous* functions needs only the simplest notion of an integral.

[17] For example, interchange of limit and integral is justified by the simplest form of Lebesgue's Dominated Convergence Theorem. It may seem to be overkill to invoke the Dominated Convergence Theorem in this context, but systematic attention to such details helps us avoid many of the gaffes of early 19^{th}-century analysis. Lebesgue's deserved fame arose exactly because he helped to overcome the murkiness of earlier theories of integration.

the derivatives f_n'. The issue is 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,

[2.0.11] Claim: With the C^k-topology, for $x \in S^1$ and integer $0 \le i \le k$, the **evaluation functional** $C^k(S^1) \to \mathbb{C}$ by

$$f \longrightarrow f^{(i)}(x)$$

is *continuous*. ///

Apply this set-up to Fourier series. Letting $\psi_n(x) = e^{2\pi i n x}$,

[2.0.12] 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 $C^k(S^1)$, and its derivative is computed by term-wise differentiation

$$\frac{d}{dx} \sum c_n \psi_n = \sum (2\pi i n)\, c_n \psi_n \in C^{k-1}(S^1)$$

Proof: The $C^o(S^1)$ norm of a Fourier series is easily estimated, by

$$\Big| \sum_{|n| \le N} c_n \psi_n(x) \Big| \le \sum_{|n| \le N} |c_n| \qquad \text{(for all } x \in S^1)$$

The right-hand side is independent of $x \in S^1$, so bounds the sup over $x \in S^1$. Similarly, estimate derivatives (of partial sums) by

$$\Bigg| \Bigg(\sum_{|n| \le N} c_n \psi_n \Bigg)^{(k)} \Bigg| \le \sum_{|n| \le N} |c_n|\, (2\pi n)^k$$

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, and, unsurprisingly, it does:

[2.0.13] Claim: The inclusion $C^k(S^1) \subset C^{k-1}(S^1)$ is continuous. [18]

Proof: The point is that, for $f \in C^k(S^1)$ the obvious inequality

$$|f|_{C^{k-1}} \le |f|_{C^k}$$

gives an explicit estimate for the continuity. ///

[18] 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.

3. *Topology on* $C^\infty(S^1)$

Next, we care about infinitely differentiable [19] functions, that is, *smooth* functions, denoted $C^\infty(S^1)$. At least as *sets* (or vector spaces),

$$C^\infty(S^1) \;=\; \bigcap_k C^k(S^1)$$

However, this space $C^\infty(S^1)$ of smooth functions provably *does not* have a structure of Banach space. Observing that a descending intersection is a *(projective) limit*, much as an ascending union is a *colimit*, [20] we should declare that

$$C^\infty(S^1) \;=\; \lim_k C^k(S^1)$$

[3.0.1] Remark: Another natural space of functions is $C^o_c(\mathbb{R})$, the space of compactly-supported continuous functions on $\mathbb{R}$. Since (as a set)

$$C^o_c(\mathbb{R}) \;=\; \bigcup_N \{f \in C^o_c(\mathbb{R}) : \text{support } \; f \subset [-N,+N]\} \;=\; \text{colim}_N \{f \in C^o_c(\mathbb{R}) : \text{support } \; f \subset [-N,+N]\}$$

$C^o_c(\mathbb{R})$ is expressed as an ascending union (colimit) of Banach spaces. [21] However, colimits of Banach spaces are even less likely to be Banach spaces than are limits. In this example, the colimit is not even *metrizable*, since, as we'll see later, a topology in which it is suitably *complete* makes it violate the conclusion of Baire's theorem. That is, $C^o_c(\mathbb{R})$ is a countable union of nowhere dense subsets. [22] We will address this when looking at functions on $\mathbb{R}$.

Returning to $C^\infty(S^1) = \lim_k C^k(S^1)$, unfortunately we are 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*. By now we know that limits of *topological spaces* exist (and are necessarily unique), but we are demanding more structure. To explain this requires a general notion of *topological vector space*.

[3.0.2] Remark: If we too optimistically speculate about what the limit might be, we can err: as it happens, this limit is provably *not* a Banach space (nor Hilbert space). [23] A limit of topological spaces has a unique topology, whatever it may be. We saw earlier that limit topologies are compatible with group structures,

[19] Use of *infinitely* here is potentially misleading, but is standard. Sometimes the phrase *indefinitely differentiable* is used, but this also offers potential for confusion. A better (and standard) contemporary usage is *smooth*.

[20] Although descending intersections are (projective) limits, with *projections* being the natural *inclusions*, not every limit should be construed as a descending intersection. Indeed, solenoids cannot usefully be construed as descending intersections. The notion of limit is broader than either of these.

[21] That the limitands are Banach spaces is easy: first, the set of continuous functions with support in $[-N,N]$ is the subspace of the Banach space $C^o[-N,N]$ defined by $f(-N) = f(N) = 0$. These evaluation functionals are continuous, so this is a *closed* subspace of a Banach space, and is Banach. Since $[-N,N]$ is *compact*, the space $C^o[-N,N]$ with the sup norm is Banach, as we saw earlier.

[22] The Baire Category Theorem, which applies both to complete metric spaces and to locally compact Hausdorff spaces, asserts that the space is not a countable union of nowhere dense subsets. A nowhere dense subset is one whose closure contains no non-empty open sets.

[23] The argument for the impossibility is not the main point. Still, it is reasonable to wonder how such a thing could be 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 its *completeness*.

and we anticipate that similar arguments will prove that the limit has a vector space structure compatible with the limit's topology.

[3.0.3] Remark: There is also the disquieting question of what *test objects* Z we should consider, with compatible mappings $Z \to C^k(S^1)$ to *define* limits in the first place. Similarly, we should wonder how large a class of vector spaces with topologies is needed to know that such a limit falls into it. That is, *existence* is non-trivial.

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

$$V \times V \to V \quad \text{by} \quad v \times w \to v + w \quad \text{is continuous}$$

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*. [24] But we require more, that the topological vector spaces be **locally convex** in the sense that there is a local basis at 0 consisting of *convex* sets. [25] It is easy to prove that Hilbert and Banach spaces are locally convex, which is why the issue is invisible in a smaller context which considers *only* such topological vector spaces. Dismayingly, there are easily constructed complete (invariantly) metrized topological vector spaces which are *not* locally convex. [26]

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), as with topological groups we can give 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). Having enlarged the class of objects sufficiently, unsurprisingly,

[3.0.4] Claim: Products and limits of topological vector spaces exist. Products and limits of locally convex spaces are locally convex. (Proof in appendix.)

[3.0.5] Remark: As usual, if they exist at all, then products and limits are unique up to unique isomorphism.

[3.0.6] Remark: Coproducts and colimits are more fragile, and we will treat them later, as needed, in discussion of function theory on non-compact spaces.

Thus, $C^\infty(S^1)$ has a (limit) topology.

But the relevant sense of bounded *cannot* be the usual metric sense. Instead, a set E in a topological vector space is *bounded* if, for every open neighborhood U of 0, there is $t > 0$ such that for all complex z with $|z| \geq t$ we have the containment $E \subset z \cdot U$. That is, sufficiently large *dilates* of opens eventually contain E. But (as we will eventually show) open balls in $C^k(S^1)$ are *not* contained in *any* dilate of open balls in $C^{k+1}(S^1)$. The definition of the limit topology then will show that $C^\infty(S^1)$ is not normable. A more detailed discussion will be given later.

[24] 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. But this is not the point.

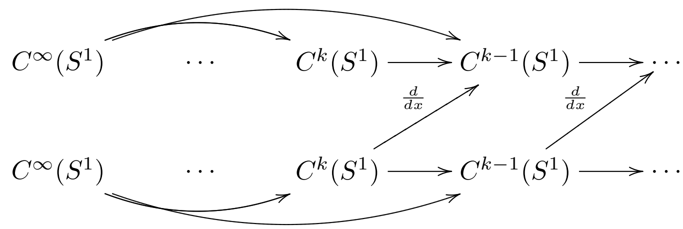
[25] This sense of convexity is the usual, namely, that a set X in a vector space is convex if, for all tuples $x_1, \ldots, x_n$ of points in X and all tuples $t_1, \ldots, t_n$ of non-negative reals with $\Sigma_i t_i = 1$, the sum $\Sigma_i t_i x_i$ is again in X.

[26] 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 interesting exercise to prove that it is not locally convex.

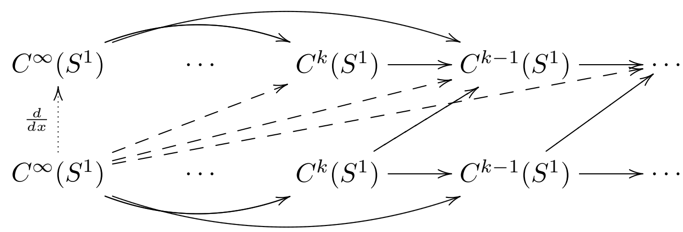
[3.0.7] Claim: Differentiation $f \to f'$ is a *continuous* map $C^\infty(S^1) \to C^\infty(S^1)$.

[3.0.8] Remark: Of course differentiation maps the smooth functions to themselves. What is not *a priori* clear is that differentiation is continuous in the limit topology.

