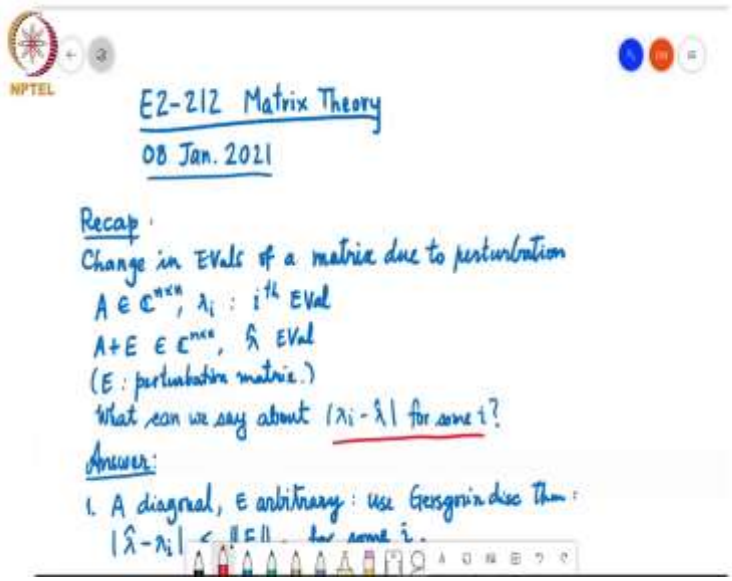
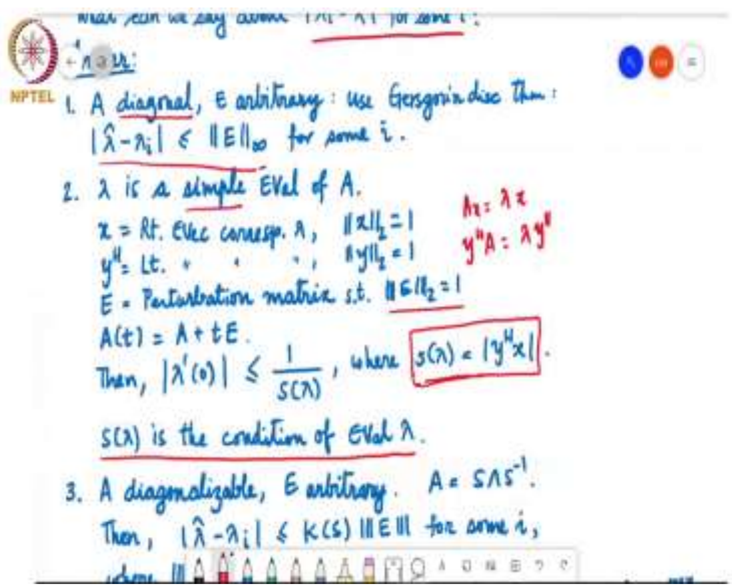
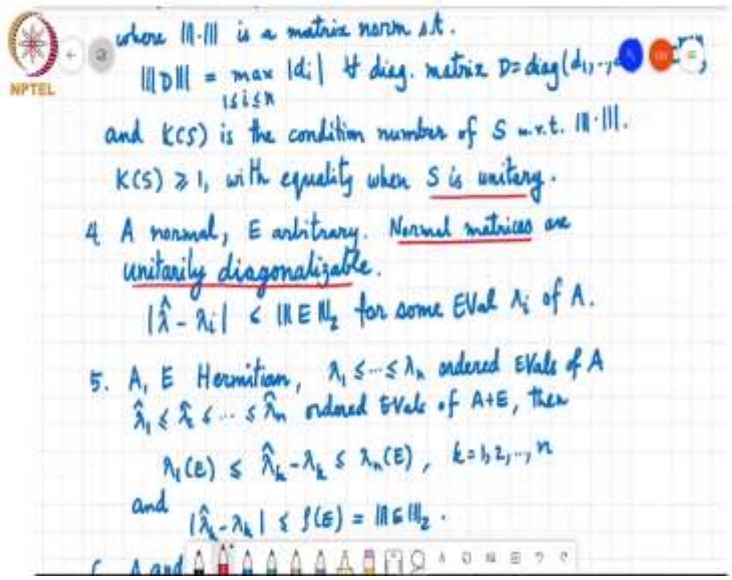
Matrix Theory
Professor Chandra R. Murthy
Department of Electrical Communication Engineering
Indian Institute of Technology Bangalore
Perturbation of Eigenvalues Birkhoff's theorem Hoffman - Weilandt Theorem

(Refer Slide Time: 00:14)







So, good afternoon and let us begin. Just to recap the last time we saw some other consequences of the Gersgorin Disc theorem and then we started discussing about perturbation location and perturbation of Eigen values namely that the eigenvalues of a matrix change when the matrix is perturbed. And so the starting point is that we are given a matrix A of size n cross n and it has Eigen values lambda 1 lambda 2 up to lambda n.

