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

Lecture - 43
Introduction to Krylov Subspace Methods

Welcome, in last few classes we started discussing over iterative methods. And I have given you some introduction to basic iterative methods and also worked on their convergence analyses which are Gauss-Seidel and Jacobi based methods. And later we have seen that convergence rate of these methods can be improved using certain methodology called successive over relaxation.

However, what we have seen there is that these methods are restricted only for diagonally dominant or irreducibly diagonally dominant matrices. That is the diagonal term in it is absolute value must be greater than equal to or sum of all the of diagonal terms in their absolute values. And at least for one row the diagonal terms absolute value is greater than not greater than equal to is greater than some of the absolute values of all the off diagonal terms. Only for these cases the matrices Jacobi and Gauss-Seidel or successive over relaxation can give a solution.

And we have seen that if we can take diagonally dominant matrix and if we do a low permutation the solution still remains same. However the matrix changes, but Jacobi or Gauss-Seidel fails for that. So, there is a particular restriction and Jacobian was settled. And the convergence rate depends on what is a maximum eigenvalue or the spectral radius of the iteration matrix g. That cannot be greater than 1, if it is less than 1 then only the method will converge. And that shows that it will converge for any diagonally dominant or irreducibility matrix.

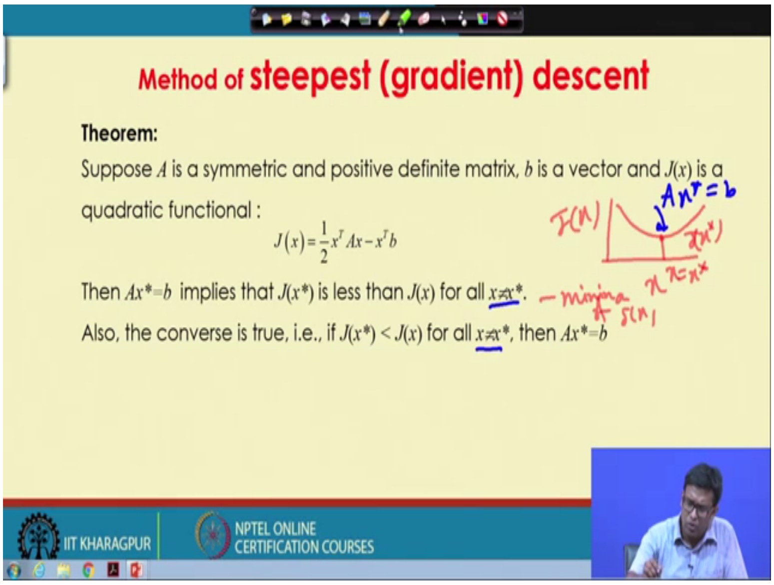
However if the maximum eigenvalue is large is less than 1, but still large the convergence rate will be slow. And we can only improve the convergence rate using successive over relaxation technique. Even there is a optimum value of omega or over relaxation factor based on which you can get highest convergence rate and if we will later discuss about a number of iterative methods. If we compare success like optimum omega or successive over relaxation Gauss-Seidel with a faster iterative solution technology; we will see that Gauss-Seidel or SOR Gauss-Seidel is still way slower than this faster iterative solvers, those solvers we will discuss.

So, the basic problem with Gauss-Seidel and Jacobi iterations are in two fold. One is that they are restricted only for diagonally dominant or irreducibly diagonally dominant matrix. And another issue is that their convergence rate is limited by the maximum value of the spectral or maximum value of the maximum eigenvalue of the iteration matrix the spectral radius of iteration matrix. If it is large even using, successive over relaxation we cannot increase it to increase the convergence rate to a very high extent it, there is an optimum omega for which it will be maximum.

So, there is some restriction on the convergence rate or there is some restriction on the number of iterations that is to be performed for solving a matrix using Gauss-Seidel or Jacobi or a SOR Gauss-Seidel. So, now, we think of looking into faster solvers and that is the very importance of this particular course; is that how can we solve large equation system using iterative methods which give us first solution. So, when we start looking into faster solution we have seen that beyond SOR Gauss-Seidel is restricted so, you look for some other solution techniques. And this solutions techniques will not be direct solution techniques; that means, that I will start with any arbitrary x 0 and only use that particular equation A x is equal to b substitute gausses value x 0 and update x.

