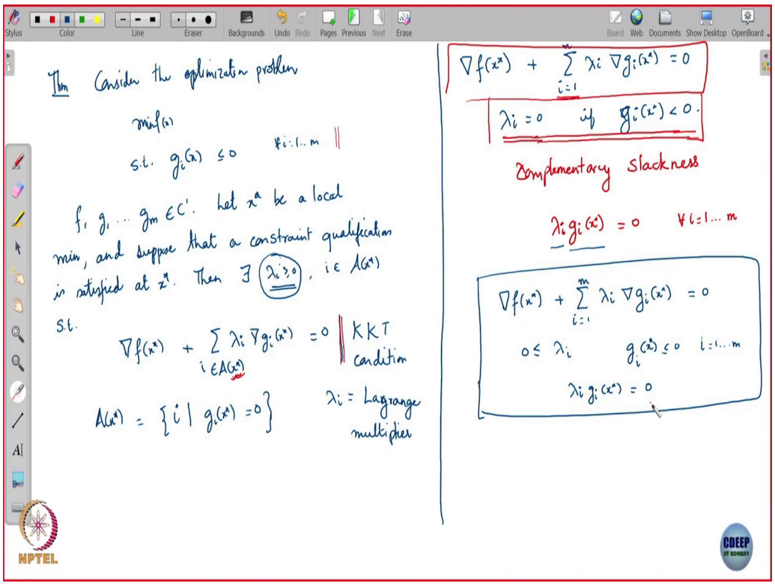
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 16A
KKT conditions

So, now, we will continue with the Theory of Optimization that we have been building up so far.

(Refer Slide Time: 00:29)



So, what we learnt in the previous lecture was we basically proved this theorem that consider the optimization problem; consider the optimization problem, minimize f x subject to g i of x less than equal to 0 for all i from 1 to m.

Student: Yes.

Where f and g 1 till g m these are all c 1. Let x star be a local minimum and suppose that a constraint qualification is satisfied at x star, ok. Then there exist lambda i greater than equal to 0, for all i in A of x star such that gradient of f at x star plus the sum of lambda i gradient of g i of x star over all i in A of x star equal 0 ok, where now A of x star we call is the active set it is those i's for which g i of x star is equal to 0.

So, this is what we derived last time. This condition is what is called was called the KKT condition due to Karush-Kuhn-and Tucker and the lambda is here called is called the Lagrange multipliers. Now, you can what we will do now is we can generalize this a little bit.

We can write the same condition here in a slightly different way. We can see if you have suppose if you have two if you have we can write this condition here in the following form. We can say let us consider this summation which is now only over A of x star. Only over the active set, let me write this summation in the following way.

I will write f of x star plus sum over lambda i gradient of g i of x star and the sum is now over all i, i going from 1 to m. This is equal to 0, but then in addition what I will do is I will put a condition that says at lambda i equals 0 if g i of x star is less than 0. So, it is says this condition simply says that lambda i is equal to 0, if your constraint is not active.

So, then what it means is effectively as a result this sum here which is over all the constraints i from 1 to m will reduce to just a sum over the active constraint. But, you can see what is happened as a result, we have now got back this condition that we have seen as part of linear programming.

I had to be; a couple of lectures ago I told you that there is this there is this condition which simply which says that this variable Lagrangian the, there is a variable of the dual problem which is which must be 0, if a particular constraint is not satisfied with equality if the corresponding constraint is not satisfied with equality and this condition was what was what we call complementary slackness.

So, this complementary slackness in the case of general optimization not this linear optimization is being precisely this condition. It says that if the constraint is not active then the corresponding value of the Lagrange multiplier must be 0, ok.

So, as a what this effectively says say thus is it asks us to now look for says that if you want to solve for look for a necessary condition for a point x star to be a local minimum what you need to solve for or you need to make sure that your constraints are satisfied.

