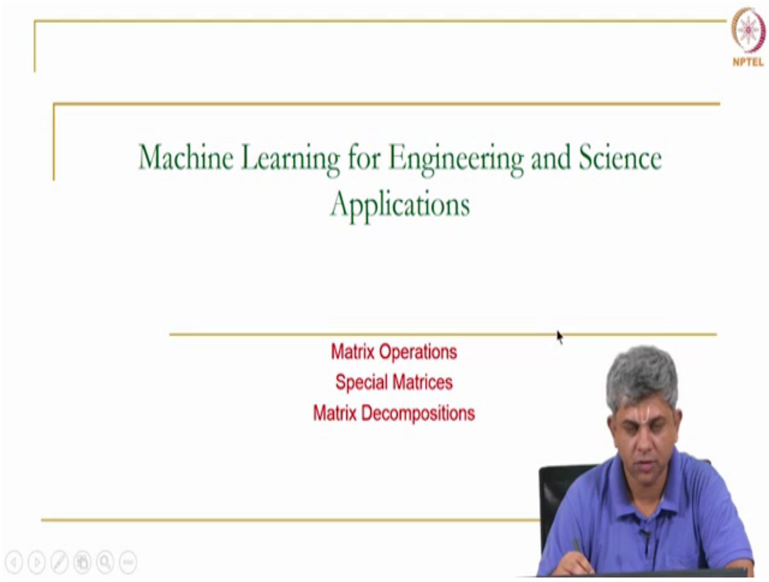
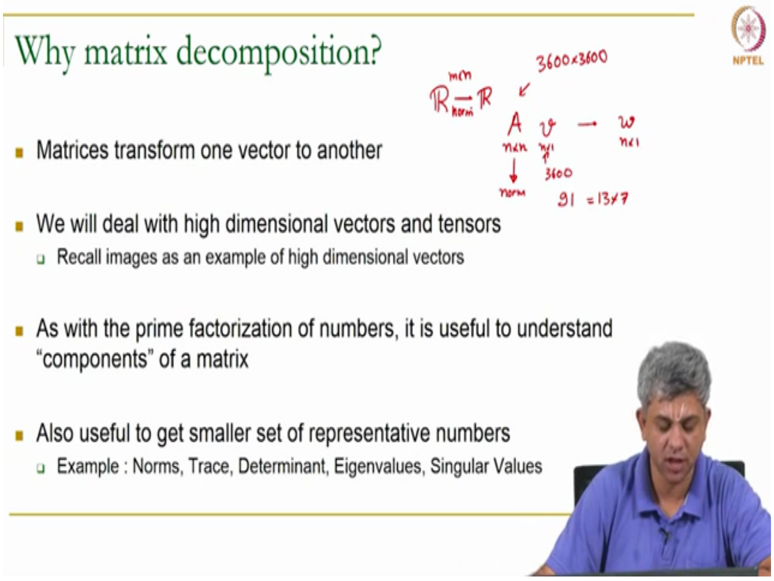
Machine Learning for Engineering and Science Applications
Professor Dr. Balaji Srinivasan
Department of Mechanical Engineering
Indian Institute of Technology, Madras
Matrix Operations Special Matrices Matrix Decompositions

(Refer Slide Time: 0:15)



In this video we will be looking at the final piece of the linear algebra portions of this course, specifically we are going to look at matrix operations, some special types of matrices and matrix decompositions, specifically within matrix decompositions we will be looking at Eigen Decomposition. Now all these ideas are ideas that you should have been familiar with, please remember this is just a recapitulation of the kind of things that you need to know for this course.

(Refer Slide Time: 0:48)



If time permits we will look at greater physical interpretations of this, but by itself linear algebra is a vast subject. So why are we looking at matrix decomposition also known as matrix factorization? To remind you in previous videos we had seen that matrices transform one vector to another. So if you pre multiply one vector by a matrix you get another vector. Now this has physical meanings which we will look at as we go on through this lecture itself.

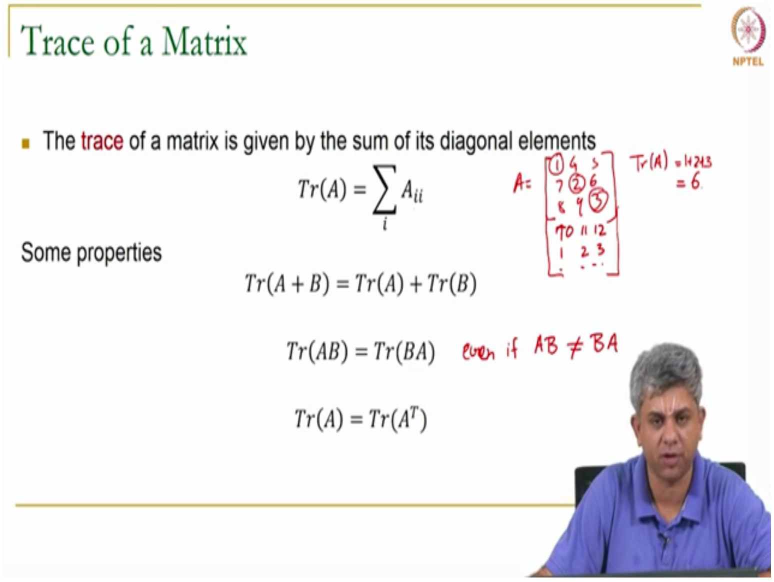
So typically as you remember we deal with very high dimensional vectors and tensors within machine learning, you might recall that if you have a 60 cross 60 grayscale image you can interpret it as if it is a single 3600 dimensional vector or one vector with 3600 components you know pixel 1, pixel 2, up till pixel 3600. So these are just examples of the size of vectors that you will be dealing with which means we are actually dealing with very large matrices.

