Matrix Theory

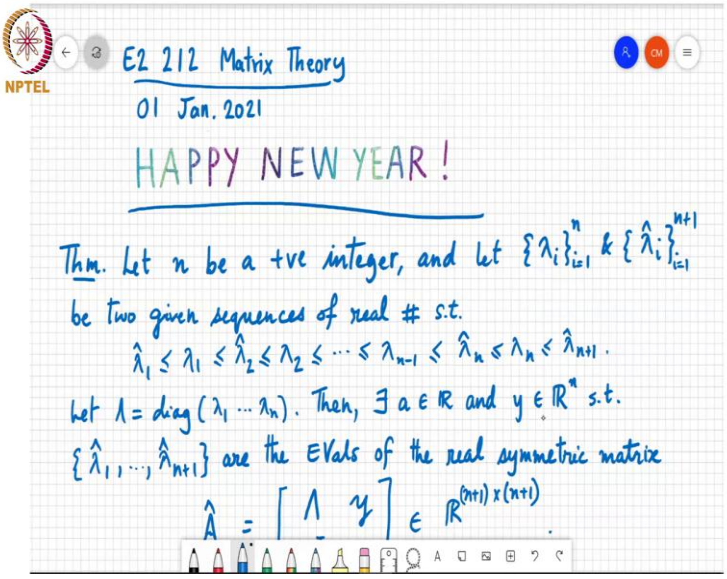
Professor Chandra R Murthy

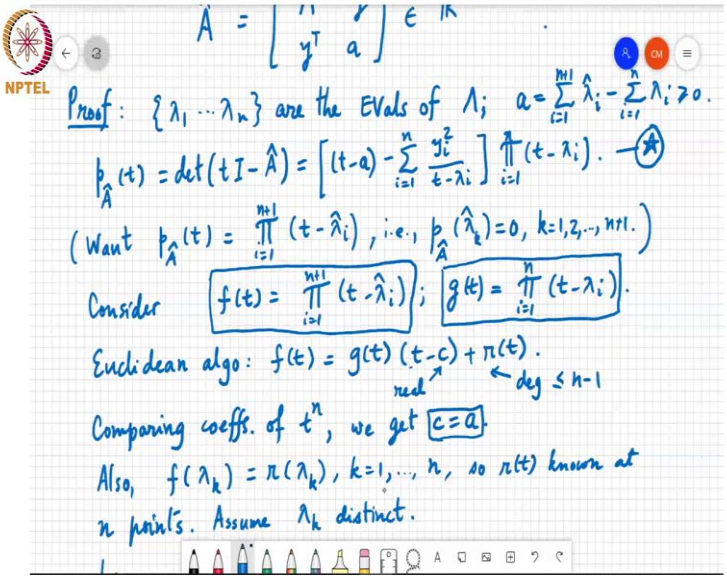
Department of Electrical Communication Engineering

Indian Institute of Science Bangalore

Interlacing theorem (Continued)

(Refer Slide Time: 00:14)





So, we begin. The last time we were looking at this theorem, which said that if n is a positive integer and lambda i going from 1 to n and lambda hat i, i going from 1 to n plus 1. Two sequences of real numbers, such that they (interlace), these two sequences interlaced with each other, meaning that lambda hat 1 is the smallest of these numbers followed by lambda 1 followed by lambda hat 2, followed by lambda 2, etcetera, up to at the very end, you will have lambda n minus 1 is less than or equal to lambda hat, lambda hat n is less than or equal to lambda n is less than or equal to lambda hat n plus 1.

So, lambda hat n plus 1 is the biggest, lambda hat 1 is the smallest and all the other numbers are in between. And if we denote the diagonal matrix containing lambda 1 to lambda n as its diagonal entries as this capital lambda, then there exists a real scalar, real value scalar, and an n length real-valued vector y such that these other numbers lambda hat 1, lambda hat 2, up to lambda hat n plus 1 are the eigenvalues of the real symmetric matrix A hat, which is lambda y, y transpose A, which is an n plus 1 cross n plus 1 matrix.

