CH5230: System Identification

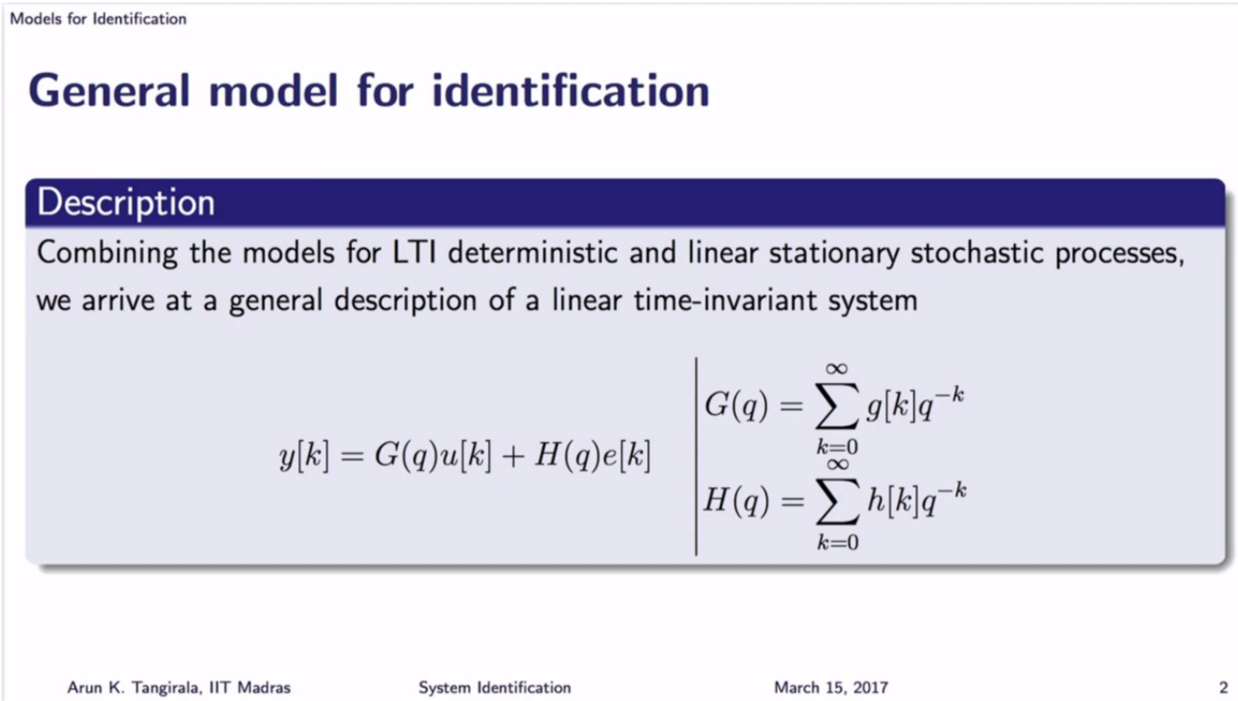
Models for Identification 1

Okay. A very good morning.Yeah, yesterday's lecture must have been quite heavy for you but it was a review. Please go through the notes. I have posted the notes on the course website, yesterday night. Primarily we went through a frequency domain analysis. What is formerly known as spectral representations and whether you have understood anything or not, the main message that you should get is that for random signals although Fourier transforms do not exist. We can still think of a power spectral density and that power spectral density is usually computed from the covariance function. Both the auto and the cross spectral density functions. And why did we have to go through that? Again that's because now we are going to fuse both the deterministic and stochastic world. Now we have at least a clear understanding of how the deterministic system is described in time and frequency domain and also the stochastic process.

So what we are going to do today is almost now we are crossing the bridge from theory to practice. So here, your journey begins. Yeah, slowly moving from theory to practice today onwards. Today we look at the different models that arise as a result of marrying the deterministic and stochastic descriptions. And then what implications does each model structure have. All right? And in that process we'll also look at a concept of quasi-stationarity, it strikes a middle ground between the deterministic and stochastic world. And then if time permits I'll briefly talk about prediction, else we'll talk about it tomorrow.

So let's get going. So here is now the composite model. Here is a composite model for y which has two effects, the effects of input and of course the stochastic effects. One thing that I did not tell you yesterday is under what conditions you can actually express the stochastic signal as a white noise passing through a filter. Right?

