Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan D

Department of Mathematics

Institute of IIT Kanpur

Lecture: 2

Welcome you all to the second lecture. So, if you recall in the last lecture, we started with an optimization problem of the following form. So, we are going to minimize $f(x)$ subject to the condition that x is in some constraint set S. And we also discussed about the solution concepts. One is the global solution, and the other is the local solution, to give a one-minute recap of the examples.

So, let us consider $S = \mathbb{R}$. So, let S equal to the real line, and if $f(x) = x$, so no global local solutions-neither a global solution nor a local solution. So, if $f(x) = (x-1)\,x\,(x+1)$, then no global solution. But a local solution at $x = 0.5$. And if $f(x) = x^2$, then a unique local solution at $x = 0$; this is also a global solution. I will write it as x^* to clarify. And if $f(x) = (x-2)(x-1)\,x\,(x+2)$.

So, we have a local solution at $x^* = 1.5$. And a global solution, which is different from this, a global solution at some $x^* \in (-2, 0)$. I mean this can be precisely found by solving this quartic equation.

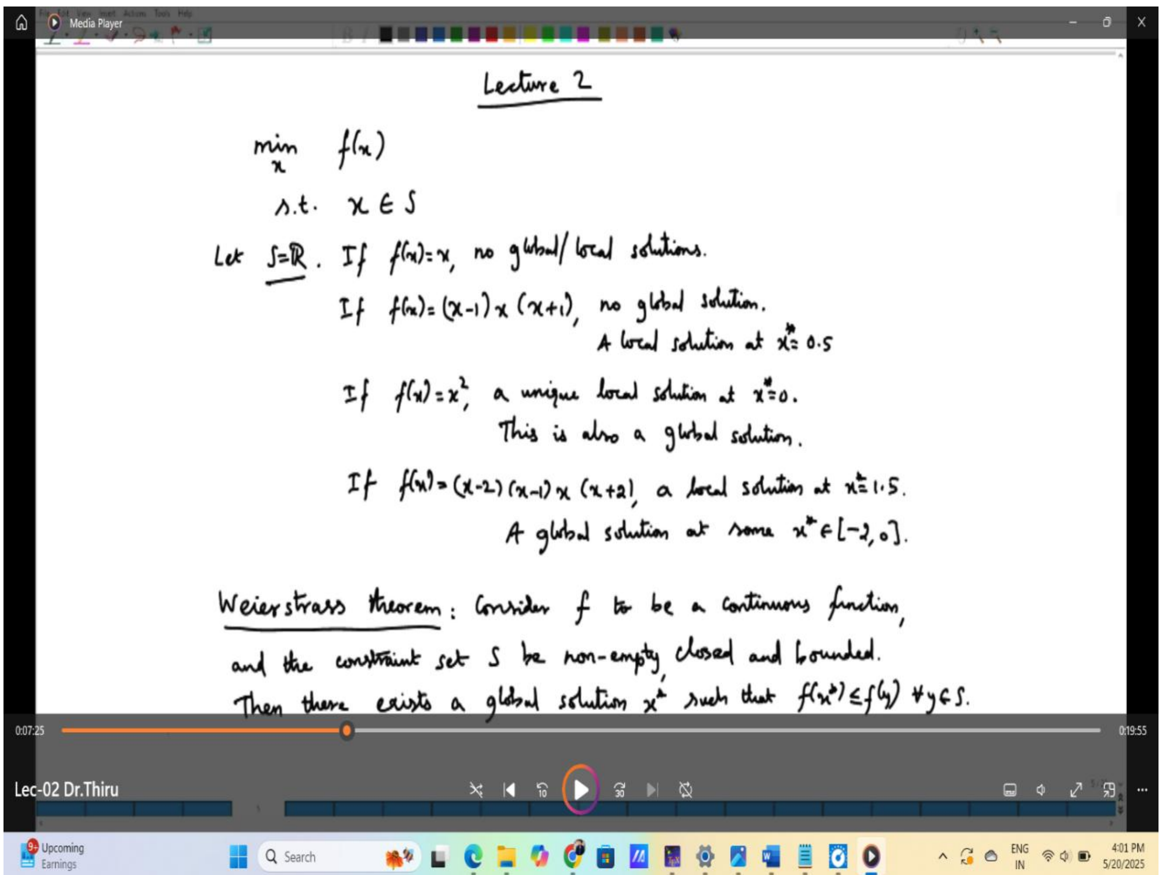
So, you can actually see that given a function $f(x)$, you can have a global solution or a local solution, or you may not have a global solution or a local solution. Even if you have both a local solution and a global solution, it could turn out that there could be local solutions which are not the global solutions. So, now we actually ask two questions. One is regarding the existence of a global solution or a local solution. Right. So, that is given a problem, how do I find whether there exists a global solution or not? So, that is one question. The other question is, when can I say that a local solution is a global solution as well?