And this matrix A is getting perturbed by a matrix E to get the matrix A plus E and suppose we compute the eigenvalues lambda hat 1 up to lambda hat n or in particular if some Eigen value is lambda hat what can we say about how close lambda hat will be to one of the eigenvalues of the matrix A that is can we say something about lambda hat minus lambda i modulus value for some i that is will is value be small for some one of the Eigen values of A and the various ways to answer this question.

First of all, if A happens to be a diagonal matrix, if A happens to be a diagonal matrix, then we can simply directly use the Gersgorin Disc theorem to say that this difference between lambda hat and one of the Eigen values is at most the L infinity norm of E, this is what we saw the last time. And the other case is when lambda is a simple Eigen value of A, meaning that it has an algebraic multiplicity equal to 1.

Then if we denote x to be the left Eigen, sorry the right eigenvector corresponding to lambda that is Ax equals lambda x and y to be a left eigenvector corresponding to lambda that is y Hermitian A equals lambda y Hermitian, then if E is a perturbation matrix such that its spectral norm equals

1, then, if we denote A of t to be A plus t times E, then we can study what happens to the, how sensitive the Eigen values of A by looking at the modulus of the derivative of the eigenvalue at t equals 0.

So, we showed that this is less than or equal to 1 over S of lambda, where S of lambda is defined to be the magnitude of y Hermitian x and S of lambda we call it the condition number or the condition of the eigenvalue lambda. So, this is for simple Eigen values. So, it is an Eigen value of A with algebraic multiplicity equal to 1.

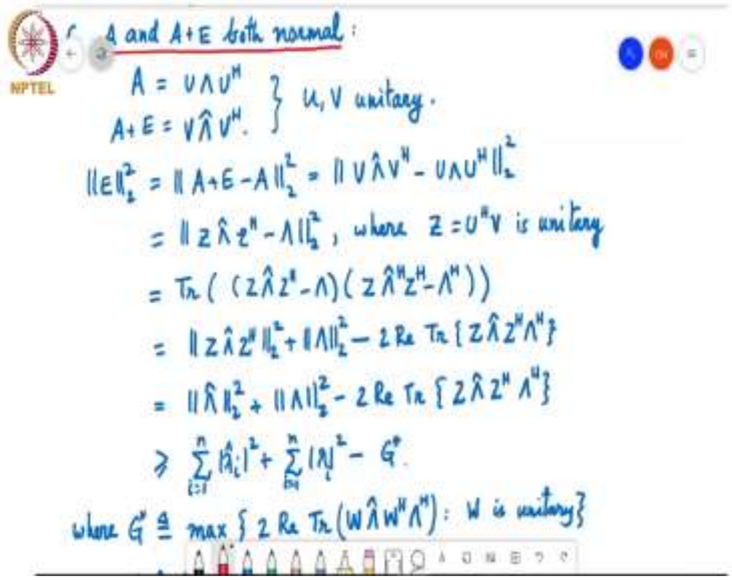
Then we looked at the case where A is diagnosable matrix and E can be anything that means A can be written as S lambda S inverse for some invertible matrix A S and a diagonal matrix lambda. Then we showed that lambda hat minus lambda i is at most K of S times the norm of E for some i, where this norm is any matrix norm such that the norm of a diagonal matrix is the maximum magnitude diagonal entry in the matrix.