And you need to look for m Lagrange multipliers lambda 1 to lambda m, so that the KKT conditions hold. But, the KKT now the KKT conditions are linear in lambda, but maybe non-linear in x star, but in addition to the KKT conditions we have to also satisfy complementary slackness meaning that you need lambda i to be equal to 0 if g i of x star is less than 0, ok.

So, what is the what is the what is the point of writing it in this way? The good the use the what is nice about it is that now when I look for when I try to solve the KKT conditions I do not have this x star sitting in this summation here. The x star which was which is telling me the indices are involved in the summation that kind of dependence is now gone and now my summation is over all i from 1 to m.

But, the complication has now arisen that now I need to make sure this new condition, which is complementary slackness that needs to be satisfied, right. So, KKT conditions will now involve a one non-linear equation like this which is this one and in addition to that a complementary slackness condition which is this.

Now, complementary slackness itself we can simplify and write in a nicer form. We can say the, that the complementary slackness can be written in this sort of form that for all i lambda i into g of g i of x g i of x star is equal to 0. Lambda i times g i of x star equal 0. Now, this and this should be true for all i, not just for the ones that are not active, this is true for all i.

The ones that are for the i's that correspond to active constraints g i of x star will be equal to 0, for the ones that are not active it necessarily means that lambda i must be equal to 0, right. So, since the product of these two is equal to 0 at least one of them must be 0. So, which means that if your g i's if your constraint i th constraint is active, then I do not care what the Lagrange multiplier is so long as it is greater than equal to 0 as written here.

And, I if my constraint is not active then I have no choice, but to make sure that my Lagrange multiplier is 0, right. So, the way KKT conditions are often written then is in this sort of comprehensive form. You have gradient of f from i equal to 1 to m lambda i g gradient of g i star equals 0 and you have lambda greater than equal to 0, g i of x lambda i greater than equal to 0 g i of x less than equal to 0; i running from 1 to m and you have lambda i times g i.

So, x star here sorry, lambda i times g i of x star equals 0. So, this is these are your KKT conditions, ok. Now, you can see what is happened in this sort of problem because of the nature of inequality constraints the whether a term will appear in this first equation here, in this first equation whether the ith the gradient with respect to the i-th constraint is going to appear or not will depend now will depend on the x star you are considering because after all that only terms that appear there are the ones that are active.

So, it will depend on the x star that you are considering. So, if it appears then this particular, so, which means, if your x star which means if your thing if your constraint is active then you do not need to worry about this complementary slackness condition. And, lambda i only thing you need to worry about is making sure lambda i is greater than equal to 0.

But, if it does not if it does not appear then it what you are effectively doing is putting lambda i equal to 0. So, essentially solving a optimization problem with inequality constraints involves making trying to first involves basically first trying to check, which constraints are actually active.

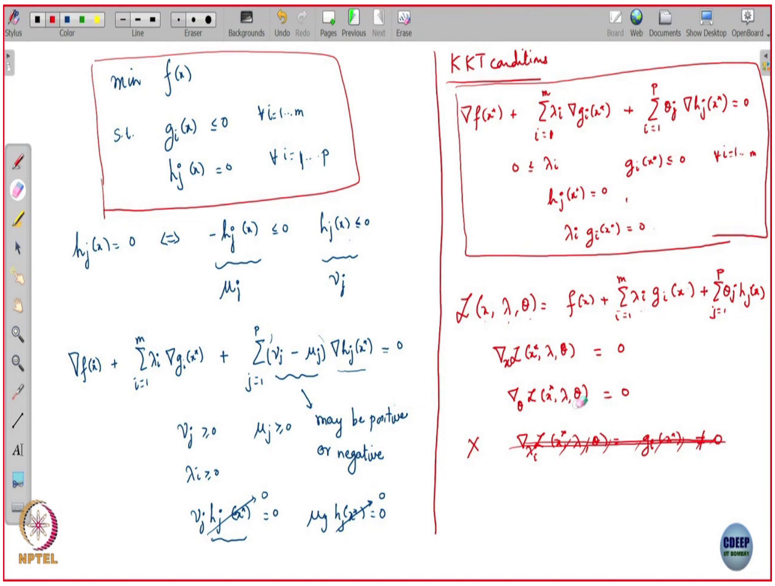
Because, once we define the active constraints then the A of x star gets fixed and then we can hope to just simply solve this equation this non-linear equation. Without the active constraints having first been determined it becomes very hard to do that. So, there are many. So, in implicitly in your in an optimization with inequality constraints is this try is this effort to try and make a combinatorial choice.

