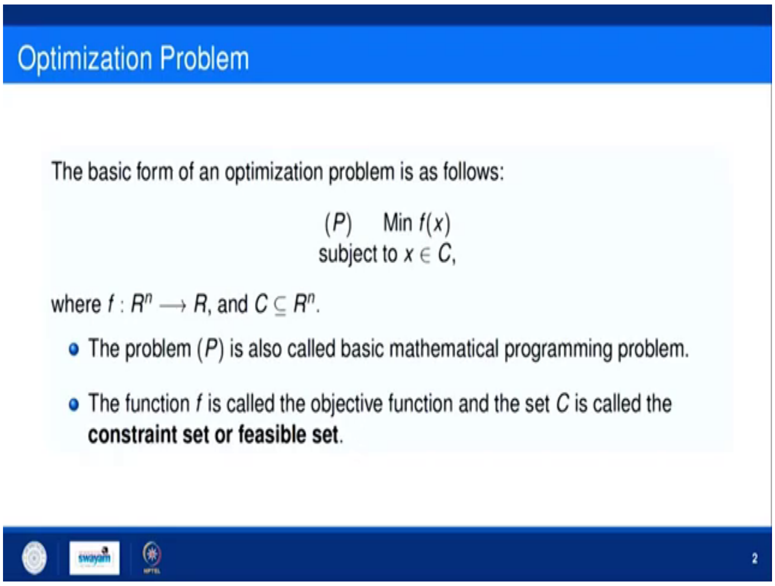
Essential Mathematics for Machine Learning
Prof. S. K. Gupta
Department of Mathematics
Indian Institute of Technology, Roorkee

Lecture - 26
Unconstrained Optimization

So, hello friends, welcome to lecture series on Essential Mathematics for Machine Learning. So, in the last lecture we have seen that if a function is given to you and you have to check whether given function is convex or not and you know that the function is twice differentiable. Then you simply find Hessian matrix of that function f and if it is positive semi definite, then it is convex. If it is negative semi definite; that means, it is concave if it is positive definite, then it is strictly convex.

Now, let us come to unconstrained optimization problems, if a optimization problem is given to you and you want to see that the where the given x bar is a point of local minimum or local maximum, then how can we check. So, that is given in this lecture on Unconstrained Optimization.

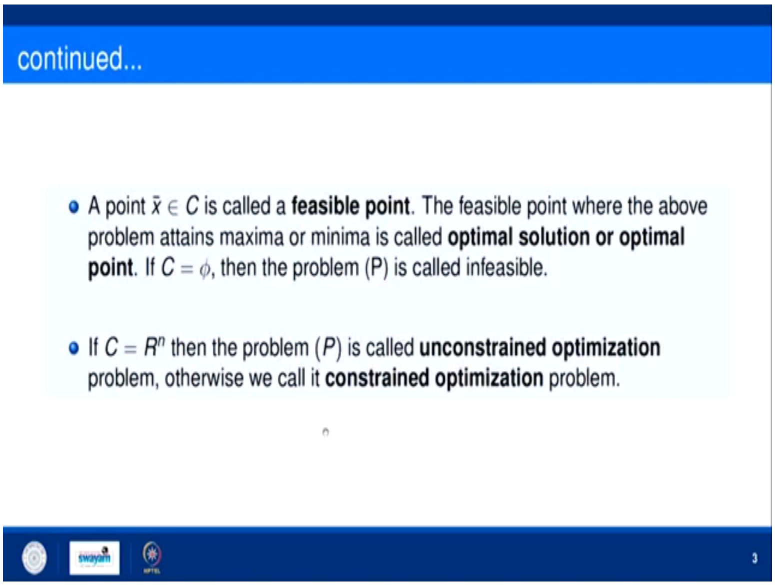
(Refer Slide Time: 01:21)



Now, first of all how we can define an optimization problem. So, optimization problem is basically maximization or minimization of a function f, which is we called as objective function, subject to some set of constraints constraint set C. Here, I am defining as C which is a constrained set. So, this function f is from R n to R a R; that means, there are n number of unknowns and this C is the subset of R n, the set of constraints.

This problem is basically called basic mathematical programming problem or basic optimization model ok, in which we have to minimize. Similarly, we one can has to maximize also. So, it is a minimization of f subject to x belongs to C the set the function f is called objective function which you have to maximize or minimize and the set C is called constraint set or feasible set ok.

