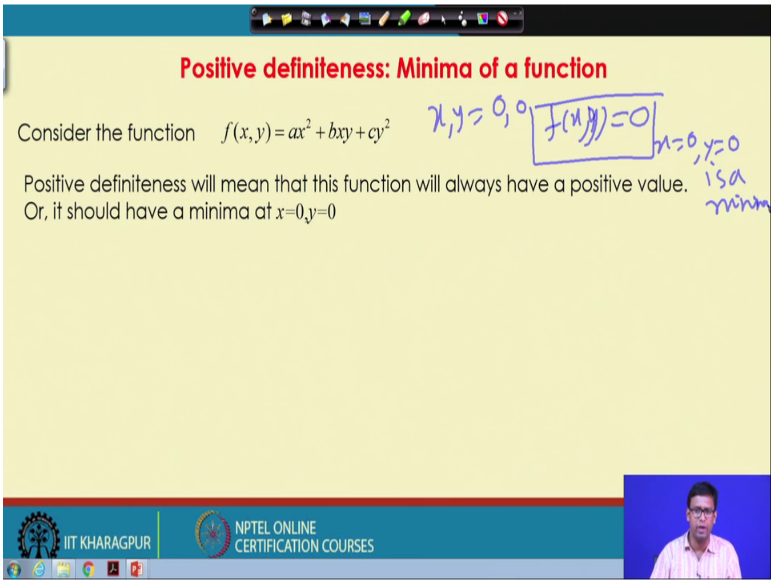
Matrix Solvers
Prof. Somnath Roy
Department of Mechanical Engineering
Indian Institute of Technology, Kharagpur

Lecture – 31
Positive Definiteness of a Matrix (Contd.)

Welcome, in this session we will continue our discussion with few of the fundamental aspects of matrix algebra. At least we will get the definitions clear so, that when they will appear in the later part of this course especially when we will deal with iterative methods, we feel our self more conversant with these terms. So, we will discuss about the few more aspects which are positive definiteness, singular value and condition number.

(Refer Slide Time: 00:48)

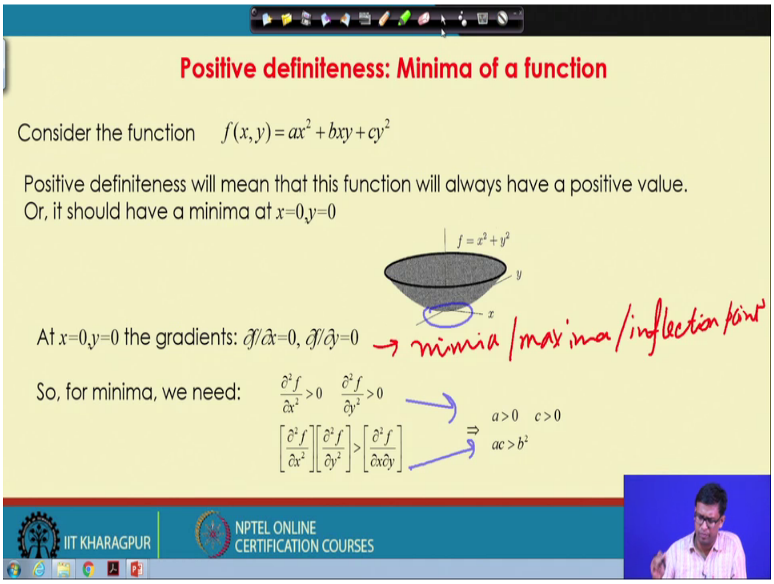


So, let us consider a function which is a function of x and y and it is a general quadratic function of x and y; f x, y is equal to ax square plus bxy plus cy square does not have any other constant in this function. So, it is entirely varies with x and y and we try to find out what is the minima of this function. Why the minima is important? Because positive if we say that this function is positive definite name, then we say that definitely this function is positive; that means it will always have a positive value.

Now, if we look into this function f x, y is equal to ax square plus bxy plus cy square at x, x, y is equal to 0, 0 f of x, y is equal to 0. So, the minimal value is not positive, it we can get at least one location where f x, y is 0. So, this function can be a positive definite function; that means it can only have positive values; definite does not have definite positive values. In a case where this f x, y at x is equal to 0 y is equal to 0 x is equal to 0 y is equal to 0 is a minima.

Then this is the minimal value of this function; that means, every other value of this function is a positive value. So, the function except x is equal to 0, y is equal to 0 everywhere has a positive value. Positive definiteness will mean that the function will always have a positive value or it should have a minima at x is equal to 0, y is equal to 0.

