Hello everyone, welcome to a fresh new lecture on the topic of wavelet transforms 1. This is Dr. Anirban Dasgupta and let us get started. So, to introduce wavelets, let us first understand what the limitations of our short-term Fourier transforms (STFT) are. So, the first is a fixed time frequency resolution. What does that mean? So, in our STFT, the time-frequency resolution is fixed because the window size is fixed.

So, if we have a larger window, say the window is defined in terms of your n, the signal length, or the window length, rather. So here, if I have a larger n, then typically we are computing the endpoint DFT. The higher the value of n, the better my frequency resolution is. Why? Because I am able to represent a greater number of frequencies for that specific window.

But the time resolution becomes poor. Why? Because the larger the window, I am not able to localize exactly where my event has occurred. For example, if my window size is 0.1 seconds, then I know that in that 0.1 seconds or a fraction of a second, that event has occurred.
But if I am analyzing in a window of, say, 5 seconds, that is a big window. So, anywhere in that 5-second duration, these frequency components are present, but exactly where is not localized in a very precise manner. So, the time resolution will be poor. So, this trade-off makes STFT difficult when both high-frequency and low-frequency components are present in the signal because each will require different resolutions. The second is capturing sharp transients and discontinuities.

Now in many signals, for example, a signal is suddenly changing like this. So, if you need, say, sinusoids, maybe you have to use a lot of sinusoids to represent this. So, this is one sinusoid of a specific frequency, then another sinusoid that is of a higher harmonic, then another fast-varying sinusoid. A lot of components will be used to approximate this kind of sharp change in the signal. And since this sinusoidal basis, which is the basis for our short-term Fourier transform, is a smooth signal.

So, it is very difficult to capture the sharp discontinuities. You will need a lot of components from a low range to a high range to capture this kind of discontinuity and a very good example is our square wave, where we will ideally need an infinite number of sinusoids to approximate this square wave. So, this brings us to the concept that I am just

giving you a hint: if my analyzing signal were instead of a sinusoid something like a square wave, then this kind of signal is easier to analyze with fewer components. Now, to understand wavelets, what is a wavelet? Let us first try to define it.

So, a wavelet is a small oscillatory function ψ(t) that is localized in both time and frequency. What does it mean? So, sine, which is the basis of the sinusoidal function, is not localized in time; it is present everywhere in the time domain. But a wavelet is a function that is of short duration; that is, it has limited duration, and that is the reason why you can capture some events that are of short-lived duration has an average value of 0; this is the same for the sign and also for the non-zero norm. So, the main distinguishing factor is that it has a limited duration: this analyzing function.

This function is usually termed the mother wavelet because from this mother wavelet, we will derive other wavelets. Now, this mother wavelet, say I have any signal that is varying over any duration. So, if I have a short-lived signal, which I am calling a wavelet, I cannot say I have a signal that is varying like this and I have a signal that is like this, and this I call a wavelet. So, I cannot, if this has a duration of, say, 0 to 10 seconds, and this signal which I am going to analyze has a duration, say, 0 to something like 50 seconds. So just with this signal, without using any operations on it, it is difficult to analyze the entire signal or the signal that is to be analyzed.

It is very difficult to get that. So, the mother wavelet is scaled. Scaled means that if you see, these are faster than the mother wavelet. So, I have to shrink them whereas there can be cases where the mother wavelet is dilated. So, I have to dilate the mother wavelet.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \, \psi_0 \left(\frac{t-b}{a}\right)$$

So, that is the meaning of scaling. And then shifting, I will analyze this region; then I will shift and analyze the next region. So typically, analyzing with wavelets means that I have a waveform like this; this is called a wavelet, and I will shrink it or dilate it, and then I will translate this version and the dilated or shrunk versions to obtain this original signal x(t), and this wavelet is called Ψ(t). So, I will analyze this x(t) using Ψ(t), its scaled versions, and its translated versions, and translations of the scaled versions. In this manner, I can capture the various aspects of the signal across different time intervals and frequency ranges.

So, this is the basic difference. So, this is a sine wave, and you see the dots, which mean it is ranging from negative infinity to positive infinity. Whereas our wavelet, here this is dB 10, the Dow-Bitches wavelet series has finite support, and outside this range, it is all 0. So, say this is a Daubich's wavelet, let us not name it Daubich's right now; say this is a wavelet and it satisfies all three norms: it has limited support, it has zero mean, and it has

non-zero norm, and I call this phi not t. Why φ_0 ? Not here means that this is the original mother wavelet, where I have not done any scaling or translation operation.

