INDIAN INSTITUTE OF TECHNOLOGY ROORKEE

NPTEL

NPTEL ONLINE CERTIFICATION COURES

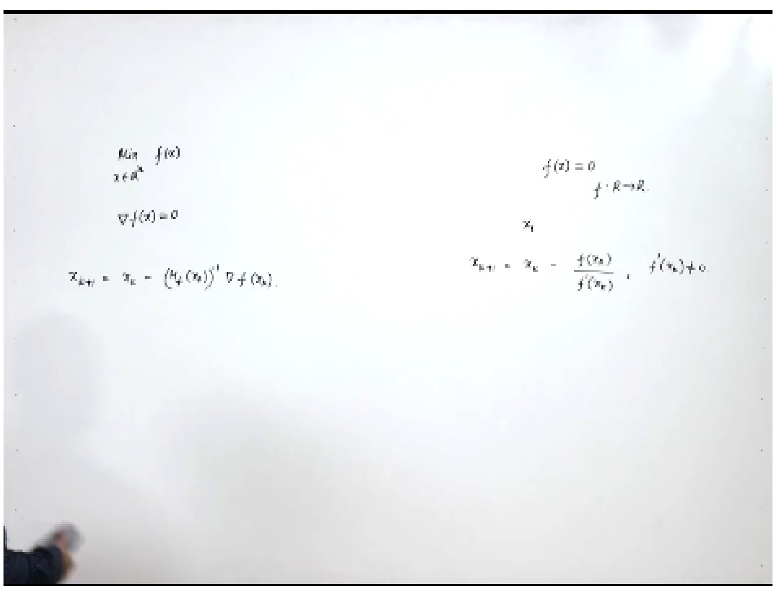
Nonlinear Programming - 1

Lec – 20
Search Techniques-III

Dr. S. K. Gupta
Department of Mathematics
Indian Institute of Technology Roorkee

So hello friends welcome to the lecture series on nonlinear programming, so this is the last lecture of the course which deal with search techniques in this lecture we will deal with two mode search technique that is Newton's mattered and conjugate direction mattered so what Newton mattered is and how it is important to solve non linear unconstraint optimization problems and what conjugate direction mattered is just see. So what is Newton mattered basically suppose you have a unconstrained optimization problem.

(Refer Slide Time: 00:54)



Problem is suppose we have to minimize a function fx where x $\in R^n$ so it is a unconstrained optimization problem this is this function we have to minimize subject to x $\in R^n$ where x R^n

okay now in numerical analysis's we have detailed with a technique called Newton raphson method.

This method is something similar to that method how now suppose you have a differentiable function f(x) = 0 f is from R → R we have differentiable function f from R → R such that F(x) = 0, now if you are interested to find out to find the roots of this equation f(x) = 0 so we sometimes use Newton raphson method what it is it is a recursive scheme bacillary in which starting from a initial gas say x1 which is given to us or sometimes we assume x1 as initial gas.

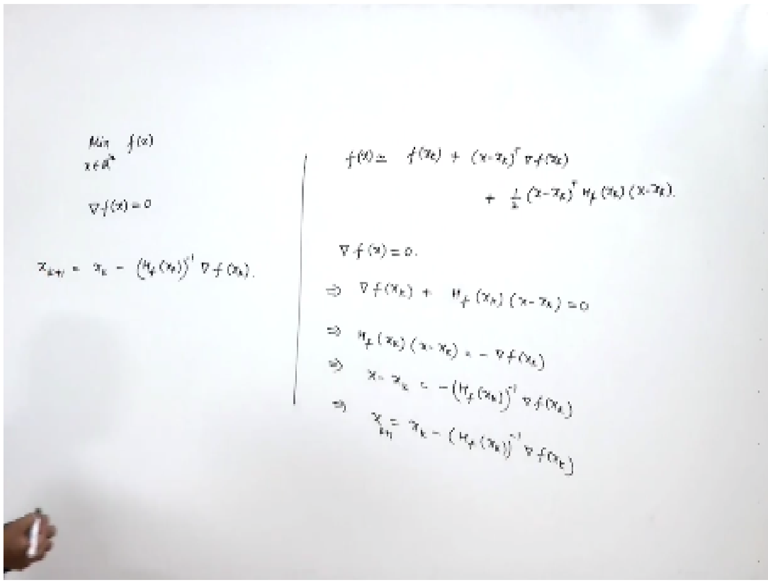
We find xk + 1 – xk – f(x)k f'xk where of course f'(xk) should not equal to 0 for all xk you should so this I am not going to the must detail of this method basically that how this method is arrived I am simply illustrating that this Newton methods to solve uncontained optimization problem is comes from Newton Raphson method what Newton Raphson method is that this matter is basically suppose we have function f from R → R okay which such that fx = 0.

