Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan D

Department of Mathematics

Institute of IIT Kanpur

Lecture: 3

Hello everyone, so we will get started with the third lecture. Recall that in the last lecture we were considering the following optimization problem, right. So, we consider a twice differentiable function whose domain is from R^n to R and a set S that is a subset of R^n, and we want to find the minimum of $f(x)$ when x is restricted to S. So, this is the optimization problem that we are considering.

And if you recall, we learned two theorems in the last lecture. The first is, so let f be continuous, continuous function and S be closed and bounded; then there exists a global solution for this problem. For the problem min $f(x)$, where $x \in S$. So, this is the first theorem that we learned and that is regarding the existence of a global solution, and the second theorem is regarding the relationship between the local solution and the global solution. So, in case f is a convex function and S is a convex set, then $x^\star$ is a local solution implies $x^\star$ is a global solution. So, these are the two theorems that we learned. To give a recap of what is a convex function. You can recall that f is convex if and only if

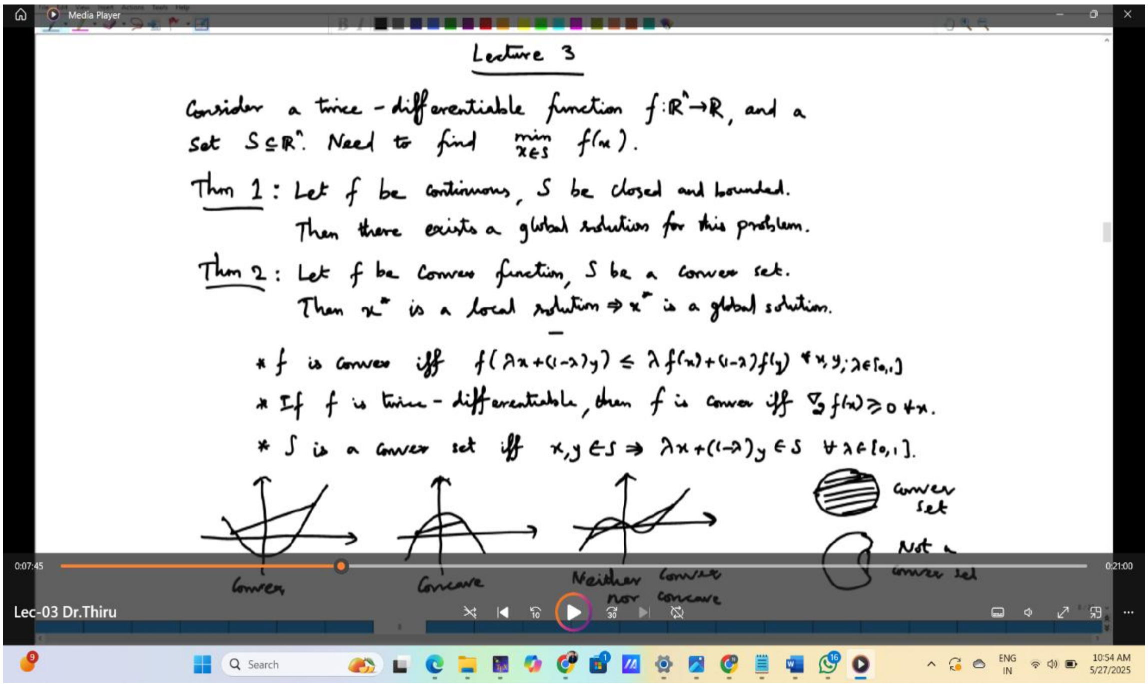
$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

for all x, y and for all $\lambda \in (0, 1)$. The other condition is that if f is twice differentiable, then f is convex if and only if $f''(x) \geq 0$ for all x. This is when $n = 1$.

If we consider any general n, this turns out to be as follows. You consider the Hessian matrix and verify whether this is positive definite for all x. This is possibly a simpler way when f is twice differentiable. Again to recall, a convex set, right. So, S is a convex set if and only if $x, y \in S$ implies $\lambda x + (1-\lambda)y \in S$ for all $\lambda \in (0, 1)$. Recalling the figures, a convex function looks like this: where, if you draw a line between two points on the curve, the line always lies above the curve. So, this is a convex function. And when you have a function of this sort where the line always lies below the curve, we call that a concave function. And when you have a function like this, you have certain set of points where the line is below the curve and certain set of points where the line is above the curve. So, this function is neither convex nor concave.

Regarding convex sets, you can recall that when you consider a set of this sort, where when you take any two points on the set and join it, you can see that all points in the line segment lie within the set. Such a set is called a convex set. Even when you have one pair of points for which this does not hold, such a set is not a convex set. Note that there is no concave set, a set is either convex or not convex.

(Refer slide time 7:55)



So, with this, let us go to the next question. Note that we are actually looking at the problem min $f(x)$ subject to $x \in S$, right. So, we have a few characterizations based on whether the solution that we are looking for is a global solution or a local solution or whether a solution itself exists for a given problem. The more relevant question for this problem is, suppose a solution exists, how do we go about finding such a solution?

