Math 212a Lecture 2.

Fejer's theorem.
Dirichlet's theorem.
The Riemann Lebesgue lemma.
Basics of Hilbert space.

Shlomo Sternberg

September 4, 2014

I will begin today's lecture with Fejer's theorem which asserts that the Fourier series of a continuous function f converges to f in the sense of Cesaro. Then I will present Dirichlet's theorem which asserts that if f is piecewise differentiable, then its Fourier series actually converges to f at all points of continuity. I then discuss the Riemann-Lebesgue lemma.

But the main mode of convergence we will use in this course is"mean square convergence" (as introduced by Bessel) which gets us into the realm of Hilbert space.

Let f be a continuous periodic function. Let $C(f,n,x)$ denote the n-th Cesaro sum of its Fourier series at x in the sense that if $a_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}dt$ is the n-th Fourier coefficient of f then

$$C(f,0,x) = a_0, \quad C(f,1,x) = \frac{1}{2}\left(a_{-1}e^{-ix} + 2a_0 + a_1e^{ix}\right),$$

$$C(f,2,x) = \frac{1}{3}\left(a_{-2}e^{-2ix} + 2a_{-1}e^{-ix} + 3a_0 + 2a_1e^{ix} + a_2e^{2ix}\right)$$

and, in general,

$$C(f,n,x) = \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} a_r e^{irx}.$$

The Fejer kernel.

$$C(f,n,x) = \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} \frac{1}{2\pi} \left(\int_{-\pi}^{\pi} f(t) e^{-irt} dt \right) e^{irx}$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} e^{ir(x-t)} dt$$

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) K_n(x-t) dt \quad \text{where}$$

$$K_n(s) := \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} e^{irs}. \text{ So, setting } y = x - t,$$

$$C(f,n,x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x-y) K_n(y) dy \text{ by the periodicity of } f \text{ and } K_n.$$

We have shown that

$$C(f,n,x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)K_n(y)dy \quad K_n(s) := \sum_{r=-n}^{n} \frac{n+1-|r|}{n+1} e^{irs}.$$

Now $\quad (e^{-is/2} + e^{is/2})^2 = e^{-is} + 2 + e^{is} = 2K_1(s)$

$$\left(e^{-is} + 1 + e^{is}\right)^2 = e^{-i2s} + 2e^{-is} + 3 + 2e^{is} + e^{2is} = 3K_3(s),$$

and, in general,

$$(n+1)K_n(s) = \left(\sum_{k=0}^{n} e^{i(k-\frac{n}{2})s}\right)^2 = \left(e^{-\frac{ins}{2}} \sum_{k=0}^{n} e^{iks}\right)^2.$$

A closed expression for the Fejer kernel.

The function K_n is called the nth **Fejer kernel**. We have shown that

$$(n+1)K_n(s) = \left(e^{-\frac{ins}{2}} \sum_{k=0}^{n} e^{iks}\right)^2.$$

For $s \neq 0 \bmod 2\pi$ we can sum the geometric sum:

$$\sum_{k=0}^{n} e^{iks} = \frac{1-e^{i(n+1)s}}{1-e^{is}}.$$

Thus $(n+1)K_n(s) =$

$$\left(e^{-\frac{ins}{2}} \frac{1-e^{i(n+1)s}}{1-e^{is}}\right)^2 = \left(\frac{e^{-\frac{i(n+1)s}{2}} - e^{\frac{i(n+1)s}{2}}}{e^{-\frac{is}{2}} - e^{\frac{is}{2}}}\right)^2 = \left(\frac{\sin\frac{(n+1)s}{2}}{\sin\frac{s}{2}}\right)^2.$$

We have shown that for $s \not\equiv 0 \mod 2\pi$

$$K_n(s) = \frac{1}{n+1}\left(\frac{\sin\frac{(n+1)s}{2}}{\sin\frac{s}{2}}\right)^2.$$

Now $K_n(0) = n+1$ from its definition, and this coincides with the limit of the above expression as $s \to 0$ by l'Hôpital's rule.

Properties of the Fejer kernel.

- K_n is continuous and periodic.
- $K_n(s) \geq 0$ for all s.
- $\lim_{n\to\infty} K_n(s) = 0$ for any $s \not\equiv 0 \mod 2\pi$.
- $K_n(0) = n + 1$.
- $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(s)ds = 1$ for all n.

This last item follows from the original definition since all the exponential terms integrate to zero.

Here is a graph of several values of K_n:

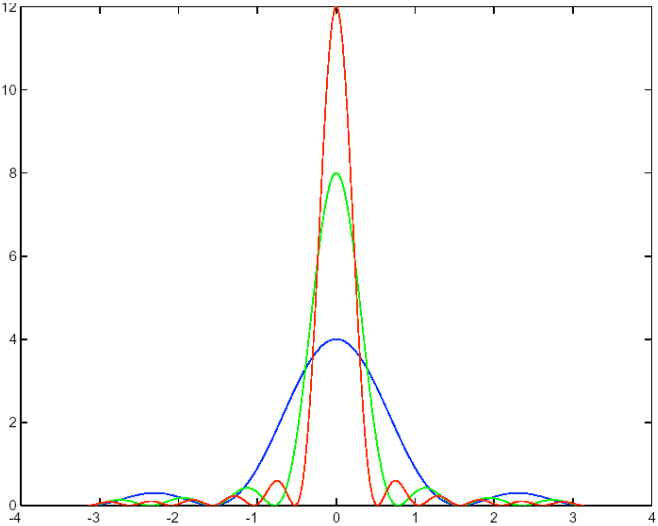


Figure 1: The graphs of K_3, K_7 and K_{11} over $[-\pi, \pi]$.

The idea behind Fejer's theorem.

Fejer's theorem asserts that if f is continuous and periodic then $C(f,n,x) \to f(x)$ as $n \to \infty$. From the preceding slide it would appear that "most of the mass" of the non-negative function $y \mapsto K_n(y)$ is concentrated about the point 0 and that therefore in computing

$$C(f,n,x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)K_n(y)dy$$

we can, approximately, assume that we are integrating over smaller and smaller neighborhoods of 0. The continuity of f guarantees that $f(x-y)$ is close to $f(x)$ on such small neighborhoods of $y = 0$.

Lemma

For any $\delta > 0$ and $\epsilon > 0$ there is an $N = N(\delta, \epsilon)$ such that

$$K_n(s) < \epsilon \quad \text{if } |s - 2\pi r| > \delta \text{ for all integers } r, \text{ and if } n > N.$$

Proof.

In the expression $K_n(s) = \frac{1}{n+1}\left(\frac{\sin\frac{(n+1)s}{2}}{\sin\frac{s}{2}}\right)^2$ the denominator $\sin\frac{s}{2}$ satisfies $|\sin\frac{s}{2}| \geq \sin\delta/2$ while the numerator $\sin\frac{(n+1)s}{2}$ is bounded in absolute value by 1. So for n sufficiently large, $|K_n(s)| < \epsilon$ for this range of s. □

Proof of Fejer's theorem, I.

