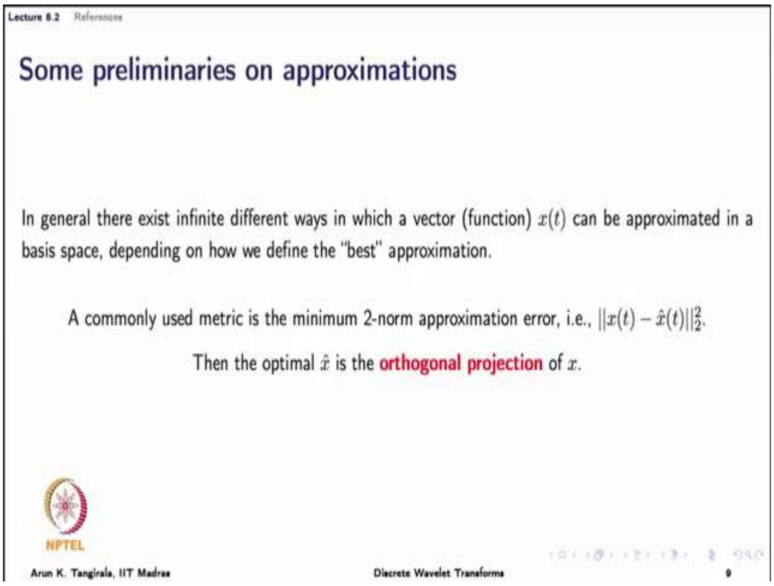
Introduction Time-Frequency Analysis and Wavelet Transforms
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 8.2
Orthogonal scaling function bases and MRA
Part 2/2

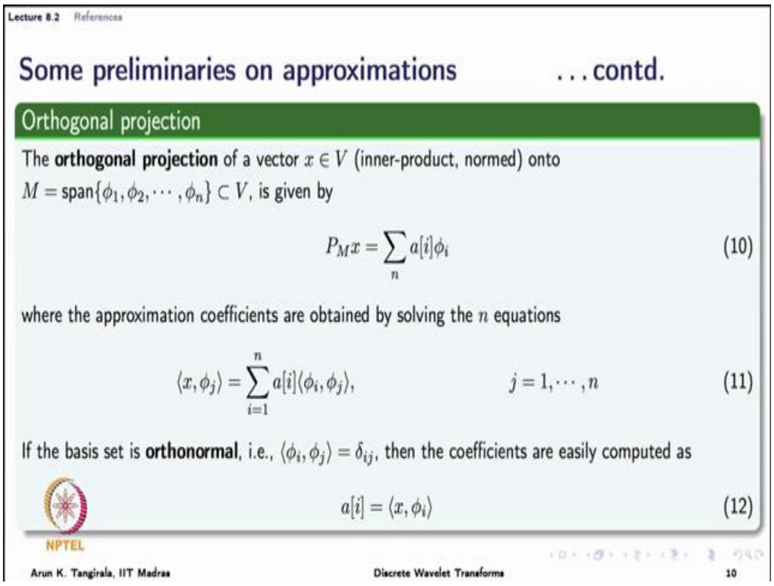
(Refer Slide Time: 00:15)



This multi-resolution approximations are very special, because a generate approximations in a nested manner, and in a hierarchical manner. Not only that, this multi resolution approximations which are very important in computer region are also useful in developing a fast algorithm for computing DWT. So, let us look at the brief theory of MRA, and that is what will conclude the lecture with. Before even we talk about approximations, all along we have been using this term approximation, but it is time to just review, formally what is mean by an approximation in vector algebra or function analysis. In general we understand what an approximation is? It is some estimate of the signal, and it will always be in some error. So, any approximation that I construct will have some error. Therefore, there are infinitely different ways in which I can construct an approximation, depending and what I want the error to satisfy; that is what characteristic error should have.

