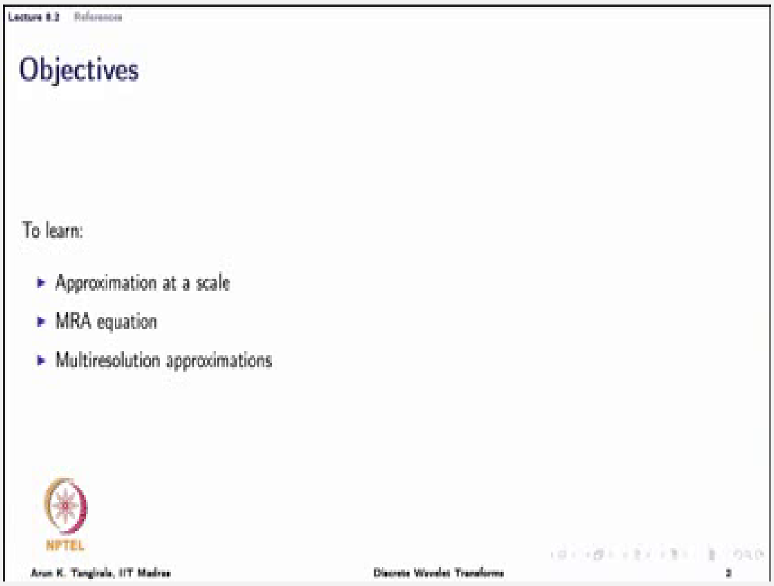
Introduction to Time-Frequency Analysis and Wavelet Transforms
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 8.2
Orthogonal Scaling Function Bases and MRA
Part 1/2

Hello friends, welcome to lecture 8.2 on the topic of Orthogonal Scaling Function Bases and Multi Resolution Approximation. So, we are in the 8th unit, which is on discrete wavelet transforms. And this is perhaps, one of the key lectures that you should watch out for, apart from 8.1. And then, the one that is coming next, which is 8.3. Because, it really establishes the base for discrete wavelet transform, the theory and also the implementation.

And also it establishes a connection between the wavelet functions, scaling functions, multi resolution approximations. And ultimately, the filters that we are going to use in implementing DWT.

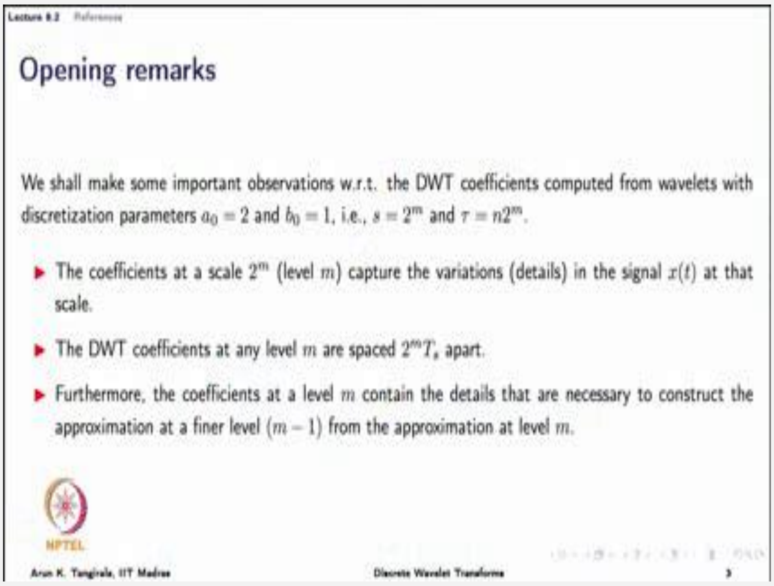
(Refer Slide Time: 01:08)



In this lecture, particularly we are going to talk about approximations at a scale. And you must have started noticing this, even in 8.1. That the moment, we started looking at DWT, we have begin to talk about approximations and so, on, rather than time frequency analysis. And will understand, why that is so. And then, we look at the multi resolution approximation equation. The basic equation, that one encounters in multi resolution approximations.

And then finally, we will move on to the formal study of multi resolution approximations. In that process, we will also take a quick review of, what is an approximation, orthogonal projection and so, on.

(Refer Slide Time: 01:58)



So, before we really dwell into the math, let we make a few opening remarks. The first thing that I want to mention, which is not really listed here in the slide, is a point that I just made. When we were looking at CWT, we were talking about time frequency analysis and using complex wavelets and. so, on and talking less about approximations. It was more about time frequency analysis. Of course, we did talk about similarity detection, regularity measurements and so, on, but, not so, much about approximations.

