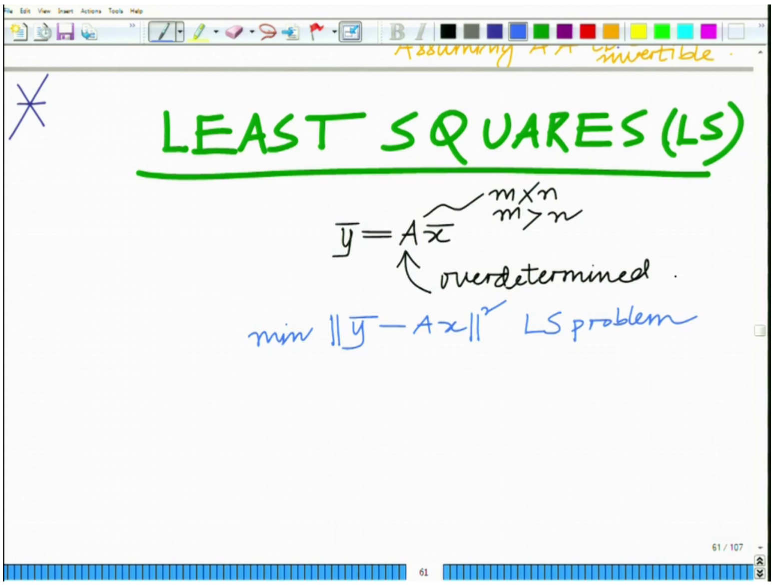
Applied Optimization for Wireless, Machine Learning, Big Data
Prof. Aditya K. Jagannatham
Department of Electrical Engineering
Indian Institute of Technology, Kanpur

Lecture – 42
Geometric Intuition for Least Squares

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 that is the solution to the least squares optimization problem.

(Refer Slide Time: 00:26)

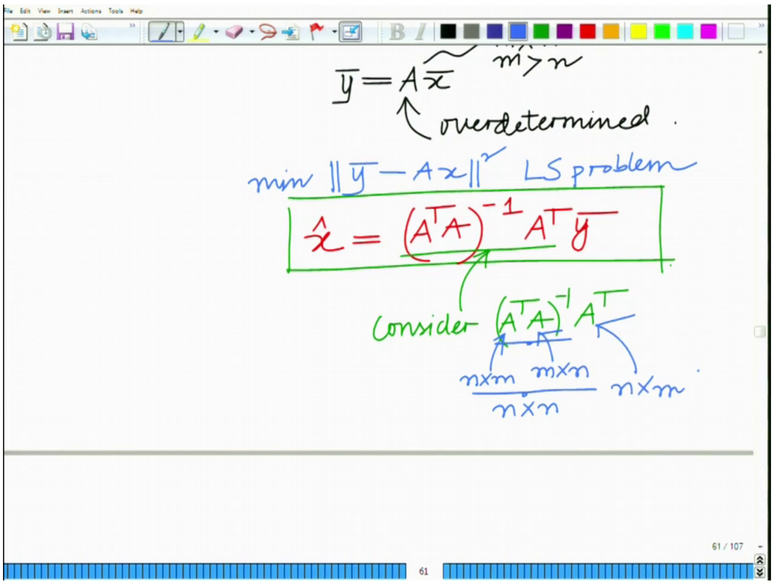


So, let us continue our discussion so, we are looking at the least squares which is a very important optimization problem. You can also think this is an approximation or modeling problem.

And what we have seen is that when we have an over determined system of equations y bar equals A x bar with a being an m cross in matrix and m greater than n this is over determined correct. And therefore, to summarize and this cannot be solved exactly therefore, what we do is we minimize norm y bar minus A x bar square this is termed as the least squares problem.

And the solution to this least square solution we have derived.

