The Fourier Transform and its Applications

The math, the majesty, the end

Last time we worked with the building blocks of periodic functions – sines and cosines and complex exponentials – and considered general sums of such ‘harmonics’. We also showed that *if* a periodic function $f(t)$ (period 1, as a convenient normalization) can be written as a sum

$$f(t) = \sum_{n=-N}^{N} c_n e^{2\pi nit},$$

then the coefficients are given by the integral

$$c_n = \int_0^1 e^{-2\pi int} f(t)\, dt.$$

This was a pretty straightforward derivation, isolating c_n and then integrating. When $f(t)$ is real, as in many applications, one has the symmetry relation $c_{-n} = \overline{c_n}$. In a story we'll spin out over the rest of the quarter, we think of this integral as some kind of transform of f, and use the notation

$$\hat{f}(n) = \int_0^1 e^{-2\pi int} f(t)\, dt.$$

to indicate this relationship.[1]

At this stage, we haven't done much. We have only demonstrated that if it is possible to write a periodic function as a sum of simple harmonics, then it must be done in the way we've just said. We also have some examples that indicate the possible difficulties in this sort of representation; an infinite series may be required and then convergence is certainly an issue. But we're about to do a lot. We're about to answer the question of how far the idea can be pushed: When *can* a periodic signal be written as a sum of simple harmonics?

Square integrable functions

There's much more to the structure of the Fourier coefficients and to the idea of writing a periodic function as a sum of complex exponentials than might appear from our simple derivation. There are:

[1] Notice that while $f(t)$ is defined for a continuous variable t the transformed function $\hat{f}$ is defined on the integers. There are reasons for this that are much deeper than just solving for the unknown coefficients as we did last time.

- Algebraic and geometric aspects
  - The algebraic and geometric aspects are straightforward extensions of the algebra and geometry of vectors in Euclidean space. The key ideas are the inner product (dot product), orthogonality, and norm. We can pretty much cover the whole thing. I remind you that your job here is to transfer your intuition from geometric vectors to a more general setting where the vectors are signals; at least accept that the words transfer in some kind of meaningful way even if the pictures do not.
- Analytic aspects
  - The analytic aspects are not straightforward and require new ideas on limits and on the nature of integration. The aspect of 'analysis' as a field of mathematics distinct from other fields is its systematic use of limiting processes. To define a new kind of limit, or to find new consequences of taking limits (or trying to), is to define a new area of analysis. We really can't cover the whole thing, and it's not appropriate to attempt to. But I'll say a little bit here, and similar issues will come up when we define the Fourier transform.

The punchline revealed

Let me introduce the notation, basic terminology and state what the important results are now, so you can see the point. Then I'll explain where these ideas come from and how they fit together.

Once again, to be definite we're working with periodic functions of period 1. We can consider such a function already to be defined for all real numbers, and satisfying the identity $f(t+1) = f(t)$ for all t, or we can consider $f(t)$ to be defined initially only on the interval from 0 to 1, say, and then extended to be periodic and defined on all of $\mathbf{R}$ by repeating the graph (recall the periodizing operation in the first problem set). In either case, once we know what we need to know about the function on $[0,1]$ we know everything. All of the action in the following discussion takes place on the interval $[0,1]$.

When $f(t)$ is a signal defined on $[0,1]$ the *energy* of the signal is defined to be the integral

$$\int_0^1 |f(t)|^2 \, dt.$$

This definition of energy comes up in other physical contexts also; we don't have to be talking about functions of time. Thus

$$\int_0^1 |f(t)|^2 \, dt < \infty$$

means that the signal has *finite energy*, a reasonable condition to expect or to impose. (In some areas the integral of the square is identified with power.)

I'm writing the definition in terms of the integral of the absolute value squared, $|f(t)|^2$, rather than just $f(t)^2$ because we'll want to consider the definition to apply to complex valued functions. For real valued functions it doesn't matter whether we integrate $|f(t)|^2$ or $f(t)^2$.

One further point before we go on. Though our purpose is to use the finite energy condition to work with periodic functions, and though you think of periodic functions as defined for all time, you can see why we have to restrict attention to one period (any period). An integral of a periodic function from $-\infty$ to ∞, for example

$$\int_{-\infty}^{\infty} \sin^2 t \, dt,$$

does not exist.

For mathematical reasons, primarily, it's best to take the square root of the integral, and to define

$$||f|| = \left\{ \int_0^1 |f(t)|^2 \, dt \right\}^{1/2}$$

With this definition one has, for example, that

$$||\alpha f|| = |\alpha| ||f||,$$

whereas if we didn't take the square root the constant would come out to the second power – see below. One can also show, though the proof is not so obvious (see Appendix 1), that the triangle inequality holds:

$$||f + g|| \leq ||f|| + ||g||.$$

Write that out in terms of integrals if you think it's obvious:

$$\left\{ \int_0^1 |f(t) + g(t)|^2 \, dt \right\}^{1/2} \leq \left\{ \int_0^1 |f(t)|^2 \, dt \right\}^{1/2} + \left\{ \int_0^1 |g(t)|^2 \, dt \right\}^{1/2}.$$

We measure the distance between two functions via

$$||f - g|| = \left\{ \int_0^1 |f(t) - g(t)|^2 \, dt \right\}^{1/2}$$

Then $||f - g|| = 0$ if and only if $f = g$.

> **Now get this:** The length of a vector is the square root of the sum of the squares of its components. This integral norm is the continuous analog of that, and so is the definition of distance.[2] We'll make the analogy even closer when we introduce the corresponding dot product.

We let $L^2([0,1])$ be the set of functions $f(t)$ on $[0,1]$ for which

$$\int_0^1 |f(t)|^2 \, dt < \infty$$

The 'L' stands for Lebesgue, the French mathematician who introduced a new definition of the integral that underlies the analytic aspects of the results we're about to talk about. His work was around the turn of the 20th century. The length we've just introduced, $||f||$, is called the *square norm* or the L^2-*norm* of the function. When we want to distinguish this from other norms that might (that will) come up, we write $||f||_2$.

It's true, you'll be relieved to hear, that if $f(t)$ is in $L^2([0,1])$ then the integral defining its Fourier coefficients exists. See Appendix 1 for this. The complex integral

$$\int_0^1 e^{-2\pi int} f(t) \, dt$$

[2]If we've really defined a 'length' then scaling $f(t)$ to $\alpha f(t)$ should scale the length of $f(t)$. If we didn't take the square root in defining $||f||$ the length wouldn't scale to the first power.

can be written in terms of two real integrals by writing $e^{-2\pi int} = \cos 2\pi nt - i \sin 2\pi nt$ so everything can be defined and computed in terms of real quantities. There are more things to be said on complex-valued versus real-valued functions in all of this, but it's best to put that off just now.

