TheFourierTransformAndItsApplications-Lecture01

Instructor (Brad Osgood):We are on the air. Okay. Welcome, one and all. And as it said on the TV when you were walking in, but just to make sure everybody knows, this is EE261, The Fourier Transform and its Applications, Fourier Transforms et al., Fourier. And my name is Brad Osgood.

Circulating around are two documents that give you information about the class. There's a general description of the class, course information, how we're gonna proceed, some basic bookkeeping items — I'll tell you a little bit more about that in just a second — and also a syllabus and a schedule, and I will also say a little bit more about that in just a second.

Let me introduce our partners in crime in this course. We have three course assistants, Thomas John — Thomas, wanna stand up? Where's Thomas? There we go. Rajiv Agarwal. Did I spell that right? Very good. Rajiv, you wanna stand up? There's Rajiv. And Nicomedus — okay so far? — M. Wanna correct that? Okay. That's Nicomedus, everybody. Thank you. All right.

And we will be setting up times for the review sessions and so on, all right? So that will be forthcoming.

We have a web page for the course. Some of you may have already visited that, but let me give you the — and the address is on one of the sheets that's being passed around, but let me write that up now so you can be sure to visit it and register for the class because it is on the web page that you will find course handouts, course information.

I will email people via the web page, all right? So you have to be registered. If I have to send an announcement to the class, post an announcement and send out an email, then that'll be done through the web page, and you have to be registered on the web page in order to get those emails. I won't be doing it through Axess, all right?

So it is at, like many of the other E classes, http:// — however you do that, wherever the colons go — where is it here? — eeclass.stanford.edu — you can find it very easily — edu/ee261, okay? Go there if you have not already and register yourself for the class. All right.

Now, let me say a little bit about the information that you have. I wanna say a little bit more about the mechanics. I'll talk more about the content in just a second. Let me say a little bit about the syllabus and schedule and the course reader. The syllabus is, as I said on the top, an outline of what we're gonna be doing, I hope a fairly accurate outline of what we're going to be doing, but it's not a contract, all right? So there will be a natural ebb and flow of the course as things go along, and when we get to particular material or what we cover in what order, this is more or less I say accurate, but it is not written in stone.

What you should use it for, however, is to plan your reading, so things will be much better for all of us if you read along with the material as the syllabus — as the schedule basically outlines, all right, because there are times when I'm gonna wanna skip around a little bit. There are times when I'm gonna derive things. There are times when I'm not gonna derive things. And you'll get much more out of the lectures, our time together, if you've read the material thoroughly before you come to class. So that's one thing I ask you to do.

We have two exams scheduled. We have a midterm exam and a final exam. I'm gonna skip to the midterm exam. Midterm exam is already actually on there, at least tentatively, sort of toward the end of October. We'll have it outside of class. That is, it'll be a sit-down, regular exam, but I wanna do it for 90 minutes rather than 50 minutes. Fifty minutes is just too short a time for material like this, so it'll be a 90-minute exam, and we'll schedule it several sessions outside of class. This is the way I've usually done it, and it hasn't been any problem. It's worked out all right for everybody. So we'll have alternate times and so on.

And the final exam is scheduled by the registrar's office. Do not come to me right before the final exam saying, "Oh, I scheduled a trip out of town. I hope that's not a problem," all right? You know what the dates are ahead of time.

We'll also have regular problem sets. None of these things that I'm saying should be new to you. You've been through the drill many times. The problem sets are gonna be — I had a startling innovation last time I taught the course where I handed out the problem sets on Monday and had them due the following Wednesday. So you actually had, like, a week and a half to do the problem sets, and there was overlap between the two. And people thought that was just a brilliant idea.

So we're gonna do that again this year except for the first problem set, all right? I decided it was not such good policy to hand out the very first problem set on the very first day of class, so I'll hand that out on Wednesday and I'll post that also — or at least I'll post it. I'm not sure I'll hand it out. It will be available on Wednesday, and it'll be due the following Wednesday. And again, these sorts of things are pretty routine for you. I'm sure you've been through them many times.

It will be practice, although again not necessarily every time without fail, to have MATLAB problems on the homework, one or two MATLAB problems on the homework. So I'm going under the assumption that people have some experience with using MATLAB — it doesn't have to be terribly advanced — and also access to using MATLAB. So if you do not have experience using MATLAB and you do not have access to MATLAB, get some experience and get some access. Won't be hard.

Okay. Now, let me say a little bit about some other things. This is the course reader for the course. It's available at the bookstore and also available on the course website, all right? It doesn't have the problems in it, but it has the material that we're gonna be covering in class.

Now, this is basically a stitched-together set of lecture notes that I've been using for a number of years in the class, and I sort of tinker with it every time I teach the class. But because it is a stitched-together set of lecture notes, the organization is sometimes a little bit odd, like you have an appendix in the middle of the chapter. And what that means is it used to be an appendix to a particular lecture that went on that particular day, and it never got moved to anywhere else, all right? So the organization can be a little bit funny.

You can help on this, all right? That is, if you find typos, if you find errors, if you find things that are less than clear in their wording, if you have some ideas for examples or other explanations, please tell me. I am working on this. I have to say that because these were written as a set of lecture notes, these are meant to be a good and I hope helpful companion for the class. That is, they're meant to be read, and they're meant to be used. So you can help, as generations of students in the past have helped, to try to refine these and turn them into something that's really a good accompaniment to the class as we go on, okay?

One other thing that's special this quarter is the class is, as always, taped, and the lectures are gonna be available to everybody. But this time for the first time, the lectures are gonna be available to the world, all right? Stanford has decided on an experimental basis — we're sort of competing with MIT here, I think — to try to make some of the materials for some classes available to the world, all right?

So the lecture notes are gonna be — everything's gonna be done through the website, but instead of needing a Stanford ID to view the taped lectures, I think anybody in the world can view these lectures. It's a little bit daunting. I have to watch my language, gotta dress well. All right. So we'll see what goes with that.

I will, however, issue a warning. I will not answer the world's email, all right? I will answer email from the class, but I will not answer — and I think I speak for the TAs here. The TAs neither will answer the world's email on this, all right? How we're gonna keep the world out of our inboxes, I'm not sure exactly whether this is gonna be a problem or not, but at any rate, that's what's happening, okay?

All right. Any questions about that? Any questions about the mechanics of the course or what your expectations should be, what my expectations of you are? Okay. All right.

Now, I always like to take an informal poll actually, when we start this class. I've taught this a number of times now, and it's always been a mixed crowd. And I think that's one of the things that's attractive about this class. So let me ask who are the EEs in this class, who are in electrical engineering, either undergraduate or graduate? All right. So that's a pretty strong show of hands. But let me also ask, who are the non-EEs in this class? All right. That's also a pretty strong show of hands.

The EEs are, as is typical, the majority of the students in the class, but there's also a pretty strong group of students in this class who are not electrical engineers by training, by desire, by anything, all right? And they usually come from all over the place. I was looking at the classes before I got to class, and I think there are some people from chemistry. Is somebody from chemistry? I thought there was somebody — yeah, I see back there. All right. And I know there were some people from earth sciences. And somebody was talking to me actually from earth sciences this morning. Somebody from earth science? Okay. Where else? I think there was an ME, couple of MEs maybe. Yeah, all right.

Now, that's important to know, I think. The course is very rich in material, all right, rich in applications, rich in content, and it appeals to many people for many different reasons, okay? For the EEs who are taking the class, you have probably seen a certain amount of this material — I don't wanna say most of the material, but you've probably seen a fair amount of this material scattered over many different classes.

But it's been my experience that one of the advantages of this class for electrical engineering students, either undergraduate or graduate students, is to see it all in one piece, all right, to put it all in your head at one time, at least once, all right, because the subject does have a great amount of coherence. It really does hang together beautifully. For all the different and varied applications, there are core ideas and core methods of the class that it is very helpful to see all at once, all right?

So if you have seen the material before, that's fine. I mean, you can draw on that and draw on your experience, but don't deny yourself the pleasure of trying to synthesize the ideas as we go along. I mean, there's nothing so pleasurable as thinking about something you already know, trying to think about it from a new perspective, trying to think about it from a new point of view, trying to fold it into some of the newer things you'll be learning.

So I have heard this from electrical engineering students many times in the past that it's a pleasure for them to see the material all together at once. It may seem like a fair amount of review, and in some cases, it will be, but not in all cases. And even if it is a review, there are often slightly different twists or slightly new takes on things that you may not have seen before or may not have thought of in quite that way. So that is my advice to the electrical engineering students.

For the students who have not seen this material before that are coming at it from a different field and maybe only heard secret tales of the Fourier transform and its uses, well, I hope you enjoy the ride because it's gonna be a hell of a ride — oh, a heck of a ride as we go along. All right.

Now, for everyone, I sort of feel like I have to issue — I don't know if I'd call this a warning or just sort of a statement or principle or whatever. This is a very mathematical class. This is one of the sort of holy trinity of classes in the information systems lab and electrical engineering. Electrical engineering is a very broad department. It's split up into a number of laboratories along research lines. I am in the information systems lab, which is sort of the mathematical part of the subject, has a lot of signal processing, coding theory, imaging, and so on.

And this course has been for a number of years taught by faculty, sort of thought of as a cornerstone in the signal processing, although it has a lot of different applications to a lot of different areas. The other courses in that holy trinity are 263, linear dynamical systems, and 278, statistical signal processing.

Actually, let me ask here because this is also very common: Who's taking 263 in the class? Also a strong majority. And who's taking 278? Yeah, okay. So there's a little bit less but still a number of people. We'll actually see, not so much with 278, well, actually with both classes, with 263 and 278, you'll actually see some overlap that I also hope you find interesting. The language will be slightly different. The perspective will be slightly different, but you see this material in this class melding over into the other classes and vice-versa. And again, I think it's something that you can really draw on and I hope you enjoy, all right?

So those classes and the perspective that we take, the faculty who are teaching this class, is a pretty mathematical one, but it's not a class in theorems and proofs. You can breathe a heavy sigh of relief now, all right? I can do that, but I won't, all right?

I will derive things — I'll derive a number of formulas — I'll derive it and I'll go through those derivations, or I'll hope that you go through the derivations in the book when I hope and I think that they will be helpful, all right? And when, in some cases, that is, there's an important technique or there's an important idea that you'll see not only in the particular instance but overall, that you'll see the same sort of derivation or the same sort of ideas be applied not only for one formula but for other sorts of formulas.

And also, in some cases, to my mind, as twisted as that may be, I sometimes think of the derivation of a formula almost as identical with the formula. I mean, to use the formula effectively almost is to know the derivation because it's to know where it applies and to know how it applies and where do we expect to use it, all right?

So that's why I will go through those things for the purpose of teaching a certain amount of technique and for the purposes of sort of having those techniques really at your fingertips so that you can apply them again in a situation that may not be quite identical with what we did but will be similar enough so that the ideas may apply in that situation. That's very important.

We will also do plenty of different sorts of applications. And again, because the field, the subject is so varied and because the clientele, because the students in the class are also varied, we'll try to take applications from different areas. We'll have applications from electrical engineering, but we'll also have applications from physics and from other areas.

I've also done in the past, and we'll see if I can get to this, some applications from earth sciences, for example. And we'll just see how that goes. So we all have to cut each other a little bit of slack, and if an application or particular area is not exactly to your liking, well, chances are it might be to somebody's liking to your right or left. So, as I say, everybody should cut each other a little slack and just enjoy the ride.

I should also say that many of the more specialized applications are found in more specialized courses, all right? So we will touch on a lot of things, and I will use the words that are used in a lot of different courses and a lot of different subjects. But we won't always do C in application to its bitter end, so to speak, or we won't do every — we certainly won't do every possible application because there are just so many of them. So you will find you will not run out of ways of using the Fourier transform and Fourier analysis techniques in any classes here. It goes on and on and on. But we'll only be able to see a certain amount of that, all right?

And actually, that leads to a very important point, at least at the start of the class; that is, where do we start, all right? That is, this subject, which is so rich and so diverse, forces you, forces me, forces all of us to make hard choices in some ways about what we're gonna cover, where we're gonna start, what direction we're going to go. And all the different choices are defensible. You will find books out there that take very different tactics towards the subject. They take different starting points; they have different emphasis; they go off in different directions. And you can make a good argument for any one of those choices. But you have to make a choice.

So for us, we are going to choose — I have chosen, not we, me — I have chosen to start the class with a brief discussion of Fourier series and go from there to the Fourier transform, all right, whereas it is also a very common choice to forget about Fourier series and maybe pick them up a little bit along the end or pick them up a little bit on the edges or, assuming that everybody's seen Fourier series, and then go right into the Fourier transform.

I don't wanna do that because I think the subject of Fourier series is interesting enough in it — we're not gonna do very much with it, but it's interesting enough in itself. Again, it's something you may have seen in different contexts, but it provides a natural transition to the study of the Fourier transform. And it is historically actually the way the subject developed, okay? So that's how we're gonna do things. We'll start with Fourier series and use them as a transition to Fourier transform.

Now, first of all, what is this concerned with overall? It may be a little bit too strong a statement, but for our purposes, I wanna identify the idea of Fourier series as almost identified with the study of periodic phenomena, all right? So for us, it's identified most strongly with a mathematical analysis of periodic phenomena.

Now, it certainly shouldn't be necessary for me to justify periodic phenomena as an important class of phenomena. You have been studying these things for your entire life pretty much. Ever since the first physics course you ever took where they do the harmonic oscillator, and then the second physics course you took where they did the harmonic oscillator, and then the third physics course you took where they did the harmonic oscillator, you have been studying periodic phenomena, all right? So that shouldn't be a controversial choice. Fourier series goes much beyond that, but it is first and foremost for us associated with the study of periodic phenomena.

The Fourier transform, although again it maybe doesn't do it justice completely, can be viewed as a limiting case of Fourier series. It has to do with a mathematical analysis on non-periodic phenomena. So if you wanna contrast Fourier series and Fourier transforms, then that's not a bad rough and ready way of doing it. As I say, it doesn't capture everything, but it captures something.

So Fourier transform as a limiting case, in a meaning that I'll make more precise later — as a limiting case of Fourier series or Fourier series techniques is identified with or has to do with — is concerned with, how about that for a weaseling way out of it? — is concerned with the analysis of non-periodic phenomena. So again, it doesn't say everything, but it says something.

And one of the things that I hope you get out of this course, especially for those of you who have had some of this material before, are these sort of broad categorizations that help you sort of organize your knowledge, all right? It's a very rich subject. You gotta organize it somehow. Otherwise, you'll get lost in the details, all right? You wanna have certain markers along the way that tell you how to think about it, how to organize it, what a particular formula, what general category it fits under, okay?

Now, it's interesting that the ideas are sometimes similar and sometimes quite different. And sometimes the situation is simpler for periodic phenomena; sometimes the situation's more complicated for periodic phenomena. So it's not as though there's sort of a one-to-one correspondence of ideas. But that's one of the things that we'll see and one of the reasons why I'm starting with Fourier series is to see how the ideas carry over from one to the other, see where they work and see where they don't work, all right? Some ideas carry easily back and forth between the two, some phenomena, some ideas, some techniques; some don't. And it's interesting to know when they do and when they don't. Sometimes the things are similar, and sometimes they're not.

Now, in both cases, there are really two kind of inverse problems. There's the question of analysis, and there's the question of synthesis, two words that you've used before, but it's worthwhile reminding you what they mean in this context. The analysis part of Fourier analysis has to do with breaking a signal or a function — I'll use the terms signal and function pretty much interchangeably, all right? I'm a mathematician by training, so I tend to think in terms of functions, but electrical engineers tend to think in terms of signals, and they mean the same thing, all right?

So analysis has to do with taking a signal or a function and breaking it up into its constituent parts, and you hope the constituent parts are simpler somehow than the complicated signal as it comes to you. So you wanna break up a signal into simpler constituent parts. I mean, if you don't talk just in terms of signals here or you don't use exactly that language, that's the meaning of the word analysis, I think, close enough; whereas, synthesis has to do with reassembling a signal or reassembling a function from its constituent parts — a signal from its constituent parts, all right?

And the two things go together, all right? You don't want one without the other. You don't wanna break something up into its constituent parts and then just let it sit there, all these little parts sitting on the table with nothing to do. You wanna be able to take those parts and maybe modify those parts, maybe see which parts are more important than other parts, and then you wanna put them back together to get either the original signal or a new signal. And the process of doing those things are the two aspects of Fourier analysis. I use the word analysis there sort of in a more generic sense.

Now, the other thing to realize about both of these procedures, analysis and synthesis, is that they are accomplished by linear operations. Series and integrals are always involved here. Both analysis and synthesis, Fourier analysis and synthesis, are accomplished by linear operations. This is one of the reasons why the subject is so — I don't know — powerful because there is such a body of knowledge, and there's such a deep and advanced understanding of linear operations, linearity.

We'll make this a little bit more explicit as we go along further, but I wanted to point it out now because I won't always point it out, all right, because when I say linear operations, what I'm thinking of here, integrals and series, all right, e.g., i.e., integrals and series, both of which are linear operations. The integral of a sum is the sum of the integrals. The integral of a constant times a function is the constant times the interval of the function, and similarly with sums, all right?

Because of this, one often says or one often thinks that Fourier analysis is part of the study of linear systems, all right? In engineering, there are courses called linear systems and so on, and sometimes Fourier analysis is thought to be a part of that because the operations involved in it are linear. I don't think of it that way. I mean, I think it's somehow important enough on its own not to think of it necessarily as subsumed in a larger subject. But nevertheless, the fact that the operations are linear does put it in a certain context; in some ways, in some cases, a more general context that it turns out to be important for many ideas, all right?

So you see — you often hear that Fourier analysis is a part of the subject of linear systems, the study of linear systems. See, I don't think that really does complete justice to Fourier analysis because of the particular special things that are involved in it, but nevertheless, you'll hear that.

Okay. Now, let's get launched, all right? Let's start with the actual subject of Fourier series and the analysis of periodic phenomenon, so periodic phenomena and Fourier series. As I said, it certainly shouldn't be necessary for me to sell the importance of periodic phenomena as something worth studying. You see it everywhere, all right?

The study of periodic phenomena is for us the mathematics and engineering or mathematics and science and engineering of regularly repeating phenomena. That's what's always involved. There's some pattern that repeats, and it repeats regularly, all right? So it's a mathematics and engineering — since this is an engineering course, I'll put that before science, or maybe I won't even mention science — mathematics and engineering of regularly repeating patterns. I'm leaving a couple of terms here — I'm leaving all these somewhat vague. What does it mean to be regular? What does it mean to be repeating? What is a pattern in the first place? But you know what I mean. You know it when you see it. And the fact that you can mathematically analyze it is what makes the subject so useful.

Now, I think, although again, it's not ironclad — the trouble is this subject is so rich that every time I make a statement, I feel like I have to qualify it. Well, it's often true, but it's not completely true, and sometimes it's not really true at all, but most of the time it's true, that it's helpful, but not always helpful, but most of the time helpful, occasionally helpful to classify periodicity as either periodicity in time or periodicity in space, all right? You often see periodic phenomena as one type or the other type, although they can overlap.

So periodic phenomena often are either periodicity in time — a pattern repeats in time over and over again. You wait long enough, and it happens again. So, for example, harmonic motion, so e.g., harmonic motion, a pendulum, a thing bobbing on a string, [inaudible] harmonic motion, or periodicity in space — I'll go over here — periodicity in space.

All right. Now, what I mean here is there is often a physical quantity that you are measuring that is living on some object in space, one dimension, two dimensions, whatever, that has a certain amount of symmetry, all right? And the periodicity of the phenomenon is a consequence of the symmetry of the object. I'll give you an example in just a second.

So here you have, say, some physical quantity, physical — not always, but often — physical quantity distributed over a region with symmetry. The region itself repeats, all right? The region itself has a repeating pattern, all right? So the periodicity of the phenomenon, the periodicity of the physical quantity that you're measuring is a consequence of the fact that it's distributed over some region that itself has some symmetry. So periodicity arises from the symmetry. Periodicity here of the physical quantity that you're measuring arises because the symmetry of the object where it's distributed, where it lives.

I'll give you an example [inaudible] from the symmetry. Matter of fact, I'll give you the example. The example that really started the subject, and we'll study this, is the distribution of heat on a circular ring, so, e.g., the distribution of heat on a circular ring, all right? So the physical quantity that you're interested in is the temperature, but it's the temperature associated with a certain region, and the region is a ring, all right? The ring has circular symmetry. It's round, okay? So you're measuring the temperature at points on the ring, and that's periodic because if you go once around, you're at the same place. So the temperature is periodic as a function of the special variable that describes where you are on the ring.

Time is not involved here. Position is involved, all right? It's periodic in space, not periodic in time, periodic in the spatial variable that gives you the position. And the periodicity arises because the object itself is symmetric because the object repeats.

This sort of example is why one often sees — and this actually turns out to be very far-reaching and quite deep — that Fourier analysis is often associated with questions of symmetry. In its sort of most mathematical form, you often find Fourier theories developed in this context and Fourier transforms developed in the context of symmetry. So you often see — so you see Fourier analysis, let me just say, Fourier analysis is often associated with problems — with analysis of questions that have some sort of underlying symmetry, so let me say often associated with problems with symmetry. I'll just leave it very general.

This is the very first problem. The problem of distribution of heat on a ring, we're gonna solve that problem. That was the problem that Fourier himself considered, all right, that introduced some of the methods into the whole subject that launched everything, all right? So again, it's not periodicity in time. It's periodicity in space.

And for those of you who have had or may have courses in this, the mathematical framework for this very general way of looking at Fourier analysis is group theory because the theory of groups in mathematics is a way of mathematizing the idea of symmetry and then one extends the ideas of Fourier analysis to take into account of groups; that is to say, to take into account the symmetry of certain problems that you're studying. And it really, as I say, it's quite far-reaching. We're not gonna do it. We'll actually have a few occasions to go into this but with a light touch, okay? I'm just giving you some indication of where the subject goes.

All right. Now, what are the mathematical descriptors of periodicity? Well, nothing I've said so far I'm sure is new to you at all. You just have to trust me that at some point before you know it, some things I say to you will be new, I hope.

But one of the mathematical descriptors of periodicity, again in the two different categories, say, the numbers, the quantities that you associate with either a phenomenon that's periodic in time or a phenomenon that's periodic in space, but periodic in time — for periodicity in time, you often use the frequency, all right? Frequency is the word that you hear most often associated with a phenomena that is periodic in time. You use frequency, the number of repetitions, the number of cycles in a second, say. If the pattern is repeating, whatever the pattern is, again if I leave that term sort of undefined or sort of vague, it's the number of repetitions of the pattern in one second or over time, all right? That's the most common mathematical descriptor of a phenomenon that's periodic in time.

For a phenomenon that's periodic in space, you actually use the period. That's the only word that's really in use in general for the particular — well, one thing at a time. So for periodicity in space, you use the period, all right? That is sort of the physical measurement of how long the pattern is before it repeats somehow, all right, the measurement of how — whether it's length or some other quantity — measurement of — let me just say how big the pattern is that repeats. They're not the same, all right? They have a different feel. They arise often from different sorts of problems.

This is probably too strong a statement, but I think it's fair to say that mathematicians tend to think in terms of mostly in periodic — they tend to think in terms of the period of a function or the period as the description of periodic behavior, whereas engineers and scientists tend to think of systems evolving in time, so they tend to think in terms of frequency. They tend to think of how often a pattern repeats over a certain period of time, all right? That's, like everything else, that statement has to be qualified, but I get tired of qualifying every statement, so I'll just leave it at that.

Now, of course, the two phenomena are not completely separate or not always completely separate. They come together. Periodicity in time and periodicity in space come together in, for example, wave motion, all right? That is a traveling disturbance, a traveling periodic disturbance. So the two notions of periodicity come together. Two notions here, periodicity in time, periodicity in space, come together in, e.g., wave motion, understood very generally here as a periodic — as a regularly repeating pattern that changes in time, that moves.

This board jumps up a little bit. I think I'd better skip it. So a regularly moving disturbance, like a group of freshmen through the quad, just they're everywhere, mostly regularly, mostly moving, all right?

Now, there again, the two descriptors come in, the frequency and the wavelength. So again, you have frequency and wavelength in a frequency nu and wavelength usually denoted by — this is for periodicity in space and then for periodicity in time frequency nu for periodicity in time. That's the number of times that it repeats in one second. This is cycles per second, the number of times that the pattern repeats in one second.

So, for example, you fix your position in space. Both time and space are involved. So you fix yourself at a point in space, and the phenomenon washing over you like a water wave, all right? And you count the number of times you're hit by a wave in a second, and that's the frequency. That's the number of times that the phenomenon comes to you. For periodicity in time, the phenomenon comes to you. For periodicity in space, you come to the phenomenon, so to speak, all right?

So I've fixed myself at a point in time. A wave washes over me at a certain characteristic frequency. Over and over again, regularly repeating, it comes to me nu times per second. The wavelength, you fix the time and allow — and see what the phenomena looks like distributed over space. So for periodicity in space, fix the time and see how the phenomena is distributed. See the pattern distributed over space, distributed — my writing is getting worse — distributed. Then the length of one of those complete [inaudible] is the period or the wavelength. Wavelength is a term that's associated with a periodicity in space for a traveling phenomena, for wave motion. So the length of the disturbance, say one complete disturbance, if I can say that, one complete pattern is the wavelength.

Now, like I say, ever since you were a kid, you've studied these things, and it's really denoted by Landen. But I bring it up here because of the one important relationship between frequency and wavelength, which we are going to see in a myriad of forms throughout the quarter. That is, in the case of wave motion, there is a relationship between the frequency and the wavelength determined by the velocity, and that can be two different phenomenon, all right?

Periodicity in time and periodicity in space may not have anything to do with each other, but if you have a wave traveling, if you have a regularly repeating pattern over time, then they do have something to do with each other and they're governed by the formula, distance equals rate times time, which is the only formula that governs motion, all right?

So the relationship between frequency and wavelength; that is, distance equals rate times time — I love writing this in a graduate course because it's the only equation in calculus actually. In all of calculus, I think this is pretty much the only equation, used in very clever ways, but the only equation.

And in our case, if the rate is the velocity of the wave, then this translates, if V is the velocity, the rate of the wave of the motion, then the equation becomes, as I'm sure you know many times, Lambda. That's the distance that the wave travels in one cycle. It's traveling at a speed, V. If it goes nu cycles in one second, then it goes one cycle in one over nu seconds.

Let me say that again to make sure I got that right. If it goes nu cycles in one second — if it rushes past you nu times in one second, then in one over nu seconds, it rushes past you once. Rushing past you once means you've got through one wavelength. So distance equals rate times time. The time it takes to go one wavelength is one over nu seconds. So I have Lambda equals V times one over nu, or Lambda nu equals V, again a formula you have seen many times.

Now, why did I say this if you've seen it many times? Because I never have the confidence that I can talk my way through that formula, for one thing, so I always have to do it. Secondly, it exhibits a reciprocal relationship between the two quantities, all right? There's a reciprocal relationship — you can see it more clearly over here where the constant of proportionality or inverse proportionality is the velocity, all right? Lambda is proportional to the reciprocal of the frequency, or the frequency is proportional to the reciprocal of the wavelength, at any rate, or expressed this way: Lamba times nu is equal to V, so there's a reciprocal relationship between the frequency and the wavelength, all right? This is the first instance when you talk about periodicity of such reciprocal relationships.

We are gonna see this everywhere, all right? It's one of the characteristics of the subject, hard to state as a general principle, but there, plain to see that in the analysis and the synthesis of signals using methods from Fourier series or Fourier analysis, there will be a reciprocal relationship between the two, between the quantities involved, all right? I'm sorry for being so general, but you'll see this play out in case after case after case, and it is something you should be attuned to, all right? All right.

So you may never have thought about this in these terms. It's a simple enough formula that you've used millions of times, all right? You may not have thought about it somehow in those terms, but I'm asking you to think about stuff you once saw in very simple contexts and how those simple ideas sort of cast a shadow into much more involved situations, all right?

The reciprocal relationship between, as we'll learn to call it, the reciprocal relationship between the two domains of Fourier analysis, the time domain and the frequency domain, or the time domain and the spatial domain, or the spatial domain and the frequency domain, and so on, is something that we will see constantly, all right? And I will point that out, but if I don't point it out, you should point it out to yourself, all right? You should be attuned to it because you will see it.

And it's one of those things that helps you organize your understanding of the material because sometimes when you're called upon to apply these ideas in some context that you haven't quite seen. You have to ask yourself — at least a good starting place is to ask yourself questions like, "Well, should I expect a reciprocal relationship here?" It might lead you to guess what the formula should be or guess what the relationship should be, so you'll say, "Well, somehow I wanna use Fourier analysis to do this problem, so I should be looking for some sort of reciprocal relationship. The quantities that I'm interested in somehow should be related in some kind of reciprocal way." And what that might mean might be more or less involved depending on the particular kind of problem, but you'll see it. Trust me, you'll see it. Okay. All right. Now, we're almost done for today.

Why does mathematics come into this in the first place? I mean, periodicity is evidently sort of a very physical-type property. Why does it allow any kind of mathematical description? Well, it does because there are very simple — maybe not too simple — mathematical functions that exhibit periodic behavior and so can be used to model periodic phenomenon. So math comes in because there are simple mathematical functions that model — that are periodic that repeat and so can be used to model periodic phenomenon.

I am speaking of course of our friends, the sine and cosine. Now, you may think, again, we've only talked about elementary things in a very elementary context. But, you know, I have a PhD in this subject, and I get excited talking about sines and cosines. I mean, and it's not just creeping old age. I mean, I think there's a lot to reflect on here and sometimes the miraculous nature of these things.

Cosine of — I'll use T as the variable. Cosine of T and sine of T are periodic of period 2 pi. That is, cosine of T plus 2 pi is equal to cosine of T for all [inaudible] of T and sine of 2 pi plus T plus 2 pi is equal to sine of T. Why? Dead silence. Because the sine and cosine are — don't tell me. I wanna do it. Because the — I'll do it over here — because the sine and the cosine are associated with periodicity in space because the sine and the cosine are associated with an object that regularly repeats. The simplest object that regularly repeats, the circle.

You didn't meet sine and cosine that way first. You met sine and cosine in terms of ratios of size, lengths of size in triangles. That's fine, but that's an incomplete definition. The real way of — not the real way, but the more sophisticated way, the ultimately more far-reaching way of understanding sine and cosine is as associated with the unit circle where a cosine of T is the X coordinate, and the sine of T is the Y coordinate, and T is a radian measure.

I'm not gonna go through this in too much detail, but the point is that the sine and the cosine are each associated with the phenomenon of periodicity in space. They are periodic because if you go once around the circle; that is to say, T goes from T to T plus 2 pi, you're back where you started from, all right? That's why. It's periodicity in space, all right?

That's the definition of sine and cosine that exhibits their periodic phenomenon, not the definition in terms of right triangles. It's not that the definition in terms of right triangles is wrong; it just doesn't go far enough. It's incomplete, all right? It doesn't reveal that fundamental link between the trigonometric functions and periodicity, and it is fundamental. If not for that, mathematics could not be brought to bear on the study of periodic phenomena.

And furthermore, it's clear — and we'll quit in just a second — that it's not just 2 pi but any multiple of 2 pi, positive or negative. I can go clockwise or I can go counter-clockwise. I can say the cosine of T plus 2 pi N is the same thing a cosine of T. And the sine of 2 pi — T plus 2 pi N is the sine of T for N — any integer — and 0 plus or minus 1, plus or minus 2, and so on and so on.

The interpretation is that when N is positive, I'm going counter — and it is just an interpretation; it is just a convention — when N is positive, I'm going counter-clockwise around the circle; when N is negative, I'm going clockwise around the circle. But it's only when you make the connection between periodicity and space and the sine and the cosine that you see this fundamental property, all right?

Now, all right, I think we made it out of junior high today. That was my goal, all right? What is most amazing and what we'll see next time is that such simple functions can be used to model the most complex periodic behavior, all right? From such simple things — from simple acorns, mighty oaks grow, or whatever bullshit — oh, excuse me — whatever stuff you learn out there — that these simple functions that are associated with such a simple phenomena can be used to model the most complex, really, the most complex periodic phenomena. And that is the fundamental discovery of Fourier series, all right? And it's the basis of Fourier analysis, and we will pick that up next time. Thank you very much. See you then.

[End of Audio]

Duration: 52 minutes

TheFourierTransformAndItsApplications-Lecture02

Instructor (Brad Osgood):All right. A few announcements – a few housekeeping announcements. Thomas checked the – Thomas John, our estimable, one of the Eskimo Bowl TA's for the course, checked the website and we only had about 80 or so students who have signed up on the website, out of 130 or 140 or so that are actually signed up for the course of access. So please register for the website; that's the way you'll be able to get email messages and important announcements and post things on the bulletin board. So go do that.

Student:[Inaudible]

Instructor (Brad Osgood):Pardon me?

Student:[Inaudible] room number, I don't –

Instructor (Brad Osgood):It should be now. I think there was a problem yesterday, briefly. It was not – there was a setting I had to change on the website, but if you haven't checked – if you haven't tried it since the first time you tried it, try it again. Okay? You should be okay. And Thomas also wanted to make an announcement about when the review sessions and office hours are gonna be. Do you need a microphone?

Student: Okay. So the review sessions have been set. The first review session will be on Friday, this coming Friday from 4:15 to 5:05, in Skilling 191, it's the room just on top. Now, we are not expecting every one of you to show up. And please, not all of you show up because we can only accommodate 30 people or so.

Now, these review sessions will be available on the SCPD website. And what we will be covering – well, main topics over the week and also you giving you hints for the homeworks.

Second – our second main – our third main thing would be the office hours for the TA's have been set. Information is available on the course website under the link of course staff. You'll see on the left-hand side there's a link to course staff, and our individual office hours have been set, and they will start on Monday, October 1st.

Instructor (Brad Osgood):Thank you.

Student: Sorry.

Instructor (Brad Osgood):Sorry.

Student:[Inaudible]

Student: Room number for the review sessions? Skilling 191.

Instructor (Brad Osgood):Skilling 190 – 191, did you say for the review session?

Student: No audible response.

Instructor (Brad Osgood):Okay. Also, I forgot to mention that the homework – first homework set has also been posted up on the web and it is due next Wednesday. Okay. Any questions about anything? Anything on anybody's mind? We haven't done much yet, so there shouldn't be that many questions. All right. So today, I wanna continue our study and begin a real – serious mathematical study of the question of periodicity. Remember that we are essentially identifying the subject of Fourier series with the study – with the mathematical study of periodicity. And last time I went on, at some length, about the virtues of periodicity, about the ubiquitous nature of periodic functions – periodic phenomena in the physical world, and also in the mathematical world. And we made a distinction, perhaps a little bit artificial but sometimes helpful, between periodicity in time and periodicity in space. Those sort of two phenomena seem to be, or often come to you in different forms, and it's sometimes useful in your own head to sort of ask yourself which kind of periodicity are you looking at? But in all cases actually, periodicity is associated with the idea of symmetry. That's the topic that will come up from time to time, and if I don't mention it explicitly, as with many other things in this course, it's one of the things that you should learn to sort of react to or think about yourself – see what aspects of symmetry are coming up in the problem, how does a particular problem fit into a more general context because, as I've said before and will say it again, one of the wonderful things about this subject is the way it all hangs together and how it can be applied in so many different ways. All right. If you understand the general framework and put yourself – and orient yourself in a certain way, using the ideas and the techniques of the class, you'll really find how remarkably applicable they can be. Okay. So I said – as I said last time – as we finished up last time, when we're sort of still just crawling our way out of junior high, a mathematical course of periodicity is possible because there are very simple mathematical functions that exhibit periodic behavior, namely the sine and the cosine. But that's also the problem because periodic phenomena can be very general and very complicated, and the sine and the cosine are so simple. So how can you really expect to use the sine and the cosine to model very general periodic phenomena? And that's really the question I want to address today. So how could we use such simple functions – sine of t and cosine of t – to model complex periodic phenomena? Now, first, the general remark is, how high should we aim here? I mean, how general can we expect this to be? So how general? All right. That's really the fundamental question here. And in answering that question, led both scientists and mathematicians very far from the original area that they were investigating. Let me say – well, let me say right now, pretty general, all right. And we'll see exactly how general – I'll try to make that more precise as we develop a little bit more of the terminology that really – that will apply and allow us to get more careful statements. But we're really aiming quite high here, all right? And we're really hoping to apply these ideas in quite general circumstances. Now, not all phenomena are periodic. All right. And even in the case of periodic phenomena, it may not be a realistic assumption. I think that's important to realize here what the limits may be, or how far the limits can be pushed. So not all phenomena, naturally, although many are, and many interesting ones are periodic. And even periodic phenomena, in some sense, you're making an assumption there that is not really physically realizable. So even for periodic phenomena or at least functions that are periodic in time, even phenomena – soon I think I'll start talking in terms of signals rather than phenomena, but phenomena sounds a little grander at this point. Even phenomena that are periodic in time – real phenomena, they just die out eventually. We only observe – or at least we only observe something over a finite period of time, whereas, as mathematical functions, the sine and the cosine go on forever. All right. As a mathematical model, sine and cosine go on forever. So how can they really be used to model to something that dies out? But a periodic function, sine and cosine – all right – go on forever, repeating over and over again. All right. So in what sense can you really use sines and cosines to model periodic phenomena when a real periodic phenomena – when it really dies out? Well, that'll take us awhile to sort all that out. Let me just give you one answer to this, and one indication of how general these ideas really are. And you have a homework problem that asks you actually to address this mathematically. All right. So if you have a finite – you can still use – still apply ideas of periodicity, even if only as an approximation or even if only as an extra assumption. So what I mean by that is, as follows. So suppose a phenomena looks like this. Suppose the signal looks like – something like this – let me write it over here. So it dies out over a period of time. So this is time and there's only a finite interval of time – it might be very large, but there's only a finite interval of time when the signal is non-zero. I'm drawing just as a – somehow generic signal here. Well, that's not a periodic phenomena; it doesn't exhibit periodic behavior. But if it dies after a finite – outside of a finite interval, then you can just repeat the pattern and make it periodic. All right. Force this – you can force periodicity. That is, by repeating the pattern. You can force extra symmetry. You can force extra structure to the problem that's not there in the beginning. So, I mean by this a very simple idea. Here's the original signal, and I just repeat the pattern. So this is the original signal, these are maybe, sort of, you know, extra copies of it that I'm just inserting artificially, and extend the function to be periodic and exist for all time. You may only be interested in this part of it, but for mathematical analysis, if you make it periodic, that'll apply to the whole thing. All right. This is sometimes called the periodization of a signal. And it can be used – and it is used – to study signals which are non-periodic to – excuse me – to use methods of Fourier series, and pumpkins is another sort of Fourier analysis to study signals which are not periodic. So there's actually a homework problem on different sorts of periodization, and it's a technique that comes up in various applications. All right. So see the first problem set, homework one. Now, the point is that periodicity and the techniques for studying periodic phenomena are really pretty general. All right. Even if you don't have a periodic phenomenon, you can make it periodic, and perhaps you could apply the techniques to study the periodized version of it, and then restrict – study the special cases of actually where you're interested in the function. All right. So it's pretty – that's the point of this remark, is that the phenomena – that you're not restricting yourself so much by insisting that you're gonna use sines and cosines, or you're gonna study periodic phenomena. So the study that we're gonna make here can be pretty general – it can apply really quite generally. Okay. Now, let's do it. Or let's get launched into the program. So, first of all, let's fix a period in the discussion. All right. So we just – just to specialize and fix ideas, let's take – let's consider periodic phenomena of a given fixed period and see what we can say about those; see how we can model those mathematically. So for the discussion, let's fix the period for the discussion. All right. And there's a choice here. A natural choice would be too high because the sine and the cosine are naturally periodic of period two pi, but I think for many formulas and for – and for a variety of reasons, it is more convenient to fix the period to be one. All right. So we're gonna look – we're gonna consider function signals, which are periodic of period one. So we'll use period one. That is, we consider signals – I'll write things as a function of time, but again, it's not only periodicity and time that I'm considering. All that I'm gonna say can apply to any sort of periodic phenomena, so we'll consider functions f of t satisfying – f of t plus one is equal to f of t for all t. All right. And as the building blocks as the basic model functions, we scale the sine and the cosine, that is, we don't just look at sine of t and cosine of t, we look at sine of two pi t and cosine of two pi t. So the model signals are sine of two pi t, that has period one, and cosine of two pi t, that has period one. All right. Simple enough. Now, one very important thing I wanna comment before – before we launch into particulars, is that periodicity is a strong assumption, and the analysis of periodicity has a lot of consequences. If you know – if you have a periodic function – if you know it on an interval of length one, and any interval of length one, you know it everywhere because the pattern repeats. If you just know a piece of the function, you know it everywhere. So if we know – if we know, and I say, if we know – if we analyze, if we – whatever formulas we derive, and so on, this is an important maxim. So I'll put it in quotes, "know." To be a sort of generic, infinite, perfect, God-like knowledge. So if we know a periodic function, say period one, on an interval, and not just a particular interval, but any interval of length one, then we know it everywhere. All right. These are all simple remarks. Okay. These are remarks that you all have seen before in various contexts, but again, I wanna lay them out because I want you to have them in your head. And I want you to be able to pull out the appropriate remark at appropriate time. And you'd be amazed how far simple remarks can lead in the analysis of really quite complicated phenomena. Now, how are we gonna – how are we gonna take such simple functions of sine and cosine individually, and model very complicated periodic phenomena? That is the sine and the cosine as endlessly fascinating as they may be, are just the sine and the cosine. But the fact is, they can be modified and combined to yield quite general results. We can modify and combine sine of two pi t, cosine of two pi t, to model very general periodic phenomena of period one. Okay. To model general periodic signals of, again, period one. All right. Now, here is the first big idea, or here's a way of phrasing the first big idea. When I talk about modifying and combining – well, let's first talk about modifying. And the maxim or the aphorism that goes on – goes with what I have in mind is, one period, many frequencies. As far as a big idea – one period, many frequencies. I think you can actually find this in The Dead Sea Scrolls. Okay. What do I mean, one period, many frequencies? Well, let me just take a simple example. I mean, for example, e.g., you have sine of two pi t, and we know what the graph of that looks like. I'll put the graphs over here. It repeats once and it has – it's a period one; it also has a frequency one that is – completes one cycle in one second. So if I think of this as the time axis, say, although, again, I'm thinking in terms of time, it's a very general mathematical statement. It repeats exactly once on the interval from zero to one. All right. If I double the frequency and look at sine four pi t – all right – then that completes two cycles. So this is period one, frequency. Sine of four pi t is period one-half, frequency one – frequency two, but period one-half also means period one. All right. And the picture looks like this, it goes up and down twice – [inaudible] – it does, in one second. All right. Zero, one, it repeats. One cycle – it goes through one cycle in a half a second, it goes through two cycles in one second, so that's frequency two. All right. But it also has period one because if you consider this as the basic pattern that repeats, and it also repeats on an interval of length one. Okay. It's true that it repeats on an interval length one-half, but the signals already contained in interval one-half, but the whole – but the signal also has a longer period. All right. It has a shorter period, but it also has a longer period. And let me do one more. If, for instance, I look at sine of six pi t, very simple. All my remarks are very simple. This is period one-third, this has frequency three, but it also has period one. You might think of this as, I don't know, the secondary period, or somehow – I don't know what exactly the best way of saying it because, really, the best description is in terms of frequency not period. And what does the picture look like? Well, this time it has three cycles per second – frequency three – so that means it goes up and down three times in one second. Let's see if I can possibly do this. One, two, three. Good enough. One, zero, and one cycle is in one – goes up and down completely once in one-third of a second, then the next third of a second it goes up and down again, the next third of a second it goes up and down a third time. But it also has period one because if you consider this has the pattern that repeats, that pattern repeats on an interval of length one. Although, in some sense, the true repetition – the true period is shorter than that. Now, what about combining them? So that's how you can modify – and you can do the same thing with cosine. That's how you modify the function – one period, many frequencies. If we combine them together, I actually have a picture of it here, but I think I'll just try to sketch it – fool that I am. What about the combination? And when I say combination, I am thinking of a simple sum. That is sine of two pi t plus sine of six pi t – excuse me – sine of four pi t plus sine of six pi t. All right. What does the graph of that look like? Well, it looks like so. And I'm gonna sketch this, and then I'm gonna – I'll make – I wanna make another comment about this in just a second, but I actually had Mathematica plot this for me. It looks something like this, it goes – this is plotted on an interval of length two. All right. I've plotted it and it goes – it's kinda nice, it goes up and then a little bit like this, and then it goes up and then down a little bit farther, and up a little bit like that, and it goes down – that's the sound it makes – up, down like that, and then up – excuse me – and that's not two, it will be two, and down and up, and then there we go. Two. And then it repeats. Here's one. All right. That's the sum. Now, what is the period of a sum? The period of the sum is one. All right. Because although the things of – the terms of higher frequency are repeating more rapidly, the sum can't go back to where it started until the slowest one gets caught up – goes to where – back to where it started. All right. The period of the sum is one. One period, many frequencies. There are three frequencies in the sum. One, two, and three. But added together, there's only one period. So in a complicated – this is a very important point, and again, I'm sure it's a point that you've seen before. That's why I say one period, many frequency, and that's why for complicated periodic phenomena, it's really better, more revealing, to talk in terms of the frequencies that might go into it, rather than the period. You are fixing the period. You're fixing the period to have length one. All right. But you want – but you might have a very complicated phenomena. That complicated phenomena, as it turns out, is gonna be built up out of sines and cosines of varying frequency. AS long as the sum has period one, then we're okay. One period, many frequencies. That's the aphorism that goes on with this –

that goes along with this. Now, in fact, what is it – what we can do more than just modify the frequency. We can also modify the amplitudes separately, and we can modify the phases of each one of those – each one of those terms. So to model complicated, perhaps, how complicated? We'll see. A complicated signal of period one we can sum – we can modify the amplitude, the frequency, and the phases of either sines or cosines, but let me just stick with the sines – of sine of two pi t, and add up the results. That is, we can consider something like this. Something of a form, say, k going from one up to n, and we can consider how ever many of them we want. A sub k, sine of two pi k times t plus – that's modifying the frequency – plus fe sub k – allowing myself to modify the phase. Fe sub k. That's about the most general kind of sum that we can form out of just the sines or – and I can do the same thing with cosines, or I can combine the two and I'll say more about that in just a second. All right. So again, many frequencies, one period. All right. the lowest – the longest period in the sum is when k is equal to one, period one. The higher terms, they're called harmonics, because of the connection with music, and it's discussed in the notes, because they model – because simple sines and cosine model musical phenomena as a repeating pattern – musical notes. The higher harmonics, the higher terms, have higher frequencies, have shorter periods, but the sum has period one because the whole pattern can't repeat until the longest period repeats. All right. Until the longest pattern is completed. Now, I'm actually gonna post on the website – I'm going to give you a little MATLAB program that allows you to experiment with just these sorts of sums. All right. That is, you choose n – it forms sums exactly like this – you can choose the a's, the amplitudes, you can choose the phases, and it will plot what a sum looks like. All right. It's called – so I have MATLAB program with a graphical user interface. I wrote this myself, actually, a couple years ago. I was very – it was the first I ever did in MATLAB, it was pretty clunky, let me tell ya. But last year, a student in the class, in 261, took it upon herself to modify it, which caused me to bump up her grade in the end, and it's really quite a nice little program. It's not complicated, but it'll show you how complicated these sums can be. All right. So there's a MATLAB program, which I'll post on the website, sine sum – actually it's called sine sum two because sine sum one was my own version, which is now on the ash heap of history. All right. To plot these sums – and it's really quite – I mean it's – talk about fun, you know – I mean to see how complicated a pattern you can build up out of relatively simple building blocks like this, it's really pretty good. So we may even – we're actually trying to see if we can do a homework assignment based on this – based on the program. There's a feature in the program that allows you too actually to play the sound that's associated with this. That is, if you consider these things as modeling – if you consider the simple sinusoids as modeling a pure musical note, then a combination models a combination of musical notes, and if you put a little button, it plays sound, you'll get something that may sound good or may not sound good. But it's interesting to try. Unfortunately, the play sound feature doesn't seem to work on the Mac, it only seems to work on Windows. It requires Windows Media Player or something like that; Bill Gates' version of death 4.2 or something, I don't know. Anyway, so we have to see if we can fix it to work on the Mac, but – everything works on the Mac fine except for playing the note. So I'll put that up on the web in a zip file, and you should fool around with it a little bit. Okay? Because it'll give you a good sense of just how complicated these things can be. All right. Now, so how complicated can they be? That's the question that I really wanna address. I mean, this sum is already more complicated than just the sine and the cosine alone, but it doesn't begin to exhaust the possibilities that we want to be able to deal with. Let me say, to advance the discussion, actually, and to really get to the point where I can ask the question in a reasonable way, how general a periodic phenomena can we expect to model with sums like this. Let me say a little bit about the different forms that you can write the sum in because algebraically – for algebraic reasons, primarily algebraic reasons, there are more or less convenient ways to write this sum. Okay. And I think it's worth commenting about it, just a little bit. So different ways of writing the sum –this sort of sum – k equals one to n of a sub – I'm gonna use a capital in here – [inaudible] – a sub k sine of two pi k t plus fe sub k. All right. Now, you can lose the phase, so to speak, and bring in – write it in terms of sines and cosines, if you use the addition formula for the sine function. That is the – the formula for the sine of the sum of two angles. So if you write sine of two pi k t plus fe k as the sine of two pi k t times the cosine of fe k plus the cosine of two pi k t times the sine of fe k, just using the addition formula, than that sum can be written in terms of sines and cosines. You can write the sum in the form – let me use – for different coefficients, say sum from k equals one to n, little a k times the cosine of two pi k t, plus little b sub k times the sine of two pi k t. All right. And you know where the – how the little ak's and the bk's are related to the capital A's in the phase just by working it out. All right. So capital A sub k times this thing, and then there's a term coming from the phase – you haven't lost information about the phase in some sense, it's still there, but it's represented differently in terms of the coefficients out in the form of the sum. All right. This is a very common way of writing these sorts of trigonometric sums. As a matter of fact, I'd say it's more common in – if you look in the applications, even if you look in the textbooks, it's more common to write the sum this form than it is to write it in this form. But there equivalent, all right? You can go back and for the between the two. And you can also allow for a constant term, you can shift the whole thing up. And that's also usually done for purposes of generality. All right. You can add a constant term. And it's usually written in the form a zero over two, the reason why a zero over two is in there is because of yet, another form of writing it. Let me just write out the rest of it: a zero over two plus, again, sum from n equals – for k equals one to n to n of a sub k, cosine of two pi k t plus b sub k, sine of two pi k t. All right. And electrical engineers always call this the dc component. I hate that. All right. But they always do. All right. Who all learned to call this the dc component? Yeah, I hate that. Not everybody, I'm glad to see that. That's because you think of a periodic phenomena, you think about alternative – I don't know what you think about. You think about alternating current or voltage somehow is a periodic part, but then there's a direct part that doesn't alternate. There's a dc part – direct current part – that doesn't alternate, and that's that term. The reason why I don't like calling this the dc component is because what if a problem had absolutely nothing to do with current, you know? You're sometimes trapped by your language, and the field again, as I will say over and over again, is so broad and so diverse that you don't want to trap yourself into thinking about it in only one way. You know. You may be modeling some very complicated phenomena, and you say, 'What is a dc component?' Is somebody gonna look at you, like, 'What are you talking about?' You know? All right. Now, so that's a very common way of writing the form of the sum, but by far, the most convenient way algebraically, and really in many ways, conceptually, is to use complex exponentials to write the sum, not the real sines and cosines. By far. And this is pretty much the last time I'm gonna use sines and cosines. Or pretty much the last time I'm gonna write the expression like this, so it's by far better. And I'll have to convince you of this. All right. Primarily, algebraically, but also conceptually, is by far better to use to represent sine and cosine via complex exponentials. And write the sum that way. All right. so what do I mean by that? Let me just remind you, of course, in either the two pi i n t, [inaudible] formula is cosine of two pi n t plus or k – or I guess I'm calling it k, let me stick with the terminology there – two pi k t plus sine of two pi k plus i times the sine of two pi k t. Oh, yes, that's something else. All right. I'm gonna announce a declaration of principle. I is the square root of minus one, in this class. Not j. Deal with it. Get over it. All right. For me, for this class, it's i. You can use j if you want. I will use i. Now, because of Oiler's formula – Oiler's famous formula – you can express sines and cosines in terms of the completive exponential ant it's conjugate, that is, very simple formula, the cosine is the real part and the sine is the imaginary part. So cosine of two pi k t is then e to the pi i is the real part. Two pi i k t plus e to the minus two pi i k t over two. And the sine of two pi k t, likewise, is the imaginary part that's e to the two pi i k t minus e to the minus two pi i k t divided by two i. There is an appendix in the notes on the algebra of complex numbers, so if you're at all rusty on that, you should review that. All right. Because you're gonna want to be able to manipulate complex numbers, and I'm thinking primarily here in terms of working with complex conjugates with real parts and imaginary parts. You're gonna wanna be able to do that with confidence and gusto. All right. So if you're at all rusty in manipulating complex number – and complex exponentials, look over the chapter. All right? Matter of fact, on the first problem set, there are several problems that you give you practice in exactly this and manipulating complex numbers. All right. Complex exponentials. Now, because of this – and I won't – I won't write it out in detail, you can obviously then convert a sum which is not gone, which looks in terms of sines and cosines – in terms of complex exponentials. So you can convert a trigonometric sum as before to the form sum – and I'll write it like this – sum from k equals minus n to n, so it includes the zeros term, the constant term, c sub k – e to the two pi i k t. All right. We are now – the C sub k's are complex numbers. Everything in sight is complex. So, All right. The c k's are complex. Now, they can be – all right – they can be expressed – I won't do this and – I was gonna give you a homework problem on this, but I decided not to. You can do this just for fun. You can see how the different coefficients are related. So start with the expression in terms of sines and cosines, make the substitution in terms of the complex exponential and see what happens to the coefficients. All right. You will find, actually, a very important symmetry property. All right. These are complex numbers, but they're not just arbitrary complex numbers, they satisfy symmetry property. And it's because of the symmetry that the total sum is real. All right. The symmetry property – this comes up a lot – and we'll see similar sort of things reflected actually, when we talked about Fourier transforms. That is c sub – the sum goes from minus n up to n – they satisfy the property, the c sub minus k is c sub k complex conjugate, c k bar. All right. That's a very important identity that's satisfied by the coefficients for a real signal like that, and it comes up often. All right. it's one of the things you have to keep in mind. All right. That's a consequence of actually making the conversion. That is, starting with a formula in terms of sines and cosines, and then getting the formula in terms of the complex numbers. All right. Conversely, conversely, if you start with the sum of this form, all right, where the coefficient satisfies the symmetry property, then the total some will be real. That's because you can group a positive term and a negative term, and because of this relationship here, you'll be adding a complex number plus its conjugate, so you'll get a real – results out of that. All right. So if the coefficients satisfy this, then the signal is real. And conversely, if the signal is real and you write it like that, then the coefficients have to satisfy the symmetry property. Yeah?

Student: What is that – [inaudible] –

Instructor (Brad Osgood):What is that line? That line is – indicates complex conjugate. All right. So for a general – all right, do I have to say anymore first? So for general complex number a plus b i, the conjugate is a minus b i. Okay. There are different notations for complex conjugates, sometimes some people use different – some people use an asterisk, a star, some people even use a dagger. All right. But I think by far – it's true, I'm not making that up. But this is the most common notation. All right. And that is another notation I will use.

Okay. Now, now, now, now. We are ready, at last, to at least ask the question that's really at the heart of all of this. How general can this be? How general can this be? I mean, I'm in the form now, algebraically – well, I'm in the form now where I can ask the question, and as we'll see algebraically, writing sums of this form is by far the easiest way to approach it.

So we can now ask the fundamental question. Why is there something rather than nothing? Let's kick that one around for awhile. The fundamental question – so again, f of t is a periodic function of period one. All right. Can we write it, f of t and that sort of sum, k going from minus n to n of c sub k e to the two pi i k t. So again, I'm assuming the signal is real here so the coefficient satisfies symmetry relation; just keep your eye on the ball here.

The fundamental question is this. You have a general periodic function, can you write it as a trigonometric sum? Can you express it in terms of sines and cosines? Can you express it in terms of the fundamental building blocks? All right. By the way, linearity is playing a role here, although again, I haven't said it explicitly until now, we're considering linear combinations of the basic building blocks. We're considering a linear way of combining the basic trigonometric functions, the basic periodic functions. All right. A linear way of doing that.

So that's the fundamental question. And the answer – I'll tell you next time. But you don't think I'd make such a big deal out of it if the answer was no. So – but there's a lot to do. And answering this question – answering this question led to a lot of very profound and far-reaching investigations.

All right. Now – but I want to get started on it. All right. Now, let me give you a little clue. Yeah –

Student:[Inaudible] earlier you were starting with one.

Instructor (Brad Osgood):Pardon? Why am I not starting with one? Well, for one thing – so the question is why does the sum go from minus n to n, and why doesn't it just go from one to n?

Well, for one thing, if it went from one to n, the signal wouldn't be real, right? Remember there's this combination of the positive terms and the negative terms, all right? And the positive terms and the negative terms – because of the symmetry relation of the coefficient, the positive and the negative terms combine to give you a real signal – to give you a real part. And it's a fact that if you start with a real signal in terms of sines and cosines, and then you use complex exponentials to express it this way, you will find that it's the symmetric sum. It goes from minus n to n. Okay.

By the way, I should have said something over here, I suppose. Note one thing, by the way, that c sub-zero is equal to c sub-minus zero. Zero being what it is. C sub minus zero is equal to c sub-zero is equal to c zero bar. What does it mean to say that a complex number is equal to its conjugate?

Student: It is real.

Instructor (Brad Osgood):It's real. All right. So the one coefficient that you know for sure is real, others may be real, it may just work out that way, but the one coefficient that you know is real for sure, is the zero of coefficient. All right. so it must be real. So c zero is real. That's just a little aside, c zero is real.

All right. You have to be a little – you have to cut me a little slack here. Like I said, we all have to cut each other a little slack. There's so many little bits, you know, to observe – little pieces, little comments and things like that. I can't make all of them, all right? I hope I put all of them, or most of them, in the lecture notes, all right, in the notes so you see these things. But, as I say, there's so many things along the way that you could point out, that you can note, that we just can't do it all because I want to keep my eye on the bigger picture. All right. But this is one thing that comes up often enough. So there'll be, you know, there'll be instances where you have to read the – read the notes carefully and try to make note of all those things. And it's hard. It's hard. You know, it's hard to know when you're gonna need this little fact or that little fact because there's so many little facts.

But you'll see when the whole thing – when you – if you keep the big picture in mind, in many cases the details will take care of themselves. Really. Now, where was I?

Yes? A secret of the universe. All right. Here's a pretty big secret of the universe, actually, coming your way. When you try to apply – when you're trying to see how mathematics works, and when you try to apply mathematics to various problems, you often have a question like this. What if – how can – what – how can something happen? All right. Is it possible to write something like this? All right. Now, a very good first approach – and I'm serious about this. When you're doing your own work and you're trying to look at a mathematical model of something, you say, can I do something like this? Often the first step is to suppose that you can, and see what the consequences are. All right. Then later on, you can say, all right, then maybe I should try this because that seems to be what has to happen. All right. And then you go backwards. All right. And mathematicians will never tell you this because they like to sort of cover their tracks. They say, 'Well, it obviously goes like this,' you know, and, 'We're obviously going to define this formula and that formula, and life is going to work out so simply.' But what they don't show you is often that first step of saying, suppose the problem is solved, what has to happen? All right. So suppose you can do this. We can write f of t equals the sum from k equals minus n to n, c sub-k e to the two pi i k t. What has to happen? All right. Now, by that I mean, if you can write this, what are the coefficients? If you can write an equation like this, then what I'm asking here is, what are the mystery coefficients in terms of f? Coefficient c sub-k in terms of f – f is given to you. All right. So the unknowns in this expression are the coefficients. And the question is, can you solve for them? All right. Suppose you can write it like that. Can you solve for the coefficients? Can we solve for the c k? All right. I'm gonna take a very naïve approach. All right. I'm gonna isolate it. What do you? It's like an algebraic equation. To start with an algebraic equation, isolate the unknown. So isolate, like, the, I don't know, Mth coefficient or something like that. All right. So isolate c m out of this. That is – it's a big old sum, right? So f of t is, you know, all these terms plus c sub m e to the two pi i m t plus all the rest of the terms. That is to say, I can write c sub-m e to the two pi i m t, is f of t minus all the terms that don't involve m, so let me write it like this. Say, sum over k different from m of c sub-k e to the two pi i k t. All right. I haven't done anything except algebraically manipulated the equation to bring the one mystery term, or one fixed term on the other side. All right. All I did here – so f of t is this big sum. One of those terms in the sum is c sub m e to the two pi i m t – I wanna solve for the unknowns. All right. So solve for the unknowns one unknown at a time. So this is the Mth term in the sum, bring that over to the other side of the equation, write c sub m e to the two pi i m t is f of t minus all the terms that don't involve m. Okay. And then, write that as – that's almost isolating c sub-m, and not quite because it's got a complex exponential in front of it. So multiply both sides – this board is not so great – multiply both sides by e to the minus two pi i m t. So c sub-m is e to the minus two pi i m t times f of t minus the sum over all k different from m of c sub-k e to the minus two pi i m t times e to the two pi i k t. You with me? Nothing on my sleeve. All right. All right. Now, that's brilliant. I have isolated one unknown in terms of all the other unknowns. All right. So, I don't know if one can say that we have really made progress here. So we need another idea. This is as far as algebra can take you. All right. Algebra says, you wanna solve for the unknown, fine. Isolate, you know, what did your eighth grade teacher tell you? Put the one unknown on one side of the equation, put everything else on the other side of the equation. Hope and pray. All right. So we put the one unknown on one side of the equation, everything else is on the other side of the equation. Hope and pray. Now, the desperate mathematician at this point looking for something to do. Let me actually take this out one more algebraic step. Let me just combine those two exponentials there, and write this as c sub-m is e to the minus two pi i m t minus the sum over all terms different from m of c sub-k e to the two pi i – two pi i k minus m times t. I'm just combining the two complex exponentials there. All right. Great.

Student:[Inaudible]

Instructor (Brad Osgood):What?

Student:F of t.

Instructor (Brad Osgood):F of t. Picky, picky. All right. F of t minus. All right. Now, good. So now, they say that we need another idea. And the desperate mathematician at this moment will think of one or two things – one of two things. [Inaudible] they differentiate or integrate. I mean, what's beyond algebra? Calculus. What's in calculus? Derivatives and intervals. All right. So, here's a clue. Derivatives won't work, but intervals will.

All right. We need another idea. And that's a good idea, it's an inspired idea. But [inaudible] because it works, and I'll show you way. So I'm gonna integrate both sides from zero to one over one period. All right. All I have to worry about here is one period. Everything's periodic at period one, so I integrate over – I integrate from zero to one. What if I do – what do I get?

Well, certainly, if I integrate zero to one of c sub-m d t, that just gives me c sub-m. All right. So what about the rest of it? So I get – I get c sub-m is equal to the interval from zero to one, e to the minus two pi i m t, f of t d t, minus the interval of the sum is the sum of the interval – let me write this, sum from all the different – all the terms k different m – and the constant comes out – c sub-k, the interval from zero to one, e to the two pi i k minus m – k minus m – t d t. That's a t there. T. Ouch.

All right. I've integrated both sides from zero to one. All right. Now, watch this. All right. Watch. I can integrate that complex exponential, that's a simple function. I can integrate that just like I integrated in calculus. All right. The interval from zero to one, e to the two pi i k minus m t d t. Now, k is different from m. All right. If k is equal to m, I'm just e to the zero here, I just get one [inaudible] so the k is different from m. So the interval of this is one over two pi i – trust me – integrating this is the same as integrating an ordinary function – two pi i, as you did in calculus.

One over two pi i k minus m e to the two pi i k minus m t, evaluated from t going from zero to one. Straightforward integration. Straightforward integration, which is equal to – we are almost done. We are almost there.

One over two pi i k minus m, that makes sense, right? Because k is different from m, so it's not a problem. E to the two pi i k minus m times one – so that's e to the pi i k minus m minus e to the zero. All right. But either the two pi i k minus m, that's e to the two pi i times an integer. That's like sine of two pi times an integer, cosine of two pi times an integer – that's one. And e to the zero is also better known as one. So this is better known as one minus one, which is better known as zero.

Nothing. All this crap integrates zero. Excuse me. All right. Incredible. What is the upshot? It all goes away. What is the upshot? The upshot is, that c sub-m – what's left? What's left is c sub-m is the interval from zero to one – that – all that – all the terms of the sum here are gone – are gone. They integrated out to zero. What remains is the interval from zero to one of e to the minus two pi i m t f of t d t.

All right. Now, in principle, this is known because you start out by assuming you knew f. Suppose I know f, what has to happen? All right. Well, suppose I'm given f, and I write f as this sum, what has to happen? Here is the answer. All right. Here is the answer. Let me summarize.

We have solved for the unknowns. All right. So given f of t – periodic of period one – suppose we can write f of t as the sum, k equals minus n to n of c k e to the two pi i k t. What has to happen? What has to happen is the coefficients have to be given by this formula. Then you must have – I'll just write c sub-k instead of c sub-m – [inaudible] of the k of coefficient is the interval from zero to one of e to the minus two pi i k t f of t d t. That's what has to happen. All right.

It's an important first step in applying mathematics to any given problem, whether it's a mathematical problem or a non-mathematical problem. Suppose the problem is solved, what has to happen? If the problem is solved, mean, suppose you have this representation then the coefficients have to be given by this formula.

All right. So next time, I'm gonna turn this around saying, suppose we give these coefficients by the formula, do we have something like that, and in what sense? And that will lead us to great things.

All right. So more on that next time.

[End of Audio]

Duration: 55 minutes

TheFourierTransformAndItsApplications-Lecture03

Instructor (Brad Osgood):I love show biz you know. Good thing. Okay. All right, anything on anybody's mind out there? Any questions about anything? Are we all enjoying our first problem set to class? I guess, Thomas posted some typos. I'll correct those. I just had a chance to look at it this morning. There was some evidently minor typo's in the problem set. So I'll look it over and repost the version of it. I don't think there's anything there that would confuse anybody. All right? Okay. Let me remind you where we finished up last time. We took an important first step in understanding the analysis and undertaking the analysis of periodic phenomena and trying to represent a general periodic phenomena by the sum of much simpler periodic phenomena under this is complex exponential? So think in terms of sines and cosines, all right?

So last time, I say we took the first step in analyzing general periodic phenomena via the sum, so several combination, a linear combination of simple building blocks, simple periodic phenomena. So let me remind you what we did because it's very important that you realize what we did and what we didn't do. We said suppose that you can write a periodic signal in a certain form what has to happen. So we start off by saying F of T is a given periodic function, periodic signal. Function, signal same thing. And just to be definite we took it to have period one, all right? And the question is can it be represented in terms of others and suppose it can be represented in terms of other simple signals of period one, namely the complex exponential.

So suppose we can write F of T as a sum, say something like this. Okay? K from minus N to N, cease of K, E to the two pi I, KT. There was a question, by the way, somebody sent me an e-mail about why the sum is symmetric and so why does it go from minus N to N and we talked a little bit about this last time. This is also discussed a little bit more in the notes. You can think in terms of sines and cosines, all right? And the idea is that if you have a real signal the coefficient [inaudible], but it bears repeating. The coefficient satisfying an important symmetry relationship. The complex numbers they cease of minus K is equal to cease of K bar and because of that the positive frequencies combines with the negative frequencies. The positive terms combine with the negative terms to give you a real part. So will give you essentially a sum of cosines. Okay?

Or sum of sines and cosines because the values are complex. And it's a symmetric sum. Instead of going from one to N or zero to N, if you use complex exponentials it goes from minus N to N. The proof of the helpfulness of this representation will just become apparent as we use it, all right? As I said before, the algebraic work in the analysis is just made incomparably easier by using complex exponentials than real sines and cosine'. Just the calculations become that much easier. All right.

So, once again, suppose we can do this. Then what we found is the coefficients had to be given by a certain formula. So then the coefficients are given by C sub-K is the interval from zero to one, E to the minus two pi I, KT times F of T, DT, all right? It's an explicit formula for the coefficients. And in principle that's known, all right? If you know the function then you can carry out the integration in principle. And I want to remind you that this formula – there were two parts of deriving that formula. There was a sort of algebraic part, where we just try to isolate the K coefficient and that certain way, but then the analytic part invoked a little calculus where to solve for the coefficient I had to integrate. This depended on a very important relationship of the complex exponentials. This I'll write over here – oh, right here.

The interval from – I'll write it over here. The interval from zero to one, E to the, say, two pi I, NT, E to the two pi minus two pi I, MT, DT. So that's interval from zero to one if I combine the two ended up two pi I and minus MT, DT. That's either one if M is equal to N. So in that case I'm just integrating to the zero, I'm just integrating one. Where it's equal to zero if M is different from N, all right? Fundamental relationship. We'll see that making its triumphant return a little later on. All right.

Now, that's fine and that was a very important first step, but it was only the first step. The second step is to turn the question around, all right? The first step says suppose we can write the function in this form, then the coefficients have to be given by this formula. The second step is to turn that around and ask the following: When is that really possible? So I want to turn this around. So, again, you're given F of T periodic of period one, all right? I know what the answer has to be, all right? In a sense, I know what the coefficients have to be. They have to be given by this, so I'm gonna define them, all right?

Define – and I want to use a different notation. I want to introduce a different notation that's quite standard in this subject. Define F hat of K to be this interval. The interval from zero to one of E to the minus two pi I, KT, F of T, DT. So, again, in principle you can compute this, all right? You're given F. You can carry out the integration of F against this complex exponential. And it's often then by F hat K and this is called the K three A coefficient of F. What did we say? K four A coefficient, all right? So it depends on F and, in fact, it can viewed as a transform of F, although we won't use that terminology – well, I won't stress that terminology quite so much now because it will become much more useful when we talk about the Fourier transform. It's not the Fourier transform. This is the Fourier coefficient.

So it's a transform of F, but evaluated on the integers. For each K, I have a corresponding number F hat of K given by this interval. And the question is, do we have – can we write the function? Can we sit beside the function in terms of its Fourier coefficients? Can we write F of T is the sum K equals minus N to N, F hat of K, E to the two pi I, KT for sum N? Is there some sum of these things that will allow us to express the given periodic function in terms of these simple building blocks? That's the question. I say, I know what the answer has to be if we can do it. The answer has to be given – such an expression has to involve this. That's what I talked about last time. The question is does it really work?

If it does, if a statement like this is true you have to believe that it gives you incredible power because, again, a general periodic function can be decomposed in this way into very simple terms. Analyzing a very complex system of periodic inputs, periodic outputs might be possible to do by analyzing what the system does to the relatively simple inputs and outputs given by the complex exponentials, all right? So if so we can analyze complex systems, simple building blocks. Not a great sentence here, but you get what I mean. Simple components, simple building blocks. I'd like another try at that sentence, but I think I'd just rather leave it there, all right? Okay.

Now, this method is only gonna be really helpful if it's fairly general. And that's always been the question I've raised a couple times. How general really is this? How general can we expect this to be if it works? All right. Is it worth putting the time and effort into even addressing this question if it's only gonna work in a few specialized cases that somehow we can handle an ad hoc basis when they came up. Well, let me give you a little warning here, all right? Let me show you the kind of things, not to be negative about it, but let me show you how high the stakes are, all right? This is a very high stakes question. And let me do that by a couple of examples, all right?

Examples that are natural enough they can come up fairly easily in applications. Let's look at some examples of some of the kind of signals that could come up. For example, you could take a signal that looks like this, all right? I'll draw it over here. I'll just draw the graph. Something that looks like this. It could model a switch that's periodically on or off, zero or one, current is flowing or it's not flowing. So something like it's one for half a second and then zero for half a second, all right? That's the basic period and then it repeats. So one down to zero. One-half and then it repeats and it goes up, then it's on, then it's off, then it's on, then it's off, and so on and so on. Okay? So one, three halves, two, and so on and so on, all right? That's a periodic function and it repeats also for the negative numbers. That's a periodic function of period one, all right? So switch on for half a second, off for half a second and repeats. All right. Can we expect to write, let's call this F of T, all right? There's the graph, and I can write down the formula for it. Can we write F in the form? I can compute the coefficients. I won't do it, but, in fact, the coefficients are calculated in the notes, all right? You can carry out the integration. It's not a complicated interval to carry out. You're just integrating the function one over the interval from zero to a half. That's all of the function times a complex exponential. So you can compute easily F hat of K. F hat of K would be, in this case, the interval from zero to a half. Let's say, I won't carry it out. E to the minus two pi I, KT times one DT, all right? It's one on the interval from zero to a half and zero on the interval from one-half to one. You only compute it on the interval on one period, on the interval from zero to one-half and that's an easy interval to figure out. And we write F of T is sum from K equals N to N, F hat of K, E to the two pi I, KT. That's the question. And the answer is, no or at least not with a finite sum. But I hate to build up all this drama asking this fundamental question only to answer it no for a relatively simple example. So the answer is no, not for a finite sum. Why? Because the complex exponentials thinking they're just sines and cosines, all right? They're continuous. And the sum of a finite number of continuous functions is continuous. It can't possibly represent a discontinuous phenomenon. You can't represent a discontinuous phenomenon by a continuous phenomenon, all right? You learned that theorem back in calculus, you know. Some jerk of a math teacher probably said something like and the sum of two continuous functions, of course, is continuous. And you said, yeah, right. Okay, fine. What's the point? Well, here's the point right now for the first time. This limits what you can do. It limits what you might want to do. You can't represent a discontinuous function by the sum of a finite number of continuous functions. Bad luck, all right? Well, what if a function is continuous at least, all right? But maybe has a corner. So let's look at the second example. Let's say a triangle wave. Something that looks like this. Up, down, up, down, up, down, and so on. So on an interval from zero to one. So it's periodic of period one and, again, the power of these, along the negative axis also are for negative values of T. So, again, we can easily compute the Fourier coefficients. That's not a problem. They exist. It's not a problem to carry out the integration and, again, that's done. I think it's worked out in the notes explicitly and if not it's easy enough to do. I will not do it in public. Again, you can compute the Fourier coefficients. Has to be integrated in two pieces. Has to be integrated. The interval from zero to a half of the function T times the complex exponential plus the interval from one-half to one of the function of whatever this is on the corresponding interval times the other times the complex interval. But you can do it, all right? It requires a little bit of work, but it's not hard. You can do it.

Can we write, I'll call this F of T, again. My generic error function is always called F, F of T. Can we write, again, the corresponding sum? Sum K equals minus N to N, E to the F head of K, E to the two pi I KAT. And, again, the answer is no. Not for a finite sum. Why? Well, because it's your jerk of a calculus teacher probably told you if your two functions are differentiable then the sum of two functions is differentiable. And you said, fine, fine, whatever. And it didn't seem to matter at the time, but now it matters because this function is not differentiable. It has a corner, all right? But the complex exponentials are just sines and cosines. They are differentiable. Their sum is differentiable. You can't represent a non-differentiable function as a finite sum of differential functions. It won't work. You just can't do it. The right-hand side is differentiable. The left-hand side is not differentiable. Okay? Now, we could go on like this with more and more bad news, all right? That is, if there is a discontinuity in –so here's the discontinuity in the first derivative, all right? There's a corner here. We could draw examples that would look smoother, but we could draw examples where there's this continuity in the second derivative or maybe the first and second derivatives are fine, but there's discontinuity in the third derivative and so on and so on. If there is any lack of smoothness, if there is any corner, no matter how smooth that corner looks, if there is some discontinuity in some high derivative you're screwed, to use a technical term. All right?

Any discontinuity in any derivative includes writing F of T is this sum. Sum from K equals minus N to N minus N to N of F hat A, E to the two pi I, KT for a finite sum because these functions are infinitely differentiable. These functions are as smooth as they can be. They're sines and cosines. And you can't take the sum of those sines and cosines and put them together and get something that's not infinitely differentiable, all right? So this great idea – we might as well quit now or take the rest of the quarter off. Because it doesn't look very general at all. I mean, the kind of signals you come up against may have jumps, may have corners, may have discontinuities, whatever. All right. Maybe we want to make an approximation that they're as smooth as can be and then we can use this. And it's an argument, but it's getting away from what we really hope to accomplish here.

Let me, before I go any further, say that there's a maxim lurking here that's important, all right? That is, if we can't represent it as a finite sum then we have to turn to infinite sums or at least larger and larger finite sums. Sums of more and more terms. And the maxim that's lurking, and I'll say it now and then we'll go back to the general discussion, is that it takes high frequencies to make sharp corners, or any corners for that matter. Maxim is it takes high frequencies to make sharp corners or really any kind of, it sounds better if you say it like this, but really any kind of corner, all right? Any time there's some kind of discontinuity in some high derivative, that means that you're gonna have trouble representing that phenomenon as a finite sum. You're gonna have to take N larger and larger to try to represent it more and more accurately. It takes more and more terms, takes higher and higher frequencies to make that bend, all right? Even for a relatively high degree of smoothness. All right? Now, by the way, you may think, again, that his is an artificial maxim that is, real signals don't have sharp corners, but that's not true. I mean, all the time when you, and later when we talk about filtering, producing sharp corners as sharp as you can is actually an important part of signal processing. Sometimes you want to take a signal and cut off after a certain point. Either cut it off in time or cut it off in frequency. Some of you, I'm sure, have had some experience with this, all right?

Sometimes your signal starts out pretty smooth, but for other reasons you want to make it somehow less smooth. You want to cut things off and cutting things off can introduce high frequencies in trying to represent the signal that is something else that has to be dealt with, all right? So there are some very tricky and very practical questions that go along with this maxim. All right. Now, like I said, at this point you say well what's happened to this grand general program that you've been announcing? If it's not gonna work and if I can't consider finite sums to represent all but the most specialized phenomena, if there's any sum slight discontinuity in there at any level of smoothness, then what good is any of this? And the come back to that is we have to consider infinite sums, all right?

To represent the general periodic phenomena, periodic signals we have to consider infinite sums, all right? That's a mathematical point. As a practical matter, of course, you can't sum up an infinite series. You can only sum up an approximation, but if you want to have some confidence in what you're doing and if you want to know what errors you might have to analyze, the first thing you have to do is realize that you have to expand your purview from finite sums to infinite sums. So to represent a more general periodic phenomena we must consider infinite sums of the form, say, for minus infinity to infinity half hat of K, E to the two pi I, KT. It may and be that not all these coefficients are non-zero. Some of them may be zero and so on, but for sure if the function has any sort of discontinuity at any level of derivatives then these coefficients are gonna go out and you're gonna have non-zero coefficients as far as you go out, all right?

Any non-smooth phenomenon signal will generate infinitely many, not just a finitely many, but infinitely many Fourier coefficients, all right? The only way you could possibly have a finite Fourier series is if the function you start out with were infinitely smooth, all right? Now, that's a problem, all right? I just want to say the stakes are high here because if you're gonna deal with an infinite sum mathematically, and even for applications, you have to talk about issues of convergence, all right? How accurate is this gonna be? If I cut it off after a finite number of terms, how accurate is it gonna be? All right? If the series is converging and I cut it off after a finite number of terms maybe I have a certain amount of confidence that I'm getting a pretty good approximation to my function, all right?

But if the series is not converging and I still try to cut it off after a finite number of terms, what confidence do I have that I'm taking a reasonable approximation to the signal that I really want? So you have to deal with issues of convergence. You are forced to if you want this theory to apply at all generally. Okay? And, again, not just for mathematical reasons, but also for practical reasons. Now, that's hard, all right? There are a lot of very hard questions here and we're not gonna go into the mathematical analysis of all of them. I do want to give you a big picture. I want to give you some good news and some hard news, but ultimately pretty good news, about how this is dealt with because it has all been sorted out. But it took generations of mathematicians and scientists and engineers working on this to finally resolve all these issues.

Why is this so – I mean, talking about convergence of series is hard anyway. It's particularly hard in cases like this because the terms are oscillating, all right? The complex exponential, again, think in terms of sines and cosines, all right? If I split this up into its real and imaginary parts, sometimes the cosine is positive, sometimes the cosine is negative, sometimes the sine is positive, sometimes the sine is negative. So you have positive terms and negative terms and adding infinitely many of them up, all right? So convergence for infinite sines like this has to depend on some type of cancellations that are going. There has to be some sort of conspiracy that's making this series converge for a given value of T, all right?

You need a conspiracy of cancellations, how about that? To make such a series converge because of the oscillation, all right? And that's hard to study. That can be hard to study in a given case. I just want you to be aware of this that the stakes are high and the issues are real. Now, here's what I want to do. I want to talk about the situation. I want to give you a number of statements, theorems that cover what the story is here. And, again, we're not going to go through the proofs of these things. It's not so crucial for us the mathematical details. A lot of them are covered in the notes, not all of them, however, and I'll say a little bit more about that as well. But I do want you aware of where the hard parts are and what the answers are, all right?

So what I want to do is, I want to have a summary of the main results. That is the convergence when the function is continuous, which happens often enough that you want to know or, let me say, smooth. It has to be an infinite series if it's not infinitely smooth, all right? The convergence, or what passes for convergence, when you have certain discontinuities and here there is a nice and helpful statement about when you have a jump discontinuity, all right? These two cases are actually relatively straightforward. It's easy to remember and it's easy to have a certain amount of confidence in what the results are, again. Although, I won't prove it. I won't go through the proofs.

And, finally, the convergence issues in general and they involve some very really quite deep changes in the perspective that you have to adopt toward this circle of problems. Convergence in general, all right? You actually do have quite a broad general statement that covers pretty much all situations that come up reasonably in practice. But the notion of convergence is a little bit different and the mathematics involved in here it took a long time to sort out, all right? So this involves a fundamental change of perspective.

As said, I'm not gonna do the tails, but I at least want to say some of the words and I'll tell you why. Because it has become so pervasive that is, it's become so – the framework for studying convergence in general, even as it applies maybe to these simple situations, has become so standard that you will see this in all the literature, all right? You'll see in the engineering literature, the terms that I'm gonna use orthogonality, mean square convergence, L2, things like that. I'll say all those words and I'll tell you what they mean, but it's become absolutely the way of talking about these things. And if you look at modern treatments of signal processing, and I'm thinking, in particular, here of wave limit analysis, which has become very popular in recent years.

You'll hear about orthonormal basis for so and so and distinguish from the complex exponentials as orthonormal basis. You'll hear all the terminology that goes along with this point of view. So I think, at least, I want you to come away from this with some understanding and familiarity with the terminology that goes along with it, all right? That's as far as I want to go with it. But even that I hope is gonna be helpful for you. All right. So let me look at these cases, because these cases are pretty straightforward and they're good to know. Convergence when the signal is continuous. Yes, good news. It converges, all right? So that is if – so continuous case, all right?

So, again, you can form the Fourier series, all right? You have that the series sum K equals minus infinity to infinity, F hat of K, E to the two pi I, KT converges for each T to the value F of T point Y. Is that means that you plug in a value of T into this sum and add it all up, then you have a series of constants, and what will it add up to? It will add up to the value of the function F of T. So that's good, all right? If the function's continuous then you know the series is gonna converge and it's gonna give you the right value, all right? Good. Not so easy to prove, all right? It takes a little work to prove that and, again, that's sort of sketched out in the notes, all right? Not all the details, but a number of them. You can at least see the broad outlines. And you will find this discussed in various levels of abstraction in most mathematical books on Fourier analysis, Fourier series, all right?

But that's what to keep in mind there. So the continuous case is good. We call this point-wise convergence, again, because you plug in a value of T, a point in time, add up the series, which is then a sum of constants and you're guaranteed that the sum will converge to the function F of T. In the case the function is smooth actually, you get a little bit more. If the function is differentiable, the smooth case. So that's if you have various degrees of differentiability. If it has one derivative, if it has two derivatives, and then there's the question about are the derivative's continuous and so on, so I don't want to split hairs on this. And, again, there's a fairly precise statement that's given in the notes, but it says this. So if it's smooth and the particulars continuous, so you know the series converges.

So, again, the series converges, the Fourier series, K equals minus infinity to infinity F hat of K, E to the two pi I, KT converges to F of T. That's the same as in the continuous case, but there's actually more to it than this that, again, can be helpful sometimes if you're trying to estimate errors. You actually get what's called uniform convergence and what that means – the way you should think about this is for different values of T you can control the rate at which the series converges. Same rate for different values of T. So, again, without trying to make it – I can make this precise, but I don't want to because it requires too much notation. So you can i.e. think of it this way, you can control, or you can estimate, the rate of convergence, how fast the series is converging.

Add different values of T depending on the degree of smoothness. What this often means is you get estimates on the size of the coefficients, you get estimates on the difference between the function, and a finite approximation, a finite version of the series, depending on the smoothness, all right? The smooth of the function is the faster it converges. That's one way of looking at it. And, again, without giving a technical definition of it. Yeah?

Student:[Inaudible]

Instructor (Brad Osgood):Well, what the uniform convergence means roughly and, again, what I don't want to write I don't want to write a lot of epsilon's and N's and things like that, is if this is the function, all right? All right? It's periodics, or the pattern, repeats. Then what I mean by uniform convergence is that if you look at a finite approximation, so if you look at a finite version of the sum just going from minus N to N, then that will track. So this is F of T, all right? That's the original function. And the approximation does something like, you know, it tracks the function along the way. This is [inaudible] sum, all right? You can estimate uniformly over the interval how far the approximation is from the function, all right?

So instead of just saying at a particular point the series is converging, all right? So, again, if I pick a particular point here then the value of the approximation is approaching the value of the function. That's fine. That's what happens for the continuous case. For the smooth case, you can say more than that. You can say uniformly how close the approximation is to the function over the entire interval, over all from zero to one, all right? And you can give an estimate for that, all right? That's what the smoothness gives for you, all right? And, again, you could write that down precisely, but I'm not gonna do that.

There is actually a statement of this theorem that's given in the notes, if you're interested. And it's interesting. I mean, and, again, it can have some practical implications because you're never gonna work with an infinite series in practice. You're always gonna work with a finite approximation. So the question is can you estimate the error? If you're called upon, can you give some reasonable estimate for how far off you are? Not just at a point, but uniformly over the interval where you're interested in making the approximation. So it can come up. So you have uniform estimate of the closeness. Okay? All right.

So that's the continuous case and the smooth case. So that's good news, all right? That's good news. That's nice. If the function is continuous the series are converging. If the function is smooth the series is converging and it actually, sort of, stays uniformly close to the given function. Now, I want to jump back, actually, to the discontinuous case because there is one situation that comes up often enough in practice that it's useful to know about and that's when you have a jump discontinuity. And, again, I'm not gonna prove this, but I think you should be aware of it and there's actually gonna be some problems that use it, or a problem that uses it. So if you have a jump discontinuity, functions can fail to be continuous in a lot of different ways. The simplest way a function can fail to be continuous if it has a jump discontinuity like the first example that I showed.

So the example of a switch where it's on, then off, all right? Something like this, all right? Where the signal jumps down or up. Okay? So EG. And there the theorem is, that is if a T knot is a point of jump discontinuity, this is a really cool result actually. Then this converges, then the sum minus infinity to infinity – well, yeah. I'll put it like this. Minus two pi I, KT, sorry, sorry, sorry. F hat of K, the Fourier series, F hat of K, E to the two pi I, KT does converge at T knot. It converges at the jump discontinuity, but the function doesn't really have a value of T knot, because it jumps. But it converges, actually, to the average value, all right? It converges to the point in the middle of the jump, to the average of the jump i.e. to one-half – let me write it like this one-half F of T knot. This is usually the way it's written. F of T knot plus F of T knot minus, all right?

So what I mean by that is you're approaching T knot from the left, that's F of T knot minus, it has this value. You approach F of T knot from the right it has this value, that's F of T knot plus, all right? They're two different values of jumps and if you look at the average of the jump, right in the middle, that's what it converges to, all right? Kind of cool. And this wasn't proved until, I think, the 1900's, sometime in the early part of the 1900's. So this was way after Fourier had done his initial work and people were struggling with a lot of these problems. And this is useful enough that it comes up in applications. So, for example, for the saw tooth, or not the saw tooth, for the switch periodic signal here, if it jumps from zero to one then at a value of discontinuity converges to one-half, all right?

So even sometimes people define this function to be one on the interval from zero to one-half, leaving out one-half. One-half at the discontinuity and zero from here to here, right? You sometimes see that definition given. And the reason why people sometimes give that definition is because they want to anticipate this result. So if they want to use Fourier's series, they say that that, you know, it's, sort of, consistent with the definition of the function in consistent with the property of the convergence of the Fourier series. This is not so easy to prove, all right? None of these things are really easy to prove. It requires a lot of work, all right? It requires a lot of estimates and careful analysis, but it's nice. It's very satisfying in some sense that it tells you – at least you know what the situation is.

I'm not saying it's easy to establish, but at least you know what the facts are, all right? That's good. That's good. Okay?

So any questions on that? This should be, sort of, part of your vocabulary. Yeah?

Student:[Inaudible] continuity?

Instructor (Brad Osgood):Then that's not a jump discontinuity, all right? What I mean by jump discontinuity, I mean it jumps between two finite values. Okay? Yeah. I'm trying to avoid – I mean, I'm talking to you in a somewhat informal way. I'm trying to avoid giving very precise definitions to all these things because you, sort of, know it when you see it, all right? You can do that, of course. You can give very precise statements here about the uniformity of approximation, about how it converges, and so on, but that's not so crucial for us. It's just you should know in general what the big picture is. Yeah?

Student:[Inaudible] point that ever function with jumps, so what about points where there are no jumps?

Instructor (Brad Osgood):Sorry?

Student: What about point where there are no jumps?

Instructor (Brad Osgood):All right. What about points where there are no jumps? So, actually, perhaps, what I should have done is give a more careful statement of the continuous case. The more careful statement of the continuous case is the series converges at any point of continuity, all right? So if a function – if it doesn't jump, if it's continuous at that point then the series converges at that point. Okay? So that would actually be a more precise statement of it. Oh, I better not stop asking for questions here, all right? As I say, I'm trying to avoid the infinite regression of making the statements as precise as I can. I can do that, but you don't want to see that. All right.

And, again, don't underestimate the effort that it took to do this, all right? When Fourier first came out with his ideas and we'll see that his original application, actually, on Monday. He was very bold. He actually said any function, not just any periodic function, because he was thinking of extending a function to be periodic by just repeating it, you know. He was thinking of functions, which die ff and then repeating it. So he made statements like any function can be represented by such an infinite series. And people were scandalized by that, especially if they were French and the French are easily scandalized. Bonjour, any function, you are a fool, monsieur.

So it caused a great deal of consternation and it caused a lot of work to get done to try to sort these things out. So don't underestimate the effort that went into getting even this much. But now, this was still ultimately not satisfying because Fourier really set his sights very high. But any periodic phenomena, never mind smoothness, whatever, could be represented in some sense by a Fourier series, by this sort of infinite sum. And to sort that out and to get to the truth of that really required an entirely different perspective. So I wanted to say a little bit about that. I'll only get to a little bit today and I'll finish it up on Monday, and then we'll see some applications of this, all right? All right.

So general case. That is not talking about continuity, not talking about smoothness, and so on, all right? In here, what's really involved is you need, we learned after decades, centuries of bitter experience, you need a different notion of convergence of the infinite series. You learn not to talk about what you think would be the most natural thing, point-wise convergence, all right? You learned by hard lessons. You learned it because it didn't work. I mean, ultimately you couldn't get an answer that was very satisfying. In these cases we're fine, but somehow under natural and fairly general situations, you learned not to ask for convergence of a sum like this, F hat – let me just put general coefficients in there. Cease of K, E to the two pi I, KTM. Even for general series, all right?

You learn not to ask for convergence of that at particular points. At values of T, all right? You've moved away from plugging in values of T and looking at the series of constants and asking whether or not converged, all right? And that was a hard step to take. Rather, what was ultimately learned was you get a satisfying answer if you asked for convergence in the mean, convergence on average. I mean, the proof that it was a good idea are the results that you can get if you take this point of view and it was a hard won point of view. So you get a better picture, more satisfying picture, if you ask for convergence in the average sense or also sometimes called convergence in the mean, and I will write that down, all right?

In engineering terms it's also sometimes called convergence in energy. You do see this term for reasons, which I'll explain, probably not today, but next time. Okay? Remind me. All right. Now, what does that mean? Well, you need to make some assumption on the function, all right? It's not maybe completely general in a sense the function's arbitrary, but the assumption you make on the function is pretty minimal. So, again, I'm assuming the function's periodic. So really everything takes place on the interval zero to one. It's only the properties on the function on the interval from zero to one that matters for us because everything is just repeated after that, all right?

So, I suppose, again, F of T is periodic period one, periodic period one, and you also suppose that it has the following property. That the interval – it's square interval. Square DT is fine. If the interval of the square of F is finite on the interval and that's not too hard a thing to insist on. It's not too restrictive a thing to insist on. There are functions that don't satisfy that. If a function goes off to infinity at a certain rate on the interval from zero to one, if it's unbounded, then it may not satisfy this. But certainly the functions that come up most often in applications are gonna satisfy a condition like this and this actually is sometimes the hypothesis of finite energy for reasons which we'll understand a little bit more later. Take this interval, it's often taken to be the total energy of the function depending on what the function represents. And so you're assuming that somehow the signal has finite energy, which is a reasonable physical assumption, all right?

Finite energy. All right. Then as it turns out, then you conform – the Fourier coefficients do exist. That's something that actually has to be proved separately because it's a question of integrability, but it's true. You conform the coefficients F hat of K equals the integral from zero to one, E to the minus two pi I, KT as before, F of T, DT, all right? Again, so that's actually now a separate issue that has to be verified because you're not assuming the function is continuous or anything else, but it can be done. And the punch line is, and it's quite a punch line, that you still get convergence of the series, of the infinite Fourier series, but not – and this board is floating up here. Let me do it over here. But not by plugging in points, but rather in an average sense as follows.

So then I'll write it like this. The interval from zero to one. I want to look at how close the function is to an approximating series, so that means a finite version of the sum. So the sum, the interval of the sum from K equals minus N to N of F hat of K, E to the two pi I, KT minus F of T squared, DT, all right? So I look at the average of the square of the difference between a finite approximation of the sum, makes perfect sense to form that, and the function. And the statement is that this tends to zero as N tends to infinity, all right? The series converges of the function in the average sense. The series converges to the function in the mean, all right?

The idea of integrating a function to get an average value is probably not so unfamiliar to you – one second. The idea of integrating the square, or the difference of the squares, is also probably more or less more familiar to you and, as a matter of fact, this is exactly, again, for reasons I can't explain today, exactly related to the least squares approximation of a function by a combination of complex exponentials. Now, you had a question?

Student:[Inaudible] convergence in the mean square?

Instructor (Brad Osgood):Pardon me?

Student: Is this mean square convergence?

Instructor (Brad Osgood):Yeah. Mean square convergence, thank you. I say convergence to the mean, maybe I should even call it mean square convergence. If you're familiar with that term by all means use it, all right? So it's mean square convergence. All right. So the result is – that I'm afraid we'll have to quit for today. That the function – the series converges to the function, right? That is, it makes sense to write, you write F of T equals it's Fourier series, K equals minus infinity to infinity, F hat of K, E to the two pi I, KT, but you have to understand what this equal sign means, all right? In this context, all right?

In this context you have to be careful here what this equal sign means. It doesn't mean pluck value of T and watch the series converge to the value of the function. It does not mean that. It means that if you compute that interval for a finite sum and let the degree go to infinity here, then that interval will tend to zero. The mean square difference will tend to zero. That's what that equal sign means, all right? That the difference between this and a finite approximation tends to zero as the approximation gets better and better. The square, interval to square.

Now, that was a big change of view, all right? That was a big change in attitude to adopt that notion of equality, that notion of convergence, and so on. And it had profound far reaching consequences, all right? And, again, it took a long time to sort out. And we're out of time for today, so I can' tell you that. So on Monday I want to wrap this up. Not in all the mathematical details, only so far as to give you what the general picture is because you're gonna see it beyond this class. You're gonna see people use this terminology, use these ideas well beyond what we do in here, and it's really quite satisfying. It's a really quite thorough and satisfying coherent picture, all right? So more on that on Monday.

[End of Audio]

Duration: 52 minutes

TheFourierTransformAndItsApplications-Lecture04

Instructor (Brad Osgood):We're on the air. Okay. A few quick announcements. First of all, the second problem set is posted on the website. Actually, I posted it last evening. So for those of you that are very eager and check on the website all the time, it was there. And secondly, the TAs are beginning their office hours this week, today in fact; is that right?

Okay, so if you have any questions for them, they will be available to help. All right, anything on anybody's mind, any questions about anything?

Student: Um.

Instructor (Brad Osgood):Yeah.

Student:[Inaudible].

Instructor (Brad Osgood):Anybody else have any issues with the online lectures? I don't know, I haven't – I'm afraid to look at myself, so I don't know what they're like.

Student:I was [Inaudible].

Student: Nothing happens.

Instructor (Brad Osgood):Nothing happens when you click on it?

Student:[Inaudible].

Instructor (Brad Osgood):It's a little trick we like to play on people.

Student:[Inaudible] which browser you are using, so in the Mac [inaudible], should be [inaudible].

Instructor (Brad Osgood):So the question may be which browser you're using. I honestly don't know; I've never tried to do it before [inaudible].

Student: If you're using a Mac, you have to use Safari; it doesn't work on anything else.

Instructor (Brad Osgood):It doesn't work on anything else, except – use the Mac, the word from over there is you have to use Safari, which is the one that comes with it. And I don't know about other ones. Anybody else have issues with this? I can find out and I can post an announcement, I suppose.

But I haven't heard, actually I haven't tried it, so I don't know that the – how do they look, the lectures?

Student:[Inaudible].

Instructor (Brad Osgood):Great. Thank you; that was the right answer. Anything else? All right, so I'll check into it, but try that on the Mac, try Safari or try other browsers. Any problem with PCs?

Student: They work fine.

Instructor (Brad Osgood):PCs work fine; okay. Don't, don't – I don't want to see. All right, anything else? All right, so today I have two things in mind today. I want to wrap up our discussion of some of the theoretical aspects of Fourier series. We're skimming the surface on this a little bit, and it really, you know, kind of kills me because it's such wonderful material and it really is important in its own way.

But as I've said before and now you'll hear me say again, the subject is so rich and so diverse that sometimes you just have to, you can't go into any – if you went into any one topic, you could easily spend most of the quarter on it and it would be worthwhile, but that would mean we wouldn't do other things which are equally worthwhile.

And so it's always a constant trade-off. It's always a question of which choices to make. So again, there are more details in the notes than I've been able to do in class, and will be able to do in class, but I do want to say a few more things about it today. That's one thing.

And the second thing is I want to talk about an application to heat flow that's a very important application historically, certainly and it also points the way to other things that we will be talking about quite a bit as the course progresses. All right, so let me wrap up and again, some of the sort of the theoretical size of things.

And I'll remind what the issue is that we're studying, and so this is our Fourier series fine, all right? Last time we talked about the problem in trying to make sense out of infinite sums, infinite Fourier series, and the important thing to realize is that that's not by no means the exception, all right?

We want to make sense of infinite sums of complex exponentials sum from K equals Minus Infinity, Infinity, cK, either the 2 pi, KT. I'm thinking of these things as Fourier coefficients, but the problem is general. How do you make sense of such an infinite sum? And the tricky thing about it is that if you think in terms of sines and cosines, these functions are oscillating.

All right, everything here in sight is a complex number and complex functions, but think in terms of the real functions, sines and cosines where they' oscillating between positive and negative, so for this thing to converge, there's got to be some sort of conspiracy of cancellations that making it work.

Of course, the size of the coefficients is going to play a role as it always does when you study issues of convergents. But it's more than that because the function is bopping around from positive to negative, see, all right and that makes it trickier to do. That makes it trickier to study.

And again, realize that this is by no means the exception, and so in particular if F of T again is periodic, period 1, we want to write with some confidence that it's equal to its Fourier series.

We want to write with some confidence, at least we want to know what we're talking about, that F of T, say is equal to its Fourier series going from minus [inaudible] 2 p i KT, and again, it's really if you want to deal with any degree of generality, it's going to be the rule rather than the exception that you'll have an infinite sum because any small lack of smoothness in the function or in any of its derivatives is gonna force an infinite number of terms.

A finite number of terms, a finite of trigometric sum will be infinitely smooth. The function and all its derivatives will be infinitely differentiatable, so if there's any discontinuity in any derivative you can't have a finite sum.

So any lack of smoothness forces an infinite sum, again, so it's not because the method is trumpeted as being so general, you have to face the fact that you're dealing with an infinite number of terms here, all right?

Now, by the way, I don't mean to say that all the terms are necessarily non-zero, that all the coefficients are necessarily non-zero. That's not true. Some of the terms may be zero.

For example, when you have certain symmetries, the even coefficients may be zero or the odd coefficients may be zero and in special cases, or a finite number may be zero or a block of them may be zero. You don't know exactly what's gonna happen.

But all I'm saying is you can't resort to only a finite sum if there's any lack of smoothness in there. All right, so again, that's the issue. Yeah.

Student:[Inaudible].

Instructor (Brad Osgood):Does [inaudible] of K, that actually is a K although it looks like a T, it's [inaudible]. All right, the K Fourier coefficient. I'll remind you what the definition is since we're gonna use that. So [inaudible] of K is the integral as a K, integral from 0 to 1 of either the minus 2 pi KT, F of T DT. Is somebody's phone ringing?

All right, now, so last time we dealt with, at least in statement, the cases where that function was smooth or is smooth and you get all that nice sort of convergents that you want. All right, so F of T is continuous, smooth or even if you have a jump continuity, then you get the sort of kind of convergents that you want. You get satisfactory convergents.

I'll just leave it at that because the precise statement we talked about last time, and it's also in the notes, you get satisfactory convergents results. So that's fine, so again, that gives you a certain amount of confidence that you write the series down, you can manipulate it and plug into it and things like that and nothing terrible is gonna happen.

But to deal with a more general functions, the more general signals that arise, it really requires a different point of view, all right, greater generality requires a different point of view, and that's where we finished up last time.

Now, it's not just for fun, even for mathematical fun. That is, this point of view turns out to have far-reaching consequences and really does frame a lot of the understanding and discussion, not only for Fourier series, but other subjects that are very similar and also in sort of everyday use in a lot of fields of signal processing.

So very generality requires a different point of view, different terminology, different language and a whole sort of re-orientation. All right, and again, I set that up last time, and I'm gonna remind you where we finished up and I wanna put in one more important aspect of it today and that's all we're gonna do, sad to say, all right?

So again, the condition is the integrability of all things. Instead of smoothness, instead of differentiability, the condition that turns out to be important is integrability to function. All right, the important condition, and a relatively easy one to verify, integrability, all right?

So you say that a function, say that F of T is squared integrable or briefly, you say that F is in L2 of the integral from 0 to 1. I'm only working on an integral from 0 to 1 here. And L2 stands for square, well the 2 stands for square and the L stands for Lebeg, and I'll say a little bit more about that in a second if the integral is finite, integral of the square is finite, F of T DT is less than –

I want to allow complex value functions here, although many of the applications are real, but I want to allow complex value functions, so I put the absolute value of F of T squared there, all right?

That's an easy condition to satisfy. Physically one encounters this condition in the context of this integral representing energy and so one also says that this signal has finite energy. That's another way of saying it, and you see that terminology.

All right? So if you have a periodic function, which is integrable, like so, all right, so again if F of T is periodic and square integrable, then you form the Fourier coefficients and they exist actually the square integrability is enough to imply that the Fourier coefficients exist.

Then your form, [inaudible], as before, either the minus 2 pi I, KT F of T DT and the sum converges the infinite Fourier series is equal to the function in the sense of mean square convergents.

And you have, and this is the fundamental result, that the integral from 0 to 1 of the difference between the function and the finite approximation between the function, the square of the difference, the integral of the square of the difference, this tends to 0 as N tends to infinity. Takes up the entire board and it deserves to. All right, it's an important result, okay?

Okay, now I really feel like have to say a little bit more here, but only a little bit more. The fact is you only get these wonderful convergents results if you not only generalize your point of view towards convergents, but you also have to generalize the integral.

All right, this actually only holds for – the whole circle of ideas really only holds for a generalization of the [Inaudible] integral. Do you have to worry about this? No. All right. But the fact is, you only get these convergence results, such convergence results. You can only really prove them in the slightly more general context of integration.

Convergence results, if you use a generalization of the integral even, which is a whole other subject due to Lebeg, and that's what the L stands for in L2, due to Lebeg. It's a French mathematician in the turn of the 20th Century, who actually in the context of these sorts of applications, trying to extend integration, trying to extend limiting processes to more general circumstances where they couldn't prove the results classically, mathematicians were worried about this, he had a more general definition of the integral.

And with that more general, more flexible definition of the integral, the limiting processes are easier to handle. You can do more than you'd like to do, but somehow don't feel justified in doing, and in that context, this is something you don't have to know about, all right. But he generalized the integral only, and by doing so, it was perfectly suited to solving these sorts of problems.

It was really a quite compelling case. It was really a beautiful theory, but, you know, it's a famous quote. The usual integral that you studied when you studied integrals in Calculus is called the [inaudible] integral. All right, and that suffices for just about every application, but there are more general integrals.

On the other hand, John Tukey who was a famous applied mathematician was quoted as saying, "I certainly don't want to fly in an airplane whose design depended on whether to function as a remount integral or a Lebeg integrable. That's not the point."

All right, the point really is a theoretical one, not a practical one. But nevertheless, somehow honesty compels me to say what's involved here. All right, so it's in that sense that you had to talk about convergents, and it's not an unreasonable condition, all right, that is also called, as I said last time, convergents in energy, convergents in the mean or mean squared convergents.

And what that means is on average this sum is converging to the functional and average in the sense that if you look at the difference between the function and finite approximation, square that, integrate it and integrating is sort of taking an average and then that tends to zero as [inaudible] infinity.

It's approximating over the entire interval on average rather than concentrating its efforts instead of approximating it at a single point. And again, you will, if you look at the literature and I'm talking about the engineering literature, you'll see these terms all the time. You'll see L2 all the time.

As a matter of fact, actually, if you take, some of you have probably had courses in quantum mechanics and if you take it all sort of an advanced course in quantum mechanics, not that I ever have, but if you do, it's also framed very much in the context of L2 [inaudible] spaces and things like that, [inaudible] spaces and L2 spaces and so on.

It's not, I'm not making this up. It really is – it has become the sort of framework for a lot of the discussion. If you want to be sure that you have some certain amount of confidence in applying mathematical formulas, you need the right kind of general framework to put them in and this is for many problems, exactly that.

Now, there is one further aspect of it, and that's as far as I'm gonna go, that brings back that fundamental property of the complex exponentials that we used to solve the Fourier coefficients. So I want to highlight that now.

So remember in solving for the Fourier coefficients way back when, I mentioned this last time and I wanted to bring it back now in a more general context again, in solving for the Fourier coefficients we used a simple integration fact that it was actually everything. It was essential that either the 2 pi, NT times the [inaudible] minus 2 pi MT DT so that we can combine those complex exponentials, that's either the 2 pi and minus MT was equal to zero if M is different from N. If M is equal to N, then it's equal to 1, all right?

That simple fact, that simple calculus fact emerges – turns out to be the cornerstone for understanding these spaces of square integral function – if we're introducing geometry into those spaces. So this simple observation, simple fact, is a cornerstone for introducing, if I can say it that way, introducing "Geometry," I put it in quotes because it's geometry like you don't even think of geometry, but the features are there by analogy.

Geometry into the space into square integral functions, L2 01. And when I say introduce geometry, again I'm reasoning by analogy here, although it's a very powerful analogy. And the thing that makes geometry, geometry as far as Euclid is concerned is the notion of no perpendicularity.

All right, that's one of the most important notions of geometry and that's exactly what gets carried over here. It allows you, it allows one to define orthogonality or perpendicularity, same word, same thing, all right, via inner product or dot product.

All right, so let me give you the definition. I'm not gonna justify it because there's justification in the book actually, in the notes. But it looks like this, so again if F and G are square integrable functions and I'm gonna assume they're complex, there's a little distinction here between what happens in the real case and what happens in the complex case, square integrable on 01, as always, that's the basic assumption we make.

Then you define they're inner product which is a generalization of the usual dot product for vectors a generalization of the dot product for vectors, by an integral formula, all right, it's defined by FG and you're right, there are various notations for it, but the common notation is just to put them in a pair, or sometimes people will write them with angled brackets or sometimes people do all sorts of things.

Sometimes physicists call it, you know, [inaudible] vectors and [inaudible] vectors and all sorts of bizarre, unnatural things. But I'll take the simplest notation, so zero to one, F of T times G of T bar complex conjugate DT. That's because I'm allowing the complex value of functions.

If G of T were a real value of function, then I would just have F times integral or F times G because it's complex for reasons which again are explained a little bit better in the notes. You put the complex conjugate in there.

Now, what you have to believe is that this is sort of a continuous infinite dimensional generalization of the [inaudible] of two vectors. How do you take the [inaudible] of two vectors? You multiply the components together, the corresponding components together and add, all right?

So it's like you're multiplying the values together, although the function times the conjugate of the other function, those are the values, and you're adding, but continuously in the sense of taking the integral instead of the sum, all right? That's sort of where it comes from.

Now, that's fine, but the real benefit is it allows you to define if you ever want to do that, when two vectors, when two functions are perpendicular. So you say that F and G are orthogonal, you could say they're perpendicular, but it sounds ever so more mathematical if you say they're orthogonal, all right.

Orthogonal and if you write it neatly – orthogonal, much neater. If they're in a product, it's 0, if FG is equal to 0 period – that's a definition, all right. That's a definition. Now, where does the definition come from, again, you know, because I can't say everything as much as I'd like to, let me refer to the notes, because it's actually not so unreasonable. This definition comes from the exactly wanting to satisfy the Pythagorean theorem or inner products.

Now, the calculation we do with the complex exponentials shows exactly the different complex exponentials are orthogonal. One more thing actually, let me introduce the length of a function or the norm of a function in terms of the inner product, or in terms of the integral, same thing.