So if we have to convert an n cross 1 vector into another n cross 1 vector, so if you have a vector v let us say this is n cross 1 and this has to go to another vector let us call it w which is also n cross 1, you will have to pre multiply by a matrix which is n cross n, okay which means if let us say n is 3600, then A is 3600 cross 3600 matrix, okay. Now it is usually useful to understand you know what these components mean and as it turns out its original form it is kind of art to understand and just like you know for a number we typically take let us say if you have something like 91, you will say 91 is 13 times 7, both of these are prime indivisible further factors.

Similarly it is useful to factorize a matrix itself and you can think of an Eigen Decomposition or other decompositions that we will be talking about as simple decomposing one big thing into smaller thing which we can understand a little bit better. It is also useful sometimes for a large matrix to be summarized by 1 or 2 or a fewer numbers rather than a large numbers. So we have seen that norms for example for a matrix we often use atleast within this course we will be using the Frobenius norm. So norms would be all this n cross n reduced to a single number, okay so mapping from an n cross m or m cross n to a single number, so that would be the norm.

Another such measure is trace all these are smaller measures obviously they do not summarize the whole matrix, a determinant which you will be familiar with, eigenvalues, singular values, etc are similar numbers which try to encapsulate some idea about what the matrix represents as we will see later on in the slides.

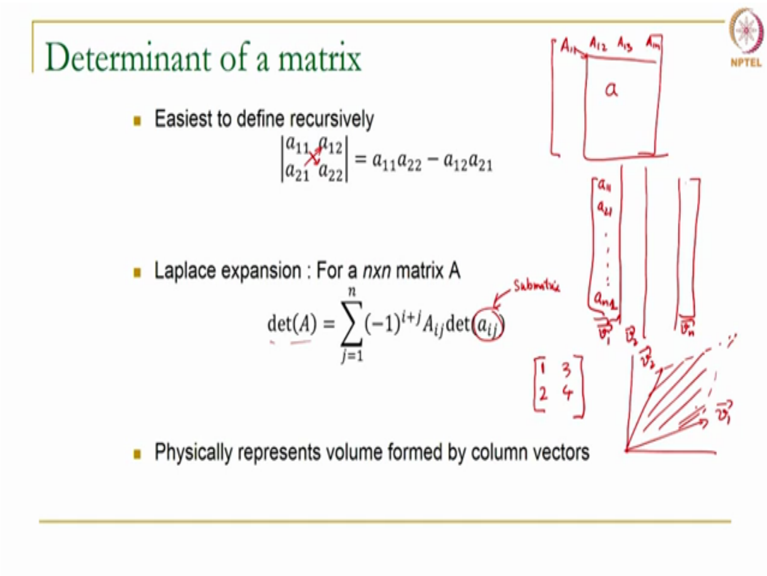
(Refer Slide Time: 3:48)



So the first idea we are going to look at is what is called trace of a matrix, it is simple a trace of a matrix is simply the sum of the diagonal elements of the matrix, okay. So it is sigma A ii, so if you have a matrix let us say 1 2 3 4 5 6 7 8 9 then trace of this matrix A is 1 plus 2 plus 3 which is 6 as shown here. Now the idea of trace you can use for non-square matrices also, so let us say this is longer you had 10 11 12 1 2 3 etc you would still look at only A ii which are A 11 A 22 A 33 it would still be 6, okay. So typically however we will be using trace for square matrices.

So the trace has certain properties, trace of A plus B is trace of A plus trace of B, trace of A times B is trace of B times A even though even if AB is not equal to BA the trace itself is a property that does not change when you commute the matrix product. Similarly trace of A is the same of trace of A transpose which directly follows from its definition these are some useful properties.

