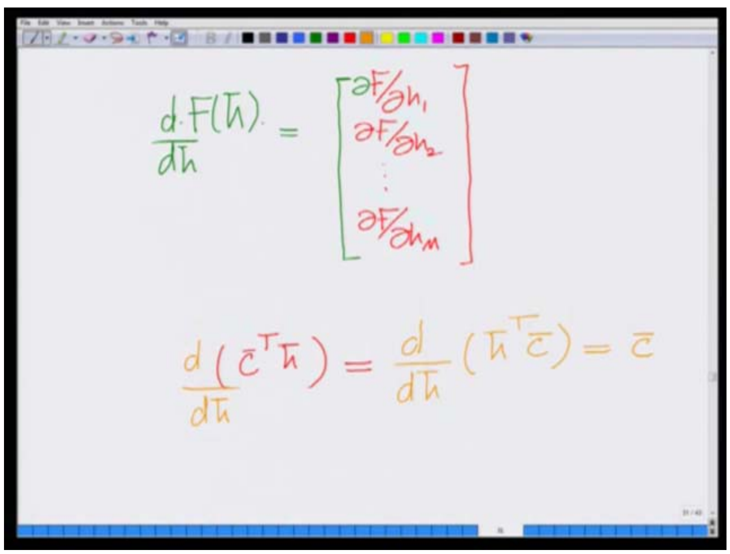
Estimation for Wireless Communications-MIMO/OFDM Cellular and Sensor Networks.
Professor Aditya K Jagannatham.
Department of Electrical Engineering.
Indian Institute of Technology Kanpur.
Lecture -16.
Least-Squares Solution, Maximum Likelihood (ML) Estimate Pseudo-Inverse.

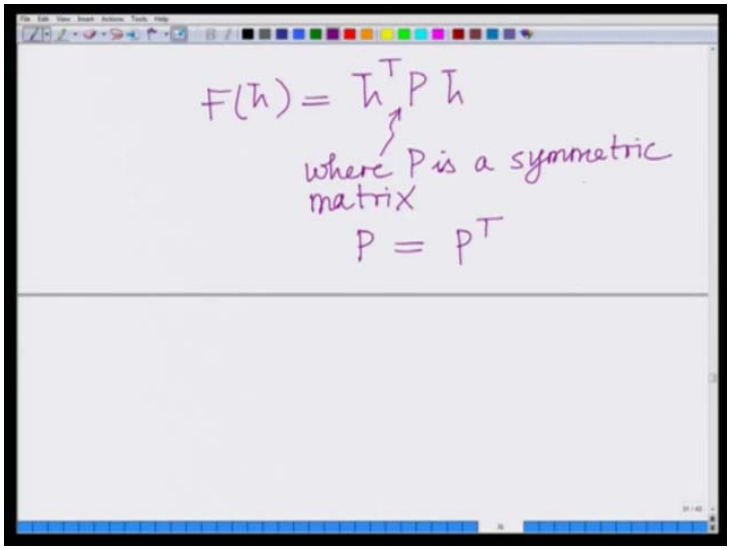
Hello, welcome to another module in this massive open online course on estimation for wireless communication. So, so far we are looking at the least squares cost function and we have simplified the least squares cost function for the estimation of a vector parameter. We said that we have to minimise this least squares cost function to find the estimate or the maximum likelihood estimate of the vector parameter H bar. And towards this end, we want to differentiate this least squares cost function and set it equal to 0, for the same purpose, we define the notion of a vector derivative or a gradient.

(Refer Slide Time: 0:46)