So, we were going over the proof of this, we had gone most of the way, but there is some part that needed to be completed. So, we will begin by filling in the rest of this proof. So, first of all, lambda 1 through, just to recall where we were lambda 1 through lambda n, of course, the eigenvalues of lambda.

And further, by noting that the trace of A hat must be the summation of lambda hat i, while the trace of lambda must be equal to lambda 1 plus etcetera lambda n, we know that this value A here, in order for lambda hat 1 through lambda hat n to be the eigenvalues of this matrix, A must be such that A is equal to summation i equal to 1 to n plus 1 lambda hat i minus the summation i equal to 1 to n lambda i.

This is always going to be greater than or equal to 0. Now, if we look at the characteristic polynomial of A hat by definition, it is the determinant of tI minus A hat. And if we substitute that and expand it out, we were able to show that it can be written in the following form where it has a factor with t minus a minus yi squared over this t minus lambda i is showing up in the denominator here, times the product of these terms, i equal to 1 to n t minus lambda i.

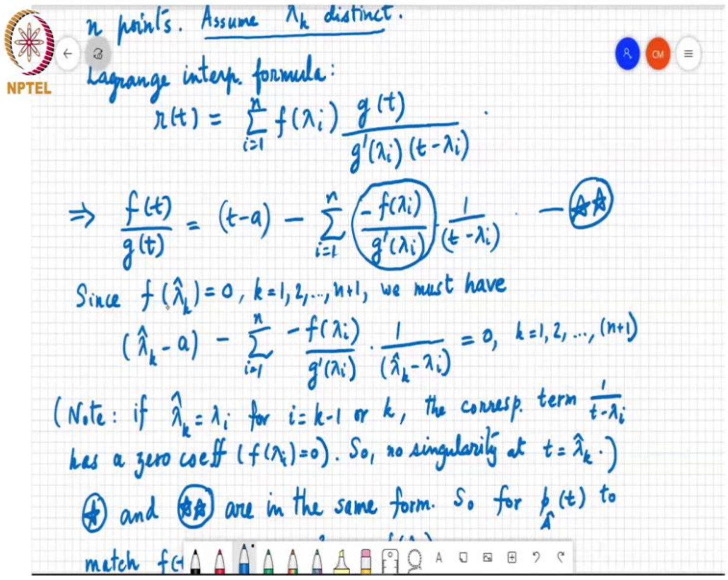
And note that this is exactly the characteristic polynomial of this matrix lambda. Now, what we want is that this pA hat of t must end up being equal to the product i equal to 1 plus 1 to n plus 1 t minus lambda hat i, that will ensure that lambda hat i going from 1 to n plus 1 are the 0s of pA hat of t. And so that this matrix we can then be assured that this matrix has lambda hat 1 through lambda hat n plus 1 as its eigenvalues. Now, to do that, we consider two functions.

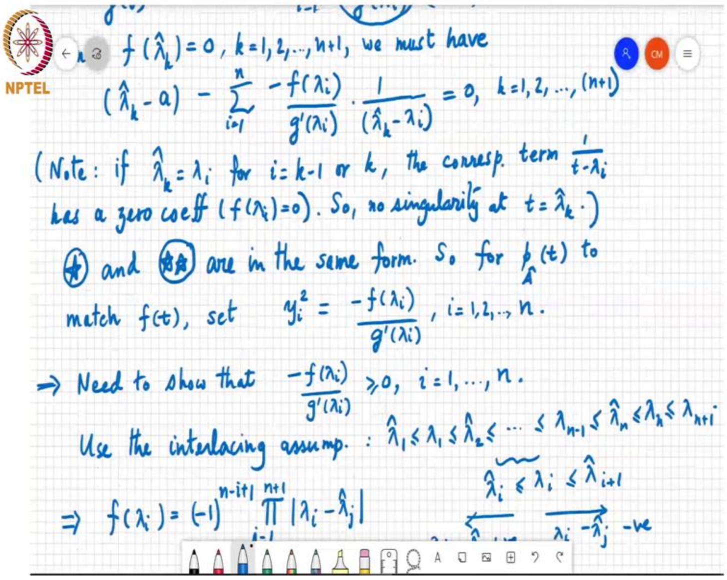
The first is f of t, which is the product this is our desired characteristic function, which is the product i equal to 1 plus 1 to n plus 1 t minus lambda hat i and g of t is the characteristic function of capital lambda. And that is just the product i equal to 1 to n t minus lambda i. Now, this is an n plus 1 degree polynomial, this is an n degree polynomial. So, we can always write f of t to be some g of t times t minus some constant c.