(Refer Slide Time: 5:18)

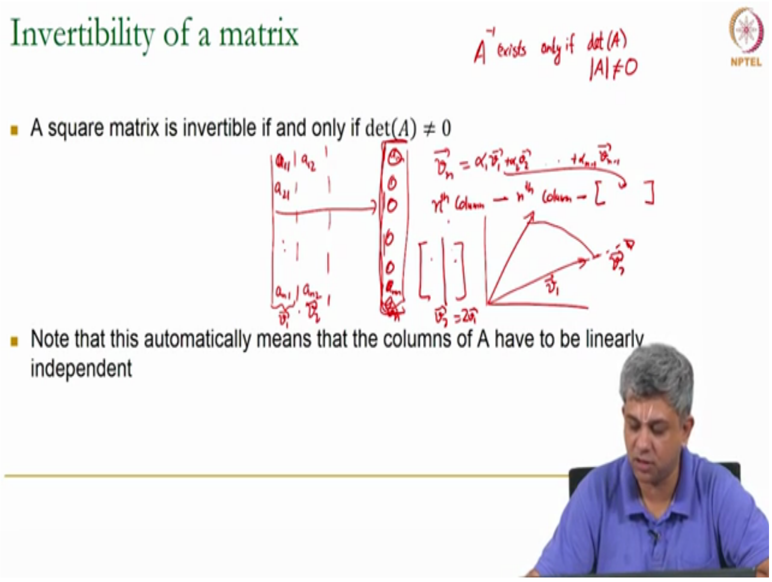


The next idea is that of a determinant of a matrix again you will be very familiar with this I just want to make the notation a little bit clear nothing much else in this slide. So we all know that if you take a 2 by 2 matrix you can simply define the determinant as a 11 a 22 minus a 21 a 12 and that the determinant of a bigger matrix is using kind of a recursion idea, so if you have something like A 11 here then and if we call this sub matrix a, then determinant of A is defined as summation over the rows or columns we can do it either way as you know minus 1 power i plus j A ij times determinant of this is the sub matrix, okay again this is something that is very very familiar to you from school.

Now more importantly the determinant actually represents the volume so if you interpret the first column let us call it a 11 a 21 a (1) n1. So if we call this v 1 vector, this as v 2 vector so on forth up till v n vector for a square matrix then so let us take a simple case, so if I have 1 2 3 4 I can now think of this as two vectors the 1 2 vector v 1 vector, another v 2 vector in that case the determinant represents the area of this parallelogram, okay.

So similarly you can extend this to higher dimensions also if you have 3 vectors it will be the volume represented by those 3 vectors 4 5 6 you can start interpreting this as n dimensional volume so this has a very interesting consequences as we will see shortly.

(Refer Slide Time: 7:44)



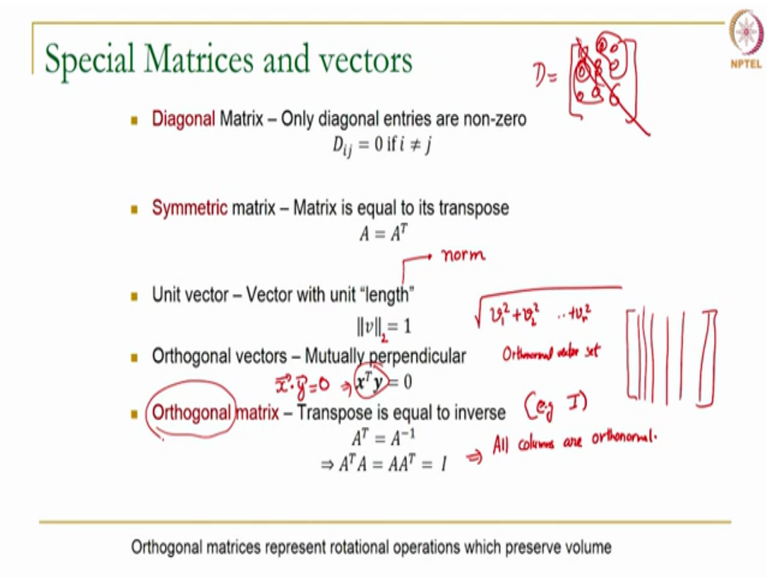
So one thing is invertibility of a matrix again as you know A inverse is defined only if you have determinant of that matrix A as non-zero. So A has a unique inverse if and only if determinant of A which we will sometimes denote as if its absolute value is not equal to 0. Now this automatically means that the columns of A have to be linearly independent, how does this follow?

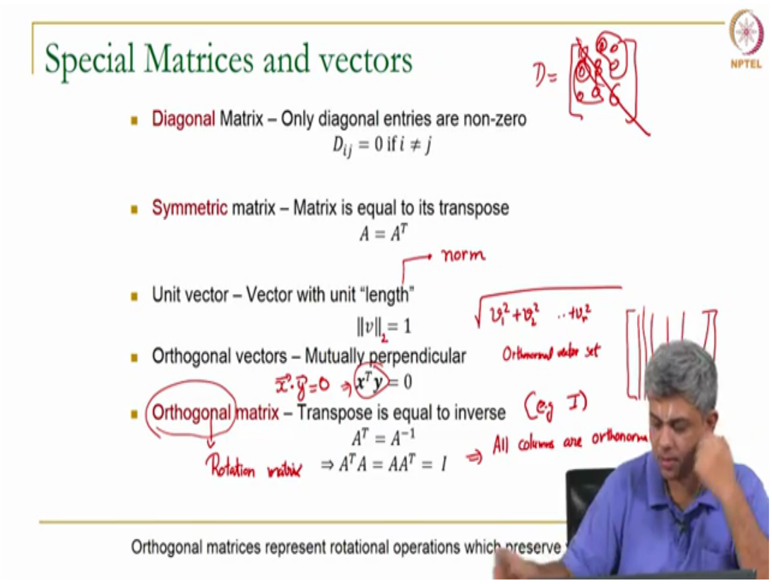
So remember I will use the same example as last time, suppose we have a 11 a 21 up till a n1 if we call this v 1 vector, v 2 vector is a 12 so on and so forth and we have v n vector which goes till a nn. Now in case one of these columns let us say this column let us say v n vector could be written as a linear combination of this was some alpha 1 v 1 plus alpha 2 v 2 up till alpha n minus 1 v n minus 1, then what does this mean? By simply doing an operation of nth column goes to nth column minus this thing, you will get all 0's this transformation as you know preserves the determinant which means the determinant will become 0, okay.

