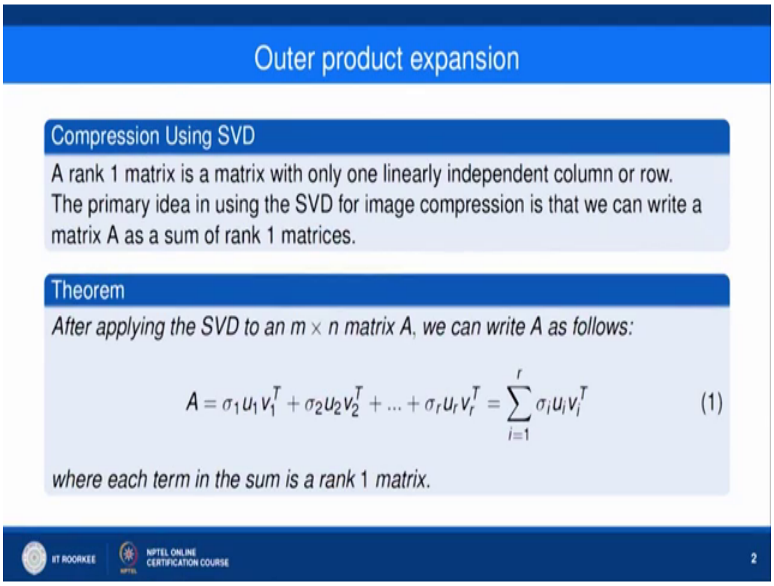
Numerical Linear Algebra
Dr. D. N. Pandey
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 49
Least square solutions- I

Hello friends, welcome to this lecture in this lecture. We will continue our study of singular value decomposition. If you recall the last result in previous lecture is the outer product expansion of a matrix A with the help of single value decomposition. In that outer product expansion of a we generally write the matrix A as the sum of rank 1 matrices. So, we will continue our study on remark based on that outer product expansion and you see that how it is important in applications.

So, we again define we recall that a rank 1 matrix is a matrix with only one linearly independent column or row, and the primary idea in using the SVD for image compression is that we can write a matrix A as a sum of rank 1 matrices.

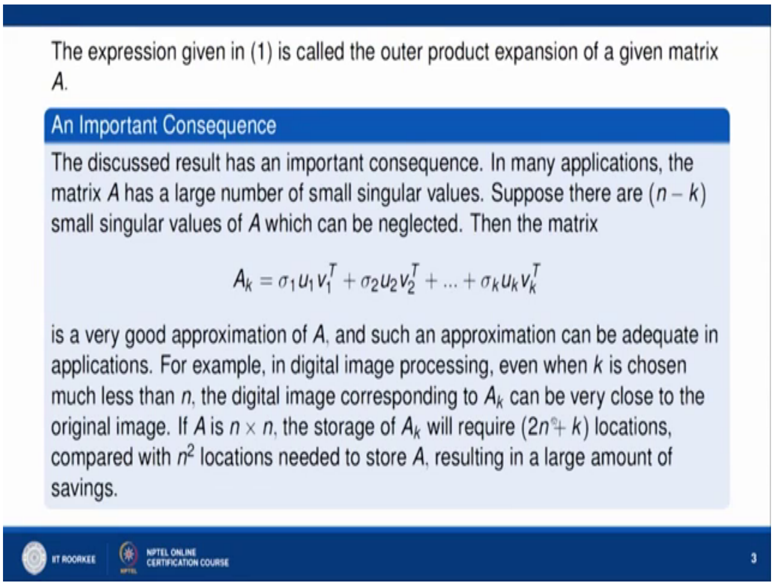
(Refer Slide Time: 01:02)



And with your theorem which we have proved in previous lecture is that after applying the SVD to an m cross n matrix A we can write A as follows. A as summation i equal to 1 to r sigma u i v i transpose where r is the rank of A and u i's are u i is are columns of the matrix orthogonal matrix U and v i's are columns of orthogonal matrix V and u i's are known as left singular vectors and v i's are known as right singular vectors.

