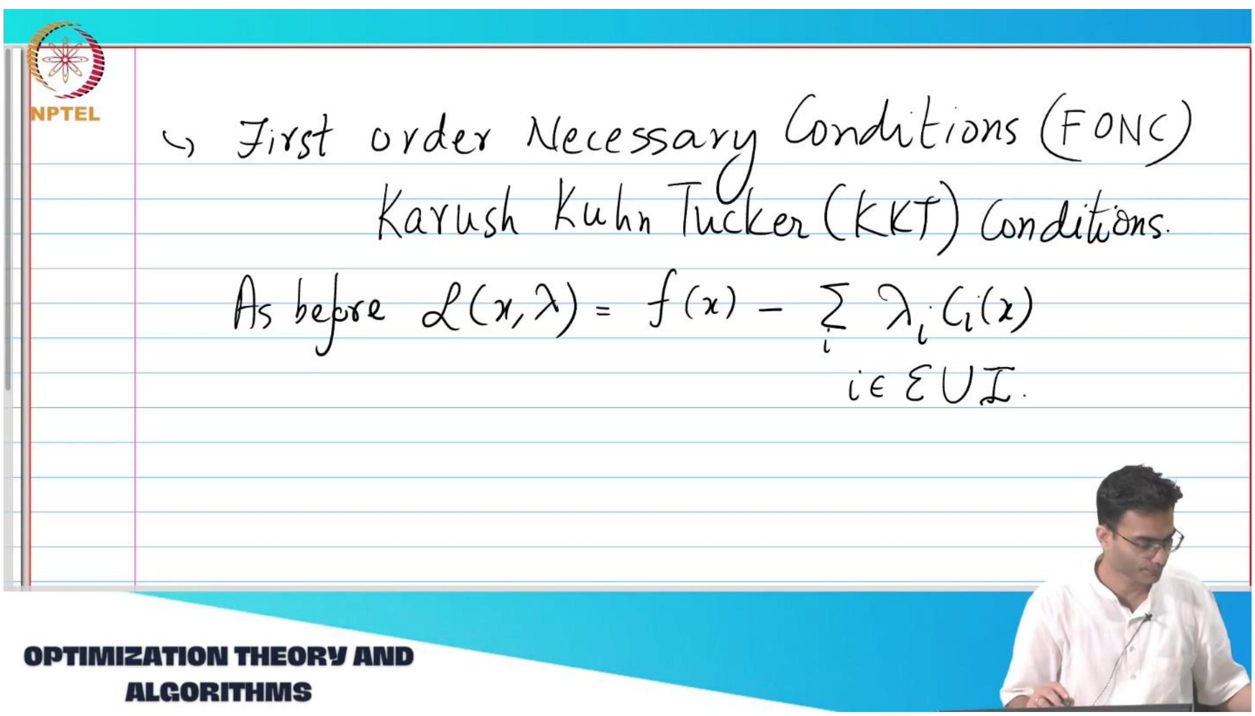
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 10
Lecture - 68

KKT conditions (First order necessary conditions)

So, that is where we will kind of bring this discussion between algebra and geometry to an end. And what we will do next is we will formalize our intuitive understanding of constraint optimization in the first theorem of constraint optimization. We are just going to put together all our intuition into something more formal. It is like the bedrock of constraint optimization. Many of you would have heard the word also for this theorem or the name for this theorem; they are called KKT conditions, Karush–Kuhn–Tucker conditions, right? So, I am sure you have come across this name in many, many papers. So, let us take a dig at this, ok?

So, they are called the first order necessary conditions, and because this is a particularly long phrase to write, we will just use the short form FONC, also called as after the people who discovered it. So, in fact, in the literature, you will find most people just call them KKT conditions; almost no one will say FONC. Ok, to start our discussion of constraint optimization, the tool that we have discovered is the Lagrangian, right. So, the Lagrangian is going to be as before.

