Thank you.
Good afternoon.
Thank you.
Today we have our first lecture on convolution.
This is a direct follow-up from previous one,
system, particularly linear system,
safety environment linear system.
And we are certainly on schedule,
and office hour-wise,
and we have office hour on the teaching day,
like today, Friday.
But whenever Friday office hour,
you could come to my office and say five minutes earlier,
and I will leave my office five minutes earlier,
because every Friday we have a lab meeting from five o'clock on.
So we just do translation.
The quality will be the same.
That's called the safety environment.
So we just save the office hour a bit earlier,
by about five minutes, something like that.
And I'm not sure how many of you did the preview,
and read the chapter with your hand.
So I know if you, okay.
Okay, very good.
If you read the chapter, then you have a better understanding.
Otherwise, you will be confused in some places.
And you need to do review anyway.
Why not just do a little earlier?
So this will be a best arrangement.
And as you see the second chapter, convolution,
and we talk about three things.
Safety environment linear system.
Then we talk about continuous convolution,
and discrete convolution.
And we keep enhancing the concepts and the scales
from different angles.
And the last time when we talk about linear system,
we already mentioned a little bit
safety environment linear system.
I mentioned safety environment system.
It's a good credibility and a good personality.
So whenever you drop a water drop into the pond,
and you will see the same ripple.
Certainly the ripple will be placed around the location
where the water drop touches the water surface.
And talking about medical imaging,
you will realize the linear system requirement,
safety environment requirements are really needed.
So suppose you do medical imaging study,
and you do it in one hospital or in another hospital.
You do it today or a few days later.
You certainly expect exactly the same finding.
Either you are okay or you got some problem.
So that is nothing but safety environment.
And also in terms of linear system properties,
and we mentioned the additivity and homogeneity.
So suppose some day you got a tumor.
At place A, you do examination.
You see the tumor A in your chest or in your abdomen.
And the other day say you got another problem.
At location B, you will see location B.
You see the pathological features.
And if you grow two tumors together,
and certainly after the scan,

you expect to see both of them.
So this is about additivity.
Homogeneity, so your tumor,
just one tumor today and maybe one cubic centimeter.
And a few months later it becomes two cubic centimeter.
Then the imaging results, imaging study,
should give you the same volume,
proportionally scaled.
So you have the exact reporting reflecting the truth.
So the linear system requirement,
additivity, homogeneity, and the safety environment
are what you expected.
Otherwise, the imaging scanner wouldn't do
what you want the imaging technology to do.
So this is just something you get a good understanding.
But now I hope all of you can remember,
can recite what is linear system additivity, homogeneity.
And you can put them together as a superposition principle.
And also safety environment is not something strange
to you, it's a very natural requirement.
And you can just try to visualize it
or remember this beautiful picture.
And in this lecture we will look into
safety environment linear system a little bit more.
And in order to do that,
I will spend a few slides talking about
the so-called impulse response,
impulse function, or Dirac delta function.
So that's something very interesting.
Let me just explain.
You can kind of treat it as a nice story.
And we know the mechanics laws by Newton.
And there are several fundamental laws.
All other things can be derived from his laws.
And one thing it says,
and all the universe, the solar system,
in nice motion, and we know the trajectories
and the interference can all be described
or predicted based on Newton's law.
And some questions were asked why we have such
coherent and harmonic arrangement.
And the rumor or legend goes
that Newton said that God kicked the whole system
at the initial time.
So everything is set into motion
because according to his law,
if something is in motion,
it will continue doing so without interference.
So the solar system is so beautiful.
Why do we have the solar system in the first place?
The modern theory says it all comes from Big Bang.
So the question can be asked why at the singular point
you have such Big Bang.
The initial push is something very mystery, right?
So we think the initial push,
and mechanically what's the initial push?
Think about a single object.
You just hit the object like a golf ball, like here.
So you got the initial push,
and the action will last for a certain amount of time.
The result is that a stationary golf ball
gains some velocity moving out.
So let's look at this picture in a little bit more detail.
We say that impulse causes momentum change.

So momentum is m times v.
Initially no momentum. It stays there happily.
Then you hit the golf ball, it changes the velocity.
So according to Newton's law, f equal to ma.
So the velocity change is due to acceleration.
Acceleration is due to force applied over time upon the object.
So force applied over time period is called impulse.
So just how impactful the hit is,
that will depend on how fast the ball will fly away.
So the mathematical description is rather simplistic.
f equal to ma.
So a is nothing but the velocity change.
So delta v, that's the velocity change,
divided by delta t, that's a, that's acceleration, right?
So f equal to m times delta v divided by delta t.
You move delta t to the left-hand side, you see f times delta t.
f is force applied upon the object over a period delta t.
So f times delta t, this is impulse.
Impulse causes the change of velocity.
And the velocity weighted by mass is called momentum.
m times v is momentum.
m times delta v is momentum change.
You got that part.
And you hit the ball, it's not something instantaneously.
Rather you imagine the golf ball will gain velocity gradually.
So you can imagine the force applied upon the ball,
initially small force, then high force,
then the force gets slower as the ball moves away.
So you can think about a lot of pictures and something like this.
So the force as a function of time, initially small force,
and acting upon the ball for a long time.
So you got this fairly fast impulse.
Or you think some very good hit.
So you can do this rather quickly, and sharp and strong,
and something like the tall one.
But what makes the final velocity depends on the area under the curve.
So you could do any of these ways as long as the area under the curve
is the same, that means this.
Means f times delta t, and all kinds of small delta t1, delta t2.
Delta t1 corresponding to maybe f1, delta t2 corresponding to f2.
So each time instant delta ti is a small rectangular shape.
You add them together.
That's the area under the force time curve.
As long as the area under the curve is constant,
you will get the same velocity change.
That is the final velocity of the golf ball will be determined by the area under the curve.
You can use rectangular shape like this.
So we say impulse as an overall effect.
As long as the overall effect is the same, the final velocity of the golf ball is the same.
Then we are happy.
We just want the golf ball flying at that speed towards the hole.
So something like that.
And this is rectangular shaped functions.
And then we can have another family of functional form like Gaussian shape, anything.
So as long as the area under the curve is the same, the impulse is the same as far as the final effect,
final velocity is concerned.
So this is just some comment.
We say the functional shapes do not matter in terms of overall effect.
Regardless of functional shape, we have the same for our purposes.
We have the same impulse.

