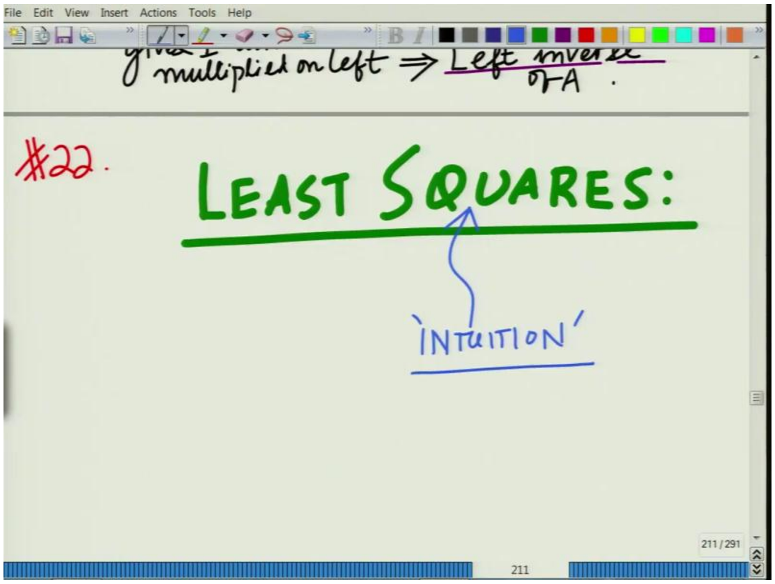
Applied Linear Algebra for Signal Processing, Data Analytics and Machine Learning
Professor Aditya K Jagannatham
Department of Electrical Engineering
Indian Institute of Technology Kanpur
Lecture 22
Matrix: Least Squares (LS) via principle of orthogonality, projection matrix, properties

Hello, welcome to another module in this massive open online course. In the previous module we have looked at the novel least squares algorithm which is a very interesting one and has several applications as we have discussed. Let us now try to get a deeper sense of the least squares technique. Let us try to develop an intuition regarding this thing, what is happening in the least squares.

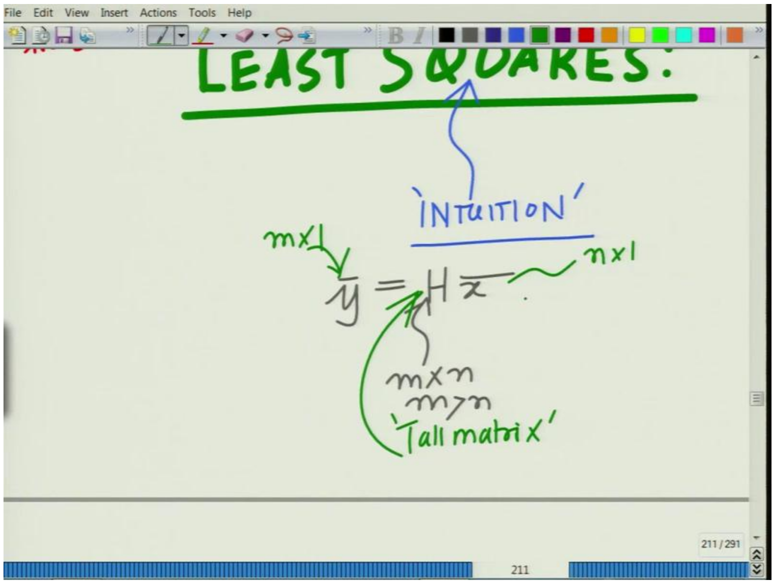
(Refer Slide Time: 00:42)



So what we want to do is we want to take a deeper look at the least squares solution, develop a deeper understanding, get an intuition regarding what is happening in the least squares. So what is the least squares all about. Why does the solution have the structure that it actually has?

One way to look at it is simply to look at it in a mathematical sense that is to write the equations and formally derive it and the other is to understand it at a much deeper level, develop an intuition and try to understand why it is, essentially how it works and why it has that particular structure. So that helps us develop newer applications and that helps us get a better sense of this algorithm.

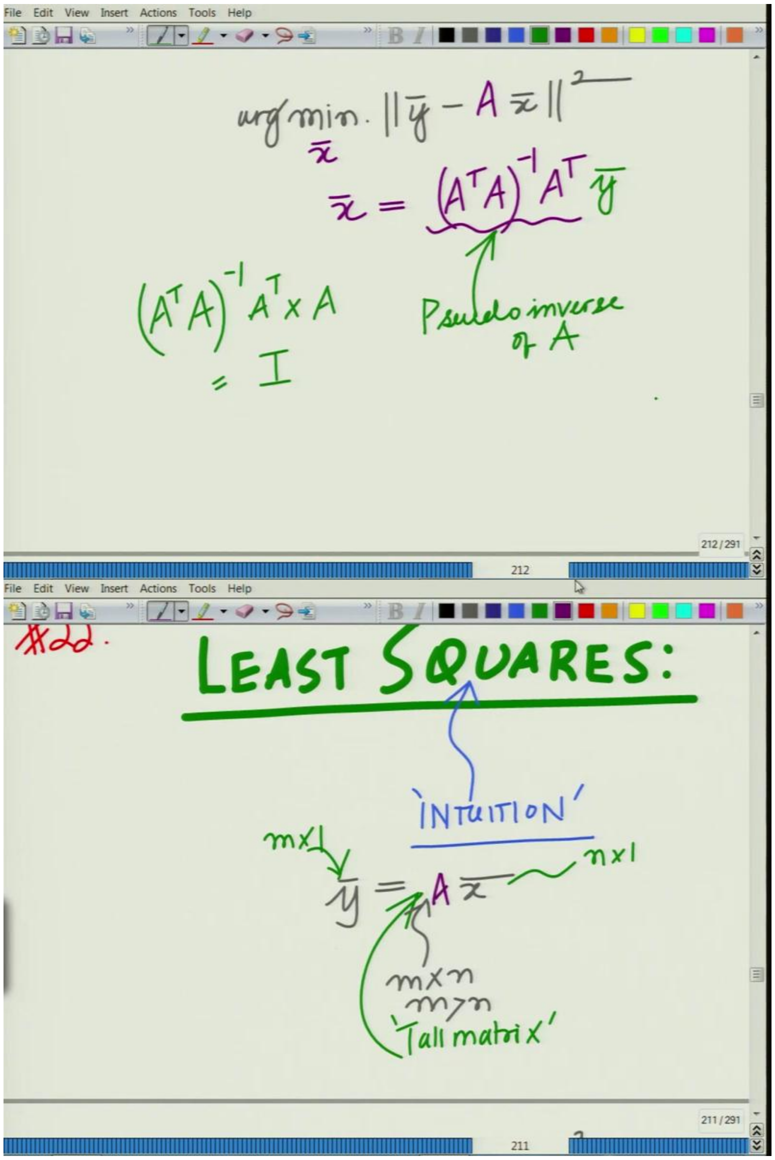
(Refer Slide Time: 01:42)