These direct iterative techniques are restricted in terms of their applicability for diagonally dominant matrix as well as in terms of the convergence rate or speed of the solution. So, you look for some other iterative techniques and what we start discussion of this other is from here we will start discussion on other iterative techniques. And a class of techniques named Krylov subspace will follow later from this discussion only. So, we will start discussion with projection based iterative methods and this with the particular method I will try to discuss in this one or two sessions is steepest descent method.

(Refer Slide Time: 05:10)



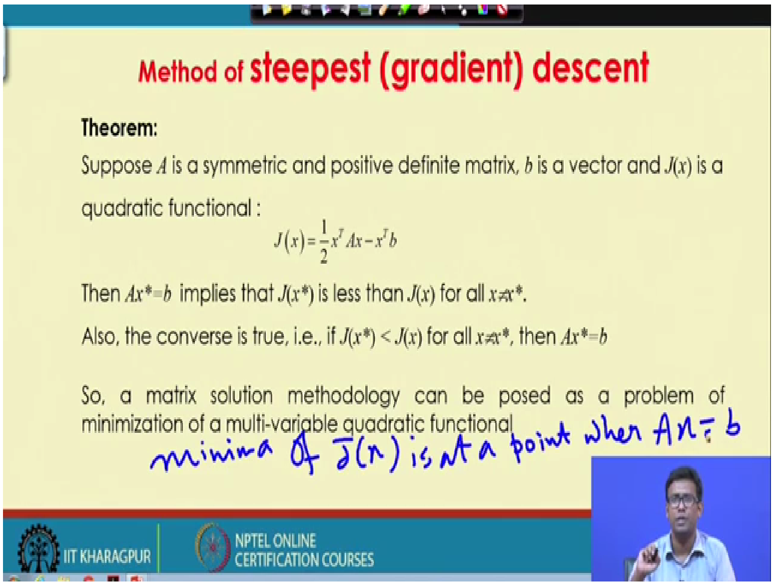
So, it is also known as gradient descent or gradient search method. So, if we look into this method it starts with the theorem; suppose there is a symmetric and positive definite matrix A and b is a vector. And now we define a quadratic functional J a quadratic functional means it is basically a quadratic function; in which all the components of the vector x is associated, x is a vector of same order as b.

So, this quadratic function is defined as Jx is half x transpose Ax minus x transpose b. If we can define this function J x then A x star is equal to b will imply that J x star for any value of x star Jx star is less than J x for all x which is not is equal to x star. So, this basically says that if we can define a functional half x transpose A x minus x transpose b we can say that Ax star is equal to b where we will get Ax star is equal to b for that particular x value of x x star Jx star is less than all Jx star Jx star is the minima.

So, if we think x to be a single value a vector of dimension 1. And we write that this is x and this is Jx so this is the Jx functional. And here at x is equal to x star J x star is less than any other Jx this is the minima or we can say that this is minima of Jx. And at this particular x is equal to x star location, Ax star is equal to b. Or Ax star is equal to b is a solution Ax star is equal to b gives us the solution is x star is only where Jx is minima. Instead of solving x star is equal to B now we can try to find out minima of Jx that that is the main philosophy behind this method.

The converse is true; that means, if we find out a minima of Jx that is Jx star is equal to Jx for all x which is not x is equal to x star. So, for one particular x star I get a value Jx star which is minima then at that x star x star is equal to b. So, when we write for all x naught is equal to x star all x naught is equal to x star that says that there is one particular extract for which this is solvable. So, this equation system; obviously, has a unique solution.

