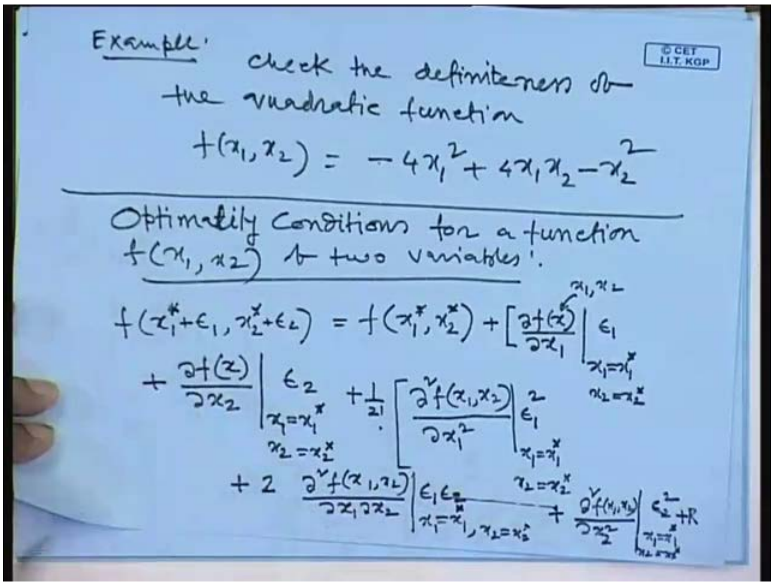
Optimal Control
Prof. G. D. Ray
Department of Electrical Engineering
Indian Institute of Technology, Kharagpur

Lecture - 5
Unconstrained Optimization Problem (Numerical Techniques)

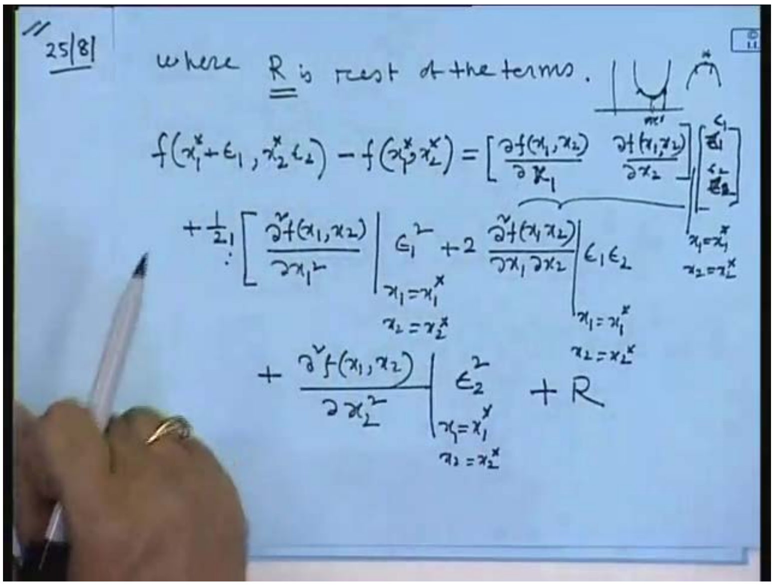
So, last class we have discussed that optimality condition for a function of two variable case.

(Refer Slide Time: 0:58)



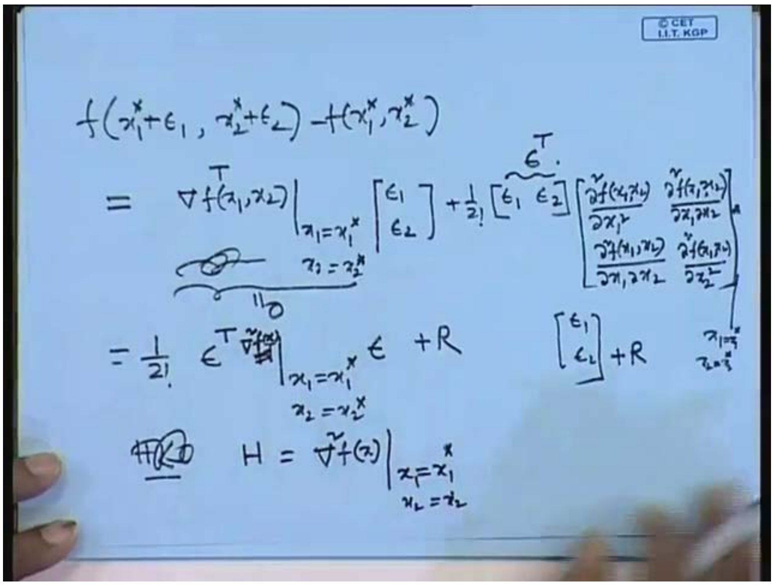
Just we will recap this things what is this one, we have a function which is a function of two variables and x star, x 1 star, x 2 star are the optimal point. Around this optimal point we have given a partavation, with x 1 with part up by epsilon 1, x 2 part up by epsilon 2, that partavation is very small. So, this if you do the Taylor series expansion, we have shown that, have seen this the function value at the optimal point and this is the Taylor series expansion, first order terms and second order terms. And the higher order terms, all other higher order terms I denoted by capital R. This R is sufficiently small, compared to the presiding terms when the partavation is near, around, very close to the x 1 star and x 2 star.

(Refer Slide Time: 01:56)



And if you do the, and if you bring it the, this in the right hand side, left hand side this is the changing function value. From the optimum point we have given a partavation and what is the change in function value, that change in function value is nothing but a gradient of that function, multiplied by that partavation vector. And if you look at this expression, we can see that this term we cannot say whether this, this is a square term, but we cannot say anything about the whether, it will be positive or negative, because epsilon can be positive and negative epsilon 1, epsilon 2. So, this term we have assigned to 0 and it is a necessary condition for this one and remaining terms is that one, that what is the quadratic term? This we can easily write it into quadratic form, these three terms we can write it in quadratic form.