This also has a nice physical interpretation, what it says is if you have one of these vectors which can be represented as a linear combination of the other vectors the volume of the thing formed by of the parallelogram or the parallelepiped formed by these vectors actually becomes 0, you can see this easily in the 2D case or even in the 3D case. So let us say you have two vectors in this case you only have a 2 by 2 matrix, okay. In case one of them is linearly dependent on the other, it simply means that both these vectors or one of these vectors let us say v 2 is equal to 2 v 1, then the area formed by these two vectors is simply going to be 0, okay. You are going to get non-zero area only if one of them is not simply scaling of the other.

Similarly if you have three vectors, if one of them is the linear combination of the other two vectors it means all three are in the same plane which means they are not going to form a non-zero volume. So there are multiple interpretations for A inverse existing and there are deep connections with the determinant of the matrix.

(Refer Slide Time: 11:10)





So we will now look at some special matrices and vectors again this should be familiar for you. The first idea is that of a diagonal matrix, a diagonal matrix is one where only the diagonal entries are non-zero, okay all other off diagonal entries are 0. So mathematically D ij if D is the matrix is equal to 0 if i is not equal to j. A symmetric matrix is a matrix which has you know it is symmetric across the diagonal, another way to say it is that the matrix is equal to its own transpose.

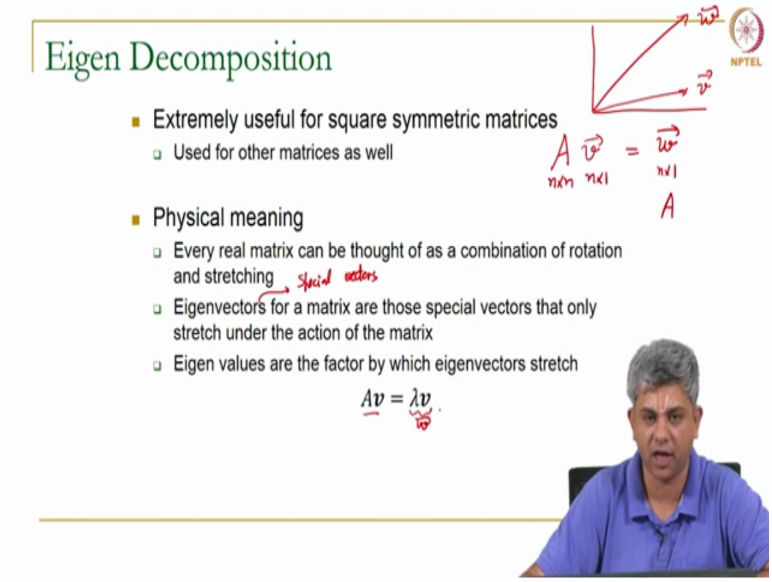
A unit vector is a vector with as all of us are familiar unit length. In our notation which we remember we had used the idea of a norm for length. So typically norm of that vector is equal to 1, which norm if you ask typically when we say unit vector we mean the 2 norm, okay please remember the 2 norm or L 2 norm is simply v 1 square plus v 2 square so on and so forth up till v n square, square root, okay but you can define a unit vector with norm 1 etc but it is usually the 2 norm that we use.

Another (usual) useful idea is that of orthogonal vectors, it simply means vectors that are mutually perpendicular which means x if x and y are mutually orthogonal mean x dot y is 0, which remember can be written in this matrix form you take x transpose y and set that equal to 0. We also have the idea of orthonormal vectors or orthonormal vector set, where you have unit vectors that are perpendicular to each other.

Orthogonal matrix is a matrix whose transpose and inverse are the same thing, which means A transpose is equal to A inverse, the simplest sort of orthogonal matrix is the identity matrix, it has some nice properties which we will discuss very shortly, but a simple thing that follows from this is that A transpose A and AA transpose is equal to I, this also means that all columns are orthonormal.

So remember if (A is equal to) A transpose is equal to A inverse when you multiply the matrix by its transpose you are actually going to get quantities of this sort which have to be 0, if x and y are not the same. Now where do we use orthogonal matrix even though this is orthonormal column vectors we still call it orthogonal matrix, an orthogonal matrix typically can always be thought of as a rotational operation, okay what that means is if I have one vector I pre multiply it by something and all it does to that vector is simply rotates it without changing the length, the matrix which should have been used would always be orthogonal this can be proved we will not have time to show all that, but please do remember it whenever you see an orthogonal matrix please think a rotation matrix, okay that is another way to think about it.

