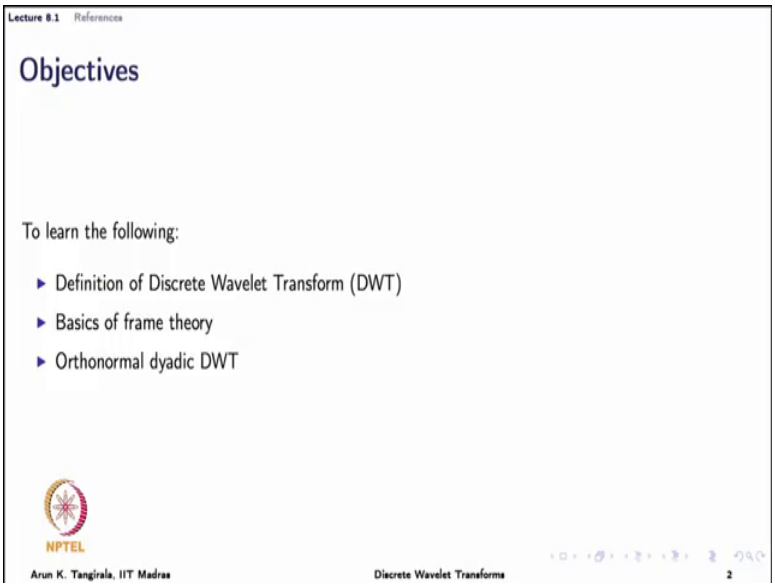
Introduction to Time-Frequency Analysis and wavelet Transforms
Prof. Arun K: Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 8.1
Discrete Wavelet Transforms
Part 1/2

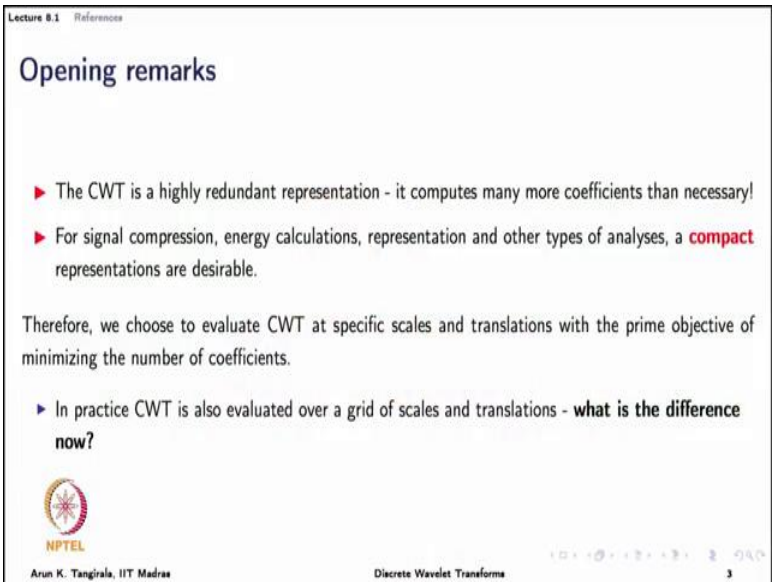
Hello friends, welcome to lecture 8.1. So, we are into the first lecture on the exciting topic of discrete wavelet transforms.

(Refer Slide Time: 00:29)



In this lecture, we shall look at definition of discrete wavelet transform. Learn and review some basics of frame theory as relevant to the DWT, and then look at the theory of orthonormal dyadic DWT. Again fairly and introductory lecture, and this lecture is primarily going to be theoretical. So, you will see a bit of math, but them mathematics is inevitable in the subject. Gradually as we did in CWT, we shall look at the few applications mat lap based demonstration and so on.

(Refer Slide Time: 01:12)



So, before we plunge into the definition of DWT, we shall ask why we should be looking at the DWT, is it different from DWT, and so on. Well first of all, the CWT is a highly redundant representation. We have mentioned this quite a few times in the lectures on continuous wavelet transform. What you mean by redundant is, it computes many more coefficients and necessary as is what we have doing with a wavelet transform or even the short time Fourier transform is, we have transforming a 1 dimension signal into a two dimensional plane. So, obviously, we are generating more numbers than necessary, but it is possible that even in the two dimensional plane, we can come up with number of coefficients; that is not more than the number of coefficients or the numbers that we had in the one dimensional plane.

