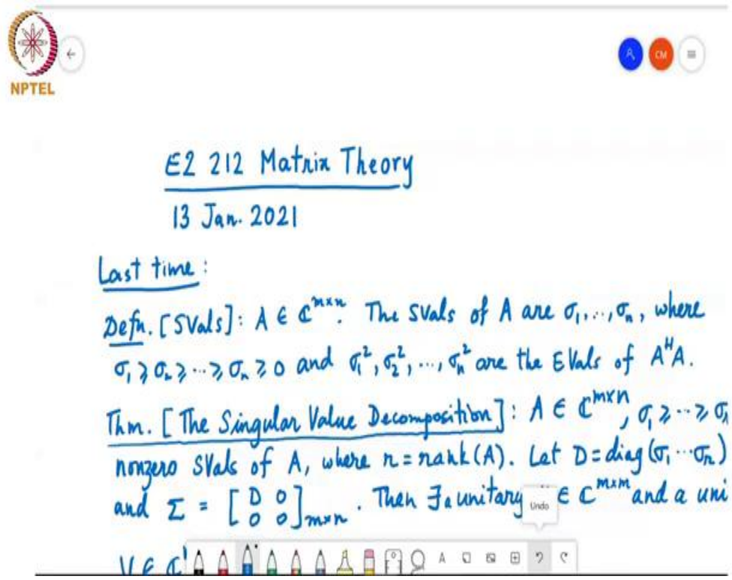
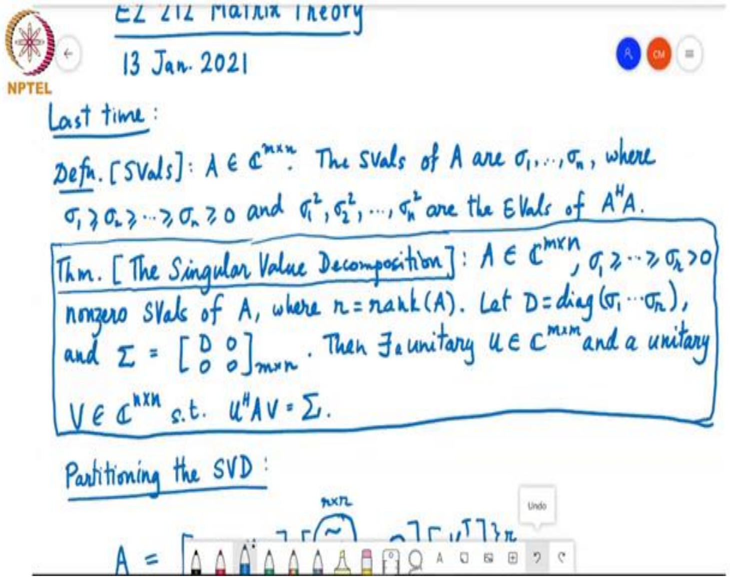
Matrix Theory
Professor Chandra R Murthy
Department of Electrical Communication Engineering
Indian Institute of Science Bangalore
Properties of SVD

(Refer Slide Time: 0:15)



The last time we were discussing about the singular value decomposition. And recall that the singular values of a matrix are sigma 1 through sigma n, where these sigma 1 through sigma n are ordered with sigma 1 being the largest and sigma n being the smallest singular value, they are all real numbers and non negative and such that sigma 1 squared, sigma 2 squared up to sigma n squared are the eigenvalues of A Hermitian A.

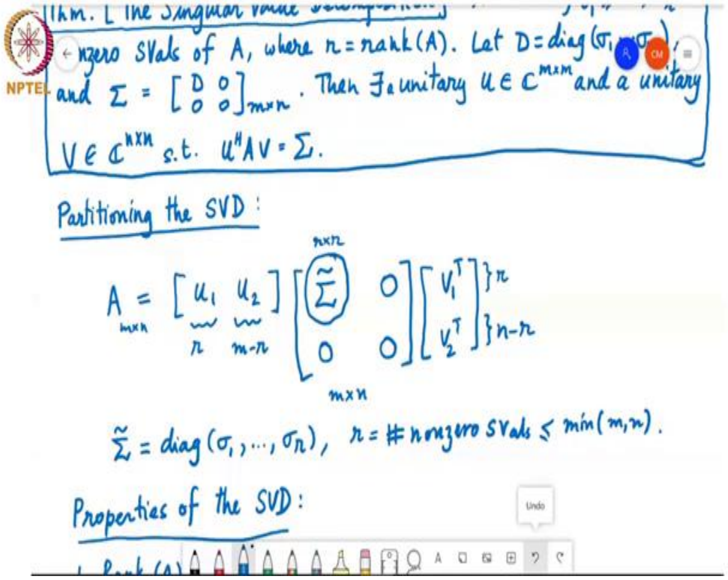
(Refer Slide Time: 0:52)



Now, we also saw the singular value decomposition theorem, which says that, if A is an m by n matrix and sigma 1 through sigma r the nonzero singular values of A, where r equals, r is the rank of this matrix A. And if we define D to be a diagonal matrix containing sigma 1 through sigma r as its diagonal entries and 0s everywhere else, and sigma to be a matrix with D as its top left r cross r block and 0s everywhere else and of size m cross n, then there exists a unitary u of size m by m and a unitary v of size n cross n such that u Hermitian AV equals this matrix sigma.

So, this is a very important theorem in the sense that I like it for its generality, it applies to any matrix A, absolutely no structural assumption made on the matrix A and you can construct a singular value decomposition for any matrix.

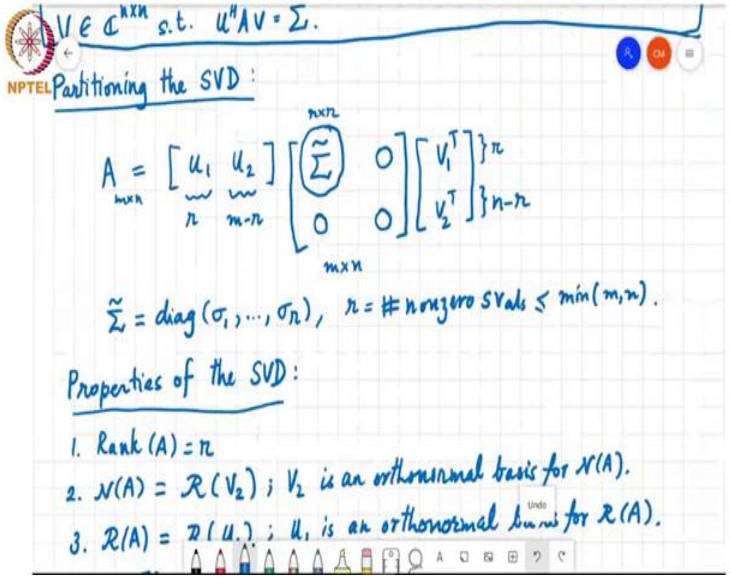
(Refer Slide Time: 2:01)



We also saw that we can partition the singular value decomposition like this, we can write it as u1, u2 where u1 contains the first r columns of the matrix u, u2 contains the remaining m minus r columns of the matrix u, this is sigma tilde, which is the r cross r matrix is the same as the D written in the statement of the theorem containing the nonzero singular values.

