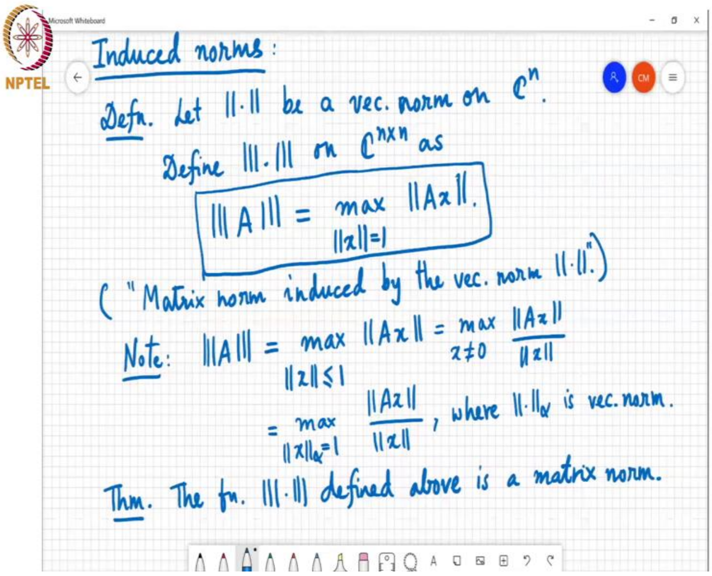
Matrix Theory
Professor Chandra R. Murthy
Department of Electrical Communication Engineering
Indian Institute of Science, Bangalore
Lecture 21
Induced norms

(Refer Slide Time: 00:17)



So, the next thing to talk about is induced norms. Basically, the idea is that we can associate matrix norms with vectors norms through this notion of induce norms. So, I will define it here. So, let be a vector norm on C to the n. So I am taking complex, like I said, a lot of the results here are applicable to complex as well. So, I may go back and forth between these two and where it is really required to make a distinction, I will tell you the difference.

Define on C to the n cross n as, so you take that norm and if I compute that norm on the vector A x and take its maximum over all vectors which have a unit norm, according to this vector norm, I will define that to be this with three bars, this function with three bars on the matrix A and so this is defined as the induced norm. So, this is the, so it is a, I will just put it in double quotes, because I have not really formally stated this yet. So, this is the matrix norm induced by the vector norm.

So, this is the basic definition of an induced norm. So one thing, you can immediately notice that if I scale x by some constant C, then the norm A x scales by the constant C and the norm of x also scales by the same constant C. Another thing is that I am trying to maximize this norm A x over all x such that norm x equals 1, I can as well maximize this over all x such that norm x is less than or equal to 1, because if there is some vector that solves, so let me write that here, so that I can explain more clearly.

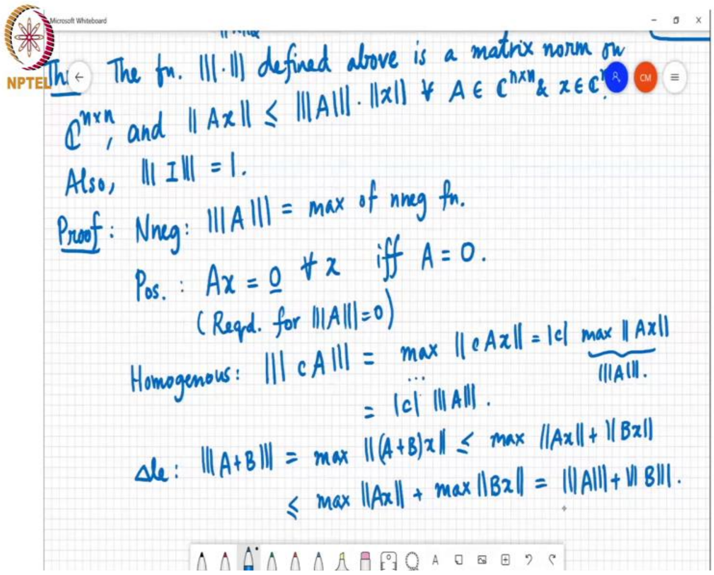
Note, is equal to max over all x such that x less than or equal to 1 of norm A x. That is because suppose you solve this problem and it gives you a solution for which norm of x is strictly less than 1, then what I can do is, I can scale that x by a factor which is greater than 1, then this norm will be made may become equal to 1. But then when I scale that x by some factor greater than 1, then this norm of A x will also scale by a factor greater than 1 and since and that will only increase its value.