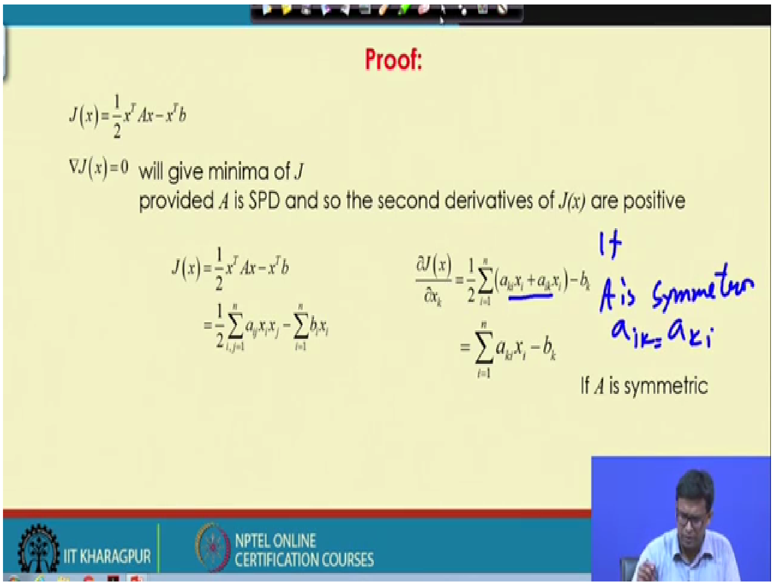
(Refer Slide Time: 08:02)



So, what this theorem mean; that a matrix solution methodology Ax is equal to b, I have to solve this. This can be replaced by finding out minima of Jx, if we can define a function J x is equal to half of x transpose Ax minus x transpose b and try find minima of Jx. The minima is a point where Ax star is equal to b is satisfied. So, matrix solution methodology can be posed as a problem of minimization of a multivariable quadratic function.

Instead of solving Ax star is equal to b will x is equal to b we will solve we will find the minima of Jx. And they that is why this is sometime called also search algorithm meaning gradient search algorithm we will look into the gradient method later because minima is associated with radiant innocence. So, the idea is that instead of solving x is equal to b try to find out minima of Jx. And then there will there will be an iterative method for doing this.

(Refer Slide Time: 09:18)

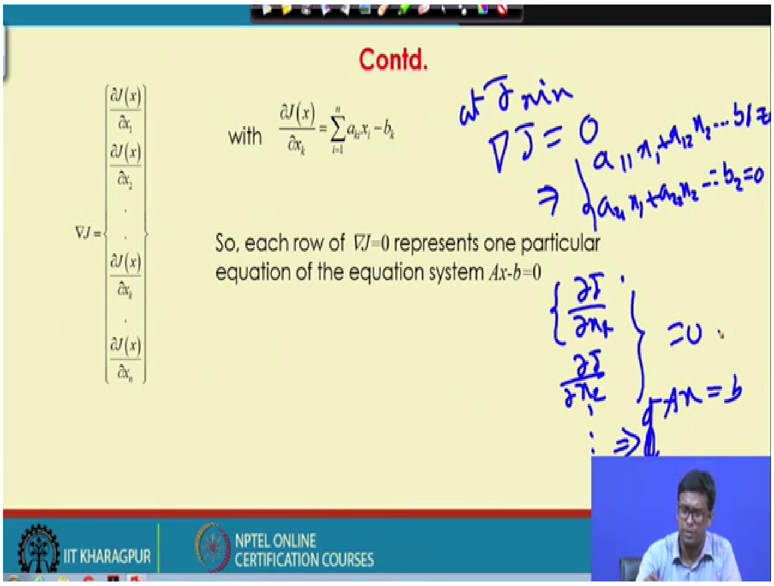


So, now I we got the theorem that Jx is equal to x transpose half of x transpose x minus x transpose b. Minima of Jx so the theorem basically says minima of Jx is at a point where Ax is equal to b. So, if we can find out the location where Jx is minima that will point to a vector for which Ax is equal to b. So, we will look into the proof of the theorem; so it starts with Jx is equal to half of x transpose x by Ax minus x transpose b. And for minima we must have the gradient of because there is a multivariable function gradient of J is equal to 0. The first order derivative is 0 and provided a is symmetric positive definite matrix.

