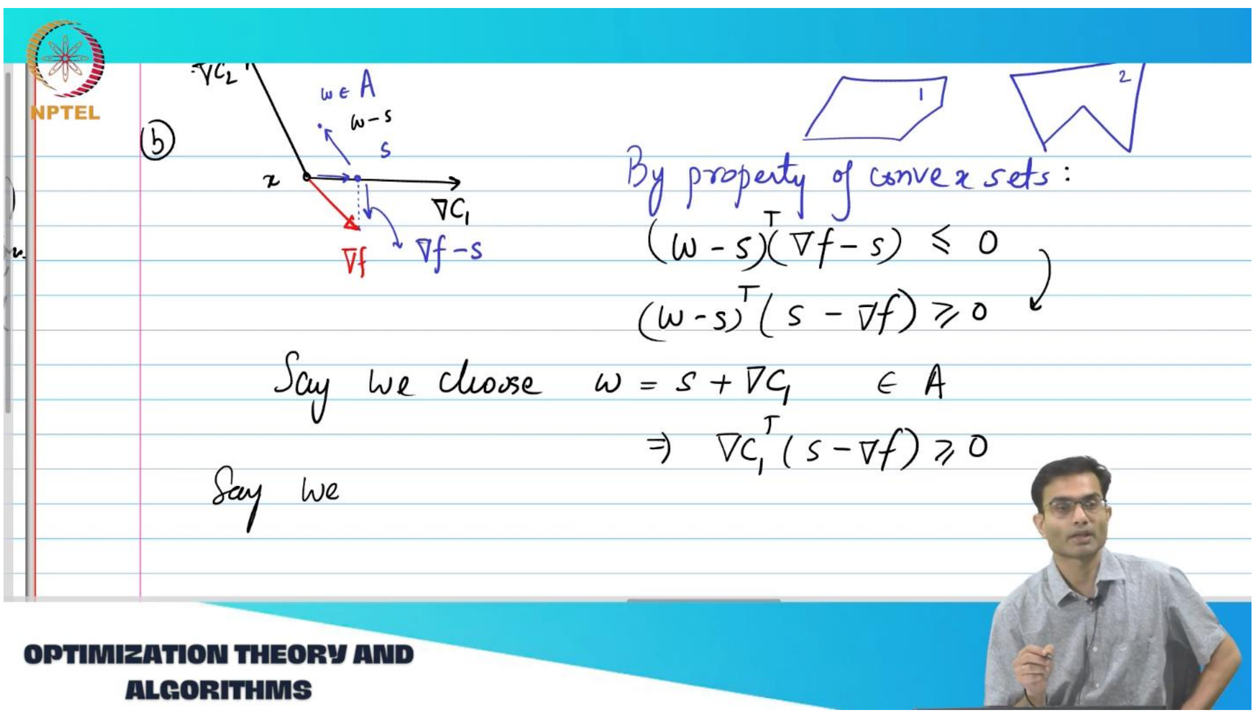
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 10
Lecture - 70

Proof sketch for KKT conditions - Part 2

Now, let us go to the next part. The next part will be to show feasibility. Okay, let us look at feasibility. I am going to draw this once again. ∇C_1 this is my x, okay, this is ∇f. Alright, let's take some point over here, $w \in A$. So, I am going to use a second intuitive property of convex sets without proof. So, what is that? If I... So, S is on the boundary of this set because it is arrived at by projection, okay?

So, if I take... let us use blue. If I take this point S and join it to a point in the interior, so for example, w is a point in the interior, I take this, okay, and I connect S to any point outside the convex set here, for example, okay? What is this vector that I have just drawn over here in the bottom vector? So, it is not $S - \nabla f$, but $\nabla f - S$, right?



So, this guy is $\nabla f - S$, okay. So, the question is, will this angle always be acute, 90 degrees, or obtuse? It is a convex set—draw some things in your mind. So, it is a convex set. That means, for example, this is a convex set, right? On the other hand, this is not a convex set, right? So, now, given that we have a convex set, this angle between a boundary point and an interior and boundary point at any point outside has to always be obtuse, right? You can see that in the second example over here. Here I can imagine that some angle can be acute, but if I am always like this inside a convex set, okay? I am not proving this formally, but you can arrive at it intuitively. This angle will always be obtuse, okay?