Now, if I translate this to 1, that is, if I delay this by 1 second, then this is my resultant waveform. If I dilate this, that is, divide it by 2, then I get this waveform. And similarly, I can translate the dilated version, and I get this waveform. So in simple terms, let us assume that a signal is a sum of these four wavelets taken in equal proportion, which is

$$\psi_0(t) + \psi_0(t-1) + \psi_0(\tfrac{t}{2}) + \psi_0(\tfrac{t-1}{2}).$$

Then these are the wavelets which are the analyzing functions and each has a coefficient of 1, and this is the wavelet analysis or the basis of wavelet analysis. And how are these analyzing functions obtained? So, this is the notation
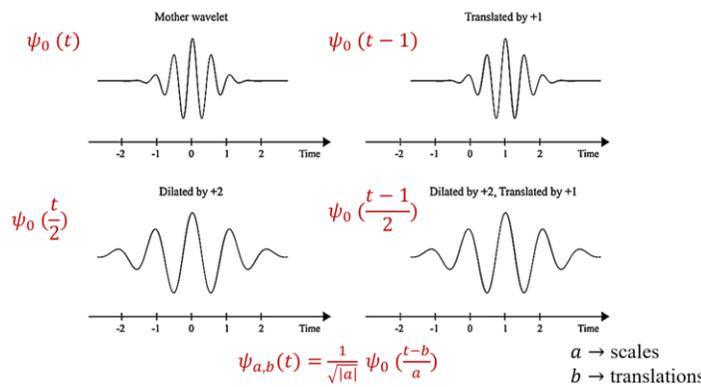
here is the translation of how much I am shifting the signal, and a is my scaling. So, what is a scaling function? So, the scaling function is different from a wavelet function, and a scaling function is typically used to define the low-frequency components of the signal. It is called a trend in the signal, and this is often referred to as the father wavelet.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \; \psi_0 \left(\frac{t-b}{a}\right)$$

So, we have a mother wavelet, a father wavelet, and then we can scale and translate each mother wavelet and father wavelet and do the linear combinations of these, and that will result in the original signal, or wavelet analysis assumes that my signal can be decomposed into the linear combinations of the father wavelet, mother wavelet, and their scaled and translated versions. So, in the context of multi-resolution analysis, scaling is associated with a space, and I call that space V_j and this is basically a subspace of the function space $L^2(R)$. What is $L^2(R)$? L^2 is the set of functions that are square integrable. So, what does this V_j mean? So, V_j is a space. What is a signal space? If you know what a vector space is, then a vector space means that if I take any two vectors from that space, do a linear combination of those two vectors, then the resultant vector will also lie in the

same space. Similarly, a signal space is a collection of signals such that if I pick any two signals and do a linear combination of them, the result will be in the same space. So, if I say V_j is a space, that means that if I take any two signals from this space V_j and do a linear combination, the result will be in the same space and these V_js form a nested sequence, do not worry if this sounds complicated.

We will analyze this slowly. So, here V_j is a subset of V_{j+1} for all j. Let us take j equal to 0. That means v_0 is a subset of v_1. is a subset of V_2; is a subset of V_3.

$$V_j \subset V_{j+1} \text{ for all } j$$

$$V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset V_3 \subset V_4$$

And what is V_0 ? So this phi t, or if I can also say $\psi_0(t)$, ψ_0 , or $\psi(t)$, it is typically the same thing. Some books will use the notation ψ_0; some use $\psi(t)$; some use $\psi_0(t)$. The concept is that this is the scaling function or the wavelet further. So, all the translated versions and their linear combinations will form a signal space. So, $\psi_0(t), \psi_0(t-1)$, similarly any $\alpha\psi_0(t-b)$, all these translations and their scalar multiplications will make a set of signals that is denoted by this space V_0.

Similarly, what is V_1? V_1 is the scaled version of $\psi_0(t)$, and that is given by a. So, V_1 is typically considered as $\psi_0(t-2)$. Similarly, you can consider V_2 as $\psi_0(t)$ by some other scale, something like that. So, these form a subspace because anything analyzed in V_0 will also be present in V_1. So, V_1 is a super space that contains V_0.

Similarly, V_2 is a super space that contains V_1 and so on. Similarly, you can have V_{-1}. So, this is what I was talking about: V_j is a linear combination of translations and dilations of $\psi(t)$. V_0 is the translation of the scaling function $\psi(t)$, which I just explained. And V_1, V_2 will be capturing increasingly finer details by dilating our $\psi(t)$ at different scales.

