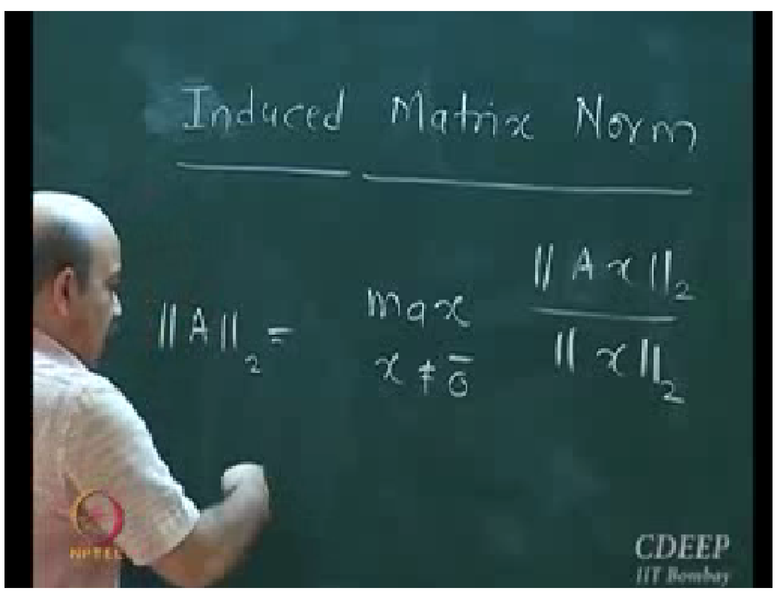
**Advanced Numerical Analysis**
**Prof. Sachin Patwardhan**
**Department of Chemical Engineering**
**Indian Institute of Technology - Bombay**

**Lecture – 30**
**Iterative Methods for Solving Linear Algebraic Equations: Convergence Analysis using Matrix Norms (Contd.)**

So we have been discussing iterative methods for solving linear algebraic equations and then after discussing algorithms, we started looking at convergence of these methods. We also derived necessary and sufficient condition for convergence. Necessary and sufficient condition for convergence was spectral radiance of matrix as inverse t should be inside unit circle and we said that it definitely gives us lot of insight into what happens.

But we need something even simpler to calculate and we have started looking at matrix norms to come up with something that is easier to compute in terms of assessing whether A matrix which are used to formulate the iterations, we can apply some quick tests to come up with convergence analysis.

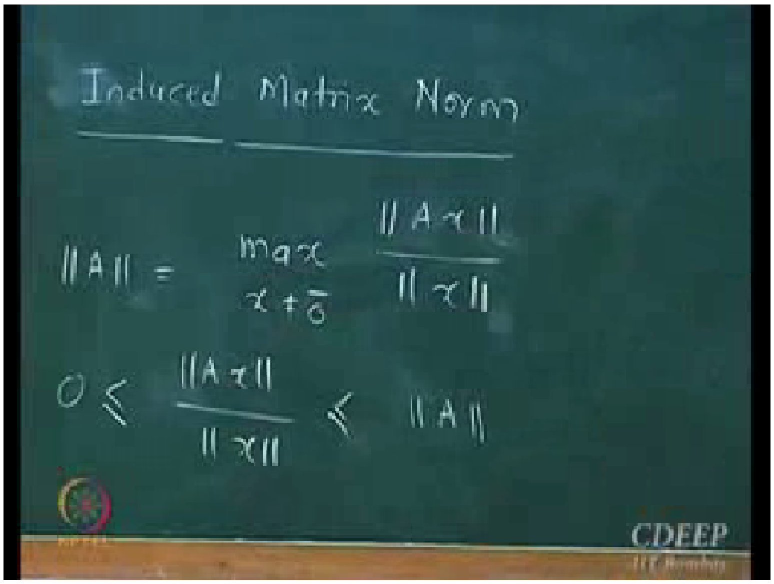
**(Refer Slide Time: 01:26)**




So Induced Matrix Norm was defined as normal of A. Let us talk about norm in which the norm induced by identical normal on both range and domain space. So this is max x != 0, norm Ax/norm x and the way I try to explain this was, this is the something like amplification power of a matrix. Other way of writing the same thing, so it is 2 normal here and 2 norm. If I want 1 normal, it will be A1 normal, 1 norm and 1 norm and so on.

This is called induced normal because this is induced by the norm defined on the range and the domain space, okay, that is why it is called induced norm.

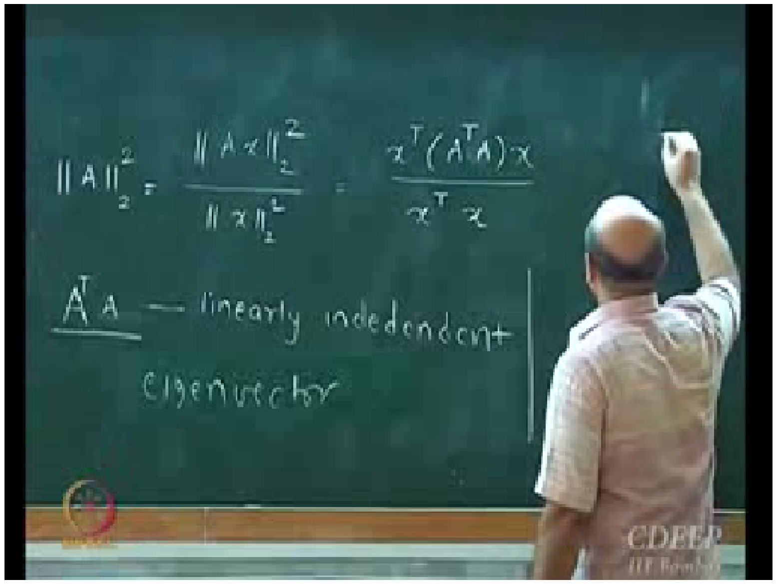
**(Refer Slide Time: 02:27)**




Other way of writing this is, this is generic thing, so I need not write 2 but we were discussing 2 norm at the end of the last class. So this inequality can also be written like this. Norm Ax/norm x is <= … So norm A is in some sense bound, upper bound on this ratio. Of course this is always > 0 or x is != 0, this is always > 0, okay or it could be always >= 0.

Suppose A is a matrix which is rank deficient and Ax is in a null space then Ax will be 0 but denominator is not 0. So this can be >= 0 and this is an upper bound, norm is an upper bound, maximum of this ratio or maximum of the gain, maximum of the amplifying power of the matrix whatever way you want to give it, okay.

