Bayesian/MMSE Estimation for MIMO/OFDM Wireless Communications

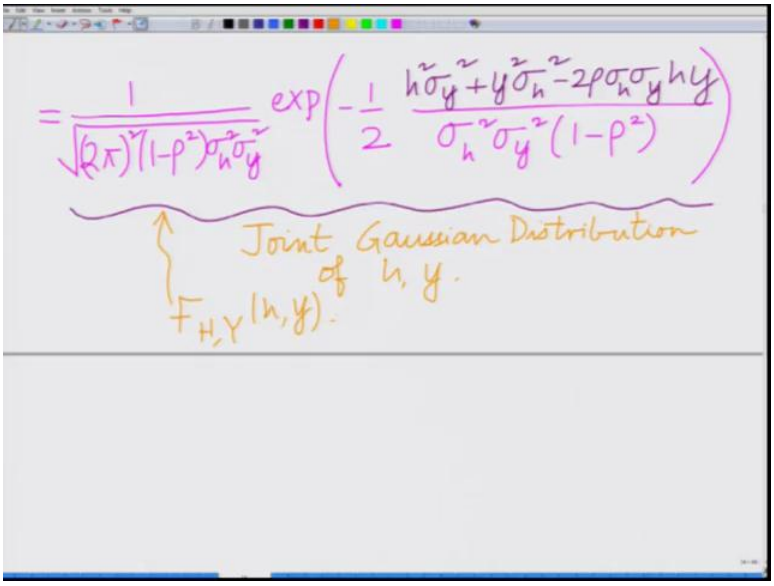
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture - 04

Derivation of Minimum Mean Squared Error (MMSE) Estimate for Gaussian Parameter – part II

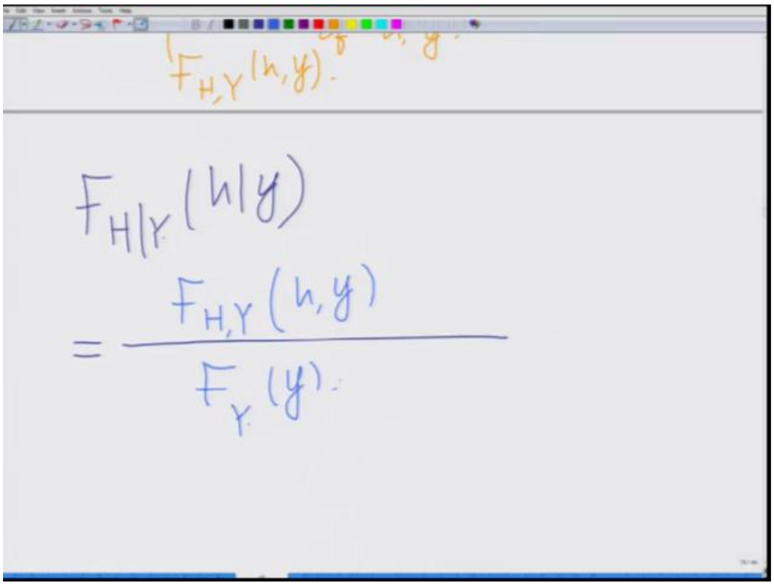
Hello welcome to another module in this massive open online course of Bayesian MMSE estimation for wireless communication. So, currently we are looking at the m MSE estimate of the parameter h and, from an observation y when the parameter h and observation are jointly Gaussian, and towards this end we have derived the joint density of this h parameter h and the observation y, and that is given by this quantity.

(Refer Slide Time: 00:43)