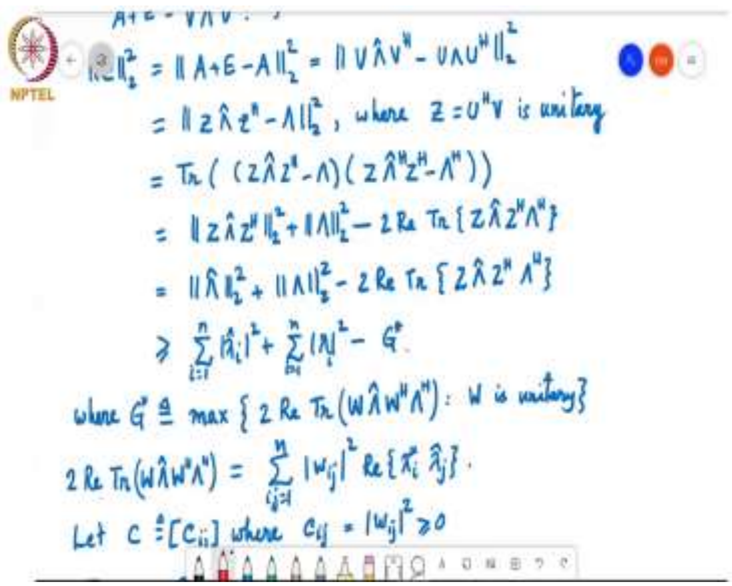
And K of S here is the condition number of this matrix S with respect to this norm used here, of course, if condition number is greater than or equal to 1 and it is equal to 1 when S is unitary. So, if the matrix is unitary diagnosable, then we have that lambda hat minus lambda i is less than or equal to norm of E itself, so this K of S is equal to 1. And so then normal matrices are unitarily diagonalizable.

So, as a consequence if A is a normal matrix and E is arbitrary, then we have that lambda hat minus lambda i is at most be here we taken the spectral norm of E for some eigenvalue lambda i of A. The next one is that suppose A and E are both Hermitian symmetric matrices. Then, if we denote lambda 1 to lambda n to be the ordered Eigen values of A and lambda hat 1 through lambda hat n to be the ordered Eigen values of A plus E, then we can lower bound lambda had k minus lambda k.

So, this is taking the kth largest Eigen value of A plus E and subtracting the kth largest Eigen value of A itself and that is at least equal to lambda 1 of E and at most equal to lambda n of E and further this magnitude of this difference is at most the spectral radius of E. So, this is where we stopped the last time.

(Refer Slide Time: 05:28)





$G^* \triangleq \max\{2\,\mathrm{Re}\,\mathrm{Tr}(W\hat{\Lambda}W^H\Lambda^H) : W \text{ is unitary}\}$

$2\,\mathrm{Re}\,\mathrm{Tr}(W\hat{\Lambda}W^H\Lambda^H) = \sum_{i,j=1}^{M} |w_{ij}|^2\, \mathrm{Re}\{\lambda_i^* \hat{\lambda}_j\}$.

Let $C \triangleq [c_{ij}]$ where $c_{ij} = |w_{ij}|^2 \geq 0$

$\sum_i c_{ij} = \sum_j c_{ij} = 1$ (since W is unitary)

$\Rightarrow C$ is a doubly stochastic matrix.

Thus, $G^* \leq \max\{2\sum_{i,j=1}^{M} c_{ij}\, \mathrm{Re}\{\lambda_i^* \hat{\lambda}_j\} : C \text{ is doubly stochastic}\}$

Linear in C.

Let $\mathcal{D}_s \triangleq \{C : C \text{ is doubly stochastic}\}$.

$\mathcal{D}_s$ is convex (show!)

The max of

$\sum_i c_{ij} = \sum_j c_{ij} = 1$

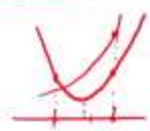
$\Rightarrow C$ is a doubly stochastic matrix.

Thus, $G^* \leq \max\{2\sum_{i,j=1}^{M} c_{ij}\, \mathrm{Re}\{\lambda_i^* \hat{\lambda}_j\} : C \text{ is doubly stochastic}\}$

Linear in C.

