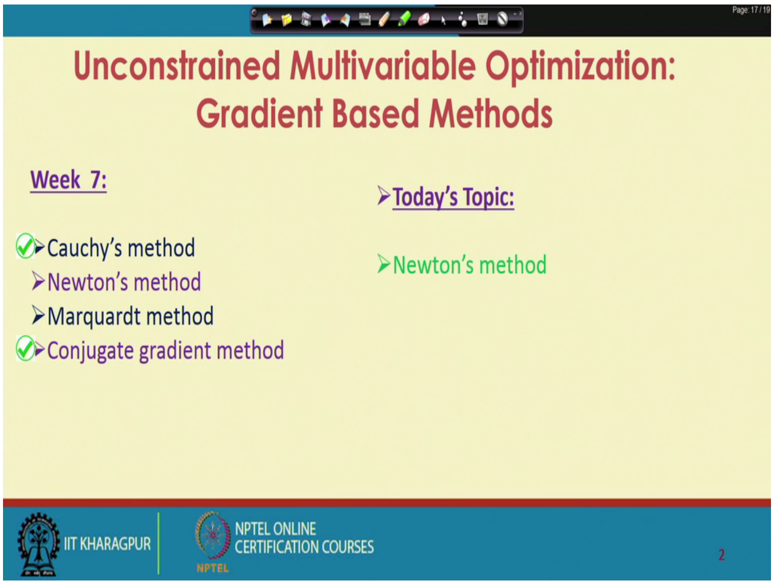
Optimization in Chemical Engineering
Prof. Debasis Sarkar
Department of Chemical Engineering
Indian Institute of Technology, Kharagpur

Lecture – 33
Unconstrained Multivariable Optimization: Gradient Based Methods (Contd.)

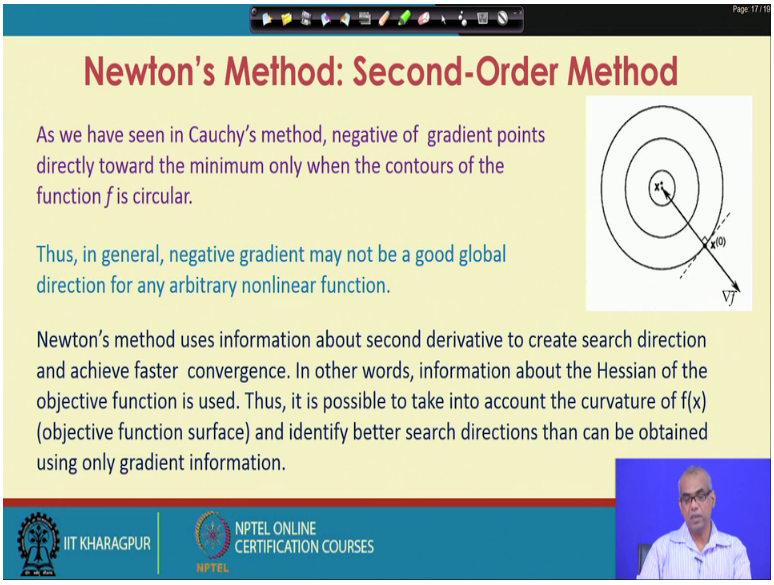
Welcome to lecture 33, this is week 7 and we are talking about Gradient Based Methods for Unconstrained Multivariable Optimization. In previous lectures, we have talked about Cauchy's steepest descent method and conjugate gradient method. In this lecture we will talk about a very popular gradient based method for unconstrained multivariable optimization namely Newton's method when converges Newton's method converges most rapidly.

(Refer Slide Time: 00:49)



We have seen Newton Raphson method, when you talked about single variable optimization or unconstrained single variable optimizations. We will see how the Newton methods were when we have unconstrained multivariable functions.

(Refer Slide Time: 01:19)



We have seen in Cauchy's method that the negative of gradient points directly towards the minimum only when the contours of the function is circular.

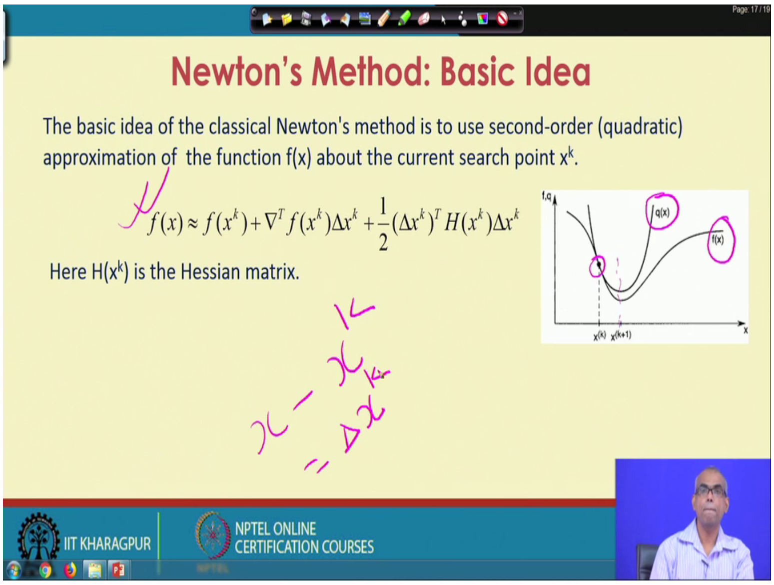
So, we have seen that if the objective function has circular contours then, the negative of the gradient will point directly towards the minimum. So, this will happen with quadratic functions such as x square plus y square. So, if you have an objective function like f equal to x square plus y square, the contours will be perfectly circular and the negative of the gradient if you consider the search direction, the search direction will point directly towards the minimum.

But if you have an objective function like say x square plus a y square f equal to x square plus a y square. it is also quadratic, but here depending on the value of a your contours may be elongated and in that case if you start from any arbitrary point the negative of the gradient will not in general direct towards the minimum part. So, in general negative gradient may not be a good global direction for any arbitrary non-linear function.

Newton's method uses information about second derivative to create search direction and achieve faster convergence. In other words, information about the Hessian of the objective function is used, thus it is possible to take into account the curvature of the objective function surface and identify better search direction then can be obtained using only gradient information. So, the second derivative tells us about the curvature of the function.

