Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan D

Department of Mathematics

Institute of IIT Kanpur

Lecture: 5

Hello everyone, this is the fifth lecture. Recall that we looked at many examples of unconstrained optimization problems in the last lecture and have also started with the constrained optimization setup.

So, when you look at the constraint setup that we had in the last lecture. So, we wanted to minimize $f(x)$ subject to some set S, and that set is not any set, but a set described by p inequality constraints and m equality constraints.

We also looked at an example, the utility maximization example that we started with in the very first lecture. So, let us first understand how to construct a general solution. We have this particular problem: minimize $f(x)$ subject to $g_i(x) \neq 0$ for all $i = 1, \dots, p$ and $h_j(x) = 0$ for all $j = 1, \dots, m$.

We first construct something called the *Lagrangian*. Let us assume that f is a function from R^n to R. So, that means x is a vector $(x_1, \dots, x_n)$.

So, L is actually a function of the x vector and certain vectors called the Lagrange multipliers for inequality constraints $\lambda_1, \dots, \lambda_p$ and the Lagrange multipliers for equality constraints $\mu_1, \dots, \mu_m$. So, Lagrangian is a function of x, λ, and μ. So, there are a total of $n + p + m$ variables.

So, what is the expression? It is

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{p} \lambda i g i(x) + \sum_{j=1}^{m} \mu j h i(x) \ ,$$

This is what we call the *Lagrangian function*. So, this is a new function that I have constructed. We have the function f, which was only a function of x, the functions g_i and h_j.
They are also only functions of x. But the Lagrangian that we have defined here is actually a function of the x vector, λ vector, and the μ vector, and it is defined as $f(x) + \sum_{i=1}^{p} \lambda i g i(x) + \sum_{j=1}^{m} \mu j h i(x)$.

Why did we define this? We are actually going to now find the set of *critical points*. So, the step 1 in any solution finding in optimization is to find the set of critical points. But what do I mean by how do you characterize the set of critical points? It is the set of points that satisfy the following set of constraints. What are they?

Take the gradient of L with respect to x, and equate it to zero. So, this will give you n constraints.

So, the x, λ, and μ that we are going to find should satisfy this constraint: when you take the Lagrangian, take the gradient of the Lagrangian only with respect to the x variables. So, that must be equal to zero. This actually means that the following must be satisfied:

$$\nabla f(x) + \sum_{i=1}^{\mu} \lambda i \nabla g i(x) + \sum_{i=1}^{\mu} \mu j \nabla h j(x) = 0,$$

where the right side is the zero vector.

You must be able to give x, λ_i's, and μ_j's in such a way that this vector equation holds. That is the first step, but you have further constraints to be satisfied.

So, the x that you give must satisfy the constraint:

$$h_j(x) = 0, \qquad \forall j = 1, \ldots, m,$$

If the constraints are not satisfied, then the point you have constructed is not in the constraint set. So, that needs to be satisfied.

The other constraint is a little tricky. You need

$$g_i(x) \leq 0,$$

and other than that, you need

$$\lambda_i g_i(x) = 0,$$

and

$$\lambda_i \geq 0, \qquad \forall \ i = 1, \ldots, p,$$

These set of constraints that we have defined are called the *Karush-Kuhn-Tucker* conditions, or in short called the *KKT conditions*. KKT stands for Karush-Kuhn-Tucker. So, the set of points (x, λ, μ) that satisfy the Karush-Kuhn-Tucker conditions form the set of critical points.

Of course, it does not stop there. We also, after finding the set of critical points, look at the *second-order conditions*. So what are the second-order conditions?

First, we look at the set of active constraints. So, what are the set of active inequality constraints? Active constraint set. Let me call that I, so

$$I = \{i : g_i(x^*) = 0\}.$$

I will also call it $I(x^*)$ since it varies with respect to x^*.