Now, the moment we have switched over to DWT, we are talking about approximations and so, on. The prime reason is, as we move from CWT to DWT as you must have noticed even in the previous lecture. We start to generate basis functions for spaces, the real number spaces or complex numbers spaces and so, on. Whereas, that was not the case with continuous wavelet transform. The analysing functions that we were using in CWT. Really it do not generate basis.

But, in DWT we have this great opportunity of generating basis functions, in particular orthogonal basis functions. And they are ideal for generating approximations. More interestingly they are ideal for generating, what are known as multi resolution approximations. And that, we shall understand both with an example and with the theory, later on in this course. So, the other points that I want to make ahead is, the coefficients that we compute. Now, we talk about coefficients. Just as the way, we talked about Fourier coefficients and so, on.

So, the transform that I compute with a wavelet, essentially yields me coefficients. And when I am computing these coefficients at a scale 2 power m or at a level m, what they essentially do is, they are capturing the details in the signal at that scale. And of course,, here scale refers to resolution. So, it does a more appropriate term in DWT. Of course, scale will always denote a frequency band. But, in DWT it is always a good idea to keep this scale versus resolution relationship, in mind as well.

So, as I am going up the scale, I am looking at courser and courser resolution. So, the analogy of a map, geographical map that we had discussed long ago, in one of the introductory lectures really applies here. So, as I am going up the scale or up the level and actually looking at a courser picture of the signal and what these wavelet coefficients give me, are the details that are left out or the details of the signal at that course at particular scale.

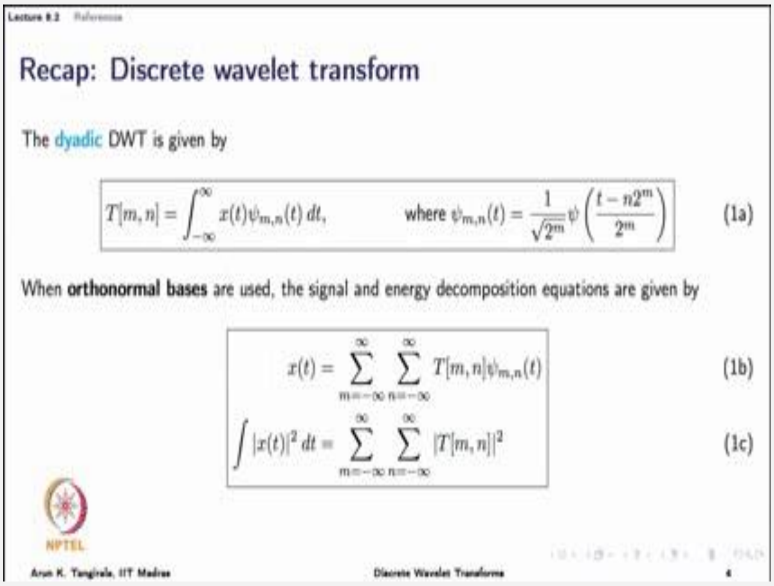
And the second point is that, the DWT coefficients at any level m are actually spaced 2 power m T as a part. Now, what does this mean? So, T s is a sampling interval of the signal. As I am moving up the scale, what is happening? I am actually losing out on the time resolution, that is what this means. So, if I move from scale 0 to scale 2 or level 0 to level 1, then the spacing in time falls off by a factor of 2. So, I have the sample signal with a sampling interval of T s.

And then, look at the DWT coefficients at scale 2 or level 1. Then, the spacing between the coefficients now is 2 T s. And then, at level 2 it is 4 T s and. So, on, which means I am losing out on the time resolution. At what cost? I am obtaining, now better frequency localization. So, that time frequency tying is always present. The only difference is, now it is in the background. And we have talked about it quite a lot, as to how wavelets really scarifies on the time localization to get an improvement over the frequency localization.

And the third point is, which is very important for multi resolution approximation. When I look at the wavelet coefficients at two different levels, what is the difference between them? Well, when I am looking at the coefficients at some level m, they contain the details that are necessary for me to move to a finer scale, which is at m minus 1 and construct an approximation at that level.

So, it is essentially what you doing is, what we are saying here is, as I move down the scales or let us say, I move from coarse to finer and I am collecting all the details. Those details become necessary to construct approximations at a finer scale. All of it may be a bit abstract for us, right now. But, as the lecture unfolds and we look at a few examples, things become a lot more clearer, hopefully.

