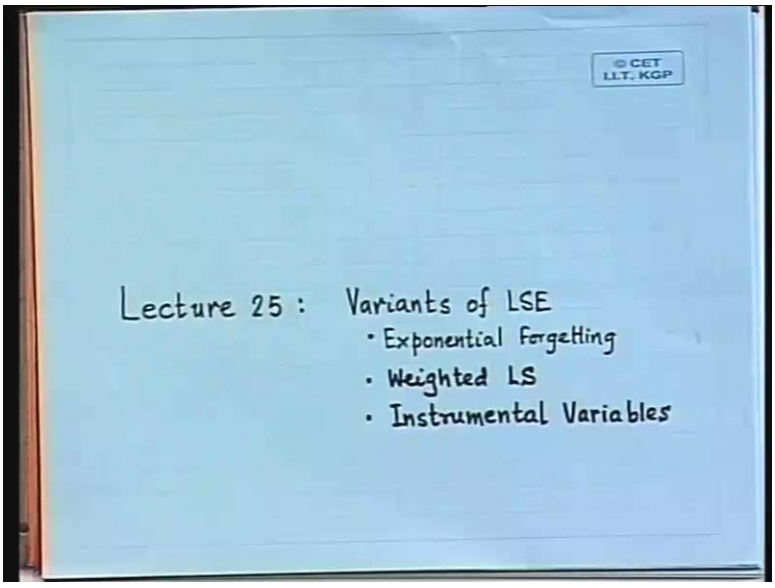
Estimation of Signals and Systems
Prof. S.Mukhopadhyay
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 25
Variants of LSE

Good morning, so what did we learn yesterday, the basic least square algorithm?

(Refer Slide Time: 00:01)

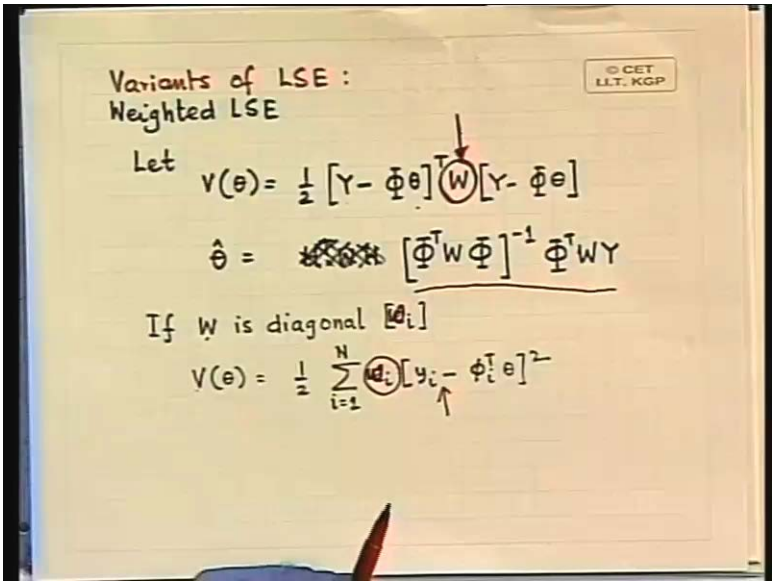


In a firm y is equal to phi transpose theta plus v, and saw some properties and found that it is supposed to work well under certain; if certain properties are satisfied by D data vector phi and the noise are error vector V, but first of all now this is the this is the basic algorithm. So what we will do today is, first of all see some variants of the basic algorithm, very common variants. And so that in certain situations where this these properties are not directly ensured, we have to do something so that in the modified formulation these properties are ensured.

So if you when when when once once we do that we will get the various radiants. We will also do another thing that is we will now today try to see, so far we have been doing just y is equal to

phi transpose theta, now we are interested in generally we are we are interested in estimating transfer functions; so we will see that under various kinds of model properties and noise properties, how does how how to form the equations so that it comes in the form y equal to phi transpose theta? Because we have seen that, the the algorithm can only be applied, provided you can cast your model description in that form, right. So that is what we will see today. So first of all we will see that, we will we will consider a slightly different performance criterion, previously we had considered this performance criterion; Y is equal to Y minus phi theta transpose, previously this was not there this we, in our ordinary list of formulation we assumed that this is identity, so it becomes a normal form of squares and we got the solution for the least square case.

(Refer Slide Time: 03:17)



Now we will put some W that is slightly more generalized, in what way is it generalized? That now we are not putting we are not putting the same amount of stress on errors at the various instance, so it turns out that this is this is useful in certain situations; in fact this is very commonly applied in cases like adoptive control, for some reason. So first of all we will see that, if we include a W here; that if we are not interested in just sum of squares, but in a weighted sum

of squares, where we are doing waiting about time that is K then then then what is the solution, okay.

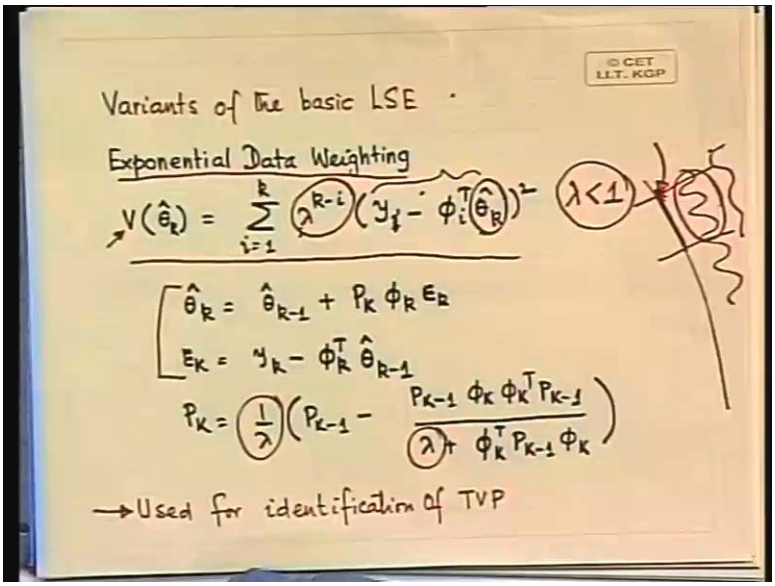
Again you can do the same thing, just just differentiate this V theta with respect to theta and set to zero, then you will get this relationship which will obviously the same as our old relationships; if you said W equal to I, you will get the least square estimate back. And if W is diagonal, if W is assumed to be diagonal then this V theta is nothing but; what I said that the errors at the various instance you have putting a weights, actually I should put okay it is a confusing notation, we should put w i suppose just to say that, these are the diagonal matrix elements.