And you are interested to find out the roots of this equation to find out root this equation we use this recursion algorithm x1 initial gas known to you suppose in a similar way suppose you want to minimize this f okay subject to x $\in R^n$ now whenever we have to minimized or maximize a function that means we have to finds out a derivative of the function put it = 0 all those point where derivative is = 0 that will give the point of maxima or minima.

Basically we are interested to find out those points where derivative is 0 because those are the stationary points we have the function we will assume we will take maximum or minimum value okay, so basically to find out to find out the points if this function is differentiable to find out the point where this attentions it is maxima or minimum where basically interested to solve ∇ of f(x) = 0 we interested to solve this equation basically we are interested to solve interested to find those x where ∇ fx = 0 because all those x will greater than f = 0 gives the straight point we are f will again maxima or minimum okay that may basically we have to solve this equation okay now this is a gradient know the derivative okay so we will use this recursion f(xk) = 1 = xk – ancient matrix of f and xk let say inverse and gradient of f at xk so basically it comes from here only here itself okay basically ensure we solve in this equation.

We are solving this ,equation these are system of equation basically because it is a gradient okay now how e obtain this result how we obtain this expression let us see, this comes from Taylor series approximation okay.

(Refer Slide Time: 05:04)

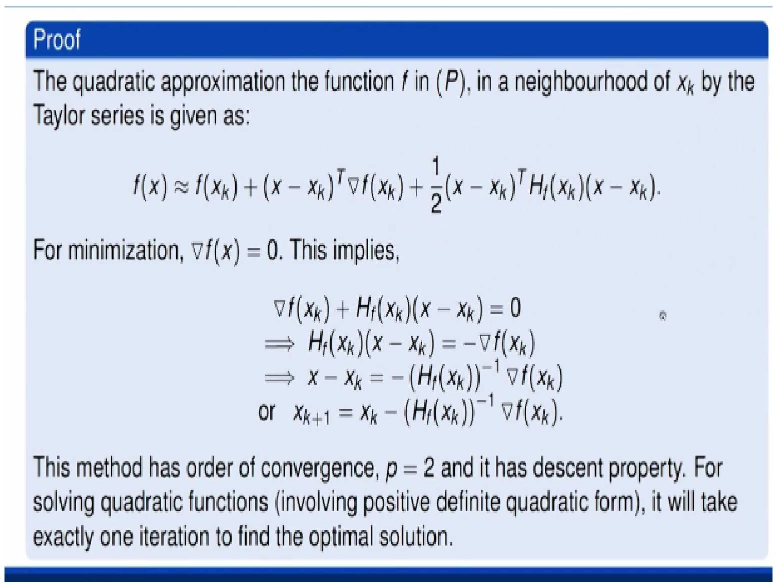


Now to find out the derivation of this expression let us find out let us write the Taylors series quality approximation of f this f let us try to find out Taylors race by Taylors series the quality of approximation of this f at xk in the neighborhood of xk now how e can write it, it is fk + x – xkT gradient of f(xk) + 1/ 2 x – x(kT) Hessian matrix of f at xk into x – xk this is the quality approximation of this function about xk okay now we know that gradient of fx must be 0, gradient of fx must be 0 for the stationary points so this implies, this implies basically this is 0.

