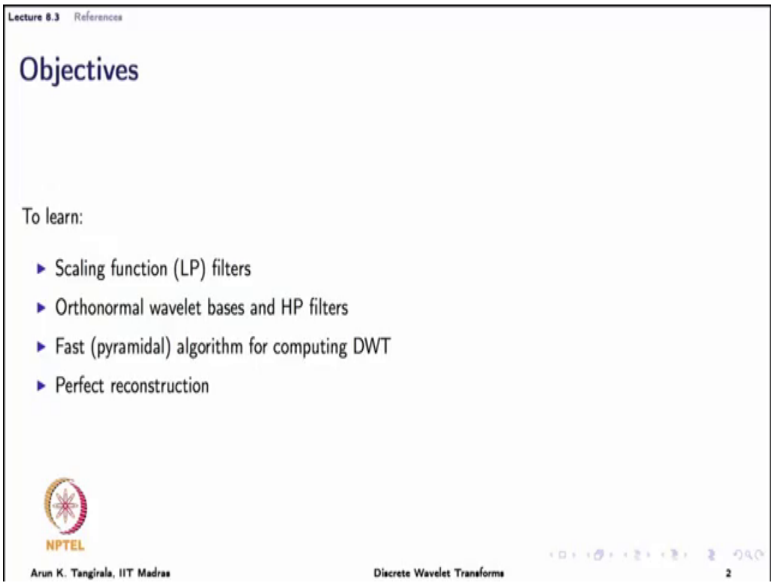
Introduction to Time-Frequency Analysis and Wavelet Transforms
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 8.3
Wavelets filter and fast DWT algorithm
Part 1/3

Hello friends, welcome to lecture 8.3 in the unit on discrete wavelet transforms. In this lecture we will continue with our discussion on multi resolution approximation. And primarily talk about wavelet filters, and also look at how the DWT is implemented using the celebrated pyramidal or fast algorithm due to Mallat.

(Refer Slide Time: 00:49)

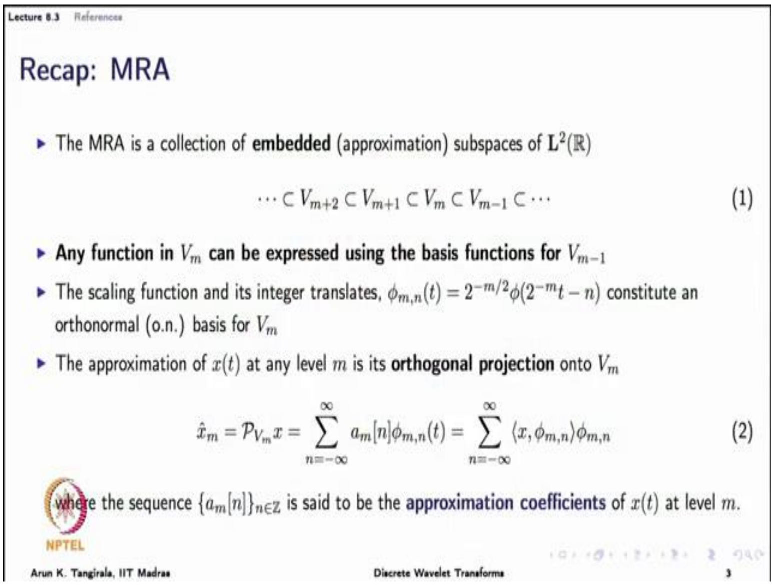


So, the specific objectives of this lecture is to discuss scaling function filters. If you recall, in lecture 8.1 we talked about DWT, as CWT evaluated on specific scales and translations. And then in lecture 8.2 that is the previous lecture we exclusively talked about approximations, the scaling functions that are responsible for generating those approximations. And in particular the multi resolution approximations generated by the scaling functions.

In the concluding slide of the previous lecture we introduced what is known as a low pass filter that is associated with the scaling function. Today we will talk more about that. And then we need to complete the picture by talking about the details. When we say details here, we are not referring to the details of the lecture, but the details of the function that are left behind during the approximations. And these details are generated by wavelet basis; in particular we are interested in orthonormal wavelet basis. And of course, associated with these are the high pass filters. So, these are complementary to that of the scaling functions on the low pass filters. We shall look at those.

