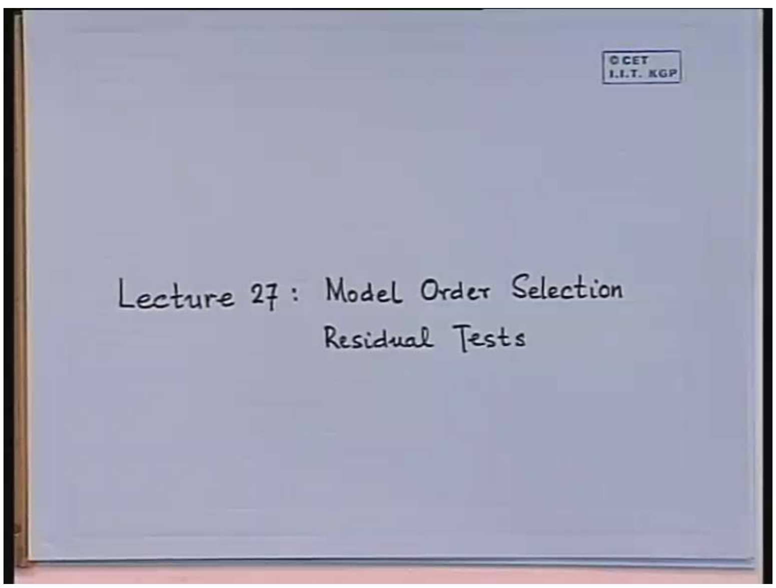
Estimation of Signals and Systems
Prof. S. Mukhopadhyay
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 27
Model Order Selection Residual Tests

So we are nearing the end of this model and the overall course slowly. So in fact at the end of the last class, one of the students asked the question, which we try to answer today.

(Refer Slide Time: 01:12)



He asked us to okay fine everything even list square etcetera, but how do you know what is the order of the model; that is a very important thing and in fact the point to note here is that, it is not that there is a there is a it may or may not be true that so called true model exists. In many cases in many cases, you are may be you are trying to trying to identify a non-linear system using a linear model. So in that case there is no there is no true linear model, there is no true parameter, it is only a question of approximating the system; the non-linear system behavior under a set of given inputs, I mean kind of inputs.

So therefore the that that is why I have use the term model order of selection; I am not saying model order of determination because it is not that you you you can I mean nicely solve out, that it it it may or may not be true that you can nicely solve out that the that the model order is going to be true rather you have to select, whether model whether selection of a model order of two solves your purpose well or selection of three is required or may be selection of one is sufficient.

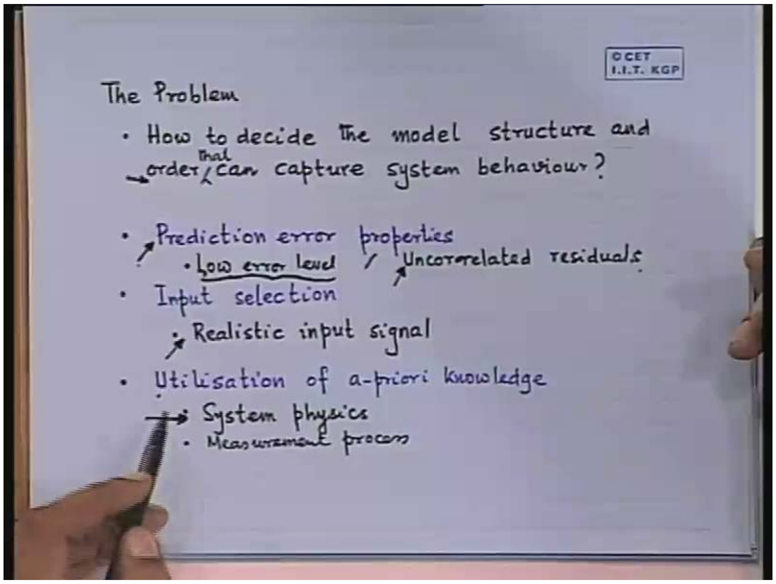
So so in general we we would, we would go by the simplest one that works for our purpose. So it is more of a question of what works good enough for my purpose. So so in that sense it is it is essential a selection problem and we will see some some you know; broadly I mean I will not going to absolute mathematical details of the procedures because there are certain things, certain basic which will be required which you have not covered in this course. So therefore, but they are they are available in very standard material on statistics or in linear algebra, but I hope that the basic approach to the selection problem will be clear and after you have selected, actually you know this process is somewhat it is it is it is like a chicken and egg problem.

So if you have after you have selected a model, I mean the proof of the putting is always in its taste. So so I mean how do you how do you know, whether you have selected a proper model; so after you have selected the model order, you have to actually go through the estimation and get some parameters based on those orders and then see whether whether your whether your estimation process was really completed, whether you could really feed the data, whether your behavior maps. If it does not map, that means if you did not get a good taste then you come back then you change your model order.

So it is not that, things can be always determined a prairie; sometimes you can carry out some tests here and there, but in many cases it is essential and iterative exercise. So you have to select something based on what you feel is could be right, then write out and see whether it is really right otherwise; you again start all over with with different model order. So for for so so essentially, there is also a kind of model validation involved, that is you select a model order do the estimation, get a model and then you validate the model basically that would examine whether it is good enough.

If you I I I I mean for doing this doing the validation, often you have to carry out some tests on residuals, so so we will some of these things.

(Refer Slide Time: 05:11)



So coming to the coming to that the basically the the the problem; is you know, how to decide the model structure and order that can capture system behavior. And by this is a rather qualitative statement but what it tries to say is that, it should be able to capture the system behaviour good enough, what do I mean what do I mean by good enough? For example, if I know that in a that in a regular scenario; for example suppose, you are if you are trying to model a model a mechanical structure, now obviously that that mechanical structure will will will perhaps never never experience forces of the order of Mega hertz, that is that is very unlikely. May be it will not even experience forces of the order of kilo hertz, may be it it will experience only forces in order of hertz; which means that you need to able to estimate a model which is good enough up to let us say hundred hertz.

You do not at all bother about, how that structure would have behave if you really excited it at ten kilo hertz; you are not concerned with that because because that is not required in the purpose, so this right. I mean in fact things have completely different model structures, depending on the use I mean I have one of my teacher's saying that, the that the that the model of the transformer; probably this I have said earlier in the course also that the model of transformer

I mean, a normal power transformer at at fifty hertz is completely different from its model at fifty kilo hertz. I mean at fifty kilo hertz; it is it is predominantly it will consist of intendent capacitors which we completely ignore, when we draw draw an equivalent circuit model of a transformer. So so so so we are we are talking about system behavior for for for which the model is going to be built and so firstly; but in that zone it should be able to model the behavior reasonably well, again reasonably well depends on what is the purpose for which you are trying to determine the model.

