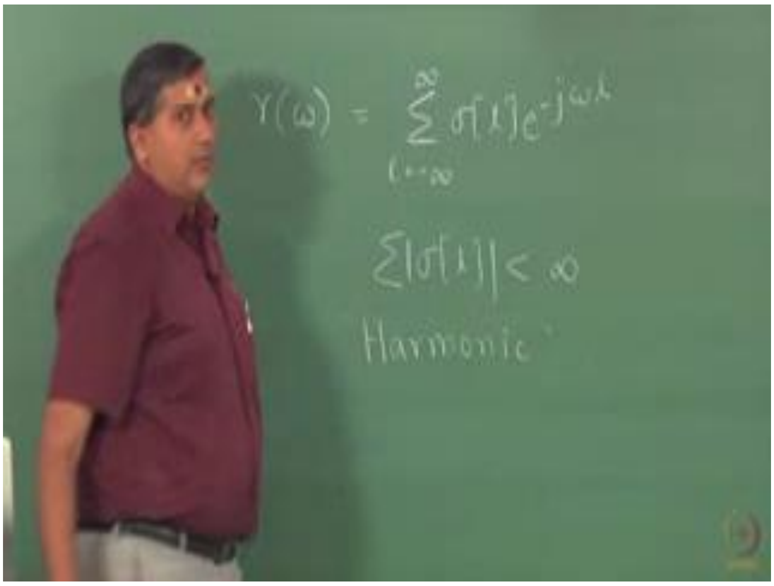
# Applied Time-Series Analysis

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

## Lecture – 38
## Lecture 17A - Models for Linear Stationary Processes-2

I will review very briefly the LTI systems theories concept that is sufficient to understand what is happening at a later time, but one of the things that I mentioned yesterday and also before is that this representation of a stationary process as a white noise passing through a filter is possible only if this spectral density of the stationary process satisfies certain conditions and what we mean by this spectral density satisfying certain conditions the full details we will state later on, but if you recall in fact. I had mentioned this yesterday as well the spectral density, which we denote by gamma you recall is the fourier transform of the auto co variance function, the discrete time fourier transform to be more specific and omega is the angular frequency that we are looking at.

(Refer Slide Time: 00:59)




This is not the definition of spectral density; please remember that many textbooks present the spectral density to be defined as the fourier transform of the auto co variance function this is not the definition, but this is the relation named after Venere and Kinton that is used in practice for computing spectral density and we will as I said we will go this definition much more in detail later on, but the point to note at this stage of discussion is that the spectral density because we are saying that the stationary process should possess a spectral density that satisfies certain conditions, but for that to happen first of all for that process this notion of spectral density should exist should be valid and if it is valid only when the auto co variance function is absolutely summable that we have already stated before.

We have said first of all the spectral density itself exists only when the auto co variance is absolutely convergent. This straightaway tells us that this concept of representing a stationary process as a white noise passing through linear filter is valid only for those stationary processes for which the auto co variance function is absolutely convergent. So, this is a an unwritten statement now, then only we can talk about the conditions that a spectral density should satisfy, there are some additional mild conditions that gamma should satisfy, but the foremost requirement is for the gamma to exist and that happens only if the ACVF is absolutely convergent.

Now, what this means intuitively is that the auto covariance should have certain characteristics. And we know that when a sequence is absolutely convergent, it has a certain characteristic, what is that? Any sequence not necessarily the auto co variance sequence, for any sequence when we say it is absolutely convergent to give you a hint at large in this case at large values of l, decays right. So, straightaway we can say that this notion of white noise passing through a linear filter can be applied to those stationary processes whose auto co variance is decay with lag l right.

What does it mean when we say auto co variance decays at large lags, what does it tell us about the correlation structure of the process? Remember the auto co variance is measuring the correlation between any two observations separated by lag l. And when we require that the auto co variance is absolutely convergent and which means that it should decay at large lags, what are we trying to state with regards to the correlations structure?

Well auto regression is just one class, we when we say decay it can also mean that the auto co variance has gone to zero already in finite lags, those sequences are also absolutely convergent right both the class of sequences that vanish after finite lags and those that decay at large lags both are admissible, but what does it tell us in general about the correlation structure? It tells us that the correlation if I take two observations that are spaced long time apart, there should be no memory there should be no influence of the present on a long time into the future, at least the effect should be diminishing, if not zero the effect of the present in the future should be diminishing and that is a very important requirement. Are they stationary processes that they do not satisfy this requirement?

Are there do you can you think of a stationary process it is still stationary, but the correlation does not die down at all, we have not discussed such processes, but I have mentioned briefly.