As we have seen, let me just briefly describe it, what is the least squares all about. Least squares is when you have this problem, system of linear equations where H is essentially an m cross n matrix, typically m is greater than n and we said that is basically it looks like a tall matrix. The colloquial way of referring to such a matrix is a tall matrix where the number of rows is greater than number of columns. X bar is naturally an n cross 1 vector and y bar is an m cross 1 vector.

y bar lies in an m dimensional space, x bar lies in an n dimensional space and the columns of H we said span in n dimensional subspace. And of course we cannot solve this exactly. We can solve this only when y bar lies in the n dimensional subspace spanned by the columns of H, otherwise this has no solution. We therefore minimize the error, in fact the norm of the error or the norm square of the error, y bar minus Hx bar and we have seen that is essentially what the least squares is.

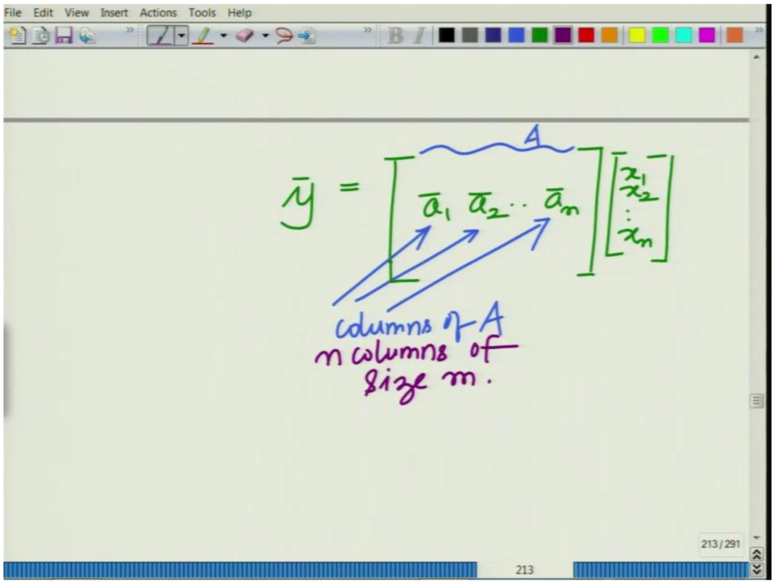
(Refer Slide Time: 02:51)



So you minimize norm y bar minus Hx bar square. In fact, we are interested in argmin that is what is the x bar that minimizes y bar minus Hx bar square and we have seen the x bar which is a solution of this. This is given as, if you look at it, in fact let me once again, I apologize for this. Let me once again make this as A and in fact this is given as A transpose A inverse A transpose x bar.

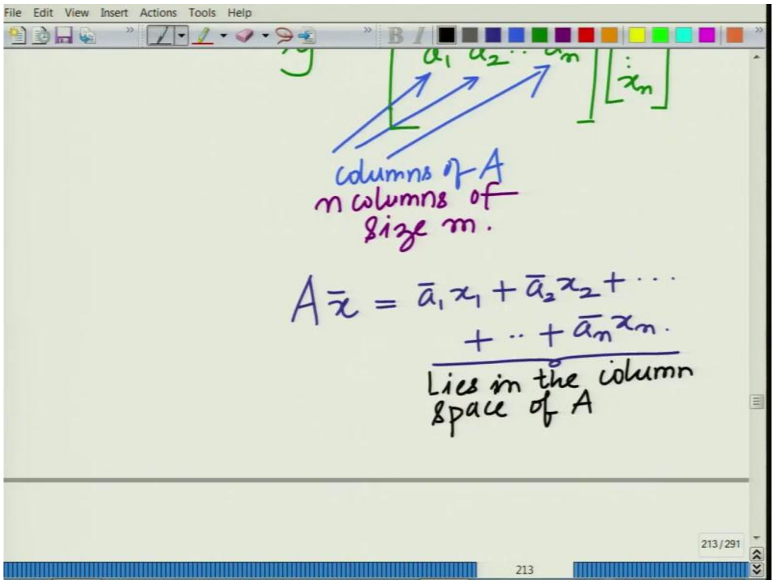
And this we said this has an interesting name, this is the pseudo-inverse of A. This is the pseudo-inverse of the matrix A meaning if you multiply this times A on the left, times A transpose into A, this is naturally equal to the identity, sorry this has to be your y bar. So this is basically your least squares solution. So, this is A transpose A inverse A transpose times y bar. And now let us try to understand why this has this particular structure. So, let us go back take a look at this thing.

