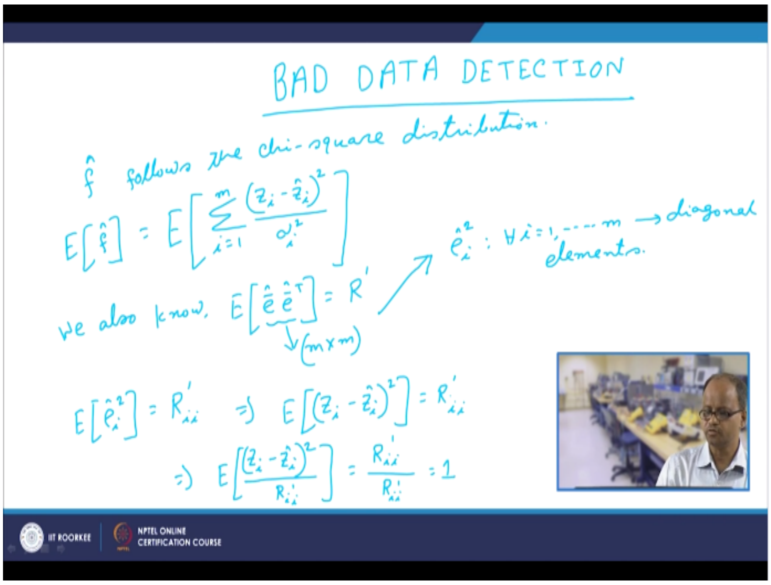
Computer Aided Power System Analysis
Prof. Biswarup Das
Department of Electrical Engineering
Indian Institute of Technology – Roorkee

Lecture - 44
Bad Data Detection

Hello friends. Welcome to this lecture on computer aided power system analysis. In the last lecture, we have discussed the concept of chi-square distribution and we said that this estimated value of the objective function follows the chi-square distribution and then we have also discussed the need of the bad data detection. So in today's lecture, we would be looking into this aspect of bad data detections.

(Refer Slide Time: 01:01)



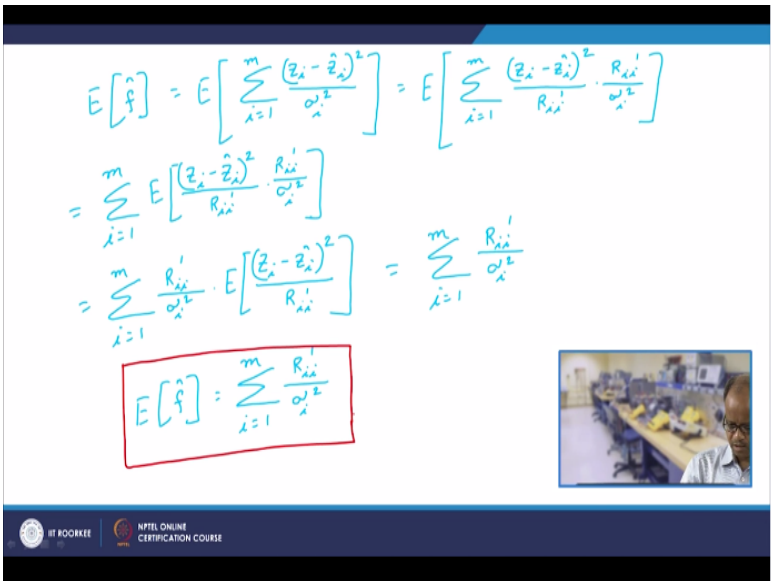
We said that this f hat, f hat we said that f had follows the chi-square distribution, now although f hat follows the chi-square distribution because it is a random value and because it follows a certain distribution probability distribution function so then therefore it will also should have a certain parameters so we would like to know that what is the mean value of f hat. So what we want to know that what is the value of expected value of f hat.

So that would be obviously something like this that would be the case. Now we also know that expected value of we have already seen in the last class or rather the last lecture that e hat transpose is actually a matrix R dash and we have also seen that this quantity that this is an m x m matrix so this is a square matrix. So then therefore and because it is a square matrix so then all quantities that ei hat square for all i is=1 to m are the diagonal elements correct.