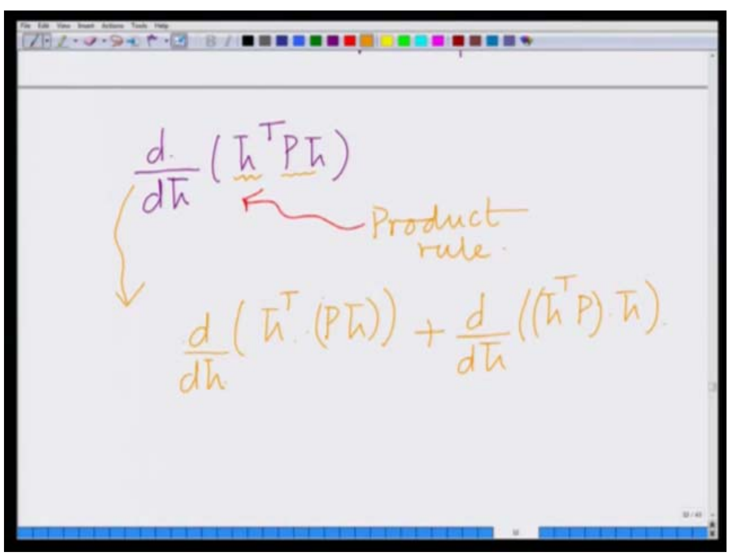
And we said for any function F of a vector parameter H bar, that is I have F of H bar dow F, that is the derivative d of this function, derivative of this function is simply given as the collection as a vector of partial derivatives with respect to the components of the vector of partial derivatives with respect to the components of the vector H. Correct. So, what we have over here is a vector derivative and we also proved the property that if my function that the partial derivative of C bar transpose H bar equals or the derivative, that is this function C bar transpose H bar is equal to the derivative of this function H bar transpose C bar which is indeed equal to the vector C bar. All right, so this is the property that we have proved with respect to the vector derivative d dF with respect to dH bar of the function F of a vector H bar.

(Refer Slide Time: 2:18)



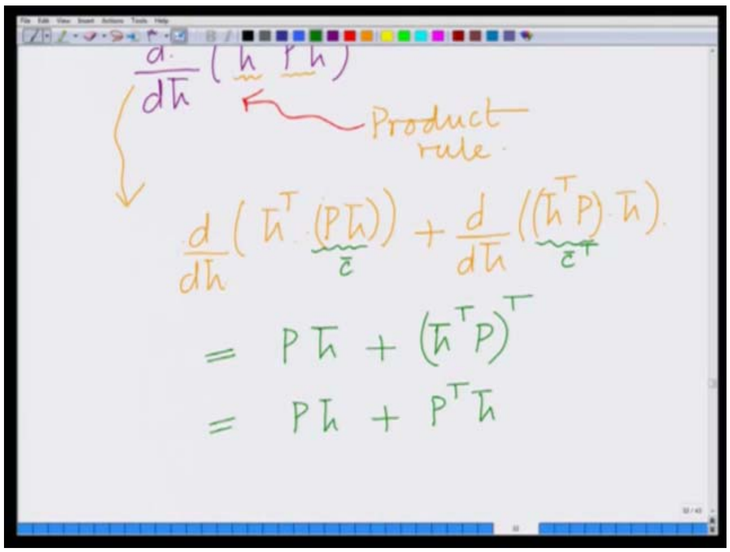
Now let us look at another function which is termed as the quadratic form and this is given as follows. That is if I have a vector H bar, the quadratic form is defined as F of H bar equals H bar transpose P H bar square matrix P is symmetric. P the symmetric matrix that is P equals P transpose.

(Refer Slide Time: 3:00)



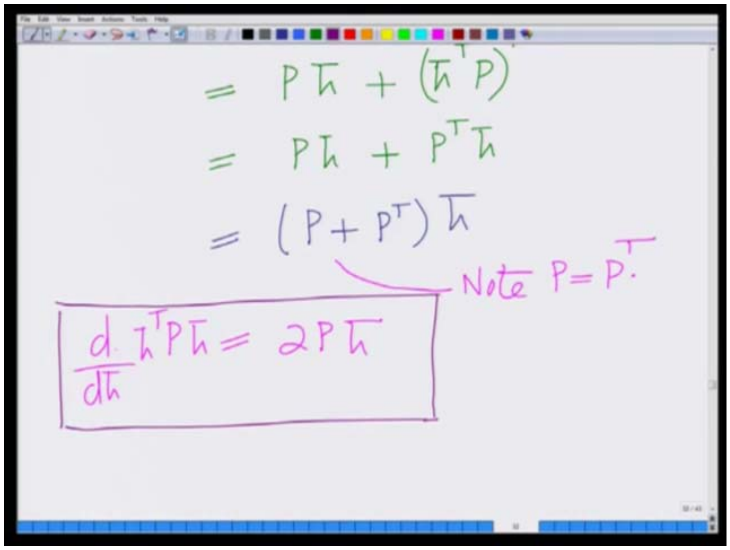
Now I were to differentiate this with respect to, that is considering the derivative of this function H bar transpose P H bar with respect to H bar, now I can do the following thing, this, I can use here, I can use the product rule for the derivatives, so I can use the product rule for computing the derivative and this is simply, therefore the derivative of well, the derivative of, this is simply well, the derivative of H bar transpose, 1st I treat a bar transpose as a constant and therefore this is the derivative of H bar transpose derivative of, that is treat H bar transpose as a constant and P H bar, therefore this is the derivative of, with respect to H bar, H bar transpose P H bar + the derivative with respect to the H bar of H bar transpose P, treat this as a constant times H bar, so basically this is your C bar, in this instance.

