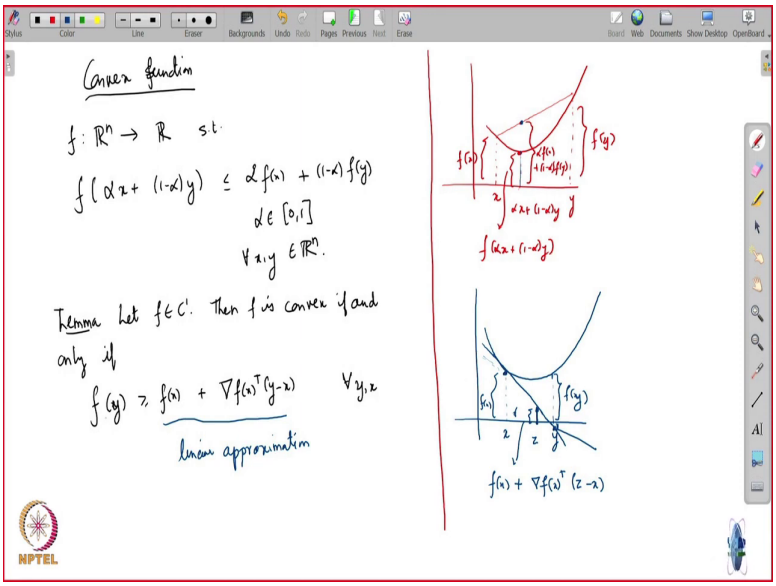
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 16B
Convex optimization and KKT conditions

So, I will let us get to this one by one. So, a function f, so, first what is a convex function?

(Refer Slide Time: 00:25)



We have seen what a convex set is. So, what is a convex function? So, convex function is a function f let us say from R n to R ok such that f of lambda x or let me use alpha, alpha x plus 1 minus alpha y is less than equal to alpha f x plus 1 minus alpha f y where alpha belongs to 0, 1 and for all x comma y in R n.

So, for every pair of vectors x and y, if you look at the convex combination of these vectors with the weight alpha and look at the function value of f at the convex combination that is less than equal to the convex combination of the values itself. So, you here on the left-hand side, you have the function value at the convex combination of the points and on the right-hand side, you have the convex combination of the function values right so, then you have an inequality between the two.

So, the left-hand side is less than equal to the right-hand side. This is what it means for a function f to be convex. Now, if f is so, if usually, if you have, you can actually say more. So, if so, let me write a lemma here. So, let f be c 1. Then f is convex if and only if you write this full if and only if f of y is greater than equal to f of x plus the gradient of the function at x transpose y minus x, this is true for all y and x.

What does this say? What this condition says is let me let us visualize this. What this condition says is so, here is my, here suppose my function f, now for this since this function to be convex, what it means is that here is suppose a point x, here is suppose a point y, this is this height here is f of x, this height is f of y and I take a point like this alpha; alpha I choose an alpha in 0, 1 so, I take a point alpha x plus 1 minus alpha y.

So, suppose I take this point right. Now, for this function to be convex, what must be the case is that if I take; if I take alpha times f x plus 1 minus alpha times f y which is if I take the convex combination of the function values, then that must be greater than equal to the function evaluated at the convex combination of the points.

So, if I take the convex combination of the function values, then that gives me this height, this gives this is actually alpha f x plus 1 minus alpha f y and what is this height? This height is simply f evaluated at alpha x plus 1 minus alpha y right and sure enough this function is convex because as you can see, we have this that this point here this value is greater than this height is greater than this height here.

The height alpha f x plus 1 minus alpha f y is greater than the height f of alpha x plus 1 minus alpha y. So, a function like this is this sort of function is convex. Now, what is this second lemma saying? So, for that let me just draw the same function here again ok. So, what this lemma is saying is well if I so, let me just draw this in a slightly different way. What this lemma is saying is take any point x and another point y.

So, here is the point x, here is the point y. Look at f x that is this, look at f y that is this. Now, in addition look at the gradient of the function at x. So, what that means is I am looking at the slope, the great the tangent drawn to the function at x ok. So, it is saying look at the gradient of the function at a, and now, extrapolate the function linearly along this gradient.

So, if you extrapolate the function linearly along this gradient so, at any point; at any point, the value of the linear at any point z like this here in between the linear extrapolation would have value exactly this, the height here right. So, this is this height here is nothing but f x plus gradient of the function at x transpose z minus x right.

