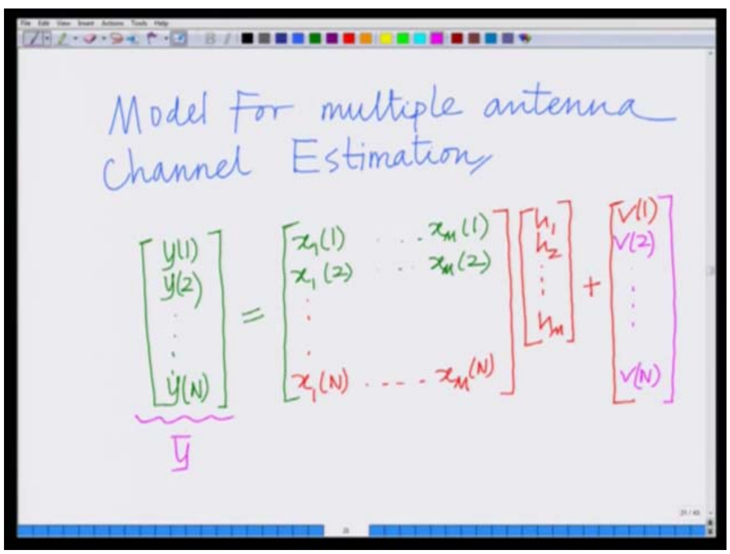
Estimation for Wireless Communications-MIMO/OFDM Cellular and Sensor Networks.
Professor Aditya K Jagannatham.
Department of Electrical Engineering.
Indian Institute of Technology Kanpur.
Lecture -14.
Likelihood Function And Least-Squares Cost Function For Vector Parameter Estimation.

Hello, welcome to another module in this massive open online course on estimation for wireless communication system. So, currently we are looking at the estimation of a vector parameter and we tried to motivate this in the context of a multiple antenna wireless communication system. So, we said the estimation model for a channel estimation in a multiple antenna wireless system. For instance, we were considering downlink ask a multiple antenna transmission system considering multiple antennas at the base station and a single antenna at the mobile. So, this, the estimation problem for the system can be written as and that is what we have showed in the previous module is that

(Refer Slide Time: 0:52)

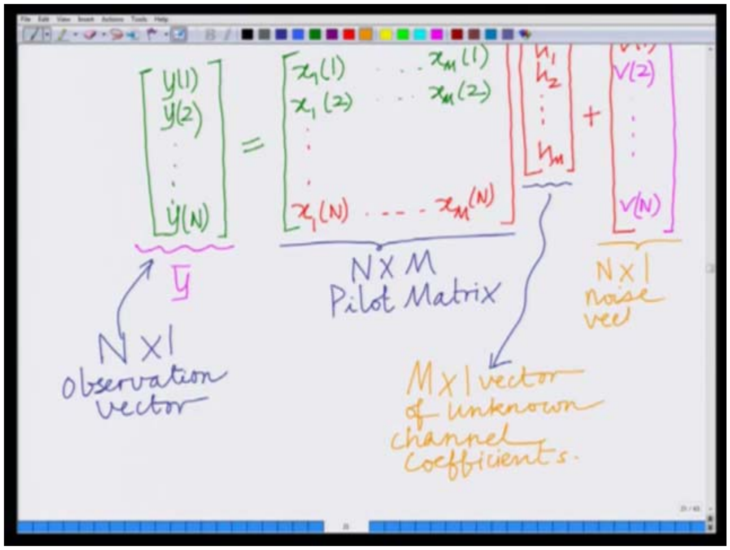


we have the model for multiple antenna channel estimation, which can be formulated as Y1, Y2… so on up to YN.

Remember these are the observations that is equal to your pilot matrix which has entries of the forms X11 up to XM or XM1, X12 up to XM2, these are the pilot symbols transmitted at the 2^{nd} time instants. And X1N to XMN, this is your pilot matrix times, let me write this in a slightly different fashion. This is H1, H2… up to HM which are the M channel coefficients +

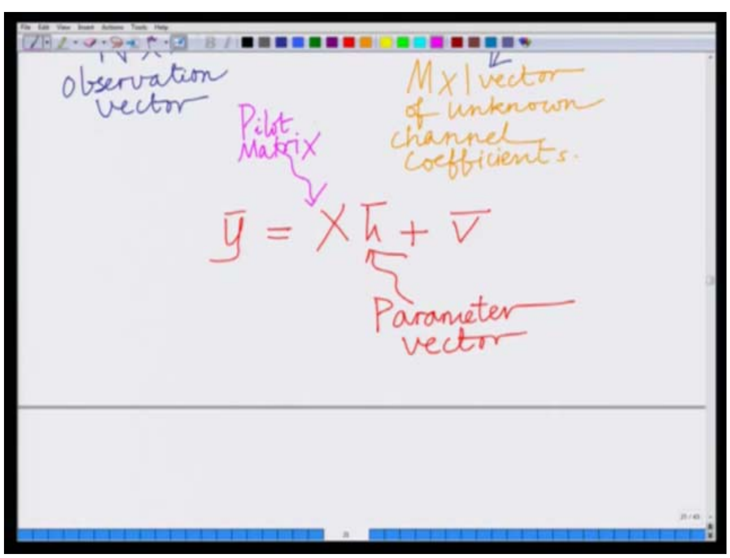
the noise samples which are basically or V1, V2… Up to VN, this is a system model for multiple antenna channel estimation where Y1, Y2… so on up to YN, these are your Y bar, this is the vector, this is the vector of N observations.