Here now is the life's work of several generations of mathematicians, all dead, all still revered:

Let $f(t)$ be in $L^2([0,1])$ and let

$$\hat{f}(n) = \int_0^1 e^{-2\pi int} f(t)\, dt, \quad n = 0, \pm 1, \pm 2, \dots,$$

be its Fourier coefficients. Then

1. For any N the finite sum

   $$\sum_{n=-N}^{N} \hat{f}(n) e^{2\pi nit}$$

   is the best approximation to $f(t)$ in $L^2([0,1])$ 'of degree N' by a trigonometric polynomial.[3] (You can think of this as the least squares approximation. I'll explain the phrase 'of degree N' in Appendix 3, where we'll prove the statement.)

2. The complex exponentials $e^{2\pi nit}$, $n = 0, \pm 1, \pm 2, \dots$ form a basis for $L^2([0,1])$, and the partial sums, above, converge to $f(t)$ as $N \to \infty$ in the L^2-distance. This means that

   $$\lim_{N\to\infty} \left\| \sum_{n=-N}^{N} \hat{f}(n) e^{2\pi nit} - f(t) \right\| = 0$$

   We write

   $$f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi nit},$$

   where the equals sign is interpreted in terms of the limit.

   - Once we introduce the inner product on $L^2([0,1])$ a more complete statement will be that the $e^{2\pi nit}$ form an *orthonormal* basis. In fact, it's only the orthonormality that we'll establish.

3. The energy can be calculated from the Fourier coefficients:

   $$\int_0^1 |f(t)|^2\, dt = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

   This is known, depending on to whom you are speaking, as Rayleigh's identity or as Parseval's theorem.

To round off the picture, let me add a fourth point that's a sort of converse to items two and three. We won't use this, but it ties things up nicely.

[3] A trigonometric polynomial is a finite sum of complex exponentials.

4. If $\{c_n\}$, $n = 0, \pm 1, \pm 2, \ldots$ is any sequence of complex numbers for which

$$\sum_{n=-\infty}^{\infty} |c_n|^2 < \infty,$$

then the function

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n t}$$

is in $L^2([0,1])$ (meaning the limit of the partial sums converges to a function in $L^2([0,1])$) and $c_n = \hat{f}(n)$.

This last result is often referred to as the Riesz-Fischer theorem.

And the point of this is, again . . . One way to think of the formula for the Fourier coefficients is as passing from the 'time domain' to the 'frequency domain': From a knowledge of $f(t)$ (the time domain) we produce a portrait of the signal in the frequency domain, namely the (complex) coefficients $\hat{f}(n)$ associated with the (complex) harmonics $e^{2\pi i n t}$. The function $\hat{f}(n)$ is defined on the integers, $n = 0, \pm 1, \pm 2, \ldots$ and the equation

$$f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n) e^{2\pi i n t},$$

recovers the time domain representation from the frequency domain representation. At least it does in the L^2 sense of equality. The extent to which equality holds in the usual, pointwise sense (plug in a value of t and the two sides agree) is a question we will address later.

The magnitude $|\hat{f}(n)|^2$ is the energy contributed by the n'th harmonic. We really have equal contributions from the 'positive and negative' harmonics $e^{2\pi i n t}$ and $e^{-2\pi i n t}$ since $|\hat{f}(-n)| = |\hat{f}(n)|$ (note the absolute values here). As you will show in the first problem set, in passing between the complex exponential form

$$\sum_{n=-\infty}^{\infty} c_n e^{2\pi i n t}, \quad c_n = \hat{f}(n),$$

and the sine-cosine form

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos 2\pi n t + b_n \sin 2\pi n t$$

of the Fourier series we have $|c_n| = (1/2)\sqrt{a_n^2 + b_n^2}$, so $\hat{f}(n)$ and $\hat{f}(-n)$ together contribute a total energy of $\sqrt{a_n^2 + b_n^2}$.

Rayleigh's identity says that we can compute the energy of the signal by adding up the energies of the individual harmonics. That's quite a satisfactory state of affairs – and an extremely useful result. You'll see an example of its use in the first problem set.

Here are a few more general comments on these results.

- The first point, on best approximation in $L^2([0,1])$ by a finite sum, is a purely algebraic result. This is of practical value since, in any real application you're always making finite approximations, and this result gives guidance on how well you're doing. We'll have a more precise statement (in Appendix 3) after we set up the necessary ingredients on inner products and orthogonality.

Realize that this gives an alternative characterization of the Fourier coefficients. Originally we said: Supposing we can express $f(t)$ as a sum of complex exponentials, then the unknown coefficients in the expression must be given by the integral formula we found. Instead, we could have asked: What is the 'least squares' approximation to the function? And again we would be led to the same integral formula for the coefficients.

- Rayleigh's identity is also an algebraic result. Once we have the proper set-up it wil follow effortlessly.
- The remaining statements, points 2 and 4, involve some serious analysis and we won't go into the proofs. The crisp statements that we have given are true *provided* one adopts a more general theory of integration, Lebesgue's theory. In particular, one must allow for much wilder functions to be integrated than those that are allowed for the Riemann integral, which is the integral you saw in calculus. This is not to say that the Riemann integral is 'incorrect', rather it is incomplete – it does not allow for integrating functions that one needs to integrate *in order to get an adequate theory of Fourier series*, among other things.

  These are mathematical issues only. They have no practical value. To paraphrase John Tukey, a mathematician who helped to invent the FFT; 'I wouldn't want to fly in a plane whose design depended on whether a function was Riemann or Lebesgue integrable.'

So do you have to worry about this? Not really, but do take note of the examples we looked at in the previous lecture. Suppose a periodic signal has even a single discontinuity or a corner, like a square wave, a sawtooth wave or a triangle wave for example. Or think of taking a smooth signal and cutting it off (using a window), thus inducing a discontinuity or a corner. The Fourier series for such a signal *must* have infinitely many terms, and thus arbitrarily high frequencies in the spectrum. This is so, recall, because if

$$f(t) = \sum_{n=-N}^{N} \hat{f}(n)e^{2\pi nit}$$

for some finite N then $f(t)$ would be the finite sum of smooth functions, hence smooth itself. It's the possibility (the reality) of representing discontinuous (or wilder) functions by an infinite sum of smooth functions that's really quite a strong result. This was anticipated, and even stated by Fourier, but people didn't believe him. The results we've stated above are Fourier's vindication, but probably not in a form he would have recognized.

