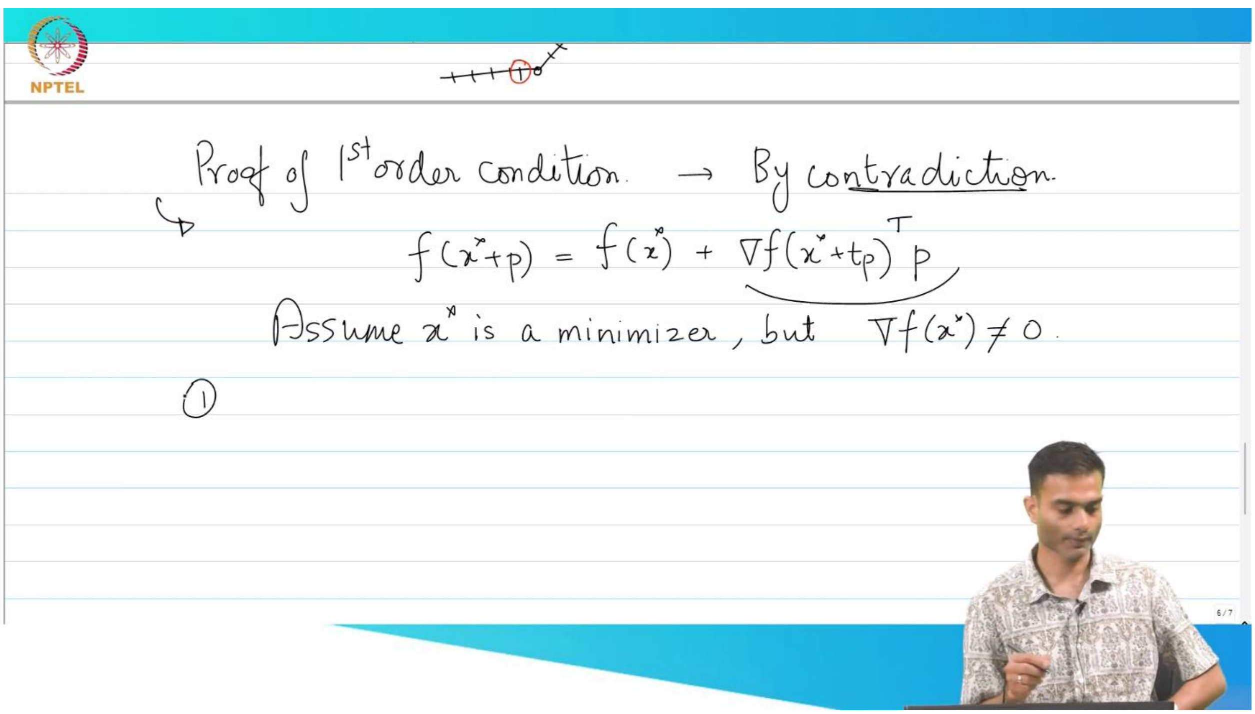
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 03
Lecture - 17

Unconstrained Optimization - 3 - Proof of 1st Order Condition

So, now let us look at the proof which will give us a lot of intuition. So, in this proof, I will introduce you to one of the oldest tricks in the math book, which is proof by contradiction. There are various other proofs possible, but we are going to go by contradiction, ok. So, the intuition behind it is going to be the following before we go to contradiction. So, I am going to write down the first order theorem which we will use, right? This is the first order theorem, right? And we are—what are we trying to investigate? Whether or not x^* is a minima of the function, right?