For example, if you are trying to use it for control then it will happen in it will happen in many cases; that that even if you have little little modeling uncertainties, the the controller will force the response, so the control in fact that is the that is the duty of feedback control that it can give performance, even if you do not know the system model very accurately. I mean especially with when when you have high gain feedback then the closed loop transfer function is g by one plus g h.

So so if h is high it becomes one by h almost because h is much much larger than one. So the so the closed loop behavior typically depends on the feedback response, so rather than pitfall response. So what it what it basically says is that, even if your model is little in accurate; if you are using it for controlling in many cases, it will not matter right but maybe may be if you are using it for prediction or for or for stimulation it might matter. So so that is a very important thing that needs to be remembered, here.

Now when shall we say that, that is how do we how how do we how do we characterize this thing called ‘can capture system behavior’, can capture system behavior means what? So that you can capture depending on how do you test this; that it can actually capture system behavior. So you will you will test it, typically by looking at prediction error properties; that is what the what the model says is going to happen and what is actually happening. So so I mean obviously speaking, this this has to look at prediction error properties. So you will check for whether the prediction error level is sufficiently low, if it is too high then you will then there is something wrong with your model.

Secondly, you in some cases you will you will get some errors, because you are trying to reject noise. So so so the so the model prediction and the measurement will be defferent. So then it is good to be different because you are trying to you are trying to get the get the get the process variable before the measurement, right. So you are trying to reject the measurement noise, but so while there may be some prediction error; you you definitely want that the that that the that the residual which you are getting is not correlated with the input because, if it is correlated with the input that means that there is the part of the system dynamic which you are which you are unable to explain using the model.

So if you if you get if you get uncorrelated residuals which are mutually uncorrelated and which are uncorrelated especially with the input; then you know that that at least whatever residual you are having that that does not contain any component of input, so you have explained all the input. So so these these are these are very important properties of the prediction error; that you usually check to know whether you have you know extracted all juice from the data. Whatever was there which was due to the input, you have extracted into the model. So whatever is remaining is actually useless, that that has no connection with the input. And not only that, this you have to do, you have to you have to choose the realistic input signal; otherwise you can get you can get very different parameters.

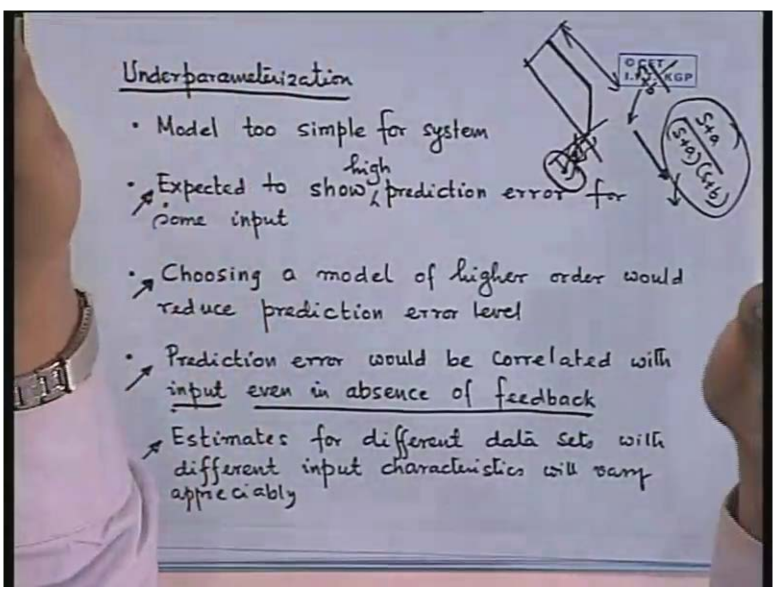
If you if you try to estimate a a model using d c, I mean d c like thing and if you put as you can see component, you can get quite different models. So you have to ensure that; the that the input data with which you have you have excited the system; which you were using for modeling spans the whole gamut of reasonable behavior, which we actually expectation we put the modeling to use in terms of amplitudes as well as frequency. And obviously to be able to select model structure, you have to use system physics. If in a circuit you know that there are there are three capacitors; then it is very likely that that, you must start with the third order model because because the because the physics tells you, that there are that there are three energy storing elements.

So the model order is likely to do with a three, now it may be more than three depending on you know some parasitic effects also; but at least your your your your structural knowledge of that modeling principle will at least make you start from three, that will be a good initial guess. So you have to use system process and you have to use also the knowledge of the measurement process. You know in many cases, it happens that you know that the the suppose you know; that the measurement process may have bias, it may have d c bias. So if you know that it may have a d c bias, often often what you do is you would like to rather than rather than working on normal data, you may like to work on difference data, that is you you create a different input which will be u k minus u k minus one, call that the input.

See as far as you are using a linear system model; you are trying to identify whether you are using difference data or whether you you are using sum data, it make as long as you are doing linear operations on the input model events are same, almost, right. So so so you may like to estimate your model based on difference data, so such such such knowledge is have to be used when you are when you are doing I mean the model order selection.

So now let us see some let us let us get more into specific.

(Refer Slide Time: 13:00)



So first of all, question is how do we know wether the model is too simple? You know there is there is always a tendency of choosing models which are; you would always like the simplest model networks. So so so how do we choose the how do we know, whether whether we have got underparameterization; typical example is that, the real model may be a second order thing, we are trying to look for a look for a model which is a first order model. So when do we know I mean, how do we know it will not work? We have to we have to actually go for second order model. So firstly that such a thing is actually expected to show high prediction error for some input.

If you really need a second order model, now if you can this you can realize by you know seeing the Vude VUDA plots; let us say they are magnitude plots. For a for a first order system is going to be like this, for a second order system it will be like this, and then like this, correct, two poles. So but as long as your input is up to this, up to this frequency is like a first order; I mean at least at least asymptotic plot, the true plot is not that let we are just we are for for for simplified argument, we are we are thinking in the terms of the asymptotic plot. So at least up to this can be approximated by a first order system. But the movement you come to this frequencies zone, if you try to feed this you can feed this part perhaps; in a given range because around a given point, but as long as soon as you get this forty degree per decade slope, this forty degree per decade slope cannot be obtained by a first order model, correct.

