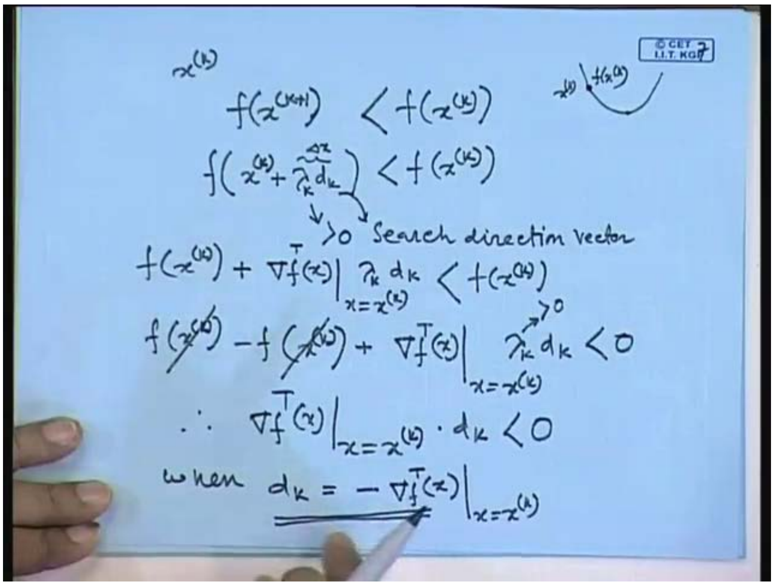
Optimal Control

Prof. Dr. Goshaidas Ray
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 6
Solution of Unconstrained Optimization Problem Using Conjugate Quadient Method and Networks Methods

So, last class we have discussed that unconstrained optimization problem in a numerical technique. And we have seen that solution through numerical technique is nothing but a iterative process. And objective of iterative optimization problem is to find out the minimum value of the cost functions agree.

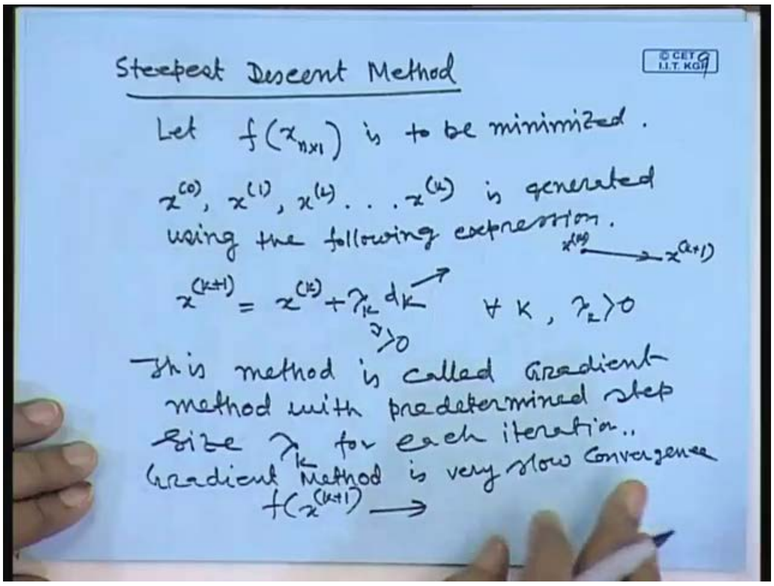
(Refer Slide Time: 01:25)



So, if you just recap what we have done in last class you see that we have a this let us call this is a function we have to minimize it and at kth iteration. There is one point is here let us call x of k and the function value of x of superscript k. Then from kth point we have to move in such a direction, so that the function value at k plus 1iteration the function value will be less than at the kth iteration value function value.

So, you have to move in such a direction we have found out with all these thing. This is the condition the direction of movement from kth instance to k plus 1th instance this is the condition. Then you have to move to in these directions, which is nothing but a minus of the transpose of the gradient of the function at that point. So, this is the condition and we have seen that.

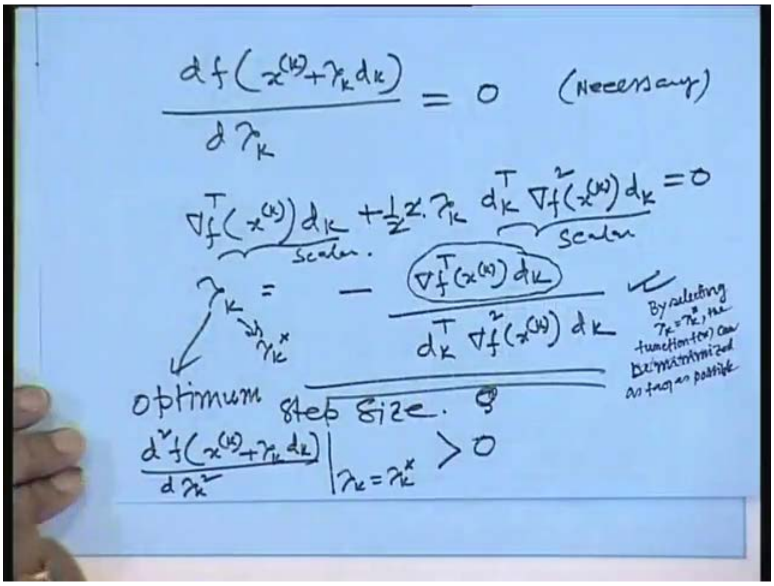
(Refer Slide Time: 02:30)