So, now, what is this inequality in the lemma saying? This the inequality in the lemma is saying that you take any point y, ok you take any point y like this so, this part of point y and look at the value of the linear extrapolation form evaluated at x or the linear approximation to the function at x evaluated at y ok. So, we will take this linear approximation, evaluate that at y and that linear approximation the its value at y must be less than the value of the function itself at y, it must be less than f of y.

So, in this figure, f of y is this height, and the value of the linear approximation is actually this it is this particular, it is this, its negative actually it is under below the axis right. So, this is so, this tells you that f of y is definitely greater than this than the value of the linear approximation.

So, what this means is, what this lemma is effectively saying is that this is true for every x and y. So, which means you can take any point x, draw a linear approximation to the function at that point, take any other point y, look at the function value at that point and look at the

value of the linear approximation evaluated at that point and compare the two, it must be the case that the function value is greater than equal to the value of the linear approximation.

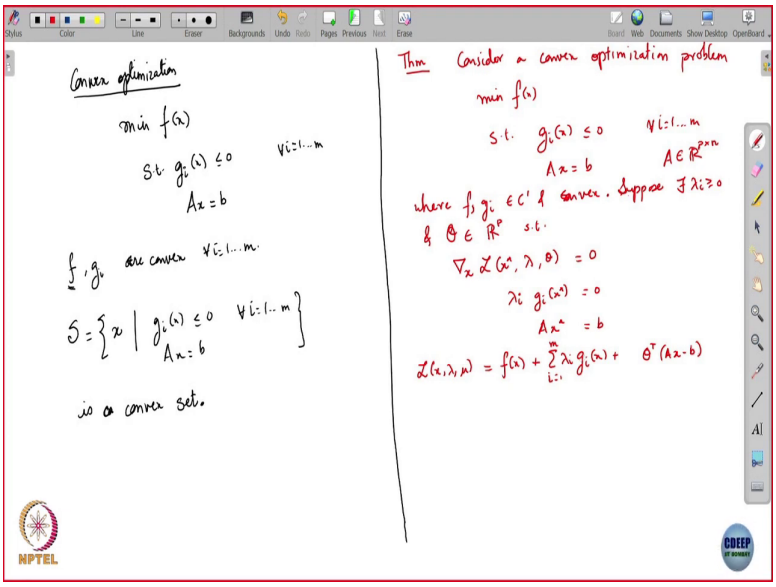
So, geometrically, what this means is that the if I take the entire linear approximation to the function at any point, then it must be that linear approximation is actually a linear under approximation. So, this linear approximation is necessarily an under approximation means it always underestimates the value of the function. Means if I wanted to estimate the value of the function using a linear approximation like this, it will always be less than the true value, the true value being f of y right.

So, what this lemma is saying is that a function is convex if and only if this holds means that if a function is convex if and only if the linear approximation to the function lies completely under the function, it under approximates the function or the and the other way around also if the linear approximation of the function underestimates the function, then it has to be that the function is convex.

If the linear approximation of the function underestimates the function at if linear approximation of the function evaluated at any point underestimates the function, then it has to be that the function is convex ok. So, this is what so, this is an equivalent you can say characterization of convexity.

Some people use this also as a definition of convexity of course, it does not hold when the gradient is not when the function is not differentiable and in that case it you wont be able to take the gradient, but for differentiable function, this is equivalent to convexity alright.

(Refer Slide Time: 10:52)



Now, what is this what does this give us? So, this gives us the following simple fact ok. So, let us write look at this all. So, I will state this theorem, maybe I should write this in red because this is an important result ok. So, before I go there, actually I should mention. So, here is a convex function, what is a convex optimization problem?

A convex optimization; convex optimization is involves minimizing a function f subject to constraints g i of x less than equal to 0 for all i from 1 to m and so, these are inequality constraints and equality constraints that are linear. So, equality constraints must take the form Ax equal to b ok.

So, here the function f and the function g i these are all convex. So, if you, it is very easy to see here from this case that this if you look at the constraints or the feasible region, the x such

that g i of x is less than equal to 0 for all i from 1 to m and Ax equal to b, this set S is actually a convex set.

So, when g i is are convex and the equality constraints are all linear, then the feasible region is actually a convex set. So, then what we are doing is effectively minimizing a convex function f over a convex set S right ok. So, that is what happens in a convex optimization alright.

