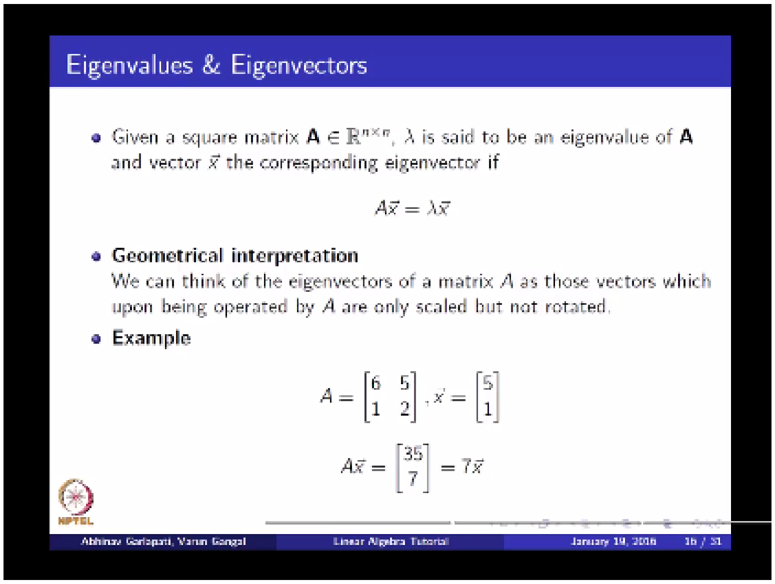
NPTEL
NPTEL ONLINE CERTIFICATION COURSE

Introduction to Machine Learning

Linear Algebra-2

(Refer Slide Time: 00:15)

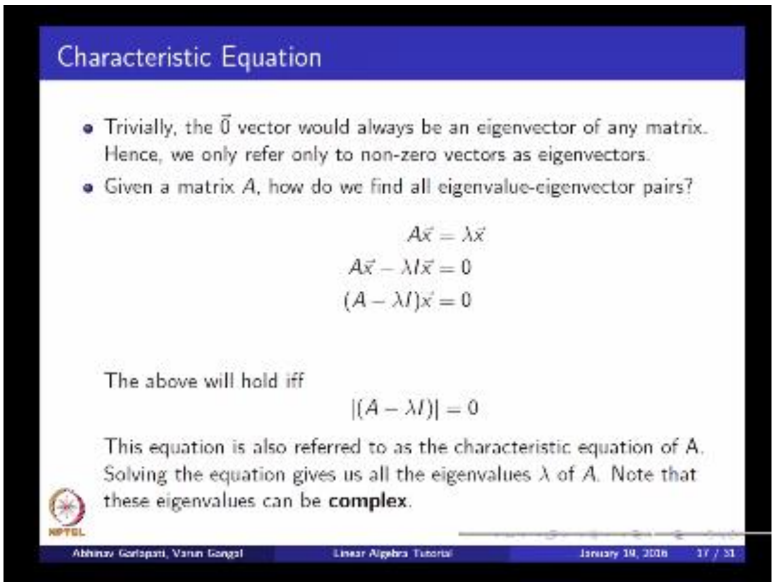


The eigenvector and eigenvalue of A spectral wave. So note here then eigenvectors and eigenvalues are tied together which means that any eigenvector has an associated eigenvalue. You often characterize square meters as head down of their eigenvector one way of looking at eigenvector is as follows.

X can be count out for the vector in R for n and the square meter is A acts like an operator which transforms X into another N dimensional vector AX. Now the eigenvectors of A are those vectors which are being transformed by A or operated upon by A are only scaled by λ but not rotated. In other words that direction does not change.

We can have a look at this example here, the 2x2 matrix A on multiplying the vector X/1 clicks back the vector X multiplied by the real value 7. So here X is an eigenvector of A and 7 is an eigenvalue of A.