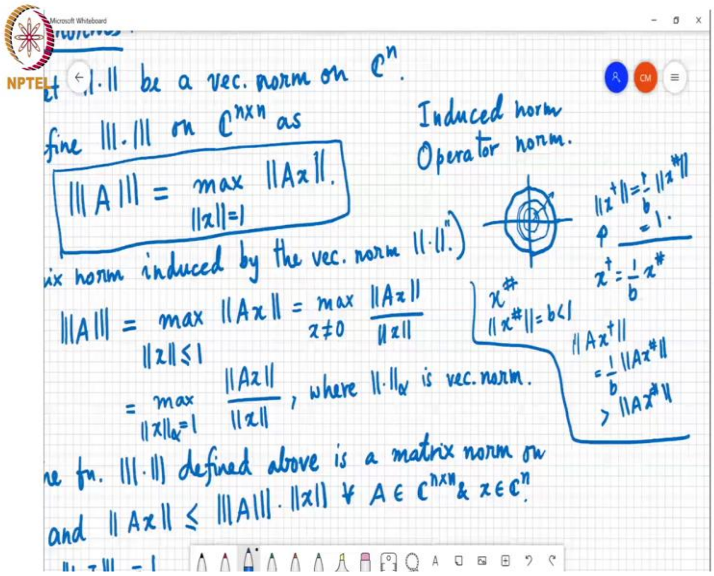
So basically, what that means is that the solution to this optimization problem will always occur at an x such that norm of x equals 1. So, there is no loss or no harm, if I include a whole bunch of other points which are inside the unit circle according to this norm. Because those points will never be this solution to this problem and this in turn can be written as the max over all x not equal to 0 of norm x over norm x.

This quantity itself does not change if you scale x, so I can as well maximize norm A x over norm x over all x not equal to 0 and finally, I can also instead of maximizing over all x naught equal to 0, I can also maximize over all x such that norm of x according to some notion of norm, I will call it x alpha equal to 1 of norm A x over norm x.

So, different ways of writing this is, is any vector norm. So, there are different ways of writing it, but they all give you the same answer, which is what we are calling norm of A. So, I have been calling it the norm of A, but actually we need to show that. So, the theorem is that the function, defined above is a matrix norm.

(Refer Slide Time: 06:23)





On C to the n cross n and norm A x, so this is a vector, so this norm is the vector norm that induced this quantity here with three buyers. This is less than or equal to the norm of A times the vector norm of x for every A and x in C to the n. So, pick any vector this norm of A is actually an upper bound on norm of A x divided by norm of x. If x equals 0, this is 0 and this is also 0. So, it holds when x equals 0 as well.

Also, the norm of the identity matrix equals 1. So, remember previously we had said that the norm of an identity, of the identity matrix is greater than or equal to 1 for any vector norm, any matrix norm, that because the norm of I squared is less than or equal to norm of I times the norm of I and therefore, but I squared equals I and therefore norm of I is greater than or equal to 1 for the identity matrix I. For these kinds of norms, induced norms, norm of I equals

1. This is also known as a so this is called induced norm or operator norm, so we will just write that here.

Student: Sir the definition of induced norm is valid for all type of vector norm?

Professor: Yes.

Student: And sir can you please explain that note part again, I could not understand that, why this max of (())(8:41) of x less than equal to 1 should be equal to norm of A.

Professor: So, by definition, this is the max over all norm x all vectors x such that norm x equals 1 and I have to find this norm A x, I search over all vectors which have unit norm and find the vector that maximizes the norm of A x. Now, instead of searching over all vectors x such that norm of x equals 1, suppose I search over all vectors for which norm of x is less than or equal to 1.