And now we think impulse can be applied over longer time and over shorter time.
And impulse as a mathematical model, we want to think the time period is very small.
So impulse is strong, is very intense.
We are considering limiting case.
Here we rely on limiting process.
And then we can define this unique function.
So initially defined as a function with parameter tau.
So you have total width is 2 tau.
The height of the function is 1 over 2 tau.
So this is just to make sure the total area under the curve is 1.
So this is the mathematical abstraction.
Then we are not staying with this parametric form.
This functional form depends on tau.
But now I mentioned that we want to check the limiting case.
And we like tau, process 0.
When tau process 0, you can visualize the picture.
And the impulse becomes taller and taller.
The support of impulse.
By support I mean the interval over which the function takes 90 value.
The support becomes narrower and narrower.
The total length of the interval or total length of the support process 0.
But the amplitude approaches infinity.
So this is the limiting process.
But in this dynamic process, what remains is the environment.
That's the total area under the curve.
So this serves as a model.
You just try to mathematically describe what is the idealized impulse.
It is so soft. It is so strong.
But the area under the curve remains the same.
So let me make some further comments.
And then maybe a little bit off track, but still relevant.
So given the desirable effect of an impulse.
So why we just deal with the impulse as it is.
Why we want to do the limiting process.
Because this is mathematical research.
You want to idealize the process to be very thorough.
So just consider the limiting case.
So we have a perfect model.
Other things we hope can be described by the perfect model.
Just like in calculus.
And you try to compute the derivative.
And you have delta y divided by delta x.
Then you take a limiting process.
So delta x becomes dx.
Delta y becomes dy.
You say dy over dx.
That's the limit.
That's the first order derivative.
So we just do the same thing for delta function.
Then we say golf ball stay here.
Then fly away at a velocity say 10 meters per second.
And then if you are curious enough, you can turn around and ask.
So we got an impulse.
And then the impulse generated the velocity change.
Then we got this final velocity.
And if you are curious enough, you want to ask.
What is the real shape of this delta function?
I mentioned this area and the curve.
That determines the final velocity.
And what is the curve?
You're curious.
You want to ask.
You may say who cares as long as the area and the curve is the same.

Like the money RPI pays me.
And I do not care how you first put $100.
Later you put $200.
As long as the lump sum is the same.
I don't care.
So you can just say you don't care.
That's OK.
But if you do care, then the answer could be like this.
So since the area is the same and any shape, you really have no way to know.
So we say all kinds of possible shapes are equally possible.
You can just think about it.
It's something like maximum likelihood inference.
And you think, OK, I see an impulse applied to my golf ball.
But what would be the shape of that particular impulse?
And all things are possible.
So you have a probabilistic description there.
And if you really want to know the detail, you may just set up a very complicated equipment that you measure.
Once you measure, you know what's a particular form.
So this is something like modern quantum mechanical view.
And many things like quantum behavior can only be described by probability distribution.
They call it a wave function.
All things are really possible.
And only if you do the measurement, you know the cat is dead or alive.
So something like that.
And this is not really medical imaging.
Just something I think would be an interesting point of view.
And why we have quantum mechanical probabilistic description.
And one way, I guess, if God really kicked the ball or something like that, then with wave form, he will follow.
Because God is a fire.
So he will make all possible, all the shapes the same thing.
He would not particularly select a particular form.
All things are possible.
But eventually you have the outcome.
And if you want to know the detail, you need to perform the measurement.
Then the probabilistic periods will disappear.
Anyway, so we talk about delta functions.
And the delta functions can have two versions.
And the continuous version and the discrete version.
We say a unit impulse function is an example of a generalized function.
And it's usually called Dirac delta function.
So we first talk about continuous version.
So the effect matters, but the shape doesn't.
So what's the value of this generalized function?
And this is kind of different from what I explained in the previous lecture.
And in the previous lecture, I say what's a function?
And one of the very early slides, I say function is a mapping.
You have a domain, you have a range.
And you pick up one element in the domain.
And then you have a corresponding element in the range.
So this is corresponding.
Specific deterministic corresponding correspondence.
But a generalized function.
So kind of you still have the correspondence.
If the variable t is not at the singularity point t0, at any other point, you know the correspondence.
And for that generality, the functional value is zero.
But interesting happens at the singularity.
What if t takes the value at the singular point t0?
At that point, what's the value?
Because we are talking about a limiting process.
That particular value at the point t0 is not well defined.

