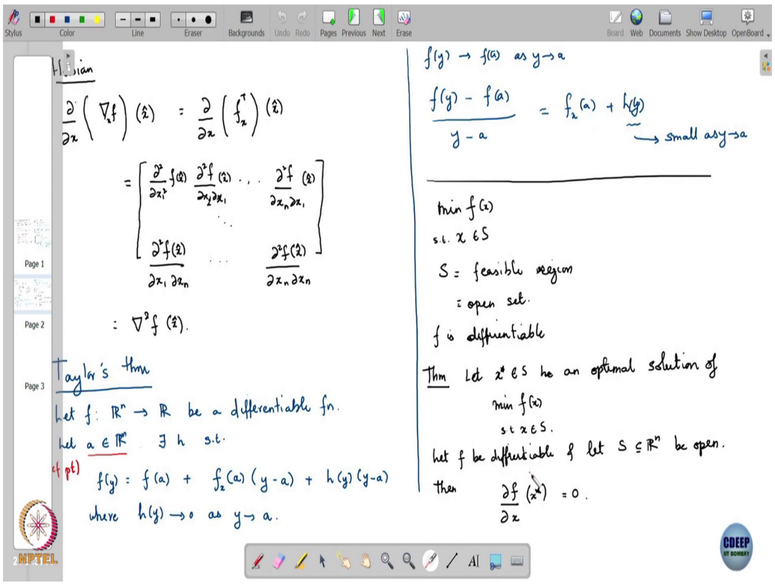
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture – 4B
First order sufficient condition

(Refer Slide Time: 00:23)



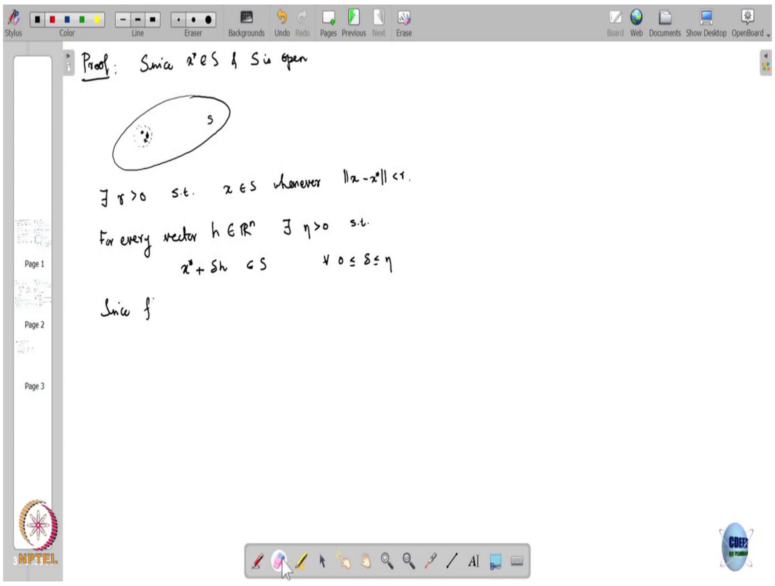
I am now going to talk about; going back to our the topic we started with which is that we want to minimize optimization of a function over a open set ok. So, we are going to talk of minimum minimizing a function f over a set x. So, over a set S subject to S x in S; so my feasible region S this is my feasible region and lets and this is an sum open set.

And I am going to assume that f is differentiable so let me write the theorem. Let x star in s be an optimal solution of the optimization problem minimum minimize the function f subject to x in s.

Let f be differentiable and let S set of R n be open then we must have this. If you look at the derivative of f at with respect to x evaluated at x star evaluated at x star then this derivative must be equal to 0 ok, that is the conclusion ok. Let me make sure that you understand what this theorem is saying.

It saying that consider suppose consider this sort of optimization problem and suppose there is an optimal solution, it is not saying there is an optimal solution it is saying suppose there is an optimal solution and the optimization problem has the properties that f is differentiable and the feasible region is at open set. Then it has to be that the derivative of the function with respect to x evaluated at x star is equal to 0 correct.