Because this is a fixed value so > this will be 0 and x into this will be the gradient of f(xk) when we take gradient both the sides and this will be +x this will be + Hessian matrix of f at xk into x – xk = 0 so this implies Hessian matrix of f at xk x – xk = - gradient of fxk and this implies x – xk now we assume the hessian matrix at xk is invertible then only we can take the inverse both the sides okay, so we are assuming that the inverse of Hessian matrix for each xk exist okay, so there is negative of Hessian matrix of xk.

Inverse gradient of f(xk) and this implies f = xk – hessian matrix of f at xk inverse gradient of f(xk) and we write it xk + 1 so in this way we can derive this result so this is basically Newton’s method to solve unconstraint optimization problems okay, so the limitation of this method is that this method will process only when the Hessian matrix at xk is invertible only then we can proceed by a Newton’s method but this convergence rate is much faster and it has the descent property descent property means for each xk the value if f decrease okay so if convergence rate to 2 and it has the descent property.

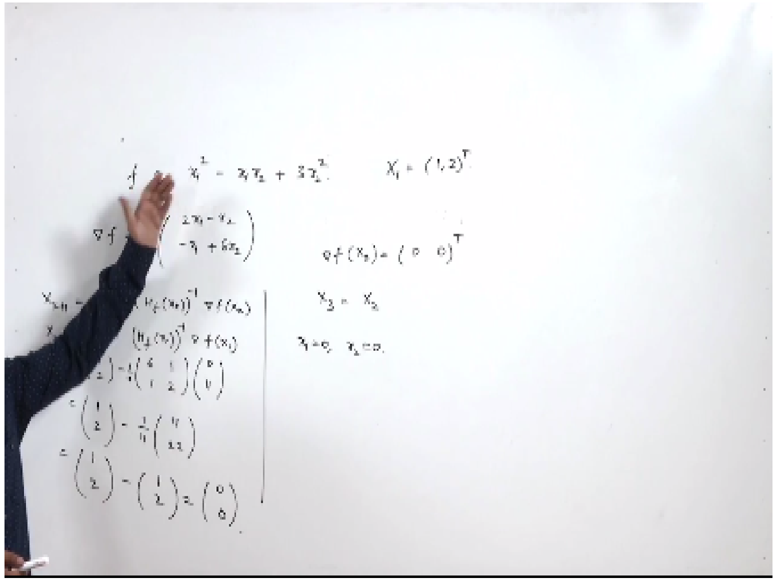
(Refer Slide Time: 08:03)



For solving quadratic functions involving positive definite quadratic form that is a hessian matrices possible definite it will take exactly 1/8 ratio to find out the optimal solution that is the beauty of this matrix that if the if we have a quadratic programming problem suppose were the hessian matrices positive definite then it will take only one iteration to find out the optimal solution okay.

Now suppose we have this problem and we want to solve this problem let us suppose using Newton matrix so how can we proceed.

(Refer Slide Time: 08:39)