It's infinitely high.
But the area under the curve remains constant.
So this is not a function in the classical sense.
That's why it is called generalized function.
So this definition is out-of-box definition.
And with delta function, later on we can see delta function can be used as a model sampling process, physical measurement process.
So this is a very cool concept beyond conventional mathematical instrument.
And the generalized function was introduced by, as you already know, Dirac.
Dirac is an engineer.
And he has the same major as myself, electrical engineering.
You see engineers can make fundamental contributions even in pure physics, theoretical physics.
So he is one of the greatest physicists all the time.
Anyway, so this is the definition of delta function.
It's a generalized function.
Very cool.
Just say the function is defined over discrete data points.
Say n equal to 0, 1, 2, 3.
And for each discrete point, you have a functional value.
So this is discrete function and variable taking values at discrete point.
And without loss of generality, you can think the index n or variable take integer values.
Could be negative, could be positive, or non-negative, depends on situation.
So you have two versions of delta function.
Continuous, discrete.
Let's look into more detail.
Before I explain the measurement sampling property of continuous delta function, let me just use one slide to review so-called integral mean value theorem.
Do you still remember that?
I'm sure you learned from your calculus 1.
Maybe calculus 2.
Doesn't matter.
You have continuous function.
It's a blue curve.
And if you do integral, finite integral from A to B, and what you are computing is an area under the curve.
This is just a green shaded area.
And the mean value theorem says somewhere in the interval AB, you have a point, there exists a point C, and you have the functional value Y equal to F of C.
And this green area equal to F of C times B minus A.
So that means if you think this is rectangular, the area of this particular rectangular happens to be the same as the area and the green curve.
And the green area and the blue curve.
So this is an integral mean value theorem.
So just a review. You know this.
Now let's look at the sampling property.
Alternative representation of a continuous function.
So we say you have a continuous function F of T, and we do multiplication.
Multiplication of delta T minus T zero and F of T.
Then you do integral.
So this product is very important for the next few lectures.
So we see this is multiplication.
And this multiplication can be expressed as a limiting process because delta function is defined by this limiting process.
So you got this one.
So this D becomes a parametric D tau.
Tau is a parameter.
So you just factorize this amplitude to 1 over 2 tau outside the integral symbol and change the limit of the integral

because this delta function is defined with respect to T zero.
So the limit of integral is T zero plus tau, T zero minus tau.
You got this one, right? Follow me.
Okay.
So this is an interval.
So this T zero plus tau corresponds to B,
I mentioned in the previous slide.
And T zero minus tau corresponds to A in the previous slide.
According to integral mean value theorem,
and within this interval, there exists a point
which doesn't care how to find the point.
The point C, here I call the point C T star.
So there is a point inside the interval.
And the functional value at that point times the length of the integral interval.
So the interval is from T zero minus tau to T zero plus tau.
The total width of that integral interval is 2 tau.
So 2 tau times functional value.
That is the height of this equivalent rectangular shape.
So you got this part, okay?
And we know that tau will approach to zero.
So you just take this 2 tau, 2 tau, they cancel out.
You got this part.
Then you let tau approach zero.
When tau approaches zero, the left limit and the right limit
all approach to the same thing.
That is T zero.
So the T star cannot be anything but T zero.
You got this part.
So this shows the sampling property of the delta function.
So if you want to get the functional value at T zero,
and you just time that function with delta function shifted to zero,
then you perform integral operation.
Then you just take that value, single out that functional value.
So this is the sampling property.
And based on that property, so you can see, okay, this.
So you do sampling of function of T at position T equal to T zero.
You got what you want here.
And you just change the variable a little bit.
You call T zero T, and you call T tau.
We just change the dummy variable, so that doesn't matter.
Now you got this one.
And through all these steps, and we return to the original object,
which is F of T.
So F of T is nothing but this integral.
So in other words, we have an alternative representation
of the original continuous function F of T.
What's the geometrical meaning of this representation?
So that basically says the F of T or F of tau
can be viewed as many, many delta functions.
So you can just add all these delta functions together.
And each of the delta functions with amplitude weighted by F of tau.
So heuristically, intuitively, you have continuous function.
What this equation means, you just cut this function,
just like you cut your vegetable into many, many thin slices.
And each slice is a small rectangular object.
And you think the continuous function is nothing
but a collection of all these small thin slices.
And each of them is a delta function.
And certainly, it's not all the same.
The delta function must be weighted by the functional value at that point.
So you just have an alternative representation of the continuous function.
So this is in the continuous case.
And in this case, you can view the original function

as a collection of rectangular bars like this.
So you define this rectangular function.
And it looks like a gate, right?
Also called a gate function.
So rectangular function you define as something like this.
At the left discontinuous point and the right discontinuous point,
we naturally define the value as half of the right limit and the left limit.
Because amplitude is 1.
So at these two discontinuous points, we define the functional value as half.
So this is a regular function, just a rectangular function.
So we think arbitrary function becomes always, again, just the same analogy.
You cut the function into many slices.
And each slice you represent as a weighted rectangular function.
So this rectangular function you think is just a finite impulse.
And you just define the discrete delta function.
We will take value 1 as a given value.
It's already an area under the curve.
Because you're talking about discrete data points, like time instances.
And you really just use that sampling point to represent a small bar.
So the small bar, what makes sense is an area under the curve.
So the area normalized to 1 as a basis function.
So you got delta n discrete version defined as these sequences.
And it will be only 1 for n equal to 0.
Otherwise, it's just 0.
So just a counterpart of a continuous definition.
And you can also shift it to any place.
So you have multiple delta pulses defined on discrete data points.
Then a general function can be expressed as a sum of rectangular or gate functions.
So this is our original continuous function in gray.
And then we can just do piecewise constant approximation.
And for each piece, you have a delta function or rectangular function.
And it's certainly scaled according to the continuous function or value.
Then this particular bar can be made equivalent to delta function.
So you got this continuous delta function.
And the area under the curve is the same as the area under a given rectangular shape.
So all these are linked together.
So you can have the discrete delta function to represent discrete.
You can use discrete delta function to represent arbitrary discrete function.
So this is a kind of arbitrary discrete function.
Say we can use discrete delta function to represent it.
You need weighting.
The weight is gk, so the functional value at the corresponding data point.
So for a specific example shown in the lower part of the slide, you see you just got three points.
The first one is a traditional discrete delta function.
Then you have a negative one just with the same amplitude, but also carrying a negative sign.
Then the third one, you got a little bit higher amplitude.
So the first one is nothing but delta n.
The second one happens at a location n equal to 1, carries a negative sign.
So this second component can be represented by minus delta n minus 1.
So you got this one.
The third one, so you got the original delta function shifted by two steps.
So you have delta n minus 2, and the amplitude should be weighted by 1.5 because this is a little higher than the original one.
So you see this weighting 1.5 may correspond to g of k.
So g of k is the functional value here.
So you can imagine for any arbitrary functions, and you can just cut it into slices.
And each slice is a weighted discrete delta function.
So now I have covered two versions, continuous versus discrete delta functions.
So a functional relationship can either be discretized, and you just have