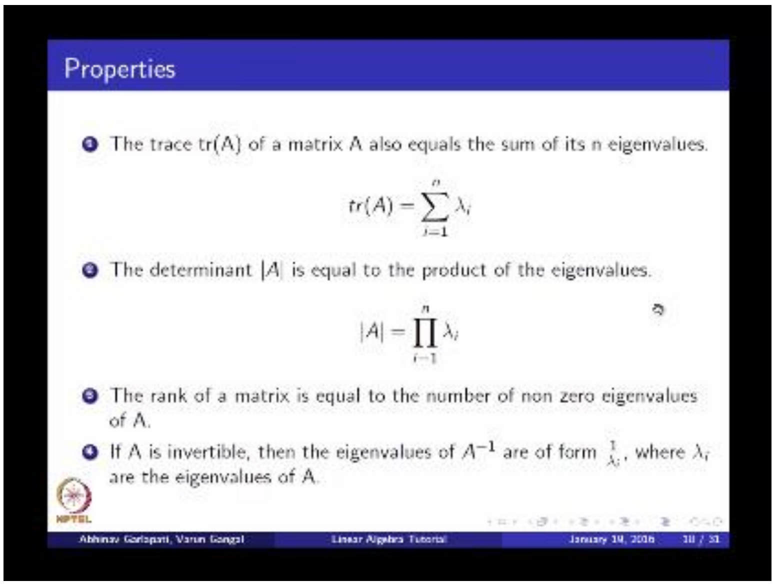
(Refer Slide Time: 01:43)



We can see that 0 would always be an eigenvector of any matrix it reasonably go by the AX= λX definition. Hence we only refer to nonzero vectors and eigenvectors. So the question is given a matrix A out as 1 mild all the eigenvalue, eigenvector bits, by simplifying a sequence λx we get A- λi to X=0.

Now since we are only looking at nonzero vectors det of x cannot be zero and x can be a zero vector which means that det of A- λi should be zero. So the equation det A- λi=0 is called a characteristic equation of A. So one designation gives this all the eigenvalues of A, one thing you should notice that even though all the value of A are real, A is a real matrix, the eigenvalues can complex.

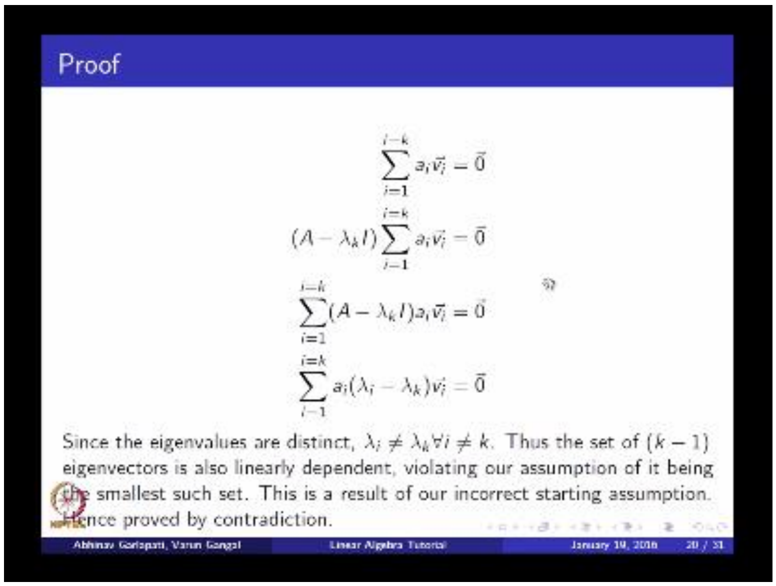
(Refer Slide Time: 02:55)



There are interesting relations between some properties of a matrix and its eigenvalues. For instance, the trace of a matrix is equal to the sum of its eigenvalue by the determinant is equal to the product. The rank of a matrix is equal to the number of nonzero eigenvalue. Note that if a eigenvalue has multiplicity greater than 1.

For instance, if two distinct eigenvectors X1 and X2 both have eigenvalue λ we would count λ twice. Also we can describe the eigenvalues of A^{-1} in terms of the eigenvalues of A provided of course, A is invertible. The eigenvalues of A^{-1} maybe of the form 1/λi, where λi is an eigenvalue of A.

