Applied Time-Series Analysis
Prof. Arun K. Tangirala
Department of Chemical Engineering
Indian Institute of Technology, Madras

Lecture – 105
Lecture 45B - Estimation Methods 2 -3

If your z were colored for example, what would be your w structurally look like? Would it be diagonal? So, it will be a full matrix then you have to estimate more Ws, now one of the ways out is when you know that z is colored, assume a time series model for z. So, suppose z as a moving average kind of a structure then what we call is a parameterization of w. So, instead of saying I am going to estimate every element of w, I will say that I know z as a moving average, how do I know that? I performed an OLS and I looked at the ACF. So, it gave me some idea of how z is evolving. So, I assume a time series model for z and from where I can write theoretically the variances of z I can essentially write the theoretical w symbolically and suppose z has an m a one then there are only 2 unknowns to be estimated the c one square and c sigma square e.

W now will be tri diagonal matrix where I would be estimating only c one and sigma square e not the entire elements of the tri diagonal matrix this is a standard idea and a standard trick when the number of a typically in any non parametric analysis you can take it from me in any non parametric analysis the number of unknowns that you will estimate are very high.

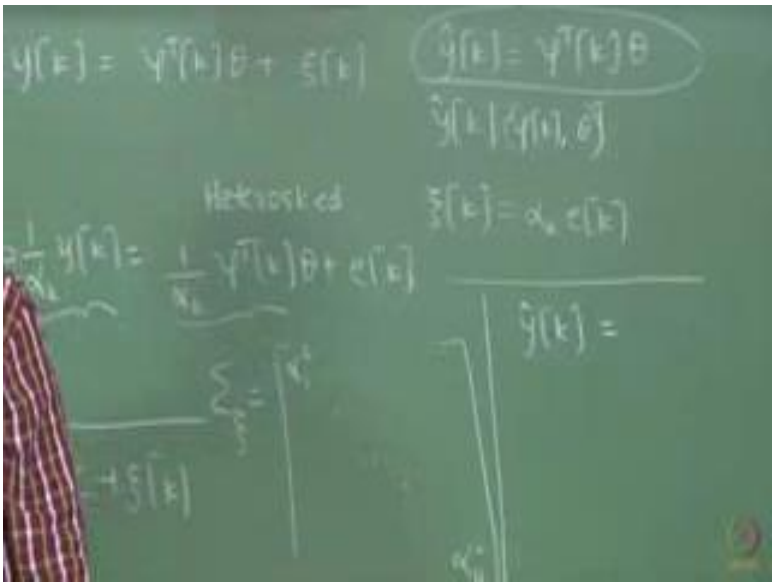
In a parametric analysis because you have poured in some information you say that I know the structure of something and you are saying that I will explain all these unknowns with the help of some function with 2 parameters or three parameters or. So, on the parameterized problems are always simpler to solve, but then somebody or somewhere you have to procure that information and that information has to be right whereas, the non parametric the advantage is you will see that when we talk of estimation of from spectral densities the same story comes about.

(Refer Slide Time: 02:29)

But then the moment you parameterize z then what you are doing is for example, if you are fitting, if z consists this, psi consists of let say past data like in AR models and z has been parameterized, let us say to an m a form and so on, essentially what you are doing is you are fitting an arma model. So, you might as well turn to the estimation of arma models and when you do that you enter the world of non-linear least squares which is what I am going to quickly discuss today.

This concludes the discussion on weighted least squares I slowly motivated you to look at no linear least squares, but there are many many situations where non-linear least squares comes into play essentially the difference between linear least squares which is what we have been discussing until now and non-linear least squares is this candidate here.

(Refer Slide Time: 03:17)



What happens now is that y hat is some non liner function of theta the issue is not non-linear functions of regressors that is not the issue at all the issue is now that this is some non-linear function you can say s or whatever, but g whatever function you want to call it, it is a non-linear function.