(Refer Slide Time: 04:50)



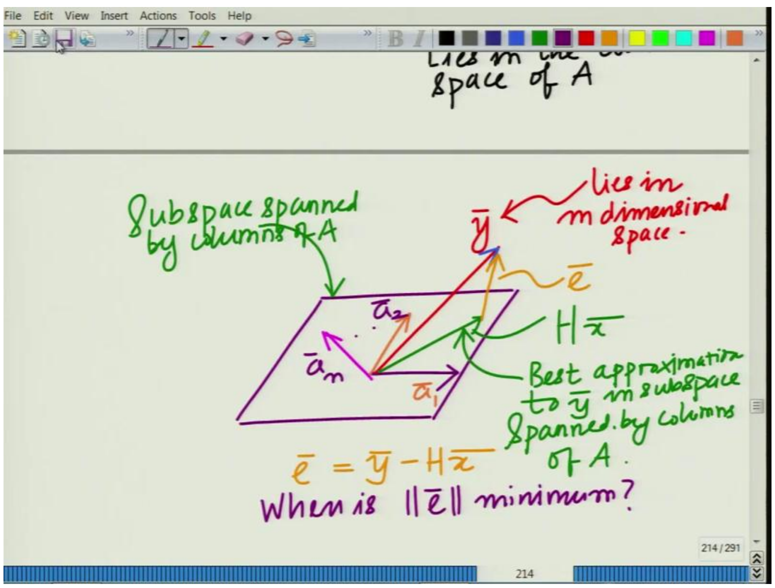
So we have y bar. Let us write it in a slightly elaborate fashion. We have y bar equals a1 bar, a2 bar, an bar and these are the columns. Now if you look at this these are the columns of, this is your matrix A which is an m cross n matrix. These are the columns of your matrix A. There are n columns, each of size m. So you have n columns of size m. So essentially they span an n dimensional subspace of the in general m dimensional space. Each of these vectors lies in an m dimensional space and they span, together they span in n dimensional subspace.

(Refer Slide Time: 06:11)



So let us denote this. Now the thing that you have to observe here is that if you look at Ax bar that is equal to a1 bar x1 plus a2 bar x2 plus so on, plus an bar xn, this lies in the n dimensional subspace and in fact this lies in the column space of A. Every such ax bar lies in the column space of the matrix A.

(Refer Slide Time: 07:24)



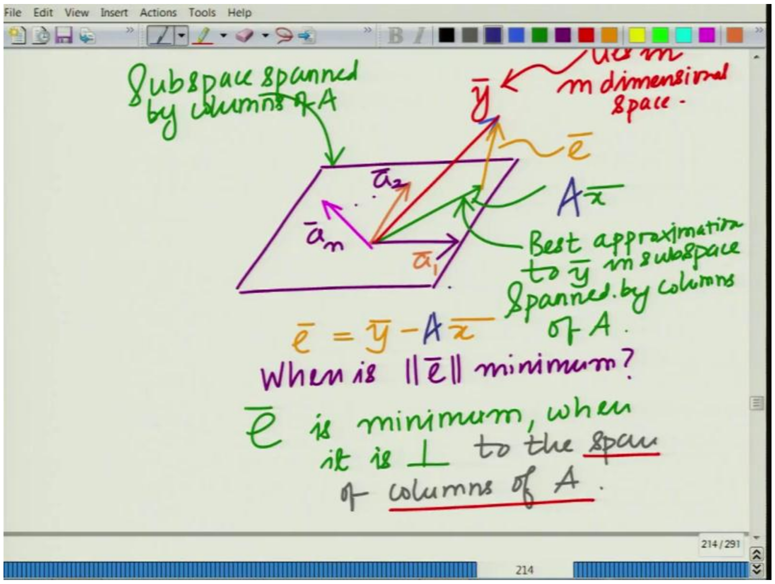
And now let us look at a figure that represents this. Let us say this sort of figure represents the column space of A that is it contains all the linear combinations. This is your a1 bar, a2 bar, so on, up to an bar. Now this is the subspace spanned by the columns of A. And this vector y bar, if you look at this vector y bar, this is some vector which lies in the general m dimensional space.

Now we are trying to find the best approximation to this y bar which lies, this the best approximation in subspace spanned by the columns this is the best approximation, so, this is essentially your Hx bar which is the best approximation to y bar and the difference between these is what we are calling as the error. The difference between these is essentially what we are calling as the error.

So, y bar, so this is your Hx bar, the error is y bar minus Hx bar and we are asking the question when is norm e bar, that is the error, when is the error minimum? And the answer to that is it is not very difficult to see if we have the vectors in a plane and then if you have a vector that lies in a plane and then you have, think of this as in a three-dimensional scenario.

If we have a two-dimensional subspace that is a plane and then you have a general vector in a three-dimensional space and you ask the question, what is the vector in the plane that is the subspace which is the closest to this vector in the three-dimensional space. Naturally it is not every difficult to see that the closest vector to that will be when this error vector is perpendicular to the plane. So, e bar is minimum when this error is perpendicular to the subspace that is spanned by the columns of the matrix A. That is an interesting thing.