If we have a function f which is a n variable functions that x 1, x 2 this function is to be minimized and we have x 0, x 1 dot dot x n . We have generated to this expressions and this d k is the direction of that we are moving from kth instance to kth iteration to k plus 1 iteration. And lambda k is scalar factor whose value is greater than 0, so in this way we have to move. And when this lambda value is pre determined then this algorithm the way we are moving the direction search direction. Then will call this is a called a gradient method on d, but lambda k is predetermined when lambda k is optimized agree. And this optimization can be done through any numerical technique any minimization technique.

Because the function when will put this x is equal to k plus 1 iteration that function value then whole function will be a function of lambda k only single variable. So, we can find out the optimized value of this function agree for lambda k is equal to lambda k star. Such that the function value will be minimized, so we can get it the minimum value function as far as possible with the choice of lambda k is equal to lambda star. So, that is we have we have how to find out optimal step size of the, this one we have derived that this is the condition for to get the optimum size of the this step size.

(Refer Slide Time: 04:21)



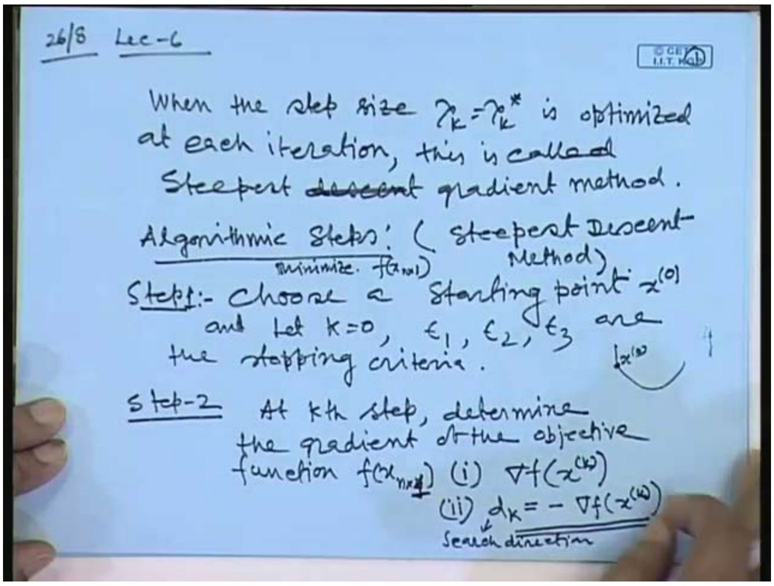
When will move from kth iteration to k plus iteration agree, then which direction it will move that is d of k. And what will be the step size and that one and that lambda k we have found out by this expression. The gradient transpose gradient of the function transpose into d k divided by that quadratic function, which is a d k transpose that gradient of the function. Once again you have differentiated with respect to x second partial derivative for the function into d k will give you that direction of this d k is the direction of the search direction with preceding minus. That is the choice of lambda k and lambda k is greater than 0 and look this expression this is nothing but a condition.

That whose direction you will move it, so that function value will decrease from kth instant to k plus 1th instance. The function value will be this is the condition and that will be less than 0. So, less than 0 means negative and that will be positive, so lambda k is a positive and this value that if you once again if you differentiate this is the derivative of function derivative of function with respect to lambda k we got it lambda k value.

And if you put this value you have to put this lambda k value in the second partial derivative in the second derivative of f of f of x. Here x is equal to x k plus lambda k d k with differentiating with a respect to lambda k square lambda k square. Then you will see if you differentiate this one forget about this one this is the what do you call derivative of the function f, which is a function of lambda k only.

Once again you have differentiated with respect to lambda k, so it will be remain only that part this part will remain because lambda k differentiate only this one. And this quantity you see this quantity is a, what is called positive definite this one. If it is this positive definite, then we have reached the final function value as far as possible minimum at that choice of lambda k. So, this is the so far we have discussed, now lets us see I told you when that if you see the this one.

