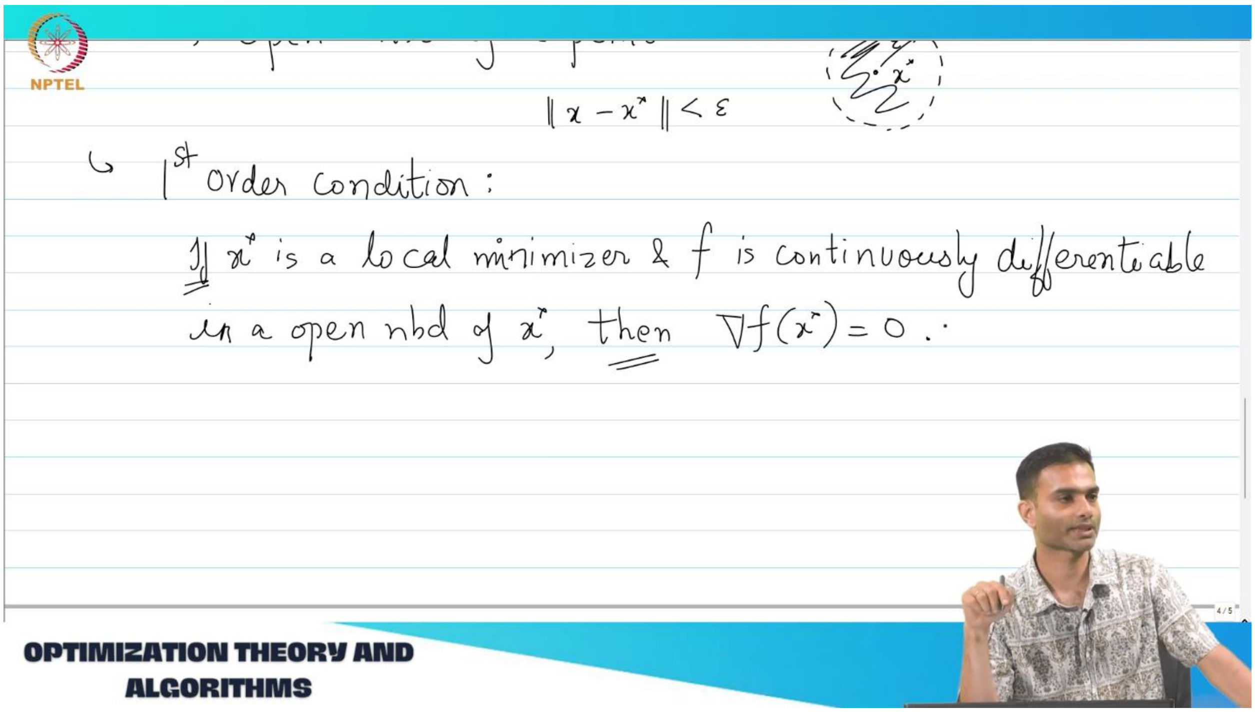
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 03
Lecture - 16

Unconstrained Optimization - 2 - Identifying a Local Minima - 1st and 2nd Order Conditions