So the norm of F, norm of a function F, is defined by the square of a norm is just the inner product of the function of itself, just like the square of the length of a vector is the inner product of a vector with itself. So the inner product of F with itself is the norm of F squared and that's the notation you use, and so what is that; that's exactly – so the norm of F squared is exactly the integral from 0 to 1 of F of T squared DT.

And let me just tell you what the Pythagorean theorem is then. All right, the Pythagorean theorem is – I'll do it over here, because that's where this comes from. That's exactly where this definition comes from and that's exactly where the property comes from. The Pythagorean theorem is exactly this, that F is orthogonal to G if and only if the norm of F plus G squared is equal to the norm of F squared plus the norm of G squared and that is if and only if the inner product is 0.

Why is that the Pythagorean theorem – because of vector addition, all right? If I wrote vectors, U and V, then this is the vector U plus V, all right, if I just write vectors and the Pythagorean theorem says, "The square of the hypotenuse is the sum of the squares of the sides for right triangle and only for right triangle." That characterizes the right triangle, all right.

So that says the norm of U plus V squared, the square of the length of the hypotenuse is the sum of the squares of the other two sides, U squared plus V squared, and that holds only exactly when the two vectors are perpendicular and that's where the definition of the inner product comes from and everything else.

Beautiful – it's beautiful, all right. So the trick is in extending that from, if you like this sort of thing, the trick is extending that from vectors to functions and reasoning by analogy, all right. It's not, you know, now let me say what you should think about and what you shouldn't think about, the analogy is very strong. In many ways your geometric intuition for what happens for vectors you can draw carries over at least algebraically to this more general situation; however, so let me say one more thing then.

These complex exponentials are exactly functions or orthogonal functions of length one. That calculation that I've just erased with complex exponentials says this, says either the 2 pi NT in a product with either 2 pi MT, all right, that's the integral of this times the conjugate of this. The conjugate of either the 2 p MT is either the minus 2 pi MT, that's where the minus sign comes from, all right.

This is equal to zero if M is different from N and it's different to one if M is equal to N. All right, these are ortho-normal vectors with respect to this inner product. They're orthogonal, their inner product is zero, when they're distinct, and they have Length One, their norm is equal to 1 when they're equal.

Okay, now, as I said your reason by analogy, you can visualize what it means for vectors to be perpendicular. You can draw this picture, all right. So you might say to yourself, I should be able to visualize this. I should be able to sit in a quite room, turn the lights off and visualize what it means for complex exponentials to be orthogonal. Yes, yes, I see it.

No, you don't. Let me relieve you of this burden. There is no reason in hell that you should be able to visualize when two functions, let alone complex exponentials are orthogonal. Don't beat yourself up trying to do that and don't say to yourself you're less of a person if you cannot visualize when two functions are perpendicular, not only complex exponentials, but sines and cosines, all sort of innocent-looking functions out there that you have worked with for all of your life turn out to have this sort of orthogonality relationship.

But you might say to yourself, like sines and cosines, so sine FT if orthogonal to cosine FT or sine of 2 p T is orthogonal to cosine 2 p T. All sorts of interesting results like that, but you might say to yourself, "Gee if I look at the graph, I should be able to visualize this." Don't bother.

All right, there's no point. There's no point. It's reasoning by analogy, all right. The fact is you establish thee formulas; you establish they're orthogonal and then you apply your intuition for orthogonal functions, for orthogonal vectors where you can visualize it to situations where you can't visualize it, all right and that's the real power of this line of reasoning, because you can apply your intuition to places where it should have no business applying somehow, all right.

All right, now, almost, almost, almost, almost, almost, the final thing then, the final piece of this approach to Fourier series is to realize the Fourier coefficients are projections of the function on to these complex exponentials, all right. So again, I want to remind you of one of the ways you use inner product is to define projections, to define orthogonal projections in particular, so you use the inner product for vectors to define and compute projections.

All right, if U and V are two vectors, univectors, say, norm of U equals 1 and norm of V equals 1, then what is the projection of V onto U, and the projection onto V onto U is just the inner product of V.U. That's how much U knows V or V knows U, all right. The projection here is the inner product of V with U, okay. And how much does U know V is the inner product of U, same thing.

All right, so the vector projection here is that's the length of the projection, all right, and then the vector that goes in the direction of U that has this length is this time, so the vector projection is inner product of V with U times U.

All right, U is the univector, so you go in that direction this length and that's how you project, and it' shouldn't be shocking. It should be somewhere in your background to realize that you've certainly had classes in linear algebra that decomposing a vector into its projections onto given vectors can be a very useful thing.

It’s breaking a vector down into its component parts, breaking a vector down into its component parts, all right. Now, what is the situation for functions, it’s exactly analogous. I don’t want to say it’s exactly the same because you can’t draw the picture, but you can write down the formulas and the formulas are a good guide. The formulas are a good guide.

What is the Fourier coefficient? The Fourier coefficient is exactly a projection, all right. If I compute the inner product of the function F with the complex exponential, either the 2pi NT, then that’s exactly the integral from zero to one of F of T times either the 2 pi NT bar DT. That is the integral from zero to one of F of T either the minus 2 pi NT DT, that is the nth Fourier coefficient. The nth Fourier coefficient is exactly the projection of the function onto the nth complex exponential.

All right, cool. Way cool. Infinitely cool. So cool. And what is writing the Fourier series? What is writing the Fourier series – to write F of T is equal to the sum from K going to a minus infinity to infinity, [inaudible] either the 2 pi KT is to write F of T is the sum from K going to a minus infinity to infinity of the inner product of F with the K complex exponential times the complex K complex exponential.

This is a number, all right, that’s the Fourier coefficient. That ‘s the length of the projection of F onto its K component, and there’s the K component. All right, it’s exactly what that is, and it’s this point of view that is so ubiquitous, all right, so ubiquitous, not only in a Fourier series, but in other versions of essentially the same ideas.

And you see this all the time in signal processing. I talk about wavelets time. I mentioned wavelets just very briefly because wavelets is such a hot topic. It’s the same sort of thing. You’re trying to decompose the function into its simpler components and in this case the simpler components are either the 2 p or the complex exponentials, all right.

So to write an expression like this down to be able to say this in the appropriate notions of convergents is to say – I’ll do it over here – is to say that the complex exponentials form an ortho-normal basis for the square integrable functions, all right.

To be able to write this statement and understand what it means in terms of convergents and all the rest of that jazz is to say, is to say, that the complex exponentials that is all of these, even the 2 pi KT, K going from minus infinity to infinity form an ortho-normal basis for these periodic functions, squared integral periodic functions, all right.

And then sometimes the game is to take different ortho-normal basis. Wavelets are nothing but, I'm not going to say nothing but because they have their own fascination, they’re another ortho-normal basis for square integrable functions.

The complex exponentials are not the only ortho-normal basis just like any vector space just doesn’t’ have, has lots of different orthogonal-normal basis. These are particularly useful ones.

So to write this synthesis formula is to express F in terms of its components, what components, the components in terms of these elementary building blocks, all right, and what are the coefficients, the coefficients, like they are for any ortho-normal basis are the projection of the vector in those directions, all right.

It's very satisfying and you should try to put this in your head, yeah.

Student:[Inaudible] in the sense that you can't get one from the other in the rotation [inaudible].

Instructor (Brad Osgood):Well, it's more complicated, so the question is are the bases like in [inaudible] dimensional spaces the bases are related to each other by a essentially a rotation of unitary or orthogonal matrix and in space, yes, you have unitary transformations linear operators that are unitary, but the definitions are a little bit more complicated.

But you have similar sets of things. All right, you can get to ortho-normal basis to another ortho-normal basis. Okay. All right, this is, like I say, so what one can say, of course, I would say much more about this. I don't want to. All right, well, I do want to, but I can't, all right.

All right, it's this point of view that is important for you to carry forward, all right. That's all I'm saying because you will see it, you will see it, all right. And again, the idea is you reason by analogy, all right. You gotta; it's hard to write down this formula, all right, that's something new. All right, writing down a formula, writing down an inner product in terms of an integral, all right, I gotta deal with that. That's something hard, and I can't visualize it, all right.

But the words you use in the case where you can visualize things are almost identical to the words you use in the situation where you can't visualize things. All right, and you can carry that intuition over from the one case to the other case and it's extremely important and I'll give you one, I keep saying one final thing.

So this time for sure, an application of this is what's also called Railey's identity, which is nothing more than to say a length of a vector can be obtained in terms of some of the sum of the squares of its components, all right.

You know how to find the length of a vector in [inaudible]. You add up the sums of the components, right? You do the same thing here. Railey's identity, and I will not derive it for you, but it is derived in the book and I even say something like, "Do not go to bed until you understand – do not go to sleep until you understand every step in the derivation."

It says the integral from 0 to 1 of F of T squared DT. You write it in terms of its Fourier coefficients – I didn't write that down, is the sum of the squares of the Fourier coefficients. Okay, going from minus infinity to infinity of [inaudible] K squared, all right. That's Railey's identify.

It says the length of a vector is the sum of the squares of its components, all right, the components of the function are its Fourier coefficients, all right. This is the length of a function; it is the inner product of the square length of a function. It's the inner product of a function with itself. It is the sum of the squares of its components. That's all this says. And it follows algebraically exactly in the same way as you would prove this using inner products for vectors, exactly, exactly.

All right, now, this was known before any of this stuff was put in place, all right, but when all this sort of general framework was put in place, this identity was known before the general framework or orthogonal functions, square integrable functions, all the rest of that jazz, and it was viewed as an identity for energy, all right, this is the energy, the function and these are somehow, you know, the individual components here.

And that's why one often says, you can compute the energy and the time domain or the frequency domain, we're gonna find an analogy for this before it transforms, all right. But it really says nothing other than the length of the vectors, the sum of the squares of the components.

How much – so let me write down, here's F of T is the sum of, here's the Fourier series, [inaudible] either the 2 pi KT, all right. All right, how much energy does F have as a signal? It has this much. All right, how much energy does each one of its components have?

Well, the energy of each one of the complex exponentials is one because they're of length one. So, how much energy, how much square energy does each one of the components have? It's the magnitude of this thing squared; it's the multiplier. It's the projection out front.

[Inaudible] squared times the energy, which is 1 of the complex exponential, so what is the total energy here, what is the square of the total energy here? It's the sum of the squares or the contributions of that energy from individual components. Each individual component is contributing an amount [inaudible] of K to the whole sum, or the energy it contributes is the square of [Inaudible] of K absolute value squared to the whole sum, the whole number, okay, pretty cool; pretty cool.

All right, here ends the sermon. Don't leave without putting something into the collection plate. All right. All right, here ends the sermon on inner product square [inaudible] and so on, okay? All I can say is trust me; you'll run into it again.

Now, I want to do and I probably won't finish it today. I'll finish it up on Wednesday, although it goes pretty fast. I want to do a classic application of Fourier series to the study of the classic physical problem, in fact, the problem that launched the whole subject. Sop I want to do an application to heat flow.

All right, this is a very important part of your intellectual heritage, all right. I'm serious. That is if you're going to be practicing scientists and engineers, you know, you want to know something about where the subject came from because again, you know, you some how wind up re-visiting a lot of these ideas in different context, but often in similar context.

And this was the problem that really started it all. This was the problem in studying how heat, how the temperature varies in time, when there's some initial temperature distribution.

All right, so you have a region in space and with an initial temperature distribution and an initial distribution F of X of temperature. I say F of X here just to indicate that X is some sort of spatial variable, all right, so some region in space, you know, the dimension of X is the dimension of the region, all right. And the question is how does the temperature [inaudible], all right.

You have an initial distribution of temperature, that's what happens at T equals zero, then as T increases, the temperature changes. All right, the heat flows from one part of the body to the other part of the body and you want to know how is that governed, all right.

How does the temperature change both in position, I should say vary, well, change is fine – how does the temperature change in position and time. All right, this was an important problem, still is an important problem actually and we're only gonna handle one very special case of it, the case of the original case that was handled by Fourier series, the problem with the Fourier study.

Where periodicity comes into the problem naturally because of the periodicity in space. So to study this problem means to say first of all, what the region is and then to say what the initial distribution of temperature is or at least say that there is given an initial distribution of temperature, all right.

So we look at a heated ring. Sound – that sound. Okay. All right, something like this, all right. Given an initial temperature, given an initial temperature say F of X, F of X – X is a position on the circle; X is a point on the circle, okay? And we let U of XT be the temperature at X at time T – at position X at time T. All right, and the question is can we say something about it. That's the function we want to find.

We want to study U of XT, all right. Now, the fact is periodicity enters into this problem, periodicity in space because the circle, the temperature here is the same as the temperature at the same place, if I increase it by 2 p if I'm going around a circle, then obviously the temperature is periodic as the position.

Okay, so the temperature is periodic as a function of X. All right, and let's normalize things so we assume the period is 1, so the circle is radius or whatever it is, or another way of looking at it is just imagine the circle is the interval from 0 to 1 and I identify the end points, all right. So let's supposed we have – just because we've been dealing with functions of Period 1, let's suppose that. Let's supposed Period 1.

Period 1, all right, so that is the function of the initial temperature distribution, F or X is periodic of Period 1 and so is UNX, not in T; it's not periodic as a function of time, but it's periodic as a function of the spatial variable. U of X plus 1 for any variable of T is the same thing as U of X at that value of T, okay?

That's fundamental and that's how Fourier got into this; that's why he introduced those ideas. You'd consider this problem where there's periodicity in space. The symmetry of the object that you're heating up and that has consequences. All right, so now, with a certain amount of confidence, with a certain amount of bravado, we write down the Fourier series.

So we write the Fourier series, U of XT is the sum from K going to minus infinity to infinity of C sub K, I'll write it like this, C sub K either the 2 pi KX. It's periodic in the spatial variable, so the variable in the complex exponential is X. Where is the time dependents? No, the time dependents is in the coefficient, C. That's where the time dependence has to be. K is just a constant; K is just an integer.

The time dependence is in the cK. That is a better way and more accurate way of writing it is like so. U of XT is the sum from K going from minus infinity to infinity of C sub K of T. I know what it is; it's the Fourier coefficient, I'll bring that in later, but let me just write in terms, let me just call it C right now.

C sub K of T times E to the 2 pi KX, all right, periodic and X varying in T. How does it vary in T; that's what we want to know, all right. The mystery here are the coefficients. We could solve the problem if I could find the coefficients, so in terms of the initial temperature distribution, so what are the CK? That's the question.

All right, now, we're gonna be able to attack this problem because independent of periodicity, independent of anything else, heat flow is governed by a partial differential equation. All right, the flow of heat on a circle or another regions – all right, this is in itself a big subject, but it's one of the basic equations, partial differential equations of mathematical physics, which you have probably seen somewhere in your life and again, you will see again, heat flow – I'll just do it in one dimension, all right.

I'm talking about one-dimensional problems here, but there are also ways of analyzing this for flow over a two-dimensional or a three-dimensional region. We have the heat equation, which says that the time derivative of the temperature is proportional and somehow maybe I should call it another constant because I've already used K, A times the second X derivative, all right. That's the one-dimensional heal equation.

All right, I'm not gonna derive that. Actually I have a derivation of that in the book that sort of follows Bracewell's discussion of the heat equation, but it's a fundamental equation of mathematical physics. The constant A here depends on the physical because this is one of these great dodges of all time.

The constant A depends on the physical characteristics of the region which no one wants to talk much about, but that affects the size of A, all right. I should say more generally, this equation governs not just the flow of heat, but it's in general is called the diffusion equation.

It governs how things diffuse. Things – what things, like charge through a wire is studied by this equation, holes through a semi conductor are governed by an equation, a higher dimensional version of the equation but the same idea, all right, so this is general governs, this is also called the diffusion equation and it governs phenomenon that are associated with diffusion.

It's a general term, but it's a term you'll probably run across. All right, now I want to choose; I'm not gonna get too far today, but I'm gonna get a little along the way, so I wanna choose – I want to apply this equation to study this function. All right, I'm gonna use this for the ring. All right, and just to simplify my calculation, although it does not make any substantive difference, I'm gonna choose the constants so that A is equal to ½.

That's a standard choice, certainly for the mathematical analysis, but it doesn't matter. You could have the constant tagging along in the whole thing and it wouldn't' affect the analysis. So I'm gonna choose A equals ½ or choose constants of A equals ½. So the equation looks like use of T is equal to ½ use of XX, all right.

Now, I'm gonna short-cut the discussion a little bit. There's one way of doing it in the book, in the notes, which is a little bit more rigorous than what I'm gonna do although both can be justified very easily. I'm going to plug that equation, that formula for – we have a formula for U in some sense, not in some sense, we have a formula for U in this sense.

And I'm gonna plug this into the equation. Plug U of XT is sum over K, CK of T, either the 2 pi KX into the equation, into the use of T is equal to ½ UXX. What happens?

What happens? Well, so use of T, if I differentiate with respect, I'm sorry; I should have said this, but I assumed everybody knew, use of T is a partial [inaudible] with respect of T, or that function over there, so what is that? Well, the only thing that depends on T here is the coefficients, C sub K. So that is sum over K, CK prime of T times the complex exponential. That stays alone -- 2 pi KX. What is UXX of this expression?

Well, here I've put the derivatives on the complex exponential and I differentiate twice, differentiating complex exponential is like differentiating an ordinary exponential. The constant here comes down if I differentiate, so I get CK of T, that's left alone because I'm differentiating with respect to X and then if I differentiate the 2 pi KX twice, with respect to X, I get 2 pi K squared times either the 2 pi KX.

Nothing up my sleeve, no tricks, no deceptions. Okay, that's one more step, that's the sum over K T sub K of T times minus 4 p K squared because I squared is minus 1 times a complex exponential either the 2 pi KX. All right, not quite the two.

Then we gotta go – damn. Equate the two sides, use the heat equation. Plug into the heat equation. So I get sum plug into UTs is equal to ½ UXX, and then I get sum over K CK prime of T, either the 2 pi KX is equal ½ times that thing, so it's gonna be sum over K minus 2 p squared K squared times C 2 p squared K squared times CK of T, times the complex exponential, either the 2 pi KX. Not hard, that's not a hard step.

Okay, all right, how would we do? Equate like terms; equate the coefficients. If I equate the coefficients, something great happens. [Inaudible] consequences I'll do next time. If I equate the coefficients, I'll get CK prime of T is equal to the coefficient, when I say equate the coefficients, I mean the coefficients of the complex exponential. CK prime into the 2 p KX, what's the corresponding coefficient over here. It's this is equal to minus 2 pi squared K squared, CK of T.

But my friends, you know, that is a simple equation. That is an ordinary differential equation for CK. I can solve this other problem. I get CK of T is the CK of 0, the initial condition times [inaudible] minus 2 p squared, K squared T. Blow me down, all right.

I've found my coefficients – pretty cool. All right, extremely cool. Very cool. And next time what I haven't done is I haven't brought back in the initial distribution of temperature and I want to manipulate this solution a little bit more and something absolutely magical happens. Wait for it on Wednesday. Okay.

[End of Audio]

Duration: 54 minutes

TheFourierTransformAndItsApplications-Lecture05

Instructor (Brad Osgood):Okay, ready to rock and roll? All right. We've been getting a lot of – as I understand it, we've been getting a lot of email questions about when and where to turn in the homework. You can certainly bring it to class, and that's fine.

But if you find yourself wanting to fill in those last minute comments, I would say that – the policy is the homework should be due by 5:00 on Wednesday, and you can turn it in to the magic filing cabinet. Across from my office – my office is 271 Packard, and there's a little hallway sort of across from there, and there are several gray filing cabinets, one of which has my name and the course number on it.

And if you open that drawer, you will see a wire basket in there with a little sign on it that says something like, "Turn in homework here." So you can just drop your homework in there, okay. So by 5:00 today the first problem assignment is due. All right. Any questions? Anything on anybody's mind?

Okay, all right. Anything else that came up that I should address? No, okay. All right. Today, I wanna finish off our discussion Fourier series. And sometimes we don't really finish the discussion of Fourier series because it will always be a touchstone for reference for some of the other things that we do.

But I wanna finish up the discussion we started last time about using Fourier series to solve the heat equation. And then I wanna talk about the transition from Fourier series to the Fourier transform, and how one gets from the study of periodic phenomenon to the study of non-periodic phenomenon, which is exactly what the Fourier transform is concerned with, by means of limiting process, all right. So that's where I wanna finish up.

But first, let me go back to the discussion we started last time about the heat equation. So this is the use of Fourier. This a classic example, the classic example one might say, of Fourier series, and also shows in a particular case, a very general principle that we will be seeing constantly throughout the course. That's really one of the reasons why I wanna talk about it.

So this is your Fourier series to solve the heat equation, one particular case of a heat equation. And I'll remind you what the setup is. We have a ring, we have a heated ring like that with an initial temperature distribution of – we're calling F of X. So the initial temperature – so I'm thinking of X here as a spacial variable.

I wanted to mention a special variable, although the ring is sitting in two dimensions. So I can think of the ring, if I want, as an interval from zero to one with the end points identified.

At any rate, the fundamental – the important fact here is that since there is periodicity in space, since the ring goes round and round and round, the function F is periodic as a function of the spacial variable, the position on the ring. And we can normalize things to assume the period is one, all right. So that's how Fourier series comes into the picture.

So we take F to be periodic of period one. Then we let the U of XT – now, the temperature is varying, both in position and in time. The temperature change is a function in time, and the temperature is different at different points along the ring.

So I let U of XT be the temperature at a position X at time T. Then U is also a function – it's a periodic function in this spacial variable. U is a periodic function of X. So U XT is periodic in X. That is U of X plus one T at any time T is U of X at the same instant of time. When T is fixed, this periodic is a function of X.

And the physical situation is described by the heat equation, which is also referred to as the defusion equation, and governs many similar processes that involve the diffusion of something through something else [inaudible]. So heat through a region charged through a wire is governed by this sort of equation. I mentioned these last time, holes through a semiconductor. It really has quite a variety of applications.

And some of the techniques that we're talking about here, although they're specialized to this case, can be applied in various forms to many different situations. So you have the heat equation [inaudible] equation, so it relates to the derivatives in time to the derivatives in X in space.

So it says UT is equal to one-half UXX. Here I'm just choosing the constant – there's a constant on the right-hand side of the heat equation. I'm just choosing things so the constant is one-half just to simplify the calculations. So this is – the first derivative with respect to time, and this is the second derivative with respect to X.

All right, now because the function is periodic in X, we can expand it as a Fourier series. So the rigger police are off duty. I'm assuming that everything converges here and there's no question about writing down the sums.

This is sort of a formal operation whose particular – the particular manipulations that I'm gonna do can be justified under reasonable assumptions, but that's not the point. The point is just to see how the techniques can be used to solve the equation.

So I can write the function as a Fourier series. Now remember, if periodic is a function of X, so the [inaudible] on T is on the coefficient. K one from minus infinity to infinity, that's – I'll call it C's of K of T times the periodic term, either the two pi IKX, ano I'm assuming period one. That's the basic assumption. Or rather, that falls from periodicity. You can expand it as a Fourier series. That's how Fourier series get into the picture.

Then you plug into the equation – we did this last time. I'm just reminding you of the setup. Plug into the heat equation and equate coefficients and you get an ordinary differential equation for the C's. The C's are the unknowns there.

So the – you get this equation, you get CK prime of T is equal to minus two pi squared, K squared, times CK of T. That's a simple ordinary differential equation, and we know how to solve it. That's a simple [inaudible].

And solution is CK of T is E to the minus – is the initial condition TK at zero times equal the minus two pi squared, K squared T. That's easy. That is easy. And this is where I think we got to last time.

Now, what is the initial condition? I haven't brought in the initial temperature here, but here's where it comes in. So what is CK of zero? All right, CK depends on time. What is the initial condition?

Well, we can see it, actually. Remember, UK – excuse me, U of XT is the sum from K equals minus infinity to infinity. CK of T times E to the two pi I, KX. So what happens to T equals zero? T equals zero, this is the initial temperature distribution. U at X, any position on the ring, at times zero is F of X. You give it some initial distribution of heat.

So that says that F of X is U of X zero. That's the sum from – if I plug into the series, sum from minus infinity to infinity of CK of zero, E to the two pi IKX.

Now, we have eyes, but if only we but see. What does this say? F is a periodic function. This is an expansion of F in terms of the harmonics, in terms of the building blocks that give you the two pi IKX. What are the coefficients of C, what are the coefficients CK of T in terms of F? What is CK in terms of F, CK of zero in terms of F?

Student:[Inaudible]

Instructor (Brad Osgood):Pardon me? It's a Fourier coefficient of F. So this must be the Fourier series of F. That is to say, i.e. CK of zero is just the K Fourier coefficient of F, F out of K.

Now, as they say, we have eyes, if only we but see. What I meant by that provocative statement is you have to be able to go from the function to the series, but you also have to be able to go from the series to the function.

Here's the function, here's the series. You have to be able to relate and say this is a series for a function, so it must be that these coefficients are the Fourier coefficients for the function.

You're used to having problems like here's the function, compute the Fourier coefficients. All right. Well, in some sense the Fourier coefficient's already computed here, so relate them to the function. Let's write that down.

That is – when I say write that down, what I mean is what does the solution look like. So U of XT, the temperature – this is already very impressive. The temperature at any time T at any position X on the circle is F out of K times E to the minus two pi squared, K squared T, times either the two pi IKX. You can see from this the dependence on time and the dependence on X.

It's very – it's a complicated expression, but it's pretty impressive. You have found that given an initial temperature distribution, this is a formula for how the heat for the temperature at any point X on the ring at any time T. You can already draw some conclusions from this. Namely, as T tends to plus infinity what happens to the temperature? It goes to zero. As T tends to infinity, this term is damping out.

Now see, there's a cute little thing to observe here. This is an exponential, either the minus two pi square, K square T. Now, K is going from minus infinity to infinity, all right. K is both positive and negative here. But K appears here squared, so K squared is always positive. So this, multiply two pi – minus two pi square, K square, this number is always positive, so minus it is always negative. It's either the minus something as T is tending to infinity, all right.

That's a little sort of check on consistency there. So as T tends to zero – this is just an observation. As T tends to infinity, U of XT tends to zero for any value of X. So the ring is cooling off eventually.

Now, this is perfectly fine. Actually, this is a perfectly fine way of writing the solution for reasons which I will explain, and for reasons which we will see often times in the course. I wanna work with this a little bit more, and I wanna write the equation in a different form. This was exactly the sort of thing I had you do in homework. Although, in homework it turned out even a little bit simpler than this. It may not have seemed it, but it did.

I wanna write this sum differently. I mean, again, we've solved the problem in the sense that we have found a formula for the temperature at any point X on the circle at any time T. All I'm saying is there's an alternate way of writing it that itself is very revealing, and in fact, has consequences that go beyond the particular problem that we're studying.

So to do that, I wanna bring back the formula for the actual – for the Fourier coefficient. On the right, F hat of K. I wanna use the explicit formula as an interval. So I wanna use F hat of K is the interval from zero to one – even though I've already used X as the variable here, so the variable for integration I'm gonna – I'll use something else, like I'll call it – I'll just use Y. It doesn't matter what I call it, as long as I'm consistent.

So I'll write this as E to the minus two pi IKY, F of YDY. All right. That's the formula for the K Fourier coefficient. And plug that into the formula for the solution.

So I get U of XT is the sum, and from K – from minus infinity to infinity, this interval is the interval from zero to one, E to the minus two pi IKY, F of YDY. E to the minus two pi squared, K squared, T. E to the two pi IKX. That's why you have to do this on a big blackboard with big chalk.

And I’m gonna combine terms and swap the interval and the sum. Again, the rigger police are off duty, and so interchanging integration and summation is something we can do with impunity.

Mathematicians can do it with impunity also, but it takes them three years of graduate courses before they really feel good about it. We don’t have the time for that.

So this is – I’m gonna swap the interval and the sum, so this is the interval from zero to one of the sum K, going from minus infinity to infinity. So I’m gonna put the terms together like this; E to the minus two pi IKY, E to the two pi IKX, E to the minus – I’m putting all the terms together that get summed. E to the minus two pi squared, K squared T. Those terms all get summed, and then what stays on the outside is F of YDY. I haven’t done anything there, I’ve just rearranged things. And I’ve swapped integration and summation.

And I’ll do one more step, and that’s the interval from zero to one, sum from K going from minus infinity to infinity of – I’ll put these two terms together because they’ve both got I’s in them. E to the to pi IKX minus Y, E to the minus two pi squared, K squared T, F of YDY.

Like I said, I haven’t done anything really, except rearrange terms. I’ve rearranged terms this way because I know what’s gonna happen, or rather, I know how to finish the discussion because I know what years of bitter experience taught the people before me finally.

That is F sum deserves to be singled out for special attention, and deserves to be given its own special name. So I’m gonna write, say, G of XT to be this sum. K going from minus infinity to infinity, E to the two pi IX, E to the minus two pi squared, K squared T.

This is actually G of X minus Y, T. That is the interval. The solution appears in this form. Pardon me?

Student:[Inaudible]

Instructor (Brad Osgood):KX, thank you. Two pi IKX.

So the solution U appears in the following form. I just use that as a shorthand notation. The solution looks like this; U of XT is the interval from zero to one, G of X minus Y, T, F of YDY. And that’s all I wanna – that’s the – I’m not gonna do anything else. I promise you, I won’t do any more rearranging; I won’t do anymore fiddling around. That’s how I’m gonna write the solution.

Now, for those of you who’ve seen this, and you actually saw this – you saw a similar sort of thing on the homework problem. This expresses – so this is an important statement that I’m about to make. Drum roll. This expresses the solution, the general solution that any time X and any time T as the convolution of the initial condition at times zero with this kernel.

This expresses UXT as the convolution of F of X, F of Y, it doesn't matter, F of X with what's called the heat kernel G of XT. So there are a lot of terms there that we haven't heard before, although you may have heard in different context the term convolution, which we are gonna be hearing all the time. And in this special case, you also call this the heat kernel. You also call this the fundamental solution or the Green's function for the heat equation.

So I will just – without saying really anything more, just to introduce you to the terminology, you call G of XT – has a variety of names. That function that I wrote down, G of XT, is called variously the heat kernel. That is it's a kernel for the heat equation. You use the word kernel often when it appears underneath and interval in the context of convolution like this.

Heat kernel is also called the fundamental solution of a heat equation, and it's also called Green's function for the heat equation.

Now, let me just ask, just to take a survey out there. First of all, who's studied this problem before? Who's studied the problem of heat flow sort of this way with Fourier series? All right, so a couple people, but not that many actually. In what class and what context? I'm just curious.

Student:[Inaudible]

Instructor (Brad Osgood):Oh, is that right? Okay. And they did it more or less like this?

Student: Pretty similar.

Instructor (Brad Osgood):Yeah. They probably worried about certain things, like convergence and things like that, right, but screw them.

And how about this terminology? So have you heard this terminology, the heat kernel, the fundamental solution or the Green's function?

Green's – people who have taken many physics courses have often heard the term Green's function for differential equations and things like that. You may have seen that. And we're not gonna – I'm not gonna make a big deal out of it, but again, it's sort of an indication of the kind of techniques and the kind of ideas that come up in this class you see everywhere.

For us, the reason why I went through the solution, and the reason why I wrote the solution this way was to bring up this idea of convolution. Convolution is a very general operation. People in electrical engineering have seen it very early on in their classes on signals and systems. We are gonna see it in all sorts of different context. Not always in terms of differential – not always associated with differential equations, but sometimes associated with differential equations. And it is the kind of thing that comes up.

The fact that you could write a solution of the convolution of two functions, in this case, the initial distribution of heat, and the special functions associated with the equation.

I'm not gonna say anything more about it now. I'm not gonna even give you the general definition of convolution. But all I wanted to do by showing you this was it comes up in a very – in retrospect, in a natural way. It's the sort of thing you should expect to see.

It's one thing – when you're working with any problem almost that has to do with Fourier analysis, it should not surprise you to see convolution coming into the picture somehow. This is an example of that.

On the homework you had another example of it. You had an example of another celebrated problem in mathematical physics, the so-called [inaudible] problem, where again, the solution ultimately could be written as convolutions.

There, the nice thing was that you actually got a closed-form expression for essentially the Green's function, or the fundamental solution for that problem. The plus on kernel is actually a closed-form expression. There's no similar closed-form expression here, or you can't do anything more with this function. It is what it is. It's just this infinite sum.

For the problem you had in homework, again, you get an infinite sum that comes in, but it collapses because it's a geometric series. And you get a nice closed-form expression for the plus on kernel.

But in principle, the principles that are operating here are very similar for the two problems. And again, you'll – it's sort of a fact that took a long time to sort out, that when you have solution – when you have partial differential equations, which govern many physical phenomenon, the solutions often appear in the form of convolution with a special solution, so-called fundamental solution, with the initial conditions, with the initial data. It's something you should expect to see.

This is a pretty big major secret of the universe, all right. So take that to heart, you'll see this. Be afraid, be very afraid. No, no, show no fear. It's what you should expect.

And with that – pretty impressive. And as I say, this is also part of your intellectual heritage, all right. You should know this solution, know this approach to the problem. It's a very famous problem. It had all sorts of far reaching consequences. It should be part of your soul.

I believe with that we bid adieu to Fourier series, although as I say we'll come back to it from time to time. And actually, maybe if not right this day where we bid adieu to it because right now what I wanna do is talk about the transition from Fourier series to Fourier transforms.

And that is the transition from periodic phenomenon to non-periodic phenomenon. So I wanna make a transition from Fourier series to Fourier transforms. And this is the transition from periodic phenomenon to non-periodic phenomenon.

Now, I've said before, and you'll hear me say it again, we have other than make a lot of choices in this classes, and choices for how to cover the material. This is not the only way of doing it, all right.

In many treatments of the Fourier transform, you don't make this – you don't do it this way. That is, you don't start with Fourier series and then try to make the transition to Fourier transforms. It's just the Fourier transform is presented as sort of a [inaudible]. God and the machine, or whatever. It's just there, and it's justified by its many uses and its important applications. That's fine, and that's quite justifiable.

I didn't wanna do that because I wanted to show you the basic phenomenon associated with periodicity. It's an important enough topic, and I wanted you to sort of have some of those things – I wanted to try to cultivate your intuition a little bit for those sorts of ideas. But you don't have to do it this way.

If you'll look at Bracewell's book, which is a very common and popular book that's been used for this course – Ron Bracewell just passed away, actually, was a professor in the electrical engineering department here for many, many years. His book starts out with the Fourier transform, with no mention of Fourier series. And then later on, it actually recovers some of the ideas of Fourier series based on the Fourier transform. And you can do that, but that's a choice, and I have made a different choice. That's all that's involved here.

Now, the transition from periodic to non-periodic phenomena and where we're gonna accomplish that is to view a non-periodic phenomena at the limiting case of a periodic phenomenon as the period tends to infinity, all right.

So we'll do this by viewing non-periodic function, and we'll say phenomena. So non-period function as sort of a limiting case of a periodic phenomenon – periodic function as the period tends to infinity, period tends to infinity.

It's a little tricky to do this, actually. It's not completely – it doesn't happen completely automatically. It takes a little work. It's not a completely automatic process.

Now, the other thing to realize is there are actually two – I'll do it over here – there are actually two aspects to this. So once again, let me remind you of the Fourier case – Fourier series case. There are two aspects to the Fourier transform, really. Two aspects, there's analysis and synthesis.

So the analysis is the Fourier series is forming – so again, if F of T is periodic, the you have the Fourier coefficients, the interval from zero to one, E to the minus two pi IKX, F of RKT, F of TDT. That's analysis. That's analyzing the function, the signal into its constituent components. Figuring out how much each complex exponential contributes to the whole by this much, this amount.

Then there's synthesis. The synthesis is writing the series. We're covering the function from its constituent components. Sum from minus infinity to infinity, F out of K, E to the two pi IKT. So that's synthesis. And both of those things generalize to the Fourier transform.

The Fourier transform is the generalization of the Fourier coefficient. The inverse Fourier transform is the generalization of the Fourier series.

For a transform is a generalization – or the limiting case if you want to think about it that way. Generalization that is limiting case in the sense that I'm talking about here is the period tends to infinity case of the Fourier coefficient. So that's the analysis decomposing a signal into its constituent parts.

The inverse Fourier transform is a generalization, or the limited case in the Fourier series. It is a limiting case of a Fourier series. That's the synthesis part of the equation, so this is part of the discussion.

Now, so how do I set this up in order to take a limit? So I keep saying that it's the limiting case is the period tends to infinity, so that means I have to give you the setup for Fourier series, for Fourier coefficients and for Fourier series, when the period is not one, but the period is some other number, capital T, that I can let tend to infinity.

So we need a setup when say F of T is periodic of period T. Then also I want to let T tend to infinity. I want to let T tend to infinity. How do I do this?

Well, you read about this, I hope, I expect. Let me tell you what the formulas are. It's not hard. The building blocks for a signal of period T are complex exponentials of period T, so that's E to the two pi IKT over T.

Or I'm actually gonna write it a little bit differently. I'm gonna write it as E to the two pi ITK over T. It doesn't matter. I just switched the parenthesis there. Same thing.

These are periodic of period capital T. And the corresponding Fourier series – the Fourier series are the form CKK going from minus infinity to infinity, C, sum K, E to the two pi IKX – KT – K – [inaudible] write like this. K over TT, or K little T over capital T. I'm gonna write it like that for reasons you'll see in just a second. That's what the Fourier series looks like.

What are the coefficients? Well, again, one can repeat all the arguments that we did when we were working with functions of period one, where actually you can use the results for the period one case to derive a general result.

At any rate, what you find is C sub K is given by one over T, times the interval from zero to T of either the minus two pi IK over T, little T, F of T, DT. Did I say here anywhere that F is little, F is periodic of period T? I guess I didn't say that, so sorry, let me say it now. F of T here is a given signal, so I'm assuming it is periodic of period capital T. So F of T, periodic of period capital T.

That's the formula the K Fourier coefficient when the function has period capital T. And actually, you use this again in the homework problem.

Okay. Now, one other thing, I'm gonna write this a little bit differently again for reasons which you will see in a moment because I have in mind – taking a limit here as capital T tends to infinity. I can also write this the function's periodic of period T, doesn't matter what interval you integrate over.

If I know the function on an interval of length T I know it everywhere, so I can also write this formula – and again, this is discussed a little bit more detailed in the notes – I can also write the formula as instead of integrating from zero to T, I can integrate from minus T over two, to T over two as symmetric interval from negative to positive. Same thing, either the minus two pi IK over capital T, T, F of TDT. Okay, fine.

Now, this is really no different than what I've done before, except I've written things a little bit more generally. Instead of a function of period one, I have a function of period one, I have a function of period capital T.

But the formula for the Fourier coefficient is perfectly [inaudible] to what I had before, and the formula for the Fourier series is perfectly [inaudible] to what I had before. So again, that's analysis, that's synthesis. You analysis the function into its constituent parts, and then you synthesis it by forming the corresponding sum.

Now, I wanna point something out here that's very interesting. How would you draw a picture of the spectrum of these cases? What's a picture of these things? What's a picture of the spectrum, picture in the frequency side, picture of the spectrum of frequencies?

Well, in the case of period one – let's take period one. Then the Fourier coefficients are given by the usual formula we had before, and you might draw the spectrum like this. There's a coefficient of zero, that's the zero coefficient, C zero. Then there's the – well, let's see.

Zero, one, two. I have frequencies at all the integers. Minus one, minus two, and so on and so on. Three, minus three. And there's space, so to speak, the harmonics, the frequencies of space are one apart. Here is like absolute value of C zero. Here is absolute value of C one. Here's – I say absolute value because they're actually complex numbers. The coefficients are complex numbers, so I can't actually plot them.

But I can plot – I can get a picture of the spectrum by plotting the absolute value. Here's the absolute value of C two, and here's the absolute value of C three, whatever it is.

And what will I get on the negative side, what will I get on the negative side? On the negative side I get the same absolute values because of the symmetry relation, C minus K is equal to CK bar. So the absolute value of C minus K is – the magnitude of C minus K is the magnitude of CK bar, and the magnitude of a conjugate of a complex number is the same thing as the magnitude of the number. So this is also the magnitude of CK.

The picture's the same on the left, so this is the magnitude of C minus one, the magnitude of C minus two, the magnitude of C three, whatever. Magnitude of C minus three, whatever it looks like. That's the picture.

And in fact, if you've ever worked with a spectrum analyzer, and if I have the chance I'm gonna bring one into class, you see pictures like this. You see a signal and you see these bars that are at the different frequencies, okay.

Now, the important thing here is – the reason I mentioned this is they're spaced one apart. Because a period one – the frequencies are also one apart, one, two, three, and so on and so on.

Now, what about for a function of period T? So the spacing here – the spacing of the frequencies is one. The spacing of the frequencies is one. They're one apart if you were gonna draw the pictures.

Now, what about if you had period T? What about if I had period T? Well, what is the picture there? The picture is like this, the Fourier series looks like this; the Fourier series for a function of period T looks like the sum from minus infinity to infinity, C sub K, E to the two pi I, K over T times T.

So the harmonics, they're indexed by K, but they're periodic of period T, so really, the harmonic you're interested in is in some sense tagged by K over capital T. So zero, one over T, two over T, three over T, minus one over T, minus two over T, minus three over T, and so on.

So if you were to draw a picture of the spectrum that would correspond to a series that looked like this, you would draw a picture that went something like this; you would draw a zero, so this is now period – so this is for period one.

For period T, you would draw the first harmonic sort of at one over T, the second harmonic is two over T, the third harmonic is three over T, and so on and so on.

Then this minus one over T, minus two over T and so on, the spacing is one over T apart. And here's the zero Fourier coefficient, here's the first Fourier coefficient, here's the second Fourier coefficient, here's the third Fourier coefficient. The ones on the negative – corresponding the negative frequencies have the same – when I say here's the first, here's the second, here's the third, I mean the magnitude because again, you can't plot complex numbers. You're just plotting the magnitude.

And on the left, I have the same picture because it's symmetric. So it's like this, like this, like that and so on. Minus three over T and so on. They're spaced one over T apart. But the spectrum has spacing one over T.

Now, this is another example – remember early on when I first talked about frequency and wavelength, and I talked about the inverse relationship, or the reciprocal relationship between frequency and wavelength, very first day of class. That's something you saw a long time ago.

This is our second related example of an inverse or reciprocal relationship between, in this case, I'm gonna say the two domains. How the function appears in the time domain, and how the function appears in the frequency domain. How this is a function of time, and how it is in terms of its constituent frequency parts.

If the period is T, then the spacing and the frequencies is one over T. There's a reciprocal relationship here between the period and the frequencies. Between the period – that's often – and what's happening to the function in time, or what's sometimes referred to as the time domain of the function, and the frequencies between the period and the frequencies.

Or viewing the function in the frequency domain, viewing it in terms of its constituent parts. Reciprocal X of a period T means frequency of spacing on the frequency is a one over T. Frequency spaced one over T apart.

All right. I didn't say anything about the size of T here, but that's a general reciprocal relationship between the two domains. And again, it's something you're gonna see throughout this course. So a reciprocal relationship between the two view of the function. And that reciprocity is exactly mediated by Fourier techniques, Fourier series in this case, or very soon the Fourier transform.

But again, it's the kinda thing – this is the sort of intuition you have to start to develop. You're viewing the function one way, you expect certain things. You view the function in the other domain you expect the reciprocal phenomenon.

Now, if T is less than one, and one over T is bigger than one, so T less than one. So a function which repeats more frequently than once a second implies larger or smaller spacing than one, implies one over T is bigger than one, so the spacing is larger. Spacing – one over T is bigger than one, spacing bigger than zero, bigger than one.

The spectrum is spread out. If T is bigger than one, if you have a long period, the spectrum is compressed. The spacing is one over T. One over T is less than one, so the spacing in the spectrum is squeezed, compressed.

In particular, as T is going – as capital T is going to infinity, which is the case ultimately I wanna deal with, the spectrum is getting more and more – the spacing of the spectrum is smaller and smaller. The frequencies are getting closer and closer together. One over T, two over T, three over T, four over T. They're getting closer and closer. One over T is getting smaller and smaller, and they're getting closer together. As T tends to infinity, the spectrum becomes sort of continuous.

I'm gonna make this more precise, or rather, I'm gonna make this more explicit if not more precise in just a minute. But the idea is the spectrum's getting closer, the frequencies are getting spaced closer and closer together because the spacing is one over T. The spacing is one over T, and if T is tending to infinity, the spacing is tending to zero. That's what I mean by the fact that the spectrum is getting continuous, so the spectrum is getting squeezed. Now, that's the formula here for the coefficient here once again.

Now, let me just start now – no, one thing at a time. So let me – once again, let me write down the formula for the coefficient. CK is one over T. This is the definition for the Fourier coefficient when the function has period T. So I'm going to integrate from minus T over two to T over two, E to the minus two pi I, K over T with a T, F of TDT.

Now, I wanna let T tend to infinity here, and use this as a way of passing from periodic to non-periodic phenomenon, and use this to pass from periodic to non-periodic.

But as I say, it's not quite straightforward. And let me tell you why. I can't just take the limit as T tends to infinity there and get – and the Fourier transform pop out of that. It doesn't work. You can't just let T tend to infinity and get the Fourier transform. You have to tickle it a little bit.

And actually, let me leave this picture up on the board and show over here, all right. Let me tell you what the setup is gonna be and what I wanna do.

So imagine I have some function. The picture's not periodic. But suppose it's fine, I can extend. So suppose F of T looks like this. Sum [inaudible] sum, sum interval going from A to B, and it's zero beyond – less than A and bigger than B, all right. That's my function.

So I take some big number, T, and I periodize this, say bigger than – so the minus T over two is less than A, and plus T over two is bigger than B. So here's my function; it's zero, less than A, it's zero, bigger than B. I take some – and I wanna approximate this thing, but I wanna imagine this is a periodic function. So I take some big period beyond where the function is zero, and I periodize it to be period T, all right. Take a big T and periodize to have period T.

Okay, fine. Now, write down the formula for the Fourier coefficient. So imagine, if the function here were fixed, if that's all I worried about, I let T go to infinity, then I'm sort of approximating that non-periodic phenomenon by a periodic phenomenon with a very big period, all right. That's gonna be my goal.

But the problem is – and I wanna see what happens to the Fourier coefficient. So write down C sub K. The Fourier coefficient looks like this, C sub K is one over T, the interval from minus T over two to T over two, E to the minus two pi I, K over T, little T, F of T, DT, okay. That's the K Fourier coefficient.

But now, F is zero, less than A and bigger than B, and those numbers are fixed. Those are sort of given to us. So this is equal to one over T times the interval from A to B. E to the minus two pi I, K over T, little T, F of T, BT. Because the function F, I'm assuming, is zero, less than A and bigger than B, okay.

Now, this interval in absolute value is gonna be bounded. The interval from A to B, E to the minus – an absolute value – E to the minus two pi I, K over T, little T, F of T, DT, the absolute value is less than the interval and the absolute value. This is less than or equal to the interval from A to B, the interval of the complex – the absolute value of the complex exponential, which is one, times the function DT. The absolute value of a complex exponential is one. I'm almost there, almost there.

So this is just equal to the interval, this absolute value. You'll see why I'm doing this in just a second, really, really, really. This is equal to the interval from A to B, F of T, DT because the absolute value of the complex exponential is just one. That's a fixed number. It's like M.

Okay. So what does that say about the Fourier coefficient? Watch what I did here. I just wrote down the formula for the K Fourier coefficient. What I wanna see – I wanna convince you there's a little bit of problem with directly letting T tend to infinity. If I directly let T tend to infinity, everything is gonna die.

Here's the K Fourier coefficient. It's one over T times this interval. What about that interval? That interval is that interval, and this interval in absolute value is bounded – say goodbye to this drawing now.

So this says that in absolute value, C sub K is less than equal to one over T times M for all K, all right. So as T tends to infinity, C sub K tends to zero. The Fourier coefficients die.

So I had this wonderful idea. I said I'm gonna approximate a non-periodic phenomenon by a periodic phenomena, a very large period, and I'm gonna let the period tend to infinity. Sounds great. Sounds like a very natural thing to do.

I write down the formula for the Fourier series, great. I got these coefficients for the Fourier series, great. I'm gonna let T tend to infinity, great. The coefficients are gonna tend to zero, not great. I'm not gonna get a formula that's gonna help me as T tends to infinity. Not great.

I think I have to quit right now. Tomorrow, on Friday, I will tell you how to save this in a very nice, easy way that's gonna lead to everything.

[End of Audio]

Duration: 53 minutes

TheFourierTransformAndItsApplications-Lecture06

Instructor (Brad Osgood):Okay, we're on.

First thing I want to say is that I made a little mistake last time in lecture. It was gently pointed out to me. When I was talking about the heat equation; the floorshow was fine, the discussion was fine, but then I said this thing about as T tens to infinity, the temperature tens to zero. That wasn't right. I forgot about the zero – because somebody said look, the fusion man, you don't lose anything, it just diffuses. So it's not right to say the temperature changes to zero, I was thinking of while it was escaping to the universe or something like that.

But in fact, what happens is that as time tens to infinity – I forgot – technically, mathematically, I forgot about the zero free coefficient, which is in the solution, so what happens is that as the tens to infinity, the temperature tens to the average temperature of the equated natial distribution of the initial distribution. So I'm going to write that up and post that on the website because I cannot stand the idea that this error is left unfixed. I meant to post that already, I just hadn't had a chance to do that, but I wanted to make sure that I made that formal apology and announcement. All right?

Any questions? Anything on anyone's mind? Other than that, any other corrections?

That's a relief. All right.

All right, so today – do you have your hand up or are you just sort of resting? Oh, no, okay.

Last time, we had gotten right to the verge of discovering the Fourier transform as a limited case of Fourier series and I just wanted to pick up where we left off and finish that off and then launch into a more formal treatment of Fourier transforms and talk about how we're going to proceed.

So we are about to get the Fourier transform as a limiting case of – we'll actually there's two aspects to it. There's a limiting case of the Fourier coefficient and the Fourier series; the analysis part and the synthesis part of the Fourier coefficient and Fourier series.

So when I talk about the Fourier transform, I'm sort of thinking of the two things together, but there's really two parts to it and they both come out of more or less the same analysis.

If I solely remind you what the set up was and who and why and what I mean by a limiting case, by I mean a non-periodic phenomena I want to model as a periodic phenomena as the period tens to infinity. All right?

And I took a special case of this, or just a case for illustration, that is, I suppose I have some function which dies off eventually so it's zero outside some interval, now that's –

ultimately we're gonna drop that restriction, I'm just taking this as a case to sort of model what the situation should look like. So I'm going to take the case and, again, this was a set up from last time. So take a case that looks like this. Have some function as F of T, which is zero outside some interval. I'm drawing it like it's positive but it can be very general, but the fact is that it dies off at some point, and then I periodize it so – you can image it's a very big interval to begin with – but finite. And the I periodize it – I take an even bigger number – Capital T, and I look at it from say minus T over T to 2 over T, that's supposed to represent one complete period, and then periodize this function. So the pattern repeats. All right?

So periodize to make periodic of period T. So think of T as big, and eventually we think of T as going off to infinity. So I won't draw the picture, but again, the idea is I just take the same pattern and repeat it over and over again.

All right then, you can write down the formula for the Fourier coefficient and you can write down the formula for the Fourier series. So what the series looks like – the Fourier coefficient looks like this: CK is 1 over T times the interval from minus T over 2 – instead in integrating from 0 over T, I integrate over any period, so I take the period from minus T over 2 to T over 2 and then the formulate is either the minus 2 PI I K over T, T, that's how I wrote it last time, FOTDT and the Fourier series, that's the analysis part, has decomposing F into its components and the complex – the corresponding complex exponentials are these.

And then the Fourier series is to recover the function from its components as a sum from minus infinity to infinity, C sub K of each of the 2 PI I, K over T, times T. All right?

And what you would like it you would want to take a simple-minded approach of saying a Fourier transform inverse Fourier transform and so on, is a limited case of Fourier series as you let the period ten to infinity you just let T ten to infinity but that doesn't quite work.

All right, so you would like to just let T ten to infinity and low and behold, the formula has emerged, but it doesn't work. Doesn't quite work. All right?

So instead, it doesn't work because you don't get anything. That is to say the Fourier coefficient tens to 0. Doesn't work because if TK tens to 0 as T tens to infinity.

All right, so what you do is a little clues job here, it tens to 0 like 1 over T, that's what I talked about last time and I won't repeat that argument. Because of that factor 1 over T in front, this interval is bounded if the function is fixed and again, if the function is 0, non 0 only in the interval from A to B, that's ultimately not getting very big, so it goes to zero like 1 over T so you scale up by T. All right?

Scale up by T. And what I mean is, I want to consider instead of T over 1 times the interval, I just want to consider the interval and I want to view that as a – to anticipate what's going to happen as a transform version of the function, not evaluated at K, although the indexing here is on K, really it's K over T that's the important thing, all right. Again, I want to let T ten to infinity so I want to emphasis that connection.

So I want to look at let me just use this notation, let me say write – F of F, add K over T, all right, so that's a new notation. That's something I'm introducing to indicate this sort of scaled interval as the interval from minus T over T over 2, E to the minus 2 PI I, K over T, on to T, F of T, DT. And the Fourier series here has to incorporate that 1 over T again because that does come into the coefficient so the Fourier series looks like, F of T is the sum from – I'll write it like this – K goes minus infinity to infinity, this transformed this coefficient K over T times the complex exponential E to the 2 PI I, K over T, times T times this factor 1 over T. This factor 1 over T is coming in there again because that occurs in the definition of the Fourier coefficient that I have to have that in there. I mean, the Fourier series is in terms of the Fourier coefficient and the 1 over T is in there. Okay?

Now I want to let T ten to infinity. This is all heuristic, all right? This is not a proof, it's an argument for what the formula otta look like as a limiting case of the Fourier series, of the periodic case.

All right so now, let T ten to infinity, and you have to make an argument as to what the formula should look like.

All right I said this last time; I want to say it again. The idea is that as T tens to infinity these numbers, K over T, of course for a fixed K, that's tenning to 0, but the idea is that K is also going from minus infinity to infinity and what's happening here is that, if you think of K over T as a discrete variable, it is getting – it is approaching a continuous variable. All right?

The space keeps getting closer and closer together. They're spaced 1 over capital T apart. One over T, 2 over T, 3 over T, 4 over T and as T is tenning to infinity, they are spaced closer and closer together so the discrete variable is approaching for all to see, a continuous variable which I'll due note by S. All right?

So the discrete variable that's not a technical term all right. I'm just reasoning and futuristically here – tens to or placed by, in the limit, a continuous variable S. And S is going to range from minus infinity to infinity. All right?

Fine. That is to say, this formula, in the limit as T tens to infinity is going to be replaced by another formula.

So you write, FO, F of S, as the interval from minus infinity to infinity, E of the minus 2 PI I S T, so K over T again is being replaced by the continuous variable S, F of T, D T.

But don't stop there, also looks what happens to the Fourier series as capital T tens to infinity and for the Fourier series, which is here, you have to recognize this as a discrete – as a sum approximating interval. All right? This is a function of evaluated at the variables K over T, K over T here and as T tenning to infinity these are being – there are pushing a continuous variable. The 1 over T here is like the Delta S, you know that comes in riding an approximating sum into an interval.

So as T tens to infinity, what happens to the Fourier series, it is replaced by an integral. Replaced by the integral from my – replaced by this formula; replaced by F of T, use the interval for minus infinity to infinity, E to the plus, 2 PI I over – I'll write it like this, I'll keep the same order of the terms – the Fourier transform of F, or the – this thing which I'm going to now call the Fourier transform, in just a second – F of – this, the Fourier transform of at S, either the 2 PI I S T D S. Okay?

Now, I really feel like the clouds otta open up at this point because something really, very momentous has happened here, if only by analogy. All right, if only by a heuristic argument saying you want to view a non-periodic phenomena as a limiting case of a periodic phenomena, this is one way of doing that, that leads to something that I'm going to spend the rest of the quarter convincing you is a good thing.

All right let me say a little bit more here, again this – this little journey here – has been a way of making the transition from a case that we studied to a case that we haven't studied. The final step in the process is to declare victory, that is to say, and this is what happens in mathematics all the time and this is what makes it very frustrating for people when they're reading a mathematics book to try to figure out what the hell is going on and where the hell do these formulas come from and so on and so on.

You sort of erase all paths of discovery and you just declare – what do you know, I'm going to give the following definition – how about that? And that's exactly what happens, so that's what we do now. We sort of declare victory.

I wanted to – I didn't want you to miss the steps in between. There they are let me write it over here, I'm making our victory blackboard over here. I didn't want you to miss the steps in between because I actually view a very important part of this course is not just going over formulas and facts, but in trying to give you a certain feeling for how the mathematics develops in the context of these problems. All right because, if you sort of get used to thinking that way, it will give you a much greater power over the problems that you're likely to confront out there where you really have to apply mathematics using your own head, using your own thoughts.

All right so what I'm describing to you is just the sort of process that you got through that ultimately results in a definition. All right, it ultimately results in the definition but the steps along the way are often hidden and I wanted not to hide them.

All right, so we define – at this point I usually say, the clear victory – so if F of T is a function defined on a whole real line, from minus infinity less than T less than infinity, you define its Fourier transform by the formula that I just wrote down.

The Fourier transform at S is this interval – minus infinity to infinity, E to the minus 2 PI I S T , F of T, D T. All right?

So here, S also is a real variable going from minus infinity to infinity but the Fourier transform itself is complex value because I'm integrating a function against a complex exponential here.

I didn't say whether or not F had to be real or complex, as a matter of fact, in general, I don't want to make that a – I want to allow either case. All right? So F can be complex in many application because of course, F is a real signal and that's fine, but it makes sense, as far as the definition goes, to allow F to be either real or complex. I won't say anything more about that.

Now there's a lot – of course, there's a lot more to say about the definition. One thing that should be stated right up front and something I'll say more details about later is, of course, this definition is only good if the interval makes sense. All right?

To write down this interval, saying this is the Fourier transform, but you, if you're going to actually carry out this innovation for a particular value of S, you know, you have to say something about when the interval converges. All right?

So you need to say something. We'll need to understand the conversions of the interval.

Just like we needed to understand the conversions of the Fourier series, at least to a certain extent we also need to understand the convergence of the interval that defines the Fourier transform.

All right, and that's an issues, right?

But not only have we been led to the Fourier transform, we have also been led to the Fourier inversion. This is analysis – all right. The Fourier transform analyses the Fourier signal, the non-periodic signal into its component parts. What are the component parts? The component parts are a continuous family of exponentials. Not a discrete family of complex exponentials but a continuous family of exponentials that equal the 2 PI I S T. All right?

The Fourier transform analyses F of T into its constituent parts.

But now we also have Fourier inversion. Fourier inversion says that we can synthesize the function from its constituent parts. And that's the second formula there. That says that if you know the Fourier transform then you can get back the function. That is, you have F of T, the signal equals the interval from minus infinity to infinity of – let me write it like this – either the 2 PI I S T, the Fourier transform at S, D S.

You often think of T as the time variable and you think of S as the frequency variable and you think of the function defining the time domain and the Fourier transform defining the frequency domain.

You can think about it that way but you don't always think about it that way, and I'll come back to that as well.

All right, so you think of generally F of T in the time domain, the Fourier transform of S in the frequency domain that is to say you think of T as a time variable, you think of S as the frequency variable that's fine but don't be weighted to that so completely that you're not willing to change your prospective.

All right, the Fourier transform is a very flexible tool, it comes up in a lot of different contexts, T is not always time, S is not always frequency and you do yourself no favor if you force yourself into thinking only in those terms. All right? It's good for many applications but not for all applications.

You will hear me say a lot, you will hear me rant, you know, about notation and sort of convention and things like that because this subject is fraught with difficulties as far as notation goes, as far as interpretation goes. Part of that is just because the richness of it. All right, it's a very rich subject and any rich subject can be abused in various ways, all right. And this is – I would say not only no exception to that, maybe a real exemplar of the abuse that can be heaved onto different symbols and their interpretation.

Okay now, I could summarize what I just said actually into what I consider a major secret of the universe. Perhaps the most major secret of the universe that you will ever lean in your lives, certainly in this class, is that every signal has a spectrum – you call S, the frequency domain or you call the value of the Fourier transform the spectrum. All right? And it's a term I'm sure you're familiar with.

So if I can summarize this, as [inaudible] major secret of the universe; probably soon to appear on YouTube all around the world. Is that every signal has a spectrum and the spectrum determines the signal.

All right, to say that every signal like all secrets of the universe, this is the paint of a little too broad of brush. All right? To say that every signal has a spectrum is, you can take Fourier transform, but of course, there are issues there. Does the interval really converge and so on. All right, to say the spectrum determines the signal means that you can always invert it like this, it means in particular that this interval always exists. All right, and that's – there are also issues associated with that. But never mind that. Let's just concentrate on the majesty and really, the enormous applicability and truth of this statement, all right? For most cases, and for the cases that you're certainly interested in, this or some version of this is true and can be a guide to happiness. Okay?

The Fourier transform, the analysis and syntheses of a function are two ways of seeing the same thing. You can look at the function in the frequency domain, you can look at it in terms of its Fourier transform or you can look at it in the time domain. You can recover it from its spectrum.

All right, the two different representation of the same thing and if you have two different representations of the same thing, you have tremendous power over it. All right? They're equivalent. Knowledge of one is equivalent to knowledge of the other.

You will not get anything more profound, I think in any class, anywhere, anything. How about that? That's a way to start the weekend off.

All right now, let me introduce one other bit of notation now so I'll have a chance to use it although I'm not going to make much more use of it today, but we'll certainly make some use of it later, and that has to do with this sort of inversion formula here.

That is, it pays to introduce a separate operator along with the Fourier transform, namely the inverse Fourier transform, which is defined in a very similar way except for a change of sign. So it's useful – so again, the Fourier transform F at S is equal to the interval for minus infinity to infinity of either the minus 2 PI I S T, F of T, D T, all right. That's sometimes called the forward Fourier transform. Again, I won't take up your time.

It's also useful to introduce the so-called inverse Fourier transform and let me call the function G. As the interval from minus infinity to infinity of either the plus 2 PI I S T, say G of S D S. All right, now be careful about how the variable – then this result, the fact that you can recover a function from its Fourier transform asserts that this really is the inverse operator of this.

All right, so Fourier inversions says – Fourier inversion says that the inverse Fourier transform for the Fourier transform of a function is the function. And it also says, for that matter, if I take the Fourier transform of the inverse Fourier transform of a function, I get back the function. Okay?

Now again, I don't want to get too far on a rant now, but, just to start, or just to give you a little warning of the things to come. You have to careful how you look at his, all right; the role of the variables here. This – take a look at the definition of the Fourier transform, all right, you're integrating either the 2 PI S T that depends both on S N T against the function of T, you're integrating with respect to T, what remains as a function of S. Right? That's why I call – use the notation, the Fourier transform of F at S. So it's an operation taking the Fourier transform evaluated at the point S.

All right, the operation is carrying out this integration but in order to write that down, you have to evaluate it at a point. So you're evaluating at a point S, it's given by this formula.

Likewise, for the inverse Fourier transform, I'm integrating either the 2 PI I S T, that's a function of two variables, S and T. If I integrate against the function of S, what remains is a function of T. It's inverse Fourier operation, to carry out the operation I have to evaluate it at a point and the variable to use here is T because I'm integrating with a function of S and T against the function of S integrating with respect to S. All right?

So note two values here actually.

All right notice, that the Fourier transform at 0, it is the operation of taking the Fourier transform and then as the operation starts [inaudible] of evaluating at a point, so the Fourier transform of F at 0 is the interval for minus infinity to infinity, E to the minus 2 PI I 0 times T, F of T, D T. I plug in S equals 0 into the formula.

And so that's just the interval from F E average value of the function, or at least what you consider the average value of the function, F of T D T. There's no – you don't divide by the length of the interval because the interval's infinite. All right, but it's the interval of the functions of the area under the curve if you want to think of it that way. All right?

So the 0 Fourier, the value of the Fourier transform of 0 is the interval of the function. This is analogous to the 0 Fourier coefficient being the average value of the function. The interval over one period. Here, I don't have a period.

And likewise, by the same token, the inverse Fourier transform of a function at 0, G at 0, is the interval for minus infinity to infinity. I integrate the Fourier transform times E to the 0 – I'm not writing at this time, I'm just giving you a final answer – so the interval for minus infinity to infinity of the Fourier transform, that gives you the inverse Fourier transform at the origin. Okay?

You have to always tell yourself, and be clear, what the role of the variables are in these expressions. Trust me, it becomes an issues. When the formulas get a little bit more complicated, as they will, how the variables are used and keeping that straight is something that you have to be careful about. Something you have to be careful about.

Okay now, so we've gotten to the stars of the show. The Fourier transform, the inverse Fourier transform, and the idea of Fourier inversion. It is now the – my responsibility for the rest of the quarter to convince you that this was worth it. All right?

That is, that these really are useful operations to consider and that they can tell you much that is worthwhile in, certainly in applications. But not just in application, the Sirline applications.

I want to tell you how we're going to proceed. Actually, we're going to proceed to develop the ideas here in a way very much like, in a path, very much like the one you followed when you were first learning calculus. All right? You cast your mind back those happy days, when the world was new; calculus was just an attraction on the horizon. When you're learning calculus, the path you follow – we learned about derivatives and intervals, all right; two basic operations of calculus. And the way you did it was you learned general – you learned specific formulas like derivatives of exponentials, derivatives of polynomial derivative functions. You used intervals of corresponding intervals and then you learned general properties. You learned how to differentiate, you learned the product rule, the chain rule, integration by parts, you learned specific formulas that are going to come up, often enough that you want to know how to differentiate a specific function.

But functions come up in combinations, products, quotions, compositions, so you had to learn, also, general rules for taking derivatives. You had to learn the product rule, the chain rule and so on.

And then, of course, you learned the applications of derivatives and intervals. All right?

Now also in connection with the Fourier transform, let me call your attention to a fact, to a problem or a challenge you face in calculus. The interval's a very rich operation. I'll come back this and I'll say this again later, but when you first learn about the interval you learn the certain interpretation of the interval. You learn that usually, as some sort of motivating [inaudible], the interval's the area under the curve, or you recover the total change in the function from its rate of change or so on.

But the interval is a very rich concept and you do yourself no favor by clinging to one particular interpretation of the interval in all case because maybe that interpretation won't really be a guide, or won't really apply.

Well again, as I was just saying a minute ago, that's the similar sort of thing with the Fourier transform and the inverse Fourier transform. It has certain interpretations that the variables often have certain interpretations, time or frequency. But you do yourself no favor if you cling to those particular interpretations and try to impose them where they don't necessarily belong.

However, we're going to follow a similar path to the one – in or however, but – analogously we're going to follow a similar sort of path when we develop properties of the Fourier transform.

That is, we're going to need to develop specific transforms. Transforms of specific signals. The kind of signals society needs. Right? That society runs on.

And then we're going to develop general properties of the Fourier transform. That is, what to do when signals are combined in certain ways. Actually in the way they're combined is a richer – in many cases – the richer operation, the richer set of operations than the ordinary functions of – that you work with in calculus, the ordinary rules of calculus.

Then, of course, we're going to talk about a lot of applications of these ideas. It's very similar in the pass somehow, to the way that you study calculus. And I say that again; because I hope it give you a way to sort of organizing your own study oven. I realize, at each stage, here I'm learning a specific formula; here I'm learning a general formula, and so on. It will help – and here I'm learning application, and this is the interpretation in this case and this is the interpretation in that case. It will help you, I hope, sort of organize it in your head, how the subject is evolving. Okay?

All right, so let's look at a few – let's start on that path and let's look at a few examples.

Basic examples; that is calculations with a Fourier transform. Now, I'm only going to do a few of these because unlike calculus – when you first learned calculus, you're teacher, I hope, you know, showed you how to do a lot of calculations in a lot of specific cases because all was new. All right, I'm not going to do that now. I'm going to do a few specific calculations to show you the kind of techniques that always come in, or often come in, but you can read the derivations, you can do the derivations, there's no new techniques involved there. All right, the only techniques are integration, either integration by substitution or integration by parts, direct integration, whatever.

That's not new, what's new are the specific formulas that come up and that's something, again, that I am going to leave to you to read, derive, learn, memorize, whatever. All right?

I want to spend a little bit more time on developing some of the general properties because there, there are some new things. There are some interesting things that you haven't seen before, or haven't seen other than in this context.

So the first example I want to take is the very simplest one, one we've seen – you've seen actually in harder problems, we saw actually in the context of the Fourier series – that's the function that models a signal that's on again and off again. Or in this case, since I'm not looking at the periodic version, it's just either on or off. And that is the so called rectangle function. PI of X, this is a notation I'm going to use and I think of it as advocated by Bracewell, it's not bad. I'll explain – actually I'll call it variable T, why not.

I'll tell you Y uses the notations to the second. It is 1 if absolute value of T is less than 1 and it is 0 if absolute value of T is greater or equal to 1.

So it's graph, as defined for all values of T – and here's 0 – excuse me, here's minus 1 and here's the 1. So, it's actually – yeah?

[Inaudible] careful here, no, see I already got it wrong because there's not universal convention here. I'm sorry. I want it to be of Width 1, not Width 2. So I'm going take ½, ½, see, I already messed it up. Damn.

Okay, it has total width 1, so it goes from minus to ½ to plus to ½. It takes a jump and it is discontinuous at the end points. All right? So it's 0 in between the end points and from ½ and minus ½ excuse me, it's 1 between the end points and it's 0 outside the endpoints.

Now, this is not – and it's called PI because PI looks like a rectangle. I think it's stupid, and I feel really, I feel so juvenile saying that, but – and it's also sometimes called the top-hat function because it's supposed to look like a top hat. As it goes up, up, up, it looks like a rectangle. So it's also called, more grandly, sometimes its call the characteristic function of the interval from minus ½ to ½. It is sometimes call the indicator function of the interval from minus ½ to ½. All these terminologies are in day-to-day use. All right, depending on the field.

I would actually tend to call it – in mathematics you tend to call it the characteristic function so I always tend to call it the characteristic function but I won't do that anymore, I'm going to call it the rectangle function.

And there's also a certain amount of debate about how to define it at the endpoints. All right, I define it a certain way at the end points. Other people would define it differently. Other people would have it to be 1 at the end points. Some people would have it have Value ½ at the end points. Because of this idea somehow at a discontinuity you have to look at the average value. This becomes a religious issue with some people. How this function should be defined at the end points. I do not want to get dragged into it. It will never make any difference for anything we ever do. All right, so this is my definition. If you don't like it, go to hell.

All right, so what is the Fourier transform? The only recourse we have in calculating Fourier transform is the definition. All right, there's nothing else to work with. So, you have to carry out the integration. One has to carry out the integration, I should say. So what does one do? One writes the definition down. The integral for minus infinity and infinity, E to the minus 2 PI I S T, PI of S, D, PI of T, D T. Careful about which variable we're using here. I'm sorry, I've already got myself a little crowded.

The interval once again, the interval for minus infinity to infinity, E to the minus 2 PI I S T , PI of T, D T. Okay?

All right now, PI T is only non 0 from minus ½ to ½. So the only thing that remains here in the interval, it's not a infinite interval, it becomes a finite interval because the function is 0 outside that interval so it's the interval from minus ½ to ½ E to the minus 2 PI I S T D T. You integrate that like you integrate any exponential, the fact they're complex number is in there does not matter. The carrying out the integration is the same as ordinary integration in calculus.

So that is minus 1 over 2 PI I S, you're integrating with respect to T. So you're regarding S, which essentially is a constant as far as the integration is concerned. So it's minus 1 over 2 PI I S, either the minus 2 PI I S T, evaluated between T equals minus ½ and T equals plus a ½. All right, that's the only thing you can do. You have to – you have no recourse here other than to use the definition. Let's do it, quickly.

So that is – what do I get? I get, at the top end point, I get minus 1 over 2 PI I S, E to the minus 2 PI I S times ½ minus, minus 1 over 2 PI I S. The only thing you can get wrong here is, you know, like minus signs and things like that. All those problems you used to have problems keeping straight in calculus, well, I'm sorry, they haven't gone away. The same sort of issues are there.

Minus 1 over 2 PI S E to the minus 2 PI I S minus ½, so that's minus 1 over 2 PI I S, E to the minus PI I S, minus a minus gives me a plus, 1 over 2 PI I S, E to the minus minus because we plus, PI I S. Or if you'll allow me to write that differently, that is 1 over PI S, 1 over 2 I, well I guess, either PI I S minus either the minus PI S divided by 2 I.

And I write it like that because you have to start to recognize how to manipulate these complex exponentials to bring in, when called for, the ordinary true functions, signs and cosigns. And the PI I S minus the minus PI I S over 2 I is the sign of PI S. This is the sign of PI I S divided by PI S and that's the formula for the Fourier transform, period. Okay?

This is the most basic example and actually, as simple as it is, it's actually one of the most important examples. It's one of the ones that comes up most often in applications. So write it down.

The Fourier transform of PI, direct single function is the sign of PI S divided by PI S and this function comes up so often in applications it is given a special name, it is called the sync function. So by definition, the sync function of S is sign of PI S divided by PI S.

Now that's a function actually, I'm sure any electrical engineer has seen this function, every electrical engineer, in fact, sees this function in their dreams.

You also saw this function in calculus let me do it over here because it's an example of sign of X over X; a famous limit. You know, what happens if X tens to 0. The limit of sign of X over X is X tens to 0 is 1 and the limit is S tens to 0 of sign of PI S over PI S is 1, so this function is actually quite nice. This function is continuous and even smooth and the graph looks something like this. It dies off; it dies off like 1 over S. It's symmetric, it's an even function. That's a damn good sync function if I do say so myself.

So here's 0, 1, 2, 3, it has zeros at the integers. Minus 1, well, that's not so good I guess, actually. It's supposed to be even. Minus 1, minus 2 and so on. Okay?

Now, you can look in the book. I have a picture of the sync function, I have many pictures of the sync function, but don't deny yourself the pleasure of going down to Sunnydale, driving down to Sunnydale, going to the Fry's in Sunnydale, Fry's I'm sure some of you are new to the area and may not know is this huge electronic store and the Fry's in Sunnydale has an enormous neon sync function. All right, in living color. It's incredible. You go into that store and you see sync functions everywhere. Right, they're on all the railings, they're on the turn styles, I mean, it's a nightmare. All right, so – and there is a picture – I saw it and I couldn't believe it – there is a picture of it in the book. All right?

Every electrical engineer knows this function. Mathematicians think sign of X over X, what is the big deal you know, but in fact, a large part of the world's economy depends on this stupid function.

So go down to Fry's, see it live. I would like to imagine the whole class standing there looking at it. You know, one Saturday afternoon and having the people from the store go out saying, can I help you, customer service, you know. What can I do, are you here to return something?

All right, who has been to that Fry's, by the way? Okay, I'm not making this up right? Yeah, but for the rest of you, oh what a treat you have in store.

All right, let me look at a couple of examples, all right. Because all I want to do, I want to show you that again, you have no recourse other than the definitions so, if you're asked to compute a specific a specific Fourier transform, sometimes there're tricks but more often than not, it's just a question of carrying out the integration. So all those, you know, hard-won skills of integration, by parts integration, but substitution and things like that, those are going to be coming into play when you actually have to calculate specific Fourier transforms.

But again, it like, you know, tables of intervals. People have done this so the, you know, the collective hard-won experience is already out there, but you have to know a certain amount of it. And a certain amount of this I think you have to do for yourself just so you have the confidence in doing the calculations, just so you see how it goes.

So related actually to the rectangle function, is the triangle function, a member of the first homework set that everybody loved so much. The basic triangle function is 1 minus absolute value of T if T is less than or equal to 1 and 0 if absolute value of T is greater than or equal to 1 and it looks like this. The graph looks like this. Called a triangle function because the graph looks like a triangle. It goes up – it's 0 outside the interval for minus 1 to 1 and then it goes up the slope 1 up to a height of 1. Okay? Nothing to it, very simple function.

But it's a very important function in itself. What about its Fourier transform? So the Fourier transform of the triangle function, evaluated S, is again the interval for minus infinity to infinity of either the minus 2 PI I S T [inaudible] of T, D T. Don't skip this step, all right. Write that down so you realize what you have to write next. What you have to write next depends on how the function is defined. The function is zero outside the interval for minus 1 to 1 so it's only the interval for minus 1 to 1 that you have to worry about but the function comes in two pieces. It has a different formula on the interval for minus 1 to 0 and on the interval from 0 to 1. And they have to be treated separately. So this is the interval for minus 1 to 0 of either the minus 2 PI I S T and on the interval from minus 1 to 0, with 1 minus absolute value of T, T is negative there so it's 1 plus T, D T, and then on the interval from 0 to 1, it's either the minus 2 PI S T times the formula for the triangle function on that interval which is 1 minus T. You cannot avoid this. All right, you cannot combine these intervals.

Well, you can do a little bit, you know, but basically you cannot combine these intervals. You have to do the intervals separately.

So how do you do them? Well, I'm not going to carry it out into detail, but I want to say just enough so you get a feeling for the kind of calculations that you have to do. Some of those calculations when you're calculating Fourier coefficients, there're only a few techniques that are in play here. All right, and in this case, the technique is integration by parts. All right? That's what comes up.

So let me remind you the formula for integration by parts. Integral from A to B of U D V. This is everybody's favorite. It is UV, sometimes what you don't remember is what integration by parts looks like for definite intervals, so the interval of U V D is U V divided by between A and B minus integral from A to B of V D U. We all know what that means. What that means is that in your integrand, you're original interval, you have to identify what the U part is and what the DV part is and I'm not here to teach you calculus, but that's what you have to do. All right?

So what do you do in this case? So let's just look at the first interval, then we'll wrap it up. So let's look at the interval from minus 1 to 0 of 1 plus T times E to the minus 2 PI I S T , D T. So looking at this you have to decide, you have to look deep in your sole and decide what is U. What is DV and what are the consequences of that. All right?

Well let me tell you. All right, U in this case, you take to be 1 plus T, and DV you take to be E to the minus 2 PI I S T, DT. All right, you do that because you look ahead, you are no fool and if you take U equals 1 plus T it's going to simply on the other part of the interval because then D is just going to be DT. DU is DT, and then if DV is equal to E to the minus 2 PI S T, V is equal to, you integrate with respect to V that means you're regarding S as a constant so it's minus 1 over 2 PI I S, either the minus 2 PI I S T.

I love to integrate. I spent a lot of weekend home alone. You know, just integrating.

It's a great ice breaker at parties. Do you like integration by parts? I do.

So if you do this, then it becomes – then the interval from minus 1 to 0 of 1 plus T, T to the minus 2 PI S T, D T becomes U V, that's 1 plus T times minus 1 over 2 PI S either the minus 2 PI I S T, evaluated between minus 1 and zero, all right? Minus the interval from minus 1 to 0 of VDU, which is V is minus 1 over 2 PI I S, P to the minus 2 PI I S T and DU is just DT. All right, so that integral has become simpler because now it's just an integral of a complex exponential. You're going to expect a T; this S is a constant out here, okay? That S is a constant out there. I have this right don't I? Yeah.

Okay now, I'm not going to take it from here. I love to integrate, but not public, so I will not carry this out any further but let me tell you what happened. You can, all right, and you probably should. I think it's carried out a little bit more in the notes. You have to do the same thing, same thing for the other integral. The two integrals, you cannot avoid this. All right, tough break, but it's just the way it is. You can carry out this integration and what do you get? So you get, it's quite a nice formula actually. You get, at the end of the day, it. That the Fourier transform of the rectangle function is sign squared of PI S divided by PI squared, S squared. That is to say, it is sync squared of S. Isn't that cool?

The Fourier transform of the triangle function is sync squared.

Now in fact – so once again, Fourier transform of the triangle function of S is sync squared S. All right now, that's its sort of fun if you like that sort of thing, to see that emerging from this picture, but I'm not going to do that. I'm not going to do that. I'm not going to carry it out.

In fact, we will, next week, actually see a different reason why this is true. I mean, it's actually not a shock. There's a deeper reason why this is true, having to do with convolution. But that's a little bit in the future, it's just but it can be done just by straight calculation, by no more than that. And again, for many Fourier transforms, for many basic functions that you need to know that you work with and you need to know the answer for, you have no recourse other than to carry out the integration to actually do it. All right?

Again, you should do a certain amount of that, it's one of these things where it's good for you to do but its true just so you keep your facilities up because there're going to be times when you have to do that. All right, no numerical work, mat-lab won't do it for you, you know you won't be able to find it in a particular table; you have to carry it out. And again, the techniques are usually not beyond what you've already done. They're typically involving either integration by substitution, integration by parts, something like that.

That's all that's involved here, and carrying out integrating the complex exponentials is the same, involves the same sort of steps, same sort of techniques as you learned in the real case when you were just learning ordinary calculus. All right?

I will wrap it up there and then on Monday, I'm going to pick up again with a couple more examples and then also develop the general properties of Fourier transform as we go on our path.

Thank you all, have a good weekend.

[End of Audio]

Duration: 49 minutes

TheFourierTransformAndItsApplications-Lecture07

Instructor (Brad Osgood):I'm gonna upload the solutions to Problem Set 1; I'm gonna do that shortly. There were a couple of people who joined the class late, so I gave them a couple of extra days. Put that Daily away. I know it's exciting about that great Cardinal team. Anyway, so I'll put it up either today – sometime over the next couple days, all right? And the latest problem set, Problem Set 3, is now posted on the website. All right. Any questions about anything; anything on anybody's mind? How do we like our Fourier Transform so far, class?

Student: Whoo.

Instructor (Brad Osgood):Yeah, all right. That's the spirit. All right. So let me remind you – as a matter of fact, let me reintroduce the star of the show. So let's recall the Fourier Transform and its inverse, and I want to make a couple of general remarks before plunging back into specific properties, specific transforms and some properties, and its inverse.

So, again, f of t is a signal and the Fourier Transform or function, same thing, the Fourier Transform, I use this notation. I want to comment about that, again, in just a second. Integral from my infinity – infinity of either the -2p I ST, F of T, VT and the inverse Fourier Transform looks very similar except for a change in sign in the exponential. So the inverse Fourier Transform of – I use a different function, although it doesn't matter. We're gonna go from -8 to 8 of either the +2p I ST, G of S, DS, okay?

All right. Now, I want to make a few general comments here. This isn't one I have to write down, but it's actually quite a complicated operation, all right? Integration is not such a simple operation. Integrating a function against the complex exponential is gonna make the thing oscillate, and computing an infinite integral from -8 to 8 brings with it all sorts of peril, all right?

Now, the rigor police are off duty, for the most part, all right? So we are not gonna be so concerned about the existence of those integrals. That is an issue, all right, and it is something we're gonna have to deal with but not right away. Right now, I just want to get a little practical, hands-on know how with using the formulas and being comfortable with the formulas.

So I'm not gonna worry about the convergence right now. We will talk a little bit more about later, although, it's never gonna be a big issue for us, or rather, we're gonna – well, we're gonna talk about it in a number of different ways, all right, but it is something to worry about.

So when I write down the definition of the Fourier Transform, what I really should have said was the Fourier Transform is defined by that integral whenever the integral exists, all right? So the question is the existence of the integral on the existence of the integral, that is to say convergence of the integral – of the integral. All right, so more on this, to some extent, later.

The set of points but just to introduce a bit of terminology here, something that I'm sure you've said, and I actually have used, I think, somewhat informally. The set of points for which the integral does exist, that is to say, for which the Fourier Transform exists, is called the spectrum, all right? So the set of S of S in R for which the Fourier Transform is defined, that is to say for which the integral exists, is called the spectrum.

So when I made the bold statement last time that every signal has a spectrum, and the signal is determined by a spectrum, I was thinking of exactly this – that is that, in many cases, the integral is not an issue; that is it will exist, all right? And the cases when it doesn't exist, of course, that poses an additional problem; you have to do some further analysis.

All right. Now, that's one comment that I wanted to make. As I said, the rigor police are off duty. I'll tell you when we have to worry about those kinds of questions. Now, that thing I wanted to say is that our definition of the Fourier Transform and the notation I'm using for the Fourier Transform is not universal. As a matter of fact, there is no universal definition or universal standard, universal notation for the Fourier Transform.

So definitions and notations vary, all right? You should be aware of this, and it's not that – and no single notation is perfect, all right? Different notations are useful in different context. So you see – you will see, for example, the notation F Hat – and really, sort of, corresponding to what we did for Fourier Coefficiency, we'll see the notation F Hat for F Hat of S, or what I'm calling, for the Fourier Transform.

You will also see – this is a pretty cute notation. You'll see an upside down Hat for the inverse Fourier Transform. I'll call it T, all right? You'll see that notation. It's not uncommon at all, all right? You also see – that's probably the cutest one, but it, sort of, gets lost in typesetting. You also see – a very common notation in engineering especially is to use a lowercase letter for the function – for the signal, and an uppercase letter for the Fourier Transform.

So you also see f of t for the signal, and F of S for the Transform, all right? The problem with this notation is that it's not so good when you talk about duality, which we're gonna talk about today, actually, because it really makes more of a distinction somehow then there ought to be made.

Anyway, but it's useful – all these notations are useful in some context, but no notation is useful in every context. There's always some ambiguity; there's always some clash, all right? And you have to be flexible, and you also have to know which notation the particular author, the reference that you're looking at, is employing because people will use different notations.

There's also a notation that people use for the, sort of, sibling relationship, which might mean more to you in just a second when I talk about duality, between a function and its Fourier Transform. So some people write things like F of T corresponds to F of S. I, sort of, use that notation, but there's this other notation, which I think is just idiotic, which is like a subset notation or something like that, right?

Like, does it go this way, or the other way, or something like that? I hate that notation, but it's used. Braceval uses this, God rest his soul, but it's a stupid notation. I'm sorry, you know? He's dead; it's gone. All right. Finally – and it may have something like that. Anyway, say no notation is perfect, but some notations are worse than others.

Finally, it's not universally agreed on how the Fourier Transform itself should be defined, all right? There are different definitions – I think, actually, let me just go back. I think the script F notation that I'm using and a lot of other people use also is the least ambiguous notation. It's sometimes awkward in certain context, but they say it's the least ambiguous, and somehow I tend to gravitate toward that more than others, but, again, you can choose what you want, and you'll see them all in use.

Anyway, I was gonna say there are also different definitions that are in use. For example, one sometimes defines the Fourier Transform of – because where do you put the 2p and where do you put the minus sign, all right? So you will see this definition. You'll see the Fourier Transform of F at S is the integral for -8 to 8 of E to the I ST, F of T, DT, with no 2p up there, or you sometimes what you see is I omega T instead of S, all right?

And you'll see the inverse Fourier Transform – excuse me, a minus, without the 2p. You'll see this notation. You'll see the Fourier Transform of F at S is 1 over 2p × the integral for -8 to 8 of either the -I ST, F of T, DT, sometimes putting the factor of 1 over 2p out front. You will also sometimes see the Fourier Transform with a plus sign and the inverse Fourier Transform with a minus sign, all right?

You'll also see the Fourier Transform F at S is integral from – or sometimes with a 1 over 2p out in front. All combinations are in use, all right? E to the +I ST, F of T, DT, and the inverse Fourier Transform is the integral for -8 to 8 be to the -I ST, F of S, DS, all right? You'll see all these notations in use, and you just have to know – somebody has to tell you or somehow otherwise you have to figure out which particular convention is in use, all right?

This one is actually especially popular in physics, all right? It's quite often in problems in optics, for example, where the Fourier Transform comes in. You often see this is the definition of the Fourier Transform, sometimes with a 2p in there. You can stick it here; you can stick it there. If you don't like it, you can stick it someplace else, and this is the definition for the inverse Fourier Transform, all right?

All I can say is be careful out there.

In the notes, I moved this in various places; I'm not sure the best place to put this. I think it's now at the end of the chapter called – the chapter on convolution, which we'll get to very shortly, there's a section called Chase the Constant or something like that, which I stole from a book by Tom Kerner.

And that tells you how the different formulas change when you, you know, we change the plus sign to a minus sign, when you change where you put the 2p and so on, all right? And so it's, sort of, it's meant to serve as a dictionary to help you translate one case from another case, or one convention from another convention when you're likely to run across them, and you will be likely to run across them. Okay. So I felt like I ought to tell you that.

Now, I think I may print that out separately and actually just give it to you, sort of, as a dictionary so you can make that translation because, as I say, it's just a pain in the neck; what can I say? But it's somehow in the nature of the subject that there's not a universal definition. I think it's fair to say that the one that I've given there is probably the most commonly – even then I'm a little bit hesitant to say that, but I think it's safe to say it's the most commonly accepted convention, but by no means, is it universal.

Okay. All right. And we also had last time – a reminder we did last time, we had two basic examples of the Fourier Transform, calculating the Fourier Transform, and, again, at this stage, the only way we have to calculate the Fourier Transform is by recourse of the definition. You have to carry out the integration if you can, all right? There's no way of getting around it. We'll learn, then, today and then going on other techniques that will allow us to calculate new Fourier Transforms from old Fourier Transforms, but, for now, right at the beginning, there's nothing to do except plug into the formula.

So the basic examples we had, which come up very often in applications, are the rectangle function, p of T, and that's the function that looks like this. It's 1 from -½ to ½ and then 0 outside the interval, and I don't care how it's defined at the end points because it doesn't make any difference in the calculation, and the Fourier Transform that is the sync function, Fp of S is sync of S, which is, in my convention, sign of pS over pS, and that's, by the way, of course, another thing that's not universally agreed upon. Some people define the sync function without the p in here. Some people just say the sign of S over S, or sign of X over X without the p. What are you gonna do? What are you gonna do?

The other example that we had, I didn't carry out the calculation completely, but, again, it was based on just calculating the integral. In this case, you had to use integration by parts. It's the triangle function has Slope 1 going up from -1 to 1, and then down from 1 to 1, 0, that's lambda of T, and the Fourier Transform there is the sync^2, nice result, sync^2 of S, okay?

All right. I want to do one more particular case that's really cool, actually, and also comes up quite a bit in applications, and then we are going to take the different path, the second path that I talked about, that is talk about some general properties of the Fourier

Transform, and how you'd use the Fourier Transform to find – how you can formulate some general properties that will allow you to find Fourier Transforms of combinations of functions or modifications of functions.

But let me do one more explicit example for you, and that is our friend the bell curve, the Gaussian. That's a very important function in many applications, and it has a remarkable property with regard to the Fourier Transform. So, as the third example, let me do the basic Gaussian F of T = E to -p T^2.

Now, again, there's a question of normalization here, all right? The shape is the familiar bell-shaped curve, which comes up in probability distributions, and I'll talk about that probably a little bit next time or the time after that when I talk about the central limit there.

I mean, it has a height of one, and with this normalization, that is putting the p in the exponent, that normalizes the area under the curve to be one. It's not the only way of doing it, but it's the way of doing it that's, somehow, most convenient – most natural when you're working with Fourier Transforms.

So the result when you make this normalization is the integral from -8 to 8, either the -p T^2, DT is 1. Now, as tempted as I am, I'm not gonna show you why that's true. I'm not gonna do the calculation. That's one of the most famous tricks in all of mathematics to get that result going back to Oilier.

You cannot do this by direct integration because the function, E to the -p X^2 has no elementary anti-derivative. So you can't do this with the fundamental theorem of calculus. It has to be done by a very devious other means, and it's done in the notes. You should look it over. If you have any questions, ask me about it because I never get tired of talking about it because it's such a famous and elegant trick, but, in the interest of time, I think I won't go through the derivation, all right, but that's an important result, and a surprising result. I mean, why Ep 1, why should they all come together in such a simple formula? But they do.

So here's what I want to show you, and to let the cat out of the bag, I'll show you where it comes from is that the Gaussian is itself – no, the Gaussian is itself. That's – I am myself too, and you are yourselves, but that's not so interesting. What's interesting is the Fourier Transform of the Gaussian is itself. That is if F of T is equal to, for this normalization, either the -p T^2, and what I want to show you is the Fourier Transform is the same function, either the -p S^2. I'm using a different variable there, but the basic fact is that the Fourier Transform of the Gaussian is the Gaussian.

Now, I mean, it's a very striking result because computing the Fourier Transform is a complicated operation, as I'm gonna go through this to show you how it works. So why a function should turn out to be itself under Fourier Transforms is really quite striking.

We’ll also see, soon, that – you remember, you’ve heard me talk about many reciprocal relationships between the time domain and the frequency domain, and we’re gonna see various examples of that, and, somehow, what this tends to say is – for reasons which you’ll understand a little bit better later – is that somehow the Gaussian is equally spread out in the time and the frequency domain; there’s no difference. Because, often, what happens is if a function is spread out in one domain, it’s stretched out and it’s squeezed together in the other domain or vice versa, all right?

We’ll see that, actually, as an example of – well, we’ll see that as a basic property of the Fourier Transform, and what this says, the fact that the Gaussian – the Fourier Transform of the Gaussian is itself, means, somehow, it is equally spread out in both time and frequency, all right? Which is something you wouldn’t occur to you to say or would occur to you to think about unless you knew this result, all right?

So why is something like this true, all right? We have, again, no recourse other than to the definition. So let me write – in this case, I think it’s a little bit easier actually to use the capital letter notation because of I’m going be differentiating and performing some other unnatural acts on the Fourier Transform, so watch.

So let me call, again, if F of T is = to either the -p T^2, let me call capital F of S its Fourier Transform. So that’s the integral for -8 to 8, E to the -2p I ST, E to the -p T^2, DT, okay? And I am actually gonna evaluate this integral, all right? Not by any, sort of, appeal to the fundamental theorem of calculus or anything like that. That won’t work, but there is a clever trick that will actually allow us to carry out the integration after an initial, little slight of hand.

That is to say I’m gonna differentiate with respect to S, all right? That is F prime of S, and this can be justified – the rigor police are off duty, but even if they’re on duty, I would have no trouble in asserting that I can find the derivative of this function by differentiating under the integral sign. So the integral for -8 to 8, the derivative of this thing is DDS. The only thing that depends on S here is this complex exponential, either the 2p I ST, and then either the -p T^2 stays the same; it doesn’t get hit by the derivative. So the derivative of this with respect to S is equal to the integral for -8 to 8 -2p I T, E to the -2p I ST, then E to the -p T^2, DT, okay?

All right. Now, watch this. You factor out the I, I’m gonna write this out as I × the integral for -8 to 8. I want to group terms in a very suggestive way. So it’s gonna be E to the -2p I ST, and then 2p T - 2p T × E to the -p T^2, DT, and I group the terms this way because it cries out to be integrated by parts, all right? Differentiating with respect to S brings down this factor of 2p T or -2p T here, all right? This absolutely cries out to be integrated by parts. This is the U. This is the DV, okay?

And if I do that, what happens? Well, if DV, once again, is -2p T × E to the -p T^2, VT, then V is E to the -p T^2, all right? Derivative of this E to the -p T^2 brings me down with respect to T, brings down to the -2p T, all right? And if U is equal to E to the -2p I ST, then DU = -2p I S, E to the -2p I ST, 2p I ST, DT, all right? I’m differentiating there with respect to T, or D-ing with respect to T, so to speak, okay?

All right. So, again, integration by parts tells us that the integral from A to B of UDV is UV, evaluated between AB - the integral from A to B of VDU, all right? So how does that work out in this case? In this case, there’s an I times the whole thing, right? So there’s an I × U × V, which is E to the -p T^2, that’s a V, -p T^2 × E to the minus – where’s U? U is – yeah, -2p I ST, and evaluated between -8 to 8, minus the integral from -8 to 8 of VDU. V is either the -p T^2, DU is -2p I S, E to the -2p I ST, DT, all right? Closed braces because there’s an I in front of the whole thing. Cool? Way cool.

All right. Now, what about these boundary terms here? What about the terms U × V between -8 to 8. What happens to that? Gone, all right, because this thing has absolutely value one. E to the -p T^2 is going to 0 at both +8 and -8, all right? So this is killing this off. It is gone.

What remains? What remains is -2p I S, a minus sign here, and then there’s an I in front, all right? So if I got all that right, it’s gonna be an I × an I, and I × I is a -1, right? So this is gonna be – right. There’s a minus, a minus, and then an I × an I gives you an extra -1. It’s gonna be minus – the integral for -8 to 8 of 2p I S – 2p S × E to -2p I ST, E to the -p T^2, DT, okay?

Now, I’m integrating with respect to T. This comes out. This depends on S. So this is -2p S integral for -8 to 8 of E to the -2p ST, E to the -p T^2, DT. Brilliant. I’ve gotten back to where I started. All those years of education, all that work, what did it get me? Back to where I started, except there’s an extra factor out front. That is this is equal to -2p S × F of S, the original Fourier Transform. That integral is the Fourier Transform of the Gaussian. Integral for -8 to 8 , E to the -2p I ST × E to the -p T^2, all right?

So what have I shown here? I have F prime of S = -2p S × F of S. Oh, but that’s just a simple differential equation for F, kids, right? So that says that F of S = the initial value F of 0, E to the -p S^2. That’s the only solution to that baby, okay?

And what is F of 0; what is capital F of 0? Capital F of 0, that’s the value 0 of the Fourier Transform actually, right? When S = 0 this is -8 to 8, E to the -2p I, 0 × T, E to the -p T^2, DT. That’s the integral for -8 to 8, E to the -p T^2, DT, which is 1, okay? So what is the actual retail value of the answer? That says that F of S, capital F of S is E to the -p S^2. This says that capital F of S is E to the -p S^2. Done. Fantastic, fantastic, all right?

So again, I take the Fourier Transform of the function E to the -p T^2, that corresponds to E if I want to use that notation for p S^2, all right? The Fourier Transform of the Gaussian, when it’s normalized this way is itself. If you change the normalization, you’re gonna get a different answer, although, we’ll figure out how to do that. We’ll figure out how to make such changes, but for this Gaussian the way it’s normalized, E to the -p T^2, its Fourier Transform is itself – quite remarkable, and quite important, all right, quite important.

Okay. All right. Now, that's about all I want to do at this stage for specific transforms. We only have three, actually: the rectangle function, the triangle function, and the Gaussian. There are other examples in the book, all right? There are other examples on those. There's a one-sided exponential decay, the two-sided exponential decay. I'm not gonna go through the calculations, all right? If this were a regular calculus class, I'd go through all those calculations, but you're beyond that. You can read the calculations; you can certainly use the results, all right?

They all have to rely on using the definition of the Fourier Transform. There's no other recourse. That's all you have to work with is the definition, all right? So that means all you have at your disposal is – well, there are tricks like this, and then but, typically, they're just integration techniques, and the integration techniques are usually integration by substitution or integration by part; that's all that's use in any of those derivations, all right? So I'll let you go through those things, and certainly use the formulas. There's no reason to memorize them especially. I mean, you can always look them up if you need them; they're there, and so on.

Instead, what I want to do now is I want to talk about some general properties of the Fourier Transform. I mentioned there are two paths that we want to follow. One is finding specific transforms, and then one is finding general properties of the Fourier Transform that will allow us to find the Fourier Transform of different combinations of the functions. So now I want to pursue that second path and look for general properties.

And the first one I want to look at often goes under the heading of duality. So I want to explain that to you. It's actually very simple, but – Fourier Transform Duality, all right? The formulas are extremely simple, but for some reason – well, for reasons which I'll explain to you, many times students really agonize over these formulas, all right? So I want to go through it in a little bit of detail.

Here, what I want to do is I want to exploit the similarity in the formulas for the Fourier Transform and its inverse. That's what this is about, between the formulas for the Fourier Transform and the inverse Fourier Transform, okay? The Fourier Transform of – let me write down the definition again. The Fourier Transform of F of S is the integral from -8 to 8, E to the -2p I ST, F of T, DT, okay? Now, remember the Fourier Transform – say this to yourself early and often – the Fourier Transform is an operation that turns one function into another function, all right? To evaluate the transform, I have to evaluate it at a variable, and in this case, I'm calling the variable S. Fine, that means I plugged S into the formula, and that tells me how to compute the Fourier Transform, fine.

What if I plug in -S into the formula, all right? The Fourier Transform of F evaluated at -S. The Fourier Transform is an operation that turns a function into another function, and to evaluate it, I have to plug in a point. I plug in the point -S. What do I get? I get the integral for -8 to 8 of E to the -2p I × -ST, F of T, DT, which is the integral from -8 to 8 of either the +2p I ST, F of T, DT, which is, if according to my formula for the inverse Fourier Transform, the inverse Fourier Transform of F at S.

I'm gonna write that down one more time. Okay. The Fourier Transform of F at -S is the inverse Fourier Transform of F at S. Fine, that's exactly what that formula says. Now, people, this gives some people just fits, all right? Why? Because this is an example of being wedded to an interpretation, all right? This is an example of – if it gives you fits, it's because you're too wedded to an interpretation and wedded to your variables, all right?

Because they say how can it be S on both sides of the equation? If I take the Fourier Transform, I get into the frequency domain. If I take the inverse Fourier Transform, I get back into the time domain, but you're calling it S in both cases, man. How can you do that? Are you allowed to do that? What kind of fool are you, man?

It doesn't make sense, all right? That's because you always think of – or people cling to the idea that they cling to one particular interpretation. You have the signal and the time domain. This Fourier Transform is something in the frequency domain. You cannot mix the two in your head or anywhere, ever, all right?

And so this formula bothers people. You take the Fourier Transform evaluator minus, it's the same thing as taking the inverse Fourier Transform, but I say to you, as I said before when I was doing it, and that's why I was making the point, the Fourier Transform is a formula, is an operation, that turns one function into another function.

To write down the formula, you have to evaluate the operation at a variable. It doesn't matter what I call the variable. I can it S. I can call it T. I can call it zippity do da. I can call it anything, all right, as long as I am consistent, all right? And there is no inconsistency; there is no error in this at all, all right? So don't think in terms of time, frequency, or whatever. Think in terms of the mathematical operation. The Fourier Transform evaluate a -S is the inverse Fourier Transform.

Now, there is a neater way of writing this. This statement actually and related statements, which I'll show you now, are – and this is exactly sometimes called Duality, or the Principle of Duality, or Duality for Fourier Transforms, or whatever. You would have a tough time – this is an example, actually, where you would have a tough time with other notations, all right? If you used the capital letter notation, all right, that would give you fits because the capital letter notation somehow forces you to distinguish between the two domains, little f of t, capital F of S, you know? And that would make it hard to write down that formula in a way that really made sense, all right? But if you use the operational notation here or the operator notation, then it's easy to write down. It's unambiguous, and there's no problem with it, all right?

Now, maybe, like I said, maybe you're looking at me like, "What is the problem? I don't have any problem with this, but, trust me, I have seen many times many people have a lot of trouble with that formula because they're trying to interpret it in ways that it can't be interpreted. It's just a mathematical statement, all right? It's a very useful mathematical statement as it turns out, but it is a mathematical statement. It's not a statement about time and frequency – not to my knowledge, all right?

Now, I'm gonna write this another way, write more neatly. I think actually it's also a good idea, I think, to try to write things in as variable a free way as you can, all right? Or try to introduce – even if you have introduce a little extra notation to try to not write your variables because in this subject writing variables and naming variables just can get things pretty muddled, all right? So I want to write this a little bit more neatly in a way that will actually allow me to write the duality formulas without any variables, where there's, sort of, no agonizing, or perhaps, no agonizing.

So I want to introduce the reverse signal. As a matter of fact, I think this was even on the first homework, first or second homework, all right? That is if F of T is a signal, then I define F - T just to be F of -T. All right. Not a big deal, all right. So you reverse – if you want to think of T as time, you're reversing time; that's why I call it the reverse signal, okay?

Then, first of all, it's a nice way of expressing evenness and oddness, actually, then in terms of the reverse signal. F is even if the reverse signal is equal to itself because to say that F is even is to say that F of -T is equal to F of T. So say if you reverse a signal, you get the same signal back again. F is odd, if you reverse the signal, you get back minus the signal. Because F is odd, if F of -T is -F of T, all right? So it's a nice simple, let's say, variable free way of writing it, if you want.

And what does this formula say; what does the duality formula say, the first duality formula say? Once again, with variables, it says the Fourier Transform evaluated at -S is equal to the inverse Fourier Transform of F evaluated at S, all right? I want to write this in a variable-free way using the reversed signal. What is the Fourier Transform of F evaluated at -S? It is the reverse of the Fourier Transform. Now, watch very carefully where I put the parenthesis. This is equal to this because the Fourier Transform of F reversed, evaluated at S is the Fourier Transform of F at -S, okay?

So this statement in a variable-free way says the Fourier Transform of F reversed is the inverse Fourier Transform of F – kind of, nice. I mean, the reverse of the Fourier Transform is the inverse Fourier Transform. That's why, in fact, some people talk about the Forward Fourier Transform and the Reverse Fourier Transform.

Actually, they don't even use the term transform and inverse. They sometimes talk about the forward transform and the backwards transform or the reverse transform, and I think what they're often thinking about is this formula, although there's this extra sign change in there. So it says the Fourier Transform of F reversed is the inverse Fourier Transform of F.

Now, what about – so that's one form – I didn't do anything new there. I just rewrote the formula in a variable-free way. It's a way that I can tend to remember a little bit more easily, to tell you the truth.

Let's look at something else. Let's look at the Fourier Transform of F-, the reverse signal, all right? Notice the parentheses there, all right? I'm not reversing the Fourier Transform of F; I'm taking the Fourier Transform of the reverse of F. What is this? Well, in order to evaluate this, I have to evaluate it at a point.

So the Fourier Transform of F- at S is, by definition, the integral for -8 to 8 – be careful here – of E to the -2p I ST, F reversed FT, DT. That's my definition of the Fourier Transform. The Fourier Transform of whatever I plug in here is what I plug into the formula. In this case it's F-, F reversed. This is the integral for -8 to 8 of E to the -2p I, ST, F of -T, DT, by definition, right?

Now, we're gonna work with that a little bit. We're gonna work with that last expression. We're gonna make a change in variable. All right. This is really an exercise in logic and notation, all right? And I'm only doing it because I can tell you that logic and notation can be a problem in this subject, all right? Finding your way through the notation, and making sure you're saying the right thing, and not saying the wrong thing can be a challenge. I have seen generations of students, very smart students, you know, get exactly balled up over these points. So that's why I'm doing them, and let's hope I can get through correctly.

I'm gonna make a change of variable. I'm gonna let U = -T, all right? That's, sort of, the obvious thing to do here. I want to get rid of this F of -T there, so I'm gonna make a change of variable in the integral, U = -T. So DU = -DT, okay? So what happens to the integral? The integral from -8 to 8 of E to the -2p I ST, F of -T, DT, becomes – let's do one step at a time, all right?

This is the integral of E to the -2p I ST, so T = -U. So it's -2p I S -U, F of U, and DT - DU, all right? Then you have to change the limitave integration also, all right? So if I let U = -T, and if T goes from -8 to 8, then U goes from 8 to -8, all right? So U goes from +8 to -8. Once again, if T goes from -8 to +8, then -T goes from +8 to -8.

But so if that's the integral, either the 2p I SU, F of U, DU, and there was an extra minus sign out front, so it's minus the integral from 8 to -8. That switches the limits of integration. That becomes the integral for -8 to 8, okay? What remains, dear friends? What is that integral? You recognize that as the inverse Fourier Transform of F evaluated at S, okay?

What is the conclusion? The Fourier Transform of the reverse signal is also the inverse Fourier Transform. If I don't write the variable in there, all right? The Fourier Transform of F reversed is the inverse Fourier Transform, all right? That's another duality theorem. Now, I want to combine this with the earlier one that I had – where was the earlier one I had?

Was the inverse Fourier Transform of F – what inverse Fourier Transform of F was the Fourier Transform of F reversed. That's this one over here, all right? That's the first duality statement that I had. If I combine this one with this one, then I get a very nice statement. I get that the Fourier Transform of F reversed is the inverse Fourier Transform of F. That's the Fourier Transform of F reversed.

So just look at this one and this one. This is my favorite statement, actually, of the duality statements 'cause it's the easiest one for me to remember.

It says that the Fourier Transform of the reverse signal is the reversal of the Fourier Transform. In words, it's so easy, and it's so nice. The Fourier Transform of the reverse signal is the reverse of the Fourier Transform, all right? That's another statement of duality, all right?

There are no inverses in there. I short circuited by putting the inverse in the middle. The Fourier Transform of F reversed is the inverse Fourier Transform of F. The Fourier inverse Transform of F is the Fourier Transform of the reverse signal. See, I can say all these words fast enough, of course I say everything fast, but I can say all these things because I've been through it a couple times, all right? And because I am not afraid not to write my variables.

So you have this statement. That's another statement of duality. So, really, there's this statement. There is the first statement that I have, the Fourier Transform of F reversed is the inverse Fourier Transform of F, all right? We use this statement to get this statement.

There is one other formula – they're all basically the same. There's one other formula that people often lump into, sort of, duality statements, and that is if you take the Fourier Transform of the Fourier Transform. Now, watch this; this is cute. One more, and that is the Fourier Transform of the Fourier Transform of F, okay? Now, you can derive this a number of ways. I think I will only give you the result of this thing and not give a derivation, but the result is you can take the Fourier Transform of the Fourier Transform and get back the reverse – you don't get back the function; you get back the reverse signal. You get back F reversed, okay?

So you can derive this; you derive this. You can derive it as a constant 'cause of what we've already shown over there. You can try to do it from scratch, just with the definition of the integral; don't worry about convergence or anything like that, just see what happens if you plug it into the definition.

And, again, people have a hard time with this because they're wedded to this interpretation of the Fourier Transform as taking you from the time domain to the frequency domain. That's not that that interpretation is wrong. I mean, in many cases you want to think in terms of two domains. You want to think in terms of times and frequency, but not always, all right, not always.

Sometimes you just want to think of the Fourier Transform as an operator taking one function to another function. I have to plug in a variable in order to write down the formula, but it doesn't matter what I call the variable. It doesn't matter how I interpret the variable, all right? It's just a statement. It's just a mathematical statement of how the thing is defined, all right?

So, say, if you think that way, then you have no trouble – it's one thing to derive this formula, but there's certainly no inconsistency here. There's nothing here that shakes your faith in the world, or shakes your faith in the distinction between the time domain and the frequency domain, you know, screw that. Excuse me. All right, it's just a mathematical statement.

Now, these are very useful, all right? These come up a fair amount, and I'll give you one example; I'll give you a quick application, all right? Quick application, this application is actually finding Fourier Transforms, all right? So, for example, let's find the Fourier Transform of the sync function, all right? We found the Fourier Transform of a rectangle function is equal to the sync function. We haven't found the Fourier Transform of the sync function.