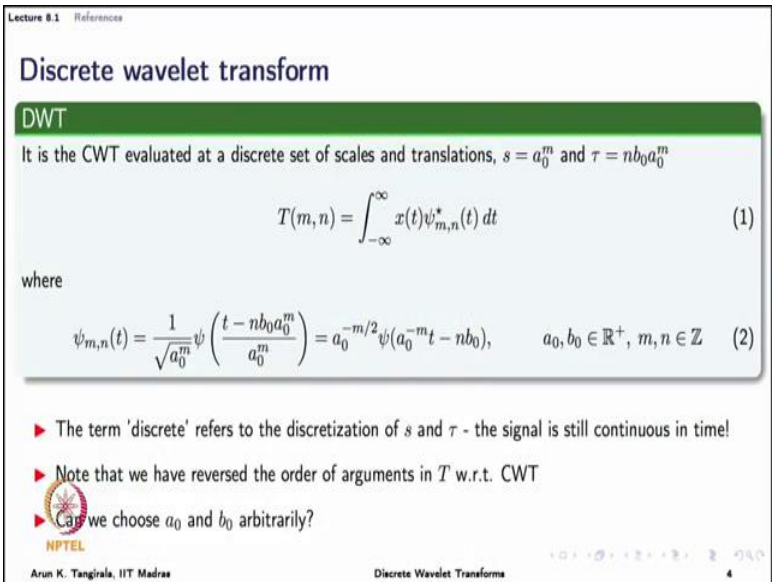
What CWT does is, that in the two dimensional plane, it uses many more numbers than necessary to represent the signal, and that is what we mean by redundancy. Now when it comes to application such as signal compression, or energy decomposition across scales, and other types of analysis. Normally one decides a compact representation. What you mean by compact? Again this just the minimum number of coefficients that are necessary in the two dimensional plane. And this minimum intuitively cannot be smaller to begin with, then the number of coefficients are the numbers that we had in the one dimensional signal, but in signal compression once we have this minimum resonation, we throw away quiet and lot and we retain quite of few to achieve very minimal representations, or even compressed representations, and we will talk about that in signal compression. So, the

main question that often arises is, when I am computing CWT.

Anyway I am going to discretize my scale and translation your seeing that many times in computation of CWT and so on. Now, what is a difference here in, we are saying discrete wavelet transform you are also mentioned this earlier; that it is nothing but is CWT evaluated at specific scales and translations. In the computation of CWT as well, we are evaluating CWT on the grid of scales and translation. So, what is a difference now. The difference is in how we choose this grid for this scales and translations. In the computation of CWT we choose an arbitrary grid. The user has the complete freedom to compute the CWT at as many scales as possible, as this the restriction is only the computational power, but otherwise I can compute CWT over a very fine grid. In DWT the user does not have or choice, and mathematically upfront we restrict our evaluation of CWT on specific scales and translations; such that something is achieved, and what is that something minimal representation.

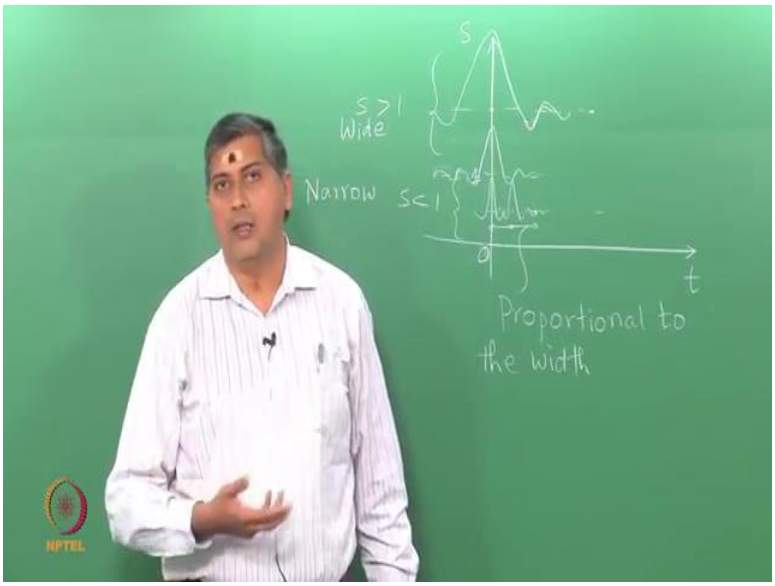
We want to have as many few coefficients as possible in the two dimensional plane or the time frequency plane, and still not loose information to begin with. So, that is the prime difference between CWT and DWT, and just this one difference opens a door to a very exiting world of transforms. It is not a new transform discrete wavelet transform, just ((Refer Slide: 04:55)) from CWT, but because of the way, we evaluate the CWT at the specific scales and translations. We achieves some very nice properties for the representation, that are useful in signal compression energy calculations, multi resolution approximation so on and that is what makes it very exciting.

