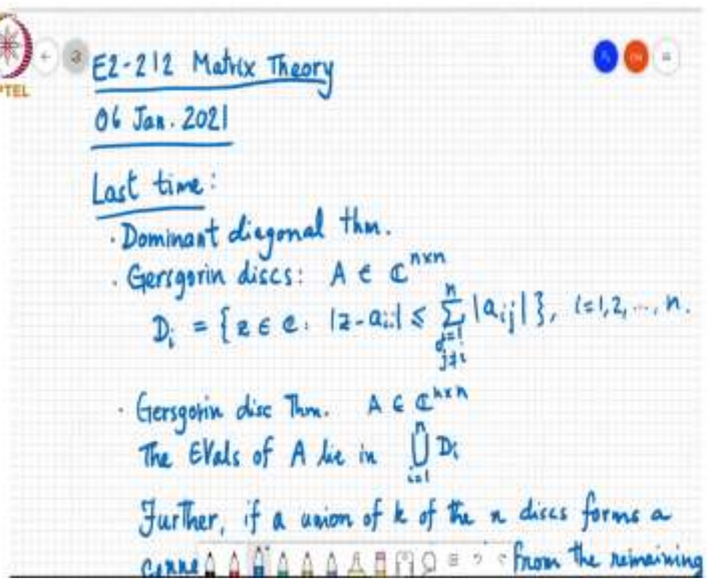
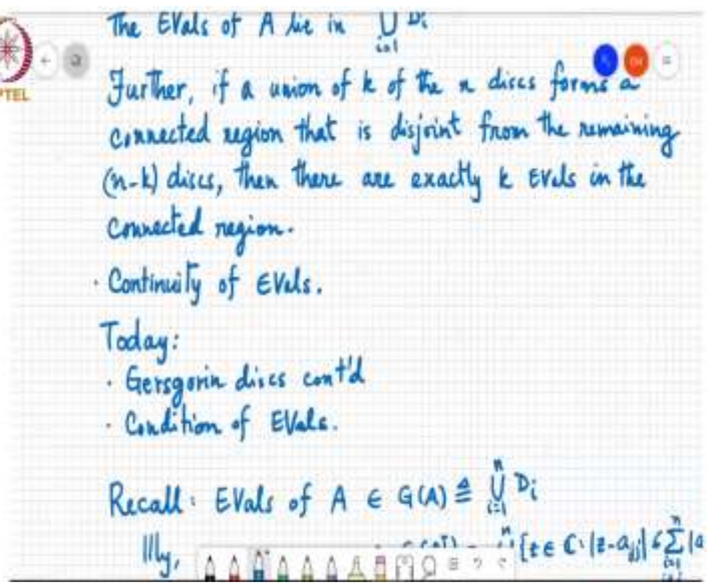
Matrix Theory
Professor Chandra R. Murhty
Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore
Implications of Gersgorin Disc Theorem, Condition of Eigenvalues

(Refer slide Time: 00:14)