So we know – do it over here. Now, I do this, actually, one way in the notes; let me do it a slightly different way here, all right? We know the Fourier Transform of a rectangle function is the sync function, all right? So the Fourier Transform of the sync function is the Fourier Transform of the Fourier Transform of a rectangle function, all right?

Fourier Transform of the rec – that's the first calculation we did. Fourier Transform of the rec function is the sync function. Fourier Transform of the sync function is the Fourier Transform of the Fourier Transform of the rec function, but the Fourier Transform of the Fourier Transform of the rec function is, according to this formula, the reversal of the rec function, but the rec function is even, all right? The rectangle function is an even function. So this is the rec function again. Lighting fast, we have derived that formula.

Now, you would be hard pressed, all right, to show – let me just indicate here what the issue is. If I were to say to you, "Find the Fourier Transform of the sync function." And you were to say to me, "I have only one recourse, and that is to the definition." All right? You would have to write down the Fourier Transform of the sync function at S is the integral from -8 to 8, E to the -2p I, ST, sign of pS ÷ pS – Oh, it's not a pT ÷ pT, DT.

Good luck. All right? Good luck. There are real serious issues of convergence here for this integral. I mean, and trying to evaluate this, and imagine, this evaluates to, if we believe this, the function which is 1 from -½ to ½ and 0 elsewhere, and – really? No. Oh, kidding? All right? You would be hard pressed, but the duality makes it easy, all right?

Now, like I said, this works because the rigor police are off duty, all right? This works because I'm assuming that there's actually no problem with the definitions, with the formulas, and so on. That all the integrals work and so on. Takes a lot of – it actually takes some work to do this, all right? I mean, it is true. This statement is true, that the Fourier Transform of the sync function is the rectangle function, but actually to do that completely rigorously requires a certain amount of effort which we're not gonna do actually. I mean, we're gonna do it but, sort of, by indirection.

But, nevertheless, with the rigor police off duty, and it's okay, you find very simply how it works, and you find a very similar – the same sort of argument will give you what the Fourier Transform of the $sync^2$ is the triangle function, same duality argument. Same duality argument gives that the Fourier Transform of $sync^2$ is the triangle function, okay? Same thing. All right.

I tell you what. I usually keep you overtime. I think today, maybe you'll even get out on time because, god forbid, we get out early. Next time what I want to do is I want to continue the discussion of general properties. I want to talk about the so-called stretch theorem, the shift theorem, and talk about convolution next time. So I want to move ahead a little bit, and make a little bit of a push to get some of the more general properties of this, okay? See you on Wednesday.

[End of Audio]

Duration: 48 minutes

TheFourierTransformAndItsApplications-Lecture08

Instructor (Brad Osgood):We're on? Okay. So today – first of all, any questions? Anything on anybody's mind? If you brought your problem sets with you, you can turn them in at the end of the period today; otherwise, put them in the filing cabinet outside my office across from my office by close of business today. Okay. Nothing? No burning issues? No burning questions? Is it the threat of rain that kept people away? It's beautiful out there now, right? Blue skies. Sunshine.

All right. Today, I want to continue down the path that we started last time, of talking about general properties of the Fourier transform. Remember, as I've said now a couple of times, there are two tracks that we're following in developing our understanding techniques of using the Fourier transform. One is to develop specific transforms, transforms of specific functions that you need to have at your fingertips, so to speak, the kind of things that come up often enough that you want to know what that answers are, what the formulas are. And the second path is to understand how to take the Fourier transform and different combinations of functions, different combinations of signals that, again, come up often.

So that's what we're gonna do today, we're going down more the second path, including an extremely important operation. So we're gonna have three big items today, each of which are important in themselves and come up all the time. One is delays, what to do with a Fourier transform when the signal is delayed. One, a formula for what happens to the Fourier transform under a stretch, and finally, a very general operation, which we have now seen a couple times in different forms, but today we're gonna see them today in the context of the Fourier transform in its full glory, so to speak, and that is convolution.

So I want to take each one of these in turn, and I want to derive what the formulas are for you. The book has – the notes have examples of how they're used in practice, and you'll also have plenty of chance to practice some of these in problems. What I want to do is show you what the general formula looks like, and how it comes about.

So let's first look at the question of delay. And there, the question is: If a signal is delayed by – I'm not sure which term you prefer, I'm not even sure what term I prefer, delayed or shifted – say shifted by an amount b, what happens to the Fourier transform? In other words, if the signal is F of t, and that corresponds to a Fourier transform F of s, so I'm gonna use the capital notation here. Then if the signal is delayed by an amount b, and for the purposes of this discussion, that means I consider F of t minus b, not F of t plus b. And, again, it's always a question of what you mean by delay, whether you take a plus or minus, but I think this is probably a fairly standard way of looking at it. Then the question is what happens over here? What happens to the Fourier transform?

Now, this is the sort of question that you have to be able to answer for yourself routinely. These are very simple cases, but there are other cases where similar sorts of things come up, not too many, fortunately. And, again, we have no recourse here, other than to deal with the definition of the Fourier transform. So the Fourier transform of F of s is the integral from -8 to 8, e to the minus 2pi, st, F of t, dt. And then the Fourier transform of the signal F of t minus v is the integral from -8 to 8, e to the minus 2pi, st, F of t minus dt. Now, when you get enough experience, and by now, you probably have that experience, when you look at something like this, the thing that cries out to be done here is a change of variable. You want this to look as much as possible like the ordinary formula for the Fourier transform, that means F of a single variable here rather than F composed with anything.

So what you do is, the simplest thing is to make a change of variable. So u = t - b, b is a constant. So du = vt, and t = u + b then. And if t goes from -8 to 8, since I'm just subtracting off a fixed constant, so does u. So in terms of the new variable, the integral becomes the integral for -8 to 8, e to the minus 2pis, t becomes u + b, and the function becomes F of u, du, in terms of the new variable.

And, now, the thing that's crying out here, is to manipulate the complex exponential. So write that as the interval -8 to 8, e to the minus 2pis, u times e to the minus 2pis times b, F of u, du, and realize that because I'm integrating with respect to u, this second complex exponential here, either the minus 2pi, sb, does depend on u, so it comes out of the integral. It's a constant, as far as integration is concerned.

So I can write that as, e to the minus 2pi, sv, integral from -8 to 8, e to the minus 2pi, su, F of u, du, and what remains is the complex exponential, this complex exponential here, times the Fourier transform of F, the Fourier transform from the original function F. That is, this is e to the minus 2pi, sb, F of s, what I called F [inaudible].

So where do we start? Where do we finish? We started by shifting F, and we discovered what happened to the Fourier transform. In other words, if the original signal F of t has a Fourier transform F of s, then the Fourier transform of F of t - b corresponds to e to the minus 2pi, sv, F of s. With a similar derivation, this is sometimes called the shift theorem, delay theorem, people call it various things; it's very simple and it comes up all the time. And for me, actually, I sometimes have trouble remembering whether it's a plus or minus up here, so that's my own personal burden. So it's actually maybe worthwhile to write down in general F of t, if I change the minus to a plus here, I'd have a plus here. So I even usually write down the formula that looks like this. The Fourier transform of F of t plus or minus b corresponds to e to the plus or minus 2pi, sb, F of s. It's the same formula, I haven't done anything different, but sometimes I like to remember them sort of distinctly. But that's my own burden. This is an example, actually, where the notation can be a little bit of a problem. That is, in this case, sort of the upper case notation is probably most helpful, indicating how the signal is paired with this Fourier transform, and if you change the signal, how does that change the Fourier transform. Because writing down with the operator notation, writing this down, is kind of awkward, because F is already evaluated at t - b, you take the Fourier transform that's supposed to be evaluated another variable. What, do I write like this? Do I write this evaluated at s? That's okay, but that's a notation only a mother would love; it's just too complicated. So we can all understand, sort of by ladies and gentlemen's agreement, what you mean when you write down the Fourier transform of F of t - b. But all I'm saying is that you have to be very careful with your variables and how you write them; what you write, what you don't write. And if you don't believe me, you'll have plenty of occasions to screw it up, trust me, because everybody does. It's just the nature of the subject, that the notation can sometimes be – like I said, there's certain ambiguity to the notation that can sometimes can cause problems. So this is sort of a bad notation in that case, although, I don't usually like to use it as probably a better notation. So the formula itself is very simple. What is the interpretation of the formula? And the interpretation is also something to keep in mind, and that is a shift in time corresponds to a phase shift in frequency. So here, I think it's actually better to use the word shift instead of delay because I can use it in both sides of the sentence. Shift in time corresponds to a shift in frequency, but a shift in frequency you think is a phase shift, corresponds to a phase shift in frequency. If I think of the signal as living in the time domain and the Fourier transform as living in the frequency domain. Phase shift in frequency.

So why do I say this? Well, because what I have in mind, with a statement like that is, that the Fourier transform is a complex number, so it has a magnitude and it has a phase. So I write F of s, the Fourier transform of the original signal as its magnitude, the magnitude of F of s, that's a real number, times e to the 2pi, say, theta of s. So theta is the phase. And then if I multiply – this is not a hard thing but it's a thing you have to keep in your head, not only the formula but any interpretation of the formula. So e to the 2pi or e to the minus 2pi, sv, times the Fourier transform of s is – the magnitude stays the same. And it's this complex exponential times that complex exponential. So that's e to the 2pi theta of s minus 2pi, s, theta of s - b. I have an s in both cases. No, I don't, 2pi theta of s - sb. In general, the phase can depend on s. That's why I write a theta of s, the phase can vary with s as well; it usually will. So there's the phase shift. The magnitude is the same. The magnitude of the spectrum does not change. The magnitude of the Fourier transform does not change. What changes is the angle. So I will let you push that interpretation however far you want, that is if you want to think about other sorts of ways, but that's the basic interpretation of the shift theorem. A shift in time corresponds to a phase shift in the frequency and the magnitude stays the same. So, again, I will let you convince yourself if that makes sense on physical intuitive grounds, if you think about a signal and how it's made up. Next, I like these formulas, and I think it's nice the way these formulas work out. This is not as exciting as some of the other things we've done, I admit. But, nevertheless, it's part of the whole package that you have to be familiar with; you have to be comfortable with. And it's not only the formulas, it's how the formulas are derived. One of the reasons why I'm going through this is just to show you, and I hope you can also show yourself, is just the kind of techniques that are involved in deriving these formulas. Because if you can see them in these relatively simple circumstances, often it's the case that you can apply them in more or less the same way in slightly less obvious or less simple circumstances. It's a pedagogical kind of thing. Now, the second general theorem, or formula, is what happens when you scale the time. And here, actually, there is a very important interpretation that I'll talk about after we get the formula, but this one I think reveals a deeper property of the Fourier transform.

Scaling is scaled – when I say scaling, I'm talking about scaling the independent variable of scaling of the time. If F of t, again, corresponds to, let me write it like this, the Fourier transform F of s, and then if I scale the time by a constant a, so I'm thinking of a as just a real number here, but not necessarily positive, then what happens to the Fourier transform? So, again, to answer this question, we have no recourse in terms of the definition. So the Fourier transform of a scale version of F is the integral from -8 to 8 of e to the minus 2pi, st, F of at, vt. And, again, the thing that is crying out to be done here, is a change of variable. But it's helpful, I think, to make a distinction here, whether a is positive or negative, because of what happens when you make that change of variable to the limits. You know, we have [inaudible] definite interval, the limits of integration also have to change, and it's helpful here to make a distinction. They change differently whether a is positive or negative. So let's take the case when a is positive first. So if a is positive, I'm gonna make, in either case positive or negative, I'm gonna make the change of variable to du = a times dt, that's clear that I want to do that. Then, of course, du = a times dt. Now, if a is positive, and t goes from -8 to 8, then so does u; u also goes from -8 to 8 because a is just a fixed constant here. So the integral becomes the integral from -8 to 8 of e to the minus 2pi, s, u/a times F of u; dt becomes 1/a times du. The a is a constant; I can factor that out of the integral. The integral is from -8 to 8 and this is minus 2pi. I'm gonna put the 1/a with the s instead of the u. And if I do that, why it's plain enough what to see, what remains is the Fourier transform of the original signal F, evaluated not at s but at s/a. So that is 1/a times the Fourier transform of the original function, not evaluated s but evaluated s/a. Fine. Done.

Now, what about if a is a negative, let's take that case. So, again, I'm gonna make the same change of variable, du = a times dt, du = a times dt. Formally, that's all the same. Now, what happens to the integral? So here you have to be a little more careful. If a is negative, and t goes from -8 to 8, then a times t goes from +8 to -8. Some people are skittish when I write integrals. They say, "You can't do that, you can't write an integral from a bigger number to a smaller number." And I say, "I just did. Nothing happened. God did not strike me dead." So you just have to be careful. So it's the integral from +8 to -8, and then the rest of the integral looks the same formally, that is minus 2pi, s, u/a, F of u, and then dt becomes 1/a times du. So the 1/a comes out of the integral. I'll do this in two steps. So it's 1/a, the integral from +8 to -8 of e to the minus 2pi, and, again, I'll bring the 1/a to the s instead of u, F of u, du. And now, if I want to recognize this as a Fourier transform, I swap it with limits of integration, instead of integrating from +8 to -8 I integrate it from -8 to +8, but then of course I have to change the sign. So this is minus 1/a times the integral from -8 to 8 of either the minus 2pi, s/a, u, F of u to u, which is minus 1/a times the Fourier transform evaluated at s/a, which is minus 1/a times the Fourier transform of the original signal evaluated at s/a. Now, you can combine these two. So that's how the two formulas look. And they're different, sort of, when a is positive and when a is negative. But, really, you can combine both cases if you realize that if a is negative, minus a is positive. In other words, this is the same thing as 1 over the absolute value of a times the Fourier transform of s of a, the Fourier transform from the original signal evaluated at s/a. And, of course, this also includes the previous formula because if a is positive taking the absolute value doesn't do anything to it, it just returns to the original value. In other words, the stretch theorem or the similarity theorem, whatever term you want to use, can be written in one piece. I had to the derivation, really, in two parts, but there's only one formula. And the formula is, that if F of t corresponds to

F of s, then F of at corresponds to 1 over the absolute value of a times the Fourier transform of the original signal at s/a. So notice the absolute value is on the outside here but not on the inside.

Now, suppose if you should say something like, of course a is not zero, but – you know, I take triviality insurance out on all my statements. If my statements are trivially true or trivially false, you can ignore them, because I know I will. Okay. That's sometimes called the stretch theorem or the similarity theorem, because the scaling is either called stretch or a similarity. Similarity is probably a better term because stretching and shrinking can be the same sort of operation, depending on whether or not a is bigger than 1 or a is less than 1. And let's talk about that point. What is the interpretation of this? Because this actually has, I think, a more fundamental interpretation than the stretch theorem. What is the interpretation? Well, let's take the case, first of all, when say, a is bigger than 1. So I'm gonna just consider a positive here. So I'll take a to be positive, and then first take a bigger than 1. If a is equal to 1, I'm not doing anything, of course. If a is bigger than 1, then how does, say, the function F of at compare with the function F of t or the graph of the function F of at compare with the graph of the function F of t? So how does F of at compare with F of t? Well, if a is bigger than 1 then – you saw this back in eighth grade, when you were first learning these things, but sometimes I have to give myself an argument for how these things look like. So if the signal is sort of like this, if this is F of t, then F of at is the same signal except it's squashed; it is compressed. Same shape, more or less, but it's sort of a squashed version. F of at is a compressed form of version of F of t. You convince yourself of that, that if the variable is bigger, like if I scale by something bigger than 1, then I'm squeezing the function. What about Fourier transform, what happens on the Fourier transform side? On the Fourier transform side, originally I had the Fourier [inaudible], and then I go to the Fourier – if I scale it, I go to 1/a, F of s/a. Now, I'm gonna draw the graph, but the graph is a little bit deceptive here because you can't draw the graph of the Fourier transform because the Fourier transform is complex value. People make this slip all the time, where they think they're gonna draw a picture of the Fourier transform, but you can't really draw a picture of the Fourier transform because you'd be plotting complex numbers. So what people actually plot, when they plot pictures of the Fourier transform, is its magnitude. So when I draw a picture here make that a realization. So if this is the original Fourier transform, if the signal is real, the magnitude of the Fourier transform is gonna be symmetric, so it's gonna look something like this. It looks a lot that, actually, but that's just because I have lack of imagination when I draw my functions. So if this is F of s, or rather say the magnitude of F of s, then if I scale here what happens? If a is bigger than 1, 1/a is less than 1. So I'm scaling s by something less than 1. If I scale it by something less than 1, I am stretching the graph out instead of compressing it, and I'm also multiplying by something less than 1 out front. So there's actually a stretching in the horizontal direction, and a shrinking in the vertical direction. And, again, this is not so easy for me to draw, but it's like a stretched out version of the signal, and shrunk.

So it's stretched out horizontally, so this 1/a, I don't have to put the absolute value there because I'm assuming it's positive, 1/a F of s/a, magnitude. So it's stretched out horizontally, and it's also squeezed down vertically. Is that right? How do I say it, decreased vertically? Squashed, those are the technical terms, squashed vertically. So that's what happens if a is bigger than 1. If a is less then 1, the situation is reversed. If a is less than 1, then if this is F of t then F of at is stretched out. And what happens to the Fourier transform, that is squeezed together. Because, again, the Fourier transform is then scaled by something bigger than 1. So, again, the original Fourier transform goes to 1 over the absolute value of a, a is positive, 1/a, F of s/a. But, again, here if a is less than 1, 1/a is bigger than 1, so the Fourier transform is squeezed. So if this is the original Fourier transform, so to speak, absolute value of F of s, then 1/a is gonna stretch things in the vertical direction, because 1/a is bigger than 1 and it's gonna squeeze things down. So maybe something like that. So it's squeezed and stretched. Now, think of the extreme cases here. Think if a is getting bigger, bigger, greater, greater than 1 or if a is getting less and less than 1, what does it mean? Let's take the case when a is bigger than 1, much bigger than 1, it's getting larger and larger. If a is getting larger and larger, then the signal is getting more and more compressed, it is getting more and more localized in time. On the other hand, what is happening to the Fourier transform? If a is getting larger and larger, the Fourier transform is getting more and more stretched out, and for that matter, also squeezed down in the vertical direction.

This is an example, an extremely important example, of this sort of reciprocal relationship. And I use that term pretty broadly here. I mean, you see reciprocals coming in, but sort of in words and in feeling you can't have a signal, which is both localized in time and localized in frequency. To localize the signal in time would be to squeeze it down. Like, look at the scaling of this F of at for a very large a. But that has a consequence and frequency of stretching things out. So if you want to concentrate the signal in time, the Fourier transform is gonna stretch out. And the reverse situation also holds. If you take a signal and stretch it out in time, by scaling by a number less than 1, you know, scale by 1/10, by 1/100, by 1/1,000, whatever, then you are stretching it out in time, but you're concentrating it in frequency. The Fourier transform is getting squeezed. That's a very important and a fundamental property of Fourier transform, of the picture and the signal, of the signal in the time domain and in the frequency domain, and it comes up all the time. It's a very important piece of intuition for you to have. And to me, that's somehow the most important consequence of the stretch theorem. This theorem is what it says about localizing the signal in time versus localizing the signal in frequency. A signal cannot be both localized in time and in frequency. This is one instance of this. There are other instances, actually, there are other ways of writing down that sort of relationship, of translating this sentence into more precise mathematical terminology.

One of the most famous ways of translating this is the Heidenberg Uncertainty Principle, that you can't localize a particle both in space and in time. You can't know both it's velocity and its position to arbitrary precision. The Heidenberg Uncertainty Principle can be proved, and there's a derivation in the notes, actually, by taking you through the Fourier transform, by using the Fourier transform. And it essentially comes down to this fact, that a signal cannot be both localized in time and in frequency. That if you stretch out of one domain you're gonna squeeze in the other domain, and vice versa. This is a very important point of intuition for you. Now, one interesting thing, if you think about what we showed about the Gaussian, we showed, as a remarkable fact, that the Fourier transform of the Gaussian is itself. So recall the situation with the Gaussian, F of t = e to the minus 2p t squared. When it's normalized this way, the Fourier transform is the same thing, F of s = e to the minus p s squared, same thing. So from interpreting the Gaussian, or putting this interpretation onto the Gaussian, you might say the Gaussian is perfectly balanced somehow in both time and frequency. It is what it is; it is perfectly balanced. It's just concentrated enough in time, and just concentrated enough in frequency or just spread out enough in time, and just spread out enough in frequency. So it's the same thing, taking the Fourier transform doesn't change it. So it's perfectly balanced. I don't know how better to say this, perfectly balanced in time and frequency.

So these things are in general use now and everyday use. You're gonna use these things all the time. People that work in signal processing can just whip these formulas out without any problem, and they are quiet familiar with these interpretations. Well, the formula that I won't write down for you, but is, again, in the notes, and there's some examples of it, and I think there's some problems of it, are what happens when you combine shifts and stretches? The formulas get a little uglier, there's nothing new involved in that, but it's just a question of combining the formulas in the right way. And part of it, actually, is a question of the variables, and how you understand parenthesis and everything else. So I'm not gonna write down that formula, because unless I practice it a couple times, I'll get it wrong. But that's the one sort of formula in this general circle of ideas that also comes in is, the formula for the Fourier transform when you both stretch and delay, so F of at + b, something like that. So I'll let you struggle with that quietly.

So one more, go out with a bang today, and that is to introduce the idea of convolution. Again, something that you've probably seen in other classes, something that we've already seen in the context of the Fourier series. But now we want to see convolution in the context of Fourier transforms, and the way that it arises in the context of Fourier transforms, which is different, most naturally different, than the way we saw it first coming up in the context of the Fourier series. There we saw it in the context of solving differential equations, and that same interpretation, or that same use of convolution, again, comes up. But that's not the way it's motivated, and that's not the most fundamental role it plays, at least for the questions of signal processing. So convolution is probably, I think it's probably safe to say, it's probably the most important operation in signal processing. I think that may be a little bit of an exaggeration, but not too much. So I'll say probably. Most frequently used, most flexible important operation in signal processing. Now, what is signal processing, since I've used that phrase, or what is the fundamental question of signal processing? Signal processing, very broadly, can be said to – you might define it as saying, how can you use one function to modify another? How can you use one signal to modify another? That's the basic question of all signal processing. Don't lose sight of this, is how to use one function to modify another or one signal?

You can ask this question or you can take this approach in either the time domain or the frequency domain. And, again, it's equivalent because we can go from one to the other by the Fourier transform, or the inverse Fourier transform. I'd say it's probably more common to address this question in the frequency domain. That is to say, when you talk about modifying a signal, most often, not always, but most often you're talking about modifying the spectrum of the signal. So most often you look to modifying the spectrum of a signal. And then at the end of the day, what it is you take the Fourier transform, you screw around with the spectrum, that's the technical term for it, and then you take the inverse Fourier transform and you modify the original signal somehow. But the operations you perform are often conceived of and carried out in the frequency domain, and then you go back and forth between the two domains, again, by the Fourier transform or the inverse Fourier transform. Now, there's a very simple example of this. For example, linearity actually gives you, in some sense, that's the simplest example of signal processing, or simplest operation of signal processing, e.g., linearity or super position. That is the Fourier transform of F + G is the Fourier transform of F plus the Fourier transform of G. So think about F as the signal that you want to modify; think of the signal G as the one that's doing the modifying. If you don't like the spectrum of F, fine, add something to it that takes away a little bit here, it takes away little bit there, adds a little here, adds a little there, and then go back. So you modify F, F, the spectrum of F, by adding the spectrum of G. You don't know what G is, but then you can take its inverse Fourier transform and find out the function that does the modifying for you. Usually, think of linearity as sort of the simplest property of Fourier transform that follows from the linear to the integral, the integral to the sums, the sums to the integrals, and you don't think it's really worth commenting on. But in some sense, it's the simplest and most basic operation of signal processing; you don't like the spectrum, add something to it. Viewed in that light – and of course I can also scale by a constant, that's even less interesting in some sense. That's the other part of linearity is I can multiply by a constant here. But, anyway, viewed in this light, a very natural question is: What if I want to multiply? So what about multiplying? That is to say: On the frequency side, I have my original function F and I want to multiply it by this Fourier transform and other function G. So I'm scaling each individual frequency, so to speak, each individual point in the spectrum by some function G that I want to construct, that does what I want it to do. And the question is: How does that come about? That is, is there some combination of F and G in the time domain, so that in the frequency domain, so within the spectrum, the spectrums multiply, the spectrum of F is multiplied by the spectrum of G. The values of the Fourier transform of F are multiplied by the values of Fourier transform G. Can I do that? Is this F of some F combined in some unnatural way with G? That's the mystery. Can I find a way of combining F and G so that the sequence of the Fourier transform is multiplied? Not so unnatural. Not so unnatural. So to answer this question, we're going to adopt a now familiar strategy. You've heard me say many times, this is how mathematics gets applied. Suppose the problem, solve, and sees what has to happen. So suppose the problem is solved – now, I want to be careful, I want to look at my notes because I want to use the same variables that I use in my notes so I don't confuse the issue, if you go back and forth between what I'm saying in class and what I say in the notes. So like I said, suppose the problem's solved? So I look at the Fourier transform of G at s times the Fourier transform of F at s. And I just play with this. What happens if you multiply the two spectrums? Well, the Fourier transform of G is the integral from -8 to 8. I have no recourse here but to go to the definition. Then -8 to 8 either to the minus 2pi, s of t, st, G of t, dt. And the Fourier transform of F at s is the integral from -8 to 8. And here I want to write this – I already used t, integrative perspective t and so I want to integrate with respect to another variable, and you'll see why in just a second, e to the minus 2pisx, F of x, dx. All right, I think I can take it from here. That's all I wanted to check.

Now, watch this. This is the product of two single integrals, they are decoupled, so to speak. So that is I can write this as one big double integral, e to the minus 2pist, e to the minus 2pisx, G of t, F of x, dt, dx. Okay. There should be no problem, you should be able to see that. And if you're having trouble seeing it, instead of going from here to here, imagine yourself going from here to here. The variables are decoupled. When I say the variables are decoupled here, I mean they're decoupled. I don't know how else to say it. There's no relationship between the variables; everything happens independently. Now, combine the complex exponentials. That is, write this, integral from -8 to 8, integral from -8 to 8. By the way, there are issues here. I mean, there are things about turning – I'm gonna swap limits of integration, I'm gonna swap dtdx, dxdt, all the rest of the rest of that stuff. As I have said before, the rigger police are still blessedly off duty. All right. So this is, e to the minus 2pi, I have a common factor of s in both exponentials, and I get a t + x. G of t, F of x, dtdx. I haven't done anything there except algebraically combine the two exponentials. Now, I want to do something a little trickier, I'm gonna write this double integral as integrated single integrals. So now the variables are coupled, really. So I'm gonna write this as the integral from -8 to 8, the inside integral, integral from -8 to 8, e to the minus 2pis, t + x, G of t, dt, and then the result is integrated against F of x, dx. Now, let's check to make sure that everything's okay here, that is actually get a number at the end of the day.

On the inside integral I have a function – this kernel here involves both, well, st and x. I'm integrating with respect to t. So what remains is a function of x, forget about the s that comes on sort of in the end, but think about this as a function of t and x because those are my variables of integration. So I've integrated this with the respect to t, this inside integral. What remains is something that depends on x. I multiply that by F of x. I integrate with respect to x. I'm done, everything's okay. So the only thing that remains here is the function of s, and that's the product of the Fourier transform. So that's what you expect. Now, once again, what this calls for, if you can only hear it calling, I'm training your ears and your eyes, is a change of variable. So I'm gonna let u = t + x, so du = dt. See, the inside integral is with respect to t, so that's the variable here, u = dt, and then t = u - x. So what happens to the inside integral? And, again, if t goes from -8 to 8, for any fixed x, so does u go from -8 to 8. So I get the integral from -8 to 8 of the integral from -8 to 8 of e to the minus 2pis, t + x is u, G of t, which is u - x, du then integrate with respect x, F of x, dx. Cool. It may seem pointless, but it's not. Now, I'm gonna swap the order of integration. Instead of integrating du, dx, I'm gonna integrate dx, du. And I'm gonna write this as the integral from -8 to 8. The inside integral becomes the integral from -8 to 8 of G of u - x, F of x, I'm skipping a step here a little bit, but not much, times e to the minus 2pisu, du. Is everything okay there? Let's check. I skipped a step here, I sort of put everything together. I could have taken one more step and put everything together, and then pulled it apart, so I have, again, the two integrated single integrals. This integral with respect to x, and then what remains – again, I integrated this with respect to x, what remains is the function of u. I integrate that against this kernel, this complex exponential, e to the minus 2pisu, with respect to u; what remains is the function of s. Want me to say that again? No. All right. So I skipped a step there. I took that integrated integral, I could have written everything together, grouped everything together, and then split it up again into to integrated integrals. All right. Now, look, say, we have eyes if only we but see. This integral inside here is a combination of G and F, and is getting integrated against a complex exponential. That is, if I now am so bright, I'm so farsighted, if I am so brilliant, I will now realize that I have solved my problem. I'm gonna define h of u to be the integral from -8 to 8 of G of u - x, F of x, dx. Right? That's what happened to the inside integral, u - x, yeah, F of x, dx. And then what I have is the integral from -8 to 8 of either the minus 2pisu, h of u, du, which is the Fourier transform of h evaluated at s.

Now, once again, where do we start? Where do we finish? I discovered that the Fourier transform of G at s times the Fourier transform of F at s, is the Fourier transform of h at s. The spectrums multiply if I define h by this formula. If we define h of u = the integral from -8 to 8 of du - x, F of x, dx. I solved my problem. Once again, what was the problem? I said is there a way of combining F and G in the time domain, so that in the frequency domain the spectrums multiply, the Fourier transforms multiply? And the answer is, yes, there is, by this complicated integral. It's not so obvious. You wouldn't get out of bed in the morning and write this down, and expect that this was going to solve that problem. How do we get it? You said, suppose the problem is solved, see what has to happen, and then we'd have to recognize when it's time to declare victory. And it's time to declare victory.

Now, being an obnoxious mathematician, you cover your tracks and say, "Well, I'm simply going to define the convolution of two functions by this formula." Let me use a different variable here to write the definition down. That's how mathematics works, you turn the solution of a problem into a definition. It irritates the hell out of some people, but that's how it works. So define the convolution of F and G by this formula, say and the special notations use G * F. Well, let me call the variable x here. So it's the integral from -8 to 8, G of x - y, F of y, dy. Define the convolution of that? And then if you're a mathematician, you'd say, "Well, sure, I've defined that. What can be more natural." And then you've proved the theorem. Then you proved separately, you proved this remarkable fact, that the Fourier transform of the convolution is the product of the Fourier transforms. But what's missing in that derivation is, the game is rigged. I mean this definition was come upon precisely so it would satisfy that property, so that the Fourier transform multiplied the two signals, so the Fourier transforms multiply.

Then we have the Fourier transform of the convolution, G * F is the Fourier transform of G times the Fourier transform F. Remarkable. But as I said, the game's rigged. I mean, I set out to solve that problem, and I solved that problem, and then I sort of covered my tracks by saying, "Well, define the convolution by this, and then we have the remarkable convolution theorem, that the Fourier transform of the convolution is the product of the Fourier transform." But of course you do because you set it up that way.

Now, I don't want it to seem like we've done something insignificant here, we've done something extremely significant. Because the idea of being able to modify a signal by multiplying the Fourier transforms is a very powerful thing to be able to do. And what this says is, there is a way of combining the signals in the time domain by a not so very obvious thing. It's a nice enough looking formula, it's sort of an elegant looking formula, but it's not obvious that this would work. And if you combine the signal this way in the time domain, in the frequency domain the Fourier transforms multiply. How cool is that.

All right. And I think on that note of coolness, we'll wrap it up for today, and then next time I'll talk about some of the properties of convolution and why it is so important to your everyday lives. See you then.

[End of Audio]

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture09

Instructor (Brad Osgood):Oh, I'm on. What a surprise. All right. Did you get the word back in the back control room that I want to show a couple pictures today? Move the camera up and down if you want to say, "Yes." Very good. All right.

Okay. Today we're gonna continue with our study of convolution. And let me remind you of the star of the show and how we got there. So last time we introduced convolution in a, what I hope you thought, was a natural way to answer a reasonable, natural question in signal processing. So we talked about how do you combine two signals in such a way that their Fourier transforms multiply. We are led to convolution by asking for the Fourier transform of a combination of f and g is the product of the Fourier transform. So the Fourier transform of f times the Fourier transform of g. And what we found was that the combination was certainly not obvious, but actually quite least compactly written. So the answer was given by, the convolution of two functions. I can either look at the convolution of g with f or f with g; it doesn't matter. The convolution is the integral from -8 to 8 of, I think just to be consistent with how I wrote it last time I believe, g of x minus y, f of y, dy. That is to say, if this combination is defined, if this is the way you combine f and g according to this integral, then the Fourier transform of the convolution is the product of the Fourier transforms, which is quite a remarkable statement. I mean, all these operations are not to be taken lightly.

Certainly, the Fourier transform is a complicated enough operation, involving an integral from -8 to 8, a complex exponential, the rest of that stuff. This integral, although, it doesn't involve any complex quantities, is certainly, again, nothing to be taken for granted. And the fact that you combine these two complicated operations, and they combine in such a simple way is pretty impressive. And not only pretty impressive, it's pretty useful. In fact, before talking anymore about any general properties of the Fourier transform, let me give you an example of just this sort of thing in action. So let me give you an example of this in filtering. And I'm gonna take a particular – I'm not gonna spend a lot of time on this, but I just want to show you that I'm not making this up. Because we're gonna return to a lot of these ideas repeatedly throughout the course, and also similar examples, and sometimes study them in greater depth as we go through the course. The example that I have in mind, though, to start with was one that I borrowed from a book by Briggs and Henson on the discrete Fourier transform. I think I put this in a list of references that's on the website, and it's called something like The DFT: An Owners' Manual. It's very well written, and has all sorts of good examples and good problems in it. And the one problem they study is the problem – they use as an example, actually of filtering, is the study of terbidity. Now, I think we actually had some people in earth sciences in the class at one point, I think I remember that. Anybody know what terbidity is? Anybody from earth science in here today?

Terbidity is sort of a study of, I don't know whether it's a measure of the clearness of water or the murkiness of water, but it has to do with measuring the clarity of water. And the idea is, that particles are suspended in water, and light scatters off of particles, and you sort of measure how light scatters; that's a measure of the murkiness. It's a measure of the more particles, the more scattering and the more murky it is. And terbidity varies over time. And one of the problems is to study how it varies over time. So they presented the results of a study of the terbidity of the subglacial waters in the Yukon territory of Canada. So you get a picture that looks something like this. I'll show it to you. Can we get a shot of that? You want to put it here? Let's see how that is. So that's a picture of terbidity. Now, the scales here aren't so important. The horizontal scale is time, and I think it's over a period of months. And the vertical scale is terbidity, whatever that means. Now, so you put sensors down in this very deep subglacial water and you measure the murkiness of light, using whatever scale and whatever techniques are involved, and it oscillates over time.

Now, in this picture you see not much. So you see a couple of examples. You see a generally periodic phenomenon, but you see a lot of jaggedness in there, you see a lot of jaggedness in the picture. So, like I said, the horizontal scale is time, I think it's a period of months, and the vertical scale is whatever it is. And you certainly see a periodic phenomenon here, but it's noisy or it's jagged. And you want to get rid of the jaggedness of this. Now, the way to do this, the way to get rid of the jaggedness, the way to smooth out the data a little bit, is to do it not in the time domain, as you see it presented, but to do it in the frequency domain exactly in the way that we were talking about. So the first step in the analysis of this data, or in the smoothing of this data, is to take the Fourier transform. Now, in fact, the data is given to you in discrete form. So what you're actually taking is the discrete Fourier transform, and we're gonna get to that shortly, but think of everything here as just sort of continuous, and think about actually taking the Fourier transform by whatever means you have.

And if you do that, you get a picture that looks like this. So this is a picture in the frequency domain; this is a picture of the Fourier transform of the signal. As a matter of fact, it's two pictures. The first picture is a set of frequencies going all the way out. And I'm only drawing the positive frequencies here. When you take the Fourier transform, of course for a real signal, you have positive and negative frequencies, but ones a complex conjugate of the other. So actually what's being plotted here is the magnitude of the Fourier transform only for the positive frequencies. And you see it goes all the way out. The high frequencies here – and this is just a section of it just showing the first 40 or so frequencies. Again, high frequencies are what's causing the jaggedness. Just as in the same case with Fourier series, the high harmonics are causing the signal to oscillate quickly, well, the same thing with the Fourier transform. Although, the spectrum is continuous it's the same principle; high frequencies are causing sharp oscillations, or rapid oscillations. So what do you want to do to smooth it? The natural thing to smooth it is to just kill off the high frequencies. Now, how do you kill off the high frequencies? The simplest way of doing it is by multiplying by rectangle function in the frequency domain. Okay. So you kill off high frequencies by multiplying, in the frequency domain, by a scaled rectangle function. That is to say, if the picture in the frequency domain is something like this, where the frequencies are going all the way out, then you just multiply it by a function, which is one, my rectangle function. Functions, which is one up to a certain point, say a cutoff frequency, from minus new c to plus new c, let's think of that as the cutoff frequency, and then a zero outside that. So you eliminate all frequencies below minus the cutoff range and above plus the cutoff range, and you keep the frequencies in between.

The other way of putting this is you are passing the low frequencies, or eliminating the high frequencies. And so this is called a low-pass filter, it passes the low frequencies. So if you do that, then in the frequency domain the result is to take a rectangle function p, I guess the way I've scaled it I'd represent it as p sub 2c, it has total width 2 new sub c, new sub c is supposed to be the cutoff frequency here, times the Fourier transform of the terbidity signal, whatever you want to call that, t. This is Fourier transform of the terbidity. That's what it looks like in the frequency domain. What does it look like in the time domain? Back in the time domain, you'd take the inverse Fourier transform of this or you ask yourself what convolution leads to the product of these two functions in the frequency domain. And we know what that is. So in the time domain this is convolution. It would be, I believe, 2 new sub c sinc 2 new sub c times t convolved with the original terbidity function t, whatever you want to call it, T of t. It's convolution in time, multiplication in frequency. And the result is, in this case, is to kill off the high frequencies. And I'll show you what the picture looks like. I'll show you what that picture looks like for two cases. One is if you keep I think the first 40 or so frequencies, and the other, I'm gonna show you the graphs in just a second, and the other I think of is if you just keep like the first 10 or 15, and you get two different pictures. This is, once again, the picture in the frequency domain. The picture in the time domain looks like this.

So what you do is, again, you carry out this multiplication in the frequency domain and then you just take the inverse Fourier transform, or you know what the result is gonna be so you just compute the convolution. It has to be computed numerically, of course; everything here inside is actually discrete data. So in the first case, here, this is I think only keeping, like, the first 10 or 15 frequencies, something like that, I can't remember exactly what they did. I'm sorry. I didn't have a chance to look it up this morning to get the precise cutoffs that they used. And this one is keeping maybe the first 20 or 30, maybe up to 40 frequencies. So you see you still have a certain amount of jaggedness in here. Here you see, very quite strongly, that the thing has been smoothed out and you can see the periodic nature of it. So this is an actual study of actual data. By the way, they rescaled here. The scale on the vertical axis is different because I think they just subtracted off the mean, so they get it to oscillate around zero instead of oscillating around whatever it was. That's why the scale is different here on the vertical axis.

All right. Now, there's a serious question involved when you're applying these techniques to real data, where something real is at stake. Namely, you might say that, let's go back to the original picture, there's a lot of noise in here that doesn't belong. Some how it's a concept of our sensors or whatever, result of a faulty experimental technique, that I'm getting all these extra jagged edges in there. You know, who knows why that's happening. So I want to filter those out. And I can filter them out very dramatically, and to get this picture, by only keeping the first however many frequencies or I can flip to the model a little bit less dramatically by allowing a certain number of high frequencies to creep in. And the question is, when are you filtering out something essential, and when are you filtering out just noise? That is to say, when are you presenting the real genuine physical phenomenon to focus on what should be understood, and when are you committing scientific fraud. Each of this must answer this question in his or her own way, I suppose. But that's the issue. You are – the original signal, everybody would believe, is imperfect. I mean, it has something attached to it, it has something that goes along with it, that shouldn't really be there. How much do you take away? That's the question. You have a lot of power when you have these mathematical techniques at your disposal. And the question is, use them wisely, young ones. So I'll show you more such examples as we go along, but I didn't want to talk about anymore sort of general properties of convolution, which I'll turn to now, without showing you how it looks in action. And you can look – I don't know whether we'll come back to this particular example when we do the DFT, but I'll refer you to it again when we talk about it. Because they actually have the, in the Briggs and Henson book, I think they have the data and they have a little bit more details about this. It's nice. And this is just one example of many possible examples.

As a matter of fact, let's stay on the subject of filtering for just a second here. Okay, we are down now with this. You may raise the screen. Okay, thank you. Filtering, or what's called filtering, is probably one of the main uses of convolution. And just in the kind of form that we were looking at. All right. You want to eliminate some frequencies, let some other frequencies go through. You do that in the frequencies domain, and then the question is: What are the consequences of that in the time domain?

And there's a little terminology that goes along with this, and, again, we'll come back to this topic a little later on in the class, but it's probably worthwhile saying it now. Many of you have, no doubt, heard this terminology and studied different aspects, different kinds of filters in different classes. And, again, we'll also come back to it. But let me just say a little bit now. Filtering is often, not always, but often, almost synonymous with convolution. There are reasons for that so-called time invariants or spacial invariants of convolution as it's associated with the filters. This is not always the case, but it's, like, almost always the case. And the idea is that the filter is defined by sort of a fixed function that you're convolving with or, in the frequency domain, a fixed function that you're multiplying the Fourier transforms with. The inputs vary, but the filter function stays that same. So you imagine a system is a system that convolves an input, which can vary, you know, one input, another input, another input, with a fixed function – or fixed signal. And the fixed signal is called the impulse response. Again, for reasons which we will understand a little bit more when we have a little bit more information about delta functions and linear systems, and so on. But there's no reason why you shouldn't learn the words now. That is to say, a filter, when it is given by convolution, is of this form, say, g is equal, let me use different [inaudible], f convolved with h. So the idea here is f is the input that can vary, you put different inputs into the system, h is fixed, and that's called the impulse response, and then what results from that is the output. Now, again, that's the picture in the time domain.

The picture in the frequency domain is what you really think of most often when you're designing filters to accomplish a certain purpose. Because the picture in the frequency domain is much simpler, it's just multiplication. So in the frequency domain, that is to say, taking the Fourier transforms, let me use the uppercase notation here, you write this as, say, G of s is equal to capital F of s times capital H of s. And in this context, capital H of s is always called the transfer function, and it's always written as capital H. I don't know why, but it always is somehow. So capital H is called the transfer function, sometimes called the system function. I am a little hesitant here, actually, because I'm not sure sometimes whether this terminology applies to the time domain or the frequency domain. Certainly, in the frequency domain, it's called the transfer function. I was gonna say it's sometimes called the system function, but I'm not sure if that refers to capital H or little h…transfer function. So to design a filter, then, is often to design the appropriate transfer function, to think about things in the frequency domain. And there's this, it's an art as much as it is a science. To design a filter is to design H, the transfer function. All right. And then let nature take it's course. Multiplying the frequency domain, convolve in the time domain. So, for example, the low-pass filter is a very simple cutoff. Low-pass is multiplied by a rec function. I won't specify the width here. But the idea is just multiply it by a scaled rectangle function of whatever the appropriate width is. And, again, the height is one here. So I'm multiplying just by one, so I'm not changing it all in the range where the function is one, where the function is non-zero, and then it kills it off completely outside that. Now, the problem with a low-pass filter, this is called the ideal low-pass filter, because it's a sharp cutoff. It cuts off exactly at the frequency.

Now, you can achieve that, actually, digitally; you can't achieve that in analog form. You can't wire this into a circuit that's gonna give you a sharp cutoff. So what people sometimes do is they have a gradual roll off. There are also consequences to cutting off very sharply, as opposed to cutting off sort of more gradually. And we will talk about some of these things. Although, they get very specialized, and it gets very, like I say, gets into sort of high art and the occult. So we're only gonna go into it to a certain extent. But it's certainly simplest to think about the ideal cases, some of the ideal low-pass filters, when you just multiply it by a rec function. And then the consequence in the time domain is just convolving with a scaled sinc function. It's not hard to say, certainly, and at least in the frequency domain it's not hard to see what's going on. Other possibilities, again, without spelling it out, is the high-pass filter. A high-pass filter would be to pass the high frequencies and filter out the low frequencies. So, for example, why would you want to have a high-pass filter? What's an example where you would want to only keep the high frequencies and eliminate the low frequencies? Actually, I'll give you a hint, it comes up a lot in imaging problems.

Student:

Edge detection.

Instructor (Brad Osgood):Edges, edge detection, exactly. And in an image, we're gonna talk about higher dimensional Fourier transforms and so on, but just imagine that edges are determined by a very rapid change of either light or dark or some rapid change in the picture, in the image. And that's characterized in the frequency domain, in the spectral picture, by very high frequencies. So if you want to emphasize the edges, whereas just sort of a placid scene, there's not much variation in the shading, not much very variation in the intensity. And edge is a very rapid variation in intensity, say, from black to white.

Whereas, just the ordinary scene, this desk or something like that might not have much of a variation. So if you want to emphasize the edges, I mean, if it doesn't have very much variation that would be typically low frequency, you want to kill those off, and then emphasize only the high frequency. That would detect the edges in an image. What does the transfer function look like for a high-pass filter? What sort of function would I multiply the spectrum by to keep only the high frequencies and to kill off the low frequencies?

Well, to do that, once again, to keep the high frequencies but to eliminate the low frequencies, I would have the function be one from a certain stage on. And I'm keeping everything symmetric, again, because remember mathematically we have both positive frequencies and negative frequencies. So, again, there's sort of a cutoff frequency, new sub c and minus new sub c. The function would be one going out ideally to 8 from new sub c, and one down here out to -8 and it'd be zero in between. You can easily write down a formula for that function. You can take a couple of rectangle functions and stretch them, and subtract them, and do all sorts of stuff with them, it's not hard, I won't do it, but I actually do it in the notes. There's some extra complications that come into this because this thing doesn't have such an easy Fourier transform. Actually, delta functions come into this. Although, the transfer function looks pretty simple, the affect in the time domain is a little bit more complicated. And we do not quite have the technology yet to deal with it. But at least at an intuitive level, understanding what you want to do, it's easy enough to draw the picture.

And another possibility would be a bandpass filter, where you pass a range of frequencies and you eliminate all of the frequencies outside that band. And, again, it's not hard to draw the picture of what that should like in the frequency domain. If you want to pass a band of frequencies then you want to multiply by a shifted rectangle function that has only a certain extent. And, again, because the frequencies mathematically are both positive and negative and symmetric about the origin, I multiply by the complete – the transfer function for a bandpass filter will look something like this. So, again, it has height one, so I'm just multiplying by one within a certain band of frequencies, whatever they are, I won't label the axis here, here's zero, and zero elsewhere. So I multiply the Fourier transform, or the desired signal, the filter that I want to signal, by a function that looks like that, that keeps only the frequencies in a certain band of frequencies and it kills off the rest. And then I take the convolution in the time domain by whatever function has this Fourier transform. And that's not hard. You did a homework problem on the modulation theorem. You know how to get the Fourier transform, or the inverse Fourier transform, of a signal that look like that. Not so bad, and it's extremely important. The whole idea of filtering, the whole idea of computing convolutions in the time domain to see what happens to the signal, whether discretely or analog, is a big industry. So for right now, actually, I'm not gonna say too much more about filters. Some explicit formulas are given in the notes. But the main idea that I wanted to get across was, really, the ease of it, and the ease of it when you think of it in the frequency domain. It's not so easy when you think of it in the time domain. And that's really the next thing I wanted to talk about.

So, I mean, of course you could be more or less sophisticated, but at least at the level we've been talking about, which really covers the essential ideas, it's easy to understand filtering, or what you're trying to get at, that is to say, convolution in frequency; not so easy in time. And that leads me to the next important thing I wanted to say about convolution in general. Now, I do want to talk a little bit about convolution in general, some of the properties of convolution in general. And I guess the first one is: How do you interpret convolution? Well, now, before I – let me talk about visualizing. So it's not so easy in time. So to see what happens to filtering in time, you would need to – to do this you need to visualize f convolved with h, where h is the given impulse response, the Fourier transform of the transfer function, or the inverse of the Fourier transform of the transfer function. You need to visualize convolution.

Now, I don't recommend this. I felt like I had to say a little bit about it in the notes. But many books spend many pages, and probably insist on you spending many hours, on trying to visualize the convolution of two given functions. And the phrase you here is flip and drag. I mean, I think it is an idiotic waste of brain cells and time to sit in a dark room quietly trying to visualize convolution. Remember when I said it was idiotic to try to visualize when two functions were orthogonal, in terms of the inner product? Well, I think it is equally idiotic to try to visualize convolution. I think the way to visualize convolution, if there is a way, is to think in terms of multiplying in the frequency domain. I mean, one of the things you start to build up is a certain amount of intuition about what Fourier transforms look like. And it's not so hard, or it might not be as hard, to visualize the product of a couple of Fourier transforms, and then, again, maybe if you know what the spectrum is like you have some sense of what the signal is like. But you tell me the truth, you be honest with me. Do you think that you can really visualize the convolution with the sinc function? I mean, for the low-pass filter the product in the frequency domain is very simple. It's the multiplication of the rectangle function times the signal. In the time domain it's the convolution of a sinc function with a signal that looks like sinc of x minus y, f of y, dy for a given function f. Now you know what the sinc function looks like. Are you trying to tell me that you can flip and drag this thing and visualize this thing? I don't think so. So don't even try. Let me just say, hard to visualize; a challenge. So I like to think that I have allowed you to put that burden down; don't do it. Now, if you can visualize, though, a fair question is can you interpret it. Is there sort of an interpretation, a natural interpretation of convolution that will lead you to know when to apply it, when you should expect to see it, what sort of features you should expect to see. So you can't visualize convolution easily. Is there a good interpretation? All right. How do you think about convolution, what is convolution really? I mean, you could write down the formula as the integral, but, you know, how do I think about that?

Now, here, too, I want to offer you some advice, but I also want to exercise a certain amount of caution in this. Convolution is a really pretty general operation, and it comes up in a lot of different ways. I think it would be a mistake to try to attach to convolution a single interpretation. You see that sometimes, and people try to do that, but I think the fact is, that it's one of those things that you get used to using, and you use it in different ways, and, consequently, you interpret it in different ways. I think the best thing to say is, "convolution is what convolution does." And you get used to using it in so many different ways that you will automatically somehow attach the appropriate interpretation when called for in the appropriate setting. So it's used in many ways. It's not subject to a single interpretation, I would say. And you do yourself no favor if you try to peg it only one way. I think it's somewhat analogous here, and I think I may have mentioned this earlier in class, to the idea of a definite integral. I mean, you learned the definite integral, when you were learning calculus, by typically a simple motivating problem, like area under a curve, something like that. But you don't always think of the integral as the area under a curve. If you always try to think of the integral as the area under the curve, you do yourself no service because in some cases, in some problems, where the integral is called for, it's not called for in the context of applying the area under a curve; it's called for in some other different context. Well, the same sort of thing happens with convolution. It's not always called for in the one, it may be called for in different context. So to try to attach one interpretation to it I think is a mistake. If you use it often enough, if you use it in a lot of different settings you get very used to it, and you get very used to sort of thinking about it and thinking about it in different ways.

Now, I do, however, want to offer a maxim that is often helpful, not universal but often helpful, for the way convolution comes up. So I feel like I'm retreating a little bit from this strong statement that it's not something to an interpretation and you shouldn't really think about it that way. I think it is fair to say that in many contexts convolution is interpreted or arises in the context of smoothing or averaging. Context convolution is associated with smoothing or averaging. Now, even that is not, again, universal. The low-pass filter smoothes; the high-pass filter does not smooth. But, actually, the difference between the two mathematically is the low-pass filter involves convolution with a function; the high-pass filter involves convolution with distributions, or delta functions, and that's not a smoothing operation. But convolution with a function is often associated with smoothing or averaging. Eliminating the jaggedness in data, like we did with the terbidity, can be thought of as smoothing the data or it can also be thought of as averaging the data. You replace a sharp jump by an average value between the two jumps. And we're gonna return to that when we talk about systems.

So, again, even this has to be qualified somewhat. But, again, we don’t have this sort of technology yet at our disposal to make that too much more precise. Although, I will say a little bit more. But, again, we're gonna see different aspects of this all throughout the different topics that we talk about in the course. Almost everything we do is gonna somehow touch convolution or vice versa. It really is that important an operation in the whole context of Fourier analysis. But in general, I'd say, if you're looking for sort of an aphorism, if you look at the convolution of two functions, f convolved with g has, together, the convolution has the best properties of f and g separately. Or you might even say that f convolved with g is at least as good as f and g separately, and it's often better; f convolved with g is usually smoother than f and g or necessarily be separately. Like all aphorisms, there are exceptions to it; it only holds when you're talking about functions. The cutoff is when you convolve with a delta function, where nothing changes at all. But for functions, the convolution of two is generally smoother than each. I'll give you an example of this. I'll give you several examples of this. One example is if I take the rectangle function and I convolve it with itself, I get the triangle function. Now, there's a problem you have to work on this actually with a scaled rectangle function, where you actually have to compute this by evaluating the integral. As we actually ask to once in your life, and probably only once in your life, you should compute a couple of convolution integrals and see how it works out. I'm not asking you to visualize, I'm asking to actually compute the integral, and to show that according to the formula for convolution is integration, the convolution of a rectangle function with itself, or a scaled rectangle function, gives you a scaled triangle function.

Now, why do I say this is an illustration of this property, that f convolved with g is smoother than f or g separately, because the rectangle function is discontinuous; the triangle function is continuous. It averages it out. But you've averaged out that jump that the rectangle function takes and actually made a continuous function out of it. So these are discontinuous on the left-hand side and continuous on the right-hand side. And, by the way, from this formula and from the convolution theorem, and I promised this was coming, and I know I mentioned it in the notes, the Fourier transform of the convolution, of course, is the product of the Fourier transforms, and the Fourier transform of the rectangle function is the sinc function. So this is sinc squared. And that is a very rapid, very quick proof that the Fourier transform of the triangle function is equal sinc squared. That's the other reason why the Fourier transform of the triangle function is sinc squared. Now, whether or not that's really a simpler proof, I'm not so sure. You could calculate the Fourier transform of the triangle function by direct integration, and there's not that much involved in it. Whereas, discovering convolution, proving the convolution theorem, establishing by hand that the convolution of the rectangle function with itself and the triangle function, and then concluding that the Fourier transform of the triangle function is sinc squared, well, that's a little bit of a long root. Even as fast I talk, it took me a long time to say. So whether or not it's a simpler way of doing it, I don't know, but at least it's a nice sort of consistency check and it sort of explains why something like that should be true. Another example of this may be even a more striking example that comes up is with regard to differentiability. And, again, you have a couple of homework problems on just these sorts of properties of convolution, actually, about periodic functions, convolving periodic functions to produce periodic function, and so on. And all that is by way of getting you to think about the fact that what properties the individual functions have are inherited by the convolution or in some cases enhanced by convolution. Another is if, say, f is a differentiable function but g is not then the convolution is differentiable, and you can say what its derivative is. Actually, the derivative of the convolution is I put the derivative on f and take the convolution as f prime convolved with g. And same thing for higher derivatives; all sorts of really interesting, wonderful formulas and properties like this hold similarly with higher derivatives.

Of course, if both f and g are differentiable then that's fine, and then I can put the derivative on either one. But the idea here is that you can take a non-smooth function and convolve it with a smooth function and the result is smooth. And not only that, it tells you how to compute the derivative. The derivative of this new differentiable function, f convolved with g, is f prime convolved with g. It's nice. All this stuff is great. I am, of course, skating over a few things here. There are always issues that convolution is defined by an infinite integral; there are issues about convergence of the integral, and so on. Those are real issues, but, again, the rigger police are banned from this room until I let them back in. I will let them back in to some extent before too long, I have to. But, for now, just think of this formally, and get some practice, get some ideas with using the properties and using the formulas because it's really just great. All right. Let me finish up today, and we're gonna talk about more properties on Monday, more applications on Monday, but let me finish up today with another important area where convolution is applied. And, actually, it harkens back to the work we did with Fourier series, where we first met convolution there in connection with solving the heat equation for heating up a ring. And I want to show you how convolution, again, arises in the context of the heat equation, but this time we'll just do it along a straight line. And you'll see how quickly convolution leads to the solution of the equation. So I want to talk about convolution of differential equations. To do that I need a general formula here. We need what's sometimes called, "The Derivative Theorem for Fourier transforms." So this you can think of as a general property of Fourier transforms. And, again, I'm not gonna write out all the assumptions very carefully here. But it says this, it says if you take the Fourier transform of the derivative of a function, it is 2pis times the Fourier transform of the original function f of s. Let me put my variables in here. The Fourier transform of f prime at s is 2pis times the Fourier transform of f at s. If the function is differentiable then it has a Fourier transform. And, again, there's a certain amount of things here that I am sweeping under the rug, but that's the main thing to understand, is the Fourier transform turns differentiation into multiplication. If you're looking for an interpretation of this in words, some that you can repeat and mention to friends in passing, it would be that the Fourier transform turns differentiation into multiplication. This is a fundamental property of the Fourier transform, and is really one of the reasons why it comes up in a lot of different applications; beyond what we're gonna talk about right now, but it comes up quite often. And similarly for higher derivatives. The Fourier transform of the f derivative is 2pis to the n, ordinary power, this is the nth derivative, times the Fourier transform of the original function f.

Now, let me show you why that's true very briefly, and I'll only do it for a special case. It actually holds quite generally, but let me give you a derivation of just the formula for the first derivative in the case when I know the function f of t tends to zero, say as t tends to plus or minus infinity. As it turns out, you can actually do it more generally, but if you make this assumption, it's a very quick derivation. It follows quite easily from integration by parts. So how do I take the Fourier transform of f prime at s. And as you've heard me say many times, you have no recourse here other than to repeal to the definition. The definition of the Fourier transform is the integral from -8 to 8 of e to the minus 2pi st, f prime of t, dt. Well, look at that integral, if that doesn't call out for integration by parts then nothing does. It's something times the derivative of something else, so for God's sake, integrate that by parts. That is to say if I let u equal e to the minus 2pi st, and dv equals f prime of t, dt, then what happens if I integrate this by parts. If I integrate this by parts then I get f of t times e to the minus 2pi st evaluated between -8 to 8 minus the integral of vdu. So that's minus the integral from -8 to 8, f of t times du, gives me a minus 2pis, I'm differentiating with respect to t, times e to the minus 2pi st, dt. Now, this term is, by our assumption, that the function tends to zero ± 8. This term is gone, the boundary terms are gone, and all that remains is the integral. But the integral, that 2pis, the minus signs cancel, minus a minus, and that 2pis comes out of the integral as a constant when I'm integrating with respect t. So this is 2pis times the integral from -8 to 8, e to the minus 2pi st, f of t, dt. And so that's just 2pis times the Fourier transform of the original function. That's all there is, that's all there is to it. So it's not hard, but it's an extremely important property. So, once again, the aphorism that goes with it is Fourier transforms turn differentiation into multiplication.

Now, let me show you how we're gonna use this. I don't know if I'll quite finish this today, but I want to give you the setup. So let's go back to the heat equation, but this time I'm gonna consider the heat equation on an infinite rod, essentially, the real line. So I want to do the heat equation on an infinite rod. Once again, the heat equation says u of xt is the temperature at a point x at time t. And I am given the initial temperature. The rod is heated up to some initial temperature that's u of x at zero, that's say, f of x. No periodicity assumptions here or anything, right, it's just everything takes place on an infinite rod, so essentially the real line. And the problem is find u. So u is the temperature – u is governed by the heat equation. And the heat equation, or the diffusion equation, the heat equation says use of t is equal to one-half uxx, that's the one-dimensional heat equation, and the one-half there is just for the calculation, just for the constants. Ordinarily, for the general heat equation, there would be a constant in here depending on the nature of the rod and so on. Now, again, u is a function of two variables. On the left-hand side is the derivative with respect to t, on the right-hand side is the second derivative with respect to s. What I'm gonna do is I'm gonna take the Fourier transform of both sides of that equation with respect to x, with respect to the spacial variable. I want to take the Fourier transform in the spacial variable. This might be a little bit easier. If lowercase u is the original function, then the Fourier transform I'll call capital U of st. So, again, the t is sort of just tagging along for the ride. It's the spacial variable that I'm taking the Fourier transform with respect to. So what about the left-hand side, what is the Fourier transform of u of t? So that's the integral from -8 to 8 of the derivative of e to the minus 2pi st, the derivative ddt of u of xt, dx, by definition of the Fourier transform. So I'm taking the derivative with respect to t here, I'm integrating with respect to x, so this derivative comes out of the integral. That is, I can write this as ddt of the integral from -8 to 8 of e to the minus 2pi st, u of dx. I'm taking the Fourier transform with respect to x, and I'm calling it spacial variable x. The Fourier transform is e to the minus 2pi sx.

If I take the Fourier transform of the time derivative, that's this. I can pull the time derivative out because the only thing that depends on time here is this u. And so that is ddt of capital U of st. That is the Fourier transform of the spacial variable of the time derivative is just the time derivative of the Fourier transform, if you want to say it in words. So that's what happens to the left-hand side. What about the right-hand side. Well, the right-hand side of the heat equation I use the derivative theorem. I have two derivatives, the second derivative with respect to x. If I take the Fourier transform of the right-hand side then what comes out is a factor of 2pis squared. So the Fourier transform of uxx, again, with respect to the spacial variable, is gonna give me 2pis squared times the Fourier transform of the original function undifferentiated at st. Again, t is sort of going along for the ride here, it's the transform in the spacial variable that counts. So that's minus 4pi squared is minus one, so it's minus 4pi squared, s squared, capital U of st.

So plug into the heat equation, there it is. Plug into the heat equation ut equals one-half uxx transforms to ddt of U st is equal to minus 2p squared, s squared times U st. Now, look at this, that's an ordinary differential equation for capital U. It's a derivative with respect to time. This is a constant, as far as time is concerned. It's ddt of U is equal to a constant times U. What is the solution? The solution is U of st is equal to U of s zero, the initial condition, times e to the minus 2p squared, s squared, t. Well, anybody could solve that equation, your little brother and sister can solve that equation.

Now, what is Us zero? That is the Fourier transform of the initial condition. That's the integral from -8 to 8 of Ux zero, e to the minus 2pi sx, dx. Ux zero is the initial temperature distribution f. So that's the integral from -8 to 8, f of x times e to the minus 2pi sx. That's the Fourier transform of f, let's call that, say, capital F of s. So what is capital U in terms of all the data that I have? So U of sx is equal to F of s times e to the minus 2p squared, s squared, t. The Fourier transform of the right-hand side is the product of two functions, The Fourier transform is the product of two functions, the product of this Gaussian with the Fourier transform of the initial data. What is the solution? The solution is the convolution of the two. Now, you have to recognize, one knows, what happens to the Fourier transform of Gaussian, actually. That is the Fourier transform – I'll just give you this fact. The square root of one over the square root of 2pt times e to the minus x squared over two t. Well, you have to use the scaling theorem here, and that's all that's involved. The Fourier transform of this is this, equals e to the minus 2p squared, s squared, t. That uses the fact the Fourier transform of the Gaussian. So what is the actual form of the solution? The form of the solution then is U of xt is the convolution, this is so cool, it is f of x convolved with that function, e to the minus one over the square root of 2pt, e to the minus 2p squared, e to the minus x squared over 2t. Convolution comes into the solution of differential equations because often in solving differential equations, if you take the Fourier transform, differentiation becomes multiplication, multiplication in the frequency domain becomes convolution in the time domain on taking the inverse of the Fourier transform.

So in the time domain, I'll write this integral out next time more fully next time so you get the full power of it, it's the convolution of the initial data with what's called the heart kernel for the infinite line, or for the rod. And that's lightening fast how the heat equation is solved and how convolution comes into it in a very fundamental way. And that's it for today. For Monday, I'll say one more thing, we're gonna talk about the central limit theorem. It is a real gem and a real jewel in this class, I think, to see how convolution applies to that theorem. So read over that material very carefully over the weekend so I don't have to do a lot of background on probability. Thank you all. Bye.

[End of Audio]

Duration: 55 minutes

TheFourierTransformAndItsApplications-Lecture10

Instructor (Brad Osgood):First thing, a quick announcement – two quick announcements. The latest homework set is posted up on the Web, so you can get that at the usual course Web site.

Also, Thomas had to change his office hours just as – just for today; is that right? Okay. He'll be in Packard 107 today from 12:00 to 2:00 for those who wanna spend a little quality time with Thomas.

Any questions about anything? Anything on anybody's mind?

All right. Big day today. We're going to talk about – we're going to do our final application of convolution. I suppose I shouldn't say "final application of convolution" because it is the kind of operation that comes up repeatedly throughout the course. But sort of as the – as a last treatment of the kind of areas we've been talking about, I wanna talk about application convolution to the central limit theorem.

And this is one of my favorite topics because it just is so – it's such an important result, and it's such, in some sense, a surprising application of convolution. And the result itself is just so – I don't know. It has this air of mystery about it that it's a real – I think it's a real treat to see how it's – see how they – see how these ideas play out.

So I wanna talk about convolution and the central limit theorem. I will describe this, but it actually takes a little while to set up. We have to develop the appropriate language to get a precise statement of what the result is, and then understand how we're going to approach it. But this – the central limit theorem is, I say, one of the cornerstones of probability. And it's not only very important from a theoretical point of view from the development probability, but it's extremely important from the practical point of view. It has to do with the ubiquity of the bell-shaped curve, or why is it that so many things are distributed according to a Gaussian.

The central limit theorem, which, like all good things, has a three-letter acronym that goes along with it – the CLT somehow explains the universal – explains the bell-shaped curve – the universal appearance of the bell-shaped curve, that is to say the Gaussian in probability.

And there is a quote, actually, on this. There's a quote that's often repeated and connection with the central limit theorem. I put it in the notes, and I just wanted to read it to you. It's by G. Litmann, who's a French physicist.

He says something like, "Everyone believes in the normal approximation."

The normal approximation is another word for the Gaussian approximation for the bell-shaped curve.

Says, "Everyone believes in the normal approximation, the mathematicians because they think that with a," – excuse me – "the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact. It's got something for everyone."

Everybody buys it. Everybody believes it. And today, you will know why it's true.

Well, what does it say? Well, again, it's gonna take us a little while – a couple of iterations – before we get to a precise statement. What it says is, it says that most probabilities – I mean, this is sort of an intuitive or informal way of putting it. Most probabilities – some kind of average, really, is the way of thinking about it – in some kind of average sense are calculated or approximated, at least, as if they were distributed – as if they were determined by a bell-shaped curve – by a Gaussian.

The key word here, as it turns out in the precise understanding of this, is averaging. In an average sense – or averaging many outcomes or many outcomes contributing in an average sense to the final outcome are distributed according to a Gaussian.

Now, the picture of this, so – before I make that precise, the picture is something like this. If you have a Gaussian, we're going to be working with a number of standard Gaussians. I – we – in the past, we've used either the minus pi x squared. We're gonna use a slightly – we're gonna use a different – one that's scaled differently today. We've used either the minus pi x squared as a standard Gaussian. And that has the advantage that it's equal to its own Fourier transform.

For the central limit theorem, the normalization is to take – more standard normalization is to take – sorry, say p of x equals 1 over the square root of 2pi either to minus x squared over 2.

Now, that has the advantage – I'm going to be using a lot of terms today, by the way, that I'm not gonna completely define. I hope you will have read – you have read or will read the section on this so that you pick up some of the terminology that we're gonna use, like standard deviation, mean, variance, probability, distributions, etc.

If we went over all of those things, I wouldn't be able to get to the really core ideas in a short enough amount of time, sorry to say. So I'm counting on you to really pick up some of the background here on your own. And I wouldn't be surprised if you'd seen a lot of these things before. A matter of fact, I'd be surprised if you hadn't seen a lot of these things before.

So the reason why you take this as the standard Gaussian in this case is that it has mean 0 – mean or expected value 0, and it has standard deviation or variance 1.

And the way it comes up in calculating probability is, again, something that you have probably seen is the probability – the idea is that this is supposed to represent the distribution of all possible outcomes to a measurement or an experiment or whatever.

And the probability that a measurement lies within a certain range is given by the area under the curve.

So the probability that a measurement – that some sort of experiment or a measurement is between two numbers a and b is given by the area under the curve – the integral for a to b of either the minus – well, 1 over squared 2pi – either the minus x squared over 2, dx.

This has to be calculated numerically because you can't integrate directly to the minus x squared – either minus x squared over 2, or it does have – doesn't have a simple anti-derivative. That's the area under the Gaussian.

Now, there is a strong sense – actually, I haven't gotten to the precise name of the theorem, but even in the informal way we've been talking about it, there's a strong sense of universality of the central limit theorem. It says that most probabilities somehow are calculated or approximated by a Gaussian.

Every – in the average sense, about any – most probabilities that you're likely to run across – most measurements that you're likely to run across, if you average them all out, you find a – you find that they're distributed like a bell-shaped curve. Now, why should that be the case? Where's this element of universality coming from?

Well, let me give you some indication and actually show you why – not why yet, but at least give you some indication how convolution is coming into this by showing you a series of quite startling pictures, I think. So I'm gonna show some slides, if I may – some drawings, if I may. And I may because I'm the teacher. Sorry, I should've warned you beforehand. Hello, back there?

But first, a song.

So here's our friend, the screen, and there is the rectangle function – see 0 all the way up and all the way down and over. This is actually an appropriate lecture for Halloween time because of what you see – because of what you will see. Ah, the rectangle function. All of this buildup for the lousy rectangle function. Boy, that is impressive.

Now, you know what happens when you convolve the rectangle function with itself? You get the triangle function. A matter of fact, I even – we talked about that a number of times. I even showed this – I even showed that – I even asked you to do that in homework. And if you read this section, you've seen these pictures, but nevertheless, you have to see them in full glory.

So this rectangle function convolved with itself looks like a triangle function. Fine. And remember, that was already in the case – that was already interesting itself. That showed somehow that the convolution is some kind of smoothing or averaging operation – that the convolution of the discontinuous function of the rectangle with itself is continuous. It's smoother than the original function.

Now, let's look at the convolution of a rectangle with itself three times. It is smoother still, and it starts to look a little bell-shaped. You start to get a characteristic shape there that starts to resemble, in your imagination, the Gaussian. And if you take the rectangle convolved with a something – I wouldn't ask you to do that. This is all calculated with MATLAB, of course. I wouldn't ask you to actually carry out the integration and try to do this calculation. That's hard. I mean, it's actually a piece-wise defined curve, but it's getting smoother, and it's getting more bell-shaped.

And if you take the rectangle convolved with itself four times, you get something that is very decidedly bell-shaped.

Now, that's already, I think, kind of impressive. But after all, the rectangle is a pretty simple-looking function. There's nothing much to it. What's really spooky is if you start with something very much unlike a rectangle function – a random just assortment of – a random collection of measurement and just start to take in the convolution random function and start taking the convolution, you're gonna find similar results.

So for example, here is, again from the book – from the notes – here's a random signal. Now, what I mean by this is I just asked MATLAB to generate a bunch of random numbers in between – I think it's just in between 0 and 1, and then join them by straight lines. So just jumps around according to what the numbers are. So that's a well-defined – well-defined, so it's a random function. Somehow in the interval, it just jumps all the way around.

But okay. You can take the convolution of this thing with itself. And here's what you get. If you take the convolution of this with itself – I call it the function f, you get something that still jumps around, but it is now very much triangular, whereas the first one looked pretty random, pretty scattered all over the place.

This one, there's sort of a "it goes up; it goes back down again." Take the convolution three times, and you get something that looks like that. Now, it's still jaggedy. It's still – it has a lot of bumps into it, but look at it. I mean, it's getting kind of bell-shaped, and it's getting kind of smoother. And if you take the convolution four times with itself, you get something that looks like that – just four times.

Now, again, you can't see it quite so much on the resolution on the screen, but if I get close enough to it, I still see some – get some jaggedy edges, no doubt, no doubt. But they're getting smoothed out. And again, the overall shape of the curve is becoming more and more Gaussian.

Now, I don't know, but I really – it's pretty spooky. And I don't think there's any other word for it. I mean, something is really going on here that you just have to understand. I mean, it has to somehow be – it's just really spooky.

That's all good. Thank you very much. You can raise the screen.

Back to live action. So I wanna explain that. The content – the mathematical content of the central limit theorem is to explain that spookiness and to explain that universality that in the limit as something or other – as convolution is repeated or as we'll make the connection, as an averaging is taken – as an average is taken over more and more measurements, the result converges to something that is distributed according to a Gaussian. That's what we wanna get to.

So here is the setup. And again, it involves a certain – unavoidably, if you wanna get – if you wanna give a precise statement, that means you have to have precise language that goes into it. And I apologize ahead of time again for not defining all these terms. They're talked about more carefully in the notes, but I did wanna go under the assumption that you've seen these things so that we can get to the central point and not talk too much about background.

So there's the setup. The primitive notion here is a random variable and how it's distributed. Primitive notion – that is to say the first thing you define in order to talk about all other things is a random variable and how it's distributed.

Now, a random variable – let me break that a little bit more neatly. A random variable, and I owe this definition to Sam Savage, who's a friend of mine in management science and engineering – a random variable is a number you don't know yet. That means that it's an outcome of a certain measurement. You're gonna perform an experiment. You're gonna make a measurement, and then you're gonna get a number. But you don't know what that number is until you've run the experiment, so you somehow wanna keep the placeholder for doing the experiment.

So you call the random variable, generally speaking, by an uppercase letter and the actual measurement that you take by the corresponding lowercase letter. This is like Fourier transform but in reverse. So you call the random variable x and the measurement – that is, the value of the random variable, little x.

So again, intuitively, you think of X as the number you don't know yet. It's like measure the length of a pin. Measure the height of a person across an entire population. Or measure – or the random variable is height. The value of the random variable is the measurement of the particular height of the particular person. So it's a number that you have to compute by running an experiment – by doing a trial.

Now, what you are interested in is the distribution of measurements. And that's given by – or it's usually called the probability density function or the distribution associated with x. So you're interested in how the measurements x are distributed. The fraction of them within a certain range, the fraction of them between 1 and 2, the fraction of them between 2 and 3, and so on, and so on. This is given by a function p of x, which is non-negative. Or if I wanna indicate the [inaudible] on x, I may put a subscript in there – piece of x, capital X.

But it's sometimes called the probability density function or the distribution function or whatever. And it has a property that – not property, but by definition, the probability that the measurement lies in a range between a and b is the integral of p between a and b – the area under the curve.

So by definition, the probability that the measurements are gonna take a value between a and b is the integral of a to b of p of x, dx.

So whatever shape the curve is in – whatever shape the curve has, it's positive, I compute the area under the curve from a to b, and that gives me the probability.

Now, there's a special property of p, but only one, really two. One is that it's positive. The second is that the total of the area is 1. The measurement's gotta be somewhere, so the probability that it's between minus infinity and infinity is 1. So equals 1 – everybody gotta be someplace, right? Every measurement's gotta be somewhere. So that translates to a property of p, the integral for minus infinity to infinity of p of x; dx is equal to 1.

So again, in mathematical abstract treatments of this thing usually start by saying – by defining what you mean by random variable, by defining by what you mean by density function. And density function is just defined by the properties. It's greater or equal to 0. And its total integral is equal to 1. And then you define probability by this sort of integral.

And again, these things are probably familiar to you, and I apologize for not really talking about them – talking about the background too much. But I do wanna get to the punch line.

Now, here is the key result. Here is how averaging starts to come into this. Averaging can really help convolutions start to come into this. Here's how averaging and convolution, I should say, are joined in this study. So here's the thing that gets it – that launches the whole argument. The key result is suppose x1 and x2 are independent random variables – more on this in just a second – with distributions p, say 1 of x1, and p2 of x2.

So to say they're independent, that means that making a measurement of 1 doesn't affect the measurement of another. They're independent events. They don't have anything to do with each other. That's – I mean, that's sort of intuitive way of saying it, but again, it's the terminology that you have no doubt heard.

But the idea, again, is that they're numbers that you don’t know yet. And knowing one number doesn't affect knowing the other number. So making a measurement of x1 doesn't affect making a measurement of x2, and vice versa.

The question is how is the sum distributed? And the answer is by the convolution, it's distributed according to the convolution of the distributions of the separate random variables. That is, the distribution of x1 plus x2 is given by the convolution p1 convolved with p2 of – well, just p1 convolved with p2. I won't write the variable.

We'll bring averaging – averaging doesn't come into this quite at this stage. Just think of this as the distribution of the sum is given by the convolution of the distributions. Very impressive result. Very beautiful result.

Now, let me show you how this goes. I'll get a proof of this for you. The probability – here is the basic observation. The probability that x1 plus x2 for any – let me start – let me add one thing here. For any t – for any value of t, the probability that x1 plus x2 is less or equal to t is given by this integral – is given by a double integral over the region x1 plus x2 – I'll draw a picture in just a second. x1 plus x2 less than equal to t, p1 of x1 times p2 of x2, dx1, dx2.

Now, there's a certain amount to say here – certain amount I will say, and a certain amount that I won't say. Here is the region in the x1, x2 plane that I'm talking about. Here's x1; here's x2. The line x1 plus x2 equals t goes, say, like that. And we are interested in integrating under this region.

Now, two things are happening here. You're making a measurement of x1, and you're making a measurement of x2. You have a distribution for the random variable x1. You have a distribution for the random variable x2. Because they're independent, and this is where the assumption of independence comes in, to calculate the probability of both – something on both x1 and x2 – to say that x1 plus x2 is less than or equal to t means the values of the sum y underneath this line, you calculate the product. So it's like x1 is less than or equal to t.

This is not quite right, but there's an "and" coming in here, like x1 is less than or equal to t, and x2 is less than or equal to t. x1 plus x2 is less than or equal to t. I'm not saying this right.

You calculate the integral of the product of the distributions of the two random variables separately over the region that you're talking about. And the region lies below the line x1 plus x2 equals t.

There's more of an argument for this, again, in the notes. But I just wanna show you how the derivation goes once you accept this fact. And it's not so – if you think about this for a while, it's actually – it's not an unreasonable statement. A matter of fact, it's quite natural.

Now, what I'm gonna do is – this integral, because the region is a little bit complicated, I can't, as I have before, uncouple this integral into two iterated single integrals, at least so easily, not the way it is written now. The variables here are coupled, in some sense, because the region of integration is a little bit more complicated. It's not just a rectangle or something bounded by a horizontal or vertical line. But I can get to that picture if I make a change of variable in the integral. I want to work with this integral by making a change of variable – a relatively simple change of variable.

I wanna make a change of variable in the integral. Now, let me write down what it is, and then I'll explain what's going on here. The change of variable that I'm gonna make is it's –

the variables are coupled. I'm gonna let u equal x1, and I have to write this down because I wanna get it right – v equals x1 plus x2. That's – that defines u and v, my new variables in terms of the old variables x1 and x2 or the old variables in terms of the new variables are x1 is equal to u, and x2 is equal to v minus u.

So the line, x1 plus x2 is equal to t is described in the other variables by x1 plus x2 is u plus v minus u is just v is described by v equals t. In other words, geometrically, the picture is this. Here is the x1, x2 plane. Here's the line, x1 plus x2 equals t, and here's the region underneath it. Here is the uv plane over here. Here's the change of variable – sort of it takes you from the uv plane from the x1, x2 plane. Here is the line v equals t. This region corresponds to this region under that change of variables.

And I'm gonna make a corresponding – now, I have to carry that through and see what happens at the integral, and actually, there's a general theorem here that tells you how you make such a change of variables in a multiple integral, and I'm not gonna go through that again. But in this case, it's particularly simple. It is something you may not have seen or may not be as familiar with as the ordinary integration by substitution that you've done many times.

There's a little bit more involved in making a change of variable in a multiple integral – the so-called Jacobian transformation comes in. But as I like to say, if you're gonna design transistors, chips with millions and millions of transistors on them, you can God damn learn how to change variables in a double integral. I mean, it's not so hard. It's not beyond your abilities.

But nevertheless, I'm not gonna go – I won't go through the formula. How does it – how do the change of variable work? Change the variables in the integral. In this case, it's particularly simple. You just get the integral from over the region x1 plus x2 less than or equal to t, p1 of x1, p2 of x2, dx1, dx2 becomes the integral.

Oh, or I could describe the regions. Minus infinity – v is going from minus infinity up to t, u is going from minus infinity to plus infinity. And it is p1 of x1. x1 is u. p2 of x2, and x2 is v minus u, du, dd. That's what it turns out to be. Now, do I have my order of integration right? I have my order of integration wrong here, I guess. Sorry.

v is going from minus infinity up to t. u is going from minus infinity to infinity, so dv, du – yeah, to describe the region in the uv plane, d is going from minus infinity up to t. u is going from minus infinity to infinity. Now, I'm gonna swap the order of integration. Let me leave that down there. I don't wanna cut it off.

So that's equal to the integral from minus infinity up to t. These are constants. The limits of integration here are all constants, so it's easy just to swap the order of integration, something I could not have done so easily in the – when the integral was written in terms of – in the x1, x2 plane. But because all the limits of integration are constant, it becomes easy to switch the order of integration. So integral for minus t for minus infinity to infinity, p1 of u, p2 of v minus u, du, and then the results is integrated with respect to v.

Now there we are. We have eyes if only we but see, as you have heard me say before, what is inside the integral here? What is inside the inner integral? Is the integral p1 of u times p2 of v minus u integrative with respect to u for minus infinity to infinity. That's just our friend, the convolution. That's the integral for minus infinity up to t of p1 convolved with p2, v, our u, du. Evaluate it. Sorry, sorry, sorry, sorry, sorry. I'll get it; I'll get it. Evaluated v, and the results is integrated with respect to v.

The inner integral is the convolution of p1 and p2 at v. Integral for minus infinity of infinity, p1 of p2, v minus u, vu.

Where do we start? Where do we finish? We see that the probability that x1 plus x2 is less than or equal to t is given by the integral for minus infinity up to t of p1 convolved with p2, vdv. And that is enough to identify what you're integrating as the distribution for the probability of that random variable. And so this identifies this result, the fact that you get a probability by integrating, some function identifies this as the probability distribution. This identifies p1 convolved with p2 as the distribution of x1 plus x2. Two exclamation points – that's really quite a striking result. And it's an extremely important result.

Now, once again, this says that if you have two independent, random variables, x1 and x2, so again, the measurement of one does not affect the measurement of the other. You ask the question, how is their sum distributed?

The answer is their sum is distributed according to the convolution of each one. That's, in words, what we just showed here. And it's a very important, very fundamental result. It links the sum of random variables to convolution. It will link averages to convolutional – averages as come as in a little bit later. In a matter fact, it'll be just a few moments.

Now, also in the notes, you should take a look at – there's a discrete form of this. As a matter of fact, I think I – I think I started off by motivating the more general result by looking at the discrete form. That is, by rolling a couple of dice – rolling one dice, rolling two dice, looking at the average, looking at the sum of two dice and so on. And you see exactly this result come in in the discrete case.

So if you're looking for a little bit more intuition about why convolution should come in when you're considering probabilities and sums, then look at the discrete case the way it's described there because it also shows you – I think pretty convincingly – why it is that convolution should come in to describe the distribution of the sum.

We are, however, dealing with it – we're have to do everything in the continuous case – simply some things, actually. Some of the calculations are actually simpler in the continuous case than in the discrete case. But it's a good motivation.

Now, I should say, by the way, that the result for two random variables also holds for a sum of any finite number random variables. So you get a similar result for x1, x2. x1 plus x2 plus [inaudible] xn is distributed according to the convolution p1 star, p2 star, p3. If you know the convolution of each, if they're independent, then you – then the convolution – then the sum is distributed according to the convolution of all of them – same argument. We see – you just work inductively.

Now, with this, I can give the setup for the central limit theorem – the actual statement. We're not quite there. We need a little bit more – a few more assumptions, which are natural, actually, which are the kind of assumptions that come up often when you're applying this – these sorts of ideas in practice and actually figuring out the distributions for actual measurements, actual experiments.

So here's the setup for the CLT. So again, I have n random variables are n independent random variables just as before. And eventually, I'm gonna let n10 to infinity here. I'm also assume that they have the same distribution – that the distributions are the same for all the xis. That is, they're distributed in the same way.

Now, that may seem like a – quite a strong assumption. It is, in some sense. But it's also an assumption that is quite natural. That is, if you're performing an experiment – if the measurements may be uniformly distributed, they're different sorts of measurements, but there's no reason to think they're distributed any other way, that making a measurement of one aspect of something – one aspect of an experiment is different or distributed differently from making a different measurement of the experiment – that they all have the same distribution.

And like anything else in this business that has an acronym associated with it, you say the – that x1 up to xn are iid. That stands for independent identically distributed. I hate these things. Independent identically distributed – but you see this terminology a lot, so you should know it.

Now, let's call the distribution p. Call it a common distribution p. And some further normalization is also possible. Let's call the distribution p of x. So it's the same distribution for each one of the random variables. Then you can normalize further. And again, this is explained in the notes.

You can assume that the mean of all the random – each one of the random variables is 0. You can assume mean 0 – that is to – that means that the integral from minus infinity to infinity of x times p of x, dx is equal to 0. That's for all the random variables because they all have the same distribution. And you can assume – that's also called the expected value. And you can assume that the center of deviation, or variance – it's the square of the standard of deviation is 1.

I'm using the term "standard of deviation." I'm a little bit more used to that, actually. So I can assume standard of deviation is 1. And that amounts to the assumption that the integral for minus infinity to infinity of x squared, p of x. Once the mean is equal to 0, then the assumption that the standard deviation is equal to 1 is this assumption.

We always have – we already have another property of p that the integral for minus infinity to infinity to p of x, dx is equal to 1. The total area is equal to 1. p of x is positive, and the total area greater than or equal to 0. And the total area is equal to 1. We can further normalize to assume the mean of 0, which means that the integral of x, p of x, dx is equal to 0. And the standard deviation is equal to 1, which means this integral, x squared, p of x, is equal to 1.

Now, what is the distribution of the sum? That's what we wanna get to. So what is the distribution xn? Well, now, that's actually – let me call this sn, the sum. That's actually not quite what we wanna look at because it doesn't have quite the same qualities as the independent – as the separate and random variables.

The mean of sn is still 0. But the standard deviation scales, by the square root of n – but the standard deviation scales by the square root of n – that is to say the standard deviation of just the sum is the square root of n of sn is the square root of n.

We want the sum to be comparable to the separate random variables that are entering into it. So instead, we look at not just the sum, but this sum, sn over the square root of n is x1 plus xn divided by the square root of n. That will have mean 0 standard deviation 1.

Now, I can give you a statement of the central limit theorem. The statement of the central limit theorem says what happens to this average – this scaled average, sn divided by the square root of n, as n tenths to infinity. Say this – it says that the limit – there are different ways of stating it.

One way of stating it is this: The limit is n tenths to infinity of the probability that a is between – well, that the events that you're measuring – the average of events is between a and b, or should I say the scaled average of events is between a and b. The limit of that thing is n tenths to infinity as you take more and more measurements of average more and more or a greater number of measurements, it is the integral – whatever 2pi the integral from a to b of the corresponding Gaussian, which also has mean 0 and standard deviation 1.

So once again, this sum has mean 0 and standard deviation 1. You take greater and greater – you take more and more measurements. You take – you're averaging out a series of measurements over a whole bunch of experiments. How are your values distributed? What is the probability that your measurements are going to lie within a certain range is you take more and more measurements and average them all out. The answer is that in the limit, it's distributed according to a Gaussian. That probability is calculated as if you were calculating the probability based on a Gaussian.

So that means event is particularly large, for example, you're probably getting a pretty good approximation if you assume that the sum – that the scale sum over there is distributed approximately according to a Gaussian.

What I'm gonna show is – I'm gonna show – so this is one version of the central limit theorem. I'm gonna show an unintegrated form of this. Namely, if p of n of x is the distribution of that sum of x1 plus xn divided by the square root of n, then I'll show that piece of n itself tenth to the Gaussian, as n tenths to infinity. And piece to x tends to whatever the square root of 2pi, either to minus x squared over 2 as n tenths to infinity. If we know that, then the integrated form follows.

Now, looks like quite a complicated statement. Well, you know what piece of n is, actually. Piece of n is easy to write down.

The distribution – well, one thing at a time here. One thing at a time. So once again, p of x is the distribution for x1, x2, up to xn. They all have the same distribution. p convolved with itself n times – so that's p convolved with p convolved with p n times of x – that's the distribution for the sum. That's the key result. It relates to sum to the convolution is the distribution of x1 plus, plus xn. But I don't have that exactly. I have the scaled version of that.

And there's something a little extra that happens here. And again, I have to refer you to the notes for a derivation of this. Let me just write down what the result is. The distribution of x1 plus xn, if I scale it, is piece is the square root of n times the n full convolution evaluate the square root times x. So it actually both scales outside and inside the variable. That's a fact. That's sort of a mathematical fact that comes from change of variables in scaling and how probability distributions behave under change of variables.

So for that, I have to refer you to the notes. So just take that on faith right now, or look it up. So in other words, piece of n of x – that's the probability distribution of the scaled 1, x1 plus xn – [inaudible] square root of n – is given by this formula, the square root of n – the n full convolution of p with itself, evaluate the square root of n times x.

Now, how the hell are you gonna show that this n full convolution of the scaled version of this n full convolution tends to a Gaussian? I mean, it doesn't look like I made the problem any simpler. And convolution is given by a pretty complicated integral, and I have an n full. I don't just have a convolution of two functions. I have the convolution of n functions. And I wanna take the limit of that thing as n tenths to infinity.

The key to analyzing this is the Fourier transform because the Fourier transform turns convolution into multiplication. Multiplication is easier to analyze than convolution.

I wanna take the Fourier transform – do it down here. The board gets too covered up. Take the Fourier transform of that function, square root of n, p convolved with itself n times. Evaluate the square root of n times x. Now, what do you get? And the idea is that multiplication – that this turns convolution into multiplication – easier, we hope, to analyze.

Now, how do we do it? Well, I apply the stretch theorem. The Fourier transform of pn of x is the Fourier transform of the scaled convolution. So it is the square root of n – that comes out of the Fourier transform of the n full convolution evaluated at the square root of n times x. Now remember, this is not each individual function evaluated. I [inaudible] little careful here. This is not each individual function evaluated. This is where you have to be careful about your variables.

So let me try to get a sentence out here. It's not each individual function evaluated at the square root of n times x. It's the convolution evaluated at the square root of n times x. So it's this Fourier transform of some big, hairy function evaluate the square root of n times x. This is the square root of n divided – times one of the square root of n if I apply the shift theorem of the Fourier transform of p star n evaluated at s over the square root of n. Just applying the shift theorem.

Now the Fourier transform of the convolution is the product of the Fourier transform. It's an n full convolution, so I have the product of the Fourier transform of p with itself n times. So the square root of n here canceled one of the square root of n there. This is the Fourier transform of p raised to the nth power, evaluated it at s over the square root of n.

In other words, it's a Fourier – [inaudible] another way – it's a Fourier transform of p evaluated s over the square root of n raised to the nth power. Make sure you follow the statement there. I mean, the one thing that I didn't show you was why the distribution of the scale average looked like this. But once you know this, then you can find this Fourier transform just by applying the shift theorem.

And you have to – again, you have to be a little careful here. You have to realize it's the n full convolution evaluated at the square root of n times x, not each individual function, not each individual term in the convolution evaluated there, but the whole convolution there evaluated the square root of n times x.

Now, we are almost there, baby. Can you feel the excitement? I'm sure that you can. The Fourier transform of p and s over the square root of n is equal to the integral of minus infinity to infinity. I have to write down the formula for the Fourier transform, believe it or not.

e to the minus 2pi i, s over the square root of n, x, p of x, dx. Now, of all things, I'm gonna use the Taylor series – yes – for the complex exponential.

This is – this, I think, goes under the heading of dirty tricks. It's integral from minus infinity to infinity. What is the Taylor series of the complex exponential – any exponential? It's 1 minus 2pi i, s over the square root – let me write sx over the square root of n, 1 plus x plus x squared over 2 plus x cubed over 3 factorial and so on and so on.

So it's this term, 1, and there's a minus sign here because either the minus 2pi s and 2 x. And the next term is square, but there's an i there. And so it's become minus 2pi square – why don't I just write it out here – 2pi i – 2pi sx squared divided by square root of n, the whole thing squared, one half of this – minus one half. There we go – I got it, I got it, I got it – plus and so on and so on and so on – plus small terms. [Inaudible] terms – times p of x.

Check me here, check me here, check me here. So again, I just use e to the x or e to the anything. e to the x is 1 plus x plus x squared plus 2 factorial plus and so on and so on. That's all I'm using.

Now, watch this. This is the integral for minus infinity to infinity, p of x minus – let me do this – dx. Let me start to take out the terms here – minus integral for minus infinity to infinity of – there are constants out in front here, 2pi i s, s comes out over the square root of n, times the integral for minus infinity to infinity of x times p of x, dx, minus – what's the next term here?

2pi squared s squared, x squared. So 2pi squared – that's 4 pi squared, s squared divided by 2, gives me a 2pi squared, s squared divided by n times the integral for minus infinity to infinity of x squared, p of x, dx, plus other small terms, or terms that can be estimated.

I will say not much more about them now, but I can give you some more details about that later. The important thing is to just keep the first couple of terms. So all I did here was multiply p of x through by the first couple of terms in the Taylor series, and then there's another integral p of x times these high-order terms. Now, use the normalization. Integral p of x is 1. The integral of x times p of x – that's the mean. That's 0. And the integral of x squared, p of x is 1 because the variance – the standard deviation is 1. This integral is 1.

What results from this dirty trick? What's left? What's left is 1 minus 2pi squared, s squared over n plus – let me just say smaller in – plus error terms. I'll just call them small.

What has happened here? We gotta go, man. We have found that the Fourier transform of p at s over the square root of n is, let's say approximately 1 minus 2pi squared, s squared divided by n. That's the Fourier transform of the function p. We want the nth power of that, right? The Fourier transform of p that's over the square root of n to the n because that's the n full convolution that comes in is this raised to the nth power. 1 minus 2pi squared, s squared divided by n to the nth power.

Now you have to remember something from calculus. This – 1 minus 2pi squared, s squared over n to the n, 1 minus something over n to the n power, as n tends to infinity goes to an exponential. This quantity to the nth power – as n tends to infinity – n tends to an exponential. 1 minus 2pi squared, s squared over n to the n is approximately and actually tends to limit e to the minus 2pi squared over s squared, as n tenths to infinity. Well, it's approximately this. And it gets better and better as n tenths to infinity.

Now you take the [inaudible] Fourier transform. The Fourier transform is tending of the convolution of the distribution – the distribution is tending to this thing as n is tending to infinity. I think it's okay. Take the inverse Fourier transform, and you get the result because you know the Fourier transform. The Gaussian is the Gaussian. You know what to do with a scaled Gaussian by applying – by invoking the theorems. And you find – if I got it right – that the result of this is that the original function – fs of n – excuse me, ps of n – the distribution of the scaled average and random variables tends to a Gaussian as n tenths to infinity. It's unbelieveable.

And that wraps it up. We brought that baby home, only five minutes late. Next time, I may wrap this up a little –

[End of Audio]

Duration: 55 minutes

TheFourierTransformAndItsApplications-Lecture11

Instructor (Brad Osgood):I wonder how long I can keep this up. All right. So the first thing – any questions? Any comments? All right. The first thing I want to do is fix a little mistake I made at the end of last time. I was upset about it for days afterwards, so I just want to correct it. This was [inaudible] central limit theorem, which I hope you will agree was a real treat. I mean, it's just an amazing fact, and the way that it brings in convolution, the [inaudible] transform of the galaxy and so on is really, I think, a real highlight of the course to see those things emerge and to – the whole spooky nature of repeated convolutions approaching the galaxy, I think, is just an amazing thing.

But I did screw it up at the end, so I just wanted to correct that. Everything was fine except for one final formula that I wrote, so I'm not going to redo it. I just want to point out where I wrote the wrong thing. It's correct in the notes. I just wrote the wrong thing on the board. This was the setup – so X1 through XN are, as they say in the biz, independent and identically distributed random variables. You don't have to write this down. I'm just recalling what the notation was. And P of little X is the distribution that goes along with all of them because they have the same distribution.

There's a single function, P of X, which describes how each one of them is distributed. So it's a distribution for each. And then I formed P of N of X was the distribution for the sum scaled by square root of N. So it's the average – excuse me. There was some assumption we made on the Xs on normalation, that is. We assume they had mean zero and we assume they had standard deviation or variance one, and then if you form the sum, the mean of the sum is zero but the standard deviation or the variant center deviation of the sum is the square root of N, so it's scaled by the square root of N.

You divide by square root of N to have this sum, SN, I called it, have mean zero and standard deviation one, and P of N of X was the distribution for this. What we found was that the Fourier transform – here we go. The Fourier Transform of P of N of S was the Fourier transform of P and S over the square root of N to the Nth power. And then the trick of the analysis – the way the proof worked was to compute the Fourier transform of P and S over the square root of N just using the definition of Fourier transform that uses a very sneaky thing of looking at the Taylor series of expansion with the complex exponential integrating the terms and so on, and it was really quite clever.

What I found was that the Fourier transform of P at S over the square root of N was approximately if N is large was one minus two Pi squared S squared over N. That was fine. That was fine. And then for some reason, I raised this to the Nth power, so the Fourier transform P at S over the square root of N to the N was then approximately this thing raised to the Nth one minus two Pi squared S squared over N to the N and then inexplicably I had the wrong approximation for this in terms of an exponential in terms of the power of E.

That is this one minus two Pi squared S squared over N is approximately E to the minus two Pi squared S squared. I'm going to look that up again to make sure I have it right this time. Sure enough. All right. Then from there, and for some reason I wrote two Pi squared over S squared last time, but it's two Pi squared times S squared. That's an elementary to the N. That's an elementary fact from calculus that you learned probably a long time ago when you were first learning about properties of the exponential function.

And then from there you take the inverse Fourier transform and you get the result. FT of this inverse Fourier transform of this galaxy and E to the minus two Pi squared S squared to get the result, which I am afraid to write down because I'm afraid I'm just going to make one more mistake, so I'm not gonna do it to get the final result of the central limit theorem. There. I feel better now that I've corrected it. I'm not sure everybody else here feels better, but I feel better. Okay.

Now to get on to the day's events. Brothers and sisters, today we have to come to a reckoning with our failings. Today we have to confront things we have not confronted but must. Now first, the good news, and there is always good news. The good news is the formulas we have derived, the applications we have done are all correct. So relax. Nothing that we have done, strictly speaking, is wrong. However, much has been left unsaid and today, we have to say it. We have to confront the convergence of intervals. It is a sin to consider integrals that do not convert, and I'm afraid there are times when we have sinned.

But in the end, nothing that we have done is wrong, so relax. But much has been left unsaid. So it is time now. Are you ready? Are you ready? Okay. Now we need to have a better, more robust definition of the Fourier transform in order to consider – I think I gotta stop this. There's only so much of this I can do. We need a more robust understanding of the Fourier transform. We need a more robust definition of the Fourier transform that will allow us, verily, to work with the signals that society needs to function – sines, cosines, even the function one – constant functions.

And the Fourier transform as we have defined it will not do the trick. More robust definition, and that's really what it amounts to – definition of the Fourier transform – to deal with common signals, ones for which the classical definition does not work or will not – is not adequate. The issue is exactly the convergence of the integral or if not the convergence of the integral for the function, applying Fourier inversion. The classical definition, let me just say, will not do.

There are two issues, really. One is the definition of the Fourier transform itself – the convergence of the integral defining the Fourier transform – when I say convergence of the integral, I mean just some way of evaluating, some way of making the integral make sense – convergence of the integral defining the Fourier transform. And the second and just as important, certainly for applications – it doesn't do you any good to take the Fourier transform if you can't undo what you've done. That is to say, you want Fourier inversion to work, also. You want Fourier inversion to apply.

It shouldn't be surprising that if there's a problem with the definition of the Fourier transform, there may be a problem also with Fourier inversion because the formulas are very similar. One involves the integral of either the minus two Pi ISX the interval of the integral of either the plus two Pi ISX but really, the issues of convergence as far as that goes are pretty much the same. Now there are two ways to approach this problem. This is a genuine problem that really has to be dealt with, and it has been dealt with in various ways over various periods of time.

Fourier analysis is a highly developed subject, and it has gone through various stages of development. There are two ways of doing this. In the early stages, which are probably a fair way of characterizing it, it was on sort of an ad hoc basis. That is, special techniques were developed for special problems that would explain away a special function or a special signal. That's one way. Two ways of dealing with these problems. That is one is so to speak ad hoc special techniques. Got a particular problematic integral? Okay, let me just work with that integral.

Having trouble with convergence of a reticular signal? Having trouble applying Fourier inversion in a particular case? Okay, let me deal with that by some trick, by some method that applies not generally but to that one problem that you're having difficult with and then you can advance until the next problem comes up. This was okay. This was all that could be done because it was all the understanding at the time led to. The second approach, and that's the approach that we're gonna take, actually, is to rework the foundations. That's the second approach.

You have to rework the foundations to give, really, a different definition of the Fourier transform, one that applies more robustly, more equally, somehow, to all signals that come up all at once. You have to rework the foundations and the definition. This came much later. This came probably in the 1940s, and by that time, as I say, Fourier analysis was highly developed, both in its theory and its applications. It was a highly developed subject. This gets into mathematically deep waters, which we will not completely tread, but I do want to give you the confidence enough to swim in them a little bit, just so you know what the issues are and know how they were dealt with.

Furthermore, the other thing that I want you to get out of this is a certain confidence that you can compute. The method wouldn't be any good, really, outside of mathematical departments and circles if it didn't offer a way of computing with confidence when that couldn't be done so much before. You can actually use this method to find specific Fourier transforms in a way that's airtight and is actually quite nice. We'll do that. Don't fret over mathematical details. What I'm going to do and I hope to do is take you to the brink of the abyss and then back off.

Say what the issues are, say what the problems are but don't deal with the mathematical proofs and don't deal with all the fundamentals that have to be analyzed, because it's quite daunting. There's a huge amount that has to be done. We don't need that. But I think what you can see is the general structure, the general overall direction to the argument. Let me give you an example to show where the issues are. Let's go back to the very first example we considered, which already shows the problem. I may have mentioned this before, actually, but now I want to say it a little bit more sternly somehow.

The problem is evident in the very first example. The very first example we had was the rectangle function in this Fourier transform. So F of T is the rectangle function, Pi of T. There is no problem at all taking the Fourier transform of that. That's not the issue. The integral converges. It's a very simple function. The Fourier transform as classically defined is fine. The Fourier transform of the rectangle function is the sinc function. Sinc of S is sine of Pi S divided by Pi S. that's fine. The problem is inversion. The problem is getting the other way around. We did this by duality.

There are really two things – there's either Fourier inversion or finding the Fourier transform of the sinc function. The problem is concluding that the inverse Fourier transform of the sinc function is the rectangle function. We use duality or more or less equivalently via what we call duality, and that was all fine. Nothing we did was wrong. It wasn't justified. The duality – what is the same thing, the Fourier transform and the sinc function, is the rectangle function. Same kind of issue. The same issue was involved here in showing this as in showing this. The problem is writing down that integral.

The Fourier transform of the sinc function of the inverse Fourier transform of the sinc function – let me just stick with the problem of Fourier inversion. The inverse Fourier transform of the sinc function is the integral from minus infinity to infinity of E to the minus two Pi IST. Let me write out the sinc in terms of sine and cosine. Sine of Pi S divided by Pi S, and I'm integrating with the respect to S, so what results is supposed to be a function of T. As a matter of fact, what results is supposed to be the rectangle function. Now I said at the time you want to try to evaluate that integral? Good luck, because it's not so easy to do.

So again, according to what we expect to be true, this integral turns out to be one when S is less than a half and it turns out to be zero when S is greater than or equal to a half. It turns out to be the rec function. Now even this actually is not quite correct because – well, one thing at a time, here. In fact, by ad hoc techniques, special techniques and there are actually a variety of techniques that can be brought to bear on this, some more complicated than others, you can actually evaluate this integral. But it's not obvious. No simple integration by par, no simple integration by substitution is going to get you this sort of result.

This result is almost okay but requires special techniques. In fact, there's a problem at the end points. There's a problem at plus and minus one with that integral. It's equal to one half, I think, when S is equal to plus or minus one half, so that's another reason for sort of defining the rectangle function that way so it doesn't jump from zero to one but has a value of one half at [inaudible] continuities. Never mind that. Special technique. So in fact, special problem at the end points – end points S equals plus or minus one-half. But anyway, the point is that it can be dealt with.

It's a little bit disconcerting that the simplest example in the entire subject already poses this problem. The most basic function already requires you to do special arguments just so you know the Fourier inversion works. That's a little bit irritating. Oh, Jesus. Yes. There. Good. Don't do that again. Same thing because it's an even function. There. Now, let me take a second example. There's a problem where it's fine to compute the Fourier transform, but the problem is with Fourier inversion. Computing the Fourier transform of the rectangle function gives you a sinc function. That's fine.

But Fourier inversion does not apply directly without some extra work. As a second example of troubles, there are very simple functions for which you can't even get started. I believe I may have mentioned these before, but let me say them again now in this context of repentance and trying to lead a more virtuous life. A second example – consider F of T equals one, the constant function. Then there is no way to make sense of in a classical way of the integral from minus infinity of its Fourier transform. The integral from minus infinity, either minus two Pi IST times one DT.

There's no way in which this integral can be made to make sense, period. It just won't work. I know you know that you write down this integral and you write, oh, that's just the delta function. That comes later, and that comes later really only as a convention because to write down this integral is to sin in the sense of expecting this to do anything. This integral does not make sense – can't be made to make sense. And slightly more generally but equally importantly, other signals do not have classical Fourier transforms in the sense that the integral defines something – the integral is well defined.

For instance, sine and cosine have the same sort of problem. Likewise for F of T equals the sine of two Pi T or G of T equals the cosine of two Pi T – there's no way to make sense. You can't make sense of the integral from minus infinity to infinity of the classical Fourier transform. You have the minus two Pi IST, say, sine of two Pi T DT. The integral oscillates. The thing doesn't die out as you're tending to infinity. Nothing good is gonna happen here. No secret combination of cancellations, no conspiracy of plusses and minuses is going to make this thing work. It just won't work.

These are the signals that society needs. What is to be done? You remember that the situation is somewhat analogous, philosophically, spiritually, heavenly. The situation is somewhat similar to what we confronted with Fourier series. That is, Fourier series was supposed to be a method that applied in great generality. You wanted to be able to write down a very general periodic function in terms of its Fourier series and then compute with it. The problem was those series didn't converge in any sense. You couldn't make those series converge, or maybe you had to make a very special argument to show that they converge and so on.

What was finally – and then finally the subject not so much collapsed of its own weight but a fundamental reworking had to be done. In order to make Fourier series apply as generally as you would like it to apply, you had to really change the notion of convergence. You had to rework the foundations. So it's analogous. I don't want to make the analogy too sharp or too precise, but in some sense it is analogous, the Fourier series.

In the foundations, we needed a new definition, new conception of convergence. There's a lot that went on with that.

You had to sort of abandon – if you wanted to advance the subject and if you wanted to talk about things in greatest generality, you couldn't talk about point wise convergence. You had to talk about convergence in the mean and so on. You had to talk about the square integral functions and all the rest of that jazz, quite far removed from the initial development of the subject. It's what had to be done, and ultimately led to a very fine theory and a very robust and applicable theory. Now we're gonna have to do a similar thing here. Again, Fourier series – the Fourier analysis, Fourier series and then the work on the Fourier transform is a very highly developed subject.

You ask yourself if I'm forced, somehow, because the thing is collapsing. It's too many ad hoc arguments, too many one off examples that have to be explained, so I really have to rework the foundations. But it's such a highly developed subject that you are confronting the problem that any scientist or engineer or mathematician confronts in a situation like that is that nature provides you with so many phenomena that have been so thoroughly studied – how do you choose the rights one upon which to base your theory?

How do you choose a relatively small number of phenomena upon which or based on which the others can be explained? So much has been done in Fourier analysis and applications of Fourier analysis and everything. Where do you start? How to choose the basic phenomena which explains the others which can be used to explain others – all others. That's the problem. It's a problem that confronts anybody doing experiments. Think of this as viewing a series of experiments. What are you measuring? Is that significant? Is it not significant? What do I pay attention to? What do I ignore?

The results are there in front of me. What do I do? Which ones do I choose? I want to explain how this was resolved. It was resolved quite brilliantly and quite decisively in the 40s. The first step, and I'll say a little bit more about the history of this as we go along. The first step in understanding this is to somehow back away from the troubles and concentrate on what works well. That's the first step. So back away from the problems. What is the best situation? Identify the best situation or what characterizes the best situation. In this case, what that means is you want to identify a class of signals for which everything you want is true.

You want to identify – when I say what is the best situation, I mean identify the best signals, the best class of functions for Fourier analysis – let me just say Fourier transforms. Now the requirements are these. What we want is two properties. We want one. Let's call this class of signals S. So think of S standing for signals but actually S is gonna stand for Schwartz, the person who isolated these particular class of functions as the best ones upon which to build a theory. So then we want to – one. If F of T is in S, then the Fourier transform of T is defined and it's also in S. Fourier transform of F is defined classically by the integral.

The Fourier transform of F is also of the same class. Now that already rules out, for example – that already tells us – well, I haven't defined S yet. You can imagine that it tells us that the rectangle function shouldn't be in that class and the sine and the cosine shouldn't be in that class or the constant one shouldn't be in that class because for the rectangle function, there's obviously a problem – the Fourier transform of the rectangle function is the sinc function. The sinc function is not as good as the rectangle function in some sense because I can't plug it back into the formula for the Fourier transform or the inverse Fourier transform.