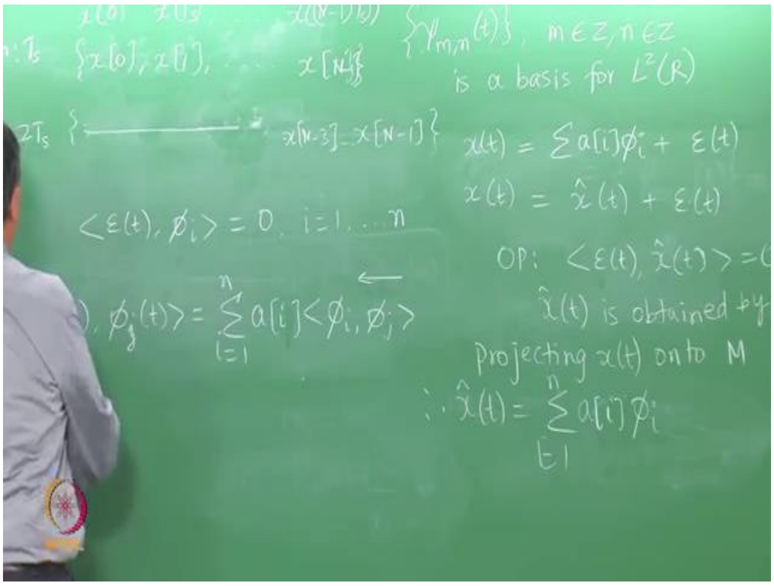
From an optimal view point, that is if I want an optimal approximation, I would say find an approximation that minimizes the zero norm of the error or one norm of the error or even the infinity norm. So, I have x of t, I have x hat of t as an approximation. For the moment we are not talking about MRA, which is talking about some approximation; x minus x hat is some error and I can say fine x hat such that the error has minimum two norm or minimum one norm and so on. Generally, what is of interest, is the approximation that yields that best two norm, or the minimum two norm of the error, and that is what we mean by best approximation here. In other words I want an approximation x hat of t of a signal or a function x of t; such that the two norm, or the energy of x minus x hat is a minimum. It terms out that the solution to that problem, is that the approximation x hat optimal x hat is nothing but the so called orthogonal projection of x on to what, on to the basis that generates x hat. First let us understand what is orthogonal projection, and then the statement will become clearer.

(Refer Slide Time: 02:38)



What is an orthogonal projection? If I take any vector x which is in a vector space for which the inner product and the norms are defined, then the orthogonal projection of a vector x projection is always about projecting a vector on to some other subspace. So, I am here projecting or vector x for which a norm exist, for which I can compute inner products, on to as subspace. So, recall the analogy that we had walking on the road or a person, 3 d object who’s shadow is on the two dimensional road or the floor or wall and so on. So, you can think of x living for example, in a three dimensional space, and your projecting that on to two dimensional space, so n is 2 here for us. In general x lives in some space, and we are projecting it on to a low dimensional subspace n, which are spanned by this basis functions. Recall in unit 2 we have given some definitions basic definitions of linear algebra, what is a span what is a basis and so on. Then the orthogonal projection is nothing but linear combination of the basis functions, with the coefficient of the linear combination derive from the signal. So, p subscript m of x always denotes orthogonal projection of x, on to the subspace m which is span by the basis function phi. So, this is nothing but x hat actually. So, p m of x that is what we meant earlier, but will talk about right now we are only talking about projection. So, if I give you x, and if I give you the basis functions phi, the orthogonal projection is calculated with a help of equation 10, where the coefficients are computed by solving this n equations in equation 11. How do I get this equation, let me quickly show you that.

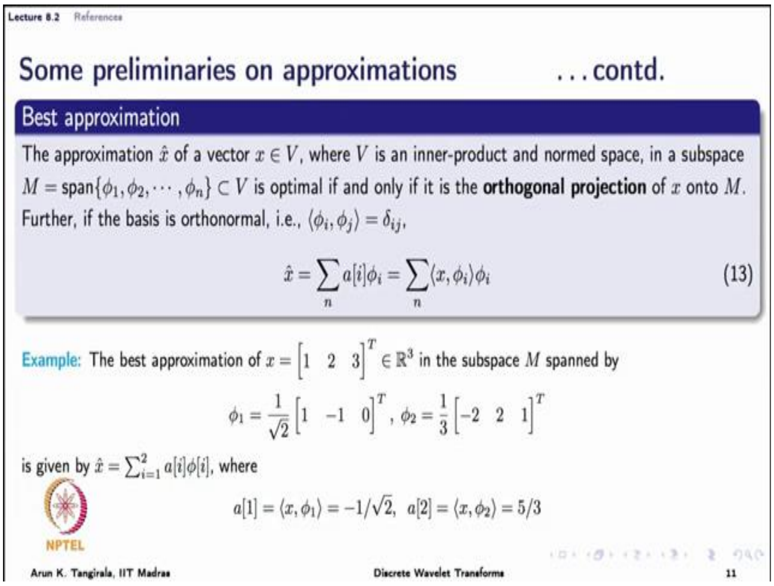
(Refer Slide Time: 04:54)