So, that at least if a is positive definite the second derivative of Jx so Nablus squared J will be positive and we will see it later. So, if minima gradient of J is equal to 0 and a is positive definite matrix, then this will indicate the minima of Jx. So, now, we write Jx is equal to half of x transpose Ax minus x transpose b if we expand it this term is particular low and of a and it is multiplied with x I and x J and then all these terms are summed up. So, these are scalars so this is half of I is equal to 1 to n J is equal to 1 to n sum of a ij x i x j, minus sum of I is equal to 1 to n bi x I.

And if I take it is derivative with just the to kth component of J we get half of I is equal to 1 to n a ki x I plus ai k xk minus b k. Now if a symmetric this has to be noted carefully A is if a is symmetric a I k is equal to a ki. So, this becomes I is equal to one to n twice aki half of twice aki so aki x I minus bi ok.

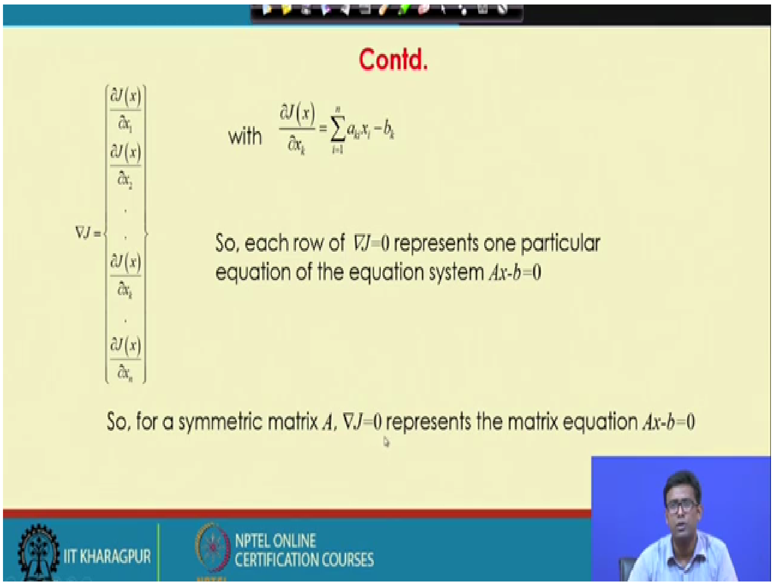
(Refer Slide Time: 12:14)



So, for if I write gradient of J that will contain all the vector del J del x 1 del J del x 2 to del J del x n. And each of this delta del x term is basically I is equal to 1 to k ki x I minus bk one particular value. So, they when we will find for minima our requirement was grad J is equal to 0 at J minima. So, grad J is equal to 0 each term will be delta del x 1 is equal to 0 del J del x 2 is equal to 0 up to del J del xn is equal to 0 which is sum of a 1 I x I minus b 1 is equal to 0.

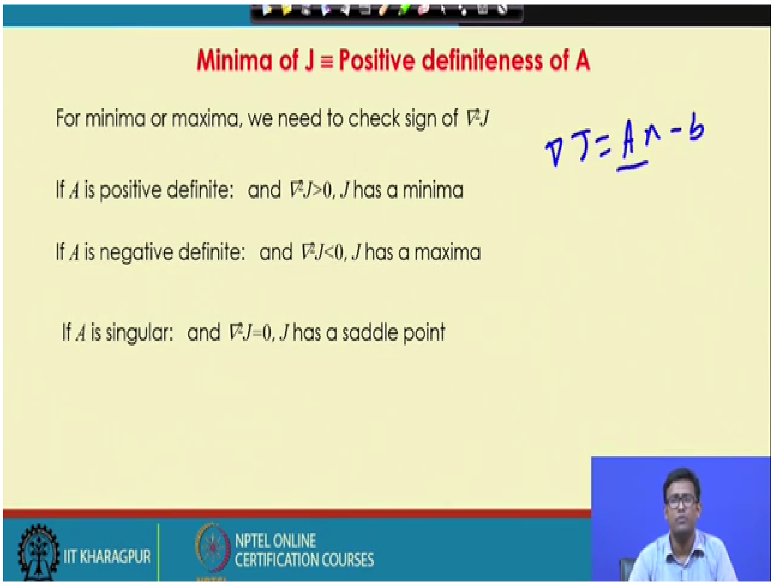
The next term is a 2 I sum of a 2 I x I my minus b 2 is equal to 0 so on. So, each of the each row of del J is equal to 0 each row of this matrix del J is equal to 0 represents one particular equation which is belongs to the equation system Ax minus b is equal to 0. So, this will finally, give me that a 1 1 x 1 plus a 1 2 x 2 minus b 1 is equal to 0. Then a 2 1 x 1 plus a 2 2 x 2 up to minus b 2 is equal to 0 so on. So, this del J del x is equal to 0 or this equation system del J del xk or del J del x 1 del J del x 2 is equal to 0 this equation system will give me nothing, but Ax is equal to b or b minus Ax is equal to 0 which is Ax is equal to b.