So, what are these two vectors I have? Okay, so this vector that I have drawn on the top is $w - S$. What is the quick way to remember which way the vector points? The destination is the first guy, the source from where you started has the minus sign. So, I will get the $w - S$ in a product. If I take it with $\nabla f - S$, what should this be? Less than or equal to 0. So, once again, as I said, this is without proof, but geometrically you can get a good idea about it. So, we have just written this down. w is up to us. We can choose w in any way that is convenient for us, agreed? So, let us just flip this around because the entire claim that we made was about $S - \nabla f$. I had $\nabla f - S$. So, if I multiply by a minus sign, I have to flip the inequality. So, I will just do that. So, we get:

$$w - S \top (S - \nabla f) \geq 0.$$

As I mentioned, w is up to us. Choose w.

So, say we choose $w = S + \nabla C_1$. Will that still be in A? Clearly, right? ∇C_1. In fact, in this example, ∇C_1 and S are along the same line. So, I am going to be. So, therefore, w I have chosen a legitimate w. It belongs to the cone A. Okay. So, if I choose w in this way, what happens to my inequality? The first term becomes ∇C_1, right? That is all that is there. $w - S = \nabla C_1$.

So, this becomes:

$$\nabla C_1^{\top}(S - \nabla f) \geq 0.$$

Fine. We can continue this logic, supposing we choose... What should I choose? $S + \nabla C_2$. Is this again legal? Does it belong to A? Right, you can see S is along ∇C_1 and ∇C_2 is there. If I take their conic combination with coefficients 1 and 1, I will still be in the set A. So, this also belongs to A. So, that gives us:

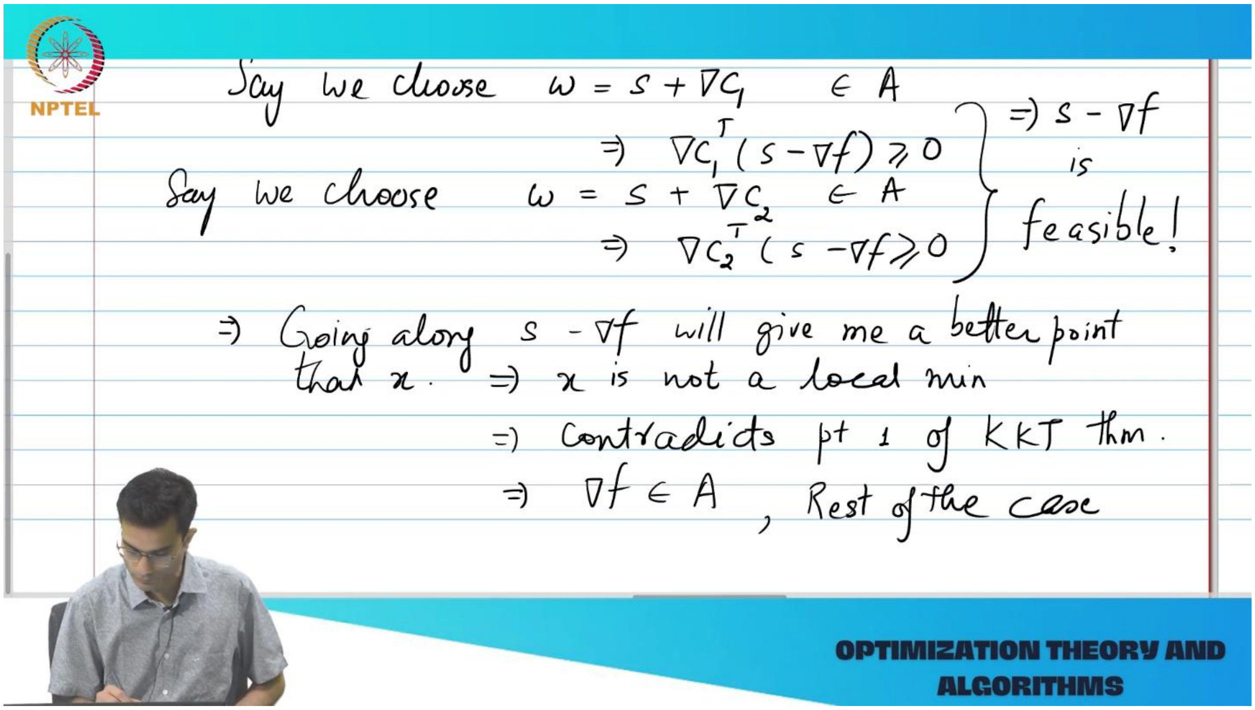
$$\nabla C_2^{\top}(S - \nabla f) \geq 0.$$