(Refer Slide Time: 2:27)



The conditions are actually about four or five but I'll tell you the most important ones. First a spectral density should exist. That is a condition for this stochastic process. Secondly it should be factorizable. What we mean by factorizable is? If you recall from yesterday's discussion we had gamma v v of omega as mod H to the minus G omega square times gamma e e of omega. Right? Now we can rewrite this as H of e to the minus j omega times gamma e e of omega times H star that is the conjugate of the FRF of the noise model. Right?

There is a reason why we have written this way. In fact you should see a very striking similarity of this expression with the one that we had written earlier for the auto-covariance generating function. So we said the auto-covariance generating function can be shown to be related to the model. In what way? What is it that we had? H of z inverse times sigma square e times H of z. Right?So you see a very strong parallel here and it should not come as a surprise in light of what we have learned yesterday that covariance functions and spectral densities are related. Correct.

In fact, what is this g sigma z, how is it defined as? This is defined as the z-transform of the auto-covariance function. Right? The bilateral run. So sometimes this is also called the spectral density because the spectral density is after all can be obtained for a class or stationary processes whose ACVF is absolutely convergent as the Fourier transform of sigma L. Okay?

But there are stationary processes whose ACVF is not absolutely convergent. Then what do you do?Does the spectral density exist? No, it doesn't exist. But this always exists. As long as the process is stationary you can always think of the auto-covariance generating function. Now let's get back to the main point here. You can write a time-series model of the formwhite noise passing through a filter for the stationary process provided, number one, the spectral density should exist.Their spectral entity has certain properties it should be non-negative at all omega. Number one. Secondly it should be symmetric. These are all qualities of a density function.

So the spectral density should have the qualities of a density function. In addition it should exist. Now it should exist would mean the auto-covariance function should be absolutely convergent. Now the third requirement is that it should be factorizable. Okay? That is when you can think, the conditions are not yet complete but this is a third condition that the spectral density should be factorizable in this way. Now do I know if it is factorizable. I wouldn't know in practice but at least it is telling you under these conditions you can write a time-series model at least. Sorry, what I mean by time-series model iswhite noise passing through a filter.

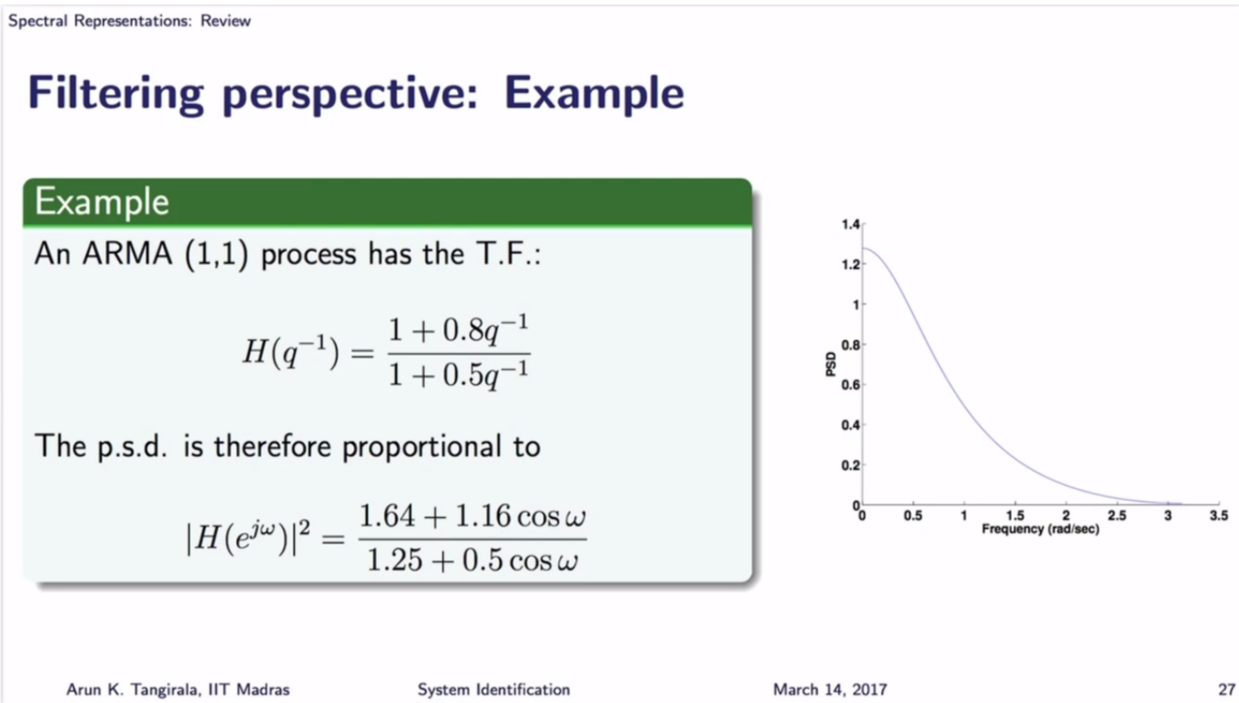
Fourth condition is, when can I develop an ARMA model? Remember, the general time-series model if you recall simply expresses H has a convolution form. Right?Then the non-parametric form.Up to the non-parametricform you have three conditions, a spectral density should exist. And then it should be factorizable, of course, the zeroth condition is that spectral density should have the qualities of a density function. There is another condition which I have not stated which has called the Wiener-Paley condition for you to construct a causal representation.