(Refer Slide Time: 02:20)

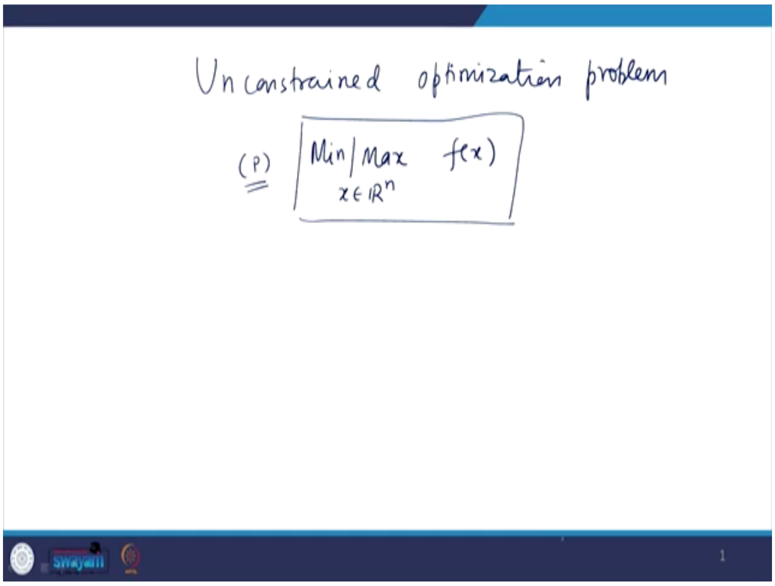


Now, a point x bar belongs to C is called a feasible point ok. Any point with constraint is called a feasible point and a feasible point where the above problem attains maxima or minima is called optimal solution or the optimal point. Here, in this case we have taken only minimization type problem.

So, if it is a minimization type problem then we say that this point x bar will be in optimal point if it minimize f and the point feasible also ok. So, if this C is phi; that means, the constraint set is empty, then this means problem is infeasible ok. If there is no x bar satisfying all the constraints; that means, problem is inconsistent, the problem has no solution ok.

Now, if this C is the whole R n, because this C here this C is basically here, this C is basically subset of R n. If this C is equal to R n, then this problem is called unconstrained optimization problem otherwise, if C is a proper subset of R n then we call it constrained optimization problem.

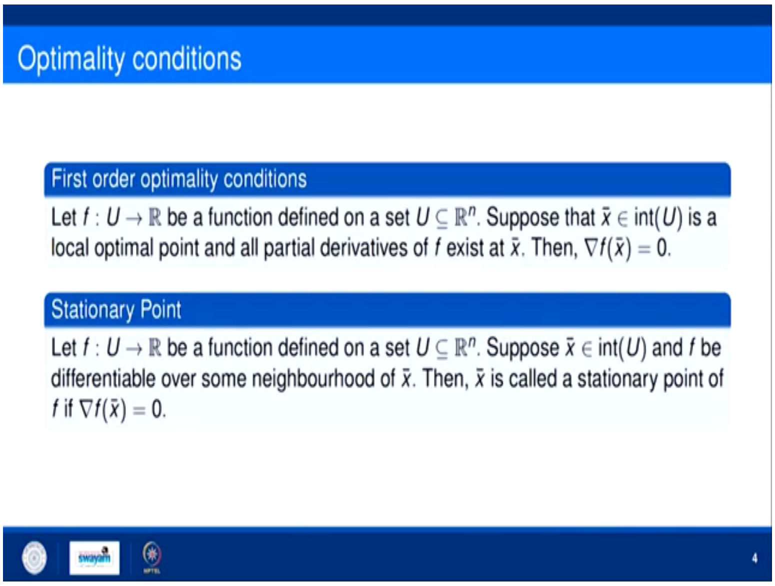
(Refer Slide Time: 03:45)



Unconstrained optimization problem can be modeled as optimization problem. So, you have to minimize or maximize fx and where x belongs to R n. So, that is basically an unconstrained optimization problem ok. Here, there is no restriction on the variables, variables maybe any x in R n ok.

So, how we can; how we can find the point of local maxima local minima for this f? What are necessary and sufficient condition for this problem P. Suppose, this problem is P. So, what are necessary condition for this problem P by which we can find out that what are the various points of local maxima local minima of this f. So, let us try to understand those conditions.