The class of harmonic processes which are essentially the periodic counterparts right the periodic counterparts of the deterministic periodic signals, we said that there are period periodic signals in the deterministic world and the periodic signals in the stochastic world. The periodic the periodicity concept in the deterministic world is fairly well understood may be we are fairly familiar with it, but the periodicity concept in the stochastic world is not so well understood, but I have said this before that a stochastic signal is said to be periodic if the auto co variance is periodic right that is one of the ways of defining a harmonic process or a periodic stochastic processes.

Now, when the auto co variance is periodic, naturally it does not meet this requirement right; obviously, because it exists for ever and no way it can be absolutely convergent, but a harmonic process can be stationary. So, which means there are this class of stationary processes and harmonic processes are quite frequently encountered, it is not a very special pathological situation and so on, it is not that we do encounter them frequently. So that means, that there do exists a class of stationary processes, that cannot be brought into this umbrella of white noise passing through a filter they cannot be brought under this umbrella. So, you have to be clear in your mind when you sit down to fit a time series model when the moment you say time series model, you are more or less implying that this kind of a model, that is a white noise passing through a filter unless you state explicitly something else yes.

Student: (Refer Time: 07:35).

Correct.

Student: so (Refer Time: 07:37).

First of all it is not at all stationary. So, we do not even discuss that here.

Student: (Refer Time: 07:43).

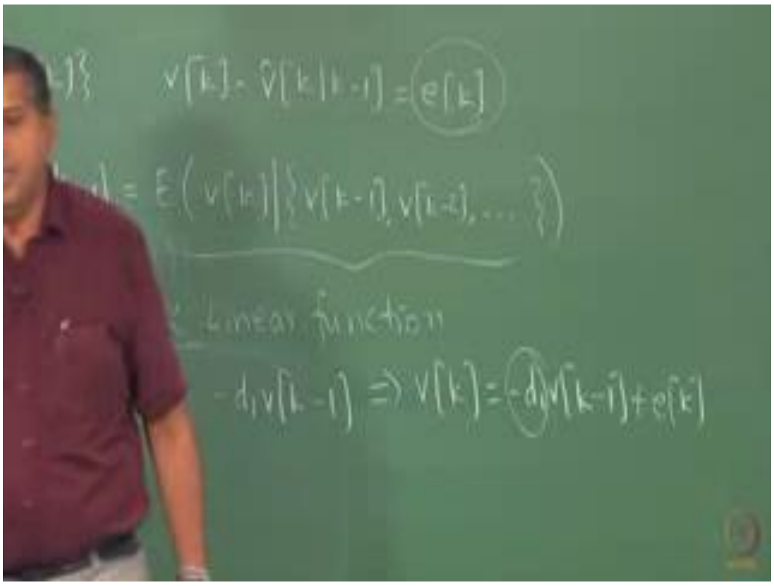
No because auto correlation that we are referring to is a theoretical one and yesterday as I said you cannot use the theoretical definition of the auto co variance that we have been working with for stationary processes, you can apply the definition of ACVF for non stationary processes. So, by the way for the audience there who are not able to hear this question; question was with reference to the signal that we talked about a periodic signal embedded in noise, does that fall into this class of processes? But and the answer is no because straightaway you can see this signal is non stationary, so this discussion eve does not even apply; well then the question is if the auto co variance is periodic, well the auto co variance that we have been working with is largely for stationary processes you cannot apply that definition, you have to go back to the non stationary one and then try to apply, but that is all a bit messy, that is a different discussion here we are talking of theoretical ACVFs.

So, we are now very clear this idea of representing a stationary process as white nose passing through a filter is applicable only of course stationary processes, but within that a subclass which satisfies this and fortunately there are well the class of the number of processes that satisfy our the kind of processes the number of processes that we encounter in practice, do more or less meet this requirement it is not an issue of course, in practice what we need to ask is what happens if I know that the spectral density exists, but it does not satisfy those mild conditions and so on; as if said yesterday in the end whatever model that you build in practice is merely an approximation, you would never know even if you are given that this spectral density exists that is the auto co variance satisfies this, you may not know if the spectral density of the process satisfies those mild conditions that we have not stated or those conditions that we have not stated.

So, you know you just assume that they satisfy and go ahead and build a time series model. What theory says is if it does not satisfy then you would be building only an approximation and that approximation gets better and better as you include more and more observations and you know more and as you increase the order of the model and so on, but we will talk about that again later on right now we are talking about theory. So, let us move on, you should keep this in mind that we shall assume hence forth, but the process satisfies this condition of absolute convergence of ACVF stationarity is always understood.

So, now we will talk about an important connection between models and predictors. As you already know many times I have stated that the time series modeling literature itself has grown out of this interest in forecasting a random process, that has been the primary objective since day one therefore, one should expect a strong connection between the model and the predictor and after all what why am I building a model for forecasting. So, there should be a connection as well and we have talked about this we have said that in fact, when we talked about when we introduced the class of auto regressive models or in fact, particularly when we talked about the role of white noise.