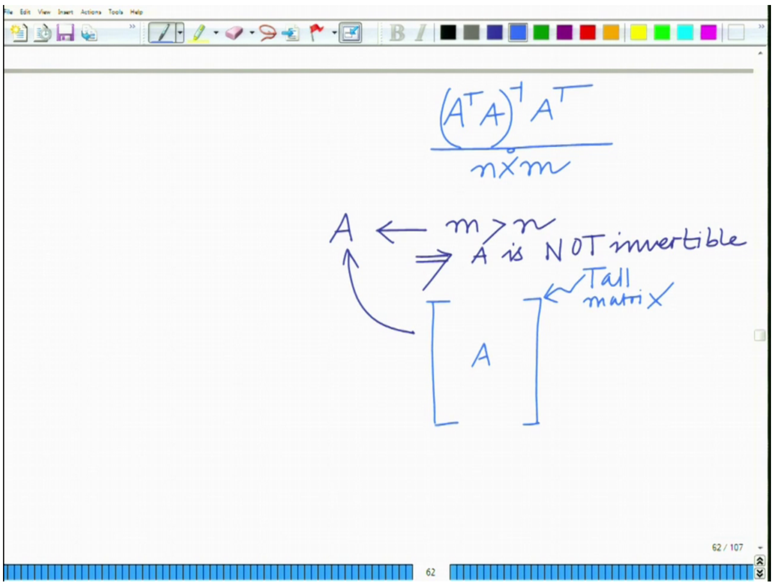
(Refer Slide Time: 01:30)



This previously that is x hat equals A transpose A inverse A transpose into y bar ok. And now, if you look the now let us look at some aspect salient aspects of this solution, now consider this matrix A transpose A inverse into A transpose ok.

Now, consider the matrix A transpose A inverse into A transpose. Now, we can see that you can easily say that this size of this matrix. So, A transpose m cross n A is n cross A is m cross n A transpose n cross m. So, this is basically your n cross, this is n cross m A transpose is this n A transpose n cross m. So, this is n cross m, this is m cross n.

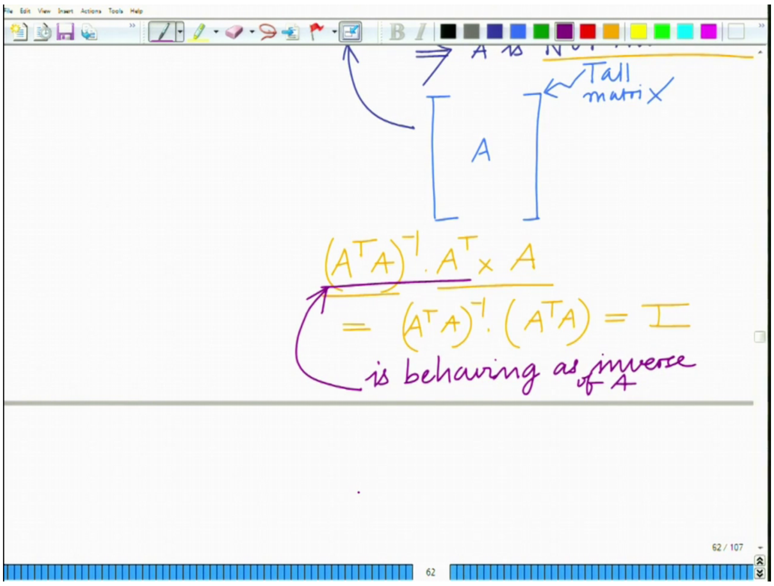
(Refer Slide Time: 03:05)



So, A transpose A’s n cross n and A transpose is n cross m implies if you look at A transpose A inverse that is therefore, A transpose A inverse into A transpose this is you can easily say this is n cross m matrix ok m greater than n. So, this has more columns then rows. But, the interesting thing here is now if you look at A for m greater than n non square matrix this implies that A is not invertible.

Now, we are considering a scenario in which m greater than n which means the number of rows is much greater than the number of columns so, A looks like this. So, this is also known as a tall matrix that is the height number of rows of the matrix is much larger than the number of columns the matrix looks like a tall matrix. Now, for non square matrix; obviously, does not have inverse, but you can observe a very interesting aspect.

