

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.

[MUSIC PLAYING]

ALAN V. OPPENHEIM: In the last lecture, we introduced the class of discrete time systems, and in particular, imposed the conditions of linearity first of all, and second, the property or constraint of shift invariance. And those constraints led us to the convolution sum representation.

In today's lecture, there are several issues that I'd like to focus on. The first is the introduction of two additional constraints that it's sometimes useful to impose, or at least consider, for discrete time systems-- namely the constraints or conditions of causality and stability.

Second of all, I would like to talk about a particular class, or at least introduce a particular class of linear shift invariant systems, namely those representable by linear constant coefficient difference equations. And finally, I'd like to introduce a representation of linear shift invariant systems that is an alternative to the convolution sum representation, and in particular, that representation corresponds to the representation in terms of a frequency response.

Well, let's begin with the notions of stability and causality, reminding you, first of all, that, as we talked about last time, we can consider a general system-- inputs and outputs are sequences-- and in general, the system just simply corresponds to some transformation from the input sequence to the output sequence.

When we impose the conditions of linearity and shift invariance, both conditions, then we can express the output sequence in this form where h of n is the response of the system to a unit sample. And this can also be rearranged in the form that I've indicated here, essentially interchanging the role of the unit sample response and the input. The sum expressed in either of these two forms, we referred to last time as the convolution sum.

So the convolution sum comes out of-- or is a consequence of-- linearity and shift invariance.

Two additional conditions are stability and causality. Stability of a system, in general, corresponds to the statement that if the input sequence is bounded-- in other words, if x of n , essentially if x of n is finite for all n , including as n goes to infinity, then a system is said to be stable if the output sequence, y of n , is also bounded. In other words, the magnitude of y of n is finite for all n .

If that's true for any bounded input sequence that the output sequence is bounded, the system is said to be stable. Now, that's a statement that applies to general discrete time systems. For the specific class of systems that we'll be dealing with in this set of lessons, namely the class of linear shift invariant systems, you can show that an equivalent statement of stability-- or a necessary and sufficient condition for this to be true-- is simply that the unit sample response of the system be absolutely summable. In other words, that the sum of the magnitudes of the values in the unit sample response are finite.

For example, if we had a unit sample response which was 2^n times a unit step, 2^n for $n \geq 0$, of course, as n increases-- as n goes from 0 to infinity-- 2^n grows exponentially, in fact. And so, obviously, this is not absolutely summable. So this is an example of a system that would be unstable.

Whereas, if $h[n]$ were a $(1/2)^n$ times $u[n]$, so that for $n \geq 0$ -- and less than 0, of course, both of these are 0-- for $n \geq 0$, this were decaying exponentially, if you sum up these values from 0 to infinity, it converges to a finite number. So that this, in fact, would correspond to a stable system.

Now, we have the option, of course, of talking about unstable systems or stable systems. Generally, it is true that stability is a condition on a system that it's useful to impose. In other words, generally, we would like our systems to be stable, although there are actually some cases where unstable systems are useful to talk about.

The second property or condition that it is useful sometimes to consider, is the condition of causality. And causality, first of all, for a general system, is a statement basically that the system doesn't respond before you kick it. In other words, if we have an input, $x[n]$, the output, $y[n]$, for some value of n -- let's say $n = 1$ -- depends only on $x[n]$ for previous values of n .

So for any $n \geq 1$, the statement is that $y[n]$ only depends on $x[n]$ for previous values of n . In other words, the system can't anticipate the values that are going to be coming in.

For a linear shift invariant system, we can show that a necessary and sufficient condition for causality is that the unit sample response of the system be 0 for $n < 0$. In other words, if the unit sample response of the system is 0 for $n < 0$, the system is guaranteed to be causal. If it's not 0 for $n < 0$, then the system is guaranteed not to be causal.

Just for example, if we had a unit sample response which was 2^{-n} for $n \geq 0$ and 2^n for $n < 0$, this, of course, would correspond to a non-causal system, since the unit sample response has non-zero values for negative values of n .