(Refer Slide Time: 03:55)



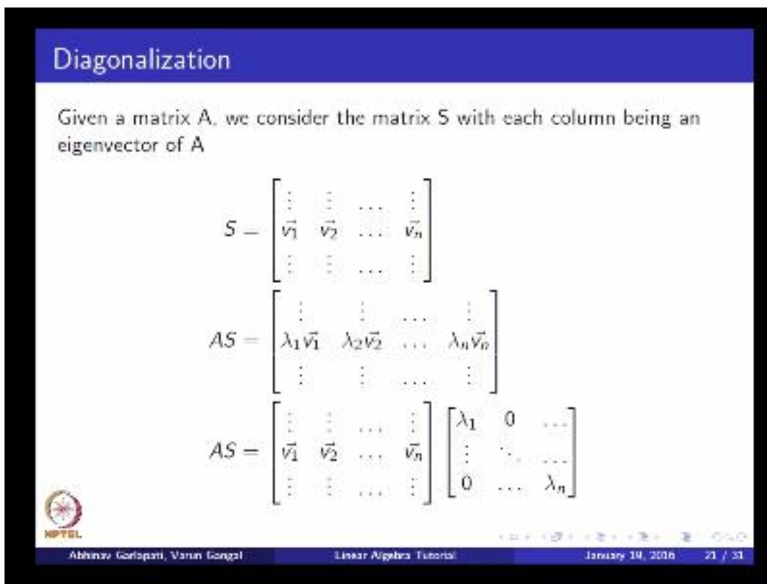
Now let us have a look at an interesting theorem about eigenvalues and eigenvectors. The theorem goes as follows. And for matrix as all its eigenvalue is distinct when its eigenvectors are linearly independent which is proof is by what is called proof by contradiction. This theorem does not hold that means there is a set of A eigenvectors such that it is linearly dependent.

That ith vector in the set BBI and the corresponding eigenvalue will have the i. Note that we are considering the smallest such set. Since the set is linearly dependent this means there exists real consonants Ai such set summation Ai Vi=0. Now let us multiply both sides of the equation by A-λk(i). Since Vk is an eigenvector of A, A-λk(i)Vk will be equal to 0, we can understand this from the characteristic equation.

Hence the term corresponding to Vk disappears from the equation since it goes to zero. Now for the remaining eigenvalues since we know they are distinct the term λi-λk cannot be equal to 0. Note that A-λkixVi simplifies to λi-λk(Vi) since AVi=IVI. For we would now, we can think of Ai(λi-λk) as a new constant Vi this means now that we have a summation running from I=1toi=k-1 such that Vi Vi=0.

However, we can assume that this is the, that the sake of size k was the smallest set of linearly dependent eigenvector. However, now we have an even smaller set, this contradictions starting assumption. Hence, such a set of k linearly dependent eigenvectors cannot exist for any k greater than equal to2. Hence all are eigenvectors are linearly independent, hence our theorem stand to.

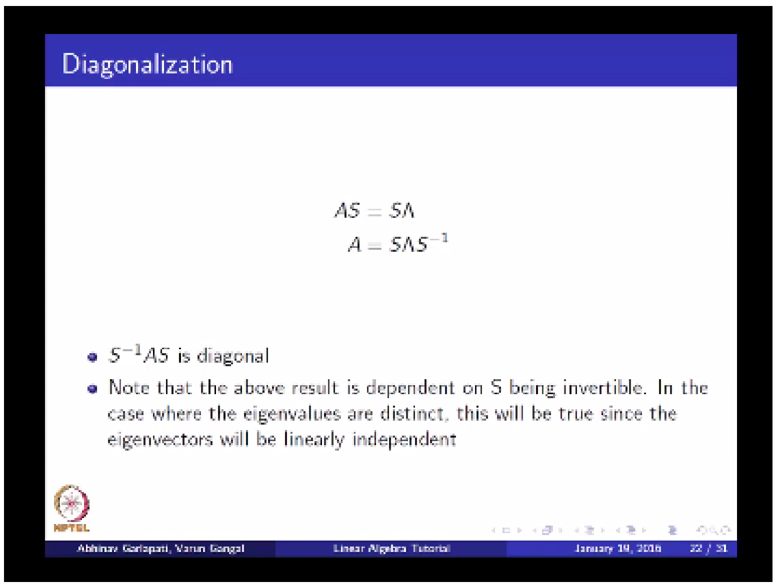
(Refer Slide Time: 06:53)



Diagonalization gives us a way of representing a matrix in terms of its eigenvalues and eigenvectors. Let us consider a NxN^2 matrix A where the amount of matrix where every column is an eigenvector of A/S. On multiplying S/A each column would get multiplied by λi since the column itself is an eigenvector of A.