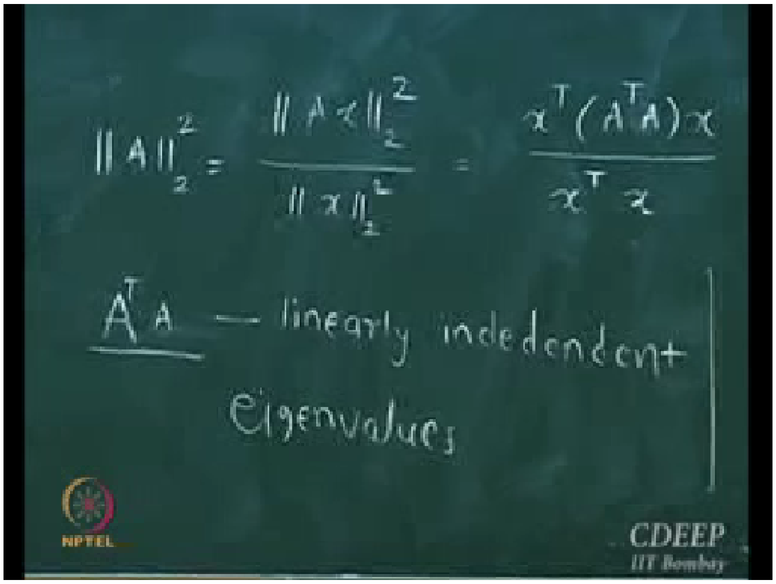
**(Refer Slide Time: 03:34)**




Now in particular, we started discussing 2 norm. So in 2 norm, I just started by putting Ax2 square/x2 square which I wrote as x transpose A transpose Ax/x transpose x. I can do all this divisions because this is a scalar, x transpose A transpose A*x is a scalar, x transpose x is a scalar, so this is ratio of 2 scalars, okay. Now this particular one, this particular norm where even though as I said that it is not very convenient for computation but gives you very nice interpretation.

So this one we showed that, we made additional assumption that matrix A transpose A, this is a symmetric matrix, positive definite matrix. We made one more assumption that A transpose A has linearly independent eigenvalues.

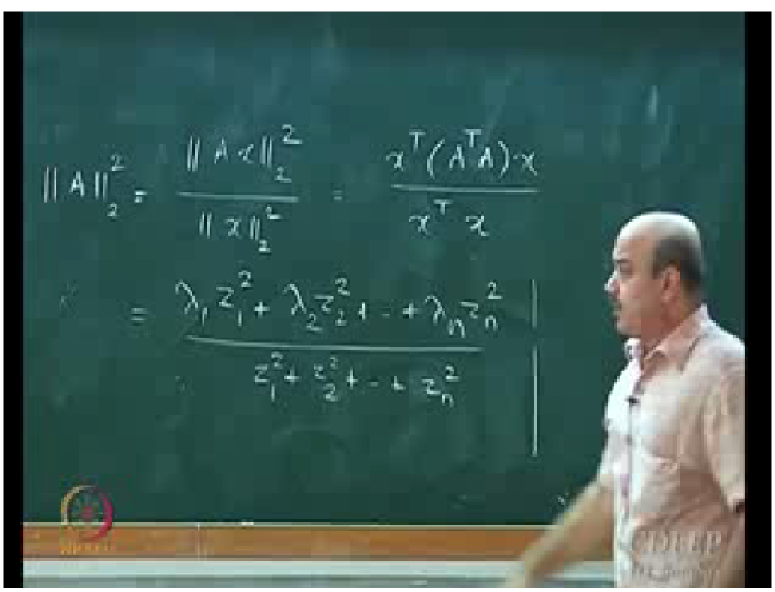
**(Refer Slide Time: 04:54)**




So we wrote this matrix A transpose A as psi lambda psi inverse. Actually what I argued was that since this is a symmetric positive A matrix, the eigenvectors are orthogonal, which is same as appearing on main diagonal and this matrix psi is nothing but matrix formed by keeping eigenvectors of A transpose A next to each other, okay. So I am just keeping eigenvectors of A next to each other.

This is a n*n matrix, not only invertible, it is an orthonormal matrix which means psi transpose psi=psi psi transpose=I, special property of this matrix, okay. And I transform this, I transform this ratio using this, I transformed this ratio to something very very interesting.

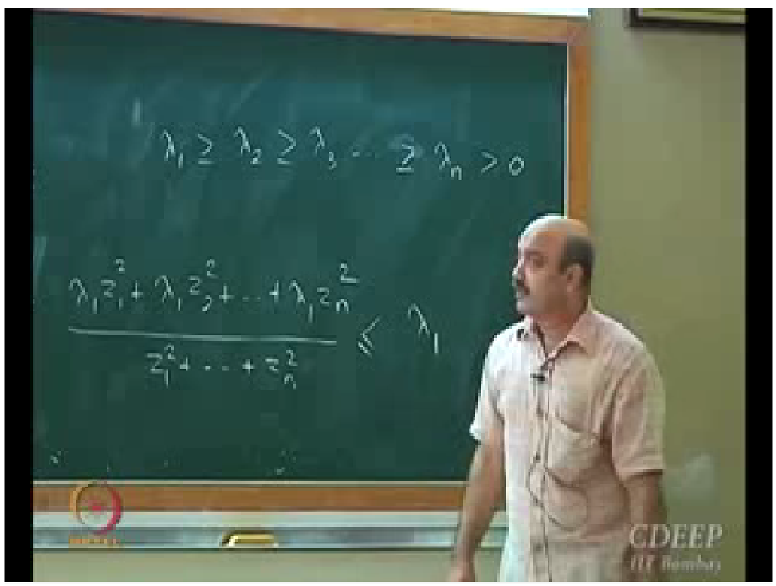
**(Refer Slide Time: 06:42)**



So I wrote this x transpose A transpose Ax as x transpose psi lambda psi transpose x… psi psi transpose x and defining y to be psi transpose x, we wrote this ratio defining this new vector, we defined y or we defined z. We wrote this ratio to be lambda1 z1 square+lambda2z2 square+lambda nzn square… So you can see that this is ratio of 2 positive quantities, is ratio of 2 positive quantities because eigenvalues of A transpose A are always positive, okay. Eigenvalues of A transpose A are always positive.

Because A transpose A is a positive defined matrix, its eigenvalues are positive z1 square z2 square, whatever is z, z1 z2 square is always positive. So this is ratio of 2 positive quantities, okay. Now if we order the eigenvalues as, all the eigenvalues are real positive for positive definite matrix.