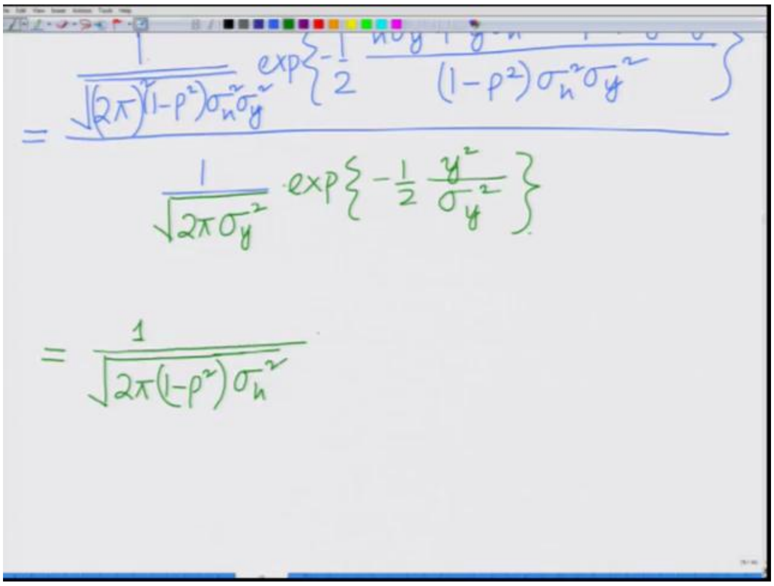
Given by this quantity that is shown here, that is done by previous module that is the joint probability density - The joint probability density function of the parameter y of the parameter h, and the observation y in the parameter h, and the observation y are the jointly Gaussian.

(Refer Slide Time: 01:06)



So, now as said before we need the conditional probability density function. I mean the conditional probability density function of in the parameter h, given the observation y, which is basically, as we have already said this is the joint probability density function of the parameter h observation y divided by the marginal probability density function of the observation y. And this now I am going to simplify by writing the exact expression for each.

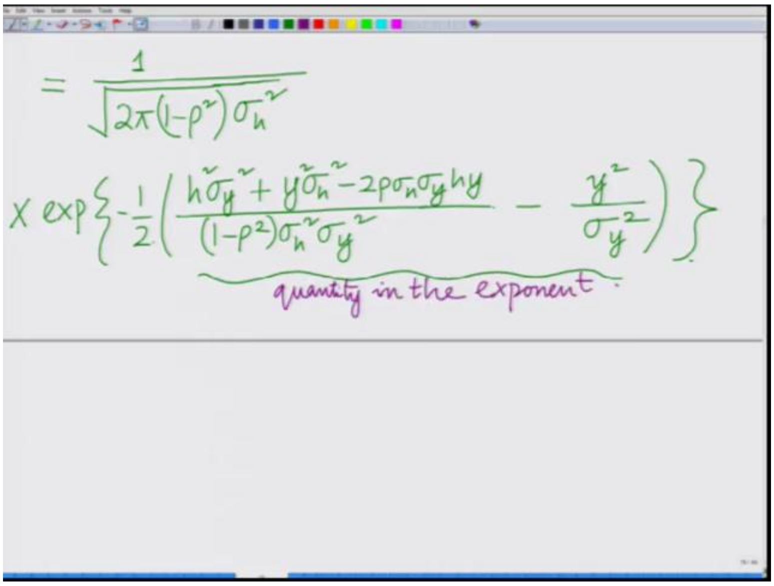
(Refer Slide Time: 01:35)



So, first is the joint probability density function that we have already derived the expressions. so that is square root of 2 pi whole square 1 minus rho square, that is your determinant r, which is sigma h square sigma y square e to the power of minus half 1 minus rho square sigma h square sigma y square h square sigma y square plus y square sigma h square minus 2 rho sigma h sigma y into h y divided by the marginal probability density function of y the observation y.

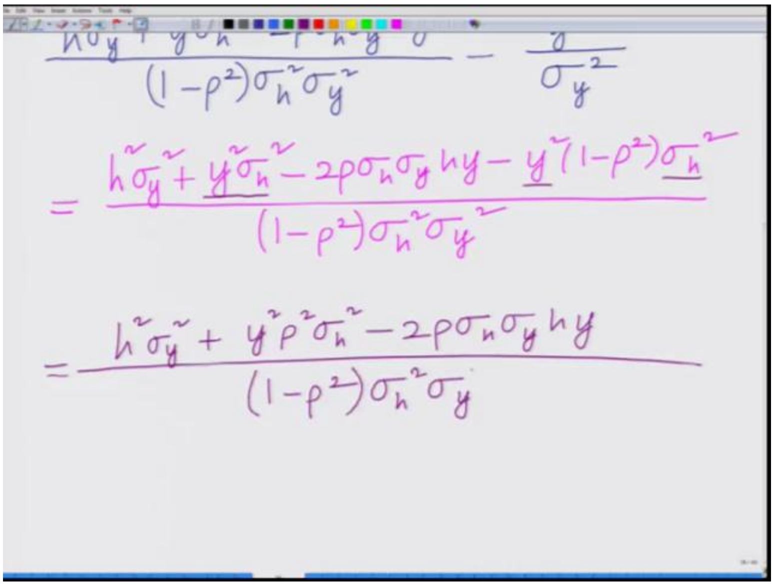
We said the marginal probability density function of the observation y, y is Gaussian land of variable with means 0 variable sigma h square; therefore, this we all ready derive this marginal probability density function of the observation y, this is given as 2 pi sigma y square e to the power of minus half y square divided by sigma y square. And therefore, now I can simplify this first letter, simplify this quantity that is the multiply factor, this is under root 2 pi 1 minus whole square sigma h square and sigma y square divided by sigma h square.