So so if you are only considering inputs in this region, then you do not be in second order model; but if you are using inputs in this region then you do need a second order model and the first order model will not do. And and then and the difference between the model model, I mean model prediction of the output and the true output will actually show up; if you if you excite it by a frequency in this region then it will show up. So so so the prediction prediction error will build up, if you give it input of high enough frequencies; that is the first thing, that the that the that that the predicted output from the model and true measured output will not match.

Second thing is that, second thing will show you that; if you are having a prediction error level then you have to you have to experiment. You know I mean, you do not know what is the model order; so you try with one see what is the prediction error level then try with two, if you find that the prediction error has suddenly dropped then you know that you should go for two. On the other hand, if you if your prediction you know if if you have a second order system and if you if you estimate it using a using a third order model, what will happen? Some some some some extra pole will appear both on the numerator and the denominator, but but as far as the prediction is concerned it will be the same; for example, s plus a by by s plus a into s plus b, this is a second order model, actual model is one by s plus b, this is the second order model, but but but if you stimulate one by s plus b and if you stimulate these two, you will get the same output.

So so so because you will get the same output, you will get same you will have the same level of prediction error. So there so in such a case by increasing your model order your prediction error will not fall, it will remain round about in the same way. So if you does not fall then you know that; you did not have any gain by actually increasing the model order, remember that if you are increasing the model order you are increasing computation, right. Not only you are increasing computation, you are actually it can be shown that your parameter variants also increase, that is another another analysis.

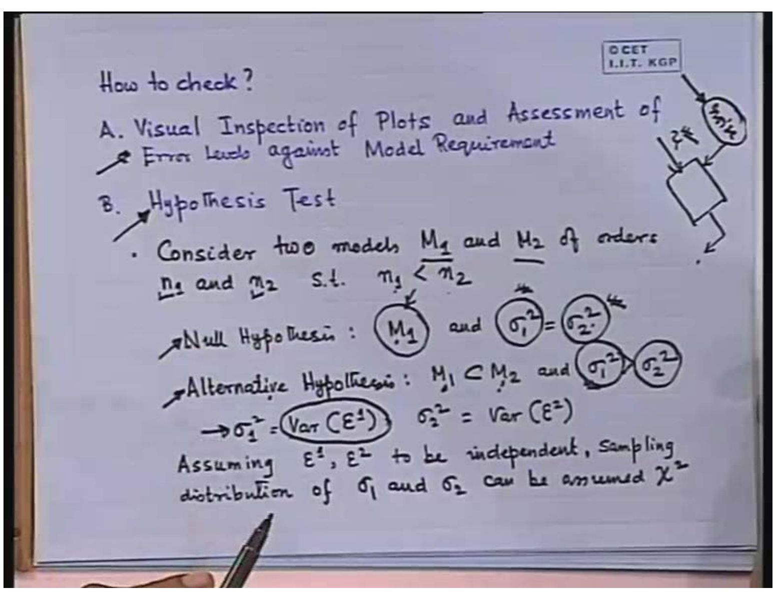
If you take a second order system and if you try to try to use it, using a tenth order model structure then your then your then your variants of the estimate will will increase; that is we I mean if you do with this data you will get something, you will if you show if you use another data, you you might get another thing. So that is why you should use this smallest structure which feeds the data. So so that is the second criterion; that you you should you should try with increasing model structures at one point you will find that, you are not getting any any further benefit.

Secondly, as I said that the prediction error would be correlated with the input even in absence of feedback; if there is feedback, prediction error will naturally be correlated with the input because the input is being computed from the output. So so this thing will hold only if only if, there is you are you are not doing closely by identification, if you are do if you are doing closely by identification this thing will not hold. And second thing, that last thing that will happen is that; if you use different data sets, for example, suppose you have used one data set which let us say contains frequencies from the order of zero zero to ten hertz, you will get one model. If you take say let us say, five to twenty hertz you will get totally a different model; but the parameters start varying too much depending on the inputs, depending on the input property.

Now, for example if you have if you have a non-linear system then you can actually linearize it around operating points, operating points depend on the amplitude of inputs typically. So if you use a very high amplitude input then you will get a linear model which is around the different operating point; therefore the therefore the linear model will be different. So if you have if you have not modeled it appropriately, then the then the estimated parameters will actually be depending on the input signal that you are using.

So if you use different different input set, you will get different different parameters that is also an indication; that your your model structure is is not good for all that input variation because its changing. So these are you know roughly ways of detecting that a model structure is too simple, it cannot and it capture the system behavior under the required conditions of the input. So we will so we will look at some of them a little more in detail.

(Refer Slide Time: 20:00)



Firstly, we obviously this thing cannot be undermined; visual inspection of plots and assessment of error levels against model requirement, this is must, you must take plots and you must see by with your eyes that is always important, it cannot should not be discounted in a practical case. And what is meant by assessment of error level is that; you may one who who has knowledge may decide, that okay even if you have a point one parameter error this this is not going to be of much problem because after all; what is my what is my accuracy of this or that this this use of model point one is okay. So somebody has to decide, so that is that is a very practical requirement but but but we will take a different we will look at straightly more more mathematical approaches.

So here one of the one of the approaches is that we use, uses a kind of technique mathematical or other statistical technique called hypothesis test. The idea is that you make a statement which is called a hypothesis and then you try to find out that, what is the probability that; see you had been you had been given some facts which come in the form of data, some values you you have got. So you make a statement and then you try to evaluate that given this this set of data to me which which has come from experiments, what is the probability that this statement is true? If that probability falls below a certain level, you say that the statement is this hypothesis is is not acceptable, so you reject the hypothesis. This is a standard method of statistics; you will find in any elementary statistics book, this is called hypothesis testing.

So here what is our what is the hypothesis, that we are trying to check? So suppose we have two models one is M 1 and other is M 2; one is of the order of the order n 1, another of the order n 2 and n 1 is less than n 2. So we are trying to we have two models; may be one is first order and another is second order. Now if you if you estimate them now you use a first order model structure, if you estimate you will get some prediction error, now this prediction error; similarly if you you use M 2 you will get another prediction error.

Now remember that, this prediction you you you have got one data set, just one data set one or two whatever you have got a finite number of data sets. Now this prediction error sum, that you have got is actually one sample of the actual prediction error of the random variable, do you realize that? That if you perform ten thousand experiments with the same input; the input deterministic input value is same. See what how is our output being generated? There is an input and there is some random elements which are which which is my noise.

So this is a random element. If I keep you and if and do ten thousand different experiments; I will get ten thousand different noise element because that is a random variable. If I do estimation with that, I will get ten thousand different prediction errors, physical sequences will be different. So so they are also random processes, now we are assuming that, let let those random processes one of them have a variance sigma 1 another has a variance sigma 2. Remember that I I am I do not know the value, at best what I can do is from the given data; I can I can calculate the variants of the prediction error, simply by taking square and then sum and then I will get one sample from the from the from the prediction error random process.