Likewise with sine and cosine, they're not good, because the Fourier transform of the sine and cosine isn't defined or the Fourier transform and the constant function one isn't defined. This already rules out the exact signals that I want to deal with. It doesn't seem like the best class to me, but in fact, it is. It's gonna be a restricted class but it's gonna serve as the foundation for a broader definition. That's the first property. So the function – if a function is in this class S, then so is its Fourier transform. The Fourier transform [inaudible] so is the Fourier transform. The second property is the Fourier inversion works.

Two. Fourier inversion is defined by the classical by the formula, and that only makes sense even to talk about two in light of one because one says if you take the Fourier transform, you wind up back in the class, so you can take the Fourier transform of this function again or the inverse Fourier transform of this function because you haven't left the class of functions for which you can define the integral. So that is to say the inverse Fourier transform or the Fourier transform of F is equal to F and same way going the other way, the Fourier transform of the inverse Fourier transform of F is [inaudible].

That's what we want. This is the best class – if we can find a class of functions S that satisfies these two properties, then that's the best class of signals that we can hope for for the classical theory. For the classical theory, this is the best thing you can hope for. There's one other property that comes up that always comes up in the discussion, and I want to just state it now. It's not a requirement of the theory, but it comes up as sort of part of the theory and is actually extremely useful for the further developments.

There's a further property – let me just say that – a further property – an identity that I haven't written down yet. I'm going to write it down now and we'll make a certain amount of use of it as we go along is what's sometimes called Parseval's identity for Fourier transforms. We have Parseval's identity, also called Railey's identity for Fourier series, and there's a corresponding one for the Fourier transform. Parseval's identity is also sometimes called a [inaudible] formula. It's even sometimes called Railey's identity for Fourier transforms, and it says this.

It says that the integral for minus infinity to infinity of if I integrate the square of the Fourier transform, it's the same thing as the integrating from minus infinity to infinity integrating the square of the function. This is not a requirement of the theory. This is not a requirement of the best class of functions to consider, but it tags along with it. It's somehow almost inseparable from it. I'm going to derive this for you a little later on today. This says this has physical interpretation in various fields depending on where it's applied.

This sort of identity comes up a lot in distinct but related fields, and it says something like you can calculate the energy. You think of the integral of the square of the functions as total energy, so you can calculate the energy in the frequency domain or you can calculate the energy in the time domain. Or sometimes you think of this as a spectral power, total spectral power. It can be computed in the spectrum or in frequency or considered in time and so on. There's various interpretations. There's various terms that go along with this depending on who you're talking to. Mathematically, it's always known as either Parseval's identity or Railey's identity – something like that.

It's analogous of what we had for Fourier series. For Fourier series, we had the sum of the squares of the Fourier coefficients. That's what we get on the left hand side. That's the spectral information and then on the right hand side we had the integral of the square of the function, but we only had it over one period. Here, it's integrated from minus infinity to infinity because the signals are supposed to live on minus infinity to infinity. I'll come back to that a little later on. As I said, that's not a requirement of defining the best class of signals, but it always comes up. Now, here's the question – it's not such a stretch to say this is what we want.

These are the functions that we want. These are the two properties that we really want, our best functions to have. The question is what is such a class and is there a simple enough way of describing it that's gonna be useful that I can actually work with and how do you find it? How do you find such a class? Again, imagine the difficulties. Imagine you are a scientist or a mathematician or an engineer confronting this problem and as I say, the field is highly developed. Nature has provided you with so many examples, so many different kinds of signals and so on. How do you choose?

What properties of the Fourier transform, what properties of the signals do you choose upon which to base your theory? It's hard. That's a hard intellectual challenge. It was met by Laurent Schwartz, who was a very famous French mathematician who taught for many years at the Ecole Polytechnique in Paris. So how to choose, how to define S? That's the question. How do you isolate a small set of phenomena along all possible phenomena that nature has provided in order to build a reasonable theory? This is solved by an absolutely brilliant and inspired work. Genius does what genius does. Who can account for it?

This is solved by Laurent Schwartz back in the 40s, and I think it's significant, actually, that Schwartz taught at Ecole Polytechnique. Ecole Polytechnique is the famous technical school of engineering in France. It was founded in the Napoleonic era to train technically competent Frenchmen – probably not so many French women at the time – to become engineers. And Laurent Schwartz taught there, and Laurent Schwartz was although mathematician by training and mathematician by published work was thoroughly grounded and thoroughly familiar with the applications.

And as a matter of fact, in his monumental book on the subject which is called the theory of distribution – these generalized functions that we're gonna talk about are often called distributions, not to be confused with probability distributions. Anyway, in his book on this, he – I think in the preface, the first sentence of the preface or something like that, he credits not mathematicians, not even scientists but he credits Oliver Heaviside with really laying the basis for the whole theory. And what he's thinking of there is delta functions that we now call delta functions and things like that, sort of operational – what then was called operational calculus.

So it was in a very applied context that Schwartz was thoroughly familiar with from which he extracted his theory of what was a more modern foundation of the Fourier transform. It's absolutely brilliant and inspired. It's very complicated. If you go through all the details, it's very complicated. We're not gonna do that. But if you allow yourself to sort of go along for the ride, it's actually not so complicated and it's a pleasure to see it work, and it's something that you can compute with. I keep coming back to that because I want you to believe me and I want you to realize that I'm not just doing this for the hell of it.

That is, we can actually compute in a very convincing way the Fourier transforms that you really want – sines and cosines, the delta function, constant function, all that stuff. It really comes out of the theory in a beautiful way, but you do have to accept a certain amount going into it. You have to accept a certain change of point of view and change in perspective. I'm going to give you the property right now, because it really is fairly simple to state. I'll tell you sort of where it comes from and why it comes from, but I'm not gonna go through the details even at that, although more details are provided in the notes.

Here's the definition. S, actually – so S is now not just standing for signals but standing for Schwartz, is the class of the best class of functions for Fourier analysis. It is the so called class of rapidly decreasing functions. S is class of rapidly decreasing functions. And it's defined by two properties – two relatively straightforward properties. First of all, any function of the class is infinitely differentiable. You don't have to worry about smoothness. It has arbitrarily high smoothness. You want to take 1,000 derivatives? Fine. You want to take a million derivatives? Fine. You want to take just one or two derivatives? Fine. Everything is just fine.

So F of X is infinitely differentiable. That's one property. That doesn't talk about the rapid decrease. The second definition is the rapid decrease, and what it says is that for any M and N greater and equal to zero integers – let me write the definition down and then tell you what it means. X to the N – the Nth derivative of F of X tends to zero as X tends to plus or minus infinity. I don't know whether I switched the rules of M and N here – what I'm writing on the board here with what I wrote in the notes, but it doesn't matter. What this says in words is not hard.

In words, this says that any derivative of F tends to zero faster than any power of X. M and N are independent here. The third derivative tends to zero faster than X to the 100.

Also faster than X. Also faster than X to one million. The 50th derivative of F tends to zero faster than X to the five millionths or whatever. M and N are independent. It says once again in words, and that's the way to think of it, is that any derivative of F tends to zero faster than any power of F. And independently – the order of the derivative of the power. Okay. Where this comes from I'll explain in just a second.

First question you might ask is are there any such functions? I mean, is it terribly restrictive? This certainly does rule out sines and cosines and constant functions. Sorry, tends to zero faster than any power of X. This already seems to rule out the functions we want, or a lot of functions we want. So it's easy to write down, but it seems highly restrictive. And again, it's a little bit surprising that one could build such a powerful theory, such a general theory on such a restricted class. But you can, and that's the story that's yet to be told. Are there any such functions? Well, I'm happy to write down at least one of them.

A galcean is such a function. Any galcean – I'll take the galcean that we've worked with so many times. It doesn't matter how it's normalized. Any such galcean will work, actually. As it turns out, of course the function tends to zero as X tends to plus or minus infinity, but in fact, the exponential is going down so fast that any derivative of the exponential is also gonna tend to zero faster than any power of X. You know what happens here If I start taking derivatives of E to the minus X squared, I'm gonna get powers of X out front, but always also – they're gonna be multiplied by this E to the minus Pi X squared, which is gonna kill them of.

And in fact, any derivative – any power of X times any derivative of this function will tend to zero. Check this out. Do some calculations. Do some experiments. Do some graphs. Work it out. Just try it. I'm not gonna prove that for you, but it's true. So that's – it's not a vacuous class. At least there are functions in the class. That's good news. Another important class of functions, as it turns out, are the smooth functions, which vanish identically as you get close to infinity. Not just tend to zero, but actually which are equal to zero beyond a certain finite interval.

So C, another class of functions – these are the infinitely differentiable functions, which are equal to zero, identically zero, outside some finite interval. Different intervals for different functions. I'm not saying all the functions vanish outside of minus one to one or minus ten to ten or whatever, but a function is in this class if it is infinitely differentiable and it goes, as they say in the biz, smoothly down to zero. So the graph is something like this. You can do whatever it does, but it does smoothly and it tends to zero smoothly. It's identically zero beyond some certain point from A to B.

And again, different intervals for different functions. I use the word C because the mathematical term for this is functions of compact support. The support of a function is the place where the function is non zero and has a bounded, closed interval, and that's called in the mathematical biz compact set or compact interval, and so these functions are called compact support. You will see that terminology certainly in the mathematics literature, and you also see it in the engineering literature. It's pretty widely adopted. That's what the C stands for, compact support.

And there are plenty of functions like this, too. It's not hard to come up with functions like that. So there are – at least the function of the class S is populated. There are functions which are rapidly increasing. It's not a vacuous theory. Where does it come from? Why these properties among all possible properties and phenomena that nature has given you? Fourier analysis is a highly developed subject. There's so many things to choose from. Why this? How did this work? Like I say, genius does – who can account for it?

I can tell you what it's based on. It's based on the relationship between smoothness and rate of decrease – it's based on the derivative formula. That turned out to be sort of the fundamental property of Fourier transforms upon which this theory could be erected and the other things could be understood. The connection here comes through the derivative theorem. There are various versions of the derivative theorem, but the one that we looked at before was the Fourier transform of the Nth derivative – I used M before, so let me use M here – was at S is two Pi IS to the M times the Fourier transform of F to S.

There's actually a similar formula for the Fourier transform of the derivative of the Fourier transform and so on. I won't write that down. Again, these are in the notes. We used this when we talked about the heat equation. I said this was actually a fundamental property of Fourier transforms. It turns differentiation into multiplication. The derivative formula [inaudible] high enough elevation if you're looking at it relates smoothness – that is to say the degree of differentiability to multiplication – to powers of S.

So saying something like this quantity tending to zero, meaning a power of S times in this case the Fourier transform tending to zero says something about how differentiable the function is. It's a relationship between differentiability and measurements in terms of powers of S of rates of decrease. I'm not really gonna say anything more about it than that. That plus about a 250 IQ will lead you to develop a fantastic theory. But anything short of 250 IQ probably will leave you short.

But this turned out to be the derivative theorem and explaining the relationship between differentiability, between the degree of smoothness and the sort of rate of decay turned out to be the crucial thing upon which this thing hinged. So it was sort of the connection from differential equations, and actually again, that's not so – that comes out a lot in the applications. We didn't do a lot of this, but in Heaviside's work, for example, where he introduced the so-called operational calculus delta functions and things like that, it was often in the context of differential equations – of solving differential equations, and he would work on solutions of differential equations by taking transforms.

Differentiation turns into multiplication under many of these transforms. You've seen probably when you were a kid the [inaudible] transform. It has the same sort of property. Other [inaudible] transforms have a similar kind of property. Again, it was one of the phenomena that was out there, but why this phenomena and not other phenomena? That's what it took a genius to see.

I want to do one more thing today, and then I'm gonna show you next time how we're gonna use this and how we're gonna proceed from this very restricted class of very smooth functions with very good properties to functions which are very wild and things that are not functions at all, like deltas and things like that and how we're gonna use this to actually define the Fourier transform for objects which are extremely general. But let me take this opportunity to finish up today with a derivation of Parseval's identity.

It's off the direct trajectory that we're talking about now, but it comes up often enough, and it's somehow always allied to the theory that I think it's worthwhile just going through the derivation. So Parseval's identity for functions in class S – for functions for which there is no issue of convergence, anything. Parseval's identity says again, the integral from minus infinity to infinity F of S squared DS is equal to the integral from minus infinity to infinity of F of T squared DT.

And by the way, for functions in the class S, there is no question about convergence of these intervals, because the functions are dying off so fast that the integral of F squared converges. And furthermore, because the class has the property that if you take the Fourier transform you stay within the class, this integral also exists. This integral also exists. I should have said that, actually. Part of this point here is that not only does the fact that this satisfies the properties that we're interested in is the function satisfies the property of rapidly decreasing, also therefore Fourier transforms do, also, and that's where this relationship also comes in.

Sorry, I should have been a little bit more explicit about that. We'll come back to it. I'm actually gonna prove a slightly more general version of this, and it'll take me but a moment. Actually, it's quite a straightforward derivation. The advantage is that everything is justified. All the manipulation I'm gonna do of integrals is completely justified. So let me prove a slightly more general one, namely the integral from minus infinity to infinity of the Fourier transform of S times the Fourier transform of G bar DS, an inner product of the Fourier transforms is the same as the inner product of the functions – integral from minus infinity to infinity F of T times G of t bar DT.

So I allow for complex value functions here, and of course the Fourier transform generally is a complex value. So how dose this follow? Well, let me use X here instead of T, not that it matters. I'm gonna invoke Fourier inversion. That is to prove this, I'm gonna write G of X as the inverse Fourier transform of the Fourier transform. I'm gonna write G of X as the integral from minus infinity to infinity of E to the plus two Pi ISX, the Fourier transform of G of S DS. I can do that because I can. In this case, I can.

If I assume that F and G are in class S, Fourier inversion works. Everything is fine. Okay. Then what about the integral from minus infinity to infinity of F of T, F of X, G of X bar DX? Well, G of X bar – sorry, let me do it over here. G of X bar is the integral from minus infinity to infinity. I take the complex conjugate of the inside.

So that'll be E to the minus two Pi ISX, the complex conjugate of the Fourier transform S DS. So I take F of X times G of X bar integral minus infinity to infinity, that's integral from minus infinity to infinity F of X times the integral from minus infinity to infinity E to the minus two Pi ISX F G of X bar DX and then the result is integrated with respect to – DS, sorry – the result is integrated with respect to X. Now, in respect to X. Now, let the rigor police through the door.

I don't give a damn because everything here converges in every possible good sense. I can interchange limits of integration. I can whip them suckers around with absolutely no fear of retribution. That is to say, brothers and sisters, that this is equal to the integral from minus infinity to infinity, the integral from minus infinity to infinity – let's put them all together and then take them all apart – E to the minus two Pi ISX F of X, G of X, F of G of S bar DS DX and now, that's putting all the integrals together. Now I take the integrals apart and turn that into an iterated single intervals again, and I can do that because everything converges in every possible good sense.

You don't scare me, you rigor police, you. Minus infinity to infinity of F of G of S bar integral from minus infinity to infinity E to the minus two Pi ISX F of X DX the result integrated with respect to S. I swapped the limits of the order of integration. You see, what's inside that integral is nothing but the Fourier transform of F. This is equal to the integral from minus infinity to infinity, the Fourier transform of G of S bar the Fourier transform of F of S DS and I am done. Where do we start? Where do we finish?

We started with the integral of the inner product – the integral of the Fourier transform of F times the Fourier transform of G – no, I didn't. Here's where I started. I started with – there I started! Okay. Up there. I started with the integral of F times G bar and then ended with the integral of the Fourier transform of F times the Fourier transform of G bar and that is what I wanted to get.

Ain't life good? And it all works and it's all perfectly justified because the class S is the perfect class for all Fourier analysis, and next time you're gonna see how we get from what seems to be a highly restricted class to all the functions that society needs to operate. All that and more, coming up.

[End of Audio]

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture12

Instructor (Brad Osgood):It's true, you know. There I was, lying there and the cop said to me, where are your clothes, pal? Oh, sorry. Okay. The saga continues. Let me remind you what we did last time. Last time, I introduced the best class of functions for Fourier transforms, or at least I asserted that it was the best class of functions for Fourier transforms, and I want to remind you what the properties are, then I want to tell you what we're gonna do with it. The best class of functions for Fourier transforms. We call that S, the class of rapidly decreasing functions, and they're characterized by two properties.

They're infinitely differentiable and any derivative decays faster than any power of X. I will write that down. First of all, Phi of X is infinitely differentiable, so as smooth as you could want, has as many derivatives as you could want and more, differentiable and secondly that, as I said, any derivative decreases faster than any power of X. For any M and N greater than or equal to zero, X to the ND N DX to the [inaudible] derivative of Phi of X [inaudible] also tends to zero as X tends to plus or minus infinity. Those two properties. This is the M. M and N are independent here, so this says – there's nothing mysterious here.

You've got to measure decay or growth some way, and the simplest way of measuring growth is in powers of X. You can say a function grows linearly or grows quadratically or grows cubically. That's a natural scale of measurement for how a function is growing, and so to talk about a function decreasing more rapidly than any power of X, you can say, well, if it decreases faster than linearly, than X times that function is going to go to zero. If it decreases faster than quadratically, you look at X squared times the function. You want that to go to zero as X tends to plus or minus infinity.

So multiplying by a positive power of X and insisting that the product of the positive power times any derivative here tends to zero says that it goes to zero faster than any power of X. It's a strong statement, but it's not an unreasonable statement. As it turns out, there are plenty of functions that satisfy this property. What wasn't obvious by any means, and again, nature provides you with so many different phenomena. How do you pick out the one to base your definition on? Why this for properties of Fourier transforms as the best class of functions for Fourier transforms and not something else?

Well, genius is what genius does, and as it turns out, this was the right class to single out. In the following sense, and even here, and it may not be completely clear that this is what you really need, and the story will spin out as we go on. Why the best for Fourier? Well, one reason is that if Phi is a rapidly decreasing function, then so is its Fourier transform. That is if the function decreases faster than any power of X and any derivative, it decreases faster than any power of X so is [inaudible] Fourier transform. Also, if the function is infinitely differential, so is its Fourier transform. All the properties are preserved.

All those analytical properties are preserved by Fourier transform. That's very important. Again, why is it so important and why these particular things work so smoothly for developing the theory, you'll see. The second property is that Fourier inversion works. That is, if Phi is an S, than the inverse Fourier transform of the Fourier transform of Phi is equal to Phi, and the same if I go the other direction – that is, the Fourier transform of the inverse Fourier transform of Phi is also equal to Phi.

The usual Fourier transform here, defined in terms of the integral, and there's no problem with that interval converging because the function is dying off. That's where we were last time. If this is such a darn fine class of signals for the Fourier transform, how come it doesn't include some of the singles we would really like it to include, like, for instance, the most basic example of all, the rectangle function? The rectangle function is not in [inaudible], but if so good, if S is so good. For instance, the rectangle function is not in S, the most basic example.

It's not in S because it's not continuous, never mind differentiable or anything else. It's not even continuous. The triangle function is not in S. It's continuous, but it's not differentiable. The examples that we started our whole study on don't even fit into this supposed class of the best functions for Fourier transforms. Not to mention the other class of functions that we might want to consider, like constant functions, sines and cosines and so on. None of those are in the class S. Constant functions, trig functions, many others are not in S.

How do you resolve this? How do you get to defining the Fourier transform or how do you get back to – how do we be sure we haven't lost anything and then what is gained by considering this very good class? How can we be assured we haven't lost anything? On the surface, it looks that we've lost something by restricting ourselves to consider this class when we've lost the rectangle function and the triangle function and who knows what else. Furthermore, and have gained and will gain even greater generality. That takes a little while to tell that story. We'll get most of the way there today.

Answering that question causes us to take up another strain of development that was happening around the same time. To answer this, we're going to have to pick up another line of development. The two will come together triumphantly, but only after we follow this path for a little while, and that is delta functions and so on. It is the idea of generalized functions. They are also referred to as distributions, and that's probably the term that I'll use. This use of the word distribution has nothing to do with the way we used it earlier when we talked about probability. It's just one of those clashes of terminology that comes up every now and then.

There are only so many words to go around. I actually don't know the origin of the word distribution in this context, but that's what's used. So generalized function and distribution are synonymous terms, and I'll probably find myself slipping into using distribution rather than generalized function, although both terms are in current use. What I mean by here is typified by the delta function, which really should be called the heavy side delta function. I am assuming, and I will remind you of some of the properties that you probably have seen.

I am assuming that you've seen the delta function in various contexts, because everybody that goes through an engineering course on signals and systems, anybody who goes through a quantum mechanics course, and I don't know where else it comes up, but it's one of those things you learn to use operationally. Maybe you feel a little queasy about it, but nevertheless, it gets the job done somehow, and you'd rather not worry about those fine points that all the statements you're making are complete bullshit. What are those statements that are complete bullshit?

You often see it defined this way. A – Delta of X is equal to zero for X different from zero, but zero is infinity. B – the integral from minus infinity to infinity of Delta of X DX is equal to one. This function, which is zero everywhere except at one point, and its total integral is equal to one. C – if I integrate the Delta against any other function F of X, I get the value at zero. Quite remarkable indeed. Everyone one of these statements is complete bullshit. There's just no way to make any of this make sense precisely. But there is something there.

People who used it with some skill were able to do so without avoiding any of the pitfalls, and there are possible pitfalls. You can manipulate Deltas incorrectly and you can make mistakes, but those classical masters, Heaviside, the others who followed him and Derek, in particular, in his applications of quantum mechanics, could make sense of these things and use them effectively. Because they got the right answers and because they were so effective operationally, nobody wanted to admit that these statements were all complete bullshit. So what is to be done?

There's only so much of that you can stand. Some people have higher thresholds than others, but at some point, it had to be cleaned up. There are still cases of this around. The ones that people cite most often now are Findamin path integrals for quantum electrodynamics is where they're used. Nobody can make sense of them mathematically, but you can't deny that they work. It is now an acute [inaudible] challenge to somehow give a rigorous foundation for Findamin path integrals. It doesn't exist yet, as far as I know. It's the same sort of thing.

In the right hands, you can effectively compute with them but feel a little queasy about it somehow. The fact is that operationally, you can understand what's going on here. Delta is supposed to represent a function, which is concentrated at a point. This was probably even said to you when you first learned about Delta. There are various ways of approaching this. You may have seen some of these. It's always involves eliminating process. It's always via eliminating process.

For example, what I mean by eliminating process is what you do is consider typically families of legitimate functions that are getting narrower and narrower and still satisfy these basic properties. There are various ways of doing it, but let me give you one very simple one. You consider a one parameter family of shrinking rectangle functions or concentrating rectangle functions. I'll write it down. I'm going to look at one over Epsilon Pi Epsilon of X. That's the family that I want to consider, and the parameter here is Epsilon, and I think of Epsilon as small as tending to zero.

What do those functions look like? The rectangle function – we know what that looks like, and it's not hard to see what happens when you scale it like that. The ordinary rectangle function, again, is one between minus half and never mind what happens at the end points. That's not important here. It's one between minus a half and a half, and it's zero outside that interval. That's Pi of X. If I scale it, one over Epsilon Pi Epsilon of X, that function is one from minus Epsilon over two to Epsilon over two, so again, I'm thinking of Epsilon as being small here, and if I multiply by one over Epsilon, I'm making it large in the vertical direction, so the height is one over Epsilon.

That's what the graph of that looks like. It's still the case that the area is one. If I integrate this function, I get one. If I integrate this function, it's the area of the rectangle. The rectangle has base Epsilon in height, one over Epsilon – the area is one. As Epsilon is getting smaller and smaller, it is approximating what you think of as an ideally concentrated function. Those properties – at least the first two properties are defining the Delta function. This of Epsilon as getting smaller. It's centering around the origin here. It's zero outside a small interval around the origin.

It's becoming steeper and steeper right at the origin, and what about the [inaudible]? It's integral is equal to one, the integral from minus infinity to infinity of Pi Epsilon of X DX is one over Epsilon. The integral from minus Epsilon over two to Epsilon over two one DX – that's the only place where it's non zero – that's equal to one. What about that final property? If I integrate this scaled rectangle function against a function Phi of X, what happens? If I look at the integral from minus infinity to infinity of Pi Epsilon of X times Phi of X DX, what happens there?

Well, let's take the case where Phi is smooth enough, say, you can actually do it more generally, but just to get a simple idea, imagine expanding Phi in a Taylor series expansion. That is, right this as – this is zero except on the interval from minus Epsilon over two to Epsilon over two. This is equal to the integral from minus Epsilon over two to Epsilon over two of Phi of X DX. Imagine writing Phi as – because Pi of X is equal to one there and it's equal to zero outside that integral. That's equal to one over Epsilon, the integral from minus Epsilon over two to Epsilon over two.

So you write Phi of X as Phi of zero plus Phi prime of zero times X plus and so on and so on. I'm thinking about the Taylor series expansion. That's assuming the function is smooth. You can write similar argument if the function's only continuous, but never mind that. I just want to see what the point is here, why it's concentrating. If I integrate that – I'll put one more term in here. Phi double prime of zero over two times X squared plus and so on. Higher order terms integrated with respect to X. What happens if I carry out the integration?

Well, the first term, that's Phi of zero – Phi of zero times the integral of one from minus Epsilon over two to Epsilon over two one of Epsilon, that just integrates to one, and then the second term, what happens here? Well, this is a constant. If I integrate X, I get X squared. I'm gonna get Epsilon squareds here times one over Epsilon. That's gonna give me a term of order Epsilon. If I integrate X squared, I'm gonna get an X cubed, and if I [inaudible] between minus Epsilon over two and Epsilon over two, I'm gonna get terms of order Epsilon cubed times one over Epsilon. That's gonna give me the terms of order Epsilon squared.

The result is that beyond the first term – beyond the constant term, I'm gonna get terms of order Epsilon or higher order terms. Epsilon, Epsilon squared, Epsilon squared and so on and so on. So what happens is Epsilon tends to zero. This term goes away. As Epsilon goes to zero, this tends to Phi of zero. That is to say, the limit as Epsilon tends to zero of this integral, one over Epsilon the integral from minus Epsilon over two Epsilon over two – let me do it like this. Let me write the whole thing down. The integral from minus infinity to infinity of Pi Epsilon of X Phi of X DX – that's the integral that I just computed – is equal to Phi of zero.

That's what's meant by concentration via eliminating process. Again, I'm assuming, actually – you can tell me if I'm wrong – that you probably saw this calculation at some point. When somebody was trying to justify the Delta function and somebody talked about it as somehow ideal concentration, they probably looked at it pretty much in this way. Again, just to make sure you understand what the issue is here, to consider this, the limit as Epsilon tends to zero one over Epsilon Pi Epsilon of X, if you consider this limit of this function, it makes no sense.

If you consider this one, it makes no sense. But to consider operationally what it means when I integrate this scaled function against an ordinary function and take the limit of the integral, that does make sense, and it produces the value at zero. This limit, Epsilon tends to zero, the integral from minus infinity to infinity – I'll do it like I did before. Pi Epsilon of X Phi of X DX, and to say that that's equal to Phi of zero, this does make sense. That's okay. The fact is by experience, the ways that Delta appeared in applications weren't so much this way, in just a limit of a sequence of functions.

Really, it occurred operationally when it was paired with another function and somehow the idea was by eliminating process, you were concentrating things and you were just pulling out the value at the origin. That's really operationally how it appeared. That was an extremely important thing to realize.

These statements somehow – again, individually, these statements just don't make sense and can't be made to make sense. But in practice, the way it was used, you replace what you think of as idealized Delta by some sequence of functions, which are concentrating, and you consider them as paired with the function via integration, and then you can do everything you want to do in a context that you have a certain amount of confidence in.

Student: Outside the integral, is there a one over Sigma?

Instructor (Brad Osgood):Where, here? No, because I – yeah, thank you. Sorry. Now, we're almost there, and once again, this is one of these tipping points where you look at the accumulated body of evidence. You say to yourself what's really going on here and again, the mathematical modus operandi somehow is to turn the solution of a problem into the definition.

Student:[Inaudible].

Instructor (Brad Osgood):You scale this thing, right? This doesn't make sense. This should be there. I messed it up somewhere else? Now is everything okay?

Student: Over there.

Instructor (Brad Osgood):I guess I was thinking when I put the scale in here that that was also scaling the outside, and I was wrong. Sorry. Is everything okay now? This was a big conceptual step. Again, it follows the mathematical modus operandi of turning the solution of a problem into a definition. We're gonna concentrate – instead of concentrating on somehow the limiting behavior and so on, the idea is to concentrate on the operational outcome of concentration in one case and then more general operations. I want to change the point of view. It really requires a fundamental change of point of view here.

To capture this idea and to include much more and how it's going to include much more I'll explain to you in just a second. We need a change of point of view. It becomes operational. It becomes an emphasis on the outcome rather than the process. The focus is on the outcome rather than on the process. What I mean by this is in the case of Delta, the outcome was at the end of the day, it concentrated in pairing this approximating sequence of functions gave you the value of the function you were interested in at zero and that had to be done – the process was taking a limit. There was a limiting process involved in that.

We want to concentrate on the outcome and actually getting the value of the origin rather than the process. Here's how you set that up. There are several aspects to it. What I'm going to do, really, is write down the definition or axioms for a class of generalized functions – a class of distributions, which are going to include the Delta function. It's going to capture the essential nature of the Delta function, and it's actually going to, as it turns out, include much more. There are several aspects to the definition. This is the definition of generalized functions or distributions.

First, you start out with a class of test functions. You start with what are called test functions. When the Fourier transform comes back into the picture, this is going to be the class of rapidly decreasing functions, but for other problems, you might consider a different class of functions, but generally speaking, these are the sort of best functions for the properties you're worried about. You think of these as the best functions of [inaudible] or the best functions for the problem at hand or the given area of application. Again, for Fourier transforms, it's going to be the Schwartz's functions – the rapidly decreasing functions.

For other functions, it may be those functions which I mentioned last time of compact support, the functions which are actually not just tending to zero outside some finite interval but which are identically zero outside some finite integral. Two, associated with these test functions is a class of what are called generalized functions of distributions. A distribution – I'll call it T – is a linear operator on the test functions that produces a number. It is a linear functional. The distribution T is a linear functional on test functions.

What that means is I give you or you give me for a test function Phi T of Phi – T operates on Phi – produces a number, and it's linear. Typically, you allow complex numbers here. T is linear. That is to say T of the sum of two functions is T of Phi one plus T of Phi two and T of – it obeys the principle of super position. T of Alpha times Phi of Alpha times T of Phi. A distribution is a linear operator on the class of test functions. You start by defining the class of test functions that somehow is going to have all the nice properties you could possibly imagine for your problem.

A distribution, also known as a generalized function, is an operator on those functions. It produces a number. The final property is – you don't want to give up taking limits completely because limits do come into the subject, and so you assume that these linear operators are continuous in the following sense. Three – the final property is the continuity property. That is if Phi N is a sequence of functions which converge to a function Phi, then that implies that if I operate on it with one of these degeneralized functions [inaudible] distributions, that that converges to T of Phi. I'll say more about this in a second.

This is the most problematic part of the definition. The continuity property means that if you have a sequence of test functions that are converging to another test function, if Phi N converges to Phi, then that implies that if I operate on the sequence with a distribution with a generalized function that produces a sequence of numbers. On the left-hand side is a convergence of a sequence of functions. That's hard. Again, I'll come back and talk a little bit more about that later. On the right-hand side is just a sequence of numbers. That's easy. It's easy to talk about a convergence of sequence of numbers.

These numbers converge to that number, because you don't want to abandon taking limits completely because it does come up in the applications. I want to introduce a little terminology here, a notation that's used in this subject, and that is you often say that a distribution is paired with a test function. Instead of saying it's operating on a test function, that's what's going on. You often say it's paired with a test function, and again, you'll see a reason for this in just a second. You often write – a notation for the pairing is often written like this with angled brackets.

T is paired with Phi – this notation is supposed to just indicate some alternate notation for writing T operating on Phi. Both notations are in use. This notation is probably a little bit more common. This is not an inner product. It's supposed to indicate that T is somehow operating on Phi to produce a number, and the operation is linear. If I take two functions, T of Phi one plus Phi two is T of Phi one plus T of Phi two, and T of Alpha Phi is Alpha times T of Phi. I know this sounds like deep waters here, but when you see how this works and you see how effectively you can compute with this, it's really quite stunning.

It was not so easy to do. Adopting this point of view to give a rigorous foundation for Delta and then actually to also develop the Fourier transform was no less revolutionary to the whole shift from classical mechanics to quantum mechanics. It required a different point of view. You had to look at things differently. The theory of distributions is a more accurate way and a more effective way of dealing with the problems you really want to deal with. That's just the way it is. Again, let's go back and recover Delta.

Let's recover Delta in the context of this definition. What is Delta doing operationally? Delta operationally – you say what is the outcome of applying Delta? It is to pull out the value of the function at zero. At the end of the day, that's what Delta is supposed to do. You wrote down this nutty integral. I'm not exactly trying to talk you out of it. I'm just trying to say that there's another way of looking at it that makes more sense. Operationally, the effect of Delta is to pull out the value of the origin. It is to evaluate the function at the origin.

So you say that to yourself a couple times – operationally, the effect of Delta is to evaluate the function at the origin. The mathematical modus operandi is turn the solution of the problem into the definition. That's how I should define Delta. Define Delta, according to this definition on test functions – it's supposed to be a linear functional on test functions. How is it defined? You give me a test function. I have to tell you how Delta operates on it.

Student:I thought the function was paired with not the same function it operates on.

Instructor (Brad Osgood):No. I mean, I don't know what you thought, but this is what I'm saying. This is a common notion of pairing. That is, Phi is a given test function. T operates on the function Phi, so instead of writing T operating on Phi, which is sort of a functional notation, you often write this notation as an alternative. It's very common. You actually also see this in physics when they talk about broad vectors and [inaudible] vectors. One sort of vector pairs with another kind of vector in physics, and here, an operator pairs with a function. There's a class of functions that it operates on. This is the operator.

What I'm saying is that the use of the word pairing there is appropriate. Like I say, let me go back to Delta here. You say to yourself, the operational effect of Delta is to evaluate a function at the origin. Turn that into a definition. Define Delta by Delta paired with Phi is what? Phi of zero. Phi is a class of test functions. That is given to you. You give me a test function. I have to tell you what Delta does to that test function, and then I have to verify that it satisfies the properties of a distribution. I say Delta operating on Phi is nothing but Phi of zero. Is it linear?

Well, what is Delta paired with Phi one plus Phi two? That is Phi one plus Phi two at zero. By definition, that's what Delta does. It evaluates Phi one plus Phi two at zero. That is, of course, Phi one of zero plus Phi two of zero, which is Delta paired with Phi one plus Delta paired with Phi two. That is Delta paired with Phi one plus Delta paired with Phi two. It's similar for the scale of multiplication property. How about continuity?

Again, without saying precisely what I mean by a sequence of functions converging, what about this statement, that a Phi N converges to Phi in a sense that you just imagine – so a sequence of functions converging to another function – does that imply that Delta Phi N converges to Delta of Phi? Does that imply that Delta paired with the sequence Phi N converges to Delta paired with Phi? Well, write out both sides. What is the left-hand side? What is Delta paired with Phi N? Delta paired with Phi N is just Phi N of zero by definition.

Well, if a sequence of functions Phi N is converging to Phi, then surely Phi N of zero is converging to the value of zero. If Phi N converges to Phi, then surely, Phi N of zero converges to Phi of zero, and so surely, it must be the case that Delta paired with Phi N operating on Phi N converges to Delta operating on Phi. It's a continuous linear functional. I want you to think about this a little bit, and I hope appreciate it, because this mysterious Delta function that was defined by these ridiculous properties – it's zero everywhere except at one point, where it's infinite.

It's total integral is equal to one and it pulls out the value of the function when you integrate it against it – those ridiculous properties have been, in effect, captured in the simplest possible distribution. This complicated limiting operation that we talked about in terms of concentration operationally is defined completely air tight by evaluation at zero. The simplest sort of operation – evaluate at zero. That captures this mysterious notion of Delta. It's very impressive. If that was all you could do, you'd say that's a lot of work.

I don't want to completely change my worldview so you can define this one little distribution that I was perfectly happy taking limits of anyway, and I know those statements were bullshit, but I'm happy enough with bullshit. I can tolerate ambiguity. Why do you really put me through this? The fact is that it goes far beyond this particular example. The fact that it captures this particular example so easily and so effectively is already a good thing. If it were the only thing, it would have withered on the vine. It allows us to define very robustly Fourier transforms and everything else.

Let me give you one other slight version of this. You have also probably seen a shifted Delta function in your work with Delta functions in other classes. You have also probably seen, I imagine, some statement that looks like this – the integral from minus infinity to infinity of Delta X minus Y F of Y DY is equal to what? F of X. You've probably seen statements that look like that. Delta sifts through the values and so on. That statement is – never mind. That doesn't make any sense. What do you want to define here operationally? You want to define a shifted Delta function.

If this is the Delta function based at zero, Delta applied to Phi pulls out the value at zero. What do you want to define to capture this statement precisely as a distribution? You give me a test function. I tell you what the pairing should be to pull out the value at number X. So what do you define? Give me a name for a distribution and tell me how it operates. You want to capture this property of a shifted Delta function. I know you're trying to think of this in terms of convolution. Don't think about it in terms of convolution. Just think about operationally trying to pull out the value of the function at some point other than the origin.

What is the distribution that will do that? Define a distribution that will do that. Right. So I define – it's a different distribution. So define Deltas of A as a distribution by the formula Deltas of A paired with Phi is Phi of A. So the case we had before is when A is equal to zero. Is that linear? Is that continuous? You can check. It's the same idea. Delta of A – this defines Delta A. You give me a test function. I have to tell you how the distribution operates on a test function to produce a number. What do I tell you?

You give me Phi. I say Delta A operating on Phi produces a value at A. Air tight. No ridiculous statement like this. No limiting processes involved. It is a straight definition based on what you want the outcome to be. Take a deep breath. Exhale. I claim that we have gained something. We've certainly gained something in clarity or rigor, if you want. This mysterious Delta function has now emerged operationally as the simplest possible distribution. Oh, the struggle. One question that you might well ask at this point is have you lost anything? Yes, you've defined Delta, so we've gained Delta. But have we lost anything?

What I mean by this is ultimately, the test functions are very restrictive. The test functions might be a very restricted class, these rapidly decreasing functions. They don't include constant functions. They don't include the rectangle function. They don't include the triangle function and so on. How are those functions going to get back into the scene? How are the rectangle function, the triangle function, trig functions etcetera going to come back in? How are they going to come back into the picture?

Delta is this bizarre thing, and true, it was defined in a pretty simple way, but I really want to get to the point where I can consider the functions that I really want to consider – triangle functions, rectangle functions, sines and cosines and so on. Can I consider those in the context of generalized functions? More to the point, when we get to it, can I actually consider those in the context of actually taking Fourier transforms, because that's what I want to get to.

I want to get to defining more general Fourier transforms so the Fourier transform of the Delta's gonna make sense and the Fourier transform of constant functions is going to make sense, so the Fourier transform of sines and cosines is gonna make sense – how can I do that? If I can't do it classically by an interval, how can I do that in this context or can I do it in this context? I want to explain now how generalized functions include, in a natural way, the sort of ordinary functions. That is to say, I haven't lost anything. I haven't lost those functions that I really want to consider. They're there.

They are in there, but they're in there in a slightly different way. You can consider – this is a question of how to consider ordinary functions in this context. Again, I want to consider, for instance, the constant function one. How do I consider the constant function one as a generalized function or as a distribution? You buy the premise, you buy the gag.

You want to consider this is a distribution. What does that mean? It always means the same thing.

If you want to consider something as a distribution, that says you give me a test function or I give you a test function and you have to tell me how your new thing here is operating on that test function. How do we pair – given the test function Phi in whatever class you're considering, how do we define a pairing of one and Phi? Well, it's actually very simple, but it's again maybe not one of those things you would – it takes its queue from really how these things grew out of the classical applications and the classical view of things. I'm going to define it by integration.

To pair one and Phi, I have to get a number. A distribution operates [inaudible] over and over again, and this is what you have to say to yourself. A distribution operates on a function to produce a number. It has to be linear. It has to be continuous. One has to operate on Phi somehow to produce a number. The pairing is by integration. That is to say, you give me Phi, I have to tell you how one pairs with Phi, and here's my definition. One pairs with Phi as the integral from minus infinity to infinity of one times Phi of X DX. A lot of big buildup for a very simple definition. That's what it is.

That certainly produces a number, and if Phi is a good enough function, this integral is going to converge. Different values of Phi will give me different numbers. In some sense, I know all about one if I know all about the integrals for different values of Phi. There's nothing about one you can ask me, somehow, that I can't answer if I integrate it against different sorts of functions. Of course, there's not much you can ask me about one that I can't answer anyway, but operationally speaking, there's nothing you can ask me about one that I can't tell you if you allow me to integrate it against any old test function Phi.

More generally, I want to include the rectangle function, the triangle function and everything else and consider those as distributions in the same way. Likewise for the rectangle function, the rectangle function pairs with a test function by integration – the integral from minus infinity to infinity Pi of X Phi of X DX. Phi of X may be very smooth. Pi of X is not smooth, but the product makes sense and the integral makes sense. That's my definition of how Pi pairs with Phi. That identifies Pi, the rectangle function, as a distribution. This definition identifies or defines the constant function one as a distribution. You can check.

Integration is a linear operation. Linearity works. So does continuity. That's a little bit more complicated. That requires certain limiting theorems for integrals, but never mind. I'm setting some of those details aside right now. In general, the same thing for a trig function. How about sine of two Pi X? You can't integrate that function from minus infinity to infinity, but you can integrate it if I pair it with a function, say, that is decreasing. By definition, sine of two Pi X can be considered as a generalized function if I tell you how it operates on a test function.

How does it operate on a test function? It does so by integration – the integral from minus infinity to infinity sine of two Pi X times Phi of X DX. The [inaudible] sine of two Pi X doesn't make sense. That integral doesn't converge. But if I multiply it by a function which is dying off, the integral will converge. For the purposes that I'm going to want to consider, sine of two Pi X can be considered as a generalized function, but for these purposes, it can also be considered as a generalized function because I can tell you how it operates on test functions. It operates by integration.

Again, this seems like a lot to absorb. You have to go along for the ride a little while longer and then you'll see how this works. How you see how to compute with this is really, I think, pretty impressive. In general, many very wild functions and some not so wild functions can be considered generalized functions by this pairing. If F of X is "any" function, you can consider F of X as a generalized function or distribution by defining its paring. F paired with Phi is the integral from minus infinity to infinity F of X Phi of X DX. F operates on Phi by integration. That's a linear operation.

Again, for not all functions will this integral converge. You can stick a really wild function in here and maybe it's not going to work. But for most garden variety functions and more than most garden variety functions for things that can be pretty wild, that sort of integral will make sense because the test functions are so nice. The nicer you make the test functions, the more wild functions you can stick in here. For Fourier transforms, we're gonna find that the Schwartz functions as test functions are just the right class. They're the ones that are gonna allow us to include sines and cosines and Deltas and all sorts of things like that.

So one says in this context that a function determines a distribution. How does a function determine a distribution? You give me the function. I have to tell you how it operates on a test function. It operates by integration. You really have to say, all right, that operation is linear. That operation is continuous. That all works out fine. We haven't lost anything in the sense of the class of generalized functions includes all the functions that society needs to function. Next time, you're going to see how this comes together to define the generalized Fourier transform and more. See you then.

[End of Audio]

Duration: 53 minutes

TheFourierTransformAndItsApplications-Lecture13

Instructor (Brad Osgood):On the air. Oh, what these people miss before the camera starts rolling. Okay. I sent out a note yesterday, or over the weekend about the mid-term coming up. So, the midterm is coming up a week from Wednesday, and without going into too much detail right now, we're going to have three sessions – it's a 90-minute exam outside of class, so we'll have class that day.

I am scheduling in three sessions: 2:00 to 3:30, 4 to 5:30, and 6:00 to 7:30, so there's a sheet circulating around, I'm asking you to sign up for one those times. This is not a contract you're signing, all right? This is just so I have an idea the relative sizes of the different sessions are, so I can find an appropriate room. And I'm hoping by having three of those times that that should pretty much take care of everybody. If there are people who can't make it any one of those times, then please let me know, we'll make some special arrangements for it. I'll have more details about the mid-term as we go on later in the week. But the general idea of it is that – I don't want it to be too computationally intensive. That is to say, I'm trying to make the exam, sort of, a little bit more on the conceptual side, so you don't get bogged down in integration by parts and things of – you can't avoid all computation of course, but the way we're going to try to write the exam is to try to keep it on the conceptual level rather than detailed calculations. And let's see, it'll be open book/open notes. Now, sometimes people ask me if they can bring any book they want and I say, 'Yeah, you can bring any book you want.' But this leads to ridiculous events where I've seen, you know, I've seen students walk into the exam with, like, a stack of 10 "Signals and Systems" books, now, which is just ridiculous. But, if you want to, hey, that's okay.

I'll provide the formula sheet; that's already posted on the web, and the blue books and all the rest of that jazz. And let's see, I've posted already also last year's mid-term exam plus the solutions, so that's there on web. And also, the most recent – the next homework assignment is also up. Okay. Any questions about anything? Any questions or comments? Nope. Okay. Big day today. Big day, big day. It's finally time to reap some of the benefits of the discussion, rather general, and I think it's safe to fair say, abstract discussion we've been having. It'll still – it may still seem a little abstract to you today, I understand that, so just ride along, all right? And you'll see today some amazing formulas that come out of this effortlessly in finding the Fourier transform of some well-known functions – things we're really gonna have to use. It's really just, I don't know, a bucket full of miracles and the fun never stops. So, I don't want to talk today about the Fourier of transform of a generalized function, or a distribution, also known as distribution Fourier transform of a distribution. So let me remind you, first of all, what the setup is, what goes into this. To define a distribution you need a class, first of all, a class of test functions. So the setup is, you first need – you first have to define a class of test functions, or test signals that usually have particularly nice properties for the given problem at hand. And it can vary from problem to problem. For us, for the Fourier transform, it's the class of rapidly decreasing functions. So these typically have particularly nice properties.

Sorry for not specifying that terribly carefully. But they come, generally, out of – again, sort of, years of bitter experience with working with problems, working with a particular class of applications and trying to decide what the best functions are for the given class. For Fourier transforms, the class of test functions is the rapidly decreasing function, so I won't write down the definition again, but I'll remind you of the main properties in just a minute when we need it. Rapidly decreasing functions – these are the functions which are infinitely differentiable and any derivative decreases faster than any power of x, so they are just as nice as you could be. Rapidly decreasing functions. All right. That's what you need. You need a class of test functions, and then by definition, a distribution or a generalized function is, to use the most compact terminology, a continuous linear functional on the set of test functions. Distribution, also called a generalized function, this term is probably not used – generalized function is probably not used as much as it once was when the subject was new. Nowadays, people just call it distribution since Schwartz is really unifying work on all this. Is – so let me give you the shorthand notation for this, a continuous linear functional on test functions. All right. So that means it satisfies two properties. So one, so you write the pairing if is a test function, and t is a distribution, we will write – most of the time, but not all of it, not always t, , with that sort of pairing for t operating on .

All right. So if you want, you can think of as some distribution of physical information, temperature, voltage, current, whatever; and t is a way of measuring it. All right. So you measure something, you get a number. All right. So t is a measuring device and is the thing you are measuring. So t operates on in such a way that it satisfies the principle's super position, or that is to say, it's linear, and it's also continuous. So t is linear, meaning t of 1 plus 2 is the same thing as t operating on 1 – I'll write up here – t operating on 1 plus t operating on 2. And the same thing for scaling. So t – apply – t operating on alpha times , is alpha operating on t of . That's the property of linearity. And the property of continuity can be phrased in different ways, but the most direct way of writing it – and I'll have a comment about this in just a second – is that if n converges to , a sequence of test functions converging to a test function of , that implies that the measurements of n also converge to the measurement of the final test function.

So, really, there's nothing wrong with trying to impart, sort of, a physical interpretation, but I don't wanna try to push that too far. But if you think of as some sort of something to be measure and t as the thing that does the measuring. Now, I wanna make a comment about the mathematical – the one – one quick comment about the mathematical for the foundation for this. The hard work in all of this is this statement right here – n converges to v. All right. The hard mathematical work is to define n converging to , what that means. Because these space of functions, although they – each individual function has particularly nice properties – infinitely differentiable, all the derivatives decay, and so on and so on – if you want to talk about a sequence of functions converging, then the more properties you insist a function having, the harder it is to control the sort of convergence. You know, if the function is infinitely differentiable then you want all the derivatives to also to converge. If the function is rapidly decreasing, you want all the derivatives to be rapidly decreasing, and so on. So it's hard to write conditions down that are gonna control this, that's what's difficult to do. And if you open a math – this is a warning – if you open a math book that says something like, 'theory of distributions or theory of generalized functions – it's gonna be, like, the first hundred pages or so, are gonna be devoted to analyzing what it means to say that n converges to . And you look at that and you'd say, 'Why would anybody learn this subject?' and I wouldn't blame you for saying that. All right. It's complicated, there's no two ways about it. If you're gonna give a very precise treatment of it, then you really have to do that. We don't have to do that, all right? I'm gonna take this sort of, again, as intuitive as you can make that. Imagine controlling the functions, all the derivatives and so – you can think in terms of the graphs, the graphs of n converge to the graph of , and everything is controlled as nicely as you want. So suffice it to say, it can be defined precisely, but that's not for us to do. All right. But I did want to point that out.

The other bit of terminology here for those of you who have seen this, for those of you have studied a certain amount of extra math, and particular the field of – it's usually called functional analysis – is that one says that the distributions are – the set of distributions is the dual space of the set of test functions. That's another way of phrasing this. So the distributions – the particular class of distributions that you're considering are the dual space – is, are – dual space of the space of test functions. Bring that up at a cocktail party some night, watch the people scatter. All right. Okay. Now, we had a couple of examples last time and I want to remind you of that. The first example we had was delta defined as a distribution, and it's the simplest distribution. This mysterious delta function that's supposed to capture this property of concentration at a point emerges as simply evaluation. And so examples: delta. All right. So, delta as a distribution is defined simply by delta operating on is of zero. All right. This replaces – now that's a pairing – once again, in all this what you're going to hear me say a lot today is, I'm gonna define a distribution. What does that mean? That means, you give me a test function, I have to tell you how to operate on it. All right. So you give me a test function of , to define a distribution, I have to tell you how that distribution operates on . So I say, 'How does delta operate on ?' Delta operating on is just of zero, nothing else. Period. All right. This makes precise – this dispenses with all of that bullshit about the classical – the classical definition of delta is zero everywhere, except at one point where its infinity and the interval is one, and so on and so on. All that is gone, all that is swept aside by this definition. Now, you know, it took awhile to get to that definition, okay, but once its there, that's what you can work with. Okay. Likewise, there's the shift of delta. I called it delta sub-a, and that's just defined by – so, again, to define a distribution you give me a test function, I have to tell you how it's defined. I have to tell you how operate. And I say, delta a operating on is of a. Period. Period.

Okay. Now, I would be remiss if I didn't indicate – there's a graphic way of picturing these deltas, all right? So the graphical picture – and we'll use this too – usually indicate the delta function by a spike at the origin, delta, all right. And since it's supposed to have infinite height, I can't exactly draw an infinite spike, so I just put a little arrow there. Everybody writes it like this, and likewise, delta sub-a is usually indicated by a spike at a, okay? Now, something else. All right. Since we've been so careful about this, and since I've made such a big deal out of the fact that it's all airtight and we can dispense with intervals and things like that, I am not going to blame you and I am not going to draw –

I'm not going to try to derive out of your system, expressions like this. Interval for minus infinity to infinity of delta of x, of x, x, is equal to of zero. If you want to write that, okay. I'll even write it from time to time. Its okay, nothing bad will happen to you. All right. Among friends, it's perfectly okay. And a corresponding thing with a shifted delta function, it's all right. All I want you to be aware of is that one can – this is just really a mnemonic in some senses – this is in some way a more complicated way of writing that, all right? But in some formulas and in some contexts it's helpful to write that and, okay. All right. Do not be afraid now that you have the hidden knowledge. All right. What's really going on because this is what's really going on? The second example we had of distribution was really an example of the whole class of distributions, and that was by way of assuring us that nothing has been lost. All right. Now, these things are sometimes called generalized functions, so if they're generalized functions they ought to include the original functions, the good old functions. And they do – that is the second example, we say a class of examples – are really distributions induced by – this is probably the best way of saying it – induced by functions. All right. So if of x is a function such that I can integrate it against the test function, such that this interval minus infinity to infinity, say of x, of x, x makes sense, the interval converges. All right.

And that will be the case for many functions because is such a nice function; may be such a nice function that, even if is a bad function, the product is gonna be such that the integral converges. All right. If that makes sense, then that's a number after all, I integrate of x against of x and that's linear, as the sum of the integral, the integral of a sum and so on, and it's also continuous. That's a little harder to show, and so that defines – that defines a pairing of and . Continuous – again, I'm not gonna show that a pairing of and . In that sense, defines or reduces a distribution. You give me a test function; I have to tell you how the distribution operates on it. How does it operate on it? It operates by that pairing. And so for shorthand, I write paired with – what I'm really thinking of here is the distribution induced by , but that's' too many words. So of x times of x would be x. Okay. So a lot of functions which are not as nice as the test functions are included in the set of distributions instead of operated – functions – things that operate on test functions. So for example, it may be the case that one pairs with , so this includes many functions. Many functions induce distributions. So it can be regarded as distributions. So e.g. – it may be the case that integral of one minus infinity to infinity, of x, x, which is the integral for minus infinity to infinity of one times of x, x – that may make sense. It should make sense; you don't want a function which is not integral for goodness sakes. And so in that sense, one reduces the distribution because I can allow one to operate on by means of this integral. Okay.

And likewise, signs and cosigns are not themselves integral, but if I multiply them by function, which is really dropping down, which is decreasing at plus or minus infinity, then the integral of sign or cosign against such a function will make sense. And in that sense, sign and cosign can be considered as distributions, that is they operate them. So let me – well, just for short, the complex exponential often works. So either the two pi i a x is – might be okay – provided the interval for minus infinity to infinity – either the two pi i a x, of x, x makes sense. Which for many functions it will. Okay. It will, even though either the two pi i a x itself is not integrable, if I multiply it by a function which is decreasing, the product will be integrable, say. All right. So that induces a distribution. How does it operate on the function? It operates on the function by integration. Okay. All right. And if you look at this example and if you think of a Fourier transform coming in, you wouldn't' be too far off, because now the big moment has arrived where I'm going to define the Fourier transform distribution. So we talk about distributions and how you'd – the definition of distributions and a couple of examples of distributions, now I wanna do something with them. And this is so cool how this works, I think. Again, just enjoy the ride. All right. And watch how these derivations go. It's almost effortless the way this all works. I mean, it's just so – it's better than sex; not that I would know, I was a math major, but –

So I want to define the Fourier transform of a distribution. So now we're gonna look – the test functions we're gonna look at now are the Schwartz functions. The rapidly decreasing functions. So let's take the test functions to be s – the rapidly decreasing functions, all right? Now, let me remind you of why they are so good Fourier transforms, all right? Why they are so good for Fourier transforms is if is in s, then so is its Fourier transform. All right. And for that matter, I guess I don't think I said this, but for that matter, the inverse Fourier transform is also an s. Okay. An inverse Fourier transform differs from the Fourier transform only by a minus sign, so if one isn't there, it's easy to see the other one is. So that's one property. And the second is that Fourier inversion works. All right. It is in that sense – those sense – let me write the second one down. That's the first property is that the function is rapidly decreasing, so is it's Fourier transform. Second is that the inverse Fourier transform of the Fourier transform of is equal to and, likewise, the Fourier transform of the inverse Fourier transform of is equal to . Those are the properties that make the right class, the best class, of the classical functions to use when you want to talk about the Fourier transform. The classical Fourier transform is defined by the integral. All right.

So these are the functions we're gonna use. The corresponding class of distributions is sometimes called the class of tempered distributions. Distributions are called the tempered distributions. All right. And what I wanna do now is I wanna show you how to define the Fourier transform of a tempered distribution to be another tempered distribution. Okay. If t is a tempered distribution, I want to define its Fourier transform. And this will be another tempered distribution. Okay. Now, again, you will hear me say over and over again today, I want to define a distribution. What does that mean? That means you give me a test function, I have to tell you how it operates. So I have to tell you how to define the Fourier transform, something I'm gonna call the Fourier transform of t operating on , or paired with . So I have to define the Fourier transform of t operating on , where is a test function – rapidly decreasing function. What shall I – how shall I do it? How shall I do it? All right. There is a guide in this. There is a guide in all things, all questions of this type. You wanna know how you should define an operation on a distribution, you ask yourself – you start off by asking yourself, 'What would I do if I were in a really good situation here?' Like, if of t were itself a rapidly decreasing function, say, or some other good function and this pairing was my integration. All right. So what if? This is another example of suppose the problem is solved, let's see what has to happen. All right.

So what if the pairing is by integration? It may not be in general, okay? I mean, delta doesn't rise by integration, but suppose it does arise by integration; what would the formula – where would I be led? Where would I be led? So I suppose that everything works out as smoothly as possible, what are the consequences of that? All right. So let's say that the Fourier transform paired with is the integral for minus infinity to infinity of a Fourier transform of t of x, of x, dx. All right. So, again, this would not be necessarily the case in general because distribution is not always given by integration. But suppose it is. And what has to happen? All right. So suppose – and again – I'm supposing that everything nice is happening, so the Fourier transform is given by an integral, it's the integral for minus infinity to infinity – e to the – I've already used x here, so let me write this – minus two pi i x y, t of y, d y, then times of x, dx. Supposing that I can do this, all right, supposing that everything is nice enough that I can do that. All right. Then – all right – you know, you buy the premise, you have to buy the gag – you have to follow your pencil all the way through and see where it leads you. So then I can combine all that – let me write this as one big integral: minus infinity to infinity minus infinity to infinity, then I'm gonna split it up again. So e to the minus two pi i x y, t of y, of x, dx, dy. Or I guess I wrote it dy, dx – dy dx – doesn't matter because now I'm gonna split everything apart again. Dy dx. All right. So now I'm gonna swap the order of integration, and I'm gonna put the – did I call it or ? Sorry. This is ; my apologies; and this is , okay. So now I'm gonna swap the order of integration. I'm gonna put the with the complex exponential and leave the t alone. Alone.

And write this as the integral for minus infinity to infinity, the interval for minus infinity to infinity – e to the minus two pi i x y, of x, dx, and then the result is integrated against t of y, dy. Now, what is inside the integral? Everything is as nice as it could possibly be, he said bouncing up and down. So this is the integral, this is the Fourier transform of , evaluated at y because I'm integrating e to the minus two pi x y, of x dx, so this is the integral from minus infinity to infinity of the Fourier transform of at y times t of y, y. Okay. Where do we start? Where do we finish? We started with – supposing everything was as nice as it could be, and the pairing was given by integration, we had this expression: the Fourier transform of t paired with , is the integral for minus infinity to infinity of t of y. The Fourier transform of y dy, and you have to look at this and look at this backwards in the sense that what appears here is the pairing of t and the Fourier transform with – of . All right. If everything is as nice as possible, and the pairing is given by integration, then what has occurred by our following our trusty pencil is that this is the pairing of t with the Fourier transform of . Okay. So now you say to yourself, 'If everything were as nice as possible, how would I pair of t with – what would the Fourier transform of t be paired with ?' It would be the same as if t were paired with the Fourier transform of . Okay. All right. Now, this right-hand side is gonna make sense regardless of what I did before I got to it. All right. Why? Because t is a tempered distribution; t operates on Schwartz functions. T operates on rapidly decreasing function. If is a rapidly decreasing function, so is it's Fourier transform. It's the property of – that's the property of Schwartz functions; that's the property of rapidly decreasing functions. So it makes sense for t to operate on this. All right. This left-hand side may not have initially made sense, but this right hand side will make sense. All right. Now, bless your souls. What is the thing to do? The thing to do is to turn the solution of the problem into a definition. All right. Turn this into a definition. Define the Fourier transform. So again, I'm telling you – so if we're given distribution t – tempered distribution t, so t is tempered distribution. I wanna define the Fourier transform; I wanna define a distribution. You give me a test function; I have to tell you how the distribution operates on that test function. I wanna define the Fourier transform of t by how it operates on a test function. The Fourier transform of t operating on is, by definition, t operating on the Fourier transform of .

On the right-hand side, this is the classical Fourier transform of , given by an integral because is as nice a function as it can be. This is a new , so to speak. This is a definition of a Fourier transform of the tempered distribution t. How do I define the distribution? You give me a test function; I have to tell you how it operates. It operates by t operating on the Fourier transform. Now, the right-hand side makes sense because if is a rapidly decreasing function, then so is of . I could smoke a quiet cigarette here, but I don't smoke. It's so cool. Now, look. You might say, 'What a cheat.' I mean, what a cheat. I mean, you're telling me after all this that the Fourier transform of t operating on is t operating on the Fourier transform? That's what I'm telling you, and that's because I'm compelled to tell you that, all right? It has to work out that way if everything else is gonna be consistent. All right. Now, how should we define the inverse Fourier transform? If we define the Fourier transform, how shall we define the inverse Fourier transform of a distribution? Well, the inverse Fourier transform of t operating on must be nothing other than t operating on the inverse Fourier transform of . Okay. What else is it gonna be?

And now, let's proof Fourier inversion. Fourier inversion would say the inverse Fourier transform of the Fourier transform of t is equal to t for any distribution. And also the Fourier transform of the inverse Fourier transform of t is equal to t. All right. It's the most important theorem in the field. All right. If you can't invert the transform, it's not gonna do any good practically, so this is the most important theorem, and it's trivial. It's an absolute triviality. Why? Because everything has been carefully defined. It is a triviality because all the terms in this expression, and how to operate with them have been precisely defined. What is the inverse Fourier transform of the Fourier transform t operating on ? The inverse Fourier transform of something is equal to the something – Fourier transform – applied to the inverse Fourier transform of . What is the Fourier transform of t operating on this test function, which is legit because the inverse Fourier transform is again a rapidly decreasing function, it is t operating on the inverse Fourier transform – excuse me – on the – excuse me – it is t operating on the Fourier transform of the inverse Fourier transform of . Fourier transform operating on something is the something operating on the Fourier transform. Its t operating the Fourier transform of the inverse Fourier transform of . By – in the space of functions, which are as nice as possible for Fourier transforms, classical Fourier transforms, Fourier inversion works. So this is t operating on . Where did we start? Where did we finish? We have found that the inverse Fourier transform of the Fourier transform t operating on is the same thing as t operating on . For every test function , and therefore, the inverse Fourier transform of the Fourier transform t must be t. Period. The most important theorem on the subject emerges absolutely effortlessly because all the terms were properly defined. The calculation just took care of itself. All right. The effort to try to prove this in the classical case is just murder. All right. It's just murder. You can prove it in the case of Schwartz functions.

You can prove it in the case of rapidly decreasing functions, and the proof actually is not in all detail, but in most of the details is given in the notes. All right. And there it goes pretty smoothly because the functions are as nice as possible. But if you try to start to prove Fourier inversion when the functions are not quite as nice as possible, you run into all sorts of mathematical snares. All right.

But if you broaden your perspective, and that's what's really required here, if you broaden your perspective on this and let these new ideas is, then it becomes just a piece of cake. It becomes a deduction immediately from the definition. All right. Now, let's have a Fourier transform hit parade. Let's calculate some Fourier transforms because one of the things I said to you, and I want to show you now in a way I hope you'll believe, is that you can calculate with this definition. It's not just for proving theorems; it's not just for making things rigorous, it's also provides you a way of calculating in a way that you can have greater confidence in your answers. Let's find the Fourier transform of delta. All right. The mysterious delta function, I mean, trying to find the Fourier transform of that, I mean, my God, you know, what could be harder in the classical case? But when you give the setup in terms of distributions, it is so simple. Watch. To define the Fourier transform of delta, by definition, the Fourier transform of delta operating on a function , is delta operating on the Fourier transform of . That's the definition of the Fourier transform. Okay. You give me a test function; I have to tell you how it operates on the test function. The Fourier transform of delta operating on , is by definition, delta operating on the Fourier transform of .

But now, how does delta operate on a test function? Delta operates on the test function by evaluating. That's the Fourier transform of evaluated at zero. Okay. But now, what is the Fourier transform of at zero? Now I write down the classical definition of the Fourier transform and evaluate it at zero. I can do that because is just as nice a function as it can be. This is, by definition, the integral from minus infinity to infinity of either the minus two pi i zero times x, of x, dx. In other words, is the integral for minus infinity to infinity of one, e to the zero, one times of x dx. And now, you have to look at this, and have to look at it with new eyes – the integral for minus infinity to infinity of one times x is the function – the constant function one paired with if I regard the constant function one as inducing a distribution. Where do we start? Where do we finish? We found that the Fourier transform of delta paired with is the same thing as one paired with , and that is true for any test function , and therefore, the Fourier transform of delta is equal to one. The simplest of all distributions has the simplest of all Fourier transforms – the constant function – one. The Fourier transform of delta is one. Airtight. Airtight. All right. Now, you may very well have seen this in other classes. You may very well have seen other, sort of, tortured derivations of this property. If the Fourier transform of delta is one, but this is the right derivation of this property. Okay. Nothing is in question here, absolutely nothing. Now, you see how easily I can compute that? All right. I've got a certain amount of practice. It's true. But none of these steps was hard, and it's worth your effort to, sort of, go through and say these things to yourself out loud. Say them out loud, all right? Because it works so nicely and so easily.

By the way, this is another example of – as a matter of fact, this may be the extreme example of the sort of dual relationship between concentrated in one domain and spread out in the other domain. All right. Delta is infinitely concentrated. All right. Sort of by definition, that's what delta is supposed to be. Delta is supposed to be the limit of functions that are concentrating at a point. Its Fourier transform, one, is uniformly spread out. Isn't that nice? That's the other thing you have to sort of get – say to yourself – every time you have one, I mean, I know it's hard because there are so many little bits and pieces like this. But that's one of the things you start to get used to when you work in this subject for awhile, when you use it, is you start to have these little checks, you know? These little interpretations that you carry with you from old situations to new situations. The old situation was in the stretch theorem, concentration in the time domain means spreading out in the frequency domain, and vice versa. All right. So here, concentration of the time domain is spread out in the frequency domain in the most uniform and the most extreme case. Let's do some more.

What is the Fourier transform of delta sub-a, the shifted delta function? All right. What about the shifted delta function, delta sub-a? All right. What is the Fourier transform? The Fourier transform of delta sub-a, I have to tell you how it operates on a test function. The Fourier transform of delta sub-a is, by definition, delta sub-a operating the Fourier transform of . But why does delta sub-a operating on anything, it evaluates the thing at a. That is, this is the Fourier transform of at a. All right. What is the Fourier transform of at a? It is the integral for minus infinity to infinity, d to the minus two pi i a x, of x, dx. And now, again you have to look at this with new eyes. And you have to say to yourself, 'This is the pairing of the complex exponential with .' This is the pairing of either the minus two pi i a x with . Where do we start? Where do we finish? We have the Fourier transform of delta sub-a paired with is equal to, either the minus two pi i a x paired with , and therefore, that holds for every test function that identifies the Fourier transform of delta sub-a as e to the minus two pi i a x. Where the right-hand side is understood as a distribution. Okay. It's the distribution which operates on a test function by this pairing. You may have seen this, you may have seen some tortured derivation of this, too, but this is the right derivation of it. And while we're at it, why don't we find the Fourier transform of the exponential – complex exponential? Hey, hey, hey. Let's find the Fourier transform of – where is it here – yes – let's find the Fourier transform of e to the two pi i a x. How do I do that? Where this is understood as a distribution. This would not exist in the classical case, right? I mean, you can't find the Fourier transform just by computing the integral, but we can by pairing.

The Fourier transform would be the two pi i a x paired with is equal to, by definition, either the two pi i a x paired with the Fourier transform of . That is the integral for minus infinity to infinity of – this is the function pairs with Fourier transform either the two pi i a x paired with the Fourier transform of x dx. Now, look at that and what do you see? What do you see? What do you see? You see the inverse Fourier – this is either the two pi i a x times the Fourier transform of , classical Fourier transform. You are computing there the inverse Fourier transform evaluated at a. This is of a because this is the classical inverse Fourier transform of the Fourier transform at a. So it is of a. But of a is the delta function at a paired with . If you use your new eyes, all right? So, where do we start?

Where do we finish? We started with that, we finished with that. The Fourier transform of either the two pi i a x paired with is the same thing as delta sub-a paired with . That identifies the Fourier transform at either two pi i a x as delta sub-a. Airtight. No question. Take the particular case – actually, when a is equal to zero. Take the case when a is equal to zero. If a is equal to zero, I have the Fourier transform of one. Either the two pi i zero – and what do I get? The Fourier transform of one is delta sub-zero. The Fourier transform of delta is one; the Fourier transform of one is delta. No doubt about it. No doubt about it.

And again, this is a nice illustration of spread out in the time domain, concentrated in the frequency domain – uniformly spread out in time infinitely concentrated in frequency. Delta infinitely concentrated, all right? It works. It works great. A couple of more, couple of more. How about – do it up here. Signs and cosigns. Signs and cosigns do not have classical Fourier transforms, but they have generalized Fourier transforms because signs and cosigns make sense as distributions. All right. For instance, cosign of two pi a x – I'll stick with the sort of scaling by a. All right. That's easy to do because that's one-half, e to the two pi i a x plus e to the minus two pi i a x. And I take the Fourier transform of each part – each piece. For the Fourier transform of cosign of two pi a x is one-half the Fourier transform of this, which is delta sub-a plus the Fourier transform of this, which is delta minus a, or plus a by minus a in this expression, so it's delta sub-minus a. Okay. So simple. So simple. Really.

There's a graphical way of representing that. The graphical picture usually goes like this. I write it by two spikes, one at a and one at minus a, here's zero. So this is usually – people usually denote the Fourier transform graphically by this. And that's fine, you know, that's fine, you can write it like that. The picture is that. How about the sign – the sign function is just as easy because the sign function can also be expressed in terms of complex exponentials, the imaginary part of the complex exponential. So the sign of two pi x – two pi a x is two pi a x is one over two i times e to the two pi i a x minus e to the minus two pi i a x. And so it's Fourier transform, which I guess I'll do over here, is the difference of two delta functions times this complex number, one over two i. The Fourier transform of sign of two pi a x is one over two i times delta sub-a minus delta sub-minus a. Nothing to it. Nothing to it. All right.

Once again, I said a little while ago that the problem with a classical Fourier transform is that it didn't make sense on the functions that you really want it to make sense on, like, the functions that society needs: trig functions, constant functions, and so on. All right. But now, if you open your mind and open your eyes, all those things work. All those things work and it's effortless. It's effortless. Now, it takes some effort before it becomes this effortless. Yeah?

Student: Do we have this more general framework; is that also how you define a Fourier transform of classical functions? That you could do before, like the rec functions?

Instructor (Brad Osgood) All right. So that's a good question. The question is what happens to all that we did before? Now, honestly, to be the most honest and rigorous about this, you should always consider tempered distribution. So that is to say, if you want to think the Fourier transform of the rec function – all right – there's two ways of doing it. You can consider it classically, that's fine. There's nothing wrong with that. Or you can say the rec function induces a distribution, right? And its Fourier transform is the sinc function also understood as a distribution. Now, that's going a little too far in some sense, in that, you know, when these functions are nice enough – when a function induces a distribution, you tend to blur the distinction between the function and the distribution that it induces. All right. But the most proper way of saying it is you only consider the Fourier transform for tempered distributions. That's where it's properly defined, and so classical functions where we computed the Fourier transform before, should be understood as distributions. And everything should be in that framework, all right? But then you have to say to yourself, 'Look, we're not kids here. All right. I think we have to take a reasonable approach to this.' And you don't want to abandon what you have done in the past. All right. So it's one of these things where you have to keep a little tension, a little cognitive dissonance in your head between the Fourier transform of tempered distributions and the Fourier transform of a function when it does make sense. All right. So – and I'm going to blur that line. All right. That is, I'm going to write the Fourier transform of the rec function is the sinc function without worrying about it, or without saying, really it has to be distributions – it has to be understood in the sense of distributions and so on.

So, for example, when you write, now, and matter of fact – and I say this in the book – you can finally justify completely that the Fourier transform of the sinc function is the rec function; or the inverse Fourier transform of the sinc function is the rec function. Same thing. All right. You can do dual; you can do all those things. The proper way of understanding this is in the sense of distributions because the Fourier transform of the sinc function exists – sinc induces a distribution; it's a fine function as far as integrating against a Schwartz function goes. All right. So the Fourier transform of the sinc function is the distribution, this has to be understood in the sense that equality of distributions, not in terms of, sort of, [inaudible] functions. But that gets a little extreme. All right. That gets a little extreme. You know? Even religion has its limits. All right. Now, I'll say one more thing. Like I said, I just wanted you to enjoy the ride a little bit. There's a little bit more of the ride to go. Next time I'm gonna talk about derivatives of distributions, how every distribution is differentiable. We're moving that flaw from the classical theory of calculus where not every function is differentiable. Every distribution is differentiable and we're gonna get some formulas out of that.

You know, why do we go through this? You know, again, you have probably seen these things derived in some tortured way in previous classes, all right? Well, like I say, things have moved on. This is the modern way of viewing Fourier transforms and how they work. And you should, as part of your education, want to know what the modern viewpoint is. You could build a radio out of vacuum tubes, but we don't teach that anymore. All right. We don't teach vacuum tube technology. All right. You should want to know, even if only to be – only to have sort of an acquaintance that is possible to do, you should want to know how the modern view of this is. And furthermore, as they say, it's not just that it's a good point of view, you can actually compute with it with confidence. I derived effortlessly these formulas. Like, the Fourier transform of one is delta; the Fourier transform of delta is one. The derivations of those in the classical case are really quite involve and not at all, I think to my mind, convincing. All right. They were – they get you the correct answer, but, you know, the cost of torturing things to do that is quite high. All right. I see people wanting to get in. We're gonna take a break. I know there was a question in the back; you can ask me when we're on the way out, okay? All right. See everybody on Wednesday.

[End of Audio]

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture14

Instructor (Brad Osgood):Okay. Let me circulate, starting over this side of the room this time, the sign-up sheet for the midterm exams. So remember, next week, next Wednesday, we have the midterm exam for the class at three sessions: 2:00 p.m. to 3:30 p.m., 4:00 p.m. to 5:30 p.m. and 6:00 p.m. to 7:30 p.m. I actually have secured rooms for that. I'll post the – I'll write them next time. I didn't bring it with me, neither the rooms nor the location of the rooms.

Anyway, I'll post it on the website and make the announcement next time. I'll say a little bit more detail about the exam. So when you're signing up there, it's just so we can have a sense of how many people are going to be in which slot. You're not signing your life away or anything, but I'm figuring that between one of those slots, from 2:00 p.m. to 3:30 p.m., from 4:00 p.m. to 5:30 p.m. and from 6:00 p.m. to 7:30 p.m., should be able to cover most everybody.

I heard a few – from a few people who have problems, but most everybody seems to be okay. So please, as that circulates, sign up, so we get an idea who's going to be where, when. Any questions about that or any other general administrative issues on anybody's mind? No?

Okay. All right, so, today, we have a few more miracles to uncover about distributions, but soon – and it's all interesting, and it's all useful, but soon we'll have to make our peace with generalities. There's a lot more detail and a lot more derivations that are given in detail in the notes, and so I will refer you to that for further reading. I'm not really gonna say too much more than what I say today, because we really do have to move on.

Again, we should not – you should not feel encumbered to derive everything. You should be, I think, satisfied, I hope, with the idea that – to look for the derivations, to try to understand some of them, and just get a general idea how the framework works, because I think it really is very satisfying.

In the past, when we've done this in class 261, many people who have seen these ideas before and worked with delta functions in various contexts in different classes have appreciated the opportunity to at least see what the more general context is and to see how the arguments work even if not – even if they don't understand all the details and haven't gone through all the derivations and details.

So really it's mostly – what we've been doing is to give you an idea of how the general framework works, and some degree of confidence that there is a firm foundation for a lot of these things, even without all the details. But you should feel free to use – as I said before, when you were first starting this, it's not that the stuff we've done before was wrong, it's not that the formulas that we used were incorrect or the applications really not well founded, so as we go forward, we'll call on those ideas and call on those formulas really without – say, without fear of recrimination.

But I think it's – I hope you found it satisfying, intellectually, certainly to see how some of these ideas play out, because it really is quite striking and it's really quite, I think, quite a remarkable accomplishment to get it all in such a beautiful form.

And then just a few more things that I wanna pick up today, but there's only so much that I think we're willing to subject each other to, all right?

So the first thing I wanna talk about, or maybe one of the final topics in the general lore of how distributions work, is the remarkable fact that, as general as they are, one other operation from calculus carries over to them, and that is the idea of a derivative. So it's possible to define the derivative of a distribution and, in fact, although I won't do it, higher order derivates. That is – derivatives turn – different – excuse me, distributions turn out to be infinitely differentiable in a natural sense.

So the derivative of a distribution. This actually turns out to be a very important operation on distributions, and one that's of widespread use. So how would we define – so if t is a given distribution, how to define its derivative, t prime?

All right, when – any time you ask yourself a question like that, if I want a carry over operation from functions to distributions, the question is how to do it. Remember, I have to tell you – it's always the case, you give me a test function, I have to tell you how t prime operates on that test function. That's always the case.

So I have to say – have to define what the pairing is. T prime paired with the test function phi. And it is always the case, or at least almost always the case, that the way you approach this question is to ask yourself, what would happen if t prime were an actual function and the pair were given by integration?

All right, as a guide to answering this question, that's what you say to yourself, and then what you hope is to see something general enough to suggest a general definition. So if t prime were given by a function, that is if t and t prime were given by a function, and the pairing is integration, we would have t prime paired with phi, I would write, say, is the integral for minus infinity to infinity of t prime of x, phi of x, d x. All right?

And you look at an expression like this and you say to yourself, that is just crying out for integration by parts, because you wanna somehow – well, one thing at a time. So t prime paired with phi, if t is given by a function then it would be given by – the pair would be given by integration integral from minus infinity to infinity t prime of x phi of x d x, and that is equal to – well if I integrate by parts, t of x, phi of x, evaluated between minus infinity to infinity, minus the integral for minus infinity to infinity of – I put – I take the derivate off of t and I put onto phi. Phi prime of x d x.

All right, now – that's just straight integration by parts. Now you use the properties of test functions, in whatever particular context you're working and, in the case of Schwartz functions or in the case of functions which are zero outside a fixed set, phi tends to zero at plus or minus infinity.

So this term is – and t you're assuming is regular enough so everything here makes sense, so that this term – the boundary terms are gone. Equals zero, because phi at plus or minus infinity is equal to zero. So what remains is just the second integral, so the integral – minus the integral from minus infinity to infinity of t of x, t prime of x, d x, which you should recognize is itself a pairing that is minus t paired with phi prime. All right?

So once again, we start off by saying, if t comes from a function then the pairing of t prime and phi is given by an integral, and that integral, in turn, is – can be written in terms of the pairing as minus t paired with phi prime.

So you say to yourself, okay, if that's how it turns out when t is given by a function, then I – then take that as the general definition, all right? That is, the right hand side, this side makes sense, even if the intermediate steps didn't make sense, all right? Because if phi is a test function then, for any decent class of test functions, phi prime will also be a test function and t can operate on that. All right?

So turn this into a definition. Into a definition. That is, you define t prime by the pairing t prime paired with phi is minus t paired with phi prime. All right? The left hand side is something new, the right hand side is something old. The left hand side is defining t prime. How do I define a distribution? I have to tell you how it operates on a test function.

T prime operating on a test function phi is minus t operating on the test function phi prime, period. The only thing that may look like a flaw – a blemish on this definition is the minus sign, that pesky minus sign out there, but that's the way it is. It comes in. You have to accept it.

All right now. For a clean definition, let me give you an example. A very nice example, something that you have probably seen before. I'll do it over here. Let's take a function that has no business having a derivative, so to speak, that is, the unit step function, a function that comes up all the time in applications. U of x is, say one the Heaviside – a unit step function also sometimes is called a Heaviside function. One for x – x derivative in zero, zero for x lesser or equal to zero or maybe sometimes people define it to be a half of zero, and again, there are religious issues here involved and I will get into it.

But you know what the graph looks like. It takes a jump at the origin. What is its derivative? U prime of x. Now that defines, actually, a perfectly good distribution. It is a function. It's not a continuous function, but it defines a distribution. Defines, determines, induces, whatever words you want to use, defines a distribution since the pairing of u with any rapidly decreasing function phi certainly makes sense.

The integral for minus infinity to infinity of u of x, phi of x. Phi x makes sense because u just – well, we'll pair that one further step. It's the integral from zero to infinity of phi of x, d x, and that integral makes sense if phi is a nice enough function. The integral exists.

All right, so again, while phi is not – while u itself is a not a particularly great function, it has a discontinuity – it has a jump discontinuity, it does define a distribution, and therefore it has a derivative, because all distributions have derivatives.

So u prime exists as a distribution. Now you have probably learned, in fact, I wouldn't doubt, you probably learned that u prime is a very well known distribution. It's the delta distribution. U prime is equal to delta. And you probably learned that because you probably said, well really now, u is equal – u is a constant on two pieces. It's sort of a piecewise constant function, it's equal to zero on the left of the origin, it's equal to one at the right of the origin, so u prime, if it had a derivative, would be identically zero here and identically zero here because the function is just a constant and it takes an infinite jump. The slope is infinite. When you go on in this direction, it goes up, and the delta function is zero, except at one point where it's infinite, and so u prime must be delta.

Now, of course, there's that thing about the integral of minus infinity to infinity if delta x is equal to one. I don't know exactly how to make sense of that, but that really can't be important. Will it? No, not necessarily. It's because u prime is equal to zero here and u prime is equal to zero here and u prime is infinite there, so it must be the delta function.

You probably said something like that, right? So many words. So many words to make that derivation, to make that justification. Why so many words? The definition is right there. Let's see how it works. Who needs words? Nothing to it. Nothing to it.

How about – let's do another example. Let's do another example. Let's do the Signum function. That's an arrow, that's implies. How about the Signum function? Signum of x is equal to, say, one when – again, the definitions are – may be varied. One when x is derivative in zero, zero, say, of x is equal to zero and minus one of x is less than zero. All right?

So the graph of that, again, it takes a jump or it takes a double jump at the origin. All right, the plot looks something like this. It's minus one down here, then it takes a jump up to the origin and then it goes out to be plus one. Not everybody defines it at the origin to be this way. It doesn't matter.

And you probably learned what the derivative of this is. You probably learned something like, the signum of x prime or signum prime is two delta. And why have you learned that? Because you say, well signum is constant over here and so its derivative is zero, and it's constant over here so its derivative is zero, so its derivative is zero everywhere except at the origin, where it takes a jump at the origin, but it takes sort of a double jump, you know? Because it jumps all the way from minus one to plus one, and that's a jump of two. So the derivative has to be really twice infinity or two delta. That's why.

And maybe the integral for minus infinity to infinity, this thing should be, I don't know, two, for some reason, because it's gotta work out that way, because somebody told me what this formula is and that's the way it is.

So many words. All right? So many words to justify that formula. We don't need those words. Although I can’t quite bring myself to the – this is a tribute to Marcel Marceau, all right? I can't do that either. All right, how do you do signum prime paired with phi? By definition, it's minus the signum function paired with phi prime, all right? So that is – that's a pairing done by integration, so that is minus the integral for minus infinity to infinity of signum of x times phi prime of x d x. Now the signum function is either plus one or minus one, it doesn't matter what happens at the origin because integration – nothing matters if you're just changing the value of the point. That is, if you change the value to a point it doesn't affect the integral.

So this is minus – the integral for minus infinity to zero, signum of x is minus one so it's minus one u prime of x d x plus the integral from zero to infinity, where the signum is plus one, so that's plus the integral from zero to infinity of plus one q prime of x d x. All right?

So if I carry out those integral – integral of phi prime is phi evaluated between minus infinity and zero, so it's minus the whole, so integral phi prime of x is phi of zero minus one. So it's minus one times this. So it's minus phi of zero minus phi of minus infinity plus phi of infinity, it doesn't matter, minus infinity, right? Plus – where am I here – plus phi prime phi of infinity minus phi of zero.

But again, phi of infinity is zero, phi of minus infinity is equal to zero, it's minus a minus phi of zero. Minus phi of zero. It is two phi of zero, if you sort out all the minus signs. But two phi of zero is just twice the delta function paired with phi. It's two delta paired with phi.

So where do we start, where do we finish? We started with u prime paired with phi – or, excuse me, signum prime paired with phi is two delta paired with phi. What is the conclusion? The conclusion is that signum prime is equal to two delta. Isn't that nice? No muss, no fuss. Airtight. Airtight.

Now these are used – these formulas are used, actually, a fair amount. Let me give you some applications to Fourier transforms. To find you some Fourier transforms that it would be very difficult to find otherwise. There are ways. There are always ways. There are arguments, there are limiting cases, all sorts of stuff like that, but it can also be done this way with very little muss, very little fuss.

So let's find the Fourier transform of the signum function and the Fourier transform of the unit step function. For that, I need the derivative theorem, actually, for distributions and Fourier transforms, and I'm gonna state that, but not derive it. This is one of those cases where the formula looks the same as it does in the classical case. The derivation is a little bit more involved, and for that I'm gonna have to refer you to the notes.

So the derivative theorem for – derivative theorem for distributions, for Fourier transforms of distributions, says this. It says the Fourier transform of t prime is equal to – I wanna make sure I have my – I wanna get my minus signs right here – is equal to two pi i s times the Fourier transform of t. Turns multiplication into – turns differentiation into multiplication.

Somebody wanna check, make sure I got the – there not be a minus sign in there? I'm a little worried, there should be a minus sign in there, but I'll check that out. Let me put that formula up there and if I have to correct it later on, I'll correct it. And the other formula is the Fourier transform of t prime is equal to the Fourier transform of minus two pi i t times t.

All right now, again, t is a distribution, t prime is a distribution, so t prime has a Fourier transform. This says – this tells you how to find the Fourier transform of t prime, and how do you do this? Well the Fourier transform of t prime has to act on a test function. You use the definition of the Fourier transform, you use the definition of the derivative, and out pops this formula, all right?

It's not hard. It takes a little bit of work, but it's not terribly hard, and it's the same for – the reason I'm not gonna derive it is because it's the same formula that we have in the classical case. The classical case was a Fourier derivative – Fourier transform of the derivative was two pi i s times the Fourier transform of the original function, and we had this formula also, okay?

Now how do you use it? As an application, we can find the Fourier transform of the signum function. So use this to find the Fourier transform of the signum function. Why? How? Well signum prime is two delta, that's what I'm just erasing, signum prime is two delta. So, on the one hand, the Fourier transform of the derivative of the signum function is the Fourier transform of two delta, which is two. Two times the Fourier transform of delta. The Fourier tansform of delta is one, all right?

On the other hand – yeah, I think there's no minus sign in there. It's okay. On the other hand, the Fourier transform of signum prime is two pi i s times the Fourier transform of signum, okay? So two pi i – so put these together. Two pi i s times the Fourier transform of signum is equal to the Fourier transform of signum prime which is equal to two, and thus the Fourier transform of the signum function is given quite nicely – is one over – I'll put an s in here although, again, things should not let a point – I'll say a few more words about this. Fourier transform of signum is two over two pi i s. That's one over pi i s. All right?

That's the correct formula. The Fourier transform of the signum function is one over pi i s, derived very quickly. Now, in fact, actually, there are a number of extra things you have to say here because there are several operations that I've done here on Fourier transforms that I haven't completely defined, and I'm not gonna do it, so, again, this is one of those cases where, for more details, I'm gonna have to refer you to the notes.

The formula is correct and the derivation is correct. Justifying – there are several steps in the derivation that actually have to be justified, so you need more – you need more of an argument, all right? And all I should say is see the notes, all right?

For one thing, this is a singularity. The function one over pi i s or one over s has a singularity. Does that really define a distribution? Can you really pair that with a smooth function by integration and so on? And that actually requires a special argument, a special definition for the pairing, called the so-called principal value distribution, so I'm not gonna talk about that but that's discussed in more detail in the notes.

The formula is correct and the derivation is, as I say, is also correct, when the proper details are supplied, none of which is hard but there are little – there are some subtleties that are involved, so I'm not gonna go through that. But, nevertheless, that's – this is a formula that you often see, and you see this formula classically. I mean this formula was derived, in some way, by some limiting process, but it follows directly once all the machinery of distributions and Fourier transforms and derivatives are in place.

How about the unit step function? There are several ways of getting that also. If you get the Fourier transform of the unit step function – remember, the unit step looks like this. It takes a jump up from zero up to one. That's u of x. The easiest way of doing that is to express that directly in terms of the signum function.

That is, u of x is one half one plus the signum of x. When x is negative, the signum of x is minus one, so one plus minus one is zero so use zero to the left of the origin. Nevermind what happens at the origin. Who cares what happens at the origin? Although this would assign it the value of one half of the origin which, again, is sort of a common convention, and when signum is – when x is positive, signum of x is one so one plus one, one half of that is one. So it takes a – it takes a jump of plus one, all right? That's the quickest way of – I mean that's an easy – a relationship between the two functions, and now that also allows us easily to find the Fourier transform, because the Fourier transform of u is then one half the Fourier transform of one plus the signum of x, and the Fourier transform of one is delta. The Fourier transform of signum we just found.

So this is one half delta plus one over pi i s, and that's it. Pretty simple, okay? That's also a very common occurring formula. These things come up a lot. They come up in terms of filters. As a matter of a fact, if you look – well we'll talk a little bit more about that later, but if you look back at the section – the chapter on convolution, we talked about highpass filters and notch filters and things like that. Delta functions come into that, and if you want to know about the transfer function or the impulse response to those, the Fourier transforms of these things come in. So these – just these expressions are actually in quite common use, and they're very – say, very – they fall out quite easily from this general framework, okay?

It's very nice. It's really nice. All right, the last – about the last thing I wanna do by way of just making you aware of the general properties is talk a little bit about multiplication and convolution in the context of distributions, and here, again, there actually are a number of subtleties which I am not gonna do in public, but refer you to the notes.

Many operations that you can apply to functions have analogues and carry over to distributions, but not all, interestingly, and maybe most interestingly, the one operation that really doesn't carry over to convolutions is multiplication. You can multiply two functions together, that's no problem, but you can't multiply two distributions, all right?

So interestingly, the multiplication of functions – it does in some cases, but not in general. Multiplication of functions does not carry over to multiplication of distributions. So this is the one caveat that I have to issue, and this is where sometimes people can make mistakes if they're a little too cavalier in thinking that everything is gonna work out just the way it should work out.

What I mean by this is, if s and t are given distributions, if s and t are distributions, then the product is generally not defined. It's okay in some cases, but generally not. Is generally not defined. And for reasons which we'll see, actually, this also has to do with the fact that convolution is a little bit more complicated also for distributions that it is for functions.

Now there is a special case, so this is just a warning and I'm not gonna explain why – where the problems are, although, again, it's discussed in a little bit more detail in the notes. There is one case where it is defined, and that's when, say, the distribution comes from a function, and that's one way of thinking about it, but that's really the proper way – I take that back. The case where it is defined is when you multiply a function times a distribution, or rather what I should say is, what – the operation that is defined is multiplying a distribution times a function, all right?

What is defined, in most cases, is f times t where f is a function, and actually this turns out to be an important operation and I'll give you a special case of it in just a second, which is extremely important, all right?

Now how is it gonna be defined? How? Well, once again, if you ask yourself, how am I gonna define a distribution, the first thing you should say is, what would it be, how it would it work if t were actually given by a function itself and the pairing by integration?

So I have to define, as always, what I mean by f t paired with phi or f t operating on a test function. Now this actually is gonna turn out to be quite simple, and actually reminiscent of some of the formulas that we had. So, again, if t is given by a function then you'd write the pairing f t paired with phi is the integral from minus infinity to infinity t of x, f of x times t of x, phi of x, d x, and I just grouped the f with the phi. That is, the integral from minus infinity to infinity of t of x times f of x times phi of x which is the same thing as t paired with f times phi. The f just moves over, all right?

Now if that's what happens if the distribution comes from a function, then you say to yourself, so that was – that must be the definition in general. That gives me a clue as to how to define it in general. But again, there's actually a little caveat here. So in general, you define f t by the formula, how does f t operate on the test function phi? By definition it's t operating on f times phi, okay?

Now here's the caveat. The caveat is this has – this works so long as f is such that f times phi is again a test function. So if phi is differentiable, then f better be differentiable. If phi is rapidly decreasing, then f better be at least such that f times phi is rapidly decreasing if you want that class. So that's the one caveat here is that you may not be able to multiply by arbitrary functions, because this expression may not make sense.

The expression on the right hand – on this side here, makes sense only if f phi is something to which – on which t can operate, so it has to be a test function, so f times phi has to have the properties that define the class. So this makes sense.

This makes sense only if f times phi is, again, a test function. So, again, that's just a caveat. It's not going to be an issue for us, but it's something you have to – one of these little flags you have to put up when you're applying some of these ideas, all right?

Now, we actually implicitly used this. We implicitly used the operation of multiplying and distribution times a function when I wrote down the derivative formulas, all right? So we used this, and it's used in the derivation when we wrote that the Fourier transform of t prime is two pi i s times the Fourier transform of t.

When I say we used this, what I mean is the right hand side makes sense because it makes sense to multiply a distribution times a function. The function, in this case, is two pi i s. The distribution is the Fourier transform of t. So that is – the expression itself makes sense, and I didn't say this at the time, but I knew this was coming, because I didn't wanna make a big deal out of it at the time, but the fact is that you – the first, if you're giving the sort of proper logical sequence of developments here, the way it's done in the notes, the first thing you have to define is this operation, operating – multiplying a function times a distribution, and then you can talk about the derivative theorem and a lot of other things, because this expression then makes sense. Okay?

And likewise, actually, the second derivative formula also made sense, provided you give that definition, because there we said the Fourier transform of t prime is the Fourier transform of minus two pi i t times t, all right? That was the second derivative formula, and that also makes sense because this expression makes sense. It makes sense to multiply t times the function minus two pi i t. All right?

Once you have that defined, then you can talk about its Fourier transform and so on. All right, now I'm actually less interested in this general property of defining a function times a Fourier transform – times a distribution than what happens in the special case of the delta function, because that's particularly interesting and particularly important for applications.

So I'm gonna give a special case of this, is multiplying a delta function times a function. It's f of x f times delta, all right? What is f times delta? Well f times delta paired with phi, by definition is – f times delta paired with phi, by definition, is delta paired with f times phi. That's the definition of how a function times a distribution pairs with a test function, but delta paired with f times phi is, by definition of a delta function of a delta distribution, this is, by definition, f of zero times phi of zero. It's the product f times phi evaluated at zero.

And now, again, you have to realize – you have to look at this and you have to reverse what you said. You have to realize that this is the same as f of zero times delta paired with phi. F of zero times delta is just a number times delta, so that makes sense. There's no special definitions required there. F of zero times phi of zero is the same thing as the pairing of f of zero times delta paired with phi. Where do we start, where do we finish?

We started with f delta paired with phi is the same thing as f of zero times delta paired with phi. What is the conclusion? The conclusion is that f times delta is f of zero times delta, and a little bit more generally – and I will – a little more generally, f times delta a, the shifted delta function, is, as you might imagine, f of a times delta sub a, okay?

It pulls out the value, in the case when you multiply by the ordinary delta function at concentrated zero, it pulled out the value at zero. If you multiply a function times the delta function concentrated at a – see, I use that terminology. I mean there's – there's nothing wrong with it. Concentrated here, concentrated there. F times delta of a is f of a times delta a. This is called the sampling property of the delta function, or the sampling property of delta, and it's very important. We're gonna make a lot of use of this.

This is the sampling property – sampling property of delta, and you've probably seen this too, all right? You probably saw this in the context of concentration, actually. You probably saw this in the context of a bunch of functions shrinking down, concentrating at a point and multiplying by a function, what happens and so on, but it's very easy and very directly – can be derived very directly from the definitions that we have, all right? So we'll make a lot of use of this.

As a matter of a fact, I think for us to say to sample – again, and you're also probably familiar with the idea of sampling, a topic that we're gonna take up next which is actually my favorite topic in the course. For us to sample means to multiply by a delta function, all right? And to use this property. That's what it means to take samples. That's the mathematical meaning of taking samples is multiplying by deltas.

All right, finally – where are we here? Is convolution. The other big operation that we've talked about that is so naturally related to Fourier transforms and, again, here there is a special caveat. Here it doesn't quite carry over quite as nicely as one might hope, or at least not in complete generality. So again, if s and t are distributions, how to define their convolution, s convolved with t, all right?

And the sad fact is it's not always defined. There are restrictions. Okay so it's not always defined. That is to say, to give a definition it's necessary to know that it – to state what it should be and to guarantee that the convolution exists, it's necessary to impose some extra conditions on the distributions, and I'm not gonna do that because it's a little bit technical, it's a little bit complicated and it's not really so – quite so crucial for us, all right?

You need extra restrictions – you need some restrictions on s and t, and, again, the definition is given in the book. You can do the definition in terms of a pairing. When everything is defined, you can approach the problem the same way you approached all these problems. If I'm gonna define it, how shall I define it? Well if it comes from a function, what would the definition be if everything here came from a function?

You write down the integral, you do a little bit of manipulating with the integral and a definition emerges, but as you see – in the course of that discussion, you see that it doesn't always work without some extra assumptions. You can do it. You can define s convolved with t via a pairing, but you need extra conditions, as I said. Extra conditions.

All right, now the good news is that there are many cases when it works without further comment, and, again, I'm not gonna make a reproduction onto this. So it's many cases when it's okay, when all is well, and one of the most important examples is when you convolve – well I'd say two distributions when one of the distributions comes from a function or just – or a little bit – or to say it a little bit differently, when you convolve a function with a distribution, that makes sense.

So e g f convolved with s, often, or f convolved with t, often makes sense – most often makes sense when f is a function, all right? I'm sorry for being so – a little bit vague about this, but the fact is that if I went – if I – I'm perfectly capable of actually giving you the detailed definition here, but it requires a little bit of extra setup and it's really not worth it, but realize, when I say f convolved with t makes sense, not as an integral, all right?

F can – everything here is – things are here – things are defined here more generally, so you can't define f convolved with t as a simple integral of f of x times t of x minus y d u x or whatever it is. You have to define it in terms of a pairing, and setting that up actually requires extra work, all right? So that's what I'm not telling you. All I'm saying is that there is an operation on – called convolution that mimics the classical operation of convolution even though the definition has to be given more generally, and that it doesn't make sense for two arbitrary distributions, it doesn't even make sense for an arbitrary function and a distribution, but it makes sense often enough for a function and a distribution that you can work with it, and, furthermore, the convolution theorem holds. Okay? And the convolution theorem holds.

That is to say, the Fourier transform of the convolution f convolved with t is the Fourier transform f times the Fourier transform of t, okay? Now, again, see the problem – and this is actually – this is related to the problem we had – I mentioned before, about the problem of – about defining multiplication. You want to the convolution theorem to hold – if you want the convolution theorem to hold for our distributions, then you'd want to be to multiply distributions, but you can't always multiply distributions, all right?

So that's – the problem here is the same. The problem with defining convolution is the same for two arbitrary distributions, is the same as the problem of defining multiplication for two arbitrary distributions. It just doesn't quite work, all right? Because you wanna have this formula, and this formula should, by all rights and by all sort of formal derivations, work, but it doesn't always work because the definition of convolution as a pairing doesn't always work and the product of two distributions doesn't always work, all right?

But it does work, most often, in these cases, because everything here is defined. As it turns out, the left hand side is defined, f convolved with t. Also, the right hand side is defined because it's a function times a distribution, and a function times a distribution makes sense. The product of two distributions may not make sense, but the product of a function times a distribution does make sense, all right?

So there's no – the most I can say here is there's no inconsistency. We haven't discovered that long awaited for contradiction in all of mathematics, and the world is not gonna crumble, all right? So everything here is consistent and everything here makes sense, and what I'm not telling you is the details about when it's true and when you can be – when you can apply it. So suffice it to say, for us it's not gonna be an issue and I will never do anything false – knowingly false, at least.

The same formulas that we used before, the same ideas, work again. In particular, the convolution theorem works. Now there is a special case of this that's most important for us, and that's when, again, you're convolving with a delta function. Again, I apologize for not giving more details here, but it just – my feeling is that there's only so much you can take, and ultimately it's not gonna do us – not gonna be so helpful to us.

We'll be able to apply the formulas, we'll be able to apply the reasoning, without really worrying about it so much. So a special case, special case, when t is equal to delta, is particularly important, and what you find is that if you convolve a function with a delta function, you get the function back. That's an extremely important formula.

But the delta function serves, in some sense, as the identity for convolution. If you think of convolution as a kind of multiplication, then delta serves as the identity element for convolution in the sense that if you convolve a function with delta, nothing happens.

Now this is not hard to derive, actually, once all the terms are properly defined. That is, and you probably saw – you probably said a lot of words, at some point in your life, or somebody said a lot of words to you, to give this argument, to give this – to justify this formula, but, in fact, you can give, as I say, sort of a wordless derivation that follows quite easily from the definitions provided you give all the definitions first, and that's what I haven't done. But one – but the – it's completely routine to show that this property holds once you have set up the mechanism for it, once you've set up the superstructure for it, and as a slight extension of this, more generally, if I convolve with a shifted delta function, I get back a shifted version of the function.

So let me write it like this although, again, I shouldn't be writing things at points, to be strictly correct, I think. Nobody's gonna strike me dead if I do this. If I convolve f with a shifted delta function, I get a shifted version of f. All right?

These are both very important – well this property is just a generalization of this property. The sampling property of the delta function, the convolution property of the delta function, are extremely important and we're gonna make constant use of them, constant use of them. So if all of this work on distributions went toward just getting those two identities, it would be worth it, somehow, because to have those at our disposal is – we'll find just constant applications of that, okay?

I'll give you one nice sort of generalization. One case where convolving two convolutions does make sense is this. It's not a special case of those formulas but it is a case where the convolution of two distributions makes sense, and that is you can convolve delta with itself. Delta with itself. And let me state the generally formula, that is – it's quite attractive. Delta a – a shifted delta function, delta function concentrated a, convolved with a delta function concentrated in b, is the delta function concentrated at a plus b.

So, again, I'm not gonna prove that. The derivation of that is given in the notes, all right? But, again, that's the sort of thing that comes up, actually, often enough that it's worth knowing. Take the delta function at a, convolved with the delta function of b is the same thing as the delta function a plus b. Makes sense, in some sense, in terms of – or at least it's consistent with the formula – with this formula, because if I shift it by a and then shift it by b, that's the same thing as shifting by a plus b.

So note f convolved with delta a convolved with delta b is like f of x minus a convolved with delta b is like f of x minus a minus b. I didn’t put equal signs in there because of where are the xs and so on, but you get the derivation. You get the sense. And that's the same thing as f convolved with delta a plus b is then f of x minus a plus b, that is to say, f of x minus a minus b, so at least it's consistent.

That's one thing, I think, that you should – again, to sort of – as you build up a set of internal checks of your understanding of the material, even if you don't know the derivations it's often a good idea to be able to sort of cross check it in cases where you can verify the formula makes sense, all right? So this is an example of – although it's not a derivation of the formula. It's sort of a consequence of the formula, and it gives you some indication that everything is consistent here.

Why should delta a convolve with delta sub b be delta sub a plus b? Well at least it makes sense if I consider that convolving a function with a shifted delta function shifts the function. Again, it's sort of an internal check of consistency – cross check of consistency, and it's a nice formula.

All right. We have one more thing today. One more property of the delta function, and then next time we're gonna use it, all right? Next time we're gonna use the delta function and some of the properties that we've derived for – properties of distributions and a study of diffraction phenomena in optics. It works – you see how the Fourier transform comes into that in really quite a nice, striking way.

But let me do one more property today, and here I'm gonna be just absolutely shameless in my derivation, and I don't know how to fit this in other than just to do it, because we're gonna need this formula, all right? And that is the so-called scaling property of the delta function. Of delta. And that is – you wanna consider what is – let me put it this way, delta of a times x. Not delta shifted to a, but if I scale the independent variable.

Now the problem with writing something like that down, and I almost gag on it when I write it down, is if I've been making all this point about delta doesn't define – delta is not defined at points. You can't look at delta of x, delta of a times x, delta of anything like that. Delta is an operation on functions, so at first blush, if I write something like this down, I have violated all of my precepts. I feel cheap and dirty. Love it! Right.

Now, in fact you can define this because it makes sense to define a scaling operator on distributions. So I'm not gonna do that again and that's done in more detail in the notes. So it is defined by defining the scaling operation, the scaling operator, on distributions, and that's not so hard. That can be done. So it makes sense, actually, in a more general context, to consider delta of a times x.

But now, if you think about what you – how you used to think about the delta function, I mean, delta is already concentrated at the origin. If you multiply it by a does that – I mean, can it be any more concentrated or what could that possibly mean?

Well, again, I'm gonna be shameless here in thinking about how I should – when I say what is delta of a x, what I want is a formula for delta of a x in terms of delta, actually. So I'm looking – I wanna look at delta of a x paired with the function phi of x, and I'm gonna write that down in terms of integration although, again, it's against all my principles, but I'm gonna do it anyway. So I'm gonna write this the way you used to write this. Delta of a x times phi of x d x, and I consider this as pulling out the value at the origin, all right?

Now this can all be justified, even these steps can be justified, without writing integrals in terms of the scaling operation, but just follow along with me here. All right, this is how you used to derive this. You used to say that – I'll make a change of variable. U is equal to a times x, and let me assume that a is greater than zero here so I don't have any trouble switching to limitless integration.

If I let u equals a times x, then d is equal to a times d x and the integral becomes, if x goes from minus infinity to infinity, then if – the integral in terms of u, u also goes from minus infinity to infinity if a is positive, and this becomes delta of u times phi of x is u over a d u, and d u is d x over – is – d x is one over a times d u, so it's one over a times d u. Sorry.

The one over a comes out of the integral, so this is one over a times the integral for minus infinity to infinity of delta of u phi of u over a d u, and now that's the ordinary, so to speak, property of the delta function. If I paired delta – if I want it like this, one over a delta paired with phi of u over a, that still pulls out the value at the origin. Phi is scaled but delta doesn't know that. Delta just pulls out the value at the origin, so this is one over a phi of zero. Again, delta doesn't care. Delta never cares, but phi is scaled to the origin. It just pulls out the value of the origin. So one of a of phi, phi is zero, so that is one over a delta paired with phi. F of a is positive, so we get – where do we start, where do we finish?

Delta of a x paired with phi of x was one over a times delta paired with phi, so the conclusion is that delta of a x is equal to one over a times delta of x, if I write the variables here and if I feel so shameless about it. F of a is positive. If a is negative, you get a very similar result, and let me just write down the final version. You get a similar argument if a is negative, and you get the scaling formula for the delta function, and then we gotta go.

That is delta of a x is equal to one over absolute value of a delta of x. So this is not the scaling theorem like in the Fourier transform because the variable over here isn't also scaled. It's only scaled out front, all right? It's only scaled out front. And, again, I'm writing this – I'm breaking the rules in the way I'm writing this, but all I'm saying is that it can be justified if you actually look at the scaling operation, apply the distributions, and then the derivation is really pretty much as we gave it, all right?

This is the cheap and dirty way of doing it. It's okay in this – in that it led us to this formula and we're gonna use – we're gonna make actually quite a bit of use out of that formula, all right? As a matter of a fact, you'll start seeing this as of, already, next time, okay? So we're gonna leave the happy world of distributions now, and we're gonna start seeing how they're applied, all right? See you then.

[End of Audio]

Duration: 55 minutes

TheFourierTransformAndItsApplications-Lecture15

Instructor (Brad Osgood):They took me by surprise again. What do you know, I think we're on the air. Where is everybody today? I mean, it's Friday, I understand that, but like no one's out there. Not that – excuse me, I didn't mean to say no one's out there, but it looks a little thin, the population.

Okay, while people are drifting in, let me call your attention to the important information up on the board. This is information for the exam for next week.

So the midterm exam is October 31st, Halloween, sorry, and there are three sessions, from 2:00 p.m. to 3:30 p.m., from 4:00 p.m. to 5:30 p.m., and from 6:00 p.m. to 7:30 p.m., and they are – we have the locations for those now, so from 2:00 p.m. to 3:30 p.m., that is in building 380, room 380 W. Building 380 is the Math Corner. That's the front right hand corner of the quad if you are looking at the quad from the oval, all right? So it's in the front of the quad. 380 W is in the basement. That seats about 50 people, something like that. Shouldn't be a problem. So both the afternoon sessions, from 2:00 p.m. to 3:30 p.m. and 4:00 p.m. to 5:30 p.m. are there, and then the session from 6:00 p.m. to 7:30 p.m. is here, in Skilling.

As far as the material goes for the exam, it goes up through deltas. You should know – have the properties of delta functions and generalized Fourier transforms. I'm not gonna go hot and heavy on the theory of distributions or anything like that, but you should be comfortable and familiar with the basic properties of delta functions, as I say, and the idea of the generalized Fourier transform, at least to the extent that you can use it for some common functions like the step function and the unit step function and the signum function, things like that. Those sorts of things are fair game.

What I won't put on the exam is the next topic that we're gonna really turn to, not today but shortly, on sampling theory. So that'll be coming up, so we won't have that on the exam.

I will provide for you – and again, as I said last time, the hope and the plan for the exam is to have it more conceptually based. That is to say, you can't avoid computations completely, naturally, but what I don't want is I don't wanna – I'm gonna try not to write questions that involve a lot of details, calculations, in the sense that you get caught in a loop of doing lots of integrations by parts or something like that.

That's not the point. The point is not to see how well you can integrate. The point is to see how well you can understand the Fourier transform.

Now again, having said that, it's also true that you can't avoid computation completely, so I wanna try to make a balance. We will provide for you, and it is already posted, has been for a while, the formula sheet. That's a formula sheet for the entire course, so we'll – I'll make copies of that and bring it to the exam. That has all sorts of helpful, useful formulas on it.

You can bring your – it's open notes, open book on this, so you can bring that with you. As I say, things get a little – things get sometimes ridiculous, where students used to bring stacks of like signals and systems books to the exam, and it was just ridiculous, but if you want to, you can do that I suppose, like Shawn's Outline Series of Signals and Systems, but hey, whatever makes you happy.