(Refer Slide Time: 03:56)



So, this gives 1 minus simply 1 minus rho square times, it is gives 1 minus rho square times sigma, 1 minus rho square times sigma h square into multiplied by. This is a large term minus half of first, which is basically your 1 minus rho square sigma h square sigma y square h square sigma y square plus y square sigma h square minus 2 rho sigma h sigma y h y minus this quantity, minus y square divided by sigma y square all right. So, this is the quantity in the exponent. So, this is multiplied by this is the multiplication symbol, and this is the quantity in the exponent. So, this is the quantity in the exponent; e to the power of minus half times this quantity.

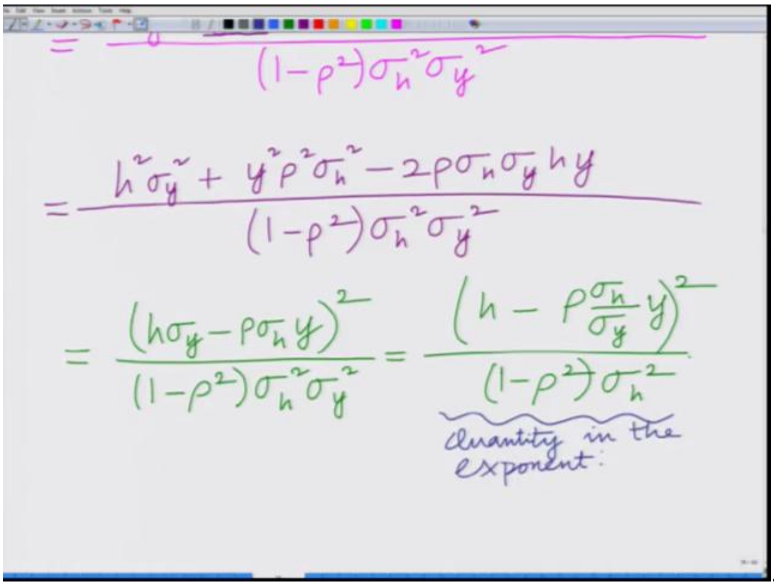
(Refer Slide Time: 05:20)



Now, let us simplify this quantity in the exponents. So, I want to first start by simplifying this quantity in the exponent, and then we can simplify the rest of the expression later. And this quantity in the exponent that can be simplified as follows. Let me just write it down, it is a bit cumbersome, but this is fundamentally important in MMSE Estimation. So, it is worth going through this derivation, at least once; that is h square sigma y square plus y square sigma h square minus 2 rho sigma h sigma y h y divided by 1 minus rho square sigma h square sigma y square minus y square divided by sigma y square.

And this is then I can combine these two 1 minus rho square sigma h square sigma y square h square sigma y square plus y square sigma h square minus 2 rho sigma h sigma y minus y square into 1 minus rho square sigma h square; that is the term I will get over here. Now observe this quantity I have y square sigma h square, and if I expand this I will have a minus y square sigma h square. I will have a minus y square sigma h square coming from these two. So, that those two terms. So, this term will cancel with this part, and what I am going to be left with interestingly if you observe this is h square sigma y square. Something very simple this y square sigma h square cancels, what I am left with is y square rho square sigma h square. So, I left with y square rho square sigma h square minus 2 rho sigma h sigma y h y 1 minus rho square sigma h square sigma y square.

