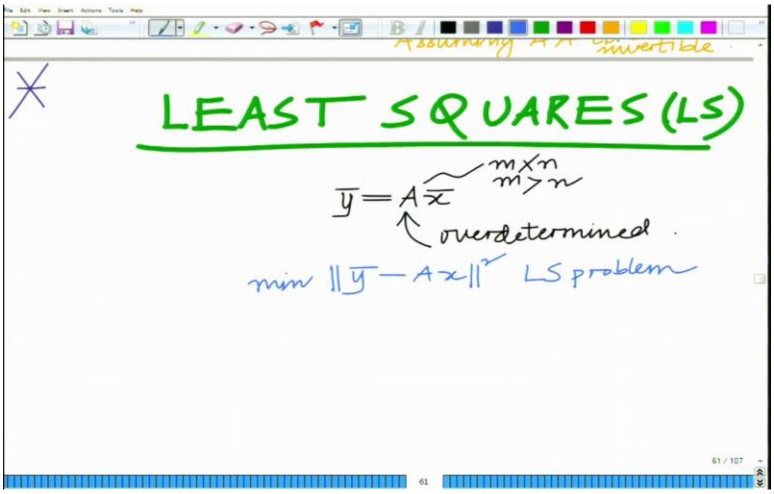
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 42
Geometric Intuition for Least Squares

Keywords: *Least Squares*

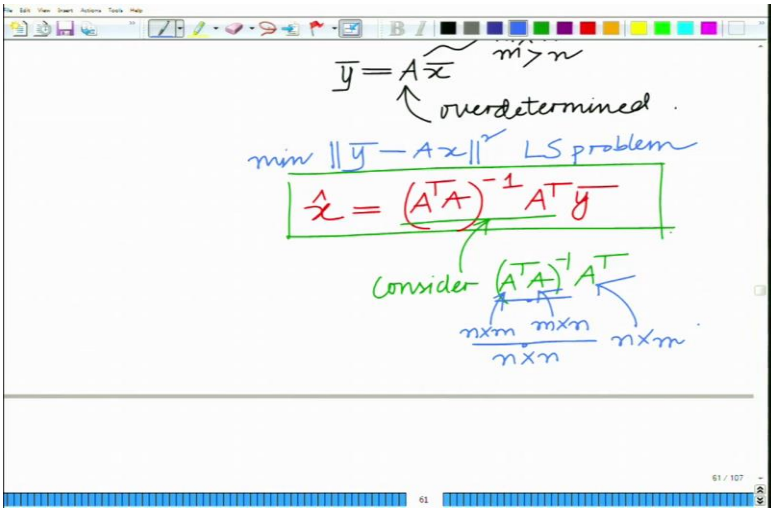
Hello, welcome to another module in this massive of online course. So we are looking at the least squares optimization problem and we also derived the least squares solution.

(Refer Slide Time: 00:26)



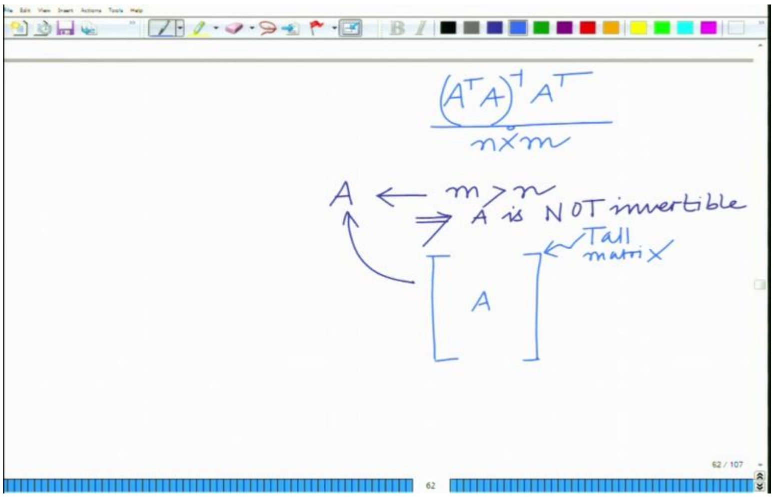
So when we have an over determined system of equations $\bar{y} = A\bar{x}$ we minimize the norm that is $\min \left\|\bar{y} - A\bar{x}\right\|^2$ and this is termed as the least squares problem. The solution to this is $x = \left(A^T A\right)^{-1} A^T \bar{y}$.