(Refer Slide Time: 15:08)



So let us now come to the matrix factorization that I was talking about Eigen Decomposition. This is typically very useful for square symmetric matrices specially square symmetric real matrices even though you can use it for other matrices as well and I am sure you would have done it before. As far as this course is concerned, we will primarily be using it for square symmetric matrices and it has we are guaranteed several things, when symmetric matrices when we have square symmetric matrices as far as Eigen Decomposition is concerned.

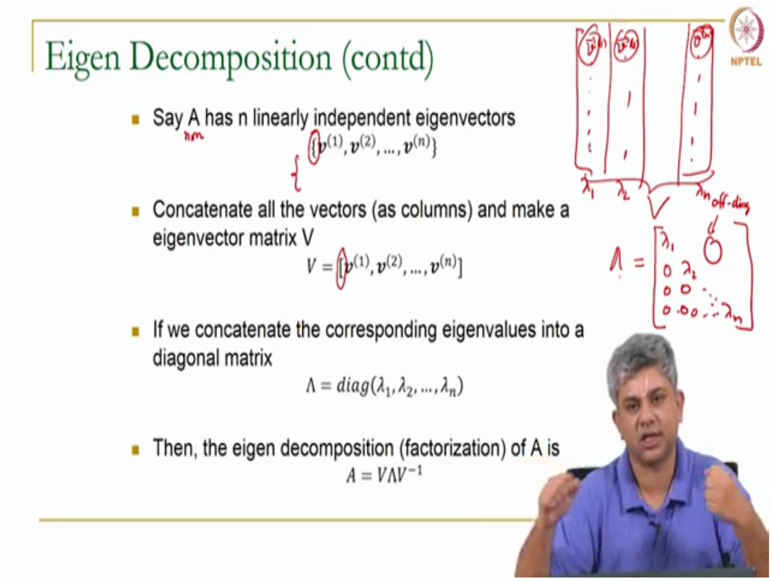
So here is the simple physical meaning that is usually useful in order for you to anchor yourself in the Eigen Decomposition. So every real matrix, remember I have talked about this before too if you have a matrix A, what it can do for now I will talk only about square matrices. So if you have a matrix A if it (multiplies) pre multiplies a vector v, it results in some other vector w.

Now you can think of A as a machine or an operator acting on v and giving you w, takes v takes it to w, okay. So let us say this is the vector v, this is some vector w and A has taken v to w. Now through physics as well as intuitively you can see that there are only two things that this matrix A can do to v, it can rotate it that is it can turn it through an angle even in 2D, 3D in any place that you can think of it turn it through an angle and the other thing it can do is it can change its length, okay.

So the length of v might not be the same as length of w, but it can stretch it, rotate it or rotate it and stretch it these are the two operations that any matrix can do as far as acting on another vector's concern. Now this is extremely useful now if you can think of every operation as if it is a matrix, then there are special vectors and the only thing that matrix will do to this vector is just stretch it.

Eigenvectors are those special vectors, so you are given a matrix A and there are a set of special vectors for that which we will again call v eigenvectors which will only stretch under the action of this matrix. What is an eigenvalue? Eigenvalue is the factor by which this vector stretches, okay. So mathematically I would write A times v is a new vector this is the new vector w but this w is not rotated it is only stretched. So Eigen Decomposition is that angle in some sense or that set of vectors which only stretch under the action of the matrix A, this is the physical interpretation, okay.

(Refer Slide Time: 18:20)

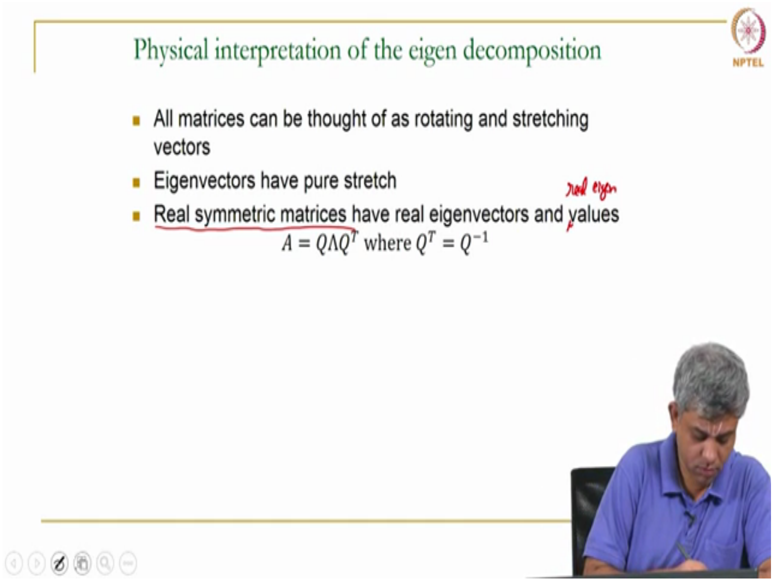


So let us say so we will now write the Eigen Decomposition, you can think of eigenvectors in some literal sense eigenvectors are essentially the coordinate system in which the matrix A looks the nicest it looks diagonal this is one way of looking at it if you do not understand this fact that is okay I will just write the mathematical expression right now. So let us say A has n linearly independent eigenvectors that is A is n cross n matrix and it has n linearly independent eigenvectors v 1 through v n.