So now, coming back to the wavelet function. So similarly, like our ψ, we can also perform translations and dilations of our mother wavelet, which is ψ_0. And wavelets are basically difference or detail approximations at different scales. So, this denotes the high frequency content, whereas the father wavelets or the scaling functions denote the low frequency components. Now, like our scaling functions from phase v, our wavelet functions will define the spaces w_j, which are associated with high frequencies at level j.

One interesting factor is that our w_j space and our scaling function spaces vj will be orthogonal spaces. And this is what I told; so v_j will capture the low frequency and w_j will capture the high frequencies. So, if we have the c_j and d_j as our approximation and detailed coefficients, respectively, then this is how our analyzing function will look. So

v_1 is j= 1, v_2 will be j=2. So, it will capture all the respective scales and the linear combinations in each space.

$$x(t) = \sum_j c_j \, \phi(t - 2^j) \; + \; \sum_j d_j \, \psi(t - 2^j)$$

Now, pictorially, if you see it, it will be clearer. So, if we consider this circle as the linear combination of the functions that I called ψ_0, or $\psi(t)$, its translated version, and their linear combinations. So, we can define another circle, which is v_1; this is v_0, and this is all linear combinations of $\psi_0(t)$ and its translations. So v_1 is a linear combination of $\psi(1.t)$; $\psi(1.t)$ means here it is the scaled version that I refer to as 2, and this is the linear combination of all this. Now, if you see, there are some signals in V_1 that are not present in V_0, which is this annular space.

So, I can call this space W_0. This may not be a signal space, but consider that this is the set of signals that is forming the annular region. So, this whole disk, if you see, is V_1, but this annular region, if you see, is W_0. And likewise, I can create my V_2, which is the big yellow disk, and the different space that is forming or excluding this V1 is my W1. And similarly, I can have a W_2, which is this annular space, grey space and if you see the

| V_0 | W_0 | W_1 | W_2 |
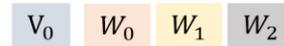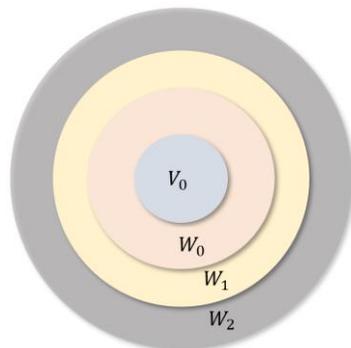
$$V_k = V_{k-1} \oplus W_{k-1}, k \in \mathbb{Z} \text{ in} [-\infty, \infty]$$

$$V_1 = V_0 \oplus W_0$$
$$V_2 = V_1 \oplus W_1$$
$$V_3 = V_2 \oplus W_2$$

whole disc as such, it is my V_3.

So, this is V_1, which is this space, this whole disc, the pink disc, consisting of the union of my V_0, which is the bluish disc, and W_0, which is the annular pink space. Similarly, V_2 is the whole yellow disc that comprises V_1. Now V_1 is this pink space, and then W_1, which is this annular ring, and this is the recurrence relation. Now we have a good amount of idea about what wavelets are. Now we will talk about the Haar wavelet, which is the simplest possible wavelet.

Now, Haar was not defined as a wavelet. Dr. Alfred Haar typically introduced the idea of this scaling and this approximation and detail function way back in 1909 and there the concept of wavelets was not present. Unless in the late 70s or probably early 80s, Dr. Ingrid Daubechies gave the solution to wavelets or the term wavelets to her famous

discovery, or rather formulation, of Daubechies wavelets and Haar wavelets can be taken as a special case of Daubechies wavelets, which is DB1.

So, what are the Haar scaling and wavelet functions? So, the Haar scaling function is this. So, it has a value of 1 from 0 to 1 and 0 otherwise.