(Refer Slide Time: 04:14)

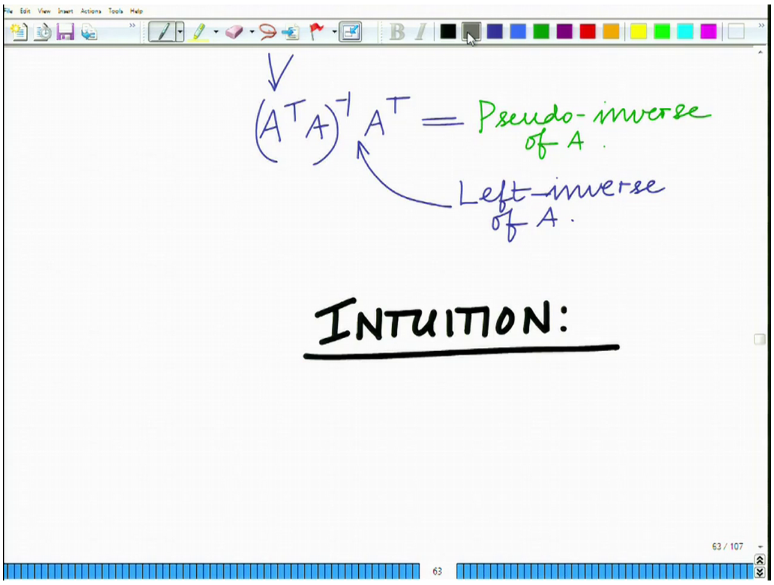


That is if you look at A transpose if you look at this matrix A transpose A inverse into A transpose and you multiply you take it is product with a now you look at this. So, A if m greater than n it implies that it is not invertible, but if you look at this A transpose A inverse into A transpose multiply this with A.

Now, look at this we have A transpose A inverse we have a transpose. So, this is basically your A transpose A inverse times A transpose A which is identity. So, it is as if all the A is not invertible, it is as if this matrix A transpose is inverse into A transpose this matrix acts is acting, is behaving you can say is behaving as an inverse of A right. When multiplied on the left with A it is giving identity.

Let us it behaving it not an inverse because A is not invertible when m is greater than n, but it is behaving as an inverse this is therefore, known as the pseudo inverse of A. Pseudo is an quantity pseudo when it is not actually the quantity, but it gives the appearance of that quantity ok.

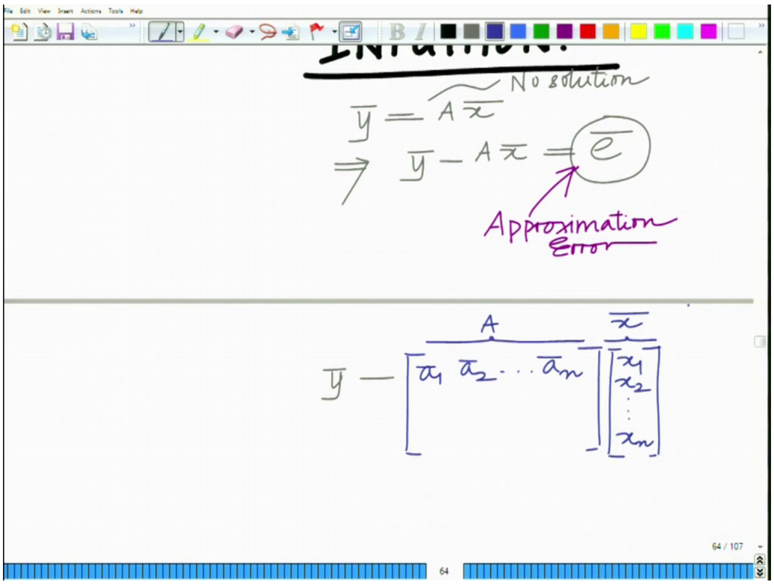
(Refer Slide Time: 05:40)



So, that is basically termed as this quantity A transpose A inverse into A transpose this is termed as the pseudo. This is termed as pseudo inverse of A and this also remember a left inverse because, it is only true when you multiply it on the left.

This is also the left not the it is a left inverse of the matrix. Now, to understand the explore of this further this nature and to understand to get some intuition behind the solution. So, what you want to do is want to get some intuition behind the least square solution and the intuition is very interesting.

