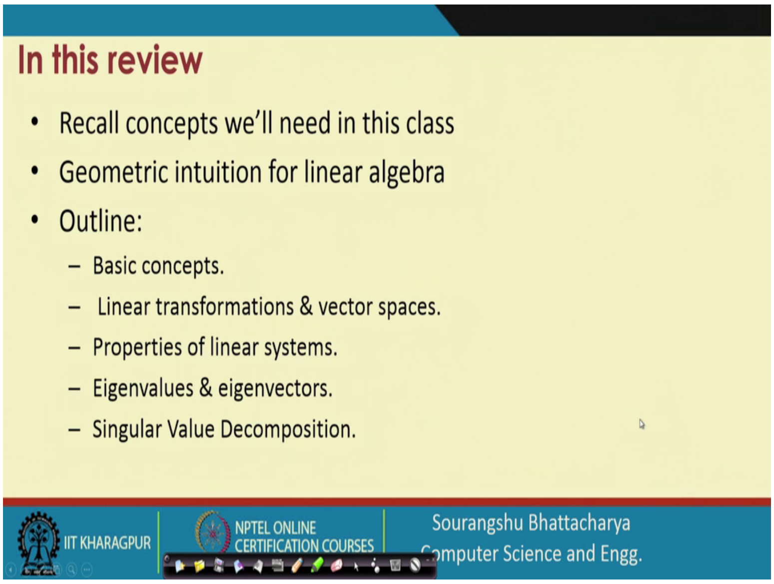
Scalable Data Science

Prof. Sourangshu Bhattacharya
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur

Lecture - 03
Background on Linear Algebra

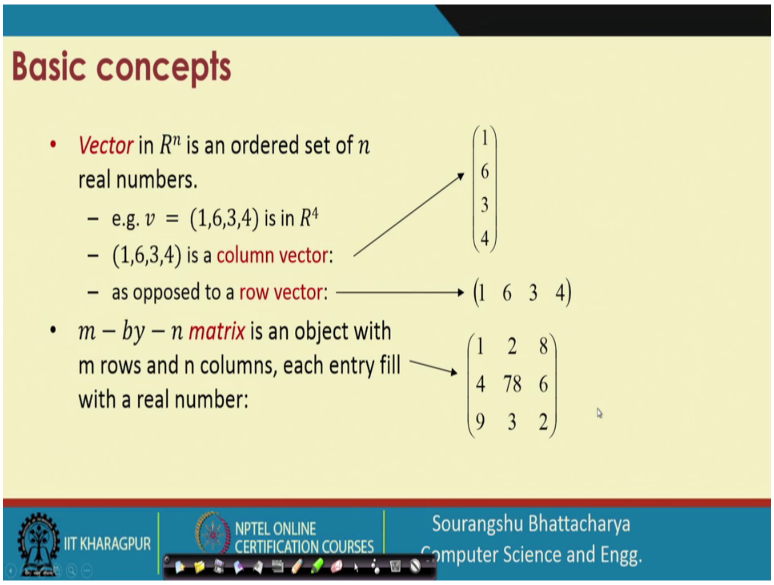
Welcome to the NPTEL course on Scalable Data Science, lecture number 3. Today's lecture is on background on linear algebra which will be needed for the course. I am Professor Sourangshu Bhattacharya of Department of Computer Science and Engineering, IIT, Kharagpur.

(Refer Slide Time: 00:36)



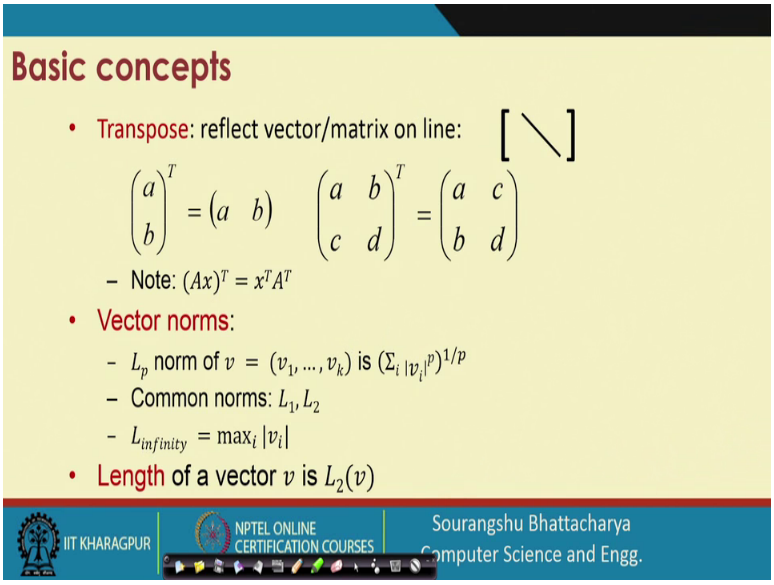
In this course, in this lecture we will recall the concepts that we need in this class. We will especially go into the geometry intuition for linear algebra. We will outline some basic concepts and then we will discuss linear transformations and vector spaces. Then we will discuss properties of linear systems, then we will discuss eigenvalues and eigenvectors of matrices and finally, we will disc briefly discuss singular value decomposition.

(Refer Slide Time: 01:12)



A vector in R n is an ordered set of n real numbers. So, for example, this is a vector in R 4 and this is a column vector, this is a row vector again in R to the power 4, m by n matrix is an object with n rows and or m rows and n columns. So, this particular matrix is a 3 by 3 matrix.

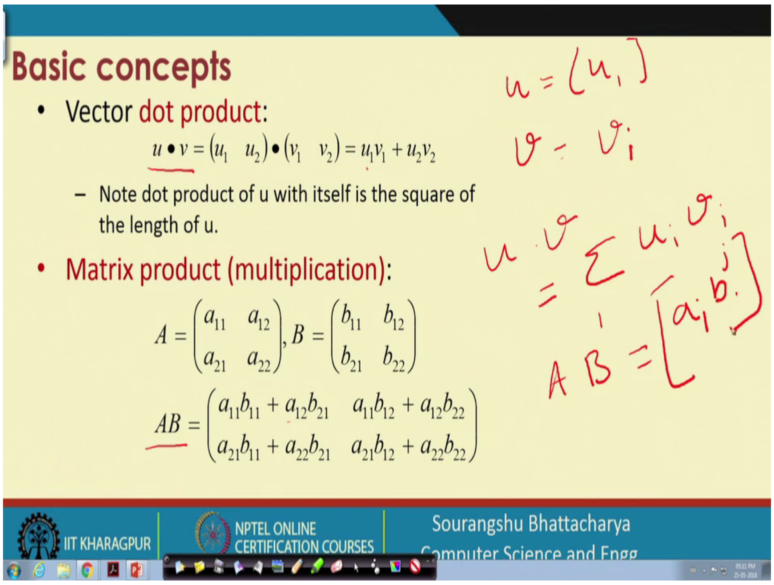
(Refer Slide Time: 01:56)



Now, transpose of a vector or a matrix is a reflection of the vector or the matrix along the diagonal line which is this particular line of the matrix. For example, the matrix a b transpose the column matrix a b transpose or the column vector a b transpose becomes the row vector a b whereas, the column, the matrix a b c d transpose becomes the matrix a c b d, ok.