So the other thing I was thinking of doing – so that's question about – that's the information about the exam. Any questions on that, on what's expected or what – how we're gonna manage it? I haven't heard from anybody. There's one person who's gonna be away on a – at a conference, and a couple other people who have some conflicts, but, by and large, having those different times seems to suit everybody all right.

We will have class on Wednesday, however. It's relentless. Relentless. Relentless, but not heartless. I'm thinking of – what do you think of having the next problem set due – there's a problem set that's out there now, and I was thinking of having that due on Friday instead of Wednesday. Yeah? Can you endorse that idea? All right. Okay. So I'll post all this information on the web, of course, but – so we'll have the current problem set, what is that? Problem set five or four?

Student: Five.

Instructor (Brad Osgood):Five, okay. So we'll have problem set five due on Friday, next Friday, instead of due next Wednesday. All right? Yeah?

Student: Does that cover –

Instructor (Brad Osgood):That has some midterm material on it, yeah, it had – because it had some stuff on deltas and had some stuff on generalized Fourier transforms, all right? Sorry, I just can't – I can't match it exactly, but I don't think there's anything on there that you couldn't sort of understand – I mean, there are enough worked examples and things like that in the notes, so I don't – I really don't think it should be a problem.

I will have the next problem set out, also, by the – I mean there will be another problem set that will be due and I'll have it posted over the weekend or by Monday, all right?

Okay, so I'll – again, I'll put all this information up on the web to make it permanent, and send everybody an email, all right? Anything else on anybody's mind?

All right, we are going to do, today, an interesting application of the Fourier transform. It will involve a little bit of deltas, and some properties of delta, and actually it will involve a physical interpretation of delta, which I think is actually quite interesting, but it's – regardless of that, regardless of the particular ways that delta enters it, but it's interesting itself, because it's a nice application of the Fourier transform. It's the kind of thing you'll see in other classes, depending on the direction you go in your further work.

That is, I wanna talk about the Fourier transform and diffraction. Now diffraction refers to a phenomenon of light, and diffraction refers – for our purposes, it refers – it is analyzed and understood in terms of the wave theory of light, so we're not gonna get into the raging battle between the wavists and the particalists. It's – diffraction for us has to do with interfere – is – I think, pretty much, I consider like identical to interference phenomenon associated with light.

So diffraction is one term for it, but it's pretty much equal to interference patterns that light makes passing through an aperture of light through apertures. This is not the only way that diffraction can be talked about, but it's the way we're gonna talk about it. Through apertures, through holes. All right?

So there are a couple things that enter into this. I'll show you a picture in just a second. You have seen pictures of diffraction patterns, I'm sure, before, but I will show you a picture just so we all know what we're talking about, but I wanna talk about the distinctions, what we're gonna do and what we're not gonna do.

So it involves light from a distant source, typically from a distant source, and we'll see how that enters into it, the fact that it's distant enters it – how the fact that it's distant enters into it. All right?

It impinges on a plane, which is usually called the source plane. I'll look for other reasons, I'll – again, I'll come back to that in a second, where – on which a number of holes are cut, and instead of calling them holes it sounds more scientific if you call them apertures.

So you have a plane with apertures that the light passes through. The rest of the plane is opaque – part of the plane is opaque and then you have holes that light passes through, and it bends around those holes, and it creates certain characteristic patterns on a plane that is some distance from this, called the image plane. So at some distance, you have an image plane, all right? And you see diffraction patterns.

Now I'll show you a picture in just a second, but right here is where I wanna make a distinction between the different types of diffraction that one studies. We're gonna make a number of simplifying assumptions.

We're gonna assume – well, first of all, we're gonna do everything in the context of the wave theory of light so, for us, a light is gonna be an oscillating, electromagnetic wave, and really we're only gonna talk about the electric field part of it, so I am making a certain number of assumptions, but are accurate enough to give you good and helpful understanding – a good and helpful understand, and good and helpful formulas.

So we'll work with – we'll assume light is an oscillating EM field, although we're really only gonna talk about the electric field, and we also assume that it's monochromatic, that there's only one frequency of light that shining, that's being diffracted. That'll simplify writing things as well. And we'll make a couple other assumptions.

The other main assumption that you make in diffraction is – has to do with this statement here. At some distance, you have an image plane upon which the light is incident, all right? There the distinction is between so called near-field and far-field diffraction. So the distance of the image plane determines two kinds of diffraction.

One is called far – one is called near-field, which means relative to the wave – and it's usually relative to the wavelength, let me just say in a second. Near-field, which means it's close in, that's called Frenel diffraction. Frenel. I think that's right. Diffraction.

And the other is far-field, so called far-field, which is Fraunhoffer diffraction. I can only hope I'm spelling those right.

And, usually, when you talk about near-field or far-field, you talk about distances measured relative to the wavelength. It is far away relative to the wavelength of the light or it is close relative to the wavelength of the light, and the other that comes in are the size of the apertures relative to the wavelength.

So you measure distances near or far, typically relative to the wavelength. Distances and the size of the aperture relative to the wavelength. I'll make comments about this. This is – we are not giving, by any means, an exhaustive treatment, all right? This is just an example of how the Fourier transform comes in in an important and interesting application.

We have a whole course, actually, called Fourier Optics, that does these things in, of course, much greater detail, and goes much farther with it, but I thought you'd be – I thought you'd be interested to see this application, because it's a very pretty application and it's nice to see some of the ideas come in, and you'll see them again, as I say. If you go in – Stanford has been a leading light in light for years and, in particular, Joe Goodman, who's an emeritus professor now, wrote a – was a very important figure in the theory of Fourier Optics, and wrote a number of very influential textbooks as well, and of course a lot of research papers, and a lot of the – the course that sort of exists now in Fourier Optics is based, to a large extent, on a lot of Joe's work.

Let me show you a picture so you know what I'm talking about. You have seen – you have seen diffraction patterns, probably, in physics classes or whatever, but the picture looks something like this, if I can. Can I show this? I'm just gonna show this one shot. I think there's a picture of this also in the notes. Looks like that. Amazing, isn't it? There we go. You can dim this a little bit, so I can – so I actually get the full effect. Can we dim the lights? And we – there we go. All right. It's just showing up a little bit. So this is – what you're seeing is the image plane. You're seeing the pattern of the image plane. The source is some distance away. It shines through, in this case, I think it's just a square aperture, a square slit, that's showing the pattern, and the things that's – was remarkable about diffraction patterns that really shocked people, was the bands of light and dark. That's the characteristic property.

What you're really seeing there is the intensity. You're seeing the magnitude of the electric field that is representing the light, all right? Actually, you're – yeah. And that's what instruments measure. They measure the intensity, or the magnitude, of the electric field. And what was so surprising to people, when they first started observing these diffraction patterns, and it goes way back into the history of physics, is the intermixing of light and dark bands, because how can light plus light make dark?

I mean that – why should – how could there be darkness, when all you're doing is shining light? How could that be? Surely this is not God's will. And it was – it was a subject of heated debates, and again – certainly between people who have advocated the particular theory of light, people advocated the wave theory of light.

The wave theory of light initially provided a much better explanation of these diffraction patterns, and that was one of the reasons why it was in ascendancy for so long, because this was a phenomenon that was considered – was not only considered important, was important to understand it that it could be analyzed by using the wave theory of light.

So you have – anybody, I'm just curious, are there – anybody ever work with these sort of diffraction patterns? There are all sorts of reasons to do it. I mean they use it for physical measurement of wavelength of light and things like that, or they – and they sometimes just have experiments. Yeah, how did you –

Student:[Inaudible].

Instructor (Brad Osgood):Yeah.

Student: The dark parts were to calculate how thick the wafer was, to calculate how hot it was, if we knew its initial temperature and things.

Instructor (Brad Osgood):Yeah? That's interesting. Anybody else? I'm just kind of curious. Even in Physics labs or something like that, when you were undergraduates, where you did diffraction patterns, diffraction experiments? It's kind of cool. It really is kind of cool.

Okay. All right. You can show that you can – you can pull up the screen, thanks. We're done. Thank you. All right.

So let me give you the setup, in a little bit more detail. And, again, what we're gonna talk about is Fraunhoffer diffraction, although the initial setup doesn't matter. You'll see exactly where this comes in. It's a certain approximation that shows you how you can approximate one quantity in terms of another quantity if the distance is large. You'll see.

But let me give you the basic setup. So, again, you have a source of light, however it's defined or however it's defined, and it is a great distance from what's – from the aperture plane. So the aperture plane is something over here. This is the aperture plane. And these holes – it is considered to be opaque except where there are holes. So the holes are the apertures. All right? And the fact that you consider the source very far away has the practical implication that, by the time it reaches the aperture plane, it does so – it's basically a plain wave coming in, and what that means is that you can assume that the aperture plane is essentially – not essentially, you can assume it's a so called wave front for the wave, which means that every point here on the plane has the same phase, all right?

So the fact that the source – so that's the first sort of simplifying assumption. A distant source means, practically, that the aperture plane is a wave front. That is, the light is coming in as a plane wave, and again, practically what that means is that everything here is the same phase, which means you can ignore the phase, okay? It's a wave front.

So the wave has the same phase at all points of the aperture plane. All these are realistic and reasonable assumptions, all right? I just wanna spell them out, which means, for all practical purposes, you ignore the phase, you can suppose the phases – you can suppose everything sort of starts from zero at the aperture plane, all right?

And furthermore, we're gonna represent the electric field – now this is the second sort of simplifying assumption, but we're gonna represent the light by a time oscillating electric field, and we're gonna use complex exponentials for that.

So you represent the light as, say, e times – e is for the electric field, for the magnitude, times e to the two pi i new t. So e is the strength of the field, and this is the light on the – this is the light on the – not at the source, but the light on the aperture plane, all right? So on the plane. Aperture, my writing's even worse, plane, as e to the pi i new t, so e is the strength of the field and new is the frequency. So this is where the other assumption comes in that I'm working with monochromatic light, so light of a single frequency. Frequency. So again, it's monochromatic.

All these assumptions can be, to a greater or lesser extent, dropped in more careful and more thorough treatments, but we're simplifying things just to get to the main punchline, which you'll see, which is pretty interesting.

And we're also gonna assume that the field strength is constant on this – on the plane, all right? So we'll assume that e, the strength of the field, is constant. So we assume e is constant, say, e naught on the aperture plane.

Okay. Now some distance away, which I can't represent too well on the blackboard, is the image plane, over here. So light starts from the source, hits the aperture plane, and then gets diffracted by the apertures and winds up on the image plane over here. And the question that we wanna address is, what is the electric field – just think of this in terms of electric field. What is the electric field at a point on the image plane, okay? P. What is the electric field at a point p on the image plane? All right? Why it passes through the slits, it gets the p from different paths, and the question is what is the – how do they all add up? Okay.

So the light gets to p. The light, that is to say the wave, i.e. the wave, gets to p along different paths, because it is bent when it passes through the – or it seems to be bent when it passes through the apertures, along different paths. How do they add up? How do the results add up? How does it all add up? I want you to put it that way.

All right. Now, classically, the way to address this problem is what's called Huygens' Principle. It goes way back, because it goes back to Huygens', and I don't know how it was initially enunciated, but the way it's – the way I'll talk about it, and again, without trying to be terribly precise here, to analyze this you approach this via Huygens' Principle. Huygens'. This is a challenge.

Huygens' Principle, which says that every point in the aperture, every point of a wave front, can be itself regarded as a new source, okay? Each point of a wave front, and by assumption, the aperture plane is a wave front, all right? So – and you have all these little – you imagine that this is made up of a bunch of little points, and each point there on the wave front – each point in the aperture can be – each point on a wave front can be regarded as a new source, all right?

You see what happens as a result of that. You see what happens over on the image plane as a result of light coming from a point here, regarded as a new source, and then you add them all up. You add up – that is to say, you integrate all these – the effects of all these sources.

Now Huygens' Principle has been criticized, and it's certainly not the modern view, but it's still applied, because it produces results that are intuitive and accurate enough. It's interesting, and I am by no means an expert on this – it's interesting to really read different treatments of this – older treatments of it, more modern treatments of it and so on, to see how Huygens' Principle has faired, to see how the whole approach to diffraction has faired, and so on, but this is the way that it was sort of classically analyzed, and it's certainly the way we're gonna do it. I'm not gonna say anything more than this, but you'll see how I'm gonna put it to use.

As a matter of a fact, I think I gave a reference in the notes to a book by Melvin Schwartz, called Principles of Electrodynamics or something like that. I can't remember. It's in the notes. And he actually has a very interesting treatment of all these things. He tries to take a relativistic approach right from the beginning, so it's a very mathematical book and it tries to – it tries to put things in a certain context, and he has sort of scathing criticisms of Huygens' Principle and so on, but he's not above using it occasionally.

So here's how we're gonna apply this. So once again, here is – so forget about the source. The source is infinitely far away. Everything now takes place between the aperture plane and the image plane, all right? So I wanna introduce coordinates on the aperture plane. As a matter of a fact, let me just focus on one aperture. So here's, say, the origin of the aperture plane, and here is a point x on the aperture plane, and here – and I take a very – imagine I take a very thin slice, so this may be part of a bigger aperture. Let me do it like this. But I take a very thin slice, say of width d x.

All right, now this is a sideways view of this. We're essentially going to turn this into a one dimensional problem. If you want, you can consider the plane – the aperture plane coming out like this, and I'm just gonna – I'm gonna take a aperture of dimension one in this direction and dimension d x in this direction, all right? So it essentially reduces this to a one dimensional problem. That's another simplification, but it's – again, it's – they're reasonable simplifications, depending on the dimensions of the actual physical problem.

All right, and so here's p over here, so light comes out of x, or out of this little slice, and reaches p, all right? The field here – the field at x is roughly – well the field – the field – let me say the field associated with the small slice is about – I'll even put the – let me write it up here. The strength of the field in the slice of size d x at x is about e naught the strength times e to the two pi i new t times d x. So the field strength is roughly – the field strength is constant on the plane, so it's e naught d x of it in the particular little slice, and then I multiply it by the – it's time varying, so it's e to the two pi i new t, okay?

Now – and so I'm assuming I have zero phase on the aperture plane. Okay, now what happens when it reaches the image plane? There's a decrease in the magnitude of e by one over the distance, but that's not the main effect, all right?

The main change in going from – that is, the main effect, really, maybe I should say, the main effect in going from x to p, over a certain distance, is the change in phase. And what we have to – yeah?

Student: Is the t up there – is that supposed to be an x or is that t for something else?

Instructor (Brad Osgood):T is time, because it's a time varying field.

Student: Okay.

Instructor (Brad Osgood):All right, so like, it's an oscillating electric field, okay? So t is time. It's actually gonna come out. That's actually not gonna turn out to be an important thing, but that's – you represent the field that way. I should have said, by the way, I'm sorry, as per usual, in this case, you represent real quantities by complex exponential, so the actual electric field would be the real part of that, but – as is common in this case. You usually sometimes call this the analytic version of the signal or the analytic version of the field when you represent it by a complex exponential. The actual field would be, say, the real part of that, but we're sort of using complex quantities. I can do it without thinking about it. We're used to representing real quantities by complex quantities.

Now – and if it reaches p from all sorts of different paths, then what you want to take in account – take into account, is the phase change that is associated with the different paths.

So say this is a distance r, say it travels a distance r, all right? So what is the phase change in traveling a distance r. R from x to p. Well how many cycles does it go through? Think in terms of – measure this in terms of wavelengths. If it goes through one wavelength, then it goes through one cycle. If it goes through two wavelengths, it goes through two cycles. If it goes a distance r, r, the wave goes through r over lambda cycles, where lambda here is the wavelength.

You can – so again, wavelength. It goes through – if r is equal to lambda, then it goes through one cycle. One wavelength goes through one cycle. If it goes through two wavelengths, it goes through two cycles, and so on. If it goes through half a wavelength, it goes through half a cycle. In general, if it goes through distance r, then the number of cycles it goes through is r compared to the wavelength. R over the wavelength. R over lambda. You can even work that out with formulas if you want, but that's intuitively what's going on here.

Okay, so what's the phase change? So the phase change – so the phase change is two pi r over lambda, all right? And then – and so the field or the – I should maybe say field – let me write it down, then I'll say something about a field, at p due to the field at x is – let me write it like this. D e is – a little bit of change in the field – I'm gonna – the differential of the field is e naught times e to the – it's still oscillating at the same rate. That's not changing. It's still the same frequency, so it's e to the two pi i new t – I'm gonna need more space here, I can tell – is e naught e to the two pi i new t is – it's still monochromatic. What's happened here is the phase change. E to the minus two pi i r over lambda. If I represent the phase change also that way, as a complex exponential, times d x. That's the key expression. That's the key expression.

There's been a drop off – drop off in the magnitude of the field, that's true, but that pales in effect in comparison to the change in the phase. Okay.

So what is the total field? Is the integral of this expression. So it's the integral over all the apertures. This is the field at p, fell from all the different sources. So it gets the p from all different source – from all different paths. E naught times e to the two pi i new t, e to the minus two pi i r over lambda d x. So I'm integrating over the aperture plane. I'm integrating with respect to x. X is a variable that describes the location on the aperture plane. So I could just say, aperture – I'm saying integrating over the aperture, I should say I'm really thinking about integrating over the aperture plane, but the only time the – you get a non-zero contribution is when – because the field – you can regard the field as being a zero other than at the apertures, because the – other than the – other than the apertures, the plane is opaque. The aperture plane is opaque.

All right, so I can pull out these constants here, and it's really – this is the part that depends on x, because r depends on x, all right? On x. So that's where the dependence is coming in in the integration. Not here. Not in the time variable. All right. So as a matter of fact, I can pull that out, and I can more or less ignore it in the rest of the discussion.

So I can write this as e equals e naught. That's a constant. I assume that's a constant – times e to the two pi i new t times the integral over the apertures of e to the minus two pi i r over lambda d x. All right.

Now I should also say that what you observe – what you see with your eyes or what you measure with instruments is the magnitude of e. So, in particular, the time varying part of that goes away because the time varying part of that, in the – under the assumptions we have, monochromatic and so on, is just a complex exponential which has magnitude one. So you see – you measure, measure, the magnitude of e.

So the quantity of interest here really is the integral, not the stuff out in front of it, okay? Is this integral over the apertures of e to the minus two pi i r over lambda d x. Now this is not a useful expression so far, it's too complicated. Here's where the approximation comes in. So this is not useful. I mean, maybe there are times when you can write it out more carefully and try to evaluate the integral, but generally, sort of as a principle, as a way to understand the phenomenon, it's not so useful – that is, you need an additional assumption to make it a useful expression, and that's where this Fraunhoffer approximation comes in.

And so now we bring in the so called Fraunhoffer approximation, and here's what that means. So let me, again, draw the picture over here. Here's the aperture plane over here, here's x, here's the origin of the aperture plane, here's the image plane, here's p, that is what – that's what we calculated. We calculated the field at p due to all the ways that light reaches it from all possible points on the aperture plane. Let's call this distance r naught. So that's the distance from zero to p. That's a fixed distance. Here's r, all right? And I wanna be sure I write this right, get this right. Here is theta. Then this distance here – I'm sorry for being careful. I'm looking at my notes carefully here, but I don't wanna screw anything up. That's x sine theta. That little length in there is x sine theta.

We're gonna assume that r, this distance, is much greater than this distance, than x. That's the Fraunhoffer approximation. That's the Fraunhoffer assumption. You're gonna assume that r is much greater than x. By that assumption, this little bit is such that r minus – if I take r naught minus x sine theta, that's approximately r. That is, the operational effect of this assumption is that r naught minus x sine theta is approximately r for all the different values of x on the aperture, so there's actually several assumptions here. The apertures shouldn't be too big, they shouldn't be too far away, but at any rate, the distance of the image plane should be very much farther than whatever x is, wherever x is varying on the aperture plane. So way far away. And under that assumption, r naught minus x sine theta is approximately – so the fixed distance from zero to p – remember, we're calculating the electric field just at a particular point in the image plane, at a point p in the image plane, from all the different contributions, and at that point, r naught minus x sine theta is approximately r. That's the expression that we plug into the integral.

So you plug this into the integral, giving the field at p, and what do you get? You get – I don't know if I'm erasing it here, but I'll write it out in just a second. Forgetting about the constants out front, you get the integral over the apertures of e to the minus two pi i r over lambda, so it's one over lambda times – and r is approximately r naught, which is fixed, minus x sine theta. So it's r naught minus x sine theta d x. D x. Right. Okay?

Split this up – again, I'm forgetting about the constants out front, because that's not so – that's not important. It's the integral, this integral, that's important, and even in this integral now, we can pull out another constant that's gonna be of absolute value one, so it won't contribute to the magnitude. So it's the integral over the apertures again of e to the minus two pi i r naught over lambda, so that doesn't depend on x, times e to the minus two pi i x sine theta – excuse me, plus two pi i x sine theta over lambda d x. I haven't done anything except plug that in.

Now pull this out, because that doesn't depend on x again. R naught is the fixed distance from zero to p, and we're calculating everything – we're calculating the strength of the field at p. This is equal to the integral over the apertures – oops, I forgot to pull up the cosile. E to the minus two pi i r naught over lambda times the integral over the apertures of e to the plus two pi i x sine theta over lambda d x. Okay.

Now, it is common, in the biz, to introduce another variable, an auxiliary variable, that replaces sine theta but by just another name. So introduce p equals sine theta over lambda. One always talk about the – one always talks about diffraction through an angle, and the angle is theta here, but this is the variable that you introduce, sine theta over lambda, so that the integral becomes – and forget about the constant out front again. The constant out front has absolute value one. You only measure the magnitude, you only see the magnitude, so forget about that constant. Let me just concentrate on the integral.

So it becomes the integral – I say equals, becomes the integral over the apertures of e to the two pi i x p v x. Okay. Almost there. Now, now, now, now, let's introduce the aperture function, so to speak. So something like a of x is one if x is in an aperture, and zero otherwise, okay? You're only integrating over the apertures. That's the only place where the field is non-zero, all right? Then write – then I could write this integral a little bit differently. I can imagine integrating over the entire aperture plane, that is to say, from x point – from minus infinity to infinity of a of x times this, because a of x is only one when I'm in an aperture, and it's zero when I'm not in aperture. So I can write the integral as the integral for minus infinity to infinity of e to the two pi i x p a of x d x, okay?

And now what do we recognize? This is, of course, nothing other than the inverse Fourier transform of the aperture function at p. The dual variable here, the variable in the Fourier transform, the form of – the two variables in the Fourier transform are x and p. P is sine theta over lambda, so p has to do with the geometry of the situation, but as far as the – as far as writing the formula goes, if you recognize this is a Fourier transform, the variables are called x and p, and this is the inverse Fourier transform at a of p.

Now, I have to make a couple of comments here, so I take a deep breath. You know, I said once that people had different conventions about what they called the Fourier transform, where they put the two pis and sometimes either where they put the plus and where they put the minus? You remember I said that? And different fields, say, have different conventions? Well, in physics, it's more common to define this as the Fourier transform, and the integral with the minus sign is the inverse Fourier transform, and so, if you were in physics and somebody wrote down this formula, they wouldn't write, this is the inverse Fourier transform of the aperture function, they would write, it's the Fourier transform of the aperture function, and as far as I know, this may be the reason, in physics, why they use that particular convention for the Fourier transform. I don't know if that's true or not, but it is a reason for it.

So if you would – if I would have defined the Fourier transform differently, if I would have defined the Fourier transform with a plus sign here instead of a minus sign, then I would have had, maybe, the more elegant result saying that the strength of the field at p, the value of the field at p, is proportional to the Fourier transform.

Now in the notes, actually, I fudged this. I wrote things a little bit differently, so I wound up with a Fourier transform, but that's really not true. So what I gave you here is – it's only a question of differing between the plus sign and the minus sign, so I haven't had the, somehow, nerve to go back and change the way I wrote it in the notes because I really want it to come out in terms of Fourier transform, but I had to be honest. So this is our result.

Now it's pretty – this is a wonderful result. This is a wonderful, intuitive result. It took us a while to get here, but what it says is that, under the Fraunhoffer approximation, so for far-field diffraction, the strength of the field at the image plane is the Fourier – think of it just the Fourier transform is the Fourier transform of the aperture function, all right?

You have an aperture plane with a bunch of slits cut into it, and you have a diffraction pattern. What do you see? You see a bunch of bands of white and black. You see light and dark. What are those bands? Analytically, they are given by the magnitude of the Fourier transform of the aperture function. That's the takeaway headline. For under the – for far-field diffraction, the magnitude, which is what you see, or the – maybe let me put it this way. The intensity of the light, because that's what you're seeing, is the magnitude of the Fourier transform of the aperture function.

That's the summary. That's the headline. And that's a very valuable thing to know if you go into this area. If you don't go into this area, who cares, but if you go into this area, or if you want to have a conversation about it, this is very important, and I was just talking with a friend of mine, Marty Fejer, who's in applied physics, just the other day, and we were talking about what I was doing in the class, and I was saying that I was gonna do this, and he was very pleased to hear that, because what he finds is that students today, when they're confronted with a diffraction problem, will try to solve things numerically, they'll set up a MATLAB program and they'll do all sorts of computation, but what he really wants them to be able to say is this sentence, at least as a starting point to get some intuition for what the field should look like, think about what the apertures will look like. You have all this experience taking Fourier transforms.

That'll tell you – that'll give you a good sense of what the – what you should be seeing on the image plane. I mean, you can do more detailed computations, numerical approximations and so on, but starting with this as the principle of what the field should look like, what the intensity – what the light should look like that you see, is given by the Fourier transform of the aperture function. It's an extremely nice results.

So let's do some examples of this. Let's do single slit diffraction, the most basic diffraction experiment. A single slit. What is a slit? A slit of width, say, a, is given by a – and let's say just center it at the origin, so this is at the origin, then the aperture function is described by the rectangle function, pi sub a of x. That's a function which is one from minus a over two to a over two, and it's zero otherwise, so the aperture is pi sub a of x. That's the function that describes a single slit.

So what do you see on the image plane? Sorry?

Student: Is that one minus?

Instructor (Brad Osgood):No, it's one – the rectangle function is one on the aperture and zero off the aperture, okay? So pi is one when x is between minus a over two and a over two, and it's zero outside that.

So what is the intensity of the light? So the light – the intensity of the light is the Fourier transform, and this is the inverse Fourier transform. This amounts to the same thing. Intensity of the light on the image plane is – what is the Fourier transform of this? Is – right down here, is a sink – let me write it down in terms of variables. A sine theta over lambda, writing p equals sine theta over – over lambda.

You see – well, excuse me. I'm sorry. The Fourier transform is given by this. The intensity is the square of that. Fourier transform is that, so the intensity you see – essentially the magnitude of this, all right? So you see the square of this. You see the magnitude, so absolute value of this. So why do you see light so – why do you see bands of light and dark? Because that's what the sink function is like. The sink function is big in some places and small in some other places, it even has zeros. Now the zeros are the places where it's dark. You take the absolute value of this thing to actually see the magnitude, so you're always seeing either some light or black, all right? But you're seeing – it decays. It's bright in the middle, then it has all these sort of little variable bands of light and dark separated by – separated by black, separated by zero intensity, and that's because it's that Fourier transform. It's given by the Fourier transform of the rectangle function. Ain't that cool? I think it's cool.

Now let me take another example. Let's tie in – I said I wanted to bring in delta functions here, so let me do that now, then I won't say too much more today. We're gonna do this again next – we're gonna take this up again next time, a little bit even more generally, and we're gonna talk about crystallography, and we'll talk about sort of one dimensional crystals. We'll do higher dimensions a little later in the course, where I'll bring the delta function in even more definitely.

Suppose I have a point source, all right? Suppose the aperture is a point source, like a pinhole, but smaller than a pinhole. A point. Then what's a good – what's a reasonable approximation of the aperture function? Delta. Then a of x is delta of x.

Now, before I write down the formula, before I write down the formula, if you have a point source on an image plane – on the aperture plane, and you put an image plane far away from it, what do you expect the illumination to be on the image plane?

Student: Circles.

Instructor (Brad Osgood):Pardon me?

Student: Circles?

Instructor (Brad Osgood):No. A circle would be a real – I mean, what you get on the image plane of circles would be a real circular aperture, but this is a point aperture.

Student: Constant.

Instructor (Brad Osgood):Constant, and the image plane is far away. You would expect –

Student: Constant.

Instructor (Brad Osgood):Constant. You would expect a constant illumination. You would expect the illumination on the image plane – if the image plane is far away, it would just be uniformly illuminated. Uniformly eliminated. Eliminated, did I say? Uniformly illuminated. That is a physical interpretation of the fact that the Fourier transform of the delta function is one.

This is a physical interpretation of the Fourier transform of the delta function or the inverse Fourier transform of the delta function, it amounts to the same thing, is one, because the intensity of the light you see on the image plane is the magnitude of the Fourier transform of the aperture function. The aperture function is given by a delta function. It's a point source, and its Fourier transform is one. You expect uniform illumination on the image plane. Ain't that cool? I think that's cool. I think that's cool.

I'll do one more example. I'll do it over here. The famous – one of the things that really got people freaked out was the so called double slit experiment of Young, and I won't talk about the experiment. I won't talk about the experiment or what the results were, but that was a famous early experiment in light, so Young's double slits. You can look this up. And, of course, talking about diffraction, light diffracting through double – through two slits, or talking about particles going through two slits, it's been a famous experiment in physics forever.

So the situation might be something like this. Might be where I have two slits that are modeled by two rectangle functions, all right? Say, a distance – say the distance between them is – I'm gonna have my – I'm gonna stick with my notes here. So say the distance between them is b, and they're each of width a. So it's given by – the aperture function would be given by the sum of two rectangle functions is pi sub a of x minus b over two pi sub a plus pi sub a x plus b over two.

So the distance is b. The distance between them – say this is at height plus b over two, this is at – this is centered at plus b over two, this is centered at minus b over two, so it's easy to write down the aperture function. It's just the sum of two rectangle functions. It's one – they don't overlap. It's one when you're in either aperture, and it is zero outside it, okay? And you're either in one aperture or the other aperture or outside it. And what is the Fourier transform? What do you get? And I think I have a picture of this. I'm sorry I didn't check. I know I have at least a graph of this thing in the notes. I don't think I have a picture of the actual diffraction experiment, but you can find it. You can track it down.

So the Fourier transform of this – we know how to do this. This is the module – this a modulated version of the – or it's a shifted version of the – it's the sum of two shifts of the rectangle function, and we know how to find that with – via the modulation theorem. In the notes, I talk about how to do this by using the delta function, actually, or you can to it this – well, I'll just write down the result, because we gotta go.

So you get – and I'll just – a sink a times p, so p is again sine theta over lambda times cosine of two cosine pi b p. So, again, here p is equal to sine theta of lambda. Theta is the angle that you are diffracting through at p, okay? Theta makes at p. And that – you should recognize this. This cosine factor comes in because of the modulation, because they're – because of the shift plus and minus, via plus b over two and minus b over two, and the sink comes in because that's the Fourier transform of the rectangle function. The stretch rectangle function.

In the notes, I mentioned – I wanted to bring – actually use this as an application of the delta function, because I wanted to write this is as a convolution of two shifted deltas, but I'll do that next time or I'll talk a little bit more generally about this next time. I just wanted to give this result too, because this is another famous experiment and famous formula in physics, what the diffraction pattern for the double slit experiment is. So look that up, I'm sure you can google this, and you'll get pictures that correspond just to this sort of function.

So it's a physical – I should have said, actually, this is a physical interpretation of the Fourier transform of the delta function is the constant function one. The single slit, or the sort of double slits, are physical interpretations, physical manifestations, of the sink function. You actually see them coming up in diffraction experiments. It's cool. All right, that'll be it for today. Then on Monday we're gonna take this a step further, talk about crystallography, and that will actually lead us into a discussion of sampling. You'll see. Thank you very much.

[End of Audio]

Duration: 56 minutes

TheFourierTransformAndItsApplications-Lecture16

Instructor (Brad Osgood):And we're on. I love it how they take me by surprise.

All right. Once again, let me call your attention to the midterm information. I mentioned this on Friday; I also posted it up on the website and sent an email to the entire class, so the exam is on this Wednesday, Halloween. Boo. There are three sessions: 2:00 to 3:00 and 4:00 to 5:30 are both in the same location, it's Building 380, that's the math corner down in the basement 380W, and then from 6:00 to 7:00 it's Skilling, that's here. Okay?

I will bring copies of the exam, of course, and the bluebooks and also the formula sheet, all right; the same formula sheet that's been posted on the Web for a long time so I'll bring copies of that. It is open books, open notes, you can bring whatever you want but don't get ridiculous, as I said.

It's gonna cover stuff on up through sort of delta functions, properties of delta functions, generalized Fourier transforms, so it won't talk about diffraction, it won't talk about sampling, which is gonna be our next topic. And I think that's about it on the midterm; any questions about that, any issues that I have to know about for the midterm exam? It's a 90-minute exam. And again, the idea is that I hope it's gonna be sort of more conceptual than computational, although I say you can't rub away all computations, obviously. So anything else?

Also, on – so we'll have class on Wednesday. It's relentless. Relentless. As I also mentioned, the homework, the current homework assignment is not due till Friday. The magnanimous Os has given you a couple of days so that'll be on Friday. And the current – the new homework assignment however, because it's relentless, is already posted up on the web, on sampling.

Yeah?

Student: Can we come to any of the sessions?

Instructor (Brad Osgood):Yeah, but only one.

Student: Okay.

Instructor (Brad Osgood):I ask people to sign up just so I can get an idea about the size of the different sessions. That was not a contract, it's not written in stone, so if you want to change your mind that's fine. But as I say, you can only take the exam once. Once, that's the only deal. Okay? But any of the sessions should be okay. I mean, you know, within some limits, I guess. I don't even if the entire – I think the entire class could fit – so if everybody shows up at 6:00, I suppose everybody'll fit in here but it might be a little tight.

Yeah? There's another question.

Student: Does the material on the midterm cover the stuff that's in the homework?

Instructor (Brad Osgood):Sorry? It doesn't – well so the stuff that's in the current homework? Only to the extent, I guess there's some generalized Fourier transforms, like you have to know the Fourier transform of the, you know, unit step function and signal function, and there were a couple of questions along those lines but nothing much sort of outside of that. Okay?

All right? Okay. All right, let's move on.

I want to talk a little bit more about diffraction actually. And as a way of actually making a transition to our next topic, this may seem a little odd way – our next topic is sampling an interpolation. And going from diffraction to sampling interpolation may seem like a little odd of way going but it's – there's an interesting connection here that I want to exploit. The topic itself that I – the general areas of diffraction, and in particular what I want to talk about today, is interesting in itself and it does make actually for a nice link, so I want to talk about the problem of crystallography.

We're gonna actually return to this when we have higher dimensional, when we talk about higher-dimensional Fourier transforms. So today, I'm only gonna talk about the one-dimensional case, which of course is not realistic but it has some essential ideas that you find in the higher-dimensional case. And as I say, it makes a nice transition to the next topic that we're gonna be talking about.

Let me remind you about the headline from last time. So headline from last time, when we talked about diffraction and the Fourier transform, is that diffraction patterns are given by or determined by the Fourier transform of the apertures that cause the diffraction. Of course, this simplifies things but that's – if you're looking for a quick summary of what our chief conclusion was last time, this is it. Diffraction patterns are determined by the Fourier transform of the apertures or the aperture function.

We had approximation; we talked about far-field diffraction, all the rest of that jazz, never mind. That was all important, of course, but this is the main thing to be carried away with it, and that's what I'm gonna be using today also.

Now here is the setup for what was troubling people, what was puzzling people when x-ray crystallography, x-ray diffraction was first invented or first brought to bear on certain set of important problems. So the setup that I want to talk about it as follows: x-rays were discovered in 1895 by Roentgen, of course, Roentgen, R-O-E-N-T-G-E-N, or some approximation of that spelling. All right?

Matter of fact, I remember actually in 1995 everybody was celebrating the 100th anniversary of x-rays, a very exciting time. And the question was what are they? Are they waves? I mean their fundamental nature was not understand, so what are they, or what were they. Are they waves, for example? It was a new phenomenon. If so, then certain considerations led them to conclude that the wavelengths should be about ten the minus eighth centimeters. All right?

If so, and the wavelength, which was too small to measure precisely, wavelength should be around ten to the minus eighth centimeters. All right? So that's too small to measure by other means – by the means that we're used to measuring different sorts of visible light, say, other sorts of waves which were diffraction gratings, too small to measure. Let me just say too small to measure, period, with any of the standard techniques, okay, for example with diffraction gratings.

On the other hand, or a different set of – a different line of questioning was crystals; 2.) Crystals. Been around for a long time and observed for a long time. What are they? In particular, it was clear from people – to people who were cutting them open, chiseling them around and making small crystals out of big crystals, that somehow the microscopic structure ought to be determined by the microscopic structure by the atomic structure. And so, the conjecture was that crystals were crystals because there was a regular periodic array of the atoms that made them up.

So microscopic structures should somehow be determined by the atomic structure, and the conjecture was the atomic structure should be a lead of some type. That is the atoms were arrayed on some – in some regular pattern. Atoms arrayed on a lattice. How to test that, experimentally?

All right, so there were famous experiments, series of experiments done in 1912 by Max von Laue, in 1912. All right, so he purposed to study the physical – the nature of crystals by using x-rays to conduct diffraction experiments, so study the atomic structure of crystals via diffraction experiments with x-rays. All right? So he had a number of hypotheses then. He had, first of all, a hypothesis that x-rays were waves. And to assume that they're waves means that they will exhibit diffraction. They'll exhibit the characteristic physical phenomena that are associated with waves: reflection, refraction, and, in this case, diffraction.

So he had – the hypotheses were 1.) X-rays are waves, some kind of waves of undetermined origin so will diffract. Now, the diffraction pattern depends on the size of the aperture relative to the wavelength. All right? So he had a second hypothesis that crystals would serve as an appropriate diffraction grating. That is to say, crystals are periodic, have a periodic atomic structure, maybe I should say lattice atomic structure, and the spacing of the – the spacing on the lattice, so the spacing of the atoms is comparable to the wavelength of the x-rays. The spacing of the atoms is comparable to the wavelength of the x-rays.

All right? So that was the setup for his experiments. And I'm gonna do a one-dimensional version of this. I want to do – talk about the mathematics that comes up and how the Fourier transform comes in, in a one-dimensional version also. Of course, everything here is three-dimensional. Crystals are three dimensions, things are much more complicated. The pictures that you get, the setup is much more complicated, but you can see some of the essential points if we already look at the one-dimensional picture.

Now before I go on any further, since I am out of my depth here, has anybody ever done any experiments like this, anybody familiar with x-ray diffraction experiments or x-ray crystallography, anybody in materials or anything out there? I have a lot of buddies who are material scientists, so I'm always afraid of one of them is gonna show up when I talk about this one day.

All right. Well you will, if you – so it's really it's a, by now of course a very standard experimental procedure, and quite refined and quite sophisticated. But this how it looked at the beginning, and there's a very important connection here with a Fourier transform, and a very fundamental fact that you have to know about the math in order to interpret the answers that you get, in order to interpret the results of your experiment. And that's really, what I want to get to.

So let's look at the one-dimensional version of this. I'll tell what you want to – I want to show what you want to measure, I want to say what you want to measure, and how you go about measuring it. Or again, what I really want to is how the math shows you an important physical property of the – physical result of the experiments.

All right, so I think of a one-dimensional crystal as an array of atoms along a line, as evenly spaced, let me just say, along a line, like so. And let's say the spacing is P. I'll get out of the way in just a second here. P. All right? So they're – the atoms are spaced P units apart, whatever the distance is. And there's effectively an infinite number of atoms because it is. All right? So it's effectively an infinite line. That's actually very important in the math, an infinite array.

Okay, now what you want to study – and the atoms are all identical. All right? What you want to study is, again for reasons, which I don't know, all right, but those who do the experiments understand this, so you want to study the so-called the electron density distribution for the entire crystal. You want to study the electron density distribution for the crystal. Now because the crystal consists of identical atoms arrayed, the idea is there's an electron density distribution for the individual atom and then that's replicated for the entire crystal. Or, as we would say given the work we've done, it's a periodized version of that.

That is, so the electron density distribution for the crystal is then a periodized version of the electron density for a single atom. All right? So let's say rho of X is the density, the electron density around a single atom. Density for an atom, all right? Then if you periodize it, the way we always periodize functions by summing up shifts, and write down a formula for that.

So again, here's the crystal with the atoms spaced P apart. Here's whatever it looks like. Imagine this a description somehow of the density, let's call it rho of X, so that describes the density of a given atom, and then that's replicated for the entire crystal, or periodized for the entire crystal like that. So let's call that rho, the density for the crystal, and that's the periodized version of the individual one. So let's say rho, so P of X indicating that spacing P apart, is the sum from K equals minus infinity to infinity of rho of X minus K times P.

Okay? That's the standard formula for periodizing a function to get a function of period P. That's what the period should be, as dictated by the physical spacing of the atoms, P. Okay?

Okay, according to our headline from last time, the diffraction pattern should be the Fourier transform of this, or should be determined by the Fourier transform of this. The diffraction pattern, which is what you want to see, a matter of fact it's what you do see when you make the experiment.

So you shine x-rays through the crystal, a bunch of spots or whatever, I shouldn't say a bunch of spots because you don't know what's gonna happen, but a bunch of spots show up on the x-ray film. And the pattern of the spots, the way the spots are arrayed, should be determined somehow by the Fourier transform. Diffraction pattern determined by the Fourier transform of rho, so I want to find that. Find this or find this in a – determine it in an effective enough way this is gonna allow you to interpret the results of the experiment.

Okay, now to do that, I'm gonna write the periodization – I'm gonna do something new with the periodization that we haven't done before. We didn't do when we were first – when we first studied it, but we have new techniques that are available to us now that allow us to carry the analysis a little bit further, a little bit differently than we did before.

So I'm gonna write the periodized version as a convolution. OP of X at a convolution, and I'm gonna invoke the convolution theorem. Now you remember I can shift a function, it's a sum – but what is the periodization? The periodization is now, as it always was, a sum of shifted versions of the function. You have a single function, and you shift it and you add them all up. I can shift a function by convolving with a shifted delta function, so that is rho of X minus KP is rho of X convolved with delta of X minus KP.

And you notice I am now happy to write variables in my delta functions, if you don't like it, tough. All right? I made a big deal out of the fact that deltas aren't functions of points, they're functions of functionals – they're functionals and so on, but I feel no hesitancy in writing that. Okay?

That's the shifted version. You shift the function rho by convolving with the shifted delta function; that's a basic property of delta functions. All right, so the periodized version is the convolution of the fixed function rho with a sum of delta functions. That is rho of X, rho P of X I'm calling it, that's the sum minus infinity to infinity a rho of X minus KP. That's the sum, K equals minus infinity to infinity of rho of X, the fixed function convolved with the shifted delta function, X minus KP.

Or if I want to write, if I want to bring the – the rho is fixed here, doesn't depend on K, so I could bring that outside the sum. That is I write this as rho of X convolved with this big sum of delta, it's shifted delta, K equals minus infinity to infinity delta of X minus KP.

Okay now this object is gonna be an object of great interest and object of our study, so let me actually introduce right now a term for it, a notation for it, so I can work with it a little bit more. Everybody with me, all right? I haven't done anything really differently than I've done before. I've taken just a slightly different point of view toward the old process/procedure of periodizing a given function.

So I want to use the notation – here, let me write it down and then I'll pronounce it for you. This is called the Shaw function or Shaw distribution, really, of spacing P. All right? That's the Cyrillic letter, not drawn very well, or it's my version of a Cyrillic letter, Shaw. And as far as I know, actually this notion was introduced and popularized by Bracewell. I don't know the history of it but is now pretty standard. All right?

And the reason why that particular symbol is chosen, or reason why Bracewell if indeed it is due to Bracewell, reason why he chose that letter was it was supposed to be reminiscent of the picture associated with this. That is, if you would draw the picture associated with the sum of shifted delta functions, you would have bunch of arrows spaced P apart. So zero P, two P, minus P, minus two P, and so on, infinitely many of them. And three of them sticking up, make a Shaw. I'm not making this up. Okay? That's the pictorial representation of the Shaw distribution.

All right, now what is the Fourier transform of this periodized electron density distribution? Why use the convolution theorem? Okay? So I write rho of P of X then is rho of X convolved with the Shaw function with spacing P of X. And if I'm interested in the Fourier transform because I want to understand the diffraction grating, I want to apply the convolution theorem. So Fourier transform of rhos of P of X is the Fourier transform of the convolution which is the Fourier transform of rho, let me just write it like this without the variable. It's the Fourier transform of rho, I'll get it, times the Fourier transform of Shaw.

So the problem becomes how do you find the Fourier transform of Shaw? All right. This is a fixed function, depending on the nature of the crystal. All right? This is something – depend of the crystals just depends on where we have the spacing. I don't have anything – I don't have any atoms in there, it just indicates that I put a delta function at each point on the line, spaced P apart, and the general crystal is defined by this convolution.

To find the diffraction pattern, I want to find the Fourier transform of the fixed function, which is gonna be the same for every atom, for each individual atom rather, and then times the Fourier transform of the Shaw function. So the question is: What is that? Ah, there's a P here. So what is, that's the question. All right now let me say a couple things about that. It's not obvious. It's not obvious and it's remarkable, actually, what happens. And we can analyze it very easily and rigorously with all that we've developed.

So first of all, does it even make sense to consider something like that? Well yes. A given delta function is, of course, a distribution, it's the simplest distribution, it's the evaluation distribution, but I'm considering a huge sum here. Does the sum really make sense? I mean does an infinite sum of delta functions make sense. That's the first thing I should comment on is that the Shaw function makes sense. Shaw function of spacing P makes sense as a distribution.

Matter of fact let me specialize here, excuse me, to take the case P equals one, so the atoms are spaced one apart, then I'll do the general case in just second. So let me take the case, take P equals one, and again I'll pass to the general case quite easily after I do this case. So and I'll draw out the subscripts. So let me just write Shaw of X is sum from K equals minus infinity to infinity of delta X minus K. All right, so the deltas are space on apart. At every integer, you put a delta.

Then this thing makes sense as a distribution. Now again, why? I tell you I have a distribution. What that means is you give me a test function, I have to tell you how this operates on a test function. Well it operates by evaluation. In this case, I add up all the values. That is the Shaw function operating on a test function phi, that's the sum of shifted deltas operating on phi, that is just the sum from K equals minus infinity to infinity of phi evaluated at the points to which the delta – where the delta function is shifted. Okay? I won't – I skipped a couple steps here but that's what the result is and you can fill that in.

Now why does this make sense? Well if phi is a test function, and phi is rapidly decreasing, then it's gonna be going down fast enough, way fast enough, so that this sum will converge. All right? Because the sum converges, the pairing makes sense. All right? So the sum converges so all is well, all right?

So if this thing makes sense as a distribution, then so does this Fourier transform so this also make sense. All right? In fact, I can say this. In fact, I can say that the Fourier transform of the Shaw function operating on phi is, by definition, the Shaw function operating on the Fourier transform of phi, classical Fourier transform, and I know what the Shaw function operating on the Fourier transform – operating on anything is, it's just the value of the thing at – the sum of the values at the integer points.

So this is the sum from K equals minus infinity to infinity of the Fourier transform of phi operating on K, evaluated to K. So the Shaw function, again, operating on phi is the sum of deltas operating on phi. Each delta operating on phi pulls out the value of phi at the place where the delta function is centered, phi of K, and I add them all up. All right?

The definition of the generalized Fourier transform is the Fourier transform of whatever distribution is operating on phi is the distribution operating on the Fourier transform of phi. So it's the Shaw function, again, operating on the Fourier transform of phi. This is the sum of shifted deltas. Each delta operates on this thing to evaluate at the place where the delta is shifted, so Fourier transform at K, and then you add them all up. So it's that result. Okay? But there is more to say.

Now you could just look at the formula for the Shaw function and try to write it down as formula for the Fourier transform because you know the Fourier transform the individual delta functions. All right? You could write down the Fourier transform of the Shaw is the sum from K equals minus infinity to infinity, the Fourier transform of these delta functions. And I know the Fourier transform of a shifted delta function is just the complex exponential. That was one of our basic, one of the first formulas we derived when we were talking about generalized Fourier transforms.

So I could write down this is equal to the sum from K equal minus infinity to infinity of E to the minus two pi IKS. And you would be right to write that down. That's okay. All right? But it misses – that's correct, but it misses the point. This series is hard to consider, actually. This series, again, doesn't converge as a classical series. I'm adding up – you think about it, I'm adding a bunch of sines and cosines here. I'm adding a bunch of complex exponentials. The coefficients in front are all one. There's no way that this series converges classically but it does make sense as a distribution. All right?

It doesn't converge classically, all right, but it's okay as a distribution and the formula's right. But okay as a distribution and it's okay to say that the Fourier transform of the Shaw function is this sum of exponentials. It's okay but it misses the point. But you're missing something. But you're missing something, all right? What you are missing actually is the deepest fact known about the integers, missing.

To make this – to see the magic, and there really is some magic here, you bring in the so-called Poisson summation formula. Okay? You need the Poisson summation formula. It is, as I say, one of my good friends who's a number theorist, analytic number theorist, refers to this as the deepest fact known about the integers. Once the Riemann hypothesis is proved, that'll be the deepest fact known about the integers, but until that happens this is pretty deep. And it's also not so hard to prove.

All right, it says the following. It says if phi is a rapidly function, there are other classes for which this applies but this will be sufficient for our purposes, then if I add up the values of phi at all the integer points, that's the same as adding up the values of the Fourier transform phi at all the integer points. Fourier transform of phi. Now this is an amazing, unexpected, not intuitive result. The individual – I should call these K, sorry. Sorry, sorry, sorry. All right?

The individual value, the value – I mean what, who cares about the integers, right? I mean, you know, why would the value of the function at an integer come in and why would the value of the Fourier transform at an integer come in? And the value of the function at a given integer has nothing to do, really nothing to do with the value of the Fourier transform at a given integer. I mean there's no relation there. You can write down the formula for the interval and so on but there's no – you shouldn't expect any sort of special relationship with how big phi is at five versus how big the Fourier transform is at 5, or 10 or 6 or 17 or anything else. All right?

But the sum, so this is a case where it's the limiting behavior and adding them all up, the sum of the values of the function phi at all the integer points is the same as the sum of the values of Fourier transform at all the integer points. Okay? Now I'm gonna prove that for you.

It is an amazing fact and remarkable fact and I'd say the deepest fact known about the integers. And in fact, if you took an undergraduate signals and systems class, you probably know this fact but you didn't know you knew this fact. Or, as I think I said in the book, the deepest fact about the integers is well known to every electrical engineer, and every material scientist for that matter, they just don't know they know it.

All right. So I'm gonna give a prove of the Poisson summation formula. How am I gonna do that? All right I'm gonna periodize proof. I think it's the first time I've actually written that word on the board but I think it deserves it. All right, so I'm gonna periodize phi. So phi is, again, little phi is a given rapidly decreasing function, periodized phi to have period one. All right, so what is that? That is phi of – I call it capital phi of X is the sum from K equals minus infinity to infinity of phi of X minus K.

That's the periodization, never mind bringing in the delta functions or anything like that, just think of the old days where we just wrote down that formula. And that periodizes little phi to have period one. Now everything is smooth here, everything is legit because phi is rapidly decreasing, the series converges, converges uniformly, it's continuous, it's everything you would want. Don't worry about it, don't ask me about it, don't get me any trouble. All right? Good.

Now capital phi being such a nice function has a Fourier series: expand phi, capital phi that is, in the Fourier series. All right? Good. All right phi, phi of X is sum K equals minus infinity to infinity. Phi hat K E to the two pi IKX. Boom, got it, great. All right? That's the Fourier series.

I can see what you're saying. I finally got a math teacher who speaks English and he's putting on an accent. All right.

Now what deal is this? All right now you showed in the homework problem that there was a relationship between the Fourier coefficients of the periodized function and the Fourier transform of the original function that you periodized. All right. You showed – we know that the Fourier coefficient, the Nth of the K Fourier coefficient, of capital phi to periodized function is the Fourier transform of the function you were periodizing, also evaluated at K. That was a homework problem.

And so let me just plug that information in now. All right? And so phi of X, its Fourier series, K equals minus infinity to infinity, it's the Fourier transform of the function you were periodizing times E to the two pi IKX. All right? That's on the one hand. That's the Fourier series for capital phi. On the other hand, phi was given actually as the periodization of little phi, so that's the sum from K equals infinity to infinity of phi of X minus K. Right?

I have two expressions for the same thing. This comes up a lot in this course, two expressions for the – if you have two expressions for the same thing you've got a lot of power over them, and the simplest thing to do here is just to evaluate at zero. Evaluate at X equals zero. All right?

On the one hand, phi of zero, if I use the Fourier series expression, if I plug in X equals zero here I'm just adding up a bunch of ones times the Fourier transform. Phi of zero is the sum from K equals minus infinity to infinity of F of phi K. On the other hand, phi of zero, if I use the other expression, K equals minus infinity to infinity is phi of zero minus K, so phi of minus K.

Well if I sum from – write this expression a little bit more neatly. If I sum from minus to infinity to infinity, summing over – summing taking sum from minus infinity to infinity, phi of minus K is the same thing as summing K equals minus infinity to infinity of phi of K. All right? Where do we start? Where do we finish? What did I do? What did I do? What did I do? I evaluated the same expression at two different – at the value zero; I get two different expressions for the same thing. That is to say, I have proved my Poisson summation formula.

So the sum of the Fourier coefficient, Fourier transform – excuse me, value of the Fourier transform, is the same thing as the sum of the values of the function at the integer points. That deserves a couple of exclamation points, at least two. Okay? Wonderful. Wonderful.

Now let's use this to find the Fourier transform of the Shaw function. Let's go back to that. That's what I wanted to get. All right? So back to the Fourier transform of the Shaw function. What did we find? We found that the Fourier transform of the Shaw function evaluated at phi was equal to the Shaw function evaluated at the Fourier – or paired with the Fourier transform of phi. And again, that's the sum from K equals minus infinity to infinity, the value of the Fourier transform of phi at the integer points.

But now I invoke the Poisson summation formula. This is so exciting. This is equal to the sum from K equals minus infinity to infinity of phi of K. But now remember, that this in turn is just the Shaw function itself evaluated or paired with phi. Where do we start? Where do we finish? The Fourier transform of the Shaw function paired with phi is the same as the Shaw function paired with phi. What is the conclusion?

The conclusion is the Fourier transform of the Shaw function is the Shaw function, three exclamation points, although maybe I should put two because it's entirely equivalent to the Poisson summation formula. Entirely equivalent, okay?

This is something that most, that many electrical engineers know. All right? This is equivalent to the Poisson summation formula. This is equivalent to the deepest fact about the integers that is known. All right? And every electrical engineer who takes a sophomore level course on signals and systems probably knows this, although they may not know that they know it. So again, it's a little surprising here, right? It's very surprising. Here's the pictorially. I mean here's the original Shaw function spaced of –

delta function spacing of the integers: minus one, two, minus two. That's the Shaw function, all right?

I take the Fourier transform I get the same thing. It's a bunch of deltas again. Not just the sum of complex exponentials or rather it is a sum of – it is an infinite sum of complex exponentials, but that's infinite sum of complex exponentials, as a distribution, is actually the sum of deltas again. So let me put it horizontally. Take the Fourier transform I get the same thing: zero, one, minus one, two, same thing. Okay, all right, very important fact, extremely important.

Now let's do the Shaw function with spacing P. That is gonna follow from scaling properties of the delta function. All right. Now let's do the Fourier transform of the Shaw function spacing P. So remember the Shaw function with spacing P of X is this sum: sum from K equals minus infinity to infinity delta X minus KP, all right, same idea just different spacing.

Okay, so let me work with the individual delta, show you what happens here. So look at delta of X minus KP. What I'm gonna do is factor a P out of this thing, all right? So I'll write this as delta of P times X over P minus K. Now you remember what happens is the scaling – an important scaling property of a delta function, this is – do I have this right. This is then one over P times the delta function at X over P minus K. Okay? Okay, did I get that right? Yes.

All right, so in other words, the Shaw function with spacing P at X is the same thing as the sum of these shifted delta functions. So it's the same thing as one over P times the sum from K equals minus infinity to infinity – well, it's the Shaw function with spacing one evaluated at X over P. All right, make sure you – I skipped a couple of steps there.

Shaw function with spacing P, that's the sum of these shifted delta functions. That's the same thing as one over P times the delta – so it'll be a sum of a bunch of delta functions with spacing one apart but not evaluated at X, evaluated at X over P. Okay? So it's one over P Shaw function with spacing X over P. Do I have that right again? I think so.

So now if I take the Fourier transform of a Shaw function with spacing P, that's the same thing as one over P, the Fourier transform of the Shaw function at X over P. It's like a scaled version of the Shaw function. Okay? All right so that is one over P – the – and I'll use the scaling theorem. The Fourier transform of X – the Fourier transform of this, right, is gonna be one over P. The scaling factor comes out with a reciprocal, so it's one over one over P, so it's times P times the Fourier transform of the Shaw function evaluated at the reciprocal of the scaling factor, so it'll be P times X. Okay?

The Fourier transform of the Shaw function itself is the Shaw function. When there's no spacing in there, when everything is spaced one apart, this is the Shaw function then at PX. All right, one more step; we are almost there.

What is the Shaw function at PX? That will be the sum from K equals minus infinity to infinity delta of PX minus K. So I'm gonna use the same scaling property of the delta function. That is, I write this as delta PX minus K, I factor out the P. That is, I write this as sum K equals minus infinity to infinity delta of P times X minus K over P. Okay? And pull that P out of the delta function. I pull that P out of the delta function. I get one over P sum from K equals minus infinity to infinity of the Shaw function at X minus K over P. Okay?

That's just using the scaling property of the delta function because it's delta P times something here, so it's one over P times the delta of the something. So it's one over P times – and I'm adding them all up. One over P times – oops, delta, delta, delta, delta. All right. All right? That is the Shaw function of spacing one over P. Right? These points are spaced one over P apart. In other words, this is one over P times the Shaw function with spacing one over P evaluated at X.

All right once again, where do we start, where do we finish? If you follow manipulations, what we showed was a very attractive formula. Thus, the Fourier transform of the Shaw function with spacing P is one over P times the Shaw function with spacing one over P. That is another very important formula; it gets two exclamation points itself and a couple underlines. Works for P equals one. For P equals one it reduces the case we had before. All right?

This is another wonderful illustration of the reciprocal relationship between the domain and the Fourier transform domain, the original domain and the original Fourier transform domain. If the spacing in the original domain is P, then the spacing in the Fourier transform domain is one over P. And they're also scaled; the height is also scaled by one over P.

So again, here's the Shaw function in the original domain, say, with spacing zero, P, minus P, two P, minus two P, and so on, going on to infinity. I take the Fourier – this is Shaw P. I take the Fourier transform and the spacing is one over P apart: zero, one over P, minus one over P, two over P, minus two over P, and so on. And the height is also scaled by one over P. Okay, extremely important.

Let's get back to our crystal, all right? Now let's get back to the crystal. All right? All of this was to help us get our Nobel Prize. There's a lot at stake. And so back to the crystal, the spacing of the atom, the electron density distribution is described by rhos of P of X is rho of X convolved with the Shaw function with spacing P. All right? That is the periodized version of the electron density distribution for the crystal. All right? The atoms are spaced P apart.

Now you do your diffraction experiment, all right? You shine your x-rays through this crystal. Hold your breath, good; get an exposure on the film. What you see on your film is, essentially, the Fourier transform. All right? You see the Fourier transform of this. So that's the Fourier transform of rho convolved with Shaw sub P. And what is that? That's the product of the Fourier transform of rho and the Fourier transform of the Shaw function with spacing P.

The Shaw function with spacing phi has a Fourier transform spaced one over P. That's the Fourier transform of rho times one over P the Shaw function at one over P, one over P. All right, let's write this out one more step. Okay? Let me write it – let me write this out with the variables in there.

Because this also involves the sampling property of the Shaw function – of the delta function, so this is F of rho of X times the Shaw function. So this is sum K equals minus infinity to infinity, there's one over P out in front of the whole thing, delta X minus K over P. All right? Now remember what happens if you multiply a function times the delta function. All right? That's one over P sum, I'll bring it inside I won't skip a step, minus infinity to infinity, the Fourier transform of rho of X times delta X minus K over P.

Multiply the sampling for the – yeah. The sampling property of the delta function is if I multiply a function times delta that evaluates the function at K over P times the delta. That is, this is equal to the sum from minus infinity to infinity, the Fourier transform of rho, one over P times the whole thing, evaluated at K over P times the delta function X minus K over P.

All right, we're almost done. What do you see in your – so this is what you see in your diffraction experiment. All right? You see this, now what are you really seeing in your diffraction experiment?

You are seeing a bunch of spots. Imagine these things – and these are impulses at spaced – at points K over P, zero, one over P, two over P, three over P, and so on, minus one over P, minus two over P. They have this intensity also scaled by one over P. So you are seeing in your picture, you see a bunch of spots of intensity, the value of – density value evaluated there and then spaced at one over P part, zero, one over P, two over P, minus one over P, minus two over P, and so on. Okay.

You had to know this if you want to claim your Nobel Prize. Why? Because you might think that you do a diffraction experiment, the spacing of the spots is proportional to the spacing of the atoms. All right, the atoms are spaced P apart. You think you do an experiment, all right, the spacing, what I – I see a bunch of spots. That must be proportional to the spacing. But it's not. It's proportional to the reciprocal of the spacing, all right?

Nature is taking a Fourier transform for you. And what you see, is you see a bunch of spots spaced on one over P apart, so you make this measurement. And then you say that's the reciprocal of the spacing in the atom, in the crystal. That's the reciprocal of the spacing in the crystal. All right? You have to know your math. If you don't know your math, kiss your Nobel Prize goodbye. Okay?

All right, that's it for today. Wasn't that exciting? And on Wednesday, we're gonna make a different use of this and talk about sampling an interpolation. Okay, thank you very much.

[End of Audio]

Duration: 53 minutes

TheFourierTransformAndItsApplications-Lecture17

Instructor (Brad Osgood):Is the screen fading and yes, Happy Halloween to everybody. Only one noble soul here came dressed as a Viking. All right. All right. I'm glad to see that sort of pioneer spirit is still with us, as ridiculous as it looks. All right. Okay. Let me call your attention to the information on the board. Mid-term exam is today. There are three sessions, from 2:00 to 3:30 and from 4:00 to 5:30 are both in building 380, room 380W. That's the basement of the math corner, and from 6:00 to 7:30 is here. Okay. We will provide you with the exam, with a formula sheet, which you already have seen as has been posted on the course website for some time, and blue books. All right. Any questions. It's an open book, open notes. You can bring your stuff with you, but as I say, don't bring these stacks of signals and systems book as I've seen in the past. You're only going to waste your time. Any questions about anything? I'm somehow not surprised to see a thin crowd out there today.

It's a shame because, actually, today, I'm gonna be talking about sampling and interpolation, which is my favorite topic in the course. It's really too bad that no one is going to be here to hear it. But I'll persevere somehow. All right. So no other questions about the exam or anything else on anybody's mind? All right. So let me remind you, again, the topic today, and we'll pick it up again next time, is sampling and interpolation. There's a lot to say and, as with many things, we can't say it all, so I want to get the main ideas across and the main phenomena's that are associated with it. Our approach to it is gonna be an application of what we talked about last time with the Shaw function so I want to remind you of that. So last time we introduced this train of delta functions, sometimes also called the Dirac comb or the Dirac train or a pulse train or an impulse train or all sorts of things. We introduced the Shaw function of spacing P some from – I'll put the variable in there, so some of delta functions evenly spaced, P apart. So K from minus infinity to infinity and delta X minus K P. All right. And it has several important properties that I will list.

So the picture is a bunch of evenly spaced delta function all up and down the line. So it's usually indicated something like this, P 2 P and so on minus P minus 2 P also and so on. All right. There's three very important – and we introduced this in the context of an important physical problem and quite an interesting physical problem that of studying X-ray diffraction. All right. The mathematical properties that allowed us to analyze that problem so effectively, are the same mathematical properties that we're gonna use today in a quite different setting, and I want to recall those for you. The three important properties – one is the sampling property, we used each one of these, but now I want to single them out. That is if I take a function and multiple it by the Shaw function, it samples the function at the points K P. So F of X times a Shaw function is the sum of F of K P delta X minus KP against [inaudible] infinity. I'll get out of the way in just a second here. All right. That sampling property to the Shaw function falls from the sampling property of the delta functions. If you multiple the shifted delta by a function, it pulls out the value at the point times the delta function. You can read that well enough here. K times P. So that's the sampling property. Allied with that, sort of the flip side of that, is the periodici, the periodizing property. And that has to do with convulsion that if I involve a function with the Shaw function of spacing P, I get a periodic function of period P. That is this is the sum of shifted versions of F, K going minus infinity to infinity, F of X minus K P. All right. So that gives you a periodic function of period P. I'm not worried about here of questions about conversions and things like that, I'm just worried about the formal properties and how they work.

They are, in some sense, flip sides of each other and we'll see that very strongly today because convulsion and multiplication are sort of swapped back and forth by taking the Fourier Transform of the Inverse Fourier Transform for you. Okay. So in fact, there's actually sort of two sides of the same coin. The final property of the Shaw function, the remarkable property that falls from this [inaudible] formula, the fact about the integers is the Fourier Transform property. That is the Fourier Transform of a Shaw Function spacing P is a Shaw function with reciprocal spacing with an extra factor of one over P out in front. Okay. That's a very deep fact. Let me just say because I'm actually gonna use the corresponding formula for the Inverse Fourier Transform, so Inverse Fourier Transform formula is the same. I'll just write it down. I won't derive it separately. The Inverse Fourier Transform of the Shaw function of spacing P is the same thing, one over P – Shaw function spacing one over P. All right. These three properties were the basis for the mathematical analysis of X-ray – the fractions we talked about last time, and they're gonna serve us also today in quite a different context. So here is the set up for the sampling and interpolation problem. In fact, let me call it the interpolation problem. Set up for the interpolation problem. It is not one problem, but rather it's a whole category general problems which have been considered in different contexts, different ways since time and memorial. Okay. All right. The problem here is – and what we're actually gonna solve in a quite remarkable way is the exact interpolation of values of a function from a discrete set of measurements or a discrete set of samples. So we're gonna consider – this is what we'll do. We'll be able to interpolate all values of a signal or a function from a discrete set of samples. All right. We won't be able to do it in this generality, that is, we're gonna have to make certain assumptions that are gonna allow us to solve it, but even under fairly general assumptions, it's surprising that you can do this at all, and it's ridiculous to expect to be able to do it at worse somehow. To ask us a general question and to expect an answer for it is really maybe asking too much, but in fact, by employing the techniques that we developed, we can actually get an answer for this in fairly great generality and that's extremely important practically. All right. Here's the set up. Here's the thing to keep in mind. Let me formulate this question again, fairly generally and then we'll get to the assumptions that we have to make in order to be able to solve it in a reasonable way. All right. So imagine you have a process evolving time say, it doesn't have to be time, but that's the easiest way to think about it.

A process evolving in time. All right. And you make a set of measurements at equal intervals – some fraction of a second. Time intervals. Equal space actually turns out to be important in this context. All right. So you have a bunch of measurements. T not, Y not, T 1, Y 1, T 2, Y 2 and so on. All right. So you can think of this as a bunch of points, and I want to formulate the question really in two ways. So here's T not, T 1, T 2 and so on. And here's the corresponding points somehow in the plain, or just plotting them, T 3. Now, you might ask – fit a curve to those points or what is equivalent – interpolate the values at intermediate points based on the measurements that you make. There's really two equivalent ways of formulating, so you might ask, one fit a curve to the points, fit a curve to the data or you might ask interpolate values of the function of the process, whatever it is, the function at intermediate points based on the sample values, based on the points you measure based on the measurements. All right. Those are reasonable – they are questions that are natural to ask and you can imagine all sorts of context for this. And there's many ways of doing it. All right. As a matter of fact, if you're doing 263, you'd probably talk about lease squares, approximations to a set of data.

We actually don't track the data, but you find somehow the line that best fits the data. But you can imagine drawing a curve like this, you can imagine doing it with straight lines, straight line – I was going to say straight line interpolation, but I'm mixing the two aspects of the problem. But something like this. You could do this or you can imagine fitting it with a polynomial or some other kind of curve where you might go up, you might go down, you might go up again somehow, but you want to actually track the values exactly. Okay. But there again, there's not a unique solution for this. There's many ways of doing it depending on the particular problem. Now, and again, it's almost silly to prefer – well, you may have a reason for preferring one over the other, but you better have an extra argument to justify one choice over another. There's many ways of doing this. Many possibilities. All right. And the question is how would you choose one over the other or choose any at all for that matter. Why would you expect to be able to do this? Well, you might want to refine, if it's possible, you might want to refine your options or narrow your options by making more measurements. All right. If possible, you might make more measurements to suggest a better fit or better interpolation or more accurate interpolation. A better fit or more accurate interpolation. Let me make sure you know what I mean here when I say interpolation, all right, and what you're actually getting. If you write down an explicit function or set of functions that tracks this data – and what I mean by interpolation is that you can find the valuate intermediate points, the reported value of the intermediate points, by plugging into your formula. All right. So you write down a formula for the straight line that goes from here to here, that means you can find every point on that straight line. You have a formula for that.

Why question the straight line from here to here and from here to here or if you have a polynomial that somehow does this or maybe does a piece wise things and maybe goes that then you can find the value at any intermediate point by plugging into that formula. That's what I mean by interpolation. All right. So all values are computable by knowing only these discrete values, and it's an approximation. You don't know whether that's doing a good job or not doing a good job, but you're making a guess and you're making what you hope is a reasonable guess. All right. But that's why I'm saying curve fitting and interpolation or equivalent sort of thing. One is a geometric way of looking at it by drawing a curve, the other is an analytic way of formulating it by actually trying to write down a formula for the function that's doing the interpolation, and then plug into that formula and seeing how well it matches. All right. So again, what I mean by this is maybe you can make more measurements and then compare those measurements to your formula. All right. See if it's working. Now, maybe you can do that, maybe you can't. What is the enemy in this or what is causing uncertainty? Well, again, if this level of generality there's more uncertainty than there is certainty. I mean, you know, you have a discrete set of values, who knows what the function is doing in between, but don't stop at saying that. Try to say something a little bit more positive. That is, the thing that is causing you difficulty, the uncertainty in the values, in the extreme case, are the oscillations. How fast the function is changing from one value to the other. The more rapidly the function might be changing, the less certainty you have about interpolating the values in between from the measured values. All right. So the more bends the function takes, the more rapid bends or the more corners a function has, the more uncertainty you have in your interpolations – in your current fitting or your interpolation. Same thing. The problem here in interpolation, in uncertainty is the more rapid the bends are, the more uncertain you are in your interpolation, the more jagged somehow the process is between sample points, the less certain you are about how to interpolate between the sample points. All right. So you somehow want to quantify the jaggedness or the bends in the process – in the function. Now, we are pretty sophisticated in questions like that. We want to regulate, maybe even get rid, but at least regulate understand quantify somehow. We want to regulate – let me put it this way – how rapidly the function, the signal is bending or oscillating – I don't quite know the right way of saying it or the right word for this, but I hope you know what I mean – oscillating – all right – between sample values. That's gonna be an extra assumption, all right, between sample measurements, between sample values. All right. Now, this is gonna take the form of an assumption, but the question is what should the nature of the assumption be? All right. Now, as I say, we've actually gotten quite sophisticated in this sort of thing. For us, we think in terms of not just the function given in the time domain, but we also think of the function in the frequency domain. We think in terms of its Fourier Transform. The Fourier Transform in representing the function in terms of its frequency components tells us something about how fast the function is oscillating.

If it has high frequencies, high frequencies are just as Fourier Transform high frequencies are associated with rapid oscillations, low frequencies with smaller oscillations, less rapid oscillations. And so if we want to make an assumption that's gonna eliminate rapid oscillations, we might make that assumption in the frequency domain, that is, it should be an assumption on the Fourier Transform. So this is governed by – this is a spectral property so it's governed by the Fourier Transform. All right. That is an assumption on how rapidly the function is oscillating, is an assumption on the Fourier Transform, that's one way of saying this a little bit more precisely. All right. Now, go right to – past the simplest assumption you can make along these lines. It takes a little experience to know this is the simplest assumption, but the idea is you want to eliminate all of – one way to think about this is you want to eliminate all frequencies beyond a certain point. All right. So that's one possibility. You're trying to analyze this. Maybe this one is good idea. Maybe there's other ones that are good ideas. I'll give you a clue. This turns out to be a good idea. One possibility is to eliminate, that is to assume, is to eliminate all frequencies beyond a certain point that is by an assumption. All right. Now, we've seen that many times. If this is what you want to get to, then make that a definition. Concentrate your attention for the interpolation problem on functions which satisfy a property like this. We are ready for an importation definition. The enemy – the problem is rapid oscillation between those measured values. Our approach to that problem to resolve that uncertainty is to say I'm gonna regulate how rapidly the function is allowed to oscillate.

And I'm gonna phrase that as an assumption on the Fourier Transform of the function, all right, because the assumption on the Fourier Transform is an assumption on the frequency components. If you eliminate high frequencies you feel like you're eliminating rapid oscillations, and I'm gonna state that as a definition. So you state a function, F of T, is band limited. If it's Fourier Transform, is identically zero outside sum, band of frequencies. That is F of S – the Fourier Transform is identically zero for S greater or equal to P over 2. I'm gonna write it like that. All right. You'll see why in a little bit. For some P. And then, the smaller such P is called the bandwidth. The smallest such P for which this is true the bandwidth. All right. So the picture is – but the Fourier Transform is identically zero outside finite interval. And I'm working with real functions here, so the Fourier Transform is symmetric so that's why I have absolute values. You can give the definition more generally, but this is the most natural setting for it. So here's the picture. He's zero, here's P over 2, here's minus P over 2 and whatever the Fourier Transform does in between, and it can't do that exactly, because it's supposed to be symmetric – it's zero outside of there. All right. This is the Fourier Transform. All right. So that is supposed to capture the idea that you are eliminating the size of the frequencies, you are limiting, in the time domain, the oscillation of the function by making this assumption in the frequency domain. Okay. Now, what I want to do is I want to show you that for band limited signals you can save the interpolation problem exactly. It's remarkable. For band limited signals, you can solve interpolation problem exactly. You get a formula for F of T for all T for all T. In terms of values of F at a discrete set of points. T – let me just call them T of C K sample values. All right. You can fit that curve exactly. That is to say, if you know the process comes from a function which is band limited, you can write down a formula for the function. All right.

Now, in the notes, there's a different sort of discussion of this. I'm gonna go right for the kill. All right. I'm gonna show you how this works and it's nothing short of remarkable. It's almost obscene the way this thing works, I think. It's the most remarkable – I think it's the most remarkable argument in the whole subject practically, and as I say, it's one of my favorite arguments because it's just obscene and it makes me feel cheap and dirty. It's great. The notes has a discussion of this, but it goes a little bit farther. I'm not gonna go through that. That is in terms of trigger metric functions, why you expect something like this to occur, why you might expect something like this to occur in different circumstances. We'll circle back and talk about some of these things, but for right now, I want to go, as I say, right for the kill. I'm gonna show you how this works. I'm gonna give you the argument. – and in evolves exactly – the periodizing property of the Shaw function and the sampling property of the Shaw function and the Fourier Transform part of the Shaw function. Those three properties come into this in a very essential way. All right. Again, here's the pictures of the frequency. I'm just gonna do this. All right. And there's no way of saying of it other than I'm going right for the kill. Just enjoy the ride. All right. Enjoy the experience because it's amazing to see this thing unfold. All right. So here's the picture of the frequency domain again. The pictures like this. So I'm assuming the signal is band limited so that means Fourier Transform only is non-zero between minus P over 2 and it might have some zeros in here, okay, all right, I'm not saying it doesn't have any zeros in between, but I'm saying that for sure, it's identically zero outside this interval. All right. Now, I'm gonna use the Shaw function with spacing P to periodize this, all right, by convulsion. That is, I look at – let me draw the picture for you. I look at the Fourier Transform of F convolved with a Shaw function of spacing P. All right.

So what is the picture? The picture is here's the Fourier Transform minus P over 2 to P to over 2. I'm gonna make it a little bit more condensed here. Minus P over 2 to P over 2. All right. Here's the Shaw function with spacing P. There's a delta function there. This is P over 2. The next delta function is a P. The delta function over here is a minus P and so on and so on. All right. Here's zero. All right. Now, what does it do when you convolve the Shaw function with this, it shifts it by P and adds it all up. All right. So the picture of the convolution of the Fourier Transform with the Shaw function looks something like this. Here's zero, here's minus P over 2, P over 2, it shifts the whole thing over to P down to minus P and so on. So you get just a bunch of repeating patterns of the thing. But there's no overlap because the Fourier Transform is zero outside of the interval for minus P over 2 to P over 2, so if you shift it by P, there is no overlap when you convolve. All right. Now you say, brilliant, you have done something and you have undone something. This takes a PhD? Yeah.

Student:[Inaudible]

Instructor (Brad Osgood):Where? No, this should not be convolution, it should be multiplication. I'm multiplying by a function which is one on the interval from minus P over 2 to P over 2 and zero outside that interval. So that gets back the original Fourier Transform. All right. The original Fourier Transform is here and it's only there. All right. It's only there. When you convolve it, you get a bunch of repeating patterns. You add them all up. But you cut those off. All right. Cut those off. And that leaves you with this, which is the original Fourier Transform. This is the whole ballgame. All right. The most important equation here is exactly this equation. I'll write it down again. The Fourier Transform F is the cutoff of the P Fourier Transform.

All right. And you say great. You've done something, you've undone something. You know, you got a PhD for this; this is why they call you Professor? Now, take the Inverse Fourier Transform. It looks like you haven't done anything, but actually, you've done something extremely significant because by taking the Inverse Fourier Transform, it swaps multiplication and convulsion. This is a picture in the frequency domain. What is happening in the time domain? So take the Inverse Fourier Transform. Well, on the left-hand side the Fourier Transform of the Fourier Transform is just F, so F is equal to the Inverse Fourier Transform or the Fourier Transform. Right? That's fine. Now, take the Fourier Transform, that gives you back the original function because I'm taking the Inverse Fourier Transform and the Fourier Transform. Right. The Inverse Fourier Transform of the right-hand side, again, it swaps, let's do this a little bit at a time here. It swaps convolution and multiplication. That is to say it takes multiplication back to convolution. This is the Inverse Fourier Transform or the rectangle function; convolve with the Inverse Fourier Transform of this. Follow along.

Enjoy the ride. Enjoy the ride. All right. All right. Let's look at this part here. Well, I'll write it out a little bit. One step at a time. The Inverse Fourier Transform of the rect function is a scaled sinc function. The Inverse Fourier Transform of the rect function of spacing P is sinc P T. If you don't believe me, figure it out yourself. All right. The Inverse Fourier Transform of this, again, is gonna swap convolution and multiplication. The Inverse Fourier Transform of involved with the Shaw function is gonna be the Inverse Fourier Transform of the Fourier Transform. What the heck. Too many Fs of the Fourier Transform of F times the Inverse Fourier Transform of the Shaw function times. Okay. I don't know if we stated the convulsion theory – I don't know if I stated it, I'm sure it's in the notes, we stated the convolution theorem in terms of the Fourier Transform, same thing holds in terms of the Inverse Fourier Transform. Okay. Swap convolution and multiplication. All right. Now, once again, the Inverse Fourier Transform of the Fourier Transform is just the function. The Inverse Fourier Transform of the Shaw function of spacing P is a Shaw function of spacing one over P. So this is F of T times one over P times the Shaw function of spacing one over P times FT. I guess I'm using T as my variable here. Okay. Now, I could combine all the formulas at once here, but let me not do that. Let me take this one step further. Now, remember now, we're gonna use the sampling property of the Shaw function. We used the periodizing property of the Shaw function in the frequency domain. When we periodize the Fourier Transform. Now, in the time domain, I'm gonna use the sampling property of the Shaw function.

All right. If I multiple F and T times this, now remember, I'll take this out one more step before I don't know remember what it is, one over P, that's a constant that just comes out in front of the whole thing. It's one over P times F of T times the sum of deltas minus infinity to infinity spaced one over P apart, T minus K over P. All right. Now, I use the shifting property of the delta function so it's this Fourier Transform function, P times time P T convolve with this sum of deltas. All right. So I'll take this out one more. The P cancels with a one over P, so this is sum K equals minus infinity to infinity. These are constants. F of K over P times a sinc function, the scaled sinc function convolved with delta T minus K over P. All right. And now you use the shifting property of the delta function and write this as the sum of the sample values and it is the sum from K [inaudible] minus infinity to infinity, F of K over P times the sinc of P, T minus K over P. Which is actually the way I prefer to write this, but many people write it like this – they multiple by through by P. Some from K equals minus infinity to infinity, F of K over P sinc of P T minus K. We're done. We're done. We found a formula for the function at all values of T in terms of its sample values. I'll write it down. This was a long chain of equalities here. We have shown that F of T is equal to the sum from minus infinity to infinity, F of K over P sinc of P – I'll write it like this, T minus K over P. We have interpolated all of the values of the function, at all values of T, in terms of sample values, evenly spaced points. All right. So my T Ks, in the way I originally formulated the problem, are one over P, two over P, three over P, minus one over P, minus two over P, minus three over P and so on and so on. If you know all these discrete values, then you can interpolate all the values of the function in terms of those. I think it's the most remarkable formula in the whole subject. So again, the assumption was – I want to go through this one more time to make sure you have this. This is called the sampling theorem, the sampling formula, sometimes called the Shannon sampling formula, sometimes call the Whitaker sampling formula. It's associated with a lot of names. Unfortunately, my name is not associated with this formula as you can tell how much I love it. All right. So again, the assumption is that if the Fourier Transform of F is zero, for S greater than or equal to P over two than you have this formula for the function, sum from minus infinity to infinity, F of K over P, sinc P times T minus K over P. All right.

Now, for me, the sampling formula is identical with the proof of the sampling formula. All right. And that's important because I think next time we're gonna talk about – I don't think I'm gonna push this into this time – we'll talk about when things go wrong, as well as when things go right, but for me, I never think of this formula alone in isolation without the miracle equation that makes it work. All right. So this is all depending on the fact that if you periodize the Fourier Transform and then cutoff, you get the Fourier Transform back. But this is the rectangle function of spacing of with P times the Fourier Transform of F convolved with the Shaw function of the spacing of T. All right. It followed, although I tried to stretch it out for dramatic purposes, it follows immediately, lightening fast, just very mechanically by applying the Fourier Transform from this formula. This is what's essential. All right. And so for me, and I'm serious about this, the sampling formula, this formula, is identical with the proof of the sampling formula. All right. To understand this formula is to understand this formula. They are the same. Okay. They are the same. It's a consequence of the Fourier Transform swapping convolution and multiplication and writing this down. It's a simple, but exceedingly cunning idea. Why do something like that? You know. It seems like you're not gonna do anything. You're periodizing and then cutting off and you get back the original function. It's like proving one is equal to one. What is the big deal? You know. But the fact is that the reason why something non trivial happens is because of this remarkable fact that the Fourier Transform is going back and forth between the two domains, swaps convolution and multiplication. You have a combination of both of them in fact. You have convolution and multiplication in the frequency domain, you take the Inverse Fourier Transform, you have convolution and multiplication in the time domain, but the roles are swapped. All right. That's why periodizing in the frequency domain turned into sampling in the time domain. All right. Because of this formula and because of the way the Fourier Transform swaps convolution and multiplication. All right. I'll tell you what. Since it's been my habit to keep people late in class, I think instead, today, we'll get out early. I want to more about the consequences of this remarkable formula next time, including, sad to say, when things go wrong, but when things go wrong, they can all be explained by when this formula is not correct, when something is not right with this formula. All right. Okay. That's it for today. I'll see everybody later at various times and –

[End of Audio]

Duration: 43 minutes

TheFourierTransformAndItsApplications-Lecture18

Instructor (Brad Osgood):And then – all right. We still friends?

Student:[Inaudible].

Instructor (Brad Osgood):Actually I have not – to be hon – I haven't started grading the exam yet, so I don't how things are gonna look. So you'll get it back some time next week. All right. And I'll also post the solutions. But I have nothing more to say right now. Okay. It's a tough week for everybody. I realize that. I mean there are a lot of exam – 263 had the exam the previous weekend. We had our exam – was it 216 also an exam or something like that, and – or one other course, and then 278's having its exam – when's that – when is that, next week?

Student: Tuesday.

Instructor (Brad Osgood):Tuesday.

Student: Should be.

Instructor (Brad Osgood):It's brutal. What can I say? Better you than me. I don't know – sorry. It's pretty [Audio breaks up] yeah.

All right. I want to spend a little more time today [Audio breaks up] theorem – sampling, and interpolation, and some of the phenomena that was associated with it. I'll remind you what the setup was from last time. And I'll almost carry out the proof again – or at least I'll give you the setup for the proof again because as I said many times in this – last time, and as I will say again today, for me the sampling [Audio breaks up] the derivation of the – the sampling theorem, the formula is identical with a proof of the sampling formula. I think the two are so closely related that to understand one you really have to think in terms of the other. All right.

So here's the setup. We look at a signal which is bandlimited. That is to say, its Fourier transform is identically zero from a certain point on. All right. So suppose – say F of T is a signal. F of T is a bandlimited signal. So I'll remind you again what that means. This was the key definition really that got the whole thing rolling. So it means that the Fourier transform is a condition on the [Audio breaks up] signal. The Fourier transform is identically zero say for S bigger than equal to sum P over two [Audio breaks up] – I'm writing it. All right.

I'm writing as [Audio breaks up] with respect to the or [Audio breaks up] and there's [Audio breaks up] such P is called the bandwidth. There may be many P's for which that is true, but the smallest one is called the band [Audio breaks up] the small such P is called the bandwidth. Bandwidth is sort of a ubiquitous [Audio breaks up] in signal processing, but I think this is – I don't know if this is the first definition of it, but I think this is the definition that's most commonly accepted. It's the smallest – you're assuming the Fourier transform is zero – identically zero from some point on, and the smallest number for which that's true is called the bandwidth.

So the picture is this. Here's zero. Here's minus P over two. Here's plus P over two. And from there on out going up to plus infinity/minus infinity the Fourier transform is identically zero. Now, remember you can't really plug a Fourier transform because the Fourier transform is complex, so – and the Fourier transform, if it's a real signal, is symmetric, and all the rest of that jazz. So this is not to be taken literally as a picture, but it's something [Audio breaks up] mind. That's the general picture you should keep in [Audio breaks up] next step [Audio breaks up] what is [Audio breaks up] the clever key idea.

You periodize the Fourier transform – FFS by comb [Audio breaks up] by [Audio breaks up] that spacing. All right. That is you look at this. The Fourier transform of F convolved with the Shah function. And what that does is because it's identically zero outside going from P over two to infinity minus P over two minus infinity, this just shifts it, and adds it up, and you get a bunch of copies of the original signal – non-overlapping copies of the original signal. So here's zero. Here is P. Here is minus P. Here is, again, P over two – I'll get out of the way in just a second. Here is minus P over two, and so on, and so on. All right. So this is a picture of the periodized version of – hello? Is something wrong?

Student: This one's cutting out.

Instructor (Brad Osgood):Oh, it's cutting out? I wonder if it's because I dropped it. Okay. Back to live action.

And then the main trick is to cut off. Then you cut off by the rectangle function of width P, and that gives you your original Fourier transform back cut off by multiplying by the rectangle function of width P. And that gives back the original Fourier transform because nothing is overlapping there because you periodized, and [inaudible] cut off.

So again, the picture's like this roughly. Here's zero. Here's minus P over two. Here's plus P over two. You multiply by the rectangle function of width one. Remember that's one between minus P over two to P over two, so when you multiply that doesn't do anything in that interval, and it sets everything to zero outside that interval. So the result is the Fourier transform of F is equal to the rectangle function times the Fourier transform of F convolved with the Shah function. This is the basic equation. Everything depends on this.

It looks like – as I say, you've done something, and then undone it, so how could it possibly lead to anything interesting, much less nontrivial. But it's the whole ballgame because now what you do is you take the inverse Fourier transform, and that swaps multiplication and convolution. And you have two instances of each here. You have an instance of each, I should say. It's multiplication here, so an inverse Fourier transform that's gonna turn that multiplication into convolution. The convolution in here – when you take the inverse Fourier transform, it's gonna turn that into multiplication. Sampling comes from this part, and the sinc part function comes from this part, and you turn the crank, and this is what you get.

So you take the inverse Fourier transform, and the turn the crank. Inverse Fourier transform on the left-hand side is just the function again. The inverse Fourier that's transformed on the right-hand side leads to the amazing formula. You get F of T is the sum from minus infinity to infinity F of K over P times the sinc of P T minus K over P. Okay. I managed to fit it all on the board. Little squeeze there – sorry. All right.

The points K over P are the sample points. They are in the time domain of course – or the sample points, and they're spaced one over P apart. All right. And what that formula does is it interpolates all values of the function in terms of the sample values. All right. So you think of these as the values that you measure say. And the question is, "Can you figure out what the values are in between in terms of the measured values?" And under the assumptions, and under the – by virtue of the derivation, yes, you can do that. All right.

So the – and the shifted sinc functions – say sinc of P T minus K over P. Sometimes you might call those the interpolating functions – something like that. Just – let me just – all right. They don't – they're independent of function F. All right. The only value – the only time F enters into the equation is through its sample values. So it interpolates all values of the function in terms of the sample values. It's a remarkable formula. It's an amazing formula. And it all depends on this. Okay.

That's why I say – and I will continue to say – that for me the sampling formula is identical with the derivation of the sampling formula. The reason why I say that, we'll see in a little bit, is because there are occasions when it goes wrong or it's misapplied or applied differently. And to understand that situation, you have to understand the derivation. Then – if you do, it's really quite straightforward. Okay.

Now, there are couple of terms that – some further terminology that's associated with this formula, and with this whole circle of ideas. You call – you also call the number P – so I'm calling it the bandwidth. That's the – that's an accepted term. You also call P when it's associated with a sampling formula like this. You call it the sampling rate. You call P the sampling rate. And you speak in terms of number of samples per second or speak in terms of – usually you – and because P has the units of frequency here, it's in Hertz. So it's something per second. So you usually speak in terms of samples per second, so you speak in terms of samples per second. All right.

And again, if it's however many samples per second, that means the cycle – the samples are spaced one over P or whatever the rate is apart. And you also refer to the sampling rate for which the theorem is true as the Nyquist rate. That's also sometimes referred to – it's named after Harry Nyquist, God of Sampling, who was one of the first to write this formula down in the context of telegraph signaling, I think actually. So you also call P the Nyquist rate. All right. At least that's what I would call P. All right. But we really understand it again, in terms of the bandwidth, and we understand it in terms of this formula. That's the most fundamental thing. All right.

Now, of course you can always sample at a rate higher than the sample – higher than the bandwidth. All right. That's okay. Then you are – your formula is still true. You can sample at a rate higher than P. All right.

What does that mean in terms of your formula? Or what does that mean in terms of the derivation? It's still true. So again, if here's the original Fourier transform – all right – going from minus P over two to P over two, and if you sample at a rate – I don't know – P prime bigger than that, then you are just periodizing that period bigger than P prime. You can still cut off by that. You're getting the original signal back. If you periodize by something like that, then you're getting something that is even – has less overlap so to speak. Looks like that. All right.

And it's still true that if you cut off you get the signal back again. All right. It still true if you cut off by the Fourier – by the corresponding rectangle function, you get the signal back again, and you can turn all the crank. That's fine. You're doing more work than you have to. That means the sample points are spaced closer together, so it's still true – go over here. Say all is okay – all is well with the rest of the derivation. You're getting more sample points than you need. The sample points are spaced closer together, and you're getting more of them so to speak, so you're working harder than you need to. The sampling rate is higher being the sampling points are spaced closer together.

Now, the real question is – and I'll come back to it in just a minute – is what happens if you sample at too low a rate? Well, this is not an interesting question so to speak. That is nothing goes wrong. Everything's okay. The question is what happens if the sample – if you take – if you sample too low a rate? All right. So we'll get back to this today.

Before doing that though I have a couple of other comments – couple other general comments about the formula. So let me write it down again. Actually it's written up there, but let me write it again.

Because there are some other interesting sort of natural phenomena that are associated with this formula, so I got to erase it, but let me write it again. So F of T is assuming that everything is all right, you've done the derivation, you know the bandwidth, you're expressing the function in terms of the sample points by this way. Now, we start off sort of in a very physical context. We said, "Suppose you have a set of measurements, and you want to fit the curve to the measurements – you want to fit a curve to the measurements. You want to pass through the points, and so on."

You will notice the sampling formula involves an infinite number of points here. Right. To interpolate the value of F by this formula involves the sum from minus infinity to infinity. They are a discreet set of points. They're a discreet set of measurements, but there an infinite number of them. Interpolation involves – depends on an infinite number of sample points. All right. So you might be quite disappointed in this, but that's the way it is. All right. So certainly for any practical application you can only use a finite sum. And that will introduce an error. All right. So – and that's another subject that we're not – really not gonna get into, but it's something to be aware of.

So a finite sum – so you need to find an approximation. And there are results that tell you a – that estimate the error, and things like that, but again, these aren't things are – those are sort of in particular applications for particular – in particular contexts, and we're not gonna talk about it. But just – it is to be aware of. So an application you'll need a finite sum in any application – in any real application. All right. So that introduces error. And you pro – so it becomes an approximation instead of an equality, and that's something you have to be aware of. That's one thing. Okay.

But actually there's a deeper – there's something act – there's something deeper going on here. This is a deeper phenomena that's related to the fact that you have an infinite number of samples here. And that is – well, let me state it as following. A signal cannot be both time limited, and bandlimited. This a – I'll write it, and then I want to – I'm not gonna – well, say – let's discuss it a little bit.

So the phenomenon is a signal cannot be both limited in time, and in frequency. All right. Now, what that means is – what I mean by this is that there's actually two parts here. All right. If the Fourier transform is identically zero for say S greater than or equal to P over two – all right. So in the situation we've been dealing with, then F of T – well, the mathematical way of saying this is it does not have compact support means that it does not – it is not identically zero at some point. All right. It's not identically zero for T sufficiently large. All right.

So if it's limited in time – if it's limited in frequency, it cannot be limited in time. I'll give you an argument for this in just a second, but let me just state the facts now. So once again, if it's in fre – a sequence cannot be limited in frequency, and limited in time. If it's limited in frequency, it cannot be limited in time. And conversely the same thing holds switching the two domains. That is if it's limited in time, it cannot be limited in frequency.

But likewise if F of T is identically zero say for T – I don't know – bigger than Q over two – something like that – just some – outside some range, then the Fourier transform of F of S is not identically zero. It does not have compact support. It can't be identically zero for S sufficiently large. All right. You can go back, and forth between the time domain, and frequency domain by Fourier transform. You'll have the same property or the related properties here in either domain. All right.

So that – in the context over here because we're assuming the signal is limited in frequency, it can't be limited in time. So you shouldn't expect to be able to interpolate with only a finite number of sample points. All right. A finite number of sample points would only go out so far on the graph, but the graph goes on forever. All right. You can't expect to interpolate values way out on the graph by only staying to part of the graph. Yeah?

Student:[Inaudible] at the post just bounded?

Instructor (Brad Osgood):Sorry?

Student: Of the Fourier transform – does everything stay compact support, and not just bounded support?

Instructor (Brad Osgood):I say – compact is sort of the technical term for – means close, and bounded. So it means it's zero outside some closed interval. That's – just for mathematical term. So even – you could think of it is – I mean I just find myself using that terminology because it's fairly standard. It's certainly standard in math, and you find – you see it more, and more in the – in nearing literature. But don't sweat it so much.

Compact means closed, and bounded. So outside a compact interval means outside some close, and bounded interval. Okay. Closed meaning it's – includes the endpoints. It's – there for tech – there are technical reasons why that's – why you have to include that – why you have to include the endpoints when you talk about things like that. All right.

So once again, let me just – to emphasize why this is not – why this has to be the way it is – why it involves an infinite number of sample points. If it's a finite extent in the – if it's a compact support in the frequency domain – if it's bandlimited, that means it can not be time limited, so the signal has to go on forever, and you can't expect to interpolate way out in the tails of the signal by only a finite number of points. All right. So it should involve an infinite number of sample points. That's just the way it is – all right – given these results.

Now, why is this true? Well, I'll give you a sort of somewhat bogus argument. There's a more precise argument that's given – more rigorous argument that's given in the notes. So let me prove this that if Fourier transform of F of S identically zero [inaudible] S bigger than or equal to P over two, then F of T goes on forever. All right.

The reason why this is true is, and it's a – I don't know –it's sort of analogous to what we did in actually deriving the sampling formula, although, with one exception. Since F – since the signal – the Fourier transform is zero after a certain point –again, here's minus P over two to P over two. Here's the Fourier transform. Then it's equal to its cut off.

The Fourier transform of F – if I cut off by multiplying by a rectangle function of that width, I haven't changed anything because the rectangle function, again, is one between minus P over two, and P over two. If I scale by P, and zero outside, so I haven't changed anything. Now, take the inverse Fourier transform on both sides. I haven't periodized here. All right. This is a different argument. I'm deriving the sampling formula. I'm proving this statement. If the Fourier transform is identically zero from a certain point on, then F of T goes on forever.

Why is that? Well, write this statement out, and then take the inverse Fourier transform. So again, it looks like I haven't done anything here because I've just written the Fourier transform in terms of the Fourier transform. I haven't done anything. But now if you take the inverse Fourier transform, the miracle that happens is it swaps convolution, and multiplication. All right. So if I take the inverse Fourier transform you get F of T is equal to the inverse Fourier transform of this which is a scaled sinc function – sinc PT convolved with F. All right. Sort of a striking statement actually. All right.

It says for that these functions the sinc function is in some sense functioning like the Delta function or functioning like an identity for convolution. All right. The sinc function won't do that for all signals, but it will do that for the bandlimited signals. All right. So F is the sinc convolved with F if the Fourier transform of F is zero beyond a certain point.

Now, here's where I'm not – I'm just gonna wave my hands, and I'm gonna say what this means is – and I think you can probably make this precise, although, there are actually better arguments for the whole circle of ideas – is that sinc function goes on forever. All right. So convolving F of T – whatever happens with F of T with a sinc function is gonna stretch out a – it is gonna smear out F so it goes on forever. It's never no – it's never identically zero beyond a certain point. And so F is never identically zero beyond a certain point.

It's a strange looking art equation, I understand. But the way that sort of – the conclusion that you draw is your convolving F with a sinc function just looking at the right-hand side. This function on the right-hand side you would say goes on forever. Does not – is not identically zero beyond a certain point. And what do you know, it's equal to F. Okay. So the right-hand side goes on forever. All right.

I don't know if you can actually make this argument completely rigorous. It's the one that's always given in electrical engineering classes. I don't like it myself, but it sort of gives you some reason for why this is true. And as I said, there's actually a more rigorous argument that's given in the notes, but I won't go through that. Okay. But nevertheless it's an important principal. All right.

Now, this is an example –think about this a little bit. If someone says – this is an example of where mathematics meets the real world, and they disagree. All right. You always – we think about mathematics as getting a pretty accurate description of the real world, and it's kind of spooky sometimes how well it works. But in the real world signals really don't go on forever. Real signals really don't go on forever. And in the real world signals really are limited in the frequency. The frequencies don't go on forever. All right.

This says that you can't have both of those. All right. You cannot have a signal which is both limited mathematically – you cannot have a signal which is both limited in frequency, and limited in time. It can't be. If it's limited in frequency, it's unlimited in time. If it's limited in time, it's unlimited in frequency. All right. But any real signal is limited in time, and limited in frequency. So it – there's a clash. All right. There's a clash between the mathematical description between the mathematical theorem, and what you actually encounter. Okay. I mean that's something to take note of. All right.

On the other hand – so that's just to say mathematical models are imperfect or the real world is imperfect, however you want to look at it. Okay. On the other hand you can also look at this in some way as like a qualitative version of the uncertainty principle. And we've seen situations somewhat like this before. That is we've seen them where they're concentrated in the time domain means spread out on the frequency domain – or constantly in the frequency domain means spread over the time domain. We've seen various instances of that, and that's a very important principle in Fourier analysis. It's very important for your intuition that realizes if something is concentrated in one domain, it's spread out in the other domain.

So here concentration is in terms of compact support. If it's identically zero beyond a certain point, it has to go on forever in the other domain. All right. The uncertainty principle in quantum mechanics is a statement about you can't simultaneously localize two physical quantities: momentum and position say. All right. And that's also was sort of shock to people's conception of the physical world, and how mathematics modeled the physical world. I mean the idea is you should be able to localize with arbitrary precision both where something is, and how fast it's going. All right. But from the point of view of quantum mechanics, that's not true. You can't do that.

And this in some says – is maybe can be viewed as related to that. All right. So maybe it's not a mathematical imperfection with the way the real world is. Maybe you have to change your mind the way the real world is. Maybe you're making it an imperfect statement when you say a real signal cannot be limited in time, and limited in – it has to be both limited in time, limited in frequency. Maybe that's really also an approximation. All right. So think about that tonight as you're lying in bed staring up into the abyss. Okay.

There is – by the way, we didn't do it. But there is in the notes actually also a proof of the uncertainty principle, and precise proof the uncertainty principle which does talk about exactly – and there are various ways of formulating it sort of in support of a function, and the support of the Fourier transform where the Fourier transform is non-zero, where the function is non-zero, and the – and relates those two in a more quantitative way. This is sort of a qualitative statement is that it can't be both local – both concentrated in time, concentrated in space or it can't be both time limited, and bandlimited. But there are ways of making these statements more quantitative. They're interesting – and they're interesting. Okay.

Now, make sure I didn't miss anything – any more profound statements here. Okay. The final thing I want to talk about – I think I told you – I mean I love this material. I think the sampling theorem, and all the material surrounding it is really fascinating, but if I had my way I'd just do it more, and more – talk about it more, and more. But I really only want to talk in a little bit more detail about one other phenomenon that's associated with sampling interpolation, and that's the phenomenon of aliasing.

That you sort of think of as the natural phenomenon associated with a misapplication of the sampling theorem or misapplication of the proof. So aliasing, and interpolation to – aliasing, and interpolation is the other comment that I wanted to – the other sort of circle of ideas that I want to talk about. All right.

So one more time the fundamental equation is F of T is bandlimited. The Fourier transform of F of S is identically zero for S greater than or equal to P over two. And you have the equation that the Fourier transform is the cut off of its periodized version. All right. Everything depends on that. Everything depends on that equation. Okay.

Now, here's a common problem for signals that you might want to interpolate for signals that you're sampling. If you know the bandwidth – if you know P – we have a pretty good idea of P, and you can always take P too large. All right. That's not a problem. You're working harder than you have to, but at least as far as this formula goes, it's okay if you're beyond – it's okay if you take a P beyond where the Fourier transform is identically zero. The problem is if you don't know the bandwidth so well, and you take a P that's too small.

So what happens if you try this for a – let's call it P prime – that's too small? Okay. All right. So you can – so here's the picture. It's bandlimited all right. So there's a real value for the bandwidth, but if you take something that's a little bit too small – here's P – minus P prime over two. Here's plus P prime over two. And actually the Fourier transform bleeds out a little bit over that. Okay. All right.

You can form the periodization. You can derive the sampling formula just like you always derive the sampling formula. So you can form the Fourier transform of F convolved with P prime – with the Shah function with spacing P prime, and you can cut it off. What do you do if you – what if you – what do you get if you form that? Well, if you form that – again, here's the real Four – here's the Fourier transform with its real bandwidth shown. If I periodize with a less per – with less of a period – with too small a period – minus P prime over two P prime over two.

Then what's gonna happen is if I shift it by P prime then I'm gonna be getting – it's not gonna be shifted off itself. It's gonna overlap. All right. It's gonna overlap, and add up. So it's not only overlapping, but the sum of – so the Fourier – this new function – the Fourier transform of F convolved with the Shah function, which is spaced a little too close together, is gonna give you something that looks like this, and so one, and so on. Now, you should convince yourself of that. All right. It's not hard to see pictorially what's going on, and that's exactly what the issue is. And you can all – but anyway, you can do it. That's fine.

And you can also cut off. All right. Then you can cut off. You persevere here. You can cut off by – you think you're doing everything right. All right. You think you're doing everything fine. So you cut off by that. You form high P prime times the Fourier transform of F convolved with the Shah function of P prime. All right.

Now, what you get is if you cut off by the corresponding – let me see if I can draw the picture here again. It's roughly like this. I'll just try to do two things. So the picture is something like – here's P prime. Here's the new curve. By the sum – this goes like that. All right. If you cut off by the rectangle function, then the curve you get is this. Right?

That's not the original Fourier transform. You don't get back the Fourier transform of F. Okay. You don't. But you persevere.

You don't know that. Right? You are a Stafford engineer. No mistake will stop you. Right? All right. You persevere. What do I mean by persevere? You take the inverse Fourier transform. You derive the sampling formula – all right – because the sampling formula is identical with the derivation of the sampling formula. All right.

So you take the Fourier transform – inverse Fourier transform of this. I P prime Fourier transform of F convolved with Shah function of P prime – just as you always did. All right. And what do you get? You get a sampling formula. You get the same formula it looked like you had before. You get the sum from minus infinity to infinity the value of F at these sample points K over P prime times the shift of sinc function sinc of P prime T minus K over P prime. That's what you get. You take the inverse Fourier transform of this expression over here, you turn the crank, and that's exactly what happens.

But you don't get F. You can't get F. All right. You can't get F because this is not the Fourier transform of F. You get something, but you don't get F. You do not get F. You do not get F of T because the derivation of the sampling formula is identical with the proof of the deri – proof of the sampling formula, and sa – and you don't have this. You have – because the Fourier transform of F is not equal to the cut off of its periodized version. Because the Fourier transform is not recovered this way, that formula can't be recovering F – can't be. All right. You don't get F, but you get something.

What do you get? Okay. What is the relationship between what you get – what do you get? Well, let's call – and so everybody got this? This is a very important point. All right. Let's call it G. You get something. It's a fine signal, so call this G. So let's say G of T is this sum. K equals minus infinity to infinity F of K over P prime sinc of P T minus K over P prime. That's it. Okay.

What does G have to do with F? Anything to do with F? Yes. It has the same sample values of F. F and G agree at the sample values. So F of T, and G of T agree – are equal at the sample values K over P prime. Why? Because G of – let's take one. So G of M over P prime. All right. That's the sample value for some value of M. All right. So what is that? That's the sum K going from minus infinity to infinity F of K over P prime sinc of P M over P prime minus K over P prime. Right?

Now, look at what –look at what's happening in the sinc function there. In the sinc function the sinc of – so the sinc of P M over P prime minus K over P prime –you'll have to remember some properties of the sinc function here. Multiply the P through. So that's sinc of M minus K. Right? That P prime's in the denominator in both cases. The sinc function –this is the sinc function of an integer. M and N – or M and K are integers here. All right.

It's one when M is equal to K, and zero otherwise. All right. This is equal to one if M is equal to K, and zero if M is different from K. And the sinc has zero crossings, if you want to think of it that way – use that terminology, at the integers – all right – except at the origin where it's one – where it's equal to one. All right. So all these terms drop out except when K is equal to M.

In other words, to write it down one more time, G of M over P prime is equal to this sum K going from minus infinity to infinity F of K over P prime sinc P M over P prime minus K over P prime. All the terms drop out except when M is equal to – when K is equal to M, and that then gives you F of M over P prime. All right. They agree at the sample points, but not all –but not elsewhere.

There may be some other accidental agreement. Who knows? Who cares? But the one thing you do know is they agree at the sample points. All right. Because of this property you say that G is an alias of F or that F, and G are aliases of each other. I don't know which way – so you say that G of T is an alias of F of T. It's not equal to F of T, but it has the same sample values as F of T.

Now, what that means is, among other things, if you know only the sample values – if that's all you know about F, you can not distinguish F from G. All right. You know some things about F, say you know the sample values, you make a guess about the bandwidth. All right. If you guess wrong, and you reconstruct G according to the derivation of the sampling formula because the sampling formula is identical with derivation of the sampling formula, then you are constructing not F, but an alias of F, but you don't know that. All right. You don't know that because you don't know anything about F other than its sample values. So you can not distinguish the two. That's why I use the term alias. It's a perfect masquerade. All right. It's G masquerading as F because they have the same sample values? Okay.

And it arises – I wouldn't say not from the misapplication of the sampling formula. I don't think that's fair. It falls from deriving the sampling formula as you would ordinarily when it doesn't apply – or rather when you don't get back – when that fundamental equation relating the Fourier transform – the periodized Fourier transform is not valid – is not satisfied. You get something, you can turn the crank, you get a formula, but you're not reproducing the original function.

Now, you've probably studied aliasing, and things like that – and not today, but I hope next time actually I bring in a demo of this. And you may have talked about various things folding in higher frequencies to lower frequencies, and all the rest of that stuff. That's fine, but for me, again, I never think of the sampling formula without thinking of the derivation of the sampling formula.

And if I wanted to understand aliasing, and what's going on in aliasing, I think about exactly this argument. What's gone wrong? Or what do you – what are you doing? The derivation was fine. Everything we did was fine, but what's not true is this fundamental equation. You don't – you have – the Fourier transform is not equal to the cut off of the periodized Fourier transform. I called it P prime here. All right.

If the Fourier transform is not equal to the cut off of the periodized Fourier transform then what you're getting back by turning the crank by writing down the sampling formula is not reproducing F. It's reproducing something else. It's reproducing an alias of F. Okay.

Now, I want to show you one explicit example. This is a fairly simple example, but a nice one. And this is done in the notes actually. All right. So I won't go all – go through all the details, but I want you to see how it works in one particular case.

So again, it's – I want to stress this that in aliasing it's not that you've done something wrong. You have reproduced a signal that had the right sample values, but it's not the signal that you hoped to reproduce because of this. All right. It's not that you made a mistake somehow – or maybe you have made a mistake, but you haven't sinned against the forces of good. It's just that you're deriving the sampling formula, pursuing the argument the natural way of pursuing it, but you're not getting back the original function because this equation is not satisfied. Okay.