Remember we are going to represent the random process as a causal combination of shock waves which means we will not consider future shock waves that condition at the moment I'm going to keep it aside. You can listen to the lectures on time series analysis and then figured out what needs to be done. Now, what is a condition there?Now, coming to the last point. When can I develop an ARMA model? Which means, I am asking the question, when can I develop a parametric model for H. Correct? So I have here. H in a general sense, H is non-parametric.

Up to that the conditions are spectral density should exist and it should be factorizable. And then there is another condition called the Wiener-Paley condition which will tell you basically whether you can write a causal form or not. In addition, if I am going to demand that a parametric form can be constructed. Then the spectral density should satisfy an additional condition. What is that, when H had a parametric form?

We already learned that gamma. That is a spectral density proportional to the squared magnitude of H of e to the j omega. So which means, what is going to be the spectral density today for this process? Whatever is given here time sigma square e by 2 pi. All right?

(Refer Slide Time: 08:18)



So which means, gamma, the spectral density for a parametric model has a certain form and that form is a rational form. In general spectral density can be any mathematical function of omega. So long as it satisfies the basic requirements. It should be non-negative. It should be symmetric and it area should give, area under the spectral density should give you the variance. All those requirements are there. But there is no restriction on the shape. On the mathematical structure of gamma of omega. What function it is of omega? Nobody says anything. It should just satisfy certain conditions.

But now if want and if I desire a parametric model should exist then an additional requirement should also exist. It's like imposing certain conditions on the impulse response coefficients. Right? Not all non-parametric models will necessarily lead to a parametric form. Only when the impulse response or the frequency response or one of the responses has a certain structure, I can derive a parametric form. So here, also in time-series modeling only when the spectral density is a rational polynomial of trigonometric functions. Which kind of trigonometric functions? Cosines only. Can I have Sine appearing in the numerator and denominator for a spectral density? Yes or no? What do you think?

The spectral density has to satisfy an important property which is that, it should be symmetric. So if I have a sinusoid there then it doesn't satisfy. Naturally therefore, the spectral density is an even function and it therefore only even functions can appear in the numerator and denominator. Specifically what kind of even functions should appear? The Cosines. We know Cosine is an even function. Correct? So to summarize, when can Irepresent a stationary process as white noise passing through a filter. Well, first the spectral density should exist with all the standard properties. It should be non-negative,it should be symmetric. It's area I should give me variants.

Secondly it should be factorizable in that form. Thirdly the Wiener-Paley condition which I have not stated here should be satisfied for us to be able to write a causal model. This is up to when I don't worry about whether parametric model or a non-parametricexit. In addition if I want to write and ARMA form then the spectral entity should necessarily have or look like, I have this functional form where it is a rational polynomial of trigonometric functions specifically Cosines. It cannot be any trigonometric functions. Under those conditions I can develop a time-series model.

