HERBERT GROSS: Hi. Welcome to our final lecture in part three of calculus. And I suppose it would be nice to have an appropriate story to tell. And the only story I can think of is, on the transatlantic flight, the stewardess coming in to the passenger's cabin midway over the Atlantic Ocean and saying to the passengers, well, I have some good news and some bad news for you. The bad news is that we're almost out of gas. And the good news is that we're making wonderful time.

And the reason I thought of this particular story is that, in a sense, we are not almost out of gas, but in terms of this being a fundamentals type of course in which we're trying to give the ingredients of what goes into all advanced mathematics courses, we have sort of come to the end of the road. We have finished the so-called ABC's of mathematics. As far as making wonderful time is concerned, I would say we've done deliberately just the opposite, that we have essentially taken one or two central themes and tried to unify all of elementary calculus in terms of these central themes. And when you do something like this, it's very difficult to come to the end of a course. You sort of don't know where to stop.

So you go back to a little trick that you learned in about the third grade when you wrote compositions where you learned that you could write anything that you wanted as long as the last line was, we all had a very nice time. So we now must look for a line along those particular thoughts. And if we recall that we began as our major theme in this course the idea of mathematical structure, the idea of what new math really meant, not new meaning brand new, but new meaning the opposite of passe, new meaning meaningful, I can think of no nicer way of ending the course than by picking as an application a topic that existed long before the so-called new math was born, and which exists in an even more enlightened way in the spirit of the new mathematics. And the topic that I have in mind to conclude with-- and it also ties in with our lecture of last time on the dot product-- is the subject called orthogonal functions, which has as a special subtopic, Fourier series.

Now the way this comes about is the following. And you have to know nothing about vector spaces to appreciate what we're going to do now, that in a sense we imitate what Fourier actually did in trying to solve a particular type of problem. We suppose that we have an infinite sequence of, functions which I'll denote by the family phi sub n of x, and a particular closed interval from a to b such that each member of my family phi sub n of x is integral on the interval from a to b.

And we'll also assume that the integral from a to b of the product of any two of these functions, two different members of the set, that that integral is 0. In other words, the interval from a to b phi sub i of x times phi sub j of x dx is 0, provided i is unequal to j. So we're assuming this particular property, remembering of course that the product of two integral functions is, again, integrable so that this thing makes sense.

Now traditionally, when a family of functions had this particular property, that family was set to be orthogonal on the interval from a to b. And I will leave it for a little bit later to show you how this jibes with the meaning of orthogonal as we defined it in terms of dot products last time. Let me show you how, for example, Fourier decided to use this type of information.

Suppose we had a given function f of x that also happened to be integrable on the interval from a to b. Or written more succinctly, we assume that the integral from a to b f of x dx exists. Let's also suppose that f of x happens to be a linear combination of the members of the orthogonal family phi sub n of x, in other words, that for the given orthogonal family, f of x is summation Cn phi n of x as n goes from 1 to infinity.

And the question is, under-- or with the knowledge that this is an orthogonal family, how can we determine each coefficient C sub n if this condition here is the whole? Or to make this more concrete and keep it as non-symbolic as possible, let's suppose I wanted to figure out here what C sub 3 happened to be. The trick is this. I recognize that, from this basic definition, if I multiply phi sub 3 of x by any other member of this family other than phi sub 3 itself and integrate from a to b, that term will be 0.

So what I do is I come back to this relationship. And I multiply both sides of this equation by phi sub 3 of x, keeping in mind if I wanted to find C sub k, I would have multiplied both sides by phi sub k of x. I'm just picking k equals 3 here. I multiply both sides by phi sub 3 of x. And because phi sub 3 of x is a constant subscript, meaning it doesn't depend on the summation n which only affects the subscript here, I can bring phi sub up 3 of x inside the summation sign.

Now the trick is I integrate both sides of this equality from a to b. In other words, I put an integral from a to b on both sides here, augment this with a dx. And I now know that this is the case.

My next gimmick, if you want to call it that-- and if you don't want to call it that, you should, because it is a gimmick, as we'll explain in the moment-- I simply interchange the order of summation and integration, remembering that the integral of the sum is the sum of the integrals. I interchange this order. And I now have the summation n goes from 1 to infinity integral from a to b Cn phi n of x phi 3 of x dx. And now remembering that the integral from a to b, phi sub n of x times phi sub 3 of x dx, by definition of orthogonality is 0 unless n equals 3, all of these terms drop out except when n equals 3.

