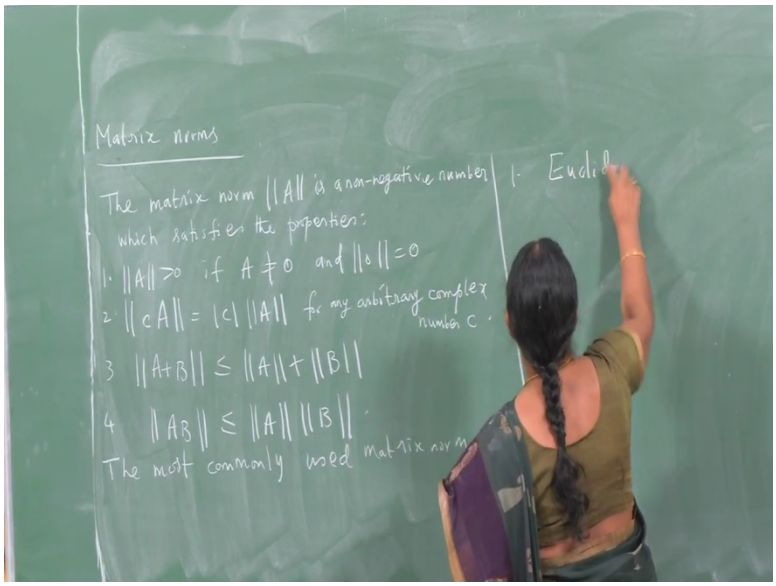
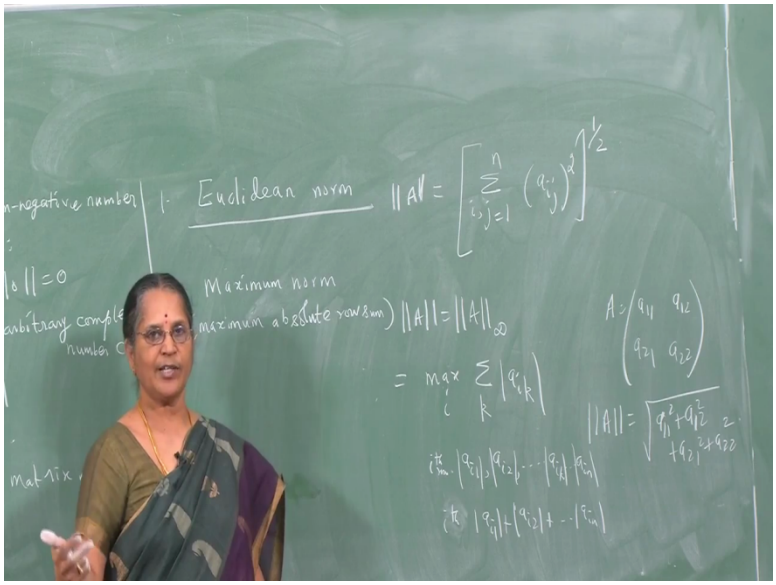
Numerical Analysis
Professor R. Usha
Department of Mathematics
Indian Institute of Technology Madras
Lecture No 44
Solution of Linear Systems of Equations -6 Error Analysis 2

(Refer Slide Time: 00:20)



So we introduced matrix norms, so the matrix norms we denoted by norm A is a nonnegative number which satisfies the following properties. Namely, norm A is greater than 0 if A is a nonzero matrix and norm of 0 matrix is 0. Secondly, norm of c into A where c is a scalar is modulus of c into norm A for any arbitrary complex number c, and norm of A + B is less than or equal to norm A + norm B and norm AB will be less than or equal to norm A into norm B, so these are the properties which are satisfied by norm of a matrix A. So now let us introduce the most commonly used matrix norm say the Euclidean norm denoted by norm A and that is Sigma i j running from 1 to n a ij square whole to the power of half.

(Refer Slide Time: 2:57)



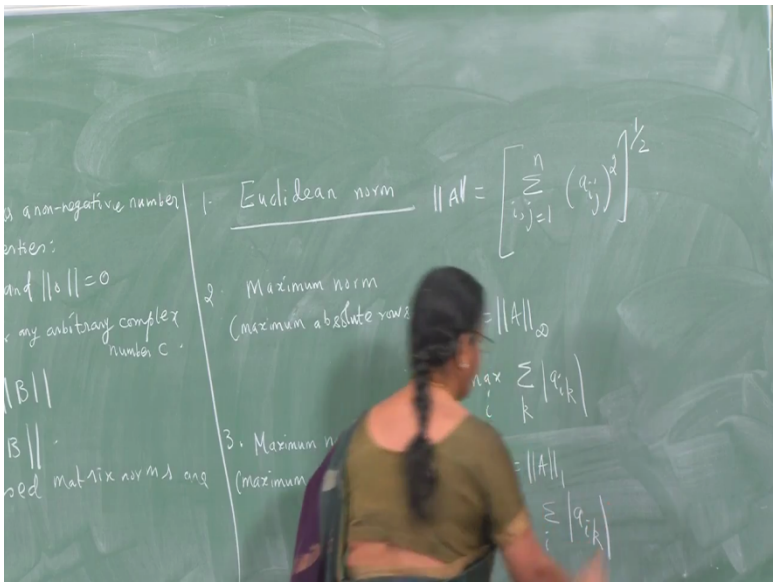
How do you compute this? you are given a matrix A whose entries are say I take a 3 by 3 matrix or a 2 by 2 matrix a 11, a 12, a 21, a 22, the Euclidean norm is computed by taking square root of the sum of the squares of all the entries, so a 11 square, a 12 square, + a 21 square + a 22 square. So take all the entries in the matrix A, square them and add them, take the square root of that number that gives you norm A. Now secondly maximum norm is very often used what is it? it is the maximum absolute row sum and it is defined as follows. And that is equal to maximum over i Sigma over K modulus of a i k, so let us try to understand this norm.