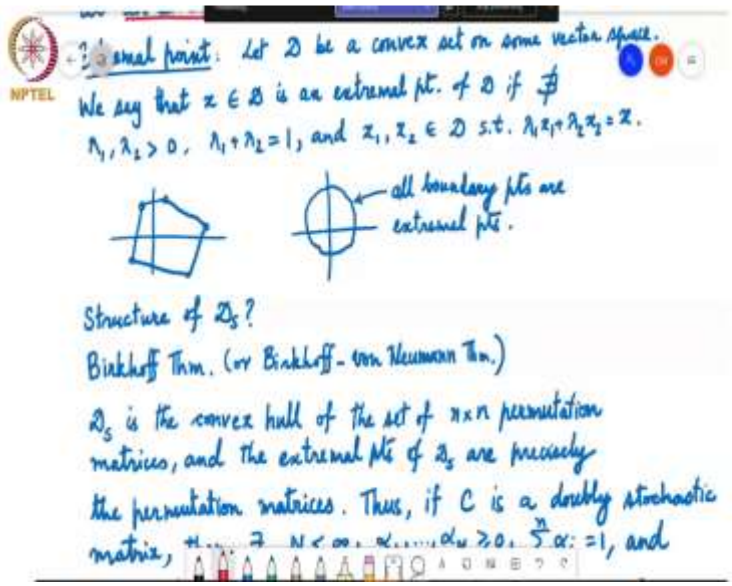
Let $\mathcal{D}_s \triangleq \{C : C \text{ is doubly stochastic}\}$.

$\mathcal{D}_s$ is convex (show!)



The max. of a convex fn. over a convex set is attained at an extremal point of the convex set.

Extremal point: Let $\mathcal{D}$ be a convex set on some vector space. We say that $x \in \mathcal{D}$ is an extremal pt. of $\mathcal{D}$ if $\nexists$ $\lambda_1, \lambda_2 > 0$, $\lambda_1 + \lambda_2 = 1$, and $x_1, x_2 \in \mathcal{D}$ s.t. $\lambda_1 x_1 + \lambda_2 x_2 = x$.



Now, we consider one more case where A and A plus E are both normal matrices. What that means is that we can write A as U lambda u Hermitian and A plus E as V lambda V Hermitian where U and V are both unitary matrices because normal matrices can be unitary diagonalized.

Now, if we look at the Frobenius norm square E2 square then that is the Frobenius norm square of A plus E minus A which I can write as U lambda hat u Hermitian just substituting for A plus E and minus U lambda U Hermitian just substituting for A Frobenius norm square and what I can do is since the Frobenius norm square of a unitary matrix is no.

So, I can pull out U on the left and the U Hermitian on the right and since the since multiplication by left or right multiplication by unitary matrix does not change the Frobenius norm of a matrix I can get rid of U and U Hermitian and write it as the norm of Z lambda hat Z Hermitian minus lambda square, where Z which is equal to U Hermitian V is a unitary matrix.

So, norm, this is a unitary matrix and this Frobenius norm square we know that it can be written as trace of so, Frobenius norm squared of a trace of A A Hermitian so, I just write that here. So, this is trace of Z lambda hat Z Hermitian minus lambda times the Hermitian of this, which is Z Hermitian Hermitian is just Z lambda hat Hermitian z Hermitian minus lambda Hermitian.

If I just expand this out, this gives me this times this which is just the Frobenius norm of Z lambda hat Z Hermitian square and then this times this is just the Frobenius norm of lambda square minus 2 times trace of the inner product between this and this, which is Z lambda hat Z Hermitian times lambda Hermitian.

Now, again this is multiplication by a unitary matrix so, I can get rid of this Z on the left and Z Hermitian on the right and write this as lambda hat 2 square plus lambda 2 square minus 2 real part of trace of this matrix the same as the previous equation. So, what are we trying to do here we are trying to see how relate this E2 square to the Eigen values of A and A plus E, so that we can ultimately bound E2 square.

So, for example, here you can already see that, this Frobenius norm squared, this is just a diagonal matrix. So, it is just the sum of the diagonal entries squared. This is also a diagonal matrix. So, it is just the sum of the diagonal entries square and then this time I will come to that in a sec. But the point is that on the right hand side, I have terms that depend only on the Eigen values of A and A plus E.

And if I can find an expression for this quantity, which only depends on the eigenvalues of A and A plus E, then now I will have a upper bound, which connects Eigen values of A and A plus E with the Frobenius norm squared of the error matrix or the perturbation matrix. So, that is the final goal is to find some expression for this quantity, which depends only on lambda hat and lambda i.

Now, this quantity itself is as something whatever it is, but if I replace it with something bigger than I am only making this whole expression smaller, so I can say that it is greater than or equal to these two terms minus a quantity G star, where I define G star to be the maximum that this can attain over all possible unitary matrices Z.