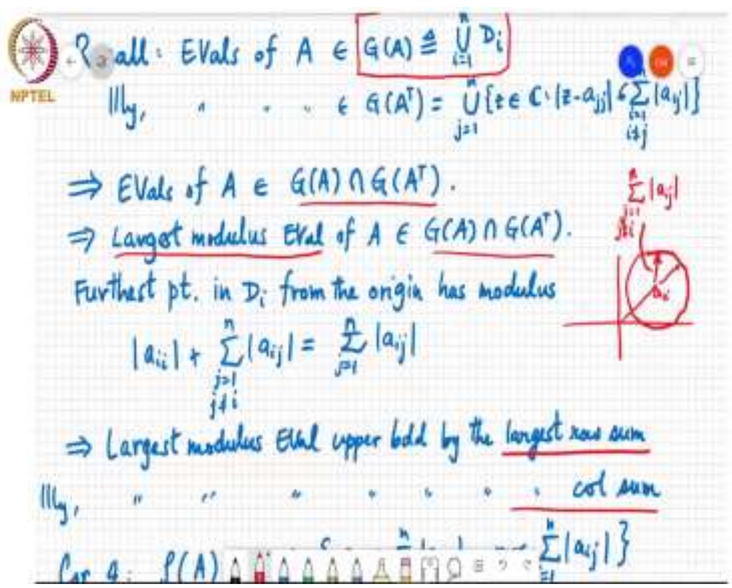


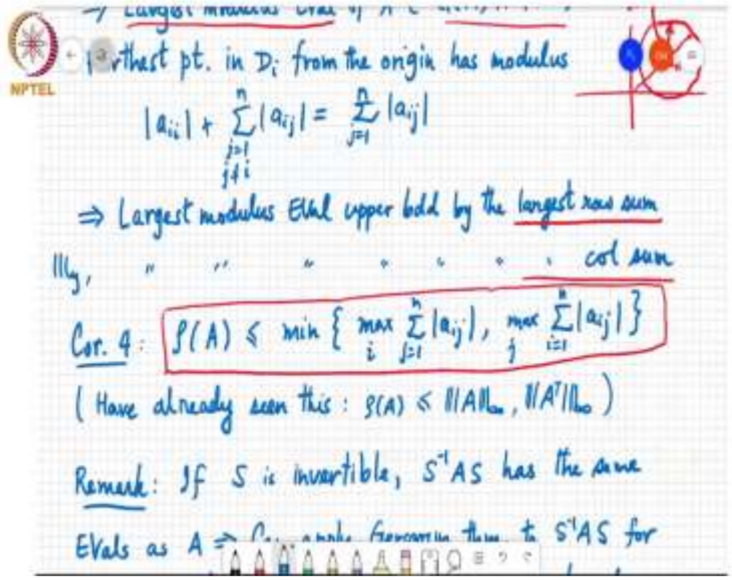
The last time we looked at this diagonal dominance theorem, and we define these Gersgorin discs, which basically are circle centred at each of the diagonal entities and radius equal to the sum of the magnitudes of all the off diagonal terms in the same row. And, we saw this very interesting theorem, which was called the Gersgorin disc theorem, which basically said that all the eigenvalues of A lie in the union of these n Gersgorin discs.

Further, if a union of k of these n discs forms a connected region that is disjoint from the remaining n minus k discs, then there are exactly k eigenvalues in that connected region. Then, we saw the end of the last class this theorem about the continuity of eigenvalues. So, today we will see some consequences of this Gersgorin disc theorem and we will also start talking about the condition number associated with eigenvalues.

Remember that we have seen the condition number plays an important role in the sensitivity of solutions to linear systems of equations, we will see similarly that there is a condition number related quantity that shows up when you are looking at sensitivity of the eigenvalue problem to perturbations in the matrix.

(Refer Slide Time: 01:49)





Now, recall that the eigenvalues of A are in the union of these Gersgorin discs, which I am going to denote by G of A. Similarly, since A and A transpose have the same eigenvalues, the eigenvalue of A lie in G of A transpose which we can define in this way, but basically it is the same definition of the Gersgorin discs, but defined with respect to A transpose, so it is the union j going from 1 to n mod of z minus ajj less than or equal to the sum of all the off diagonal entries in the ith column of A.

So, because the eigenvalues must lie in this set as well as in this set, it must they must all lie in the intersection of these two sets, so this can give us a tighter region within which we can locate the eigenvalues of A. And in particular, note that the largest modulus eigenvalue of A must also lie in this set, but then these are just circles, z minus ajj less than or equal to this.

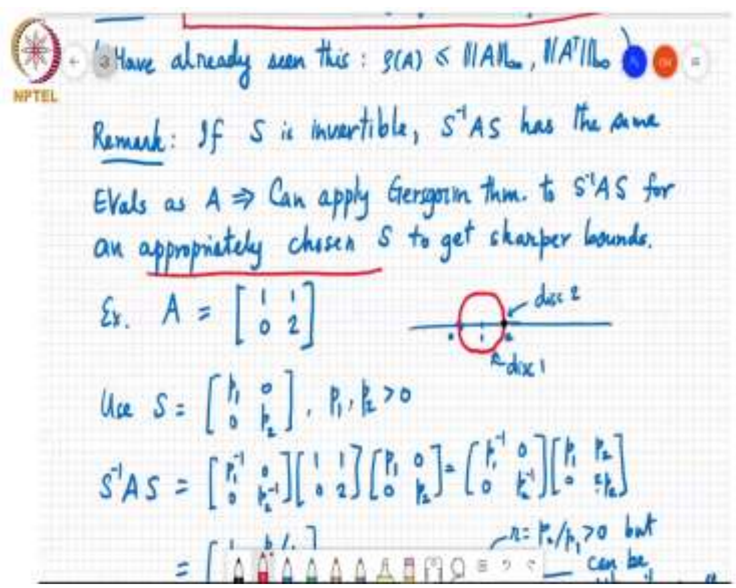
So, we can actually find out the furthest point from the origin in this circle, the furthest point in Di from the origin it has the module is equal to, so the this the circle the Di the circle is basically a circle that is centred at aii, so this is some complex number aii and it has a radius equal to the sum of i or j equal to 1 to n, j naught equal to i mod aij, that is the radius.