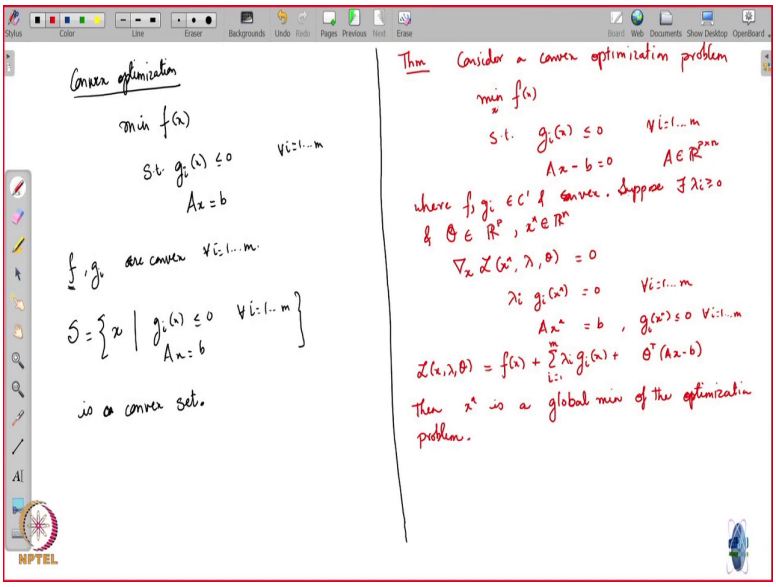
So, now, let us look at what happens to KKT conditions for convex optimization. So, theorem. Consider a convex optimization; convex optimization problem minimize f x subject to g i of x less than equal to 0 for all i from 1 to m and Ax equal to b ok where f and g i these are in c 1 and convex.

Now, suppose so, suppose there exists lambda i greater than equal to 0 and say so, I will and theta; theta j ok so, the matrix A here, let us suppose the matrix A is in R p cross n ok so, there exists a theta in so, let us suppose this theta belongs to R p such that I take the gradient of the Lagrangian with respect to x evaluated at x star, lambda, theta, this is equal to 0. Lambda i g i of x star equals 0 and Ax star equals b.

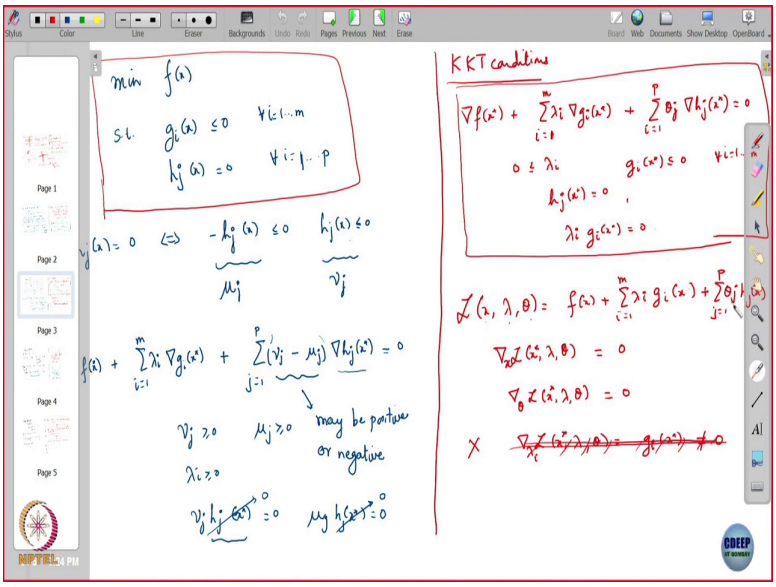
So, I have taken the gradient of the Lagrangian with respect to x star with respect to x evaluated at x star, lambda and theta. Now, the Lagrangian here will be just for clarity let me write the Lagrangian also so, you have f plus summation lambda i g i of x plus now, the way we wrote it earlier is we took every constraint and multiplied it with its corresponding Lagrange multiplier.

A simple; the simple thing to do in this case would be to simply do theta transpose Ax minus b but that effectively does the same thing. So, this constraint Ax equal to b it can be equivalently written as a x minus be equal to 0.

(Refer Slide Time: 16:08)



(Refer Slide Time: 16:14)



And once it is written in the form a x minus b equal to 0, it comes in the form of h j of x equal to 0 in this sort of form and then, we have to just multiply the h j by theta j so, what I am effectively doing, I am defectively doing that by taking theta transpose Ax minus b. So, that gives me a sum of theta j times the jth row of Ax minus b alright ok. So, this is my; this is my Lagrangian for this theta ok.

