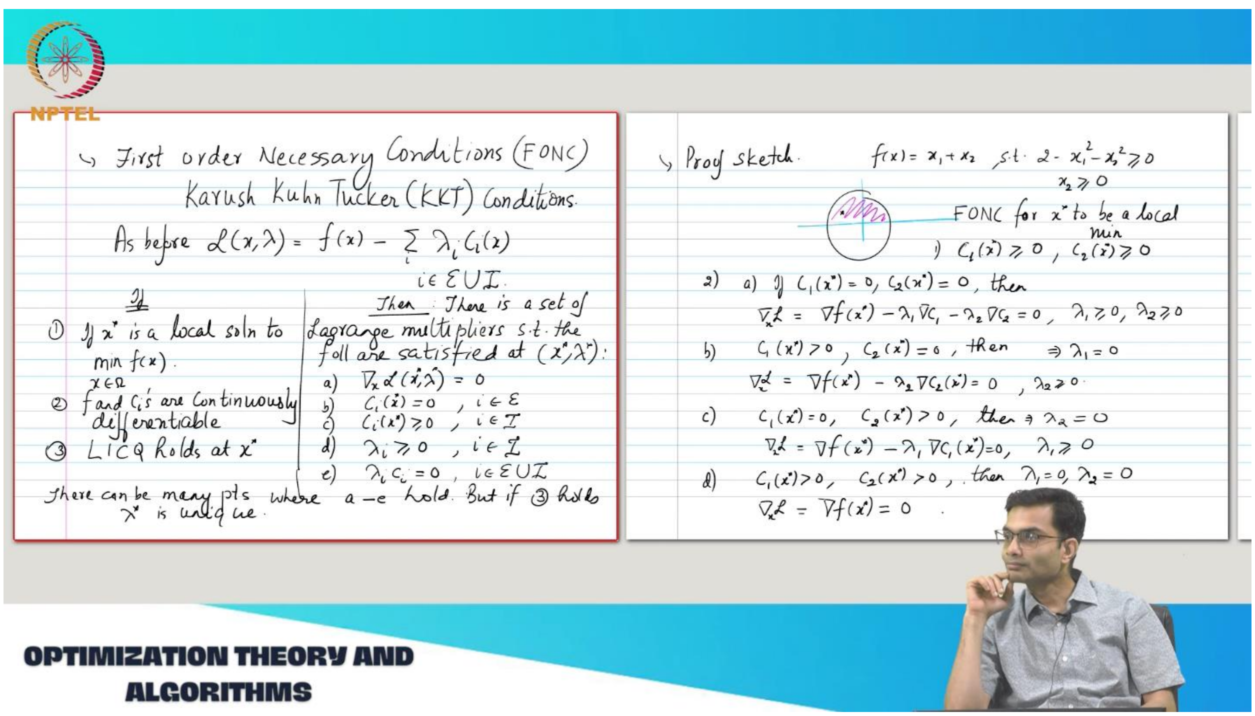
Course Name: Optimization Theory and Algorithms
Professor Name: Dr. Uday K. Khankhoje
Department Name: Electrical Engineering
Institute Name: Indian Institute of Technology Madras
Week - 10
Lecture - 69

Proof sketch for KKT conditions - Part 1

Tangent Cone and Dimensionality Discussion

Let us start with the one and only question on the doubt sheet. Can we say something about the dimensionality of the tangent cone? Everyone remembers what the tangent cone is? How do we arrive at it?

How do we start with this whole idea of the tangent cone? The first thing is, we start with the function f. Everyone forgotten after eating nice? How do we start with the tangent? Priyanka, how do we start with a tangent? What is the first object you have to start with? I start with a feasible sequence, right? A feasible sequence was defined as a limiting direction, meaning I am standing at a point in the feasible set, and I am watching a sequence approach me. I take that vector and divide it by a scalar sequence whose only property is that its length tends to 0 as I approach the point, right? And I take that limit, and I get a tangent—the limiting direction is the tangent. So, I got one tangent. From one tangent, how do I arrive at a tangent cone? I collect all possible feasible points and, therefore, collect all possible tangents, right? So, that is how I get my tangent cone. So, that is revision about the tangent cone.