(Refer Slide Time: 3:00)



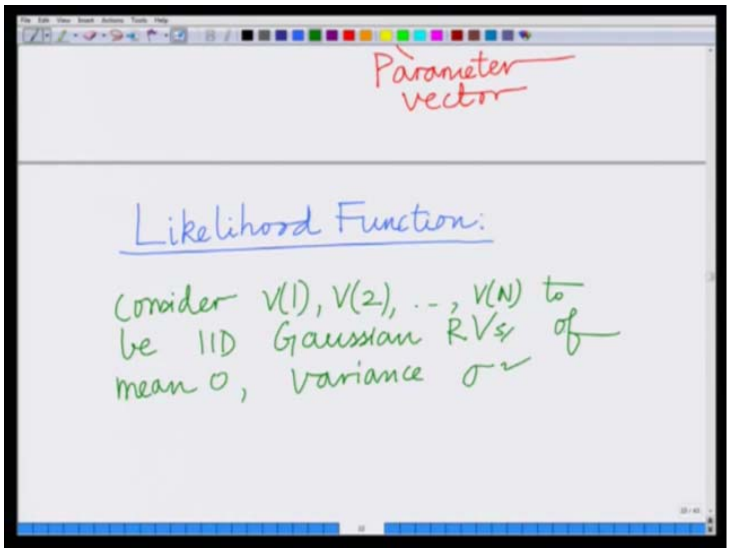
This is the N x 1, this is, let me write this clearly, this is your N x 1 observation vector, this is the N x M pilot matrix, that is the matrix of the pilot symbols and this basically, these are the unknown, this is M x 1vector of unknown channel coefficient and this is your N x 1 additive noise vector or simply noise vector and then we had said this system model therefore can be written as

(Refer Slide Time: 4:06)



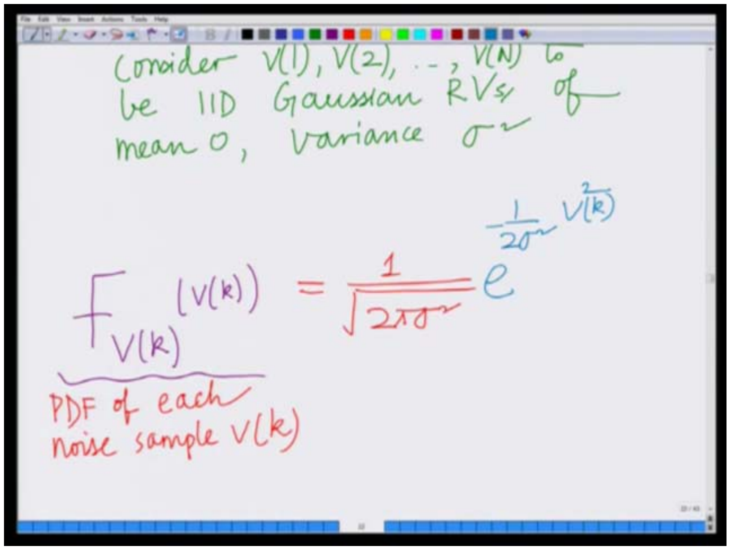
your Y bar or Y equals X times H bar which is the unknown channel vector or the parameter vector to be estimated class + V bar. So, this is the parameter vector, correct. The parameter vector which has to be estimated which has to be estimated, so we have Y bar equals H X times H bar + V bar, where H bar contains these M channel coefficients, the M unknown channel coefficients which have to be estimated, X is the matrix of the pilot symbols, this is also known as the pilot matrix.

(Refer Slide Time: 5:04)