(Refer Slide Time: 07:20)

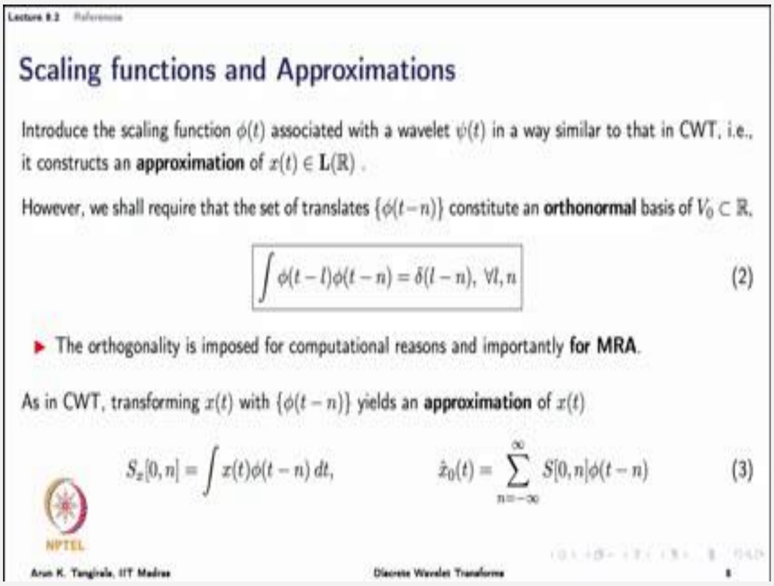


So, let us begin with a recap of, what a discrete wavelet transform is, the dyadic discrete wavelet transform. We will not use the dyadic prefix all the time. It is understood that, it is dyadic. Essentially, s is 2 power m. That is, what dyadic means. And that is given by the equation that, we have seen in the previous lecture. It simply the wavelet transform, the standard CWT, but, now with evaluated with this specific wavelets, which are now positioned at the dyadic scales and translations proportional to these dyadic scales.

At this point, that is in equation 1 a, the family of wavelets that are generated by spanning m and n. That is, when m and n take on different values. They are not necessarily orthonormal. However, if we design a wavelet, such that the family of wavelets are ortho normal. Then, as we have seen in the last lecture, the reconstruction equation of the synthesis equation and the energy decomposition equation takes some very nice forms.

Notice and recall from lecture 8.1. Whenever, I use an ortho normal basis, then the reconstructing or the synthesizing function is the same as the analyzing function. And that is, what is reflected in 1 b. And 1 c gives me a Parseval's decomposition equivalent of the energy, as we saw in the Fourier domain as well. So, things are much nicer when we make sure, that the wavelets that I am looking at for different values of m and n are an ortho normal.

(Refer Slide Time: 09:02)



Now, let us move to scaling functions, like the way we did in see CWT because, we are going to talk about approximations a lot more, than we did in CWT. So, we introduced scaling function way ahead, in DWT. What is a role of this scaling function? It is exactly the same as in CWT. We call from CWT lectures that, the role of the scaling function is to construct an approximation, if you look at from an approximation view point.

If you look at from a filtering view point, it act as a complementary low pass filter to that of the wavelet filter, which is a band pass filter. So, either way it is complementing the wavelet function. It complements from a functional view point, it provides approximations, whereas, wavelets provide the details. And from a signal processing view point, this scaling function act as a low pass filter, whereas, the wavelet acts like a band pass filter.

So, we will look at the scaling function now, but with an additional requirement. Remember, we just mentioned that ortho normal basis have certain advantages, which is that the synthesis equation become simpler. The energy decomposition is a lot more easier to interpret. And later on, we will learn also that the calculation of the coefficients, whether it is scaling function coefficients or wavelet coefficients. They also become a lot easy, numerical.

So, we shall require that the set of this scaling functions and it is translates. So, we are still looking at scale 1, that is m equal 0. Exactly, the way we did in CWT. When we introduced scaling function CWT, we used the reference point s equals 1. So, here also we are using the reference point s equals 1. And denoting, this scaling function by phi of t. Now, additionally requiring that at this scale the scaling function and it is translates, constitute an ortho normal basis.

And that is specified in equation 2. The delta here is a standard chronicle delta function, which assumes a value of 1, when the index is 0 and 0 otherwise. So, again as in CWT, we can say that transforming x of t with this scaling function and it is translates, essentially constructs an approximation or a low pass filter version. We call these coefficients now. S at 0 comma n, 0 stands for the value of m. Because, we are looking at scale 1, m is 0 pretty much. And then, ((Refer Time: 11:50)) induction being indexed by n.