The random process is a general thing but based on that data, if I assume one prediction error if I compute then I will get one sample from that random process. So the so the variance of the prediction error, that I calculate it will be a sample, right. So that also may be what we are trying to say is that; the prediction error that you have got now for for for this data set, you may get a different prediction error in a in a in a different data set, so you must be aware of its uncertainty as the random variable and then try to decide, fully keeping in mind that this value that I am using for decision is a random process, is a is a random variable.

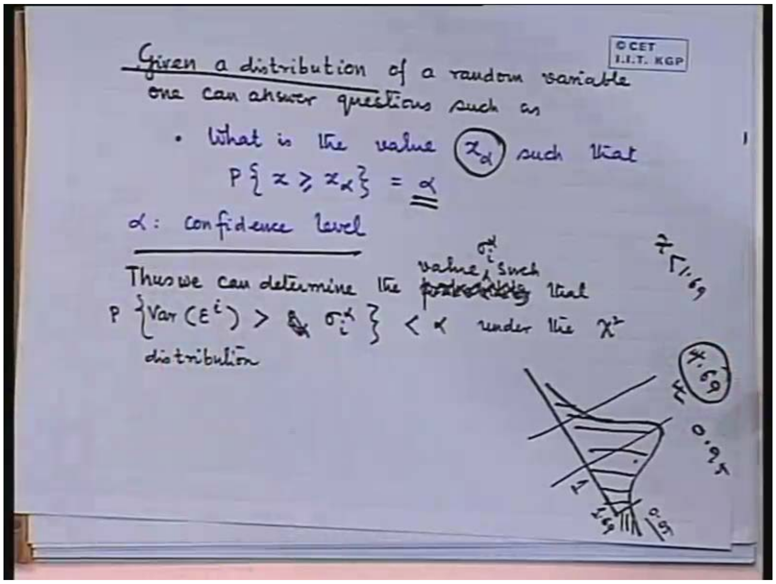
So I am trying to so what is by null hypothesis, null hypothesis means; the basic hypothesis which I want to reject out exit. That the models that the true model structure is M 1 and the variants of the prediction error that will happen; if I use the model structure of M 2 that is sigma 1 and sigma 2 are same, that is which means that M 1 is a is a is a sufficient order model, please try to follow this carefully. Suppose, the model was was was truly M 1, the first order model then if I choose a first order model structure; I will get some prediction error, if I choose a second order model structure, I will get the same prediction error on an average, is it not? Because its true model is first order. So again my parameters will match. Now so so I am trying to say that, I will accept this model to be the true model; if these variants are the same that is the hypothesis, I mean. But I do not know these variants; I know only a sample from its random process because I have got one set of data, right.

So based on those samples one sample from this random process, another sample from this random process, I have to make my decision because I cannot conduct ten thousand experiments. So and what is my alternative hypothesis? That is if this is rejected, then my my alternative hypothesis is that M 2 is actually a more general model than M 1. So M 1 is a M 1 the clause of models described by their structure is actually subset of the clause of models; when a first order models are a subset of second order model structures, all first order models can be described by a second order model structure, but there are second order model structures which cannot be described by first order model structures, therefore this is subset of this. So if this is rejected then my then my alternative hypothesis is that, M 1 is is actually a subset of M 2; and sigma 1 square, that is the variants that I am getting from M 1 is actually greater than that from M 2.

So there is a possibility of reducing the prediction error, correct. Now there is some there there there is some theories; which I have some what you know complicated, we say that if you have a set of independent Gaussian random variables, say and if you take their squared sum then that variable which is their squared sum, it is a it is a function of number of Gaussian random variables. So so so it will have a it will have a distribution, that will not be Gaussian; Gaussian variables if you just add them, linear operations keeps Gaussian, but if you square them and sum them it will not remain Gaussian. First it will firstly that random variable will not have any negative number, obviously because it is a squared sum.

So such such a variable it can be shown as a distribution which is called chi squared distribution. But, so this is a this is a this is a very common distribution which is which is used for especially for testing Gaussian of residual whether whether whether some random variable is Gaussian or not; typically people use chi squared test, right. So is this this also again a standard statistical theory. So here what I am trying to say is that, what is if I compute variants, what I am doing in variants? I am taking squares of individual prediction errors and then sum it, how do you compute variants? So so it is not unreasonable to assume that, this distribution of the variants is going to be chi squared.

(Refer Slide Time: 28:55)



Now, now you now you what you what you try to if you if you if you accept that; that the prediction error variants random variable is going to be chi squared distribution variable, if you if you if you accept that then you can ask that what is the value of chi l? That you you got a value of variants, right; now what is the probability that it has actually come from a chi square distribution of a given type and if that probability falls too low, if that probability becomes less than ten percent then you say no no no then then then then this hypothesis because the chi squared distribution is is calculated under that hypothesis.

So then the this hypothesis led to a chi squared distribution which cannot generate, this kind of a variants; when the probability is too low, so my so my hypothesis is wrong this is the kind of approach that you take. So what you what you can do is actually given a distribution of for example; you can you can easily understand this in the in the case of Gaussian, suppose somebody tells you that if you do an experiment, if you if you pick an object or if you pick a value or a number, it will come from a from a from a distribution of mean one and variant one, right.

So you know that the distribution is like this, this one. It he tells you that it comes from comes from a distribution, then it comes from mean one and variant one. Now you draw something, you got a value one point six nine, what is the probability that that is? What is likelihood that that this number has come from this distribution? Then you have to find out that see if if this is 1.69 then the probability; that that if you draw something from that distribution, the probability that that x will be less than or equal to 1.69 is the is the is the area under this curve.

This is the density curve, for a distribution curve the probability that, x is less than or equal to 1.69 is the area under this curve. If the area under this curve becomes 0.95, that means the area under this is 0.05, right. So that means the the probability that you got, you will get anything get greater or equal to 1.69 is is just five percent. So if you got a number 1.69 it is quite unlikely; that it has come from this Gaussian process, it is very unlikely I mean, I am not I am not using consistent numbers, I do not know for a for a for a standard deviation of one, what what is what is the ninety-five percent, yes ninety-five percent shows to two sigma kind of thing.