You can see that the critical point x^* may satisfy this particular constraint $g_i(x)$

either with equality or it could be strictly less than zero. So, we could have

$$g_i(x^*) = 0,$$

or

$$g_i(x^*) < 0.$$

So, we pick those sets of constraints from 1 to p for which $g_i(x^*) = 0$. So those constraints are called *active constraints*, and the others are called *inactive constraints*, basically. You will see why the set of active constraints is needed shortly.

Now, I am going to construct a set called the *tangent space*. The tangent space T is the set of all directions d that satisfy

$$d^\top \nabla h_j(x^*) = 0, \quad \forall j,$$

and

$$d^\top \nabla g_i(x^*) = 0, \qquad \forall i \in I(x^*),$$

Now, we have a similar set of conditions like what we had in unconstrained optimization.

So, if

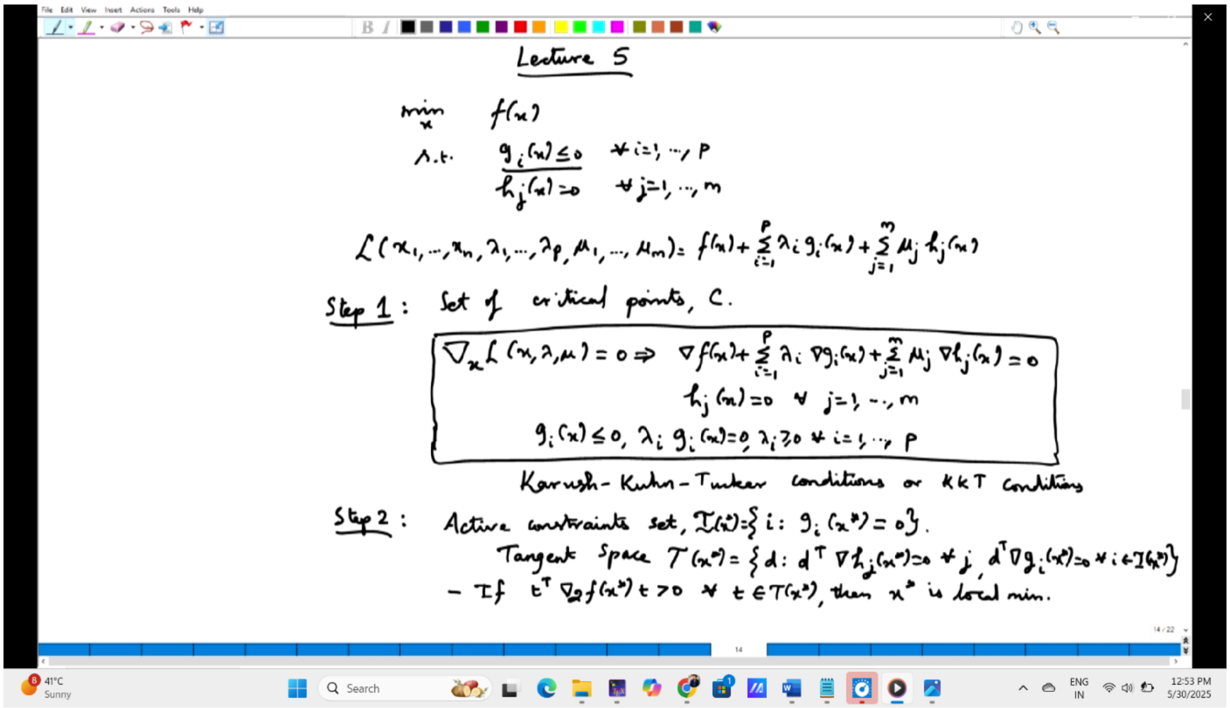
$$t^\top \nabla_2 f(x^*) t > 0, \qquad \forall t \in T(x^*),$$

then x^* is a local minimum, and the constraints will be similar.