(Refer Slide Time: 4:30)



And this is your C bar transpose and we have already seen the derivative of H bar transpose C bar is nothing but C bar, so this is basically P times H bar + the derivative of C bar transpose H bar is also C bar, so this is basically H bar transpose P whole transpose which is basically P H bar + P transpose H bar which is equal to

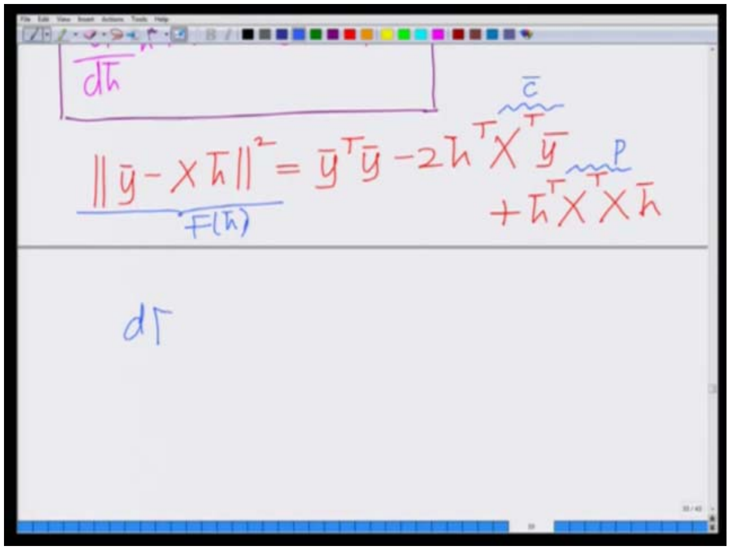
(Refer Slide Time: 5:04)



P + P transpose H bar. But we already, but note at this point note that matrix P is symmetric, which means P equals P transpose, therefore this is simply 2P times H bar, so the derivative of the quadratic form H bar transpose P H bar with respect to H bar is basically 2P times H bar. And this is another result that we have derived, that is a vector derivative of H bar transpose P H bar with respect to H bar is 2P H bar for a symmetric matrix P.

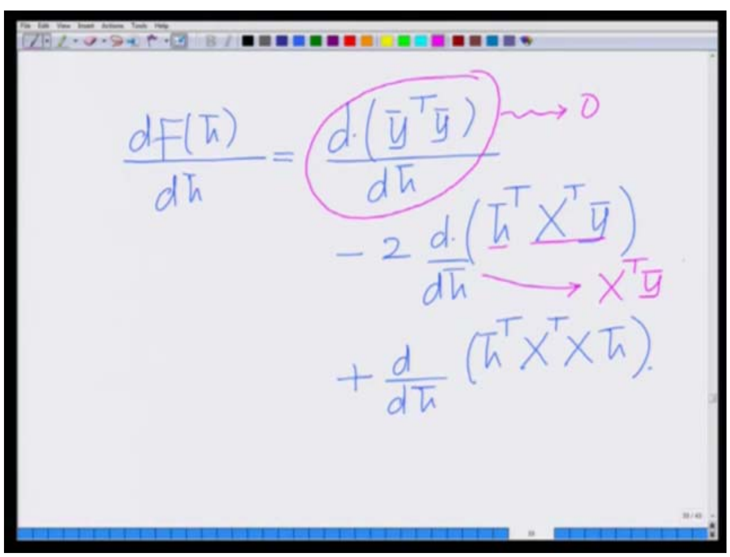
That is where matrix P is equal to P transpose. Okay. And now we have the results to basically differentiate the likelihood, the likelihood function or simplified or basically the least squares cost function towards minimising, towards minimising it so that we can find the estimate of the parameter vector H bar. And we have already seen, basically the least squares cost function can be simplified as, remember in the previous module we have already seen

