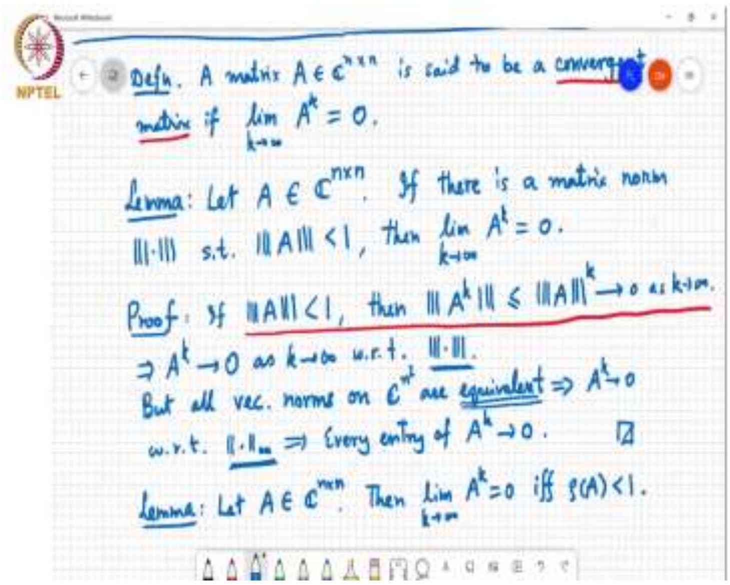
Matrix Theory

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

Convergent matrices, Banach lemma

The reason why we are looking at many of these are because ultimately when we use, when we develop algorithms for optimization, we have to deal with matrices and these matrices, you often end up with powers of these matrices and you want to know whether your algorithm converges and so, sort of a fundamental question is whether A power k is something that converges as k gets large. So, basically, let me maybe first say give this definition.

(Refer Slide Time: 01:08)



A matrix A c to the n cross n is said to be convergent matrix if k tending to infinity of A power k equals 0, so this is the n cross n all 0 matrix. So, this is a convergent matrix and what we will be interested in is, under what conditions can we say that a matrix A is going to be convergent. Obviously, you can imagine that in particular applications, you may not be, you may not have a matrix A such that A power k goes to 0.

But what we will see is that by suitably scaling the matrix, you can ensure that the matrix will be a convergent matrix. What I mean by that is if you scale a matrix, you know that all its Eigen values also scale by the same amount and so I can potentially scale a matrix such that all of its Eigen values become smaller and I can choose that scaling factor to be appropriate, so that all its Eigen values in magnitude get bounded between 0 and 1.

And so then all its eigenvalues are less than 1 and what we will see momentarily is that if, so the scale matrix has a spectral radius, which is less than one and if a matrix has a spectral radius less than 1, then it is convergent. That is what we will see. So, just so let us get there. So, we have the following lemma. It is going all over the place today for some reason.

Lemma let A be an n cross n matrix, if there is a matrix norm, such that is less than 1, then. So, what this means is that every entry of A power k will go to 0 as k goes to infinity. So, we will just quickly see, this is a one line proof, it is just that if norm of A is less than 1, then norm of A power k by sub multiplicativity it is less than equal to norm A power k and this quantity is less than 1.

So, when you take it to the large enough power, this quantity will go to 0 as k goes to infinity. So, basically what it says is that the norm of A power is going to 0 as K goes to infinity and so this matrix A, now you can I mean there is various ways to argue it. But one way is that this is a matrix norm, you are taking a matrix norm of some matrix called B.

Let us say B is equal to A power k, then if the norm of, this is a non-negative quantity and this A power k is less than or equal to something that is going to 0. So as k gets large, the norm of this matrix is going to go to 0 and if the norm of the matrix goes to 0, by property of matrix norm, that the positivity property of a matrix norm, if the norm of a matrix goes to 0, it must mean that the matrix itself must be going to 0.

So, that is one way to argue it. The other way to argue it is to say that, so A power k goes to 0 as k goes to infinity with respect to, this is something we saw again in a previous class, but with respect to this particular norm. But then, a norm is also a vector norm on C to the n squared, any matrix norm is also a vector norm of C to the n squared and all vector norms on C to the n square are equivalent.