In fact, this is yeah, these are the nonzero singular values. And v is also partition row wise with the first r rows of v being denoted by, of v transpose denoted by v1 transpose. And the remaining n minus r rows of v transpose being denoted by v2 transpose. So, sigma 1 is therefore, a diagonal matrix containing sigma 1 through sigma r.

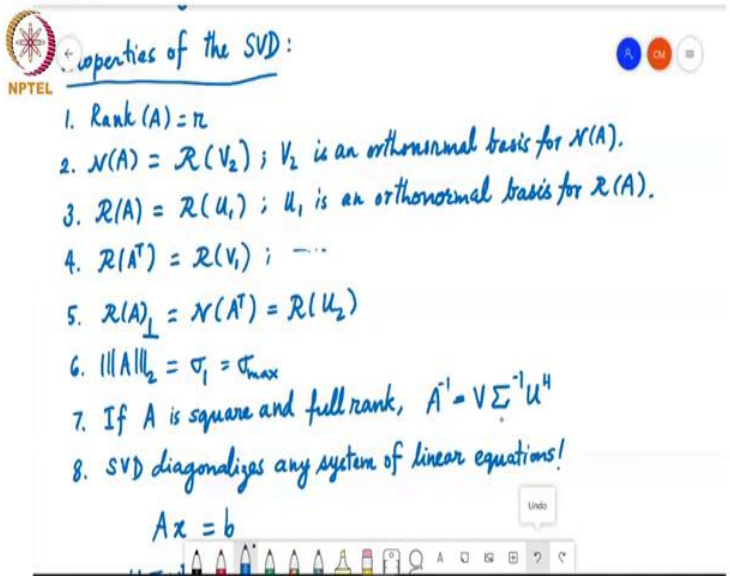
(Refer Slide Time: 2:58)



So, we were looking at several properties of the singular value decomposition. The first was that the rank of the matrix A equals the number of nonzero singular values. And we saw this because this is true, because, if we write it like this, then we can multiply these together and we see that we can write A to be, so, we can write this as u1 sigma tilde v1 transpose and we also recall that this is called the economy singular value decomposition.

So, basically the columns of A are linear combinations of the columns of the matrix u1 with the coefficients given by the columns of this matrix b. And this sigma tilde has linearly independent columns and v1 tilde has, v1 transpose as linearly independent rows. So, basically the columns of this matrix are linearly independent. So, they are exactly r linearly independent columns in this matrix A. So, in fact, the converse is also true that is if the rank of A equals r then it has r nonzero singular values. This comes from the SVD theorem itself. If you look back at the proof of the theorem.

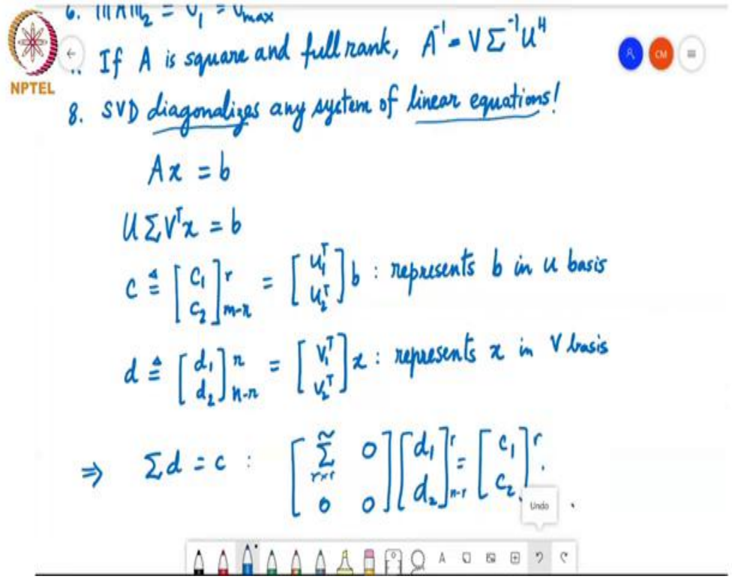
(Refer Slide Time: 4:27)

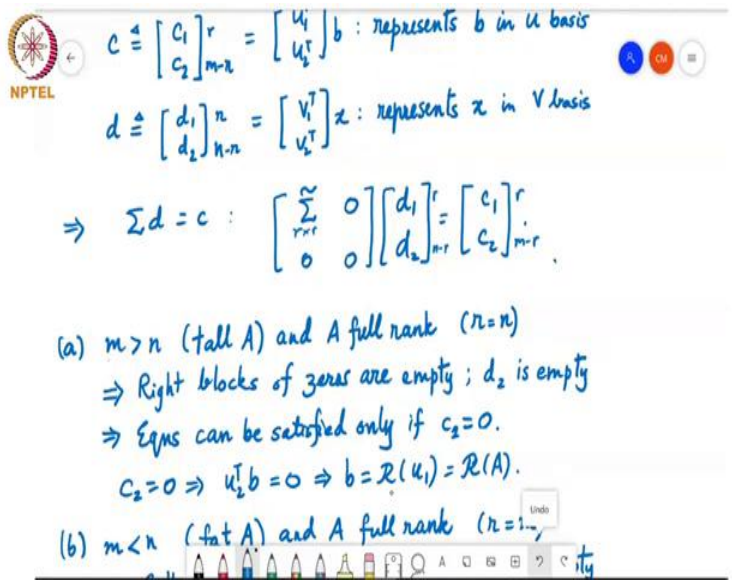


Other properties are that the null space of A is the same as the span of the columns of v2 this matrix here. And v2 is actually an orthonormal basis for the null space of A, the range space of A is the span of the columns of u1, u1 is this matrix here. And u1 is an orthonormal basis for the range space of A and the range space of A transpose is the span of the columns in v1 and the orthogonal complement of the range space of A which is the same as the null space of A transpose is the span of the columns in u2.

So basically, the SVD reveals the four fundamental subspaces associated with any matrix, it gives you an orthonormal basis for all of the four subspaces. Also, the spectral norm of A is equal to sigma 1, which is the largest singular value. And if A is square and full rank, we can easily determine A inverse once we know the singular value decomposition, A inverse is simply v sigma inverse and u times u Hermitian. And since sigma is a diagonal matrix, inverting it is trivial, you just have to invert the each of the diagonal entries.

(Refer Slide Time: 5:54)





Another interesting property of the SVD is that it diagonalizes any system of linear equations. So, for example, if we are given Ax equals b, this is a system of linear equations, then substituting A equals u sigma v transpose we get u sigma v transpose x equals b. And if we define C to be the matrix u transpose times, the vector u transpose times v, so I am pre multiplying this by u transpose. And what I get on the right side u transpose b, and I call that my C.

