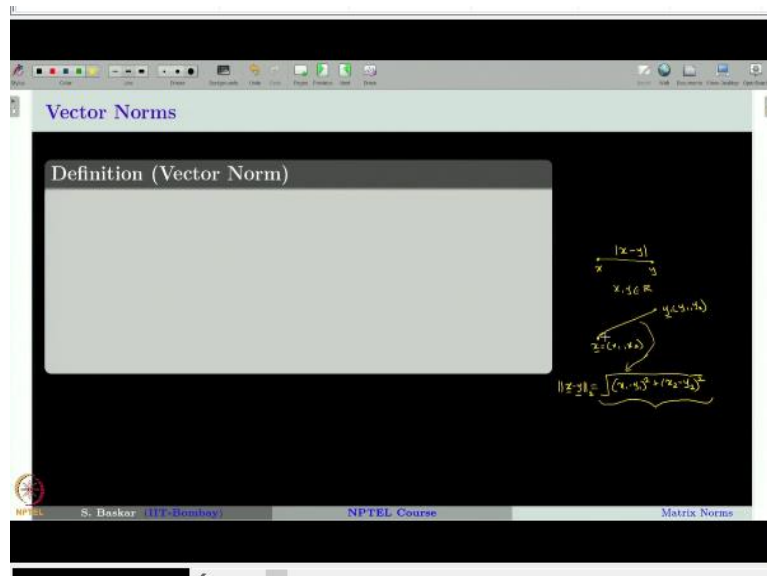


Numerical Analysis
Prof. Dr. S. Baskar
Department of Mathematics
Indian Institute of Technology-Bombay

Lecture-15
Matrix Norms (Subordinate Matrix Norms)

Hi, in the last few lectures we have finished the direct methods for solving linear systems. As the next topic we are supposed to take up iterative methods for linear systems, but we will take a small deviation from our usual stream and learn a tool called matrix norms. This tool is very important for us to do error analysis for iterative methods.

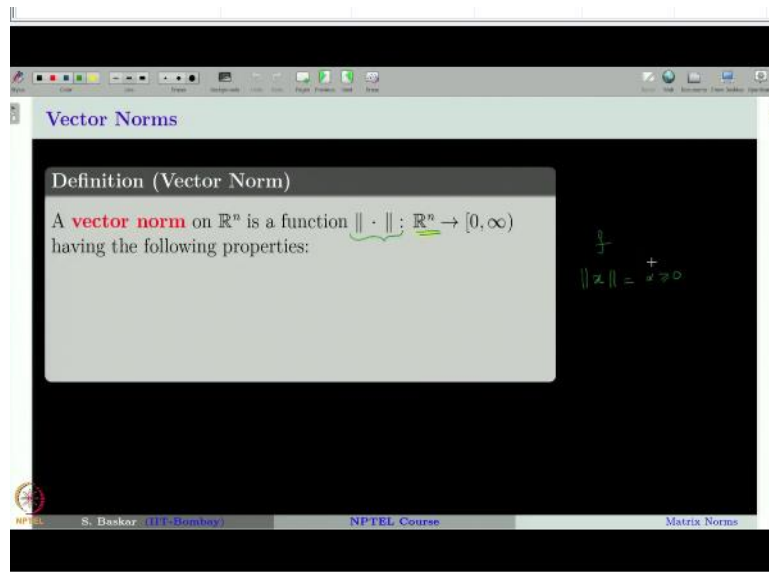
(Refer Slide Time: 00:49)



We will start our discussion with vector norms. Well, first let us understand what is meant by a norm? A norm is an abstract version of a distance concept. We all know that if we have two points x and y , say both x and y are real numbers then the distance between x and y is given by $|x - y|$. Similarly, if x is a vector in the plane given by (x_1, x_2) and similarly y is a vector on a plane given by (y_1, y_2) .

Then the distance between these two points is given by $\sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$. We know that this is called the Euclidean distance or Euclidean norm. So, we generally denote it by $\|x - y\|_2$. This is the physical distance between the two points in whatever unit that we measure. Now our interest is to generalize this idea and in that way we can bring in more functions into this class which can mimic this distance idea. That is what the aim of defining norms.