This right hand side can then be simplify and the product of two matrices. The first one mean is itself by the second one B the diagonal matrix where the i^{th} a diagonal matrix the eigenvalue λi. Remember the DLH is AS.

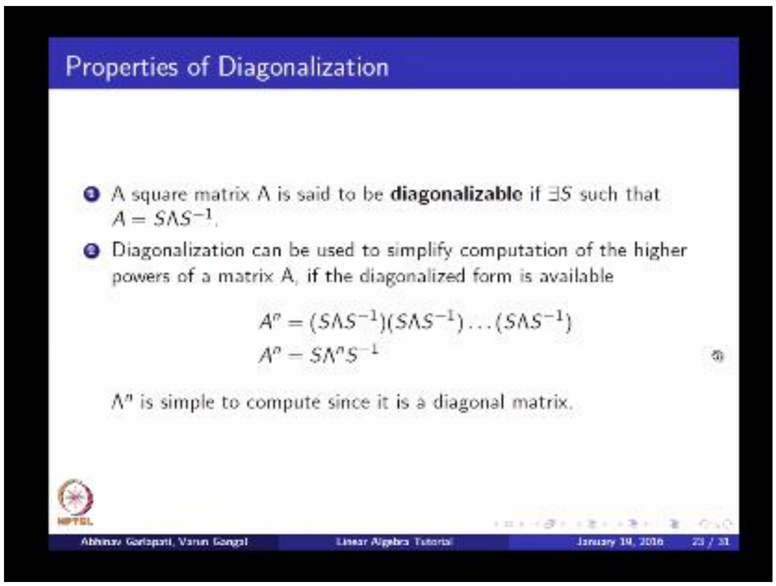
(Refer Slide Time: 07:50)



Now we have the equation AS-=Sλ where λ is the diagonal matrix of eigenvalues. On simplifying this we get A=SλS=, this is a diagonalization of A. Note that $S^{-1}AS$ is a diagonal matrix since $S^{-1}AS$ is nothing but λ the diagonal matrix of eigenvalues. This result is dependent on S being invertible.

It will bored with the eigenvalue of a matrix at distinct. Since the eigenvectors would then be linearly independent. This would mean the problems of S would be linearly independent, and hence S would be pulled and as a consequence invertible.

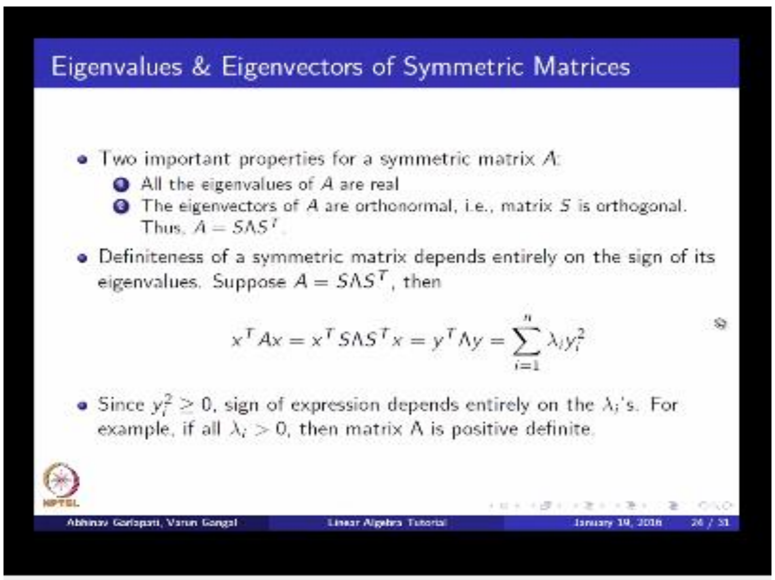
(Refer Slide Time: 08:59)



Then do we see that the square matrix is diagonalizable. Well when such a diagonalization exist we saw that we needed S to be invertible for the diagonalization to exist. Another advantage of diagonalization is that it simplifies the process of computing A^n, the first represent every A in diagonalized form.

