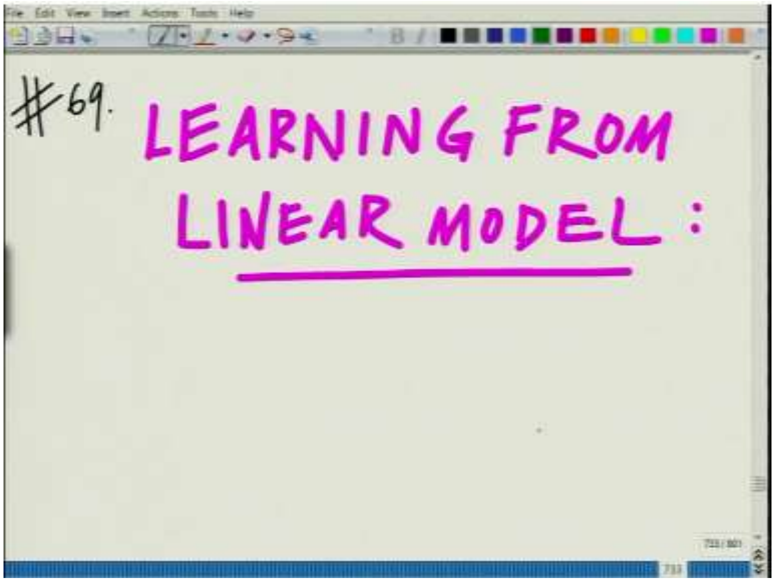
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur
Lecture No. 69
Scalar Linear Model for Gaussian Estimation

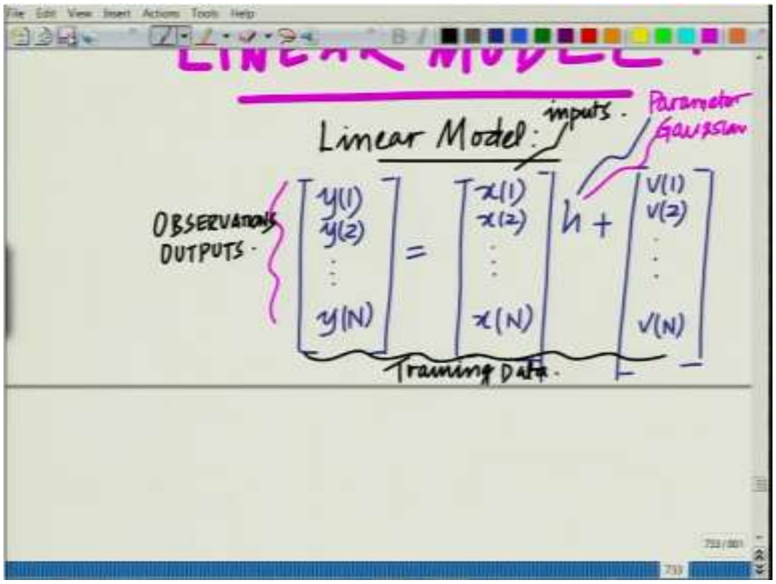
Hello everyone, welcome to another module in this massive open online course. So, today let us extend the concept that we have learned regarding the conditional Gaussian and inference from the conditional Gaussian model, in particular let us talk about an application related to learning or extracting the information, learning a parameter from a linear model.

(Refer Slide Time: 0:35)



So, we want to talk about using the results of the conditional Gaussian, let us talk about learning from linear. So, an application regarding, so how do you infer more knowledge about a parameter from a linear model, how do you extract information from a parameter.

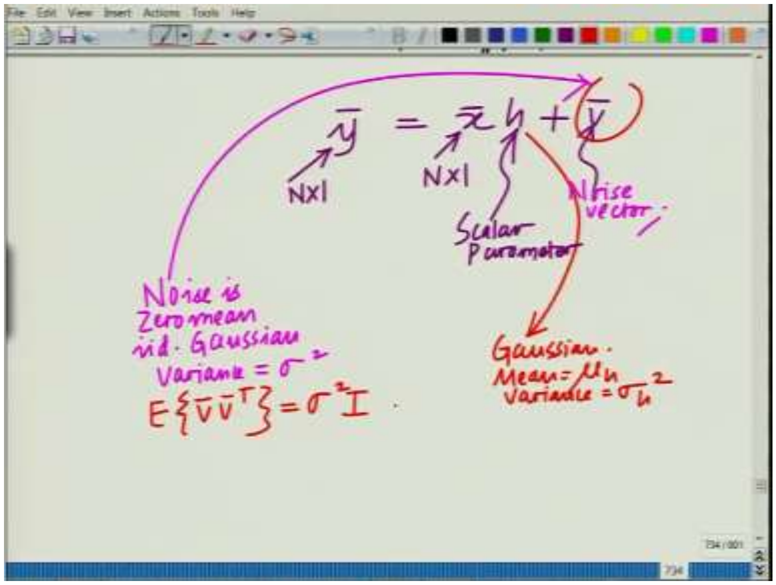
(Refer Slide Time: 1:11)



