

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

PROFESSOR: in the last lecture, we discussed sinusoidal and real and complex exponential signals both for continuous time and discrete time. And those signals will form very important building blocks when we return to a discussion of Fourier analysis in a later lecture.

In today's lecture, I'd like to introduce some additional basic signals, specifically the unit step and unit impulse signal. Let's begin with discrete time and the discrete-time unit step and unit impulse.

The discrete-time unit step is a sequence as I've indicated here, specifically a sequence which is 0 for negative values of its argument, and equal to 1 for positive values of its argument and 0. So mathematically, the unit step sequence is 1 for n greater than or equal to 0 and 0 for n less than 0.

The unit impulse sequence, likewise, is defined in a straightforward way. The unit impulse sequence is a sequence which is 0 for all values of its argument except for $n = 0$. So the unit step and unit impulse sequence are defined in a straightforward way mathematically.

And in fact, they are also related to each other in a straightforward way mathematically. Specifically, the unit impulse can be related to the unit step through the relationship that I've indicated here-- delta of n , the unit impulse, equal to a unit step minus the unit step delayed. So mathematically, the relationship is what is referred to as a first difference.

And to see the validity of this expression, we can simply look at the unit step and its delayed version. So here, we show the unit step, $u[n]$. Here, we show the unit step delayed by 1. So it's 0 for n less than or equal to 0.

And clearly, if we subtract the delayed step from the original unit step, everything subtracts out except at $n = 0$, at which point the difference is equal to 1. And so the difference between $u[n]$ and $u[n-1]$ is simply the unit impulse, sometimes incidentally also referred to as the unit sample.

Now, in a similar way, we can express the unit step in terms of the unit impulse. And there are several ways of doing this. One way is through a relationship referred to as a running sum. What I mean by that is the following expression.

If we think of forming the sum from minus infinity up to some value n of a unit impulse or unit sample, then this running sum, in fact, is equal to the unit step. And we can see that in a fairly straightforward way, by simply observing that in this expression, when n is less than 0, there's nothing accumulated in the sum. And we can see that graphically as I've shown here.

So for n less than 0, so we accumulate no terms in the sum. Whereas for n greater than 0, we accumulate 1 non-zero value in the sum, namely the value of the unit sample at $n = 0$. So we have, then, one expression for the relationship between the unit step and the unit sample.

We can also develop another relationship by observing in essence that if we look at the unit step sequence, as I've returned to here, we can, in effect, think of the unit step sequence as a succession of unit impulses, one following another. So if we consider forming a sum of delayed impulses, as I indicate mathematically here, and as I indicate graphically down below, we have an impulse here at $n = 0$ and an impulse here at $n = 1$, an impulse here at $n = 2$, et cetera.

And when we continue to add these up, then what they add up to is the unit step sequence. And so mathematically, then, that would correspond to an impulse at $n = 0$ plus an impulse at $n = 1$ plus an impulse at $n = 2$, et cetera.

Now, in continuous time, we have a very more or less similar situation. We will find it equally useful to talk about a unit step continuous-time signal and a unit impulse continuous-time signal. Let's begin with the continuous-time unit step.

The continuous-time unit step function is graphically indicated as I've shown here. It's a time function which is 0 for t less than 0. And it's 1 for t greater than 0. And so mathematically, what it corresponds to is $u[t]$ defined as a time function which is 0 for t less than 0, 1 for t greater than 0.

Now, an obvious question is, what happens at $t = 0$? And the difficulty here-- which is not a difficulty that arises in the discrete-time case-- is that at $t = 0$, the units step function is in fact discontinuous, which generates a variety of mathematical problems. And one can define the unit step at $t = 0$ in a variety of ways, but the essential point is that the unit step function is discontinuous.

So in effect, what we need to do is think of the unit step function as the limit of a continuous function. And so we can define a function, which I specify here as $u_{\Delta}(t)$ (u sub Δ of t). And $u_{\Delta}(t)$ is a time function which is 0 for t less than 0, linearly increases to time Δ which would correspond to this break point, and then 1 following that. And so we can think then of the discontinuous unit step as the limiting form of $u_{\Delta}(t)$ as Δ goes to 0.