(Refer Slide Time: 07:01)



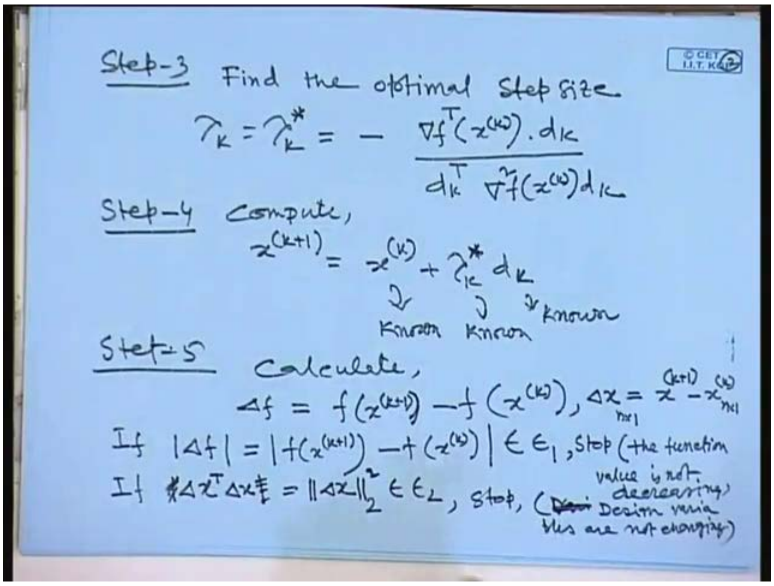
When the step size is optimized at each iteration when the step size lambda k is equal to lambda k star is optimized at each iteration. We call it is a steepest descent method at each iteration then this is called steepest gradient method. So, lambda k we have to optimize, so that the function value will be whose a function value means the function value of lambda k, which will be as far as possible minimum the function value. So, let us see this one algorithmic steps for algorithmic steps for steepest descent method agree. So, what is this first step to get the starting point of our iteration that means initial guess of step 1 choose a starting point. Starting point x super script 0 super script indicate iteration number of iterations that 0th iteration or 0th starting point.

This value of x is equal to x superscript start that to chooses and let k is equal to 0. And epsilon 1, epsilon 2 and epsilon 3 are the stopping criteria are the stopping criteria of the algorithms. This is step 1 and step 1 once you know this one at that point let us call this is the function I know this point x of super script. This at this point you find out this gradient. Step 2 at kth step interval at kth step determine the gradient of the objective function f of that is our objective function which a n cross 1 variables. So, our job is if you see minimizations of a that objective function and our objective function is f of x. This one minimize this is our objective minimize this function using steepest descent method.

Descent means one iteration to another iteration when you will move it the function value should decrease from the previous just previous iterations. The function value should decrease is a descent value of functions downward able. So, the objective determine the objective of the gradient function. That means in other words you can say one the gradient of this function calculate at kth instance. This one and next what is the your what is called the gradient we got it at k is equal to kth iteration k superscript of k again x superscript of k second. You calculate the descent direction that whose direction you have to move, so that the function value will decrease is from the just previous value of the function.

So, that is d k d k is equal to minus delta f of super script of k, that is the search direction, so this is called search direction. So, that is we have to find out in Step 2 then one step three. Once you know this value then you can find out what is the value of our decision variable x at kth k plus 1, iteration that we can easily find out agree.

(Refer Slide Time: 12:57)



So, step 3 but in that expression you see we have a x super script k plus 1 is equal to x k x superscript k plus lambda k into d k, d k we have calculated x superscript k we know that what is initial or kth iteration. The value of this decision variable and lambda k and lambda k is if you determined pre determined it is a gradient descent method. Gradient method in short or if use the steepest descent method, then you have to optimize the value of lambda k.

In other words you have to optimize the step size of the scalar quantity lambda k. So, how to do the step size of lambda k, you have to minimize in what sense that, so that function value at lambda k is equal to lambda k star. Some value of lambda k star the function value will be as far as possible minimum in that iteration and that agree.

So, this we have shown it in our last class that how to find out the optimal step size of lambda k. Find the optimal step size if we recollect we have derived this one by putting the value of x is equal to x super script k plus 1. And where x superscript k plus 1 is equal to x super script k plus lambda k into derivative d k that is us directions. So, that we have got it last class we remember it is nothing but a the gradient transpose of the function at kth iteration.

This function transpose into d k divided by d k transpose and the Hessian matrix of the function at kth iteration d k and this is the optimum size of the lambda k, which will give you the function value as far as possible minimum at that iteration. So, step 4 once you know this one I know what is this value of x superscript k plus 1 is equal to x superscript k plus lambda k star. Note that each iteration that lambda k value will change agree we are optimizing the function value which is the function of lambda k only agree. So, d k this is known this is just in step 3 we have calculated known and this is the our k iteration the decision variable value that is also known. So, you know k plus 1 iteration what is the value of decision variables this, so this way you have to repeat this process. Then where when you should stop this process means which process that is the iterative process when you will stop it.