So then you have weighted the errors at the various instances, that is what you have done. Now why should you weight? Why should you weight the errors at the various instances, under what situation it may be necessary? One case is that, when you you expect that the parameters; see you are looking for parameters based on data, now the question is that if the parameters are constant then all the data has been generated from one constant parameter, and you are looking for it. So it makes sense to put the same weight on all the data points, but suppose now the parameter is slowly changing it it itself varies with time; so if it if it is varying with time and if you are at the thousandth data point, suppose K is equal to thousand then is it meaningful to give the same weight to the error which is being to the error which happens at nine hundred and ninety-ninth instant and the error which happened at the first instant?

Actually you should not put so much weight on the error in the first instance, because that error will actually cause the different parameter; within this thousand instance the parameter have also changed, so the parameter that we are looking for now, you should find out from the recent data because the long pass data has been actually have been generated from the different parameter, so you should not put so much weight on that that data. So in other words you should you should give more weight to to recent data and less weight to the pass data, then then what will happen is that the is that the estimate that you have as K is passing; so you are continuously getting new and new parameters. So then you will find that the parameter that you are getting will actually try

to track the variation of the in the parameter, because because that will give only give give weightage to the most recent data, right.

(Refer Slide Time: 07:32)



So based on this philosophy we would like to, you will like to find an optimal estimate, according to this sort of a performance criterion. See what am I doing? This is this is Y k, i is equal to one to K. So what I am doing is that, this is the parameter that I have that I have got at the fourth instance. Suppose so I have data up from one to K; now using that I am trying to predict what is Y I, so I am trying to predict y one, y two as well as y k minus one, k minus two all that. Now since the parameter were slowly changing, so so it is not likely that you will be able to predict y 1 using theta hat k; because y 1 was generated using a line pass parameter, that has changed.

So it is very likely that, this error for long past data points, are may be large but then you should not change your you should not change your estimates because the long pass parameters are large; because the long pass parameters are likely to be large because because a real parameter which has generated, the data changed. So you should discount the past. So so so this error should not get for much weightage when when when i is equal to one rather that that mean K is

equal to say, at K is equal to k i or K is equal to k minus one or K is equal to k minus two, this recent past they are should not be larger then your parameter is wrong. So you should put large ((00:09:04 min)) weight on these recent errors.

So that is what we are doing. We are writing these errors by a factor called lambda to the power k minus i, where lambda is less than 1. So if lambda is less than one suppose, lambda is point six; so first one is getting so you are giving full, when when i is equal to k it is one, when i is equal to k minus one is point six and point three six, then it is point two one six. So after then seven eight samples there is there is hardly any weight of the error in the, I mean past which is beyond the seven eight samples.

So you are so which means that; you are doing this optimization only based on the last say its seven, eight, ten, twelve samples whatever you choose, what value of lambda you will choose depends on what at what weight you expect your parameter varying. Usually, actually this this lambda is chosen very close to one, I mean they are chosen some of the points nine nine, point nine eight because generally parameters variations are supposed to be slow, why? What do you mean by parameter after all?

If the parameter is varying, why did you call it a state? So actually if I if if if you have two parameter; if you have two two signals one of them is varying very fast, another is varying very slow then you can consider that for any reasonable timing stand, this is constant and this is your signal which is varying. So for so actually that is you call the parameter, otherwise if if this one also varying fast, you you cannot call that parameter you should be called as state. So by the fact you are calling it parameter, even it is varying it is it is variations must be much smaller than much slower than the signal which you are calling states, right.

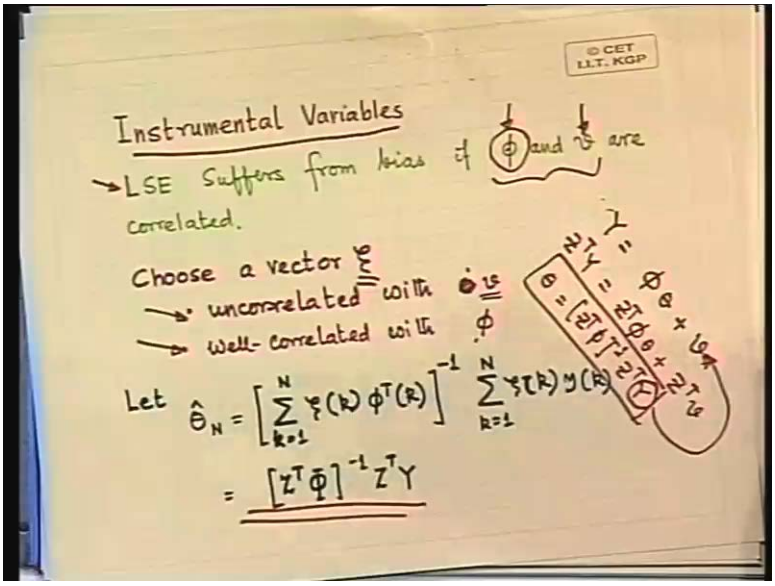
So therefore generally parameter variations are much slower, compared to your your sampling instance; generally probably vary by ten percent or may be five hundred samples. So so typically your your your data should not be five, six, seven, eight samples on which you guess your estimate; it will be some I mean a few hundred samples, so that is why this lambda should be

chosen is usually chosen very close to one. But it can be less than one and if it is may be even if it is made point nine nine, it has significant impact on the estimate.