So these are all diagonal elements and this matrix is also an m x m matrix that also we have already seen. So then therefore this is also an m x m matrix so this is also a square matrix. So then therefore we can say that expected value of ei hat square is actually Rii dash that is the diagonal element of this. So then therefore, we can say that expected value of zi-zi hat square is=R dash ii that we can say.

So then therefore I can say that expected value of zi-zi hat square/Rii dash that is=Rii dash/Rii dash is=1. Now if I utilize this into this expression, so from there what do I get?

(Refer Slide Time: 04:37)



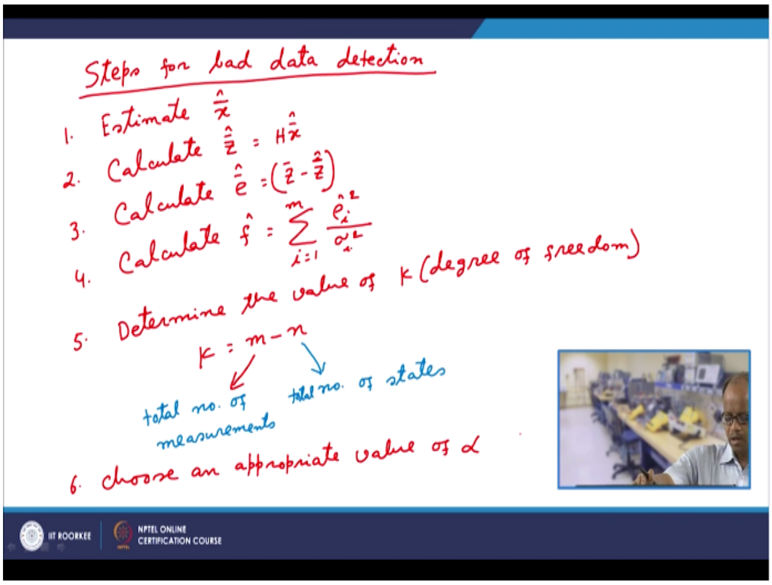
We get that expected value of f hat so then therefore expected value of f hat is expected value of 1 to m zi-zi hat square/sigma i square that is we have already seen. Now we write that this we can write as Rii dash*Rii dash/sigma i square which is entire quantity is within this summation. So this is the expected value of this, so then therefore this is nothing but we can write down that i=1 to m expected value of zi-zi whole square/Rii dash*Rii dash/sigma i whole square.

Now this is something into this so then I can take it out so then therefore I can write it down i is=1 to m Rii dash/sigma i square*E zi-zi this would be hat sorry zi hat square, I must write it more properly to expected value of zi-zi hat square/Rii dash. Now this quantity we have just now seen that this quantity is=1 so then therefore it turns out to be that i is=1 to m Rii dash/sigma i square.

So then therefore this is a very important relation expected value of f hat that is the estimated value of the objective function is given by i is=1 to m Rii dash/sigma i whole square to this is a very important relation anyway. So by this we can calculate that what would be the mean or average value of the estimated value of the objective function.

Now let us again come back to our topic of discussion that we would like to estimate or rather we would like to find out that which particular measurement data is bad. Bad means that which particular measurement data has got unacceptably large value of error.

(Refer Slide Time: 08:37)