So, Newton's method uses Hessian information or the second order information to obtain a search direction and a better search direction is obtained and therefore, we obtain faster convergence, but we will see later then this is true only when we start very close to the optimal point. So, Newton's method is a second order method because it makes use of second order information that is Hessian of the objective function is used.

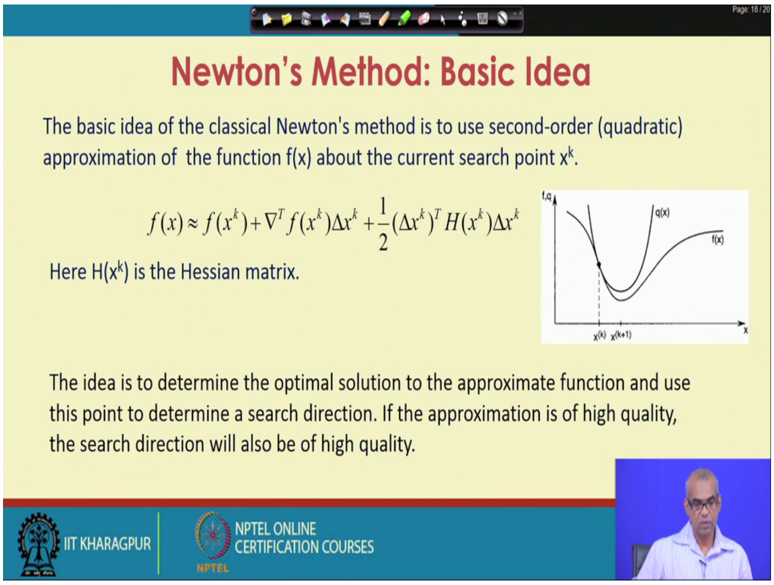
(Refer Slide Time: 04:37)



The basic idea of the classical Newton's method is to use second order or quadratic approximation of the objective function about the current search point xk. So, if you look at the figure, this is the function fx, this is the point xk and at point xk about the point xk, I approximate the objective function fx by a quadratic function qx. Look at here, the point xk is such that the quadratic approximation qx and the function fx has more or less same minimum point.

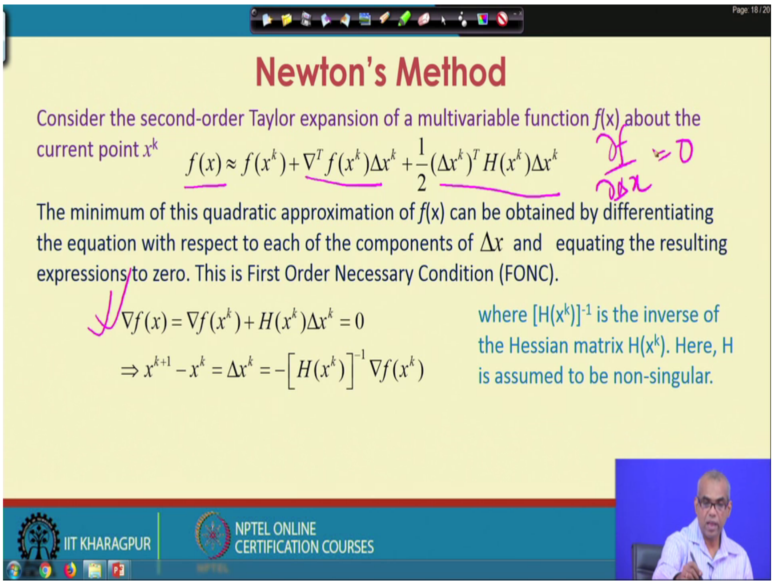
So, the function fx can have a quadratic approximation as shown, fx equal to f of xk plus gradient of fxk into del xk plus half del xk transpose H xk del xk; that means, you expand fx in terms of Taylor series about the point xk. So, x minus xk is taken as del x k.

(Refer Slide Time: 06:51)



So, here H is the Hessian matrix. The idea is to determine the optimal solution to the approximate function and use this point to determine a search direction. If the approximation is of high quality the search direction will also be of high quality.

(Refer Slide Time: 07:13)



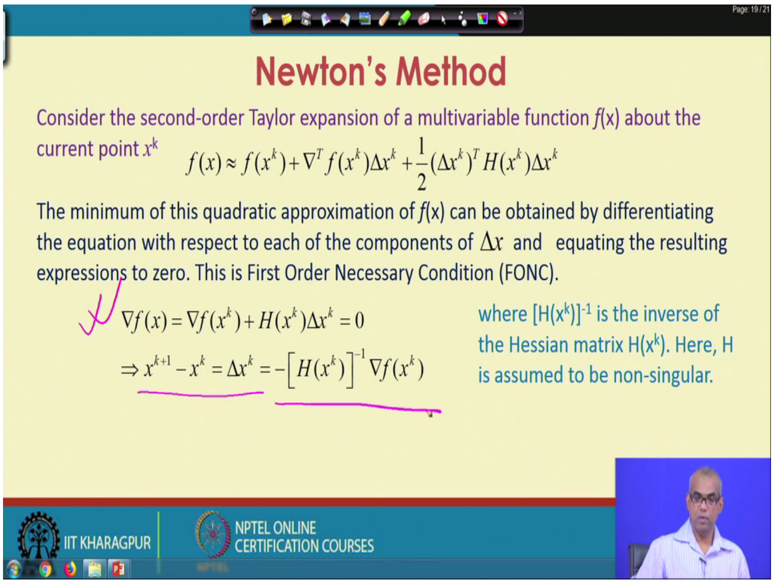
So, consider the second order Taylor expansion of a multivariable function fx about current point xk and you obtain the approximation as fx equal to f of xk plus gradient of f at xk into del xk plus half into del xk into H at del xk at H into xk into del xk. So, this is

Taylor series expansion of a multivariable function fx about the current point xk and you have retained only up to second order terms.

Now, the minimum of this quadratic approximation of fx can be obtained by differentiating the equation with respect to each of the components of del x and equating the resulting expression to 0. Note that this is first order necessary condition. So, basically what we are saying is that, we can take del f del del x equal to 0. So, if I do that I will get this, gradient of fx is equal to gradient of f xk plus H at xk into del xk.