(Refer Slide Time: 02:35)

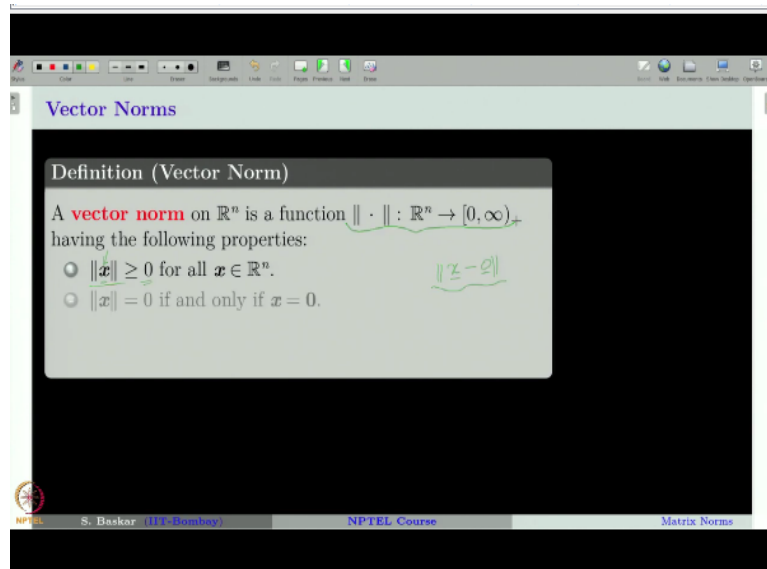


Let us see the definition of vector norm. A vector norm is basically a function defined on \mathbb{R}_n , such that when you apply that function on a vector x in \mathbb{R}_n it gives a non negative real number. That is the basic need of a distance. That is distance between two points is always a non-negative real number. Let us denote this norm by this notation generally a function means we use notation f, g and so on.

But here we have this notation where you just give any point x that gives us a number which is say α which is greater than or equal to 0 and not only that this function has few properties what we do is we just carefully observe some of the important properties that a distance function satisfies that is the Euclidean distance in the case of \mathbb{R}_2 and similarly this idea can be generalized to any \mathbb{R}_n and in one dimension it is just the modulus.

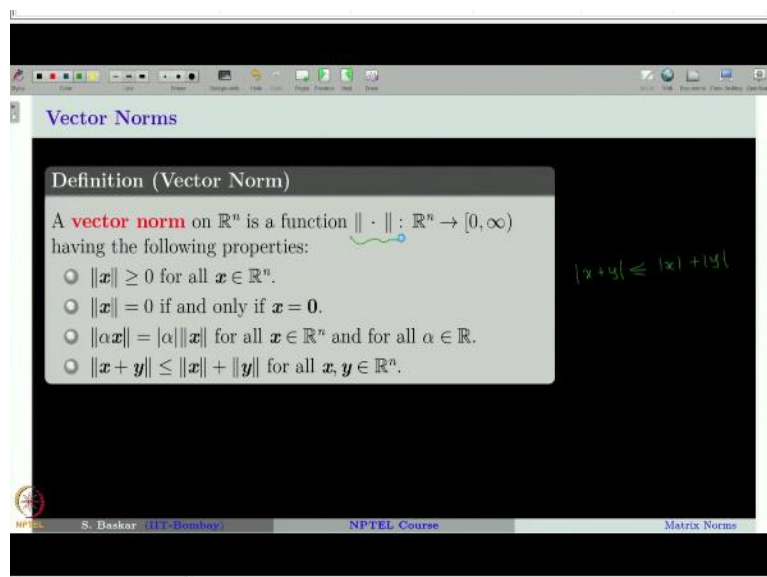
So, we will carefully observe some of the important properties of this distance functions and just impose those conditions on this function and call the resulting function as a vector norm. That is the idea, obviously if you take any vector x in \mathbb{R}_n and if you find the distance between x and the 0 vector.

(Refer Slide Time: 04:30)



That is what we denote by this. So, this will now indicate in some sense a distance function. So, you should see the definition of norm by keeping the distance concept in mind. So, the norm should give some number for a given vector x which is non-negative. That is the first condition that we want our function to satisfy. Well, that is already inbuilt in the way we have defined our function.