Since $\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)dy = 1$ and $K_n \geq 0$, $|C(f,n,x) - f(x)| =$

$$\left|\frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-y)K_n(y)dy - \frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)f(x)dy\right|$$

$$\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} |f(x-y) - f(x)|K_n(y)dy = I_1 + I_2$$

where in I_1 the integration is over $|y| \leq \delta$ and in I_2 the integration is over $|y| \geq \delta$. Since f is continuous and periodic, it is bounded, say $|f| \leq M$ and is uniformly continuous. So for any $\epsilon > 0$ we can find a $\delta > 0$ so that (for all x)

$$|f(x) - f(y)| \leq \frac{\epsilon}{2} \quad \text{if } |x-y| \leq \delta.$$

Proof of Fejer's theorem, II.

We have shown that

$$|C(f,n,x)-f(x)| \leq \frac{1}{2\pi}\int_{-\pi}^{\pi}|f(x-y)-f(x)|K_n(y)dy = I_1 + I_2$$

where in I_1 the integration is over $|y| \leq \delta$ and in I_2 the integration is over $|y| \geq \delta$, and have chosen δ so that

$$|f(x)-f(y)| \leq \frac{\epsilon}{2} \quad \text{if } |x-y| \leq \delta.$$

Thus $I_1 \leq \frac{\epsilon}{2}\frac{1}{2\pi}\int_{-\pi}^{\pi} K_n(y)dy = \frac{\epsilon}{2}$. By the Lemma, we can choose $N = N(\delta, \epsilon/4M)$ so that $K_n(y) \leq \frac{\epsilon}{4M}$ if $n > N$ on the range $|y| \geq \delta$.

So $I_2 \leq \frac{\epsilon}{2}$. □

Some consequences of Fejer's theorem.

- The trigonometric polynomials are dense in the space of continuous periodic functions in the uniform topology.
- If two continuous periodic functions have the same Fourier coefficients then they are equal.
- The Weierstrass approximation theorem: Any continuous function on a compact interval can be uniformly approximated by polynomials. Indeed: We can extend the function to be periodic, approximate the extended function by trigonometric polynomials and then use the Taylor expansion of each exponential to approximate by polynomials.

Lipót Fejér

Born: 9 Feb 1880 in Pécs, Hungary
Died: 15 Oct 1959 in Budapest, Hungary

Dirichlet's theorem.

Dirichlet's theorem asserts that the (symmetric sums of the) Fourier series of a piecewise differentiable function f converges at all x to $\frac{1}{2}(f(x_+) + f(x_-))$. Let

$$s_n(f,x) := \sum_{-n}^{n} a_k e^{ikx} \quad a_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-ikt}dt.$$

Write s_n as $s_n(f,x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_n(x-t)dt$ where the **Dirichlet kernel** D_n is defined as

$$D_n(t) := \sum_{-n}^{n} e^{ikt}.$$

Evaluating the Dirichlet kernel.

For $t \not\equiv 0 \bmod 2\pi$, we have $D_n(t) = \sum_{-n}^{n} e^{ikt} = e^{-int} \sum_0^{2n} e^{ikt}$, and we can evaluate the geometric sum to obtain

$$\begin{aligned} D_n(t) &= e^{-int} \cdot \frac{1 - e^{(2n+1)it}}{1 - e^{it}} \\ &= \frac{e^{-int} - e^{(n+1)it}}{1 - e^{it}} \\ &= \frac{e^{-i(n+\frac{1}{2})t} - e^{i(n+\frac{1}{2})t}}{e^{-i\frac{1}{2}t} - e^{i\frac{1}{2}t}} \\ &= \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}. \end{aligned}$$

$D_n(0) = 2n+1$ which is the limit of the above expression as $t \to 0$.

So D_n is continuous, and from the definition, $\int_{-\pi}^{\pi} D_n(t)dt = 1$.

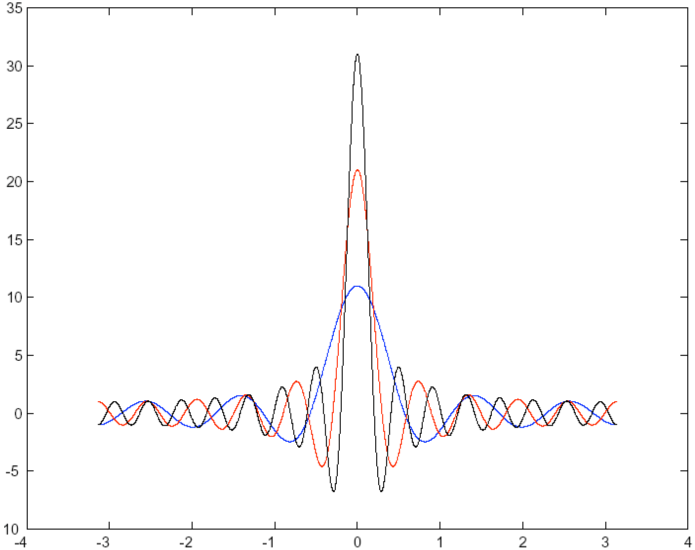
As n increases,

$$D_n(t) = \frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}$$

becomes more and more oscillatory outside any fixed interval about 0 (on $[-\pi, \pi]$). But, in contrast to the Fejer kernel, its amplitude does not tend to zero there. See the next slide where D_n is plotted for $n = 5, 10, 15$. The issue of the convergence of

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_n(x-t)dt \to f(x)$$

is more subtle than the convergence in Fejer's theorem.



The Dirichlet kernel D_k for k= 5,10,15

Although I could give a direct proof of Dirichlet's theorem involving nothing more than integration by parts, I want to put the proof in a more general context.

I will consider functions on $\mathbb{R}$ which are "integrable" and have the property that for any $\epsilon > 0$ there is a step function g such that

$$\int_{\mathbb{R}} |f - g| dx < \epsilon.$$

This definition depends on the meaning of the word "integrable". Eventually, when we study the Lebesgue integral, we will find that all functions which are Lebesgue integrable will have this property. Clearly any f which is piecewise continuous and vanishes outside a finite interval has this property. If f is only defined on some interval $[a, b]$ and has this property there, we just extend f by declaring it to be zero outside $[a, b]$ and the extended function still has this property.

The averaging condition.

For the present, the Riemann integral will do. We will denote our class of functions by $L_1(\mathbb{R})$.

A bounded integrable function h is said to satisfy the **averaging condition** if

$$\lim_{c\to\infty} \frac{1}{c}\int_0^c h(t)dt = 0.$$

For example, $h(t) = \sin t$ satisfies the averaging condition.