And we could also examine stability. It would turn out that this corresponds to a stable system, although in the previous example, we had talked about 2^n for n positive corresponds to an unstable system. The point is that if you look at 2^{-n} as the index n runs negative, then that is-- as n runs negative, it's an exponential that's decaying in negative values of n . So then, in fact, this is absolutely summable-- that makes it stable, but it's obviously non-causal.

Now, I want to stress-- it's a very important point to stress-- that causality is, of course, a useful thing to inquire about about a system, it's useful to ask, is the system causal or is it not causal? But generally, it is not necessarily true that causality is a condition that we'll want to impose on the system. There are many examples of useful, non-causal systems, and in many instances, we'll want to talk about systems which are non-causal.

So again, it's useful to inquire about whether a system is causal or not causal, but it is not generally useful to constrain ourselves to talk only about causal systems.

The story is somewhat different for stability in the sense that an unstable system is somewhat less useful than a non-causal system.

So these are two additional conditions, or properties, that we'll sometimes want to inquire about when we talk about a system.

In general, for linear shift invariant systems, there is a wide latitude in terms of the description of those systems. And that is a similar type of statement that applies in the continuous time case, also. Just as in the continuous time case, it is useful to concentrate, in many cases, on systems that can be implemented with r 's, l 's, and c 's-- and consequently are representable by linear constant coefficient differential equations. In the discrete time case, it's often useful to concentrate on systems that are describable by linear constant coefficient difference equations.

So I would like to just briefly introduce the class of systems which are representable by linear constant coefficient difference equations. And the discussion in this lecture is only a brief

introduction, and we'll, in fact, be returning to this class of systems several times over this set of lessons.

We refer to an N th order-- linear constant coefficient difference equation-- as being in the form that I've indicated here. And what it consists of is a linear combination of the delayed output sequences equal to a linear combination of delayed input sequences. A differential equation, of course, in continuous time involves linear combinations of derivatives. The corresponding situation in the discrete time case, is a linear combination of differences.

So this is, then, a general N th order, linear-- in the sense that it's a linear combination-- constant coefficient-- meaning that these are constant numbers, as opposed to being functions of n -- difference equation-- meaning it involves differences of the input and output sequence. The order of the difference equation generally is used to refer to the number of delays required in the output sequence. In general, the number of delays in the input sequence, M , does not have to be equal to N . But it's generally convenient to refer to the order as corresponding to the number of delays involved in the output sequence.

For the 0th order difference equation, the solution if it corresponds to a representation of a linear shift and invariant system, the solution is trivial-- very straightforward. Particular, let's examine what this difference equation reduces to, if N is equal to 0. And just for convenience, we'll choose the coefficients to be normalized so that a_0 is equal to 1. Obviously anything I say is straightforwardly generalized for a 0 not equal to 1, but that's just a convenient normalization to impose.

All right, if N is equal to 0, and a_0 is equal to 1, then what this equation looks like is just on the left hand side, y of n , and the right hand side as it is. So we have y of n is equal to this sum. And that looks suspiciously like the convolution sum-- it involves a linear combination only of delayed input sequences.

And so, in fact, this is identical to the convolution sum. If we thought of this, say, as h of r , h of r , then, is equal to b sub r . So the unit sample response corresponding to the 0th order difference equation is just this set of coefficients, and, of course, r runs only from 0 to M , so it's the coefficients b sub n for n between 0 and M . And it's equal to 0, otherwise.

So that's very straightforward. We can pick off the solution, essentially, by recognizing this as the convolution sum. Obviously, another way of obtaining the unit sample response corresponding to this system is simply to plug in a unit sample for x of n . And you'll see, in fact,

that what comes rolling out are these coefficients b_r , or b_n .

So if this is used to describe a linear shift invariant system, the unit sample response corresponds simply to the coefficients in the difference equation.