Proof: We already know that differentiation d/dx gives a continuous map $C^k(S^1) \to C^{k-1}(S^1)$. Differentiation is compatible with the inclusions among the $C^k(S^1)$. Thus, we have a commutative diagram



Composing the projections with d/dx gives (dashed) induced maps from $C^\infty(S^1)$ to the limitands, which induce a unique (dotted) map to the limit, as in



This proves the continuity of differentiation, in the limit topology. ///

[3.0.9] Corollary: When a Fourier series $\sum_n c_n \, \psi_n$ satisfies

$$\sum_m |c_n| \, |n|^N \; < \; +\infty \qquad \text{(for every } N\text{)}$$

the series is a smooth function, which can be differentiated term-by-term, to see that its derivative is

$$\sum_m c_n \, (2\pi i n) \, \psi_n$$

Proof: The hypothesis assures that the Fourier series lies in C^k for every k. Differentiation is continuous in the limit topology on C^∞. ///

[3.0.10] Remark: This continuity is necessary to define differentiation of *distributions*, in the following.

---

4. *Distributions (generalized functions)*

Although much amplification is needed, having topologized $C^\infty(S^1)$ we can give the bare definition: a **distribution** or **generalized function** [27] on S^1 is a *continuous linear functional* [28]

$$u : C^\infty(S^1) \longrightarrow \mathbb{C}$$

[27] What's in a name? In this case, this expresses our *intention* to think of distributions as extensions of ordinary functions, rather than as abstract things in a dual space.

[28] 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*.

Why a dual space? Unsurprisingly, as soon as we have a notion of integral on S^1 (in the next section), a function $\varphi \in C^o(S^1)$ gives rise to a distribution u_φ by **integration against** φ,

$$u_\varphi(f) \;=\; \int_{S^1} f(x)\,\varphi(x)\,dx \qquad\qquad (\text{for } f \in C^\infty(S^1))$$

Thus, we relax our notion of *function*, no longer requiring *pointwise values*, but only that a function can be *integrated against*. Then it makes sense to declare functionals in a dual space to be generalized functions. The vector space of distributions is denoted

$$\text{distributions} \;=\; C^\infty(S^1)^* \;=\; \text{Hom}^o_{\mathbb{C}}(C^\infty(S^1), \mathbb{C}) \;=\; \text{continuous dual of } C^\infty(S^1)$$

That is, given a reasonable notion of integral, we have a continuous imbedding

$$C^o(S^1) \;\subset\; C^\infty(S^1)^* \quad \text{by} \quad \varphi \longrightarrow u_\varphi \quad \text{where (again)} \quad u_\varphi(f) \;=\; \int_{S^1} f(x)\,\varphi(x)\,dx \quad (f \in C^\infty(S^1))$$

[4.0.1] Remark: One further benefit of defining *generalized functions* (distributions) as functionals in a dual space is that this describes the totality of them at one stroke. Alternatives, such as various notions of *completion*, are less clear and less decisive.

Duals of limits do *not* behave well. Duals of *colimits do* behave well. It *is* true that in any reasonable situation

$$\text{Hom}(\text{colim}_i X_i, Z) \;\approx\; \text{lim}_i \text{Hom}(X_i, Z)$$

But $C^\infty(S^1)$ is a *limit*, not a colimit. Luckily, using particulars of the situation, we obtain the following useful result, that *the (continuous) dual of a limit of Banach spaces is the colimit of the (continuous) duals.*

[4.0.2] Theorem: Let $X = \lim_i B_i$ be a limit of Banach spaces B_i with projections $p_i : X \to B_i$. Any $\lambda \in X^* = \text{Hom}^o_{\mathbb{C}}(X, \mathbb{C})$ 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,

$$(\text{lim}_i B_i)^* \;\approx\; \text{colim}_i B_i^*$$

Proof: 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 $\mathbb{C}$,[29] by the continuity at 0. By properties of the limit topology[30] 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[31] with $j \geq i_t$ for all t, and

$$V_j' \;=\; \bigcap_{t=1}^n p_{i_t,j}^{-1} V_{i_t} \subset B_j$$

[29] The choice of the *unit* ball is arbitrary. Any fixed ball will do, since all maps are linear.

[30] The limit $X = \lim_i B_i$ is the closed subspace (with the subspace topology) of the product $Y = \Pi_i B_i$ consisting 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 $i_1, \ldots, i_n$.

[31] 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 $k > j$.

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 [32] 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. [33] Simplifying notation, let $\lambda : X \to \mathbb{C}$ and $p : X \to N$ be continuous linear to a normed space N, with

$$\lambda(p^{-1} V) \subset \text{unit ball in } \mathbb{C} \qquad \text{(for some neighborhood } V \text{ of 0 in } N)$$

We claim that λ factors through $p : X \to N$ as a (continuous) linear necessarily continuous) 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

$$\lambda\left(\bigcap_n \frac{1}{n} \cdot p^{-1} V\right) \subset \frac{1}{m}\text{-ball} \qquad \text{(for all } m)$$

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

$$\mu(\varepsilon \cdot V) = \varepsilon \cdot \mu V \subset \varepsilon\text{-ball in } \mathbb{C}$$

---

[32] 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*.

[33] 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 *not* all of B_j. For example, B_j may be a *completion* of $p_j X$.

proving the continuity of $\mu : N \to \mathbb{C}$. [34] This proves the claim.

Applying the claim, we obtain a 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 [35] to the closure of $p_j X$ in B_j. The Hahn-Banach theorem asserts that a continuous linear functional on a closed subspace (such as the closure of $p_j X$) extends [36] to a continuous linear functional on B_j, which is the desired map. ///

[4.0.3] Remark: The same proof shows that a continuous linear map from a limit of Banach spaces to a *normed* space factors through a limitand.

[4.0.4] Corollary: The space of distributions on S^1 is the ascending union (colimit)

$$C^\infty(S^1)^* \;=\; \big(\lim{}_k C^k(S^1)\big)^* \;=\; \mathrm{colim}_k C^k(S^1)^* \;=\; \bigcup{}_k C^k(S^1)^*$$

of duals of the Banach spaces $C^k(S^1)$. ///

The **order** of a distribution is the smallest k such that the distribution lies inside the dual $C^k(S^1)^*$. Since the space of all distributions is this colimit, the order of a distribution is well-defined. [37]

Thinking of distributions as generalized *functions*, we should be able to *differentiate* them 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 a suitable notion of integral on S^1 (as in the next section), for *functions* f, g, by integration by parts,

$$\int_{S^1} f(x)\, g'(x)\, dx \;=\; -\int_{S^1} f'(x)\, g(x)\, dx$$

with no boundary terms because S^1 has empty boundary. Note the negative sign. Motivated by this, define the **distributional derivative** u' of $u \in C^\infty(S^1)^*$ to be another distribution defined by

$$u'(f) \;=\; -u(f') \qquad\qquad \text{(for any } f \in C^\infty(S^1))$$

The continuity of differentiation $\frac{d}{dx} : C^\infty(S^1) \to C^\infty(S^1)$ assures that u' is a distribution, since

$$u' = -(u \circ \frac{d}{dx}) \;:\; C^\infty(S^1) \to \mathbb{C}$$

[34] 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.

[35] The extension by continuity is unambiguous, since λ_j is *linear*. In more detail: let λ be a continuous linear function on a dense subspace Y of a topological vector space X. Given $\varepsilon > 0$, let U be a *convex* neighborhood of 0 in X such that for $y \in U$ we have $|\lambda y| < \varepsilon$. We may also 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. Using 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.

[36] When we know that $C^\infty(S^1)$ is *dense* in $C^k(S^1)$, there is no need to invoke the Hahn-Banach theorem to extend the functional λ_j.

[37] The *Riesz representation theorem* asserts that the dual of $C^o(S^1)$ is *Borel measures* on S^1, so zero-order distributions are Borel measures. For example, elements η of $L^2(S^1)$ can be construed as Borel measures, by giving integrals $f \to \int_{S^1} f(x)\,\eta(x)\,dx$ for $f \in C^o(S^1)$. 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 $C^k(S^1)$'s don't have such a classical interpretation.

5. *Invariant integration, periodicization*

We need a notion of *(invariant) integral* on the circle $S^1 = \mathbb{R}/\mathbb{Z}$. The main property required is **translation invariance**, meaning that, for a (for example) continuous function f on S^1,

$$\int_{S^1} f(x+y)\,dx \;=\; \int_{S^1} f(x)\,dx \qquad\qquad (\text{for all } y \in S^1)$$

This invariance is sufficient to prove that various important integrals *vanish*.

For example, let $\psi_m(x) = e^{2\pi i m x}$. Then we have an instance of an important idea:

[5.0.1] Claim: For $m \neq n$,

$$\int_{S^1} \psi_m(x)\,\overline{\psi}_n(x)\,dx \;=\; 0$$

Proof: Since $m \neq n$, the function $f(x) = \psi_m(x)\overline{\psi}_n(x)$ is a *non-trivial* (not identically 1) group homomorphism $S^1 \to \mathbb{C}^\times$. Thus, there is $y \in S^1$ such that $f(y) \neq 1$. Then, since the *change of variables* $x \to x+y$ in the integral does not change the overall value of the integral,

$$\int_{S^1} f(x)\,dx \;=\; \int_{S^1} f(x+y)\,dx \;=\; \int_{S^1} f(x)\cdot f(y)\,dx \;=\; f(y)\int_{S^1} f(x)\,dx$$

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 $t \neq 1$. ///

[5.0.2] Remark: This *vanishing trick* is impressive, since nothing specific about the continuous group homomorphism f or topological group (S^1 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*[38] is missing for general groups. Nevertheless, when we *do* have an invariant measure, the same argument will succeed.