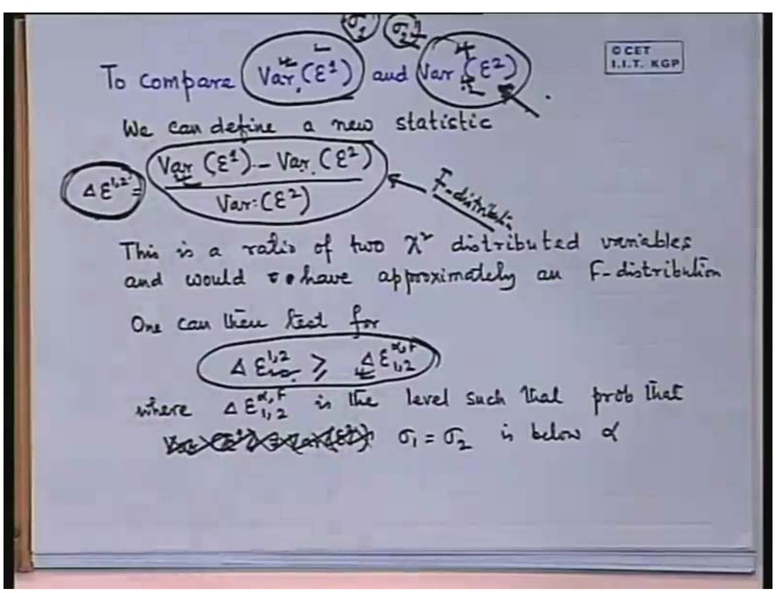
So so so this is the this is the argument; so given a distribution, if you if you know the distribution then you can and if you are given a value; you can determine that what is the probability that the value came from this distribution, this you can determine. This this what I am trying to say simply by calculating the area under that, area under that distribution curve. So now you are given a chi squared distribution and and from the data you got a value, you got a value of the prediction error variants. So now you have to under you have to calculate, that what is the probability that this prediction error variants which I computed from my data has come from that?

Has come from that chi squared distribution? If that probability falls, say less than some chosen value alpha which you can chose as ten percent, five percent depends on how sure you want to be. Then this alpha is typically called a confidence level, if you want to be ninety-nine percent confident then you will say that this alpha is zero point zero one, okay. So this is the this is the approach, that you first make a hypothesis; based on the hypothesis you try to characterize, that under this hypothesis some quantity which you are going to test, that quantity is generally called as statistic, not statistics, it is called as statistic. A statistic is a function of former which is which is which is computed from random variable and which does not contain any unknown parameter.

So if you take variants as the statistic then you can find out that; whether this variants came from this distribution, basically this is the approach, that there there are I mean I am skipping the mathematical details because that will involve into guesses, okay. So essentially basically I mean we are we are just trying to find out, whether basically the same thing; we are trying to find out whether the variants of the error is too large or not, but we are doing it in a in a in a mathematical way. And not only that and keeping in mind, that I have just picked one value; so it is a random variable, it could be straightly off I am not and I am not using a steady threshold because the threshold has been computed best on the randomness properties of the prediction error process.

So the threshold is computed using the fact, that you might get a get a straightly off value when you are choosing a random variable.

(Refer Slide Time: 34:35)



So this is the principle. Similarly, if you want to compare; whether the variants from M 1 and the variants of M 2, whether they were whether variants of the second process is actually variants of the variants of the first process? It is again, you will say that why is it I mean that it is trivial; because we have already computed this and we have already computed this, so one will adjust, so what is there big thing to test? One will be a larger number, remember that we are not trying to test that, we have got two distributions.

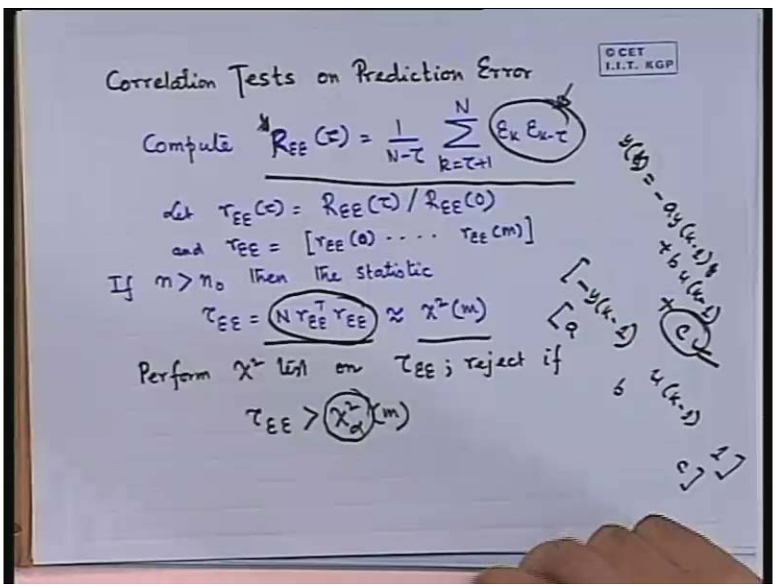
Once this random process actually, I should not have written this; this is this is this is strictly speaking, wrong, I should have written in that whether I can compare sigma1 and sigma 2. That is this process this is random variable, it has a standard deviation. This is a random variable it has standard deviation; that is standard deviation is is unknown, what I when I have calculate the, computed these two, I have just got a sample from this, I have just got a sample from this distribution and a sample from that distribution.

So I am trying to find out that, if I get this sample from this distribution and this sample from this distribution; then what is the probability that the variants of this distribution is actually greater than the variants of this dis this distribution. So from two samples, I am trying to assess what is the probability that the actual variants is a is larger than that is the that is the settle thing. So again you can define a variable like this and in this case; it turns out that if you take that in this case you see, if the if these are individually chi squared then their sample will also be a chi squared, somewhat difference.

So and this is the ratio of two chi squares, so it turns out again theory of statistic; that this kind of ratio of chi squares variable have a distribution which is called an F distribution, is the these are very well known distributions, F t chi square. So then you have to do, so now you know that if you can if you can compute this quantity; then this comes from F distribution, so again you check that whether under the under the F distribution property whether this will should actually come out to be what you got. So principle is the same, you you you compute a statistic, find its property, find the the property of the distribution and then check; whether this value that you got, how probability is to get it from that distribution, the same thing again.

So then you you can again test whether; what is the probability that this is greater than this? If if this is greater than this then you reject the hypothesis, saying that it cannot be so large with with alpha level of confidence you can say that it cannot be so large. So this is a way of checking because you know variants are; the why we have to do so much because you have to always remember that, if you are seeing a plot if you are seeing a plot for only one data set, what is the guarantee that if you take took took took another data set, you would not see the different plot? So you cannot you you that plot is not certain; you have to always be aware of its uncertainty as the as the random process.

