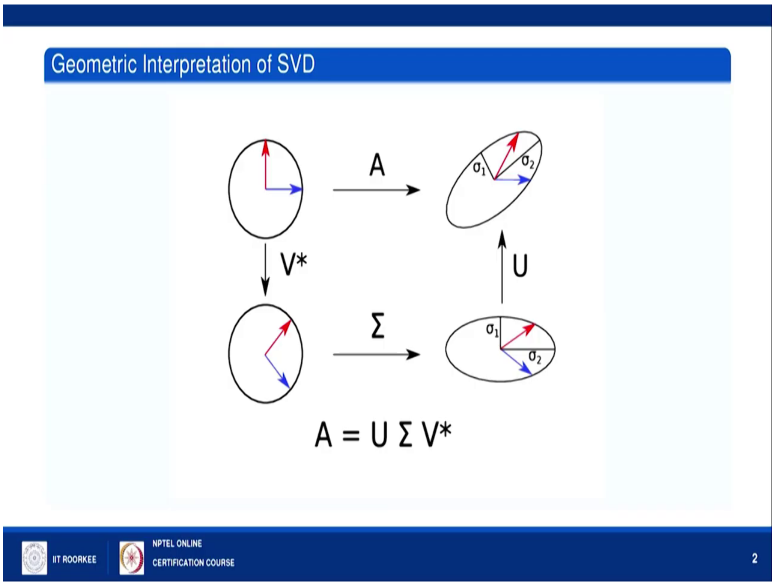
Essential Mathematics for Machine Learning
Prof. Sanjeev Kumar
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 12
SVD: Properties and Applications

Hello friends. So, welcome to lecture 12 of this course Essential Mathematics for Machine Learning. In last lecture, we have learn about singular value decomposition. In this lecture, we will continue singular value decomposition and we will see some of the Properties and Applications of singular value decomposition. In the later part of this lecture, we will define various matrix norm.

(Refer Slide Time: 00:49)



So, first let us understand the geometrical interpretation of singular value decomposition. So, just look here. Suppose, I am having a unit circle and these two vectors; one is blue and another one is red. So, I am having these two vectors. If I apply a transformation A which may be a rectangular matrix also, then this particular circle transformed into this ellipse, having this orientation. Now, what is meaning of that in terms of singular value decomposition?

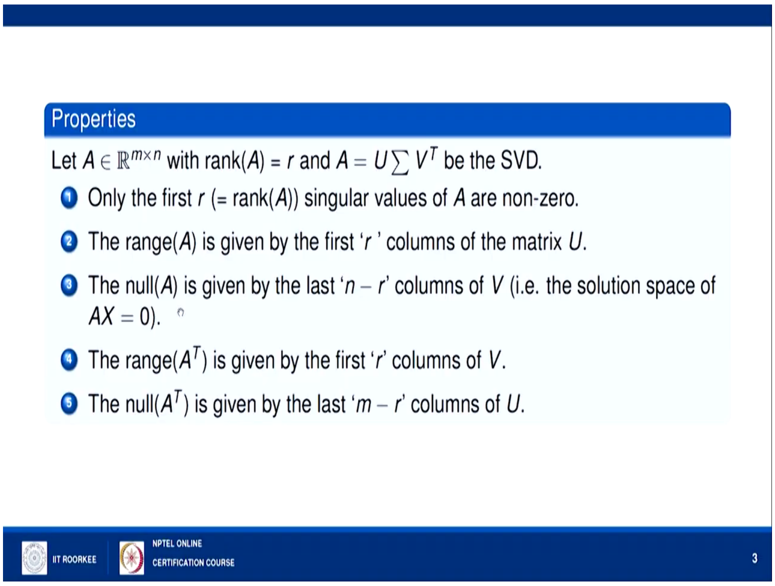
So, as you know that singular value decomposition is a equals to U sigma V transpose. So, V star; star is transpose in case of real matrices. Now, when I apply this A onto this circle, so it will become let us say this object is my someone B. So, U sigma V star into B. So, first I will apply V star on it. So, once I apply V star on it which is an orthogonal matrix because V is an orthogonal matrix. So, V transpose, then orthogonal matrix rotates the object.

So, what will happen? It will rotate this object by some angle based on the value of V transpose. So, now, I am having the same circle, but orientation is different. Now, I will apply sigma on it. Once, I will apply sigma. So, sigma is having scaling factors and though these scaling factors are proportional to singular values, so it will scale this circle and it will deform into an ellipse.

If both singular values are equal, sigma 1 equals to sigma 2, it will remain as a circle; otherwise, it will become an ellipse, where this distance will be sigma 2 and this distance will be means ratio will be sigma 1. After that, what will happen? I will apply U on it. U is again an orthogonal matrix and it will rotate this ellipse in this way. So, this is the geometrical interpretation of singular value decomposition.