So, suppose there exist lambda, this lambda greater than equal to 0, theta in R p and an x star in R n such that these KKT conditions are satisfied. So, what are the KKT conditions here? Gradient of the Lagrangian is with respect to x at x star is equal to 0, complementary slackness holds between for all constraints and equality inequality constraints.

Sorry I forgot inequality constraints are all satisfied and equality constraints are also all satisfied. So, in short KKT conditions hold ok then x star is a global minimum of the

optimization problem. So, let us take a moment and absorb this. This is what is this theorem saying?

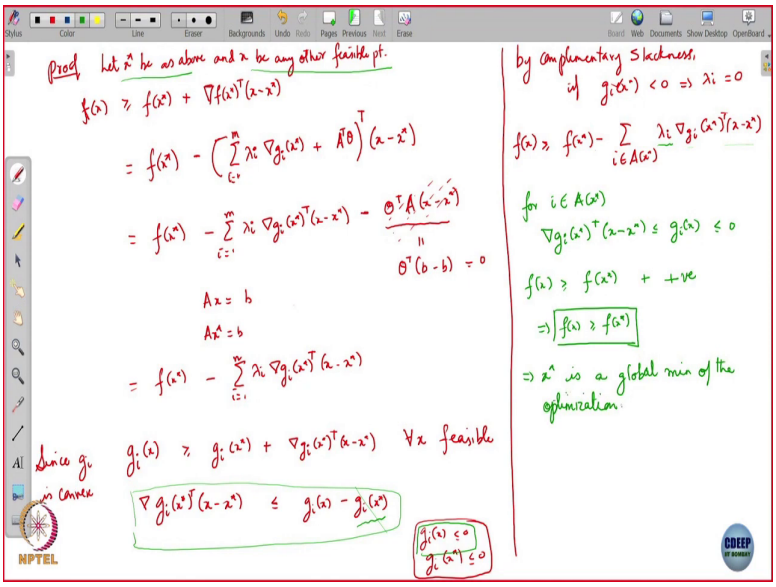
Take any convex optimization problem written as minimize f x subject to g i of x less than equal to 0s, Ax and Ax equal to b where f and g i s are continuously differentiable and convex and suppose you can find Lagrange multipliers and a feasible point x star such that the gradient of the Lagrangian with respect to x evaluated at x star, lambda and theta is equal to 0.

Lambda i x i; lambda i times g i of x star is equal to 0 that means, complementary slackness holds and the point x star is feasible in that case, it must we must have that x star is a global minimum of the optimization problem. So, notice that I have not said a word about constrained qualifications. Constraint it does not matter whether constraint qualifications hold or do not hold. What matters is that you are able to somehow solve the KKT conditions.

If you can solve the KKT conditions, the problem it must be that you have a; you have a global minimum ok. So, you have to you so, for that all you need to do is make sure you have a point that is feasible and that and you have Lagrange multipliers, appropriate Lagrange multipliers that satisfy complementary slackness and gradient of Lagrangian equal to 0.

Once you have that; you once you have that and you are you have solved your KKT conditions, you have a global minimum ok. This is obviously, an extremely powerful result, but it turns out the proof is actually very simple.

(Refer Slide Time: 19:50)



So, let us quickly do the proof of this ok. Let us do the proof of this. So, we just wrote that f of x at any point x is greater than equal to f of x star plus gradient of f at x star transpose x minus x star ok. So, let x star b as above and x be any other feasible point, then we must have that f of x is greater than equal to f of x star plus gradient of f at x star transpose x minus x star this is what we just showed because f is c 1 and f is convex, this must be the case.

But then, what is gradient of f at x star? Well, that can be written through the Lagrangian. The Lagrangian is telling you that the gradient of f at x star; the gradient of f at x star is equal to the negative of if you look at this here, the gradient of f at x star must be the negative of the sum of Lagrange multipliers times the gradients of the constraints right.

So, just putting that in will now give me that this is equal to f of x star plus summation lambda i gradient of g i at x star i equals 1 to m plus now, I need to take the gradient of the

equality constraints so, if you can verify that, this actually becomes the same as A transpose into theta ok that is what this becomes ok. So, this is equal to this what I have written here is just the gradient of f I take x star and I need a negative sign outside this. So, let me put a negative here times, then I have x minus x star alright.