And here if you look at the each sum in this summation sigma u i v i transpose this has a rank 1 matrix of size m cross n this we have seen in previous lecture and each term is known as modes. So, we can write that a can be written as say sum of modes here.

(Refer Slide Time: 02:12)



And the expression given in 1 is called the outer products. So, this is known as outer product expansion of the matrix A.

Now, based on this outer product expansion we you just try to see what is the important consequences of this. So, the discus result has an important consequences in many applications the matrix A has a large number of small singular values. So, it means that if A is very large and if we calculate the say all these singular values of matrix A, it many times it happened that the large number of small a large number of single values are having a small values.

So, let us suppose that there are n minus k small singular values of A which can be neglected. So, out of n your n minus k singular values are very very small compared to the other k singular values and we can say that we can neglect n minus k small singular values and which result the approximation of A by this A k, where A k is given by sigma 1 u 1 v 1 transpose plus sigma 2 u 2 v 2 transpose up to sigma k u k v k transpose.

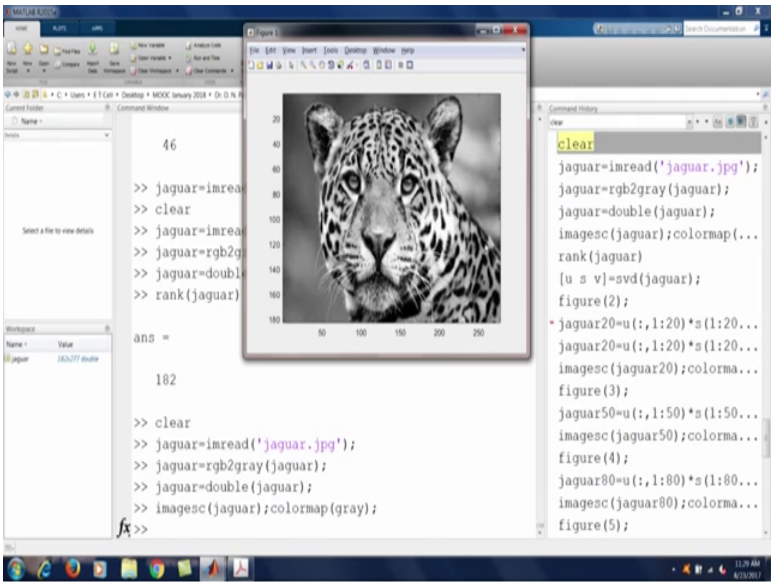
So, this A k if you simplify it is nothing, but US k and V transpose. I hope you remember that we have defined the matrix A k as US k V transpose and we know that the rank of the matrix A k is A and that we have seen that this A k approximate a closely approximate best approximation of the matrix A in the set of the matrices of rank k. So, A k we have already seen and we have seen that it is a very good approximation of A

Now, we say that if n minus k are small then this A k will very will approximate of a and such an approximation can be adequate in application. So, many a times only this much approximation will be helpful in dealing the application part of the problem. So, for example, in digital image processing even when k is chosen much less than n the digital image corresponding to A k can be very close to the original image that we are going to see that and how this is this can be seen.

Now, if A is n cross n matrix the storage of A k will require only 2 n plus k location. So, here we can say that the storage of A k requires 2 n k plus k locations and if we compare with any square then it is quite lesser than n square. So, it means that we will save a large amount of space and time, if we cal if we do calculation with the help of this A k.

Now, let us take one example and see how this approximation of A by A k will help in one example of image processing. For that let us consider one example. So, again we had to use a MATLAB.