(Refer Slide Time: 02:39)

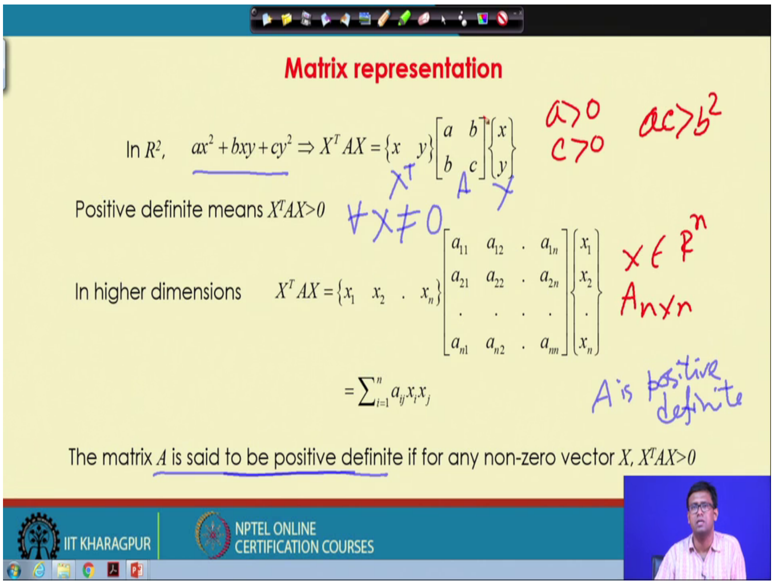


So, this can be a function like this; concave function where the minima is that x is equal to 0, at y is equal to 0. And how do I ensure that the minima is at x is equal to 0 and y is equal to 0? We will take the gradients will consider the gradients. So, you can see that at x is equal to 0 y is equal to 0, then radiants are already 0; del f del x is equal to 0 del f del y is equal to 0 because there is a quadratic function of x and y without any other constants clear.

So, these things say that there can be three cases and we know that if first derivative is 0, it can be a minima, it can be a maxima or it can be a an inflection point. So, what are the tests that it should be a minima? The test is that that the second derivatives should be greater than 0 and not for like with respect to x and y and not for being an inflection point d 2 f dx square into d 2 f dy del 2 f del x square into del 2 f del y square should be greater than equal to del 2 f del x del y.

Now, if we impose this conditions on the function, this function will get that the first two will give me a is equal to 0's; a is greater than 0 and c is greater than 0 and the third one will give us ac is greater than b square. So, if these conditions hold, then will say that x f x, y is a positive definite function; that means, it will definitely have positive value will know where it will have a negative value.

(Refer Slide Time: 04:39)



So, we can say in R 2 function f x, y is x square plus bxy plus cy square it is positive definite; that means, it only has a positive value except x is equal to 0, y is equal to 0; that is the only location where it has a negative value. So, if we can represent this function as x transpose AX where x y is the x vector. So, this is X transpose this is A and this is X.

Now, positive definite means it has a 0 value at the a at x is equal to 0, y is equal to 0 which is the minima of this function. So, if x is a non-zero vector, nowhere it has a 0 value everywhere it is a it is a positive function is definitely a positive function. And in that case we write that positive definiteness means X transpose AX is equal is greater than 0 for all X which is a non-zero vector.

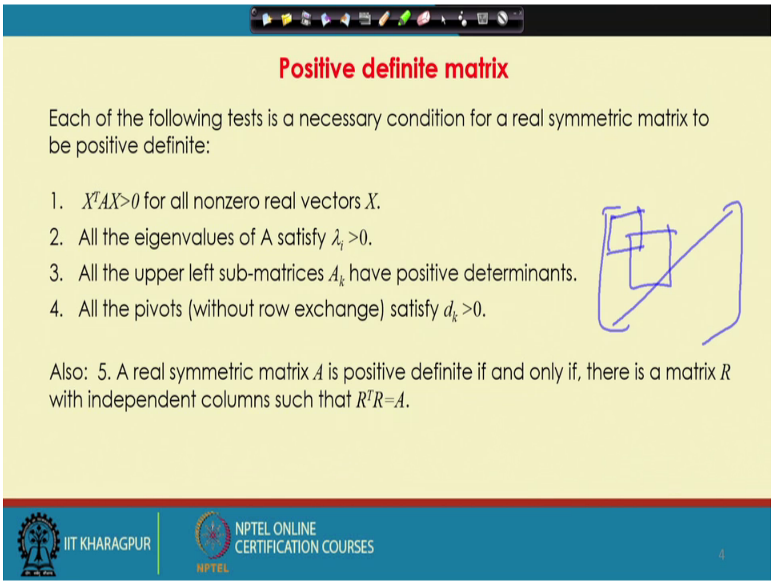
And in a higher dimensional space this can be instead of x be x in R 2 and a being a 2 by 2 matrix, this can be a multi dimensional matrix X belongs to or we can write this with the vector X belongs to R n and A is A n into n matrix. So, this X transpose AX will give us a ij x i x j and it will also be a positive definite matrix, if X transpose AX for any non zero vector AX X transpose AX is greater than 0.

So, the idea is similar that we have a matrix a which is multiplied pre multiplied by transpose of a non zero vector and post multiplied by the non-zero vector and the product which is a quadratic function of different x and y or the quadratic function of different components of the vector of vector x. This product is always greater than 0, that x is a non zero vector if x is a 0 vector, it is 0, but f x is a non zero vector; this product is always greater than 0 or definitely this product is positive. Then the matrix A is also said to be a positive definite matrix. In that case we write that A is positive.

Now, earlier we have seen that there are few conditions based on which like when we took the function ax square bx y plus c x square plus bxy plus c square cy square. There are few conditions based on which we can say that x is equal to 0 x is equal to 0 is a minima and therefore, it is a positive definite function x x squared plus bxy plus cy square will have a non zero value for any other values of xy except the 0 0. And these conditions are a is greater than 0, c is greater than 0 and we got ac is ac is greater than b square.

