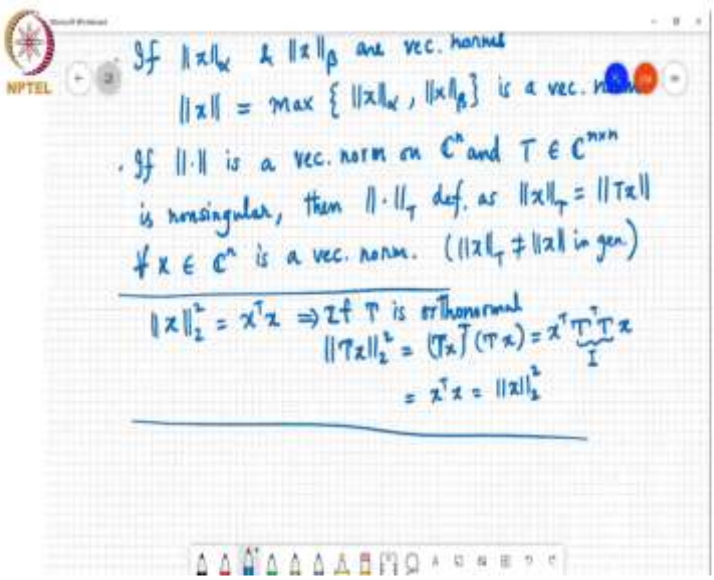
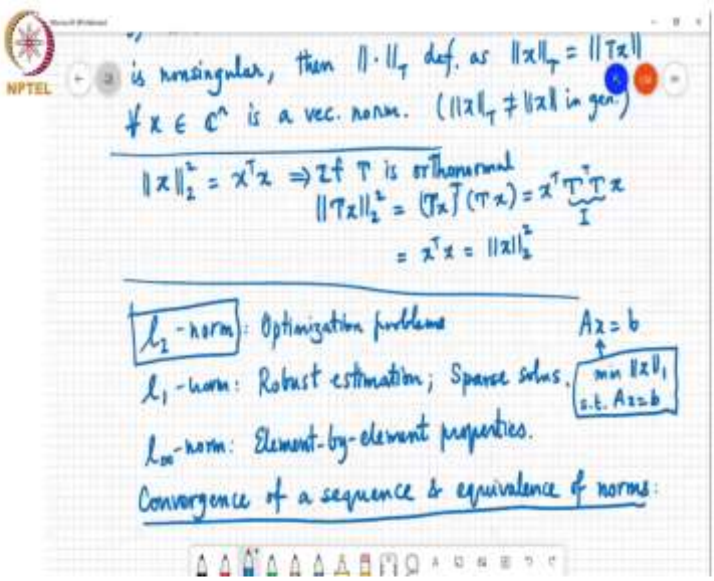
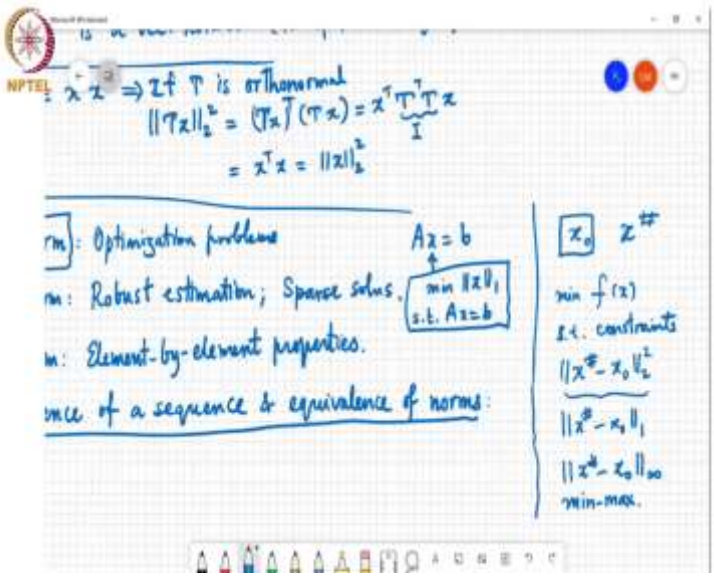
Matrix Theory
Professor Chandra R. Murthy
Department of Electrical Communication and Engineering
Indian Institute of Technology, Bangalore
Applications and equivalence of vector norms

(Refer Slide Time: 00:14)







So, typically we use just discussing about, where we use different types of norms. The reason why the l2 norm is the most popular norm is because it is it can be written like this, the l2 norm squared can be written as x transpose x. So, it has lots of good properties. So, it is typical to use the l2 norm and optimization problems.

