Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 15B
Constraints qualification, Farkas' lemma and KKT conditions

So the trouble that we what we learnt was that basically the same geometric object can be represented in multiple ways using formulae or using algebra basically. And our way of calculating normals is by tooking a derivatives of all of the algebraic function, but then that may not be the right give may not give you the true normal.

Student: (Refer Time: 00:42).

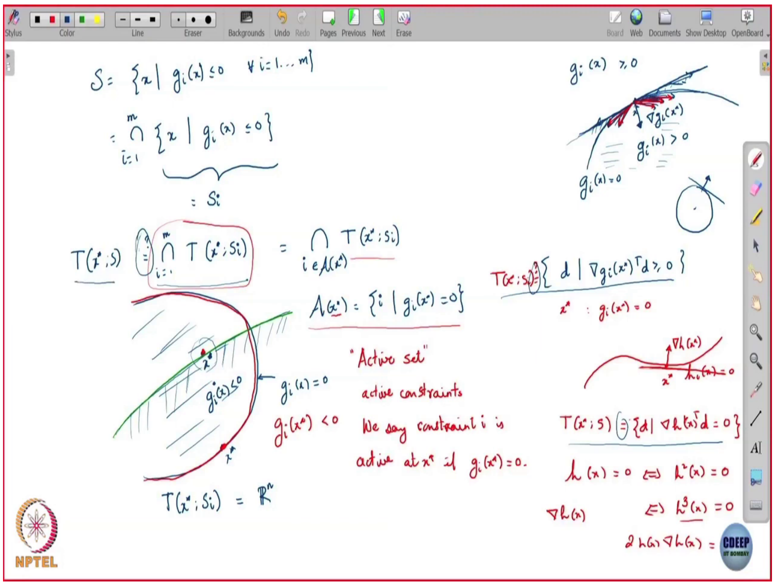
Yes.

Student: (Refer Time: 00:45).

Could have arisen there, could have arisen there. And I will come to that. So, I will tell we I will put together you will remember there that there was a condition I had used about the gradients of the constraints must be linearly independent right, and that is the; that is the condition that will save us. So, it turns out. So, I will come to so good that you brought this up.

So, what is it that we need to fix this essentially? So, this is what this conditions that will basically ensure thing that, so, if I. So, what do we want? Let us come back to this again.

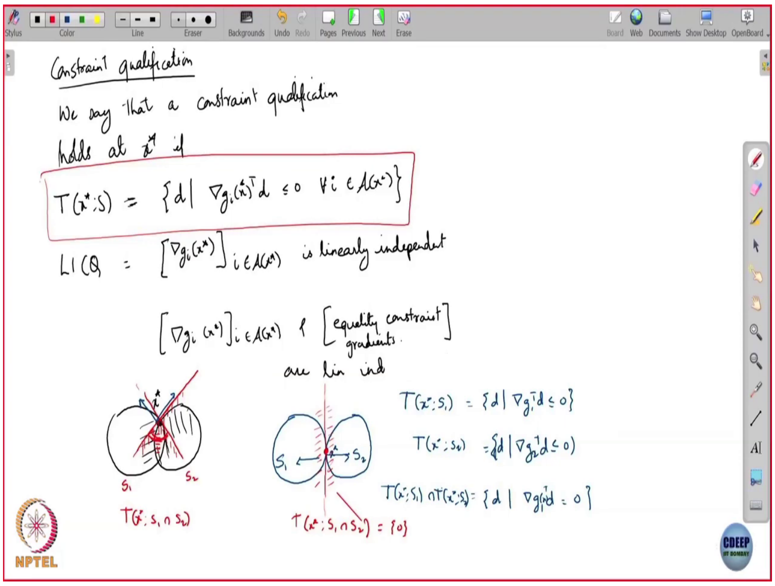
(Refer Slide Time: 01:22)



So, we wanted we would have liked that tangent cones can be described using formulae like this, using these sort of formulae right, this is what we want ok. And if they are describable using that sort of formulae, then an additional thing an additional thing that we want is that there is equality here.

If you have; if you have equality here in these cases and then you have equality here, then we are done. Then you have a formula eventually for the tangent cone of the of this of the set as a whole right. So, conditions that ensure that right are what are called constraint qualifications.