What are we supposed to take? you compute absolute values of the entries so that in the ith row you have those entries to be a i1, a i2, a etc, a ik, etc, a in if a is an n cross matrix. Where did you take, we took these entries in the ith row and then take the absolute values of all these entries and then sum them up so it is mod a i1 + modulus of a i2 + etc + modulus of a i n so you sum up all those entries, this sum is for the ith row, do that for all the rows namely start with the first row compute the absolute value of the entries in the first row and then go to the second row, compute the absolute values of the entries in the second row, add them up and do that for all the n rows.

From among those n values that you have got, you pick that which is the maximum namely, you have picked that which is the maximum absolute row sum and that is your norm A. I hope it is clear, what should you do? you take typically ith row, look at its entries, take the absolute value of these entries, sum them up, do this for all the rows if A is an n cross n

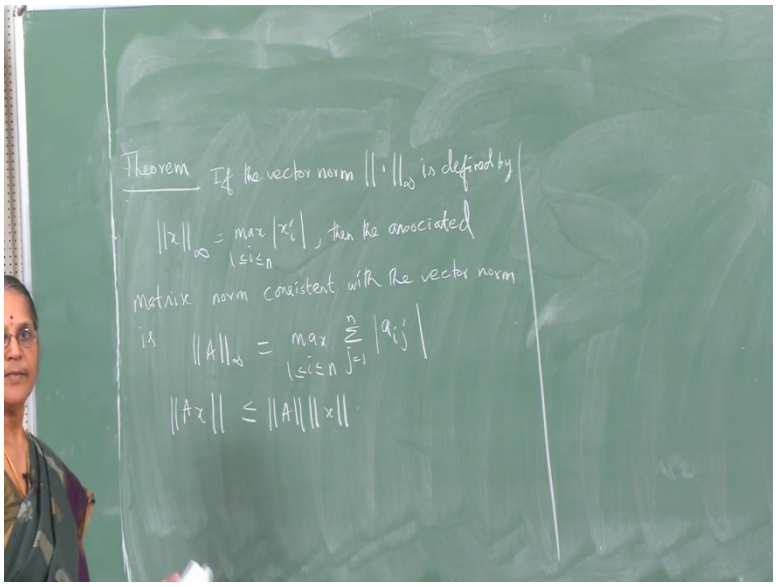
matrix, there are going to be n such rows so do this for all the n rows, you have now n values. From among these values pick that which is the maximum and that gives you the maximum absolute row sum. One can also define the maximum norm as follows.

(Refer Slide Time: 6:52)



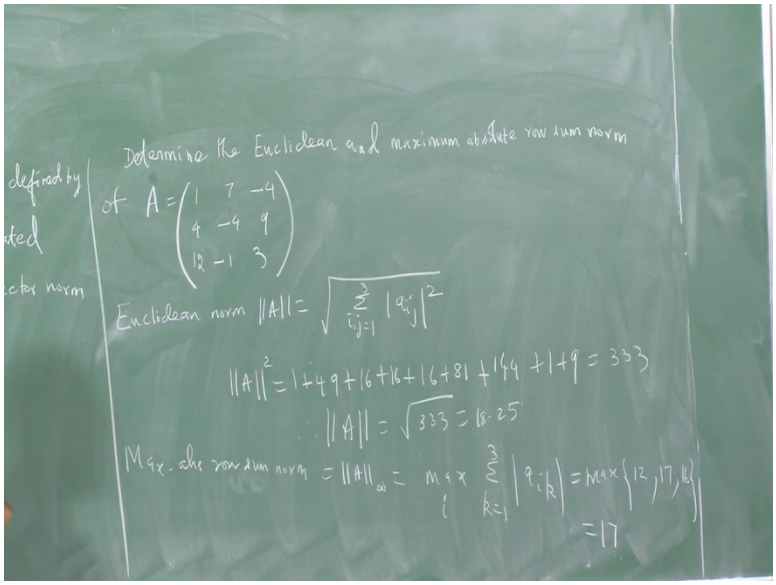
In this case it is the maximum absolute columns, so norm A is going to be norm A 1 and that is maximum over k Sigma over i mod a i k, so we must have the matrix norm to be such that it is consistent with the vector norm that we choose so we have the following result which can be used to appropriately take the matrix norm associated with a vector norm.