And I partition c as like this with the c1 containing the first r entries in c and c2 containing the remaining m minus r entries in c. This is the same as representing this vector b in the basis defined by u, so change of basis, because u is an orthonormal matrix, multiplying by u transpose corresponds to a change of basis. And c is basically the same as b but in a new basis, represented in a new basis.

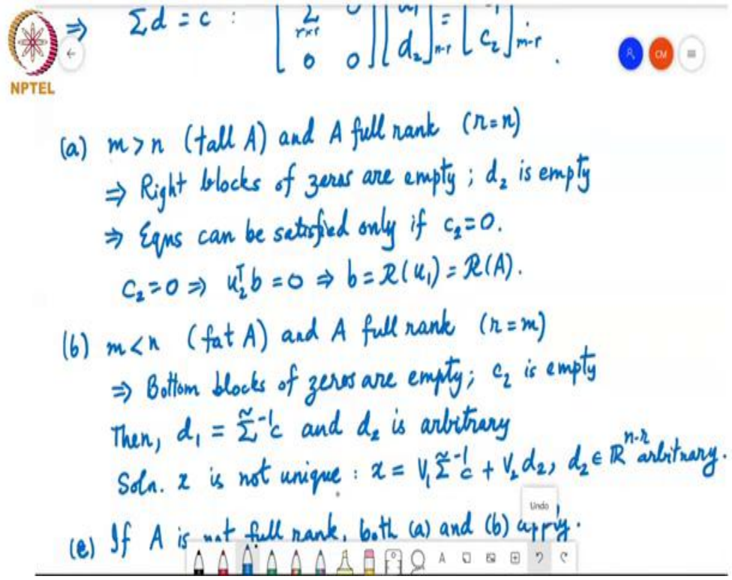
Similarly, if I define d to be this v transpose x, and I partition it as d1 d2 where d1 contains the first r entries, and d2 contains the remaining n minus r entries, this is a representation of x in the basis v. So, if I do these substitutions, then the system of linear equations reduces to sigma times d equals c, which is, which looks like this, it is sigma tilde 0 0 0 times d1 d2. So, this is r cross r. And this has r entries, this has n minus r entries, this has r entries, and this has m minus r entries.

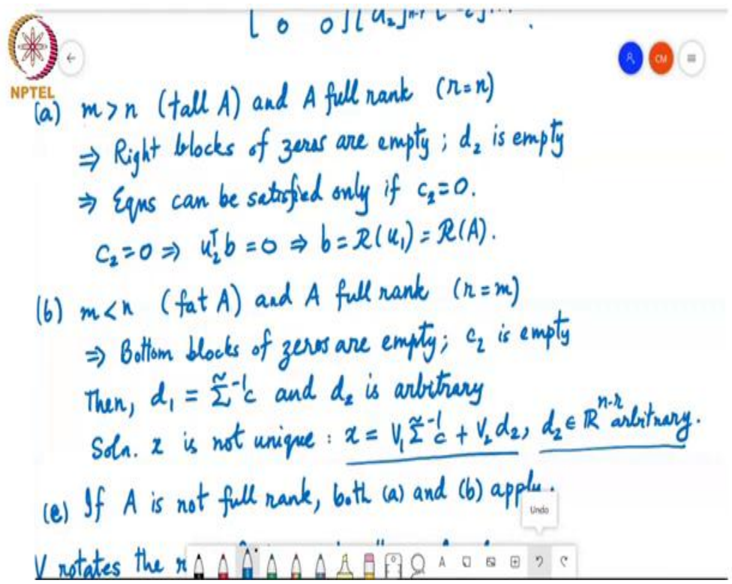
So, this is now a very simple system of linear equations. And in fact, we can visually solve this system of linear equations. In particular, if m is greater than n, that is, the number of rows is more than the number of columns, then the matrix A is tall. And if you look back at the system of linear equations, this means you have more equations then you have variables that you need to solve for.

And in particular, if you assume A is full rank, that is, the rank of this matrix is equal to the number of columns in A which is n, then, then there will not be any right block here because r is equal to n, there is no such thing as d2, and there is no right block over here. So, it is just sigma tilde times d equals c1 c2.

And of course, these equations because d is only multiplying 0s in these rows, these equations can only be satisfied if c2 equals 0 and c2 equals 0 in turn means that, it means that u2 transpose times b equals 0 because c2 is just exactly equal to u2 transpose times b. And if u2 transpose b equals 0, it means that b must belong to the range space of u1 or in other words the range space of A, we already saw that the range space of A is this the span of u1. So, that is clear right if you have more equations, then if you have more equations, then you have unknowns, there exists a solution only if b lies in the range space of A.

(Refer Slide Time: 10:01)



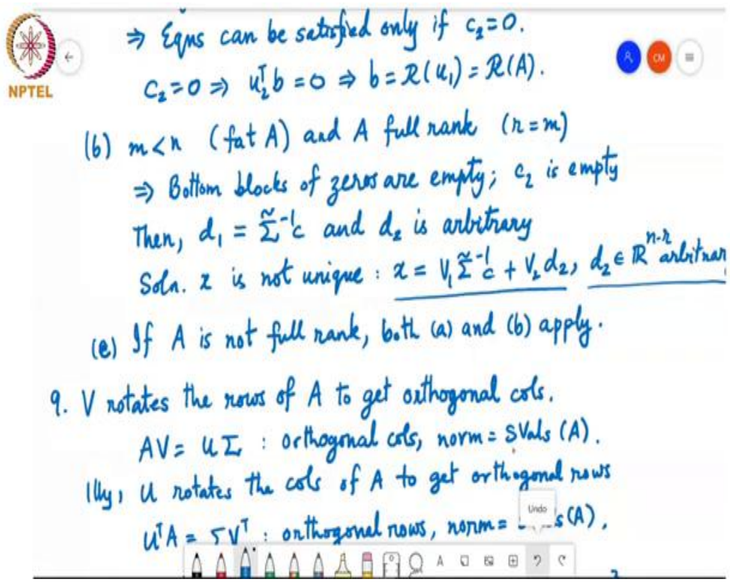


Similarly, if m is less than n that is the matrix A is fat, then you have more variables, then you have equations. And if A is full rank that is r equal to m, then there are no bottom blocks here. So you will have sigma tilde or d1 d2 is equal to c, there is no bottom, there is no c2 here. Then, if we define d1 to be sigma tilde inverse times c, this is allowed because we have assumed A is full rank. So, sigma tilde has all nonzero entries.

And so, if we, so, in that case, we can d1, the solution d1 is just sigma tilde inverse times c and d2 can be anything because it is anyway just going to multiply 0 here, and so, the solution is not unique. And in general we can write the solution to be v1 times sigma tilde inverse times c plus v2 d2, where d2 is arbitrary. So, this is the complete set of solutions of this system of equations.