(Refer Slide Time: 03:16)



So, now let us let us just prove this ok. Since, x star lies in S and S is open then what do we know? If x star is a point in S and the set it set S is open then what must be the case?

Student: (Refer Time: 03:37).

Then there is a ball around x star that lies completely in S right. So, the picture you should have in mind is; so suppose imagine this is your set s this is your point x star and there is a ball around it like this that lies completely here ok what that means, is; so there exists radius r greater than 0 such that x belongs to s whenever the distance between x and x star is less than r right ok.

Now what is unique about of the ball. What is so interesting about? Ok there is a ball around the point x star which lies completely in s. So, what if there is a ball around? So, what is so great about the ball?

Student: (Refer Time: 04:43).

So, there are many interesting things about the ball, but as we think that that is of relevance here is that remember the unit any ball ok of unit ball or any ball of any radius positive radius is essentially like all of r n shrunk down to a down to a smaller radius. So, what that means is; if you look at x star and you look at any other point in r n ok look at the direction starting from x star going towards that other point ok.

Let us call that other point 1. Look at this direction that starts from x star going towards y that direction ok this is a ray that starts from x star goes towards y. What I can do is without changing the direction I can simply shrink this ray all the way down to the point where now that ray is present completely in the ball right.

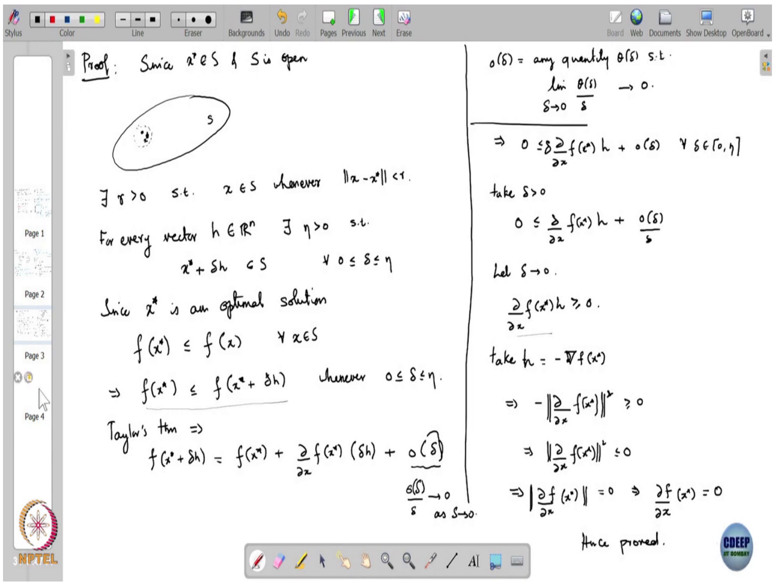
So, any direction I can conceive of which is present in starting from x star towards which I can go in r n, all of those directions are also present inside the ball itself; its asked form for one particular form x star towards some point in the ball right.

Student: Yes sir.

So, what that means, is; so for every vector h in R n I can find some eta positive ok such that if I look at the vector x plus delta h ok. I will tell you what delta is; this vector x plus delta h belongs to s for all delta that are greater than equal to 0 and less than equal to eta.

So, take any direction h right take any direction h and what I can do is take that direction and scale it down by multiplying it with the factor delta such that now x plus delta times that direction lies in the ball alright. So, this lies in the ball and therefore, in s. This is clear, ok? So, for every vector h there exists an eta such that one can do this ok.

(Refer Slide Time: 07:51)



Now, since we know that x star is optimal is an optimal solution. So, what does that mean? That means, that f of x star is less than equal to f of x for all x in S ok. But then I can in particular take a point like this take a point like x star plus delta h which is also lies in S in particular f of x star is less than equal to f of x star plus delta h whenever ok.

And now remember we that delta can be is the way you what is the factor by which you are shrunk the vector h ok. So, now, we are going to make delta smaller and smaller. So, if delta is smaller and smaller then an x star plus delta h is very close to x star itself and then we are in the regime of Taylors theorem ok.