Now, we also want to define a unit impulse function. And it had a fairly straightforward definition in discrete time. In continuous time, things get slightly more difficult.

And to motivate the definition, let me return to the discrete-time definition-- or rather, the discrete-time relationship between the unit step and the unit impulse function. In discrete time, we saw that the unit impulse function is the first difference of the unit step function. Well, similarly in continuous time, we can talk about an impulse function, which is the first derivative of a unit step function.

So the unit impulse, as we want to define it, is the derivative of the unit step. Of course, we just finished discussing the fact that the unit step function is, in fact, discontinuous at $t = 0$. But we can think of its derivative as related to this approximation to the unit step. And specifically, we will think of the continuous-time impulse as the derivative of $u_{\Delta}(t)$ as Δ goes to 0.

So to define the unit impulse, we think of the derivative of this approximation to the unit step and then observe what happens as delta goes to 0. We have, then, the definition of the unit impulse function more or less formally defined as the first derivative of the unit step, or thought of as the limiting form of the derivative of the approximation to the unit step in the limit, as delta, the duration of the discontinuity, goes to 0.

Well, let's look at that. If we think about the derivative of $u_{\delta}(t)$, the derivative, of course, is 0 for t less than 0. It's equal to a constant during this linear slope, and then, it's 0 for t greater than delta. So the derivative of $u_{\delta}(t)$ will then be as I've indicated here. And it's simply a rectangle with a height, which is $1 / \delta$, and a width, which is equal to δ . And observe that no matter what the value of δ is, the area is always equal to 1.

Now, as we let δ go to 0, what happens is that the width of the rectangle gets smaller, the height of the rectangle gets bigger, the area still remains 1. As δ goes to 0, of course, the width goes to 0, and the height goes to infinity. And graphically, we choose to depict that as an arrow, where the arrow indicates the fact that we have an impulse occurring at $t = 0$, and the height of the impulse is used to represent what the area of the impulse is. And in this case, since we took the derivative of a unit step, the height is equal to 1.

So the impulse has 0 width, infinite height, area 1. It's mathematically not terribly comfortable because what we've done is taken the derivative of the unit step, which has a discontinuity at the origin, and there's some mathematical difficulties in doing that. We'll in fact return to another interpretation of the impulse later to emphasize some of the discomfort with the impulse.

I remember something that Sam Mason used to say that his students said about the unit impulse. His definition was the unit impulse is something that's so small every place I can't see it except at one point where it's so big I can't see it. In other words, I can't see it at. Well, accept some informality with the unit impulse function in the continuous time case, and generally, we'll see that we won't get into particular

difficulty.

OK, now the impulse is the derivative of the step. We saw in the discrete-time case that the step could be recovered from the impulse through a running sum. In continuous time, if the impulse is the derivative of a step, we would more or less reasonably expect that the step would be like an integral of the impulse. And indeed, that's true.

In fact, mathematically the relationship is that the unit step function in continuous time is the running integral of the unit impulse function. And so it's the integral from minus infinity up to time t , where t is the argument at which we're examining $u(t)$.

And so, just as with the discrete-time case, if we are looking at t less than 0, there's no area accumulated in the integral. If we're looking at t greater than 0, then we accumulate the area under the impulse.

OK, now we'll shortly be returning to a further discussion on use of impulses and step functions. And in particular, what we'll see is that they provide a very convenient and powerful mechanism for describing a particular class of systems referred to as linear time-invariant systems.

To lead up to that discussion, which will be the principal focus of the next lecture, let's for the remainder of this lecture talk about systems in general and then some properties of systems. And as the lecture proceeds, the specific properties that I want to get to are properties of linearity and time invariance.

So let's first talk about systems in general. And a system in general, in its most general definition, is simply a transformation from an input signal to an output signal. So in a continuous-time case, we would have a continuous-time system, the input, $x[t]$, and the output, $y[t]$. And the box, in essence, is used to denote a transformation from $x[t]$ to $y[t]$. And sometimes, we'll also use a shorthand notation along the lines of indicating that the input, $x[t]$, is transformed to the output, $y[t]$.

Now, we have exactly the same kind of definition in the discrete-time case. In the discrete-time case, of course, the inputs are sequences, and the outputs are

sequences. And our shorthand notation is similar.