So I want to do an example. So again, these are in the no – this is in the notes. Example of aliasing. All right. So I'm gonna take just a cosine. So the spectrum is just two Delta functions. G – F of T is a cosine of nine Pi over two times T. All right. So this says frequency nine fourths – Hertz. Now, I know what the Fourier transform looks like. The Fourier transform of F looks like two Delta functions. Looks like one half Delta S plus nine fourths plus Delta S minus nine fourths. All right.

So what's the picture? The picture is this. The picture is – here's the Fourier transform. Here's zero. There's a Delta at nine fourths. There's a Delta at minus nine fourths. Now, a little tricky here – this is something that again, that's discussed in the notes, and I'm not – I'm gonna sort of skate over now. How would you – now, so you can – if you want to reproduce this function by a sampling formula somehow, right? So what do you do? You periodize this. You cut off, and so on.

So how would you periodize this? What is the appropriate P to take? Well, the appropriate P to take should be anything strictly bigger than nine fourths because you want to shift this completely off its self. The problem here is that there's a spike right at the endpoint. All right. And it takes – it drops sharply. All right. So there's a little bit of an issue here at the endpoints, but I'm not going to – let me just say – sort of skate by that, and I'll – you can look at the notes on that for a little bit more detail.

So proper sampling rate would be anything greater than nine fourths because shifting this by – convolving this thing by a Shah function with that spacing – anything bigger than – not actually half – twice this, right? The whole bandwidth would be okay. All right.

So you would want to convolve – so P here – P over two is nine fourths. So P is equal to nine halves, but again, you want to take P – anything bigger than that to get – to shift the spectrum off its self convolving with the Shah function of spacing P. All right. And again, there's a little bit – some more to say here with the endpoints, but never mind that right now.

What I – the phenomenon that I want to talk about is what if you can't do that. What if your sampler is stuck in low, and you can sample it, and you sample it too low a rate? All right. So what if we can't do this? What if the sampler is stuck in low? All right.

So, for example, E G sample at rate P is equal to one. Okay. So in order to reproduce this thing you'd want to take samples at a rate greater than nine halves – Hertz, but you can't always do that. You can only take samples at a rate of one per second. All right. Play along. Play along. The sampling formula is identical with the derivation of the sampling formula.

What do I have to do? To sample at a rate one means to convolve with Shah sub one. Sorry – I should've – probably should've said that a little bit more forcefully when I was talking about the original derivation. So sample at a rate at P means to convolve with the Shah function of spacing P. To sample at a rate one means to convolved with the Shah function of spacing one. Right? To sample at P equals one means to form the cut off, and the periodization. And the cut off means to form Pi one of the Fourier transform of F convolved with the Shah function of spacing one. Very good. Right? That's what it means because the sampling formula is identical with the derivation of the sampling formula.

Now – all right. You can form – and I won't do it. You can easily form that convolution one half Delta S plus nine quarters plus Delta S minus nine quarters convolved with the Shah function K equals minus infinity to infinity Delta X minus K spacing one. All right. What you use is – I'll remind you of formula for convolving Delta functions – shifted Delta functions, you use Delta A convolved with Delta B is Delta A plus B. Okay. All right. And you find if you do that you get a bunch of arrows. All right.

You get a bunch of arrows, but you find only that two of them lie within – I mean there are two parts to this derivation of the sampling formula. That is periodizing, and there's cutting off. All right. So you get a bunch of arrows, but then you have to ask yourself, "This is Pi one. Which arrows are within the interval for minus a half to a half?" Only two arrows – only two Deltas within minus a half less than S less than a half. That is what you find is – you find that – I want to be sure I got this right. You find only – matter of a fact at a quarter to minus a quarter. Only – S equals only Delta S minus a quarter, and Delta S plus a quarter.

That is to say if I take the rect – if I cut off by a rectangle function of width one, the Fourier transform F convolved with the Shah function of spacing one, the only thing that remains are these two Delta functions. That is it's one half Delta S plus one quarter plus Delta S minus one quarter. All the other Deltas are outside the range. Okay. All the other Deltas are outside the range. You just check that out algebraically just by convolving using this formula. Okay. All right. Play along.

What does the derivation of the sampling formula say? The derivation of the sampling formula says take this formula, and takes its inverse Fourier transform. Derivation of the formula says to take the inverse Fourier transform of I one of the Fourier transform of F convolved with Pi one. That'll give you the function. That'll give the so called interpolating function.

So what do you get? You get the inverse Fourier transform of that. That's easy. That's a cosine. That's another cosine. This is the inverse Fourier transform of one half Delta S plus a quarter plus Delta S minus a quarter. And that is the cosine of Pi over two times T. All right. Deriving the sampling formula at the sampling rate one, which is too small, has led to the function cosine of Pi over two T, not to the function cosine of nine Pi over two times T. All right. Deriving a sampling formula, once again, leads to the function G of T is cosine of Pi over two times T, not to F of T is equal to cosine of nine Pi over two times T. All right.

Not that you've done something wrong, you haven't made any mistakes along the way other than the mistake of taking too low a sampling rate, but you've turned the crank, and you found a function, but you didn't find the function you wanted. All right. But you did find a function which agrees with the function that you wanted at all the sample values.

And you find that G of T equals F of T at the sample values at – which sample values? The samples values corresponding to taking samples every second – at zero, plus or minus one, plus or minus two, and so one, and so on. Okay. You have found an alias of the function, not the function. You've done that by deriving the sampling formula with a sampling rate that's too small. All right. In the sense that the fundamental equation – the cut off of the periodized Fourier transform does not give you the Fourier transform back. All right – that that fundamental equation rela – is not satisfied. All right. And so what you've come up with is something different. Okay.

That's the main – this is sort of the natural phenomenon this assoc – aliasing is a way to sort of speak the natural phenomenon that's associated with the sampling theorem. All right. But it is nothing other than doing the right work for – in the wrong cause, so to speak – all right – by taking the wrong sampling rate – by taking a sampling rate such that the fundamental equation's not satisfied. Okay. All right.

That's all about I wanted to say on this actually as much as I love this topic. So next time actually we're gonna use it to give a derivation of the discreet Fourier transform, and we're gonna move in the modern world from the analog to the digital. All right. Next time.

[End of Audio]

Duration: 52 minutes

TheFourierTransformAndItsApplications-Lecture19

Instructor (Brad Osgood):We're on the air. Pay attention. You should – I think the exams will be back sometime today. I'll send out a note when everything is all sorted out. I don't really know how things are gonna turn out yet. I graded my part over the weekend. I graded the first part problem and people did actually quite well on that. I think Rajib is working on it now, and at some point – anyway, I'll send you a note out when everything's all set for you guys to pick up. All right? I am going to do a little demo on aliasing today that I didn't want you to miss. Unfortunately, I loaned my strobe light to somebody, and I can't find it. I can't remember who I lent it to, so I can't give you a demonstration of the spinning fan, which is the classic demonstration of aliasing. If I can get my strobe light back, then maybe we'll have a chance to do that a little bit later. So I'm gonna give you a demonstration, a slight demonstration, a brief demonstration on what happens when you under sample music, that is when you turn analog music into digital music that we're all so used to these days, and when you do it the wrong way, or you do not do it accurately enough. So how would you – how rapidly do you have to sample a piece of music in order to be able to interpolate it, in order to be able to reconstruct it digitally? Well, for that we need – sampling. We need to know something about human hearing. That is how high a note you can hear, and that's roughly about 20 – the bandwidth of – or rather – gonna start that again. You can hear roughly up to 20,000 hertz. That is on the high end. You can hear – dogs go up much higher of course. You can hear up to roughly a note with a pitch 20,000 hertz. That's way beyond anything you'd really hear musically of course, way beyond where music goes, but that's roughly the range of the human hearing from about 20 hertz to about 20,000 hertz.

So according to the way we do things then, the bandwidth or the spectrum of – a slice of spectrum of music would go roughly up to say 20,000 hertz and then down to minus 20,000, so the frequency would be – and beyond that it's essentially zero. At least as far as you're concerned it's zero. You can't hear anything. So if that's a picture of the spectrum of a slice of music, then it's between minus 20,000 and 20,000, so the way we write things, that would be P over 2. The bandwidth is 20,000, so P is about 40,000, which means that if you wanna sample and reconstruct music, you should do it roughly at a rate of 40,000 hertz. That is to say you should sample every 40,000th of a second. So this should be the sampling rate for music, roughly 40,000 times per second, or sampled at a rate of one over 40,000 per second in order to be able to have confidence that you are sampling rapidly enough so that your reconstruction based on the sync function is gonna interpolate the actual music. Now in fact, as people probably know, they sample at a rate – for CDs, for example, they sample at a rate higher than that. Do you know what it is off the top of your head? 44.1 or something like that? Is that right? Somewhere around there? In fact, the sampling rate for CDs is whatever it is. I think it's 44.1 kilohertz. And also, as far as I know, and some of you may know the history of this better than I do, this is – it's gotta be something bigger than 40,000, all right? But as far as I know, the precise number comes from just the equipment that they were using when they were making the transition from analog to digital. I think the way the tapes – I'm sorry, do you know?

Student:[Inaudible].

Instructor (Brad Osgood):Is that right? Oh, really? The story that I had heard – I tracked this down [inaudible], but I haven't looked at it really carefully was that everything was set up for analog recording, of course. And the way the machines were set up, it was just most natural to sample at that rate somehow. They got the best – they got the most accurate – most reliable sampling if they did it roughly at that rate, but it was not – that number comes out of practical considerations, not out of any sort of theoretical inspiration. Anyway, that's how fast you should sample, okay? So here is a well-known piece – and if you sample at less than that rate, then you're not gonna get an accurate reproduction of the music. So here is Gershwin's Rhapsody in Blue sampled at a high enough rate, sampled at 44.1 kilohertz. It has a famous clarinet gliss at the beginning. Everybody remember? Everybody heard this? Okay. All right, now – so then – isn't that nice? Nice way to start the day. So then what we did was sample this at a low rate, I think around 16 hertz – 16 kilohertz. All right? Not 16 hertz, 16,000. Roughly a little less than half of that. Now remember what that means in terms of the picture. For me, as I said many times, the sampling theorem is identical with the proof of the sampling theorem, so I always try to imagine what things look like in the frequency domain, in the spectrum. So that means if the real spectrum goes to that range, you are cutting of like here. So that means you're cutting off a bunch of high frequencies, so to speak. And what does the proof of the sampling theorem say? The proof of the sampling theorem says to periodize this thing by a Shaw function that has spacing that is too narrow and then cut off.

So if you just take the first shift, if you shift thing over by the corresponding Shaw function, then you're doing this. You're getting like that. And then shifting it over the other way, you're doing that, so when the two signals add, when this adds with its shifted version, you're getting something that looks like that. And then you cut off – [inaudible] I'm not doing this right. And you cut off like that, so you're getting some signal – the frequency of the signal that you're reconstructing, instead of the original spectrum, looks something like that. And they talk about low frequencies being aliased as high frequencies, and high frequencies being aliased as low frequencies. What they mean by that is the lower part of the spectrum here is getting shifted over, and is getting added in as if it were part of the high frequency. So that means the lower frequencies are being aliased as high frequencies because you're shifting things inadequately to shift this spectrum off itself. So you're reproducing a signal based on this picture of the frequency instead of this picture of the frequency, and here is what you get. [Inaudible] lower parts are not as affected by [inaudible] distortion is more [inaudible]. You get the idea. The distortion really is more on the high end of things because it's the high end of things where the overlap's occurring. You're cutting it off by not enough. You're shifting things over. You're adding up. And there's more mistake that's made in the higher part of the spectrum than the lower part of the spectrum. This part of the spectrum is pretty much unchanged. It's the higher part of the spectrum that's getting messed up by an inadequate shifting or too low a sampling rate.

And there are all sorts of experiments and tricks you can do with this. Aliasing should not always be thought of as the enemy, by the way. There are times when you wanna alias on purpose. As a matter of fact, there's a problem I believe on the problem set on the basis for the so-called sampling oscilloscope where you try to sample – if you have a very fast signal to sample, that's faster than – the frequencies are higher than an ordinary oscilloscope can actually handle, then there are ways of aliasing effectively to get a picture of the signal at a lower – by sampling at a lower rate. So it's not that aliasing is always the enemy necessarily, but it's something that certainly has to be understood. Then again, as far as I'm concerned, understanding aliasing is understanding the proof of the sampling theorem, which is what we've been talking about. Okay. The other famous example with aliasing is one you had a problem on also, and if I can find my strobe light, we'll do it sometime because it is sort of fun to do where you freeze a fan going – or you freeze a rotating object that's rotating periodically. It's described by a periodic signal. If you sample it inappropriately, say at too low a rate once again, you can freeze it. You've probably seen demos like that. It's very impressive. Okay. And then what I'd really like to do – I have done this before – is do some other examples of sampling and actually spectrum analyzing with a real, honest to god spectrum analyzer, bring my trombone in, bring some other musical instruments in, so you see how they look in the frequency domain. You can also do a bunch of interesting demos on that. We'll see if we have the time for that. Any questions about that? Anything on anybody's mind? Okay. So today, it is welcome to the modern world. Today, we're gonna make the transition from the continuous to the discrete, from the analog to the digital, the modern world. I want to introduce – you've actually been using it in – any time you've used MATLAB with a Fourier transform you've been using it, but now we're gonna really talk about why it works, and how it works, and where it comes from. I wanna introduce the DFT, the discrete Fourier transform. The DFT – that is to say we're not saying farewell to the continuous and the analog world, but we are now having a meeting between the two, the old world and the new world. So we wanna move from the continuous and the analog to the discrete and the digital. So we're moving from the continuous and analog to the digital and discrete, to the discrete and digital.

Now like so much else in this course, there are choices to make. Some people would say this is not a transition at all. You should never have been in a continuous world in the first place, you Neanderthal you. The modern is digital, the modern world is discrete, and you should always understand things in terms of exactly those sorts of operation. That's a defensible choice, and there are courses actually on digital signal processing where you don't really talk so much about the continuous case, but everything is discrete and digital right from the beginning, and all the formulas go that way, all the demonstrations go that way, all the theorems go that way, and so on. So that's defensible. I don't like it, and I don't think it's ultimately that – I think you miss a number of things by doing it that way, and I don't think it always simplifies things, but nevertheless, that's a defensible choice. For us however, we wanna make a transition from the continuous to the discrete, or from the analog to the digital. And in doing so, I wanna leverage I hope on all the hard won skill and intuition that we built up on the continuous Fourier transform. We spent a lot of time and before this class, I'm sure you spent a certain amount of time also on just developing those ideas, understanding analog signals. And what I wanna convince you of is that a lot of that skill that you've build up, and a lot of the intuition that you built up carries over. So much so that even there's a similarity in the formulas. Even the formal aspects of the subject are similar enough in the discrete case that you can carry over what you learned then to what you're gonna learn now. That's the other reason for doing it this way is so we can use all that we have learned so far in the continuous and analog case to understand, and help us analyze, and help us work with the discrete and the digital case. So that's the choice that we made, but it is a choice. It's not necessarily the only way of doing things, by no means. So here's how we're gonna proceed. Here's the plan. It comes really in three parts. One – so I have – you're starting off with continuous signals, so FFT is a continuous signal as usual. There's a continuous signal. When I say that, I'm not meaning it in a formal mathematical sense. I'm thinking of it as a function of a continuous variable T, and that's what's gonna get replaced by a discrete. F is gonna get replaced by a discrete function. All right? Think of this as the analog case.

So the first step is I wanna find a reasonable discrete approximation to FFT. Secondly, I wanna find a reasonable discrete approximation to its Fourier transform. All right? I know how to take the Fourier transform and make a continuous signal. The question is can I discretize that in a reasonable way that's providing a reasonable approximation. And the third, which really combines the first two steps immediately when you see how the development goes, is to find a reasonable way from passing from one to the other, from the discrete approximation of F to the discrete approximation of the Fourier transform that is the most natural thing to do in this case. That is it approximates the passing from the continuous case, the continuous function of the continuous Fourier transform. So find a reasonable way of passing from the discrete form of F to the discrete form of the Fourier transform that mimics – I won't write the rest of the sentence down – that mimics the way you pass from the form of F to the signal to its Fourier transform, the continuous version of that. Okay? And again, I wanna make this look as much like the continuous case as possible. There are different ways of making this argument also. There are different choices for how you can execute this. We're actually gonna base this on the sampling theorem, misapplied a little bit or – yeah, misapplied is probably the right word. So we're gonna base this – that is to say, I'm gonna base the reasonableness, the test of reasonableness, on sampling. So base this on misuse of the sampling theorem.

Well, the sampling theorem let me say – rather than saying based on a misuse of the sampling theorem, I'm not gonna write down the sync interpolation. I'll say I'm gonna base it on the misuse of sampling. Here's what I have in mind. Okay? I'm gonna make some assumptions that I know are wrong right from the outset. I'm gonna assume – but play along – assume first of all that F of T is limited – it's a time limit signal to zero less than or equal to T, less than or equal to L. Okay. And I'm gonna also assume that the Fourier transform of F is limited to zero less than or equal to S, less than or equal to 2B, so B is the bandwidth, and I'm writing B for bandwidth instead of P. Now both of those assumptions cannot hold together. So you can't have both together, but play along. This can't happen. This cannot be because the signal cannot be both time limited and band limited. That's one thing. The other thing here is that I'm writing the Fourier transform as if it's only on the interval from zero to 2B. Sorry, you can't see it very well here. So F of S is limited by zero less than or equal to S, less than or equal to 2B. All right? Now we know the Fourier – that's what I just wrote there. The Fourier transform is symmetric. You talk about frequencies going from minus B to plus B. The only reason I'm saying from zero to – as between zero and 2B is because of the indexing of the discrete variable that's gonna go along with this. You'll see what I mean. All right? So we're saying this –

it's purely a formal statement just to make the notation a little bit easier in a minute – saying this to make indexing of the discrete variable easier. All right. You'll see. So just play along.

All right, now – so I have a sample that's limited in time and limited in frequency. We know that can't happen, but let's play along. And what would the sampling theorem tell us? If I wanted to reconstruct F, how many samples would I need? Or rather what should the sampling rate be? Well, the sampling rate to consider myself getting a reasonable approximation of F from its samples is dictated by the properties of F in the frequency domain, that is to say the bandwidth of it. So to get a reasonable discrete approximation of F by sampling, I take samples spaced one over 2B apart. That is the sampling rate should be 2B, and so the samples should be one over 2B apart in the time domain. Okay? So in the time domain, here's zero, 2B, and the sample should be spaced one over 2B apart. So let's say there are N of them. Let's say take N samples. All right? Oh sorry. This is the time domain, so the length of the interval is L. In the time domain, the function is limited from zero to L. I take the sample space, one over 2B apart, so I should have N over 2B is equal to L. That is if I take N samples, and the relationship between N and the number of samples that I take, the spacing, and the length of the interval that they're supposed to fill up is that, and it's equal to 2B times L if I take N samples. Okay? And let's call the samples say T [inaudible] the sample points, T not zero, T1 is equal to one over 2B. T2 is equal to two over 2B. And then I go all the way up to [inaudible] N minus one over 2B. Okay? Those are my sample points. There are N of them indexed from zero to N minus one. So what is the sampled form of F? For us, the sample formed of F [inaudible] function is always multiplying a function by given – by delta function with that spacing. Take F of T times the sum from say K equals zero to N minus one of delta, T minus TK. That's what we always mean by sampling. All right? So again, the TKs are zero, one over 2B, two over 2B, all the way to N minus one over 2B, so the sum goes from zero to N minus one. So this is sum F of TK delta T minus TK, zero up to N minus one. That's a sample. We'll just call that F sample.

That's still a function of a continuous variable, but it's recording what we think of as a reasonable approximation to F. It's a reasonable approximation to F in the sense that the sampling theorem would tell us that if I base my interpolation on those values, I should be able to reconstruct it exactly. That's why it's reasonable. And as those points weren't just chosen arbitrarily, those points were chosen with that spacing because that's what the sampling theorem would tell us to do. Okay, now the sampled version of F is still a function of a continuous variable. I can take its Fourier transform. All right? That's not so hard. The Fourier transform of F sample is also a function of the continuous frequency variable S, and that's just the sum from K equals zero to N minus one of F of TK times the Fourier transform of the delta function which E to the minus – shift the delta function – E to the minus 2p I S T K. Okay? All right, so where are we here? I've got what I think is a reasonable approximation to the continuous function in the time domain. I've taken this Fourier transform, but the Fourier transform is not yet discretized itself. That's a function of the continuous variable. So I want to sample or discretize the Fourier transform of F, the Fourier transform of the sample function. Okay? How do I do that? Well, think of the sampling theorem. What would the sampling theorem say again viewed somehow going from the frequency domain back to the time domain? That is think of this as a function that I wanna sample. How rapidly should I sample – never mind that it's the Fourier transform of something. How rapidly should I sample this thing so the sampling theorem will tell me I'm getting a reasonable discrete approximation of it? That is I'm getting enough sample points so that if I wanted to interpolate it, I'd be getting the function back again.

How to sample this in the frequency domain so you get a reasonable approximation, let's say – so you get a reasonable let's say discrete version if I start using that sort of terminology. All right. So in the frequency domain the signal is limited to zero to 2B. All right? How rapidly to sample in that domain is dictated to us by what happens in the other domain where the signal is limited to zero L. How far apart should the sample points be spaced in the frequency domain in order that I get a reasonable approximation if I take samples of those points? It was either an answer or a cell phone ringing. One over L, all right? In the time domain, remember the signal is limited to one over L. All right. In the frequency domain, it's limited to one over 2B. How fast the sample in one domain is dictated by the properties of the function in the other domain. All right? So the answers will go back and forth here between the function and its Fourier transform, but you should be able to say one or the other in either – I mean talk about one in terms of the other in either direction. So once again, let me say it. To think about how fast a sample a signal in this domain should have to do with what its properties are in the other domain. So I'm not speaking so much in terms of time and frequency here. I'm speaking in terms of the one domain and the other domain. The other domain, the function is limited to zero to L, so in this domain the sample should be taken one over L apart. All right? That's what the sampling theorem would tell us to do if we were applying the sampling theorem in this direction.

If you take samples spaced one over L apart, spaced one over L, okay – and that's supposed to cover an interval of length 2B. Okay? How many sample points would I take in the frequency domain? So again, if I have M sample points in the frequency domain, then I have M times – so M points say – then I have M times one over L is equal to 2B. That is M is equal to 2B L. Oh, but that's N, all right? The number of sample points in the time domain. That is to say – I shouldn't say N. Suppose M points, then M is equal to N. That is to say I take the same number of sample points from the time domain as in the frequency domain. So again you – that's what I meant by saying again here. So again in the frequency domain, you again take N sample points. What are they? Then they are space one over L apart, and sample points spaced one over L apart, so they are like S zero is equal to zero, S1 is equal to one over L, S2 is equal to two over L, and so on going on to SN minus one is N minus one over L. Okay? Again, one feels by an appeal to the sampling theorem that if you are sampling the function often enough at those points here, you're getting a reasonable enough approximation in the sense the sampling theorem would tell you you could reconstruct the function exactly from those sample points. Never mind that all this is being misapplied. Play along. All right? Okay. So what is the sampled version of the Fourier transform of the sampled function? What is the sample form of the Fourier transform of the sample version of F? Well, how do I sample a function? I multiply it by the corresponding delta function. So this would be the Fourier transform of the sample function, F sampled, times the sum of the corresponding delta function I've gotta call – what did I call it over there? I called it K in that sum. Let me call it something else, M. M equals zero to N minus one. Again, I have N points delta S minus S K where these are the S Ks, those are sample points. Okay. That's the sampled version of the Fourier transform. What is that? Let's write it down. I'm erasing this, but I'm gonna write it again. Oh, there it is. Okay, so the Fourier transform of F sampled once again at S is this sum: sum from K equals zero up to – K equals zero up to N minus one of F of T K, E to the minus 2 p I S T K. All right? So I multiply that by the corresponding delta function – sum of delta functions rather. So this K equals zero up to N minus one, N minus one of F of T K, E to the minus 2 p I S T K times the sum of deltas M equals zero up to N minus one. I'm calling M, right? Just to be consistent here. Yeah. Delta S minus S K gives me what? Gives me these terms times these terms, so it gives me the sum over K and L going from – K and M going from zero up to N minus one. F of T K, that's a constant. I get an exponential function, a complex exponential function times the delta function. We know what that does. That pulls out the value of the exponential at the point where the delta function is shifted and then times the corresponding delta function. So F of T K, either the minus 2 p I S M T K times delta S minus S K. Cool. Okay? Cool, cool, cool. Pardon me?

Student:[Inaudible].

Instructor (Brad Osgood):S, S M. Now it's cool. Is it cool?

Student:[Inaudible].

Instructor (Brad Osgood):Are we cool? Okay. All right. So what are the sampled values of the Fourier transform? It is there before us, actually. All right. I wanna repeat again how we got to this point. We got to this point by misapplying the sampling theorem that has led to something that actually turns out to be reasonable and useful. And again, as is the tradition here, once you reach this point you sort of cover your tracks and just say, "All right. Now I'm gonna turn this into a definition."

How do we get to this point? We said we wanted to choose sample points of F that we thought were gonna be reasonable – sample points of the function little F in the time domain that we thought were gonna be reasonable – provide a reasonable approximation of F. We take the Fourier transform of that. That gives us a signal. That gives us a continuous signal, function of a continuous variable in the frequency domain.

Then I wanna sample that, and you say to yourself, "How do I sample that?" Well, you sample that according to what the properties are in the other domain. That says I take sample values one over L apart, and I sample according to that criteria. And then the sample version of the Fourier transform of the sampled function is this. I'm gonna say that one more time to make sure I've said it right. The sample version of the Fourier transform of the sampled function looks like this.

So what are the sampled values of the Fourier transform of the sampled function are – let me call them – let me just use a different notation, [inaudible] call it capital F – are F of S zero, say that's the sum from K equals zero to N minus one. If M is fixed at zero here, the first entry, the zeroth entry is F of T K E to the minus 2 p I S is not – S zero times T K. F of S1 is the sum from K equals zero up to N minus one of F of T K, E to the minus 2 p I S1 T K, and so on. I get N sample values of the Fourier transform, and so on down to F of S M minus one is the sum from K equals zero up to N minus one of F of T K, E to the minus 2 p I S N minus one T K. All right? That's the discrete approximation to the Fourier transform of the discrete version of F – discrete approximation to the Fourier transformation of the discrete approximation of F. All right. I approximate F by taking N samples in the time domain. I approximate the Fourier transform N values in the frequency domain according to this rule, according to this formula. All right. Let me take stock here because we are actually – we are there almost. So again F – F of T is approximated by – or discretized by [inaudible] better here – F is discretized to F of T not F of T N minus one, and the Fourier transform of F of S is discretized to F of capital F S not F S N minus one where the formula relating capital F to little F to the sampled – to the values here, to the values here, are the formula that I just have. F of – capital F of S M is equal to the sum from K equals zero to N minus one of F of T K, E to the minus 2 p I S M T K.

Okay. Now there's one more step in actually defining the Fourier transform as people usually think of the Fourier transform. This is still – a continuous side of things is still in sight. All right? We start off by looking at continuous functions and try to approximate it. You still see the continuous variables here. You still see the continuous picture. All right? And in fact, that's most often where the Fourier transform comes up when you're actually gonna apply it. There's usually some continuous process working in the background that you are approximating discretely. That's where it comes from. But the definition of the discrete Fourier transform as it's usually given is purely in terms of discrete data and discrete signals. So the final step in defining the DFT is to sort of eliminate the continuous completely, and define – and everything is defined in terms of discrete signals and use only discrete signals, or digital signals, however you wanna call it, discrete signals. Now how do you do that?

Well, it's not hard actually. According to our spacing, according to the way we set this up – so in our set up, T K – the Kth value is K over 2B, right? The points are spaced 2B apart, one over 2B apart in the time domain. The sample points in the time domain are spaced one over 2B apart. It's zero, one over 2B, two over 2B, three over 2B, and so on and so on. Those are the sample points in the time domain. In the frequency domain, the sample points S M are M over L. All right? They're spaced one over L apart, zero over L, one over L, two over L, and so on and so on. All right? T K times S M – if I use this how they are in relation to the spacing, then it's K times M over 2B times L, but you remember 2 B times L in the way we did the set up is actually the number of sample points. 2 B L equals N, the number of the sample points. All right? So in the complex exponential – so what do I have in my complex exponential? If I still keep the shadow of the continuous variable, I have 2 p I S M T K, but then if I use this, I can write this as E to the minus 2 p I K M over 2 B L. That's K M over N. All right? Only the discrete indices K and M now appear in that complex exponential. No more do you see on the right hand side where those points – where the K and the M so to speak came from. You just see the indices K and M. All right. They came from the Mth index of the continuous variable S, the Kth index of the continuous variable K, but there's a relationship here. That is we sampled at a certain rate. T K was K over 2 B. S M was M over L. The product is K M over 2 B L and 2 B L is the total number of sample points you take in either domain. All right? So the product there is just minus 2 p I K M over N. That's one – that's sort of the first step in eliminating the continuous variable from the picture. The last step in eliminating the continuous picture from the variable is to identify the value of the function in this Fourier transform at the sample points with just the index that determines that value. So finally, you identify the values of F of T K with the value of a discrete signal with a value – say let me call it this: F K of a discrete signal. That is discrete signal F is F – I need a nonsum notation for this – F zero up to F N minus one. And I'm following the sort of common notation here of using brackets when you talk about discrete variable rather than parentheses when you talk about a continuous variable. Where these entries here are nothing but the values of the measured signal – of the continuous signal at the point T K. F K is equal to F of T K, so you have to be a little careful here. I'm trying to make a distinction somehow between the discrete signal indexed by K coming from the values of the continuous signal at the sample point T sub K, okay? And likewise, in the frequency domain, you replace in your mind or in the definition the continuous S K or S M by the index that determines it. So likewise, we replace S M by the index M, i.e. F by the discrete signal, capital F, underline – I'm trying to use that as an indication that I have a discrete signal – of this n-tuple F zero, capital F, N minus one where its value at the Kth index is before what I was calling F of S K. All right?

The right hand side made sense from what I had before. It's the value of this approximate Fourier transform at the Kth sample point. But identify the Kth sample point just with this index, and I consider that to be defining a discrete signal, F0, F1, F2, F3, and so on and so on. Okay? And if I do that, then all traces of the continuous variable are gone, and I now have a transformation from one discrete signal to another discrete signal. If I do this, then all traces of a continuous variable are gone, and you have a transformation from one discrete signal to another discrete signal. That is to say if you start off with the signal little F, discrete signal, which I'm writing like that – if you think of it as an n-tuple, F0 up to F N minus one. All right? So just given by N discrete values, indexed here from zero to N minus one, then it's discrete Fourier transform, the DFT of this discrete signal is the discrete signal capital F, which is again an n-tuple indexed from zero to N minus one the way I've indexed things.

And what is the definition? The definition is F of M is the sum from K equals zero up to N minus one of F of K. I'll write that a little bit nicer. E to the minus 2 p I M K over N. Right? All traces of the continuous have disappeared, and some would consider this a great step forward. All right? Oftentimes, when you see treatments of the discrete Fourier transform, if you just look at a book that's devoted to the discrete Fourier transform, or sometimes even if you look at other books on Fourier analysis and they talk about the discrete Fourier transform, they often jump right to this definition. All right? And that's defensible. You can say the Fourier transform is for continuous function – is functions of continuous variables, and so on. The discrete Fourier transform is for functions of discrete variables. It's for discrete signals. Here's the definition of the continuous case. Here's the definition in the discrete case. Don't bother me. All right? It's a definition. I get to do what I want. But what you often miss in that treatment – [inaudible] it's a choice. And I chose to do it this way because I wanted to make the connection with the continuous case because I think actually that in applications, that's how most often you see it. You wanna pass from the continuous to the discrete. You wanna pass from a continuous signal to a discrete approximation, and you still want the tools of Fourier analysis that you worked so hard to learn in the continuous case to be available to you in the discrete case.

So what you have to do most often in applying this is to say to yourself, "All right. That's the definition. Fine." That's what the discrete Fourier transform looks like. I have a discrete signal, little F. I have its discrete Fourier transform, capital F. This is how they're related. The indices of capital F, the values of capital F at the discrete points, 0, 1, 2, 3, and so on, are related to the values of little F, so on and – according to this formula, everything is discrete, but what it comes from is – you have to realize that these are coming from values of a continuous signal approximated at a discrete set of points. These are values of the approximation of the Fourier transformation, the continuous Fourier transform at a discrete set of points. Okay? That is the burden of my remarks. Okay, now what we're gonna do is – like I say, I'm gonna try to make the discrete Fourier transform look as much as possible like the continuous Fourier transform. I'm gonna try write the – we're gonna wrap up today. I'm gonna to write the formulas for the discrete Fourier transform in a way that look like the formulas for the continuous Fourier transform. I'm gonna write the theorems for the discrete Fourier transform to look as much as possible like the continuous Fourier transform. The inverse of the discrete Fourier transform to look like the inverse of the Fourier transform – there's a catch there as it turns out. But by and large, you can do this quite – you can take this quite far. You can take the analogy quite far. Now again, that's a choice. You can just study this object the way it is. You can apply it the way it is to the particular cases that come up. But I think again from my experience, you gain much more from trying to make a connection to the continuous case than you lose by – you gain much more by doing that than by just taking this as the accepted definition and working with it as it is. All right? So that's gonna be our modus operandi. Now there are lots of little points along the way. There are lots of little observations of the similarities and the differences. I don't want to talk about all of them because there are – it's difficult to sort of see the whole picture at once, so I wanna try to do this at a level where you can see the whole argument go. So be sure to read through these sections carefully. Before you come to class, you should always do this naturally. Just so there's some small points – I won't always talk about this or always talk about that, but I wanna assume you'll have some familiarity with how some of the calculations go. I'm gonna do a certain number of them so you can see how the techniques go, and so you can see how the formulas come about, but there are a number of sort of little points along the way that have to do with the discrete, and then the continuous, and so on that I don't always wanna emphasize. So I'm asking you to sort of keep up with that as we go ahead and develop some of the properties of this. All right? So next time, we're gonna start to unwind this a little bit, and see how it works, and see how it works analogous to the continuous case. Thank you.

[End of Audio]

Duration: 52 minutes

TheFourierTransformAndItsApplications-Lecture20

Instructor (Brad Osgood):There we go. All right, the exams – remember the exams? I think they've all been graded, and the scores have all been entered, although I don't think we've made the scores yet visible on the website, so I will do that after I get back from class today. And you can pick up your exams from Denise Murphy, my admin in Packard 267. She has all of them. I'll also send that announcement out to the class. Any questions or comments on that? Good. Today we're gonna continue with our discussion of the DFT. This is "Getting to know your DFT, your discrete Fourier transformation." Now the subtitle I would say actually for this should be "You already know it." The point of the way we're talking about the discrete Fourier transform is that it can be made to resemble the continuous Fourier transform in a great many ways. So the intuition you built up for the continuous Fourier transform, the formulas that you've learned to work with and so on, really all have analogs in the discrete case. Now not in all cases. There are some things that don't quite match up, and that's interesting, too, but to realize what doesn't match up so carefully, and to realize why it's interesting I think goes along with seeing in many ways how much things are the same.

So we're gonna take the point of view that – and I'm gonna try to make the – to take the route to make the discrete Fourier transform look as much as possible like the continuous Fourier transform. That's our point of view. You don't have to do it this way, but I think it's most satisfactory, and again it allows us to leverage off of what we have done in the continuous case. Now before doing – so let me recall the definition. Here's the definition, and where we ended up last time was sort of a definition that where the continuous case, the idea of sampling a continuous signal to get a discrete signal, sampling its Fourier transform to get a discrete Fourier transform, all that had pretty much vanished by the final definition. So the definition that we wound up with – ultimately wound up with – ultimately had makes the continuous case almost invisible, vanish. It looked like this. You have a discrete signal. So I'm using sort of both the signal notation and vector notation here, and I'll continue to do that, sort of mix the two up because I think they're both useful. So the idea is you have either – you can think of this an n-tuple of numbers, or a discrete signal whose value at the nth point is just the nth component here. Okay? Oops, that doesn't look very – that's not much of a statement. And since you can either consider it as a discrete signal that's defined on the integers, where the integers from zero to N minus one, or you can think of it as an n-tuple or a vector. If you have a discrete signal and its Fourier transform is another discrete signal, its DFT is the discrete signal – I'll call it capital F, but I'll also use the notation to make the connection with the continuous case a script F with a little underline under it indicating it's supposed to be a vector quantity or something a little bit different from the continuous case.

And it's defined by its nth component, so the nth component of the Fourier transform is the sum from say N equals zero to N minus one of the Nth component of F times E to the minus two pi I N M over N. All right? So everything is defined here in terms of the indices in the exponential, and these are the values of the discrete function at the index points, F of zero, F of one, F of two and so on. That's the definition. They say you don't see at all the fact that in our derivation this came from starting with a continuous signal, sampling it, sampling the Fourier transform, and then somehow ultimately leading to this definition. Here it's just as an operation on one discrete signal producing another discrete signal. And that's pretty much how we're gonna deal with it. But before embarking on that path forever, I wanna have one nod back to the continuous case, and talk a little bit about how the DFT is employed and the kind of things you have to know when you're gonna use it in practice because you already have to some extent, and you certainly will more in the future. So one look back at the continuous case to talk about one additional – one sort of phenomena of reciprocity that comes – the reciprocal relationship between the time domain and the frequency domain that comes up also in the discrete case. So one more look back at the continuous case, continuous roots of the DFT, and reciprocity of the two domains, time and frequency.

So how did it work? We had a signal in the time domain that we discretized, and we had a signal in the frequency domain that we discretized. So imagine that you have two grids. You have a grid in the time domain. You have N sample points say that are delta T apart, N points. I'm running out of room to write things, so let me just write it here and then write some of the notation up above. And this is the frequency domain, and we have spacing in the frequency domain say delta S apart. So we have N sample points T zero up to T N minus one spaced delta T apart, and N sample points in time that is, and we have N sample points in frequency, S not up to S N minus one spaced delta S apart. So there are three quantities here. There's the spacing in the time domain, the spacing in the frequency domain, and the number of sample points N, and there's a relationship between them, so you have three quantities of interest, delta T, delta S, and N. And they're not independent. That's the important point. Let me remind you of again – I'm not gonna go through the derivation again, but let me remind you what the relationship is. We had at N delta T – that's the spacing in the time domain that has to be [inaudible] – just let me sure I get this right. I want to say this right – is L, that's the spacing in the time domain. I want to use the notation I used last night. This is a sort of a time limitedness. And we had N times delta S is the band limitedness that we called that two B. That was the band limitedness, bandwidth.

So delta T times delta S is equal to delta T is L over N times two B over N. That's L times two, that's two B L over N squared, but you will recall that when we set up the sampling, two times B times L was equal to N, the number of sample points. This is N over N squared. That is one over N. Okay? The way we did the sampling, making this not completely justifiable, appeal to the sampling theorem and so on gave us a relationship between how you sample in the time domain, how you sample in the frequency domain, the number of sampling points you took, and it was exactly that two B L was equal to N. It was sort of surprising or not clear a priori that if you carry out this procedure of sampling in the time domain and sampling in the frequency domain, you're gonna take the same number of sample points in both domains, but that's what we found. There's a relationship there. So here let me highlight it. That is delta S delta T or delta T delta S is one over N. The spacing in the frequency domain and the spacing in the time domain are reciprocally related to the number of sample points. This is called the reciprocity relationship. This again has practical significance and practical consequences when you're applying the DFT. You can imagine you have a continuous signal. You wanna sample it. What can you choose, and what is forced upon you? Well, you can imagine choosing how frequently you sample. Right? That's the delta T. And you can imagine how many measurements you make, how often you sample, how many samples you take, which is N. So you can imagine choosing delta T – that's how frequently you sample or the sampling rate – and N, the number of samples.

And if you do that, once you do that then delta S is fixed, and delta S is determined. The spacing in the frequency domain is determined. That is to say maybe one way of putting this is the accuracy of the resolution or how fine the resolution is in frequency is fixed by the choices you make in the time domain. So let's say – I don't know how you would say the resolution in frequency is determined – is fixed by the choices you make in time. Conversely, you could say I want a certain accuracy in – or I want a certain resolution in frequency. I want delta S to have a certain fineness. That's gonna determine then how many sample points I want to take and how I space them in the time domain. Or at least you have to choose two of those things and the third one is determined. And you can imagine if you're doing this with real data, you have a certain freedom here, but the freedom also carries with it certain restrictions. Ain't that life? Ain't that the way things go? The freedom you have typically is how many measurements you make and how frequently you make them. And once you do that, then that determines what the resolution is like in the frequency domain. There's ways of getting around this. There's zero padding and special – not getting around it, but ways of understanding it or massaging it. So-called zero padding, you have a problem on that. You'll have other chances to experiment with that and other sorts of things, but it's built into the system.

This sort of reciprocity relationship is another example of the same sort of thing we've seen so many times in the continuous case now carried over to the discrete case, the sort of stretched in one domain means shrunk in the other domain, or reciprocity between time and frequency. I say we've seen many different instances of that in the continuous case, and here's an example of how that carries over to the discrete case. So I wanted to say that because it is the sort of thing that you meet and you have to understand when you actually apply the DFT in – most often in most common contexts where it's associated somehow with some sort of continuous process that you're sampling. You wanna take the Fourier transform. You don't have the formula for it. You have to use numerical algorithms to actually compute it, and that means your sampling and these – the reciprocity relationship puts certain restrictions or limitations on what you should expect to get. So that is my final nod at least for now back to the continuous case. So let's go back now to the discrete world and pretty much stay there. So back to this formula and its consequences. Let me erase it and write it down again. We're back to the discrete world, discrete setting.

Again, my goal here is to make the discrete Fourier transform look as much possible like the continuous Fourier transform. So here's how to do that. I wanna make the DFT and associated formulas look like the continuous case. Maybe I should say we're not really abandoning the continuous case, but we're doing it – we're situating ourselves firmly in the discrete side of things. Now to help out in that, partly it's a matter of notation. Even more so, it's what you do with the notation and the consequences of thinking about things a certain way. So the first thing I wanna do is introduce a symbol, introduce a way of thinking about the exponentials that occur in the definition of the DFT. So let me write down the formula one more time. So again F is a discrete signal, F zero up to F N minus one. And its Fourier transform is another discrete signal whose Nth component is given by this formula: sum from N equals zero N minus one, the Nth component of F, little F times E to the minus two pi I N M over N. Okay, that's a two pi I. The first thing I wanna do is introduce the notation. I wanna view these complex exponentials as themselves coming from a discrete signal, as values of a discrete signal. This turns out to be a very helpful thing to do for a number of reasons. It gives you a compact way of writing things. It also gives you a way of enunciating certain properties of the discrete Fourier transform that would be very difficult to do otherwise. So I want you to realize the complex exponentials that come in the definition here as arising from also a discrete signal, the discrete or vector complex exponential. People either refer to it as the discrete complex exponential or the vector complex exponential. You pick what you wanna call it.

So let me give you the definition. I'm gonna write – I'll leave that board up there, so let me write it over here. I'm gonna write omega as a vector or as a discrete signal to be the n-tuple of powers of the complex exponential that appear in the definition. So the zeroth coefficient, the first entry is one, that's E to the zero. Then it's E to the two pi I – there's a minus sign up here, so let me define it in terms of a plus sign. Then I'll start taking powers to get a minus sign. So E to the two pi I over N. We'll see where this is coming from in just a second. E to the two pi I times two over N, and so on and so on, all the way to final term. The Nth term is going to be E to the two pi I N minus one over N. That's the basic vector or discrete complex exponentials. So its Nth component – so omega, the Nth component of this is E to the two pi I M over N. Okay? It doesn't take a great leap in the imagination to do this. Now I wanna define powers of this, powers of omega. This just collects in one place the powers of the complex exponential that appear in the definition of the DFT after the minus sign, but it views it slightly differently. Its self is a discrete signal, so you can either view it as an N vector, or you can view it as a discrete signal defined on the integers from zero to N minus one. I wanna take powers. Now you say to yourself, what does it mean to take powers of a vector? Well, it doesn't make sense to take powers of a vector, but if you believe in MATLAB it does, I suppose. And it certainly makes sense to take powers of a discrete function, so omega to the – I wanna use the same notation I have in the notes here. Omega to the N is just the discrete signal whose entries are the Nth powers of the entries of omega, so it's one E to the two pi I N over N, E to the two pi I two N over N, up to E to the two pi I N N minus one over N. And of course, if I can take positive powers, I can also take negative powers. N is not meant to be either – we're sticking to the positive here, but just to write it down, omega to the minus N same thing is one E to the minus two pi I N over N, E to the minus two pi I two N over N, then all the way up to E to the minus to pi I N times N minus one over N, just replacing N by minus N.

Okay, so with this notation, the discrete Fourier transform looks a little bit more compact. It's not the only reason for doing it, but it's not a bad reason for doing it. The way this notation – we're defining the vector or the discrete complex exponential this way, we can rewrite the definition of the discrete Fourier transform. So with this, you can write the

DFT as the Fourier transform of F, the Nth component is the sum from N equals zero up to N minus one of the Nth component of little F, the function you're operating on, times omega to the minus N, the Nth component of that. I haven't done anything different here. I've just rewritten in terms of this discrete vector exponential. Omega to the minus N of M is the Nth entry of that quantity, omega to the minus N. The Nth component of this is E to the minus two pi I M N over N. Okay? Or – let me put my underlines there so we realize we're taking discrete case here. Everything's discrete. Or even more compactly, this is the Fourier transform evaluated at M. The Fourier transform of F sort of written just as an operator is the sum from N equals zero up to N minus one of F N, the values of the signal you're inputting at N times omega to the minus N. This expression is nothing but this expression evaluated on the index M. Now I actually find this actually a pretty convenient way of writing the DFT. It's a question about when you write your – it's always a question in this subject, whether it's the continuous case of the discrete case, when you write your variables and when you wanna avoid writing your variables. And the same sort of thing applies in the discrete case as well as the continuous case, so this is sort of as far as you can go in writing the DFT without writing variables. In this case, the variable that we're not writing is the Mth index. Again, this expression is nothing but this expression evaluated at the index M, or evaluated at the component – finding the Mth component.

So we're gonna work with this expression a fair amount or expression evaluate it, and try to make the – so this is probably about as close as you can make the discrete Fourier transform look like the continuous Fourier transform. The integral's replaced by the sum. The continuous exponential E to the two pi I – minus two pi I S T in the continuous case is replaced by the discrete exponential. That's about as close as you can get, and it's pretty close actually. For a lot of practical purposes, for a lot of formulas, for a lot of computations, it's pretty close. We're gonna make a lot of use of it. Now I told you last time that there are just lots of little things that you have to kind of digest, most of which are analogous to properties in the continuous case. So I'm afraid that really what I have to do now is go through almost a list of different points, little points that come up, and I can't do all of them. So again as I pleaded last time that I want you to read through the notes and sort of hit those points, some of which we'll talk about, some of which we won't talk about, and I decided actually to give a slightly – to reorder things slightly from the notes just for a little variety here. So I wanna derive many of the – I'll derive all the same formulas. Nothing's gonna be different, of course, but I wanna do it from a slightly different tack. I mean I like the way I actually wrote it all out in the notes, but it does take a little more time, and I wanted to here in class try to sort of hit the high points a little bit more quickly. So again we have – there's no way of getting around this. We just have to have a certain list of properties that come out as consequences of that formula that we need in order to be able to sort of use the DFT day to day. So the very first one – so list, many properties, little properties.

The first one actually is something – I'll mention it now, but I'll talk about it next time – is something that's different between the discrete case and the periodic and the continuous case, and that is the periodicity of the inputs and outputs. So I'll do this next time. I'll talk about this next time, but I wanted to mention it here because it really is sort of the first thing you should establish about the DFT. And here, despite all my big build up about how similar they are, this is a difference actually between continuous and discrete cases. Now what I mean by this, and again I'll say this in more detail next time, is that initially you have a discrete signal you feed in, and you get a discrete signal out. You have a signal defined from zero – indexed from zero to N minus one. The signal that comes out of that is also indexed from zero to N minus one, but as it turns out, you are really compelled by the definition of the DFT to extend those signals to be period of period capital N, that is to say to be defined on all the integers. So we'll leave that up there because it really depends on the formula. The definition of the DFT compels you to regard the input little F and the output capital F, its discrete Fourier transform as not just defined on the integers from zero to N minus one but as periodic discrete signals of period N. That is so because the vector exponential itself turns out to be naturally a periodic signal of period N, discrete periodic signal of period N. But again, I'll do the details next time, but I wanted to highlight this because this is how – this is sort of the approach taken in the notes, and it really [inaudible] high points [inaudible] this is so [inaudible] omega itself, its discrete complex exponential should be – is naturally a periodic discrete signal of period N. And more on this next time.

This actually turns out to have significant consequences. The fact that you really have to – you know, it's all – whether or not you use it in a particular problem or a particular setting, it's lurking in the background. If the input signals are periodic and the output signals are periodic. And in some cases actually, it has consequences for computation. Some computations, if it's not taken into account, you can get results that are not in accord with what you might expect, and it usually has to do with not taking into proper account the periodicity. Sometimes you have to shape a signal a little bit differently to take that into account and so on, so more on this next time, but I did wanna mention it. The main thing that I wanna talk about today is the orthogonality of the complex – of the vector – of the discrete complex exponentials and its consequences. So two, the other sort of little fact – not so little fact, actually – is the orthogonality of the discrete complex exponentials. Most sort of non-trivial or slightly less than trivial – most interesting properties of the DFT can be traced back to what I'm about to talk about now. I'll put that in quotes. Properties of the DFT involve this property. Now what is the property? So let me give you a set up, and I'll tell you what I mean. This would be harder to – or not as – couldn't be formulated in as easy a way if I didn't introduce this vector complex exponential. If I stayed writing just the exponential terms themselves, either the two pi I, blah blah blah, then it would be a little more difficult, a little more awkward to formulate the property, but if you introduce the discrete complex exponential, there's a very nice way, a very nice property that they have that turns out to be very important for a number of reasons. So again, omega is one E to the two pi I over N up to E to the two pi I N over N – N minus one, sorry, over N. And the powers that we look at are omega to the K say is one E to the two pi I K over N, E to the two pi – I'll write one more term out here – two pi I two K over N, up to E to the two pi I K times N minus one over N. The orthogonality of the discrete complex exponentials is really the orthogonality of the powers of these complex exponentials. That is to say if K is different from L, then omega to the K and omega to the L are orthogonal. Now here when I say they're orthogonal, I mean not so much thinking of them as discrete signals, but thinking of them as N vectors. Now I wanna show you why that works, and also what happens when K is equal to L because that's where much heartache comes.

So let's compute omega to the K – to say they're orthogonal means I have to compute their inner product. All right, we haven't looked at inner products much since the very beginning of the course when we talk about Fourier series, but they're coming back now, and I'll remind you of the definition in this particular case. So I'm taking omega to the K dot omega to the L. It's a complex inner product of these two vectors, so what is that? That's the sum from say N equals zero to N minus one. Let me just write it, then I'll say more about it in a second. It's the Nth component of omega to the K times the Nth component of omega to the L conjugate. If I use this notation to indicate the Nth component, then the inner product of the two vectors is the sum of the products of the components, but in the complex inner product, you take the complex conjugate of the second term. So what is this? This is equal to the sum from N is equal to zero to N minus one. Let me write it through with the complex exponentials now, the usual complex exponentials. E to the two pi I K N over N times E to the two pi I L N over N conjugate which just puts a minus sign in the complex exponential. Keep me honest on my algebra here so I don't make any slips. So that's the sum from N is equal to zero to N minus one of E to the two pi I K N over N times E to the minus two pi I L N over N. Right? And you have to realize that what you have here is a geometric series. That is this is – let me write this a little bit differently. This is the sum from N is equal to zero to N minus one of E to the two pi I K minus L raised to the N – oh, divided by N, sorry. K minus L divided by N all raised to the N. Exponentials being what they are, this is either the two pi I K N over N, either the two pi I – minus E to the two pi I L N over N. I group those terms two pi I K minus L divided by N raised to the Nth power.

That's a geometric series, finite geometric series, and we know what its sum is. Okay? We know how to sum that. So if K is different from L – if K is not equal to L then this sum is equal to one minus E to the two pi I K minus L divided by N raised to the Nth power – let me write it like this – divided by one minus E to the two pi I K minus L divided by N. Nothing up my sleeve. Make sure I did this right. I believe so. Make sure I wrote it neatly enough. So it's one plus R plus R squared plus R cubed up to the R minus one where R is this exponential. So it's one minus the thing to the Nth power at N terms – zero to N minus one is N terms – raised to the Nth power divided by the thing that's getting raised to the power, one minus the thing that's getting raised to the power. There's no problem with the denominator here because K is different form L, and K minus L is always less than N, so these are distinct. This is never equal to one down below, but on top, if I raise it to the Nth power, it is one minus E to the two pi I K minus L divided by one minus E to the two pi I K minus L divided by N. [Inaudible] an integer, E to the two pi I K minus L is one, so this baby is zero. So if K is different from L, omega to the K dot omega to the L, the inner product of the discrete complex exponentials is equal to zero. What happens if K is equal to L? If K is equal to L, go back to the sum here. If K is equal to L, then I'm getting E to the zero, so I'm just getting a sum of a bunch of ones. How many ones do I have? N of them. So if K is equal to L, omega to the K dot omega to the L is equal to N.

To summarize, omega to the K dot omega to the L is equal to zero if K is different from L, is equal to N if K is equal to L. In the notes, this is done a little bit more generally because I allow for periodicity, and instead of saying K is equal to L or K is different from L, you say K is different from L modulo capital N. Here you say K is equal to L modulo capital N. It doesn't really – that follows out from the general discussion. We'll get to that next time, but this is the sort of encounter with it. This is probably the easiest way of thinking about it, and it's an extremely important property, the orthogonality of the vector exponentials. Now it's a fact or it has been a fact – traditionally in the electrical engineering department – I say this. Listen very carefully. It's a Quals tip. Bob Gray, my dear colleague Bob Gray used to always ask Quals questions that somehow reduced to the orthogonality of the discrete complex exponentials. Somehow, that was always involved in his Quals question. All right. So now you know. I don't know if he still does it since I've been making a big deal out of this the last couple years, but somehow he would always manage to ask a question that reduced to this or somehow involved this in a crucial way. It's very important. Now the thing that makes – now here's another difference between the continuous case and the discrete case that shows up in a lot of formulas, and it leads to much heartache and much grief – the fact that they're orthogonal but not orthonormal, all right? The length of the vectors is not N, it's the – is not one, it's the square root of N. What this says is to put this another way the length, the norm of this vector exponential, discrete complex exponential is the inner product of the vector with itself, so omega K squared is omega to the K dot omega to the K, and that's equal to N. So the length squared – and the length is equal to the square root of N.

This fact causes an extra factor of N or one over N to appear in many formulas involving the DFT. It can all be traced back to this. It is a royal pain in the ass. I'm sorry to have to report that to you, but it is, and it always traces back to this. Causes a factor of N or one over N to appear in many formulas. Sorry. The way we define the DFT it does. There are ways of sort of getting around it, but it's sort of awkward, and once again like in most things, there's no particular consensus as to what is best here, but the way we're defining the DFT which is pretty standard, you have this sad fact that the vector complex exponentials are orthogonal but not quite orthonormal. They cause that extra factor to come in. Now let me give you the first important consequence of this. The first important consequence of the orthogonality of these discrete complex exponentials is the inverse DFT. So a consequence of the orthogonality is a simple formula for the inverse DFT. Now again in the notes, I did this a little bit differently. I mean I wound up with the same formula, but I did it a little bit differently, and I sort of liked it again because it was a way of discovering what the formula should be independent actually of the discrete – although that comes into it – of the orthogonality of discrete complex exponentials and so on. But again just to do things a little bit differently so you have two different points of view here, let me just give you the formula and show why it works. I don't like doing that because I don't like the sort of deus ex machina aspect of it where you write down this formula and say, "Son of a bitch, it works." I sort of instinctively don't like that, but I'm gonna do it anyway. Son of a bitch, I'm gonna do it. And that is – I don't wanna – because I wanna show you how it's a consequence of this orthogonality relationship. It's quite nice. Here's what we find. The inverse DFT is given by the inverse Fourier – I'll just write down the formula. I'll write it over here in its full glory. Again, I'll put an underline under it, so again apply to some signal F is the Mth component of that is one over N times the sum from N equals zero to capital N minus one of little F and N, and this time you take positive powers, E to the plus two pi I N M over N. Or written a little bit more compactly in terms of vector exponentials, the inverse Fourier transform of F is one over N – at M is one over N times the sum from N is equal to zero to N minus one of F N – omega to the N, its Mth component, same thing, omega to the positive power N. Or written more compactly still if I drop the variable, inverse Fourier transform of F is the sum from N is equal to zero to capital N minus one of F N times omega to the Nth. Oh, I forgot the painful factor one over N.

So this is as close as you can get for the inverse discrete Fourier transform to look like the inverse continuous Fourier transform, and the difference among other things is this irritating factor of one over N, and it comes in there exactly because the vector complex exponentials are orthogonal but not quite orthonormal. Okay? But now I'll show you why it works. I'll show you why this is the inverse. Again, if you look at the discussion in the notes, this formula emerges from an analog of the continuous case. We talk about reverse signals, and duality, and all those things we're gonna talk about – well, maybe not talk about all of them, but I want you to read all of them because it's another series of little points that you have to sort of digest. And if you do that, if you pursue the analogy actually with the continuous case via duality and all the rest of that stuff, actually this formula emerges quite nicely from that, independent of first talking about the orthogonality of the vector complex exponential. But since we've done it this way, let me show you how they come in. All right, so I have to show you if I have time here that the inverse Fourier transform of the Fourier transform of F is equal to F. Or I have to show you the inverse Fourier transform of the Fourier transform – let me put a variable in here – at the M, the Mth component is the Mth component. That's what it means for the thing to be the inverse Fourier transform. Okay. So let's do that. We have no recourse here other than to appeal to the formula, and I want you to see how this orthogonality comes in. I have to be careful with my indices here. So the inverse Fourier transform of the Fourier transform of F is one over N times [inaudible] at M – is one over N times the sum from N is equal to zero up to N minus one of the coefficients of the thing I'm putting into it which is the Fourier transform times omega two – or maybe I'll write in terms of complex exponentials here – times E to the two pi I N M over N. Right? Okay. Right? Is that correct? That's what the Fourier – the inverse Fourier transform looks like that?

Student:[Inaudible].

Instructor (Brad Osgood):Right. Okay. Now I can begin the formula for the Fourier transform. I'm just sort of debating in my own mind here, which you can't hear whether or not I should write it in terms of the complex exponential, or write it in terms of the omegas. I started with this, so I'll keep this us. It may be a little cleaner if I use the other notation for it, but never mind. Let's go forward. So now I have to write this sum from N equals zero up to capital N minus one, so I need to bring in the formula for the Fourier transform. That's the sum say of K equals zero up to N minus one of F of K. The Nth component of the Fourier transform is the sum from K equals zero to N minus one F of K E to the minus two pi I K N over M, capital N, and then times E to the two pi I N M over capital N. Check me. Make sure I haven't slipped anything past you here. Did I remember my factor one over N in front? I did. So once again, all I did here was substitute the formula for the Fourier transform, the Nth component of the discrete Fourier transform of F is this sum. The power here N means I'm taking the Nth component of the Fourier transform of the left. Good. And now you can combine everything here. What I'm gonna do – guess. I'm gonna swap the order of summation. Well, I'm gonna put everything together and then swap the order of summation. So that's the sum from – it's one over N times from the N equals zero – K equals zero one over N – sum from N equals zero to N minus one, sum from K equals zero to N minus one, and F of K times E to the two pi I N, K over N times E to the two pi I N M over N, and I see I got this wrong. This is a minus here, right? Minus, okay.

Okay, now swap the order of summations, like swapping the order of integration, a technique we used many times. So this is one over N sum from K equals zero of the sum from L and the sum from N equals zero. Now this depends only on F of K. It does not depend on N, so I'm gonna bring that out of the sum, F of K times the sum from N equals zero up to N minus one, and what remains are these products of complex exponentials, E to the minus two pi I N K over N times E to the two pi I N M over N. Wonderful. You have to have a certain taste for this. Now what you should recognize in that sum is the inner product of – this is the powers of the vector complex exponential times its conjugate. E to the plus two pi I N M over N, E to the minus two pi I N K over N, it is the Mth vector complex exponential – the Mth power inner product with the Kth power. That's what this sum is summed over N, the Kth power of omega inner product with the Mth power of omega. And that is either equal to zero if M is different from N – M is different from K, or N if M is equal to K. So that sum, N equals zero up to N minus one of E to the minus two pi I N K over N times E to the plus two pi I N M over N is either equal to zero if K is equal to M, and it's equal to N if – if K is different from M, and equal to N if K is equal to M. So only one term survives here, and then it's the outside sum here. So the only term that survives is when K is equal to M in which case you get N times the one over N, and if the only term that survives is when K is equal to M, you get F of M. And we're done about. So the only surviving term is K equals M, and you get the inverse Fourier transform of the Fourier transform of F at M is equal to F of M, and that is to say the inverse Fourier transform given by this formula really is the inverse of the Fourier transform. Works like a charm, and it depends crucially on the orthogonality of the complex exponential, crucially on that, like many of the properties. And the only thing that makes it a little bit painful is this extra factor of N that comes in but that's just life. That's just the way it goes. That's it for today. Next time, we'll do more little interesting facts about the DFT.

[End of Audio]

Duration: 53 minutes

TheFourierTransformAndItsApplications-Lecture21

Instructor (Brad Osgood): Here we go. All right. The exams, you remember the exams, I think they've all been graded, and the scores have all been entered. Although, I don't think we've made the scores yet visible on the web site. I will do that after I get back from class today. You can pick up your exams from Denise Murphy, my admin in Packard 267. She has them all, so I'll send that announcement out to the class. Any questions or comments on that?

Good. Okay. Today, we're going to continue with our discussion of the DFT. This is getting to know your DFT. Your Discrete Fourier Transform. Now, the subtitle, I would say for this, should be – you already know it. The point of the way we're talking about the Discrete Fourier Transform is that it can be made to resemble the Continuous Fourier Transform in a great many ways. So the intuition that you built up for the Continuous Fourier Transform, the formulas that you've learned to work with and so on really all have analogues in the Discrete case. Now, not in all cases. There are some things that don't quite match up, and that's interesting, too, but to realize what doesn't match up so carefully and to realize why it's interesting, I think, goes along with seeing in many ways how much things are the same. So we're going to take the point of view that – and I'm going to try to take the route to make the Discrete Fourier Transform look as much as possible like the Continuous Fourier Transform. That's our point of view. You don't have to do it this way, but I think it's most satisfactory, and it allows us to leverage off what we have done in the continuous case. Now, before doing this, let me recall the definition. Here's the definition. Where we ended up last time was sort of a definition that, for the Continuous case, the idea of sampling a continuous signal to get a discrete signal, sampling a Fourier Transform to get a Discrete Fourier Transform, all that had pretty much vanished by the final definition. So the definition that we ultimately wound up with and ultimately had makes the continuous case almost invisible, vanish.

It looked like this. You have a Discrete signal. So I'm using both the signal notation and vector notation here, and I'll continue to do that, sort of mix the two up, because I think they're both useful. So the idea is you have either an N-tuple of numbers or a Discrete signal whose value at the nth point is just the value here, FM. Okay? Oops, that doesn't look right. That's not much of a statement. So you can either consider it as a Discrete signal who's defined on the integers or the integers from zero to N minus one, or you can think of it as an N-tuple or as a vector. So you have a Discrete signal, and the Fourier Transform is another Discrete signal. It's DFT. It's the discrete signal – I'll call if capital F, but I'll also use the notation to make the connection with a continuous case, a script F with a little underline under it indicating it's supposed to be a vector quantity or something a little bit different from the continuous case. It's defined by its nth component. So the nth component of the Fourier Transform is the sum from N equals zero to N minus one. The nth component of F times E to the minus two pi I, NM over N. So everything is defined here in terms of the indices in the exponential, and these are the values of the Discrete Function at the index point, F of zero, F of one, F of two and so on.

That's the definition. You don't see at all the fact that in our derivation, this came from, starting with a Continuous signal, sampling it, sampling the Fourier Transform and then somehow, ultimately, leading to this definition here. It's just as an operation on one Discrete signal producing another Discrete signal. That's pretty much how we're going to deal with it. But before embarking on that path forever, I want to have one nod back to the Continuous case. I'm going to talk a little bit about how the DFT is employed and the kind of things you have to know when you use it in practice. You already have to some extent, and you certainly will more in the future. So we want to look back at the Continuous case to talk about one additional phenomenon of reciprocity that comes – the reciprocal relationship between the time domain and the frequency domain that comes up also in the Discrete case. So one more look back at the Continuous case, the Continuous roots of the DFT and reciprocity of the two domains, time and frequency. So how did it work? We have a signal on the time domain that we discretize, and we have a signal in the frequency domain that we discretize. So imagine that you have two grids. You have a grid in the time domain. You have N sample points that are delta T apart. I'm running out of room to write things. Let me just write it here, and then I'll write some of the notation up above. This is the frequency domain, and we have spacing in the frequency domain of delta S apart. So N sample points, T zero up to TN minus one, space, delta T apart, and N sample points in time that is. And we have N sample points in frequency. S one [inaudible] say up to SN minus one, spaced, delta S apart. So there are three quantities here. There's the spacing in the time domain, the spacing in the frequency domain and the number of sample points, N. There's a relationship between them. You have three quantities. Quantities of interest. Delta T, delta S and N. They're not independent. They're not independent. That's the important point. Let me remind you again – I'm not going to go through the derivation again, but let me remind you what the relationship is. We had an N delta T. That's the spacing in the time domain that has to be – I want to say this right – is L. That's the spacing of the time domain. I want to use the notation I used last night.

This is sort of a time limitedness. We had N times delta S is the band limitedness that we called that two B. That was the bandwidth. So delta T times delta S is equal to – delta T is L over N times two B over N. That's L times two. That's two BL over N squared. You will recall that when we set up the sampling, two times B times L was equal to N, the number of sample points. This is N over N squared. That is one over N. Okay? The way we did the sampling, making this not completely justifiable, appeal to the sampling theorem and so on, gave us a relationship between how you sample in the time domain, how you sample in the frequency domain, the number of sampling points you took, and it was exactly that two BL was equal to N. It was sort of surprising or not clear, a priori, that if you carry out this procedure of sampling in the time domain and sampling in the frequency domain, you're going to take the same number of sample points in both domains.

That's what we found as a relationship there. Let me highlight it. That is delta S delta T, or delta T delta S is one over M. The spacing in the frequency domain and the spacing in the time domain are reciprocally related to the number of sample points. This is called a reciprocity relationship. All right. Now this again has practical significant and practical consequences when you're applying the DFT. You can imagine, when you have a continuous signal, you want to sample it. What can you choose and what is forced upon you? Well, you can imagine choosing how frequently you sample, right? That's the delta T, and you can imagine how many measurements you make, how often you sample, how many samples you take, which is N. So you can imagine choosing delta T. That's how frequently you sample, or the sampling rate, and N, the number of samples. If you do that, once you do that, then delta S is fixed. Delta S is determined. The spacing in the frequency domain is determined. One way of putting this is the accuracy of the resolution, or how find the resolution is in frequency, is fixed by the choices you make in the time domain.

So let's say the resolution in frequency is detailed, is fixed, by the choices you make in time. Conversely, you could say, I want a certain accuracy in – I want a certain resolution in frequency. I want delta S to have a certain fineness. That's going to detail, then, how many sample points I want to take and how I space them in the time domain. At least, you have to choose two of those things, and the third one is detailed. You can imagine, if you're doing this with real data, you have a certain freedom here, but the freedom also carries with it certain restrictions. Ain't that life? That's the way things go. The freedom you have is typically how many measurements you make and how frequently you make them. Once you do that, that determines what the resolution is like in the frequency domain. There are ways of getting around this. There's zero padding and special techniques – not getting around it, but there's ways of understanding it or massaging it. So-called zero padding, you'll have a problem on that, and you'll have other chances to experiment with that. It's built into the system. This sort of reciprocity relationship is another example of the same sort of thing we've seen so many times in the Continuous case, now carried over to the Discrete case.

It's stretched in one domain means shrunk in the other domain. Or it's reciprocity between time and frequency. As I said, we've seen many different instances of that in the Continuous case, and here's an example of how it caries over to the Discrete case. All right. I wanted to say that because it is the sort of thing that you meet, and you have to understand, when you actually apply the DFT in most common context where it's associated somehow with some sort of Continuous process that you're sampling. You want to take the Fourier Transform and you don't have a formula for it, you have to use numerical algorithms to actually compute it. That means you're sampling. The reciprocity relationship puts certain limitations on what you should expect to get. Okay. So that is my final nod, at least for now, back to the Continuous case. Let's go back now into the Discrete world and pretty much stay there. So back to this formula and it's consequences. Let me erase it and write it down again. Back to the Discrete world, the Discrete setting.

Again, my goal here is to make the Discrete Fourier Transform look as much as possible like the Continuous Fourier Transform. So here's how to do that. I want to make the DFT and associated formulas look like the Continuous case. So maybe I should say we're not really abandoning the Continuous case, but we're situating ourselves firmly on the Discrete side of things. Now, to help out in that, partly it's a matter of notation, but even more so, it's what you do with the notation and the consequences of thinking about things a certain way. So the first thing I want to do is introduce a symbol, introduce a way of thinking about exponentials that occur in the definition of a DFT. So let me write down the formula one more time. Again, F is a Discrete signal. F zero up to FN minus one. Its Fourier Transform is another Discrete signal whose nth component is given by this formula. N equals zero, N minus one, the nth component of little F times E to the minus two pi I, NM over N. Okay. That's a two pi I. First thing I want to do is I want to introduce the notation. I want to view these complex exponentials as themselves, coming from a Discrete signal, as values of a Discrete signal.

This turns out to be a very helpful thing to do for a number of reasons. It gives you a compact way of writing things. It also gives you a way of enunciating certain properties of the Fourier Transform that would be very – the Discrete Fourier Transform that would be very difficult to do otherwise. So I want to realize the complex exponentials that come in the definition here as arising from also a Discrete signal, the Discrete or vector complex exponential. People either refer to it as a Discrete complex exponential or the vector complex exponential. You pick what you want to call it. So let me give you the definition. I'll leave that word up there. Let me write it over here. I'm going to write omega as a vector or as a discrete signal to be the N-tuple of powers of the complex exponential that appear in the definition. So the zeroth coefficient, the first entry is one. That's E to the zero. Then it's E to the two pi I – there's a minus sign up here. Let me define it in terms of a plus sign, and then I'll start taking powers to get a minus sign. So either the two pi I over N. You'll see where this is going from in just a second. E to the two pi I times two over N, and so on, so on, all the way up to the final term. The nth term is going to be E to the to pi I, N minus one over N. All right?

That's the basic vector or Discrete complex exponential. So its nth component – so omega, the nth component of this, is either the two pi I, M over N. Okay? Doesn't take a great leap of the imagination to do this. Now, I'm going to define powers of this. Powers of omega. This just collects in one place the powers of the complex exponential that appear in the definition of the DFT up to the minus sign, but it views it slightly differently. Itself is a Discrete signal. So again, you can either view it as an N-vector, or you can view it as a discrete signal to find out the integers from zero to N minus one. I want to take power. Now you say to yourself, what does it mean to take powers of a vector? Well, it doesn't make sense to take powers of a vector, but if you believe in mat lab it does, I suppose. It certainly makes sense to take powers of Discrete function. So omega to the – I want to use the same notation I have in the notes here. Omega to the N is just the Discrete signal whose entries are the nth powers of the entry-level omega. So it's one E to the two pi I, N over N, E to the two pi I two, N over N, up to E to the two pi I, N-1 over N.

Of course, if I can take positive powers, I can also take negative powers. N is not restricted to be positive here, but just to write it down, omega to the minus N, same thing, is one E to the minus two pi I, N over N, E to the minus two pi I, two N over N. Then all the way up to E to the minus two pi I, N times N minus one over N. Just replacing N by minus N. Okay. So with this notation, the Discrete Fourier Transform looks a little bit more compact. It's not the only reason for doing it, but it's not a bad reason for doing it. So with this notation, we're defining the Discrete complex exponential this way, we can rewrite the definition of the Discrete Fourier Transform. So with this, you can write the DFT as the Fourier Transform of F. The nth component is the sum from N equals zero up to N minus one of the nth component, a little F, the function you're arbitrating on, time omega to the minus N, the nth component of that. I haven't done anything different here. I've just rewritten it in terms of this Discrete vector exponential. Omega to the minus N of M is the nth entry of that quantity. Omega to the minus N, the nth component of this is E to the minus two pi I, MN over N. Okay?

My underlines there so we realize we're taking Discrete case here. Everything's Discrete. Or even more compactly, this is the Fourier Transform evaluated in M. The Fourier Transform of F sort of written just as an operator, is the sum from N equals zero up to N minus one of FN. The values of the signal you're inputting at N times omega to the minus N. All right? This expression is nothing but this expression evaluated on the index, M. Now, I find this, actually, a pretty convenient way of writing the DFT. It's always a question in this subject whether it's the Continuous case or the Discrete case. When you write your variables and when you want to avoid writing your variables. The same sort of thing applies in the Discrete case as well as the Continuous case. So this is sort of as far as you can go in writing the DFT without writing variables. In this case, the variable that we're not writing is the nth index. Again, this expression is nothing but this expression evaluated at the index M, or evaluated at the component, finding the nth component.

Okay. Now, we're going to work with this expression a fair amount, or the expression evaluated and try to make the decision – so this is probably about as close as you can make the Discrete Fourier Transform look like the Continuous Fourier Transform. The integral's replaced by the sum. The Continuous exponential, either the minus two pi I ST in the Continuous case is replaced by the Discrete exponential. That's about as close as you can get. It's pretty close, actually. For a lot of practical purposes, for a lot of computations, for a lot of formulas, it's pretty close. We're going to make a lot of use of it. Now, I told you last time that there are lots of little things that you have to digest, most of which are analogous to properties in the Continuous case. So I'm afraid that what I have to do now is just really go through almost a list of different, little points that come up, and I can't do all of them. So again, as I pleaded last time, I want you to read through the notes and sort of hit those points, some of which we'll talk about, some of which we won't talk about. I decided, actually, to reorder things slightly from the notes, just for a little variety here. So I'm going to derive many of the – I'm going to derive all the same formulas. Nothing's going to be different, of course, but I want to do it from a slightly different tack. I mean, I like the way I wrote it all out in the notes, but it does take a little more time, and I wanted to, here in class, try and hit the high points a little bit more quickly.

So again, there's no way of getting around this. We just have to have a certain list of properties that come out as consequences of that formula that we need in order to be able to use the DFT day to day. So the very first one, mini-properties, little properties. So the first one, actually, is something I'll mention now, but I'll talk about it next time. It's something that's different between the Discrete case and the Continuous case. That is the periodicity of the inputs and outputs. So I'll do this next time. I'll talk about this next time, but I wanted to mention it here because it really is probably the first thing you should establish about the DFT. Here, despite all my big build-up about how similar they are, this is a difference, actually, between Continuous and Discrete cases. Now, what I mean by this, and I'll say this in more detail next time, is that initially, you have a Discrete signal you feed in, and you get a Discrete signal out. You have a signal defined or indexed from zero to N minus one. The signal that comes out of that is also indexed from zero to N minus one. But as it turns out, you are really compelled by the definition of the DFT to extend those signals to be periodic of period capital N. That is to say to be defined on all the integers.

So I'll leave that up there because it really depends on the formula. The definition of the DFT compels you to regard the input, little F, and the output, capital F, as not just defined on the integers from zero to N minus one but as periodic Discrete signals of period N. All right? That is so because the vector exponential itself turns out to be a Discrete periodic signal of period N. Again, I'll do the details next time, but I just wanted to highlight this because this is the approach taken in the notes, and it [audio cuts in and out]. This is so because omega itself is a Discrete complex exponential is naturally a periodic Discrete signal of period N. Okay? More on this next time. This actually turns out to have significant consequences, the fact that you have to – I mean, whether or not you use it in a particular problem or a particular setting, it's lurking in the background. The input signals are periodic, and the output signals are periodic. In some cases, actually, it has consequences for computation. Some computations, if it's not taken into account, you can get results that are not in accord with what you might expect. It usually has to do with not taking into proper account the periodicity. Sometimes you have to shape this thing a little bit differently to take that into account and so on. So more on this next time, but I did want to mention it.

The main thing that I want to talk about today is the orthogonality of the Discrete complex exponentials and its consequences. The other sort of little fact – not so little fact, actually, is the orthogonality of the Discrete complex exponentials. Most non-trivial or slightly less than trivial – most interesting properties of the DFT can be traced back to what I'm about to talk about now. I'll put that in quotes. The properties of the DFT involve this property. What is the property? Okay. Let me get you set up, and I'll tell you what I mean. This couldn't be formulated in as easy a way if I didn't introduce this vector complex exponential. If I stayed writing the exponential terms themselves, either the two pi I, blah, blah, blah, then it would be a little more difficult, a little more awkward to formulate the property. But if you introduce the Discrete complex exponential, there's a very nice property that they have, that turns out to be very important for a number of reasons. So again, omega is one E to the two pi I over N, up to E to the two pi I, N minus one over N. The powers that we look at are omega to the K, say, is one E to the two pi I, K over N, E to the two pi I, two K over N, up to E to the two pi I, K times N minus one over N.

The orthogonality of the Discrete complex exponentials is really the orthogonality of the powers of these complex exponentials. That is to say if K is different from L, then omega to the K and omega to the L are orthogonal. Now here, when I say they're orthogonal, I mean not so much thinking of them as Discrete signals, but thinking of them as N vectors. All right. Now, I want to show you why that works and what happens when K is equal to L because that's where much heartache comes. All right. So let's compute omega to the K. So to say the [inaudible], they have to compute their inner product. We haven't looked at inner products much since the beginning of the course, when we talked about Fourier series, but they're coming back now. I'll remind you of the definition in this particular case. So I'm taking omega to the K dot omega to the L. It's a complex inner product. It's the inner product of these two vectors. So what is that? That's the sum from N equals zero to N minus one. Let me write it out in terms of – well, I'll just write it, and then I'll say more about it in a second. It's the nth component of omega to the K times the nth component of omega to the L conjugate. If I use this notation to indicate the nth component, then the inner product of two vectors is the sum of the products of the component. But with a complex inner product, you take the complex conjugate of the second term. So what is this? This is equal to the sum from N is equal to zero to N minus one.

Let me write it in terms of the complex exponentials now. E to the two pi I KN over N times E to the two pi I LN over N, conjugate, which is – it just puts a minus sign in the second complex exponential. Keep me honest on my algebra here so I don't make any slips. So that's the sum from N is equal to zero to N minus one of E to the two pi I KN over N times E to the minus two pi I of LN over N. You have to realize that what you have here is a geometric series. That is – I'm going to write this a little bit differently. The sum from N is equal to zero to N minus one of E to the two pi I, K minus L raised to the N – oh, divided by N, sorry. K minus L divided by N, all raised to the N. Exponentials being what they are, this is E to the two pi I, KN over N, minus E to the two pi I, LN over N. I group those terms, two pi I, K minus L divided by N, raised to the nth power. That's a geometric series, finite geometric series.

We know what its sum is. We know how to sum that. So if K is different from L, if K is not equal to L, then this sum is equal to one minus E to the two pi I, K minus L, divided by N, raised to the nth power. Divided by one minus E to the two pi I, K minus L divided by N. Okay? Nothing up my sleeve, make sure I did this write. I believe so. Make sure I wrote it neatly enough. So it's one plus R plus R-squared plus R-cubed up to the R to the N minus one, where R is this exponential. So it's one minus the thing to the nth power. I've [inaudible]. Zero to N minus one is N terms. Raise it to the nth power, divided by the thing that's getting raised to the power, one minus the thing that's getting raised to the power. It's no problem with the denominator here because K is different from L, and K minus L is always less than N, so these are distinct. This is never equal to one down below. But on top, if I raise it to the nth power, it is one minus E to the two pi I, K minus L, divided by one minus E to the two pi I, K minus L, divided by N. K minus L is an integer. E to the two pi I, K minus L, is one. So this baby is zero.

So if K is different from L, omega to the K dot omega to the L, the inner product of the Discrete complex exponentials is equal to zero. What happens if K is equal to L? If K is equal to L, go back to the sum here. If K is equal to L, then I'm getting E to the zero, so I'm just getting a sum of a bunch of ones. How many ones do I have? N of them. So if K is equal to L, omega to the K dot omega to the L is equal to N. To summarize, omega to the K dot omega to the L is equal to zero, if K is different from L. It's equal to N if K is equal to L. In the notes, this is done a little bit more generally because I allow for periodicity. Instead of saying K is equal to L or K is different from L, you say K is different from L, modulo, capital N. Here you say K is equal to L, modulo, capital N. That falls out from the general discussion, but we'll get to that next time. This is the first sort of encounter with it, and this is probably the easiest way of thinking about it. It's an extremely important property. The orthogonality of the vector exponentials. Now, it has been a factor traditionally in the electrical engineering department – I say this. Listen very carefully. It's a quals tip. My dear colleague, Bob Gray, used to always ask quals questions that somehow reduced to the orthogonality of the Discrete complex exponentials. Somehow, that was always involved in his quals question.

So now you know. I don't know if he still does this since I've been making a big deal out of this the last couple of years, but somehow, he would always manage to ask a question that reduced to this or somehow involved this in a crucial way. It's very important. The thing that makes – here's another difference between the Continuous case and the Discrete case that shows up in a lot of formulas, and it leads to much heartache and much grief. The fact that they're orthogonal but not orthonormal. The length of the vectors is not one. It's the square root of N. What this says is, to put this another way, the length, the norm of this vector exponential, Discrete complex exponential, is the inner prerogative of the vector itself. So omega K squared is omega to the K dot omega to the K, and that's equal to N. So the length squared is N. The length is equal to the square root of N.

This fact causes an extra factor of N or one of N to appear in many formulas involving the DFT. It can all be traced back to this. It is a royal pain in the ass. I'm sorry to have to report that to you, but it is. It always traces back to this. It causes a factor of one over N to appear in many formulas. The way we define the DFT, it does. There are ways of sort of getting around it, but it's awkward. Once again, like in most things, there's no particular consensus as to what is best. But the way we're defining the DFT, you have this sad fact that the vector complex exponentials are orthogonal but not quite orthonormal. They cause that extra factor to come in. Now, let me give you the first important consequence of this. The first important consequence of the orthogonality of these Discrete complex exponentials is the inverse DFT. So a consequence of the orthogonality is a simple formula for the inverse DFT.

Now again, in the notes, I did this a little bit differently. I mean, I wound up with the same formula, but I did it differently. I sort of liked it because it was a way of discovering what the formula should be independent of the orthogonality of Discrete complex exponentials and so on. But again, just to do things differently so you have two different points of view here, let me just give you the formula and show why it works. I don't like doing that because I don't like this [inaudible] aspect of it where you write down this formula and say, son of a bitch, it works. I instinctively don't like that, but I'm going to do it anyway. I'm going to do it. I want to show you the consequence of this orthogonality relationship. It's nice. Here's what we find. The inverse DFT is given by the inverse – I'll just write down the formula in its full glory. I'll put an underline under it. So apply to some signal F is the nth component of that. It's one over N times the sum from N equals zero to NM minus one of little F and N. This time, you take positive powers. E to the plus two pi I NM over N.

Or, written a little bit more compactly, in terms of vector exponentials, the inverse Fourier Transform of F is one over N at M. It's one over N times the sum from N is equal to zero to N minus one of FN, omega to the N, its nth component. Same thing. Omega to the positive power of N. Or written more compactly still, if I drop the variable, inverse Fourier Transform of F is the sum from N is equal to zero to capital N minus one of FN times omega to the N. I forgot the painful factor, one over N. So this is as close as you can get for the inverse Discrete Fourier Transform to look like the inverse Continuous Fourier Transform and the difference, among other things, is this irritating factor of one over N. It comes in there exactly because the vector complex exponentials are orthogonal but not quite orthonormal. But now, I'll show you why this is the inverse. Again, if you look at the discussion in the notes, this formula emerges from an analogue of the Continuous case. You talk about reverse signals and duality. All those things, we're going to talk about. Maybe not talk about all of them, but I want you to read all of them because it's another series of little points that you have to digest. If you pursue the analogy with the Continuous case, via duality and all the rest of that stuff, this formula emerges quite nicely from that, independent of first talking about the orthogonality of the vector complex exponentials.

But since we're done it this way, let me show you how they come in. So I have to show you, if I have time here, that the inverse Fourier Transform or the Fourier Transform of F is equal to F. Or I have to show to the inverse Fourier Transform or the Fourier Transform – let me put a variable in here – at the nth component is the nth component. That's what it means for the [inaudible] inverse Fourier Transform. Let's do that. We have no recourse here other than to appeal to the formula. I want you to see how this orthogonality comes in. I have to be careful of my indices here. So the inverse Fourier Transform of a Fourier Transform of F at M is one over N times the sum from N is equal to zero up to N minus one, up to the coefficients of the things I'm putting into it, which is the Fourier Transform, times omega to – maybe I'll write it in terms of complex exponentials here. Times E to the two pi I NM over N, right? Okay. Is it correct? The inverse Fourier Transform looks like that. Okay. Now, I can begin the formula for the Fourier Transform. I'm just sort of debating in my own mind, which you can't hear, whether or not I should write it in terms of the complex exponential or write it in terms of the omega. I'll keep this up. It might be a little cleaner if I use the other notation form, but never mind. Let's go forward.

Now I have to write this sum from N equals zero up to capital N minus one. So I have to bring in the formula for the Fourier Transform. That's the sum of K equals zero up to N minus one of F of K. The nth component of the Fourier Transform is the sum from K equals zero minus one, F of KE to the minus two pi I, KN over M. Then times E to the two pi I, NM over M. Check me. Make sure I haven't slipped anything past you here. Did I remember my factor one over N in front? I did. All right. So once again, all I did here was substitute the formula for the Fourier Transform. The nth component of the Discrete Fourier Transform of F is this sum, right? The power here, N, means I'm taking the nth component of the Fourier Transform of little F. Good.

Now, you can combine everything here. What I'm going to do, guess. I'm going to swap the order of summation. I'm going to put everything together, and then I'm going to swap the order of summation. That's the sum from – it's one over N times the sum from N equals zero to K equals zero. One over N, sum from N equals zero to N minus one. Sum from K equals zero to N minus one, F of K, times E to the two pi I, N – K over N times E to the two pi I. NM over N. And I see I got this wrong. This is a minus here. Now, swap the order of summation. It's like swapping the order of integration, a technique we've used many times. So this is one over N, sum from K equals zero of the sum from LN – the sum from N equals zero. Now this depends only on – F of K does not depend on N, so I'm going to bring that out of the sum. F of K times the sun from N equals zero up to N minus one. What remains are these products of complex exponentials. E to the minus two pi I, NK over N, times E to the two pi I, NM over N. Okay? Wonderful. You have to have a certain taste for this. Now, what you should recognize in that sum is the inner product of – this is the powers of the vector complex exponential times its conjugate. E to the plus two pi I, NM over N. E to the minus two pi I, NK over N. It is the nth power inner product with the Kth power. That's what this sum is, summed over N. All right? The Kth power of omega inner product with the nth power of omega. That is either equal to zero if M is different from K, or N if M is equal to K. So that sum, N equals zero up to N minus one of E to the minus two pi I, NK over N, times E to the plus two pi I, NM over N, is either equal to zero, if K is different to M. And it's equal to N if K is equal to M. Okay?

All right. So only one term survives here, and then it's the outside sum here. So the only term that survives is when K is equal to M, in which case you get N times one over N, times – and if the only term that survives is when K is equal to M, you get F of M. And we're done. The [inaudible] surviving term is K equals M, and you get the inverse Fourier Transform of the Fourier Transform of F at M is equal to F of M. That is to say the inverse Fourier Transform given by this formula really is the inverse of the Fourier Transform. Works like a charm, and it depends crucially on the orthogonality of the complex exponentials. Crucially on that, like many other properties. The only thing that makes it a little bit painful is this extra factor of N that comes in, but that's just life. It's just the way it goes. Okay. That's it for today. Next time, we'll do more interesting facts about the DFT.

[End of Audio]

Duration: 53 minutes

TheFourierTransformAndItsApplications-Lecture22

Instructor (Brad Osgood):All right, by popular demand, today, I'm going to talk about the basics of the fast Fourier transform algorithm, the famous FFT. We're not gonna do it in all detail, that is, I'm not gonna carry it out to the bitter end. And in fact, it's actually a highly refined art. It's almost more art than science these days. There's actually, these days, not one single FFT algorithm. There are a number of them. They share common ideas, they share similar properties, but they are often tailored to particular applications for each context.

But what we're gonna talk about today is the main idea that underlies any of them and it's the one that was sort of originally put forth by Cooley and Tukey back in the '50s I guess it was, something like that. I can't remember the dates on this.

It's very interesting. You should read about the history of these things and as you've probably heard, I mentioned in the notes, the idea behind the fast Fourier transform algorithm actually was known to Gauss way back when, when he was trying to carry out calculations of estimating the orbit of an asteroid. So a lot of interesting history.

Now, what this means in particular, I have to give you a little caution on this, is that there's sort of one more general topic on this fast Fourier transform that I would ordinarily talk about and that is convolution, circular convolution, so I'm not going to do that. I will leave that up to you to read. There's not much. There wouldn't be much to do other than talk about the formulas and basic applications. The applications we'll do in the context of linear systems coming up, so you'll see some of that later.

But the definition of convolution and some of the basic properties of convolution again, is they're similar to the continuous case. I'll let you just read the notes about them, all right? And I think there are some problems on that. I think there are already some problems on that, actually.

Okay, so that's the plan. This is going to put us a little bit behind because I was going to start talking about it sooner. I was gonna start talking about linear systems, linear time and variance systems, so we'll try to trim things a little bit after we leave the discreet world, after we leave this discussion of DFT.

So let me explain what's going on here. Let me set the stage for you. Here's the FFT algorithm. Now, let me remind you that – well, there are a couple ways of looking at it. Maybe the simplest way of understanding what the challenge is, as I mentioned last time, to write the DFT as an N x N matrix. So you can write the DFT as an N x N matrix. It's a very simple entrance, very simply described. F is just Omega to the minus NM, where remember I'm using this notation now just to simplify things so I don't have to constantly write complex exponentials.

So Omega is the primitive N root of unity, so Omega to the minus nm is just e to the minus 2pi nm over capital N. And the way we index things is from 0 to N minus 1. Indexed from 0 to N minus 1.

So to compute discreet Fourier transform, the discreet signal is to multiply the matrix by the column vector, is to regard F as a column vector and multi-matrices. All right, so there is an end-by-end matrix that would seem to require N2 computations. Never mind that you just also have to add up the N elements, but the main computation and intensive aspect of it is the multiplication.

So this requires – would seem to require – so let me just use the big O notation here, meaning that some multiple of N2 to the dominant term is N2, O(N2,) capital O(N2), big O(N2) operations.

And how could it require any less than this? I mean everything has to be accounted for here. I mean you can't just throw things out while computing the DFT. It would seem to require every element of the matrix be involved, when in fact, what the FFT algorithm does is reduce this enormously. The FFT algorithm reduces this to big O of N log N operations, usually meaning log base 2 here, but it doesn't really matter.

Now, if we're saying why something like this might be possible because it's not obvious, you have to realize that that's an enormous savings in computation. To go from N 2 to N log N is a huge savings. For example, it's a huge savings – so N is equal to 103, if N is equal to 1,000, and of course, N2 is a million, that N log N is 103 log 103, log of 103 is 3, so it's just 3 times 103. So it's a thousand operations, or a constant times a thousand operations, versus a million operations. All right, that's an enormous savings. And of course, the larger N is, the savings are even more dramatic.

Now, why should you be able to do this? Well, there are a couple of different sort of intuitive reasons why this should be the case. The intuition is one thing; carrying it out is something else. Let me tell you one way of thinking about it that we're not going to do, although it's discussed a little bit in the notes, and this is a very common way of approaching the DFT and the FFT algorithms, and that is sticking with a matrix, notation of the matrix idea.

One way is that there are a lot of structures to the matrix and there's a lot of structure to the DFT, so when you reduce it, you want to exploit the structure in the DFT. Well, that's not – that's a vacuous statement. Of course, you're going to exploit the structure in the DFT. It depends on only parameter. It depends on the parameter Omega and this complex exponential, and that has a lot of nice algebraic properties via ray of properties of the complex exponential. That's the main thing that's involved.

Now, in the matrix approach, what this amounts to, the reason why there can be such a dramatic reduction is that you can factor the DFT into a product of other matrices, which have lots of zeroes in them. So one approach is to write the DFT matrix as a product of simpler matrices; that is, to use the algebraic properties of the exponential, to carry out this matrix factorization. The simpler matrices have lots – as it turns out, the simpler matrices have lots of zeroes in them and when matrices have lots of zeroes in them, they effectively don't count in the calculation. When you multiply by zero, you're not doing anything. Simple matrices have lots of zeroes, so fewer multiplications are required. Fewer multiplications are required to carry out the matrix multiplication, to carry out the product. That's one reason, that's one approach to the FFT algorithm, and in fact, that's explained in the notes. I'm not gonna do this; I'm not gonna go through that particular approach, but that's often done, all right? And in some of the sort of highly refined approaches to the FFT, you often find they're presented in terms of matrix factorizations, and if any of you have taken 263, or are taking 263, you know that matrix factorization is a big business. Writing a matrix, writing a complicated matrix as a product of simpler matrices is something that can often, not only simplify a calculation, but can also often really illuminate some ideas that you wouldn't see, that are a little bit, that are somewhat subtle. They can't be somewhat subtle.

So it's a standard approach that we're not gonna do, but you can easily find many references to the literature on this approach. We are gonna take really a more directly algebraic approach that's based on the formula. So although I'm thinking in terms of writing a matrix, although we did the basic counts of the number of operations in terms of the matrix, I'm actually not going to work with a matrix to develop the algorithm. I'm just gonna work with the formula for it for the DFT as we have written it before and I wanna use the – I wanna exploit the algebraic properties of the complex exponential to write it in a more – I don't know – simple form or just – well, you'll see. You'll see.

It's one thing to have the idea and it's not an obvious thing to have the idea. It's not a thing to really carry it out. It's really quite beautiful and I say this as someone who's orientation in life is not Fourier algorithms necessarily, so I really have a grand appreciation for it and the whole idea behind it.

All right, now, let's write down the formula. So once again, as always, the Fourier transform of the discrete signal is given by – we write it like this. The F component – when we write it out like that – is the sum from – so I'm indexing from 0 to N – 1, so the sum from n = 0, N – 1 of F(N), the N component of F, times – if I use this notation for the complex exponential, just the Omega to the minus N x N. It's a simple formula. All right, now, what we're gonna do is we are going to write the sum as a sum over even and odd powers. We're gonna write the complete sum as a sum over even and odd indices and then also, therefore, powers of Omega. Well, let me just say even/odd indices. And in doing so, it's the algebra that's gonna allow us to do this in a helpful way. You can say I'm gonna do that. It's one thing to, again, it's one thing to say that; it's another thing to actually give you any resulting savings. You'll see in just a sec.

The purpose of this is to try to write the N for order DFT as a combination of two DFTs of order N over 2, capital N over 2. The purpose is to write the order N DFT as a combination, really a sum – there's a little bit more to it than just a straight sum – as a combination of two DFTs of order N over 2, all right, and then iterate. And as you write these DFTs of order N over 2, you apply the same rules to the DFT of the order N over 2 to write them as combinations of DFT of order N over 4 and then iterate and so on and so on and so on. Okay? Now, so first of all, to make this work, you have to assume that N is even and if you want to keep iterating, you have to assume actually that N is a power of 2. So to do this, you need to assume that N is even. Even to get started, we need to assume it's even. To iterate, you need to assume that N is a power of 2.

So this is where all these special things come in when you have a signal and you wanna apply it and you wanna compute numerically a discrete Fourier transform, and your signal, your discrete signal doesn't happen to have the length of a power of 2. What do you do? They zero padding; they do all sorts of things like this. There actually are other Fourier transform discrete Fourier transform algorithms, or DFT algorithms, that don't require this, but they all require something like this. So sometimes you have to modify your signal and again, I'm not gonna get into this. I'm just gonna talk about the algorithm per se assuming all these things. But there are ways of trying to address this and they can introduce some extra complications, so if the signal is not a power of 2, you add a bunch of zeroes, so-called "zero padding," until you get it to be a power of 2. Of course, that's not the original signal, so you've changed things a little bit and have those changes affected things? Well, that's something that depends, again, on the particular context. So we're gonna make this assumption. So sue me. I'm gonna assume this now. I'm gonna assume that N is even and then in general to iterate, I'm gonna assume that N is a power of 2. Okay, now, it depends – the algorithm and this way of arranging things, way of rearranging things, depends on the algebraic properties of the complex exponential. I want to introduce some notation here. Sorry; you know I'm always introducing some extra notation, just so I can keep track of things without – a little bit more easily, a little bit more neatly. I wanna keep track of things, so I use a notation to keep track of powers, really, of the complex exponential. That's what's involved. All right, so let me write – sorry, but let write this. I'm only using it for this performance only. Omega P (q) is e to the 2p q over p, because what's involved is both the denominator here and the power that you're raising it to.

I'm gonna use that notation and I'm gonna write my Fourier transforms in terms of that notation. Now in terms of this, how we write down the usual properties of complex exponentials because I'm gonna use them all, well, if I put down p, say q1 + q2, that's taking e = 2pi(q1+q2) over p, and of course, complex exponentials being what they are, that multiplies. So this is Omega p (q1) times Omega p (q2). Nothing there. Absolutely everything I do today is gonna be – is dead simple, okay? A high school student could do this if you explain to them at every stage what to do. But there's nothing here that's at all involved. It's just extremely clever. Now, let's look what happens when I change the power, or change the order maybe I should say. First, by the way, let me just comment here. So Omega, for the terms that are entering into the Fourier transform, DFT, Omega and n and m, minus nm is what is the power that occurs here, so that's Omega to the minus nm that is equal to each of the e = 2pi nm over N. Okay, so what is our – I wanna follow my notes here because I don't wanna get this wrong. So Omega, so I'm assuming that N is even, capital N is even, so what is Omega, the sort of N over 2nd order when I have such a power here, which is what it comes up nm in the transform, so that is – let's work it out – e to the minus 2pi nm divided by N over 2. So it's 2p nm over N over 2.

The two comes upstairs. This is e to the minus 2pi times 2nm divided by N. That is to say, if in terms of my notation, this is Omega, the F order complex exponential weighs to the power of minus 2nm. So in words, a power of the half order exponential is an even power of the four exponential. That's one way of saying it. If I think of N as sort of the order of exponential that I'm worried about, then in words, say a power, in this case nm, half order exponential, that's N over 2, is an even power, that is, 2nm, of the 4 exponential that's N.

I think I'm developing tennis elbow, by the way, over the course of this quarter. I'm in a lot of pain, but I can play through it. All right. So that side is the even powers. What about odd powers? Well, odd powers – I'm thinking about N being odd. Let's go over to odd powers, the power being 1. So then I'm gonna look at Omega – careful, careful here; careful, careful here – Omega N of 2n + 1. Now be careful. I don't want to get any wrong. 2n minus – 2n + 1 because all of my powers are minus m. So that's an odd power of N. See, once again, powers of N over 2 here led to even powers of N, so I wanna look at what happens to odd powers of N and relate those to certain powers of N over 2. So what is this? This is Omega N – 2nm – m, okay, right. Now, remember, complex exponentials being what they are, this is the product of the two, the product of the N for exponential N raise to this power and raise to this power. So that's to say this is Omega to the nth order exponential to the power -2nm times the nth order exponential to the power of minus m. I did get my minus signs right over here, right? Yeah, okay.

All right, now, this you've already seen. This is the first term is Omega, the half-order exponential at the power nm, minus nm, times – this one I leave alone; nothing I can do with that. Okay, so once again, what happens to even order powers? An ordinary power of N over 2 is an even power of N, all right? Or you can say it this way. An even power of nth order exponential is an ordinary power of the N of the half-order exponential. Here, an odd power of the 4 exponential is a product of an ordinary power of the half-exponential, half-order exponential times this extra factor here. Okay. All right, now, let's go back to the formula for the DFT and plug this in. I'll put this into the formula for the DFT. All right, so what is that? That is the DFT, the nth component once again, is to just write it down again, is sum from N=0 to N-1 of the nth component of F times the power of the complex exponential, so I'm writing this in the form Omega N – nm. Okay?

All right, now, I'm gonna write this sum as a sum over even indices and odd indices. All right, so I'm gonna write this as sum over even indices plus the sum over odd indices. So what does that mean? Well, that means – I mean what does it mean in symbols? That means that the discreet Fourier transform is gonna be – I'm gonna take the sum from N=0 to N over 2 minus 1. We'll write it down and make sure we have it right. So this is N over 2 minus – let me write this a little more – give us a little more room here. The Fourier transform of F, the nth component is gonna be the sum over even indices. That's half the indices are even; half the indices are odd. So if I go to the half indices, there are N over 2 of them here, from 0 to N over 2 minus 1 or N over 2, those are the even indices and that is gonna be F of 2n, right, since I'm summing over the even indices, times Omega N, -2nm plus – to be continued on the next board – the sum over the odd indices, and again, half of the indices are odd, so it's gonna be the sum N = 0 up to N over 2 minus 1, the odd indices. So an odd number is the form 2N plus 1. Omega N, minus 2n plus 1m. Now everything's accounted for there, right? Everything's accounted for. These indices, 0, 2, 4, 6, 8, and so on, these are the indices. The components 1, 3, 5, 7, 9 and so on and so on; they're all there. I haven't lost anything. Okay, now, watch this. Put that up there and we put this up there. How about that for professionalism? And now – almost. Okay, now use the way that I have written the nth order complex exponential in terms of complex exponentials of order N over 2, the kind of stuff that I'm just about to erase.

So the sum over the even indices gives me the sum from $N = 0$ to N over 2 minus 1. F of 2n and this is the ordinary power N nth power of the half-order complex exponential, capital N over 2. Plus the sum of the odd indices over 2 minus 1 F of 2n plus 1. So what happened here, here was Omega n minus m times Omega N over 2 minus nm, if I had that correct. Do I have that correct? I believe I do, so that is what happens when you look at this complex exponential for an odd index, for an odd power. Now this doesn't depend on the sum over N, this doesn't depend on N, so that comes out of the sum. So let me just rewrite this quickly. So this is the sum from $N = 0$ to N over 2 minus 1 F of 2n times Omega N over 2(n-nm) plus Omega N minus m times the sum – very similar sum – and equals 0 to N over 2 minus 1 F(2n) plus 1 Omega of half-order exponential minus nm. Make sure I got all this right. Okay. Okay, nm, mn, doesn't matter. Again, everything is there, right? Everything is accounted for, mn, everything is accounted for. All the indices are there; all the components are there.

All right, now, this is looking like you've expressed an nth order DFT in terms of two DFTs of order N over 2 because the DFT of order capital N over 2 involves exactly the complex exponentials of power, to the power of the denominator or the power you're dividing by is capital N over 2. So this is looking like – this is almost, but not quite, for reasons you have to understand, this is almost the Fourier transform of – let me write it like this – of N, all right. The nth order of DFT looks like the N over 2nd Fourier transform of the even indices. If I take F, but I just divide it by the even indices, plus this power Omega and minus m, the DFT of order N over 2 applied to the odd indices, which is what I'm trying – I'm trying to get something like this. I'm trying to say you can evaluate an nth order DFT in terms of a combination of two DFTs of order half that. Not quite right. It's not quite right because – you have to think about this – the DFT of order n accepts an n-tuple and returns an n-tuple. The DFT of order N over 2 accepts an N over 2 tuple and returns an N over 2 tuple. So how are you getting an n-tuple out of two N over 2 tuples, and how would you like to say that a couple times fast? F (N over 2) accepts an N over 2 tuple and it returns an N over 2 tuple. So writing this statement can't be right, all right? You're missing something because if I just had the right hand side, the right hand side makes sense because F even is an N over 2 tuple, F odd is an N over 2 tuple, but what returns are two N over 2 tuples, not an n-tuple, which is the nth order Fourier transform of F. So we have to tickle this a little bit. We have to tickle the formula. Tickle the formula. Sounds so cute. Sounds so human. All right, now, I have to find – I have to get the formula for the Fourier – for the nth order Fourier transform of F for m going all the way from zero to N minus 1. I need to get all N components of it.

Now, for the first N over 2 minus 1 components, these are okay. That formula would apply. For m = 0 up to N over 2 minus 1, we can use the above formula, the formula above. That is to say – let me just rewrite it; make sure I haven't missed anything here – we can write, and all is well, the formula of the Fourier transform, the discreet Fourier transform of F, the N for this Fourier transform is exactly the N over 2nd, the Fourier transform of order N over 2, DFT order of N over 2 at m plus Omega(n-m), the N over 2 half-order Fourier transform of F odd, the nth component of that. All right, that's correct. That's correct for m going from zero to N over 2 minus 1. Okay? Now, that gets half of it. So what about the second half? So what about indices m indices from N over 2 up to N minus 1? All right. Write these in the form m plus N over 2, where m goes from zero up to N over 2 minus 1. That is, if m is ranging over the indices that index the half-order DFT, write the indices that index the second half of the nth order DFT in this form, all right? That is, the indices from N over 2, N over 2 plus 1, up to N minus 1 are of the form – if I started with zero here, I get N over 2, N over 2 plus 1. Then I would get N over 2 minus 1 plus N over 2 gives you m minus 1. It covers all of them. Okay? That covers all of them.

All right, now, what I mean by this, you have to write right now, what happens if I substitute this expression, an index in this form, into the expression that I have for the nth order DFT, which I am now stupidly erasing? Then the nth order Fourier transform at an index of the form m plus N over 2 is according to what I've already written down, which is valid for all of these indices, sum from N = 0 to N over 2 minus 1, F of 2n – this is what I had before – Omega N over 2, m plus N over 2, N times N plus over – minus – there were 2n, right? Plus Omega N minus m plus N over 2 times the sum from N = 0 to N over 2 minus 1 of F of 2n plus 1 the odd indices Omega – what was here? The same sort of thing, right? N over 2 minus m plus N over 2n. All right? Once again, let me make sure I have not screwed the pooch on this. Not, good, all right. Everybody with me? So that was the formula for the DFT writing it as the sum over even indices plus the sum over odd indices, but now I'm seeing what happens if you plug in an index in the form m plus N over 2, and again here, m is gonna range from zero to N over 2.

Well, now, let's just see what happens to these complex exponentials. That's what we have to answer. All right, but it's not hard. A high school student could do this if you told him what to do at every stage. Omega – a lot of research is like that. It's a big secret. N over 2 minus m plus N over 2n, so that's Omega N over 2 minus mn plus N times N over 2 – minus N times N over 2. Now once again, these exponentials being what they are, this multiplies. This is Omega of N over 2, minus mn times Omega N over 2, minus N times N over 2. But this is a complex exponential. This is an N over 2 root of unity raised to the N over 2nd power, or an integer times N over 2nd power. So this is equal to 1. This is like e = 2pi over N to nth power. So this is 1. So what remains there is just Omega to the N over 2 minus nm. Omega to the N over 2 minus nm – mn, same thing. And that comes up in both these sums, right? Yeah, it's the same thing in both these sums, so in the sum over the odd indices and the sum over the even indices, the only other thing you have to figure out is this up front. That bracket out front. This comes up in both sums, even and odd indices. This expression. The only other thing that comes up is the factor out front, which is Omega – what is it? Omega N minus m plus N over 2. So that's Omega N minus m minus N over 2, and again, complex exponentials being what they are, this is Omega N minus m times Omega N, N over 2. Minus, it doesn't matter, minus.

Now, what is that second expression? This expression, fine, that's what we had before actually. That's what we had before as a factor out front. What about this expression? Well, this is e = 2pi over N raised to the minus N over 2 power. So what happens is the N cancels out the N, the 2 cancels out the ½ there and it's e = pi, which is also known as minus 1. So the factor out front changes to minus what is was before. So Omega, in other words, Omega N, minus m plus N over 2 is just minus Omega N minus m. Correct? Correct. All right, so what is the actual retail value of the expression? So we have Fourier to nth order of the DFT of F and an index to the form m plus N over 2, which is what I'm worried about here, for little m going from zero to 1, 2, 3, up to capital N over 2 minus 1 is equal to the sum over one half of the first sum – the first sum didn't really change very much – over N over 2 minus 1, F(2n) – that's the sum of the even indices – Omega N over 2, minus nm. Was that right? Was it right, the first sum didn't change, the first factor didn't change? Help me here. Thank you. Yes, I believe that's the case that the first sum didn't change. The second sum – the only thing that happened in the second sum was this factor out front and the factor out front changed to minus itself. So minus Omega nm times the sum from N = 0 to N over 2 minus 1, F of 2n plus 1 Omega N over 2 minus nm.

That's the result of plugging in an index of the form little m plus N over 2 into the original form we had to decompose the nth order DFT into the sum over even indices plus sum over odd indices. Okay. But look! Once again, this is okay – m is ranging from zero to N over 2 minus 1 here. This is again a DFT of order N over 2 returning an N over 2 tuple and this is a DFT of order N over 2 returning an N over 2 tuple. So this says – it's exactly what we want – this says that Fourier transform, the discreet Fourier transform of order N at m plus N over 2 is equal to the discreet Fourier transform of order N over 2 of F of the even indices applied to m minus Omega N, minus m, the Fourier transform of N over 2 of F applied just keeping the odd indices at m, and this is again, m from zero up to N over 2 minus 1. All the components of the nth order DFT of F have now been accounted for and they are computed by finding half-order DFTs. The first half of them – let me summarize. The first half of them are computed by the first formula that we had up there. The second half of them are computed by this formula and the only difference between the two is this is a minus sign here instead of a plus sign. So in summary, I'm gonna erase this, but I'm gonna write it again. Glutton for punishment that I am. So to summarize, for m equal zero, 1, up to N over 2 minus 1. The Fourier transform, the nth order Fourier transform of F at m, the first half of the coefficients, is the Fourier transform of order N over 2 applied to the even indexed elements of the original input plus Omega nm, the Fourier transform order N over 2 of the odd indexed version of the original input, and the second half – so that's the first half of the indices. For the second half of the indices, the nth order Fourier transform of F at m plus N over 2. Again, m is just ranging from zero to N over 2 minus 1. That's the index range for the half-order DFT. This is equal to the half-order Fourier transform F applied to the even indices, evaluated m, minus Omega minus m of the half-order Fourier transform of F the odd indices evaluated at m. And that's all of them, okay? That's all of them. That writes the

DFT as a combination of two half-order DFTs or maybe I should say two DFTs of half the order.