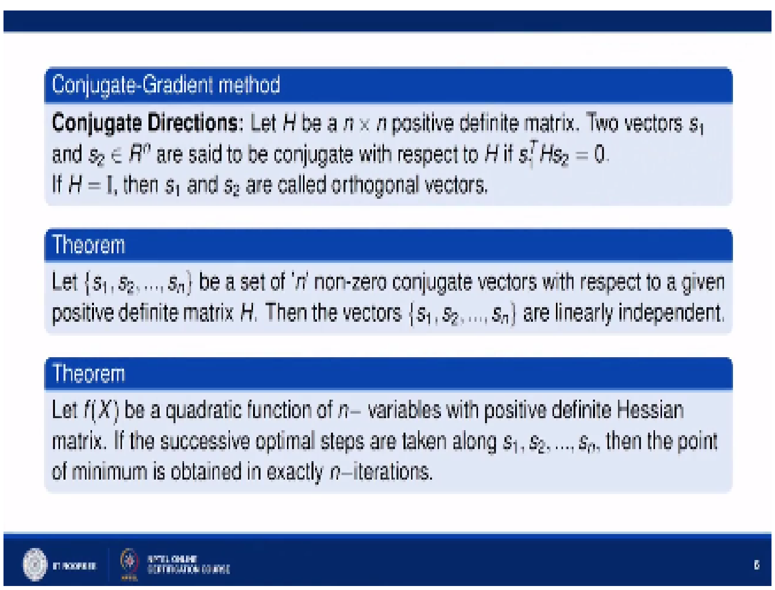
So problem is basically f = $x1^2$ – x1 x2 + $3x2^2$ okay so this is the problem and initial approximation which we can take is 1, 2 it is capital x1 which we can take as 1, 2 okay. So using Newton method first we find ∇f, ∇f will be it is 2x1 – x2 and it is –x2 + 6x1 it is Δf/Δx1, Δf/Δx2 okay and what is hessian matrix of f it is 2,-1,-1, 6 and is it positive definite so when you take it is 6x2 okay it is x1 sorry when you find ∇f(x), Δf/Δx1 is this thing, Δf/Δx2 is x1 + 6x2 okay from 6x2 from here, now when you find hessian matrix of f it is 2 it is -1,p-1 and 6 okay, now this hessian matrix when you compute D1 it is two which is greater than 0 and D2whne you compute it is 12 – 1 which is 11 and again greater than 0.

So the hessian matrix is positive definite, okay. So hessian matrix positive definite and it is free from x1 and x2 it is a fixed matrix, so what is what is it inverse is it invertible, yes it is invertible because is a positive definite matrix and every positive definite matrix is invertible, okay. It is invertible so what is the inverse of this matrix, inverse will be a 6,2,1,1 and determinant is 11/11. So this will be the inverse of this matrix.
So we will use this recursion this recursive algorithm to solve this problem okay, take a transpose of this okay, you need a column vector here so it is okay, so when you x2 it is x1-Hf matrix of f at x1 it is inverse gradient of f(x1), now it is equals to x1 is the first initial we access 1 and 2- okay, now inverse of Hf matrix is 1/11 it is 6,1,1,2 and gradient of f(x1) is when you take x1, x1 is 1 and 2 it is 1 and 2 that is 0 when you take 1 and 2 here it is 11, so gradient of f(x1) is 0,11.

And it is 1,2- it is 1/11 when you take this row this column it is 11, this row this column it is 22 and it is nothing but (12)-(12) which is 0,0 so H2 is 0 and 0 now when you compute for x3 here when you compute for x3 what is gradient of f(x2) when you put 0,0 it is 0,0 and of course when it is 0,0, 0,0 into this will be 0 so xk+1 will be xk that is x3 will be x2, so that means in the iteration coincide with the previous iteration that means the optimal solution is 0 and 0.

So x1=0 and x2=0 is the optimal solution okay, so it exactly takes one iteration to solve to this positive definite quadratic programming problem okay, so in this way applying Newton's method we can solve NN constrained optimization problems. Now the next method is conjugate gradient method, so what this method is let us see, so before setting the method we have some terms let us define those terms first is conjugate directions.