Out of the m constraints which K are actually active? We are trying to always decide the set of active constraints either directly or indirectly. Directly means that you actually try to keep searching over active constraints or indirectly means that you try to somehow try to discover them through complementary slackness meanings through this condition.

So, remember this that the that optimization problems with inequality constraints have a basically a combinatorial flavor to them. Because the nature of the problem changes from whether you are whether the constraint is active or not active. If it is not active essentially the tangent cone is R n.

You do not need to worry about you have plenty of room around a particular point to see what to perturb the function and to perturb the point whereas, if you are if your constraint is active, then you have then your when the nature of the problem changes because you are constrained to move only in certain types of directions ok, alright.

(Refer Slide Time: 12:52)



So, now let us add a little bit of for completeness sake let us add a little bit of complexity here. So, let us allow make this problem consider this problem minimize function f subject to g i of x less than equal to 0 for all i from 1 to m and also h j of x equal to 0 for all i from 1 to p.

This for this sort of problem the what we can do is we can simply look at the constraint h j of x equal to 0 as been two opposing inequalities; so, minus h j of x less than equal to 0 and h j of x less than equal to 0. You with these two opposing inequalities we can again write out the write out the tangent cone conditions, write out the KKT conditions.

And, what that would give us is you would now get a Lagrange multiplier for this constraint and you will get a Lagrange multiplier for this constraint, right. So, suppose the Lagrange multiplier for this constraint is mu j and the Lagrange multiplier for this constraint is suppose nu j.

So, then in that case the KKT conditions will read as gradient of f at x star plus summation lambda i gradient of g i of x star i from 1 to m plus now.

Student: (Refer Time: 14:30).

Summation nu j minus mu j gradient of h j at x star j going from 1 to p, right. So, how did I get this? I just wrote it wrote out the KKT conditions from the previous slide. But, now I will I consider in place of this equality constraint h j of x equal to 0, I am going to consider an inequality constraint like this minus h j of x equal to 0 and also h j of x equal to 0.

So, with the h j of x equal to 0, I get a nu j as my Lagrange multiplier with minus h j of x I have mu j as my Lagrange multiplier. So, I get nu j minus mu j, alright. Now, what we know is that these both these Lagrange multipliers nu j should be greater than equal to 0 mu j should be greater than equal to 0 just like lambda i is should also be greater than equal to 0. But, then the difference here nu j minus mu j this may be positive or may be negative this may be positive or negative, right.

So, also it must be that ether since eventually for a point to be feasible and since both these constraints must be satisfied it has to be that for a point x star to be feasible it has to be that h j of x star is h j of x star is equal to 0 and so, consequently any complementary slackness type condition is actually meaningless for this sort of constraint.

Because even if I put it I would we end up saying something like this nu j into h j of x star equals 0 and mu j into h j of x star equals 0. This kind of condition automatically holds since h j of x star itself is equal to 0, right. So, as a consequence two things happen one is that whatever is multiplying this gradient of h j has cannot be constrained in sign we cannot say it is greater than equal to 0 or less than equal to 0.

And, secondly, the complementary slackness conditions for h j of x are automatically satisfied. So, the way we can summarize all this in terms of KKT conditions is to write that the now the KKT conditions for this sort of problem, KKT conditions for this sort of problem are gradient of f at x star plus lambda i, i equals 1 to m plus say let me introduce another Greek character here.

So, let us say theta i this equals 0, where now lambda i must be greater than equal to 0 g i of x star must be less than equal to 0, this is for all i from 1 to m h j of x star must be equal to 0, and lambda i times g i of x star must be equal to 0. There is no sign restriction on theta i; theta i can be sorry, theta j I should have written this theta j.