So, together, what does this imply? If you look back at what we had wanted, right, we had said we had made a claim of a feasible and descent direction. We already showed it is a descent direction, and here we are showing that $\nabla C_1^{\top} d \geq 0$ and $\nabla C_2^{\top} d \geq 0$. Therefore, $S - \nabla f$ is feasible. Which, again, I am just repeating myself, this implies that going along $S - \nabla f$ will give me a better point than x, implying that x is not a local minima, right? This contradicts point 1 of the KKT theorem. Okay? And therefore, our starting proof by contradiction follows through. Therefore, ∇f belongs to the set A. We have shown a contradiction.

So, you notice that this is the reason why this is called a proof sketch. We did not assume anything particular about ∇C_1 or ∇C_2, right? We did not say that, oh, this proof follows only because $\nabla C_1 = -2x_1 - 2x_2$ or blah blah blah. We just used it as a framework for us to visualize this, and I chose two constraints. Why? Because I can draw it on a plane. On the screen, if I had chosen three constraints or n constraints, it is very hard to visualize. One is too easy, two is the best, right? So, that is why this is a proof sketch, and we have shown it for the case when both the inequalities are active.

For the other cases, the proof will follow a similar kind of logic, except that steps are simpler. Okay, so I am not going to do the proof of b, c, and d. They are much simpler than this, but this is the logic that you would have to apply to arrive at it. Actually, for b and c, d is just like unconstrained optimization; we do not have to... the constraints are inactive, they are not playing any role in this thing. So, that part can be skipped. Okay? So, we will just write this. The rest of the cases can be done. So, I will recommend trying b or c as an exercise.

This will give you good practice in also making sure that you have understood this correctly. Any questions on what we did? Pretty straightforward, right? The only things you kind of had to take on faith are two slightly intuitive properties of convex sets. The first property was what? The length of the projection is less than or equal to the length of the original vector, and strictly less because I said ∇f does not live in A. That was the first assumption, and the second assumption was about the angle between the angle subtended between an interior point and an exterior point. This angle is obtuse, okay?



That also, by drawing a few sketches, you can convince yourself that that is true, okay? So, let me just make that note over here. Otherwise, later you will wonder what that is. So, convex, not convex, okay, and that gave us feasible and descent direction. So, to make sure that we really understand this, I am going to take a couple of examples. So, we will work through a couple of examples. The first one is very similar to what we have done; the second is a totally different example, and that will give us some practice, okay?

So, we are going to continue with the same problem: $f(x) = x_1 + x_2$, and

$$C_1(x) = 2 - x_1^2 - x_2^2 \geq 0, \quad C_2(x) = x_2 \geq 0,$$

same problem, okay? And we are going to use this KKT theorem as a tester, right? So, we will give it a few points, and we will see how to use the KKT theorem, okay? This is your sort of first experience in giving a theorem—how do we use it, okay? So, let us... I am going to take two points: point A, which is $(-1,0)$, and point B, which is $(-\sqrt{2}, 0)$.

These are the two points, and we will subject our points A and B to this test, okay? So, let us take point A. Okay, you can also write down ∇C_1, that is pretty straightforward, and $\nabla C_2 = (0,1)$. At point A, let us go step by step and list out what are the active and inactive constraints because that helps us figure out which Lagrange multipliers to worry about or not. So, at point A, which of the constraints are active?