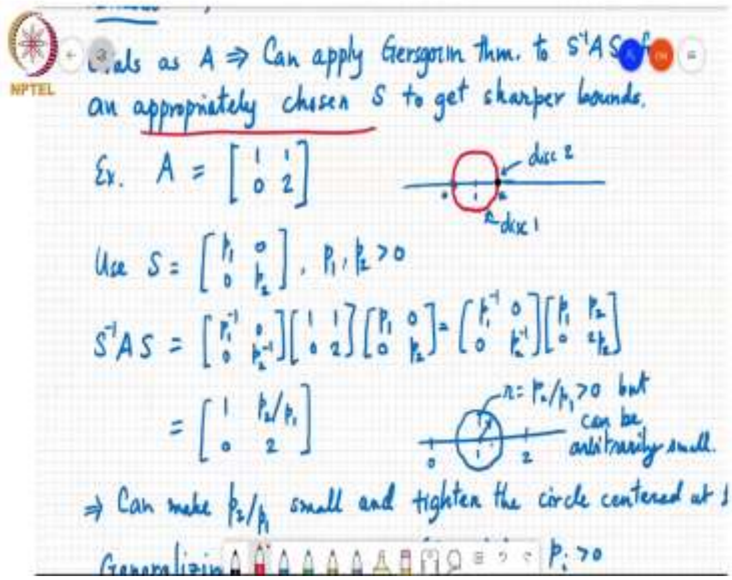
So, if I look at the look for the furthest point, I have to go all the way up to aii and then I have to go further this radius distance, so basically the furthest point has modulus aii plus summation j naught equal to i mod of aij, which is just simply the summation j equal to 1 to n mod aij. So, basically the largest modules eigenvalue of A must be at most this much distance from the origin.

So, the largest modulus eigenvalue of A, but then of course the eigenvalues lie in the union of these Gersgorin discs and so the largest eigenvalue in modulus of A is upper bounded by the largest row sum, it is also upper bounded by the largest columns are and so from that we get the following corollary that the spectral radius of a is less than or equal to the minimum of the largest column sum and the largest row sum.

You already seen this, we have seen it in this form, actually I should write it with three bars, we have already seen this, but this is a more geometric view of the same result that the spectral radius of A is less than or equal to the mean of the largest row sum now and the largest column now.

(Refer Slide Time: 05:20)





Now, another remark is that if S is an invertible matrix, then S inverse AS has the same eigenvalues as A and so now we can apply Gersgorin disc theorem to S inverse AS and choose S you know we can try to choose S cleverly to get sharper bounds. So, here is an example. Suppose I have this matrix A is equal to 1 1 0 2, there are two Gersgorin discs, the Gersgorin disc the first Gersgorin disc is centred at 1 and it has a radius equal to 1.

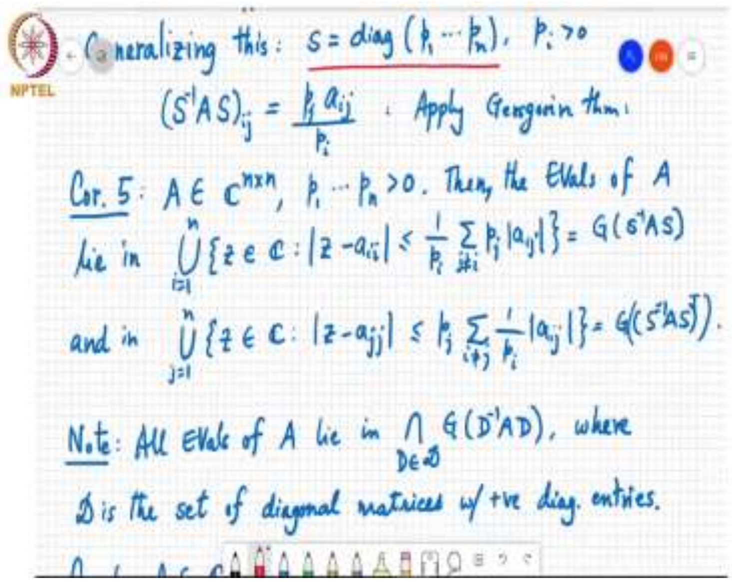
The second Gersgorin disc is centred at 2 and as the radius equal to 0. So, the first disc is shown in red here, it is centred at 1 and it has a radius 1 and the second disc is shown in black here, it is centred at 2 and it has a radius of 0. So, the eigenvalues of A in fact for this we can read off the eigenvalues there 1 and 2 and of course they lie in the union of these Gersgorin discs.

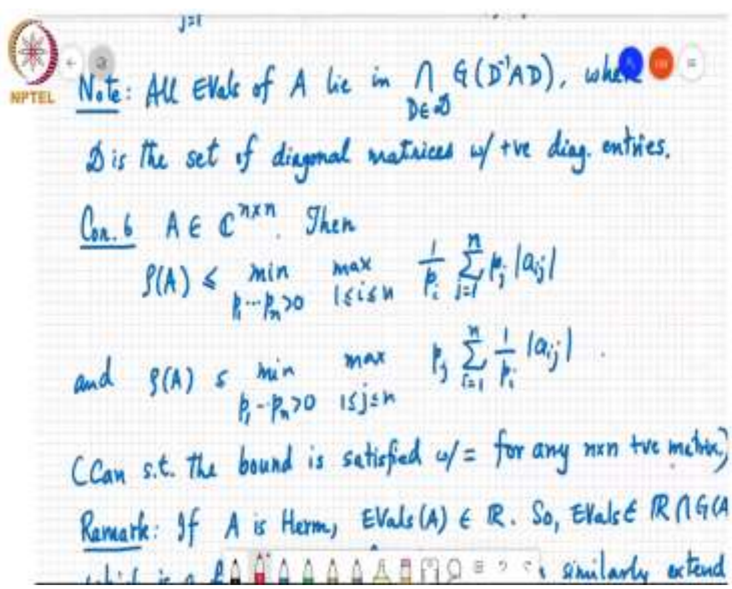
Now, suppose we use S equal to this diagonal matrix with p1 and p2 on the diagonal and p1 p2 being positive numbers, then if I work out what S inverse AS is, it is p1 inverse p2 inverse times this A matrix times p1 p2, so if I solve this product I get p1 p2 0 and 2p2 and then if I multiply that with this matrix p1 p1 inverse cancel and so the diagonal entry will remain 1, and the diagonal entry here will remain equal to 2, but the off diagonal term becomes p2 over p1.