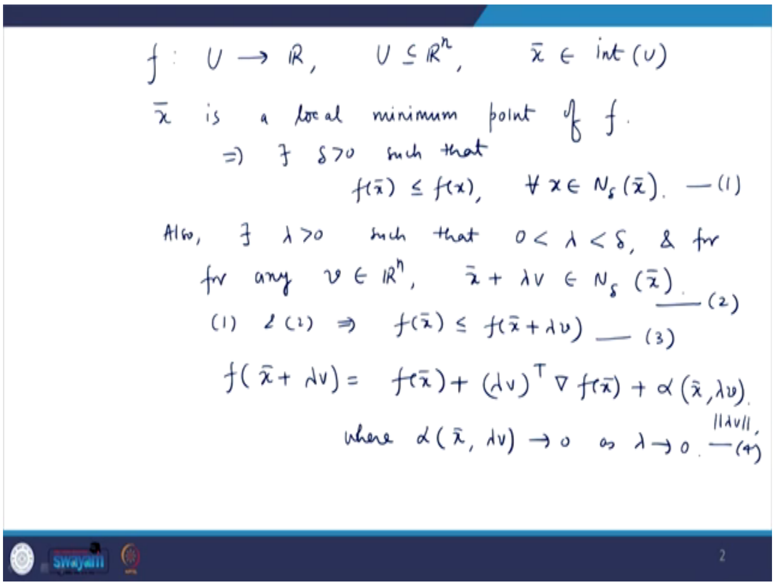
(Refer Slide Time: 04:47)



So, first of all first order optimality conditions, what is that? Now, let f from U to R be a function defined on a set U subset of R n. Suppose, that this x bar belongs to interior of U interior of this U it is not on the boundary; however, it is it is belongs to interior of U. Then this x bar belongs to interior of U is a point of local optimal and all the partial derivative of f exist at x bar.

So, we are supposing this, then gradient of fx at x bar is always 0. This is a necessary condition that if x bar is a x bar belongs to interior of U and is a point of local optimal local optimal point, then and all partial derivative exist at x bar then gradient of f at x bar will be 0. So, how can we; how can we show this? So, let us try to prove this.

(Refer Slide Time: 05:54)



So, here f is a function from U to R, U is some subset of R n and x y is a point which belongs to interior of U and also it is given to us that all first order partial derivatives of f at x bar exist. So, we have to show that gradient of f at x bar equal to 0. So, what has given to us that x bar, it is also given to us x bar is a point of is a local minimum point of f.

So, here I am taken the local minimum point. Similarly, we can show for local maximum point also ok. So, if it is a point of local minimum of f for our unconstrained optimization problem, then what does it mean? It mean that there will exist some delta greater than 0 such that fx bar is less than equals to fx, for every x belongs to delta neighborhood of x bar.

This is a meaning of local minimum that you have a neighborhood of f at x bar, you have a neighborhood of x bar such that for every x belongs to that neighborhood f of x bar will be less than equals to fx. So, let us suppose this is 1 also, there will exist some lambda greater than 0 such that, for 0 less than lambda less than delta and for any v and for any v belongs to R n x bar plus lambda v will belongs to delta neighborhood of x bar.

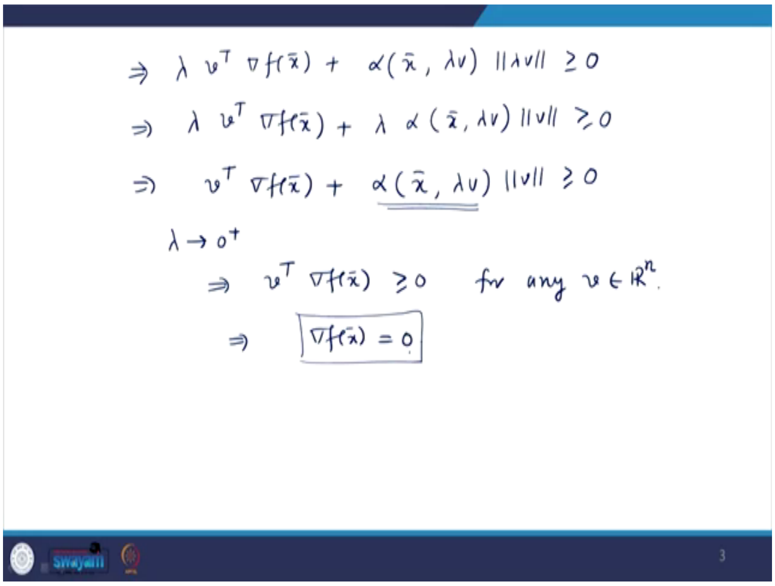
See, no matter what this lambda is ok. So, for any v for any v belongs to R n if x bar if x bar is in x bar is already in delta neighborhood of x bar ok. So, for any v belongs to R n and for sufficiently small lambda x bar plus lambda v will also belongs to delta neighborhood of x bar. So, what we have conclude from 1 and 2?

From 1 and 2 we have concluded implies that f of x bar will be less than equals to f of x bar plus lambda v also, because this point is in delta neighborhood of x bar and this inequality hold for every x in delta neighborhood. So, if it is in delta neighborhood of x bar this inequality will hold for x bar plus lambda v also ok.