So, this you can say is the discretized equivalent of s comma tau. The continuous wavelet or the continuous wavelet transform equivalent of the scaling function that, we saw in CWT. So, what is this here? It is simply, the transform of x of t with phi of t minus n again, the same way as we have seen in CWT. When I try to recover this signal from the scaling function coefficients, then what I end up with is, an approximation of this signal..

Why do I end up with an approximation? Look, I am transforming x of t. If everything is perfect, then I should be able to recover x of t. But, everything is not perfect here. Why? Because, if you look at the top of this slide, we have assumed that the signal leaves in the real space. But, in this L to R actually, it should be L to R. What this mean is, finite energy signals in the real space. What about the scaling functions? What space do they span? Do they span the entire real number space? No, not really, not necessarily.

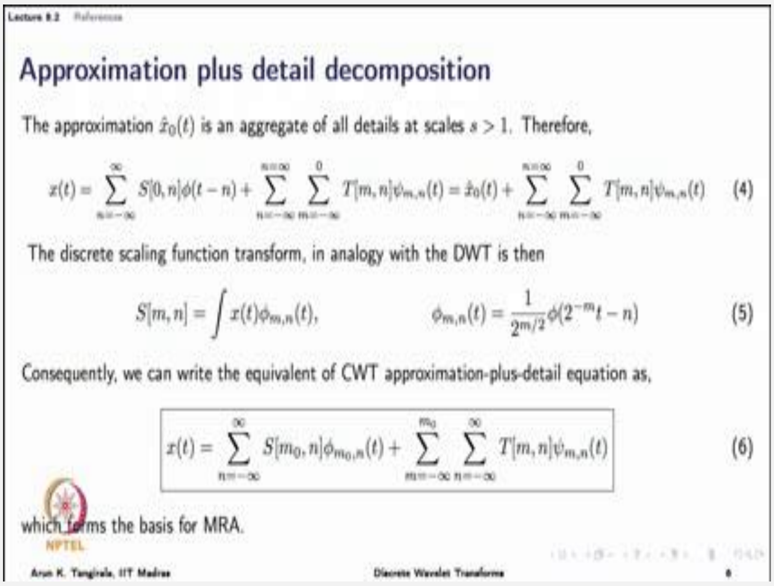
They span a sub set of the real numbers space. And we denote the sub space by V naught. So, phi t and phi of t minus n. This entire family, span only a sub space of the real numbers space, whereas, the signal itself leaves in the larger dimensional space, which is the full real number space. Therefore, when I am computing the coefficients here, what these coefficients are essentially is, resulting from projecting the signal onto your low dimensional space.

So, x of t leaves in a larger dimensional space. Phi of t leaves in a lower dimensional space. Therefore, when I recover;; obviously, from those coefficients, I am going to lose out the information. So, it is like this. As an example, I am walking on the road. Any person, who is walking on the road is a three dimensional object from a vector algebra view point, from a topology view point.

When you look at the shadow of this person on the road or on the wall, when light is impinged on the person. Then, the shadow leaves in a two dimensional space. So, what is being done is that, the three dimensional object is being projected on to the two dimensional space. Not that here, phi of t actually spans a two dimensional. But, the example essentially says, that you are projecting an object, which leaves in three dimensional to a lower dimensional space.

So, I give you the shadow. Now, if I ask you to recover the three dimensional figure, just from the shadow, you will; obviously, not be able to recover the three dimensional object but, only an approximation of that. So, that is what exactly is happening. So, x naught of t is the approximation of x of t at scale 1 or m 0. If I want perfect recovery, then I need to have basis functions that is also span the entire real space and we will talk about that.

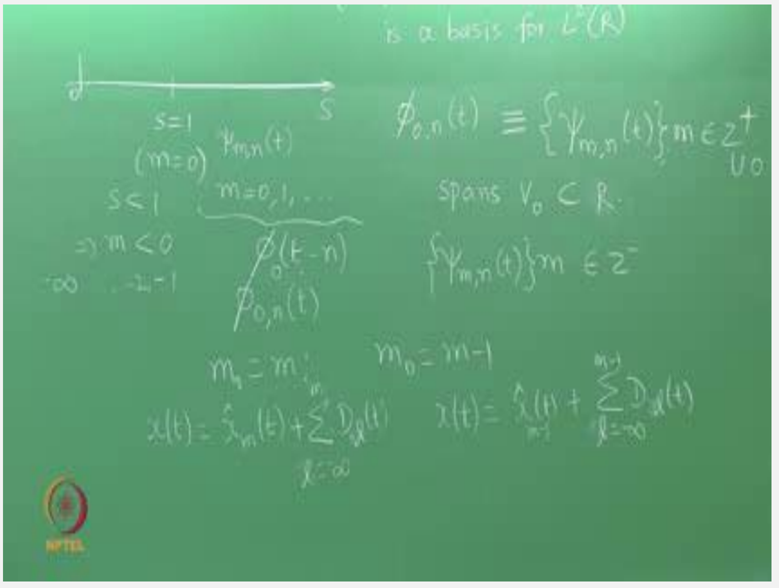
(Refer Slide Time: 15:19)