(Refer Slide Time: 14:19)



So, for and this is only for symmetric matrix because if it is a non symmetric matrix then I have to write half of a ki plus half of a ikx x I is equal to 0. So, for a symmetric matrix a gradient of J is equal to 0 actually represents the matrix equation Ax minus b is equal to 0.

(Refer Slide Time: 14:44)



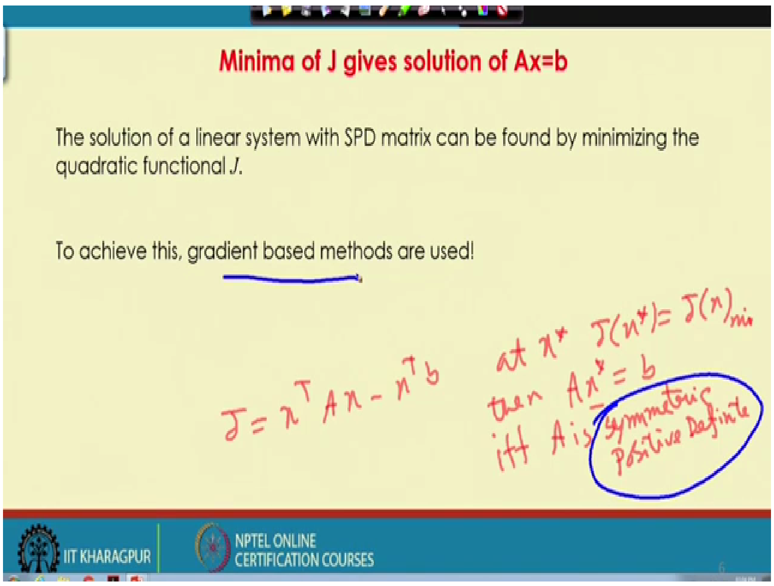
So, for minima or maxima that has to be decided from the sign of Nablus square J is equal to 0 and J is equal to. So, we got gradient of J is equal to Ax minus b right. So, second derivative of g or Laplacian of J will be the matrix norm of matrix A only. So, if

A is positive definite then Nablus square of J is greater than 0 the matrix norm of a is greater than 0 and J has a minima. If J is negative definite then J has a maxima and incase J is singular, then Nablus squared J is equal to 0 or J has saddle point. So, you can ponder upon the question that how a saddle point will be in a multi dimensional space. However, like we call it and in a 2 d space we get an inflection point if second derivative is 0 so, it is something like that.

So, if A is symmetric then we got gradient of J is equal to 0 repress similar to x is equal to B. And if a singular positive a is positive definite then Nablus square J is greater than 0. So, at gradient of J is equal to 0 Nablus square J is also greater than 0; that means, J is minima and they are Ax is equal to b. So, we have to look into the case where A is the positive definite matrix and Nablus square J is greater than 0; that means, J has A minima.

So, is a positive definite matrix and; that means, J has a minima at that particular location where del J del x is equal to 0. And if a is a symmetric matrix then this particular location where del gradient of J is equal to 0 represents the matrix equation Ax minus b is equal to 0. So, if A is symmetric and positive definite then Ax minus b is at a particular location where J has a J is x transpose Ax minus x transpose b as the minima. And that proves the theorem where you start started with.

