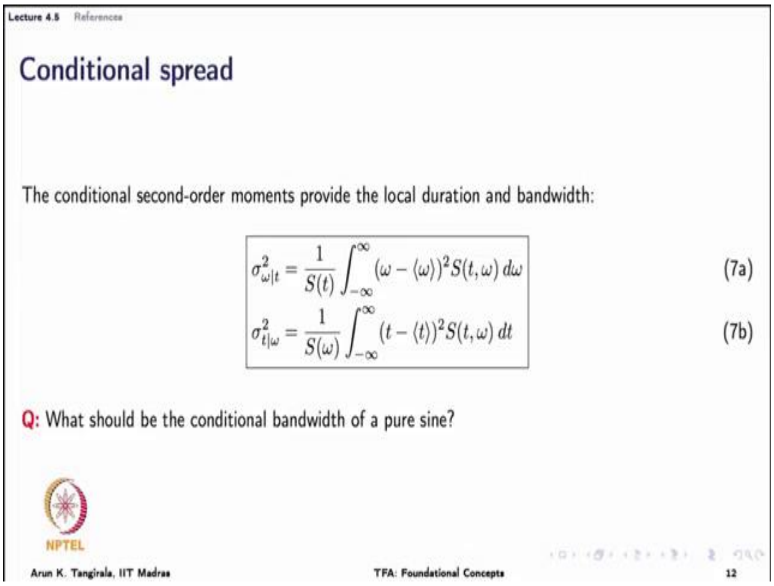
Introduction to Time-Frequency Analysis and Wavelet Transforms
Prof. Arun. K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

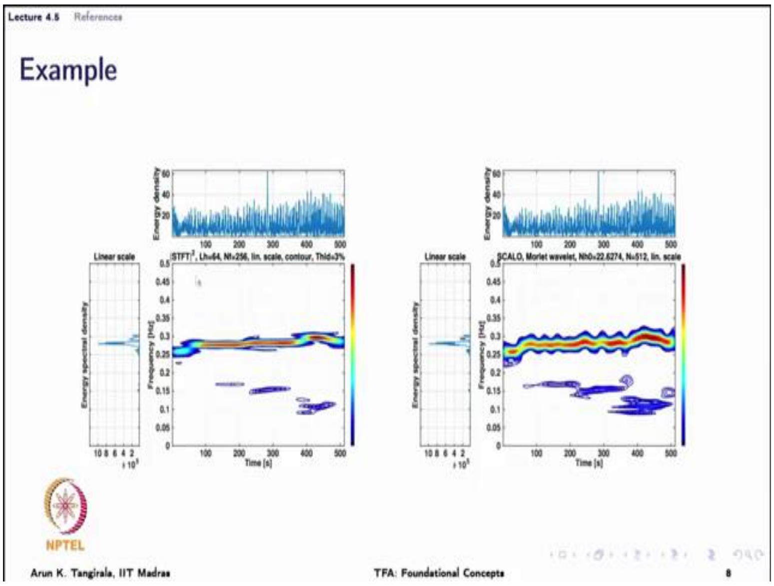
Lecture - 4.5
Requirements of Time-Frequency Analysis Techniques
Part 2/2

(Refer Slide Time: 00:15)



Associated with this conditional means, we have conditional spreads. That means, now I am asking the local bandwidth. We have talked about global duration and global bandwidth, this is something very interesting that you should keep in mind. Whenever I am doing Fourier analysis and doing a global analysis and that is what you will see in many texts, I am looking at the global property. I am, I do not care when this frequency occurred in time. I am only interested in what frequency components are present in the signal. Of course, the problem with that is, it loses track of time and that the bandwidth has contributions from both amplitude and frequency modulations and so on, and that is what motivates us for the, motivates us to the time-frequency analysis.

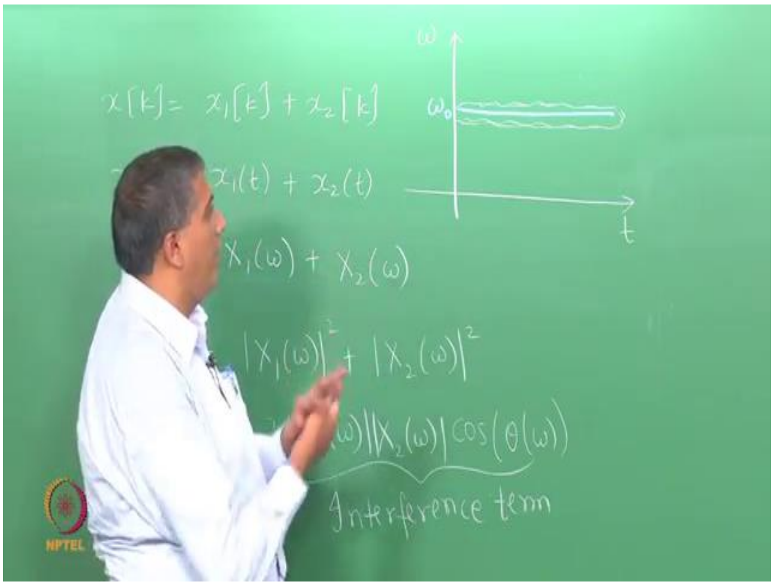
(Refer Slide Time: 01:10)



