

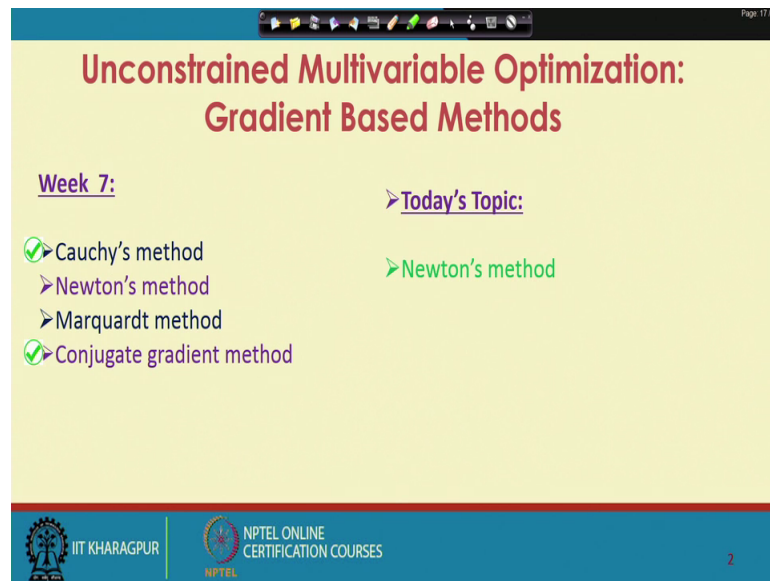
Optimization in Chemical Engineering
Prof. Debasis Sarkar
Department of Chemical Engineering
Indian Institute of Technology, Kharagpur

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)



The slide features a yellow background with a blue header and footer. The title is in red. The content is organized into two columns: 'Week 7:' and 'Today's Topic:'. The 'Week 7:' column lists four methods, with the first and last having green checkmarks. The 'Today's Topic:' column lists one method with a green arrow. The footer contains the IIT Kharagpur and NPTEL logos.

**Unconstrained Multivariable Optimization:
Gradient Based Methods**

Week 7:

- ✔ Cauchy's method
- Newton's method
- Marquardt method
- ✔ Conjugate gradient method

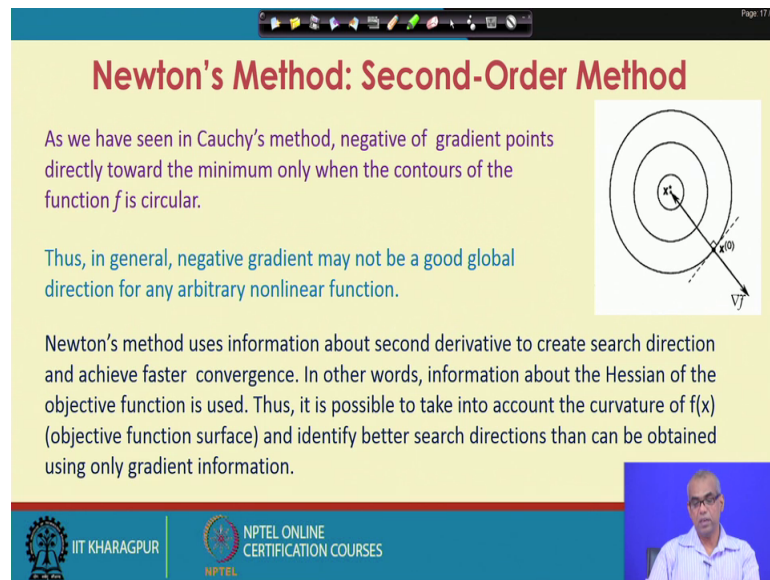
Today's Topic:

- Newton's method

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

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)



Newton's Method: Second-Order Method

As we have seen in Cauchy's method, negative of gradient points directly toward the minimum only when the contours of the function f is circular.

Thus, in general, negative gradient may not be a good global direction for any arbitrary nonlinear 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 $f(x)$ (objective function surface) and identify better search directions than can be obtained using only gradient information.

The diagram shows three concentric circles representing contours of a function. A point $x^{(0)}$ is marked on the outermost circle. A dashed line represents the negative gradient direction, which points towards the center of the circles. A solid arrow labeled ∇f points away from the center, representing the direction of the gradient.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

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 $x^2 + y^2$. So, if you have an objective function like f equal to $x^2 + y^2$, 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 $x^2 + ay^2$ f equal to $x^2 + ay^2$. 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)

Newton's Method: Basic Idea

The basic idea of the classical Newton's method is to use second-order (quadratic) approximation of the function $f(x)$ about the current search point x^k .