(Refer Slide Time: 17:07)



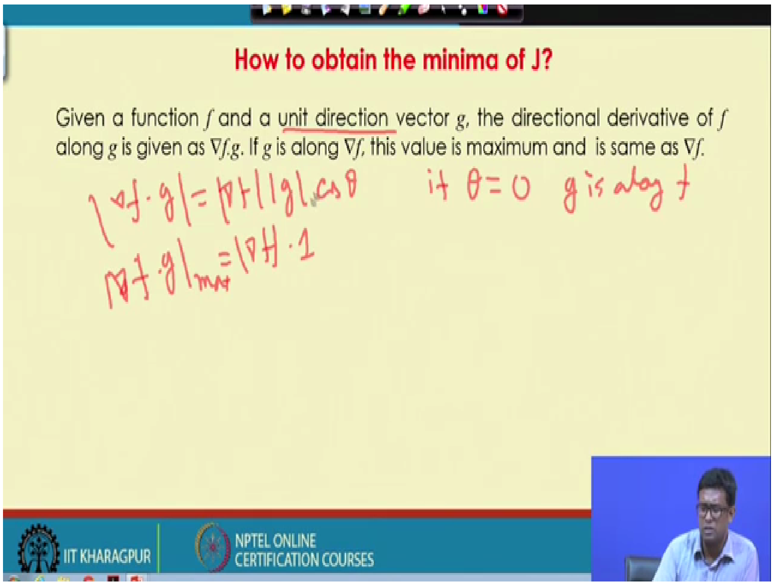
And further tells us that minima of J will give us a solution of Ax minus b The solution of a linear system with symmetric positive definite matrix can be found by minimizing the quadratic functional J. Or we can write that J is equal to x transpose may be I will write it here J is equal to x transpose Ax minus x transpose b; at x star J x star is equal to Jx min, then Ax star is equal to b.

So, if I can find out an x star for which J is minimal that particular a x star will give me the solution of Ax star is equal to b. So, instead of solving Ax star is equal to b we can now think of finding minima of Jx star. And this is this is if and only if A is symmetric. So, this particular methodology we are discussing which we are only discussing this for symmetric positive definite. So, later we should also focus how we can improve this to a general matrix which may not be symmetric mineral, because it is definite how we can improve this methodology for that particular case.

However at this stage we are only considering symmetric positive definite matrix and we observed that solving the matrix equation can be substituted by finding out minima of a quadratic functional. So, our idea will be now because we have seen solving matrix equation by directly trading methods and at the beginning of the class I tried to list out the issues with the direct iterative methods or whatever where are they restricted. Now our goal will be instead of solving this using direct iterative methods we will use an iterative method for finding out minima of J.

Instead of solving x star is equal to b we will try to find out how we can achieve minima of J and the location where we will achieve minima J will say that this is the location where Ax star is an Ax is equal to b is also satisfied. So, we will say that the solution now is posed as a problem of finding minima. And to achieve these in and iterative methods we will use something called a gradient based method. The method we will use called a gradient based method and then we will explore this method in the subsequent slides.

(Refer Slide Time: 20:17)



Now, the question is how to find the minima of J? In this class at least in this class we have always discussed about matrices, how to do work with matrices, how to find out solution of matrix, when we can have solutions etcetera. But now we are in our little different paradigm, where you have to find out minima of a functional. It is little different than what we are doing in the matrix of equations. However, if you can remember the idea of symmetric positive definite matrix when we discussed that we have looked into minima of a function and etcetera.

So, now, our question is how to find this minima in practical application. So, given when for example, when I am trying to find out a minima I have a function and I have to see where the function is minimum. So, I will start with the value of the function. If I am going in the right direction I am approaching towards minimum of the function because it is a continuous function which you are discussing with, I will see that the value of the function and is reducing. It will reduce until it reaches a minima and then it will increase.