(Refer Slide Time: 02:52)

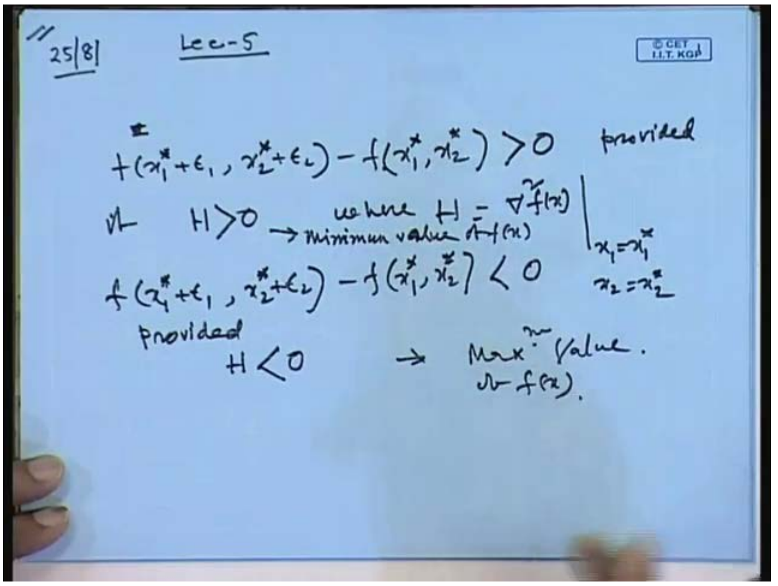


And which we have written like this way, if you see this, we have written this term we assign to 0 and this a half, this we can say it is a epsilon transpose. Epsilon is a vector, which components are epsilon 1, epsilon 2 then this is the Hessian matrix, second derivative of the functions, evaluate at x 1 is equal to x star, x 2 is equal to x star, x 2 is equal to x 2 star. Then this is a, your quadric from and ultimately this is nothing but a you can write it, this is nothing but a gradient this is the what is called Hessian matrix. In the second gradient you can write, gradient of a vector f of x, gradient of a second derivative of this function.

So, f of this an evaluate this value x 1 is equal to x star and x 2 is equal to x 2 star and this we can consider, this can be neglected, that one. So, this function value will be positive because we have assume that x 1 star, x 2 star is the optimum point. So, this function value, if it is a positive, this positive indicates that, that the point x 1 star, x 2 star is a optimum point, in other side of this partavations, optimum point we got it, that means function as reach to this optimum value.

So, this we can say if we consider this, if we consider h is equal to del square f of x evaluated at x is equal to x 1 is equal to x 1 star and x 2 is equal to x 2 star and depending upon the matrix value at this this h. If h is a positive definite matrix, then we can say that 1 by 2, factorial 2 epsilon transpose in to h.

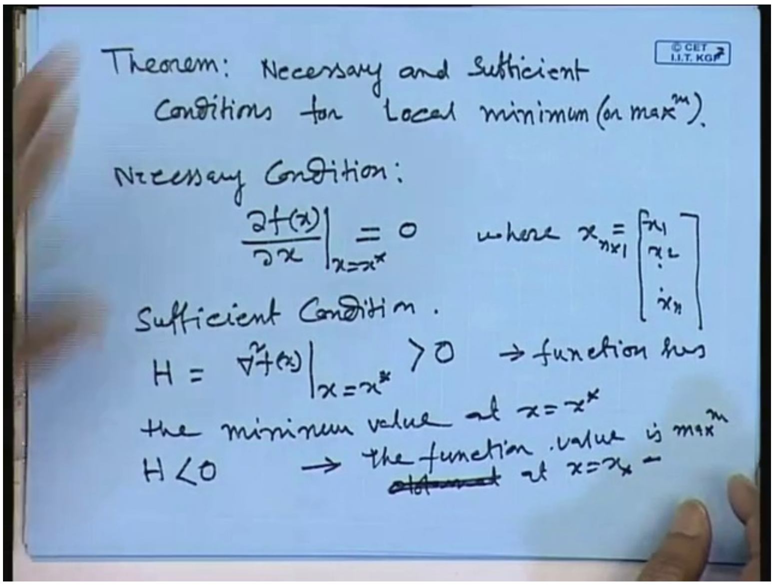
(Refer Slide Time: 5:00)



That means we can write this, that epsilon this equal to f of x 1 star plus epsilon comma x 2 star plus epsilon 2, minus this f of x 1 star comma x 2 star, this will be less than. If it is, if I write this this value is greater than 0, this indicates this will be greater than 0 provided h is greater than 0, means h is positive definite matrix. Where h is, h is nothing but a, the derivative of the gradient or Hessian matrix evaluate at x is equal to x 1 star and x 2 is equal to x 2 star.

If this value is positive definite matrix, this indicates that function value we have reach to the minimum value of this function. If then f, again f x 1 star plus epsilon 1 plus x 2 star plus epsilon 2, minus x 1 star x 2 star this, if it less than 0,if is this a less than 0, this indicates this will be less than 0, provided h is negative definite matrix. This indicates, this indicates we got the what is call, maximum value of the function at x is equal to x star, means x 1 is equal to x 1 star and x 2 is x 2 star. This is a maximum value and this condition is for minimum value of the function, value of f x, maximum value of f x.

(Refer Slide Time: 7:27)

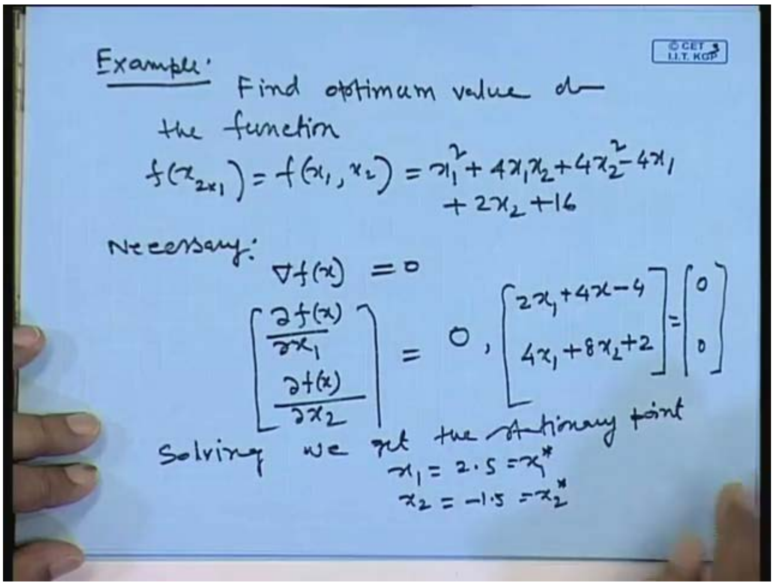


Now we can restate the, our problem that what is call, the necessary and sufficient condition like this way that, theorem the necessary and sufficient condition, sufficient condition, the theorem. The necessary and sufficient condition for function to be sufficient conditions for local minimum or maximum is like this way or necessary condition is like this way necessary condition, first assign the gradient of this vector. That is what is call gradient of this vector del f of del x, with respect to x you assign this is equal to 0. This is our necessary condition, where x is equal to, in general now we write it dimension in n cross 1, which is x 1, x 2 dot dot x n.