The question is, can we say something about the dimensionality of the tangent cone? So, in the example we took, it was one-dimensional. Remember the tangent cone was a line parallel to the y-axis at the point $(-1,0)$ or $\left(-\sqrt{2}, \ldots\right)$. Is it always the case that the tangent cone will be one dimension less than the feasible set of dimensions? So, the set was two-dimensional because I drew the half-circle. So, the feasible set consists of points in two dimensions, and we found this tangent cone to be one-dimensional, which is a line. Is it always the case? Will it always be the case that the tangent cone will be one dimension less? Can it be even less than that?

Right. So, if you take just one or two constraints, we got what we got. But I can have multiple constraints in such a way that the feasible set is a point, right? That is also possible. It will be a very silly problem to solve because if there is only one point in the feasible set, there is nothing to optimize, that is the solution, right?

In general, the dimensionality is going to be less than the overall space in which we live, obviously, but we cannot say it will always be one less than the overall dimension.

KKT Conditions

Let us restart our discussion about the KKT conditions. So, remember, just to refresh your memory, what do the KKT conditions of the FONC tell us? It is a test—the test that you subject a candidate point to. If the candidate point passes that test, then we say that to first order, we believe this to be an optimal point. That is the meaning of this test. So, for doing this test, it is like when you show up for a blood test—you have to give blood, right? Without blood, there is no test. So, you have to provide a point to the test to say evaluate it for this guy—that is how it works.

Now, we said that instead of proving this in full generality, we will prove this theorem and give a sketch of it. Everyone remembers now, right? Let's zoom in a little bit. These were the if-then conditions for the theorem. If what? Obviously, I am considering a point x^* to be a local solution to this problem. f and C are continuously differentiable, and most importantly, LICQ holds. These three points have to be satisfied for us to start the test.

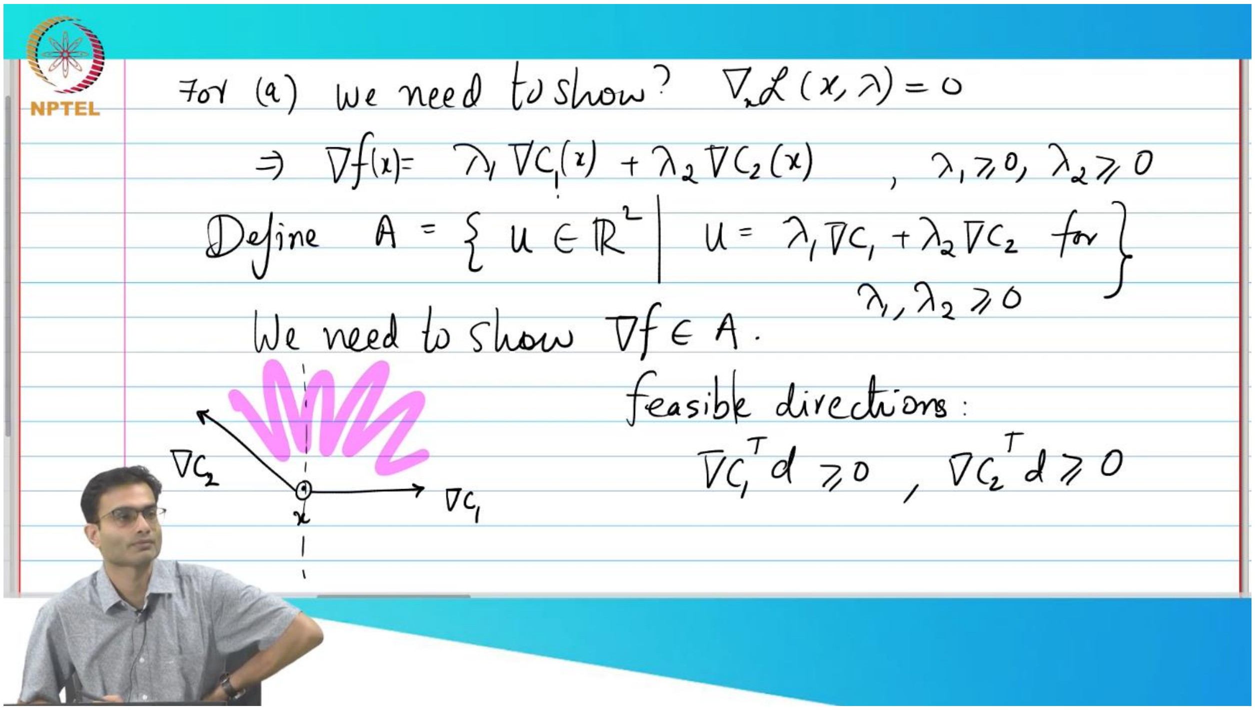
If some of these are not true, there is no sense in applying the "then" part of this theorem—it is not applicable. And this is again a classic mistake: you do not verify, for example, that LICQ holds, then you solve your KKT implications, get some λ, and end up with a contradiction because you did not verify the first part.