So, that means, the matrix a should pose a certain properties based on which, this will be decided that x transpose ax is greater than 0 for any x except x is equal to 0. And if we look into this properties that there will be few tests on the matrix A and will see how certain parameters for example, eigenvalue or the determinant or the pivots of the matrix appear. So, that we can say that X transpose AX is a is positive definite functional and A is a positive definite matrix.

(Refer Slide Time: 08:47)



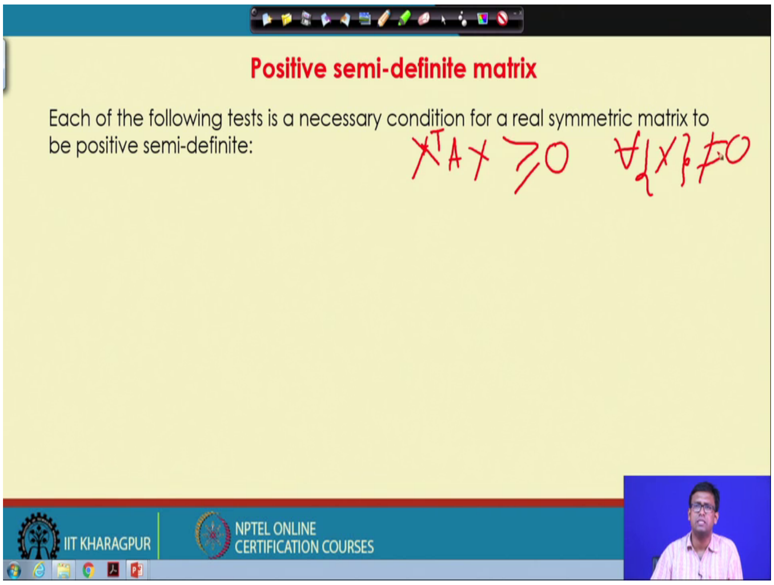
So, there are few tests. Each of the following tests is a necessary condition for a real symmetric matrix to be positive definite. So, each of this has to be satisfied X transpose AX is greater than 0 for all non zero real vectors X. All the eigenvectors x all the eigenvalues of A, must satisfy these are not necessary all these are necessary and sufficient conditions. So, if at least one of this condition is satisfied, the rest are also satisfied and A is a positive definite matrix; that means, X transpose AX is greater than 0 for all non-zero real vectors x that is from the first condition which is same which will be also satisfied if eigenvalues of a satisfy that all the eigenvalues are greater than 0. There is no non-zero, no 0 eigenvalue, A is not a singular matrix and also there is no negative eigenvalue of A.

All the upper left sub matrices of a for example, if we have a large matrix A all the upper left sub matrix, this matrix or maybe this matrix. All the upper left sub matrix of A, if we take this diagonal any matrix above this will have a positive determinant. All the pivots without doing row exchange must satisfy d k is greater than 0. So, all the pivots should be positive pivots.

Fifth is that that it is for a real symmetric matrix A. It is positive definite if and only if there exists a matrix R which as independent columns. So, that R transpose R is equal to A. So, R is also square matrix there can be a square matrix which has independent columns. So, that A can be classified as A can be decomposed as R transpose R in and that is for a symmetric matrix with real values; then A is a positive definite matrix.

So, if any of these properties hold A must be a positive definite matrix. The last one is only for symmetric matrix. Symmetric positive definiteness has certain very interesting properties and few applications of matrix solvers are specifically for symmetric positive definite matrices. We will discuss it later.

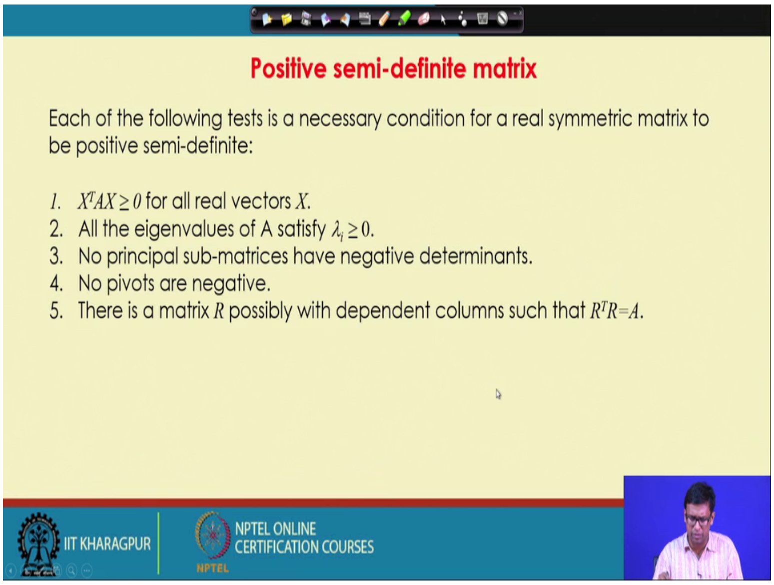
(Refer Slide Time: 11:24)



There is another term called positive semi-definite matrix and positive semi-definite means this is more negative, X transpose AX X transpose AX is greater than equal to 0 for all x which is non not equal to 0.

