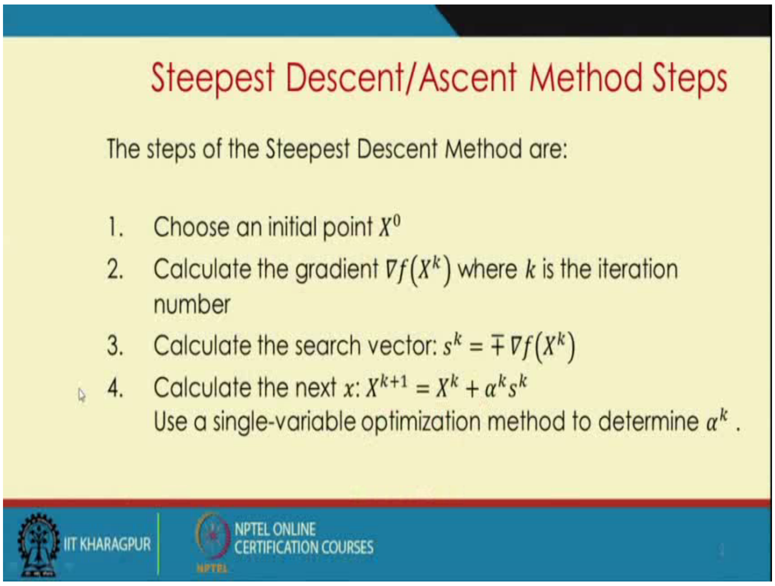
Constrained and Unconstrained Optimization
Prof. Debjani Chakraborty
Department of Mathematics
Indian Institute of Technology, Kharagpur

Lecture – 43
Unconstrained Optimization

In continuation to my previous class, today I will speak more on steepest descent or ascent method. Now there are certain properties of these unconstraint programming technique for non-linear programming. Now there are few things just I would like to mention in this connection.

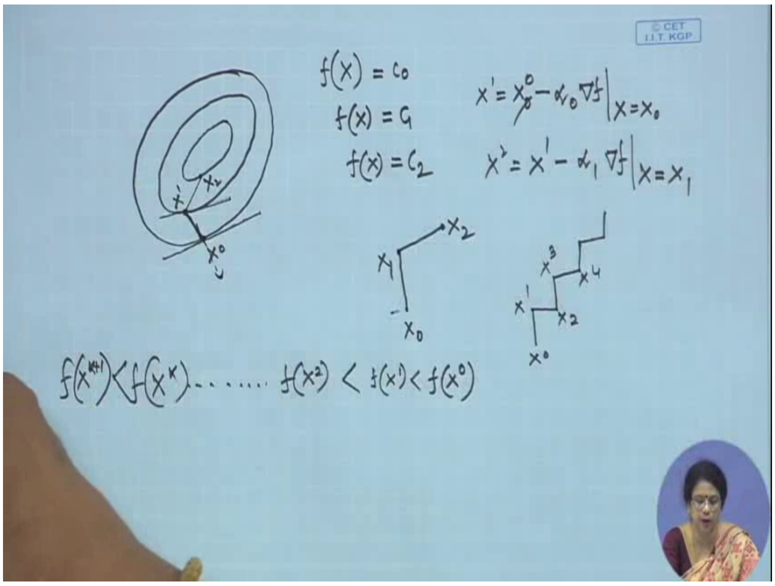
(Refer Slide Time: 00:41)



If I just recapitulate the process once again for you. Now the process is that there is a function given to us, the function is must be continuous in nature because we are using the gradient technique, gradient of a function at a point that is why that is very necessary function has to be continuous.

Now, the process tells us that we will start from initial guess point that is x naught. Now in the x naught once we are guessing that should not be the optimal solution at all for the problem now we will proceed in which direction shall we proceed for the maximization problem we will proceed through the gradient direction from that point and if it is a minimization problem we will just proceed to the reverse gradient direction. That is the idea that is why it has been explained to you that if x naught is the initial point then we will just find it out what is x 1 that is the next point that is equal to x naught, x 1 is equal to x naught plus alpha k, alpha here we have considered the step length at the kth iteration and sk is the direction, sk is positive if it is maximization, it is negative if it is minimization. Now let us consider one example for this. Now for example, we are having a function like this.