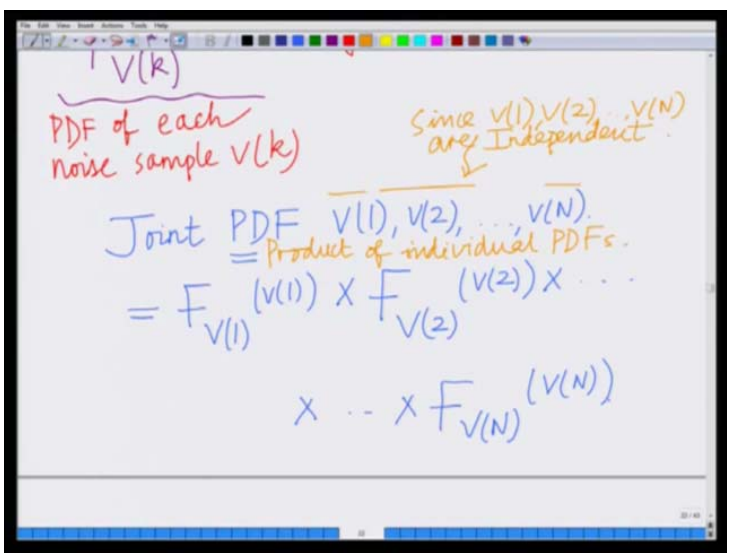
So, X is known as, this is known as the pilot matrix. So, similar to what we have done before, we have to develop for estimation of the vector parameter H bar, we have to develop the likelihood function, so any estimation, we have to start by developing the likelihood function for the **est**, likelihood function corresponding to the parameter vector H bar and again we will start with the same other mentioned that is we will start by considering the noise samples V1, V2 up to VN, the N noise samples to be IID Gaussian, to be independent identically distributed Gaussian random variables of mean 0 and variance Sigma square each. So, consider V1, V2… Up to VN to be IID Gaussian RVs of mean 0 variance Sigma square, which means the PDF of each, the probability density function of each noise sample VK, what is this, this is the PDF, let me write this again clearly a little bit.

(Refer Slide Time: 6:23)



This is the PDF of each noise sample VK equals 1 over square root of 2 Pie Sigma square E raised to -1 over 2 Sigma square, in fact this is the mean is 0, VK square. So, this is the probability density function of each noise sample VK which we said is 1 over square root of 2 pie Sigma square E raised to -1 over 2 Sigma square times P square K. Alright, because this is a Gaussian noise sample of mean 0 and variance Sigma square. Further, realise now that all the noise samples are independent, therefore the joint probability density functions,

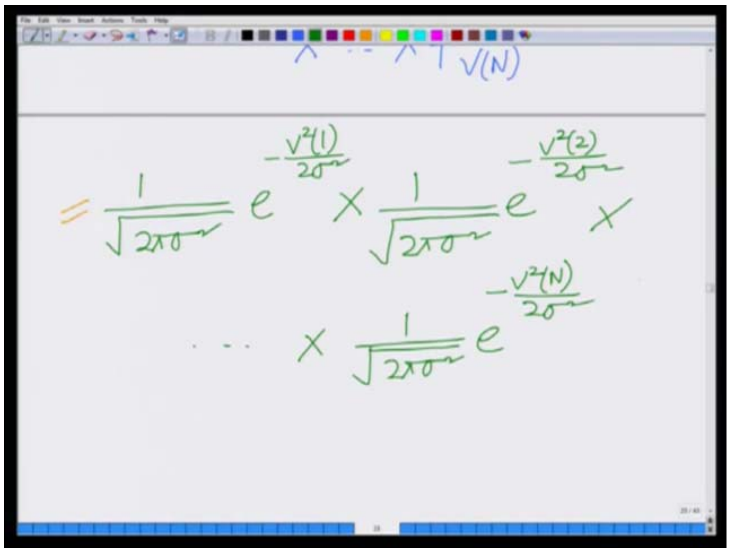
(Refer Slide Time: 7:53)



function of these noise samples V1, V2… Up to VN is the product of the individual probability density functions and this is also something which we have seen before.

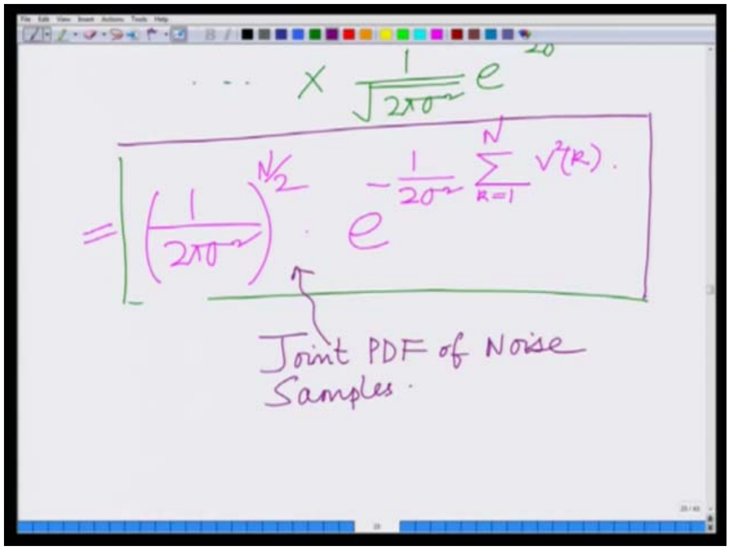
Therefore the joint probability density function, again just to follow joint PDF, again just to repeated so as to be very clear, although many of you must be familiar, the joint PDF of V1, V2… Up to VN equals the product of the individual probability density functions F of V1, F of V2 so on until F of VN, the joint PDF equals product of individual PDFs, let us write this down. And why is this, because since up to VN are independent. These are independent random variables, and therefore

(Refer Slide Time: 9:17)



I can write this as basically 1 over under root 2 pie Sigma square E raised to - V square 1 by 2 Sigma square times 1 over square root of 2 pie Sigma square E raised to - V square 2 by 2 Sigma square, so on and so forth until 1 over square root of 2 pie Sigma square E raised to - V square N by 2 Sigma square