So, G star is the max of this thing 2 real trace of W lambda hat, W Hermitian lambda summation over all matrices W which are unitary. Now, if I simply expand this out and take into account the fact that lambda and lambda Hermitian are both diagonal matrices, I can write this 2 real trace of W lambda hat W Hermitian lambda Hermitian as the summation over i, j going from 1 to n of the modulus of Wij square times the real part of lambda i star times lambda j hat is there a question. So, this is like this.

Now, suppose we define a matrix C with entry Cij, Cij is equal to this quantity this coefficient of this term here. Now, what is this matrix C it has non-negative entries and if you take the sum of the rows the entries in any given row or the sum of the entries along any given column, they all add up to 1 because W is a unitary matrix, such a matrix is called a doubly stochastic matrix and a matrix with non-negative entries where every column adds up to 1 and every row adds up to 1 is called a doubly stochastic matrix.

And so, what we see is that whenever W is unitary then this matrix C will be a doubly stochastic matrix. However, the converse may not be true in the sense that if I take C which is doubly stochastic, it there may not exist a unitary W such that mod Wi squared is this doubly stochastic matrix.

So, if I replace the maximisation over W with maximisation over all matrices C which are doubly stochastic, then I am only expanding my potentially expanding my space expanding the space which I am doing this optimization. So, this is further upper bounded by the maximum of 2 Cij. So, Wij squared is equal to Cij times this real part of lambda i star times lambda j hat over all possible doubly stochastic matrices.

Now, this objective function is linear in this matrix and the entries of this matrix C and further you know a linear function is both a convex and concave function of these variables. And so, as a consequence this is this amounts to maximising a convex function over the space of doubly stochastic matrices.

So, in order to solve this optimization problem, we need to know something about this constraint space what is the space of doubly stochastic matrices. So, let us precisely define that. So, suppose Ds is the set of all doubly stochastic matrices. It is a small exercise to show that Ds is a convex set that is you take any two doubly stochastic matrices and if you take a convex combination of those two matrices you will get a doubly stochastic matrix.

So, you can show this very easily. So, the set Ds is actually convex. Now, we use one very fundamental result from optimization that the maximum of a convex function over a convex set is always attained at what is called an extremal point of the convex set. Similarly, if you want to minimise a concave function over a convex set the minimum will occur at an extremal point of the convex set.

So, just to give you the idea suppose I have a convex function like this and if I have a convex set in on this real line which is an interval and if I want to in fact, if I want to maximise this convex function over this convex set then the maximum will occur at an extreme point this is also true if I take a convex function that looks like this for instance.

The maximum will be either this point or this point depending on which one is higher the minimum could be at some point inside but the maximum will always be at an extreme point of the convex set. So, then what is extreme point is like the endpoints of the convex set but the generalisation is as follows.

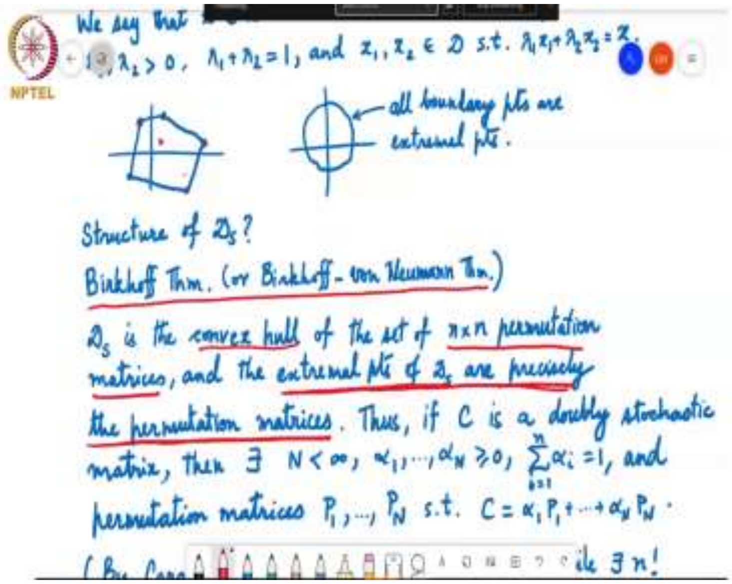
So, if D is a convex set on some defined on some vector space, we say that x a point x in D is an extremal point or an extremal point of D if there is no lambda 1 lambda 2 greater than 0, which add up to 1 and x1, x2 belonging to D such that lambda 1 x1 plus lambda 2 x2 is equal to x.