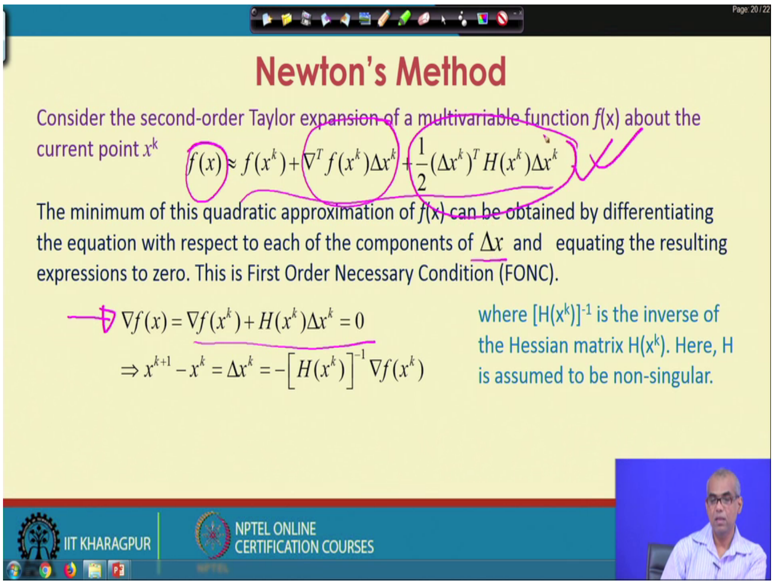
So, basically what I am doing is del f del del x equal to 0.

(Refer Slide Time: 09:25)



So, in that case you will get this expression and this is a first order necessary condition. So, this we will we will set equal to 0 and you can solve for del xk which is nothing, but xk plus 1 minus xk, which is minus of Hessian inverse into gradient both evaluated at xk. So, inverse of the Hessian matrix H is required we assume Hessian matrix H to be nonsingular. So, let us look at one more time.

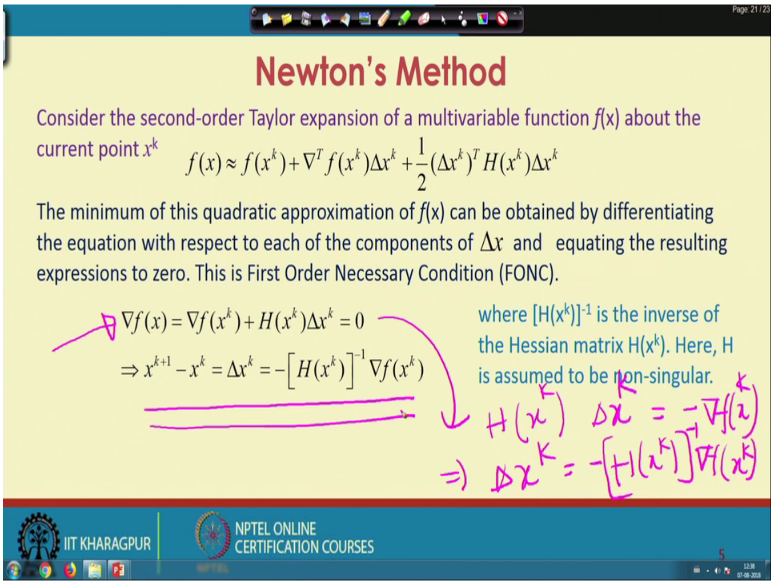
(Refer Slide Time: 10:17)



This is the Taylor series expansion of a multivariable function about current point xk. So, this is a quadratic approximation of any non-linear general function fx. Now I want to find out the minimum of this quadratic function, which is an approximation of the original function fx. So, I make use of first order necessary condition.

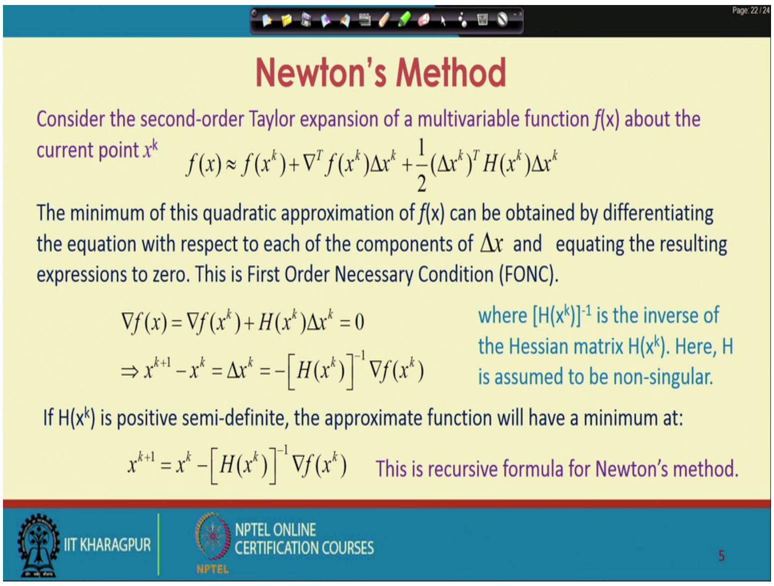
So, I take the gradient and set that equal to 0; that means, we have to differentiate the function with respect to each of the component of del x, we have to differentiate the function with respect to each of the component of delta x and we have to set the resulting expressions to 0. If we do that we get the gradient of fx as gradient of f evaluated at xk, which comes from here and Hessian evaluated xk multiplied by delta xk, which comes from here equal to 0.

(Refer Slide Time: 12:05)



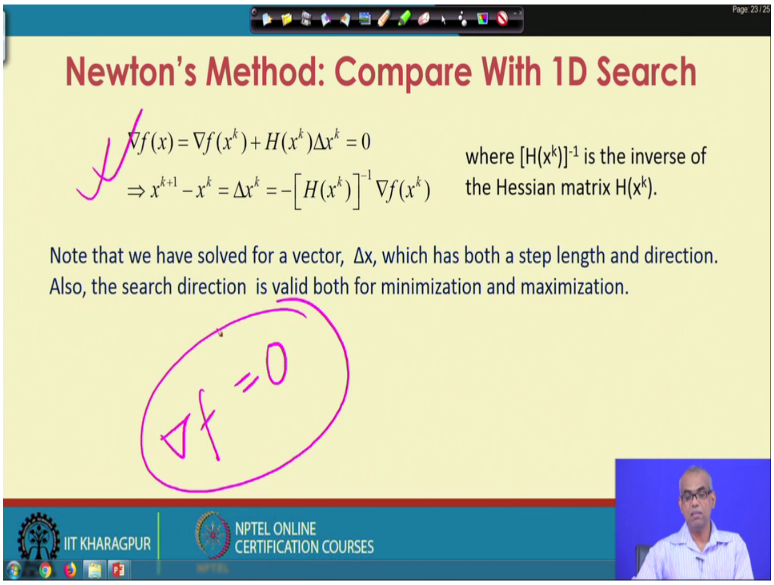
So, now you solve this for delta xk delta xk nothing, but xk plus 1 minus xk. So, from here you get H at xk delta xk is equal to minus delta f at xk. So, delta xk is obtained as minus H xk inverse delta f xk. So, this expression is obtained which is the recursive formula for the increment of the current estimates of the minimum.