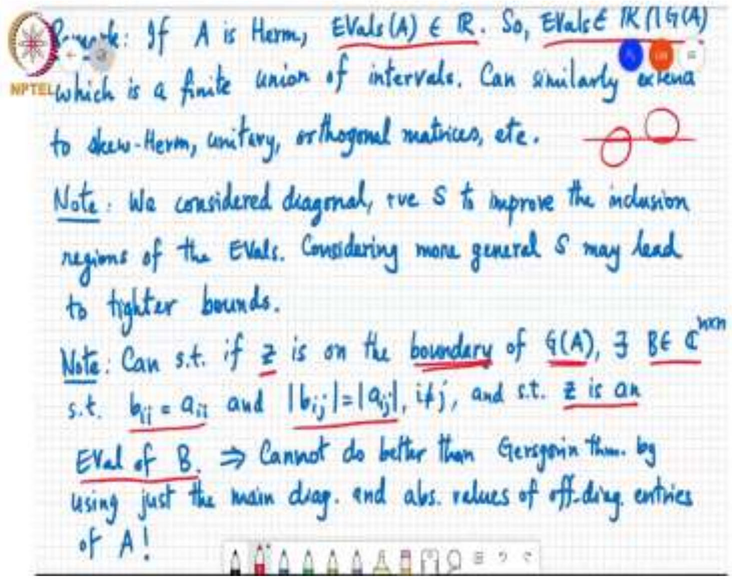
And so now the first circle, the second circle remains at centre 2 and radius equal to 0, but the first Gersgorin disc will now have be centred at 1 and have a radius p2 or p1 and p2 over and p1 p2 can be chosen such that p2 over p1 is positive but arbitrarily small and so you can actually locate the eigenvalues much more accurately, you know that 1 is in the in some tiny neighbourhood around 1 and the other eigenvalue is in a tiny neighbourhood around 2.

Again, of course in this case, the matrix is upper triangular and so it is kind of trivial you already know that the eigenvalues are 1 and 2, so you do not have to approximately locate them, but for more complex matrices, this could be a useful trick. So, let us generalize this.

(Refer Slide Time: 07:45)







So, suppose S was the diagonal matrix containing p1 to pn along the diagonal with all this pi's being positive, then if you look work out the ijth element of S inverse AS that is going to be equal to pj time's aij divided by pi. And now we apply Gersgorin theorem to this, this matrix S inverse AS, if we do that we get the following corollary A is an n cross n matrix, and p1 to pn are some numbers which are greater than 0, then the eigenvalues of A lie in so the when you do this, the diagonal entries remain unchanged.

So, the centres of a circle remain unchanged, so it is mod z minus aii is less than or equal to the sum of the off diagonal terms when I am summing over j, pi does not depend on j, so I can bring it out to the summation and so I have summation j naught equal to i pj times mod aij and this is basically G of S inverse AS my notation above, and also in union j equal to 1 to n this is the column version, z minus ajj less than or equal to…

When I am summing over i this does not depend on i, so I can pull it out, I will get pj times summation over i going from 1 to n i naught equal to j, 1 over pi times mod aij. So, this can give us some more sharper bounds on the location of the eigenvalues. And specifically I can think about optimizing pi and pj that is p1 o pn such that these bounds are as tight as possible.

So, essentially all the eigenvalues lie in the intersection of all such choices over all the choices of this p matrix, so is in the intersection of D belonging to script D of the G that is the union of

Gersgorin discs, corresponding to D inverse AD where D the script D is a set of diagonal matrices with positive entries.

So, that is basically this Corollary is 6, which is that the spectral radius when applied, when this is applied to the spectral radius, we get that the spectral radius is at most the min over p1 to pn being positive of the max row sum norm or the max column sum norm, which is basically max over this is the sum across columns, and this is the sum across rows. And then you are free to minimize that over p1 to pn and this minimum value of this is still an upper bound on row of A.