So, we will start working on answering that question. Recall that optimization problems are classified into unconstrained optimization problems and constrained optimization problems. It is unconstrained when $S = \mathrm{R}^n$ for some n. So, we start with an unconstrained optimization problem. So, consider the problem min $f(x)$ subject to $x \in \mathrm{R}$. We are actually considering an unconstrained problem where $n = 1$.

So, f is a function from $\mathrm{R} \to \mathrm{R}$. We want to find the local solution or global solution of a general unconstrained optimization problem. The first step is to find the set of critical points. So, what is the set of critical points? We define the set of critical points, call it C, as the set of all x such that $f'(x) = 0$.

The first derivative of f vanishes. So, find all those points where $f'(x) = 0$. We call that set the set of critical points. If a solution is the local minimum or local solution of the above problem, then it must be a critical point. You must have learned this as the necessary conditions or the first order conditions. The necessary first-order condition is that the local solution $x^\star$ satisfies $f'(x^\star) = 0$. So, if a solution does not satisfy $f'(x) = 0$, then such a point cannot be a critical point.

Note that this is only a necessary condition for a point to be the local solution. But it is not the sufficient condition. What is the sufficient condition? A critical point $x^\star$ is a local minimum if $f''(x^\star) > 0$. And it is a local maximum if $f''(x^\star) < 0$. You can ask what happens if $f''(x^\star) = 0$. In such a case further probe will be needed.

Let us consider $f(x) = x^2$. In such a case, we first find the set of critical points that is x such that $2x = 0$, which is just the singleton $\{0\}$. There is only one critical point, 0.

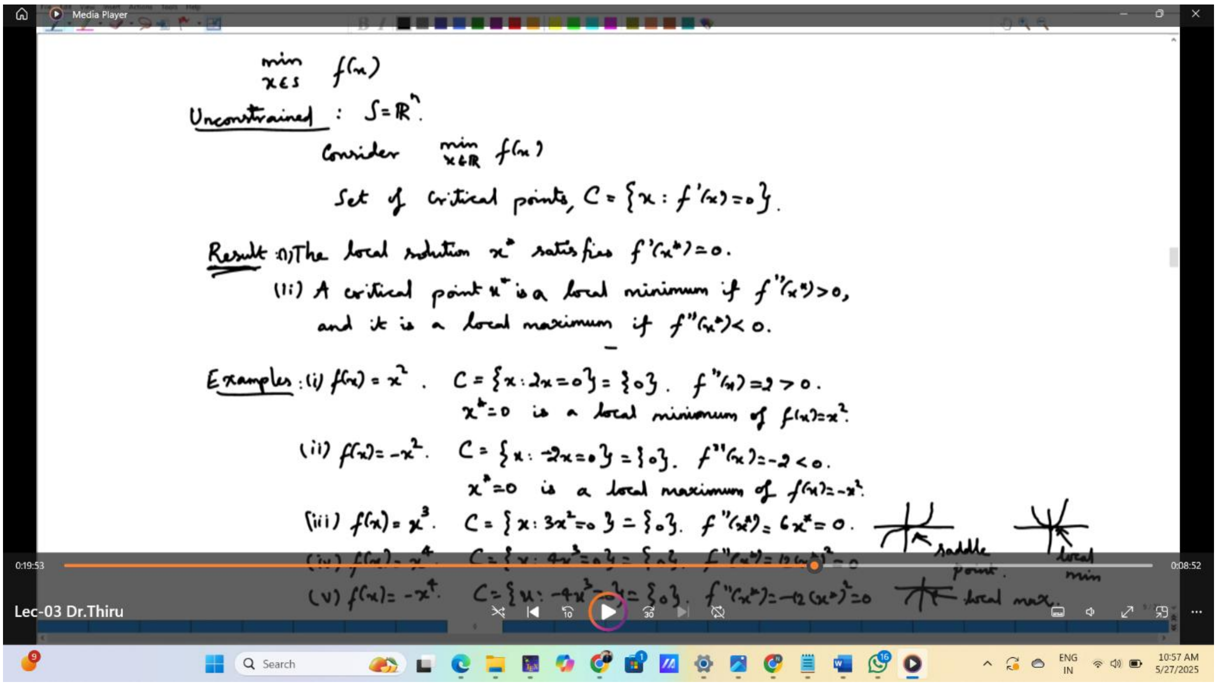
$f''(x) = 2 > 0$. This implies that $x^\star = 0$ is a local minimum of $f(x) = x^2$.

Now, consider the example $f(x) = -x^2$. The critical point is x such that $-2x = 0$, again $\{0\}$. $f''(x) = -2 < 0$. Here, $x^\star = 0$ is a local maximum of $f(x) = -x^2$. This result should not be surprising because when you are maximizing a function, you can consider minimizing the negative of the function.

Now consider $f(x) = x^3$. The critical point is x such that $3x^2 = 0$, again $\{0\}$. $f''(x) = 6x$, so $f''(0) = 0$. In this case it turns out to be neither local minimum nor local maximum. It is actually termed as a saddle point.

