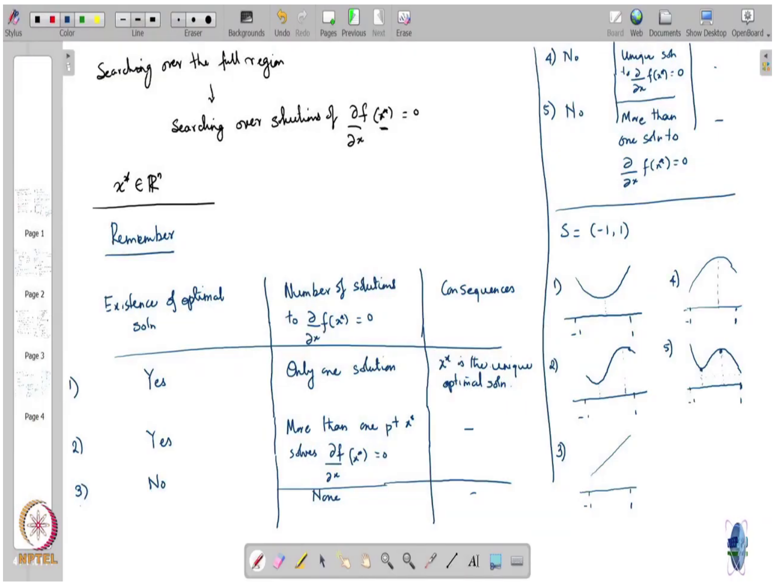
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture – 4C
Second order necessary condition

(Refer Slide Time: 00:16)



So, now let me come to the point that one of you asked, what about existence? Now, remember Weierstrass theorem gave us existence under what condition?

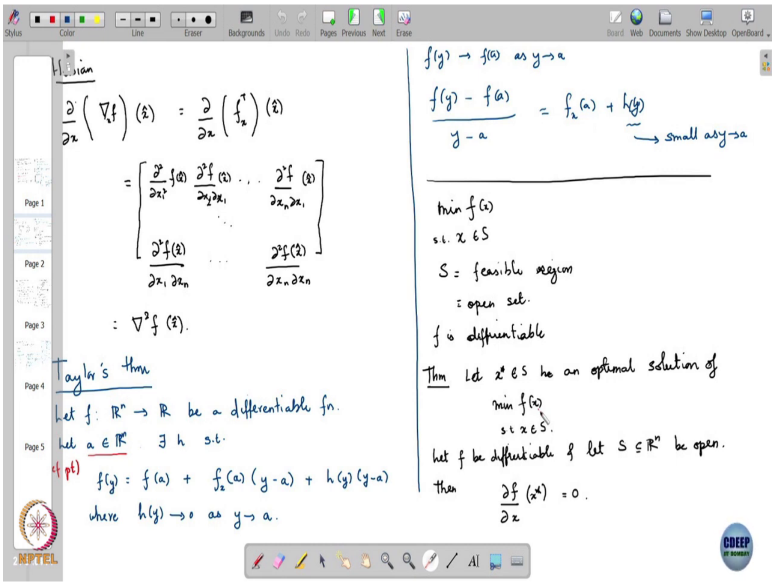
Student: (Refer Time: 00:27).

The objective function has to be continuous and the feasible region has to be closed and bounded. A feasible region that is open is not closed and is it could be bounded, but it is not necessarily closed ok. So, there are these strange sets that are both closed and open; let us not bother about them, basically there is no guarantee of being closed ok.

So, then therefore, you need another an independent way of verifying that the solution actually exist. It also brings me back to the first point I had made, we need the reason the importance of Weierstrass theorem in optimization is that it let us you check for a that a solution exist without actually asking you to find one right.

So, now, you without having a guarantee of way of claiming that a solution exist you can be in this kind of situation. You just you know try and find one and then you find all these points, but then none of them; there is no guarantee about any of them right, there could be solutions could be not nothing can be said.