Now, what does this the whole thing transpose? Now, let me expand this out, I will by doing that I get f of x star minus now, I get sum lambda i, i equals 1 to m gradient of g i at x star transpose x minus x star plus theta transpose A into x minus x star sorry there should be a minus, minus theta transpose A into x minus x star ok.

Now, remember x is any other feasible point. So, Ax should therefore, be equal to b, x star is feasible so, therefore, Ax star is also equal to b. So, this term theta transpose A into x minus x star is simply so, this term is actually equal to theta transpose b minus b so, this is actually equal to 0 ok.

So, this term is equal to 0. So, this term is not present in our problem ok. So, we are left with therefore, f of x star minus sum from i equal to 1 to m lambda i gradient at x of x star gradient of g i at evaluated at x star transpose x minus x star ok. Now, gradient of g i evaluated at x star transpose x minus x star.

Now, this is a similar to the term that we had about here which came from the convexity of the function f right. So, now, we remember g i itself is convex. So, consequently, since g i is convex so, we know that therefore, g i of x is greater than equal to g i of x star plus gradient of g i evaluated at x star transpose x minus x star right. So, this and this is true for all x feasible since g i is convex, this must be the case alright.

Now, look at now, let us look at this term therefore, so, take this gradient of g i evaluated at x star transpose x minus x star that is therefore, less than equal to g i of x minus g i of x star ok. So, now, what can we say about this difference g i of x minus g i of x star? The problem you can see that all I know is that x star is feasible and g i of x star is also feasible.

So, x star is feasible, and x is also feasible, these are the only two things that I know. So, therefore, for these since both of these points are feasible, what I do know is that g i of x is less than equal to 0, g i of x star is also less than equal to 0, but they are both less than equal to 0 that does not mean that the difference is the is also less than equal to 0 or greater than equal to 0, nothing can be said about the difference from here right.

So, we need a little bit more information and that little piece of additional piece of information actually comes from complementary slackness. So, we already, we also have complementary slackness. So, far we have used the fact that x star is feasible and that the gradient of the Lagrangian is equal to 0, we will now use complementary slackness.

So, complementary slackness so, by complementary slackness; by complementary slackness, if g i of x star is less than 0, then the corresponding Lagrange multiplier is equal to 0. So, consequently, if I look at; if I look at this summation here, this summation actually involves only those lambda i's for which g i of x star is exactly equal to 0. So, I am going to rewrite the summation now.

So, that gives me f of x is greater than equal to f of x star minus sum i in A of x star lambda i gradient of g i at x star transpose x minus x star. Now, A of x star remember is the those indices, those constraints which are active ok. Now, go back to; now, if a now go back to this equation here that we wrote out let me put this in a box, let me wrote we wrote this out. Go back to this green boxed equation and write now, look at this equation in the case when x star is actually a point for which g i of x star is active.

So, look at right look at this equation for those constraints that are active which means g i of x star must be equal to 0 for such point, for such constraints. So, if g i of x star is equal to 0, then this term actually disappears; then this term disappears, this term is gone ok.

So, if g i of x star is equal to 0, this term is gone and then, what I am left with is therefore, gradient ok for i in A of x star I have gradient of g i evaluated at x star transpose x minus x

star less than equal to g i of x and what is g i of x? x being feasible ensures that g i of x is less than equal to 0. So, this is indeed less than equal to 0.

Now, lambda i's are greater than equal to 0, there is a minus sign outside and then, g i of x x star transpose x minus x star is less than equal to 0, what is put together means that f of x is greater than equal to f of x star plus something that is positive which means that f of x is greater than equal to f of x star.

Now, what did we do? We said let x star be as above and x be any other feasible point, it can be any other feasible point and from there, we got that f of x is greater than equal to f of x star. So, x star was my point that satisfied the KT conditions, x was any other feasible point, and we got that this inequality must hold. So, consequently, it has to be that x star is a global minimum.

So, if we are able to verify KKT conditions, we automatically get that the point that satisfies the KKT conditions is in fact, a global minimum of your optimization problem. So, this is; so, this is how you reverse the implications that have led us to KKT conditions.

You use an optimize you work with if you are working with an optimization problem that is convex, all you need to do is simply check if the KKT conditions hold and then, you get that the point that the point at which the KKT conditions hold are is a solution of your optimization problem ok.

So, with this I will end today's lecture, we can then I will tell you this the same fact in a little bit more abstraction next and I will also tell you about convex optimization, duality in convex optimization in the next lecture ok.