The Riemann-Lebesgue lemma.

Theorem

[The Riemann-Lebesgue lemma.] *If* $f \in L_1(\mathbb{R})$ *and* h *satisfies the averaging condition, then*

$$lim_{\omega \to \infty} \int_a^b f(t)h(\omega t)dt = 0$$

for any interval $[a, b]$.

Clearly it is enough to prove this for $a = 0, \ b = \infty$.

Proof.

If $f = \mathbf{1}_{[c,d]}$, where $0 \leq c \leq d < \infty$ then $\int_0^\infty f(t)h(\omega t)dt$

$$= \int_c^d h(\omega t)dt = \frac{1}{\omega}\int_0^{d\omega} h(x)dx - \frac{1}{\omega}\int_0^{c\omega} h(x)dx \to 0.$$

By linearity, the theorem is true for step functions. Choose C such that $|h(x)| \leq C$ for all x. Let $f \in L_1(\mathbb{R})$. Choose a step function g such that $\int_{\mathbb{R}} |f-g|dx \leq \frac{\epsilon}{2C}$. Choose Ω such that for all $\omega > \Omega$ $\left|\int_0^\infty g(t)h(\omega t)dt\right| < \frac{\epsilon}{2}$. Then for $\omega > \Omega$ we have

$$\begin{aligned}\left|\int_0^\infty f(t)h(\omega t)d\right| &\leq \int_0^\infty |f(t)-g(t)||h(\omega t)|dt + \left|\int_0^\infty g(t)h(\omega t)dt\right| \\ &< \frac{\epsilon}{2C}\cdot C + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

Applying Riemann-Lebesgue to Dirichlet's theorem

Proposition

If $f \in L_1([0,\pi])$ *and* $0 < r < \pi$ *then*

$$\lim_{n\to\infty} \int_r^\pi f(t)D_n(t)dt = 0.$$

Proof.

On $[r,\pi]$ the denominator $\sin \frac{t}{2}$ of $D_n(t)$ is bounded from below so $t \mapsto g(t) := \frac{f(t)}{\sin \frac{t}{2}} \in L_1$ and so by R-L

$$\int_r^\pi f(t)D_n(t)dt = \int_r^\pi g(t)\sin\left(n+\frac{1}{2}\right)tdt \to 0. \quad \square$$

The Fourier kernel.

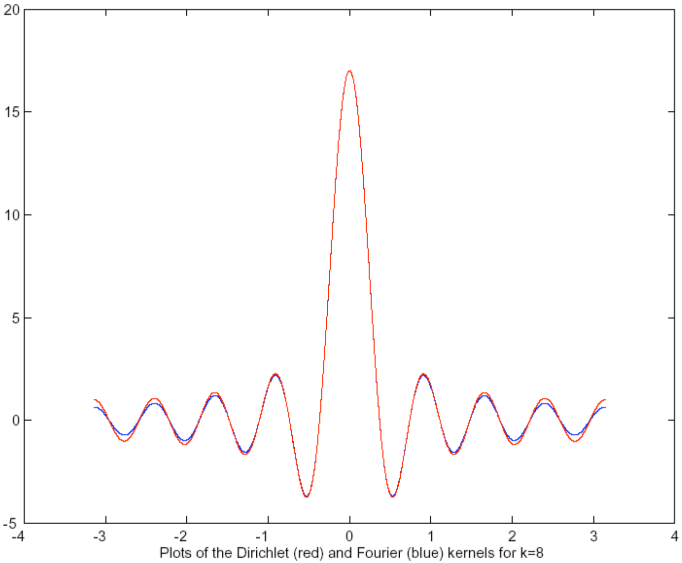
The **Fourier kernel** F_n is defined as

$$F_n(t) := \frac{\sin\left(n+\frac{1}{2}\right)t}{\frac{t}{2}}.$$

the same argument shows that $\lim_{n\to\infty}\int_r^\pi f(t)F_n(t)dt = 0.$

On the interval $[-2,2]$ the Fourier kernel is very close to the Dirichlet kernel for large n. For example, both are plotted on the next slide for $n=8$:

Applying Riemann-Lebesgue to Dirichlet's theorem



Proposition

Let $f \in L_1([0,\pi])$ *and* $0 < r < \pi$. *If one or the other of the limits*

$$\lim_{n\to\infty} \int_0^r f(t)D_n(t)dt \quad or \quad \lim_{n\to\infty} \int_0^r f(t)F_n(t)dt$$

exists, then both limits exist and are equal.

Proof.

$$\lim_{t\to 0}\left(\frac{1}{\sin t}-\frac{1}{t}\right)=\lim_{t=0}\frac{\frac{1}{3!}t^3-\cdots}{t^2-\cdots}=0.$$

(More formally, apply Hôpital's rule twice.) Set

$$g(t):=\frac{1}{\sin\frac{t}{2}}-\frac{1}{t/2}.$$

So g is continuous at 0 if we set $g(0)=0$. The difference between the two integrals in the Proposition is

$$\int_0^r f(t)g(t)\sin\left(n+\frac{1}{2}\right)tdt$$

which tends to zero by Riemann-Lebesgue. □

Back to Dirichlet's theorem.

We want conditions on f which guarantee that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_n(x-t)dt \to f(x).$$

Now

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)D_n(x-t)dt = \frac{1}{2\pi}\int_{t-\pi}^{t+\pi} f(s+x)D_n(s)ds$$

by the change of variables $s = t - x$ and the fact that D_n is even. Since both f and D_n are periodic, this last integral equals

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} f(s+x)D_n(s)ds = \frac{1}{2\pi}\int_{0}^{\pi} [f(x+s)+f(x-s)]D_n(s)ds.$$

Completion of the proof of Dirichlet's theorem.

We are interested in examining the limiting behavior of

$$\frac{1}{2\pi}\int_0^\pi [f(x+s)+f(x-s)]D_n(s)ds.$$

By the preceding two propositions, this will approach a limit c as $n\to\infty$ if and only if there is some $r>0$ such that

$$\lim_{n\to\infty}\frac{1}{\pi}\int_0^r [f(x+s)+f(x-s)-2c]\frac{\sin\left(n+\frac{1}{2}\right)s}{s}ds\to 0.$$

In particular, if f is differentiable from the right and from the left at x then we get convergence to $c=\frac{1}{2}[f(x_+)+f(x_-)]$ which is Dirichlet's theorem. □

Bessel.

Although Fejer's theorem gives one interpretation of Fouriers claim that "every" periodic function can be expanded into a Fourier series, and Dirichlet's theorem gives another justification of this claim, an entirely different approach to Fourier's claim derives from the work of the astronomer Bessel, an approach derived from the method of "least squares" of great use in observational astronomy. The modern (i.e. mid 20th century) setting for this approach is the concept of a Hilbert space and the notion of an orthonormal basis.

