

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue offer high quality educational resources for free. To make a donation or to view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at ocw.mit.edu.

PROFESSOR: I just started talking a little bit last time about viewing L^2 , namely this set of functions that are square integrable as a vector space. And I want to reveal a little bit about why we want to do that. Because after spending so long worrying about all these questions of measurability and all that, you must wonder, why do we want to look at it a different way now.

Well, a part of it is we want to be able to look at orthogonal expansions geometrically. In other words, we would like to be able to draw pictures of them. You can draw pictures of a function, but you're drawing a picture of one function. And you can't draw anything about the relationship between different functions, except by drawing the two different functions. You get all the detail there, and you can't abstract it at all. So somehow we want to start to abstract some of this information. And we want to be able to draw pictures which look more like vector pictures than like functions. And you'll see why that becomes important in a while.

The other thing is, when you draw a function as a function of time, the only thing you see is how it behaves in time. When you take the Fourier transform and you draw it in frequency, the only thing you see is how it behaves in frequency. And, again, you don't see what the relationship is between different functions. So you lose all of that. When you take this signal space viewpoint, what you're trying to do there is to not stress time or frequency so much, but more to look at the relationship between different functions.

Why do we want to do that? Well, because as soon as we start looking at noise and things like that, we want to be able to tell something about how distinguishable different functions are from each other. So the critical question there you want to ask, when you ask how different functions are, is how do you look at two functions

both at the same time. So, again, that's -- all of this is coming back to the same thing. So we want to be able to draw pictures of functions which show how functions are related, rather than show all the individual detail of function.

Finally, what we'll see at the end of today that it gives us a lot of capability for understanding much, much better what's going on when we look at conversions of orthogonal series. This is something we haven't had any way to look at before we talked about the Fourier series, the discrete time Fourier transform, and things like this. But we haven't been able to really say something general. Which, again, is what we want to do now.

I'm going to go very quickly through these axioms of a vector space. Most of you have seen them before. Those of you who haven't seen axiomatic treatments of various things in mathematics are going to be puzzled by it anyway. And you're going to have to sit at home or somewhere on your own and puzzle this out. But I just wanted to put them up so I could start to say what it is that we're trying to do with these axioms. What we're trying to do is to say everything we know about n -tuples, which we've been using all of our lives. All of these tricks that we use, most of them we can use to deal with functions also. The pictures, we can use. All of the ideas about dimension, all of the ideas about expansions. All of this stuff becomes useful again. But, there are a lot of things that are true about functions that aren't true about vectors. There are lots of things which are true about n -dimensional vectors that aren't true about functions. And you want to be able to go back to these axioms when you have to, and say, well, is this property we're looking at something which is a consequence of the axioms. Namely, is this an inherent property of vector spaces, or is it in fact something else which is just because of the particular kind of thing we're looking at.

So, vector spaces. Important thing is, there's an addition operation. You can add any two vectors. You can't multiply them, by the way. You can only add them. You can multiply by scalars, which we'll talk about in the next slide. But you can only add the vectors themselves. And the addition is commutative, just like ordinary addition of real numbers or complex numbers is. It's associative, which says v, u plus w in

parentheses.

Now, you see why we're doing this is, we've said by definition that for any two vectors, u and w , there's another vector, which is called u plus w . In other words, this axiomatic system says, whenever you add two vectors you have to get a vector. There's no way around that. So u plus w has to be a vector. And that says that v plus u plus w has to be a vector. Associativity says that you get the same vector if you look at it the other way, if first adding v and u and then adding w . So, anything what you call a vector space has to have this property. Of course, the real numbers have this property. The complex numbers have this property. All the n -tuples you deal with all the time have this property. Functions have this property. Sequences, infinite length sequences have this property, so there's no big deal about it. But that is one of the axioms.

There's a unique vector 0 , such that when you add 0 to a vector, you get the same vector again. Now, this is not the 0 in the real number system. It's not the 0 in a complex number system. In terms of vectors, if you're thinking of n -tuples, this is the n -tuple which is all 0 's. In terms of other vectors, spaces, it might be other things. Like, in terms of functions, 0 means the function which is 0 everywhere.

And finally, there's the unique vector minus v , which has the property that v plus minus v -- in other words, v minus v -- is equal to 0 . And in a sense that defines subtraction as well as addition. So we have addition and subtraction, but we don't have multiplication.

And there's scalar multiplication. Which means you can multiply a vector by a scalar. And the vector spaces we're looking at here only have two kinds of scalars. One is real scalars and the other is complex scalars. The two are quite different, as you know. And so when we're talking about a vector space we have to say what the scalar field is that we're talking about.

So, for every vector. And also for every scalar, there's another vector which is the scalar times the vector. Which means, you have scalar multiplication. You can multiply vectors by scalars in terms of n -tuples. When you're multiplying a scalar by

an n -tuple, you just multiply every component by that scalar.

When you take the scalar 1, and in a field there's always an element 1, there's always an element 0. When you take the element 1, and you multiply it by a vector, you got the same vector. Which, of course, is what you would expect if you were looking at functions. You multiply a function. For every t you multiply it by 1, and you get the same thing back again. For an n -tuple, you multiply every component by 1, you get the same thing back again. So that's -- well, it's just another axiom.

Then we have these distributive laws. And I won't read them out to you, they're just what they say. I want to talk about them in a second a little bit. And as we said, the simplistic example of a vector space, and the one you've been using for years, partly because it saves you a lot of notation. Incidentally, one of the reasons I didn't list for talking about L^2 as a vector space is, it saves you a lot of writing. Also, just like when you're talking about n -tuples it saves you writing. It's a nice, convenient notational device. I don't think any part of mathematics becomes popular with everyone, unless it saves notation as well as well as simplifying things.