For n not equal to 0, it's not quite as straightforward as that. First of all, let me point out that for N not equal to 0, the left hand side of this equation involves y of n , y of n minus 1, y of n minus 2, et cetera, up to y of n minus N . We can take all of the terms except the one involving y of n over to the right hand side of the equation, and end up with an equation which expresses y of n as a linear combination of delayed inputs, and a linear combination of past output sequences.

If we rewrite this equation in this form-- and again, just for convenience, we'll choose a 0 equal to 1, simply normalizing that coefficient. Then, the difference equation that we end up with is in the form that I've indicated here, y of n is a linear combination of delayed input sequences minus-- because we brought that from the other side of the equation-- minus a linear combination of past-- of the previous output values.

What that says is that we presumably can iterate this equation, or equivalently, what it says is that the difference equation corresponds to an explicit input/output relationship for the system, because if somehow we could get the equation going, then we can solve for y of n -- we can solve for y of n if we have the right previous values for y of n . And then we can solve for y of n plus 1, y of n plus 2, et cetera. In other words, we can continue to iterate the equation.

What's required to do that are that we have to know the input-- and we assume, of course, that we know that-- and we have to know some previous output values. And to know the previous output values, then, means that there is some additional information that we have to specify, and those we'll generally refer to as the boundary conditions, or the initial conditions.

For example, let's see how this equation would work if we focused on the first order case. So the first order case, N is equal to 1. We have the equation, y of n minus a , y of n minus 1 equals x of n . And let's find the unit sample response for the system-- in other words, choosing x of n equal to a unit sample. And for the initial conditions, let's assume that, for a unit sample input, the output is equal to 0 for n less than 0.

Now, when we do that, we're making a specific assumption about causality-- we're imposing the boundary conditions that the unit sample response has to be 0 for n less than 0-- that's

exactly the necessary and sufficient condition that we need for causality. So basically, we're saying that we're imposing on the solution that we're about to generate, the additional condition of causality.

All right, well now, rewriting this equation by taking this term over to the right hand side, and taking account of the fact that we're considering the input to be a unit sample, we have the output sequences unit sample plus a times the output sequence delayed. And now let's run through an iteration of this equation, obtaining some successive values.

y of minus 1 we can get simply by referring to the initial condition. We stated there y of n is 0 for n less than 0, and so that means, of course, that y of minus 1 is 0. y of 0 is equal to delta of 0, but that's equal to 1, plus a times y of minus 1, which was 0. So y of 0 is equal to 1.

y of 1 is equal to delta of 1, which is 0-- the unit sample is only non-0 at n equals 0. So that 0 plus a times y of 0-- y of 0 is equal to 1-- so y of 1 is equal to a times 1, or a . y of 2, if we follow that through, is equal to a squared.

And if you run through a few more of those, you'll see rather quickly that what we get for the unit sample response, imposing the condition of causality, is that the unit sample response is a^n for n greater than 0. Of course, it's 0 for n less than 0, because of our initial condition. So it's a^n times u of n .

And it's obviously causal, because that's what we imposed. It may or may not be stable, depending on the value of a . In particular, if the magnitude of a is less than 1, then this sequence will be decaying exponentially as n increases. So for the magnitude of a less than 1, this corresponds to a stable system.

Now this isn't the only initial condition that we can impose on the system and still have it correspond to a linear shift invariant system. We could, alternatively, impose a different set of initial conditions, which is the-- or boundary conditions, really, because they're not, for this case, initial conditions-- boundary conditions that state instead of the statement that the system is causal, the statement that the system is totally non-causal.

What I mean is, let's take the same example-- the same difference equation-- let's again choose the input to be a unit sample, but let's impose another condition on the solution, boundary condition. Namely that the unit sample response is 0 for n greater than 0, rather than the boundary condition that says that it's 0 for n less than 0.