Here, what we are having? We are having 2 rotations and 1 scaling. So, in this way, we can interpret singular value decomposition that it is a sequence of transformation; first rotation, then scaling and then rotation. So, this is the geometrical interpretation of singular value decomposition.

(Refer Slide Time: 03:37)

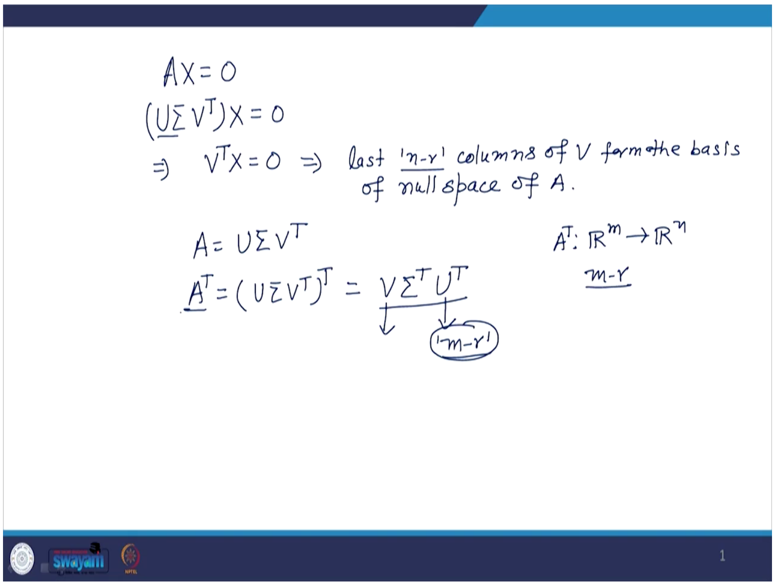


Now, we will discuss some of the properties of singular value decomposition. So, let A be a m by n real matrix means having real entries and rank of A equals to r. Let us consider that the singular value decomposition of A is U sigma V transpose. Now, if the rank A equals to r, then the first r singular values of A will be nonzero. So, what I want to say that rank of A matrix equals to number of nonzero singular values. The range of A that is the range space of A is given by the first r columns of the matrix U.

So, as you know that A is a m by n matrix, so it is a linear transformation from Rn to Rm ok. So, what will happen? Range of A is given by the first r columns of the matrix U. What will be the size of matrix U? It will be of m by m. So, what will happen? We will be having m-dimensional vectors.

And how many vectors? r vectors that is the that is the rank of A. Those m-dimensional r vectors, those are the first r columns of m by n matrix U will give the range space of A, that is they will form the basis of range space of A. Similarly, the null space of A is given by the last n minus R columns of V that is the solution space of AX equals to 0.

(Refer Slide Time: 05:31)



So, what I want to say; what I want to find out? AX equals to 0 and AX equals to 0 means U sigma V transpose into X equals to 0. So, from there, I can say that it will be 0, when V transpose X equals to 0. Because when this matrix will operate with a 0 vector, it will give you 0 and from here, what I want to say that last n minus r columns n is the size of V n by n is the size of V and r is the rank. So, last n minus r columns of V form the basis of null space of A.

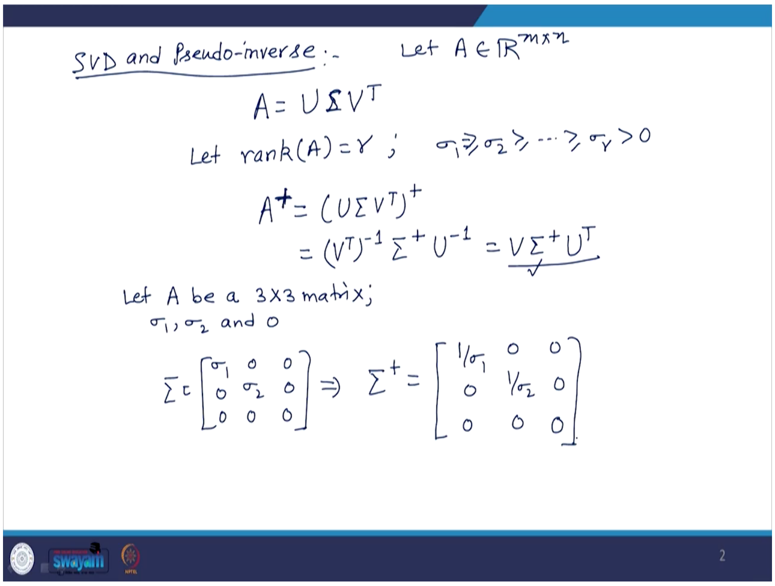
Now, what I want to say in other word that last n minus r columns means the columns those are associated with the 0 singular values will give you the null space of A. Similarly, if you talk about the range space of A transpose. So, you can see easily A equals to U sigma V T. So, A transpose will become transpose of U sigma V transpose that is V sigma T U transpose.

