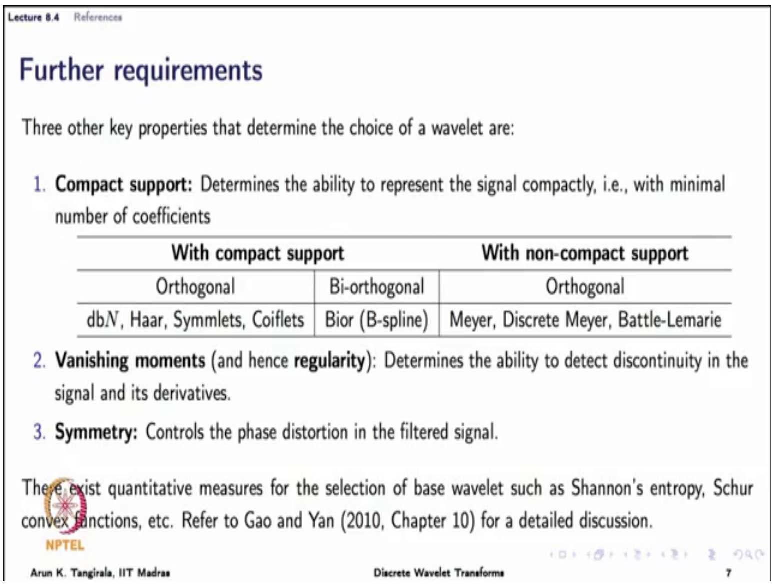
Introduction to Time-Frequency Analysis and Wavelet Transforms
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 8.4
Wavelets for DWT
Part 2/3

(Refer Slide Time: 00:15)



So, let us now return to the classical DWT. We are going to primarily restrict ourselves to the orthogonal DWT. As long as we understand how orthogonal DWTs work, the theory is more or less the same with when it comes to properties of the wavelets. The properties, that we are discussing here, of course are independent of any wavelet that we use. The property is namely, compact support vanishing moments, and therefore regularity of wavelet and symmetry. These three are independent, whether you are going to look at orthogonal wavelets or not, that does not matter, but what matters is how the compact support and the vanishing moments are tied together for orthogonal wavelets, whereas for biorthogonal wavelets, they are not tied together. This is what I was mentioning earlier, that when I want the wavelet to have certain vanishing moments, there is also a certain restriction on the width of the wavelet for orthogonal wavelet. So, let us look at this in detail.

So, the first property of wavelet in general is compact support. And what we mean by compact support? It is a, it is a technical term for the width of the signal, how effect, how long the scaling function or the wavelet is active or non-zero, right. Obviously, we want it to be as minimal as possible so that we can capture the local features of the signal in time very effectively and this effect, that is, the width of the scaling function and the wavelet essentially determines the ability to represent the signal compactly, that is, with as few number of coefficients as possible.

Now, there are a number of wavelets, that have compact support and there are a number of wavelets that do not have non, that do not have compact support. What do we mean by do not have compact support? That is, they do not die down in finite time. So, what are those examples here? In under orthogonal wavelets you have the daubechies wavelets of different vanishing moment, dbN, Haar wavelet, of course it is a special case of the daubechies, Symmlets and Coiflets. Whereas, for biorthogonal wavelets you have the B-spline wavelets, that have compact support. Among orthogonal wavelets, that do not have a compact support are the Meyer wavelets, which die down only exponentially and so are Battle-Lemarie wavelet and so on. So, that is the main, one of the fundamental properties. The compact support plays a big role in the choice of wavelet.

The second property that plays a role in choosing the wavelet is the vanishing moments. We have discussed this at length in CWT. We have discussed its effect on the regularity of the signal. Higher the vanishing moments, more smooth or regular is the signal. And the vanishing moments property of the signal determines the, sorry, vanishing moments properties of the wavelet determines the ability of the wavelet to detect the discontinue, discontinuity in the signal and its derivatives.