(Refer Slide Time: 05:11)



The next condition is that the distance between x and the origin 0 is 0 if and only if that x itself is in the origin. That is also something which you can easily understand without any problem. The same condition is also imposed for our norm function. Next is the scaling property; you multiply α with x and then find the distance. That should be same as you find the distance between α and 0 .

And then multiply with $|\alpha|$. They both should be the same. So, this condition is also imposed on the vector norm and finally the well-known triangle inequality, which is satisfied by the Euclidean distance also in 1D, modulus satisfies this that is $|x + y| \leq |x| + |y|$ is a very simple property that we all know. We are imposing that property also on to our function. And if all these conditions are satisfied by this function, then we will call that function as vector norm.

(Refer Slide Time: 06:37)

Vector Norms (contd.)

Notation: $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \mathbb{R}^n$.

Example:

- **Euclidean norm** on \mathbb{R}^n is denoted by $\|\cdot\|_2$, and is defined by

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$$
- **l_∞ norm**, which is also called **maximum norm**, on \mathbb{R}^n is denoted by $\|\cdot\|_\infty$, and is defined by

$$\|\mathbf{x}\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$$
- **l_1 norm** on \mathbb{R}^n is denoted by $\|\cdot\|_1$ and is defined by

$$\|\mathbf{x}\|_1 = |x_1| + |x_2| + \dots + |x_n|$$

NPTEL S. Baskar (IIT-Bombay) NPTEL Course Matrix Norms

Well, we will see some examples. First is the physically realistic distance function called Euclidean norm and that is given like this when x belongs to \mathbb{R}_n . We wrote it for $n = 2$ which will be the same for any n that we take and this is the Euclidean norm and that is the physically realistic way of finding distance between a given point x and the origin. Now we can extend this distance concept to something abstract and say that there are also other functions that can mimic our distance function.

One example is what we call as l_∞ norm and that is denoted by this notation with the suffix infinity and its definition is, you take the modulus of each component of the vector x and then take the maximum of all this guys and that is what we will assign as infinite norm of x . You can verify that all the important properties that we listed in the previous slide that is satisfied by the physically realistic distance function is also satisfied by this function.

See the number that this gives for a given x and the number given by this function for a given x may be different but this function will satisfy all the properties that this function satisfies. That is the only idea behind this abstract concept of norm. The next example is what we call as

l_1 norm and that is denoted by these parallel lines with a suffix 1 and its definition is that you take modulus of all the components of the vector x and then sum them up.

You can again verify all the properties listed in the definition of vector norm by this function and therefore it is also a vector norm. Remember these are the three norms that are commonly used in numerical analysis while doing error analysis which involves any vectors. However, there are many other ways that we can define vector norms, but we will not go into listing all of them.

Whatever we do in our course we will use one or the other of these three norms only. Mostly we will use l_∞ norm, because it is very convenient to use. You may ask the question when we have the physically realistic distance concept why are we putting effort to make an abstract idea out of this distance concept and get all these definitions which are not physically realistic; they just mimic the distance concept but they are not going to give you physically what comes as the distance.

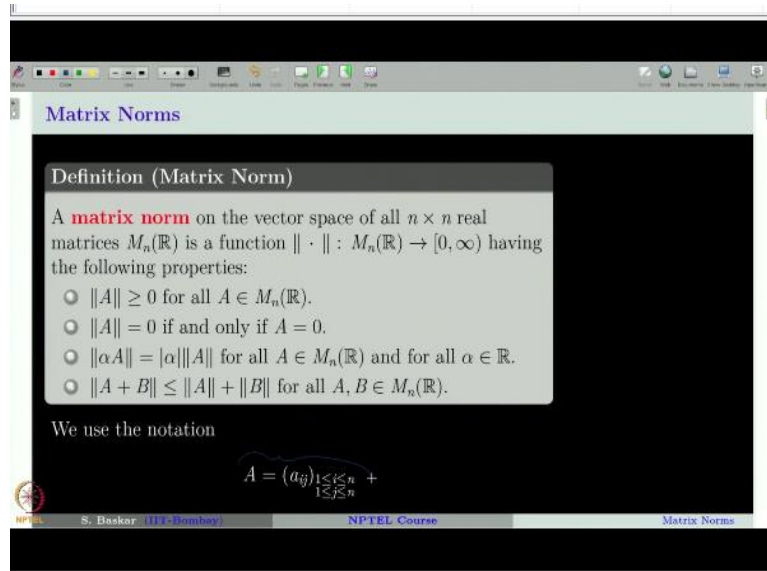
So, that is so why are we even worrying about this abstract concept? Well, there are two reasons for why we are interested in making this distance concept abstract and considering some of this which are actually physically not realistic. There are two reasons; one is that we need to find distance not only between two vectors we will also come across situations where we want distance between two matrices.