So, now I can extent this idea of approximation to, at general reference point. Exactly, the way we did in CWT. We started off with a reference point being s equals 1. And then, we said we look at now, other an arbitrary reference point s equals s naught. Now, m naught is the corresponding reference point for me. So, s naught is 2 power m naught. So, what I do here is, first if you look at equation 4. Now, I am expressing x of t as a sum of it is approximation at scale 0 and all the details that are left out, in this approximation..

So, here we said in equation 3, I only obtain an approximation of the signal at scale 0 or scale 1. What about the details? Well, the details are contained in these wavelet coefficients at finer scales. Remember, the approximation at scale 0 is an aggregate of all the details at coarser scales. So, let me just show you graphically.

(Refer Slide Time: 16:36)



So, if I just considered this scale. Let us say, this is a scale space. Our reference point is s equals 1 or m equals 0. What my scaling function is doing? So, let us say here is, a scale 0. What my scaling function is doing is, it is actually constructing an approximation of the signal by aggregating all the details that you would have obtained, using the wavelet basis functions, at larger scales.

So, there are many wavelets here. I can say s, m comma t, m comma m, t. Where, m runs from 0, 1 and so, on. And here, m actually assumes negative values. Because, s is 2 power m, remember if I want s less than m so, s less than 1. So, this the space s less than 1, which means m actually is less than 0. Here I have minus 1, minus 2 up to minus infinity. That is what, those are the values of m, spanning this space here.

All these wavelets are that x levels m 0, 1, 2 and. So, on are being replaced with a single phi of t, which is at scale 1 or m 0. So, we call this as phi naught of t. And in fact, we also allow translates. So, t minus m or you can say phi naught n comma t. That is, what is happening right now. When I look at the approximation x naught of t, then what essentially it contains is, it contains all the details in this scales. But, if I want to recover x of t fully, then I need the details or the projections on these scales, as well.

Because, then only together you can they constitute the basis for the real space. That is psi m, n comma t, m belonging to the integers space. That is, m running from minus infinity to infinity. And n belonging to the integers space is a basis. This entire family is a basis for the real number space, which have finite energy or the functions in the real numbers space. In the real space, which have finite energy.

By selecting only those wavelets, we corresponding it to m greater than or equal to 0. What essentially have done is, I have chosen an subspace. I have only chosen a subspace here. So, phi of t or phi naught n comma t is an aggregate. It is equivalent to with only m belonging to z plus. Obviously, what I have done is, I taken this huge infinite family, which spans the real space. And I have taken only some members of the family and constructed the scaling function.

Therefore, this phi will span V naught which is a sub space of L to R, you can say. Normally, it is just sufficient to write R at this point. If therefore, if I recover x of t only from the projections on to this, I require only an approximation. If I want to recover x of t completely, I need the other wavelets. The projection of the signal on to other wavelets, which I have left out, which is for m belonging to… In fact, it should not be z plus, z plus and 0.

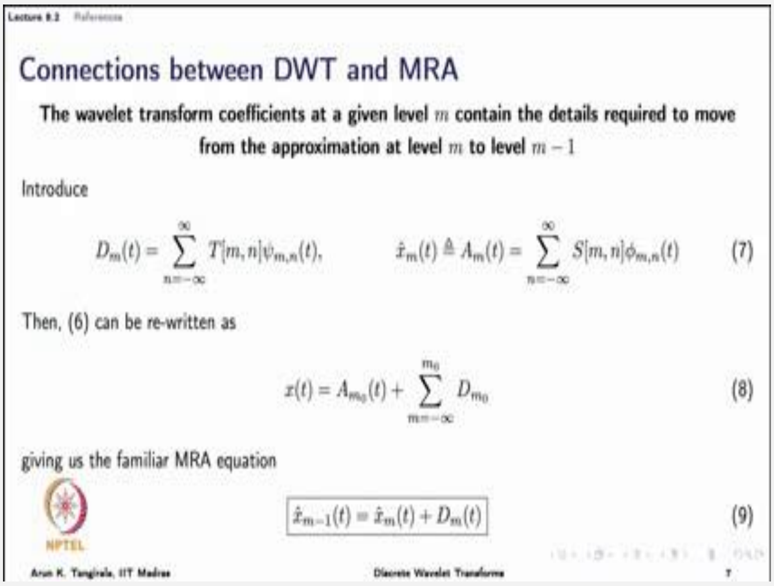
So, I need the wavelets that belonging to, that are spanned by allowing m to take on values in a negative integers space. That is m starting from minus 1, minus 2 up to minus infinity. And that is exactly, what that equation at the top is telling you. It says that, to recover x of t, I need the projections or the coefficients that I obtain. Not necessarily the projections, the projection coefficients that, I obtain with the wavelets at the remaining scales or at finer scales.