Less critically than the invariance, we want a **normalization**[39]

$$\int_{S^1} 1\,dx \;=\; \mathrm{vol}\,(S^1) \;=\; \mathrm{vol}\,(\mathbb{R}/\mathbb{Z}) \;=\; 1$$

Then

$$\int_{S^1} |\psi_n(x)|^2\,dx \;=\; \int_{S^1} 1\,dx \;=\; 1$$

Thus, without any explicit presentation of the integral or measure, we have proven that the distinct exponentials are an *orthonormal set* with respect to the inner product

$$\langle f, g\rangle \;=\; \int_{S^1} f(x)\,\overline{g}(x)\,dx$$

[38] 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.

[39] We are not required to normalize the measure of the circle to be 2π, which would suggest presenting the circle as $\mathbb{R}/2\pi\mathbb{Z}$. Our normalization of S^1 as $\mathbb{R}/\mathbb{Z}$ is comparably natural.

We expect an **integration by parts** formula, presumably with no *boundary terms* since $S^1 = \mathbb{R}/\mathbb{Z}$ has empty boundary. Indeed, without *constructing*[40] the invariant integral, we prove what we want from its *properties*:

[5.0.3] Claim: Let $f \to \int_{S^1} f(x)\,dx$ be an invariant integral on S^1, for $f \in C^o(S^1)$. Then for $f \in C^1(S^1)$

$$\int_{S^1} f'(x)\,dx \;=\; 0$$

and we have the *integration by parts* formula for $f, g \in C^1(S^1)$

$$\int_{S^1} f(x)\,g'(x)\,dx \;=\; -\int_{S^1} f(x)'\,g(x)\,dx$$

[5.0.4] 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 $\int_{S^1} f'(x)dx = \int_0^1 f(x)\,dx = f(1) - f(0) = 0$ because f is periodic. The latter argument is valid, but fails to show a mechanism applicable in general.. 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*.[41]

$$\int_{S^1} f'(x)\,dx \;=\; \int_{S^1} \frac{\partial}{\partial t}\big|_{t=0}\; f(x+t)\,dx \;=\; \frac{d}{dt}\big|_{t=0} \int_{S^1} f(x+t)\,dx \;=\; \frac{d}{dt}\big|_{t=0} \int_{S^1} f(x)\,dx \;=\; 0$$

by changing variables in the integral. Then apply this to the function $(f \cdot g)' = f'g + fg'$ to obtain

$$\int_{S^1} f'(x)\,g(x)\,dx + \int_{S^1} f(x)\,g'(x)\,dx \;=\; 0$$

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 $\mathbb{R} \to \mathbb{R}/\mathbb{Z} = S^1$ of $\mathbb{R}$ we expect a **compatibility**[42]

$$\int_{\mathbb{R}} F(x)\,dx \;=\; \int_{\mathbb{R}/\mathbb{Z}} \left(\sum_{n\in\mathbb{Z}} F(x+n)\right)\,dx$$

for at least *compactly-supported* continuous functions F on $\mathbb{R}$.

[40] It is usually better to *characterize*, not *construct*.

[41] The argument bluntly demands this interchange of limit and differentiation, so *justification* of it is secondary to the act itself. In the near future we will definitively justify this and other interchanges via *Gelfand-Pettis* (also called *weak*) integrals. In the present concrete situation elementary (but opaque) arguments can be invoked.

[42] In contrast to many sources, this compatibility is *not* about choosing representatives in $[0, 1)$ or anywhere else for $\mathscr{S}^1$. Rather, this compatibility would be required for a group G (here $\mathbb{R}$), a discrete subgroup Γ (here $\mathbb{Z}$), and the quotient G/Γ (here S^1), 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.

Indeed, we can *define* integrals of functions on S^1 by this compatibility relation, by expressing a continuous function f on S^1 as a **periodicization** (or **automorphization**)

$$f(x) \;=\; \sum_{n\in\mathbb{Z}} F(x+n)$$

of a compactly supported continuous function F on $\mathbb{R}$, and *define*

$$\int_{S^1} f(x)\,dx \;=\; \int_{\mathbb{R}} F(x)\,dx$$

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 S^1 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 $q:\mathbb{R}\to\mathbb{R}/\mathbb{Z}=S^1$ the unit interval $[0,1]$ maps surjectively, with the endpoints being identified (and no other points identified). In traditional terminology, $[0,1]$ is a **fundamental domain** [43] for the action of $\mathbb{Z}$ on $\mathbb{R}$. Then define the integral of f on S^1 by

$$\int_{S^1} f(x)\,dx \;=\; \int_0^1 (f\circ q)(x)\,dx$$

with usual (Lebesgue) measure on the unit interval. The verification of the *compatibility* with integration on $\mathbb{R}$ is silly, from this viewpoint.

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. [44] Take $y\in\mathbb{R}$, and compute

$$\int_{S^1} f(x+y)\,dx \;=\; \int_0^1 (f\circ q)(x+y)\,dx \;=\; \int_{-y}^{1-y} (f\circ q)(x)\,dx$$

$$=\; \int_{-y}^{0} (f\circ q)(x)\,dx + \int_0^{1-y} (f\circ q)(x)\,dx \;=\; \int_{-y}^{0} (f\circ q)(x-1)\,dx + \int_0^{1-y} (f\circ q)(x)\,dx$$

since $(f\circ q)(x)=(f\circ q)(x-1)$ by periodicity. Then, replacing x by $x+1$ in the first integral, this is

$$\int_{1-y}^{1} (f\circ q)(x)\,dx + \int_0^{1-y} (f\circ q)(x)\,dx \;=\; \int_0^1 (f\circ q)(x)\,dx$$

[43] 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. Instead, the *quotient* should be treated directly.

[44] While suppressing our disgust.

6. *Hilbert space theory of Fourier series*

The spaces of functions $C^k(S^1)$ do not have Hilbert space structures, only Banach space structures. [45] The rich geometry of Hilbert spaces gives a strong motivation to find a way to compare these Banach spaces with Hilbert spaces. Fourier series provide a good device for this comparison. [46]

As discussed above, we do not need an *explicit* or *formulaic* integral on S^1, only the *property* of *translation invariance*, choice of *normalization*, and the property of *compatibility* with integration on $\mathbb{R}$. Nevertheless, for expediency, we may integrate functions on $S^1 = \mathbb{R}/\mathbb{Z}$ by pulling back to $\mathbb{R}$ and integrating on $[0,1]$.

As usual, $L^2(S^1) \approx L^2[0,1]$ is the collection of measurable functions f on $[0,1]$ such that

$$\int_0^1 |f(x)|^2 \, dx \ < \ +\infty$$

modulo the equivalence relation of almost-everywhere equality, with the inner product

$$\langle f, g \rangle \ = \ \int_0^1 f(x) \, \overline{g}(x) \, dx$$

This definition makes $L^2[0,1]$ a *Hilbert space.* [47]

As observed in the previous section, the exponential functions

$$\psi_n(x) \ = \ e^{2\pi i n x} \qquad\qquad (\text{for } n \in \mathbb{Z})$$

form an orthonormal *set* in $L^2(S^1)$. It is not clear that they form an orthonormal *basis*. That is, we should show that the finite linear combinations of the exponential functions ψ_n are *dense* in $L^2(S^1)$. The **Fourier coefficients** of an L^2 function f are the inner products

$$n^{th} \text{ Fourier coefficient of } f \ = \ \langle f, \psi_n \rangle \ = \ \int_0^1 f(x) \, \overline{\psi}(x) \, dx$$

The **Fourier expansion** of f in $L^2(S^1)$ is

$$f(x) \ \sim \ \sum_{n \in \mathbb{Z}} \langle f, \psi_n \rangle \, \psi_n(x)$$

We do not write *equality* here, for several reasons, clarified in the sequel. We *do* expect to have an equality of (equivalence classes of) functions in $L^2(S^1)$.

[45] It is not trivial to prove that there are no Hilbert space structures on these spaces of functions, without knowing the technical ramifications of the geometry of Hilbert spaces. Even if the given Banach space structure is not Hilbert, that does not preclude the existence of a *different* Banach space structure that *does* come from a Hilbert structure. More to the point is that the *natural* structures are Banach.

[46] Fourier series and Fourier transforms play a central role in the work of both Sobolev (1930's) and Schwartz (1940's) on generalized functions.

[47] The completeness requires proof. The kernel of the idea is that a Cauchy sequence of L^2 functions has a subsequence that converges pointwise off a set of measure 0. This gives a pointwise definition of the candidate limit. A different definition of $L^2(S^1)$, namely as the completion of $C^o(S^1)$ with respect to the L^2 metric has the virtue that the completeness is defined into it, but the workload has been shifted to proving that these two definitions are the same.

Before we prove that the ψ_n are an orthonormal *basis*, we need

[6.0.1] Claim: *(Riemann-Lebesgue lemma)* For any $f \in L^2(S^1)$, the Fourier coefficients of f go to 0, that is,

$$\lim_n \langle f, \psi_n \rangle \;=\; 0$$

Proof: *Bessel's inequality*[48]

$$|f|^2_{L^2} \;\geq\; \sum_n |\langle f, \psi_n \rangle|^2$$

from abstract Hilbert-space theory applies to any orthonormal *set*, whether or it is an orthonormal *basis*. Thus, the sum on the right converges, so the summands go to 0. ///

Even before proving that the exponentials form an orthonormal *basis*, we prove a limited (but useful) *pointwise convergence* result. The hypotheses of the following claim are far from optimal, but are sufficient for our purposes, and are readily verifiable.

