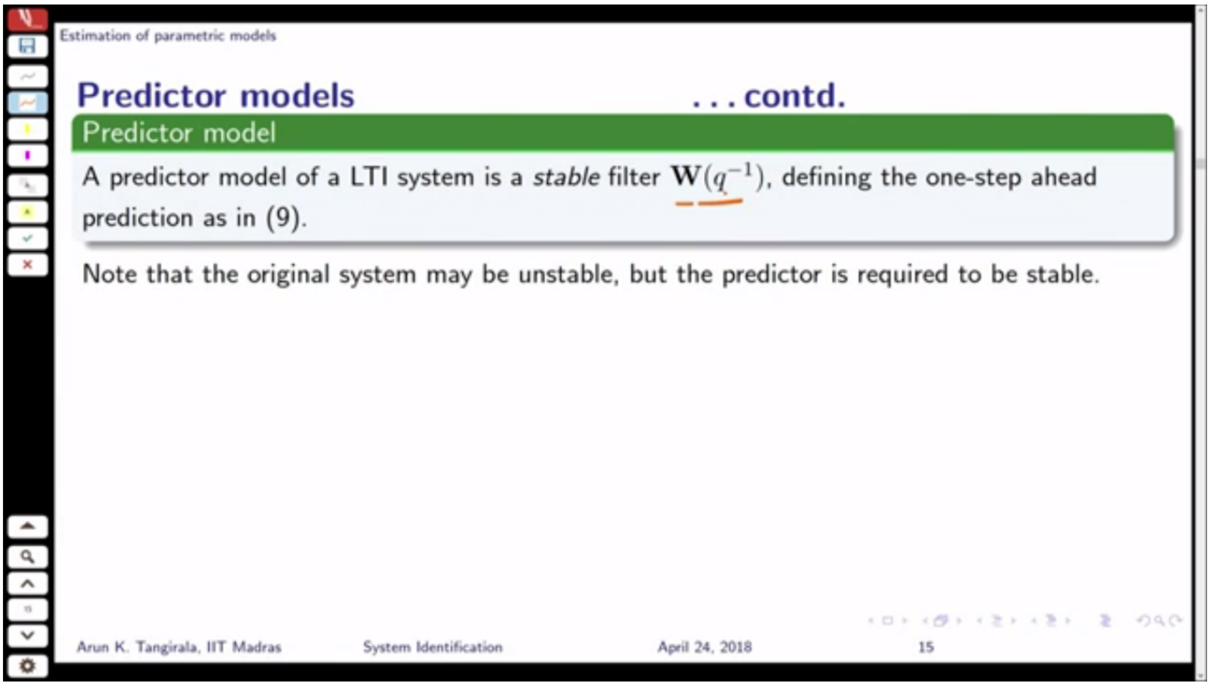
CH5230: System Identification

Estimation of parametric model

Part 2

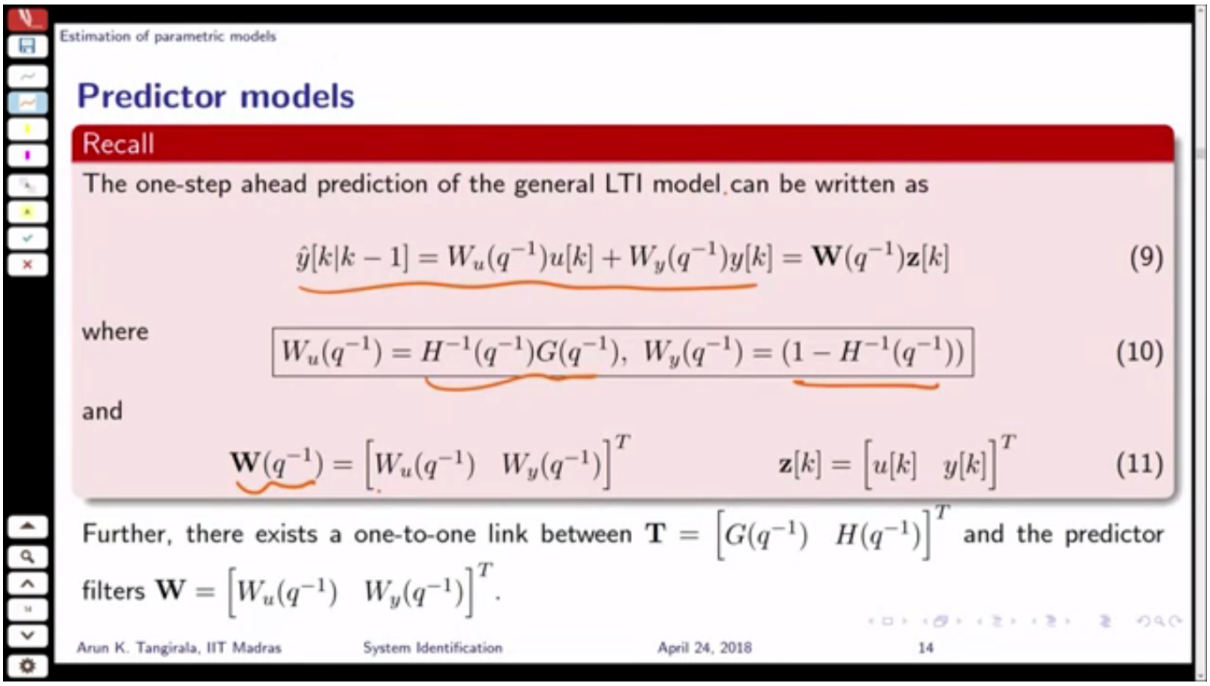
So the definition of a predictor model is that essentially this filter W of q inverse.

(Refer Slide Time: 00:21)



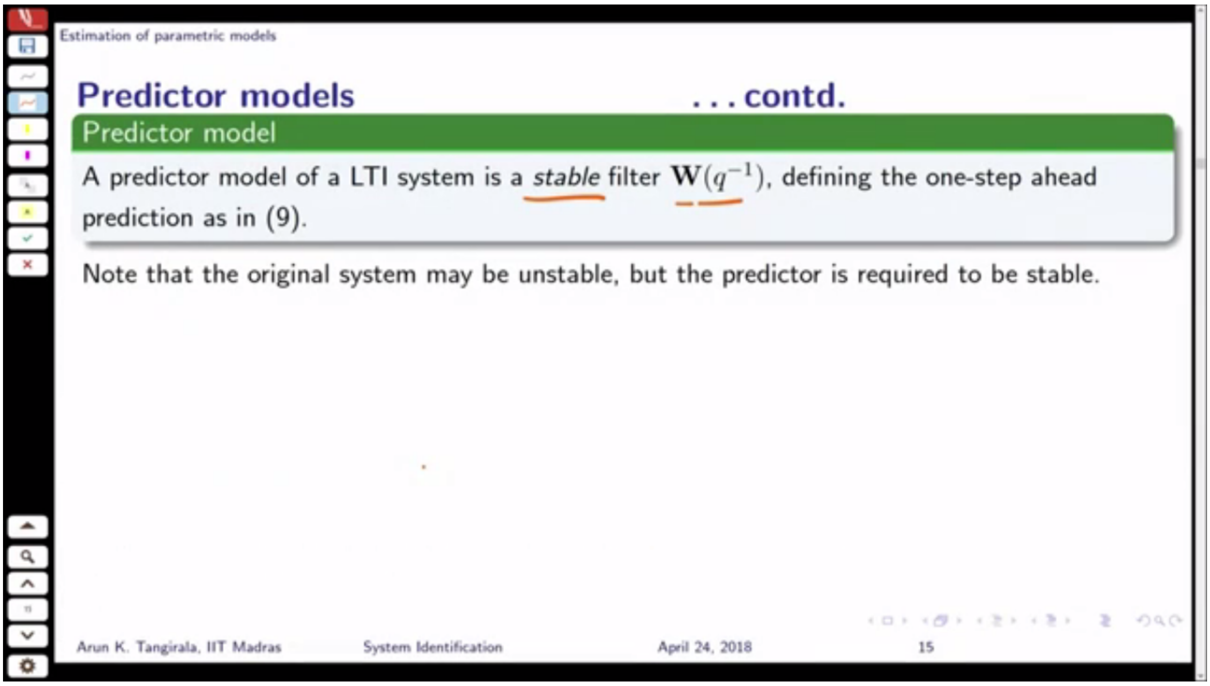
It's a set of two filters.