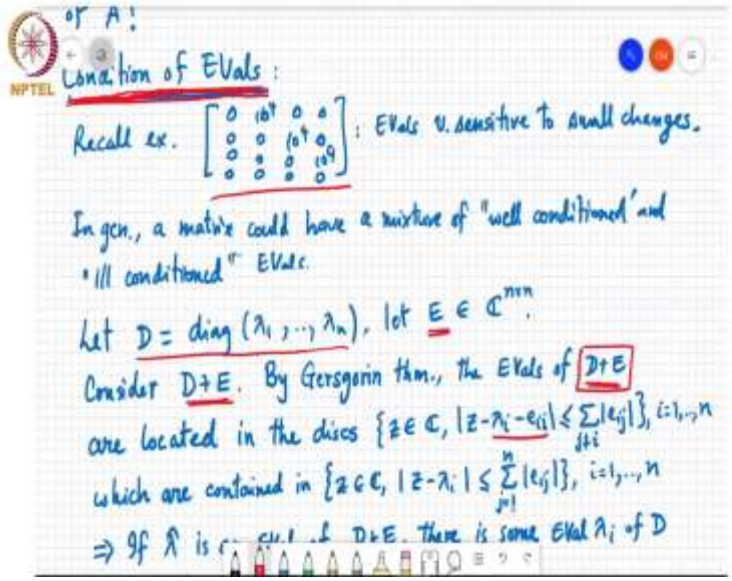
It turns out that this upper bound is actually tight for any matrix with positive entries and there is a proof in the text that you can look at. Now, so far we have been discussing matrices that are arbitrary not necessarily Hermitian, but suppose A was Hermitian, then the eigenvalues of A are real valued and so then we can specialize the Gersgorin theorem to say that the eigenvalues belong to the real line intersection with the union of Gersgorin disc.

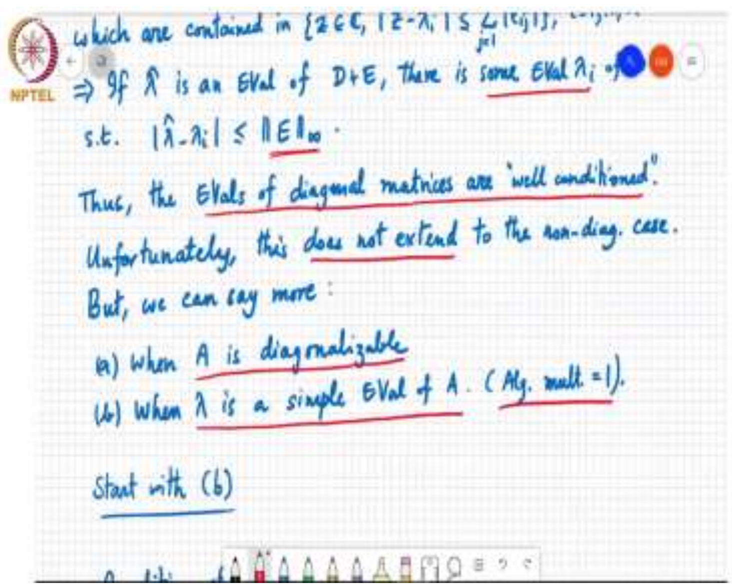
So, if I take these Gersgorin discs which could be located wherever et cetera, and then I take the union of that with the real line, I just get these line segments. So, this is a finite union of intervals. You can similarly write out tighter bounds when the matrix has additional structure like skew Hermitian, unitary or orthogonal et cetera. Now, we also looked at a diagonal positive S to improve the inclusion regions of the eigenvalues, but it is possible to get tighter bounds perhaps by considering more general S, but we will not look at that in this in this course.

Now, one related question is, can you do better than the Gersgorin disc theorem or is that the tightest uncertainty region within which you can you within which the eigenvalues of the matrix A are guaranteed to lie? The answer is no, because there is also this is also there in the text, but it can be shown that if z is some complex number on the boundary of G of A, boundary of these Gersgorin discs, then you can find a matrix B, which matches A in the diagonal entries and matches A in magnitude in the off diagonal entries and such that this z is an eigenvalue of B.

And so, basically the point is that in Gersgorin's theorem, we are only using the main diagonal entries and the absolute values of the off diagonal entries, and that is indeed the tightest bound you can get on the uncertainty region within which eigenvalues lie. So, if you want a tighter bound than the Gersgorin disc theorem you will need to take into account the phase angles or the signs of the off diagonal terms.

(Refer Slide Time: 13:11)





Now, we move on to a different topic, which is the condition of eigenvalues. This was actually a topic that sort of initiated all this discussion location and perturbation of eigenvalues, this whole chapter is about that. Now, I come back to this example we discussed in the first class we had on this chapter, where we looked at this matrix and we said that these eigenvalues are very sensitive to small changes to this matrix.

We added a 10 to the minus 2 here and we found that the eigenvalues became plus or minus 100 and plus or minus 100i. So, this is very sensitive to small changes in the eigenvalue. And in general, the matrix could have here there is only one eigenvalue 0 and this matrix is not a well-conditioned matrix with respect to these eigenvalues, in the sense that it is a small perturbation can lead to a large perturbation in eigenvalues.

So, by and large this is the definition we will use for an eigenvalue being well conditioned or ill condition namely, that if you apply a small perturbation of size measured in some norm, if you apply a perturbation of size epsilon to the matrix A then the perturbation in the eigenvalue should also be of the order epsilon.