Orthogonality

The aspect of Euclidean geometry that sets it apart from geometries which share most of its other features is *perpendicularity* and its consequences. To set up a notion of perpendicularity in settings other than the familiar Euclidean plane or three dimensional space is to try to copy the Euclidean properties that go with it.

Perpendicularity becomes operationally useful, especially for applications, when it's linked to measurement, *i.e.* to length. This link is the Pythagorean theorem.[4] Perpendicularity becomes austere when mathematicians start referring to it as *orthogonality*, but that's what I'm used to and it's another term you can throw around to impress your friends.

[4]How do you lay out a big rectangular field of specified dimensions? You use the Pythagorean theorem. I had an encounter with this a few summers ago when I volunteered to help lay out soccer fields. I was only asked to assist, because evidently I could not be trusted with the details. Put two stakes in the ground to determine one side of the

Vectors To fix ideas, I want to remind you briefly of vectors and geometry in Euclidean space. We write vectors in $\mathbf{R}^n$ as n-tuples of real numbers:

$$\mathbf{v} = (v_1, v_2, \ldots, v_n)$$

The v_i are called the components of $\mathbf{v}$. The length, or *norm* of $\mathbf{v}$ is

$$||\mathbf{v}|| = (v_1^2 + v_2^2 + \cdots + v_n^2)^{1/2}.$$

The distance between two vectors $\mathbf{v}$ and $\mathbf{w}$ is $||\mathbf{v} - \mathbf{w}||$.

How does the Pythagorean theorem look in terms of vectors? Let's just work in $\mathbf{R}^2$. Let $\mathbf{u} = (u_1, u_2)$, $\mathbf{v} = (v_1, v_2)$ and $\mathbf{w} = \mathbf{u} + \mathbf{v} = (u_1 + v_1, u_2 + v_2)$. If $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ form a right triangle with $\mathbf{w}$ the hypotenuse then

$$\begin{aligned}
||\mathbf{w}||^2 = ||\mathbf{u} + \mathbf{v}||^2 &= ||\mathbf{u}||^2 + ||\mathbf{v}||^2 \\
(u_1 + v_1)^2 + (u_2 + v_2)^2 &= (u_1^2 + u_2^2) + (v_1^2 + v_2^2) \\
(u_1^2 + 2u_1v_1 + v_1^2) + (u_2^2 + 2u_2v_2 + v_2^2) &= u_1^2 + u_2^2 + v_1^2 + v_2^2
\end{aligned}$$

The squared terms cancel and we conclude that

$$u_1v_1 + u_2v_2 = 0$$

is a necessary and sufficient condition for $\mathbf{u}$ and $\mathbf{v}$ to be perpendicular.

And so we introduce the (algebraic) definition of the *inner product*, or *dot product* of two vectors. We give this in $\mathbf{R}^n$:

If $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ and $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ then the the inner product is:

$$\mathbf{v} \cdot \mathbf{w} = v_1w_1 + v_2w_2 + \cdots + v_nw_n$$

Other notations for the inner product are $(\mathbf{v}, \mathbf{w})$ (just parentheses; we'll be using this notation) and $\langle \mathbf{v}, \mathbf{w} \rangle$ (angle brackets; for those who think parentheses are not fancy enough. The use of angle brackets is especially common in physics where it's also used to denote more general 'parings' of vectors that produce real or complex numbers.)

Notice that

$$(\mathbf{v}, \mathbf{v}) = v_1^2 + v_2^2 + \cdots + v_n^2 = ||\mathbf{v}||^2.$$

Thus

$$||\mathbf{v}|| = (\mathbf{v}, \mathbf{v})^{1/2}.$$

There is also a geometric approach to the inner product which leads to the formula

$$(\mathbf{v}, \mathbf{w}) = ||\mathbf{v}||\,||\mathbf{w}|| \cos\theta$$

field. That's one leg of what is to become a right triangle – half the field. I hooked a tape measure on one stake and walked off in a direction generally perpendicular to the first leg, stopping when I had gone the regulation distance for that side of the field, or when I needed rest. The chief of the crew hooked another tape measure on the other stake and walked approximately along the diagonal of the field – the hypotenuse. We adjusted our positions – but not the length we had walked off – to meet up, so that the Pythagorean theorem was satisfied; he had a chart showing what his distance should be. Hence at our meeting point the leg I determined must be perpendicular to the first leg we laid out. This was my first practical use of the Pythagorean theorem, and so began my transition from a pure mathematician to an engineer.

where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$. This is sometimes taken as an alternate definition of the inner product, though we'll stick with the algebraic definition. For a few comments on this see Appendix 1.

We see that $(\mathbf{v}, \mathbf{w}) = 0$ if and only if $\mathbf{v}$ and $\mathbf{w}$ are orthogonal. This was the point, after all, and it is a truly helpful result, especially because it's so easy to verify when the vectors are given in coordinates. The inner product does more than identify orthogonal vectors, however. When it's nonzero it tells you how much of one vector is in the direction of another. That is, the vector

$$\frac{(\mathbf{v}, \mathbf{w})}{||\mathbf{w}||} \frac{\mathbf{w}}{||\mathbf{w}||} \quad \text{also written as} \quad \frac{(\mathbf{v}, \mathbf{w})}{(\mathbf{w}, \mathbf{w})}\mathbf{w}$$

is the projection of $\mathbf{v}$ onto the unit vector $\mathbf{w}/||\mathbf{w}||$, or, if you prefer, $(\mathbf{v}, \mathbf{w})/||\mathbf{w}||$ is the (scalar) component of $\mathbf{v}$ in the direction of $\mathbf{w}$. I think of the inner product as measuring how much one vector 'knows' another; two orthogonal vectors don't know each other.

Finally, I want to list the main algebraic properties of the inner product, I won't give the proofs – they are straightforward verifications. We'll see these properties again, modified slightly to allow for complex numbers, a little later.

1. $(\mathbf{v}, \mathbf{v}) \geq 0$ and $(\mathbf{v}, \mathbf{v}) = 0$ if and only if $\mathbf{v} = 0$. (positive definiteness)
2. $(\mathbf{v}, \mathbf{w}) = (\mathbf{w}, \mathbf{v})$ (symmetry)
3. $(\alpha\mathbf{v}, \mathbf{w}) = \alpha(\mathbf{v}, \mathbf{w})$ for any scalar α. (homogeneity)
4. $(\mathbf{v} + \mathbf{w}, \mathbf{u}) = (\mathbf{v}, \mathbf{u}) + (\mathbf{w}, \mathbf{u})$ (additivity)

In fact, these are exactly the properties that ordinary multiplication has.