Now what happens here? When n equals 3, this term here is C sub 3. This is phi sub 3 squared x dx integrated from a to b. And all the remaining terms are 0, because the integrand is 0. In other words, the right-hand side simply becomes-- well, since C3 is a constant, I can take it outside the integral sign. It's C3 integral from a to b C sub 3 squared of x dx. And now you see from this relationship here, I can solve for C sub 3. Namely, I simply divide both sides of this equation by this number here.

Now what I said by gimmick is simply this. When we said that the integral of a sum is the sum of the integrals, that was only for a finite sum. You may remember in part one of our course when we talked about uniform convergence, one of the things that we pointed out was that for infinite sums, you weren't always guaranteed that you could interchange the order of summation and integration. Or as I usually put it, you could always interchange them, but you may not get the same answer.

So the first question that comes up is, how do I know that I am allowed to interchange this order? By the way, the second question that comes up is-- and this is a more subtle question, but going back to what we're talking about before-- how do I even know that there is a linear combination of the phi sub n's that comes out to be the given f of x? You see you notice that our demonstration here presupposes that f of x can be written this way.

You see the two questions that come up is, given an f of x, first of all, can it be written this way? And secondly, even if it can, do we have the right to interchange the order of integration and summation? By the way, assuming for the moment that we can do this mechanically at least, notice that what C3 turns out to be is the integral from a to b f of x phi sub 3 of x dx divided by the integral form a to b phi sub 3 of x squared times dx, and again, replacing the 3 by a k. This is how every one of these coefficients could have been determined.

And now you see, as a very, very brief aside which I will go into much more detail in the study guide when we talk about the exercises and the various discussion of the exercises, given a series of functions which is integrable, a sequence of functions integrable on a given interval from a to b, one usually defines the dot product of these two functions to be the integral from a to b the product of the two functions times dx.

You see, remember last time we showed that all you needed to have a dot product was a rule that showed you how to combine two members of your set to give you a number? And certainly, this thing here is a number. And it also had to have three or four basic properties. And as one of the exercises, we will show that those properties for a dot product are obeyed here.

But this is where the word "orthogonal functions" come from. Notice that in terms of this definition of a dot product, two vectors are orthogonal if their dot product is 0. And now they're saying that this equals 0 is the same as saying that this integral is 0. And that's precisely what the definition of orthogonal was. So you see, this is the connection between the vector space interpretation of orthogonal and the traditional interpretation of orthogonal. And as I say, I'll leave further discussion of that for the notes.

But the important point is this. If I mechanically solve for these coefficients-- in other words, suppose I'm given f of x. And I'm given a family of functions phi sub n of x which is orthogonal on the interval from a to b. Suppose that without even knowing whether f of x can be represented as a linear combination of the phis, and that without even knowing whether I can interchange these two operations, the order in which I perform these two operations, let's suppose that I mechanically invent as my coefficients the C's in this particular way.

In other words, regardless of whether it means anything or not, I can always multiply f of x by phi sub k of x, integrate from a to b. You see, the product of two integrable functions is always integrable. I can always divide that by the square of the function phi sub k of x from a to b, keeping in mind, by the way, that as long as phi sub 3 is not identically 0 or if it's not 0 other than on a set of points of measure 0, what have you, notice that the square of a number can't be negative.

So if phi sub 3 of x is different from 0 in enough places, by squaring it, I have a positive integrand. And when I integrate a positive integrand I can't get 0 for an answer. So I don't have a zero denominator here.

But the point is this. If I choose the C's as I did here, see, with the C's as just defined, we write that f is x-- and then we put a little squiggle here. See, don't know if we had the right to assume that this series here is equal to f of x, but we write this little squiggle mark here. We call the resulting infinite series capital F of x. And what we say is that capital F of x is the Fourier representation of little f of x with respect to the family phi sub n of x. You see, we don't say that these two are equal.

But when we choose these coefficients the way I just mentioned, this is, in any event, called the Fourier representation of the function little f of x. I see there are many questions that come up. And in the form of an overview, let me just state them.

First of all, how do we know that capital F of x is even a convergent series? After all, this is an infinite series. How do we know it even converges? Secondly, if it does converge, how do we know that it converges to f of x? The answer is we don't know. We really don't know.

But there is a rather remarkable result. And the result is not that difficult to prove. But in the form of an overview, I prefer to omit the proof here. It will be an exercise in the study guide, but I think it's important to see what the statement of the result is. And then we can talk about the ramifications of that. But the proof would actually obscure the beauty of the result.

And the aside, which I leave as an exercise, is simply this. Suppose you took a linear combination of the first n of the terms phi 1 of x up to phi n of x. If you choose the C sub k's to be the so-called Fourier coefficients the way I just mentioned, what is interesting is this, that even if the Fourier representation doesn't converge to f of x, it has the property that it gives you the best least square approximation.