**(Refer Slide Time: 08:16)**



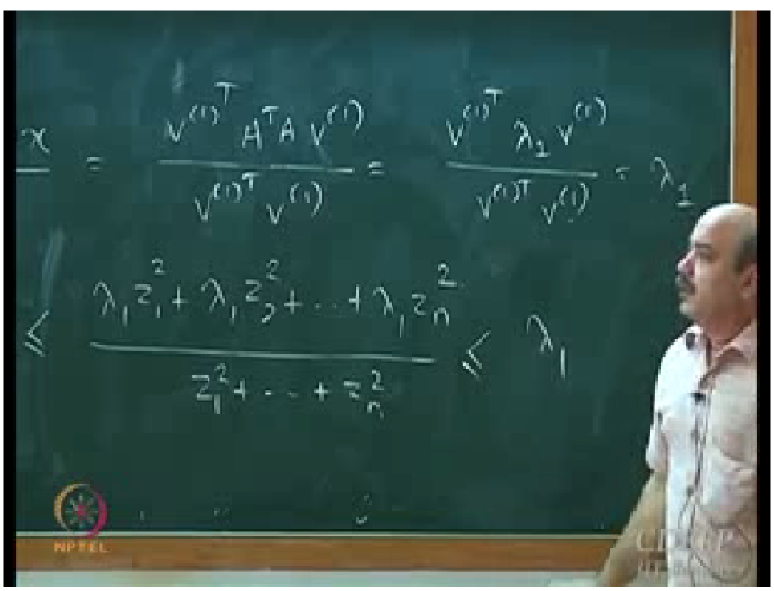

And then if I say that lambda1 is greater or equal to lambda2>0, okay. If this holds, well there is an implicit assumption when we started all this analysis that A is full rank because we are assuming that when you are solving this we are talking about a full rank matrix. So that is why you get minimum eigenvalue > 0; otherwise, you may have possibilities some eigenvalues will be = 0. Now if I number my eigenvalue such that lambda1 corresponds to of the largest.

Numbering, it depends upon me what I call 1 and what I call 2 matrix as eigenvalues. They do not come numbered. We number them. So I am numbering lambda1 to be the largest, okay. So it is very easy to see that this ratio will always be less or equal to lambda1z1 square+lambda1z2 square+lambda1zn square/z1 square up to zn square. Is everyone with me on this. I am just replacing lambda2 by lambda1, lambda3 by lambda1, lambda 4 by lambda1, okay. Lambda1 is the largest magnitude eigenvalue of A transpose A, okay.

So I am allowed to do this, denominator is same, numerator, I am replacing by the larger value for every term by term, okay. Very easy to see that you can take lambda common, okay. So this ratio is independent of, this is lambda1, okay. So this ratio can never exceed lambda1. This ratio can never exceed lambda1, very very nice property, okay. What did we start with? We started with this, we wanted to find out maximum of this ratio.

What is the maximum of this ratio? Lambda1, okay. This ratio can never exceed eigenvalue of maximum magnitude eigenvalue of A transpose A, okay. So what is 2 norm of matrix A? Lambda1. This is the upper bound and you can show that this upper bound is attained when eigenvector is aligned in a particular direction, okay. What happens when x=v1? When will this be an equality? For which direction?

**(Refer Slide Time: 11:22)**

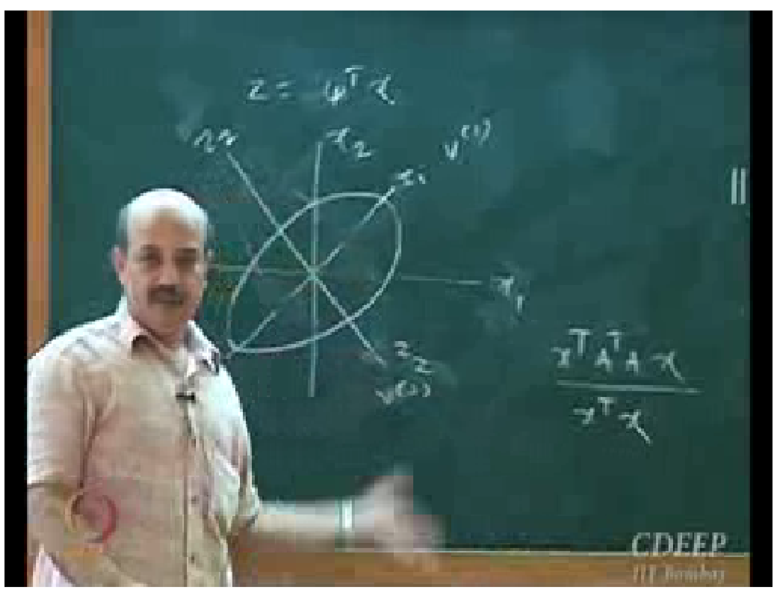



V1 transpose , A transpose AV1/V1 transpose V1, okay but V1 can be chosen to be orthonormal. So V1 transpose V1 will be unity, okay. So this will be V1 transpose lambda1V1/V1 transpose … even if it is not chosen orthogonal, even then this ratio will be equal to lambda1, okay. So when x is aligned along the direction which corresponds to the eigenvector associated with lambda1, eigenvector of A transpose A, not eigenvector of A.

Eigenvector is defined for a square matrix, okay. Eigenvector is defined for a square matrix. A transpose A is a square matrix. A in general need not be a square matrix but A transpose A is always a square matrix. So of course right now we are dealing with square matrices, we are dealing with square matrices, so there is no question of non-square matrices. We are dealing with solving Ax=b where A is square, in fact we also have a problem where is A is invertible, otherwise.

So this ratio becomes equality where x is aligned along eigenvector of A transpose A. First eigenvector, first in the sense that corresponds to the maximum magnitude eigenvalue. So this is equality, okay but this ratio for any other x is not equality, okay. For any other x, this ratio will not be equality, this ratio will be smaller, okay. That is why the maximum amplification power of a matrix using 2 norm is given by this, okay. So I think I have this picture somewhere drawn for a 2-dimensional case.

**(Refer Slide Time: 13:41)**



Actually when you have defined this z=psi transpose x, you have actually defined a transformation which is rotation, okay. So suppose this is your x, y, then this is your z. So this is x1 x2. Let us say this is x1 x2 and this will be your z1 z2 coordinate space. So this is an invertible transformation. You can go from one to other, okay. So actually you are just rotating your coordinate axis when you are multiplying, okay and what happens in the rotated coordinate axis?