That is because the t power n plus 1 term has a coefficient of 1 and the t power n has a coefficient of 1 here. And so, it can be written as G of t times t minus c plus some r of t this is a remainder polynomial and this has a degree at most n minus 1. Now, if so, obviously, the t power n plus 1th coefficients here match already by construction, but if we compare the coefficients at t to the n, we did this the last time we ended up with saying that c must be equal to a.

Now, further, because g of lambda k is equal to 0 for k going from 1 to n because there is a t minus lambda i factor here, what we have is f of lambda k must be equal to r of lambda k. So, what this means is this r of t is known to us at n points because we know what f of t is, it is just this polynomial, so we can substitute lambda k, k going from 1 to n. And we can calculate the value of f of lambda k, that tells us what the value of r of lambda k is, and it is a degree at most n minus 1 polynomial, and it is known now at 1 k at n different points.

(Refer Slide Time: 05:54)





Now, if we assume that lambda ks are distinct, and I, as I mentioned will, at the end, talk about what happens when there are repeated lambda ks. But if they are distinct, we can write the following Lagrange interpolation formula. So, knowing the value r of lambda k at k equal to 1 to n, we can directly write out what r of t is, r of t is this polynomial here. And notice that it is, it has a g of t divided by t minus lambda i, and g of t itself is this product of all such factors.

So, each factor will cancel one of these factors here. And so, each of these ratios is a degree n minus 1 polynomial. And that is getting weighted by f of lambda i divided by g dash of lambda i, and then added together. So overall, this has a degree at most n minus 1. But this is the expression for r of t. So, if you find r of lambda k, k going from 1 to n, you will find that it is equal to f of lambda k.

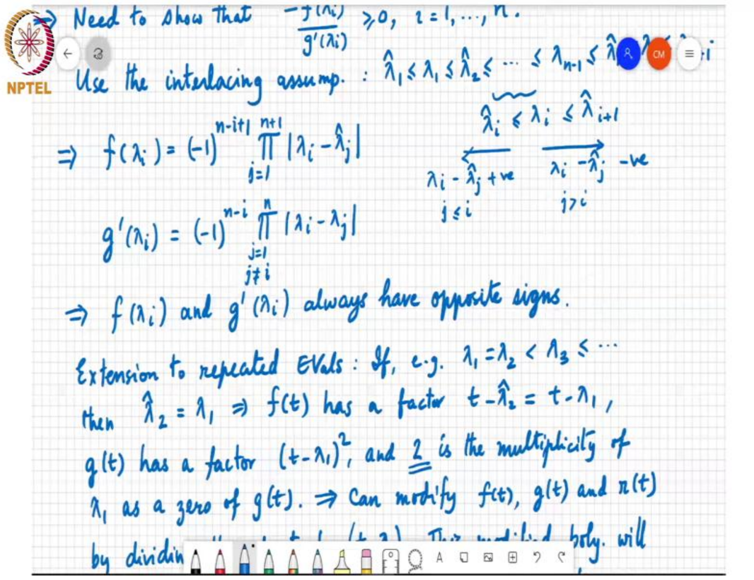
So, it meets this constraint that we have set on these on r of t. Now, if we use this formula here, f of t equals g of t into t minus c plus r of t, and we divide throughout by g of t, then we get that f of t divided by g of t is equal to t minus a, c is equal to a minus this whole thing divided by g of t and the g of t, g of t will cancel and so you are left with, so here it was plus r of t, so I write it as minus of minus, minus i equal to 1 to n, f minus f of lambda i divided by g dash of lambda i times 1 over t minus lambda i.

