NPTEL

NATIONAL PROGRAMME ON TECHNOLOGY ENHANCED LEARNING

IIT BOMBAY

CDEEP
IIT BOMBAY

ADVANCE PROCESS CONTROL

Prof. Sachin Patwardhan

Department of Chemical Engineering,
IIT Bombay

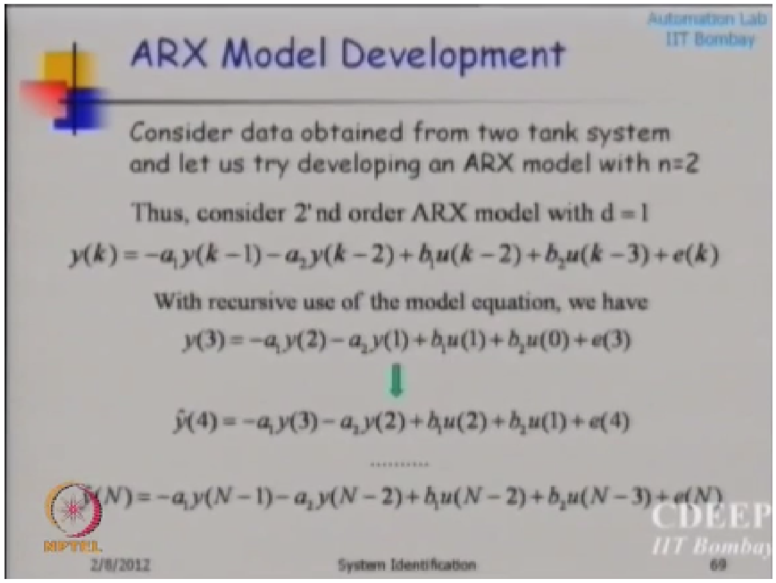
Lecture No - 09

Development of Control Relevant Linear Perturbation Models (Part 2)

Sub-topics:
Statistical Properties of ARX models and Development of ARMAX models

Okay, so let us have a short recap of what we have done. We have been looking at this ARX models, AR stands for auto regressive, X stands for exogenous inputs, external inputs. This terminology comes from time series modeling as I said time series modeling is used everywhere not just in control, so for us X would mean measured input, loan input, manipulated input okay. Exogenous for us would be input that is loan okay.

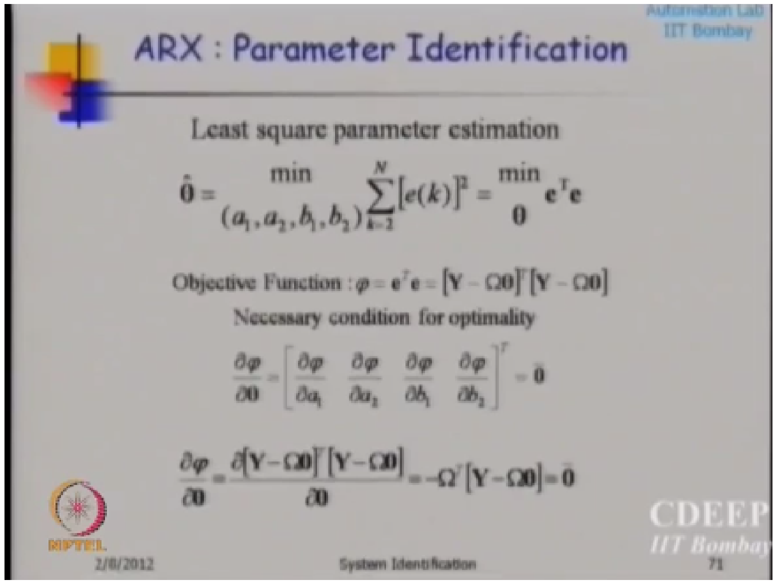
(Refer Slide Time: 00:55)



We looked at this example, I just very quickly go over it, and then I will reduct it to something that we have been looking at correlations okay. So we started looking at this model, we said that YK the measurement that instant K is some function of past measurements YK-1, YK-2 it is also function of past inputs okay. So a dynamical system very clear to see here the dynamical system as a memory.

What happens now is a function of what has happened in the past and in also function of the inputs, the new inputs that you are giving to the system. PK here we wanted it to be white noise. Well, the problem was estimating a1, a2, b1, b2 from data, we have data for Y1Y, we have a data for inputs Y and we want to estimate model parameters a1, a2, b1 and b2. So I started writing these equations and then we got this set of equations. We have large number of equations, we conducted that into matrix form.

(Refer Slide Time: 02:13)

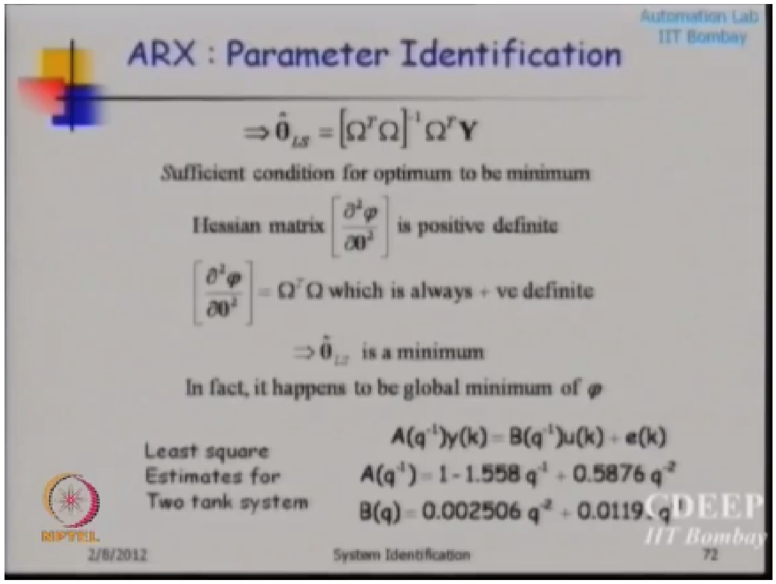


There was some error in my last time when I presented this slide, I have corrected it. So now I have this matrix ω times θ+E is the error vector, this is equal to Y, the Y capital is the measurements started in time. I have taken large number of measurements in this particular in 250 measurements. In the real problem I would take 1000, 2000, 10000 depending upon, how do I choose how many number of measurements are required. I am going to talk about it today okay.

What is the relevance of this large numbers, why do we need large number of measurements okay. So in this matrix equation this Y here, Y vector = ω times θ plus error vector. And then we developed these square estimate parameters θ. In this particular case the model is linear in parameters as I explained in my last lecture. And you can estimate parameters θ, you can find optimum value, the global optimum value analytically okay.

Is analytical solution was obtained by setting the gradient equal to 0 okay, we said this e^Te is nothing but some of the square of errors okay. We said that some of the square of errors derivative of this with respect to the parameters is said to 0, that gives us this equation. And then this particular equation is the simple matrix equation.

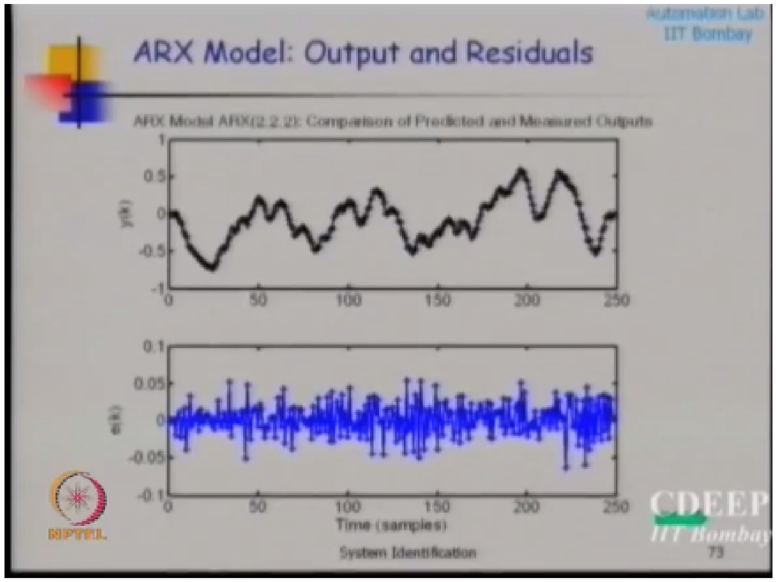
(Refer Slide Time: 04:04)



It can be solved like this well we assume here that matrix ω has full column rank, if it has full column rank then you can give out this $\omega^T\omega$ and that you can get these square estimates of parameters. A sufficient condition to know that this is a minimum use to find the second derivative of the Hessian, and the Hessian is always positively definite. So we are guaranteed to reach the minimum. So those of you know about necessary conditions and sufficient conditions for optimality this $\omega^T\omega$ is positively definite, you reach the global minimum.

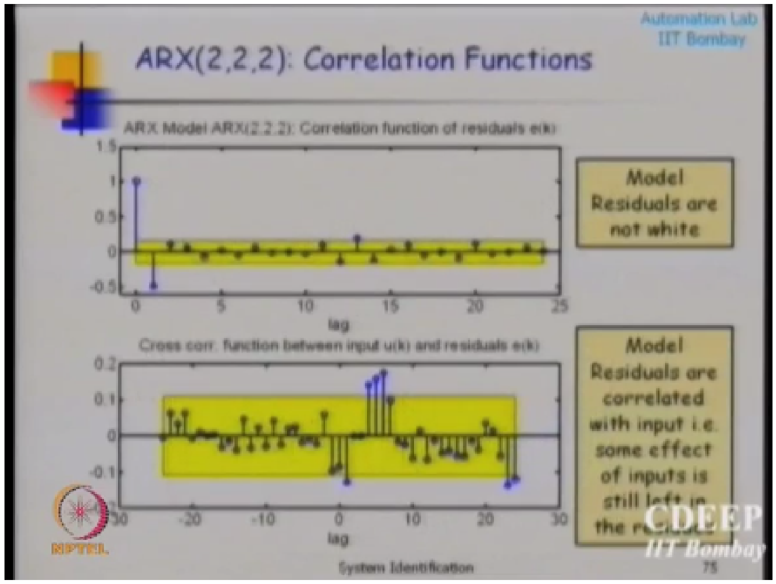
And then I found out this particular set of values from using MATLAB, I can actually construct this ω matrix $\omega^T\omega$ this give me this particular model.

(Refer Slide Time: 05:10)



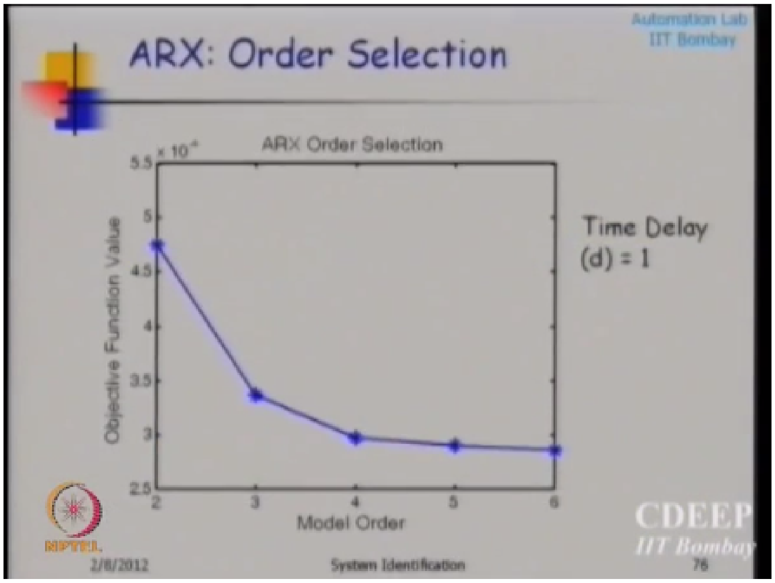
We said that this model was okay, but the error was the right, we wanted this error sequence to be white noise.