So, let us say we have the following model, this is our linear model, in this model we have the observations y1, y2 so on up to yn, this is our vector of observations, or this is our training data you have x1, x2 so you can think of this as our training set. So, these are the training data and this is the parameter h that we would like to learn, so you have the parameter h plus we have v1, v2 up to vn.

So, this is the parameter that we would like to learn and let us assume that this is a Gaussian parameter, so you would like to learn more information about this Gaussian parameter, these are the observations, so these are your observations or you can think of this as the outputs and these are the inputs to the model and together you can think of this as the training data basically, this is your training data in your learning problem, in your machine learning problem, you can think of this as your training data.

(Refer Slide Time: 3:08)

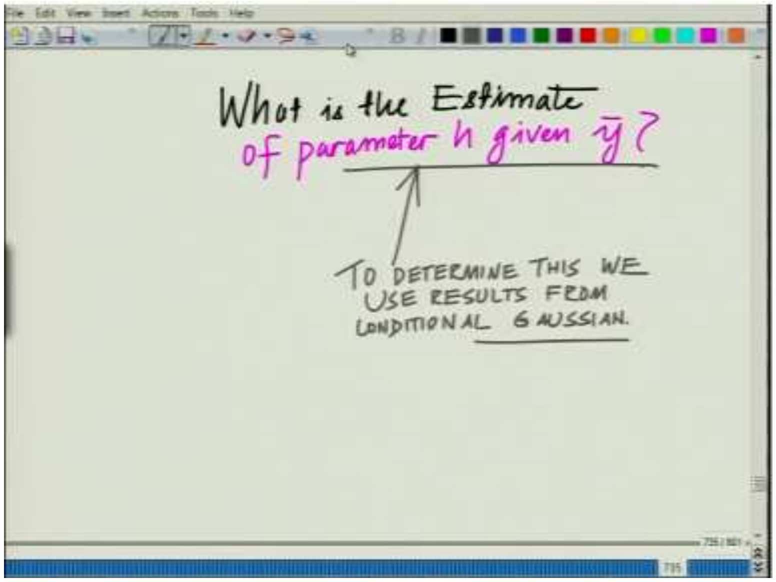


And so this you can write this in this compact fashion, you can write it in the compact fashion y bar equals x bar h plus v bar, where this is of course, we have seen this is an n dimensional vector is N cross 1, this is also an N dimensional vector this N cross 1, this is a scalar parameter, 1 cross 1 you can think of it, this is a scalar parameter, this is the noise vector and what we have said is the following thing, so this is Gaussian in nature.

Remember, we are considering Gaussian inference this is Gaussian in nature, let us say the mean equals mu h and the variance of this is sigma square and the noise as usual simple model for the noise, the noise is i.i.d., that is the samples are independent identically distributed there is a correlation is 0, each noise is 0, each noise sample vi is 0 mean and the variance is sigma square. So, the noise is 0 mean i.i.d. Gaussian therefore, variance let us say the noise samples variance equals sigma square therefore if you can look at the covariance matrix that becomes expected value of v bar v bar transpose.

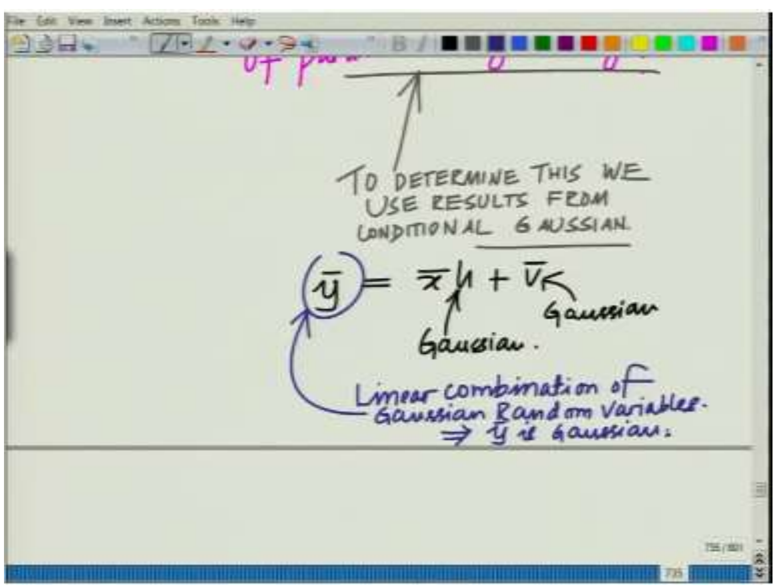
You can write this as this is equal to sigma square times identity, this is what we know this is we know, this is what we call as the covariance matrix of the noise, that is when the noise samples are independent identically distributed the covariance matrix is essentially proportional to the identity matrix because the off diagonal terms are 0 under the cross correlate, the correlation between the different noise samples 0, all the diagonal terms the variance is essentially sigma square. So, this is basically sigma square times the identity. So, this is the noise covariance matrix, this is the noise covariance matrix and we already determined we already talked about the parameter.