Orthonormal bases The natural basis for $\mathbf{R}^n$ are the vectors of length 1 in the n 'coordinate directions':

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, \ldots, 0), \ldots, \ \mathbf{e}_n = (0, 0, \ldots, 1).$$

These vectors are called the 'natural' basis because a vector $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ is expressed 'naturally' in terms of its components as

$$\mathbf{v} = v_1\mathbf{e}_1 + v_2\mathbf{e}_2 = \cdots + v_n\mathbf{e}_n.$$

One says that the natural basis $\mathbf{e}_1$, $\mathbf{e}_2$, $\ldots$, $\mathbf{e}_n$ are an *orthonormal basis* for $\mathbf{R}^n$, meaning

$$(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij},$$

where δ_{ij} is the *Kronecker delta* defined by

$$\delta_{ij} = \begin{cases} 1, & \text{if} \quad i = j \\ 0, & if \quad i \neq j \end{cases}$$

Notice that

$$(\mathbf{v}, \mathbf{e}_k) = v_k,$$

and hence that

$$\mathbf{v} = \sum_{k=1}^{n} (\mathbf{v}, \mathbf{e}_k)\mathbf{e}_k.$$

In words:

$\mathbf{v}$ is decomposed as a sum of vectors in the directions of the orthonormal basis vectors, and the components are given by the inner product of $\mathbf{v}$ with the basis vectors.

Since the $\mathbf{e}_k$ have length 1, the inner products $(\mathbf{v}, \mathbf{e}_k)$ are the projections of $\mathbf{v}$ onto the basis vectors.[5]

Functions All of what we've just done can be carried over to $L^2([0,1])$, including the same motivation for orthogonality and defining the inner product. When will two functions be 'perpendicular'? If the Pythagorean theorem is satisfied. Thus if we are to have

$$||f + g||^2 = ||f||^2 + ||g||^2$$

then

$$\begin{aligned}
\int_0^1 (f(t) + g(t))^2 &= \int_0^1 f(t)^2\,dt + \int_0^1 g(t)^2\,dt \\
\int_0^1 (f(t)^2 + 2f(t)g(t) + g(t)^2)\,dt &= \int_0^1 f(t)^2\,dt + \int_0^1 g(t)^2\,dt \\
\int_0^1 f(t)^2\,dt + 2\int_0^1 f(t)g(t)\,dt + \int_0^1 g(t)^2\,dt &= \int_0^1 f(t)^2\,dt + \int_0^1 g(t)^2\,dt
\end{aligned}$$

If you buy the premise, you have to buy the conclusion – we conclude that the condition to adopt to define when two functions are perpendicular (or as we'll now say, orthogonal) is

$$\int_0^1 f(t)g(t)\,dt = 0$$

So we *define* the inner product of two functions in $L^2([0,1])$ to be.

$$(f, g) = \int_0^1 f(t)g(t)\,dt.$$

(See Appendix 1 for a discussion of why $f(t)g(t)$ is integrable if $f(t)$ and $g(t)$ are each square integrable.)

This inner product has all of the algebraic properties of the dot product of vectors. We list them, again:

1. $(f, f) \geq 0$ and $(f, f) = 0$ if and only if $f = 0$.
2. $(f, g) = (g, f)$
3. $(f + g, h) = (f, h) + (g, h)$
4. $(\alpha f, g) = \alpha(f, g)$

We have

$$(f, f) = \int_0^1 f(t)^2\,dt = ||f||^2.$$

Now, let me relieve you of a burden that you may feel you must carry. There is no reason on earth why you should have any pictorial intuition for the inner product of two functions, and for

[5] Put that other way I like so much, the inner product $(\mathbf{v}, \mathbf{e}_k)$ is how much $\mathbf{v}$ and $\mathbf{e}_k$ know each other.

when two functions are orthogonal. How can you picture the condition $(f, g) = 0$? In terms of the graphs of f and g? I don't think so. And if (f, g) is not zero, how are you to picture how much f and g know each other? Don't be silly.

We're working by analogy here. It's a very strong analogy, but that's not to say that the two settings – functions and geometric vectors – are identical. They aren't. As I have said before, what you should do is draw pictures in $\mathbf{R}^2$ and $\mathbf{R}^3$, see, somehow, what algebraic or geometric idea may be called for, and *using the same words* make the attempt to carry that over to $L^2([0,1])$. It's surprising how often and how well this works.

There's a catch There's always a catch. In the preceding discussion we've been working with the *real* vector space $\mathbf{R}^n$, as motivation, and with real-valued functions in $L^2([0,1])$. But, of course, the definition of the Fourier coefficients involves complex functions in the form of the complex exponential, and the Fourier series is a sum of complex terms. We could avoid this catch by writing everything in terms of sine and cosine, a procedure you may have followed in an earlier course. However, we don't want to sacrifice the algebraic dexterity we can show by working with the complex form of the Fourier sums, and a more effective and encompassing choice is to consider *complex-valued* square integrable functions and the *complex inner product.*

Here are the definitions. For the definition of $L^2([0,1])$ we assume again that

$$\int_0^1 |f(t)|^2 \, dt < \infty.$$

The definition looks the same as before, but $|f(t)|^2$ is now the magnitude of the (possibly) complex number $f(t)$.

The inner product of complex valued functions $f(t)$ and $g(t)$ in $L^2([0,1])$ is defined to be

$$(f, g) = \int_0^1 f(t)\overline{g(t)} \, dt.$$

The complex conjugate in the second slot causes a few changes in the algebraic properties. To wit:

1. $(f, g) = \overline{(g, f)}$ (*Hermitian* symmetry)
2. $(f, f) \geq 0$ and $(f, f) = 0$ if and only if $f = 0$ (positive definiteness – same as before)
3. $(\alpha f, g) = \alpha(f, g)$, $(f, \alpha g) = \overline{\alpha}(f, g)$ (homogeneity – same as before in the first slot, conjugate scalar comes out if it's in the second slot)
4. $(f + g, h) = (f, h) + (g, h)$, $(f, g + h) = (f, g) + (f, h)$ (additivity – same as before, no difference between additivity in first or second slot)

I'll say more about the reason for the definition in Appendix 2. As before,

$$(f, f) = \int_0^1 f(t)\overline{f(t)} \, dt = \int_0^1 |f(t)|^2 \, dt = ||f||^2.$$

> From now on, when we talk about $L^2([0,1])$ and the inner product on $L^2([0,1])$ we will always assume the complex inner product. If the functions happen to be real-valued then this reduces to the earlier definition.