So we are now going to embark on the basics of Hilbert space. Much of this will be in the nature of a review for most of you.

Friedrich Wilhelm Bessel

Born: 22 July 1784 in Minden, Westphalia (now Germany)
Died: 17 March 1846 in Königsberg, Prussia (now Kaliningrad, Russia)

Semi-scalar products and scalar products.

Let V be a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.

Semi-scalar products and scalar products.

Let V be a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.
2. $(g, f) = \overline{(f, g)}$. This implies that (f, g) is anti-linear in g when f is held fixed. In other words. $(f, ag + bh) = \overline{a}(f, g) + \overline{b}(f, h)$. It also implies that (f, f) is real.

Semi-scalar products and scalar products.

Let V be a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.
2. $(g, f) = \overline{(f, g)}$. This implies that (f, g) is anti-linear in g when f is held fixed. In other words. $(f, ag + bh) = \overline{a}(f, g) + \overline{b}(f, h)$. It also implies that (f, f) is real.
3. $(f, f) \geq 0$ for all $f \in V$.

Semi-scalar products and scalar products.

Let V be a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.
2. $(g, f) = \overline{(f, g)}$. This implies that (f, g) is anti-linear in g when f is held fixed. In other words. $(f, ag + bh) = \overline{a}(f, g) + \overline{b}(f, h)$. It also implies that (f, f) is real.
3. $(f, f) \geq 0$ for all $f \in V$.

Semi-scalar products and scalar products.

Let V be a complex vector space. A rule assigning to every pair of vectors $f, g \in V$ a complex number (f, g) is called a **semi-scalar product** if

1. (f, g) is linear in f when g is held fixed.
2. $(g, f) = \overline{(f, g)}$. This implies that (f, g) is anti-linear in g when f is held fixed. In other words. $(f, ag + bh) = \overline{a}(f, g) + \overline{b}(f, h)$. It also implies that (f, f) is real.
3. $(f, f) \geq 0$ for all $f \in V$.

If 3. is replaced by the stronger condition

4. $(f, f) > 0$ *for all non-zero* $f \in V$

then we say that $(\ ,\ )$ is a **scalar product**.

Examples: $\mathbb{C}^n$.

$V = \mathbb{C}^n$, so an element $\mathbf{z}$ of V is a column vector of complex numbers:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$$

and $(\mathbf{z}, \mathbf{w})$ is given by

$$(\mathbf{z}, \mathbf{w}) := \sum_1^n z_i \overline{w_i}.$$

$\mathcal{C}(\mathbb{T})$.

V consists of all continuous (complex valued) functions on the real line which are periodic of period 2π and

$$(f,g) := \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)\overline{g(x)}dx.$$

We will denote this space by $\mathcal{C}(\mathbb{T})$. Here the letter $\mathbb{T}$ stands for the one dimensional torus, i.e. the circle. We are identifying functions which are periodic with period 2π with functions which are defined on the circle $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$.

Doubly infinite sequences of complex numbers.

V consists of all doubly infinite sequences of complex numbers

$$\mathbf{a} = \ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots$$

which satisfy

$$\sum |a_i|^2 < \infty.$$

Here

$$(\mathbf{a}, \mathbf{b}) := \sum a_i \overline{b}_i.$$

The Cauchy-Schwarz inequality.

This says that if (,) is a semi-scalar product then

$$|(f,g)| \leq (f,f)^{\frac{1}{2}}(g,g)^{\frac{1}{2}}. \tag{1}$$

Proof.

For any real number t condition 3. above says that $(f - tg, f - tg) \geq 0$.

Proof.

For any real number t condition 3. above says that $(f - tg, f - tg) \geq 0$. Expanding out gives

$$0 \leq (f - tg, f - tg) = (f, f) - t[(f, g) + (g, f)] + t^2(g, g).$$

Since $(g, f) = \overline{(f, g)}$, the coefficient of t in the above expression is twice the real part of (f, g). So the real quadratic form

$$Q(t) := (f, f) - 2\text{Re}(f, g)t + t^2(g.g)$$

is nowhere negative.

Proof.

For any real number t condition 3. above says that $(f - tg, f - tg) \geq 0$. Expanding out gives

$$0 \leq (f - tg, f - tg) = (f, f) - t[(f, g) + (g, f)] + t^2(g, g).$$

Since $(g, f) = \overline{(f, g)}$, the coefficient of t in the above expression is twice the real part of (f, g). So the real quadratic form

$$Q(t) := (f, f) - 2\text{Re}(f, g)t + t^2(g.g)$$

is nowhere negative. So it can not have distinct real roots, and hence by the $b^2 - 4ac$ rule we get

$$4(\text{Re}(f, g))^2 - 4(f, f)(g, g) \leq 0$$

or

$$(\text{Re}(f, g))^2 \leq (f, f)(g, g). \tag{2}$$

$$(\mathrm{Re}(f,g))^2 \leq (f,f)(g,g).$$

This is useful and almost, but not quite what we want. But we may apply this inequality to $h = e^{i\theta}g$ for any θ. Then $(h,h) = (g,g)$. Choose θ so that

$$(f,g) = re^{i\theta}$$

where $r = |(f,g)|$. Then

$$(f,h) = (f, e^{i\theta}g) = e^{-i\theta}(f,g) = |(f,g)|$$

and the preceding inequality with g replaced by h gives

$$|(f,g)|^2 \leq (f,f)(g,g)$$

and taking square roots gives (1).

For any semiscalar product define

$$\|f\| := (f,f)^{\frac{1}{2}}$$

so we can write the Cauchy-Schwarz inequality as

$$|(f,g)| \leq \|f\|\|g\|.$$

The **triangle inequality** says that

$$\|f+g\| \leq \|f\| + \|g\|. \tag{3}$$

Proof.

$$\begin{aligned}
\|f+g\|^2 &= (f+g, f+g) \\
&= (f,f) + 2\mathrm{Re}(f,g) + (g,g) \\
&\leq (f,f) + 2\|f\|\|g\| + (g,g) \quad \text{by (2)} \\
&= \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 \\
&= (\|f\| + \|g\|)^2.
\end{aligned}$$

Taking square roots gives the triangle inequality (3). □

Proof.

$$\begin{aligned} \|f+g\|^2 &= (f+g, f+g) \\ &= (f,f) + 2\mathrm{Re}(f,g) + (g,g) \\ &\leq (f,f) + 2\|f\|\|g\| + (g,g) \quad \text{by (2)} \\ &= \|f\|^2 + 2\|f\|\|g\| + \|g\|^2 \\ &= (\|f\| + \|g\|)^2. \end{aligned}$$

Taking square roots gives the triangle inequality (3). □

Notice that

$$\|cf\| = |c|\|f\| \tag{4}$$

since $(cf, cf) = c\overline{c}(f,f) = |c|^2\|f\|^2$.