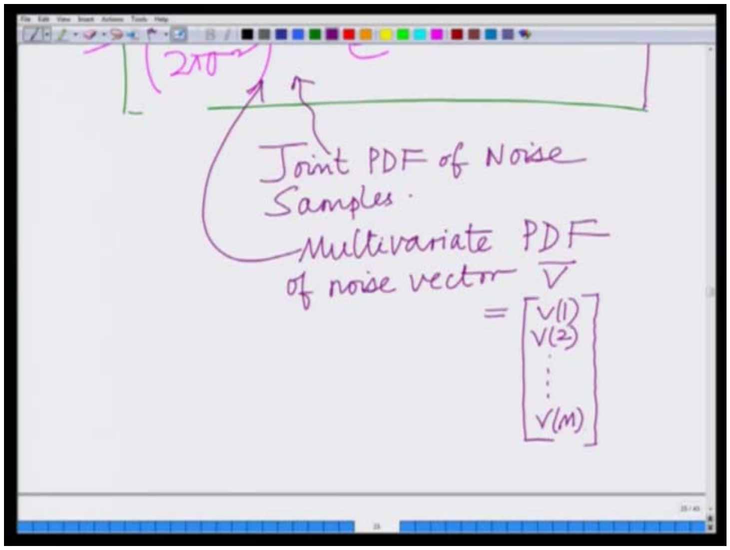
(Refer Slide Time: 10:01)



which is basically equal to 1 over 2 pie square Sigma square raised to the power of N by 2E raised to -1 over 2 Sigma square summation of K equals 1 to N V square K and what is this, this is basically your joint PDF.

This is basically the, what is this, this is basically the joint PDF of the noise samples. This is basically the joint PDF of the noise samples V1, V2… Up to VN but the noise samples are nothing but, they are in the vector V bar, that is V1, V2 V bar that is V bar that is the noise vector equals basically V1, V2… Up to VN, the n-dimensional vector with elements V1, V2… Up to VN, therefore this can also be viewed as the multivariate density of these of these noise samples V1, V2… Up to VN,

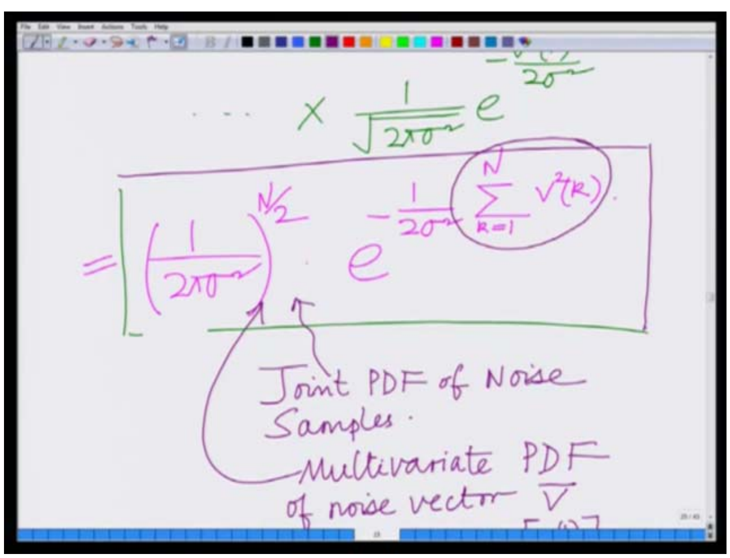
(Refer Slide Time: 11:30)



that is the multidimensional probability density function, the joint probability density function of V1, V2… Up to VN, typically this is also known as the multivariate density of the noise vector V bar.

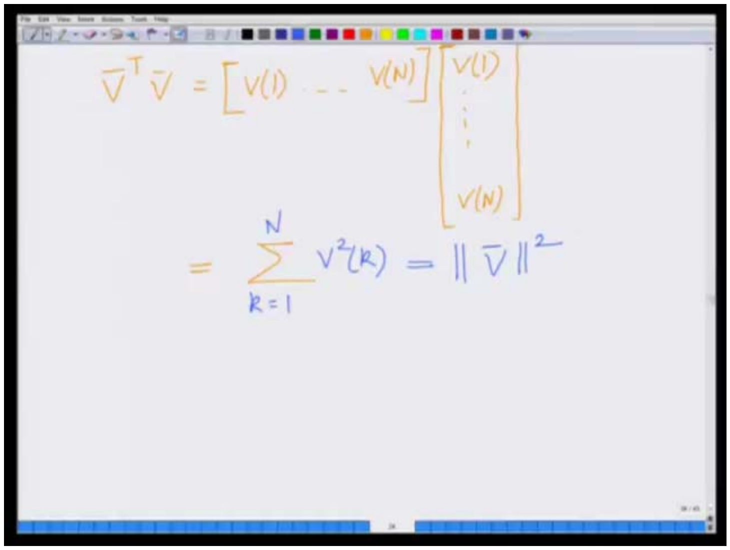
So, this is the probability density, PDF of the noise vector or the multivariate PDF of noise vector V bar. Which is nothing but, basically this is nothing but contains, again the noise samples V1, V2… Up to VN, it is simply a different way of stating this. Now, if you observe this quantity, the essence lies in this quantity here,