The complex exponentials are an orthonormal basis Number two in our list of the greatest hits of the theory of Fourier series says that the complex exponentials form a basis for $L^2([0,1])$. This is *not* a trivial statement. In many ways it's the whole ball game, for in establishing this fact one sees why $L^2([0,1])$ is the natural space to work with, and why convergence in $L^2([0,1])$ is the right thing to ask for in asking for the convergence of the partial sums of Fourier series.[6] But it's too much for us to do.

Instead, we'll be content with the news that, just like the natural basis of $\mathbf{R}^n$, the complex exponentials are *orthonormal*. Here's the calculation; in fact, it's the same calculation we did when we first solved for the Fourier coefficients.

Write

$$e_n(t) = e^{2\pi int}.$$

The inner product of two of them, $e_n(t)$ and $e_m(t)$, when $n \neq m$ is:

$$\begin{aligned}
(e_n, e_m) &= \int_0^1 e^{2\pi nit}\overline{e^{2\pi imt}}\,dt \\
&= \int_0^1 e^{2\pi nit}e^{-2\pi imt}\,dt \\
&= \int_0^1 e^{2\pi i(n-m)t}\,dt \\
&= \frac{1}{2\pi i(n-m)}e^{2\pi i(n-m)t}\Big]_0^1 \\
&= \frac{1}{2\pi i(n-m)}(e^{2\pi i(n-m)} - e^0) \\
&= \frac{1}{2\pi i(n-m)}(1-1) \\
&= 0
\end{aligned}$$

They are orthogonal. And when $n = m$

$$\begin{aligned}
(e_n, e_n) &= \int_0^1 e^{2\pi nit}\overline{e^{2\pi int}}\,dt \\
&= \int_0^1 e^{2\pi nit}e^{-2\pi int}\,dt \\
&= \int_0^1 e^{2\pi i(n-n)t}\,dt \\
&= \int_0^1 1\,dt = 1.
\end{aligned}$$

Therefore the functions $e_n(t)$ are *orthonormal*:

$$(e_n, e_m) = \delta_{nm} = \begin{cases} 1, & \text{if} \quad n = m \\ 0, & \text{if} \quad n \neq m \end{cases}$$

[6] An important point in this development is understanding what happens to the usual kind of pointwise convergence *vis a vis* $L^2([0,1])$ convergence when the functions are smooth enough.

What is the component of a function $f(t)$ 'in the direction' $e_n(t)$? By analogy to the Euclidean case, it is given by the inner product

$$(f, e_n) = \int_0^1 f(t)\overline{e_n(t)}\, dt = \int_0^1 f(t)e^{-2\pi int}\, dt,$$

precisely the n'th Fourier coefficient $\hat{f}(n)$. (Note that e_n really does have to be in the second slot here.)

Thus writing the Fourier series

$$f = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{2\pi nit},$$

as we did earlier, is exactly like the decomposition in terms of an orthonormal basis and associated inner product:

$$f = \sum_{n=-\infty}^{\infty} (f, e_n)e_n.$$

What we haven't done is to show that this really works – that the complex exponentials are a *basis* as well as being orthonormal. We would be required to show that

$$\lim_{N\to\infty} ||f - \sum_{n=-N}^{N} (f, e_n)e_n|| = 0.$$

We're not going to do that. It's hard.

What if the period isn't 1? Remember how we modified the Fourier series when the period is T rather than 1. We were led to the expansion

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{2\pi int/T}.$$

where

$$c_n = \frac{1}{T}\int_0^T e^{-2\pi int/T} f(t)\, dt.$$

The whole set-up we've just been through can be easily modified to cover this case. We work in the space $L^2([0, T])$ of square integrable functions on the interval $[0, T]$. The inner product (complex) is

$$(f, g) = \int_0^T f(t)\overline{g(t)}\, dt.$$

What happens with the T-periodic complex exponentials $e^{2\pi int/T}$?. If $n \neq m$ then, much as before,

$$\begin{aligned}
(e^{2\pi int/T}\,,\,e^{2\pi imt/T}) &= \int_0^T e^{2\pi int/T}\,\overline{e^{2\pi imt/T}}\,dt \\
&= \int_0^T e^{2\pi int/T}\,e^{-2\pi imt/T}\,dt \\
&= \int_0^T e^{2\pi i(n-m)t/T}\,dt \\
&= \frac{1}{2\pi i(n-m)/T}\,e^{2\pi i(n-m)t/T}\Big]_0^T \\
&= \frac{1}{2\pi i(n-m)/T}\,(e^{2\pi i(n-m)} - e^0) \\
&= \frac{1}{2\pi i(n-m)/T}\,(1-1) \\
&= 0
\end{aligned}$$

And when $n = m$:

$$\begin{aligned}
(e^{2\pi int/T}\,,\,e^{2\pi int/T}) &= \int_0^T e^{2\pi int/T}\,\overline{e^{2\pi int/T}}\,dt \\
&= \int_0^T e^{2\pi int/T}\,e^{-2\pi int/T}\,dt \\
&= \int_0^T 1\,dt \\
&= T
\end{aligned}$$

Aha – it's not 1, it's T. The complex exponentials are orthogonal but not orthonormal. To get the latter property we scale the complex exponentials to

$$e_n(t) = \frac{1}{\sqrt{T}}e^{2\pi int/T},$$

for then

$$(e_n, e_m) = \begin{cases} 1 & , \quad n = m \\ 0 & , \quad n \neq m. \end{cases}$$

This is where the factor $1/\sqrt{T}$ comes from, the factor mentioned in the previous lecture. The inner product of f with e_n is

$$(f, e_n) = \frac{1}{\sqrt{T}}\int_0^T f(t)e^{-2\pi int/T}\,dt.$$

and then

$$\begin{aligned}
\sum_{n=-\infty}^{\infty}(f, e_n)e_n &= \sum_{n=-\infty}^{\infty}\left(\frac{1}{\sqrt{T}}\int_0^T f(s)e^{-2\pi ins/T}\,ds\right)\frac{1}{\sqrt{T}}e^{2\pi int/T} \\
&= \sum_{n=-\infty}^{\infty} c_n e^{2\pi int/T},
\end{aligned}$$

where

$$c_n = \frac{1}{T}\int_0^T e^{-2\pi int/T} f(t)\,dt,$$

as above. We're back to our earlier formula.