(Refer Slide Time: 7:49)



The results says, if the vector norm infinity is defined by norm infinity of a vector x is maximum of modulus of x i for i lying between 1 and n then the associated matrix norm consistent with the vector norm is norm A infinity so it is the maximum of j = 1 to n modulus of a ij for i lying between 1 and n so it is the maximum absolute row sum norm. So if you use the vector norm as infinity norm then the associated matrix norm consistent with the vector norm is maximum absolute row sum norm and in this case norm A x will be less than or equal to norm A into norm x.

(Refer Slide Time: 9:55)



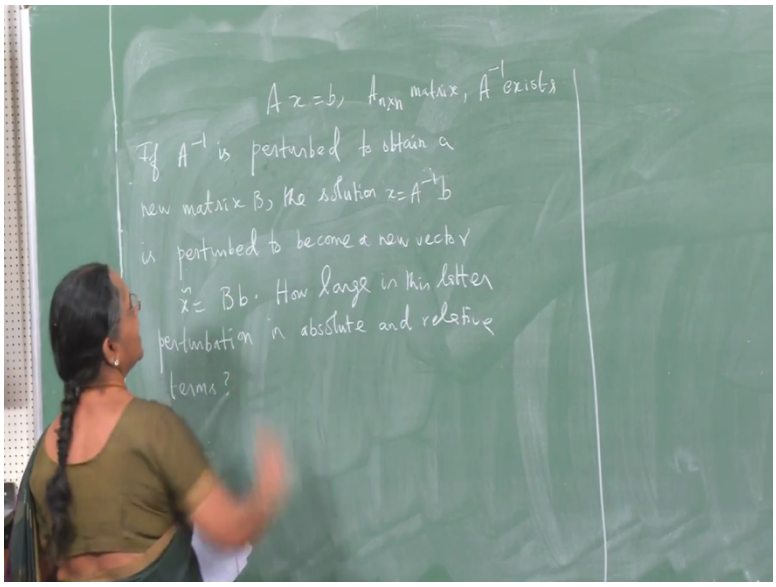
So let us now consider the following example, so let us determine the Euclidean and maximum absolute row sum norm of the matrix A which is 1, 7, – 4, 4, – 4, 9, 12, – 1, 3. So we want Euclidean norm, so what is Euclidean norm given this matrix A, it is norm A which is equal to square root of Sigma i j taking values 1 to 3 of modulus of a ij square, so let us compute norm A square so it is 1 square + 7 square + – modulus of – 4 the whole square + 4 square + again 4 square + 9 square + 12 square + 1 square + 3 square so that turns out to be 333 and therefore the Euclidean norm or Euclidean matrix norm A is root of 333 which is 18.25.

Let us now compute the maximum absolute row sum norm. It is by definition is denoted by norm A infinity and by definition it is maximum over i Sigma K = 1 to 3 mod a I k, so what should I do? I should take the first row, take the absolute values of the entries in the first row, add them up and do that for all the 3 rows from among them pick that which is the maximum. So this is going to be maximum of what is the absolute row sum here? Mod 1 + mod 7 + modulus of – 4 and so that is 12, 4 + 4 + 9 so 8 + 9 that is 17, for the third row 12 + 1 + 3 and

therefore it is 16, so I have 3 such values from among them I should choose that which is the largest namely 17 so this gives me maximum absolute row sum norm for the given matrix A.

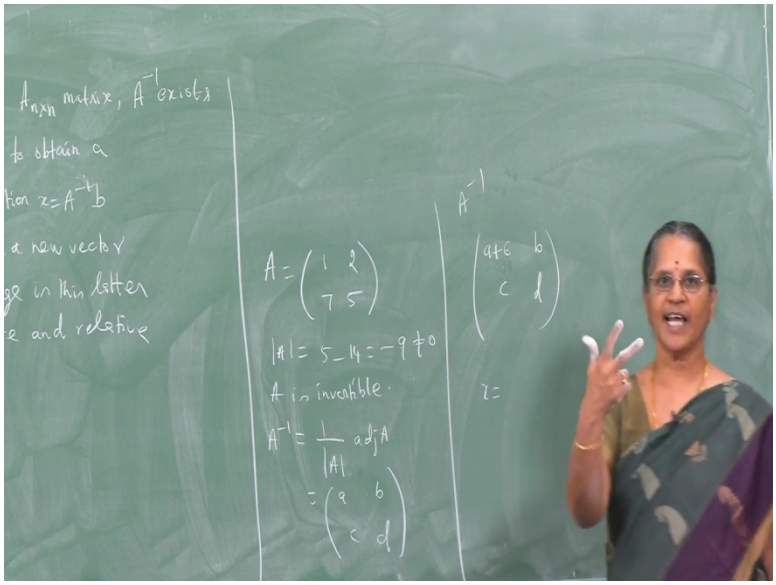
So let us use this notion of norms of a matrix and the given vector and try to perform error analysis of the direct method, so let us work out the following example which is going to give us the bound on the absolute error if during our computation some amount of error has been incorporated, so let us consider the following example.