First, say that a function f on $S^1 = \mathbb{R}/\mathbb{Z}$ 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.

[6.0.2] Claim: Let f be piecewise C^o on S^1. 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,

$$f(x_o) = \sum_{n \in \mathbb{Z}} \langle f, \psi_n \rangle \; \psi_n(x_o) \qquad \text{(a convergent sum)}$$

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 $f(x_o) = 0$. Note that

$$\int_0^1 f(x + x_o)\, \overline{\psi}_n(x)\, dx \;=\; \int_0^1 f(x)\, \overline{\psi}_n(x - x_o)\, dx \;=\; \psi_n(x_o) \int_0^1 f(x)\, \overline{\psi}_n(x)\, dx \;=\; \psi_n(x_o) \langle f, \psi_n \rangle$$

The left-hand side is the n^{th} Fourier coefficient of $f(x + x_o)$, that is, the n^{th} Fourier *term* of $f(x + x_o)$ evaluated at 0, while the right-hand side is the n^{th} Fourier *term* of $f(x)$ evaluated at x_o. Thus, simplify further by taking $x_o = 0$, without loss of generality.

Write out a partial sum of the Fourier expansion evaluated at 0

$$\sum_{-M \leq n < N} \langle f, \psi_n \rangle \;=\; \sum_{-M \leq n < N} \int_0^1 f(x) \overline{\psi}_n(x)\, dx \;=\; \int_0^1 f(x) \sum_{-M \leq n < N} \overline{\psi}_n(x)\, dx$$

[48] It is easy to prove Bessel's inequality: given a *finite* orthonormal set $\{e_i\}$ and a vector v in a Hilbert space,

$$0 \;\leq\; |v - \sum_i \langle v, e_i \rangle e_i|^2 \;=\; |v|^2 - 2 \sum_i \langle v, e_i \rangle \overline{\langle v, e_i \rangle} + \sum_i |\langle v, e_i \rangle|^2 \;=\; |v|^2 - \sum_i |\langle v, e_i \rangle|^2$$

Given an *arbitrary* orthonormal set, the sum is the sup of the finite sub-sums, so the general Bessel inequality follows.

$$= \int_0^1 \frac{f(x)}{\psi_{-1}(x)-1}\,(\overline{\psi}_N(x) - \overline{\psi}_{-M}(x))\,dx \;=\; \langle \frac{f}{\psi_{-1}-1},\,\psi_N\rangle - \langle \frac{f}{\psi_{-1}-1},\,\psi_{-M}\rangle \to 0-0=0$$

by Riemann-Lebesgue, whenever $f(x)/(\psi_{-1}(x)-1)$ is in $L^2(S^1)$. 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, ignoring the endpoints, which do not contribute to the integrals, $f/(\psi_{-1}-1)$ 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 $0 = f(x_o)$. ///

[6.0.3] Corollary: Fourier series of a function in $C^1(S^1)$ converge pointwise to the function's values. ///

[6.0.4] Remark: Pointwise convergence is not L^2 convergence. Thus, we have *not* yet proven that the exponentials are an orthonormal *basis* for the Hilbert space $L^2(S^1)$. The pointwise result just proven is suggestive, but not decisive.

---

7. *Completeness in* $L^2(S^1)$

There are many different proofs that the exponentials ψ_n form an orthonormal basis in $L^2(S^1)$. At one extreme, a completely analogous result holds for any *compact abelian* topological group, without further structure. Even the abelian-ness can be dropped without much harm, though with complications. Further, formulation and proof of that general result uses completely wholesome methods, so we will give it later. But, for the moment, we take advantage of the particulars of the circle to give a slightly old-fashioned argument, roughly following Fejer. We *do* take this opportunity to introduce the notion of *approximate identity* (below). [49]

[7.0.1] Theorem: The exponentials ψ_n are an orthonormal basis for $L^2(S^1)$.

[7.0.2] Remark: The proof does *not* prove that Fourier series converge pointwise (which they often do not).

[7.0.3] Remark: The pointwise convergence of Fourier series of C^1 functions does *not* instantly yield L^2 convergence, nor does it instantly yield *uniform* pointwise convergence, which *would* imply L^2 convergence. Modifying the earlier idea to achieve these goals *gracefully* motivates introduction of an *approximate identity* in the proof.

Proof: First, by *Urysohn's Lemma* (see appendix), the continuous functions $C^o(S^1)$ are dense in $L^2(S^1)$. Thus, it suffices to prove that C^o functions are approximable in L^2 by finite sums of the exponentials ψ_n. For the latter it suffices to prove that finite sums of exponentials approximate C^o functions in the C^o topology. Indeed, the total measure of the space S^1 is 1, so the L^2 norm of a continuous function is less than or equal its sup norm. Thus, the sup norm gives a topology at least as *fine* as the restriction to C^o of the L^2 norm topology. Therefore, density in sup norm implies density in L^2 norm. Fortunately, proving that C^o functions are approximable in the sup norm by finite sums of exponentials does *not* require that we prove that Fourier

[49] The notion of *approximate identity* is standard and important, but may not be visible in introductory sources.

series of continuous functions converge. That is, we are *not* compelled to prove that the partial sums of the Fourier series are the approximating sums.

The most obvious sequence of candidates for approximating a continuous function would be the sequence of partial sums of its Fourier series. But it turns out that this cannot possibly succeed. [50] Still, we should review the situation. As in our proof of pointwise convergence for C^1 functions, the N^{th} partial sum of the Fourier series of a function f on $S^1 = \mathbb{R}/\mathbb{Z}$ can be described usefully as an *integral operator*[51] by

$$\sum_{|k|\le n} \langle f, \psi_k\rangle\, \psi_k(x) \;=\; \int_{S^1} f(y) \sum_{|k|\le n} \overline{\psi}_k(y)\, \psi_k(x)\, dy \;=\; \int_{S^1} f(y) \sum_{|k|\le n} \psi_k(x-y)\, dy$$

$$= \int_{S^1} f(y)\, \frac{\psi_{n+1}(x-y) - \psi_{-n}(x-y)}{\psi_1(x-y) - 1}\, dy$$

by summing finite geometric series. Let $K_n(x)$ be the summed geometric series

$$K_n(x) \;=\; \frac{\psi_{n+1}(x) - \psi_{-n}(x)}{\psi_1(x) - 1} \;=\; \frac{\psi_{n+\frac{1}{2}}(x) - \psi_{-(n+\frac{1}{2})}(x)}{\psi_{\frac{1}{2}}(x) - \psi_{-\frac{1}{2}}(x)} \;=\; \frac{\sin 2\pi(n+\frac{1}{2})x}{\sin \pi x}$$

so

$$\sum_{|k|\le n} \langle f, \psi_k\rangle\, \psi_k(x) \;=\; \int_{S^1} f(y)\, K_n(x-y)\, dy$$

Granting that it is futile to prove that the partial sums converge pointwise for continuous functions, we might try to see what related but different integral operators would work better.

This is our excuse to introduce **approximate identities** in this situation. [52] The *rough* idea of *approximate identity* is of a sequence $\{\varphi_n\}$ of functions φ_n approximating a *point-mass measure* [53] at $0 \in S^1$. [54] Precisely, a sequence of continuous functions φ_n on S^1 is an approximate identity if

$$\varphi_n(x) \;\ge\; 0 \quad \text{(for all } n, x) \qquad \int_{S^1} \varphi_n(x)\, dx \;=\; 1 \quad \text{(for all } n)$$

and if for every $\varepsilon > 0$ and for every $\delta > 0$ there is n_o such that for all $n \ge n_o$

$$\int_{|x|<\delta} \varphi_n(x)\, dx > 1 - \varepsilon \qquad \text{(equivalently, } \int_{\delta \le |x| < \frac{1}{2}} \varphi_n(x)\, dx < \varepsilon)$$

[50] Early examples of continuous functions whose Fourier series diverge at a point appear in Fejer, L, (1910) *Beispiele stetiger Funktionen mit divergenter Fourierreihe* Journal Reine Angew. Math. 137, pp. 1-5. Existential arguments are not hard to give based on the *Baire category theorem*. But we don't need the details of the failure.

[51] The notion of *integral operator* T is very general, and perhaps has no genuine precise definition. The *form* is that $Tf(x) = \int K(x,y)\, f(y)\, dy$, and the function $K(x,y)$ is the *kernel* of the operator. The class of functions in which the kernel lies, and in which the input and output lie, varies enormously with context.

[52] Various versions of *approximate identity* are useful and important on *topological groups* generally.

[53] A *point-mass* measure at a point x_o is a measure which gives the point x_o measure 1 and gives a set not containing x_o measure 0. These are also called *Dirac measures*.

