Introduction to Time-Frequency Analysis and Wavelet Transforms

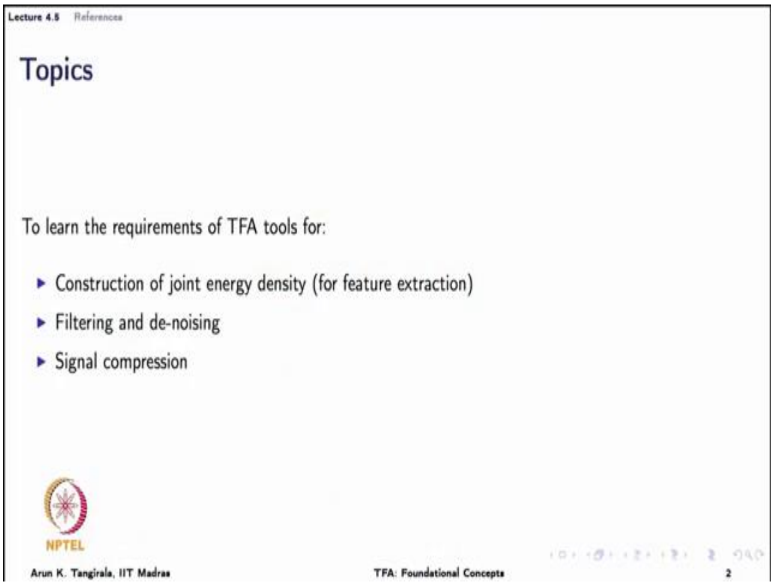
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture - 4.5
Requirements of time-frequency analysis techniques
Part 1/2

Hello friends, welcome to lecture 4.5 on the course on introduction to time frequency analysis and wavelet transforms. So, in this lecture, we are going to talk about the requirements of time frequency analysis techniques. Until now, that is in the previous, until the previous lecture we have studied the theoretical properties of energy densities. In particular, we have studied the duration band width principle; of course the definitions of duration and bandwidth as well.

So, now it is time to plunge into the techniques themselves; especially, the short time Fourier transform, the Wigner Ville distribution and the wavelet transform. But before we plunge into these tools, we should ask what do we except of them? What is typically expected of a time frequency analysis tool that I am going to use?

(Refer Slide Time: 01:08)



Lecture 4.5 References

Topics

To learn the requirements of TFA tools for:

- Construction of joint energy density (for feature extraction)
- Filtering and de-noising
- Signal compression

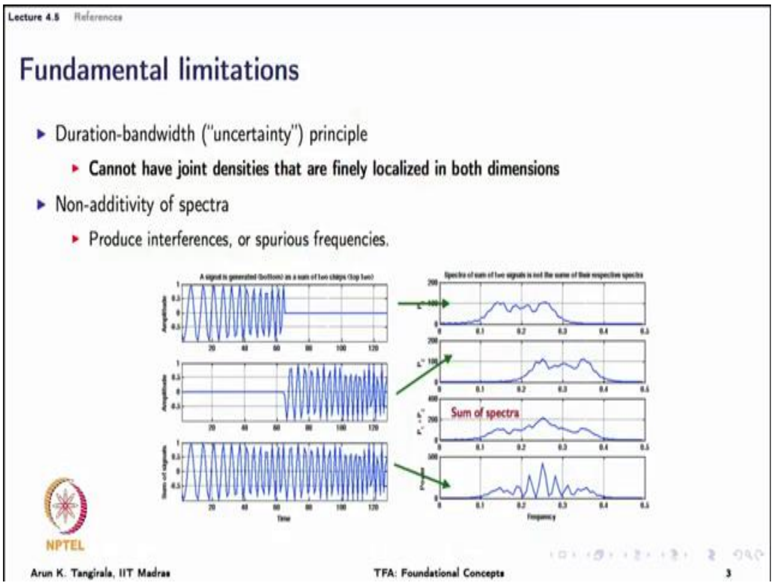
Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 2

And we are going to not list out all the requirements; we will list out the requirements with respect to specific applications. But these are broad applications namely feature extraction, secondly - filtering and de-noising, and thirdly - signal compression. The last

2 are closely related that is the filtering and signal compression, but feature extraction is a totally different set of application. Now, in feature extraction, normally we construct joint energy density, pretty much like what we did in Fourier analysis; when I want to extract the oscillatory features, I construct this spectral density and I search for peaks in the spectrum, and so on.

Likewise here in time frequency analysis when I am looking at feature extraction then I would construct a join energy density, and search for the required features in the joint energy density plot. Now, of course, once again there are several ways of constructing this joint energy density depending on the tool that you are using. But then to be able to make meaningful analysis from the joint energy density plot, we should require that these techniques theoretically satisfy certain properties. And what we will learn at the end of the lecture is that there is no tool out there which will satisfy all the requirements that we stipulate.