So, if I do $C_1(x_A)$, what do I get? Substitute this coordinate inside, what do I get? $C_1(x_A) = 1$, which is strictly greater than 0. For $C_2(x_A)$, what is it equal to? 0, right? So, this is equal to 0; that means this is active, and this is inactive. Can I say something about λ_1, λ_2? λ_1 has to be 0 because of the complementarity conditions, right? So, my Lagrangian of x, λ is going to be $f - \lambda_2 C_2$, why? Because $\lambda_1 = 0$, okay?

Now, is this an optimal point or not? If this is... if x_A is optimal, this is point number 1 of the "if" condition of the KKT theorem, right? We have to say, if x_A is optimal, then everything else is going to follow. f and C are differentiable, all of that has happened. ∇C_1, you can see ∇C_1 and ∇C_2, if you substitute that, will they be linearly independent? Yes, why? What is $\nabla C_1 = (2,0)$ and $\nabla C_2 = (0,1)$? Are they linearly independent? Yes, LICQ satisfied.

So, 0.1, 0.2, and 0.3 are satisfied. Now, let us see the "then" part, right? If x is optimal, we should have $\nabla \mathcal{L} = 0$, right? So, then let us write that:

$$\nabla_x \mathcal{L} = 0.$$

So, ∇f must be equal to $\lambda_2 \nabla C_2$, this must hold true, okay? What is ∇f? f is here; what is ∇f? $(1,1)$, right? So, this is $(1,1)$ should be equal to $\lambda_2 \cdot (0,1)$, okay?

So, if x_A were to be an optimal point, we should find a set of Lagrange multipliers such that $\nabla \mathcal{L} = 0$ holds true, right? For this to hold true, is there any value of λ_1 and λ_2 that will satisfy this? No. Do what you want, $1 = 0$ can never be satisfied. So, this implies no value of λ_2 works, right? Since I cannot find any lambda, this is where my test works, right?

So, this says that x_A is not optimal. This is how you use your KKT theorem: you subject your point to the test, the test says $\nabla \mathcal{L} = 0$, complementarity conditions, blah, blah, blah, but for $\nabla \mathcal{L} = 0$, I am not able to come up with a λ_1, λ_2 that does it, therefore fail.

Let us look at point B. What is C_1 at B? What is $B = \left(-\sqrt{2}, 0\right)$? So, therefore, it is on the corner over there. So, $C_1 = 0$, and $C_2 = 0$. That means both are active. If both are active, can I say anything about λ_1, λ_2? They need not be 0, right? So, I would then have to have:

$$\nabla f = \lambda_1 \nabla C_1 + \lambda_2 \nabla C_2.$$

We have to find out whether there exist such λ_1 and λ_2 that solve or satisfy the situation. Let us see. So, I get $(1,1)$ from ∇f, why? Because $\nabla f = (1,1)$, since $f(x) = x_1 + x_2$.

Now, what is the value of ∇C_1 at x_B? $\nabla C_1(x_B) = \left(2\sqrt{2}, 0\right)$. So, for the Lagrange multiplier equation, we write:

$$\nabla f = \lambda_1 \nabla C_1 + \lambda_2 \nabla C_2.$$

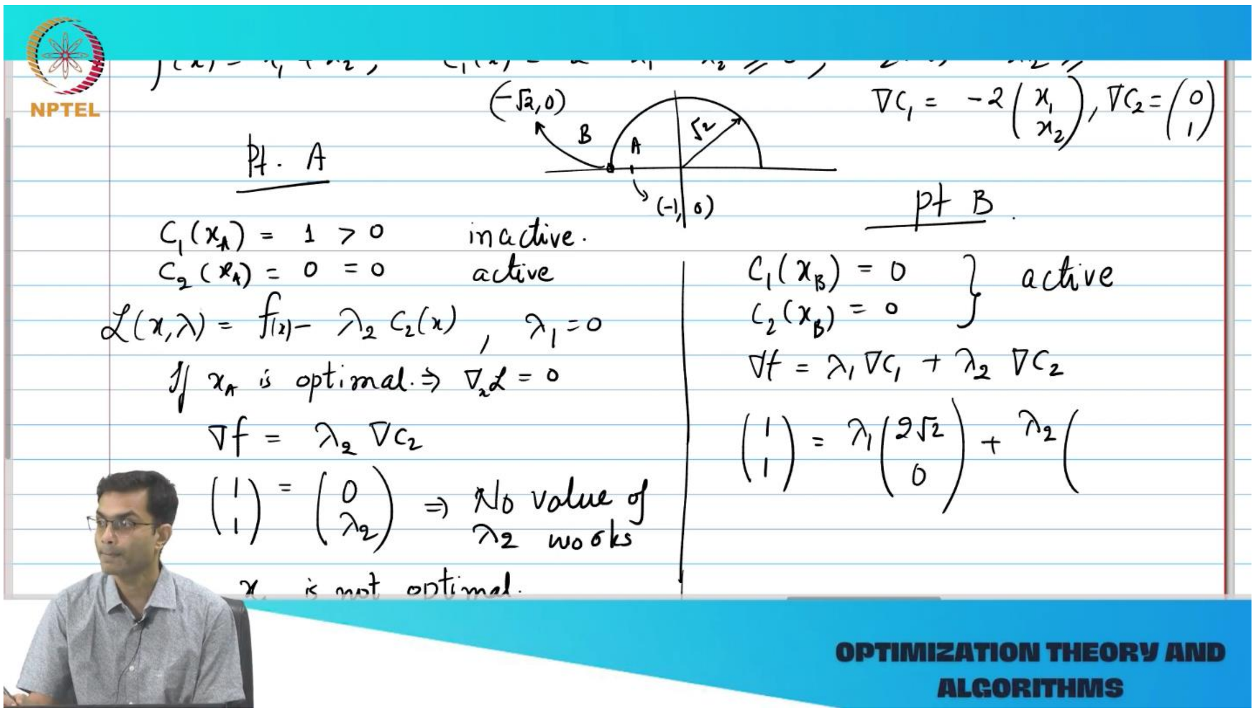
This gives:

$$(1,1) = \lambda_1\left(2\sqrt{2}, 0\right) + \lambda_2(0,1).$$

Does this give a value for λ_1 and λ_2? Yes, straight away:

$$\lambda_1 = \frac{1}{2\sqrt{2}}, \quad \lambda_2 = 1.$$

That's not enough. What further thing do I have to check? I found a set of λ_1 and λ_2 that satisfy $\nabla\mathcal{L} = 0$. But what were these constraints? Were they equalities or inequalities? Inequalities, right? Therefore, there is a further stipulation on the Lagrange multipliers: they should be non-negative.



Are they non-negative? Yes. So, I have to make sure that because if you reach this point and one of them is negative, you're finished. Therefore, the final note is:

$$\lambda_1 = \frac{1}{2\sqrt{2}}, \quad \lambda_2 = 1 \quad \text{are non-negative.}$$

Therefore, to first order, x_B is an optimal point that satisfies the KKT conditions. So, everyone has got the logic that we need to follow here, right? You verify points 2 and 3 of your "if" condition; they are satisfied. Then proceed to the implications of this. If it works, then you have found a point that satisfies the KKT conditions.

So, you can see that this is also quite easy to code up, right? You would have to have some way to solve for λ_1 and λ_2, and then check if there is a solution to this system of equations. In fact, it is a very simple system of equations to solve. Is it a linear system of equations? Yes, right? Because we can write it as:

$$\nabla C(x)^{\top}\lambda = \begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix}.$$

This whole thing can be written compactly as:

$$\nabla C(x)^{\top}\lambda = \begin{pmatrix}\lambda_1 \\ \lambda_2\end{pmatrix}.$$

You can solve for λ using any numerical technique and then check if the solution is valid, i.e., if it is positive, all of that. If everything is satisfied, we can say, "Okay, to first order, this point passes the test and satisfies the KKT conditions."

So, this is really one of the most foundational bedrock theorems of constrained optimization, the KKT conditions. I have mentioned this previously. You will come across this theorem in books sometimes with a sign flipped. The Lagrangian will be defined as $\nabla f + \lambda_1 C_1 + \lambda_2 C_2$, so all the signs will get flipped. Just make sure you know what the starting point is before you get confused by the subsequent results.

Is this clear? All right.