[54] Unfortunately, and dangerously, it is easy to give a too-naive formulation of the notion of approximate identity, which *fails*. In particular, the requirement that φ_n form an approximate identity is *strictly stronger* than the condition that φ_n approach a Dirac measure in a *distributional* sense. Specifically, the *non-negativity* condition on an approximate identity is indispensable.

where we use coordinates in $\mathbb{R}$ for $S^1 = \mathbb{R}/\mathbb{Z}$. That is, the functions φ_n are non-negative, their integrals are all 1, and their *mass bunches up* at $0 \in S^1$. It is not surprising that the integral operators made from an approximate identity have a useful property:

[7.0.4] Claim: For $f \in C^o(S^1)$ on S^1 and for an approximate identity φ_n, in the topology of $C^o(S^1)$,

$$\lim_n \int_{S^1} f(y)\, \varphi_n(x-y)\, dy \;=\; f(x)$$

Granting this claim, making an approximate identity out of finite sums of exponentials will prove that such finite sums are dense in $C^o(S^1)$. [55]

Proof: *(of claim)* Given $f \in C^o(S^1)$, and given $\varepsilon > 0$, by *uniform* continuity of f on the compact S^1, there is $\delta > 0$ such that

$$|f(x) - f(y)| \;<\; \varepsilon \qquad\qquad (\text{for } |x-y| < \delta)$$

Let n be sufficiently large such that

$$\int_{|x|<\delta} \varphi_n(x)\, dx \;>\; 1 - \varepsilon$$

Since the total mass of φ_n is 1,

$$\int_{S^1} f(y)\, \varphi_n(x-y)\, dy - f(x) \;=\; \int_{S^1} (f(y) - f(x))\, \varphi_n(x-y)\, dy$$

$$= \int_{|y-x|<\delta} (f(y) - f(x))\, \varphi_n(x-y)\, dy + \int_{\delta \le |y-x| \le \frac{1}{2}} (f(y) - f(x))\, \varphi_n(x-y)\, dy$$

The second integral is easily estimated by

$$\Big|\int_{\delta \le |y-x| \le \frac{1}{2}} (f(y) - f(x))\, \varphi_n(x-y)\, dy\Big| \;\le\; 2|f|_C^o \int_{\delta \le |y-x| \le \frac{1}{2}} \varphi_n(x-y)\, dy \;\le\; 2|f|_C^o \cdot \varepsilon$$

Estimation of the integral near 0 uses the positivity of the φ_n:

$$\Big|\int_{|y-x|<\delta} (f(y) - f(x))\, \varphi_n(x-y)\, dy\Big| \;\le\; \varepsilon \int_{|y-x|<\delta} \varphi_n(x-y)\, dy \;\le\; \varepsilon \int_{S^1} \varphi_n(x-y)\, dy \;=\; \varepsilon \cdot 1$$

This holds for all $\varepsilon > 0$ and uniformly in x, so the integrals approach $f(x)$ in the C^o topology, proving the claim. ///

As noted above, to prove the completeness, we need only exhibit an approximate identity made from finite sums of exponentials. A failing of the kernels [56]

$$K_n(x) \;=\; \frac{\sin 2\pi(n + \frac{1}{2})x}{\sin \pi x}$$

that yield the partial sums is that these kernels are not non-negative. Still, the fact (proven above) that a stronger hypothesis of C^1-ness gives a limited result in the direction we want *suggests* that the masses of the

[55] Similarly, to prove a tangible form of the Stone-Weierstraß theorem that polynomials are dense in $C^o(K)$ for compact K in $\mathbb{R}^n$, one exhibits an *approximate identity* consisting of *polynomials*.

[56] These are often called *Dirichlet kernels*.

K_n *do* bunch up near 0. Indeed, the expression for K_n in terms of sines *does* show that these functions are *real*-valued. Thus, a plausible choice for an approximate identity φ_n is

$$\varphi_n \;=\; K_n^2 \times (\text{constant depending on } n)$$

where the constant is chosen to give total mass 1. Computing directly in terms of exponentials,

$$\int_{S^1} K_n(x)^2\,dx \;=\; \int_{S^1} \big(\psi_{-n} + \ldots + \psi_n(x)\big)^2\,dx \;=\; \int_{S^1} \big((2n+1)\cdot\psi_0(x) + (\text{non-trivial})\big)\,dx \;=\; 2n+1$$

where *non-trivial* denotes exponentials ψ_k with $k \neq 0$, all of whose integrals are 0. Thus, take

$$\varphi_n(x) \;=\; K_n(x)^2/(2n+1)$$

By the discussion so far, we have non-negativity and total mass 1. We must show that the mass of the φ_n's bunches up at 0. For this, we revert to the expression for K_n in terms of *sines*, giving, for $\delta \le |x| \le \frac{1}{2}$,

$$\varphi_n(x) \;=\; \frac{\sin^2 \pi n x}{(2n+1)\sin^2 \pi x} \le \frac{1}{(2n+1)\,x^2} \qquad (\text{since } |\sin \pi x| \ge |x| \text{ for } |x| \le \tfrac{1}{2})$$

For x bounded away from 0 we get inequalities such as

$$0 \le \varphi_n(x) \;\le\; \frac{1}{(2n+1)\,x^2} \;\le\; \frac{1}{(2n+1)\,n^{-2/3}} \;\le\; \frac{1}{n^{1/3}} \qquad (\text{for } |x| \ge n^{-1/3})$$

That is, on the part of S^1 covered by $n^{-1/3} \le |x| \le \frac{1}{2}$, we have a sup-norm estimate for φ_n, so

$$\int_{n^{-1/3} \le |x| \le \frac{1}{2}} \varphi_n(x)\,dx \;\le\; 2 \cdot \frac{1}{2 - n^{-1/3}} \cdot \frac{1}{n^{1/3}} \;\le\; \frac{2}{n^{1/3}}$$

Thus, given $\varepsilon > 0$ and $\delta > 0$, take n large enough so that $n > (2\varepsilon)^3$, and $n > \delta^3$ to meet the criterion.
///

Therefore, having shown that the exponential functions form an orthonormal basis, for $f \in L^2[0,1]$ we have an L^2-equality

$$f \;=\; \sum_n \langle f, \psi_n\rangle\,\psi_n \qquad (\text{in } L^2(S^1))$$

and *Plancherel's theorem* from general Hilbert space theory gives

$$|f|^2 \;=\; \sum_n |\langle f, \psi_n\rangle|^2$$

[7.0.5] Remark: Again, L^2 convergence says *nothing* directly about *pointwise* convergence. Nor is there anything to deny the possibility that a Fourier series *does converge* at a point, but converges to a value *different* from the value of f there. For example, we proved pointwise convergence of C^1 functions, but what about C^o?

[7.0.6] Remark: Since L^2 functions are only defined almost everywhere *anyway*, pointwise convergence would distinguish a special function in the equivalence class in $L^2[0,1]$, which would be suspicious.

8. *Sobolev inequalities, Sobolev imbeddings*

The simplest L^2 theory of Fourier series does *not* address continuity and differentiability. 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*. Sobolev succeeded in making a useful comparison. [57] The idea is to compare C^k norms to norms coming from adroitly contrived Hilbert spaces (the Sobolev spaces).

First, we have an easy estimate for the C^k norms, where we use the variant *sup* version of the C^k norm, rather than the *sum*,

$$\Big|\sum_{|n|\le N} c_n\,\psi_n\Big|_{C^k} \;=\; \sup_{0\le j\le k}\,\sup_x\Big|(2\pi i)^j\sum_{|n|\le N} c_n\,n^j\,\psi_n(x)\Big|_{\mathbb{C}} \;\le\; (2\pi)^k\sum_{|n|\le N}|c_n|\cdot(1+n^2)^{k/2}$$

all for elementary reasons. [58] Then for any $s\in\mathbb{R}$

$$\Big|\sum_{|n|\le N} c_n\,\psi_n\Big|_{C^k} \;\le\; (2\pi)^k\sum_{|n|\le N}|c_n|\cdot(1+n^2)^{s/2}\cdot\frac{1}{(1+n^2)^{(s-k)/2}}$$

$$\le\; (2\pi)^k\Big(\sum_{|n|\le N}|c_n|^2\cdot(1+n^2)^s\Big)^{1/2}\cdot\Big(\sum_{|n|\le N}\frac{1}{(1+n^2)^{s-k}}\Big)^{1/2}$$

by the Cauchy-Schwarz-Bunyakowsky inequality. Now

$$\sum_{n\in\mathbb{Z}}\frac{1}{(1+n^2)^{s-k}} \;<\; +\infty \qquad\qquad (\text{for } s>k+\tfrac{1}{2})$$

so, for any $s>k+\frac{1}{2}$ we have a **Sobolev inequality**

$$\Big|\sum_{|n|\le N} c_n\,\psi_n\Big|_{C^k} \;\le\; (2\pi)^k\Big(\sum_{n\in\mathbb{Z}}\frac{1}{(1+n^2)^{s-k}}\Big)^{1/2}\cdot\Big(\sum_{|n|\le N}|c_n|^2\cdot(1+n^2)^s\Big)^{1/2}$$

$$\le\; (2\pi)^k\Big(\sum_{n\in\mathbb{Z}}\frac{1}{(1+n^2)^{s-k}}\Big)^{1/2}\cdot\Big(\sum_{n\in\mathbb{Z}}|c_n|^2\cdot(1+n^2)^s\Big)^{1/2}$$

Thus,

$$\Big|\sum_{n\in\mathbb{Z}} c_n\,\psi_n\Big|_{C^k} \;\le\; (2\pi)^k\Big(\sum_{n\in\mathbb{Z}}\frac{1}{(1+n^2)^{s-k}}\Big)^{1/2}\cdot\Big(\sum_{n\in\mathbb{Z}}|c_n|^2\cdot(1+n^2)^s\Big)^{1/2}$$

[57] Sobolev's work dealt with more complicated situations, but the germ of the idea is illustrated by Fourier series. Further, there are L^p versions of Sobolev results, which for $p\neq 2$ are significantly more complicated, but less important for our program.