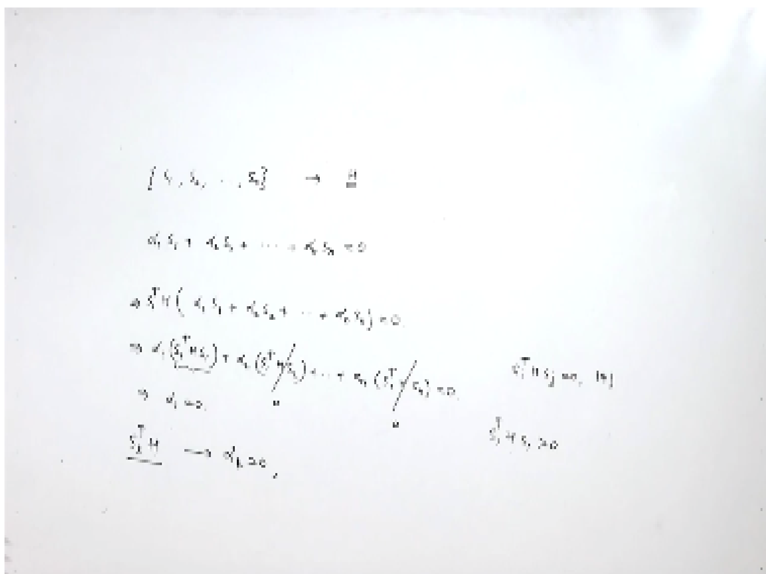
(Refer Slide Time: 13:48)



So let H be a nxn positive definite matrix okay, the two vectors as shown in s2 belongs to R^n are said to be conjugate with respect to h, if $s1^T h_s = 0$ if this equation hold that is $s_1 T\ h\ s_2 = 0$ where h is a positive definite matrix then the vector s_1 s_2 are called conjugate vectors respect to the matrix h or conjugate directions. Now if h = identity matrix then $s_1 s_2$ are orthogonal vectors we already know that if it is identity matrix then this is $s_1^T s_2 = 0$ that means vectors are orthogonal vectors.

Now we have a theorem let s_1 s_2 up to s_n be a set of n non 0 conjugate vectors with respect to a given positive definite matrix such okay, then the set of vectors s_1 s_2 up to s_n are linearly independent.

(Refer Slide Time: 15:13)



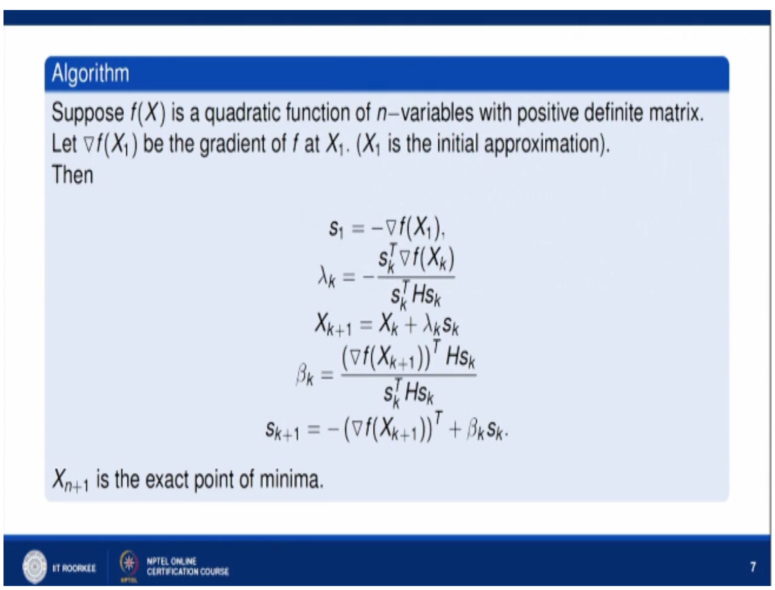
So if we have a set of vectors are s_1 s_2 up to s_n these are set of vectors which are conjugate vectors with respect to a given positive matrix such okay. Then we have to show that these vectors are linearly independent. So how to show take a linear combination of these vectors put it equal to 0 and if anyhow we prove that all α is a 0 this means vectors are linearly independent okay, so it implies that all α is are 0 that means vectors are linearly independent. Now suppose you multiply both side to this equation is given to you and you have to prove that all α is a 0, so you multiply both sides by suppose s_1^T h if you multiply both side by this factor so what we will obtain?