Remember, as m with respect to s equals 1. Whenever, m takes on values 1, 2, 3 and. So, on, then we are looking at coarser scales. And whenever, m takes on values of minus 1, minus 2 and. So, on, then we are looking at finer scales. So, I am looking now, to using the coefficients at finer scales to recover x of t. Now, if I chose an arbitrary reference point n naught, then I get to equation 6. Where, what I am doing is here, I am introducing s of m naught comma m.

Earlier, my m naught was 0, in the previous equation. Here in equation 3, m naught is 0. Now, we are shifted from m equals 0 to m equal m naught. So, I have s of m naught comma n times phi m naught comma n of t. So, my phi of m naught comma n of t is once again representing all the wavelets that are at scales higher than or greater than or equal to 2 power n naught. And obviously, once again if I want to recover x of t, then I need the information at finer scales, finer than m naught.

If I choose to ignore the second term in equation 6 and only look at the first term, then I end up with an approximation of x of t at scale, 2 power n naught or at level n naught.

(Refer Slide Time: 23:17)



And that is the basic idea in the equations that you see here. The basic message from this equation is that, the wavelet transform coefficients which are the t's, your t of m comma n at any level m. What do they contain? How do you interpret them? They contain the details required to move from an approximation at level m to level m minus 1. This is the statement that we made at the beginning of this lecture. So, to understand this point, let us look at the inner summation here, at a specific scale m. And we shall call that as D subscript m of t.

What is this D subscript m of t? Momentarily, when we are looking at equation 6 here, particular the second term, how do you evaluate? You fix the value of m. And then, you evaluate the inner term. Then, you take the next level m. And then, you evaluate the inner term. That is, how you proceed at any level m. When the inner term is evaluated, what we you are doing is, you are recovering the detailed component of that signal at that scale. That is what it means. That is, at scale 2 power m.

And when you put together all those details at finer scales the m naught, then you get all the complimentary information that was not present in your approximation at scale m naught. So, the inner summation is always going to give me, the detail at that scale at level m naught. And of course,, by our previous notation, we call this summation, that is the first term here in x of t has nothing,, but the approximation at scale 2 power m or at level m.

Why is this? Because, we are going by the same convention here. When our reference point was m equals 0, we said whatever I recover from the scaling function coefficients at level 0 is only going to be an approximation at level 0. Now, since we have chosen an arbitrary reference point. And that reference point is either m naught or m. Whatever I recover from the scaling function coefficients is simply going to be an approximation at that scale m, 2 power m or at level m.

We shall use this term scale and level interchangeably. But, you should understand that scale is exactly 2 power m and level is m. So, please do not get confused. Now, with the introduction of these terms, I can write equation 6 as A m naught of t. That is, approximation at a scale m naught or a level m naught, plus all the details at scales finer than m naught.

Now, if I write this equation at two difference values of m naught. M naught equals m and m naught equal m minus 1. And subtract, then what I end up with is x hat of t at m minus 1 is x hat of p at n plus D m of t. How do you get this? Well as I said, write your equation 8 at two successive scales. So, I have x of t as x hat of t at m naught. Let us choose two different values of m naught. M naught equals m and m naught equals m minus 1. That is I can choose two different reference points. If that is the case, then the first equation yields approximation of x at level m plus, the summations that you have starting from m equals minus infinity to m D m. And then, when you have m naught equals m minus 1, the breakup of the signal is x hat of t at level m minus 1 plus sigma m equals minus infinity to m minus 1 D m up. Well, if you are confused with the index, we can change this to l. So, that it is lot more clear to you. They are just dummy variables anyways. So, now, when I subtract these two, So, here, then what I get is of course,, on the left hand side I have the signal.