Another example: $f(x) = x^4$. The critical point is again $\{0\}$, $f''(x) = 12x^2$, $f''(0) = 0$. Plotting $f(x) = x^4$, you see $x = 0$ is a local minimum.
Now $f(x) = -x^4$, again critical point is $\{0\}$, $f''(x) = -12x^2$, $f''(0) = 0$. Here $x = 0$ is a local maximum.

(Refer slide time 19:54)



What this tells us is: if you consider the problem min $f(x)$ subject to $x \in$ R, the first step is to find the set of critical points, x such that $f'(x) = 0$. Then for each $x^\star \in C$, compute $f''(x^\star)$. If $f''(x^\star) > 0$, then $x^\star$ is a local minimum. If $f''(x^\star) < 0$, then $x^\star$ is a local maximum. If $f''(x^\star) = 0$, then further probe is needed.

This is the case when $n = 1$, i.e., minimizing $f(x)$ where $x \in$ R (a uni-variate function). Now suppose $n > 1$. We want to minimize $f(x_1, x_2, \ldots, x_n)$, where each $x_i \in$ R. So, how do we approach this problem?

Step 1: Find the set of critical points x such that $\nabla f = 0$. That is, all partial derivatives $\frac{\partial f}{\partial x}$ =0 for $i = 1$ to n.

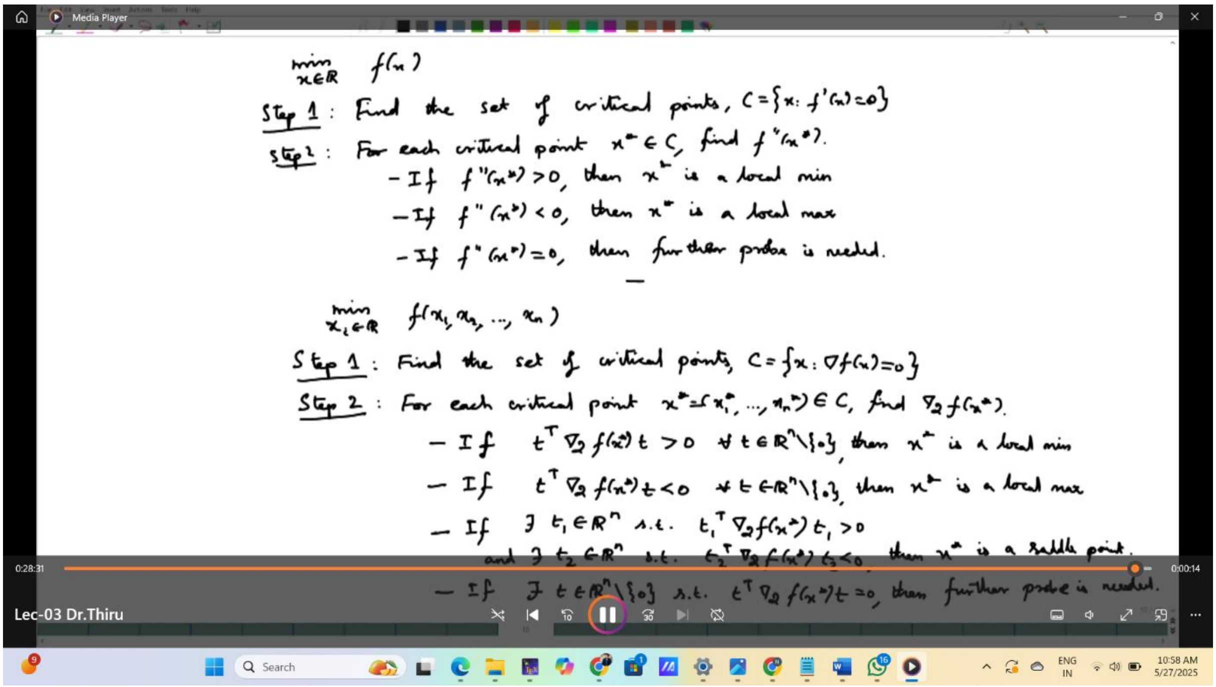
For each critical point $x^\star = (x_1, x_2, \ldots, x_n) \in C$, find the Hessian matrix $\nabla^2 f(x^\star)$. This consists of all second derivatives.

To classify $x^\star$, consider the quadratic form $t^T \nabla^2 f(x^\star) t$ for $t \in$ R$^n \setminus \{0\}$.

- If $t^{\top}\nabla^2 f(x^{\star})t > 0$ for all $t \neq 0$, then $x^{\star}$ is a local minimum.
- If $t^{\top}\nabla^2 f(x^{\star})t < 0$ for all $t \neq 0$, then $x^{\star}$ is a local maximum.
- If $\exists\ t_1, t_2$ such that $t_1{}^{\top}\nabla^2 f(x^{\star})t_1 > 0$ and $t_2{}^{\top}\nabla^2 f(x^{\star})t_2 < 0$, then $x^{\star}$ is a saddle point.

If there exists $t \neq 0$ such that $t^{\top}\nabla^2 f(x^{\star})t = 0$, then further probe is needed. So, you can see that the second-order conditions in multivariate settings are more complex than in the uni-variate setting. We will explain this further in the next lecture.

(Refer slide time 28:35)



Thankyou.