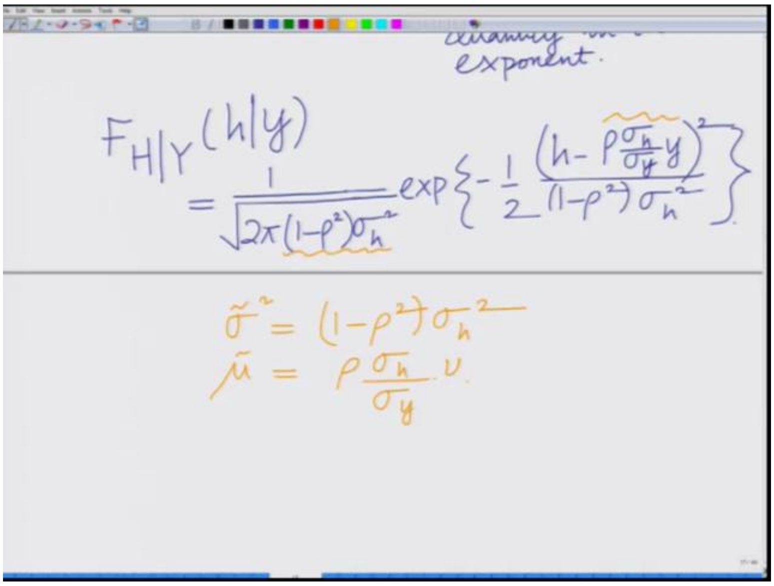
(Refer Slide Time: 08:06)



And now if you look at this quantity in the numerator that is nothing, but, observe that this quantity is nothing, but h sigma y minus rho sigma h y whole square, because look at this is h square sigma y square plus rho square sigma h square y square minus 2 rho sigma h sigma y h into y. So, this is nothing, but h sigma y minus rho sigma h y whole square.

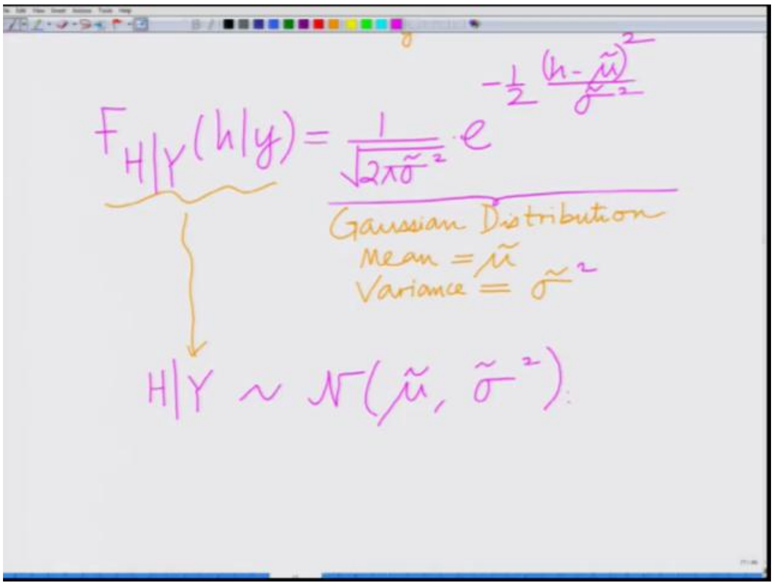
And once we have that the rest is very simple, this is 1 minus rho square sigma h square sigma y square. and now take sigma y out common from the denominator and numerator that is sigma y square common from the numerator and denominator, what I am going to have is h minus rho sigma h by sigma y into y whole square divided by 1 minus rho square, because sigma h square cancels from the numerator and denominator, 1 minus rho square sigma h square. This is basically your quantity in the exponent. Quantity in the exponent will be e to the power of minus half times this. This is your quantity the exponent, therefore, now the conditional, putting all together. Remember we set out to derive the conditional probability density function of parameter h giving the observation y.

(Refer Slide Time: 09:50)



Therefore now the conditional probability density function of the parameter h given the observation y, conditional probability of density function this will be given has follows. This will be given as, this is the multiplying factor is 1 over square root 2 pi 1 minus rho square time sigma h square e to the power of minus half. Now you can look at this e to the power of minus half h e to the power minus half h minus rho sigma h by sigma y y whole square divided by 1 minus rho square sigma h square. So, this is the joint probability density function. Now, what I am going to do is, basically I am going to set this quantity as sigma tilde; that is I am going to define two new quantity sigma tilde equal or sigma tilde square equals 1 minus rho square sigma h square and mu tilde equals this quantity over here. I am going to set mu tilde, mu tilde equals rho sigma h by sigma y sigma y times y.