So now if you formulate this then it it is again simple; you can just again differentiate and I mean get the solution for this case, and similarly, actually this means that in our old version w, this w i is equal to lambda to the power, w i lambda to the power K minus i, correct. So we are so this since this is an exponential, so this kind of which this kind of a strategy is called exponential data weighting, sometimes call exponential forgetting; because you forget the past in an exponential rate.

So if you if you just may correspond this performance criterion, if you again rederive the the least square algorithm then again reclassify it in the in the in the same old manner, you will get the same equation. These two equations are absolutely similar, this equation is different. So this one so covariance update equation is different at these two points, other than that the equations are same as the least square equation. so they are very simple to implement, its nearly the same. And it is used for identification of slowly time varying parameters; which is typically used in adoptive control, because why I am try to adopt your control because you expect your parameters to change, it may change depending on set points because some operating point to operating point it may change, right. So so this is used for such things, so this is our variant of least square number one.

(Refer Slide Time: 13:45)



We have so when you are expect your time varying parameters to change, you you can use this. Next one is instrumental variable; this is a this this is very very suffer passed approach works really well provided you have you can you can see that instrumental variables properly, we will we will come to that. What is the idea there is, what form this idea is come? The idea is come because sometimes, it so happens that we want we we had seen that the least square least square algorithm gives an unbiased estimate only when this phi and v; there is element of phi and the elements of the air are uncorrelated, they are correlated it will tend to bias, right.

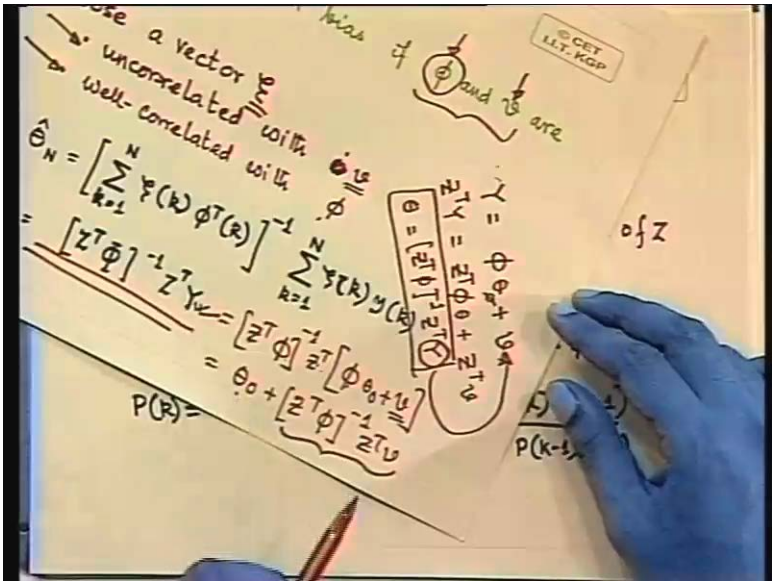
So now now the question is that, how can they be made uncorrelated? If if it is so happens, then phi and v are correlated, what do you do because I mean if you re square you get biased parameter. So the question is that, can you got use some other parameter some other get in place of phi? Now how do you have to ensure that, when you when what what sort of vector you should choose? Firstly, first property is that they could be uncorrelated; see you have to choose this new vector which you want to use in some way in such a manner that, that that vector and v will be uncorrelated, fact number one.

Fact number two is that, in your in your least square estimate you have you have something like a phi transpose i inverse, that inverse should not get disturbed, right. So phi transpose, phi inverse should not suddenly become very large; may be phi terms phi may be is very large then your parameter also get very large, so it should remember phi transpose phi. So in other words guideline is choose a vector, new vector which is uncorrelated with v and very strongly correlated with phi. So if it is strongly correlated with phi, means basically its like phi; so if it is like phi, phi transpose phi inverse, I mean property you maintain while that factor is going to be uncorrelated with v.

So you have to choose certain vector and then use it, how? So the whole idea that, you are basically your basic equation; Y is equal to phi theta plus v this is your model, so so what did we do? We we actually multiplied it in in in obtaining least square; we actually multiplied this side by phi transpose then it will get become phi transpose phi then we multiplied by phi transpose phi inverse, that is how we get the least square, correct. Than lets that is multiplied by a by new matrix system Z transpose y, then it will be z transpose phi theta plus Z transpose v. So now what is your theta? So now take theta as Z transpose phi inverse, Z transpose Y.

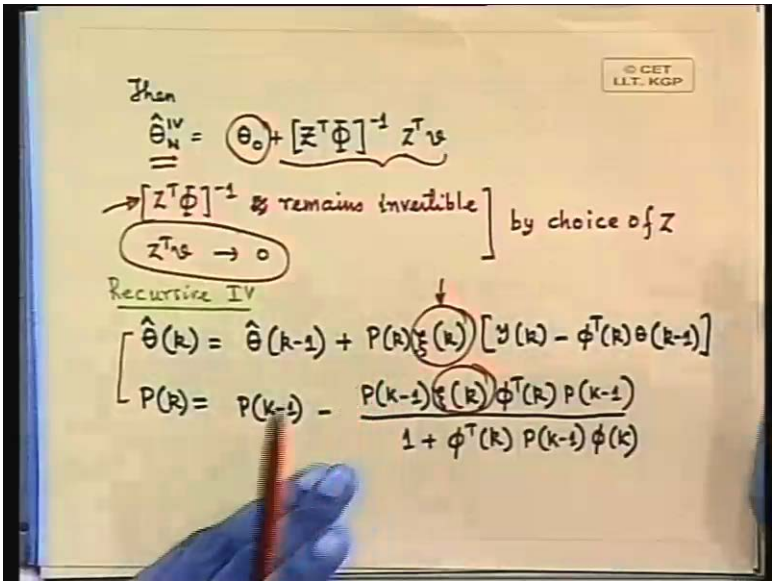
Suppose you calculate theta like this, so so this one, why why did have to do this; because now this can be shown this theta had n, so if if in place of Y you put this equation back again then you will find that, theta i v is the true parameter theta naught, if you are if your data were generated using the same model. So if your data were also generated by this system using the same model form but using a new but using a some true unknown theta naught; if if that is correct then the initial then the instrumental variable estimate, can be shown easily to be just just by substituting that y is equal to,

(Refer Slide Time: 17:50)



so you will get Z transpose phi inverse, Z transpose into phi theta plus v phi theta naught plus v; this is how Y has been generated, right. So now using that Z transpose phi inverse, Z transpose phi, so you get theta naught and then you get z transpose phi inverse, Z transpose v that is your error term, parameter error. This is your theta. This is the amount by which your parameter, estimated parameter is of from the real parameter, if this system used.