And as I mentioned earlier, we will look at the fast algorithm for computing DWT, the pyramidal algorithm. And finally talk about perfect reconstruction that is how to synthesize the signal, back from its approximations and details. In this process we shall look at a couple of examples. This lecture again is largely theoretical with some examples, illustrated examples. In the next lecture I am going to show you how to compute or how to perform the discrete wavelet transform in MATLAB.

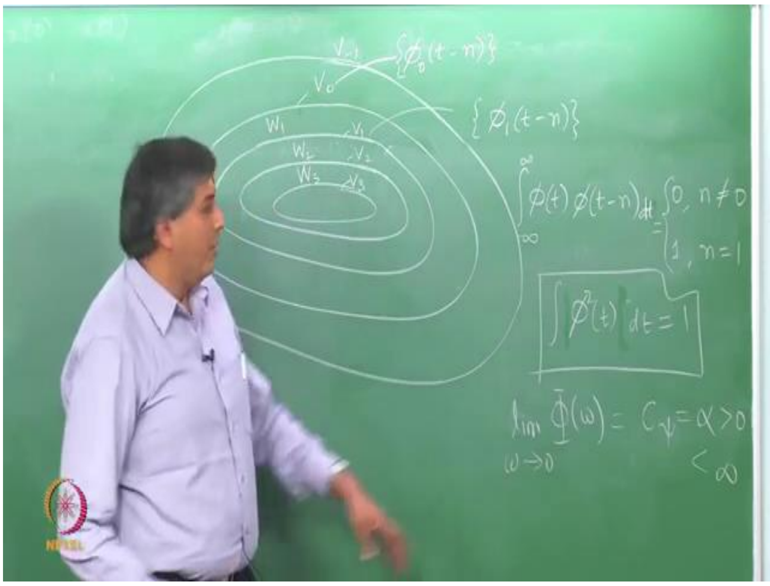
(Refer Slide Time: 02:47)



So, let us quickly recap the concept of multi resolution approximation. Essentially, multi resolution approximation is a collection or a set of embedded approximation. So, these are not collection of any arbitrary approximations, but so called embedded approximations, embedded in resolution spaces that are differing by a factor of 2. And we also call them as nested subspaces. So, if you look at any scale m or any level m, then the V m is the subspace corresponding to the approximation at that scale.

So, all functions in V m are approximations of a signal x (t) at that level m; that is how you understand the subspace. This V m is embedded in V m minus 1, because m minus 1 corresponds to a finer resolution or a finer scale. And in turn this V m contains a coarser approximation V m plus 1 and so on. You extend this nested subspaces from minus corresponding to m running from minus infinity to infinity.

(Refer Slide Time: 04:09)



So, if I were to simply illustrate graphically, this is how it would look like. So, I have here a subspace; let us say this is my subspace v 0; within this subspace is v 1. So, this entire thing is v 1. And then you have v 2 and so on. And in fact v 0 itself is contained in v minus 1 and so on. That is the subscripts are the values of m. As I mentioned yesterday, normally we assume that the signal that is given to us in the sample form is living in this subspace v 0. And we can only construct coarser and coarser approximations.

To get a finer approximation than the sample signal that I have, I need to sample faster so that I get a signal that is an approximation of x (t), the continuous time signal that lives in minus 1 and so on, v minus 1. Now, today what we are going to talk about is the gap that is left out when you move from v 0 to v 1. This space we call as v 1. And once again when I move from v 1 to v 2, the details that are left out are contained in the space w 2 and so on.

So, as you can see, these spaces are inclusive; which spaces? V 0, v 1, v 2, v minus 1, they are inclusive. And as m goes to minus infinity it would consume the entire real space. In contrast here, that is as far as inclusion is concerned w 1s, w 2s, and in fact if I draw another subspace v 3, then associated with this approximation space is a detail spaces w 3, they are all mutually exclusive. So, w 1 is not contained in w 2, and w 2 is not contained in w 3, you can design your wavelets in such a way. These vs are generated by the scaling function.

So, v 0 is generated by the scaling functions and the translates at level 0; v 1 is generated or expand by the basis functions phi 1 of t and their translates and so on. So, the specialty of each of the scaling functions is that they are orthonormal. So, that is the main feature of a multi resolution approximation. So, that is the second point that we are making; the scaling functions and its integer translates constitute an orthonormal basis for v m.