Now how will you even imagine finding distance between two matrices or in general we may have to also find distance between two operators or two functions something like that. So, it is not very clear how to define distance between two such mathematical objects? So, one needs to put this distance concept in abstract form so that that can be used to adopt some distance between two objects such as two matrices.

That is one reason for going for an abstract definition of distance. The next one is often physically realistic distance concept like this are not very easy to handle, they are very difficult to handle. Therefore, to understand the errors and their behaviour we may have to simplify our problem by going for some other distance concepts which are perhaps equivalent to studying such error analysis with the physically realistic distance concept.

That is why we also want to take up some of the other ways of defining the distance which may not be physically realistic, but is some way equivalent to our physically realistic distance. These are the two reasons why we go for putting the distance concept in an abstract way.

(Refer Slide Time: 12:26)



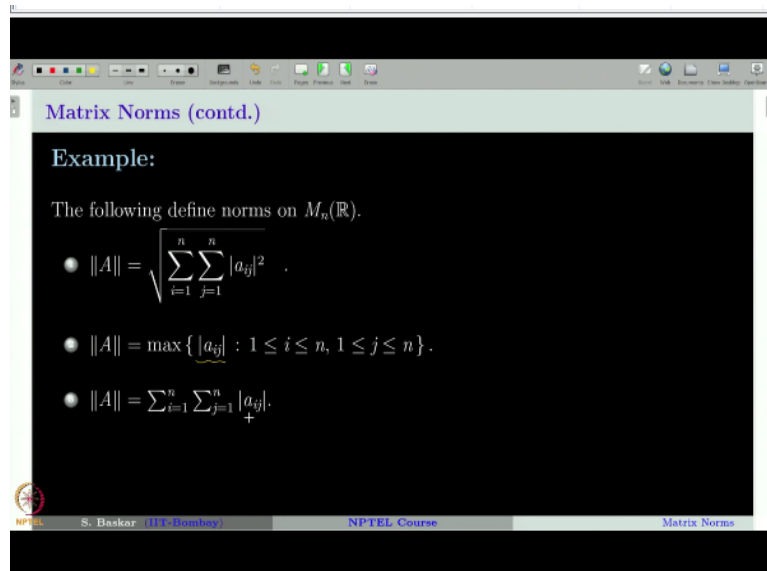
Well, we understood this and now let us go to define what is meant by matrix norm. Remember wherever I say norm you have to keep in mind the distance concept and see norm as a generalization of the distance concept. Now we are trying to make a tool to measure distance between two matrices. Well once you understand how we measure distance between two vectors.

Now we can just mimic that to define what is meant by matrix norm. Again, we will use the same notation for matrix norm; we will generally not distinguish the matrix norm with vector norm through its notation because the argument itself will tell us whether we are having a vector norm or a matrix norm. So, a function denoted by this symbol which is now going to be a function between the set of all $n \times n$ matrices with real entries to non-negative real numbers.

And that function should have the following properties in order to be called as matrix norm. The first property is of course given here that you take any matrix A , plug in into this function that function should generate a number that should be greater than or equal to 0. The next property is if that number given by the matrix norm is 0 it means you have given the 0 matrix and the third condition is the scaling property.