If that happens, we say that the matrix the eigenvalue lambda is a well-conditioned eigenvalue, otherwise we say that it is an ill conditioned eigenvalue. So, generally what happens is that these eigenvalues could be some of the eigenvalues I mean the matrix A could be well conditioned with respect to some of its eigenvalue values and ill condition with respect to the other eigenvalues. Now, in particular…

Student: Sir?

Professor: Yes.

Student: Sir, suppose if I apply this epsilon perturbation to diagonal entries, then it is not necessary for the eigenvalue to be ill conditioned, what I am essentially trying to say is it also depends on where we apply the perturbation.

Professor: Yes.

Student: So, how can we conclude it is well conditioned in general? So, Should not it be a function of the position also?

Professor: It depends on it, so what we will be doing, we will look at perturbations like this by a matrix E and what you are given is a matrix D plus E and an adversary is allowed to choose whichever entries in E they want to perturb and your eigenvalues should remain stable, no matter which entries, the adversary perturbs as long as the overall magnitude of perturbation is less than some epsilon, that is the kind of guarantees that we will look for.

So, I will explain that further as we go along. So, now in particular if you start with a matrix D which is diagonal and you let E be some perturbation matrix and consider D plus E, now by Gersgorin theorem the eigenvalues of D plus E the diagonal entries become lambda i plus eii, so those become the new centres of these discs and the radius is the summation j naught equal to i of eij, so that is basically a Gersgorin disc.

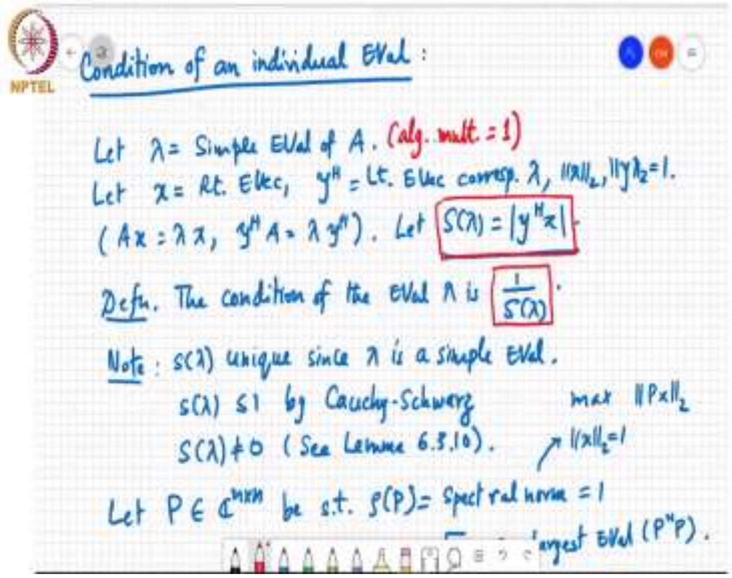
And the union of this over i going from 1 to 1 is where the eigenvalues are guaranteed to be located, this is the, these are the eigenvalues of the perturbed matrix D plus E. Now, what I can do is to add eii and use triangle inequality and I can show that these eigenvalues these discs are actually contained in the this set of discs, which is centred at lambda i and as radius mod eii plus this right hand side here which is summation j equal to 1 to n mod of eij.

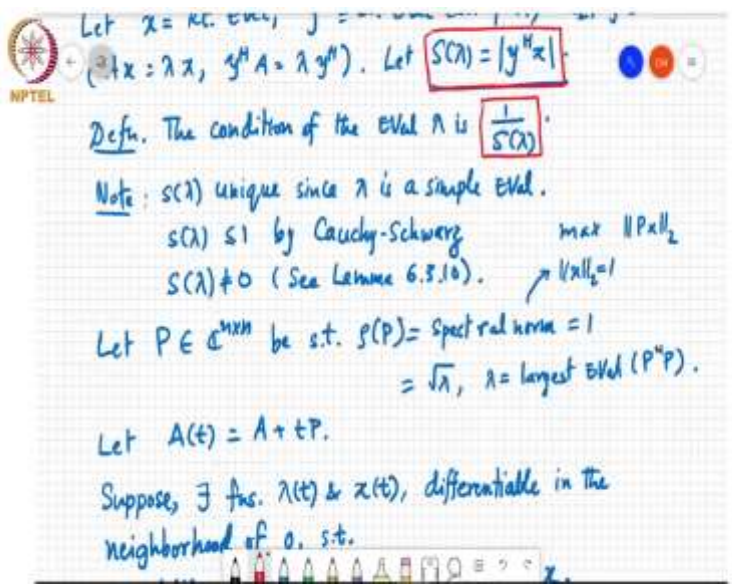
So, what that means is that, if lambda hat is an eigenvalue of D plus E, then it must hold that when I substitute lambda hat for z here that lambda hat minus lambda i less than or equal to this should hold for at least one of these i's. And so, that means there is at least there is some eigenvalue lambda i have of this matrix D such that lambda hat minus lambda i is less than or equal to the max of all these radius, which is equal to the l infinity norm of E.