So, let us suppose this inequality is 3 ok. Now, let us, apply, because first order partial derivative exists; that means, function is once differentiable. So, let us apply the definition of once differentiability of f at x bar. So, by that definition what we what we are having; that f of x bar plus lambda v is equals to f of x bar plus lambda v transpose gradient of f at x bar plus alpha x bar into lambda v function of this and this and norm of lambda v where this alpha x bar lambda v will tends to 0 as lambda tends to 0 ok.

Suppose, it is suppose this inequality is 4 now from 3 and 4 since, this fx bar plus lambda v minus fx bar from three is greater than equal to 0 so; that means, this expression is greater than equal to 0 ok.

(Refer Slide Time: 10:04)



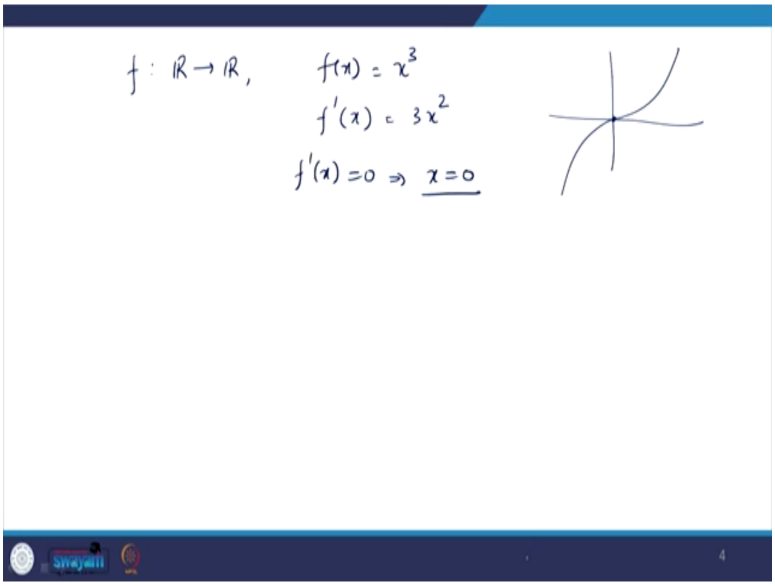
So; that means, lambda into v transpose gradient of fx bar plus alpha times x bar lambda v norm of lambda v is greater than equal to 0, this implies. So, this further implies lambda v transpose gradient of fx bar, this lambda can be taken out from the norm ok, lambda it is alpha x bar it is norm of v is greater than equal to 0.

So, this implies v transpose gradient of fx bar plus alpha times alpha is the function of x bar plus x bar and lambda v norm of v is greater than equal to 0 ok. Now, take lambda tending to 0 plus we are assuming lambda is tending to 0 plus. So, if lambda is tend to 0 plus this will tends to 0 and this will tend to 0 means this is greater than equal to 0. So, this implies v transpose gradient of fx bar is greater than equal to 0 for any v belongs to R n.

So, if this inequality is holding for any v belongs to R n this will happen only when gradient of fx bar itself is 0, because this v is an arbitrary vector in R n and this inequality is holding for every any v in R n, this may be anything. So, this will be happen this will happen only when this gradient itself is 0, because 0 into this will be greater than equal to 0. So, this proves the necessary condition for local optimality of f at x bar.

So, hence we can say that if x bar is a local minimum or local maximum point of f at x bar and function is once differentiable, then gradient of f at x bar equal to 0, but this condition is not sufficient.

(Refer Slide Time: 12:12)

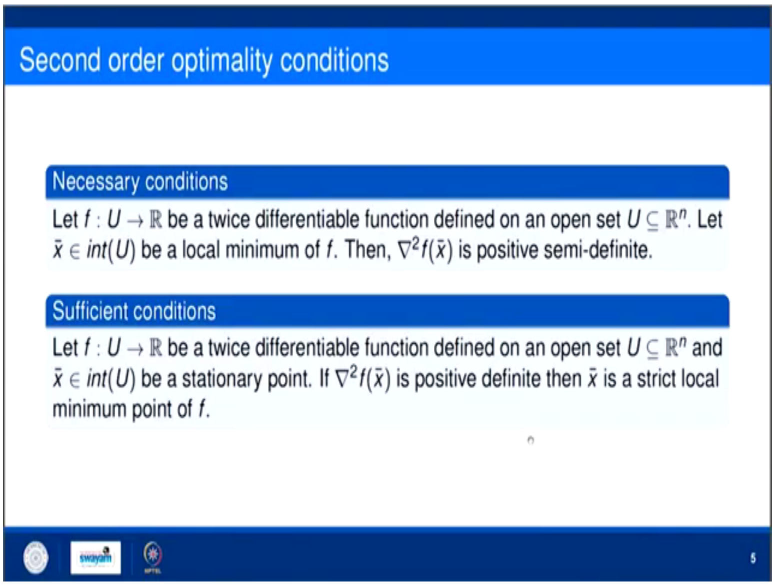


For example for example, suppose you have function from R to R and the function is given as x cube. So, first derivative is equal to 3 x square and this derivative is 0 implies x equal to 0, but x equal to 0 is not a point of minima and neither a form point of maxima for this function x cube. You can you can check from the graph. These are 0 point and this 0 point is neither a point of maxima nor the point of minimum for this function. So, this gradient of f at x bar is only a necessary condition.