(Refer Slide Time: 6:25)



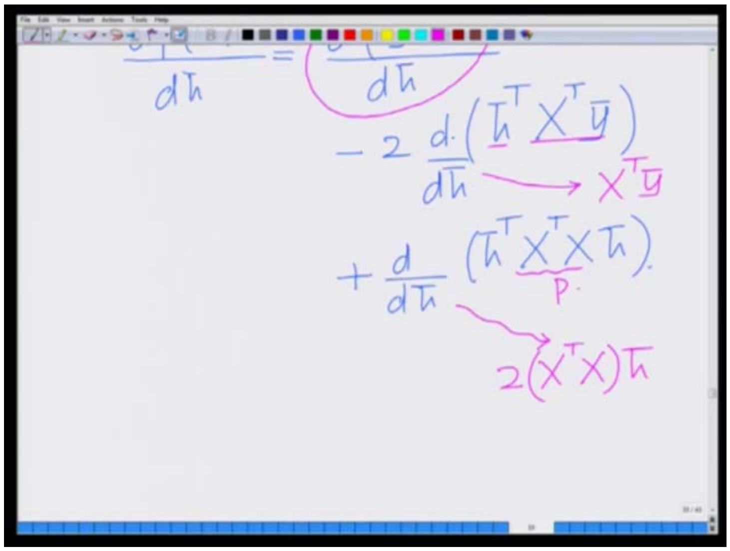
that the least squares cost function can simplified as Y bar transpose Y bar - 2H bar transpose X transpose Y bar + H bar transpose X transpose X times H bar. And now if you look at this, let us treat this, in this summation, let us treat this X transpose Y bar as your vector C bar and X transpose X, you can see this is a symmetric matrix, so this is P, therefore if you call this as your function, if you treat this as your function H bar, remember this is the function we are interested in differentiating.

(Refer Slide Time: 7:27)



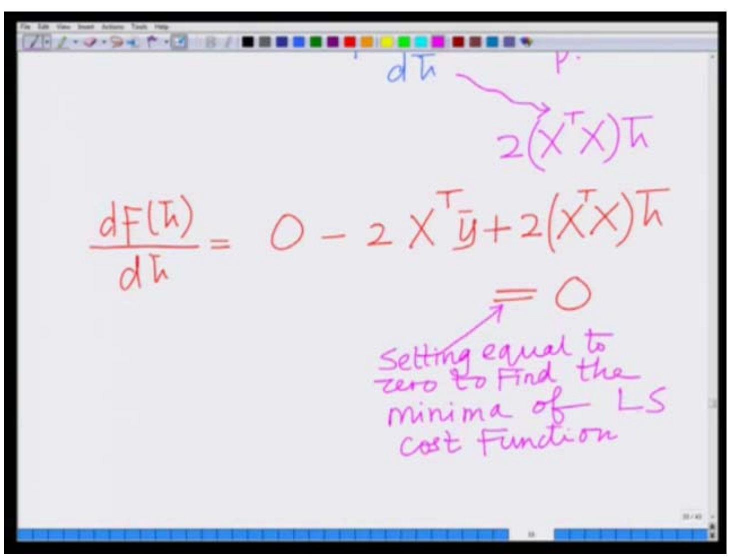
So, dF of H bar where F is your least squares cost function dF of H bar with respect to dH bar is equal to, well I have d of Y bar transpose Y bar with respect to H bar - twice d H bar transpose X times transpose Y bar with respect to dH bar + the derivative of, derivative of H bar transpose X transpose X H bar with respect to H bar. And now if you look at this, this quantity Y bar does not depend on your H bar, this is Y bar transpose Y bar, so this quantity is basically 0. Now this quantity is H bar transpose C bar, so the derivative of this is equal to X transpose Y bar and this quantity over here

(Refer Slide Time: 8:44)



is basically H bar transpose P times H bar which means this derivative is twice P, that is twice X transpose X times H bar.