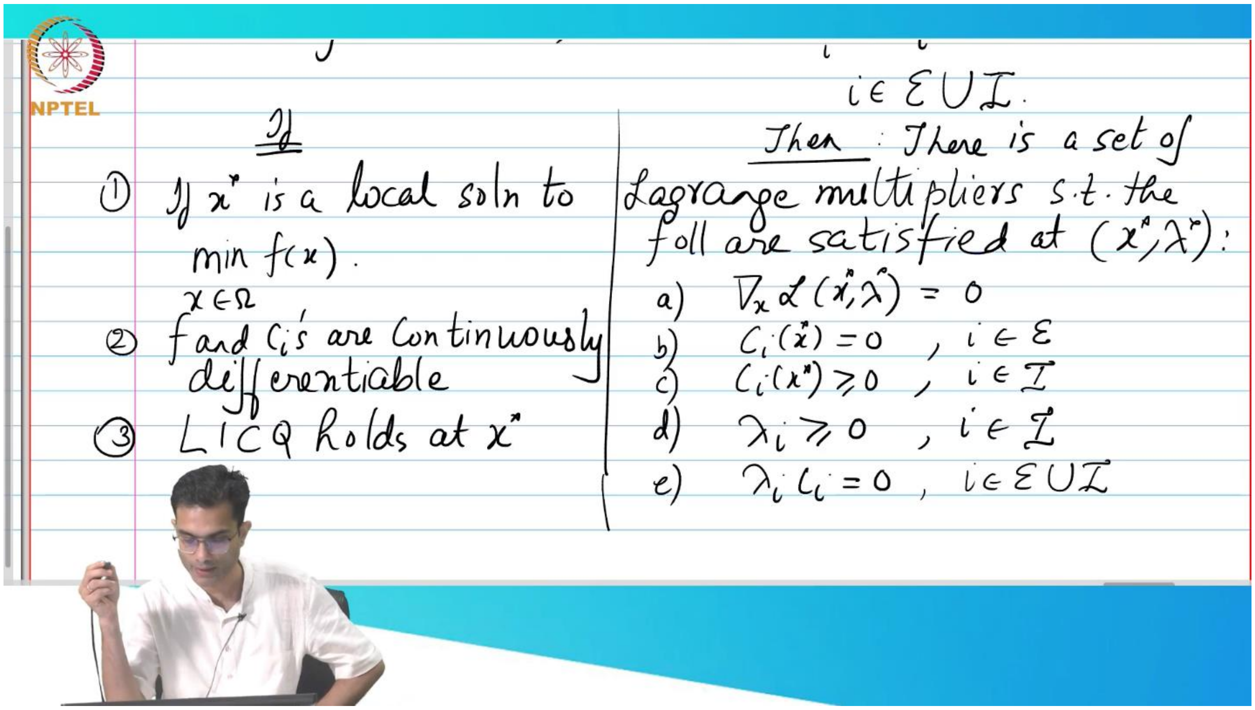


The Lagrangian is a function of x and λ, simply written as:

$$L(x,\lambda) = f(x) - \sum_i \lambda_i C_i(x)$$

where the sum includes both equalities and inequalities. Ok. So, it is a relatively long sequence of statements that we have to write to put this KKT theorem on paper. Ok, so let us do that with this Lagrangian in mind. So, the theorem has a set of conditions: "if" a bunch of things are true, then these things follow; we will write it as "if" and "then." So, what are the "if" parts?

This is the first-order necessary condition for what? No, for constraint optimization, but for what? To check whether or not x^* is a stationary point. That is what this is a test for; it is not an algorithm. This is a test that if you give me a point x, I will subject it to this test and tell you whether or not it satisfies the conditions. So, that is the first part. If x^* is a local solution to:



$$\min_x f(x) \quad \text{subject to} \quad C_i(x) = 0, \quad i \in \text{equality set}, \quad C_i(x) \leq 0, \quad i \in \text{inequality set},$$

what are the reasonable conditions to apply? Since we want to do calculus, what should we assume about f and C_i? Differentiability is to be expected, right. So, f and C_i should be continuously differentiable. Ok, what is the third condition you think? It is something very related to what we just did. Louder, LICQ, right? LICQ holds at x^*. Ok, so if you give me all of these three conditions, then comes the "then" part of the theory. Ok, then there is a set of Lagrange multipliers λ^* such that the following are satisfied at the optimal point.