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

Here $H(x^k)$ is the Hessian matrix.

Handwritten notes:
 $f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$
 $x - x^k = \Delta x^k$

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 x_k . So, if you look at the figure, this is the function $f(x)$, this is the point x_k and at point x_k about the point x_k , I approximate the objective function $f(x)$ by a quadratic function $q(x)$. Look at here, the point x_k is such that the quadratic approximation $q(x)$ and the function $f(x)$ has more or less same minimum point.

So, the function $f(x)$ can have a quadratic approximation as shown, $f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$; that means, you expand $f(x)$ in terms of Taylor series about the point x_k . So, $x - x_k$ is taken as Δx^k .

(Refer Slide Time: 06:51)

Newton's Method: Basic Idea

The basic idea of the classical Newton's method is to use second-order (quadratic) approximation of the function $f(x)$ about the current search point x^k .

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

Here $H(x^k)$ 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.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

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)

Newton's Method

Consider the second-order Taylor expansion of a multivariable function $f(x)$ about the current point x^k

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

The minimum of this quadratic approximation of $f(x)$ can be obtained by differentiating the equation with respect to each of the components of Δx and equating the resulting expressions to zero. This is First Order Necessary Condition (FONC).

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0$$

$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$. Here, H is assumed to be non-singular.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, consider the second order Taylor expansion of a multivariable function $f(x)$ about current point x^k and you obtain the approximation as $f(x)$ equal to f of x^k plus gradient of f at x^k into Δx^k plus half into Δx^k into H at Δx^k at H into x^k into Δx^k . So, this is

Taylor series expansion of a multivariable function $f(x)$ about the current point x^k and you have retained only up to second order terms.

Now, the minimum of this quadratic approximation of $f(x)$ can be obtained by differentiating the equation with respect to each of the components of Δx 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 $\nabla f(x)$ equal to 0. So, if I do that I will get this, gradient of $f(x)$ is equal to gradient of $f(x^k)$ plus H at x^k into Δx^k .

So, basically what I am doing is $\nabla f(x)$ equal to 0.

(Refer Slide Time: 09:25)

Newton's Method

Consider the second-order Taylor expansion of a multivariable function $f(x)$ about the current point x^k

$$f(x) \approx f(x^k) + \nabla f(x^k)^T \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

The minimum of this quadratic approximation of $f(x)$ can be obtained by differentiating the equation with respect to each of the components of Δx and equating the resulting expressions to zero. This is First Order Necessary Condition (FONC).

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0$$

$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$. Here, H is assumed to be non-singular.

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 Δx^k which is nothing, but $x^{k+1} - x^k$, which is minus of Hessian inverse into gradient both evaluated at x^k . 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)

Page 20/22

Newton's Method

Consider the second-order Taylor expansion of a multivariable function $f(x)$ about the current point x^k

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

The minimum of this quadratic approximation of $f(x)$ can be obtained by differentiating the equation with respect to each of the components of Δx and equating the resulting expressions to zero. This is First Order Necessary Condition (FONC).

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0$$
$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$. Here, H is assumed to be non-singular.

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

This is the Taylor series expansion of a multivariable function about current point x^k . So, this is a quadratic approximation of any non-linear general function $f(x)$. Now I want to find out the minimum of this quadratic function, which is an approximation of the original function $f(x)$. 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 ∇x , we have to differentiate the function with respect to each of the component of Δx and we have to set the resulting expressions to 0. If we do that we get the gradient of $f(x)$ as gradient of f evaluated at x^k , which comes from here and Hessian evaluated x^k multiplied by Δx^k , which comes from here equal to 0.

(Refer Slide Time: 12:05)

Page 21/22

Newton's Method

Consider the second-order Taylor expansion of a multivariable function $f(x)$ about the current point x^k

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

The minimum of this quadratic approximation of $f(x)$ can be obtained by differentiating the equation with respect to each of the components of Δx and equating the resulting expressions to zero. This is First Order Necessary Condition (FONC).

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0$$

$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$. Here, H is assumed to be non-singular.