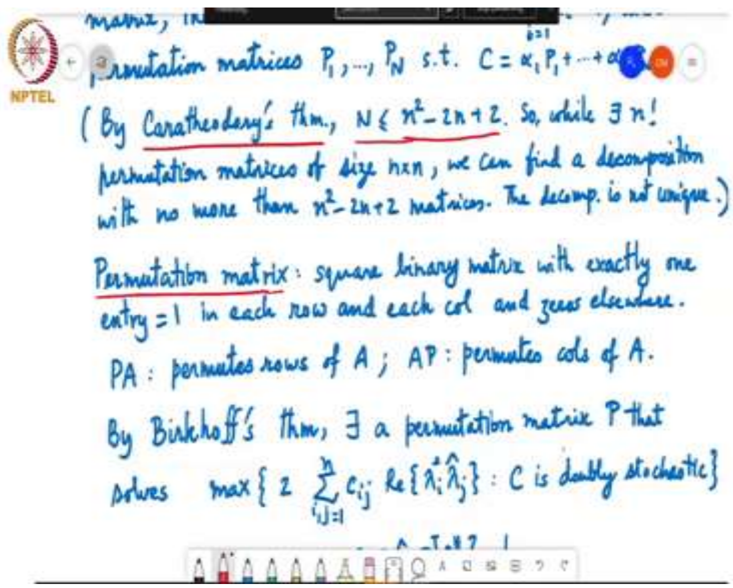
In other words, you cannot find two points that are internal to the set D where if you take a convex combination of those two points you will get this point x. So, that you can see that from this interval also, if I take either this point or this point I cannot write this as write this point or this point as a convex combination of two points that are internal to this convex set.

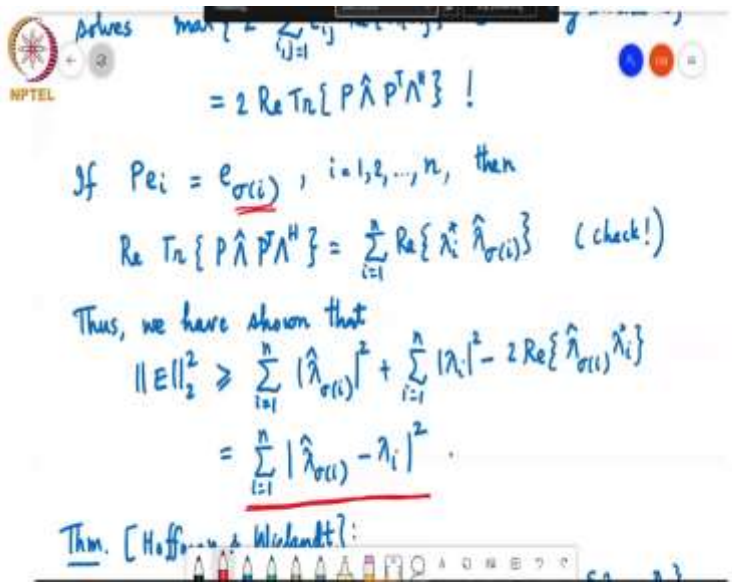
Whereas, if I take a point here I can write it as a convex combination of these two extreme points, so, this is not an extreme point. So, for example, in two dimensions if I have a polygon like this then the corners of the polygon are all extreme points or if I have a circle in two dimensions like this the entire boundary of the circle is a are extreme points.

So, the solution to this to this optimization problem will be at one of the extreme points of this convex set. So, all we need to do is to identify what are the extreme points of this convex set, substitute those extreme points and then pick the best the highest value we can get. So, we need to understand what would be the extreme points of this particular convex set.

(Refer Slide Time: 16:15)







So, the extreme points of this convex set are actually given by a theorem, which is known as Birkhoff's theorem or the Birkhoff's Von Neumann theorem. This is one of the this is a theorem I will not be proving in the class, it is a little digress too much from what we want to do plus where we do not have time to prove this theorem here.

But basically what it says is that the set Ds is the convex hull do not worry about if you do not know what this convex hull means, is the convex hull of the set of n cross n perturbation permutation matrices and the external points of Ds are precisely these permutation matrices that is the punchline.