(Refer Slide Time: 12:51)



So if suppose A inverse is perturbed to obtain a new matrix B, the solution x = A inverse b is perturbed to become a new vector x tilde which is B into b, so the question is how large is this latter perturbation in absolute and relative terms. If a small change is given in the input then how large is the deviation in the output is what the question is. So we are given a system of equation A x = b where is an n cross n matrix and A is non-singular so that A inverse exists or in other words A is an invertible matrix. The question is, if suppose in the matrix A inverse you give a small change, we have already seen that.

(Refer Slide Time: 14:37)



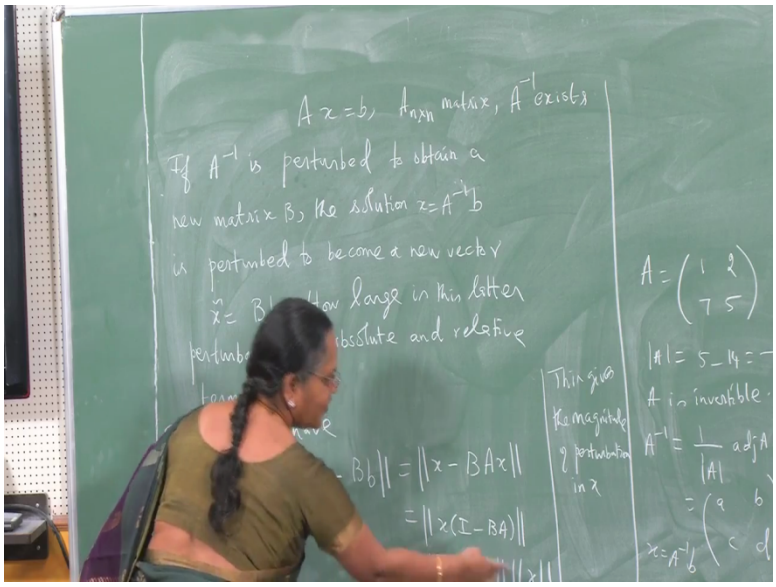
Suppose I am given a 2 by 2 matrix 1, 2, 7, 5, this matrix is non-singular since determinant A is going to be 5 – 14 that is – 9 so A is invertible. And I compute A inverse which is 1 by determinant of A into A joint of A. And suppose I get the inverse matrix to be a, b, c, d, now if I give a small change in anyone of these entries so that the entries are say a + Epsilon, b, c, d, so my A inverse which is based is perturbed namely slightly changed because of the change in one of the entries so a has become say a + Epsilon, I continue with the computation of getting the solution of this system and taking x to be equal to A inverse b, actually my A inverse b should be such that it should take this as A inverse but I take this as my A inverse and workout the solution.

So the solution that I get is something different so the question is how large is this difference if the actual solution is x and the computed solution is a say x tilde, what is the difference x – x tilde what is its magnitude, since it is a vector what is the norm of this vector x – x tilde that is what the question is, so let us understand the question again. If A inverse is perturbed that is some slight change is given to any one of these entries to obtain a new matrix B so this is my new matrix B, the solution x is equal to A inverse b so if I take this as A inverse and compute my solution as A inverse b that solution is perturbed to become a new vector x tilde.

While I am going to use this as my A inverse and obtain the solution so it is B times vector b so with this my solution is B b and I should have got this solution as A inverse b, so I have got it as B b so I have got a new vector x tilde. So how large is this change in absolute and relative terms so compute what is x – x tilde compute its magnitude that is going to give me

in absolute terms this is the magnitude that results or in relative terms I am going to get the norm of this vector x – x tilde by x so this is what we are asked to work out. So let us write down the solution of this problem.

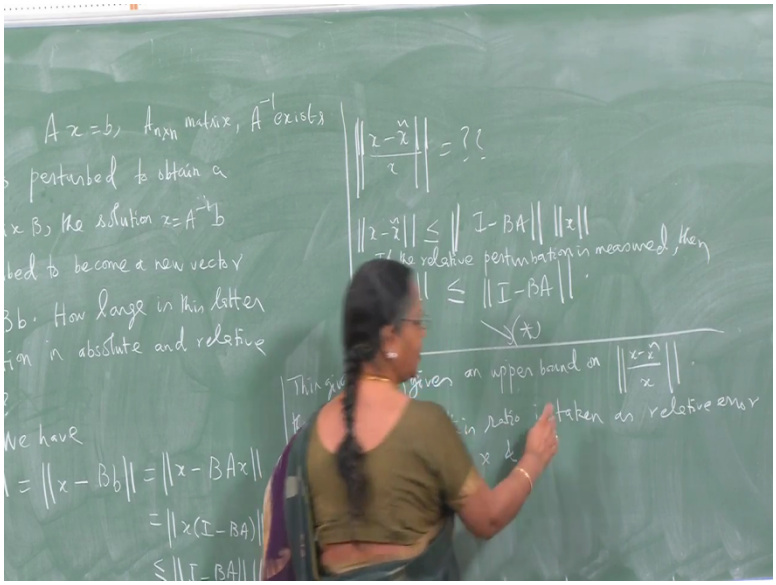
(Refer Slide Time: 18:07)