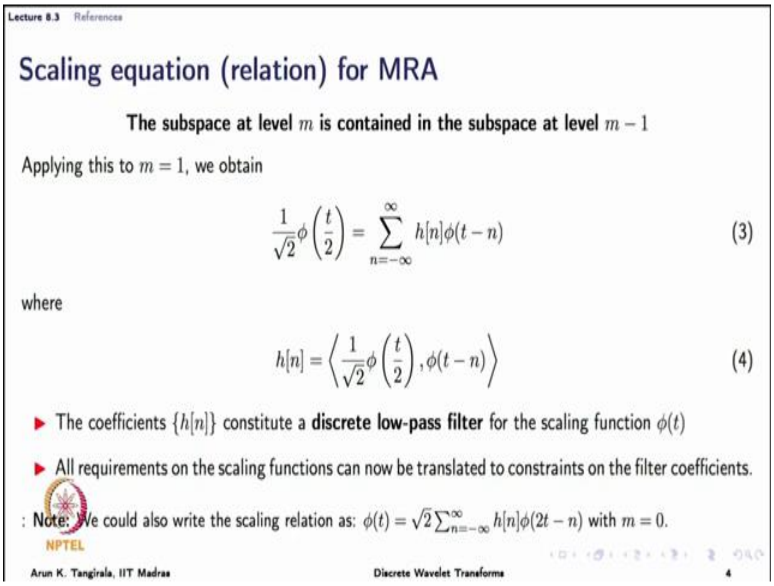
And when you want to obtain an approximation of x (t) at any level m then you implement the orthogonal projection. In the previous lecture we had look at the definition of orthogonal projection, and we also learnt that orthogonal projection yields the best approximation of the signal on to a subspace v, here the subspace is v m. So, how do you compute this?

Well, using the same expression that we had learnt earlier, x hat, or that is the convention that we follow to denote an approximation; the subscripts m denotes level at which you are approximating that is the orthogonal projection of x on to the subspace v m. Essentially, what we are trying to do is you are reexpressing your signal x in terms of the basis functions that span v m.

And the coefficients of this expansion of x on to the basis functions at span v m are denoted as, a subscript m n. So, n keeps track of the index, the translation index; and m is keeping track of the scale at which you are approximating or level at which you are approximating. And because the phis are orthonormal we know that these coefficients themselves, the, a m subscript n, which are known as the approximation coefficients; x hat is known as the approximation; a subscript m n is known as the approximation coefficients.

They are computed simply as inner product between x and the basis functions themselves. That is again you should go back and refer to the previous lecture for this expression. So, the expression that we use for computing the approximation. And the approximation coefficients themselves are computed as inner product between x and the basis functions. We will see a similar expression for the detail component that is left out as a result of this approximation, where we will have what are known as detail coefficients. And the scaling functions would be replaced by their corresponding wavelet counter parts.

(Refer Slide Time: 08:58)



One of the fundamental equations that comes out of MRA which is what we had seen in the previous lecture, is a so called scaling equation or the scaling relation. And this stems from the property of the MRA which says that the subspace at level m is contained in the subspace at level m minus 1; pretty much like saying v 1 is contained in v 0, and v 2 is contained in v 1 and so on.

So, you look at 2 successive subspaces and write the relation between the basis functions that span these subspaces. So, the basis if I choose m equals 1, then we can say v 1 is contained in v 0; and the basis functions that span v 1 are the phi subscript 1, as I have written on the board, phi 1 of t and its translates. And phi 1 of t, by definition, if you recall, phi 1 of t by definition is, 1 over root 2 phi of t by 2, right; and then of course their translates.

But, we can pick one basis function in v 1 and then rewrite what it means by saying v 1 is contained in v 0 is, as I have explained earlier, any function in v 1 can be expressed as a liner combination of the basis functions in v 0. So, any function in v 1 is also the basis function. So, I pick, 1 over root 2 t of phi by 2, as one of the basis functions and express that as a liner combination of the basis functions in v 0. The basis functions for v 0 or phi of t or phi 0 of t, we do not use the 0, phi of t and it is translates.

And the coefficients of this linear combination or your h (n); notice that h is a discrete sequence, whereas, phi (t) is a continuous valued function. So, there is big difference between these two. And how are this h computed? Well, because phi is orthonormal we know that, we can straight away write h as the inner product between the 1 over route 2 phi of t by 2 and phi of t minus n. That is the beauty here because we have an orthonormal basis expression for h (n) is very simple.