(Refer Slide Time: 02:23)



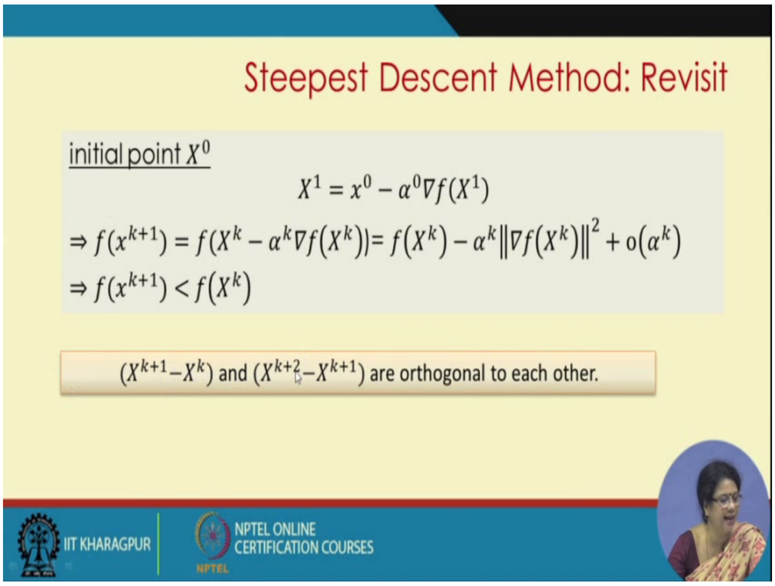
Let this is the function f x capital X is the vector it can be in r 2 if we consider similarly it can be in r3 that is the surface etcetera if I just consider the simplest case that is in r 2 x then this is a level set of the function; that means, f x will have certain value at this position. Now we can have another level set like this where the functional value would be different, since c 1 there can be another level set like this where the functional value would be c 2 ok.

If we consider the initial guess point is here, then if this is my x naught then if I take the tangent at this point then this direction is the gradient direction and this direction is the negative gradient direction, all right. That is why in this point the function will become the gradient direction, actually this is not the case it should be the 90 degree from here and at this point again it will touch to x 2, where after a certain step length that is alpha naught again here there will be again we will consider the tangent again it will move like this, all right. That is why x naught will move to x 1 then x 1 will become x naught minus alpha naught sorry x naught alpha naught, then gradient of f at x is equal to x naught all right x 2 will be is equal to x 1 minus alpha 1 gradient of f x is equal to x 1.

Then we will move to x 2. Now if we consider 3 points x naught x 1 then if we consider x 2, what do see that at this point in the x naught we consider the gradient direction again from x 1 we are considering the gradient direction that is why it will just take a zigzag path like this always to reach to the optimal solution at different iteration x naught x1, x2, x3, x4 like this, all right in connection to this just I will tell you few more things. So, that we can choose steepest descent or ascent method slightly different way and there are certain properties.

First property is that if we just move from x naught to x 1 what is our expectation, since this is the negative gradient direction then we can say that we have to prove that f x 1 must be lesser than f x naught all right similarly f x 2 must be lesser than f x 1. This way we will get a sequence in the x kth point and we will see that f xk plus 1 functional value is much smaller than f x naught maybe that could be the optimal solution all right for checking the optimal solution there is a process.

(Refer Slide Time: 06:22)