(Refer Slide Time: 02:38)



Lecture 4.5 References

Fundamental limitations

- Duration-bandwidth ("uncertainty") principle
  - **Cannot have joint densities that are finely localized in both dimensions**
- Non-additivity of spectra
  - Produce interferences, or spurious frequencies.

Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 3

Regardless of the application that we are looking at, that is whether it is feature extraction, filtering and de-noising, or signal compression, or approximation and so on, there are fundamental limitations to what we can do. And we have discussed this before; we have seen this before. The first limitation, obviously, being governed by the duration band width principle, which essentially says that I cannot construct a joint density that is very finely localized in both time and frequency.

If the joint density is narrow in time, it has to be broad in frequency and vice versa. This does not mean that I will not be able to locate the frequency component and so on. This is unfortunately the machine reputation that it leads to. How you locate a particular frequency, how you extract this information, depends on further coarse processing that you do of the joint energy density.

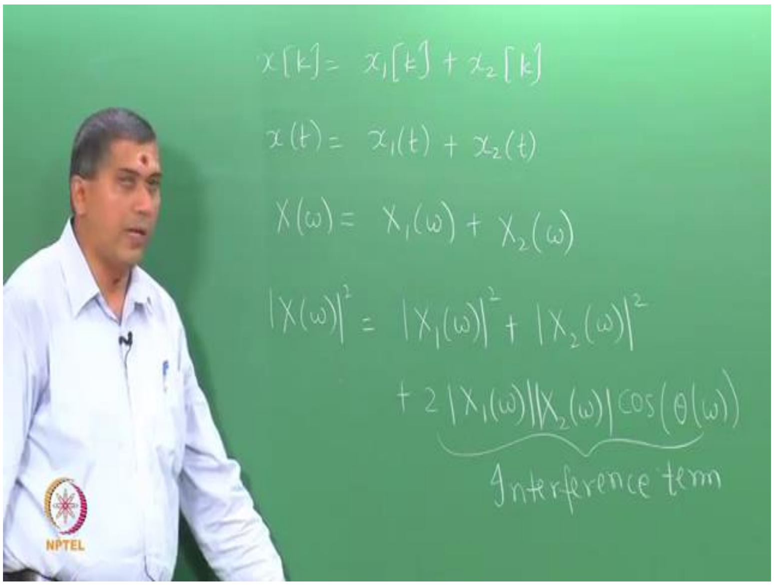
But, ((Refer Time: 03:37)), what is happening is that, the joint density is going to be finely localized only in one dimension. Now, of course, different techniques will give you different trades on the localization. And which technique is suited for a particular application depends on the signal characteristic and so on. This aspect will become clearer when we discuss these specific techniques.

The second limitation is governed by the non-additivity of spectra which I had mentioned in one of the previous lectures. Here, I have an illustration for you. I have here, on the left top 2 panels, I have 2 signals which are chirps, one dice down after while and the second one starts of after while before being inactive. Now, when I look at the spectra of these 2 signals I have these spectra; they are, obviously, they have different frequency content.

Let us say, I add up these 2 signals and I look at the spectrum of the added signal; that is now I have just simply numerically added these 2 signals. So, this spectrum of the added signal or the spectral density is actually shown here at the bottom. Above this bottom panel on the right hand side, I have the sum of spectra of the individual signals.

If things were ideal in an ideal world, I would expect spectra to add up, so that when I look at these spectrum of the added signal there is no difference between the added spectra and spectrum of the added signals, and then there are no artifacts. Unfortunately, when I add 2 signals, I have what are known as interferences, right. And let me explain that to you; why, what is this interference term? Why does it arrive?

(Refer Slide Time: 05:37)



$$x[k] = x_1[k] + x_2[k]$$

$$x(t) = x_1(t) + x_2(t)$$

$$X(\omega) = X_1(\omega) + X_2(\omega)$$

$$|X(\omega)|^2 = |X_1(\omega)|^2 + |X_2(\omega)|^2 + \underbrace{2|X_1(\omega)||X_2(\omega)|\cos(\theta(\omega))}_{\text{Interference term}}$$

