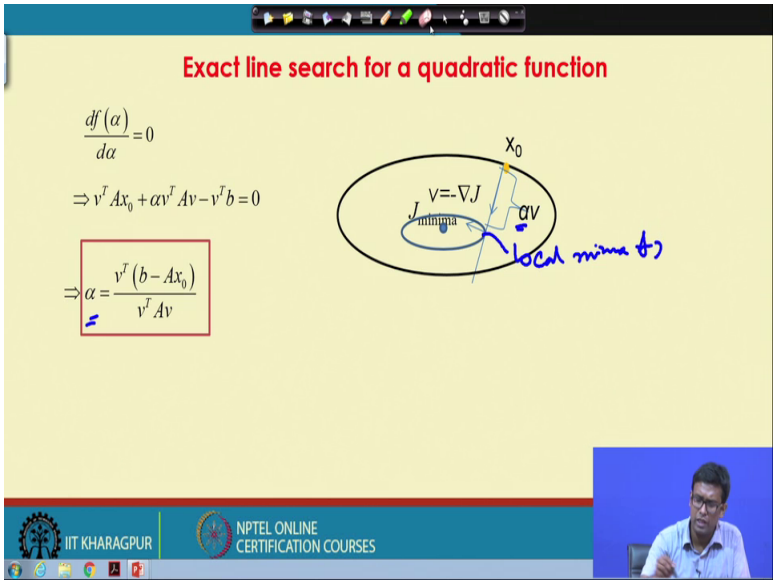
Matrix Solvers
Prof. Somnath Roy
Department of Mechanical Engineering
Indian Institute of Technology, Kharagpur

Lecture – 38
Steepest Descent Method: Algorithm and Convergence

Welcome so, we are discussing about the Steepest Descent Algorithm where we have seen that solving a problem Ax is equal to b is equivalent to find out minima of the functional J X where J X is defined as x transpose Ax minus x transpose b. And, in order to propose an iterative method for solving for finding the minima of J X we discussed about the gradient such algorithm that choose any value X 0 find J X 0 there and then find out the gradient of J and move along minus gradient of J.

And move up to a certain distance where this along with J reduces on that particular line and this distance is measured by parameter alpha we discussed about how to find out alpha till which J reduces along that line. And then when you see that J has reached a local minima along that particular line change the direction and go to the another take another direction.

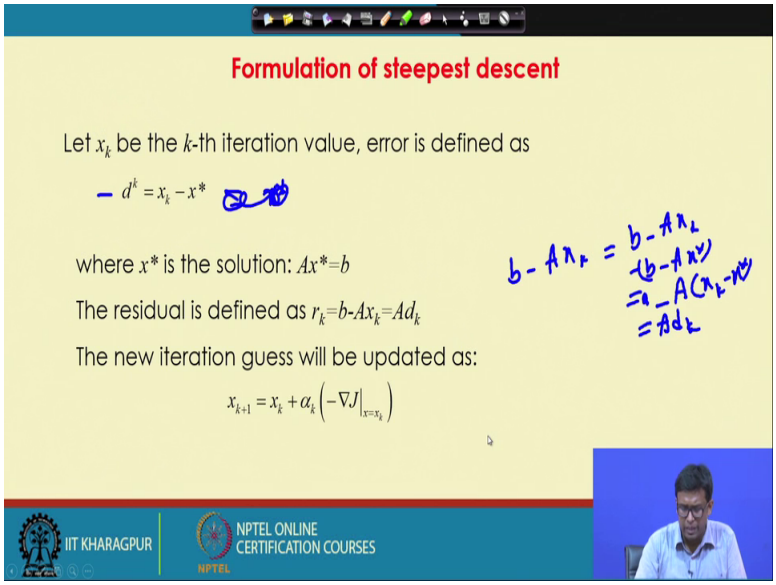
(Refer Slide Time: 01:20)



So, roughly it is this particular method that start with one particular X 0 move along one direction and then see that, this is a local minima of J from here you take another direction and that is how you approach the global minima.

And this is the parameter alpha that; distance should be covered in particular one particular search direction. So, that the search is optimum and in least number of steps we reach the local minima. Now we thought about taking this into a solver and writing and proposing an algorithm for iteratively solving x is equal to b or finding J minima using this.