Haar Wavelet Scaling Function $\varphi(t)$ $\qquad\qquad\qquad \varphi(t) = \begin{cases} 1, & 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases}$

Haar Wavelet Function $\psi(t)$ $\qquad\qquad\qquad \psi(t) = \begin{cases} 1, & 0 \leq t < 0.5 \\ -1, & 0.5 \leq t < 1 \\ 0, & \text{otherwise} \end{cases}$

Haar Wavelet Scaling Function

Haar Wavelet Function
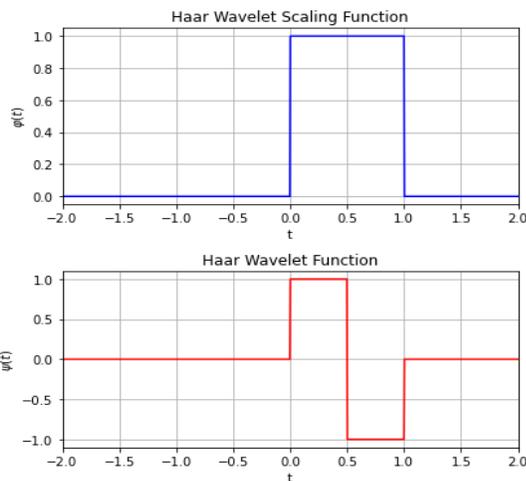
What is the Haar wavelet function? It has a value of 1 from 0 to half and minus 1 from half to 1, and it is 0 otherwise. Now, the first thing is that initially, when I said, "say we want to analyze square waves," if I use a sinusoidal basis, it will take an infinite number of sines and cosines. But if we analyze with this function, then it will be much easier because we have these sharp boundaries.

The second thing is that if you see if this is my $\psi(t)$. And if I take a signal that is just a compressed version of this, like if this is my $\psi(t)$ and if I take a signal that is just compressed, say $\psi(2t)$. And if I do a subtraction of these two signals, what do I get? So typically, this portion should be here, but there is also scaling because there is a 1 by mod a kind of thing. Technically, you will get this as a different kind of signal.

So, in simple words, this is like averaging out. If you convolve with this signal, you are averaging a portion of the original signal. Whereas if you convolve with this signal, this is basically adding this part, adding this part, and then subtracting them. So, this is like a difference in the signal. So, this captures the high frequency zone and this captures the low frequency zone.

So, this is what I was talking about. So, if I scale this function, this is my hard wavelet. So, if I scale the wavelet, what do I get? I get this; this is at a scale of 2. If I do it at a scale of 4, that is a dilation of 4. I get this waveform, and if I dilate by a factor of 1/2, that means it is shrinking. So, this is the waveform, and by using different scales and translations, you would be able to analyze any signal, of course, with the linear combination.

Each of these wavelets and their shrunken or scaled versions will have some weights, and those are your hard coefficients. So now, let's talk about the continuous wavelet transform or CWT. So, the CWT is given by the formula,

$$CWT(a,\ b) = \sum_n x[n] \cdot \frac{1}{\sqrt{|a|}}\ \psi\ [\frac{n-b}{a}]$$

and here I am talking about a discrete-time sequence. So, this is convolving it with the different wavelets. And here is why this continues. So, it is continuous in scale and translation. So, I can use any factor of a and b and then translate it. So, what is the importance of CWT? CWT, the scale factor is the major factor that determines the frequency content of a signal. So, if your scale factor a is small, that means it is determining high frequencies because they are compressed wavelets, and here you get good time resolution because you have smaller versions.

So, these are smaller versions. So, picking up highly varying or fast varying signals. Whereas large values of a are related to low frequencies because these are kind of dilated versions. So, capturing slow changes in the signal and of course, these are the scaled versions of the wavelet functions; we will also have the scaled versions of the scaling functions, or the father wavelets and in this manner, it will give you a continuous mapping of the signal's time-frequency content.

This is very useful for non-stationary signals; this is what I have already discussed while introducing time-frequency analysis. And similarly to the forward transform, we should also be able to recover our signal from the coefficients, and that is given by this relation, which is also called the inverse CWT.

$$x[n] = \frac{1}{C_0}\sum_a\sum_b CWT(a,b) \cdot \frac{1}{\sqrt{|a|}}\ \psi\ [\frac{n-b}{a}]$$

And here we have C, which is our normalization factor. $CWT(a,b)$ are the wavelet coefficients at scale a and shift b and $\psi[n]$ is the mother wavelet and one thing to be noted is that these normalization factors or all these scales are selected such that the

energy content or the Parseval theorem is valid. Our energy content is the same in both the wavelet domain and the time domain.