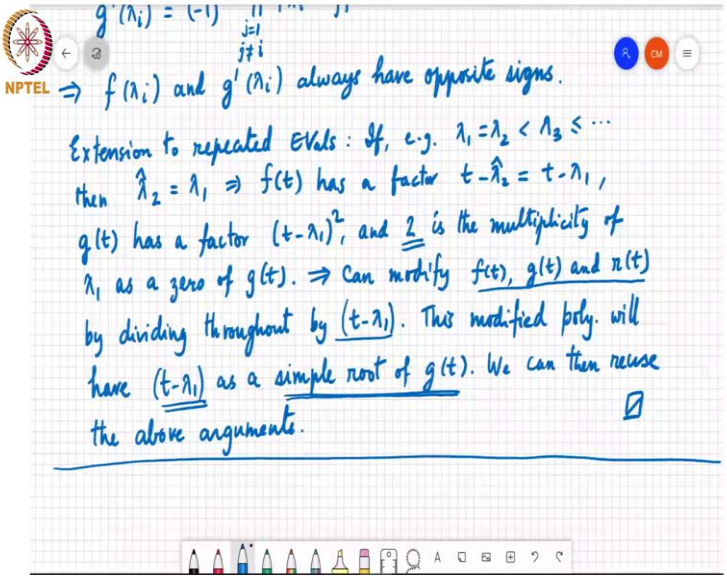
Now this, this thing is, so if I write it this way, f of t is equal to this whole coefficient times g of t, and g of t is this product here. And so now, if you look back at pA hat of t, that also has the same form, it is t minus a minus something times the product, this thing, which is exactly equal to my g of t. So, f of t is now equal to the same is, f of t is now in the same form as this pA hat of t.

And so, for these two polynomials to completely match, we just need to choose yi squared to be equal to this coefficient here, minus f of lambda i over g dash of lambda i. So, yeah, so since f of lambda k hat equals 0, for i equal to 1 to n plus 1, we must have that if I substitute lambda k hat here, this must be equal to 0 for i k equal to 1 to n plus 1. This is another condition that this polynomial ratio here will satisfy.

And as I said this two are the same forms. So, if you want these two to match, all we need is to set yi squared to be equal to this coefficient minus f of lambda i over g dash of lambda i. So, for you to be able to find a real-valued vector y, such that yi squared equal to the negative of this, we need that these numbers should be positive numbers, then I can find a real-valued yi such that yi squared equals this thing. So, we just need to show that these are all greater than or equal to 0 for i equal to 1 to n.

(Refer Slide Time: 09:42)





So, we now finally use the interlacing assumption. Notice that in the proof, so far, we are not used this fact that this lambdas and lambda hats are an interlacing set of numbers. So, we will use this interlacing assumption. And now, if I take the, so you can see here that lambda 1 is between is lower bounded by lambda hat 1 and upper bounded by lambda hat 2 and so on. So, more generally, lambda i is lower bounded by lambda hat i and upper bounded by lambda hat i plus 1.

So, what this means is that if I take any lambda hat j, where j is less than i, then lambda i minus lambda hat j will be a positive number this minus something which is on this side. And similarly, if I take any j which is greater than i, then lambda i minus lambda hat j is going to be negative. This is a smaller number. These are all the lambda hats on this side that are bigger numbers.

And so, based on that, if I look at what f of lambda i is, f of lambda i is the product of lambda i minus lambda j hat j going from 1 to n plus 1, but out of these all these numbers are positive and all these numbers are negative. So, there are n minus i plus 1 negative numbers over here. And so, if I pull that out, I can write this as minus 1 power n minus i plus 1 product j going from 1 to n plus 1 the modulus of lambda i minus lambda j hat.

And similarly, g dash of lambda i is just the same product. So, g dash is this t minus lambda j but not including the ith term because you are taking the derivative and evaluating it at lambda i. And so, this product and again all these terms will be positive and all these terms will be negative. So, there are n minus i such negative terms. So, I can write it as minus 1 to the n minus i times the product of all positive numbers, mod of lambda i minus lambda j.

