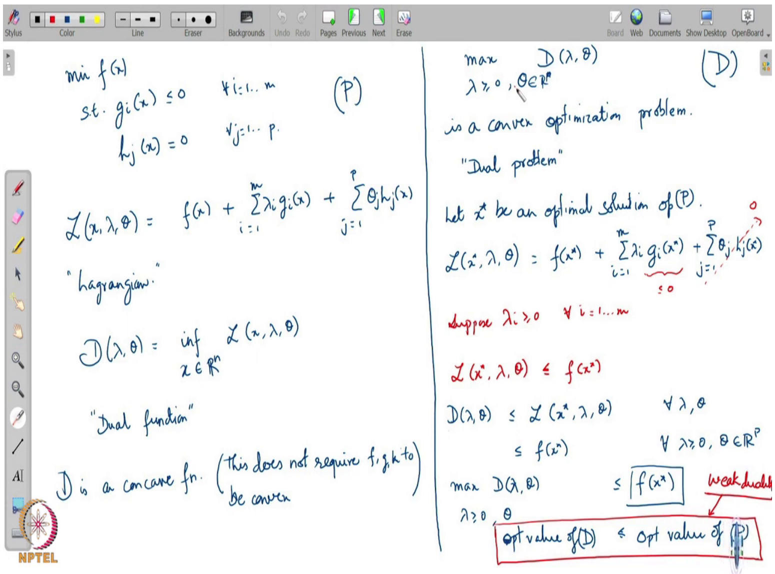
Optimization from Fundamentals
Prof. Ankur Kulkarni
Department of Systems and Control Engineering
Indian Institute of Technology, Bombay

Lecture - 17B
Weak duality in convex optimization and Fenchel dual

(Refer Slide Time: 00:22)



So, where does weak duality come from? So, suppose now let us say let x star be so I will call this problem P ok. Let x star be an optimal solution of P and optimal solution of let x star be an optimal solution of P. And now let us look at lambda of x star comma this comma this ok.

So, this is equal to f of x star plus all this, i going from 1 to m for the inequality constraints and j going from 1 to P for the equality constraints. Now what is it that you can observe here? So, if you look at this term here, these 2 terms what can one say about these? Well x star is an optimal solution of P ok. If it is an optimal solution of P then it has to at the very least it has to be feasible.

Which means that h j of x star has to be equal to 0, which means this term is actually 0 ok. All these guys this term here is actually 0 ok, alright. What about this term here? What about g i of x star? Well g i of x star being again x star being feasible, it means that this guy is less than equal to 0, alright. Now if this guy is less than equal to 0, then so we have that then that f of x star is greater than equal to lambda i times something that is less than equal to 0.

Now, I have not yet said anything about lambda i itself ok. I have not told you that lambda i should be greater than equal to 0 or anything like that. We did put lambda i greater than equal to 0 in this optimization here when we were defining the dual problem. So, let us go ahead with that.

Let us so suppose we restrict lambda i to be greater than equal to 0 for all i equal to 1 to m. Then in that case what would we get? Then we would get that this here ok, is less than equal to f of x star, alright, you would get that this is actually less than so this quantity here is less than equal to f of x star. Why? Because this is less than equal to 0 this is less than equal because this that these the g is are actually less than equal to 0 and these terms have disappeared ok.

So, let us play around with this in a different sort of way. So, suppose so let suppose x star is the optimal solution of P, alright. Now if and this is my; this is my this is the value of my Lagrangian. Now remember D of remember that, D of lambda comma theta it is actually the infimum of the Lagrangian. So, D of lambda comma theta is therefore, always greater than equal to L of x star comma lambda comma theta, alright.

So, D of lambda comma theta is greater than equal to L of x star comma lambda comma theta right. So, if this is sorry my mistake here. So, D of lambda comma; so, D of lambda comma theta is the infimum of the Lagrangian. So, if it is in its the infimum of the Lagrangian it follows that D of lambda comma theta is always less than equal to L of x star comma lambda comma theta.

And this holds for all lambda; for all lambda and for all theta. This is just I do not even need to restrict my sign here. If this holds for all lambda and all theta right. Now, what does this mean? And this it is this quantity L of lambda x star comma lambda comma theta was itself less than equal to f of x star. So, for and this and in getting to this inequality we needed that lambda is greater than equal to 0 and theta is in R P.

So, when lambda is greater than equal to 0 and theta and for any theta we all we have that D of lambda comma theta is less than equal to f of x star. So, this is true for all lambda greater than equal to 0 and all theta and all vectors theta. What does this mean? You look back here then again at this optimization. What is this optimization? Well it is a maximization of D of lambda comma theta over precisely the lambdas and thetas that are mentioned here.

So, what does this mean? It means then that the maximization of D of lambda comma theta, over lambda greater than equal to 0 comma any and any theta also has value less than equal to f of x star. And now what is this f of x star here? Remember x star, x star was an optimal solution of P right. It was an optimal solution of P. So, this is in fact the optimal value.