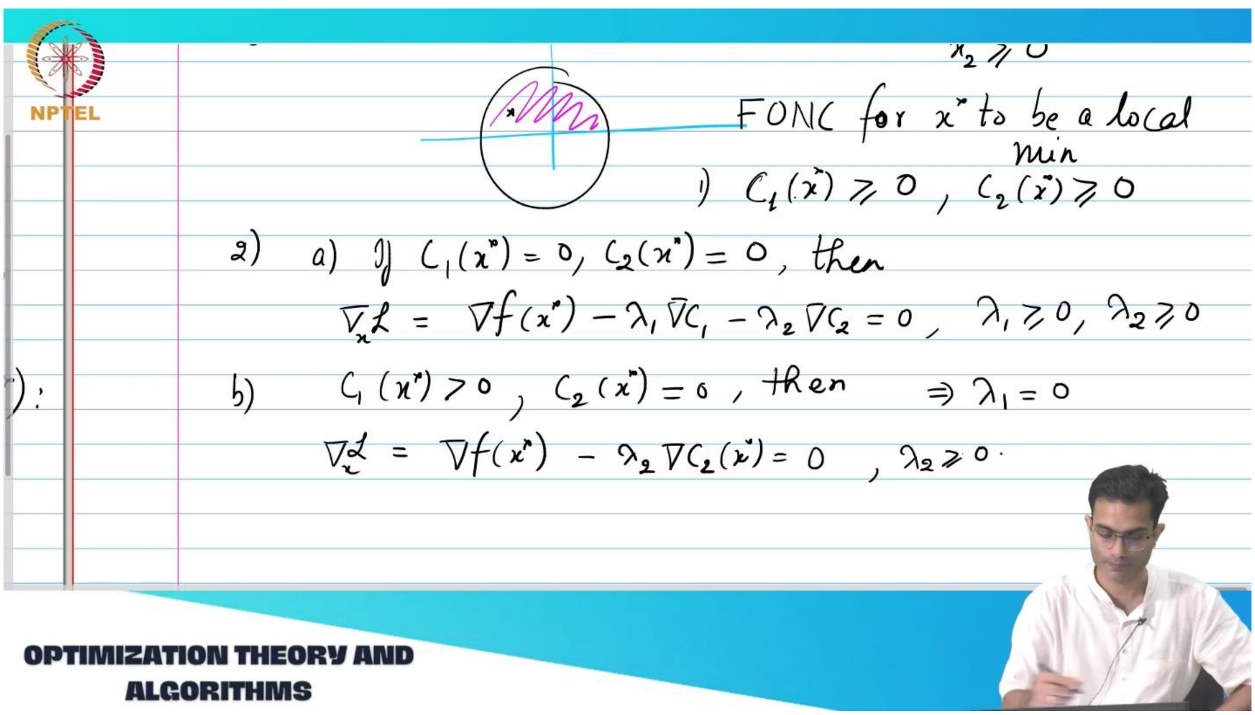
So, what are these things that are satisfied? This is what we have already studied, we are just going to formalize this. What is the first condition in constraint optimization? Gradient of what should be 0? The Lagrangian. That is the first condition: the gradient of the Lagrangian with respect to x at the optimal point should be 0. Ok, what is the other thing? The other two steps are very, very obvious: we should be on the feasible set, which means equalities should be satisfied. So, $C_i(x^*) = 0$, x^* should satisfy for all i in the equality set. What about inequalities? Right. So, $C_i(x^*) \geq 0$, i belonging to the inequality set.

Now, what should be true about the Lagrange multipliers λ_i? Look, it is just an analog of the equality case. What would it be? Inequalities, right. So, $\lambda_i \geq 0$, for i belonging to the inequality set. Ok, and finally, comes our complementarity condition: $\lambda_i C_i(x^*) = 0$ for all i. So, you have these conditions (a), (b), (c), (d), (e) which will hold if these 1, 2, and 3 are satisfied. Ok.