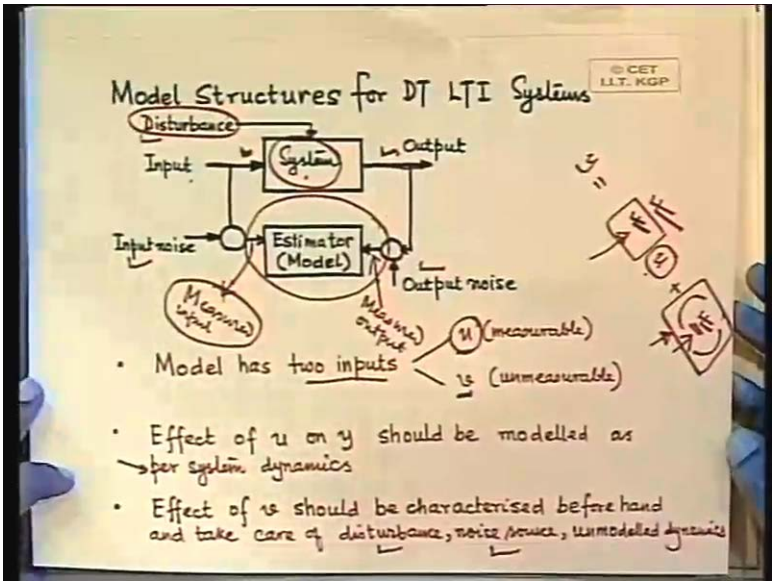
(Refer Slide Time: 18:30)



So so obviously now so what do you want to do? You reduce this, so if you want to reduce this; you have to choose Z. Now you have to choose Z, such that this inverse invertible and not too close to singular. And this should be zero, that is what you want. So how to get that, so what Z you should use, that is a that is a main question. So the algorithm is exactly same, least square algorithm is exactly same only in so the least square got to be exactly same, only in some places in place of phi you have to use this weight, that is all.

Now the question is how would you choose this? That in cracks of the matter, so the what is the idea? The idea is that the now we will see there are that could be various ways of choosing this; and we will see how to choose this in the context of our dynamics system, shortly. So I will stop here and we will again come back towards the end of this lecture as to how to choose the x i, when you are doing transfer function identification. You may choose x i in various ways for different kinds of things; it is not that already trying to do a transfer function identification, but but then that is our main focus in this course.

(Refer Slide Time: 20:00)



So we will come back to it towards end of the lecture, before that let us let me first introduce that now, all this i transpose theta as a what fallows, what theta what theta we should use for transfer function identification; because we are interested in identifying dynamic linear time variant, discrete time dynamic models, okay. If you are that that we will discuss now, we will have also remarks of suppose it is a continuous time model, how to do? So those are variants, so the basic idea is how to be discrete time model? Continuous time model is nearly similar.

So first of all remember that, again coming back to model structure that this is what we are doing; actually you have a system, you have an input and you have an output. And this is your estimator and in the estimator; what is what is what are you sending to your estimator? In the estimator you are sending measured input and measured output; remember that they are not the same as this and this because you have to use transmitter, that means introduce measurement noise. You may you may find that, you are giving the input for our computer; so you will say that I know the value, after all I am giving the input, that the value which is there in the computer and the value which is actually going to the plant may be somewhat different because of several facts, because of convertor, characteristics, because of actuated dynamics, so many things may happen.

So in general first of all thing, I want to say want to say is that the the that the data related system is generating this inputs outputs; this physical input and output which are there in the system, they are not the same as you are using in your estimator, remember that. And you are you are going to build the model based on these two, not based on these two, okay. So so now other force or factor that coming here apart from the fact that, the plan itself has have disturbances which we have not covered. So in general all these affect this disturbance, this noise, this noise any any any dynamics which you have ignored; so all such things you are going to.

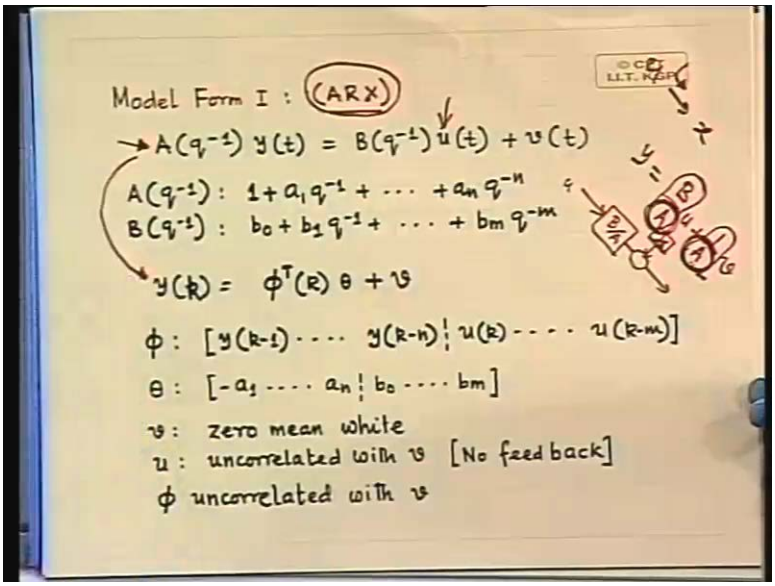
So basically what you do is in your model, you try to get it say y is equal to some function of u which is your controlled known input; this one that we use plus what are the rest, you said that is some some some some noise down and you want to characterise that noise. So that means that you have to you have to choose such a model, such a a so so so when you are trying to estimate; you have to choose, what kind of thing is this, what structure is this and you have to choose what is the property of this, all these that you are not modelled that you are called and all these that you are chosen to describe by the term noise. What is its characteristic? Does it have a daffy component or not?