something shown here, or continuous.
For continuous function, you can give an alternative expression in terms of continuous delta function.
And for discrete function, you can do the same thing, but with discrete delta function.
And after all, we are dealing with functional relationship, whether it is continuous or discrete, depends on situation.
And philosophically, the things can be fundamentally continuous or discrete, we don't know.
Just like a water flow, and if you see water drop, that's discrete.
But you have water flow, you think it's continuous.
If you use magnifier, powerful enough, you see a lot of water molecules, that becomes discrete.
And if you check further, and the water molecule, you see a lot of quantum wave functions, that becomes continuous probability wave.
That becomes continuous.
So you can keep asking, who knows, on Earth, continuous or discrete.
But anyway, we know they are interchangeable, and we can convert a function either as a summation of discrete bars,
or you think the even loop is a really continuous function, you want to model it as a continuous counterpart.
So just two things are really equivalent.
So you remember the first slide of this lecture.
On the left hand side, you have summation, right hand side, you have integral.
So they are really the same thing, if you understand the process to link them.
Now we talk more about safety environment system.
Safety environment system is so important, because if you know the system response to impulse,
then you know everything about system behavior.
Why I say that? Because the system is safety environment.
So you have an impulse at a given location, then you can infer what the system will react when the impulse is given at a different location.
You know the system response when you give an impulse at an initial time instant.
Then you can infer what the system will react when you give the input at a later or earlier time instant.
So you know the behavior at any time instant.
Or you have multiple places, multiple time instances.
So you just give input as a combination of all these things.
Because the system is linear, you can combine these things together.
So if you follow me so far, you would understand.
If I give you system's impulse response, then for any input, you can infer what will be the system's output.
That's why the impulse response is so important for safety environment system.
Let me just give you one slide to show how you find the system's output to arbitrary input
in terms of system's response to an idealized impulse function.
This is, I didn't put no space, but do you have a question? No?
So this is a safety environment linear system.
I underlined this. I don't have space to write.
But this is a safety environment linear system output.
And this slide is the key of this whole lecture.
So for safety environment linear system, we want to find out the output.
We are engineers. We want to find the results. We want to find the output.
Given arbitrary input, you gave system input. What is the output?
This is your question. You want to find out.
What do you know? You are given the system's impulse response.
So that means if you give system impulse at a time equal to zero, the system will respond like this.
This is all you know.
And now you are asked to find the output for arbitrary input.
I do not ask you to do this from nothing. I give you critical information.
I gave system a specialized standard input, which is delta function at time t equal to zero.

The system will respond like this.
I gave you this piece of knowledge.
I ask you, if you gave system a general input, it's not a delta function.
The delta impulse, and I already know, if you gave system delta impulse, the output will be this.
But what if I gave system input something like this?
What will be the output? How you can compute that?
Compute the safety environment linear system output.
You can do so this way.
Just follow me. Then we have a little rest.
See here.
So this is a safety environment linear system with characteristic h.
Meaning, if you gave system a delta function, this creates a delta function.
The system will have output h of n.
h of n is not a single point. It's just a sequence.
This is a series of numbers shown here.
This is what I gave you.
What I ask you to do, if you gave system an input called x of n,
x of n is a series of numbers.
This is a sequence of numbers.
What will be the system output y?
This is your task.
How do you deal with this?
Let's just go line by line.
Also, good review of what we learned so far.
Linear system, additivity, homogeneity, or scaling, and the safety environment.
See the third line.
So if the system response is delta response, impulse response, is h of n,
what will be the system's output?
If you shift the input from delta of n to delta of n minus k,
you really just shift the input by the amount of k.
And because I underlined that the system is a safety environment,
so you shift the input by k, then the output will be saved by the same amount of k.
So if you have this delta n minus k on the right-hand side,
you must have h of n minus k.
Why you have the third line?
By safety environment.
Now we say, what if I scale the input with the weighting factor x of k?
And then we say the output must be scaled accordingly.
Why is that?
Because the system is a linear system.
Linear system satisfies the homogeneity.
So you got this one.
Then how about we do summation?
We add a bunch of numbers, a bunch of input together,
indexed by variable, discrete variable k.
So you have a finite number added together.
Then on the output side, so all these things,
the output will be linearly combined in the same way.
And then we further assume this summation can be done over an infinity number.
Certainly you assume the summation converges to a finite number.
Then that is feasible to perform.
Then you got this summation from minus infinity to positive infinity.
And on the right-hand side, you see same thing.
And we mentioned that with delta function,
you have an alternative way to represent our original function,
either in the continuous or discrete case.
And I showed you step by step.
And with delta function, you can sample functional values at any point
so that continuous function can be represented as a convolution, as an integral.
And the same thing here.
This summation on the left-hand side is equivalent to the original input function

because this summation for given n,
this summation will only make sense when k equal to n.
Then this only xn come out.
The rest will be zero out.
So this part goes through these steps,
really return to the original form of the input, x of n.
So this is what you asked us to do at the very beginning.
This is really the output of the system.
Now the unknown yn becomes known
because this expression contains x and h.
Both x and h are known.
x is input.
h is the impulse response of the system,
which is given in the first place.
So this whole process just shows you the importance of
impulse response function for safe environment linear system.
And this summation, we call it convolution.
So this is the title of this theme of this lecture.
Now we have about 10 minutes rest,
then come back to finish the class.
Let's continue.
You see this slide, we have a red diamond.
That means very important.
Try to understand.
Convolution has been very confusing concept
for engineering student and maybe also for mathematical student
because you have things kind of complicated.
It's not a direct multiplication.
And you got two things.
One is the negative k and the parameter n.
And the easy way to remember,
you think this is a central thing,
you got a positive n, you got a negative n,
you got a positive k, negative k kind of symmetric
with respect to n.
So this kind of convolved relationship may cause confusion.
But if you view the derivation step by step,
you see how this h of n minus k comes from.
This is important.
If you think linear system is important,
system is important.
Linear system is a simple class of general system.
So for linear system, when you have safe environment,
you can just go step by step.
So for any linear safe environment system,
output can be computed from input
in terms of impulse response h of n.
And because of the logic shown here,
you got to have this form.
So you know this form is fundamentally important
for linear system theory and engineering.
Because it is so important,
and in the book draft,
and also in this PowerPoint file,
it gave you continuous version.
So just do the same thing again,
this time a little quicker.
You have input, output.
Given input, you want to find output.
What you know is continuous response to a delta function.
You know this.
Then you do shift.
And based on the safety environment,
you got this part.