So, I have x of t living in some subspace, some space we and we are breaking this sub into an approximation plus some error all right orthogonal projection has this specialty we call orthogonal passion o p if the main feature of the orthogonal projection is that the inner product between epsilon and x hat is 0. So, the inner product between epsilon and x hat is 0 and x hat itself is generated by projecting x on to m and m itself is span by this basis functions phi. Therefore, this is a span of the basis. So, we are saying, some combination of the basis functions in this space m will generate an approximation of x. Now I plug this in to here, and then I use this result. So, what I do is, I take inner products of x with x hat of t. The only step that I do is, instead of taking the inner product of epsilon with x hat I take inner products of epsilon, with a basis functions are generate x hat. It is one and the same; therefore, I require this. And once I write this equations I get those n equations that you seeing equation eleven. So, what I do is, I take inner product of x of t with phi i. On the right hand side I have. Sorry let me use a different dummy variable here so that you are not confused.

The second term goes to zero by our orthogonal projection requirement, and therefore, I get these equations, and I like j run from 1 to n. Now, the big advantage of choosing an orthonormal phi. In the original statement the phi's are not orthonormal necessarily, but they are a basis, which means we have to be linearly independent. But now if I require that the phi's are orthonormal; that is the inner products or either between similar basis function, identical basis function is 1, or if they are not identical it is zero, then the coefficients can be readily computed; that is the big advantage of orthonormal. If you look at equation 11. So, I have n equations and n unknowns, and I have to perform matrix inversion, but if the phi's are orthonormal then I can solve coefficients individually, and that is a big advantage of choosing an orthonormal basis. So, the equation twelve is a much simpler version of eleven, because your only solving for the individual coefficients there are no inversions involved. And that has come about, because I have assume phi's to be orthonormal.

(Refer Slide Time: 08:52)



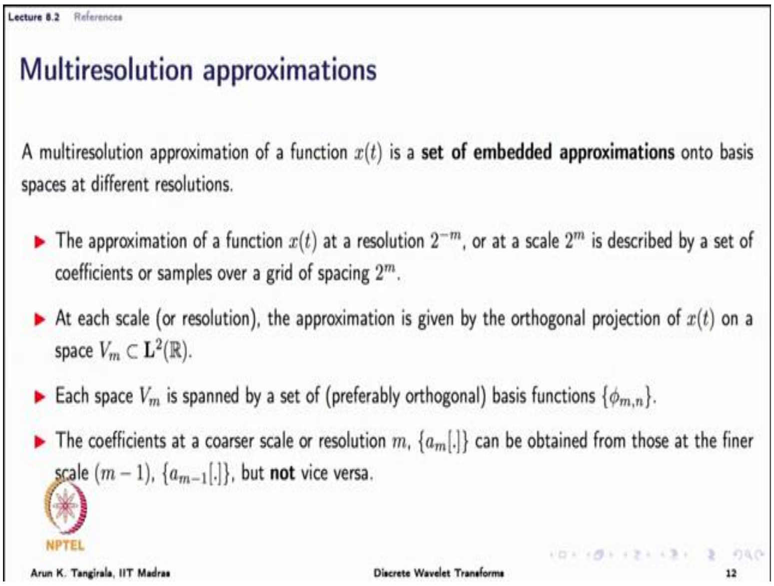
Now, let us written to the best approximation situation, the statement essentially says what we have said earlier. The best approximation of a vector x using basis functions;

that is of x living in a vector space, obtained by projecting it on to a subspace m, which are spanned by basis functions phi is, it is best only if it is orthogonal projection; that is if I want to if I have a signal x and I want to construct a best approximation of that, using some basis functions. Then the best approximation in the lease square sense, in the two norm error minimizing sense is nothing but the orthogonal projection itself. Therefore, the best one is x hat sigma a i phi i. Exactly what we are seen on the board earlier, all I have done is I assume the phi's orthonormal then I can use equation twelve in place of a i's to construct my orthogonal projection and that is what exactly I have. So, let us look at an example quickly. So, the best approximation of x; suppose I have a vector.