(Refer Slide Time: 05:19)



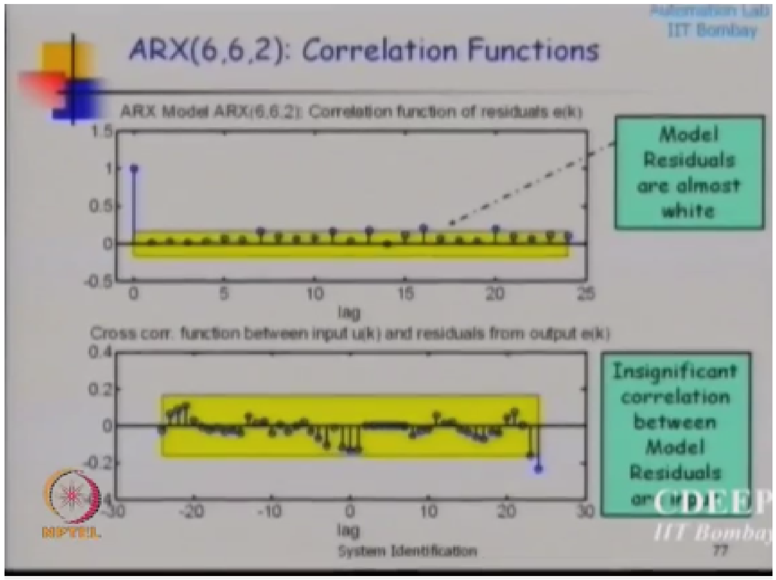
We check for the correlations, so we check for correlations, auto correlation of this error, we found that this error is auto correlated which means, you know this model is not accept with, because if you want an ARX model error Y-Y predicted this error should be a white noise, it should be complete in a way, free of any relevant signal. So we also found that there is a correlation between the residual and the inputs. There is a correlation between the residuals and the inputs.

(Refer Slide Time: 05:54)



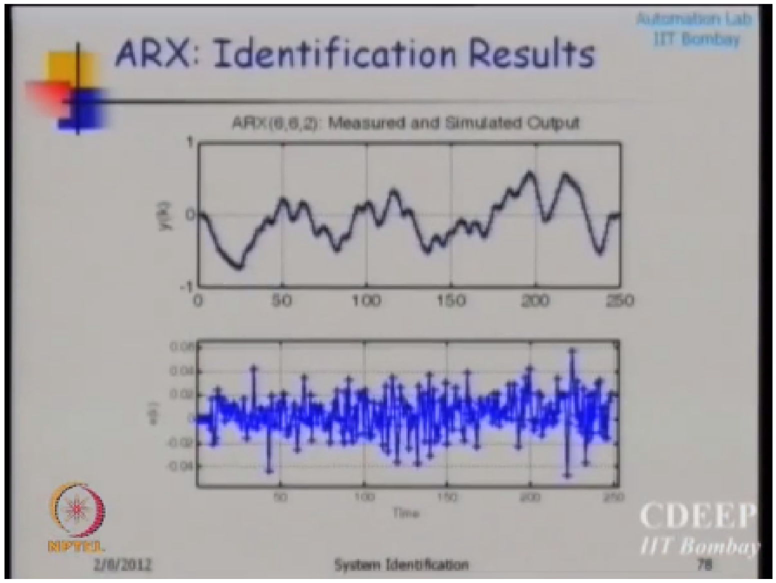
So this model is not acceptable, we moved ahead and said well, we should develop higher order models. So I develop third order, fourth order, fifth order, I realize that beyond sixth order model there is not much gain by increasing the order, so I stopped at sixth order.

(Refer Slide Time: 06:11)



And then for sixth order the innovations or errors are almost white noise sequence, there is earlier the correlation between inputs and the innovations or the errors, these errors are also called innovations.

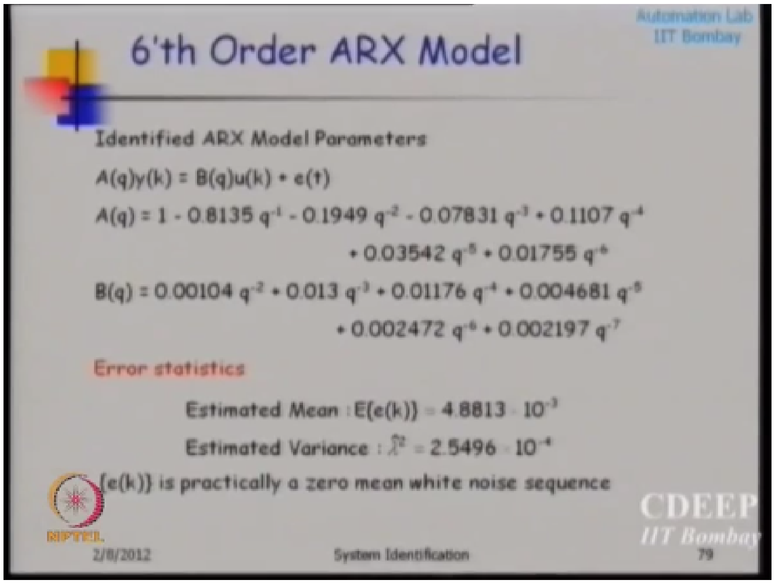
(Refer Slide Time: 06:29)



So this is an acceptable model sixth order ARX model. I want innovation to be a white noise, because I want to remove anything that is like a signal, that is like a signal means something that is correlated in the past. A white noise is something which has no relation with the past and future. It is like complete, what you will say dirt in the signal, you just started doing the third factor from the signal okay.

So there is nothing what you can call a signal left once it is white noise okay. All important dynamics of the process is captured okay, all the important dynamics. Now this has two parts.

(Refer Slide Time: 07:17)

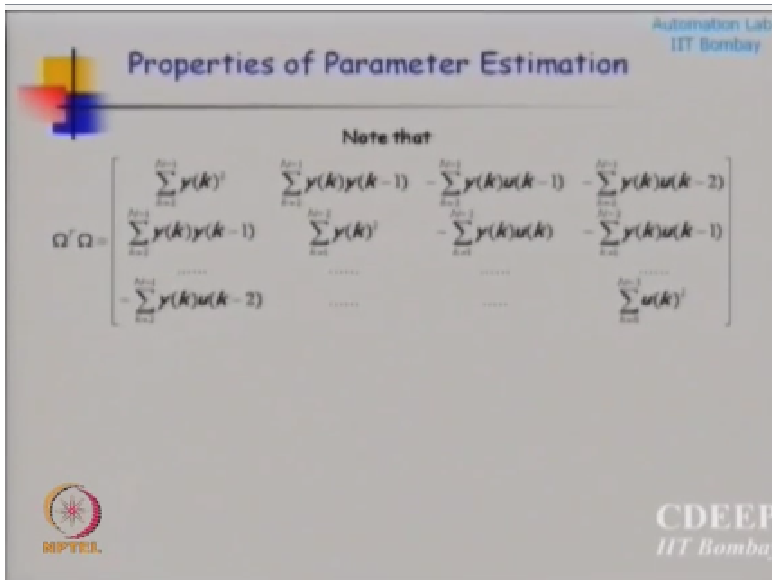


Actually when you have developed this model it is simultaneously captured deterministic as well as stochastic component. Everything that is signal okay, is captured in this okay. So I talked about new model structure, but I am going to come back and talk about something else right now okay. I am going to talk about the properties of the V^2 estimation.

And these properties of V^2 estimation are deeply related to how do you plan our experiment okay. It is not that you will see here lot of interplay between, your understanding of maths behind this and the practice. And I am going to explain that in next few slides. First thing is we begin by talking about auto correlation, cross correlation right. We talked about auto correlation between the signal, within the signal stochastic process, cross correlation between the two signals.

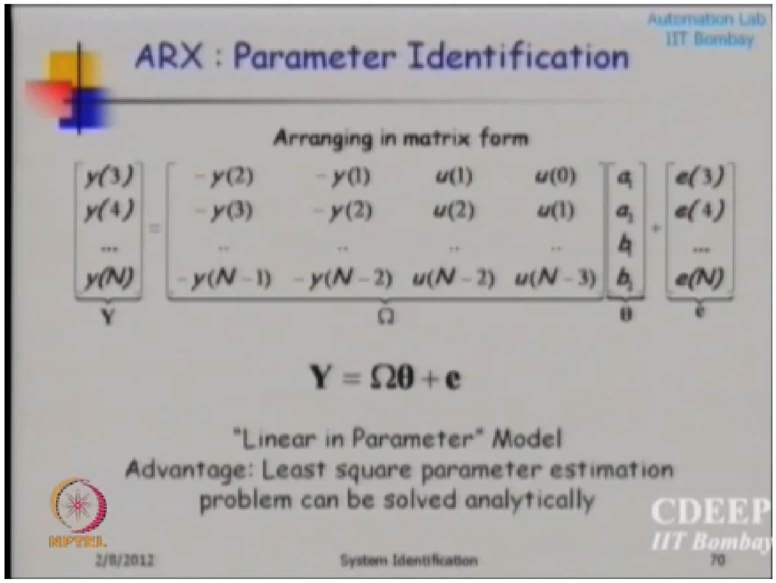
And when I developed ARX model I suddenly, you know did something else, I never talked about auto correlations, cross correlations. So is there some relationship between auto correlation, cross correlation business and ARX models. This is a natural question, you know I talked, I spent almost three hours talking about auto correlations, cross correlations. What is the relevance, it suddenly seems for disappeared when you, but it is not so, that is all hidden there and I will show you what is where it is appearing okay.

(Refer Slide Time: 08:54)



Now I am going to look at this matrix, maybe if you want just try on your notebook, you can try this.

(Refer Slide Time: 09:05)

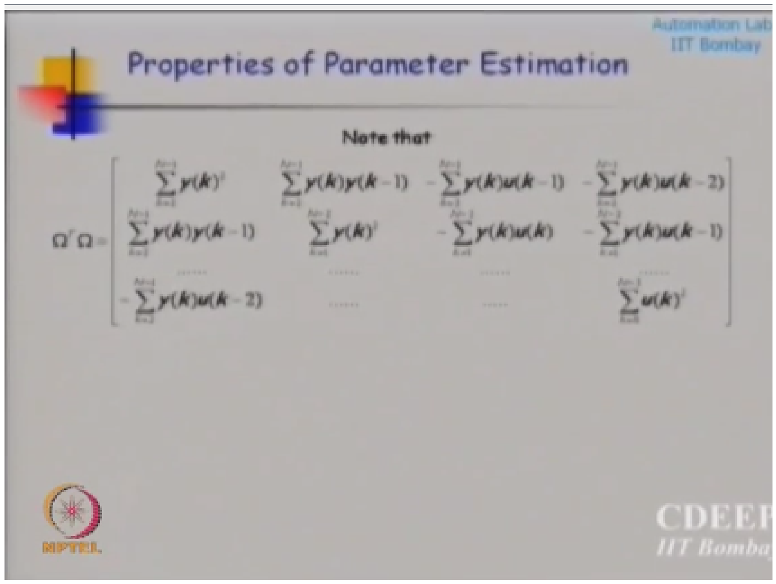


That if you, what is this ψ matrix, what is the ω matrix, just go back and check, ω matrix, let us go back and check ω matrix. Ω matrix first column is –Y2-Y3-Y4 okay, this is, what is the dimension of this matrix ω it has four columns on the right. How many rows? N-2 rows, so it is a tall matrix right, it is a tall matrix.

It has 250 rows and it has only four columns okay, what should be $\omega^T\omega$, what should be dimension of $\omega^T\omega$. It should be 4x4 $\omega^T\omega$ should be 4x4, just imagine what will be first element of this matrix, it is this column transpose into this column right. It will be this column transpose ω^T will be transpose of this column into this column. So the first column will be $Y2^2+Y3^2+Y4^2$ okay.

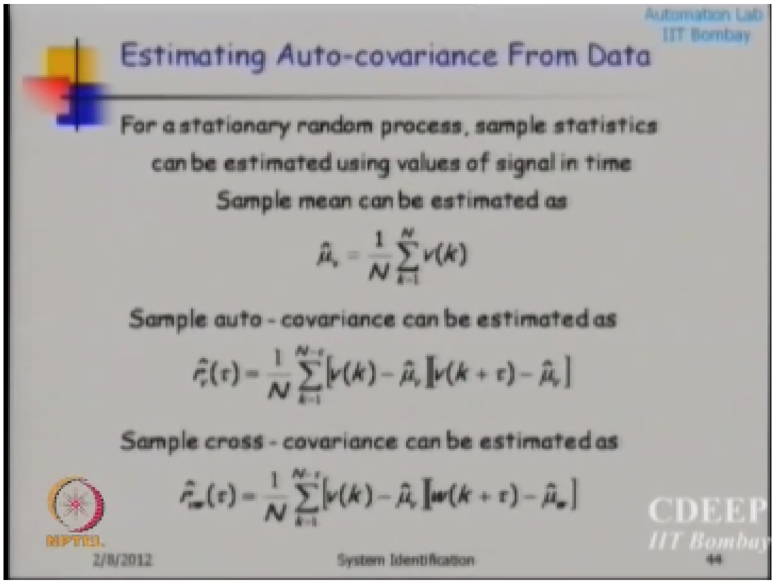
What will be the second the element on the first row, it will be multiplication of this column and this column okay. If I sit down and be little bit patient and then work out, I will get this matrix.