(Refer Slide Time: 00:23)



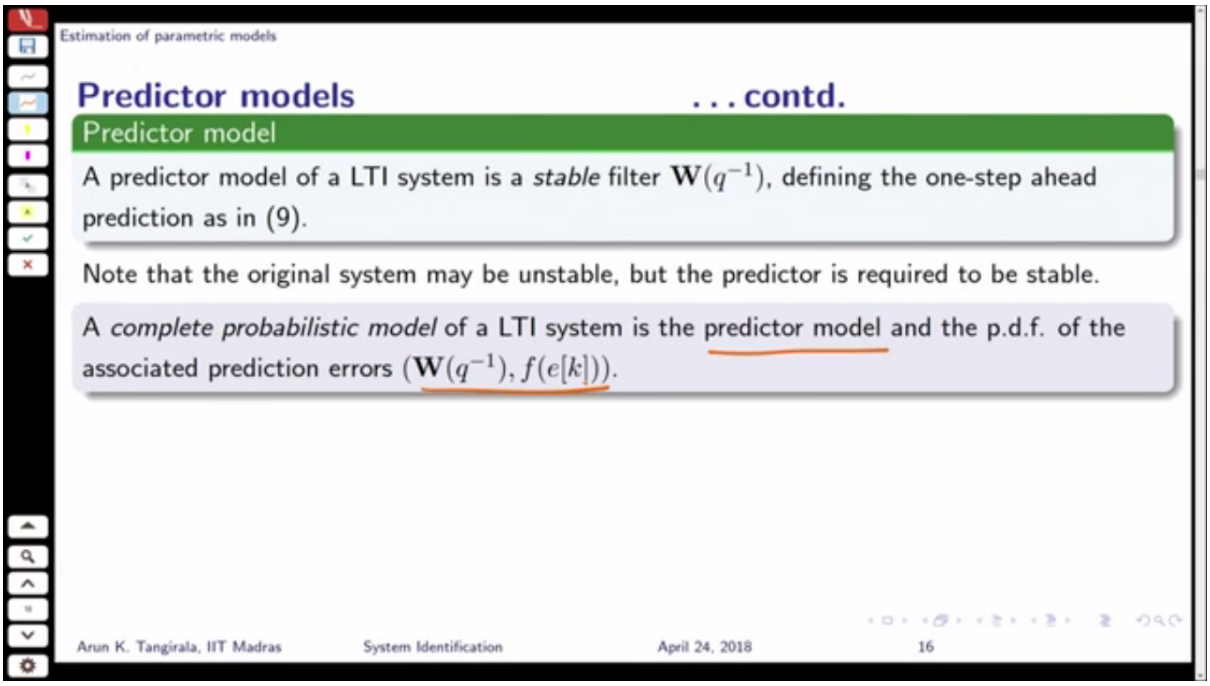
One that acts on the input the other that acts on the output. And what kind of filter should it be stable. Why are we imposing stability.Because I want the predictions to be stable. I don't want the predictions to run away. And it can happen that the original system may be unstable. But the predictor is still required to be stable. Okay.So there is that weird requirement in some cases it looks weird. But by and large we want the filters to be stable because also you're going to really work with gradients of this and so on. So there are other requirements that call for stability. Essentially the predictor model is nothing but two filters W u and W y, which are stable.

(Refer Slide Time: 01:21)



And you know what a stable filter does essentially the same theory that we learn in LTI systems. And the model is complete, when you specify the predictor model and the p.d.f of the associated prediction errors. That's all. So if you don't understand any of what I have just said, you can just think that this is an alternate way of describing your system.

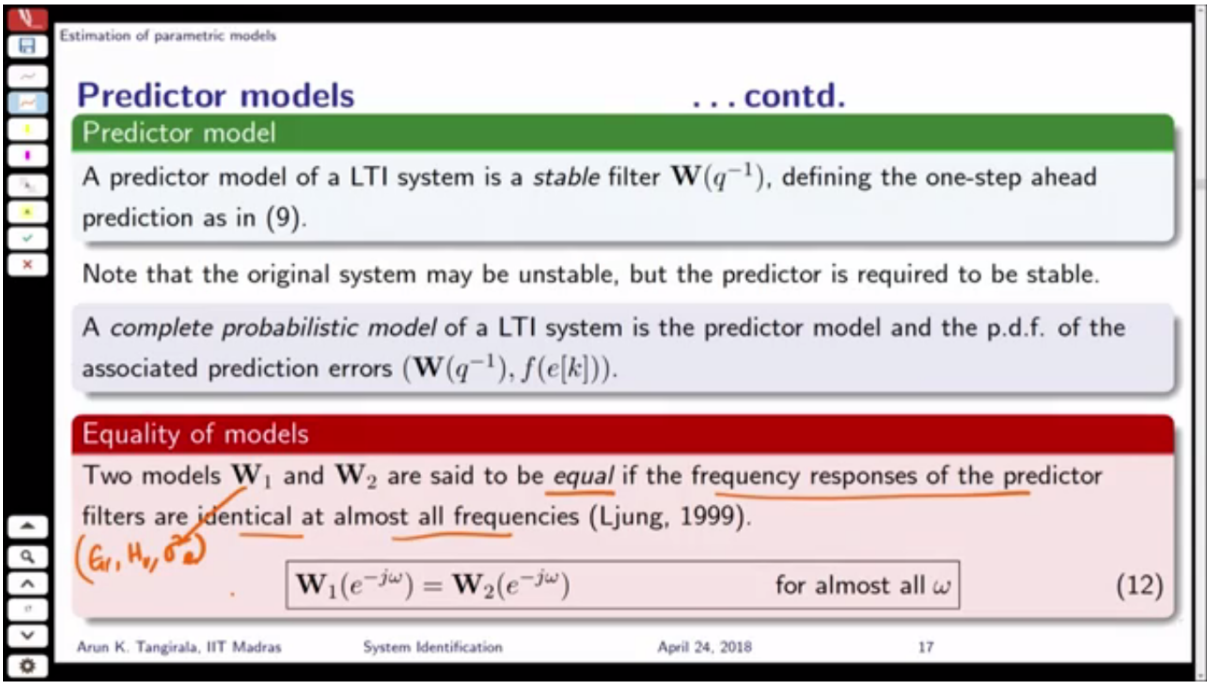
(Refer Slide Time: 01:45)



You could describe by means of G H and sigma square e, or alternatively you can describe in terms of W u W y and F of E. So it's one in the same. And as I had said earlier, these predictor filters become come very handy in talking of equality of models. An equality of models is a very central concept in deriving certain formal results of identification. One such result we have said is, identifiability an input design.

So two models, when we say two models, W1 and W2. You should read in the background some G1 H1 sigma square e 1. And G2 H2 sigma square e 2. Because associated with this W1there would be a G1 H1 and sigma square e. And likewise associated with W2 there would be a G2 H2 sigma square e. There they are set to be equal if the frequency responses of the predictor filters are identical at almost all frequencies. So some one or two exceptions are okay. They cannot be unequal equal over a band of frequencies. This is what sets the notion for equality and also the system belonging to M and so on. Whatever we will require later on. Okay.

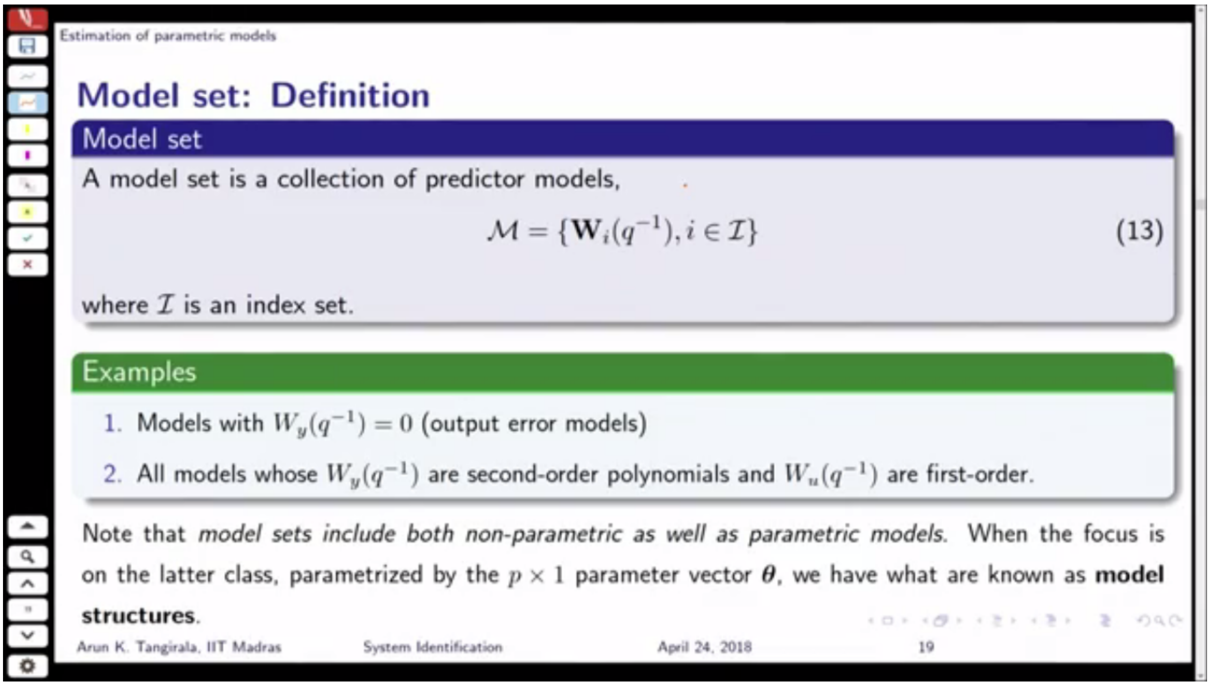
(Refer Slide Time: 03:11)



So I can be given two models, all I need to do is to check if they're equal. I compute the frequency responses of the predictor filters and if they're equal at almost all frequencies then we say two models are identical.

And now we come to the notion of a model set because this M is what that appeared in the consistency conditions for PEMestimates. Look at how beautifully all the formalism is set up. It was only with the help of this setup that you could talk of consistency of PEM estimates and so on. So now a model set is essentially a collection of predictor models. You should notice that model set includes both parametric and non-parametric. This predictor models like your G and H could be in parametric form or non-parametric form. Okay. Therefore this model set essentially is a collection of all the models that you were having by over some index I what you have of the predictor models. All right.So, crudely speaking it's a collection of different class, different models which yield different predictions perhaps they don't have to be equal.