So, now what I want to say that the in the similar way that the range space of AT is given by the first r columns of V and the null space of A transpose is given by last n minus r earlier it was n minus r. Here, it will be m minus r. Because now, it will be a transformation from Rm to Rn, that is A transpose and by the rank nullity theorem, rank is R which is equals to a rank of A equals to rank of A transpose.

So, here nullity will become m minus r. So, that is why I am taking these m minus r columns and these columns are orthogonal. In fact, orthonormal because the matrix U and V are orthogonal matrices. So, they are obviously, linearly independent ok.

So, last m minus r columns of U will form the basis of null space of A transpose and the range space of A transpose is given by the first r columns of U sorry V. So, these are some of the properties, you can directly if you are having the singular value decomposition of a transformation or a matrix, you can directly find out range space, null space, rank nullity are these kind of thing. Another application is how to what is the relation between SVD and pseudo inverse?

(Refer Slide Time: 09:02)



So, I am having singular value decomposition as A equals to us V transpose. Let A is m by n matrix. So, here U will be m by m orthogonal matrix and V will be n by n orthogonal matrix and S is a m by n matrix. Let rank of A equals to r. So, what I will be having? I will be having sigma 1, sigma 2 sigma r greater than 0 and then, rest of the singular values will be 0. Now, what will happen if I calculate A inverse? Because A is a rectangular matrix. So, how pseudo inverse, I will define here? It will become U. So, I am writing let us write sigma.

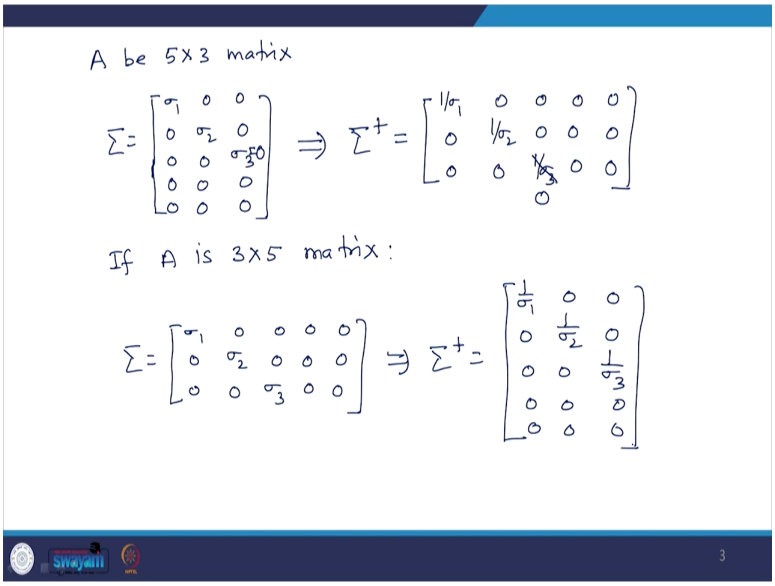
So, U sigma V transpose this become V T inverse sigma plus that is I am writing for pseudo inverse and U inverse as you know V is an orthogonal matrix. So, V transpose equals to V inverse. So, V transpose inverse will become V transpose and I can write it V. Then, sigma plus and U inverse again since U is an orthogonal matrix. So, U inverse will become U transpose.

So, if U and V transpose are given you can easily find out V and U transpose. Now, how to write this sigma plus. So, I will tell you how to write the sigma plus. So, let A be a 3 by 3 matrix. I am taking some examples and singular values are sigma 1, sigma 2 and 0. So, sigma 1, sigma 2 are not equals to 0 here, means I am having two nonzero singular values and one 0.

So, in this case, what will be sigma? Sigma will be sigma 1 0 0 0 sigma 2 0 and 0 0 0 because your sigma 3 is 0. Now, in this case what will be your sigma plus? It will become 1 upon sigma 1 0 0, 0 1 upon sigma 2 0, 0 and like in usual cases if we are going for inverse the diagonal matrix is having entries like a 1, a 2, a 3 then inverse will be having diagonal entries 1 upon a 1, 1 upon a 2 and 1 upon a 3. But here, it will become 1 upon 0 in this case. So, simply I will replace it by 0.

So, if it is less than a very small entry or 0, close to 0 or 0; in that case, simply I will replace it by 0, not by the reciprocal value because that will not be defined or that will become a very large value. So, this is the case of square matrix.