The extremal points of Ds are precisely the permutation matrices. So, basically, if C is a doubly stochastic matrix, then there exists in n less than infinity, alpha 1 to alpha n non-negative and adding up to 1 and permutation matrices P1 to PN such that you can write C to be alpha 1 P1 plus et cetera up to alpha N PN.

So, any doubly stochastic matrix can be written as a convex combination of permutation matrices, that is the punchline and these permutation matrices are essentially vertices of this convex set Ds. In fact, some side notes here that there is another famous theorem called Caratheodory's theorem, which says that you can choose N to be at most n squared minus 2n plus 2, that is n minus 1 the whole squared plus 1.

So, while there exist n factorial permutation matrices so size n cross n, we do not need to use all n factorial permutation matrices to decompose a given doubly stochastic matrices we can make do with at most n squared minus 2 plus 2 matrices. And further does this decomposition is not unique there are there various ways in which you can decompose if I go back to this example here.

So, if I take a point over here, obviously, you know, I can write this as a convex combination of these extreme points in multiple ways. For example, I can take these three points and find a convex combination get me here or I can potentially take these three points and write it as a convex combination of these three points and so on. So, there is no unique way to do it. But in this case, you can see that for any point I can take, I can reach any point by taking a convex combination of three points.

So, when n equals 2 what happens to n squared minus 2n plus 2 n squared 4 minus 4 plus 2 in fact, two points are enough that is what Caratheodory's theorem says, but this is not this, this is not the convex set corresponding to permutation matrices. So, if you look at the two cross two doubly stochastic matrices, you can write any two cross two doubly stochastic matrix as a convex combination of at most two permutation matrices.

So, then what is a permutation matrix it is basically a square binary matrix with exactly 1 entry equal to 1 in every row and 1 entry equal to 1 in every column and 0's is everywhere else and we have also seen these matrices previously we know that PA permutes the rows of A and AP permutes the columns of A.

So, now if we use this Birkhoff's theorem, then there exists a permutation matrix P that solves this optimization problem, because these, the solution to this optimization problem is an extreme point and the extreme points are all permutation matrices.

So, for that permutation matrix, we go back to the earlier way of writing this expression, and we write it as 2 times real part of trace of now, P is a permutation matrix, so, P lambda hat P transpose because it is zeros and ones you do not need a Hermitian there times lambda Hermitian and now P lambda hat is a permutation of the columns P transpose lambda hat Hermitian is a permutation of the rows of lambda Hermitian.

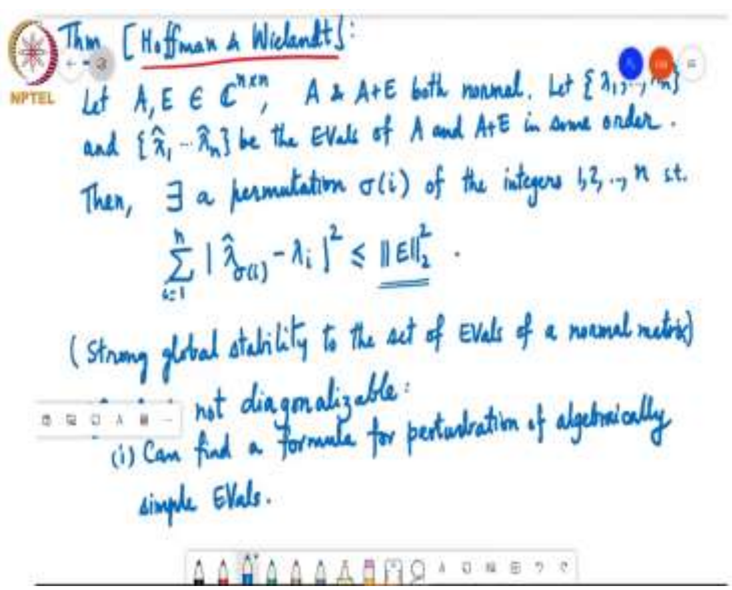
Now, so, if Pei is equal to e of sigma i, so, sigma i represents the permutation. So, basically what it is saying is that the index i is getting mapped to index sigma i, that is what this permutation matrix is doing. So, there are different ways of writing out a permutation matrix one is to write it as a matrix and the other is to specify this sigma of i the permutation function, which maps indices i equal to 1 to n 2i equal to 1 to n.