$H(x^k) \Delta x^k = -\nabla f(x^k)$
 $\Rightarrow \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, now you solve this for delta x^k nothing, but x^k plus 1 minus x^k. So, from here you get H at x^k delta x^k is equal to minus delta f at x^k. So, delta x^k is obtained as minus H⁻¹ at x^k delta f at x^k. 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)

Page 22/24

Newton's Method

Consider the second-order Taylor expansion of a multivariable function $f(x)$ about the current point x^k

$$f(x) \approx f(x^k) + \nabla^T f(x^k) \Delta x^k + \frac{1}{2} (\Delta x^k)^T H(x^k) \Delta x^k$$

The minimum of this quadratic approximation of $f(x)$ can be obtained by differentiating the equation with respect to each of the components of Δx and equating the resulting expressions to zero. This is First Order Necessary Condition (FONC).

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0$$

$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$. Here, H is assumed to be non-singular.

If $H(x^k)$ is positive semi-definite, the approximate function will have a minimum at:

$$x^{k+1} = x^k - [H(x^k)]^{-1} \nabla f(x^k) \quad \text{This is recursive formula for Newton's method.}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

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

(Refer Slide Time: 13:45)

Newton's Method: Compare With 1D Search

$$\nabla f(x) = \nabla f(x^k) + H(x^k)\Delta x^k = 0$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$.

$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

Note that we have solved for a vector, Δx , which has both a step length and direction. Also, the search direction is valid both for minimization and maximization.

$\nabla f = 0$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, this is what we just seen the update rule in case of Newton's method, which is $x_{k+1} - x_k = -[H(x^k)]^{-1} \nabla f(x^k)$, both evaluated at the current point x_k .

Note that we have solved for a vector Δx or Δx_k , which has both a step length and direction. So, Δx 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)

Newton's Method: Compare With 1D Search

$$\nabla f(x) = \nabla f(x^k) + H(x^k)\Delta x^k = 0$$
$$\Rightarrow x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$$

where $[H(x^k)]^{-1}$ is the inverse of the Hessian matrix $H(x^k)$.

Note that we have solved for a vector, Δx , which has both a step length and direction. Also, the search direction is valid both for minimization and maximization.

Compare with one-dimension search (single variable unconstrained search):

$$x^{k+1} = x^k - \frac{f'(x^k)}{f''(x^k)}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

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

(Refer Slide Time: 16:03)

Newton's Method: Modified Update Rule

Classical Newton's Method: $x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$

If $f(x)$ is actually quadratic, only one step is required to reach the minimum of $f(x)$.

For a general nonlinear objective function, however, the minimum of $f(x)$ cannot be reached in one step. The Newton step change may be large when we are far from optimum and there is a chance of divergence. Thus, we can modify the above equation (search direction) by introducing the step length parameter (α):

$$x^{k+1} - x^k = \Delta x^k = -\alpha^k [H(x^k)]^{-1} \nabla f(x^k)$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Now for the classical Newton's method update rule is $x_{k+1} - x_k$ equal to minus Hessian inverse into gradient of the objective function.

Now, if the objective function $f(x)$ 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 $f(x)$ cannot be reached in one step. The minimum of $f(x)$ 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 step length parameter α .

So, instead of taking $x_{k+1} - x_k$ equal to minus Hessian inverse into gradient of f , I can introduce a step length parameter α and can write as, $x_{k+1} - x_k$ equal to minus α times 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)

Page: 28/28

Newton's Method: Modified Update Rule



Classical Newton's Method: $x^{k+1} - x^k = \Delta x^k = -[H(x^k)]^{-1} \nabla f(x^k)$

If $f(x)$ is actually quadratic, only one step is required to reach the minimum of $f(x)$.

For a general nonlinear objective function, however, the minimum of $f(x)$ cannot be reached in one step. The Newton step change may be large when we are far from optimum and there is a chance of divergence. Thus, we can modify the above equation (search direction) by introducing the step length parameter (α):

$$x^{k+1} - x^k = \Delta x^k = -\alpha^k [H(x^k)]^{-1} \nabla f(x^k)$$

By performing one-dimensional line search, we can choose α^k such that $f(x^{k+1})$ is minimized. This will ensure descent: $f(x^{k+1}) < f(x^k)$

 IIT KHARAGPUR
  NPTEL ONLINE CERTIFICATION COURSES

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)

Newton's Method: Algorithm

Step - 1: Estimate reasonable starting point $x^{(0)}$; Maximum number of iterations, K_{\max} ; Termination parameter ϵ ; Set iteration counter $k = 0$