So, in this case, there is always a solution in fact, there are always infinitely many solutions. If a is not full rank, then both of these apply. So, you have to you use both these conditions. So, there may or may not exist solutions.

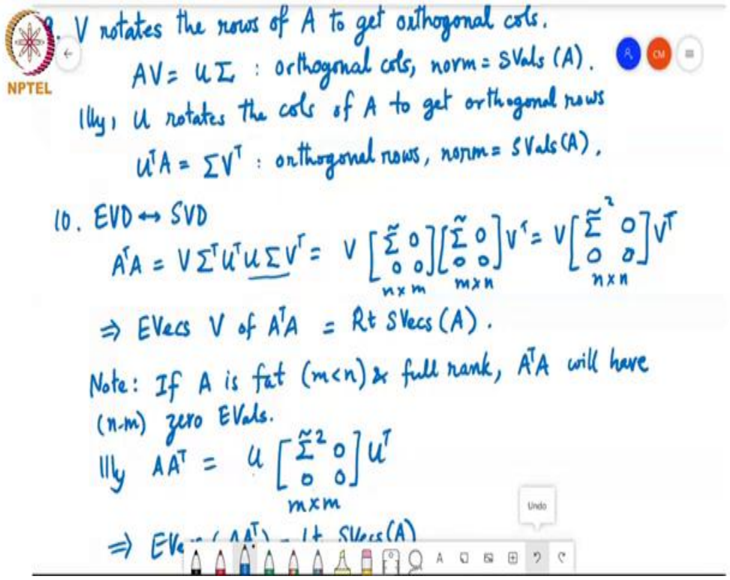
(Refer Slide Time: 11:38)



Another property is that v; it rotates the rows of A to get orthogonal columns. So, in particular, since A equals u sigma v transpose, if we write multiply by v, A times v becomes equal to u times sigma. Now, the columns of u are orthonormal. And then this is a diagonal matrix. So, this matrix u times sigma has orthogonal columns with each column norm being equal to 1 of the singular values of this matrix A.

Similarly, if I take u transpose A, A is u sigma v transpose, so u transpose u becomes the identity matrix. So, what I am left with is sigma v transpose, which in turn has orthogonal rows with the row norms being equal to the singular values of this matrix A.

(Refer Slide Time: 12:33)



So, the last time somebody asked me a question about the connections between the eigenvalue decomposition and the singular value decomposition, the explicit connection is exactly this that if I consider the matrix A transpose A, again A is equal to u sigma v transpose, then A transpose will be v sigma transpose u transpose.

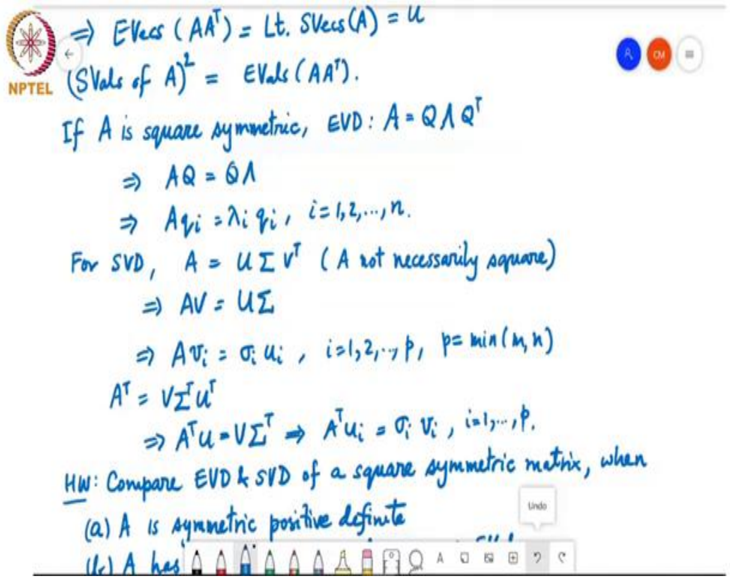
So, A transpose A is v times the sigma is sigma tilde 000. And since I am writing it as sigma tilde, I do not need a transpose here this is just a diagonal matrix its transpose is itself, but the matrix overall is of size n by m, not m by n. Whereas, this is the sigma matrix which is sigma tilde 0 0 0, which is of size m by n times v transpose.

Now, if I carry out this multiplication, I will get the matrix sigma squared as the top left r cross r matrix and then 0s everywhere else, so, that the overall size of the matrix is n cross n and I have v, v transpose. So, from this we see that the eigenvectors. So, basically this is the eigenvalue decomposition of A transpose A. So, the Eigenvectors v of A transpose A are the right singular vectors of A.

And of course, this also shows that if A is fat then A transpose A will have and if it is actually full rank, then all these entries will be nonzero. So, there will be m nonzero Eigenvalues for A transpose A and the remaining n minus m Eigenvalues will be 0s. Similarly, if you write out what AA transpose is, that is u times sigma tilde squared 0 0 0, but now, this is of size m by m.

So, the 0s are of appropriate dimensions, so that the overall dimension is m by m times u transpose.

(Refer Slide Time: 14:31)



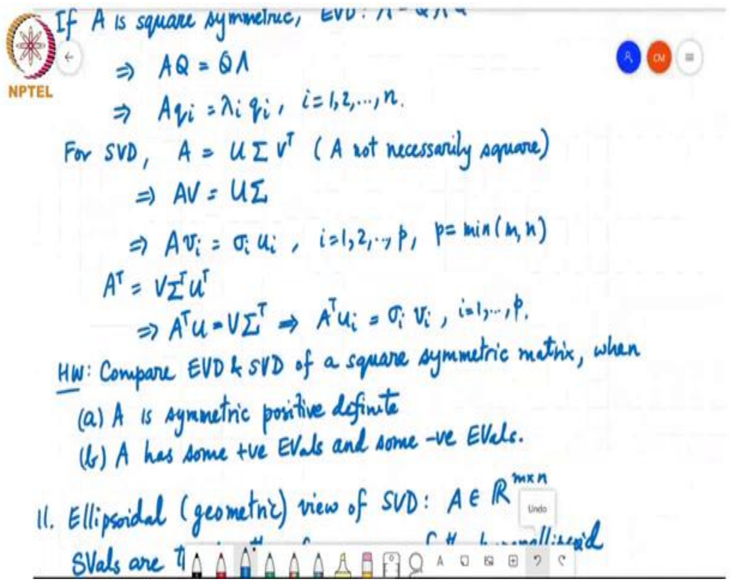
And so the eigenvectors of AA transpose of this u, which are equal to the left singular vectors of A. And of course, the singular values of A square are the eigenvalues of AA transpose. So, basically the price paid for being able to diagonalize an arbitrary matrix is that we need two matrices to diagonalize the matrix instead of one, if it was A, if A was square and symmetric, we can find the eigenvalue decomposition like this Q lambda Q transpose, where lambda is a diagonal matrix containing the Eigenvalues and Q is an orthonormal matrix containing the eigenvectors.