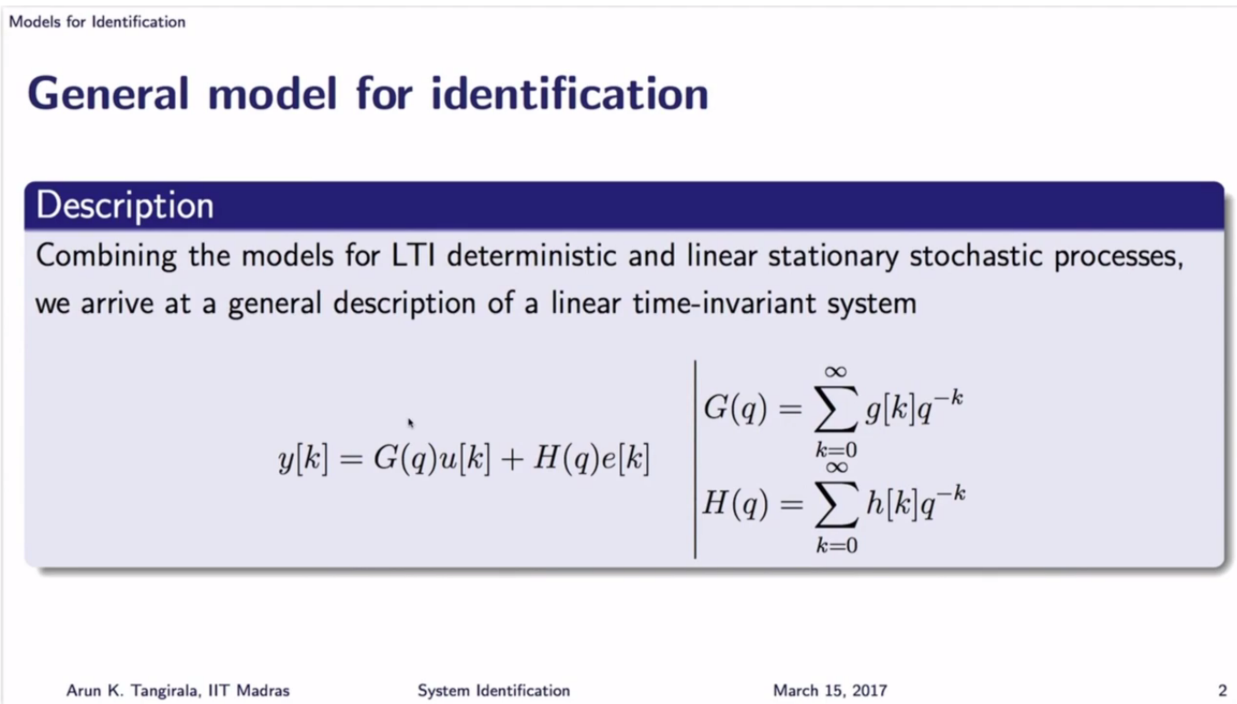
No one point to take home and an extremely important point which I have mentioned earlier is I can only identify the time-series modelcorrectly up to a phase. The other way of stating this, I will not be able to resolve the delay. Correct? And you can now see that either through this equation or this equation. Finally, if you look at the way the models are estimated. The time-series model is estimated, they're all estimated from the auto-covariance function or the spectral density. One of those.

Now the question is, whether the auto-covariance function or the spectral density is sensitive to delays in H. That means, if I replace H with some other H which only differs by some phase. In other words here, suppose I multiply H with I have a new H with some other phase e power j Phi, then the conjugate also is altered by e to the minus j Phi. And that cancels out. So which means, there is always this ambiguity in H of each of the jPhi, that is up to a phase. And that can never get resolved. Again if you ask why, because the input is unknown. If I know the input then this issue doesn't arise. We never spoke about this in the deterministic world. Because they knew the input. Everything was fixed by the input side.

Here the input is fictitious, onlyits statistical properties are fixed. And therefore, I will never be able to resolve the phase factor. In other words I man would be able to resolve the delay. So that is something to keep in mind. So now you know under at least I have given you the list of conditions under which you can write or you can express or represent the stationary process as white noise passing through a linear filter. And the conditions are straight away on the spectral density. So this together whatever I have given you in the form of condition is known as a spectral factorization result which is the milestone result in linear random processes.

Only after that result came about. People said, okay. Now I know the theoretical framework. I have a very strong framework and I have the support to build a time-series model, linear time-series models of the form that we build today. Okay. So that is a theoretical support. So let's get back to the discussion here. Now here is a composite model G and H are in general non-parametric but you can choose to parameterize.