Here I am looking a local property. So, at a given time t, I would like to know what is this spread. So, again going back to this figure here, I pick a certain time t and I ask what is the spread of frequencies, that is what are the band of frequencies that are present. Here the main drawback of this spectrogram or scalogram again by virtue of the duration bandwidth principal, that information is not going to be really accurate. What we mean by this is, at a given time t, let us say there is a certain band of frequencies in the signal. And you turn to the spectrogram to determine this answer, that is I pick this time t and I ask I realize that these are the duration, this is the band of frequencies that are present. This answer will not exactly match with the reality there, but it is much better than what Fourier analysis can give you, because Fourier analysis cannot give you any information of what is happening in time t.

Why does this occurred is spectrogram and scalogram and so on? Because we are windowing, the moment you window the signal, that means your segmenting and windowing. The properties of the window and the signal, they actually convolve with each other, as we will show later on, as a result of which the local spread that you see is not truly the spread of the signal at the time t, but is also masked is in fact, masked by the spread of the windows. So, the window adds on or spoils the game here to a certain extent. Therefore, what you see truly here, at least visually is not necessarily the true spread of frequencies in that signal, in that signal at that time t, and so is the case with scalogram as well. So, if I ask the simple question, what should be the conditional bandwidth of a pure sine wave, right? What, this is a very fairly easy question to answer. I have a pure sine wave, it has the same frequency at any time t, and ideally I should get a flat line in the energy density. So, let me just show that to you on the board.

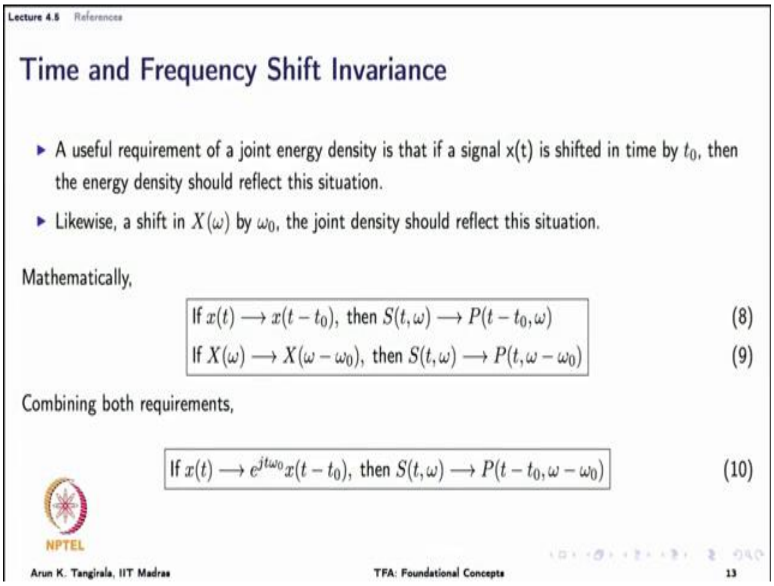
(Refer Slide Time: 03:29)



So, if I am looking at the joint energy density of a sine wave, ideally I should see a flat line. So, let us say this is the frequency present in a sine wave. This is what I should see coming out of a spectrogram and scalogram and so on. Unfortunately, in the spectrogram and the scalogram I will not see this line. In fact, I will see masking of this line and I would see a band. We have seen this in other slides as well. So, you would see a band of frequencies.