(Refer Slide Time: 5:59)



Now we want to ask the question what is the estimate of the parameter or how can you learn what is, we ask the question what is the estimate of your parameter h given the observation vector y bar, what is the estimate of the parameter h given y bar? And for that to determine this we use to determine this, to determine this we use results from the conditional Gaussian properties, to determine this we use results from the conditional Gaussian.

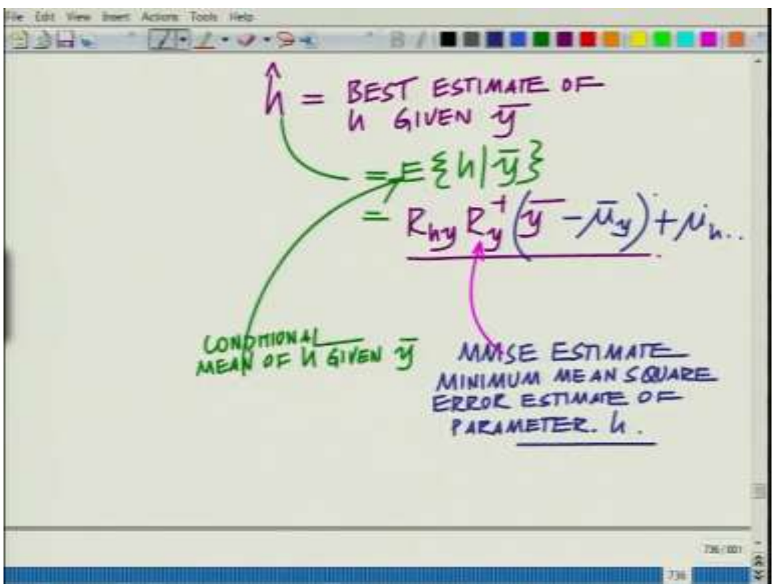
(Refer Slide Time: 7:22)

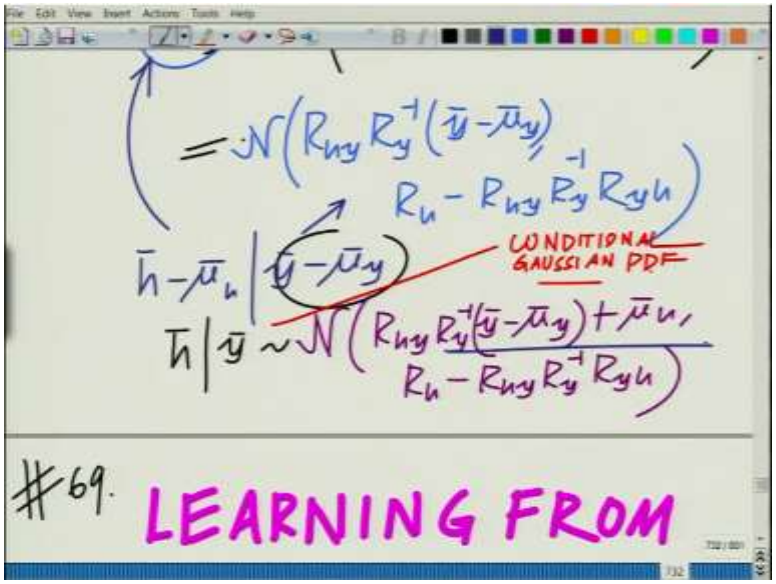


So, let us look at this, so we have go back to our model y bar equal to x bar h bar plus v bar where this is Gaussian, this is Gaussian and therefore now what we have is this is a linear or linear combination of Gaussian random variables. So, you can see y bar, y bar that is the observation y vector, y bar is essentially a linear combination of two Gaussian random variables that is your h, parameter h and the noise vector y bar. So, naturally y bar that is observation with vector itself is also Gaussian nature, because linear combination of Gaussian random variables is in turn Gaussian. So, this is a linear combination of Gaussian random variables implies y bar is Gaussian.