[58] The awkward expression $(1+n^2)^{1/2}$ can be read as being approximately n. However, for $n=0$ we cannot divide by n, and replacing n by $(1+n^2)^{1/2}$ is the traditional stunt to avoid this small trouble.

Existence of this comparison makes the right side interesting. Taking away from the right-hand side the constant

$$\omega_{s-k} \;=\; (2\pi)^k \left(\sum_{n\in\mathbb{Z}} \frac{1}{(1+n^2)^{s-k}} \right)^{1/2}$$

that does not depend upon the coefficients c_n gives the s^{th} **Sobolev norm**

$$s^{th} \text{ Sobolev norm} \;=\; \Big|\sum_{n\in\mathbb{Z}} c_n\,\psi_n\Big|_{H_s} \;=\; \left(\sum_{n\in\mathbb{Z}} |c_n|^2\cdot(1+n^2)^s\right)^{1/2}$$

Paraphrasing, we have proven that

$$|\;|_{C^k} \;\le\; \omega_{s-k}\cdot|\;|_{H_s} \qquad\qquad (\text{for any } s > k+\tfrac{1}{2})$$

Therefore, for $s > k+\frac{1}{2}$, whenever we have a finite Sobolev norm

$$|\sum_n c_n\,\psi_n|_{H_s} = \left(\sum_{n\in\mathbb{Z}} |c_n|^2\cdot(1+n^2)^s\right)^{1/2} \;<\; +\infty$$

then

$$\sum_n c_n\,\psi_n \;=\; C^k \text{ function on } S^1$$

because the partial sums are *Cauchy*[59] in H_s, therefore Cauchy in C^k, which is a Banach space, so *complete.*

For $s\ge 0$, the s^{th} **Sobolev space** is[60]

$$H_s(S^1) \;=\; \{f\in L^2(S^1) : \sum_n |\langle f,\psi_n\rangle|^2\cdot(1+n^2)^s < +\infty\}$$

The inner product on $H_s(S^1)$ is

$$\langle \sum_n a_n\,\psi_n,\; \sum_n b_n\,\psi_n\rangle \;=\; \sum_n a_n\,\overline{b}_n\,(1+n^2)^s$$

[8.0.1] Remark: It is clear that the exponentials ψ_n are an *orthogonal* basis for $H_s(S^1)$, though not *orthonormal* unless $s=0$. In particular, the collection of finite linear combinations of exponentials is *dense* in $H_s(S^1)$.

[8.0.2] Remark: This definition of $H_s(S^1)$ defines a useful space of functions or generalized functions only for $s\ge 0$, since for $s<0$ the constraint $f\in L^2(S^1)$ is stronger (from the Plancherel theorem) than the condition defining $H_s(S^1)$ in the previous display.

[8.0.3] Claim: The s^{th} Sobolev space $H_s(S^1)$ (with $0\le s\in\mathbb{R}$) is a Hilbert space. In particular, the sequences of Fourier coefficients of functions in $H_s(S^1)$ are *all* two-sided sequences $\{c_n : n\in\mathbb{Z}\}$ of complex numbers meeting the condition

$$\sum_n |c_n|^2\cdot(1+n^2)^s \;<\; +\infty$$

[59] The *definition* of convergence of an infinite sum is that the sequence of the partial sums is Cauchy. An infinite sum of *positive* terms that is *bounded above* must converge, because the sequence of partial sums is *increasing.*

[60] 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 $L^2(S^1)$ contains $C^o(S^1)$ and all the $C^k(S^1)$'s.

[8.0.4] Remark: Again, we do want to *define* these positively-indexed 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 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 $H_s(S^1)$ 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 $s \geq 0$,

$$\sum_n |c_n|^2 \;<\; +\infty$$

and, by Plancherel,

$$\sum_n c_n\, \psi_n \;\in\; L^2(S^1)$$

This shows that $H_s(S^1)$ is a Hilbert space for $s \geq 0$. ///

[8.0.5] Corollary: *(Of Sobolev inequality)* For $s > k + \frac{1}{2}$, we have a continuous inclusion

$$H_s(S^1) \;\subset\; C^k(S^1)$$

Proof: Use the Sobolev inequality $|f|_{C^k} \leq \omega \cdot |f|_{H_s}$ applied 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 $H_s(S^1)$, has a sequence of partial sums that converge in $C^k(S^1)$. That is, by the completeness of $C^k(S^1)$, 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. ///

The following comparison in the other direction is elementary.

[8.0.6] Claim: For $s \geq 0$, for $k > s + \frac{1}{2}$,

$$C^k(S^1) \;\subset\; H_s(S^1)$$

Proof: Integrating by parts k times, for $f \in C^k(S^1)$

$$|\langle f^{(k)}, \psi_n\rangle| \;=\; |\langle f, (2\pi i n)^k \psi_n\rangle| \;=\; (2\pi)^k \cdot |n|^k \cdot |\langle f, \psi_n\rangle|$$

The Fourier coefficients of $f^{(k)}$ are at least *bounded*, so

$$|n|^k \cdot |\langle f, \psi_n\rangle| \;=\; \text{bounded}$$

and there is a bound B such that

$$|1+n^2|^k \cdot |\langle f, \psi_n\rangle|^2 \;\leq\; B$$

Thus,

$$\sum_n |\langle f, \psi_n\rangle|^2 \cdot |1+n^2|^s \;=\; \sum_n |1+n^2|^k \cdot |\langle f, \psi_n\rangle|^2 \cdot \frac{1}{|1+n^2|^{k-s}} \;\leq\; B \cdot \sum_n \frac{1}{|1+n^2|^{k-s}}$$

which is finite for $k - s > \frac{1}{2}$. ///

[8.0.7] Corollary: We have, for example, continuous inclusions

$$C^{k+2}(S^1) \subset H_{k+1}(S^1) \subset C^k(S^1)$$

by combining the Sobolev imbedding and the previous elementary result. ///

[8.0.8] Remark: Apart from having the virtue of giving inner-product structures, the expressions appearing in these Sobolev norms are *natural* insofar as they have meaning in terms of L^2-norms of derivatives. For $f = \sum c_n \psi_n \in C^k(S^1)$, by Plancherel

$$\text{(norm via derivatives)} = |f|^2 + |f'|^2 + |f''|^2 + \ldots + |f^{(k)}|^2$$

$$= \sum_n |c_n|^2 \cdot (1 + (2\pi n)^2 + (2\pi n)^4 + \ldots + (2\pi n)^{2k}) \leq (2\pi)^{2k} \sum_n |c_n|^2 \cdot (1+n^2)^k$$

Conversely,

$$(1+n^2)^k \leq C_k \cdot (1 + n^2 + n^4 + n^6 + \ldots + n^{2k}) \qquad \text{(for some constant } C_k)$$

so

$$\text{(norm via Fourier coefficients)} = \sum_n |c_n|^2 \cdot (1+n^2)^k$$

$$\leq C_k \cdot \Big(|f|^2 + |f'|^2 + |f''|^2 + \ldots + |f^{(k)}|^2 \Big)$$

Thus, the two definitions of 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.

[8.0.9] Remark: The interested reader might work out the corresponding inequalities for Fourier series in several variables, aiming at 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 Sobolev theory.

[8.0.10] Remark: It is important to realize that even on S^1, *extra* L^2 differentiability[61] is needed to assure comparable *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 show that the gap widens as the dimension grows.*

---

9. $C^\infty = \lim C^k = \lim H_s = H_\infty$

For our larger purposes, the specific comparisons of indices in the containments

$$H_s(S^1) \subset C^k(S^1) \quad (\text{for } s > k + \tfrac{1}{2})$$

$$C^k(S^1) \subset H_s(S^1) \quad (\text{for } k > s + \tfrac{1}{2})$$

are not the point, since we are more interested in *smooth functions* $C^\infty(S^1)$ than functions with *limited* continuous differentiability.

Thus, the point is that the Sobolev spaces and $C^k(S^1)$ spaces are *cofinal* under taking *descending intersections*. That is, letting $H_\infty(S^1)$ be the *intersection* of all the $H_s(S^1)$, as *sets* we have

$$C^\infty(S^1) = \bigcap_k C^k(S^1) = \bigcap_{s\geq 0} H_s(S^1) = H_\infty(S^1)$$

[61] Since L^2 functions do not *a priori* have pointwise values, a discussion of L^2 differentiability is considerably more delicate than classical differentiation.

Realizing that descending nested intersections are *limits*, we have a way to keep track of the topologies effortlessly, and have

[9.0.1] **Theorem**: As topological vector spaces

$$C^\infty(S^1) \;=\; \lim_k C^k(S^1) \;=\; \lim_{s\geq 0} H_s(S^1) \;=\; H_\infty(S^1)$$

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,

[9.0.2] **Corollary**: The space of distributions on S^1 is

$$C^\infty(S^1)^* \;=\; \mathrm{colim}_k C^k(S^1)^* \;=\; \mathrm{colim}_{s\geq 0} H_s(S^1)^* \;=\; H_\infty(S^1)^*$$

(and the duals $H_s(S^1)^*$ admit further explication, below). ///

Expressing $C^\infty(S^1)$ as a limit of the Hilbert spaces $H_s(S^1)$, as opposed to its more natural expression as a limit of the Banach spaces $C^k(S^1)$, is convenient when taking *duals*, since by the *Riesz-Fischer theorem* [62] we have explicit expressions for Hilbert space duals. We exploit this possibility below.