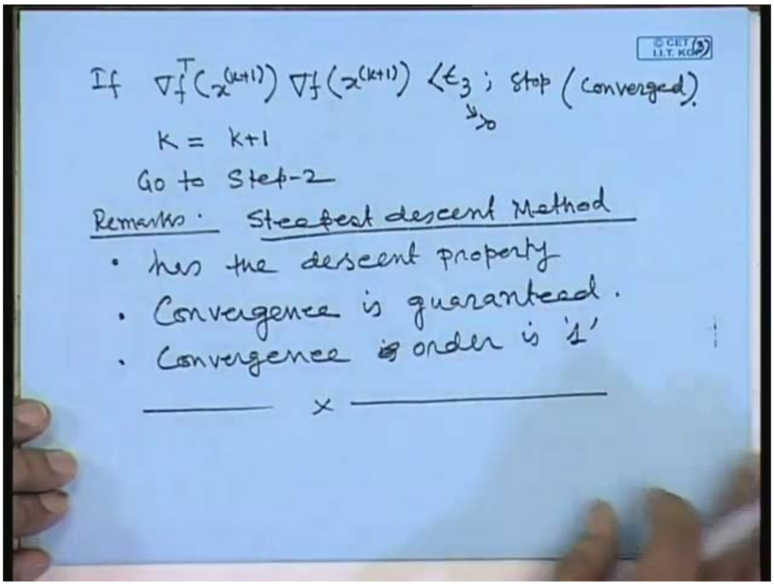
Step 5 here you can write as compute then step 5 is calculate that delta f mind it delta f is what the change in function f at k plus 1th iteration minus function value at kth iteration. So, it is a x k plus 1 the function value minus the function value at k the iteration, so this is delta f. Also calculate delta x that this is the value of decision variable at k plus 1th instant minus x of super script k kth instance, so this is delta x. So, this is a scalar quantity as you know his is a vector n plus 1, because our decision variables we have a n variables x 1, x 2 dot dot x n. So, we have a, we can stop if we like this way if mod of absolute value of f delta f. Since, it is a scalar quantity we are taking absolute value of that one is equal to nothing but a if you see f of x super script k plus 1.

Agree bracket is here this 1 minus f of x superscript k bracket this one absolutely this. If it less than epsilon 1 which is a pre assigned value that value is a positive quantity very small. If it is less than this one stop it indicates that the function value is not decreasing agree. Now, it indicates the function value the function a value is not decreasing it may be another criteria if delta x delta x is a vector of dimension n plus 1. So, we have to write that epsilon or the absolute value of that one agreed or it is a nothing but a distance of the vector. This from k plus 1th instance what is the decision variable to minus kth instance, what is the decision variable value this difference is delta x. So, if you just write delta x transpose of this one, this one is a positive quantity.

So, need not to write that one or in short I can write it this one it will use the norm delta x norm epsilon norm it is called epsilon norm square if it is less than epsilon 2 again the epsilon 1 and epsilon 2 and epsilon 3. We have considered in the at the beginning that means step 1 they are pre assigned value which value is very small. If it is less than this one stop it this indicate that the decision variable are not changing much. So, this indicate that decision variable again the variables design variable or decision variable design variables are not changing this. And you see there may be some function which function value is changing very slowly agreed. But, it has not reached to the optimum value of the function, so if you stop this one then you are land up with a wrong answer.

So, there may be a some situation delta x that delta x when it is changing very small from one point to another point. And what function value may be is very large function value may be very large. So, if you stop it here then there is a problem, so one can do the both the criteria combinely taken into account to stop the iterative process. So, another criteria is convergence criteria you can say that.

(Refer Slide Time: 21:28)



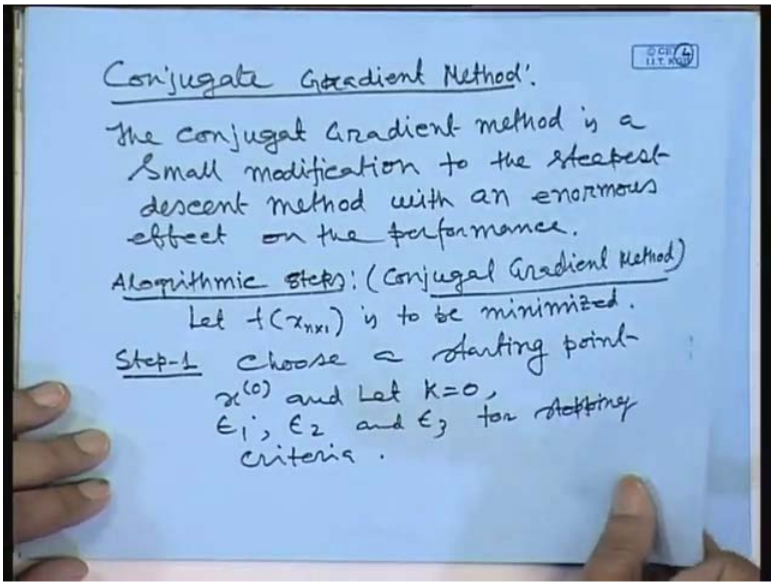
If that gradient of that function at k plus 1th iteration again k plus 1th iteration into gradient transpose and that x of k plus 1 transpose. This quantity is a gradient of function is a column vector and transpose is a row vector. This two quantities will be a scalar quantity and that scalar quantity is a what is called positive quantity. Just like a norm of a vectors, so this quantity is less than epsilon 3. And epsilon 3 is a positive quantity you can write this is greater than 0, but very small quantity again then stop that means it is that our algorithm converged. So, here you can all these I mentioned it this will be a positive quantity which is very small lambda epsilon 1 and epsilon 2 greater than 0 positive quantity and very small.