(Refer Slide Time: 13:04)

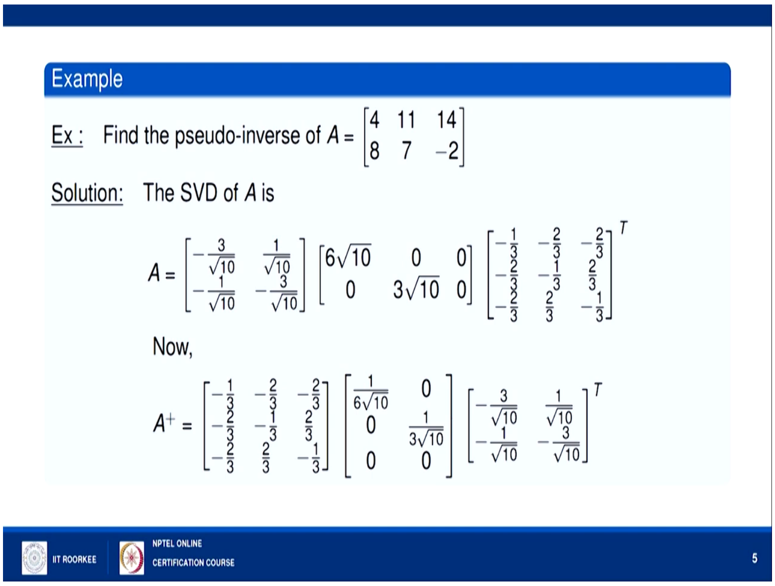


If I am having a rectangular matrix, let us say A be 5 by 3 matrix ok. In this case, sigma will be something like this; sigma 1 0 0, 0 sigma 2 0, 0 0 sigma 3, 0 0 0, 0 0 0. So, this will be your sigma. Now, what will be pseudo inverse of sigma in this case? That is sigma plus it will become a 3 by 5 matrix and it will be first entry will be 1 upon sigma 1, if sigma 1 is nonzero; if sigma 1 is 0, it simply that will be 0 matrix only. Then, 0 1 upon sigma 2 0 0 0, 0 0 1 upon sigma 3 0 0.

So, if sigma 3 0, let us say it is 0 here. In this case instead of 1 upon 0, I will write simply here 0. Similarly, if a is 3 by 5 matrix. In this case, sigma will be 3 by 5 that is sigma 1 0 0 0 0, 0 sigma 2 0 0 0, 0 0 and then sigma 3 0 0. In this case, pseudo inverse of sigma will be something like this 1 upon sigma 1 0 0, 0 1 upon sigma 2 0, 0 0 1 upon sigma 3, 0 0 0, 0 0 0 and again, the same rule. If it is 0 or very close to 0, some of the singular value, we will replace in the pseudo inverse by 0 only.

So, in that way, you can write sigma plus, rest you know how to find out if V transpose is there, how to write V and transpose of U and then, by multiplying these V sigma plus and U transpose, you will get the pseudo inverse. Left, right, in both the cases; only difference will be how to write sigma ok. So, this is the application of singular value decomposition for calculating the pseudo inverse.

(Refer Slide Time: 15:50)



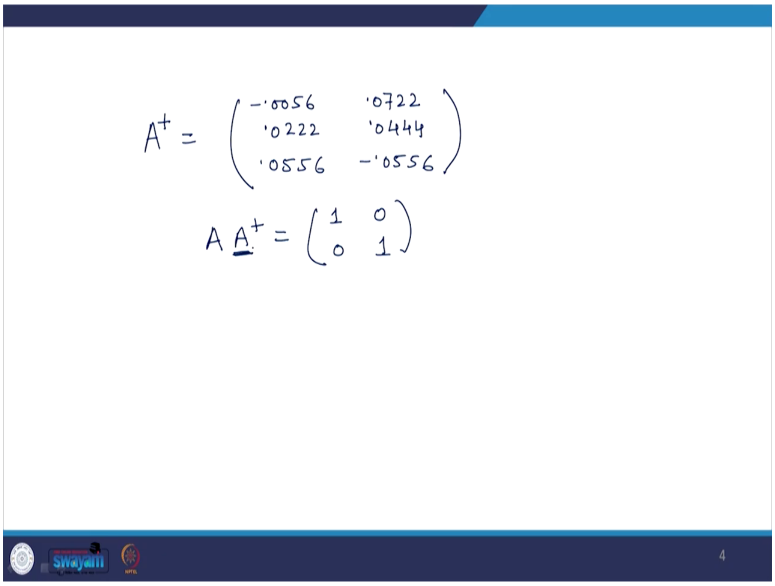
Let us see an example of this. So, find the pseudo inverse of A, where A is a 2 by 3 matrix; entries of A are 4 11 14, 8 7 and minus 2. So, first we will find out the singular value decomposition of A which is given by this is my matrix U, this is the matrix sigma and this is the matrix V.