(Refer Slide Time: 02:15)



Which is called the steepest descent algorithm let k be the k th iteration value and the error is defined as d k is x x k minus x x star where, x star is the solution of Ax star is equal to b, x star is the where we have the solution x star is equal to b that is the location where J is minima.

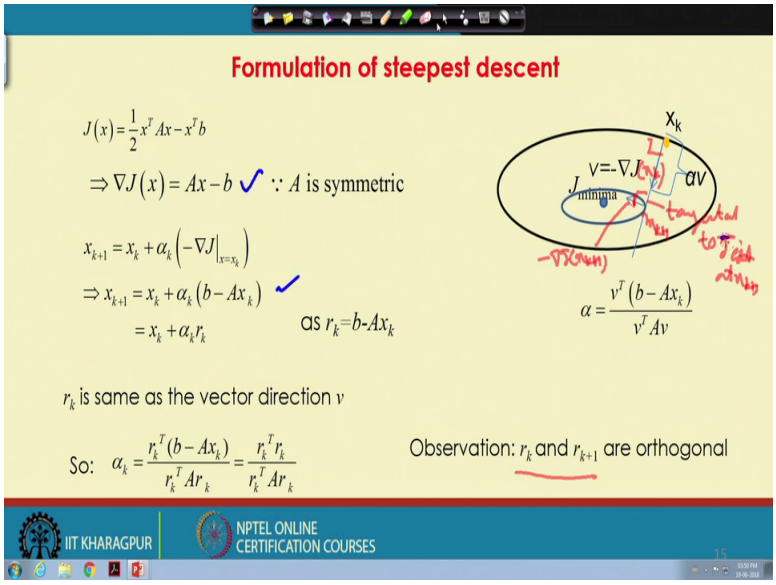
So, this let x star be the point where J is minimal and d k is the error; that means, we considered a it guess value x k or k th iteration value x k d k is the error which is x k minus x star. The residual is defined as b minus Ax k, if x k is equal to x star we reach the right value b minus Ax is equal to 0.

So, residual is 0 till it is not reached this is the non 0 value and we define residual as b minus Ax k or this is equal to b into A into d k why? Because b minus Ax k is equal to b minus Ax k minus b minus Ax star because this is 0. So, this is if you duel is d k is equal to sorry, there is a there is a small sign convention small sign convention issue r k is b minus Ax k and this is so.

So, let us defined this is equal to x star minus x this is minus d k then it should come out. So, this is minus of A into x k minus x star which is A into d k. So, let us define minus d k is x k minus x star, new iteration guess will be updated as x k plus 1 is equal to x k plus alpha k into minus gradient of J x is equal to x k.

So, it started with a value of x k and will update it to the new iteration x k plus1. And this is this is how because we are also when we are in thinking of solving x is equal to b we are also trying to find out the minima of J. So, you should go along minus grad J x k and we found out there is a parameter alpha k by which we should go in that direction.

(Refer Slide Time: 05:09)



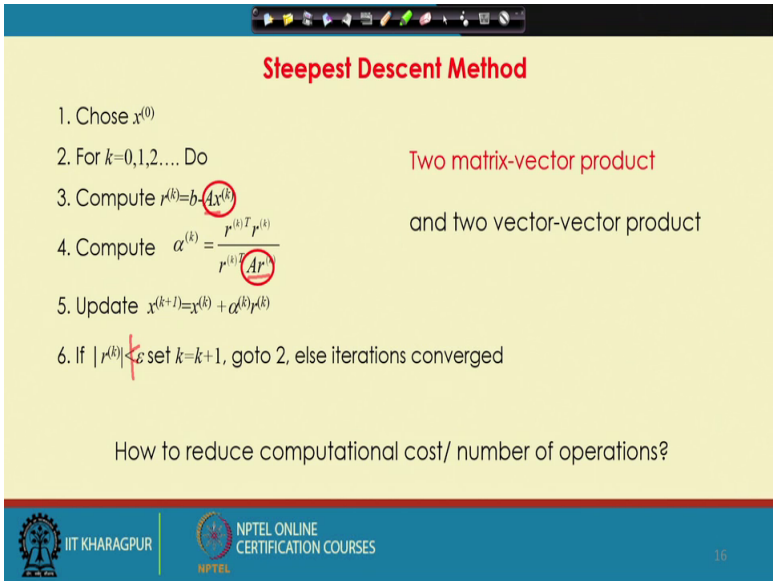
Now, have J x is equal to half of x transpose Ax minus b grad J is equal to Ax minus b as A is a symmetric matrix, x k plus 1 is equal to alpha k into minus grad J of x is equal to x k which is evaluated there x ks in that earlier and minus grad J is b grad J is A x minus b.