(Refer Slide Time: 01:42)

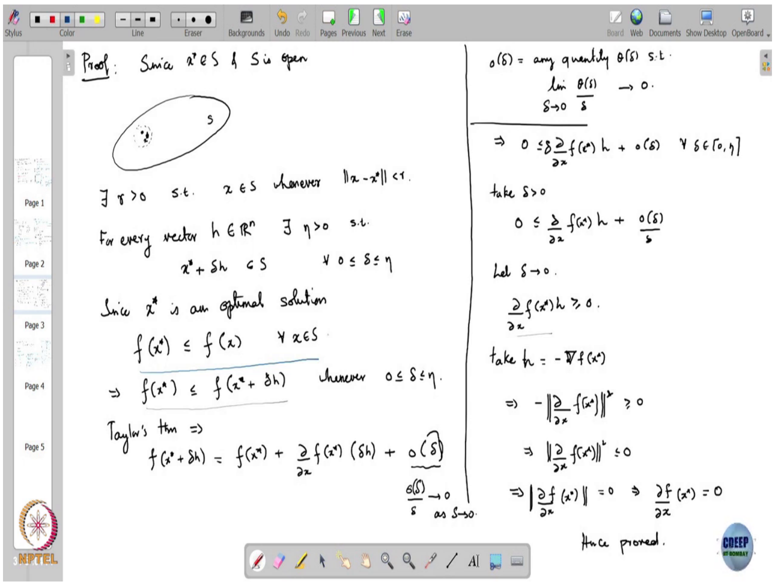


Now, also let us go; if you go back to the argument, the way we prove this right. So, I will say we start we started with this theorem saying let x star be an optimal solution of this; this particular optimization problem and what did we do? We said we took this particular optimization problem and we said let us look at this optimization.

Let us look at every direction in R n, shrink that direction down to the point where now it lies inside the ball. And therefore, inside the set and then you can say something about the inner product between the direction and the derivative, correct.

Student: (Refer Time: 02:18).

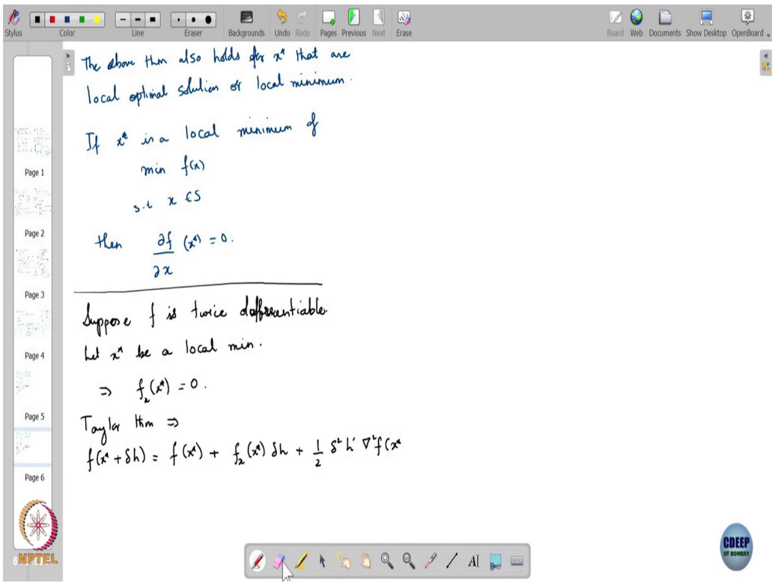
(Refer Slide Time: 02:16)



So, this was; this now this kind of argument can also be done if you are; if you are talking of not global optimal solution, but a local optimal solution right. So, here I was referring to optimal solution means a global optimal solution I was; I use this particular thing, this particular property right. I use this that f of x star is less than equal to f of x for all x in S, but then we do not need to do that; we can also work with the local optimal solution and there, entire local optimal your, the earlier ideas will continue to work.

I can still look for a vector that is in my local neighborhood that neighborhood can; I can shrink that vector further down to the point where it lies completely in it and again do complete repeat the same argument.

(Refer Slide Time: 03:26)



So, what this means is this particular result actually also gives us; so this theorem the above theorem also holds for x star that are local optimal solutions or local minima or local minimum ok. So, which what this means is; if x star is a local minimum of minimizing this function f x in S evaluated at x star is equal to 0.