So, then sufficient condition, condition is our Hessian matrix that h, h that is equal to del f square del x, x is equal to x 1, x star. If suppose, if this quantity is greater than 0, means positive definite, then function least is minimum value. The function is, the function has the minimum value at x is equal to x star, we can put it here is that, put at this luster ad if I knew, yes.

Similarly, if h is less than 0 means negative definite, this is negative definite, this implies the function, function value is obtained at x is equal to x star. It is he minimum value, maximum value for some function value is maximum function, function value is maximum at x is equal to x star. So, this is the necessary and sufficient condition for a function of n variables x is equal to dimension n cross 1, so this necessary and sufficient condition, for this one. So, let us take one simple example.

(Refer Slide Time: 11:00)

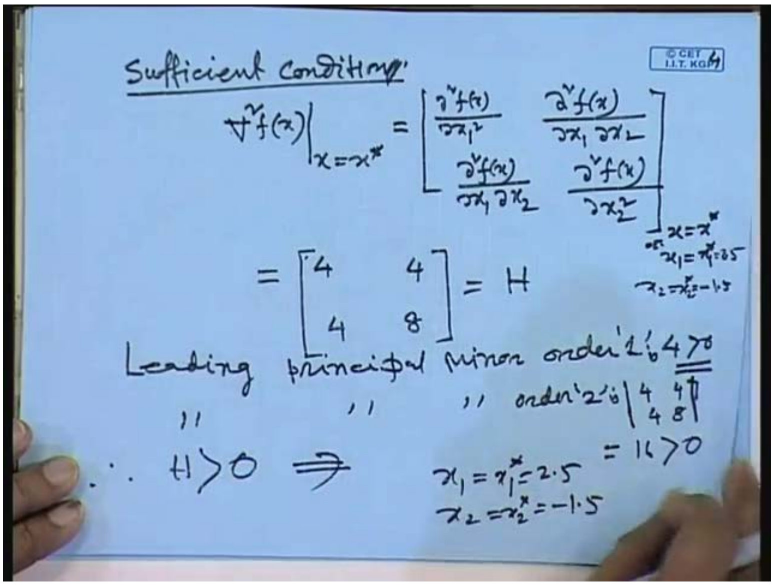


Quickly we just take this example. Suppose you are asked to find out optimum value of this function, optimum value of the function, either it is a minimum or maxim below of the function. So, the function is given f of x which is in our example it is 2 by 2, means it is function of x 1 and x 2. That is x 1 square plus 4 x 1 x 2 plus 4 x 2 square minus 4 x 1 plus 2 x 2 plus 16. So, our necessary condition according to the theorem we proved, necessary condition means gradient of f of x is equal to 0. What is gradient of this one? del f, del x 1, del x 1 and del f, del x 2, this is a gradient of this, is equal to 0.

So, if you differentiate this with except to x 1, it is coming 2 x 1 plus 4 x 2 minus 4, this and if you differentiate the second part f with respect to x 2, then it will be 4 x 1 plus 8 x 2 plus 2 is equal to 0 and 0. Solving this set of equation may two equation, algebraic equation if you, if solved, solving the equations, solving we get the stationery points or we get the stationery point. x 1 is equal to 2.5, let us call consider the x 1 star x 2 is equal to minus 1.5, now it is good to them.

Now, we have to see whether the hessian matrix value of this one, whether it is a positive definite matrix or negative definite matrix or negative semi definite matrix or positive semi definite matrix. So, if you find out the value of the Hessian matrix so this the necessary condition we got, the out stationery point here. At this point the function may be minimum, maximum, positive semi definite, negative semi definite all these things.

(Refer Slide Tim: 14:05)



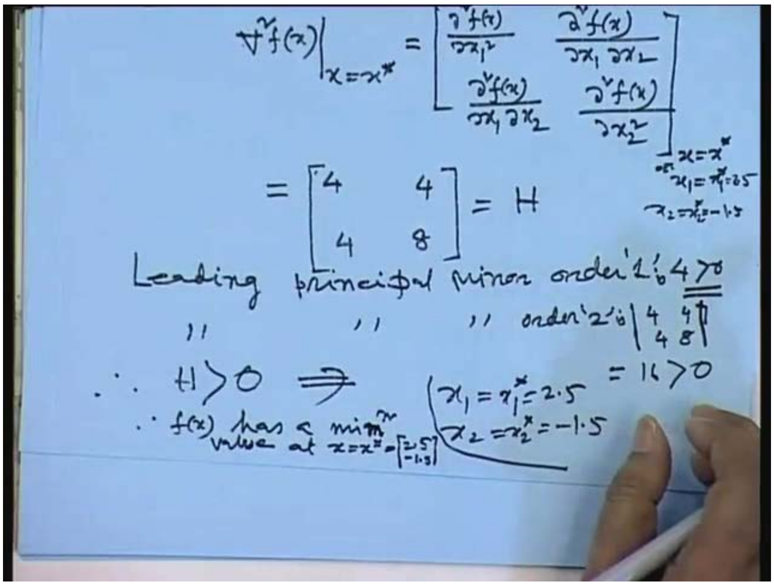
So, you have to further check with the sufficient conditions, sufficient condition. So, our the Hessian matrix are, the differentiate the gradient of a vector, the gradient of the function, gradient of, gradient of, gradient with respect to x, once again or this is the Hessian matrix, we find out the value x is equal to x star. Differentiate the gradient once again with respect to x, which is nothing but a Hessian matrix, this value you find out. If you find out this value, we have already see this one, the what is our Hessian matrix, if you see this one, this is nothing but a del square f of x, del x 1 square del square f of x. This is already we have written so many times so just without explaining in details I am writing the expression for the Hessian matrix.