So, minus grad J is grad J is A x minus b therefore, minus grad J is b minus Ax. So, x k plus 1 into alpha k into b minus A x k and we have defined b minus A x k was r k. So, this is x k plus 1 is equal to x k into x k x k plus 1 is equal to x k plus alpha k r k, r k is the same vector as the vector direction as the same vector in the direction v r k and v are same. It because we have written earlier v is equal to minus grad J now we can see that v is same as r k.

So, I have seen that alpha is equal to v if v is minus grad J, now we have seen that minus grad J is r k. So, r k and v are same alpha is equal to v transpose b minus Ax k v transpose Av. So, we will get alpha is equal to r k transpose r k and v are same b minus Ax k by r k transpose Ar k which is r k transpose r k into r k transpose Ar k. And there is one observation which is r k and r k plus 1 are orthogonal why? Because, grad J is the direction in which J reduces first test, grad J is perpendicular to the J contour where it was evaluated.

Now, we move till J is minima or grad J this is tangential to J contour to tangential through J contour at x k plus 1 this is x k plus 1. The new direction is minus gradient of J x at x k plus 1 is perpendicular to the tangent. So, this grad J x k and grad J x k plus 1 must be perpendicular to each other. Therefore, r k and r k plus 1 are orthogonal vectors. So, nevertheless we found out that for one particular iteration how to find out alpha k?

(Refer Slide Time: 08:16)



So, the steepest descent method will be start with one guess value x 0 and then do an iteration k is equal to 0 1 2 until it converges compute r r k is equal to b minus Ax k compute alpha k which is r k transpose r k by r k transpose Ar k. Update x k plus 1 is equal to x k plus alpha k r k and check if r k is less than an epsilon is a small value if the residual b minus A I have to see whether b minus ax is equal to 0 if b minus Ax is A very small value.

Then sorry if not a very small value if it is a very small value, then it is iterated if it is not a very small value. Then set k is equal to k plus 1 and go to the 2 and do the try for the convergence of the iterations.

If else if this value is very small r k is a very small number, then you say that iterations are convergence you have reached the right solution. So, this is roughly the steepest descent algorithm what you discussed here and for a symmetric positive definite matrix, we can utilize this b irrespective of the way it has been represented as a diagonal dominant matrix or not only.

Only we have to see that this matrix is symmetric as well as positive definite, then we can utilize this method. This is in general a faster method for symmetric positive definite material this is a faster method, then up to gauss Seidel gauss Seidel SOR or Jacobi method.

However, when we look into this method, we see that if we try to think do a computer program here there are two matrix vector multiplication here A into x and A into r. So, if I have a million by mill 10 to the power 6 by 10 to the power 6 matrix, each of these steps needs multiplication with each element of the matrix with the each component of the vector.

So, in a sense 10 to the power 6 into 10 to the power 6 multiplications are needed, each row needs 10 to the power 6 multiplications and they 10 to the power 6 rows.

So, this particular method needs in gauss Seidel in each iteration you only have to do one Ax multiplication because, xi is equal to b minus sum of ai ijz j at the older value except the diagonal term. So, there is only one matrix vector multiplication there are two matrix vector multiplication.

So, if I try to do a computer program write a computer program out of it though the number of steps will be less than gauss Seidel; however, the calculations will be very high for large matrices. Because they are doing in each iteration the calculations will be very high for the large matrices because, they are doing lot of the Gauss Seidel is on doing only once matrix vector multiplication whereas, the steepest descent will do it twice. So, you need to modify this algorithm.