Now, what is the greatness about this equation? Well, this is a fundamental equation governing DWT because as you will see later on it allows me to compute approximations. In fact, approximation coefficients, a m of n, from, m a minus 1, or, a m plus 1, in from, a m of n, and so on, because the basis functions themselves satisfy this scaling relation, the coefficients which are computed as the inner products between the signal and these basis functions will also share a similar relation. That is the main idea. And this gives raise to the fast algorithm that exits for computing DWT.

This h (m) which is a discrete sequence can be given the interpretation of a filter, primarily because if you look at the equation 3, it is a convolution equation. So, h (m), it is a convolution phi (t) with h (n). So, you are pumping in phi (t) or injecting phi (t) into a system whose impulse response is h (m), and that produces 1 over route 2 phi of t by 2. So, that is how you view this as.

Therefore, from linear systems theory we know h (n) can be given a filter, the impulse response of a filter interpretation. What kind of a filter are we looking at? Well, low pass filter because we know that as I move from one subspace to another subspace I am approximating further; and every approximation is a low pass filtering operation; and therefore, h (n) acts as a low pass filter.

In fact, the other way of checking whether it is a low pass filter or not, is to see if h of omega which is a Fourier frequency response function of this filter, if that is non 0 at 0 frequency. If that is true then h acts as a low pass filter. It turns out that, the moment you say phi (t) is a low pass filter which means phi of omega itself, at omega equals 0 should be non-0. If you translate that condition to a condition on h, then you can show that h is a low pass filter. We will not show all of that, but we will just talk about those results.

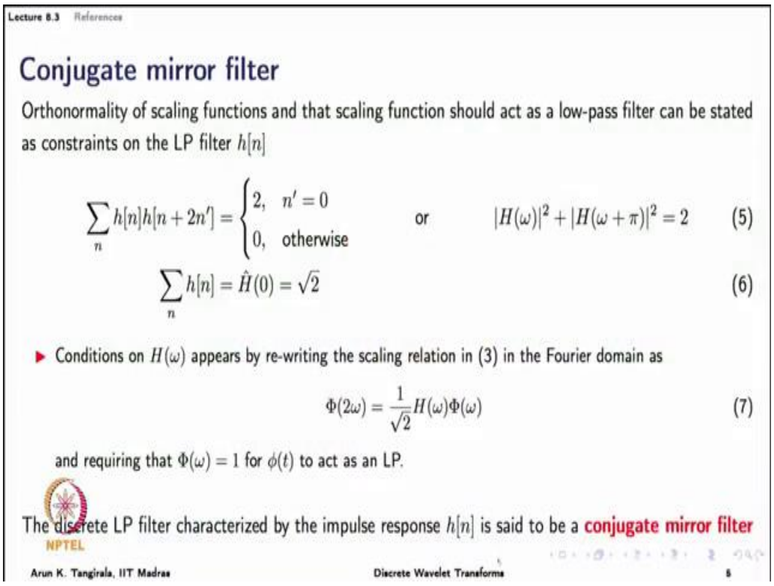
Now, the key point here that makes it complete distinct from CWT is, in CWT we have primarily used directly the wavelets, or in fact, we hardly worked with scaling functions; we only worked with wavelet functions. And we directly use the wavelet functions. We never talked about the associated filters. But now in DWT, instead of using the phi (t) to construct an approximation or the psi (t) to construct the detail, we shall use the corresponding filters instead. That is now we have h (n) as the low pass filter associated with the scaling function.

We need to only work with h (n), we do not have to really work with phi (t), why? Because, all requirements on the scaling functions can be translated to constraints on the impulse response or the low pass filter impulse response coefficients. And that is the beauty here. By the way, before even we talk about how the constraints on phi translate to constraints on h, you may also note that equation 3 could be written for any subspace, 2 successive subspaces in general.

For example, I could write the expression, I could write the basis function for v 0 in terms of basis functions for v minus 1. So, that is the equation that you see here, phi (t) is root 2 sigma h (n) phi of 2 t minus n. Notice that the h remains the same. Whether you are looking at basis functions for v 0 in terms of v minus 1, or basis functions for v 1 in terms of v 0, or in general v m in terms of v m minus 1, the coefficients or the filter that relates 2 successive approximations is the same; and this is what you should remember. But this is only true for 2 successive approximations.

Now, let us see how the scaling function constraints translate to those on the impulse response coefficients h. Now, we avoid all the proofs; this is the, probably the longest and the central result. I mean the proof associated with this result is one of the longest one if you look at Mallat’s book. And this a central result because it helps you move from the scaling function space to filter space, and thereby implementing all your approximations using some filtering algorithms, special filtering algorithms.