So likewise, so far, I have just given you an example of the haar. Similarly, there are other wavelets that are available, and they are sometimes or often called wavelet families because there is not a single wavelet but a family of wavelets. So, Haar, we have already discussed.
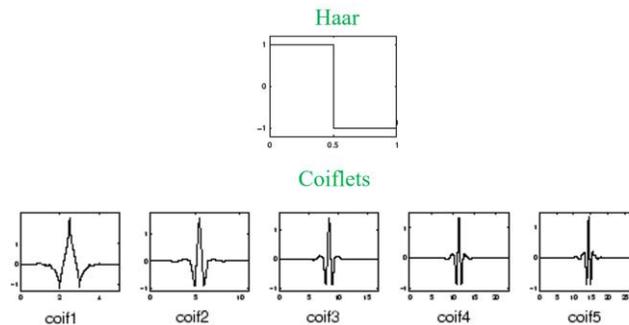

I also named this Haar wavelet after Dr. Ingrid Daubichis. Then there are wavelets like Symlets, which are variations of the Daubechies. There are Coiflets, biorthogonal wavelets, and many other wavelets like Modlet, and you can create your custom wavelets as well, which follow the rules of wavelets. So, the best thing is that whenever you are analyzing a signal with wavelets, you should try all the wavelets and see which one is best for your specific signal or application. Because each signal will have a specific pattern that may be captured by a specific set of wavelets. So, just a pictorial representation of these wavelets like this is already introduced as Haar.
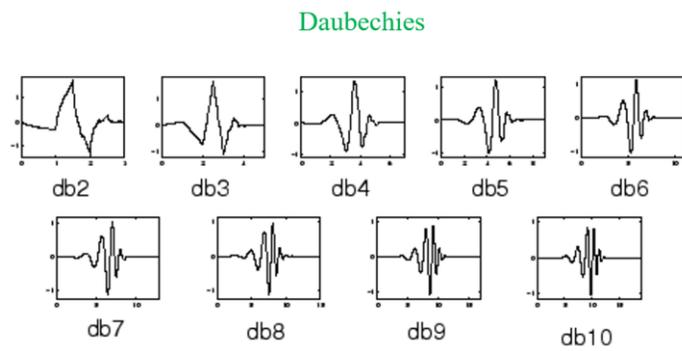
These are Coiflets, and you can see the pattern and how Coiflets 1, 2, 3, 4, and 5 are evolving. Similarly, the Dobichies wavelet family consists of these families. We have Symlets, and this is how Symlets look, and these are the biorthogonal wavelets. So, these

have been obtained from MATLAB. So, what is this CWT algorithm? So, the first step is to select your mother wavelet.
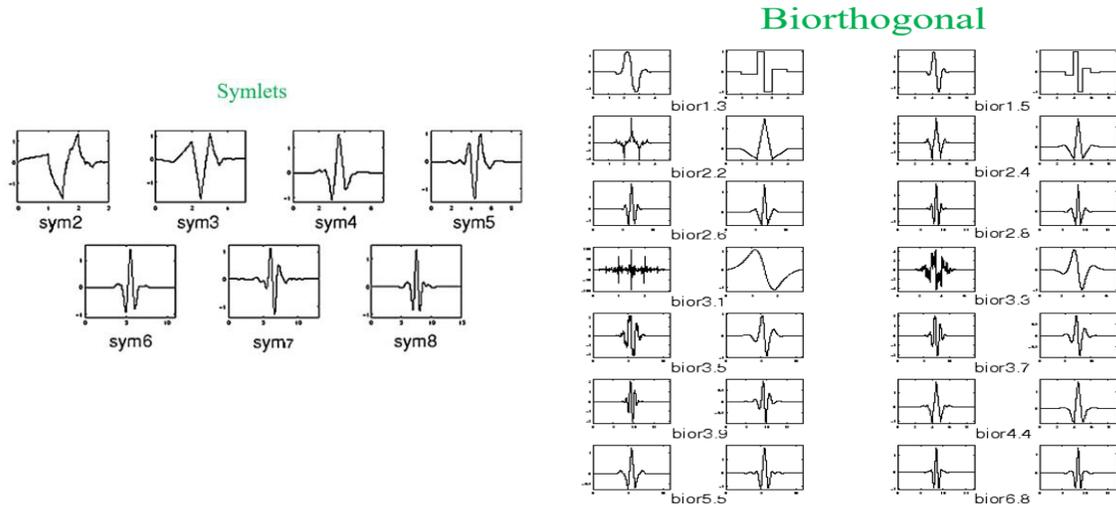
Now you can also define your own custom wavelet, but that is a challenging task. It is better to use a standard wavelet that is available. The next step is to scale your mother wavelet by a, then shift it by b, create such filters, and convolve them with the signal and each of your convolutions will result in different outputs that denote the time and frequency aspects of the signal and when you repeat it for different values of a and b, that is your time-frequency representation of your signal.

So, we will stop here and come back to wavelets later. Thank you so much.