(Refer Slide Time: 04:41)



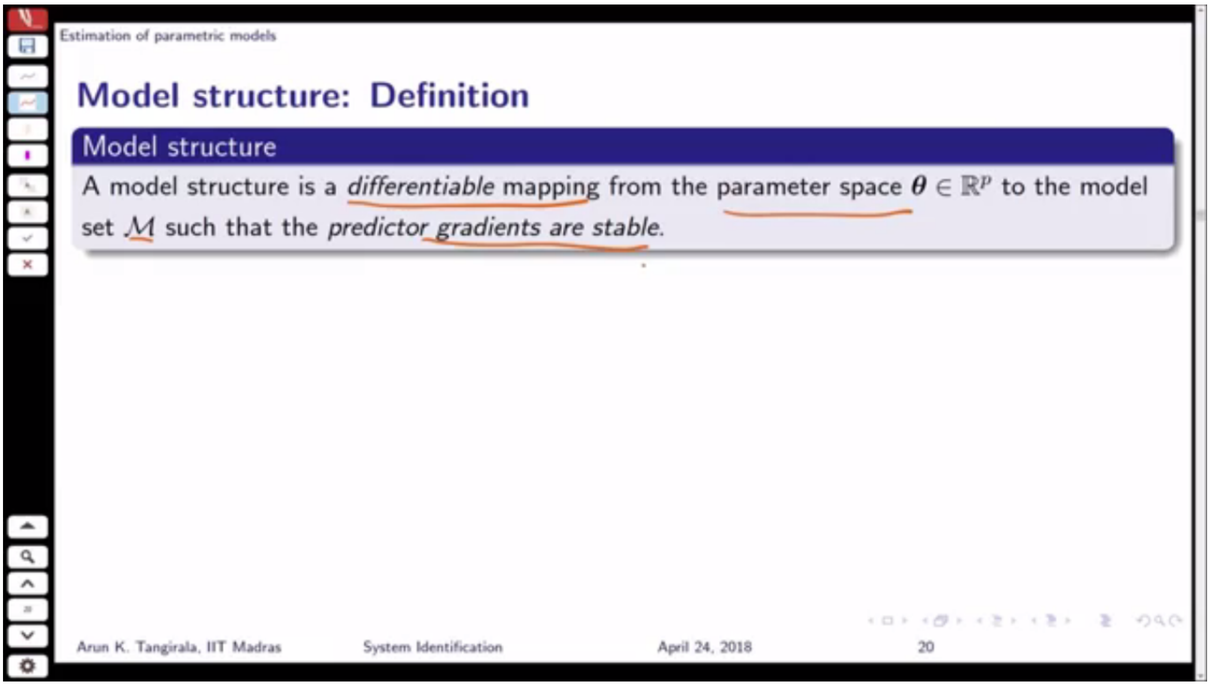
For example, a model set is all those models with W y of of q inverse equals0. So, remember your output error model, what is the prediction error expression for output error model. Y hat of k givenk minus 1, for the output error model, what is this expression?

Correct. G of q inverse u k.That means in this case W y 0and Wu is simply G of q inverse. So your model set is, all those models that you generate by choosing different G. Okay. Or you can say my model set is essentially all consists of all those models who'sW y second order polynomials and Wu are first order. You can do any kind of restricted model set collection.And as I mentioned earlier, model sets includes non-parametric as well as parametric models.

When the focus is only on parametric models, that means when you start parameterizing your models then you introduce the name model structure. Why do we use a different term now? Why are we using structure against the set? What this statement says is, the moment you start parameterizing your model, we are going to call now, this model set as a model structure or structures. Why is that? That's all. That's a very simple because parameterization gives you simple structures to those models.Right.Whereas non-parametric models do not have specifically any structure.

So what is the model structure?Again, formally it's a differentiable mapping from the parameter space theta to the model set M. Such that the predictor gradients are stable. So essentially what happens is, yes.

(Refer Slide Time: 06:51)

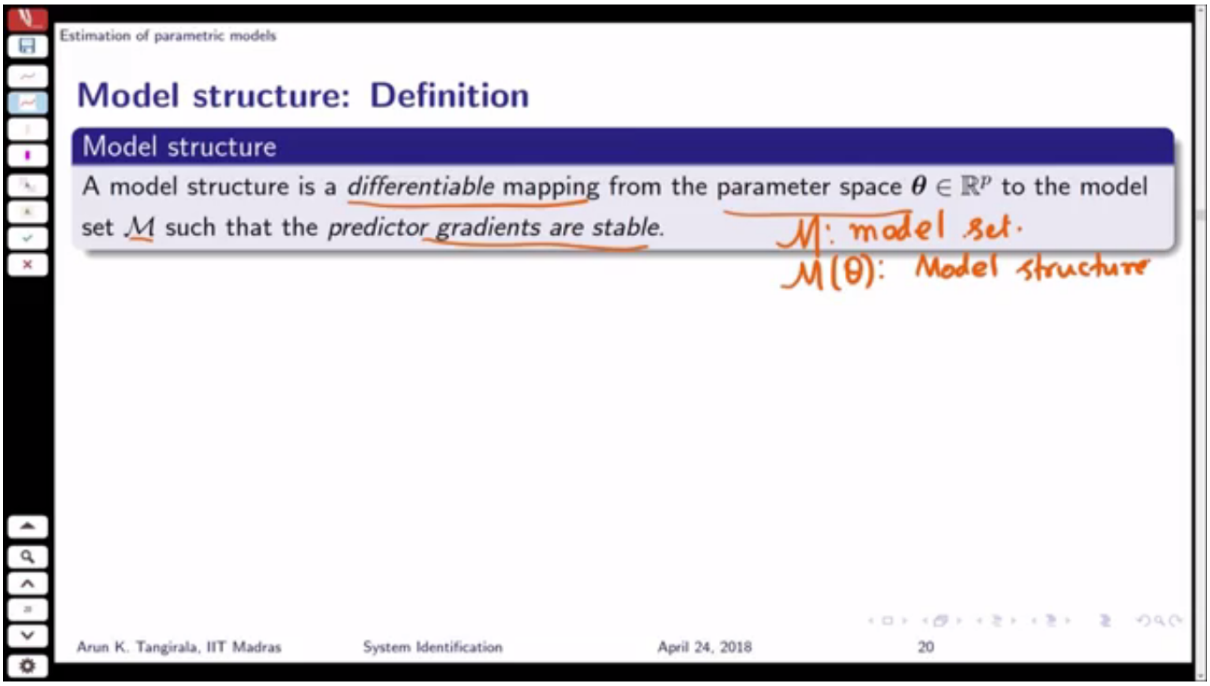


It can be anything. It can be anything. As long its stable it's okay. Correct. Linear systems can have any form of impulse response coefficients or frequency response functions. They don't have to be rational. You remember, even in time series modelling the story is the same. We say that there is a spectrum as long as a spectral density exists. You can fit the linear time series model. But if you're fitting an ARMA model then you are assuming that is spectral density has rational form. Likewise here, LTI doesn't mean, you should remember, LTI doesn't mean that G and H have a specific structure. They just have to obey, possess the properties of linearity and time invariance.

If there is a specific structure, then you would like to exploit that. And that is what you are trying to do in parametric modelling. But remember that your original model system can be LTI but not have any structure. In which case if you are fitting a parametric model it'll only be an approximation.

So earlier we dealt with M, which is a model set. Now we are dealing with m of theta. Now M is a specifically a function of some parameter theta. Now there is a mapping, so which means that for every choice of theta thereexists a model in your model structure. That is what we mean by mapping. The differentiable mapping is to ensure there is continuity. That means as I change the parameters,there should be a model here. It should not be that your model structure does not contain a particular member that you generate by choosing a particular theta.

(Refer Slide Time: 08:50)



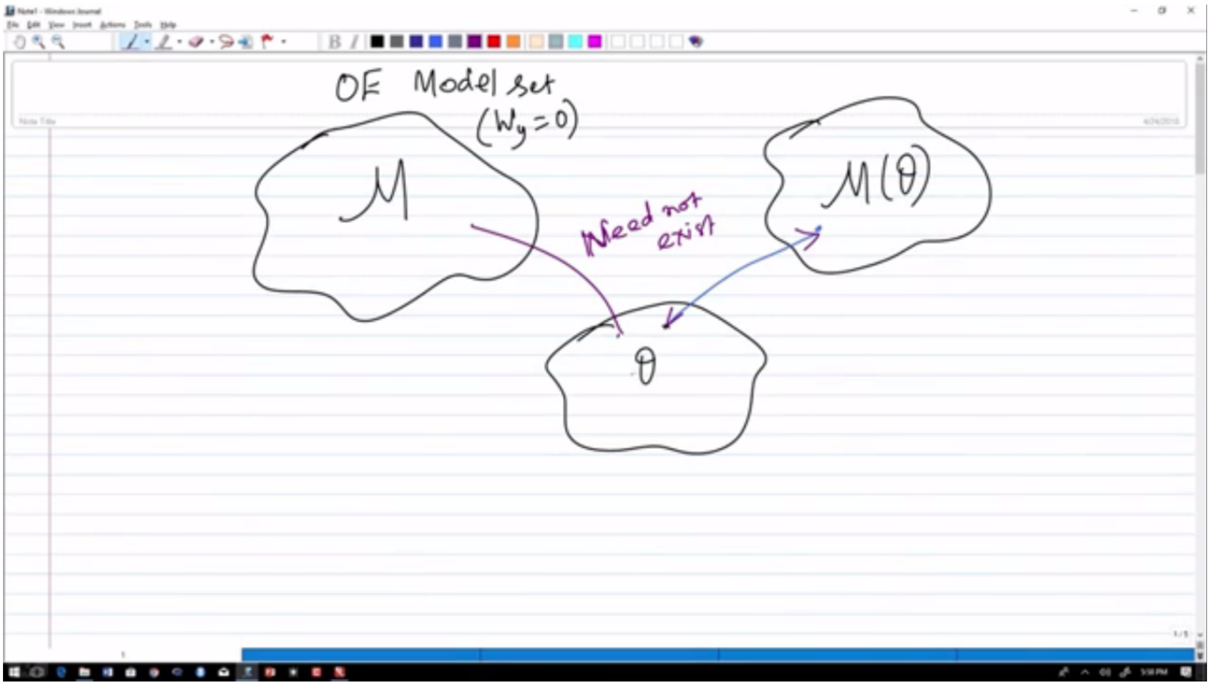
For every choice of theta there should exist a model in your model structure. A model structure is also set. A set this broader because it may not correspond to a particular theta.So you understand right. Let me draw that here.