In that case, it's convenient to generate the solution iteratively by running the difference equation backwards, in other words, by expressing y of n minus 1, in terms of y of n and x of n -- this is just simply another rearrangement of the difference equation. And with this initial condition, if we look at y of 1-- of course, y of 1 is equal to 0 by virtue of the boundary condition that we've imposed, so this is equal to 0-- y of 0 corresponds to n equal to 1 in this equation. So n equal to 1 here, we have 1 over a times y of 1, which is 0 minus Δ of 1 which is 0 . So y of 0 is again equal to 0 .

y of minus 1 corresponds to n equal to 0, so we have 1 over a times y of 0, which was 0 , minus Δ of 0, which is 1 . So we get minus a minus to the minus 1. y of minus 2, we would substitute n equals minus 1. y of minus 1 we had as minus a to the minus 1. And Δ of minus 1 is 0 , so we have here minus a to the minus 2, and it continues on that way.

And if you generated some more of these, what you'd see rather quickly is that this is of the form minus a to the n times u of minus n minus 1. In other words, it is an exponential to form a to the n , as we found for the causal solution, also. But it's 0 for n greater than 0 , and it's of the form a to the n for n less than 0 .

What that means, in particular, is that if we again inquired as to whether it was stable or unstable, for the magnitude of a less than 1 , what we would find, in this case, is that it's unstable rather than stable.

So what we've seen, then, is that a linear constant coefficient difference equation, by itself, doesn't specify uniquely a system-- it requires a set of initial conditions. And depending on the initial conditions or boundary conditions that are imposed, it may correspond to a causal system, or it may correspond to a non-causal system. And in some situations, we might want it to correspond to either one of those two.

Something that I haven't stated explicitly, but there is some discussion of this in the text, is the fact that it is not for every set of boundary conditions that a linear constant coefficient difference equation corresponds to a linear shift invariant system, but it does in particular for the two sets of boundary conditions that are imposed here.

More generally, what's required of the boundary conditions, so that the difference equation corresponds to a linear shift invariant system, is that the boundary conditions have to be consistent with the statement that if the input, x of n were 0 , 0 for all n , that the output would

also be 0 for all n . And there is some additional discussion of this in the text.

I indicate also that we'll be returning from time to time to further discussion of linear constant coefficient difference equations, and discussing, in particular, other ways of solving this class of equations.

For the remainder of the lecture, I'd like to focus on an alternative representation of linear shift invariant systems, alternative to the time domain, or convolution sum representation, that we dealt with in the last lecture. In particular, the very useful alternative is a description of linear shift invariant systems in terms of its frequency response. In other words, in terms of its response either to sinusoidal excitations, or to complex exponential excitations.

Well, the basic notion behind the frequency response description of linear shift invariant systems is the fact that complex exponentials are eigenfunctions of linear shift invariant systems. What I mean by that is that for linear shift invariant systems, if you put in a complex exponential, you get out a complex exponential. And the only change is in the complex amplitude of the complex exponential-- the functional form is the same, so complex exponential in gives you a complex exponential out.

We can see that rather easily by referring to the convolution sum description of linear shift invariant systems, as I've indicated here. In particular, suppose that we choose an input sequence which is a complex exponential. And let's substitute this into this expression, so that the output is then the sum of $h(k) e^{j\omega(n-k)}$.

This term can be decomposed into a product of two exponentials, $e^{j\omega n}$, and $e^{-j\omega k}$. And since $e^{j\omega n}$ doesn't depend on the index k , we can take that piece outside the sum, and what we're left with is $y(n) = e^{j\omega n}$ times the sum of $h(k) e^{-j\omega k}$, as I've indicated up here.

So we started with a complex exponential sequence going into the system. None of this stuff over here depends on n , so when we sum all that up, that's just a number-- it's a function of ω , depends on what complex frequency we've put in. But it doesn't depend on n . And, in fact, notationally, we'll refer to this as the $H(e^{j\omega})$. So consequently, the output sequence, due to a complex exponential input, is this number, or function of ω , times $e^{j\omega n}$.

The change in the complex amplitude, then, is $H(e^{j\omega})$, but the functional form of

the output is the same as the functional form of the input. And that's what we mean when we say that the complex exponential is an eigenfunction of a linear shift invariant system.