And so, you will see that this is multiplied by minus 1 to the n minus i plus 1 this is multiplied by minus 1 to the n minus i times a positive number. And so, what that means is that f of lambda i and g dash of lambda i will always have opposite signs because there is a plus 1 extra here. And so, their ratio will be negative, or negative of that ratio will be positive.

So, that establishes that we can choose yi squared to be minus f of lambda i over g dash of lambda i. So, the only thing left is to handle the case where there are repeated eigenvalues. So, this is a very simple argument. Suppose, for example, that lambda 1 equals lambda 2 which is strictly less than lambda 3, and so on. For all other cases, the argument is very similar.

So, the you can consider an example like this and see what happens. Now, if that is the case, because it has this interlacing property if lambda 1 equals lambda 2 then lambda 2 hat must be equal to lambda 1 and which is in turn equal to lambda 2. So, that means that f of t which is the product of t minus lambda hat i, it has a factor t minus lambda 2 hat, which is the same as t minus lambda 1.

Similarly, g of t has two factors which are equal to t minus lambda 1. So, it has the factor t minus lambda 1 square because the first term and the second term are both t minus lambda 1 and (t), so the first term and second term at t minus lambda 1 into t minus lambda 2, but since lambda 1 equals lambda 2 you have a factor like t minus lambda 1 squared. And so, the multiplicity of lambda 1 as a 0 of g of t is exactly equal to 2.

And so, what we can do is to modify our f of t, g of t, and r of t by dividing throughout by t minus lambda 1 and this modified polynomial will be exactly like before, it will have t minus lambda 1 as a simple root of g of t. And that was the reason why we assumed that the eigenvalues are distinct, so, that the eigenvalues turned out to be simple roots, and then we can reuse all of the above arguments.

So, that is the is the proof. So, the proof directly extends to the case where there are repeated eigenvalues. So, this argument the main point of this argument here is that g of t always has one greater degree for repeated root compared to f of t and that means that we can remove these common factors, and then we will be left with distinct eigenvalues and we can reuse the same arguments as we used to establish this.

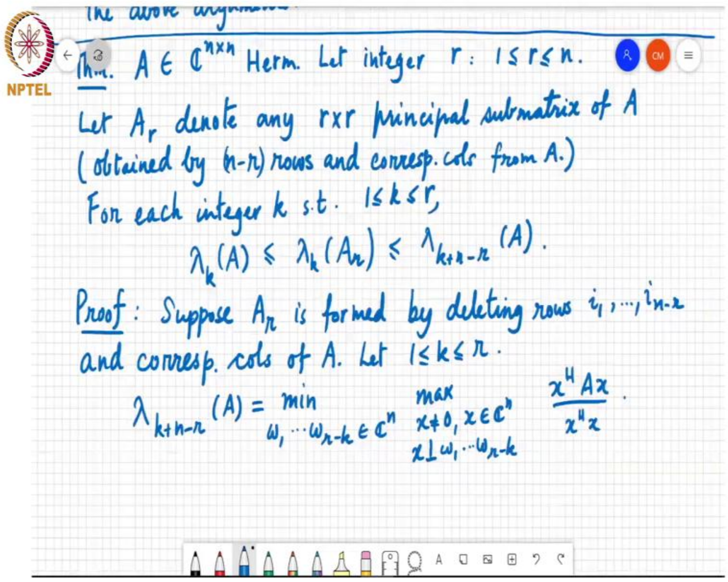
Any questions about this proof? So, what did this result tell us, it told us that if you are given a set of interlacing numbers, we can construct matrices such that the first set of numbers is are the eigenvalues of a matrix. And the second set of numbers are the eigenvalues of a matrix that is obtained by taking the first matrix and bordering it on the right and below by y and y transpose respectively and on the bottom right by some number a.