That includes, so I guess, first of all, a very intuitive and simple way of thinking about it is, suppose I consider the this norm to be the l 2 norm. Then basically when I say norm x equals 1, I am asking you to search over points lying on this circle and A is some matrix and among all the points on the circle, I am asking you which point maximizes norm A x, then basically as you go further and further out on the circle, this norm of A x will also increase. So, if I take a point x here versus is a point x here, the norm of A x over here will be bigger than the norm of x over here, if these are on the same line, the same direction.

So, if for example, when you solve this problem, if you get an if you say my solution is some particular x hash, such that norm of x hash is equal to some value say B, which is less than 1. So, then this maximum is happening at an interior point, then what I can do is I can propose a better solution, I can say my solution is x dagger, which is equal to your 1 over b times x hash.

If I do this, then if I look at what happens to basically norm of A x, this is equal to 1 over b times norm A x hash. So, whatever maximum you found, I am able to find an x dagger, notice that this also satisfies my constraint is actually equal to 1 over b times the norm of x hash, which is equal to b and so, this is equal to 1.

So, this satisfies the constraint that norm of x should be less than or equal to 1, but the value of the objective function A x dagger is going to be strictly greater than A x dagger itself. So, whatever solution you found, I am able to find a solution that beats your solution it achieves that higher maximum and as a consequence of that the solution to this problem will always occur at a point where norm of x equals 1 and so, you do not lose anything by saying I will maximize overall x such that norm x is less than or equal to 1, it will always give you the same solution as the earlier optimization problem.

Student: Thank you sir.

Professor: Yeah, I do not have to restrict to x less than or equal to 1, I can maximize over all x not equal to 0 by simply including norm x and the denominator of this objective function. That is the A. So, this is the theorem. So, basically whatever we defined earlier is in fact a matrix norm.

So, let us check this. So, again we need to verify non-negativity positivity, homogeneity, triangle inequality and sub multiplicative at. Now, obviously.

Student: Sir?

Professor: Yeah.

Student: Sir, some of this is matrix norm is with double bars and in some cases with triple bars, could you explain that?

Professor: It is always triple bars, except when I was giving a couple of examples here and I or somebody else already asked the question, I suppose you were not paying attention and I explained that, I am going to use a different, I am going to define the l 2 norm of A with three bars in a different way and so I am saving the triple bar here for that purpose and so here is a l 1 and l 2 I have defined the two bars because I am going to use a different, I am going to define a different norm for the l 1 norm and l 2 norm. Otherwise matrix norms are with triple bars.

Student: Okay, sir.

Professor: So now, these are the four properties. So first, we will take non-negativity. So obviously, by definition, the norm of A is the maximum of a non-negative quantity. This norm of A x, this is a vector norm and it is a non-negative quantity. So, when I take the maximum of a non-negative quantity, I cannot suddenly get a negative quantity. So, it is non-negative.

The positivity is because the, this norm of A will be equal to 0 if and only if A x equals 0 for all x x naught equal to 0. So, for that if you the simplest way is to consider this version of writing out norm A is the maximum over all x naught equal to 0 of norm A A x over norm x. So, for this maximum value to be equal to 0, the value itself must be equal to 0 for all x.

Now, if norm A has to be if norm of A x is 0 for all x, it means that A x itself this is a vector now, it means a x equals 0 for all x. So, and this is true if and only if A is equal to 0. Why is that true? Why should A x equals 0 for all x mean that A is equal to 0?

Student: Positivity property of the vector norm sir.

Professor: No, there is no norm here, I am just saying that A matrix is such that if I find A x, it gives me the 0 vector for all x and I am saying,

Student: We can express the columns of the, We can express the columns of the matrix as the transformation of the basis elements.

Professor: I do not follow that.

Student: Any metrics, if we add the matrix on the basis elements, then we will get the columns of the matrix and if all of them are 0, then the matrix will be 0.

Professor: So, you are saying that you want, you want to think of a basis in say C to the n and you want to project the columns of the matrix on this basis and given that somehow conclude that A x equals 0 means A equals 0.

Student: Yes, sir.

Professor: Well, I can think of a simpler way of saying that. So, suppose I.

Student: Range space of A 0 means is A 0, always.