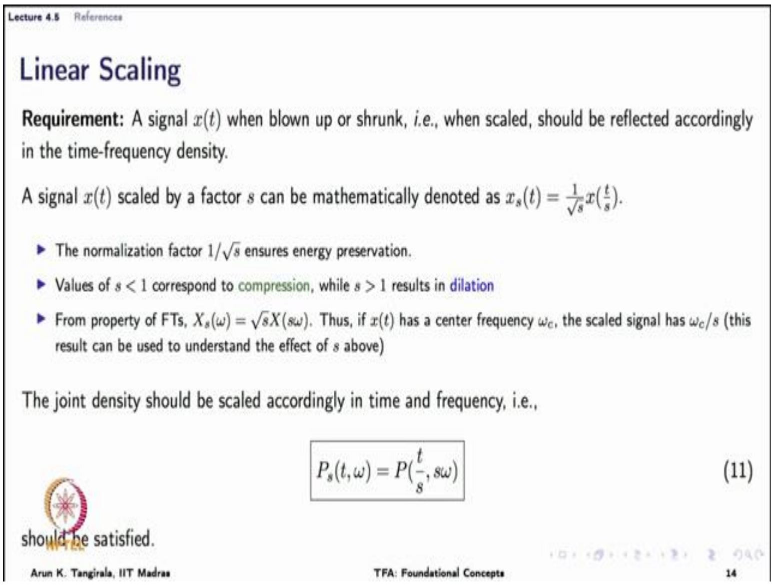
So, why does this band of frequencies arise? Because of the presence of the window, you are looking at a signal though a window. Therefore, your perception of the signal properties is masked by the properties of the window. So, you have to unravel or uncover this signal features knowing the properties of the window that you can use. You can be successful only to a certain extent, that is again the limitation coming by virtue of the duration bandwidth principle. In other words, whatever properties you really try to infer from this spectrogram or scalogram and so on, are not directly or truly the properties of the signal. There is always this windowing effect coming in to the picture, right. And in fact, towards the end of this lecture, I have a slide that shows you what is happening, what I mean by this, but this is fairly easy example to understand.

(Refer Slide Time: 05:04)



So, the other property that the joint energy density should satisfy is a time and frequency shift invariants. What this means is, if I shift the signal or the feature of the signal in time, the joint energy density should also reflect that. And if I have a shift of frequency, then the joint energy density should also show that. Together I can state this condition, I can say, that if the signal is shifted in time and frequency modulated, then the joint energy density should reflect the shifts accordingly, that is, it should show a shift in time t naught and a shift in frequency omega naught. Once again, here please read the Ps as S. It is just a typographical error.

(Refer Slide Time: 05:51)

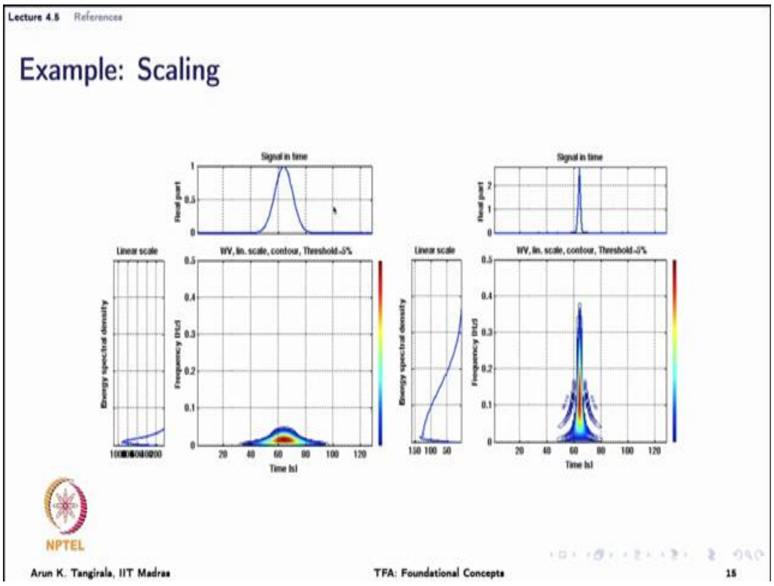


The next property of interest is linear scaling. This is a very important property because if the feature, let us say, existed for a certain time t and again and appeared later on, but in a compressed manner. So, the feature that existed only for a short a period of time, then the joint energy density should reflect that. We call this as a linear scaling property.

And we have discussed the scaling earlier. We denote the scale signal as X subscript s. And we know, x of t over s is what essentially is a scale version. In addition, I have a 1 over root of s. This factor is only to ensure, that the scale signal has same energy as the original sector.

So, what do I want when I perform a joint energy density of this scale version? That it should actually show up here as accordingly, there should be a, if, if this, the feature is dilated in the signal, then the joint energy density should reflect this dilation. And remember, dilation or compression results in a frequency shift, therefore the joint energy density again should show this frequency shifts.

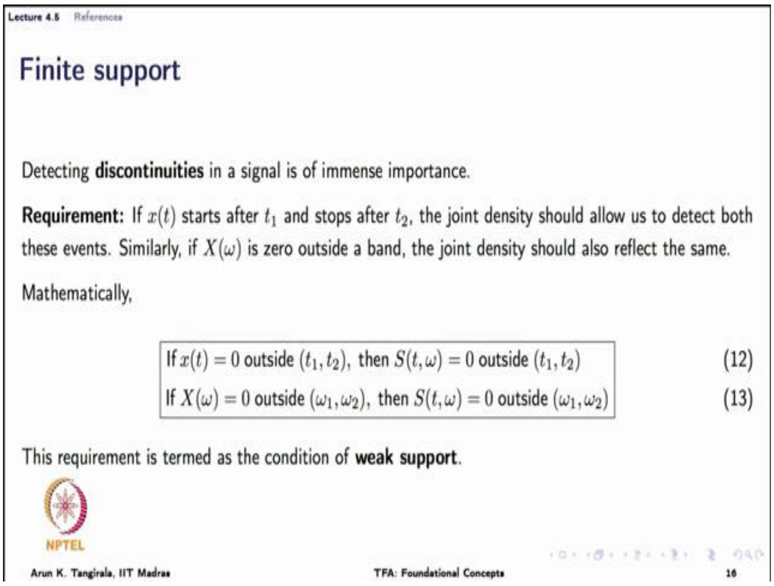
(Refer Slide Time: 07:13)