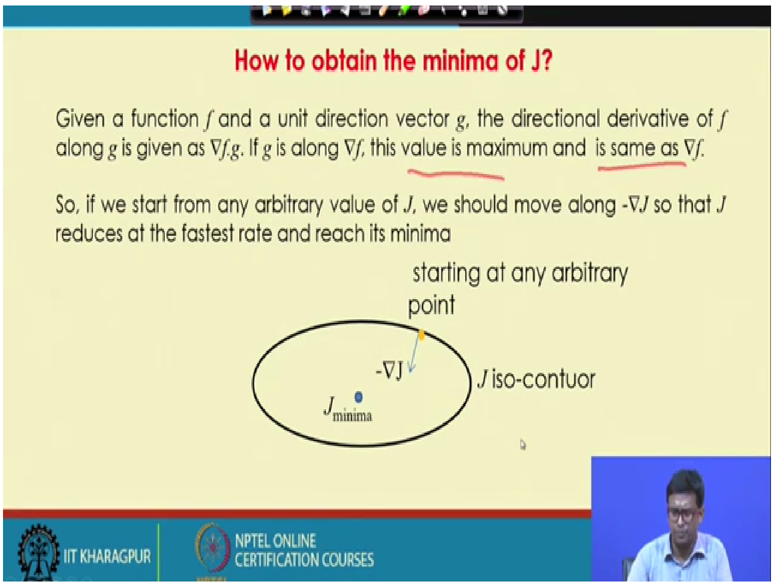
So, if I am going in an iterative method say I am going in a trial and error method, I found out a value a particular value of J. And next it I find out another value of J this is reducing then I am approaching towards the minima. When I have approached the minimal del J del x is equal to 0 so at that particular location if I change my primary variable little bit there is no change in the value. But later I will see that the value of J is increasing with changing the primary variable. So, I will start with a value and I will see if the value is reducing then I am approaching towards minima.

And in order to reach minima first I will should see that the value of the functional is also reducing in the first test manner. So, we need to see when the value is reducing, and how we can reduce the value as fast as possible. So, look little bit into vector calculus given a function f and a unit direction vector g the directional derivative of f along g is given by gradient of f dot g. And this is directional derivative means rate of change of f along the direction g. If g is along gradient of f then g is an unit vector. So, gradient of way if dot and unit vector in that direction and that is the maximum value of gradient of f dot g and this value is same as gradient of f.

So, so this is probably straightforward that gradient of f dot g is basically it is magnitude is gradient of f into g into cos theta. If theta is equal to 0; that means, g is along f and mod g g is a unit direction vector so mod g is equal to 1. So, this is gradient of f into 1 and this is the maximum value of f dot g because if theta is nonzero then there is this is or cos theta is always less than 1 and this is a smaller value..

So, given a function f and a unit direction vector g the directional derivative of f along g will is give grad f dot g and grad f is the grad f is the maximum value of reduction of change of f along any of the direction. It is maximum along the direction of grad f and the maximum value is grad f.

(Refer Slide Time: 24:15)

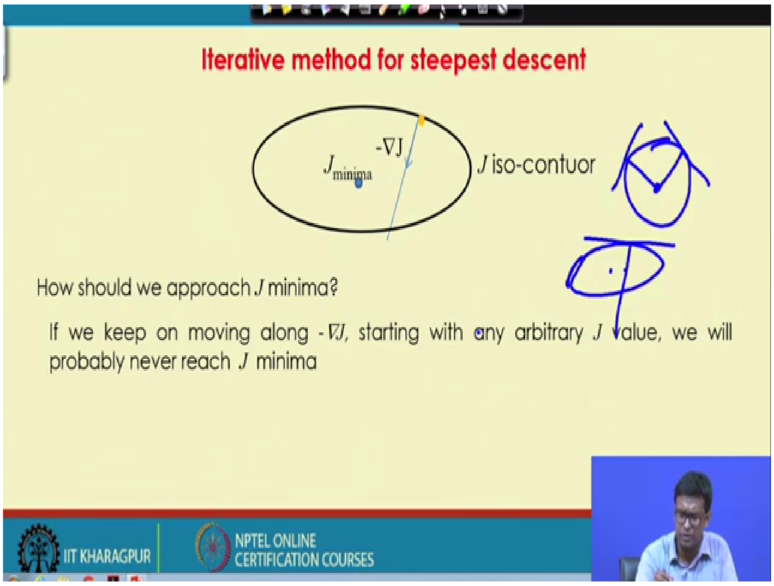


So, if we start from an arbitrary value of J we should move along minus grad J. So, that the reduction is if we move along grad J there rate is maximum. So, if we go along minus grad J the rate is minimal or the reduction is maximum. So, if we start we will start from an arbitrary value of J and we will move along gradient of minus of gradient J. So, that J reduces at the fastest rate and it will reach it is minima.