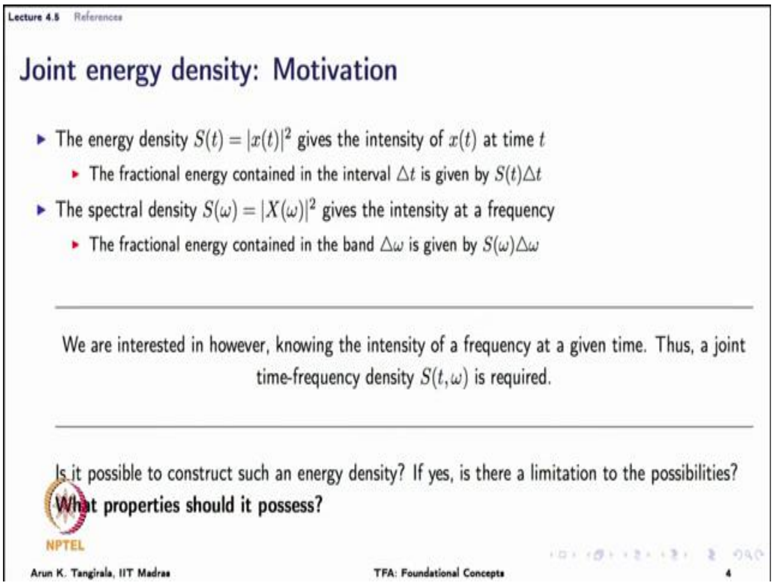
So, let us take a signal that is made up of 2 signals, x 1 and x 2; it could be even in continuous time as well, does not matter. Now, when I take the Fourier transform of each of these signals, discrete time or continuous time, by the linearity property the Fourier transform of the composite signal is the sum of the respective Fourier transform of the individual ones.

Until here, things are alright, but when I am looking at the spectral density of the composite signal, and when I workout the math for this spectral density from the theory of complex numbers, I know that when I add 2 complex numbers, x 1 and x 2, are complex valued numbers. In the right hand side I will have the respective squared magnitude plus this term here. The cross term, let us say, I have here, theta of omega. This theta of omega is the phase between these 2 signals at that frequency. Unless this theta is 90 degrees, this is cross term will not vanish.

In other words, if the signals are orthogonal or uncorrelated at that frequency, you will not have this spectrum of the, spectral density of the composite signal being the sum of the respective spectral density. So, this is the interference term that causes the appearance of spurious features in the spectral density, spectral plot of the composite signal that you see at the bottom panel on the right hand side.

So, the appearance of this interference terms, in fact, causes problems in interpretations, because now I have spurious features which are purely artifacts of mathematical operations. They are not necessarily a part of the composite signal at all. And this is going to be the case in all time frequency analysis techniques because I am going to have typically non stationary signals where the frequency is changed with time. And this example is a representative of all those situations; and therefore you will run into interferences and so on.

(Refer Slide Time: 08:23)



Joint energy density: Motivation

- The energy density $S(t) = |x(t)|^2$ gives the intensity of $x(t)$ at time t
  - The fractional energy contained in the interval Δt is given by $S(t)\Delta t$
- The spectral density $S(\omega) = |X(\omega)|^2$ gives the intensity at a frequency
  - The fractional energy contained in the band $\Delta\omega$ is given by $S(\omega)\Delta\omega$

We are interested in however, knowing the intensity of a frequency at a given time. Thus, a joint time-frequency density $S(t, \omega)$ is required.

Is it possible to construct such an energy density? If yes, is there a limitation to the possibilities? **What properties should it possess?**

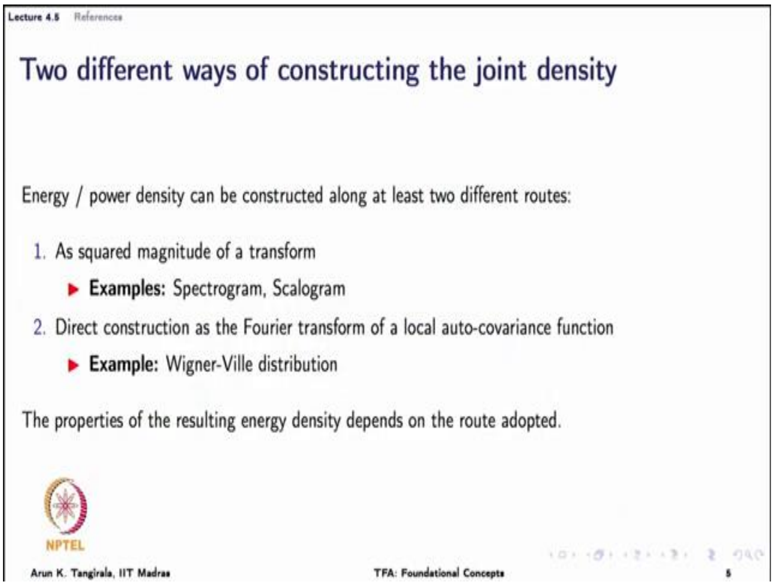
So, let us look at first the joint density or the joint energy density. The motivation for looking at joint energy density is fairly obvious by now. In time, if I look at the energy density in time it gives me the intensity of the signal at time t. And if I am looking at the spectral density it gives me the intensity at particular frequency. And then the respective fractional energies in the interval or in the band can be computed. However, in practice, quite often I would be interested in knowing the intensity of frequency at a given time, and that is why I would need the joint energy density.

As I mentioned early, on assuming that I am interested in extracting the features I am constructing the joint energy density. If I am interested in filtering, maybe I do not need to construct a joint energy density at all. I simply work with the transforms, perform certain operations in the transform domain, reconstruct, and I get the filtered component. There is no need to look at joint energy density at all. However, the principles governing the construction of energy densities or filtering or signal compression and so on, more or less are the same; that is the limitations governing these are more or less the same.