So, now, what does Taylors theorem tell you? Taylors theorem implies that f of x Taylors theorem tells you that f of x star plus delta h is equal to f of x star plus this thing this derivative of f at with respect to x evaluated at x star times now x star minus x star plus delta h minus x star. So, that is just delta h right plus something that was going to 0 ok. Now we go back to what kind of factor it was it was going.

So, Taylors theorem said that there was ok my slightly bad choice of notation because here there is this I have used h for the function, but I hope you will still be able to get this. This factor here this factor here which was going to 0 it is such that h itself goes to 0 as y goes to 0, y minus a also goes to 0 as y goes to 0.

But because h goes to 0 as y goes to 0 what this means is this residue that is left is such that if I divide that by y minus a that also goes to 0 right. So, another way of writing Taylors theorem would be to say something like this. Would be to say that this is equal to this f of x star plus delta h is equal to f of x star plus this derivative times delta h plus something which I will write like this small o of delta.

What is what do you mean by small o of delta? Small o of delta is just a generic notation it stands for the following small o of delta is just a is any quantity such that any quantity say theta of any quantity theta of delta such that if I look at any quantity. So, theta of delta.

Student: Yes.

Such that I look at the limit small theta of delta divided by delta this goes to 0. As delta goes to 0, ok. So, its small o of delta means that it approve not only does it itself go to 0 even after dividing by delta it goes to 0 alright. So, now, the this is a way of writing things in optimization we make this kind of short hand we dont really need to know what this exact thing is; so long as we know at what rate it goes to 0 ok.

A quantity you can go to 0, but it, but it can. So, you can have a quantity that theta of delta that goes to 0. But then theta of delta divided by delta may not go to 0 right, but in this case this is a quantity the where theta of delta itself goes to 0 and upon dividing by delta also its going to 0. So, this is that is the significance of this.

So, what Taylors theorem is basically telling us is that f of x star plus delta h is equal to f of x star plus this derivative times delta h plus this quantity which is this, which means this is something such that small o of delta divided by delta goes to 0 as delta goes to 0 ok alright. So, if the e since that is the case now what we get is.

So, now f of x star is we know that f of x star is less than equal to f of x plus f of x star plus delta h so consequently we get that. So, if I look at this difference f of x star plus delta h minus f of x star that difference is greater than equal to 0. So, consequently I get that this is.

Now I this holds for all delta this holds for all delta in some neighborhoods 0 to eta. So, what I can do is take delta greater than 0 that allows me to divide throughout by delta in that case I have that this is true. Now since this is true I can now let delta go to 0 then that would give me that the derivative of f with respect to x evaluated at x star times h is greater than equal to 0 ok.

So, remember this derivative is a row vector, h is a column vector, this times this the scalar is greater than equal to 0 right ok; now I can do this ok. So, this is greater than equal to 0 now what did we want to show if you want to if you go back to the pre to the claim of the theorem, it was saying that the derivative is actually equal to 0 right.

But remember now what I have got is that this inner product. The inner product that is written here the derivatives times x this guy is e greater than equal to 0 for every vector h every vector h in r n right because I started off with h as an arbitrary vector and I got that this thing should be greater than equal to 0. So, since this has to be greater than equal to 0 for every vector h ok. What does that mean?

Student: (Refer Time: 16:37).