Often when we talk about systems, we'll want to talk about interconnections of systems. And we'll see again in later lectures that interconnections become very important and powerful. And in the way of introducing terminology, let me introduce the terminology for a few basic and important interconnections of systems.

The first is what's referred to as a cascade of systems, or sometimes as a series interconnection of systems. And putting two systems in cascade, as I've indicated here, means taking the output of one system-- let's say system 1-- and using that as the input to the second system, which I've denoted as system 2. So in this cascade with system 1 first and system 2 second, we have the output of system 1 going into the input of system 2.

Now, we could, of course, take those two systems and cascade them in the reverse order. Namely, take the output of system 2 and put it into the input of system 1. And I've indicated that here. And we have system 2 first in the cascade followed by system 1.

It's important to keep in mind that in general, for general systems, the order in which you cascade the systems is very important. And in fact, the overall system transformation will be different depending on which system came first and which system came second.

For example, if system 1 was, let's say, a system for which the output doubles the input, and system 2, the output is the square root of the input, clearly doubling first and then taking the square root is different than taking the square root first and then doubling.

Now, kind of amazingly, what we'll see again when we get to this issue of linearity and time invariance-- which are a class of systems that we'll focus in on-- somewhat amazingly, it turns out that for that specific class of systems, the overall system transformation is independent of the order in which the systems are cascaded. And we'll see that, of course, in more detail later on.

All right, well, that's a cascade or a *series* interconnection. Let's look at another interconnection, which is a *parallel* interconnection of systems. And here, what's meant by the interconnection is that the input is fed simultaneously into system 1 and into system 2. And then, the outputs of these two systems are added to give the overall system output.

Now, in this particular case, in contrast to a cascade, no matter what the systems are, it turns out that a parallel combination, the order in which they're put in parallel is-- the overall transformation is independent of the order. And of course, that follows from the fact that we're simply adding up outputs. And the outputs can be added in any order because of the fact that addition doesn't care in which order you're adding things up.

OK, so that's the parallel interconnection. And let's look at one more interconnection, which is what's referred to as a *feedback* interconnection. And this, again, is an interconnection that will become a very important topic much later in the course. But let me just indicate at this point what I mean by it.

What a feedback interconnection means is that we have one system, system 1, with an input and an output. The output of system 1 is the output of the overall system. And that output is fed into system 2. The output of system 2 is then added to the input to the overall system.

So in essence, what happens in a feedback interconnection is we have one system. The output of that system is fed back through system 2, added to the overall input. And then, that sum is what forms the input of system 1. And we'll see that there are lots of uses for feedback and, of course, also lots of ways that feedback gets in the way. In fact, probably some of you are already familiar with some of the issues in feedback-- for example, in audio systems or whatever. But this will be a topic that we'll devote a considerable amount of time to later in the course.

And then, of course, there are lots of other interconnections of systems. And as the course progresses, we'll see lots of ways in which systems get interconnected both in series and in parallel and feedback interconnections, et cetera, to achieve a wide

variety of things.

Now, what I've said so far relates to systems in general. And you can't say much about systems when you try to treat them in their most general form. So it's useful and important to focus in on properties that a system may or may not have.

So what I'd like to do now is turn our attention to system properties. And we'll define a number of them. Some of them, we'll want to impose on a system. Some of them, we may not want to impose on a system. But as things progress, we'll tend to find it useful to ask whether a system does or doesn't have a certain property.

Well, let's begin with a property which I refer to here as *memoryless*. And what I mean by a system being memoryless is that the output at any given time, t_0 , depends only on the input at the same time. And so what I'm suggesting is that a memoryless system is one for which the output at a specific time depends only on the input at that time. And that statement is true, or that definition applies, both for continuous time, as I've indicated here, and also for discrete time, as I've indicated here.

And so we have a similar definition in discrete time, that the system is memoryless if the output at any given time depends only on the input at that time. Well, I have a number of examples here. Let's look at the first example in which the output is the square of the input. And that is a possible system either in continuous time or discrete time.

And that system, as you would expect, is commonly referred to a *squarer*. And since the square of a signal at any time depends only on the value of the signal at that time, clearly a squarer is a memoryless system. So in fact, we can indicate here that this system is memoryless.