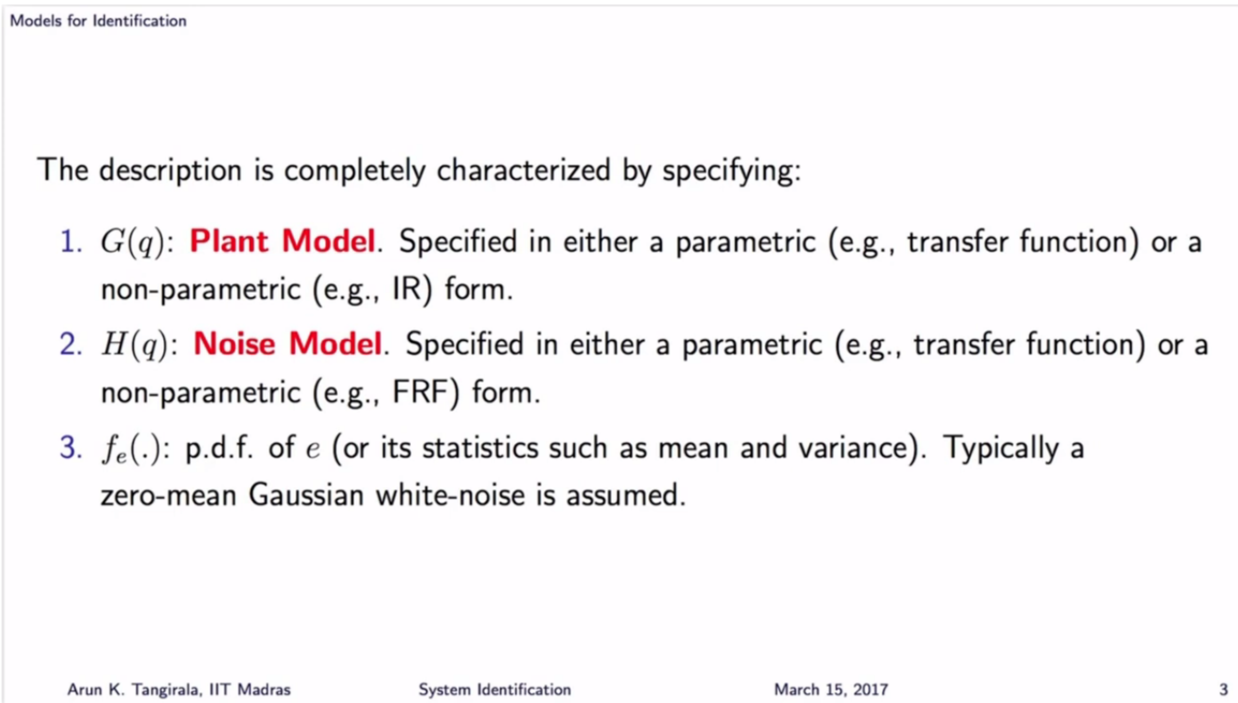
(Refer Slide Time: 14:12)



We'll study both those situations or scenarios. So there's some terminology we know in G of q is known as a Plant Model or the deterministic model. Again as I said it can be either in a parametric form or a non-parametric form andH is known as a noise model. It essentially, is a representative of the channel through which the uncertainties are entering by. Although, we keep using the word noise. I think it is better to use the word uncertainties. So v is, what you say consolidated effect of all the uncertainties in your system. At this moment we do not talk about what is a source of v. We know that the sources could be many, unmeasured disturbances,sensor noise,modeling errors,perhaps and so on.