(Refer Slide Time: 38:07)



Next is a correlation test on prediction error. So why because we know that prediction error are supposed to be white. If you have taken out all information related to input and assuming that, you know these are all assumptions; that noise processes are generally of white nature and they are not correlated with the input. If if we if we have if we have unable to explain the whole input then this will also be uncorrelated with itself; if the original noise is uncorrelated with itself which was an assumption. Remember that, if you are if your measurement process includes a d c term will happen? It will never happen; so which means that, but if your if your measurement process includes the d c term, then what will you do?

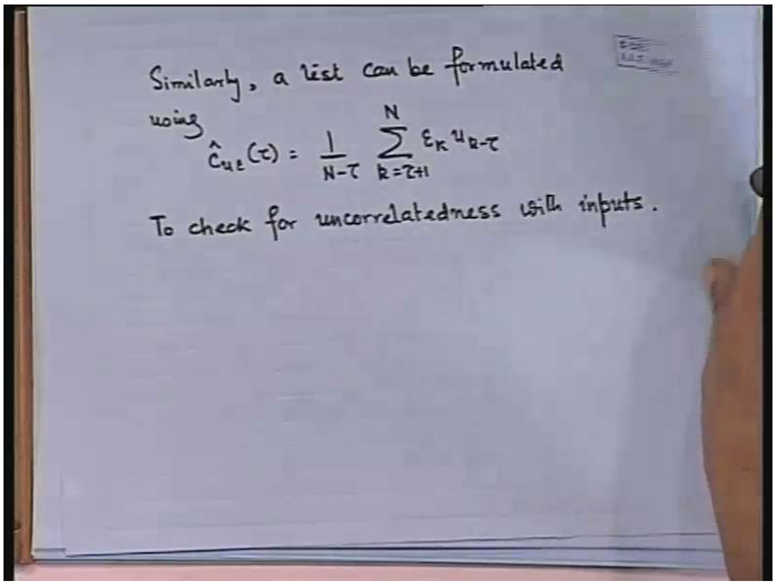
Then what you should do is that; in your model you should contain a d c term then you should write, that y t is equal to minus a y k minus 1 t plus b u k minus 1; not t y k, u k minus 1 plus some c. If you think that your output measurements, contain a d c then you should in your model structure you should include the d c and then say that; my my my data vector should be minus y k minus 1, u k minus 1 and 1. And my parameter vector will be a b and c. If you think that you have a you have d c bias in your measurement; you should model that bias because if you do not model that bias, this property will never be satisfied, okay.

So it is it is expected that, you have the regular properties you have all modeled then whatever will be remaining should be uncorrelated, that is the expectation. So again remember that if you if you compute this from your data; you will get a you will get a sample, is it principle is the same? You you want to test for whiteness of this that is for only for R epsilon epsilon tau equal to zero, you will get some correlation. If tau is anything greater than zero the correlation should fall, sharply, okay then it is uncorrelated.

So again you have to you have to test those things; that you have to you have to form a statistic and then check whether it is whether the whether this quantity is chi squared distribution; because again you are checking for correlation which is like a which is like a product of variables, so again the distribution is going to chi square. So here you are checking, so you see that the same kind of test by by different; so what are you what are you trying, what are you guessing? That you are what what what are the steps, that depending on the property that you want to check; you have to formulate a suitable statistic, in this case this is the statistic.

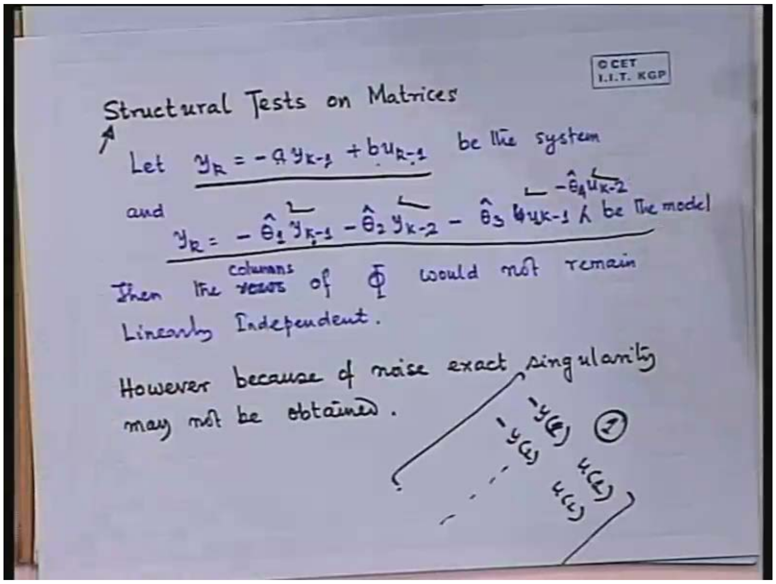
And then you have to choose it in such a manner that; it that it gives you a known kind of distribution, for example in this case it gives chi squared distribution. And then check the value against the properties of that distribution using hypothesis test, right. So just wanted to explain the principle, similarly you can you can say this is this test whether the residuals are auto correlated or not epsilon, epsilon. Similarly, you could test whether the whether the residual are correlated with input in which case you have to test for cross correlation.

(Refer Slide Time: 41:56)



So you could also formulate something like this, epsilon k u k minus tau. And then again formulate suitable statistic, such that this will follow a particular distribution and then check for hypothesis testing.

(Refer Slide Time: 42:13)



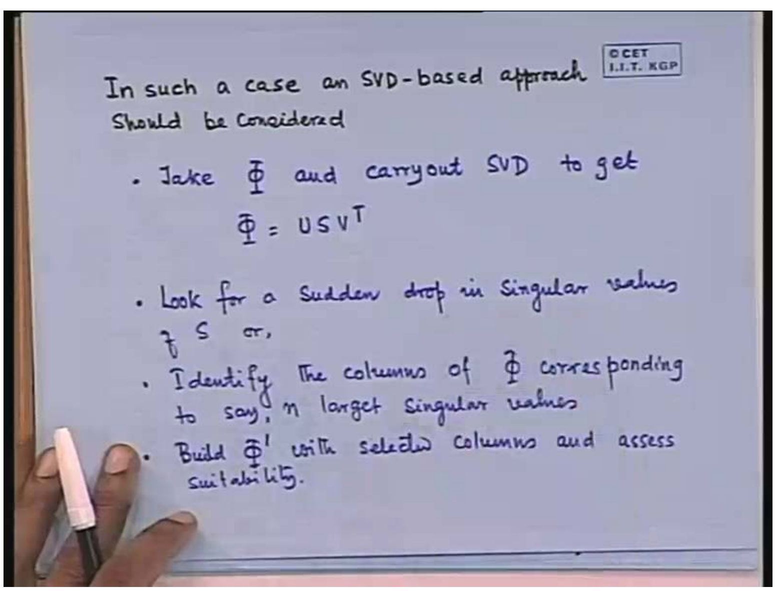
So this is a this is a this is a general approach; which is followed in in checking very much used in model order checking. Then the the other approach which is simpler, conceptually is to do structural tests on matrices. Suppose, you have a data vector, right; so you have minus y k minus 1, say minus y 1 u 1, minus y 1 u 1 and so on. This is your data vector phi, right; this is this is your data vector for a first order model.