Based on scaling or homogeneity property,
you can scale the delta function.
You got scaled output.
You add the things together,
so you got this integral.
And this integral, and I used one slide to show,
this integral will give you the original function,
continuous function, x of t.
So you got this part.
And the right-hand side is nothing but the output,
which is what you want.
The output is a convolution of input,
and the system impels raise bounds.
So see the expression.
You think this is a function of t,
and this t in the middle,
and you have the dummy variable,
tau minus tau.
You do integral with respect to tau.
So this is the system output.
So the linear system,
safety environment linear system,
the output can be computed as a convolution,
either in discrete or continuous cases.
And I express as discrete or continuous convolution.
So heuristically, this is to say,
you can express input as many impulses.
And for each impulses,
you know the corresponding output component.
And the input is summation of all these impulses.
Then you do the summation or integral the same way
on the output port.
So you sum all the responses to individual impulses.
Scale the safety of all these components
added together to form your final output.
This is your final output in the discrete case.
This is your final output in the continuous case.
So this is the formulation.
So as far as formulation goes,
we have covered everything.
So this is a way to compute the system output in any case.
And as I said, I know based on experience,
convolution can be confusing,
so we better look at a few examples
so you have a better understanding.
Let me give you a hands-on example.
Just think about two discrete functions.
In each function, you have just five data points.
You just see your hands.
So left hand, you have five data points.
Right hand, you have five data points.
And according to the derivation of convolution operation,
you know you have this minus sign for this k.
So the minus sign, if you have a function,
just forget about n at this moment.
For function, say f of k, you have certain function.
Then you just change the sign of the variable.
The functional form will be the same,
except you flip it around.
So you just think you have two functions.
Then because of the negative sign,
negative k, one function just flip around like this.
So you see the back of your right hand.
So you got something like this.

So this is original right function,
original green function.
Because of this negative sign,
you need to fold one function over to flip your hand.
So you got this green function here.
So this green function is after flipping.
You got this.
Suppose n equal to zero.
Now after n equal to zero,
so this is right function.
This is flip green function for n equal to zero.
You got things like this.
Then you see the definition of the convolution.
When n equal to zero,
you really need to do a lot of partial product,
x of k times h of minus k.
So you really do pairwise multiplication.
So you do pairwise multiplication.
This green value is paired with some value here,
but this is zero.
We know it's only non-negative for these five data points.
It's zero, zero.
But at this point, you can do partial summation.
Non-trivial.
You got number.
So general case, you really do partial summation,
partial product, many terms.
Then you add things together.
Then you got the final result for a particular n.
So for n equal to zero,
you flip over, do matching, obtain partial product,
then add all the partial product together.
And most of the partial products are zero,
and only the middle one is non-zero.
You got a value there.
Then for n equal to one,
so this green series of number will be shifted
by one step towards the right-hand side.
So you see this one is shifted this way.
Then the computation happens the same way.
So after the shifting, so you got n equal to,
here is n equal to one.
Go back.
N equal to one.
For n equal to one,
you still do this pairwise multiplication.
Many of them added together.
And in this particular case,
only two partial products are non-zero.
So this one and this one.
And others, you think a lot of green zeros here,
and the right zeros on the left half of the axis,
and a lot of red zeros from the fifth point on towards the right.
So the matching up result,
zeros matched to certain values
will give you zero partial product.
But for these two data points,
you have non-zero partial product.
You added it together.
That is a value of the discrete convolution
when n is equal to one.
You got this one.
So you keep doing n equal to two.
This shifted one more step further towards the right.

You do the matching.
You got a result.
You keep adding n.
Let n be three, four, five, six, seven, eight.
So you do the matching.
So the most interesting thing for n equal to four.
So you have this four match up.
You will have a maximum value of convolution.
Then you move out.
Finally, you got all zeros.
So you have two functions.
And each function has five non-zero points.
You flip over.
Then you do safety multiplication, summation.
You keep going.
So you got non-zero value here.
You got maximum value here.
You move it out.
You got a small value, non-zero value again.
So when you have two discrete functions of length five
in this case, so non-zero value will be five plus five
minus one.
Because you do the matching.
You just do your hands-on experiment.
You will realize that you can extend this example
into a more general case.
And then say we got a particular example.
Our left-hand figure, suppose your thumb takes a value five.
Then the index finger, four, three, two, one.
And here, the small finger, one, two, three, four.
The right thumb, five.
You flip.
And then you do the discrete convolution step by step.
You got all the values.
Altogether, you have nine non-zero values.
So suppose you have two discrete functions with length n1.
The other has length n2.
So you do convolution.
The total non-zero length in general
will be n1 plus n2 minus 1.
So just think about this.
This is the example, hands-on example.
You play around.
You see how things work together.
So the flipping, safety, multiplication, addition,
first time may confuse you.
But the convolution is so important for you to master.
I argued the case imaging system by definition.
And we hope it should be ideally safe to environment and linear.
And if the system is safe to environment and linear,
the output, the general output, is nothing but a convolution.
If you want to find out imaging system solution,
convolution is something fundamental.
So you have to know.
So this is the first simple hands-on example.
You play around.
You have a very good visual impression.
Now let me give you a second example.
The same idea, but you just need to see the same thing
multiple times before it becomes part of us.
And this is about injection of contrast.
Why you want to inject contrast like iodine into your bloodstream?
Because in actual imaging, the vessel, blood, tissue all look similar.
So you want to see if the vessel gets narrowed down, got blocked.