(Refer Slide Time: 01:30)



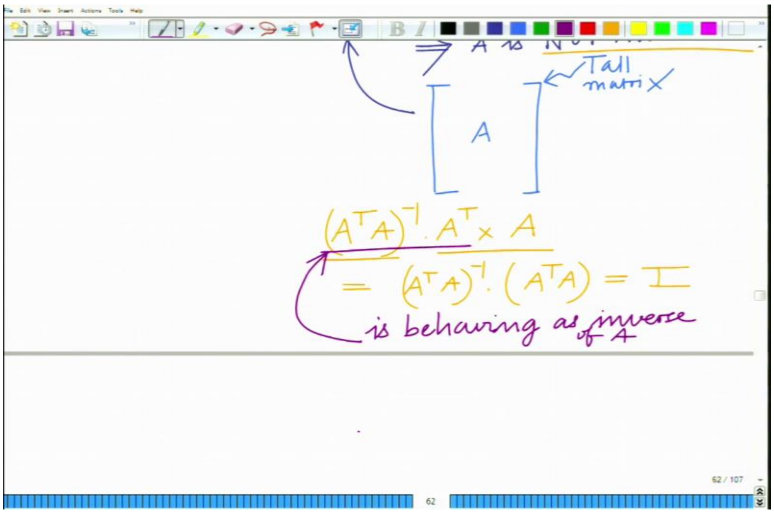
Now, consider the matrix $\left(A^T A\right)^{-1} A^T$, we can see that the size of this matrix is $n \times m$.

(Refer Slide Time: 03:05)



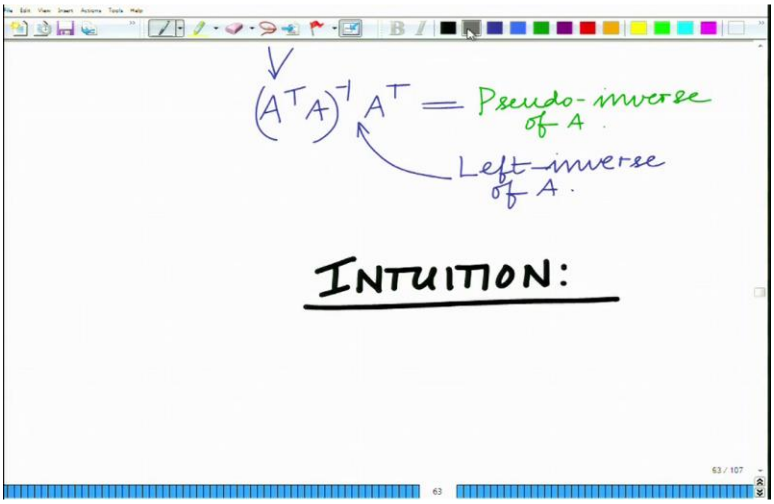
So this has more columns than rows and for m > n this implies that A is not invertible. Now, we are considering a scenario in which the number of rows is much greater than the number of columns, so this is also known as a tall matrix.

(Refer Slide Time: 04:14)



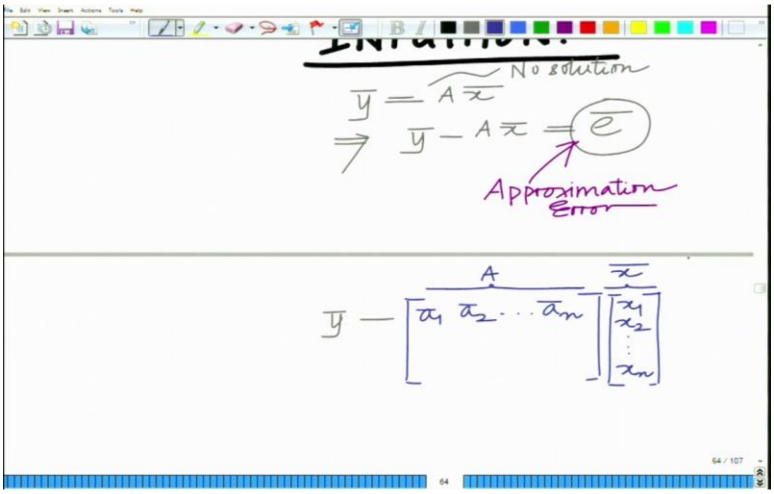
If you look at $(A^T A)^{-1} A^T$ and you take its product with A we get an identity matrix. So it is as if $(A^T A)^{-1} A^T$ is acting as an inverse of A when multiplied on the left because A is not invertible when m > n but it is behaving as an inverse and this is therefore known as the pseudo inverse of A.