So, what that means is that A power k goes to 0 in the sense of, with respect to the infinity norm as well or norm infinity, which means that this is the maximum magnitude entry, which implies every entry of A power k converges to 0. So, it was not quite a one line proof. But nonetheless, the core part of the proof is really this step here. Now…

Student: Sir? Sir what does this equivalent mean? Is it related to the homework assignment problem?

Professor: So, what yeah it is, probably it was the homework problem on this also, but it is a property that I discussed again in a previous class, that you can always bound; given a particular norm, you can bound any other norm in terms of, in terms of that norm. So there are constants, C small m and C capital M, such that the first norm is bounded between C small m times the second norm and C capital M times the second norm for any vector.

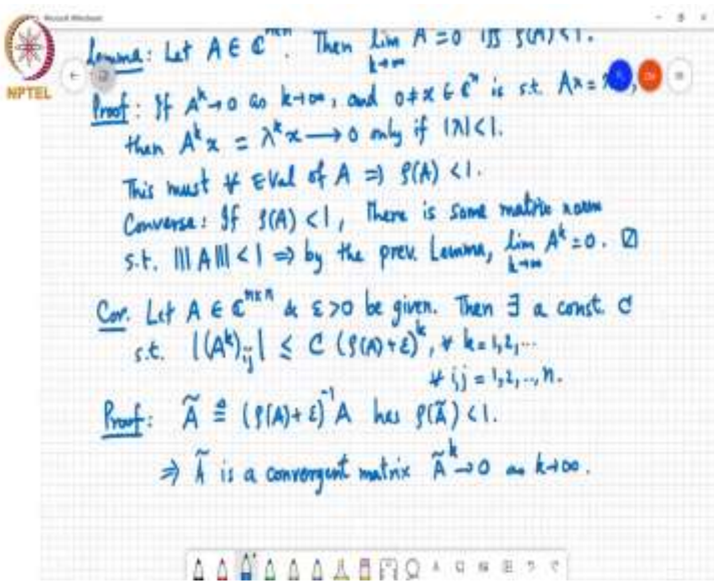
And so instead of this sum norm we have used here, we can instead replace that with this norm and since an upper bound on this norm is going to go to 0, then the infinity norm must also go to 0, which means that every entity of A power k must go to 0.

Student: Yes. Thank you, sir.

Professor: So, now, we saw. So, we know that if there is a norm for which this is less than 1 then limit is 0, then we also have seen that rho of A is a lower bound on any possible norm. So obviously, if rho is less than 1, I know that I can find a norm such that the norm of A is within rho of A plus epsilon.

So, if rho of A is strictly less than 1, then I can always find a norm under which the norm of A is strictly less than 1 and therefore, such a matrix will be convergent. So, that is this lemma here. Let A C to the n cross n. Then limit k going to infinity A power k equal to 0, if and only if rho of A is less than 1. So, this is thing more than what I just said it is an if and only if statement, but the other part is very easy.

(Refer Slide Time: 10:24)



If a power k if A power k goes to 0, as k goes to infinity and so 0 not equal to x in C to the n is such that A x equals lambda x then, so this is one of the eigenvectors of the matrix, then if I look at A power k x, this is just repeated multiplication of x with A each time I multiply I will get another lambda factor. So, this is going to be lambda power k times x and for this to go to 0 x is a nonzero fixed vector, so the only dependence on k is coming through lambda power k and so, this will go to 0 only if mod lambda is less than 1.

So, and this must hold for all eigenvalues of A, which implies that row of A is less than 1. So, that is the if part and for the converse if rho A is less than 1, less than 1 then there is sum norm such that norm of A is less than 1 and so, this implies that by the previous lemma limit k can be tend to infinity, A power k equals 0. So, if rho of A is less than 1 then this limit A power k as k goes to infinity will be equal to 0 that shows the other side.

So, I mentioned that in algorithms, we often want to consider scaled versions of matrices that will allow us to bound the entries of a matrix as you take higher and higher powers and this is a corollary that helps us bound the size of the entries of A power k as k goes to infinity, so that is this corollary.