Step - 2: Compute $\nabla f(x^{(k)})$

Step - 3: Check for convergence: if $\|\nabla f(x^{(k)})\| \leq \epsilon$, Stop
Else if $k \geq K_{\max}$, Stop; Else go to Step - 4

Step - 4: Compute $\alpha^{(k)}$ such that $f(x^{(k+1)}) = f\left(x^{(k)} - \alpha^{(k)} \left[\nabla^2 f(x^{(k)}) \right]^{-1} \nabla f(x^{(k)})\right)$ is minimum

Step - 5: If $\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k)}\|} \leq \epsilon$, Stop; Else set $k = k + 1$ and go to Step - 2

Handwritten notes in pink:
 $H^{-1} = \alpha^{(k)} - \frac{1}{H(x^{(k)})} \nabla f(x^{(k)})$
 H

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 x_k , 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 x_k plus 1 is minimum. So, what is f of x_k what is f of x_k plus 1? f of x_k plus 1 is x_k minus alpha k into gradient or Hessian inverse into gradient.

So, $x_k + 1$ is obtained as $x_k - \alpha_k \text{Hessian}^{-1} \text{gradient}$. 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 $x_k + 1$ as $x_k - \text{Hessian}^{-1} \text{gradient}$. If you are using step length at this step we will find out α_k such that f of $x_k + 1$ is minimum and f of $x_k + 1$ is f of $x_k - \alpha_k \text{Hessian}^{-1} \text{gradient}$.

(Refer Slide Time: 22:43)

Newton's Method: Algorithm

Step - 1: Estimate reasonable starting point $x^{(0)}$; Maximum number of iterations, K_{\max} ; Termination parameter ϵ ; Set iteration counter $k = 0$

Step - 2: Compute $\nabla f(x^{(k)})$

Step - 3: Check for convergence: if $\|\nabla f(x^{(k)})\| \leq \epsilon$, Stop
Else if $k \geq K_{\max}$, Stop; Else go to Step - 4

Step - 4: Compute $\alpha^{(k)}$ such that $f(x^{(k)} - \alpha^{(k)} [\nabla^2 f(x^{(k)})]^{-1} \nabla f(x^{(k)})$ is minimum

Step - 5: If $\frac{\|x^{(k+1)} - x^{(k)}\|}{\|x^{(k)}\|} \leq \epsilon$, Stop; Else set $k = k + 1$ and go to Step - 2

Note that $x_k + 1$ is nothing but this and the classical Newton's method is obtained when he said α 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 $x_k + 1$ and the value of x_k are very close to each other, we will stop assuming convergence has been achieved, otherwise we will set k equal to $k + 1$ and go to step 2, where we again compute the gradient at this current estimate now $x_k + 1$. So, this way we proceed iteratively.

(Refer Slide Time: 23:47)

Newton's Method: Example - 1

Minimize $f(x_1, x_2) = 4x_1^2 + x_2^2 - 2x_1x_2$ starting from the point $\mathbf{X}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ $\alpha = 1$

$$\nabla f(\mathbf{X}^{(0)}) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \end{bmatrix} = \begin{bmatrix} 8x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$H(\mathbf{X}^{(0)}) = \begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix}, H^{-1}(\mathbf{X}^{(0)}) = \begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix}$$

With $\alpha = 1$,

$$\Delta \mathbf{x}^0 = -[H(\mathbf{X}^{(0)})]^{-1} \nabla f(\mathbf{X}^{(0)}) = -\begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

MATLAB
H = [8 -2; -2 2]
inv(H)

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 $4x_1^2 + x_2^2 - 2x_1x_2$ and you start from \mathbf{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 \mathbf{x}^0 equal to $[1 \ 1]$. So, my function is $4x_1^2 + x_2^2 - 2x_1x_2$. So, $\partial f / \partial x_1$ $\partial f / \partial x_2$ are the component of the gradient, which can be computed as $8x_1 - 2x_2$ and $2x_2 - 2x_1$, putting the value of \mathbf{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 $\begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix}$. Find the Hessian inverse and you obtained as $\begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix}$. 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 $\begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix}$, 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 2×2 matrix and you obtain the inverse as this.