(Refer Slide Time: 10:40)



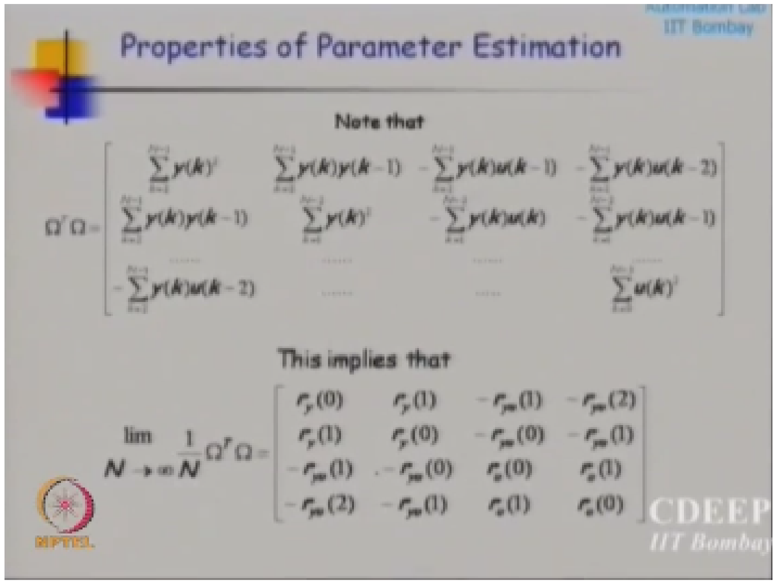
If I am little bit patient and work out these multiplications I will get this matrix, this is a 4x4 matrix. You can see here the element 11 is $Y2^2+Y3^2+Y4^2$ up to n-1 right. Likewise, I have done all possible multiplications of columns and rows okay I will get this 4 x 4 matrix does this look similar to something we talked about auto correlation and cross correlation do these elements look similar to auto correlation we had estimate of auto correlation estimate of cross correlation what is missing 1/ n is missing if you go back.

(Refer Slide Time: 11:35)



Just recall that auto correlation within a signal okay and cross correlation between 2 signals okay what are the signals here yk and uk these are 2 signals yk are the measurements uk are the inputs okays o here are called w and v, w let us say let us call v as u and w as y okay so this matrix you will realize if you correlate with this particular slide.

(Refer Slide Time: 12:14)



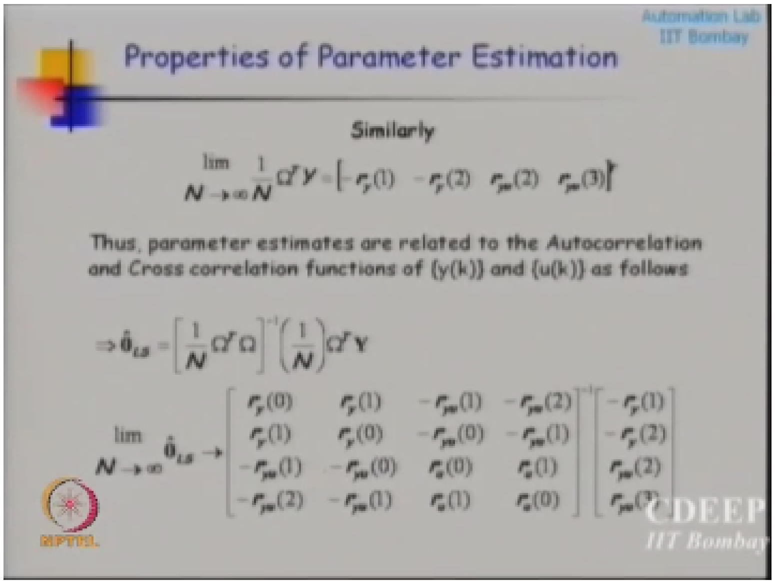
And the go back to and look carefully at all the 16 terms in this matrix you will see that this is very close to the definition of auto correlation cross correlation except 1/n is missing okay 1/n is missing see in cross correlation I will have this term 1/n into the first term will have 1/n here 1/n n is or 1/n – 1 whatever is the relevant number you will have that particular number coming up what I can do is I can multiply this matrix by 1/n and take limit as n goes to infinity okay.

What you will see that this matrix actually consists of auto correlations and cross correlations this matrix actually consists of auto correlations and cross correlations just look carefully is everyone convened about this slide see all that I am doing is am taking this matrix multiplying by 1/ n and they think n go large value I am getting n go to large value for large n I can approximate this matrix if I take n to be say 10000 okay.

This particular matrix will approach auto correlation 0 of y auto correlation 0 auto correlation 1 of y okay cross correlation 1 between y and u cross correlation 1 and 2 between y and u and so on this is one more thing you will notice this matrix is symmetric you always symmetric okay.

So in 4 x 4 matrix \how many independent elements are there say it is symmetric matrix not 4, 12 do not know 10 so this upper triangular 6 + 4 diagonal so there are 10 elements which are it is a symmetric matrix so first of all auto correlation and cross correlation is very much appearing when you are actually identifying this model okay.

(Refer Slide Time: 14:39)

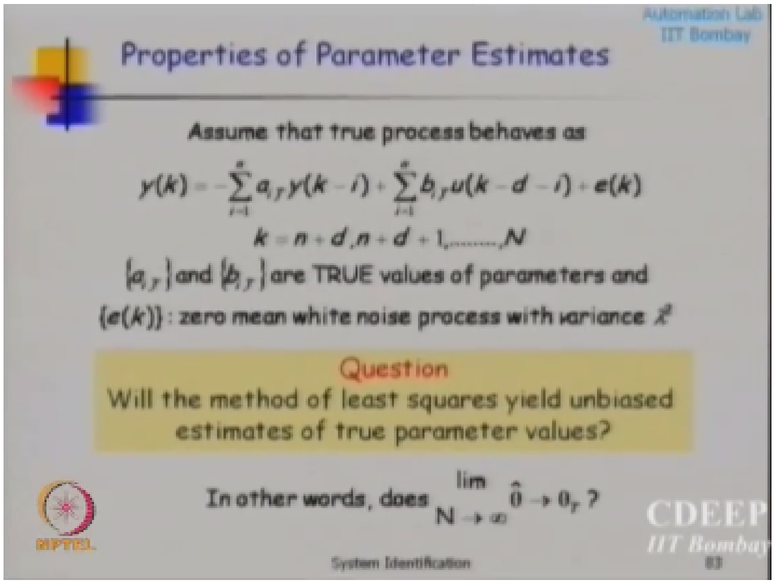


Next part is that what about you can also show that Ω transpose y okay I can pre multiply that by 1/n and actually Ω transpose y asymptotically tends to auto correlation 1 auto correlation 2, cross correlation2 cross correlation 3 yeah why is 0 that is appearing all the 0's because you will get σy^2 will be identical actually it is only time shift see you will get here y y^2 okay going from 2 to n – 1 this is y^2 from n2 n-2.

Okay when you get n go to infinity this actually varies which one in this particular case they are not so we will actually fills still have lower dimensional yeah okays so now this θ least square I am rewetting do you see this what was θ least square Ω transpose Ω inverse okay into Ω transpose y I will just insert this 1/ n okay and that gives me a completely different interpretation okay that gives me interpretation from the view point of auto correlation and cross correlation.

So these coefficients of this model which I am getting are actually deeply correlated to auto correlation and cross correlation business okay coming from auto correlations and cross correlations it is not all that I talked about between 2 stochastic process you know auto correlation and cross correlation all that is at work here that is what I want to show okay so we have actually used also those concepts while deriving this model okay.

(Refer Slide Time: 16:45)



Next fundamental question okay let us say you know there is a true process in course which exactly behaviors like this see when I am using a method to extract parameters from data okay the question that I should ask is this method correct will this give the correct parameters see because I have a process I have system I am collecting data and I am using some least square estimates I am using some method to estimate the parameters.

What is the guarantee that estimates are close to the truth okay now forget about the real world where the real world is actually non linear when you develop a model for real world system arx model is an approximation forget about all that let us say you know I am writing a programming mat lab okay I am creating data suing this equation arx model okay.

I created a data and I gave it to and I asked you to construct the model I told you this of this is 2^{nd} order process which I have simulated the white noise is like this and so on I give you all the information I just give data I just tell you that there are 4 parameters okay question is will list where give you that correct values the true values is a valid question because if my method does not leave me true values.

My method is useless is giving to give me wrong values then you know I am going to get a wrong model if I am going to get a wrong model I cannot use it for control right so first thing is to check weather if a method make sense okay I want to actually assess whether this is method itself is good or bad so if I do a ideal world computer experiment in which I generate y and I give some known input u okay.

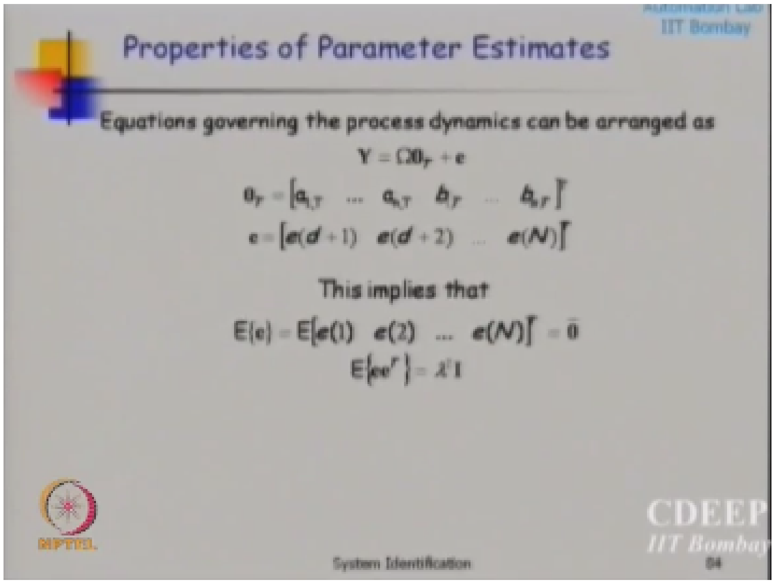
I give some known you know white noise sequence and I generate y and there are some true values of this parameters okay suppose some point 9.155.4 there are some true values will least square estimation give me 0.1, 0.2, 0.4, 0.9 is the question if it does not this method is useless okay so we have to ask this first question okay in statistical terms I am going to say that will this method give me un bias estimates of parameters oaky what does it mean.

What is a unbias estimate when I am getting an estimate we said that estimate is a random number okay estimate is a random number so I am never going to get exactly the true value okay so what is the meaning of unbias estimate okay what I want to know is this if I have start collecting more and more data points okay I write now collected 250 data points if collect 1000, 10000, 50000, 100000 okay as enters to infinity will the estimates tend to the truth the ia m guaranteed that this question is correct okay.

That is a first fundamental question I am going to ask okay so will the estimated value tend to the truth okay now right now we are again let me remind you right now we are in a ideal world where we have created the data okay we know that ek is a white noise we know what are the true parameters I mean I know and I have given you data and I asked to construct estimate model parameters using least square if you want if you want 1000o points I will give you 10000 data points.

If you want 1 million data points I will give 1 million data points okay the question is whether your estimate will tend to the truth okay that is the question I am going to ask.

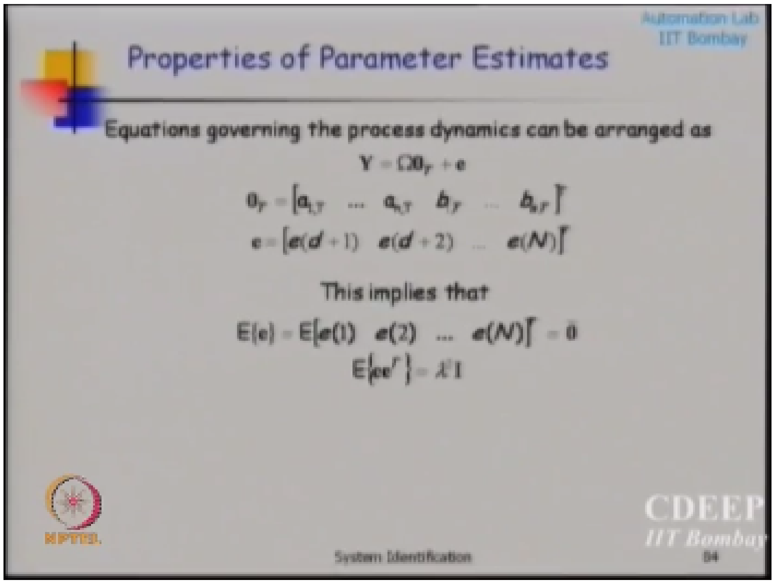
(Refer Slide Time: 20:54)



Okay some preparation for doing this I have rewritten as a so my true plant my true process which is computer simulated right now okay behaviors according to this matrix equation okay y as all the measurements okay then Ω as all the past state measurements y and u θ true are the true values okay and e is the white noise sequence which we have introduced into the system okay so I am writing this θ true okay.