So, here if you see that U is minus 3 upon square root 10, 1 upon square root 10, minus 1 upon square root 10 and minus 3 upon square root 10. Here, I am having singular values of a as 6 root 10 and 3 root 10 and these are coming from the eigenvalues of A transpose A. So, if you calculate our A A transpose.

So, A A transpose will be a 2 by 2 matrix and eigen value of that will be 360 and 90. If you are going for A transpose A, it will become a 3 by 3 matrix and eigen value will be 360 again, 90 and 0 because one dimension is increasing and it is having rank 2. So obviously, third eigen value will be 0. So, I am having this sigma 1 0 0, 0 sigma 2 0 and then this is my matrix V. So, if you carefully see the columns of these matrices are orthonormal that is pair wise orthonormal, they may they form a set of orthonormal vectors.

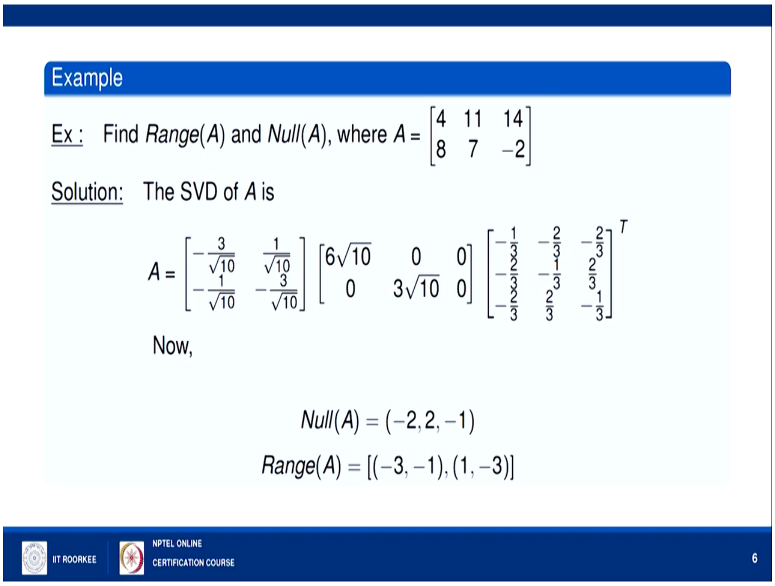
Now, in the previous slide we have seen that the pseudo inverse of A is given by V sigma plus U transpose. So, here I am writing U and transpose here. This is my matrix V, so I am writing as such and then, this will become it is of size 2 by 3. So, sigma plus will be of size 3 by 2 and it will become 1 upon 6 root 10 0, 0 1 upon 3 root 10, 0 0 and if you multiply these matrices, you will get pseudo inverse.

(Refer Slide Time: 18:19)



And it comes out to be minus 0.0056 0.0722, 0.0222 0.0444, 0.556 and minus 0.556 ok. And if you calculate here A into pseudo inverse of A, you will get you can verify that you will get the identity matrix of order 2, that is why it is the right pseudo inverse here in this case. So, this is the example means application of singular value decomposition for calculating pseudo inverse.

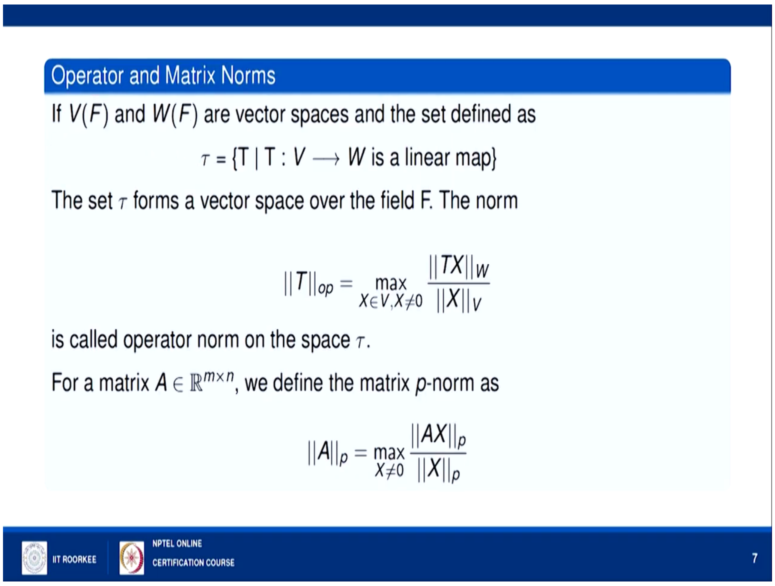
(Refer Slide Time: 19:17)



Another application the same matrix, I am having and find the range space of A and null space of A. So, if you recall from one of the previous slide, the range space of A is given by the first r columns of matrix U, where U is a orthogonal matrix and R is the rank of A. So, hence, this is my matrix A. So, hence, range space of A is given by just I have now taken normalized one minus 3 minus 1 that is the first column and the second column is 1 and minus 3. So, these two vectors in R 2 are linearly independent and they are forming the range space of the matrix A.