(Refer Slide Time: 05:46)

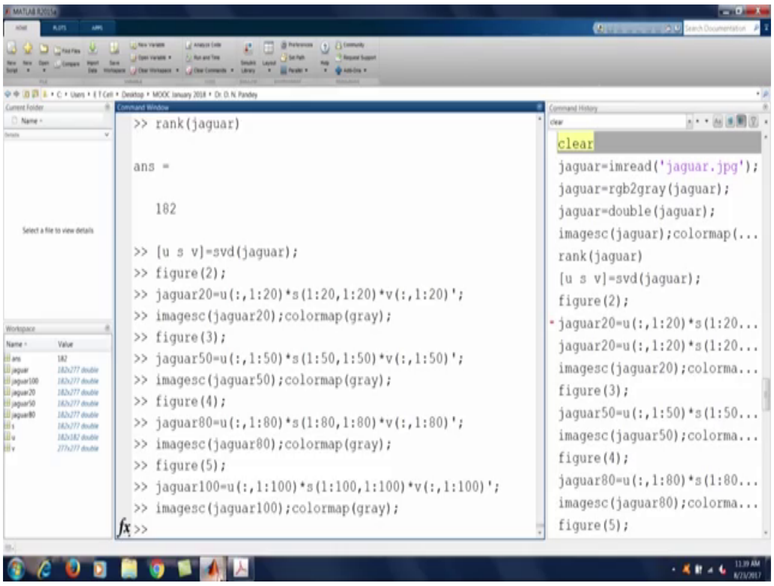


So, we start first by reading the image file. So, let us I am just denoting the file name as Jaguar. So, let us write a Jaguar as image read and the file name file name I am writing jaguar dot jpg I am using the jpg format. So, let me write it like this and then we write the we are only using the grey map.

So, we simply write jaguar as rgb to gray and jaguar and then we double it because we want to work with say SVD. So, jaguar is double jaguar right. And then if you look at what file we are dealing with. So, let us screen it image sc and jaguar and we are using simplest let us say use color map green. So, let me write it here and if you look at this the file original file in grey color and we want to approximate this file with the help of singular value decomposition and outer product expansion of A.

So, we if you look at what is the rank of this of data matrix here jaguar represent the matrix our data saved in a matrix form and it is 182 cross 277, size of the matrix is 182 cross 277.

(Refer Slide Time: 07:40)



So, what is the rank of this? So, let us find out the rank of an jaguar and rank of jaguar is coming out to be 182. So, it means that the here we have singular values we can say that 182 positive singular values we have.

Now, what we want to, we want to approximate with the help of outer product expansion here. So, what we try to do let us find out first of all the singular value decomposition of this. So, you write USV as SVD of the file jaguar. So, here we write like this and then. So, this is once we have find out say singular value decomposition of the matrix A jaguar then we try to find out the approximation. So, we start first by figure 2, we want to see the output in this figure file. So, let us call this figure 2 which is right. Now, it is a blank screen. Now, we want to approximate this jaguar matrix by A lesser singular values.

So, let us start with the jaguar say we can write it 20 we let us start with 20 and here we consider only; so first 20 singular values of this matrix jaguar. So, let me write it u and then we are considering the first 20 columns into S and here we take 1 to 20 rows n comma 1 to 20 columns. And then product and here we and again we are taking only same 1 to 1 to 20 columns of this and transpose and then we suppress it right output we are not considering. So, that a jaguar 20 is the 20 first twentieth approximation of you can write this as that jaguar 20 is a 20 of the matrix jaguar.

So now, we want to see that this jaguar 20 will give how much better or better approximation of A matrix or we can say jaguar matrix. So, let us see what this represent. So, image screen jaguar 20 and again we are using color map gray and if you look at it then we have this file. And if you look at this file is not a very good approximation of your original file. So, let us consider more a number of approximation.

So, rather than considering only the first 1 T singular values let us consider say 50 singular values. So, to find out 50th approximation let us write it jaguar 50 and it is. So, here to find out the 50th approximation you write 1 to 50, 50 and here we have 50 and here we have 50. So, that will give you the a 50 it is an approximation of the rank 50.