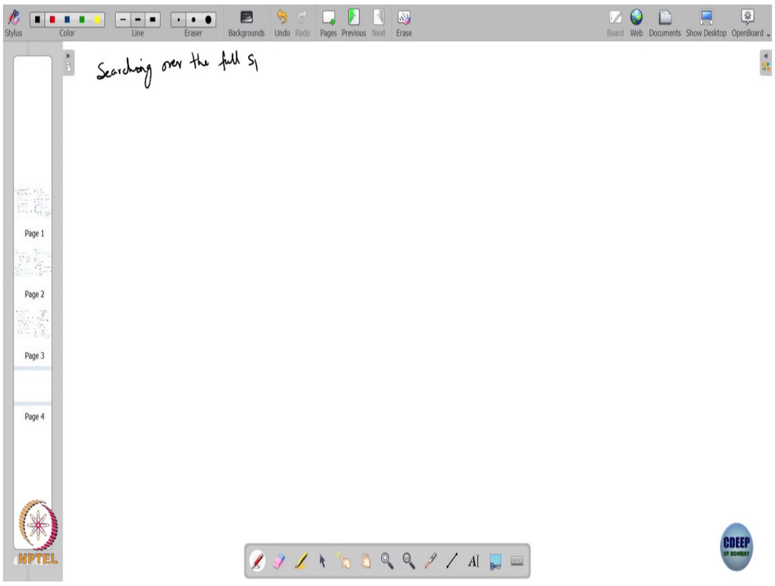
Yeah. So, this must also be true for minus h right or it must be true in particular if I take h to be this negative of this derivative itself right. So, I can take for instance h to be the negative gradient of f and so, that would then give me that my negative of this derivative evaluated at x star, the norm of the whole thing squared should be greater than equal to 0. (Refer Time: 17:21).

So, now, negative of the norm squared is greater than equal to 0 that just means that the norm of this derivative x squared is less than equal to 0, but it cannot be strictly less than 0 which means it has to be equal to 0; but, when if the norm is equal to 0 means the vector itself is equal to 0. So, ok so that proves the theorem, hence proved.

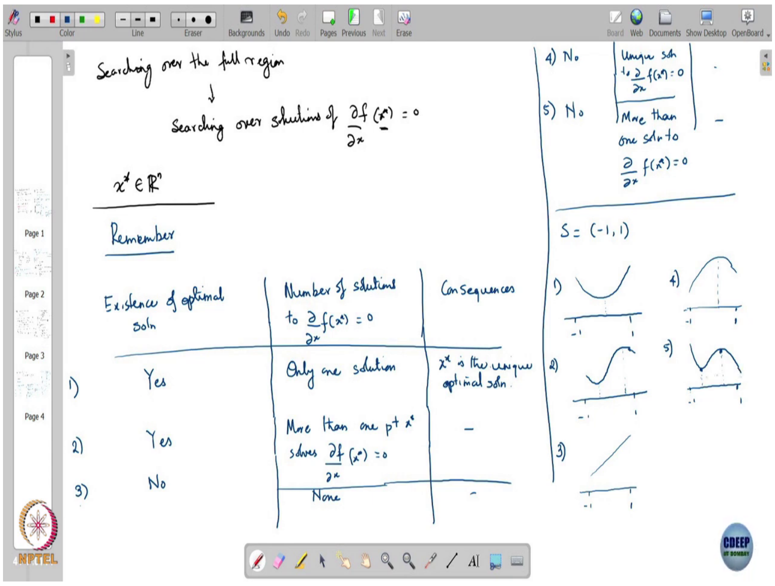
So, now let us understand the consequences of this and also understand what it what the limitations are. So, the consequence main consequence is remember that when you are searching for a minimum of a function over an open set, there are potentially there are infinitely many possible options there infinitely many possible choices.

What this is saying is, what this theorem is saying is that well the if you wanted to look instead of looking over all of them you can actually limit yourself to do the to searching over those points for which the derivative is equal to 0. Why does it say that? It is saying that because this is a necessary condition any minimum must satisfy this right.

(Refer Slide Time: 19:15)



(Refer Slide Time: 19:29)



So, the problem so searching over the full space full feasible region has become has been reduced to searching over solutions of this equation solutions means solutions x star of this equation right. Now what is what kind of equations are these? How many variables and how many equations? x star itself was in r n and this is the derivative of f with respect to x right.

So, there are n equations n variables now usually the difficulty arises that these equations are not necessarily linear. So, you need some numerical methods to find it to find the solutions, but it is still better than having to search you know in the entire space with it right.

So, this is how optimization works. You first try to find a mathematical result which constraints yourself and then you come up with an algorithm to go and find search in that particular region ok. So, this is otherwise there is you know you would have without a result you would be searching everywhere and that is too expensive ok alright.