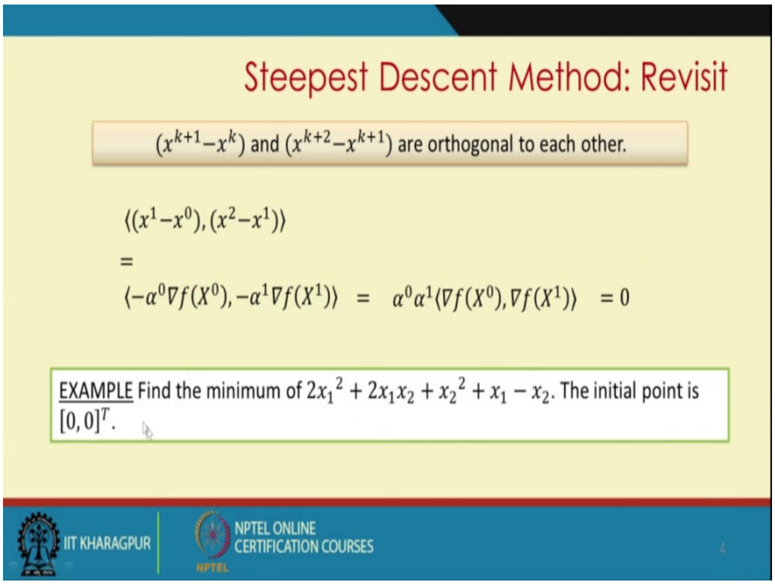


Now, I will discuss more on that part, but in this connection I will show you few properties, as I said my initial point is x naught what we are going to prove we are going to prove that f x k plus 1 is lesser than f x k where k is the kth iteration in k plus 1 is the next iteration and this is the minimum of optimization we are considering. That is why if it is x naught x 1 would be this way and f x k plus 1 would be f x k minus alpha k grad of f x k all right. If I just expand in Taylor series then what we will get here f of x at point xk then we will get f x k minus alpha k, then grad of f x k square plus third order term plus 4th order term that way we can say this is the order of alpha k we will get several terms ok.

Because in the next it would be alpha k square, in the third it would be alpha k cube, why we are considering alpha k because we want to determine alpha k here and where alpha k is the optimal step length. Now from here what we see that you see this is the square term and the step length is always we are considering the positive step length because direction we are considering the negative direction for minimization unless we consider the step length that is the length we are considering the positive step length. That is why alpha k is positive this is positive, that is why we can see from here that f x k plus 1. If we just ignore the other higher order terms then we can say that f x k plus 1 must be lesser than f x k.

What tells we can see from here just look at the picture, we are moving from x naught to x1, x1 to x2. Now here we can say that these 2 vectors are orthogonal to each other, this makes the 90 degree that is why if we consider the inner product of 2 vectors always that will be 0 we can also prove that part. That is why in the next I am going to show you that xk plus 1 minus xk and xk plus 2 minus xk plus 1 orthogonal to each other, if we consider k is equal to here 0 then certainly we can say that x1 minus x2 and x2 minus x1 both are orthogonal to each other we are going to prove it ok.

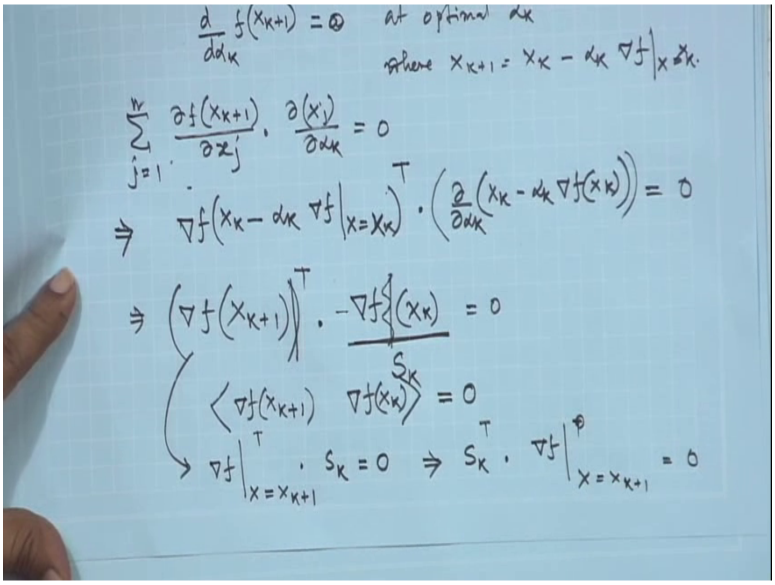
(Refer Slide Time: 09:06)



For proving that what we have to do we have to consider the inner product of 2 vectors, what are the vectors if we just take x 1 minus 0 and x 2 minus x 1 these 2 vectors we are considering and we know x 1 is equal to x naught plus sorry minus alpha naught grad f x at x is equal to x naught. That is why if we can write it minus alpha naught, this difference and this difference we can write with this way all right.