So, it can be X transpose AX is possible to have X transpose AX 0 also, but it is it is not up what I will say that it is not necessary that X transpose AX is always positive it can be 0 also, but it is never negative. And for real symmetric matrices, there are few tests for positive semi-definite matrix. Second we came to the paradigm of symmetric matrices.

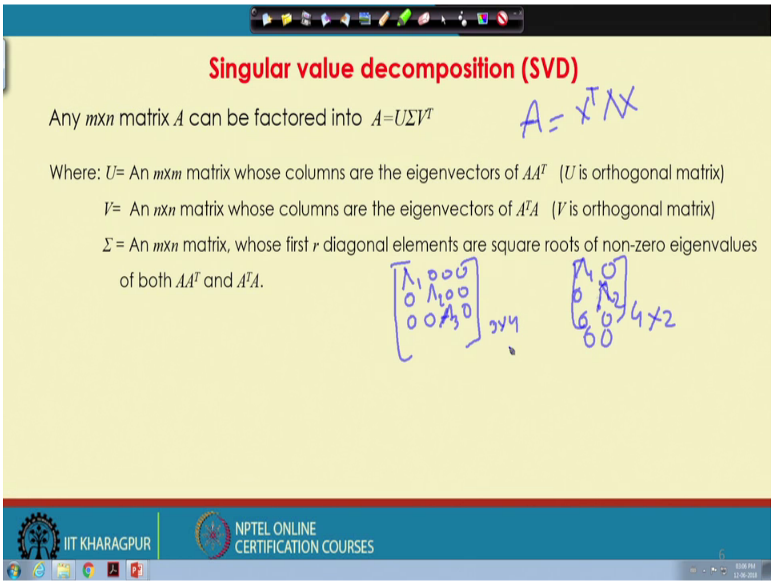
(Refer Slide Time: 12:17)



That X transpose AX is greater than equal to 0 for all real vectors X. All the eigenvalues satisfy that the eigenvalues are non-negative; it is either greater than 0 or is equal to 0.

So, A can be a singular matrix also. No principal sub matrix has negative determinant, the none of the pivots are negative. There can be 0 pivot the pivot, but none of them are negative. So, it can be a singular matrix that is why there can be 0 pivot also. And there possibly exist a matrix R there exists a matrix R, we possibly have dependent columns. So, that we can decompose A as R transpose R and so, few of these things of similarity with positive definite matrices; only thing that positive semi definite matrix a can be a singular matrix too, but positive definite matrix has no eigenvalue which is 0; all eigenvalues are greater than 0. So, it is definitely a non singular matrix.

(Refer Slide Time: 13:29)



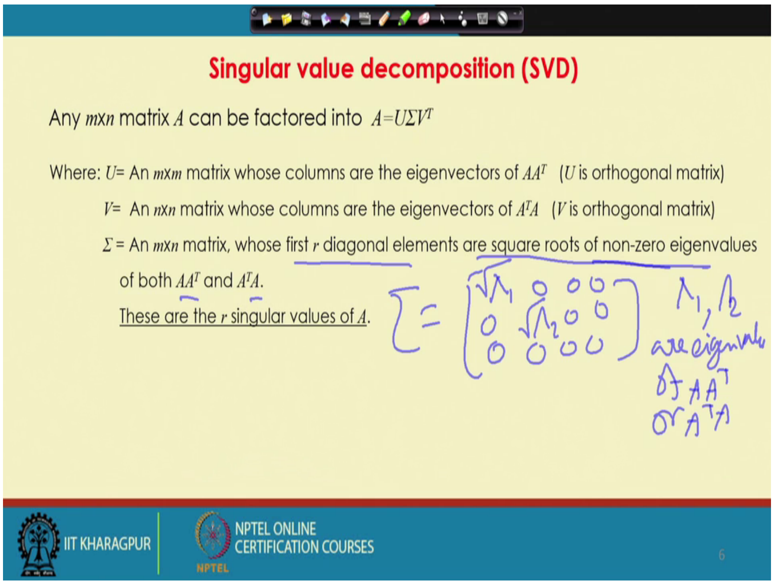
So, you go to the next set of the next property which you are interested in or this is this is not a property these decomposition of matrix is one particular operation on a matrix. And for if we can remember that for square matrices, we obtained that A is equal to; we can decompose any matrix A is equal to vector X transpose lambda X, where X is a orthogonal matrix X is or X is the matrix combining the eigenvalues, lambda is the eigenvectors, lambda is the eigenvalue matrix.

In case of a rectangular matrix, a similar decomposition is possible which is given as A is equal to U sigma V transpose. Instead of X transpose A and X where using 2 different matrices U and V transpose and instead of the lambda or the eigenvalue matrix. We are also using another different matrix sigma, where U is an m into m matrix whose columns are the eigenvectors. Now, they are not eigenvectors of A because A is a rectangular matrix. We cannot have eigenvalue and Eigen vector of a, but we can get eigenvector of A transpose. A is a U is a m into m matrixes columns are the eigenvectors of U u transpose U is and then U is also an orthogonal matrix, why because A transpose is a symmetric matrix. Therefore, it is eigenvectors must be orthogonal to each other.

