Course Name: Optimization Theory and Algorithms

Professor Name: Dr. Uday K. Khankhoje

Department Name: Electrical Engineering

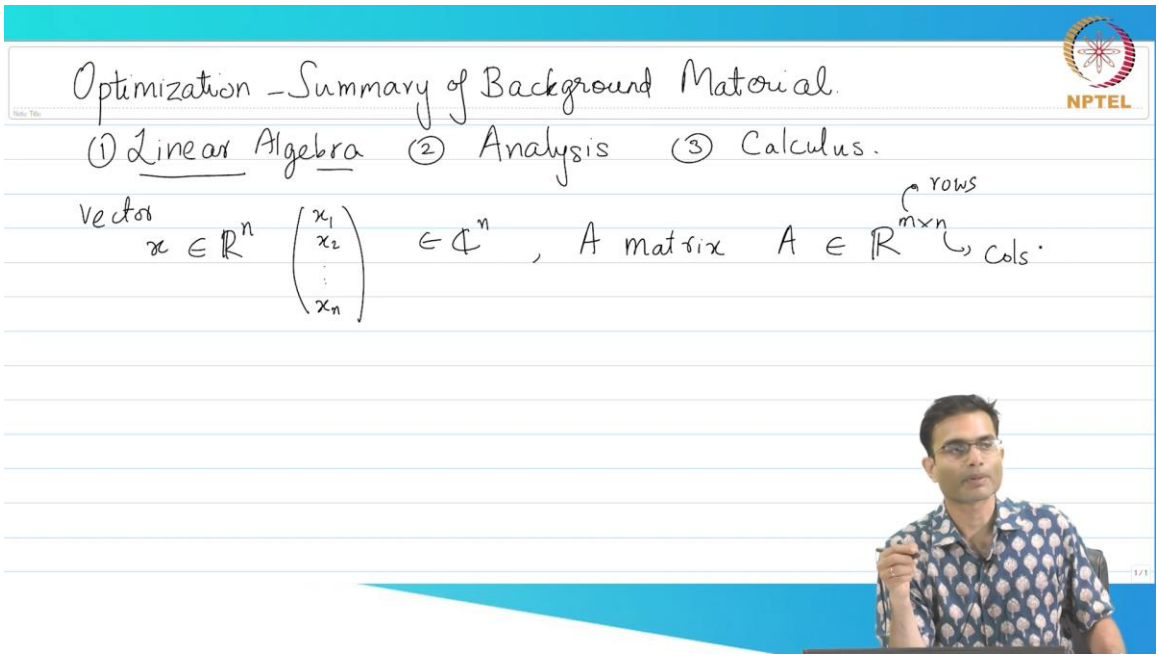
Institute Name: Indian Institute of Technology Madras

Week – 01

Lecture - 04

Summary of background material - Linear Algebra 1

Alright so today's class is going to be just like a very fast run through of background material which we will need so that we can understand the rest of the course on optimization. So, I shall warn you in advance it may sound a little boring because you're not deriving, we're just going to state one after the other just as a means of refreshing your mind okay. and roughly there are three things that we need to review ok. So, those are linear algebra, second is analysis and the third is let us say calculus ok. You have all sort of covered this in some way or the other at some basic level alright. So, let us start with the first topic of linear algebra ok.



Now, in general when I write a symbol like x you are used to it being just a scalar, but in general here x is going to be either I mean something like this