(Refer Slide Time: 13:17)



So, if the Hessian evaluated xk is positive semi definite the approximate function will have a minimum at xk plus 1 equal to xk minus H inverse into gradient. So, this is the recursive formula for Newton's method.

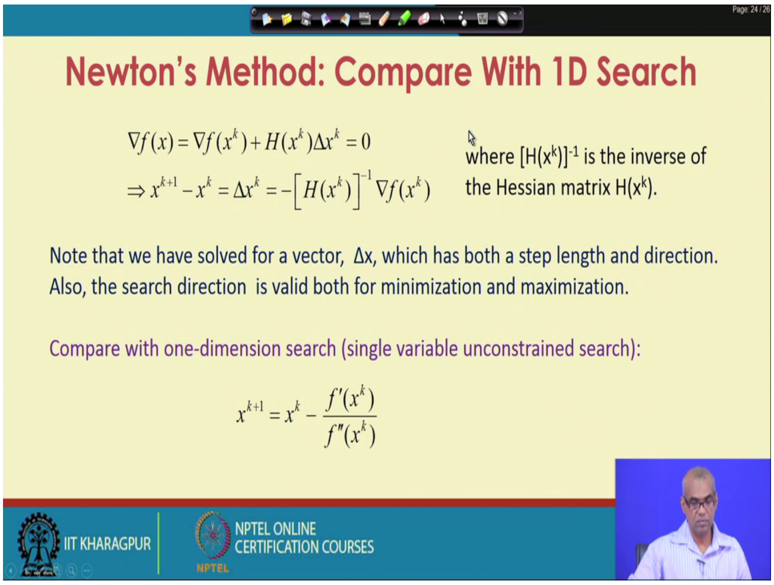
(Refer Slide Time: 13:45)



So, this is what we just seen the update rule in case of Newton's method, which is delta xk equal to xk plus 1 minus xk equal to minus Hessian inverse into gradient of f, both evaluated at the current point xk.

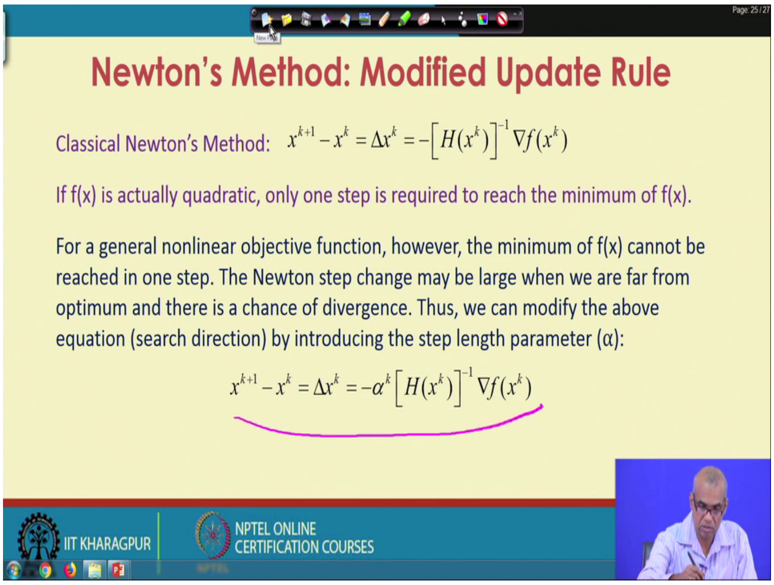
Note that we have solved for a vector delta x or delta xk, which has both a step length and direction. So, delta x has both step length and direction also the search direction is valid both for minimization and maximization, because for both the cases the gradient of f equal to 0 the first order necessary condition we have, remember we have obtained this equation by setting equal to 0. Now, this is the first order necessary condition for optimality and first order necessary condition for minimization as well as maximization.

(Refer Slide Time: 15:31)



So, the search direction is valid for both minimization and maximization. You compare this, what we learned for one dimensional search? So, that was xk plus 1 equal to xk minus first derivative divided by second derivative. So, here you have gradient and Hessian for multivariable.

(Refer Slide Time: 16:03)



.

Now for the classical Newton's method update rule is xk plus 1 minus xk equal to minus Hessian inverse into gradient of the objective function.

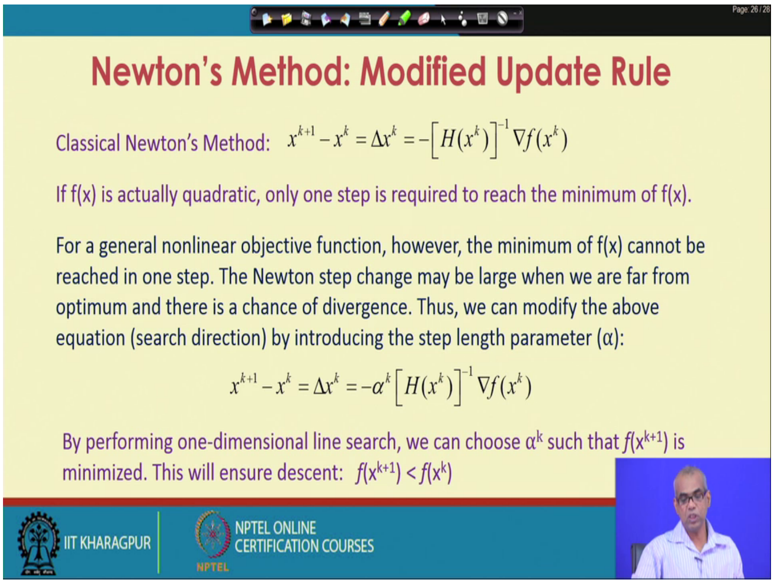
Now, if the objective function fx is actually quadratic only one step is required to reach the minimum point of the objective function. This is obvious because update rule for the classical Newton's method is obtained by considering a quadratic approximation of the objective function. Now if the objective function given itself, is quadratic function then the approximation is exact. So, update rule will give you the minimum in one step.