(Refer Slide Time: 11:23)

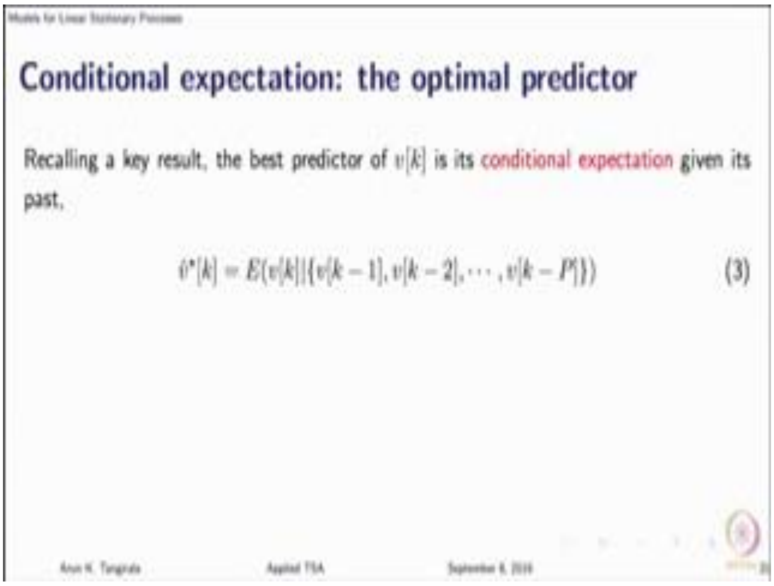



We went through this discussion, we said that the goal in prediction is to develop a predictor for this random process and we said that usually one works with a step ahead prediction, such that the residual has white noise characteristics, right we assume this v hat is optimal in that sense that whatever is the difference between the truth and the prediction that would turn out to be a having a white noise kind of characteristics.

And from here we develop this notion or this perceptive of this e k being the input as well right not only playing the role of white noise or this ideal residual, but also as an input if you recall and we will go through that more in detail now and we will see that repeatedly white noise appears as an input either in the form to the moving average process or the auto regressive process and so on.

In fact, that interpretation of white noise being the input that we have also seen in the previous schematic is actually central to the definition of what is called as a linear random process. So, there is an additional restriction now that we are going to bring in; one restriction is stationarity, the other restriction is that the auto co variance should be absolutely convergent, now the third restriction is the linearity of this class or subclass of stationary processes.

(Refer Slide Time: 13:06)




So, again what you will see on the screen is what we have already discussed. So, I am going to go fast and rip and then basically emphasized that now this e k will act as the input to the white noise process. So, to re quickly recap the discussion we had the ideal prediction that one can construct is the conditional expectation. Now we will slowly march towards linearity, the ideal prediction as we know is the conditional expectation of v k given it is past, so given v k minus 1, v k minus 2 and so on.

Here it since it is theory we assume we have the information from the infinite past this is our ideal predictor or optimal predictor and as we know to compute this conditional expectation, we need to know the joint period apart from that we do know that this conditional expectation is in general are non-linear function of the past or of the variables on which your condition if you recall.

(Refer Slide Time: 14:11)


And I have I am looking at expectation of Y given X; we know that this is nothing but some non-linear function of X because we are averaging over the outcome space of Y.

Now, do we want to work with this non-linear function? One limitation is I do not know the joint periods that itself actually limits the use of this resulted practice, the other aspect of this result is that this is the non-linear function and typically we do not want to work with non-linear functions straightaway because estimation becomes difficult, the interpretations become difficult and you know implementation becomes difficult online implementation there are so many demerits of working with a non-linear model at least in the first place.

If the process calls for such no problem we will consider that, but maybe many processes can be brought under this linear into this linear framework therefore, we want to replace this conditional expectation with a linear model that is the goal now. So, see that you should see that we are approaching this linear framework only for the sake of convenience, not because we believe that processes are linear, always even if you go to the deterministic world this requirement of linearity is largely motivated by a mathematical convenience than anything else because no realistic process can be thought of as linear.

So, the there are two points when we work with a linear model; first point is; obviously I am sacrificing some optimality when I approximate this with some linear function, that is the first point that is I am sacrificing the optimality, but the good news as we have also seen earlier many times is that this linear function gives me the same prediction as a conditional expectation, if these observations follow a joint Gaussian distribution.