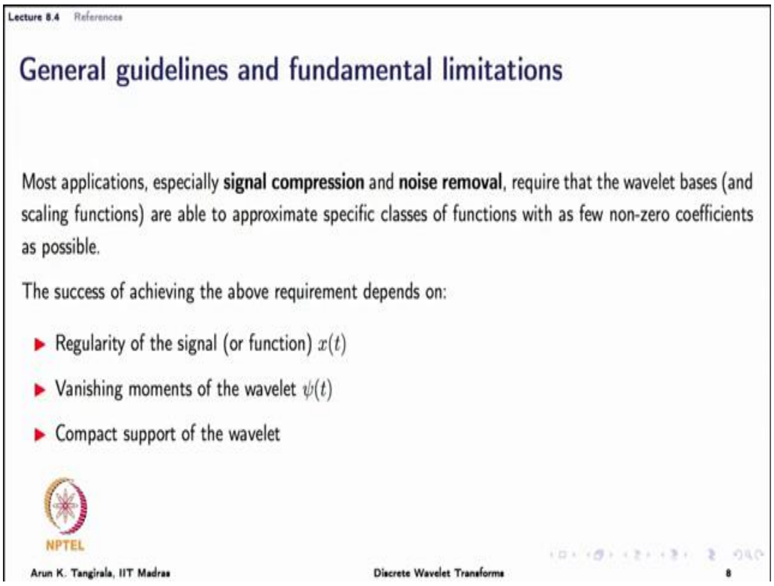
A wavelet with smaller number of vanishing moments, like the Haar wavelet has one vanishing moment. It is able to detect the discontinuity in the signal, whereas daubechies wavelets of higher order, higher number of vanishing moments are good at detecting discontinuities in the signal derivatives, and we will see examples of this in the MATLAB session. Qualitatively what we can understand is, that the vanishing moments determine the ability to detect discontinuity signal and their derivatives. In CWT, it is got to do with regularity measurements also.

Thirdly, it is a symmetry property that is of interest, as I had described earlier, the symmetry of the wavelet controls the phase distortion of the signal. Now, when I say wavelet, it is also with respect to the scaling function as well. And as I mentioned earlier, there are many applications in which I do not want phase of the signal to the distorted. Which signal are we talking about the filtered signal. Remember, I decompose a signal and then from the decomposed coefficients, I filter, I, I reconstruct the particular component of the signal, that is nothing but your filtered signal.

I do not want the sharp changes in the signal to be lost, for example, because they may carry some vital information. And if I want to really preserve that as much as possible, then I should not have asymmetric wave, I should not be using asymmetric wavelets because they lead to phase distortions. Now, these are qualitative measures. It is very hard, of course, vanishing moments has a number associated with it. Compact support also has number associated with it, which is the width of the wavelet or the filters, but there exist more quantitative measures, that allow you to select a particular wavelet for a given application. And these quantitative, quantitative measures are based on ((Refer Time: 05:40)) entropy ((Refer Time: 05:42)) convex functions and so on.

To discuss these quantitative measures is outside the scope of this course, but I recommend that you refer to this book by ((Refer Time: 05:55)). It is a fantastic book, which talks about this, in particular chapter 10, which discusses in detail the measures themselves and how these measures are calculated and certain applications. So, you will find this book very enlightening in that respect.

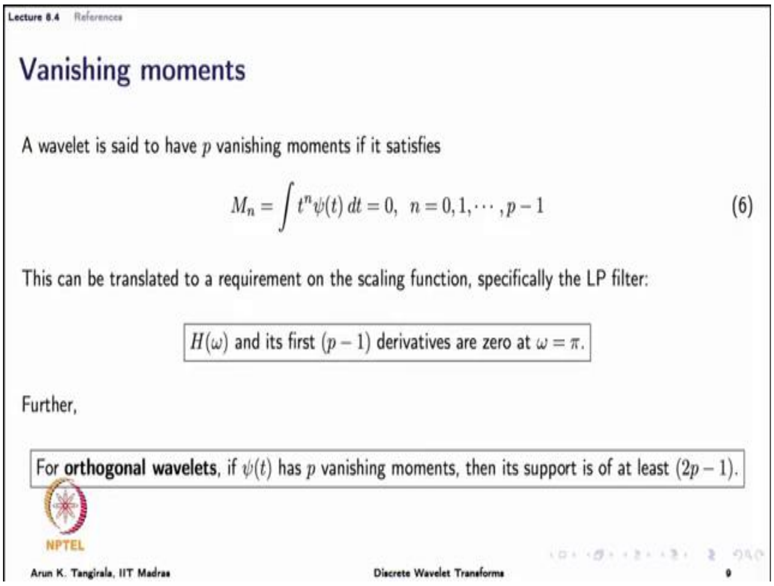
(Refer Slide Time: 06:15)



So, let us now talk about some general guidelines. Again, these are based on the properties that we just studied. In most, if you look at the applications of the wavelets, most of them are for signal compression and noise removal, that is, signal de-noising and what do they require. They require, that the wavelet basis and of course, the scaling functions basis represent the signal in as few coefficients as possible, that is, the scaling function should be able to capture the trend in as few coefficients. And the wavelet coefficients should be able to capture the abruptly changing features or the high frequency features in as few coefficients as possible.