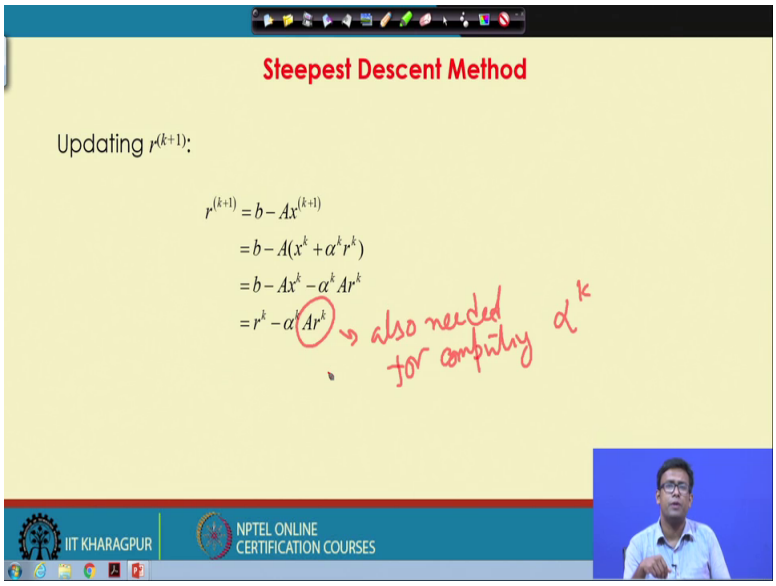
So, you seen that there are two matrix vector multiplications and two vector vector product vector vector is r k transpose r k and r k transpose Ar k which is in again another vector is the vector vector product.

However, they are less time consuming because if there are 10 to the power 6 rows their main only 10 to the power 6 operations are there or matrix vector is 10 to the power 6 into 10 to the power 6 10 to the power 12 operations are there is a very computer computationally costly operation if there is a matrix vector multiplication it is done twice.

So, we have to see how can we reduce the cost of the computation or number of operations, how can we improve the algorithm little bit betters. So, that you can at least avoid one matrix vector multiplication here. So, that for million by million matrix 10 to the power 12 floating point operations are saved. If I can do one matrix vector multiplications here seen to look into some modifications some possible modifications into the steepest descent method before we proposed an algorithm for computer programs.

So, the idea is again if I go back to the previous slide that ah; every why I need to matrix vector multiplication one is needed to find alpha, another is needed to find r and every time we need to do one matrix multiply vector multiplication to find the updated value of r, r k is equal to b minus Ax k once we update x into row matrix vector multiplication here that is actually replaced here.

(Refer Slide Time: 13:31)

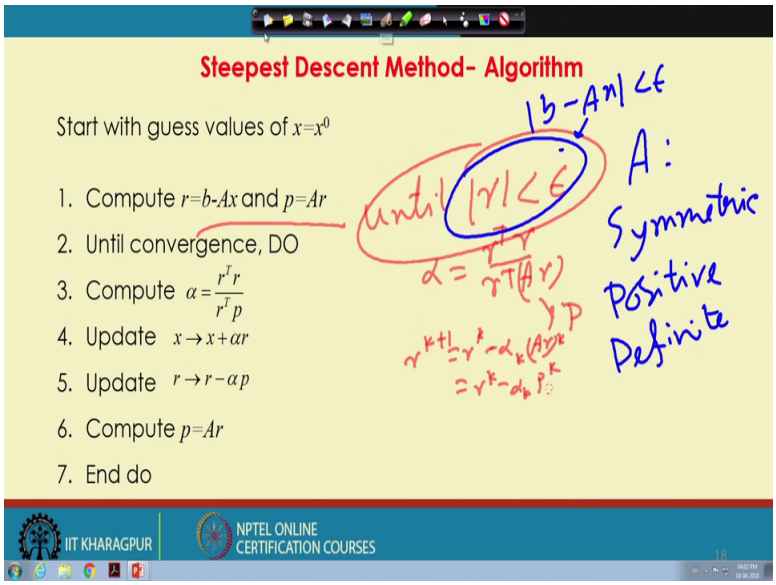


Updating r k plus 1 is equal to r k plus 1 is equal to b minus ax k plus 1, which is b minus ax kb alpha k r k is b minus Ax k alpha k ark that is r k minus alpha k Ar k.

And if I again I go back to the previous slide Ar k is needed also for computing of alpha k plus 1. So, if r k plus 1 can be computed using Ar k, then this is the only matrix vector multiplication which is common both to calculation of r k plus 1 as well as to the computation of alpha k.