Rayleigh's identity As a last application of these ideas let's derive Rayleigh's identity. It states that

$$\int_0^1 |f(t)|^2\,dt = \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2.$$

This is a cinch! Expand $f(t)$ is a Fourier series:

$$f(t) = \sum_{n=-\infty}^{\infty} \hat{f}(n)e^{2\pi nit} = \sum_{n=-\infty}^{\infty} (f, e_n)e_n.$$

Then

$$\begin{aligned}
\int_0^1 |f(t)|^2\,dt &= ||f||^2 = (f,f) \\
&= \left(\sum_{n=-\infty}^{\infty} (f,e_n)e_n, \sum_{m=-\infty}^{\infty} (f,e_m)e_m\right) \\
&= \sum_{n,m} (f,e_n)\overline{(f,e_m)}(e_n,e_m) \\
&= \sum_{n,m=-\infty}^{\infty} (f,e_n)\overline{(f,e_m)}\delta_{nm} \\
&= \sum_{n=-\infty}^{\infty} (f,e_n)\overline{(f,e_n)} \\
&= \sum_{n=-\infty}^{\infty} |(f,e_n)|^2 \\
&= \sum_{n=-\infty}^{\infty} |\hat{f}(n)|^2
\end{aligned}$$

This used

1. The algebraic properties of the complex inner product;
2. The fact that the $e_n(t) = e^{2\pi nit}$ are orthonormal with respect to this inner product;
3. Know-how in whipping around sums

Do not go to sleep until you can follow every line in this derivation.

Writing Rayleigh's identity as

$$||f||^2 = \sum_{n=-\infty}^{\infty} |(f,e_n)|^2$$

again highlights the parallels between the geometry of L^2 and the geometry of vectors: How do you find the squared length of a vector? By adding the squares of its components with respect to an orthonormal basis. That's exactly what Rayleigh's identity is saying.

Appendix 1: The Cauchy-Schwarz inequality and its consequences

The Cauchy-Schwarz inequality is between the inner product of two vectors and their norms. It states

$$|(\mathbf{v}, \mathbf{w})| \leq ||\mathbf{v}||\, ||\mathbf{w}||.$$

This is trivial to see from the geometric formula for the inner product:

$$|(\mathbf{v}, \mathbf{w})| = ||\mathbf{v}||\, ||\mathbf{w}||\, |\cos\theta| \leq ||\mathbf{v}||\, ||\mathbf{w}||,$$

because $|\cos\theta| \leq 1$. In fact, the rationale for the geometric formula of the inner product will *follow* from the Cauchy-Schwarz inequality.

It's certainly not obvious how to derive the inequality from the algebraic definition. Written out in components, the inequality says that

$$\left|\sum_{k=1}^{n} v_k w_k\right| \leq \left\{\sum_{k=1}^{n} v_k^2\right\}^{1/2} \left\{\sum_{k=1}^{n} w_k^2\right\}^{1/2}.$$

Sit down and try that one out sometime.

In fact, the proof of the Cauchy-Schwarz inequality in general uses only the four algebraic properties of the inner product listed earlier. Consequently the same argument applies to any sort of 'product' satisfying these properties. It's such an elegant argument (due to John von Neumann, I believe) that I'd like to show it to you. We'll give this for the real inner product here, with comments on the complex case to follow in the next appendix.

Any inequality can ultimately be written in a way that says that some quantity is positive. There aren't many things that we know are positive: the square of a real number; the area of something; and the length of something are examples.[7] For this proof we use that the norm of a vector is positive, but we throw in a parameter.[8] Let t be any real number. Then $||\mathbf{v} - t\mathbf{w}||^2 \geq 0$. Write this in terms of the inner product and expand out using the algebraic properties; because of homogeneity, symmetry, and additivity, this is just like multiplication – that's important to realize:

$$\begin{aligned} 0 &\leq ||\mathbf{v} - t\mathbf{w}||^2 \\ &= (\mathbf{v} - t\mathbf{w}, \mathbf{v} - t\mathbf{w}) \\ &= (\mathbf{v}, \mathbf{v}) - 2t(\mathbf{v}, \mathbf{w}) + t^2(\mathbf{w}, \mathbf{w}) \\ &= ||\mathbf{v}||^2 - 2t(\mathbf{v}, \mathbf{w}) + t^2||\mathbf{w}||^2 \end{aligned}$$

This is a quadratic equation in t, of the form $at^2 + bt + c$, where $a = ||\mathbf{w}||^2$, $b = -2(\mathbf{v}, \mathbf{w})$, and $c = ||\mathbf{v}||^2$. The first inequality, and the chain of equalities that follow, says that this quadratic is *always non-negative*. Now a quadratic that's always non-negative has to have a *non-positive* discriminant: The discriminant, $b^2 - 4ac$ determines the nature of the roots of the quadratic – if the discriminant is positive then there are two real roots, but if there are two real roots, then the quadratic must be negative somewhere.

Therefore $b^2 - 4ac \leq 0$, which translates to

$$\begin{aligned} &4(\mathbf{v}, \mathbf{w})^2 - 4||\mathbf{w}||^2\, ||\mathbf{v}||^2 \leq 0, \quad \text{or} \\ &(\mathbf{v}, \mathbf{w})^2 \leq ||\mathbf{w}||^2\, ||\mathbf{v}||^2. \end{aligned}$$

[7] This little riff on the nature of inequalities qualifies as a minor secret of the universe. More subtle inequalities sometimes rely on convexity, as in the center of gravity of a system of masses is contained within the convex hull of the masses.

[8] 'Throwing in a parameter' goes under the heading of dirty tricks of the universe.

Take the square root of both sides to obtain

$$|(\mathbf{v},\mathbf{w})| \leq ||\mathbf{v}||\,||\mathbf{w}||,$$

as desired. (Amazing, isn't it – a non-trivial application of the *quadratic formula*!)[9] This proof also shows when equality holds in the Cauchy-Schwarz inequality. When is that?

To get back to geometry, we now know that

$$-1 \leq \frac{(\mathbf{v},\mathbf{w})}{||\mathbf{v}||\,||\mathbf{w}||} \leq 1.$$

Therefore there is a unique angle θ with $0 \leq \theta \leq \pi$ such that

$$\cos\theta = \frac{(\mathbf{v},\mathbf{w})}{||\mathbf{v}||\,||\mathbf{w}||},$$

i.e.

$$(\mathbf{v},\mathbf{w}) = ||\mathbf{v}||\,||\mathbf{w}||\cos\theta.$$

Identifying θ as the angle between $\mathbf{v}$ and $\mathbf{w}$ we have now reproduced the geometric formula for the inner product. What a relief.