So, what is the problem if this is non-linear, why is it? So, special why cannot I use the methods that I have used in linear least squares, what is the difficulty? this is not an estimator, this is only a predictor, I am saying if the predictor is a non-linear function that is an issue we cannot use the exe the solutions that we have used until now and why how did you solve the linear least squares problem how did we solve this problem let say there is no lambda how did we solve this?

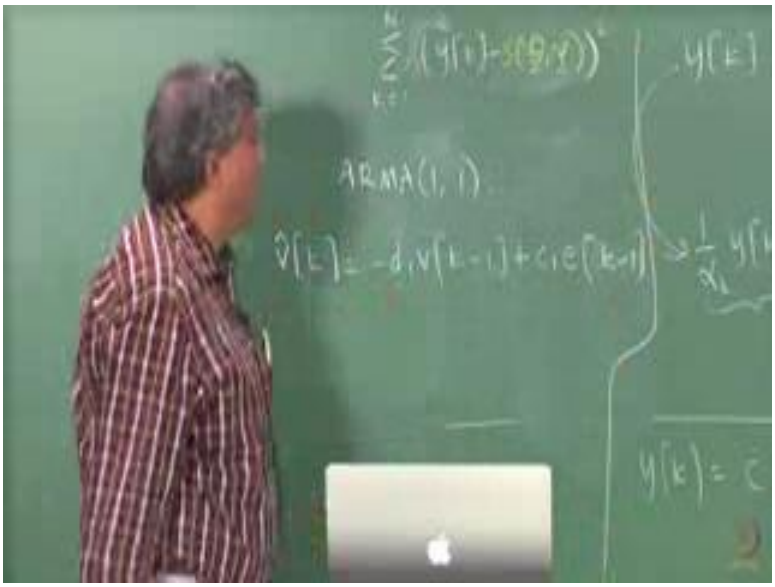
Student: (Refer Time: 04:16)

We have used projection theorem, fine suppose I did not use the projection theorem, what is the natural approach? Take the derivative when you take the derivative of this function here objective function what do you get and you take p derivatives for p parameters p partial derivatives what kind of equations you end up with?

Student: linear equation.

Linear equations p linear equations p unknowns life is happy ever after correct, but unfortunately life is not happy ever after with the non-linear one why because now instead of the psi transpose k theta here now.

(Refer Slide Time: 05:03)



I have a non-linear function sitting, I mean the k's are there, but now you should see what is happening when you take the derivative of this objective function with respect to theta do you end up with this happy merry go around linear equations you do not right you will end up with a bunch of non-linear equations and we know already the pains associated with solving a bunch of non-linear equations.

And that is the first problem with this non-linear least square says that I do not end up with a bunch of linear equations which means I cannot first of all write an analytical solution that is the big problem when it comes to computation there is no unique solution I have to use non-linear optimizers and that is optimizers meant for solving non-linear least squares problems I may get local minima and so on.

the side effects of not having analytical solution come up and that is the main drawback of working with non-linear least squares, but you have no choice if you know it is a a non-linear predictor it. So, be it you cannot do anything about it if possible you would work like to work with a linear one, but if the situation demands. So, for example, if you have to fit an arma model what happens to the predictor if I am working with an arma

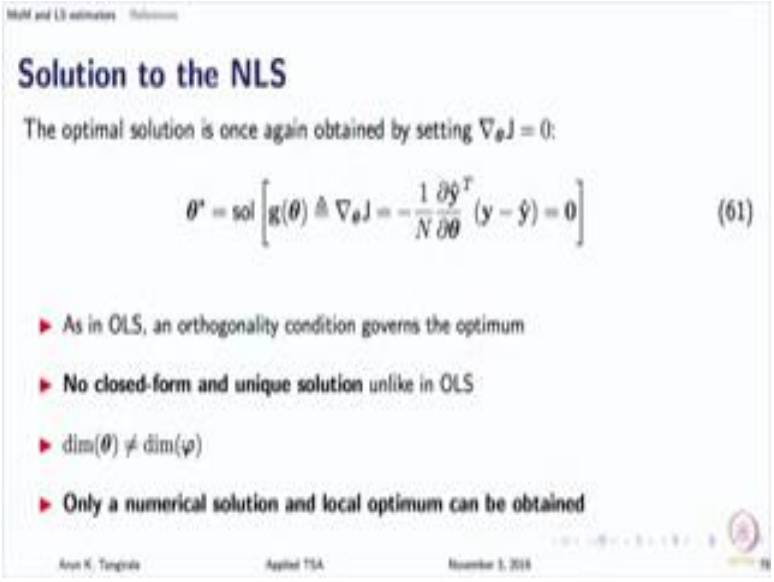
one, then I know the predictor v hat of k is going to be minus d one v k minus one plus c one e k minus one right this is going to be my predictor.