So, there is no sign restriction on the theta j; theta j can be positive, theta j can be negative does not matter or then as because there is also no reason to impose any complementary slackness since h j of x is already equal to 0. So, this what I have written here are your comprehensive KKT conditions for all kinds of optimization problem.

So, any problem that can be written like this and if you have that the constraint qualification is satisfied, then it must be that at x star all of these constraint these conditions if x star is a local minimum then these conditions must hold that is what this means ok, alright.

So, now what we can do now is ask the following question that what if constraint qualifications are not satisfied, what if the KKT conditions hold, but I do not know if I am at how do I know that I am at a local minimum etcetera etcetera ok. Before I do that I am I forgot one thing, let me mention this function l this introduce just for you to remember this is this function is what is called the Lagrangian.

So, this function written with a file fancy L is what it is called Lagrangian. It is a function of both x as well as the two Lagrange multipliers involved the Lagrange multipliers for the inequality constraints, the Lagrange multipliers for the equality constraints, right.

So, the first equation therefore, in the KKT conditions are actually simply saying that if you take the Lagrangian and differentiated with respect to x the gradient and take. So, the gradient of the Lagrangian with respect to x must be equal to 0. That is what the first condition is saying.

Sometimes people also write this second condition here this second condition or this feasibility condition here that h j of x is less than equal to 0; sometimes people also write it as saying that gradient with of the Lagrangian with respect to sorry, this should not be mu this should have been theta the gradient of the Lagrangian with respect to theta is equal to 0.

That that is correct, because then if you take the gradient of the Lagrangian with respect theta it actually gives you h j of x equal to 0. And, put that equal to 0 you get h j of x x star should be equal to 0, which is precisely what is written here. However, you should be careful in not taking this too far, do not.

Now, if you take the gradient of the Lagrangian with respect to lambda if you take the gradient of the Lagrangian with respect to lambda. So, this should be at x star what is that equal to? Well, that what gradient of the Lagrangian with respect to lambda i what that is equal to what that is equal to is simply g i of g i of x star.

And, this is not necessarily equal to 0. This could be 0, could be less than 0. If it is less than 0, well then in that case in that case the lambda i corresponding to it must have been 0, alright. So, you have to be extremely careful in applying this particular thing or where you take Lagrange multipliers where you take gradients of the Lagrangian with respect to the Lagrange multipliers.

So, there is remember that there is no such thing for with respect for the inequality constraints, if you can use that kind of mnemonic you like with for the equality constraints only. So, now with this let us now answer the questions of when what happens if my KKT conditions are satisfied or what if my constraint qualifications do not hold and so on. So, if my constraint qualifications do not hold can I say something about the problem?

If my KKT conditions are satisfied can I still say something about the problem? Maybe may how can I know I just know that the KKT conditions are necessary conditions? How do I know that how what can I say about anything about the above the point itself? So, there are so, these kind of questions let us try to answer them now.

All of these questions basically come down answering these kind of questions come down to one key property which is convexity. The convexity is the heart of optimization because it lets you go reverse this chain of implications. So, we so far the way we worked was we said here is an optimization problem suppose here is a local minimum and if it is if here is a point x star that is a local minimum and we said if x star is a local minimum, then here is a condition that x star must satisfy.

Now, if you wanted if you wanted to go the other way round where you suppose found a point x star which satisfies the KKT conditions what can you say about x star? Is x star a local minimum, is x star a global minimum etcetera, to answer these kind of questions the one of the key property that comes to use is convexing.

So, when you have a convex optimization problems this chain of implications can be reversed means that you if you find an x star that satisfies the KKT conditions, then that x star is the solution of the optimization problem. No questions asked.

It does not matter if your constraint qualifications hold or do not or do not hold it does not you do not in have to check if that x star was is a local minimum to begin with. You just simply solve the KKT conditions once the KKT conditions are solved the problem is solved, alright. So, this is the beauty of convex optimization.