So, this is the criteria of this one, so you have to check each iteration of this one. So, once you check it then your iteration is updated k plus 1 is updated to this k. I am writing this symbol by k then what will go you go to step 2 here you have to come back see this one the. So, that k plus 1th iteration that decision variable value at k plus 1th iteration x to the power of k plus 1, that you know it. Now, you come to the step and proceed step 2 to what is called step 5 until unless this criteria is satisfied. So, this the iterative process agree, so this is the algorithm. So, remarks one the remarks on the steepest descent method agree, so one we can write that the steepest descent method has the descent property.

Why because we have selected the search direction again in such a way that condition is satisfied. What condition we are showing if you recollect that gradient of that function transpose into d k, d k is the gradient must be lesser than 0. So, with that condition we have selected the d k, so that is why he descent property is satisfied. That means at kth instance kth iteration what is the value of the function again, that function value is greater than at k plus 1th iteration. The function value again that means the function value at k plus 1th position function value is less than the function value at kth iterations Agree this is, so that descent property is satisfied second is convergence is guaranteed.

And third is regarding convergence rate convergence order is 1, so this is the thing. So, convergence is very slow for this one, but it is much better than the gradient method again. So, next will see what is called the some other technique which is a much faster than the steepest descent method that is called what is called conjugate gradient method.

(Refer Slide Time: 26:27)



We have noticed that our crucial factor is the selection of search direction. And that will our I mean main point how to select the search directions, so that the function value at k plus the function value at k plus 1th iteration will be less than the function value at kth iteration, but it how first how the function value as far as possible will be a minimum agreed. So, our conjugate gradient method is just a steepest descent method only a minor modification. And helps our performance of the algorithm drastically it improves it improves performance of conjugate gradient.

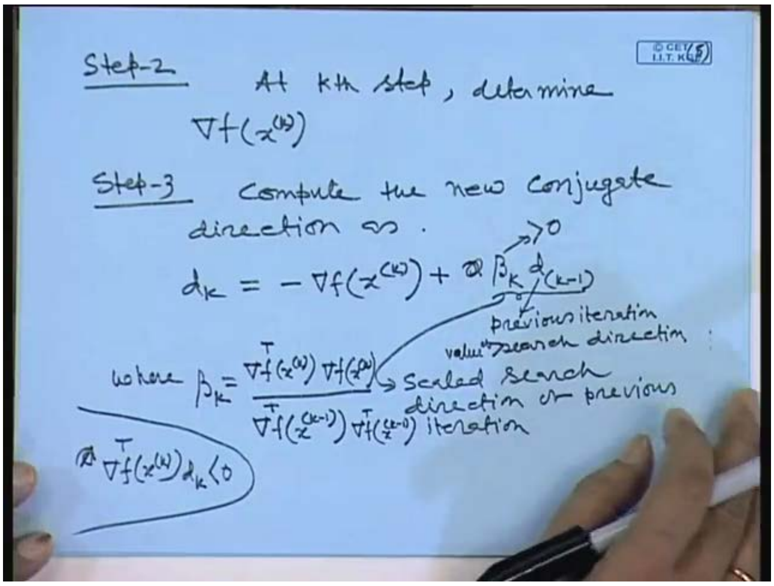
Just slightly we change the algorithm in the, what is called in the steepest descent method. So, the conjugate gradient method you can say just small change in the steepest descent method. The conjugate gradient method is a small modification to the steepest descent method with an enormous affect on the performance. And the what sense the enormous affect on the performance the rate of convergence of the conjugate gradient method is much faster than the steepest descent method. So, what we made it here we took the some history of the previous search direction to find out the present search direction that is we are taking into account.

So, if you take this into account we can write it now straight way the algorithm this way. So, I will just explain through algorithm that what is the conjugate gradient method and that rate of convergence is faster than the steepest descent method due to only one reason. That we are taking the history of the previous steepest descent previous search directions while we will competing the search directions algorithm. Algorithmic steps for conjugate gradient method, so our problem is minimize this function of f x. Let the function f x which is a function of n cross 1 is to be minimized that is our problem.

Because why we are calling it as always minimize because it is a descent direction we are moving from kth iteration to k plus 1th iteration in such a way the function value at k plus 1th iteration is less than the function value at kth iteration. And it is descent direction, so it is minimization problem where if you do move like this way we will achieve to the minimum value of the function. So, this idea can be applied for maximum value of the function. How just maximum value of maximization of the problem can be reformulated into a what is called if I multiplied by minus 1 of this cost function which is the maximization of the function, then after multiplying by minus one you can think of it as the minimizations of the minus of f x, so that we can do it.