(Refer Slide Time: 02:05)



So, a constraint qualification, so we say that a constraint qualification holds that a constraint qualification holds at x star if the tangent cone ends up being equal to this. So, since I am looking at only equality constraints, let me look at; let us look at I will just write it for sorry inequality constraints, I will just write it for inequality constraint. So, this is equal to sorry for all i in the active side.

So, a constraint qualification holds if somehow you get that the tangent cone is equal to the intersection of the individual tangent cones, and the individual tangent cones in addition are given by the formula that you expect them to do given by ok, then we say that a constraint qualification only.

So, where, so what kind of constraint qualifications? Well, there are many different conditions out there for example one of them is as that the gradients of these constraints should all be linearly independent ok. So, for example, there is a constraint qualification called the linear independence constraint qualification which simply says that this if I look at all the constraints constraint normals, then over a collection of inequality constraint normals that are formed by the active constraints, then these this set is linearly independent.

If you have in addition inequality if you also have equality constraints, then you need that this the this in addition to and all of this and the equality constraint gradients. These are all linearly independent ok. This is one example. But this is just one out of many several constraint qualifications that are out there. Eventual aim of all of them is to get equality here. They eventually ensure that you get that the tangent cone is given by this formula, what, there are many ways by which this can be attained ok.

So, what is. So, I will let us go back and look at the two examples we had seen. See, one of the this let us look at the linear independence constraint qualification. Linear independence automatically ensures that you will not get equality constraint gradients, the gradients of all the equality constraints. I will write out write this out more clearly ok.

So, we had seen two examples right. So, one example was just in the previous slide where I said that suppose you have it is possible that h of x you have a constraint like this h of x is equal to 0, but then that is the same as x square of x equal to 0, x cube of x equal to 0 and so on.

Now, if you look at the gradients of x square, x cube, etcetera, these gradients will all end up being 0 necessarily by just h being on the constraint, because you will get 2 h of x gradient of 2 h square, 3 h square of x gradient of x etcetera etcetera. So, they being 0 will mean that you will have a problem right. You cannot have a 0 vector in a linearly independent side, so that is not. So, this prevents these kind of cases for mapping.

Another case which we had seen in the previous lecture was this case where you had two circles. Remember I showed you this case I had we had these two circles that intersected in this sort of way. And we looked at a point x star like this. And you at a at this sort of point x star you said that the tangent cone would be you can look you can think of the tangent cone in this way you have this, and you have this, and the common region here this part would be the tangent cone at x star right.

So, this is. So, here what was the logic? Well, here the logic was you said we will take the tangent cone with respect to one set tangent cone with respect to the other set ok, alright. And look at the intersection of the two. But then the intersection logic failed when we looked at this example, the same circles I move them apart a little bit in such a way that now they are just touching each other at x star.

What would happen in this case, the tangent cone with respect to one set would be this, tangent cone with respect to the other set ok, so the tangent cone at x. So, we are talking of tangent cone at x star in the common region right. So, tangent cone of x star with respect to S 1 intersection S 2. A tangent cone again of x star with respect to S 1 intersection S 2, that tangent cone should have been 0.

But then if you look at the intersections of the individual tangent cone, then what you get is an entire line. And you see you can see here also what is happened is that the what is failed is linear independence. So, if you had, if the gradients of the two sets, so in this case the gradient in this with respect to a gradient of the constraint S 1 here was normal here, gradient with respect to S 2 was normal here right, and their intersection the intersection of the individual tangent cones was giving us the correct tangent cone.

And the reason that was all working out nicely was because the gradients were actually because the gradients were linearly independent. Whereas, in this case if you look at the gradients while the here the gradients are actually because the sets that sets are perfectly tangent at that point the gradients actually are collinear opposite direction, but collinear right. So, again linear, so here again linear independence has failed.

So, linear independence constraint qualification is one sweeping constraint qualification that actually takes care of both these issues. It takes care of issues of representation; it takes care of issues of that making sure that the tangent cone of the intersection is the intersection of the tangent cone right.

So, both of these gaps which is the first one being ensuring that there is equality in these two places, and the second one ensuring that there is an equality here both of these gaps are plugged once we have linear independent ok. And more generally any other constraint qualification ok, so that. So, constraint qualification ensures that you have equality throughout here.

Student: (Refer Time: 10:49).

No, these two constraints, so the s, so here are my sets right S 1 and S 2, the gradients with respect. So, I can look at the tangent cone of S with respect to S 1 as the tangent cone with respect to S 1 as d such that this is less than equal to 0.