Because you must at least get some idea of these components or otherwise; that noise property that you are assuming, that this this part will be zero mean why it and all those things will not come or they will give you bad estimates on these. So so you must when you are selecting a model structure you must must put what and what should be? What you should put here and on on how you are going to characterise these; these are the two things that have to be thought about. Then you have to put design number model structure then after you decide the model structure, you have to do parameter estimation in that stretch.

So typically model has two inputs u and v; this one u explicitly used in your in your estimation equation, this one you do not know so you have to do something based on the based on the characteristics of these that you can find out may be by doing offline experiments, something something. So effects of u and y should be modelled as per system dynamics; obviously you must think that, this is like a this is like a third order system from physical consideration, from

from I L C circuit, from motor model whatever. Then this part obviously you would like to characterise by third order system because that that, so you choose the third order model for estimation and this part you should use depending on what what are value you have about disturbances, about noise sequences, right. So that is how you have you have to choose, so what are the typical model structure which was which are used.

(Refer Slide Time: 24:56)



For example, the simplest model structure that is used for estimation is this one; now we are talking about dynamics system estimation, now you are not seeing phi transpose theta, we will we will see what is phi and what is theta exactly in this model. So in this model people assume that y and u, that is the measured y and the measured u are related with by this measurement; it is an assumption, so so you can you can have certain observations about this about this. For example, you can say that, you can see that if you do this you will find that y is equal to B by A u plus 1 by A v from this just just simply divide by A. So this is your input transfer function and this is your ((00:25:59 min)) transfer function.

So you are assuming as if the data is generated by u here B by A and here v; also coming from one by u, so this A A assuming to be common why? That means, you have if you you if you use

this model structure you you must have some basics of thinking that the model structure that the that the any model that may be coming is also coming to the same set of poles; which means that it it may be some model structure, now what is what is that signal which gets filtered through the ((00:26:42 min)) it is the state because what is that what is that what is the state two input? What is that state two input transfer function; that has s i minus a inverse. So that means that the the a polynomial, which is which is the determinant of s i minus y, in this case Z i minus A, is is coming in the denominator; So which you by which you mean that, this large must have just like u four x, u two x has that whole one by A.

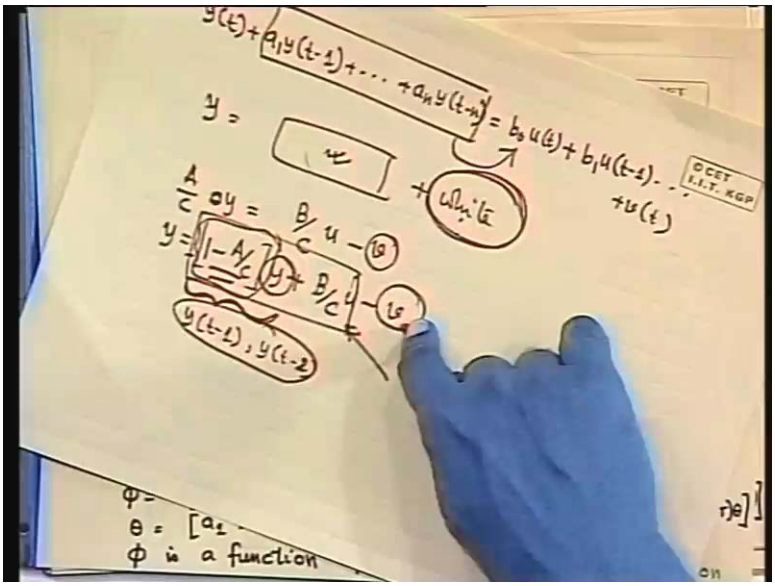
Now v too we are also adding this one by A; that means we we must have entered along with you, so you are kind of assuming that noise is also being being being processed by the system. So you know people sometimes there are people you you may not always be able to justify such things, but actually what you should do is in a in a in a practical situation what people do is that; people do trial and error, sometimes they will they will try something. They they will first make some selection based on their their understanding, what kind of noise they have, what kind of calibration experiments they have done etcetera.

And then they will also try some numerical experiments; that is if they not satisfied with the with the kind of parameters that they have been obtained, then they will change the model and then again try; so it it is not that you will be able to arrive at the structure always and these structure are popular because they, it is also affect that sometimes very very simplistic structures are dealt with in text books because they give rise to very nice analysis.

So fine but they also will reasonably work. So so let us first look at this this model structure which is which is the simplest; it is called A R X or autoregressive with exogenous input, A for auto, R for regressive and X for exogenous input, exogenous input is u. So you have so when you have these are these are terms which are used by the by this ((00:29:09 min)) community. So in such a case, you can you can easily understand that how can I cast this model in to this phi transpose theta form? Because you need to cast so our algorithm use it in this form, so you see that you can this model; you can nicely cost in that form. For example, phi can be this, what is

this? If you write it in time domain, it will become what? It will become A is see; one of the parameter of n v you can normalise, so I have chosen to normalise A C dome as one. A A by B by any one parameter can be normalized, so I chose to normalize A C that is the usual custom. So in that case, what you have?