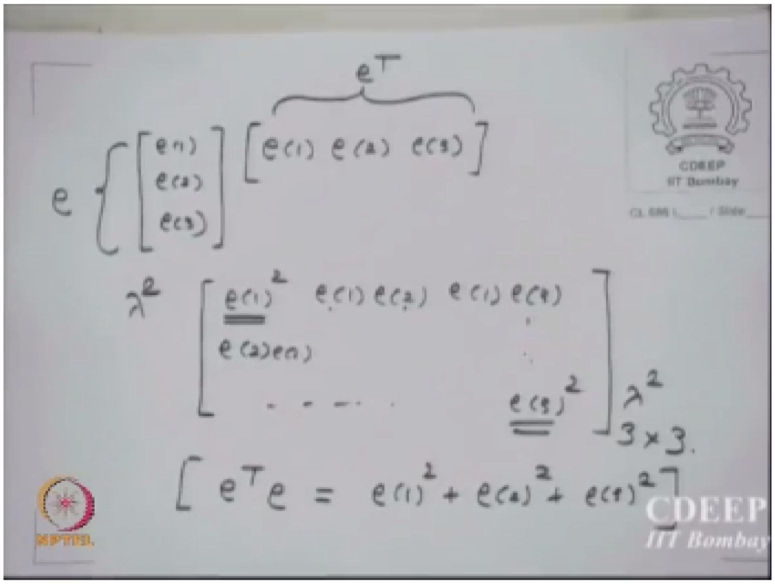
What is e is white noise sequence 0 mean Gaussian white noise sequence with standard deviation $= \lambda^2$ okay it is a white noise sequence so this e this capital E vector oaky consists of you know all the 0 mean white noise numbers okay, what is the expected value of this E it 0.

(Refer Slide Time: 22:14)



If you take e, e^t what is ee^T will it be a vector will it be matrix, will it be a scalar, e^Te will be a scalar ee^T will be a matrix let me just move here and show you.

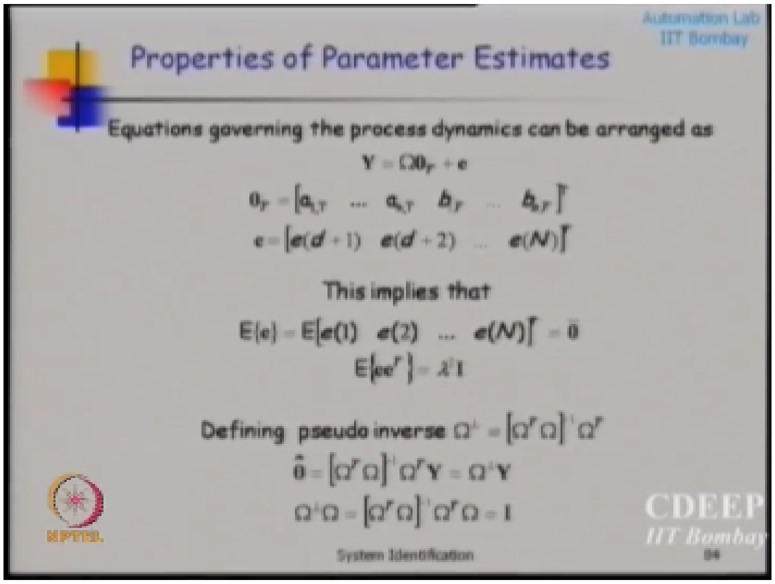
(Refer Slide Time: 22:33)



See this, this is my e vector e^T is so this is e^T this is my e vector so if I take ee^T I will get $e(1)^2$, e(1)e(2), e(1)e(3), e(2)e(1) and so on, do you agree $e(3)^2$ this will be 3x3 matrix do not confuse this with e^Te, what is e^Te that is $e(1)^2+e(2)^2+e(3)^2$ this is inner product eTe is inner product eeT is outer product okay, eeT is outer product. Now if this is a 0 mean white noise, if this is 0 mean white noise, if this is a 0 mean white noise okay, what is expected value of $e(1)^2$ $\sigma^2\lambda^2$ what is expected value of $e(3)^2$ λ^2 what is expected value of e(1) e(2), 0, okay.

e(1)e(3) is 0 and so on okay, so that is what I have written here that expected value of eeT is nothing but λ^2i which is just my preparation for you know coming up with the properties.

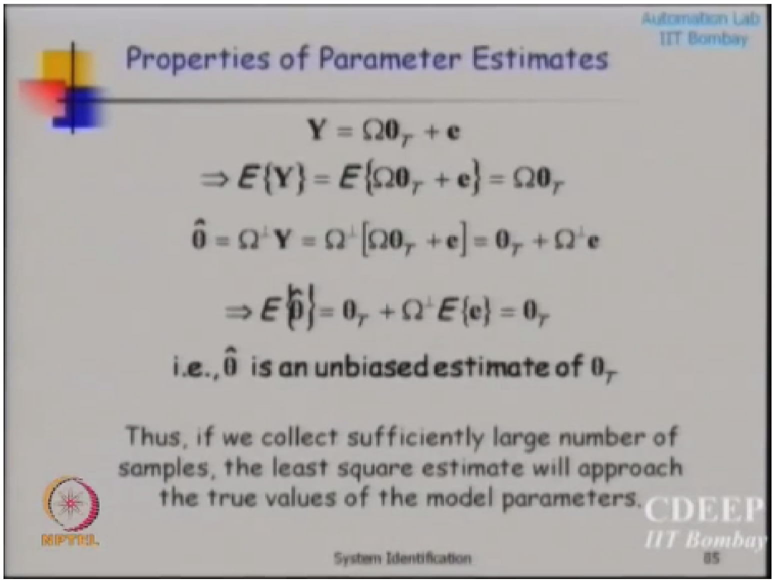
(Refer Slide Time: 24:32)



Okay, to make the notation simple I am going to define this matrix ΩT, Ω-1, ΩT as $\Omega^{\perp}$ here okay, this is only a notation because later at some point I am going to get multiplication of these matrices and it becomes very compressible right, so this is only a notation okay, notation that I am introducing that is ΩT this not transpose this $\Omega^{\perp}$ is nothing but ΩT Ω-1x ΩT what happens if I post multiply this by Ω what happens to this matrix I will get identity matrix.

So actually this $\Omega^{\perp}$ is called as left inverse of Ω, Ω remember is a τ matrix it has fear columns than rows it is a tall matrix okay, so it does not have normal inverse which has either left inverse or right inverse so this is a left inverse of this matrix so this is just a notation, okay let us move ahead.

(Refer Slide Time: 25:48)



Okay, so my model is actually given by Y= Ωxθ true okay, the data is generated by this equation right, this is how the data is generated plus error. If I take expectation of Y okay, if I take expectation of Y then I will get expectation of Ωxθ true +e what is the expectation of e, 0 okay. so this is expectation of Y is nothing but Ωxθ true okay how this is going to help us. what is θ, θ is Ω^+ Y, then what is Y, Y is Ωθ true.

Yeah, I will just these square estimate it is understood that is least square estimate to make notation simple I have removed a less okay, we are going to talk about least square estimate in this course mostly then I am going to talk about other one norm, two norm, one norm infinite norm and so. so what is Ω^+ Ω identity matrix okay, so you get this θ least square estimate is nothing but θ true plus Ω^+ into error, okay.

Now I am going to take expected value of θ what is the expected value, suppose I do you know hundreds of experiments okay numerical experiments. From each experiment I estimate θ I can do that okay, I can generate data from computer I can give you each one of you what data set and ask you from the same process, from the same stochastic process I am generating datas each one of you generate a θ okay, then I will take a mean of all of that okay, I can take a mean of all of that.

The mean should tend to the truth right, mean should tend to the truth that is what I mean here if I take expected value of θ what should I get expected value of θ is θtrue+ Ω^+ into expectedvalue of e, what is the expected value of e 0, so this simple analyze tells me that Li^2 estimate is going to give me the true values if I collect large number of experiment, large number of data points, okay this analyzes tells me that if I collect huge number data points least square estimate will tend to the truth. At least in the ideal world where you know, we know the true parameters okay, this is how this question we are asking about you know goodness of the method itself of a estimating θ, okay is this clear, is the idea clear.

Any doubts at this point yeah, yeah I have right now I shown that you know perfectly the order okay, you know perfectly the order I am giving you this information that this data is collected from stochastic process of order four that is you know Y order is four U order is four I given you this, okay. Now the question is suppose you know perfectly you know order but you do not know either they are A1A2xB1B2 you just have data for Y and U okay. from data will you get the truth okay, see if that itself is not happening then you know you should change the method okay, so now I am guaranteed that even though I will not get the truth from one experiment or ten experiments if I conduct large number of experiments I will tend to the truth that much I am guaranteed.

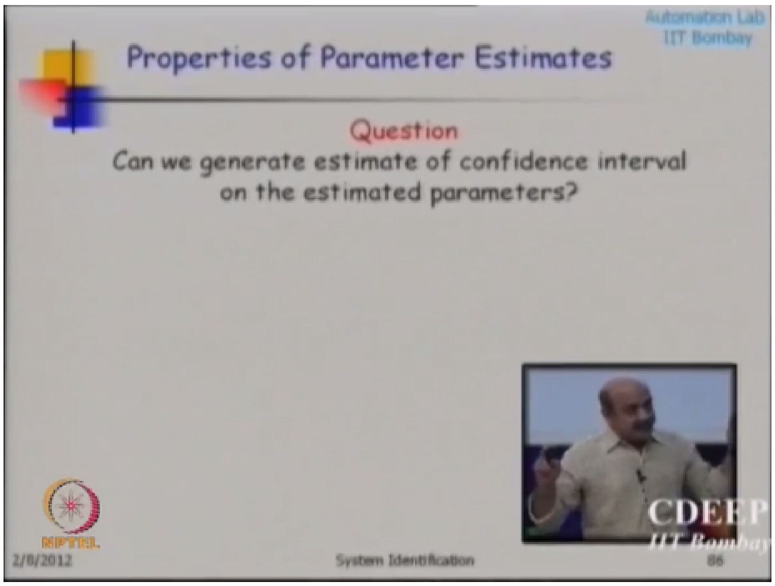
So this method is a good method okay, if I patients to do you know large data collect large data now in reality collecting large data has some other implications. See you know this mathematical truth that you know if you collect large data you will get a good model okay, but if you want to collect large data you have put up your system your plan for a longer time. If you are put up in your plan for a longer time it means you know you are disturbing the production okay, if you disturb the production okay, then your company will say well to get your model you know your disturbing may applied, okay.

So you have to convince your management look if you want a good model I need to put, but U as a control engineer has to strike a balance and say well I need significant amount of data points at the same time I should not cost too much loss economic loss to my company okay, so this understanding of this particular thing is deeply related to where you are going to conduct an experiment okay. so this analysis is not just mathematical analysis you know it tells you that you know well there is a trade off.

If you collect bigger set of data if you put up the plan for a longer time you can develop a good model at the same time you know that if you put up the plan for a longer time then you loss of production, so your do you know strike a balance as to. Yeah, we should actually so remodeling, re-identification when do you will re-identify you know how much time you put up all these things are I mean golden questions.

If you know how to answer them as a modular you will be very, very highly you know. So one thing I am guaranteed that if I collect sufficiently large number of data points the estimates will approach the truth that I am guaranteed about of least square method.

(Refer Slide Time: 32:20)



At least for air x model okay, well you can show this your for the other models. The second thing is suppose I am not able to do this, so you know I am not able to collect a millions of data points I am going to collect 1000 data pints or 250 data points or whatever I am going to collect small number of data points okay, then what I know is my estimates are you know not close, not equal to the truth they are slightly away from the truth.