Now you can see that the S^{-1} of the first term and the S of the second term would multiply to give us time. Similarly for the second toward, third, fourth and so on, in this way by regrouping the terms we get $A^n=S\ \lambda^n S^{-1}$. Note that it is very easy to compute the nth power of a diagonal matrix. Since you just have to realize every diagonal element to the power of n. In this way the diagonalization has left us simply by the process of computing A^n without the simplification we would have needed to multiply a non-diagonal matrix 10 times.

(Refer Slide Time: 10:30)

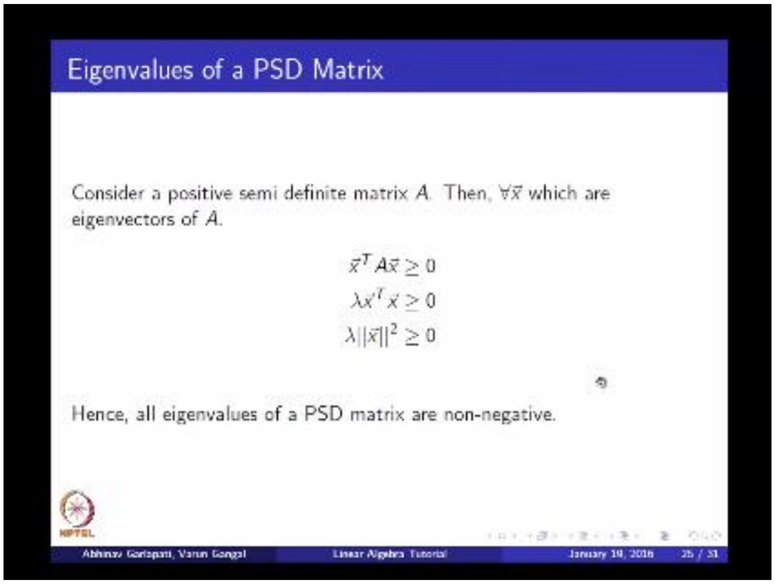


If a matrix is symmetric then all its eigenvalues are real numbers. Also if its eigenvectors are also normal that is they are mutually orthogonal and normalized. This means that the matrix of eigenvectors S is also orthogonal. We have seen that for orthogonal matrices that inverse and a transpose are the same.

Hence we can write $A=S\lambda S^T$ as when the diagonalization we defined earlier. For symmetric matrices the definiteness can be inferred from the signs of their eigenvalue. Suppose that $A=S\lambda S^T$ now taking the quadratic form with respect to A for the vector X, X^TA X simplifies to Y^T λy, thereby is S^TX.

This further simplifies to sum over iλiyi^2. Now for a matrix to be positive definite this term must always be positive. Since yi^2 is always greater than 0 anyway this i of this term depends on the eigenvalue. All the eigenvalues are positive, the matrix is positive definite.

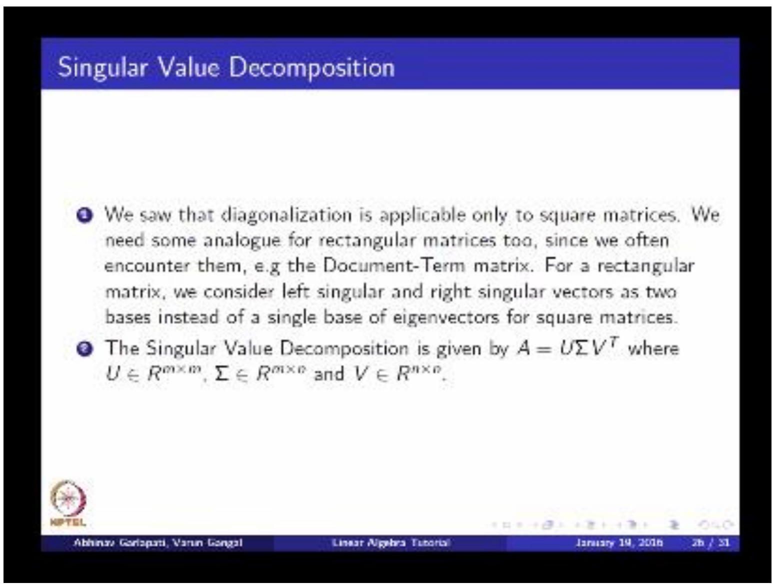
(Refer Slide Time: 12:12)