So, coming back to the energy density, is it possible to construct an energy density which will give me this information? That is I would like to know the intensity of a frequency at a given time t. So, I am asking for a very fine localization in time and in frequency domain; and obviously, that is not going to be possible, given the duration bandwidth principle. That is the basic limitation that just comes into the picture immediately.

And so therefore, I have to now step back and say, well, I am going to restrict myself to an interval in time or band of frequencies. Even there, what are the possibilities? What are the tradeoffs involved? And most importantly, what properties should the joint energy density possess? By properties what we mean is if there is a certain change in the signal that should be reflected accordingly in the joint energy density and so on. This will become clearer as we move along.

(Refer Slide Time: 10:47)



Lecture 4.5 References

Two different ways of constructing the joint density

Energy / power density can be constructed along at least two different routes:

1. As squared magnitude of a transform
   - **Examples:** Spectrogram, Scalogram
2. Direct construction as the Fourier transform of a local auto-covariance function
   - **Example:** Wigner-Ville distribution

The properties of the resulting energy density depends on the route adopted.

NPTEL

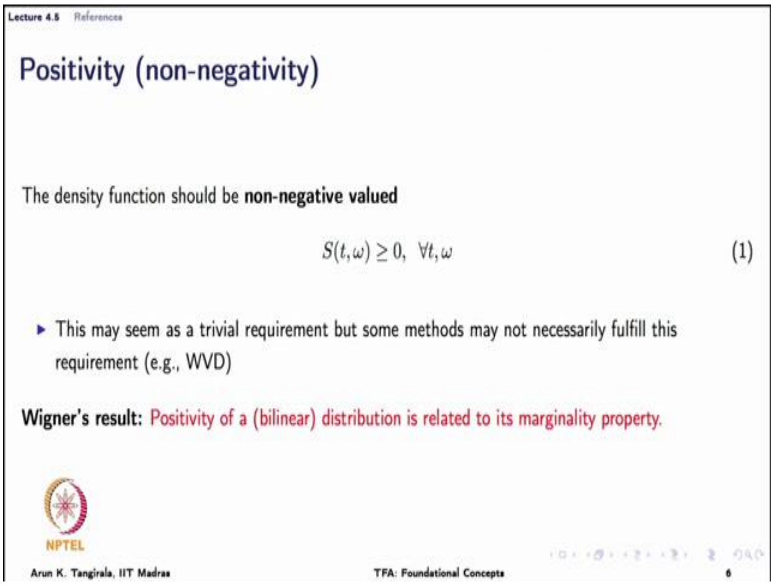
Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 5

Before we look at the requirements of a joint energy density, what is important to keep in mind is, there are at least 2 different ways of constructing the energy density or even the power densities. One as a squared magnitude of a transform; we have seen this earlier as well in the Fourier analysis or a Fourier domain. We take the Fourier transform and then take the squared magnitude, that will give me the density. And this is the principle that is followed in constructing the spectrogram and scalogram.

The other alternative is, by first constructing the auto covariance function and then taking the Fourier transform. In Fourier analysis, that is in pure frequency domain analysis, we would just construct the ordinary auto covariance function and then take the Fourier transform, and this is what we referred to us ((Refer Time: 11:38)) theorem and so on. A similar idea exists in time frequency analysis as well where we construct what are known as local auto covariance functions.

Now, the auto covariance function is not going to be simply a function of a single quantity lag. It is going to be a function of 2 things, lag and the local time, but still I can take a Fourier transform of that and that will give me the local energy density in the time frequency plane. This is the idea that is used in Wigner Ville distribution, alright. Now, does it matter really how I construct these densities? Well, yes, because the properties of the resulting energy density can vary.

(Refer Slide Time: 12:22)



Lecture 4.5 References

Positivity (non-negativity)

The density function should be **non-negative valued**

$$S(t,\omega) \geq 0, \ \ \forall t,\omega \qquad (1)$$

- This may seem as a trivial requirement but some methods may not necessarily fulfill this requirement (e.g., WVD)

Wigner's result: Positivity of a (bilinear) distribution is related to its marginality property.