In other words, if I were to compute the difference between f of x and the sum of the first n terms in the Fourier representation where the C's were chosen according to the method I just mentioned to you before and square that difference-- you see, this is why it's called the mean square idea. You see, large errors and small errors, in other words, a large negative error can cancel a large positive error because the algebraic signs tend to cancel one another. But by squaring large errors, all of these become what? Squares of large magnitudes are large positive magnitudes. By squaring, I really magnify the errors. The positives can't cancel the negatives.

And what this says is that if I use this linear combination of the phi sub n's, that the square approximation error is less using these coefficients than if I had used any other possible coefficients, that if I picked any other coefficients in the whole world other than the ones that I found by the method indicated just a little while ago, if I summed these terms, subtracted that from f of x squared and integrated the results from a to b, I would get a number which was at least as large. And in most cases in fact, unless the d sub k's were chosen to equal the C sub k's, I would get a number which was larger than this.

Now you see what this thing says? What this says is that there can't be any real interval where this gets very, very far away from f of x, because you see, if it did, if it did get very, very far away, this would be a large magnitude. Squaring it would give me a positive large magnitude. Integrating that from a to b would make this thing still very, very large. And why should it stay smaller than any other linear combination that I can get?

And I'll emphasize this more from a concrete point of view in a few moments. But for the time being, what the impact of this beautiful result is-- and notice, it doesn't look like much when you look at it like this. It just looks like some abstract remark. But the beauty of this remark the impact of it is that, when you write the so-called Fourier representation, as opposed to power series, which I'll also mention at the end of the lecture what the contrast really is, the Fourier representation, if it does converge at all, converges to the function, what we call in the large, that over the whole interval it starts to fit the function very nicely. Now I'll come back to that in terms of a more specific example in a few moments.

Let me point out that, Fourier, being basically an engineer and a scientist, worked with one particular set of orthogonal functions, a set of orthogonal functions that were very, very natural to anybody working with wave motion or things of this type, as you'll see in a moment. The special case that I have in mind is this. Consider the family of functions-- well, I can call this cosine 0x, which is another way of writing 1. But I have 1 cosine x, cosine 2x, cosine 3x, cosine 4x, et cetera, and also included in this family, sine x, sine 2x, sine 3x, et cetera.

I claim that this family of functions is orthogonal on the interval from negative pi to pi. What I mean by that is what? That if I take any two different members of this family, multiply them together, integrate from minus pi to pi, the result will always be 0.

Now the-- I won't go through this whole thing. That will also be one of the exercises in the study guide will be to verify this. But let me give you a head start in doing this. Let me pick two members of the cosine, two different numbers of the cosine family here, and multiply them together, and integrate from minus pi to pi. In other words, let me take integral from minus pi to pi cosine mx cosine nx dx where m is unequal to n.

Now again, without belaboring this point, remember that the formula for cosine a cosine b comes from looking at the formulas for cosine a plus b cosine of a minus b and adding these two results. And the sine a sine b terms drop out. To make a long story short, or somewhat shorter, cosine mx times cosine nx is another way of writing one half the sum of cosine m plus nx plus cosine m minus n x.

Now you see, to integrate this, very simply, this just gives me what? A 1/2 here. This is sine m plus n x over m plus n. This is sine m minus x-- m minus n times x over m minus n. And those, by the way, the fact that m is unequal to n, guarantees me that my denominator here can't be zero.

And at any rate, noticing that when I evaluate this from x equals minus pi to pi, that these are integral, meaning whole-number multiples of pi, both these terms are zero. And we've established the fact that this is zero. So at least the members of the cosine family are orthogonal. One can also show the members of the sine family are orthogonal. And one can also show that any member of the cosine family taken-- multiplied by any member of the sine family integrated from minus pi to pi also gives zero.

So the point then is, fair enough, that this family here is an orthogonal family on the interval from minus pi to pi. And the beauty of this result-- and let me point out that in textbooks on advanced mathematics, the multitude of pages I use to prove the results, but the statements of the terms that you prove are not that difficult to keep track of. Let me just give one major result, a result that is very, very crucial, and one that I will emphasize in the study guide. And that is this.

Suppose that I am given a particular function little f of x, which is piecewise smooth on the interval from minus pi to pi. And remember what I mean by piecewise smooth. I mean it's a differentiable function except that it may have a finite number of jump discontinuities, that if I plot the graph y equals f of x, I get a smooth curve with the possible exception that there may be a few jumps in it, by few meaning what? A not-too-great infinite number, a set of measures zero type thing. But we won't belabor that point right now.