The triangle inequality,

$$||\mathbf{v}+\mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$$

follows directly from the Cauchy-Schwarz inequality. Here's the argument:

$$\begin{aligned} ||\mathbf{v}+\mathbf{w}||^2 &= (\mathbf{v}+\mathbf{w},\mathbf{v}+\mathbf{w}) \\ &= (\mathbf{v},\mathbf{v}) + 2(\mathbf{v},\mathbf{w}) + (\mathbf{w},\mathbf{w}) \\ &\leq (\mathbf{v},\mathbf{v}) + 2|(\mathbf{v},\mathbf{w})| + (\mathbf{w},\mathbf{w}) \\ &\leq (\mathbf{v},\mathbf{v}) + 2||\mathbf{v}||\,||\mathbf{w}|| + (\mathbf{w},\mathbf{w}) \quad \text{(by Cauchy-Schwarz)} \\ &= ||\mathbf{v}||^2 + 2||\mathbf{v}||\,||\mathbf{w}|| + ||\mathbf{w}||^2 \\ &= (||\mathbf{v}|| + ||\mathbf{w}||)^2. \end{aligned}$$

Now take the square root of both sides to get $||\mathbf{v}+\mathbf{w}|| \leq ||\mathbf{v}|| + ||\mathbf{w}||$. In coordinates this says that

$$\left\{\sum_{k=1}^{n}(v_k+w_k)^2\right\}^{1/2} \leq \left\{\sum_{k=1}^{n} v_k^2\right\}^{1/2} + \left\{\sum_{k=1}^{n} w_k^2\right\}^{1/2}.$$

For the inner product on $L^2([0,1])$ the Cauchy-Schwarz inequality takes the impressive form

$$\left|\int_0^1 f(t)g(t)\,dt\right| \leq \left\{\int_0^1 f(t)^2\,dt\right\}^{1/2} \left\{\int_0^1 g(t)^2\,dt\right\}^{1/2}.$$

You can think of this as a limiting case of the Cauchy-Schwarz inequality for vectors – sums of products become integrals of products on taking limits, an ongoing theme – but it's better to think in terms of general inner products and their properties. For example, we now also know that

$$||f+g|| \leq ||f|| + ||g||,$$

[9] As a slight alternative to this argument, if the quadratic $f(t) = at^2 + bt + c$ is non-negative then, in particular, its minimum value is non-negative. This minimum occurs at $t = 0$ $(a > 0)$ and leads to the same inequality, $4ac - b^2 \geq 0$.

i.e. that

$$\left\{ \int_0^1 (f(t) + g(t))^2 \, dt \right\}^{1/2} \le \left\{ \int_0^1 f(t)^2 \, dt \right\}^{1/2} + \left\{ \int_0^1 g(t)^2 \, dt \right\}^{1/2}.$$

Once again, one could, I suppose, derive this from the corresponding inequality for sums, but why keep gong through that extra work?

Incidentally, I have skipped over something here. If $f(t)$ and $g(t)$ are square integrable, then in order to get the Cauchy-Schwarz inequality working one has to know that the inner product (f, g) makes sense, *i.e.* that

$$\int_0^1 f(t)g(t) \, dt < \infty.$$

(This isn't an issue for vectors in $\mathbf{R}^n$, of course. Here's an instance when something more needs to be said for the case of functions.) To deduce this you can first observe that

$$f(t)g(t) \le f(t)^2 + g(t)^2.$$

With this

$$\int_0^1 f(t)g(t) \, dt \le \int_0^1 f(t)^2 \, dt + \int_0^1 g(t)^2 \, dt < \infty,$$

since we started by assuming that $f(t)$ and $g(t)$ are square integrable.[10]

Another consequence of this last argument is the fortunate fact that the Fourier coefficients of a function in $L^2([0,1])$ exist. That is, we're wondering about the existence of

$$\int_0^1 e^{-2\pi i n t} f(t) \, dt,$$

allowing for integrating complex functions. Now,

$$\left| \int_0^1 e^{-2\pi i n t} f(t) \, dt \right| \le \int_0^1 |e^{-2\pi i n t} f(t)| \, dt = \int_0^1 |f(t)| \, dt,$$

so we're wondering whether

$$\int_0^1 |f(t)| \, dt < \infty,$$

i.e. is $f(t)$ *absolutely integrable* given that it is *square* integrable. But $f(t) = f(t) \times 1$, and both $f(t)$ and the constant function 1 are square integrable on $[0,1]$, so the result follows from Cauchy-Schwartz. We wonder no more.

> **Warning:** This casual argument *would not work* if the interval $[0,1]$ were replaced by the entire real line. The constant function 1 has an infinite integral on $\mathbf{R}$. You may think we can get around this little inconvenience, but it is *exactly* the sort of trouble that comes up in trying to apply Fourier *series* ideas (where functions are defined on finite intervals) to Fourier *transform* ideas (where functions are defined on all of $\mathbf{R}$).

[10] And where does that little observation come from? From the same positivity trick used to prove Cauchy-Schwarz:

$$0 \le (f(t) - g(t))^2 = f(t)^2 - 2f(t)g(t) + g(t)^2,$$

hence

$$2f(t)g(t) \le f(t)^2 + g(t)^2.$$

This is the inequality between the arithmetic and geometric mean.

Appendix 2: More on the complex inner product

Here's an argument why the conjugate comes in in defining a complex inner product. Let's go right to the case of integrals. What if we apply the Pythagorean Theorem to deduce the condition for perpendicularity in the complex case, just as we did in the real case? We have

$$\begin{aligned}
\int_0^1 |f(t)+g(t)|^2 &= \int_0^1 |f(t)|^2\,dt + \int_0^1 |g(t)|^2\,dt \\
\int_0^1 (|f(t)|^2 + 2\,\mathrm{Re}\{f(t)\overline{g(t)}\} + |g(t)|^2)\,dt &= \int_0^1 |f(t)|^2\,dt + \int_0^1 |g(t)|^2\,dt \\
\int_0^1 |f(t)|^2\,dt + 2\,\mathrm{Re}\left\{\int_0^1 f(t)\overline{g(t)}\,dt\right\} + \int_0^1 |g(t)|^2\,dt &= \int_0^1 |f(t)|^2\,dt + \int_0^1 |g(t)|^2\,dt
\end{aligned}$$

So it looks like the condition should be

$$\mathrm{Re}\left\{\int_0^1 f(t)\overline{g(t)}\,dt\right\} = 0.$$

Why doesn't this determine the definition of the inner product of two complex functions? That is, why don't we define

$$(f,g) = \mathrm{Re}\left\{\int_0^1 f(t)\overline{g(t)}\,dt\right\}?$$

This definition has a nicer symmetry property, for example, than the definition we used earlier. Here we have

$$(f,g) = \mathrm{Re}\left\{\int_0^1 f(t)\overline{g(t)}\,dt\right\} = \mathrm{Re}\left\{\int_0^1 \overline{f(t)}g(t)\,dt\right\} = (g,f);$$

so none of that Hermitian symmetry that we always have to remember.