(Refer Slide Time: 12:13)



summation K equals to 1 to N V square N. Now, if you observe this quantity,

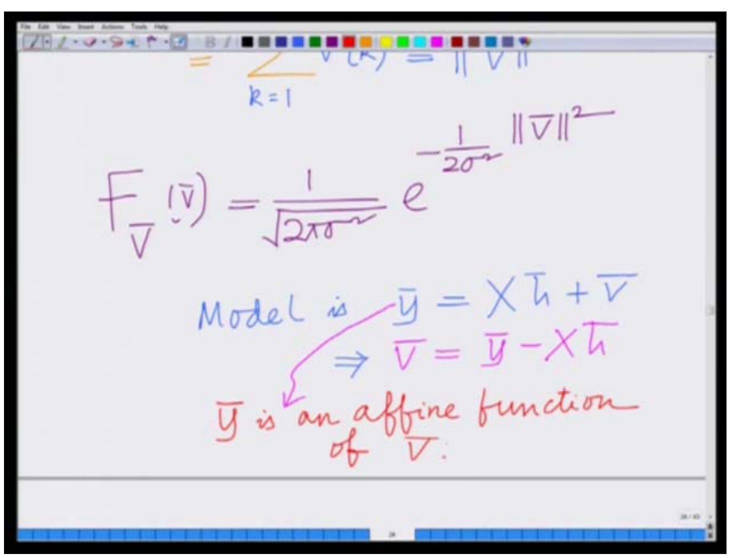
(Refer Slide Time: 12:22)



we know that our noise V bar equals V1, V2… Up to VN, therefore if you **are** now consider V bar transpose V bar, this is basically a row vector V1, V2… Up to VN times your column vector V1, V2… Up to VN, which you can now see is nothing but basically your summation K equals 1 to N V square K.

And which is also basically norm of V bar square. Summation K equals 1 V square K is basically nothing but V bar transpose times V bar and it is also basically norm V bar square. Where norm V bar is the norm of the vector of the Euclidean norm of the vector.

(Refer Slide Time: 13:33)

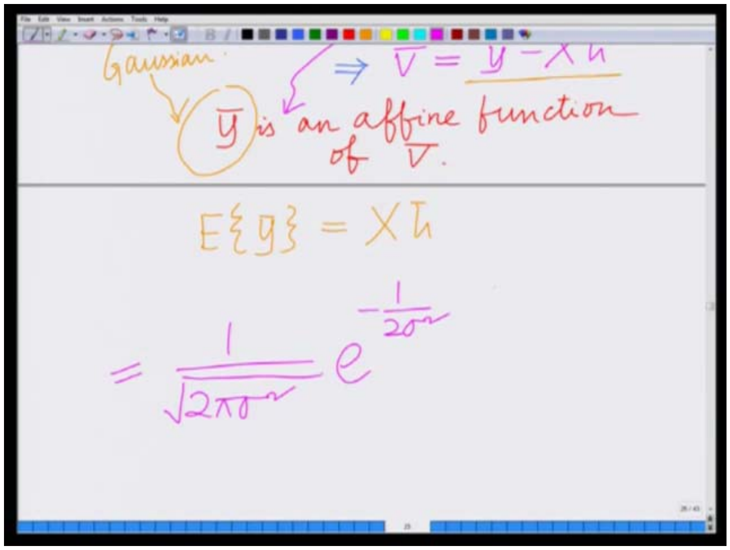


Therefore I can write the joint probability density function, that is this multivariate probability density function, in fact I can now write this as the multivariate probability density function F of V bar of V bar equals 1 over square root of 2 pie Sigma square E raised to -1 over 2 Sigma square, the summation of VK square, I am going to replace this by norm of V bar square. And there is an important point, notice that the summation, small point, but it is fairly important, replacing the summation K equals to 1 to N V square K by norm V bar square, where norm V bar is the Euclidean norm of the vector or the L2 or also known as the L2 norm of the vector V bar.

Now, once I replace this, now you can see basically, now go back to our system model, look at, go back to our system model, our system model is our model, model is Y bar equals H X bar class V bar is which means look at this, which implies basically your bringing V bar onto this side, V bar equals Y bar - X H bar, now one thing you can observe here is Y bar is linearly related or rather related in and affine fashion, Y bar is an affine function of V bar. Yah, V bar is a Gaussian random vector Y bar is basically simply a linear transformation + some shift, shift by a constant. If you have a linear transformation, it is a linear function but this is a linear transformation + the shift by this quantity X times H bar.