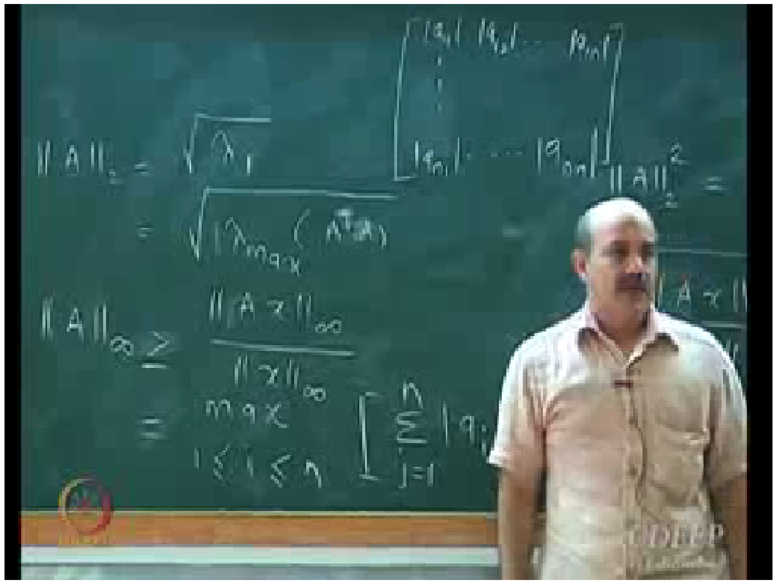
In the rotated coordinate axis, okay, this x transpose, if you draw this x transpose A transpose A into x that is which is same as…, okay. If you draw this inside a region where norm x is < 1, you also had other interpretation of norm, right. You remember we did one more interpretation of norm, max of Ax cap, where x cap is a selective unit circle, okay. If you draw that in a locus of points, then you will see here that this actually corresponds to a ellipse, actually corresponds to an ellipse, okay and these coordinates will be nothing but this will be v1, I have drawn I think wrong.

The ellipse will be like this. The ellipse will be like this. This z1 will be actually aligned along your direction v1, z2 will be aligned along direction v2, okay. It will be an ellipse. This ellipse will be, major axis will be along the eigenvector corresponds to maximum eigenvalue. Minor axis value along the direction which is smallest eigenvalue. All other axis are in between. So it will be an ellipse drawn in 3-dimension.

It will be an ellipse drawn in n-dimension depending upon what kind of matrix we are looking at. I have drawn a picture somewhere in the… you can have a look at it. But again, the problem with this 2 norm is you have to find out eigenvalue of A transpose A. If A is large, it does not help us, okay. So we could actually come up with a criterion which says that convergence will occur if 2 norm, 2 norm is nothing.

But eigenvalue of A transpose A which is strictly < 1 but as I said, it does not really help. So actually this what we have found here is ratio of the squares. It is square of this by square of this. So what is the 2 norm.

**(Refer Slide Time: 17:12)**



So 2 norm, it turns out is that, is square root of lambda1 that is square root of lambda max of A transpose A. That ratio is far as square of, see this ratio was found for A2 square, okay, A2 square=lambda1. So what is A2, square root of lambda1 and this will always be positive because A transpose A will always have positives eigenvalues, okay. So this number will be positive, okay.

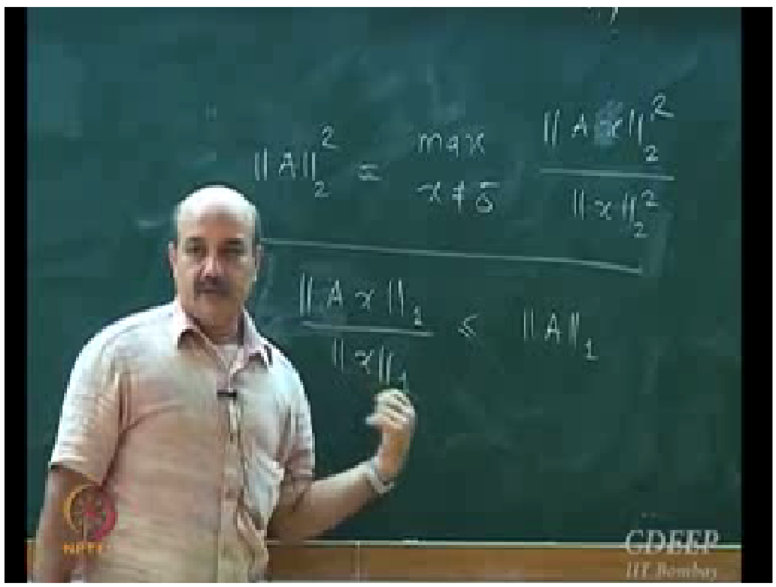
In fact, eigenvalues of A transpose A are called as singular values of A and square root of the maximum magnitude singular value that is what is… What do you expect when A is symmetric? There is a problem in the problem sheet. If A is symmetric, then it will be A transpose=A. So A transpose A will be equal to A square.

What is the relationship of eigenvalue of A and A square? Square, it is very easy to show that if lambda is eigenvalue of A, lambda square is eigenvalue of A square, okay. So if A is symmetric, A transpose A will be… and if A is symmetric positive definite, then it is much easier. A is symmetric positive definite, then it is lambda max of A square, okay. Lambda max of A square is lambda square of A and you can reduce for a symmetric positive definite matrix, just look at its eigenvalue and then you can.

Maximum eigenvalue will directly give you the norm but then this norm is again inconvenient because you have to compute eigenvalue, okay. Now I am going to state 2 other norms without deriving. The derivations to some extent required for these norms are there in the, as a part of your exercises. So if you look at, you should try to work out and then you can see whether you are able to derive that or you end up into some difficulties.

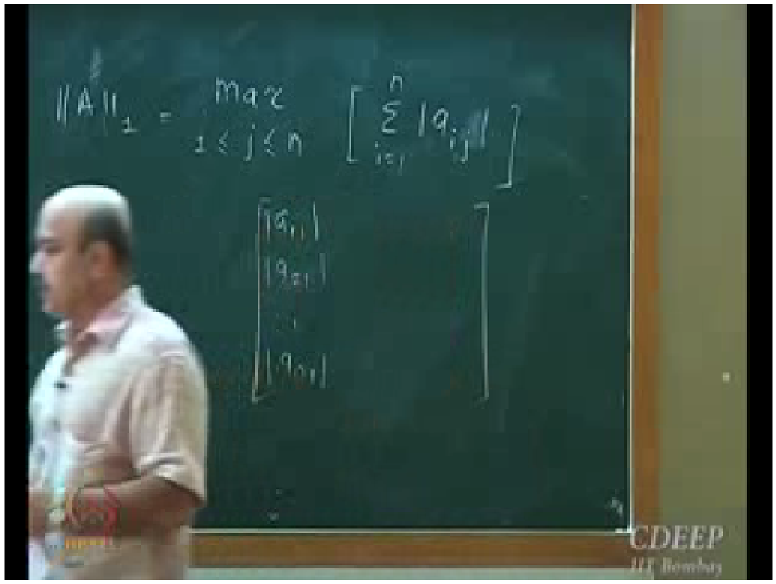
I am just going to write this final statements for other norms, okay. So now even 2 norms, though it has some nice geometric interpretations, it is not quite convenient for me for computing. So I am going to talk about one norm. So one norm is nothing but max over this. One norm is nothing but max over this, okay. Well one small correction. I realise that I made a small mistake here. When I wrote the earlier expression, I will just correct it.