Now from here what we see that this is equal to alpha naught and alpha 1 if we just take outside then we are consider, we are getting 2 vectors because alpha naught and alpha 1 these are not the vectors, these are the scalars these are the step length. That is why 2 vectors if we can prove that the gradient of these 2 vectors are perpendicular to each other rather they are orthogonal to each other then for in general k we can prove it that is why we need to prove that this is equal to 0.

(Refer Slide Time: 10:37)



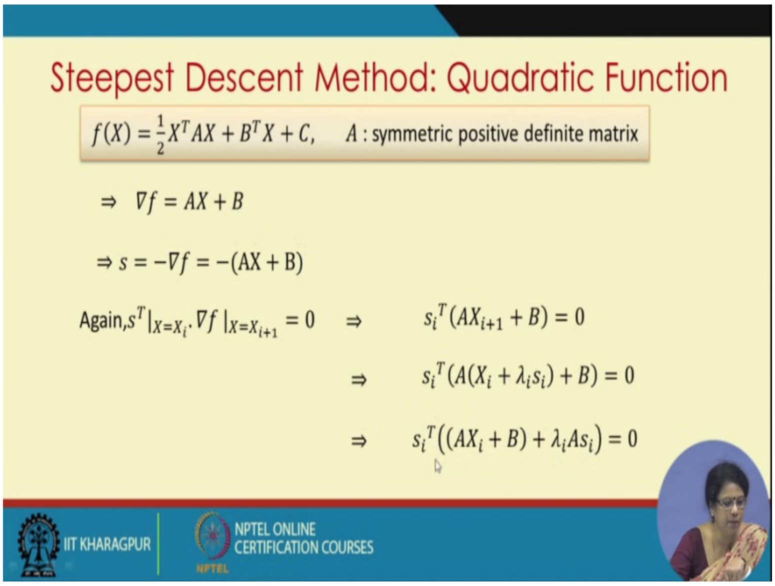
Let us see the calculation for it, what we know that d of d alpha k f xk plus 1 is equal to 0 at optimal alpha k all right where xk plus 1 is equal to xk minus alpha k then grad of f x is equal to xk all right this is the case. Now once we are, this is you see x is a vector there are several components for x we can have n number of components there that is why if we just use the chain rule here we are getting that del of f xk plus 1 divided by del of xj, j is equal to 1 to n into del of xj by del of alpha k that must be is equal to 0. From here we can say through the chain rule all right and this is the necessary condition for optimality of alpha k.

Now, from here we can say that grad, this is the grad value just if we consider this as a vector then what is this value this value is nothing, but grad of f where xk plus 1 that is why we can write xk minus alpha k grad of f x is equal to xk, this is a vector and if we just consider that vector as a transpose and this can be written as del of del alpha k xk minus alpha k grad of f x k. There is no harm to it we can write it this is again a vector altogether all right, what is this vector, this vector is the direction vector I can write let me write it down in the next grid of f xk plus 1 if we just consider the transpose of this vector dot grad of f sorry xk this is is equal to 0 and this term is nothing, but the direction that we are considering as sk.

We can consider minus as well this should be minus of grad fk that is why this is coming as xk for the minimization problem, we are getting from here then what exactly we are getting, we are getting from here that grad of f xk plus 1. If I consider the inner product grad of f sk is equal to 0 clear just you put k is equal to 0 then we are getting grad of f x 0 inner product of grad of f x 0 and grad of f x 1 is equal to 0, am I clear, I do not think I need to repeat once more this one.

Now, from here if we just get this condition then we can say that both the vectors are orthogonal to each other and we can very easily can prove for in general k the same fact all right. Now we use this concept for this problem, just after modifying steepest descent method a little bit I will tell you what is the modification in the next.