But the idea is this. Suppose that I have this piecewise smooth function little f of x. And suppose capital F of x is the Fourier representation of little f of x relative to my sines and cosines. In other words, capital F of x is the linear combinations of the cosine terms and the sine terms where the coefficients are chosen by that little recipe that I showed you at the beginning of the lecture, and which, as I say, I'll reinforce as we do the exercises.

At any rate, let's call this capital F of x. And in this particular case, in other words, what particular case? If little f of x is piecewise smooth and the orthogonal functions are the sines and cosines on the interval from minus pi the pi, then the Fourier representation of f of x, capital F of x, is given by the following. And this is a remarkable result. It says that capital F of x at any value of x in the interval from minus pi to pi is little f of x plus plus little f of x minus over 2.

Now, what does f of x plus mean? That means the limit of f of x as you approach the given value of x from the positive side. And this f is x minus is the limit of f of x as you approach the given value of x from the negative side. In fact, to keep this easier to understand so I don't use x in two different ways, what I'm saying is capital F of x sub 0 is computed as follows.

See what little f of x sub 0 is as x approaches x0 from the left. See what f of x0 of f of x approaches as x approaches x0 from the right. Add these two results and divide by 2. I hope that you can see that what I'm doing here is taking the average of these two results. Half the sum is the average.

Also observe that if f happens to be continuous at x0, f of x0 plus is the same as f of x0 minus. You see, the only time that these two things can be different is if there is a jump in the function at x0, because if there is no jump, as I approach x0 from the left and approach x0 from the right, the two values I get must be the same. In other words, notice that at the points where f is continuous, this is nothing more than f of x0 itself.

And what this famous result says is that, except at the points at which little f of x is discontinuous, capital F, the Fourier representation, is little f itself, that the Fourier representation converges to the function itself. And what it does that's remarkable is that where there is a jump in the function, it equals the average of the jump. Well, let me show you that in terms of an example.

Let me pick a very simple function to begin with. Let me define the function little f of x to be negative 1 when x is in the interval from minus pi to 0. And let it be, say, positive 1 when x is in the interval from 0 to pi. So graphically, you see how this thing is going to look. It goes like this.

By the way, notice that this is what mathematically we call an odd function. F of x is the negative of f of minus x. Notice that the sine terms are odd. The cosine terms are even. You see, cosine of x is the same as cosine of minus x. But sine of x is minus sine minus x.

So the sine terms are odd. The cosine terms are even. So the first hint or the first nice simplification that takes place by choosing this example is that because this is an odd function, the cosine terms will all be absent.

And it turns out that if I use the procedure indicated earlier in the lecture-- and again that will be one of the exercises so that you get the drill that's necessary here, again, I only want to give you the overview-- it turns out that the Fourier series representation of this particular function, capital F of x, turns out to be 4 over pi times the summation sine n x over n for n odd. And what that means is it's 4 over pi times sine x plus sine 3x over 3 plus sine 5x over 5, et cetera.

By the way, these are very difficult things to draw. You're not used to working with these. I should point out this, that if there is anything that works nicely in the laboratory, it's sine and cosine terms, because those are things that you can produce very nicely on an oscilloscope.

Analytically, these are very nasty terms to handle. But electronically, it's a very nice way of getting a feeling for Fourier series. But I don't have electronic chalk here, so I just talk about it and draw these things for you.

But for example, to show you what I mean here, let me just take the first two terms. Let me plot the curve y equals 4 over pi times the quantity sine x plus sine 3x over 3. And if I plot that curve, which I've done here in the accentuated chalk, look at how nicely that curve starts already to fit the graph y equals f of x.

You see how it fits over here? No perfect fit any place, but notice that, by and large, it fits the curve very nicely without too big an error all the way along here. It has roughly the right shape. And notice also it goes through the origin, because when x is 0, capital F of x is 0.

And that's exactly the average of the jump over here. So you jump from minus 1 to 1. So the origin is the average over here. By the way, even though I don't draw very well, what I have tried to do here is show what happens if I tack on just a couple of more terms.

I've tacked onto this the term sine 5x over 5 and sine 7x over 7. And then the graph would look something like this. You see how nicely this starts to fit? It fits very nicely along here. But wherever there's a jump, it jumps to-- it averages out this way.

By the way, one other observation-- obviously, the function is I'm constructing here is periodic with period 2 pi. If I replace x by x plus 2 pi, I don't change the sine any. And so obviously, the function which I'm calling capital F of x represents more than this given function. Rather, it represents the given function reproduced infinitely in both directions with period 2 pi.