So, I will just write them down anyway:

$$t^\top \nabla_2 f(x^*) t > 0, \qquad \forall t \in T(x^*).$$

(Refer Slide time 11:06)



If

$$t^{\top}\nabla_2 f(x^*)t < 0,$$

then x^* is a local maximum.
If there exists t_1 such that

$$t_1^{\top}\nabla_2 f(x^*)t_1 > 0,$$

and there exists t_2 such that

$$t_2^{\top}\nabla_2 f(x^*)t_2 < 0,$$

then x^* is a saddle point.

Finally, if there exists $t \in T(x^*)$ such that

$$t^{\top}\nabla_2 f(x^*)t = 0,$$

then further probe is needed.
Second-order conditions in constrained optimization is quite difficult to verify analytically in practice.
But the only thing that you need to note is that the difference between the conditions that you saw here — what you had for unconstrained optimization and the ones that

you are having for constrained optimization — is that here you are actually considering every non-zero t, or sometimes the zero as well, if required.
But here you are considering t only those vectors t that are in the tangent space $T(x^*)$.
You do not need to consider the other directions.
So, if that happens for this restricted set of vectors, if you have the condition

$$t^\top \nabla_2 f(x^*) t > 0,$$

then x^* is a local minimum.
It is less than zero, then it is a local maximum.
If you have $t_1 \in T(x^*)$ such that the quantity is greater than zero and $t_2 \in T(x^*)$ such that the quantity is less than zero, then it is a saddle point.
And if you have $t \in T(x^*)$ such that

$$t^\top \nabla_2 f(x^*) t = 0,$$

then further probe will be needed.
These are harder constraints to verify, but nevertheless, I am giving this for the completeness of analysis.
Given that we have characterized the solutions, we will now look at an example.
So, let us consider the same example that we had given. We are minimizing, so we just wrote a general utility function.
We will consider minimizing $-x_1x_2$

subject to the condition

$$p_1x_1 + p_2x_2 - w \leq 0, \qquad -x_1 \leq 0, \qquad -x_2 \leq 0.$$

If you actually plot this to see what this constraint set actually means, you will get the following set.
This line is:

$$p_1x_1 + p_2x_2 - w = 0,$$

Since we want this to be less than or equal to zero, we are considering this region.
And so this is $x_2 = 0$, and this is $x_1 = 0$.
Since you want $x_2 \geq 0$ or $-x_1 \leq 0$, you consider the intersection of all these three sets, you will actually get this particular set.
So, x_1 and x_2 can take values only in this particular set, the shaded region in the figure here, and you are expected to minimize $-x_1x_2$ or maximize x_1x_2.

Let us find out with the theory whatever we have developed until now.
There is a minor modification in whatever I have written down.

We consider minimizing -x_1*x_2, subject to the constraints:

$$p_1 * x_1 + p_2 * x_2 - w \leq 0$$
$$-x_1 \leq 0$$
$$-x_2 \leq 0$$

This defines a feasible region bounded by the lines $p_1 x_1 + p_2 x_2 = w$, $x_1 = 0$, and $x_2 = 0$. The optimization is to minimize -$x_1 x_2$ or equivalently maximize $x_1 x_2$ in this feasible set.

Constructing the Lagrangian:

$$L(x_1, x_2, \lambda_1, \lambda_2, \lambda_3) = -x_1 x_2 + \lambda_1 (p_1 x_1 + p_2 x_2 - w) - \lambda_2 x_1 - \lambda_3 x_2$$

KKT Conditions:

1. Stationarity:

$$\partial L/\partial x_1 = -x_2 + \lambda_1 * p_1 - \lambda_2 = 0$$
$$\partial L/\partial x_2 = -x_1 + \lambda_1 * p_2 - \lambda_3 = 0$$