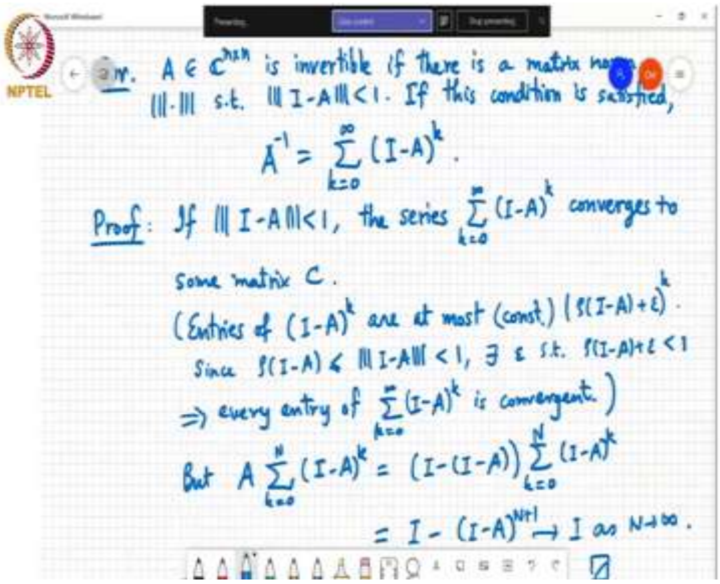
So, let A C to the n cross n and epsilon be some positive number, then there exists a constant C such that mod A power k i jth entry is less than or equal to this constant C times rho of A plus epsilon power k for k equal to 1, 2 and i j being any entry. So, that is the current ruling. So, that it the corollary, so what is says is that there is a way to base on this rho of A, you can bound the i jth entry of A power k in terms of some constant times rho of A plus epsilon power k.

So, proof is very simple again. These bonds come in very useful when you are analysing the convergence behaviour of algorithms. So, suppose I define A tilde to be rho of A plus epsilon inverse times A, then I claim that rho of A tilde is less than 1. Why is this true? I just scaled every entry of this matrix A, by row of A plus epsilon inverse and so, the Eigen values also get scaled by row of A plus epsilon inverse and in particular, the largest Eigen value in magnitude also gets scaled by rho A plus epsilon inverse.

And so, the largest eigenvalue then in magnitude becomes rho of A divided by rho of A plus epsilon which is going to be less than 1. So, this has rho of A tilde less than 1 which means A tilde is a convergent matrix. that is A tilde power k goes to 0 as k goes to infinity. So, one fundamental property of any convergent sequence.

So, now think of it as there is a sequence of matrices and this sequence of matrices which is A tilde A tilde square A tilde cube, A tilde power 4 et cetera, this sequence of matrices are converging, I can also think of this as there are n squared sequences, each sequence corresponding to a distinct entry of this matrix A tilde and all these n squared sequences are converging. One fundamental property of a convergence sequence is that every convergence sequence is bounded.

(Refer Slide Time: 18:04)



So, that means in words, the entries of A tilde K are bounded. So, what bounded in turn means is that there exists some constant C greater than 0 such that mod A tilde i j A tilde power k i j is less than or equal to this constant C and this is true for all k and for all i j and so, that implies now I will just remove the scaling that I applied and take the scaling factor to the other side.

So, that means that mod of A power k i j is less than or equal to this constant C times rho of A plus epsilon for every k equal to 1, 2 et cetera and for every i j, i j equal to 1, 2 to n. Now, if you remember two classes ago, we were discussing about invertibility of matrices and we made a small remark. Let me just find that here.

We said that in terms of convergence, if x is a scalar, let us say real or complex and mod x is less than 1, then 1 minus x inverse can be expanded as 1 plus x plus x squared plus et cetera and we were asking the question of it suggests this kind of a formula that if I want to find I minus A inverse, I can write that as I plus A plus A squared, plus et cetera.

And we asked when is this valid, this is valid for mod x less than 1 and we said that this is true if norm of A is less than 1 and any matrix norm will do and now we know that if rho is less than 1, then I can find a norm under which this condition will hold and so the equivalent condition is that rho of A should be less than 1, so that is the next lemma.