So, this value you evaluate x is equal to x star, in our case x 1 is equal to, if you see this nothing but r or x is equal to x 1 star, which we got it 2.5, 2.5, x 2 is equal to x 2 star, we got minus 1.5. So, put this value in this one and this values are, if you see this, if you differentiate, already we have differentiated in gradient of a function, we have find out here, we are you differentiate this with respect to x 1.

Again you differentiate with respect to x 2 so this a, if you differentiate this one, that is x 1, this is missing the subscribe, this is the x 2. So, if you do this one and put this limit, it will come 4, 4, 4 and 8 and let us call this matrix is our H. Now we check it whether this is a positive definite matrix or not, since the diagonal elements are all positive so you can proceed further positive definite matrix first.

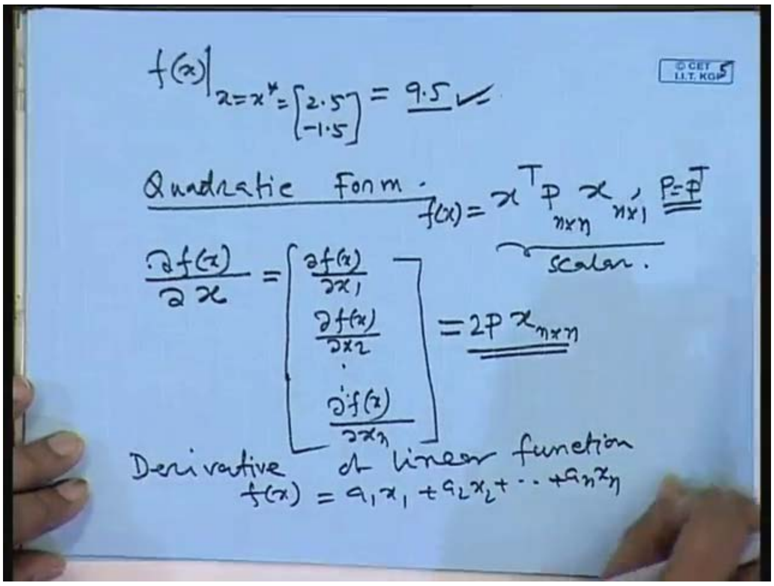
So, leading principle minors, as you know the leading principle minor of order 1 is 4, which is greater than 0, is 4 which is greater than 0. Leading principle minor, leading principle minor of order 2 is matrix itself, the way we are calculating leading principle minors. That means determinate is, determinate of 4, 4, 4, 8, that determinate of this one, which is equal to 32 minus 16, will be 16 which greater than 0. Therefore, our H is positive definite matrix implies, implies what? that function is, attend is minimum value at x 1 is equal to x 1 star means, 2.5 and x 2 is equal to x 2 star 1.5. At this point, the function value attend is minimum value of the function. Since this is a positive definite function, positive definite matrix.

(Refer Slide Time: 17:56)



So, we know at this moment that, what we call, our conclusion is now f of x has a minimum value, has a minimum, minimum value at x is equal x star, which is equal to 2.5 minus 1.5 at this point. And its value is, if you put this value in that expression, f x expression the value of the function.

(Refer Slide Time: 18:35)



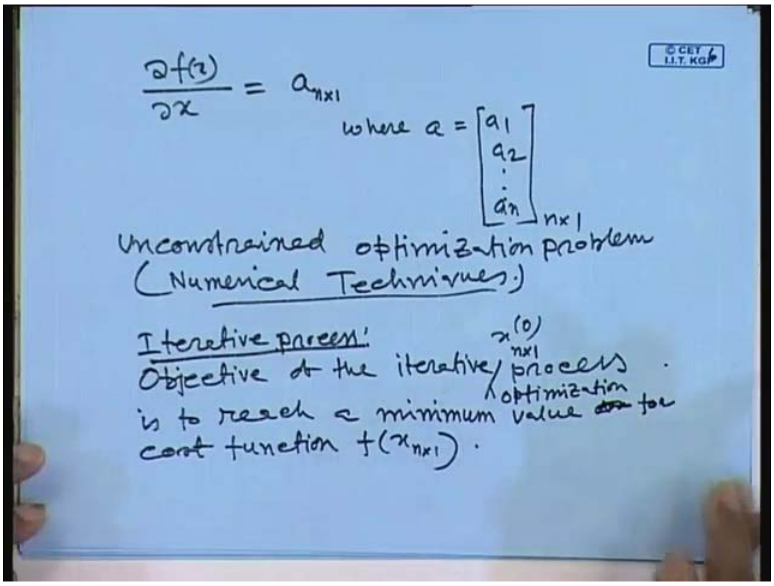
You will get f of x, at x is equal to x star, means is equal to 2.5 minus 1.5, this is the out stationery point we got it, if you see earlier. So, that value will get 9.5, please check it so we know at this moment how to find out the function, which is a more than one variable or multi variable functions of dimension n. How to find out its local optimum or local optimum, means local minimum and local maximum value of the function. So, before we proceed further, first we see that what is call, that if you have a quadric form is there.

If quadratic form, quadratic form let us call x transpose p x, it is a n variables are there this matrix, immediately the dimension is n cross n and this is a scalar form. So, if you differentiate and this is a, is a function of n variables x 1, x 2 dot dot x n. So, if we, if you differentiate this one with respect to x, that differential let us call this I am denoting by f of x is equal to that one. So, if you differentiate this a scalar function with respect to a vector x, that results you know this is nothing but a dell f of x dell x 1 similarly, dell of x dell x 2, dot dot dell f of x dell x n.