So, this here is an optimal value of P; optimal value of P. And what is the one on the left-hand side? Let us call this problem D the one on the left hand-side is simply the optimal value of D. And what have we got here? Then we have got that the optimal value of the dual is less than equal to the optimal value of the primal. So, this statement here is nothing, but the statement of weak duality.

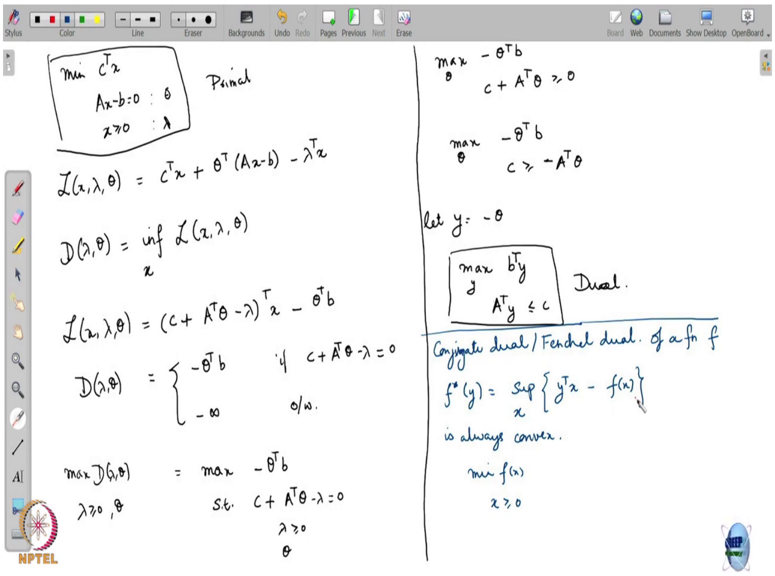
So, what does this mean? To summarize you can take any optimization problem like this you write it is Lagrangian. Lagrangian is formed by taking a linear of the objective and a linear combination of its of the constraints. Then for the constraints that are inequality constraints you restrict the multipliers the Lagrange multipliers here to be greater than equal to 0. For the constraints that are equality constraints you do not need to have any such restriction.

Then you look at the least possible value of the Lagrangian over the entire space, define that as D of lambda comma theta and then maximize that D of lambda comma theta over the Lagrange multipliers. The way you have restricted the Lagrange multiplies. Lambda to be greater than equal to 0 and theta n theta can be anything. In that case what and then what do we get?

We get that the optimal value of this is what we call the dual problem. You call this is the dual problem, you call this is the primal problem ok. The optimum what we get is weak duality. So, the optimal value of the dual problem is always less than equal to the optimal value of the primal alright ok.

So, now I will show you now that actually what you found calculated as the dual of a linear program in fact appears as a special case of this. So, that is not very hard to see. So, let us just go through this.

(Refer Slide Time: 09:19)



So, just look at consider our linear program in standard form C transpose x Ax equal to b and x greater than equal to 0. Now, I am going to create a Lagrangian of this lambda of L of x lambda comma theta. Now let us be careful here lambda corresponds to the inequality constraints. So, that is what it is going to so lambda is going to multiply my x and theta corresponds to the equality constraint.

So, theta is going to multiply my Ax minus Ax equal to b and since we want it in this in the form that we had for the optimization problem in the previous page. So, what I will write I will write this as Ax minus b equal to 0 ok. So, this is now so my Lagrangian therefore is C transpose x, C transpose x plus theta transpose Ax minus b.

Now if you go back here I wrote this problem with inequality constraints only since now, but now I am going to allow for equality constraints here in sorry I am going to allow sorry I wrote this problem with less than equal to constraints here the. So, if you go back to this problem I this problem has been written with less than equal to type constraints whereas, here my constraints are greater than equal to type of constraint.

So, I can effectively just multiply both sides by minus 1 and that would flip the direction of the inequality of the constraint. Alternatively I can what I need to do is just compensate for that in the definition of my Lagrangian itself. So, this is now my Lagrangian.

Now, let me write the dual function. The dual function which is D of lambda comma theta that is the infimum of the Lagrangian over the entire space. Now if you look at the Lagrangian function as a function of x. If you look at this as a function of x this is clearly a linear function in x ok. For each fixed lambda and theta this is a linear function in x and what you are doing here is you are taking the infimum of this linear function over this entire space.

Now, a linear function if you minimize this over in an unconstrained without any constraints, then you would get and you the optimal value is going to be minus infinity. Except in the case when the coefficients of the linear function are actually 0. The coefficients in if the coefficients involved are 0 then the linear function would evaluate to something that is just a constant.

So, now to do that to evaluate this more clearly let me let us just put gather together the coefficients of x. So, let us write this L of x comma lambda comma theta in this sort of way C minus A transpose theta minus lambda. The whole thing transpose or sorry the C plus this the whole thing transpose x then there is a and then I am left with a minus theta transpose b ok, alright.

So, now if I then if I take the; if I take the; if I take the infimum of the Lagrangian then that tells me that D of lambda comma theta should be equal to one of these. So, it is equal to either minus theta transpose b if C plus A transpose theta minus lambda is exactly equal to 0 and otherwise it is minus infinity.