So we have these n -tuples, which is what we mean by \mathbb{R}^n or \mathbb{C}^n . In other words, when I talk about \mathbb{R}^n , I don't mean just a vector space of dimension n . I haven't even defined dimension yet. And when you talk in this generality, you have to think a little bit about what it means. So we're just talking about n -tuples now. \mathbb{R}^n is n -tuples of real numbers. \mathbb{C}^n is n -tuples of complex numbers.

Addition and scalar multiplication are component by component, which we just said. In other words, $w = u + v$ means this, for all i . Scalar multiplication means this. The unit vector, e_i , is a 1 in the i 'th position, a 0 everywhere else. And what that means is that any vector v which is an n -tuple, \mathbb{R}^n or \mathbb{C}^n , can also be expressed as the equals summation of these coefficients, times these unit factors. That looks like we're giving up the simple notation we have and going to more complex notation.

This is a very useful idea here. That you can take even something as simple as n -tuples and express every vector as a sum of these simple vectors. And when you

start drawing pictures, you start to see why this makes sense. I'm going to do this on the next slide. So let's do it.

And on the slide here, what I've done is to draw a diagram which is one you've seen many times, I'm sure. Except for the particular things on it. Of two-dimensional space, where you take a 2-tuple and you think of the first element in the 2-tuple, v_1 , as being the horizontal axis. The second element, v_2 , as being the vertical axis. And then you can draw any 2-tuple by going over v_1 and then up v_2 , which is what you've done all your lives.

The reason I'm drawing this is to show you that you can take any two vectors. First vector, v . Second vector, u . One thing that -- I'm sure you all traditionally do is you view a vector in two ways. One, you view it as a point in two-dimensional space. And the other is you view it as a line from 0 up to v . And when you put a little arrow on it, it looks like a vector, and we call it a vector. So it must be a vector, OK? So either the point or the line represents a vector. Here's another vector, u . I can take the difference between two vectors, u minus v , just by drawing a line from v up to u . And thinking of that as a vector also. So this vector really means a vector starting here, going parallel to this, up to this point, which is what w is. But it's very convenient to draw it this way, which lets you know what's going on. Namely, you represent u as the sum of v plus w . Or you represent w as a difference between u and v .

And all of this is trivial. I just want to say it explicitly so you start to see what the connection is between these axioms, which if I don't talk about them a little bit, I'm sure you're just going to blow them off and decide, eh, I know all of that, I can look at things in two dimensions and I don't have to think about them at all. And then, in a few days, when we're doing the same thing for functions, you're suddenly going to become very confused. So you ought to think about it in these simple terms.

Now, when I take a scalar multiple of the vector, in terms of these diagrams, what we're doing is taking something on this same line here. I can take scalar multiples which go all the way up here, all the way down there. I can take scalar multiples of u which go up here. And down here. The interesting thing here is when I scale v down

by alpha and I scale u down by alpha, I can also scale w down by alpha and I connect it just like that, from alpha v to alpha u.

Which axiom is that? I mean, this is the canonic example of one of those axioms.

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: What?

AUDIENCE: Distributed.

PROFESSOR: Distributed, good. This is saying that alpha times the quantity u minus v is equal to alpha u minus alpha v. That's so trivial that it's hard to see it, but that's what it's saying. So that the distributive law really says the triangles maintain their shape. Maybe it's easier to just say distributive law, I don't know.

So, for the space of L2 complex functions, we're going to define u plus v as w, where w of t equals u of t plus v of t. Now, I've done a bunch of things here all at once. One of them is to say what we used to call a function, u of t, we're now referring to with a single letter, boldface u. What's the advantage of that?

Well, one advantage of it is when you talk about a function as u of t, you're really talking about two different things. One, you're talking about the value of the function at a particular argument, t. And the other is, you're talking about the whole function. I mean, sometimes you want to say a function has some property. A function is L2. Well, u of t at a particular t is just a number. It's not a function. So this gives us a nice way of distinguishing between functions and between the value of functions for particular arguments. So that's one more notational advantage you get by talking about vectors here.

We're going to define the sum of two functions just as the point y sum. In other words, these two functions are defined at each t. w defined at a t is the sum of this and that. The scalar multiplication is just defined by, at every t, u of t is equal to alpha times v of t. Just what you'd expect. There's nothing strange here. I just want to be explicit in saying how everything follows from these axioms here. And I won't

say all of it because there's too much of it.

With this addition and scalar multiplication, L^2 , the set of finite energy measurable functions is in fact the complex vector space. And it's trivial to go back and check through all the axioms with what we said here, and I'm not going to do it now.

There's only one of those axioms which is a bit questionable. And that is, when you add up two functions which are square integrable, is the sum going to be square integrable also.

Well, you nod and say yes, but it's worthwhile proving it once in your lives. So the question is, is this function here less than infinity if the integral of u of t squared and the integral of v of t squared are both integrable. Well, it's a useful inequality, which looks like a very weak inequality but it, in fact, is not weak. It says that u of t plus v of t . This is just at a particular value of t . So this is just an inequality for real numbers and complex numbers. It says that this u of t plus v of t , quantity squared, magnitude squared, is less than or equal to 2 times u of t squared plus 2 times v of t squared. You wonder, at first, what's the 2 doing in there. But then think of making v of t equal to u of t . You have 2 times u of t squared, which is 4 times u of t squared. Well, that's what you need here to make this true. So, in that example this inequality is satisfied with equality.

To verify the inequality in general, you just multiply this out. It's u of t squared plus v of t squared plus two cross-terms and it all works.

This vector space here is not quite the vector space we want to talk about. But let's put off that question for a while and we'll come to it later. We will come up with a vector space which is just slightly different than that.