But for a general non-linear objective function, the minimum of fx cannot be reached in one step. The minimum of fx in one step can be achieved only for a quadratic function not for any general non-linear objective function. The Newton step change has obtained from the classical Newton's method update rule may be large when you are far from optimum and thus there is a chance of divergence. Thus, we can modify the search direction by introducing a statement parameter alpha.

So, instead of taking xk plus 1 minus xk equal to minus Hessian inverse into gradient of f, I can introduce a step length parameter alpha and can write as, xk plus 1 minus xk equal to minus alpha k into Hessian inverse into gradient of the objective function. So, the only difference between the classical Newton's update rule and this update rule is this that, we have added a step length parameter.

So, this is a simple modification to the classical Newton's method sometimes also known as modified Newton's method.

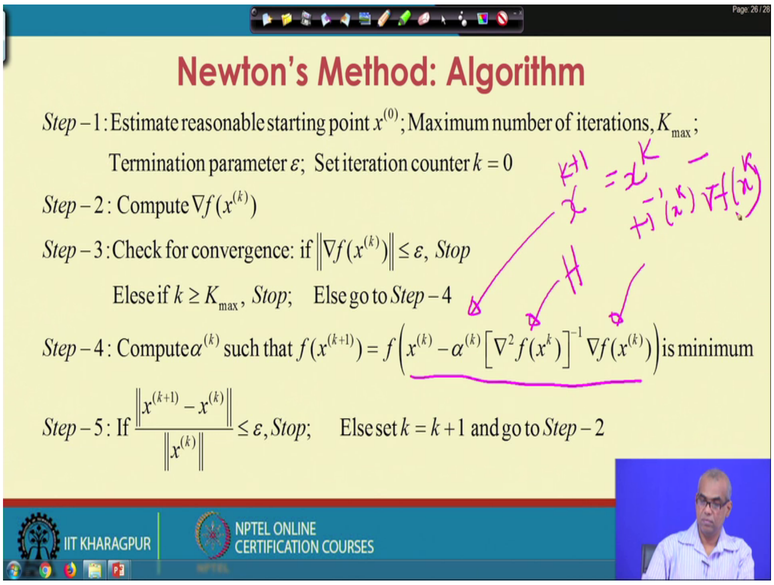
(Refer Slide Time: 18:39)



By performing one dimensional line search we can choose alpha k such that f of xk plus 1 is minimized again remember the 1 dimensional line search method we can perform one dimensional line search and can choose alpha k such that, f of xk plus 1 is minimized if we do this it will ensure that you are moving in the same direction.

So, descent is ensure; that means, the function value at k plus 1 th iteration will be less than the function value at k th iteration; that means, f of xk plus 1 will be less than f of xk.

(Refer Slide Time: 19:35)

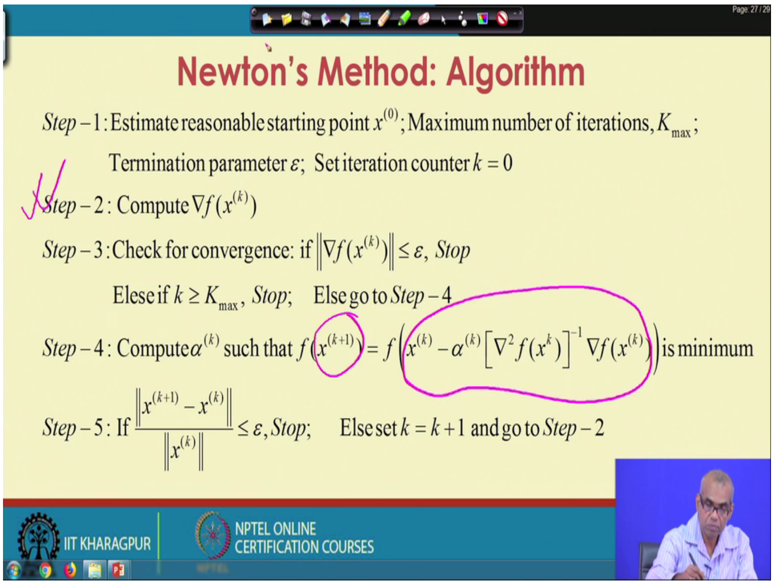


So, this is the algorithm for the Newton's method in the step one we estimate a reasonable starting point x 0, we define maximum number of iterations k max, we define termination parameter epsilon a small value and we said the iteration counter k equal to 0.

So, first thing we do is we compute the gradient of the objective function at xk, which is the current estimate. We check for convergence in the third step that you can do by looking at the norm of the gradient; that means, the magnitude of the gradient if very small will stop, will also stop if you have exceeded the maximum allowable number of iterations. So, if k is greater equal to k max will stop otherwise, we go to next step, in the next step I compute alpha k the step length such that f of xk plus 1 is minimum. So, what is f of xk what is f of xk plus 1? F of xk plus 1 is xk minus alpha k into gradient or Hessian inverse into gradient.

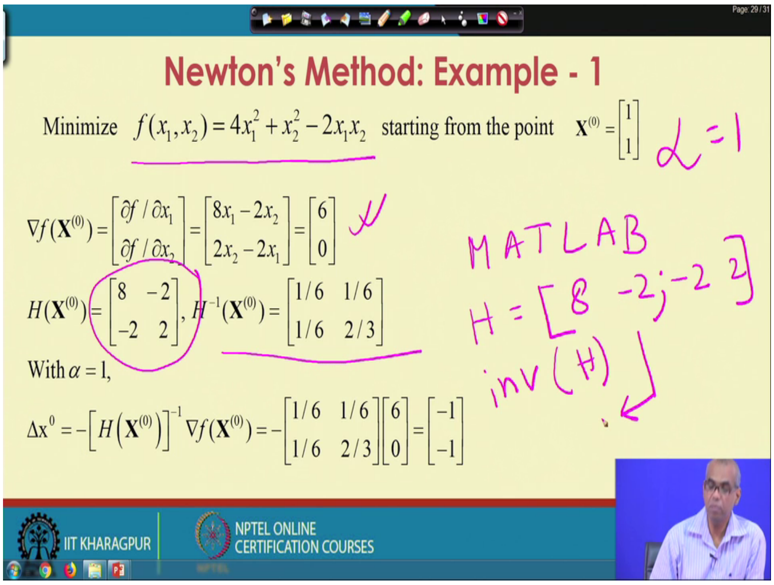
So, xk plus 1 is obtained as xk minus alpha k into Hessian inverse into gradient. So, if you do not have step length; that means, if you are using classical Newton's method. In this case you will find out xk plus 1 as xk minus Hessian inverse gradient xk. If you are using step length at this step we will find out alpha k such that f of xk plus 1 is minimum and f of xk plus 1 is f of xk minus alpha k into Hessian inverse into gradient.