2. Primal feasibility:

$$p_1 x_1 + p_2 x_2 - w \leq 0$$
$$-x_1 \leq 0$$
$$-x_2 \leq 0$$

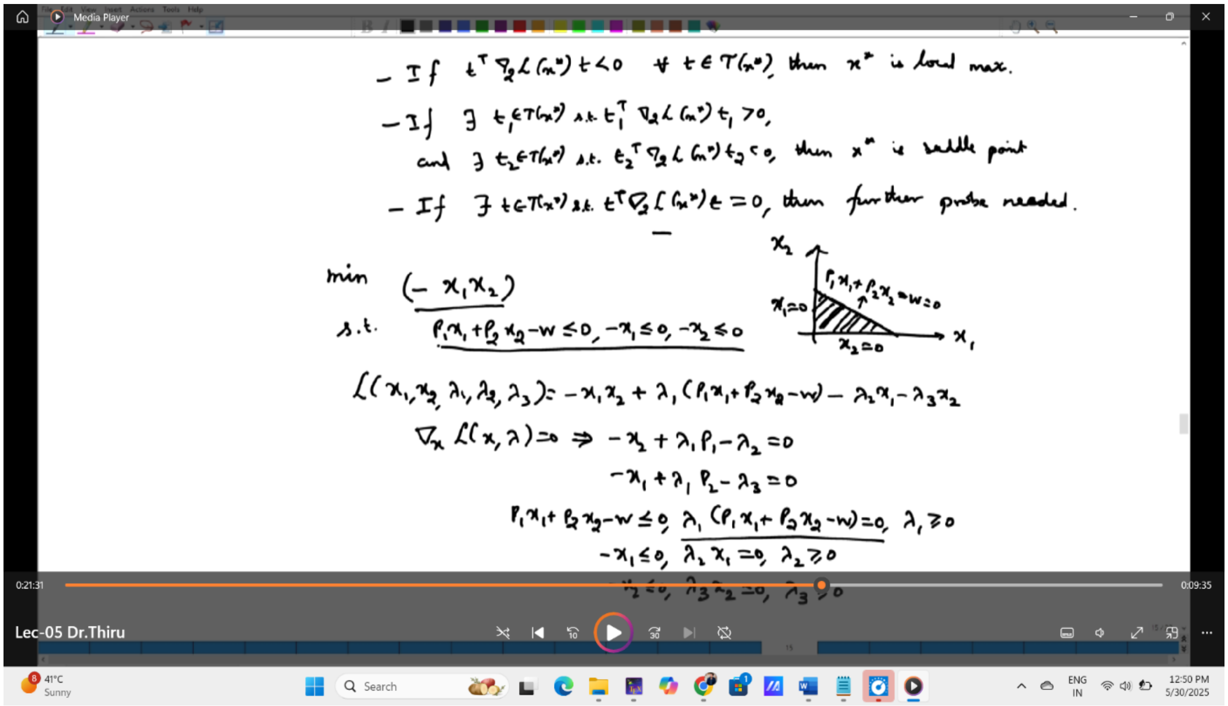
3. Dual feasibility:

$$\lambda_1 \geq 0, \lambda_2 \geq 0, \lambda_3 \geq 0$$

4. Complementary slackness:

$$\lambda 1(p_1 x_1 + p_2 x_2 - w) = 0$$
$$\lambda 2 x_1 = 0$$
$$\lambda 3 x_2 = 0$$

(Refer slide time 21:35)



Assuming $\lambda^*{}_2 = \lambda^*{}_3 = 0$, we get:

$$x_2^* = \lambda^*{}_1\, p_1$$
$$x_1^* = \lambda^*{}_1\, p_2$$

Substituting into the constraint:

$$p_1 x_1^* + p_2 x_2^* = \text{w}$$
$$\Rightarrow \lambda 1^* p_1 p_2 + \lambda 1^* p_1 p_2 = \text{w}$$
$$\Rightarrow \lambda^*{}_1 = \text{w} / (2 p_1 p_2)$$

Therefore:

$$x_1^* = \text{w} / (2p_1)$$
$$x_2^* * = \text{w} / (2p_2)$$
$$\lambda^*{}_1 = \text{w} / (2 p_1 p_2)$$
$$\lambda^*{}_2 = 0$$
$$\lambda^*{}_3 = 0$$