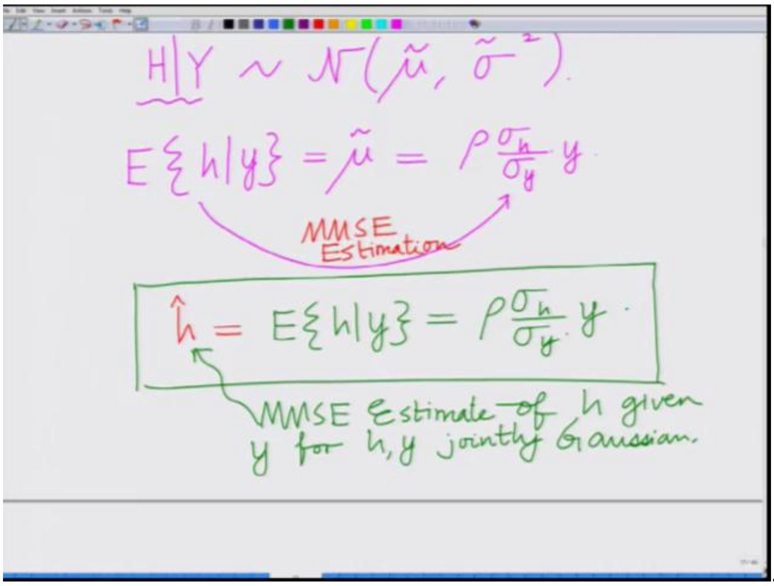
(Refer Slide Time: 11:36)



And now you can see something very interesting. now if I write this joint, if I write this conditional probability density function what I am going to observe is that; f of h given y of h given y equals 1 over 1 over, what you are going to observe is something very simple, 1 over square root 2 pi sigma tilde square e raise to minus half e raise to e is nothing, but expand, e raise to minus half h minus mu tilde whole square minus mu tilde whole square. And look at this, this is again your sigma tilde square divided by sigma tilde square. And now if you look at this, you will realize that this is a Gaussian distribution with mean mu tilde and variance sigma tilde square, and that is the interesting aspect of this.

So, this is basically, again now when you see this, that this is you can see that this is basically a Gaussian distribution with mean equals mu tilde variance equals sigma tilde. Therefore, what you can say is that the posterior. remember this also known as the posterior, the conditional probability density function of the parameter h given the observation y this is also the a posterior probability density function. This is Gaussian, so h given y h given y is basically Gaussian distributed with mean mu tilde variance sigma tilde square.

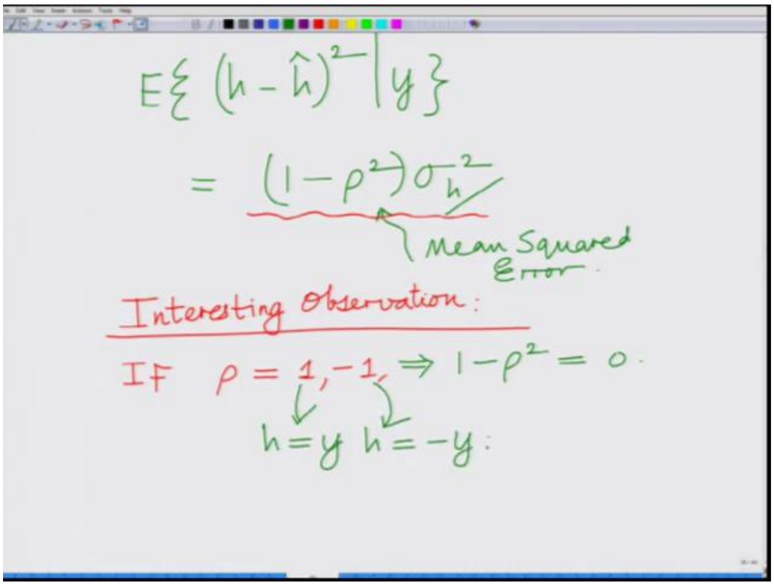
(Refer Slide Time: 13:31)



And therefore, now you can clearly see expected value of h y, and this is very straight forward, now you can see the expected value of h given y equals mu tilde which is nothing, but now you can see what is mu tilde. mu tilde is basically this quantity over here; that is your rho sigma h by sigma y times y, so this is your rho sigma h over sigma y times y. And indeed now we are found the MMSE estimate, because what is the MMSE estimate. MMSE estimate is the minimum mean squared error estimate, remember the optimal minimum mean squared error estimate; that is the estimate which minimizes the mean squared error, is nothing, but the expected value of the parameter h given the observation y. Therefore, this quantity mu tilde is nothing, but the MMSE estimate of h given y for that scenario when h and y are jointly Gaussian; that is something in mind.