(Refer Slide Time: 28:45)

So, that goes to 0. And on the right hand side, I have x hat of t at m minus x hat of t at m minus 1 plus sigma l equals minus infinity m D l of t minus. Here, I have m minus 1 D l of t. This entire thing is 0. So, I just flipped the sides, because x of t minus x of t is 0. As a result what I have is, x hat at x hat of t at m minus 1. That is the approximation of the signal at level m minus 1 is the approximation of the signal at level m plus D m of t.

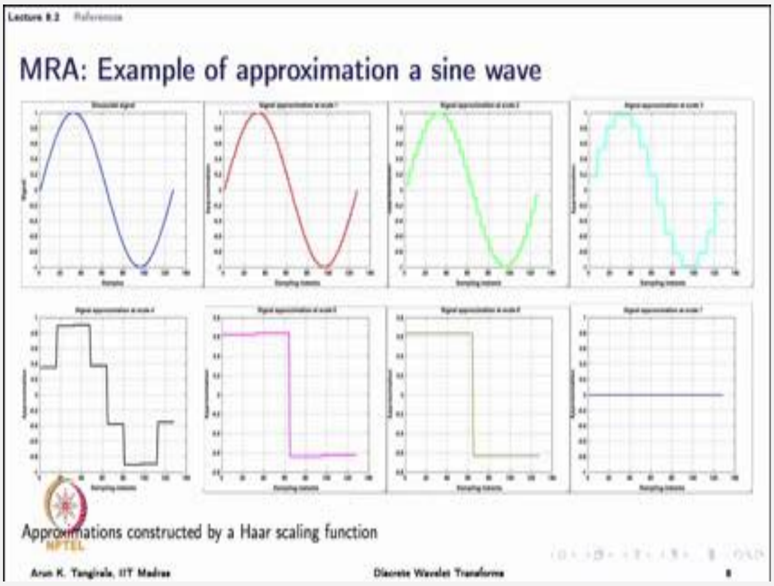
This is the fundamental equation for multi resolution approximation. That is, why is this fundamental equation for MRA? That is, because as I move from m to m minus 1, the resolution is improving. Because, the scales are becoming finer. Therefore, the time grade now is much more finer than at m. In fact, it is finer by factor of 2. So, what this equation tells me is, the approximation of the signal at resolution 2 power minus m minus 1.

Resolution is always inverse of scale or you can say the approximation of signal at 2 power m minus 1 is nothing but, the approximation at a coarser scale, which is 2 power m plus the details which are obtained by projecting the signal on to the wavelets, at that scale. When you put them together, you get a better approximation. So, now if I can apply this recursively, I can say x hat of t at m minus 2 is x.

Remember, associated with this approximation is a detail D m minus 1, at that scale. So, when I look at x hat of t at m minus 2 that would be… If I rewrite this equation, I would have x hat of m minus 2 of t as x hat of t at m minus 1 plus D m minus 1 of t. What is D m minus 1 of t? Whatever details that you have obtained by projecting the signal on to the wavelets at this scale or at this level. When you put together this, you get x hat of m minus 2.

When do I recover this signal at all? Do I will I recover this signal? Yes, if I consider all these scales, which is actually the equation that we have seen, earlier 4 or even equation 6 so, to speak. If I want the complete signal, then I need to consider all these scales. Because, only then I would have considered all the basis functions, that span the real number space.

(Refer Slide Time: 32:02)



So, let us now look at an example of multi resolution approximation. That is, now what we just said is, I can now construct approximations of a given signal at different scales, from fine to coarser scales. Why do I talk about only fine to coarser scales? Let us look at this example, here. So, what I have here is a sine wave. And I want to construct multi resolution approximations of this. What is exactly MRA? Will try to now understand this with a figure and then with this example. Then, we will look at the formal theory.

First of all, I show you the sine wave in blue colour here, which is at the left top. Although, I show this as a continuous signal. Strictly speaking, I have only samples of x of t at a sampling intro in T s. So, ideally I should be showing you spikes. But, because a plot may look a bit uglier. I actually have not shown you that. For all our discussion, x of t is not a continuous time signal. It is a sampled signal.

Now, I do not have x of t at all, until the equation 9. We have been discussing with respect to theoretical signals x of t. But, in practise I only have sample signals. So, how does the situation change? Well, it does not change much. But, just for one important thing, which is I have a bunch of numbers, which is the sample signal. Now, what do I assume this numbers to be? That is, what resolution do I assume to be? At what scale do I have this?