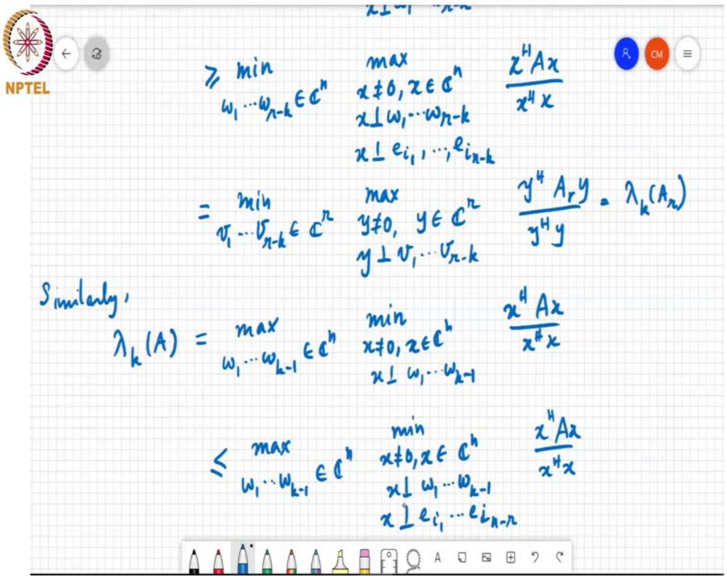
So, and the result prior to that said that if you take a matrix A and then you border it by a vector y and on the right and y Hermitian below and small a on the bottom right, you will get an n plus 1 cross n plus 1 matrix and the eigenvalues of the n cross n matrix and the n plus 1 cross n plus 1 matrix interlaced with each other.

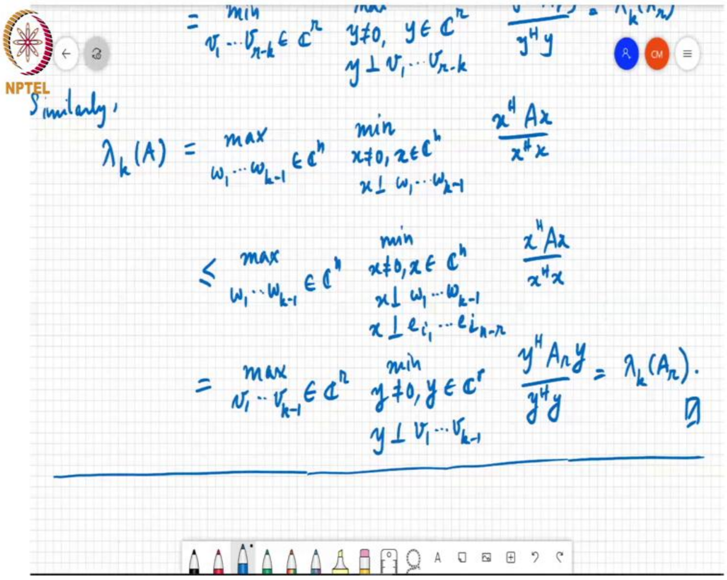
Well, there is no nothing very special about adding a row and column to a matrix, the same result applies to deleting a row or a column or in fact, the last row or column of a matrix. When you delete the last row or column, the eigenvalues of the reduced matrix will interlace with the eigenvalues of the original matrix. And of course, if you think about it, there is not no sanctity in deleting the last row or column nothing special about the last row or column, the results apply equally well if you delete any row or column.

So, when you delete any row or column you, so, if you take an n cross n matrix and delete a particular row and column you get what is called an n minus 1 cross n minus 1 principle sub matrix of that matrix and the eigenvalues of the principles of matrix interlaced with the eigenvalues of the original matrix and you can apply this repeatedly and then you will get interlacing results for that (happens), you will get results related to the interlacing of eigenvalues when you delete say r rows or say k rows and columns of a matrix. And so, this kind of results are called inclusion principles. And here is one such example result.

(Refer Slide Time: 17:46)







So, this is theorem. You can prove this theorem by using this result on adding a row and column and converting it to deleting a row and column and applying it repeatedly. But then it is equally easy to prove this result directly which is what we will do. So, the theorem says that A is an n cross n Hermitian matrix and let r be an integer such that 1 is less than or equal to r is less than or equal to n.

Let Ar denote any r cross r principal submatrix of A. So, then you obtained the principles of submatrix by deleting n minus r rows and the corresponding columns. So, you have to delete the same row and column index, so then for each integer k 1 less than or equal to k less than or equal to r lambda k of A less than or equal to lambda k of Ar is less than or equal to lambda k plus n minus r of A. Again, we are arranging these eigenvalues in increasing order.