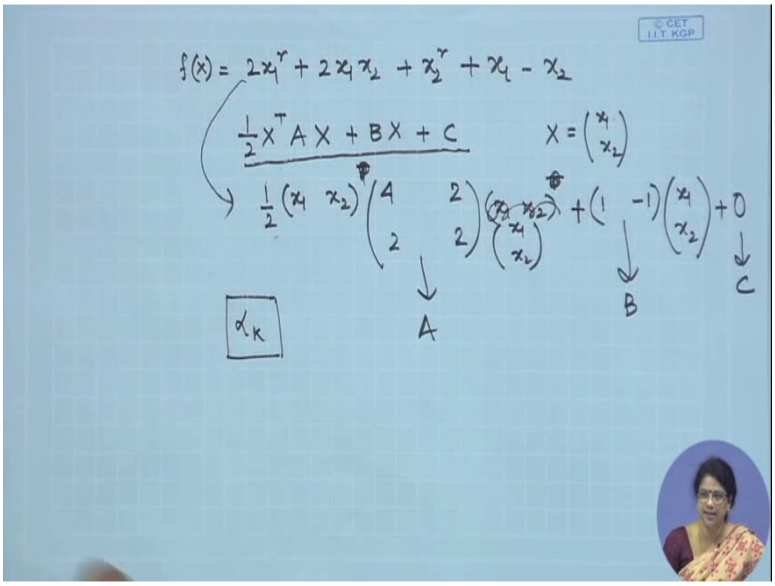
(Refer Slide Time: 14:58)



Now, you see whenever we are having any non-linear function as you have learned in the interpolation method that any non-linear function can be approximated by the quadratic function.

Why we are more interested in optimization about the quadratic function because quadratic function is of order 2 that is why whatever optimality we will get for that quadratic function that would be global optimal. That is why if any non-linear function can be approximated nicely with a quadratic function and if we find out the optimal value of that quadratic function then very easily we can say that that local optimal is the global optimal that. So, I the steepest descent method has been redefined for the quadratic function by employing the concept I explained to you, what is a quadratic function? Quadratic function is a function of degree 2 any quadratic function I can just name a 1 quadratic function like this.

(Refer Slide Time: 16:07)



Fx is equal to say just I showed you before that is a quadratic function just to see 2 x 1 square plus 2, x1, x2 plus x 2 square plus x1 minus x2 this is a quadratic function of degree of 2 variables and of degree 2.

And in general any quadratic function can be written in this fashion, where a is the positive symmetric matrix, plus Bx plus C here the vector x has been considered as the 2 dimensional vector in are 2, x1, x2 that is why this function can be written in this form as 1 by 2, x1, x2 if I consider the since we have taken half here. That is why the a can be constructed from these 3 components which are of degree 2, x1 square would be 4 because 4 by 2 would be 2 then it will be 2 because I will separate 2, x1, x2 into 2 parts 1 is the 2, x1, x2 plus 2, x1, x2 divided by 2 that is why these 2 terms will come like this.

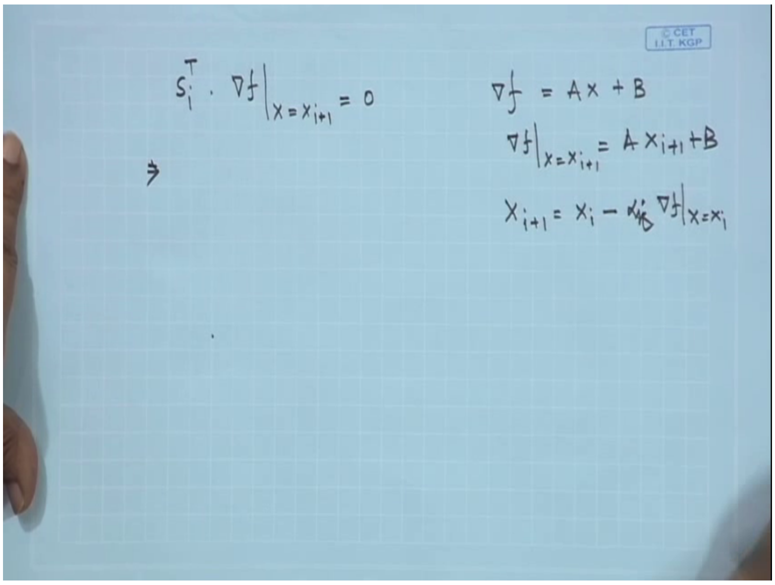
And the third term will be here its 1 that is why we will put 2 just double it because of this half as I say it and these matrix would be 1, minus 1, x1, x2 plus 0 all right that is why you see this if I can write this function in the quadratic form where this is my a, this is my b, this is my c, all right this must be x1 x2 transpose all right yeah that is transpose if I consider x1, x2 this way this is transpose that is all right and no this is transpose and that is it must be x 1 x 2 all right then it is a in this form.