So, these are two very important questions, and it has been answered many years back. So, the first question was answered by Weierstrass, and it is famously called as the Weierstrass Theorem, and what does it say? It says that consider f to be a continuous function and the constraint set S be non-empty, closed, and bounded. If all of these conditions are satisfied, then there exists a global solution x^* such that $f(x^*) \le f(y)$ for all $y \in S$. **This is called as the Weierstrass theorem.**

What this says is very simple. It says that the function f needs to be continuous, and the constraint set has to be a non-empty set. It has to be a closed set and a bounded set. If all of these happens, then there will certainly exist a global solution. We have actually, assumed initially that f will be a twice differentiable function. So, it is certainly continuous, we know that, and the continuity of the function is not a problem in all the examples that we have given here.

But if you see here, you can see that the constraint set S is actually the real line itself. So, it is certainly non-empty, it is certainly closed, but it is not bounded. It is neither bounded on the left nor bounded on the right. So, since the condition of boundedness is violated, that is the reason why we actually have examples where there is no global solution. This actually gives a sufficient condition for the optimization problem to admit a global solution.

(Refer Slide time 7:40)



Let us now move on to the second question. When does it happen that the local minimum turns out to be a global minimum? There again we have a theorem. Let f be a convex function. The objective function which we are minimizing is assumed to be a convex function, and let the constraint set S be a convex set. Then the local solution x^* is also a global solution.

Maybe I will spend a few minutes explaining what is a convex set and a convex function. When is a set called a convex set?

A set X is convex if $x, y \in X \Rightarrow \lambda x + (1-\lambda)\, y \in X$ for all $\lambda \in (0, 1)$.

To explain what this is, consider that we have the following set. It is a circle not drawn very well, but I suppose you understand, and the set X is all points within this circle and also on the boundary. I claim that this set is a convex set. Why is that so? Take any two points on this set, for example, this point and this point, and draw a line in between them. All points on this line segment is within the set. This set is actually a convex set, and that is exactly what this definition says.

Take any two points, say x and y. Then $\lambda x + (1-\lambda)y$ actually represents the points on the line segment joining x and y. You can see that when you put $\lambda = 0$, you get the point y, and when you put $\lambda = 1$, you get the point x, and if you put anything in between, say $\lambda = \frac{1}{2}$, you get the midpoint $\frac{x+y}{2}$, and similarly you can get all the other points.

This is an example for a convex set. An example for a non-convex set: let us say I have this kidney-shaped set where I have all the points on the boundary as well as the interior. Now, this is not a convex set because if I join these two points, say this is a point on the set and this is another point on the set. If I join these two points, the line segment in between does not lie within the set. Such a set is a non-convex set.

Now, **what is the definition of a convex function**? Consider a function $f : X \to \mathrm{R}$ where X is a convex set; then f is a convex function if

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y), \qquad \forall x, y \in X,\ \forall \lambda \in (0, 1).$$

This is the definition for a convex function. What is the physical meaning of this? Consider this particular function and consider two points x and y, and join them. If this line segment lies above the curve, then the function is a convex function. This should happen for any x and y that you take.