**(Refer Slide Time: 20:31)**




So it is not norm 2 square = this equal to max x != 0. So earlier when I started, I had forgotten this max, max operator is there. Without max operator you cannot proceed and then I have simplified this quantity, okay. So moving on to one norm. So this one norm is induced by one norm on the range space and one norm on the domain space, okay. One norm on the domain space, maximum of this ratio.

**(Refer Slide Time: 21:22)**



Now you can show that one norm. So what is one norm? How do you interpret this? What is this row, what is this summation? Is it a column sum or it is a row sum? It is a column sum. So you take all the elements in one column a11 a21, okay. So one is summation over 1. So I am taking summation of mod of each column, okay and max over that. So I find out… First of all, I take a matrix which is consisting of only absolute values, fully positive numbers, okay.

Then I find the column sums, okay. I find the column sums, max over the column sum is nothing but this ratio, max over this ratio, that you can show, it is not very difficult to show this. So max over column sums, absolute of column sums that is one norm and what you can show is infinite norm. So infinite norm is nothing but max over absolute of row sums, okay. So these norms, you can see here, computationally is much more easy to… See what you have to do when you want to compute one norm or infinite norms? You create a matrix a which is a11, a12…

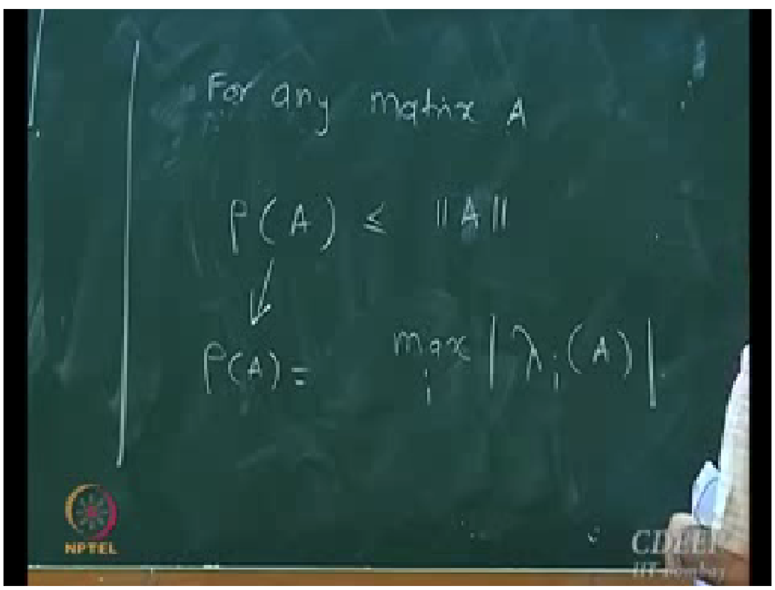
You create this matrix which is absolute value of each number. All of them are positive numbers now, okay. If you take all column sums, find max over it, you will get one norm. Take all row sums, okay. Find max of the row sums, that will give you infinite norm. This is much much easier to compute… one norm or infinite norm are much much easier to compute. Let us close the… See this one norm and infinite norm are much much easier to compute than computing A transpose A and its eigenvalues, though much more complex business than doing this.

This is very, very easy, okay. So I want to take some easy way of computing norms. Now where am I going to use this, okay? Why am I computing norms? Because our condition necessary and sufficient condition was spectral radius. Spectral radius is nothing but eigenvalue, okay. So what is the relationship between norm and the eigenvalue. So that is the next part of the puzzle. Is this clear?

Now that we have 3 different ways of computing norms, 2 norm, 1 norm and infinite norm. Among these 1 norm and infinite norm are computationally preferable, okay. And now comes the point, is why am I talking about norms, yes. **"Professor - student conversation starts"** (()) (24:52) lambda1 was nothing but eigenvalue of lambda I, we have this A transpose AVi=lambda iVi. So lambdas are eigenvalues of A transpose A, okay.

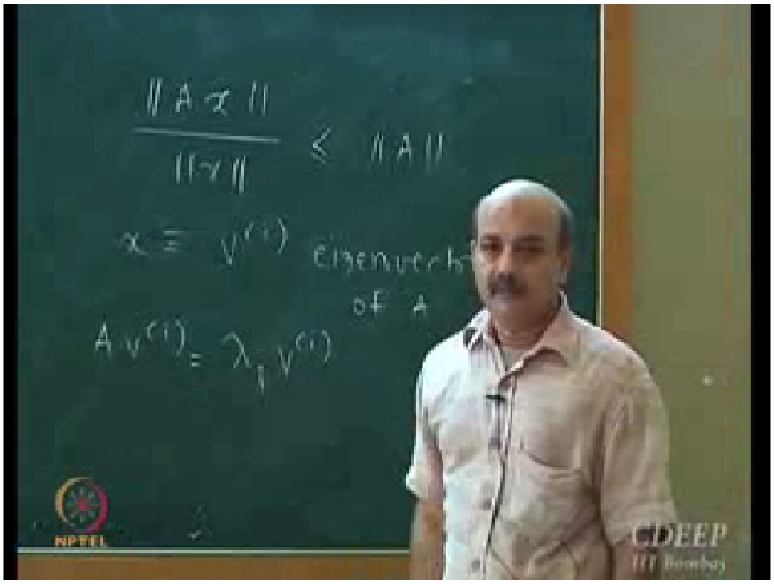
Now I said that I have numbered the eigenvalues, okay such that lambda is >= lambda2 is >= lambda3… I have numbered them, okay. So lambda1 is nothing but another way of saying lambda is lambda max, okay. Instead of giving a number and remembering, it is easier to remember this as a formula, lambda max of A transpose A. That is why I called it, okay. **"Professor - student conversation ends"** So this theorem which I am going to state is now the crux of the matter.