With respect to that what we would like have is a wavelet, that has small compact support that is of very narrow width like the Haar wavelet and may be few vanishing moments. And how well we can actually succeed in representing the signal compactly also depends on the regularity of the signal itself, how smooth the signal is and so on. But that is not typically in our hands. The signal is given to us and then we have to choose a wavelet. So, we are talking about the wavelet itself.

(Refer Slide Time: 07:40)



Just to recap what the vanishing moment property is, I have given you the definition in 6. It is nothing but integral t power n psi of t dt, that is some moment and these, this moment should vanish for n equal 0 to p minus 1. And then we say, that wavelet has p vanishing moments because we have realized, that talking about scaling functions or wavelets is as good as talking about filters. One can actually translate this requirement of a wavelet to have p vanishing moments to that of the, to a constraint on the low pass filter itself. At least for orthogonal wavelets we can do that.

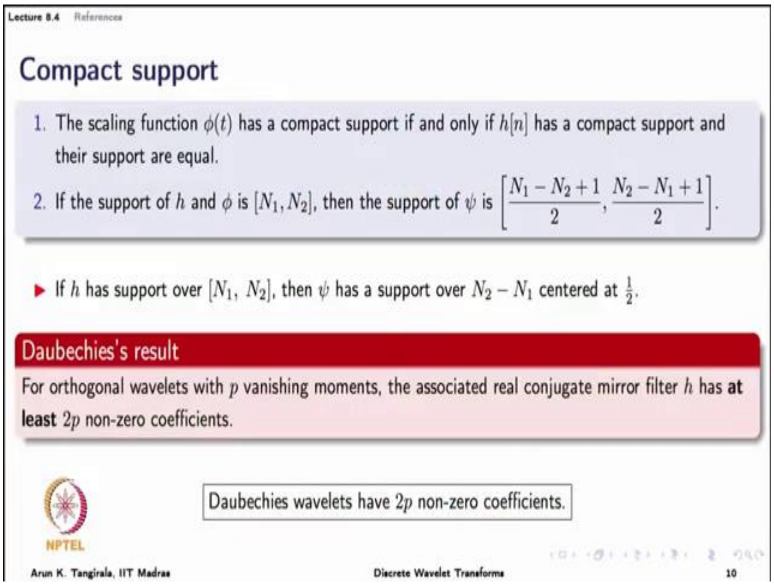
Remember, in DWT first we begin with the low pass filter, you construct phi of t and then from the low pass filter you construct the high pas filter and then we construct your psi of t. So, all the properties of the scaling function and wavelets are actually derived from the low pass filter itself. Therefore, it is natural, that we can rewrite this vanishing moments condition in 6.

In terms of constraints on the low pass filter what is constraint? The constraint is, that the Fourier transform of the low pass filters impulse response sequence, that is, a frequency response function of the low pass filter and its first p minus 1 derivatives are 0 at omega equals pi. I am avoiding proof of this. A proof of this is given in many text books, but what you should take with you is every required constraint, that you place any constrain that you place on scaling functions or wavelets can be translated to constraint on the filters and that becomes important in the wavelet design as we shall see shortly.

Now, the most important result that we shall also study later on due to the daubechies is that this vanishing moments property is tied to the compact support. In general, the vanishing moments property and the compact support, that is, the width of the wavelet and the vanishing moment property are independent of each other, but for orthogonal wavelets alone the width of the wavelet is tied to the vanishing moments. If I want, what this statement means is, if I want higher vanishing moments, I want, I, I need to have a wavelet that is wider and I, I have also talked about this even in the context of CWT, right. But this is true only for orthogonal wavelets. Biorthogonal wavelets are not really bound by this requirement.