So, the first three conditions have to be checked. Then I can apply the "then" part. Then several things follow—these are all kind of formalizations of the intuition which you already had. The gradient of the Lagrangian should be 0, etc., and then we get the complementarity condition.

We said, okay, let us look at a toy problem. The toy problem had an objective function which is linear, along with two inequality constraints. The feasible set was the pink shaded region, and then we said, let us take some point in the feasible set. What all will be the implications of this theorem?

Can we say that f and C are continuously differentiable? Yes, right. What is the second point that I have to verify? That is the third point... second? I thought that was the third. Okay. For x^*

to be a local minimum, and the third is LICQ. Do you think LICQ will hold? It will hold because:



$$\nabla C_1 = (-2x_1, -2x_2), \quad \nabla C_2 = (0,1)$$

So, for most points, these should be linearly independent. We can verify it in particular for specific points.

For any point x^* inside this feasible region, these are the tests that we have to subject this point to. The first one is very trivial—obviously, it should be feasible. So, I say $C_1(x)$ and $C_2(x)$ both should be greater than or equal to 0. If that is not true, there is no conversation. Then we come to the various possibilities. The possibilities are simply whether the equalities are strict or whether the inequalities are active or not active. That is how we will separate these two things. Why? Just so that we understand clearly how to approach the proof.

We will split it into four possibilities: either both are 0, one of them is 0, or neither are 0. That is how we are going to split this. It turns out that if you look at case (2a), that is the case where the Lagrangian has the most complicated expression. Will you agree?