So, the results you can easily verify writing the details expression of this one and expand it, write the polynomial quadratic form with in terms of x 1, x 2 and p assume a matrix of ten by n matrix, with a, assume that p is a symmetric matrix, then only p is a symmetric matrix. Then results is p into 2, p into x and dimension is that, assume that p is a symmetric matrix. So, this you expand it in terms of x 1, x 2 just product it and p elements you consider the p 11, p 12 dot dot p 1 n and the second elements have p 21 in place of p 21, you write p 2 because this is a symmetric matrix. Then differentiate each element, each f of x in terms of x 1, x 2 ultimately the results will come this one.

So, this is a quadratic form where p is a symmetric matrix, if you differentiate, you that function with respect to n, that results is 2 p x. Next is your the derivative of linear function, this is the quadratic function, the derivative of, derivative of linear function f of x is equal to a 1 x 1, a 2 x 2 plus dot dot a n x n. And this is called linear function, even there may be a constant term C. So, when C term is there, constant term we call it is a affine function, will come details in later. So, let us call for the timing this is so you differentiate this thing with respect to x. Then what is the results? That you straight away can find out there is f of x is a scalar function, but which is a linear. Now you differentiate with respect to x vector.

(Refer Slide Time: 22:38)



So, that gradient of that, that is you are finding out, if you do this one I am not writing that expression, you know that, there gradient of function, each element you know how to do this. This will come a cross n so what is a n? This expression, if you see this expression I can write into row vector and column vector form dot dot a n, again this is row vector multiplied by x 1, x 2 dot dot x n, which I can write it is equal to a transpose x, x is a n plus 1 and a is where, where a is I can write it, where a is equal to u a 1, a 2 dot dot a n, whose dimension is n plus 1.

So, in short now can tell if you have a linear function is there, if you differentiate with respect to x, x is a vector x 1, x 2 then results is nothing but a, that means function which is expected a transpose x. The inner product of two vectors, this a transpose x, if you differentiate with respect to x, then result should be not a transpose, it will be a and that you can easily verify this one.

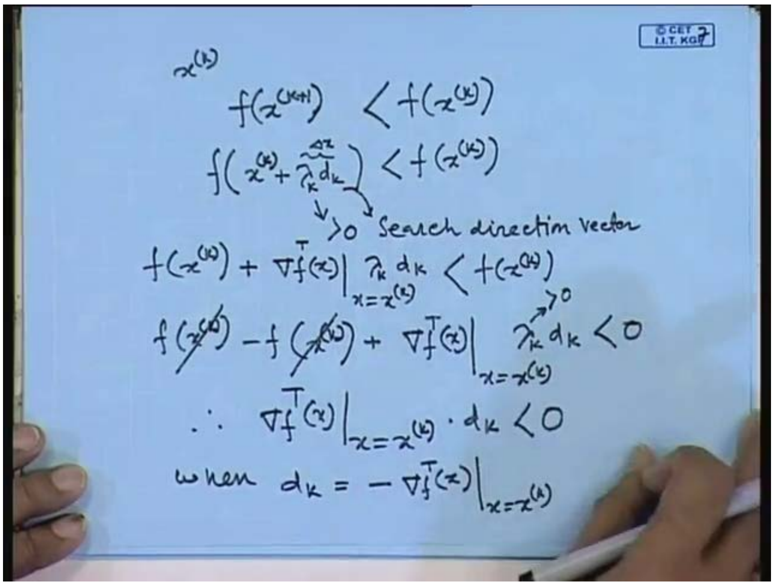
So, keeping these two results in mind, then we can what is called proceed further that, on constraint optimization problem. At the beginning of the class I think, let first lecture we have discussed what do you mean by the constraint optimization problem, unconstraint optimization problem. That means, objective function is given and there is no constraint subject to any conditions or and as well as there is no side constraints are there. So, this is called unconstraint optimization problems.

So, you are now discussing is a unconstraint optimization problems, but using some numerical techniques, numerical techniques, but the numerical techniques is say, it is a iterative process, iterative. The objective of iterative process is like this way, you guess some value, some value of x then find out the function value f of x. But our problem is minimization value, minimization of a function so you take a initial value of a, that variable x, let us call x super script of 0, you take this one.

And now next iteration, next iteration you got the improvement of x, where x is a n variables and if the function value, function value is improving. Means if our problem is minimization, the function value is decreasing that means, slowly we are, each illustration you are going to approach to minimum value of the function. So, our iterative process you can say iterative process is nothing but a, the objective of iterative process, of the iterative process, iterative optimization process, iterative, optimization.

So, in process is to reach a minimum value of the function, minimum value of the function, minimum value or you can write minimum value for the cost function or objective function, function f of x, which is n variables. Suppose that, suppose that at kth iteration, kth iteration we have not reached to the minimum value of the function. Next k plus 1th iteration of our value, that k plus 1th iteration, the function value if it is, it should decrease it. Then we are saying that we are approaching to slowly, if it is each iteration value is decreasing from the previous iteration value that means we are approaching to the, our minimum point, minimum value of the functions.

(Refer Slide Time: 28:01)



So, mathematical we can just write it here f of x, f superscript x, superscript indicates the iteration, kth iteration the value of the function, kth iteration the value of the variable, not function, value of the variable, when you put f x of k, it indicates that function at kth iteration. So, x superscript, x superscript k indicates the value of the decision variable at kth instant, if you put the values of x in the function expression, then you will value of the function at kth iterations, value of the function. So, if you assume at kth iteration the function has not reached to the optimum value, then k plus 1th iteration, the function value will be less than this, because we are approaching to the minimum value.

So, you can write it k plus 1th iteration value is less than this value. I can now write it, what is k plus 1th iteration value, it is value from kth iteration value, plus some partavation about kth iteration values, kth iteration, the decision variable at kth iteration, what is this variable from there? There is the partavations. Let us call that partavation, I will be writing it dell k, lambda k d k which I am writing delta x.

So, this value and kth iteration the decision variable value is x to the power of k and function value is this one, at k plus 1th iteration, that kth iteration what is the value of decision, value, value from there is a some partavation is delta x. And that, at that value in, what is called iteration, what is the function value f of this one and that value if it is less than this quantity, that means we are approaching towards the minimum point, each iteration should decrease it value. So, lambda is a scalar quantity whose values is greater than 0 and dk, naturally you can f, the x superscript k, that decision variable I am adding another vector. Where lambda k is scalar quantity, dk must be vector so that dk is called the search, the search direction vector.