The main thing you can do with vector spaces is talk about their dimension. Well, there are a lot of other things you can do, but this is one of the main things we can do. And it's an important thing which we have to talk about before going into inner product spaces, which is what we're really interested in.

So we need a bunch of definitions here. All of this, I'm sure is familiar to most of

you. For those of you that it's not familiar to, you just have to spend a little longer with it. There's not a whole lot involved here. If you have a set of vectors, which are in a vector space, you say that they span the vector space if in fact every vector in this vector space is a linear combination of those vectors. In other words, any vector, u , can be made up as some sum of α_i times v_i .

Now, notice we've gone a long way here beyond those axioms. Because we're talking about scalar multiplications. And then we're talking about a sum of a lot of scalar multiplications. Each scalar multiplication is a vector. The sum of a bunch of vectors, by the associative law is, in fact, another vector. So every one of these sums is a vector. And by definition, a set of vectors spans a vector space if every vector can be represented as some linear combination of them. In

other words, there isn't something outside of here sitting there waiting to be discovered later. You really understand everything that's there. And we say that v is finite dimensional if it is spanned by a finite set of vectors. So that's another definition. That's what you mean by finite dimensional. You have to be able to find a finite set of vectors such that linear combinations of those vectors gives you everything. It doesn't mean you have a finite set of vectors. You have an infinite set of vectors because you have an infinite set of scalars. In fact, you'd have an uncountably infinite set of vectors because of these scalars.

Second definition. The vector v_1 to v_n are linearly independent -- and this is a mouthful -- if u equals the sum of $\alpha_i v_i$ equals 0, only for α_i equal to 0 for each i . In other words, you can't take any linear combination of the v_i 's and get 0 unless that linear combination is using all 0's. You can't, in any non-trivial way, add up a bunch of vectors and get 0. To put it another way, none of these basis vectors is a linear combination of the others. That's usually a more convenient way to put it. Although it takes more writing.

Now, we say that a set of vectors are a basis for v if they're both linearly independent and if they span v . When we first talked about spanning, we didn't say anything about these vectors being linearly independent, so we might have had

many more of them than we needed. Now, when we're talking about a basis, we restrict ourselves to just the set we need to span the space.

And then the theorem, which I'm not going to prove, but it's standard Theorem One of any linear algebra book -- well, it's probably Theorem One, Two, Three and Four all put together -- but anyway, if it says if v_1 and $v_{\text{sub } n}$ span v , then a subset of these vectors is the basis of b . In other words, if you have a set of vectors which span a space, you can find the basis by systematically eliminating vectors which are linear combinations of the others, until you get to a point where you can't do that any further. So this theorem has an algorithm tied into it. Given any set of vectors which span a space, you can find the basis from it by perhaps throwing out some of those vectors.

The next part of it is, if v is a finite dimensional space, then every basis has the same size. This, in fact, is a thing which takes a little bit of work proving it. And, also, any linearly independent set, v_1 to $v_{\text{sub } n}$, is part of the basis. In other words, here's another algorithm you can use. You have this big, finite dimensional space. You have a few vectors which are linearly independent. You can build a basis starting with these, and you just experiment. You experiment to find new vectors, which are not linear combinations of that set. As soon as you find one, you add it to the basis. You keep on going. And the theorem says, by time you get to n of them, you're done. So that, in a sense, spanning sets are too big. Linearly independent sets are too small. And what you want to do is add the linearly independent sets, shrink the spanning sets, and come up with a bases. And all bases have the same number of vectors. There many different bases you come up with, but they all have the same number of vectors.

Not going to talk at all about infinite dimensional spaces until the last slide today. Because the only way I know to understand infinite dimensional vector spaces is by thinking of them, in some sort of limiting way, as finite dimensional spaces. And I think that's the only way you can do it.

A vector space in itself has no sense of distance or angles in it. When I drew that

picture before for you -- let me put it up again -- it almost looked like there was some sense of distance in here. Because when we took scalar multiples, we scaled these things down. When I took these triangles, we scaled down the triangle, and the triangle was maintained. But, in fact, I could do this just as well if I took v and moved it down here almost down on this axis, and if I took u and moved that almost down on the axis. I have the same picture, the same kind of distributive law. And everything else.

You can't -- I mean, one of the troubles with n -dimensional space, to study what a vector space is about, is it's very hard to think of n -tuples without thinking of distance. And without thinking of things being orthogonal to each other, and of all of these things. None of that is talked about at all in any of these axioms. And, in fact, you need some new axioms to be able to talk about ideas of distance, or angle, or any of these things.

And that's what we want to add here. So we need some new axioms. And what we need is a new operation on the vector space, before the only -- the only operations we have are addition and scalar multiplication. So that vector spaces are really incredibly simple animals. There's very little you can do with them. And this added thing is called an inner product. An inner product is a scalar valued function of two vectors. And it's represented by these little brackets. And the axioms that these inner products have to satisfy is, if you're dealing with a complex vector space. In other words, where the scalars are complex numbers, then this inner product, when you switch it around, you have Hermitian symmetry. Which means that this inner product is equal to $u \cdot v$ complex conjugate.

We've already seen that sort of thing in taking Fourier series. And the fact that when you're dealing with complex numbers, symmetry doesn't usually hold, and you usually need some kind of Hermitian symmetry in most of the things you do.

The next axiom is something called bilinearity. It says that if you take a vector which is α times a vector v , plus β times a vector u , take the inner product of that with w , it splits up as α times $v \cdot w$ plus β times $u \cdot w$.

How about if I do it the other way? See if you understand what I'm talking about at all here. Suppose we take $w\alpha u + \beta v$. What's that equal to? Well, it's equal to $\alpha wu + \beta wv$. Except that's wrong, it's right for real vector spaces, it's wrong for complex vector spaces. What am I missing here?