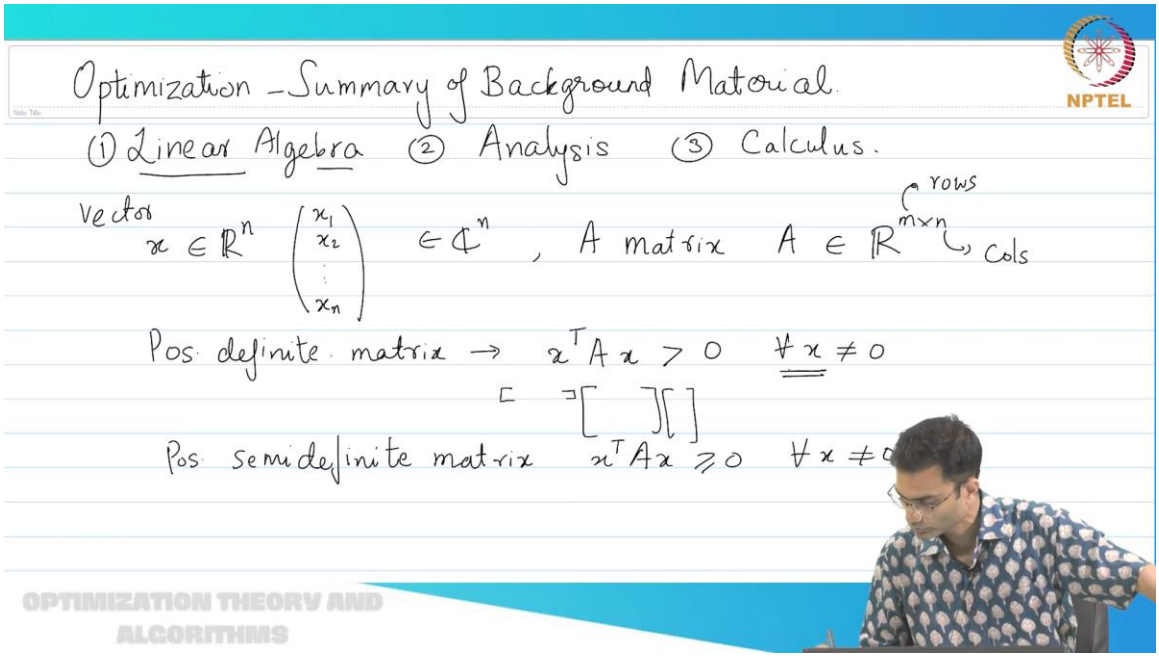
$$R^n$$

right. So, everyone knows what this means a set of n real numbers right and typically the most common way of representing it would be like a column vector right. If I wanted to talk of complex numbers, I would simply do

$$C^n,$$

ok. Now most of this course we will stay with real numbers.

There is a perfect analog for complex numbers as well, once we master real numbers complex numbers will not be very difficult, ok. So, that is our very basic notation, ok. A matrix, so this was a vector, a matrix on the other hand how do I represent? So, I will write something like A belongs to right so that when I write this it should tell you that I have m rows and n columns right that is the usual the same convention ok. Now, there is one particular type of matrix that we will come across very often in this course I will mention it over here which is a positive definite matrix ok. So, let us see a positive definite.



Can someone remind me what the definition of a positive definite matrix is?

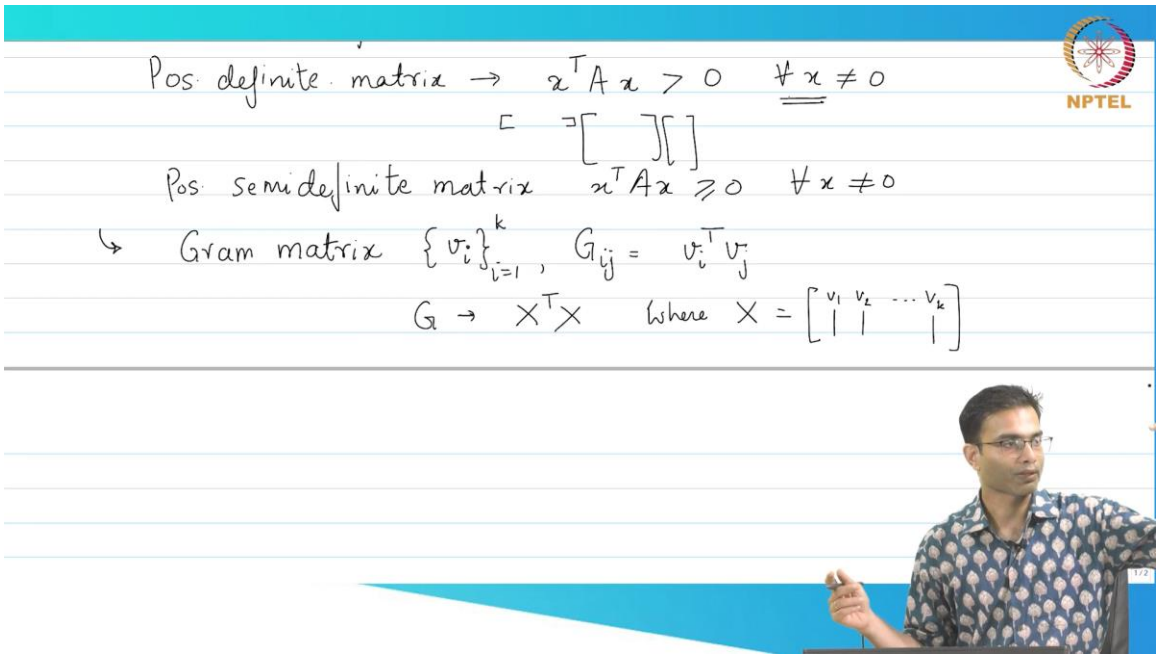
$$x^T A x > 0 \;\; \forall \, x \neq 0$$