You want to see clearly.
Then you introduce iodine.
Once you introduce iodine into bloodstream,
the vasculature gets lighted up.
So you see vasculature very clearly.
So iodine into bloodstream and through vasculature
may go through the system in multiple paths.
And some paths, the contrast travels faster,
may correspond to major vessel.
And the other part may be micro vasculature.
And it goes slowly.
So we will reach to an output port at a different time point.
So there's some delays, different delays along various paths.
So the gamma camera will take a picture of the whole thing.
Because of time delay, the injection will first capture as H1,
then the second one as H2, H3.
So this is an impulse response.
Suppose you inject ideal impulse,
and you just have a very powerful injector.
You inject delta function into bloodstream.
And this is what happens due to physiological
or pathological impedance or delay.
So this is the output, the impulse function.
If you use gamma camera,
and you will follow the dynamics as a time curve shown here.
But in reality, you cannot just all of a sudden inject
a given amount of iodine into bloodstream.
Rather it takes time.
So this is an ideal input.
So initially you inject more, later you inject less.
So over a time course, you have an area under the curve.
That amount of iodine injected into bloodstream.
And this non-ideal input can be cut into multiple idealized delta functions.
So initially you have this blue delta function,
red delta function, yellow delta function.
So altogether it's a continuous curve.
You are not able to do this magic.
All of these things put into your body instantaneously.
You cannot do that.
You can only do it over a certain amount of time.
So that is decomposed into multiple delta functions.
Now we already know for a given delta function,
the system response, the time curve,
can be monitored by gamma camera like this.
Then for this particular component,
and we have the delta function look like this,
to be consistent with the impulse response shown here.
So you know this response is due to this single rectangular bar as input.
Then the next component, the red component,
and it will be, if you just look at all these red components together,
and just remove all other components,
you will have the same shape of response.
But all these are red.
So you got this red contribution.
And likewise you have a blue, sorry, yellow,
and certainly the yellow impulse function is delayed by two time units.
So the response is also delayed by the same amount of time units.
So you got all the blue responses.
Green, purple, and the dark blue, and things like that.
So you got all these output components.
The real output will be summation of all these things.
So when you sum all these things together,
you got the overall output.
So each component's response is already scaled by the input function

at a given point.
So for this one, it's a little less.
So the overall response for red input, red output,
so you have a smaller amplitude overall.
So just add all these together.
So this is a visualization of discrete convolution operation.
So these are very heuristic,
and may take your time to let the idea sink into your mind.
Just take your time.
So now let me give you some more abstract examples.
You really need to think about calculus.
It's not hard, but this is a continuous example.
But whenever you deal with convolution,
you flip, save, and a few more steps,
not so much straightforward,
and things become complicated,
so you may feel confused.
Anyway, so just let me show you an example.
A lot of technical details, and not so hard,
so I wouldn't spend too much time.
Just show you the steps,
and later on you can review for yourself.
So you have two functions, f of t, g of t,
so f, red function, g, the purple function.
Then to compute the convolution for given time t,
remember the hands-on example,
first flip one of the functions,
then keep saving by the amount you want.
At that amount, you can evaluate the value of convolution.
And here, the purple function is symmetric.
It's symmetric, and we want to flip that.
It's easy because it's symmetric, so it's just the same thing.
Then after flipping, you're concerned about the amount,
you do the shifting for the purple function.
These are finitely supported functions,
and we need to discuss the result case by case,
so we really divide it into several cases.
So as shown here, in the first case,
t is less than minus 2,
minus 2 means the negative value,
means you really move the purple function left-wise.
So when t is less than minus 2,
so it was moved so left-way,
and this purple function and the red function,
they do not mess up each other.
So we say we do shifting, then we do multiplication.
Finally, we do summation or integral,
and then the multiplication is point-wise multiplication.
Like the finger example, you do the point-wise matching.
You find all the parts of the product,
then you add them together.
So this is the idea.
The continuous case and the continuous discrete
are really the same thing.
They can be viewed interchangeably.
So in this case, case 1, t is less than minus 2,
all the matching up are very simple,
and the non-zero value is matched to zero values
of the other function,
so the partial product will be all zero,
so the result is zero.
I hope everything is as simple as this,
so we are done.
The second case, the t is between negative 2 and 0,

so in this case, you have a little overlap here.
So this purple function,
given this saved arbitrary amount t in this interval,
so the right limit is 2 plus t,
so this is 2 plus t,
and here is zero,
so only in this interval,
you have non-zero partial product.
Beyond the upper limit,
so you got zero matched to right,
and before the zero point here,
the zero value from the right function
is matched to purple function,
so the integral will be zero,
so only this part,
you need to compute the area under the curve,
so you do multiplication,
you got non-zero value,
you compute the area under the curve,
so from the lower bound,
zero all the way up to 2 plus t,
you do integral,
then this integral is multiplication of the two functions,
so the purple function has amplitude,
purple function has amplitude 3,
and this red function has linear form
minus tau plus 2,
this is a partial multiplication,
you do integral and several steps,
you got this part,
you got the value,
the final result here,
as a function of t,
so this is result,
I wouldn't give you all the detailed derivation,
you can try yourself,
so this is for cases 1 and 2,
then we have three more cases,
cases 3 are shown here,
so you have this purple function,
contain the right function completely,
so you just do the computation,
from 2,
integral from 0 to 2,
you got constant value 6,
in the case 3,
and you move out,
so this purple gate is no longer able to cover
the right function completely,
then you need to do the partial product integral
for this non-trivial interval,
you got this part,
and finally the purple function is moved out,
you got zero again,
so the final total whole solution,
it just put all these five cases together,
as summarized here,
so you just check for yourself,
this is continuous example,
and another example,
and your homework is related to this,
this is RC circuit,
those of you who are not familiar with RC circuit yet,
do not worry,