Similarly, V is an n into n matrix whose columns are the eigenvectors of A transpose A, V is also symmetric matrix. So, the V sorry A transpose A is also symmetric matrix. So, eigenvectors are orthogonal to each other therefore, V is also an orthogonal matrix. And sigma is an m into n matrix whose first R diagonals are the square roots of non-zero eigenvectors of both A transpose and A transpose A. Depending on how many indepe[ndent], what is the rank of a or how many independent columns or rows do a have, we can get non-zero eigenvalues of A transpose and A transpose A and they will come in the matrix sigma.

So, sigma will look like a square matrix say, will have lambda 1, 0 0 0 0 lambda 2 0 0 may be 0 0 lambda 3 lambda 3 0 if sigma is 3 into 4 matrix. So, depending on and this three are the independent rows number of independent rows and columns of A. So, depending on that sigma, we will have non-zero non-zero eigenvalues; non non-zero diagonal terms. For example, if I have something like a 4 into 2 sigma a is 4 into 2 sigma will also be 4 into 2. So, this will be lambda 1 0 0 lambda 0 sorry 0 lambda 2 0 0 0 0. Only these 2 diagonals will have these 2 rows will have non-zero diagonal terms.

(Refer Slide Time: 17:30)

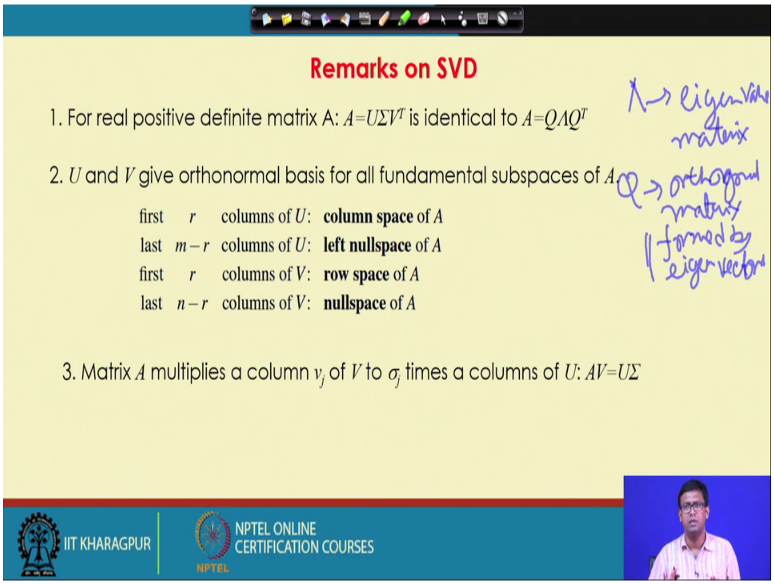


And these R diagonal elements which are the square roots of non-zero eigenvalues of A transpose and A transpose A. So, rather we should write that sigma will have the form of root over of lambda 1 0 0 0 0 root over of lambda 2 0 0 may be 0 0 0 0 0. And lambda 1 and lambda 2 are Eigen values of A transpose or A transpose A.

So, A transpose and A transpose they essentially have same eigenvalues, same non-zero eigenvalues because one of them will have a larger dimension; however, the there will be 0 eigenvalues corresponding to that dimension nevertheless. So, we can get the first R diagonal elements which has square roots of the non-zero eigenvalues of A transpose and A transpose A and these are called the singular values of A. The non-zero I have to find out the non-zero eigenvalues of A transpose or A transpose a given A is a rectangular matrix and square root of the non-zero eigenvalues will be called as the singular values of A.

And when we can decompose the matrix A as U sigma V transpose where sigma is the matrix comprising the singular values of a which are the square root of A transpose and A transpose A. This is called singular value decomposition. And interestingly if we can take a square matrix A, we will see that this is very similar to a X sigma X transpose type of decomposition of A. Where the matrix sigma will now compose the eigenvalues of A, if a is a square matrix. It is defined for a general matrix which is a rectangular matrix and we get singular values which are square root of the eigenvalues of A transpose and A transpose A.

(Refer Slide Time: 19:45)



So, there are few remarks for a real positive definite matrix A. Positive definite means it is a non-singular matrix and it does not have a negative eigenvalue also. A is equal to U sigma V transpose is identical to A is equal to Q lambda Q transpose where we can write lambda, we have seen this form earlier lambda is the eigenvalue matrix. It is a diagonal matrix whose diagonal elements are the eigenvalues only and Q is an orthogonal matrix formed by Eigen vectors. Q is an orthogonal matrix is formed by eigenvectors that depends on, if the if A is symmetric; then Q is formed by eigenvectors.

If U and V give orthonormal basis of all fundamental subspaces of A; so, you have seen that sigma has R eigenvalues are non-zero Eigen increase non-zero entries in sigma that would A transpose and A transpose A will have eigenvalues non-zero eigenvalues which is exactly same as the dimension of the matrix as a rank of the matrix A. Therefore, U and V will have same number of because U and V are the Eigen vector matrices will have same number of independent eigenvectors.

So, similar that will be determined first r columns of U will give the column space of A which is the number of independent columns which will number of independent columns of A is basically r. So, first for r columns of A, we will give the will be the basis of column space of A and m minus r columns will be; obviously, U is a m into m matrix m minus r columns will; obviously, give the complement orthogonal complementary space of A which is the left null space of A.