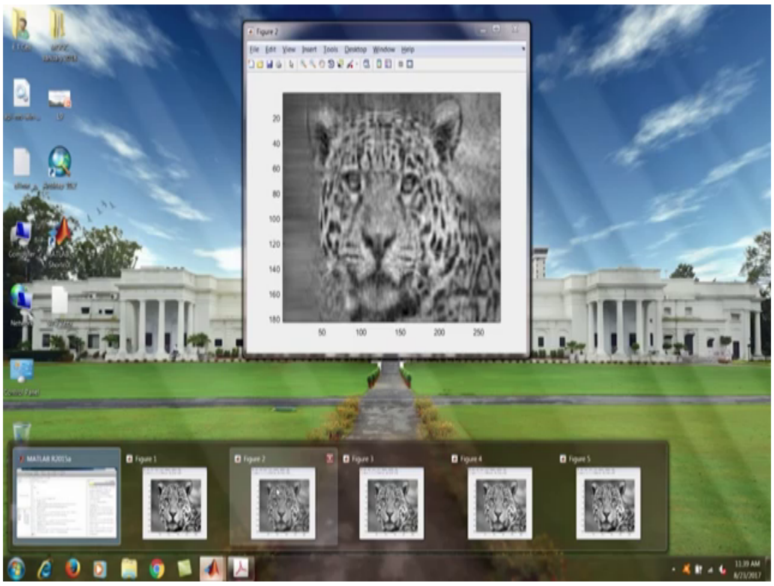
Now, again let us see, but before that we have to 5 figure 3 here and then we write it image 50 here and see what you will get. So, if you look at the figure this represent the approximation of the original figure by using the first 50 singular values and this represent the approximation of the original figure by taking only say 20 singular values.

So, now, let us move to more accurate approximation. So, let us consider figure 3 and figure 4 which represent say more accurate result. So, here this is again a blank screen. Now, let us approximate using same 80 you can say or you can say that 80 let us say start with 80 one. So, you take as 80, so we are approximating our image with the first 80 singular values. So, we are using outer product expansion up to say 80 values and we are neglecting 102 singular values and see how you, so, now, let us screen it.

So, again it is 80 here when we look at the approximation by taking the first 80 singular values and if you look at the image this will give you the original figure. And this is the approximation of the original figure by taking first 20 singular values, and this is the approximation of the original figure by taking the first 50 singular values, and if you look at this is the approximation of the original figure by taking say 80 singular values.

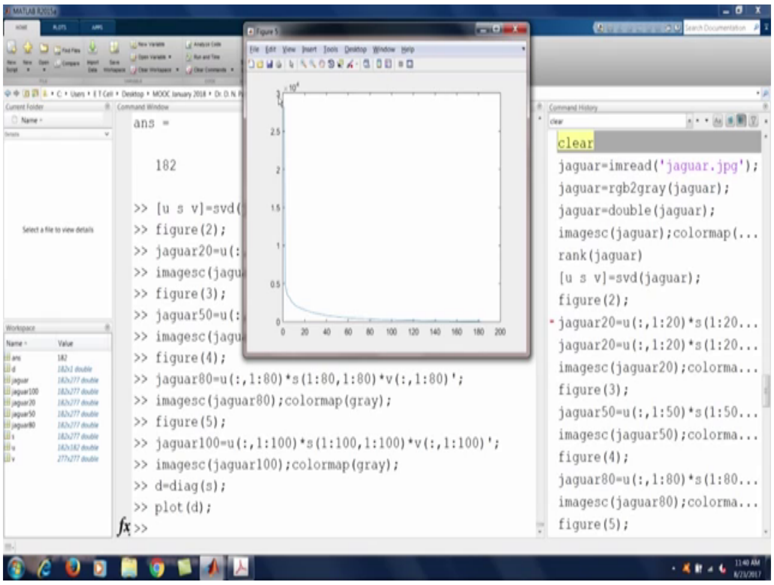
Now, let us take one more approximation of this. So, let us take a figure 5 here and this is a blank screen right. Now, let us consider 100. So, we are taking 100 now, so 1 to 100 and it is 1 to 100 and here it is 1 to 100, and then we screen it. And if you screen it and here you will get, now, again I am reiterating this is the original figure this is the 20th first 20th singular values and then it is 50, 80 and this is your 100.