So, this is basically your MMSE estimate. So, let us also write that down clearly. It requires some manipulation, but the end result is very simple and clear. So, this says that the MMSE estimate h hat equals expected value of h given y equals rho sigma h divided by sigma y times y, this is the MMSE estimate of. So, this is the MMSE estimate h given y for h comma y; that is the parameter and the observation, when the parameter and observation are basically jointly Gaussian random variation all right. And now what is this point is sigma tilde square, this quantity, this is your mean squared error; this is the mean squared error.

(Refer Slide Time: 16:08)

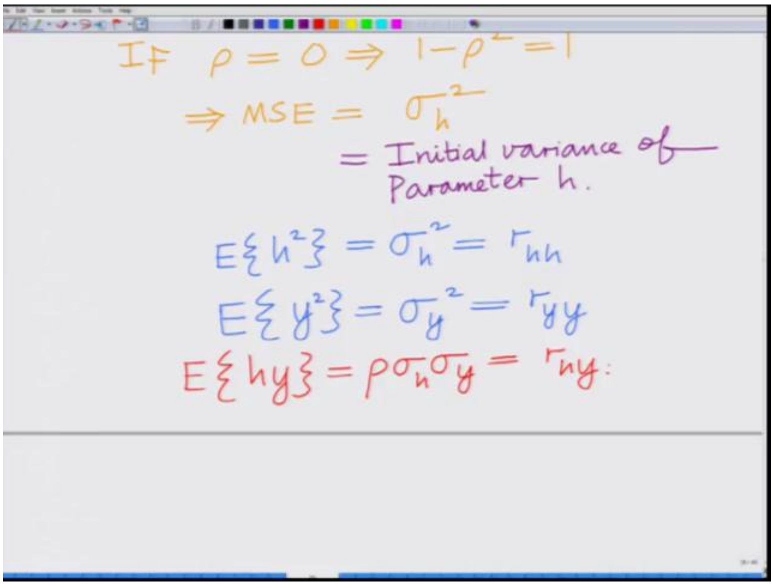


So, this is the mean squared error, what is this. This is the expected value of the parameter h minus h hat whole square given of course, the observation y. And what is this quantity. This quantity you can see 1 minus rho square times sigma h square. So, this is the mean squared error. And something very interesting can be observed from this mean squared error. What is the interest thing that we can observe from this mean squared error? Consider two particular cases, if. So, let us write this as interesting observation. Consider the simple case, if rho equals 1. Remember the correlation coefficient rho we said can lie between minus 1 or 1. Now correlation coefficient of 1 basically means y and h; that is the observation y and the parameter h are statistically identical, because the correlation is perfect; that is a correlation is the maximum value of one.

On the other hand if rho is minus 1 there also statistically identical in different way; that is h is basically minus y and y is minus h. So, basically in both these cases we see; that is correlation coefficient rho is equal to 1, or the correlation coefficient rho is minus 1 rho square is equal to 1. Therefore, 1 minus rho square is 0 which means MSE is 0, and that is naturally true because if the correlation coefficient is 1, on observing y you know that h is identically equal to y. And when correlation coefficient is minus 1 on observing y you know that h is equal to minus y, so there is no error, in that sense the mean squared error is 0. When there is perfect correlation between the observation y and the parameter h, the mean squared error is 0, and that is the interesting implication from this derivation.

So, rho is equal to 1 or for that matter minus 1, we have this implies 1 minus rho square equal to 0. So, this case, what we can see is in this case h is basically equal to y or h is basically equal to minus y, h can be perfectly estimated.