Pre-Hilbert spaces.

Suppose we try to define the distance between two elements of V by

$$d(f,g) := \|f-g\|.$$

Notice that then $d(f,f)=0$, $d(f,g)=d(g,f)$ and for any three elements

$$d(f,h) \leq d(f,g) + d(g,h)$$

by virtue of the triangle inequality. The only trouble with this definition is that we might have two distinct elements at zero distance, i.e. $0 = d(f,g) = \|f-g\|$. But this can not happen if $(\ ,\ )$ is a scalar product, i.e. satisfies condition 4.

Pre-Hilbert spaces.

Suppose we try to define the distance between two elements of V by

$$d(f,g) := \|f-g\|.$$

Notice that then $d(f,f)=0$, $d(f,g)=d(g,f)$ and for any three elements

$$d(f,h) \leq d(f,g) + d(g,h)$$

by virtue of the triangle inequality. The only trouble with this definition is that we might have two distinct elements at zero distance, i.e. $0=d(f,g)=\|f-g\|$. But this can not happen if $(\ ,\ )$ is a scalar product, i.e. satisfies condition 4.

A complex vector space V endowed with a scalar product is called a **pre-Hilbert** space.

Normed Spaces.

Let V be a complex vector space and let $\|\cdot\|$ be a map which assigns to any $f \in V$ a non-negative real $\|f\|$ number such that $\|f\| > 0$ for all non-zero f. If $\|\cdot\|$ satisfies the triangle inequality (3) and equation (4):

$$\|cf\| = |c|\|f\| \quad \forall c \in \mathbb{C}$$

it is called a **norm**. A vector space endowed with a norm is called a normed space.

Normed Spaces.

Let V be a complex vector space and let $\|\cdot\|$ be a map which assigns to any $f \in V$ a non-negative real $\|f\|$ number such that $\|f\| > 0$ for all non-zero f. If $\|\cdot\|$ satisfies the triangle inequality (3) and equation (4):

$$\|cf\| = |c|\|f\| \quad \forall c \in \mathbb{C}$$

it is called a **norm**. A vector space endowed with a norm is called a normed space.

The pre-Hilbert spaces can be characterized among all normed spaces by the parallelogram law as we will discuss below.

Limits.

The reason for the prefix "pre" is the following: The distance d defined above has all the desired properties we might expect of a distance. In particular, we can define the notion of "limit" : If f_n is a sequence of elements of V, and $f \in V$ we say that f is the limit of the f_n and write

$$\lim_{n\to\infty} f_n = f, \quad \text{or} \quad f_n \to f$$

if, for any positive number ϵ there is an $N = N(\epsilon)$ such that

$$d(f_n, f) < \epsilon \quad \text{for all } n \geq N.$$

If a sequence converges to some limit f, then this limit is unique, since any limits must be at zero distance and hence equal.

Cauchy sequences.

We say that a sequence of elements is **Cauchy** if for any $\delta > 0$ there is an $K = K(\delta)$ such that

$$d(f_m, f_n) < \delta \quad \forall, m, n \geq K.$$

If the sequence f_n has a limit, then it is Cauchy - just choose $K(\delta) = N(\frac{1}{2}\delta)$ and use the triangle inequality.

Cauchy sequences.

We say that a sequence of elements is **Cauchy** if for any $\delta > 0$ there is an $K = K(\delta)$ such that

$$d(f_m, f_n) < \delta \quad \forall, m, n \geq K.$$

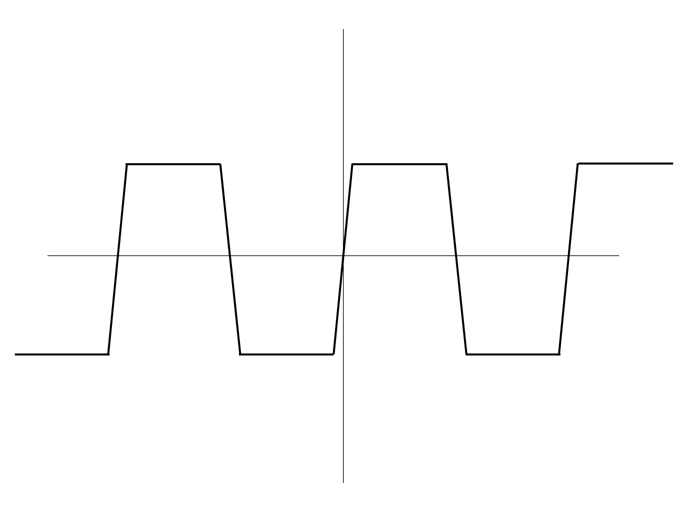
If the sequence f_n has a limit, then it is Cauchy - just choose $K(\delta) = N(\frac{1}{2}\delta)$ and use the triangle inequality.

But it is quite possible that a Cauchy sequence has no limit. As an example of this type of phenomenon, think of the rational numbers with $|r - s|$ as the distance. The whole point of introducing the real numbers is to guarantee that every Cauchy sequence has a limit.

So we say that a pre-Hilbert space is a **Hilbert space** if it is "complete" in the above sense - if every Cauchy sequence has a limit.
Since the complex numbers are complete (because the real numbers are), it follows that $\mathbb{C}^n$ is complete, i.e. is a Hilbert space. Indeed, we can say that any finite dimensional pre-Hilbert space is a Hilbert space because it is isomorphic (as a pre-Hilbert space) to $\mathbb{C}^n$ for some n. (See below when we discuss orthonormal bases.)

The trouble is in the infinite dimensional case.

The trouble is in the infinite dimensional case, such as the space of continuous periodic functions. This space is not complete. For example, let f_n be the function which is equal to one on $(-\pi + \frac{1}{n}, -\frac{1}{n})$, equal to zero on $(\frac{1}{n}, \pi - \frac{1}{n})$ and extended linearly $-\frac{1}{n}$ to $\frac{1}{n}$ and from $\pi - \frac{1}{n}$ to $\pi + \frac{1}{n}$ so as to be continuous and then extended so as to be periodic.(Thus on the interval $(\pi - \frac{1}{n}, \pi + \frac{1}{n})$ the function is given by $f_n(x) = \frac{n}{2}(x - (\pi - \frac{1}{n}).)$ If $m \leq n$, the functions f_m and f_n agree outside two intervals of length $\frac{2}{m}$ and on these intervals $|f_m(x) - f_n(x)| \leq 1$. So $\|f_m - f_n\|^2 \leq \frac{1}{2\pi} \cdot 2/m$ showing that the sequence $\{f_n\}$ is Cauchy. But the limit would have to equal one on $(-\pi, 0)$ and equal zero on $(0, \pi)$ and so be discontinuous at the origin and at π. Thus the space of continuous periodic functions is not a Hilbert space, only a pre-Hilbert space.