And in that light we will try to modify it will see that this is also needed for computing alpha k. So, while computing alpha k will store a Ar k and you will utilize this for updating r k plus 1.

(Refer Slide Time: 14:53)



The final steepest descent algorithm method algorithm will be start with a guess value of x is equal to x 0 compute b minus Ax and r is equal to b minus Ax and define a new variable p which is Ar p will be utilized later. And until convergence; that means, convergence means that until r is less than epsilon, epsilon is a very small number do compute alpha which is r transpose r into r transpose p.

So, alpha is equal to r transpose r r transpose Ar and now Ar has been substituted by p update x is equal to this is because the same variable x which is overwritten as x is equal to x plus alpha. And that is why we written x arrow is equal to x plus alpha r; this is a way to write something which is getting overwritten. Update x is equal to x as x plus alpha update r so, r k plus 1 is equal to r k minus alpha k Ar evaluated at k th level.

So, this is r k minus alpha k p k p k. So, update r is equal to r minus alpha p, compute p is equal to Ar and then if the convergence is not done end do means you again come back here and further with this the new value of p compute alpha update x update r and repeat this loop until you see that r is less than mod r is less than epsilon or convergence has been achieved.

So, this is the steepest descent method algorithm and this is applicable only for A symmetric what happens if A is not symmetric positive definite. Then the problem we are solving if this problem what we are solving here is not finding x is equal to b solving or J is minima. The problem we are solving is not J is minimize not now solving x is equal to

b something different is, if it is not symmetric we are solving half of A plus a transpose x is equal to b a different problem.

So, we are trying to find out minima of something which is not A x is solving x is equal to b. However, where convergence is based on mod r is less than epsilon so, either it will not converge, but if it converges then it will take us to the same solution that mod r is less than epsilon so, x is equal to b.

So, in case the matrix is not symmetric or matrix is little not the asymmetricity is not very high a i k is probably a k i plus a small number. In that case, it still converges, but it takes lot of lot many iteration does not take less number of iterations because we are not using the right solution algorithm for that cases.

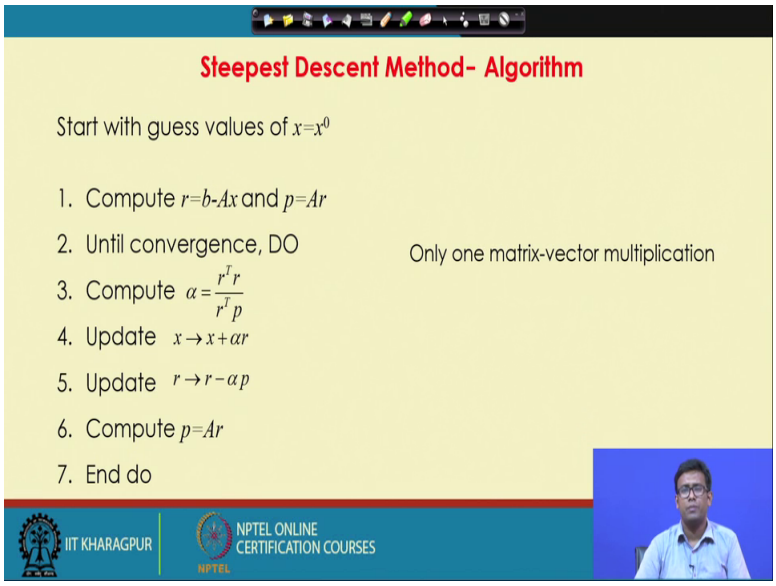
However, if only the matrix is symmetric positive definite then using this method is advantageous in terms that this the formulation is complicated one, program is also probably little complicated than the Gauss Seidel method. However, this is advantageous because it sub times the number of steps are small number of iterations are small some calculations in each relation is comparable to Gauss Seidel.

So, overall computational cost is less if we apply this method, only the if the matrix is symmetric positive definite, then application of this method is worthwhile. Otherwise it might give us the right result because, we are looking for this criteria that mod r less than epsilon; that means, b minus Ax is less than epsilon, we are looking into this right in here. So, it might give if it converges it will give us right it shows that it will take as the x is equal to b location.