Similarly, null space of A is given by the last n minus r columns of the matrix V. Here n is 3, r is 2. So, last column, last one column; so, last one column is minus 2 by 3, 2 by 3, minus 1 by 3. So, if I am not normalizing it, I can write it simply minus 2, 2 and minus 1. So, hence, rank of A is 2; nullity of A is 1. So, range plus nullity equals to the vector space R 3 which is A because A is a linear transformation from R 3 to R 2. So, in that way the singular value decomposition is given, you can easily find the basis and basis for range space as well as null space.

(Refer Slide Time: 21:02)



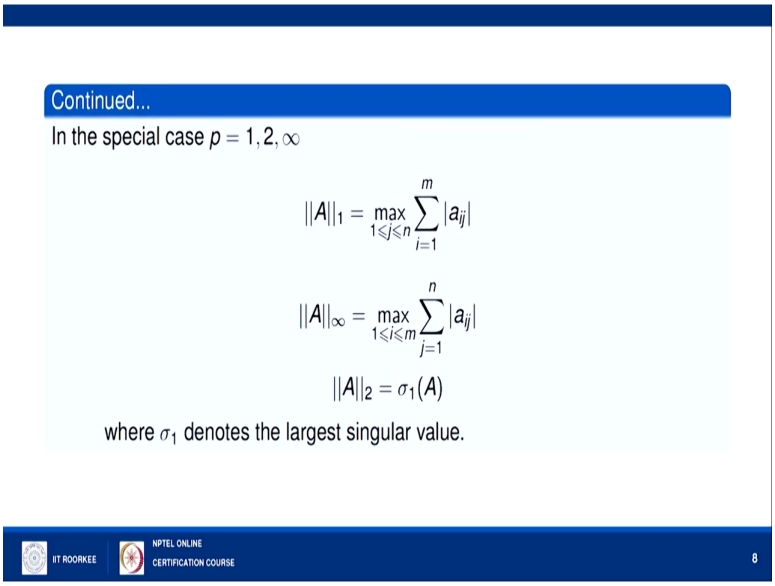
Now, come to the second part of this lecture. So, just again consider that you are having two vector spaces V and W defined over the field F and now, define a set of linear transformation tau for having all the linear transformation from V to W.

So, tau is the set containing all the linear transformations defined from vector space V to vector space W. So, if you carefully see the set tau forms a vector space over the field F, you can verify it easily. Because zero transformation will be there. It will be closure; it will form an abelian group with respect to addition of linear transformation and so on..

Now, define the operator norm T operator norm as maximum X belongs to V and X is nonzero norm of TX with respect to W upon norm of X with respect to V because TX will be the vectors in vector space W. This norm is called the operator norm on the space tau, where tau is a space of linear transformations.

So, let us understand it in other way, in terms of the matrix. So, I am having a matrix A which is of order m by n having real entries. We define the matrix p-norm as, so p-norm of A equals to maximum X not equals to 0, AX p upon X and here, this particular we can generalize it for different values of p.

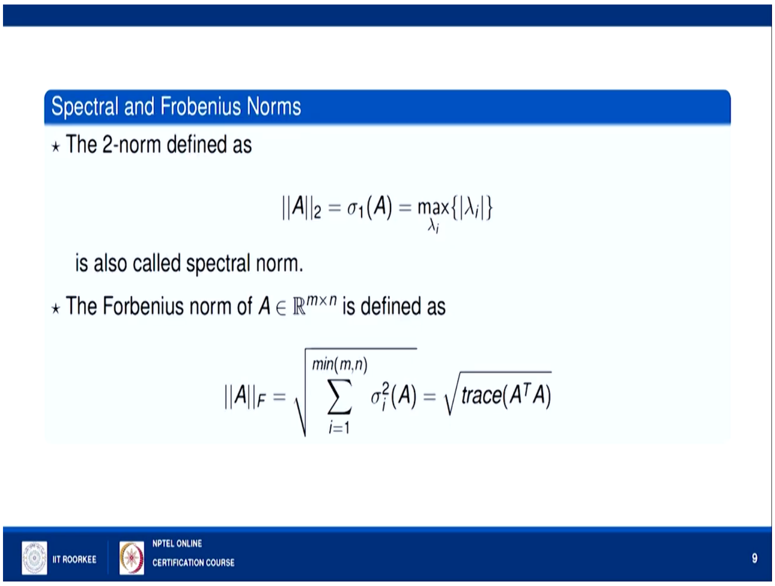
(Refer Slide Time: 22:53)



For example, if I take p equals to 1, then this become the matrix 1 norm and it is given by the maximum 1 to j to n summation i equals to 1 to m and absolute ai j. So, what this one is saying? So, if you carefully see it, I am running this summation over i from 1 to m. So, I am taking the sum of the columns and what kind? If you are having negative entry, I am taking the absolute value for each columns, I am taking the sum of the absolute entries of individual columns and then, I am taking the maximum one.