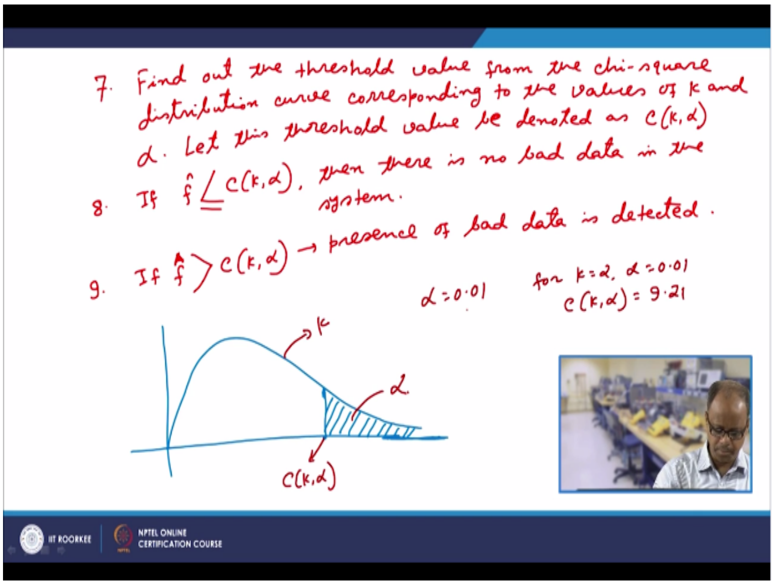
Now what we do is for that we do follow some simple steps, so we write that the steps for bad data detection is like this. Steps for we have already covered all the theoretical background so now we are ready to write down the steps for bad data detection. What are the steps? Step 1 is estimate x hat that is we get the measurement and etc, estimate x hat or rather calculate x hat.

We can write down calculate x hat, from that we can calculate z hat, z hat is H*x hat. From that we can calculate e hat, e hat is z-z hat can calculate. Once we get e hat then we calculate f hat, that we know that how to calculate 1 to m so calculate f hat. Once I calculate f hat then we now consult our chi-square distribution curve. To consult our chi-square distribution curve first, we determine what is the, determine the value of k.

What is k? K is degree of freedom. Now how do I determine the value of k? k is nothing but m-n. What is m? m is the number of measurement, it is the number of measurement and what is n? We should say total number of measurement and we should say it is total number of states. So we know that I mean what is the total number of states and we already know that how many meters I have put so k is=m-n.

Next step is choose an appropriate value of alpha. What is alpha? Alpha is basically the probability which I wish to maintain. Alpha is the value of the probability which I wish to maintain.

(Refer Slide Time: 12:01)



So then what is the seventh step? Find out the threshold value from the chi-square distribution curve corresponding to the values of k and alpha. So we find out that what is the threshold value. We have already shown in the last lecture that corresponding to any particular combination of k and alpha, there is a certain threshold value. So we find out the threshold value.

Then what you say? If let this threshold value is denoted by, let this value be denoted let us say this is c k, alpha. This c stands for this chi-square distribution curve and k and alpha are those values. If so now after this we do if the value of f hat is <c k, alpha then we say that there is no bad data in the system and if there and if this value if f hat is>c k, alpha then presence of bad data is detected.

Now what is the philosophy for this, so I mean these are the very simple steps. Now what is the philosophy for it? Please understand that again I mean let us look at this curve. Let us look at this curve and this curve looking like this. Let us say this is this, this is the curve and we have some value of alpha and this is my value of alpha, so this area is alpha and this value is c k, alpha and this is alpha and let us say this is particular value of k.

So what does it mean? Now suppose if say alpha is=0.01 suppose if I say for example in our table we said that for k is=2 and alpha is=0.01, c k, alpha is=9.21 that we have already seen in the last lecture. So what does it mean? It means that the calculated value of f hat has got only 1% chance that it will go, it will have a value of more than 9.21. Now we need to understand that what is the occasion at in which this value of f hat will keep on increasing.