So let us see what is norm x – x tilde, what is x, I write x as it is but what is x tilde, it is B times b and is norm of x – B into what is b, b is A x because I am given the system A x so x – b A into x (())(18:31) into – B A. So I have something like a vector multiplied by a matrix, I is the identity matrix of the same order as that of A. So we have seen that norm of A matrix satisfies the property that norm A B is less than or equal to norm A into norm B. So here I have x into a matrix and therefore this will be less than or equal to norm of I – B A multiplied by norm of x so this is the amount of perturbation in x that is what we have.

So if x is disturbed and the resulting value that you get resulting solution that you get is x tilde, then the magnitude of that change is less than or equal to this mainly norm of I – is BA into norm x. So this gives us the magnitude of perturbation in x, what do you mean by perturbation in x? The change in x the disturbance given to x so you have in absolute terms the magnitude of the change x that this cannot be greater than this value, so let us discuss the details when we want to give the result in terms of relative magnitude.

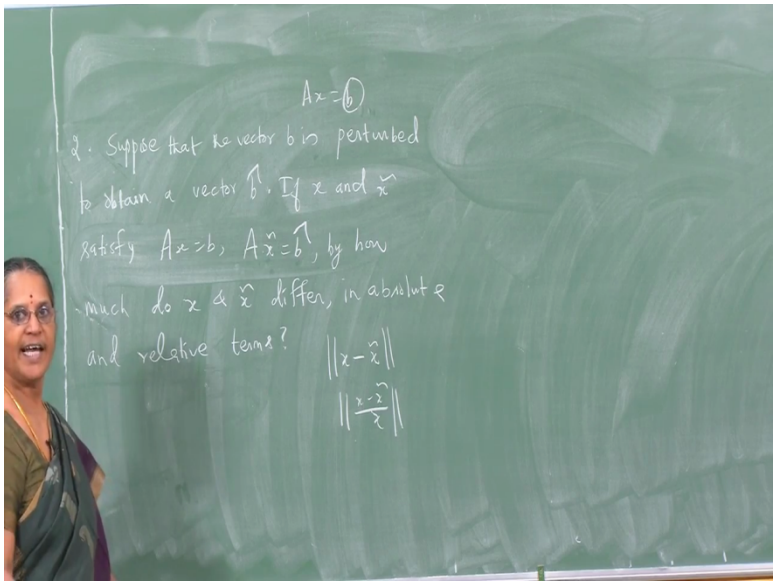
(Refer Slide Time: 20:36)



So I want norm x – x tilde divided by x, what is it that I know? I know norm x – x tilde is less than or equal to norm of I – B A into norm of x. I mean I divide throughout by norm x so that will be less than or equal to norm of I – B A. So if the relatives perturbation is measured then we have the relative change is such that it cannot exceed norm I – B A, so we can appropriately choose the norm of a matrix that appears here and then give the relatives error due to the upper termination given in A inverse, which results in perturbation in the actual solution.

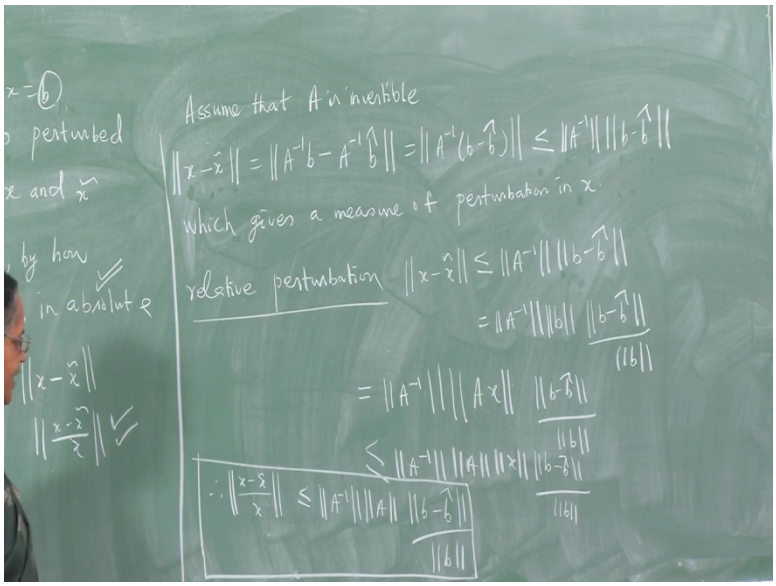
So these are very useful results, namely this gives you I call this as star, star gives an upper bound on the relative magnitude of the error and this ratio is taken as relative error between the actual value and the computed value namely between x and x tilde, let us consider another problem.