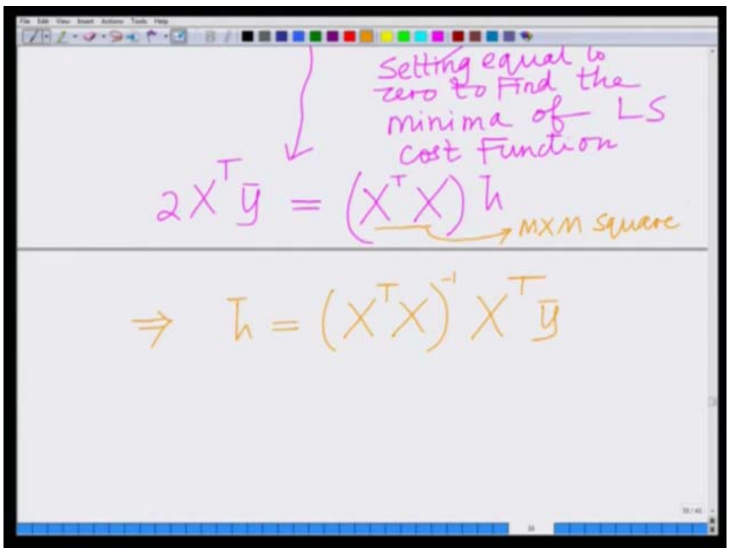
(Refer Slide Time: 8:57)



Which means that derivative now can be simplified as, the derivatives equal 0 - 2 X transpose Y + 2 X transpose X H bar which we are setting equal to 0, remember we have to set in derivative equal to 0 to find, we have remember, that is what we said, we have to differentiate this with respect to the vector H bar and set it equal to 0 to find the minima of this cost function. That is the parameter vector H bar where the derivative is where the derivative is 0 corresponds to the least squares cost function and that corresponds to basically the maximum likelihood estimate of the parameter vector H.

So, basically here you are setting equal to 0 **to 0** to find the minimum or the minima, let us put it that way. To find the minima of the least squares, that is your LS cost function.

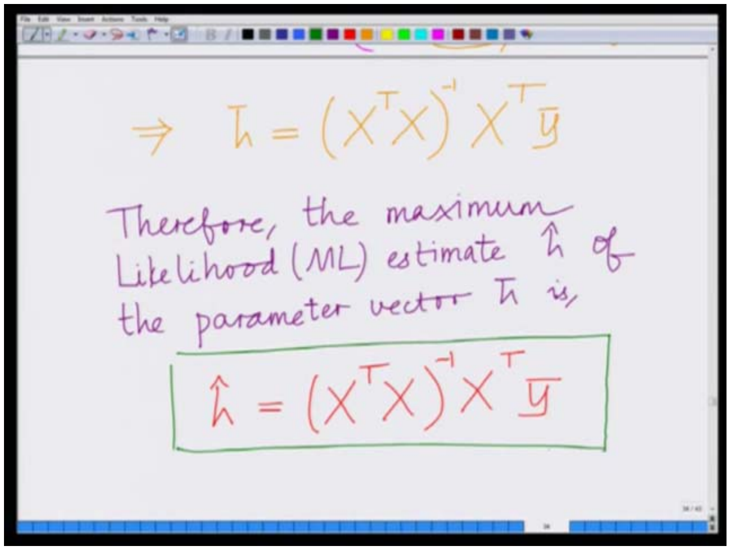
(Refer Slide Time: 10:28)



And this implies basically your 2X transpose Y bar is equal to X transpose X times H bar and you can see a beautiful relation, which basically implies the minima occurs where H bar is equal to taking this matrix X transpose X H X inverse times X transpose Y bar. So, this X transpose X, remember X is N x M, so this X transpose X is then M x M matrix, square matrix and if this is invertible, this is N x, this is an M x… X transpose X is an M x M square matrix, which means if this is invertible, I take it on the left.