So, this is your M, let us say. And this M contains all kinds of models, it could be parametric, non-parametric and so on. You can say for example. Let M be all those models with W y 0. Which means you're looking at output error modelset. So when you say output error model set it doesn't mean that necessarily G is parameterized. You should also remember that. So as an example M could be simply an output error model set. Which means that W y is 0.

Now specifically when I choose to parameterizeWu, which is nothing by G. Then takes birth an M theta. And what is this M of theta? M of theta essentially, there is a space of theta here for every point in theta there exists a point here. Okay. So there is a mapping.

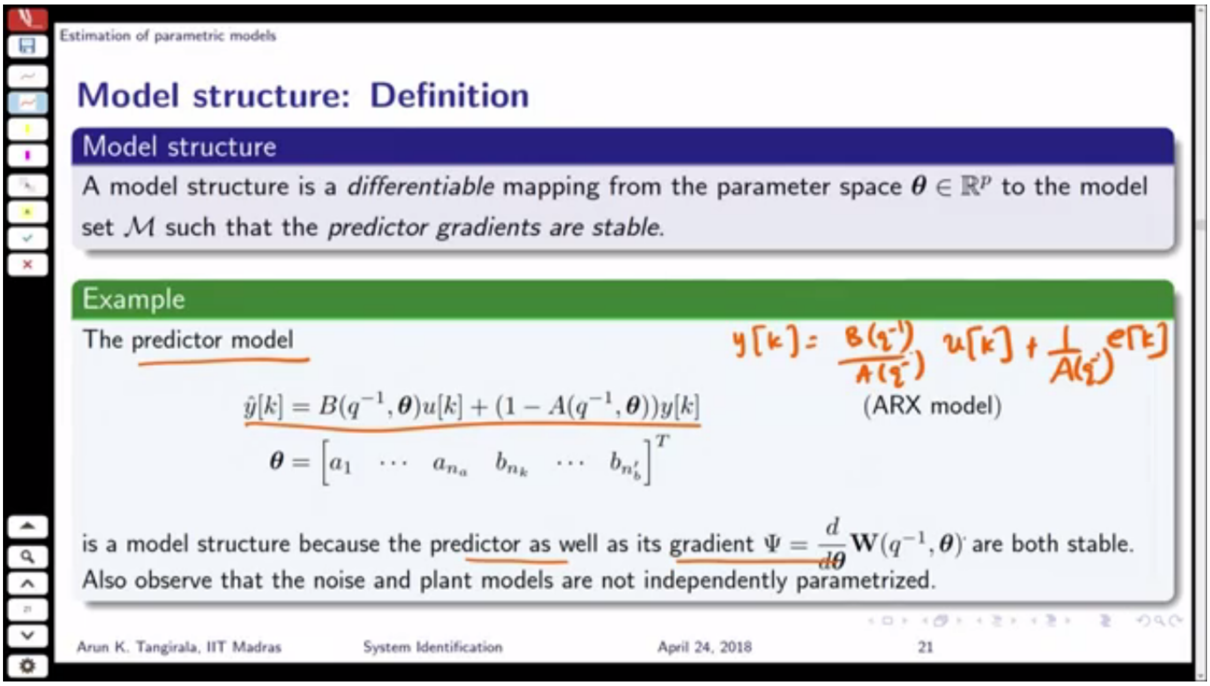
But for every point in theta here there need not exist, this need not exist. Yes.Correct. Subset of M, so that means, you know, what this means is that you may not be able to find them theta backwards. So one way of mapping may exist. Okay. For every point in theta you may be able to find here some model but not every point in M will have a point in theta. Yeah. That's a good point. So it's not a bijective mapping. So you may be able to find a theta. I mean, for one choice of theta,you may be able to find M here but not for every choice of M,you will find the theta.Whereas here this is bi-directional. Okay. So that is what is the difference between the model set and the model structure.

(Refer Slide Time: 11:28)



Correct. So now as an example, we have the predictor models here, y hat of k. We know, what this predictor model is. Or look at the word that I've used. I said the predictor model, I've given the expression only for the predictor. From this you should be able to guess that actually I'm referring to an ARXS model. Normally when I write an ARXS model, I would write y k equals B over A, U plus 1 over A, e, right. This is what I would write. But instead of writing, that's why, I have given the prediction expression .Of course is understood, it's a one step ahead prediction. It is a model structure because the predictor as well as its gradient at both stable. That is also important because the gradient of the predictors play a critical role in your parameter estimation. Remember this is what we work with in our optimization.

(Refer Slide Time: 12:37)



But you should also observe that the noise and plantmodels have not independently parameterized. We know that anyway. So we'll talk about that a bit later.

So now let's come to the notion of this true system and model. We know why we assume a true system because obviously we want to make some qualifying statements of the above the goodness of the estimator. So, we have this true system being denoted with a G0 and H0 and E0. And remember that this G0 and H0 are not necessarilyparameterizedforms. They're essentially LTI. That's all is the assumption. And E0 is a stationary sequence. Or you can say stationary white noise signal. Whereas the model that I'm looking at is a parameterizedone.

Now, how do I say that S belongs to M. Right.Because S seems to be superior.S can be outside the purview of M. Why? Because you may not be able to find the theta that exactly finds G0. So, if I substitute the value of theta here in G, it should get me Go.But that is where we are talking of equality of models and we have defined fortunately what is equality of forms. What do we mean by equality of models?Essentially, when we say S belongs toM. This is what we want. We could have even stated in terms of predictor models but essentially that is what it is.These two statements stem from the equality of models.What is the equality of models say, that two models are said to be equal if the frequency responses of the predictor filters are identical. So which means if I were to denote W S as the predictor filter for the system and W M as a predictor filter for this model, then what do I mean by S belonging to M of theta, essentially it means that W S of E to the J omega. This W is a vector. It consists of Wy Wu, should equal W u of E to the J omega.Remember W M is a function of theta. Because S is not parameterized, W M is parameterized. And when you when you equate them, you will have, remember W S contains W y and W u, W M also contains W yWu, when you equate everything then these are the conditions that followed. Essentially this system is said to belong to M, simply if the frequency response of the system can be generated their existing theta, that is what it says.

If you can find one value of theta for your model structure such that, if I plug in that value of theta parameters into the model structure, it would generate the true frequency response. If I'm not able to do that then we say that the system is outside the purview of M. That's all. We are not comparing structures. That is what you should understand. S may or may not have a structure. G0 and H0,may or may not have a structure. Whereas I know G and H have a structure because I'm saying clearly it's a parametric model. So obviously, I cannot sit down and compare structures but what I can definitely compare is the frequency responses.

So I'm bringing them onto some platform where the comparison is possible and that comparison is based on the frequency response. You could do this even in terms of impulse response. But then you will have to do it at all times. Okay.Because impulse responses are also in general in non-parametric forms. So that is the idea. You should remember now, when we say that a system belongs to model means, model structure means essentially that their exists one value of parameter theta. They may exist multiple, we don't know but they should exist at least one. That will generate the frequency response of the plant and noise models. Yes,correct. That we will. But first of all they should. Okay. Now, I'm sorry.

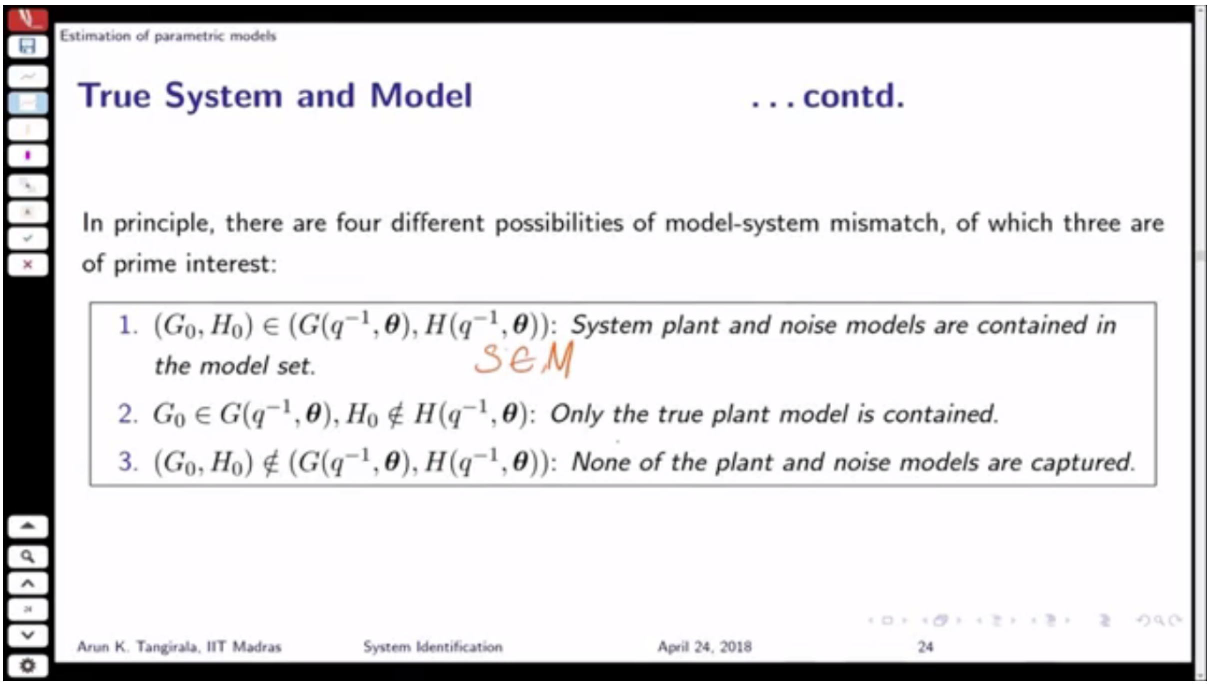
What happens is your model structure. Yes. For identifiability there should exist only one value of theta. But the system belonging to M is not worried about identifiability for now. When we talk of consistency, then they will talk of a fixed value and so on. But to begin with essentially S should belong to, even you say S belongs to M, there should be one, at least one theta. Okay.