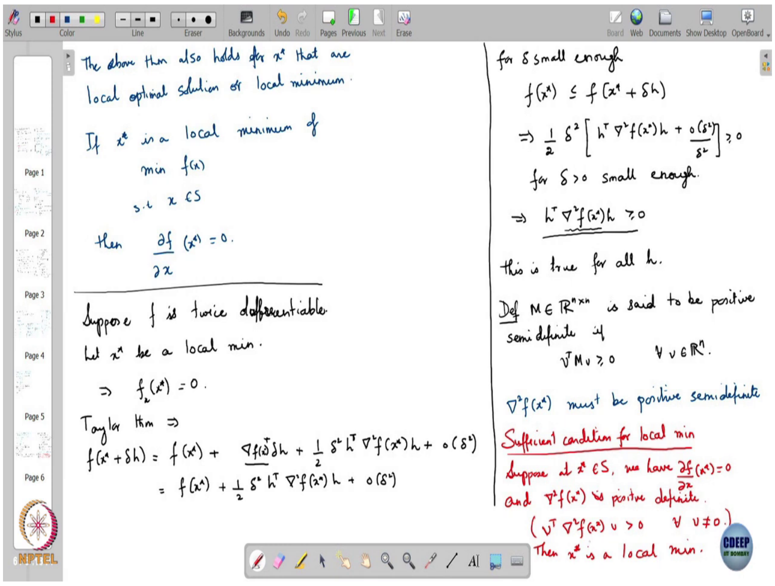
So, now suppose you get to this situation where you now have; you have you can you are in one of these situations, you have you know that the solution exist. But then there are multiple points where there are multiple points more than one point x star that solves the; that derivative equal to 0 or gradient equal to 0. So, what so; can we eliminate some of these ok? So, that is what we can; we will look at now. So, let us consider this; so, suppose f is twice differentiable ok.

Let x star be a local minimum then it must be that the derivative is equal to 0 and now we can actually use a stronger version of Taylor's theorem. So, when a function when you have twice when the function is twice differentiable, Taylor's theorem, you can use Taylor's theorem to get there is a stronger version of Taylor's theorem that makes second use of also the second derivative ok.

So, the Taylor's theorem I wrote here this use the only the first derivative and it was; it said that you can come construct a linear approximation of the function near that point. But if you also have second derivative of information, then you can actually construct a quadratic approximation ok; so that is the; that is a theorem that we will use.

So, by Taylor's theorem; Taylor's theorem actually implies that f of x star plus delta h is equal to f of x star plus now the. So, f of x star plus f x plus half delta square h transpose.

(Refer Slide Time: 07:42)



So, let me use just let me be consistent with one kind of notation; let me write plus now the quantity; the residual quantity that comes is now small o of delta square. So, you can construct a quadratic approximation and a quantity that the residual error that remains after that quadratic approximation is small o of delta square.

What this means is this is now a function that after dividing by delta square also goes to 0 ok; it is that sort of quantity alright. But now, remember this because f is a; x star is a local minimum, this gradient here is equal to 0. So this first term is gone now; so, you are left with just f of x star plus half delta square h transpose del square f at x star h plus some small o of delta square right. And now once again we have in a small neighborhood f of x star plus delta h will be greater than equal to f of x star.

So, for delta small enough what this imply is that you have your half delta square h transpose; for delta positive and small enough. Then what that means is as if you can divide throughout by you can divide by delta and then let delta go to 0, that gives that h transpose is greater than equal to 0.

Now, this must be true for this that you get this condition; the h transpose, Hessian of h; f at x star times h this should be greater than equal to 0, you get that you get this condition. This should be true this is true for all h; this is true for all h. Since this is true for all h, this just a matrix like this; a matrix like this that satisfies this kind of inequality a matrix of this kind which satisfies an inequality like this for all h, this sort of matrix is called positive semi definite.

So, let me just define that for you here; M n cross n is said to be positive semi definite if v transpose M v is greater than equal to 0, for all v in R n ok. So, with this definition what we are; what we get is that, you look at this Hessian matrix evaluated at x star; this must be positive semi definite.

So, this must be positive semi definite. So, if your f is twice differentiable; we can say more we can say that, if your x star is a local minimum not only must the derivative be equal to 0; then amongst the points where the derivative is equal to 0, you further check.

Student: (Refer Time: 12:44).

You can further narrow down. You can say well there are these points where the Hessian is positive definite and then amongst; so those are your optimal solutions must be amongst those points right because this is now a necessary condition. Now, what about sufficiency? Sufficiency means what? Necessary means simply that once it is a solution, this is a condition that the solution must satisfy.