**(Refer Slide Time: 25:52)**



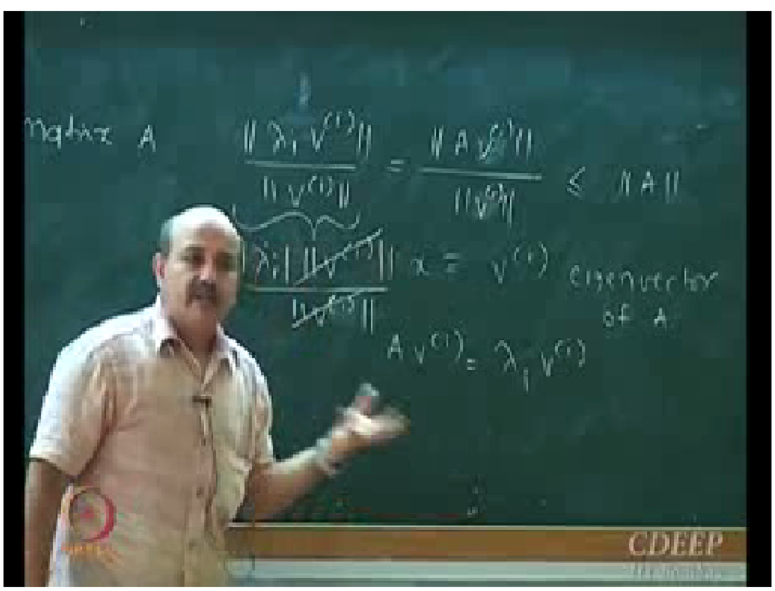

For any matrix B, any square matrix B, okay, sorry, any matrix A, for any matrix A, okay, its spectral radius is always <= any induced norm, okay. The spectral radius is always <= induced norm. So can you prove this? How will you prove this? What is spectral radius? Spectral radius of A is max over I or let us use the new notation that we have lambda max… Spectral radius is nothing but max over this, right. Now what is induced norm of a matrix?

**(Refer Slide Time: 27:04)**




For any induced norm, this is true. This is the definition right. This is the induced norm definition, okay. So this thing also holds when x corresponds to eigenvector of A. Let us say Vi is eigenvector of…, okay. Vi=lambda iVi, right. So I am going to write this is equal to norm…

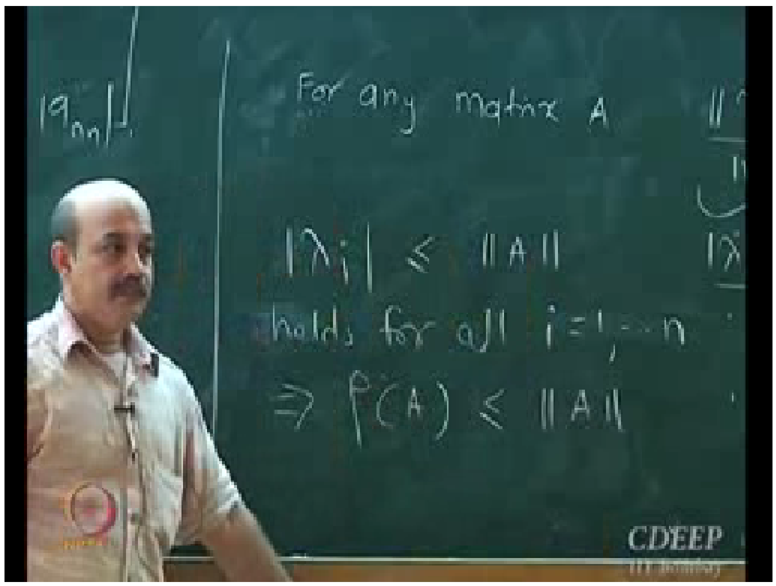
**(Refer Slide Time: 27:48)**




So AVi and if I substitute Vi here, this is nothing but lambda iVi…, right, right. But what is this quantity? When lambda comes out of the numerator, what happens? Mod lambda I norm Vi/norm Vi, right. Mod lambda I norm Vi/norm Vi, so this cancels, okay. What remains is? Mod lambda i. This holds for every eigenvector. So this also holds for maximum magnitude eigenvector, right.

See this holds for every eigenvector, okay which means it also holds for that eigenvector which has a maximum magnitude, okay. But what is the maximum magnitude eigenvalue? Maximum magnitude eigenvalue is nothing but the spectral radius, okay.

**(Refer Slide Time: 29:16)**

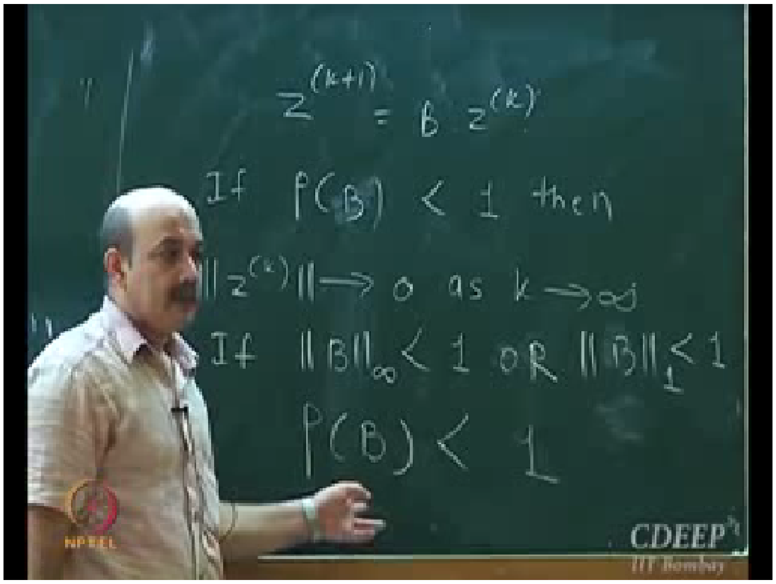



So this inequality that is mod lambda i is <= norm A. This holds for all i 1 to n and this implies that spectral radius of A is less or equal to norm of A. Every one with me on this, okay. So far so good. So now we have developed concept of matrix norms. We have expressions for computing matrix norms. Of course in Matlab, if you give a matrix and say give A matrix and say 2 norm, it will give you 2 norm which is nothing but this.

If you say i and f, infinite normal, it will give you infinite norm which is nothing but this and so on. So computing using a software these days for any huge matrix is just very very simple. Of course computationally for a large-scale matrix, this is much much easier. It just has to take absolute sums of rows and find the max, very very easy as compared to doing this. And we have a very nice relationship here, okay.

Now I am not to exploit this. I am going to use this relationship to come up with sufficient conditions for convergence, okay. Is this alright, we have matrix norms, we know how to compute them and now we know what is relationship between the spectral radius and the matrix A, its eigenvalue.

**(Refer Slide Time: 31:04)**




We were at one point analysing behaviour of systems of the time zk+1, just about 2 lectures back, linear difference equations. We were analysing behaviour of this and we said that if spectral radius of A is strictly < 1, then what happens? Then norm zk goes to 0 as k tends to infinity, okay. As k tends to infinity. But spectral radius for a large matrix is difficult to compute. You have to compute eigenvalues, okay.