But, for all this time is minimization because we are looking for the search direction in the descent direction. Search direction is the descent direction, so this we have to minimize. So, first step is same as earlier choose the starting point x of superscript 0 and let the same steps k is equal to 0. Then epsilon 1, epsilon 2, epsilon 3 for stopping criteria this is same as that what is called the steepest descent method. Step 1 and step 2 you will see what we are doing is this one I told you there is a small modification in steepest descent method.

(Refer Slide Time: 33:38)



Then what is the modification step 2 at kth iteration. At kth step determine gradient of that function at kth iterations means at kth iteration what is value of x value put it in that gradient of that functions. This again step 3 compute the new conjugate direction as d k is equal to minus del f gradient of that function at kth iteration. Look if it is there it is nothing but it is our what is called steepest descent method. But, we are adding some other term that will take the history of the previous iteration of search directions again lambda k we will use a different notation beta k into that d k minus 1. So, it is previous iteration information of search direction, so this is the you can this is the you can say the previous iteration value of search directions.

So, this and this beta k is a constant which is greater than 0 this one, so all this two products is nothing but a scaled. This together it indicates the scaled of previous iteration search direction agreed. So, this total product is nothing but a scaled search direction of previous situation agree. So, this now you see naturally I am taking some information of previous iteration search direction and adding with the present iteration search directions. So, this will this d k will improve the rate of convergence of this algorithms agree. So, where beta k is equal to how it will be delta f transpose of super script k this into delta f x super script k of this.

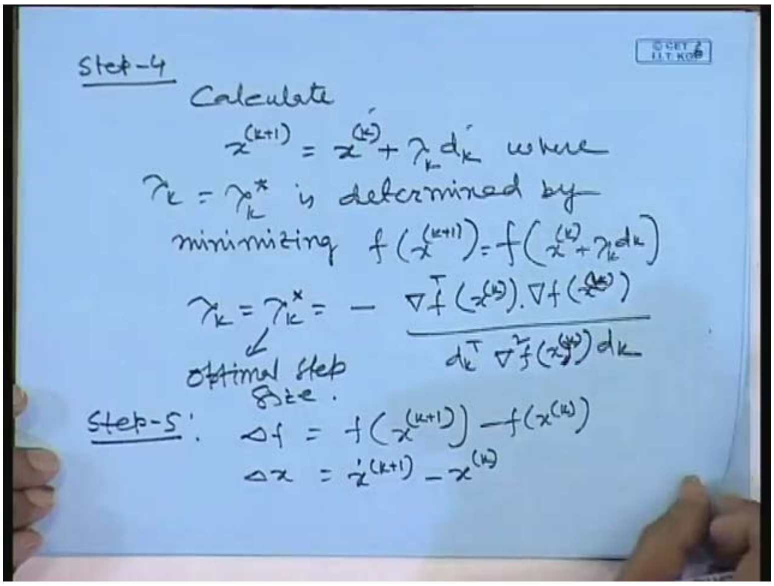
So, this a scalar quantity of this one that means you find out the gradient norm of his gradient kth iteration what is this one you find out norm square. This one this divided by previous iteration with kth iteration, and take the previous iteration that is f of x k minus

1. This then delta f of x k minus 1 this iteration that beta k, so this is a positive quantity. This is also a positive quantity results is a positive quantity, so beta k is you can write it as greater than 0 agreed. So, if you look this at this expression because what is the guarantee that d k if you move in the direction what is the guarantee that we are moving in the descent direction. That means function value at k plus 1th instant k plus 1th iteration the function value will be less than at kth iteration.

The function value of kth iteration what is the guarantee, so this value you can check it by multiplying by both sides by gradient transpose of that one. That is what is the condition we know descent direction if we want to do this one that our condition is recollect that delta f of x k whole transpose into d k must be less than 0. The function value will always at k plus 1th iteration will be less than the function value at kth iteration. This will be true provided this condition is satisfied, so if you multiply by both sides delta f transpose of f for x k both quantity multiply. This side with this one this quantity is positive preceded with minus, so this quantity is negative.

So, what about this one I just mentioned beta is greater than 0 and d k in previous iteration also you have seen it is multiplied by delta f k agreed. So, this quantity what is this quantity that condition of that one is it will be negative this one. So, whole quantity is becoming negative that multiplied by delta k this whole quantity is becoming 0. So, we are moving to the descent direction.

(Refer Slide Time: 40:48)



