

## The Fourier Transform and Its Applications - Lecture 09

**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 turbidity. 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 turbidity 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 turbidity 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 turbidity 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 turbidity. 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 turbidity, 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 turbidity signal, whatever you want to call that,  $t$ . This is Fourier transform of the turbidity. 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 turbidity 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  $\infty$  from new sub  $c$ , and one down here out to  $-\infty$  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 turbidity, 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  $2\pi i$  times the Fourier transform of the original function  $f$  of  $s$ . Let me put my variables in here. The Fourier transform of  $f'$  at  $s$  is  $2\pi i$  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  $(2\pi i)^n$ , ordinary power, this is the  $n$ th 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  $f'$  at  $s$ . And as you've heard me say many times, you have no recourse here other than to repeat to the definition. The definition of the Fourier transform is the integral from  $-8$  to  $8$  of  $e^{-2\pi i s t}$ ,  $f'$  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  $e^{-2\pi i s t}$ , and  $dv$  equals  $f'$  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  $e^{-2\pi i s t}$  evaluated between  $-8$  to  $8$  minus the integral of  $v du$ . So that's minus the integral from  $-8$  to  $8$ ,  $f$  of  $t$  times  $2\pi i s$ ,  $e^{-2\pi i s t}$ ,  $dt$ . Now, this term is, by our assumption, that the function tends to zero  $\pm 8$ . This term is gone, the boundary terms are gone, and all that remains is the integral. But the integral, that  $2\pi i s$ , the minus

signs cancel, minus a minus, and that  $2\pi$  comes out of the integral as a constant when I'm integrating with respect to  $t$ . So this is  $2\pi$  times the integral from  $-8$  to  $8$ ,  $e$  to the minus  $2\pi st$ ,  $f$  of  $t$ ,  $dt$ . And so that's just  $2\pi$  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  $u$  of  $t$  is equal to one-half  $u_{xx}$ , 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  $2\pi st$ , 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  $2\pi st$ ,  $u$  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  $2\pi sx$ .

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  $2\pi$  squared. So the Fourier transform of  $u_{xx}$ , again, with respect to the spacial variable, is gonna give me  $2\pi$  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  $4\pi$  squared is minus one, so it's minus  $4\pi$  squared,  $s$  squared, capital  $U$  of  $st$ .

So plug into the heat equation, there it is. Plug into the heat equation  $u_t$  equals one-half  $u_{xx}$  transforms to  $\frac{d}{dt}$  of  $U$  is equal to minus  $2\pi^2 s^2$  times  $U$ . 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  $\frac{d}{dt}$  of  $U$  is equal to a constant times  $U$ . What is the solution? The solution is  $U$  of  $t$  is equal to  $U$  of  $s=0$ , the initial condition, times  $e$  to the minus  $2\pi^2 s^2 t$ . Well, anybody could solve that equation, your little brother and sister can solve that equation.

Now, what is  $U$  of  $s=0$ ? That is the Fourier transform of the initial condition. That's the integral from  $-8$  to  $8$  of  $U_x$  of  $s=0$ ,  $e$  to the minus  $2\pi i s x$ ,  $dx$ .  $U_x$  of  $s=0$  is the initial temperature distribution  $f$ . So that's the integral from  $-8$  to  $8$ ,  $f$  of  $x$  times  $e$  to the minus  $2\pi i s x$ . 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  $s x$  is equal to  $F$  of  $s$  times  $e$  to the minus  $2\pi^2 s^2 t$ . 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  $2\pi t$  times  $e$  to the minus  $x^2$  over  $2t$ . 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  $2\pi^2 s^2 t$ . 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  $x t$  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  $2\pi t$ ,  $e$  to the minus  $2\pi^2 s^2 t$ ,  $e$  to the minus  $x^2$  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 heat 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