Now we will do what we have been doing so far, I will write v 1 as if it is the first column, v 2 as if it is the second column remember v itself is a vector therefore it has n components and we will go till v n which also has n components. Now you concatenate or put them together and you get one large matrix the eigenvector matrix V. So this notation here if you have a curly bracket it is a set if I put this I have put them together, there is no comma separating this this is actually a set of numbers put together as a matrix.

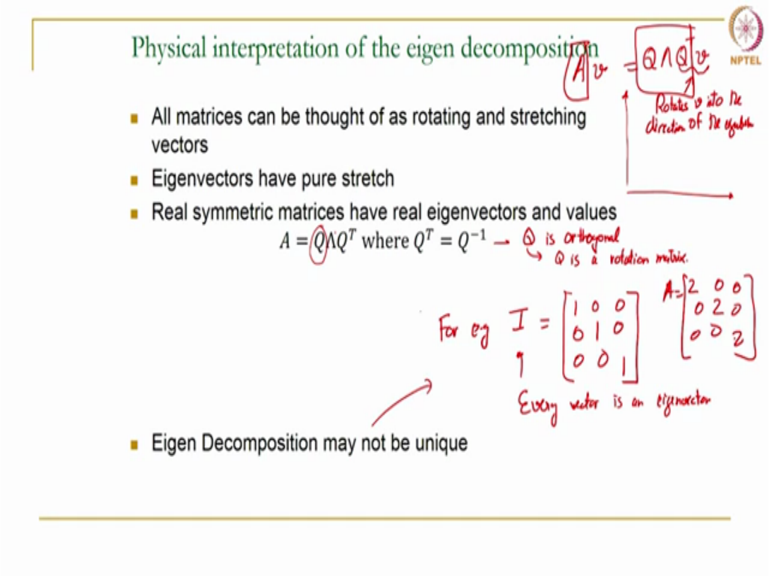
Now similarly if each of these has a corresponding eigenvalue lambda 1, lambda 2, lambda n and I put them together into one giant matrix lambda which is a diagonal matrix remember it is a diagonal matrix so all off diagonal elements are 0, then we can write the factorization of A as A can be written as a product of 3 matrices V multiplied by this diagonal matrix lambda multiplied by V inverse, physically what it means is we have sort of rotated into the coordinate system which is defined by all these eigenvectors and these eigenvectors purely do stretching, okay.

(Refer Slide Time: 20:46)



So as I said before all matrices can be thought of as rotating and stretching vectors. Eigenvectors are those vectors that are simply purely stretching. Now what we know is real symmetric matrices and this is where we will use them have real eigenvectors and real eigenvalues, okay this is not necessarily true of all matrices even if you have a real matrix and it is not symmetric, it might or might not have real eigenvectors and real eigenvalues, okay.

(Refer Slide Time: 21:30)



In case you do have a symmetric matrix there is a nice factorization for it, remember we have V lambda V inverse in the case of a real symmetric matrix you can write it as Q lambda Q transpose, where Q transpose would be the same as Q inverse which means Q is orthogonal that was our definition of an orthogonal matrix which also means Q is a rotation matrix. So physically what this means what this factorization means is so if I have A v and I am trying to determine what the action of this matrix A is on v and let us say A is a symmetric matrix, we know from here that I can write it as Q lambda Q transpose v, okay.

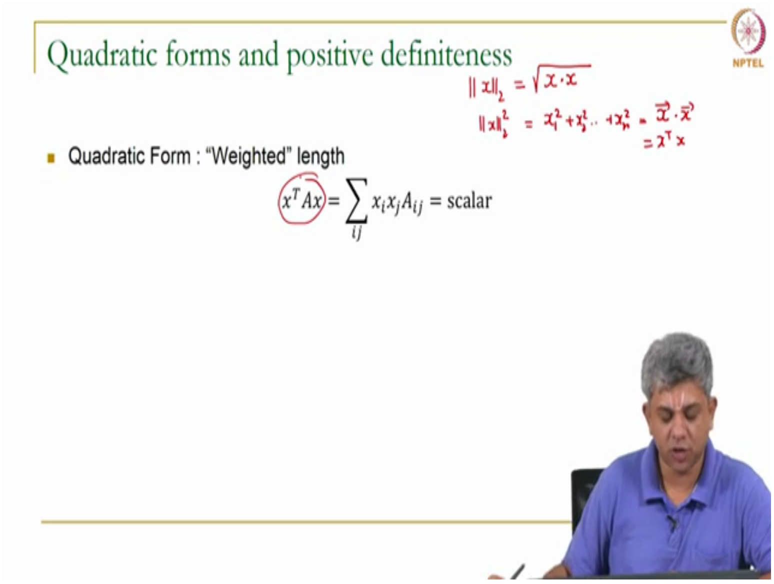
So let us say we have some eigenvalues or an eigenvectors, what this action does what Q transpose v does is it rotates v into the direction of the eigenvector, okay so that is what it basically does. So you have two actions going on there is rotation and then there is stretching, so what the Eigen factorization does cleverly is this rotation is first rotated into the eigenvector directions, then you stretch it through the lambda and after that you rotate it back so that the net rotation and the net stretching are put together into one matrix A and which can be written as Q lambda Q transpose.