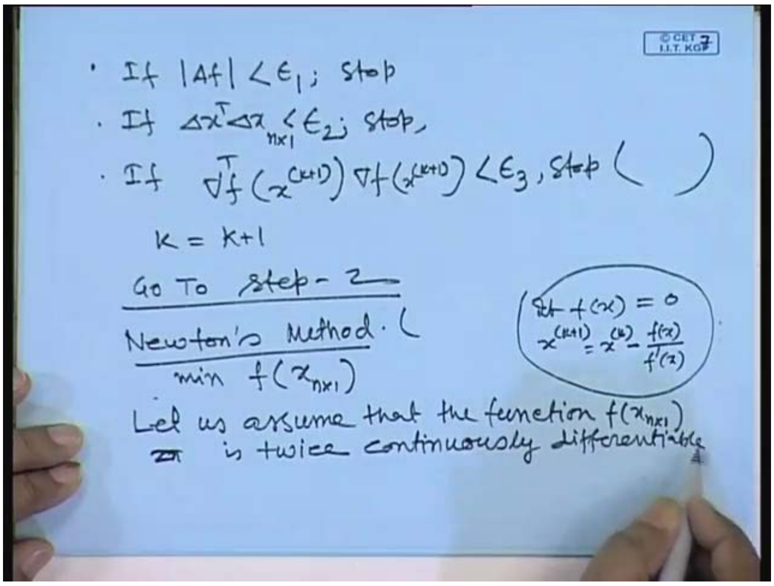
So, this step 3 so this way you update the not update this way you find out the descent direction of that one agree. Next is Step 4 and your see this quantity if you look at this expression this expression it is always delta if you multiplied by this delta f transpose of x to the power of kth iteration. This product both side you multiply by this quantity just now I mentioned it is a negative quantity.

That means, we are moving to the what is called descent directions step 4. Calculate now I can calculate comfortably that value of descent value of variable at k plus 1th iteration is equal to x of superscript k plus lambda k into d k that expression. And in this state we have to find out the optimal choice of step size, so that the function value will be as far as possible minimum at that iteration.

So, that we know how to select the our, what is called lambda k is equal to lambda k star that we know. So, this where lambda k is equal to lambda k star is determined by minimizing f of x superscript k plus 1 is equal to f of x superscript k plus lambda k d k. Now, this is the function of only lambda k that you know what is the choice of that one this implies is lambda k is equal to lambda k star is equal to minus what we derived earlier. That is nothing but x of superscript k into delta f of superscript k this k this divided by d k. This is the Hessian matrix or second derivative of this function at super script k this one agree into d k. So, that at that step you have to find out the optimum step size optimal step size that you have to calculate.

Once you calculated this one then you know because you know this one you know this one this I have calculated. And that will give you x k plus 1 what does it mean with this value that we will move from kth iteration to k plus 1th iteration and the function value will be as far as possible minimum with this choice. So, this next is ours stopping criteria and the stopping criteria is same as we have discussed already here in steepest decent method. So, you define delta f is equal to f x superscript k plus 1, this minus of f superscript of k. And delta x is x super script k plus 1 minus x superscript of k this is delta x and then you just check this criteria’s.

(Refer Slide Time: 44:44)



So, one criteria is if so delta f is the change in value is a scalar quantity I can write delta f of this if it is less than epsilon the stopping criteria which is a positive and very small quantity may be 10 to the power of minus 6 minus 4. As you consider the pre assigned tolerance value agree, so this stop and you know the meaning of this one. The function value is not changing with it from one iteration to another iteration if this delta x transpose of delta x which is an cross 1 if less than epsilon 2 stop. This indicates that decision variable or design variable are not changing. Similarly, if this is a your gradient of that function transpose x at k plus 1th iteration direction this. And gradient of this meaning is that distance square gradient is a vector gradient of the function is a vector.

That vector square that vector distance square if it less than this then stop means it is converged. So, this means it is epsilon is a very small quantity I told positive quantity very small let us get into the four. This indicates the delta f is very small delta f is means delta f gradient of this one is small that means the gradient is equal to 0. That means we have reached to the optimum value of the function f at that iteration. So, this is our algorithms we have a, now what we have to do it that one you update your iterations check a bit iterations. This things and then go to step 2 and step 2 if you see in this conjugate gradient method is that one will come and repeats step 2, 3, 4, 5 and so on.

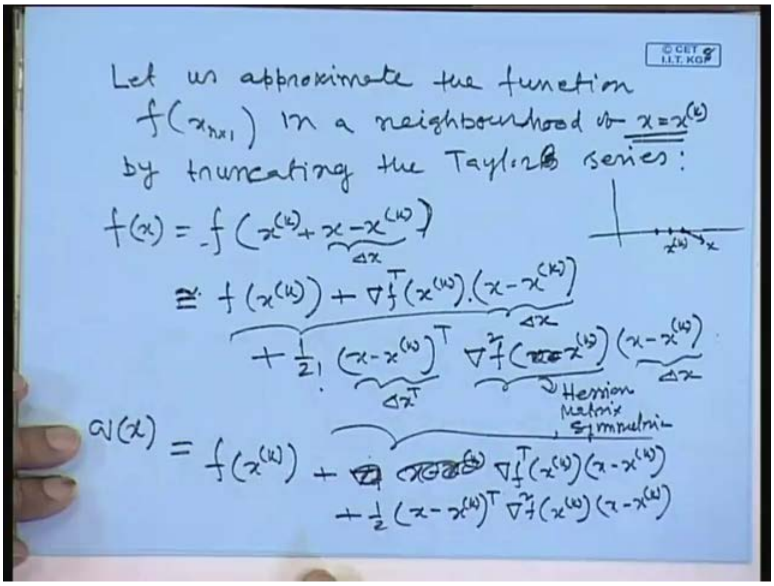
So, our difference from if you look at this one the difference from the steepest descent method is this step. This step 3 means present iteration we are taking the value of what is the gradient k that is the search direction. We are taking it in addition to we are taking the scaled search direction of previous iteration k minus 1 d in your case this iteration we are taking into account. So, this d k will ensure that value of the function if you move like this way reach to the k plus 1th iteration values .The function value will decrease from the function value at kth iteration this indicate this way that I told you multiplied by both side gradient of f transpose. Then this will this see this expression will be less than 0 that is the our necessary condition for descent directions.