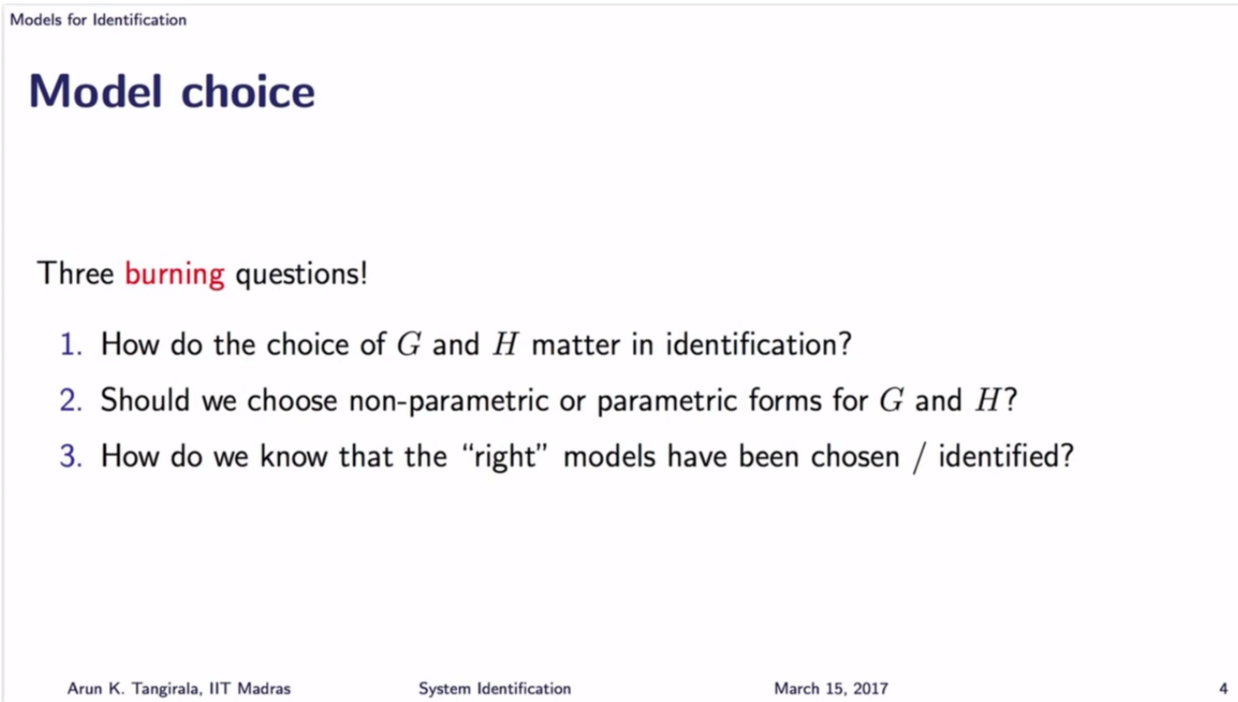
We don't really care at this point as long as v k stationary, that is all we need. And then the third one is of course the pdf of e k. Right? We assume the pdf typically to be Gaussian. Once you specify G H and the pdf of e k, give that e k is white noise. The description is complete.

(Refer Slide Time: 15:36)



There is nothing more to state about this LTI system. You have given away everything. Nothing more is required to construct the optimal prediction and so on of y. All right? So these are the three elements. Now in practice what is the identificationproblem. Therefore, given y and u data, I have to identify G, I have to identify H and I also have to identify the pdf. But we have already fixed the pdf to be joined Gaussian. Therefore, we don't have to do that. Instead of estimating the pdf now we estimate the parameters of the pdf. And since, it is white I and it is stationary. I only have to estimate the variance assuming it to be zero mean. If the e k doesn't have a zero mean we'll include that as an additional term in y or we will make sure that we do some pre-processing. We'll subtract the means from y and u accordingly and work with zero mean or mean center data. So that is identification for u. Now there are three burning questions when it comes to choosing the models.So slowly we are coming to practicality, fine. Now I know all the model structures, it is like I've gone to a shop the shopkeeper has, to buy a cellphone, the shopkeeper has shown me all the cell phone models. Right? Now the three burning questions is, how do I care? What G and H are? I mean does it matter really what model they choose for G and H in identification. It matters a lot. Particularly in your ability to explain a class of processes.

(Refer Slide Time 17:15)



If you choose a certain model for G and the certain model for H, then remember I told this a long ago, I gave you this analogy when we were discussing the liquid level case study. G and H are like your left and right hands. And you have to get a lot of work done. So how I choose my G and H. Translates to how I position my right and left hand. Right? Depending on that I get my work done. Certain choices of G and H may actually end up putting G and H together that you are parameterizing particularly or maybe fixing one hand and doing your work with only the other hand.