Now, when I look at it is it linear in the unknowns is the question what do you think remember this is known to me in time series this is the parameter these are the parameters of interest this is not known to me. So, I do not have this. So, I do not know what is this also?

But we know already when through our discussions on invertibility that e k minus one is an it has to be can be re written in terms of the past v and when rewriting that the model again comes into play that is your c one and d one will come into play; that means, now the right hand side is a complicated function of your theta because this e k minus one is now if I were to rewrite this in terms of theta and knowns, knowns are always my data and if were to rewrite this in terms of theta and data they rhyme well, but unfortunately you end up with a non-linear function and this is a standard thing even if you did not have the a r component with the moving average model you would end up with the ah with a non-linear function that is even if this was absent.

in time series modeling you will routinely run into non-linear least squares there is no escape therefore, now we will have to ask how to solve this there is nothing much.

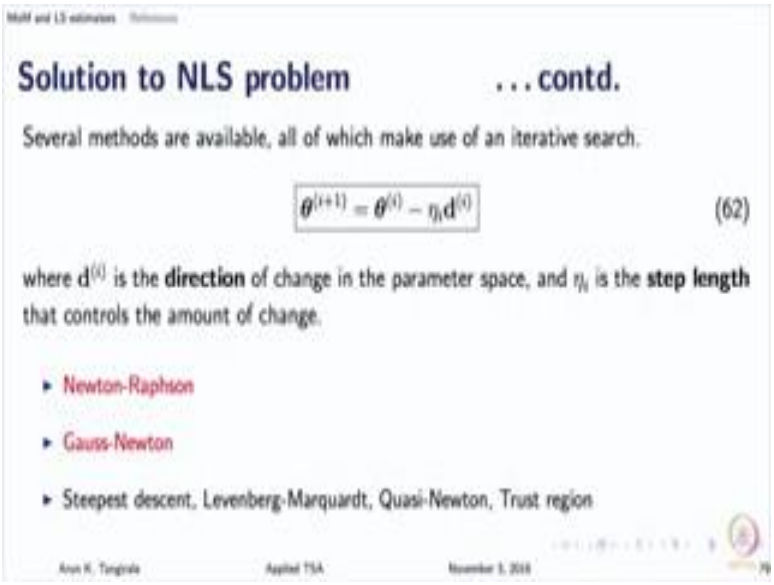
(Refer Slide Time: 08:30)



Your standard Newton, Raphson method, Gaussian, you have a bunch of non-linear equations. So, question is how to solve this non-linear equations right. So, you can see where do these equations come from I have taken the derivative of the objective function with respect to theta set them to zero your dou y hat by dou theta you should place attention on that it is the gradient of the predictor with respect to theta in the linear case what is it the regressor, your psi which is independent of theta am I right. So, therefore, I would get a bunch of linear equations the difficulty lies with this gradient here.

Now, as far as the solution to non-linear least squares is concerned one uses standard methods like Newton Raphson method or a Gauss Newton method. In fact, a modified Gauss Newton method or a Liebenberg Marquardt method, all of them, then are, they are iterative algorithms only they essentially set up, they start with an initial guess refine the parameters.