(Refer Slide Time: 14:33)



Now, if you look at this last figure 100 using the 100 first 100 singular values. And look at the original figure it is almost say approximating the original figure. So, here this figure is the original figure and this is the figure we obtained using the first 100 singular values of the matrix. And we can say that this figure 5 is quite good approximation of the original figure, and here we have used only say first 100 singular values of the matrix. And we have seen that we have 182 singular values for this particular image.

(Refer Slide Time: 15:31)



Now, if you look at the singular values diagram, so d represent the diagonal of S here and that will give you the diagonal entries of S means the singular values on the matrix jaguar

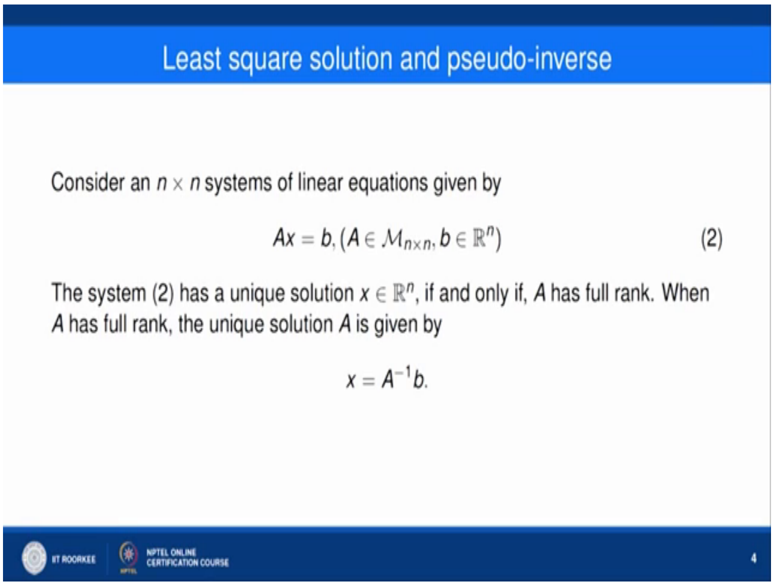
Now, let us plot this d and see what you will get. And if you plot this you will get this figure and it says that if you look at this plot or graph carefully then this represent the number of singular values and here we have 182 singular value, so here we have this 182. And then these represent the modulus of or modulus value of the singular values.

So, if you look at the largest singular value somewhere here that it is between 2.5 to 3 to 3 into 10 to power 4. And if you look at the first 20 the first few your singular values are quite large, but after say 20 it is somewhere between 0.5 to 0 into 10 to power 4, and if we look at and this the modulus value of the singular values is going to be use in decreasing order. And you can say that after say 100 or something it the contribution of the singular values are almost nil, or you can say that the after say 100 and the singular values are quite say negligible compared to all other singular values.

So, you can say that the first say up to 100 singular values carrying most say most data of the file and that we have already seen that if we approximate our original figure that is this, by say 100 you will get a good approximation of the original file. Though 20 is a blurred version, but if you look at the 100 one it is say almost very a close version of this and that is indicated here. That in this matrix jaguar we have say first few are very large and rest of are all neglected compared to the largest singular values.

And we can say that you can take 100 or say 100 and 10 singular values to approximate the original figure. And this is the idea where we can utilize this singular value decomposition and outer product expansion of a matrix to image processing or image a speech processing and so other, so many other applications.

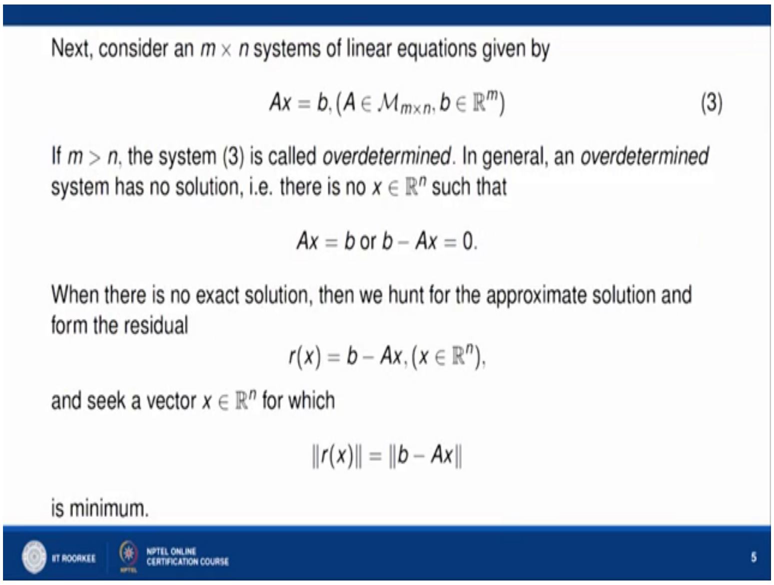
(Refer Slide Time: 18:40)