(Refer Slide Time: 29:57)



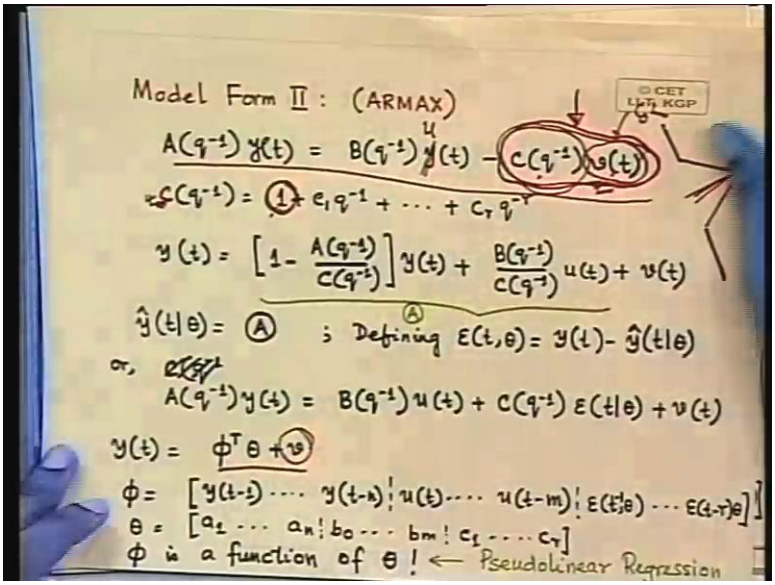
You have if you if you write this equation in time domain you have y t plus a 1 y t minus 1 q inverse plus plus a n y t minus n is equal to b 0 u t this term may or may not have; this is this is the direct term, if you have a d matrix then you will you get this, u t minus 1 and so on, plus v t. So if you bring these terms to the R H S and then arranged in terms of this form; you will get so so you get minus a 1 k minus 1 minus a n y K minus n, this minus either you make this minus or you make this minus it does not matter. And then b zero u k and b n u k minus m. So this is your phi terms plus k plus v.

So you see that, you see now now you are now so this system; you are now able to cast in this form and then you can apply least square, if this v that the question is you can always like it like that, you got some y and u you can always look for modelling that form. When will it work? Only when you think that this v is going to be zero mean white, so this this this this v is not zero

means white; then this this is not going to work firstly. Secondly only v means zero mean white will not do the job, v should be uncorrelated with u because; v should be uncorrelated with phi and and phi contains this. So v should be uncorrelated with u, when does it happen? When the system is not operating in close loop; if the system if the system is operating in close loop then you are computing u rest on y and y contains v, so your u is u contains v. If you are using a feed back control law then you are calculating u rest on y, y contains v so therefore u contains v.

So therefore this and this and v are going to be correlated. So if you make these assumptions, that they are uncorrelated then phi will be uncorrelated with v v and then you can expect good results from by applying ordinary least square; but not otherwise. So what do you do if you are not if that does not happen? Then you have to use, that means that yours this description assumption if it is wrong then you have to see, how you can improve this description; if this really does not turn out to be zero mean white then what is it? Find find what it is and try to model it.

(Refer Slide Time: 32:57)



So you can get this. So sometimes people will say that, no no I will assume that now this term previously which we were been assumed white; now they will assume that no no this may not be white but it may be a term like this, where this is white. So what does it mean? Now you are

assuming that the that the that the and this this term which you previously assume white is now follow of some white terms which means that its autocorrelation function is not like delta; but it will state little bit depends on what is up what is the order of the c polynomial because you are taking weighted average of r white lines, white signals.

So now your autocorrelation function is is going to spread, so it is not like a delta not ideally white; but it it will not be it will not be correlated with the toss circuit, but it will obviously not be correlated with the with thirty odder samples may be. So now you are assume such a thing; so you are relaxed your noise model assumption, because you found that the previous model does not work. So now how do you now now again you see, what we have do now? This is the model description and this is how data is generated, you are assuming. Now you have to find, you have to cast this kind of description into a five terms pose theta form; such that the residual is white unless the residual is white, you are not going to get two parameters that is the sure.

So now how do you cast this into phi transpose theta form; such that what are will be remaining with phi transpose theta, that part will white that may not be exactly this that. So what do you do is; you try to do some manipulations that is you first divide this this thing. You know you know you have to you have to this is the model; so you have to whole whole idea of main case is that you have to write the model as y is equal to something plus white, you have to expect this model like this. So first of all how do you express something plus white? This is not this is white, but this is c q inverse is there. So first I will divide, so I will get A by C; see this was my equation, at first I will try to make this term white, let me try to make this part then let's see what comes here.

So I get A by C y is equal to B by C; sorry this is u, u minus v. So first of all brought it into a form where this is white, now I want to write it as y is equal to something plus white, so I will now write y is equal to this this side I will write 1 minus A by C y. I can write that, can I write that? No no it should be minus plus B by C u minus v. Now what is the what is the duty of this 1 by I mean is just 1 minus? I I actually if you if you compare see incidentally this c matrix I have chosen this as 1, why? Why it cannot be some arbitrary number c naught?

After what am I what am I trying to do? I am trying to what with this c that is, what I am trying to do? I am trying to model the power spectrum of v, remember that v is the stochastic variable it does not have value; but you do not know what is the value, so all you can do is that you can modelling power spectrum. Now if your modelling its power spectrum, you can always put this as one and then adjust the variants of v; nobody is saying that v is going to be a white signal, in a white signal of what variants? I am not saying that the white signal of unit variants. So so if this is c naught make its variant c naught square and then make this final as 1.

So so the idea is that, you can always choose it as 1 and then adjust the variants of v; such that the noise power is 1. So that is people can choose this as 1, so if you choose it at 1 remember that; this this thing that is 1 minus A by C only contains past terms of y, so it will contain because here is 1, c c naught is 1 and n naught is 1. So when you do one minus A by C in the numerator you do not have any direct term, so you have only past terms. So which means that, this is now going to be a vector of past terms; so this will only contain y t minus 1 t minus 2, you you can just experiment with any example just put values and try. So so so this is going to be a function of past values of y; this is going to be a function of u, this is going to be a function of v and this is going to be y.

So now I have to write this equation; in term if I can write this as phi transpose theta, this part then I am done because then what what is whatever remaining is white then I can apply L A C, right, so so that is the idea. So now what they have done is but the main problem is that, how do you write this one because this is not, this is not a linear term; this is not, can you write it as phi transpose theta? That is the c q inverse at the bottom. So you have to do some clever manipulation, such that it can be written as phi transpose theta.