Obviously depending on how you position,your right and left and what freedom you have will determine the kind of work that you can do. So likewise here with G and H, the choice of G and H as you will see shortly will govern or determine your ability to explain a small class of processes or a large class of processes. That is point number one. Secondly, very importantly the choice of models in G and H have a large implication in the effort of estimation. Estimation effort, whether I can estimate the parameters easily or not because as you will see shortly, if I choose a particular model structure, the predictor terms are to be linear in the unknowns. If I choose some of that model structure the predictor turns out to be non-linear in unknowns. Accordingly always solving a linearly squares or a non-linearly squares. And we know the efforts are different in these two.

Thirdly,which is very important, it's got nothing to do with estimation effort. It has got to do now or ability to arrive at the so-called true model. Whether I can figure out what is a true G and what is a true H. That again depends on the choice of G and H.As I am speaking you should recall the liquid level case study that we went through. Of course we will go through it now in more in detail, we'll study a lot of interesting results but the answer to the first question is, rather thansaying how does a choice. Yes it does. So the second burning question is whether it should be to a parametric form or a non-parametric form.

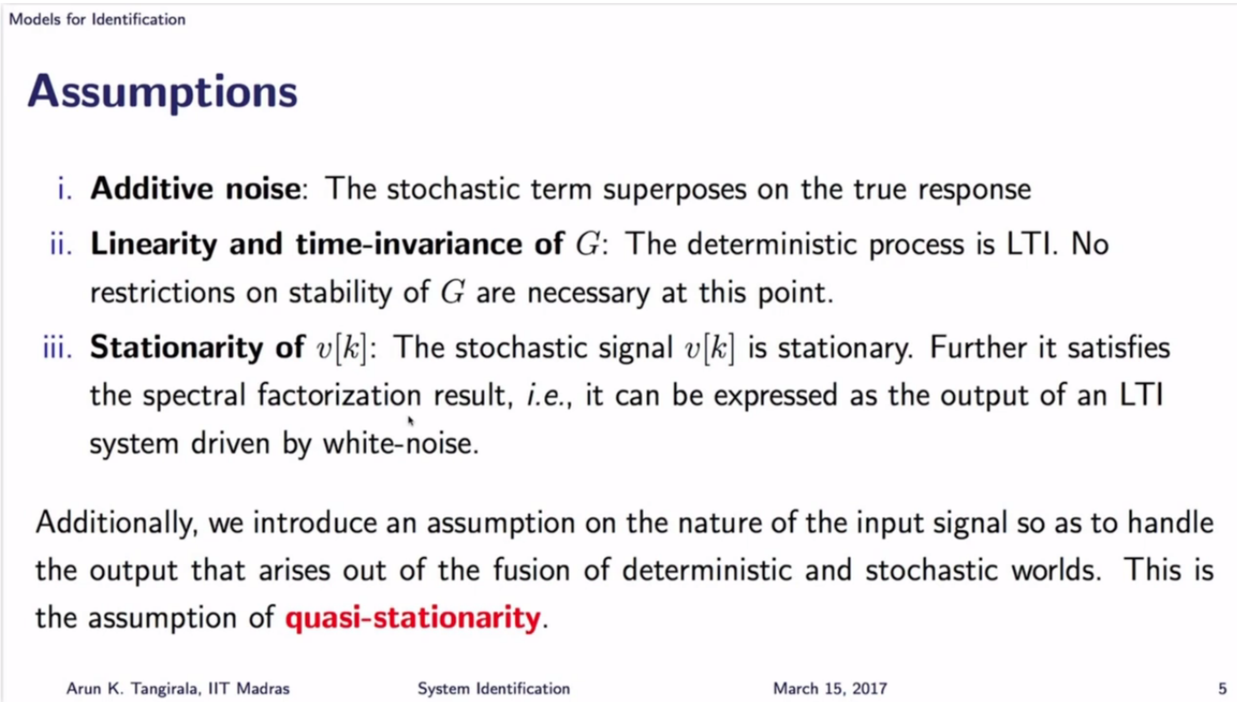
What is your suggestion on that, what are your thoughts on that? First question we have only answered partly in detail. We learn more details as we go along. We should choose the parametric form. Any other ideas, in other words should I choose a impulse response or transfer function form. What do you think? Transfer function form, we decided. Okay.
Anybody else? Correct. First we go fornon-parametric because if I have to choose. That's correct. And so eventually we'll head to transfer function. But there is an intermediate junction. That I have to get down from the train, in my journey to make a stopover.figured out now, how long my final destination looks like. Where do I want to go eventually?