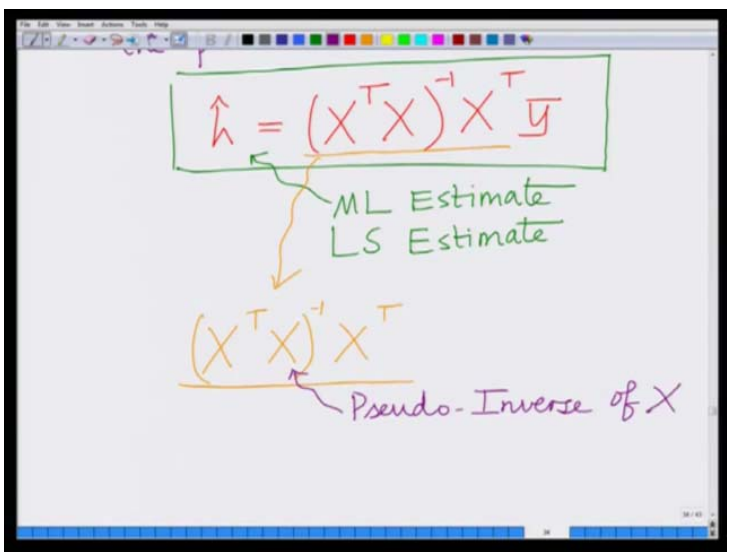
Therefore I have H bar equals X transpose X inverse times X transpose Y. This is the value of H bar where the derivative of the least squares cost function is 0, therefore this is the minima of the least squares cost function, therefore this value of H bar minimises the least squares cost function and hence it corresponds to the maximum likelihood estimate. I am going to denote this by H hat,

(Refer Slide Time: 11:54)



therefore, the maximum likelihood estimate, therefore the maximum likelihood, the maximum likelihood estimate H hat of the, of the parameter vector H bar is H hat equals X transpose or let me write it a little bit more clearly, with a different colour to highlight this.

(Refer Slide Time: 13:07)

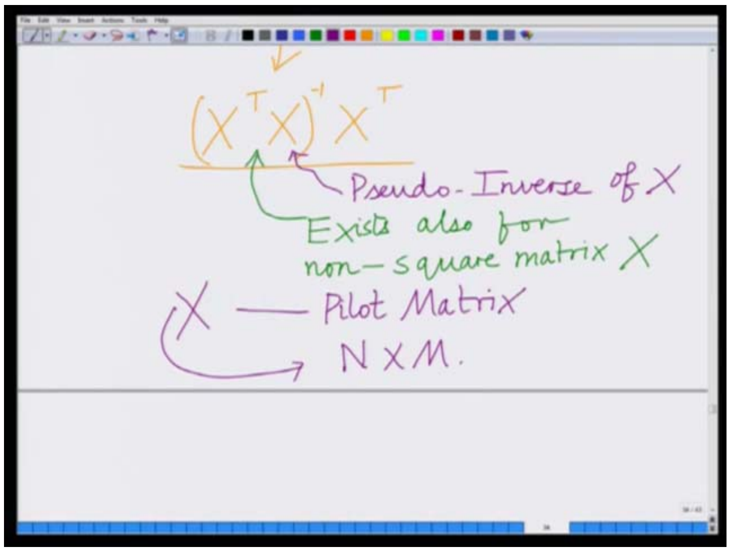


H hat equals X transpose X inverse X transpose Y bar. This is your least squares estimates. This is the ML estimate, also known as the least squares estimate, which in this case is also your LS estimate or the least squares estimate for the, this is also termed as the ML estimate, the maximum likelihood estimate of the channel vector H bar or also is the LS estimate where

LS stands for least squares. Since we said this is the least squares cost function, this also refers to as the least squares estimate of the channel vector H bar.

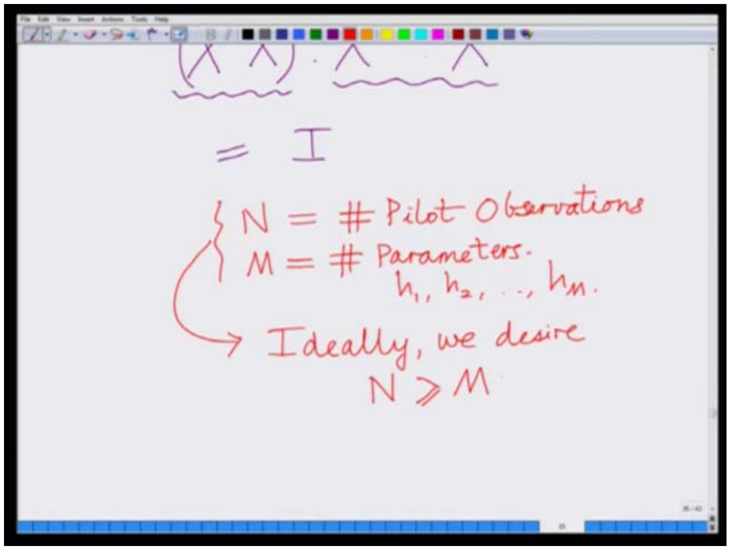
Let us now look at another interesting property of this matrix over here, if you look at this matrix over here, let us look at another interesting property of this matrix X. Or this matrix which is a function of X, that is X transpose X inverse X transpose, this matrix is termed as a pseudo-inverse, pseudo-inverse of it.