So, what is a sufficient condition? So, first let us understand what stationary point is. So, let f is from U to R be a function defined on a set U subset of R n. Suppose, x bar belongs to interior of U and f be differentiable over some neighbourhood of x bar then x bar is called stationary point of f where gradient of f f at x bar equal to 0.

So, here first order derivative is 0 here first order derivative is 0. So, those points are called stationary point. So, if gradient of f at x bar equal to 0, all those x bars are called stationary point of f.

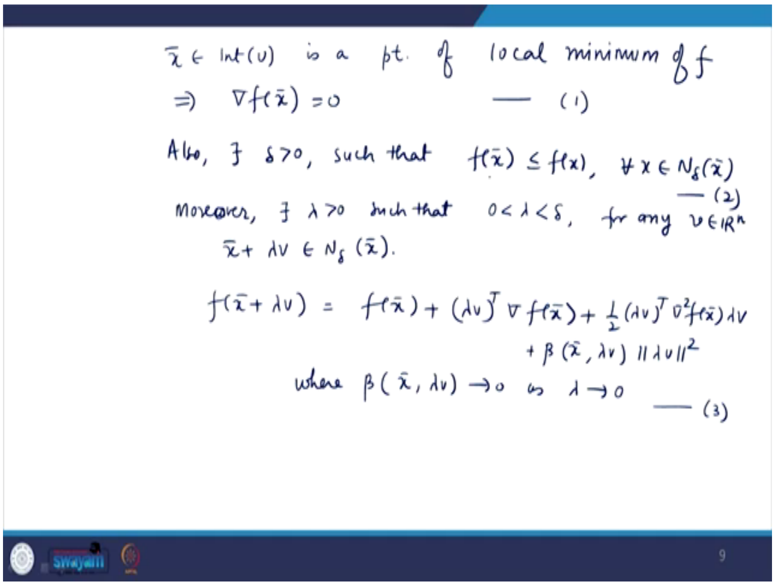
(Refer Slide Time: 13:41)



So, now what is the second order necessary condition of f to be a point of local minimum? So, this is a second order necessary condition that if f is twice differentiable function on an open set U of R n subset of R n and let x bar belongs to interior of U be a local minimum point of f, then Hessian matrix of f at x bar is positive semi definite. This is also a necessary condition not sufficient.

So, first necessary condition is a point x bar is a point of local minimum, this implies gradient of f x bar equal to 0 and the second condition is that Hessian matrix of f at x bar is positive semi definite, only necessary not sufficient. So, so how can we show this? The proof again for this is easy. So, what we have to show? We have to show that if x bar is a point of local minimum of f, then Hessian matrix of f at f x bar is positive semi definite. So, this we have to show. So, let us try to prove it quickly.

(Refer Slide Time: 14:45)



Now, it is given to us that x bar which is belongs to interior of U is a point of local minimum ok. So, this implies, the first of all this implies gradient of fx bar is equal to 0 and also there will exist some delta greater than 0 such that f of x bar is less than equal to fx for every x belongs to delta neighborhood of x bar, this is condition 2.

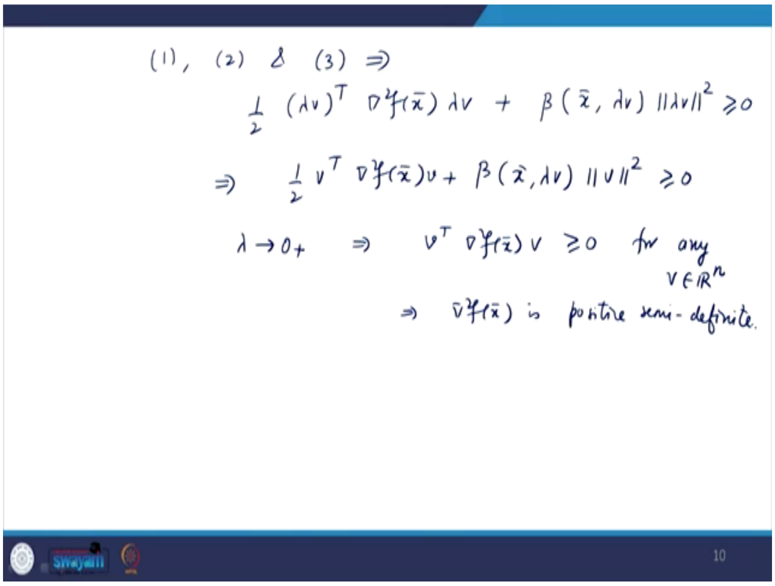
And also we can say that more over, there will exist some lambda greater than 0 such that such that 0 less than lambda less than delta and for any for any v belongs to R n x bar plus lambda v will belongs to delta neighborhood of a bar as we did in earlier theorem ok.