All this is equally applicable to both functions and vectors. I have a vector x here; 1 2 3 transpose, living in the three dimensional real space. And I want to approximate this vector with two basis functions only, which means now the. And the two basis functions are orthonormal as you can verify this span a subspace of r 3, because they are only two basis functions here, this span only the two dimensional subspace they are orthonormal. So, they are basis, but for a subspace of r 3.

So, I want to obtain an approximation of x, on to this two dimensional subspace span by phi 1 and phi 2, by the above result the best approximation is this here, where a i's are given by this co-efficiencial. In other words approximation itself is x hat. These a i's are called approximation. In the world of wavelet transforms one of the key things that you should be a distinguishing is between approximations and approximation coefficients. a i's are call approximation coefficients, they are called a signal representation in the subspace m, and x hat is a approximation. So, that is what we will look.

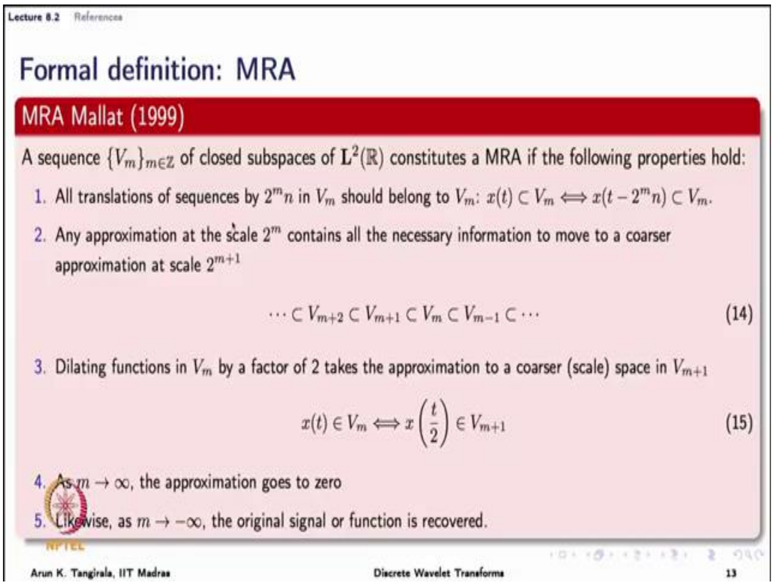
(Refer Slide Time: 11:16)



Now we will extend is idea of approximations to multi resolution approximations. What is multi resolution approximation. It is also an approximation, but it is a collection of approximations, over different spaces. Just now is spoke about approximation on to some subspace m, but now we are going to talk about approximations on several subspaces, and what is special about is subspaces. The subspaces or span by basis functions, that are defined over coarser and coarser grids. The grid from finer to coarser grids, and the grids spacing varying by factor of two. So, there are some important features of an MRA; one that the approximation of the function at a resolution, 2 power minus m is describe by set of coefficient or samples over a grid of spacing 2 power m. So, the coefficients that you are computing at a certain resolution are not defined over an arbitrary grid. They are defined over a grid which has spacing 2 power m.

And at each scale the approximation is given with orthogonal projection naturally, because we are seeking the best approximation. So, what I do is, to construct MRA’s, I take the basis functions which are the scaling functions, at level m, and project, compute the orthogonal projection of x of t over the scaling function at that scale. And further each space; this v m essentially is span by the set of preferably orthogonal basis function. In general this basis functions at any scale, that span the subspace at any scale need not be orthogonal, but we just not saw, what is the advantage of choosing orthogonal basis function. And the last point is something that we have mentioned earlier. The coefficient at a coarser scale can be obtained from those at the finer scale. We have seen that in terms of approximations earlier with the sine wave example. So, in the sine wave example I did not show you coefficients, I showed you approximations.