(Refer Slide Time: 16:04)



So, you can show that the 2 primary constraints that we have on the scaling function - one that they should be orthonormal and two that it should be a low pass filter, right. So, orthonormal would mean, integral phi (t); phi (t) minus n is 0 for all n not equal to 0. So, orthonormality on phi (t), that is at any scale let us pick v 0, orthonormality would mean this integral from minus infinity to infinity is 0 when n is not equal to 0, and 1 when n equals, sorry there should be d.

This 1 essentially means that we have normalized the scaling function to have, in fact we can say, phi square because phi is real. It is normalized to have unit energy. So, this is one of the prime requirements on the scaling function even for us to be able to generate the MRA. And then the other requirement is that phi of omega should be some constant alpha greater than 0, but less than infinity. This requirement is stems from, this condition stems from the requirement that phi is a low pass filter. So, these are the 2 prime conditions.

Now, these 2 conditions translate to the conditions that we have on the slide on the impulse response coefficients. As I said we avoid the proof, the condition of orthonormality can be written directly in terms of the impulse response coefficients or in terms of the f r f of the filter; f r f meaning frequency response functions. So, the h of omega which is the discrete time Fourier transform of the impulse response, as usual, as with a standard definition of a frequency response function, that is what we have.

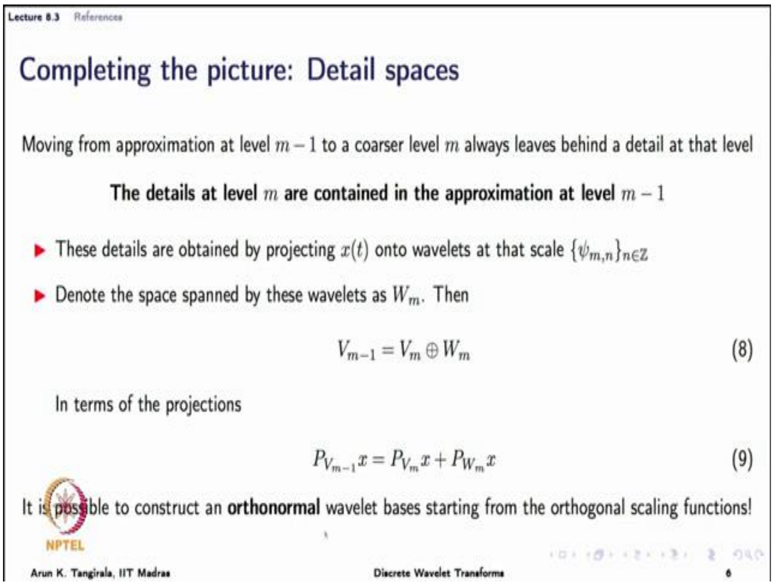
So, any sequence that satisfies this conditions, and then there is another mild condition that is sufficient, not really necessary, constitutes or qualifies to be a low pass filter that can be associated with the scaling function. So, I can start off with this low pass filter, and I can generate my scaling function. Now you understand how the Daubechies wavelets are arrived at. Remember, when we talked about scale to frequency conversion in CWT, we talked about Daubechies filter and we said there are no close form expressions for Daubechies filters; instead, they are generated by specifying the low pass and the high pass filter. So, this is a connection now.

The Daubechies filters are arrived at by additionally imposing some requirements on h. But the basic idea is to design h first, and then from this h you design your phi (t), or you compute your phi (t) iteratively. That we will talk about in one of the subsequent lectures - how to arrive at phi (t), how to compute the scaling functions starting with the low pass filter. What we are looking at with now? At now is how constraints on phi (t) translates to constraints on h.

Now, the constraint here that you see in terms of h of omega for orthonormality of scaling functions can be arrived at by writing the scaling relation that we saw earlier in the Fourier domain. So, this the scaling relation that we had in the time domain. If I take the Fourier transform on both sides then I can arrive at the equation 7 which is relating the Fourier transforms of the scaling functions. Because, you are taking Fourier transform of phi (t) by 2, the Fourier transform becomes phi of 2 omega; whereas, the Fourier transforms of phi (t) would be phi of omega, and t minus n would be phi of omega multiplied by modulating factor or an exponential.