Even if you want to use other norms, it is not uncommon to try to reduce it to an l2 norm, and then solve a sequence of problems where you are working with the l2 norm, and then hope that you will be able to solve the problem involving other norms. l1 norm is typically used when you want to find what are called robust estimators.

And it is also used very heavily in compressed sensing, which I teach them the next term and it is promotes past solutions. Sparse solutions as solutions for these vectors, where lots of entries of the vector are equal to 0, and the many applications where you want to solve for example, an equation like Ax equals b.

But this suppose this has many solutions, you want to find a solution which has the maximum number of zeros in x, and for such things solving for minimizing l1 such that Ax equals b this optimization problem will lead to sparse solutions for x, l infinity norm is useful when you want to, you care about element by element convergence or properties.

However, as I mentioned, l2 norm is by far the one that is most amenable to optimization. So, the norm in which the norm that is most natural to a given problem may not be the most mathematically convenient or tractable one and so, if you use a different norm to solve the problem, we want to, ideally we want to know how it is related to the original how to solve it.

And so, for example, if you are considering a sequence of vectors, and you want to look at this is a sequence of vectors output by a particular algorithm, and if you monitor say the l2 norm of these vectors and you find that the l2 norm is converging, or you take the difference between consecutive outcomes of this iterative algorithm, and you find that the l2 norm of that difference is converging, does it mean that the vector itself converges or not?

So, to answer these kinds of questions, there is a very strong property that norms satisfying which is that essentially, if a sequence of vectors converges according to a given norms, it in fact, converges to the same point with respect to any other norm that you wish to use. And so, I will just discuss that aspect a little bit.

Student: Sir, what is meant by robust estimation? Is it like, (())(04:25) perform well in the presence of noise or…

Professor: Yes, certainly you want it to perform well in the presence of noise, but other norms also will give you good properties in recovery in the presence of noise. However, what happens is that, if you think about it, suppose, just go off to the side a bit here, this is a side note. So, suppose you have a certain point x naught and you have an algorithm, where you hope that the algorithm will return this optimal point x naught.

But it returns a different point call it say x hash. And now there is a distance between these two and your algorithm is returning x hash, because you have sort of said I want to minimize, say something like x, some function of x, subject to some constraints and effectively, if you are looking at say the l2 norm, what this is doing is it actually taking the difference between all entries have x hash and the corresponding entries of x naught, it is squaring them and adding them up and then finally taking a square root.

So, if I consider the square of this Euclidean norm, you can see that if there are one particular pair of entries in x hash and x naught, where this difference is very large, because you are squaring it, the distance or the Euclidean norm will end up becoming a very big number. And so this really penalizes the most mismatched entry the most.

And the penalizes the least mismatched entries less, but if you take x hash minus x naught and one norm, then what this is doing is just looking at the magnitude of the error between x hash and x naught and so, this essentially penalizes all the errors more or less equally. So, that is what is called robust estimation.

So, you are not giving undue importance to incurring a large error in some components and incurring a small error in other components, all errors are equivalent to you. And that is what is referred to as robust estimation.

Student: (())(07:15)

Professor: So, there are many parameters.