So, I did not write that explicitly, but lambda 1 of A is the smallest eigenvalue of A and lambda n of A is the largest eigenvalue of A, and lambda k of Ar is the kth largest eigenvalue of the principal submatrix Ar. So, proof. So, the essential ideas of this proof we have already seen, so, we can quickly run through the proof. So, suppose Ar is formed by deleting rows, it will delete n minus r rows.

So, i 1 up to i n minus r and the corresponding columns of A and let 1 less than or equal to k less than or equal to R. Now, just using the Courant-Fischer theorem, we have that. So, first of all, we want to show a lower bound like this. So, lower bound, as you remember, as you might remember, we will use the min-max formulation of the result. So, lambda k plus n minus r of A is equal to the min over w 1 through w n minus this index, and n minus this index is n minus k minus n plus r, which is r minus k.

These are in C to the n, the maximum over x not equal to 0, x in C to the n, and x perpendicular to all these vectors of x Hermitian Ax divided by x Hermitian x. Now, this trick is something we saw earlier, this is greater than or equal to the minimum over the same vectors w 1 through w r minus k in C to the n, the maximum over x not equal to 0, x in C to the n, x perpendicular to w 1 through w r minus k.

Now, I will throw in an extra set of constraints, x perpendicular to ei1 up to ein minus k. So, these are columns of a n cross n identity matrix corresponding to columns i 1 through i n minus k. So, I am adding extra constraints here. So, this maximum, the solution to this maximization problem may not be as big as the solution to this maximization problem. And that is why we have a greater than or equal to sign here.

So, the objective function is the same, x Hermitian Ax over x Hermitian x. Now, if x is perpendicular to all these vectors, then what I can do is to simply delete that corresponding row and column of A and then consider a reduced vector and then solve this over that reduced space. So, that reduced vector I will call it y and y will be perpendicular to these vectors w 1 through w n minus k but with the indices, i 1 through i n minus k delete it, and those I will call v 1 to v r minus k.

So, this is exactly equal to the minimum over vectors v 1 through the r minus k in C to the r, the maximum over y not equal to 0, y in C to the r, and y perpendicular to v 1 through v r minus k of y Hermitian Ary over y Hermitian y, which is exactly by the Courant-Fischer theorem itself equal to lambda k of Ar. So, that proves this part of the inequality. And similarly, lambda k of A. Now, we want to show an upper bound, so we will use the max-min version.

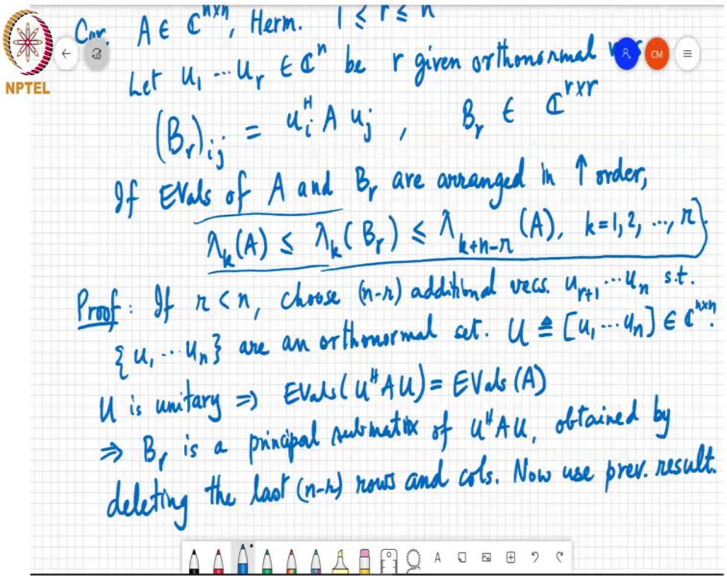
So, this is the maximum over w 1 through w k minus 1 in C to the n of the minimum over x not equal to 0, x in c to the n, x perpendicular to w 1 through w k minus 1 of x Hermitian Ax over x Hermitian x. And now, I use the same trick again that this is less than or equal to if I throw an extra constraint on the minimum, I may not be able to minimize it as well as I am doing it here.