Similarly, you could have a data vector for a for a second order model. So so let suppose the suppose the system is truly first order; now I am assuming that the system is exactly like this and you have chosen a model structure which is like this which is second order. We have taken two past inputs. Now, obviously you can you know that y k minus 1 is the function of y k minus 2 and y k minus 2 and u k minus 2; because this because this system is like that, the true data which you are putting in the matrices, their y k minus 1 will be equal to minus a y k minus 2 plus b u k minus 2.

So this row this row and this row are going to be correlated, linearly dependent; if you if you take this row multiply by a, if you take this row multiply by b, you will get this row, row not row column, right. So therefore, the rank of this matrix will fall in general; what is the rank of this matrix? This the rank of rank of a matrix of dimension m by n is minimum of m and n, provided when a maximum it can be is is minimum of m or n. So generally the number of data points is much much larger than the number of then the then the then the then the then the number of columns; so you have may be you have thousand rows and you have only three columns, so that kind of difference. So therefore the the rank is always is always determine by the number of columns, theoretically, practically, rather.

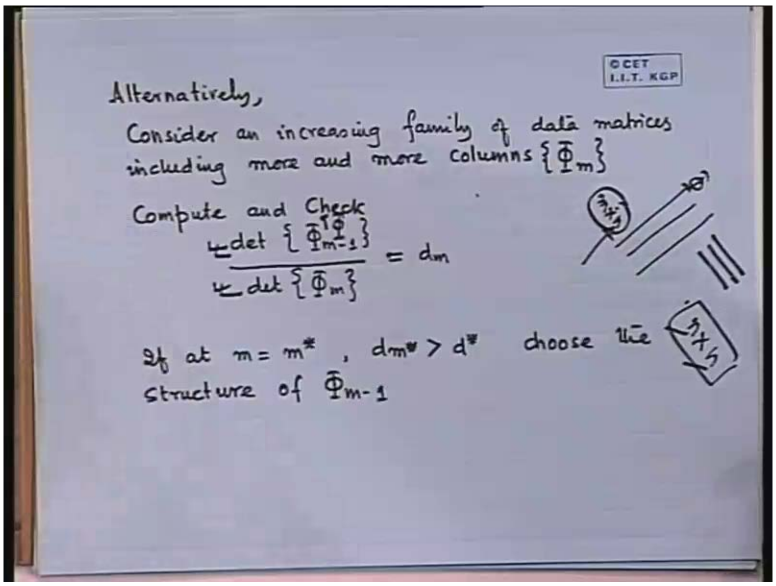
So what will happen is that is that, is now some columns become become linearly dependent on on other columns; the rank of this matrix will be less than the number of columns, so that immediately shows, that you are now getting a rank deficient matrix. So that means that so many columns are not needed; some columns can be generated from earlier columns. So you do not need to complicate your model structure. Now the only the catch is that, because of noise, exact singularity cannot be obtained because all these, exactly this will not match, even even in the measurement because of noise. So therefore exact singularity you will never get, therefore you should try other things.

(Refer Slide Time: 45:21)



So you have a singularity of a rank test of matrix is actually a very very numerically susceptible thing; you have little layer at here and there you will not get rank, as I mean you must still get a rank when I, you may get a higher rank though the matrix is singular. So in such cases what you should do is you can do two things; let me let me discuss the first one, sorry second one, first because that is simple.

(Refer Slide Time: 45:42)

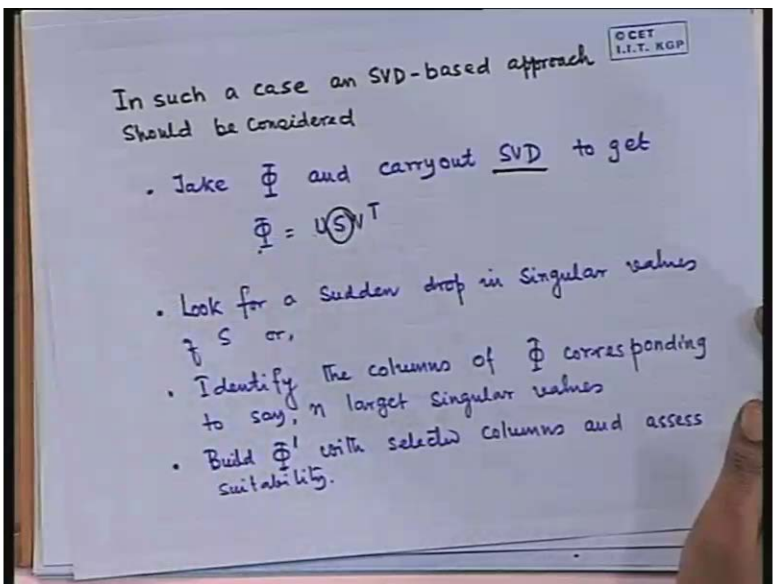


If you consider, if you construct at a determinant; actually this should not be phi, this should phi transpose phi, phi is a non-square matrix it cannot have a determinant. So if you take that, now if you have a a matrix; this is phi and this phi transpose, so therefore the overall matrix will be now rather this is phi. So you this phi transpose, so you are having phi transpose phi; so it will be if if if this is m into n phi then phi transpose phi will be n into n.

Now because the because the rank of this is less than n; so therefore phi transpose phi will not have m into n full rank, so therefore phi transpose phi will tend to become singular, it will come very close to singularity. Coming close to singularity means what? Determinant coming close to zero, when does matrix will become singular; if the determinant is zero. So if you start expanding the phi matrix at some points, suddenly the determinant will come close to zero. So now if you if you compute checking; let us say first have two columns; first order model compute phi transpose phi compute determinant, you get somewhere.