Let me just show that with an example for the Wigner-Ville case. So, here I have the signal whose Wigner-Ville distribution is shown here. On the left you have this spectral density, but let us look at the Wigner-Ville distribution. On the right hand side I have a compressed version of the same signal, now it exists for a finite time. Look at what compression does.

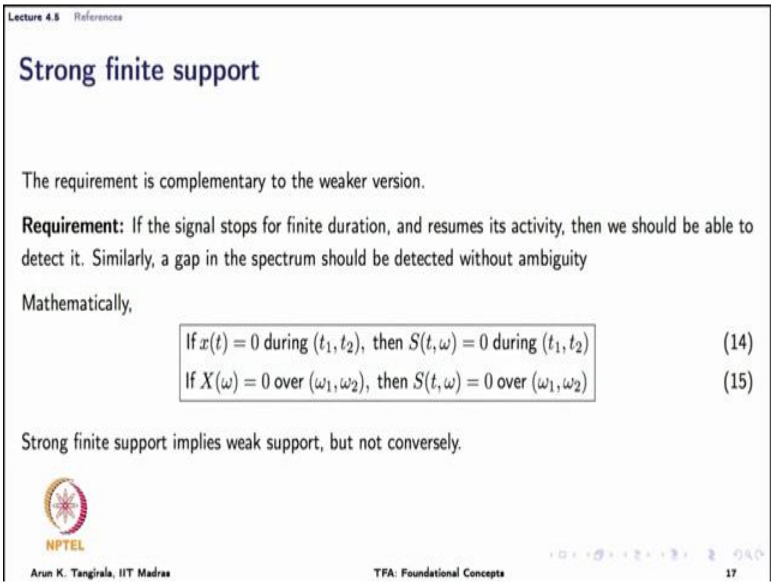
It actually produces a broader set of frequencies, right, that should be accepted now and also a shift in centre frequency, but let us look at the Wigner-Ville distribution. Is it sensitive to this compression? Yes, so the moment I compress this signal, the density also gets squeezed in time, but also is spread in frequency. That you cannot escape again by virtual of the so called uncertainty principle. But that is what we want. Every technique should be sensitive to this compressions and dilations in the signal. How sensitive each technique is depends on that particular method.

(Refer Slide Time: 08:13)



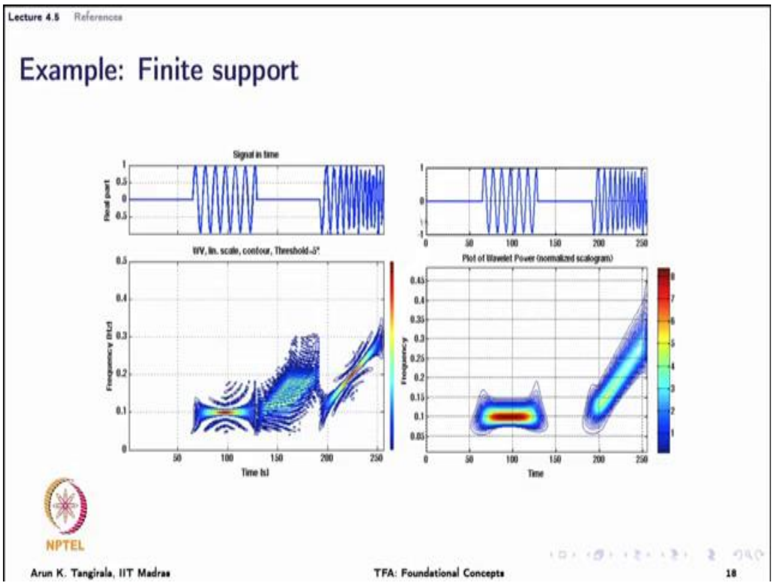
And the last property when it comes to energy densities is that of the finite support. There are two versions of this. One is called a weak support requirement, other is called a strong support requirement. In the case of weak support what I require is, if the signal exists only for the finite time and is 0 outside an interval, then so should be the energy density. Likewise, if the signal is band limited within omega 1 and omega 2, then the energy density should also exactly be band limited within that interval.