But, it can take a very high number of iterations in the matrix is not symmetric positive definite. So, this if I have a symmetric positive definite matrix, this is the algorithm through which we will see later like Gauss Seidel have or SOR how a computer program can be developed. And we can also demonstrate that the number of iterations and as well as computational time is much small than Gauss Seidel or Jacobi or SOR method for a symmetric positive definite matrix.

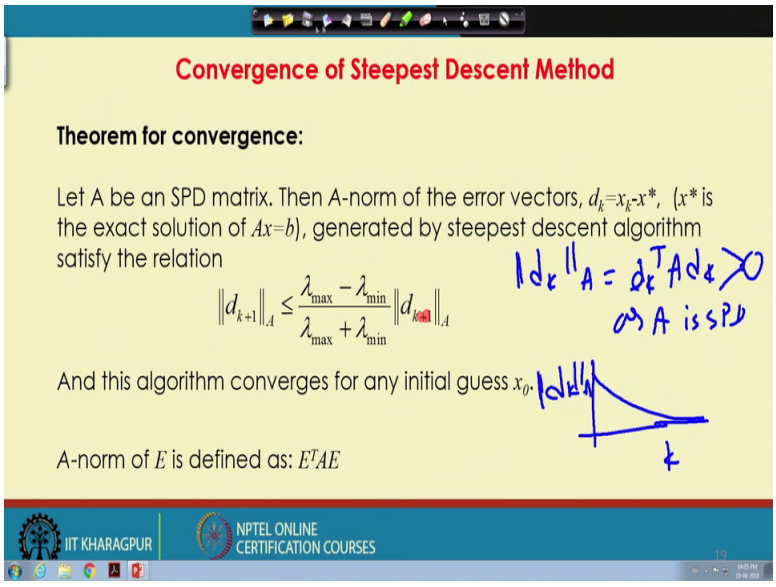
So, we will see in later class that how we can develop a computer program using this and this.

(Refer Slide Time: 20:05)



And interestingly this is only one matrix vector multiplication earlier, we had two matrix vector multiplication. So, one has been reduced there is only one matrix vector multiplication here.

(Refer Slide Time: 20:17)



Now, we need to look into the convergence of this method that is for two d I can give you some visualization that iteratively we if we change the size direction you should approach to the J minima or this method should converge.

Now, there are there are certain theorems and analysis which can show that the error convergence means; the error will we smaller x minus x star is smaller than is there or x k minus extent this error. Will finally, reduce to a very small number given certain conditions, that let a be the theorem for convergence tells that let A be and SPD matrix, that is a symmetric positive definite matrix.

Then the a norm of the error vector which is x k minus x star x star is the exact solution of x is equal to b. Generated by steepest descent algorithm satisfy the relation mod d k plus 1 A norm, A norm of any vector is defined as E transpose AE.

So, a d k plus 1 transpose d k plus so, this is basically d I am sorry issue d k plus 1 transpose A A norm of d k plus 1 or let us take d k it is A norm is basically given as d k transpose Ad k. This norm is satisfy the relation is less than equal to lambda max minus lambda min divided by lambda max plus lambda min of d k A.

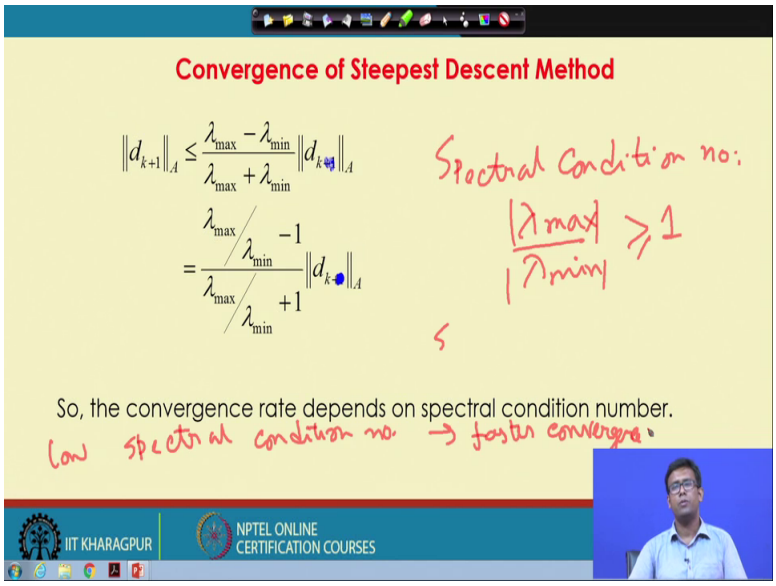
So, if we have any always d k plus 1 is less than d k A, the error x k minus x star at any iteration is always less than the error in the previous iteration. Therefore, it should converge to a finally, this error should be a small value and it should converge to a small small number it should converge to the right solution. However, the rate of convergence here depends on not on the spectral radius rather the maximum and minimum lambda.