And the tangent cone with respect to S 2 as this, but then if I look at the common region of these two because g 1 and g gradient of g 1 and gradient of g 2 are collinear, the common region of this is simply gradient of g such that sorry d such that gradient of g 1 wherein this transpose d is equal to 0 that would be the tangent cone with respect to S 1 in that is what sorry that would be the intersection of the two tangent cones.

So, the intersection of the two tangent cones would turn out to be this line, this halves hyper plane, but the true tangent cone is actually just 0.

Student: (Refer Time: 12:29).

That constraint that concern is also solved that is what I am saying. So, that concern is also solved by the linear independence.

Student: (Refer Time: 12:38).

So, then in that case you would not have the linear independent effectively. So, constraint qualifications take care of all of this. So, all, so that is what I said. So, there are two gaps here, two possible gaps, one gap is out here making sure that this holds with equality, the other possible gap is here making sure that this holds with equality. So, once we have both these plug, we will get equality throughout.

Student: (Refer Time: 13:13).

But power terms are one possible reason, the point is the unifying one unifying way of dealing with all of this is linear independence.

Student: (Refer Time: 13:26).

Yes, but see for example, but how would you know that you are at the minimum power term, or see power is one kind of reason why this happens.

Student: Yeah.

It was just a way for you need to demonstrate that the same geometric region can be expressed with multiple different algebraic constraint, but it is not that easy to know whether this is you are at the minimal nor or not right ok. So, as that is what this constraint this condition is ensuring. It is automatically taking care of such matters al right.

So, now go to a fresh page, so alright. So, this, so what did we conclude? We said that well if you have a constraint qualification, then constraint qualification basically ensures this ok. This is what I want you to remember that constraint qualification ensures that this holds.

Student: (Refer Time: 14:26).

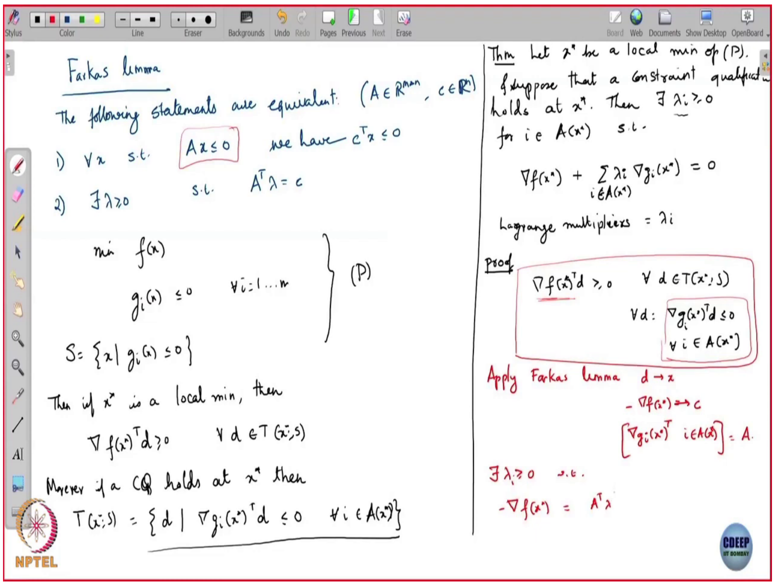
Formulate the, yes. So, you can put it in the for there are many ways you can apply this. One way is that you should see if you can get rid of constraints or redundancies and so on. So, that you have left with only a linearly independent set ok, that is one possibility. Other, possibility is to see if you can formulate the problem to begin with in such a way that linear independence holds I mean the so somehow you need to ensure that this you know the some constraint qualification is holding.

Linear independence is usually not that easy to guarantee, because you have to check it point at a each point, some then there are other constraint qualifications also that are you know that are of a slightly different kind, but then they will be more restrictive. They may not always they may not apply for that many problems.

Student: We need to (Refer Time: 15:25)?

No. So, the way we way this is usually applied is that one checks this whether this holds for all x star, all points in the set, all everywhere is your gradient are your gradients linearly independent that is what you would like to check ok. So, now, let us move forward.