No matter how small this lambda is there will exist this lambda such that for any v x bar plus lambda v belongs to delta neighborhood of fx x bar. So, now, let us apply second order differentiability of f at x bar. So, what we obtain fx bar plus lambda v is equals to fx bar plus lambda v transpose gradient of f at x bar plus 1 by 2 lambda v transpose gradient square of fx bar lambda v plus beta x bar lambda v and norm of lambda v whole square and where? Of course, this beta x bar lambda v will tends to 0 as lambda tends to 0.

So, suppose this is 3. Now, since this inequality 2 is holding for every x belongs to delta neighborhood of fx delta neighborhood of x bar. So, this will hold for x bar plus lambda v also, because x bar plus lambda v belongs to delta neighborhood of x bar. So, let us substitute it here.

So, so, this f x bar f of x bar plus lambda v minus fx bar will be from this inequality will be greater than equal to 0 and this gradient of gradient of f at x bar is equal to 0 from 1. So, this is 0 and this is greater than equal to 0. So, we can say that this quantity is greater than equal to 0.

(Refer Slide Time: 17:37)



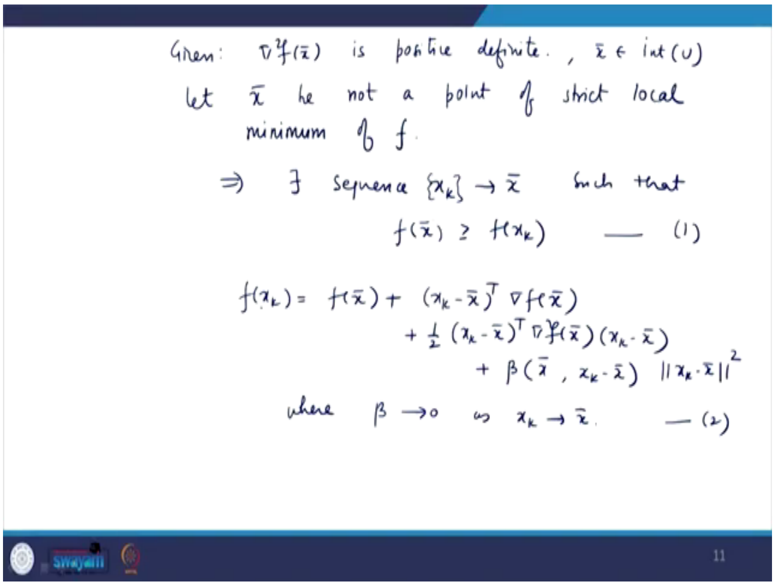
So, from 1 2 and 3 what we have obtained ? 1 2 and 3 implies 1 by 2 lambda v transpose gradient square of f x bar lambda v plus beta x bar lambda v norm of lambda v whole square is greater than equal to 0.

Now, lambda square can be cancelled. So, now, this implies 1 by 2 v transpose gradient square fx bar plus beta x bar lambda v norm of v whole square is greater than equal to 0. Now, take lambda tending to 0 plus. So, this implies v transpose this v is missing here, v transpose gradient square of fx bar into v is greater than equal to 0 for any v belongs to R n.

So, this implies gradient square fx bar is positive semi definite ok. So, in this way we have derived the second order necessary condition also. So, there are two condition basically; if x bar is a point of local minimum then the first condition is gradient of f at x bar equal to 0 and a second condition is Hessian matrix of f at x bar is positive semi definite.

So, both are necessary condition, but not sufficient. So, what is a sufficient condition? The sufficient condition is if f is twice differentiable function on an open set U subset of R n and x bar belongs to interior of U be a stationary point, then if f is Hessian matrix of f at x bar is positive definite then x bar is a strict local minimum point of f ok. So, let us quickly try to prove this result.