(Refer Slide Time: 06:48)



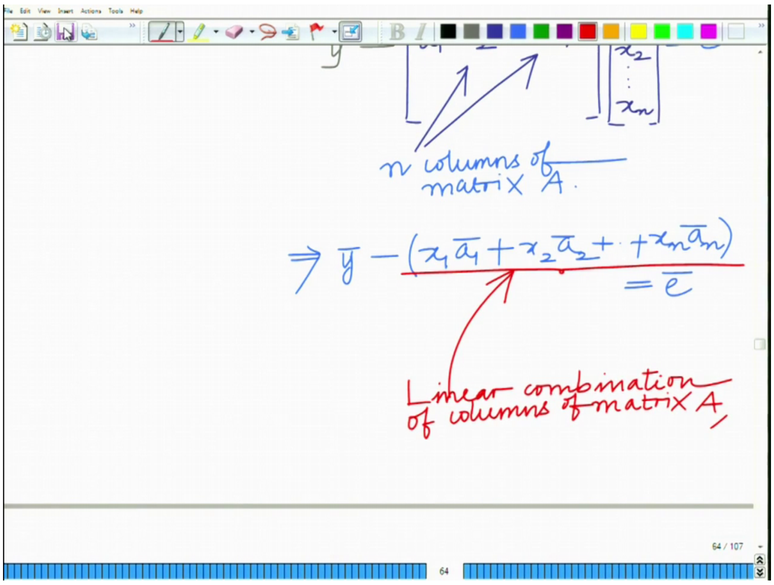
The intuition behind the least square solution now if you look at our problem we have y bar equals A x bar this is our problem.

But, we know that they does not exist any x bar that is this does not have any solution. So, this is does not have any solution which implies that no matter that x bar you choose they it will not satisfy y bar equal to y bar equals A x bar. Which means y bar minus A x bar will always be non-zero that can be denoted by vector e bar.

So, there is no vector x bar so, this is an over determined system. Remember that is what we said unless and if you consider 3 equations 3 equations that 2 variables and you have 3 lines and unless they all intersected a point, it does not have any solutions. So, there will always be an approximation here let us denote that by e bar. So, this is the approximation error.

So, this is the approximation error. Now, therefore, y bar minus A x bar equals e bar this is a approximation error now let me write it a little explicitly. So, this is will be y bar minus A is an m cross n matrix which means it has n columns. So, these I can denote by a 1 bar a 2 bar upto a n bar times x 1 x 2 up to x n. So, this is your matrix A and this is x bar.

(Refer Slide Time: 08:38)

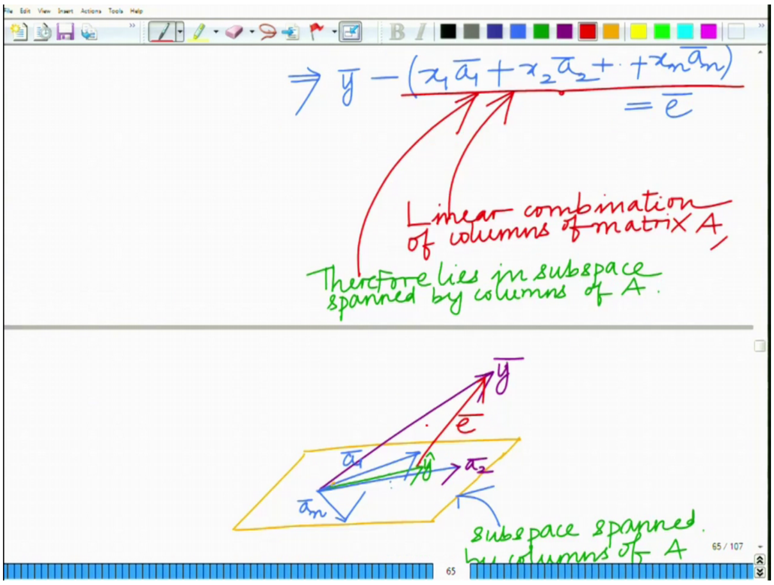


So, these are the n columns n columns of the matrix A and so, this is equal to e bar and. So, this implies that y bar minus now we should multiply this out. So, you will get a 1 bar times x 1 times a 1 bar plus x 2 times a 2 bar.