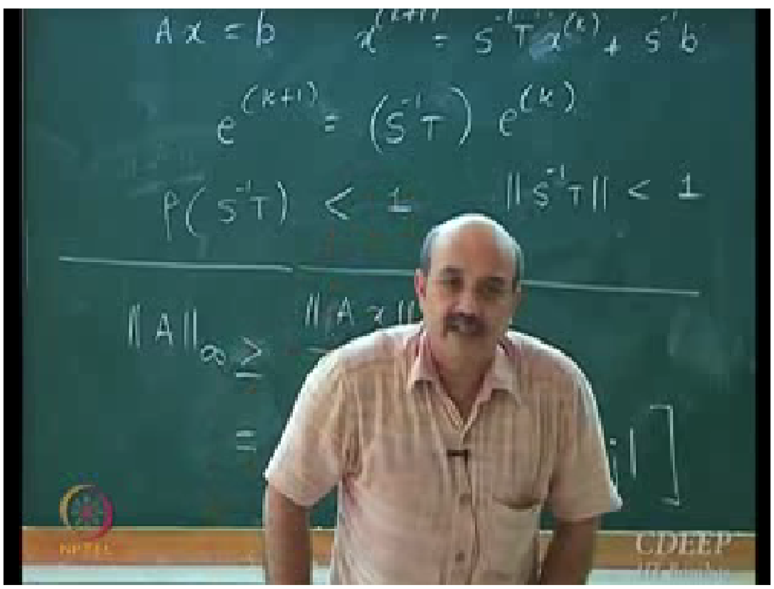
But from this inequality, what I know is that spectral radius is always < induced norm, okay. Now suppose I take A matrix, okay, compute its induced norm, I compute its infinite norm, okay or not A matrix. Here we are talking about, sorry. This should be B matrix here. I am really sorry. Just I stand corrected. This should be B matrix here. B should be strictly < 1. I take my B matrix, okay. So I take my B matrix, compute its norm, say 1 norm or infinite norm and that norm turns out to be < 1.

What can I say about the spectral radius, right? So if its norm of B, say infinite norm is < 1 or I am not going to compute 2 norm, I am going to compute only 1 norm or infinite norm. So this infinite norm or norm of B1 norm, if this is strictly < 1, either of them are strictly < 1, okay. From this theorem, what I know is that spectral radius of B should be strictly < 1, okay.

So which means a sufficient condition for convergence of this zk sequence to 0, norm of zk sequence to 0 is that take B matrix, find its 1 norm or find its infinite norm, if that norm turns out to be < 1, strictly < 1, I am done, okay. I know that zk is going to go to 0 irrespective of what happens to, what is your initial condition. It does not depend upon what is your z0. Z0 can be large, z0 can be small, okay, z0 can be arbitrary.

I know that if this condition holds then the spectral radius is always < 1 because induced norm gives the upper bound on the spectral radius, Induced norm gives the upper bound on the spectral radius and if upper bound is smaller than 1, obvious spectral radius is < 1 and then, okay. So I think we had lot of side stories. Now let me go back to solving Ax=b. So far so good. Is the line of arguments clear, okay? So let us go back and look at what we were doing.

**(Refer Slide Time: 34:41)**




Finally, we are back to what we wanted. I wanted to solve Ax=b using an iteration method xk+1=S inverse Txk+S inverse b, okay. We have this iteration method and we said that the error behaves according to ek+1=S inverse Tek, okay. S inverse Tek and then we had this condition that spectral radius of S inverse T, now S and T are different depending upon whether it is a Jacobi method or Gauss-Seidel method or relaxation method and so on.

Now typically S is a simple matrix. S is a diagonal matrix or S is s lower triangular matrix and inverting that matrix is not that difficult. Well we will come up with conditions which even do not require inversion further but right now even if you may want to do it by group force by actually inverting the S matrix even though it is not difficult but now we have this condition spectral radius < 1 for error convergence. This is necessary and sufficient condition.

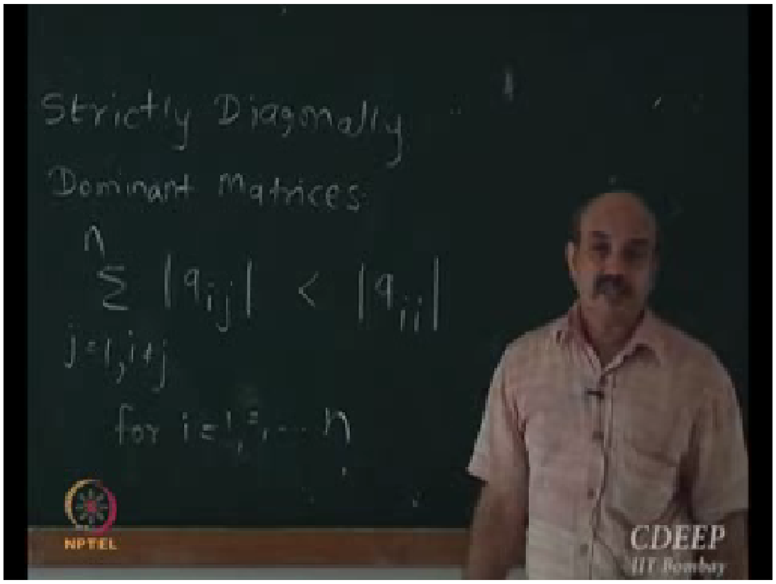
This is necessary and sufficient condition. I am now giving you a sufficient condition that if induced norm of S inverse T is strictly < 1 okay, then spectral radius is < 1 and… Why this is just sufficient? Why this is not necessary? It can happen that this is > 1 and this is < 1. See this inequality says it is like saying 0.1<5 or 0.1<0.9, okay. So if you get norm of A to be 5, quite likely that spectral radius of A could be 0.1, you do not know.

But if norm of A is 0.9, I surely know that spectral radius of A<1, okay. If norm of A comes out to be 1.1, I cannot say anything about being < 1 or this being… 1, sorry if induced norm okay, if we compute 1 norm or infinite norm, if this is < 1, definitely this is < 1, okay. But if this is > 1, we cannot say anything about what is this, okay. Inequality just says that this quantity is less than this quantity.

We are particularly interested in this number 1 whether this is < 1, so this is < 1, we are sure that this is < 1 (()) (38:02) okay. So I can actually use the norm computation, 1 norm or repeat norm computation to come up with a sufficient condition for convergence, okay. So now let us actually apply this and come up with more practical condition because I will be talking about 1000*1000 matrix or 10,000*10,000 matrix.

