

The Fourier Transform and Its Applications - Lecture 12

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, Φ 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 N $D^M X$ to the [inaudible] derivative of Φ 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 Φ 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 differentiable, 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 Φ is in S , then the inverse Fourier transform of the Fourier transform of Φ is equal to Φ , and the same if I go the other direction – that is, the Fourier transform of the inverse Fourier transform of Φ is also equal to Φ .

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 Feynman 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 Feynman 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 Π of X . If I scale it, one over ϵ Π ϵ of X , that function is one from minus ϵ over two to ϵ over two, so again, I'm thinking of ϵ as being small here, and if I multiply by one over ϵ , I'm making it large in the vertical direction, so the height is one over ϵ .

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 ϵ in height, one over ϵ – the area is one. As ϵ 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 ϵ 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]? Its integral is equal to one, the integral from minus infinity to infinity of Π ϵ of X DX is one over ϵ . The integral from minus ϵ over two to ϵ 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 Φ of X , what happens? If I look at the integral from minus infinity to infinity of Π ϵ of X times Φ of X DX , what happens there?

Well, let's take the case where Φ is smooth enough, say, you can actually do it more generally, but just to get a simple idea, imagine expanding Φ in a Taylor series expansion. That is, right this as – this is zero except on the interval from minus ϵ over two to ϵ over two. This is equal to the integral from minus ϵ over two to ϵ over two of Φ of X DX . Imagine writing Φ as – because Π of X is equal to one there and it's equal to zero outside that integral. That's equal to one over ϵ , the integral from minus ϵ over two to ϵ over two.

So you write Φ of X as Φ of zero plus Φ 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. Φ 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 Φ of zero – Φ of zero times the integral of one from minus ϵ over two to ϵ over two one of ϵ , 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 ϵ squared here times one over ϵ . That's gonna give me a term of order ϵ . 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 interval. 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 Φ T of Φ – T operates on Φ – 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 Φ_1 plus T of Φ_2 and T of $\alpha \Phi$ is α times T of Φ – it obeys the principle of superposition. T of $\alpha \Phi$ is α times T of Φ . 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 Φ_n is a sequence of functions which converge to a function Φ , then that implies that if I operate on it with one of these degeneralized functions [inaudible] distributions, that that converges to T of Φ . 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 Φ_n converges to Φ , 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 Φ – this notation is supposed to just indicate some alternate notation for writing T operating on Φ . 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 Φ to produce a number, and the operation is linear. If I take two functions, T of Φ_1 plus T of Φ_2 is T of Φ_1 plus T of Φ_2 , and T of $\alpha \Phi$ is α times T of Φ . 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 bra 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 Φ_N converges to Φ 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 $\Delta \Phi$? Does that imply that Δ paired with the sequence Φ_N converges to Δ paired with Φ ? Well, write out both sides. What is the left-hand side? What is Δ paired with Φ_N ? Δ paired with Φ_N is just Φ_N of zero by definition.

Well, if a sequence of functions Φ_N is converging to Φ , then surely Φ_N of zero is converging to the value of zero. If Φ_N converges to Φ , then surely, Φ_N of zero converges to Φ of zero, and so surely, it must be the case that Δ paired with Φ_N operating on Φ_N converges to Δ operating on Φ . It's a continuous linear functional. I want you to think about this a little bit, and I hope appreciate it, because this mysterious Δ function that was defined by these ridiculous properties – it's zero everywhere except at one point, where it's infinite.

Its 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 Δ . 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 Δ function in your work with Δ 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 $\Delta(x - y) f(y) dy$ is equal to what? $f(x)$. You've probably seen statements that look like that. Δ 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 Δ function.

If this is the Δ function based at zero, Δ applied to Φ 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 Δ 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 Δ of A as a distribution by the formula Δ of A paired with Φ is Φ 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. Δ of A – this defines ΔA . 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 Φ . I say ΔA operating on Φ 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 Δ 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 Δ , so we've gained Δ . 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?

Δ 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 Δ '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 Φ in whatever class you're considering, how do we define a pairing of one and Φ ? Well, it's actually very simple, but it's again maybe not one of those things you would – it takes its cue 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 Φ , 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 Φ somehow to produce a number. The pairing is by integration. That is to say, you give me Φ , I have to tell you how one pairs with Φ , and here's my definition. One pairs with Φ as the integral from minus infinity to infinity of one times Φ 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 Φ is a good enough function, this integral is going to converge. Different values of Φ will give me different numbers. In some sense, I know all about one if I know all about the integrals for different values of Φ . 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 Φ .

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 Φ of X Φ of X DX . Φ of X may be very smooth. Φ of X is not smooth, but the product makes sense and the integral makes sense. That's my definition of how Φ pairs with Φ . That identifies Φ , 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 Φ 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 Φ 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 Φ X times Φ of X DX . The [inaudible] sine of two Φ 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 pairing. F paired with Φ is the integral from minus infinity to infinity F of X Φ of X DX . F operates on Φ 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