We note that we can have a matrix vector product which is A x the transpose of which becomes x transpose a transpose. Also note the definition for the vector norms. The L p norm of a vector v with components v 1 till v k is denoted by summation over i, mod of v i to the power p whole raise to the power 1 by p. We will commonly use the L 1 and the L 2 norms. The L infinity norm of a matrix is the max over all the components the absolute value of all the components of a vector. Also note that the length of a vector v is the L 2 norm of the particular vector.

(Refer Slide Time: 03:51)

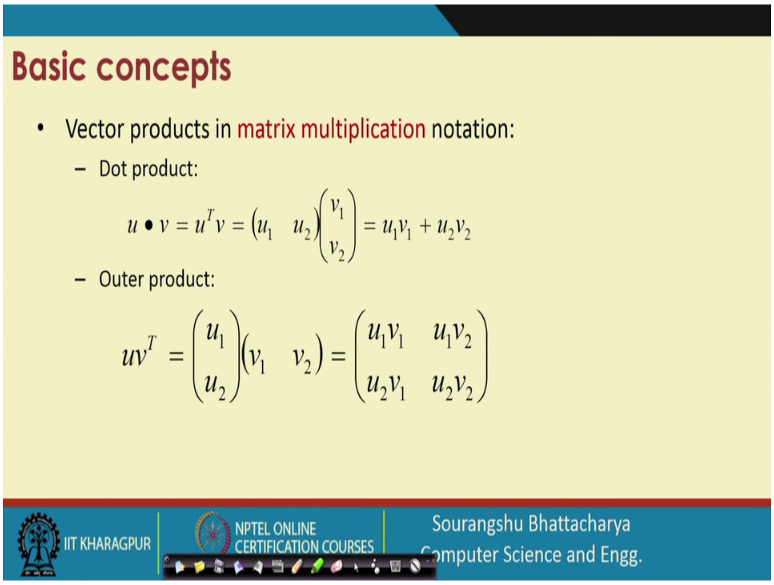


Next we go into the dot products. First we define the dot products between 2 vectors u and v with components u 1, u 2 and v 1, v 2 as sum over u 1, u v 1 plus u 2, v 2.

So, in general if you have a vector u with components u i and another vector v with components v i, then u dot v is equal to sum over i u i times v i. Going forward we can also define the matrix product as given in this example. So, if a matrix A has a 11, a 12, a 21 and a 22 and similarly the matrix B then the entries of the matrix AB are of this form a 11 times b 11 plus a 12 times b 12 and so on and so forth.

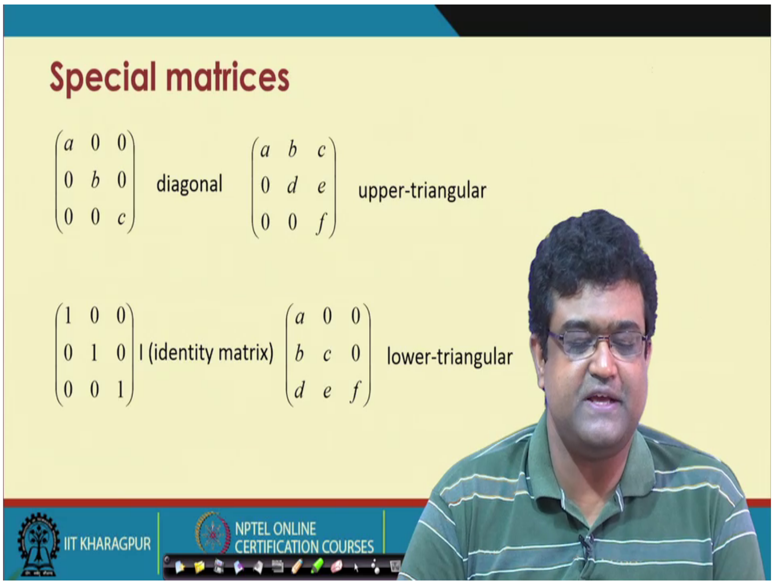
In other words the product the entries of the matrix A times B are so the i jth entry is the ith row of a matrix times jth column of the matrix B ok.

(Refer Slide Time: 06:00)



We can also define the dot product of two matrices we can represent the 2 vectors in terms of column matrices or column vectors. And then we can represent the dot product of 2 vectors u and v in terms of the matrix product u transpose v which is the same. Similarly we can define the outer product between the 2 vectors as given here.

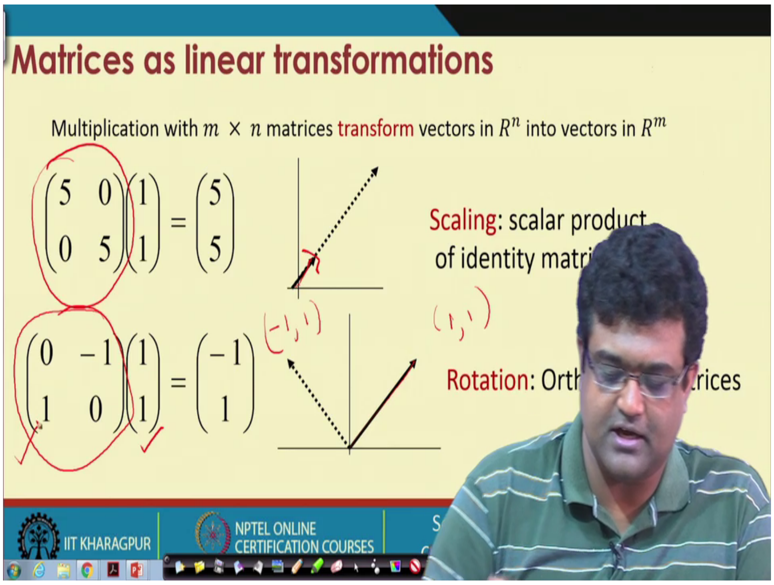
(Refer Slide Time: 06:43)



There are some special matrices which we will come across in our discussions; the first one is the diagonal matrix which is given here. As you can see only the entries on the diagonal have values all the off diagonal entries are 0. Similarly we can define an upper triangular and a lower triangular matrix, in the upper triangular matrix only the entries upwards of the diagonal are nonzero all the entries below the low below the diagonal are nonzero and vice versa.

Another important matrix is the identity matrix which is just the matrix of all 0 entries and once only in the diagonal positions.