I need complex conjugates here and complex conjugates here. I wanted to talk about that, because when you're dealing with inner products, I don't know whether you're like me, but every time I start dealing with inner products and I get in a hurry writing things down, I forgot to put those damn complex conjugates in them. And, just be careful. Because you need them. At least go back after you're all done and put the complex conjugates in. If you're dealing with real vectors, of course you don't need to worry about any of that.

And all the pictures you draw are always of real vectors. Think of trying to draw this picture. This is the simplest picture you can draw. Of two-dimensional vectors. This is really a picture of a vector space of dimension two, for real vectors. What happens if you try to draw a picture of complex two-dimensional vector space? Well, it becomes very difficult to do. Because you're really talking about the real part of, if you're dealing with a basis which consists of two complex vectors, then instead of v_1 , you need a real part of v_1 and imaginary part of v_1 . Instead of v_2 , you need real part of v_2 and imaginary part of v_2 . And you can always draw this in four dimensions. And I even have trouble drawing in three dimensions, because somehow my pen doesn't make marks in three dimensions. And in four dimensions, I'm a blinking 12. And have no idea of what to do. So you have to remember this.

If you're dealing with \mathbb{R}^n or \mathbb{C}^n , almost always you define the inner product of v and u as the sum of the components with the second component complex conjugated. This is just a standard thing that we do all the time. When we do this, and we use unit vectors, the inner product of v with the i 'th unit vector is just v_i . That's what this formula says. Because e_i is this vector u , in which there's a 1 only in the i 'th position, and a 0 everywhere else. So $v \cdot e_i$ is always the v_i , and $e_i \cdot v$ is always v_i complex conjugate. Again, this Hermitian nonsense that comes up to plague us all the time.

And from that, if you make v equal to e_j or e_i , you get the inner product of two of these basis vectors is equal to 0 for $i \neq j$. In other words, the standard way of drawing pictures when you make it into an inner product space, those unit vectors become orthonormal. Because that's the way you like to draw things. You like to draw one here and one there. And that's what we mean by perpendicular, which the two-dimensional or three-dimensional word for orthogonal.

So we have a couple of definitions. The inner product of v with itself is called inner product v squared, which is called the squared norm of the vector. The squared norm has to be non-negative, by axiom. It has to be greater than 0, unless this vector, v , is in fact a 0 vector. And the length is just the square root of the norm squared. In other words, the length and the norm are the same thing. The norm of a vector is the length of the vector. I've always called it the length, but a lot of people like to call it the norm. v and u are orthogonal if the inner product of v and u is equal to 0. How did I get that? I defined it that way. Everybody defines it that way. That's what you mean by orthogonality. Now we have to go back and see if it makes any sense in terms of these nice simple diagrams.

But first I'm going to do something called the one-dimensional projection theorem. Which is something you all know but you probably have never thought about. And what it says is, if you have two vectors, v and u , you can always break v up into two parts. One of which is on the same line with u . In other words, is colinear with u . I'm drawing a picture here for real spaces, but when I say colinear, when I'm dealing with complex spaces, I mean it's u times some scalar, which could be complex. So it's somewhere on this line. And the other part is perpendicular to this line. And this theorem says in any old inner product space at all, no matter how many dimensions you have, infinite dimensional, finite dimensional, anything, if it's an inner product space on either the scalars \mathbb{R} or the scalars \mathbb{C} , you can always take any old two vectors at all. And you can break one vector up into a part that's colinear with the other, and another part which is orthogonal. And you can always draw a picture of it. If you don't mind just drawing the picture for real vectors instead of complex vectors.

This is an important idea. Because what we're going to use it for in a while, is to be

able to talk about functions which are these incredibly complicated objects. And we're going to talk about two different functions. And we want to be able to draw a picture in which those two functions are represented just as points in a two-dimensional picture. And we're going to do that by saying, OK, I take one of those functions. And I can represent it as partly being colinear with this other function. And partly being orthogonal to the other function. Which says, you can forget about all of these functions which extend to infinity, in time extend to infinity, in frequency and everything else. And, so long as you're only interested in some small set of functions, you can just deal with them as a finite dimensional vector space. You can get rid of all the mess, and just think of them in this very simple sense. That's really why this vector space idea, which is called signal space, is so popular among engineers. It lets them get rid of all the mess and think in terms of very, very simple things.

So let's see why this complicated theorem is in fact true. Let me state the theorem first. It says for any inner products space, V . And any vectors, u and v in V , with u unequal to 0 -- I hope I said that in the notes -- the vector v can be broken up into a colinear term plus an orthogonal term, where the colinear term is equal to a scalar times u . That's what we mean by colinear. That's just the definition of colinear. And the other vector is orthogonal to u . And α is uniquely given by the inner product $v \cdot u$ divided by the norm squared of u .

Now, there's one thing ugly about this. You see that norm squared, you say, what is that doing there. It just looks like it's making the formula complicated. Forgot about that for the time being. We will get into it in a minute and explain why we have that problem. But, for the moment, let's just prove this and see what it says. So what we're going to do is say, well, if I don't look at what the theorem is saying, what I'd like to do is look at some generic element which is a scalar multiple times u .

So I'll say, OK let v parallel to u be α times u . I don't know what α is yet, but α 's going to be whatever it has to be. We're going to choose α so that this other vector, v minus $v \cdot u$, is a thing we're calling v perp. So that that, the inner product of that and u , is equal to 0. So what I'm trying to do is to find a vector --

strategy here, is to take any old vector along this line and try to choose the scalar α in such a way that the difference between this point and this point is orthogonal to this line. That's why I started out with α unknown. α unknown just says we have any point along this line. Now I'm going to find out what α has to be. I hope I will find out that it has to be only one thing, and it's uniquely chosen. And that's what we're going to find.