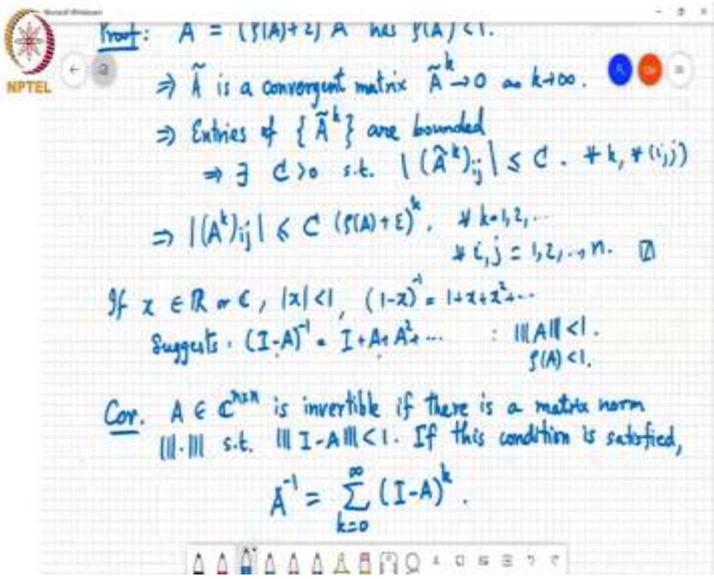
It is actually a corollary. So, A in C to the n cross n is invertible if there is a matrix norm, such that norm of I minus A is less than 1. If this condition is satisfied, then like I can write A inverse is equal to sigma k equal to 0 to infinity I minus A power k. So, notice that I wrote this in terms of a but now I am writing it in terms of I minus A.

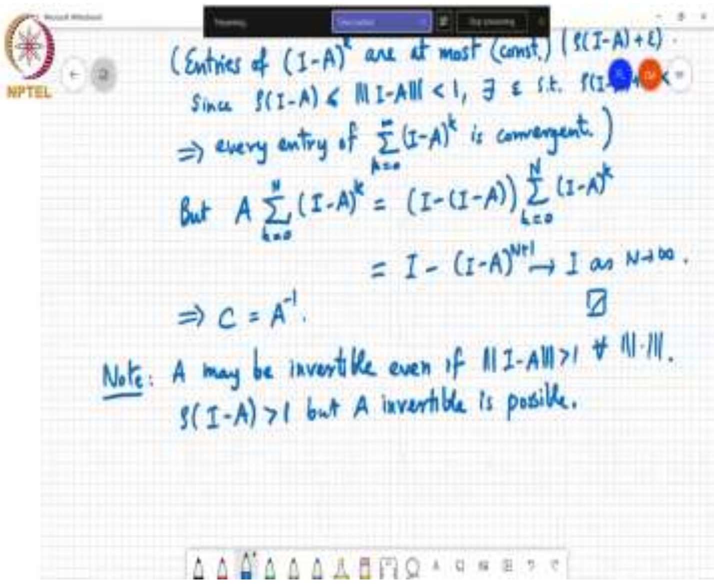
So, effectively you replace A by I minus A here, then this becomes A inverse and this becomes I plus I minus A plus I minus A whole squared plus et cetera and that is what this formula here says. So, in this condition here, gets replaced with norm of I minus A less than 1. So, I it is exactly what we said earlier.

At that point, it was a conjecture, but now we actually formally stating the result. So, given what we have seen so far again, the proof is very simple. So, if this norm of A, I minus A is less than 1, then the series sigma k equal to 0 to infinity, I minus A power k converges. Why? So, remember that I am not just looking at I minus A power k here, I am looking at the summation of such terms.

So, we have already seen that if the norm is less than one, we go here, it is here. So, we have already seen that if normal is less than 1, then limit a A power k is 0. We have seen that but what we are doing now is kind of adding up these A power K type of terms from k equal to 0 to infinity and what we are saying here is that this kind of a summation, it actually converges.

(Refer Slide Time: 24:44)





Let us say converges to sum matrix C. That is because the entries of this matrix I minus A power k of I minus A power k are at most we just saw this result that the magnitude of the entries of I minus A power k are at most some constant times rho of A rho of I minus A plus some small number epsilon power key.

And since rho of I minus A is some number that is less than or equal to this norm of I minus A, which is less than 1, it means that there exists some small enough epsilon, such that rho of I minus A plus epsilon is also less than 1 which then in turn implies that every entry of I minus A power k summation I minus A power k is convergent.

Now, so now all I have to do is to show that whatever this thing converges to is actually A inverse. So, that is very easy. So, I just do A times whatever is supposed to be A inverse. So, sigma k equal to 0 to n. So, I will take n terms and then make n go to infinity I minus A power k. This is equal to I will write a as I minus I minus A sigma k equal to 0 to n, I minus A power k.