And now, iterate. If capital N is a power of 2, then you cut each one of these things down, okay? You cut each one of those things down and so on and so on and so on till you're down to a single 2. Now, here's what I'm not gonna do because we're almost out of time, although that's never stopped me before, but it's what I'm not gonna do. So I'm not gonna show you – so this is the FFT algorithm, or rather maybe I should say this is the basis for the FFT algorithm. All right, again, the idea is that half the work is twice the fun somehow. I mean you're cutting down the nth order DFT to 2 N over 2 DFTs and then iterate. Now, so what I'm not gonna show you is why, theoretically and practically, this gives you the savings going from N2 to N log N. That's discussed in the notes and if you have any experience with what they call computational complexity, you'll see how the argument goes. If you don't, we can work through it some time, but I'm not going to go through that part. I don't want to somehow detract from this beautiful arrangement of the algorithm we have here. The second thing that I'm not gonna – so I'm not gonna do that, but this savings, cutting it down by 2 each stage, reduces the number of steps from N2 to N log N. I'm also not going to show you the sort of final form of the algorithm. If you look – and I actually don't have this in the notes, but if you look in other books, it's quite common to display, to sort of carry the step out, and there's this gorgeous – or not, depending on your point of view – diagram called the "Butterfly Diagram" about how the individual inputs, original inputs, of the discreet signal sort of snake their way through the algorithm if you keep iterating it like this.

You know, what happens to the zero of output, what happens to the first output, what happens to the even inputs – it shouldn't be even – but the zero input, the first input, and so on, what happens to the even input, what happens to the odd input and so on. So I'm also not gonna show you that and that I actually don't do in the book, in the notes, but you can find that elsewhere if you're interested in it. And finally, I'm also not gonna show you – I'm sorry for all these things I'm not gonna show you – I'm also not gonna show you how this leads to the matrix factorization, all right? But that's exactly what happens. I mean you can translate from one to the other, that is, the fast Fourier transform algorithm leads to a factorization of the DFT matrix into a product of matrices where there are lots of zeroes. And the fact that you have lots of zeroes is another reason why there's such a savings in computation because when you have zeroes in your matrix, that's effectively not an operation, not a multiplication that happens. So those are the things I'm not gonna show you. But what I did want you to see is I wanted you to see how this arrangement works. The crucial first step in going from an nth order DFT to an order to 2 DFTs to order N over 2, that's really the crucial observation here and how this was done. It really is quite striking and of course, how much of the modern world's economy depends on this algorithm. That's also really – I think it's right up there at the top of the list for the number for the algorithms that are in constant day-to-day use. So we say goodbye right now to the DFT and its properties. Next time, I'm gonna talk about linear systems. We have to do a little rearranging here because I wanted to be sure to cover this, by popular demand, and then we'll move on to a discussion of linear systems, linear time variance systems, and then higher dimensional Fourier transforms. Thank you all.

End of Audio

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture23

Instructor (Brad Osgood):Are we on? I can't see. It looks kinda dark. I don't know. It looks a little dim there.

All right. So today – assuming this is working – or assuming even it's not working – we are going to spend a little bit of time over the next couple days talking about linear systems, particularly linear time invariance systems because those are the ones that are most naturally associated with the Fourier Transform and can be understood and analyzed – some aspects of them – in terms of the Fourier Transform.

But before doing that, we wanna talk about the general set up – the idea of linear systems in general – and talk about some of their general properties, as fascinating as they are.

Now it's a pretty limited treatment that we're gonna do of this. So I would say this is more an appreciation rather than anything like a detailed study. It's a vast field, and in many ways, I think it was one of the defining fields of the 20th century. The 20th century, in many ways, was a century of – I think I even said this in the – made this bold statement in the notes. The 20th century was a century of linearity in a lot of ways.

The 21st century – I say this as a sweeping bold statement, but I stand by it. The 21st century may be the century of non-linearity. We don't know yet, but non-linear problems are becoming increasingly more trackable because of computational techniques. One of the reasons why linear problems were studied so extensively and were so useful is because a lot can be done sort of theoretically even if you couldn't compute. And then, of course, later on when computational techniques – computational power was there, then they became even more – they were able to be exploited even more. What I wanna get to is the connection between the Fourier Transform and linear systems, and that's gonna be primarily along the lines – so we definitely wanna see how the Fourier Transform applies to linear systems, again in a fairly limited way.

And here, the main ideas are talking about the impulse response and the transfer function. These are the sort of major topics that I wanna be sure that we hit. The impulse response and the transfer function – these are terms, actually, we've used already, but now we're gonna see them a little bit more systematically and a little bit more generally. And, again, they're probably terms and probably ideas that you've run across before if you've had some of this material earlier in singles and systems. And the other thing is – again, somewhat limited and maybe even to a lesser extent – is to talk a little bit about complex exponentials appearing as IGen functions of certain linear systems – time invariance systems. So we'll put that up here.

Complex exponentials as IGen functions. I'll explain the term later if you haven't heard it, although I suspect many of you have. IGen functions of linear time invariance systems. All right. So this is, I guess, sort of a preview of the main things that we wanna be – we wanna discuss. But before getting that – before doing that, I do have to do a certain amount of background work and frame things in somewhat general terms. So let's get the basic definitions in the picture. First a basic definition of a linear system. So a linear system for us is a method of associating an output to an input that satisfies the principle of super position. All right? So it's a very general concept. It's a mapping from inputs to outputs. In other words, it's a function. But this is usually the engineering terminology that's associate with it. Outputs that satisfies the principles of super position. And you know what that is, but I will write it down. Super – not supervision. Super position. I get it. I'll get it. Super position.

So you have – you think of the linear system L as a black box. It takes an input V to an output W, and to say this has the principle super position says that if you add signals, then the – add the inputs, and the outputs also add. If you scale the inputs, then the outputs also scale. So it says that L of V1 plus V2 – whatever their nature – is L of V1 plus L of V2, and it says that L of alpha times V is alpha times L of V. By the way, it's sort of a common convention here, when you're dealing with linear systems, not to write the parenthesis because it's supposed to be reminiscent of matrix multiplication where you don't always write the parenthesis when you're multiplying by a matrix. As a matter of fact, I'll have more to say about that in just a little bit. All right?

That's the definition of linearity. To say that it's a system is just to say that it's a mapping from inputs to outputs. Again, that doesn't really say very much. Everything we study is sort of a mapping from inputs to outputs, but this extra condition of linearity is what makes it interesting. And it took a long time before this simple principle was isolated for special attention, but it turned out to be extremely valuable. I mean, nature provides you with many varied phenomena, and to make some progress, you have to somehow isolate what's common to the various phenomena. And again in mathematics, the way it works is – in the applications of mathematics, you wanna turn that around and turn that into a – turn around the solution of what you observe and turn that into a definition. So the definition that came from studying many different phenomena in many different contexts was this simple notion of linearity or super position – same thing. All right? So it really is quite striking how fundamental and important these simple conditions turned out to be in so many different contexts. And that's really, I say, almost defines a lot of the applications of mathematics – practical applications of mathematics in the 20th century – just isolating systems that satisfy this sort of property.

Now, there are additional properties that it might satisfy, and we'll talk about some of them. But the basic property of super position is the one that really started the whole ball rolling. Okay? I should say, as an extension of this, if you have finite sums, then I can take L applied to, say, I equals one to N of alpha I times VI, then that's the sum of – so that's a linear combination of the inputs, and what linearity says is linear combination of the inputs go with the linear combination of the outputs. This is a sum from I equals one to N of alpha I times L of VI. Now it's also true, in most cases, that this extends to infinite sums. But any time you deal with infinite sums, you have to deal with questions of conversions and extra properties, the operators – we are not gonna make a big deal out of this. I won't tell you anything that's not true, I hope. But again, I'm not gonna always take the assumptions carefully. You can extend these ideas to infinite sums and even to intervals, which we will talk about a little bit. But generally, that requires additional assumptions on the operator, L. Which again, I'm not gonna make – and usually, assumptions are fairly mild, all right? That are gonna be satisfied in any real applications. The basic assumption that you often make, and again, without really talking about it in detail, is you assume continuity – some sort of continuity properties. Any time limiting operations are involved – we've seen this in a number of instances – there has to be some extra assumption on the operations you're working with. And it's generally some sort of continuity assumption that allows you to take limits. So you assume some kind of continuity.

But the problem is defining what it is – defining what continuity means and so on and so – I'm not gonna get into it. And again, it's not gonna be an issue for us, but I thought I ought to mention it by – to be honest. Now what's a basic example of anytime you learn a new concept, you should have – or even revisit a familiar concept – you should have examples in mind. What is an example of a linear system? There is actually only one example of a linear system. They're all the same. It is the relationship of direct proportionality. The outputs are directly proportional to the inputs. L of V is equal to alpha times V. All right? That is certainly linear. It certainly satisfies the properties of super position. L of V1V2 is alpha times V1 plus V2. So that's alpha times V1 plus alpha times V2. So that's L of V1 plus L of V2. And likewise, if I scale, I actually call that alpha there. But I'm thinking of alpha as just a constant here. L of, say, A times V is equal to A times L of V for the same reason. All right? The relationship of direct proportionality is the prototype – the archetype – for a linear system. In fact, it's the only example. All right? All linear systems essentially can be understood in terms of direct proportionality. That's one of the things that I wanna convince you of. That's one of the things that I wanna try to explain. It's the only example.

And that's sort of a bold statement, but I stand by it. Maybe a little shakily, but I stand by it. Say all linear systems say [inaudible] back somehow to the operation of direct proportionality. All right? So don't lose sight of that. So for example – now, it can be very – it can look very general, all right? It can look very general. Direct proportionality is also known as multiplication. So any linear system that is given by multiplication – any system that is given by multiplication is a linear system. All right? So a little bit more generally is multiplication. That is to say you can think of multiplying by a constant, but if your signal is not a constant, but a function of T or a function of X, then I can multiply it by another function. So L of V of T, I can multiply by alpha T times V of T. Okay? The constant proportionality doesn't have to be constant. It can also depend on T. But nevertheless, the relationship is one of direct proportionality. And for the same simple reason as up here, that defines a linear system – linear.

So when there are many such examples of that – practical examples of that – a switch! A switch can be models of a linear system. If it's on for a certain duration of time, then that's multiplication by, say, a rectangle function of a certain duration. So EG – a switch – L of V is – the L of V of T is, say, a rectangle function of duration T times VT, a duration A. All right? So you switch on for duration A. Then you switch off. On for duration A. Now you don't necessarily think of flipping the switch as a linear operation, but it is. Why? Because it's multiplication. You could – somebody could say to you, "Verify that the act of switching on a light bulb is a linear operation." But the fact is that it's modeled mathematically by multiplication by a function, which is one for a certain period of time and zero for the rest of the time. And as multiplication, it is the operation – it is the principle – it is just expressing direct proportionality, and that's always linear.

Sampling is a linear operation. Sampling at a certain rate L of V of T would be – could be a Shaw function of spacing T times V of T because [inaudible] spacing P times V of T. All right? It's multiplication. It's direct proportion. It's linear. So again, somebody could say to you, "Say, is it true that the sample of the sum of two functions is the sum of the sample functions?" And you might be puzzled by that question, or you might – that might take you a while to sort it out. You might try to show something. I don't know what you might try to show. You might try to show that sort of directly, but in fact, yes it must be true that the sum of two sample functions – the sample of the sum of two functions is the sum of the sample functions. All right? But that's true because the relationship – because sampling – the act of sampling is a linear operation – is a linear system. Okay? It's multiplication. It's direct proportion.

Now, a slight generalization of direct proportion is direct proportion plus adding up the results. That is, to say, matrix multiplication. I should say slight, but important, generalization. Generalization is – well, let's say direct proportion plus adding two linear operations plus adding. And what I have in mind here is matrix multiplication. So i.e. matrix multiplication. All right? If I have an N-by-N matrix, say A is – and let me see if I can even do this more generally – say an N by M matrix, all right? So it's N rows by M columns and V is an M vector, so it's a column vector with M rows. Then A times V is an N vector, and the operation of multiplying the matrix, say, by the column vector V is a linear operation. It is a combination exactly of direct proportion or multiplication with adding. So what is it? If you write A as the matrix, say, AIJ so it's mixed by columns and rows, then A times V the [inaudible] entry is a sum over J – J equals one to N – J equals one to N – AIJVJ, isn't that right? Probably not. Let's see. Do I have an M? I go across the number of pies, and N by M – I hate this stuff. Man, I can never get this right.

N by M matrix, so – yeah. Right, M columns. Right, okay. That's fine. That gives you all the entries. If it does, fine. If it doesn't, then switch M and N. Okay. Each component is multiplied by J – that's direct proportionality – and then they're all added up. And as you know, the basic property of matrix multiplication is that A applied to the sum of two vectors is A of V plus A of W. And A of a scale times V is alpha times A of V. Okay? It's a slight generalization, but it turns out to be, actually, a crucial generalization. And it comes up in all sorts of different applications.

Those of you who are taking 266 [inaudible] have done nothing but study matrix multiplication. Well, that may be a little bit of an extreme statement, right? So EGE 263 where you study the linear dynamical system X dot is equal to AX, and you solve that. Say X of zero is equal to V in initial condition. All right? Then solve by X of T equals E to the T times A times X of zero, which is V. All right? It's a matrix times the fixed vector V gives you how the system evolves in time. All right? And you wanna be able to compute that, and you wanna be able to study that. And you spend your life doing that.

Many people do. Now, again, without going into detail, now – and we'll say a little bit more about this later – the property of linearity is extremely general. There are special cases that are important, some of which I'm sure you've seen. So let me just mention special linear systems – let's just stick with the case of matrix multiplication right now. All right?

So special – linear systems with special properties derive from the special property of the matrix of A. So for example, some of the main examples are – some of the most important examples are, for example, if A is symmetric, then you sometimes call it a self-adjoint system or a symmetric system. So A – to say that A is symmetric is to say that it's equal to its transpose. So for example, if A is symmetric, that's a special type of linear system. As a matter of fact, I'll tell you why that's important in just a second. [Inaudible] transpose is equal to A. A can be – or hermission is the complex version of this, where the condition is A star is equal to A. So this is the complex case. All right? That is A star is equal to the conjugate transpose. These are both very important special cases. They come up often enough, so again, it was important to single them out for special study.

Or another possibility – those are, maybe, the two main ones. Another possibility is A can be unitary or foginal. So if A unitary means that A times its conjugate transpose – its adjoint – is equal to the identity or A star times A is equal to the identity. I'm talking about – I should say here, I'm talking about square matrices – a N-by-N matrix. So it's square. Okay? Now, a very important problem and a very important way of – and we'll – again, we're gonna talk about this when we talk more about general linear systems – a very important approach to understanding the properties of linear systems is to understand the aspect of their IGen values and IGen vectors associated with them. I'm saying these things to you fairly quickly because I'm going under the assumption that this is largely – by and large review. All right? That you've seen these things in other classes and other contexts.

So you often look for IGen vectors and IGen values of matrix A. All right? And we are going to, likewise, talk about IGen vectors and IGen values for general linear systems, and that's where the Fourier Transform comes in. But just to remind you what happens here in this case, just to give you sort of the basic definition – so you say V is an IGen vector if A times V is equal to lambda times V for some V. So V is a non-zero IGen vector – non-zero. If there's some non-zero vector that's transformed into itself. So there you see – you really see the relationship with direct proportionality, all right? For an IGen vector, the relationship is exactly direct proportionality. A times V is just a scale version of V. The output is directly proportional to the input. All right?

Now it may be that any – that you have a whole family of IGen vectors that span the set of all possible inputs – that form a basis for the set of all possible inputs. If you have IGen vectors, say, V1 through VN with corresponding IGen values lambda one through lambda N that form a basis for all the inputs for all the input, all the system, all the signals that you're gonna input into the system, then you can analyze A – the action of A easily. All right? That's because, if it forms a basis for all the inputs – if V is any input – so let's say yea. It's V any input – then you can write V is some combination alpha I times VI. I equals 1N. That's what it's saying – that that's what it means to say that they form a basis for it. And then A operating on V by linearity – I can pull that to – I can pull that A inside the sum and have it operate on the individual scaled IGen vectors. So A of V is A of the sum is the sum of I equals one to N of A of alpha I times VI. But again, the scaler alpha I comes out by linearity. That's the sum from I equals one to N of alpha I times A times VI. But A just takes V to itself or a scaled version of itself – VI to a scaled version of itself. So this is sum I equals one to N of alpha I times lambda I times VI. The action of A on an arbitrary input is really here. You see you're getting direct proportionality plus adding. It's very simple to understand. Each component is stretched, and then the whole thing is scaled by whatever initially scaled the inputs. All right? If the inputs are scaled by alpha I, the outputs are also scaled by alpha I. In addition, they're scaled by how much the individual IGen vectors are stretched. Okay?

It's a very satisfactory picture and an extremely useful picture. So the question is, for example, when do linear systems have a basis of IGen vectors? When can you do this? And that's when these special properties come in. All right? That's when these special properties come in. So for example, and I won't – I'm not gonna – this is sort of, again, fundamental linear algebra that I assume you've probably seen in some context. But because this is so important, you gotta ask yourself when can you actually do this. And the spectral theorem in finite dimensions for matrices says that if A is a hermission operator or symmetric operator in the real case, then it has a basis of IGen vectors, and you define a basis of IGen vectors. The spectral theorem says when you can do this. If A is symmetric or, in the context case, hermission, then you can find a basis – actually, an orthonormal basis – even better – basis of IGen vectors. All right?

Now if you're thinking that this doesn't look – that this looks sort of vaguely familiar or this is somehow – I'm using the similar sorts of words to when we talk about fourierciaries, and I talked about complex exponentials forming an orthonormal basis and so on. It's very similar. All right? It's very similar. And the whole idea of the Fourier Transform and diagonalizing the Fourier Transform – applying IGen vectors, IGen functions – in that case, you call them for the Fourier Transform or how they come up in Fourier series is exactly sort of what's going on here. Okay? These are simple ideas, right? We – all that we started with this idea of super position – that the sum of the inputs goes to the sum of the outputs and a scale version of the input goes to a scale version of the output. And the structure that that entails is really quite breathtaking. It's really quite astounding. All right? Now, there's one other important fact about the finite dimensional case – the case of just finding N-by-N square matrices – that's very important. And also, we're going to – all these things have some analogue in the continuous case and sort of the infinite dimensional continuous case, which is where we're gonna spend most of our time. All right? But this you should sort of know. This should be your touchstone for understanding the more – what happens more generally or sort of what happens in the case of N by N matrices and what happens in the finite dimensional case – what happens from what you learned in linear algebra. So one more property – it's not that matrix multiplication is just a good example of linear systems. All right? This is like – it's not just direct proportionality as an example of linear systems. Direct proportionality is the only example of linear systems. All right?

Well slightly more generally, it's not that – or it's not just that matrix multiplication is a good example – a natural example of – let's call it finite dimensional linear systems. All right? So it's like an N-by-N matrix operating on an N vector – whatever. It's the only example. All right? Now you learned this in linear algebra, although you may not have learned it quite that way. What that means is that any linear operator – I'll say it very mathematically that'll just give you an example – any linear operator on a finite dimensional space can be realized as matrix multiplication. And I'm gonna give you a problem to think about. Any linear system – let me put it this way. Any finite dimensional – so finite number of degrees of freedom – a finite number of ways of describing any input – described by a finite set of vectors – a finite set of signals, inputs. Any finite dimensional linear system can be realized as matrix multiplication. All right? It's not just that it's a good example. It's the only example.

Now let me just say – let me just take a little poll here. Raise your hand if you saw this in linear algebra – saw this theorem in linear algebra. Not so widespread. All right. Well, you did. If you took a linear algebra class, you probably saw this result. All right? Maybe not phrased quite this way, but this is sort of one of the fundamental results of linear algebra. Now mathematicians are quick to say, "Yes, but we don't like matrices. We would rather stay with the linear operators, per say. Beautiful and pristine as they are, to introduce matrices is an obscene act."

Went out like that. All right? We find it useful to manipulate matrices. We find it useful, often, to have this sort of representation. I'll give you one example you can try out for yourself. So for example, some of you may have – an example you may have done – example: let me look at all polynomials of degree less than or equal to N. All right? That's the space of inputs. Inputs are polynomials of degree less than or equal to N. So N is fixed. All right? So any input looks like A0 plus A1 times X as a constant term, a coefficient of X, a coefficient of X squared up to a coefficient of X to the N. Not exact – I'll allow myself to have some zero coefficients in here. So I don't go up – I don't necessarily have to go all the way up to N, but I go up, at most, to N – to X of the N. All right?

So any input looks like that. Now what is a familiar linear operator? A familiar linear operator on polynomials that takes polynomials to polynomials is the derivative. If I differentiate a polynomial, I get a polynomial of lower degree. So take L to BDDX. All right? That's the linear operator. That's a linear system. All right? As such, in the space of polynomials is a finite dimensional space. It can be finite degrees – a finite number of degrees of freedom. The degrees of freedom are exactly described by the N plus one coefficients. They're N plus one because they have a constant term – order one up to order – up to degree N. All right? So L can be described as an N by one by N by one matrix as – by an N plus one by N plus one matrix. Find it.

Any linear operator in a finite dimensional space can be described as matrix multiplication – can be written in terms of matrix multiplication. All right? There's a linear operator on a finite dimensional space. Doesn't look like a matrix, but it can be described as a matrix. Find the matrix. Yeah. Thank you. And no – well, it can – yes, actually – yeah. So I'll leave it to you to think that. That's right. It actually drops the degrees by one. So you can describe it either – if you do N plus one by N plus one – let me give – I'll give you a hint. You're gonna have either a row or column of zeros in there. All right? But in general, if I'm thinking of it just sort of as a map from N plus one degree polynomials and N plus one degree polynomials, it'd be a square matrix. All right? So I'll let you think – I'll let you start this out. This is a problem – actually, so let me take a poll again – a brief poll again. Anybody do this problem in linear algebra class?

Yeah. Okay. You probably hated it then. You may hate it now. But – and again, it's a sort of scattered minority response out there. All right? But it just shows, again, that this idea that – it's not just that it's a good idea. It's the only idea. Representing linear operators matrix – interpreting linear operator – linear system on finite dimensional space as a matrix multiplication – it's not just a clever thing. It's not just a nice example. It's the only example. All right? And in fact, we're gonna see that that same statement, more or less and for our purposes, holds an infinite dimensional continuous case. That's what I wanna get to. I won't quite – I don't think I'll quite get there today, but I'll get a good part of the way. All right?

I wanna see that a similar – very similar statement – but there is an analogous statement – the infinite dimensional continuous case – very satisfactory state of affairs. There is an analogous statement for the infinite dimensional continuous case. All right? So let's understand that now. So I'll understand that first, in terms of an example rather than a general statement, the example that generalizes matrix multiplication is integration against a kernel – or what I should say is the operation that generalizes matrix multiplication is integration against the kernel. Something we have seen. Something I will write down now for you. So the operation – the linear system that generalizes matrix multiplication is the operation of integration against the kernel. That would – that's the phrase that you would use to describe it.

So what it is? What do I have in mind here? Well, again, the inputs this time are gonna be functions. We'll do it over here. All right. So the inputs are, instead of just a column vector, are going to be functions – is a function, say, V of X. All right?

And the kernel – a fixed kernel for the operator – the things defines the operation is a function of two variables. So the kernel is a function – let's call it K – K for kernel – K of XY. All right? Integration against the kernel is – the operation is L of V. So I can – it's gonna be producing a new function. I'll say that's also a function of variable X. There's also a little bit of a problem here like there is in this whole subject with writing variables, but let me write it. It's gonna go from minus infinity to infinity K of XY, V of Y, VY. All right? K's a function of two variables. I integrate K of XY against V of Y, VY. What remains is a function of X. That by definition is the output devaluated at X. All right?

So L of V is another function. What is its value at X? I integrate K of XY against V of Y, VY. What remains in this integration is the function of X – depends on X. Okay? That's what I mean by integration against a kernel. The kernel K defines the operation – defines the linear system. So it is linear because integration is linear. The integral of the sum of two functions is the sum of the integrals. The integral of a scaler times the function is the scaler times the integral of the function and so on. So that's the first thing. I won't write that down, but I will say it. So L is a linear system. So L is linear. All right? Now first of all, if you sort of open your mind a little bit, you can really think of this as a sort of infinite dimensional continuous analogue of matrix multiplication. It's the infinite dimensional continuous analogue of matrix multiplication.

Why? What do I have in mind by a statement like that? Well, what I have in mind is it's like you think of V as, somehow, an infinite continuous column vector. All right? So it's like you think of V as – I mean, you can even make this more precise if you actually use [inaudible] sums, but I don't wanna do that. I don't wanna write the – well, let me just write it out like this – as an infinite column vector. All right?

And think of this operation integral from minus infinity to infinity K of XY, V of Y, VY – what's going on here? So this is like a column vector. K of XY is like a matrix – a doubly infinite continuous matrix. X is the index of the row. V is the index of the column. You are, like, summing across the columns of the matrix time – summing across a row of a matrix – that's integrating with respect to Y – times the corresponding column entry V of Y. So this is like a column index. This is like a row index. And an interval, of course, is like a sum. Okay? This is exactly what's going on. Exactly what's going on.

K of XY – you're summing across the X row, right? XYYYYYYYY times VY VYYYYYYYY and you're adding them all up according to the integral, and you're getting the component – the X component of V. All right? Now see, analogue – now what else is true? Or what else is true? If it's such a good analogue, are there analogues to the other statements that went along with the finite dimensional case? Well again, just as in the finite dimensional case, there are special linear systems that are characterized by special properties of the matrix. So too, in the sense of [inaudible] continuous case, there are special properties of the systems that are characterized by special properties of the kernel. All right? And although I'm not gonna use them now, I at least wanna mention them because I wanna continue this sort of analogy between the finite dimensional discreet case and the infinite dimensional continuous case.

So special linear systems arise by extra assumptions on the kernel – on K of XY. All right? So for example, you might assume – now what do you think is the analogue to the symmetric case? For a matrix, it's that the transpose of the matrix is equal to the matrix. So what do you suppose the transpose of – or the analogue of the transpose is for a kernel K of XY? What should the condition be? What should the symmetry condition be?

Yes. Be bold. I'll help you. I won't help you. All right. What should a symmetry condition be that's sort of analogous to a matrix being equal to its transpose? If K of XY is the analogue of the matrix where X is the row and Y is the column, how do you get the matrix? You interchange the column and the row.

Student:[Inaudible].

Instructor (Brad Osgood):Pardon me?

Student: Time invariance?

Instructor (Brad Osgood):No, not time invariance. We'll get to that.

Student:[Inaudible].

Instructor (Brad Osgood):Right. I think I heard it there. All right.

Symmetry – or self-adjointess – is the property K of XY is equal to K of YX. If the kernel satisfies this property, you say it's a symmetric system – symmetric or sometimes you call it a self-adjoint linear system. They have special properties. I'm not gonna talk about the properties now, but again, I'm just pursuing the analogy between the discreet case and the continuous case. All right? Or – and what's hermission symmetry? Hermission symmetry, in the case of a complex kernel – and I won't allow the case of a complex kernel – would be K of XY would be K of YX bar. Okay? Complex conjugate. This is all [inaudible] and so on and so on, and I won't gulf into that very much now. Now we have seen many examples of linear systems that are given by the integration against a kernel. What is an example of – what is a fundamental example in this class of a linear system that is given by integration against a kernel? A Fourier Transform. Good. So for example, a Fourier Transform – FF of S – is the integral from minus infinity to infinity E to the minus two pi IST, F of T, DT is exactly integration against a kernel. What is the kernel? The kernel is K of ST is a minus two pi IST. All right? It fits into that category. It has special properties – many special properties. That's why we have a course on it. Okay? But nonetheless, it fits under the general category of a linear system. And actually, you can check that K of ST is equal to K of TS is actually symmetric. All right? If I switch S and T, the kernel doesn't change. So it's a symmetric linear system and so on. What is another example of an important linear system that can be described by a – by integration against a kernel? What's another example that we have studied extensively and use everyday almost – on good days?

Student:[Inaudible].

Instructor (Brad Osgood):Convolution. All right? All right? Fix a function H, all right? Then if I define L of V to be H convolved with V, that is a linear system. That's a linear system. Convolution is linear, but what is that in terms of the operator? L of V of X is the integral from minus infinity to infinity H of X minus Y, V of Y, DY. All right? Convolution is a linear system that falls under the general category integration against the kernel. It's special one, actually, and it's what – as it turns out, it's a very important special case because the kernel here doesn't depend on X and Y separately. It depends only on their difference. All right? So note for convolution – all right. For convolution – that is, for a linear system given by convolution, the kernel depends on X minus Y. [Inaudible] is a function only of one variable or the difference between the two variables X minus Y instead of X and Y separately – and not X and Y separately. All right? Now for reasons which you've probably seen, actually, and which we'll talk a little bit more about detail – this particular special case leads this – and this property leads to a so-called shift invariance or time invariance. All right?

So in particular, if we shift X and Y by the same amount – A, say – some number A – A. So X goes go X plus A or X minus A if I delay it by A. Y goes to Y minus A. And then, of course, X minus Y is equal to X minus – goes to X minus A minus Y minus A is X minus Y. It's – the difference is unchanged. All right? So the convolution is unchanged. If I shift X and Y. All right? And this leads to – I'm not – I don't wanna say too much more about it than that, but this is what leads to the so-called shift invariance or time invariance of convolution. This leads to convolution. That is this observation leads to the phrase you hear – and we'll talk about this – convolution as a linear shift invariant or time invariant. People usually say time invariant, but it's really better to say shift invariant somehow. It's more descriptive – linear time invariant system. All right? But we'll get back to that. The fact is that, again, convolution is, of the form, integration against the kernel, but it's a special kernel because it depends only on the difference of the variables, not on their – not on the variables separately. Okay?

In general, integration against the kernel is integration against the function of two variables. Now it's not just that this is a good idea. It's not just that this is a good example of linear systems. I'm not talking about convolution here. I'm talking about generally integrating against a kernel. All right? So the words that I said, like, ten minutes ago, I'm gonna say again. But in this different context. So it's not just that integration against the kernel is a good example of linear systems – in this case, continuous linear systems – infinite dimensional linear systems – just like it's not just that matrix multiplication is a good example of finite dimensional linear systems. It's the only example. Okay? It's the only example. Any linear system – now this is statement has to be qualified because there are assumptions you have to make and so on, but that's not the point. The point is that any linear system can be realized somehow as integration against the kernel. Yeah.

Student:[Inaudible] manifest in a matrix operator?

Instructor (Brad Osgood):Oh, that's a good question, and we'll come back to that, actually. It's in the notes. The matrix has had special properties

Student:[Inaudible].

Instructor (Brad Osgood):Circulant, actually. It's a little bit more than Tarpowitz. Yeah. RTFN, man. It's in the notes. Okay? We'll come back to that.

All right. For now, don't spoil my drama. All right?

Again, it's not just that matrix – in the finite dimensional case, it's not just that matrix multiplication is a good example. It's the only example. In the infinite dimensional – the continuous case, it's not just that integration to kernel – against the kernel is a good example. It's the only example. All right?

Any linear system can be realized as integration against the kernel. All right?

Now on that fantastically provocative statement, I think we will finish for today. And I will show you why this works next time.

[End of Audio]

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture24

Instructor (Brad Osgood):That was quick. Looks like the crowd's a little thin out there. This looks more like the Wednesday before Thanksgiving than – instead of the Monday before the Wednesday before Thanksgiving. Okay.

That was the big fear that – when Stanford switched to giving the entire week off of Thanksgiving, they just sort of – people would take off earlier and earlier. And sure enough – all right.

Anyway, I want to continue with our discussion of linear systems – big day today, actually. It's really – the state of affairs, as you'll see when we wrap up the discussion, is really quite satisfactory and understanding linear systems and the structure of linear systems. And then we'll make further use of it next time when we talk about – more essentially about time invariant systems as a special case.

So let me remind you sort of where we finished up last time. This is linear systems part deux. And I was sort of making the comparison between the continuous case and the discreet case. We're gonna spend most of our time today talking about the continuous case, but one of the places where I made the connection or the analogy was in the discreet case, any linear system has to be given by multiplication by a matrix. All right?

So in the finite dimensional, I said – when I talk about discreet case, I'm sorta thinking about N-by-N matrices and discreet signals and so on. So in the discreet finite dimensional case, as opposed to the continuous infinite dimensional case where the inputs and outputs are functions – so here, they're just finite vectors. Any linear system is given by multiplication of the matrix by a matrix. All right? That's a fundamental fact that you will learn in linear algebra.

The matrices are different. You get a matrix by choosing a basis of the space of inputs, and then you express the matrix in terms of what happens to the bases. And I'm not gonna go through this because I'm assuming that you've seen this in linear algebra, although you may not have thought about it quite in these terms.

And what I wanna see is that an analogous result holds in the continuous case, and it's really quite striking.

So an analogous result holds in the continuous infinite dimensional case when the inputs and outputs are functions of a continuous variable instead of discreet functions finite – just a finite list of numbers – a vector – infinite dimensional case.

All right, now this gets into – to get a completely precise statement of this. And to see to what extent this really holds gets into pretty deep waters, all right? And that's not my intention to go there, but I do want you to sort of see what the point is.

All right. And what it amounts to is understanding a little bit better about the special case of a linear system is given by integration against the kernel. So we'll see a linear system, in this case – in the infinite dimensional continuous case – is given by integration against the kernel – against a kernel.

This was sort of a example of – one example of a linear system in the infinite dimensional continuous case K of XY, V of Y, DY. That's L operating on V is of that form, and that we talked about last time as an analogy to matrix multiplication where K is playing the role of the matrix. And what I wanna see is that this is not just a good example of a linear system, it's the only example of a linear system. That is any linear system essentially looks like this.

So to do that, I have to tell you what the K is, and I have to actually produce the K – the kernel – against what you integrate. So I have to produce a K for a general linear system. All right? In order to show something like this.

All right. Now to do this, I need a brief digression into the idea of cascading linear systems – following one linear system with another. So we need a little digression here, and I'm not gonna take this very far, I just wanna – in a particular case. So we need a digression on cascading or composing. Engineers tend to use the word cascading. Mathematicians tend to use the word composing linear systems.

The idea being that if you have an input into two linear systems, one linear system following another, say LM. This is V and this is W. Then if L and M are each linear, so is their cascade. If L and M are each linear, then M – L followed by M, W equals ML applied to V. First L then M – you can see by the way I've written it there in that diagrammatic way – is also linear. All right.

I'm not gonna check that. You have seen this before. It's not interesting to check. But it comes up often, all right? The idea of following one linear system with another, and the fact that linearity is preserved under a system like that.

Now, the only way – the only place I wanna use this is when one of the systems is – the first system is given by a integration against the kernel, and then you follow that with another system. So I wanna look at a special case. When – say L is given by integration against the kernel. So L of V of X is, say, the integral for minus infinity to infinity. Once again, K of XY, V of Y, DY. All right?

What happens if you follow that with another linear system L – M, excuse me. So what is M applied to L at V?

All right. That's not hard to see. And again, it's very – it's quite a satisfactory result. The result of this, actually, is applying M to K. So in fact, I'll give you the punch line, and then I'll tell you why it's true. M applied to L at V of X is given by the integral from minus infinity to infinity of M applied to K. Now, let me get a little bit more careful about this. Let me write it down, and I'll tell you what I mean.

I put a little subscript X here indicating M acts on functions of one variable, so to speak. It takes inputs to outputs, where the input is a function and the output is a function. So when I'm writing this, I'm meaning that M is operating on a function K of XY in the X variable, so to speak. All right?

So it's not a great notation, but it's the best I can do. I just want to indicate this. There's a little – something a little bit that has to be said here because K is a function of two variables. So you think of K of XY operate it – operate on it as if Y were fixed, and then operate it with respect to the X variable. And this is what happens when you operate on the system is given by integration against the kernel.

Now let me show you why that's – that's pretty nice, right? That's a nice result. And let me show you why that's true – again, without all the details.

All of these things, if you wanna give a very careful mathematical statement of them, require extra assumptions – require taking limits, knowing when you can take limits and so on. And again, this is beyond our concern. They say the rigger police are off duty.

But why does this work? Well, the idea is you approximate the integral by a sum and then apply ordinary linearity. So K of XY, D of Y, DY is approximately – the integral's approximately the sum. Let me not write the [inaudible]. Just sum over I, K of XYI, V of YI, delta YI. All right? The usual way you approximate an integral by the sums. So I'm leaving X here as the continuous variable and replacing Y by a discreet set of measurements Y1, Y2 – all of them. Whatever. Okay? Because I'm integrating with respect to Y.

Now, if you operate with M on the sum – and here you'll see why – what I mean by M operating in the X variable. So M operating on the sum K, XYI, VYI, delta YI – and it's linear. So a linear operator applied to a sum is the sum of M applied to the components in here. So this is the sum over I of M applied to all this inside – K of XYI, V of YI, delta YI. But again, as far as M is concerned – if M is operating in the X variable, then V evaluator YI and delta YI are constants as far as M is concerned. All right?

So here is where – and here's what I meant. Here's where you can see what I meant by M operating on the X variable. That is, again, because it's linear, the [inaudible] position is M of a constant times a function is a constant times M of the function. In this case, the constants are these V of YI and the delta YI. So this is the sum over I of – let me write it like this: M operating the X variable K of XYI times V of YI delta YI.

And now, think of this as the approximation going the other way. That is, this is the sum approximating the integral. We started off as the integral being approximated by the sum. Now think of this as the sum approximating an integral. So this is approximately the integral for minus infinity to infinity M operating the X variable K of XY, V of Y, DY. All right?

So once again, just to summarize – so if L looks like this: L of V of X is the integral from minus infinity to infinity K of XY, V of Y, DY, then M of L of V of X is the integral from minus infinity to infinity of M operating on K of XY, V of Y, DY. Okay?

This would take a lot of work to actually make this into a completely rigorous proof. And again, I don't wanna do that. Okay?

But it is this sort – and I wouldn't expect you to do it. But it is – but I would expect you to be able to believe the argument. And this is the sort of argument that you should be able to do more and more on your own to see what should the properties be, how should it work and so on to see that it's reasonable. All right? To see that something like this is – try to – if somebody gives you a statement like this, ask yourself why is it reasonable? And to give yourself an argument like that to justify it, I think, is a good thing to develop – to develop being able to do. All right?

Okay. Now, that's all I wanted to say about cascading. Cascading is a big – there are lots of examples. There are lots of different things you can do with it. But I'm only mentioning it toward the end of establishing this result – this really striking result that any linear system is integration against some kernel – can be given as integration against some kernel. So it's back to live action.

Back to the main plot – the main plot being that any linear system is given by integration against a kernel. All right. Now, I'm gonna play – I'm gonna bring back distributions into the picture. All right? And I'm actually gonna write distributions in terms of integration. So when I say integration against a kernel when we follow through using that language – and at the end, I'll say a little bit more generally how you would make this more precise in terms of pairings and distributions and all the rest of that stuff. But put that aside for the moment and just think about distributions and, particularly, delta functions in the way you probably first learned how to use them. That is to say if V of X is any function, I know I can write V of X as the integral from minus infinity to infinity delta X minus Y – it's like convolution with a delta function – V of Y, DY. All right? That expresses V in terms of a delta function. You know what this is, actually – is this is the continuous analogue of expressing a vector in terms of the sum of its components. All right? If you think of the discreet delta functions as just one in one slot and zeros every place else, then writing a vector as V1 times the first basis vector plus V2 times the second basis vector plus V3 times the third basis vector is exactly the continuous analogue of writing V as the integral against a shifted delta function here. It's exactly the same thing.

So again, if I were doing this – and I'll come back to this later. If I were doing this precisely with distributions, I wouldn't talk about integration. I'd talk about a pairing. But think of it in terms of integration now. All right? Because it's the way you're probably a little bit more familiar with, and it allows us to keep up the analogy with the finite case. All right. Well now, if L is any linear system – I'll call it L instead of M – I wanna apply L to V by applying L to that integral. All right? So if L is a linear system, then I find LV by applying L to the integral. All right? That is to say L of V of X is L applied to the integral from minus infinity to infinity V of Y delta – let me write the delta first, there –

delta X minus Y, V of Y, DY. And from what I just said, you can already view that as a linear system – as integration against a kernel, the kernel being the delta function. So what happens is L comes inside and gets applied to the delta function. That is – this is the integral from minus infinity to infinity of L of delta X minus Y. And if I wanna be a little careful here, it's like L applied in the X variable. All right? Delta X minus Y times V of Y, DY. And if you'll look at it, and if you believe in fairies – if you believe in magic, we're done. That is to say now set – say H or K – that K of XY to be L of delta X minus Y, L operating in the X variable. I won't keep writing that, but that's always what's going on here. All right. Then it's exactly – then the system is given exactly in the form of integration against a kernel.

Then we have L of V of X is the integral from minus infinity to infinity of K of XY, V of Y, DY. Amazing, actually. It's really quite striking. Any linear system is given by integration against a kernel. What is the kernel? The kernel is what the system does when you feed it a delta function. Now there's terminology that goes along with this that, again, I'm sure you have heard. Delta is called a respond – delta is called an impulse. When you feed an impulse into the system, the system responds. And so you call K of XY the impulse response. All right. That's the terminology that is always used here. It is how a system responds to feeding it an impulse. It's shifted impulse, not just a delta function of zero. But the delta function at Y – that is to say delta of X minus Y. The system L responds to the input – to an impulsive input, namely delta X minus Y. Okay? And I think I've already broken protocol here because the impulse response is always denoted by H instead of K. So sue me. All right. Before, I was calling things K instead of H. You always – you almost always, for some reason – history seems to dictate that you call the impulse response H of XY instead of K of XY. All right. Maybe I'll go back to that. Now, play it a little fast and loose with the facts here. All right? That is we were writing a delta function. We were writing how do you use a delta function. Well, you use it in terms of integration. We wrote sort of the V of X is equal to the integral from minus infinity to infinity delta X minus Y, V of Y, DY and so on. We applied L to not just an ordinary function, but to a distribution. Does that really all make sense and so on.

Now this whole method and – this method of writing – of working with linear systems was very well to engineers and was used very effectively in a lot of different applications before mathematicians decided to get their hands on it and try to make it rigorous. And again, it was Laurent Schwartz – the same person who founded the rigorous theory of distributions – who also found a rigorous treatment of just this result. All right? And it would probably be considered – I think it's probably fair to say it would be considered the deepest theorem – the hardest theorem in his whole theory of distributions. All right? And I won't write it out in detail, but I think it's worth noting here. And I also want to recall that Schwartz taught at the École Polytechnique in Paris. All right? That's France's leading engineering school, and so I'm sure – quite sure – I don't know the history of this really, and I'd be interested to know – that he was thoroughly familiar with the applications and thoroughly familiar with sort of how engineers viewed things. All right? So I wouldn't be surprised at all if he understood that that was in the context. But being a mathematician, and so socially deviant, he felt like he had to put this on a rigorous foundation, which he did. And it's usually called the Schwartz Kernel Theorem because the deepest fact in the – I think the way I described this in the notes is the deepest fact in the entire theory of distributions is well known – sort of been used every day by every electrical engineer – namely this sort of – the system is given by integration against the impulse response. Schwartz would say it something like this: if M – if L is a linear operator on distributions – so it sends one distribution to another distribution.

And again, you have to make certain assumptions here on continuity or boundedness, but I think the assumptions are generally pretty mild. If L is a linear operator on distributions, then there is a unique – actually, part of the contribution was the unique kernel, K, which is another distribution, so that L of V is given by the pairing of K with V. Okay? He would say it something like this. And you have to understand the nature of the pairing and so on. All right? So there is a precise statement of what I've written down here using delta functions and integrating against delta functions and the whole theory of distributions – I'm not gonna go – I'm not gonna really say anything more about this other than this is – what we said so far can be justified. Although we'll use it in practice like I've written it down there, in fact, it does fit into a more general framework. And it's really nicely and fully established, but it's very – it's quite hard. All right? This is, I think, considered – I think it's fair to say this is probably considered the hardest theorem to prove it in its full generality in the whole theory of distributions. All right? And for us, again, in many cases, we consider the pairing to be given by integration. Now, we know it's not always given by integration. There's a pairing – that's how distributions are defined in terms of a pairing – but in many circumstances, certainly practically working with it, you often view that pairing as given in terms of integration. So that's where the connection is. All right? So this is called the Schwartz Kernel Theorem. I hope I'm spelling his name right. I can never quite remember – Kernel Theorem. Okay? It's a big deal. It's a big deal.

Now, let's do a couple of examples of this. Okay? And again, by the way, the kernel in Schwartz's theory works out to be L applied to the delta function or the delta distribution. All right? So that – it's not as though you sort of prove it in one case not very rigorously and then do something completely different in the case where you're trying to prove it rigorously. Again, in Schwartz's theory, K turns out to be L applied to the delta, but things have to be understood in terms – things have to be understood properly in terms of pairing and what you assume about the operator and so on. But let's look at some examples. All right. What is the impulse response for the Fourier Transform? View it as a linear system. There's a curveball for ya. We've been studying now for a couple of months. Now if you view the Fourier Transform as a linear system, what is its impulse response?

Now you can answer – actually answer this several different ways. All right? You can answer it based on the theorem, which is how I [inaudible], or you can answer it based on the properties of a Fourier Transform that we have seen, namely the Fourier Transform and the definition of the impulse response that we have seen. By definition, the impulse response is the system applied to a shifted delta function. All right? So the Fourier Transform – the impulse response of the Fourier Transform is how the Fourier Transform responds when you apply it to an impulse. So you have to know the Fourier Transform delta X minus Y. And we actually know that. We figured this out. It is E to the minus two pi IXY. Okay?

And actually, that was a fairly straightforward calculation that we did on the basis of the definition – the more general definition of the Fourier Transform. However, let me also point out that you know the Fourier Transform is given by this formula. The Fourier Transform – I'll use variables X and Y instead of S and T – is integral from minus infinity to infinity. Even the minus two pi IXY, F of Y, DY. Okay? Now this exhibits the Fourier Transform as integration against a kernel. All right? The Fourier Transform is a linear system. When you write it like that – is realized as integration against a kernel. Here's the kernel. Okay? The kernel K of XY or H of XY – I'll call it H so – to keep my response – to keep my traditions up – is E to the minus two pi IXY. Schwartz's theorem says – and this is the aspect of Schwartz's theorem that I wanted to mention – is it says there is a unique kernel so that the linear operator is given by the pairing of the kernel with V. All right? And that kernel has to be the impulse response. That kernel has to be L applied to a delta function, and so the fact that the kernel – the fact that the Fourier Transform is given by integration against this kernel implies – and in fact, the kernel is unique – has to be the impulse response – actually implies that the Fourier Transform applied to delta X minus Y, which is how you get the impulse response, has to be given by this complex exponential. You get the reasoning there? All right? It's not circular – complicated way of getting the fact that the Fourier Transform over shifts the delta function as a complex exponential, but it's – it sort of shows how all these ideas are consistent with each other and kind of circle around – not circular reasoning, but how all the things I just circled around – let me just say it one more time.

You know by – so we know that any linear operator has to be given by integration against the impulse response. Let me just not call it kernel – the integration against the impulse response. All right? And furthermore, the impulse response is unique. So if you express – this can sometimes be useful. If you can express your linear operator as integration, then you have found the impulse response. All right? I'll say that one more time. If you – if by – if somebody gives it to you or you can figure it out how your linear system can be written as integration – as an integral of H of XY times the input, then that H of XY has to be the impulse response. All right? So in this case, the Fourier Transform is given by integration against a kernel. That kernel has to be the impulse response. That has to be the Fourier Transform applied to delta X minus Y. Cute. It's cute. I like these cute things.

Let me look at another example. Actually, let me give an example for you to think about. Let me go back briefly. I won't say anything more about it other than to raise the question. We start off by saying this is analogous to the discreet – this is the continuous analogue to the discreet case – the infinite dimensional analogue to the finite dimensional case. Well, what about the finite dimensional case? Finite dimensional case of the discreet finite dimensional discreet case. All right? That is to say if L is a linear operator, L of V plus VA times V where A is a matrix. All right? Any linear system – any finite dimensional linear system is given by multiplication by a matrix. What is the impulse response? All right. All the same words make sense, and all the same reasoning makes sense in the discreet case as it did in the continuous case. You have discreet delta functions. Integration is replaced by sonation. The continuous variables were replaced by discreet variables. It makes sense to talk about – I'm not gonna give the definition for you. All right? I think I actually – it's written down in the notes, but just think about now yourself. All the definitions make sense. It would be good for you to actually reason this through. All right? Yourself in a quite room, sitting there in the dark. What is the impulse response?

Well, I'll give you the answer, but you should be sure that you understand the answer and think it through yourself. In this case, the impulse response is the system itself. This says the impulse response is A – is A. If a system is given by multiplication by A, then A is also the impulse response for this finite dimensional continuous case. So reason that through. Reason that through. That is to say ask yourself and answer to yourself: what do you mean by the impulse response? What is the analogy here? Does it really work the same way? And you'll find – I work with a discreet delta function instead of a continuous delta function, integration has been replaced by summation and so on. All the same words apply in pretty much the same form. It makes sense to talk about the impulse response for a discreet system. What is it? In this case, if it's given by matrix multiplication, then it's the matrix A.

Let's do another example. Let's do an example of the switch. All right? The switch was – one way of reviewing it is V is – L of V is equal to the rectangle function times V. Okay? You switch on for a second. You switch off after that. All right? So this is just a rectangle function center to the origin of duration one half. Okay? What is the impulse response? And let's check to see that it works, actually, also. Let's check to see they actually get the interval – get the system back by integrating against the impulse response. Okay?

So what is it? Well, L of delta X minus Y. I have to find the impulse response, and I have to compute. I have to see how the system responds to an impulse – to the shift in impulse. So that is equal to pi of X times delta X minus Y if I put my variables in there. Right? Now you will remember – and the sort of L operating in the X variable. Now you remember the property of multiplying a function times a delta function. It's the sampling property of the delta function. That is this is pi of Y times delta X minus Y. All right? That's the sampling property of the delta function. So it appears that H of XY – the impulse response to the switch system as a linear system – and it is a linear system because it's multiplication. It's direct proportion. All right?

So H of X minus Y is pi Y times delta X minus Y. And that's the impulse response. Now if you really want – now, if you're a little skittish about this, we can test it. All right? Does it really work? Does it really work meaning is the system really given by integration against the impulse response. So let's check this. Check this that is what if I get the integral from minus infinity to infinity of K – of H of XY, V of Y, DY. That's integrating X against the impulse response is the integral from minus infinity to infinity of pi Y, delta X minus Y, V of Y, DY. Now put the pi together with the V – that is the integral from minus infinity to infinity, delta X minus Y, pi of Y, V of Y, DY. And again, that's sort of the convolution property of the delta function. If I integrate delta against this function: pi of Y times V of Y, I know what I get. I get pi of X times V of X. Okay? And that's the system. So it works. Was there ever any doubt? I ask you. I'm a professional. Okay?

There are many other examples. You'll have some examples in homework. There are other examples that are mentioned in the notes. I think this example is actually done in the notes. Let me look at a special case of integration against the kernel – a very important special case of convolution. All right. Everything that we've said before applies in great generality. Now, I wanna look at the special case where the system is given by convolution – special case of convolution. All right – where L of V is equal to H convolved with V. So H is a fixed function here. The input is V. The output is H convolved with V where H is fixed. All right? Which is, of course, so L of V of X is given by the integral from minus infinity to infinity H of X minus Y, V of Y, DY. All right? So now, if you want, we can appeal the Schwartz Kernel Theorem that says the system is given as a integration against a kernel. That kernel must be the impulse response. All right? This H of X minus Y must be the impulse response. So you conclude. There are other ways of seeing this, as well, but I wanna conclude it from that theorem. You conclude that L of delta X minus Y is equal to H of X minus Y.

Okay? That must be the impulse response. Now you can also check this directly. I'm gonna say that directly from the definition that this is so. All right? Again, I'm concluding it from this uniqueness statement in this very general Schwartz Kernel Theorem. Once again, if the linear system is given by pairing with a kernel or integration against a kernel, then that kernel must be the impulse response. All right? Now, there is a special property – and I alluded to this before, and I know you've seen this. There's a special property of convolution that makes it particularly important for linear systems. I mean, we've seen the importance of convolution in all sorts of different aspects – in many guises. All right?

For linear systems, it's the relationship between convolution and delay that turns out to be worth singling out. All right? There's a relationship – a simple relationship between convolution and delay or shift – same thing. All right? So let me remind you – all right – let me remind you of – I wanna write it in terms of this delay operator. All right? So tal sub A of V at X is just delay V by A – V of X minus A. Incidentally, you could ask – I may even ask you at some point – is tal sub A as the delay operator a linear operator? And what's its impulse response? I may ask you that. But now – don't worry. Never mind that right now. It's a delay operator. We've seen this before. And you showed on homework that the delay of the convolution is the convolution of the delay in words. Very nice to say that in words. You showed that the convolution of a delay – of a delayed signal is the delay of the convolution of a signal – of the convolution. That is to say, you showed that if I take H convolved with a delayed signal – this is – so I'm writing this without writing any variables here. But that's to say if I first delay V and then take the convolution, that's the same thing as taking the convolution and then delaying the whole thing.

This is tal sub A of H convolved with V. Okay? You showed that. I can first delay the signal and then convolve, or I can convolve and then delay the signal, and I get the same result. Okay? Now, I just wanna reinterpret this in terms of linear systems. That's a standard result, and it's a very important result. I'm not gonna say anything different here. I'm just gonna give it, now, a different interpretation – reinterpret in terms of linear systems. All right? It says if the system is given by, again, convolution – here's the system: L of V is equal to H star V. So it's given by convolution. All right? It says if V is delayed – so let's say W is the output. W is equal to L of V – that's H star V – H convolved with V. All right?

So it says if I delay – a delay of the input – that is V going into V of T minus A – or X minus A, all right – causes an identical delay of the output. That is if V goes into W by L – that's H star V – all right? And if V gets delayed to V of T minus A – that's tal sub A applied to V – then the delayed output goes to the delayed input – the delayed input goes to the delayed output W of T minus A. This is H convolved with the delayed signal, but that's the same as – this is tal sub A of W. You like this? That's quite a diagram. That's the same as tal sub A of H star V. All right? Make sure you understand all the different arrows in this diagram and what's happening here. All right? V goes to W by convolution. That's what the system is doing. All right? If V is delayed by applying the delay operator, that means V goes to T minus A. All right? What happens to the output? Well, this is the fundamental relationship between delay and convolution. The convolution of a delay is the same as the delay of the convolution. All right? So if I delay V, and I convolve with H, that's applying L to the delayed signal. Convolving H with the delay of V is applying L to the delayed signal. All right? So this is L of tal sub A of V.

Applying H to the delayed signal is the same thing as applying the delay to the original – to L of the signal. All right? And that produces, therefore, a delay of the output. So delayed inputs go to delayed outputs by the same amount in the case of convolution. And because of that property, you say that the system is time invariant or – it's better to say shift invariant because you don't always think of the variables representing time. But nevertheless, it's almost always referred to as a time invariant system. And I know you guys have seen this before, probably in a signals and systems class. There is a mistake of assuming that every single – every system you see is time invariant because it is such an important class of systems, but it's not the only class of systems. It is the class of systems that is associated intimately with convolution. So again, what have we shown here? So you say that L is, as a system, time invariant or shift invariant. All right? So again, what that means is that if W is equal to L of V, then W at X minus A is equal to L applied to V of X minus A. Shifted inputs go to shifted outputs. That's the definition of time invariance. All right?

Now that's a general definition, and what we just showed was that convolution is a time invariant or shift invariant system. This definition can apply to any linear system. All right? You can ask the question of any linear system: is it time invariant? Is it shift invariant? In which case, you are asking is that statement true? If I delay the input, does that cause a corresponding delay in the output? All right? And what we just showed was that if the system is given by convolution, then that is the case. All right? So we just saw that if the system is given by convolution, then you have time invariance. All right?

Now, the remarkable thing is the converse is also true. I'm just tingling to think about it, but that's me. All right? The converse is also true. What does the converse say? The converse says if you have a time invariance system, then it must be given by convolution. I.e., if L is time invariant, then L is given by convolution. All right. Now, I'm gonna – and I'm gonna show this to you in just a second. This is another indication – one that we haven't seen earlier and maybe the last sort of one we'll see. The general property of convolution indicates why convolution is such an important – and in some case – in many ways such a natural operation. I mean, it was an odd thing to write down – when we first wrote down convolution – where the hell did that come from? And why would anybody think of writing anything like that down? But it's not so – from the point of view of linear systems, it's not so odd to ask it be time invariant. It's not so odd to ask that a system be such that, if you delay the input, you also get a delay of the output by the same amount. That's not – that's a reasonable thing to expect of a system. All right? If I run a program today, I get an answer. If I run a program tomorrow, I'm gonna get the same answer delayed by 24 hours – delayed by a certain amount of time, but I'm gonna get the same result, but I'm gonna get the same result delayed. All right?

If I flip a switch today and something happens, if I flip the switch tomorrow, the same thing happens, we hope. So the input is the same as the output but delayed. Actually flipping the switch may not be a good example because it's not time invariant. Never mind. But you know what I mean? You get the idea that it's not an unnatural thing to say. And the fact is, a time invariant system has to be given by convolution. Now why. All right. We know that it's given by integration against the kernel. All right? So we know that L of V of X is given by the integral from minus infinity to infinity of the impulse – of the impulse response integrated against the function L of delta X minus Y of V of Y times V of Y, DY. We know that for any linear system. All right? That's the impulse response. This is L applied to delta. Okay? So the question is what is that?

So what is L of delta X minus Y? It's gonna be some general thing, but it's gonna be a general thing of a very – of a special type. That is instead of the following, all right? So let me let L of delta X – the unshifted delta function – be H of X. All right? I'm gonna define H of X by that. Okay? Then L of delta X minus Y – think of it this way – is L applied to the shifted delta function. Delta X of Y applied to delta. Okay? But if it's a time invariant system, a shift in input goes to the shift in output. The shift in input goes to the shift in output. That is this is the same thing as tal sub Y of L applied to delta – perhaps I should have said that over here when I was talking about generalities. All right? To say that a shifted input – matter of fact, let me just say that. It says that if a shift – I mean, I sort of said this, but I didn't write it down. If I shift the input by A, that's the same thing as shifting the output by A. All right? So symbolically, I would write it like that. I'm sorry. I probably should have said that. All right?

So this says L applied to the shifted input is the same thing as shifting the output. So another way of saying that, actually, quite nice is that L commutes with shifts. That's the only way of saying – that's probably the quickest mathematical way of saying what a time invariant system is. It commutes with a delay operator. The linear system commutes with the delay operator. First delay then apply L as the same thing as first applying L and then applying delay. All right. So back to live action, once again, if I shift the inputs, that's shifting tal by Y. That's the same thing as shifting the output. That is tal sub Y apply – of H of X. That is H of – we'll put an X in there. That is H of X minus Y. Okay? Now, it doesn't look like I did much there. That's where the assumption of time invariance comes in – right at this stage. Okay? It makes sense. This definition makes sense. All right? To define – I'm defining function H by how L responds to the impulse centered at zero. But if I know the system is time invariant, and if I know how the system responds to an impulse at zero, I know how it responds to any delayed impulse. It responds to the delayed impulse by delaying the output – delaying the response. Right?

That is L of the delayed signal is the delay of L of a signal – L of the impulse H of X minus Y. All right. So what does that say? You know it's given by this. So that says that L of V of X is equal to the integral from minus infinity to infinity of H of X minus Y, V of Y, DY because it's always given by the integral against the impulse response. Where H of X is equal to L of delta X and H of X minus Y is equal to L of delta X minus Y. The difference between a time invariant system and a general linear system is the impulse response depends only on the difference of the variables X and Y rather than on X and Y separately. All right? It depends only on the difference. And I mentioned this, actually, when we were first talking about linear systems and so on and delays. All right? The difference between a general linear system and a time invariant system is the impulse response is not a function of X and Y independently, but is rather a function of X minus Y. Now this is – so for example, you can see that the switch is not time invariant. All right. Let's go back to that example.

All right, back to the switch. V of W is equal to pi times V. All right? Or L of V is equal to the rectangle function times V. What was the impulse response? The impulse response was, if I remember right – what was it? Was it pi of Y times delta X minus Y. Right? Is that right? Remind me. Yeah. Okay. Is that a function only of X minus Y? No. All right? Not quite. It's not a function of X minus Y alone. It's pi of Y times delta X – this is a function only of X minus Y, but this isn't. All right? Not of the form function of X minus Y alone. Okay? The switch is not a time invariance system. This can be a little disappointing, right? The simplest linear system – the relation of direct proportion multiplication is actually not a time invariant system. It's only when you actually complicate sort of direct proportion and multiplication into this operation of convolution that you actually get a time invariance system. All right. Now, let's summarize. This is a very – I think it's a very – it's a beautiful state of affairs intellectually. I mean, it's a very satisfactory state of affairs. Okay? There are two things that we've really shown today. We've shown that any linear system is integration against a kernel.

As a matter of fact, let me just say integration – well, integration against a kernel – and the kernel's the impulse response – any linear system. All right? The special case of the time invariance system is also the integration against a kernel, but it's a special integration. It's convolution. The impulse response is a function of X minus Y – the difference of the variables, not the variables separately. Two, a system is time invariant if and only if it is given by convolution. I think that's just gorgeous. All right? I think intellectually, that's just very satisfying. Every electrical engineer knows this somehow.

Right? You're all taught this. Maybe not quite in the sweeping grandeur of general linear systems, but it's a very satisfactory state of affairs. You have – any linear system is given by integration against a kernel. That has to be if you really want to understand the theory of it. It gets quite complicated. But never mind. All right. There's the impulse response – linear systems given by integration against the impulse response. In the special important case of time invariance systems, the impulse response takes a special form. The linear system takes a special form. The integration takes a special form of convolution. And if nothing else, it's yet another indication of how fundamental an operation of convolution really is. Okay? Now – one second. We're going. We're about to go. We are about to get a week off. All right? The fact that convolution is entered into the picture now means that the Fourier Transform cannot be far behind because as soon as somebody says convolution to you, your trained behavior at this point should be to salivate and say, "I'll take the Fourier Transform. I know I should. I should do that. Yes. Yes. Yes. Yes. Yes. I'll take the Fourier Transform." And so we will. All right? Anytime convolution comes into the picture, the Fourier Transform cannot be far behind, but it will be a week delayed. All right? So I hope everybody has a wonderful holiday. I will see everybody after Thanksgiving.

[End of Audio]

Duration: 57 minutes

TheFourierTransformAndItsApplications-Lecture25

Instructor (Brad Osgood):Hey. Jesus. Man, I just back and I'm tired. And welcome back everybody. Let's see if we can get our head back in the game. It's not so easy, somehow. I'm sure I speak for all of you. Anyway, I hope everybody had a very good holiday, either here or elsewhere, and we can now sprint to the finish. All right.

I want to wrap up today, the discussion on LTI systems. There are a lot of topics to do a lot of little things to do, a lot of big things to do, and like many things in this class, it goes off in a lot of different directions, and is often the subject of very specialized courses. I was going to do some material on a little bit more on filters, on digital filters today, but I decided not to that. It's discussed in the notes, and there's, of course, we have an entire course of digital filters. So you have plenty of opportunity to see it. If you don't see it there you'll see it elsewhere, worked into other courses.

So I thought I do some fairly general things. Do a couple sample calculations and talk about their relationship, the connection between the LTI, Linear Time Invariant systems and the 4A transformer, which is the main sort of important fundamental foundational information that I think everybody should maybe know.

So I want to remind you where we finished up last time. Last time we got a pretty satisfactory answer about the general structure of linear systems in terms of the impulse response. So if L is the linear system, I'll label it linear system, and we introduce the impulse response, a function of two variables separately, L of XY is L of delta X minus Y. And then the basic result, which is known in the theory of distributions of the Schwartz quantum theorem, but for us, it's something you probably heard about when you had your first course in signals and systems, that the output of the system can be given in terms of integrating the impulse response against the input. All right.

So if W of X is equal to LV of X, then you actually get W of X is [inaudible] for minus infinity of H of XY, Z of Y, DY. All right. So once again, the output of the system is obtained by taking input of the system and integrating that against the impulse response. It's a very satisfactory result.

Now, in the special case where we have an LTI system, the integrating reduces the convolution. So let me remind you what an LTI system is. You say that L is time invariant or shift invariant, invariant, let me get the word better here, or shift invariant if the following happens: Say if W is equal to L of V of X, then W of X minus Y is equal to L of V minus XY. I write it in symbols. It's easier to say it in words. L has time invariant if the delay of input was also in corresponding delay of the output. All right. So if V is the output of L, V is the input and W is the output, V of X minus Y is the delayed form of V, then W of X minus Y is the delay formula of W and its delayed input is in correspondence with its delayed output. And a system is time invariant or shift invariant if and only if the impulse response is given by or the interval is given by convolution. Okay. So in this case, the impulse response for an LTI system is a little bit easier. Delta of X, I just set to the L of – I'll get it again, I'll get it right. H of X is L of the un shifted delta function, say delta of X, and then by the time invariants, H of X minus Y is equal to L of delta X minus Y, so that is the impulse response. And the action of the system is given by W of X the output is the interval from minus infinity to infinity, H of X minus Y, DY, DY. Same form, right, integrating the output of systems by integrating against the impulse response, but now the impulse response is of a special form. It doesn't depend on X and Y separately, it depends only on the difference. And we recognize this as a convolution. H star V FX, and in fact, this characteristic of time invariant systems. That is to say, a system is time invariant, a linear system is time invariant if, and only if, it's given by convolution. So system is time invariant if, and only if it's given by convolution. All right. That's where we finished up last time. All right and it's in a very satisfactory state of affairs, as far as a structure of linear systems go. Any linear system is given by integration against the impulse response this is the time invariant if, and only if, that integration reduces to a convolution. So it's another indication how fundamental convolution is in the whole theory, all right, or just as an operation.

We're going to see how 4A transformer comes into this in just a second because anytime, should be one of the great lessons of this class, is that anytime anybody mentions convolution, bells should go off your head suggesting that you take 4A transformer. But wait! We'll do that in just a minute. I did want to comment, all of this in the context of continuous time systems, but I did want to comment, as a matter of fact just write down a simple system, that the same sort of consideration holds when you have discrete systems. The same considerations hold for discrete systems. All right. Now, any discrete system, remember, is multiplication by a matrix. All right. If W is equal to L of V, then this is given by – I'm thinking about this W and V as vectors here, is given by multiplication by a matrix. All right. We talked about that before. Any linear operator is given by multiplication by matrix. Any functional operator is given by multiplication by matrix. And it is the definition of shift in variance, of time variance is the same as before, except this time you're shifting a discrete variable instead of a continuous variable. And again, you have a system that is time in variant or shift in variant if, and only if, it is given by convolution. In this case, vector convolution or matrix convolution, well, vector convolution. So L is an LTI if, and only if, W is equal to H star V. Okay?

Now, in this case, so here is the impulse response, D is the input vector and W is the output vector. All right. So again, H is L of delta. All right, the M shifted to delta function. And H of M minus N is L of delta, so let me write this, minus M is L of the delta function shifted to M. Now, it's interesting, I just wanted to share this example. I want to do a couple calculations today so you'll feel sort of comfortable with how these things work out. The matrix A that realizes though the linear system has special form in the case of a timing variance system. It's cute and actually, it's very important in a lot of number calculations. So again, the operator is given by matrix multiplication. If we write a system as a matrix multiplication, say again, let me write it A times V, all right, where A is a matrix. Then A is special form for time and variance systems. Let me just do it, rather than try and give you the state of general fear here, let me just do an example so you see how it works.

All right. So let's take e.g., let's take just a four by four system. So I'm going to take H to B and the matrix or the vector, one, two, three, four. All right, just to take a random example that I happen to work out in detail before I got here so I wouldn't make any mistakes. So if W is equal to A times V, which is also given by H convolved with V, the question is what is A? All right. So I'm telling you the system is given by convolution. All right. H star V, where H is this vector. So even when your system's given by matrix multiplication, the question is what is a matrix? All right. So it's clear what I'm asking here? Well, how do you find the matrix A? How do you find any matrix A? You have to find the image of the basis factors. All right. The columns of A are given by A of – well, the first basis vector is just what we're calling, in fancy language, delta naught. Delta naught is 1000, right? That's the first column. The second column is A delta one, delta one is 0100, if I use the language of delta functions instead of the language of lie algebra. Second column, the next column, is A delta two, and the next column is A delta three. So A delta two, delta two is 0010. And the next thing here, remember there's always this problem when you're working in the sort of context of linear systems, DF etc., wherein the index is usually from zero to N minus one, instead of one to M. All right. So delta zero is this is the zero slot, the first slot, the second slot, the third slot, 0123, and so on. And the final column is A delta three, where delta three is the last basis vector, that's 0001.

All right, so how do I compute all these things? Well, I compute them by convolution. All right, because by definition the system is given to you as convolution with the vector H. So A of delta to naught is H convolved with delta naught. All right, now, what is H convolved with delta naught? Wait! Don't tell me, I know, it's H. All right. Convolving it with the un shifted delta function doesn't do anything to the vector, it's H. All right, so that's one, two, three, four. That letter's a column. Okay. What about A of delta one? What is A delta one? It's H convolved with delta one. Now, what do you get if you convolve H with the shifted delta function? It's a shifted H. All right. H convolved with delta one at the index M is H of M minus one. Okay. Just like the continuous case, just like the continuous case, all right. So what is that? Well, now, here's where you have to say something a little extra. All right, if H is the vector of one, two, three, four, what is H shifted by one? All right. Now, you have to use the fact that H has to be assumed. Any time convolution comes into the picture, we haven't brought the DFT in, although we will, but anytime any of that sort of stuff comes in, you always have to assume that your signals, you're discrete signals are extended to be periodic. All right. So that it makes sense to consider H for values other than the index, and to see zero, one, two, three, remember it's index one, two, three, all right. So it makes sense to consider H defined for all integers and you just keep repeating the pattern. All right. So what is H convolved with delta one as a vector, he looks at his note to make sure. I just forget what you shifted. It's shifted like one to the right, right? You know, if I look at F minus one, it's like shifts the function over. Well, one, two, three. All right, make sure you see this, okay?

Again, you have to assume that H is extended to be periodic, and it's shifted down by one. So if it's shifted down by one the four goes up top. Or you can think about it this way, the zero component here is H of minus one. Right? Delta one convolved with H at zero is H of minus one. But H of minus one is the same thing as H of three because of the periodicity. And H of three is the third component, remember, we're indexing zero, one, two three, so that's four, and so on. Okay? What about the rest of them? Now you see what the pattern is. What is A of delta two, is that where I am now? Yes, that's H convolved with delta two. So that's H shifted by two or if I shift this thing one more, so what would this be? I ask. Pardon me?

Student:[Inaudible].

Instructor (Brad Osgood):Be bold.

Student:[Inaudible].

Instructor (Brad Osgood):Thank you! All right. Shift it down again. And finally HA of delta three, is H involved with delta three, that's just H by three, and that is two, three, four, one, right? Yeah. All right.

Now, again, those are the four columns of the matrix A. So what is the matrix A? What is the matrix A? Or simply what is the matrix? Neo. A is one, two, three, four, that's the first column. The second column is, what do I have here, four, one, two three. Third column was three, four, one, two, four by four matrix. And the fourth column is two three, four, one. All right.

So again, and you can check, you can check that this is a different description of the system. As the system is given by convolution but it's also given by matrix multiplication. W is equal to A times V. That is A times V, multiplying matrices out, is the same thing as convolving with H.

Now, this is kind of cool matrix. All right, if you look at this matrix, this is what's called a circulant matrix in the biz. And I think I actually mentioned this once before. Someone asked a little bit about this. Circulant matrix. All right, this is a special case of more general matrices called Toeplitz matrices. They come up in a lot of different applications in discrete systems, all right. Circulant matrix is constant on the diagonals. The columns are periodic, as they are in this case, so the pattern just repeats the cycles around and each column is obtained form the previous column by a shift. And consequently, it's constant on the diagonals, so it constant on the main diagonals, all ones, fours here, threes here, twos, twos, threes, four, and so on. Okay. And it's called circulant. That sort of property, if it's constant on the diagonal, and I'm a little hesitant to give you the general definition here, because I don't want to get it wrong. But there's standard terminology of the standard matrices that come up a lot in various applications. Typically it's Toeplitz matrices and circulant matrices and they have the circulant matrices are like Toeplitz matrices except they have the additional property that columns are periodic. Okay. But each one is obtained from the previous one by shift. Okay. Bob Gray, in our department, has a whole book on, he has a whole book on a lot of things, actually, but you know there's a whole book in particular on Toeplitz matrices and their applications. So they come up a lot. We'll come back to this matrix a little bit later. All right. It's kind of cool.

And it's the sort of calculation you should be able to do, all right, you know. You should be able to take this result in the continuous case, and bring it over to this brief case and realize what form it takes, and realize that it's not so different than what you're doing in the continuous case. We set it up this way just so the formalism just so that symbols and everything else would look, just as much as possible, like the continuous case. It's nice. Okay.

Now, at long last, let's bring back the 4A transform for LTI systems. Okay. LTI systems get convolution, bells should go off in your head, buzzers should go off in your pocket, who knows what else should happen, but whatever happens, you should think of the 4A transform. Bring in the 4A transform. All right, so with an LTI system we have convolution. So if W is equal to 8 involved with V, V is the input, W is the output, H is the impulse response, so H is fixed and V varies over different inputs. and if you take the 4A transform you, of course, get via the convolution theory 4A transform value is the transform of H times the 4A transform of V, or as it is universally written in uppercase letters, capital W is equal to capital H of S times capital V of S. And in this context, in terminology, again, I'm sure you have heard, capital H of S is called the transfer function. Little H is called the impulse response capital H is called the function. You have to be a little careful here. That's all right, never mind. Sometimes what you call the impulse response, what you call the impulse response H of X or H of X minus Y, I guess that's the only question. But it's not important in this case.

Anyway, the standard terminology I gave you, again, that I am sure you have heard, I think we even used back when we talked about filters when we first started talking about part of convolution, is that capital H is called the transform function. It's also sometimes called the system function. Any other terms of this anybody knows, just out of curiosity. Other than capital H function, I don't know. Either call it the system function or the transform function. Now, I want to point out something here, again, sort of in the spirit of this such beautiful structure involved in linear systems. When we start talking about linear systems, I said, I made the bold statement, that the most basic example of linear system is, is the relationship in direct proportionality. The output is directly proportional to the input. All right. And I said, boldly, that any linear system is somehow some generalization of that, or somehow you can trace back the idea of direct proportionality or you can redirect proportionality into any linear system. And for LTI systems, this is staring you in the face. All right. Because what this says is that for an LTI system in the frequency domain it's exactly described by multiplication. It's exactly described by direct proportionality. Okay. In the frequency domain, the system is really given by direct proportion, the relationship of direct proportion. In the time domain, it's a little more complicated. In the time domain, it involves convolution, but in the frequency domain, it really is the relationship between the input and the output is given by direct proportion, the most basic relationship that underlies linearity. All Right.

And of course, again, we have you know part of the point of this course is that the time domain and the frequency domain in some sense are equivalent. You can pass back and forth between the two. They're different pictures of the same thing, different views of the same phenomena. You can use one to study the other. All right. So again, I just want to point this out because I think it's just another, I don't know, an example of how beautifully unified and coherent this subject is when you talk about linearity and frequent of time and variance of convolution, this whole idea, again, direct proportion comes out very strongly here. Okay. It's not just almost there, it's there! It's right in front of you. Now, the importance of bringing the fully transform to LTI systems is the fact that it would not be obvious if you didn't do it with 4A transforms that complex exponentials are agin functions for all LTI systems. Let me write that down and then I'll explain what I mean. This is sort of a last general fact that we're going to talk about for linear systems and LTI systems in particular. For the last gray fact on LTI systems is the complex exponentials are agin functions. Now, this actually is an extremely important result but we are not going to take it anywhere. All right. I've got to say, just because again, this goes off in a lot of different directions and becomes, you see this more in special applications. And I would be surprised if you didn't see this in special applications. But for us, I just want to make sure you understand where it comes about, and why, and how it happens, what the basic definition is. So we're not going to do any particular application with this because to do one application is to do dozens of applications, probably, and again, you'll probably see these more in other courses that come up. Quantum mechanics comes up in a lot of various aspects of sequel processing, but I just want to make sure you see the basic fact. So here's what I mean by this, so then L of V is given by H star V. Okay. So it's a timing variance system, let's call it W on the left side. W is the output, V is the input, and if I take the 4A transform I get W of S is equal to H of S times V of S. Now, what happens if I input a complex exponential into the system? Input V of X is E of the two pi I nu times X, any nu. Okay. And the question is what is L of V of X? I think sometimes people call this the frequency response because you're inputting a pure frequency or a pure harmonic. All right, but I tend not to use that term.

Anyway, what is it? Well, first of all, what is a 4A transformer? A 4A transformer either a two pi I nu X – ladies and gentlemen, let's work in the frequency domain. That is, work in the relationship between the 4A transformer of the output function of the 4A transformer, the input function, and the transfer function. All right, so the 4A transform of the, you know, the two pi nu X is delta X minus nu. Okay. It gives you the shift of delta function. And so the output is W of S, the output in the frequency domain is W of S is H of S times delta S minus nu. But now, there's the fundamental sampling property of the delta function. H of S times delta S minus nu is H of nu, a constant times delta S minus nu. All right. Now, take the universe 4A transformer on both sides. Go back to the time domain. If I go back to the time domain, then H of nu, that is to say, if I take the 4A transform, H of nu just comes out, it's kind of along for the ride because it's a constant, the universe 4A transform of delta, the shifted delta function is a complex exponential, again, so I get W of X is equal to H of nu times E into pi to I nu X. All right. In other words, remember, the input was into the nu pi X and then the output is a multiple of that, mainly the value of the transfer function at nu. So i.e. L of E of the two pi I nu X is equal to H of nu times into the two pi I nu X. All right. That says exactly, that either the two pi, the complex exponential, either the two pi nu X is an agin function with agin value H of nu. That's exactly what that statement means. Okay. This says, all right, this says exactly, either to two I pi nu X is an agin function, agin function, of any LTI system, of any LTI system. Now, it doesn't always have the same agin value because the LTI system depends on the particular transfer function. The agin value for a given LTI system are values of the transfer function at the frequency nu. The agin values, request for the agin value is H of nu, the value of the transfer function at nu. All right. That's a fundamental fact. And again, some people interpret this, you know, often, I think I probably even put this thing in the notes, time invariance is a further indication of how natural and important convolution is from the fact that complex exponentials are agin functions is, I don't know, either a further statement of how important shifting invariant or linear timing invariant systems are, or how important complex exponentials are. All right. The fact that they enter into the theory of linear systems in such an important way, and I mean, it's very important. And if you're analyzing a linear system, is I'm sure you've seen in various classes, to know the possible agin values an agin vectors, and to know that there are agin values and agin vectors. I say agin function here, instead of agin vector, because I'm thinking of functions instead of vectors. That is, a function of the continuous invariable. So I'll do a discrete action, actually. But same idea, it's the same terminology in linear algebra but just applied, as you use your linear algebra, but just applied to sort of the continuous case. Okay.

So any LIT system has complex exponentials as, actually, as it turns out, as a basis of agin functions, and that turns out to be very important. That allows you to diagonalize the operators associated with LTI systems and understand how they operate in a much more natural way. But again, as I say, to do one application is to do a lot of applications so I would rather let that wait for other occasions. But it's also important to realize that, you know, we often talk about how you can work with complex exponentials but, you know, but really you're sometimes thinking about real signals so you take the real part and put a sign, you know, properties of complex exponentials are the same as the properties of signs and cosigns, and so on and so on. But this is a case where that's not true. All right. That is, it's the complex exponentials that are agin functions of linear time invariant systems, not the sign and cosign separately. So let me show you that. So it's not true without additional assumptions. There are some cases where it's true, but generally, it's not true that sign and cosign are themselves agin functions of an LTI system. All right. Watch what happens here. So let's take cosign for example. So e.g. take Z of X to a cosign, cosign of two pi nu times X. All right. And the question is, what is L of VX. If it's an agin function L of VX has to a multiple of itself. It has a be a multiple of V of X. All right. Well, is it or not? We can calculate this by expressing to cosign in terms of complex exponentials as the real part. So L of cosign of two pi nu X is L of ½ E to the two pi I nu X plus E to the ½ of the sum, plus E to the minus the two pi I nu X. All right. L will apply to the sum. The ½ comes out and L is linear, so L of each one of those sum of the complex exponentials is the sum of L applied to the complex exponentials. And we know what happens in that case, actually. So this is ½ L of E to the two pi I nu X plus L of E of the minus two pi I nu X. All right. And the complex exponentials are agin functions. All right. So this is ½ H of nu times E to the two pi I nu X plus H of minus nu, you know, the minus two pi nu X is E to the two pi minus nu times X of H minus nu X equal minus two pi I nu X, on half of that whole thing, and now, you're stuck. All right. You are stuck. Because unless you have additional assumptions, all right, you can't combine these terms. You have H of nu, H of minus nu, and without further assumptions, they don't have anything to do with each other. Okay. So you are now stuck without further assumptions. Okay. So it's an agin function without further assumptions.

Now, there are assumptions even made. Actually, one of the most natural assumptions, which almost gets you there, but not quite, is to assume that H is real. L of H, the impulse response, which is a natural assumption, is real, as it will be in most cases, all right. If H is real, then what symmetries does the 4A transform have? If H is real, then H has asymmetry, H of minus nu is equal to H of nu bar. That's the basic symmetry of the 4A transform, when you have the 4A transform of the real signal. Okay. So if I plug this in, let's go with that assumption. And if I go with that assumption, then L of again, cosign of two pi nu X is equal to ½ H of nu either the two pi I nu X plus H of nu bar, and let me write this as, actually, E to the two pi I nu X bar. All right, either minus two pi nu X is a congregant of E to the two pi nu X. So that is the real part of H of nu E to the two pi I nu X. Now, you're still not quite there. Right? You're still stuck. If H of nu were real, then you'd be okay. Right? But I'm not assuming H of nu is real, just that H satisfies the symmetry property. There's a little more you can do, all right. Let' write, so you're still stuck in the sense that it's not an agin function, it's just not. So don't say that it is. Don't make me mad. All right? So if you write, though, there's a little bit more that you do on it, and it's the common thing to do, is if I write H in terms of its polar form, H of nu, the magnitude times E to say IC. You've got to say E to the IC before you can say it, so in terms of the magnitude of the phase. Then H of nu times E to the two pi I nu X is I to the value of H of nu times E of the I five times this is E to the two pi I nu X plus E. Okay. Or put the two pi, yeah, I guess I'll write it like that.

So two pi I – I'll get it, I'll get it, don't panic. I times two pi nu X, here we go, plus V. Okay. So the real part of this does give you a cosign, but it's a phase shifted cosign. The real part of H of nu times E to two pi I nu X is going to be absolute value of H of nu times cosign of – so the real part of this is, this is already real, the real part of this is a cosign with a phase shift. So it's cosign of two pi nu X plus V. Okay. All phase shifts always drive people out. All right. So it's still not an agin function, right, but it's as close as you're sort of going to get. This says L of cosign of two pi nu X is equal to absolute value of H of nu times cosign of two pi nu X plus C. All right, the cosign is not an agin function but it's close. Okay? Yeah.

Student:[Inaudible] even.

Instructor (Brad Osgood):Well, then you can tend your business. The more assumptions you can put on this, because then you know you'll be okay. All right?

There are extra assumptions you can put on here, but all I'm saying is that if you don't put those extra assumptions on, it's just not the case that the signs and cosigns are separately agin functions for LTI systems. Only the complex that's really interesting. I mean, it's almost really, sort of the fundamental difference between the complex exponentials world and the real world is that for any LTI system, complex exponentials are agin functions, but not the real imaginary parts are not agin functions without extra assumptions. Okay.

All right, let's finish up. Let's do the discrete case. This is the discrete version of this. Again, same considerations hold for the discrete case. A discrete case is W is equal to L of V, which is H involved with V, but everything here now is a discrete signal. Again, W of M is the 4A transform of H times the transform of V, that everything is discrete here. Okay. And again, discrete complex exponentials are agin functions. Discrete complex exponentials are agin functions. All right. Maybe in this case I should call them agin vectors, I suppose. All right.

So for example, what if I input V equals omega to the K for any case. Or maybe, as a discrete vector complex exponential, right, that we use many times, but let me put it to the omega K. Well, the 4A transform of omega to the K is, if you will recall, I don't recall so I have to write it down, is N times delta K. All right. There's that extra damn factor of N in there that comes in. What are you going to do? It just is, it's a pain in the neck, but here it is.

So again, what is L of VK? So to find L of omega to the K, I thought that was EK, thought it was Ek, omega to the K, look in the frequency domain, same argument as before. And I get W of, say, N is equal to H of N times – let me just write it like this, without the indexes, W is equal to H times this, N times delta K. All right. So that's equal to H of K times N times delta K. Because of the sampling property of the discrete delta function, same property, same damn thing. It's the same damn thing over and over, again. It's the same argument

So back in the time domain, it's the same thing. W is equal to H of K times omega to the K. Okay. That is to say, i.e., L of omega to the K, the discrete complex exponentials of the K, the discrete complex exponential is H of K times omega to the K.

Now, an interesting thing happens here because whereas, in the continuous case you had sort of a infinite family discrete complex exponentials, even when the two pi nu X where nu can be anything and a continuous variable between minus infinity and plus infinity. Here these powers cycle, right. Up to zero minus to the first power and then they start repeating. All right. So what you have is, this is maybe a difference or this is a special feature of the discrete case, so here, now, you see that one omega to the one, omega square, up to omega to the end minus one, form a basis of agin vectors for any linear time invariant system. They're independent, they're [inaudible], they're each agin vectors, they actually form a basis, and almost a normal basis. They're not quite worth normal, right, because the lengths are N no one. Damn. All right. Form the basis of agin vectors for any LTI system. All right. Now, I want to make sure you understand, again, what this says and what this doesn't say. All right. Any LTI system, these discrete complex exponentials, one, this is a vector will all ones in it, all right. One omega, omega square up to omega and minus one, any LTI system, these are agin vectors and they form a basis of agin vectors. All right. The agin values are different. The agin values depend on the system. Because the agin values are the values of the transfer function at the index K and that's going to be different for different systems. All right, because H is going to be different for different systems. To define a system, a discrete LTI system, is to give the H, and so that gives the agin value in these terms. But the vectors, themselves, one, omega, omega square, four omega N minus one are agin vectors for any LTI system. Another way of putting this is, any LTI system, so a system given by convolution in this discrete case, is diagonalized by the complex exponentials. All right. This is another important property, makes for good quals questions. All right. It's an important property of discrete complex exponentials that they form a basis of agin vectors for any LTI system.

All right, now, I'll do one more calculation for you. Let's take that system we had before. Let's do a 263 problem in a slightly way. All right. Let's take, again, W is equal to H involved with V, in the discrete case where H was the vector one, two, three, four. All right. And we found that this was given by matrix multiplication, W is equal to A times V where A was this matrix, right. A was the matrix I don't have to write it down again. One, two, three, four, that's the first column, one, two, three, four. And I start shifting. Four, one, two, three; then three, four, one, two; he said looking at his notes desperately, then two, three, four, one. All right. All right. That's the matrix. All right. Now, agin vectors of the system, then therefore, are agin vectors of A. All right. Of the system are agin vectors of A. I should say agin vectors and agin values are agin vectors and agin values of the matrix A. All right. Let's find them. Now, you know how to do that, actually, by matrix methods where you look at A minus 11 times the identity, and figure out the determinant, and you know, figure out the roots and the characters of the components. No, no, that's the thing where you just plug in the map lab and let the map lab chug away, and so on, and so on. Right. So you can do this. But let's solve this problem by using what we know. Okay. So let's do this via a theory of LTI systems. The agin values are given by values of the transfer function. All right. So we need the transfer function. Agin values are H of zero, H of one, remember I'm using that same one, two, three, H of two, and H of three. Those are the agin values. I mean, actually, I know what the agin vectors are. The agin vectors are, of course, are complex exponentials. All right.

So how do I find those numbers? Well, H is the 4A transform of little h, of course. All right. So I can calculate this directly. The 4A transform of little h is the sum from I equals zero to three of the values of H HMI omega to the minus I. Maybe I shouldn't use I, maybe I should use K. K, omega minus K. All right. That's the discrete 4A transform of H. All right. Now, H is easy. H is given explicitly. This is the sum from K equals zero to three of K plus one. Right, H of zero is one, H is one of two, three, four, so it's K plus one times omega in the minus K. All right. Write this out. Write this out in terms of vectors. Write this out. But I will do it. Okay. It's a sum of vectors. K plus one times, this is a zero omega to zero's all ones, and omega to minus one, omega minus two, and so on, and here's what you get. You get very easily, very quickly you get this equals, I'll write it out for you, ten minus two plus two I, minus two and minus two minus two I. All right. Just by evaluating the sum, all right, very, very easy, you do it by hand. Okay. That's what you get. And that tells you exactly the agin values. The agin values are exactly minus ten, minus two plus two I. That is the agin values of A. The matrix A are given by minus ten, minus two plus two I minus two and minus two minus two I. Okay. It drops to like a piece of light fruit. And in fact, I wasn't so sure my – well, I was sure of myself, of course, but I decided to check this and if you put this into matrix mathematics, and sure enough, if you ask for the agin values in that matrix this is what you get. All right. So it works like a charm. Once again, how does it work? An LTI system given by convolution, the convolution is also realized by matrix multiplication this is the matrix, therefore, agin vectors of the system are agin vectors of the matrix. But I know the matrix corresponds to an LTI system so the agin vectors are power of complex exponentials one, omega, omega squared, omega tubed, all right. And the agin values are the values of the corresponding transfer function at the value zero, one, two, three, all right.

So all I have to do is find the transfer function to find the agin values of the matrix. And to find the transfer function I calculate the 4A transform discrete function directly from the definition of discrete 4A transform. All right. It's just the component of H times omega to the minus K, add up those vectors, and you get this vector and the entry here happen to be the agin vectors of the system, okay, of the matrix. It's cute. All right. It's pretty cute. All right. We will now leave the theory of linear systems, as much more as there is to do, and there's plenty more to do. I want finish up the course with a discussion of how two dimensional transforms work, so we'll start on that on Wednesday. As always, please sort of read around and read ahead in the section so, again, our pursuit is going to try to make it look as much like one dimensional cases, as possible. And that will mean more to you if you sort of read ahead a little bit and familiarize yourself with how the formulas look so I can jump back and forth more easily. Okay. See you on Wednesday.

[End of Audio]

Duration: 53 minutes

TheFourierTransformAndItsApplications-Lecture26

Instructor (Brad Osgood): Relax, but no, no, no, the TV is on. It's time to hit the road. Time to rock and roll. We're going to now turn to our last topic of the quarter, and that is space, the final frontier, a quick trip to some of the fascination and mysteries of the higher-dimensional Fourier transform. Actually, one of the things that I want to convince you of is that there are no mysteries to the higher-dimensional Fourier transform, or at least not so many because our goal is to try to make the higher-dimensional case look as much as possible like the one-dimensional case. So all that you learned, that hard one knowledge – sort of the same thing we did when we were looking at the Discrete Fourier transform, to try to carry over your intuition from one setting to another setting. So all the stuff you learned, all the formulas you learned, we'll find counterparts in the higher – all the stuff you learned in the one-dimensional case, we'll find counterparts in the higher-dimensional case. This is not just – so here's the topic. Higher-dimensional Fourier transforms. By that, I really mean Fourier transforms as functions of several variables. That's what I mean when I talk about higher-dimensional Fourier transforms, i.e. Fourier transforms of functions of more than one variable. Now this is not an idle generalization by any means. It's not a generalization for generalization's sake. These days, you're as likely to find applications of Fourier analysis, and single processing and so on to functions of more than one variable as you are functions of one variable. At least two or three variables. I think a leading example of this, for instance, is in the Fourier analysis of imaging. There are many applications of a higher-dimensional theory, higher-dimensional Fourier transforms. For example, EG images or the Fourier analysis of images. Spectral analysis. Let me put it that way. Spectral analysis and what is almost the same thing, signal processing for images.

For what is an image, after all? What is a mathematical description of an image? Well, at least not a two-dimensional image. At least mathematically, it's given by a function of two variables, say X1 and X2. Function F of X1, X2, where X1 and X2 are varying over some part of the X1, X2 plane. At each point, what the function prescribes is the intensity. I'm thinking about black and white images here. So you think F of X1 and X2 as a range of numbers from zero to one, from black to white. So you think of F of X1, X2 as the intensity from black to white, say, at the point X1, X2. And then as X1 and X2 vary over region the plane, the intensity varies, and that what makes an image. That's what makes a black and white image. Color's more complicated, but at least for black and white images, that's what you get. Now, you're used to the more Discrete version of this in the digital version where the [inaudible] quantals at 256 levels and so on. But you can think of this first in the continuous case where X1 and X2 were varying continuously over a region. The intensity is varying continuously from black to white. That's what describes an image.

Now, the question is – this you might think of as the spacial description of an image. It's what you see with your senses. It's like the temporal of the time description of sound, what you hear with your ears. So this is the spacial description of an image. It's what you perceive. The question the Fourier analysis raises is, is there another description? Is there a spectral description? Can you describe an image in terms of its component parts like you can describe a sound in terms of its component parts? Is there a spectral description? Can you analyze an image into simpler components, like you can analyze a sound into simpler components. If you can, of course that gives you a certain amount of control over it. If you can manipulate the component parts, you can manipulate the image. In the same way as if you can manipulate the harmonics that go into a sound, you can manipulate the sounds, to pick one example of it. There are other sorts of examples like this, but just keep this one in mind because it's a very natural example. It's a very important example. The answer, of course, is yes, and it's provided, exactly, in this case, by the two-dimensional Fourier transform and in the higher dimensions, similar sorts of problems are provided by the higher-dimensional Fourier transform. So yes, the spectral analysis, spectral description, is provided by the two-dimensional Fourier transform. The component parts are the two-dimensional complex exponentials. It's a very close analogy to what happens in the one-dimensional case, and that's really what I want to push. The components are the 2D complex exponentials. The same things hold in higher dimensions. So again, as we understand these things, and as we go through the formulas, the thing that I want to emphasize is that the two-dimensional case and the higher-dimensional case can be made to look very much like the one-dimensional case. Furthermore, there's no essential difference between the two-dimensional case and the higher-dimensional case.

There are differences in going from one to two dimensions. Some of the formulas are a little bit different. There are new phenomenon that come in, in going from one dimensions to two dimensions. You have two degrees of freedom instead of one degree of freedom. Naturally, some things are going to be richer in the two-dimensional case. That's true. But there's more similarity than there is difference, I'd say, between the one-dimensional case and the two-dimensional case and certainly there is very little difference between the two-dimensional case and the higher-dimensional case. So we want to make the 2D case and higher look like the 1D case. There are differences, and as a matter of fact, by making them look as similar as we can, it'll make the differences also stand out when there are differences. But certainly, there's more similarities than there are differences, and there are very few differences between the two-dimensional case and higher-dimensional cases. So if you get that down, then you're in pretty good shape. I consider this good news. That is, the things we've learned so far will have counterparts and analogues in the higher-dimensional setting. Now, the key to doing this, the key to making it look very similar, is really how you write the formulas, and the way you write the formulas is with vector notation. So the key to doing this, that is, to making the cases look that same, is to use – whenever possible, is to think in terms of vectors. Write the formulas, etc et al, with vectors. There are different ways of motivating this. I'm going to really pretty much jump into it. So let me write down the one-dimensional case, the one-dimensional formula for the Fourier transform and see what has to be replaced and what has to change and again, how I can make the higher-dimensional case look like the one-dimensional case. Here's the one-dimensional Fourier transform as we've written down many times. Fourier transform function is the integral from minus infinity to infinity of either the minus two pi I ST, F of T, DT. Wonderful.

Now, what goes into that? So I have one-dimensional continuous variables as a T, or, so to speak, 1D continuous variables. The function if I take the Fourier transform as a function of that one variable, T, and the Fourier transform is a function also, so that one variable S. F of T is a function of T. The Fourier transform, F of S, is a function of S. Of course, the other thing that happens here, the other big element that enters into the definition of the Fourier transform is the complex exponential. There, it's the product of the variables, S and T, that come in. That's what it takes to define the Fourier transform. Remember, of course, there was the integral and everything else. It's not that many symbols, but it's a pretty complicated expression. In either the minus two pi I ST, it's the product of S and T enters. Now, what about the higher-dimensional case? Let's just do the two-dimensional case. What do I replace the variables by? What do I replace the functions by, and so on? So instead of single functions of a continuous variable, I have a vector function X, a vector variable X, is a function of two variables. That'll be the so-called spacial variable. So you can think of it as just a pair of numbers, X1 and X2, but again, I'm going to write that as a vector in most cases. Then the frequency, I'm going to write with a Greek letter, xi. That's a pretty standard term, and if you haven't learned to write a Greek letter, xi, this is your big chance.

That's a pair of numbers, xi 1, xi 2. That'll be thought of as the frequency variable. The functions that I'm going to be taking the transform of is not a function just of one variable alone, but it's a function of a vector variable, X, or it's a functions of X1 and X2. Function to be transformed is F of X. If I write it just like that in vector notation, or F of X1, X2. The Fourier transform is likewise, going to be a function of the frequency variable, which is the pair, xi 1 and xi 2. The Fourier transform will be something like the Fourier transform of F, I use the same notation of the vector variable, the frequency variable, xi, or if I write it out as a pair, xi 1, xi 2. Now the big [inaudible] how will I actually make the definition? What happens to the complex exponential? How do I replace multiplication of the one-dimensional variables S and T in the one-dimensional case by a higher-dimensional analogue? So what happens to the complex exponential? You replace the product, ST, in the one-dimensional case, by the dot product or the inner product of the two variables that are into the higher-dimensional case. That is the spacial variable and the frequency variable. You replace this by X dot xi, the inner product. So that's X1 times xi 1 plus X2 times xi 2.

So the complex exponential, then, is E to the minus two pi I, X dot xi – or xi dot X. It doesn't matter, it's the same thing – which is E to the minus two pi I, X1 times xi 1 plus X2 times xi 2. That's what replaces the complex exponential. If you write it like this, in vector notation with the inner product, it looks pretty much, as much as you can make it, like the one-dimensional case. Putting all this together, what is the definition of the Fourier transform? Now, realize here, by the way, this is still – I'm computing a scale here. I haven't done any – this is the ordinary exponential function because X dot xi is a number, is a scaler. So I'm taking E to the minus two pi I times a scaler. That's nothing new. The fact that that scaler is arising as the inner product of two vectors, well, okay, that's something new. But nevertheless, I haven't introduced a new function here. I've only replaced the multiplication of the one-dimensional variables, S and T, by the inner product of the two-dimensional variables, X and xi. So with all this, what is the definition of the Fourier transform?

Let me write it in vector form first, and then I'll write it out in terms of components. The Fourier transform of F at the vector variable, xi, is the interval over R2, interval over the entire plane, of E to the minus two pi I, X dot xi, F of X, DX. All right? Looks as much like the one-dimensional case as I can make it look. I've replaced scaler variables, one-dimensional variables, by vector variables, X and xi. For that matter, this same definition holds in – well, let me write it out in components before I say anything about higher-dimensions. In components, this says that the Fourier transform F at xi 1, xi 2, the pair of frequency variables, that integral over R2 becomes a double integral. That's the integral from minus infinity to infinity. The integral from minus infinity to infinity of E to the minus two pi I, X1 times xi 1 plus X2 times xi 2, times F of X1, X2 times DX1, DX2. Same formula, but now everything is written out in terms of components. Now, which would you rather write? This, which took the entire blackboard, or this, which took the entire blackboard, but a little bit less of the entire blackboard? This, you can recognize as the same form as the one-dimensional Fourier transform. This sort of looks the same, but it's a little bit more complicated. You need both. This we're going to need to write things out in terms of components when we actually compute for specific Fourier transforms, but if we want to understand the Fourier transform in higher-dimensions, and we want to work with it conceptually, then it's by far better to write things in vector form.

The formulas, the theorems that govern the higher-dimensional case into the extent that we can make them analogous of the one-dimensional case, and it really is, to a great extent, they are much better understood if we use a vector notation in that sort of form for the Fourier transform. So let me just say that now the higher-dimensional case is exactly an extension of this, from two dimensions to N dimensions. N dimensions, the spacial variable is a function of N variables. The frequency variable also is an N-tuple, xi 1 up through xi N. Again, multiplication's replaced by the inner product. X dot xi is X1 times xi 1 plus – up to XN times xi N. So this time, the function you're taking the transform of is the function of N variables. The Fourier transform is also a function of N variables. But the definition looks the same, and I will not write it out in terms of components. Let me just write out the vector form of it. So the Fourier transform of F at xi, the vector variable, is the integral over RN of same thing, E to the minus two pi I, X dot xi times F of X, DX. Same thing. Same formula, except this time, I'm integrating over all of N-dimensional space instead of just the plane. I will demure from trying to write this out in components, X1 through XN, but you can.

You see, the economy of writing things in the vector notation. It just makes things much simpler to write. Again, I contend, and I hope to show you, that it also makes it much simpler to understand things conceptually, understand how the formulas look conceptually if you stick with the vector case. I should mention that of course one also has the inverse Fourier transform and by analogy, it looks very similar to the right of the Fourier – to the forward Fourier transform, except I changed the plus to a minus rather than a minus to a plus in the complex exponential. So the inverse Fourier transform, let's say it's the inverse Fourier transform of G. This function of the spacial variable is the interval over RN. I'll do it in the N-dimensional case for the thrill of it. E to the plus two pi I, X dot xi, say G of xi D xi, which looks the same as the one-dimensional inverse Fourier transform. Now, I'm not going to – when time comes, I'll talk about Fourier inversion and all the rest of that stuff, but it's going to work out the same way as in the one-dimensional case. There are some differences, again, to be sure, but the basic phenomenon that hold in the one-dimensional case are also going to hold in the two-dimensional case and the higher-dimensional case. It works out just beautifully. I'll say one other thing here. What about the dimensions of the variables involved – dimensions, not units so much. So again, there is a – and we'll see different instances of this. There's a reciprocal relationship between the spacial domain and the frequency domain. In the one-dimensional case, we refer to the time domain and the frequency domain. In the higher-dimensional case, we think more in terms of X as a spacial variable rather than a time variable, naturally, because that's where the problems come from. There is a reciprocal relationship between the spacial domain and the frequency domain. So this time, you're talking about the two domains as a spacial domain and a frequency domain.

What is the first instance of a reciprocal relationship between the spacial domain and the frequency domain? Now, we'll see various instances of this. Again, most of the analogates to the one-dimensional case, but even in the two-dimensional case, what do I mean by this? Well, if I think of the vector variable X as X1 through XN as the spacial variable, then I would imagine the XIs each have dimension length. So if this is a spacial variable, then XI has dimension length, right? So what dimension should the frequency variable have? If you think physically? Mathematically, we never care about dimensions. The people who work in physics, especially in applications, they always like to attach units to their variables as a way of sort of checking the formulas make sense. So to form X dot xi, and have it make sense physically, that's X1 times xi 1 plus X2 times xi 2 and so on, to XN times xi N. To make sense of this physically, you would want the xis to have dimension one over length. So it's length times one over length gives you a number, gives you a scale so you can add them all up. You would want xi I to have dimension one over length. That's the reciprocal relationship between space and frequency. If the Xs have dimension length describing the spacial variable, then in the frequency domain, the variables have dimension one over length. Its analogous to time and frequency, time having the dimension time, or units of seconds. And frequency having units of one over second, or hertz. Okay? Same thing.

So this is sort of the first natural example of the reciprocal relationship between the two domains, again, analogous to what happens in the one-dimensional case. Cool? I think so. I like this stuff. Now, let me talk about – this is the formula. Again, I presented this to you as directly analogous, as close as I can make it, to the one-dimensional case. Of course, there are ways of – you can get to the higher-dimensional Fourier transform in a similar way as you can get to the one-dimensional Fourier transform by considering higher-dimensional Fourier series and sort of limiting cases of higher-dimensional Fourier series and so on. I'm not going to do that. There's a lot of water under the bridge between where we started and where we are now, so I'm not going to revisit all those ideas. I'm not going to talk about – there is a section in the notes on a nice application, I think, of higher-dimensional Fourier series, but it's not something I want to talk about in public. So we're going to really go pretty much directly for the main ideas and the main applications of this thing without unfortunately somehow taking some of the many fascinating little side routes that you could go on. Again, there are ways of motivating the definition of a higher-dimensional Fourier transform. That is, there are ways of motivating why you replace multiplication of the one-dimensional variables, S and T, with inner product. Why that's a natural thing to do rather than just my presenting it as my [inaudible], that this is how we're going to do it.

Now, let me write down the formula again in vector form. I want to say something more about – again, as I said, if you consider that what we want to do is try to develop the spectral picture of higher-dimensional signals, then the aspects of that are defined [inaudible] Fourier transform, and understanding what the components are. The components, in this case, are these higher-dimensional complex exponentials, so I would like to spend a little bit of time making sure we understand and helping to get a feeling for what the higher-dimensional complex exponentials are. How you can understand then geometrically, how you can – if you can't quite visualize them, at least how you can put them in your head in some reasonable way other than just the formula. So again, let me write down the formula for the Fourier transform. Let me just do the two-dimensional case. So in the two-dimensional cases, the integral over the plane, E to the minus two pi I, X dot xi, F of X, VX. What I want to talk a little bit about is the nature of the 2D complex exponential. E to the plus or minus two pi I, X – see, I put plus or minus in there because it's a minus sign that comes in for the Fourier transform, it's a plus sign that comes in for the inversed Fourier transform. It doesn't matter. They have the same nature, whether or not you consider plus or minus two pi I times X dot xi. Okay. Now you can't draw – they're complex functions, so you can't really draw a picture of them. You can't draw a graph of these things, but there are ways of understanding it geometrically. So again, let's think of –

You can't draw a graph like you can draw a graph of sines and cosines, but you can understand it geometrically. In particular, you can understand frequency almost in terms of – you get a very intuitive sense of tactiles sense of frequency, what it means to have high frequency, what it means to have low frequency and so on. You can get an understanding, you can get a feeling for this sort of vector frequency. Now, let's go back for a second to the one-dimensional case. I'm just going to set the stage for this because again, I want to pursue by analogy here. So let's go back to the 1D case where I'm looking at either the – I'll look at the plus sign, okay? Just so I don't have to write the minus sign in there. So E to the two pi I ST where S is the frequency variable and T is the time variable. So imagine – fix the frequency. Fix S. Fix the frequency S, and look at this as a function of T. Look at E to the two pi I ST as a function of T. Then it is periodic of period one over S. As a function of T, it's periodic of period one over S. That's E to the two pi I, S times T plus one over S is E to the two pi I ST times E to the two pi I, S times one over S. E to the two pi I, E to the two pi I. It's just E to the two pi I ST. S is fixed here, two, three, five, twelve, whatever.

So that already tells you something about how rapidly – can't draw a picture here, but you can say the words, how rapidly is this complex exponential oscillating depending on S?

Well, if S is large, then one over S is small. The period is small, so it's oscillating fast. It's returning to the same value over and over again very quickly. If S is small, one over S is large. The period is large. So that gives you a sense of how fast the function's oscillating, how fast the function E to the two pi I ST is oscillating as a function of T, depending on the size of S. Or another way of looking at this is E to the two pi I ST will be equal to one at the point – well, T equals zero. T equals one over S or minus one over S, so plus or minus one over S. T equals plus or minus two over S, and so on and so on. So although it's a complex function and you can't draw the graph, you can say, well, it's returning to the value one at point space one over S apart. All right? So E to the two pi S is equal to one, E to the two pi I ST is equal to one at points spaced one over S apart. So again, if S is large, if you have a high frequency, then those spacings are very small. If S is small, you have a low frequency, then one over S is large. The points are spaced far apart. So it gives it a tactile description of how fast the function is oscillating, how often it assumes the value one. This is all in the one-dimensional case.

I don't know if I ever really spoke in these terms when we spoke about the complex exponentials, but should not be too much for you to get your head around. We worked with one-dimensional complex exponentials a lot. If I didn't say this before, maybe I should have. Now what about the two-dimensional case? The 2D case, where my complex exponentials will form E to the two pi I X dot xi, or if I write it out in components, that's E to the two pi I, X1 times xi 1 plus X2 times xi 2. Okay? So again, I'm going to fix a frequency and look at this as a function of X1 and X2. So I want to fix xi. Xi 1, xi 2, a given frequency. I'm not going to talk about high frequencies or low frequencies yet, but I will. But just imagine fixing a frequency, and look at [inaudible] complex exponential as a function of X1 and X2. That is, it's a function on the X1, X2 plane. Look at E to the two pi I – I'll write it out in components – xi 1 plus X2 times xi 2 as a function of X1 and X2. Now, by analogy to what I did in the one-dimensional case, when does a complex exponential equal to one? When does this function of a function of – xi 1 and xi 2 are fixed. When does this function assume the value one? Where is E to the two pi I, X1 times xi 1 plus X2 times xi 2 equal to one? Well, don't get too far ahead of things here. That will be true for points X1 and X2. So again, xi 1 and xi 2 is fixed. I'm asking this as a function of X1 and X2.