So, basically A Q is equal to Q lambda or A times the ith column of Q is equal to lambda i times Qi, i equal to 1 to n. So, this is the eigenvalue decomposition. So, you need only one matrix to diagonalize this A, but for a single, for an arbitrary matrix, which is not necessarily square, we can find an SVD, a singular value decomposition like this A equal to u sigma v transpose. And of course, this means that A times v equals u times sigma or A vi equals sigma i ui. So, these are, vi are unit norm vectors and ui are unit norm vectors.

So, these are kind of like Eigenvectors of the matrix but not really because this ui is not equal to this vi, but they are unit norm vectors, they are in fact of different dimensions vis are of dimension n by 1 whereas, uis are of dimension m by 1. Similarly, A transpose equals v sigma u transpose taking the transpose of this and this matrix I should actually write it as sigma transpose u transpose.

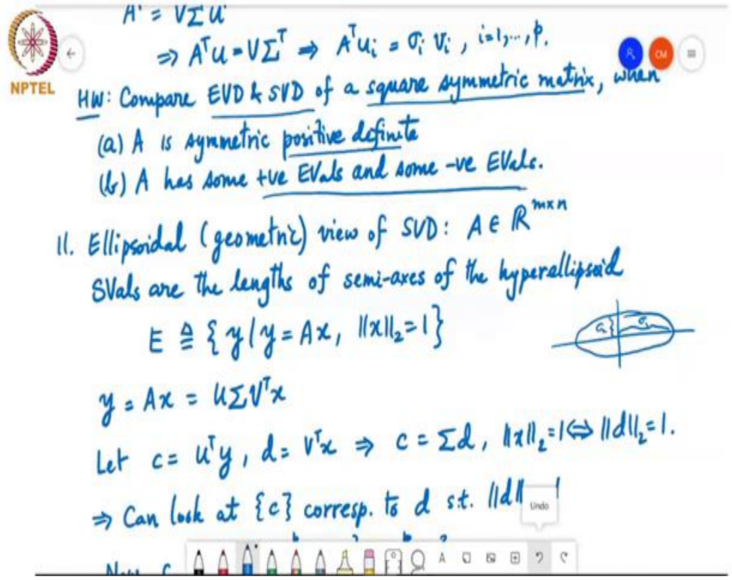
So, then if I right multiply by u I have A transpose u equals v sigma transpose then A transpose ui is equal to sigma i times vi. And so, again so this is the kind of relation you have A vi equals sigma ui and a transpose ui equal sigma i vi that is the relation you have for the singular value decomposition.

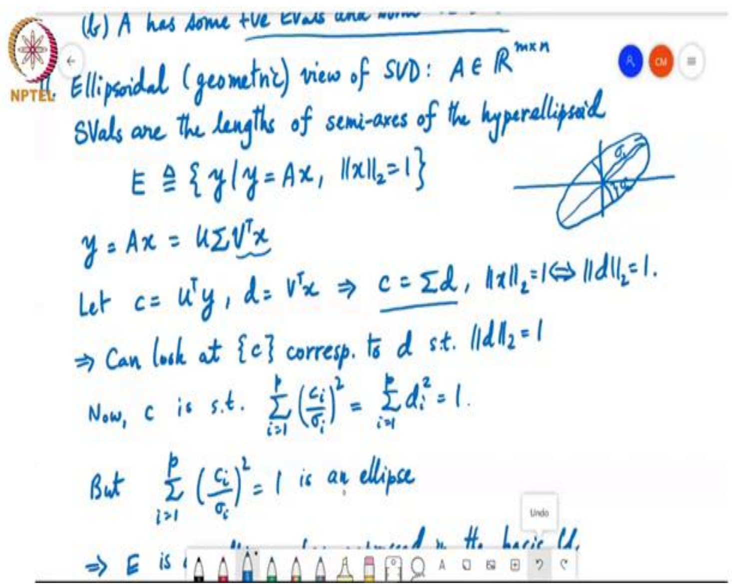
(Refer Slide Time: 17:02)



What you should do on your own is to compare these two, compare the Eigenvalue decomposition and the singular value decomposition when the matrix is square symmetric, and positive definite, or in the case where it is indefinite that it has some that is it has some positive and some negative Eigenvalues.

(Refer Slide Time: 17:30)





There is another geometric view or the ellipsoidal view of the singular value decomposition. So, again take a matrix which is of size m by n, let us look at this space defined by the vectors y, where y is equal to Ax for some x, which is of unit l2 norm. And let us look at what this, what the space of y looks like. It turns out that the space E.

So, that is you take any vector which is of unit l2 norm, you compute Ax and you look at, you plot that point in the m dimensional space, and look at all the points y that you can reach by this operation. It turns out that set of yi describe an ellipse are actually in m dimension, it is called a hyper ellipsoid. And these singular values of the matrix A, are the length of the semi axis of this hyper ellipsoids.

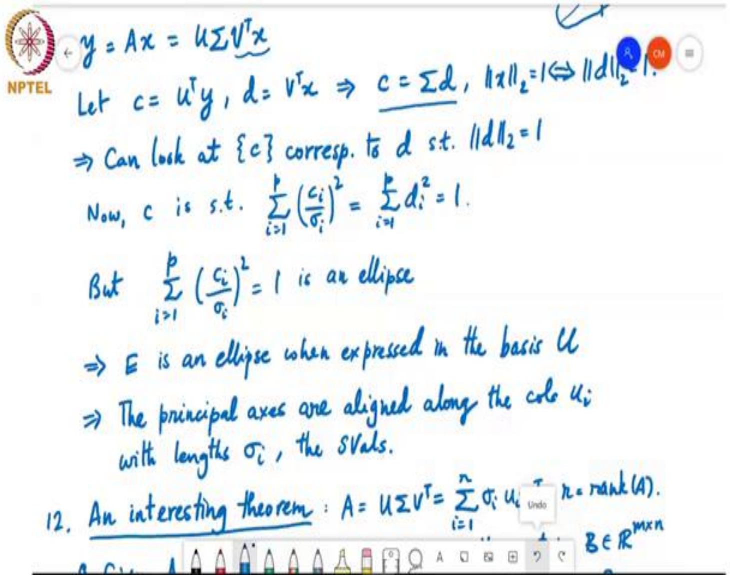
What I mean by that is for example, if I take a two dimensional ellipse, then these lengths are the lengths of the semi axis, actually just to be a little more clear, let me draw it in a different way. Suppose, there is an ellipse that goes like this, then this is called the semi major axis and this is called the semi minor axis and this length will be sigma 2 and this length will be sigma 1.