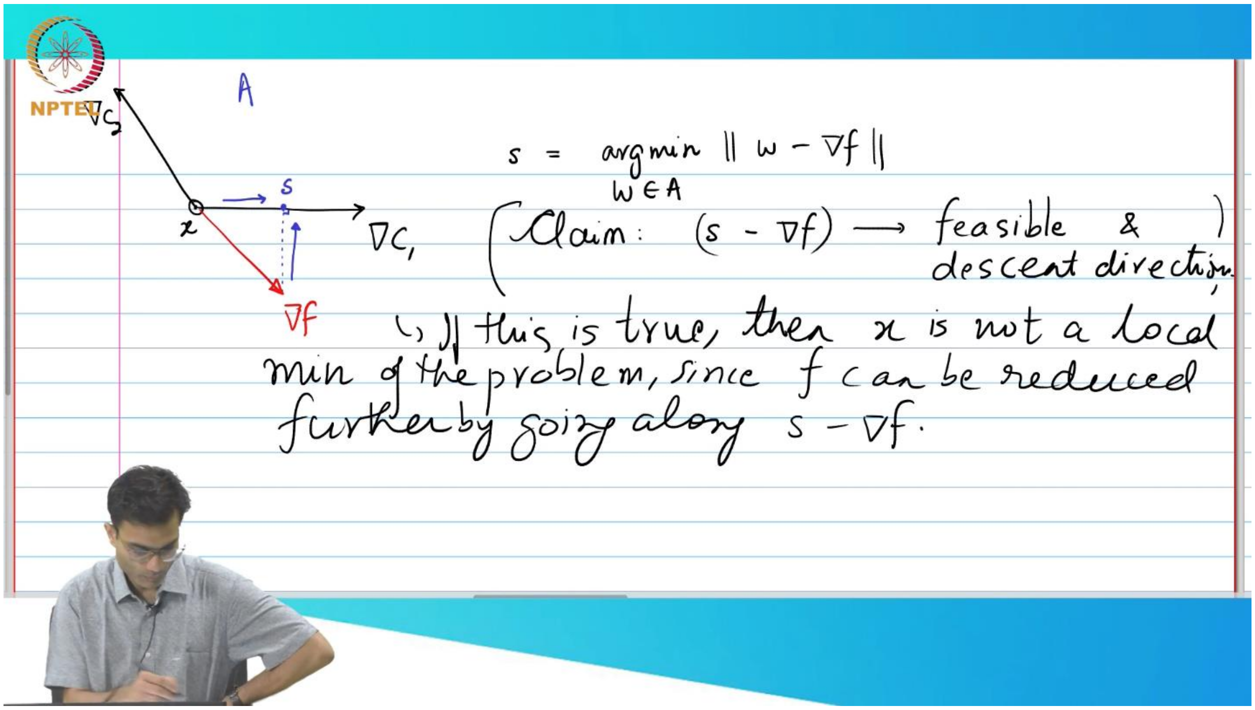
Case (a), the Lagrangian is:

$$\nabla f - \lambda_1 \nabla C_1 - \lambda_2 \nabla C_2$$

So, this is like the most general possible thing that you can think of because I have three vectors to worry about: ∇f, ∇C_1, and ∇C_2, and they all have to sort of work in harmony to get this expression.

Geometric Interpretation of the KKT Conditions and Tangent Cones

The simplest case would be d. What does d remind you of? It reminds us of an unconstrained optimization problem, right? Why? Because the inequalities are inactive. A very simple way to understand or visualize this is by considering the following problem:



$$\min x^2 + y^2 \quad \text{subject to} \quad x^2 + y^2 \leq 1.$$

What is the feasible set? $x^2 + y^2 \leq 1$ is the constraint. What does that look like? It is the interior of a circle, and I am asking you to minimize $x^2 + y^2$. Where is the solution? The origin. Is the inequality active there? No, right? We are clearly inside the circle, not on the boundary.

In this case, the Lagrangian is simply ∇f, and you could have started at any interior point, performed gradient descent, and arrived at the solution. So, this is basically what d is telling us. d in some sense is the simplest, while d and C are similar. There is an additional ∇C_1, but for A, there is both ∇C_1 and ∇C_2.

Now, let us look at how to handle this with our toy example in mind. For A, what do we need to show? We need to show that the gradient of the Lagrangian at x, λ is equal to zero:

$$\nabla_x \mathcal{L}(x, \lambda) = 0,$$

which in turn implies that:

$$\nabla f(x) = \lambda_1 \nabla C_1(x) + \lambda_2 \nabla C_2(x).$$

Any other conditions that the KKT conditions tell us must hold true? Look at the conditions (a), (b), (c), (d), and (e). Is there anything else I need to tell you? Or rather, is there anything else that the KKT theorem tells us? A little louder—$\lambda \geq 0$, because these are inequality constraints. Since these are inequality constraints, I have to say that the Lagrange multipliers should be greater than or equal to zero. This is both necessary and sufficient.

Looking at the KKT theorem, these are the conditions that must hold true. Now, let us keep flipping between algebra and geometry. This is clearly the meaning of $\nabla\mathcal{L} = 0$, right? That is what algebra is telling us. Now, I want to draw your attention to the right-hand side of this equation.