So, the final answer may turn out to be larger than whatever I get here. So, this is the max over w 1 through w k minus 1 in C to the n of the minimum x not equal to 0, x in C to the n x perpendicular to w 1 through w k minus 1. And now, x also perpendicular to ei1 through ein minus r of x Hermitian Ax over x Hermitian x. What is x is perpendicular to all these the entries i1 through in minus r of x are always equal to 0.

So, those particular rows and columns from A can just be deleted, because there is nothing to optimize over there. And correspondingly, I can delete those indices in w 1 through w k minus 1 and call them, call the resulting r-dimensional vectors as v 1 through v k minus 1. So, this is exactly equal to the maximum over v 1 through v k minus 1 in C to the r the minimum over y not equal to 0, y belonging to C to the r and y perpendicular to v 1 through v k minus 1 of y Hermitian Ar y over y Hermitian y which is exactly equal to lambda k of Ar, which is the result we wanted to show.

So, that completes this proof. So, one immediate consequence of this, this result is something called the Poincare separation theorem. And this is very useful for example, in quantum mechanics in situations where one has information about ui Hermitian Auj for orthonormal vectors ui and uj. So, you do not get to observe the whole matrix A, but you get to observe projections of this matrix using a set of orthonormal vectors, and then what can we say about the eigenvalues of the original matrix in terms of the eigenvalues of the matrix obtained by forming these projections.

(Refer Slide Time: 28:58)



So, here is the corollary, A is in C to the n cross n Hermitian and 1 less than or equal to r less than or equal to n. So, r is some number between 1 and n let u1 through ur in C to the n be r given orthonormal vectors. And define the matrix Br with ijth element equal to ui Hermitian Auj. And notice that Br is a matrix in C to the r cross r. If the eigenvalues of A and B are arranged in increasing order lambda k of a is less than or equal to lambda k of Br is less than or equal to lambda k plus n minus r of A.

So, what this says is that instead of observing A, if we get to observe ui Hermitian Auj and then you arrange that in a matrix Br and then you find the, so this is a smaller matrix possibly a much smaller matrix of size r cross r, then you can say something about the eigenvalues of A.

So, specifically, the kth eigenvalue of A is at most the kth eigenvalue of Br and the k plus n minus r eigenvalue of A is at least equal to the kth eigenvalue of Br. So, this allows you to bound the eigenvalues of A in terms of lambda k of Br. So, the proof is very short it is just pointing to the previous results that we are going to use.

So, if r is strictly less than n, so, if r equals n, that is almost trivial but if r is less than n, I mean the next step that I am going to say is not required if r equals n, because u1 to un if r equals n, u1 to un are form an n cross n orthonormal matrix together when you stack them into a matrix, but if r is less than n, we can choose n minus r additional vectors ur plus 1 up to un such that this set u1 through un form an orthonormal set. And let u be defined this is matrix u1 through un in C to the n cross n.

Now, u is unitary. So, which implies that the eigenvalues of u Hermitian Au are equal to the eigenvalues of A and that means that the given Br is a principal submatrix of u Hermitian Au, just obtained by deleting the last n minus 1 or n minus r rows and columns. So, that is it. So, now, we can use the previous theorem, just deleting n minus r rows and columns. And we can use the previous theorem and that is exactly what the previous theorem said that the eigenvalues have this relationship.

So, in fact, this result itself is very useful, we can use it to show many more results on many more variational results on eigenvalues and there are lots of results in the text. We just do not have the time to systematically cover them in this course, but you can look at the text and find many more interesting results.

Now, recall that the diagonal elements of Hermitian symmetric matrix are always real because if it is summation symmetric, A equals A Hermitian and if we equate the diagonal elements, it means A must be, Aii must be equal to Aii star. In other words, diagonal entries must be real-valued. Also, the eigenvalues of a Hermitian symmetric matrix are real-valued, we have seen that also. Furthermore, the diagonal entries and the eigenvalues have the same sum that is the trace of a matrix is equal to the sum of its eigenvalues.