Professor: Yeah, that is a, so I am masking why is that true and so one very simple way to see that is that if I take x equals the vector 100000, then what that will do is to pull out the first column of A, since A x equals 0 for all x, it is also true for this vector 10000. That means that the first column of A must be 0.

Next, if I consider x equals 01000, that will, that A x for such a vector will be equal to the column A 2, but that is also equal to 0, so the second column is also equal to 0 and so on. So, if A x equals 0 for all x, then A is equal to 0. Similarly, I mean, on the contrary is trivial. Because if A is equal to 0, then of course, x is equal to 0 for all x.

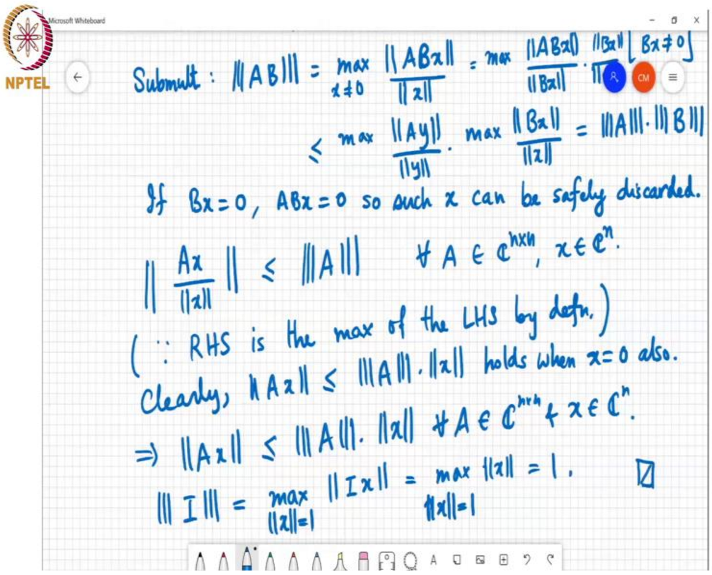
Student: Yes, sir same proof.

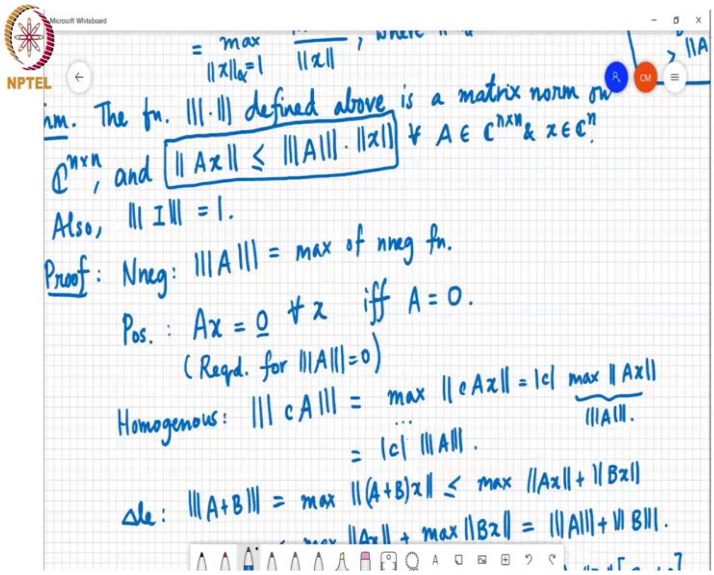
Professor: So, the next property is homogeneity. So, we need to show that. So, if I consider the norm of C times this matrix A, this is equal to the max of, so I would not write the constraint here just because it is just repetitive. It is the same constraint x norm of x equals 1, so I would not write that.

C A x and this is a vector norm, so I can pull out the C and write this as mod C and this mod C does not depend on x. So I can in fact, pull it out of the max also times the max of A x and this is actually equal to more C times norm of x. So, it does satisfy the homogeneity property. This by definition is the norm of A.

Then triangle inequality. So, I take A plus B this is equal to the max of A plus B times x which is equal to. So, this is the norm of A x plus B x and if I use the triangle inequality this is less than or equal to the max of norm A x plus non B x in turn this is less than or equal to, so instead of maximizing the sum of these two quantities, I can individually maximize these two terms and add them up that can only increase the value. So, that is max of norm A x plus max of norm B x and this by definition is norm of A plus norm of B. So, it satisfies the triangle inequality.