The same set of properties that we have listed for vector norm is also listed here that is all. αA you take and then take the norm that should be equal to $|\alpha| \|A\|$. That should hold for any real number α . That is important and finally the well-known triangle inequality should also be satisfied. Well, we will generally use this usual notation for matrices.

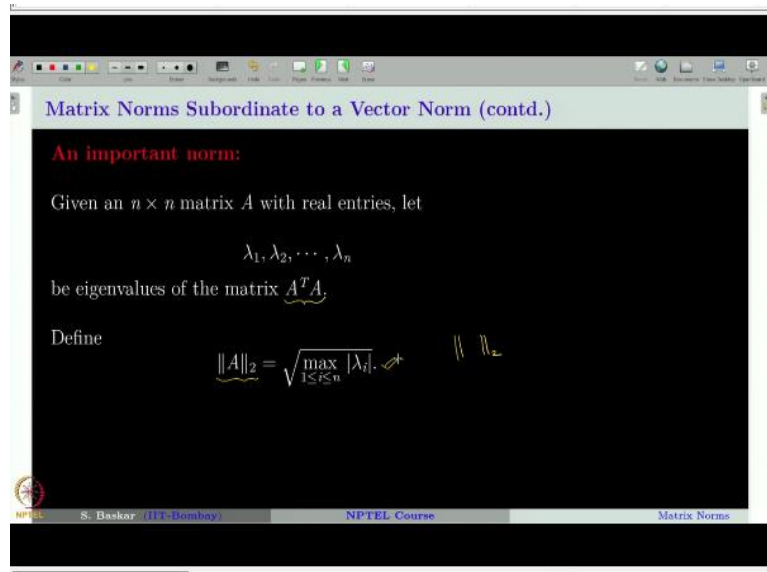
(Refer Slide Time: 14:46)



Just keep this notation in mind and let us see some examples of matrix norms. First thing is if you define a function which takes a matrix A and gives this as the number, you can see that clearly it is a non-negative number and you can also check that all the other properties of the matrix norm is satisfied by this expression. Therefore, a function given by this for a given matrix is indeed a matrix norm.

Similarly, you take the modulus of all the elements of the matrix A and then take the maximum of them. That will also define a matrix norm and finally take all the absolute values of the elements of A and then sum them all up that also will define a matrix norm. These are some of the examples of matrix norms. There are also other examples.

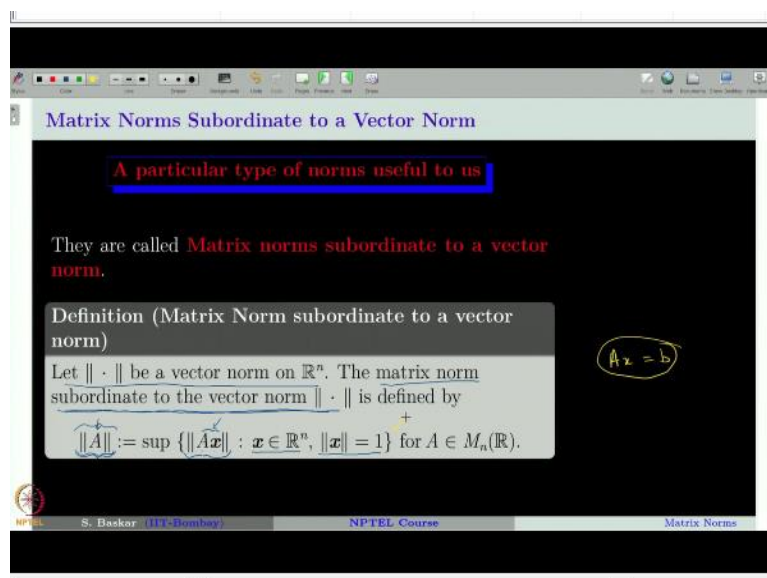
(Refer Slide Time: 15:57)



An important example of a matrix norm which is quite common in applications; I will tell you why it is so important little later. Now let me just define this matrix norm, it is generally denoted by this notation that is norm with a suffix 2 and it is defined as the maximum of the modulus of all the eigenvalues of the matrix $A^T A$, it is not the eigenvalues of A , but it is the eigenvalues of $A^T A$.