(Refer Slide Time: 08:53)



The stronger requirement is, if the signal is 0 for the finite period of time. Earlier, we said signal is non-zero for a finite period of time. Now, we are saying if the signal becomes inactive over a certain interval of time, then so should be the energy density. Likewise, for frequencies, if the signal does not have frequencies within a certain band, then energy density should also be reflective of the same.

(Refer Slide Time: 09:20)



So, let us look at this example here. Again, now I have on the left Wigner-Ville distribution, but on the right I have scalogram. This tells us, gives some insights into how these two methods behave with respect to this property. So, I have a sinusoid. So, the signal start from 0, remains 0 for a while, remains idle, then a frequency sets in diced on, abruptly remains. The signal remains inactive for a while, and then a linear chirp is generated.

So, what does the Wigner-Ville distribution do for me? It, it is exactly non-zero over this time interval, right, and also tracks the frequency changes exactly over the time interval it exists. So, that means, it has perfect finite support, it is very good. There is no ambiguity with respect to time localization at all here.

However, I have interferences. This is one of the other disadvantages or demerits of a Wigner-Ville distribution, that I will have interferences between signal components. This interference, again, is on the same lines as we talked about earlier with when we talked about non-additivity of spectra. You can show, that there exists an interference term or a cross term whenever I evaluate Wigner-Ville distribution of a sum of signals. So, you can imagine, that I am adding up two signals here.

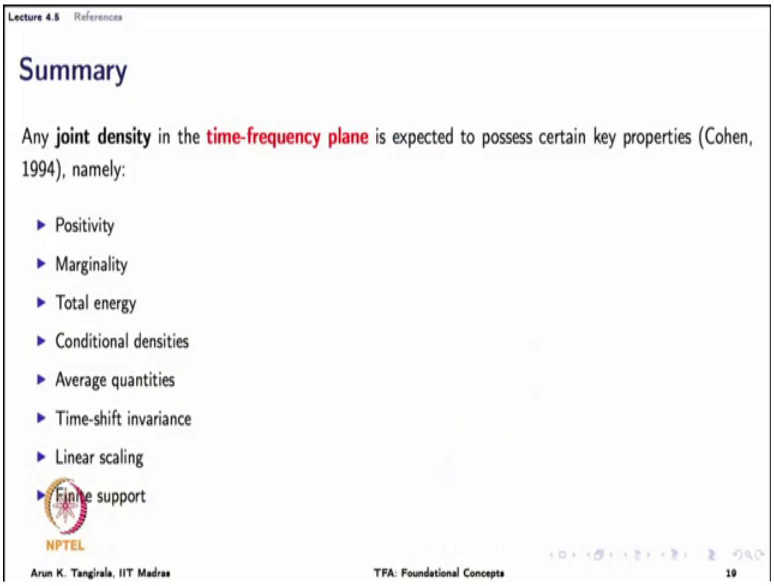
Turning to scalogram, for the same signal the interference term is gone. However, what has happened, the time ambiguity section, because now the energy density is actually smeared in time, the energy density is non-zero even when the signal is 0 and remains non-zero even after the signal has died down, which means, it does not perfectly satisfy the finite support requirement. But two advantages, one the interference term is gone and secondly, the energy density is guaranteed to be non-negative for all classes of signals.

Now, what is interesting, that we will learn later on is, let us say, I start with Wigner-Ville distribution and I say, wow this is very nice. I have finite support properties and I would like to retain this, but I only want to get rid of this interferences. If only I can get rid of these interferences, I would have a wonderful distribution, which localizes the time activity of the signal perfectly. However, it turns out, that when I sit down to remove this interferences I end up spoiling the finite support property, that is, the trade of that you have and that is again taking you back to the Wigner’s result.

If I want 0 interferences, then I have to sacrifice on the time ambiguity and a few other properties as well. In fact, you can show with a certain choice of method for removing these inferences, you can end up with scalogram. What this means is a very unifying result. You can say, at the centre of all the joint energy densities I have the Wigner-Ville distribution, which has certain very nice properties except positivity and this interference business. And if I start modifying this Wigner-Ville distribution in a bit to remove the interferences and also guaranty positivity, if I go in one direction, I will end up with scalogram. If I go in another direction, I can show I will end up with spectrogram. So, this is the unifying result that came about.