Completing a pre-Hilbert space to get a Hilbert space.

Just as we complete the rationals to get the real numbers, we may complete any metric space to get a complete metric space. The completion of a normed vector space will be a complete normed vector space, which is called a Banach space.

Completing a pre-Hilbert space to get a Hilbert space.

Just as we complete the rationals to get the real numbers, we may complete any metric space to get a complete metric space. The completion of a normed vector space will be a complete normed vector space, which is called a Banach space.

From the parallelogram law discussed below, it will follow that the completion of a pre-Hilbert space is a Hilbert space.

Completing a pre-Hilbert space to get a Hilbert space.

Just as we complete the rationals to get the real numbers, we may complete any metric space to get a complete metric space. The completion of a normed vector space will be a complete normed vector space, which is called a Banach space.

From the parallelogram law discussed below, it will follow that the completion of a pre-Hilbert space is a Hilbert space.

For example, we *define* the Hilbert space $L_2(\mathbb{T})$ to be the completion of the space $\mathcal{C}(\mathbb{T})$ of continuous periodic functions under the norm coming from scalar product introduced above.

Let V be a pre-Hilbert space. We have

$$\|f+g\|^2 = \|f\|^2 + 2\mathrm{Re}(f,g) + \|g\|^2.$$

So

$$\|f+g\|^2 = \|f\|^2 + \|g\|^2 \quad \Leftrightarrow \quad \mathrm{Re}\,(f,g) = 0. \tag{5}$$

We make the definition

$$f \perp g \;\Leftrightarrow\; (f,g) = 0$$

and say that f is perpendicular to g or that f is orthogonal to g. Notice that this is a stronger condition than the condition for the Pythagorean theorem, the right hand condition in (5). For example $\|f+if\|^2 = 2\|f\|^2$ but $(f,if) = -i\|f\|^2 \neq 0$ if $\|f\| \neq 0$.

Orthogonality implies independence.

If u_i is some finite collection of mutually orthogonal vectors, then so are $z_i u_i$ where the z_i are any complex numbers. So if

$$u = \sum_i z_i u_i$$

then by the Pythagorean theorem

$$\|u\|^2 = \sum_i |z_i|^2 \|u_i\|^2.$$

In particular, if the $u_i \neq 0$, then $u = 0 \Rightarrow z_i = 0$ for all i. This shows that any set of mutually orthogonal (non-zero) vectors is linearly independent.

Orthonormal sets.

Notice that the set of functions

$$e^{in\theta}$$

is an **orthonormal** set in the space of continuous periodic functions in that not only are they mutually orthogonal, but each has norm one.

The theorem of Apollonius.

Adding the equations

$$\|f+g\|^2 = \|f\|^2 + 2\mathrm{Re}\,(f,g) + \|g\|^2 \qquad (6)$$

$$\|f-g\|^2 = \|f\|^2 - 2\mathrm{Re}\,(f,g) + \|g\|^2 \qquad (7)$$

gives

$$\|f+g\|^2 + \|f-g\|^2 = 2\left(\|f\|^2 + \|g\|^2\right). \qquad (8)$$

This is known as the **parallelogram law**. It is the algebraic expression of the theorem of Apollonius which asserts that the sum of the areas of the squares on the sides of a parallelogram equals the sum of the areas of the squares on the diagonals.

If we subtract (7) from (6) we get

$$\mathrm{Re}(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right). \tag{9}$$

Now $(if,g) = i(f,g)$ and $\mathrm{Re}\{i(f,g)\} = -\mathrm{Im}(f,g)$ so

$$\mathrm{Im}(f,g) = -\mathrm{Re}(if,g) = \mathrm{Re}(f,ig)$$

so

$$(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2\right). \tag{10}$$

If we complete a pre-Hilbert space, the right hand side of (10) is defined on the completion, and is continuous there. It follows that the scalar product extends to the completion, and, by continuity, satisfies all the axioms for a scalar product, plus the completeness condition for the associated norm. Thus the completion of a pre-Hilbert space is a Hilbert space.

The theorem of Jordan and von Neumann.

This is essentially a converse to the theorem of Apollonius. It says that if $\|\cdot\|$ is a norm on a (complex) vector space V which satisfies

$$\|f+g\|^2+\|f-g\|^2 = 2\left(\|f\|^2+\|g\|^2\right) \qquad (8),$$

then V is in fact a pre-Hilbert space with $\|f\|^2 = (f,f)$. If the theorem is true, then the scalar product must be given by

$$(f,g) = \frac{1}{4}\left(\|f+g\|^2-\|f-g\|^2+i\|f+ig\|^2-i\|f-ig\|^2\right) \quad (10).$$

So we must prove that if we take (10) as the definition, then all the axioms on a scalar product hold.

Verifying $(g,f)=\overline{(f,g)}$.

$$(f,g)=\frac{1}{4}\left(\|f+g\|^2-\|f-g\|^2+i\|f+ig\|^2-i\|f-ig\|^2\right) \quad (10).$$

Indeed, the real part of the right hand side of (10) is unchanged under the interchange of f and g (since $g-f=-(f-g)$ and $\|-h\|=\|h\|$ for any h is one of the properties of a norm). Also $g+if=i(f-ig)$ and $\|ih\|=\|h\|$ so the last two terms on the right of (10) get interchanged, proving that $(g,f)=\overline{(f,g)}$.

Verifying $(if, g) = i(f, g)$.

$$(f, g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2\right) \quad (10).$$

Indeed replacing f by if sends $\|f+ig\|^2$ into $\|if+ig\|^2 = \|f+g\|^2$ and sends $\|f+g\|^2$ into $\|if+g\|^2 = \|i(f-ig)\|^2 = \|f-ig\|^2 = i(-i\|f-ig\|^2)$ so has the effect of multiplying the sum of the first and fourth terms by i, and similarly for the sum of the second and third terms on the right hand side of (10).