(Refer Slide Time: 22:33)



Suppose that the vector b is perturbed to obtain nectar b cap, so entries in the right-hand side vector for a given system A x = b are slightly changed so that the new right-hand side vector is b cap if x and x tilde satisfy A x = b and A x tilde = b cap. The question is by how much do x and x tilde differ and give this result in absolute and relative terms. So the question is, get the magnitude of the absolute error and the magnitude of the relative error, I hope it is clear it has small change in the entries in the right-hand side matrix b is given so that you solve the system A x tilde = b cap so this x changes to x tilde because right-hand side vector has changed from b to b cap. So you obtain your solution as x tilde whereas, the actual solution is denoted by x because you want to solve the system A x = b.

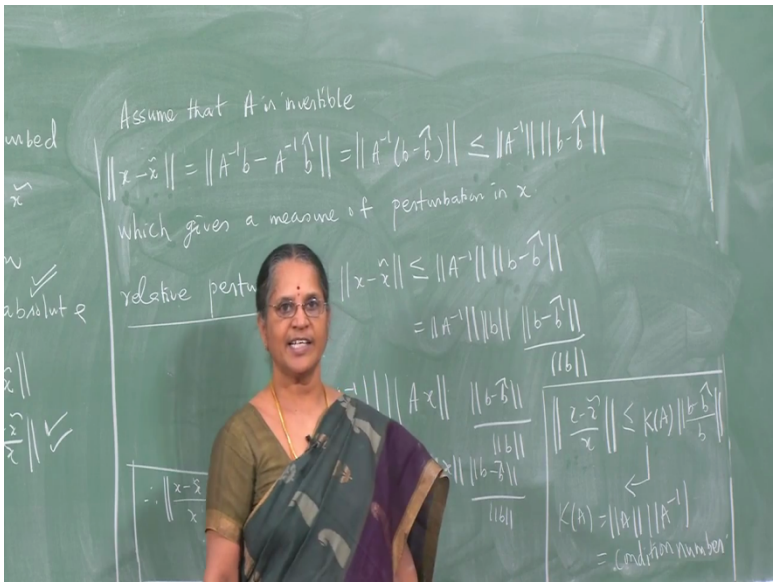
(Refer Slide Time: 25:07)



So the question is compute what is the error x – x tilde and get its magnitude, so norm x – x tilde so that you have magnitude of the absolute error. Also compute the magnitude of the relative error in this case if there is upper termination given to the right-hand side entry, so let us compute the result. So we assume that A is invertible, so we compute the magnitude of the absolute error which is norm x – x tilde, what is x? x is A inverse b, what is x tilde? It is A inverse into b cap, so it is norm A inverse into b – b cap, so by the property of the matrix norms this is less than or equal to norm A inverse into norm b – b cap. So this gives you the measure of perturbation in x, this gives a measure of perturbation in x or the change in x, what is the magnitude of the deviation?

So we have computed what we want namely in absolute terms the result is available. Now we want to get in relative terms of measure of perturbation in x. So we start with computation for relative perturbation so we start with the result that we already have namely, norm x – x tilde is less than or equal to norm A inverse into norm b – b cap, so this is norm A inverse into I shall multiplied by norm b and therefore I shall divide by norm, so I have not altered anything in this step, I have multiplied and divided by the same county. So this will be equal to norm A inverse, but what do I know about norm b, A x is b so norm b is norm A x so into norm b – b cap by norm b.

(Refer Slide Time: 30:20)

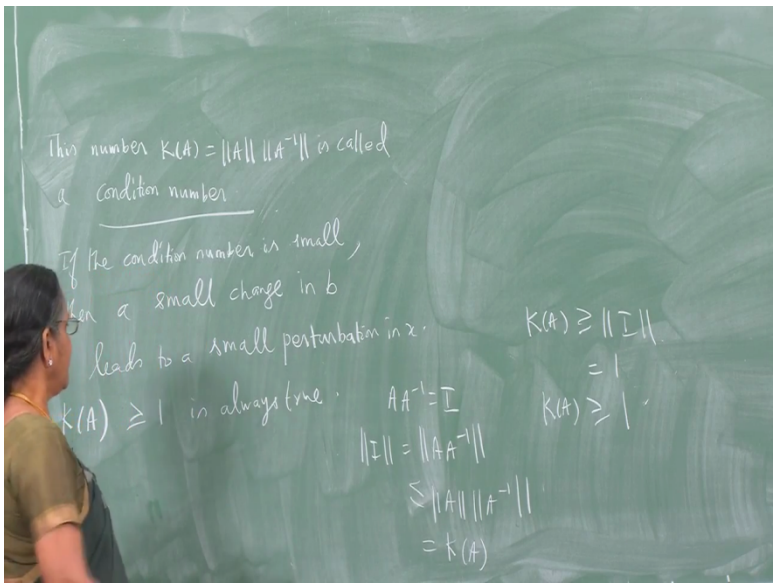


So I know norm A x is less than or equal to norm A into norm x, so I use that property and right down the result. This multiplied by Norm b – b cap by norm b. So I divide throughout by norm x so I get norm x – x tilde by norm x is less than or equal to norm A inverse into norm A into norm b – b cap by norm b. So I have result which gives me the magnitude of the

relative change in x is less than or equal to norm A into norm A inverse into the magnitude of the relative change in the right-hand side vector. I denote norm A into norm A inverse as k of A, so k of A multiplied by norm b – b cap by b, what is k of A? k of A is norm A into norm A inverse, which I call as the condition number.