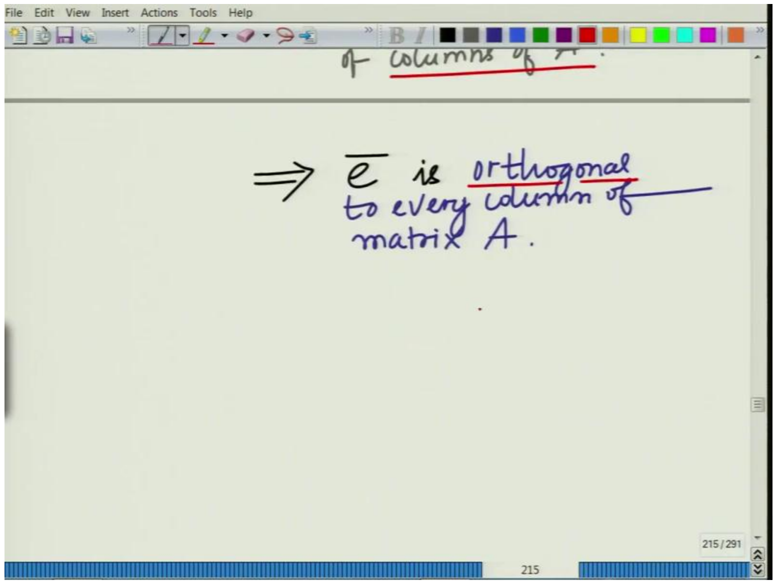
(Refer Slide Time: 11:39)

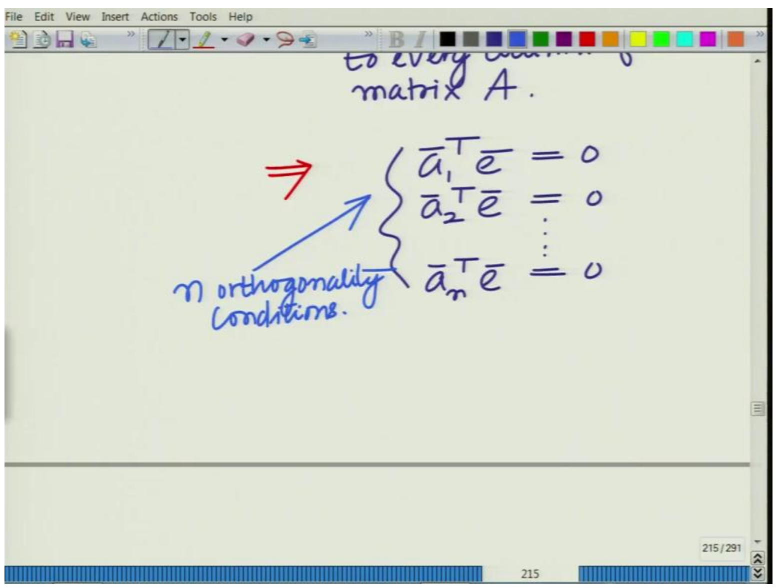


So, this e bar is minimum, the intuition here is e bar is minimum when it is perpendicular to the span of columns of A. So that is the interesting aspect. The error is minimum when it is perpendicular to the subspace, to this n dimensional subspace that is spanned by the columns of A.

And remember, the subspace that is spanned by the columns of A, the columns of A are the basis to that subspace which is spanned by the columns of A. Naturally e bar is perpendicular to the subspace if and only if it is perpendicular to every vector, every basis vector of that subspace which means e bar is perpendicular to every column of the matrix A.

(Refer Slide Time: 13:00)

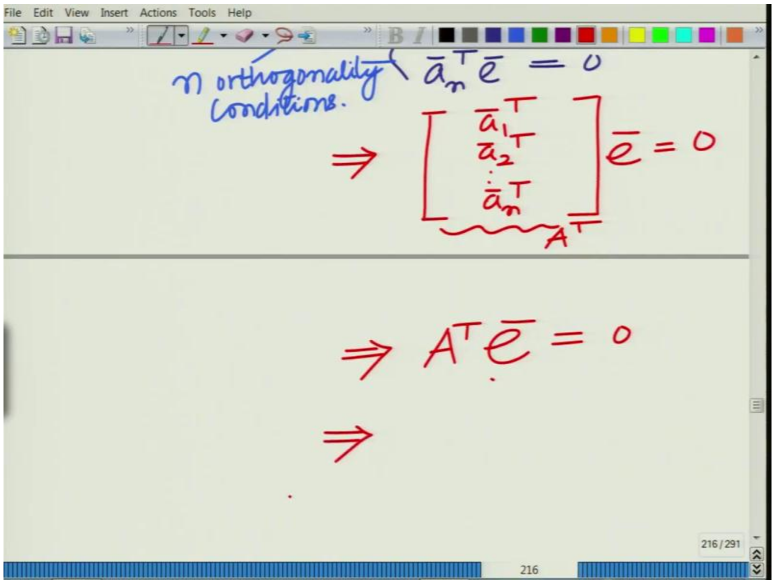




This essentially implies e bar is perpendicular, I hope you understand the symbol, right, e bar is perpendicular or essentially you can say e bar is now orthogonal because we have seen the concept of perpendicularity nothing but orthogonality in the context of vector, e bar is orthogonal, e bar is perpendicular to every column of A, e bar is orthogonal to every column of A which implies, now you write it down the columns of A, the condition for orthogonality will be a1 bar transpose e bar equal to 0, a2 bar transpose e bar equal to 0 so on and so forth, an bar transpose e bar equal to 0.