And since we use orthogonal wavelets frequently, this statement is very important if I want to use a very smooth wavelet. What does it mean, smoothness has got to do with high vanishing moments? So, very smooth wavelet will have high number of vanishing moments and as a consequence of this statement, here it will also be wide. If I want a discontinuous wavelet, then I can offer to have a narrow wavelet and that is what is a Haar wavelet. Haar wavelet is a discontinuous wavelet. It has only one vanishing moment and in fact, it is the shortest or the narrowest wavelet that you can design, orthogonal wavelet that you can design.

(Refer Slide Time: 11:11)



So, some more details on the compact support technical details, which is very important. So, if I remember, why, why the purpose of this discussion is to be able to select a wavelet based on a compact support vanishing moments property and of course, the symmetry property. When you are looking at orthogonal wavelets, symmetry has to be sacrificed somewhat. But if you want symmetry, then you turn to biorthogonal wavelet, but the key properties are the vanishing moments and the compact support.

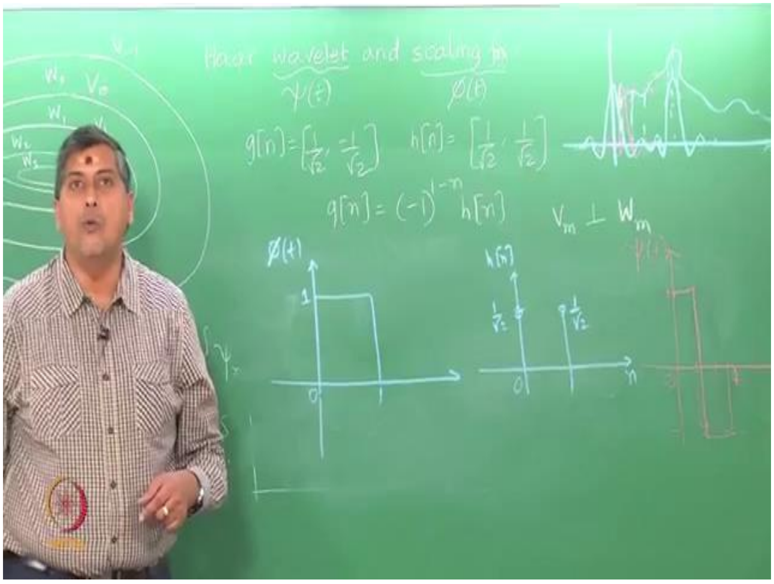
And compact support if I want to design is scaling function that has a certain width, then what results do we have in the literature? The first result is, the scaling function has a compact support, if one, if and only if the filter has a compact support. So, once again this establishes a connection between the scaling function, and the low, and the low pass filter. Remember, we have said this repeatedly, any constraint on the scaling function or the wavelet can be translated to constraints on the respective filters.

So, I want a compact support for phi of t. I want it to die down in finite time, like the daubechies scaling function, that I have, or the Haar scaling function, that I have, which dies down. The Haar scaling function is simply a box, it dies down in finite time. What does it mean on the filters? The filters should also be finite length that means, they should be finite impulse response filters. That is why, the daubechies filters and the Haar filters are fir filters. In fact, if you just recall, we saw the Haar filter, the coefficients are 1 over root 2 1 over root 2, the length is 2, right, which is finite.

And the second result says, if the impulse response has a support, what we mean by support is the duration over which the impulse response is non-zero. If the support of the h is (n 1, n 2), then of course, phi will also have the same support. That is what exactly the first result is saying.

First result is saying two things, if I want the scaling function to have compact support, then the filter should also have a compact support. And then it says, secondly, that whatever is a width of my filter will be the width of the scaling function, alright. If my filter goes from, has only two coefficients, then the scaling function associated will run from 0, will only exist in the interval 0 to 1. So, let me just show that to you on the, for the Haar scaling function.

(Refer Slide Time: 13:59)



So, how does the Haar scaling function look like? It is a box function over the interval 0 to 1. And because we normalize the scaling function to have unit energy, it has value of 1 here. And if you look at the Haar impulse response coefficients, that is, the low pass filter impulse response quotient. This is a continuous function, this is a discreet sequence. If you were to plot h of n, well, typically n start from 0 here and goes up to 1. So, I have 1 over root 2, this is one impulse response. And the next impulse response is 1 over root 2. So, the support of h of n is over 0, 1 and phi of t is also having support in 0, 1. The only difference is, that h of n is a discrete sequence, is a sequence and this is a continuous value function.