Think of λ_1, λ_2 as non-negative variables, and think of ∇C_1 and ∇C_2 as fixed vectors. Why? Because x is fixed, and I am evaluating the whole thing at a point x. What kind of geometric object does the right-hand side remind you of? It is a linear combination. Can we be more specific? Yes, it is also a convex combination. Can we get even more specific? It is a parallelogram.

But more importantly, in terms of what we have recently been reading about, what else could we say? Hint: ice cream. It is a cone—not tangent, but a cone! This geometric object is a cone. So, what I am saying is that ∇f should actually live in a particular cone. And what is this cone defined by? It is specified by ∇C_1 and ∇C_2.

We had defined a cone last class. What was the definition of a cone? If a vector lives in a cone, then any positive scalar multiple of that vector should also live in the cone. You can verify that the definition will easily hold true over here.

Let us now define this cone properly. Before proceeding further, let us draw it. Here is my point x, which is in the feasible set, and just for definiteness, let me draw ∇C_1 and ∇C_2 at this point x. I have chosen arbitrary directions for ∇C_1 and ∇C_2, pointing here and there.

Now, which way is my cone? Where is my cone? It is the upper half of the region here. To quickly check this: supposing I set $\lambda_1 = \lambda_2 = 1$, where will I be pointing? I will just join the parallelogram and come to that point. This point has to belong to the cone. Put $\lambda_1 = 0$, and $\lambda_2 = 0$, I will get the two ends of it. To get the other side, I would need to use negative values, which are not allowed. So, the cone is defined like this.

This is a quick way to sketch it in your mind. So, we are saying that ∇f should belong to this cone. Let us now sketch the feasible directions. Does anyone remember how we define the feasible directions given ∇C_1 and ∇C_2? What do I start with? I need ∇C_1 and then the orthogonal to the linearized feasible directions.

What are the directions in which I can move so that I still stay feasible? What was it? Wasn't it $\nabla C_1^T d \geq 0$ and $\nabla C_2^T d \geq 0$? If this was an equality instead of an inequality, what would change? This would become equal, but in this case, it is inequality.

Now, $\nabla C_1^T d \geq 0$ defines what kind of geometric object? A half-space, right?

So, I would draw something like this and am I to the left or the right of this? I am to the right of this. Similarly, I take $\nabla C_2^T d > 0$, also a half-space, right. So, I sketch one thing like this, right. So, what are my legitimate feasible directions? They are in this intersection over. Anyway, this was just to give us some clarity on which would be a correct direction to go to. This is the linearized feasible direction. After that, I would also have to consider what else? When I think of an algorithm, this has given me a set of regions where, if I go, I will continue to be feasible.

What is the other thing that I have to worry about? Not length. This has no mention of what? The objective function, right. It is $\nabla C_1, \nabla C_2$ who are dictating this. What is the other thing I will have to worry about? Descent direction. Is this a descent direction? These two things together will help me to go forward in any algorithm, but we are not talking about an algorithm right now, okay.

So, we are going to use one of the oldest tricks in the book to prove this, which is proof by contradiction, okay. So, proof by contradiction means I will assume that ∇f is not in this cone, okay. So, I need to prove this. So, where can I draw my ∇f? Anywhere, but inside the cone, right. So, I am going to draw a ∇f like this, okay.

∇f is not in the cone. So, I have taken some arbitrary point over here, okay. Now, I am going to do an interesting thing. I am going to ask you first in words, and then we will write it mathematically. What is the point? So, this is a what is the point in the set A, the cone A, which is nearest to ∇f? How would I draw this or figure it out? I drop an orthogonal onto this, right. So, it would be something like this, right.

This point I am going to use a lot. I am going to call this guy S, okay. S is as close to ∇f as A can get. That is the meaning of this guy, okay. Now that you have gotten it by geometry, let us try to write it in fancy language. So, S is S, the solution to an optimization problem? If I write it, if I think of it a little bit formally, is S the solution to a small little optimization problem? Yes, right. So, what is the variable that I am allowing from? What am I allowed to choose to come up with this S? Any point in A, okay.