So, what this means then is that as long as I am allowed to only perturb the matrix A D by a matrix E whose infinity norm is bounded, then the perturbations in the eigenvalues are also bounded by the same quantity. So, in other words what this shows is that the eigenvalues of diagonal matrices are well conditioned.

Unfortunately, this argument does not extend to the non-diagonal case, but we can say more in two important special cases, the first being when A is diagonalizable and the second being when lambda is a simple eigenvalue of A, that is an eigenvalue whose algebraic multiplicity equals 1. Let us, start with the second case. So, it is the simple eigenvalue and so we will look at the condition of that specific eigenvalue.

(Refer Slide Time: 18:55)





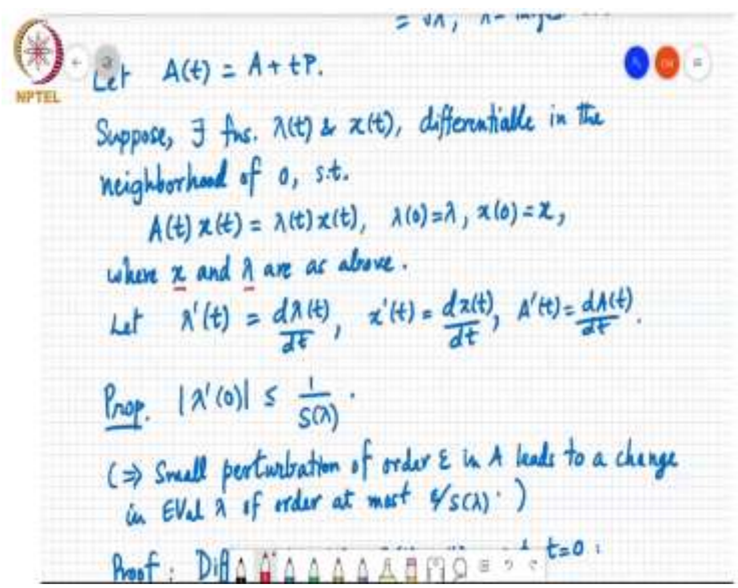
So, let lambda be a simple eigenvalue of A which means that it has algebraic multiplicity equal to 1 and let x be a right eigenvector and y be a left eigenvector corresponding to lambda and both being unit now, that means Ax equals lambda x and y Hermitian A equals lambda y Hermitian and define S of lambda to be equal to the mod of y Hermitian x, the inner product between y and x, the left eigenvector and the right eigenvector, then we define the condition of the eigenvalue lambda to be 1 over S of lambda.

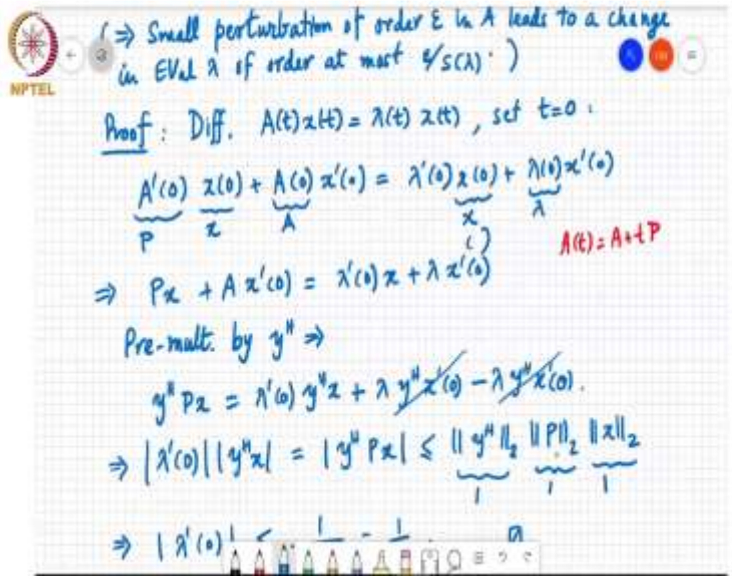
Now, because a lambda is a simple eigenvalue of A, is S of lambda is unique and S of lambda is at most equal to 1 by the Cauchy Schwarz inequality and it also possible to show that S of lambda is not equal to 0, so it is a number between 0 strictly greater than 0 and less than or equal to 1. Now, let p be any matrix whose spectral norm equals 1, that is square root of spectral norm is square root of lambda, where lambda is the largest eigenvalue of p Hermitian p or it is also equal to the max over all unit l to norm vectors of the l to norm of px.