I would like to know how much away they are from the truth, what is the confidence interval, what is the bound, what is the error bound on the estimate of parameters that is what I going to know, okay. So in statics terms can we estimate parameter interval of you know confidence intervals on the parameter estimates way to do this is to find out the co variance we find out the mean okay what is the second moment. We showed that mean of estimated parameters will tend to the truth okay next thing I am going to do is to estimate the co variance of the estimated parameters, this will help me to put bounds on the estimates so I want to find out this co variance matrix.

That is θ estimated { - θ ^true okay in to θ – θ ^ transpose I want to find out expected value of θ^ and a diagonal elements of this matrix will tell me individual variances for each parameter estimate okay there is some algebra which is I am going to leave it to you to figure out there is nothing it looks complex it is very simple matrix multiplication okay, okay what is θ^ - θ true that will be you know it is Ω pert in to y – θ true okay, which is if you just do a little bit of algebra you will realize that this is nothing but well I will just go back I show you this we have this expression here okay that is θ^ = θ true + Ω pert e okay same thing I am writing here.

I am just saying that θ^ - θ true is Ω pert t okay is the same thing which I have written here okay, so what is this vector in to this vector transpose will give me Ω perl in to ee transpose okay if I take expectations of this Ω perl is a matrix ones you have data it is constant matrix okay co variants of θ will be Ω perl in to expected of ee transpose okay which is nothing but λ i2 we have already work this out okay.

So finally if I do this multiplication so this is a scalar I can move it out this is identity matrix okay so actually it falls down to Ω pert in to Ω pert transpose okay and then you can show that this Ω pert in to Ω pert t is nothing but Ωt inverse this identity you can work out if you are not convinced with this go back and work out. So the point is this last equation it tells me that co variants of θ is λ2 in to Ωt Ω inverse okay Ω t this is my co variants of θ.

There is a trouble here okay it involves variants of e what is e here? Why it is not okay if I had given you information about variants of e see I am generating the data if I had given you that variants of e is one or whatever 0.1 you will be able to estimate this co variants because I have given you data for y hat u from data you can generate Ωt Ω this matrix you can generate you can generate this inverse this is 4 + 4 matrix by the way this matrix is nothing but all those auto correlations cross co relations that we looked at right, so if give you this information about λ2 you can generate those co variants in reality we do not what is λ2 so we need to estimate even λ2 okay how will you estimate λ2.

What I am going to do is I am going to find out e ^ estimate of e okay which is y measured – Ω x θ hat, okay this is give me e^ estimate of e not the true e this is an estimate of e when we will look at true e if you collect large number of data point if you let n tend to infinity which is not going to happen actually so you will get an estimate of e, if I get estimate of e vector I can find out its variables, this is the expression for finding out the virginals of okay there are d there are d just go back and check y have written here it should be p here there is a small error.

There are p parameters to get unbiased estimate you need just correct in your notes well y unbiased estimate require 1/ n – p you have to go back and look at your basic statistics book I am not going to explain that, if there are p parameters you have to divide by 1/ n – p to get correct estimate okay. so which means I can generate co variants estimate of n – p okay using this expression this is my final expression fro estimate of parameters if you see on the right hand side every ting is known to you Ω matrix is known to you it consist of data okay Ωt Ω inverse is known to you right.

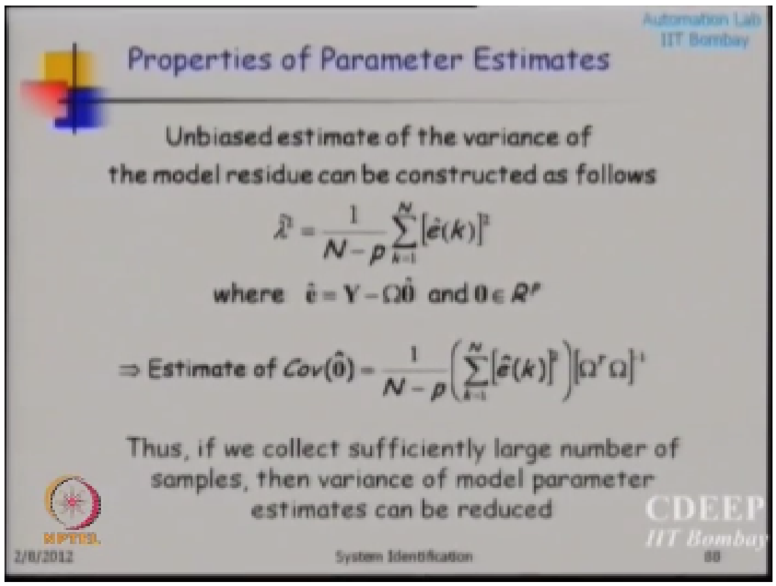
Then innovations e you have calculated from that you have calculated this variants which you are multiplying here okay, all that I have done is I have replaced λ 2 / λ^2 you realize this, what is this λ^ estimate of λ, λ2 okay. Is this expression clear? Okay now I want to do interpret this expression. I will show you the interpretation, you tell me what are the interpretation? How will you interpretative? Somebody should think about this not just look at it why I am writing this, how will you reduce co variants?

Number of samples increases the number of samples with relation to what? For the given set of parameters if you take if you have a model with two parameters okay then probably you can sue 30 samples but if you have model with 10 parameters you know you should use in relation to 10 parameters you should use larger number of samples this tells you that if you have larger number of parameters you need larger number of samples okay then also tells you that variants in the estimate becomes smaller and smaller if you take more and more data points.

The estimates of parameters will become sharper and sharper if you take more number of data points okay, but as I said there is a trade off. The tradeoff between getting good parameters and perturbation the plan for a longer time okay, so this is very crucial decision when you actually perturbation plan you know you should know the mats how long should I pert up.

Why when should I stop you have to taken to consideration the you know need to get more data as a modeler as a same time you have to consider your company is you know should based product or should not disturb the plan for a longer time. So you have to strike a behave and you have to strike the balance you have to understand this equation of variance and mean okay.

(Refer Slide Time: 41:33)



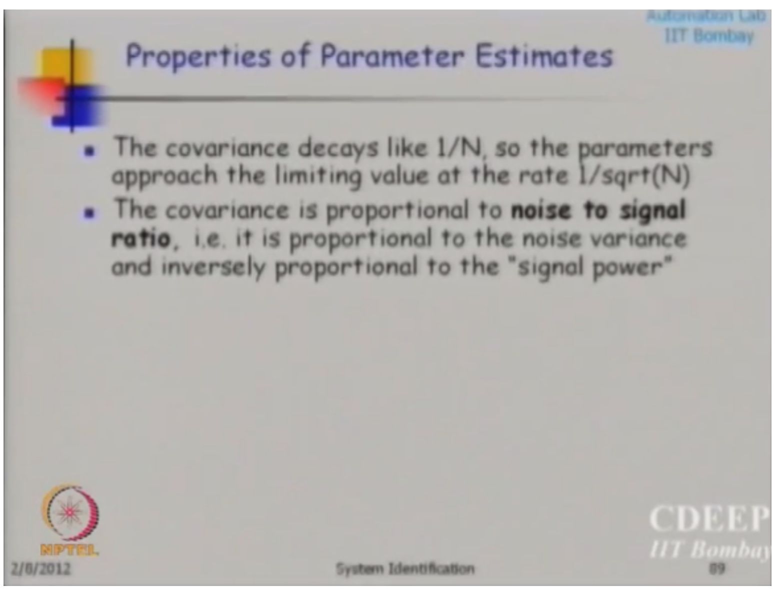
So basically if I increase the number of data point I can get better and better estimates now we develop two models you remember you develop output error model output error model was giving me good estimate or deterministic component of using how many parameters? Only four parameters, a 1 va2 b 1 b2 for ARX model how many parameters we required 12 parameters 6 + 6 12 parameters so almost three times when I went from oe to ARX I needed 12 parameters so the data points I need to you know roughly I will tell you roughly if I take 100 data points for output error model I would need 300 data points to get a similar positive model for ARX model.

So ARX model is bridging that gap giving me a good model but there is a cost it is not free there is no free ventures, okay you want to good model you better pay for it by probating the plan for a more time okay. so now you know my search is for a model which has less number of parameters but will also give me good noise model okay, so I want you know to look for a model which has less number of parameters, what is the significant of less number of parameters I have a pert for smaller time I can save my companies money okay.

So ARX model is good trouble is you need large number of parameters in real industrial problems ARX model would require something of the order of 80 to 100 parameters to get good model between one input output okay large number of parameters are required to get good ARX model it is very easy to develop just these squares okay very easy to develop. But the numbers of parameters are large number of parameters are large means you need large data set okay and if you happen to use ARX model without knowing this fundamentals that means if you conduct a experiment for a shorter period of time okay develop an ARX model of high order you have committed large errors in your model, please understand this oaky.

So unless you know this math's you will be not be able to appreciate because map lap will give you some numbers you know you give if you have to estimate you know 100 parameters if you give 200 data points mat lab will give you some numbers but that number might be meaningless because there will be large errors reduce the errors.

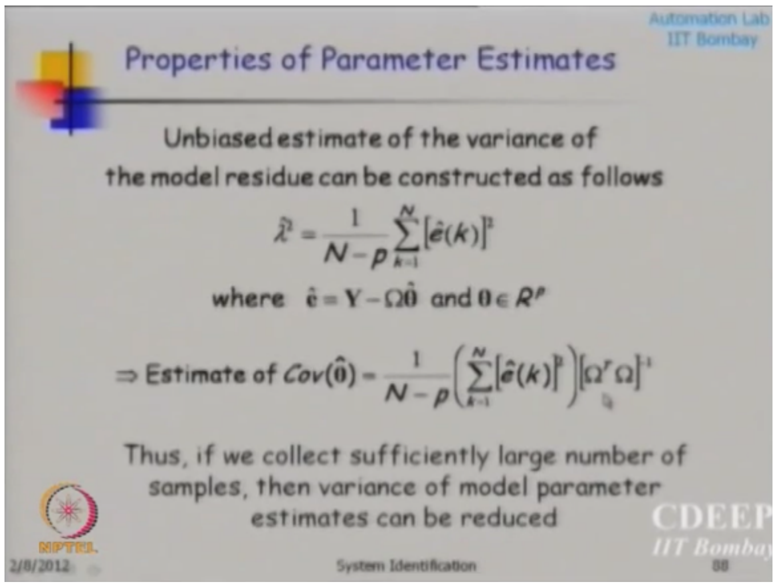
(Refer Slide Time: 44:27)



You have to increase the number of the data points okay so the properties the first properties is that the covariance decays as 1/N so the variance of each parameters actually decays by square root of N that is the first lesson to learn when N becomes large you can neglect that p which is the appear in okay.

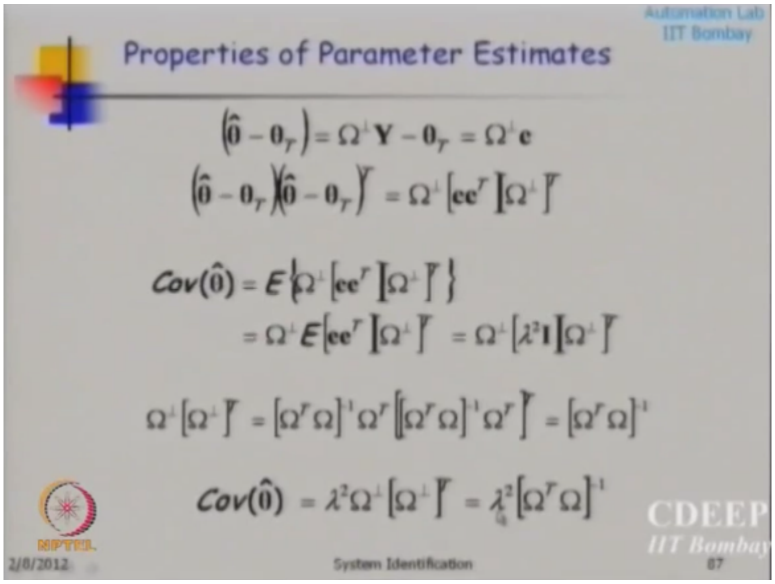
And this variance is also proportional to what is called as signal to noise ratio okay this thing is going to haunt you if you are going to continue into a system identification signal to noise ratio signal to noise ratio is power in the input that you are to part to be okay which part of this is a talking about power of signals.