So finally, H of $e^{j\omega n}$ is given by this sum-- I've just changed the index of summation from what I had up here, but obviously that's no important change. And we will refer to this, or to H of $e^{j\omega n}$, as the frequency response of the system.

One of the reasons why the frequency response is important-- of course, one of the facts that you can see right away is that it's easily obtained directly from the unit sample response. One of the reasons why the frequency response is useful is because it's essentially the frequency response that allows us to obtain quite easily the response of the system to sinusoidal excitations. And, as we'll see in later lectures, starting actually with the next lecture, essentially arbitrary sequences can be represented either as linear combinations of complex exponentials, or as linear combinations of sinusoidal sequences.

And so, if we know what the response is to a complex exponential or to sinusoidal sequences, we can, in effect, describe the response to arbitrary sequences. We'll see all of that coming up in the next lecture. But to see how this frequency response relates to the sinusoidal response, the relation essentially pops out from the fact that if we have a sinusoidal excitation, a sinusoidal excitation can be represented as a linear combination of two complex exponentials-- as I've indicated here.

So $\frac{a}{2} e^{j\phi} e^{j\omega_0 n}$, this is one complex exponential with a complex frequency ω_0 , and a complex amplitude, $\frac{a}{2} e^{j\phi}$. Then its complex conjugate term, those two added up give us a complex exponential.

We can find the response to each of these simply from the frequency response. We know that all that happens to a complex exponential is that its complex amplitude gets multiplied by the frequency response. And so this is one complex exponential, and another one.

We're talking about linear systems, so the response of the sum of these two is the sum of the responses. And so if we express the frequency response in a polar form, as I've indicated here, in terms of magnitude and phase, and if you track through what happens when you multiply each of these complex exponentials by the appropriate frequency response-- this one at a frequency ω_0 , this one at a frequency $-\omega_0$ -- and add the terms back together again, the resulting output sequence has a change in real amplitude given by, or dictated by the magnitude of the frequency response, and a change in phase dictated by the

angle of the frequency response. In other words, by this θ of ω .

So the frequency response, when thought of in polar form-- magnitude and angle-- the magnitude represents the change in the real amplitude of a sinusoidal excitation. And the angle, or complex argument, represents the phase shift. And that is exactly analogous, exactly identical to what we're used to in the continuous time case.

Well, let's just look at an example of a simple linear shift invariant system, and the resulting frequency response. Let's return to the first order difference equation that we talked about just a short time ago. We saw that there were two solutions that we could generate for this, depending on whether we assumed that it was a causal or non-causal system. Let's focus on the causal case.

If the system was causal, we generated, essentially, iteratively the solution that the unit sample response is $a^n u[n]$. And let's assume that we're talking about a stable system, so that $|a| < 1$ -- of course, this could be negative and still be stable between minus 1 and 0. I'll assume in the pictures that I draw that a is, in fact, positive, and then also less than 1.

Now, the expression for the frequency response of the system, from what we derived just above, is the sum overall n of $a^n e^{-j\omega n}$. Because of the unit step in here, the limits on the sum change from 0 to infinity-- in other words, all of this is going to be 0 from minus infinity up to 0, because of the unit step.

For $n > 0$, we have $a^n e^{-j\omega n}$, so we have $a^n e^{-j\omega n}$ to the-- surely there's a minus sign there-- $e^{-j\omega n}$. If we sum this, this is just the sum of a geometric series.

In other words, it's of the form, the sum of α^n -- α equals 0 to infinity. Sums of this form are equal to $\frac{1}{1-\alpha}$. And α , in this case, is $a e^{-j\omega}$. So that sum, then, is $\frac{1}{1-a e^{-j\omega}}$.

Well, as we saw-- at least for the sinusoidal response-- it's useful to look at the magnitude and phase of this, in other words, to look at it in polar form. So if we want the magnitude of this frequency response, we can obtain that by multiplying this by its complex conjugate.