just believe this is input as a function,
this is output,
the output is related to input,
and the system impulse response,
that means if X of t is just a standard input,
delta of t,
is a delta function,
then the output of Yt will be H of t,
described as this,
so this is something like this delta function,
the system got excited,
then just decay away,
this is output,
if you do not know it electronically,
at least you understand it mathematically,
then you can do MATLAB programming,
so you know the delta response,
or impulse response,
therefore arbitrary input,
you can cut it into many pieces,
you decompose the continuous general input function,
as a bunch of delta function,
and as we argued before,
the output for this arbitrary input,
will be a convolution of the input function,
and the system delta response,
you just do convolution,
continuous function,
how do you do convolution,
and I gave you an example earlier,
same idea, certainly the functional forms are not the same,
but the computation is the same,
so you try to compute,
and then you try to derive,
try to compute,
then you have an idea,
how you can find linear system output from input,
and the systems impulse response,
you play it around,
and you see the curve make sense to you,
then you get a better understanding,
so far I have been explaining to you,
convolution in one dimensional case,
and the convolution is a nice operator,
and once you understand it well,
in one dimensional case,
it's not so hard for you to understand it,
in two dimensional, three dimensional,
two three dimensional are relevant,
because we are talking about medical imaging,
so convolution can be extended,
to higher dimensional case,
for example,
talking about two dimensional convolution,
you see the two dimensional function,
will be defined on xy plane,
then convolution,
you have x and y,
so with respect to x,
you have minus n1, n1,
with respect to y,
you have minus n2, n2,
so kind of same rule,
you can remember,

this is in summation,
so discrete case,
and in continuous case,
the summation becomes integral,
so you have variable x,
with respect to x,
you see minus tau y, then tau y,
so for y, you have minus tau 2,
then you have tau 2,
so these two symbols,
summation and integral,
for engineers,
they are very much the same,
and the two dimensional convolution,
has direct practical imaging implication,
for example,
you have input image,
something like this,
for example,
the imaging system is never perfect,
will blur things out,
no imaging system will give you all the details,
that is not possible practically,
so the blurring process,
can be modeled by this averaging mask,
so you got to say,
three by three,
the blurring mask,
and you got all the values,
so basically you do convolution,
this mask normalized by one over nine,
because you have nine elements all together,
so basically it says,
and you do convolution,
with this blurring mask,
over the input image,
you just lay out this blurring mask,
this way centralized with a given pixel,
in this case,
the given pixel is three,
so you just flip the mask,
this symmetric mask doesn't matter,
you flip it,
you just save it over the field of view,
at each given location shown here,
and you do pairwise matching,
and after matching up,
you compute all the partial product,
then you add it together,
so the pure effect is,
you just add all the nine elements under the mask together,
then you do normalization,
divided by nine,
so this is just the averaging operation,
over all these pixels,
nine pixels around the central pixel,
so this averaging, as you understand,
will blur things out,
so any sub-peak or deep valley,
you average out,
so the detail becomes blurry,
you do so pixel by pixel,
because the convolution,
you just smoothly save the filters,

from left to right,
then go to next line,
do it again,
so it's just roaming around,
so the picture is smoothed out,
so this is image blurring process,
and this is pretty much a reasonable model,
for optical imaging system,
you have optical lines of finite aperture,
the system response function,
something like this,
is a finite size,
a single bright small dot,
will appear like something like this,
so this point spread function,
will smooth things out,
if two small points,
too close,
then you have one peak,
the other peak,
they overlap,
so altogether,
you would not be able to tell them apart,
so this is blurring process,
of the optical camera,
so any imaging system,
by definition,
is not perfect,
and idealized delta function,
will appear as a small base,
or finite size small structure,
so that will be a blurred version,
of idealized high contrast dot,
so each small bright dot,
can only be focused,
into certain feature,
greater or more blurry,
and you will not have idealized,
picture recovered,
so this is blurring process,
so now we say,
okay,
image deblurring operation,
is just to undo the blurring process,
and something like,
you have a camera,
take a picture of the plate,
you cannot read the number clearly,
because motion blurring,
because the camera is not of high resolution,
you got blurred thing,
so the blurring is due to convolution,
so the real object,
convolved with system,
point spread function,
then use some advanced algorithm,
you can take the blurring out,
this is called deconvolution,
and we will see clearer and clearer,
and the convolution is advanced form,
of multiplication,
and this deconvolution,
is advanced form of division,
so when we learn mathematics,

initially we have addition, subtraction, multiplication, division,
and now you are in college,
advanced form of summation,
and I would say it's integral,
advanced form of subtraction is derivative,
now you have convolution, deconvolution,
they are advanced versions of multiplication and division,
and look at this,
this is a real image,
you have a system, point spread function,
you convolve together, you got this blurry version,
and from next lecture on,
we will talk about Fourier analysis,
so in terms of Fourier transform,
so image can be transformed into Fourier transform,
and the point spread function can have Fourier spectrum,
so convolution is equivalent to multiplication in Fourier domain,
so the deconvolution is just the original function,
and divided by the Fourier transform of point spread function,
so this is really just the backup of the point,
and you do not fully understand now,
but you will after we finish the foundational part of this lecture,
so now you are well educated,
and you are not like kids,
just know how to add things together,
even algebraically,
x plus y,
you know how to add things in very sophisticated fashion,
that is integral,
and the differential is just subtraction,
and about multiplication and division,
really this is fancy form,
convolution, here called deconvolution,
you will know that more later,
and now we return to one dimensional case,
and in multi-dimensional cases,
and the claims remain about convolution properties,
because I said convolution is advanced form of multiplication,
and you can expect all the generic properties
for multiplication also applies to convolution,
that's cool,
so if you convolve function one or function two,
that's the same as you do convolution,
with function two and function one,
so all that doesn't matter,
that is called community,
so multiplication satisfies the same property,
and here, so, convolution,
and associative, so you have three functions
convolved together,
it doesn't matter,
so how you group the convolution,
you do the convolution for the first to first,
or for the last to first,
the result is the same,
and also you have distribution property,
so you can just factor out the common function,
or just distribute a factor into the summation,
so that's the same,
so you can try to prove these properties,
and I wouldn't spend too much time on the proof,
but I will list one proof,
like how do you prove community property for convolution,
so all these nice properties,