Therefore, the best estimate of h bar, so y bar is Gaussian and we have seen the best estimate of h given y bar is the conditional mean of h, which follows from the a posteriori probability density function of h given the observation vector y bar, that is from the theory that we have seen with respect to the conditional Gaussian probability density function.

(Refer Slide Time: 9:05)



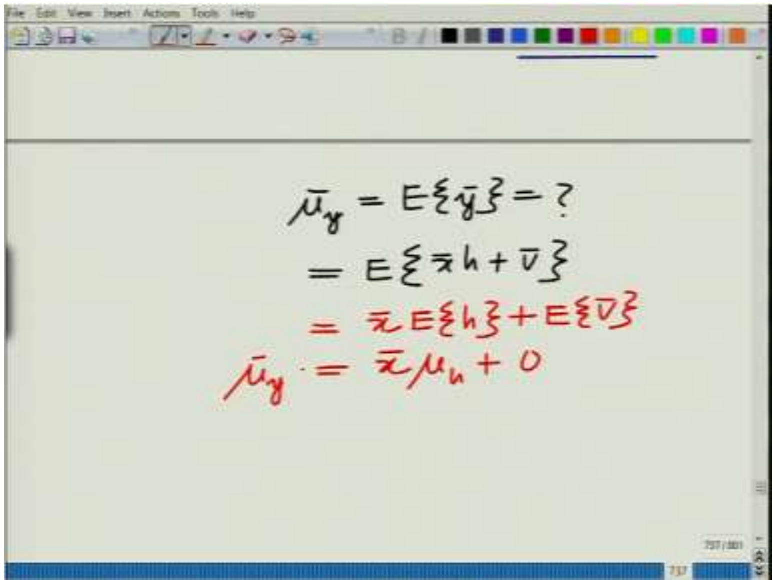


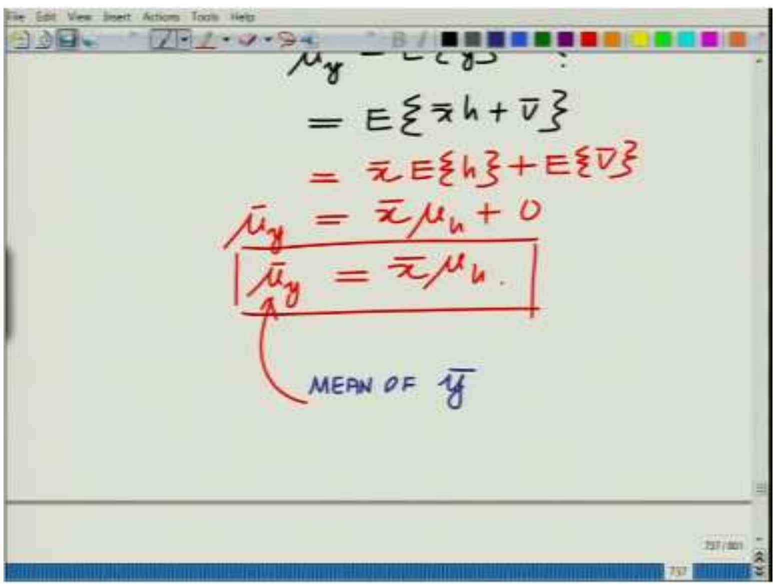
So, if we ask the question what is h hat that is the best and we have seen best in the MMSE, best estimate of h given y bar this is essentially equal to expected value of h given y bar which we have seen this is essentially nothing but the conditional expectation, conditional mean, this is the conditional mean of h, given conditional mean of h given y bar and this as we have seen is given by Rhy, the cross covariance between h y times Ry inverse into y bar that is essentially equal to your h hat and this is essentially that is what we also termed as the MMSE estimate.

This is in terms of learning or estimation this is essentially the MMSE estimate, that is the minimum mean squared error estimate of the parameter h, this is the MMSE estimate, that is the minimum mean square error estimate of the parameter h. Now, therefore the estimate of h given y bar of course, you have to also take into account the mean, so this is y minus mu y bar plus mu h because this is non zero mean quantity, so you have to also remember we are using the relation corresponding to the non zero mean quantities, that is if you go all the way back and you look at this, this is essentially the result that we are using.