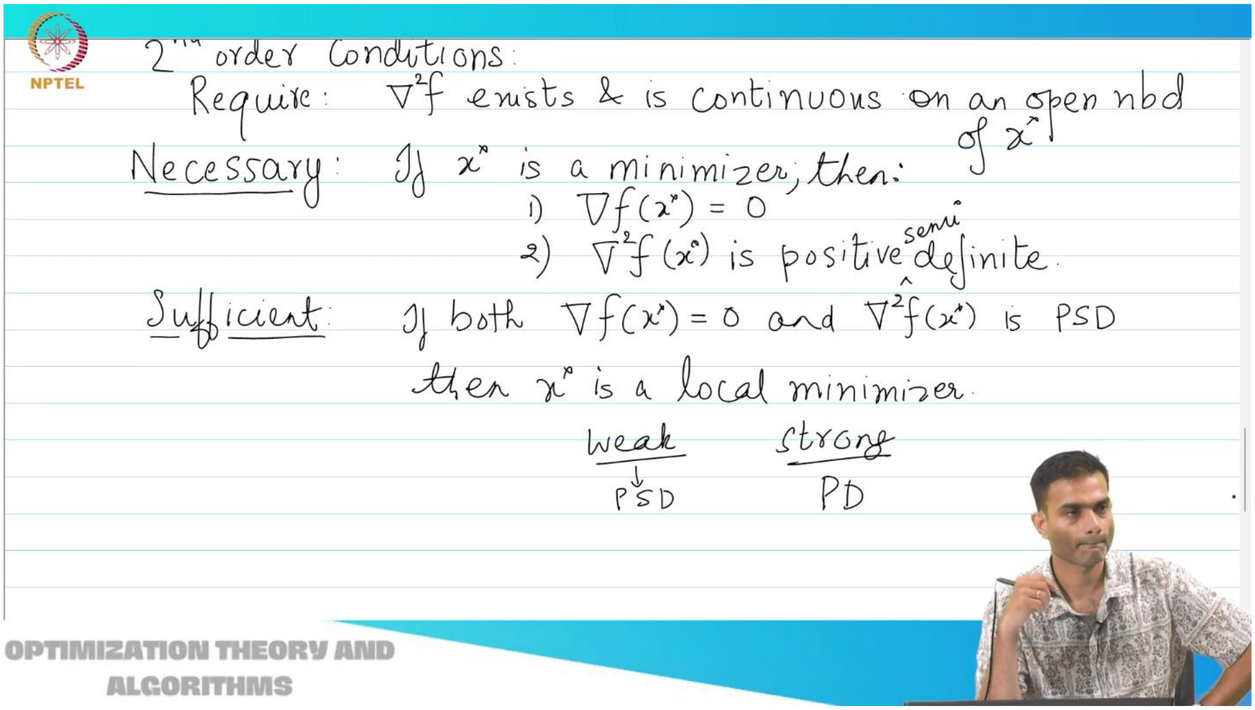
Less than equal to include the boundary; that would be a closed neighborhood. All right. So, we have all the nuts and bolts in place; let us write down our first order condition. So, let us—I will write it down and then we will interpret it. It is very straightforward. It is the generalization of what you had expected from calculus of single variable. What was in calculus of a single variable? What was your identifier of a minima? $f' = 0$.



What do you expect is going to happen in the multivariate case? $\nabla f^* = 0$, right? That is it. There are no surprises there. But let us write it down formally, ok. So, if x^* is a local minimizer, but I also have to say that f has to be differentiable, right? It is okay, but where in an open neighborhood of x^*, right? Then $\nabla f(x^*) = 0$, right? So, this is the "if" and this is the "then."

So, let me take a very quick example. Supposing I wrote my function f as a function that captured a chess game, and to win chess I have to minimize f. Can I use this theorem to identify a possible solution to a chess game? f is not continuously differentiable because, first of all, the variables themselves are discrete. I cannot make a continuous move, right? I can make a discrete move in one step, two steps, or whatever the way the horse moves. So, be careful of where you apply this; that is why this requirement that f is continuously differentiable is not a formality. You need to make sure that your function is differentiable.

Now, chess, you may say, okay, we are never actually going to write an optimization problem for chess, but it may happen that in your engineering problem you are designing some antenna surface and you have discretized it into discrete variables, and you try to apply this; it will not work, right? So, this is your first order condition; it is also a necessary condition. Further, if this holds true, then there is a special word—another special word reserved for x^*; it is called a stationary point. I am just—the only reason I am mentioning this is because some books will call it a stationary point. So, you should just know that there are various words by which this is referred.



So, this is quite straightforward; we will prove this. The proof is quite simple and gives us insight; it will give us actually insight into our first algorithm. So, I will go into the proof. Before we go into the proof, I want to write down the second order condition, ok. Any questions about this?

Yeah. It is not specified, ok. So, within whatever radius you specify, which is non-zero, that function should be continuously differentiable, right? And within that radius, if you find $\nabla f(x^*) = 0$, you are fine, right? Now, it may happen that within this neighborhood, you do not find any point where $\nabla f(x^*) = 0$. Then you would have to expand your neighborhood and check if the function is continuously differentiable. As long as the function is continuously differentiable, you have a whole set of techniques available for you to move from a starting point to a finishing point, right?