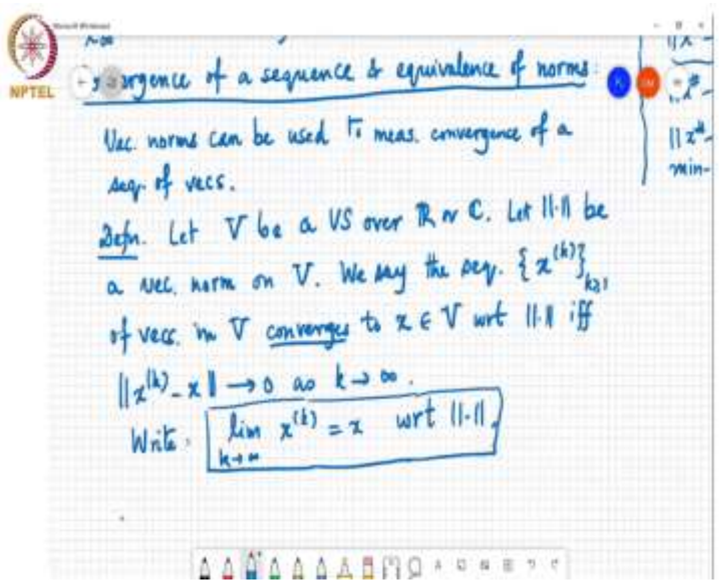
Student: But how does it translate in l infinity norm, there we are only optimizing with respect to only one this maximum element or…

Professor: You are looking in l infinity norm, you are looking at the largest entry of x hash minus x naught. So, this is typically used in what are called min max type of problems, where what you want to minimize is the maximum deviation across all entries between x hash and x naught. And when that is important to you, then you would use the l infinity norm, so yeah.

Student: Thank you sir.

Professor: So, let me define what I mean by convergence of a sequence, a sequence of vectors. So, the main point, I will first write down the punch line here, and then I will discuss further. So, the punch line is that vector norms can be used to measure convergence of a sequence of vectors.

(Refer Slide Time: 08:33)



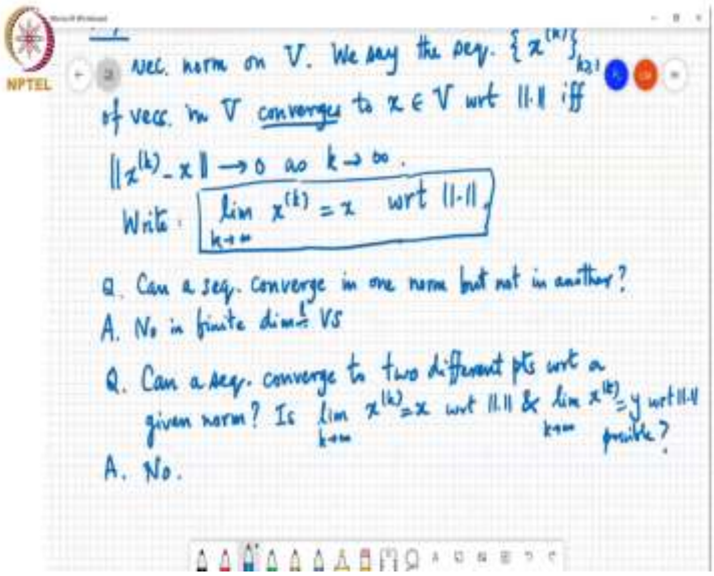
So, let me define convergence first. So, let V be a vector space over R or C, and let be a vector on V. So, we say the sequence xk. So, this is a common notation for denoting a sequence you write curly braces xk. And sometimes you write k greater than or equal to 1. If you want to say k goes from 1, 2, 3 up to infinity, this of vectors nV converges to x this is also in V with respect to this norm defined like this, if and only if xk minus x goes to 0 as k goes to infinity.