So, $w \in A$, what do I want to minimize? The distance to what? ∇f. So, what should I write over here? $\| w - \nabla f \|$, right. So, this is a very simple example of a projection operation. It looks fancy when you write it, but it is just going step by step from left to right. I come up with this, okay. So, this is it. It looks a little arbitrary that I started with ∇f and suddenly I have proposed this projection point, but there is a reason I am doing this. The claim is if I look at this new vector $S - \nabla f$.

Okay, $S - \nabla f$, which way will $S - \nabla f$ point? It will point from ∇f towards S, right. So, if I draw it, this is what the vector looks like, and you can check it very quickly, right. If I take the red arrow, which is ∇f, plus the blue arrow, which is $S - \nabla f$, what should I be left with? S, which is what I get, right, because S is this vector. So, the claim is that this $S - \nabla f$ is a very interesting vector because, well, you will see, it is both a feasible direction and a descent direction. We will show this. We will show this using simple geometry, that this new funny object that we have created from ∇f, we got the projection point, and then I got this difference vector $S - \nabla f$. This is both a feasible and descent direction. So, while the problem is still fresh in your mind, what are we trying to prove? We said that for the implication of KKT to hold, I should have that ∇f must live in this cone. Then we said, let us do proof by contradiction. That means ∇f should not be in this cone.

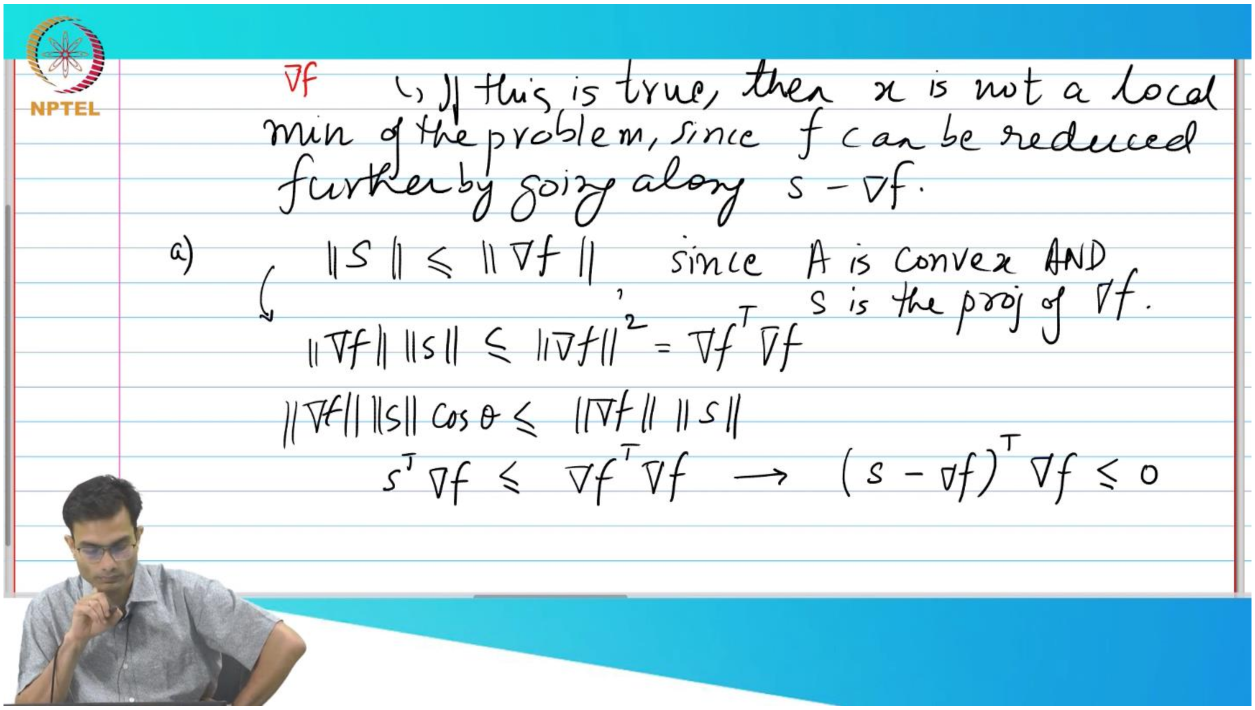
Now, that is what we started out with. Now, if I find that $S - \nabla f$ is both feasible and a descent direction, what does that imply? Yes. So, in your case, this is why paying attention to every line of the theorem pays off. What is the first line of the if condition of the theorem? x should be a solution to, I mean it should be a local minima of the optimization problem. That is what I started with. Only then the implications follow. Is that somehow being violated? Right. It is being violated because I have found a direction in which, if I go, not only do I continue to be feasible, but I also descend, meaning improve the objective function f. That means x is not a local minima. Therefore, the rest of the theorem does not hold. Therefore, the contradiction is there, and therefore, ∇f must belong to the cone.