All right. So let's move on and now talk of this true system and model and the different possibilities that may arise. So one possibility is that both G0 and H0 belong to G and H, which is the case we just described. So we say that S belongs to M. This is a case that S belongs to M. The other possibility is that, I have captured the true system in my true plantin my plant model set. But I made a mistake in the noise model. So the possibilities that we are discussing is that, my net was big enough to capture the G0 and H0. They say other possibilities, second possibility is that, oh, I did a good job of guessing the plant model structure but I did not do a good job of the noise modelling.

For example, in OE, suppose the original process is ARXS. The data generating process is ARXS and I am fitting in aOE model. Then which scenario does this belong to at least possibly. Definitely not one. Because I am assuming an OE that H is 1. Whereas a data generating process is 1 over A. There is no way I can actually explain that frequency response function using my all OE model.So the possibilities perhaps two. Perhaps because that depends on now what I have assumed for my G.

The third possibilities is I made a mistake in both. Of course you can ask. What about the fourth possibility? Why didn't we consider the fourth possibility where G0 does not belong to G.And H not belongs to H. That is usually an art of so much interest in identification. Right. We are not so much worried about that because the focus is on G, all the time. So we want to talk of whether we have made a mistake in guessing getting G0 or not. And the result depends on whether you have guessed the Hrightly or not and parameterization, how you parameterized. So that is why we don't talk about the fourth possibility otherwise the fourth possibility also can be stated in principal.

(Refer Slide Time: 20:32)

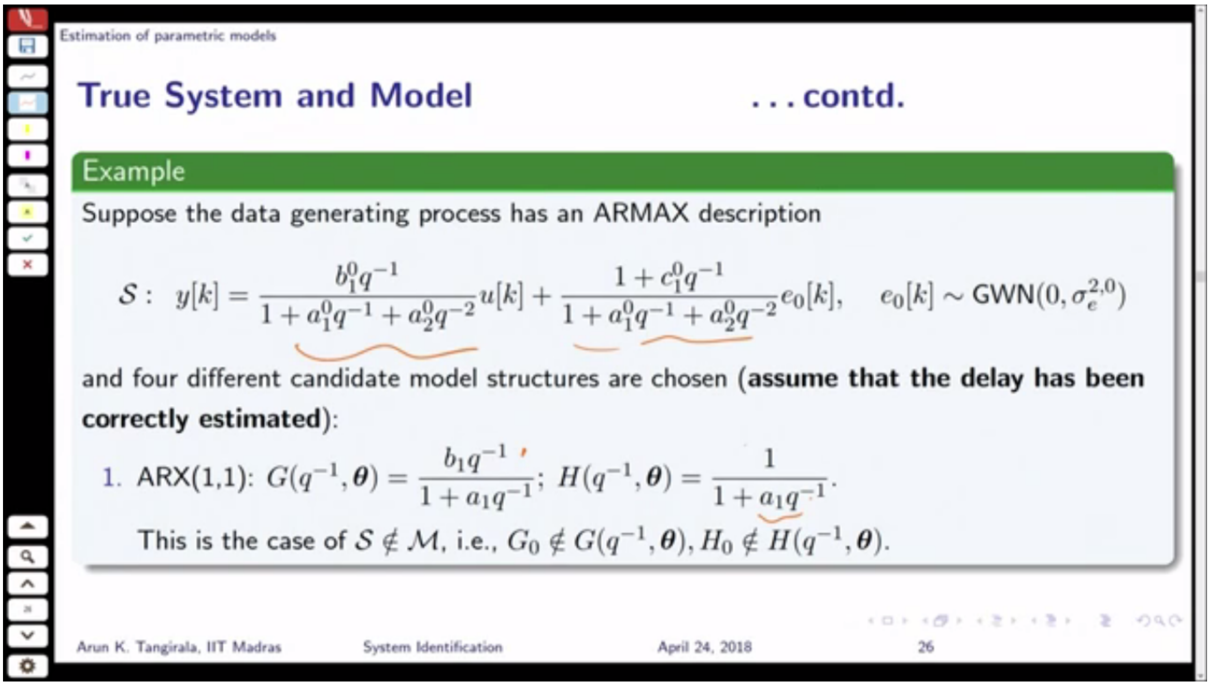


So these are the different possibilities and consistency of the PEMestimator essentially depends on these possibilities. So let's first understand this three different possibilities. And then we will go back to the consistency results.What do the results say. So, suppose a data generating process has a ARMAX description. And we consider four different candidate model structures. Okay.Then the purpose of this exercise is for us to train ourselves in answering whether S belongs to M or which scenario are we dealing with.

Let us assume that the delay has been correctly estimated. That is another assumption that we shall make. What is the delay in this system.Right. Okay.So let's assume that has been stated. Now first, let us assume that the first model that I'm going to fit is this. It's an ARXS model. First order in G and naturally first order in H. There is no freedom in specifying H. This is clearly the case of S not belonging to M. Why?

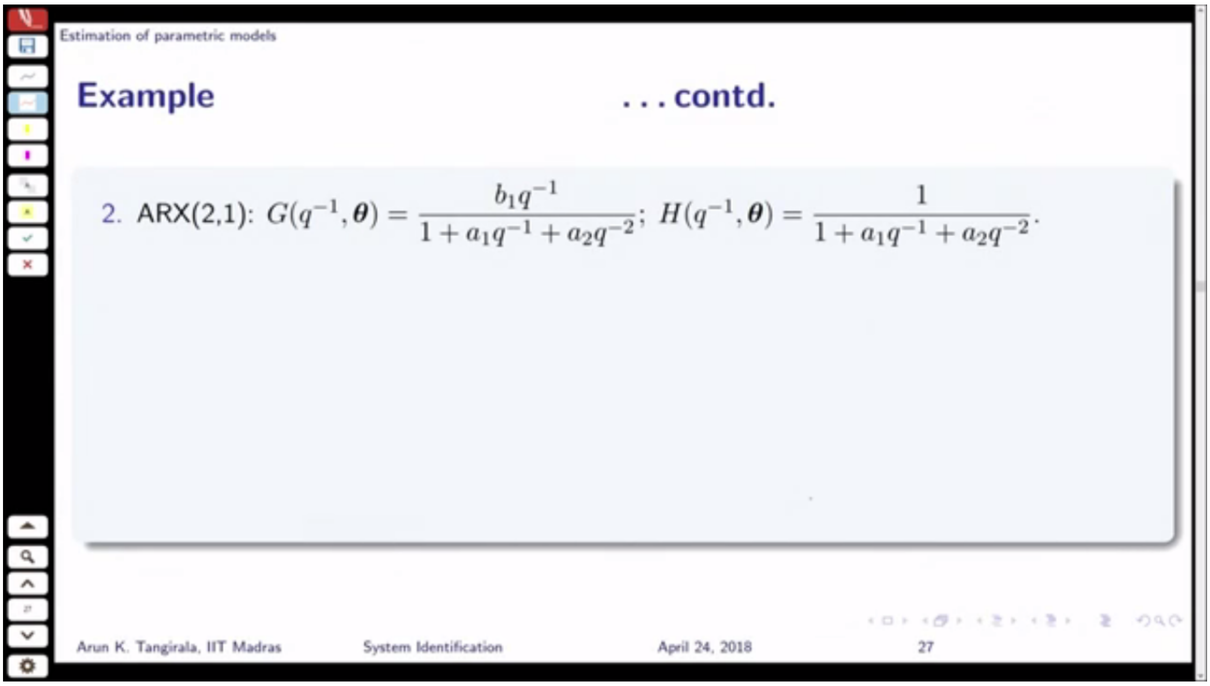
Because. I cannot explain G0 using this. Right. There is no way, I will be able to explainthis system using this transfer function, unless a 2 not(0) is 0. Likewise, I cannot explain this system using this model. So G0and H0 do not belong to the models set or model structure. In this case we say S does not belong to M, fully.

(Refer Slide Time: 22:10)



What about this case.