Now, as I told you that any non-linear function can be approximated by the quadratic function that is why we can revisit that thing in this way. I should write not a I should write this as a this way forget about transpose all right then it is. Now we have considered x, x transpose is this one that much we wanted to say. Now if this is. So, let us try to revisit the whole steepest descent method, what are the things we need to find out? We need to find out the gradient direction for this quadratic function and we need to find out the step length, you will see for quadratic function finding out the optimal step length there is a nice process and that is a function of the gradient of the function at the point xk that part I am going to show you show you in the next.

Just you see if we consider the function as a f x is a quadratic function grad of f would be is equal to x plus b there is no doubt about it therefore, any direction for the minimization problem minus grad f that is reverse direction is it is equal to minus ax plus b, from here we can see that st x is equal to xi this is equal to 0. Will you object with it because just now if you remember we where proving this fact that, this fact we proved what is that? Grad of f x is equal to from here x is equal to xk plus 1 this is the transpose dot sk is equal to 0 all right otherwise if I just take the transpose of it of the whole case then we are getting xk dot grad of f t x not t we are considering the t here all right. We are taking the transpose, transpose of transpose would be the same is equal to x k plus 1 is equal to 0 we can say it what is xk, xk is nothing, but s t x is equal to xk we are considering the separate index here i.

That is why from here we can directly say this thing all right once we could say it you see again let me just elaborate these calculations for you sit.

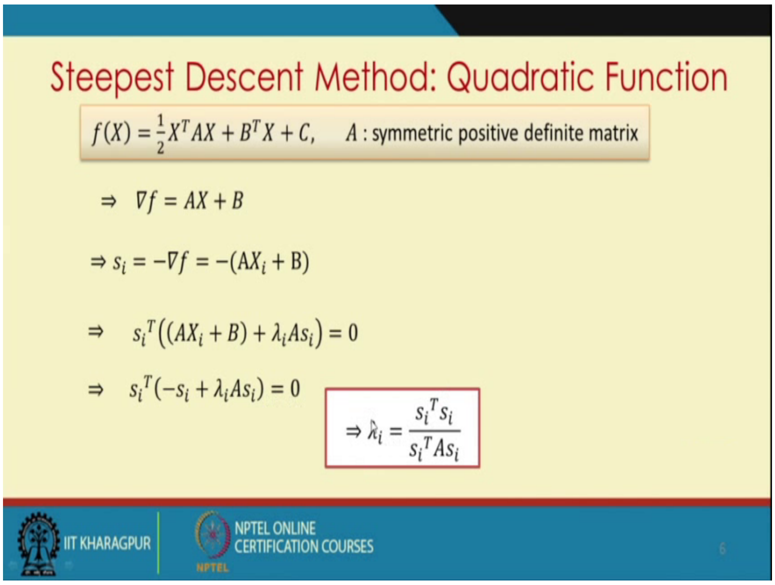
(Refer Slide Time: 22:19)



If I write it sits x equal to x xi grad of f at x is equal to xi is equal to 0 it implies that if you remember, just now we proved grad f is equal to ax plus b all right that is why grad of f at x is equal to xi would be, here we are considering I plus 1. I plus 1 it would be ax I plus 1 plus b no problem that is why here we can say xit, a xi plus 1 plus b is equal to 0, what else we know? We know that xi plus 1 is equal to xi minus optimal alpha k into grad of f at x this is I sorry, x is equal to xi we know that that is why if we just substitute that fact here then we can write it a of xi plus 1, instead of alpha here it has been considered as lambda this way there is no harm to it all right here si is it is equal to minus grad of f the same thing ok.