If we know that the matrix is positive semi definite or PSD then what can we say about its eigenvalues. Since the quadratic form of a PSD matrix is non-negative for any vector X this should hold for the eigenvectors too. Now since AX= λXX^{T}AX simplifies to λ norm of X^2 greater than equal to 0.

Since eigenvectors are nonzero by definition the square of the norm is always positive. This means that every eigenvalue of A is non-negative.

(Refer Slide Time: 13:03)

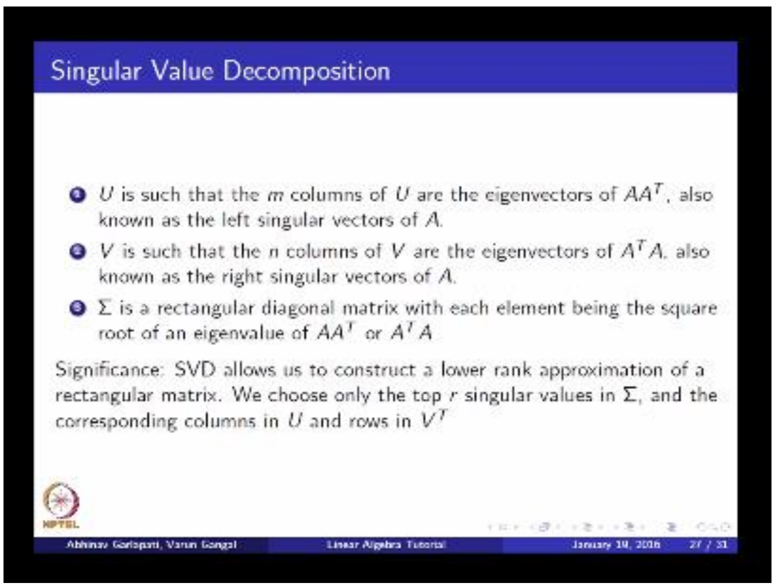


We looked about diagonalization which stood in our square matrix of size nxn and represented it in terms of its eigenvectors. However, we cannot directly apply by the same diagonalization for rectangular matrices, since the notion of eigenvector is defined only for the square matrix. They need a diagonalization for rectangular matrices since it come to them often.

For instance, the matrix of N data points out integers or the matrix of n documents and R terms. For the rectangular matrix A or size mxn we can represent it in terms of the eigenvectors of AT^T

and A^TA out of which our square matrices. This is known as the singular value decomposition A is represented as $U\Sigma V^T$ where U is a mxm matrix Σ is a mxn matrix and V is a nxn matrix.

(Refer Slide Time: 14:20)