Another important system is what is referred to as an integrator. The output is equal to the integral of the input. Here, as I've indicated it, it's not just integrating $x(t)$, but it's integrating the square of $x(t)$. And whether we square before we integrate or not, the essential point is that since we're integrating the input, the value of the output at

any time is an accumulation of past history of the input.

Well, an accumulation of past history in essence implies memory. To get the output at a given time requires the input over an interval, specifically for longer than that time. So this system, in fact, is not memoryless.

And now, I have a third system defined here. The third system is a system, a discrete-time system, in which the output is equal to the input but not quite. The output at some time-- let's say for example at $n = 0$ -- is equal to the input at one time sample, or instant, or value of the index before that. And so this, in fact, is a system for which the output is simply the input delayed or shifted by one sample. So this system is referred to as a unit delay.

And now, the question is, is that system memoryless? Well, the output depends only on the input-- the output at any instant depends only on the input at one instant. But since it depends on an instant *prior* to the time at which we're looking, or *different* than the time at which we're looking, it violates the definition of memoryless that we introduced. And so, in fact, the unit delay is a system that has memory, and so let's indicate that here. So this is not a memoryless system because of the fact that there is 1 unit of delay. And in essence, delay requires memory.

OK, so that's the issue of memory and memoryless systems. Let's now turn to another property which a system may or may not have, the property referred to as *invertibility*.

Now, essentially what invertibility means is that given the output of the system, you can figure out uniquely what the input was. That's one definition for invertibility. Said another way, invertibility means that given the output, there's only one input that could have caused it. That's another common definition for invertibility.

Another way of looking at it is, in fact, to look at it in the context of a cascade of systems. So let's consider a system. And here is a system which I refer to as system A. And this could be continuous-time or discrete-time. It has an input, $x_1(t)$ or $x_1[n]$, depending on whether it's continuous-time or discrete-time that we're talking

about, and an associated output. And here, we have system B with its associated input and output.

And now, let's put these two systems in cascade. So we'll take the output of system 1 and feed it into the input of system 2-- or system B. So the output of system A goes into the input of system B. And if system A is invertible and system B is its inverse, then the consequence is that the output of system B is equal to the input of system A.

Now, I know there are a lot of inputs and outputs and inverses in there. But essentially, what we mean by what I've just said is that if we have system A and it's invertible, and if we cascade it with its inverse, system B, then the overall cascade of these two systems is simply what's referred to as the identity system.

And the identity system is simply a system which if you put a signal into it, you get the same signal out of it. In other words, the overall system is no transformation at all. And clearly, of course, for a system to be able to be cascaded with another system to generate the identity system requires that the first system, system A, be invertible.

So let's look at some examples. If we had system A as I've indicated here, where now the output is the running integral of the input-- and remember that we saw the running integral when we talked about the relationship between steps and impulses-- if this happened to be an impulse, then the output would be a step. This system is referred to, of course, as an *integrator* since the output is the running integral of the input.

And the integrator, in fact, is an invertible system. And its inverse is a system for which the output is the derivative of the input. So the inverse of system A, if system A is an integrator, is a system for which the output is equal to the derivative of the input which, not surprisingly, is referred to as a *differentiator*. So an integrator is invertible. Its inverse is a differentiator.

What you might want to think about is the question of whether a differentiator is

invertible. Now, to answer that, what you would ask yourself is, if you always knew what the derivative of the signal is, would you necessarily know what the signal was? In other words, if you have a differentiator and you have the output of the differentiator, could you always figure out what the input was? If you could, the system would be invertible. If you couldn't, the system would not be invertible.

So you might just want to think about that. I guess I won't tell you right now. But I'm sure that you'll think about that more, particularly with the guidance of the video manual.

OK, well let's look at one last system. I've indicated here a system for which the output is related to the input through this curve. And what I mean by this curve, which wasn't done quite as well as I might have, is that the output is equal to the square of the input.

So this system is our squarer as we talked about before. We saw previously, or discussed the fact previously, that the squarer is a memoryless system. And now the question is, is a squarer an invertible system? Well, the question then is, if you're given the square of a signal, can you figure out what the signal is? And as I'm sure you've already suspected, the answer to that is no, because obviously if the signal was negative or positive, you wouldn't be able to figure that out after you've squared. So in fact, the squarer is not invertible.