(Refer Slide Time: 05:40)



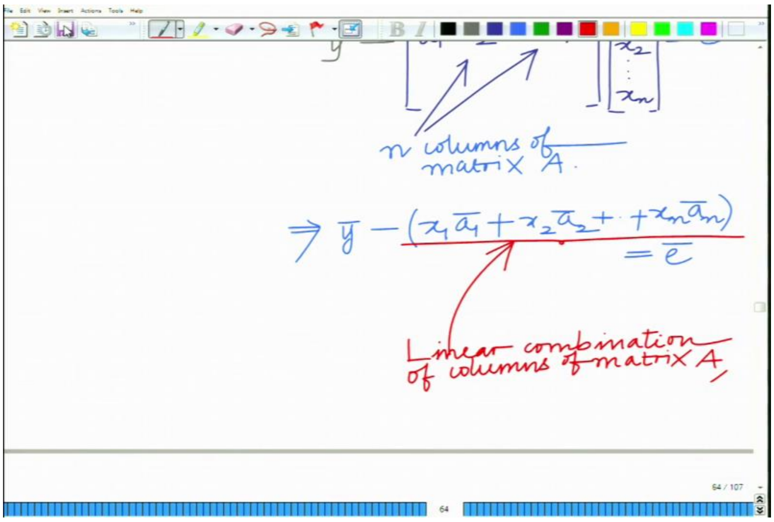
So this is a left inverse because it is only true when you multiply it on the left.

(Refer Slide Time: 06:48)



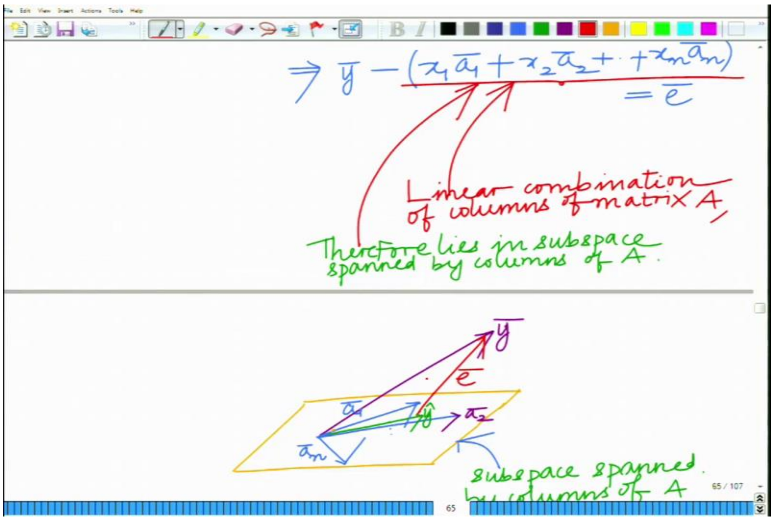
We have the least squares problem $\bar{y} = A\bar{x}$ and we know that there does not exist any solution. So this implies that no matter what $\bar{x}$ you choose, it will not satisfy $\bar{y} = A\bar{x}$ which means $\bar{y} - A\bar{x} = \bar{e}$ and will always be non-zero. So this is the approximation error and this is represented as shown in slide.

(Refer Slide Time: 08:38)



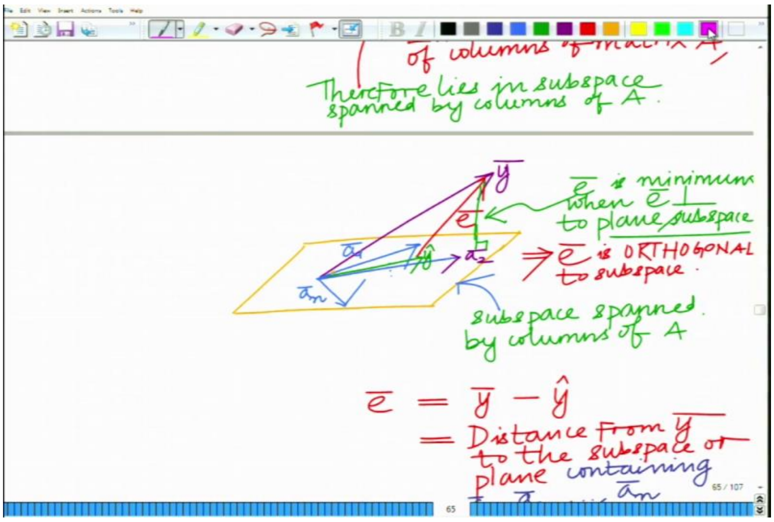
So these are the n columns of the matrix A and this is $\bar{y} - \left(x_1\bar{a}_1 + x_2\bar{a}_2 + \ldots + x_n\bar{a}_n\right) = \bar{e}$ and on expansion of this we get a linear combination of the columns of matrix A which implies that this always lies in the subspace spanned by the columns of A.