That is the point that one has to keep in mind. So, this particular norm is very important you can check that it is indeed a matrix norm.

(Refer Slide Time: 16:51)



Now we will discuss a very useful class of norms called matrix norm subordinate to vector norms. The idea is you give me a vector norm, I will generate a matrix norm with the help of that vector norm that is the idea. How am I going to generate a matrix norm for a given vector

norm? Let us see. As I told you are given a vector norm on \mathbb{R}_n , now I am going to give you a matrix norm which is so called subordinate to the vector norm and it is defined as $\sup \{\|Ax\| : x \in \mathbb{R}_n, \|x\| = 1\}$.

That is, you take all the unit vectors in \mathbb{R}_n , find Ax and then take the vector norm of all this vectors Ax . Remember whenever we put these two parallel lines on both sides of a vector it denotes a vector norm and the same if you put on both sides of a matrix then it is a matrix norm. Subordinate matrix norm is very useful because this particular way of defining a matrix norm for a given vector norm is going to satisfy three important properties which are very useful in our error analysis.

Also, when we are dealing with the error analysis of an iterative method for solving $Ax = b$, you will naturally see that the equation is some way combining a vector and matrix. Therefore, you have to have some way of linking your matrix norm with the vector norm in order to get some feasible results. That is why we are coming up with this kind of ideas. That is, we would like to work with a matrix norm that is generated from the vector norm in this way.

In other words we will consider a particular vector norm and then we will take the corresponding matrix norm subordinate to that vector norm and we will use these two norms in our error analysis for iterative methods in the next section. So, for that we have to understand the concept of subordinate norms.

(Refer Slide Time: 19:41)

Matrix Norms Subordinate to a Vector Norm (contd.)

Lemma

For any $A \in M_n(\mathbb{R})$ and a given vector norm $\|\cdot\|$, we have

$$\|A\| = \max_{z \neq 0} \frac{\|Az\|}{\|z\|}.$$

Proof: For any $z \neq 0$, we have $x = z/\|z\|$ as a unit vector. Hence

$$\begin{aligned} \max_{\|x\|=1} \|Ax\| &= \max_{\|z\| \neq 0} \left\| A \left(\frac{z}{\|z\|} \right) \right\| \\ &= \max_{\|z\| \neq 0} \frac{\|Az\|}{\|z\|}. \end{aligned}$$

S. Baskar (IIT-Tirupur) NPTEL Course Matrix Norms

Let us put the definition of subordinate norm in a little different way that is more useful for us in working with them. That is, you are given any matrix A and a given vector norm. Now you know at least mathematically how to define a subordinate matrix norm, you can see that that subordinate norm can also be written in this way. Remember by definition we have to take the supremum over unit vectors.

But now what we are doing we are taking the supremum over all non-zero vectors. That is the only difference, but this should not be a difficult thing to understand because you take any non-zero vector z you can always write $x = \frac{z}{\|z\|}$. Then you can see that x will become a unit vector. Therefore, your subordinate matrix norm which is by definition given like this, you are taking the supremum over all unit vectors, that can be now written as, simply substitute this expression into x and that gives you maximum over all $z \neq 0$, A , instead of x now I am putting this expression.

Just take z outside and then you will get this expression. That is what you would like to show here. So, it is a very simple proof and its very useful way of defining the subordinate norms rather than the norm that we have defined. That is only for the definition sake but in most of the results we will be using this definition for subordinate norm.

(Refer Slide Time: 21:53)

Matrix Norms Subordinate to a Vector Norm (contd.)

The matrix norm subordinate to a vector norm has additional properties.

Theorem

Let $\|\cdot\|$ be a matrix norm subordinate to a vector norm. Then

- $\|Ax\| \leq \|A\| \|x\|$ for all $x \in \mathbb{R}^n$.
- $\|I\| = 1$ where I is the Identity matrix.
- $\|AB\| \leq \|A\| \|B\|$ for all $A, B \in M_n(\mathbb{R})$.