you observe from this result that if the condition number is very large then the relative change in x is going to be very large in which case the system will be a well conditioned system. So we can immediately conclude about whether the system is an ill conditioned system or a well conditioned system by computing the condition number and looking at its magnitude, if the magnitude namely of a condition number turns out to be very very large then this result tells you that the relative error in x namely the output that you get as x tilde is such that it is deviated very much from x if the condition number is very very large so that the system becomes an ill conditioned system.

(Refer Slide Time: 30:31)

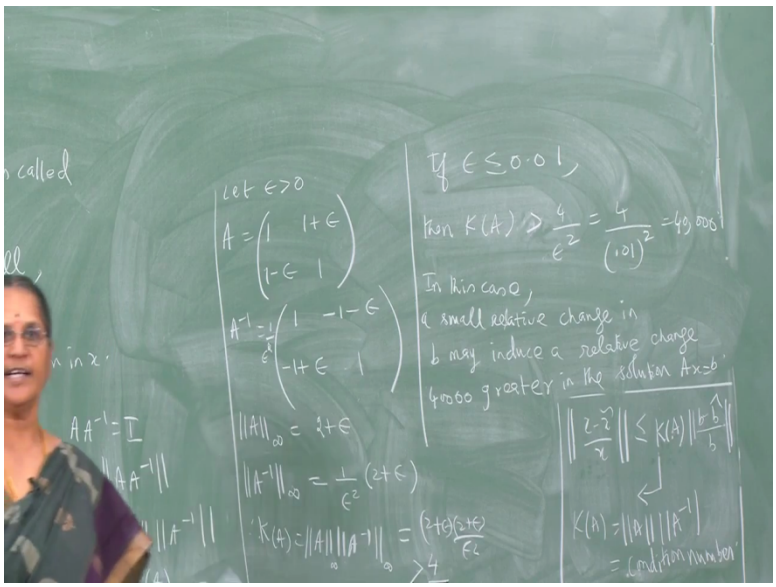


This number k of A which is equal to norm A into norm A inverse is called a condition number and the results says that if the condition number is small then a small change in b or a small change or perturbation in b leads to a small perturbation in x. The condition number is small then small change in the right-hand side vector will result in a small change in the output namely the solution vector for the problem. And what can you say about this k of A, k of A have to be greater than or equal to 1 and this is always true, how do you show that that k of A must be greater than or equal to 1 is always true. So let us take A A inverse what is it? It is identity matrix.

So norm of I must be norm of A A inverse and that is less than or equal to norm A into norm A inverse, but what is it and that is what we call as the condition number of the matrix A. So condition number of the matrix A must be greater than or equal to norm of I, what is I, I is an identity matrix so this will be 1 and therefore k of A the condition number has to be greater than or equal to 1. So this result is always true side of the condition number that you get here for a matrix A to be such that it is very close to 1 then a small change in the right-hand side vector will result in a very small change in the solution vector and so the system will be well conditioned system so that the solution that you obtain at the end will be a reliable result.

On the other hand, if the condition number k of A is very large and it is very much deviated from the value 1 say something like 40908 and so on then your output will be such that it will be very much deviated from the actual solution so your solution cannot be a reliable solution and the system is an ill conditioned system in this case.

(Refer Slide Time: 34:14)