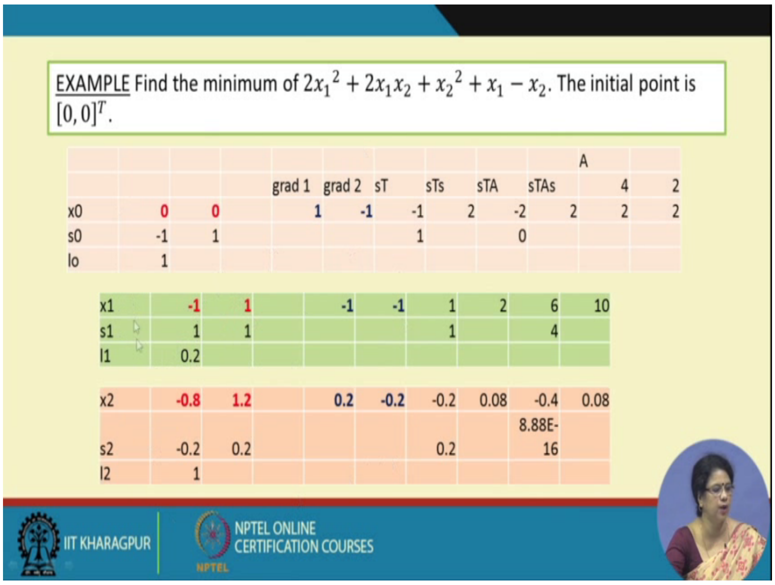
Now, if we just readjust state we are getting this fact axi plus b together and this is 1, then from here we can calculate 1 the value for lambda I lambda I or alpha I whatever you want to say that is the optimal value of step length in the at the I th iteration all right.

(Refer Slide Time: 24:10)



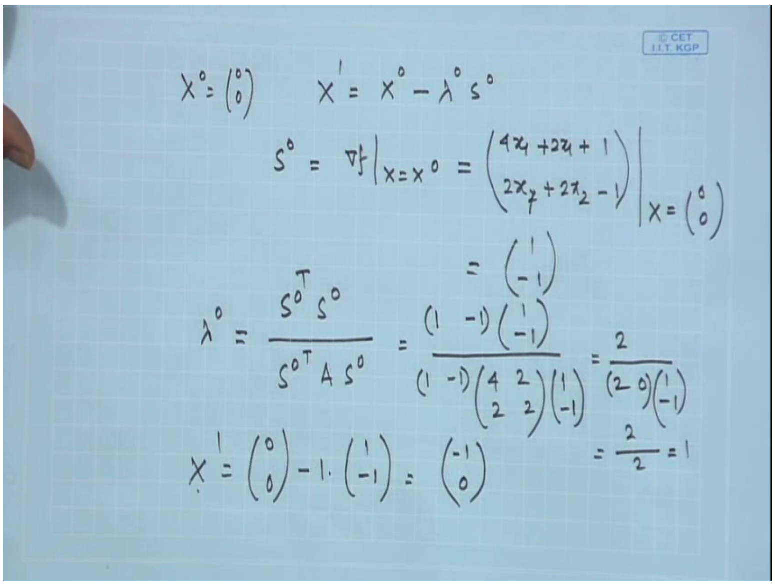
This can be calculated as this way, we can very easily we can calculate this as this 1 just, I repeated the same calculations for you and from here you see we are getting the optimal value of the step length at any iteration is equal to si to the power t s is the, s is the gradient direction negative gradient direction for the minimization problem. Transpose of feet si divided by sit, a is the positive definite matrix that symmetric matrix I was referring for that quadratic function si all right. That is why we need not to formulate the optimization problem I showed you before in the last class to find out the optimal step length we will simply calculate this formula and we will get the solution of it.

(Refer Slide Time: 25:08)



Now, for this problem if we just start it from 0, 0, the starting point it 0, 0 then my x 0 is equal to 0, 0, what is my x 1? x 1 would be is equal to x 0 minus here.