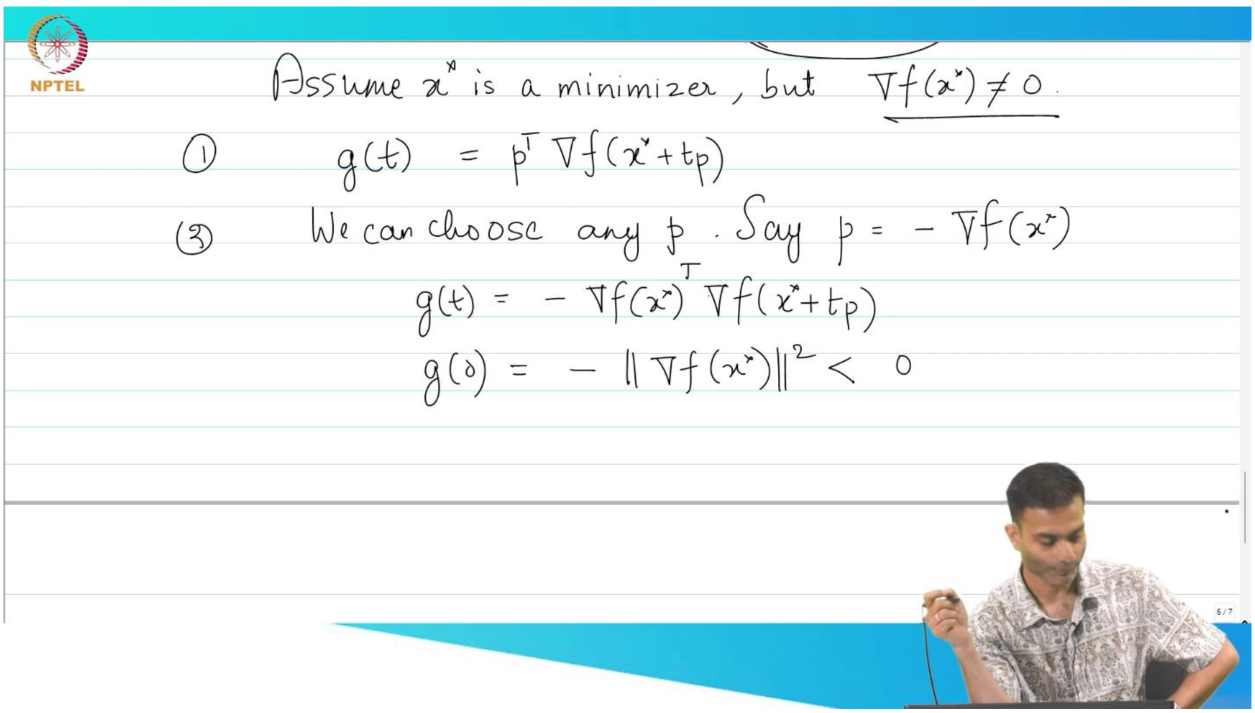
What does this remind you of? We did this in the previous class; this looks like a directional derivative, like the rate of change of the function in the direction p, right? That is what is the meaning of this guy. Now, if x^* is going to be a minima, then the rate of change of the function in any direction away from this point should be what? Positive. The function should always increase as I go away from this point. There should be no direction in which the rate of change is negative.



That means there exists a lower point if I were to find such a direction, right? So, it is basically—if you look back at this cone and this bowl over here, the rate of change of the function, no matter what direction you look, it is always increasing if I if I were at a true minimum. On the other hand, if I was sitting over here, there is one direction in which the rate of change is taking me to a lower point, right? So, that is the kind of intuition you can keep in mind as we go to the proof; it will become clear now, ok. So, let us start.

Yeah, in a direction. At x is of course, 0, but notice this is ∇f at x^* plus tp, right? Not at x^*, $\nabla f(x^*)$ should be 0. So, what is the contradiction that we will assume? We will— which says that assume x^* is a minimizer, but ∇f at this point is not 0. So, this is the contradiction that we are going to assume, ok. So, let us go step by step to see where this contradiction lands us, ok.

It is always easier to work with scalar functions, so I am going to cook up a scalar function, and the hint or the intuition for that scalar function actually is coming from this first term sitting over here; that is why I gave you the intuition first. So, I am going to call $g(t)$, t is a scalar, remember, from 0 to 1. So, $g(t)$ I am going to define to be precisely this term; I will just take the transpose the other way, fine. Now I am going to try to see if you know what good this does us, ok. So, now p is left unspecified; I have not said anything about p, p can be any point that I want.