And now to start with we have to therefore determine to begin this process we have to determine what is the mean of y bar, we know that v bar is 0 mean, parameter h has mean mu h, we ask the question what is the mean of y bar, that is what is mu h bar, what is the expected value.

(Refer Slide Time: 12:25)

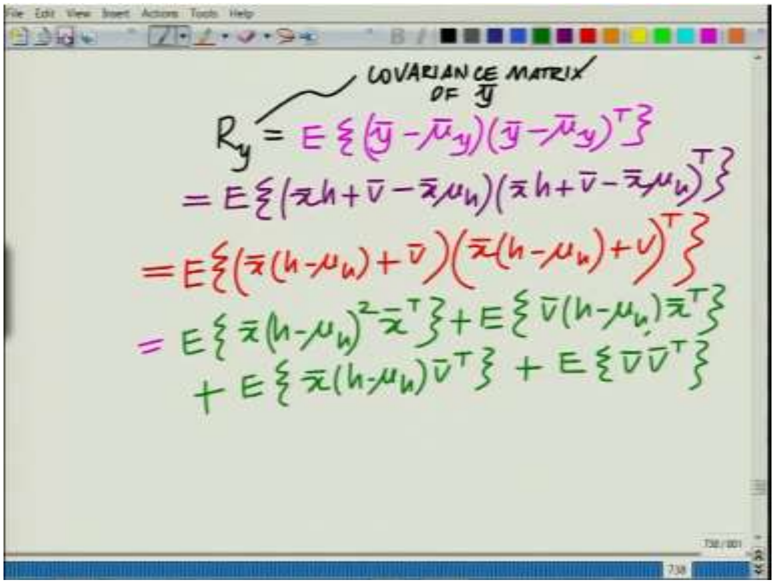




So, if we denote that by mu bar y, which is equal to expected value of y bar, what is this quantity? This quantity is simple, this quantity is expected value of x bar h plus v bar, which is if you think about this simply x bar times the expected value of h plus expected value of v bar which is equal to x bar equal x bar times, x bar times mu h plus 0. So, mu bar y, this is expected value of x bar times mu h, so this is x bar times mu h, this is your mu bar y, this is the mean of, essentially the mean of y bar, mean of the observation vector y bar.

Now we ask the question, remember the other thing that we need is Ry, which is the covariance matrix of the observation vector y bar, that is expected value of y bar minus mu y bar times y bar minus mu y bar transpose.

(Refer Slide Time: 13:57)

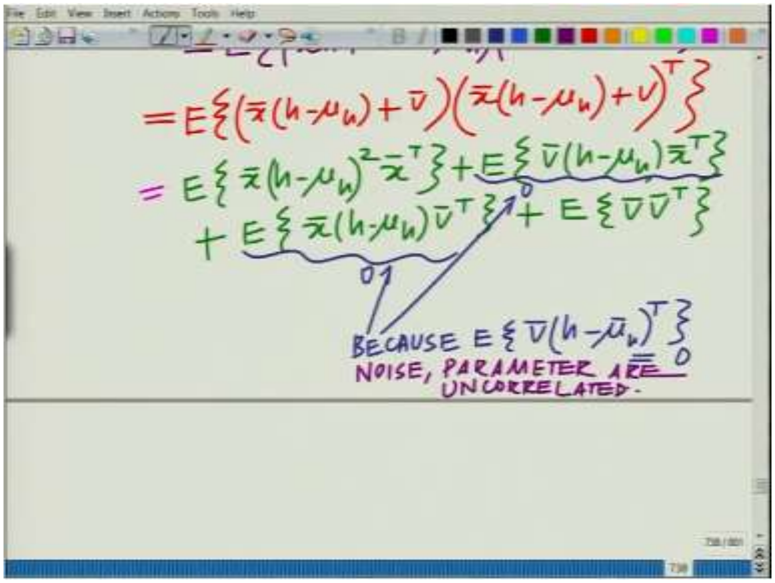


So, we need Ry for the conditional mean or inferring the parameter h, the other quantity that we need is Ry which we are calling as the covariance matrix of y bar which is equal to expected value of y bar minus mu bar y into y bar minus mu bar y transpose which is, now if you simplify this, this is expected value of well y bar is x bar h plus v bar minus mu y or minus mu bar y which is x bar mu h times, the same thing x bar.

So, let me just write this again, this is equal to the expected value of x bar h plus v bar minus mu bar y which is x bar mu h times x bar h plus v bar minus x bar mu h transpose, which you can now write as the expected value of x bar h minus mu h plus v bar into x bar h minus mu h plus v bar transpose, which if you simplify this, this is expected value of x bar into h minus mu h of course, h minus mu which is a scalar quantity, so I can write it as h minus mu h square times x bar transpose plus.