So, let us see that. So, suppose if I consider y equals Ax, and now I use the singular value decomposition is u sigma v transpose x, then once again if I define u transpose y to be c and v transpose x this part to be d, then this equation reduces to c equal to sigma times d. And further, if x is of unit l2 norm because v is an orthonormal matrix, it means that, instead of searching over or instead of going over all points such that x2 equals 1 I can as well go over all points such that d2 equals 1.

So, we can look at the set of c vectors that I get corresponding to d satisfying norm d2 equals 1. Now, because c equals sigma times d norm d2 equals 1 simply means that summation i equal to 1 to p, where p is the number of nonzero singular values in the matrix A ci over sigma i square, which is the same as sigma equal to 1 to p di square. This is diagonal, that is why I am able to write this is equal to 1.

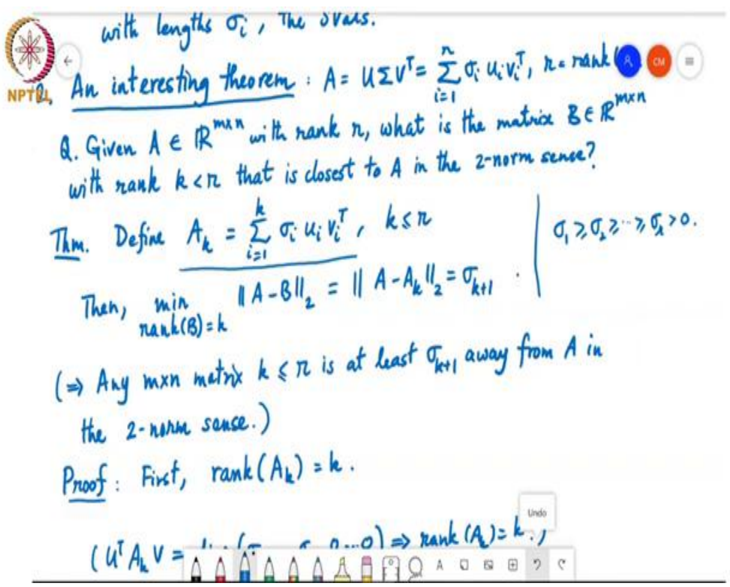
Now, the set of points c is actually exactly described by the c that satisfies this equation. And this equation, sigma i equal to 1 to p ci over sigma i square equals 1 is an ellipse or a hyper ellipsoid.

(Refer Slide Time: 20:45)



So basically, E and this going from c to y is just all it takes is multiplying by u, which is an orthonormal matrix, it is just a orthonormal change of basis. So, E is also an ellipse or a hyper ellipsoid when expressed in the basis u. And the principle axis of this, when you write an ellipse, equation of ellipse in this elementary form like this, then sigma i are the semi major axis of this hyper ellipsoid. So, the principle axes are aligned along the columns ui, and have length sigma i, which are the singular values of the matrix A.

(Refer Slide Time: 21:35)



There is one more very, very interesting and useful theorem associated with the singular values, singular value decomposition. So, suppose A equals u sigma v transpose, which we can, if we expand this product out, we can write it as sigma i equal to 1 to r sigma i ui vi transpose. So, in fact, we see that the matrix A can be written as the sum of r matrices of rank 1, ui vi transpose is a matrix of size m by n and it has rank 1. And so A can be written like this.

And the question we ask is given this matrix A with rank r, what is the matrix B with rank equal to k which could be less than r which is in fact less than r that is closest to A in the two norn-sense? So, that the answer to this is given by the following theorem. Define Ak to be the summation i equal to 1 to k sigma i ui vi transpose that is I am stop stopping this summation from i equal to 1 to r instead of going all the way up to r I am only going up to k. And that matrix I define as Ak.

Then the minimum over all B, matrices B of rank at most k, rank equal to k of A minus B l2 norm is equal to the norm of A minus this particular matrix Ak which is in turn equal to sigma k plus 1. In other words, Ak is the rank, it is the matrix of rank k which is as close to A as possible in the two norm sense. And how far is it from A? It is sigma k plus 1 away from A.

Recall that for all this to I mean the assumption here is that sigma 1 is greater than or equal to sigma 2 is greater than or equal to etcetera up to sigma r is greater than 0, these are ordered. So, it means, what this means is that in words what this theorem is saying is that the closest rank k matrix to A in the two norm sense is given by Ak, this matrix here.

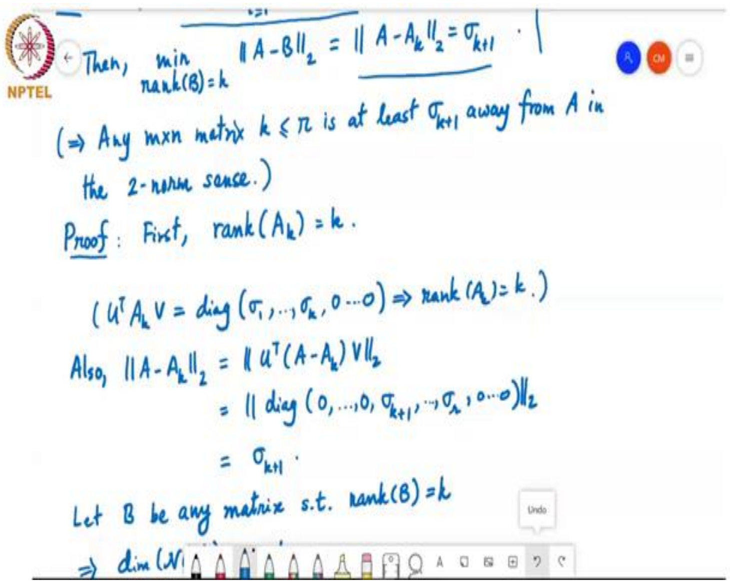
And this matrix is found by excluding the contribution of the smallest singular values from, in the expansion of this matrix A given here. It also means that any m by n matrix of rank k less than or equal to r is at least sigma k plus 1 away from A in the two norm sense, so you cannot find a matrix of rank k which is strictly less than r that is arbitrarily close to A, it will be at least sigma k plus 1 away from A.

Now, the proof of this theorem is actually very interesting, but before I walk you through the proof, note that what this is saying is that the solution to this optimization problem is Ak. This by itself is actually fairly, it is not obvious how you would show something like this, it is saying that you need to show that when you solve this optimization problem, you will get Ak as the solution.

So, the way you actually prove theorems like this often is that you show a lower bound on this quantity, if you look at all matrices of rank equal to k, what is the smallest value that A minus B 2 can achieve? And then you show that this Ak as defined here actually attains this lower bound. So, that basically establishes the theorem.

So, this kind of tricks can be used when you somehow have a guess of what the answer to the optimization problem is. So, I show that kind of problem is actually always, almost always easier than find the solution to this optimization problem.