We are looking for, that means from kth, kth iteration the decision variable x k form they are we are moving in the such a direction, such that the function value is decrease at k plus 1th iteration, function value should decrease. And we have to move in a such a direction and that direction is denoted by dk.

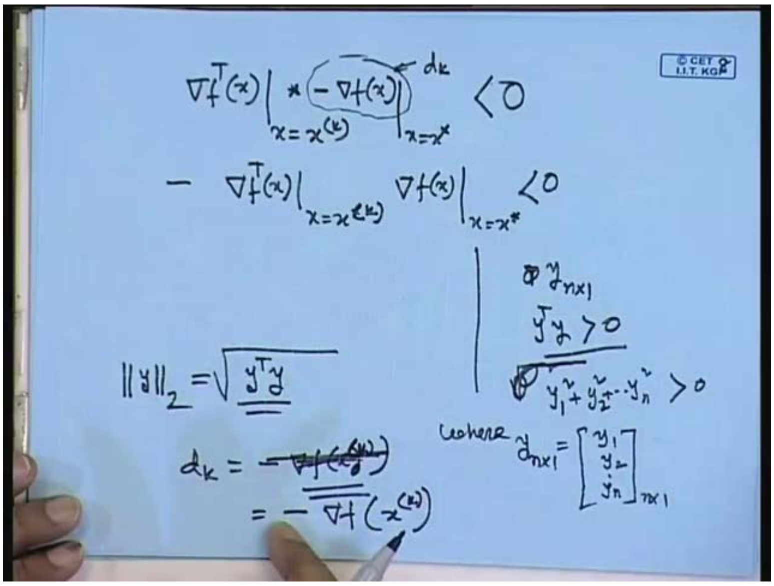
So, if you do this one Taylor series expansion because I told you that delta x is the partavation from the kth iteration, that decision variable, decision variable value at kth iteration, form there the partavation is delta x. If you do the Taylor series expansion up to first order, then we will see f of x of k this plus gradient of that vector f transpose of x k and put the value x is equal to x superscript kth iteration value. Because, I am doing the Taylor series expansion around this point where, kth iteration decision variable values from there into delta k lambda k into dk.

So, that and other what is called terms I have neglected them because I am considering delta x is very very small considering, that value if it is less than this one, then we are approaching in the right direction. So, that function value each iteration, when iteration from k to k plus 1th iteration, the function value is decreasing so this. Now you will bring it in this side, this equation if you bring it this one, this function x of superscript k minus f of superscript of k this plus, gradient of this function of f of x at x is equal to x superscript k in to lambda k dk, must be less than 0. Because, this part I have taken here this and this cancel.

So, this quantity should be less than 0, this is a scalar quantity see this is a means, row vector, lambda is a scalar quantity whose value is greater than 0, positive quantity. And dk is a vector and that vector is a column vector. So, you have to select dk in such a way, that product because lambda k is a positive quantity so it will affect that value of, what is this quantity less than 0, it will not accept. So, you can write therefore, gradient of this one f of x, x is equal to x superscript of k into dk must be less than 0, negative quantity.

So, there is a lot of different choices are dk is there, may exist, one of the choice is obvious that, it will be a negative quantity, that this product will be negative quantity when, we select or choose the dk value is equal to minus that f of this value. When dk value if you select that value, it will be a that, gradient transposing to that gradient with ((Refer time: 34:52)). So, this product, this into this you could write it x is equal to x k so this into this, is a norm of the vector and that is always norm means, distance of vector, distance is positive quantity and divided by the minus.

(Refer Slide Time: 35:13)



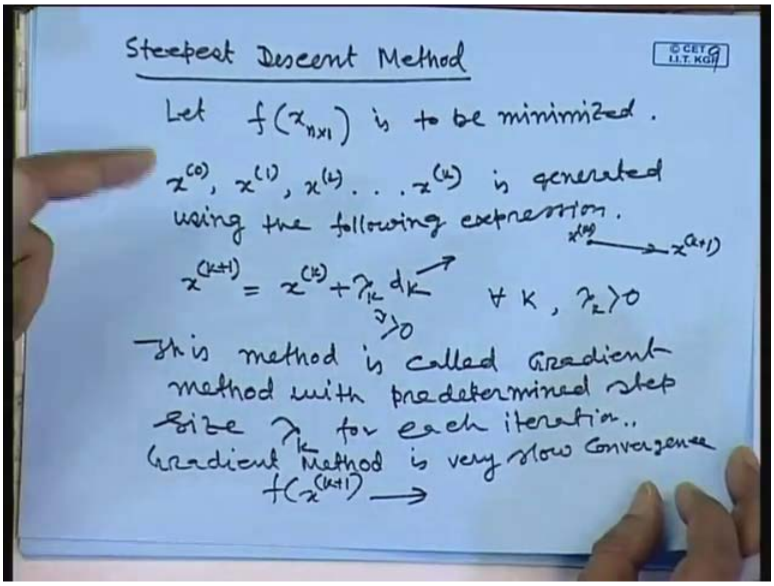
So, if I put it here this trace, then it will coming like this way see dell f this x, x is equal to what is called x superscript k, kth iteration value, multiplied by delta f of x ((Refer Time: 35:33)), because this multiplied by delta. So, this is we call, this we have selected as a dk so this and put this value x is equal to x star and this value always less than 0. Because, this is nothing but a f transpose of x into this quantity x is equal to x star, x superscript k into delta f of x, x is equal to x star and this quantity is less than 0.

If x, if y is a let us call y, y is a vector n cross n, this then if you write it y transpose y is nothing but a scalar quantity and this scalar quantity is always greater than 0. This means, if you see this companying is nothing but a if you do this, this one if you just multiply it is nothing but a y 1 square plus y 2 square plus dot dot y n square is greater than 0. Where y is a vector of dimension n cross 1, is elements are y 1, y 2 dot dot y n. So, this is and physically it is nothing but a distance square, this is nothing but a distance square from the origin of the vector.