NPTEL

Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 6

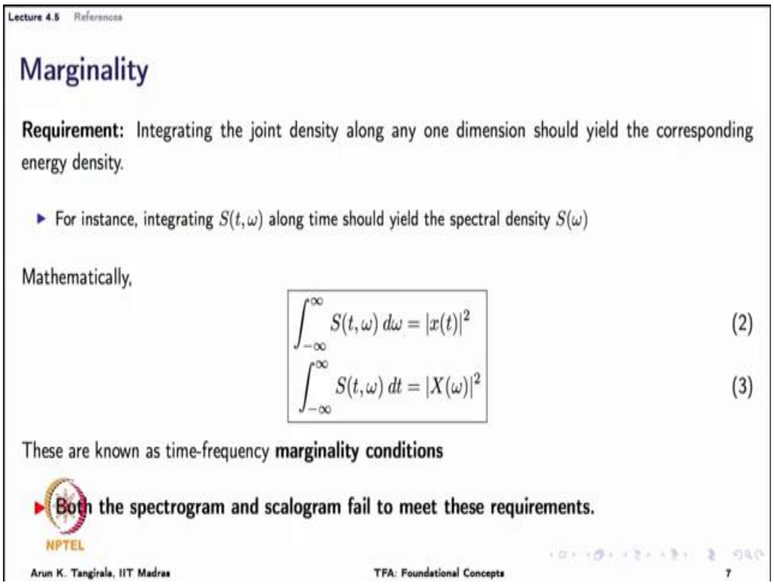
And one such property is the positivity which is the most important or non negativity we may say, because the spectral density can be 0. So, the appropriate term is non negativity. Now, the fundamental requirement for any density function to be meaningful is that it should be non negative value. And this depends on that is whether this property is satisfied by the joint energy density depends on how you have constructed the energy density.

So, as we have just discussed, there are 2 different ways of constructing this. In the first method, I am manually constructing the energy density which means I take a transform and then take the squared magnitude; what is it mean? It means that I am guaranteeing non negativity because squared magnitudes of a complex number cannot be negative. Therefore, this property is immediately satisfied by spectrogram and scalogram.

On the other hand, there is no guarantee that the direct construction method will give you a non negative spectral density in the time frequency plane. In the pure frequency plane, the real value and positivity is guaranteed by certain properties of the auto covariance function. But, in the case of the time frequency analysis, this is not necessarily guaranteed. In fact, the fundamental result proved by Wigner states that the positivity of a bilinear, bilinear means you can say quadratic; that is squad quantities of the transforms or where signals are appearing in multiples as products.

Then, there when you look at the distribution that is constructed in a bilinear manner, there is no way actually you can guarantee positivity and other properties to be satisfied simultaneously. Particularly what Wigner said is, showed is, that the positivity of distribution or a density is very closely related to what is known as the marginality property, that we will study next. And this was the fundamental result, and it governs all the bilinear or quadratic time frequency distributions. The consequence of this will become more clear when we study the Wigner Ville distribution, but let us study the marginality property first.

(Refer Slide Time: 14:44)



Lecture 4.5 References

Marginality

Requirement: Integrating the joint density along any one dimension should yield the corresponding energy density.

- For instance, integrating $S(t,\omega)$ along time should yield the spectral density $S(\omega)$

Mathematically,

$$\boxed{\begin{aligned}\int_{-\infty}^{\infty} S(t,\omega)\, d\omega &= |x(t)|^2 \quad (2)\\ \int_{-\infty}^{\infty} S(t,\omega)\, dt &= |X(\omega)|^2 \quad (3)\end{aligned}}$$

These are known as time-frequency **marginality conditions**

- **Both the spectrogram and scalogram fail to meet these requirements.**

Arun K. Tangirala, IIT Madras   TFA: Foundational Concepts   7

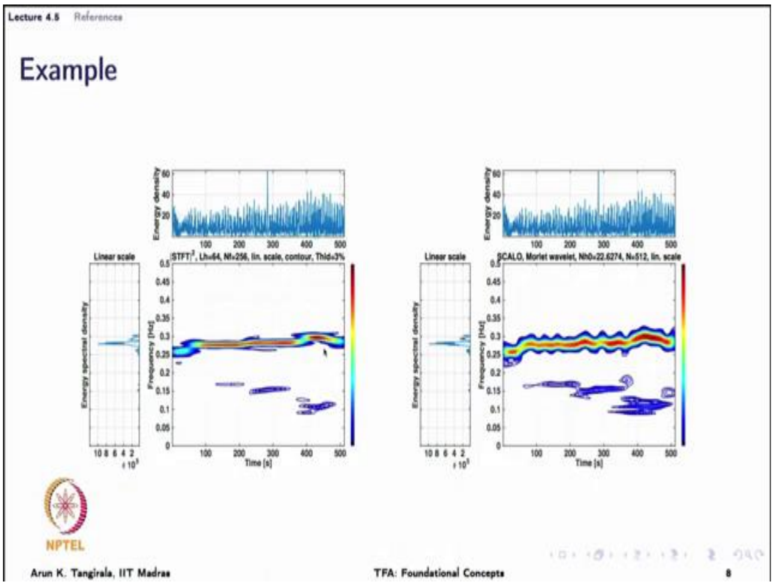
So, what does marginality mean? Marginality essentially means that if I sum up the energy density in the time frequency plane along one dimension I should be able to recover the energy density in the other dimension. This term marginality is also encountered in probability theory and theory of random variables and so on.