(Refer Slide Time: 09:43)



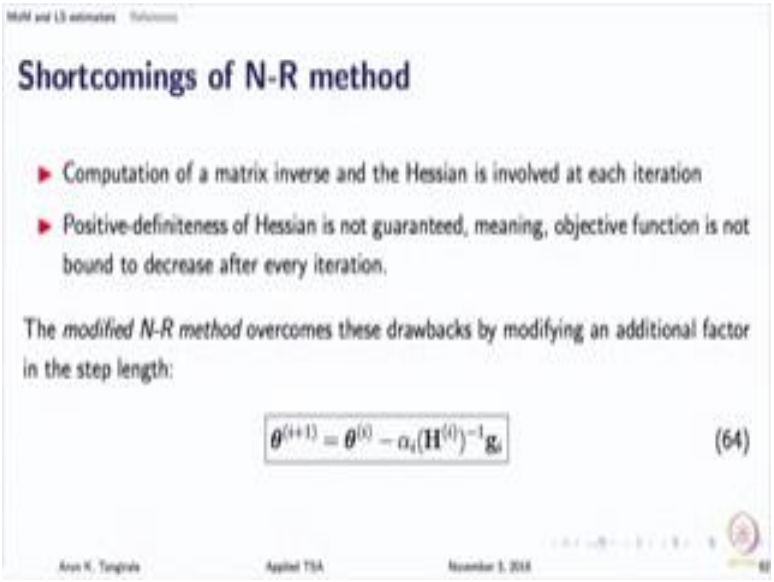
You have Newton Raphson method which search. So, the generic form of this update equation in these methods is of this theta I plus one is theta I minus or plus does not matter eta times some direction method. So, it needs a direction to search Newton Raphson method offers one suggestion for the direction Gauss Newton method says no no look at in this direction steepest descend will look we will ask you to look at in some other direction and so on.

Essentially the formula is the same and eta is a parameter, user defined parameter that has been tuned and people have studied, a lot of PhDs have gone into this in tuning these parameters to the effect that you have a very good algorithms today, but very good does not mean that they will get you unique solutions and so on, it is essentially the solutions sometimes, some of the most sophisticated solutions make themselves robust to the initial guesses, 1 of the drawbacks of this non-linear solvers is at sensitivity to initial guesses and if you are able to generate a very good initial guess then you are guaranteed that you will get a good solution, but if you have generated a poor initial guess then it may be stuck there within the vicinity of it, sometimes you may even get absurd solutions.

There is a big area that of lit research which focuses on generating initial guesses next week when we talk of how to estimate time series models using non-linear least squares even m l e gives rise to non-linear optimization problems we will talk about generating good initial guesses particularly for estimating arma models there are some algo for example, I can use a Yule Walker's method, we have talked about this Yule Walker's method also gives rise to a bunch of non-linear equation, but they are a bit easier to solve. So, I solve the Yule Walker problem and then feed that to my non-linear least squares and we find the guess, but whatever you do the general ah formula that is used for updating theta is this.

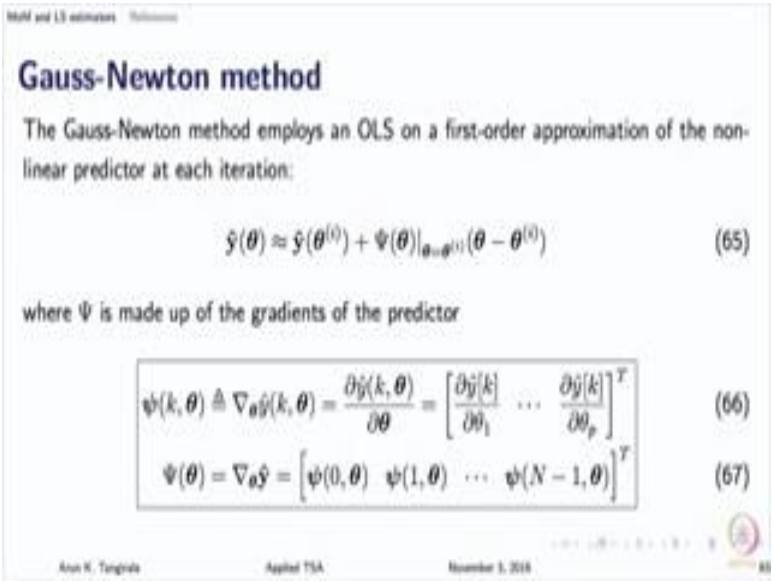
Therefore, you just you can just go through a quick review of the Newton Raphson and Gauss Newton method.