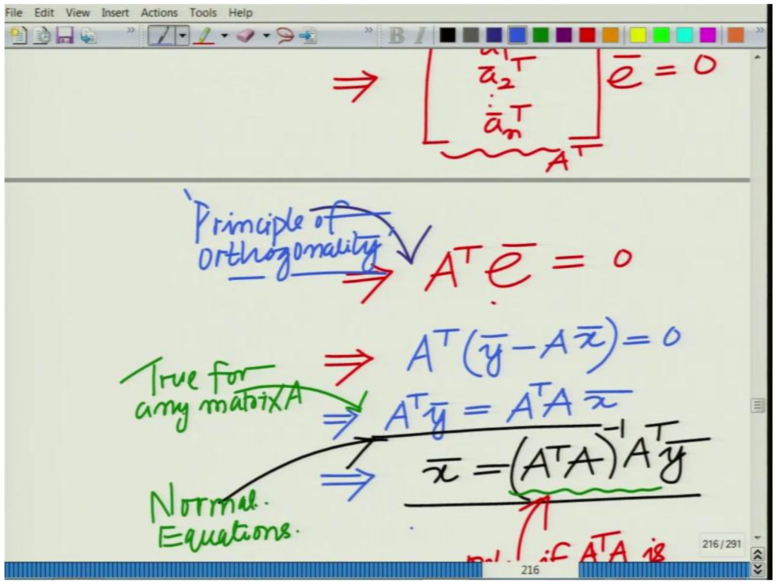
So you have the n orthogonality conditions that is essentially orthogonal to a1 bar, orthogonal to a2 bar, orthogonal to n bar, that is it is orthogonal, the error vector e bar is orthogonal to each column of the matrix A.

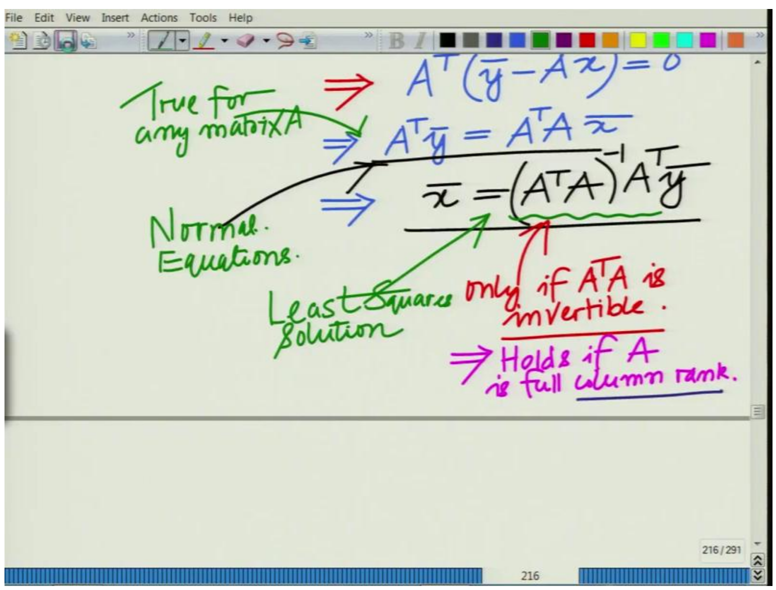
(Refer Slide Time: 14:56)



Now if we consolidate these and write it in the form of a matrix, what you will get is, this implies a1 bar transpose, a2 bar transpose, so on, an bar transpose e bar equal to 0. This is nothing but, you can see this is a transpose. This implies A transpose times e bar equal to 0.

(Refer Slide Time: 15:27)





And now substitute for e bar. We know what is e bar. e bar is nothing but y bar minus ax bar. So interestingly this gives y bar minus ax bar equal to 0. This implies A transpose y bar equals A transpose Ax bar. This implies x bar equal to A transpose A. In fact, I can now make this Hermitian without, I can now also talk about a general complex matrix. It is okay, we can leave it. A transpose A inverse A transpose y bar. So essentially this is the expression that we have and these equations, these are known as the normal equations.

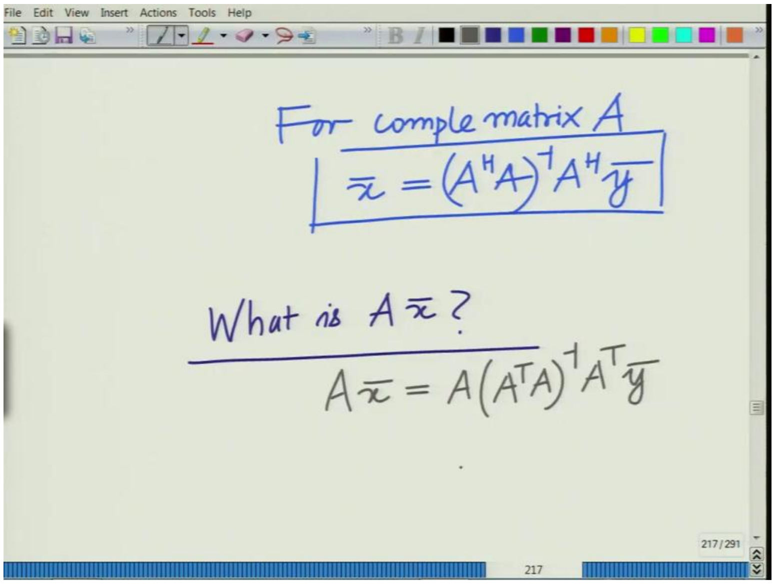
These equations, these are known as the normal equations. These are true for any matrix A. Now this holds as we said there is a caveat, only if A transpose A is invertible. You can write it in this form, only if inverse A transpose A exists, that is only if A transpose A, naturally, only if A transpose A is invertible which essentially implies that we said, we will not show this explicitly if, holds if A is full column rank. This A is holds that is the rank of A equal to n where is the number of columns.

And this essentially is termed as the principle of orthogonality. What we have written over here that is A transpose E bar equal to 0. This is essentially very important in fact signal processing. This is essentially termed the principle of orthogonality that is the error is minimum when it is perpendicular to the span of the columns of the matrix A.

If you use that condition that essentially gives A transpose e bar equal to 0, substitute for e bar equal to y bar minus ax bar, simplify it, get exactly our least squares solution without doing any of the rather complicated mathematical derivation, the analysis. This is very intuitive.