Then you choose model order test two, so you choose a phi with four columns; immediately what will happen is that the determinant of phi transpose phi with with four columns will sharply fall, it may not be exactly zero but it sharply fall because it will come will become very close to singular. So immediately that will show of this ratio. So the idea is that you go on increasing the model order, every time compute determinant of phi transpose phi; and check when when the determinant is falling rather sharply, and stop at that level, do not increase, that is a that is a simple numerical approach.

(Refer Slide Time: 47:48)

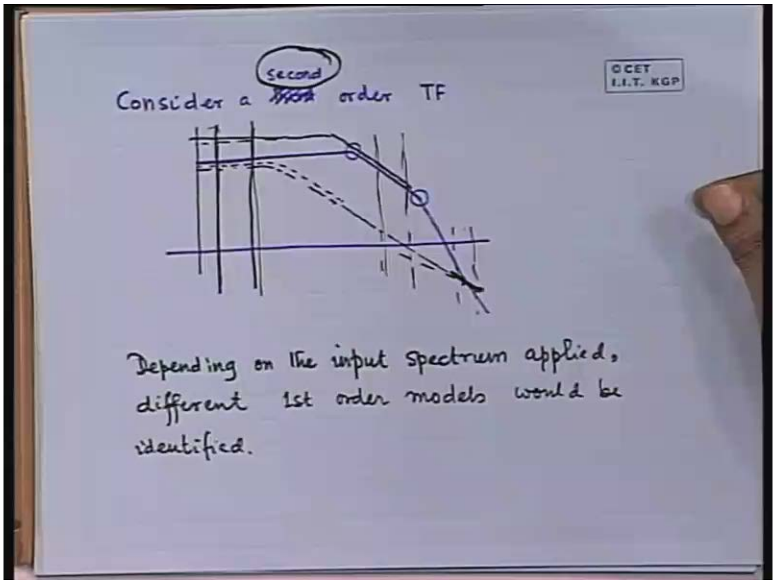


Similarly, you could do singular value decomposition rather than typically typical way of detecting; whether a matrix is very close to singularity, may not be exactly singular, if then is is to compute singular values. So you take a matrix and you do a singular value decomposition which is a very standard numerical practice; basic basically break it up into u s v transpose then this matrix is actually diagonal and its elements are called singular value.

If if the matrix is close to singular, then what will happen if that; may be suppose the matrix is phi by phi, if it has three singular values which are you know nice and good may be ten, fifteen and suddenly the fourth and fifth become point zero one and point zero two then you know that the rank of the, that the real rank of the matrix is three, it is not five. So you you check the singular values and you reject the rows corresponding to the one's which has very very low singular values; these are you know I mean, numerical procedures. So you can find a a a good idea of the order by looking at the data matrix properties, that is what I am trying to get at.

So by look by taking the matrix phi and then computing either either determinant ratio or identifying the singular values of the of that matrix; you can get a very good idea of what is the like t order.

(Refer Slide Time: 49:15)

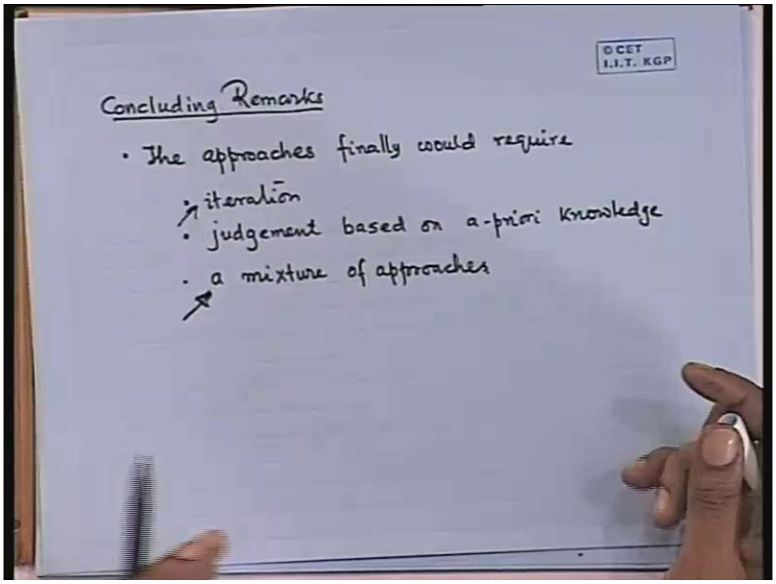


So we are coming to close to the end. This this another point I tried to make without becoming do mathematical is that; if you for example, suppose this old point that I am trying to make, if you have a second order transfer function, if you are using an input spectrum which is here then you will probably get the first order model like this. If you choose an input spectrum which is here then you will get a first order model may be like this. And if you choose it here then it will it will just try to match as close as possible around this zone; it cannot any way approximate forty degree per decade, this thing. So it will be somewhere like this.

So what I am trying to say is that, this because the true system is in second order; if you are using a low order model then the then the parameters that you will be getting will be highly dependent on the kind of input that you choose, but if you had obtained the second order model if you have taken a second order model structure, whatever input you give you would you are likely to get this form, the the the parameters will not change, right.

So that is also a that could also as a indicator, that your model structure is not same; I mean every time it is trying to match, I mean some property somewhere, so it is moving. Since since it has a you know there is a big thing, system is big and you are using a small torch light to actually know what the system is; so it is now focusing on this, now focusing on that, it cannot show you the whole thing. It is showing only that part which you are exciting, right. So these are I mean I mean I am giving very qualitative explanations because the mathematics involved with this is quite substantial, no more get into this but I want you to get the idea.

(Refer Slide Time: 51:14)



So while concluding, I just like to say that finally if you really want to do in practice; you will necessarily have to iteration, you have to this is this this is a trial error procedure, do not think you will get the data even put it in a nice algorithm and then it will say second order, second order nothing like that is going to happen. We have to do do multiple tests, you may have to do multiple tests, you may have iterated. You first time think it is second order; for another test, it might say that, no it is not the second order its first order.

And finally you may have to decide based on judgment; typical example, is suppose you find that most of the singular values are I mean, you have you know that best on computing time you can utmost choose the fifth order model, three of the singular values are, four of the singular values are nice and strong, the other two are lower but they are of the same order, which one do you reject? So so finally you will have to adopt judgment. You may not always get a nice clear answer which which which always happens in in all engineering exercises; mathematics is not a substitute for that, not yet, it helps but not always.

So that is all, we so we will in the in the next class; we will discuss what is what are the that is the various practical aspects which you must follow, if you really want to identify a model in practice, right from choosing your sensors, to doing signal conditioning, to doing and choosing a b c's all sorts of things. So we will we will discuss on that in the next class, thank you very much.