Now, this for all x is important right it says pick any x you want if I stick it in in this way a nice way to visualize this is this is going to be x^T is a row vector then I have a matrix

and then I have a vector over here right. Now $x^T Ax$ looks like a big expression, but it is boiling down into a single scalar. Therefore, I can compare a single scalar against 0 and say greater than 0 right. So, this is a positive definite matrix as I mentioned we will come across it quite often in this course.

Slightly related concept is positive semi definite matrix right. So, positive semi definite matrix what is the only difference? greater than equal to 0. So, let us just put this concept into use, just to sort of revise your linear algebra. So, there is something called a Gram matrix. So, let us write that down.

Has anyone heard of a Gram matrix? Probably not. Again, a very useful concept over here. So, how do I define a Gram matrix? First, I am going to give you a sequence of vectors v_1, v_2, v_3 up to whatever. and the way I define the sequence is like this. If you see curly bracket curly brackets it means there are more than 1.



So, I can write i is equal to 1 to k. It simply means that there are k vectors v_1, v_2, v_3 up to v_k shorthand notation is in curly brackets that is the notation. So, I have given you let us say k vectors and how do I define this gram matrix by defining you for you the ij^{th} element. So, how do I define g_{ij}, ok. This I will write simply

$$G_{ij} = v_i^T v_j.$$

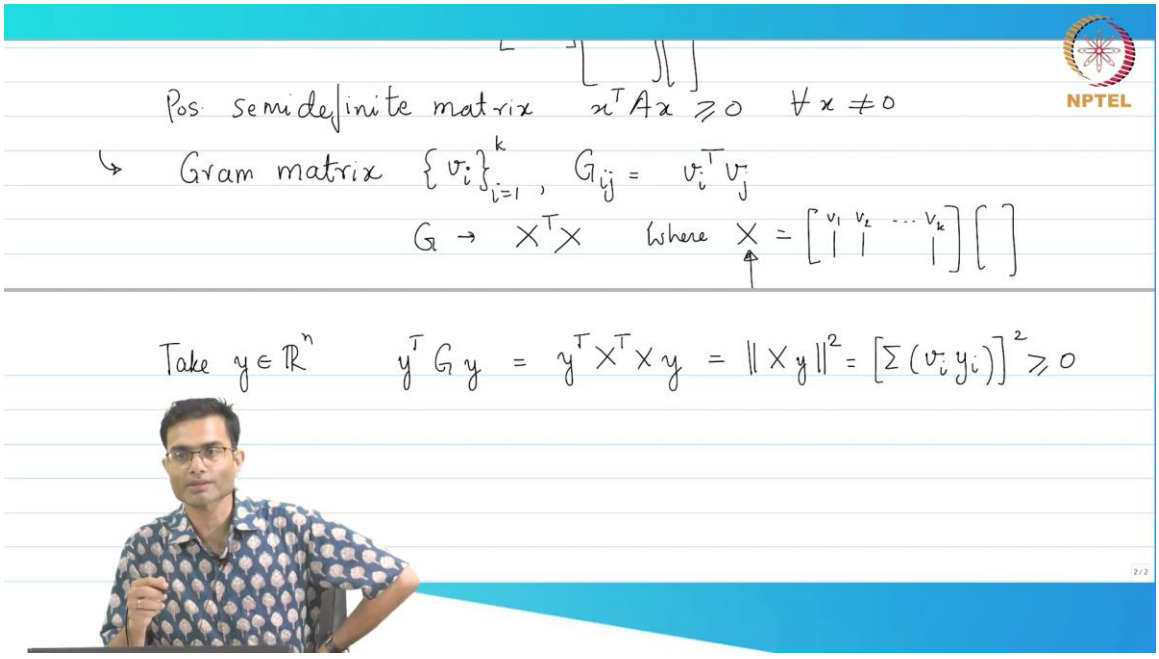
So, it is the inner product right everyone is familiar with this kind of a thing vector transpose times vector is an inner product is it a scalar or a vector. The scalar right. So, I am giving you the ij^{th} element of this matrix and it is this scalar quantity ok. Now, let us try to see is there something special about this matrix is it for example, positive definite is it positive semi definite or we cannot say anything what do you think? So, before we answer. So, let us approach this systematically.