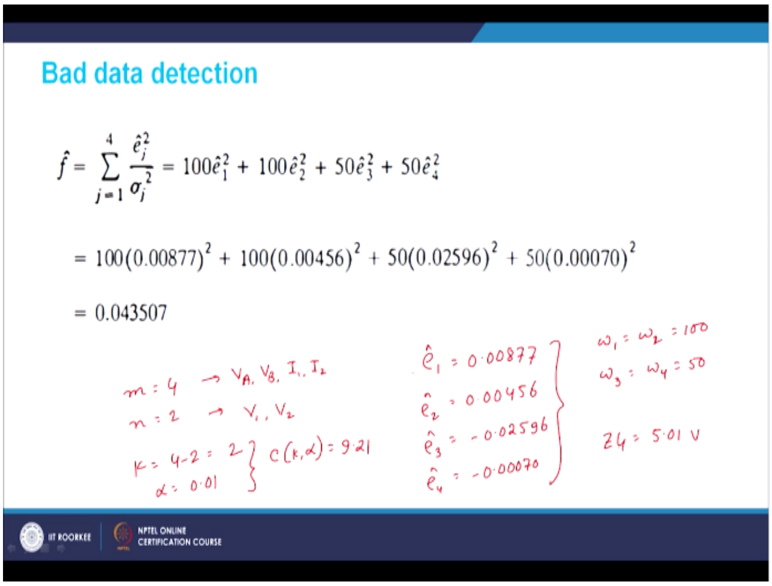
Please note that f hat is nothing but this quantity so then therefore this value of f hat will only increase if this value of ei hat increases. After all sigma i square is constant, so then therefore f hat only increases with this value of ei hat increases, ei hat will only increase if there is large difference between the actual measurement value.

And the estimated measurement value and this large difference will only happen if and only if some measured data is absolutely erroneous and that particular erroneous data makes all the estimated states to be quite erroneous which are not at all close to their actual value and because this estimated states are quite erroneous so then therefore this estimated value of the measurements become quite erroneous.

And because this estimated measurement values become quite erroneous so then therefore this vector e hat will also become quite erroneous or rather large so then each and every element of ei hat would be large in magnitude so then therefore f hat would be large, so then therefore this probability is pretty small, it is just 1% so then under normal condition it is most likely that the calculated value of f hat will be always less than this particular threshold value of 9.21.

But if this value calculated value of f hat crosses this threshold value that means that this calculated value of f hat is quite high or in other words at least some value of ei hat is quite high. So that means there has to be this probably at least 1 meter which has gone bad. So it detect the presence of bad data but now the question is how do I detect that which one has got bad data? So for that what we do is that we simply calculated the value to detect bad data.

(Refer Slide Time: 18:43)



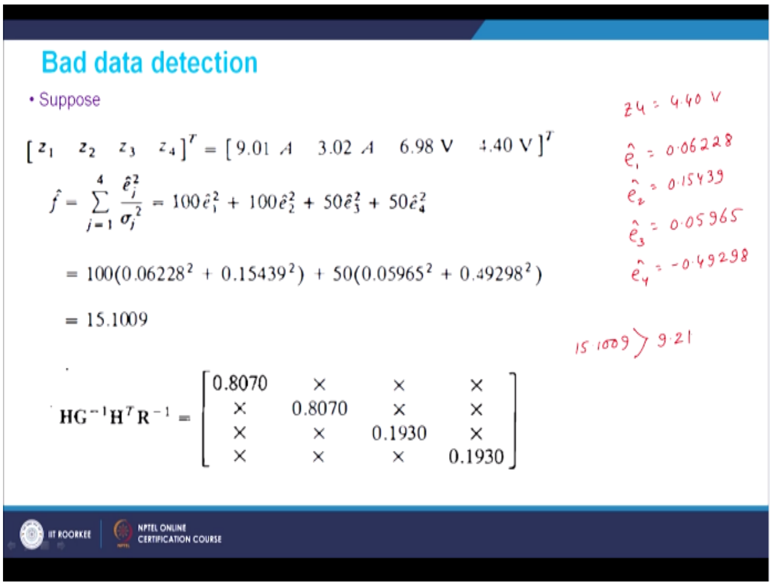
Then, what we do is so in this example if we look at if you do remember in the first case basically our measurement values were so in the first case in the basically last to last lecture we have considered two cases, you have considered a simple circuit and we have considered two cases. In the first case, there were some meters and there were some values and for that values we have calculated all the state vectors.