(Refer Slide Time: 21:17)





And sub multiplicativity. So, if I consider a norm of A B, this is equal to max of I will consider the other form, so I will write it like this A B x over norm x, over all x not equal to 0. So, this is equal to, so for a moment, let us say B x is not equal to 0, then I can multiply and divide by B x by norm of B x and now, I have the max of the product of two terms and what I can do is, I can individually maximize these two terms, they are non-negative terms.

So, instead of maximizing the product of these two terms, I can individually maximize these two terms and then take the product of the two maxima that will only be greater than or equal to this term. So, I can write it like this and of course, I can consider B x to be some other vector y and maximize over all y not equal to 0.

So, max of norm A y over norm y times max of B x over norm x and this is equal to B. But of course, if B x is equal to 0, then A B x is also equal to 0. So, the this quantity you are trying to maximize this will be 0 and so, such kinds of xs can be discarded safely from this maximization problem.

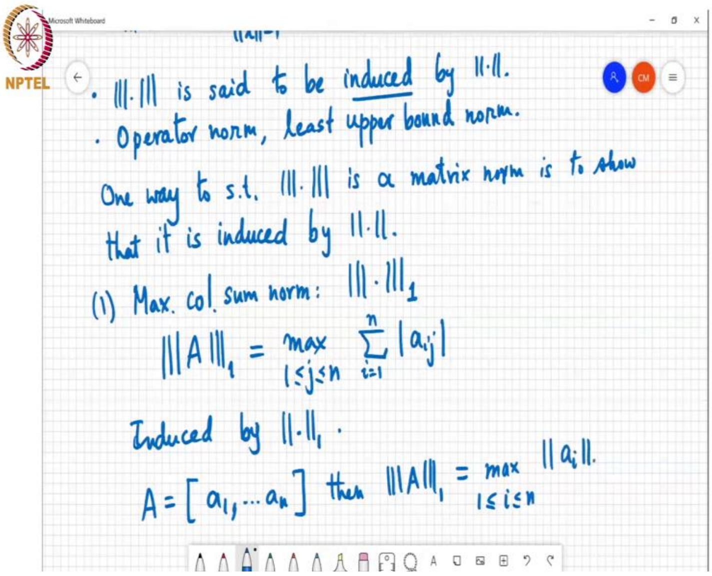
They would not be the maximum. In fact, they are achieving the minimum value that this can achieve which is 0 and the final part is that there is this, the statement of the theorem also says one inequality which is that the norm of A x is less than or equal to the matrix now of A times the vector norm of x for every A and x, we need to show that.

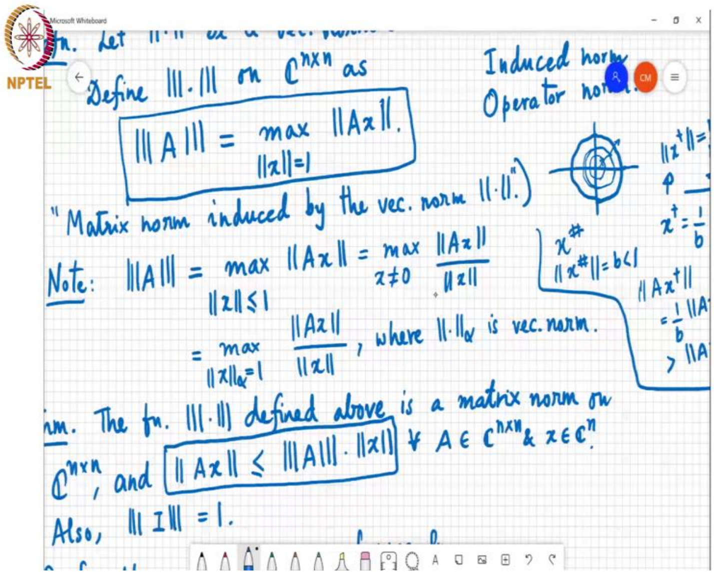
Of course, that is trivial because if you consider norm of A x over norm x, this is norm of A x divided by norm x and this is less than or equal to the matrix norm of A because by definition, by definition the right hand side is the maximum value of the left hand side over all x not equal to 0 and so, for any A and x, this is true and so, I will write that just for completeness, the RHS is the max of the left hand side by definition.