As going back to the world of random variables in general g of X is a non liner function of X; however, if Y and X are jointly Gaussian distributed, g of X is a linear function of X. So, assuming now that the series or this process has a joint Gaussian distribution, we can straightaway postulate a linear predictor and start working with that, that is the idea what if the process does not fall out of a joint Gaussian does not follow joint Gaussian distribution; then there are two things that are going to happen one; obviously, you what will happen is your prediction will fall short of the optimal one, but how much it will fallen short of depends on the extent of deviation from joint Guassianity

And that is hard to really quantify; if you will have to look at the distribution and then talk about the extent of deviation of violation of joint Guassianity, what this tells us is now another requirement because we are going to work with linear models, it is more or less necessary to test upfront if this whether the series is falling out of a joint Gaussian distribution, how do we do that? So, it is look at the work the restrictions you know first we said stationarity, this is the big world right if you say this is the world of stationarity then within this we said we will auto co variance should be absolutely convergent.

So, which means you know we have now a restricted space to work with, what else was the restriction? Sorry linear or something else.

Student: (Refer Time: 18:19).

So, one is stationarity then you have. So, let me maybe some of you are unable to see I will draw it again here. So, this is the world of stationarity this big circle. So, let us denotes this by S and then we have then we have this restriction that the auto co variance should be absolutely convergent then only the spectral density can exists, further we said the spectral density should satisfy some conditions then there is a slightly smaller space to work with and now further we are saying that I would like to work with linear predictors or linear models more or less say they are almost one and the same.

Now; that means, I have a restricted space and the restricted space is this. So, this innermost circle contains processes that are stationary, whose auto co variance function is absolutely convergent and a spectral density satisfies certain mild a certain conditions and the joint distribution is Gaussian. So, there are the series of restrictions that one has to impose even before we think of hitting this ARMA model. The question of course that arises in our minds is in practice does it become too restricted? Are we really restricting ourselves to a very nice, cozy set of processes and that is a comfort zone that we are developing; well it is restricted there is no doubt about it, but it is not highly restrictable in the sense that it can be that is this idea of ARMA model, ARMA modeling can be applied to a large class of processes, but there exists a larger class of processes which do not fall within this framework.

So, when you are working with a series that has fallen out of a non Gaussian distribution then perhaps it is coming from this area, assuming that the remaining three are satisfied. So, here is stationarity, here you have the absolute convergence criterion and then your spectral density satisfying some conditions. So, if the spectral density satisfying conditions are denoted by SD here and then you have the joint Gaussian distribution.

So, if your process actually is coming out of this shaded area or the series is falling out of this shaded a process belonging to this shaded area then an and you are still working at a linear model, then yes you will be making some optimal predictions. So, how do I check upfront that the series has fallen out of a Gaussian joint Gaussian distribution? There are tests for joint Guassianity one can actually cons conduct test for joint Guassianity, normally one does not do that, but strictly speaking you should do it at least visually you should plot the histogram of the data that you have right the histogram in some sense gives you an idea of the joint p d f; in r you have the heist command which allows you to plot the histogram in the form of a density function and you should be at least visually convinced that yes this data does fall out of a joint Gaussian distribution, it is very hard it is very very hard because what you are going to work with is a realization.

So, it is a very crude thing that you are testing for, but it is much better than not doing anything. It is not a correct representative of the joint Guassianity; what it tells you is whether these variables are falling out of a Gaussian distribution that is all it is not a correct test of joint Guassianity, if you are truly worried about joint Guassianity you should use a test that are available for joint Guassianity. But anyway most of the series that we will work with we will assume that they satisfy joint Guassianity, but you should keep this at the back of your mind and if something goes wrong and if your predictions are fallen way short of the reality then you may want to come back and question this assumption of joint Guassianity.

So, this framework is essentially telling us where to search when things go wrong; when we may things go wrong when the predictions go wrong, ultimately how would I know if things have gone wrong? I do not know the truth I do not know the true model. So, the only way I can assess the goodness of my model is through predictions. So, if my predictions are fallen short then one or more of these assumptions have been violated right. So, this is why it is important to discuss the theoretical framework.

Anyway coming back to the point here we will work linear functions assuming joint Guassianity and it is with that idea that we went through this example of an auto regressive kind of model. So, one of the possibilities is that the prediction is a linear function of it is immediate past only of course, example that we took up the other day was where the predictor was a linear function of past two observations, but suppose this is the case then we have straightaway from this equation and this, this is the optimal one now that we are proposing, we will throw away the star it is understood that it is sorry we are always working with optimal predictors, so that I can write this evolution model for v k; this is only an example do not think that all processes have this kind of an evolution equation. So, let me actually indicate that clearly.

We have then what is known as an auto regressive model of the first order. Now you can extend this idea to an auto regressive model of p th order. Basically you assume that a prediction is a linear function of the past p observations, so that you see the model you get the model that you see on the screen, where you have now a summation of the essential linear combination of the past p observations. So, of course, some minor things we have used a negative sign and I have said this before the reason for using this negative sign is; so that when I write a transfer function operator all the coefficients have a positive they are all represented as additions that is the only reason.