There is one final piece to this, which talks about the uniqueness of λ^*. Remember, λ^* also has to have an optimal choice, right? So, there can be many points, let us just note that over here, there can be many points where (a) through (e) hold, but if you give me LICQ, which is condition 3 over here, then the λ^* is also unique. So, but if 3 holds, λ^* is unique.

So, these are, in very, very brief, your famous KKT conditions. To satisfy them, there is not a very stringent condition. What do you have to make sure? The objective function and constraints are differentiable, continuously differentiable; you need to check LICQ. Right. To check LICQ, what do I have to do? Do I need to do feasible sequences? No. What do I have to do? Look at all the active gradients, put them together in a set, figure out whether or not they are linearly independent. I can check that.

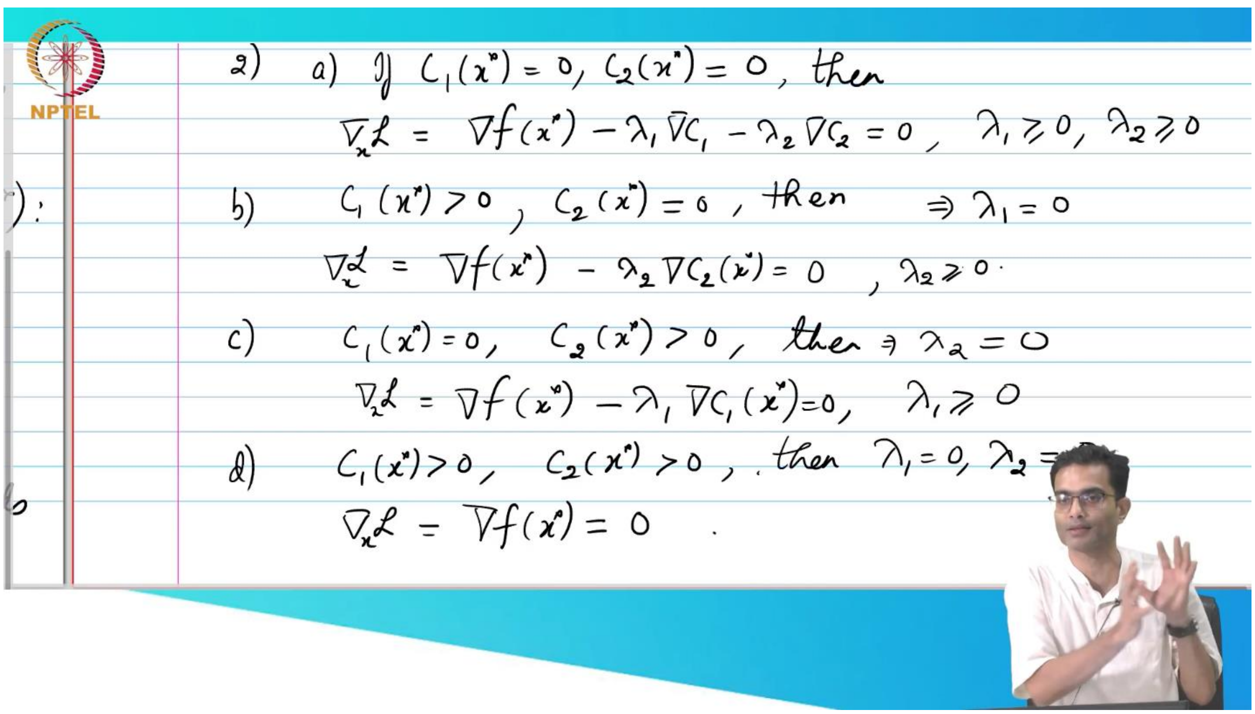
So, if all of these three criteria are satisfied, right, then this theorem applies, and only then this theorem applies, then there will be a set of Lagrange multipliers λ; they may not be easy to calculate, that is a different matter, but all of this will hold true: gradient of the Lagrangian is 0, right? You are on the feasible set (point B and C), and the complementarity condition holds. So, there is... I have not yet written a proof over here. In the examples that we did (equality constraint, inequality constraint), also I did not give a proof, right? So, we cannot run away from giving a proof for too long.