What this says is that the right-hand side here is the line segment joining $f(x)$ and $f(y)$, and the left-hand side here is the function value at that particular point. So, the function value is less than or equal to the line segment that joins the points x and y. When this happens for any x, y in the domain of f, then we call such a function as a convex function. If the exact reverse inequality holds, then such a function is called as a concave function. If

$$f(\lambda x + (1-\lambda)\, y) \geq \lambda f(x) + (1-\lambda)\, f(y), \;\; \forall x, y \in X, \;\; \forall \lambda \in (0, 1)$$

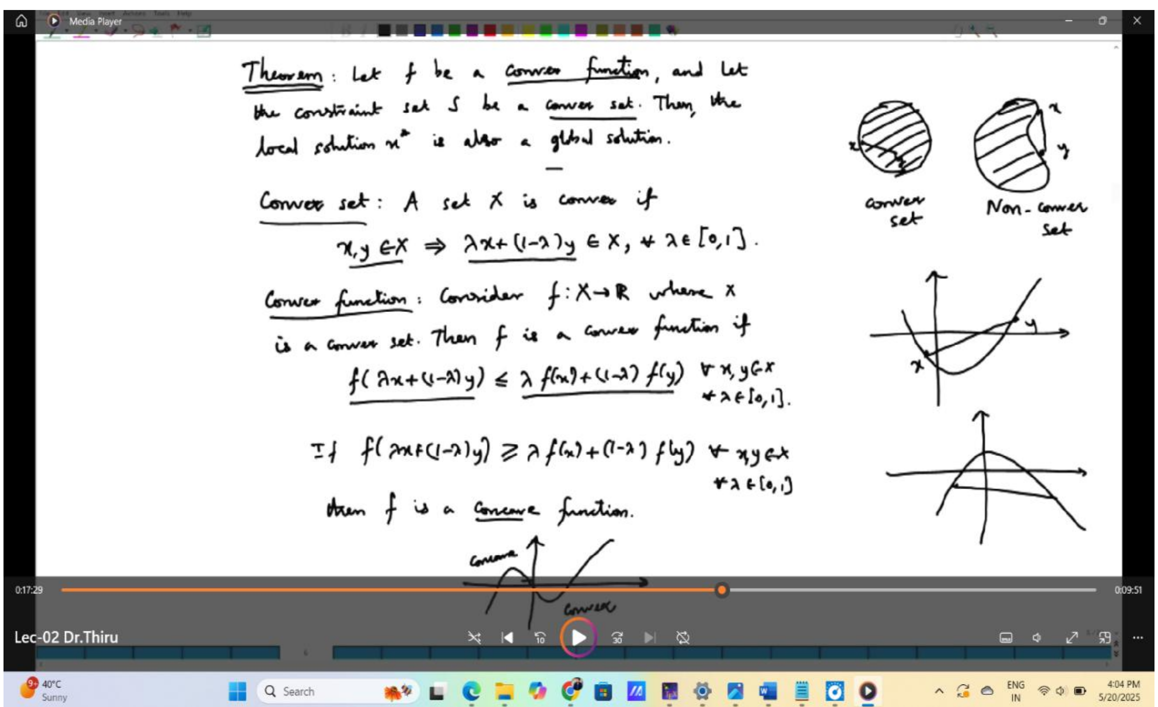
then f is a concave function.

If you consider the inverted curve of what I have drawn here, something like this, then the curve always lies above the line segment, or the line segment lies below the curve. In such a case, it is called as a concave function. We could have functions which are neither convex nor concave. In certain parts it is convex, and in certain parts it is concave. For example, you can have functions like this: in this part up till here it is concave, and in this part it is convex.

Let us come back to the theorem. What does this say? It says that if f is a convex function and the constraint is a convex set, then the local solution x^* of the optimization problem is also a global solution.

Let us check which of these functions is actually a convex function. $f(x) = x$ is actually a convex function. Another way of verifying whether a function is convex is using the Hessian matrix. If the domain of $f \subset$ R, then you can do it with the second- order derivative.

(Refer Slide time 18:24)



Suppose that f is twice differentiable. Then f is convex if and only if the Hessian matrix is positive semidefinite for all x in the domain of f.

The Hessian matrix is the matrix consisting of the second-order derivatives of f:

$$H(f) = \begin{bmatrix} \partial^2 f/\partial x_1^2 & \partial^2 f/\partial x_1 \partial x_2 & \dots & \partial^2 f/\partial x_1 \partial x_n \\ \partial^2 f/\partial x_2 \partial x_1 & \partial^2 f/\partial x_2^2 & \dots & \partial^2 f/\partial x_2 \partial x_n \\ \dots & \dots & \dots & \\ \partial^2 f/\partial x_n \partial x_1 & \partial^2 f/\partial x_n \partial x_2 & \dots & \partial^2 f/\partial x_n^2 \end{bmatrix}$$

This is an $n \times n$ matrix. We call this positive semi-definite if the following condition is satisfied:

$$t^T H t \geq 0 \quad \text{for all } t \in \mathrm{R}^n$$

If this condition is satisfied, then we say that $\nabla^2 f(x)$ is positive semi-definite.