(Refer Slide Time: 22:43)



Note that xk plus 1 is nothing but this and the classical Newton's method is obtained when he said alpha equal to 1. Now, I again check for convergence by saying if the current estimate is changing or not changing appreciably or not; that means, if the value of xk plus 1 and the value of xk are very close to each other, we will stop assuming convergence has been achieved, otherwise we will set k equal to k plus 1 and go to step 2, where we again compute the gradient at this current estimate now xk plus 1. So, this way we proceed iteratively.

(Refer Slide Time: 23:47)



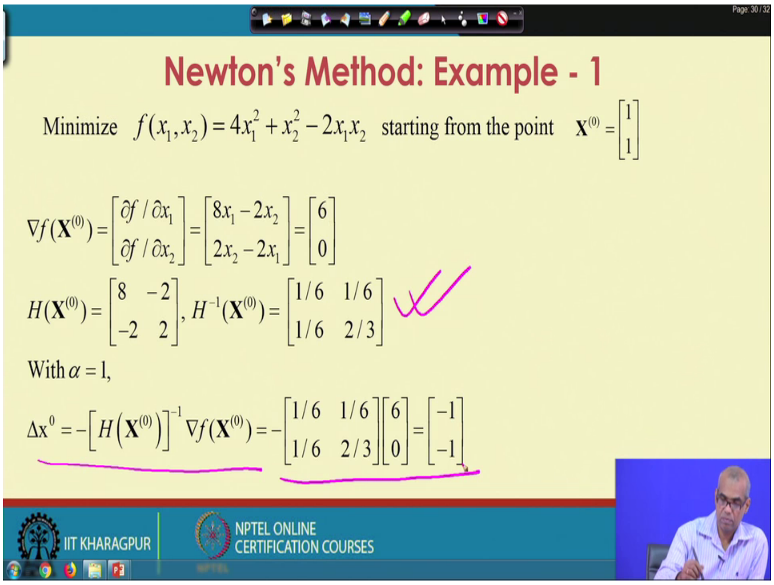
So, let us now consider an example and solve using Newton's method and we will use Newton's classical method; that means we will consider alpha equal to 1. So, you considered a quadratic function 4 x 1 square plus x 2 square minus 2 x 1 x 2 and you start from x 0 equal to 1 1. So, this is a quadratic function and you expect that the conversation will be achieved in one step. So, we have considered alpha equal to 1; the step length alpha equal to 1.

So, the first thing we do is we find out the gradient of the objective function at the given starting point x 0 equal to 1 1. So, my function is 4 x 1 square plus x 2 square minus 2 x 1 x 2. So, del f del x 1 del f del x 2 are the component of the gradient, which can be computed as 8 x 1 minus 2 x 2 and 2 x 2 minus 2 x 1, putting the value of x 0 1 1, I get the gradient as 6 0. Find out the Hessian. Look at the objective function is quadratic function.

So, the Hessian will be a constant matrix and that matrix is 8 minus 2 minus 2 2. Find the Hessian inverse and you obtained as 1 by 6 1 by 6 1 by 6 2 by 3. At this stage I would like to inform that you can find out some software's to find out this inverse as well. For example, if you are using MATLAB, you define the Hessian as 8 minus 2 semicolon minus 2 2, then you use inv stands for inverse of H return you will get the inverse of H.

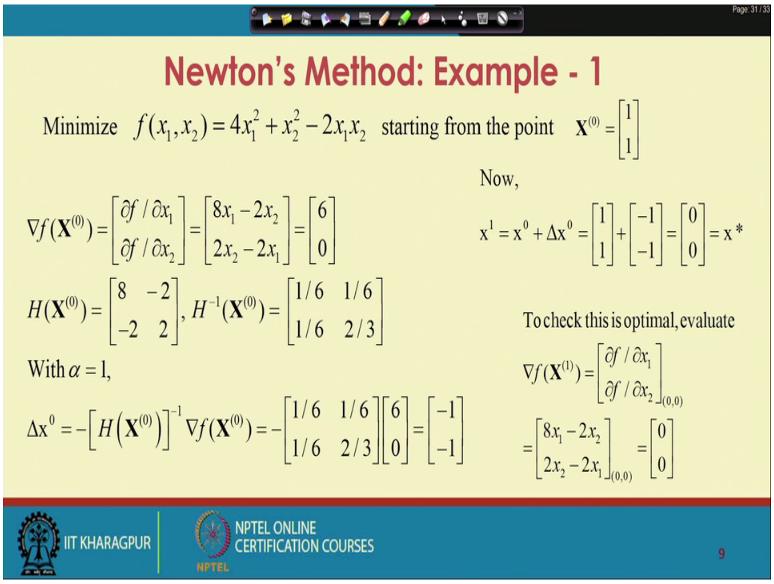
So, you can also analytically find out by hand what is the inverse of H by 2 by 2 matrix and you obtain the inverse as this.

(Refer Slide Time: 27:04)



Now, if alpha equal to 1 xk plus 1 minus xk which is del x 0 is minus H an inverse into gradient. So, this is obtained as minus 1 minus 1. So, delta x 0 is minus 1 minus 1.