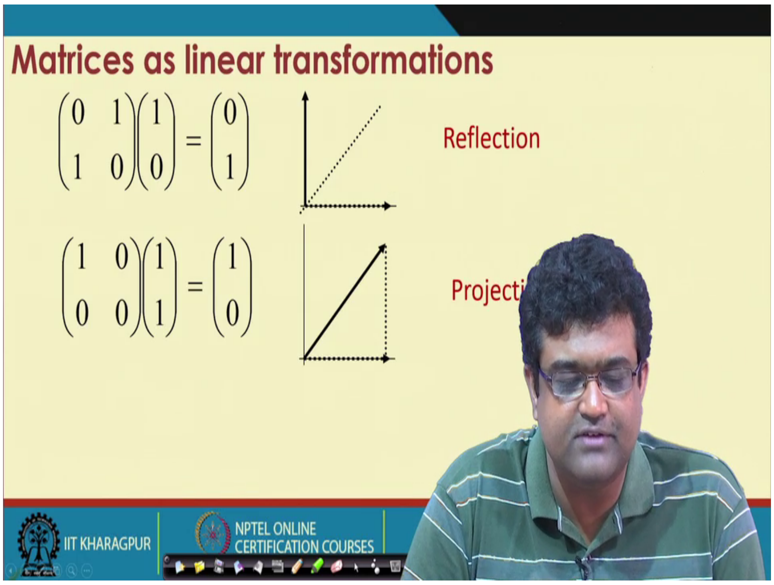
(Refer Slide Time: 07:43)



Next we discuss matrices as linear transformations. The first transformation that we discussed is the scaling transformation. You see that the matrix a here transforms this vector 1 1 which is the vector like this as 5, 5 which is just a scaled version of the original vector, hence this transformation is called the scaling transformation.

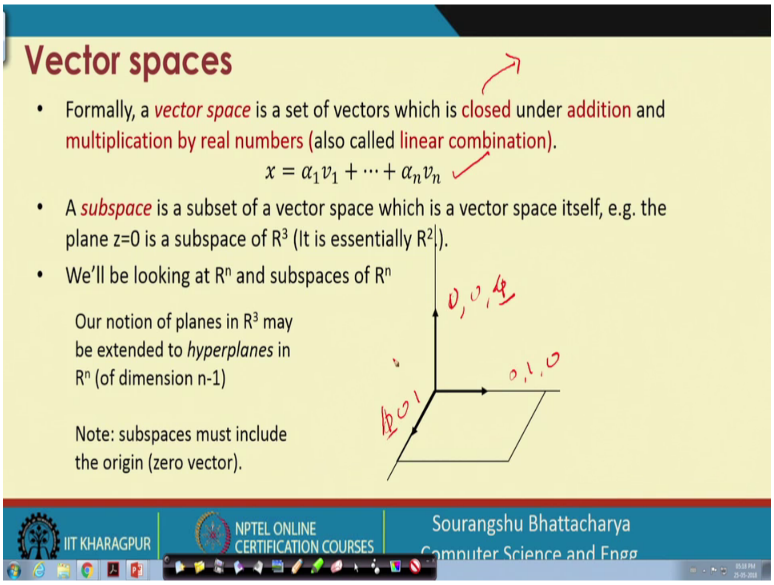
Next we discuss the rotation transformations. Note that the matrix 0, minus 1, 1, 0 rotates this vector. So, when matrices when multiplied with this vector 1 1 produces the vector minus 1 1, this is the vector 1 1 and this is the vector minus 1 1. Note that the length of both the vectors are same however, this vector has been rotated by 90 degrees. So, this matrix is called a rotation matrix.

(Refer Slide Time: 09:16)



Similarly, we can define a reflection matrix which takes the vector 1 0 which takes the vector 1 0 and transforms it to the vector 0 1. In other words it is performing a reflection about the line for about the 45 degree line another matrix is the projection matrix which takes for example, this vector 11 and projects it on to the x axis. Note that only the there is only there is a one in the first entry and all the other entries are 0. Such a matrix would take any vector and project it on to the first dimension.

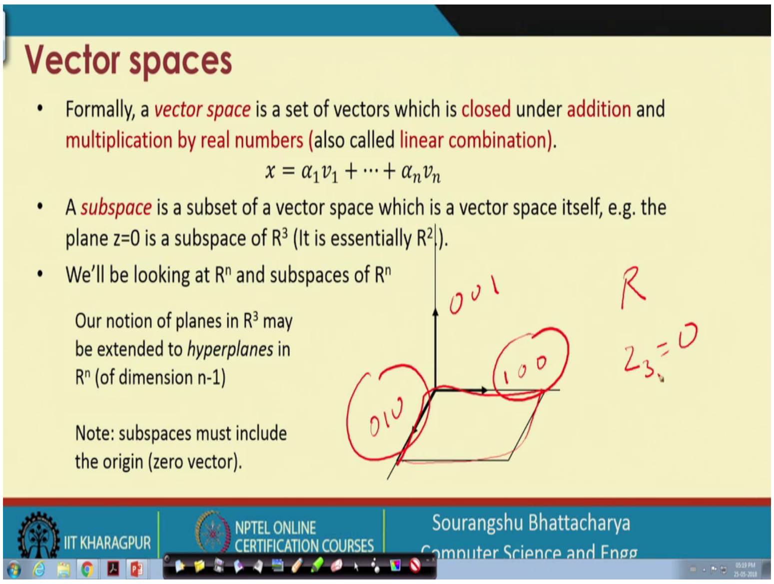
(Refer Slide Time: 10:27)



Next we describe an important concept of vector spaces formally vector spaces are a set of vectors which are closed under addition and multiplication by real numbers. This is also sometimes called a linear combination of the vectors. So, for example, here we have shown that the vector x is a linear combination of vector v 1 till v n with where v 1 till v n are vectors and alpha 1 till alpha n are the corresponding real numbers which they are multiplied by.

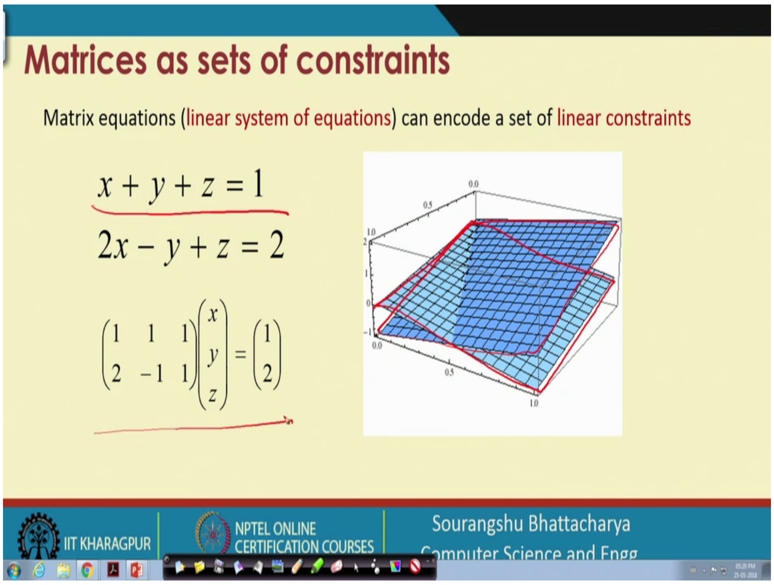
Closure means that any element in this set can be generated as a linear combination of elements other elements in the set. A sub space is a subset of a vector space which is a vector space in itself. For example, you can see here that this coordinate axis denotes the 3 dimensional vector space. So, for example, if you take the vectors 1 0 0 0 1 0 and 0 0 1, as the may be I will make it 0 1 0.