The next few slides will be devoted to showing that (f, g) is linear in f.

$$\|f+g\|^2+\|f-g\|^2=2\left(\|f\|^2+\|g\|^2\right). \quad (8)$$

$$\operatorname{Re}(f,g)=\frac{1}{4}\left(\|f+g\|^2-\|f-g\|^2\right). \quad (9).$$

$$(f,g)=\frac{1}{4}\left(\|f+g\|^2-\|f-g\|^2+i\|f+ig\|^2-i\|f-ig\|^2\right) \ (10).$$

Now (10) implies (9). Suppose we replace f,g in (8) by f_1+g, f_2 and by f_1-g, f_2 and subtract the second equation from the first. We get

$$\|f_1+f_2+g\|^2-\|f_1+f_2-g\|^2+\|f_1-f_2+g\|^2-\|f_1-f_2-g\|^2$$

$$=2\left(\|f_1+g\|^2-\|f_1-g\|^2\right).$$

$$\mathrm{Re}(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right). \quad (9).$$

$$(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2\right) \quad (10).$$

$$\|f_1+f_2+g\|^2 - \|f_1+f_2-g\|^2 + \|f_1-f_2+g\|^2 - \|f_1-f_2-g\|^2$$
$$= 2\left(\|f_1+g\|^2 - \|f_1-g\|^2\right).$$

In view of (9) we can write this last equation as

$$\mathrm{Re}\,(f_1+f_2,g) + \mathrm{Re}\,(f_1-f_2,g) = 2\mathrm{Re}\,(f_1,g).$$

$$\text{Re }(f_1+f_2,g)+\text{Re }(f_1-f_2,g)=2\text{Re }(f_1,g). \tag{11}$$

$$\text{Re}(f,g)=\frac{1}{4}\left(\|f+g\|^2-\|f-g\|^2\right). \qquad (9).$$

Now the right hand side of (9) vanishes when $f=0$ since $\|g\|=\|-g\|$. So if we take $f_1=f_2=f$ in (11) we get

$$\text{Re }(2f,g)=2\text{Re }(f,g).$$

We can thus write (11) as

$$\text{Re }(f_1+f_2,g)+\text{Re }(f_1-f_2,g)=\text{Re }(2f_1,g).$$

So we have:

$$\mathrm{Re}(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2\right). \quad (9).$$

$$(f,g) = \frac{1}{4}\left(\|f+g\|^2 - \|f-g\|^2 + i\|f+ig\|^2 - i\|f-ig\|^2\right) \quad (10).$$

$$\mathrm{Re}\,(f_1+f_2,g) + \mathrm{Re}\,(f_1-f_2,g) = 2\mathrm{Re}\,(f_1,g). \quad (11).$$

Substitute $f_1 \mapsto \frac{1}{2}(f_1+f_2), \quad f_2 \mapsto \frac{1}{2}(f_1-f_2)$ in (11) yielding

$$\mathrm{Re}\,(f_1+f_2,g) = \mathrm{Re}\,(f_1,g) + \mathrm{Re}\,(f_2,g).$$

Since it follows from (10) and (9) that

$$(f,g) = \mathrm{Re}\,(f,g) - i\mathrm{Re}\,(if,g)$$

we conclude that

$$(f_1+f_2,g) = (f_1,g) + (f_2,g).$$

$$(f_1 + f_2, g) = (f_1, g) + (f_2, g).$$

Taking $f_1 = -f_2$ shows that

$$(-f, g) = -(f, g).$$

Consider the collection $\mathcal{C}$ of complex numbers α which satisfy

$$(\alpha f, g) = \alpha(f, g)$$

(for all f, g). We know from $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$ that

$$\alpha, \beta \in \mathcal{C} \Rightarrow \alpha + \beta \in \mathcal{C}.$$

So $\mathcal{C}$ contains all integers.

If $0 \neq \beta \in \mathcal{C}$ then

$$(f,g) = (\beta \cdot (1/\beta)f, g) = \beta((1/\beta)f, g)$$

so $\beta^{-1} \in \mathcal{C}$. Thus $\mathcal{C}$ contains all (complex) rational numbers. The theorem will be proved if we prove that $(\alpha f, g)$ is continuous in α.

The triangle inequality $\|f+g\| \le \|f\| + \|g\|$ applied to $f = f_2, g = f_1 - f_2$ implies that $\|f_1\| \le \|f_1 - f_2\| + \|f_2\|$ or

$$\|f_1\| - \|f_2\| \le \|f_1 - f_2\|.$$

Interchanging the role of f_1 and f_2 gives

$$|\ \|f_1\| - \|f_2\|\ | \le \|f_1 - f_2\|.$$

Therefore

$$|\ \|\alpha f \pm g\| - \|\beta f \pm g\|\ | \le \|(\alpha - \beta) f\|.$$

Since $\|(\alpha - \beta)f\| \to 0$ as $\alpha \to \beta$ this shows that the right hand side of (10) when applied to αf and g is a continuous function of α. Thus $\mathcal{C} = \mathbb{C}$. We have proved

Theorem

[P. Jordan and J. von Neumann] *If V is a normed space whose norm satisfies (8) then V is a pre-Hilbert space.*

Notice that the condition (8) involves only two vectors at a time. So we conclude as an immediate consequence of this theorem that

Corollary

A normed vector space is pre-Hilbert space if and only if every two dimensional subspace is a Hilbert space in the induced norm.

Actually, a weaker version of this corollary, with two replaced by three had been proved by Fréchet, *Annals of Mathematics*, July 1935, who raised the problem of giving an abstract characterization of those norms on vector spaces which come from scalar products. In the immediately following paper Jordan and von Neumann proved the theorem above leading to the stronger corollary that two dimensions suffice.

Orthogonal complements.

We continue with the assumption that V is pre-Hilbert space. If A and B are two subsets of V, we write $A \perp B$ if $u \in A$ and $v \in B \Rightarrow u \perp v$, in other words if every element of A is perpendicular to every element of B. Similarly, we will write $v \perp A$ if the element v is perpendicular to all elements of A. Finally, we will write $A^{\perp}$ for the set of all v which satisfy $v \perp A$. Notice that $A^{\perp}$ is always a linear subspace of V, for any A.

The problem of orthogonal projection.

Now let M be a (linear) subspace of V. Let v be some element of V, not necessarily belonging to M. We want to investigate the problem of finding a $w \in M$ such that $(v - w) \perp M$. Of course, if $v \in M$ then the only choice is to take $w = v$. So the interesting problem is when $v \notin M$. Suppose that such a w exists, and let x be any (other) point of M. Then by the Pythagorean theorem,

$$\|v - x\|^2 = \|(v - w) + (w - x)\|^2 = \|v - w\|^2 + \|w - x\|^2$$

since $(v - w) \perp M$ and $(w - x) \in M$. So

$$\|v - w\| \leq \|v - x\|$$

and this inequality is strict if $x \neq w$. In words:

If we can find a $w \in M$ such that $(v - w) \perp M$ then w is the unique solution of the problem of finding the point in M which is closest to v.

Conversely, suppose we found a $w \in M$ which has this minimization property, and let x be any element of M. Then for any real number t we have

$$\|v-w\|^2 \leq \|(v-w)+tx\|^2 = \|v-w\|^2+2t\text{Re }(v-w,x)+t^2\|x\|^2.$$

Since the minimum of this quadratic polynomial in t occurring on the right is achieved at $t = 0$, we conclude (by differentiating with respect to t and setting $t = 0$, for example) that

$$\mathrm{Re}\,(v-w,x)=0.$$

By our usual trick of replacing x by $e^{i\theta}x$ we conclude that

$$(v-w,x)=0.$$

Since this holds for all $x \in M$, we conclude that $(v-w) \perp M$. So to find w we search for the minimum of $\|v-x\|,\ x \in M$.