(Refer Slide Time: 05:19)



So, let us now look at the definition of DWT. Again, we know that DWT is the CWT evaluated at specific scales. Now what kind of scales have we looking at; exponential scales a naught rise to m where m is an integer a naught is a number that I am free to choose, and also at specific translations; here n b naught times a naught to the m. Now you can quickly recognize a naught to the m is nothing but the scale itself. This n b naught essentially says I am moving the wavelet window in the analysis of the signal, proportional to the scale of the wavelet. Remember scale determines width of the wavelet. If I am looking at large scales, high scales then the wavelet is wide, and therefore, I should be taking large steps. Whereas, when I look at final scales, lower scales then the wavelet window is narrow, so I should be taking smaller steps. So, let me just show you graphically on the board, what is the qualitative difference between CWT and DWT.

(Refer Slide Time: 06:35)



So, let us say this is the time, and here we have let us say the time scale plane to like it, then at lower scales. So, remember unlike the regular time frequency plane, here we are drawing the time scale plane. Let us assume that this is scale one just for our reference, or let us say this is scale zero, and here we have scale one. So, here we have s greater than 1 which means I have wide wavelets, related to one, and here I have s less than 1, so narrow with in time. So, what kind of wavelet do I see if I pick a certain scale here, and I will have to draw the wavelet, let us say it is going to be this wide. Whereas, if I pick a scale here, a wavelet centered at the origin could look this wide. So, is all qualitative sketch, so do not try to quantitatively compare the widths. Just tell you here you have narrow and wide, relative to what you would see at s equals 1. So, may be at s equals 1 you have, all of these wavelets are normalized unit energy. What we are doing in CWT is we are marching a head in time, with an arbitrary step size.

And also we are choosing wavelets that us paste in an arbitrary wave in scales; that is I choose a scale here let us say point five, and then I could choose a scale let us say 0.51, depending on how fine a grid I want on the scales. Now in DWT what we are doing is, we are not really going to at any scale, we are not going to march a head in time over in arbitrary step size, but we are going to choose the step size propositional to the width of the wavelet; that is the prime difference. So, if I look at this wavelet I would choose the next time interval; that is where I should place the center here, the center could be some way here. So, that is approximately. In fact, for a certain choice of a naught and b naught

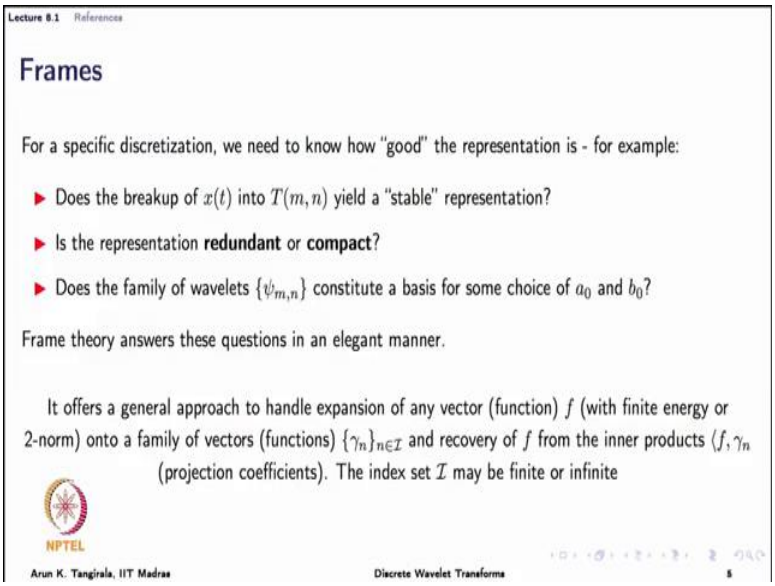
I will have non over lapping. I would have generated non overlapping wavelet windows. So, it is possible that in DWT also these wavelets can overlaps. So, in my next center would be here.