(Refer Slide Time: 22:13)

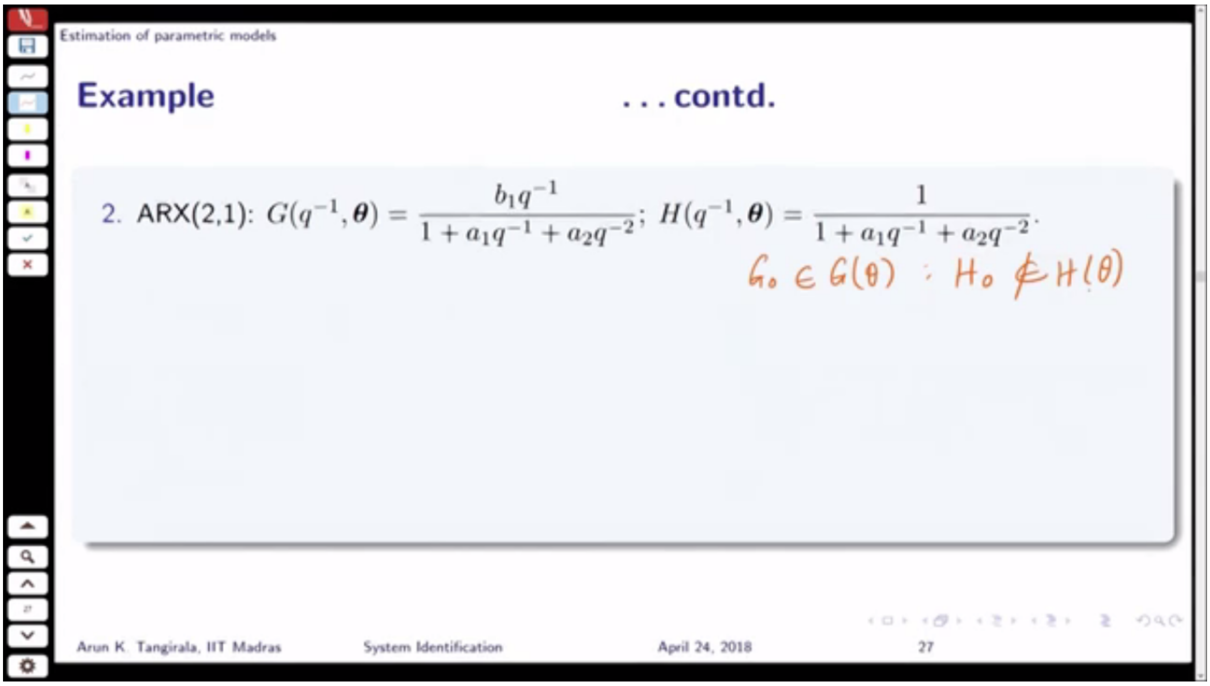


What happened.Correct. Right.That's true. That's true. So you have somehow managed to choose such a high dimensional parameters set. You are parameters is in such a way that you are able to. Here we are talking of equality in fact strictly speaking we should talk of within some arbitrary accuracy. We'll talk about that a bit later. For now we will restrict ourselves because there is--Otherwise there is no way you can talk of consistency and so on. Because we are assuming the true ones to be LTI and that they can be parameterized. If we don't assume that way then it will always lay in a higher dimensional space or outside the purview of your-- then what consistency can you talk about. Correct.

So in practice, yes, you will generate an approximation only. There is no doubt about it. Because we want to say whether you get a consistent estimator or not, you have to assume that you're G0 and H0 canbe parameterized. But that's a very good point that you make. Since we're the purpose of this discussion is for consistency, we have to assume that there exists a true theta not or true J0, otherwise the entire discussion falls apart. Okay. Okay. So what about this case?

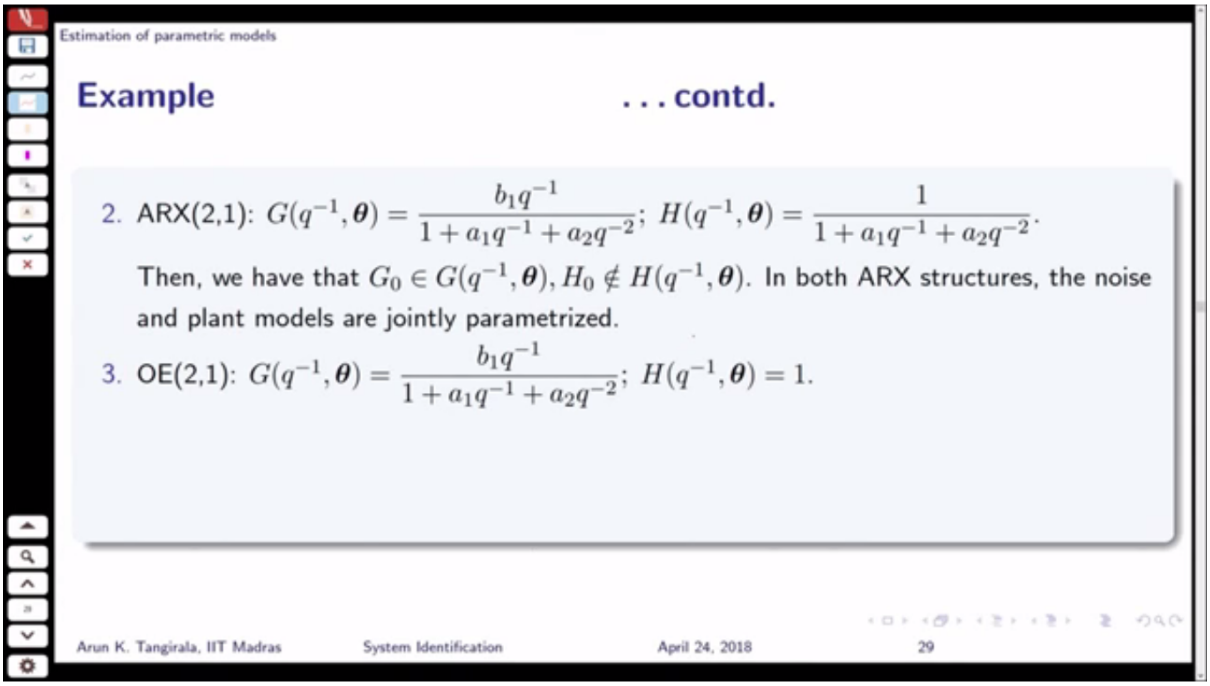
How do you qualify this scenario. I've chosen this model. The process is the same. Nowhere do you say? Does S belong to M? Correct. So we can say that G0 belongs to G of theta but H0does not belong to.

(Refer Slide Time: 24:27)



Clearly because the model structure says so,or the system structure has them. Correct. So we have clearly this situation. And in addition we also see that the plant and noise models are jointly parameterized. Then the third candidate.

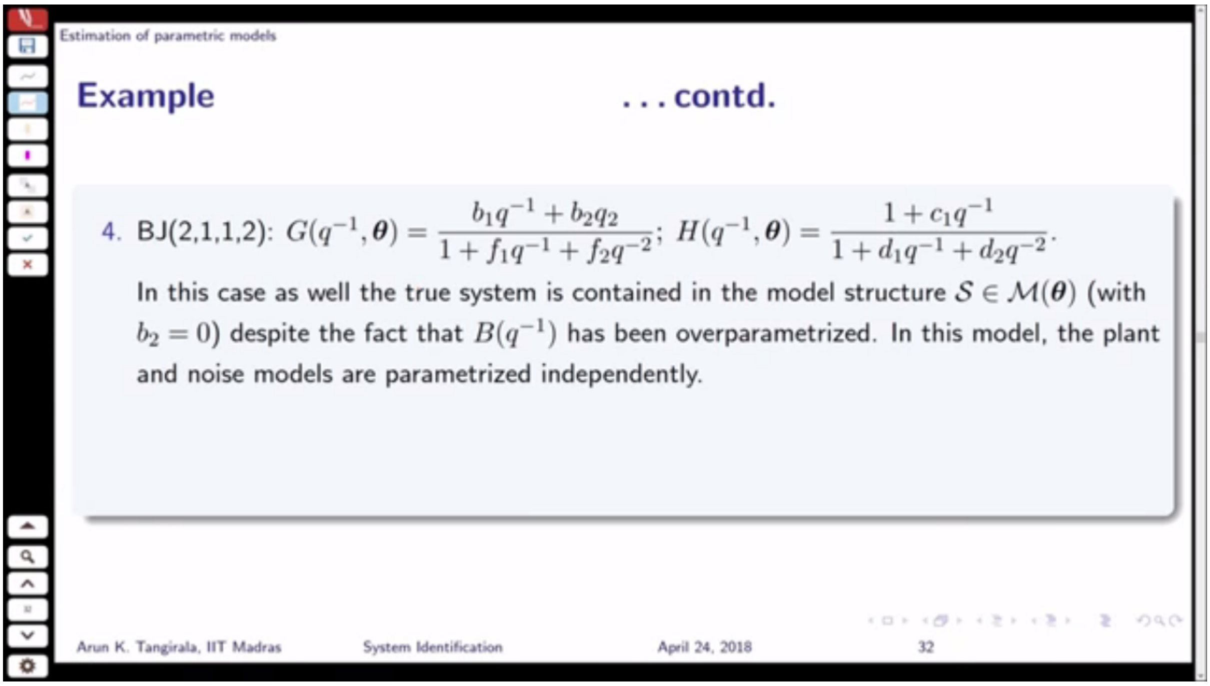
(Refer Slide Time: 24:45)



Sorry. OE model. What about this? What scenario does this correspond? Same, right.So once again we have G0 belonging to G of theta and then H0not belonging to H of theta. In fact it there is theta for H. But what is the difference between scenario two and three? They are not jointly parameterized. That point you should observe. As they say in the court please observe, take note of this my lord. Okay. The plant and noise models are parameterized independently.

And the fourth scenario.Is that we are working with a Box–Jenkins model. Now, what do you see? Do you remember the true system by know or forgotten? Yeah, we are only two slides away. In fact this should not be q2, it should be q to the minus 2.We'll correct that. There is no new operator that I've come up with. What do you think now? S belongs to M. Right. If you said b2 to 0, then you will get. So there exists one combination and that's it. So these are the different scenarios. Now you're able to understand what we mean.