(Refer Slide Time: 12:14)



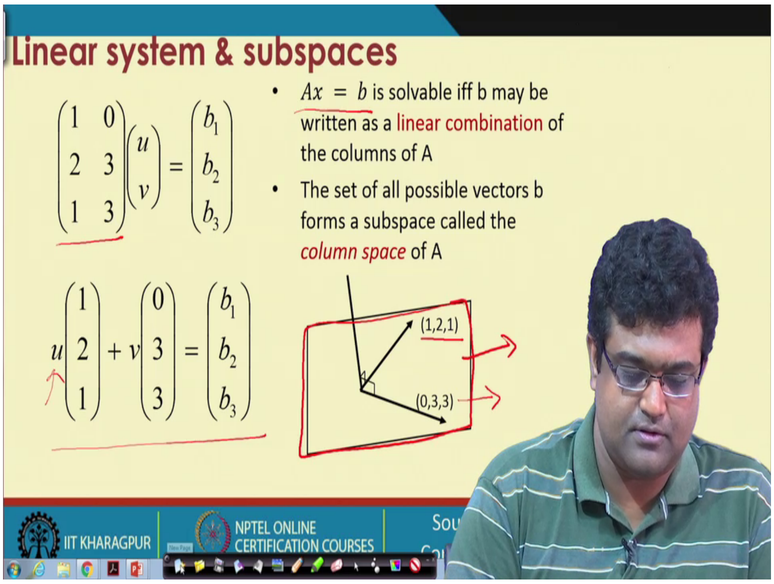
So, if we take the vectors 1 0 0 0 1 0 and 0 0 1 this can generate all the vectors in R 3. A linear combination of these 3 vectors can generate all the vectors in R 3 however, if we set z 3 equal to 0 then this and this are sufficient to generate this subspace for which z equals 0. We can also extend the notion of planes to n dimensional sub spaces using hyper planes which we will discuss next.

(Refer Slide Time: 13:13)



This is linear equations, so matrix equations or systems of linear equations can also encode a set of linear constraints. So, for example, the first equation encodes a linear constraints which is denoted by this hyper plane ok. And the second one encodes another set another linear constraints which is denoted by the second hyper plane ok, together they can be written as a system of linear equations in matrix notation as given below.

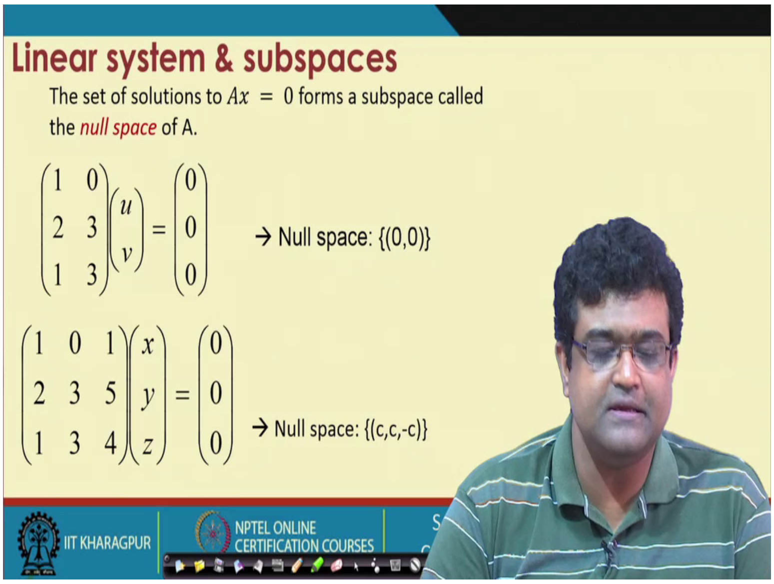
(Refer Slide Time: 14:13)



Now, a system of linear equations can be generally written or compactly written in matrix notation as A x equals b where a is the coefficient matrix which is in this case this matrix, x is a column vector and b is another column vector. If we want to solve this system of equation A x equals b in terms of x it is solvable only if the vector b can be written as the linear combination of columns of A. So, for example, in this case the columns of a matrix are 1 2 1 and 0 3 3. So, only those vectors, which can be written in this form for u times 1 2 1 plus v times 0 3 3 equals b; therefore, the sub space given by this hyper plane.

The set of possible vectors generated by generated by a linear combination of columns of a matrix is called column space of the matrix A.

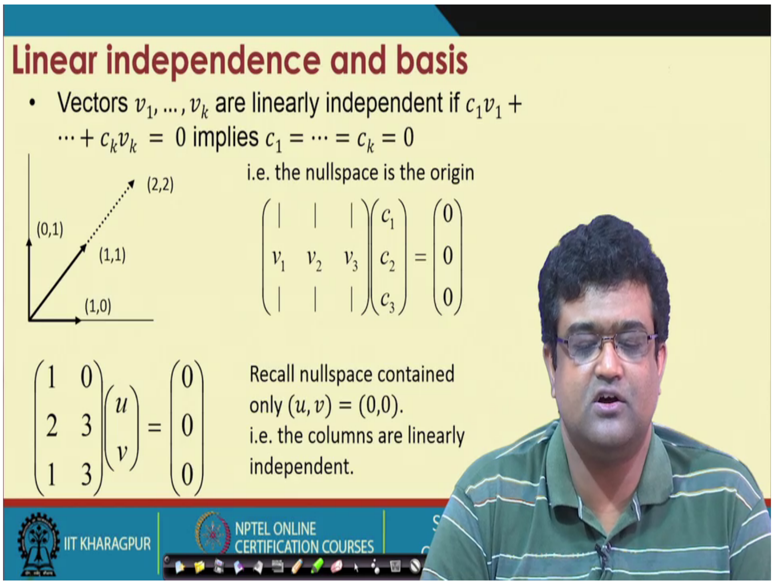
(Refer Slide Time: 16:01)



Another important concept is the null space of a matrix A. The null space of a matrix A is all those vectors which are solutions of equation A x equals 0 or A x is equal to 0 vector. So, for example, 0 0 is a part of null space is always a part of null space or the origin is always the part of null space because any vector multiplied by 0 will give 0.

You can check that for this coefficient matrix all vectors of the form c c and minus c will actually generate the 0 vector, hence this set of all vectors of this form c c and minus c, where c can vary c can be any number constitute the null space of this particular matrix.

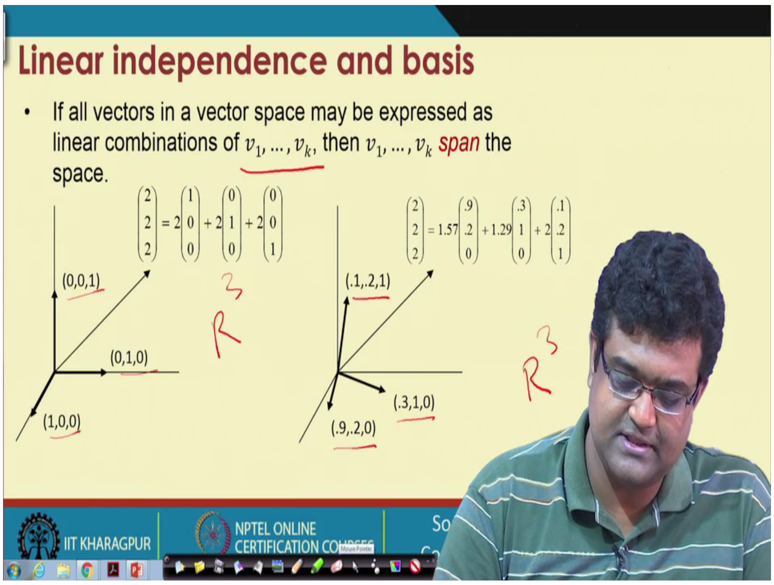
(Refer Slide Time: 17:17)



Next we discuss the concept of linear independence, a set of vectors v 1 to v k are linearly independent if a linear combination of this vectors equals to 0 implies that the coefficients are all 0. In other words the set of vectors are linearly independent if their null space is the origin.