And then requiring that phi of omega equals 1 at omega equals 0, infact we should read phi of 0 equals 1, not phi of omega equals 1. At 0 frequency, this scaling function should have a Fourier transform value of 1, so that the scaling functions axis are low pass filters; that is it. So, instead of working with phi (t), now we will work with the impulse response coefficients that satisfy the conditions given in 5 and 6. And this discrete filter is known as the conjugate mirror filter. We will learn the reason for this name when we talk of reconstruction shortly.

(Refer Slide Time: 21:13)



As I said one of the prime things that we shall do in this lecture is to complete the picture that is we have talked about approximations until now, now we are going to look at detail spaces. And as I explained to you earlier, in this sketch we had v s and w s, now we will ask what basis functions allow me get those details? Well, I already know the answer; they are the wavelets. But can I come up with some special properties of the wavelets; what are the, how do, how are this wavelets related to the scaling functions? I do not want them to be loosely, or I want them to be tied together that is the main thing.

Now, that tie comes through the fact that the details at level m are contained in the approximation at level n minus 1. So, this is to say why is this true? Because, when I move from approximation at level m minus 1 to a coarser level m then I always leave behind some detail. That detail is associated with the approximation at level m. So, we call that as a detail at level m.

And therefore, we say that the detail at level m are contained in the approximation at level m minus 1 because m minus 1 is a finer approximation. How are these details computed? Well, they are computed by projecting x (t) on to the basis that span this detail space. And the basis that span this detail space are the wavelets because they are essentially band pass filters and so on. And so, the wavelets are that scale and their translates; and that is what this means here, psi of m comma n, n belonging to z.

So, psi of m is the basis function at scale m and it is translates, together constitute the family of basis functions for what we call as a W m which is a detail space. We have introduced this notation earlier as well. So, denoting now the detail space at level m by W m, I can write V m minus 1 as the orthogonal sum; this symbol here means orthogonal sum. I am actually, or you can say, W m is the orthogonal complement of V m. At the moment we are not imposing orthogonality, but what we are saying is that we sum up these 2 spaces not in the regular addition sense, but in a linear algebra sense. Then we obtain V m minus 1.

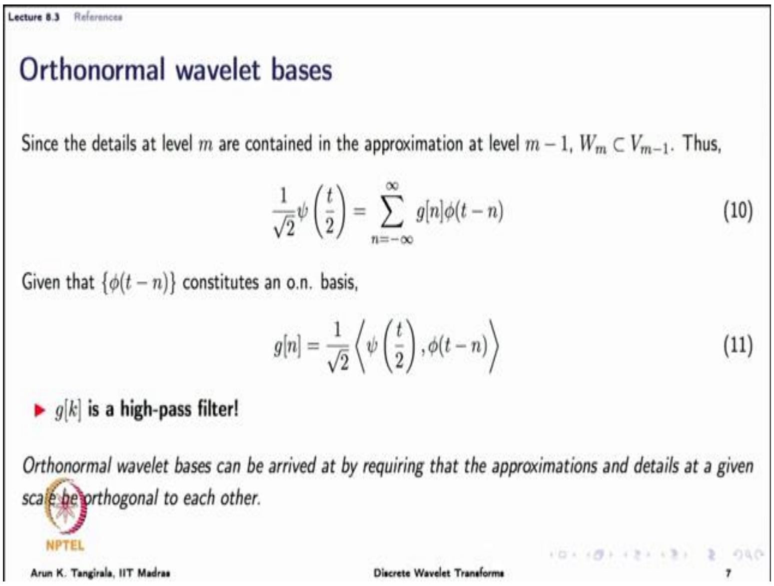
Intuitively what this means is, when I take the approximation and detail at level m and put them together in a particular manner, not by direct addition, then I get the approximation at a finer level V m minus 1. In terms of the projections, the approximation at m minus 1, we saw this relation yesterday as well; x hat of m minus 1 is x hat at m. So, x hat at m minus 1, is a projection of x on to V m minus 1, is x hat of m which is the projection x on to the m plus, the detail at scale m or level m which is the projection x on to w. This is a same equation that we have seen in the previous lecture as well.

Until now, we have not talked about any orthonormality of this wavelet basis. Once again to compute projections, it is ideal to have this basis as orthonormal. And the fact is there it is possible to construct an orthonormal wavelet basis as well from the orthonormal scaling functions; so, which is nice. I have already orthonormal scaling functions.