(Refer Slide Time: 26:11)

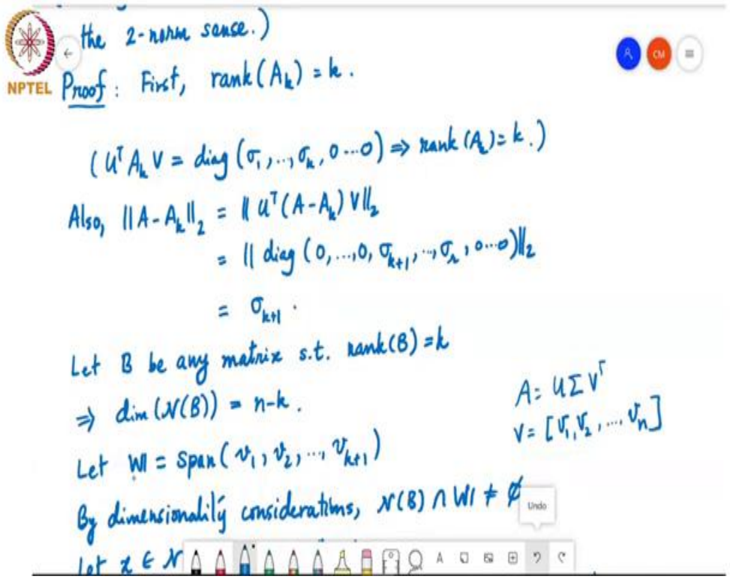


So, the proof goes like this, first of all, the rank of Ak as defined here is equal to k, that is simply because if I do u transpose Ak times v, I get this diagonal matrix containing sigma 1 through sigma k along the diagonal, and now this clearly has rank equal to k. So, rank of Ak equals k.

Also, if you look at the l2 norm of A minus Ak, this l2 norm is invariant to left or right multiplication by unitary matrix, so, I can consider u transpose A minus Ak times v which is substituting for A to be u sigma v transpose and Ak to be u some sigma dash v transpose, you can see that the first k singular values will just cancel off and what you will be left with is a diagonal matrix containing sigma k plus 1 through sigma r.

And of course, the largest Eigenvalue of this diagonal matrix is just sigma k plus 1. So, this norm is exactly equal to sigma k plus 1. So, that shows that, that shows this part of the theorem, A minus Ak is sigma k plus 1.

(Refer Slide Time: 27:52)

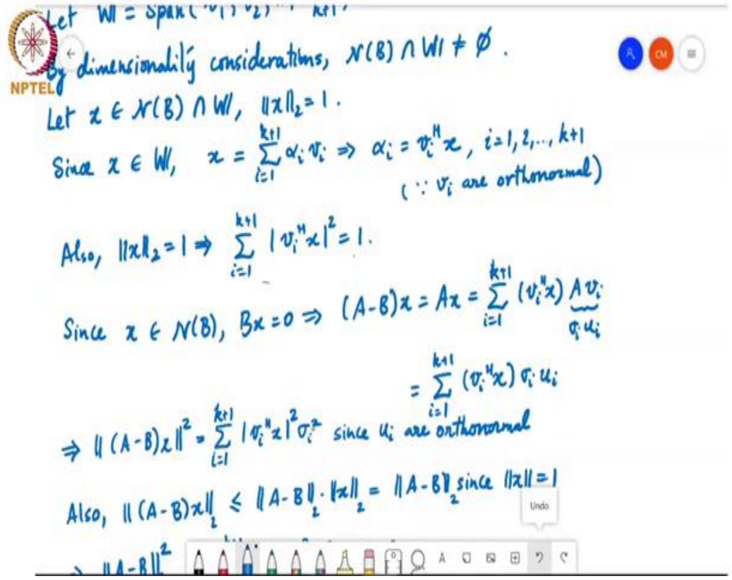


Now, what we need to show is that there is no other matrix that can have an l2 norm of the difference being bigger than sigma k plus 1, sorry being smaller than sigma k plus 1. In other words, Ak itself is a matrix that is closest to A among all rank k matrices in this l2 norm sense.

So, suppose B is a matrix such that rank of B equals k which in turn means that by the rank nullity theorem, the dimension of the null space of B is equal to n minus k. Define W to be the span of v1 through v k plus 1, the first k plus 1 columns of v. So, A is equal to u sigma v transpose and v is v1, v2, vn. So I take these first k plus 1 columns here. And the span of this is the subspace W.

Now, the dimension of this subspace W is k plus 1, the dimension of the null space of B is n minus k. So, the sum of the two dimensions is more than n. And therefore, by dimensional reconsiderations, we see that the null space of B intersection W cannot be the null set, there has to be some nonzero vectors in it.

(Refer Slide Time: 29:24)



So, suppose we pick one such vector x, which belongs to the intersection of these two sets subspaces which is of unit norm x2 equals 1. Now, this x of course belongs to the null space of B and it belongs to W. So, since it x, x belongs to W, we can write x to be a linear combination of these k plus 1 columns here, so, we will call that i equal to 1 to k plus 1 alpha i vi.

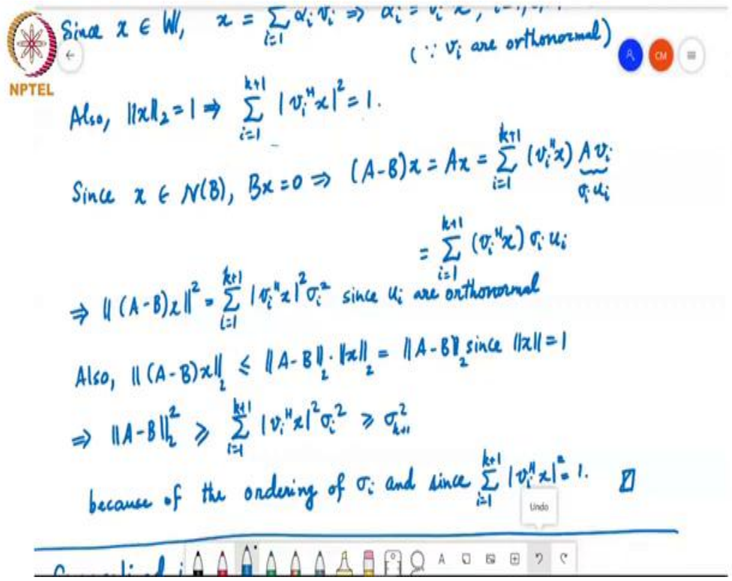
And these vi are orthonormal. And so, we can actually directly find these alpha i to be vi Hermitian times x, i equal to 1 to k plus 1. Now, since the l2 norm of x equals 1, it means that if I take vi trans, vi Hermitian x square these when you square and add them, that should also be equal to 1, you can simply take this Hermitian times itself, all the cross terms drop off because these vi are orthogonal to each other. And you are left with sigma, alpha i, mod of alpha i square, which is the same as sigma vi Hermitian x squared, i is equal to 1 to k plus 1 and that should be equal to 1.