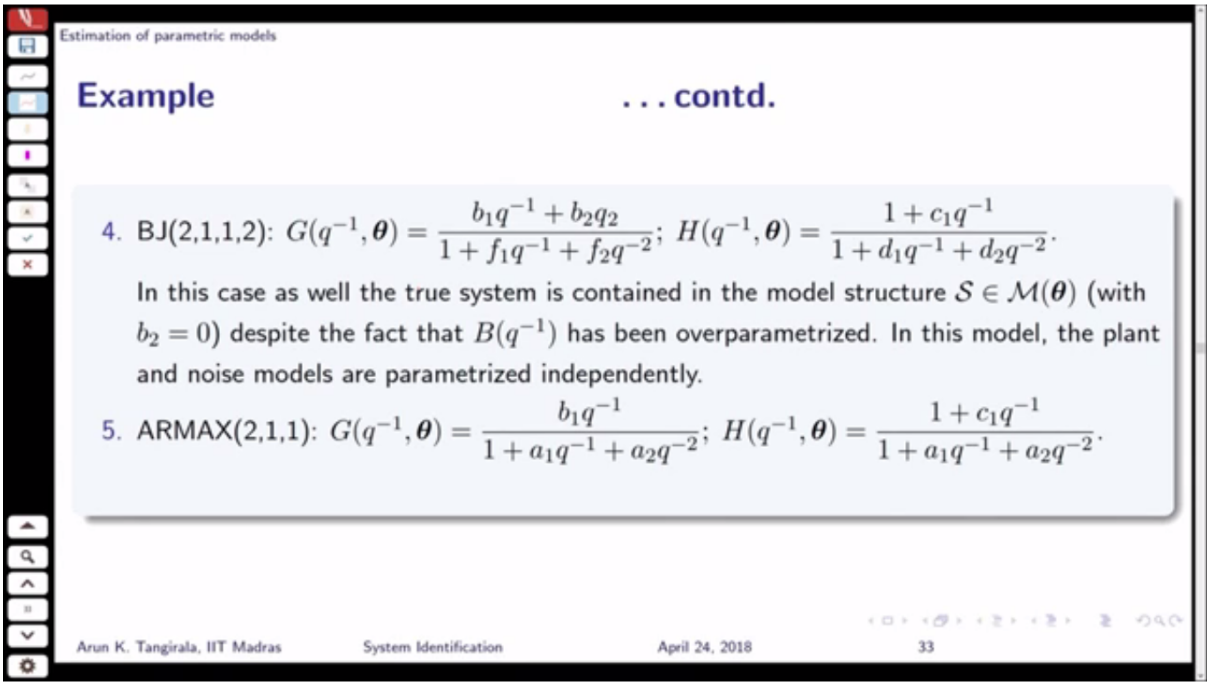
(Refer Slide Time: 26:21)



Of course B has been over parameterized, there is no doubt about it. But the other thing is also that the noise and the models have been parameterized independently.

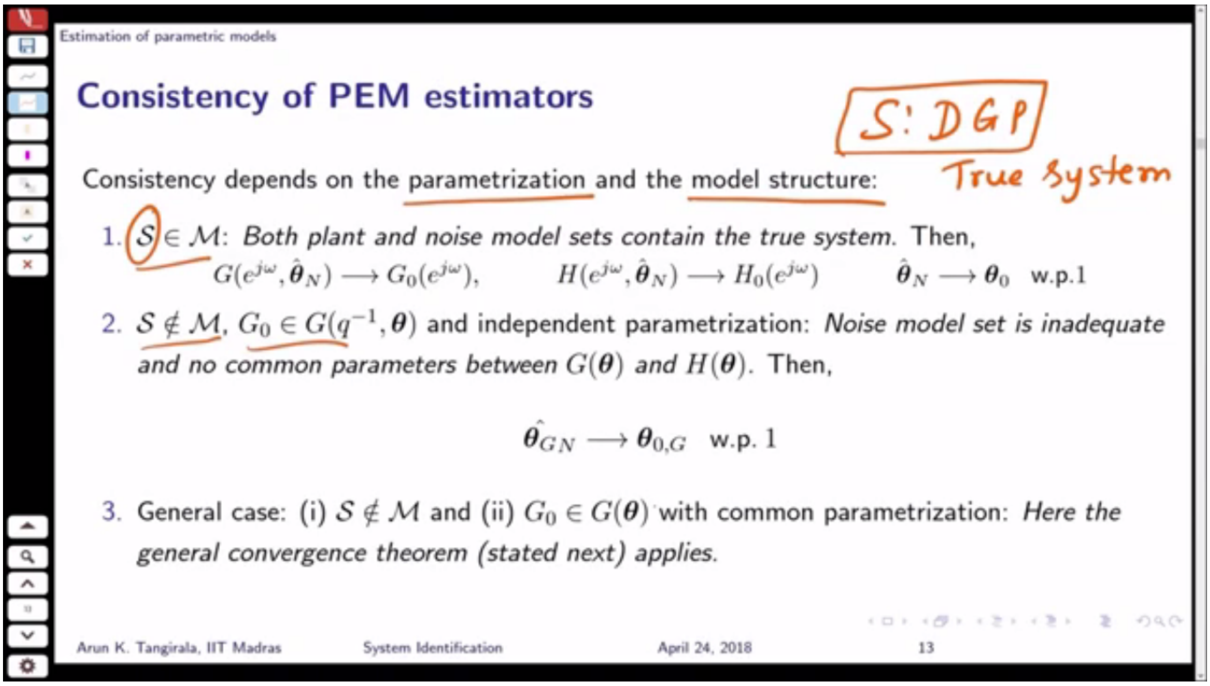
What about the fifth scenario.

(Refer Slide Time: 26:40)



S belongs to M. Correct. But what is the difference between four and five? Joint para-- So these are just-- now you have become an expert in figuring out whether s belongs to M by parameterization and join them. So now you can go back to the consistency thing. Yeah.

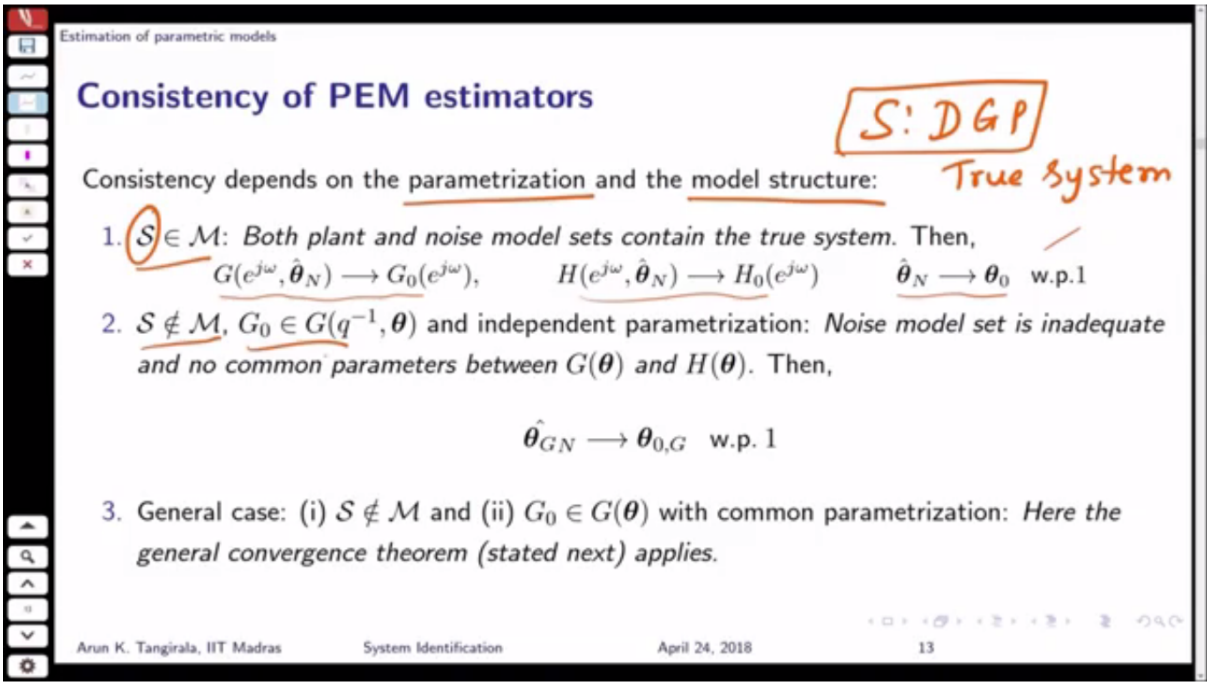
(Refer Slide Time: 27:03)



So the consistency results are given for three different scenarios. One is S belongs to M. Both plant and noise models contain the true systems.Here, there is no imposition on parameterization. As long as S belongs to M, whether the parameterization is joined or independent it doesn't matter. Your net the two nets that you have cast have captured both the plant and noise model. Then you are guaranteed consistency with probability 1. So which is almost sure convergence.

So look at the statements, the statement says that G of theta can converges to G0 not H of theta converges to H0 and theta hat also converges to theta 0. So one talks of models converging and the other the statement is of parameters converging. Correct. So this is the answer. So if I don't assume that there exists a true theta 0. That means that there exists a parameter that can parameterize G0 and H0. Then this entire discussion falls apart. Okay.