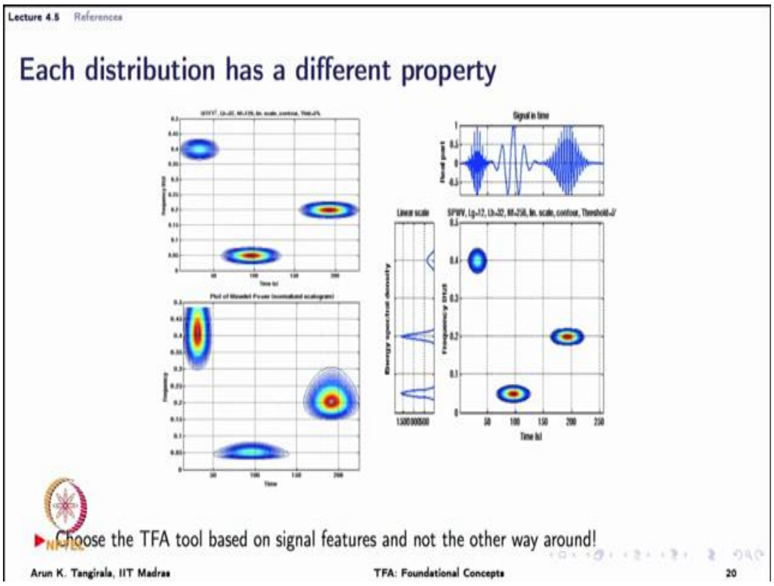
And largely due to Cohen's work, then you have what have what are known as Cohen’s class of distributions. What this means is, you can view scalogram, spectrogram as a special case of Wigner-Ville distribution, or you can argue the other way round. You can say, I can start with scalogram and then require, that it has perfect finite support and all the other properties, that Wigner-Ville has marginality properties and I start modifying this scalogram, I will end up with the Wigner-Ville, fine. So, you cannot really prove, that one is the father of the rest. You, you can always have one as a starting point and arrive at the remaining. That is something that we will learn later on and also notice for the chirp signal. Wigner-Ville distribution is very fine localization in frequency, whereas the scalogram is whites smearing, right. So, you have to now decide what signal features are important to you or what is the, what are the characteristics present in the signal and pick the corresponding technique. In fact, one of, there is a slide where I will ask you to think about it by means of an example.

(Refer Slide Time: 14:05)



So, to summarize, the joint energy density should satisfy and number of properties, positivity, marginality, total energy and so on. And the main point is, there does not exist a technique which satisfies all of this. You have to sacrifice one for a few others.

(Refer Slide Time: 14:20)



And this is what exactly I meant earlier. You will you have to choose the time-frequency analysis based on signal properties. So, here I have the signal made up of three different features. I have amplitude modulations, but different frequencies. They exist only over finite time roughly.

And what I am showing you here is a spectrogram on the top, below here is a scalogram in the left and then I have the Wigner-Ville distribution on the right. Now, you think which distribution? It is not exactly Wigner-Ville distribution, by the way. This is called a pseudo smooth or smooth pseudo Wigner-Ville distribution, smooth in a sense have removed the interferences and pseudo because I use some window function and so on. Whatever it is, it is some variant of the Wigner-Ville distribution and so are scalogram and spectrogram. But typically, we do not refer to scalogram as a variant of Wigner-Ville because the scalograms are constructed directly through wavelet transform.

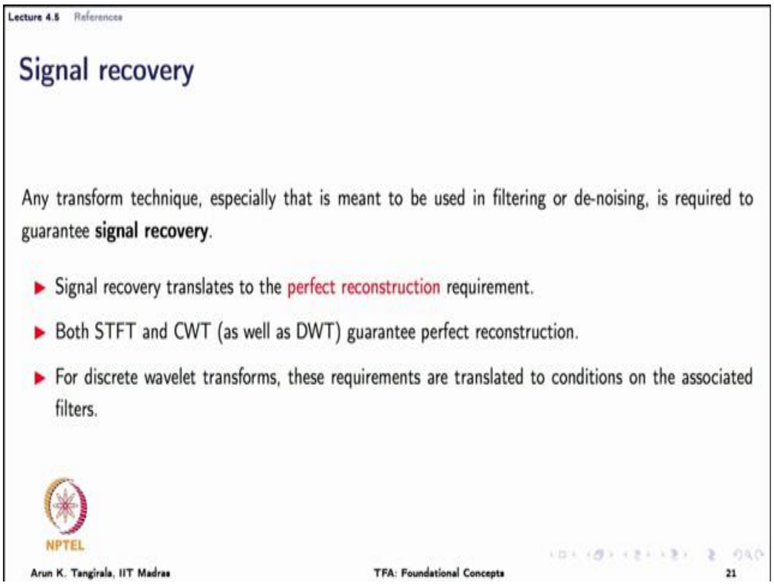
So, the question that I have for you is which distribution is good here for this signal? Obviously, the Wigner-Ville distribution gives you very nice localizations, but this plot also tells you what is happing in the scalogram. Look at this. In the scalogram I have very fine localization. Relatively speaking, I have fine localization of frequencies in the lower frequency region, whereas a broad smearing of frequencies in the higher frequency region. We have pointed this out in lecture 1.2 where we talked exclusively about wavelets, at least gave an overview.