So, we will give you a proof of this by means of an example. Ok, so let us see how much time we have. So, let me start out with the statement, and we will not do the full proof, but we will start out with the statement and we will take an example. Our famous, favorite example is what we are going to take, right? So, this is not a full proof, let us call this a proof sketch, and for that, our example was $f(x) = x_1 + x_2$ and the constraint was that I am going to take two constraints: the interior of the circle.

Right, so we did this example once before. This was two inequality constraints. This is that example, right? So, right, and here what is the feasible set? The upper disk, right? The upper disk is my feasible set. So, what we want to check is if I give you some candidate point. So, first-order necessary conditions for x^* to be a local minimum, this is what we want to check. Ok, so we will apply it to this, and that will help us to formulate the proof. So, $C_1(x)$, $C_2(x)$, obviously, at this point x^*, the inequalities must be respected.

The second part is, see here, that the trick or the difficulty with inequalities is that they may be active or inactive, and I have two of them. So, how many possibilities are there? Actually, four possibilities. They could, you know, right? Depending on greater than or equal to, there are four possibilities. So, we will have to check for all of these four possibilities and then see depending on the point, right? So, let us note those down.



Ok, so if $C_1(x^*) = 0$, I have not specified it, I am leaving it general. If both of these guys are 0, right, then what all should happen? So, the constraints are active, right? What does that say about the Lagrange multipliers? They have to be greater than or equal to 0 because it is an inequality, right? So, $C_1(x) = 0$, $C_2(x) = 0$. Does that say $\nabla C_1 = 0$? No, obviously not. So, the gradient of the Lagrangian will give me:

$$\nabla f(x^*) - \lambda_1 \nabla C_1 - \lambda_2 \nabla C_2 = 0$$

This should happen, and $\lambda_1 \geq 0$, $\lambda_2 \geq 0$. This is the first possibility, and I have to verify whether or not this happens. Yeah, I mean you can, I am not assuming it. Look at the f and C; are they continuously differentiable? Yes.

And let us say I am taking my candidate point x^* and I am going to investigate. Now, without specifying which point I am going to take, I am listing out all the investigations I have to do. x^* being a candidate point for local minimization. For, yeah, exactly, LICQ. We will verify, we will verify.

Actually, with the constraints that you have taken, you can see LICQ will be satisfied because there are no funny whole square terms happening there. So, LICQ will also be satisfied. So, the "then" parts of the theorem give me this, right?

What if I had case B where $C_1(x^*)$ is strictly greater than 0, $C_2(x^*) = 0$? Then what should I have to verify? If $C_2(x^*) = 0$, gradient of the Lagrangian has to be 0 anyway, right? So, but if $C_1(x^*) > 0$, what does that say about λ_1? What should happen? λ_1 should be 0.

So, the gradient of the Lagrangian will give me:

$$\nabla f(x^*) - \lambda_1 \nabla C_2(x^*) = 0$$

and is there any condition on λ_2? $\lambda_2 \geq 0$. Ok, so this is, I am giving you practice in inferring what does this mean, how to use this theorem, and what are the various conditions that this theorem tells me to check for. Ok, so that is why I am going step by step over here.

Condition C is the symmetric one: $C_1(x^*) = 0$, $C_2(x^*) \geq 0$. Then what happens? Which of the lambdas will go to 0? $\lambda_2 = 0$ because of complementarity, right?

So, $\lambda_2 = 0$, gradient of the Lagrangian has to be 0. So, this will simply be:

$$\nabla f(x^*) - \lambda_1 \nabla C_1(x^*) = 0$$

and $\lambda_1 \geq 0$. This is what the theorem statement is.

And last, but not the least, what would it be? Both the inequalities are inactive. So, I have $C_1(x^*) > 0$, $C_2(x^*) > 0$. Ok, what happens to my Lagrange multipliers? Both have to be 0, $\lambda_1 = 0$, $\lambda_2 = 0$. The gradient of the Lagrangian will simply become:

$$\nabla f(x^*) = 0$$

Anything else? No. Right, so that is it. So, I am not assuming it. From here on, I have to prove these statements, right?