Now, what this results says is, that if phi has a support in n 1, n 2 and so does h, then the wavelet has a support over n 2 minus n 1 and is centered at half, that is, if I were to draw the associated wavelet for this, that is a Haar wavelet, then this is, this is how the Haar wavelet would look like, 0, 1. So, the support of the wavelet, I just drew to make sure, that these widths are identical at least qualitatively. The support of wavelet is over 0, 1 and centered at half. If you look at the results say, so this is exactly it reaches to the negative axis in the y-direction at 0.5 and this is true for any, any fir file, alright.

And you should ask yourself, why is this result important. Well, it is important because it tells me how to design a wavelet when I want it to have certain compact support. So, I cannot really arbitrarily choose a low pass filter and also arbitrarily design a wavelet, which has a compact support. The moment I choose a low pass filter, everything is fixed; the compact support is fixed, the vanishing moments are fixed. Everything is related to the low pass filter as far as orthogonal wavelets are concerned.

And biorthogonal wavelets, you have more freedom. You can choose your low pass and high pass filters independently, but then the synthesis filters are, or you can say low pass and the synthesis filters are, because the synthesizing low pass filter and the analyzing high pass filter are tied together and the synthesizing low high pass filter and the analyzing low pass filter are tied together, so you have some freedom there. In orthogonal wavelets you do not have much freedom. The only freedom that you have is the low pass filter. And these results help me understand how the properties of the scaling function and the wavelets used in orthogonal DWT are dependent on the properties of the low pass filter. So, I have to choose my low pass filter carefully.

And the very famous result in Daubechies, which was kind of ground breaking result in the design of compact support orthogonal filters until this result came about, until

Daubechies really worked out the this associated theory, the only known orthogonal wavelet that had compact support was the Haar wavelet. However, it was discontinuous. The Daubechies essentially established and showed how you could design continuous phi of t that is, scaling functions, that are still orthogonal and that have compact support.

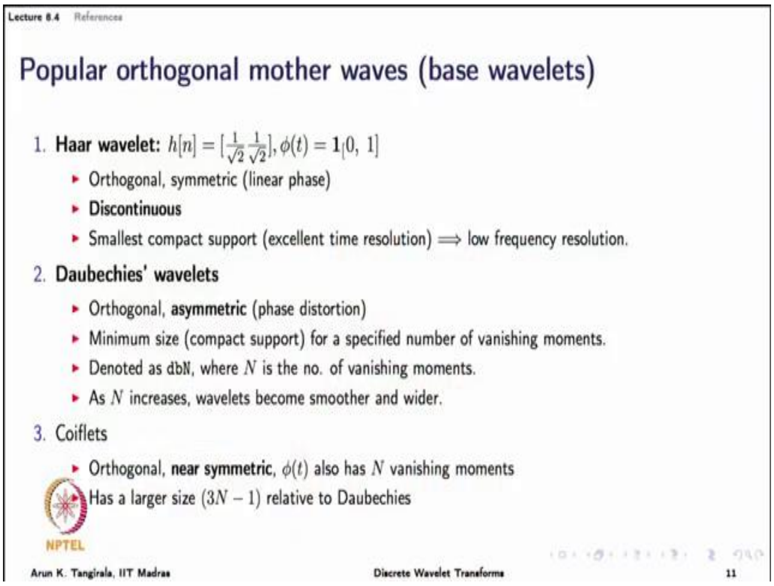
So, what does the result say? It says, for orthogonal wavelet with p vanishing moments, the associated conjugate mirror filter, that is, a low pass filter has at least 2 p non-zero coefficients. Now, this result established the connection between the vanishing moments and the compact support only for orthogonal wavelets. Again, for biorthogonal wavelets this is not true. So, this once again tells me, that the moment I choose a low pass filter of a certain width, then the number of vanishing moments is also decided. So, for the Haar scaling function, once again, what is the number of non-zero coefficients that I have? I have 2 non-zero coefficients, that means, it will have 1 vanishing moment, right.

Now, you understand the dbn, that n is a number of vanishing moments, the length of the filter is 2 n that is the general convention. But there is also another convention where n means a filter length itself. Therefore, half of that would be the number of vanishing moments. We have been following the convention, that when we say dbn, then n refers to actually p naught, the number of filter coefficient. So, if there is a confusion, then you should sort it out.