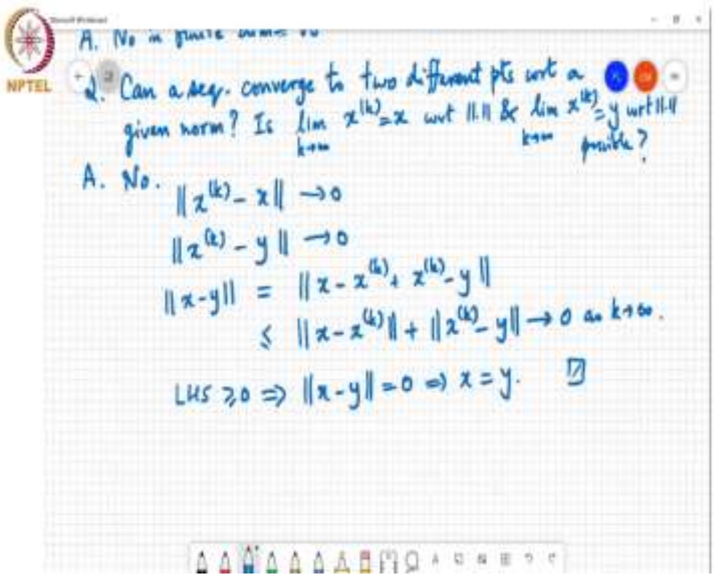
And we will write this as, limit k tends to infinity, xk is equal to x. Now, again I have to write with respect to this norm, like this. So, this is the definition and two aspects sort of immediately come out. One is that, it seems that in order to define convergence of a sequence of vectors, I need to tell you with respect to which norm I am asking for this convergence. The second is that, if I change the norm, it is possible that this xk will converge to a different point x dash, because it is dependent on which norm I am specifying here.

So, the two related questions are one is, is it possible that this given sequence xk converges with respect to one notion of norm, but not in another? And the second question is that, can a sequence converge to two different points with respect to a given norm? So, it turns out that the answer to the first question to both questions is no in finite dimensional space, but it is possible that a sequence converges with respect to one norm, but not in another infinite dimensional vector space.

There is an example in Holland Johnson, which shows that shows that sequence can converge to two different points with respect to two different norms. But we would not discuss that here because the focus of this course is on finite dimensional vector spaces.

(Refer Slide Time: 12:35)





So, the first question is, can a sequence converge in one norm but not in another? And the answer is no finite dimensional vector space, and we will see why this is true in a minute, but before that, let me write the other question which is actually easier to show. So, can a sequence converts to two different points with respect to a given norm?

So, the answer is no. So, that is this limit k tends to infinity xk equal to x and limit k tends to infinity xk equal to y with respect to the same norm possible, and the answer is no. And that is, that is very easy to see. And I guess some of you may have already been able to figure out why, and the reason follows from triangle inequality.

So, if so, what we are told is that xk minus x, this goes to 0 as k goes to infinity. And similarly, xk minus y also goes to 0 as k goes to infinity. So, what that means is that if I take the norm of x minus y, so that is equal to the norm of x minus xk plus xk minus y, which is less than or equal to the norm of this is triangle inequality, xk plus xk minus y, which both of these terms are going to 0 as k goes to infinity, so, this itself goes to 0 as k goes to infinity.

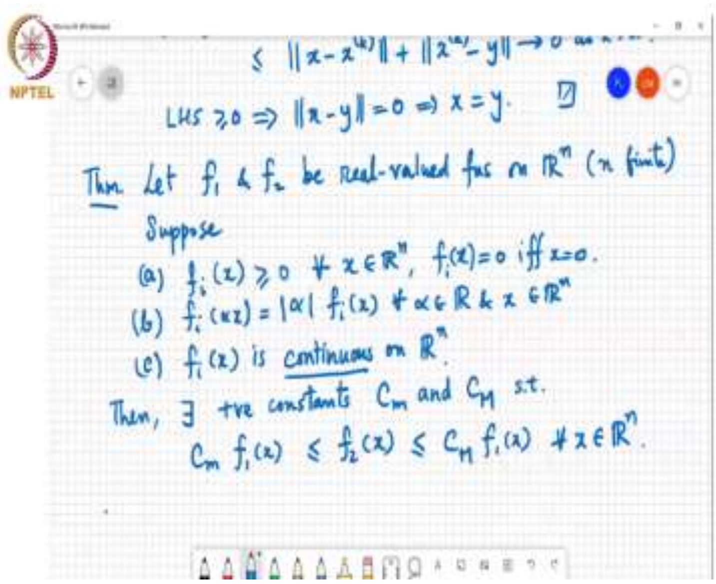
So, but the left-hand side is greater than or equal to 0. And which implies that the norm of x minus y and so this is non negative and but this is a norm. So, if this becomes equal to 0, it implies that x equals 1. So, it has to converge to the same point.