So, if there is a J minima in the in this 2 d space anywhere and this is a Ji. So, contour now I start from one particular location here because there is an iterative method I have to start with something I start from this method and I have to reach J minima. So, how should we go? I will see that in which direction J is reducing fastest and the reduction is along gradient of minus J. So, I will move along from starting from some value x here I will change the values along this line and see where the value is minima because it will reduce fastest in this particular direction.

Now when I am in this location it is reducing fastest, but when I came to here then J has been changed. So, minus grad J is probably different it is not reducing in the fastest way. So, it might here it is not reducing in the fastest it might not ever dj minima will see later so you to use an iterative method for that. However if we start from one particular point I have to go along minus grad J so the reduction is fastest.

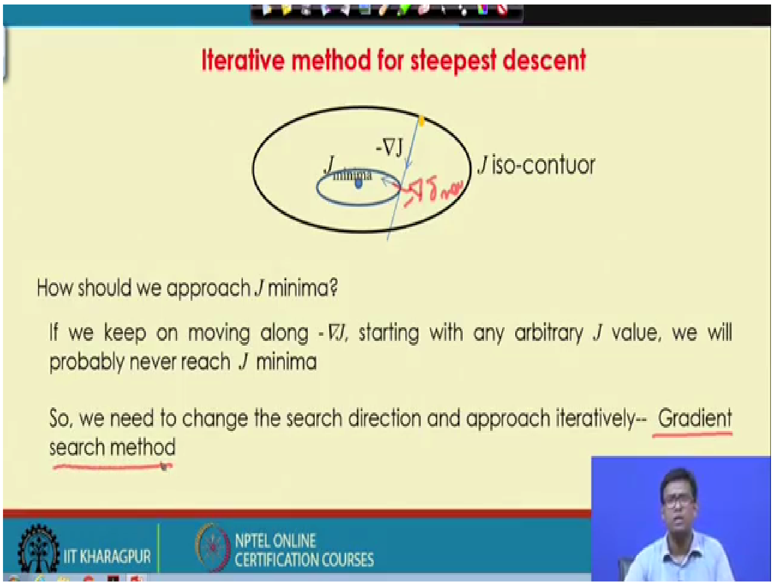
(Refer Slide Time: 26:01)



So, and the question is so I know that I started from an arbitrary point I am going along minus grad J and I have to count come to J minima. How should we reach that? So, if we move along this particular line we might not be ever able to reach the J minima. Because along this line J is first reducing and then we can see if we keep on minus grad J we can see that it is clearly missing J minima starting from any arbitrary J value we will possibly never reach J minima. We will reach J minima in only in one case that is the functional of J is a perfect circle. So, if we start from any point the normal should take me to the center.

But if it is not a circle we if it is an ellipse and I start from somewhere here I will the minima is here I will never reach the minima go somewhere else. So, it probably never reach minima. However, we can see this we will start we if we draw this is the J iso contour; that means, J constant here. So, minus grad J is normal to that. So, we can see that up to certain level will J will reduce because, J is initially reducing here. And there will be local minima on this, we will probably go up to there and then see how we can again move.

(Refer Slide Time: 27:46)



So, we need to change the search direction and approach iteratively we will see we will go along minus grad J and reach to one particular level. And then again I will calculate the new minus grad J here and move along this direction. I will probably again reach somewhere from there I have to again calculate minus J grad say and move along that direction that is the iteration technique necessarily here. So, you need to change the search direction and approach iteratively. And this method is called the Gradient search method. In the next class we will look into more detail of the gradient search method.

Thank you.