(Refer Slide Time: 12:10)



These are some standard methods that you should have learnt in some numerical course, the Newton Raphson method works with what is known as a hessian the direction is the hessian the hessian is essentially the second derivative of what of the objective function with respect to theta it is searching in this direction.

(Refer Slide Time: 12:32)



Gauss Newton method solves a locally linear least squares problem that is why it is very attractive, it says you have given me initial guess, I will approximate; I will construct a linearized version of s around that theta that is the beauty that is the 1 way of looking at

Gauss Newton method, there are 2 perspectives to Gauss Newton method, one is that it is going to replace the Hessian calculations in Newton Raphson method with something else, but the other perspective that I prefer is that it is solving a locally linear least squares problem at every guess of theta it is linearizing your predictor solving a least squares problem generating improving your guess around the guess again it is also linear least squares problem and continuous that is the basic idea that is why you see this equation in 65 theta I is the ith guess at the ith iteration and what I am constructing is y hat of theta using Taylor series expansion, I am constructing a first order I am only taking the first order terms and constructing a linear one. So, now, this is linear you see, but what is appearing here this big this is now not your phi do not get confused between phi and psi.

At some point in time if you have not yet felt that the course is heavy it is my duty to make sure that you feel that way and the literature is also such to bombard you with too many symbols. So, here I have psi and then there is phi, phi is the matrix of regressors, this big psi is a matrix of gradients they are one and the same in the least square in the linear least squares in the non-linear least squares case they are not ok.

The more general set of regressors is this big phi that you also see sometimes in quantum mechanics and since it is not as deadly as that, but essentially the point is when you move from linear least squares to non-linear least squares you are moving from the matrix of regressors to the matrix of gradients, but this gradient is not of the objective function this gradient is of the predictor with respect to theta if you are working with a linear least squares problem the predict this gradient of predictor with respect to theta is the same as the regressor.

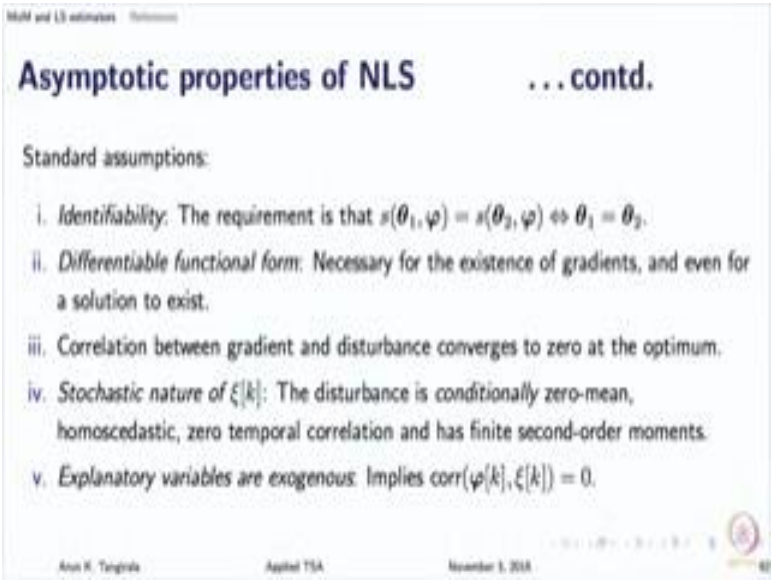
Why am I even focusing on this course on emphasizing this part so much because ultimately when you look at the properties of the non-linear least squares estimator such as consistency efficiency and distributional properties you will see that they bear a striking resemblance with those in the linear least squares case.