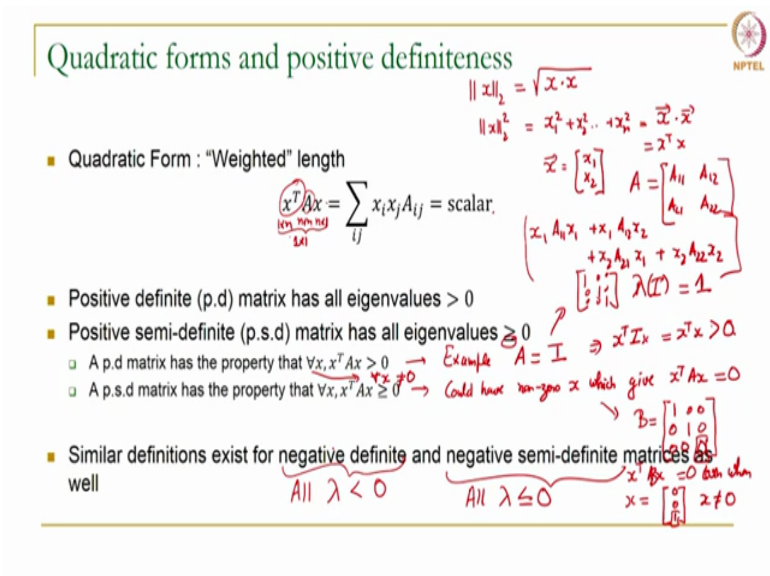
So this of course will take a lot of visualization, I just sort of summarize it, if time permits we will give some bonus videos towards the end of this course so that you can visualize it too and maybe we will give some bonus codes that you can run to see this. So one important thing for us to remember is the Eigen Decomposition might not be unique. For example if you have the matrix the identity matrix, so let us take a 3 cross 3 identity matrix for the identity matrix every vector is an eigenvector, why is that?

Because the identity matrix has only one action, it does not even stretch it basically keeps the vector as it is you can think of it as a stretch by a factor of 1, I could also make up another matrix let us call it A, so there is no rotation at all for this matrix A all it is doing is stretching and it will do so for every vector.

Now one can think of a counter part for this if we think of a rotation matrix it is not going to have stretching at all, it will not have stretching at all which means really speaking that you cannot really have a real Eigen Decomposition because an Eigen Decomposition only tries to find out those vectors which are actually going to purely stretch so if I have a pure rotation matrix or a pure orthogonal matrix it is not going to have a real Eigen Decomposition, you can try this out for yourself.

(Refer Slide Time: 25:16)





So we are going to look at one very important idea that of a quadratic form, okay. Remember when we are trying to find out the length of some vector let us say x, the 2 norm is square root of x dotted with x and if I look at 2 norm square that is going to be x 1 square plus x 2 square up till x n square which I can write as x dotted with x or x transpose x. Now the quadratic form is a slightly weighted form of this, I will show you what I mean by that it is written as x transpose Ax.

So let us look at what this means. Suppose I have A as n cross n matrix and x as n cross 1 vector then x transpose is going to be 1 cross n which means all put together you are going to get a 1 cross 1 number which is a scalar. So a quadratic form is something that takes a matrix or takes a vector x and gives back a scalar much like the length plus except it has a factor of A in between.

Now what does this do, if you write it out if you write out this matrix product you will basically get combinations of all sorts of terms. For example let us say if x vector is x 1, x 2 and A is A 11 A 12 A 21 A 22, then x transpose Ax is simply going to be x 1 A 11 x 1 plus x 1 A 12 x 2 plus x 2 A 21 x 1 plus x 2 A 22 x 2, okay it is a sum of every possible you know linear combination of this as I have written here x i x j A ij a simple summation of that is called the quadratic form we will see several uses of this as we go on through the course.