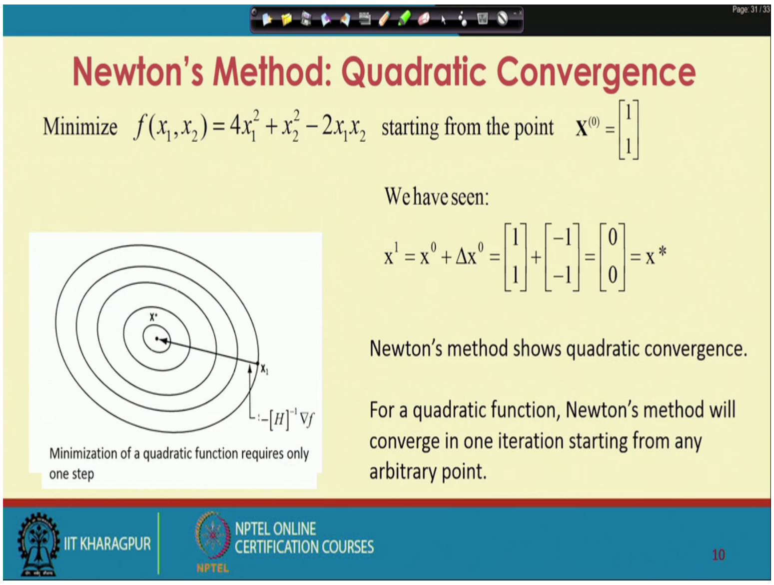
(Refer Slide Time: 27:31)



So, then x 1 equal to x 0 plus delta x 0, because delta x 0 is nothing but x 1 minus x 0.

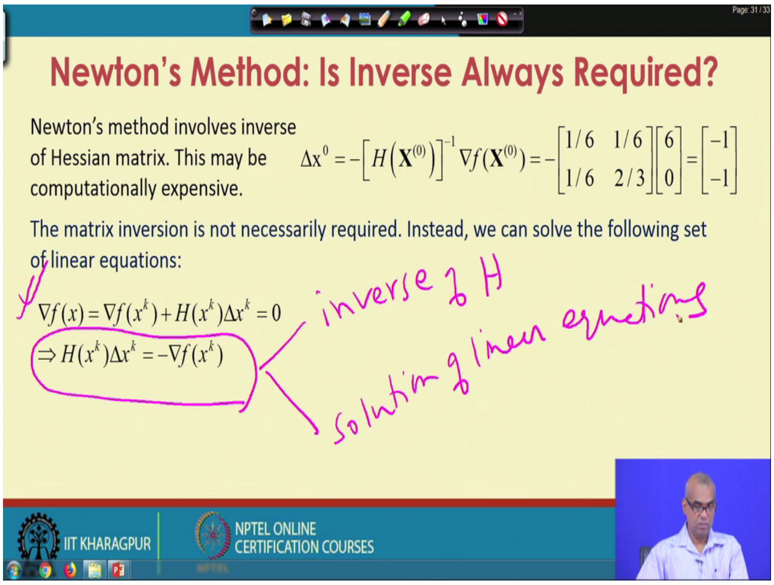
So, x 1 equal to x 0 plus delta x 0 which is nothing but 0 0; 0 0 is the optimal point of the given quadratic function and you see we also get the convergence in one iteration. To check that 0 0 is actually a optimal, evaluate the gradient at this point x 1 which is 0 0 and you see that we get the gradient as 0 0, which satisfies the first order necessary condition. So, Newton's method converges in one iteration for a quadratic objective function.

(Refer Slide Time: 28:35)



So, the figure schematically shows the minimization of a quadratic function in just one step.

(Refer Slide Time: 29:00)

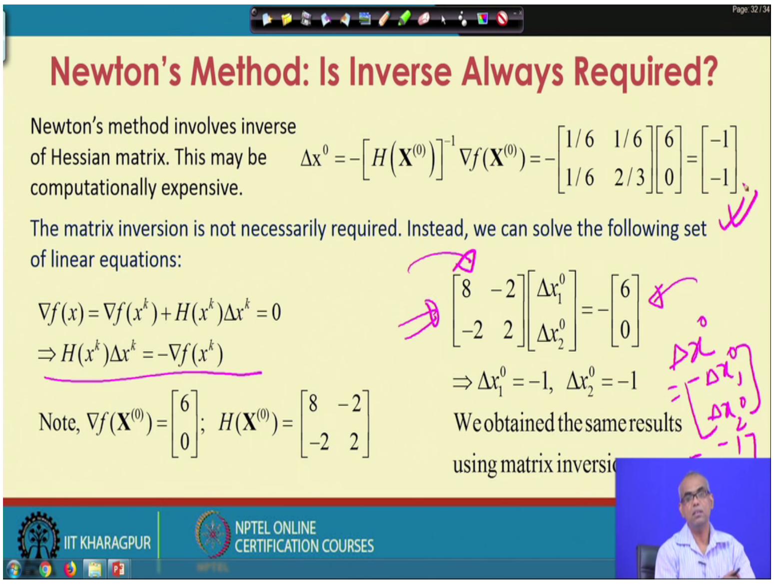


Now, we have seen that inverse of Hessian is required, but is inverse always required Newton's method involves inverse of Hessian matrix and this may be computationally expensive. For example, in the previous case delta x 0 is obtained as minus Hessian inverse into gradient of f and this was obtained as minus 1 minus 1.

Now, the matrix inversion is not necessarily required instead of this, we can also solve the set of linear equations that are involved. So, this is the equation from which the update rule of the classical Newton's method is obtained by applying the first order necessary condition to the quadratic approximation of the objective function. So, I get Hessian into delta x is equal to minus gradient of f. So, this is a set this gives you a set of linear equations which can be solved for delta x. So, there are two ways of solving this equation, one is inverse of Hessian another is solutions of linear equations.

So, a computation of inverse of Hessian is expensive we can solve the set of linear equations.

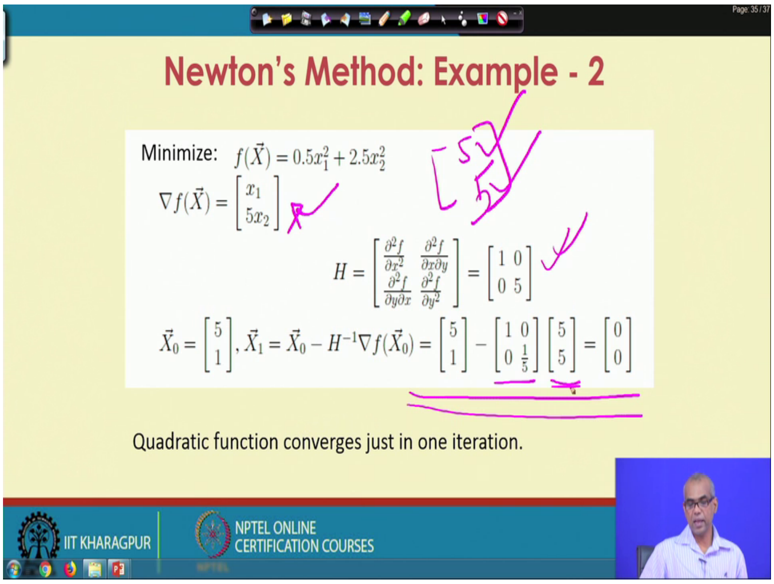
(Refer Slide Time: 31:03)



How let us see. Now look at this expression, Hessian into delta x equal to minus gradient of f. So, gradient of f is 6 0. So, my right hand side is minus 6 0 Hessian is 8 minus 2 minus 2 2 and delta x have two component delta x 1 0 and delta x 2 0. So, this is set of two linear equations in delta x 1 0 and delta x 2 0. So, if we solve I obtain as delta x 1 0 equal to minus 1 delta x 2 0 equal to minus 1.