In fact, it is not easy to come arrive at those properties people have broken their heads on arriving at the distributional properties they are only available for large sample cases. So, you should read some classic books by Ameya and so on, they are very widely cited or the more recent ones I have given a few references the point is whatever results that we

have seen for example, if I look at variance of theta hat in OLs, we had sigma square e times phi transpose phi inverse

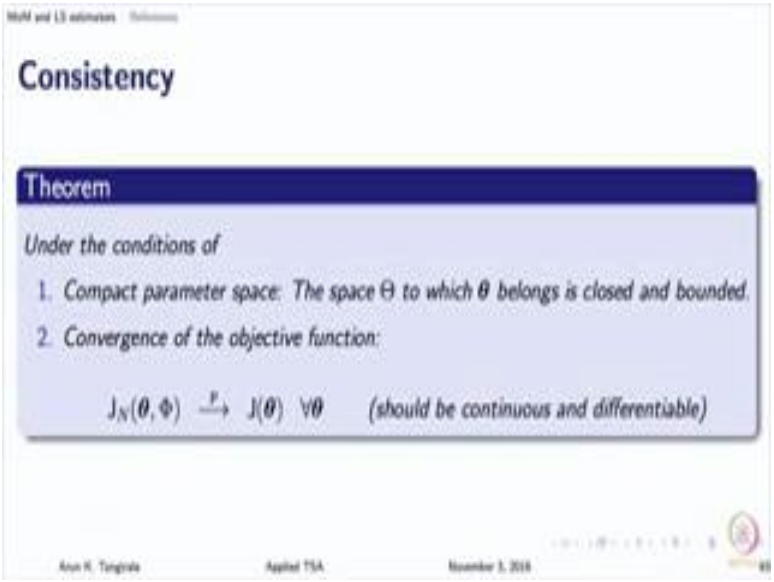
In the non-linear least square case also you will see a similar result if I were to say what is the variance of the theta hat of the estimate that you have obtained it is once again sigma square e times not phi transpose phi, but the psi, but the big psi transpose psi inverse that psi is the gradient of the predictor evaluated at the final estimate that you have obtained that is all assuming that the final estimate is the divine one is the final solution that you wanted it is not we know very well it is a local optimum, but there is no other choice to be practical. So, that is why I am just going to skip this Gauss Newton method I have explained the concept I just want to I have spoken about the gradient I just want to conclude the lecture with the asymptotic properties.

(Refer Slide Time: 16:45)

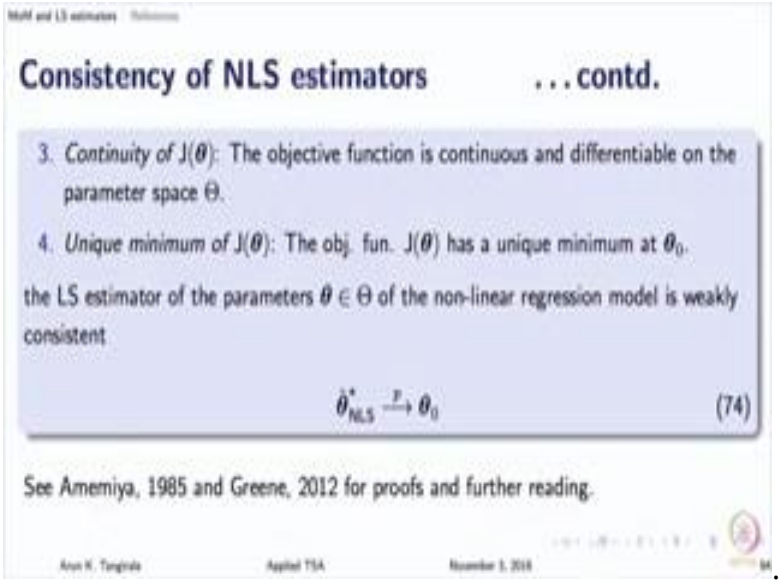


There are certain conditions standard conditions, you know the most important thing is that the correlation.