And once we calculated all the state vectors then we have calculated this e hat vectors and those e hat values were e1 hat was=0.00877, e2 hat was given by 0.00456, e3 hat was -0.02596 and e4 hat was -0.00070. So this is what the estimated errors in the first case and we have also said that and our (()) (20:14) omega w1=w2=100 and w3 is=w4=50 that we have already seen, we have already taken.

So when I utilize this values, we get this is the value of f hat so it was 100 e1 hat square+100 e2 hat square+50 e3 hat square+50 e4 hat square and if you do substitute these values we get these values. Now in our example what we had? We have got total number of measurement was 4 because there are 4 meters, 2 meters and 2 old meters and my total number of state was 2, so this measurements were V1 I think this was VA, VB and I1 and I2 and these states were V1 and V2.

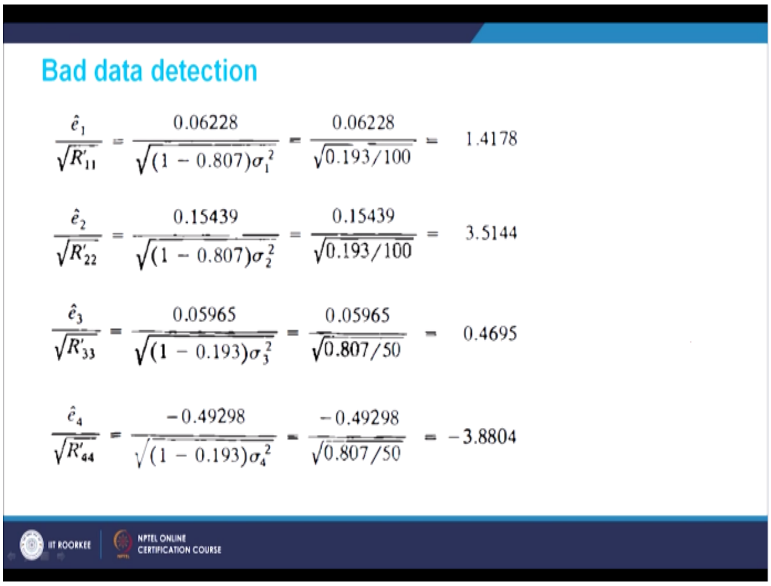
So then therefore total number of measurement was 4 and then number of states is 2, so then therefore k is=4-2 that is 2 and you have taken let alpha is=0.01 and corresponding to this c k, alpha is 9.21. Now you can see that the calculated value of f hat is much less than this threshold value 9.21. So then therefore we say that there is no bad data is detected. Now this was the case when our last measurement value was 5.01 volts, so this is for VB.

(Refer Slide Time: 21:52)



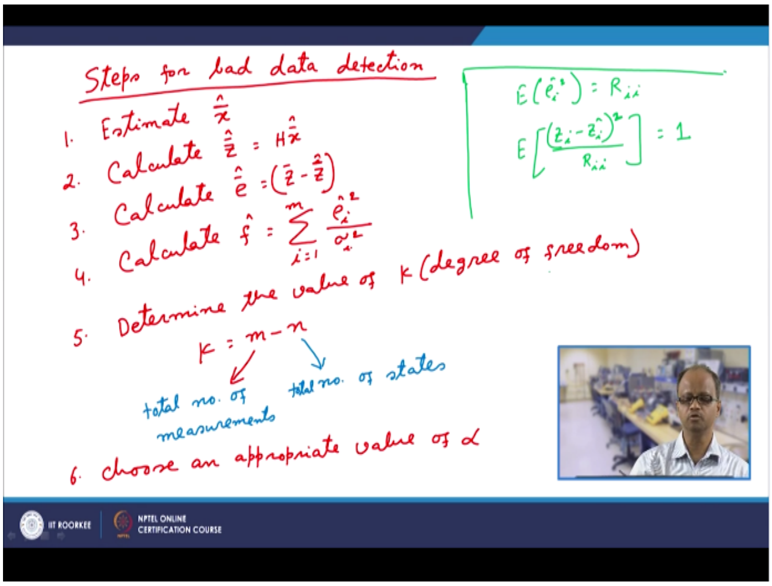
Now suppose you have got another case, we have shown another case where all the other measurements were the same but now z4 was 4.40 volt. So for these we have also shown that the value of estimated errors were 0.06228 then e2 hat is 0.15439, e3 hat is 0.05965 and e4 hat is -0.49298. So these were the values and now when I put these values, I get that f hat is 15.1009 and 15.1009 is much greater than 9.21, so then therefore bad data is detected. So now the question is how do I calculate, how do I find out that which one is the bad data?