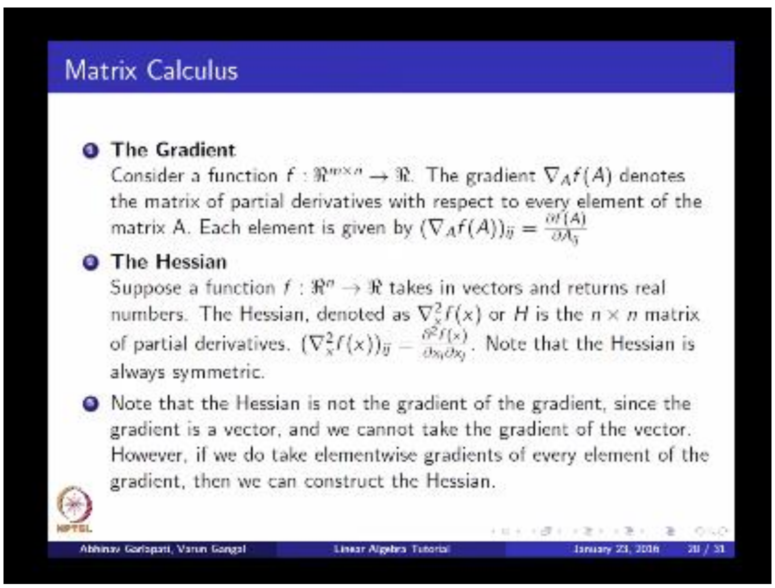


The three N is UΣV are as follows. In U every column represent an eigenvector of AA^T, in V every column represents as eigenvector or A^TA Σ is a rectangular diagonal matrix if each elopement being described of an eigenvalue of AA^T or A^TA. Now note that AA^T and A^TA have different eigenvectors with the set of eigenvalues is the same.

This is because suppose $A^TAX=\lambda X$ for some eigenvector X and eigenvalue λ. Now multiplying both side by A we get AA^T whereas $AX=\lambda AX$ hence AX is an eigenvector of AA^T while λ is also an eigenvalue of AA^T, this is why AA^T and A^TA have the same set of eigenvalues. The significance of this decomposition is that we all know in U, V and Σ such that the eigenvalue is magnitude is larger come first both in U and V at the column or and also along the diagonal in Σ.

Then we can drop everything greater than index R to get a R dimension and load and approximation of the original matrix A. Since approximate form of A we represented as U which is an mxr matrix, Σ which is a rxr matrix, and V which is a nxr matrix.

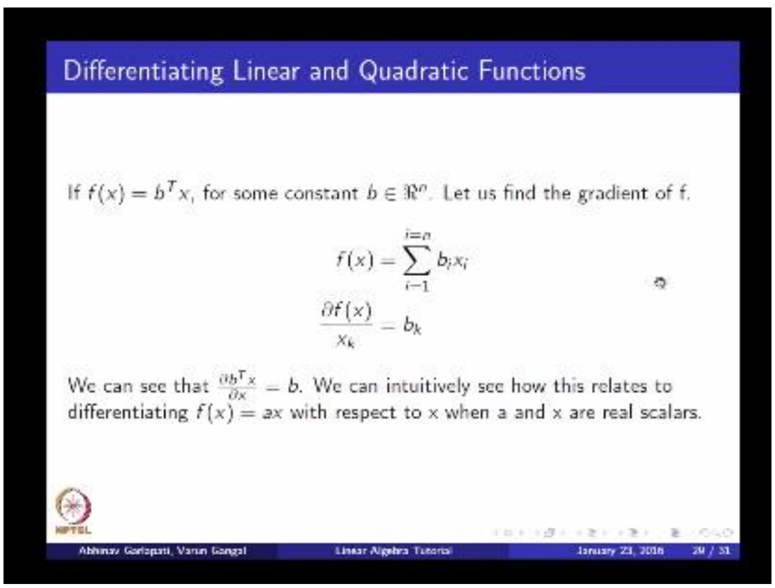
(Refer Slide Time: 16:26)



Consider function F which takes in matrix systems of dimension mxn and outputs real of course. The gradient is the matrix of partial derivatives. The i,j element of ΔF(A) or the gradient of F(A) is the partial derivative of F(A) with respect to Aij. Consider it with time of function which takes in at the in dimensional vector and returns a real number.

The Hessian for this function is defined as follows, the i,j the element of the Hessian is given by first differentiating F(X) with respect to the j^{th} component of X, Xj and then the ith component Xi. We can see that the Hessian would be nxn matrix.