(Refer Slide Time: 16:49)



(Refer Slide Time: 17:03)



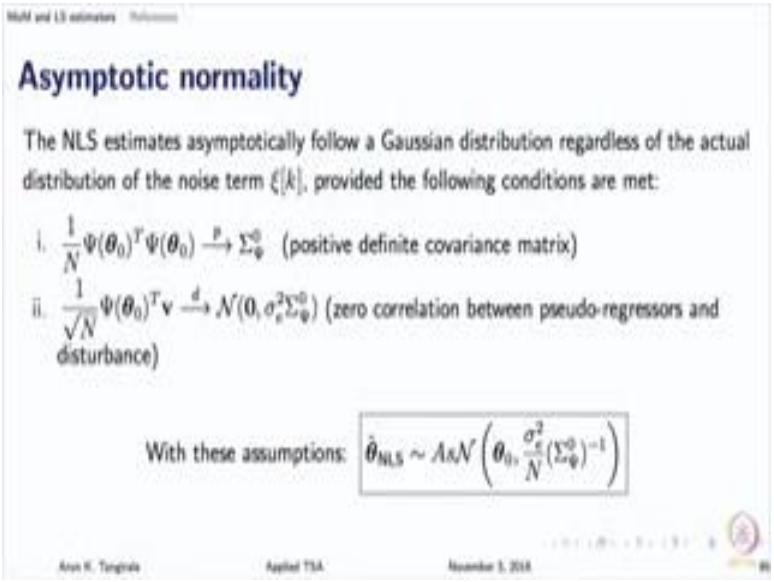
For example, if you look at consistency which is the most important thing it says that forget about the first one, the essentially it talks about the parameters, they should be in a closed sub space, but the more important thing is that the co regressors should be uncorrelated with the residuals that also applies here and it says under these several conditions which are mostly conditions demanding how the objective function should be it should be continuous and that there should be a unique minimum all of this has to be guaranteed.

If a unique minimum exists then theta hat star will go and sit at that unique minimum as n goes to infinity provided the again whatever you have left out is not correlated with what you have included and there is one more subtle aspect that you should understand in non-linear least squares in linear least squares we have used the terms regressors and explanatory variables and we have said more or less they are the same, but in non-linear least squares they can be different their explanatory variables are let us say you know pressure all or some volume and so on. So, suppose I take the ideal gas law I have temperature and pressure and volume readings the relationship is non-linear right, but when I look at the explanatory variables for temperature they are pressure and volume, but when I look at it as a regressor when I ask what is a regressor the regressor is not pressure and volume the product of pressure and volume.

That is the difference that you should observe. So, what happens when you start off the problem? You have pressure and volume, you have 2 explanatory variables in a linear world you will have as many parameters to estimate as the number of explanatory variables, but in the non-linear world you need not have in a non-linear world depending on the number of regressors in a ideal gas law case p v is your regressor and there is only one coef parameter estimate which is a universal gas constant that is all.

The number of parameters and the number of explanatory variables need not match. So, I will I will just close the discussion with this.

(Refer Slide Time: 19:04)



With this result, here on the distribution property, you can see they are quite similar to what we have seen in the OLs except that now we are saying that the regressor mate instead of saying regressor matrix should have should be should have a non singular covariance matrix here we are saying that the gradient of the predictor matrix should be having a non singular covariance matrix and it should be uncorrelated between the regressor the pseudo regressors and the disturbance under these 2 conditions you are guaranteed that fortunately even in the non-linear least squares case the estimates follow a Gaussian distribution this is what is used in calculating the errors for you when you are estimating arma models we have already seen what is the case for AR models.

All you have to do is a non-linear least squares, if you do not understand anything, simply keep go to the linear least squares and keep replacing phi with the pre gradient of the predictor with respect to theta that you have obtained at the optimum that is all, the rest of the solutions will properties everything have a striking resemblance, as long as you stick to that hang on to that perspective, things are simple, do not try to get into technical proofs, if you are interested of course, you can, but final result is this. So, with this we come to a close on least squares.