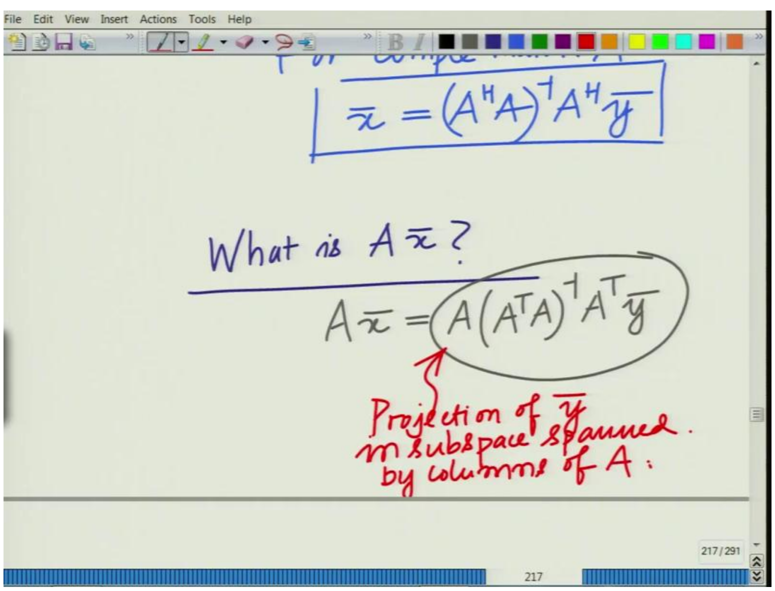
This gives us a much deeper understanding of least squares. All it is saying is whatever is this error, it is minimum only when it is perpendicular to the span of the columns of A which implies that it is perpendicular to each of the columns of the matrix A. And this is basically our least squares solution, not very difficult to see. This is our least squares solution.

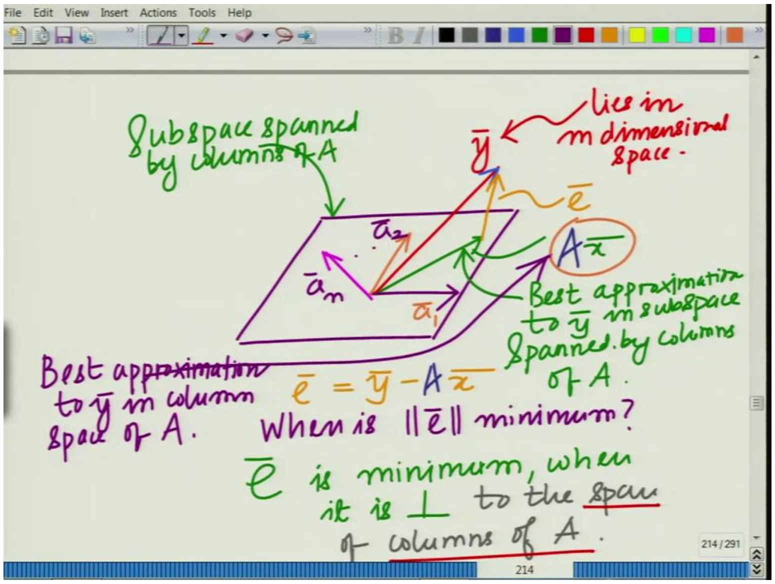
(Refer Slide Time: 19:56)



And for a complex matrix it is not very difficult to see. You have replaced the transpose by Hermitian. You have a Hermitian A inverse. In fact, complex matrices or what occur more frequently in communication, wireless communication is an example because the channel is essentially a complex quantity. When we look at the channel in the (()) (20:31) so the matrices you can see naturally also be complex.

(Refer Slide Time: 21:14)