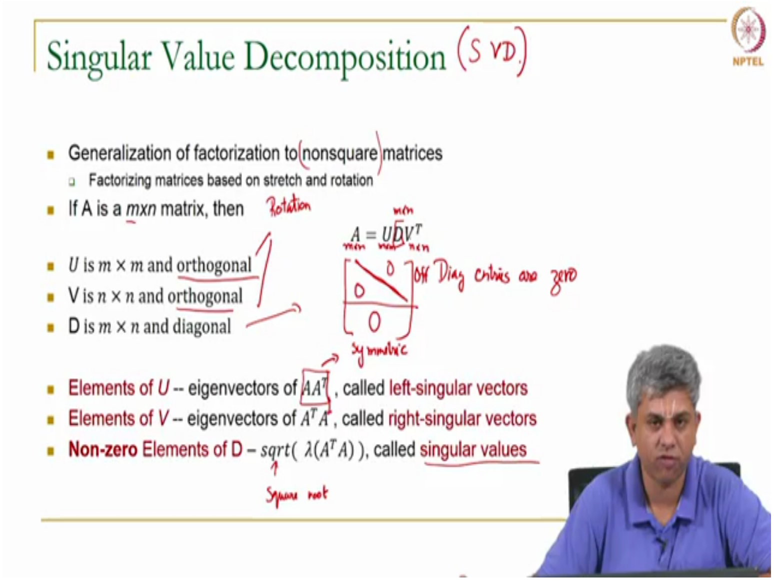
Now one important definition that sort of comes from the quadratic form is that of a positive definite matrix. A positive definite matrix is any matrix that has all completely positive eigenvalues, we also have another definite called positive semi-definite, a positive semi-definite means not just greater but greater than equal to 0. A positive definite matrix has very nice property which is all quadratic forms so if you take any x at all it does not matter which x you take it is your choice x can be positive or negative, which ever x you take x transpose Ax will always be positive.

A simple example is the matrix A's identity, notice identity is already a diagonal matrix 1 1 1 which means lambda all eigenvalues of identity are 1 which means it is a positive definite matrix since all are positive, this will give us x transpose Ix since I is the matrix is the same as x transpose x, so this is always positive as you can say. So for all x of course for all x not equal to 0 if I multiply it by 0 of course it is trivially 0, so a positive definite matrix has the property that for all non-zero x, x transpose Ax will always be positive.

A positive semi-definite matrix has the property that for all x, x transpose Ax is greater than equal to 0. So you could have non-zero x which give x transpose Ax is equal to 0. So you can see a simple example of this let us say a matrix B which is 1 1 0, we can find out some x transpose Ax or x transpose Bx which is equal to 0 for x is equal to 0 0 1, okay this is a non-zero x but if you multiply it this term will be 0 if you write it out in this way you will you can automatically check that x transpose Ax or x transpose Bx is equal to 0 even when x is not equal to 0.

So similarly you can define negative definite and negative semi-definite matrices, so a negative definite matrix has all eigenvalues less than 0, negative semi-definite means all eigenvalues are less than equal to 0 and similarly the quadratic form x transpose Ax for a negative definite matrix will always be less than 0 and for a negative semi-definite matrix x transpose Ax will be less than equal to 0.

(Refer Slide Time: 30:58)



Finally we come to another decomposition, we will not be using it very often in this course but it is a very useful idea and it is a useful idea when we discuss decompositions, this is the generalization of the idea of factorization that we have just used you can think of it as a generalization of the idea of the Eigen Decomposition itself, but we can apply this idea to non-square matrices, okay.

So factorizing matrices we had done it based on stretched and rotation, so is the same idea which is applied for singular value decomposition also. So let us say A is m cross n matrix, m which is not necessarily equal to n so it can be a non-square matrix. So then you can write the factorization of A as UDV transpose, where each of these have the following properties, U is a m cross m matrix, okay so remember A is m cross n, U is m cross m, V is n cross n, so obviously if we need to match the sizes we need a m cross D is a m cross n matrix, okay.

Now U and V have properties remember both of them are orthogonal which means both of them are can be interpreted as rotation matrices, D is a diagonal matrix by diagonal what does it mean it means that only the diagonal entries are non-zero in case it is not square even these elements will be 0, okay so off diagonal entries are zero that we know for sure. Now there are certain terminology here, the elements of U this matrix U are called the left singular vectors and they can be calculated as if they are eigenvectors of AA transpose.

Now notice one thing about AA transpose in case A is real, AA transpose is symmetric and since it is symmetric we know if it is real in symmetric then it has real eigenvalues and real eigenvectors, okay so this will always be real U's elements will always be real for any real A. Similarly elements of V they are just a switch they are eigenvectors of A transpose A and they are called the right singular values.

The non-zero elements of D the elements here are given as the square root this is the square root of the eigenvalues of A transpose A, okay so these are called singular values. A singular value decomposition has very similar not the same but very similar interpretation to what I talked about in terms of the interpretation of the Eigen Decomposition you take a vector rotate it into by the V transformation you rotate it and D simply stretches and then you rotate it back, okay.

So you can think of any matrix A as again doing simply a rotation as well as stretching and that is the significance of a singular value decomposition atleast as far as we will discuss singular value decomposition often called SVD as far as this course is concerned if time permits towards the middle or end of the course we will probably provide a few bonus videos with which you can actually visualize Eigen Decomposition and singular value decomposition.

This ends the discussion of linear algebra atleast the separate discussion of linear algebra for this course, in the next series of videos in next week we will be looking at probability which is the next half of mathematics (that we are) that we require for this course, thank you.