So, the relation the difference of the maximum eigenvalue and minimum eigenvalue of this lambda max and lambda min are there is no iteration matrix are the eigenvalues of the matrix A only. So, it depends on the eigenvalues of the matrix A. So, this algorithm and this shows that this algorithm will converge for any guess value x 0, because this is always a number less than 1 therefore, d k plus 1 is always A radius always reducing.

So, finally, the error should go down it will with start with some value it will keep on the magnitude modulus of error is always reduced reducing. So, it will be a small number and this is this is a positive number right because, A is SPD this is always greater than 0, as A is SPD.

So, this number will be greater than 0 and reducing; that means, that if this is d k A and this is k this will asymptotically approach 0. It will never be exactly 0 it will be always greater than 0, but it will be a very small number and then it will be finally, 0. So, this algorithm will converge for any initial guess x 0.

(Refer Slide Time: 24:07)



This is a mistake here and this should be d k, I should correct it in the original notes also, this is d k. So, you can see d k plus 1 in our means lambda max minus lambda min of by lambda max plus lambda min of d k, which is lambda max by lambda min minus 1 by lambda max plus lambda min plus 1 of d k.

And spectral condition number group defined it earlier of any matrix is the ratio of lambda max and lambda min, now we have said that as; small as the condition number. Therefore, lambda max and lambda min are closer it is easier good for matrix solver. So, you are also seeing here and this is always greater than so, basically this is absolute all these are absolute values. So, this is always greater than 1, all this should be in absolute value.

So, this is always greater than equal to 1. So, as this value is close to 1, this number is smaller is 1.000 it is very small number. So, in very few steps the d k plus 1 there should approach 0. So, if convergence spectral radius is the convergence rate depends on the spectral condition number. Spectral condition number low spectral condition number; that means faster convergence.

And when discussing about condition number we have discussed that, if the condition number is small then the convergence is faster. If lambda max and lambda min are closed at the condition number are small convergence is faster and we can see if this number is

smaller will reach the convergence faster. So, low spectral condition number will give a faster convergence here.

So, this is the first time we discuss earlier discussed about condition number, we are seeing one example of condition number in first in the rate of convergence of the matrix particular matrix solver, which is the steepest descent method matrix solver ok.

So, this method will converge; that means, is started with some geometric functional, now we can see that once we have derived the algorithm. This algorithm converge starting with any guess value x 0, this algorithm should take us to the convert solution of x is equal to b and that is only for symmetric positive definite matrix.

Now, the question is that the entire exercise is only devoted for symmetric positive definite matrix. A matrix may not be symmetric in reality we deal with number of cases where we get a symmetric matrices, for example, if we think of a finite difference equation that we are discussing earlier and if we use non uniform grid spacing the matrix will be asymmetric.

So, how can we modify this equation for an asymmetric matrix, as well as in a general case there can be a negative definite matrix, for positive, positive definite matrix that is a solution if the matrix is not positive definite. If there is a negative eigenvalue what is how to solve this matrix? This method does not cover those matrices. Therefore, it is still now restricted only to a closed class of method, a small class of method we use symmetric positive definite.

Now, next goal will be if we can extend this method for general matrices, which are not symmetric and non positive definite matrix. And what we will discuss for that there is called general projection methods. Instead of having a method for searching to gradients such approaching the minimum value of a functional will see a more generalized method where probably the functional is little different we are trying to find out minima of some other function. But solving a matrix which is not symmetric or which may not be positive definite also, extension of this method will take us to general projection methods, which can solve a larger class of matrices we will so, see that in the next classes.

Thank you.