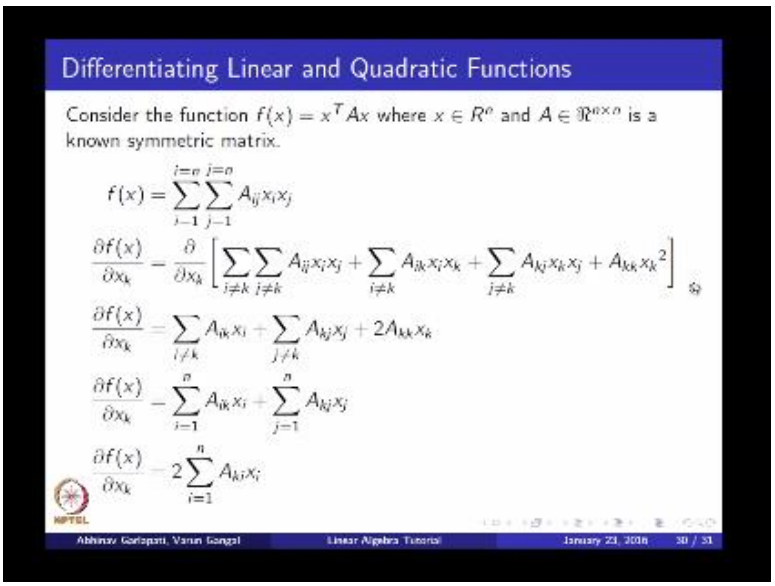
(Refer Slide Time: 17:27)



Now let us study how will you find the gradient for some simple vector functions. Consider the function $F(X)=B^TX$ where X is an in dimensional vector and B is also an in dimensional vector. F(X) can be written down as sum over i=1 to i=n BiXi. On differentiating this with respect to the 8^{th} component of the vector X we can do F(X) by ∂XA=Bk.

The gradient of F(X) is given by the vector V, you can see how this intuitively remains to the first derivative of the scalar function F(X)=AX which is equal to A.

(Refer Slide Time: 18:27)

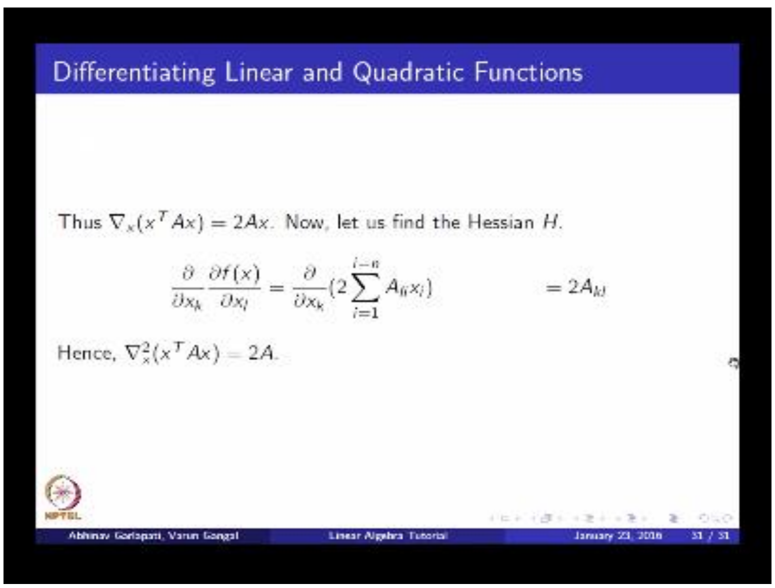


We had earlier looked at a type of function called the quadratic form defined for an nxn matrix A. The quadratic form with respect to matrix A is a function $F(X)=X^TAX$ so it takes in a in dimensional vector X. Now let us have a look at how one can find the gradient and Hessian for the quadratic form of an known symmetric matrix A.

They can write down F(X) as sum over I^{-1} 1 to n, sum over j=I to n AijXiXj. We can split up this summation into four terms based on whether i and j are equal or not equal to k. Finally we get

∂F(X) for ∂XK=Y sum over i=1, i=n AkiXi. Note that the simplification from the second last step with the last step can only be done if A is symmetric.

(Refer Slide Time: 19:41)



Thus we get the gradient of X^TAX = AX. Similarly, on further differentiating every element of the gradient by XK we can drive the Hessian of the function. The Hessian of this function comes out to be 2A.

IIT Madras Production

Funded by

Department of Higher Education