So, this mathematically is, if we were writing like this way, y norm of this two is nothing but a this is a distance square, is nothing but a square root of y transpose y. So, this quantity, if you see this quantity is a always a positive quantity and preceded with minus. It is always this product of this iteration, this would be negative so if this condition is satisfied. In other words, if you consider the or such direction d k with reverse sign of a gradient of a that function, then the function value form kth iteration to k plus 1 iteration, the function will decrease. And you have to select that dk, the relation of this search vector direction, you have to choose that way. So, what is our choose of search direction? dk will be a minus delta f of x, that is our choice. So, dk because we in kth iteration so you write it this one k so dk I am writing once again.

So, minus gradient of, find out the gradient at kth iteration, this is a direction, search vector direction we have to move next, to go from x superscript k, k point to k plus 1, we have to move in such a way so the function value is decreased, when you will move from kth point to k plus 1 point. And you have to move like this way because function is known, gradient is noted with negative sign, reverse sign of that one. So, this way you have to move it. So, now our, we will discuss keeping this thing in the mind that how to how to select the search vector or direction vector, then we can proceed that, how to solve a optimization problem, optimization problem of function, which is a n variable function by using the iterative process.

(Refer Slide Time: 39:57)



So, our next is we can that Steepest descent method and this is based on this, what is called that concept is the, this concept will be used here. So, let us consider we have a function f of x and x is a n variable, the function f is a function of n variable x 1, x 2 dot dot x n, is to be minimized, that is our problem. So, you consider we have a sequence of point x of 0 x of 1 x of 2 dot dot x of a, this is the, we have a sequence of point is generated. How it is generated? using, using the following expression, following expression. Now look at this expression, this one our basic, that concept what we consider is this, this one that means our problem is minimization.

Now if we are, if we are in kth point, we move to k plus 1th point in such a way, in such a direction so that, the function value kth iteration, at kth point, the function value at kth and the function value at k plus 1, that difference will be negative. So, that is the basic one. With this thing keeping in the mind, I can write x k plus 1th iteration, that x of k and which relation from k, kth iteration and from kth point which direction I have to move in the dk direction. And dk I know, what is dk? Minus of gradient of that function, this function.

So, this is plus lambda k into dk, if you move in this direction let us call you have a point is x k, x superscript k, you move in the direction dk and dk what gradient of that vector, that function f of x with negative sign, you move that way, that. So, you will get it x superscript of k, at this value, at this value the function value will be less than what is the value of the function at kth iteration.

So, this is a expression so this I told you is, value is greater than 0, positive quantity and this if, you can choose the direction in such a way for all values of k, this is true. And also lambda k, we have selected greater than 0, then ultimately we will reach to the what is called minimum value of the function. So, this method is called, this method, this method is called, is called or known as the gradient method. Because, the whole, this method that we are moving in such a direction, the function value is decreased from one point to another point.

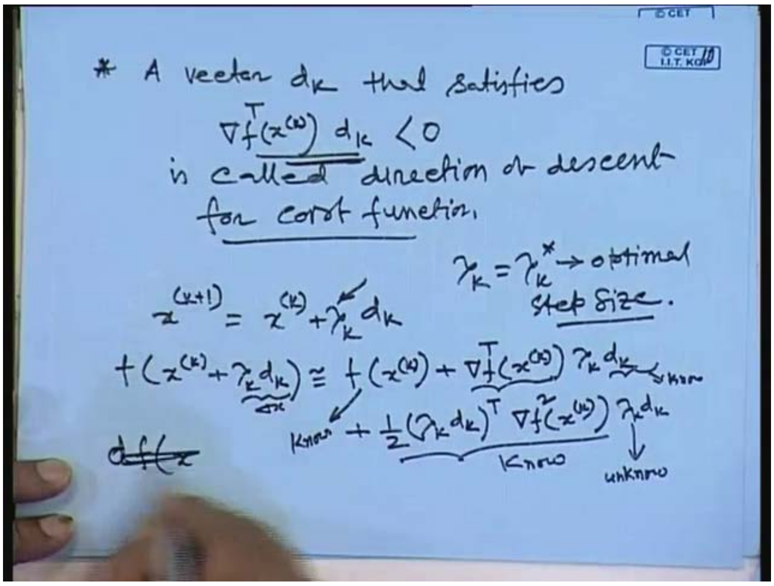
So, this method is called gradient method with predetermined step size, lambda k for each iteration, for each iteration lambda k is predetermined and it is kept constant throughout the iterative forces, for each iteration so that is gradient method. Now steepest descent method, is just extension of gradient method. In this case, we have fixed that lambda k which is nothing but a step size, we will move in dk direction, but what is a step size, we will move it. Now in steepest descent method, the lambda k is optimized, lambda k is in step size is optimized, at each iteration k is equal to, k is equal to let us call 0, k is equal to 1, k is equal to 2, k is equal to 3, in each direction lambda k is optimized, step size of this one is optimized.

So, this iteration, this is optimized and then it is called, it is a steepest descent method, the difference between this and this only, this is the basic expression. So, this one and this is the search direction, search direction and the gradient method the lambda k is predetermined, step size is predetermined and each direction, it is taking the same value. Whereas in the steepest descent method, the lambda k the step size is from one step to, one iteration to another iteration is optimized and it is used to find out the value of decision variable at k plus 1th instant, this is the only difference.

So, let us call in gradient method one is, what is call is a very simple gradient method is very simple no doubt, but convergence is very slow. Gradient method is slow, is very slow convergence, but easy to implement. But whereas, in the steepest descent method it is simply simple, but little competition burden is increased when we are going to find out the step size of lambda k, optimally. So, that is little competition burden is there, but convergence is fast because we are, each iteration we are finding out the, what is called step size, optimal step size of the lambda k in order to get the function value as minimum as possible. Because, when you put this value in the function, it is a function of only single variable lambda k because x of superscript k is known to us, dk is known to us.

So, if you put this value that f of x k plus 1 value in this expression and this expression this is the function of lambda k only and if it is a function is a, that is the is only a single variable case. We know how to find out the optimum value of the function f, which is function of lambda k only, we know find out its minimum value of the function. So, that way the convergence, concept convergence is more faster than the gradient method. So, and you can see this one from x, that if you are at k, if you are on the kth point. And if you move in the search direction, in the proper direction in search direction is slight partavation in that direction if you move it, there is lot of improvement in the function value, function value is the decreasing.