All right, so we've introduced several properties. And by the way, as we've gone through it, also introduced some systems that will turn out to be useful and important systems. And now, let's continue with some other properties.

A property that we'll find useful to make reference to, from time to time, and will, in fact, play a fairly important role in a variety of discussions during the course, is a property which is referred to as *causality*. Now, in essence what causality means is the following.

A system is set to be causal if, as one way of saying it, it only responds when you kick it. Is another way of saying it, its response at any time only depends on values

of the input prior to that time.

So a causal system, both continuous time and discrete time, is defined sometimes as a system which has the property that the output at any time depends only on the input prior or equal to that time. Essentially what we're saying is that the system can't anticipate future inputs. And so that, in fact, is another possible definition for causality, that the system is causal if it can't anticipate future inputs.

Finally, an alternative way of saying it mathematically is as follows. If I have two signals, $x_1(t)$ and $x_2(t)$, with their associated outputs, $y_1(t)$ and $y_2(t)$, then a system is said to be causal if and only if it has the property that if those two inputs are identical up until some time, then the outputs are identical up until the same time.

So if we have two signals that are exactly the same up until some time and perhaps do something different later on, causality requires that the outputs not anticipate the fact that those inputs are at some future time going to do something different. And that, in fact, is the most useful mathematical definition of causality. And of course, I've written that here for continuous time. And the same definition of causality also applies for discrete time.

OK, well let's look at an example. Let's take an example which is a system which is the following discrete-time system. The output at any given time is the sum of $x[n]$, $x[n]$ delayed, and $x[n]$ anticipated. And this is, in fact, a system that's very useful and referred to as a *moving average*.

And so if we think of a moving average, if we have here a sequence, $x[n]$ and $x[n]$ with other values going off in both directions, for any value, n_0 , at which we're computing the output-- and this is $y[n]$, the output-- we take $x[n_0]$, $x[n_0-1]$, and $x[n_0+1]$. And so to form that moving average, we would take these three values and combine them together, adding them and then dividing by 3 to get that.

Well, is the system causal? One way to answer that is to determine whether the output at any given time depends on future values of the input. And clearly, if you

look at this, what you see is that the output at time n_0 depends both on past values and on future values. As opposed to another moving average, which I've indicated here, where I simply shifted the values that I combined together.

And in this case, because of the way in which I picked the values, the output depends on the value at n_0 , the value at $n_0 - 1$, and the value at $n_0 - 2$. So $y[n_0]$ would depend here on $x[n_0]$, $x[n_0-1]$, and $x[n_0-2]$. And so in this case, the system is not causal. And in this case, the system is causal.

All right, now let's turn to another system property, the property referred to as *stability*. Now, there are lots of definitions of stability, and some of them get very mathematical and formal. But we've chosen, and what we'll use as our definition of stability, is what's called bounded-input bounded-output stability. And essentially, the definition is that a system is stable if and only if for every bounded input, the output is bounded.

So the notion is if you have a system and the input never gets above some finite value, then stability requires that the output also stay within some bounded values. And I'm sure that stability and instability are things that you're kind of informally familiar with. Let me just emphasize the point with something that I borrowed actually from my son with some reluctance on his part.

If we, for example, take a system like this, which is in essence a pendulum, this system as I'm holding it here is stable because if I put in a bounded input, which is a displacement, the output, which is the movement of it, remains bounded.

Now on the other hand, if I put the system like this, which is, in fact, what's referred to as an inverted pendulum, although we could conceivably get this to balance, just a slight displacement because of the fact that the pendulum is inverted, a slight displacement and the output becomes unbounded.

Now, an interesting thing with the inverted pendulum, by the way, which I'm sure all of you, if you were anything like me, were intrigued with as a kid, was the notion that you could take an inverted pendulum and in effect turn it back into a stable system

by using what I'm doing right now, which is feedback. What I've done in that case is I've stabilized the system by using feedback, visual feedback, from my eye to my hand.

And in fact, one of the very important things that we'll see about feedback when we talk about feedback systems much later in the course is that one of their very important applications is in stabilizing unstable systems. By the way, one of their problems is that if not used correctly, it can destabilize stable systems.