So v minus this projection term, this is called a projection of v on u , is equal to $v - \alpha u$ minus a projection inner product with u . So it's equal to this difference here. This is equal to the inner product. Since we have chosen this term to be α times u , we can bring the α out. So it's α times the inner product of u with itself. That's the norm squared of u . So this is inner product of v and u minus α times the norm squared. This is 0 if and only if you set this equal to 0. And the only value α can have to make this 0 is the inner product of v and u divided by the norm squared.

So I would think that if I ask you to prove this without knowing the projection theorem, I would hope that if you weren't afraid of it or something, and you just sat down and tried to do it, you would all do it in about half an hour. It would probably take most of you longer than that, because everybody gets screwed up in the notation when they first try to do this. But, in fact, this is not a complicated thing. This is not rocket science.

Now. What is this norm squared doing here? The thing we have just proven is that with any two vectors v and u , the projection of v on u , namely that vector, there which has the property that $v - \text{proj}_u v$ is perpendicular to u , we showed, is this inner product divided by the norm squared times u .

Now, let me break up this norm squared. Which is just some positive number. It's a positive real number. Into the length times the length. Namely, the norm u is just some real number. And write it this way, as the inner product of v with u divided by the length of u times u divided by the length of u .

Now, what is the vector u divided by the length of u ?

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: What?

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: It's the same direction of u , but it has length 1. In other words, this is the normalized form of u . And I have the normalized form of u in here also. So what this is saying is that this projection is also equal to the projection of v on the normalized form of u , times the normalized form of u . Which says that it doesn't make any difference what the length of u is. This projection is a function only of the direction of u . I mean, this is obvious from the picture, isn't it?

But again, since we can't draw pictures for complex valued things, it's nice to be able to see it analytically. If I shorten u , or lengthen u , this projection is still going to be exactly the same thing. And that's what the norm squared of u is doing here. The norm squared of u is simply sitting there so it does this normalization function for us. It makes this projection equal to the inner product of v with the normalized form of u , times the normalized vector for u .

The Pythagorean theorem, which doesn't follow from this, it's something else, but it's simple -- in fact, we can do it right away -- it says if v and u are orthogonal, then the norm squared of u plus v is equal to u squared plus v squared. I mean, this is something you use geometrically all the time. And you're familiar with this. And the argument is, you just break this up into the norm squared of u plus the norm squared of v , plus the two cross-products, the inner product of u times the inner product of u with v , which is 0, and the inner product of v with u , which is 0. So the two cross-terms go away and you're left with just this.

And for any v and u , the Schwarz inequality says that the inner product, the magnitude of the inner product of v and u , is less than or equal to the length of v times the length of u . The Schwarz inequality is probably -- well, I'm not sure that it's probably, but it's perhaps the most used inequality in mathematics. Any time you

use vectors, you use this all the time. And it's extremely useful. I'm not going to prove it here because it's in the notes. It's a two-step proof from what we've done. And the trouble is watching two-step proofs in class. At a certain point you saturate on them. And I have more important things I want to do later today, so I don't want you to saturate on this. You can read this at your leisure and see why this is true.

I did want to say something about it, though. If you divide the left side by the right side, you can write this as the magnitude of the inner product of the normalized form of v with the normalized form of u . If we're talking about real vector space, this in fact is the cosine of the angle between v and u . It's less than or equal to 1. So for real two-dimensional vectors, the fact that the cosine is less than or equal to 1 is really equivalent to the Schwarz inequality. And this is the appropriate generalization for any old vectors at all. And the notes would say more about that if that went a little too quickly.

OK, the inner product space of interest to us is this thing we've called signal space. Namely, it's a space of functions which are measurable n square integrals. In other words, finite value when you take the square and integrate it. And we want to be able to talk about the set of either real or complex L^2 functions. So, either one of them, we're going to define the inner product in the same way for each. It really is the only natural way to define an inner product here. And you'll see this more later as we start doing other things with these inner products. But just like what when you're dealing with n -tuples, there's only one sensible way to define an inner product.

You can define inner products in other ways. But it's just a little bizarre to do so. And here it's a little bizarre also.

There's a big technical problem here. And if you look at, it can anybody spot what it -- no, no, of course you can't spot what it is unless you've read the notes and you know what it is. One of the axioms of an inner product space is that the only vector in the space which has an inner product with itself, a squared norm equal to zero is the zero vector. Now, we have all these crazy vector we've been talking about,

which are zero, except on a set of measures zero. In other words, vectors' functions which have zero energy but just pop up at various isolated points and have values there. Which we really can't get rid of if we view a function as something which is defined at every value of t . You have to accept those things as part of what you're dealing with.

As soon as you started integrating things, those things all disappear. But the trouble is, those functions, which are zero almost everywhere, are not zero. They're only zero almost everywhere. They're zero for all engineering purposes. But they're not zero, they're only zero almost everywhere. Well, if you define the inner product in this way and you want to satisfy the axioms of an inner product space, you're out of luck. There's no way you can do it, because this axiom just gets violated. So what do you do? Well, you do what we've been doing all along. We've been sort of squinting a little bit and saying, well, really, these functions of measure 0 are really 0 for all practical purposes. Mathematically, what we have to say is, we want to talk about an equivalence class of functions. And two functions are in the same equivalence class if their difference has 0 energy. Which is equivalent to saying their difference is zero almost everywhere. It's one of these bizarre functions which just jumps up at isolated points and doesn't do anything else.

Not impulses at isolated points, just non-zero at isolated points. Impulses are not really functions at all. So we're talking about things that are functions, but they're these bizarre functions which we talked about and we've said they're unimportant. So, but they are there.