(Refer Slide Time: 45:12)



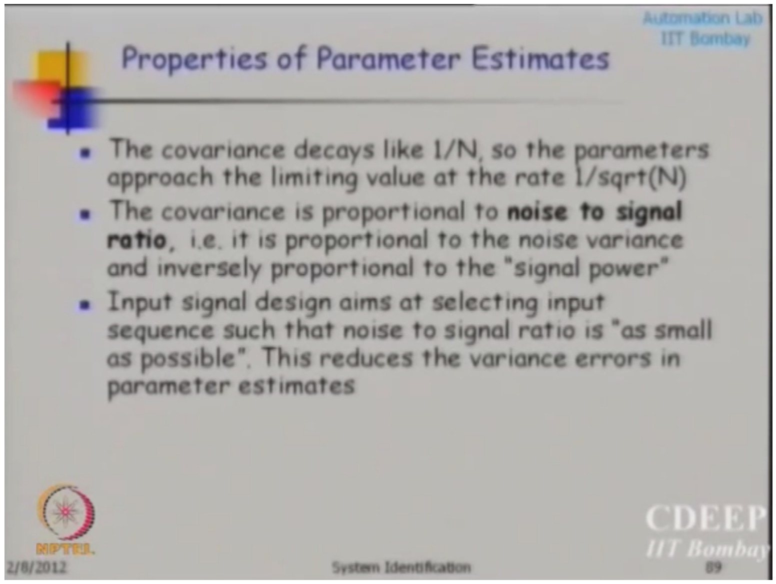
In this expression one is this N what about this guy Ω transpose Ω how can you reduce variance that you make Ω transpose Ω inverse as small as possible see there are two measure variances I want to play with one possibility okay but this Ω transpose Ω there is an information about signal you can note this particular model strong let us understand we talk about parameters so make it in such a you chose a matrix in such way that this ratio you chose this matrix,

(Refer Slide Time: 46:31)



You chose this matrix in such a way that λ^2 multiplies this Ω in difference small so I can play with other vectors and modify properties of Ω okay change signal and you take this as noise and this is the signal competence and it is okay.

(Refer Slide Time: 47:04)



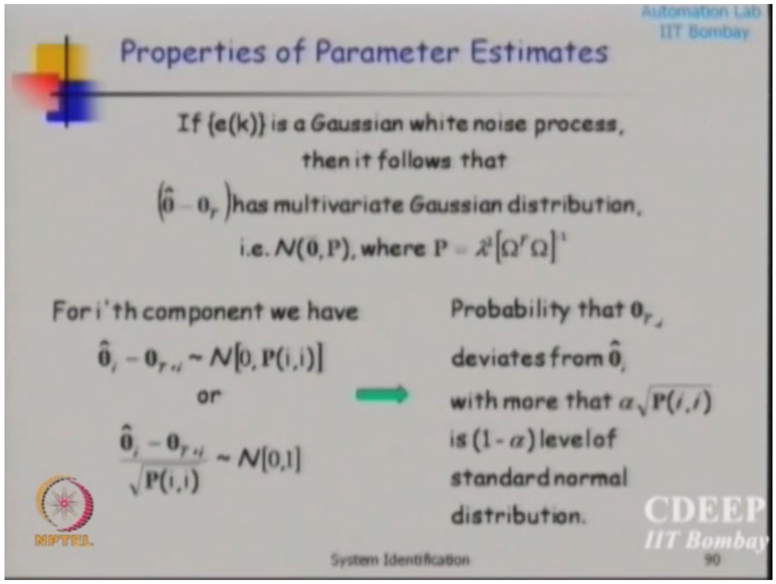
See actually, actually in the real system when you have to identify model from data how do you plan it becomes low you have to low physics of the problem okay and then gently introduce that there is you know it is like you know see doctor when we wants to let so say some stomach problem and then nowadays you put some endoscope it painful for you.

But to get an information how much okay so depending upon fluctuations conditions what you judge unless you cannot moderate okay so it has an implications how well would be part okay how does the things re activate okay and where it underline mass you cannot use mat lab package to develop models and mat lab what you realize and you develop the model.

I want this ARX model to9 flash to develop a model unless you understand they measure the ARX model better collect what have we have said okay so I have to you know well there is a good noise to signal ratio you have to increase the signal properties which means you are integrate the quadrapations of the model if you introduce rather quadrapations plan gets collapsed you know there is no disturbances in the model okay.

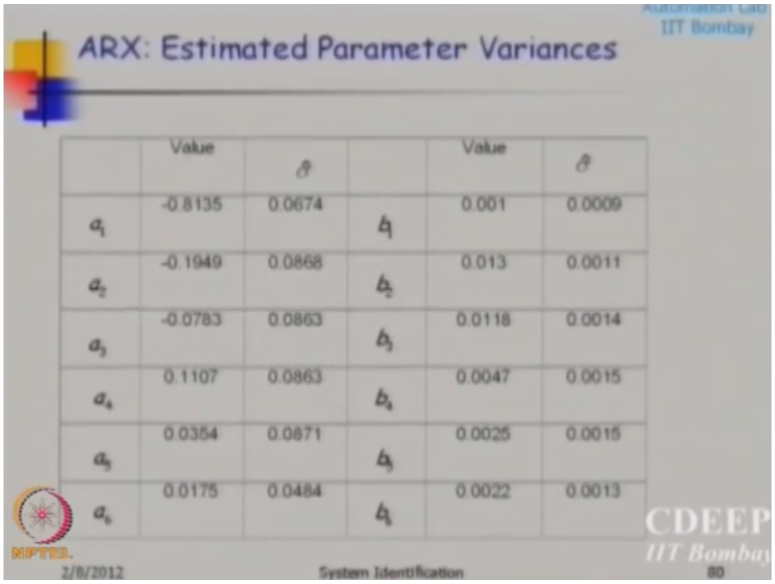
So again in fact to get a good model you are get to quadruple more if you quadruple more you are disturbing the quantity okay you are disturbing that plan to go so we have to understand Maths is important so managing you know systematic properties.

(Refer Slide Time: 49:46)



Well how given here somehow to estimate the Gaussian white noise process so assume that you can take the identical of the model and calling this matrix p here p is the thing but λ^2 in the inverse here or the λ^ whatever it is from that we can estimate for the each parameters and so actually mat lab will give you all this I will show you mat lb will give you confidence parameters.

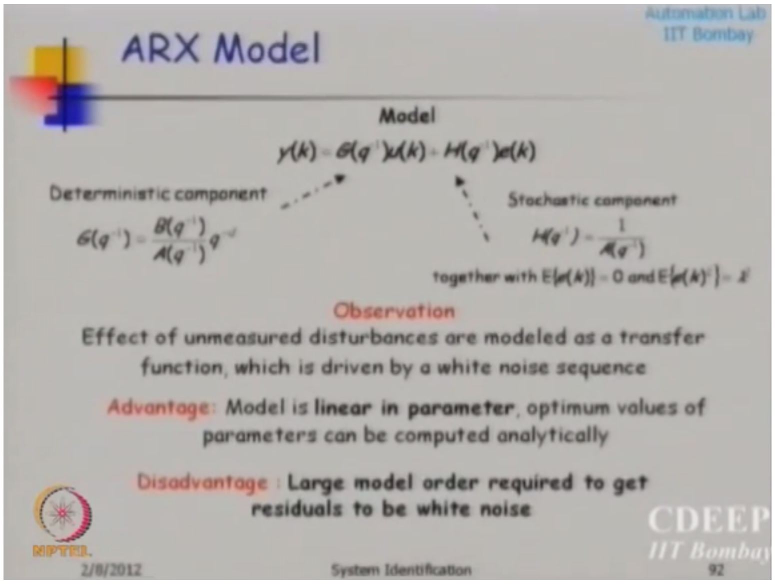
(Refer Slide Time: 50:47)



	Value	σ		Value	σ
a_1	-0.8135	0.0674	b_1	0.001	0.0009
a_2	-0.1949	0.0868	b_2	0.013	0.0011
a_3	-0.0783	0.0863	b_3	0.0118	0.0014
a_4	0.1107	0.0863	b_4	0.0047	0.0015
a_5	0.0354	0.0871	b_5	0.0025	0.0015
a_6	0.0175	0.0484	b_6	0.0022	0.0013

Mat lab will give you sigma square on this parameters okay and mat lab using model okay so this estimates are allowable for the model okay let us move on to okay.

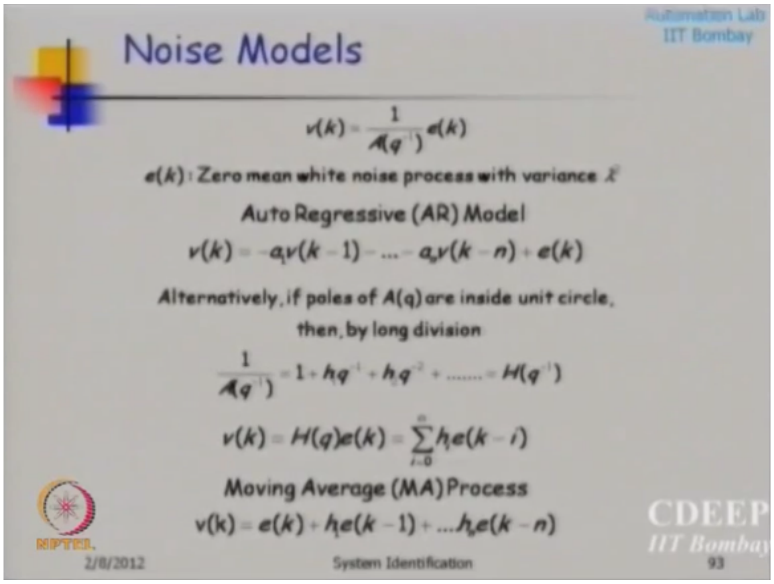
(Refer Slide Time: 51:29)



So could we said that after developing ARX models we have this known of model which consists of two components model that is disturbance of the component because it is available for you manipulate the variables which are known numbers okay we have one more component which is stochastic process is stochastic problem is this model ahs to be stochastic as a combination of two things model is entered okay and available in the innovation sin it these two are jointly estimate the variables okay.

So effect of unmeasured disturbances are modeled as a transfer function which is driven by a white noise sequence this is something which I discussed in last lecture at the end of the last lecture.

(Refer Slide Time: 52:38)



And so in the last lecture if I put this u take as 0 then I get this model okay and I showed you that this model super auto regressive model okay so auto regressive model is a very nice simple way of modeling unknown disturbances stochastic disturbances okay which are arising from unknown sources we are played trick you do not know what the source is passing through this 1/q okay it is giving me this decay which is unknown disturbances okay please remember.

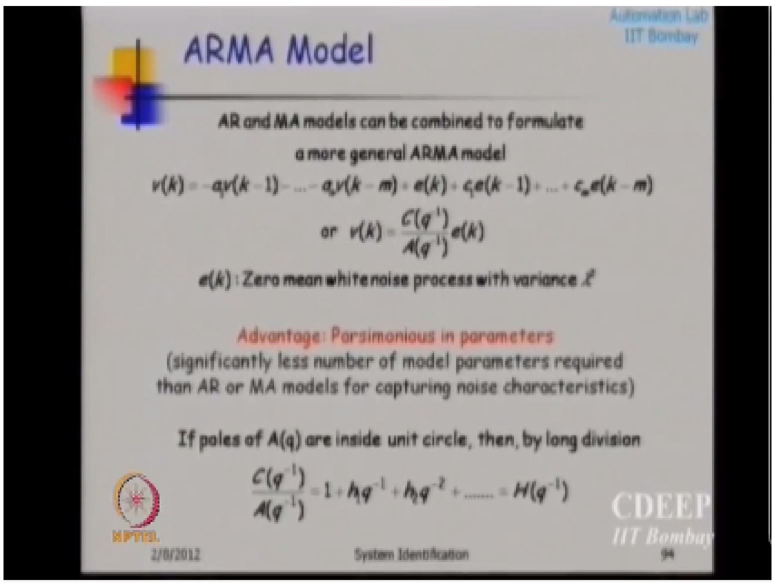
That this ek either imposter this is not the truth this is not a truth this is a model for explaining relationship within that unknown disturbances okay this is the trick to, this is something which behaves similar to unknown disturbances, there is no real source called ek in the real, this is the model okay. Well when you do the exercise you will do this actually with the poles of Aq are inside the unit circle then you can actually expand this and then get the moving process.

This is also I talked about last time that we can convert the auto regressive process into in the moving average process or moving average process to auto regressive process we can do back, so these are only two different way of expressing the same thing, moving average or auto regressive process okay.

I can go from one form to other form in the, I have uploaded the new tutorial sheet, which is your mixture is going to be based on the new tutorial sheet okay majority of it. So please start working with the tutorial sheet unless you start working with tutorial sheet you will not understand what is happening in the class, to work out those problems.