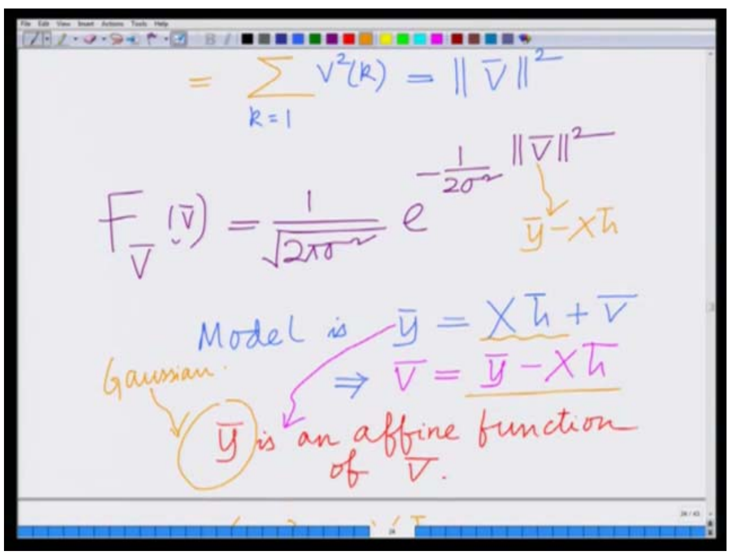
So, it is related in an affine fashion to this vector V bar which is a Gaussian vector, so Y bar is in turn, it is a Gaussian random variable, that is if you take a scale of Gaussian random variable and you shift it by some constant, then what you get is basically a Gaussian random variable, correct. So, Y bar is also a Gaussian random variable, in fact a Gaussian random vector in this context because it is a vector, so Y bar is Gaussian in nature and also observe that V bar is 0 mean, so the mean of Y bar is X H bar.

(Refer Slide Time: 16:10)



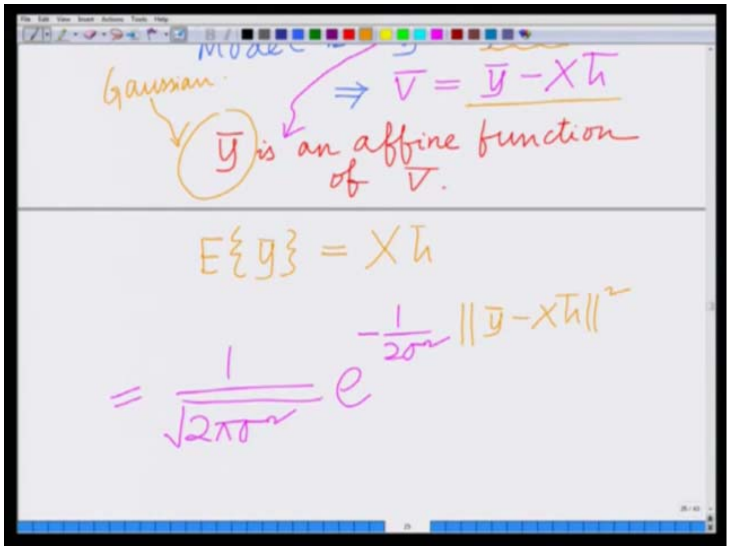
So, mean of Y bar or expected value as we can look at this,

(Refer Slide Time: 16:42)



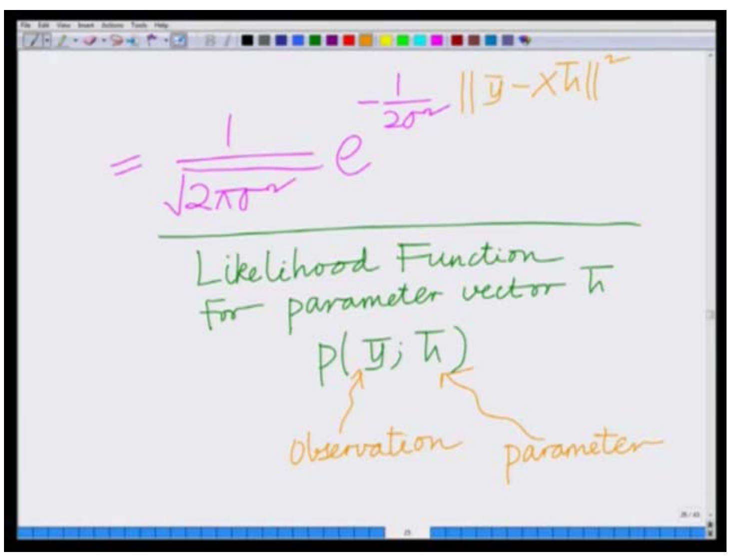
the mean of Y bar equals X H bar. Now what you can see is if I use this religion, V bar equals - Y bar - H X H bar and substitute this in a multidimensional probability density function, I have F of, in fact I can write the probability, joint probability density function as 1 over 2 pie Sigma square E raised to -1 over 2 Sigma square and look at this, this norm, this V bar, I can replace this by Y bar - X H bar.

(Refer Slide Time: 16:52)



So what I am going to have over here is basically I am going to have Y bar - X H bar square. This basically says Y bar is Gaussian with mean X H bar and this now remember, this is the probability density function but this probability density function, when I view it as a function of the unknown parameter vector H bar, this is now the likelihood function, the likelihood function for the parameter vector H bar.