So, delta x 0 which is delta x 1 0 and delta x 2 0 is again obtained as minus 1 minus 1, which is same as what you obtain using inverse of Hessian. So, inverse of Hessian is not necessarily always required instead you can solve a set of linear equations.

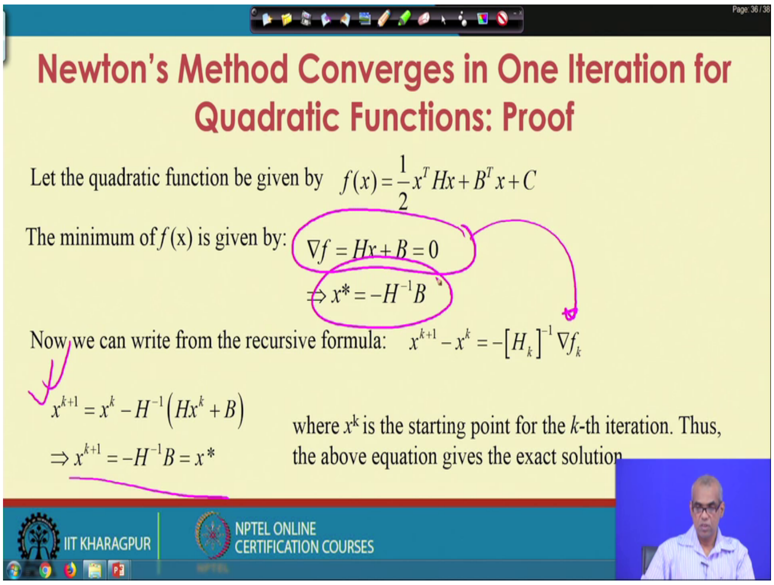
(Refer Slide Time: 32:30)



Let us consider another example again a quadratic function 0.5 x 1 square plus 2.5 x 2 square. The gradient is obtained as x 1 5 x 2. Hessian will be a constant matrix because it is a quadratic function it is obtained as 1 0 0 5 is the positive definite symmetric matrix, my starting vector is 5 1.

So, if it is 5 1 the gradient will be 5 1, x 1 equal to 5 x 2 equal to 1, Hessian inverse will be 1 0 0 1 by 5. So, you see that we again get the convergence in one iteration, the gradient will be 5 5, because 5 x 1 5 x 2. So, x 1 equal to 5 an x and next component is 5 x 2 x 2 equal to 1. So, it is 5. So, the convergence is again achieved in one iteration.

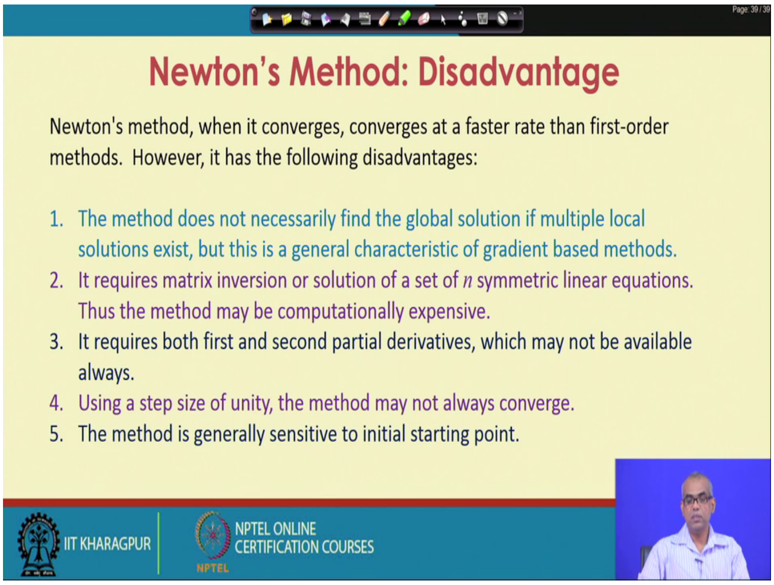
(Refer Slide Time: 34:17)



We can prove that Newton's method convergence in one iteration for quadratic functions easily, let the quadratic function be given by fx equal to half x transpose Hx plus B transpose x plus C the minimum of this function fx is given by the first order necessary condition gradient of f equal to Hx plus B equal to 0, you can solve this equation for x, so x star equal to minus H inverse B.

Now, let us write the recursive formula for the new classical Newton's method, which is xk plus 1 minus xk equal to minus H inverse delta gradient of f. So, xk plus 1 equal to xk minus H inverse Hxk plus B. So, if you simplify this you get xk plus 1 equal to minus H inverse B, which is actually the optimal solution. So, for a quadratic function the Newton's method will converge in one iteration starting from any arbitrary point.

(Refer Slide Time: 35:59)



Newton's method, when it converges at a faster rate than first order methods. However, it has the following disadvantages. The method does not necessarily find the global solution if multiple local solutions exist, but this is a general characteristic of gradient based methods.

It requires matrix inversion or solution of a set of n symmetric linear equations. Thus the method may be computationally expensive. The method requires both first and second partial derivatives which may not be available always. Using a step size of unity, the method may not always converge. The method is generally sensitive to initial starting point.

If the initial starting point is close to the optimal point the method works very well, but if the starting point is very far from the optimal solution, the Newton's method can go either uphill or downhill; that means, it can go in the direction of minimization of function or the maximization of function. You remember the search direction is valid for both minimization and maximization of the function. So, the method is generally sensitive to initial starting point. With this we stop our discussion on Newton's method in today's lecture.