Note that the origin is always content in the null space.

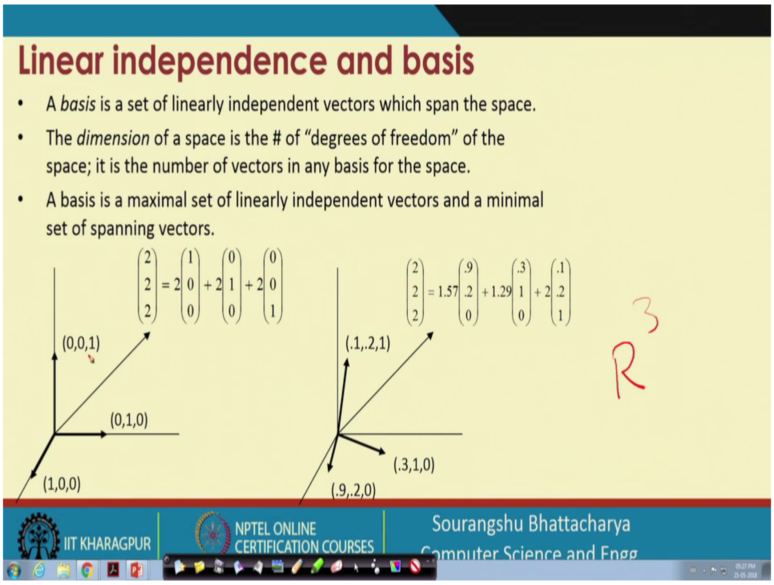
(Refer Slide Time: 18:08)



If all vectors in a vector space can be expressed as linear combination of a set of vectors say v 1 to v k then the vectors v 1 to v k are said to span the entire vector space. So, as discussed this 3 vectors span the entire vector space of R 3.

Similarly, you can see that this, this and this vector together also span the vector space of R 3 and any vector can be generated as a linear. For example, the vector 2 2 2 in this case can be generated as a linear combination of this vector.

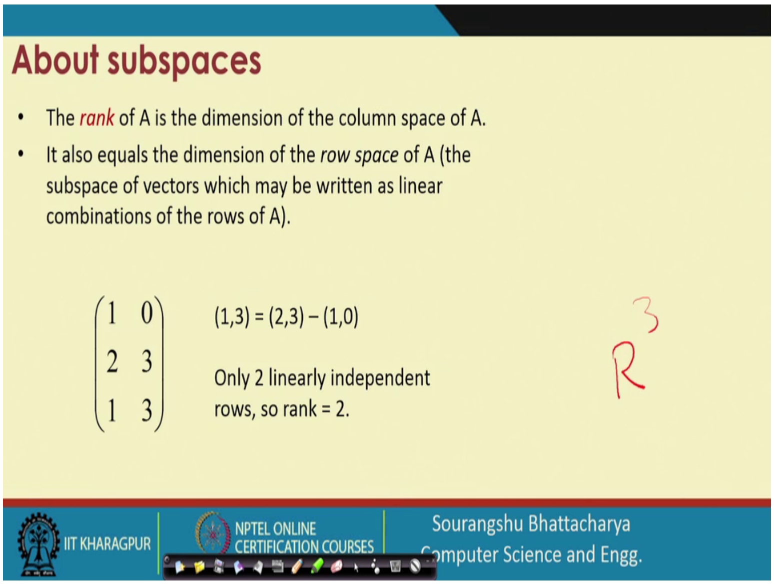
(Refer Slide Time: 19:11)



A basis of a vector space is a set of linearly independent vectors which span the vector space. The dimension of a vector space is the number of degrees of freedom of a vector space or it is the maximal set of linearly independent vectors which can generate the vector space or alternately it is the minimal set of spanning vectors of the vector space this two are equivalent.

So, the vector space R 3 has dimension 3 as we know that 3 vectors are sufficient to generate all the points in R 3 and more than 3 vectors are not necessary, rather less than 3 vectors are not sufficient to generate all the points in R 3.

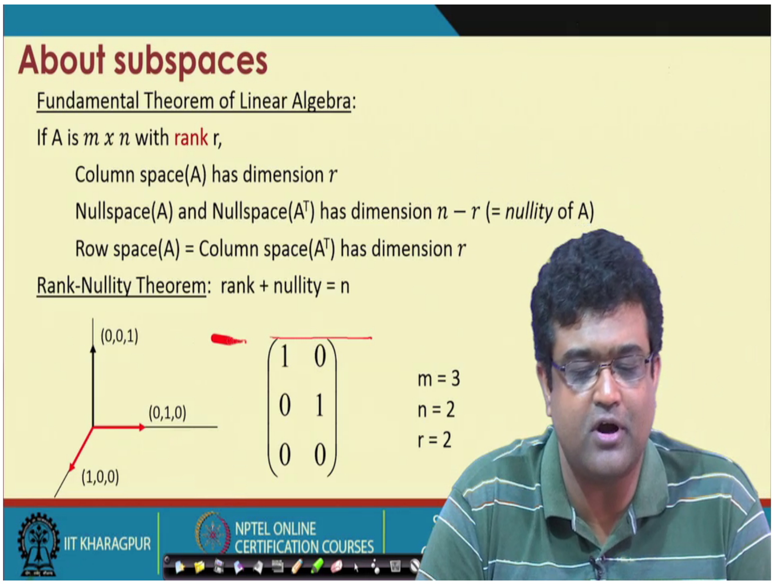
(Refer Slide Time: 20:37)



Next we discuss the concept of rank of a vector space, rank of a matrix sorry the rank of a matrix is the dimension of the column space of the matrix. For example, this matrix has a rank 2 because it has the dimension of the space generated by its columns is 2. In this case, note that it also has 2 columns and both the columns are linearly independent.

You may note that this also equal to the dimension of the row space. So, for example, in this case even though there are 3 rows you can check that one of the rows can be generated by a linear combination of the other two rows. For example, 1, 1 and 3 can be generated as a linear combination of 2 and 3 in this manner.