(Refer Slide Time: 13:30)



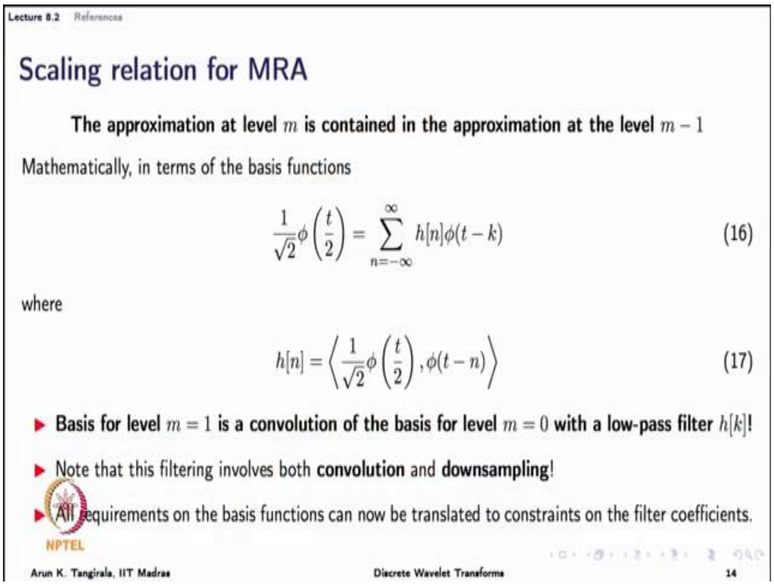
So, let us conclude the discussion with the formal definition of MRA. We will take the essence of this and the main equation, and then move along. The first point says that if I look at collection of subspaces, when can I call it as an MRA. Remember multi resolution approximation is a collection of approximations of a signal on two subspaces. So, what subspaces will qualify, to generate multi resolution approximation, is what Mallat had prescribed long ago, and this was the turning point in the DWT. The subspaces should satisfy certain properties, and those are the five key properties that have been listed here. All of them are important, but I will highlight the points 2 and 3. The 0.2 says, that the subspaces should be embedded, which means as I move down m, then I should get finer and finer approximations.

And I say move as I increase m; that is I move up in the level of m then I get coarser and coarser approximations, and these approximations are nested, the subspaces are nested. Therefore, approximations can also be set to be nested. What nested means is given in equation fifteen; that is 3 really tells you what is nested means, it says that as I move from v m which is the subspace at level m, to a coarser subspace at v m plus 1, then if x of t belongs to v m then x of t pi 2 that is dilating. Here x is some function, it needs not be the signal that we are talking about.

So, if I take any function that lives, that is span by the basis functions in v m. Then a dilated version of that function, the moment I dilated I am moving to a coarser subspace; that is what means. And this is the fundamental equation for us, both fourteen and fifteen are fundamental in deriving, what is known as the scaling relation in DWT. Then point four and five are essentially saying, that as I move down, as I move up; that is coarser and coarser and coarser approximation scales, the approximation essentially goes to zero. And as I moved down; that is go and finer and finer and finer and finer scales the approximation will converts to the original functions x of t by…

So, I will be able to recover the original signal. This is ensures that spanning essentially four and five together are essentially saying, that as I move from m equals minus infinity to plus infinity, I am spanning the entire real space of finite energy functions that is what it means. So, this is the last point that will make in this lecture, and in fact, this is also the transition for the next lecture. How do I uses MRA now, what kind of basis functions to I use. Can I use any scaling function. So, look at what has happened, we talked about DWT in the previous lecture, and now we have moved to scaling functions and we are taking about approximations in particular MRA. Now, what we want to know is, given this properties of MRA, how do I choose my scaling function. Do I always work with scaling functions, or is there a better way, how may I go to implement.