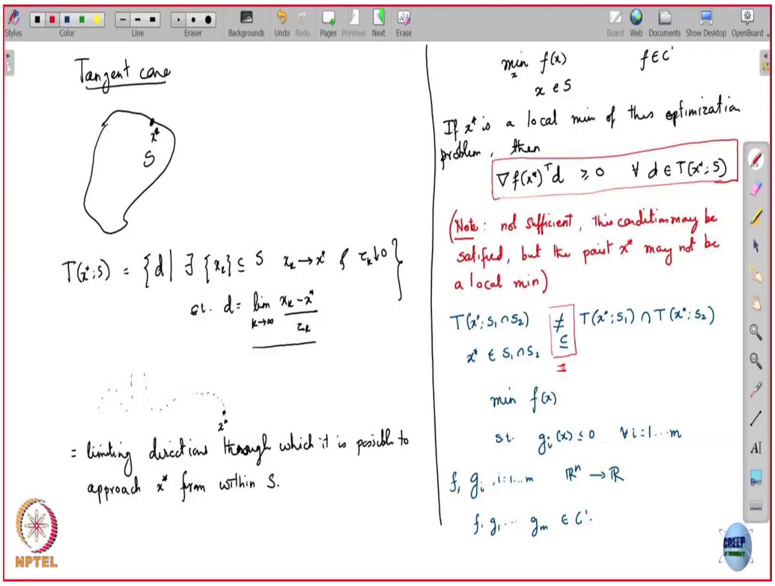
(Refer Slide Time: 15:48)



So, then now recall this recall Farkas lemma that I had mentioned to you couple of lectures ago. So, what did it say well it say that the following statements are equivalent for all x such that Ax is less than equal to 0, we have c transpose x less than equal to 0. And the 2nd is the statement that there exists a lambda greater than equal to 0 such that A transpose lambda is equal to c.

So, x for all x such that Ax is less than equal to 0, we must we have c transpose x less than equal to 0, and there exists lambda greater than equal to 0 such that A transpose lambda is equal to c. So, here recall this is. So, here A is just a matrix and R m cross n, and C is a vector in R n ok. So, that is the context of this alright. So, now, what we will do is, we will just apply Farkas lemma to the two things that we have, what we have.

(Refer Slide Time: 17:29)



We have what I proved in the previous lecture which is this statement which is that the gradient of f at x star transpose d is greater than equal that if x star is a local minimum then of this optimization, then gradient of f at x star transpose d is greater than equal to 0 for all d in the tangent cone. And moreover if I if my optimization problem is given in this sort of form, if my optimization problem is given in this sort of form, then the tangent cone is given by this expression right.

So, let me summarize this. So, if let us, so consider once again this optimization, minimize function f subject to g i of x less than equal to 0 for all i going from 1 to m ok. And let S be the feasible region, then x star if x star is a local minimum, then gradient of f at x star transpose d is greater than equal to 0 for all d in the tangent cone. Moreover, if a constraint qualification holds at x star, then T of x star is the same as d such that gradient of g i transpose d is less than equal to 0 for all i in the active set ok.

So, we I have just summarized what was there on the previous two slide. So, we have one condition here which says that the gradient of the objective if you have a local minimum x star, then the gradient of the objective must make an acute angle with all directions in the tangent cone. The other condition which gives a formula for that tangent cone.

It says that well the tangent cone is comprised of d such that gradient of the constraints make acute angles with the directions in the tangent cone for all active constraint, sorry, obtuse angles with the directions of the tangent cone for all active constraint right ok. So, now look at these two things and compare them with the Farkas lemma that is written right up there.

Farkas lemma says that for all x such that this Ax less than equal to 0, we must have c transpose x less than equal to 0 that is equivalent to saying that there exists the lambda greater than equal to 0 such that c can be written as A transpose lambda ok. So, let us apply let us do this. So, I will just state for you the final theorem. So, suppose this is let us call this optimization problem P.

Let x star be a local minimum of this optimization problem P ok. And suppose that a constraint qualification holds at x star ok. Then there exist lambda i greater than equal to 0 for i in A of x star such that gradient of f at x star plus the summation lambda i grad g i of x star i equals i in A of x star this must equal 0.

Now, what is, what does this say? Well, if x it says that if x star is a local minimum of this optimization problem and suppose that are constraint qualification hold that x star, then you can find lambda i greater than equal to 0 for i in the active set of pertaining into x star such that the gradient of the objective is plus a linear combination of the gradients of these constraints is equal to 0 ok.

Now, this you will this of course, will remind you of something that you have already seen in the case of when we were talking of optimization over equality constraints. What are these lambdas? These are Lagrange multipliers. Then lambda is are actually the Lagrange multiplier ok. You will we will make this slightly more general in a moment, but for that I will quickly show you how this is; how this is derived ok.