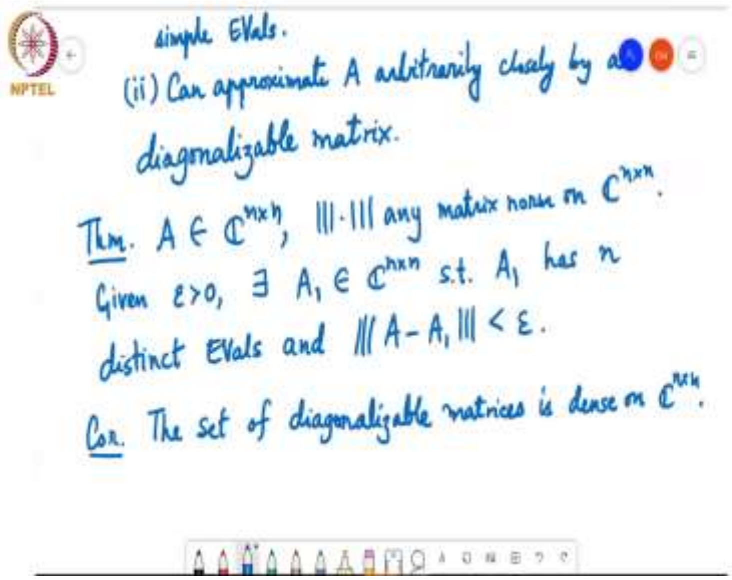
So, if we do this and then we substitute that into this expression, you can actually simplify this formula to this expression here. So, we now see that what we are left with is just lambda i star times lambda hat of sigma i. So, it is just products of permitted versions of lambda hat. So, thus what we have shown is that this E2 squared, the l2 norm of E square is lower bounded by the summation i equal to 1 to n lambda hat of sigma of i square.

So, it was lambda hat i squared earlier, but sigma of i is just a permutation. So, all the entries will get included if I use sigma of i instead of i itself. So, this is also okay plus the summation i equal to 1 to n lambda i square minus 2 times the real part of sigma hat of lambda hat of sigma of i times lambda i star.

Now, this expression here is nothing but the modulus of lambda hat of sigma of i minus lambda i squared added up i equal to 1 to n. So, now, this is beautiful, because, I am now shown that this quantity here, which is the sum of the squared differences of the eigenvalues of A and A plus E these are the eigenvalues of A these are the eigenvalues of A plus E but written in some other order sigma of i, this sum is at most equal to the Frobenius norm squared of E.

(Refer Slide Time: 23:02)





So, what we have shown is what is it is known as the Hoffman Weilandt theorem, which says that if A and E are both n cross n matrices A and A plus E both being normal matrices and lambda 1 to lambda n and lambda hat 1 to lambda hat n are the eigenvalues of A and A plus E respectively and in some order, then there exists a permutation sigma of i of the integers 1 to n such that summation i equal to 1 to n lambda hat of sigma of i minus lambda i squared is at most the norm of E squared.

So, basically what this theorem does is to show that there is a strong global stability, it is strong because it just depends on E2 square and it is global because it does not matter which E you pick even if you choose adversarially this thing is at most E2 square and these are stable to stability of the set of Eigen values of a normal matrix. So, this is one more result that we have about the perturbation of eigenvalues of a matrix.

So, if A not diagonalizable, we seem that we can find a formula for perturbation of algebraically simple Eigen values and we can also do one other thing, which is that we can approximate A arbitrarily by a diagonalizable matrix, so we can approximate this matrix A arbitrarily closely by a diagonalizable matrix and then everything we said about diagonalizable matrix are is applicable and we can then say something about how the matrix A will get perturbed.

What we mean by we can approximated arbitrarily closely by a diagonalizable matrix is something we have already seen before. So, I will just state that to recall what we said. So, A is an n cross n matrix and so suppose this is any norm then given epsilon greater than 0 there exists a matrix A1 say to the n cross n such that A1 has n distinct eigenvalues and A minus A1 is actually less than or equal to epsilon.

In fact, you can even make this strictly less than is that most epsilon and a corollary to this the set or matrices is dense. So, given a matrix A when say to the n cross n, I can find a diagnosable matrix which is arbitrarily close to this matrix A. So, it is useful because one useful way of approaching perturbation related problems is to first solve for diagonal matrices, then solve for diagnosable matrices and then use the approximation and some limiting process to say something about what happens in the non-diagonalizable case.