---

10. *Distributions, generalized functions, again*

One reason to discuss *distributions* or *generalized functions* in the context of Fourier series is that the Fourier coefficients afford a tangible descriptive mechanism. [63]

Since the exponential functions ψ_n are in $C^\infty(S^1)$, for any distribution u we can compute Fourier coefficients by

$$(n^{th} \text{ Fourier coefficient of } u) \;=\; \widehat{u}(n) \;=\; u(\psi_{-n})$$

Write

$$u \;\sim\; \sum_n \widehat{u}(n)\cdot\psi_n$$

even though pointwise convergence of the indicated sum is not expected. Define Sobolev spaces for *all* real s by

$$H_s(S^1) \;=\; \{u \in C^\infty(S^1)^* \;:\; \sum_n |u(\psi_{-n})|^2\cdot(1+n^2)^s < \infty\}$$

and the s^{th} **Sobolev norm** $|u|_{H_s}$ is

$$|u|^2_{H_s} \;=\; \sum_n |u(\psi_{-n})|^2\cdot(1+n^2)^s$$

---

[62] 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.

[63] The analogous discussion of distributions on the real line $\mathbb{R}$ is more complicated, due to the non-compactness of $\mathbb{R}$, and, thus, for example, to the fact that *not* every distribution is the Fourier transform of a *function*. In fact, the larger set of distributions which *admit* Fourier transforms, called *tempered* distributions, is a proper subset of all distributions on $\mathbb{R}$. By contrast, every distribution on the circle *does* have a Fourier series expansion.

For $0 \le s \in \mathbb{Z}$, this definition is visibly compatible with the previous definition via derivatives.

[10.0.1] Remark: The formation of the 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 $C^k(S^1)$ 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 [64]

$$\ell^{2,s} \;=\; \{\{c_n : n \in \mathbb{Z}\} \;:\; \sum_{n\in\mathbb{Z}} |c_n|^2 \cdot (1+n^2)^s < \infty\}$$

with the weighted version of the usual hermitian inner product, namely,

$$\langle \{c_n\},\, \{d_n\}\rangle \;=\; \sum_{n\in\mathbb{Z}} c_n\, \overline{d_n} \cdot (1+n^2)^s$$

[10.0.2] 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

$$\ell^{2,-s} \;\longrightarrow\; (\ell^{2,s})^* \quad \text{by} \quad \{d_n\} \to \lambda_{\{d_n\}} \quad \text{where} \quad \lambda_{\{d_n\}}(\{c_n\}) = \sum_n c_n\, d_{-n}$$

[10.0.3] 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

$$|\sum_n c_n \cdot d_{-n}| \;\le\; \sum_n |c_n|\,(1+n^2)^{s/2} \cdot |d_{-n}|\,(1+n^2)^{-s/2}$$

$$\le\; \left(\sum_n |c_n|^2\,(1+n^2)^s\right)^{1/2} \cdot \left(\sum_n |d_n|^2\,(1+n^2)^{-s}\right)^{1/2} \;=\; |\{c_n\}|_{\ell^{2,s}} \cdot |\{d_n\}|_{\ell^{2,-s}}$$

proving the continuity. [65] 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

$$d_n \;=\; \overline{a}_{-n} \cdot (1+n^2)^s$$

[64] The Hilbert space ℓ^2 is the collection of sequences $\{a_1, a_2, \ldots\}$ with $\sum |a_i|^2 < \infty$.

[65] Continuity of a linear functional λ on a Banach space B is equivalent to an estimate $|\lambda(b)|_{\mathbb{C}} \le C \cdot |b|_B$ for some constant C.

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 $\ell^{2,s}$. ///

[10.0.4] Claim: The map $u \to \{\widehat{u}(n)\}$ on $H_s(S^1)$ by taking Fourier coefficients is a Hilbert-space isomorphism

$$H_s(S^1) \;\approx\; \ell^{2,s}$$

Proof: That the two-sided sequence of Fourier coefficients $u(\psi_{-n})$ is in $\ell^{2,s}$ is part of the definition of $H_s(S^1)$. The more serious question is *surjectivity*.

Let $\{c_n\} \in \ell^{2,s}$. For $s \geq 0$, the s^{th} Sobolev norm dominates the 0^{th}, so distributions in $H_s(S^1)$ are at least $L^2(S^1)$-functions. The definition of $H_s(S^1)$ in this case makes $H_s(S^1)$ a Hilbert space, and we directly invoke the Plancherel theorem, using the orthonormal basis $\{\psi_n \cdot (1+n^2)^{-s/2}\}$ for $H_s(S^1)$. This gives the surjectivity $H_s(S^1) \to \ell^{2,s}$ 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 $H_s(S^1)$, by

$$u(f) \;=\; \sum_n \widehat{f}(n) \cdot c_{-n} \qquad\qquad (f \in C^\infty(S^1))$$

By the Cauchy-Schwarz-Bunyakowsky inequality

$$|\sum_n \widehat{f}(n) \cdot c_{-n}| \;\leq\; \sum_n |\widehat{f}(n)|\,(1+n^2)^{-s/2} \cdot |c_n|\,(1+n^2)^{s/2}$$

$$\leq \left(\sum_n |\widehat{f}(n)|^2\,(1+n^2)^{-s}\right)^{1/2} \cdot \left(\sum_n |c_n|^2\,(1+n^2)^{s}\right)^{1/2} \;=\; |f|_{H_{-s}} \cdot |\{c_n\}|_{\ell^{2,s}}$$

This shows that u is a continuous linear functional on $H_{-s}(S^1)$. For $s < 0$, the test functions $C^\infty(S^1)$ imbed continuously into $H_{-s}(S^1)$, so u gives a continuous functional on $C^\infty(S^1)$, so is a distribution. This proves that the Fourier coefficient map is a surjection to $\ell^{2,s}$ for $s < 0$. ///

[10.0.5] Remark: After this preparation, the remainder of this section is completely unsurprising. The following corollary is the conceptual point of this story.

[10.0.6] 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

$$\lambda_u(f) \;=\; \langle f, u\rangle \qquad\qquad (\text{for } f \in H_s(S^1))$$

[10.0.7] 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 $H_s(S^1) \approx \ell^{2,s}$ and $H_{-s}(S^1) \approx \ell^{2,-s}$ with the pairing of $\ell^{2,s}$ and $\ell^{2,-s}$ given just above. ///

[10.0.8] Corollary: The space of all distributions on S^1 is

$$\text{distributions} \;=\; C^\infty(S^1)^* \;=\; \bigcup_{s\geq 0} H_s(S^1)^* \;=\; \bigcup_{s\geq 0} H_{-s}(S^1) \;=\; \text{colim}_{s\geq 0} H_{-s}(S^1)$$

thus expressing $C^\infty(S^1)^*$ as an ascending union of Hilbert spaces. ///

[10.0.9] Corollary: A distribution $u \sim \sum_n c_n\, \psi_n$ can be evaluated on $f \in C^\infty(S^1)$ by

$$u(f) \;=\; \sum_n \widehat{f}(n) \cdot \widehat{u}(-n)$$

Proof: The thing that needs to be said is that u lies in some $H_{-s}(S^1)$, so gives a continuous functional on $H_s(S^1)$, which contains $C^\infty(S^1)$. Then 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

$$|c_n| \;\leq\; C \cdot (1+n^2)^N \qquad \text{(for all } n \in \mathbb{Z})$$

[10.0.10] 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 $|c_n| \leq C \cdot (1+n^2)^N$,

$$\sum_n |c_n|^2 \cdot (1+n^2)^{-(2N+1)} \;\leq\; \sum_n C^2 \cdot (1+n^2)^{2N} \cdot (1+n^2)^{-(2N+1)} \;=\; C^2 \cdot \sum_n (1+n^2)^{-1} \;<\; \infty$$

That is, from the previous discussion, the sequence gives an element of $H_{-(2N+1)}(S^1) \subset C^\infty(S^1)^*$. ///

[10.0.11] Corollary: For $u \sim \sum_n c_n\, \psi_n \in H_s(S^1)$ the derivative (for any $s \in \mathbb{R}$) is

$$u' \;\sim\; (-2\pi i) \cdot \sum_n n \cdot c_n \cdot \psi_n \;\in\; H_{s-1}$$

Proof: Invoke the definition (compatible with integration by parts) of the derivative of distributions to compute (for $f \in C^\infty(S^1)$)

$$u'(f) \;=\; -u(f') \;=\; -\sum_n \widehat{f'}(n) \cdot \widehat{u}(n) \;=\; -(-2\pi i) \sum_n n\, \widehat{f}(n) \cdot \widehat{u}(-n) \;=\; (-2\pi i) \sum_n \widehat{f}(n) \cdot (-n)\, \widehat{u}(-n)$$

as claimed. The Fourier coefficients $(-2\pi i) \cdot n \cdot \widehat{u}(n)$ do satisfy

$$\sum_n |(-2\pi i)\, n\, \widehat{u}(n)|^2 \cdot (1+n^2)^{s-1} \;\; \leq (2\pi)^2 \sum_n (1+n^2)\, |\widehat{u}(n)|^2 \cdot (1+n^2)^{s-1}$$

$$= (2\pi)^2 \sum_n |\widehat{u}(n)|^2 \cdot (1+n^2)^s \;=\; (2\pi)^2\, |u|^2_{H_s} < \infty$$

which proves that the differentiation maps H_s to H_{s-1}. ///

[10.0.12] 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.

[10.0.13] Corollary: The collection of finite linear combinations of exponentials ψ_n is dense in every $H_s(S^1)$, for $s \in \mathbb{R}$. In particular, $C^\infty(S^1)$ is dense in every $H_s(S^1)$, for $s \in \mathbb{R}$.