Can I have defined for you the ij^{th}element, but can I define the entire matrix in one like in one compact way? Right. So, I have the correct answer over here. It is like the transpose of a matrix with another matrix. So, what is that matrix? You can write

$$G = X^T X, \text{ where } X = [v_1\ v_2\ v_3]$$

The matrix of vectors that I have right.

So, v_1, v_2, v_k. Does that make sense? Now you can visualize $x^T$is making them into row vectors and then when I multiply it with x, I am going to get exactly G, ok. Now, let us tackle the question of whether it is positive definite or not right. So, to see whether it is positive definite or not what do I need to do? I need to stick one vector and another vector the same vector right. So, supposing I take let us take a vector any vector which is non-zero and I do $y^T Gy$. This is what I have to look at in order to conclude whether or not it is positive definite and then this is going to get simplified as ok.

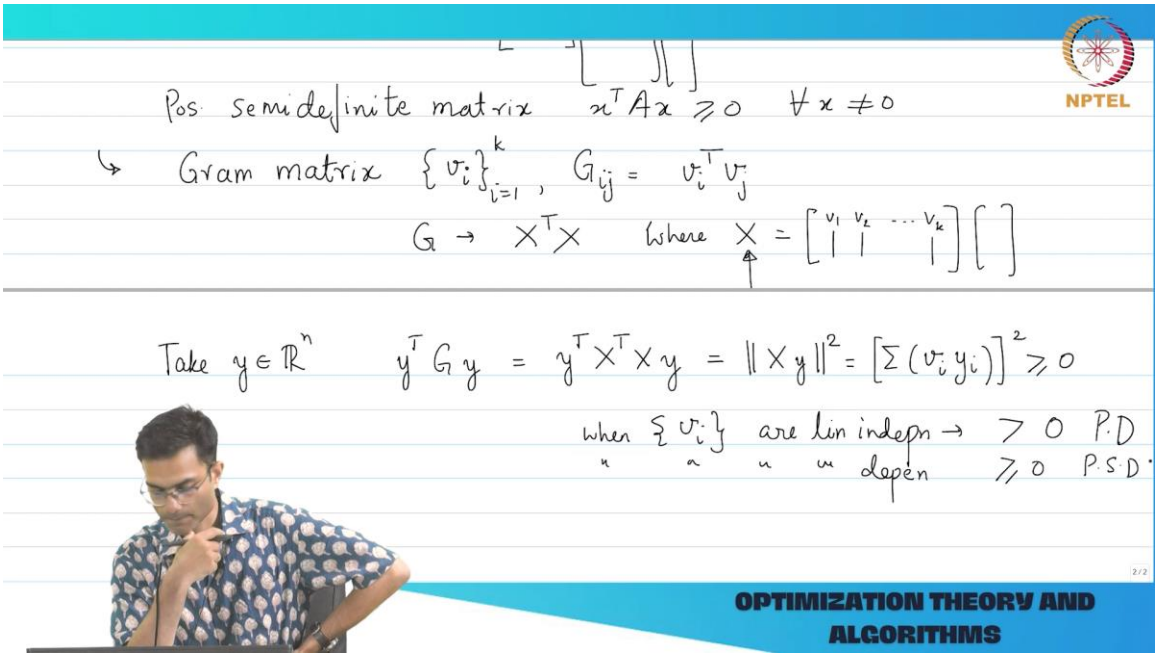


So, does anyone see a pattern over here? What does this look like? Norm of norm or norm squared? Norm. Norm squared of what?

$$y^T G y = y^T X^T X y = \|Xy\|^2 = [\sum(v_i y_i)]^2 \geq 0$$

Now Xy can I simplify X into y what is it going to be? y is a vector everyone agrees right. So, this is the column picture of matrix multiplied by a vector.