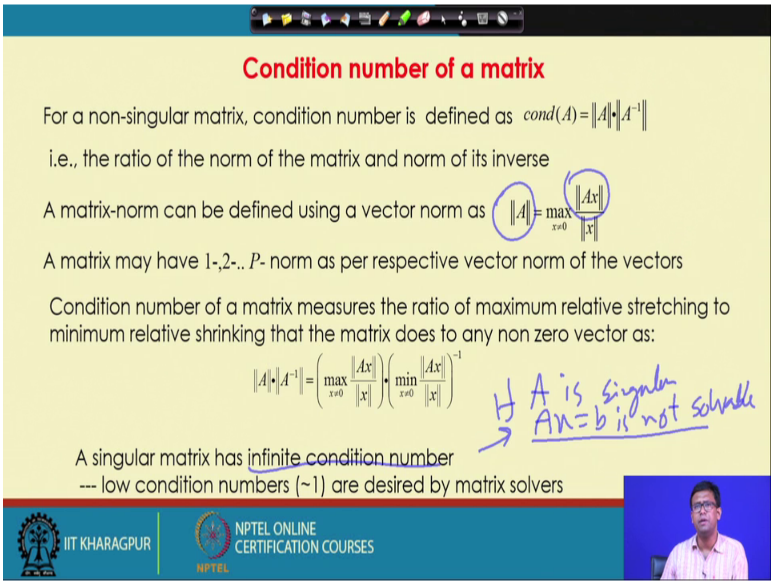
Similarly, first r columns of V will give the dimension of basis for row space of A and last n minus r columns which will be the complementary part, which is null space of A. So, if we can do a singular value decomposition not only we get the singular values which are eigenvalues of A transpose, but we can also get the fundamental basis for orthonormal basis for fundamental subspaces of A. So, there orthonormal basis; that means, there independent spanning vectors as well as each vector has an unit length and each vector is perpendicular to the to the rest of the basis vectors. We do not need to do a gram smith type of operation to get an orthogonal set and they already orthogonal set.

And matrix A and this is a very interesting property matrix A multiplies a column of column v j of V to sigma j times of a column of U. Sigma j is what is sigma j? Sigma j is the eigenvalue square root of eigenvalue of A transpose or the singular value of A. So, A multiplied with V will give U multiplied by the singular value matrix as any matrix. So, AV is equal to U sigma.

So, we will go to another third concept which I am supposed to discuss to the condition number and we have discussed singular value mostly in 2 purposes. The main purpose is that that when we look into condition number which is lot of importance in matrix solvers that condition numbers can be estimated using singular values. We will not do the detailed algebraic proof for that, but we will show the formula will show at the end, how to express condition number using the singular values. This is the main purpose; however, there is another purpose which is that any rectangular matrix can also be decomposed into 3 matrices; 2 matrices with orthogonal column vectors and the third matrix which is a kind of a diagonal matrix with few non non-zero diagonals and remaining terms are diagonal in this matrix.

And this, by this decomposition, we can get the basis for the fundamental subspaces of a column space, row space. Space left null space, bases for all these spaces can be obtained. And also AV is equal to U sigma, this relation says that I can transform one space to I can transform column space to row space or vice versa using this particular decomposition and the matrix. This is in linear algebra singular value decomposition pose a interest and lot of lot of theorems and applications of linear algebra depends on singular value decomposition. However, we try to cover it very briefly so, that we are conversant with the definitions that that is the main purpose of this discussion here.

(Refer Slide Time: 25:29)



So, we will go to the third topic which is condition number. For a non-singular matrix condition number is defined as the matrix norm of the condition number of matrix A matrix norm of the matrix A into it is dot product or its mean to the norm of the inverse of the matrix. It is not simply multiplication matrix because norm is a scalar quantity. So, multiplication between 2 scalars one is the matrix norm of the matrix A, another is the matrix norm of the in inverse of A.

A matrix norm we have looked into vector norms. A matrix norm is defined as a vector norm of as defined using a vector norm as matrix norm of A is the ratio of the vector norm of Ax, x being any non-zero vector divided by the norm of the vector norm of the vector x and maximum of that. So, we take any non-zero vector x and multiply with A and the ratio of the norms vector norm Ax by x and find out which one is a maximum of that and we call that as a matrix norm of A. This in a sense can be thought if we have a vector x, how much to which extent like if I have a vector x and when we multiply a vector with a matrix this is again another vector.

So, what is the norm of the product vector divided by the norm of the original vector or how much the vector is being stretched when it is multiplied with that matrix? So, it is kind of a measurement of the stretching which is accomplished by multiplying any vector with the matrix. And what is the maximum amount of stretching when we do multiplication of the matrix? This is that is defined as the matrix norm.