Because, if I look at the theoretical analysis, x of t is at the finest scale. Because, it is a continuous signal. Whereas, now I am looking at a sample signal. I cannot say, it is at the finest scale. Had I sample faster, then I would have obtain better x of t. That is better resolution. Had I sampled even faster, than I have even finer scale representation of x of t. But, now I have decided on some sampling interval. Therefore, I have x of t at some scale.

The nice thing is, I do not nearly need to worry, what scale this is? We will arbitrarily say and we can do that, without any loss of generality in the sense of theory. That we will say that, this sample signal is available to me at scale 1, that is at m equals 0. I cannot get a better representation of the signal, unless I go back and resample. So, now, what all I can do is, I can only construct approximations of the signal at coarser scales.

In fact, the crudest approximations would be to simply down sample. What kind of approximations? Multi resolution approximation. Now, I am interested, not in some arbitrary approximation. I am interested in approximation, that are at coarser and coarser time scales. Multi resolution approximations are specific approximations, which where the resolutions fall by factor of 2. So, I have the signal at resolution of T s in time.

Now, the next resolution that I can look at is 2 T s. Because, we are looking at dyadic scales or because the resolution is inverse of scale. I can move from T s to 2 T s. As I said, crude approximation of the signal would be of the signal at a resolution 2 T s would be simply to thrown away every sample. So, let me just show you that on the board and then, we will come back to the example.

(Refer Slide Time: 36:12)

Suppose, I have a sequence x of 1, x 2 and so, on. Let say, I have a bunch of samples like the one that I have here, for the sine wave. Then, this is at a resolution 2 T s although, I write x of 1 here. What this means is, I have x at. In fact, I can say x of 0, x of 1, x of n minus 1, if you like it. Then, this is at x 0 x at p s and x at n minus 1 T s. Notice that, I move from square brackets to parenthesis. The moment I move from sample indices to actual time.

So, this is at a resolution T s. Now, I want to construct an approximation of the signal at 2 T s. So, the next resolution I am looking at, it is a coarser resolution 2 T s. As a said, a crude way of constructing an approximation of this signal at this resolution is to throw away every other sample. So, I would have for an example, x of 0 and then I have x of 2 and. So, on. May be let us assume that, m is power of 2. Therefore, I have here.

So, the previous sample would be x of n minus 3, nothing in between. So, I have thrown away. But, this is a crude approximation. The approximation that you see on the figure is a better approximation than this one. And I will tell you, better in what sense? A better approximation is to construct an average of x 0 and x 1. And say that, over this interval 0 to 2 T s,. So, here I have x of 2. What I am going to do is, this x of 0. If I were to plot, I have 0 T s and then 2 T s, 3 T s and so, on. This is my time t.

What I am going to do is, I am going to say that an approximation of the signal at 2 T s would be average of x 0 and x 1 and held constant over this. So, this would be the value here, initially. In fact, in the example that is what you see, in the second plot. You can see, now that we have plotted this as a piece wise constant over the interval 2 T s. And again, over the next courser interval, I would construct an average. So, I would construct an average of x 2 and x 3 and continue in this fashion.

And this is how I continue. And when I want to construct the approximation at a coarser resolution, that is now, remember the resolution should fall of by factor of 2. So, the next resolution that I am going to look at is 4 T s. So, that is what exactly I am doing here. So, I am taking a sine wave at resolution T s. And then, the next course approximation is at 2 T s and then 4 T s. Then, I have 8 T s, 16, 32, 64, 128. As I move down the approximations, the resolution approximation becomes coarser and coarser, as you can see.

In fact, the crudest approximation is simply the average of the sine wave over that entire interval. That is the crudest representation that you can give for the sine wave that has been sample at T s. So, here I am constructing averages over intervals of 2 T s. Here, I am constructing averages over 4 T s interval, averages over 8 T s, averages over 16 T s. That is why, it is constant for a longer period of interval. Then, averages over 32 T s I have 128 points.

So, here the signal becomes more and more piece wise constant over larger intervals. And then, I have here over 64 T s and 128 T s. So, what you have now from all these signals that you have here, all the approximations that you see at constitute, what are known as multi resolution approximations. What is special about this multi resolution approximation? As you go down, you are actually constructing resolutions that are nested, that are embedded or you can say, the first feature is that the resolution falls off by factor of 2.

The other important feature is that, the approximation that I have generated... For example, if I look at the red one, that is the 1 comma 2 plot. That has been constructed from an approximation of x of t. Remember, your sampled signal is also an approximation. And we said, this approximation that is scale 0 or scale 1 or m 0 level 0. So, the approximation at level 1 is derived from level 0. The approximation at level 2 is derived from level 1 and so, on. So, one approximation is derived from the other and they are said to be nested.