(Refer Slide Time: 23:18)



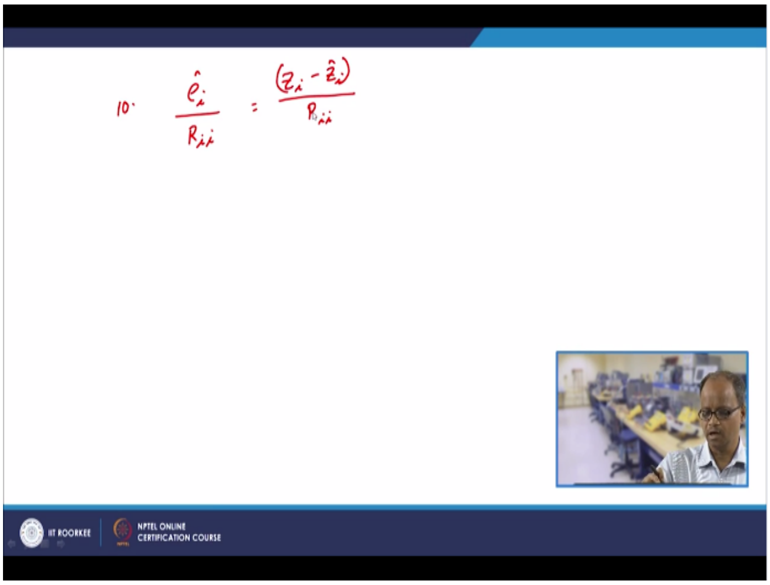
So for that what we do is we actually calculate so that is what we are seeing. We actually calculated the value of e1 hat/root over R11 dash.

(Refer Slide Time: 23:32)



Now here we need to remember that expected value of ei hat square is actually Rii. So we said that expected value of we said that zi-zi hat square/Rii is=1 that we have already seen. So then what we are trying to do but this is the expected value.

(Refer Slide Time: 24:20)



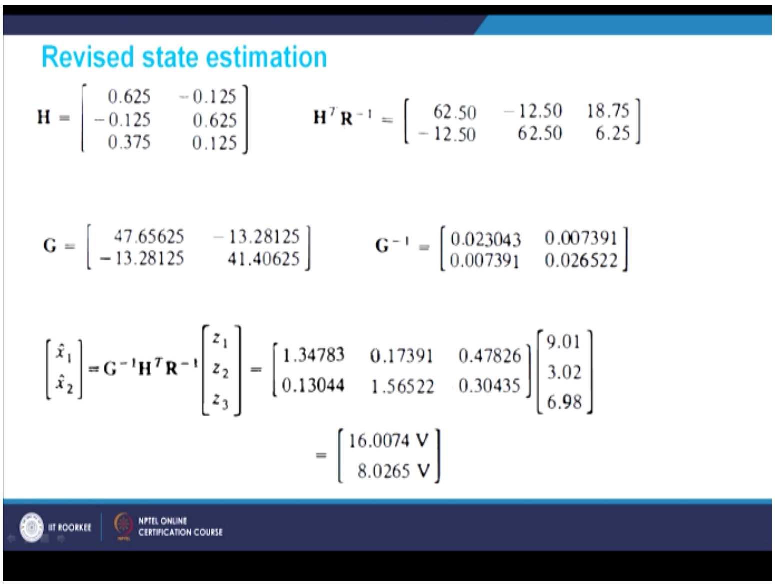
But then what is but in that case what we calculate is we actually calculate when there is bad data is detected we actually calculated something called ei hat/Rii so we calculated zi-zi hat/Rii. Once we calculated this value and then we find out that which one is the highest and which one is the highest that one we do take as the. So now what we do is that we do calculate e1 hat/root over R11 dash.