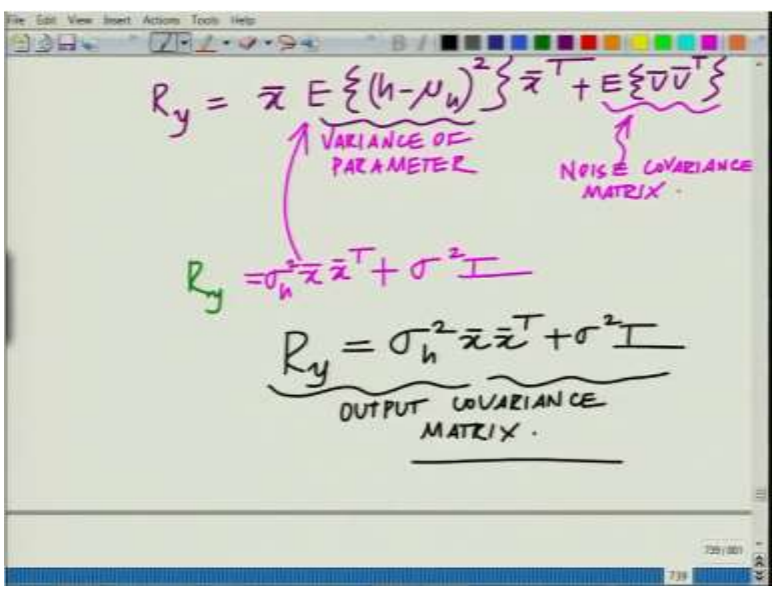
Now, we are going to use a property, now if you look at this, let me just write this terms that is expected value of v bar h minus mu h x bar transpose plus expected value of x bar h minus mu h times v bar transpose plus expected value of v bar v bar transpose, plus expected value v bar v bar transpose. Now if you look at these two quantities, the quantities in the middle, these are related to the cross covariance between the parameter and the noise and typically the parameter is uncorrelated with the noise, that is the parameter that you are trying to infer is typically uncorrelated with the noise.

(Refer Slide Time: 17:20)



So, essentially what is happening over here is this pair, these two terms, these are 0 because these are typically 0 because expected value of v bar into h minus mu h transpose this equal to 0, why is that? Because noise and the parameter to be learnt are uncorrelated or often in fact independent, these are uncorrelated.

(Refer Slide Time: 18:10)