So the solution is to associate vectors with equivalence classes. And d of t and u of t are equivalent if the v of t minus u of t is zero almost everywhere. In other words, when we talk about a vector, u , what we're talking about is an equivalence class of functions. It's the equivalence class of functions for which two functions are in the same equivalence class if they differ only on a set of measure zero.

In other words, these are the things that gave us trouble when we were talking about Fourier transforms. These are the things that gave us trouble when we were

talking about Fourier series. When you take anything in the same equivalence class, time-limited functions, and you form a Fourier series, what happens? All of the things in the same equivalence class have the same Fourier coefficients. But when you go back from the Fourier series coefficients back to the function, then you might go back in a bunch of different ways. So, we started using this limit in the mean notation and all of that stuff. And what we're doing here now is, for these vectors, we're just saying, let's represent a vector as this whole equivalence class.

While we're talking about vectors, we're almost always interested in orthogonal expansions. And when we're interested in orthogonal expansions, the coefficients in the orthogonal expansions are found as integrals with the function. And the integrals with different functions in the same equivalence class are identical. In other words, any two functions in the same equivalence class have the same coefficients in any orthogonal expansion.

So if you talk only about the orthogonal expansion, and leave out these detailed notions of what the function is doing at individual times, then you don't have to worry about equivalence classes. In other words, when you -- I'm going to say this again more carefully later, but let me try to say it now a little bit crudely. One of the things we're interested in doing this taking this class of functions, mapping each function into a set of coefficients, where the set of coefficients are the coefficients in some particular orthogonal expansion that we're using. Namely, that's the whole way that we're using to get from waveforms to sequences. It's the whole -- it's the entire thing we're doing when we start out on a channel with a sequence of binary digits and then a sequence of symbols. And we modulate it into a waveform. Again, it's the mapping from sequence to waveform.

Now, the important thing here about these equivalence classes is, you can't tell any of the members of the equivalence class apart within the sequence that we're dealing with. Everything we're interested in has to do with these sequences. I mean, if we could we'd just ignore the waveforms altogether. Because all the processing that we do is with sequences. So the only reason we have these equivalence classes is because we need them to really define what the functions are.

So, we will come back to that later. Boy, I think I'm going to get done today. That's surprising.

The next idea that we want to talk about is vector subspaces. Again, that's an idea you've probably heard of, for the most part. A subspace of a vector space is a subset of the vector space such that that subspace is a vector space in its own right. An equivalent definition is, for all vectors u and v , in the subspace α times u plus β times v is in s also. In other words, a subspace is something which you can't get out of by linear combinations.

If I take one of these diagrams back here that I keep looking at -- I'll use this one. If I want to form a subspace of this subspace, if I want to form a subspace of this two-dimensional vector space here, one of the subspaces includes u , but not v , perhaps. Now, if I want to make a one-dimensional subspace including u , what is that subspace? It's just all the scalars times u . In other words, it's this line that goes through the origin and through that vector, u . And a subspace has to include the whole line. That's what we're saying. It's all scalar multiples of u . If I want a subspace that includes u and v , where u and v are just arbitrary vectors I've chosen out of my hat, then I have this two-dimensional subspace, which is what I've drawn here. In other words, this idea is something we've been using already. It's just that we didn't need to be explicit about it.

The subspace which includes u and v -- well, a subspace which includes u and v , is the subspace of all linear combinations of u and v . And nothing else. So, it's all vectors along here. It's all vectors along here. And you fill it all in with anything here added to anything here. So you get this two-dimensional space. Is 0 always in a subspace? Of course it is. I mean, you multiply any vector by 0 and you get the 0 vector. So you sort of get it as a linear -- what more can I say?

If we have a vector space which is an inner product space; in other words, if we add this inner product definition to our vector space, and I take a subspace of it, that can be defined as an inner product space also, with the same definition of inner product that I had before. Because I can't get out of it by linear combinations, and the inner

product is defined for every pair of vectors in that space. So we still have a nice, well-defined vector space, which is an inner product space, and which is a subspace of the space we started with.

Everything I do from now on, I'm going to assume that V is an inner product space. And want to look at how we normalize vectors. We've already talked about that. If I have a vector in this vector space that's normalized, if its norm equals 1. We already decided how to normalize a vector. We took an arbitrary vector, u , divided by its norm. And as soon as we divide by its norm, that vector, v , divided by the norm of v , the norm of that is just 1.

So the projection, what the projection theorem says, and all I need here is a one-dimensional projection theorem, it says that v , in the direction of this vector ϕ , is equal to the inner product of u with ϕ times ϕ . That's what we said. As soon as we normalized these vectors, that ugly denominator here disappears. Because the norm of ϕ is equal to -- because the norm is equal to 1.

So, an orthonormal set of vectors is a set such that each pair of vectors is orthogonal to each other, and where each vector is normalized. In other words, it has norm squared equal to 1. So, the inner product of these vectors is just δ_{jk} . If I have an orthogonal set, v_j , say, then ϕ_j is an orthonormal set just by taking each of these vectors and normalizing it. In other words, it's no big deal to take an orthogonal set of functions and turn them into an orthonormal set of functions. The Fourier series was natural to define that, and most people define it, in such a way that it's not orthonormal. Because we're defining it over some interval of time, t , that has a norm squared of t and, therefore, you have to divide by square root of t to normalize it.

If you want to put everything in a common framework, it's nice to deal with orthonormal series. And therefore, that's what we're going to be stressing from now on.

So I want to go on so the real projection theorem. Actually, there are three projection theorems. There's the one-dimensional projection theorem. There's the

n-dimensional projection theorem, which is what this is. And then there's an infinite-dimensional projection theorem, which is not general for all inner product spaces, but is certainly general for L^2 . So I'm going to assume that ϕ_1 to ϕ_n is an orthonormal basis for an n-dimensional subspace, s , which is a subspace of v . How do I know there is such an orthonormal basis? Well, I don't know that yet, and I'm going to come back to that later. But for the time being, I'm just going to assume that as part of the theorem. Assume I have some particular subspace which has the property that it has an orthonormal basis.