So, this minus x 1 times a 1 bar plus x 2 times a n a 2 bar plus so on x n times a n bar this is equal to e bar where a vector and now if you look at this, if you look at this x 1 times a 1 bar explicit x 2 times a 2 bar times so, on until x n times m. And now this is nothing, but a linear combination of the columns of matrix A. What this is? Is this represents a linear combination a linear combination of the columns of the matrix A.

So, now, you have this linear combination of the columns which implies that this approximation that is A times x bar always lies in the subspace spanned by the columns of A. Remember linear combination you can consider all the linear combinations of the these columns you get the subspaces.

(Refer Slide Time: 10:27)

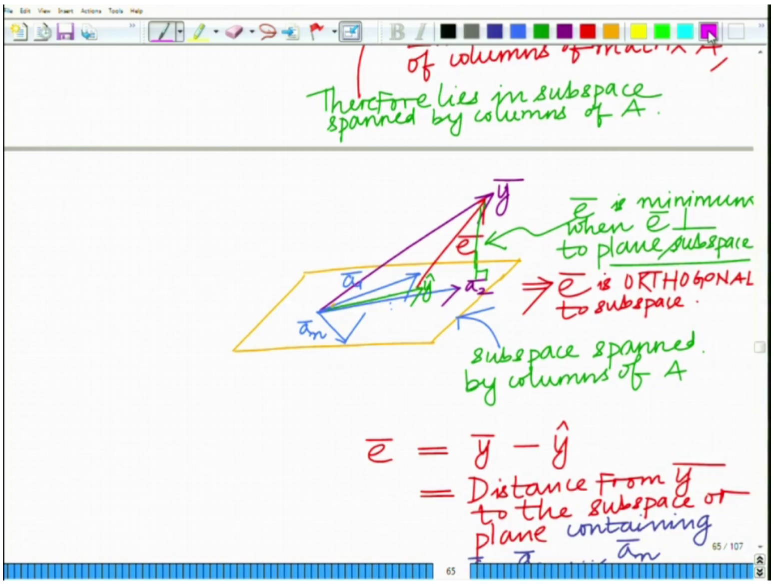


So, which means therefore, this lies in the subspace therefore, this lies in the subspace spanned by columns of A. what you have over here now to represent this pictorially if you take let us say this plane remember a subspace a plane is nothing, but a subspace. So, let us say you have the subspace which is spanned just to give you an idea I am presenting the subspace by a plane.

So, let us say this is the subspace spanned by columns of A. Let us say this subspace by columns of A and you have your vector y bar which does not necessarily lie in the subspace. And you are trying to form an approximation which lies in this subspace. So, this is your approximation let us denote this approximation this approximation by y hat.

Ah this approximation let us say this approximation is let us say you denote this by y hat and this approximation and let us say you are let me just. So, this is your a 2 bar and this approximation is let us say you denote this by y hat. Now, this y minus y hat this is a if you look at this, this is the corresponding error.

(Refer Slide Time: 12:51)