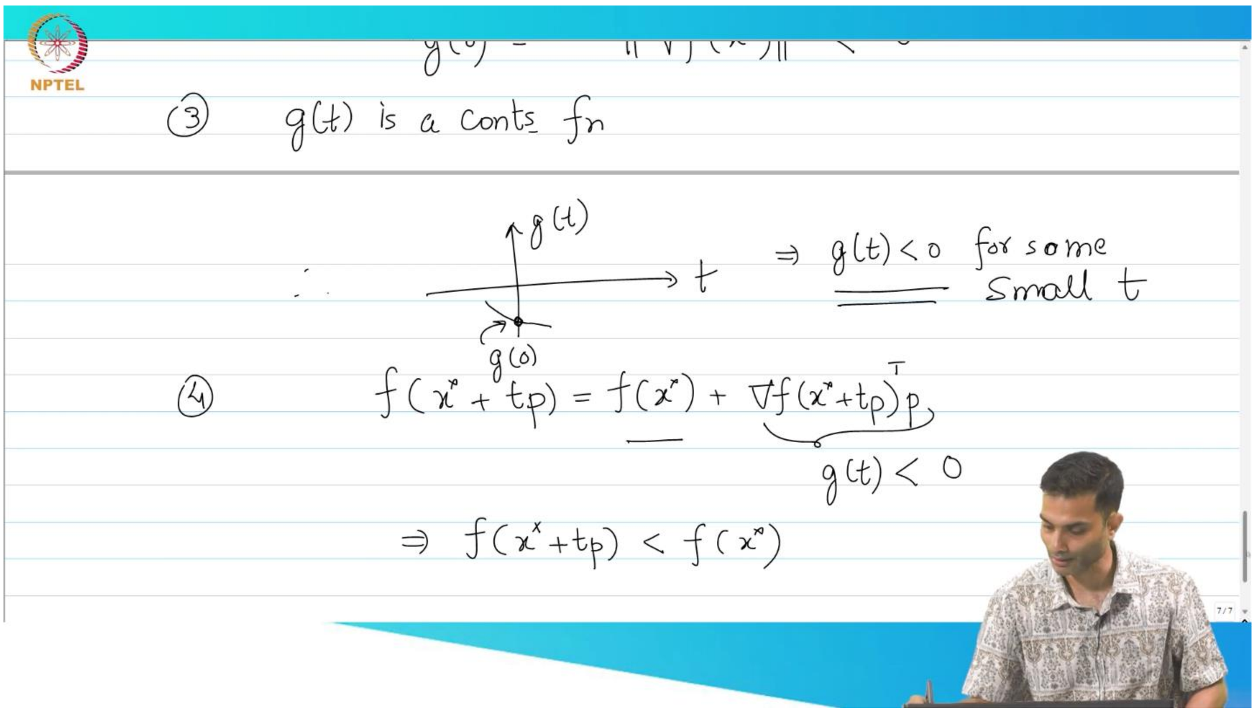
So, it is up to me to choose this p, right? So, we can choose any p, right? Supposing I choose p to be the negative gradient of f at x^*. So, what happens to my $g(t)$? Simply becomes $\nabla f(x^*)^T \nabla f(x^*) + tp$, right? Which it does not look that interesting; you might look at it and say, "so what?" But let us see what happens: what is $g(0)$? $-\| \nabla f(x^*) \|^2$.

Now, this is something that I can say something about. What can I say about this? Is it weakly or strictly? Strictly, strictly less than 0 because I have assumed that $\nabla f(x^*) \neq 0$. So, this is strictly less than 0. Now, because of what I had assumed about the function being continuously differentiable, right? So, here is the catch over here; let us note that down in step 3: $g(t)$ is a continuous function, right.

Therefore, if I were to graph $g(t)$, a simple graph like this, this is $g(t)$. At what point should I mark $g(0)$—above the origin or below the origin? $g(0)$ is strictly negative, so $g(0)$ must be some point over here—has to be, right? Further, $g(t)$ is a continuous function; that means that it will be something like this, let us say, or any shape; the point is that it cannot jump to plus infinity in a delta tending to 0 distance, right? That is the meaning that it is continuous, right?

So, it implies that $g(t)$ is still less than 0 for some small t; I can always find a small enough t such that $g(t) < 0$, right? So, now, we are beginning to tighten the noose.

Now, what is $g(t)$? We have already written it over here; this is my $g(t)$, ok. Now, as the phrase goes, our goose is almost cooked. Let us go back to our statement of the first order theorem. I had written that; that was the statement of the first order theorem, right? This is exactly—does this not look like $g(t)$?



It looks like $g(t)$. p is not specified. So, I specify some p to be $-\nabla f$. So, this is essentially nothing but $g(t)$, right? And this $g(t)$, as I have shown you over here by simple logic, is less than 0, right? Do you see the contradiction now?

I have $f(x^*) +$ a negative quantity is equal to $f(x)$ at some other point. What does that imply? It implies that now I have got f at some point given by t and p which is less than $f(x^*)$ strictly; its function value is lesser—its $f(x^*) -$ something, therefore, it is lower.

Right. So, this contradicts our original statement which said that assume x is a minimizer. Therefore, I have found a new point which is given by $x^* + tp$ which has a lower function value; that means x^* was not a minimizer, right? So, this would imply that $x^* + tp$ is a better minimizer, let us just say. Or you can say that x^* is not a minimizer, right?

So, that is the contradiction, yeah. Gradient, exactly, which is why I needed it in the theorem statement. Correct, exactly, right? So, you see that every part of a theorem statement gets used later; without it, you could not have done it, right? If it was just that it exists, but it is not continuous, then I cannot make this argument; then the gradient can suddenly jump like in the case of the kink, right? That is why continuously differentiable is needed. Very good, ok.

So, this is your first order theorem. I am not going to prove the second order theorem; it is there in Nocedal and Wright if you want to have a look—little lengthy, so we will not do it, ok. So, this is how the simplest test to search for a local minima is: find its gradient and set it equal to 0 and check whether or not it is equal to 0, which is a necessary condition. This is not sufficient for sufficiency; we have to go to second order conditions.