Now, if I expand this out, so I will get I minus A I minus A squared et cetera when I multiply with i, but then I will get also negative terms which will start with I minus A and then you will get a minus A squared I minus A cube and so on, that the first term, the first I minus A term will cancel the I minus A term I get when I multiply I with this summation here.

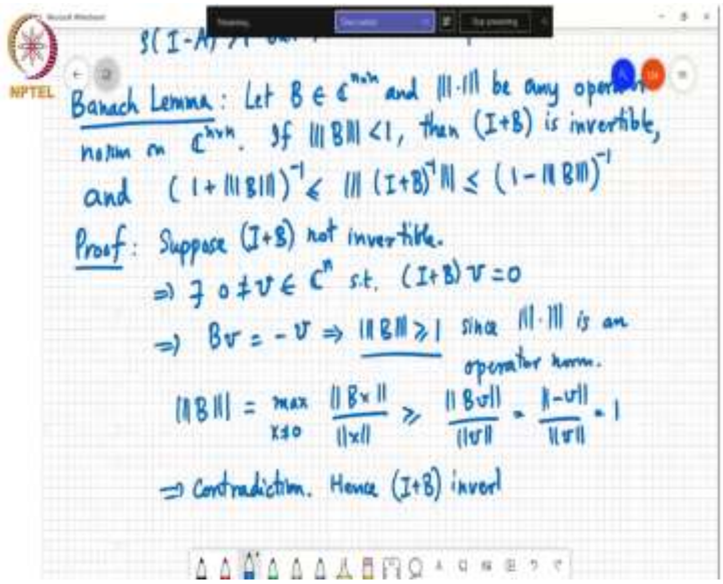
So, this is what is called a telescoping sum. So, all the alternate terms will cancel off and this is going to be equal to I minus I minus A power N plus 1 and this I minus A power N plus 1 goes to 0 the 0 matrix as N goes to infinity, so this converges to identity as N goes to infinity. So that is why when I to, so that is why I minus A, the summation k equal 0 to 0 to infinity I minus A power k will be equal to A inverse.

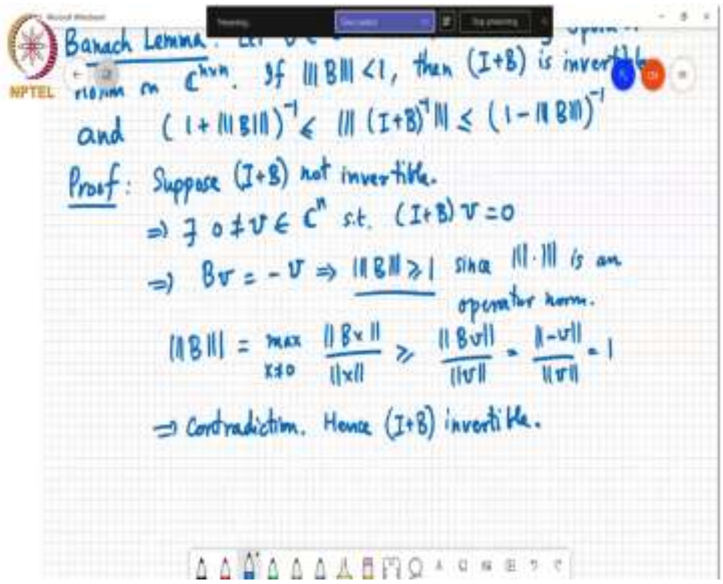
So, this is one lemma about how we can use the fact that the norm of I minus A is less than 1 to write A inverse as a series. But keep in mind that if norm of I minus A is greater than 1, it does not say anything about the invertibility of the matrix A, A could well be invertible even if norm of I minus A is greater than 1, you can easily see that.

So, obviously, rho of A could be greater than 1, if rho of a is greater than 1, then every matrix norm of, So, let me see actually to avoid confusion, let me write that note. A may be invertible. So, this is a one way result is all I am trying to say, even if I minus A is greater than 1 for all norms.

So, for example, rho of A rho of I minus A could be greater than 1, in which case every norm of I minus A will be greater than 1, but A invertible is possible, very closely related result is something quite the Banach lemma.

(Refer Slide Time: 30:53)





And I like to state it because the proof is slightly different. So, let B and C to the n cross n and be any operator norm, that is it is induced by a vector norm on C to be n cross N, norm of B is less than 1, then I plus B is invertible and 1 plus norm of B inverse less than or equal to norm of I plus B inverse is less than or equal to 1 minus norm of B in this.