So, this step size that I see here; the marching step size, and then next one here, let us say next center is a. This here is proportional to the width, and you can show a singular think here. Here I would probably place. So, here is a center of the wavelet to begin with. The next center could be somewhere here depend, qualitatively could be this for. So, you can see that wide wavelets are taking larger steps. It is like long person be able to take long steps ahead, and person who is short, he is actually taking more frequent steps, so that is about it. So, it is a way you grid, you choose your grid points for this scale and DWT, and that is with respect to the translation, and scaling is also a naught to the m it is not a linear discretization is exponential discretization.

Of course, normally you see a naught equals 2 and b naught equals 1 being very the common choice, but will come to that, at the movement we are studying more general problem. The rest of the story is a same t of m comma n. So, now we have turn away tau and s and you replace them with m and n, the only difference you should notice is, we are reverse the order of indices here when we were looking at CWT, we wrote t of tau comma s. So, time translation parameter appear fist and scaling parameter appear next, but here we have m comma n, where m denotes the scale and n denotes the translation.

So, that is generally the convention DWT, but you are most welcome to choose your own convention. On the right hand side you see the familiar wavelet transform. In fact, in almost all DWT is, you will never use complex wavelets, so you can really throwaway this complex conjugate here on top of the psi. And these wavelets are nothing but again children of the mother wave, evaluated at specific scales and translations. Do not forget that we have a normalization factor, to ensure that the wavelet as the same energy as the mother wave itself. So, an important point to keep in mind, which often leads to a lot of confusion, is that the term discrete refers to the discretization of the scales and the translation parameter, not the signal itself. If you notice in equation one this signal is still continuous, it is not discrete yet. We will look at a discrete time version of this integral in 1, even we talk about computation of DWT and practical aspects and so on. At this moment will keep the signal continues. Now the question, the big question that we have in front of is, can I choose a naught and b naught arbitrary is it possible.

(Refer Slide Time: 12:26)



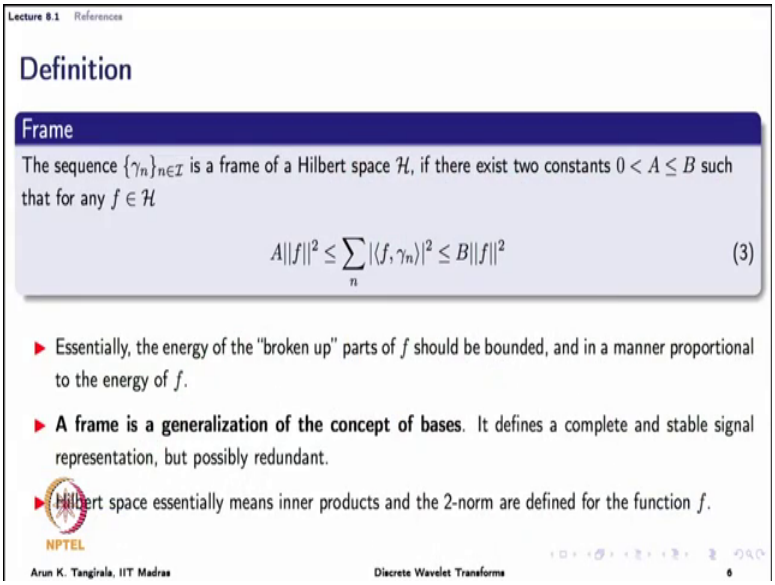
Well this is where the frame theory which was developed in computer version in a couple of researches, comes to our rescue. The frame theory was developed in computer version. In fact, in signal processing where 1 was interested in knowing whether it is possible to recover a signal from it is irregularly sample version. So, signal is being evaluated at some irregular samples. Is it possible to evaluate the reconstruct the signal from it is irregularly sample version. So, that is in signal processing, those are the origins. Those irregular samples of a signal you can think of the projection of the continuous time signal on to some functions.

And therefore, frame theory essentially asks or address is a problem of expanding any vector or a function I like, which has finite energy are too long on to another family of vectors. So, I take a signal and I break it up into some components, and what frame theory tells me is, whether breaking this up will yield stable or a redundant, or an orthonormal or a tight representation and so on. So, what kind of representation do I get when I take a signal and break it up in to pieces. And how am I breaking it up, with the help of these functions I project the signal in on to some functions.