This will be so at points X1, X2 where the inner product, X1 times xi 1 plus X2 times xi 2 equals an integer, N. So N can be zero plus or minus one, plus or minus two and so on. At those points, in the X1, X2 plane, this complex exponential will be equal to one. That's just the property of the ordinary exponential. See, we've [inaudible] E to the two pi I, N. Now, what are these points? X1, X2. It's not hard to see. Let me show you how this goes. It's very pretty. You get a very pretty picture here. Again, this is a picture – if you've studies imaging already, you have probably seen this picture. I don't know how it's presented in those courses, but let's present it here. I'm going to push it here as an analogue of the one-dimensional case. Let the case then equal zero. What about the points X1, xi 1 plus X2, xi 2 equals zero? What are those points? Again, xi 1, xi 2 are fixed, like five times X1 plus three times X2 equals zero. Three times X1 plus five times X2 equals zero, minus X1 plus four times X2 equals zero and so on. So again, if xi 1 and xi 2 are fixed, what does this set of points look like in the complex plane? Well, you have to remember a little bit of analytic geometry here. You need to remember analytic – and let me just say, analytic vector geometry. That is, the set of points in the X1, X2 plane where X1 plus xi 1 plus X2 times xi 2 is equal to zero is a line through the origin. So here's the X1, X2 plane. If I wrote it like this, three X1 plus five X2 is equal to zero or minus X1 plus four X2 equals zero, you'd probably recognize – and if I said, what is that figure, you would probably say it's a line through the origin.

That's fine. But you also have to remember the relationship of xi 1 and xi 2 to that line. What is the relationship? It's a line through the origin with the vector, xi, as the normal vector. X1 times xi 1 plus X2 times xi 2 equals zero is a line through the origin in the X1, X2 plane with xi as the normal vector. Not the unit normal vector. It doesn't add length to one. I'm not assuming that, but it's perpendicular to the line. It's normal to the line. So going back to what we're describing here, for example, E to the two pi I, X1 times xi 1 plus X2 times xi 2 is equal to one along that line because along this line, X1 times xi 1 plus X2 times xi 2 is equal to zero. So it's E to the zero is one. Now, what about all the other places where it's equal to one? That is, X1 times xi 1 plus X2 times xi 2 is equal to N for N equals zero, plus or minus one, plus or minus two and so on. What are those configurations in the X1, X2 plane? So now what is a description of all of the points where X1, xi one plus X2 times xi 2 is equal to N? N equals zero, plus or minus one, plus or minus two and so on. Well, each one of these things – let me do just one more example. X1 times xi 1 plus X2 times xi 2 is equal to one is another line with the same normal vector, xi, but not through the origin.

As a matter of fact, the picture would look something like this. Here's the X1, X2 plane. Here's the line, X1. Let me write it in vector notation. X dot xi equals zero, and here is the line, X dot xi equals one. It's parallel to this line because it has the same normal vector, but it doesn't pass through the origin. This is xi, and this is xi. How far apart are these lines? What is the spacing? How far apart are these lines? So the two lines, this one and this one. There's a point here. How do I find the distance between the two lines? Well, I'm going to find it using vectors. I take a point, X1, X2 on that line, and that point satisfies X1 times xi 1 plus X2 times xi 2 is equal to one. That's the defining property of the line. Call this angle theta. Then what I want is I want the – let me call this vector X. Then the distance between the two lines is the magnitude of X times the cosine of theta for any point, X1, X2, on the line. But what is the magnitude of X times the cosine of theta in terms of inner products? Remember, X dot xi is equal to the magnitude of X times the magnitude of xi times cosine of theta. That basic geometric property of dot products. That is to say the magnitude of X times cosine of theta is X dot xi divided by the magnitude of xi. But what is X dot xi? One. Anywhere along this line, X dot xi is equal to one. So it is equal to one over the magnitude of xi. How far apart are the lines spaced? The lines are spaced one over the magnitude of xi apart.

So again, I'm going to draw the picture. Here's the line, X dot xi equals zero. Here's the line X dot xi equals one. How far apart are they? The spacing, distance, is equal to the reciprocal of the length of xi. Now, same thing holds for all those other lines. Which other lines? I mean, when the inner product is equal to an integer. Zero plus or minus one, plus or minus two and so on and so on. That gives you a family of parallel, evenly-spaced lines in the plane. So X dot xi equals N, for N equals zero plus or minus one, plus or minus two and so on, gives you a family of parallel – they all have normal vector xi – lines in the plane. Parallel, evenly-spaced lines. What is the spacing? The spacing is one over xi. So the picture is something like this – I can only draw so many of these things. Here's the line through the origin. Here's the line corresponding, say, N equals one. Here's the line corresponding to N equals two. Here's the line corresponding to N equals minus one, minus two, whatever. Here's X dot xi equals two. Here's X dot xi equals one. Here's X dot xi equals zero and so on.

They're all parallel. They all have the same normal vector, and the spacing is the same. The spacing between any two lines, any two adjacent lines, is one over xi apart. Now, I'll remind you – let's go back to the complex exponential. So that says E to the two pi I, X dot xi, is equal to one along each one of these lines. So it's equal to one here. Then it oscillates, and it's equal to one, equal to one, equal to one and so on. It's a complex function. You can't draw the picture, but you can say that it's somehow – it's oscillating up and down. It's not really the right thing to say because it's complex. You can't talk about up and down with complex, but it's oscillating, and it's equal to one. Then it oscillates. Then it's one, one, one and so on. As a matter of fact, to be a little more precise here – I'll let you sort this out. It's in the notes – you can say that E to the two pi I X dot xi is periodic in the direction xi with period one over the magnitude of xi. I will leave it to you. I think this is discussed in the notes, but I will leave it to you to read that or make that precise, but that's the closest analogy you can have in the two-dimensional case to the one-dimensional case. If I just imagine myself in the X1, X2 plane, going off in a certain direction, and if I go off in that direction, this function's going to be periodic, and it's going to periodic of period one over xi. It oscillates. It goes up to one, down. Again, I can't resist saying up and down, but the idea is that is oscillates and it returns to the value on along each one of these lines. If you look back, this is not a bad analogy. It's a pretty close analogy to the way we look at the complex exponential, for the way you can visualize or imagine the complex exponential in the one-dimensional case.

In particular, it gives you a sense of what it means to have a high frequency and low frequency harmonic. So if you think of these complex exponentials as a two-dimensional harmonic, you get a sense of what it means to have a high-frequency harmonic or a low-frequency harmonic. A high-frequency harmonic would be if the magnitude of xi is large. High frequency means the magnitude of xi is large. That means one over xi is small. That means these lines are spaced close together. That means there's rapid oscillation. So close spacing of the lines, rapid oscillation of the complex exponential. And low frequency would meant the magnitude of xi is small. That means that one over the magnitude xi is large. That means the line spacing is large, and I have slow oscillation. So far spacing of the lines, and I have a slow oscillation of the complex exponential. Now, again, you have to be a little more careful. It's richer in two dimensions than it is in one dimension because harmonics don't just have a magnitude associated with it, don't just have a – you can't just say how fast or how slow it's oscillating. You also have the specify the direction. The frequency is a vector quantity in two dimensions and in higher dimensions, not a scaler quantity. So you have to talk about it's oscillating slowly in a given direction. It'll be different depending on the frequency, right? If you change the frequency, you're changing – you might change both its magnitude and its direction. So you can change those two independently. It might be oscillating rapidly in one direction and slowing in another direction. So it's a richer picture.

We'll talk about this, and there are pictures of it, also, that are given in the notes. But this is the geometric interpretation. Again, you can't – I hesitate to say this is how you visualize complex exponentials – vector, two-dimensional complex exponentials because unless you're really visualizing them because they're complex functions. But again, you are seeing from this how they're oscillating, in what ways they're oscillating, in what way they generalize the one-dimensional complex exponential. In this case, you have lines where the functions are equal to one. In some areas, in optics, for example, usually refer to these lines as lines of constant phase. They complex exponential's real on those lines, so we sometimes refer to that as constant phase. It varies. I think the terminology varies. I don't necessarily want to attach any one particular interpretation or any one particular terminology to it. I think the picture itself holds, regardless of the interpretation and regardless of the setting, and that's what you should think of. Okay? All right. So we have now done – we have the basic formula for the Fourier transform, the components to go into it, that is the particular elements that are added into it, and some understanding, I hope, of the components into which a higher-dimensional signal is broken. That is, these complex exponentials and what they represent. Next time, we will get [inaudible] higher-dimensional Fourier transform, and you will see how the theorems in the one-dimensional case carry over to the higher-dimensional case, where they're similar and where they're different. All that is waiting for us. See you then.

[End of Audio]

Duration: 54 minutes

TheFourierTransformAndItsApplications-Lecture27

Instructor (Brad Osgood):Hello. I hate it when I do that. All right. Here's some information on the final exam. Some of this you know. Some of this you don't know. What you do know or what you should know is that it's on Thursday the 13th from 8:30 to 11:30 in the morning. The new information is the location for it. The Dinklespiel Auditorium. I'm not sure if I spelled Dinklespiel correctly, but that's where it's going to be. I'll post that to the website, and I'll announce it many times.

It'll be open notes, open book as before. I will post last year's final exam. Maybe I'll do that today, actually, so you get a chance to look at that. I will also provide you with a formula sheet. Same one as we had for the midterm. It was a comprehensive formula sheet. You can also print that off, of course. We'll talk more about the exam and the mechanic of it in due course, but we already have a location established for it. Any questions or comments on that?

All right. So today, we're going to continue our discussion of the higher dimensions. We're going to continue our trip in the outer reaches and talk about getting to know your higher-dimensional Fourier transform. The main point that I started to make last time, and I want to repeat, is that you already know your higher-dimensional Fourier transform because you know your one-dimensional Fourier transform so well.

The point [inaudible] getting to know your higher-dimensional Fourier transform. As I said, you know it. You have to convince yourself of that. There are differences, and I'll highlight some of them a little bit today but even more so next time. Again, the point of using the notation that we've used and the approach that we're using is to make the higher-dimensional transform look as much like the one-dimensional transform as possible.

The formula's carried over. We'll start to see some of those today. The ideas carry over. There's no reason to think of it as a completely different subject. Let me remember the definition here. I'm going to remind you of the definition and the elements that earn interest. So you have a spacial variable and a frequency variable and then a function and a transform from one domain to the other domain. So I usually write the spacial variable like this.

We have the frequency variable. I call it xi one through xi N. Same dimension. We have a scaler value function which could be real or complex. But it's not vector value, so it's a real value function or a complex value function. Its Fourier transform is a function of the frequency variable. It's defined by this integral, which we tried to make look as much like the one-dimensional case as possible. You integrate all [inaudible] space, E to the two pi I, minus two pi I, X dot xi, F of X, DX. I'm even making it look as much like the one-dimensional case as possible. I still couldn't fit it all on one line.

Let me write it over here again. The Fourier transform of F at xi – I'm just going to rewrite it – is the integral over all space, E to the minus two pi I, X dot xi, F of X, DX.

The new future here is the dot product. You take the dot product of the two vectors, X and xi. That gives you a scaler. You take the complex exponential and apply it to that. Writing it this way makes it look as much like the one-dimensional case as you can make it look.

We'll even see that the formulas look similar, and even the approach to proving the similars is also very similar. Now, I do want to follow a similar path to what we did in the one-dimensional case. That is, looking at specific transforms and specific functions. I'll do less of that, and then also looking at general formulas, versions of the shift theorem and the stretch theorem and so on. I'll do many not quite as much of that as I did in the one-dimensional case where we saw it in more detail. There, actually, it's in the general formulas that you start to see a few differences that start to crop up between the one-dimensional case and the higher-dimensional cases.

This is where some of the interest is, but let's do a couple of simple examples first, of computations. There's an important class of functions for which you can computer a higher-dimensional transform by computing a series of lower-dimensional transforms. There's a class of examples for which you can compute a higher-dimensional transform via one-dimensional transforms. Maybe I should just say lower-dimensional transforms. Typically one-dimensional transforms.

What I'm taking here are what are sometimes called separable functions. I don't think there's a uniform term to describe them, but let me give you and example here. These are separable functions. They come up often enough that there's enough example of them. They come up often enough in applications that it's worthwhile just calling to attention to them.

The easiest and important example is like the two-dimensional – let's just do this in two-dimensions, but you can do the same thing in higher-dimensions. Let's work in 2D. One of the most basic examples and important examples is the two-dimensional rec function. So I'll call that pi of the rectangle function of X1, X2. The graph would look like the two-dimensional version of the graph of a one-dimensional rectangle function. That is to say – let's see if I can draw a picture of this thing here. So here's the X1, X2 plane. It's a box. It's a box. This is one of those IQ tests. Can you draw this box. The answer obviously is no.

This is a function which is one over a rectangular region in the plane. Very good. Perfect. That's fine, Mr. Osgood. Now try doing this basket for us. It's supposed to be equal to one over this box in the plane. It's equal to zero outside the box. Let me try to draw perspective here.

So here's the X1, X2 plane. Here's the box that goes from minus one-half to plus one-half in the two different directions. The two-dimensional rectangle function is one over that box and zero outside the box. You can decide, depending on your religion, what you want to do on the boundary. It's equal to one if X1 in absolute value is less than one half, and X2 is less than one half. It's equal to zero otherwise.

That's fine. You could write down the formula and definition that way, and that generalizes the one-dimensional rectangle function. That's a really pathetic drawing, isn't it? What's maybe not so obvious is that the two-dimensional rectangle function can be written as a product of one-dimensional functions. That's the point I want to make.

So you can write the two-dimensional rectangle function as a product of two one-dimensional rectangle functions. In fact, it's easy to write, and I will let you convince yourself of this, that it's just the rectangle function of the one variable, X1, times the corresponding ordinary rectangle function of the variable X2. So this is one between X1. If X1 is between a half and a half and zero otherwise, and this function is one if X2 is between minus one half and a half and zero otherwise. If you take the product of those two functions, that describes exactly this condition. It's equal to one if both variables, X1 and X2, are between minus a half and a half. It's equal to zero if either one is outside that interval. If either X1 or X2 is outside the interval from minus one half to one-half.

Not all functions can be decomposed this way, but when you can decompose a function this way, then it's easy to compute the Fourier transform. It's something to watch for. Again, it happens often enough, and it's worth calling attention to it as an important special case because now you have a Fourier transform of F at xi one, xi two. It's equal to by definition – so I'll write it out in coordinates – the interval from minus infinity to infinity. E to the minus two pi I, X1, xi one plus X2, xi 2 times pi, X1, X2, DX1 and DX2. That's just the definition of the Fourier transform. It take the entire, damn board.

That's only two dimensions. Now you can split the variables up. You can write this as – let me combine two steps here. Complex exponential, the exponential function being what it is, you can split up that inner product in the exponent to be minus two pi I, X1 times xi one times E to the minus two pi I, X2 times xi two. Then, as I just pointed out, a two-dimensional rectangle function can be written as a product of two one-dimensional rectangle functions. DX1, DX2. Now you sort of look ahead a little bit and see what happens. Things decouple.

The X1s and X2s are independent. That is, you can write this as a product of two integrals which are, then, just one-dimensional Fourier transforms. So this is equal to the integral from minus infinity to infinity, E to the minus two pi I, X1 times xi 1, times pi of X1, DX1 times the integral from minus infinity to infinity, E to the minus two pi I, X2 times xi two, pi of X2, DX2. We've done this many times going the other way. We've done this many times in a variety of circumstances, taking the product of two integrals and converting them to a single double integral here. I'm going the other way around.

I'm taking the double integral and because the variables are decoupled, I can write that as a product of two single integrals. Now you just [inaudible]. This is the Fourier transform of the rectangle function, but you have to pay attention to what the variables are named. That is, it's the product of – so what do you get for the first one? You get the sync function of xi one because I'm taking the Fourier transform with respect to X1 and the dual variable here is called xi one. Then the second integral, it's a Fourier transform of the rectangle function again.

But the dual variable, the variable I'm integrating, is X2, and dual variable in the frequency domain is xi two. So it's just the product of [inaudible] functions. Sync xi one times sync xi two. Are there other tongue twisters we can try?

You can look at the notes for a picture of that. You get a nice picture of the product of two sync functions. We were able to do this because we were able to rely on our one-dimensional knowledge to get a two-dimensional result because the function was separable. That happens often enough that, as I say, it's worth pointing out. In general, what is the situation?

In general, if you can write a function of N variables, and you can't always do this. I want to emphasize this. You can't always do this, but in some cases, you can. If you can write it as a product of functions of one variable or functions of a lower dimension, say F1 times X1 times F2 times X2 for some functions. I'm not saying there's a procedure for doing this, and I'm not saying it's always possible. It's not always possible, and there's not always procedure to do it. You can.

Then the Fourier transform of the function of several variables can be written as the product of the Fourier transforms of the functions of one variable. So if you can do that, then you can write – and I won't derive this – the Fourier transform of the big function at the big frequency variable is the product of the Fourier transforms of the little functions at the single frequency variables, Fourier transform of F1, Fourier transform of F2, xi two times xi N.

There are various combinations of this. Maybe you can write it as a function of one variable times a function of two variables, if F is a function of three variables, say. X1, X2 and X3, maybe you can decompose it as the product of a function of two variables times the product of a function of one variable or whatever. I'm not saying that there's any blanket thing that's going to work or that you might expect to happen, but it is the sort of thing to look for.

That's what I mean by separable functions. Let me give you another example, actually, that will lead to another important class of functions. We'll handle it one way, and we can also handle it another way. So another example – really, I'm just going to do most of these examples in two dimensions because I don't have the patience or the strength to write anything out in higher dimensions. Another example is the 2D Gaussian. The 2D Gaussian, I call it G. Say G of X1, X2 would be E to the minus pi, X1 squared plus X2 squared.

Again, if you drew the graph of that as a surface in space, and I think the graph is drawn, you'd see a two-dimensional bell-shaped curve or surface. Now, that's also a separable function because – in fact, it's a product of two one-dimensional Gaussians just because of the property of the exponential function. Exponentials being what they are, you can write this as E to the minus pi times X1 squared times E to the minus pi times X2 squared. So it's a product of two one-dimensional Gaussians. We'll call it G1 of X2. G2 of X2, although it's really the same function in both cases where G1 is E to the minus pi one X squared. G2 is E to the minus pi two X squared.

So the Fourier transform of the two-dimensional function is the product of the Fourier transform of a one-dimensional function. So the Fourier transform of G and xi one, xi two is the Fourier transform of G1 at xi one. The Fourier transform of G two at xi two. We know what happens if I take the Fourier transform of a Gaussian. You get it back. That's the remarkable property of the Gaussian. That's equal to E to the minus pi xi one times E to the minus pi xi two squared.

That's fine. Now, actually, the nice thing is – so that uses a separability. Now you can recombine these. You can leave it like that. That's a perfectly fine answer, but a finer answer is to recombine. That is, write this as E to the minus pi xi one squared plus xi two squared. Again, exponentials being what they are, and you're back to where you started. That is the Fourier transform of two-dimensional Gaussian is another two-dimensional Gaussian.

So F of G, xi one, xi two, is equal to E to the minus [inaudible] itself. Xi one squared plus xi two squared. It's nice.

So our one-dimensional knowledge and expertise, in this case, lends itself to understanding a higher-dimensional phenomenon. You could do the same thing in higher dimensions. You could have a higher-dimensional Gaussian. You should be able to imagine instances where you want to consider higher-dimensional Gaussians. If you think of a Gaussian, the bell-shaped curve describing a set of measurements, then instead of taking one measurement, you're taking many measurements. They can be distributed normally, so distributed according to a higher-dimensional Gaussian. That's okay. You want to know the higher-dimensional Fourier transform of a higher-dimensional Gaussian.

Of course, you have higher-dimensional versions of the central limit theorem and all that stuff. It all carries though, and it all carries through pretty much the same because you can make the higher-dimensional case look like the one-dimensional case. You're not seeing any new phenomenon. They're the same sort of things that come up at higher dimensions have already come up at one dimension. That's the power of this approach.

Now, one reason why I wanted to do the Gaussian is because not only is the gaussian separable, the gaussian actually has another property that is an illustration of another important class of functions that come up a lot in applications. So consequently, there's a lot of special things you can say about them as far as the Fourier transform goes. Namely, the gaussian is radial. So the gaussian is also an example of a radial function.

Now by that, I mean if you introduce polar coordinates, there's some circular symmetry to the gaussian. So if you introduce polar coordinates, that is R equals the square root of X squared plus Y squared, or R squared is equal to X squared plus Y squared. Theta is equal to R10 of Y over X. Going the other way, X equals – let me use X1 and X2. Let me be consistent here.

X1 squared plus X2 squared – sorry, I'll get out of the way in just a second. X2 divided by X1 – let me write it over here. So I'm introducing polar coordinates in the X1, X2 plane. So X1 is equal to R, cosine of theta. X2 is equal to R sine theta. Then of course, the gaussian is just E to the minus pi R squared. It depends only on R, the radial distance from the origin.

E to the minus pi, X1 squared plus X2 squared is E to the minus pi R squared. That's what you mean when you say it's a radial function. If it depends only on R. This depends on R squared, but this depends on the distance from the origin, not X1 and X2, independently, so to speak, but on X1 squared plus X2 squared.

So it depends only on R, not on X1 and X2 independently, if you know what I mean. I'm sort of caught in a language here, independently. So to put quotes around that because from what I've said before, it's separable if it depends on X1 and X2 separately but not independently. Okay? Sorry for the clash of language. I don't know how else to say that.

So this is something that comes up really in two dimensions or in higher-dimensions that wouldn't come up in one dimension. In one dimension, you have only one variable. One variable's describing the position or the time or whatever. In space, it's a little bit different. You have a little bit more freedom. You have two degrees of freedom instead of one degree of freedom. So it makes sense to consider the class of functions that depend only on, say, the distance from the origin. Those are the radial functions. You don't see that in one dimension.

So radial functions, in general, depend on the distance R from some origin. You can fix the origin. You see many examples like this. They come up all the time in applications. The important thing is, as far as Fourier transforms are concerned, is that the Fourier transform of a radial function is also a radial function. That's what I wanted to show to you.

Let me write it over here. So I want to see – this is actually a little tricky. The Fourier transform of a radial function is also a radial function. Now, this is a little tricky. There are actually a lot of interesting things to say around these ideas. I don't think we're going to be able to get into this, but let me make a comment here. We're going to get a somewhat complicated formula for the Fourier transform of a radial function. So we'll get a formula for this. We'll convert the usual definition of a Fourier transform into a radial Fourier transform. They have another name. We call it Henkel transform.

In many applications, radial functions are a more natural way of describing the phenomenon. The problem is that the Fourier transform as its given, the definition of the Fourier transform, is a very cartesian thing. I'll say this more precisely when I actually write down the formula for it, but I just want to set the stage. I want to mention something we may not have a chance to talk about, but you should be aware of it.

So we're going to get a formula for this. It's going to be a bit more complicated. This causes all sorts of problems in applications because the Fourier transform itself is set up in sort of a cartesian form. It's sort of the definition involves an inner product, and the inner product somehow looks very cartesian. Consequently, the Discrete Fourier transform and the algorithms for evaluating the Discrete Fourier transform are also very cartesian in nature.

To implement a Discrete Fourier transform and the algorithms that go with it for two dimensions and higher dimensions, it requires that you put down a cartesian grid. In many cases, the functions your interested in don't fit naturally on our cartesian grid. They fit much more naturally on a polar grid. You don't have analogous formulas for the Discrete Fourier transform and for the FFT algorithm for polar grids, and it's a problem.

Where is it a problem? One big area where it's a problem is medical imaging where the way you take samples and the way you actually work with your functions is much more naturally described in polar coordinates than it is in cartesian coordinates. You want to do a lot of serious computation, but it's not set up for it. It's not set up for it because the formulas somehow, on the one hand, the [inaudible] algorithms, the efficient algorithms are all set up in the cartesian world for rectangular coordinates.

The phenomenon that you want to compute are much more naturally described in polar coordinates. That's a problem. It's a hack job to get around it, actually. I may not know the state of the art, but I believe that it has to be gotten around but not really dealt with, theoretically. There's no way of doing it other than hacking, to tell you the truth.

So now let me show you what the issue is here. So what I want to do – the only thing I can do is convert the formula for the Fourier transform into polar coordinates. So the formula for that, and I'm just going to do that in two dimensions. The formula for the Fourier transform is the interval from minus infinity to infinity, E to the minus two pi I, X1 times xi one plus X2 times xi 2. F of xi one, xi two. Sorry. X1, X2. I'll get it. DX one, DX two. So all I want to do is convert this to polar coordinates. That's all I want to do, but that's where complications come in. Some of this is straight forward.

We're assuming that F is a radial function, so I want to convert to polar coordinates in both planes, actually. Both the X1, X2 plane the xi one, xi two plane. So let me write X1 is equal to R cosine theta. X2 is equal to R sine theta for polar coordinates in the X1, X2 plane. But I also have the frequency plane, so let me introduce polar coordinate there. I got to call it something else, so I'll call it row, cosine of phi, and xi two is going to be row sine of phi. Same interpretation of polar coordinates, but instead of calling it R and theta, I'll call it row and phi.

As we will see, the Fourier transform of F, if F is a radial function meaning that it's a function of R, the Fourier transform of F will be a function of row. That's what we're going to see. So we'll see that if F is radial, a function of R, then the Fourier transform of F is a function of row. That's what it means to say that the Fourier transform of a radial function is again, a radial function. That's what it means.

So I'll just do this. So part of the conversion is easy. That is, F of X1, X2 is converted into F of R. So I'm assuming it's a radial function. So it needs the function of R, whatever the formula is. I know what happens to the area element. With DX1, DX2, the area element gets converted into RD, RD theta. That's the area element and polar coordinates in the X1, X2 plane.

It gets more complicated in converting the inner product. That's when I say it's very cartesian. That's what I mean when I say it's very cartesian. You have to convert the inner product. What is the inner product? It's X1, xi one, plus X2 times xi two. So I just write this out. It's R cosine theta times row cosine phi plus – that's X1 times xi one times X2 times xi two. It gives me R sine theta times row sine phi. So the R and the row factor out of both terms. That's R times row, cosine theta, cosine phi. Plus sine theta, sine phi.

If you remember your addition formulas for the cosine, and I'm sure we all do, that is R row times cosine of phi minus theta. So this is R row cosine of theta minus phi. What happens to the Fourier transform? I said it's complicated, the complications are about to come up. It's straight forward just to plug formulas in and convert mechanically from cartesian coordinates to polar coordinates. So I get the Fourier transform of F becomes what?

The Fourier transform becomes the integral – I only put the [inaudible] of integration in. So I'm going to have to describe the whole plane. So to describe the whole plane, theta's going from zero to two pi, and zero is going from zero to infinity. I'm going to integrate RD, theta DR. It doesn't matter. So what happens to the exponential?

It becomes E to the minus two pi I, R row, cosine of theta minus phi, times F of R, times R, D theta, DR. Do I have it all? I have it all. That's the formula for the Fourier transform. That's what you get mechanically converting the two-dimensional cartesian Fourier transform into polar coordinates. Now, you have to work with this a little bit more.

So there's something important to notice here. It's discussed in the notes, but I'm not going to prove it for you. Here's what you find. You find there's a fact that the integral from zero to two pi – I want to make sure I stay with my notes here. So there's an R integral and there's a theta integral. The only part the depends on – the F doesn't depend on theta. That's what it means to be radial, so it's this part here that'd depending on theta, but it doesn't just depend on theta alone. It also depends on phi, the angle in the other plane, the angle in the frequency plane.

But the fact is that this integral, E to the minus two pi I, R row, cosine of theta minus phi, D theta, actually is independent of phi. That's a fact that you can check if you look at the notes. I'm not going to verify that. Actually, it's not hard to verify, but it's one of these things you have to look at and know. If you didn't know that, your formula would become – you wouldn't be able to make any progress, as we're about to.

So to say that's independent of phi means that in fact, that integral is the same as if phi were not there. So in fact, the integral fro zero to two pi of E to the minus two pi I, R row, cosine of theta minus phi, D theta is equal to the integral from zero to two pi, E to the minus two pi I, R row, cosine of theta, D theta. I just leave the phi out. That's just grand. You see there's no F in there or anything. In some senses, it's just an ordinary integral. It's some function of theta integrated with respect to theta. There's no more phi in there.

The problem is, this is not an elementary integral. There's no formula for anti-derivatives that's going to allow you to evaluate this integral in closed form. You don't have a closed-form expression for this. So what do you do if you're a mathematical jerk at a point like this when you don't have a closed-form expression? You define your way out of trouble. You say, well, that just defines a new function. That's what that does. In fact, that's what's done. It's what pisses engineers off.

So define – and I'll give you the general definition. Let me write it down here. I have to copy it from my notes because I don't know it off the top of my head by any means. One over two pi. This is the integral that comes up – let me write it down, and then I'll say something about it.

E to the minus IA, cosine of theta, D theta. A is a real number, here. You can see that what I have here is a version of that. I have two pi R row and so on. This is called – then the integral that I have, so we have – what do we have? We have two pi times [inaudible] of two pi R times row. That's what we're dealing with here, with this integral. J-not is called the zeroth order bessel function of the first kind.

Be very afraid. There are several terms that ought to make you afraid. Zeroth order. That would suggest that there's more than one order, like first, second, third, fourth, and first kind would suggest that there's more than one kind. All that's true. There are bessel functions of many orders and many kinds. They always come up – bessel functions always come up in the context of radial functions. Any time you're dealing with radial – and be aware of this. If you're dealing with problems that have circular symmetry, chances are you're going to come across bessel functions. The other reason for that is often there are partial differential equations involved that are describing the phenomenon.

Often, one does a separation of variables that assumes the solution is a product of an angular function and a radial function. Very often, the radial function or the angular function has solutions which are bessel functions. Equations are of a certain type, and the solutions are bessel functions.

So if you were a British student in the early part of the 20th century, you would learn all about bessel functions of eery order and every kind because that's what all British students had to learn for the competitive exams they had to take. There is a famous book by G.N. Watson called, A Treatise on the Bessel Function, which runs something like 800 pages or something like that.

The publisher wanted Watson to do a second addition of that book, but he didn't have it in him to do a second edition. There was a second printing, and I think the second printing says something like, since the publication of the first edition, my interest in the bessel function has waned. One might expect, right? But what can I say? These are special functions of mathematical physics, as they're called, they just come in the applications. Now they're all hard wired into mathematics and hard wired into Matlab and things like that so that you can draw graphs of these things. You can compute with them and so on. You don't have to have, somehow, the same kind of intimate knowledge that was the terror of British students' lives in the early part of the 20th century, but they still do come up.

So now, what is the consequence of this for us? So now we conclude the Fourier transform of F of R, a radial function, is – I will call it. And we see that it's going to be a function only of row, not of phi because the phi is gone. F of row is the integral – I'm going to write it down here. I have to be careful. Two pi times the integral from zero to infinity of F of R times [inaudible] of two pi R row, DR. No phi. The phi is not in there because of that fact.

This integral is independent of phi. I skated over that, but that is sort of the whole ball game. That says the Fourier transform of a radial function is also radial. I didn't do anything here except substitute the formula – substitute the definition of J-not for what our inner integral was when I was computing the Fourier transform. I converted the Fourier transform into polar coordinates.

This, too, has a name. Question?

Student:[Inaudible] for the DR? Instructor:

Oh, thank you. RDR. Thank you. Polar area. RDR. Now, in fact, this has a name. They call the zeroth order Henkel transformation.

So again, you can be a little afraid here because this is a zeroth order Henkel transform. You can believe there are higher-order Henkel transforms. I'm sorry to break it to you, but there are. We don't have to deal with it. So for us, the zeroth order Henkel transform is nothing other than the Fourier transform of a radial function.

The headline from this calculation is that the Fourier transform of a radial function is another radial function. That's important. I don't tell you this for the fun of it. This comes up in all sorts of different applications just because radial functions come up in all sorts of different applications. It's the kind of thing you have to know. You don't have to be intimately familiar with bessel functions and all the rest of that jazz anymore, but you do have to know these sorts of things. The basic fact you should walk away with is that radial functions go to radial functions under Fourier transforms.

Also, the thing to be aware of is that this can cause some complications on the computational side of things because some algorithms and some computations are set up in cartesian form, and they don't convert so easily into polar form. So that causes problems. Full stop, something new, something different. We're going to have – I'm not going to go through the same hit parade of Fourier transforms in the higher-dimensional case like I did in the one-dimensional case because the point is that what you know in one dimensions carries over to higher dimensions. I do want to go over how the theorems look because there, I want you to see how both things look similar and how things look different.

We don't have much time today, so I don't want to get caught in the middle of something. Let me say a little bit about what happens to convolution and maybe the shift theorem. I may have to say the shift theorem and certainly the stretch theorem is where something very interesting happens that doesn't happen in one dimension. But let me mention convolution because here's something where it looks exactly the same as the one-dimensional case. So let me at least record that fact.

What is the definition of convolution of two functions? The definition of convolution of two functions can be made to look exactly the same as the one-dimensional case if you write it in terms of vectors. F convolved with G, at X is equal to the integral over RN of F of Y, G of X minus Y, DY. That's the same looking formula. Remember, F and G are scaler-value functions. So that's the ordinary product. There's nothing special there.

Now, you would not want to write this out in coordinates. You are much happier writing this out in vector form. Think about writing this down in the two-dimensional case. You would write F convolved with G of X1, X2 is equal to the integral from minus infinity to infinity of F of Y1, Y2 times G of Y1 minus X1, Y2 minus X2, DY1, DY2. That's what you would write out in the two-dimensional case. You certainly wouldn't want to write out the three-dimensional case. You wouldn't want to do that. Nobody would.

It's so many variables, so much space, so much pencil lead, but that's what it is. It's much easier to write out the definition in vector form. Now, here's something else you would not do. I tried to scare you off this, anyway. There's a whole file-and-drag way of life. The whole flip-and-drag religion of trying to visualize convolution. Flip this, drag that, screw you.

You would not flip and drag these higher-dimensional functions. Nobody in their right mind would try to flip G, drag F or do anything else when the graphs of these things are surfaces. The only case you could do it would be functions of two variables where you could visualize the graph as surfaces in R3 in three dimensions. Then you might think, I'll flip this, drag that, move it over the whole plane. I'll add it all up, and I can do that. That's worth my time.

Forget it. There's no way you could do anything like that in higher dimensions, but it is true that the standard interpretations that we have of convolution still hold. All the operations and all the interpretations of convolutions hold because the form is the same. The form is the same. It can be viewed as a smoothing operation. Maybe searing out the good properties of one function are inherited by the convolution. Just don't flip and drag. All formulas, properties, interpretations continue to apply.

That's good news. In particular, the Fourier transform property still applies. The so-called convolution theorem still applies. The Fourier transform of a convolution of two functions is the Fourier transform of F times the Fourier transform G. The Fourier transform of the product of two functions is the convolution of the Fourier transforms. All right? So the basis of signal processing is still with us.

Signal processing in higher dimensions amounts to, in many cases, taking the convolution of two functions in the spacial domain corresponds to taking the product of two functions in the frequency domain. To do filtering – now, one of the places this comes up, and you have a problem on this, actually, is images. We talked about this last time. What is an image? You can think of an image as a function of two variables, X1, X2, where the value, F of X1, X2, corresponds to the intensity.

Think about black-and-white images. [Inaudible] from zero to one. How do you do filtering, how do you do signal processing on images? Most often, you do it by some sort of multiplication in the frequency domain. The spectrum of the image times some thing, some mask, some filter that you are cooking up. You take the product in the frequency domain. That corresponds to convolution in the spacial domain. That's not the only way it comes up, but that's a typical way it comes up. The fact is, again, that convolution corresponds to multiplication if you take the Fourier transform. Same thing holds. I'm not going to prove it. It's the same proof, the same argument, as before.

You could've discovered higher-dimensional convolution the same way we discovered one-dimensional convolution by asking is there an operation, is there a way of combining two functions which corresponds to the Fourier transforms multiplied? That's what we did in one dimension. You could do the same thing in higher dimensions. Same thing because we've made everything look, in the higher-dimensional case, as much as possible like the one-dimensional case.

So it goes through. It's very powerful. All that hard-won intuition, all those formulas are going to go through. We'll quit for today. Next time, I am going to go through where things are a little bit different. I want to go through the shift theorem, and in particular, the stretch theorem. I love the higher-dimensional stretch theorem because it's new phenomenon that come up but actually have very important physical applications and physical implications. So we'll do that next time.

[End of Audio]

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture28

Instructor (Brad Osgood):All right. Let me remind you of the exam information as I said last time. I also sent out an announcement to the class this morning via the website. So the exam is a week from Thursday, the 13th crack of dawn, 8:30 - 11:30 in the morning and it is in Dinkelspiel Auditorium. It seats 700 or something like that, so it shouldn't be any problem. And as before, as I said last time, it is open books, open notes and all the rest of that jazz. And I will provide you with the formula sheet that you already have from the mid-term exam and is already posted on the class website. Okay. Also, as I sent out the announcement this morning by email, I changed the due date of the last problem set from Wednesday to Friday, so it's due on Friday instead of due on Wednesday. Okay. Wanted to have a chance to have a couple more groups to make sure we finish the relative material and give you a couple extra days. Okay. What a guy. All right. Any questions about anything? Anything that makes money? Okay. Wish we had one more day, I'm sure you're saying the same thing. I wish we had one more day of this class, but we don't. So it may be a little bit of a rush to try to get to the stuff on medical imaging. I wanted to finish up with that. We'll have to see if we can.

I won't be able to do as much of that as I wanted to, so I'm trying to arrange the material that we can do as much of that as I can because I think it's such a nice application. And out of a lot of the ideas that we've seen, certainly in the higher dimension, two-dimensional case, but to get there, we've got to do a few more foundational results on the higher dimensional Fourier Transform. So again, this is under the general heading of getting to know your higher dimensional Fourier Transform and again, the fact is you already know a lot about your higher dimensional Fourier Transform because you know the one dimensional version very well. Today, I want to do two other examples of that. At least two other examples of that. Maybe a little bit more. One where the situation is very much the same. The formula looks very much the same as the one dimensional case, and one where there's a difference where you can see traces, certainly, of the one dimensional formula, but there are [inaudible] phenomena in higher dimensions, already in two dimensions, that show up in the formula and it allows you to make a new maxim or a new aphorism that explains a little bit that has to do with the relationship between the time domain and the frequency domain and reciprocality in higher dimensions. All right. So I want to do a couple examples of that. The first one that I want to talk about looks very similar to the one dimensional case and that is the so-called shift theorem that we've seen and made much use of. Shift theorem. So again, I'm gonna remind you of the one dimensional version. A shift in time is reflected by a phase shift in the frequency domain. That's a way of saying it in words, and in formulas, it says if F of T corresponds to [inaudible] capital F of S. This is the signal in the time domain; this is the Fourier Transform and then what happens if you make a shift.

If you make a shift by B then that corresponds to E to the minus 2 pie ISB, that's the phase shift times the Fourier Transform of the original function. All right. That's easy result. That's one of the very first results that we proved when we were talking about general properties of the Fourier Transform and it follows, like many other formulas, just by making a change of variable in the interval that defines it. All right. Interval defines a

Fourier Transform. Well, what does the situation look like in higher dimensions? As a matter of fact, let's look at it in two dimensions because already there you can see what the general pattern is and the argument in higher dimensions is exactly the same. What's a 2D shift? When I say a 2D shift, I mean a function of 2 variables, and what does it mean to shift it? You can shift each variable independently. So X1 can get shifted to X1 minus B1 and X2 can get shifted to X2 minus B2 and so the function F of X1, X2 can get shifted to F of X1 minus B1 on an X2 minus B2. All right. And the question is what happens to the Fourier Transform if you make that shift? Well, we have no recourse other than to the definition. So the definition of the Fourier Transform and the shifted function is, and again, I'm just doing the two dimensional case so I have to write this out in coordinates. It goes from minus infinity over minus infinity E to the minus 2 pie I X1 times C1 plus X2 times C2 times F of X1 minus B1, X2, minus B2, DX1 and DX2. Man, is it tough to do higher dimensional Fourier Transforms on the board. All right. Now, there's really only one thing to do here and that is make a change of variable. The techniques in a lot of these results are making a change of variable interval and that's where some tricky things can come in, and some not so tricky things.

This is a not-so tricky case where it is an easy to make a change of variable. So you can make a change in variable, an easy change. And in this case, the area element, the X1 DX2 changes the DU1, DU2. It's the change in the area element. Not a change in DX1 and DX2 separately. That's one of the things that's a complications in higher dimensions, but for such a simple case, it's just a direct change. So DX1 and DX2 is equal to VX1, VX2 and the interval becomes the limited integration stay the same, again, going from minus infinity to infinity minus infinity to infinity. So that's E to the minus 2 pie I, so X1 becomes U1 plus B1 times C1 plus U2 plus B2, C2, F of U1, U2, DU1, DU2. All right. Now, this is not hard to manipulate [inaudible]. I could write this as minus infinity to infinity minus infinity to infinity, E to the minus – what I would do is I want to split up the complex exponential. All right. So E to the minus 2 pie, U1 times – let me put together here. Let me do this. B1 times C1 plus B2 times C2 times E to the minus 2 pie I, U1 times C1 plus U2 plus C2 times F of U1, U2, DU1, DU2. I split it up that to make sure I got everything right there. I took all the exponents into account. And then what remains is just the Fourier Transform. I did make a mistake. There's a minus sign there. Okay. So that's E to the minus B1, C1 plus B2, C2 times the interval, double interval minus infinity to infinity minus infinity over infinity, and what remains is just the Fourier Transform, but the U1 and U2 are replacing X1 and X2, so U1 times C1 plus U2 times C2, F of U1, U2, DU1, DU2 so what is the answer [inaudible] the answer is oh, well, nag, nag. Minus 2 pie I. What? What? What? Took my argument. Minus 2 pie I. Is that all? All right. So what's left? E to the minus 2 pie I, C1 times B1 plus C2 times B2 times the Fourier Transform of F at C1, C2. All right.

Now, there is, after all this work, a much easier way of writing it or a more compact way of writing it that shows the shift theorem in higher dimensions to look exactly the same really as the shift theorem in one dimension. All right. Once again, here's the phase shift term out in front, but you can see that this exponent here, what's in the exponential is the inner product of the vector C1, C2 with a vector B1, B2. All right. So if I write B as B1, B2 and of course C as before is C1 times C2 then F of X1 minus B1, X2 minus B2 is just, in vector form, I'd write that as just F of X minus B. You can make it look as much like the one dimensional case as possible. And that formula holds in any dimensions. All right. Any number of dimensions. It's pretty. All right. It's the same formula in higher dimensions as it is in one dimension because you're writing things in vector notation. The vector notation allows you a very compact and very easy way of writing it and a very easy way of remembering it. All right. I think it's much easier, I think, to remember this than it is to remember this. Whether or not you can remember your 2 pie I's. Okay. All right. So this is an example. I wanted to go through the derivation just to show you how it worked. The main technique, the only technique involved was a change of variable in the multiple interval. And I wanted to go through this example so you could see how that worked and as an example of how something looks very much the same in the two dimension as it does in one dimension or even higher dimensions as it does in the one dimension. Now, let me look at one of the formulas where a new phenomena comes in. And that's the stretch theorem. So there's two basic theorems that we've used a lot. One is the shift theorem that I've just derived, and the other is the stretch theorem when the variables are scaled.

Okay. And again, we've used that many times in many different contexts. And the interesting, sort of physical aspect of the theorem or the thing that comes up often is there's sort of a reciprocal nature between stretching in one domain and shrinking in the other domain. All right. That is the reciprocal relationship here between what happens in the time domain scaling by A and the frequency domain scaling by one over A. So this sets up this reciprocal relationship between scaling the two domains. I don't know if that's a good sense or not, but you know what I mean, scaling in the two domains. All right. Two domains. Long quarter, huh. Two domains. All right. Scaling by A in the time domain means scaling by one over A in the frequency domain. And again, the proof of this was integration was the change of variable in the integration. There's only really sort of two techniques that ever come up. One is a change of variable and the other is integration by parts. Those are the two techniques that come up all the time. Now, what about in the higher dimensional case? So again, let me look in the two dimensional case and I'll state one formula that is directly analogous to the two dimensional to the one-dimensional formula and then I want to do something that's a little bit more general. All right. Let's look at the 2D case. And again, we can scale the variables independently. So we scale, so we have a function of X1 and X2 and we can scale X1 and X2 independently. All right. And the question is what happens to the Fourier Transform? All right. I'm not going to go through the derivation. But I'll give you the result. The derivation is a change of variable, a simple change of variable because you can change the variables. You can make a change of variable independently so it's not difficult to change the variables in the multiple intervals that compute the double interval that computes the Fourier Transform of A1 and A2.

Here's what you get. Let me give you the special case first. It looks like this. Again, if F of X1, X2 transforms to F of C1, C2, all right, the special domain and the frequency domain, then F of A1, X1, A2, X2 transforms to one over absolute value of A1 times 1 over F of value of A2 times F of C1 over A1, C2 over A2. Just in pretty close correspondence to the one dimensional case. Okay. If the one dimensional variables X1 and X2 scale separately, then so do the frequency variables, C1 and C2 and it's a reciprocal relationship again. Okay. So again you have reciprocal relationships. There is a more general notion of scaling in higher dimensions and requires a little bit more complicated derivation and a little bit more complicated looking formula, but it's a very important phenomena that comes up when you do this, and it's very similar to what we were talking about when we talk about linear systems. All right. The basic result of linear systems is direct proportionality. That's scaling. All right. But the more general way of looking at linear systems is say multiplication by a matrix. And likewise in this case, a more general notation of scaling or stretching is not just stretching X1 and X2 independent but allowing it to mix up what happens to X1 and X2 by multiplying by a matrix. And here's where two dimensions and higher dimensions are richer than one dimension. There's more degrees of freedom or general notation of scaling. All right. Instead of just X1 going to A1, X1 and X2 going to A2 times X2 independently; you could have something like X1 goes into a culmination. To write that a little more compactly, [inaudible] could go the matrix ABCD times X1, X2. There's more degrees of freedom in higher dimensions that there are in one dimensions. That's the richness, and some cases, the complication of the field, but there's not much of a difference between two dimensions and higher dimensions here.

So in two D and up a general scaling is multiplication by a matrix A, say, and I want to assume that A is non singular, all right. So there's no collapsing here. We'll assume the determinant of A is not zero. So in other words, F of X gets scaled to F of A times X. All right. And the question is what happens to the Fourier Transform if you do this kind of scaling? Now, there is actually a very attractive formula and I want to show it to you. And a very important formula. Again, we have no recourse here other than to appeal to the definition of the Fourier Transform and to see what we can do. What's more complicated here is – and the technique is pretty much the same to make a change of variable in the interval, but the change of variable is a little bit more complicated. All right. So the Fourier Transform of F of AX – I'm gonna do this in higher dimension just for the thrill of it. I'm gonna do this in RN. All right. The Fourier Transform of F of AX is the interval over all of RN E to the minus 2 pie I X.T F of AX DX. That's just the definition of Fourier Transform where F is scaled by A. Okay. Now, I want to make a change of variable [inaudible] U equals AX. All right. That's fine. But it introduces a number of complications, so the variables are not changing independently here. What I mean is they're all coupled. Okay. Now, what you have to know, and I'm certainly not gonna derive this and I don't know your experience in dealing with the changes of variables in multiple intervals, so I'm gonna give you the formula and if you would ever like to have a quiet conversation about how to change variables in multiples, say, over a drink sometimes, which is what it takes, I'd be happy to do that. All right. But it looks like this. What happens to the volume element is DU becomes the absolute value of the determinant of A times DX.

When you make a linear change of coordinates, when you make a linear change of variables, the volume scales by the absolute value of a determinant. All right. So if you think of DU as the volume element DU1, DU2 up to DUN, and DX as the volume element in the X variable DX1, DX2 and so on, then the volume scales by multiplication by the determinant of A. I wouldn't be surprised if you've seen that as a basic geometric fact about linear transformations, what do linear transformations do, they scale the volume by the determinant. Okay. There's something else that goes on here. Okay. There's something else. So the other thing that happens is in the exponential there. So if I want to make a change of variables, what happens to the exponential? What about a complex exponential. So I have X.C, that's the thing I'm worried about there, what happens to X.C? Well, again, I'm assuming the matrix is non singular; the determinant of A is different from zero so if U is equal to AX then X is equal to A inverse U. I want to express everything in terms of U. Everything in the variable has to be expressed in terms of U. So if U is equal to AX then X is equal to A inverse you, that's fine, then I have to look at X.C is equal to A inverse U.C. All right. Now, something more actually can be done with this. All right. There is a general relationship, and this is very important, there's a general relationship between matrix multiplication and dot products. And it says this. Let me just give you the general statement. It says [inaudible], but there's a general relationship that's going on here. So it says that if I have a matrix B then B of X.Y is the same thing as X.B transpose of Y. All right. You can shift the matrix from one variable to the other and in so doing, the transpose comes in. All right.

Now, I want to plug this into the formula for the Fourier Transform. Okay. Plug all this into the formula for the Fourier Transform. Okay. So where are we? It's an integral still over all space because you're making a change of variable by a non singular matrix space goes the space and not a lower dimensional space. [Inaudible] E to the minus so I'm making a change of variable here in this interval. And everything should be written in terms of U. So it's an interval over all of RN, E to the minus 2 pie X.C becomes E to the minus 2 pie IU.A inverse transpose of C. All right. F of AX becomes F of U. That was the whole point of making the change of variable. All right. That becomes F of U. All right. And DX, the volume element in the X coordinates becomes one over the determinant of A [inaudible] of the U coordinates, DU. One over determinant of A DU. All right. Now, you see that interval is nothing but the Fourier Transform of F evaluated at this variable. Not a value of C, but evaluated at A inverse transpose of C. All right. So this is one over the determinant of A times the Fourier Transform of F evaluated at A inverse transpose of C. That is the formula. What is the formula? Let me summarize it for you. So the general and stretch theorem is very attractive really, if you like this sort of thing.

But here's the stretch theorem. All right. So F of X corresponds to, say, capital F of C, all right, so this the special function, this is its Fourier Transform. Then F of AX, I want to write it so it looks like the one dimensional case as much as possible, it belongs to one over the Fourier Transform. This is a determinant to of A times the Fourier Transform of F evaluated at A inverse transpose at C. A inverse transpose at C. That's the formula. All right. That looks a little different, right, than the classical stretch theorem, but it includes the one dimensional stretch theorem. It also includes the theorem that I had first. Let me say a couple things about this. All right. That's how the Fourier Transform changes, deal with it. All right. So now there's particular cases. This scaling is given by this matrix times X1, X2, A1 times X1 zero plus zero is zero times X1 plus A2 times X2. All right. And of course, so this A then the determinant of A [inaudible]. All right. What is A inverse? Well, for a diagonal matrix, A inverse – and I have to assume that A1 and A2 are different from zero here. I'm assuming the determinant is different from zero. So for a diagonal matrix, A inverse is equal to one over A1, 1 over A2, 00 and it's diagonal again so it's transpose is equal to itself, A inverse transpose is equal to the same thing, 1 over A1 00 1 over A2 and if you unpack all this, you will find that applying the general stretch theorem, in this special case, leads exactly to the theorem as I had stated it before. Now, there's another important special case. We haven't lost anything. And we've gained something, but in particular, we've included the formula that I had before.

Another special case, let's just do it in two dimensions, is when coordinates are rotated. That is two dimensional rotation matrix. Two D rotation. That is A equals – it's usually written this ay – cosign of theta minus sign theta, sign theta, cosign theta. All right. That is a rotation through an angle theta clockwise, I think. I always have to think about that for a couple minutes, but I don't have a couple minutes. All right. That's a rotation by an angle theta. Now, this matrix has a special property. This matrix is orthogonal in the following sense. So rigid motion, it just rotates. To say it's orthogonal is to say that A transpose times A is the same as AA transpose is the 2 by 2 identity matrix. Okay. So there's a couple of consequences of that. One consequence is the determinant is absolute value 1 because the determinant of A transpose A, on the one hand, is equal to the determinant of the identity, which is 1. On the other hand, the determinant of A transpose A is the determinant of a transpose times the determinant of A. The geometric meaning of that is that A preserves volumes. And it does. It's just a rotation. It's a rigid motion. So it is not distorting volumes. Volumes are distorted under linear transformation by the absolute value of the determinant. But in this case, the absolute value of the determinant is one. The further consequence is [inaudible] A transpose times A is equal to the identity. That says that A transpose is the same thing as A inverse because the inverse of a matrix is the identity so that identifies a transpose as an inverse or the other way of putting that is A inverse transpose is equal to A. That is A inverse transpose is equal to A. Oaky. All right. Plug that into the formula in the case of rotation matrix. In the case of a matrix like this, plug this into this general formula.

Plug in the formula. If you plug into the formula, F of AX goes to 1 over the determinant of A times the Fourier Transform at A inverse transpose at C. all right. The determinant of A is one, A inverse transpose is A, so this is the Fourier Transform of A applied to C. That is F of AX goes to the Fourier Transform if I evaluated it AX. It goes to the Fourier Transform at AC. All right. In words, a rotation of the special domain corresponds to the same rotation of the frequency domain. This is a rotation of the special coordinate. This is a rotation of the frequency coordinate. So this says rotation of the spatial domain corresponds to the same rotation of the frequency domain. If you rotate an image, that's the same thing as rotating its spectrum. Okay. You take the image, you take its spectrum, you rotate the image, you rotate the spectrum by the same amount. All right. So this is an important result in imaging. All right. It tells you that the spectrum somehow is rotated along with the original image. Okay. Very nice. Very consistent. Finally, there's a new aphorism that goes with this theorem. It's also very important. And this is something new. All right. This is something that you probably have not said to yourself as you've been drifting off to sleep at night. Once again, if F of X corresponds to F of C then F of X corresponds to times F of A inverse transpose C. Now, if you believe that scaling in the one domain has to do with – there's a reciprocal relationship between scaling the one domain and scaling the other domain. All right.

If you believe that from a one dimensional case and if you want to maintain that intuition that there's a reciprocal relationship between the two domains then what you have to say that is in higher dimensions, reciprocal relationship means A inverse transpose. That's something new that you have to say to yourself. That's a new interruption of the world reciprocal. All right. You buy the [inaudible], you buy the gag as they say in show biz and if you believe in this formula – you have to believe in the formula because I just derived it, but if you believe that there ought to be a reciprocal relationship between scaling in one domain and scaling the other domain then what you have to say to yourself is in higher dimension, as tough as it is for me, I have to believe that reciprocal means, somehow, means, inverse transpose. Okay. Now, let me just say a few words about delta functions. Good news. First of all, we don't have to deal with rapidly decreasing functions, [inaudible] functions, distributions or any of that stuff, that all carries over. All right. So I don't really feel the need and you don't have to be afraid of a whole secondary treatment of rapidly decreasing functions in the theory of distributions. So we're gonna use these things pretty much as before, and delta functions also go through us before. [Inaudible] okay, delta functions, deltas et al. So again, delta is defined the same way – you can think of this in terms of distribution. Delta operates on a test function, you can write it out in terms of the interval if you wish, but the basic definition or the basic properties of delta functions are as before. As in 1D. So for example, delta paired [inaudible] a higher dimensional delta function paired with thee is thee of zero. So thee is a function of N variables and zero is just the origin. So delta operating on thee just pulled out the value of the origin. And I can look at a shifted delta function. I can write it as delta X minus D or I can write it as delta sub B if I want. So B is our vector here so it's shifted to another point in RN and the basic relationship is the same as in one dimension. That is delta sub B paired with a function fee is fee evaluated at B. Okay. So B here is B1 up through B10. Okay. [Inaudible] and it's also true, good news, that the Fourier Transform behaves this same way. You have the same formulas for the Fourier Transform. That is to say the Fourier Transform of the delta function is 1, the constant function 1 and the Fourier Transform of the shifted delta function, delta sub B is E to the minus 2 pie I B.C. All right. So if you use the vector notation, it looks the same as the one dimensional case.

Okay. It's a complex exponential. That's all great. And again, it's also true that if I multiple delta times – I'm sorry, I should've mentioned this before, if I multiply delta times a function, it just samples the function. Multiplication – I should've mention this when I was talking about the basic pairing here – is if I take a function F times delta as F of 0 times delta and if I take F of 0 this is a vector function now. A function of N variables, and likewise, if I take F of delta sub B, that's the same thing as F of B times delta sub B. Same results as before. Okay. Everything is the same. All right. Now, where are things different? Well, again, when you start to bring matrix's in, actually, things becomes a little bit different. When you start to talk about scaling. And I think this will be the last thing I say today. I probably won't derive it. I'm gonna use it more next time.

Okay. But let me mention now the scaling properties. Scaling. The scaling properties of delta so here I actually have to write the formula in terms of a variable. The basic phenomena is – well, again, what was the one dimensional case?

The one dimensional case was if I take delta and scale it, then delta AX is 1 over absolute value of A times delta of X is the way I allow myself to write the variable delta of X instead of thinking of it as a distribution. Okay. That's a result we used in various context. All right. The tricky thing here is that there's no scaling in the inside here. All right. It's not the stretch formula for the Fourier Transform, it's an independent fact about how delta scales. Okay. And what do you suppose the result is in higher dimensions? If I scale the variable inside by A, there are ways of making this precise without bringing the variable in, but I'm not gonna do that. It takes a little more effort. So if A is a matrix here, then this is 1 over the determinant of A – it's really analogous to the one dimensional case – times delta of X. That's how delta scales. We're gonna need this result. The derivation is actually given in the notes, but I'm not gonna derive it. We are gonna use it however. I won't try to push ahead today. We're gonna use this next time when we talk about higher dimensional Shaw functions. That's where this sort of lattice is reciprocal lattice, there's actually higher dimensional sampling theory, it all comes in. And we'll have a chance to do a little bit of it, so we'll see this in connection with higher dimensional Shaws.

It's really cool. All right. So read through that material if you would. You don't have to read through the sampling part of it because I don't think we're gonna have a chance to do that although it's my favorite stuff, but I do want to talk about the crystallography, the application of the Shaw stuff to crystallography because that's where you see this wonderful fact about reciprocal means inverse transpose come in in a very important setting. All right. So we'll do that next time and I hope to actually push forward a little bit and start talking about medical imaging. All right. See you all then.

[End of Audio]

Duration: 50 minutes

TheFourierTransformAndItsApplications-Lecture29

Instructor (Brad Osgood):Does somebody know – don't you occasionally get bad reactions to flu shots like about a week later? I got a flu shot last week, and I sort of had all these flu symptoms yesterday, like I thought I was gonna die for example. So anyway, I won't be my usual perky self today, and of course, only half the class is here, so I guess perkiness is in short supply these days. All right. I wish we had one additional day. Well first of all, let me call your attention one more time to the information about the final exam. It's next Thursday. That's a week from tomorrow from 8:30 to 11:30 in Dinkelspiel Auditorium. I'm sure you can find that. It is open books, open notes, and I'll supply the blue books, and I'll bring copies of the formula sheets, and so on, and so on. Does anybody have any questions about the exam, about the mechanics of it or anything like that? No? Everybody brimming with confidence, or about shocked and doesn't really wanna think about it right now? Okay. Like I say, I wish we had one additional day in this quarter because I [inaudible] don't quite have enough time to do what I wanna do in full detail, or even if not in full detail, just enough detail so you can see some of the finer points of the arguments. But nevertheless, I still think it's important to try to go over the main highlights of two more topics: 1.) a continuation of what we talked about last time, the general stretch theorem and actually, sort of an application and a physical manifestation of the general stretch; and then the last topic which I'll begin today and then finish up next time on medical imaging and application of the two-dimensional Fourier transform to medical imaging.

And again, I'm sorry. I really do regret not having one more day somehow that we could do this in a little bit more detail. But to give you a little – at least a sense of some of the directions that some of this material goes. So today I wanna talk about – and actually, this hearkens back both to yesterday, last time, and stuff we did earlier when we first started talking about the Shah function. I wanna talk about Shahs, lattices, and crystallography. Now I'm doing this because I really think this is a really interesting application. It's very pretty, and it just shows you somehow how some of these ideas come into play in unexpected ways. This is an application – although again unfortunately, we can't see all of the details. I just don't time to give all the details, although it's discussed in the notes. This is an application and say physical manifestation of this idea that in higher dimensions in reciprocal means inverse transpose. That's what we talked about when we talked about the general stretch theorem, and I'll remind you what that says. It's a physical manifestation – physical example. Manifestation is too long a word for a man with a grip – physical example of the idea is reciprocal means inverse transpose. We saw that phrase, that aphorism in the context of the generalized stretch theorem, so I'll remind you what that says. And that comes into the derivation of what I'm going to be doing. So we saw this in the generalized stretch theorem which told us how to take the Fourier transform one and make it change a variable by a matrix, and not just scale the variables independently.

So it said this: it said the Fourier transform of F of AX – all right, so you change the variables X by a matrix A, a nonsingular matrix A – is one over the determinate of A times the Fourier transform of F evaluated at A inverse transpose at the frequency variable C. Okay? It's a very interesting formula. We derived it last time, and it's complicated. It's more complicated than the one-dimensional stretch case, but it includes the one-dimensional stretch case, but what you don't see in one dimensions is this new phenomenon as I say that reciprocal somehow means inverse transpose. That's something new. That's something different on the scene. All right, so now I wanna show you how this comes in in these contexts. So first of all, I wanna talk about how we might generalize the Shah function to higher dimensions, so let me remind you of the Shah function in one dimension, and some of its main important properties. So I want to generalize the Shah distribution to the sum of deltas. Let me say evenly spaced deltas. So in one dimension we had the model case. That was a Shah function where the deltas are spaced by one, so you put a delta at each integer point. The model case is the basic Shah function where you put a delta function at each integer, zero, one, two, minus one, minus two, and so on. And you define the Shah as just the sum of these delta functions. Shah of X is the sum of K going from minus infinity to infinity at delta X minus K, so there's a delta function at each integer point.

And the remarkable property – and as we said when we were talking about this, this is the deepest property known about the integers. Its equivalent to this fact in number theory called the Poisson summation formula is the Fourier transform of the Shah function is itself. The Fourier transform of Shah is Shah. And that was fundamental in a lot of the work we did. It was fundamental also into the sampling theorem, and also into other applications. We actually looked at this in a slightly more general version, which I'll remind you of just now, in the context of crystallography, sort of as a way of motivating it, and I'm gonna talk exactly about that in just a second. So we generalize this slightly. We allow for spacing P for other spacing other than the integers, so spacing P. Doesn't matter what I call it. So instead of space to the integers, it's space – I put deltas at the points zero, P, 2P, and so on, and so on, minus P, minus 2P. So this time the Shah function is the same sort of definition but the spacing is different. So again it's the sum from minus infinity to infinity of delta X minus K times P. Here the fundamental result is that when you take the Fourier transform, the spacing becomes reciprocal. That is the Fourier transform of – let me use the subscript here P to indicate that that's what the spacing is. Sorry. The Fourier transform of the Shah function with spacing P is actually – there's two places where the reciprocal comes in. It gets scaled out in front by one over P, and then the spacing of the deltas in the frequency domain – the spacing of the deltas when you take the Fourier transform is spaced by one over P.

That's pretty cool. It's a very interesting result, and it's a very important example of this sort of reciprocity relationship between the two domains. And this is what I wanna generalize. All of these properties I wanna generalize. Now, I'm gonna stick in two dimensions, but really – and there's a change from one dimension to two dimensions, but there's very little change from two dimensions to higher dimensions. I can draw the pictures more easily in two dimensions, but really all I have to say will hold in higher dimensions, and in particular three dimensions, which is really the appropriate setting for studying crystals. But as I say just so I can draw the pictures a little more easily, I'm gonna write things and speak about things just in two dimensions. So what's the generalization of Shah let's say to 2-D. What should it be? Well, what I wanna do is –

there are two parts to the Shah function. I put a bunch of delta functions, but before I put the delta functions down, I have to have a set of evenly spaced points. That was the whole idea behind the Shah function, so you need – that's the first thing that you want. You want evenly spaced points, and you have to decide what that means, evenly spaced points in the plane, in R2. Now there are a lot of different possibilities here, but again we started with a Shah function with a model case where the deltas or the points were spaced one apart. So for my model of evenly spaced points in the plane, I'm gonna take all the points in two dimensions whose coordinates are integers. So I'll take a grid in the plane whose points have integer coordinates like so.

So that's gonna be my model for evenly spaced points in the plane. And of course, in the higher dimensions it would be a similar sort of thing. I would take points with integer coordinates in R3 or in higher dimensions. So a point here, a grid point has coordinates K1 and K2 where K1 and K2 are integers, all these points. Those points are said to form a lattice. That's not an unfamiliar term. So this forms what's called the integer lattice in the plane. It's denoted by Z2, sort of a bold-faced letter Z. I can't remember if I've used that notation before, Z for the integers, to stand for the integers. Z stands for the German word zahlen, which means number. It's a universally adopted term in – pretty much a universally adopted term in certainly mathematics and more and more so in engineering in application just to denote the integers. So Z2 meaning it's in two dimensions. So this is the set of all points, nothing but the set of all points and the lattice points in the plane with integer coordinates. And my Shah function is gonna be defined by putting a delta at each one of those points. Put a two-dimensional – put a spike at each one of those points. So define by Shah by putting a delta at each one of those points, each grid – I'll call it lattice points, okay? That is I'll define the Shah function for this integer lattice as exactly the analog of the one-dimensional case of X. It's gonna be the sum – let me write it like this: K in Z2 of delta X minus K. So again, what I'm trying to do here is I'm trying to make this look as much like the one-dimensional case as possible. So if I write the vector K, what I mean by that is it's a pair, K1, K2 where K1 and K2 are integers. So if I write it like that again it looks like the one-dimensional case. Here again K is just shorthand notation K1, K2, integers.

The basic phenomenon – again, generalizing the one-dimensional case is what happens to the Fourier transform of the Shah function for the integer lattice. And this is unfortunately something I don't feel like I really have the time to derive, and it just kills me, but I'll give you the result. So just as in the 1-D case, you have with the Fourier transform – and for the same reasons really, it's a straightforward – you set things up so it looks exactly the same as in the one-dimensional case. You have the Fourier transform, the two-dimensional Fourier transform of this lattice – the Shah function based on the integers is itself. The proof really again, the derivation of that is the same as in the one-dimensional case. It depends on – actually, it depends on higher dimensional Fourier series, but it's really very much the same. You can use the same words and really follow the same argument. The argument is very much the same, very much like the 1-D case. Now, things get more interesting when you ask yourself, "How do I change the spacing?" In the line, in one dimension, you only have one degree of freedom. The only way to change the spacing is to stretch or shrink in one direction. So how do I change the spacing for a 2-D lattice? What should we mean by that? We still want the points to be somehow evenly spaced, but not the same. Well again, in two dimensions, we have more degrees of freedom than in one dimension, but for our purposes, there's a very natural way of looking at this. In two dimensions, we obtain a different lattice. Informally, I'll call it an oblique lattice. I'll draw a picture in just a second. That is we change the spacing by applying a linear – a two by two matrix say A to the integer lattice.

That's what it is in words. Let me draw a picture so you see what I mean here because it's not hard to see what I mean. So here let me draw a picture of the integer lattice over here. So this is Z2. And as a matter of fact, let me take – let's consider this the origin, and let's take this to be the natural basis of R2, so the two natural basis vectors one zero and zero one. Now I take a linear transformation, a two by two matrix A nonsingular, so nothing's getting collapsed. Then they're gonna take those basis vectors E1 and E2 to some other vector, say V1 and V2. And they're gonna take – now all the points in the integer lattice are just integer combinations of these basis vectors. If I wanted to reach the Point 2 over here, then I just take two times E1. If I wanna reach the Point 2 up here, I take two times E2. If I wanna reach this point, I take E1 plus E2. Okay? One times E1 plus one times E2, and so on and so on. So all the points in the integer lattice are just integer combinations of the basis vectors E1 and E2. Under a matrix or linear transformation, they go over to integer linear combinations of V1 and V2, so what I have over here is a picture that looks like this where the lattice points over here correspond exactly to the lattice points over here, and so on and so on.

So the lattice is sheared if you wanna think of it that way. The square lattice of the integers is made into an oblique lattice by taking a matrix that takes the basis vectors over here, the natural basis vectors and some other pair of vectors. So A of Z2 – A applied to – I just mean by this applying A to all the different basis – all the different points in Z2 results in an oblique lattice L. Okay? And L is just the combination – the points in L are combinations of the basis vectors K1 times V1 plus K2 times V2 where K1 and K2 are integers. And here V1 say is A of E1 and V2 is A of E2. Stay with me. Now, one more geometric – actually, there are a couple of other things I have to introduce. It's very common – as a matter of fact, crystallographers do this – do we have anybody who has actually done any crystallography in here? By the way, I'm just curious. Then this is all new.

But if you do any kind of imaging, you're gonna see these terms and see these ideas, especially these days in molecular imaging and things like that where these things are coming in in really interesting ways. As a quick geometric description – so to jump ahead a little bit, and I'm gonna come back to this in a second, crystals come into this because you consider crystals to be modeled on lattices where there's an atom at each lattice point. And of course, the idea of considering oblique lattices is because some crystals are modeled on a rectangular lattice or a cubic lattice, but a lot of crystals aren't. There's sheer. They're oblique. There's a geometric – just to get an idea of somehow just to attach a number that describes a lattice in some loose but helpful sense, you often talk about the area of a lattice. So you often speak of the area of really what's called a fundamental cell, or fundamental parallelogram, fundamental cell in the lattice. And that's just the area of one of the parallelograms spanned by the basis vectors. So again, if here's the lattice, and say this is V1, V1 at V2, then the area of the lattice is the area of the – we're gonna come back – we're gonna need this concept in just a second. You talk about the area of the lattice as just the area of the parallelogram determined by the basis vectors, so the area of L is area of the parallelogram.

The reason why crystallographers think about that is because they think of that as the fundamental cell or fundamental building block for the rest of the crystal. That region in space is the fundamental building block for the rest of the crystal, and the question is how big is that. They talk about the volume of the fundamental cell. Everything in crystallography of course is in three dimensions. I'm only doing things in two dimensions, so I'm talking about the area in terms of volume, but that's what they do. They talk in those terms. So the area of the integer lattice is one because each one of these basis vectors is the length of one, and they're perpendicular. The area of that square is just one. Now I wanted to introduce that, but the real thing I want to get to of course is what happens if we consider the Shah function for an oblique lattice, and what is its Fourier transform. So again considering the Shah function for oblique lattice is like considering a model for a crystal where the underlying lattice structure or underlying crystal structure is on an oblique lattice instead of on a square lattice, so we can consider – let's call it the Shah function of lattice L. I'm struggling here with my script letters.

So let me just write it like this, it's the sum over all points in the lattice, so those are all the lattice points, the grid points – I put a delta at each one of those points, say delta of X minus P. You put a delta function at each lattice point P. Now the real question is what is the Fourier transform. What happens to the Fourier transform? What is the Fourier transform of the Shah function associated with the oblique lattice? A lot's going on here, right? You have to have the notion of a lattice. You have to have the notion of an oblique lattice. You have to have the notion of a Shah function on the lattice. And now you're asking about the Fourier transform. I mean what levels of complexity have been added to this simple conversation. And the reason why – and again, I'll come back to this in a second, but the reason why you wanna know about the Fourier transform is because in honest to God scientific experiments on crystallography, when you're shining X-rays through a crystal, you get a bunch of spots, and the pattern of the spots is determined by the Fourier transform of the crystal that's doing the diffracting.

So if the crystal that's doing the diffracting is modeled by the Shah function for a lattice, what you know about is you wanna know about the Fourier transform of that Shah function for the lattice because that's what you're seeing. Those are the spots that you're seeing. So now finally, here's where this reciprocal relationship comes in and it's just great. And again unfortunately, and I really do feel badly about this, I don't have the time to talk too much about the motivation for this, or even derive many of the formulas, so I just wanna show you in some sense what the punch line is without the whole set up, and it just kills me, but I've got to. So I'm gonna put two other lattices on that board, but here's the integer lattice Z2. I have two linear transformations. I have my linear transformation A that's going to go down to the oblique lattice L. That's the sort of the lattice that I'm starting with. Here's the oblique lattice L. And then the crystallographers

– and even mathematicians, too, but for completely different reasons as far as I can tell – define what they call the dual or reciprocal lattice, L star. So L is A of Z2. L is the transformation A. Any lattice can be realized as a linear transformation applied to the integer lattice. So L is A applied to Z2. Then the crystallographers the reciprocal or dual lattice to be – guess what – A inverse transpose applied to Z2. So they take A inverse transpose – [Crosstalk]

Now that's also another linear transformation. It goes into another lattice, and I don't have great confidence in actually how I draw it, but something like this say, L star. And they call that the reciprocal lattice. And in fact, as you can easily check, the area of L star – that is to say the area of one of these fundamental polygons, polyhedral is – so I certainly didn't draw it very well because it doesn't look quite so easy – is the reciprocal of the area of the other lattice. Now this is one definition of it, and this is the cleanest definition. It's completely unmotivated. I'm sorry. There are different ways of motivating it. There's one discussion in the notes. That is it's a completely unmotivated definition. Why the hell would you define a dual lattice or reciprocal lattice this way? You just have to take my word for it or look around for other motivations for it, other definitions for it. In fact, people who work in materials and people who work in crystallography actually have a geometric construction of dual lattice. Given a lattice L, they have a way of constructing geometrically the dual lattice, like you know a ruler and compass construction. So it's a time honored important construction in materials to pass from the lattice to the reciprocal lattice.

Now, why? What is the big deal? Because the fundamental on the Shah function is that the Fourier transform for the Shah function for a lattice is the Shah function for the dual lattice except scaled by the area, so that's a fundamental fact. The Fourier transform – I've got too many Fs and Ls and everything else in here. The Fourier transform of the Shah function of a lattice L is – it's a beautiful formula – it's one over the area of L times the Shah function of the dual lattice. Now again in higher dimensions which is where the real applications of crystallography go, you'd have the same result except you'd have volume here instead of area, but it's the same kind of result, and the definition of a dual lattice is the same, except instead of a transformation in R2, instead of everything happening in R2 here, everything would be happening in R3. You'd have the lattice of integer points in R3. You'd have a three by three matrix going over to an oblique lattice here in R3. You'd have the inverse transpose going into another oblique lattice of R3 and so on. And you'd have the same result. This is exactly – if you believe it – analogous to the one-dimensional formula. This is the analog of the one-dimensional formula. This is the analog to the Fourier transform of the one-dimensional Shah function of with spacing P is one over P times the Shah function with spacing one over P.

There reciprocal just means one over. But see, the thing is you miss something. If you just look at that, you either miss something in the one-dimensional case, or something new and deeper is revealed in the higher dimensional case because in the one-dimensional case, reciprocal means reciprocal. It just means one over. In the higher dimensional case, reciprocal means inverse transpose. That's the change in point of view. That's sort of the change in intuition. It's very important. That's something you don't see in the one-dimensional case. That's a difference between the one-dimensional case and the two-dimensional case in how to understand the notion of reciprocity, to understand what reciprocal means. It's a mathematical fact, and not only a mathematical fact, a fact of nature that in higher dimensions reciprocal often means – can be interpreted as involves, ergo it [inaudible] reciprocal transpose. Now so what about crystals in all of this? I made allusions to this a lot, so let me just say quickly what the situation is for crystals and why this is so important or where this comes up. So again for crystals – so again when we did this in the one-dimensional case, the idea is you study a crystal by studying the electron density distribution of a crystal. What is the electron density distribution of a crystal? You see, you conjecture, you measure how the electrons are distributed about an atom, and then you periodize that. Crystals have a periodic structure, so row of X in two dimensions now is the electron density for an atom in a crystal. The electron density distribution for the whole crystal is a periodized version of that. For the crystal as a whole, you take a periodized version. That's the whole point about crystals is they have periodic structure. How do you take a periodized version? You convolve with the appropriate Shah function.

So if this is the crystal whose atoms are what you model as a bunch of atoms at lattice points as L, then the density for the lattice, the density for the crystal is the density for a single atom convolved with the Shah function. That's sum over all the points in the lattice of row of X minus P, same as in the one-dimensional case. So if there's a little – if the density looks like that there, then that pattern's just repeated for the whole lattice, and again I can't draw this too well, but you get the picture. This is now periodic with respect to the lattice. This is also a different phenomenon. It has two periods. It's periodic in that direction, and it's periodic in that direction. It's periodic on the whole lattice. The pattern repeats on the whole lattice. Now X-ray crystal experiments, X-ray diffraction experiments measure the Fourier transform. That is an X-ray diffraction experiment produces the magnitude of the Fourier transform, but just basically think of it as the Fourier transform – produces the Fourier transform of row of the lattice. That's what it does. You see a bunch of spots. What you're seeing is the Fourier transform of the lattice, and what is that? Well, that's the Fourier transform of the convolution, so that's the product of the Fourier transform of row with the Fourier transform of the Shah. This is the Fourier transform of row times the Fourier transform of the Shah function of the lattice.

But we know what that is. It's the dual lattice. That is this is the Fourier transform of row times one over the area of L times the Shah function of the dual lattice. And if I take the product of the function times a bunch of delta functions, what do I get? I get sum over the points – let me call it P star in the dual lattice. They're just delta functions at the lattice points of the dual lattice of the Fourier transform of row at – I don't know, S times delta of S minus P star. So the Fourier transform of the density function times the sum of deltas, and remember what happens when you take a function times delta? It just pulls out the value – sorry, P star. So once again, you do your experiment, and unless you know the math, you're not gonna be able to draw the right conclusion. You might think that you do – you pass a bunch of X-rays through the crystal, and the spacing of the lattice points that you're seeing should be proportional to the spacing of the lattice points in the crystal, but it's not. Nature for whatever mysterious reason is taking a Fourier transform. X-ray crystallography, X-ray diffraction takes a Fourier transform for you physically. And what you're seeing is you're see the points at – you're seeing the little spots that appear on your X-ray film at points on the dual lattice. And unless you know the math, unless you know that in higher dimensions reciprocal means inverse transpose and so on, you can kiss your Nobel prize goodbye.

But if you sort of follow your pencil through this, and see how the math plays out here, you can draw conclusions about the crystal by knowing this. I think it's really cool. I think it's a really interesting application and physical manifestation of this fact that in higher dimensions inverse transpose means reciprocal. So I don't have time to derive this formula, and I'm sorry for that. The derivation of this formula – and that shouldn't surprise you given how all these terms are coming – involves exactly the generalized stretch theorem. The reason why inverse transpose comes in here, and the reason why the dual lattice comes in is because in deriving this formula, it uses exactly the generalized stretch theorem. That's how it – it comes into it in deriving that formula. Isn't that nice? So I say I wish we had one additional day because I would like to go into just a little bit more detail about this, but we don't. We don't because I spent all that time on the damn fast Fourier transform. Everybody wanted to see that damn fast Fourier transform. So full stop, speaking about things I can't spend enough time on, there's one more topic I won't be able to spend enough time on. The final topic in the class is application of the two-dimensional Fourier transform to the problems of medical imaging, and particularly to the problems of tomography. So we'll leave this fascinating subject behind, and pass to another fascinating subject. I'll start off today, and then we'll finish it up next time.

So the final application is – the final topic is the application of the two-dimensional Fourier transform to medical imaging, and in particular the problem of tomography. I will tell you what the problem of tomography is in just a second – the fundamental problem of tomography. Now let me say at once that what I'm gonna be looking at is really in the realm of CAT scan, of computer tomography, of tomography not MRI. But as it turns out this is actually – this is not MRI. As they say in the biz, the modalities of imaging for the two kinds of – two approaches are different. However, it's also true for whatever reason – well, for a reason that we can't really get into that the final formula that I'm gonna get to solve the problem of tomography is the same as the formula you get when you're using an MR system to do the imaging. So what we're doing here, although it's in the realm of tomography, CAT scans and things like that, actually the final formulas that you get and a lot of the considerations that we're gonna talk about actually apply in the MR case, magnetic resonance case, NMR case, like nuclear magnetic resonance, but again I cannot get into it. Here's the basic setup. You have a 2-D region, a two-dimensional region, think of it as a slice of your body filled with goop, blood, bones, organs, you know. There you are, and there's all that goop inside you. And that goop inside you is described by a function of variable density. You can imagine describing the density of the goop. So describe the goop by a function mu of X1 X2. So this is taking place in the X1 X2 plane, and mu gives the density of the goop at each point.

So if you know mu, you know you, so to speak that is to say you wanna get you by taking measurements. You want to recover mu X1 X2 by X-ray measurements or let me just say by whatever. That's what you want. You want to reconstruct the function mu. Now here's how it's done in tomography. The approach to tomography is the following. Tomography means – if I get this right. I forgot to look it up to get the precise – I think it means section, so the idea is you are taking sections, and in particular actually what I'm gonna do is I'm gonna take a one-dimensional section. That is I'm gonna pass a line through this region. So the approach via tomography – a line meaning I'm gonna pass an X-ray. The approach via tomography is you shoot an X-ray through the region along a line and it comes out the other end. It gets diffracted. It get attenuated by all this goop in the middle. So here's the X-ray, and there's all this stuff in here, and you measure what happens to it when it comes out. So you know how – you know what it was going in, and you measure what it is coming out. Say it goes this way, so you know the intensity of the X-ray going in, and then it gets attenuated, so you measure the intensity – let's call this I not, the initial intensity. You measure the intensity going out on the other end. Let's call that I.

Now then it is not hard to show as a reasonable approximation – and again, this is derived in the notes. You can show that there's a relatively simple – well, relatively simple. We'll see. There's a relationship between the intensity going out and the intensity coming in. This is I not going in, I coming out, and it's attenuated – it drops exponentially, so it's given by intensity coming out is the intensity going in, and then I say it drops exponentially. How? It drops exponentially according to the integral of mu along the line, sort of like the average density. Mu is a variable density, but the total drop in intensity is given by E to the minus integral – so integral of L over mu – I'm using a shorthand notation here – is the line integral of mu along L. Again, you can sort of think of that as the average density or whatever. There are different ways of making that argument, and again I don't wanna take the time now to derive that formula for you, but it's actually not that hard to derive, and it's not even that hard to believe.

Now again, you know I not. You measure I. So all those numbers are known. What you don't know is you don't know mu, and in particular you don't know the integral of mu along a line. What you do is you make lots of measurements. You send X-rays through along all different lines, or along whole families of lines. You send X-rays through along many lines, this way, that way. Usually you think of a parallel line, but then you change the angle. You send them in this way. Don't think of this as a lattice. I'm just changing my X-rays here. You send a bunch of lines in. Then you know, and at every time you know the intensity going in, you measure the intensity coming in, so that means you know the integral along all those different lines of mu. Once again, you write this down. I equals I not E to the minus integral of mu over the line L for all different L. You know this. You measure this. So then you know all the numbers integral over L of mu. When I say all the numbers, you know them along all these different lines. It's because you made those measurements.

And the fundamental problem of tomography as it is often stated is can you recover mu by knowing all these line integrals. That's the problem. So again, I wanna make sure you understand here at least what you know and what you're trying to figure out. You know the value of these integrals because you know this formula, and you know the intensity going in, and you measure the intensity coming out, so that tells you what this. You just solve for it in terms of log. So the question is – the fundamental problem of tomography is can you recover mu by knowing the integral of mu along L for all L, for all lines L. That's the question. And the answer is affirmative. The answer is in the affirmative. The answer is yes, and it involves the two-dimensional Fourier transform. That's how it comes into the picture. I'm not gonna get to that today, but I'll tell you how we're gonna approach it actually. So the answer is yes. It's not obvious, and it's certainly not obvious that the Fourier transform's gonna come into this in any sort of way, but here's the approach. This is the approach that we're gonna take, and here's how it works. And as I say, it's not obvious. It was a brilliant idea. You consider these numbers as defining a transform of mu. What I mean by that is you start with a line – what's that a function of? It depends on the line. The number depends on the line.

So starting with a line L, you compute the number, the integral of mu along L. L goes to – that's a line – goes to the integral over L of mu. That's a number. When I say transform is you wanna consider this a transform of mu evaluated on the line L. Now that's sort of mind expanding for you, but that's the way you look at it. That is write this as R of mu, the transform of mu at a particular line, so again that is by definition that integral. That's the integral along L of mu. That's called the Radon transform of mu. We're almost out of [inaudible]. It's called the Radon transform of mu. The fundamental problem of tomography can be stated as saying can you invert the Radon transform. R of mu gives you something [inaudible] L. The question is can you find mu given all the values of R of mu. Can we invert? Is there an inverse Radon transform to find mu? And the answer is yes, and the answer is in terms of the two-dimensional Fourier transform. It's amazing. It's just amazing, but that's what's gonna happen, so that's why the two-dimensional Fourier transform, and that's why Fourier transform techniques come into medical imaging because of exactly this connection.

So we will finish up with that next time, and me, I'm gonna go take a pill and collapse. Thank you all. I'll see you all on Friday.

[End of Audio]

Duration: 51 minutes

TheFourierTransformAndItsApplications-Lecture30

Instructor (Brad Osgood):And let's see, review so – oh, I'm on. Man, give me a chance here, will you? Where is the review session? The usual place? Review session today – what time?

Student:[Inaudible].

Instructor (Brad Osgood):4:15 – and where is it?

Student: Skilling 191.

Instructor (Brad Osgood):Skilling 191. Okay. All right. Let me call your attention to the various announcements on the board, some of which you have seen before. So the final exam is a week from yesterday, next Thursday [inaudible] 11:30 in the morning, Dinkelspiel Auditorium. Be there or be dead. Now, I know there are a couple people who have conflicts, so I'm working out some times with them. Once again, it's open notes, open book. We'll provide the blue books. We'll provide the formula sheets. You provide the knowledge, and the answer, and the correct answers. Please make sure your answers are correct. It makes the exams much easier to grade. Let's see. Any questions on that? Watch the website for postings of various things. I'm not sure if the solutions for last year's final are posted yet or not. Actually – you don't have them? Anyway, if they're – I posted the final from last year. I'm also gonna post the solutions if it hasn't been done yet, so that'll be up there, and there'll be other things, too. I think I owe you some homework solutions and things like that that will be up, and other sorts of announcements. So watch the webpage. Watch your e-mail for other exciting announcements. There is a review session today – the usual review session, I guess, at 4:15 in Skilling 191, and also the TAs will have the regular office hours next week. I'm gonna have to check on my office hours because I know I have some meetings and things like that that I have to go to, so I have to let you know on that. So again, watch the webpage for other sorts of announcements.

Also, you should have gotten an announcement from the registrar's office that the online teaching evaluation form is open, has been open, and will be open through next week, I guess, so I urge you please to go to whatever the appropriate site is on Axess. Don't leave this room. Where do you think you're going? Oh, too late. Go to the – sit down. All right. Go to the appropriate website on Axess and please fill out your teaching evaluation. Now, every time I teach a course at the end of the quarter, as a public service to the students in the class, I offer a few tips on how to fill out the teaching evaluations because I know you're very busy, and sometimes stressed, and sometimes are stuck for something to say, so let me give you some sample adjectives you might wanna use, like for instance, brilliant, dazzling, never have I had a richer more fulfilling intellectual experience, and of course, close-up please, handsome. All right. Okay. Dolly back. So now I wanna finish things up with a treatment of inverting the Radon transform, so that's how I wanna finish things up. And again, I'm sorry that we don't have a little bit more time to do a little bit more detail and go into the finer points of this because it's really – and show you actually some of the ways that it's implemented because it's interesting, but there are plenty of ways of following this up, certainly in courses on medical imaging and so on with any – and as you know, I'm sure we have a very active medical imaging group in the department.

So this is tomography and inverting the Radon transform. So I'll remind you of the setup, what the Radon transform is, and why this is an interesting problem, and why the question is formulated this way. So the setup is we have this two-dimensional slice full of gunk, 2-D region – let's just put it delicately – of variable density, density mu of X1 X2. And the idea is that if you can find mu, if you can describe – you don't know mu, but it's variable and the question is if you can reconstruct mu then you can tell what's inside. So you pass X-rays through the region, and you measure the integral of mu along various lines. So you're restricting mu to different lines, and that's actually what gets measured. You think of that as the intensity of the X-ray, or related to the intensity of the X-ray as it exits the region. So it enters with a certain intensity that you know. It gets scattered, it gets attenuated by all the junk that's in there, and that's what you measure. And from that, what you're measuring actually is the integral of L along mu. So you restrict mu to a line, you integrate it, and that's what you're measuring. So this is along various lines, along line L through the region. So you measure that.

And the question is if we know this, these values, along all lines, many lines, whatever – along lines through the region, can we get mu? Can you reconstruct mu by knowing those integrals? Now you think about this as a transform question, and already that's not so obvious. That's not an obvious step. As a matter of fact, none of this is obvious. It took I think a lot of insight, intuition, luck, whatever to formulate the problem this way and actually figure out the solution because as I say, when you see how this is solved, it's just gobsmackingly amazing. So you think of this as a transform problem. And what I mean by that is with mu fixed so it's unknown, we have a correspondence of lines to numbers given by this formula. That's a line, and the integral gives you a number. And that correspondence you think of as a transform of mu evaluated on a line L. That is to say – and I write it R because it's the Radon transform – the Radon transform of mu evaluated on a line L is exactly that formula, the integral along L of mu, so it's called the Radon transform.

It was introduced – I actually knew the history of this a little bit more thoroughly, and I cannot recall it now. It was certainly not introduced in the context of X-ray tomography or anything else. It was introduced for purely mathematical reasons, for interesting geometric reasons. The idea was to sort of study the geometry of a region by knowing integrals of sections through it just as a purely mathematical question. I don't think there were any practical implications that were anticipated or attempted certainly at the time it was introduced. So our question is – so you know all these values. All right? You know all these values, and the question is can you invert the transform? Can you find mu, given that you know all the values of its transform? Knowing all values, R mu of L – can we invert – or another way of putting it is can we invert R? Okay. Now, this looks pretty abstract, so to make this tractable I have to introduce coordinates. All right? I wanna write things in coordinates. And what I mean by that is not coordinates on the plane, not something replacing the X1 and X2 coordinates, but I want a coordinate description of the line, and actually more precisely what I want is I want a way of coordinatizing all possible lines. I want a coordinate description of the family of all lines. I wanna write this – I wanna write R mu L in coordinates, so script R.

What I mean by that is I want coordinates that describe family of the lines L in the plane. Now that's actually not outlandish at all. And in fact, you've actually seen possible coordinates when you first learned about writing equations of lines, so let me give you an example, not the example that we're ultimately gonna use, but an example that you're very familiar with, and that's what we call at least in America the slope intercept form for the equation of a line, so e.g. for example, you learned to write – You can write equation of a line as Y equals MX plus B. That's one of the first things you learned. And you know that M is the slope and B is the intercept – B and slope M. Now you can think of the pair M and B as giving the coordinates of a line. If I specify M and I specify B, that determines the line. So as the pairs M and B vary, I describe lines in the plane. So MB gives coordinates for describing a line. It describes a line, gives the line. Let me just say describes the line, the unique line Y equals MX plus B. So that's a set of parameters. That's a set of coordinates that describe the lines in the plane. It's not the best one to use certainly for our problem, and there's a problem with this set of coordinates because it omits – it doesn't give you a description of the vertical lines of infinite slope. So it's not the best – not a good set of coordinates because it'll miss the vertical lines. But certainly, as M and B vary, M can vary between zero and infinity, can equal to zero but not equal to infinity – zero is horizontal lines. Infinity is vertical lines, so that's a problem. And B can vary between minus infinity and infinity. As M and B vary over that range, you're describing all the possible lines through the origin, which should be all the possible lines in the plane, except the vertical lines. So this omits – so here again, minus infinity less than B less than infinity, and M bigger than or equal to zero less than infinity describes all but the vertical lines.

So that's not so good. You can sort of fool around with it a little bit and try to make it better, but for our problem there's a better set of coordinates. It shouldn't surprise you that some set of coordinates to a particular problem than another set of coordinates. Polar coordinates is better suited to describing problems when there's circular symmetry because the equations are simpler in polar coordinates when there's circular symmetry than they might be in Cartesian coordinates. So you wanna choose the coordinates that make the equations as simple as possible, or somehow that reveal the essential structure or symmetry of the problem, and do so in a way that's helpful for the calculation. All right. So for us actually, although it may not be immediately – it may not be evidence from what I've said, when you pass lines through the region, certainly in applications, when you're sending X-rays through, you tend to send X-rays through along parallel lines. So for us, a natural configuration of lines in the problem is something like this. Here's the region, and you may wanna pass through a bunch of – you don't send a single X-ray through, but you send a bunch of X-rays through along parallel lines, all making the same angle but parallel lines. And then you change the angle and send another bunch of parallel lines through, something that looks like this. Okay? So a family of parallel lines.

And is there a coordinate description of the family of all lines that makes it easy to write down lines like that? Well, yes. And it's actually – if you phrase it this way, it's not so unnatural. As a matter of fact, I'd like to think it's natural, although everything looks a little unnatural when you first see it. What is common to these lines? The thing that's common to these lines is they all have the same normal vector, or they make the same angle with the horizontal axis, say something like that. So that's gonna be one of our parameters is the parameter that describes the orientation of the normal vector. As a matter of fact, let's start with a single line and describe a single line in a way that's gonna allow me to easily describe families of lines like this. So for a single line, I could take something like – as a matter of fact, let me take a single line through the origin. As a matter of fact, let me make it look like it goes through the origin. Make a single line through the origin, all right? Like so. And that line is determined by its normal vector, or what is the same thing? It's determined by the angle that the normal vector – I'll call it phi – that the normal vector makes with the X1 axis. So I consider I fixed the two axes X1 and X2, and then the line's determined by phi.

The normal vector which I'm gonna need, so let me right it down – as a matter of fact, I'll take the unit normal vector is just cosine of phi sine of phi. That's a vector of length one that's perpendicular to the line. The unit normal vector is say N is cosine of phi sine of phi, but I don't need two numbers. I don't need the cosine and the sine to describe it. I just need the angle phi. That describes it. And here, what is the range of phi? Phi should vary from zero to pi, and you strip the inequality at least on one of the sides, let's say on the right-hand side. Phi equals zero means the line is vertical. It means it's making an angle of zero with the horizontal axis, so that's straight up. And then as phi moves, the line rotates, and when phi is equal to zero it's vertical. When phi is equal to pi, it's also vertical, but you don't wanna have the same line described twice, so you have a strict inequality on this side. So phi equals zero describes a vertical line. Phi equals pi also describes a vertical line. You don't wanna describe the same line twice, so you have a strict inequality here.

So that's a single parameter that describes all lines through the origin. How do I get other lines? Well, I take a line with the same normal vector – and again, I'm interested actually – I have in mind later on I'm interested in families of parallel lines, so I take a line with the same normal vector, and the other parameter that determines it uniquely is its distance from the origin, so a line not through the origin. The picture would be something like this. It has the same – here's the line through the origin with that normal vector. So here's the same normal vector. Here's the angle phi. But there's also a distance rho from the origin. I'll call it rho. So it's determined by phi and a distance rho from the origin, and that determines the line almost. I'll say a little bit more here. There's a slight subtle point here because I don't just wanna take the distance. How do I distinguish between that line – we draw two lines, and I wanna be able to distinguish between the two. Here's one line, and here's a parallel line. They're the same distance from the origin. Here's a normal vector, and here's a normal vector. The normal vector's unambiguous because I decide that phi goes with – the direction of the normal vector's unambiguous because I decide goes between zero and pi, so if the normal vector – if this line goes that way, the normal vector for this line also goes that way. They're both the same distance from the origin.

How do I distinguish between the two? I don't consider just the distance, but I consider the sine to distance. So I consider the sine distance.

So what I mean by that is rho is positive if I get from the origin to the line in the direction of the normal vector, and rho is negative if I get from zero to the line in the direction opposite the normal vector. Rho is positive if I get from the origin to the line in the direction of the normal vector. So that would be the case here. The normal vector's going that way. I get from the origin to the line by moving in the direction of the normal vector. Here rho is negative because I go the other way. Here rho is positive. Here rho is negative. So rho is just – I'll write it down. Rho is negative. Rho is equal to zero of course if I have a line for the origin. So rho is negative if I move – if I get from origin to a line by moving in the direction opposite to the normal vector, like this case. So rho can vary between minus infinity and infinity, so the range of my parameters, the range of my coordinates for the lines is – and I get all lines this way – is zero less than or equal to phi less than pi and minus infinity less than or equal to – less than rho less than plus infinity. I describe all the lines in the plane by those two coordinates. So a give rho and a given phi tell me a given line, specify a given line.

And you'll notice that actually families of parallel lines are described very nicely in this coordinate system because a family of parallel lines means phi is fixed and rho varies. So a set of lines like this – something like that or at a different angle would be phi fixed, rho varies. Okay? That's nice. Now how do you write the equation of a line? I wanna write the – transforms are set up generally speaking – and we'll see actually. We're gonna bring the Fourier transform into this in a way that's just unbelievable – in Cartesian form, so I wanna write a Cartesian equation of the line in these coordinates – of a particular line. So what is the Cartesian equation of the line specified by a given pair rho, phi? Notice here I have a distance and I have an angle, but these aren't polar coordinates. These aren't polar coordinates in the plane. These are line coordinates. These are coordinates of lines. Well, that's not hard to see, and I will refer you to the notes for a little bit more of a derivation of this. Let me just say what the formula is, and again it's something not actually so different from what you've seen before most likely, but maybe not written in quite the same terminology or quite the same language. It's given by in vector form X dot N equals rho, very simple. N is equal to cosine phi sine phi. That's the Cartesian equation in vector form of a line which is given by a given normal direction, a given unit normal, and a distance rho from the origin, sine distance plus or minus.

And if I write that on coordinates, it would be cosine of phi times X1 plus sine of phi times X2 equals rho. That's the equation. Now there's more magic. And again here, unfortunately I don't have time to derive this for you, and I'm gonna see – I have a set of notes on this, but it gets a little involved. Not so bad, it's quite interesting. It's another example of how things in higher dimensions can be a little richer. The variety of objects you find at higher dimensions can be richer than in one dimension, and this has to do with evaluating a line integral via a delta function. So I wanna use this to evaluate an integral like this – an integral a function of two variables, X1, X2, along a line via a delta function. What you do is you consider – so here I just have to state a fact for you. And then with this, we are set to go, we are good to go for the rest of our trip. You consider – I hate that, like a mathematical weasel word, “consider.” Like what could be more natural in the world to consider the following ridiculously complicated looking object. Consider a delta function along the line. You can consider a delta function along the line described this way, so describe the way I wrote it. Let me write it like this: rho minus X1 cosine phi minus X2 sine phi equals zero. That is to say delta of rho minus those things, X1 cosine phi minus X2 sine phi.

Now you can think of this – so this is so-called line impulse. So you can think of this intuitively as a delta function which is concentrated along the line, so it’s infinite – we don’t think these ways because we’re very sophisticated. It’s a distribution, blah blah blah, but if you wanna think about it easily, think about it as concentrated along the line. So it’s infinite along the line, zero off the line, and it has a special property with respect to integration. So it’s infinite along the line, zero off the line, and the key property is for us is what it does under integration. It’s analogous to the one-dimensional case, but it’s perfectly suited to our problem. So zero off the line like all good deltas, infinity on the line – I can write this shamelessly, right? And for the integral what happens is – I’ll give this a board of its own. This is really a key step because it allow – well, you’ll see. One thing at a time. The line integral of mu along the line – so imagine mu is – when you’re integrating mu along the line, it’s as if you want to concentrate mu along the line or just restrict mu to the line. That’s given by integrating over the whole plane mu against the delta function concentrated on the line. This is mu of X1 X2 integrated against the delta function of the line or the line impulse associated with the – the line impulse for the line. X – rho minus X1 cosine phi minus X2 sine phi VX1 VX2. Now that’s a fact that I’m not gonna prove for you. You have to know a little bit more carefully how line impulses are defined and so on, but it is analogous to how delta functions work in connection with integration if you think of them in terms of integration with the one-dimensional case, or you integrate a function against the delta function – it consecrates the function at a point.

Here the delta function is concentrated along a line. To integrate a function against the delta function concentrated on a line, it sort of drops at one dimension so to speak and integrates the function along the line. You haven’t seen this formula, but it is analogous to the kind of formulas you have seen before, and if you really want, I’ll post a set of notes on line impulses. I haven’t had a chance to incorporate them into the notes because they need a little work, but it’ll give you a rough idea if you really wanna know, but it’s dangerous stuff. Watch this. We are set. I’m gonna invert the Radon transform. It is unbelievable. So again here’s the setup. I’m imagining myself – although you won’t necessarily see it in the calculation, I’m imagining that I’m fixing phi and I’m letting rho vary. You wanna use this to invert the Radon transform mu. So again, mu is this unknown function. I know the values of R of mu because I know the values of the line interval. Remember, R of mu – I’ll write it down one more time. R of mu – I could even write it in – R of mu in coordinates would be – say rho phi would be the integral of the line specified by rho and phi of mu. It’s the integral along the line. And I know how to concentrate that. I know I can write this out. I can write this as – that line integral can be written in turn as a double integral – let me write the whole thing out here.

The integral from minus infinity to infinity of mu X1 X2 delta rho minus X1 cosine phi minus X2 sine phi DX1 DX2. Great, so I've now replaced a simple looking formula by a complicated looking formula, and if you like that, you're gonna love what's gonna happen now. Just follow along for the ride. Enjoy it. Don't even take notes. Just sit back. Relax. The hostesses will be by soon to serve nuts and drinks. Now think about phi fixed to let rho vary. Think effectively about passing a family of parallel lines. So think of phi fixed and rho varying. So think about effectively computing this thing along a whole bunch of parallel lines. So you're computing all these values as a function of rho. So with phi fixed and rho varying, you think about R mu rho phi as a function. Think of – as a function of rho with phi fixed. Now I'm going to before your very eyes take the one-dimensional Fourier transform of this function with respect to rho. Like, of course. Like what else would you do? I will take the 1-D Fourier transform of this – that function with rho varying – with respect to rho. In other words, I wanna computer the integral from minus – let's call it something like this. Let's call it the Fourier transform in the rho variable of R mu rho phi.

Now I need a dual variable, right? Because the Fourier transform always has a dual variable in the space domain and the frequency domain, so let me call it R. So let me write this as the integral from minus infinity to infinity of E to the minus two pi I R rho. Let me write it up here so it's – let me write it bigger. Phi just tags along for the ride. As a matter of fact, I'd better write this a little bit more carefully already. It's hard to write. We're gonna get caught a little bit here in problems of notation and problems of variables. It's the usual thing in a subject that you know – there are problems here with variables. There are problems with naming your variables. I'm gonna take the Fourier transform. Phi is fixed. Rho is varying. I'm gonna take the Fourier transform with respect to Rho, and I need to say what the variable – I need to give the variable in the spatial domain a name, so I'm gonna call it R. So that would be like the integral from minus infinity to infinity E to the minus two pi I R rho, Radon transform of mu at rho and phi integrated with respect to rho. Then what pops out is a function of R and phi because phi just tags along for the ride. So phi is fixed, but that's an additional parameter that's entering into the definition here, but I'm integrating with respect to rho. So what pops out is a function of R. Think of it as a function of R, but then I say that phi is sort of tagging along for the ride.

All right. Now be not afraid. Write this out. Be not afraid. Be of good cheer. The holidays are close. What is this? This is the integral from minus infinity to infinity E to the minus two pi I R rho – oh, of this double integral minus infinity to infinity, minus infinity to infinity of mu X1 X2 delta X – delta rho minus X1 cosine of phi minus X2 sine of phi DX1 DX2, and then – that's the R. And then the whole thing is integrated with respect to rho. That's just rewriting this equation substituting what I had before over there writing the line integral of mu in terms of the integral against the line impulse, the integral against the delta function. Okay? Be not afraid. Be not afraid. Flip the integrals. Do all sorts of unnatural acts. Write this as the integral from minus infinity to infinity, minus infinity to infinity mu of X1 X2 – I wanna bring the other stuff inside – the integral from minus infinity to infinity of E to the minus two pi I R rho times delta of rho minus X1 cosine phi minus X2 sine phi. That inside stuff is integrated with respect to rho, and then the whole thing is integrated with respect to X1 and X2.

So all I did was I sort of flipped the order of integration here. Inside here, first I was integrating with respect to X1 and X2, and then integrating with respect to rho, and there was this complex exponential outside. Now I'm flipping this thing around. I'm bringing this mu X1 X2 outside. I'm putting this with that. I'm integrating with respect to rho, and then the whole thing gets integrated with respect to X1 and X2. Be not afraid. Now look at that inside integral. We've seen integrals like that before. We have. Think about this as an engineer, not as a damn mathematician. As an engineer you're integrating a function against the delta function. Integrating a function against the delta function evaluates the function where the delta – it's the shifted delta function. You're integrating a complex exponential with respect to rho against the shifted delta function rho minus – here's the shift. All right? Here's the shift. So what is that integral on the inside? The integral from minus infinity to infinity of E to the minus two pi I R rho delta rho minus X1 cosine phi minus X2 sine phi D rho is E to the minus two pi I R X1 cosine phi minus X – or plus X2 sine phi. That's it. You're integrating a function in a complex exponential against the delta function integral phi of X times delta – integral of F of X times delta X minus Y is F of Y. You know what I'm talking about here. It's the integral. It's the convolution. It's the integral of the function against the shifted delta function. So that substitutes – the shift, it's rho minus – let me write it like this. Rho minus this, so it substitutes that into the exponent. Okay?

Let me write that out a little bit more. This is E to the minus two pi I R – let me write this X1 times R cosine phi plus X2 times R sine phi. And now, I'm gonna introduce some new variables here. I'm gonna introduce sort of dual variables to X1 and X2. So introduce C1 is equal to R cosine phi. C2 is equal to R sine phi. Now these are not polar coordinates on the C1 C2 plane, but you can sort of think of them that way. [Inaudible] because R is the dual variable to rho when you're doing Fourier transform, and phi is an angle in the X1 X2 plane, so this is sort of – these are not exactly introducing polar coordinates in the C1 C2 plane, although there is an interpretation of them that way. But if I do that then that exponential becomes E to the minus two pi I X1 times C1 plus X2 times C2. Now all this was evaluating this inside integral here, the inside integral here. The integral of E to the minus two pi I R rho times this line impulse. And that emerges to be this complex exponential E to the minus two pi I X1 times C1 plus X2 times C2 when I make this change of variable.

Plug this in to my formula. Plug this in. Plug this into what? Plug that into that. This says the Fourier transform in rho of R – this Radon transform of R mu rho phi is equal to the integral from minus infinity to infinity, the integral from minus infinity to infinity mu X1 X2 – that's what was leftover here – times this integral which I just computed to be E to the minus two pi I X1 C1 plus X2 C2 DX1 DX2. What do you see? You see the two-dimensional Fourier transform of mu. This equals the Fourier transform of mu at C1 C2, 2-D Fourier transform. All right? Let's recap. Fix phi. Let rho vary. Take the 1-D Fourier transform of the corresponding Radon transform. That is take the one-dimensional Fourier transform of this function of rho phi with respect to rho. What does it produce? It produces the two-dimensional Fourier transform of mu. Now in principle, the problem is solved. You know this. You know this expression. You measure these values. This is what you're measuring. You're measuring the Radon transform, that's the line integral of mu along this whole family of lines. As rho is varying and phi is fixed, you are measuring these values. You know this function.

Because you know this function, you can compute in theory its Fourier transform. You can compute its Fourier transform with respect to rho. Now what does that tell you? Computing the Fourier transform of this thing with respect to rho gives you the two-dimensional Fourier transform of mu. That tells you you can find mu by taking the inverse two-dimensional Fourier transform of what you have found. You can now find mu, so you compute once again the Fourier transform in the rho variable of this R mu rho phi. And that gives you a function. Now you have to make the changes of variable and all this jazz. You have to substitute C1 equals R cosine phi and C2 equals R sine phi and all the rest of that stuff. Okay, but what that results in at the end of the day is a function of two variables, C1 and C2, so you make the change of variables. C1 is equal to R cosine of phi. That's known. C2 is equal to R sine of phi. That's known. So this thing results in a function of two variables.

Just wanna make sure you understand exactly why this can solve the problem and in what form. So you compute the one-dimensional Fourier transform of this. You make this change of variables. This results in a function – let's call it G – of C1 and C2. Again, you know this function. You can compute its Fourier transform. You make these changes of variables. That gives you some function that you're calling G of C1 and C2, and you know that that function is nothing other than the two-dimensional Fourier transform of mu. So you know that G of C1 C2 is the two-dimensional Fourier transform of mu of C1 C2. And you have solved your problem because now you take the inverse two-dimensional Fourier transform of this function which you have computed, and that gives you mu. Now get mu is the inverse two-dimensional Fourier transform of G. Let me just write it like that. That recovers mu by knowing all those line integrals. Done. Amazing. It's absolutely incredible. So this is the mathematical basis of CAT scans, of recovering the density function by knowing all the one-dimensional slices. Now again, I set this up this way – although the modality, although the imaging techniques in the sort of whole physics behind it is different for MR, at the end of the day actually for amazing reasons, surprising reasons I think, the same thing holds.

That is the idea of finding the – when you try to do MR imaging, you're confronted with the same sort of problem and it's solved the same way. Ultimately, you compute a two-dimensional inverse Fourier transform of like a one-dimensional Fourier transform of some auxiliary function that you introduced. It comes about in a different way, but it amounts to the same thing. It's just amazing. Now again, none of this was known when Radon introduced this idea way back in I think either the end of the 19th century or the beginning of the 20th century – maybe in the '20s, 1920s. I'm not sure when. I have to look it up again. And then found this absolutely stunning and incredibly important application. I think it's nothing short of amazing. I think it's just absolutely amazing. Now of course, to make this practical and to implement this requires a lot of work.

Everything has to be done discretely. If you're actually gonna calculate this numerically, you have to implement discrete versions of this. There are a lot of computational issues. There are very interesting issues actually with the playoff between Cartesian and polar coordinate representations of the different quantities. There are lots of things that have to be done to actually make this practical, but this is the basis for it. This mathematical derivation, this way of going from the one-dimensional Fourier transform in the one variable to the two-dimensional Fourier transform in the other variables and allowing that to solve the problem is the basis for the entire thing. And I say I think it's just absolutely remarkable.

And I think with that we are done. So I wanna thank you very much for a stimulating quarter. I hope you enjoyed it as much as I did. I think this is wonderful material. I really do think it's every day another miracle almost. And I'm sure you'll find this material as you go on in whatever you direction you go on in your studies. Whether it's electrical engineering or other different fields, you're gonna find the Fourier transform and its techniques used absolutely everywhere. And it has been my pleasure to give you what I hope has been a good introduction to it, so thank you very much.

[End of Audio]

Duration: 50 minutes