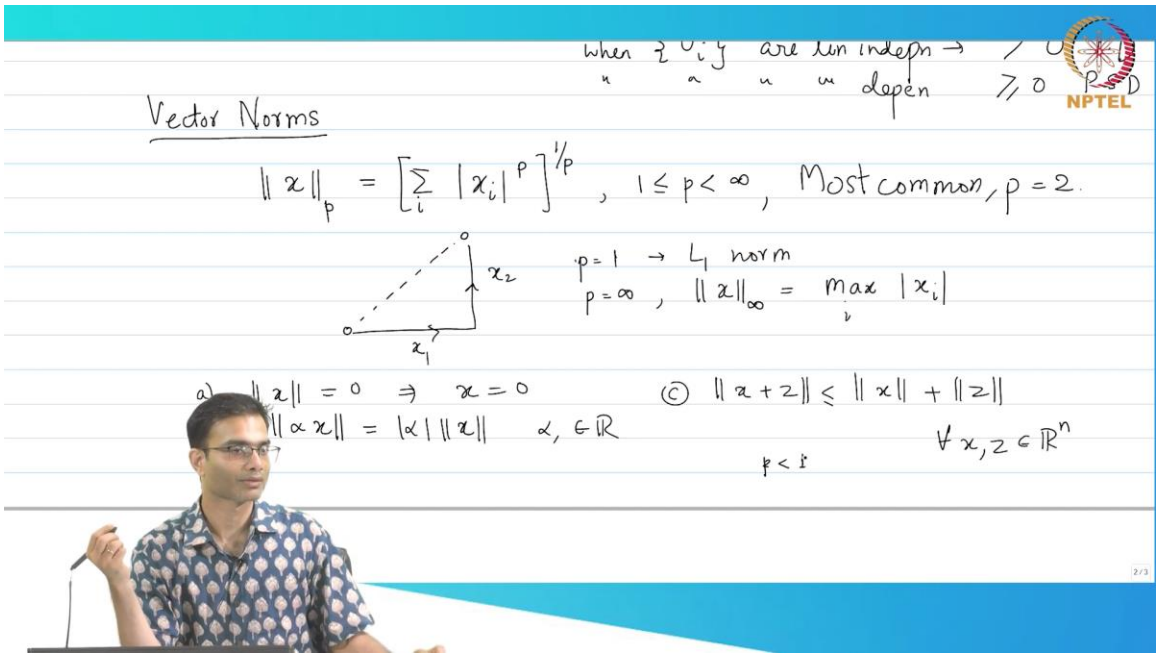
So, can I simplify this further? Can I open up this norm squared what would it look like? Visualize X in this form over here and this is getting multiplied by a vector y. What would this summation open I mean what would this norm squared open up to? $y_1 v$ right exactly I am going to get something like this I am going to get v_i multiplied by y_i right and this is going to be squared is it the whole thing squared where have I put the summation the squared correctly nowhere does it go outside right so it is actually going to go like this. Now, what can we say about this? It is greater than equal to 0 ok; let us we can even make it little bit more precise ok. So, when will it be greater than greater than 0 and when will it be greater than equal to 0? So, this is related to the linear independence of vectors, if I give you a set of vectors that are linearly independent then can their linear combination sum to 0? No. No, unless the y_i's are 0 right, but I am not taking y_i equal to 0 anyways trivial.



So, I am not going to consider that. So, this is so, when v_i are linearly independent. What do I get? I will get greater than 0. So, positive definite right and when these are linearly dependent then I can come up with some set of coefficients y_i such that this linear

combination sums to 0 right. So, this will become greater than equal to 0 positive semi, any questions on this? So, this is a very simple kind of practice exercise for us to get familiar with positive definite, positive semi definite, how do I multiply vectors and matrices and so on ok.

Now, we are going to define the next thing which are vector norms. In fact, we already used the concept of a vector norm over here right, but I am going to give you a very general definition of a vector norm ok. The first sign that we are talking about norms is the fact that I did not put an absolute value sign up because there are two double bars over there. So, the most general definition is what is called a p norm ok. So, the



norm how do I define it?

$$\|x\|_p = \left[\sum_i |x_i|^p\right]^{1/p}, 1 \le p \le \infty,$$

There is one more important restriction what is that? Can p take any values? Right. So, this is 1 greater than equal to p greater than infinity right. So, when this is satisfied this is defined as the p norm of a of a vector ok. Now, what is the most commonly encountered norm? $p = 2$ right. People particularly in engineering disciplines they love $p = 2$ for a very simple reason what is that? Because it relates to energy right, energy has squared quantities and this is going to give me sum of x_i squared and then the square root of that.