We do not call this as basis functions in frames here, because we do not know if these functions are linearly independent or not. They are just some set of functions, and will look at an example to understand this. There some set of functions, and what I would like know is about there a representation and also recovery of the function from this

projections. Now, projections are of in computed as inner products. Projections are nothing but shadows. So, every projection of a signal onto some function will get me some component of it, not the entire thing. So, what frame theory tells me is, also whether I can recover the function from the inner products or the projection coefficients. Now this m subscript n here. I am sorry if there is a confusion. This n unfortunately coincides with the n that we had in the DWT, but please keep them separate. This n false out of some interior set and this set could be finite or infinite. What is means essentially is, that the functions on to which or the vector on to which I am projecting the signal of interest, could be finite or infinite that is all.

(Refer Slide Time: 15:17)



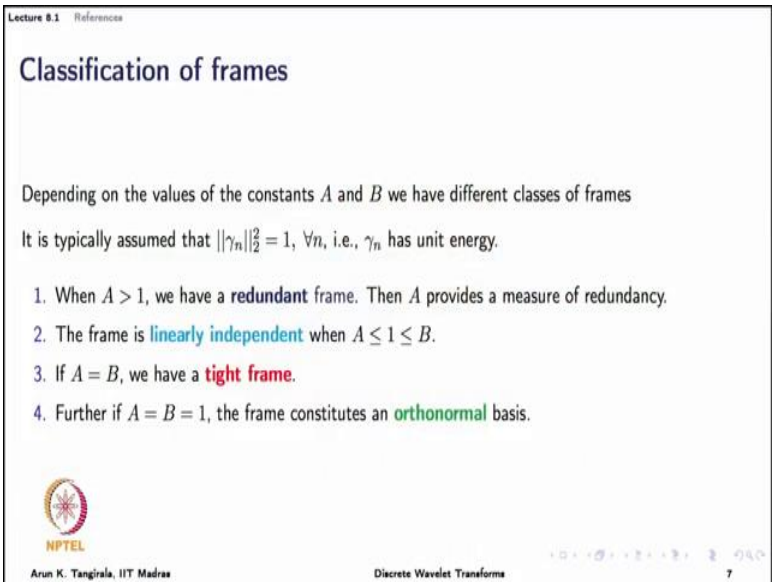
Now, let us look at the frame theory bit formally. First let us learn what a frame is. Before we look at equation three, I have written bold place here, that a frame is a generalization of the concept of basis. We know, we all know that any signal or any vector, can be represented as a linear combination of some basis for that signal, but if you recall the definition of basis requires linear independence. Of course, orthonormality is very nice, but linear independence is the minimum requirement for basis. What if I have functions that are not linearly independent, then what do I do. that means, I cannot expand the function the signal on to this so called analyzing functions, I can that is what frame theory essentially does. But then you cannot really expand the signal on to any set of functions, and expect that your broken up components are going to be stable or enough, or that you can recover exactly.

So, now, let us look at the definition three, what constitutes a frame, what set of functions constitute of frame. A frame is a nice term, if you can relate this to the construction of a room or house. If you have visited the constriction, at the time of constriction you see that there are some frames, holing the roof; that is being constructed or some walls that have being constructed, depending on what kind of wall you are constructing and so on. So, these are the supporting frames. You need a minimum set of supporting frames, you can have more than that, but you cannot have less than that that. If you have less than the necessary supporting frames, the wall or the roof is going to fall down, so it is not stable set of frames. So, if you draw that analogy here, I have a set of functions which I call as a frame in a Hilbert space. Now Hilbert space should not scare you.

All that means is that I am working with a set of signals and functions, that have inner products define for them, and a norm is also define for them; that is very important, because we are dealing with signals that have energy, and some mathematical support has to be there, to be able to say that energy can be calculated for this signal. So, a sequence of functions, whether this could be finite or infinite set of functions is called a frame, if and only if there exist two constants a and b; such that both are greater than 0, but b is greater than or equal to a; such that the energy of the broken up components; that is how you view the modulus of the inner product of f with gamma n square. So, the term that I am summing up is the energy of the individual broken up component, and I put them together; that is the overall energy of the broken up pieces.