(Refer Slide Time: 10:27)



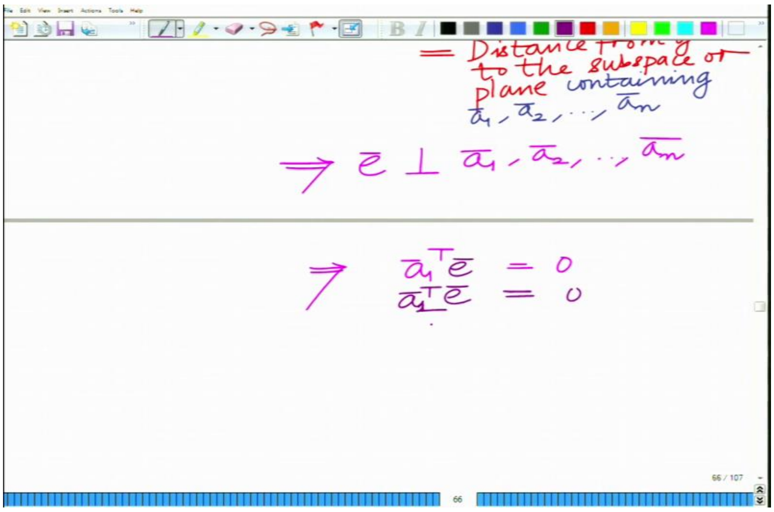
So this can be represented pictorially if you take a plane which is nothing but a subspace. So let us say this is the subspace spanned by columns of A and you have your vector $\bar{y}$ which does not necessarily lie in the subspace. So you are trying to form an approximation which lies in this subspace and let us say you denote this by y. So $\bar{y} - y = \bar{e}$ is the corresponding error.

(Refer Slide Time: 12:51)



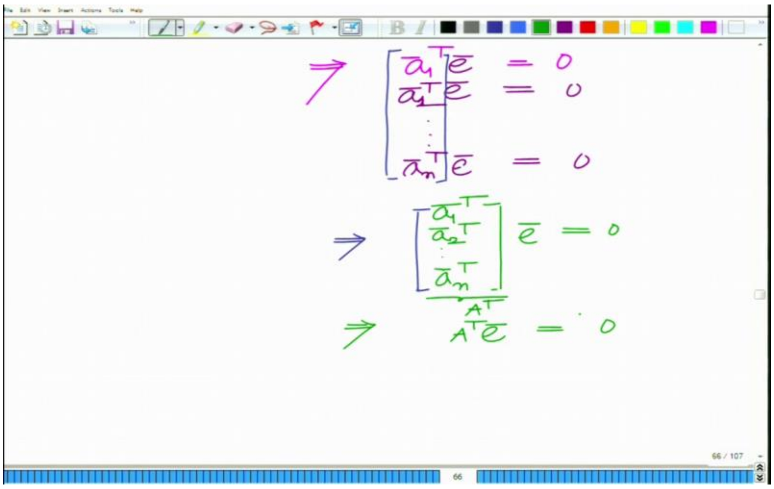
So this is the distance from $\bar{y}$ to the subspace or the plane that is spanned. Now we want to find the error and we want to minimize this error. And now you can see that this error is minimum when the error is perpendicular to the plane or we can say that this error vector is orthogonal to the subspace.

(Refer Slide Time: 15:35)



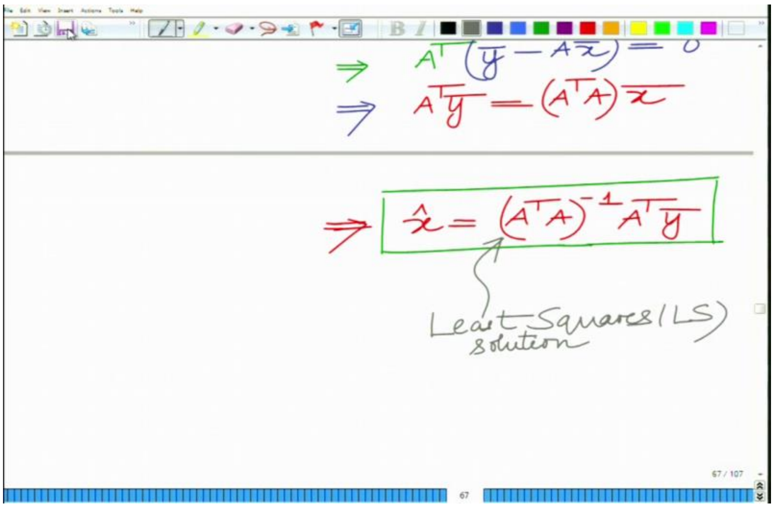
So this $\bar{e}$ is orthogonal to the subspace, this implies that $\bar{e}$ has to be orthogonal to each of the vectors in the subspace as shown in slide. We know that two vectors are orthogonal when their inner product is 0.