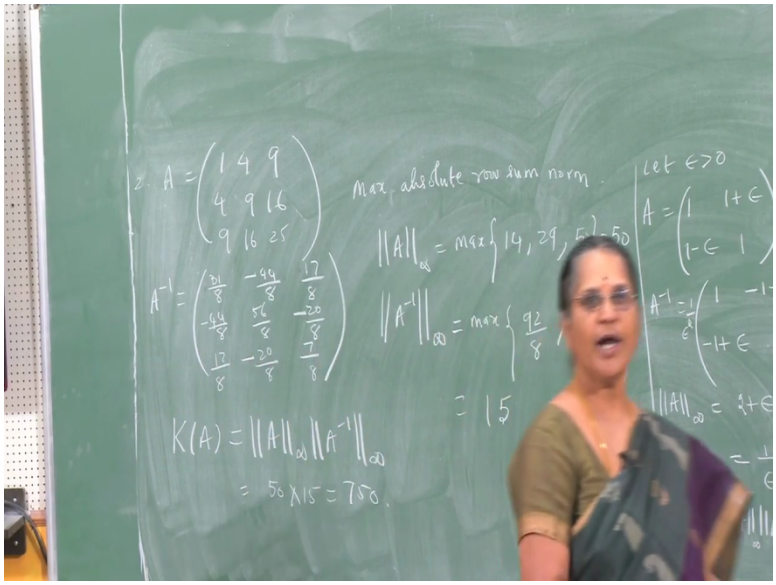


Suppose say Epsilon is positive and I give you a matrix A which has entries 1, 1 + Epsilon, 1 – Epsilon and 1 then compute A inverse which is 1 by Epsilon square into 1, – 1 – Epsilon, – 1 + Epsilon and 1, so let us compute norm A infinity. So it is the maximum absolute row sum so it is 1 + 1 + Epsilon so 2 + Epsilon and here it is 2 – Epsilon maximum so we take that to be norm A infinity. Similarly we compute norm A inverse infinity so that is 1 by Epsilon square into 2 + Epsilon again, so therefore k of A is norm A into norm A inverse.

I choose the infinity norm for computing norm A infinity and norm A inverse infinity, so it turns out to be 2 + Epsilon into 2 + Epsilon by Epsilon square so it is greater than 4 divided

by Epsilon square and therefore if Epsilon is less than or equal to 0.01 then k of A which is greater than 4 by Epsilon square will be 4 by 0.01 the whole square so will be greater than 40,000. So in this case what happens, in this case small relative perturbation in b may induce a relative perturbation in x when you solve a system A x = b and therefore the system will be an ill conditioned system.

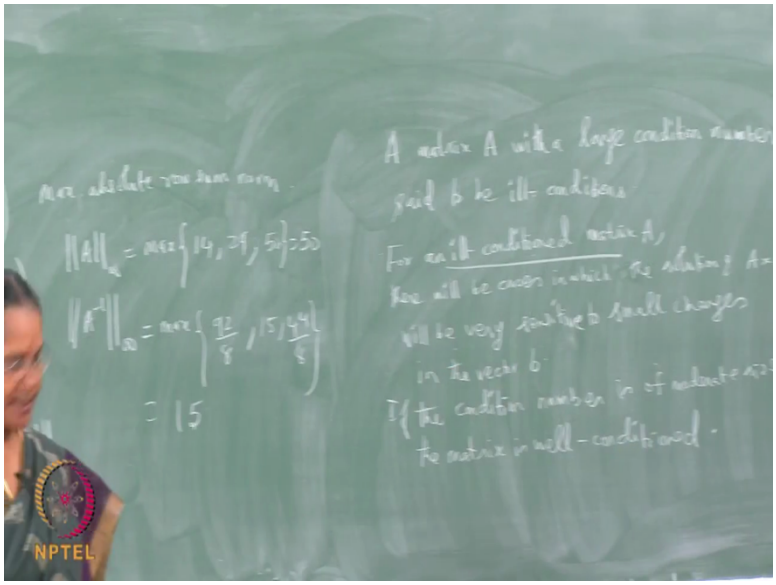
(Refer Slide Time: 36:43)



Suppose say matrix A is 1, 4, 9, 4, 9, 16, 9, 16, 25, use the maximum absolute row sum norm and compute the condition number. So I require A inverse, how do you compute A inverse? you already know, apply Gauss Jordan technique and compute A inverse and show that A inverse is 31 by 8, – 44 by 8, 17 by 8, – 44 by 8, 56 by 8, – 20 by 8, 17 by 8, – 20 by 8 and 7 by 8 so I first have to compute norm A infinity which is it is the maximum salute row sum norm, so I must take the maximum of the sum of the absolute values of the entries in each of these rows so it is going to be 14, 29 and 50, the maximum is 50.

Next I will compute norm A inverse infinity, so that is going to be maximum of 31 + 1 dealers of 44 + 17 by 8 so that turns out to be 92 by 8. Similarly I compute the absolute row sum values for the second row and the third row and it turns out to be 15 and 44 by 8 so the maximum is 15. So we evaluate the condition number of the given matrix A and that is norm A infinity into norm A inverse infinity so it is 15 into 15 that is 750 which is large as compared to a value which is 1 or close to 1. If you are going to solve a system of equations with A as the coefficient matrix then the corresponding system is going to be an ill conditioned system because the condition number of A is large.

(Refer Slide Time: 39:40)



A matrix A with a large condition number is ill conditioned, so in this case so for an ill conditioned matrix there will be cases in which the solution of the system A x = b will be very sensitive to small changes in the vector b. Our second example shows this namely if you have a small perturbation in the right-hand side vector b and if the matrix A is an ill conditioned matrix because the condition number of the matrix is very large then if you see the solution of the system A x = b with A having condition number to be very very large then the solution of the system will be very sensitive to this small change in the vector b. On the other hand, if the condition number is of moderate value then the matrix is said to be well conditioned.

The question now is now that we have performed the error analysis, how are we going to make use of this information in improving the accuracy of the solution that we have obtained using direct methods. Is there a way to make use of the results that we have derived in the error analysis so that we can improve the accuracy of the solution of the system of equation A x = b which are obtained using direct methods? Yes it is possible and we shall try to discuss these details in the next class.