If I compute the overall energy the way we are doing here, it should remain bounded, between this lower bound and upper bound, and the lower bound and upper bound are themselves propositional to the energy of the signal. Clearly you know it is possible that a and b can be 1. If a and b or equal and equal to 1; that means, the energy of the broken up composition exactly matches the energy of the function itself. Now we shall ask different situations, we shall examine quickly different possibilities. Obviously, a and b are bother greater than 0, and we will also assume that the gamma n’s that I have, have been normalized have finite energy.

(Refer Slide Time: 19:07)

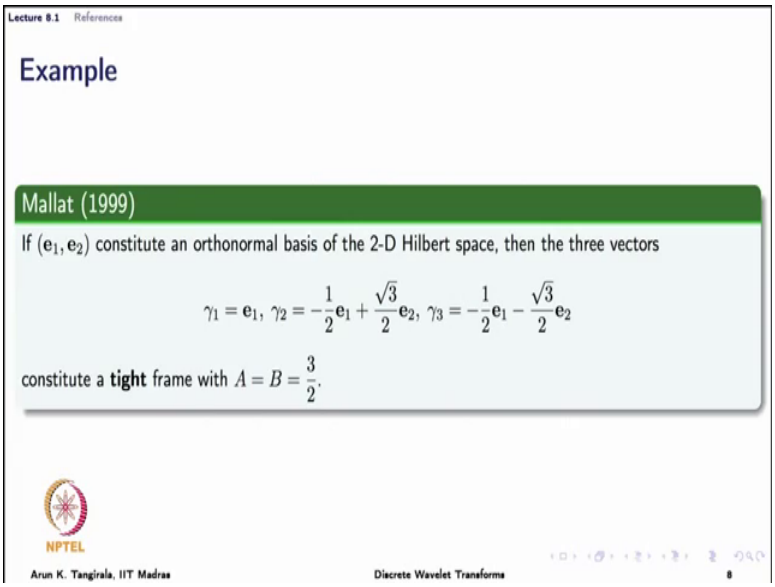


So, the first frame that we have, is a redundant frame, and a is greater than 1 what does a greater than 1 mean. We go back to this inequality here; obviously, is mean that the energy of the components put together, is much greater, or is greater than the energy of the signal itself. a is greater than 1, b could be any think, but that is a is a lower bound. So, which means I have more components then necessary. So, that is why the term redundancy comes in. Now, when a and b are at the most one. They could be equal, both can be equal, but a is never greater than 1, and b is also never greater than 1, but then b can be greater than a. Then we have what are known us linearly independent, we have what is known us a linearly independent frame. Why is it linear independent why not you know perfect tight frame, because a is not equal to be necessarily. When a equals b we have a tight frame. What this means is, that it is a very snug fit for the signal, that you have really found, really broken up the signal in a very tight manner. It does not mean that you have a minimal representation; that is very important.

It is simply says a equals b; the lower and upper bounds are equal. When the lower and upper bounds are equal to 1, both then we have an orthonormal basis, where you have exactly the minimal set of gammas or the functions frames to represent your signal f, and then we can say that we have a basis. In both situations situation 2 and situation 4, we have gammas as basis, because in case two a frame is linearly independent; that means, now I have a basis, and in frame four I have an orthonormal frame which is also linearly independent therefore, I have a basis. But in cases one and two I do not have basis,

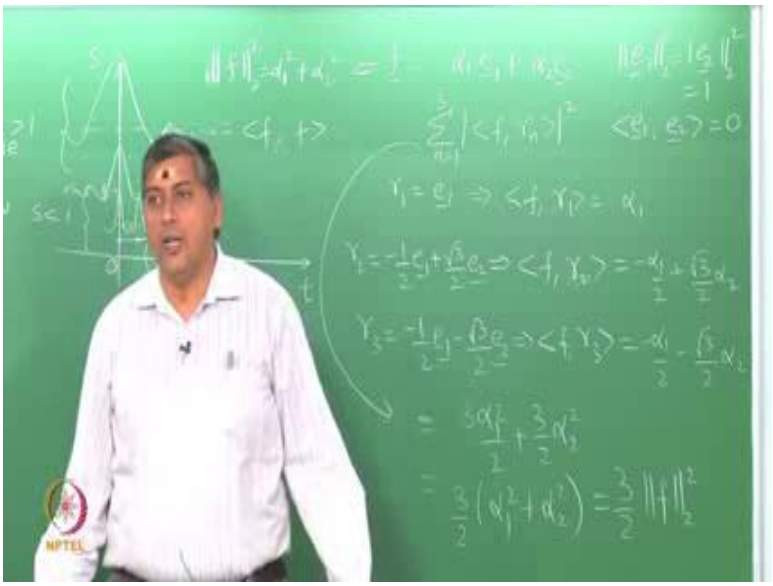
because linear independents is not being guaranteed. So, these are the different possibilities that you can have. Now you can clearly see why frame is a generalization of the concept of basis.