We obtain this in to α_1 s_1 + α_2 s_2 and so on α and s_n = 0 this implies since α_1 is a scalar so it can be take an out α_1 it s s_1^T hs_1 α_2 is a scalar can be taken out s_1^T is s_2 and so on α_n s_1^T h_sn, now since s_1 s_2 up to s_n are conjugate vectors are conjugate directions respect to a fix positive definite matrix H so this means s_i conjugate h s_j = 0 if i_0 = 0 okay, this means this is 0 then this is 0 all are the values are 0 only this value will be non 0 okay.

And since h is the positive definite matrix so s_1^T as h_1 will be significant 0 so s_1 ^{T} hs_1 will be significant and 0 okay. so it is non 0 value so this implies $\alpha_1 = 0$, so similarly if you multiply both sides by s_2 ^{T} h then it will give $\alpha_2 = 0$ and similarly if you multiply both sides by s_k^T h so this will give this will implied $\alpha_k = 0$ where k may be 1 2 3 up to L so in this way we obtain all α = 0 and hence we can say that the set s_1 s_2 up to s_n is linearly independent okay.

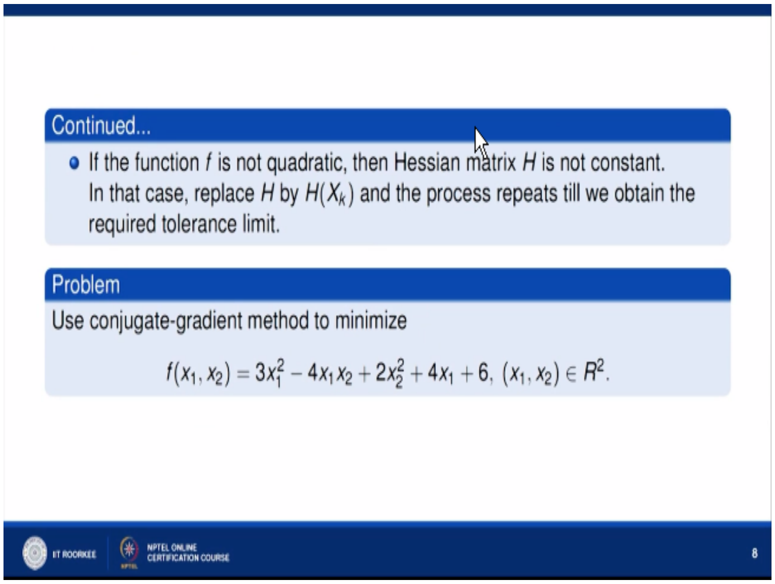
So hence we have this result, now the next result let f be a cultic function of n variables with positive definite hessian matrix okay if the successive optimal steps are taken along s_1 s_2 up to s_n then the point of minimize obtained exactly in n iterations. So this is the theorem I am not going to in the proof of this theorem okay, suppose f a cultic function of n variables with positive definite matrix okay.

(Refer Slide Time: 19:08)



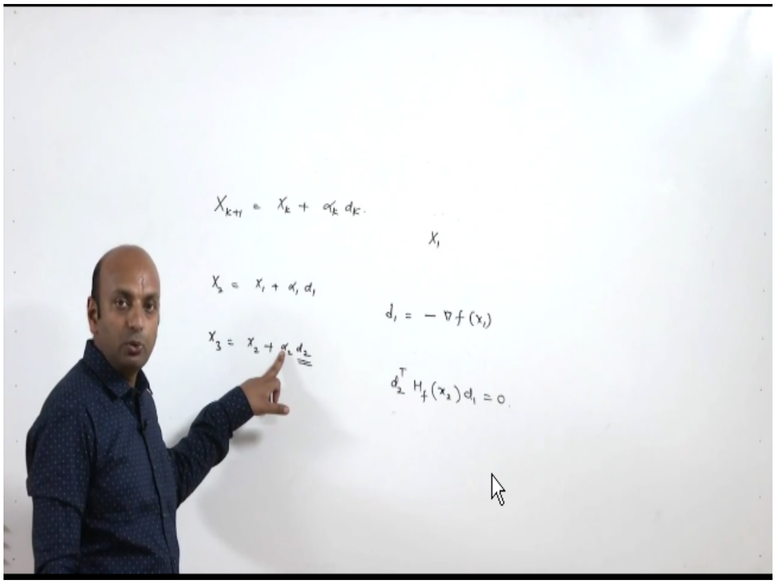
Let gradient of f at x_1 with the gradient of f at x_1 where x_1 is an initial approximation then we have all this expression basically what is the main idea of this conjugate gradient method.