(Refer Slide Time: 27:04)

Page 30/32

Newton's Method: Example - 1

Minimize $f(x_1, x_2) = 4x_1^2 + x_2^2 - 2x_1x_2$ starting from the point $\mathbf{X}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$\nabla f(\mathbf{X}^{(0)}) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \end{bmatrix} = \begin{bmatrix} 8x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$H(\mathbf{X}^{(0)}) = \begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix}, H^{-1}(\mathbf{X}^{(0)}) = \begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix}$$

With $\alpha = 1$,

$$\Delta \mathbf{x}^0 = -[H(\mathbf{X}^{(0)})]^{-1} \nabla f(\mathbf{X}^{(0)}) = -\begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

Now, if alpha equal to 1 x_k plus 1 minus x_k which is \mathbf{x}^0 is minus H an inverse into gradient. So, this is obtained as minus 1 minus 1. So, $\Delta \mathbf{x}^0$ is minus 1 minus 1.

(Refer Slide Time: 27:31)

Page 31/33

Newton's Method: Example - 1

Minimize $f(x_1, x_2) = 4x_1^2 + x_2^2 - 2x_1x_2$ starting from the point $\mathbf{X}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

Now,

$$\mathbf{x}^1 = \mathbf{x}^0 + \Delta \mathbf{x}^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{x}^*$$

$$\nabla f(\mathbf{X}^{(0)}) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \end{bmatrix} = \begin{bmatrix} 8x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix} = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$$H(\mathbf{X}^{(0)}) = \begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix}, H^{-1}(\mathbf{X}^{(0)}) = \begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix}$$

With $\alpha = 1$,

$$\Delta \mathbf{x}^0 = -[H(\mathbf{X}^{(0)})]^{-1} \nabla f(\mathbf{X}^{(0)}) = -\begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

To check this is optimal, evaluate

$$\nabla f(\mathbf{X}^{(1)}) = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \end{bmatrix}_{(0,0)} = \begin{bmatrix} 8x_1 - 2x_2 \\ 2x_2 - 2x_1 \end{bmatrix}_{(0,0)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

So, then x_1 equal to x_0 plus Δx_0 , because Δx_0 is nothing but x_1 minus x_0 .

So, x_1 equal to x_0 plus Δx_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)

Newton's Method: Quadratic Convergence

Minimize $f(x_1, x_2) = 4x_1^2 + x_2^2 - 2x_1x_2$ starting from the point $\mathbf{x}^{(0)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

We have seen:

$$\mathbf{x}^1 = \mathbf{x}^0 + \Delta\mathbf{x}^0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{x}^*$$

Newton's method shows quadratic convergence.

For a quadratic function, Newton's method will converge in one iteration starting from any arbitrary point.

Minimization of a quadratic function requires only one step

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES | 10

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

(Refer Slide Time: 29:00)

Newton's Method: Is Inverse Always Required?

Newton's method involves inverse of Hessian matrix. This may be computationally expensive.

$$\Delta\mathbf{x}^0 = -[H(\mathbf{X}^{(0)})]^{-1} \nabla f(\mathbf{X}^{(0)}) = - \begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

The matrix inversion is not necessarily required. Instead, we can solve the following set of linear equations:

$$\nabla f(x) = \nabla f(x^k) + H(x^k)\Delta x^k = 0$$

$$\Rightarrow H(x^k)\Delta x^k = -\nabla f(x^k)$$

inverse of H
solving linear equations

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

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 Δx^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 Δx 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 Δx . 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)

Newton's Method: Is Inverse Always Required?

Newton's method involves inverse of Hessian matrix. This may be computationally expensive.

$$\Delta x^0 = -[H(\mathbf{X}^{(0)})]^{-1} \nabla f(\mathbf{X}^{(0)}) = -\begin{bmatrix} 1/6 & 1/6 \\ 1/6 & 2/3 \end{bmatrix} \begin{bmatrix} 6 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

The matrix inversion is not necessarily required. Instead, we can solve the following set of linear equations:

$$\nabla f(x) = \nabla f(x^k) + H(x^k) \Delta x^k = 0$$

$$\Rightarrow H(x^k) \Delta x^k = -\nabla f(x^k)$$

Note, $\nabla f(\mathbf{X}^{(0)}) = \begin{bmatrix} 6 \\ 0 \end{bmatrix}$; $H(\mathbf{X}^{(0)}) = \begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix}$

$$\begin{bmatrix} 8 & -2 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} \Delta x_1^0 \\ \Delta x_2^0 \end{bmatrix} = -\begin{bmatrix} 6 \\ 0 \end{bmatrix}$$

$\Rightarrow \Delta x_1^0 = -1, \Delta x_2^0 = -1$

We obtained the same results using matrix inversion.