So what is that manipulation? The manipulation is that; if any but you know by by largely this much you should know that, if something is some quantity plus white then it will then then then what is it is what is conditional estimate? Only that part because this is a function of y t minus 1 t minus 2. So therefore this is given and this is all known; so this is given, this is white. So what will be if if if A B C are given; then what what is the what is the conditional estimate, only upto

this part, white part you cannot you cannot estimate anyway. So this is your conditional estimate, this A marked part, these two terms. So then then what is your prediction error? Given given the past data and theta, it is this much. So now what you can write; so now see that, you can write this equation, the same equation in term like these. Well, this is a prediction error and this is the v; once you have white theta in this form, then you know that you can put it in phi transpose theta. What is the catch? The catch is how do you know this? Who gives you theta, theta, you want to estimate?

So so so nobody will give you theta A B C parameter values, who will give you? How will you calculate this? If you want to put it as phi transpose theta; this will become elements of phi but for for computing this, you need theta and you do not have theta. So the whole idea is that, if you do not have theta then you then you use what you have; at least what are the estimates of theta you have, you have you are going on improving your estimate. So at any point you do not have you you you may not have the true estimate; but you have some estimate, use that that will also improve the picture.

So what do you do is, so people what are its latest estimate of theta available; people will use that to compute its L A C, so then you can put it in phi transpose theta form. Remember that now very interesting thing comes, that is phi itself is now a function of theta; because elements of phi are now functions of theta. So the basic remains us linear regression is now valid. It is this equation y is equal to phi transpose theta does not remain linear in theta.

Previously it used to remain linear in parameters, why because the because what are the elements of phi at least it does not contain theta. Now that is our ledge because we are now computing elements of phi, using theta; that is why it is called symbol linear regression. It looks like linear but actually it is not linear. But anyway whatever it is at least you have done something, you have tried to model this thing; and and you put whatever you know best all about theta, to be able to compute it in a phi transpose theta form.

Now hopefully; as theta estimates will improve, this competition will also improve and that I will give further boost to your theta estimation. So that is like record; see you know as as the estimation of theta improves, the this estimation improves therefore the estimation of theta improves. So that hopefully there will aide each other rather than destruct each other, and that is rather difficult thing to prove, that that it will actually aide each other and result in a better estimates of A and B.

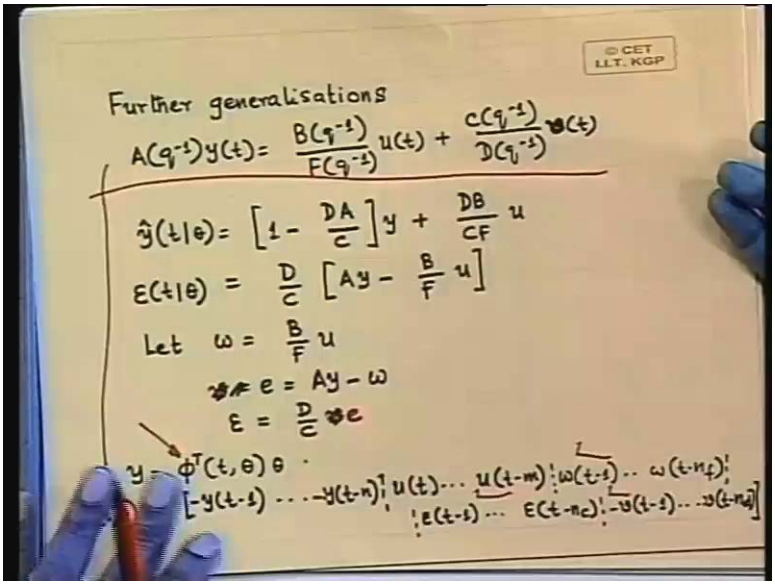
You are actually interested in mainly interested in the estimation of A and B, you are modelling C; because you want to improve your estimate of A and B, that is why you are doing it. You are generally not generally noise models are not directly used, especially in contour. So so this is this model is called A R M A X, previous one was A R X, why? Because it contains an autoregressive component, it contains a moving component; now it is an average of white noises and it contains exogenous input. So A R M A and X, previously it was only v; therefore it was only A R and X N X, now it is A R M A X, okay.

Student <for also we had these>

[Conversation between Student and Professor – Not audible ((00:43:42 min))]

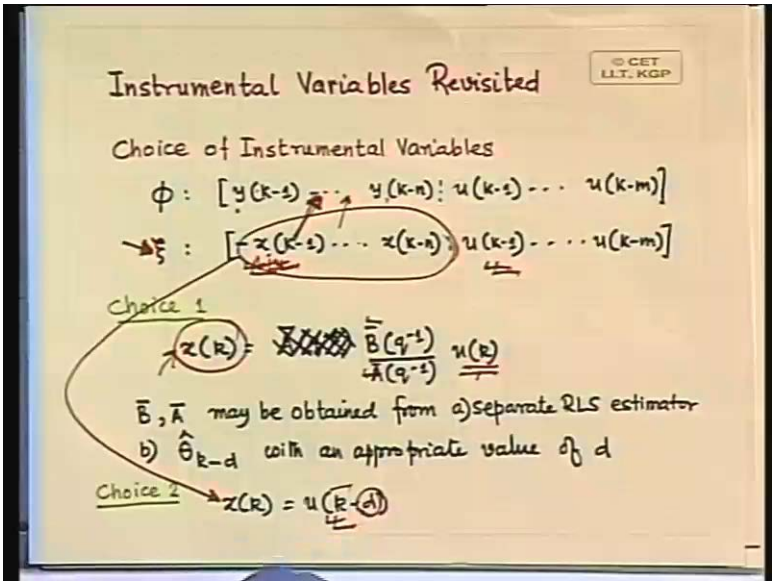
In early model, we have v but we did not have moving average of v; there we assume that the input is only a white noise process, now we are considering an average of white noise process. So I am generating by auto correlation assumption. You know previously I am I mean I put a sting requirement, that we should have a delta like auto correlation function; it should be uncorrelated with even the earlier sample. Now I think that no no it can be little correlated with the past two three samples that is very likely, I mean very difficult to get a totally uncorrelated signal.