So, the maximum of a certain function is always greater than or equal to the value obtained by it at any particular point. So, this of course, assumes that x is not equal to 0, but clearly this inequality is valid when x equals 0 also, clearly this A x less than or equal to norm A norm x holds when x equals 0 also.

So, basically this in turn, I can simply take this norm of x to the other side and hence norm A x less than or equal to the matrix norm of A times the vector norm of x and finally, if I consider the norm of the identity matrix, this is equal to the max over all x such that norm of x equals 1 of the norm of identity times x, but this is just norm of x itself. So, the constraints that says norm of x, you are maximizing overall x such that norm x equals 1 of norm x itself and so this is equal to 1, so that concludes this proof. Any questions?

(Refer Slide Time: 27:32)





So, this matrix norm is said to be induced by the vector norm. So, it is also called the operator norm or it is also called the least upper bound norm. This is basically says the least upper bound norm, this name comes from the fact that the by definition, this norm is the maximum over all x equal to 1 of the norm of A x. So in other words, it is the smallest upper bound you can place on norm of A x over all x satisfying norm x equals 1 or if you look at this definition here, it is the smallest upper bound you can place on norm A x over norm x all x not equal to 0. So, that is why it is called also called the least upper bound norm.

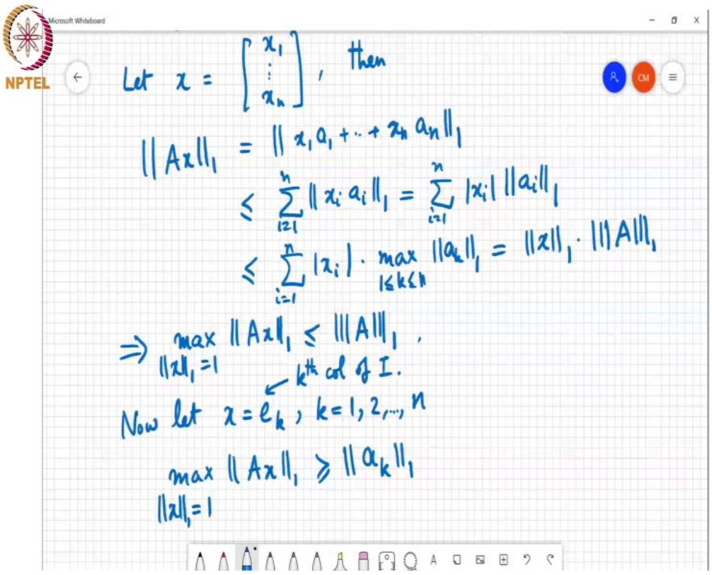
So, basically the fact that the operator norm or the induced norm is a matrix norm is a general property and it is true for all vector norms and so, one way to show that some particular definition of a norm is indeed a matrix norm, is to show that it is in fact induced by some vector norm and so now we have another way to show that.

So, one way to show something is a matrix norm is to show that it is induced by some vector norm. So, here is an example. So, now I am going to introduce the l 1 norm with three bars. So this is called the max column sum norm. So, I am going to denote this by with three bars and the 1 next to it. So, it is defined as max 1 less than or equal to j less than or equal to n of sigma i equal to 1 to n mod a i j.

So, this, so what I am doing here is, I am taking the some of the magnitude of entries in each column of the matrix. So, when j equals 1 it is the first column, I am taking this l 1 norm of the first column and I am looking across all columns first, second, third, up to the nth column, I am asking, I am asking which column has the maximum l 1 norm and I am defining that to be the matrix l 1 norm of this matrix A.

So, this is induced by the vector norm l 1. So, to quickly see that, if I have A equal to a 1 a n, then this norm a 1 is equal to the max 1 less than or equal to i less than or equal to n of the l 1 norm of a i, as I just mentioned this norm here is taking the max column max of the l 1 norms of the columns of A, the largest l 1 norm among the columns of A is this matrix.