Handwritten notes in pink: $\Delta x^0 = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$

How let us see. Now look at this expression, Hessian into Δx 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 Δx have two component Δx_1^0 and Δx_2^0 . So, this is set of two linear equations in Δx_1^0 and Δx_2^0 . So, if we solve I obtain as Δx_1^0 equal to minus 1 Δx_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)

Newton's Method: Example - 2

Minimize: $f(\vec{X}) = 0.5x_1^2 + 2.5x_2^2$

$$\nabla f(\vec{X}) = \begin{bmatrix} x_1 \\ 5x_2 \end{bmatrix}$$

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 5 \end{bmatrix}$$

$$\vec{X}_0 = \begin{bmatrix} 5 \\ 1 \end{bmatrix}, \vec{X}_1 = \vec{X}_0 - H^{-1} \nabla f(\vec{X}_0) = \begin{bmatrix} 5 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Quadratic function converges just in one iteration.

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)

Newton's Method Converges in One Iteration for Quadratic Functions: Proof

Let the quadratic function be given by $f(x) = \frac{1}{2}x^T Hx + B^T x + C$

The minimum of $f(x)$ is given by: $\nabla f = Hx + B = 0$
 $\Rightarrow x^* = -H^{-1}B$

Now we can write from the recursive formula: $x^{k+1} - x^k = -[H_k]^{-1} \nabla f_k$

$x^{k+1} = x^k - H^{-1}(Hx^k + B)$
 $\Rightarrow x^{k+1} = -H^{-1}B = x^*$

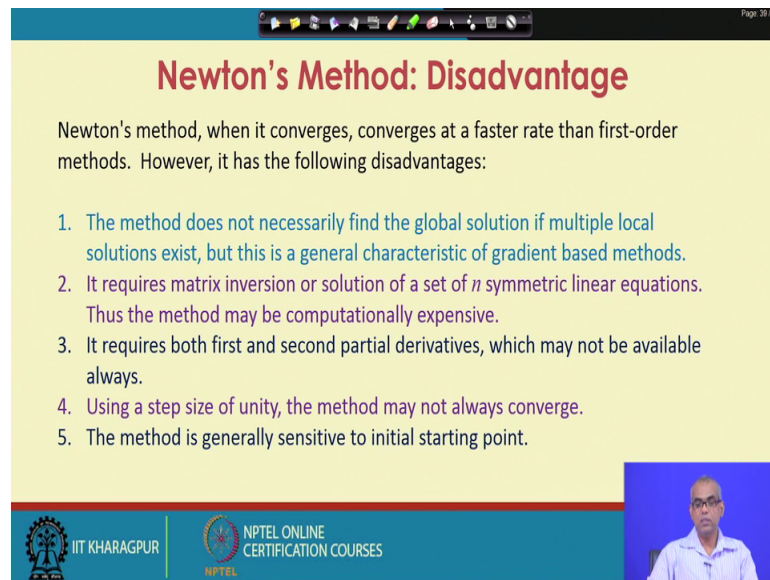
where x^k is the starting point for the k -th iteration. Thus, the above equation gives the exact solution

IIT KHARAGPUR | NPTEL ONLINE CERTIFICATION COURSES

We can prove that Newton's method convergence in one iteration for quadratic functions easily, let the quadratic function be given by $f(x)$ equal to half x transpose Hx plus B transpose x plus C the minimum of this function $f(x)$ 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^* equal to minus H inverse B .

Now, let us write the recursive formula for the new classical Newton's method, which is $x_{k+1} - x_k$ equal to minus H inverse Δ gradient of f . So, x_{k+1} equal to x_k minus H inverse Hx_k plus B . So, if you simplify this you get x_{k+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)



The slide is titled "Newton's Method: Disadvantage" in red text. Below the title, it states: "Newton's method, when it converges, converges at a faster rate than first-order methods. However, it has the following disadvantages:". A numbered list follows:

1. The method does not necessarily find the global solution if multiple local solutions exist, but this is a general characteristic of gradient based methods.
2. It requires matrix inversion or solution of a set of n symmetric linear equations. Thus the method may be computationally expensive.
3. It requires both first and second partial derivatives, which may not be available always.
4. Using a step size of unity, the method may not always converge.
5. The method is generally sensitive to initial starting point.

The slide footer includes the IIT KHARAGPUR logo, the NPTEL ONLINE CERTIFICATION COURSES logo, and a small video inset of a man in a white shirt.

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.