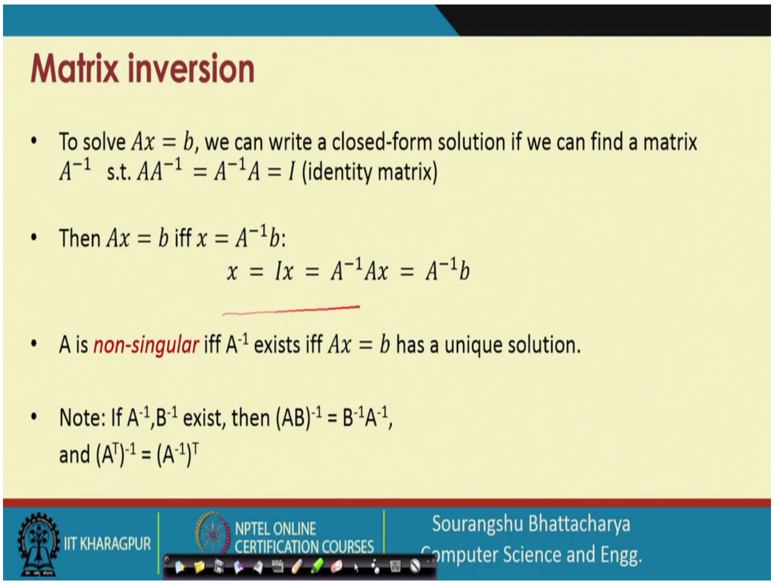
(Refer Slide Time: 22:12)



Next we discuss an important result in linear algebra which is call also called the fundamental theorem of linear algebra which is given a matrix A let us say it is m by n matrix with rank R, column space of a has dimension R. This is coming from the definition of rank which says the rank of a matrix is the dimension of the column space of a matrix. Also the null space of a or the null space of a transpose has dimension n minus R which is also sometimes called the nullity of A, ok.

Hence and also the since the dimension of row space and the dimension of the column space are same hence the dimension of the row space or which is also the dimension of the column space of A transpose is also R hence the important result that given the n dimensional space the rank plus nullity of the matrix A is equal to n.

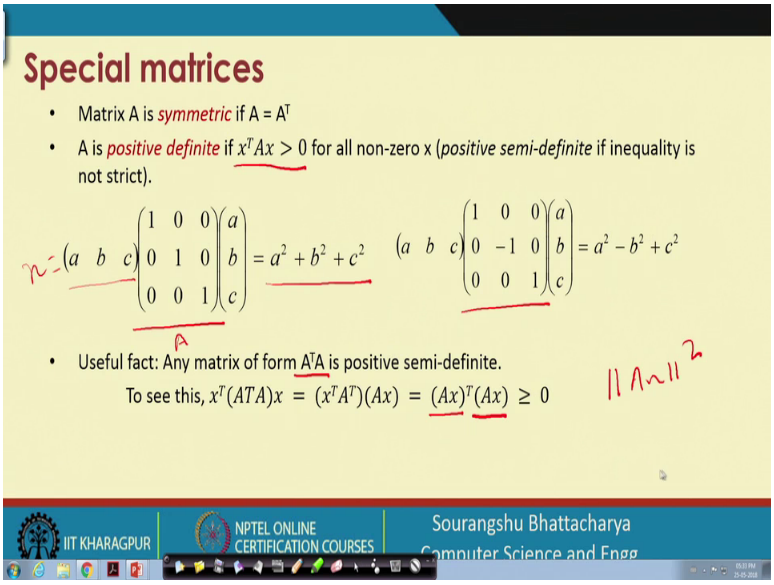
(Refer Slide Time: 23:48)



Next we discuss the concept of inverse of a matrix. In order to solve the system of linear equations A x times b, A times x equals b we can write a closed form solution if we can compute a matrix called A inverse which has the property that A times A inverse equals A inverse times A equals identity. Then we can simply write x as A inverse times b. We can verify that this is indeed true because you can write identity here you can write x has A x and then you can write the identity as A inverse A and then you can write A x as b.

The matrix A is called non singular matrix if and only if such matrix A inverse exist and if that happens the system of linear equations A x equals b has a unique solution.

(Refer Slide Time: 25:30)

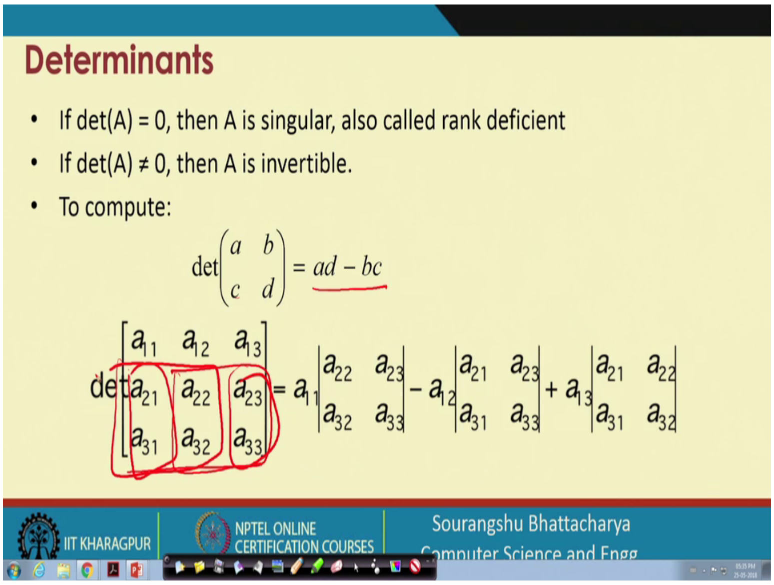


Some special definitions a matrix A is called symmetric of A equals A transpose. Furthermore, a matrix A is called positive definite if for any vector x x transpose A x is greater than 0 and if the inequality is not strict then it is called positive semi definite.

So, let us see a common positive semi definite matrix which is the identity matrix. Note that if we have any vector a b c then if we call this vector as x then x transpose A x is nothing but a square plus b square plus c square which is positive for all values of a b and c. Hence all vectors x, hence this matrix is positive semi definite whereas is not necessarily a positive semi definite matrix, ok.

A very useful fact is any matrix of the form A transpose A is always positive semi definite because you can write x transpose A transpose A x as A x transpose A x which is nothing but the norm of A x or rather norm of A x square L 2 norm of A x square.

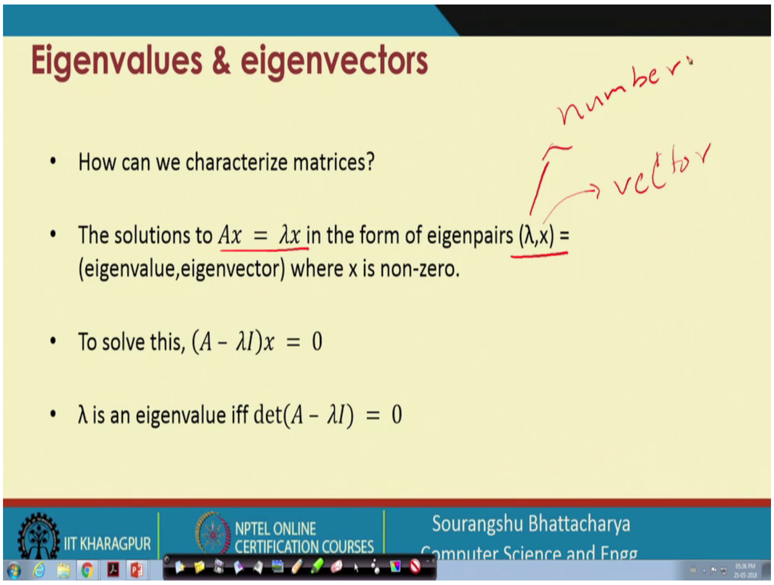
(Refer Slide Time: 27:30)



We also discuss the notion of determinant. Determinants are numbers which are calculated from a matrix for a 2 by 2 matrix of with entries a b c d the determinant value is given by ad times dc, ad minus bc sorry. And for a 3 cross 3 matrix the determinant is given by eh is computed recursively as a 11 times determinant of this sub matrix minus a 12 times determinant of the sub matrix formed by taking these two rows and plus a 13 times the determinant of this sub matrix.

