

The Fourier Transform and Its Applications - Lecture 14

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 derivatives. 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 ϕ . 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 ϕ , I would write, say, is the integral for minus infinity to infinity of t' prime of x , ϕ of x , dx . 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 ϕ , 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 ϕ of x dx , and that is equal to – well if I integrate by parts, t of x , ϕ of x , evaluated between minus infinity to infinity, minus the integral for minus infinity to infinity of – I put – I take the derivative off of t and I put onto ϕ . ϕ' prime of x dx .

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, ϕ 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 ϕ 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 ϕ prime. All right?

So once again, we start off by saying, if t comes from a function then the pairing of t prime and ϕ is given by an integral, and that integral, in turn, is – can be written in terms of the pairing as minus t paired with ϕ 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 ϕ is a test function then, for any decent class of test functions, ϕ 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 ϕ is minus t paired with ϕ 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 ϕ is minus t operating on the test function ϕ 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 $x > 0$ 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 ϕ certainly makes sense.

The integral for minus infinity to infinity of u of x , ϕ of x . ϕ x makes sense because u just – well, we'll pair that one further step. It's the integral from zero to infinity of ϕ of x , $d x$, and that integral makes sense if ϕ is a nice enough function. The integral exists.

All right, so again, while ϕ is not – while u itself is 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' exists as a distribution. Now you have probably learned, in fact, I wouldn't doubt, you probably learned that u' is a very well known distribution. It's the delta distribution. u' is equal to δ . 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' , 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' must be δ .

Now, of course, there's that thing about the integral of minus infinity to infinity if δ 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' is equal to zero here and u' is equal to zero here and u' 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 δ . 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 δ . 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 πi 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 πi 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 δ , that's what I'm just erasing, signum prime is two δ . So, on the one hand, the Fourier transform of the derivative of the signum function is the Fourier transform of two δ , which is two. Two times the Fourier transform of δ . The Fourier transform of δ 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 πi s times the Fourier transform of signum, okay? So two πi – so put these together. Two πi 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 πi s . That's one over πi s . All right?

That's the correct formula. The Fourier transform of the signum function is one over πi 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 than 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 ϕ 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 ϕ is the integral from minus infinity to infinity t of x , f of x times t of x , ϕ of x , dx , and I just grouped the f with the ϕ . That is, the integral from minus infinity to infinity of t of x times f of x times ϕ of x which is the same thing as t paired with f times ϕ . 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 ϕ ? By definition it's t operating on f times ϕ , okay?

Now here's the caveat. The caveat is this has – this works so long as f is such that f times ϕ is again a test function. So if ϕ is differentiable, then f better be differentiable. If ϕ is rapidly decreasing, then f better be at least such that f times ϕ 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 ϕ has to have the properties that define the class. So this makes sense.

This makes sense only if f times ϕ 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' is $2\pi i s$ 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 $2\pi i s$. 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'' is the Fourier transform of $-2\pi i t'$ 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 $-2\pi i t$. 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 $f(x) \delta(x)$, all right? What is $f(x) \delta(x)$? Well $f(x) \delta(x)$ paired with ϕ , by definition is $f(0) \phi(0)$, $f(x) \delta(x)$ paired with ϕ , by definition, is $\phi(0)$ paired with $f(x) \delta(x)$. That's the definition of how a function times a distribution pairs with a test function, but $\delta(x)$ paired with $f(x) \phi(x)$ is, by definition of a delta function of a delta distribution,

this is, by definition, f of zero times ϕ of zero. It's the product f times ϕ 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 δ paired with ϕ . F of zero times δ is just a number times δ , so that makes sense. There's no special definitions required there. F of zero times ϕ of zero is the same thing as the pairing of f of zero times δ paired with ϕ . Where do we start, where do we finish?

We started with f δ paired with ϕ is the same thing as f of zero times δ paired with ϕ . What is the conclusion? The conclusion is that f times δ is f of zero times δ , and a little bit more generally – and I will – a little more generally, f times δ_a , the shifted delta function, is, as you might imagine, f of a times δ 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 δ_a is f of a times δ_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 dx 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 the convolution theorem to hold – if you want the convolution theorem to hold for our distributions, then you'd want to be able 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 x s 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 b be delta 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 times x, what I want is a formula for delta of a times x in terms of delta, actually. So I'm looking – I wanna look at delta of a times 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 times x 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