So, there are several nested levels of logic happening here, right. So, if you are confused, let us revise it once again. Is there anyone who would like me to go over this once more? Okay. So, what is the first statement of the if condition? x must be a local minima. That means in the neighborhood of x, there is no better x. Then I said, okay, let us prove by contradiction. Let us assume that ∇f does not live in the cone.

Okay, that cone came from the KKT theorem. Now, ∇f does not live in the cone, fine. So, then we have constructed something new, and I am claiming, which I will prove shortly, that this new vector $S - \nabla f$ is both feasible as well as a descent direction. That means, standing at x, I can improve myself in terms of the objective function value. That is my contradiction. What I started with is wrong. Therefore, ∇f will live in the cone. So, that is the general direction that I want to go into.

So, let us just make a note over here. If this is true, then x is not a local minimum of the problem since f can be reduced further. Okay, that is the general idea. I want to give you the entire story that we are going to do before we get into the nitty-gritty details. Okay. So, there are two claims. So, there are going to be two sets of reasoning for it.

So, let us take the first one. Okay. Again, it pays to use some properties of convex sets without proof, but they are very intuitive properties. Okay. The first property: the existence of S comes only once you give me ∇f, agreed? Once you give me ∇f, I do a projection operation and I arrive at S. What can you say about the norm of ∇f and the norm of S? Norm is a scalar; can I compare scalars? Yes. Can you think of an intuitive relation between the norm of ∇f and the norm of S? The length of the projection should have a smaller length in general than the original point. Okay. And remember, I am saying convex sets. You can have funny things happen with non-convex sets. Okay.



So, I can say that $\| \nabla f \| \geq \| S \|$ since A is convex and S is the projection of ∇f. Okay. So, everyone agrees with this? Okay. So, from here I can do a little bit of tricks. I can, if I multiply $\| \nabla f \|$ on both sides, it is a positive number, so the inequality will stay as it is. Right. So, I can also write this as:

$$\nabla f^{\top} S \leq \| \nabla f \|^2$$

Can I further reduce it by multiplying some cosine of something? Yes, so I can say that $\nabla f \cos\theta \leq \| \nabla f \|$, agreed? Cosine of theta is less than or equal to 1.

This is fine. Now, what θ? This sounds like a funny exercise I am doing just multiplying numbers on the left and on the right, but can I choose this $\cos\theta$ to be the cosine of the angle between ∇f and S? Up to me. So, if I choose it, what does the left-hand side become? $\nabla f^{\top} S$, right? So, this becomes:

$$\nabla f^{\top} S \leq \| \nabla f \|^2$$

Can I combine this now? I can take ∇f common. What will I be left with? $S - \nabla f$. This implies:

$$S - \nabla f \text{ is a descent direction.}$$

That was quite easy to prove, right?

So, the first part of the puzzle we have solved. We have found that $S - \nabla f$ is a descent direction. Okay, but as you know, just proving a descent direction is not enough because it has to also be feasible. It has to make sure that I stay in that set. Everyone clear with this argument?