And this result says, at least 2 p non-zero coefficient, that is, the moment you say I want a wavelet with, that is, say 2 vanishing moments, what does 2 vanishing moments mean? My wavelet should be able to detect the discontinuity in the signals derivative, right. Haar wavelet detects a discontinuity in the signal itself because it has 1 vanishing moment. There are 2 vanishing moments, then the wavelet can detect discontinuity in the signals derivative. So, I want such a wavelet, let us say, for some application.

Then Daubechies result says, you have to choose a filter that is at least 4 coefficients long at least. Meaning, you can choose any other filter, filter, which has more number of coefficients, but not less than 4. And daubechies showed how you could design such minimum width, minimum compact support for a given set of number of vanishing moments that is the specialty of daubechies wavelets. They are continuous, they are orthogonal and they have minimum support, that is, minimum width for a given number of vanishing moments. That is why, they are used very widely.

(Refer Slide Time: 21:09)



So, let us just briefly go through a few popular orthogonal mother waves or as they are known as base wavelets. The first one is a Haar wavelet. We have discussed, we have been discussing this quite extensively and as I mentioned, it is orthogonal. It is only known symmetric orthogonal, which has a smallest compact support. All the other orthogonal continuous wavelet are asymmetric, as you see in item number 2 on daubechies wavelets. The only drawback if you say about Haar wavelets is, that it is, it is discontinonus, but then if you want to call it as drawback, it is drawback, but it is very usefully in detecting discontinuities. So, it is very good.

Whereas, when you come to the daubechies wavelets, of course, they are orthogonal, but they are asymmetric, which means, they result in phase distortions and they have a minimum compact support, as I said, for a given number of vanishing moments. And the 3rd point is something that we have discussed and 4th as well, as the number of vanishing moments increases, wavelets becomes smoother and wider.

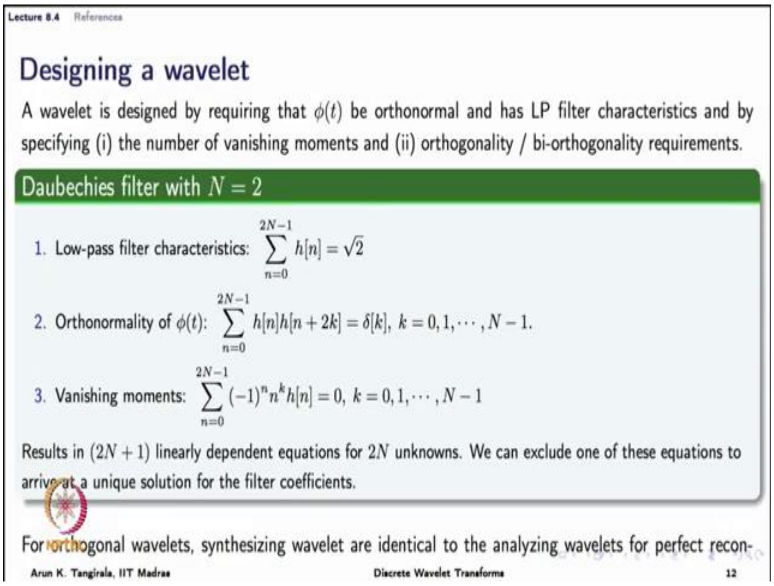
Now, after daubechies design Haar wavelets, there was also request to design daubechies like wavelets, which are near symmetric and has the same number of vanishing moments, but with an additional requirement, that the scaling function should also have certain vanishing moments. Normally, until know we have been talking about vanishing moments only on the wavelets, which tells me how the wavelet is capable of detecting a discontinuity in the signal and its derivatives. Now, we are talking about a scaling function being, having p vanishing moments and there are some applications where that is useful.

Then Coifman devised these functions or low pass filters that satisfy the additional requirement, that the scaling function has n vanishing moments. But it turns out, that it has a much larger number size, not much larger, but definitely significantly larger size related to daubechies. See, with the daubechies wavelet if I want n vanishing moments, its width is 2 n minus 1. Whereas, with coiflets if I want to have n vanishing moments for the wavelet, its width is 3 n minus 1. So, it is wider one. What does it mean? That the time resolution, if you in fact look at the haar wavelet, the haar wavelet has the smallest compact support. So, it has the smallest time resolution, but then it has wide spread, wide bandwidth.

Again, if you recall duration bandwidth result, sigma square t is small. So, sigma square omega has to be large daubechies wavelets relative to haar wavelets. As the number of vanishing moments increases what happens? The result says, that it should be wider, which means, sigma square t is increasing, but sigma square omega is decreasing. Therefore, its ability to localize frequencies is much better compared to the haar wavelet and coiflets have even better lower band, that is, frequency domain characteristic because they are wider, but then they have much poorer time resolution.