(Refer Slide Time: 19:42)



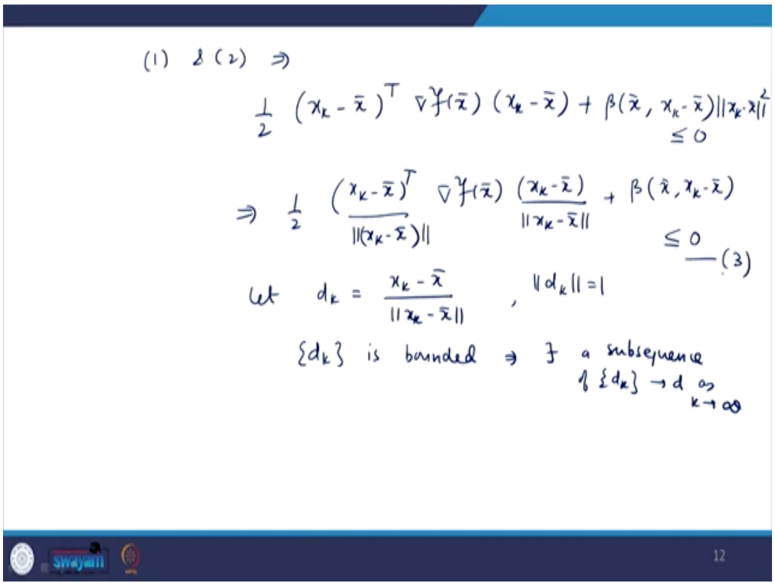
Now, it is given to us that gradient square of f at x bar is positive definite x bar belongs to interior of U ok. We have to show that x bar is nothing, but a strict local minimum point of f. So, let us try to prove this result by contradiction. Let us suppose that x bar is not a point of not a strict local minimum point of f. Let x bar be not a point of strict local minimum of f.

So, if it is not a point of strict local minimum of f this implies, there will exist a sequence x k which converse to x bar such that such that for fx bar will be greater than equal to f x k, because there will be a cluster of points in the neighborhood of x bar, a sequence I should say, sequence of sequence x k. Now, let us apply the second order differentiability of f at x bar. So, what we will get by the second order differentiability of f at x bar?

It is fx k is equals to f x bar plus x k minus x bar whole transpose gradient of f at x bar plus 1 by 2 x k minus x bar whole transpose gradient square of fx bar x k minus x bar plus beta x k into x k minus x bar and norm of x k minus x bar whole square where so, this is I think is this is x bar where this beta function will tends to 0 as x k tends to x bar.

So, this is second condition now, this f x k is minus f x bar is less than equal to 0 from one fx k minus f x bar is less than equal to 0.

(Refer Slide Time: 22:41)



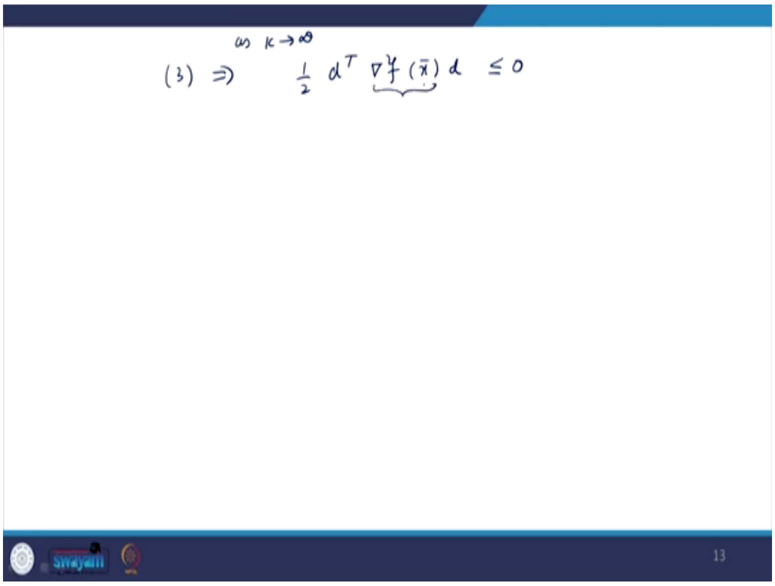
So, from 1 and 2 what we obtain so, from 1 so, and also for the statement is a stationary point. The stationary point means gradient of f at x bar equal to 0. So, this gradient term is 0 and fxk minus fx bar is less than equal to 0. So, from 1 and 2 what we obtain? From 1 and 2 we obtain we obtain that this quantity is less than equal to 0 1 by 2 x k minus x bar whole transpose gradient square of f x bar x k minus x bar plus beta x bar x k minus x bar norm of xk minus x bar whole square is greater than equal to 0.

So, this further implies 1 by 2 x k minus x bar whole transpose upon norm of x k minus x bar ok, this norm. So, so, we can check the expression from here. x k minus x bar whole transpose and norm of this thing beta beta times x bar x k minus x bar and norm of this square gradient square of fx bar x k minus x bar upon norm of x k minus x bar plus beta x bar x k minus x bar is greater than equal to 0.