Student: Sir.

Professor: Yeah?

Student: Is there a particular sequence can converge to different points, if we did… the different norms?

(Refer Slide Time: 16:45)



Professor: So, that is what I want to show now. To, show that a sequence cannot converge to different limits for different norms, is one other theorem that we will need. And so, this theorem is actually a theorem from real analysis will outline the proof, but there is one step that we will need from real analysis, which I would not go into here.

So, f1 and f2 be real valued functions on R to the n, for n finite. When suppose, three properties hold, first is that fi so, fi of x is greater than or equal to 0. So, these will be replacing f1 and f2 with norms later on. So, this is true for every x in R to the n and fi of x equals 0 if and only if x equals 0. And property b, is that fi of alpha x is equal to mod alpha times fi of x for every alpha belonging to R and x belonging to R to the n.

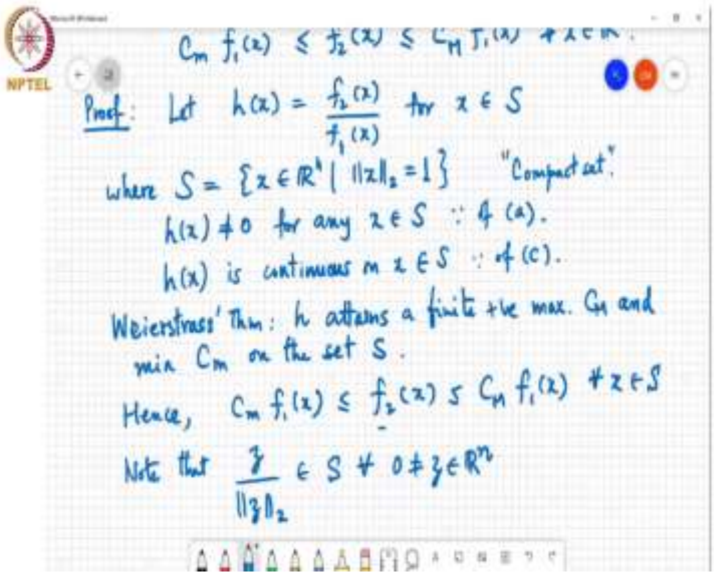
And property c is, fi of x is continuous on R to the n. So, notice that I am not using I do not require the triangle inequality, which is part of the definition of a norm, I just need these three properties. But instead of or not instead of, but I do not need the triangle inequality. But I do need this continuity property, on R to the n, I will just leave it like this, I think all of you have some idea of what it means for a function to be continuous.

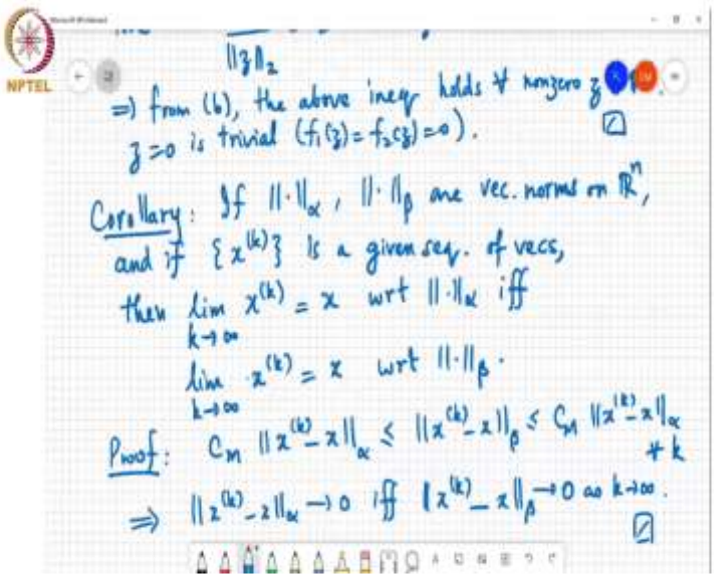
I would not go into the definition of continuity and so on here. In fact, this continuity is really used only in when we use another famous theorem from real analysis called Weierstrass theorem, which is used in the proof but other than that, you know, let us not get into the notion of contiguity in this right now.