Similarly, you are having the infinity matrix norm and just what we are doing in this case? Instead of column sum, I am taking the row sum and out of this whatever row is giving me the maximum sum, I am taking that as the infinity norm. Another norm is that is called a Spectral norm, that is given by the largest absolute eigen value or the singular value, largest singular value. So, spectral norm is nothing just the spectral radius of the matrix. Because what will be the spectral radius? It will be the largest absolute eigen value.

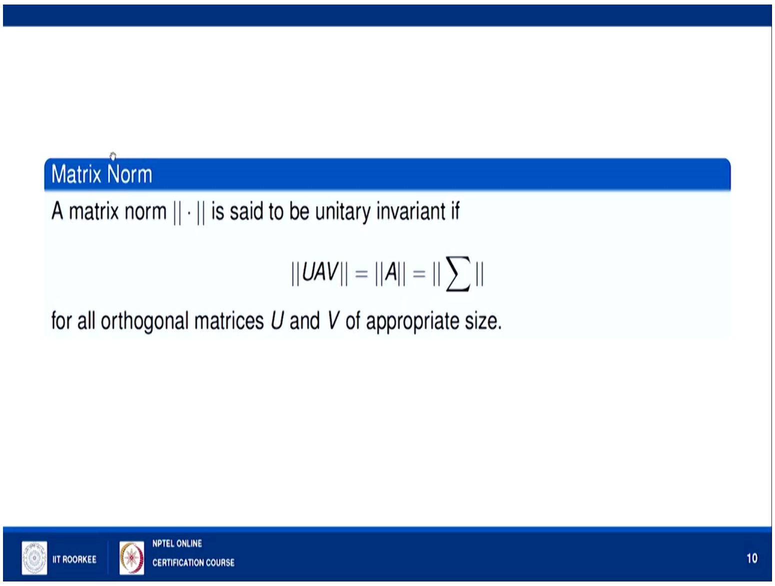
(Refer Slide Time: 24:22)



So, as I told you, this is in terms of eigen value we can write in this way and this two norm is also called a spectral norm as I told you. We are having Frobenius norm of the matrix A and

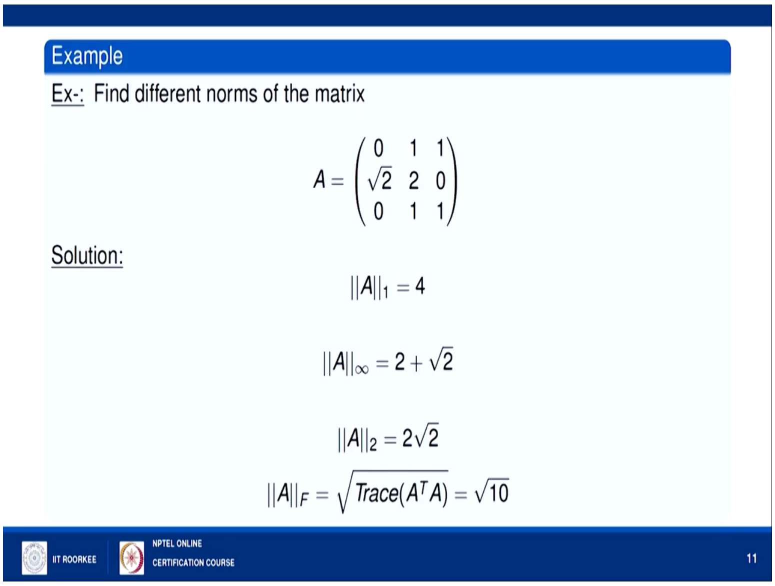
Frobenius norm is nothing just the square root of the sum of the square of the singular values and that will be nothing just a square root of the trace of A transpose A, as I am giving here. So, these are different norms which we are defining over the matrix and all these norms are having own importance, we will see later.