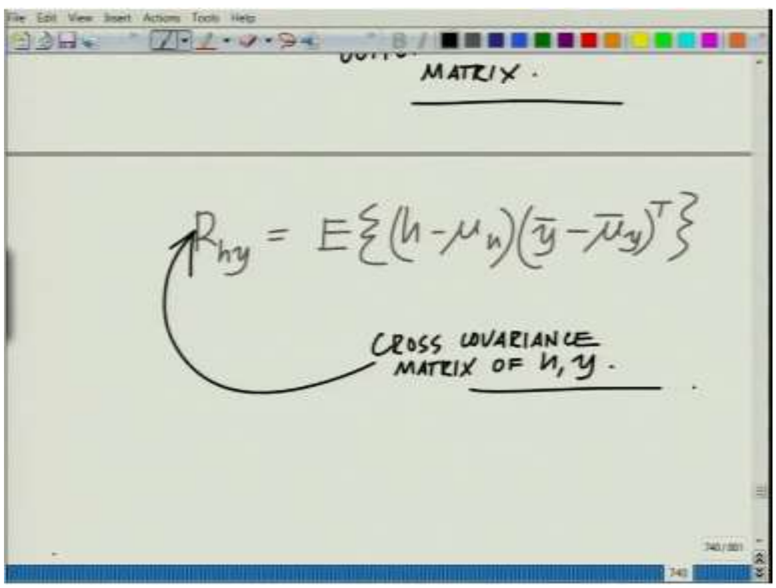


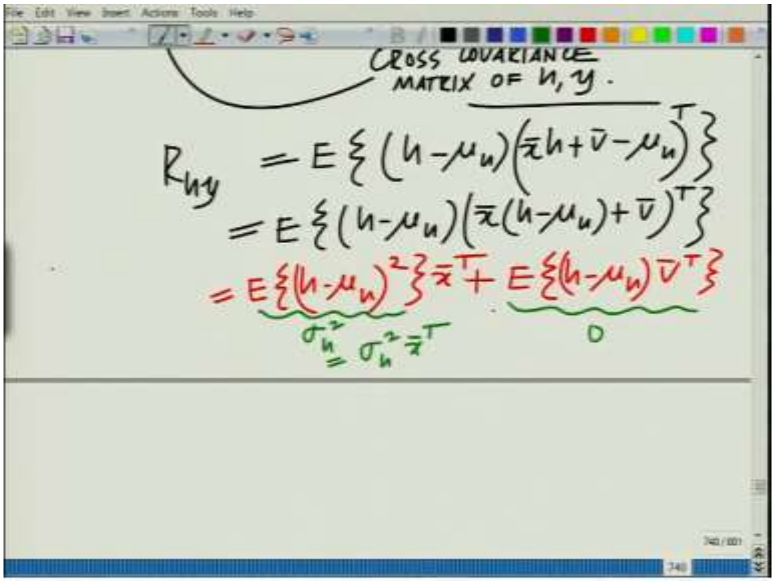
So, these two terms are 0 and therefore we are left with only the outer terms which is Ry equals x bar expected value of h minus mu h square x bar transpose plus expected value of v bar v bar transpose, this we know expected value of h minus mu h square this we know is nothing but the variance of the parameter. So, this quantity is sigma h square and expected value of v bar v bar transpose that is essentially the covariance matrix of the noise.

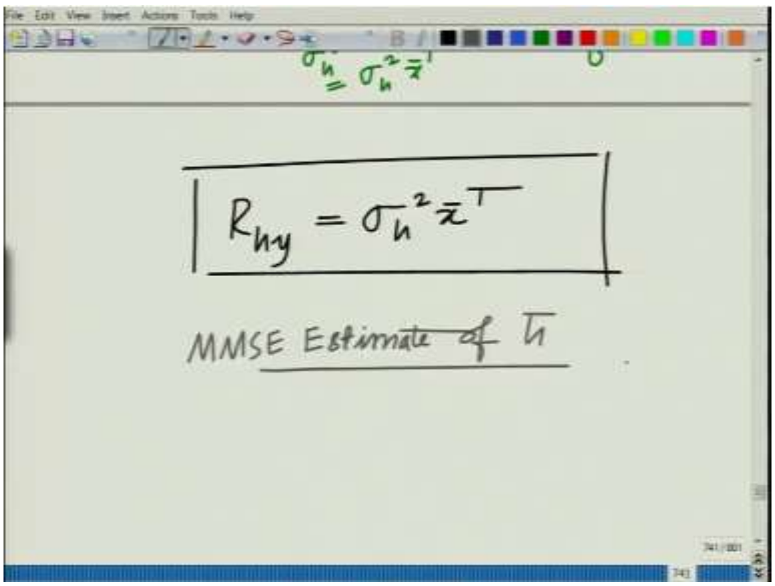
So, this is essentially the variance of the parameter and what about this, this is the noise covariance matrix and therefore, you can write this as this is equal to x bar, this is sigma h square x bar x bar transpose plus sigma square times identity. So, this is essentially your Ry, this is the covariance matrix of the output.

So, let me write this again over here clearly. So, Ry equals sigma h square x bar x bar transpose plus sigma square times identity, so this is your output covariance matrix, this is one of the quantities, important quantities that we are going to use to learn the parameter h. So, that is Ry in the formula if you remember we have the term Ry inverse.

(Refer Slide Time: 20:28)





