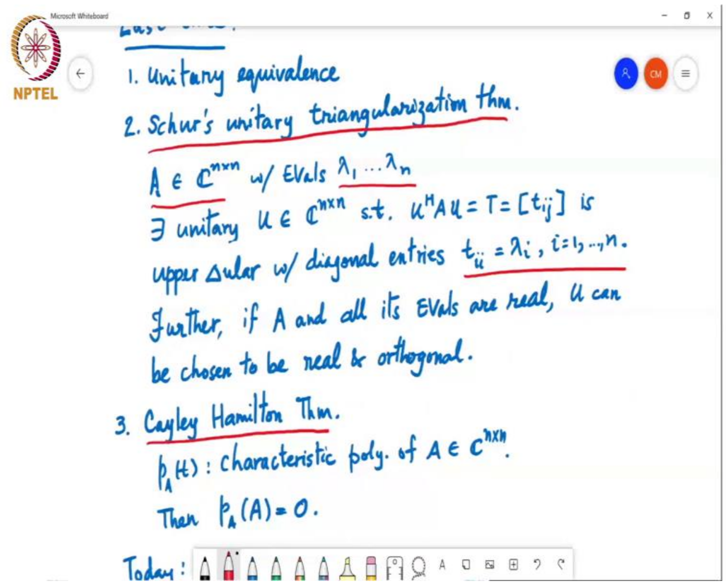
Matrix Theory

Professor Chandra R. Murthy
Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore

Lecture 42

Use of Cayley-Hamilton Theorem and Diagonalizability Revisited

(Refer Slide Time: 00:14)

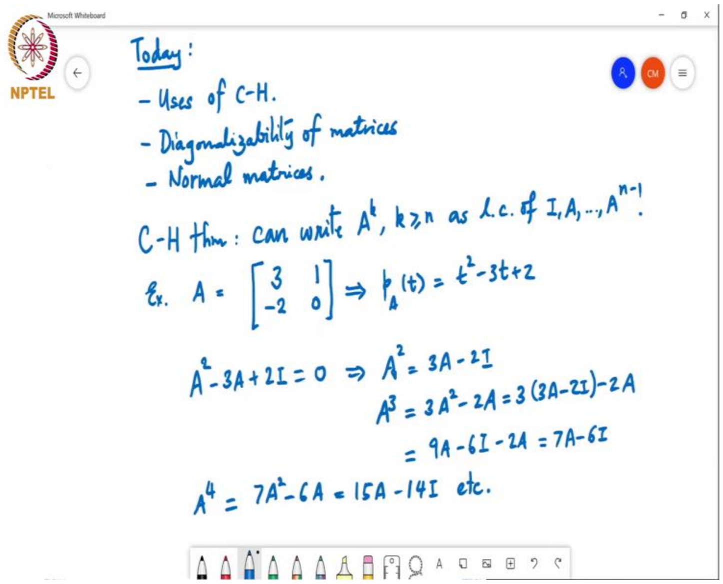


The last time we looked at the notion of unitary equivalence. We completed the discussion. Then we presented or discussed this very important theorem which is Schur's unitary triangularization theorem, which basically says that given any matrix, there is no restrictions, it can be any complex minus valued matrix of size n cross n and it has eigenvalues lambda 1 to lambda n.

An n cross n matrix will always have n eigenvalues. Then there exists a unitary matrix u such that A is unitarily equivalent to a matrix T which is upper triangular with diagonal entries equivalent to these n eigenvalues lambda 1 to lambda n. Of course if this A and all its eigenvalues are real u can be chosen to be real orthogonal matrix which will also be real orthonormal matrix. That is the, I mean generality of the theorem is what makes it very important. It is applicable under no restrictions on the matrix A.

One application of this triangularization theorem that we saw was the Cayley-Hamilton theorem which basically says that any matrix A satisfies its own characteristic polynomial. And we saw the proof of this theorem and that is where we stopped the previous time.

(Refer Slide Time: 01:41)



The next thing today what I want to discuss is some uses of the Cayley-Hamilton theorem and then some points about diagonalizability of matrices. And then I may be start the discussion on normal matrices. So, we will begin with the first point that is uses of Cayley-Hamilton theorem. So, the Cayley-Hamilton theorem can be used to express A power k for k a greater than or equal to n as a linear combination of lower powers of k. So, this is an application that I am… suppose most of you have seen in your undergraduate program. So, we can write A power k, k greater than or equal to n as a linear combination of I, A, A power n minus 1.