So, it gives me energy what else does it give me in I mean if I were to ask you this when you were in class 12 what would you say? Distance right, Euclidean distance not just any distance Euclidean distance. But then as we found out that Euclidean distance is not the only type of distance there are other distances. For example, if you were, well if you were a crow, you would be interested in Euclidean distance right distance from one tree to another tree you do not care which way the roads are. But if you were ah doing an internship in South Bombay where there are all these skyscrapers and you want to go from your office to Starbucks. not going to go the way the crow flies you are going to go either this way or this way right.

So, that is another type of distance what do you think that corresponds to? $p = 1$ right. So, for example, if I am here and this is another point then this is how I go right. So, from here to here. So, this could be this is x_1 this is x_2, you do not go like this unless you are a crow right. So, this is $p = 1$ also called the L_1 norm, other people call it the Manhattan norm ok.

$p = \infty$ is also an interesting norm it comes up often ok and it is simply the maximum value of x_i overall i ok. So, these are useful things. So, in this course for the most part we are going to use $p = 2$, towards the end when we go to constraint optimization, we will briefly use $p = 1$, ok. And sometimes in certain proofs you might see infinity norm ok, nothing else other than that. Now, the choice of norm is often determined by what makes my math easier.

Ok. That is many times the choice of norm is dictated by what is going to simplify the derivation right. So, $p = 2$ is most convenient ok. So, there are certain requirements of a norm which we will just note down and they are very very basic ok, which is what in fact prevents p from being less than 1. So, I am going to just note down those properties. They are quite intuitive, the first property is

$$\|x\| = 0 \rightarrow x = 0$$

So, all of this is very intuitive if you think of norm as some type of distance measure ok. Then I have if I give you a scalar α multiplied by a vector x,

$$\|\alpha x\| = |\alpha|\|x\| \;\; \alpha \in R$$

it simply means that α belongs to real and the third property is actually the make-or-break property of a norm which is the triangle inequality, right. So, if I say

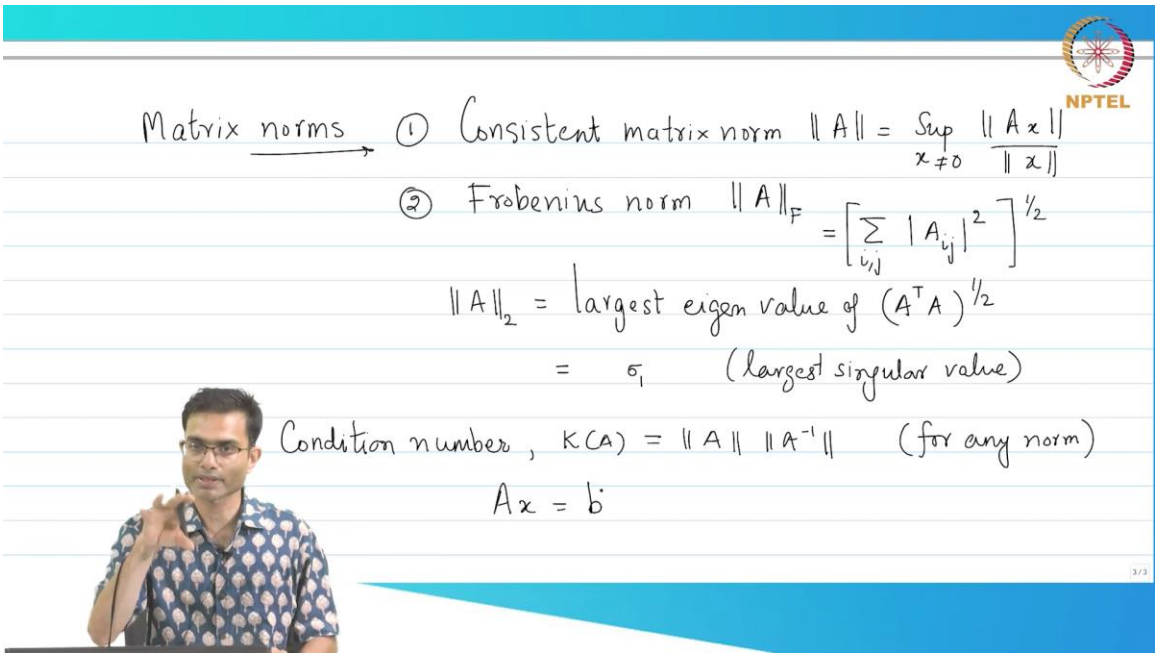
$$\|x + z\| \leq \|x\| + \|z\|$$

So, in fact, you will find out that if I choose $p < 1$ this property gets violated. So, that is why it is not a norm if $p < 1$, ok.