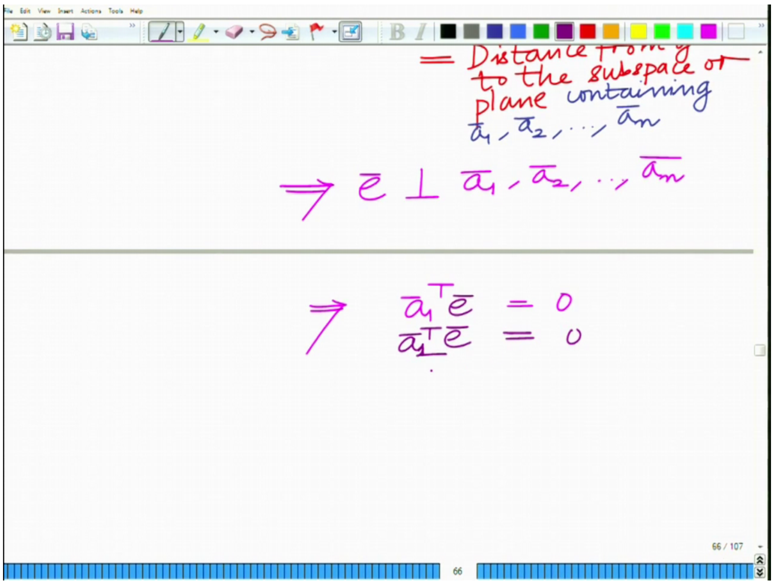


Ok, So, this is your corresponding error vector ok. So, this y minus y hat this is your corresponding error. And therefore, now e bar equals y bar minus y hat and what is this? This is the distance from y bar to the subspace or you can say the plane, the plane that is spanned plane or plane let us make it simple plane containing a 1 bar a 2 bar up to. So, what do you think of this as basically you have a vector y bar and you have this plane that contains y, a 1 bar a 2 bar a n bar already in this plane contain different possible approximations y hat.

Now, what is the error? Error is the distance between this vector y bar and y hat which lies in the plane. And we want to find the error the vector error I mean we want to minimize this error that is we want to make the error vector e bar such that it has the minimum. Or in other words the distance of y bar from this plane has to be minimum.

And now you can see this error is the distance of y bar from this plane is minimum when the error is perpendicular to the plane that is the whole point. So, this error which is nothing, but the distance geometrically you can see this error is minimum when e bar is perpendicular to the. This is the important ideas so this error vector is minimum when it is perpendicular to the subspace that is spanned by a 1 bar a 2 bar up to a n bar. Or we can also say that this error vector is orthogonal to the subspace this is a big idea.

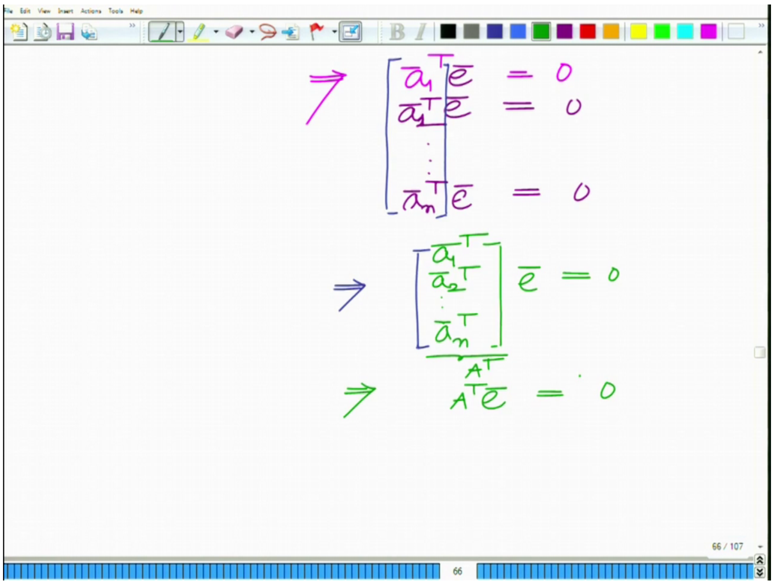
(Refer Slide Time: 15:35)



So, this e bar is orthogonal when is this orthogonal to the subspace this implies that e bar has to be orthogonal to each of the vectors in the subspace this means e bar has to be perpendicular to a 1 bar, a 2 bar, up to a n bar. So, e bar has to be perpendicular to the subspace and.

We know the condition for orthogonality the condition 2 vectors are orthogonal when their inner product is 0. Which means we must have we must have we must have a 1 bar transpose e 1 bar equal to 0, a 1 bar transpose or a 1 bar transpose e bar sorry there is only e bar, e bar equal to 0, a 2 bar transpose e bar equal to 0.