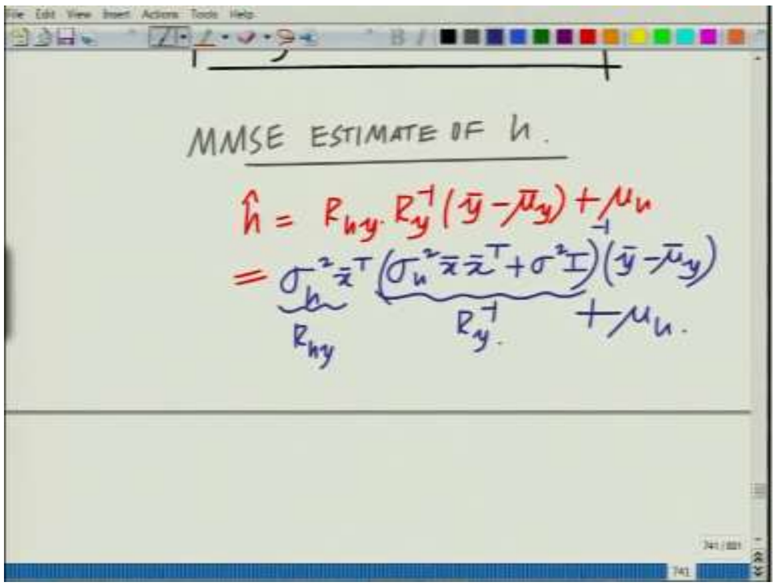


Now comes the other quantity which is the cross covariance, now we have to evaluate the other quantity which is Rhy which again if you remember the definition, this is h minus mu h, this is scalar quantity times y bar minus mu bar y transpose, this is the cross covariance matrix of h comma y. And this is another quantity that is needed to determine the estimate of the unknown parameter h and how what is the formula to compute this we already stated the formula and therefore this is equal to Rhy.

This is equal to expected value of h minus mu h times y bar is x bar h plus v bar minus mu h which again if you take a look at it this is essentially equal to h minus mu h into x bar h minus mu h plus v bar transpose, transpose which is equal to expected value of h minus mu h square times x bar transpose plus expected value of h minus mu h into v bar transpose and as you know this quantity is equal to 0, and this quantity, well this is of course, the variance this is sigma h square, so this quantity evaluates as sigma h square times x bar transfer.

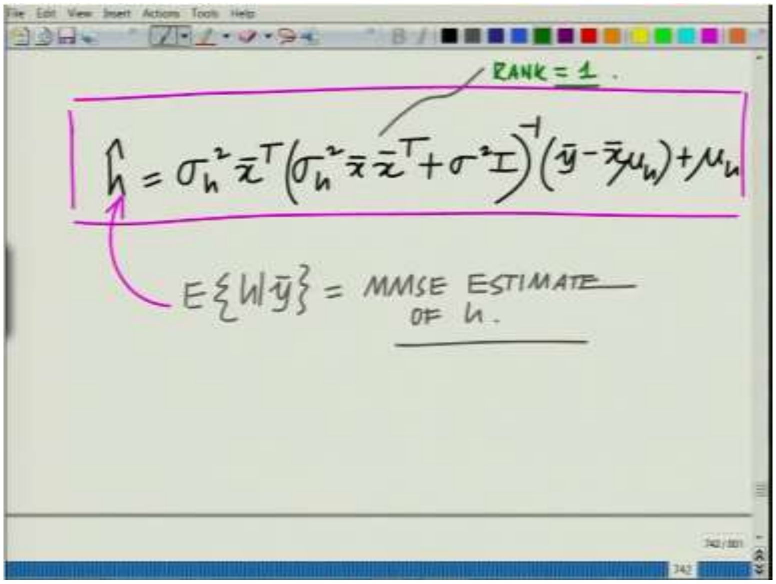
So, the cross covariance matrix Rhy in summary Rhy evaluates as sigma x square times x bar transpose, this is the cross covariance matrix of this thing and now what we can do is essentially we can determine the MMSE estimate, now we have the MMSE estimate.

(Refer Slide Time: 23:19)



So, the MMSE estimate of h that is given as h hat equals Rhy times Ry inverse y bar minus mu bar y plus mu bar or mu of h which is equal to, now substitute for these quantities Rhy, so this is you have your sigma h square x bar transpose times sigma h square x bar x bar transpose plus sigma square I inverse y bar minus mu bar y plus mu of h. So, essentially you are substituting this Ry that we have determined over here and we have the expression for the Rhy that we have determined over here so if you look at this, this is essentially Rhy and this is basically your Ry inverse, this is essentially your Ry inverse.

(Refer Slide Time: 24:54)



So, therefore just to write it out again once again clearly. So, you have the MMSE estimate that is your Rhy which is sigma h square x bar transpose times sigma h square x bar x bar transpose plus sigma square identity inverse into y bar minus x bar mu h plus mu h. So, this is the MMSE estimate. So, this is essentially if you look at this, this is the MMSE estimate of h given y bar, this is essentially expected value of h given y bar which is equal to the also calling as the MMSE estimate of h, this is also the estimate of h.

So, this can be simplified further, this expression can be simplified further, in fact using interesting property, remember the matrix inversion identity that we have seen earlier in one of the earlier modules this can be further simplified and in fact, you can compute it very efficiently because this x bar x bar transpose, if you look at this, this is a rank one matrix. So, this can be computed much more efficiently which we will look at in the subsequent modules.

So, efficiently evaluate this and also look at what is the variance of estimation, what is the accuracy of estimation of this unknown parameter h that we are trying to learn. So, with that let us stop here and let us continue this discussion in the subsequent (())(27:01).