So this we calculate, so from this we calculate this, from this we calculate this, so then we are simply weighted value of this estimated errors we have calculated. This is called weighted values. So then we calculated these weighted values e1 hat/root over R11 dash, we do to calculate e2 dash/root over R22 dash, e3 dash/root over R33 dash, e4 dash/root over R44 dash. Remember these quantities are nothing but the R11 dash, R22 dash, R33 dash and R44 dash are nothing but the diagonal elements of the R dash matrix.

Because these are the diagonal elements of the R dash matrix so then we can easily calculate them so this is already known. So from these we calculated these values and we find out that these are the values. Now out of this which one is the highest, out of this magnitude wise this one is the highest although this is positive this is negative but as far as this magnitude is concerned this is highest.

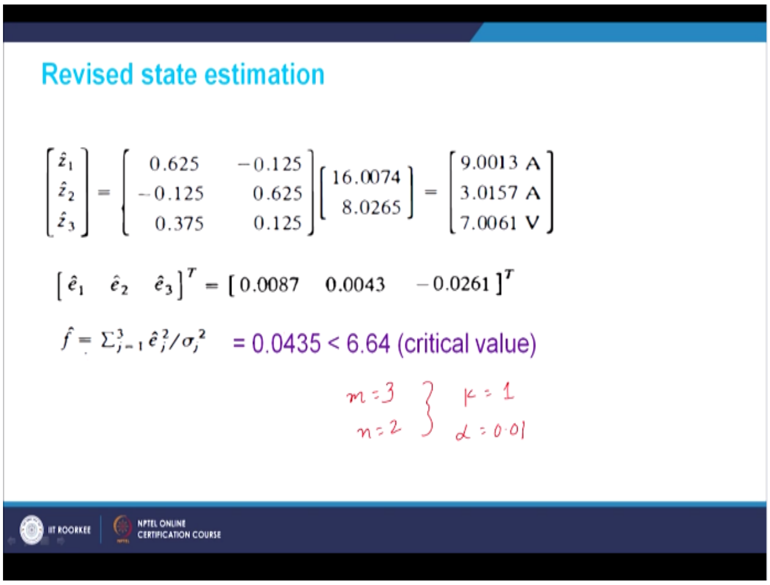
Because this is highest so then therefore we conclude that this is the bad data. So this data must be the erroneous so we simply discard them.

(Refer Slide Time: 26:33)



So once we discard them then after that so now we have got actually 3 measurements with us, so because we have got now 3 measurements with us so then we have got now this H matrix is only 3 so after that we again do all these calculations.

(Refer Slide Time: 26:51)



Once we do this x1 calculation, x1 hat x2 dash after that we calculate e1 hat, e2 hat, e3 hat this and then we again calculate f hat is equal to this which is 0.0435. Now what happens in this case, in this case now my m is actually, now my m is 3 and n is=2 so my k is=1 and if I keep alpha is=0.01 so then therefore if I look at this graph so for k is=1 and alpha is=0.01 this threshold value is 6.64 because this threshold value is 6.64 and this is less than this so now there is no bad data.

So therefore these estimated values are reasonably close. Please note that when we have done our estimation with 4 measurements and we said that there is no error, these estimated values were coming out to be quite close to be 16 and 8, their true values are 16 and 8 and now when we have already detected one bad data and removed it and done this particular state estimation with this value of only 3 measurements available.

So then therefore we have got again these same values which are again close to this estimated quantities or rather which are close to this true values. So that is how we do this bad data detection. So we will continue this particular aspect in the next lecture especially the reason for this calculation is not still explained in a great detail. So in the next lecture, we would be looking at this aspect as I mean we would be actually looking at this aspect in the next lecture in more detail. Thank you.