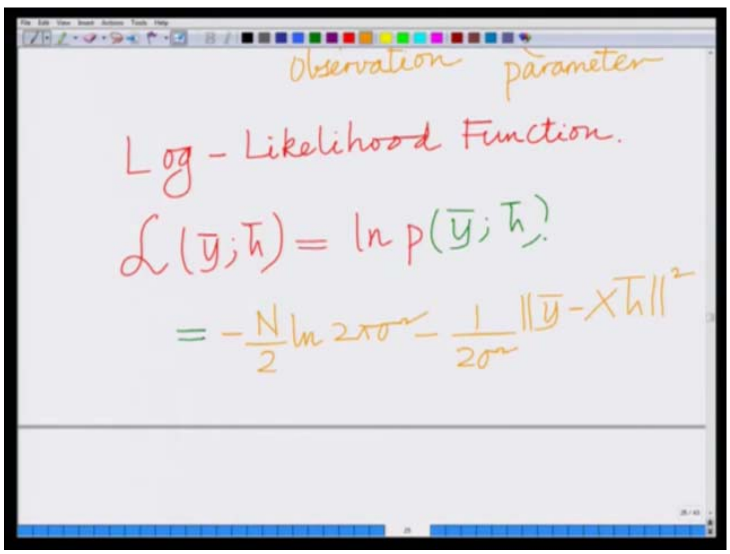
(Refer Slide Time: 17:22)



So, this probability density function is now basically your function for, for the parameter vector H bar. And I can denote this as similar to what we have doing been doing before, that is P of basically Y bar which is your observation vector parameter is by H bar, this is the likelihood function P bar P of the observations Y bar this is the observations parameterised by the parameter vector H bar.

The likelihood function of the observations Y bar parameterised by H bar which is the parameter vector. So, this is the, this to be extra clear, this is your observation vector and this is your parameter vector. This is your parameter vector, yah, so now what we have done is basically **we** we have achieved the aim which we initially set out set out for that is to basically develop a likelihood function towards the estimation of this parameter vector H bar. And now similar to what we have done before, we have the likelihood function, now we can consider taking the log, basically we can get the log likelihood function which is more amenable to the process of estimation, all right, that is what we said.

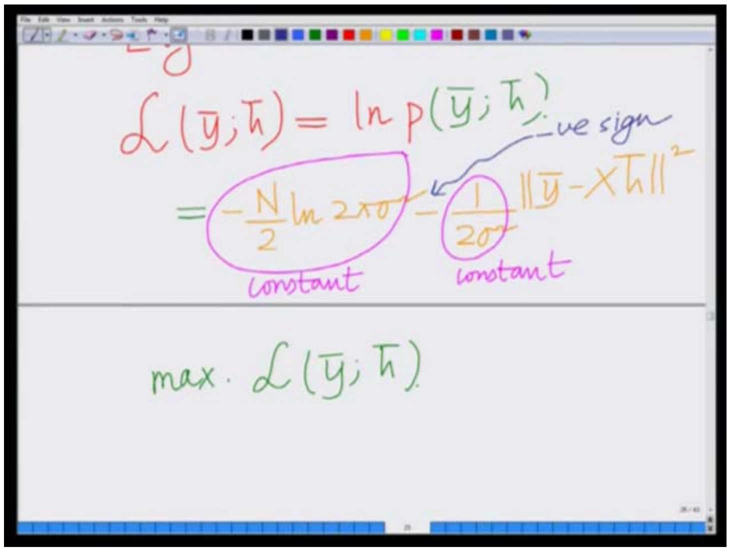
(Refer Slide Time: 19:02)



So, the log likelihood function, that is if you will now look at, again traversing exactly the same steps as we have done before, the log likelihood function, log like you would function of the observation parameterised by the parameter vector H bar is nothing but the natural logarithm of the likelihood function of that is P Y bar; H bar.

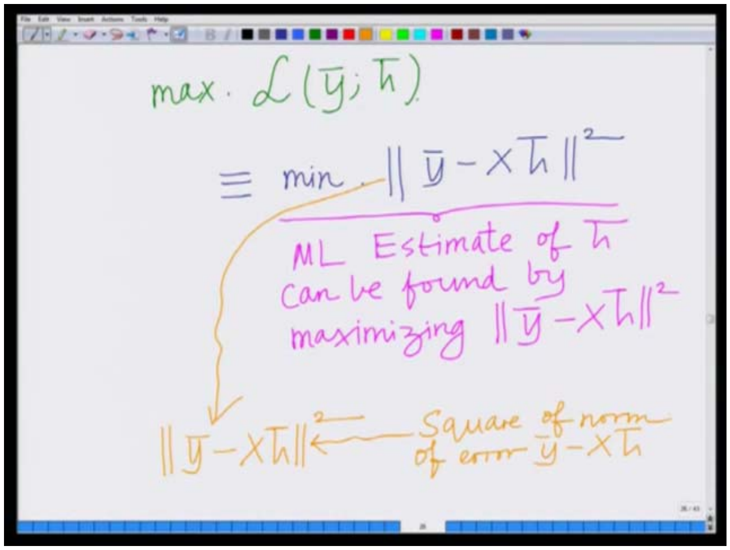
Remember that semi-colon plays an important role which is now if you take the logarithm of the above likelihood function, you can see, this is - N by 2 natural logarithm of 2 pie Sigma square -1 over 2 Sigma square norm Y bar - X X X, norm Y bar - X H bar whole square. Correct, so this is the log likelihood function corresponding to the parameter vector H bar, all right, we developed the likelihood function, taking the logarithm, we get the log likelihood function. Now we have to maximise this in order to obtain, we find, we have to find the value of H bar, that is that vector H bar for this log likelihood function, the likelihood function or basically the log likelihood function is maximise, that is the maximum likelihood estimate of the parameter vector H bar. So, what I am going to do is I am going to maximise this log likelihood function, I am going to maximise this log likelihood functions.