We said, wavelets are ideally suited for low frequency components where they are long lived and high frequency signals that are short lived, which means, it will give you good localization of energy in time for high frequencies, but poor localization of energy for the same component in the high frequency reign and vice versa for the low frequency. That is exactly what you see here.

The short-time Fourier transform or the spectrogram has uniform localization throughout the time-frequency plane. It does not matter whether you are working in low frequencies or high frequencies. This smearing of energy is more or less the same, right. It, it is, in fact, if it is same, it does not matter and that is one of the drawbacks that is considered for a spectrogram because when you move to high frequency regions. Typically, you want very good time localization because of the assumption, that high frequency components are short lived, but if that is not the case, then it is different story. Typically, high frequency components, that means, sudden changes, abrupt changes are short lived. If that is the case, then spectrogram is not right because there I need good time localization of the energy and that is not happening here.

Look at the time localization of the scalogram. It is much better than that of the spectrogram, but of course, it has sacrificed on the frequency localization, that is exactly the duration bandwidth principle working for you there. The Wigner-Ville distribution has another trade-off for the time frequency localization, but it is very nice. In fact, it is ideally suited for these kinds of amplitude modulations and so on. So, I would pick this distribution when it comes to the signal, but it is not necessarily the right choice for all classes of signals. If you want to detect discontinuities for example, wavelet transform is perhaps the best among these three. So, think about it.

(Refer Slide Time: 17:58)

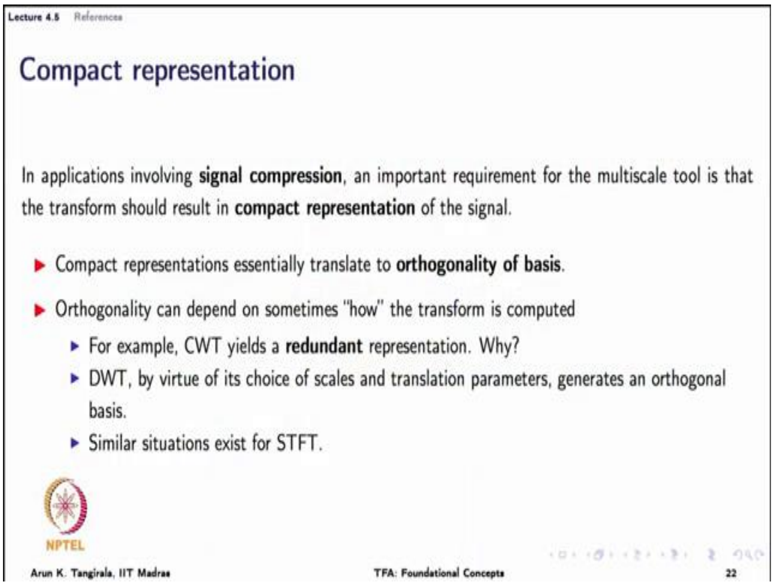


We will conclude this lecture with two other requirements that we talked about early on. One is filtering and de-noising. One of the key requirements for using a particular tool in multiscale filtering or multiscale de-noising is the requirement of signal recovery or what is known as a prefect reconstruction. What this means is, I transform the signal whether it is short-time Fourier transform or wavelet transform, I transform, perform certain operations in the transformed domain or if I do not do anything and I take an inverse transform, then I should be able to recover. In filtering and de-noising what is done is, you transform the signal, perform certain operations in the transform domain and then recruit and inverse transform, you should be able to get the filtered version.

The perfect reconstruction requirements says, that if you do not do anything in the transform domain and then simply apply the inverse transform, you should be able to recover the signal and both short-time Fourier transform and continuous wavelet transform guarantee perfect reconstruction under some mild conditions on the analysing functions.

For discrete wavelet transforms, you can translate this requirements of perfect reconstruction to the associated filters that you use in DWT and we will talk about it when we talk of DWT itself. Here, there is nothing about energy density and I am not even talking about Wigner-Ville distribution because Wigner-Ville distributions directly compute the energy. So, do not expect to use WVD for filtering although there are some applications reported in literature. Main point is, the phase of the signal is lost in constructing the Wigner-Ville distribution and therefore, you will not be able to perfectly recover the signal with Wigner-Ville's.

(Refer Slide Time: 19:48)



Finally, we have this notion of compact representation. Remember, when I am transforming the signal and representing the signal in a new domain. What is the purpose? What is the motivation? May be, this new representation gives me certain desirable features.

So, if I take a sine wave and I look at its Fourier representation, its Fourier representation has three numbers: amplitude, frequency and phase. We say now these three numbers represent the entire sine wave, which exist forever that is beautiful because now I have at enormous level of compression here. Instead of representing the infinitely long sine wave in time and storing it in my hard disk, I only need to store three numbers. So, that is a perfect or ideal compression that I have.