(Refer Slide Time: 19:23)



It's something like sleeper decent matter but in sheet of moving along negative or gradient of f we move along conjugate directions now what is there in this matter let us see the main algorithm of this matter basically technical is scheme.

(Refer Slide Time: 19:48)



That is it is $x_{k+1}=x_k+\alpha_k d_k$ we are d_k is an direction in which we should proceed so that we will be having decent property that is the value of f decreases okay α_k is an optimal step size that how much we should move in that direction that is optimal steps of α_k and using this we will get next iteration that is x_{k+1} okay.

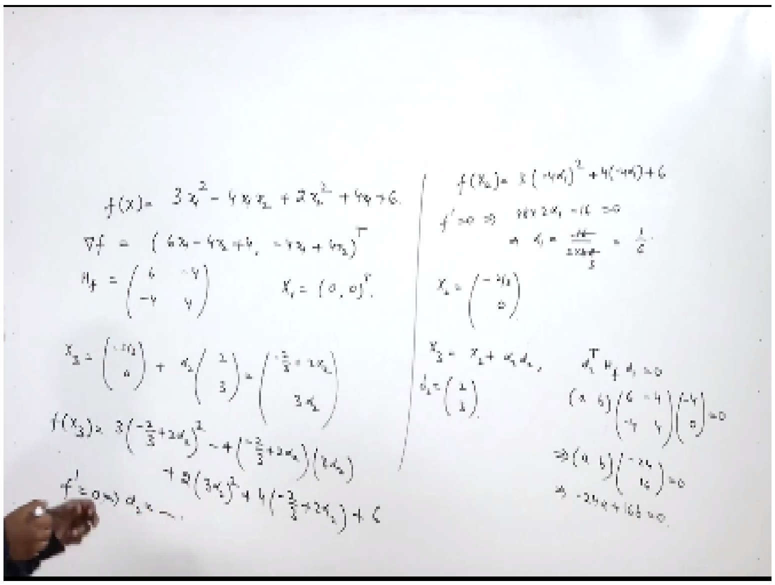
Now initial approximation, initial x is known to you so x_2 will be $x_1+\alpha_1 d_1$ for the finding of the next iteration x_2 now here for the first direction in this method conjugate gradient method in this method for the first direction we take first direction of negative of gradient of fx_1 first direction we take like this okay.

We move long negative gradient of f at x_1 because we know that along this direction the value of f will decrease okay and to find the optimal step size we substitute this x_2 in the function find out the derivative of f put it equal to 0 that will give the optimal step size α_1 using α_1 and d_1 x_1 is known to us we can find x_2 now to find x3 which is $x_2+\alpha_2 d_2$ to find this d_2 instead of moving long the negative of gradient of f at x_2 we move along to find this d_2 basically d_2 is simply d_2 transpose H of f at x_1 x_2 okay.

And d_1=0 basically we find the initial matrix of f, f is a key one function which we have to optimize okay we find the initial matrix of f at x_2 and we find the conjugate direction d_2 respective to d_1 in this way we find d_2 substitute d_2 over here to find out the optimal step size α_2 we again substitute x_3 in the function put derivative is equal to 0 find optimal step size of α_2 and the process repeats.