Now $\|v-x\| \geq 0$ and is some finite number for any $x \in M$. So there will be some real number m such that $m \leq \|v-x\|$ for $x \in M$, and such that no strictly larger real number will have this property. (m is known as the "greatest lower bound" of the values $\|v-x\|,\ x \in M$.) So we can find a sequence of vectors $x_n \in M$ such that

$$\|v-x_n\| \to m.$$

We claim that

Proposition

The x_n form a Cauchy sequence.

Indeed, $\|x_i - x_j\|^2 = \|(v - x_j) - (v - x_i)\|^2$, and by the parallelogram law this equals

$$2\left(\|v - x_i\|^2 + \|v - x_j\|^2\right) - \|2v - (x_i + x_j)\|^2.$$

Now the expression in parenthesis converges to $2m^2$. The last term on the right is $-\|2(v - \frac{1}{2}(x_i + x_j)\|^2$. Since $\frac{1}{2}(x_i + x_j) \in M$, we conclude that

$$\|2v - (x_i + x_j)\|^2 \geq 4m^2$$

so

$$\|x_i - x_j\|^2 \leq 4(m + \epsilon)^2 - 4m^2$$

for i and j large enough that $\|v - x_i\| \leq m + \epsilon$ and $\|v - x_j\| \leq m + \epsilon$. This proves that the sequence x_n is Cauchy. □

The essential role of the completeness of M.

Here is the crux of the matter: If M is complete, then we can conclude that the x_n converge to a limit w which is then the unique element in M such that $(v - w) \perp M$. It is at this point that completeness plays such an important role.
Put another way, we can say that if M is a subspace of V which is complete (under the scalar product $(\ ,\ )$ restricted to M) then we have the orthogonal direct sum decomposition

$$V = M \oplus M^{\perp},$$

which says that every element of V can be uniquely decomposed into the sum of an element of M and a vector perpendicular to M.

For example, if M is the one dimensional subspace consisting of all (complex) multiples of a non-zero vector y, then M is complete, since $\mathbb{C}$ is complete. So w exists. Since all elements of M are of the form ay, we can write $w = ay$ for some complex number a. Then $(v - ay, y) = 0$ or

$$(v, y) = a\|y\|^2$$

so

$$a = \frac{(v, y)}{\|y\|^2}.$$

We call a the **Fourier coefficient** of v with respect to y. Particularly useful is the case where $\|y\| = 1$ and we can write

$$a = (v, y). \qquad (12)$$

Orthogonal projection

Getting back to the general case, if $V = M \oplus M^{\perp}$ holds, so that to every v there corresponds a unique $w \in M$ satisfying $(v - w) \in M^{\perp}$ the map $v \mapsto w$ is called orthogonal projection of V onto M and will be denoted by π_M.

Linear and antilinear maps.

Let V and W be two complex vector spaces. A map

$$T : V \to W$$

is called **linear** if

$$T(\lambda x + \mu y) = \lambda T(x) + \mu T(Y) \quad \forall\, x, y \in V, \quad \lambda, \mu \in \mathbb{C}$$

and is called **anti-linear** if

$$T(\lambda x + \mu y) = \overline{\lambda} T(x) + \overline{\mu} T(Y) \quad \forall\, x, y \in V \quad \lambda, \mu \in \mathbb{C}.$$

Linear functions.

If $\ell : V \to \mathbb{C}$ is a linear map, (also known as a linear function) then

$$\ker \ell := \{x \in V|\ \ell(x) = 0\}$$

has codimension one (unless $\ell \equiv 0$). Indeed, if

$$\ell(y) \neq 0$$

then

$$\ell(x) = 1 \quad \text{where } x = \frac{1}{\ell(y)} y$$

and for any $z \in V$,

$$z - \ell(z)x \in \ker \ell.$$

If V is a normed space and ℓ is continuous, then ker (ℓ) is a closed subspace. The space of continuous linear functions is denoted by V^*. It has its own norm defined by

$$\|\ell\| := \sup_{x \in V, \|x\| \neq 0} |\ell(x)|/\|x\|.$$

Suppose that H is a pre-hilbert space. We have an antilinear map

$$\phi : H \to H^*, \quad (\phi(g))(f) := (f, g).$$

The Cauchy-Schwarz inequality implies that

$$\|\phi(g)\| \leq \|g\|$$

and $\quad (g, g) = \|g\|^2 \quad$ shows that $\quad \|\phi(g)\| = \|g\|.$
In particular the map ϕ is injective.

The Riesz representation theorem.

The Riesz representation theorem says that if H is a Hilbert space, then this map is surjective:

Theorem

Every continuous linear function on H is given by scalar product by some element of H.

Proof of the Riesz representation theorem.

The proof follows from theorem about projections applied to

$$N := \ker \ell :$$

If $\ell = 0$ there is nothing to prove. If $\ell \neq 0$ then N is a closed subspace of codimension one. Choose $v \notin N$. Then there is an $x \in N$ with $(v - x) \perp N$. Let

$$y := \frac{1}{\|v - x\|}(v - x).$$

Then

$$y \perp N \quad \text{and} \quad \|y\| = 1.$$

For any $f \in H$,

$$[f - (f, y)y] \perp y$$

so

$$f - (f, y)y \in N$$

or

$$\ell(f) = (f, y)\ell(y),$$

so if we set

$$g := \overline{\ell(y)}y$$

then

$$(f, g) = \ell(f)$$

for all $f \in H$. □

What is $L_2(\mathbb{T})$?

We have defined the space $L_2(\mathbb{T})$ to be the completion of the space $\mathcal{C}(\mathbb{T})$ under the L_2 norm $\|f\|_2 = (f, f)^{\frac{1}{2}}$. In particular, every linear function on $\mathcal{C}(\mathbb{T})$ which is continuous with respect to the this L_2 norm extends to a unique continuous linear function on $L_2(\mathbb{T})$. By the Riesz representation theorem we know that every such continuous linear function is given by scalar product by an element of $L_2(\mathbb{T})$. Thus we may think of the elements of $L_2(\mathbb{T})$ as being the linear functions on $\mathcal{C}(\mathbb{T})$ which are continuous with respect to the L_2 norm. An element of $L_2(\mathbb{T})$ should not be thought of as a function, but rather as a linear function on the space of continuous functions relative to a special norm - the L_2 norm.

The Riesz representation theorem (and its various variants all due to Riesz) will turn out to be the main hero of this course - despite the fact that we were able to give an easy proof in the second lecture!

Frigyes Riesz

Born: 22 Jan 1880 in Györ, Austria-Hungary (now Hungar
Died: 28 Feb 1956 in Budapest, Hungary

Summary.