How do I know, how do I compute S inverse. If it is non-regular matrix, computing < 1 is again not a great idea, okay. I still want simpler conditions, okay. So what is this going to be. Now I am going to design a special class of matrices called as diagonally dominant matrices, okay. I am going to define a special class of matrices called as diagonally dominant matrices or strictly diagonally dominant matrices.

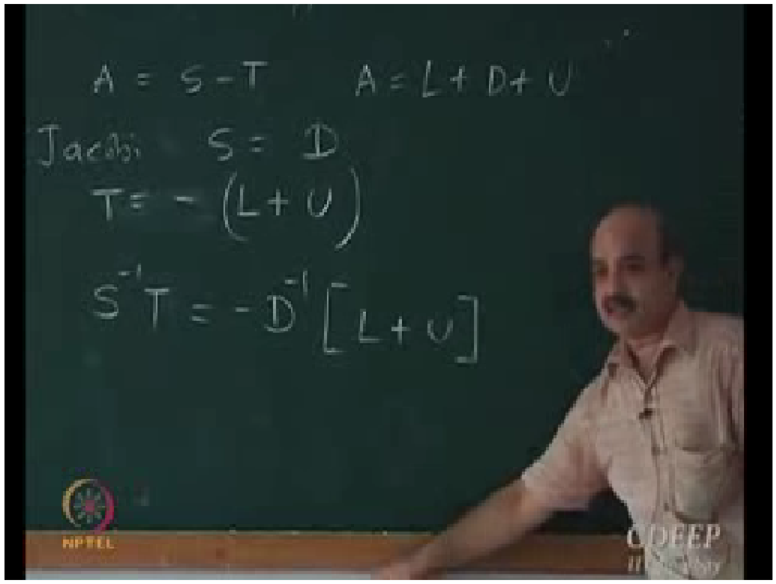
**(Refer Slide Time: 39:13)**




I am finding the summation, I am getting a matrix, okay. I am taking summation from J=1 to n. So I am making rows of, I am excluding 1 element from the rows of, which is that element? **"Professor - student conversation starts**" Diagonal. The diagonal element, okay. **"Professor - student conversation ends"** Now this mod of the diagonal element, okay, if mod of the diagonal element is strictly greater than sum of the remaining elements, absolute sum of the remaining elements, okay. Then and this should hold for every i, okay.

This is for mod i=1, 2, …n. So if these inequalities hold for each I, then such a matrix is called as diagonally dominant matrix. Such a matrix is called as diagonally dominant matrix. Now where (()) (40:27) I will move on here. I hope you have all these in your notebook.

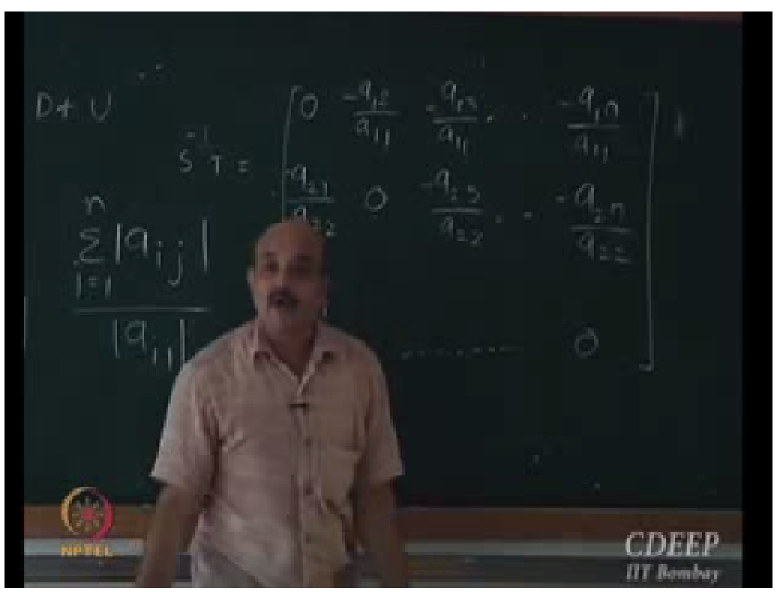
**(Refer Slide Time: 40:41)**




Let us go back and see in the Jacobi method, what was, we wrote A=S-T, okay. For Jacobi method, we had S=, and we also wrote this that A=L+D+U. Then for Jacobi method, S=D and T=-(L+U), okay. Now just think of think S inverse T which is -D inverse L+U. What is D inverse? D is a diagonal matrix, all the elements, inversely just 1/diagonal elements, okay.

Now if matrix a is diagonally dominant, if matrix a is diagonally dominant, okay, then what happens? aii are all > the summation of all the row elements, absolute of those elements. What is this matrix? Can you just write down, what is this matrix?

**(Refer Slide Time: 42:24)**




For Jacobi method, S inverse T will be, what will be this matrix? 0 -a12/a11 -a13/a11 …

-a1n/a11, -a21/a22 0 -a23/a22 … -a2n/a22, right. This S inverse T will be a matrix which has 0 on the diagonal, all diagonal elements will be like this. What will be its infinite norm? Row sum, row sum. Row sum will be nothing but, what will be each row sum? Summation aij, j going from 1-n, divided by aii, aii is divided each row, just look at here, okay. What is the value of the matrix which means this sum is strictly <… right?

This sum is strictly < this. If this sum is strictly < this, what does it mean? That all these ratios are strictly < 1. What does it mean? infinite norm is strictly < 1. If infinity norm is strictly < 1, what can you say about the spectral radius of this matrix? Okay. So now I have reduced checking whether Jacobi method will converge or not just to see whether a matrix, whether this diagonally dominant or not.

If matrix a diagonally dominant, okay, my iterations will converge irrespective of where I start from, okay irrespective of where I start from, my iterations will converge if my matrix a is diagonally dominant. So after bringing all these juggling lot of arguments, (()) (45:00) matrix norms and then spectral radius all that, you have come up with very simple criteria for finding out whether these Jacobi iterations will converge or not. I will reduce some more theorems.

I will not get you the roots of each one of them (()) (45:17) but I will give you some more theorems which are very elegant and from which you can ascertain whether the iterations will converge or not or you can modify your problem such that your iterations are diagonally converge, okay. So that is what we will see in next class. We will see that then we will move on some other norms. But this is where you can see and know I can value spectral radius and norms, everything is actually.

Well, we will get into analyse the behaviour qualitatively without actually having to solve it, just looking at diagonal dominance, I can come to the conclusion for any initial case, our iterations will converge, okay. (()) (46:01).