So this is an orthonormal basis for this n-dimensional subspace, s and v . For each vector in v , s now is some small subspace. v is a big subspace out around s . What I want to do now is, I want to take some vector in the big subspace. I want to project it onto the subspace. By projecting it onto the subspace, what I mean is, I want to find some vector in the subspace such that the difference between v and that point in the subspace is orthogonal to the subspace itself. In other words, it's the same idea as we used before for this over-used picture. Which I keep looking at.

Here, the subspace that I'm looking at is just the subspace of vectors colinear with u . And what I'm trying to do here is to find, from v I'm trying to drop a perpendicular to this subspace, which is just a straight line. In general, what I'm trying to do is, I have an n-dimensional subspace. We can sort of visualize a two-dimensional subspace if you think of this in three dimensions. And think of replacing u with some space, some two-dimensional space, which is going through 0. And now I have this vector, v , which is outside of that plane. And what I'm trying to do here is to drop a perpendicular from v onto that subspace.

The projection is where that perpendicular lands. So, in other words, v , minus the projection, is this v perp we're talking about. And what I want to do is exactly the same thing that we did before with the one-dimensional projection theorem. And that's what we're going to do. And it works. And it isn't really any more complicated, except for notation.

So the theorem says, assume you have this orthonormal basis for this subspace.

And then you take any old v in the entire vector space. This can be an infinite dimensional vector space or anything else. And what we really want it to be is some element in L_2 , which is some infinite-dimensional element. It says there's a unique projection in the subspace s . And it's given by the inner product of v with each of those basis vectors. That inner product, times $v_{\text{sub } j}$. For the case of a one-dimensional vector, you take the sum out and that was exactly the projection we had before. Now we just have the multi-dimensional projection.

And it has a property that v is equal to this projection plus the orthogonal thing. And the orthogonal thing, the inner product of that with s equal to 0 for all s in the subspace. In other words, it's just what we got in this picture we were just trying to construct, of a two-dimensional plane. You drop a perpendicular two-dimensional plane. And when I drop a perpendicular to a two-dimensional plane, and I take any old vector in this two-dimensional plane, we still have the perpendicularity. In other words, you have the notion of this vector being perpendicular to a plane if it's perpendicular whichever way you look at it.

Let me outline the proof of this. Actually, it's pretty much a complete proof. But with a couple small details left out. We're going to start out the same way I did before. Namely, I don't know how to choose this vector. But I know I want to choose it to be in this subspace. And any element in this subspace is a linear combination of these $p_{\text{sub } i}$'s. So this is just a generic element in s . And I want to find out what element I have to use. I want to find the conditions on these coefficients here such that v minus the projection -- in other words, this v_{perp} , as we've been calling it -- is orthogonal to each $\phi_{\text{sub } i}$. Now, if it's orthogonal to each $\phi_{\text{sub } i}$, it's also orthogonal to each linear combination of the $\phi_{\text{sub } i}$'s. So in fact, that solved our problem for us.

So what I want to do is, I want to set 0 equal to v minus this projection. Where I don't yet know how to make the projection, because I don't know what the $\alpha_{\text{sub } i}$'s are. But I'm trying to choose these so that this, minus the projection, the inner product of that with $\phi_{\text{sub } j}$ is equal to 0 for every j . This inner product here is equal to the inner product of v with $\phi_{\text{sub } j}$. I have this difference here, so the inner product

is the inner product of v with ϕ_j minus the inner product of this with ϕ_j . Let me write that here. $v - \sum \alpha_i \phi_i$, ϕ_j is equal to $\phi_j - \sum_i \alpha_i \phi_i \cdot \phi_j$. Which is equal to this. All of these terms are 0 except where j is equal to i . Where i is equal to j .

So, α_j has to be equal to the inner product of v with this basis vector here. And, therefore, this projection is equal, which we said was $\sum \alpha_i \phi_i$, that's really the sum of $v \cdot \phi_j$, this inner product, times ϕ_j ,

Now, if you really use your imagination and you really think hard about the formula we were using for the Fourier series coefficients, was really the same formula without the normalization in it. It's simplified by being already normalized for us. We don't have that $1/t$ in here, which we had in the Fourier series because now we've gone to orthonormal functions. So, in fact that sort of proves the theorem.

If we express v as some linear combination of these orthonormal vectors, then if I take the norm squared of v , this is something we've done many times already. I just express the norm squared, just by expanding this the sum of -- well, here I've done it this way, so let's do it this way again. When I take the inner product of v with all of these terms here, I get the sum of α_j complex conjugated, times the inner product of v with ϕ_j . But the inner product of v with ϕ_j is just α_j times 1. So it's a sum of α_j squared.

OK, this is this energy relationship we've been using all along. We've been using it for the Fourier series. We've been using it for everything we've been doing. It's just a special case of this relationship here, in this n -dimensional projection, except that there we were dealing with infinite dimensions and here we're dealing with finite dimensions. But it's the same formula, and you'll see how it generalizes in a little bit.

We still have the Pythagorean theorem, which in this case says that the norm squared of vector v is equal to the norm squared of a projection, plus the norm squared of the perpendicular part. In other words, when I start to represent this vector outside of the space by a vector inside the space, by this projection, I wind up with two things. I wind up both with the part that's outside of the space entirely, and

is orthogonal to the space, plus the part which is in the space. And each of those has a certain amount of energy.