(Refer Slide Time: 48:46)



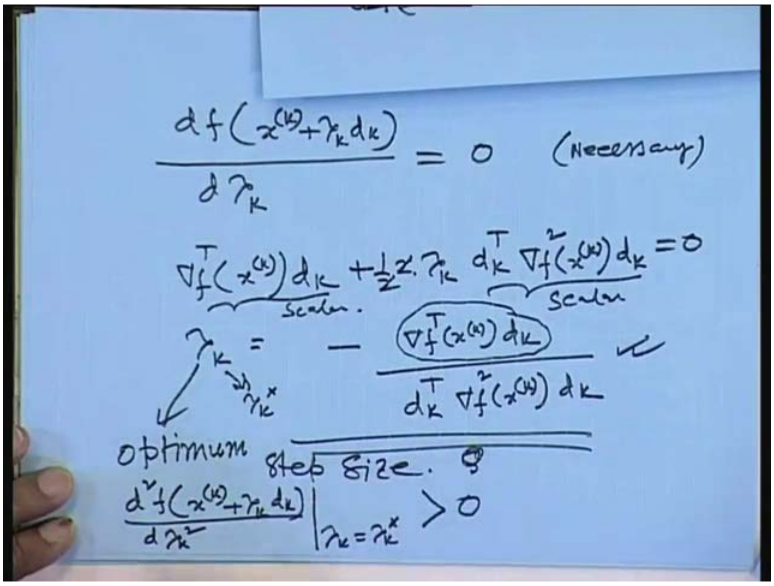
So, a vector, who is satisfy a vector dk, so our conclusion, final conclusion is a vector dk, that satisfies this condition, what this condition? Delta f what is called gradient transpose of this, superscript k into dk, if it is less than this 0, if a vector that satisfy this one, is called, is called direction of descent for the cost function. In short if this condition is satisfied, then if you move from to x k to x k superscript k plus 1, the function value will decrease at k plus k superscript k plus 1, the function value will decrease from, what is called, function value at kth iteration, at kth iteration function values.

So, this is the condition, this is the important condition so now question arise how to find out the optimum step size of lambda k? Lambda k is equal to lambda k star, that means optimal step size, while we go from k plus 1 to, we will go from k to k plus 1 it is iteration, then what is the optimal size of that one? That means that one, what is optimum size of that one, lambda k? When you move from kth iteration while we update k plus one decision variables, how to move this one. So, let us see this one so let us call it is a f of k plus lambda k, we move from kth iteration to k plus1th iteration dk.

So, this nearly equal to by Taylor series expansion I can write f of superscript k this plus f transpose of superscript k of this gradient transpose this. And what is this one, I am just writing Taylor series expansion, this is lambda k into dk, this is delta x, you can say delta x simply. Then half lambda k, delta k transpose, lambda k dk whole transpose then what is called gradient, that Hessian matrix, differentiation of gradient of function, that function, with respect to x once again. So, that is that one, is that, that why at what point kth point this and then lambda k dk and I neglected the higher terms of this one.

Now you see this one, this is known to us, if you see this is, this quantity is known to us, known, this quantity also known, this quantity also known, this is also known and this is also known. This whole thing is known, except unknown is here, unknown lambda k, lambda k. Now this is, if you see this one in this, the function expression if you put it that x k plus 1, that k plus 1th iteration the value of x, if I put it here, the function is, that function, objective function or the cost function is the function of lambda k, only. You see in this expression is lambda k only, lambda k. So, now I what should we change of lambda k so that, the function value is minimum, for this one. So, you have to differentiate this with respect to x, since it is a function of single variable I will write it d f, which is x of, next page I am writing.

(Refer Slide Time: 53:56)



This is nothing but a df x of superscript k plus lambda k dk that differentiate with respect to lambda k, is assigned 0, this is a necessary condition for this one. So, I have to find out the function value is minimum for what choice of lambda, this is the necessary conditions and if you solve this one, you will get the level value of lambda. So, if you differentiate this, this one you see, if you differentiate this, this one with aspect to lambda, this is the known constant. So, this will be 0 and this is what lambda k differentiating with aspect to lambda k, this is the, this vector is row vector, this is a column vector and this is scalar quantity.

So, if you differentiate this one you will get same thing here so delta f transpose of f of k whole dk, this is the first term it will come and this one you see lambda k is a scalar quantity. There are two lambda k, lambda square would come and it will be dk transpose, the gradient, that is where Hessian matrix then dk. So, it will becoming plus half 2, I am differentiating with respect to lambda so it is a lambda k, lambda square, lambda k square so the 2 is coming here, then it is a dk transpose, then Hessian matrix or second variable, derivative of the partial derivative of the function.

So, this have superscript k this into dk is equal to 0 so this is our necessary conditions then lambda you will find out, lambda k is equal to, if you just do this one, this, this cancel lambda k is equal to, you will get, this is the scalar quantity mind it, x transpose p x is a same thing, scalar quantity. And this is a also scalar so if you take it this, that side is a minus gradient of, gradient transpose x of k d k divided by, I can divide because it is a scalar quantity, divided by d k transpose delta f square of x k in to d k.

So, you see this is scalar quantity, this is scalar quantity and this quantity the product of this one, I told you lambda k is greater than 0. So, this product, if you see this product there is a condition for this and direction, the function value will decrease from kth point to k plus 1th point, if this product is negative. So, negative, negative here is negative so positive lambda k is positive quantity. So, this is the choice of lambda k, for which function value will decrease, if you take the lambda value, lambda k value some other value, other than this one function value is decreased. But if you select this one, the function value will decrease as much as possible.

So, this is the choice of optimum, this is the optimum, optimum step size, size. So, this now we can check it whether, function value also optimum, they check whether this value is del square f x of superscript k plus lambda k d k, differentiate with respect to d lambda k square. This value, putting lambda k is equal to lambda k star, that is we call this equal to lambda k star. Put if this value is greater than 0, this is a scalar quantity, is greater than 0. That means, function is value is decreasing as minimum as possible for the, for the choice of lambda k. So, we will stop it today, next will continue with this one, will write the, what is call algorithm, then how to implement this one in digital computer, for this is one.

Thank you.