So, if you have an application where you are not worry about the time localization, but you want some time localization, not as bad as a Fourier transform, but you want very good frequency localization of the signal features, then you can use coif lets, right.

(Refer Slide Time: 24:51)



Let us conclude this lecture with talking about how to design a wavelet and then we will go on to the MATLAB session. As I said, it is essentially consolidation of what we have discussed until now. There are certain key requirements. One, for orthogonal wavelets I want orthogonality of the scaling function. Two, I want the scaling function to have low pass filter characteristics. Three, I am going to specify some number of vanishing moments and already I know, the moment I specify vanishing moments, the width of the filter is also specified. So, I do not need to separately specify how wide a filter I want or how wide the scaling function I want.

So, let us look at the case with n equals 2, daubechies filter with n equals 2 that means, 2 vanishing moments. The first criterion is, that it should be, it should, the scaling function should be a low pass filter. And we have seen this constraint in the previous lecture that the sum of the impulse response coefficients of the low pass filter should be root 2 and for orthogonality also we have seen this constraint in the previous lecture.

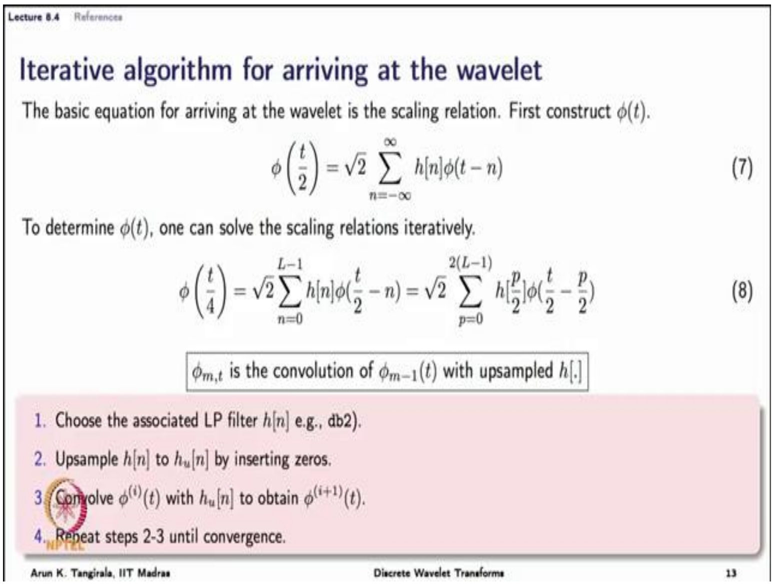
So, this is just a constraint that is coming from there again on the low pass filter coefficients, and now the constraint on the vanishing moments of the wavelet is being expressed as a constraint on the low pass filter. I have not shown you this expression before until now. We have seen the expression, integral t power k psi of t dt to be 0, k running from 0 to p minus 1. Now, we have changed that to a constraint on the low pass filter that is natural. We have been saying, all the properties of the scaling function and the wavelets can be translated to a requirement on the filter. So, that is the requirement. Again I avoid the proof, that this is indeed the condition corresponding to the vanishing moments requirement, that it is now.

I have, how many constraints do I have, how many equation to I have? I have one equation here from number 1, requirement 1. Requirement two, I have N equations because I have k running from 0 to N minus 1 and I have another N equations here. So, in total I have 2N plus 1 equations. And how many unknowns do I have? I have N filter coefficients, right, sorry, 2N filter coefficients because N is a number of vanishing moments. So, I have 2N plus N equations, 2 N unknown. So, obviously, I have more equations than the number of unknowns.

Fortunately, these are linearly dependent, that means, I can throw away one of this equations and get a unique solution, right. So, we can throw away one of the equations if you wish and then arrive at unique solution. The choice is yours. Typically, one throws away, for example, the squared. When I said k equal 0 in the orthogonality requirement, I get a constraint on the sum square of the impulse response coefficients, because when I said k equal 0, what do I get? I get sigma x square of n to be 1 because this is delta k, right. You can actually leave aside that equation or you can leave or you can leave aside the first one whichever works for you and retain 2N equations and solve for 2N unknowns.