So, I want to also this is one consequence, but also some cautionary remarks ok. Now remember what we said remember the theorem said that if there is a solution then this must hold ok. So, there are many different cases and sub cases that arise because of these kind of statements. So, let us go through all of them.

So, the first case so let me try and the first case is. So, let me write it this way existence of solution of optimal solution. Second thing is number of solutions to this equation ok and then you have consequences ok. So, let us take this case by case. So, suppose there is an optimal solution ok.

Now if there is an optimal solution then we know that it must satisfy it must satisfy the derivative that its derivative is equal to 0 ok. If there exists some optimal solution this must be true and now suppose there is only one solution to this bunch of equations. So, there are these non-linear equations there is we know that there is this these non-linear equations must be satisfied by the optimal solution, but there may also be other things other points that solve these non-linear equations right.

But suppose there is only one solution to these non-linear equations ok. So, suppose only one solution then what does that mean? That has to be the optimal solution right. So, every optimal solution must satisfy this these set of equations, but then these set of equations have only one solution then it has to be that solution is the optimal ok.

So, then the x star that solve these set of equation is the unique optimal solution. Now suppose there is a solution to this set of equations sorry to the optimization problem there is an optimal solution, but when you look at these non-linear equations ok you will find more than one point satisfying it.

So, more than one point solve this more than one point x star solve this. Now if there are more than one point that is that solves this what does that mean? Does it does not mean anything it any of those any of those could be optimal solution right ok. No particular nothing no particular conclusion that can be done ok.

Now suppose there is no optimal solution suppose there is no optimal solution and no solution here also then of course, then again nothing can be said. Suppose there is again going further suppose there is no optimal solution ok, but there is a unique point at which you can solve this system of equations.

What can one conclude about this? Again we cannot say anything see remember because the theorem started with the premise that you have a solution, if there is no solution the theorem will not tell you anything ok and again once again there is suppose there is no optimal solution and then there is more than one more than one solution to this that also does not same.

I will one second I will let me illustrate this for you to do some examples. So, let us do the first case first the first case is when there is an optimal solution and there is a unique solution ok. So, what I will do is, I will take s to be say minus 1 to 1 by the open interval minus 1 to 1 alright and let us draw some functions illustrating this case is minus 1, here there is 1 ok. So, I am looking I am minimizing this, there is and I find that there is a unique point here like this where the derivative has become equal to 0 ok.

Then it has to be that this is also the optimal solution that is. So, this is the happy case the second case is once again I am on minus 1 to 1 like this and I get I have a function that looks like this. So, now, if you look at the derivative of this function the derivative of the function has become 0 here and its also become 0 here.

So, there are now more than is there is more than one point where the derivative has vanished ok. So, we need to dig deeper in order to conclude what the optimal solution is. So, in this case we have to dig deeper though that is say that is a separate matter. If the theorem does not help us conclude anything from here about this ok.

What does case 3 look like? No optimal solution and no derivative also does not vanish. So, let us take so let us take for example, we are taking the I want to keep the domain as minus 1 to 1. So, I just take say I will just a line like this right. So, there is this has there is no optimal solution the infimum is not attained in this, it would have been attained at minus 1 if minus 1 was part of the domain, but it is not right.

So, there is the domain is open. So, there is the infimum is not attained and also that function shape is such that its derivative does not vanish in anywhere here ok. Let us look at 4. So, 4 is 4 is actually a case that is very tricky because so here you have a unique solution to the equation that derivative is equal to 0.

But there is no guarantee that the optimal solution actually does not exist. So, what does that look like? Yeah, so, just invert this figure 1, if I invert this figure 1 let us see for example, you do this look at this sort of figure. There is no there is no the infimum is not attained in this case, but there is this point here where the derivative has vanished ok.

So, this is case 4 is often responsible for many errors because before checking that a solution exists people jump into checking for solutions of derivative equal to 0 and then often come up with answers that are not actually solutions of the optimal solutions of the problem ok.

And then equation 5 would look say something like this, I have been talking of minus 1 to 1. Again there are now multiple points where the derivative is vanishing right, but there is no solution you know let me draw this see and further down yeah. This is yeah, solution does not exists.