However, if you look into the research literature you will find people do use $p < 1$ and it is called a quasi-norm and it gives some very interesting mathematical results ok. So, even though it is not a norm people use it, but you they use it very carefully and you get some very surprising results. So, just letting you know that even though I have this is the standard definition, once you know what you are doing you can always break the definition and you know get surprising results. So, this is yes question.

Yes, this is the definition of the norm. And, this definition of a norm gets used across if you look at the whole field of what is called differential geometry where relativity etc. is also spoken about these become very important because there people define totally new kinds of norms not just p norm there are other kinds of norms ok. So, this was about this is about vector norms right. Now, often we will also come across norm of a matrix right. Now, what is the norm of a matrix? So, it is there are in fact two different definitions of the norm of a matrix.

So, let us look at that. So, matrix norms. The first type of norm is a norm which is derived from the definition of the vector norm ok. So, it is called a consistent matrix norm. consistent simply means consistent with the definition of a vector norm and it is defined like this.



So,

$$\|A\| = Sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

So, this is not a very intuitive looking definition, but what is it saying? It is saying that take any x you want stick it into A and calculate this ratio norm Ax by norm x and of this get the supremum. So, in some sense the lowest upper bound of this and that is defined as the matrix norm. Notice that to calculate this I just have to rely on the definition of the vector norm because A into x is a what is it? Is it a scalar or a vector or a matrix? It is a vector. It is a vector. So, I just need to use a definition of the vector norm.

So, if I chose $p = 3$ as my vector norm, I can derive the matrix norm for $p = 3$. So, a straight forward definition ok. So, this is the consistent norm. Now, the other type of norm is, I will give you an example of a norm which is not a consistent matrix norm, but is used all over the place. It is called the Frobenius norm of a matrix and many of you may have heard of this.

$$\|A\|_F = \left[\sum_{i,j} |A_{ij}|^2\right]^{1/2}$$

So, Frobenius norm is denoted with a F, ok. This is actually a very intuitive definition which you would have expected this to be the actual definition of a matrix norm, but it is not. It is simply going to be it looks like the 2 norm of a vector right. So, it is it is it is exactly that ok, but we are putting the the symbol F over there to make sure that we are not talking about a consistent matrix norm ok. On the other hand, if I talk of the consistent matrix norm of with $p = 2$, this is something that we will use throughout the course. This has a very special interpretation; this is going is related to the singular values of a matrix ok.

So, this is the largest eigenvalue of $(A^T A)^{1/2}$, ok, which is also going to be equal to σ_1. σ_1 denotes what? The largest singular value of a matrix. Now, related very closely once I have the definition of a norm there is something which we which gets used a lot in linear algebra and also in optimization which is something called the condition number ok. We will talk about condition number as we go into the course right now it might seem a little abstract, but this is number denoted by κ and this κ is defined as

$$\kappa(A) = \|A\| \, \|A^{-1}\|$$

for any norm and what this condition number right now just looks like strange definition, but it is telling you how much for example, let me just give you an intuitive idea.

Supposing I am solving the system of equations

$$Ax = B$$

Now, this is thought of this as a real-life engineering problem where B, for example, could be coming from some measurements of some system, A is your system matrix, x is your set of variables right. In the real world I mean ok, let us start with the ideal world. In the ideal world, your measurements are exact that means there is no error in B. You know that once you go into the real-world B will have some noise, it can be noise in the measurement errors in the measurement.

You might be interested in knowing that if there is an error in B, how much error is there is going to be in x. That is important because supposing 1 percent error in B gives me 100 percent error in x. What is my conclusion? Common sense conclusion. A is very sensitive therefore I need not bother doing this experiment right if I if such a small error in B is going to give me 100 percent error in x that means I may as well just flip a coin

and determine I mean get x right. So, how much is this error amplification happening is that is captured by this condition number.

So, it is a very important quantity in numerical analysis, optimization and all because if your optimization system is has a very very high condition number that means it is amplifying error so much. you know. So, you need to know this and it will be a very important part of the proofs of our optimization algorithms ok. So, we will talk about this in detail as we go in this course.