For higher order matrices the determinants can be defined similarly. And important thing to note is that if a matrix A is singular or rank deficient if and only if its determinants is 0, if its determinants is not 0 then the matrix is invertible.

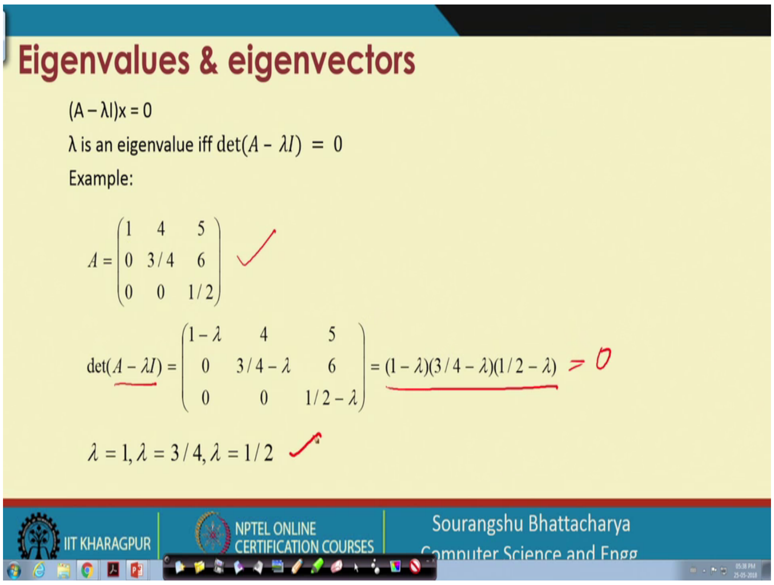
(Refer Slide Time: 29:02)



Next we discuss the important concept of eigenvalues and eigenvectors. So, the question comes how can we characterize a set of matrices. One way to characterize the set of matrices is to find solutions of equation of this form A x equals lambda x, for all x ok. So, all alls all pairs of here note here that here lambda is a value and x is a vector. So, lambda is a number and x is a vector, ok.

So, of course, if x is equal to 0 and lambda is equal to 0 this equation is trivially satisfied. So, we are seeking solutions which are nonzero or where x is both x and lambda are nonzero,. One way to solve this is to look for solutions to this linear system of equations which is possible only if the determinant of this coefficient matrix is 0, ok.

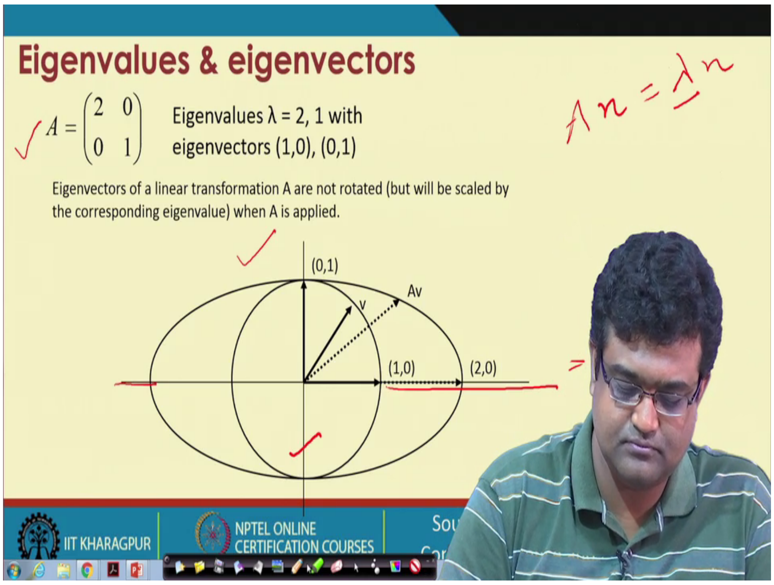
(Refer Slide Time: 30:49)



Using this concept we can solve the eigenvalue problem for example, given this particular matrix we can compute or given this matrix a we can compute the matrix A minus lambda I. And then we can compute the determinant of this matrix which in this case turns out to be this particular number. In general we will have a polynomial in lambda we can solve this equal to 0 to get the get eh the degree of the polynomial times values. And in case of n cross n matrix this turns out to be n the degree of this polynomial or which is also called the characteristic polynomial turns out to be n.

By solving this we can get the lambda after which we can solve for x.

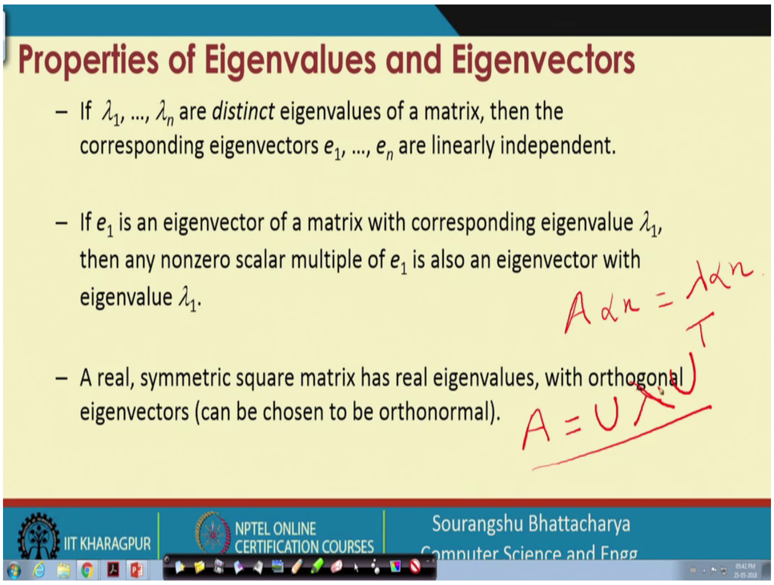
(Refer Slide Time: 32:00)



Now, the significance of eigenvalues and eigenvectors is that the eigenvalues and eigenvectors scale the the or or the eigenvectors are the directions in which any vector when transform by this matrix A is just rescaled or scaled by the eigenvalue. This is because as you can see for the eigenvectors direction let say a your A x become n a becomes nothing but a value lambda times x.