Now, just a word of caution, although I keep giving you the analogy with the probability theory, it does not mean that it is a prerequisite for understanding this course. I am just giving you this analogy, so that if you are familiar with probability theory you may be able to easily follow these things. But, you do not have to have a base in probability theory necessarily to understand any of the techniques that we are talking about.

So, going back to the analogy of probability theory, we speak of joint densities in probability theory as well when we are looking at more than 1 random variable. There we have joint densities and the concept of marginal densities. When I integrate the joint density along one random variable dimension then I recover the marginal density for another dimension. So, the requirements are stated in equations 2 and 3. If I integrate the energy density along the frequency domain I should recover the energy density in time, and likewise for frequency.

At this moment it may appear strange that there could be techniques that do not necessarily satisfy this. Why should they not satisfy? One quick way of understanding this is, remember techniques such as spectrogram and scalogram or energy densities resulting from spectrogram and scalogram are coming out of a manual construction; that is, you are actually constructing the energy densities. And remember, there is a certain windowing that is involved. And both these windowing and the manual construction of the energy density are the ones that can spoil these requirements. And they do, in fact, spectrogram and scalogram fail to satisfy the marginality conditions. Let me explain this to you with a simple example here. It is not a complete example but I will a take the support of this example to illustrate the point.

(Refer Slide Time: 17:04)


Lecture 4.5 References

Example

Arun K. Tangirala, IIT Madras

TFA: Foundational Concepts

8

So, now, I have this signal coming from an industrial process. It is coming from a control loop. It is, the plot on the top is not the signal. It is actually the energy density in time such as squared signal itself. And the plot on left here of the left figure itself, is the spectral density. And here, the big plot, the 2 dimensional plot is the energy density in the time frequency plane where the energy is represented by the color.

Now, obviously, the spectral density tells me that the signal is oscillatory, and of course, and also that there are some other accompanying frequencies but to a minor extent. The joint energy density here constructed from spectrogram, tells me something more than what this spectral density can tell me. It tells me that yes there are these frequencies present in this signal over these times, but the intensity of this frequency varies with time.

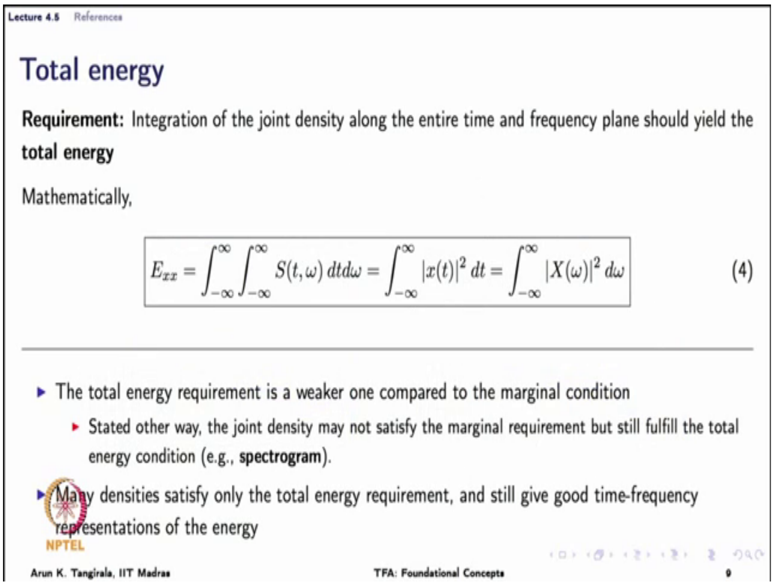
So, you can see from changing colors here that the intensity is maximum here in this interval, and it is low here. In fact, there is a shift of frequency from this region to the main dominant frequency, and then another mild shift in the frequency. So, what is marginality requirement here mean, if I, let us say, I pick this dominant frequency, and I walk along time, add up all the energy densities that I have from spectrogram, I should be able to recover this peak here. Now, it turns out that it does not. And I will show you this in more in details when we talk about spectrogram, but that is what marginality requirement means.

In fact, the other requirement is I pick a certain time t, and I walk along the frequency domain axis, add up all the energy densities, I should be able to recover the value of the energy density at that point in time. That is also not satisfied by spectrogram. Likewise, for the scalogram here as well; on the right hand side you have the scalogram. I have constructed this using the Morlet wavelet. Both these plots are generated using the time frequency tool box. The scalogram also fails to satisfy the marginality requirement.

And of course, the other thing that you should also notice here is, the loss of localization in frequency. Look at how finely this spectral density is localized in frequency that is a Fourier spectral density. Whereas, whether it is a scalogram or the spectrogram there is a certain smearing of the energy; there is a spread of energy in the frequency domain. The localization is not as fine as what you see in spectral density.