(Refer Slide Time: 18:54)

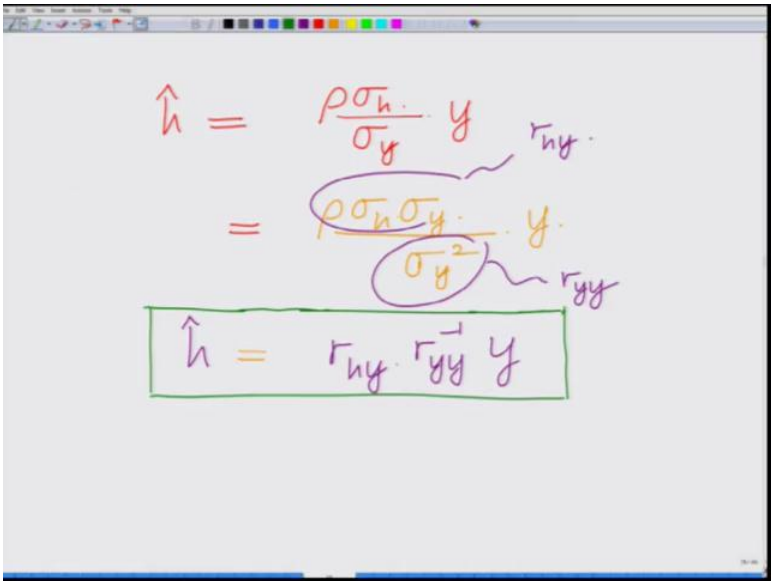


So, look we are saying is h can be, h can be perfectly estimated on y, since the correlation is minus 1 or 1. The correlation is 1 h is equal to y, if the correlation is minus 1 h is equal to y h is equal to minus 1. Now consider the other scenario, if rho is equal to zero; that is implies the correlation between h and y is 0 implies 1 minus rho square equals 1 implies MMSE equals what is the MSE, you can see 1 minus rho square is equal to 1 MSE rho is equal zero, MSE simply sigma h square the mean squared error simply sigma; that is initial variance which is equal to, which is equal to initial variance of the parameter h. it means having the, because the correlation between the observation y and h is zero, the observation y does not convey any information about a.

So, observation y cannot be used in any meaningful sense, to get an estimate on the unknown parameter h. Therefore, the variance of h; that is the there is the parameter h is sigma h square to begin with, even after making the observation y the variance of h means that sigma h square. Therefore, having observed y does not make any impact, or does not convey any information about the unknown parameter h. These are two extreme cases which interesting insights into the behavior of this MSE estimate. And now we can see for different values of rho between 0 and 1, and the variance 0 can go all the way from sigma h square; that is when rho is equal to 0 the variance is sigma h square. When rho is equal to 1 the variance is zero, and between when rho increases, as rho increases from 0 to 1 the variance decreases, or the MSE decreases from sigma h square to zero. So, that is the interesting observation that one can make from the MSE; the mean squared error per say, of this MMSE estimate that we derived from the jointly Gaussian scenario.

So, now let us also simplify this MMSE estimate and the variance to give alternative expression, which are both convenient. So, we have already said, what have we already said? We have already said if you go back to our nomenclature, and if you look at our definitions. From our definitions what we have, is basically, we have that expected h square equals sigma h square. Let us denote this by r h h, expected y square equals sigma y square, let us denote this by r y y. just denoting this by the variance by r h h, and the variance of y by r y y, and the cross correlation between h and y. This we can denote by expected h comma by why we denote by r h comma. So, let me just repeat those definitions over here. So, that it becomes convenient. let us define your sigma h square as expected value of the parameter sigma h square equals expected h square sigma h square equals r h h. Let us define expected value of y square equals sigma y square as r y y. And let us define the expected value of h y which is equal to rho sigma h sigma y as r h r h y.