(Refer Slide Time: 44:05)



You can, this page I will not going to you can do further generalisations. You can assume a model like this which is which is the most general model you used, totally different CD, totally different B F, you can use. And then you can go along with the same approach; you have and still cast it in phi transpose theta form that, I mean you will have various other vectors. And these are generally not called for they have they it is not in the that simple, I mean they you might experience convergence problems; especially when you have white I mean large amount of noise. So you know, you can do it but it is not you have to do it with caution; just by generalising the noise model you will not get always get more improved estimates. We will skip this, is just to show that, it is possible to even further generalising, it is done in the literature.

(Refer Slide Time: 44:59)



Now we will come to the instrumental variables model, that so so now remember that we wanted to choose a vector x i; which will have that property, it is well correlated with phi and not correlated with v, that is a property that we said if we we have then our parameter are within low. And what is the advantage? That I do not have to extend my phi, remember that my phi phi is the same; only in place of phi and in some place, phi is more than u. See in the earlier approach also I am I am trying to take care of this correlation between phi and u phi and v, but there are modelling v and extending phi and then making my approach more complex; all my vectors are going to be hard dimension, computation load is going to be larger. Here that is not going to happen, here is just contain y and u so nice and small but I have to choose; but you know nothing comes free, so I have to compute this instant, these elements.

So so all so what did what could be generally do is that; just in case of y and u they will here they they generally use different values, because this is anyway uncorrelated with phi. If you are having uncorrelated with v, if you are having open look experience; if you are having close look experience, you have to do totally different approach that we discuss later. We are always assuming that our u and y, u and y have come from open look experience, not close look experience, okay.

So what is so you have to choose this such that, it is well correlated with this but uncorrelated with v that is approach, three eleven thirty-two minutes. So one approach is why do not you put x is the same as u, but little delayed, why do we have to delay? Because, if you if you make if you just repeat this, then this matrix will become rank deficient, you cannot just repeat this columns will become same. So you have to so just before that ranking, you just delayed but in the system how much will you delay?

What is this d? You want this to be well correlated with y. So you have to know, that suppose supposed your system's time constant is twenty seconds; so you know that at least y now is does have some effect of u k minus ten, because time constant is twenty seconds, at least over the last five ten seconds the u and y are going to be correlated. So if you choose the u here and if you choose d of you know, typically less than the time constant of the system; now the point is that, how do you know the time constant?

You you you are trying to estimate the model and you are making assumptions about time constant, this itself is a contradiction but it usually happens and after all the I mean you you have some idea about the system, time I mean estimation of time constant you can do order of time Constance; not exact value you can do very easily, for that you do not need any just just a simple state response experiment will tell you, what is the time constant. What is the order of the time constant, whether it is ten seconds or whether it is hundred seconds or whether it is one second that is very easily determined.

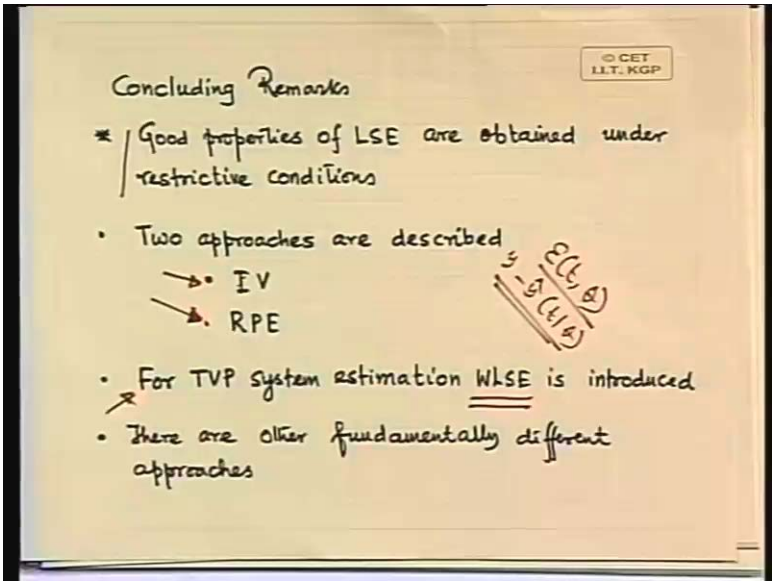
So if you have an idea of that, then you can one very simple choice is just just use delayed inputs; so now this vector only contain inputs, so it is uncorrelated with v that requirement is satisfied and since u k minus d is is naturally going to be correlated with these because you are done the shift within time constant. If you have done the shift, if you maintain this too large then it may become uncorrelated, right.

The other one is that, why do not you first first you do a simple recursive least square? You know that you will get some parameter bias, fine. So you got a biased parameter B bar and A bar,

now use this B bar and A bar to stimulate x using the same u. So obviously it will not match with y; remember that now it now it becomes a pure function of u, B bar and A bar are constants once you have chosen that. So now this x is the pure function of u, it does not use y. So once you use it will become the pure function of u and it is again uncorrelated with v, fact number one established. And second is that even though even though these parameters are slightly erroneous, this x is highly going to be highly correlated with y obviously.

Then there may be some error little bit error, after all the least square by the volts; so another approach is just use use some approximate least square and then go on stimulating x along with y, using the same u and then use this x here. It can give you a significant link better parameter estimate when you have correlated noise this this algorithm really works well, okay. So we will stop here to just to see that, we we only did that that good properties of l c are obtained under restrictive conditions. So we saw that under some other condition, how are we going to work?

(Refer Slide Time: 50:19)



We saw basically two approaches, one is called instrumental variable another is called recursive prediction error approach. Now this prediction that is everywhere we are trying to use epsilon t even theta, which is y minus y hat t even theta. This we are using, for for for various model

structures. So such methods this clause of methods are called, regressive prediction error methods. So we saw some examples of this and this. And for time varying parameter systems; we saw that you have to have a weighted least square in some name, some exponential. But but if I stop here, then it will appear to you as if least square algorithm is the only estimated in the world that is it is not the case; there are they are they are estimators which you can derive under other from from from other philosophies.

So I think it is important to you know, I mean otherwise you always trying to feel that; I mean the least square is the only estimation approach that is not a fact. So we will try to look at at least some other estimation philosophies, like maximum likelihood, like maximum opostelian methods, like Bayesian estimation methods. That is all for today, thank you.