So, we will just illustrate this with an example. So, suppose A was the matrix 3, 1, minus 2 and 0. Then its characteristic polynomial p A of t will be t minus 3 times t plus 2. So, that is going to be equal to t square minus 3t plus 2. And since the matrix A satisfies its characteristic polynomial we have A square minus 3A plus 2I equals the all 0 2 cross 2 matrix which in turn implies A squared equals 3A minus 2I. So, I can write A squared in terms of A and the identity matrix.

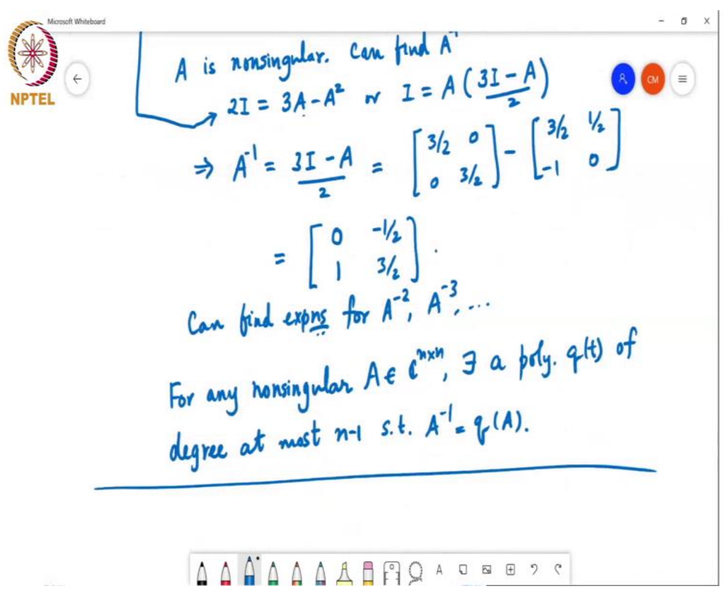
Now, A cube, I just have to multiply this by A. I get 3A squared minus 2A and I can substitute for A squared from this. So, that becomes… I just do it 3 times 3A minus 2I, minus 2A which is equal to 9A minus 6I minus 2A which is equal to 7I minus 6, sorry 7A minus 6I. Similarly A power 4, again you multiply this by A and you substitute for A square and you get 7A square minus 6A which then is equal to 15A minus 14I and so on.

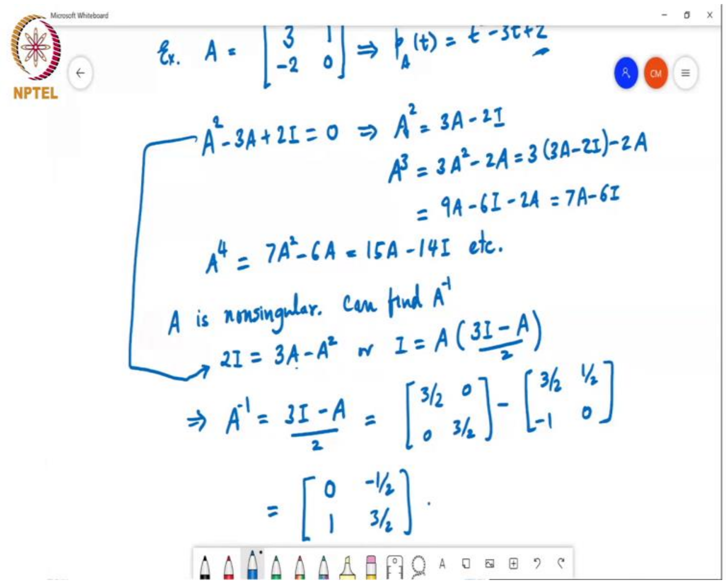
So, we can write all the higher powers of A as a linear combination of lower powers of A. Again this is a very interesting observation and to me it is not obvious why if you take higher and higher powers of A you should always be able to write it as linear combination of the first n powers including 0 of A.