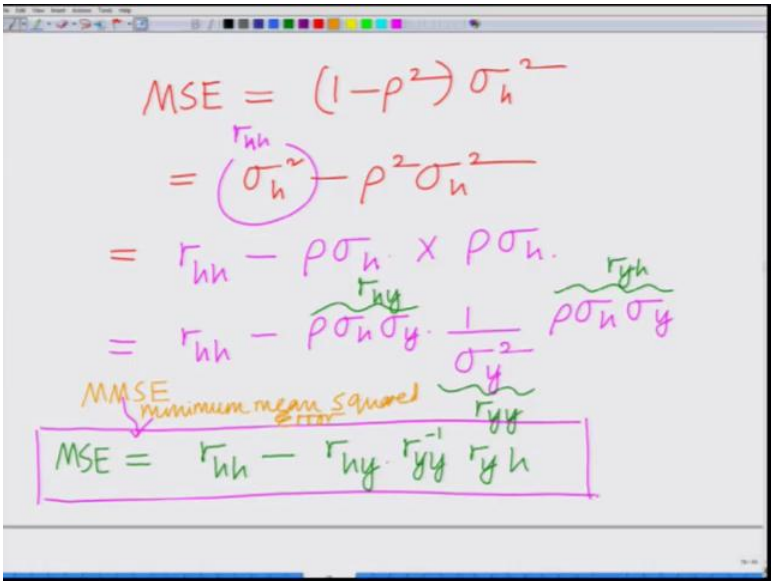
(Refer Slide Time: 23:15)



Now if you look at the MMSE estimate, the MMSE estimate is h hat which is the expected value of h given y, which is also you can see for this rho sigma h divided by sigma y times y. So, this is basically your rho sigma h divided by sigma y times y. And now I can simplify this has follows; that is rho sigma h sigma y multiplying and dividing by sigma y I have rho sigma h sigma y divided by sigma y square times y, which is now if you look at this, this quantity here, this is basically your r h y, and this quantity in the denominator basically your r y y. So, I can write this as something that is going to be helpful in our later discussion r h y times 1 over r y y, which is basically the r y y inverse times y.

And this is a very interesting expression, because the same expression can be carried over also for the matrix scenario; that is when h and y r vectors. So, I can write the estimate h hat as r h y times r y y inverse times y. Now of course, when h and y are vectors the interest thing is I can replace r y y by the covariance matrix of y r h y by the cross covariance matrix between the parameter vector h and the observation. So, this is the very interesting way to write this, because is a very convenient expression to remember, and that can also be used in various other much more complicated scenarios, such as vector scenario, when y is a vector when h is also a parameter vector. Although the expression that we have derived which is rho sigma h divided by sigma y times y is equally valid, this is much more convenient and much more standard form to remember this expression.

(Refer Slide Time: 25:19)



And the expression for the MSE can be simplified as, so this is the expression for the MMSE estimate. And the expression for the MSE this can be simplified as follows. So, this is the expression for MMSE estimate and the expression for the MSE this can be simplified as follows this equal to 1 minus rho square sigma h square, which is basically your sigma h square minus rho square sigma h square. Now of course, this quantity here we already know what this quantity is. This quantity is r y y, so this is, I am sorry this quantity here is basically your r h h. So, this is basically r h h minus. Now this quantity I can write as rho sigma h times rho sigma h. Now, this quantity I will again manipulate as by multiplying and dividing by sigma y square rho sigma h sigma y into 1 over sigma y square into rho sigma h sigma y.

Now you can this quantity is r h y, this quantity is here r y y, and this quantity here I can write it as r y h, because for the scalar scenario r h equals r y h which might not to be same as the vector scenario. In fact, r h y is r y h transpose for real vector. So, this is r h y into r y y inverse into r y h, and this is your convenient expression for the MMSE m s. In fact, this is not just the MMSE, but this is the MMSE this. in this case this is the minimum mean squared error, it is not just the MMSE, but for this particular scenario this is the MMSE; that is your actual minimum mean squared error. So, we have found two things; one, we have found the MMSE estimate; that is rho sigma h sigma y divided by sigma y square into y which is basically r h y times r y y inverse into y. And also we have found the actual value of this minimum mean squared error; that is the MMSE which is r h h that is sigma h square minus r h y into r y y inverse into r y h, this is your means squared error for this scenario.

So, that basically completes this derivation where we have started with jointly Gaussian random variables h and y, where h is the random parameter, y is the random observation, we considered into the jointly Gaussian in nature. And following this we basically derived the conditional probability density function of h given y, using the joint probability density function of h and y, and the marginal probability density function of y, and basically from that we inferred the conditional probability, the conditional mean of the unknown parameter h given the observation y which is in fact, the MMSE estimate. And we have also found the mean squared error or rather in this case also the minimum mean squared error.

So, we will stop this module here, and we will continue the other aspects in subsequent modules.

Thank you very much.