(Refer Slide Time: 14:15)



And exists even if N is not a square matrix, this exists, the only condition is that X transpose X has to be invertible, this exists also for non-square, remember we have X which is the pilot matrix, our X is basically this is the pilot matrix and this is an N x M matrix, correct, so this is the pilot matrix which is N x M matrix and it is not necessary that N be equal to M. We can have number of pilot vectors or observations N greater than M, in fact it would be desirable, we would like to have as many observations or as many pilot observations as possible, we would like to have as many pilot observations as possible. So, we can consider the case where N is in fact simply greater than M and therefore what we are going to show now is

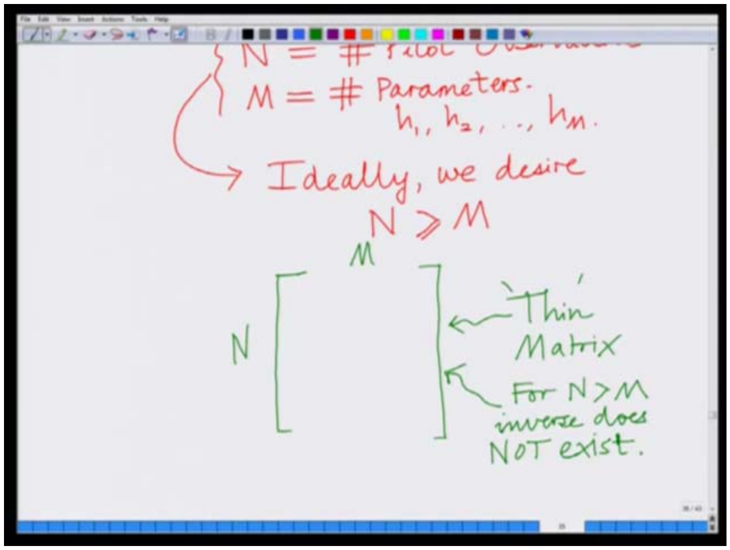
(Refer Slide Time: 15:26)



if you look at this matrix X transpose X inverse X transpose times the matrix X where X is the pilot matrix times the matrix X, look at this, we have X transpose X inverse, we have X transpose X, so therefore this is equal to the identity matrix.

So, we have X which is not necessarily I square matrix, X is an N x M matrix N is possibly greater than M, N is the number of pilot observations and M is the number of unknown parameters, so ideally we would like to have N as high as possible, which means the vector X, the matrix X has a large number of rows, much more, that is the number of rows is much larger than the number of columns, all right. Here we can consider a scenario that N is greater than M, remember N is equal to number of pilot and M is equal to number of parameters in the parameter vector. That is your H1, H2,… up to HM and ideally we desire N to be greater than, that is N to be greater than or equal to M and remember,

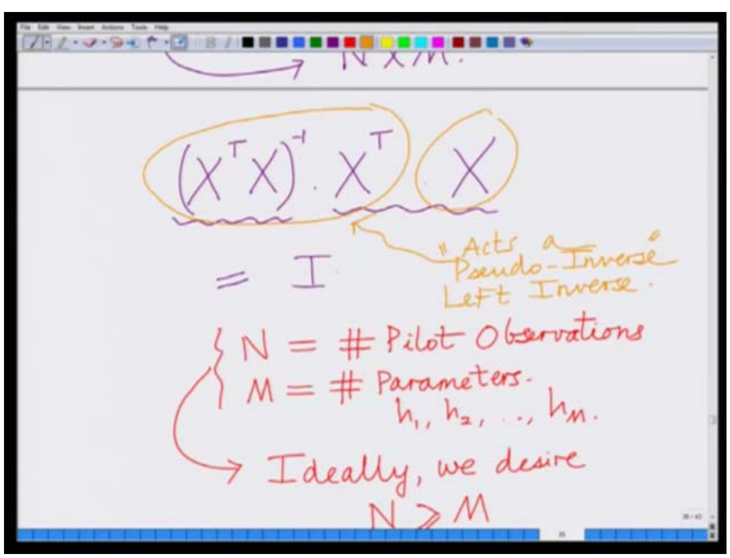
(Refer Slide Time: 17:20)



when N is greater than or equal to M, our matrix X is N x M and this matrix is basically also known as a thin matrix.