Specifically, if you take an n cross n matrix it is an object that is living in n squared dimensional space. And so I can always write an n square dimensional vector as a linear combination of n squared linearly independent vectors all sitting in the n squared dimensional space. So, the fact that you can, I mean so if I had said that I can write A power n as a linear combination of n squared matrices I, A, S squared up to A to the n squared minus1 that would not have been surprising.

But what is surprising here is that you can write A power k as a linear combination of I, A, A squared up to A power n minus1 only. So, you only need n of these matrices and all other powers of A can be written as linear combination of these n matrices. That is what is surprising here.

(Refer Slide Time: 06:49)



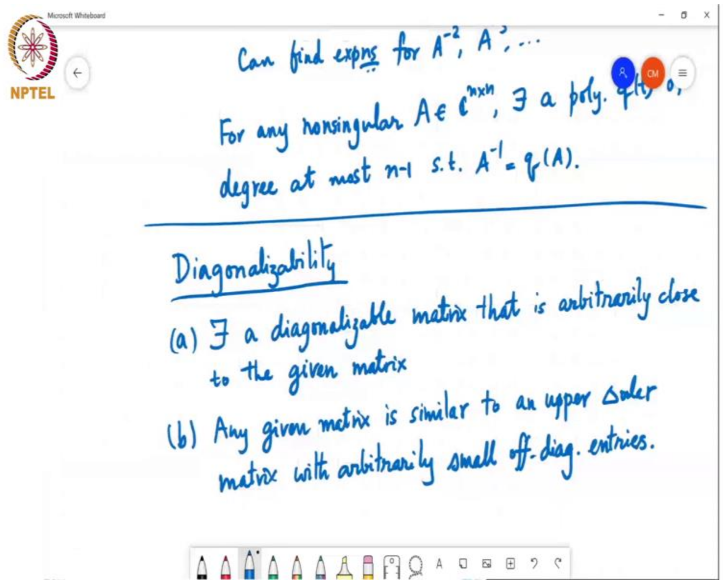


Now, so basically one other small thing is that this constant term here is actually determinant of A and that is not equal to 0. So, A is nonsingular. So, A is nonsingular. So, this allows us actually to write or find A inverse like this. So, basically what I do is I take this equation and write 2I is equal to 3A minus A square. Or I is equal to 3 over 2, or I will write it this way. I will take one A common out between these two and write it as A times 3I minus A over 2. Now, I multiply both sides by A inverse. And that gives me A inverse is equal to 3I minus A over 2.

So, which I can write as 3I is 3 over 2 0, 0, 3 over 2, minus A over 2 is 3 over 2,1 over 2, minus 1,0. So, then that gives me A inverse is equal to 0, minus half, 1 and 3 over 2. So, basically Cayley-Hamilton theorem allows you to also compute A inverse when A is nonsingular. And in fact you can find expressions for A to the minus 2, A to the minus 3 and so on also. So, you can try this. So, all you have to do is to multiply this by A inverse and then substitute for A inverse from this.

So, you will get 3 over 2 times A inverse minus 1 half the identity matrix. But you already have an expression for A inverse. You substitute for that. You will get the expression for A to the minus 2 and so on. And this observation is true for any nonsingular matrix. And so we can say that for any nonsingular A in C to the n cross n there exists a polynomial q of t of degree at most n minus 1 such that A inverse equals q of A.

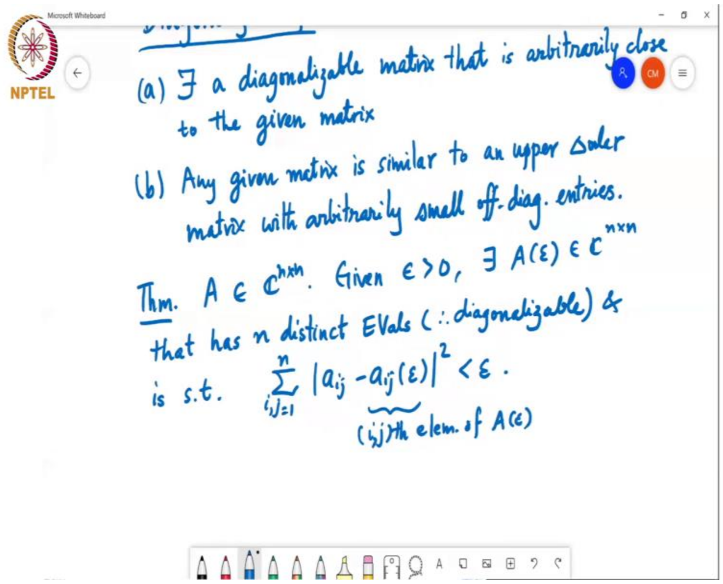
(Refer Slide Time: 10:55)