So we will have some session in this week and also in the next week other than the lectures to discuss about the tutorial problems okay I will announce that in the end of the class. So after move 4 between you know moving representation or auto regressive representation I can combine the two okay.

(Refer Slide Time: 55:42)

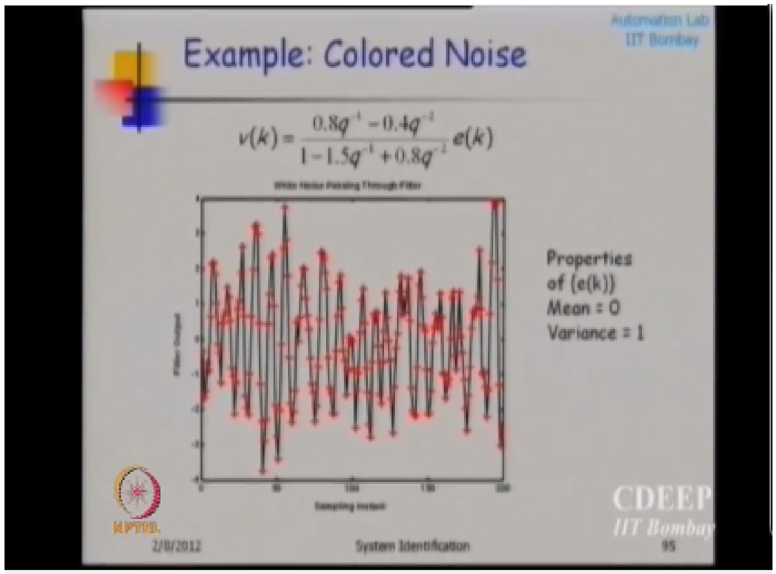


Auto regressive and moving average into one model this is called as ARMA model, auto regressive moving average process okay, so this as auto regressive terms which are coming from past vk it as moving average terms which are coming from past ek okay. This kind of model is called as the ARMA model it is more convenient to use this model because you need less number of parameters okay.

Wherever you need less number of parameters the parameter estimation is difficult, so there is the price to pay. So wherever you need the things are easy you need larger there is the price to pay in terms of you know loss of production men so on. I can actually think of this combined model which is auto regressive and moving average combined together gives me the power of both the ideas together.

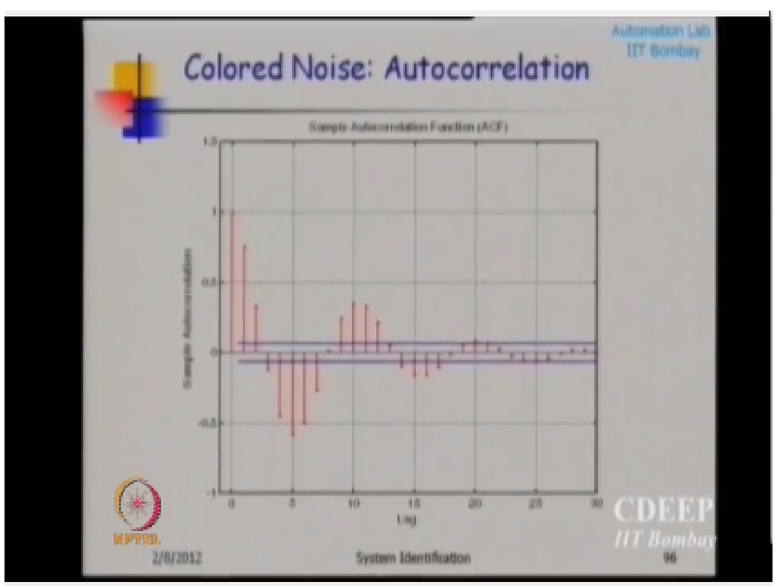
So you can actually have models which require less number of diameters and then well I can actually move between one forms to another forms.

(Refer Slide Time: 56:51)



I am just showing you an example here of a how to do a noise will look like I have fabricated one ARMA process in the computer, this is driven by a white noise signal Gaussian white noise signal, with mean 0 variants 1 how will the output look like it will look like this. Just to give you a visual feel of how the output look like. What is the auto correlation of this signal?

(Refer Slide Time: 57:22)

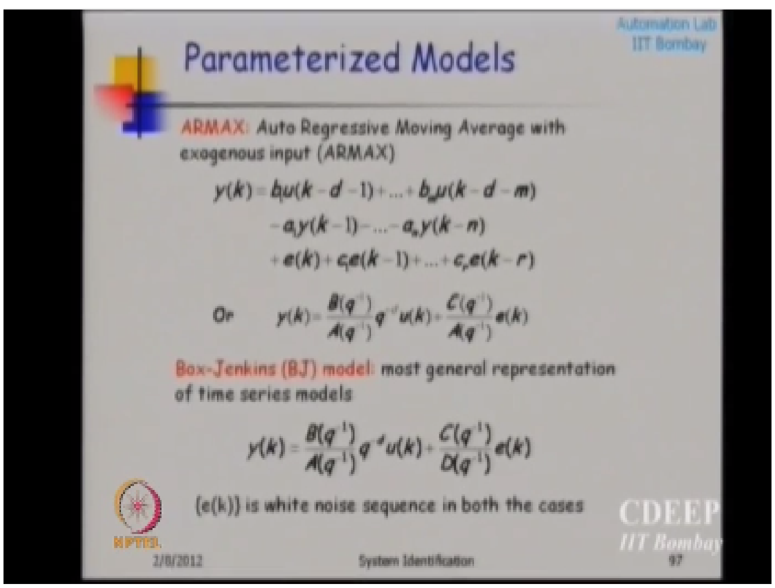


It is a colored noise okay; strongly auto correlated colored noise so ARMA process can be used to model colored noise okay. Because normally if I use only AR or only ARMA only moving average the number of parameters I need becomes large okay. What I can do probably with only 4 parameters here 1, 2, 3, 4 okay see if you do a long division I will come to that actually if you convert this into only moving average only AR model okay.

Then the order of the model that you need to get good disturbances model it becomes large, so you know same if the same model is converted is into AR model probably you will require 20 parameters instead of 4 parameters. So why I want to go 4 parameters because N is small okay, now for the time being you have to trust me that this is done require. I have actually given exercise in which you convert from one form to other form.

You know when you convert from ARMA to AR or moving average okay the number of parameters should be estimated becomes large, number of parameters becomes large perturbation time is large means you know loss of production. So we want models which I have less number of parameters yet are able to capture the noise properties that are why I want to go for ARMA model.

(Refer Slide Time: 59:11)



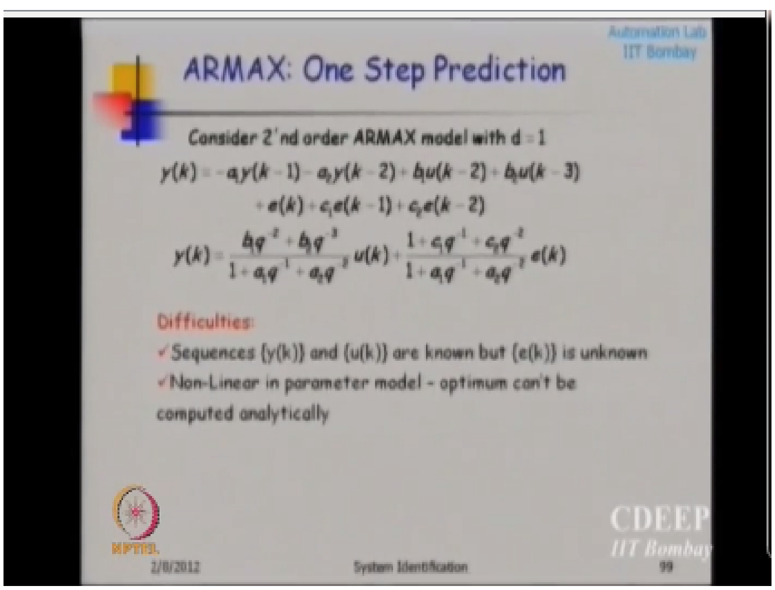
So now I am going to introduce a more complex form called as ARMAX okay, so what is ARMAX, AR stands for auto regressive MA stands for moving average and X stands for excess inputs okay. so this is you know it has terms coming from past u, past y auto regression and moving average. E has to be a wide noise sequence okay, so in the transfer function form it will look like this.

What ahs change from ARX model to this model? See for ARX model = 1 in this model it is not equal to 1, then you might say why am I asking you to have same denominator polynomial here okay. I can have a model which as you knows b/A for a excess input, some C/D for you can do that no problem, this particular model is named after Box Jenkins, you know very famous book types of modeling by Box Jenkins.

Any one of you interested in pursuing this as a serious thing probably you should have this book, this book is like bible of times series modeling. So Box Jenkins model is a more generous model one can think of, you can convert this general form you know if you multiply if you cross multiply then you can convert this into ARMAX if it is the special case of list to 1 and d = A it becomes ARMAX model.

This is most the generous form; one can think of okay, more the general form is less number of parameter more difficult is to estimate the parameters okay. so moment we go for this form it becomes more and more difficult to.

(Refer Slide Time: 1:01:45)



So I know I can, I do not have to identify only ARMAX model or okay output error model I can develop ARMAX model I am going to see how to develop the ARMAX model. So what is my problem, I have this data which is collected y0 to yn I have this input sequence u0 to un and I am going to use some parameterization A is the 2nd order B is 2nd order c whatever and I am going to parameterization this models.

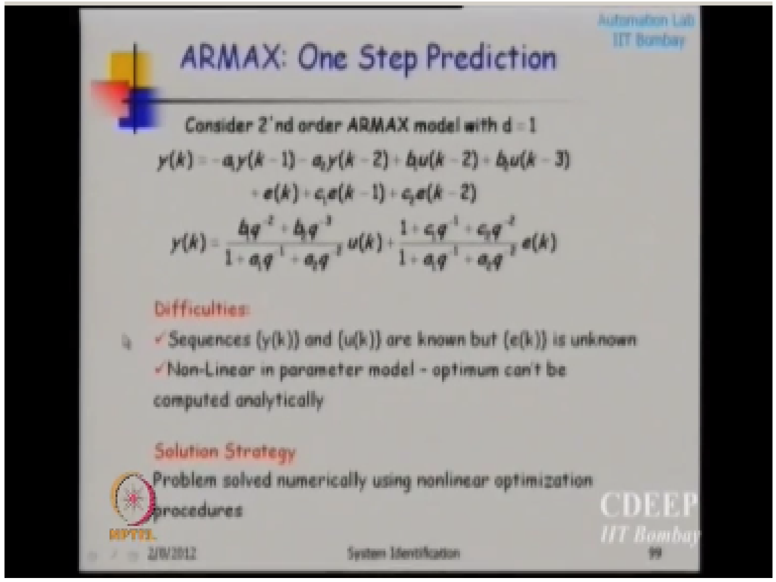
And I want to estimate the coefficients of Aq, Bq, Cq using data okay that is our problem how do you do this? you do this in such a way that you are going to use that b^2 paradise which means some of the square of the errors, what is the errors here ek, should be minimized with respect to the parameters, same idea which we use for ARX life is more complex here because the model is very, the big problem is that ek is not known.

But we have requirement that ek should be a white noise sequence the trouble here is that we have a model in which okay I have this model I can convert this into different equation right all of you know how to convert this model into a difference equation. So this is my difference equation I do not know A1, A2, B1, B2 okay, I do not know ek, c1, ek – 1, c2 ek -2 all these are quantities are unknown to me.

What is known to me is yk, yk -1, yk -2 , uk -2 , uk -3 these things are known to me okay and this etc things that is ek – 1 ek – 2 did not appear in ARX model but they are appearing now here. Now there is the trouble okay how to estimate this model why? Because first of all only yk and uk is known ek is not known okay so how do I estimate the parameters this model is what is called as non linear in parameter model.