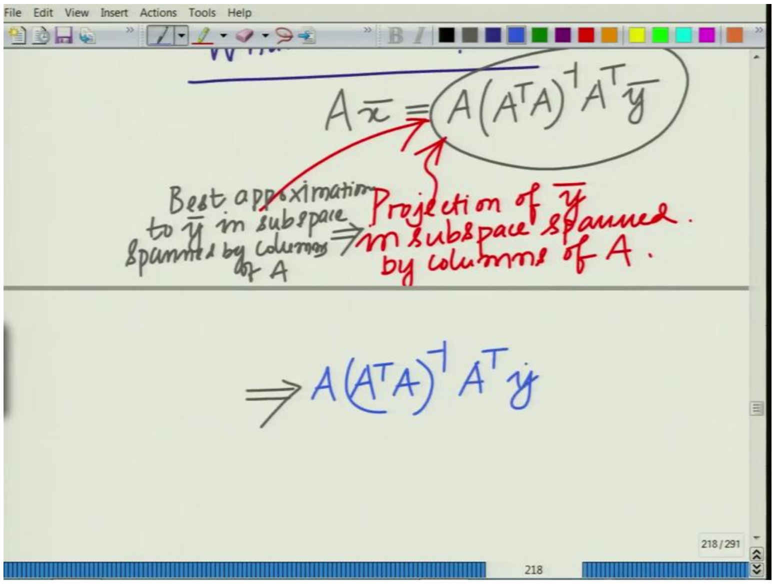


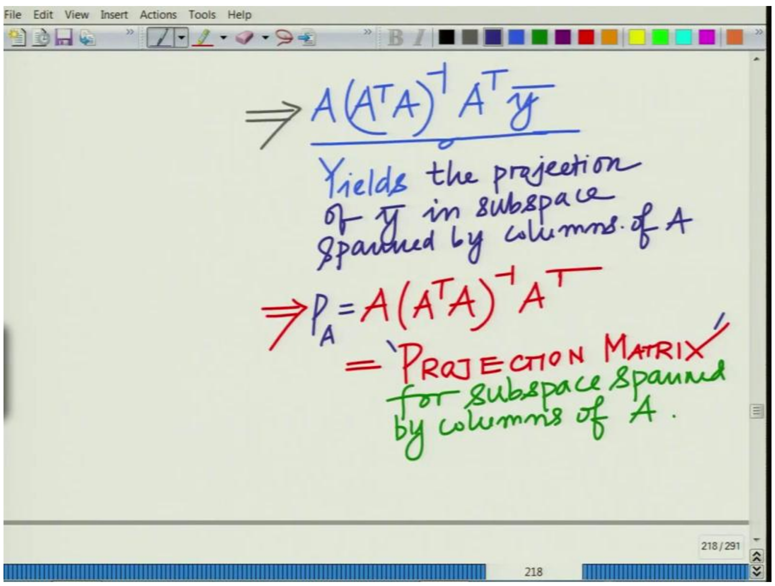


The other interesting aspect of this is further observe, sorry this has to be ax bar, this has to be ax bar. Now further observe that we ask the question what is ax bar? Ax bar is which is equal to A times A transpose A inverse, A transpose y bar for any vector y bar. What is this quantity? Remember if you go back and take a look at this, what is this quantity. You take a look at this, what is this quantity. This is the best approximation to y bar. If you ask this question, what is this quantity, this is the best approximation.

This is the best approximation to y bar that lies in the subspace spanned by the columns of A and what is the best approximation to any vector in a particular subspace, that is nothing but the projection of that vector in that subspace. That is why the error vector is perpendicular. So A x bar which is A, A transpose A inverse, A transpose y bar is the projection of y bar in the subspace that is spanned by the columns of A. That is the interesting aspect. So this here, this is the projection of y bar in the subspace spanned by columns of A. This is the projection of y bar in the subspace spanned by the columns of A.

(Refer Slide Time: 23:51)

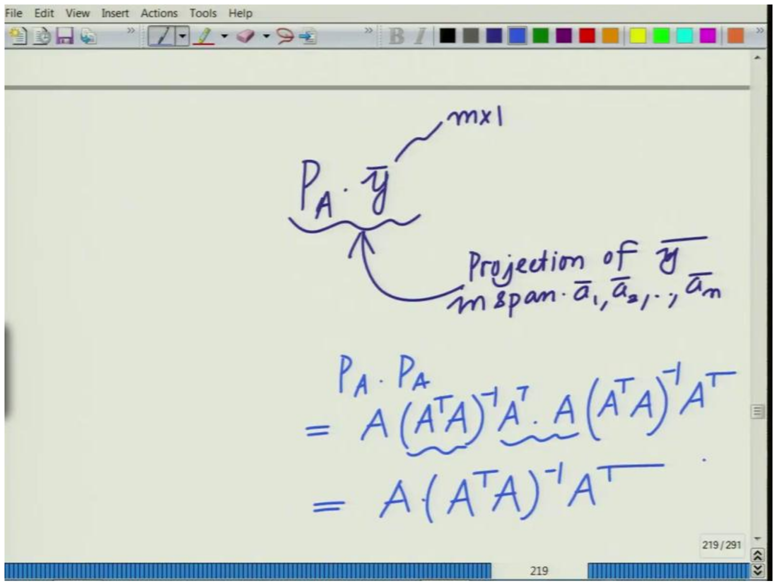




So A x bar is the best approximation, this is the best approximation to y bar in subspace spanned by the columns of the matrix A which implies that it is the projection of y bar in the subspace spanned by the columns of the A. This implies A times A transpose A inverse, A transpose into y bar. What is this? This yields the projection of y bar in subspace spanned by the columns of A which implies this matrix A A transpose A inverse A transpose, this is equal to the projection matrix for the subspace spanned by the columns of A.

This is the projection matrix for the subspace spanned by the columns of A. In fact, this is a very interesting matrix. Let us call this as P of A equal to A A transpose A inverse A transpose, this is the projection matrix spanned by the columns of A and this has very interesting properties.

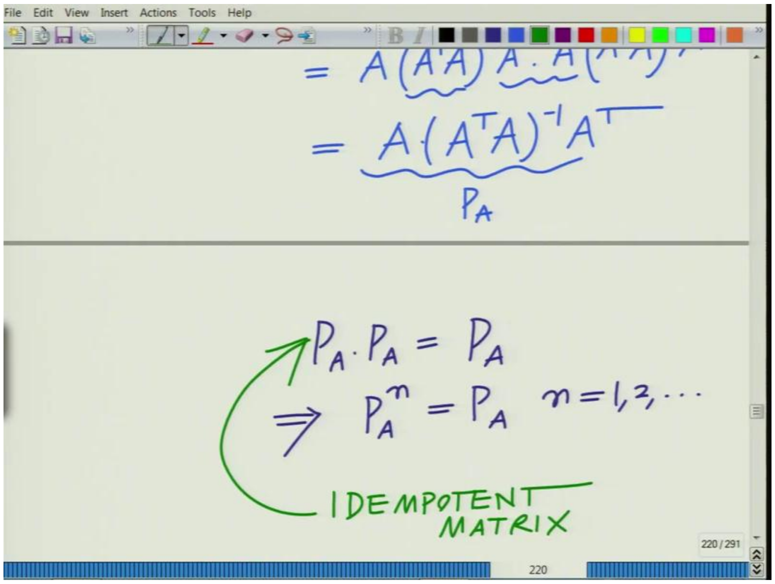
(Refer Slide Time: 26:25)



So if you multiply, what do we mean by the projection matrix for the column space of A which means if you consider P bar into any vector y bar y bar that is if you multiply this any vector y bar which is of size as we have seen m cross 1, this yields projection of y bar in the span a1 bar, a2 bar, so on up to an bar that is the columns of A.