Consequently, the magnitude of the frequency response is given by $|H(e^{j\omega})|$ -- the minus sign is conveniently there for us-- multiplied by its complex conjugate, which is $1 - a e^{j\omega}$

a e to the minus-- I'm sorry, $1 / (1 - a^2 \cos^2 \omega)$.

If we multiply these together, then what we obtain is $1 / (1 - a^2 \cos^2 \omega)$. And the phase angle of this is equal to the arctangent of-- if you just simply work this out-- the arctangent of $a \sin \omega$ divided by $1 - a \cos \omega$. I suspect, actually, that because of that algebraic sign error I made, the minus sign down here, actually I think that this comes out with a minus sign. But I won't guarantee that on the spot right now-- you can simply verify that, but I suspect that there should be a minus sign there.

OK, well this, then, represents the phase shift that would be encountered by a sinusoidal input going through the linear shift invariant system. This would represent the square of the magnitude change of a sinusoidal input. And if we were to sketch this, then what we see is that at ω equal to 0, $\cos \omega$ of course at ω equal to 0 is 1. So this is $1 / (1 - a^2)$, or $1 / (1 - a^2 \cos^2 \omega)$. We're assuming that a is between 0 and 1, and so we're indicated that by this value.

When ω is equal to π , $\cos \omega$ is equal to minus 1. This, then, comes out to be $1 / (1 - a^2)$, which for a between 0 and 1 is less than this value. So the frequency response, then, between minus π and plus π , would have the shape that I've indicated here.

Now, what happens if we continue on in frequency-- ω can, of course, run from π to 2π , and on past that. You can verify in a very straightforward way from this expression that, from π to 2π , the frequency response will look exactly like it did from minus π to π . It will, then, look like that.

And for minus π to minus 2π , it will look exactly as it did from π down to 0. So it will look like this. And in fact, it's straightforward to verify for both the magnitude and the phase that both of those are periodic in ω , with a period of 2π . So this is one period, say, from minus π to π , and then another period would start from π to 3π . And it would continue on like that.

So, in fact, if we were to sketch this out over a wider range of ω , we would see this periodically repeating. For this example, that's straightforward to verify.

So there are, then, two properties of the frequency response which I would like to call your attention to in preparation for our discussion next time. One of them is the fact, very important to keep in mind, that the frequency response, as we're talking about it, is a function of a

continuous variable, ω . ω , I hadn't stressed this point previously, but ω is a variable that changes continuously over whatever range we're talking about, as opposed to sequences which are functions, obviously, of a discrete variable.

Here we're talking about a function of a continuous variable, ω . As we saw, for our example, the frequency response is a periodic function of ω . And the period is equal to 2π .

Now, the reason that it's a periodic function of ω is, in fact, somewhat obvious. Suppose that we take a complex exponential, $e^{j\omega n}$, and inquire as to how the complex exponential itself behaves if we change ω over an interval of more than 2π .

Suppose that we replace ω by $\omega + 2\pi k$, and now if we decompose this into a product, we have $e^{j\omega n}$, times $e^{j2\pi k n}$. This is an integer multiple-- this exponent is an integer multiple of 2π . And so this is simply equal to unity.

Now what that says, is that a complex exponential-- once we've varied ω over an interval of 2π , and we go past that, there are no more no new complex exponentials to see. We'll see the same ones over and over and over again.

And consequently, the system response has to be periodic in ω with period 2π , because we're putting in, essentially, the same inputs over and over and over again. This is a point that I'll be mentioning from time to time, and it's, in fact, somewhat important to keep in mind. Also, it's discussed in some detail again in the text.

So these, then, are some properties of the frequency response. There is a generalization of the frequency response, which is, in fact, very important for describing both signals and systems. The generalization is what we'll refer to as the Fourier transform, which plays the identical role in the discrete time case that the Fourier transform did in the continuous time case.

And so in the next lecture, we'll be going on to a discussion of the Fourier transform taking off from the set of ideas that we've developed here, with regard to frequency response. Thank you.

[MUSIC PLAYING]