So, now, we stop this example and move on to next topic that is least square solution and pseudo inverse, that how this singular value decomposition will help us to find out least square solution, but before that let us understand what is this least square solution is. So, if we start with the n cross n system of linear equation which we have we have already discussed in many platforms and here it is A x equal to b where A is a n cross n a square matrix and b is a n cross 1 vector in R n.

And the system 2 this is the system 2 has a unique solution if and only if A has full rank. So, if A has full rank here we have a square matrix, so A has full rank means a is invertible. Then in this case we get a unique solution as x as A inverse b. So, in this case when we have full rank we have a unique solution and it is given by A inverse b.

(Refer Slide Time: 19:43)



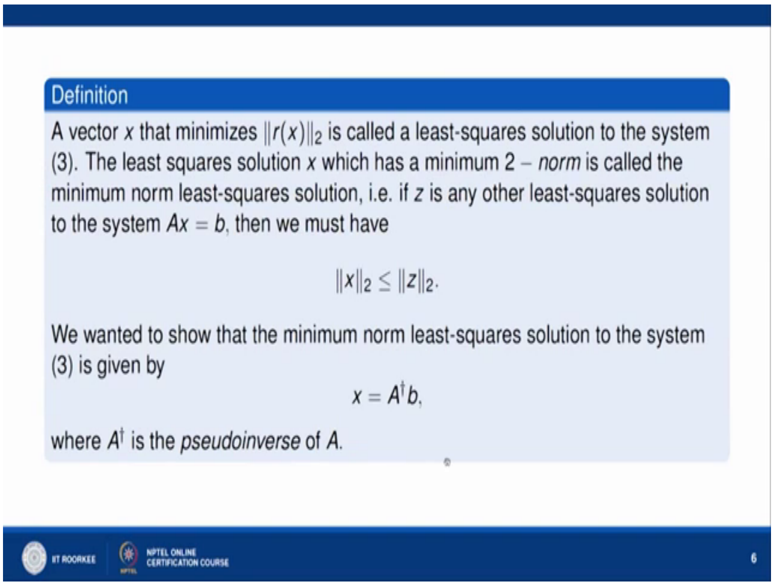
Now, consider the case when A is not a square matrix, but it is a rectangular matrix A m cross n. So, we have a system of linear equation A x equal to b, where A is a m cross n matrix and b is a vector in R n.

Now, if m is a strictly bigger than n then system 3 is called over determined. In general we can say that an over determined system has no solution why because if we look at this is a what we have m equations and in n variables, where b is in R n. So, here we have x is from n cross 1. So, we have unknowns are only n and we have more number of equations. So, more number of equations and we have very less number of unknowns.

So, in general we this system may not have any solution at all. So, it means that there is no x in R n such that A x equal to b or b minus A x is equal to say 0. Similarly in the case when m is less than n then it means that we have less number of equation and we have more number of variables. So, in that particular case we may not have any unique solution rather than we may have infinitely many solutions. So, in case of when m is greater than n we may have the possibility that there is no solution at all, but if m is less than n then there is more likely that this solution the system has a infinitely many number of solution and we do not have any unique solution.