So, will take this on faith that we know what is continuous and what it means for a function to be continuous. So, when this is true, then for there exists positive constants which will call C small m and C capital M, such that C small m times f1 of x is less than or equal to f 2 of x is less than or equal to C capital M times f1 of x for every x in R to the n.

That means, now translating this into norms, what this is saying is that, if you take a different norm, the norm of x with respect to this, the second norm of x is sandwiched between some constant times the first norm and some other constant times that same first norm. So, that is this result.

(Refer Slide Time: 21:02)





So, the proof goes like this. So, let h of x be defined as f2 of x over f1 of x, for x in some set S, where I will define this set as S is the set of x in R to the n, such that x2 so I am using the Euclidean norm here, but you can actually use any other norm here, it does not matter. So, the reason I mean the only thing you need is that the set S must be a compact set and it does not include the 0 vector.

So, that is again this is another notion from real analysis that is beyond the scope of this course, but, for your reference, you can note that the compact set. Then what we have is that h of x is certainly not 0 for any x belonging to S, because there is the 0 vector is not here and a fi of x is positive strictly positive for any x naught equal to 0.

And so, both these numbers as f1 f2 of x and f1 of x are both strictly positive numbers and so, their ratio is also some strictly positive number and so, h of x is not 0 for any x in S, and also h of x is continuous on x belonging to S because of this property is c, the ratio of two continuous functions is also continuous.

So, now what Weierstrass theorem says, it says that h a function h, which has these two properties it attains finite positive max, maximum C capital M and minimum C small m on the set S. So, that implies that we have C small m times f1 of x is less than or equal to f2 of x. So, h of x is between is bounded between C small m and C capital M and h of x is just f2 over f1. I am taking f1 to the other side.

And then I have C capital M times f1 of x for every x in S. But we have that if I take z over norm z. This it always belongs to S for every nonzero z in R to the n. So, then what I can do is, if I want to show that this holds for every x, I just replace x with x and R to the R, if I want to show that this holds for every z in R to the n, I just replace x with z over norm of z for any nonzero z then by property B, which is homogeneity property.

This non z can come out of this and then it will cancel throughout because there will be a 1 over norm z, here 1 over norm z, here 1 over norm z, here it comes out throughout. So, from being the above inequality holds for every nonzero z belonging to R to the n, but then if z equal to 0 the case is trivial because this is 0 and this is 0 this is also 0.

So, it is already true. So, that concludes this proof. So, now the consequence of this is that, if you have two different norms are vector norms on Rn and if xk is given sequence of vectors, then limit k tends to infinity xk is equal to x with respect to alpha if and only if limit k tends to infinity xk is equal to x with respect to beta.

So, it does not matter which norm you consider if it can be converges with respect to a given norm then it converges to with respect to any other norm and in fact, it converges to the same point. So, the proof is one line. So, we have by the previous theorem that Cm times xk minus x alpha is less than or equal to xk minus x beta.

So, there exists exist constant Cm and C capital M such that this holds xk minus x alpha and this is true for every k that implies that the xk minus x alpha can only go to 0 if and only if see this quantity is sandwiched. This quantity is sandwiched between these two quantities. So, if you want this to go to 0, then this side will also go to 0. And that is only possible if this guy is also going to 0, if this is going to some nonzero quantity, then you cannot have it being sandwiched between these two if this is going to 0.

So, this is true if and only if xk minus x beta goes to 0 as k goes to infinity. So, basically this implies that in the finite dimensional vector space, all norms are equivalent in the sense that whenever xk converges to x with respect to 1 norm, then it converges to the same x with respect to any other norm. So, I think we are out of time for this class. So, we will stop here and continue the next time.