(Refer Slide Time: 32:09)



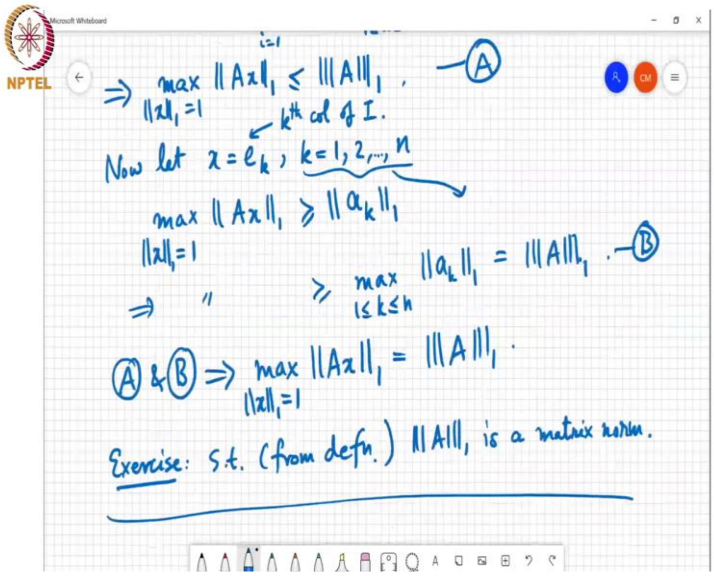
Now, if we let x to be equal to this vector, x 1 through x n, then if I look at A x l 1, this is equal to the l 1 norm of x 1 a 1 plus plus x n a n, the entries of this vector x, multiply the corresponding columns of A and this in turn is less than or equal to I am using triangle inequality, sigma i equal to 1 to n l 1 norm of x i a i.

So, I have just taking the norm inside the summation, that is triangle inequality, which is equal to now x i is a scalar here, so I can pull that out, sigma i equal to 1 to n mod x i times the l 1 norm of a i, which in turn is less than or equal to. So, I can replace all these by the largest column norm and that will only increase the value. So, sigma i equal to 1 to n mod x i times the max 1 less than or equal to k less than or equal to n norm of A k and this by definition is the l 1 norm as defined here, it is the max column sum.

And so this is equal to the l 1 norm of x, the sum of all the entries, it does not depend on i anymore does not come out of the summation, times the matrix norm of A. So, what this means is that for any x, the l 1 norm of A x is less than or equal to, so I will write that this implies. This is true for any x the l 1 norm of A x is less than or equal to the norm of x, the l 1 norm of x times the matrix norm of A and so, if I, it is also true that if I take, since that is true for every x, if I take the max over all x such that x l 1 equals 1 of the norm of A x, this is going to be less than or equal to the matrix norm A.

But I want to show that these are actually equal that, this definition that I have here is in fact induced by this for that I need to do a little bit more work. So, now let x is equal to e k k equal to 1 2 up to n. So, this e k is the kth column of i. So, then what I have is that max norm x equals 1 of A x is at least equal to the value it takes at 1 of these an example vector which is for example, e 1 if I take e 1 that satisfies l 1 norm of e 1 equals 1 and I can, that will give me the l 1 norm of the first column of A. So, in general if I take the kth column, then it will be A k l 1. So, this is also true and so, then this is true for all of these k, k equal to 1 to up to n.

(Refer Slide Time: 36:05)



Which implies that this maximum of this is greater than or equal to the max 1 less than or equal to k less than or equal to n of norm A k l 1 which by definition is the matrix normal of A. So, I have shown that this max of this is less than or equal to norm of A l 1, I have also shown that this max of this is greater than or equal to norm of A l 1. So, if I call this step A and this step B, these two together imply max 1 A x l 1 equal to.

So, this shows that this may, the max column some norm as I defined it is in fact induced by the l 1 norm. So, it is an exercise for you to show that this is a matrix norm from by showing that it satisfies the four properties for a matrix norm. So, that is all I am have time for today. I am actually a little over time.