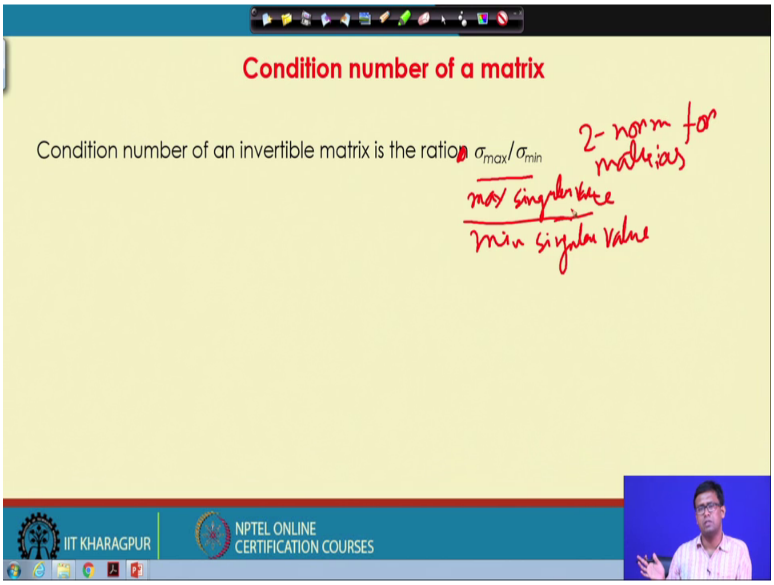
A matrix may have one first norm, second norm P-norm, infinity norm same as the vector norm of the vector. So, what definition are we using on the vector norm that will define that what definition are we will be using for the matrix. Now, if we use l 1 norm for the vector, it is the first norm of the matrix. If we use l 2 norm of the vector is the second norm of for the matrix. If we use l infinity norm that is the largest component of the vector will use the infinite norm for the matrix.

So, it is it will go hand in hand. Condition number of A matrix measures the ratio of maximum relative stretching which is norm of A to minimum relative shrinking that the matrix does to any non-zero vector. Such as a matrix norm of A into matrix norm of A inverse is maximum of Ax by x which is maximum stretching of the vector x by this matrix into minimum of Ax by x inverse 1 by minima Ax by x that is the minimum stretching possible relative shrinking using that particular matrix.

A singular matrix will have infinite condition number why because A inverse singular matrix means the in inverses A has A there is some 0 in A. So, multiplying A vector by a singular matrix will probably result in 0 stretching of that matrix and A inverse will be 1 by 0. And therefore, the matrix will have infinite condition number.

So, interestingly we can say, if A is singular, Ax is equal to b is not solvable. So, if A matrix is infinite condition number, Ax is equal to b is not solvable. A matrix which has very high condition number, it will be difficult to solve that matrix using any matrix solver. Low condition number or condition number close to 1 are designed by the matrix solvers. If the condition number is high, we call them to be ill conditioned matrix and we will we can quickly show it later also as a matter of fact any matrix solver will probably not work well with that with low high condition number matrices.

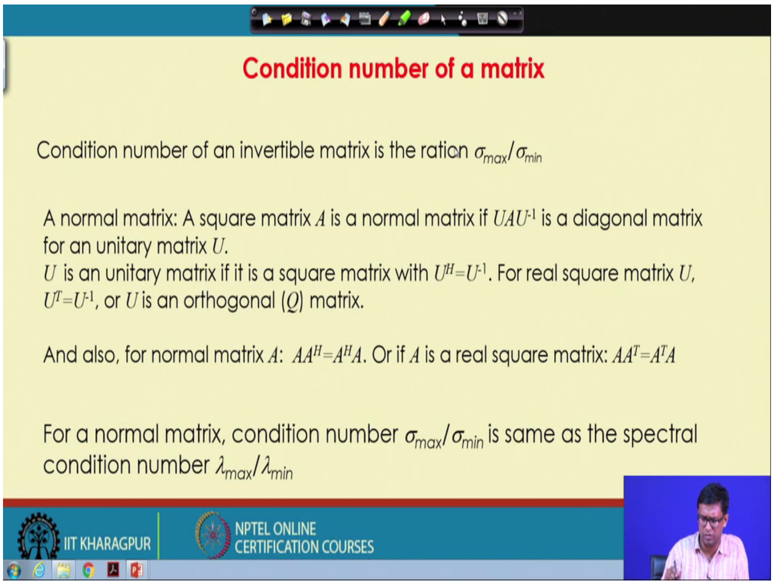
(Refer Slide Time: 30:20)



If we use l 2 norm for the vectors or if you use 2- norm for matrices, if discussed in detail about the norm of A vector. So, if we use l 2-norm that is norm of A vector is square root of the square of all the components of vector. It can be shown that for an invertible matrix, the condition number is same as the ratio sigma max by sigma min or the ratio of the maximum singular value by minimum singular value.

So, that there is one being importance of doing singular value decomposition. If we find maximum singular value and minimum singular value for any matrix A, it is ratio is a good measure of condition number and it is actually the second norm of the if you use second norm for the matrices. It is the condition number which comes to the definition.