So, it is it is less than equal to 0 I think, because you see if you see fx k minus fx bar from this expression is less than equal to 0. So, it is. In fact, less than equal to 0 ok. Now, let dk is equal to x k minus x bar upon norm of x k minus x bar and of course, norm of dk is 1 ok. So, so, this dk is a bounded sequence, because norm is 1. So, this dk is a bounded sequence.

So, there will exist a subsequence of dk such that dk will converse to T for when K tending to infinity. So, we can write that dk the sequence of dk is bounded. So, this imply there will exist a subsequence of dk which will converse to d as k tending to infinity. So, so what this expression this inequality implies then? This is now 3.

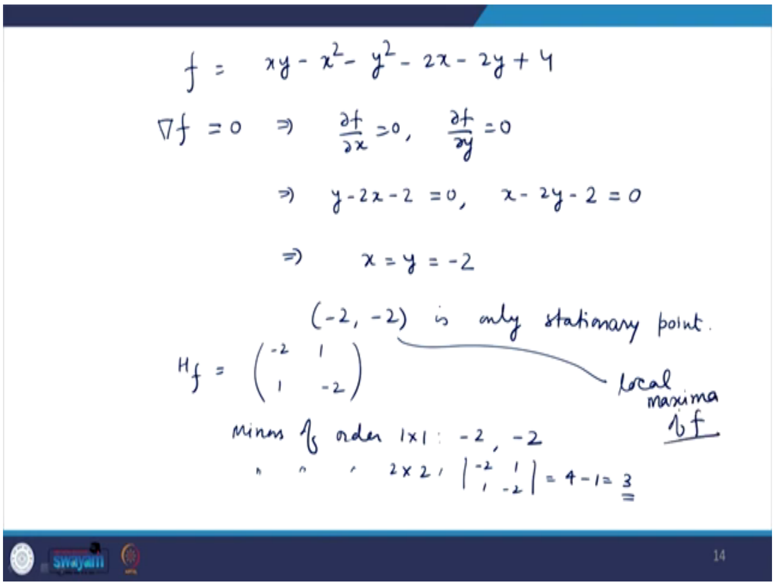
(Refer Slide Time: 25:12)



So, now 3 implies 1 by 2 d transpose gradient square of fx bar into d plus beta x bar into x k minus x bar is less than equal to 0 and as k tending that is it is as k tending to infinity, I should write, as k tending to infinity. So, as k tending to infinity this term is tending into 0. So, there is no need of this term. So, we can simply say that this is less than equal to 0.

So, this means this means this matrix is negative semi definite. So, this contradicts that Hessian matrix is positive definite. So, hence we have done. So, hence we can say that x bar is nothing, but a point of a strict local minimum of f. So, these are sufficiency condition. So, using this condition one can easily check that this point x bar whether it is a point of local minimum or not.

(Refer Slide Time: 26:18)



So, let us discuss it quickly by one example. So, take one example let f equal to x y minus x square minus y square minus 2 x minus 2 y plus 4. Now, suppose you are interested to see that what are the point of local minima or local maxima of this f if exist.

So, first we will find the critical points or stationary points. How to find critical point or stationary point? We take the gradient of f and put it equal to 0. So, what are gradient of f? Here, it is del f upon del x equal to 0 and del f upon del y equal to 0, only 2 variable problem. So, this implies y minus 2 x minus 2 equal to 0 and this implies x minus 2 y minus 2 equal to 0.

So, solve these two questions. So, if you solve these two equations we get we get x equal to y equal to minus 2. You can easily verify so; that means, minus 2 minus 2 is the only critical point is the only stationary point, because this is a necessary condition, this must hold. Gradient of f at x bar must be equal to 0 for the point to be a point of local maxima or local minima.

Now, to check whether this point is a point of local maxima or local minima find the Hessian matrix of f. So, what is the Hessian matrix of f? This is nothing, but minus 2 minus 2 and 1 and 1 and what are minors of this? Now, minors of 1 cross 1 are minors of 1 cross 1 are minus 2 minus 2 diagonal elements and minors of order 2 cross 2 are determinant of matrix itself, which is 4 minus 1 3

So, by this we can say that this matrix is negative definite and since it is a negative definite so; that means, this point is the point of local maxima. So, this point is the point of local maxima of f, because same theorem can be stated like this also if Hessian matrix of f at x bar is negative definite then x bar is a strict local maximum point of f.

So, in this way if we are if unconstrained optimization problem is given to you, then you first find the stationary point or the critical point of f and then find the Hessian matrix of f. If Hessian matrix comes out to be positive definite; that means, that point is a point of local minima if it is come out to be negative definite; that means, that point is a point of local maxima. In the next lecture, we will see how we can find out the optimality conditions for constraint optimization problems also.

Thank you.