So, the thing that enables all this magic in some sense is differentiability. If you do not have it, you are stuck. Then you have to go to other things like evolutionary algorithms or whatever, right? For this entire course, we are going to assume at least first order differentiability. Ah, when you talk about convex functions, etc., they are already differentiable. So, it covers a vast range of engineering topics already. So, it is not restrictive by any means, ok.

Ah, linear programming, if you have heard of linear programming, for example, again, differentiability, etc., is there. So, it covers a large range of optimization. Ok. So, now let us note down our second order conditions. These are conditions for what? For checking whether the point I have got is a local minima or not, right?

And why do we need a second order condition? Can someone point out looking at this page? Because I have only got a necessary condition; I have not got a sufficient condition. So, the second order condition is actually what is going to give me a sufficient condition. Since it involves second order, it may be a little bit more difficult to evaluate, ok.

Now, for the second order conditions, we are going to use second order derivatives. So, my requirements are a little bit more stringent on the function. So, what do I require? Let us note that down. Obviously, the Hessian should exist; not only should it exist, it should be continuous. Exists and is continuous—and again we will use the formal phrasing on an open neighborhood. All right.

So, we will start with the necessary part, right? So, it says that if x^* is a minimizer, right? Then what all happens? First is what we expected from the first order condition: $\nabla\nabla f(x^*) = 0$, right? No surprises here. The second condition in I think a few previous classes we had had a brief discussion about since it involved second order derivatives, which was that someone wants to take a guess what might be the second condition. If I were doing this in single variable scalar calculus, what would be the condition for—what else would I need to have?

$f'' > 0$. Now, I have to generalize that to the Hessian being positive definite, right? Exactly. So, this gives me the necessary conditions—quite straightforward. Now, let us come to the sufficient conditions. Sufficient will work the reverse way; right? Here I had an "if" and a "then." Now I will start with the "then" and come to the "if" part, right?

So, if both this and the Hessian at x^* is positive definite, then x^* is a local minimizer. This is flipping it the other way. There is one small correction that we need to make here; can anyone point it out? The hint is weak strong, right? If I want my minimizers to include weak minimizers, then I will have to relax things a little bit and make this positive semi-definite, right? And then this will be positive semi-definite, right?

So, let us just mention that weak—if I am talking about weak, then it has to be positive semi-definite. If I am talking about strong, it should be positive. Any intuition as to why it should be positive definite? The intuition, if you want a geometric intuition, is again similar to the one-dimensional case. It is that if you look at—ok, so when I write a second-order approximation, second-order Taylor's approximation, what am I approximating my function as? A parabola.

Right, a parabolic equation because I am stopping at x^2. So, something $A + Bx + Cx^2$ in general is a parabola, right? Now, if I want my point to be a local minimizer, what can it be—any kind of parabola? No, it should be a parabola that is upward-opening; then that lower part of that parabola is the minimizer. Now, when I come to multiple dimensions, that same intuition—now instead of a line, think of a bowl in n dimensions, right?