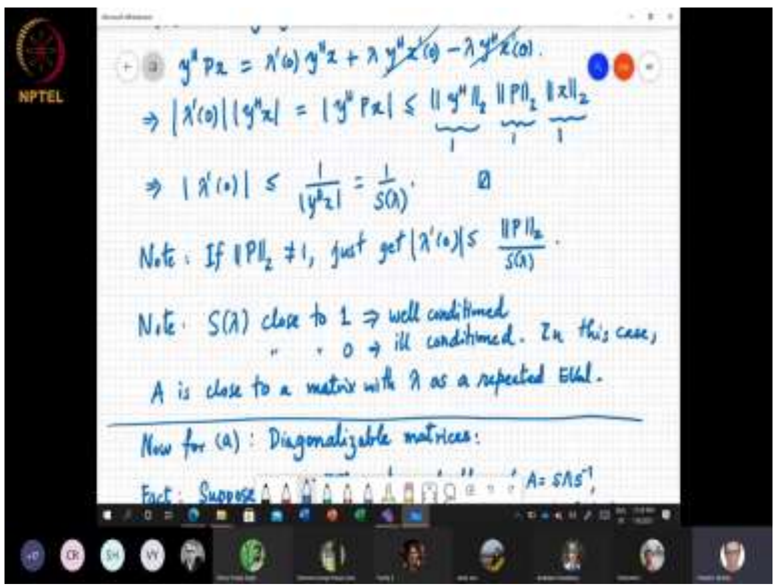
So, p be a matrix such that it has spectral norm equal to 1 or spectral radius equal to 1, then define A of t to be A plus tp, in other words we are looking at perturbing the matrix A by a unit spectral norm matrix multiplied by some coefficient t and think of t as being a small number, so if t is small enough you are really applying a small perturbation on this matrix A.

So, I am, so what am I doing here? I am trying to show you why we consider 1 over S of lambda to be the condition of the eigenvalue lambda. In other words, when you perturb the matrix like this, A of t equals tp, then the perturbation in the eigenvalue lambda will be of the order 1 over S of lambda times the perturbation that you applied, which is like t, that is what we will show.

(Refer Slide Time: 21:38)







Now, suppose lambda of t and x of t, both being differentiable in with respect to t in the neighbourhood of 0, the eigenvectors and eigenvalue of A of t, that is a of t times X of t equals lambda of t, x of t and know that when I said t equals 0, I get lambda of 0 and that lambda of 0 equals lambda and x of 0 equals x where lambda and x are as I defined above, so lambda is a simple eigenvalue, and x is its corresponding eigenvector.

And define dash with to mean the derivative. So, in other words lambda dash of t is d lambda of t over dt, x dash of t is dx of t over dt and A dash of t is dA of t over dt. So, we have the following proposition which says that lambda dash of 0 in magnitude is at most 1 over S of lambda. What this means is that a small perturbation of order epsilon in A leads to a change in eigenvalue lambda of order at most epsilon over S of lambda that is what lambda dash of 0 being at most 1 over S of lambda means.

So, how do you show this? So, we start by differentiating this equation A of tx of t equals lambda of tx of t with respect to t, and then we said t equals 0. So, if I differentiate this using chain rule, I have A dash of t x of t plus A dash of so plus A of t x dash of t is equal to lambda dash of t x of t plus lambda of t x dash of t, so and then I am substituted t equals 0 to get this equation.

But remember that A dash of 0, A of t is equal to A plus tp. So, A dash of 0 is just p and x of 0 equals x, A of 0 is just A, and here x of 0 equals x and lambda of 0 equals lambda. So, substituting all that I have px plus A x dash of 0 is equal to lambda dash of 0 times x plus lambda x dash of 0. And now we pre multiply by y Hermitian, so I will get y Hermitian px and here I have y Hermitian Ax dash of 0, but y Hermitian A is lambda y Hermitian.

So, I will have a minus y Hermitian, so I will have a minus lambda y Hermitian x dash of 0 on the right hand side. And this is lambda dash of 0 times y Hermitian x, and this is lambda y Hermitian x dash of 0, which exactly cancels with the last lambda y Hermitian x dash of 0 coming from the left hand side. So, these two cancel and what I am left with is lambda dash of 0 times mod y so if I take the modulus of this equation, so just this equals this is all I am left with.

So, we take the modulus on both sides mod of lambda dash of 0 times mod y Hermitian x is equal to y Hermitian px magnitude, which by the sub-multiplicativity of and compared you use the idea of compatible norms and sub-multiplicativity to write it as the product of the norm of y Hermitian times the norm of p times the norm of x.

And we have started out by assuming that these both all three of these are equal to 1 and so we have lambda dash of 0 is less than or equal to 1 over y Hermitian x, which is 1 over x of lambda. Even if norm p2 is not equal to 1, we would just have a p2 with sitting in the numerator here, the result looks more elegant if you use assume p2 equals 1 and write it as 1 over S lambda.

So, what this means is that if S of lambda, which is mod y Hermitian x the inner product between the left eigenvector and the right eigenvector, if that is close to 1, then the matrix is well, the eigenvalue lambda is a well-conditioned eigenvalue, if it is close to 0 then iis an ill conditioned eigenvalue.

In this case, it actually means that A is close to a matrix with lambda as being a repeated eigenvalue, so when you have repeated eigenvalues, it is possible that you can apply small perturbations and make large changes in the eigenvalues, not necessary that is possibly, the result does not apply to that case.