So, this is a lemma that allows you to bound the norm of I plus B inverse in terms of the norm of B. It is at least 1 plus norm of B inverse and at most 1 minus normal B inverse. So, we will just quickly see how this is done. So, contradictions, so suppose I plus B is not invertible, then that means that there is a nonzero V in C to the n which lies in the null space, I plus B times V equals 0, because I plus B is a singular matrix.

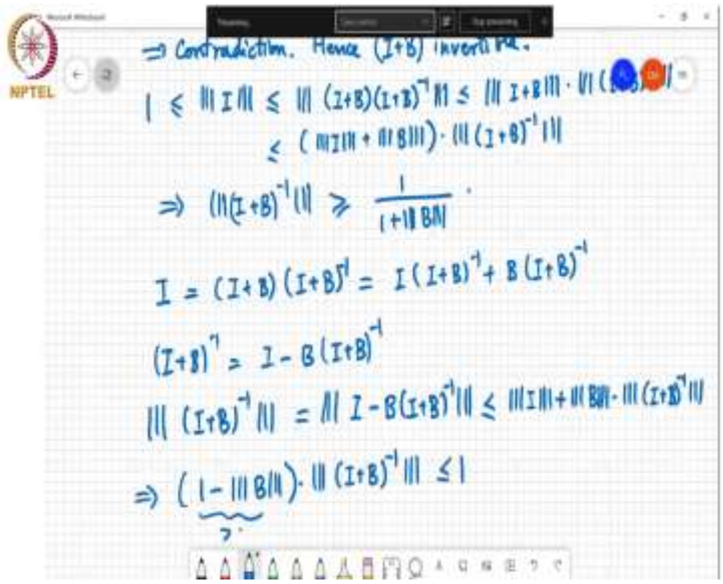
So, if I expand this out this means that B V equals minus v. So, that means that there is at least one Eigen value whose magnitude is 1. So, that means the largest magnitude Eigen value must also be greater than or equal to 1. So, norm B is greater than or equal to 1 since this norm is actually an induced norm. So, let me, it is not because of Eigen value being equal to 1 in magnitude, but because this is an induced norm.

Student: Sir?

Professor: I am just explaining this. So, if B V equals minus V, this norm here, norm B, it is a operator norm. So, this is actually equal to the max of let us say x not equal to 0 of norm this whatever norm that induced this norm B x divided by norm x. Now, if I choose a particular value of this, for this x that will only give me a lower bound on this, so this is going to be greater than or equal to norm of B V over norm V, but B V equals minus v.

So, that is equal to norm of minus V over norm V and norm of minus V is equal to norm V and this is one. So, that is basically this claim here. But then this is a contradiction, because we started out by saying that if norm B is less than 1. So, basically what that means is that norm B is less than 1, then I plus B must be invertible. Now, for the last step note that one.

(Refer Slide Time: 35:50)



We know that any norm for any norm the norm of the identity matrix is at least equal to 1. This follows from the sub multiplicative at of norms. In fact, for operator norms these are equal and this is in turn less than or equal to the norm of I plus B times I plus B inverse, which is less than or equal to the norm of I plus B times the norm of I plus B inverse, sub multiplicativity and I can further upper bound this by using triangle inequality and say that this is norm of I plus norm of B times norm of I plus B inverse.

So, we got one part already norm identity is equal to 1. So, that means norm of I plus B inverse is greater than or equal to 1 over 1 plus norm of B and similarly, for the second half again I is equal to I plus B times I plus B inverse, which I can write as I times I plus B inverse plus b times I plus B inverse and now I take norms. Rather, I will actually bring this to the other side and so, I plus B inverse is equal to I minus B times I plus B inverse.

So, now I take norm on both sides, I plus B inverse is equal to the norm of this quantity I minus B times I plus B inverse, again triangle inequality, norm of I and sub multiplicativity for the second term, norm of B times norm I plus B inverse. Now, I take this to the other side.

So, that means 1 minus norm of B times norm of I plus B inverse less than or equal to 1. So, just take this to the other side, this is greater than 0, so you can take this to the other side. So, norm of I plus B inverse is less than or equal to 1 over 1 minus norm of B. So, that concludes the proof. So, that is all we have time for today and we will continue again on Friday.