Proof: The exponentials are an orthogonal basis for every Sobolev space. ///

[10.0.14] 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.

---

11. *The confusing 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** [66]

$$s(x) \;=\; x - \frac{1}{2} \qquad\qquad (\text{for } 0 \le x < 1)$$

made *periodic* by demanding $s(x+n) = s(x)$ for all $n \in \mathbb{Z}$, so

$$s(x) \;=\; x - [\![x]\!] - \frac{1}{2} \qquad\qquad (\text{for } x \in \mathbb{R})$$

where $[\![x]\!]$ is the greatest integer less than or equal x. Away from integers, this function is infinitely differentiable, with derivative 1. At integers it jumps down from value to $1/2$ to value $-1/2$. We do not attempt to define a value *at* integers.

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 $\psi_n(x) = e^{-2\pi i n x}$ by integrating by parts

$$\int_0^1 s(x)\cdot e^{-2\pi i n x}\,dx \;=\; \begin{cases} \frac{1}{-2\pi i n} & (\text{for } n \neq 0) \\ \\ 0 & (\text{for } n = 0) \end{cases}$$

Thus, *at least* as a distribution, its Fourier expansion is

$$s(x) \;=\; \frac{1}{-2\pi i} \sum_{n \neq 0} \frac{1}{n} \cdot e^{2\pi i n x}$$

---

[66] As earlier, the subtraction of $1/2$ is convenient, since then the integral of $s(x)$ is 0.

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,

$$\begin{aligned} s'(x) &= -\textstyle\sum_{n\neq 0} e^{2\pi inx} \\ s''(x) &= -2\pi i \textstyle\sum_{n\neq 0} n\, e^{2\pi inx} \\ &\ldots \\ s^{(k)}(x) &= -(2\pi i)^{k-1} \textstyle\sum_{n\neq 0} n^{k-1}\, e^{2\pi inx} \end{aligned}$$

and so on, just as successive derivatives of smooth functions $f(x) = \sum_n c_n\, e^{2\pi inx}$ are obtained by termwise differentiation

$$f^{(k)}(x) \;=\; (2\pi i)^k \sum_{n\neq 0} n^k\, c_n\, e^{2\pi inx}$$

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 $f(x) \sim \sum_n c_n\, e^{2\pi inx}$, the imagined integral of f against $s^{(k)}$ should be expressible as the sum of products of Fourier coefficients

$$\text{imagined } \langle f, s^{(k)}\rangle \;=\; \sum_{n\neq 0} c_n \cdot \left(\frac{(2\pi in)^k}{-2\pi in}\right)^{\text{conj}}$$

(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,

$$\sum_{n\neq 0} c_n \cdot \left(\frac{(2\pi in)^k}{-2\pi in}\right)^{\text{conj}} \;=\; (-1)^k \sum_{n\neq 0} (2\pi in)^k\, c_n \cdot \left(\frac{1}{-2\pi in}\right)^{\text{conj}} \;=\; (-1)^k \int_{S^1} f^{(k)}(x)\, \overline{s}(x)\, dx$$

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

$$(-1)^k \int_{S^1} f^{(k)}(x)\, \overline{s}(x)\, dx \;=\; (-1)^k u(f^{(k)}) \;=\; u^{(k)}(f)$$

It is in this sense that the sum $\sum_{n\neq 0} c_n \cdot \frac{(2\pi in)^k}{-2\pi in}$ 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, and was not the point.

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*.

12. *Appendix: products and limits of topological vector spaces*

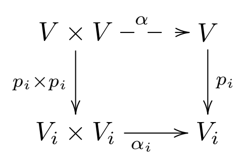
Here we carry out the diagrammatical proof that products and limits of topological vector spaces *exist*, and are locally convex when the factors or limitands are locally convex. The arguments are similar to those in proving that products and limits of topological *groups* exist. Nothing surprising happens.

[12.0.1] 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.

[12.0.2] 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. [67]

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

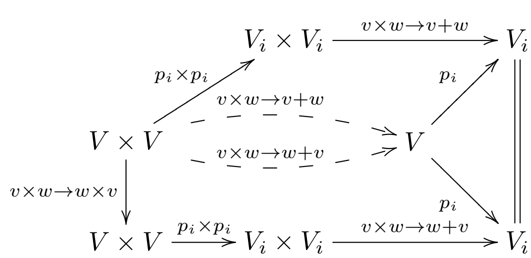


This defines what we will show to be a *vector addition* on V. Similarly, the scalar multiplications $s_i : \mathbb{C} \times V_i \to V_i$ composed with the projections $p_i : V \to V_i$ give a family of maps

$$s_i \circ (1 \times p_i) \; : \; \mathbb{C} \times V \;\longrightarrow\; V_i$$

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



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

[67] 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.

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.

[12.0.3] 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. ///

---

13. *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. [68] Thus, for example,

$$C^\infty(S^1) \;=\; \bigcap_k C^k(S^1) \;=\; \lim_k C^k(S^1)$$

is a Fréchet space, *by definition*.

Despite its advantages, the present definition is not the usual one. [69] We make a comparison, and elaborate on the features of Fréchet spaces.

[68] Of course, it suffices that a limit have a countable cofinal subfamily.

[69] 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

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.

[13.0.1] Claim: A Fréchet space is locally convex and complete (invariantly) metrizable. [70]

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,

[13.0.2] Theorem: Every locally convex topological vector space is a *subspace* of a limit of Banach spaces (and vice-versa).

[13.0.3] 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

$$\begin{array}{rll} |\alpha \cdot v| = |\alpha|_{\mathbb{C}} \cdot |v| & & \text{(homogeneity)} \\ |v+w| \le |v| + |w| & & \text{(triangle inequality)} \\ |v| \ge 0, & \text{(equality } \textit{only} \text{ for } v=0) & \text{(definiteness)} \end{array}$$

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

$$\begin{array}{rl} p(\alpha \cdot v) = |\alpha|_{\mathbb{C}} \cdot p(v) & \text{(homogeneity)} \\ p(v+w) \le p(v) + p(w) & \text{(triangle inequality)} \\ p(v) \ge 0 & \text{(semi-definiteness)} \end{array}$$

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 $p(v) > 0$.

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.

[70] 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.

A separating family Φ of semi-norms on a complex vector space V gives a *locally convex* topology by taking as local sub-basis [71] 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.

[13.0.4] 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. [72]

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

$$\alpha \cdot w \in U \quad \text{(for all } \alpha \in W \text{ and } w \in U')$$

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,

$$v \in t \cdot U \qquad \text{(for all } t > \varepsilon^{-1})$$

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

$$\alpha \cdot U = U \qquad \text{(for } \alpha \in \mathbb{C} \text{ with } |\alpha| = 1)$$

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.

[71] 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.

[72] 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.

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

$$U' \;=\; \{\alpha \cdot w :\; |\alpha| \le \frac{\varepsilon}{2},\; w \in W\} \;=\; \bigcup_{|\alpha| \le \varepsilon/2} \alpha \cdot W$$

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 $t \ge t_1$ and $w \in t \cdot U$ for $t \ge t_2$. 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

$$p_U(v + w) \;\le\; p_U(v) + p_U(w)$$

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 $i \ge j$ when $i \supset j$. For $i > j$ we have

$$p_i(v) \;=\; \sum_{p \in i} p(v) \;\ge\; \sum_{p \in j} p(v) = p_j(v)$$

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

$$\{v \in V : p_U(v) < \varepsilon\} \qquad \text{(for } \varepsilon > 0 \text{ and convex balanced open } U \text{ containing } 0)$$

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. ///

[13.0.5] Remark: We do *not* assert that an arbitrary locally convex topological vector space *is* a limit of Banach spaces. Such an assertion is equivalent to various *completeness* hypotheses, which we will investigate later.

---

14. *Appendix: Urysohn and density of C^o*

Urysohn's lemma is the technical point that allows us to relate *measurable* functions to *continuous* functions. Then, from the Lebesgue definition of *integral*, we can prove the density of continuous functions in L^2 spaces, for example.

[14.0.1] 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 $0 \leq f \leq 1$ 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 $0 \le r \le 1$ 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.* [73] 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, $f(x) = g(x)$. ///

[14.0.2] Corollary: For a topological space X with a regular Borel measure μ, $C_c^o(X)$ is dense in $L^2(X, \mu)$.

Proof: The *regularity* of the measure is the property that $\mu(E)$ is both the *sup* of $\mu(K)$ for compacts $K \subset E$, and is the *inf* of $\mu(U)$ for opens $U \supset E$. From Urysohn's lemma, we have a continuous $f_{K,U}(x)$ which is 1 on K and 0 off U. Let $K_1 \subset K_2 \subset \ldots$ be a sequence of compacts inside E whose measure approaches that of E from *below*, and let $U_1 \supset U_2 \supset \ldots$ be a sequence of opens containing E whose measures approach that of E

[73] 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.

from *above*. Let f_i be a function as in Urysohn's lemma made to be 1 on K_i and 0 off U_i. Then Lebesgue's Dominated convergence theorem implies that

$$f_i \longrightarrow \text{(characteristic function of } E) \qquad\qquad (\text{in } L^2(X,\mu))$$

From the definition of integral of *measurable functions*, finite linear combinations of characteristic functions are dense in L^2 (or any other L^p with $1 \le p < \infty$). Thus, continuous functions are dense. ///