So, let us go next is our what is called Newton’s method, so next is our Newton’s method. As you know earlier that if you have a function f of x which is a scalar function. Let us call f of x is a and it is a function of single variable and if you recollect this one that. Suppose, if we are asked to f of x is the function which is equal to 0 I asked to find the roots of this function. Then what we do we can do analytically or you can do iterative process that is one of the iterative process is the Newton Raphson method. That what we do we take some initial guess, then we find out that the input value of this variable x is equal to x super script k.

In earlier iteration that variable value minus f of x k divided by derivative of x k I asked I told you that this is the function of a single variable case agreed. So, this things we can do it and you repeat this one. And when it is converged the same convergence criteria you can put it then it indicates the converse value of that x is the solution or roots of the that function roots of this function or roots or solutions of this function. So, that is we know same concept we can apply here to find the what is called minimum value of the function f of x. So, our job is to now let us assume that our minimize f of x, now our function is a variable of this is a function is variable x is a variable this has a n component is function x 1, x 2 dot dot x n.

So, this function you have to minimize then what we are doing it look this expression. Let us assume that the function our job is to minimize this objective function or cost function using Newton’s method. But, to whole concept I will apply what is the how to find out the roots of their, what is called the function. The same concept we can use it here to find minimize this function. Let us assume that a function f of x which is n cross 1 this function is differentiable twice. That means the second derivative of this function can be computed agree. Assume this function is twice differentiable continuously differentiable and it is also valid for other algorithms also that we have discussed here agree.

(Refer Slide Time: 51:50)



Now, let us what will do just see let us approximate the function f x n cross 1. This function we approximate this function in the neighbourhood of x is equal to x super script k, let us approximate this function in a neighbourhood of x is equal to x super script. This means at kth iteration what is the design variable or the decision variable values at this end we want to approximate this function value. So, that we can do by using by truncating the Taylor series expansion agreed. So, we know that f of x is what I am the approximate the function value in the neighbourhood of x of k agreed. So, that I can write it now x of if you see this one x of this is the thing and then I writing x minus x superscript k, so that I considered as a delta x.

Suppose, we have a x just to scalar case think of if it is a x of super script k neighbourhood of this one is delta x either in this side or this side. That is the I am telling it if you call this as x, so x minus x super script k is delta x either this side or that side agree. So, this I can write this equal to till now nearly equal to this I can write it I am truncating the series after second order. So, I can write f of super script k this plus gradient of this function f agree. Transpose x superscript k into this one x minus change in the decision variables this is what we can write it this is what delta x plus i equal truncate after second order after second order.

So, equal to x minus x superscript k whole transpose then second partial differentiation of this function at x is equal to x superscript k means kth iteration. What is the decision variable value you put it x is equal to this. So, you can write x of this into del minus x superscript k, now see this one this is the nothing but this is approximately that one. So, this is the quadratic function you see this is something like a I can write this is a delta x transpose this is the matrix which is a Hessian matrix, which is a symmetric matrix. Hessian matrix and symmetric matrix and this is a delta x, so this is a quadratic function you can say.

And this is also quadratic function is a something like x transpose p x then it is x transpose that some matrix into x agreed some vector row vector into x. And that is f of x scalar quantity of all this things are scalar quantity, so it is a altogether this is quadratic form of function in term of x. Because x k you know it this is also something you know it a x transpose b because this scalar quantity I can out the transpose both side. Then it will be a x minus x transpose into delta x, so this will be this is a scalar quantity constant known scalar quantity. So, this is quadratic form, so if I consider is a x this whole quantity now I am denoted by this here and here I denoted by f is equal to q of x is equal to f x superscript k plus delta f transpose of or I can write it does not matter.

If I write it like this way agree or let us call write this that way f transpose x superscript k into minus x super script k plus half x of x superscript k whole transpose delta f x superscript k into x of super script k. So, this is the quadratic function, so this is a function of x only of you see because I know x of k value all this things that our kth iteration value that a decision variable value. So, it is a function of only x so I can minimize this function minimizing this function means approximately I am minimizing f of x. So, this function minimizations is what is first necessary condition is gradient of that one is 0 you have to put it. So, then once you get it gradient of that one from there you will find out what is our variable. Here that delta x or x you can say from there you will find out what is the value at k plus one’th iteration decision variables value agreed. So, I will discuss this you this portion next class in details.

Thank you.