How did I achieve this compression? Because this representation achieve that and we say, then the Fourier representation for sine waves gives me the most compact representation. And how does Fourier representation achieve that? Of course, by choosing sine waves as the basis, but more importantly, this property of orthogonality is the one that renders compact representation.

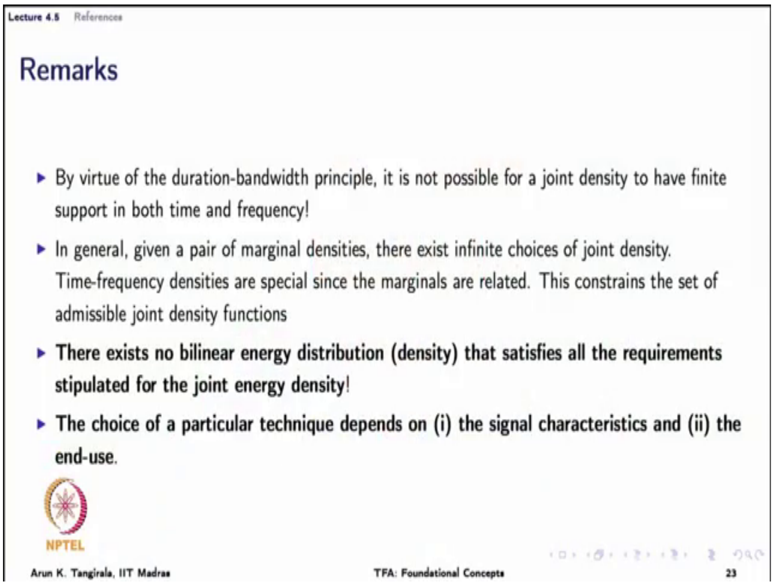
So, if I take continuous wavelet transform it yields what is known as a redundant representation. That means, what I am doing is whether it is CWT or short-time Fourier transform, I am moving from the one-dimensional case signal to a two-dimensional time frequency plane. Therefore, I am generating more numbers than what I have for the signal.

Now, the question is, now what is, what we mean by redundant is, I have many more numbers than required to represent the signal. But with DWT, I can actually bring down this redundancy by choosing to evaluate the CWT only at certain scales and translations, and that is what we also discussed briefly in lecture 1.2 when we talked about DWT. It generates an orthogonal family and there is a theory called frame theory, which talks about all of this and you can show, that DWT gives you the most compact representation among all the other choices of scale and translations and that is why in all compression applications you will see DWT being used very widely, and the same case holds for short time Fourier transform.

What are we doing in short-time Fourier transform? I am segmenting the signal, taking the Fourier transform of the segmented signal and I move along. Now, how I march along in time will determine whether I have a compact representation or a redundant representation. If I am segmenting the signal in such a way that I generate highly overlapping segments. That means, I move the window one sample at a time, then I will have a highly overlapping set of segments, which means, I will repeat the analysis many times, which is not necessary, then I will have a redundant representation.

But if I choose the segmentation in such a way, that I segment the signal in non-overlapping segments and then perform the Fourier transform, then I will have a compact representation. So, whether you have a compact representation or a redundant representation does not depend only on the transform itself, but your choice of translations and scales in the case of short-time Fourier transform. It is only translations in the choice in the case of wavelet transforms, it is both scales and translations. So, keep this in mind. With this we will close this lecture and I would like to again make a few remarks, retreating the points that I have been mentioning in this lecture.

(Refer Slide Time: 23:27)



And particularly, I would like to point, draw your attention to these two points here, that I have here in bold phase font. There exists no bilinear energy distribution that satisfies all the requirements, we have seen that. And secondly the choice of a particular technique depends on two things, the signal characteristics. Of course, you would not know a priory what signal characteristics are there. Therefore, there is some trial and error involved depending on how much prior knowledge you have. And secondly, the end use of the method. What we mean by end use is, whether you are using this analysis for feature extraction or signal compression or filtering.

If I am use, if I am looking at filtering applications typically one would turn to either a short-time Fourier transform more, in fact typically turn to wavelet transform or a short-time Fourier transform, because they allow signal recovery. But if you, if I am looking at energy densities, then the Wigner-Ville distribution also comes as a very good competitor variants of it. So, examine and think before you really jump to choosing a particular technique that is the message that I am trying to give you here.

So, this closes the topic of the foundations and concepts that I want to give you on time-frequency analysis. In the next unit we will start off on short-time Fourier transforms. We will go into the math, discuss what properties it has, where it can be used and also discuss one application of the short-time Fourier transform. So, hope you enjoy this lecture, this is going to be divided in two parts; part a, and part b. So, make sure you have listened to both parts before you move on.

Thank you.