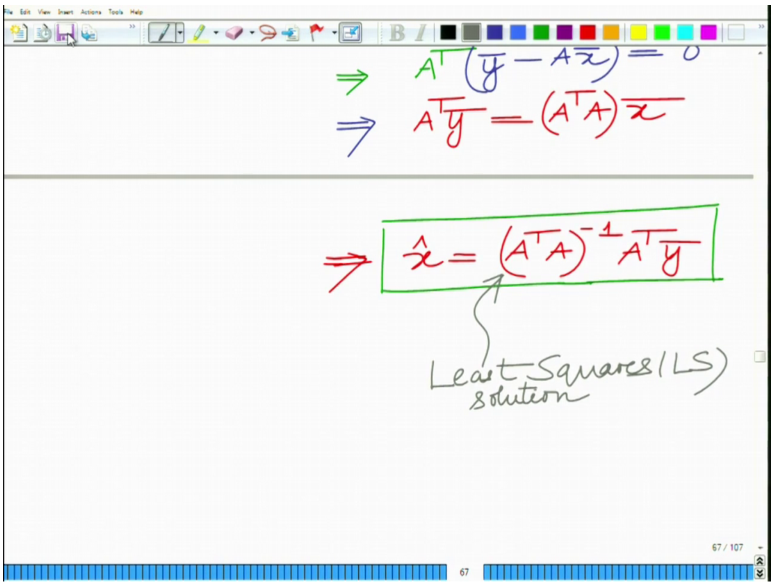
(Refer Slide Time: 16:23)



So, on a n bar transpose e bar equal to 0, that is e bar has to be orthogonal to all these vector a 1 bar, a 2 bar upto a n bar. And now you can write this you can put write this as a matrix. So, this implies basically now you can concatenate these condition as a matrix. So, this implies a 1 bar transpose, a 2 bar transpose so, on upto a n bar transpose into e bar equal to 0.

And this implies well this is nothing, but now you can see this in nothing, but the matrix A transpose. So, this implies A transpose e bar equal to 0.

(Refer Slide Time: 17:13)

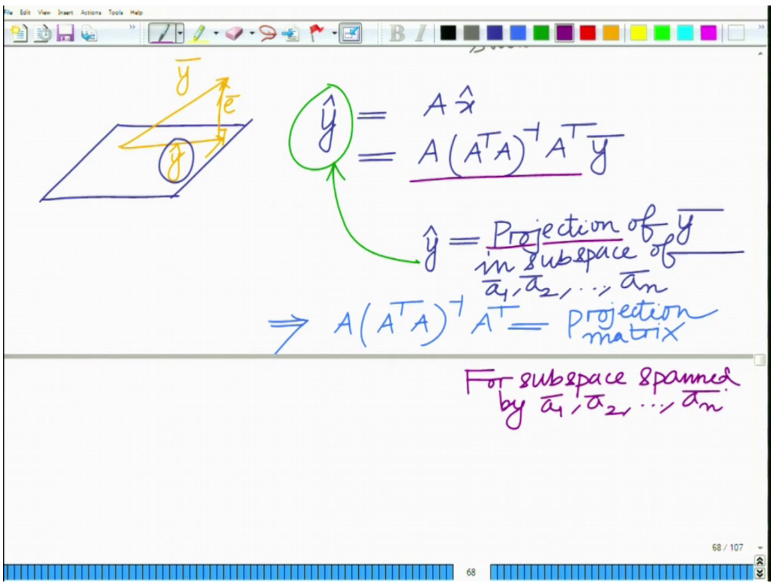


But, e bar is y bar minus A x bar so, this implies A transpose times y bar minus A x bar equal to 0. This implies that now you observe something interesting A transpose y bar equals A transpose A into x bar which is a condition that we have already seen which implies that x bar or the best vector x bar that minimizes the error that is x hat equals A transpose A inverse A transpose into y bar ok. So, this implies that x hat equals A transpose A inverse into A transpose into y bar. So, this implies this is nothing, but again you get the least square solution which is basically exactly.