By the way, when we talked about MRA, one of the things that we did not mention is the result which says there exists an orthonormal basis that for computing the MRA or constructing the MRA; that is very important; we have avoided those theoretical proofs. But it is proved already that given an MRA you can always find some orthonormal scaling function that will allow you to generate them.

So, we have the orthonormal scaling function, now we want orthonormal wavelet basis. And the fact is we can find such an orthonormal wavelet basis; very important. Until, in the equations 8 or 9, there is no need that the basis, wavelet basis are orthonormal but now we want the orthonormality.

(Refer Slide Time: 25:54)



And how do we achieve that? This is the subject of the next couple of the slides. The first thing that we note, the first point that we note is W m is a sub space of V m minus 1. W m is contained in V m minus 1; that is the details are contained in the, details at level m are contained in approximation it level V m minus 1. Once again, as we argued earlier, what this means is, the basis functions for W m can be expressed as a linear combination of the basis functions in V m minus 1.

So, the basis functions for W m, let us say m equals 1, so, W 1 is a sub space of V 0. The basis functions for W 1 or psi (t) by 2 and it is translates; and the basis functions of V 0, as usual, phi (t) and its translates. So, I pick one of the basis functions for W 1 which is psi (t) by 2; in fact, 1 over root 2 psi of t by 2. And express that as a linear combination of the basis functions for V 0 which is phi (t) and phi of t minus n, its translates.

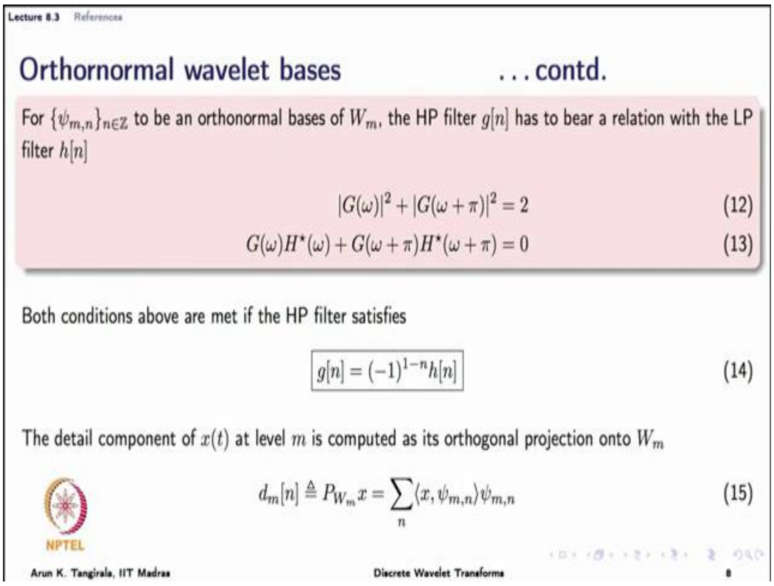
These coefficients now are complementary to that of h; that is the beauty here. Earlier I had h, and now I have g; the notation is also chosen so that you can see the complementary characteristics of g and h. Naturally, you should expect now g to act as a high pass filter for 2 reasons – 1, because I know wavelets are high pass filters, and 2, because the details are complementing approximation. And approximations are characterized by low pass filters, therefore details should be characterized by high pass filters.

And in terms of the imposition on psi, for psi to be a low pass filter, we have already discussed that many times in CWT. The value of the Fourier transform of the wavelet at omega equals 0 should be 0. Now, this g can be computed once again as an inner product between the wavelet at scale 2 or level 1, and the scaling functions at level 0 because phis are orthonormal I have a nice expression for g. Once again this is what to do with the projection property associated with the orthonormal basis. So, g is a high pass filter.

Now, what remains to be done is to make sure that my wavelets are orthonormal. How, what kind of constraints now exits on g? As we said, as we observed earlier, constraints on the scaling function can be translated to constraints on the low pass filter. Now, requirements on the wavelet can be now translated to constraints on the high pass filter. So, h for phi, g for psi; whatever I want for phi, I can dictate certain conditions on h; whatever I want for psi, I can dictate certain conditions on g.

What do I want now? I want psi to be orthonormal, and I want psi to be a high pass filter, right. So, we will just look at the orthonormality requirement. The high pass filter requirement is naturally satisfied the moment you formulate, express a scaling function as a linear combination of the, sorry, the wavelet function as a linear combination of scaling functions. We are more worried about orthonormality.