(Refer Slide Time: 17:20)



This the slide that allows to transit from scaling functions to filters, low pass filters. And the basic idea, is again derived from the property of MRA. The approximation at level m, is contained in approximation at a finer level m minus 1. How do I construct approximation at level m. Well by projecting the signal on to the basis functions that span level m. What are the basis functions that span level m, then those are phi m of t, and it translates. Now this is the generic statement; m can be 0 and m can be 1 and m can two whatever. So, if I choose m as 0 then or m is 1 let us say then what I am doing is I am looking at a scale two, and when I and then what this statement means is the approximation at level 1, is contained in the approximation at level zero; that means, in terms of basis functions. The basis functions that span the subspace v 1, are can be generated by the, or can be spanned by the basis function that span v naught.

What are what are the basis functions that span v naught phi of t and it is translates. What is the basis function for v 1 the subspace v 1 phi of t by 2, because v 1 is a subspace that is essentially a dilated version of the functions in v naught. So, phi of t, if I move if I am looking at m equals 1, then I have phi of t by 2 at v 1 the basis functions and phi of t for the basis functions for v naught. So, this statement essentially means I can take any function in v 1 include in a basis function, and express it as a liner combination of the basis functions at a finer resolution. By fixing m equals 1 what I have here is, v 1 and v naught, and I am relating the basis functions for v 1. I am expressing the basis function in v 1 in terms of the basis functions at v 0.

(Refer Slide Time: 19:48)

In general you can also write these relations. Let me do that on board for you. So, what that statement essentially means is, v 1 the basis functions for v 1, can be spanned by basis functions for v naught. In general basis for v m can be span by the basis for v m minus 1. So, what is the basis for v m. The basis for v m the subspace v m is nothing but your phi 2 power m n of t, and basis for v m minus 1 or phi 2 power m minus 1 n of t. So, what we have done is, we have said m equals 1 and written that equation. You said the basis. Well although written here phi 2 of 2 power m, we know that phi 2 power m is nothing but scaled version of phi. So, phi of t by 2 for example. In fact, when I move from phi of t to phi of 2 power m, then that is nothing but 1 over root 2 power m times phi of 2 power minus m t minus n, because m equals 1 we have 1 over root 2. I am picking any basis functions for v naught. In fact, I can write another basis function also in v naught which will be that translate of 1 over root 2 phi of t by 2. I am just saying one basis function any basis function, in v naught which is 1 over root 2 phi of 2, can be expressed as a liner combination of the basis for v naught, v naught basis functions are phi and that is what we have essentially.

So, these h are the coefficients that allow me to go from v naught to v 1. And how are these coefficient obtained. They are obtained by taking the inner products of phi of t by 2 with phi of t minus n; that is again by our assumption that phi’s are orthonormal. So, all I have to do is. That is the beauty here, because we have assumed phi’s to be orthonormal. There is a slight mistake there in equation sixteen. It should be t minus n not t minus k. Because phi’s are orthonormal, the coefficients are obtained by simply taking the inner product between 1 over root 2 phi of t by 2 and phi of t minus n.

Now, what you see here, the equation sixteen is actually looking like a convolution equation, but there is a difference. On the left hand side you are not obtaining some function of t, but your obtaining function of t by 2. So, there is a convolution followed by some down samplings. In fact, exactly down sampling by factor of 2. So, this is different from the regular convolution and this is what we will notice in the pyramidal algorithm. But for now what is important is to remember that these coefficients h are nothing but our low pass filter coefficients.

Why these are low pass filter coefficients, because what we are doing here, as we go from v 0 to v 1 we are constructing another approximation, but a courser approximation, and approximation are always generated by low pass filters. So, the most important point, is that instead of talking of phi of t's which becomes boring or difficult later on, we shall talk in terms of these h which are the coefficients. So, by specifying the low pass filters coefficients h here, I would be indirectly specifying phi of t itself, and will look at all of this in the next lecture; that is essentially the connections between scaling functions phi of t, and the filters coefficients h of n. And also remember we have talked about approximations, but to complete the picture we need the details.

So, we need to bring in the wavelets as well. So, how do I bring in the wavelets, so that the picture is complete, and that is where we will end up looking at the high pass filters, and eventually put together the complete picture for us. So, will meet in the next lecture, where will talk about connections between scaling functions and filters, and also the connections between wavelets and filters, and then discuss the pyramidal algorithm or the fast algorithm; that is used in computing DWT.