(Refer Slide Time: 20:40)



But when you maximise this log likelihood function, you observe that this is a constant - N by 2 log 2 pie Sigma square, this one over 2 Sigma square is also, this also is the constant, again, once again observe that there is a negative sign here, basically that changes the negative sign, so you remove the constant, you invert the negative… Because of the negative sign, the maximisation of the log likelihood function, basically you can see is equal into the minimisation of this part.

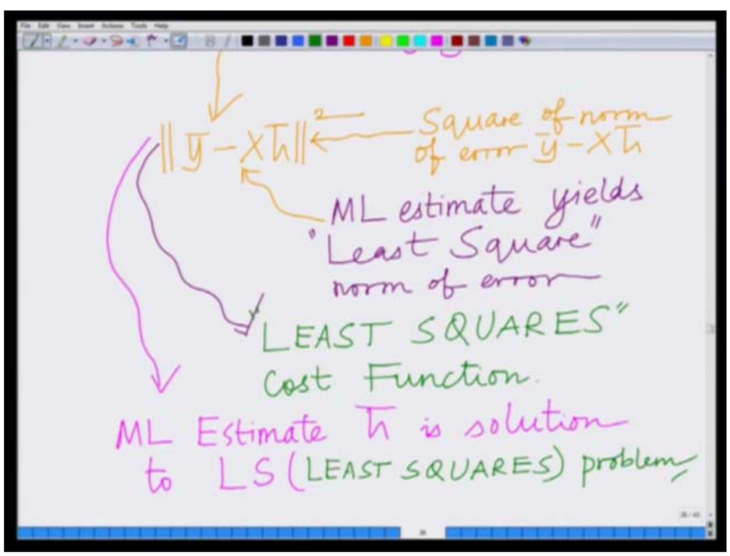
(Refer Slide Time: 21:27)



This is equivalent, that this is the symbol for equivalent, this is equivalent to minimising norm Y bar - X H bar whole square, that is one can find the estimate, that is**,** the ML estimate of H bar, H bar can be found by maximising norm Y bar - XH bar square.

And this basically, if you look at this quantity here, this is, look at this, this is basically norm of Y bar - X H bar whole square, that is basically the square of norm of the error Y bar - X H bar, this is the square of the norm of the error.

(Refer Slide Time: 22:56)



And we are trying to find the H bar which yields a least squared norm of the error, so H bar, ML estimate, the ML estimate yields least square that is the least square norm of the error.

Therefore this is frequently, this is very popular cost function, this is termed as your least squares or let me write that in bold letters, this is termed as your least squares. This arises very frequently in practice and this is termed as a least squares problem or the**,** the least squares cost function.

Right, so this arises very frequently in practice, this is known as the least squares cost function that is Y bar - H X bar Y bar - H X bar norm square, this is the norm square of the error, that is Y bar - H X bar and we are trying to find the X bar which minimises the norm square of this error. That is which finds the least squared, which finds that H bar which has the least square and, this is known as the least squares problem, this cost function is known as the least squares cost function. Therefore the ML, the maximum likelihood estimate of the channel vector H bar is given as the solution of this least squares problem.

And that is important to realise, so colloquially, so ML estimate is solution to LS, basically LS is the abbreviation of the least is solution of the least squares cost function or the least squares problem. Already, so basically what we have done in today's module is basically we have started with this system model, this vector system model Y bar equals X H bar + V bar for channel estimation in this multiple antenna system. And then what we have developed is we have developed the joint probability density function for the noise samples, from that we have derived the Gaussian nature of the observation vector Y bar and we have also developed the likelihood function for this unknown parameter vector, likely function for the estimation of the unknown parameter vector H bar and we have seen that maximising the log likelihood is basically equivalent to minimising norm of Y bar - H X bar square, this is known as the least squares cost function or the least squares problem because it basically corresponding to findings the H bar which yields the least squared norm of the error Y bar - H X bar.

Alright, so we have developed the likely function and basically formulated the least squares problem for estimation of the parameter vector H bar. In the subsequent modules we are going to solve the least squares problems and actually compute the estimate of the parameter, the maximum likelihood estimate of the parameter vector H bar. So, we will stop this module here, thank you very much.