This is the unique critical point. Now, verifying second-order conditions:

Hessian of Lagrangian:

$$\nabla_2 \text{L}(x^* \;\; \lambda^*) = [\ 0 \ \ \text{-}1\ ]$$
$$[\ \text{-}1 \ \ 0\ ]$$

Tangent space defined by:

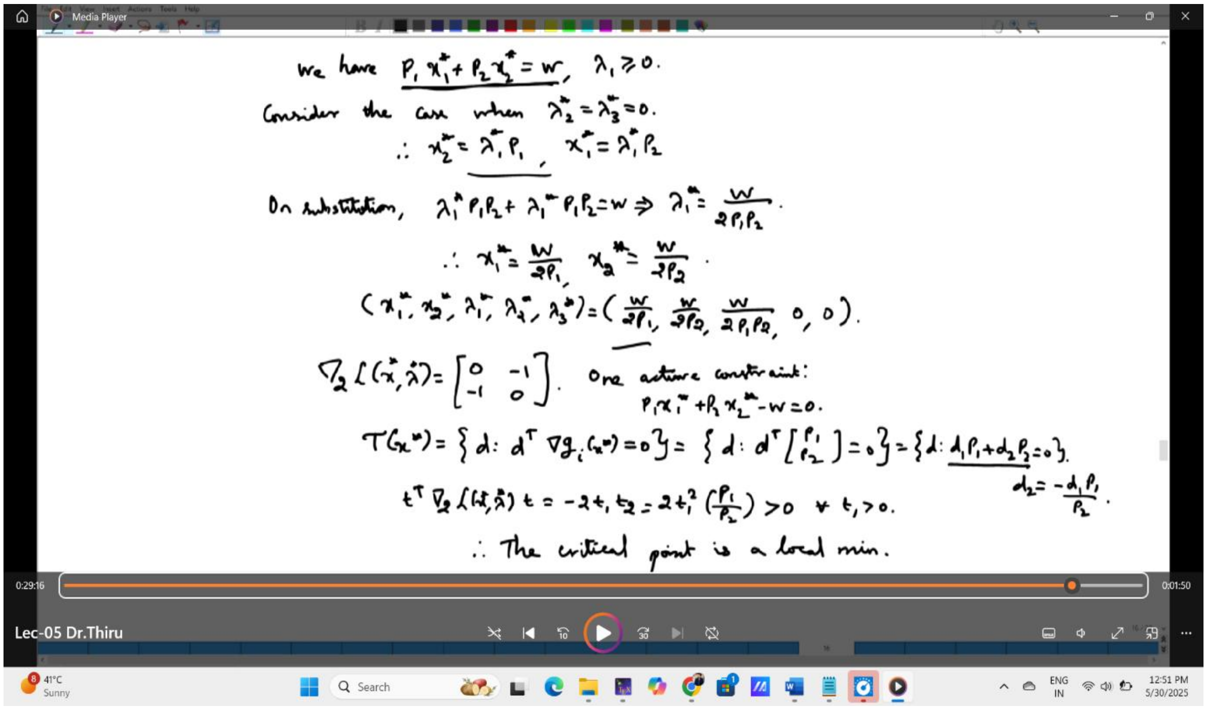
$$d_1 p_1 + d_2 p_2 = 0$$
$$\Rightarrow d_2 = \text{-}d_1 p_1 / p_2$$

Then:

$$t^T \nabla_2 L(x^*, \lambda^*) = -2t_1 t_2 = 2t_1^2 (p_1/p_2) > 0 \text{ for } t_1 \neq 0$$

So, the critical point is a local minimum.

(Refer slide time 31:05)



This concludes the primer on constrained optimization. Week 1 ends here. Next week, we will begin learning the basics of Python and understand the necessity of computational tools in optimization.