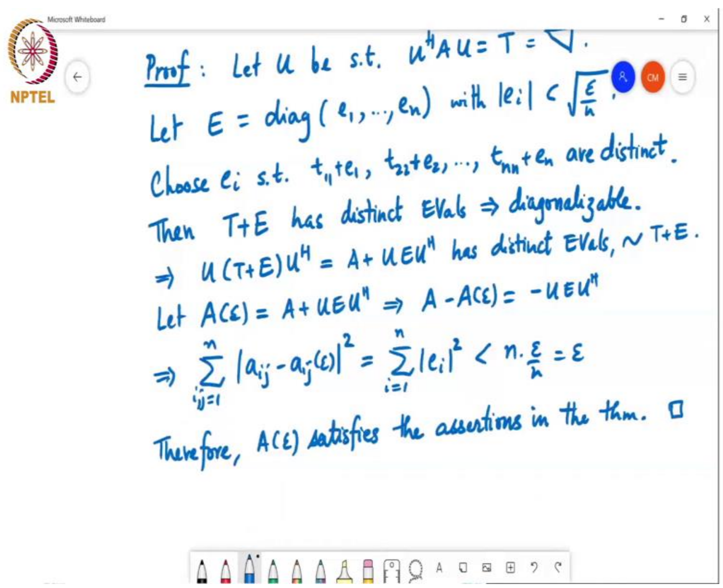


So, now we move on to the next point which is that we know that not all matrices are diagonalizable, but how close can we get? Can we get a matrix that is, can we take a matrix that is not diagonalizable and express it through a similarity transform or unitary equivalence for a matrix that is almost diagonal.

So, there are two ways to answer this. So, the first way is to consider that is, we can say, can we find or… so I will just write the answer. So, there exists a diagonalizable matrix that is arbitrarily close to the given matrix. And I will make the sense in which I am saying arbitrary closed clear in a minute. And b, any given matrix is similar to an upper triangular matrix with arbitrarily small off-diagonal entries that is almost diagonalized.

(Refer Slide Time: 13:02)





So, basically this is… There are two theorems that basically make this assertion. So, the first one is like this. So, given the matrix A in C to be n cross n, given any small number epsilon greater than 0 there exists an A of epsilon which is the matrix which is going to be close to A, in C to the n cross n that has n distinct eigenvalues and therefore diagonalizable and is such that sigma I j equal to 1 to n, a ij minus a ij of epsilon, this is the difference between corresponding entries of a and a of epsilon and added up, squared values added up over all the entries is less than epsilon. So, here a ij of epsilon is the ijth element of a of epsilon. So, this is the result. So, basically it does not matter if a matrix is not diagonalizable. You can find a matrix that is arbitrarily close to it and is also diagonalizable.

Proof is actually quite straightforward. So, since A is n cross n there is a u such that u Hermitian A u is equal to T which is upper triangular. That is Schur's theorem. Let E be a matrix, a diagonal matrix en, with ei in magnitude being less than square root of epsilon over n. So, ei's, each of them, none of them is bigger than square root of epsilon over n in magnitude. And we choose these ei's such that t11 plus e1, t 22 plus e2, tnn plus en are distinct.

So, can this be done? Can you always find e1 e2 upto en with magnitudes less than square root of n such that t11 plus e1, t22 plus e2 etc up to tnn plus en are distinct numbers? Of course you can because there are infinitely many numbers between 0 and square root of epsilon over m. We just have to pick some numbers such that, and you also can choose the phase angles of these numbers to make them all distinct.