(Refer Slide Time: 16:23)



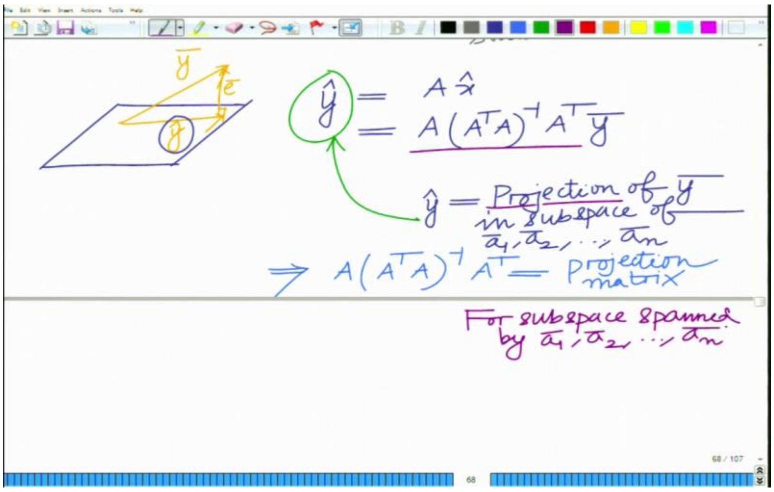
So this implies $A^T \bar{e} = 0$ as shown in slide.

(Refer Slide Time: 17:13)



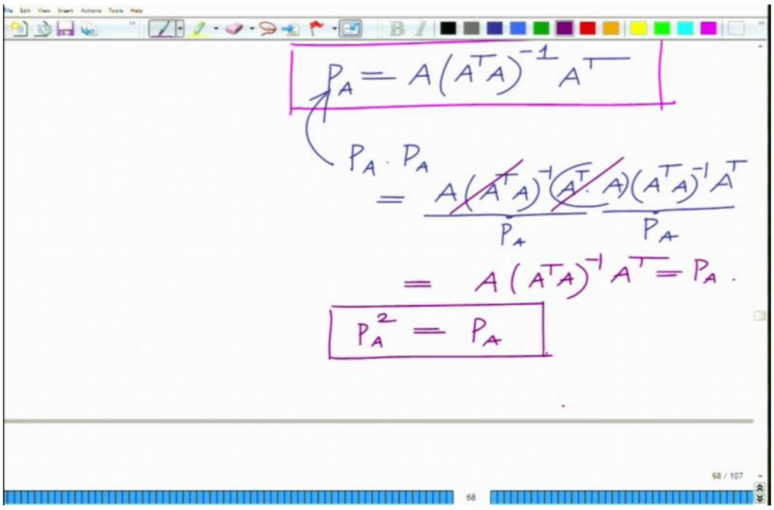
But, $\bar{y} - A\bar{x} = \bar{e}$ so this implies $A^T(\bar{y} - A\bar{x}) = A^T\bar{e} = 0$. Now you observe that the best vector $\bar{x}$ that minimizes the error is $x = (A^TA)^{-1}A^T\bar{y}$. So this is nothing but again the least square solution, so intuitively what the least square solution is doing is basically finding the best approximation to $\bar{y}$ in the subspace that is spanned by the columns of the matrix A. And therefore the distance of $\bar{y}$ to the plane is minimum when the error vector is perpendicular to the plane. Now $y = Ax = A(A^TA)^{-1}A^T\bar{y}$.

(Refer Slide Time: 19:15)



Now, if you look at this plane again you can also say that y is the projection of $\bar{y}$ in the subspace spanned by the columns of A, so when you multiply this matrix by $\bar{y}$ you get the projection. So this matrix is the projection matrix, that is $P_A = A\left(A^T A\right)^{-1} A^T$ is the projection matrix for the subspace that is spanned by the columns of A.

(Refer Slide Time: 21:40)



So one of the most interesting properties of the projection matrix is that $P_A^{\ 2} = P_A$, in fact, $P_A^{\ n} = P_A$ for any integer $n \geq 1$. So this is the intuition behind the least square solution. So we will stop here and continue in the subsequent modules. Thank you very much.