Sufficiency means the other way around right; sufficiency means the other way round. Here is the condition, if I get; if I can verify then definitely that is a solution ok. Now, how what would be a sufficient condition in this case?

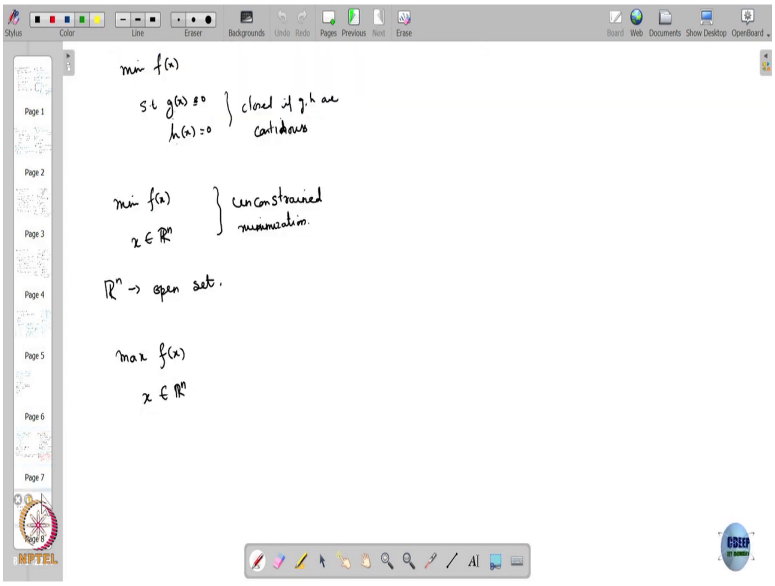
Student: (Refer Time: 13:30).

So, sufficient conditions are stronger than necessary conditions right; because sufficient conditions would imply the necessary conditions correct yeah. So, what were what would be the what would be the sufficient condition in this case? So, you must have that the derivative is equal to 0; that is necessary ok.

You must also that and, but in addition to this there the second derivative being or the Hessian being positive semi definite ok, if we ask for something more. If we ask for the that is that the Hessian is strictly positive definite or positive definite, then that ensures that these that the x star is a local minimum ok.

So, sufficient condition; this strictly is positive definite. So, what does that mean? That means that; so if I take v transpose del square f x star v, this is positive for all v that are not 0. So, if you have; if you have an x star in S such that the derivative is equal to 0 and the Hessian is positive definite; that means, v transpose the Hessian times v again is strictly positive for all v not equal to 0 ok, then that x star is a local minimum.

(Refer Slide Time: 16:27)



So, the optimization over open sets is as it does not always is not a very common problem that occurs. Usually the kind of when one writes an optimization problem like this you are writing a minimization of a function f subject to some g of x less than equal to 0 and h of x is equal to 0. The less than equal to 0 and h of x equal to 0; these kinds of constraints if these functions g and h are continuous, then these constraints ensure that your set is actually closed.

The feasible region is actually closed, but this closed; if g comma h are continuous, but a class of problems where it is very important to consider where we have every naturally encounter open sets are those where there are no constraints ok. So, the problem of minimizing f x where x just is in R n; all of R n ok; so this is what is called unconstrained minimization. So, unconstrained minimization this is actually a minimization over an open set because R n itself is an open set right.

So, all the previous results that I just mentioned; they all applied to minimization over an open set or so over a over an; for a minimization of in an unconstrained set. So, you have of unconstrained minimization problem alright. You can another point to note is that; suppose you had instead maximization instead of minimization, how would the results change?

Student: (Refer Time: 18:20).

Yes. So, how would my, our conclusions change? Our conclusion, the conclusion of this theorem here that the derivative is equal to 0, this conclusion will continue to work, even if you had a maximization; derivative would is still has to be 0. The conclusion of this theorem of the this conclusion here right that this blue line here that the Hessian must be positive semi definite that the Hessian should be positive semi definite that gets changed to Hessian should be negative semi definite ok.

And the sufficient condition will change to Hessian should be instead of positive definite Hessian is negative definite. Negative definite simply means that in place of in place of negative semi definite would just switch this inequality to a less than equal to and this inequality also, it was strictly less than ok. I am skipping over these details you can easy to work out on your own ok. So, we will end the class here.