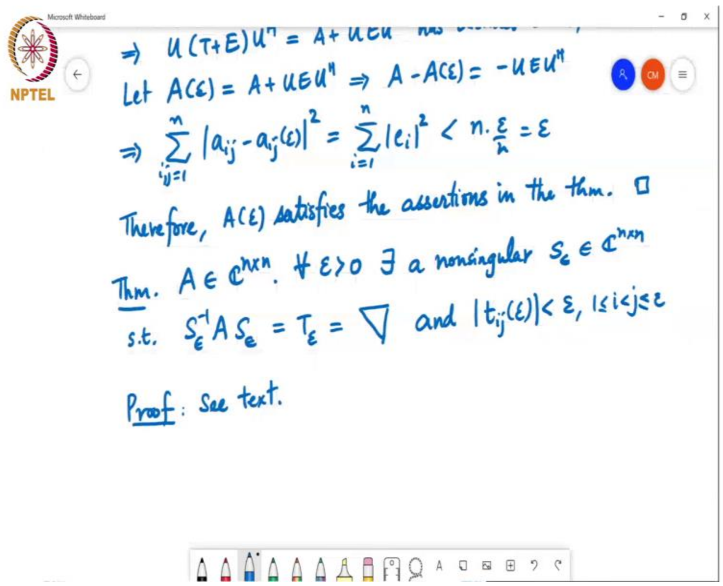
So, it is really very easy to choose n numbers such that t11 plus e1 up to tnn plus en are all distinct numbers. Then the matrix T plus E has distinct eigenvalues. It means that it is diagonalizable. We have already seen that before that the matrix that has distinct eigenvalues is always diagonalizable. And so basically we then have that if I consider u times T plus E times u Hermitian, so I am undoing this operation here. This will be equal to A plus u e u Hermitian. This matrix, this is similar similarity transform, it preserves the eigenvalues, so this matrix has distinct eigenvalues which implies it is diagonalizable.

So, I will so I am, it is similar to T plus E. So, that tells what we should choose as A of epsilon. So, let A of epsilon be equal to A plus u e u Hermitian which implies A minus A of epsilon is going to be minus u E u Hermitian. So, and we have already seen that A of epsilon is diagonalizable. We just need to show that the Frobenius norm of A minus A of epsilon, the Frobenius norm square of this would be less than epsilon. That would satisfy this last requirement of the theorem.

So, then we have that summation i j equal to 1 to n mod a ij minus a ij of epsilon square. This is the Frobenius norm and this is invariant under unitary equivalence, and so, or in fact invariant under similarity transforms and so this is equal to Sigma i equal to n. So, I just need to consider the Frobenius norm of this quantity e which is diagonal. So, I just need to add up over diagonal entries mod ei square.

But each of these eis is at most square root of epsilon over n in magnitude. So, ei square is at most epsilon over n. And so if I add up n of these I get that this is less than n times epsilon over n, which is equal to epsilon. So, basically therefore A of epsilon satisfies the requirements of the theorem.

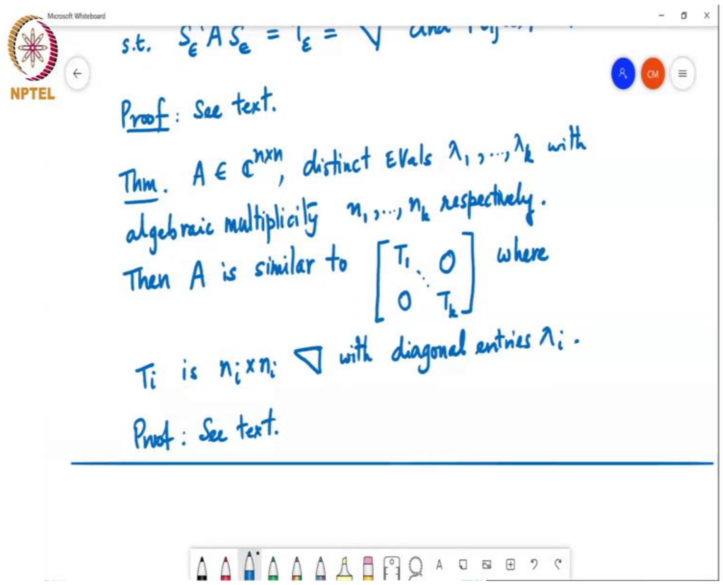
(Refer Slide Time: 21:25)



The other theorem, it goes like this. So, again A is n cross n matrix. Then for every epsilon greater than 0 there exists a nonsingular S epsilon belonging to the C to the n cross n such that S epsilon inverse A S epsilon is equal to T epsilon which is upper triangular and mod of t ij of epsilon is less than epsilon for 1 less than or equal to i less than j less than or equal to epsilon. Of course t ii you cannot restrict it to be small because t iis are eigenvalues of A and so those may not be small. But the off-diagonal terms can be made arbitrarily small.