(Refer Slide Time: 21:40)



So, let us look at a quick example here, which I borrower from Mallat's book. Suppose and looking at the two dimensional Hilbert space; that is two dimensional real space where the inner products are defined, could be real or complex. And we know that these vectors e 1 and e 2. What are these e 1 and e 2; they constitute, they are this 1 0 0 1 orthonormal vectors, we know that they constitute an orthonormal basis. But instead of choosing e 1 and e 2, if I choose gamma 1 gamma 2 and gamma 3 as my new basis, when I do not want to called basis as a, but a new set of representing functions then what kind of a frame do I can. Now it turns out that we get a tight frame with a equals b equals 3 by 2, and let us see why this is the case.

(Refer Slide Time: 22:35)



So, let us take any functions in the two dimensional Hilbert's space, and since e one's are, the e 1 and e 2 are the basis. Let us say assume that is function is vector; e 1 and e 2 are both two dimensional vectors, and what I have there is; gamma 1 gamma 2 and gamma 3, in order to determine what kind of a frame I have, I need to evaluate this summation.

So, the inner product has to be evaluated, and here n runs from 1 to 3, because I have three functions in my frame, sorry what is gamma 1? Gamma 1 is e 1 itself. Therefore, the inner product of f with e 1, we know that e 1 and e 2 are orthonormal which means the inner product between e 1 and e 2 is 0. So, it is clearly by substituting for f here, and applying the given information. What is the given information? It is orthonormal which means is square 2 norm of e 1 and e 2 are both 1, and the inner product between e 1 and e 2 is 0. So, this the given information to us. I just evaluate the inner product to a, I get alpha 1.

And then I have gamma 2 as minus 1 over 2 e 1 plus root 3 over 2 e 2. So, now I am going to evaluate the inner product of f with e 2, what do I get? I have minus of e 1 plus root 3 over 2 e 2, sorry the gamma 1, and gamma 2. So, I have here f in a product of f with gamma 2 f is alpha 1 e 1 plus alpha 2 e 2. And therefore, I have minus alpha 1 over 2 plus root 3 over 2 alpha 2. And then finally, the second one is fairly straight forward. Once you understood the second case gamma 3 I have here minus alpha 1 over 2 minus

root 3 over 2 alpha. Now all I have to do is, I have some square these quantities and what do I get when I square this, I get alpha 1 square, and here I get alpha 1 square over 4 alpha 1 square over 4. So, when I add them up I get alpha 1 square over 2. So, this summation here produces 3 alpha 1 square over 2. Then likewise here I have 3 by 4 alpha 2 square and 3 by 4 alpha 2 square. So, I get 3 by 2 alpha 2 square. Then I have a cross terms which cancel out each other. This is nothing but 3 by 2 alpha 1 square plus alpha 2 square and alpha 1 square plus alpha 2 square is nothing but a energy of f.

So, this means essentially that the energy of f is alpha 1 square plus alpha 2 square. All you have to do is, simply take the inner product of this with itself. Remember the square to norm is nothing but inner product of any function with itself by the property of e 1 and e 2. You can show this. So, this is nothing but 3 by 2 that is it. So, now, you understand why we have a tight frame a equals b equals 3 by 2. So, what is this means? Obviously, in this example it is value. Obvious it is a two dimensional space. I only need two basis functions two's span the entire space, but I have three representing functions gamma 1 gamma 2 gamma 3, that have been constructed out of e 1 e 2 e 3.

Therefore, it is going to be redundant and the only nice thing about this is, its a tight frame. So, tight frame have certain advantage, but we have redundancy. How much redundancy do I have, and that is what is measured by this factor a, whenever you have a tight frame a equals b. So, the redundancy is measured by a. So, I have about redundancy of 3 by 2. In fact, strictly speaking half, because when a equals 1 then I have minimal, so that extra half is what is say measure of redundancy, does not mean that I have a half function in excess, but it is just a measure of redundancy, higher the value of a more the redundancy that I have.