On the other hand, since x also belongs to the null space of B, it means that B times x equals 0, which means that A minus B times x is equal to Ax because Bx is equal to 0. And if I write out what Ax is, so all I am doing here is I am just substituting sigma i equal to 1 to k plus 1 alpha i vi for x. So, I get sigma i equal to 1 to k plus 1, alpha is just a scalar. So I can bring it out front. And it is equal to vi Hermitian x times A vi.

But we already saw that A vi is the same as sigma i ui. And so I can write it as i equal to 1 to k plus 1, vi Hermitian x times sigma i times ui. And so basically, if I look at norm square of A minus vx, this is equal to sigma i equal to 1 to k plus 1 vi Hermitian x squared times these vectors are orthonormal.

So, once again, if I take this thing squared all the cross terms will drop off and all the direct terms will have ui Hermitian ui, which is equal to 1 there. So, I have vi Hermitian x squared times sigma i squared.

(Refer Slide Time: 32:04)



And so and further by the sub multiplicativity property, A minus B x l2 norm is less than or equal to the l2 norm of A minus B times the l2 norm of x. But the l2 norm of x equals 1, we started out by saying that let x be a vector such that l2 norm equals 1. So this is less than or equal to the l2 norm of A minus B.

And so this is bigger than or equal to this, which has in turn equal to this. And so A minus, or rather the square of this. So it is equal to the square root of this. So, A minus B l2 norm squared is greater than or equal to the square of this, which is equal to this, sigma i equal to 1 to k plus 1 vi Hermitian x squared times sigma square.

Now, these sigma i are in decreasing order. And this vi Hermitian x squares add up to 1. So if I replace all of these sigma i squares with sigma k plus 1 the smallest value here, then sigma k plus 1 can come out. But this summation i equal to 1 to k plus 1 vi Hermitian x squared is equal to 1, that is what we wrote over here.

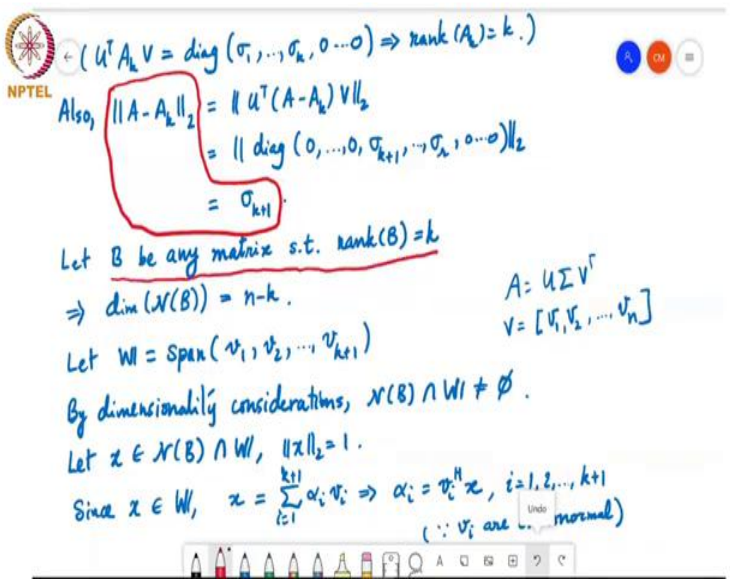
And so this is greater than or equal to sigma squared k plus 1. So, that means that if I had taken any other B, which is a rank k, then A minus B squared is at least equal to sigma k plus 1 square. And we have found the matrix Ak which is exactly at sigma k plus 1 squared distance from A and so that proves the result. Any questions so far?

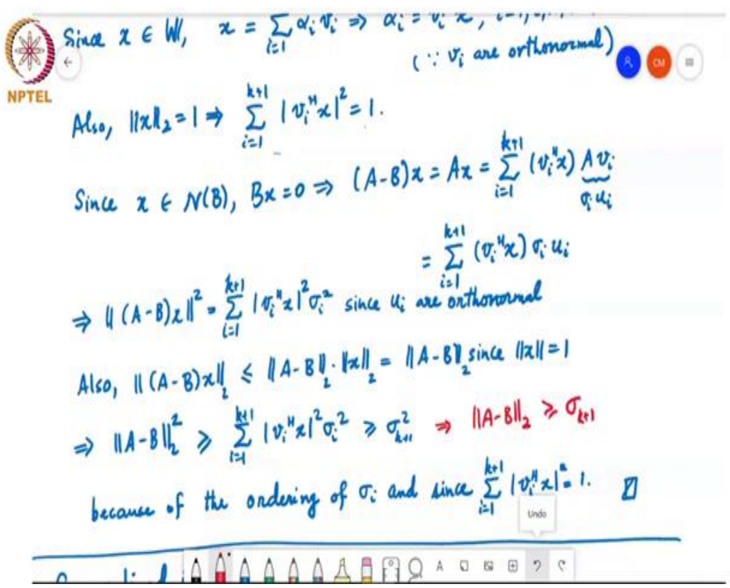
Student: Sir.

Professor: Yes.

Student: You have not proved the first part, the matrix B which minimizes this problem is Ak.

(Refer Slide Time: 34:15)





Professor: No, what I am saying is these two points together, we made two points. The first point is that let me go back here. The first point is that A minus Ak l2 norm is equal to sigma k plus 1. So, Ak is exactly at it is a matrix of rank k, which is exactly at sigma k plus 1 distance from A. And then we said let B be any other matrix such that it has rank equal to k. Then what we showed is that this matrix B is at least equal to sigma k plus 1 distance.

Student: Yes sir. Okay, I understand.

Professor: So, now, basically, I can challenge you to say find me a matrix, which is of rank k which is less then sigma k plus 1 distance away, you will not be able to, because I already proved a lower bound, that any other matrix, any matrix B of rank k is at least sigma k plus 1 away from A, can we found the matrix Ak which is at sigma k plus 1 distance.

And therefore, this matrix Ak that we found is a solution to the optimization problem of finding a matrix B of rank equal to k which is as close as possible to A in the l2 norm sense. Of course, there may be other solutions, but this is at least, this is certainly one solution.

Student: Yes sir. Okay. Thank you.

Professor: So, if there are no other questions, I move on to the next topic which is generalized inverses of matrices.

Student: Sir.

Professor: Yeah.

Student: Sir A minus B into norm for fixed A and variable B if it is convex function then it is the only solution, we can say.

Professor: Yes.

Student: Okay Sir.

Professor: So, you need two things to prove that this uniqueness based on convexity, you need that the cost function is convex. And if I take A minus B l2 norm squared that is indeed convex in B. However, you also need that the constraint set is a convex set. And clearly you can see that if I look at the space of all matrices of rank or matrices B of rank k, if I take a convex combination of two such matrices that need not be ranking. And so, this space is a complicated space, it is not a convex space. And because of that, you cannot use convexity based arguments to say that this is a unique solution.

Student: Okay Sir I understand. Thank you.

Professor: Welcome.