(Refer Slide Time: 28:14)



The second scenario is when, S does not belong to M but G0 belongs to G and this condition. The models have to be parameterized independently. And the noise model set there essentially is inadequate and no common parameters between G and H, that is what it amounts to. In this case you're guaranteed that theta hat G will converge to theta not G. That means G will be estimated consistently. And this is the case that you'll see in OE models. And that's why in the liquid level system, even if you had colourednoise, in the assignment you've worked out a coloured noise K.Even if you had assumed an OE model, you would have estimated the G correctly. Correct. So that's the beauty of output error models. All you have to do is make sure that you choose the orders for the numerator and denominators for G correctly. Don't worry about noise modelling. You are guaranteed that you will-- And the way how do you keep refining the orders for the numerator and denominator for G? Sorry. So you just look at cross correlation between residuals and the inputs and you will know. It will tell you whether the model is adequate or not. That's all .Yes.

Correct. So what is the issue. Yeah. Yeah. Yeah. G and G0. Correct. So, G will tend to G0 is probably 1, H will tend to H0is probably. Yeah. That is kind of understood.

So the third senior is, S does not belong to M and G0 belongs to G but with common parameterization. So what is this business of parameterization. Why on earth this parameterization is making such a big difference? Right. Intuitively as I've always said G and H arelike your two hands. That you can use to get your work done. What is a work done? Predicting or approximation. Right. That is what you want to do.

If I leave two hand freely, if I give you, you know, enormous ample amount of freedom. And I say both of your hands are free to do whatever in whichever way they want to move. Of course, you can actually get a lot of work done. And there is a chance that you will accomplish the work fully.

On the other hand if I say that, so that is what essentially is joint parameterization. If I parameterizethem independently, that means I have let my both hands free. And they've been given full freedom. On the other hand, with output error models, right. One hand is fixed. It says I don't care. H is 1. So this will remain like this. And I'm going to actually use only my left hand, one of my hands, left or right hand and get my work done. Yes, the interesting results is even with one hand being free, you can at least get the deterministic part right. That is what the interesting. So the first thing is not such a great because its intuitive, what is most interesting is the second one always. And that's my opinion.

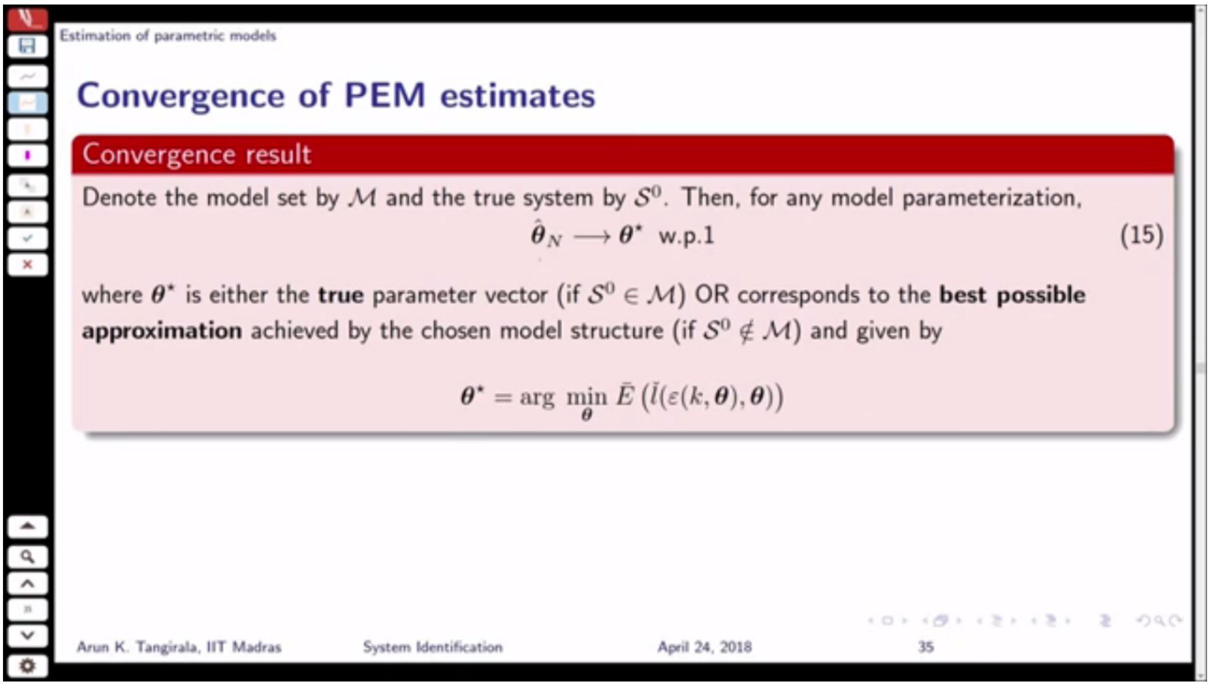
That it says even though you made a mistake in the noise model, don't worry. Don't worry about that.Under open loop conditions only by the way.All this results under open loop conditions. As long as because there are certain other conditions for consistency, I'm not stated that, persistent excitation, all of that is required on the input, identifiability and so on. It says you can get your work done as far as a deterministic model is concerned.

The third case is, when S does not belong to M. Yeah. You've been given you know, your net is not big enough. You don't have too many degrees of freedom. But one hand is able to actually capture the G, G0 belongs to G. Yes. But common parameterization.This hand is not fixed anchored to one point that is independent of this. This hand is actually anchored to this particular hand and as a result it says that you will not be able to estimate consistently. Okay.

So in this case consistency is not guaranteed. So this is the case of ARXS, remember. It can be the case. ARXS can belong fall into this category. If you go back to the liquid level example. We chose first order If you recall, in which case G0 belonging to G,satisfy. But we chose the wrong model structure, that is also, okay. That is why S does not belong to M but this applies. And as a result, I could not actually estimate the model consistently. There was a bias in the estimate. So Lume, Wheton and others proved not only proved this, they went along and asked the other question. Suppose this is a scenario, yes I know that they'll get a biased estimate. What will be the optimal estimate?

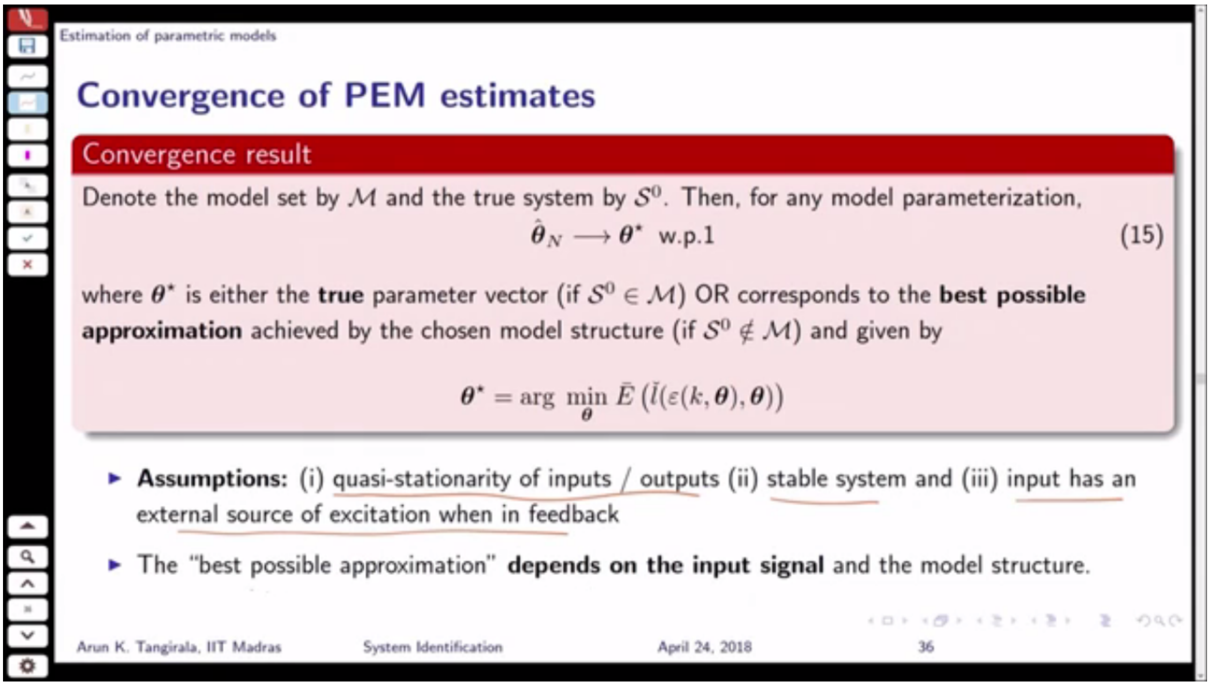
Can I say that to what estimate does it convert to.Yes. Now I know it doesn't converge to the truth. That means this is no longer going to hold. Theta hat G will not converge to theta 0. What will it converge to, is the question? And that is the general result. That I'll talk about now. So the general convergence result is stated again due to Lume.

(Refer Slide Time: 34:32)



Is that theta hat N will converge to some theta start with probability 1.Which theta star will it converge to? It will either converge to the true parameter vector if S belongs to M. Okay. Orcorresponds to the best possible approximation achieved by the chosen model structure. And that best possible approximation is determined by solving this.I will work out an example and I'll show you what this means, then you can understand. And remember I said there are some assumptions for this convergence. One is that input and output are quasi-stationary, it’s a stable system. Input has an external-- when feedback there should be a deter signal. When you're operating under closed loop conditions there should be an external source of excitation in input. Otherwise the excitation becomes endogenous. All the excitation in the input is derived purely from output. That's not good.

(Refer Slide Time: 35:30)



Now,let's work out the example and we'll understand what we mean by best possible approximation. So consider this example. Where have ARX model for an OE process. So what is the system here? The system, the data generating process is OE and is being given here.

(Refer Slide Time: 35:49)