And again, this is the consequence of the duration bandwidth density. Because, I want one foot in time, I can only keep one foot in frequency domain, and therefore, I have this loss of localization. Remember, I do not use the term resolution; I am constantly using the term localization. So, hope, now that the marginality property requirement is clear.

(Refer Slide Time: 20:20)



Lecture 4.5 References

Total energy

Requirement: Integration of the joint density along the entire time and frequency plane should yield the **total energy**

Mathematically,

$$\boxed{E_{xx} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} S(t,\omega)\, dt d\omega = \int_{-\infty}^{\infty} |x(t)|^2\, dt = \int_{-\infty}^{\infty} |X(\omega)|^2\, d\omega} \qquad (4)$$

- The total energy requirement is a weaker one compared to the marginal condition
  - Stated other way, the joint density may not satisfy the marginal requirement but still fulfill the total energy condition (e.g., **spectrogram**).
- Many densities satisfy only the total energy requirement, and still give good time-frequency representations of the energy

Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 9

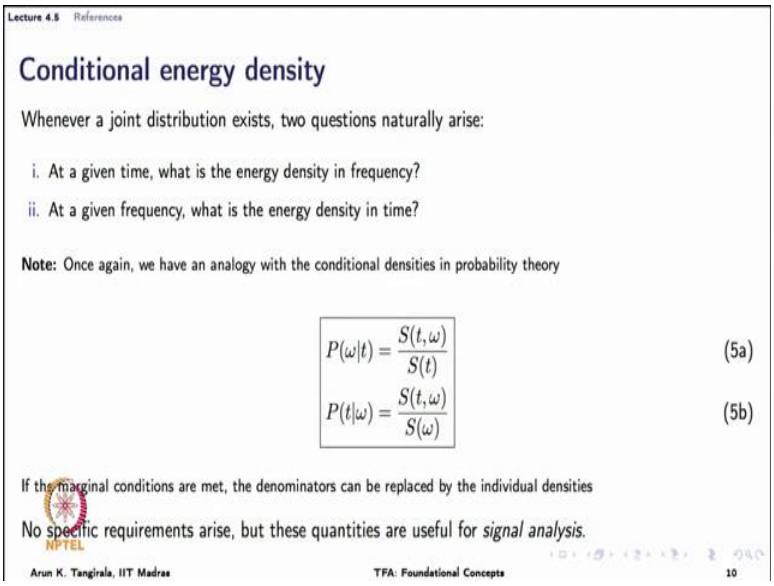
A weaker requirement is that the area under the energy density should yield the total energy which is generally satisfied by all these techniques that we are looking at, whether it is spectrogram, scalogram or Wigner Ville, and so on. In fact, I should mention that Wigner Ville distribution satisfies the marginality property. I did not show you the Wigner Ville distribution for that example, but that is the beauty of the Wigner Ville distribution.

But the biggest problem with it is, it is not, does not guarantee non negative spectral or the energy spectral energy densities. And that is what essentially Wigner’s result says, if a particular energy density satisfies the marginality requirement it cannot be guaranteed to have non negative value energy densities and vice versa. if it is satisfies the positivity or non negativity requirement then it cannot satisfy the marginality requirement.

So, coming back to the total energy requirement, this is a weaker requirement because if the area under the energy density adds up to the total energy it does not necessarily mean that the marginality properties are satisfied. So, going back to this figure here if I add up the energy densities here and taking into account the windowing that I must have done in either in short time Fourier transform or constructing the or the wavelet transform.

Once I take into account the windowing effect and I construct the total energy density, still, I will, then I will able to satisfy this requirement; it will add up to the total energy, but that does not necessarily mean that the marginality property is satisfied. And many densities satisfy the total energy requirement. This is a very important requirement; energy preservation is quite important; otherwise, then you would get absurd results as well.

(Refer Slide Time: 22:17)



Lecture 4.5 References

Conditional energy density

Whenever a joint distribution exists, two questions naturally arise:

i. At a given time, what is the energy density in frequency?
ii. At a given frequency, what is the energy density in time?

Note: Once again, we have an analogy with the conditional densities in probability theory

$$\boxed{\begin{aligned} P(\omega|t) &= \frac{S(t,\omega)}{S(t)} \qquad (5a) \\ P(t|\omega) &= \frac{S(t,\omega)}{S(\omega)} \qquad (5b) \end{aligned}}$$

If the marginal conditions are met, the denominators can be replaced by the individual densities

No specific requirements arise, but these quantities are useful for *signal analysis*.