satisfied by conventional multiplication,
works for convolution,
there's another angle to indicate,
convolution is an advanced form of multiplication,
related to convolution,
and another definition is called cross-correlation,
it's just a variant of convolution,
so this is a minus sign,
so remember,
with respect to x, you have u, minus u,
and the order doesn't matter,
but if you change the minus u to plus u,
that's called cross-correlation,
you just think, you flip the function,
the function is arbitrary,
so you flip it, it's still arbitrary,
the flip is not essential,
but we have a different name,
called cross-correlation,
so just let you know,
just let you know this general idea,
and you have this cross-convolution,
cross-correlation,
and graphically,
and you can see this function f,
function g, you do correlation,
and cross-correlation,
and one function is flipped,
the minus sign is changed to plus sign,
so the result is flipped, shown here,
if two functions are the same,
so the function is convolved,
or auto-correlated to itself,
and again, we have a special terminology,
we call it auto-correlation,
so auto-correlation,
you do the trick with the function itself,
so you get a curve,
the final curve, something like this,
and the cross-correlation is very important,
and it can be used for signal detection,
because of so-called quasi-schwarz equality,
so for two functions,
so given the norm,
so given the integral shown here,
and we can have this inequality,
and you take time to read,
and I explained in my book draft,
and you have this relationship,
I believe you showed this before,
but anyway, that shows this kind of product,
you do point-wise or element-wise matching,
then you compute partial product,
add things together,
this is more generally called inner product,
so this inner product form will be always less or equal
to the product of the norm,
so because all these squares added together
give you an idea of the vector lines,
it's a norm, so you read it,
it shows the inner product form will reach the maximum value
only when the two vectors, they are similar,
ak equal to cbk,
so always like that,

that means one vector is a scaled version of the other vector,
they look the same,
this value reaches the maximum value,
so this can be used for signal detection,
so you have a desired signal form,
like a train of sinusoidal pulse in red,
so we call this a matched filter,
this is the functional form we want to find,
but this kind of signal is buried in very noisy background,
like a radar echo,
so all the blue signals,
so in the blue signals you have two components,
one is something very similar to this sinusoidal oscillation
over a finite period of time,
like you send the radar echo into the sky,
the radar pulse into the sky,
the pulse hit an airplane,
will come back,
because that is what you send,
you expect to receive an echo,
something like this red signal,
but what you receive is very messy,
all the noisy background,
and the signal is weak,
buried in the background,
so you have this blue signal,
then you do cross-correlation,
you do the pairwise matching,
multiplication,
added together,
that's your output,
when this output reaches maximum value,
that means at that point,
the echo signal is there,
so this is used for signal detection,
and also in imaging case,
you can use cross-correlation or convolution to detect a feature,
so this is the signal,
so you see this is a feature,
or matched filter,
and this has a shape like an edge,
so you got the positive one,
zero, negative one,
so you do the convolution,
or cross the field of view,
so whenever you have edge,
it will give you higher value back,
so you do this matching filter,
convolution or cross-convolution,
or the field of view,
you will detect edges,
whenever you have edges,
it will give higher value,
certainly each filter emphasize only one direction,
you need to use multiple directions,
so this is a topic very popular in image processing,
you have filters,
emphasize edge of different orientation,
you just do edge detection,
the edge things together will give you overall edge enhanced picture,
there are many kinds of filters,
and it could be multi-scale, multi-stage,
and all different directions,
the edge can have different orientation,

but the idea is that the matching filter,
when you do convolution,
or you do it cross-correlation,
according to Cauchy-Schwarz equality,
when the waveforms match,
you have higher value,
so this is a good utility of cross-correlation,
the original motivation for us to learn convolution,
is to find safety invariant linear system response,
then based on Cauchy-Schwarz inequality,
we know that the same operation can also be used to detect the edge,
so it all indicates this kind of special mathematical operation,
you have a function of continuous discrete,
you do this point-wise or element-wise matching,
you got all the parts of product added together,
so this kind of computation plays a fundamental important role in information science,
I think in biological science,
there's a matching,
then you have all these things together,
that's just a DNA sequence,
you do the matching,
you have double helix,
but here the convolution,
cross-correlation,
or more generally, later on I will explain,
the inner product,
in calculus you learn the inner product,
the inner product really you have vectors or functions,
you do the pair-wise matching or point-wise matching,
do partial, do multiplication,
then add it together,
so this structure is not only important in biology,
also important in information processing,
so just think about that,
so summary,
we talk about linear system,
particularly safety environment system,
this is add-on to linear system,
imaging system,
we want the imaging system,
safety environment in general,
and the convolution,
and this is a way to find the output of safety environment linear system,
this is the main motivation,
and additionally I say,
convolution can be viewed as cross-correlation,
cross-correlation can be used for feature detection,
that's just a second,
but also very important application,
so convolution in this case is cross-correlation,
this x prime is something like this,
in continuous case,
x minus a,
so try to understand the convolution,
do not get confused,
this is not an easy thing,
and not uncommon,
students say the convolution makes me really headache,
but you need to spend time,
convolution will be your good friend,
so we talk about decomposition of general function as delta function,
then in terms of delta function,
you have impulse function,

impulse function and input come out together,
give you output of linear system,
so this is your homework,
pretty much like the slides I showed you,
it's RC circuit,
if you are not familiar with RC circuit,
at least you just treat it mathematically,
you just think this is a black box,
the input, output, delta function,
how you compute for output, given input,
and I will ask TA to upload,
again whenever you have homework,
you have one week to finish,
that's all for today.