So, in the case when we do not have exact solution we have seen that mm is greater than n or m is less than n there is a possibility that we have no solution or infinitely many solution or you can say that we do not have any exact solution. Then we hunt for the approximate solution and from the residual r x as b minus a x we want to see that how this b minus A x is differ from 0. So, look at the residual, residual we are obtained by writing b minus A x and we want to show that this r x should be a very small. So, we want to find out a vector in R n for which this norm of the quantity r x is minimum, and if we are able to find out a vector like this then we say that it is the solution in the sense of minimum minimum least square solution.

(Refer Slide Time: 22:29)

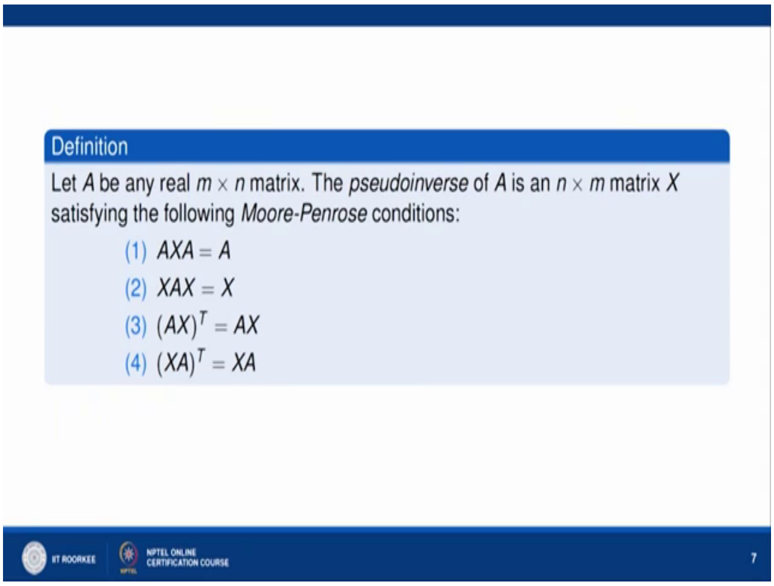


So, here let me define it whether here. So, we define a vector x that minimizes the r x in 2 norm is called a least square solution to the system 3. And the least square solution x which has minimum 2 norm is called the minimum norm least square solution.

So, first thing we define as least square solution. So, least square solution means which minimizes the residual in 2 norm, and among all the solutions which minimizes the least residual r x in 2 norm and having minimum 2 norm is called the minimum norm least square least square solution. That is if we have any other least square solution to the system A x equal to b then we must have 2 norm of x is less than or equal 2 norm of z. In this case we say that x is known as minimum 2 norm least square solutions and we wanted to show that the minimum norm least square solution of the system 3 is given by x equal to A dagger b. Now, what is this A dagger? A dagger is known as the pseudo inverse of A.

Now, how we define A dagger and what are the properties of A dagger that we are going to study. So, our aim of studying this system is that how this singular value decomposition is going to help to find out the minimum norm least square solution of the system A x equal to b and that can be given as x as A dagger b, where A dagger is pseudo inverse. So, we will take the help of singular value decomposition to find out the pseudo inverse of the matrix A.

(Refer Slide Time: 24:26)



So, now, we wanted to know what is this A dagger is. So, we define A dagger as Moore-Penrose inverse of the matrix A,. And we define it like this that let A be any real m cross n matrix the pseudo inverse of A is an n cross m matrix A and it satisfy the following Moore-Penrose condition that is AXA equal to A, and XAX is equal to X and AX is symmetric and it means that AX transpose is same as AX and similarly XA is also symmetric. So, XA transpose is equal to XA.

So, this definition is not a very constructive definition, but you can say that any matrix which satisfy these 4 condition we can call that as Moore, as pseudo inverse of the matrix A. And we will see that once a matrix b satisfy all these 4 properties then that matrix is called as pseudo inverse of A. And you will see that pseudo inverse is going to be unique, so we can call that matrix as the pseudo inverse of the matrix A.

So, I will stop here and in next lecture we will see the properties of pseudo inverse of A and how this is going to be helpful in finding the minimum norm least square solution of the system A x equal to b. So, thank you for listening us.

Thank you.