OK, well let's continue on with our property of stability. I have here, again, the example of an integrator. And as I indicate here, if we have an integrator and we put a step function into it or a step signal into it, the output is what's referred to as a *ramp signal*. It linearly increases.

Now, the question is, is a ramp unbounded? The input is bounded. The step input is bounded. The ramp is unbounded because if you try to establish any bound on it, you can always go out far enough in time so that the output will exceed that bound. So in fact, the integrator is not stable.

OK, now finally, I'd like to turn to two properties that we'll make considerable use of as the course goes on, the properties of time invariance and linearity. Time invariance, in essence, says that the system doesn't really care what you call the origin. In other words, it says if you take the input and you shift it in time, all that you've done is taken the output and shifted it in time by the same amount.

Somewhat more formally as I've indicated here, if in continuous time we have an input, $x(t)$, which generate an output, $y(t)$, then time invariance requires that if the input is shifted by any amount of time, the output is shifted by the same amount of time. And exactly the same applies in discrete time.

For example, if we have a system which is a system I've shown here, which by the way is the system that we talked about previously to go from a step sequence-- I'm sorry, from an impulse sequence to a step sequence. We called it a running sum. And actually, what it's also often called is an *accumulator*. What it does is

accumulate past values of the input.

Well, is the accumulator time-invariant? The best way to establish that is to work through the equations and verify that it either does or doesn't satisfy the formal definition of time invariance. Informally, if you think about it, it makes intuitive sense that the accumulator is time-invariant because if you're accumulating values and if you delay the values that you're putting into the accumulator, then the associated values that come out will be delayed by the same amount. The accumulator doesn't care really if you shift the input. It'll just simply shift the associated output.

But more generally, if you're trying to test time invariance, it's important to return to the definition. And that's what you're required to do in the examples in the video manual.

OK, well, I indicate another example. We had the example of an accumulator. Here's another example which, in fact, as we'll see later is a system which is a modulator. The output is the input, modulated. And although you might think at first that this system is time invariant, in fact it is not, because the input shifted generates an output which is the input shifted times the same modulation function.

Whereas if we were to take the output of the system, we have $x(t)$ is the input, then what that would correspond to is $\sin(t-t_0) * x(t-t_0)$. And since these two are not equal, this system is not time invariant. And this is an example that often causes a slight amount of difficulty because it seems like when you look at it ought to be. And so I strongly encourage you, in the context of working problems in the manual, that you think very carefully about this and at least believe that what I told you is the right answer.

OK, now the final property that I want to introduce today is the property of *linearity*. And linearity is defined in a manner similar for continuous time and discrete time. And what it says is that if we have some inputs with associated outputs, let's say $x_1(t)$ and $x_2(t)$, then a system is linear if it has the property that the output to a linear combination of those inputs is the same linear combination of the associated outputs.

And so that's what I've indicated here, that if we now put into the system a linear combination of those inputs, then for linearity, we require that the output is the same linear combination. And exactly the same applies in discrete time. And you can show from this definition that if a system is linear with two inputs, then it's linear in terms of an arbitrary number of inputs.

I have a number of examples. And these are examples that, again, I ask you to think about as you look at the video manual. Just to suggest the answer-- well, not to suggest but to tell you the answer, the integrator as we have here is linear.

This system in which the output is double the input plus a constant, you would kind of think it's linear because it's a straight line. But one has to be careful. And in fact, as it turns out, this is not linear. There is a qualifier attached to it because it has a property referred to as incrementally linear, which is discussed somewhat more in the text. And finally, we have a system which is the squarer that I've indicated again here. And squaring it is definitely not a linear operation.

OK, so what we've done, then, is to introduce a number of properties of systems. And we've also, by the way, as I've stressed previously, as we've gone along introduced also a number of important and useful systems, like the accumulator, the integrator, the differentiator, et cetera. What we'll see is that the properties of linearity and time invariance in particular become central and important properties.

And in the next lecture, what we'll show is that with systems that are linear and time-invariant, the use of the impulse function, both in continuous time and discrete time, provides an extraordinarily important and useful mechanism for characterizing those systems. Thank you.