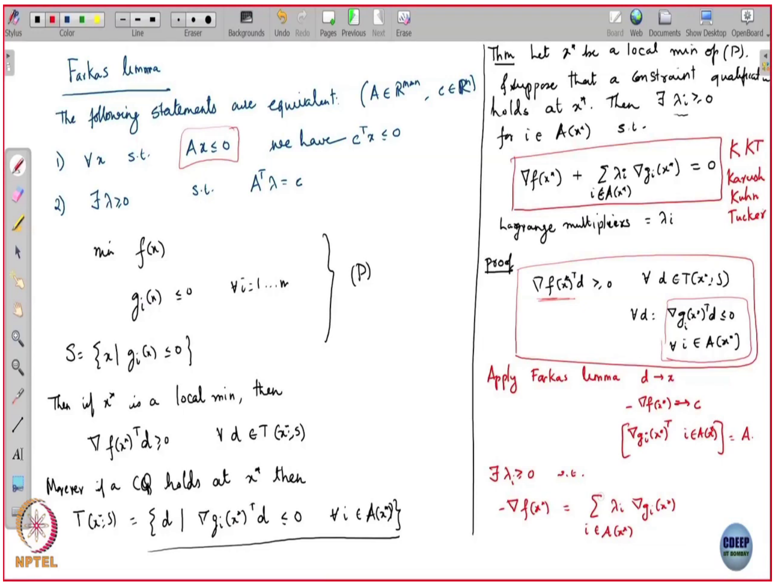
So, I know that this is greater than equal to 0 for all d in the tangent cone. And tangent cone is the same as a tangent cone is given by this particular expression which means that this is greater than equal to 0 for all d such that gradient of g i transpose d is less than equal to 0 for all i in way of x star right.

So, now what I have all I need to do is just apply the Farkas lemma. How do I apply Farkas lemma here? So, Farkas lemma the condition this condition where g grad g i less of grad g i transpose d less than equal to 0 for all i in A of x star that is analogous to this condition Ax less than equal to 0 ok. So, and my grad f transpose d greater than equal to 0 that is analogous to this. All I need to do is just change do the change the sign. So, c becomes minus f, and then I get back this condition right.

So, what is. So, effectively I am. So, my the so apply Farkas lemma in the following way. So, your d becomes x, the your minus grad f at x star becomes c and all these g i of x stars for i in this if I put these together they form my matrix A sorry, they this transpose put these together they will form my matrix A, they are the rows of my matrix A right. So, I put that together and I can apply Farkas lemma.

Now, then what does Farkas lemma tell me? This is this it tells me that this thing in the box holds that is equivalent to there that there exists lambda grad lambda i greater than equal to 0 such that my c, c is minus grad f minus grad f at x star is equal to A transpose lambda. And A transpose lambda is what, that is just that is basically this summation.

(Refer Slide Time: 27:04)



That is this summation lambda i grad g i of x star i in a of x star. If I take my grad f to the other side, I get the condition that I am looking for right. So, this here is what I just wrote this is what is called you have this is called the KKT condition. KKT stands for Karush Kuhn Tucker, Karush Kuhn and Tucker conditions for the optimal.

So, these give you necessary conditions for a x star to be a local minimum of an optimization problem assuming a constraint qualification holds. If a constraint qualification holds and what should be what must that x star satisfy that is what is given by these KKT conditions. Such that while x star is a local minimum and a constraint for qualification hold, then you must be able to find Lagrange multipliers lambda i greater than equal to 0, such that this condition in this boxed condition is satisfied alright.

So, this is now we can this can be general put made little bit more general, and I will explain that in the next in the next lecture. So, what I have not allowed for so far are equality constraints in this definition of this optimization problem (Refer Time: 28:47) that is not very hard to do because any equality constraint can also be written as two inequality constraints opposing inequality constraint.

So, those can be taken into account, and that will give putting all that together will give you a comprehensive, one comprehensive condition which can be applied for any type of optimization ok. So, we will do that in the following lecture ok. So, any questions? You can say well a constraint qualification will be used is used to get you to this point, this equality here. See without constraint qualification you would not have got concluded this right that is the; that is the that is why the problem.

Student: (Refer Time: 29:36).

Linear independence is one of them, there are others I will give a homework based on those also. So, there are other type of constraint qualification, linear independence is one simple one. Look at all the gradients and check if they are linearly independent.