This is the interpretation which is shown here. So, for example, in this case this matrix scales any vector in this direction, as in this direction by 1 and any vector in this direction by 2, ok. Hence this is a eigenvector and this is a another eigenvector and the corresponding eigenvalue for this eigenvector is 2 whereas, the corresponding eigenvalue for this eigenvector is 1.

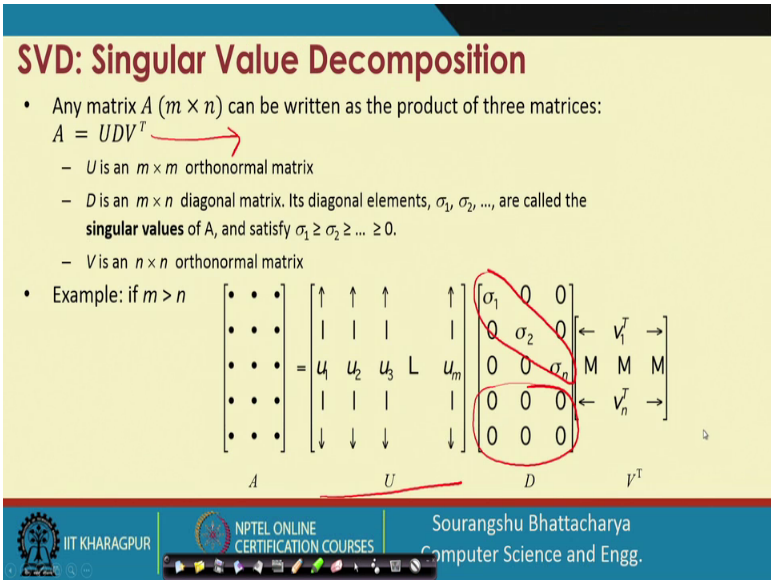
(Refer Slide Time: 33:34)



Now, eigenvalues and eigenvectors have very interesting properties. So, the first property is if lambda 1 till lambda n are distinct eigenvalues of a matrix then the corresponding eigenvectors e 1 till e n are linearly independent. Next if e 1 is a eigenvector of a matrix with corresponding eigenvalue n then any nonzero scalar multiple of e 1 is also an eigenvector with eigenvalue lambda 1. This is very clear because you can check that A of let say alpha times x is also equals lambda of alpha times x.

The third property is also very useful any real symmetric square matrix has real eigenvalue with orthogonal eigenvectors which can also be chosen to be orthonormal. So, in other words your set if eigenvectors forms a orthonormal matrix and the set of eigenvalues are real and symmetry. In fact, if you calculate all the eigenvectors you can write the matrix A as a matrix of eigenvectors u times lambda times u transpose. So, this is also called the eigenvalue decomposition of a matrix, and where u is the matrix of all eigenvectors and lambda is diagonal matrix of all eigenvalues.

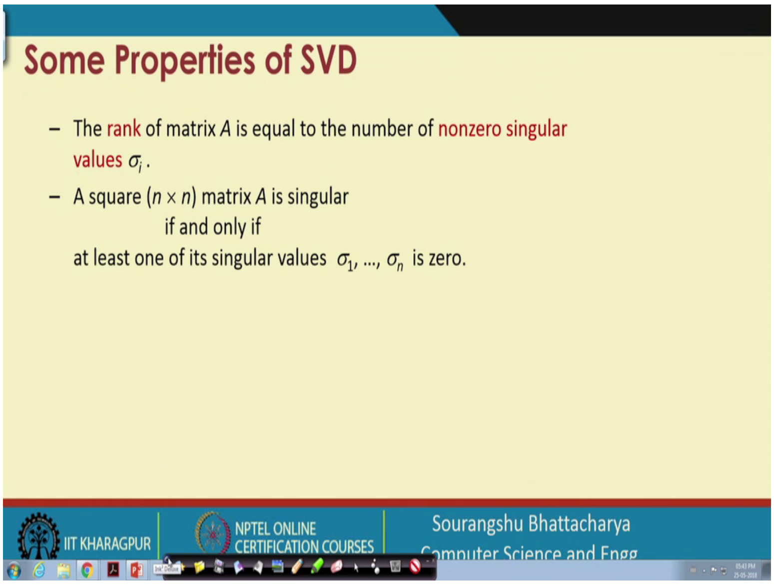
(Refer Slide Time: 35:58)



Extending this is another concept called the singular value decomposition. So, given any matrix a which is an m cross n matrix it can be written as a product of 3 matrices U which is an m cross m orthogonal matrix which or orthonormal matrix D which is an m cross n diagonal matrix with. So, this is the matrix U which is an m cross m orthogonal matrix or orthonormal matrix the matrix D is m cross n diagonal matrix. So, in this case let say if m is greater than n. So, it will only have n many diagonal values rest all the values will be 0 and v which is another n cross n orthonormal matrix.

This decomposition or this way of writing a matrix A is called the singular value decomposition.

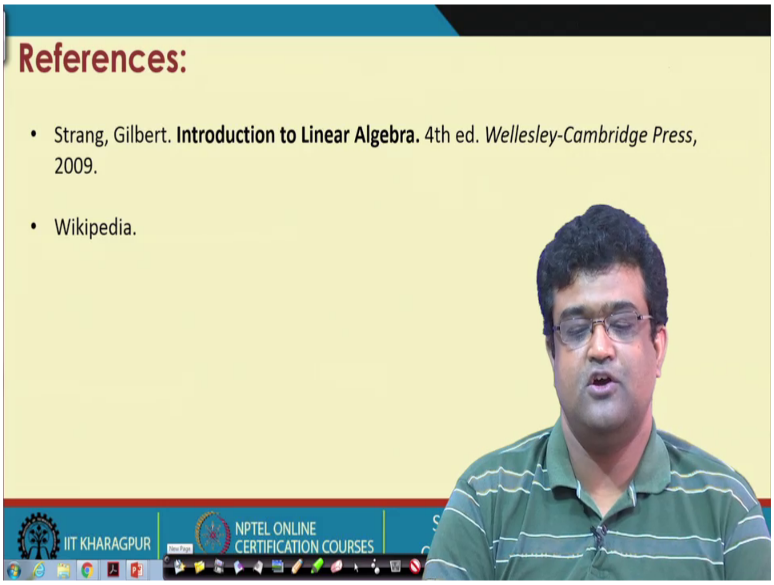
(Refer Slide Time: 37:26)



The singular value decomposition has some very in important properties, so very important and useful property. So, for example, the rank of matrix A is equal to the number of nonzero singular values of the matrix A. So, every matrix may not have eigenvalue decomposition in real numbers but every matrix will have a singular value decomposition and once you calculate the singular value decomposition, the number of nonzero singular value determines the rank of a matrix.

For a square matrix A you can see that it will have for a n cross n square matrix A it will have sigma 1 till sigma n many singular values and it will be non singular it will be non singular only if none of the singular values sigma 1 till sigma n is 0. In a other words the matrix will be singular if at least one of the singular value sigma 1 till sigma n is 0.

(Refer Slide Time: 38:48)



So, with this we come to an end of this introduction to linear algebra. The reference for this a good reference for this is the introduction to linear algebra book by Gilbert Strang and another very nice reference is the Wikipedia.