(Refer Slide Time: 29:17)



You can once again show that the orthonormality requirement for the wavelet at any scale m, at any scaling if I take psi of m subscript m comma n, if I want it to be orthonormal, if this basis should be orthonormal, then the associated high pass filter should satisfy these conditions given in 12 and 13. Once again I avoid the proofs and directly giving you the conditions.

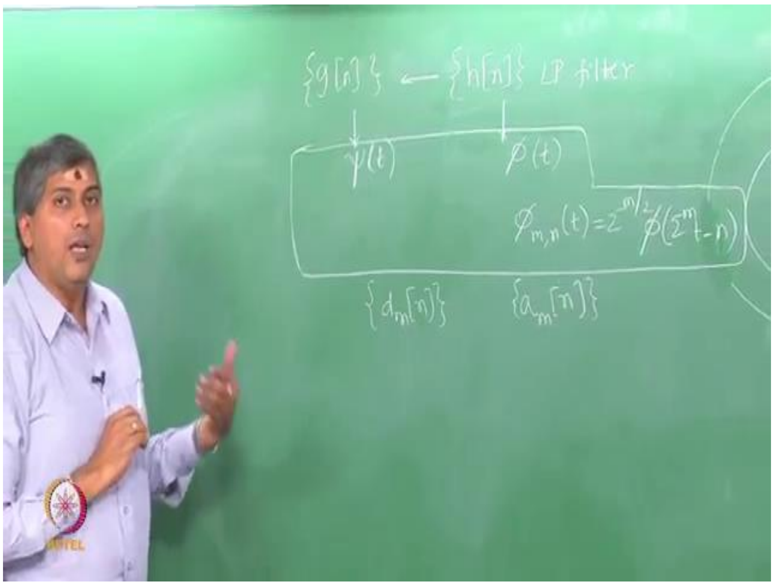
Now, these conditions are in terms of the frequency response function of the high pass filter. It turns out that you can show both these conditions 12 and 13 are satisfied, the moment the high pass filter shares this beautiful relations with the low pass filter. It is like a reflection with the sign reversed; all yours, that is well known in filtering literature, the standard filtering theory literature. If I want to construct a high pass filter, I take a low pass filter and I can reflect it and so on.

But, this is a special reflection. And this g and h are now tied together. This tying together is coming due to the fact that I have required W m to be one - that is 2 conditions - W m to be a subspace of V m minus 1, and W m to be orthogonal to V m. So, these 2 conditions have produced this 14 on the filter coefficients. So, now, you should see the big picture; we start off with CWT where you only talk about wavelets, then we move to DWT where we saw that the approximations that I construct has some special features, MRAs. So, now, I am interested in the scaling functions.

And then we said, instead of scaling functions I can talk about the associated low pass filters, and then we said we want to complete the picture by talking about the details as well which are produced by the wavelets. So, the scaling functions and wavelets are complementary; the low pass filters and the associated high pass filters are also complementary.

When I want orthonormal wavelet basis then these filters, obviously, intuitively, have to be tied together, and that is given a equation 14. That is it. So, I start with h. All I have to do is I have to design h; then g is fixed. Once h and g are fixed I just compute my approximations; I mean, I compute my phi (t) and psi (t), and then compute my approximations and details. So, this is, let me just illustrate that work flow for you in a schematic.

(Refer Slide Time: 32:10)



So, I have h; I start with the design of this low pass filter. And if I want orthonormal decomposition that is approximation and detail should be orthogonal, then that condition automatically fixes the associated high pass filter. So, from h (n) I design g (n), the high pass filter. And this low pass filter generates the scaling function, from this high pass filter also I can generate the wavelet, not only at scale 0, but at all scales. Of course, once I know phi (t) and once I know psi (t), the wavelets the scaling functions and wavelets or other scales are automatically known, because of the relation, right. I know phi m, n of t is nothing but 2 to the minus m by 2 phi of 2 to the minus m t minus m, I know this; and likewise for the wavelets. So, once I have this, and once I have these as well, the scaling function lets and the wavelets, then I can construct a m of m, the approximation coefficients at any scale, and the associated coefficients, detail coefficients at any scale.

What we shall learn is, a very important fact, that we can compute these approximations and details, completely bypassing this step. We do not have to really compute the scaling functions and wavelets at all. Although the expressions for a m and d m are given in terms of projections of x on to the respective basis that is phi of p and c or phi m comma n and c m comma n, we can avoid this step completely and that is what is the beauty of the fast algorithm that exists for DWT.