When I expand this by this relationship here -- I'm not doing that yet -- what I'm doing here is what's called a norm bound. Which says both of these terms are non-negative. This term is non-negative in particular, and therefore the difference between this and this is always positive, or non-negative. Which says that 0 has to be less than or equal to this, because it's non-negative. And this has to be less than or equal to the norm of v . In other words, the projection always has less energy than the vector itself. Which is not very surprising. So the norm bound is no big deal.

When I substitute this for the actual value, what I get is the sum j equals 1 to n of the norm of the inner product of v , with each one of these basis vectors, magnitude squared, that's less than or equal to the energy in v . In other words, if we start to expand, as n gets bigger and bigger, and we look at these terms, we take these inner products, square them. No matter how many terms I take here, the sum is always less than or equal to the energy in v . That's called Bessel's inequality, and it's a nice, straightforward thing.

And finally, the last of these things -- well, I'll use the other one if I need it. The last is this thing called the mean square error property. It says that if you take the difference between the vector and its projection onto this space, this is less than or equal to the difference between the vector and the other s in the space. Any other -- I can always represent v as being equal to this plus the orthogonal component. So I wind up with a sum here of two terms. One is the difference between -- well, it's the -- it's the length squared of the projection. Write it out. v minus v_s -- let me write this term out. v minus s is equal to v , the projection, plus v perpendicular to the subspace s minus this vector s . This is perpendicular to this and this, so -- ah, to hell with it. Excuse my language. I mean, this is proven in the notes. I'm not going to go through it now because I want to finish these other things. I don't want to play around with it.

We left something out of the n -dimensional projection theorem. How do you find an

orthonormal basis to start with? And there's this neat thing called Gram-Schmidt, which I suspect most of you have seen before also. Which is pretty simple now in terms of the projection theorem.

Gram-Schmidt is really a bootstrap operation starting with the one-dimensional projection theorem, working your way up to larger and larger dimensions. And each case winding up with an orthonormal basis for what you started with. Let's see how that happens.

I start out with a basis for an inner product subspace. So, s_1 up to s_n as a basis for this inner product space. First thing I do is, I start out with s_1 . I find the normalized version of s_1 . I call that ϕ_1 . So ϕ_1 is now an orthonormal basis for the subspace whose basis is just ϕ_1 itself. ϕ_1 is the basis for the subspace of all linear combinations of s_1 . So it's just a straight line in space.

The next thing I do is, I take s_2 . I find the projection of s_2 on this subspace s_1 . I can do that. So I find a part which is colinear with s_1 . I find the part which is orthogonal. I take the orthogonal part, and that's orthogonal, to ϕ_1 . And I normalize it. So I then have two vectors, ϕ_1 and ϕ_2 , which span the space of functions of linear combinations of s_1 and s_2 . And I call that subspace S_2 , capital S_2 . And then I go on. So, given any orthonormal basis, ϕ_1 up to ϕ_k of the subspace s_k generated by s_1 to s_k , I'm going to project s_{k+1} onto this subspace s_k , and then I'm going to normalize it. And by going through this procedure, I can in fact find an orthonormal basis to any set of vectors that I want to, to any subspace that I want to.

Why is this important, is this something you want to do? Well, it's something you can program a computer to do almost trivially. But that's not why we want it here. The thing we want it here is to say that there's no reason to deal with bases other than orthonormal bases. We can generate orthonormal bases easily. The projection theorem now is valid for any n -dimensional space because for any n -dimensional space we can form this basis that we want.

Let me just go on and finish this, so we can start dealing with channels next time. So

far, the projection theorem is just for finite dimensional vectors. We want to now extend it to infinite dimensional vectors. To countably infinite set of vectors. So I'm given any orthogonal set of functions, ϕ_i , we can first generate orthonormal functions as ϕ_i , which are normalized. And that's old stuff. I can now think of doing the same thing that we did before. Namely, starting out, taking any old vector I want to, and projecting it first on to the subspace with only ϕ_1 in it. Then the subspace generated by ϕ_1 and ϕ_2 , then the subspace generated by ϕ_1 , ϕ_2 and ϕ_3 , and so forth. When I do that successively, which is successive approximations in a Fourier expansion, or in any orthonormal expansion, what I'm going to wind up with is the following theorem that says, let ϕ_n be a set of orthonormal functions. Let v be any L^2 vector. Then there exists an L^2 vector, u , such that $v - u$ is orthogonal to each ϕ_n .

In other words, this is the projection theorem, carried on as n goes to infinity. But I can't quite state it in the way I did before. I need a limit in here. Which says the limit as n goes to infinity of u , namely what is now going to be this projection, minus this term here, which is the term in the subspace of these orthonormal functions. This difference goes to zero.

What does this say? It doesn't say that I can take any function, v , and expand it in an orthonormal expansion. I couldn't say that, because I have nothing to know whether this arbitrary orthonormal expansion I started with actually spans L^2 or not. And without knowing that, I can't state a theorem like this. But what the theorem does say is, you take any orthonormal expansion you want to. Like the Fourier series, which only spans functions which are time limited. I take an arbitrary function. I expand them in this Fourier series. And, bingo, what I get is a function u , which is the part of v , which is within these time limits and what's left over, which is orthogonal, is the stuff outside of those time limits. So this is the theorem that says that in fact you can --

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: What?

AUDIENCE: [UNINTELLIGIBLE]

PROFESSOR: It's similar to Plancherel. Plancherel is done for the Fourier integral, and in Plancherel you need this limit in the mean on both sides. Here we're just dealing with a series. I mean, we still have a limit in the mean sort of thing, but we have a weaker theorem in the sense that we're not asserting that this orthonormal series actually spans all of L^2 . I mean, we found a couple of orthonormal expansions that do span L^2 . So it's lacking in that.

OK, going to stop there.