(Refer Slide Time: 25:16)

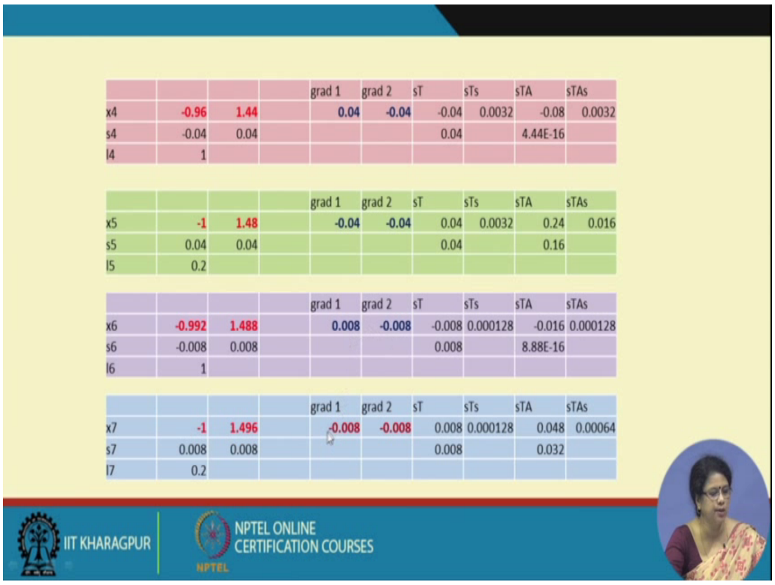


We have considered the lambda 0 and s 0 what is my S 0? S 0 is equal to grad f at x is equal to x naught, what is my grad f from here the function is 2x1 square plus 2x1x2 this way that is why the first component of grad it would be the partial derivative of these with respect to x 1, that is why it is equal to 4x1 plus 2x1 plus 1 and the another component with respect to x2 we are going to consider 2x2 plus 2x, I am sorry, 2x1 plus 2x2 minus 1 this is the grad f at x is equal to 0 0 that is why this value will come as 1 minus 1.

Once we could calculate this value very easily we can calculate the value of lambda 0, what is that? S 0 t s 0 divided by s 0 t A s 0 here s 0 is equal to 1 minus 1, on1e minus 1 divided by 1 minus 1 and here 4, 2, 2, 2 we considered for that function and is equal to again 1 minus 1. If we just calculate the value see 2 plus 1 plus 1 that is coming 2 divided by here 4 minus 2, that is 2 minus 2 plus 2 0 1 minus 1 this is is equal to 2 divided by 2 is equal to 1.

Therefore we can say my next x 1 would be is equal to 0, 0 minus 1 into 1 minus 1, that is why my next x 1 is minus 1 0 clear, this is the calculation for you. Now this calculation I have done through excel just you see x 0 is 0, 0, minus 1 plus 1 sorry this is minus 1 plus one. Now we have calculated the gradient of first component, gradient of the second component and everything we have calculated and the next x 1 we are getting minus 1 1 all right and these are the gradient values for us, gradient of the first term, gradient of the second term, what is the stopping criteria for this whole process, whole iterative process? If we see the respective iterative values of the decision variable that is x 1 x 2 rather capital x is very near by then we will stop otherwise if we see if the grad f in that point is 0; that means, there is no provision for improvement then we will stop our process ok.

(Refer Slide Time: 28:35)



We I have repeated the process 1 iteration, second iteration the same process I did for the 3rd iteration 4th, 5th and you see the values x seven is coming minus 1 and 1.5 almost 1.5. What did you see the gradient values are coming almost 0 here in the previous case also almost 0, even if we just repeat the process we will see the gradient value will come more 0 here that is why the symptom tells me that you can stop the iteration here probably the optimal solution, approximate optimal solution for this problem is minus 1 1.5, why did I say 1.5 because they stay approximated the value.

Now, you can repeat the process. Now this process the calculations whatever I say to you that I have done through excel what is suggest to you repeat the process on your own for the same problem and just do the same thing for other problems which are give, are supplying to you through our assignment sheets you do the same process and see how the iteration is going on. How really we are reaching to the optimal solution and one thing I must mention here, since see this is method this method steepest descent method to be applied for quadratic function and due to the very good nature of the quadratic convergence of this, quadratic function of this function we are getting a global optimal, that is why we can declare that this is the global optimal value for the given problem.

Thank you for today.