The problem is that this definition doesn't give any kind of homogeneity when multiplying by a *complex* scalar. If α is a complex number then

$$(\alpha f, g) = \mathrm{Re}\left\{\int_0^1 \alpha f(t)\overline{g(t)}\,dt\right\} = \mathrm{Re}\left\{\alpha\int_0^1 f(t)\overline{g(t)}\,dt\right\}.$$

But we can't pull the α out of taking the real part unless it's real to begin with. If α is not real then

$$(\alpha f, g) \neq \alpha(f,g).$$

Not having equality here is too much to sacrifice. (Nor do we have anything good for $(f, \alpha g)$, despite the natural symmetry $(f,g) = (g,f)$.) We adopt the definition

$$(f,g) = \int_0^1 f(t)\overline{g(t)}\,dt.$$

A helpful identity A frequently employed identity for the complex inner product is:

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

We more or less used this, above, and I wanted to single it out. The verification is:

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

Similarly,

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

Here's how to get the Cauchy-Schwarz inequality for complex inner products from this. The inequality states

$$|(f,g)| \le ||f||\,||g||,$$

and on the left hand side we have the magnitude of the (possibly) complex number (f,g). As a slight twist on what we did in the real case, let $\alpha = te^{i\theta}$ be a complex number (t real) and consider

$$\begin{aligned} 0 \le ||f-\alpha g||^2 &= ||f||^2 - 2\,\mathrm{Re}(f,\alpha g) + ||\alpha g||^2 \\ &= ||f||^2 - 2\,\mathrm{Re}\{\overline{\alpha}(f,g)\} + ||\alpha g||^2 \\ &= ||f||^2 - 2t\,\mathrm{Re}\{e^{-i\theta}(f,g)\} + t^2||g||^2 \end{aligned}$$

Now we can choose θ here, and we do so to make

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

Multiplying (f,g) by $e^{-i\theta}$ rotates the complex number (f,g) clockwise by θ, so choose θ to rotate (f,g) to be real and positive. From here the argument is the same as it was in the real case.

It's worth writing out the Cauchy-Schwarz inequality in terms of integrals:

$$\left|\int_0^1 f(t)\overline{g(t)}\,dt\right| \le \left\{\int_0^1 |f(t)|^2\,dt\right\}^{1/2} \left\{\int_0^1 |g(t)|^2\,dt\right\}^{1/2}.$$

Appendix 3: Best L^2 approximation by finite Fourier series

Here's a precise statement, and a proof, that a finite Fourier series of degree N gives the best (trigonometric) approximation of that order in $L^2([0,1])$ to a function.

Theorem If $f(t)$ is in $L^2([0,1])$ and $\alpha_1, \alpha_2, \ldots, \alpha_N$ are any complex numbers, then

$$\left\| f - \sum_{n=-N}^{N} (f,e_n)e_n \right\| \le \left\| f - \sum_{n=-N}^{N} \alpha_n e_n \right\|$$

Furthermore, equality holds here only when $\alpha_n = (f,e_n)$ for all n.

It's the last statement, on the case of equality, that leads to the Fourier coefficients in a different way than solving for them directly as we did originally. Another way of stating the result is that the *orthogonal projection* of f onto the subspace of *Ltwo* spanned by the e_n, $n = -N, \ldots, N$ is

$$\sum_{n=-N}^{N} \hat{f}(n)e^{2\pi i n t}.$$

Here comes the proof. Hold on. Write

$$
\begin{aligned}
\left\| f - \sum_{n=-N}^{N} \alpha_n e_n \right\|^2 &= \left\| f - \sum_{n=-N}^{N} (f, e_n)e_n + \sum_{n=-N}^{N} (f, e_n)e_n - \sum_{n=-N}^{N} \alpha_n e_n \right\|^2 \\
&= \left\| \left(f - \sum_{n=-N}^{N} (f, e_n)e_n \right) + \sum_{n=-N}^{N} ((f, e_n) - \alpha_n)e_n \right\|^2
\end{aligned}
$$

We squared all the norms because we want to use the properties of inner products to expand the last line. Using the identity we derived earlier, it's equal to:

$$
\begin{aligned}
\left\| \left(f - \sum_{n=-N}^{N} (f, e_n)e_n \right) + \sum_{n=-N}^{N} ((f, e_n) - \alpha_n)e_n \right\|^2 &= \left\| f - \sum_{n=-N}^{N} (f, e_n)e_n \right\|^2 \\
&\quad + 2\,\mathrm{Re} \left(f - \sum_{n=-N}^{N} (f, e_n)e_n, \sum_{m=-N}^{N} ((f, e_m) - \alpha_m)e_m \right) \\
&\quad + \left\| \sum_{n=-N}^{N} ((f, e_n) - \alpha_n)e_n \right\|^2
\end{aligned}
$$

This looks complicated, but the middle term is just a sum of multiples of terms of the form

$$
\begin{aligned}
\left(f - \sum_{n=-N}^{N} (f, e_n)e_n, e_m \right) &= (f, e_m) - \sum_{n=-N}^{N} (f, e_n)(e_n, e_m) \\
&= (f, e_m) - (f, e_m) = 0,
\end{aligned}
$$

so the whole thing drops out! The final term is

$$
\left\| \sum_{n=-N}^{N} ((f, e_n) - \alpha_n)e_n \right\|^2 = \sum_{n=-N}^{N} |(f, e_n) - \alpha_n|^2.
$$

We are left with

$$
\left\| f - \sum_{n=-N}^{N} \alpha_n e_n \right\|^2 = \left\| f - \sum_{n=-N}^{N} (f, e_n)e_n \right\|^2 + \sum_{n=-N}^{N} |(f, e_n) - \alpha_n|^2.
$$

This completely proves the theorem, for the right hand side is the sum of two positive terms and hence

$$
\left\| f - \sum_{n=-N}^{N} \alpha_n e_n \right\|^2 \geq \left\| f - \sum_{n=-N}^{N} (f, e_n)e_n \right\|^2
$$

with equality holding if and only if

$$
\sum_{n=-N}^{N} |(f, e_n) - \alpha_n|^2 = 0.
$$

The latter holds if and only if $\alpha_n = (f, e_n)$ for all n.

The preceding argument may have seemed labor intensive, but it was all *algebra* based on the properties of the inner product. Imagine trying to write all of it out in terms of integrals.