So, intuitively what the least square solution is doing is basically finding the vector y hat which is the best approximation to y bar in the subspace that is spanned by the vectors a 1 bar a 2 bar up to a n bar which are basically nothing, but the columns of this matrix A. And therefore, what now what is this so, therefore, that so, therefore, which implies the error vector e bar which is the distance of y bar to y hat or basically distance of y bar to the plane is minimum when the error vector is perpendicular to the plane which implies that the error vector has to be perpendicular to all the vector a 1 bar, a 2 bar, a n bar which spanned the plane.

Now, in addition what is y hat? Now, we can see y hat that is.

(Refer Slide Time: 19:15)

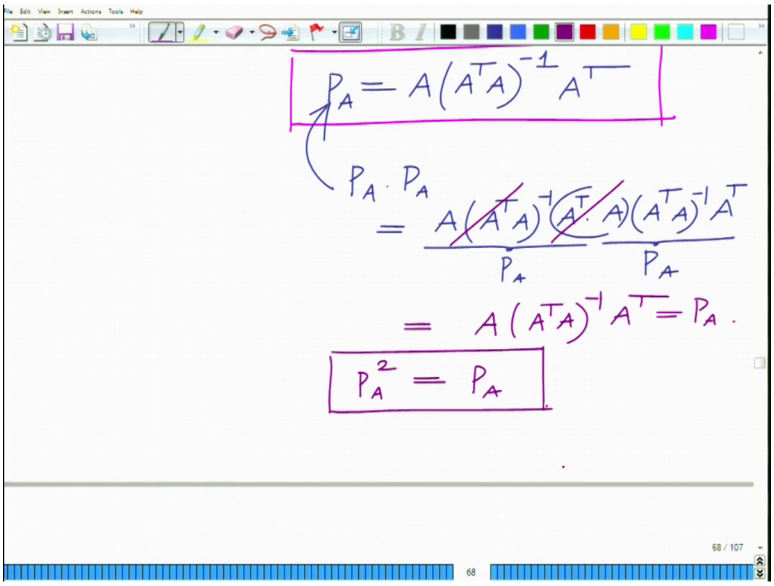


Approximation to y, y hat is nothing, but a times x hat which is equal to A times A transpose A inverse A transpose into y. Now, what is y bar? Now, what is y hat? Remember, y hat is the best approximation. Now, if you look at this plane again, go back and look at this plane again this is your vector y bar, this is your vector y hat and the error vector is orthogonal. And the resulting error vector is orthogonal ok.

And therefore, what is y hat now y hat you can see is the best approximation to y bar in the plane or you can also say y hat is the projection of y bar in the plane or subspace containing. So, y hat equals projection in the best approximation, in the subspace of y. y hat is a projection of y bar in the subspace or spanned by a 1 bar a 2 bar a n bar this matrix which is giving you the projections. So, when you multiply this matrix by y bar you get the projection. So, this implies that this matrix is the projection matrix.

This implies that A into A transpose A inverse A transpose is the projection matrix and the projection matrix for what? Projection matrix for the subspace that is spanned by a 1 bar, a 2 bar, a n bar that is it.

(Refer Slide Time: 21:40)



So, this matrix which is very interesting, this matrix which is the projection matrix that is spanned this is basically given as A equals A transpose A inverse into A transpose. This is the projection matrix corresponding to the subspace that is spanned by the columns of the matrix A ok.

So, this is basically your projection matrix and this you can see is very interesting properties. One of the most interesting properties of the projection matrix is that if you look P A times P A that gives you A A transpose A inverse A transpose times multiplied by this A A transpose A inverse into A transpose.

So, this is A multiplying it by P A and you can see now you have A transpose A transpose A inverse. So, these things cancel and what you are left with is again you can see A A transpose A inverse into A transpose which is P A. So, this satisfies the property P A square equals P A.

In fact, P A square equals P A, P A raise to the power of n correct any integer n right any integer n greater than or equal to 1 is P A. So, this is the projection matrix and this is basically the intuitive or the intuition behind the least square solution which sheds which basically very conducive to sort of intuitively understanding the reasoning and the methodology behind the least square solution all right. So, we will stop here and continue in the subsequent modulus.

Thank you very much.