I have merely taken the theorem, applied it to this problem, and figured out what are all the possibilities that will happen. If my x^* happens to be on the boundary of one constraint, not on the boundary of one constraint, not on the boundary of any constraint, whatever these four possible combinations are, this is what the theorem tells me. So, I have to go in and painstakingly prove each one of these to be true and prove this to be true in a way that is not dependent on this problem that I have taken, but in general.

So, that is what I am going to do in the next session. So, this is just to tell you what to expect.

You would have a λ_3 if you had a third constraint, equality or inequality, you would have a λ_3 as well too. You said it is an equality or inequality? Equality. So, it will always be equal to 0. I mean it is always $C_3(x) = 0$.

So, you would have to maintain that. λ_3 will be greater than or equal to 0. λ_3 need not be greater than or equal to 0. It can be anything. It can be anything because the greater than or equal to 0 condition is for inequality.

For equality, it can be anything. For, I mean if it is inactive, then it has to be 0, yeah. But why do we put that positivity condition for one lambda for an active inequality? We got that intuition from the example that we took, right? We saw that in the case where I had either parallel or anti-parallel configuration, in one case I had no feasible direction, and in the other case, I had a feasible direction. So, I was interested only in the one where there was no further direction to go to, and for that, I needed parallel, and parallel made it such that λ was positive.

That is because if you define the inequality as $2 - x_1^2 - x_2^2$, correct? Now, if it was the other way around, like the exterior of the circle... Exterior, very good. So, maybe that is what we can end this discussion with.

So, we have been talking about, for example, a single inequality constraint as the interior of the circle, $f(x) = x_1 + x_2$, and I want the minima of this. Now, he is flipping the inequality, he is saying $x_1^2 + x_2^2 - 2 \geq 0$. So, before, this is really important before you begin to solve the problem, you should ask yourself: is this a sensible problem to solve? Is it? It is not because this problem's objective function now is not bounded from below. The minimum you can achieve is $-\infty$.

Right, so do not even attempt to solve it, right? It is like saying, "What is the minimum of $\log(x)$?" Right, it is going to blow up. So, check that before you start solving a problem. After doing all of the, you know, hard work, you realize, "Oops!" I need not have bothered. Ok, intuitively, okay.

So, how did I get the tangent cone to be... Yeah, yeah, this one.

So, when I took the sequence z_k, I got a vector that is pointing in the +1 direction, I mean along the y-axis. When I took the sequence w_k, the vector is pointing in the -y direction. The directions are +y or -y. So, I could have, first in the first cut, written it as $(0, \pm 1)$, but then you notice that this t_k is in my hand. It only needs to satisfy the property that for large values of k, it tends to 0. So, with t_k—if t_k is defined like this, the norm of the vector—then I get ± 1. Supposing t_k was defined as 2 times the norm of the vector, what is the vector that I get? The tangent vector is $\left(\frac{1}{2}, \pm\frac{1}{2}\right)$.

So, it is entirely in my hand. So, like this, you can keep choosing different multipliers over here, different t_ks, and you will be able to get the entire... So, that is why I am writing as $(0, d_2)$, where d_2 belongs to real numbers. Yeah, you can take, yeah. So, a sequence is not uniquely defined; there are infinite possible sequences. So, this is left to you. That is why I kind of cheated by taking an easy case, right? If I choose $C_1(x) = 0$, I am on the boundary of the circle. It is relatively easy for me to define this feasible sequence. Now, if you say that I live inside the circle, then I have a lot more freedom and creativity to define any kind of crazy sequence that gets me...

So, it is—you have to define it. The trouble is yours, but it can be done. Correct. So, you want the proof?

So, the intuition part is clear to you. Yeah. So, that... so, as far as intuition goes, it is clear to you why it only holds true for the active inequalities. For equality, there was no reason or need to constrain in that way, right? That, you got from geometry, that you are coming. The proof part is what we will do in the next session and why it is forced to be positive or... In a way that is a good way of looking at it.

Whether I write the constraint as $C_1(x) = 0$ or $-C_1(x) = 0$, does not matter. So, by flipping C_1 to $-C_1$, ∇C_1 also flips a sign.

So, it does not really matter. It matters because you change the half-space from here to there.