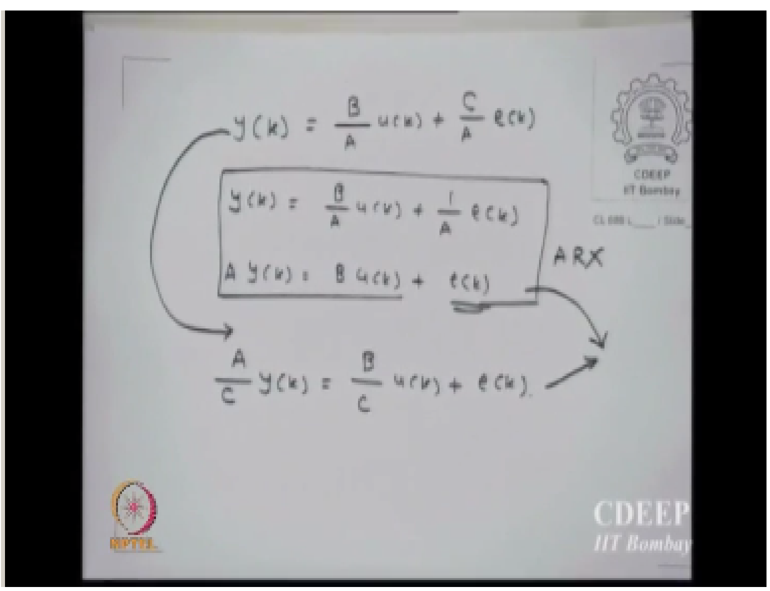
Because see you do not know ek you do not c1 so you have to guess ek you have to guess c1 okay so it is a linear in parameter model it is difficult to estimate you cannot use linear squares business because it is unfortunately does not work here you have to use something else so you cannot untypically compute this solution you have to use non linear optimization tools to solve this problem okay.

(Refer Slide Time: 01:04:46)



So there is a trouble when you want to do this what is known to you now let me, let me go back and tell you what is my aim now for the next step this model is expressed in terms of u and e right but e is not known to me okay I want to convert this into a form which is like this

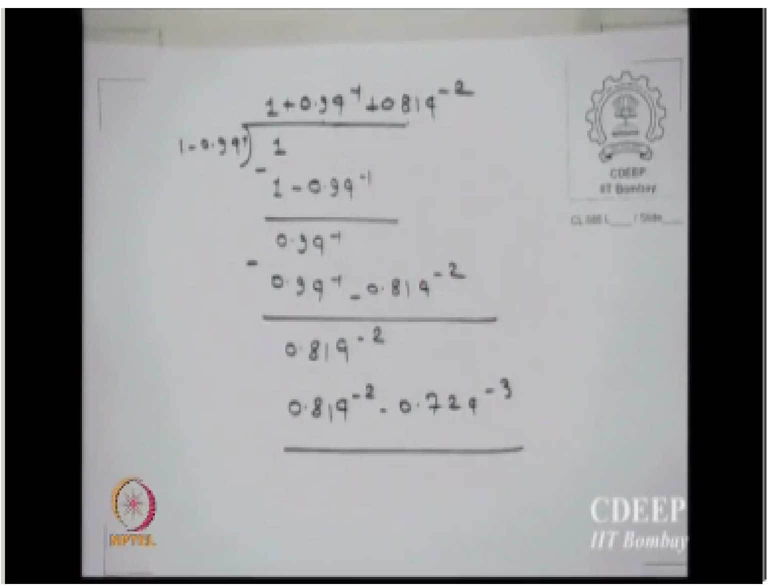
(Refer Slide Time: 01:05:26)



See this model is okay this is fine now, I just the step is ARX model was yk=b/auk+1/aek remember the nice thing about ARX model because I could write this as a yk=buk+e only ek applied right there was in a repressor only known things where there, there was no unknown things okay can I convert this model into something similar so this is my ARX okay I want to convert into something okay which is having only ek can you think of doing that here right now multiply by a/c good.

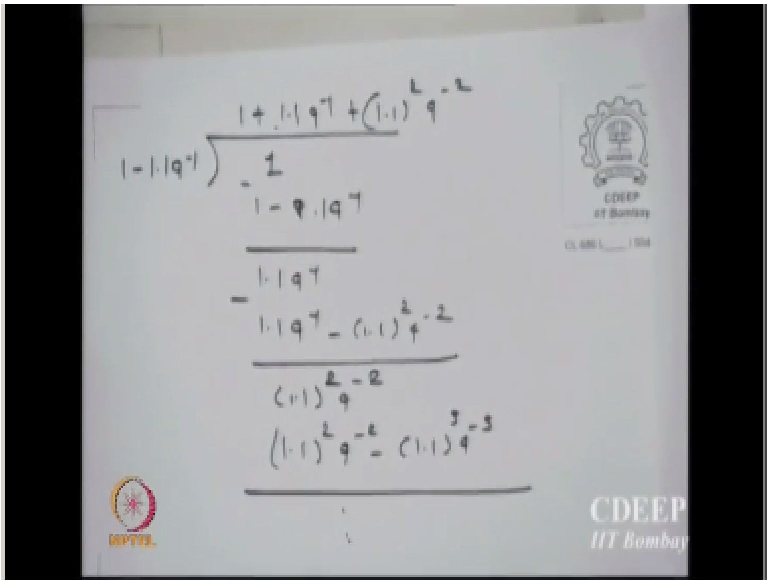
So I will say a/cyk=b/cuk+ek I am multiplying by a/c okay this look similar right except this 1/c is appearing what was c in the case in radius model why life was simple but right now c=1 okay there is one more thing that becomes very cruel here you get this 1/c okay 1/c so you want to find out inverse of c okay it is required that is inverse of c should be should be stable okay inverse of c should be stable i9f inverse of c is not stable we have trouble okay let us take one particular example of c can you do this 1/0.9 q-1 what will this be can you just solve this you get two situations 1/1.1.147 what will be the first case can you do the long division and tell me how will you do this

(Refer Slide Time: 01:08:46)



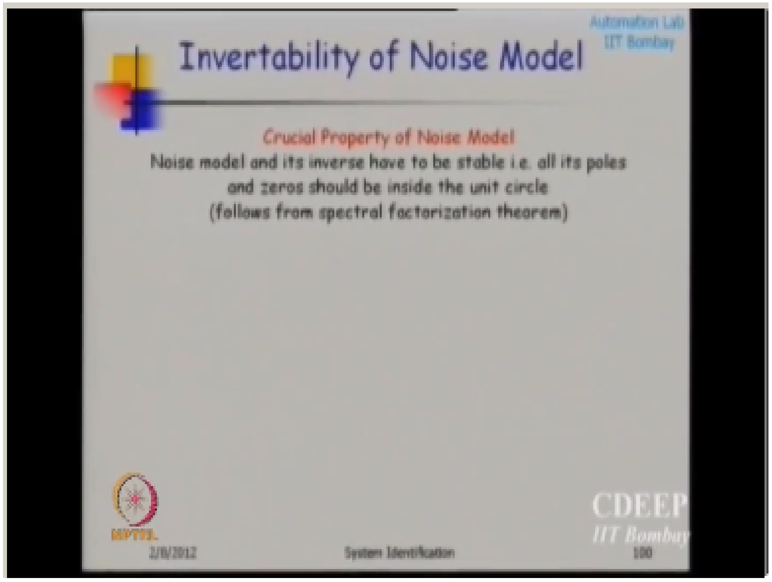
Okay so this is $1/1-0.9q^{-1}$ right so my first thing is one here $1-.09q^{-1}$ I subtract I get $0.9q^{-1}$ then what should I do multiply by $0.9q^{-1}-0.8q^{-2}$ what will I get 0.819^{-2} what is next $0.81q^{-2}-0.72q^{-3}$ and so on you can just go on doing this what is happening here 0.9 to 0.81 what will happen next it will go on reducing okay it will go on reducing so after sometime I can truncate this series right and then I can you know I do not need to worry about the higher order terms because they are becoming smaller and smaller okay just construct this with the other situation.

(Refer Slide Time: 01:10:36)



Now let us say you have $1.1-1.1q^{-1}$ so what will happen here so this should be $1-1.1q^{-1}$ so if I subtract this we will get $1.1q^{-1}+1.1q^{-1}$ so I will subtract this I will get 1.1^2q^{-2} then this will be 1.1^2q^{-2} so what will happen to this parameters here radius start growing okay if they start growing okay which will go to ∞ after some time this coefficient will go to ∞ and.

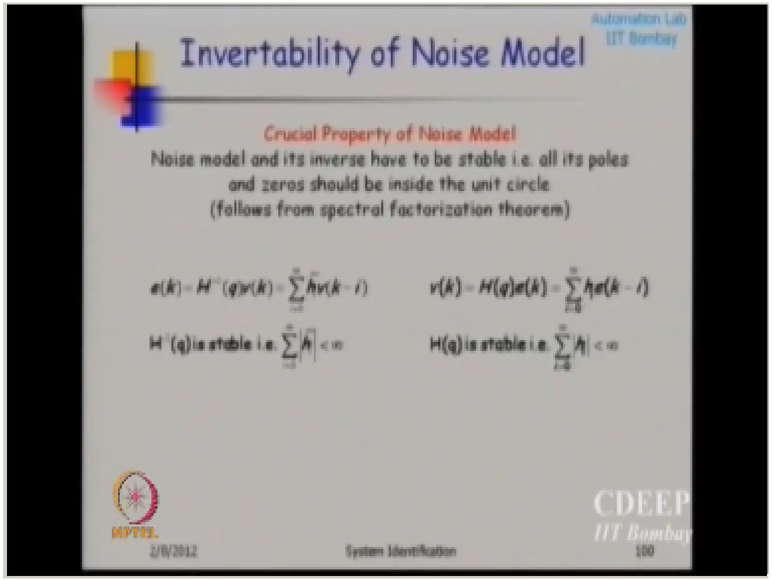
(Refer Slide Time: 01:11:53)



This thing can cause trouble when comes to, so what is required what is critical when you do noise modeling and taking of very short cut? There is a fundamental theorem called spectral theorem and I am not sending that theorem and I am giving you essence of it in finding in simple words okay it is necessary that if you want to model a stationary process okay as a colored noise.

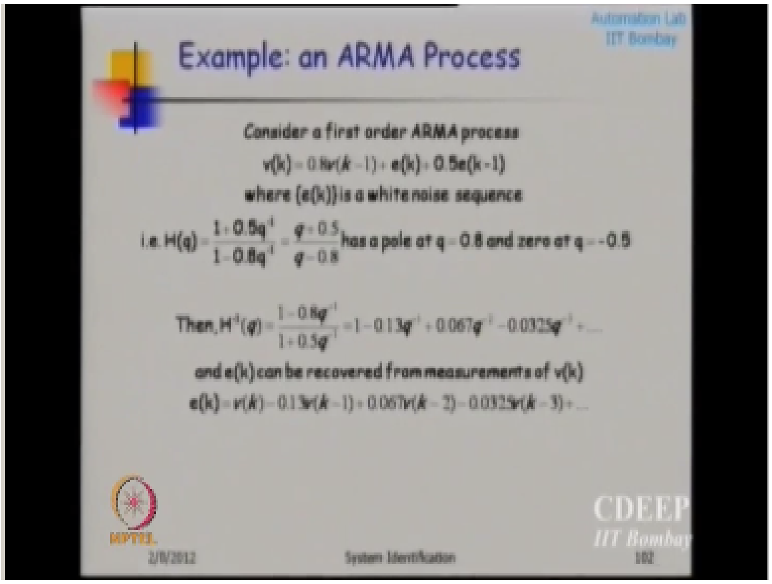
Okay it should be invertible inversely stable it should be stable and inversely stable okay so which means if I am constructing a noise model c/a all the poles of c should be inside the unit circle all the poles of a should be inside the circle if this condition is not stratified we cannot develop a model here because what will happen is if the noise model is non invertible the terms will start growing and it will cause problem in the parameter estimation okay

(Refer Slide Time: 01:12:49)



It is very, very important that the noise model in invertible I should be able to go from you know e to v, v to e it should not be a problem so h should be stable and h inverse should be stable this is the fundamental requirement of modeling when you do aromatics models or model and if you does not happen you cannot develop models I will just show you one a example.

(Refer Slide Time: 01:13:29)



This is a ARMA model okay and h is given by 1+0.5q and 1-0.8 this has 0 at -0.5 this as pole at 0.8 and if you do this long division you will get this terms here if you do this long division then what you can see is you know, this coefficient will start dimension fast this coefficient will dimension fast and look at the appreciate first few terms in the series.

That is very, very curial when it comes to voice modeling will go over second in the net cross and connect it to estimation of the model parameter curial property of ARMA model box model is that could be it should be stable and inversely stable without that you please go head and look at the rule it is not possible that you just listen to this you know or you go back and start solving the problems okay.

NPTEL Head IIT Bombay
Prof. R.K.Shevgaonkar

CDEEP Head
Prof.B.L.Tembe

Producer
Arun Kalwankar

Project Manager
M.Sangeeta Srivastava