Now coming back to the examples. When $f(x) = x$, we have $f''(x) = 0$, that is convex, but it does not have a local solution at all.

Now when you take the example $f(x) = (x\text{-}1)x(x + 1)$, there is a local solution at x^* but it is not a global solution. So that means that this function may not be convex. Let us check:

$$f(x) = x^3\text{-}x, \qquad f''(x) = 6x$$

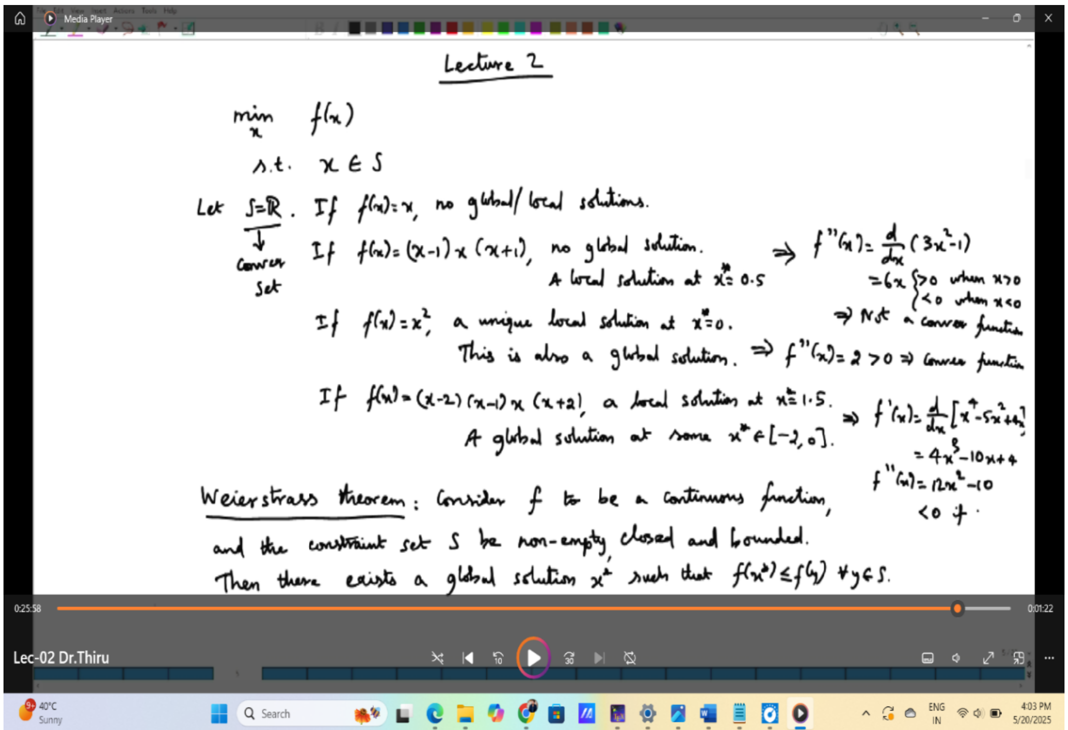
This is positive when $x > 0$, negative when $x < 0$, so not convex.

Next, $f(x) = x^2$, here $f''(x) = 2$, always positive, so convex function. Since this is a convex function defined on a convex set, and R is a convex set, every local solution is also a global solution. Now for the quartic function at the end:

$$f(x) = x^4 - 5x^2 + 4x, \quad f'(x) = 4x^3 - 10x + 4, \quad f''(x) = 12x^2 - 10$$

This is less than 0 if $|x| < \sqrt{\frac{5}{6}}$, again, not a convex function.

(Refer Slide time 26:20)



Coming back to the theorem, if f is a convex function and the constraint set is a convex set, then the local solution x^* of the optimization problem is also a global solution.

In the next class we will look at, given an optimization problem, how to find a local solution or a global solution, what are the characteristics of a global solution or a local solution for both a constrained optimization problem as well as an unconstrained optimization problem. Thank you.