(Refer Slide Time: 31:41)

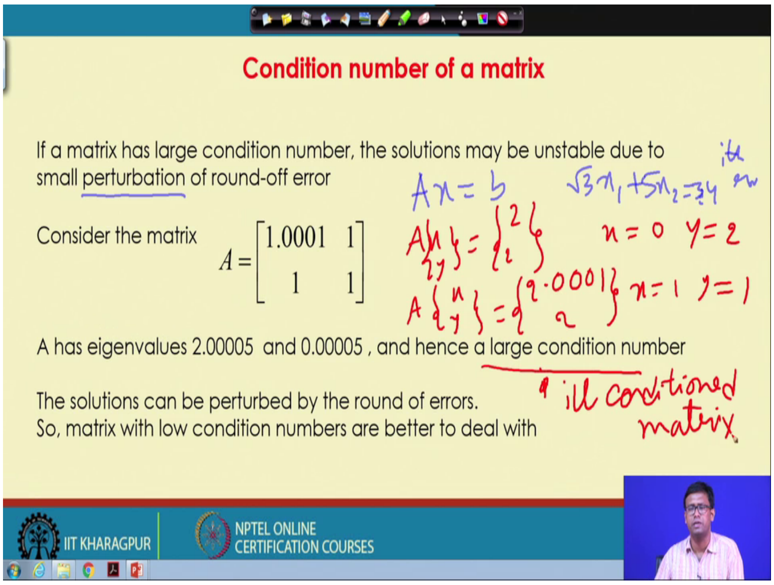


A normal matrix is defined as a square matrix A is called a normal matrix if UA U inverse is a diagonal matrix for any unitary matrix u. So, will trying to define A matrix; normal matrix which needs definition of another type of matrix which is an unitary matrix. And U is an unitary; U is a unitary matrix U is a unitary matrix. If it is a square matrix with U Hermitian is equal to U inverse. For real square matrix U transpose is equal to U inverse or U is an orthogonal Q matrix, then U is called an unitary matrix. So, unitary matrix is a complex equivalent of a orthogonal matrix.

So, if A is x, A if U AU inverse is a diagonal matrix or for real matrix Q A Q inverse Q A Q transpose is the diagonal matrix, then A is called A normal matrix. A is a And if A is a normal matrix if Q A Q transpose is the diagonal matrix A must be a symmetric matrix or if A is a normal matrix A Hermitian is equal to A Hermitian A for real matrix A transpose is equal to A transpose A, then A is a normal matrix.

And if it is so, then the condition number is same as the spectral condition number or the ratio of the maximum eigenvalue and minimum eigenvalue of A. And will very quickly check that what are the importance of having condition number or having high spectral condition number or low spectral condition number of a matrix or high condition number or low conditional number.

(Refer Slide Time: 33:39)



If a matrix has large condition number, the solution may be unstable due to small perturbation or round off error. For example, if I am solving some equation Ax is equal to b and A has something like root 3 x 1 which is ith row plus 5 x 2 is equal to 3.4. Now this root 3 or even 3.4 when we use computer to define it, it never defines the value which is exactly same as root. We discussed it earlier like root 3 is a rational number. So, it will have infinite decimals going on to define exact value of root 3.

However when you use a computer, it truncates it after certain values. So, what we write into as what computer write says, root 2 and what root 3 is very small difference between them. Therefore, if I look into the ith equation when I write it in a computer program, it is not the exact right equation. There are some small values taken out of some of the coefficient or some small values added to some of these coefficients and we call that this is a per perturbed equation. The equation is not the exact equation, there is small perturbation on the equation.

Now, keeping this in mind that we are never solving the exact equation which are writing in pen paper. Whenever translating it into a computer program, whatever be the precision of the computer, whatever be the accuracy of the computer program; there will be small perturbations. Because due to round off error because root 3 will be replaced by some well which is closed very close to root 3, but not exactly root 3; 1 by 3 will be replaced by some value which is 0.3333333 and it will be cut somewhere it is not exactly 1 by 3 very close to 1 by 3 same for all real numbers except probably except integers.

So, the equations are little perturbed and not solving the exact equation. Now, if we think of a matrix with very high consider condition number which is a is equal to 1.001111 and we try to find out the eigenvalues. Eigenvalues are 2.005 and 0.00005. Therefore, it has a large condition number condition number to the order of 2 into 10 to the power 4 into 10 to the power or something.

Now, if I look into the equation say Ax is equal to or A x, y is equal to say 2 2 and we solve it we will get x is equal to 0 y is equal to 2. And if we have another equation A x y is equal to 2.00012, we solve it we will get x is equal to 1, y is equal to 1. So, for very small difference in one particular place, the equations are giving entirely different results.

And this and why they are giving entirely different result because it is a large condition number matrix. So, if we have some small perturbation instead of 2.0001, if it is 2.00005; we can see a very different solution. So, with very small changes in the matrix or in the b vector, the equations change to a great extent and they are called an unstable equation system or the matrix solver operating on that will be unstable. Though the solution in pen paper can be obtained very easily, but the solutions are heavily perturbed by the round off error and this happens for matrices with large condition numbers. And we call them to be ill conditioned matrix that sense matrices with low condition numbers are better to deal with.

So, condition number or finding out singular value and finding out the ratios of singular value and finding condition number is an important task when looking for matrix solvers. And later we will see lot of, we will discuss we about few more matrix solvers in subsequent classes and we will see the lot of matrix solvers actually do not work for matrices where the condition number is bad or works very inefficiently for very high in condition matrices or high condition number matrices. These are a few important concepts; positive definiteness, condition number, singular values which will appear in the next course, next classes with you at certain locations. And we have to utilize these concepts to understand the applications in where the matrix solvers are involved.

Thank you.