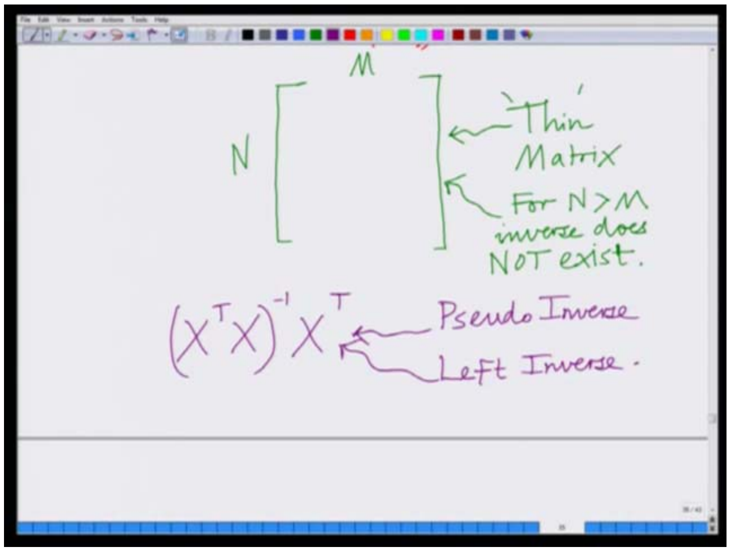
Colloquially, this matrix is also known as a thin matrix. Which means the number of rows is much larger than the number of columns. It is, basically it means that if N is greater than M, it is not a square matrix, therefore X is not invertible or it does not have an inverse. Remember, inverse exists only for a square matrix. So, if N is greater than M, where X cannot be, X is a non-square matrix, hence the inverse does not exist. So, for this thin matrix, for N greater than M, your inverse does not, the inverse does not exist.

(Refer Slide Time: 18:26)



However, if you look at this X transpose X inverse X transpose times X equals identity, so this matrix, this acts as a pseudo-inverse, that is not an inverse but acts as an inverse, that is known as a pseudo-inverse. Or basically also known as the left inverse of X because when multiplied on the left, so when this matrix multiplies X on the left, then you get identity, this matrix is also known as the left inverse.

(Refer Slide Time: 19:06)



So, let us summarise that, I have this quantity, this is also known as the pseudo-inverse or also known as the left inverse of your matrix X. Alright, so basically that is the least squares estimate. So, what we have derived is basically in this module what we have looked that is a very important aspect of vector parameter estimation. Basically we have differentiated, we have considered the vector derivative of the least squares cost function, set this vector derivative equal to 0 and found the value of that parameter vector H bar where this vector derivative is0. Or in other words, where the least squares cost function is minimum. That we have shown is basically given by X transpose X inverse X transpose Y bar where X is basically the pilot matrix, Y bar is the better of observations and this we have come to be the maximum likelihood estimate.

Since this corresponds to a minima of the least squares cost function, this is the maximum likelihood estimate, it is also, since it is also basically a solution that minimises the least squares cost function, this is also known as the least squares estimate. This is the least squares estimate of the parameter vector H bar. Remember the parameter vector H bar is the vector of channel coefficients corresponding to a multi-antenna, remember, that is what we started with, it basically contains the vector of channel coefficients corresponding to a multi-antenna transmission scenario where the base station is transmitting to a mobile, when the base station has M antennas, the mobile has a single antenna and H1, H2… HM corresponds to basically the channel coefficients of these M antennas. Right, so basically this gives the least squares or the maximum likelihood estimate of this parameter vector H bar and this is also the maximum likelihood estimate. So, we will stop this module here and explore other properties of the maximum likelihood estimate in the subsequent modules, thank you very much.