Handwritten proof for the third property:

$$\|AB\| = \max_{x \neq 0} \frac{\|ABx\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|A\| \|Bx\|}{\|x\|} = \|A\| \max_{x \neq 0} \frac{\|Bx\|}{\|x\|} = \|A\| \|B\|$$

Let me list all the three properties that I told previously which are very important for us in doing error analysis in the next section. The properties of the matrix norm that we are interested

are here. First is $\|Ax\| \leq \|A\| \|x\|$. Remember this is a vector norm, this is a matrix norm and again this is a vector norm. Let us see how to prove this first result.

Well, if $x = 0$ it is very clear because in this case Ax is 0, therefore norm of 0 is 0. On the right-hand side norm of 0 vector is again 0. Therefore, you have $0 = 0$, therefore if $x = 0$ this equality holds in the first case. Now let us take $x \neq 0$ then you can write $\frac{\|Ax\|}{\|x\|} \leq \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$.

Because this is just any vector from this set and the right-hand side is obtained by taking maximum over all such vectors. Therefore, this value should surely be less than or at most equal to this and that is nothing but $\|A\|$ by our previous Lemma. From here we can immediately get the first inequality. The second inequality can be easily obtained from the first inequality.

That should not be a problem. Let us take the third property. In the third property what you can do is you just write $\frac{\|ABx\|}{\|x\|}$ and take the maximum over all $x \neq 0$, that is nothing but your $\|AB\|$ by definition. Now this can be seen as A into some vector therefore by first property you can write this as $\max_{x \neq 0} \frac{\|A\| \|Bx\|}{\|x\|}$.

Again, you apply the first inequality for this term $\|Bx\|$ that will be less than or equal to $\max_{x \neq 0} \frac{\|A\| \|B\| \|x\|}{\|x\|}$. Now you see x and x gets canceled and now what remains is independent of x . Therefore, that is equal to $\|A\| \|B\|$ and that is the proof of the third inequality. Well, these are some of the properties that we will be using quite often in our error analysis of iteration methods.

(Refer Slide Time: 25:31)

Matrix Norms Subordinate to a Vector Norm (contd.)

Theorem (Matrix norm subordinate to the maximum norm)

The matrix norm subordinate to the l_∞ norm (also called **maximum norm**) on \mathbb{R}^n

$$\|x\|_\infty = \max\{|x_1|, |x_2|, \dots, |x_n|\}$$

is denoted by $\|A\|_\infty$ and is given by

$$\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

The norm $\|A\|_\infty$ given by the above formula is called **maximum-of-row-sums norm**.

S. Baskar (IIT-Bombay) NPTEL Course Matrix Norms

Now the interesting part is, I will give you some vector norm how can I generate a matrix norm subordinate to that vector norm. If you carefully look at the definition it is very abstract you cannot make any sense of how the matrix norm will look like for a given vector norm. But interestingly you can get a very nice formula for subordinate matrix norm for the important three vectors norms that we have given at the beginning of this lecture.

That is l_2 norm, l_∞ norm and l_1 norm. Let us give the formulas for the matrix norm subordinate to these three vector norms, we will not prove these theorems but we will just consider them as formulas. Suppose your given norm is l_∞ norm, that is recall your vector now is given like this, then the corresponding subordinate matrix norm is defined like this. How will you do that?

If you carefully see you take each row and find the sum of the absolute values of the column elements of that row and then sum that. Now in a $n \times n$ matrix you have n such rows, therefore you have n such numbers like this. Now take the maximum of that. So, this is called the maximum of row sum norm.

(Refer Slide Time: 27:08)

Matrix Norms Subordinate to a Vector Norm (contd.)

Example:
 Let us now compute $\|A\|_\infty$ for the matrix

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{pmatrix}.$$

$$\sum_{j=1}^3 |a_{1j}| = |1| + |1| + |-1| = 3, \quad \|A\|_\infty = 5$$

$$\sum_{j=1}^3 |a_{2j}| = |1| + |2| + |-2| = 5,$$

$$\sum_{j=1}^3 |a_{3j}| = |-2| + |1| + |1| = 4.$$

S. Baskar (IIT-Bombay) NPTEL Course Matrix Norms

Let us see an example. Suppose you are given this matrix, from the first row you get the number 3 by taking the absolute value of each column of that matrix and then sum them up. Then the second row will give 5 and the third row will give 4. Therefore, the matrix norm subordinate to the l_∞ norm denoted by $\|A\|_\infty$ is given by 5 because you have to take the maximum over these three numbers.