So, the difference between the two theorems is that in this case what we are doing is instead of trying to diagonalize A we are trying to diagonalize a nearby matrix A epsilon and we say that there is a nearby matrix A epsilon that is diagonalizable. And in this theorem what we are trying to do is instead of bringing A to the diagonal form we are bringing it to the upper triangular form with arbitrarily small off-diagonal entries. So, we are getting closer and closer to E. So, I will not go over the proof of this theorem. It is a, it is some detail which will take me a long time to complete but you can see the text.

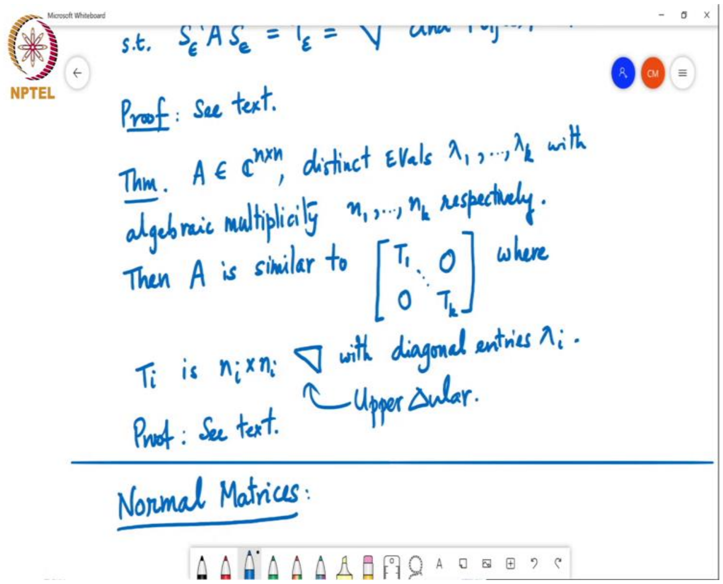
(Refer Slide Time: 23:51)



And in order to just, there is a last point in this particular discussion. There is one more theorem which is actually another extension of Schur's theorem and is useful for the Jordan Canonical form which we will discuss a bit later. So, this is, so again A is n cross n matrix. Then and it has distinct eigenvalues lambda 1 through lambda k. It has k distinct eigenvalues, which and k can be less than n, k can at most be equal to n with algebraic multiplicities. So, algebraic multiplicity is the number of times it occurs as a 0 of the characteristic polynomial n1 through nk respectively.

Then A is similar to matrix T1 Tk, 0 everywhere else where this matrix is Ti is upper triangular ni cross ni upper triangular with diagonal entries lambda i. So, this theorem again I will not show the proof here. But the proof it basically first involves Schur's theorem to get an upper triangular form and then using a series of carefully chosen non unitary similarity transforms that produce this kind of upper, block upper triangular form and especially the zeros in the off-diagonal terms without changing the diagonal or the upper triangular structure of the matrix T. But this is going to be used later when we discuss the Jordan Canonical form.

(Refer Slide Time: 27:09)



So, next we will discuss about normal matrices.

Student: Sir, this matrix is a diagonal, right?

Professor: Which matrix?

Student: T1 to Tk this Matrix, similar matrix to A, this would be diagonal, right, because…

Professor: This is upper triangular. It is upper triangular so again keep in mind that this is a result which applies to any A which is of size n cross n. So, A need not be diagonalizable for the result to hold. If, I mean it is possible that you will end up with Ti's all being diagonal which is possible if A is diagonalizable.

But if A is not diagonalizable but it has these distinct eigenvalues lambda 1 to lambda k then A is similar to this kind of a block, upper triangular triangular matrix, concatenation of upper triangular matrices along the diagonal where each Ti is upper triangular with diagonal entries equal to the corresponding eigenvalues lambda i.