(Refer Slide Time: 25:02)

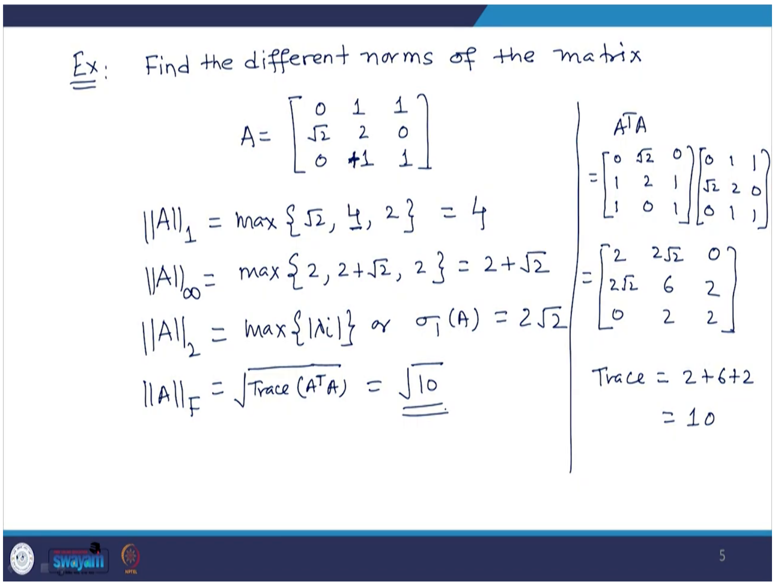


A matrix norm is said to be unitary invariant, if the norm of UAV equals to norm of A and that is norm of sigma, in singular value decomposition of A, for all orthogonal matrices U and V of appropriate size. Appropriate size means where this product is defined. So, let us take an example.

(Refer Slide Time: 25:25)



(Refer Slide Time: 25:32)



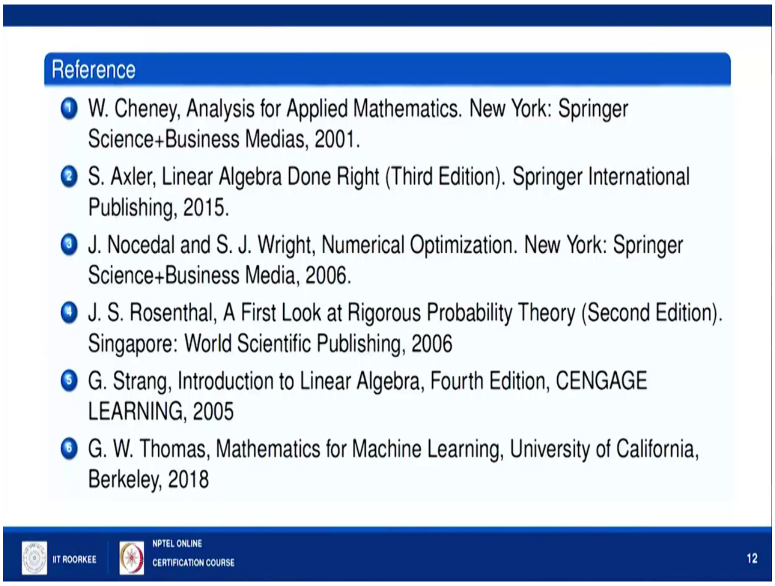
So, find the different norms of the matrix. So, my example is of the matrix and what are the entries of matrix 0 1 1. So, first column is 0 1 1, sorry first row and second row is root 2 2 0 and third row is again 0 1 1. So, let us say this is my matrix A. So, first let us take one norm. So, one norm will be maximum of column sum. So, what is the sum of first column? Root 2, sum of second column is 1 plus 2 plus 1, 4 and third column is again 1 plus 1, 2. So, here this comes out to be 4. Please note that if I am having minus 1 here, still it will become the 4 because I am taking the absolute value.

So, at this moment, I am taking is as plus ok. Now, infinity norm will become the row sum. So, it will be maximum sum of first row is 2, sum of second row is 2 plus root 2 and sum of third row is again 2. So, it comes out to be 2 plus root 2. Similarly, the spectral norm will be the maximum absolute eigenvalue of A or sigma 1 of A that is the largest singular value of A that is the equal; both are the same thing. So, in this case, if I check here it comes out to be 2 root 2. So, you can verify it easily.

Now, if I talk about the Frobenius norm, so it will be square root of trace of A transpose A. So, let us first calculate A transpose A. So, A transpose A equals to 0 1 1, root 2 2 0 and 0 1 1 and then, I am having A; A will be 0 1 1, root 2 2 0 and 0 1 1. So, this comes out to be 2 2 root 2 and then, 0 1 0 2 root 2 1 plus 4 plus 1, 6 and this will be 2, then 0, this will be 2 and this will become 2. So, your trace of this comes out to be 2 plus 6 plus 2 that is 10. So, it will become a square root 10 that is the Frobenius norm of the matrix A.

So, in this lecture, we have learned some applications of singular values that how to find out range space, null space of a given matrix using the singular value decomposition, how to calculate pseudo inverse using singular value decomposition and then, we have defined some of the matrix norm and we have seen by the help of an example that how to calculate all those. In the next lecture, we will again see a very important application of singular value decomposition in machine learning that is the low rank approximation. These are the references for this lecture.

(Refer Slide Time: 30:05)



Thank you.