That means for the first direction we take negative gradient of fx_1 for the second direction we take a direction conjugate with the direction d_1 with respect the initial matrix h so in this way if you proceed finally obtain the optimal solution with the very faster rate compare to the steeper method okay.

(Refer Slide Time: 23:12)



So let us try this problem basically using conjugate gradient matter for the illustration so let us discuss the problem this fx equals to the problem is$3x_1^2$ $-4x_1x_2+2x_2^2$ then it is $+4x_1$ +6 okay so this is our main problem basically now suppose we want to find out the optimal solution of this problem using conjugate gradient method term okay so first we are finding gradient of fx it is $6x_1-4x_2+4$ and it is $-4x_1+ 4x_2$ ^{T} then what is initial matrix of f is 6 – 4, -4 , 4 and it is definite because it is 6 > 0 and the determination is 24 – 16 8 which is also > 0 initial matrix is positive definite okay, let the initial guess as initial approximation as 0, 0 if it is not given to us so we can take 0, 0 as initial approximation okay, so to find out x_2 it is x_1 + α_1 d_1 where d_1 is nothing but negative of gradient of f at x_1 which is 4,0 okay and x1 is 0, 0 + α1 times – so it is negative of this – 4, 0 and this implies it is - 4 α1 and 0.

So it is x_2, now to find $\alpha_1$substitute this f_2 in the given function put its derivative = 0 find out the optimal step size $\alpha_{1,}$ so what is $f(x_2)$, you replace x_1 by - 4 α_1 x2/0 here so it is 3x -4 $\alpha1^2$ 0, 0 then 4 x 4 α1+x put derivative = 0 this implies 48 x 2 α1 because we are taking derivative also and

-16 = 0, so this implies α1 = 16/2 x 48 that is 1/6 okay, α1 = 1/6, so what will x_2 you substitute α1 here it is – 2/3 and 0. So this would x_2 okay.

Now to find x_3 it is x_2 α_2 d_2 and to find d_2 it is d_2^T matrix of f x d_1 x 0 that means we are now finding a direction which is consecutive to d_1 with respect to ration matrix h okay, so suppose this direction is a, b initial matrix is 6-4 and -4, 4 and the d_1 direction is -4, 0 = 0 so this implies b and it is – 24 and it is 16 should be 0 and this implies – 24 a + 16 b should be 0, so you can choose any ab satisfied in this equation, so if you take a as 2 so be can be taken as 3 , so d_2 will 2,3 okay and again to find α_2 so what will be x_3? Now x_3 will be x_2 is we have just find -2/3 0 + α_2 + 2 and 3.

So that is – 2/3+ $2\alpha_2$ and it is $3\alpha_2$ again we will substitute this x_3 in this equation in this function, what will be fx_3, fx_3 x –(2/3+ $2\alpha_2)^2$ -4 x_1 is – 2/3+ $2\alpha_2$ it is $3\alpha_2$ + $2x_2$ it is $3\alpha_2$ + 4 x_1 is – 2/3+ $2\alpha_2$ and + 6. Now put it derivative = α_2 and this α_2 will give next approximation that is x_3 okay. So in this way we can find out the optimal solution using conjugate gradient method. Now here this finish the course basically non linear programming, I have tried cover the various topics in this course thank you very much.

For Further Details Contact
Coordinator, Educational Technology Cell
Indian Institute of Technology Roorkee
Roorkee – 247667
E Mail: etcell.iitrke@gmail.com. etcell@iitr.ernet.in
Website: www.iitr.ac.in/centers/ETC, www.nptel.ac.in

Camera
Jithin. K
Graphics
Binoy. V. P
Online & Video Editing
Mohan Raj. S

Production Team
Sarath Koovery
Arun. S
Pankaj Saini
Neetesh Kumar
Jitender Kumar
Nibedita Bisoyi

An Educational Technology cell
IIT Roorkee Production

© Copyright All Rights Reserved