But again, I'll leave that for the exercises. To summarize what's happened over here, you see how nicely this thing fits? If I actually took capital F of x and computed this limit for each value of x, what the graph would look like-- now obviously, I can't get this on the oscilloscope, because no matter how many terms I use, no matter how large a number, n is still finite.

But if I actually took the infinite sum, the graph that I would get mathematically is I would actually get the straight line x equals-- I'm sorry, y equals negative from minus pi to 0. It would be y equals positive 1 from 0 to pi. And at 0 itself, the function would be 0. You see, it splits the gap. It splits the jump. And then of course, the function would repeat itself at a regular interval of-- regular period of 2 pi, that this is what capital F of x would look like if you graphed it.

By the way, to give you an example of an even function, if we were to use this same technique for getting the Fourier series for f of x equals the absolute value of x on the interval from minus pi to pi, it turns out that capital F of x would be pi over 2 minus 4 over pi summation over n odd cosine nx over n squared. By the way, if this looks a little bit like the previous problem, notice that this is more than just coincidence.

If you differentiate the square root of x-- I'm sorry, if you differentiate the absolute value of x, what is the derivative? It's minus 1 for x negative and positive 1 for x positive. It's the function that we were just talking about before, only it doesn't exist at x equals 0, you see?

What I'm driving at over here is that it's not surprising that the terms in here are the integrals of the terms that were in our previous summation. But notice that the absolute value of x is an even function. Consequently, it's the cosine terms which appear here rather than the sine terms.

And again, I've run the risk of drawing my own diagrams here. I have plotted this for n equals 3, meaning I've plotted y equals pi over 2 minus 4 over pi times the quantity cosine x plus cosine 3x over 9. And that's represented by this curve drawn in the accentuated chalk over here. And again, notice how even for just a small number of terms, we have already captured in the large the shape of the curve y equals the absolute value of x, which by the way contrasts very sharply with power series.

Remember how power series worked? First of all, you had to require that the function that you were talking about possessed derivatives of all orders. Secondly, when you wanted to fit the function by its power series, you picked a particular point. Remember what you got? You first got a tangent line approximation. Then you got a quadratic approximation that fit the curve in a neighborhood at this point better than the straight line did. But after a while, the thing would sag off.

Then you added a cubic term. And that fit even better for a while. And then again something weird might happen. And you kept on going like this so that ultimately, even though you had uniform convergence, what you were doing was you were getting a splendid fit near one particular point and slowly but surely, and actually very, very slowly, and also actually, very surely, but you may not have noticed it, these approximations were starting to fit better and better as you went out here. But the least square approximations would not be too good. In other words, out here someplace, there was starting to become some drastically large errors.

On the other hand, the Fourier representation does what? It fits the curve that you're talking about very, very nicely very, very quickly. And this is one reason why Fourier series are as powerful as they are.

And by the way, again, to end on a nice theme, Fourier series probably have as much space devoted to them in pure mathematics as they do in applied mathematics. It's a beautiful topic from both points of view. But I felt that for our lesson today, since it is a finishing touch to essentially the equivalent of three-plus semesters of calculus, that we should juxtaposition the old and the new and show the peaceful coexistence versus this idea of polarization where you either have to be a traditionalist or a modernist, where you either have to be an applied man or a purist, that all of these ideas mingle side by side. And about the only choice you have as a human being depending on what your interests are is what proportion you're going to mix them in.

Well, at any rate, this brings us to the part of the course that always gives me a little bit of remorse. And that is, I hate to let my captive audience go. We have now been through, as I say, the equivalent of a couple of years of mathematics together. It has been a great pleasure for me. I hope it has been for you.

And once again, I would like to thank everybody who has worked with me on this, but in particular, the usual three people who have been with me through thick and thin on this, John [? Fitchar, ?] project self-study manager, who has also done a lot of work with me in helping me write the study guide, who has given me a lot of advice on how to give these lectures and how to prepare them for the taping, and has also been our director for the taping, to Miss Elise [? Pelletier, ?] who in addition to her usual secretarial work, has typed much of our study guide and has worked as a video technician for us, and to our chief-- I'll just call him chief of audio visual effects at the center here, Charles [? Patton. ?] I am particularly indebted to these three people, indebted to a lot of other people. I thank you all for being with me this long. And I hope that our paths will cross again.

Funding for the publication of this video was provided by the Gabriella and Paul Rosenbaum Foundation. Help OCW continue to provide free and open access to MIT courses by making a donation at ocw.mit.edu/donate.