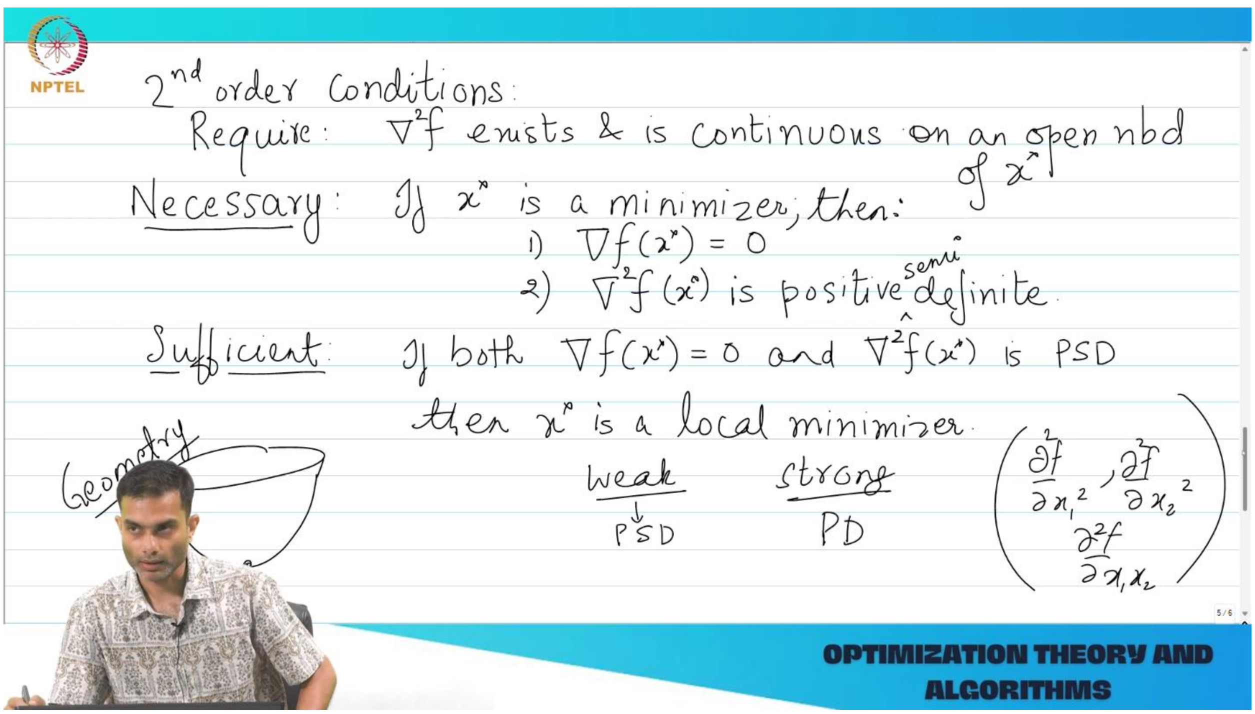
It is hard to visualize beyond 3 dimensions, but imagine a bowl in 3 dimensions—in n dimensions, and what you are saying and that whether or not it is a bowl as opposed to a downward bowl, a saddle hump—it can take various shapes. So, it can take some really arbitrary shape, but if I want that point to be a minimizer, that means there is no point of the function below that point. It has to be a bowl, and that geometric property is captured by the Hessian being positive definite.

So, it is a one-to-one geometric translation, ok. So, if you want, we can draw this. So, the geometric intuition—because if you remember, the second derivative of a function talks about curvature, right? If the curvature is always positive, then it is always one kind of shape. Correct?

It is always—yeah. Do not think—I mean, it is a matrix, alright, but keep the geometric—in I mean, what is a matrix? Is this a collection of—in this case three values, three mixed derivatives, right? So, there is—for example, if I have x_1, x_2, it is going to be this. This is essentially what is there inside the Hessian, right? And all of these second-order derivatives are talking about the curvature of the function in these dimensions. That is pretty much correct, correct. And the intuition which you are mentioning we will get actually—we will open it up properly in the next few slides, ok? Good.

Yeah, it is always good to have this geometry in mind. So, having stated this, does anyone have any other questions on the statement of the second order condition?

Yeah. At that in the neighborhood of that point.



Well, if you are talking about whether it is continuous at just one point and not at any other points, is that what you are saying? No, if it is—see, if you are saying that the derivative—I mean the Hessian exists, and if it exists, you are saying therefore, it is continuous, right? Well, I can have a second-order derivative which exists, but need not be continuous; I can cook up functions like that, right?

Can you think of an example? For example, a function—let us come down one derivative—let us talk about the first-order derivative. This function is continuous; is it differentiable? So, continuity does not imply differentiability, right? So, there is going to be one point where the derivative is not defined.

Well, the derivative here exists at all of these points, but not at this point. But what I—what this theorem is saying is that in the small neighborhood of this point, you check whether it is continuous; that is what I need to check. So, what is the motivation to have a neighborhood around that point? Is that this will come to the next—what we are going to talk about is come up with some algorithm.

We do not see—this is just a check or a test for a minima. That doesn't really help me because when I start my problem, I don't have any candidate; I have a starting point. There is not much fun in checking the starting point; most likely I would have started at the wrong point. I need some way to move, and if I want to move, I need this open neighborhood in which I can move. If that open neighborhood does not exist, then what use is it? It is a nice theorem, but it doesn't help me.

So that is why the open neighborhood allows me to move to a better point, to a better point. That is the motivation.