So, whenever this is not equal to 0 you can choose a suitable x to drive the value down to minus infinity, alright. So, it is equal to some real value which is minus theta transpose b for x for those for so long as theta and lambda satisfy this equation otherwise, it is equal to minus infinity right.

So, now if I am looking if I now look to maximize D of lambda comma theta. Remember now I need to do this over lambda greater than equal to 0 and over all theta then what is where would my maximum be attained? Well my maximum cannot be minus infinity; obviously, it since it is a maximum.

So, my maximum is going to be attained over this this region, what do you mean by this? What do I mean by this region? I am looking for the maximum over these over the lambda comma theta such that they satisfy this right. So, effectively my the maximum is going to be equal this is going to be equal to maximize minus of theta transpose b, subject to C plus A transpose theta minus lambda equal to 0 lambda greater than equal to 0 and any theta.

Now, if you play around with this a little bit what do you realize? If you play around this with a if you play around this; play around this with a little bit you realize that well my lambda does not appear in the objective at all. I can absorb this the fact that lambda is greater than equal to 0 and there is a minus lambda here then it is simply that the lambda is simply appearing here as a slack variable.

So, effectively this constraint here can be written in this form. That I can simply write this as maximum maximizing minus theta transpose b C plus A transpose theta greater than equal to 0 and I am my theta is unrestricted. So, my lambda can plays no role lambda can be removed from this by just observing that these 2 inequal these 2 equations here they are effectively saying C plus A transpose theta equals lambda and lambda is greater than equal to 0 right.

So, its maximize maximizing this bit. So, let I can simplify this even further and write this more neatly. So, I can say maximize minus of theta transpose b C so I will write this in the in a following way I can C is greater than equal to negative of A transpose theta ok. So, now and I am maximizing over theta.

Now notice how something that we can do. Since theta does not is sign is has no sign constraints maximizing this particular thing this particular minus theta transpose b is subject to C greater than equal to minus A transpose theta. This can be this since theta has no sign constraints I can absorb I can just replace theta by minus theta.

And the optimization problem would remain optimal value would remain should remain the same or alternatively I take this minus sign minus I just define theta dash as minus theta or define y as minus theta here. If I just let y equal to minus theta then what I am what I have is our familiar form of the dual which is b is maximize b transpose y subject to A transpose y less than equal to C ok.

I am not multi remember I am not multiplying anything by negative sign I am just changing my notation. I am doing a change of variables. I am expressing y as my I am expressing minus theta as y and then substituting and that gives me this and this is actually nothing, but the familiar dual problem of this particular problem.

So, this was our primal and this is what we had learnt as the dual ok. So, if you work with the Lagrangian and you follow this follow the routine that I mentioned in the on the previous slide, in fact you get back the dual ok. So, what this does is this way of this entire whatever is there here right on the that is mentioned here which is that you define the Lagrangian. You then define the dual function by taking the infimum of the Lagrangian over the whole space.

And then take the then maximize the dual function subject to with constraints on the Lagrange multipliers. If you do all of that actually gives you if you do that for a linear program that gives you back the dual that we had defined earlier. So, this way of defining the dual is now is a way of is basically generalizing the duality formulation of for linear programming to problems that are potentially non-linear ok.

So, this is our going to be our vehicle for analyzing the for talking of duality for convex optimization ok. I should also tell you that there is another; there is another connection here let me mention that. So, there is another the word dual as is one that is being used by multiple people in different senses.

So, there is another dual what is which is called the conjugate dual or conjugate dual or what is also called the Fenchel dual. Thanks to Werner Fenchel ok. So, what is this dual? This dual is of the conjugate dual of a function f, ok f is denoted in of a function f it is denoted as f star. So, f star of y is defined in this way it is defined as the supremum over all x of supremum over all x of y transpose x minus f x ok.

So, you what you do is you take you subtract from f subtract or rather subtract f from a linear function whose slope is the parameter that you control the slope here is y. So, you subtract f from this linear function and you look at the maximum value of that you can get with a certain slope.

So, what is the maximum departure of this function from a linear function and study that as a function of the slope? That if you look at that as a function of the slope that quantity is what is called gives you is what is called the dual function. Now the dual function just like your just like D is has the D was always concave then this f star is always convex, right.

And so now what is the connection between D and f star? Well there is a connection in the following sense that you can see that this here as a what you are; what you are taking the supremum of has a resemblance to the Lagrangian in some way has a resemblance to the Lagrangian.

So, it is inside it is sort of closely related to the to an optimization to a certain type of optimization problem and that optimization problem is basically you can say consider you can consider the optimization problem where you are minimizing f the if you so if you are minimizing f subject to say x greater than equal to 0, alright.

So, in that case you the kind of the quantity you would encounter would end up becoming something like this. But this is just for you to mention for you to know there is no the this is another notion of the dual and it should not be confusing this with the Lagrange; with the Lagrange dual alright. So, what we will be working with is the Lagrange dual.