(Refer Slide Time: 27:45)

Matrix Norms Subordinate to a Vector Norm (contd.)

Theorem (Matrix Norm Subordinate to the l_1 -norm)
 The matrix norm subordinate to the l_1 -norm on \mathbb{R}^n

$$\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|. \quad (1)$$

is denoted by $\|A\|_1$ and is given by

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|. \quad (2)$$

The norm $\|A\|_1$ is called the **maximum-of-column-sums norm** of A .

S. Baskar (IIT-Bombay) NPTEL Course Matrix Norms

Now for the matrix norm subordinate to l_1 norm it is just row should be replaced by column in the previous definition that is you take each column and take the absolute values of the elements of each row in that column and then sum them. Like that you have n numbers now. For each column you have one number, like that n numbers. Now you take the maximum of all those numbers and that is called the maximum of column sum norm. And this is the matrix norm subordinate to l_1 norm; you can just compute the l_1 norm of the previous matrix.

(Refer Slide Time: 28:30)

Matrix Norms Subordinate to a Vector Norm (contd.)

Theorem (Matrix norm subordinate to the Euclidean norm)
The matrix norm subordinate to the l_2 norm (also called **Euclidean norm**) on \mathbb{R}^n

$$\|x\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2}$$

is denoted by $\|A\|_2$

$$\|A\|_2 = \sqrt{\max_{1 \leq i \leq n} |\lambda_i|}$$

where $\lambda_1, \lambda_2, \dots, \lambda_n$ are eigenvalues of the matrix $A^T A$.

NPTEL S. Baskar (IIT-Bombay) NPTEL Course Matrix Norms

Now coming to the l_2 norm that is the Euclidean norm. That is if you are given the vector norm as the Euclidean norm the physically realistic norm given by this if you recall. Then the corresponding subordinate norm is the one which I have already introduced that is maximum of all the eigenvalues of the matrix $A^T A$. This is very important; it is the eigenvalues of the matrix $A^T A$ and then take the square root here.

And that is the matrix norm subordinate to the physically realistic norm that is the Euclidean norm. Now you can see that if you have to work with the physically realistic norm then the corresponding matrix norm involves the computation of the eigenvalue of this matrix. So, it involves the product of these two matrices and then finding the eigenvalues of this matrix. So, that is computationally very costly.

That is why we often do not prefer to work with l_2 norm rather we will try to work with l_∞ or l_1 norm because you can see that the subordinate norms in the case of l_1 and l_∞ are very easy to compute. So, that is the reason why also we want to look for some equivalent way of doing our error analysis rather than going for this physically realistic distance concept.

(Refer Slide Time: 30:12)

Matrix Norms Subordinate to a Vector Norm (contd.)

Example:

Let us now compute $\|A\|_2$ for the matrix

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 2 & -2 \\ -2 & 1 & 1 \end{pmatrix}$$

The eigenvalues of $A^T A$ are

$$\lambda_1 \approx 0.0616, \lambda_2 \approx 5.0256 \text{ and } \lambda_3 \approx 12.9128.$$

Hence $\|A\|_2 \approx \sqrt{12.9128} \approx 3.5934$. \square

S. Baskor (IIT-Bombay) NPTEL Course Matrix Norms

Again, let us take an example and compute the two norms for the matrix A . For that you have to find the eigenvalues of $A^T A$. In this case they are given by these numbers and clearly the l_2 norm of this matrix is $\sqrt{12.9128}$ and that is given by this number. This is the l_2 norm of the matrix A given by this. In the next class we will see a concept called condition number of a matrix which will tell you how sensitive it is for us to perform any computation of a matrix on a computer.

That in fact involves matrix norm and therefore it is very important for us to understand the meaning of matrix norm and also you have to understand how to compute the matrix norm subordinate to a given vector norm in particular, we will be restricting ourselves to one of these three norms. Thank you for your attention.