Typically, as the number of vanishing moments grows, you may have to use a computer to solve it. I am showing you this for N equals 2, how would this equations look like for N equals 1, which is 1 vanishing moment. What is the daubechies wavelet with 1 vanishing moment? A Haar wavelet. What equations would result when with N equals 1? I am going to write that for you. I am going to get h 0 plus h 1 equals root 2, alright. And the other equation that I am going to get is h 0 minus h 1 to be 0. This comes from the vanishing moments property. In total, I will have 3 equations, but we only take 2 equations, and the solution to this is straight forward. Solution to these set of equations is nothing but what we have here, right. 1 over root 2 plus 1 over root 2 is root 2 and they are identical, like now the moment you have h, you have your g and the moment you have h and g, you have a, you, phi of t and psi of t. But how do we generate phi, that is, the scaling function and the wavelet once I design the filter is the only point that is left for discussion.

(Refer Slide Time: 29:46)



So, given a low pass filter, how do I arrive at phi of t? In fact, let me again reiterate. As far as the computation of the dwt is concerned, as far the signal decomposition, signal compression, all the application of the DWT, DWT are concerned, I do not need to compute the phi of t or phi of t and psi of t. It is sufficient to work with h, h of n and g of n. But as I mentioned at the beginning of this lecture, in many situations I would like to know how the phi of t are, the phi of t and the psi of t look like.

So, there are two ways of arriving at the scaling function given the low pass filter. One is an iterative algorithm, which uses a scaling relation. This is a famous scaling relation that the scaling functions satisfy for them to generate an MRA. What we can do is, we can rewrite. So, this is the scaling relation between v 1 and v 0. I can write another scaling relation between the subspace v 2 and v 1.

So, let me go through equation 8 here. This is the basic scaling relation between v 2 and v 1 subspaces. I can rewrite the summation here by introducing a dummy variable, which is P by, P, where P is 2n. So, I am saying, n is P by 2 and l is a length of the filter that I am looking at.

So, what does this new equation mean? This new equation means, that phi of t by 4 is the convolution of phi at a previous scale, that is, at a finer scale. Phi of t by 2 is actually at a scale m equals 1, at level m equals 1. Phi of t by 4 is the, is the basis function for v 2, that is, at m equals to 2. It is the convolution of phi of m minus 1 t with an upsample h, that part. This upsampling I leave it to you to figure out why this is upsampled h. This is not simply h of n, this is h of P by 2 because look at this. Now, I have gone from length of a, length l h to length to l minus 2, twice the length. How do you get twice length h? By up sampling h in such way, that I insert 0s alternatively.

So, if I take Haar wavelet, I have 1 over root 1, 1 over root 2 if I have upsampled any sequence. Typically, upsampling means inserting 0s like the way we do in our reconstruction from wavelet coefficients and approximation coefficients that is it. So, choose the associated low pass filter, that is, what we have in the beginning such as db 2 or whatever is known to you, upsample that by inserting zeros and start off with a guess for your phi of t, right, that is not given here unfortunately. In the first step, you can you have to initialize your phi of t. So, initialize your phi of t and set the iteration count to 0, i equal 0. So, once you have upsampled the filter sequence, convolve that with h of n to obtain next phi of t.

So, when you look at phi of t by 4 here, rather than looking upon this as a dilated phi in the computation of scaling function, you take a different perspective. What is that perspective? The perspective is phi of t by 4. So, suppose I take t equals 1, I know at scale, at level m equals 1, I know phi of t by 2. In the sense, I know, at phi at 0.5, but I may not know phi at 0.25, right. I may know t, t let us say, in, in practice I can only compute phi over a grid that is something you should understand.

So, I start off with the specification of phi only over a very coarse grid. So, I know phi only, let us say, at one point, I am, I am guessing, that phi at two points, let us say for is, is some guess, right. And then when I move to phi of t by 2, I am actually filling the missing information t by 2 would mean, suppose I guess initially, that phi is known at 0 and 1 and then the next instances I will get phi at 0.5 and 1. Then, I get 0.25, the values of phi at 0.25, 0.5, 0.75 and 1, and so on. How I am getting this equation this values of phi, through these scaling relations.

So, you should not look upon here phi as a continuous function, because I am constructing my phi, and I can only construct phi over a grid. There is no closed form expression at all, that is the most important thing that you should remember. And I will show you, once you, I will show you MATLAB code. The way we run this iterative algorithm, how things change.