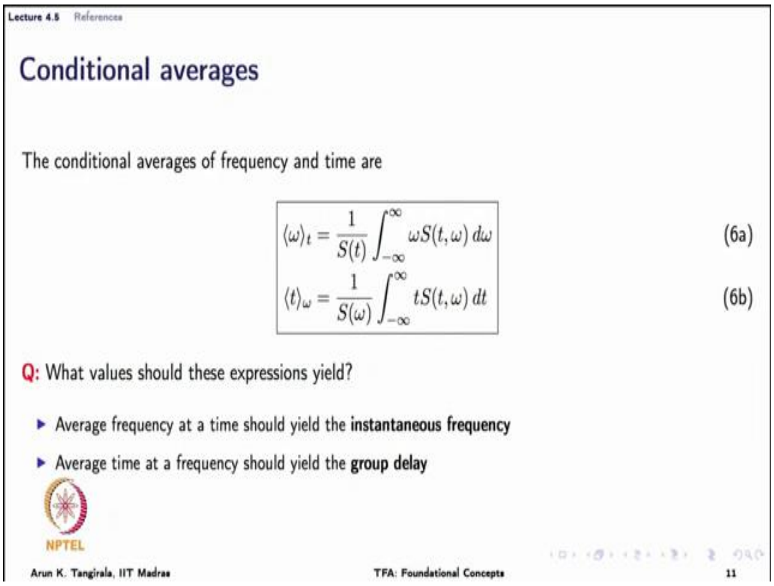
Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 10

Now, very often, in fact the purpose of time frequency analysis is to find the local properties of the signal in the time frequency plane. That is the most important purpose of time frequency analysis; otherwise, why would I break my head in evaluating this 2 dimensional transform; I would simply live with the Fourier transform. Therefore, now it is time to talk about the local properties. What are these local, how are these local properties evaluated, in terms of the conditional properties.

And specifically, you want to know, at a given point in time, what is the energy density in frequency? And if I pick a certain frequency, what is the energy density in time? So, I am specifying locally what is, I am asking locally what is happening. And to be able to construct this, we construct what are known as the local or the conditional energy densities. So, you can see here; please read s in place of p. So, I have here the local energy density of frequency at a given point in time.

And this definition is along the same lines as what we have in probability theory once again, where the conditional density, f (x) given y, is f (x comma y) divided by f (y); that is exactly the expression that I have here; in place of the probability densities, I have the energy densities. Now, typically there are no specific requirements on these densities themselves, but these densities are used in computing the conditional quantity such as local averages and local spreads.

(Refer Slide Time: 24:01)



Lecture 4.5 References

Conditional averages

The conditional averages of frequency and time are

$$\langle \omega \rangle_t = \frac{1}{S(t)} \int_{-\infty}^{\infty} \omega S(t, \omega)\, d\omega \qquad (6a)$$

$$\langle t \rangle_\omega = \frac{1}{S(\omega)} \int_{-\infty}^{\infty} t S(t, \omega)\, dt \qquad (6b)$$

Q: What values should these expressions yield?

- Average frequency at a time should yield the **instantaneous frequency**
- Average time at a frequency should yield the **group delay**

NPTEL

Arun K. Tangirala, IIT Madras TFA: Foundational Concepts 11

For example, I would like to know at a given time, what is a average frequency, right. If I pick a certain time t, I would like to know, what is the average frequency. Let me take you back to the example here, at a given time t; so let us pick this time here, 100, and I would like to know at this time what is the average frequency. Or, I pick another time here, 400; I would like to know, what is the average frequency? What is the use of this?

This is of great use in fall detection in feature extraction; I would like to know how the features of signal are changing. And if I have a certain template in mind for normal operating condition and I compare these features with the template then I will able to detect the presence of any anomalies or falls and so on. So, that is the local average of frequency at a given point in time.

Likewise, I can pick a frequency. So, I can pick for example, this frequency here, and ask what is the average time spend by this frequency, right. That is also important because I would like to know how long a frequency component persisted. It is very useful in biomedical signal analysis, process data analysis, all these features are very nice and they can be exploited in a very beneficial manner; that is exactly what we are evaluating here; the average frequency at a given time t and the average time spent by as a frequency component.

If the technique is right; that is, if your energy density is good and so on, and has been constructed an appropriate manner, then these expressions should yield what are known as the instantaneous frequencies and group delays that we have seen already; obviously, because the average frequency at a given time is instantaneous frequency itself. Of course, instantaneous frequency is actually the frequency at a given time t.

But, if I have a multi component signal that is if I have a many frequencies at a given time t, then the average frequency is nothing but the instantaneous frequency itself. If I have a single component at a given time t, it should actually yield instantaneous frequency itself; they should all match.

Likewise, the average time spent by a frequency component is what we had defined as group delay before if you recall the particular lecture, then you should able to be verify that the group delay is nothing but the average time spent by that particular frequency component. And again we expect that the particular method that I am using for time frequency analysis will give me these results correctly.

So, what I do is I construct the energy density, evaluate the right hand side; of course, replace the integrals with summations because I would not be able to compute the energy density over a continuum. So, I compute the right hand side and then I also compute the instantaneous frequency independently, and see if they both tally. So, that is how I would do it. Of course, the way instantaneous frequencies are computed with joint energy densities are slightly different, particularly for spectrogram and scalogram, using what are known as ridges and so on. We will talk about it later.