So this fitting in non-parametric model like we did in the liquid level case study will give us some insights into what is a delay, what should be the numerator order, denominator order? Right? Whether a nice model should be fit and so on. So we'll get answers to a lot of questions that we need to answer when choosing with transfer function form. Okay. So that is a general strategy. And the third one is very important. How do we know we have the right model.Right? That we have learned in the liquid level case study. And even when we were going through the sample time-series model that we have to do a residual analysis. And also we have to do a test for over fitting.

So many things but we also need a theoretical backing not only will the residual analysis and the test for over fitting do the job. More importantly I need to know for a given system whether choosing G and H in a particular way will guarantee that I will get to the truth. We will learn later on that if I choose G and H in a particular way doesn't matter what the system is. There is no guarantee that they will get the true model even though I'll do theresidual analysis and so on. The residualanalysis and a test for over fitting will help me in getting a working model, whether that working model is in line with the true model. That is a generating process depends on how we choose G and H. So there are quite a few questions to be answered here. Okay. We will answer them as we go along.
Now let's get back and ask, what are the assumptions? I'm just listing some of the assumptions which you already know. Firstly, we have assumed additive nice. That is one. Secondly, we have assumed linearity

and time-variance of G. And on the noise part or the stochastic component we have assumed stationarity. So this is our framework. Right.

(Refer Slide Time: 22:50)



Now these three assumptions alone are not enough. Individually they have been assumed to be like this, that is individually G has been assumed to be LTI, individually v k has been assumed to be stationary. And then we have said, how they come together which is additive. They might have, they could have come together in a multiplicative also. But we not considered that option. Now that we have brought together these two systems in an additive manner, we need an additional condition to make sure things are consistent. What we mean meetings are consistent is, I'm going to estimate parameters. I'm going to build certain estimators where I will use some, let's say some expressions and so on for estimating the model parameters, estimating the variance of v k and so on.

Those expressions that I use, those estimators that I use may not necessarily work all the time. I have to actually impose an additional condition. For example, you look at v k, we have assumed to be stationary. Fine but when I look at y. So if I look at y k. y k's G of q u plus v. v is stationary, doesn't mean why is stationary. Why should I care whether the output is stationary? By the way now the output is a stochastic signal. It's no longer a deterministic signal. So I'll have to, I will turn to the auto-covariance or the cross-covariance and so on.All these tools to estimate the parameters of G and H.

But those expressions that I'm going to use should require that y k be not stationary. It should be stationary. But if you look at this y here, what is the expectation of y? One of the first tests that we conduct for stationarity is mean to check if mean changes with time or not. Correct. So what is the expectation of y? G of q u k, because that's deterministic plus expectation of v k. Typical we assume it to be zero mean. Now you tell me whether y is stationary or not. What do you think?What do you think? Is y stationary?Why is it so difficult? It isn't? Becauseit's a function time, even if you take the most simple form of G of q. Let's say an FIRmodel of order two. Even a static process, G of q is simply a constant.

Even then they input keeps changing with time. Which means the local mean of y is changing with time. Therefore,y k is not stationary. No that is not good news. However it turns out that this requirement of second order stationarity can be further relaxed. To what is known as quasi-stationarity. Okay? Which means that-- What is a relaxation that we have done while going from strict stationarity to second order stationarity. That only the first two moments be invariant with time and that they are finite. Correct? Mean variance should be finite. They should be invariant with time, auto-covariance function should only be a function of like.

This is what we meant by second order stationarity. The other woman can do whatever they want. They can go to the movie they can do whatever they want. We don't care really. Now we are going to relax that a bit further. By saying that, all right, the mean can change with time. No problem. Variants can change with time. No problem as long as they're bonded. I'm cutting known all the theory to show that this quasi-stationarity is sufficient but I'll discuss what is quasi-stationarity shortly. But the first thing message is that you should get across is. We can still relax a bit more that requirement of second order stationarity and. allow time varying means time waiting variances. Also auto-covariance can change with time but with some additional requirements. Okay? What are those requirements? And those requirements aren't summarized in these requirements of quasi-stationarity.