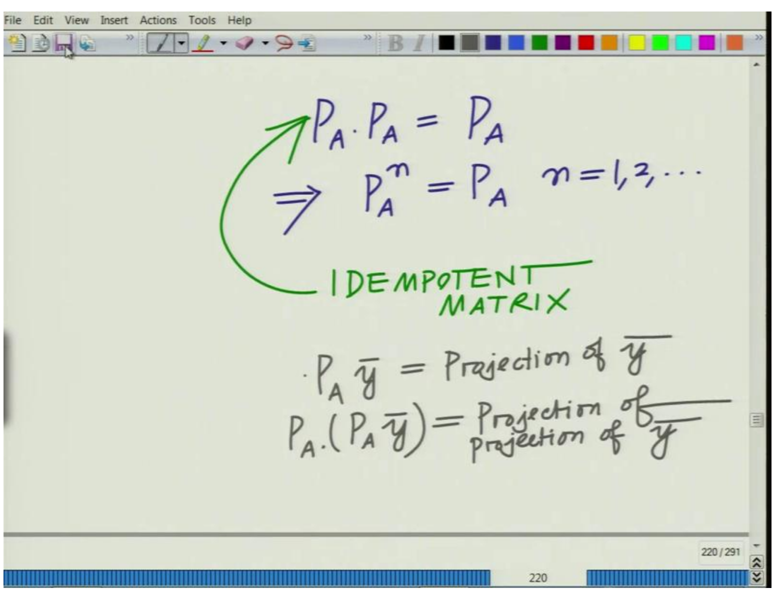
In fact, you will notice a very important property that is if you consider this PA into PA, this is equal to A times A transpose A inverse A transpose times A into A transpose A inverse into A transpose and now you see A transpose A inverse into A transpose A, this is identity so this reduces to A A transpose A inverse into A transpose.

(Refer Slide Time: 28:09)



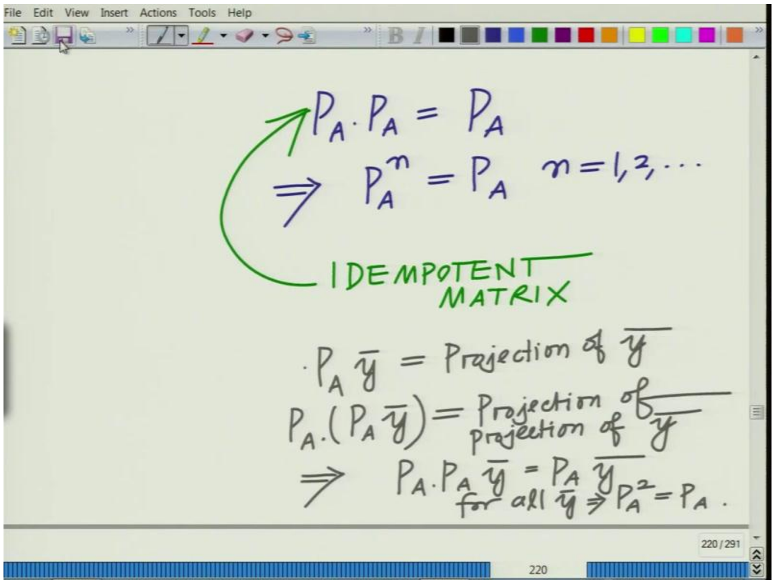
And if you can see this is nothing but P of A and therefore this satisfies the property P of A into P of A equal to P of A. In fact this also implies P of A raise to the power of n equals P of A for any n equal to 1, 2 and so on and so forth and this is known as the idempotent matrix. This is such a matrix which satisfies this property is basically idempotent. This is known as an idempotent matrix P raise to PA into P, PA into PA equal to PA.

(Refer Slide Time: 29:05)



And that is naturally not difficult to see because PA into y bar equal to projection of y bar. Now you take this PA and multiply this to once again PA into y bar, this is the projection of y bar. But once you have projected y bar, PA into y bar, once you have projected y bar into the subspace, any further projection will leave it unchanged because it is already in the subspace which means if you multiply PA into y bar and you take that and multiple it again by PA, the vector should remain unchanged because it already lies in the subspace that is spanned by the columns of A.

(Refer Slide Time: 30:13)



It therefore implies that PA into PA into y bar equal to PA into y bar for all y bar which naturally implies PA square equal to PA. This naturally implies that PA equal to PA, PA square. This is again an intuitive justification for this idempotent property, idempotent property of this matrix A which is the projection matrix corresponding to the column space of A.

So these are I think very, very interesting properties, interesting facts you can call them about the least squares solution. So we have looked at this least squares solution. I have already told you it is something is very, very important, vast applications in machine learning, signal processing, wireless communication, so on and so forth and in fact all fields and some fields of science and engineering, probably across the board in humanity, social sciences and so on and so forth.

In this module what we have looked at is we have looked at an intuitive, the intuition behind the least squares to deeper understanding, to develop a deeper understanding of it, derive it without using any deep rigorous mathematical analysis but simply intuitive principles. What is the principle that we have used which is it is basically the principle of orthogonality that is the error vector, the approximation error to be minimum, has to be perpendicular to the subspace that is spanned by the columns of A.

That has also given us the least squares solution but in a much more intuitive fashion and then we have developed this notion of the projection matrix corresponding to the column spaces. So these are very interesting aspects of the least squares so please go through these modules again and try to understand them thoroughly. Thank you, thank you very much.